UNIVERSITÀ DEGLI STUDI DELLA TUSCIA

**Università degli Studi della Tuscia di Viterbo**

**Dipartimento per la INNOVAZIONE NEI SISTEMI BIOLOGICI, AGROALIMENTARI E FORESTALI**

**Corso di Dottorato di Ricerca in**

**SCIENZE, TECNOLOGIE E BIOTECNOLOGIE PER LA SOSTENIBILITÀ - XXXIV**
**Curriculum Ecologia Forestale e Tecnologie Ambientali**

# HARMONIZATION OF DESIGN-BASED MAPPING FOR SPATIAL POPULATIONS

(SSD SECS-S/01)


**Tesi di dottorato di:**
Dott.ssa Agnese Marcelli

Agnese Marcelli

**Coordinatore del corso**
Prof. Andrea Vannini

**Tutore**
Prof. Lorenzo Fattorini

**Co-tutore**
Prof. Piermaria Corona

Piermaria Corona


**A.A. 2020/21**

## Short Abstract

The mapping of a survey variable throughout a continuum or in finite populations of areas or points is usually performed from a model-dependent perspective. Nevertheless, when a sample of points/units is selected by a probabilistic sampling scheme, the complex task of modelling can be avoided by using the inverse distance weighting interpolator and deriving the properties of maps in a design-based perspective. Conditions ensuring consistency of maps can be derived mainly on the basis of some assumptions about the pattern of the survey variable throughout the study region as well as from the feature of the sampling scheme adopted to select points/units. However, in a design-based setting, the totals and averages of the survey variable for a set of domains partitioning the study region are commonly estimated by the well-known Horvitz-Thompson estimator or related estimators. That necessarily gives rise to different estimates with respect to those achieved from the resulting maps by the interpolated values within domains. In order to obtain non-discrepant results, a harmonization of maps is here suggested, in such a way that the resulting totals and averages arising from maps coincide with those achieved by traditional estimation. The ability of the harmonization procedure to maintain consistency of maps is theoretically argued and checked by a simulation study performed on some real and artificial populations. An application of the procedure for harmonizing maps is also reported.

**Key words:** spatial maps; inverse distance weighting interpolation; Horvitz-Thompson estimator; difference estimator; simulation study; case study.

# Extended Abstract

A methodology for the harmonization of design-based spatial maps is described in order to match estimates of totals and averages of forest attributes arising from the estimated maps and the traditional design-based Horvitz-Thompson criterion. The ability of the harmonization procedure to maintain consistency of maps is theoretically argued and checked by a simulation study performed on some real and artificial populations. A case study is also presented.

## 1. Introduction

The knowledge of the spatial pattern of forest attributes is fundamental for making forest management decisions. Commonly, the available resources render impractical the complete census of the population under study, which can be a continuum of points, a finite collection of spatial units or a finite collection of units spread throughout a study region. Accordingly, the distribution of the interest attribute constitutes an unknown parameter and sampling procedures are needed for its estimation. Spatial maps estimation is traditionally approached in a model-based framework. As an alternative approach, Fattorini et al. (2018a, 2018b and 2019) proposed mapping continuous populations, finite populations of spatial units and finite populations of points in a design-based framework, avoiding the complex task of modelling the surfaces to be mapped. However, pure design-based methods cannot be used to produce wall-to-wall maps because of the impossibility to estimate non-sampled values without making any assumption. Therefore, Fattorini et al. (2018a, 2018b and 2019) proposed to work in a model-assisted framework in order to fill the lack of information. The authors considered as assisting model the well-known Tobler's first law of geography (Tobler, 1970). Accordingly, they proposed the adoption of the so-called inverse distance weighting interpolator (IDW) (e.g., Henley, 1981) and determined the conditions ensuring design-based unbiasedness and consistency. IDW maps can be exploited for the estimation of totals and averages of forest attributes for a set of domains partitioning the study region, which is traditionally performed adopting the Horvitz-Thompson (HT) estimator or some modifications able to exploit auxiliary information, when available (e.g., Särndal et al., 1992, chapter 6). In this perspective, an unresolved problem arises: obviously, the two resulting sets of estimates will differ. Therefore, the aim of this work is to harmonize IDW maps in such a way that totals and averages of single estimates within domains will match the traditional design-based estimates, avoiding unsuitable discrepancies in final results.

## 2. Materials and methods

Consider a study region $A$ and let $f$ be a bounded measurable function related to the values of a survey variable $Y$ and defined on a subset $B$ of $A$. Depending on the features of the spatial populations, the IDW mapping of the interest attribute can be performed in the following three settings.

*Continuous populations*: $B$ coincides with $A$ and $f(\mathbf{p})$ is the value or the density of $Y$ at $\mathbf{p} \in B$. Therefore, mapping necessitates the knowledge of $f(\mathbf{p})$ for each $\mathbf{p} \in B$. Let $\mathbf{P}_1, \dots, \mathbf{P}_N$ be $n$ locations selected by means of a probabilistic sampling scheme. Then, in accordance with Fattorini *et al*. (2018a), the IDW interpolator of $f(\mathbf{p})$ is

$$\hat{f}(\mathbf{p}) = \sum_{i=1}^{n} w_i(\mathbf{p}) f(\mathbf{P}_i), \mathbf{p} \in B \tag{1}$$

where $w_i$s are weights decreasing with distances to the location under estimation.

*Finite populations of areas*: $A$ is partitioned into a finite population $U$ of $N$ spatial units of extents $\lambda_1, \dots, \lambda_N$. $B$ is the set of centroids of the areas, $y_j$ is the amount of $Y$ within unit $j$ and $f_j = y_j/\lambda_j$ is its density. Therefore, mapping necessitates the knowledge of $y_j$ or $f_j$ for each $j \in U$. If $S$ denotes the set of labels identifying the selected areas, then, in accordance with Fattorini *et al*. (2018b), the IDW interpolator of $f_j$ is

$$\hat{f}_j = Z_j f_j + \left(1 - Z_j\right) \sum_{i \in S} w_{ij} f_i, j \in U \tag{2}$$

where $w_{ij}$s are weights decreasing with distances to the area under estimation and $Z_j$ is the random variable equal to 1 if $j \in S$ and equal to 0 otherwise.

*Finite populations of marked points:* a finite population $U$ of $N$ points is scattered onto $A$, $B$ is the set of point locations and $y_j = f(\mathbf{p}_j)$ is the value of the survey variable for the point $j$ (*mark*). Therefore, mapping requires the knowledge of $y_j$ for each $j \in U$. If $S$ denotes the set of labels identifying the selected points, then, in accordance with Fattorini *et al*. (2019), the IDW interpolator of $y_j$ is

$$\hat{y}_j = Z_j y_j + \left(1 - Z_j\right) \sum_{i \in S} w_{ij} y_i \, , j \in U \tag{3}$$

where $w_{ij}$s are weights decreasing with distances to the point under estimation and $Z_j$ is defined as in (2).

The asymptotic properties of (1), (2) and (3) are derived in Fattorini *et al.* (2018a), Fattorini *et al.* (2018b) and Fattorini *et al.* (2019), respectively. Without going into technical details, whatever the population setting, the IDW interpolators are design-based unbiased and consistent under reasonable conditions that seem to match most real situations encountered in environmental surveys. Commonly, once the $n$ locations, areas, or points are selected, the population total of the survey variable $Y$ is estimated by means of the Horvitz-Thompson (HT) criterion or related criteria able to exploit auxiliary information. However, total estimates naturally arise from the resulting IDW maps as the integral, in the continuous case, or the sum, in the discrete cases, of the interpolated values, thus achieving total estimates that invariably differ from those achieved by traditional, HT-based techniques. To eliminate these unsuitable discrepancies, the matching of the two total estimates can be performed by rescaling the IDW interpolation for each location, area, or point. Therefore, for continuous populations, the rescaled maps are given by

$$\hat{f}^*(\mathbf{p}) = \frac{\hat{T}}{\hat{T}_{IDW}} \hat{f}(\mathbf{p}), \mathbf{p} \in B \tag{4}$$

while, for both finite populations of areas and finite populations of points, the rescaled maps are given by

$$\hat{y}_j^* = \frac{\hat{T}}{\hat{T}_{IDW}} \hat{y}_j, j \in U \tag{5}$$

respectively. In all cases, $\hat{T}$ denotes the total estimate obtained by means of commonly adopted HT-based techniques and $\hat{T}_{IDW}$ denotes the total estimate obtained from the IDW map. Consistency of (4) and (5) is derived under the most common sampling scheme adopted in environmental surveys.

Total estimates of an interest attribute can be required also for $D$ subpopulations partitioning the population, usually referred to as domains (e.g., Särndal *et al.*, 1992, section 10.3). For continuous populations, the rescaled map ensuring the matching of total estimates for each domain, as well as the matching of the overall total estimates, is given by

$$\hat{f}^*(\mathbf{p}) = \frac{\hat{T}_{(d)}}{\hat{T}_{IDW(d)}} \hat{f}(\mathbf{p}), \mathbf{p} \in B_d, d = 1, \dots, D \tag{6}$$

where $B_1, \dots, B_D$ are the $D$ subsets partitioning $B$.

For finite populations of areas and finite populations of points, the rescaled map ensuring the matching of total estimates for each domain, as well as the matching of the overall total estimates, is given by

$$\hat{y}_j^* = \frac{\hat{T}_{(d)}}{\hat{T}_{IDW(d)}} \hat{y}_j, j \in U_d, d = 1, \dots, D \tag{7}$$

where $U_1, \dots, U_D$ are the $D$ subsets of areas or points, respectively, partitioning $U$.

## 3. Results and discussion

### 3.1 Simulation studies

For each population setting, a simulation study was performed to investigate the performance of the procedure described in the section above.

*Continuous populations*:

- Survey region: square-shaped region of 129,600 ha in North-Western Tuscany.
- Survey variable: precipitations (mm) occurred between 3rd January 2021 and 3rd February 2021.
- Sampling scheme: uniform random sampling (URS), tessellation stratified sampling (TSS), systematic grid sampling (SGS).
- Sample size: $n = 16, 36, 64, 100$ locations.

*Finite populations of areas*:

- Survey region: rectangular region of 212 ha in Calabria.
- Populations: three populations were constructed by partitioning the survey region into grids of $N = 250, 1{,}000, 4{,}000$ areas.
- Survey variable: densities of growing stock volumes ($m^3$/ha).
- Sampling: simple random sampling without replacement (SRSWOR), one per stratum stratified sampling (OPSS), systematic sampling (SYS).
- Sample size: $n = N/10$ areas.

*Finite populations of marked points:*

- Survey region: forested region of 4.8 ha in Val di Sella (North-Eastern Italy).
- Populations: three populations of $N = 219, 365, 608$ trees (diameter at breast height $> 17.5$ cm.
- Survey variable: wood volumes (m$^3$).
- Sampling: 3P sampling.
- Expected sample size: $n = N/10$ trees.

To set an example, Figure 1 reports the map of the continuous population considered in the simulation.

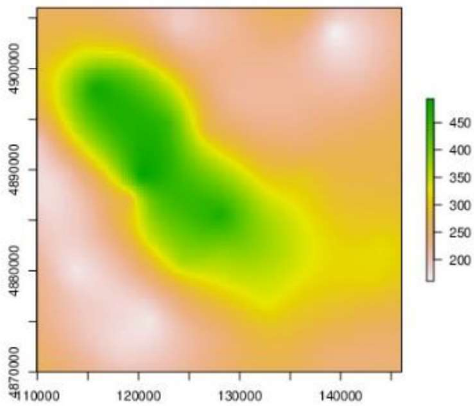


**Figure 1.** Map of precipitations (mm) occurred between 3$^{rd}$ January 2021 and 3$^{rd}$ February 2021 in a square-shaped region in North-Western Tuscany.

For each combination of population, sampling scheme and sample size, sampling was replicated $R = 10{,}000$ times. At each simulation, run $r = 1, \dots, R$, the harmonized maps were achieved by means of equation (4) for the continuous population and equation (5) for finite populations of areas and points. Additionally, each population was partitioned into $D = 2, 4, 8$ domains. For each combination of population, sampling scheme, sample size and domain partition, sampling was replicated $R = 10{,}000$ times. At each simulation run, the IDW harmonization by domains was performed by means of equation (6) for the continuous population and equation (7) for finite populations of areas and points. The absolute bias (AB) and root mean squared error were calculated for IDW maps as follow:

Continuous populations: ($k, l$: nodes of the grid for interpolation)

$$AB(\mathbf{p}_{k,l}) = \left|\frac{1}{R}\sum_{r=1}^{R} \hat{f}_r(\mathbf{p}_{k,l}) - f(\mathbf{p}_{k,l})\right| \qquad RMSE(\mathbf{p}_{k,l}) = \left[\frac{1}{R}\sum_{r=1}^{R}\left[\hat{f}_r(\mathbf{p}_{k,l}) - f(\mathbf{p}_{k,l})\right]^2\right]^{1/2}$$

Finite populations: (areas or points)

$$AB_j = \left|\frac{1}{R}\sum_{r=1}^{R} \hat{y}_{j,r} - y_j\right| \qquad RMSE_j = \left[\frac{1}{R}\sum_{r=1}^{R}\left[\hat{y}_{j,r} - y_j\right]^2\right]^{1/2}, j \in U$$

The same performance indicators (AB* and RMSE*) for harmonized maps were computed as in expressions above by substituting $\hat{f}_r\left(\mathbf{p}_{k,l}\right)$ with $\hat{f}_r^*\left(\mathbf{p}_{k,l}\right)$ and $\hat{y}_{j,r}$ with $\hat{y}_{j,r}^*$, respectively. The spatial patterns of performance indicators were reconstructed both for IDW and harmonized maps (see e.g., Figure 2). Finally, the level of matching between HT-based techniques and IDW total estimates, as well as between IDW and harmonized maps, was quantified by the following correction factor:

$$RATIO_r = \hat{T}_r / \hat{T}_{IDW,r} \;\; r = 1, \ldots, R$$

Similarly, performance indicators and correction factors were obtained for the harmonized maps, considering a partition of the populations into $D = 2, 4, 8$ domains.

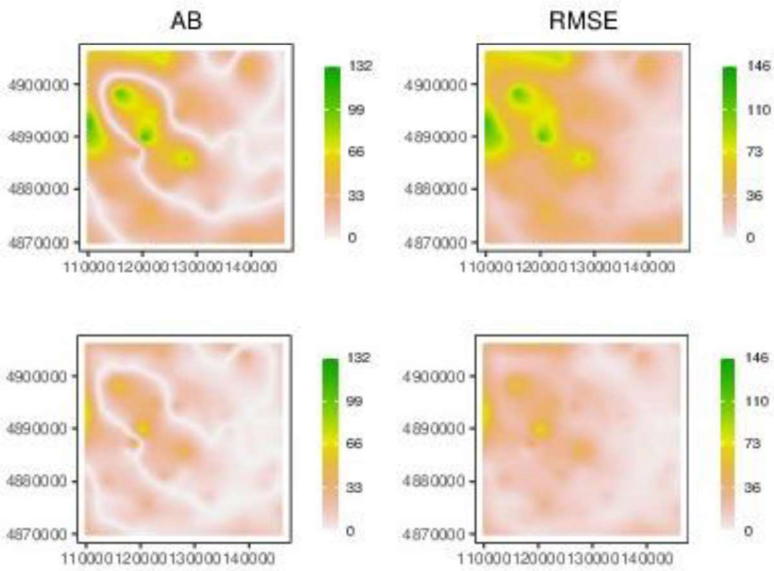


**Figure 2:** Maps of AB and RMSE of the IDW interpolator of precipitations (mm) in a square-shaped region located in Tuscany, under SGS and sample size $n = 16, 64$.

For all the populations and sampling schemes, simulation results show that harmonized and IDW maps are comparable in terms of AB and RMSE, suggesting that harmonization can be performed without relevant loss in precision. Additionally, in all cases, the correction factor quickly approaches 1 as the population and/or sample size increase. For the sake of brevity, Table 1 reports only simulation results on the continuous population, for each sampling scheme and sample size. As to harmonization by domains, RMSE increases as the number of domain increases for continuous and finite population of areas. This result was expected, since with the increase of number of domains, the number of selected locations or areas within becomes lower and lower. This tendency is less evident in populations of marked points, presumably because of the efficiency of the difference estimator adopted to estimate totals under good predictions.

**Table 1.** Minima, maxima and averages of AB, AB*, RMSE, RMSE* and RATIO achieved for precipitations in a square-shaped region located in North-Western Tuscany, for each combination of sampling scheme and sample size.

| Scheme | $n$ | AB*(AB) | | | RMSE*(RMSE) | | | RATIO | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max | min | mean | max |
| URS | 16 | 0.00 (0.00) | 30.62 (31.18) | 137.06 (130.51) | 16.81 (13.51) | 49.89 (49.55) | 150.69 (144.91) | 0.87 | 0.98 | 1.08 |
| | 36 | 0.00 (0.00) | 21.71 (21.89) | 113.41 (109.37) | 10.83 (7.82) | 35.97 (34.87) | 126.00 (121.83) | 0.91 | 0.99 | 1.08 |
| | 64 | 0.00 (0.00) | 16.80 (16.89) | 97.01 (94.44) | 7.92 (4.99) | 28.11 (26.89) | 108.76 (105.88) | 0.92 | 0.99 | 1.07 |
| | 100 | 0.00 (0.00) | 13.55 (13.59) | 85.87 (84.07) | 6.18 (3.30) | 22.88 (21.74) | 96.70 (94.49) | 0.94 | 1.00 | 1.06 |
| TSS | 16 | 0.00 (0.00) | 26.14 (26.28) | 126.81 (122.98) | 7.77 (7.27) | 40.50 (40.38) | 138.70 (135.40) | 0.94 | 1.00 | 1.04 |
| | 36 | 0.00 (0.00) | 17.76 (17.81) | 101.17 (99.00) | 4.04 (3.60) | 27.29 (27.11) | 111.79 (109.77) | 0.97 | 0.99 | 1.02 |
| | 64 | 0.00 (0.00) | 13.19 (13.22) | 81.06 (79.66) | 2.52 (1.76) | 20.19 (20.07) | 88.29 (87.05) | 0.97 | 1.00 | 1.02 |
| | 100 | 0.00 (0.00) | 10.52 (10.53) | 77.41 (76.52) | 1.54 (1.07) | 16.09 (16.00) | 85.94 (85.09) | 0.98 | 1.00 | 1.01 |
| SGS | 16 | 0.00 (0.00) | 23.90 (24.11) | 118.38 (114.81) | 3.09 (3.95) | 35.19 (35.75) | 126.99 (125.80) | 0.96 | 0.98 | 1.00 |
| | 36 | 0.00 (0.00) | 15.57 (15.60) | 95.65 (93.48) | 1.21 (1.22) | 22.92 (23.01) | 104.05 (104.03) | 0.98 | 0.99 | 1.00 |
| | 64 | 0.00 (0.00) | 11.41 (11.42) | 80.53 (78.97) | 1.04 (0.81) | 16.80 (16.81) | 87.63 (87.27) | 0.99 | 1.00 | 1.00 |
| | 100 | 0.00 (0.00) | 8.99 (8.99) | 70.17 (69.18) | 0.91 (0.66) | 13.25 (13.24) | 76.67 (75.76) | 0.99 | 1.00 | 1.00 |

### 3.2 Real case application

The described procedure was adopted to provide the harmonized map of densities of living wood volume in Rincine (Central Italy), a forest complex of approximately 290 ha. The survey area was tessellated by a grid of 5,449 quadrats of size $23 \times 23$ m$^2$. For sampling purposes, population was partitioned into 50 blocks of contiguous cells. After the partition, 340 cells were discarded from the population since they were located outside the forest, in artificial areas or waters. Therefore, the resulting population to be sampled was constituted by $N = 5{,}109$ quadrats partitioned into 50 blocks of different sizes. One quadrat was randomly selected within each block by means of OPSS, for a total sample of $n = 50$ cells. The total living wood volume of the sampled cells was reckoned by the allometric equations from the Italian National Forest Inventory (Tabacchi *et al.*, 2011) and its density was adopted to interpolate the total living wood volume density within each spatial unit of the population by means of the IDW interpolator (2). RMSE estimates were achieved by means of the asymptotically conservative estimator proposed by Fattorini *et al.* (2018b). Total living wood volume estimates achieved by the traditional HT estimator and by the IDW map were computed. Finally, harmonized estimates were achieved by means of equation (5) and the corresponding RMSEs estimates were computed adopting, *mutatis mutandis*, the estimator proposed by Fattorini *et al.* (2018b). Total living wood volume estimate achieved by means of the HT estimator was equal to 133,603 $m^3$, whereas the total estimate arising from the IDW map was equal to 135,054 $m^3$. The correction factor adopted to achieve the rescaled estimates resulted to be approximately 0.989. Since the correction factor is very close to one, harmonized estimates of wood volume density and the corresponding RMSEs estimates turn out to be quite similar to the non-harmonized ones. Figure 3 reports the harmonized map of living wood volume densities together with the map of the RMSEs estimates.

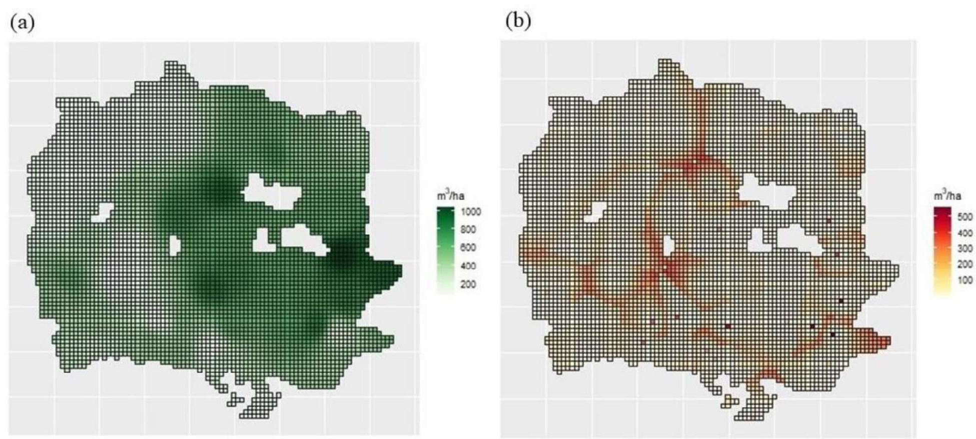


**Figure 3.** Harmonized map of living wood volume densities (a) and map of the corresponding RMSEs estimates (b).

## 4. Conclusions and future perspectives

The recent papers by Fattorini *et al.* (2018a, 2018b and 2019) proposed a novel approach for mapping continuous populations or finite populations of areas and points in a design-based framework. However, in the design-based approach, an unresolved problem arises. Since totals and averages throughout the whole survey region or within domains are traditionally achieved by the HT or related criteria, these estimates invariably differ from the estimates achieved by the resulting maps. For avoiding these unsuitable discrepancies, we propose to rescale the resulting maps in such a way that the estimates arising from maps will match the traditional estimates. We also prove the asymptotic convergence of the two maps. Since IDW and harmonized maps seem to result very similar, we may argue that harmonization is a useless procedure. However, simulation results showed that, in some situations, the correction factor may vary widely from one, especially when the estimation of domain totals is involved.

It should be noted that we neglected the problem of estimating the precision of the harmonized maps because it can be performed, *mutatis mutandis*, by the naïve procedure proposed by Fattorini *et al.* (2018a, 2018b, 2019) for the three types of populations.

### References


Fattorini L, Marcheselli M, Pisani C, Pratelli L (2018a) Design-based maps for continuous spatial populations. *Biometrika*. **105**:419-429.

Fattorini L, Marcheselli M, Pratelli L (2018b) Design-based maps for finite populations of spatial units. *J Am Stat Assoc*. **113**:686-697.

Fattorini L, Marcheselli M, Pisani C, Pratelli L (2019) Design-based mapping for finite population of marked points. *Electron J Stat*. **13**:2121-2149.

Henley S (1981) Nonparametric Geostatistics. London: Applied Science Publishers.

Särndal CE, Swensson B, Wretman J (1992) Model Assisted Survey Sampling. New York: Springer.

Tabacchi G, Di Cosmo L, Gasparini P (2011) Aboveground tree volume and phytomass prediction equations for forest species in Italy. *Eur J For Res*. 130:911-934.

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr*. **46**:234-240.

## Contents


**Chapter 1**

**Introduction ..................................................................................... 4**

**Chapter 2**

**State-of-the-art ..................................................................................... 7**

2.1 Model-based methods ..................................................................................... 7

2.1.1 Kriging ..................................................................................... 7

2.1.2 Locally weighted regression ..................................................................................... 9

2.1.3 Nearest neighbor techniques ..................................................................................... 10

2.2 Design-based methods ..................................................................................... 11

**Chapter 3**

**Materials and Methods ..................................................................................... 14**

3.1 Notation and setting ..................................................................................... 14

3.1.1 Continuous populations ..................................................................................... 14

3.1.2 Finite populations of areas ..................................................................................... 14

3.1.3 Finite populations of marked points ..................................................................................... 15

3.2 Harmonization with the overall population total ..................................................................................... 16

3.2.1 IDW harmonization for continuous populations ..................................................................................... 17

3.2.2 IDW harmonization for finite population of areas ..................................................................................... 19

3.2.3 IDW harmonization for finite populations of marked points ..................................................................................... 20

3.3 IDW harmonization by domains ..................................................................................... 22

**Chapter 4**

**Simulation Studies ..................................................................................... 24**

4.1 Populations and sampling ..................................................................................... 24

4.2 Simulation ..................................................................................... 31

4.3 Performance indicators ..................................................................................... 32

**Chapter 5**

**Case Study** .................................................. **34**

5.1 Forest complex of Rincine .................................................. 34

5.2 Sampling procedure and estimation .................................................. 35

**Chapter 6**

**Results and Discussion** .................................................. **37**

6.1 Simulation results .................................................. 37

6.2 Case study results .................................................. 71

**Chapter 7**

**Conclusions and Future Perspectives** .................................................. **72**

**Acknowledgements** .................................................. **74**

**References** .................................................. **75**

**List of papers** .................................................. **81**

# Chapter 1

# Introduction

Forests provide several ecosystem goods and services to human welfare, such as timber stock as well as non-timber forest products (Krieger, 2001; Holmgren and Persson, 2002; Köhl *et al.*, 2006), e.g., conservation of biodiversity, mitigation of climate change, carbon sequestration, water regulation, and soil protection. Decisions related to the maintenance and enhancement of the multiple functions provided by the forests need to be based on objective, reliable information (Corona, 2016). Forest inventories based on statistically sound estimation of forest attributes are regarded as effective tools for assessments of forest resources (Corona *et al.*, 2011). The earliest national inventories were developed in the 20$^{th}$ century to find out the amount of available wood. Besides timber production, no other benefits or uses from the forest were considered. For example, Sweden began a national forest inventory (NFI) in 1923 because the country feared the beginning of a wood shortage (Holmgren and Persson, 2002). Nonetheless, forest inventories gradually evolved towards multipurpose resource surveys (Lund, 1998; Corona and Marchetti, 2007; Tomppo *et al.*, 2010), including variables not directly related to timber assessment, wood volume growth and carbon-related issues, e.g., biodiversity attributes and non-traditional communities, e.g., urban forests and trees outside forests (Corona, 2016).

Spatial phenomena frequently need detailed information, especially in natural resources management. Thus, a crucial goal of multipurpose inventories is to perform spatially explicit estimation providing wall-to-wall maps depicting the spatial pattern of an interest attribute (Corona, 2010). Indeed, mapping is becoming a burning issue in forest surveys and inventories. For example, in the recent conference on "A century of national forest inventories – informing past, present and future," Sundvolden (NW), May 20–23, 2019 (https://nibio.pameldingssystem.no/nfi100years), the need for mapping has arisen as one of the major necessities in forest studies.

The population to be surveyed can be a continuum of points, a finite collection of spatial units, or a finite collection of units spread throughout a study region. In most cases, the available resources render impractical the complete census of these populations. Therefore, the survey variable is recorded only for a subset of locations/units, and an estimation criterion is adopted to estimate the values of the survey variable within non recorded locations/units and obtain wall-to-wall maps of the interest variable throughout the whole population.

Usually, mapping is approached in a model-based framework: locations/units where the variable is recorded are considered as fixed, while the population values are assumed to be outcomes of a super-population probability model (e.g., Cressie, 1993). The most common model-based techniques for spatial interpolation, such as the kriging (e.g., Cressie, 1993) and k-nearest neighbors (k-NN) (McRoberts *et al*., 2007), are described in Chapter 2. As an alternative approach, Fattorini *et al.* (2018a, 2018b and 2019) proposed mapping continuous populations, finite populations of spatial units, and finite populations of points in a design-based framework. In this scenario, population values are viewed as fixed constants, and the probability distribution of any sample statistic is determined from the uncertainty entailed by the probabilistic sampling scheme adopted for selecting spatial locations/units. In their works, the authors highlighted as any design-based mapping is challenging. Indeed, when estimating the values of the survey variable in a single spatial location or unit, either the spatial location/unit has been sampled and therefore there is no need for estimation, or the location/unit is unsampled and therefore we do not have any information to perform estimation. Consequently, even in a design-based framework, the use of an assisting model seems to be the only way to fill the lack of information. As a solution, Fattorini *et al.* (2018a, 2018b and 2019) considered as assisting model the well-known Tobler's first law of geography, i.e., locations/units close in space tend to have more similar values than locations/units that are far apart (Tobler, 1970). Accordingly, they proposed the adoption of the so-called inverse distance weighting interpolator (IDW) (e.g., Henley, 1981), and determined the conditions ensuring design-based unbiasedness and consistency. Some details on IDW interpolator are given in Chapter 3 together with details on the three different population settings (continuous populations, finite populations of spatial units and finite populations of marked points). Nevertheless, it should be acknowledged that frequently, as it occurs in forest inventories, estimation of totals and averages of interest attributes for a set of domains partitioning the study region is traditionally performed from a design-based perspective adopting the Horvitz-Thompson (HT) estimator or some modifications able to exploit auxiliary information (e.g., Särndal *et al.*, 1992, chapter 6). However, totals and averages estimates for these domains can be achieved also from the resulting maps by the IDW interpolations within domains. Obviously, the two resulting sets of estimates will differ. Therefore, the aim of this work is to harmonize IDW maps in such a way that totals and averages of single estimates within domains will match the traditional design-based estimates, avoiding unsuitable discrepancies in final results that may seem perplexing, especially in a reporting phase. Because averages are achieved as totals divided by the size of the study region in continuous populations or as totals divided by the population size in finite populations of spatial units and points, that are necessarily known for mapping, harmonization concerning totals or averages are equivalent from a statistical point of view. Therefore, this work will refer to harmonization with respect to totals only.

The thesis is organized as follow. Chapter 2 introduces the most common model-based techniques for spatial interpolation, together with the design-based method proposed by Fattorini et al. (2018a, 2018b and 2019). Details on IDW interpolator, population settings, and the procedure for harmonizing IDW maps are described in Chapter 3. Moreover, in order to check the capacity of the harmonization procedure to maintain consistency of the resulting maps, a simulation study is performed on a set of populations from artificial and real study regions and is presented in Chapter 4. An application of the procedure for harmonizing the map of living wood volume in a forest complex in eastern Tuscany (Central Italy) is illustrated in Chapter 5. Results of both simulation studies and the case study are discussed in Chapter 6. Lastly, conclusions and future perspectives are reported in Chapter 7.

# Chapter 2

# State-of-the-art

## 2.1 Model-based methods

Most statistical methodologies applied for wall-to-wall mapping rely on model-based inference. The crucial reason for this choice is the impossibility of estimating non-sampled values without any assumptions (Di Biase *et al.*, 2018). Model-based inference views the population values as the realization of a random process called super-population, making assumptions about the mechanism generating the super-population, and considers the sampled locations/units as fixed, i.e., purposively selected. When model assumptions hold, it is possible to achieve optimal predictors and their related model-based properties (Corona *et al.*, 2014).

We can distinguish three types of model-based methods: parametric, semi-parametric, and non-parametric. Parametric methods require to fully specify the distribution of the survey variables, semi-parametric methods are mixed approaches using fully specified distributions of the survey variables subsequently corrected by data-driven procedures, non-parametric methods do not require to fully specify the distribution of the survey variables.

Statistical literature presents several model-based techniques for forest mapping, therefore only the most popular methodologies applied in forest studies are reported in the following paragraphs. We chose kriging and its modifications among parametric methods, locally weighted regression among semi-parametric methods, nearest neighbor techniques among non-parametric methods.

These model-based methodologies involve severe assumptions, therefore results are effective in quite unrealistic settings.

### *2.1.1 Kriging*

Kriging is a minimum-mean-squared-error method of spatial prediction (Cressie 1993, p. 106), estimating the values of the survey variable at non-sampled locations as linear combinations of the values related to the sampled locations.

Kriging predictor relies on the assumption of intrinsic stationarity of the spatial process, i.e., the super-population is generated from a spatial process that involves expected values that do not depend on spatial locations and variances of increments only depending on the spatial gaps between two locations.

The mathematically simplest version of kriging predictor is the ordinary kriging. The main assumption of this methodology consists in the presence of a "pure" spatial autocorrelation trend, however, this condition rarely occurs in forestry inventory data. Therefore, ordinary kriging is seldom adopted in forest mapping. Mostly, it is found in literature for performing comparisons and checking the performance of different techniques. For example, Freeman and Moisen (2007) used ordinary kriging for improving the point prediction accuracy of the nationwide forest biomass map, but the results showed that neither the field biomass nor the residual biomass turned out to be good candidates.

A generalization of ordinary kriging is the universal kriging (Matheron, 1969). This technique is vastly applied in forest surveys and its main feature is that the mean is "driven" by auxiliary data known for the whole study region and in addition to the sample locations. Therefore, universal kriging presumes that expectations are linear combinations of some functions of spatial coordinates. When no auxiliary information is available, universal kriging reduces to ordinary kriging. Universal kriging was implemented by Lochhead *et al.* (2018) in order to provide wall-to-wall information on forest attributes. The authors found out that universal kriging was very accurate, but computational demanding when implemented for macroscale.

Another modification of kriging is regression kriging (Odeh *et al*., 1995), which combines a regression model with kriging, similarly to universal kriging. Indeed, many authors (e.g., Deutsch and Journel, 1992; Hengl *et al.*, 2003; Wackernagel, 2003) use the term "universal" when the deterministic trend is supposed to be a function of the coordinates only, whereas the term "regression" is used when the trend is supposed to be a function of some auxiliary variables (Di Biase *et al*., 2018). In regression kriging, trend and residuals are usually estimated separately and summed afterward (e.g., Ahmed and de Marsily, 1987). Specifically, the residuals are predicted by means of ordinary kriging and then added to the estimated spatial trend (Goovaerts, 1997).

Regression kriging was adopted by Brus *et al.* (2012) for performing a statistical mapping of forest tree species across Europe.

An additional variation of kriging exploiting auxiliary information is the cokriging. This technique assumes that both auxiliary and interest variables are outcomes of intrinsically stationary random processes. Hudak *et al.* (2002) compared cokriging with other kriging predictors. Cokriging results turned out to be the ones with the smallest bias, however, the vegetation patterns were poorly reproduced.

Meng *et al.* (2009) performed a comparison between ordinary kriging, universal kriging, regression kriging, and cokriging to map basal area, exploiting Landsat ETM+ images as auxiliary information in loblolly pine (*pinus taeda*) and slash pine (*pinus eliotii*) stands. Results showed that regression kriging proved to be the best geostatistical method for spatial prediction.

### *2.1.2 Locally weighted regression*

Locally weighted regression (LWR) (Cleveland, 1979) is one of the most popular semi-parametric methodologies for spatial interpolation. Under this technique, the data are supposed to be generated from a semi-parametric model: the survey variable is assumed to be a smooth function of the covariates, and errors are presumed to be normally distributed independent variables with zero-mean and constant variance, as in the most common regression frameworks. For any unsampled location, LWR estimates the survey variable exploiting the covariates of a neighborhood set of sampled locations that are close to the location under estimation. Practically speaking, the survey variable value of an unsampled location is predicted using only data that are "local" to that point. Additionally, each location of the neighborhood set is weighted according to its distance from the location under estimation, therefore higher weights correspond to closer locations. It is worth noting that the number of neighborhood locations divided by the overall number of sampled locations, usually referred to as span value, covers a role of great relevance in the prediction (Cleveland and Devlin, 1988). Indeed, a too large span produces an over-smoothed function that may not lead to a suitable fitting of the data. On the other hand, a too small span entails prediction affected by much noise (Corona *et al.*, 2014).

LWR turns out to be particularly suitable when the absence of stationarity in the relationship between interest and auxiliary variables reduces the efficiency of kriging predictors (e.g., Brunsdon *et al*., 1996; Foody 2003; Maselli *et al.*, 2014).
LWR has been often implemented for investigating spatial trends or obtaining insights on single tree growth trends. For example, LWR was adopted by Wang *et al.* (2005) for deriving a net primary production regression model for forest ecosystems in China, obtaining better results compared to the more classic least squares regression. Additionally, Zhang and Shi (2004) examined the spatial heterogeneity of multivariate relationships between tree growth and diameter at breast height. The authors assessed spatial variation by calibrating a multiple regression model fitted at each tree in a sample plot, adopting a distance function from the subject tree in order to weigh all neighboring trees.

### *2.1.3 Nearest neighbor techniques*

Non-parametric models represent an alternative to parametric models (e.g., universal or regression kriging) accounting for spatial heterogeneity. Nearest neighbor (NN) techniques are among the most popular non-parametric methodologies, especially in forest applications (McRoberts *et al.*, 2010b; Baffetta *et al.*, 2012). These techniques, introduced by Fix and Hodges (1951), were adopted for the first time in the Finnish NFI (Tomppo, 1991). Under NN methodologies, unsampled values are predicted as linear combinations of sampled observations that are nearest or most similar, in a space of auxiliary variables, to the value under prediction. Two relevant advantages of NN methods concern: i) the possibility to use them both for univariate and multivariate prediction and ii) no assumptions regarding the distribution of the variables (McRoberts *et al.*, 2010a).
The most general NN method is the k-NN technique, and it is likely the most utilized in forest operational cases (Corona *et al*., 2011). k-NN predicts the unsampled values by a weighted average of the k nearest values (concerning a fixed metric in the covariate space) among those included in the sample. Implementation of k-NN requires three primary selections: (i) a distance metric, (ii) a neighbor weighting scheme, and (iii) a value for k (McRoberts, 2012). As to distance metrics, very common criteria are Euclidean or Mahalanobis distances. Additionally, neighbor weights can be constant but usually they

are inversely proportional to their distance to the element under estimation. Finally, regarding the number of neighbors, the choice of a small k is commonly preferred. However, McRoberts (2012) remarked that the value of k may be selected according to other optimization criteria.

k-NN techniques are widely adopted in forest applications (e.g., Franco-Lopez *et al.*, 2001; Reese *et al.*, 2003; Chirici *et al.,* 2008; Ohmann *et al.*, 2011), nevertheless poor theoretical studies on the model-based properties of k-NN predictions have been carried out (Corona *et al.*, 2014). At least to our knowledge, the only theoretical effort has been attempted by McRoberts *et al.* (2007), however statistical properties are derived under severe assumptions, e.g., the realizations of the interest variable are considered as random variables with same expectation and variance and a parametric structure of their covariances (Corona *et al*., 2014).

Two variations of k-NN technique are most similar neighbor (MSN) (Moeur and Stage, 1995) and gradient nearest neighbor (GNN) (Ohmann and Gregory, 2002). Both techniques exploit only one neighbor to predict the value of the interest variable. MSN procedure chooses the most similar neighbor using a similarity function derived from canonical correlation analysis (Hotelling, 1936), whereas GNN procedure models the relationship between survey and auxiliary variables with direct gradient analysis (Gauch, 1982) using stepwise canonical correspondence analysis (ter Braak, 1986).

## 2.2 Design-based methods

Contrary to model-based inference, design-based inference views the population values as fixed, therefore no assumption is made about the process generating them, and sampled locations/units as random, i.e., the outcome of a random selection entailed by a probabilistic sampling scheme. Differences between model-based and design-based inference are well delineated in statistical literature (Smith, 1994, 2001; Gregoire, 1998; Thompson, 2002, chapter 10; Schreuder *et al*., 1993; Little, 2004). Thompson (2002) argued that "the advantage of design-based methods is that properties, such as design-based unbiasedness and consistency, do not depend on assumptions about the population itself". Additionally, Särndal *et al.* (1992) stated that the main appeal of 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, as mentioned in the Introduction (Chapter 1), pure design-based methods cannot be used for making wall-to-wall maps, because of the impossibility to predict unsampled values without assumptions. Specifically, either the spatial location/unit has been selected in the sample and thus the estimation is not needed, or the location/unit is not selected so that we do not have any information to perform the estimation. Therefore, as remarked by Fattorini *et al.* (2018b), the use of an assisting model to estimate unsampled values exploiting auxiliary information seems the sole way to fill the information vacancy. The use of assisting models is effective for estimating totals or averages, however, the same remark cannot be extended to the estimation of single population values. The use of assisting models allows achieving design-unbiasedness of estimators of population totals and averages, or approximate unbiasedness up to the first term of approximation, as well as exact or approximate design-based variance expressions and suitable variance estimators. Indeed, the assisting model is used to predict each population value, but the total of errors produced by the model is estimated from the sample and then adopted to correct the total or the average obtained from the predictions, according to the criterion provided by the difference estimator (Särndal *et al*. 1992, Chapter 6). On the contrary, when estimating single values, there is no way to correct the errors invariably provided by the model prediction. Thus, any model-assisted estimator of a single population value and the subsequent map are destined to be design-biased (Fattorini *et al.*, 2018b). Therefore, conditions ensuring at least design-based asymptotic unbiasedness and consistency are needed to produce reliable design-based model-assisted maps. In recent works, Fattorini *et al.* (2018a, 2018b, 2019) derived these conditions when dealing with continuous populations, finite populations of spatial units, and finite populations of points, respectively. Since information provided from space is always available while auxiliary variables related to the survey variable may not, the authors adopted as assisting model Tobler's first law of geography, asserting that location/units are more similar to nearby location/units than to those further apart (Tobler, 1970). Based on Tobler's law, the IDW interpolator is adopted to estimate the values of the unsampled locations/units as a weighted sum of the sample values with weights inversely decreasing with distances to the locations/units under estimation. Conditions ensuring design-based unbiasedness and consistency of IDW interpolator concern: i) the spatial pattern of the survey variable, ii) the shape of spatial units or the enlargement of points when dealing with finite population or units of finite population of points, respectively, iii) the sampling design, iv) the distance function. The IDW interpolator regarding the three different population settings is described in detail in the next Chapter.

# Chapter 3

# Materials and Methods

## 3.1 Notation and setting

Consider a study region $A$, that is supposed to be a connected and compact set of $\mathbb{R}^2$, and let $f$ be a bounded measurable function related to the values of a survey variable $Y$ and defined on a subset $B$ of $A$. Moreover, let $\| \; \|$ be a norm in $\mathbb{R}^2$ and $\phi: [0,\infty) \to \mathbb{R}^+$ a non-increasing continuous function on $(0,\infty)$, with $\phi(0) = 0$ and

$$\lim_{d \to 0^+} \phi(d) = \infty \tag{1}$$

Depending on the features of the spatial populations, the IDW mapping of the interest attribute can be performed in the following three settings.

### *3.1.1 Continuous populations*

$B$ coincides with $A$ and $f(\mathbf{p})$ is the value or the density of the survey variable $Y$ at $\mathbf{p} \in B$. Therefore, mapping necessitates the knowledge of $f(\mathbf{p})$ for each $\mathbf{p} \in B$. To this purpose, let $\mathbf{P}_1, \dots, \mathbf{P}_n$ be $n$ random variables with values in $B$ that represent the $n$ locations selected from $B$ by means of a probabilistic sampling scheme. Then, in accordance with Fattorini *et al.* (2018a), if the existence of the continuous probability density function of $(\mathbf{P}_1, \dots, \mathbf{P}_n)$ is assumed, the IDW interpolator of $f(\mathbf{p})$ is almost certainly equal to

$$\hat{f}(\mathbf{p}) = \sum_{i=1}^{n} w_i(\mathbf{p}) f(\mathbf{P}_i)\,, \mathbf{p} \in B \tag{2}$$

with weights

$$w_i(\mathbf{p}) = \frac{\phi(\| \mathbf{P}_i - \mathbf{p} \|)}{\sum_{i=1}^{n} \phi(\| \mathbf{P}_i - \mathbf{p} \|)}, i = 1, \dots, n$$

### *3.1.2 Finite populations of areas*

$A$ is partitioned into a finite population $U$ of $N$ spatial units $a_1, \dots, a_N$ of extents $\lambda_1, \dots, \lambda_N$. In this case, $B$ is the set of centroids $\mathbf{b}_1, \dots, \mathbf{b}_N$ of the areas and $y_j$ is the amount of the survey variable $Y$ within $a_j$. Therefore, mapping necessitates the knowledge of $y_j$ for each $j \in U$. However, since the area size $\lambda_j$ is usually known for each $j \in U$, the knowledge of

$y_j$ is equivalent to the knowledge of density $f_j = f(\mathbf{b}_j) = y_j/\lambda_j$. Accordingly, if $S$ denotes the set of labels identifying the selected areas, then, in accordance with Fattorini *et al.* (2018b), the IDW interpolator of $f_j$ is given by

$$\hat{f}_j = Z_j f_j + (1 - Z_j) \sum_{i \in S} w_{ij} f_i \, , j \in U \tag{3}$$

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

$$w_{ij} = \frac{\phi(\| \mathbf{b}_i - \mathbf{b}_j \|)}{\sum_{h \in S} \phi(\| \mathbf{b}_h - \mathbf{b}_j \|)}, i \in S$$

Once the $\hat{f}_j$s are achieved, the interpolation of the $y_j$s is simply given by

$$\hat{y}_j = \lambda_j \hat{f}_j, j \in U \tag{4}$$

While interpolators (3) and (4) are equivalent for finite $N$, the interpolation of densities is more suitable for working in an asymptotic scenario as that considered by *Fattorini et al.* (2018b) in which the $\lambda_j$s decrease and then the $y_j$s invariably approach zero (see section 3.2.2).

### *3.1.3 Finite populations of marked points*

A finite population $U$ of $N$ points is scattered onto $A$, $B$ is the set $\{\mathbf{p}_1, \dots, \mathbf{p}_N\}$ of point locations and $y_j = f(\mathbf{p}_j)$ is the value of the survey variable for the point $j$, referred to as the mark of unit $j$. Therefore, mapping requires the knowledge of $y_j$ for each $j \in U$. Therefore, if $S$ denotes the set of labels identifying the selected points, then, in accordance with Fattorini *et al.* (2019),

$$\hat{y}_j = Z_j y_j + (1 - Z_j) \sum_{i \in S} w_{ij} y_i \, , j \in U \tag{5}$$

where $Z_j$ is defined as in (3) and

$$w_{ij} = \frac{\phi\left(\left\| \mathbf{p}_i - \mathbf{p}_j \right\|\right)}{\sum_{h \in S} \phi\left(\left\| \mathbf{p}_h - \mathbf{p}_j \right\|\right)}, i \in S$$

The asymptotic properties of (2), (3), and (5) are derived in Fattorini *et al.* (2018a), Fattorini *et al.* (2018b), and Fattorini *et al.* (2019), respectively. Without going into technical details (that are provided in those papers), whatever the population setting, the design-based unbiasedness and consistency of IDW interpolators concern: i) a sort of smoothness of the function $f$ onto the study region with jumps that occur in sub-sets of measure 0; ii) when dealing with finite populations of spatial units or finite populations

of marked points, the regularities in the shape of spatial units or the regularities in the enlargements of the points populations, respectively; iii) the capacity of the sampling design to asymptotically achieve a spatial balance of the selected units; iv) the use of a distance function $\phi$ satisfying (1).
It is worth noting that the four conditions seem to match most real situations encountered in environmental surveys. Indeed, the smoothness assumption i) is very common and it is at the basis of most interpolation techniques (e.g., Cressie, 1993, section 3.1). Moreover, assumption i) is also reasonably valid in many natural scenarios where the density of an attribute changes smoothly throughout space (continuity) and when it changes abruptly, that usually occurs along borders delineating variations in the characteristics of the study region (e.g., forest-meadows). Therefore, borders may be realistically approximated by curves well approaching the theoretical condition of discontinuity over a region of zero measure. Regarding the other conditions, assumption ii) concerning the regularity of the shape of spatial units is ensured by the fact that in most cases, especially in forest and vegetation surveys, spatial units are regular polygons, while assumption concerning regularity in the enlargements of point population simply ensures the absence of many isolated points in the population, a situation unlikely to occur in dense natural communities. Finally, iii) and iv) do not constitute assumptions because both the sampling scheme and the distance function are selected by the user. Moreover, as emphasized in the subsequent section, the asymptotic spatial balance required by iii) is ensured under the schemes usually applied in environmental surveys. Finally, assumption iv) is simply satisfied adopting negative powers of distances such as $\phi(d) = d^{-\alpha}$ for any choice of $\alpha > 2$.

### 3.2 Harmonization with the overall population total

In environmental and ecological studies, the total of a survey variable for the whole study region frequently covers a role of great interest. As typical examples, estimation of the total amount of a pollutant in a lake is crucial for achieving information on environmental changes in the surrounding zones (e.g. Greaver *et al.*, 2016), estimation of the total erosion extent in a region is fundamental for the management of cultivations (e.g. Kelley, 1990), estimation of the total timber volume in a forest is essential for determining carbon storage and sequestration capacity (Food and Agriculture Organization of the United Nations, 2010, Global Forest Resources Assessment 200, Chapter 2).

Once the $n$ locations, areas, or points are selected by means of a probabilistic sampling scheme, the population total is commonly estimated by means of the Horvitz-Thompson (HT) criterion or by related criteria able to exploit the presence of auxiliary information. These criteria constitute consolidated, widely applied strategies as they were unbiased or approximately unbiased with design-based variances with known analytic expressions or approximations. Therefore, variances can be estimated on the basis of those expressions (Gregoire & Valentine, 2008).

However, total estimates naturally arise from the resulting maps as the integral, in the continuous case, or the sum, in the discrete cases, of the interpolated values, thus achieving total estimates that invariably differ from those achieved by traditional, HT-based techniques. To eliminate these unsuitable discrepancies, the matching of the two total estimates can be performed by rescaling the IDW interpolation of the interest attribute for each location, area, or point. Therefore, for continuous populations, the rescaled maps are given by

$$\hat{f}^*(\mathbf{p}) = \frac{\hat{T}}{\hat{T}_{IDW}} \hat{f}(\mathbf{p}), \mathbf{p} \in B \tag{6}$$

while, for both finite populations of areas and finite populations of points, the rescaled maps are given by

$$\hat{y}_j^* = \frac{\hat{T}}{\hat{T}_{IDW}} \hat{y}_j, j \in U \tag{7}$$

respectively, where, in all cases, $\hat{T}$ denotes the total estimate obtained by means of a commonly adopted HT-based technique and $\hat{T}_{IDW}$ denotes the total estimate obtained from the resulting map.

It should be noticed that this harmonization also constitutes a way to exploit first-order inclusion probabilities in the interpolation, stated that they are not directly involved in the IDW interpolator.

### *3.2.1 IDW harmonization for continuous populations*

In the case of continuous populations, the spatial total can be expressed as

$$T = \int_B f(\mathbf{p}) d\mathbf{p} \tag{8}$$

(see e.g., Stevens, 1997; Gregoire & Valentine, 2008, Chapter 10), while its design-based estimation $\hat{T}$ can be obtained by means of the extension of the HT estimator to the continuous case

$$\hat{T} = \sum_{i=1}^{n} \frac{f(\mathbf{P}_i)}{\pi(\mathbf{P}_i)} \tag{9}$$

where $\pi(\mathbf{P}_i)$ denotes the strictly positive inclusion density function on $B$ at the sample locations $\mathbf{P}_i$ $(i = 1, \dots, n)$ (see e.g., Cordy, 1993).

Because of the integral representation (8), the estimation of $T$ may be also approached as a Monte Carlo integration. Interestingly, the most common Monte Carlo integration methods such as crude Monte Carlo integration, modified Monte Carlo integration and random grid Monte Carlo integration are equivalent to Uniform Random Sampling (URS), Tessellation Stratified Sampling (TSS) and Systematic Grid Sampling (SGS), respectively, that are the most common sampling schemes adopted in environmental surveys for continuous populations (e.g., Barabesi, 2003).

In these cases, the estimator (9) reduces to

$$\hat{T} = \frac{\lambda(B)}{n} \sum_{i=1}^{n} f(\mathbf{P}_i) \tag{10}$$

Where $\lambda(B)$ is the extent of $B$. Consistency of (10) under URS, TSS and SGS has been proven by supposing a sequence of designs, each of them selecting an increasing number of points. In particular, the relative efficiency of TSS with respect to URS has been proven for finite samples, also proving that efficiency approaches infinity as the sample size increases because TSS variance goes to 0 more quickly than under URS (Barabesi & Franceschi, 2011; Barabesi *et al.*, 2012, 2015).

Moreover, consistency of (10) under SGS has been proven by Fattorini *et al.* (2020b).

On the other hand, from the resulting map, the total estimate turns out to be

$$\hat{T}_{IDW} = \int_B \hat{f}(\mathbf{p}) d\mathbf{p} = \sum_{i=1}^{n} f(\mathbf{P}_i) \int_B w_i(\mathbf{p}) d\mathbf{p} \tag{11}$$

Fattorini *et al.* (2018a) derived conditions for ensuring design-based consistency of interpolator (2) and hence of estimator (11), proving that it holds under the same asymptotic scenario and the same schemes (URS, TSS and SGS) that ensure consistency of (10).

Joining the two consistency results for the estimators (10) and (11), the rescaling constant $\hat{T}/\hat{T}_{IDW}$ in equation (6) obviously converges to 1, in such a way that the harmonized interpolator $\hat{f}^*(\mathbf{p})$ converges to the IDW interpolator $\hat{f}(\mathbf{p})$. That proves consistency of (6) for continuous populations under URS, TSS and SGS.

*3.2.2 IDW harmonization for finite population of areas*

In the case of finite populations of areas, the population total can be expressed as

$$T = \sum_{j \in U} y_j \tag{12}$$

while its design-based estimation can be obtained by means of the HT estimator

$$\hat{T} = \sum_{j \in S} \frac{y_j}{\pi_j} \tag{13}$$

where $\pi_j$ denotes the first-order inclusion probability of the area $j$ induced by the adopted sampling scheme. As the extents of areas partitioning the study region decrease in such a way that their number and sample size increase, Fattorini *et al.* (2020b) derived conditions ensuring design-based consistency of (13). In particular, they proved that consistency conditions are satisfied when Simple Random Sampling Without Replacement (SRSWOR) and One Per Stratum Sampling (OPSS) are adopted. Moreover, they proved consistency also under Systematic Sampling (SYS), that however necessitates further assumptions.

On the other hand, from the resulting map, the total estimate turns out to be

$$\hat{T}_{IDW} = \sum_{j \in U} \hat{y}_j \tag{14}$$

Fattorini *et al.* (2018b, Supplementary Material, Appendix B) derived conditions for ensuring design-based consistency of (14), proving that it holds under the same asymptotic scenario and the same schemes (SRSWOR, OPSS and SYS) that ensure consistency of (13). Joining the two consistency results for the estimators (13) and (14), the rescaling constant $\hat{T}/\hat{T}_{IDW}$ in equation (7) obviously converges to 1, in such a way that the harmonized interpolator $\hat{y}_j^*$ converges to the IDW interpolator $\hat{y}_j$ under SRSWOR, OPSS and SYS. It is worth noting that this result does not entail consistency of (7), because only consistency of the $\hat{f}_j$s can be proven, while nothing ensures consistency of the $\hat{y}_j$s for finite populations of areas.

*3.2.3 IDW harmonization for finite populations of marked points*

In the case of finite populations of marked points, the population total can be expressed as

$$T = \sum_{j \in U} y_j \tag{15}$$

while the HT estimator of (15) turns out to be

$$\hat{T} = \sum_{j \in S} \frac{y_j}{\pi_j} \tag{16}$$

where $\pi_j$ denotes the first-order inclusion probability of point $j$ induced by the adopted sampling scheme. Supposing a sequence of nested populations of points, in such a way that both population and sample size can increase, Fattorini *et al.* (2020b) derived conditions ensuring the design-based consistency of (16).

In particular, they focused on 3P sampling because it represents the sole scheme of practical relevance when mapping natural populations. Indeed, for natural communities over large study regions (e.g., plants, shrubs, or trees), the list and locations of population units are usually not available. Therefore, maps of marked points are precluded. The sole cases in which mapping is possible occurred for forest stands located on surfaces of limited size, i.e., patches of no more than 10-15 hectares in which 3P sampling can be performed (Iles 2003, p. 594). Indeed, in these cases it becomes possible to visit (and then to locate and list) all the population units by a team of experts that also gives a prediction $y_j^0$ of the survey variable for each $j \in U$. In those situations, Grosenbaugh (1964) proposed to adopt a Poisson sampling, independently including units in the sample with probabilities proportional to the predictions, i.e., with first-order inclusion probabilities $\pi_j = y_j^0/L$, where $L$ is an upper bound for marks ensuring that $\pi_j \leq 1$ for any $j \in U$ (Gregoire & Valentine, 2008). In forest surveys, this scheme is usually referred to as 3P sampling. Under this scheme, Fattorini *et al.* (2020b) proved consistency of (16) if also a lower bound for marks, say $l > 0$, exists, ensuring that $\pi_j \geq l/L$ for any $j \in U$. This requirement usually occurs in forest surveys, because in most cases only trees with values of the survey variable greater than a threshold are admitted in the population.

Since the predictions from a crew of experts obtained to perform 3P sampling are likely to represent good proxies for the survey variable, it seems convenient to exploit this auxiliary information and to estimate (15) by means of the difference estimator (Särndal *et al.*, 1992, section 6.3). In practice, in presence of predictions, the population total (15) can be rewritten as

$$T = \sum_{j \in U} y_j^0 + \sum_{j \in U} e_j = T^0 + E \tag{17}$$

where $e_j = y_j - y_j^0$ are the prediction errors, $T^0$ is the total of predictions and $E$ is the total of errors. Therefore, because $T^0$ is known, the estimation of $T$ reduces to the estimation of $E$ by means of the HT criterion, in such a way that the difference estimator turns out to be

$$\hat{T} = T^0 + \sum_{j \in S} \frac{e_j}{\pi_j} = T^0 + \hat{E}_{HT} \tag{18}$$

where $\hat{E}_{HT}$ denotes the HT estimator of $E$. Being a simple translation of an HT estimator, the difference estimator (18) is consistent under the same asymptotic scenario ensuring consistency of (16). Moreover, if predictions are good, a relevant gain in precision can be achieved with respect to (16).

Similarly, under 3P sampling, Fattorini *et al.* (2019) suggested to exploit predictions also for mapping, proposing the IDW interpolation of prediction errors instead of the values of the survey variable. In practice, the authors propose to adopt the IDW interpolator of errors

$$\hat{e}_j = Z_j e_j + (1 - Z_j) \sum_{i \in S} w_{ij} e_i \tag{19}$$

where the $Z_j$s and $w_{ij}$s are as in (5), in such a way that the interpolated marks are obtained as

$$\hat{y}_j = y_j^0 + \hat{e}_j \tag{20}$$

Under good predictions, Fattorini *et al.* (2020a) proved the formidable gain in precision that can be achieved adopting (20) instead of the direct interpolation of the survey variable obtained by means of (5). It should be noticed that (20) keeps on being a genuine interpolator for $y_j$ because $\hat{e}_j = e_j$ when $j \in S$, in such a way that $\hat{y}_j = y_j^0 + e_j = y_j$. Therefore, from the resulting map (20), the total estimate turns out to be

$$\hat{T}_{IDW} = \sum_{j \in U} \hat{y}_j = T^0 + \sum_{j \in U} \hat{e}_j = T^0 + \hat{E}_{IDW} \tag{21}$$

where $\hat{E}_{IDW}$ denotes the IDW estimator of $E$. Fattorini *et al.* (2019) proved the design-based consistency of IDW interpolator under 3P sampling and the same asymptotic scenario that ensures consistency of HT estimator and difference estimator as well. Accordingly, $\hat{E}_{IDW}$ is a consistent estimator of $E$, in such a way that $\hat{T}_{IDW}$ is a consistent estimator of $T$ under 3P sampling and the same asymptotic scenario that ensures consistency of (18).

Joining the two consistency results for estimators (18) and (21), the rescaling constant

$\hat{T}/\hat{T}_{IDW}$ in equation (7) obviously converges to 1, in such a way that the harmonized interpolator $\hat{y}_j^*$ converges to the IDW interpolator $\hat{y}_j$. That proves consistency of (7) for finite populations of points under 3P sampling.

### 3.3 IDW harmonization by domains

It frequently occurs that total estimates of an interest attribute are required also for $D$ subpopulations partitioning the population, usually referred to as domains (e.g., Särndal *et al.*, 1992, section 10.3). For example, a survey region may be divided into $D$ domains by administrative bounds (e.g., regions, counties, municipalities) and we could be interested not only in mapping and in the overall total of an attribute, but also in its sub-totals within the domains.

In the case of continuous populations, denote by $B_1, \ldots, B_D$ the $D$ subsets partitioning $B$, whose totals are of interest. In this case, estimation of totals within domains can be performed simply defining, for each domain $d = 1, \ldots, D$, the function

$$f_d(\mathbf{p}) = \begin{cases} f(\mathbf{p}) & if\ \mathbf{p} \in B_d \\ 0 & otherwise \end{cases},$$

in such a way that the total for the domain $d$, say $T_{(d)}$, is obtained as in (8) by substituting $f(\mathbf{p})$ with $f_d(\mathbf{p})$. Similarly, the Monte Carlo estimator for the $d$-th domain, say $\hat{T}_{(d)}$, can be performed as in (10), once again substituting $f(\mathbf{p})$ with $f_d(\mathbf{p})$. On the other hand, from the resulting map, the total estimate for each domain $d = 1, \ldots, D$ , say $\hat{T}_{IDW(d)}$, turns out to be as in (11) with the integral extended to $B_d$, instead of the whole $B$. Therefore, for continuous populations, the rescaled map ensuring the matching of total estimates for each domain, as well as the matching of the overall total estimates, is given by

$$\hat{f}^*(\mathbf{p}) = \frac{\hat{T}_{(d)}}{\hat{T}_{IDW(d)}} \hat{f}(\mathbf{p}), \mathbf{p} \in B_d, d = 1, \ldots, D \qquad (22)$$

Consistency of (22) arises, *mutatis mutandis*, from the same consideration performed in section 3.2.1.

In the case of finite populations of areas, denote by $U_1, \ldots, U_D$ the $D$ subsets of areas partitioning $U$, whose totals are of interest. In this case, estimation of totals within domains can be performed simply defining, for each domain $d = 1, \ldots, D$, the values

$$y_{j(d)} = \begin{cases} y_j & if\ j \in U_d \\ 0 & otherwise \end{cases},$$

in such a way that the total for the domain $d$, say $T_{(d)}$, is obtained as in (12) by substituting $y_j$ with $y_{j(d)}$. Similarly, the HT estimator for the $d$-th domain, say $\hat{T}_{(d)}$, can be performed as in (13), once again substituting $y_j$ with $y_{j(d)}$. On the other hand, from the resulting map, the total estimate for each domain $d = 1, \dots, D$ , say $\hat{T}_{IDW(d)}$, turns out to be as in (14) with the summand extended to $U_d$, instead of to the whole $U$. Therefore, for finite populations of areas, the rescaled map ensuring the matching of total estimates for each domain, as well as the matching of the overall total estimates, is given by

$$\hat{y}_j^* = \frac{\hat{T}_{(d)}}{\hat{T}_{IDW(d)}} \hat{y}_j, j \in U_d, d = 1, \dots, D \quad (23)$$

Convergence of the $\hat{y}_j^*$s to $\hat{y}_j$s arises, *mutatis mutandis*, from the same consideration performed in section 3.2.2.

Finally, in the case of finite populations of points, denote by $U_1, \dots, U_D$ the $D$ subsets of points partitioning $U$, whose totals are of interest. In this case, estimation of totals within domains can be performed simply defining, for each domain $d = 1, \dots, D$, the values

$$y_{j(d)} = \begin{cases} y_j & if\ j \in U_d \\ 0 & otherwise \end{cases}$$

as well as the predictions

$$y_{j(d)}^0 = \begin{cases} y_j^0 & if\ j \in U_d \\ 0 & otherwise \end{cases},$$

in such a way that the total for the domain $d$, say $T_{(d)}$, is obtained as in (17) by substituting $y_j$ with $y_{j(d)}$ and $y_j^0$ with $y_{j(d)}^0$. Similarly, the difference estimator for the $d$-th domain, say $\hat{T}_{(d)}$, can be performed as in (18), once again substituting $y_j$ with $y_{j(d)}$ and $y_j^0$ with $y_{j(d)}^0$. On the other hand, from the resulting map, the total estimate for each domain $d = 1, \dots, D$ , say $\hat{T}_{IDW(d)}$, turns out to be as in (21) with the summand extended to $U_d$, instead of the whole $U$. Therefore, for finite populations of points, the rescaled map ensuring the matching of the overall total estimates is identical to (23) and its consistency arises, *mutatis mutandis*, from the same consideration performed in section 3.2.3.

# Chapter 4

# Simulation Studies

For each population setting, simulation studies are performed to empirically check if and how much the harmonization procedure deteriorates the performance of IDW maps as well as to check the rate of convergence of the harmonized maps to the original ones.

## 4.1 Populations and sampling

As to the continuous populations setting, the survey region considered for the simulation was a quadrat of 129,600 ha located in North-Western Tuscany (Central Italy). The population values to be mapped were the precipitations (mm) occurred between 3rd January 2021 and 3rd February 2021, which were artificially achieved by means of an ordinary kriging prediction performed from the values recorded on 46 rain gauge stations present in the survey region. The population average was $\bar{Y} = 282$ mm for the whole region (see Figure 1). Moreover, the region was partitioned into $D = 2, 4, 8$ domains of equal sizes, as depicted in Figure 2. Precipitation averages (mm) within the two domains were $\bar{Y}_{(1)} = 282$ and $\bar{Y}_{(2)} = 283$, precipitation averages within the four domains were $\bar{Y}_{(1)} = 305$, $\bar{Y}_{(2)} = 264$, $\bar{Y}_{(3)} = 282$ and $\bar{Y}_{(4)} = 279$, precipitation averages within the eight domains were $\bar{Y}_{(1)} = 247$, $\bar{Y}_{(2)} = 254$, $\bar{Y}_{(3)} = 369$, $\bar{Y}_{(4)} = 294$, $\bar{Y}_{(5)} = 222$, $\bar{Y}_{(6)} = 258$, $\bar{Y}_{(7)} = 316$ and $\bar{Y}_{(8)} = 299$.

Sampling was performed by selecting $n = 16, 36, 64, 100$ locations on the quadrat by means of URS, TSS and SGS, i.e. selecting $n$ locations completely at random (URS), partitioning the quadrat into $n$ sub-quadrats of equal size and selecting a location at random within each of them (TSS), or selecting one location at random in one of them and then repeating the location in the others (SGS).

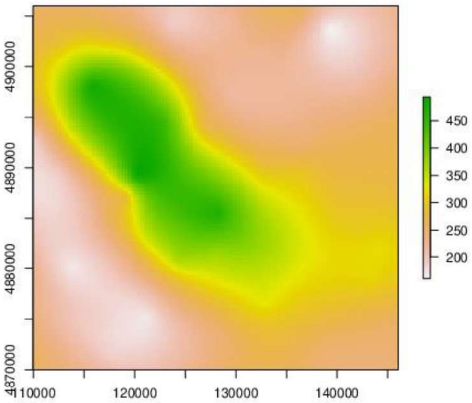


**Figure 1.** Precipitations (mm) between 3rd January 2021 and 3rd February 2021 in a quadrat region located in North-Western Tuscany (Central Italy).

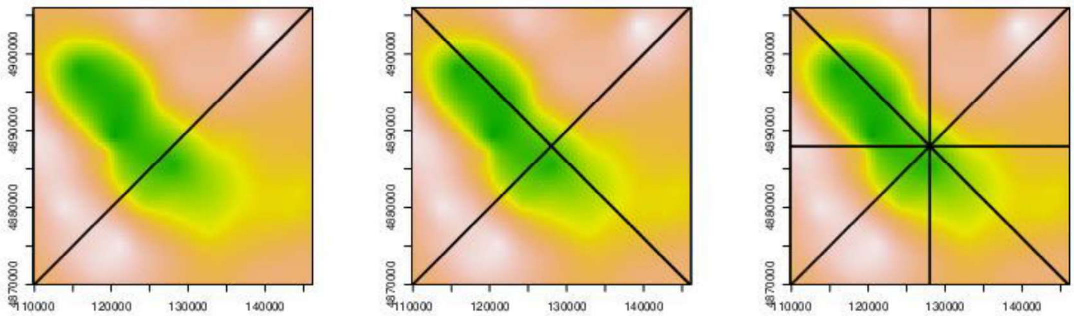

**Figure 2.** Partition of the quadrat region located in North-Western Tuscany into $D = 2, 4, 8$ domains of equal sizes.

As to finite populations of areas we considered both artificial and real populations. Regarding artificial populations, three surfaces on the unit square, referred to as surface 1, surface 2, and surface 3, were considered. The three surfaces were respectively defined at any point $\mathbf{p} = (p_1, p_2)$ by

$$f(\mathbf{p}) = C_1(sin^2 p_1 + cos^2 p_2 + p_1)$$

$$f(\mathbf{p}) = C_2(sin 3p_1 sin^2 3p_2)$$

$$f(\mathbf{p}) = \begin{cases} C_3 p_1 p_2 & \min(p_1 p_2 < 0.5) \\ C_3(1 + p_1 p_2) & otherwise \end{cases}$$

where the constants $C_1, C_2$ and $C_3$ ensured a minimum density of 0 and a maximum density of 10. The three surfaces were chosen to represent the major characteristics of spatial populations. Surface 1 and surface 2 were continuous, indeed the values of the survey variable in neighboring locations tended to be similar. Surface 1 showed an increasing trend moving from the left to the right part of the unit square while surface 2 showed an increasing trend toward the center of the square. Lastly, surface 3 presented a discontinuity at the edge of the upper right quadrant of the unit square (see Figure 3).

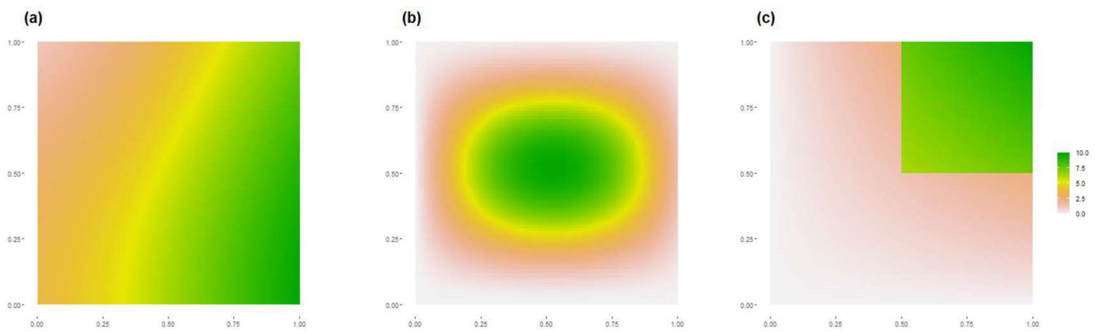


**Figure 3.** Maps of the three artificial surfaces: (a) surface 1, (b) surface 2, (c) surface 3 on the unit square.

For any artificial surface, five populations of $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$ areas were constructed by partitioning the unit square into grids of $10 \times 10, 20 \times 20, 40 \times 40, 80 \times 80, 160 \times 160$ quadrats, respectively (e.g., see Figure 4). Population values were the integrals of the surfaces onto the quadrats. Accordingly, the population totals resulted of $T = 5.54, 3.50$ and $2.50$ for surface 1, surface 2 and surface 3, respectively.

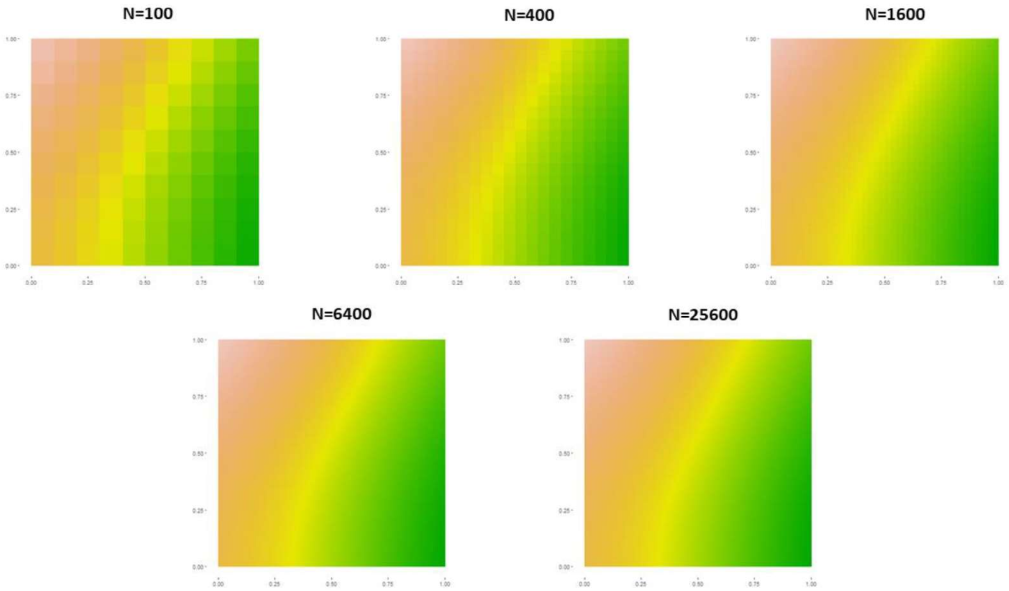


**Figure 4.** Maps of surface 1 partitioned into the five populations of $100$, $400$, $1{,}600$, $6{,}400$, $25{,}600$ areas.

Moreover, for any surface, the unit square was partitioned into $D = 2, 4, 8$ domains constituted by an equal number of areas, as depicted in Figure 5. Totals within domains for the three surfaces were reported in Table 1.

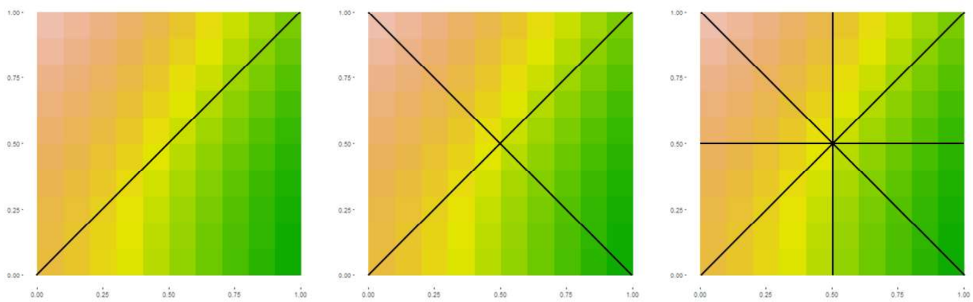

**Figure 5.** Partition of the unit square of $N = 100$ areas into $D = 2, 4, 8$ domains constituted by an equal number of areas.

**Table 1.** Values of totals within the $D = 2, 4, 8$ domains for surface 1, surface 2 and surface 3.

| | surface 1 | surface 2 | surface 3 |
|---|---|---|---|
| $D = 2$ | 3.13 | 1.59 | 0.31 |
| | 2.41 | 1.91 | 2.19 |
| $D = 4$ | 0.92 | 0.89 | 0.09 |
| | 1.03 | 0.84 | 0.97 |
| | 1.92 | 1.11 | 1.32 |
| | 1.67 | 0.66 | 0.12 |
| $D = 8$ | 0.66 | 0.26 | 0.03 |
| | 0.49 | 0.48 | 0.05 |
| | 0.42 | 0.57 | 0.08 |
| | 0.37 | 0.32 | 0.16 |
| | 0.68 | 0.34 | 0.96 |
| | 0.94 | 0.68 | 0.99 |
| | 0.98 | 0.57 | 0.17 |
| | 1.00 | 0.28 | 0.06 |

Regarding real populations, the survey region considered for the simulation was a rectangle of about 212 ha located in Calabria (Southern Italy). Three populations of $N = 250, 1{,}000, 4{,}000$ areas were considered by partitioning the region into as many quadrats of size 8,464, 2,116 and 529 m$^2$, respectively. The population values to be mapped were the densities of growing stock volumes (m$^3$/ha) within the areas that were artificially achieved by means of a random forest imputation technique (see Chirici *et al.*, 2020 for details) from the ground data recorded in 2000-2007 during the Italian National Forest Inventory (Fattorini *et al.*, 2006). The population total resulted of $T = 63{,}524.6$ m$^3$ for the whole region (see Figure 6).

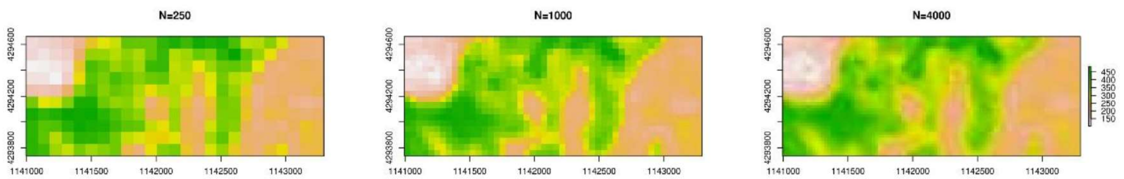


**Figure 6.** Maps of the growing stock volume densities (m$^3$/ha) in a rectangular region located in Calabria (Southern Italy) partitioned into three populations of 250, 1,000, 4,000 areas.

Moreover, the region was partitioned into $D = 2, 4, 8$ domains constituted by an equal number of areas, as depicted in Figure 7. Totals (in m$^3$) within the two domains were $T_1 = 34{,}915.5$ and $T_2 = 28{,}609.1$, totals within the four domains were $T_1 = 16{,}830.6$, $T_2 = 18{,}084.9$, $T_3 = 12{,}634.2$, and $T_4 = 15{,}974.8$, totals within the eight domains were $T_1 = 8{,}891.6$, $T_2 = 10{,}553.7$, $T_3 = 6{,}276.9$, $T_4 = 8{,}755.0$, $T_5 = 9{,}329.9$, $T_6 = 6{,}350.5$, $T_7 = 6{,}283.8$, and $T_8 = 7{,}083.3$.

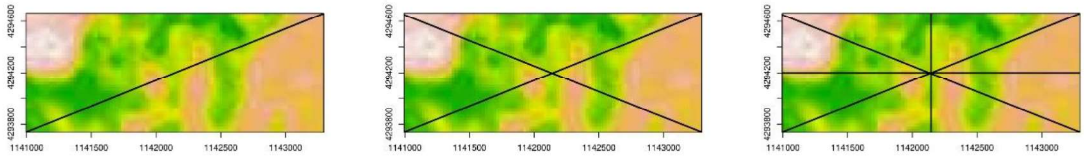

**Figure 7.** Partition of the rectangular region of $N = 4{,}000$ areas located in Calabria into $D = 2, 4, 8$ domains constituted by an equal number of areas.

For both artificial and real populations, sampling was performed by selecting samples of $n = N/10$ areas by means of SRSWOR, OPSS and SYS, i.e. selecting $n$ areas at random without replacement (SRSWOR), partitioning the populations into $n$ blocks of $2 \times 5$ contiguous areas, and selecting an area at random within each of them (OPSS), or selecting one area at random in one of them and then repeating it in the others (SYS).

Finally, as to the finite populations of marked points, three nested populations of trees settled in a forested region of about 4.8 ha located in Val di Sella (North-Eastern Italy, Southern Alpine area) were considered for simulation. The largest population was constituted by the $N = 608$ trees settled in the region with diameter at breast height greater than 17.5 cm and wood volume greater than $l = 1.5$ m$^3$. The middle-size population of 365 trees was achieved from the largest one by randomly discarding 243 (about 40%) trees, and the smallest population of 219 trees was achieved from the middle-size one by randomly discarding 146 (about 40%) trees. The populations values were the wood volumes (m$^3$) evaluated by means of the local double-entry volume tables currently used in the forest management practice (Corona *et al.*, 2014). The population totals resulted of $T = 843.4, 1{,}353.5$ and $2{,}223.4$ m$^3$ for the smallest, middle-size, and largest population, respectively (see Figure 8).

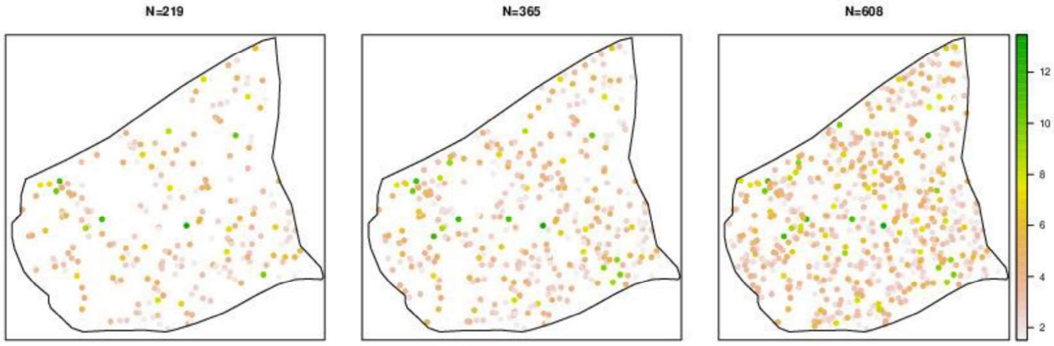


**Figure 8.** Maps of wood volumes ($m^3$) for three nested populations of $219, 365, 608$ trees settled in a forested region of Val di Sella (North-Eastern Italy).

Moreover, the forested region was partitioned into $D = 2, 4, 8$ domains of approximately equal extents, as depicted in Figure 9. Totals (in $m^3$) within domains for the three populations were reported in Table 2.

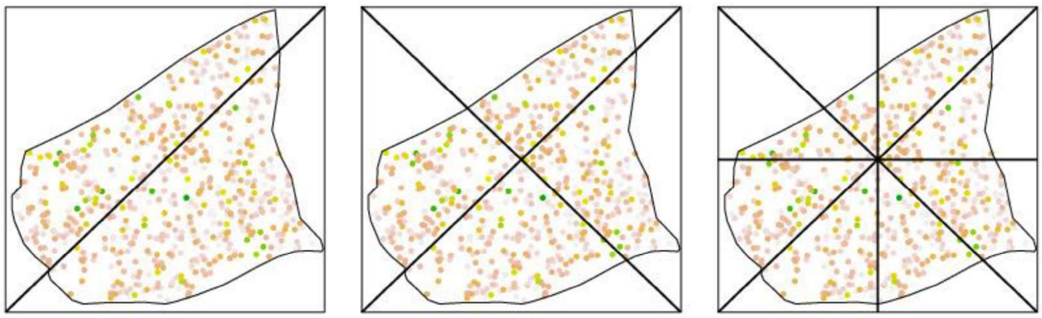

**Figure 9.** Partition of the forested region of Val di Sella into $D = 2, 4, 8$ domains of approximately equal extents.

**Table 2.** Values of totals of wood volumes (m$^3$) within the $D = 2, 4, 8$ domains of approximately equal extents for the three nested populations of $219, 365$ and $608$ trees settled in a forested region of Val di Sella.

| | $N = 219$ | $N = 365$ | $N = 608$ |
|---|---|---|---|
| $D = 2$ | 346.3 | 557.9 | 966.0 |
| | 497.1 | 795.6 | 1257.4 |
| $D = 4$ | 205.4 | 333.1 | 553.4 |
| | 140.9 | 224.9 | 412.5 |
| | 226.6 | 375.9 | 596.3 |
| | 270.5 | 419.7 | 661.1 |
| $D = 8$ | 133.6 | 198.9 | 359.0 |
| | 151.0 | 237.6 | 390.5 |
| | 54.5 | 95.4 | 162.9 |
| | 40.9 | 55.4 | 105.0 |
| | 100.0 | 169.5 | 307.6 |
| | 63.9 | 112.8 | 190.3 |
| | 162.7 | 263.1 | 406.0 |
| | 136.9 | 220.8 | 302.1 |

3P sampling was performed in the three populations independently selecting each tree with probability $\pi_j = y_j^0/L$, where $L = 30.6$ was an upper bound for marks that ensured $\pi_j \leq 1$, and an expected sample size of about 10% of each population size. In turn, expert predictions $y_j^0$ were generated in accordance with Fattorini *et al.* (2020a), assuming that predictions increased linearly with marks values with a maximum error rate $r = 0.15, 0.10, 0.05$ occurred at the extremes $l = 1.5$ and $L = 30.6$.

### 4.2 Simulation

For each combination of population, sampling scheme, and sample size, sampling was replicated $R = 10{,}000$ times. At each simulation run, the estimated maps were obtained by means of the IDW interpolators (2), (3) and (5), for the continuous populations and the finite populations of areas and points, respectively. In all cases, interpolation was performed by using $d^{-3}$ as distance function. In the case of the continuous population, mapping was performed computing the estimates onto a regular grid of $120 \times 120$ nodes on the quadrat region of Figure 1. Then population totals were estimated adopting HT or difference criterion by means of equations (10), (13) and (18), as well as from the resulting maps by means of equations (11), (14) and (21) for the continuous populations and the finite populations of areas and points, respectively. In the case of continuous population, average estimates were achieved by dividing the total estimates by the size of the study region. On the basis of these estimates, the harmonized maps were achieved by means of equation (6) for the continuous population and equation (7) for finite populations of areas and points.

Moreover, each population was partitioned into $D = 2, 4, 8$ domains as depicted in Figures 2, 5, 7 and 9 and, for each combination of sampling scheme, sample size and domain partition, sampling was replicated $R = 10{,}000$ times. At each simulation run, if the sample was empty for at least one domain, the sample selection was newly repeated until all the domains contained at least one sample unit. Then, IDW mapping was performed as described above, total estimates within domains were performed by means of the HT or difference criterion as well as from the resulting maps as described in section 3.3, and the harmonized maps were achieved by means of equation (22) for the continuous population and equation (23) for finite populations of areas and points. Once again, in the case of continuous populations, average estimates were achieved by dividing the total estimates by the size of the domains.

### 4.3 Performance indicators

Denote by $\hat{f}_r(\mathbf{p}_{k,l})$ and $\hat{f}_r^*(\mathbf{p}_{k,l})$ the IDW and the harmonized interpolation occurred at the node $k, l$ of the $120 \times 120$ grid within the quadrat region of Figure 1 at the $r$ simulation run. Then, the absolute bias (AB)

$$AB(\mathbf{p}_{k,l}) = \left|\frac{1}{R}\sum_{r=1}^{R} \hat{f}_r(\mathbf{p}_{k,l}) - f(\mathbf{p}_{k,l})\right| \tag{24}$$

and

$$AB^*(\mathbf{p}_{k,l}) = \left|\frac{1}{R}\sum_{r=1}^{R} \hat{f}_r^*(\mathbf{p}_{k,l}) - f(\mathbf{p}_{k,l})\right| \tag{25}$$

as well as the root mean squared error (RMSE)

$$RMSE(\mathbf{p}_{k,l}) = \left[\frac{1}{R}\sum_{r=1}^{R}\left[\hat{f}_r(\mathbf{p}_{k,l}) - f(\mathbf{p}_{k,l})\right]^2\right]^{1/2} \tag{26}$$

and

$$RMSE^*(\mathbf{p}_{k,l}) = \left[\frac{1}{R}\sum_{r=1}^{R}\left[\hat{f}_r^*(\mathbf{p}_{k,l}) - f(\mathbf{p}_{k,l})\right]^2\right]^{1/2} \tag{27}$$

were computed for each node $k,l$ for $k,l = 1,\ldots,120$.

Similarly, denote by $\hat{y}_{j,r}$ and $\hat{y}_{j,r}^*$ the IDW and the harmonized interpolation of the $j$ unit in the populations of areas or points at the $r$ simulation run. Then, the absolute bias (AB)

$$AB_j = \left|\frac{1}{R}\sum_{r=1}^{R}\hat{y}_{j,r} - y_j\right| \tag{28}$$

and

$$AB_j^* = \left|\frac{1}{R}\sum_{r=1}^{R}\hat{y}_{j,r}^* - y_j\right| \tag{29}$$

as well as the root mean squared error (RMSE)

$$RMSE_j = \left[\frac{1}{R}\sum_{r=1}^{R}\left[\hat{y}_{j,r} - y_j\right]^2\right]^{1/2} \tag{30}$$

and

$$RMSE_j^* = \left[\frac{1}{R}\sum_{r=1}^{R}\left[\hat{y}_{j,r}^* - y_j\right]^2\right]^{1/2} \tag{31}$$

were computed for each $j \in U$.

Finally, denote by $\hat{T}_r$ and $\hat{T}_{IDW,r}$ the total estimates (overall or within a domain) achieved at the $r$ simulation run by means of the HT or difference estimator and from the IDW maps, respectively. Then, the quantity

$$RATIO_r = \frac{\hat{T}_r}{\hat{T}_{IDW,r}} \quad r = 1,\ldots,R \tag{32}$$

is the correction factor and quantifies the level of matching between the two total estimates as well as between the IDW and harmonized maps. It should be noted that in the case of continuous population, ratios of averages estimates coincide with (32).

# Chapter 5

# Case Study

## 5.1 Forest complex of Rincine

The IDW interpolator (3), as well as its harmonized version (7), was adopted to provide the map of densities of living wood volume in Rincine (Figure 10), a forest complex in eastern Tuscany, Central Italy (43.9 N, 11.6 E). The forest spreads over an area of approximately 290 ha, and it is representative of the forest ecosystems typically found in the central Apennines. The dominant species are mixed oak, chestnut, and beech stands, as well as coniferous reforestations. The site is under the management of the Union of Mountain Municipalities of Valdarno and Valdisieve.

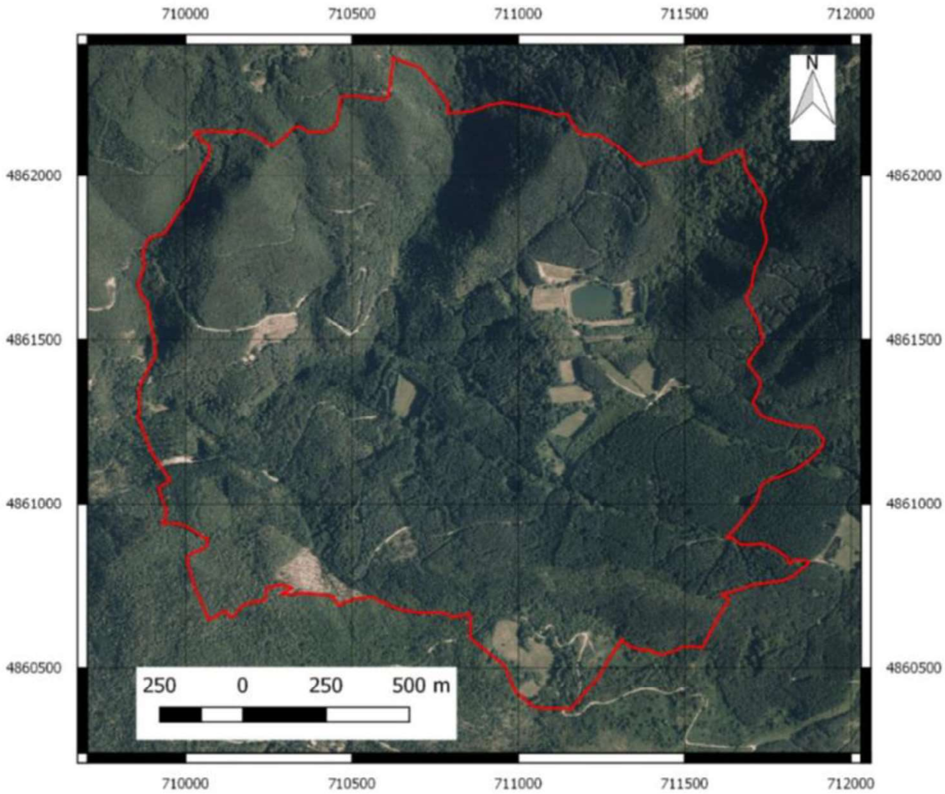


**Figure 10.** Survey area: the forest complex of Rincine.

## 5.2 Sampling procedure and estimation

Sample information was achieved from a survey performed in 2016 by the Union of Mountain Municipalities of Valdarno and Valdisieve. The survey area was tessellated by a grid of $5449$ quadrats of size $23 \times 23$ m$^2$. For sampling purposes, the population was partitioned into 49 blocks of $109$ contiguous cells each and one block of $108$ cells, for a total of $50$ blocks. After the partition, it was realized that $340$ cells were located outside the forest, in artificial areas or waters. Therefore, these cells were discarded from the population. Accordingly, the resulting population to be sampled was constituted by $N = 5109$ quadrats partitioned into blocks of different sizes. The largest block was made up of $109$ units, the smallest one of $57$ units. One quadrat was randomly selected within each block by means of OPSS, for a total sample of $n = 50$ units (see Figure 11).

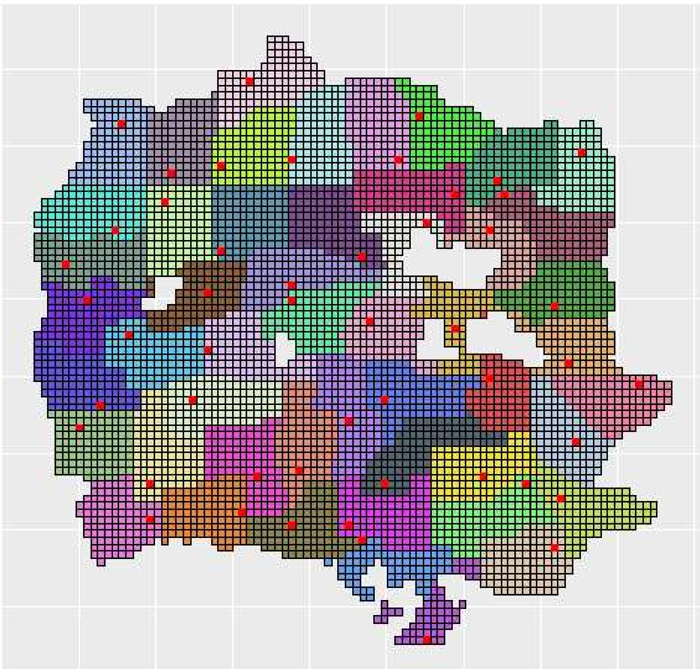

**Figure 11.** Stratification of the study area and spatial distribution of field sampling units. Non-forest cells are white, sampled cells are red.

Field measurements on the sample quadrats were carried out from June to November 2016. In any sample unit tree species, stem diameter at breast height (DBH), tree height, four crown radii, and tree status (live or dead) of all the trees with $DBH > 2.5$ cm were recorded. The wood volume of each tree in the selected cells was reckoned by the allometric equations from the Italian National Forest Inventory (Tabacchi *et al.*, 2011) to achieve the wood volume density by three size classes, i.e., class $Y1$ for small stems ($DBH \leq 12$ cm), class $Y2$ for medium stems ($12 < DBH \leq 27$ cm), class $Y3$ for large stems ($DBH > 27$ cm). The total living wood volume $Y$ was achieved by summing the volumes of the three size classes $Y1$, $Y2$ and $Y3$ in any sample quadrat and its density was adopted to interpolate the total living wood volume density within each spatial unit of the population by means of the IDW interpolator (3). The distance function adopted was $d^{-\alpha}, \alpha = 3$. RMSE estimates were achieved by means of the asymptotically conservative estimator proposed by Fattorini *et al.* (2018b). The total living wood volume estimate in the whole study area was computed by using the HT estimator (13). Additionally, total living wood volume estimate arising from the IDW map was achieved by means of (14). Finally, harmonized estimates were achieved by means of equation (7), and the corresponding RMSEs estimates were computed adopting, *mutatis mutandis*, the estimator proposed by Fattorini *et al.* (2018b).

# Chapter 6

# Results and Discussion

## 6.1 Simulation results

Tables 3, 5, 7, 9, 11, and 13 contain the minima, the averages, and the maxima of ABs and RMSEs arising from the IDW interpolation for each population, sampling scheme and sample size. Figures 12-29 show the spatial pattern of these performance indicators. The same indicators are reported in Tables 4, 6, 8, 10, 12, and 14 for the harmonized interpolation with respect to the overall total or average estimates, together with the minima, the averages, and the maxima of the correction factors (32), for each population, sampling scheme and sample size. Figures showing the spatial pattern of these indicators are not reported because they resulted graphically indistinguishable from those achieved using the crude IDW interpolation.

Finally, Tables 15-20 contain the minima, the averages, and the maxima of ABs and RMSEs arising when harmonization is performed with respect to the total or average estimates within domains, together with the minima, the averages, and the maxima of the correction factors (32), for each population, sampling scheme, sample size and number of domains.

Simulation results show that harmonized and IDW estimates are comparable in terms of AB and RMSE for all the populations and sampling schemes, suggesting that harmonization can be performed without relevant loss in precision. Furthermore, in all the situations, the correction factor quickly approaches 1 as the population and/or sample size increase, as apparent from minima, averages, and maxima of the correction factor (32). Moreover, for the continuous population and for the populations of spatial units, RMSE increases as the number of domains increases. This result was expected since, when the number of domains increases, the number of selected units within becomes lower and lower, thus invariably reducing the precision of the domain estimators from which harmonization is performed. This tendency is not evident in the populations of marked points. That may be because, under good predictions, the difference estimator of totals is so efficient to cover the loss in precision due to the small size of domains and the subsequent scarcity of sample information.

**Table 3.** Minima, averages, and maxima of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of precipitations (expressed in mm) in a quadrat region located in North-Western Tuscany, for each combination of sampling scheme and sample size ($n$).

| Scheme | $n$ | AB | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max |
| URS | 16 | 0.00 | 31.18 | 130.51 | 13.51 | 49.55 | 144.91 |
| | 36 | 0.00 | 21.89 | 109.37 | 7.82 | 34.87 | 121.83 |
| | 64 | 0.00 | 16.89 | 94.44 | 4.99 | 26.89 | 105.88 |
| | 100 | 0.00 | 13.59 | 84.07 | 3.30 | 21.74 | 94.49 |
| TSS | 16 | 0.00 | 26.28 | 122.98 | 7.27 | 40.38 | 135.40 |
| | 36 | 0.00 | 17.81 | 99.00 | 3.60 | 27.11 | 109.77 |
| | 64 | 0.00 | 13.22 | 79.66 | 1.76 | 20.07 | 87.05 |
| | 100 | 0.00 | 10.53 | 76.52 | 1.07 | 16.00 | 85.09 |
| SGS | 16 | 0.00 | 24.11 | 114.81 | 3.95 | 35.75 | 125.80 |
| | 36 | 0.00 | 15.60 | 93.48 | 1.22 | 23.01 | 104.03 |
| | 64 | 0.00 | 11.42 | 78.97 | 0.81 | 16.81 | 87.27 |
| | 100 | 0.00 | 8.99 | 69.18 | 0.66 | 13.24 | 75.76 |

**Table 4.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the overall average estimate) of precipitations (expressed in mm) in a quadrat region located in North-Western Tuscany, for each combination of sampling scheme and sample size ($n$).

| Scheme | $n$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max | min | mean | max |
| URS | 16 | 0.87 | 0.98 | 1.08 | 0.00 | 30.62 | 137.06 | 16.81 | 49.89 | 150.69 |
| | 36 | 0.91 | 0.99 | 1.08 | 0.00 | 21.71 | 113.41 | 10.83 | 35.97 | 126.00 |
| | 64 | 0.92 | 0.99 | 1.07 | 0.00 | 16.80 | 97.01 | 7.92 | 28.11 | 108.76 |
| | 100 | 0.94 | 1.00 | 1.06 | 0.00 | 13.55 | 85.87 | 6.18 | 22.88 | 96.70 |
| TSS | 16 | 0.94 | 1.00 | 1.04 | 0.00 | 26.14 | 126.81 | 7.77 | 40.50 | 138.70 |
| | 36 | 0.97 | 0.99 | 1.02 | 0.00 | 17.76 | 101.17 | 4.04 | 27.29 | 111.79 |
| | 64 | 0.97 | 1.00 | 1.02 | 0.00 | 13.19 | 81.06 | 2.52 | 20.19 | 88.29 |
| | 100 | 0.98 | 1.00 | 1.01 | 0.00 | 10.52 | 77.41 | 1.54 | 16.09 | 85.94 |
| SGS | 16 | 0.96 | 0.98 | 1.00 | 0.00 | 23.90 | 118.38 | 3.09 | 35.19 | 126.99 |
| | 36 | 0.98 | 0.99 | 1.00 | 0.00 | 15.57 | 95.65 | 1.21 | 22.92 | 104.05 |
| | 64 | 0.99 | 1.00 | 1.00 | 0.00 | 11.41 | 80.53 | 1.04 | 16.80 | 87.63 |
| | 100 | 0.99 | 1.00 | 1.00 | 0.00 | 8.99 | 70.17 | 0.91 | 13.25 | 76.67 |

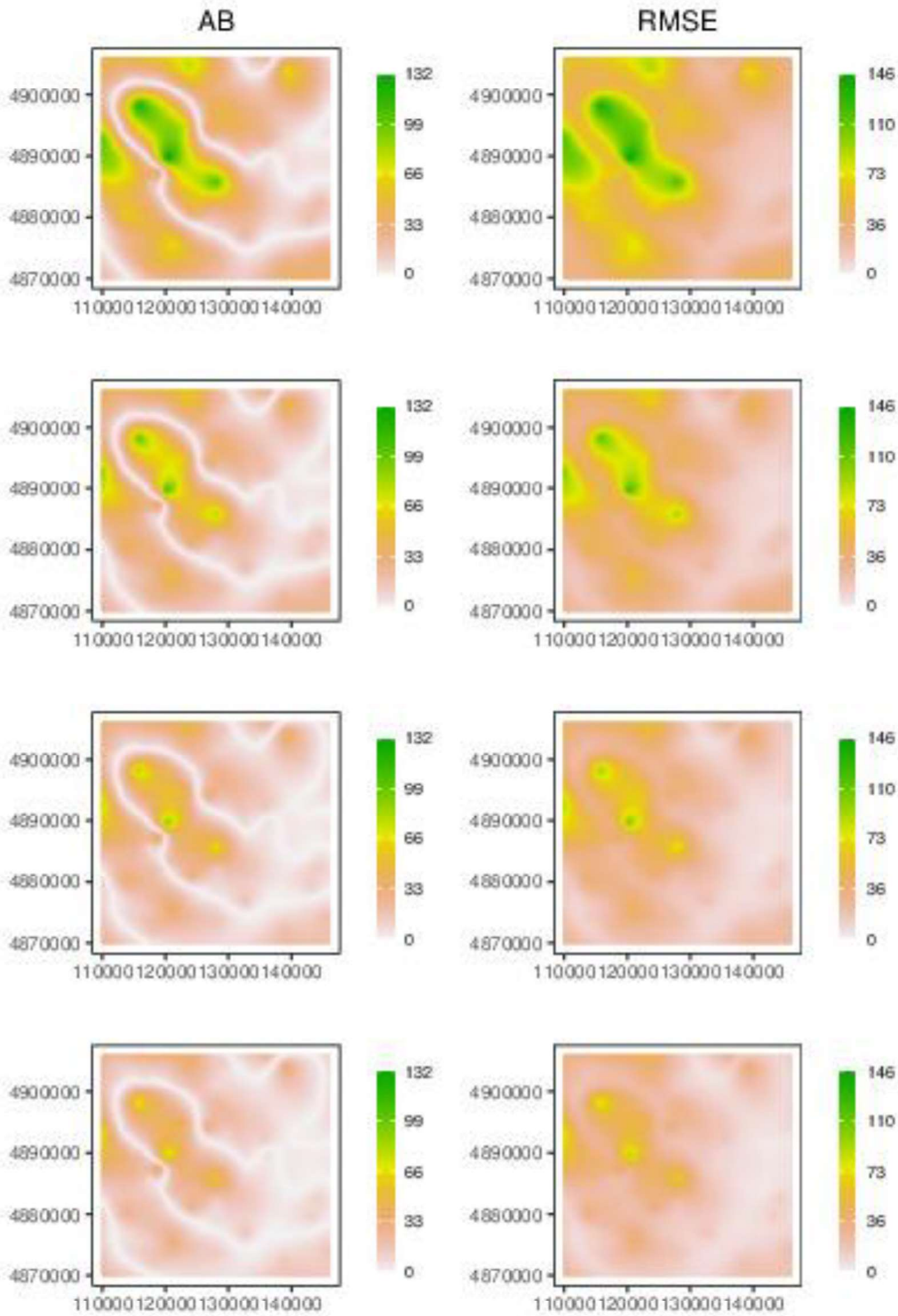


**Figure 12.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of precipitations (expressed in mm) in a quadrat region located in North-Western Tuscany, under URS and sample size $n = 16, 36, 64, 100$.

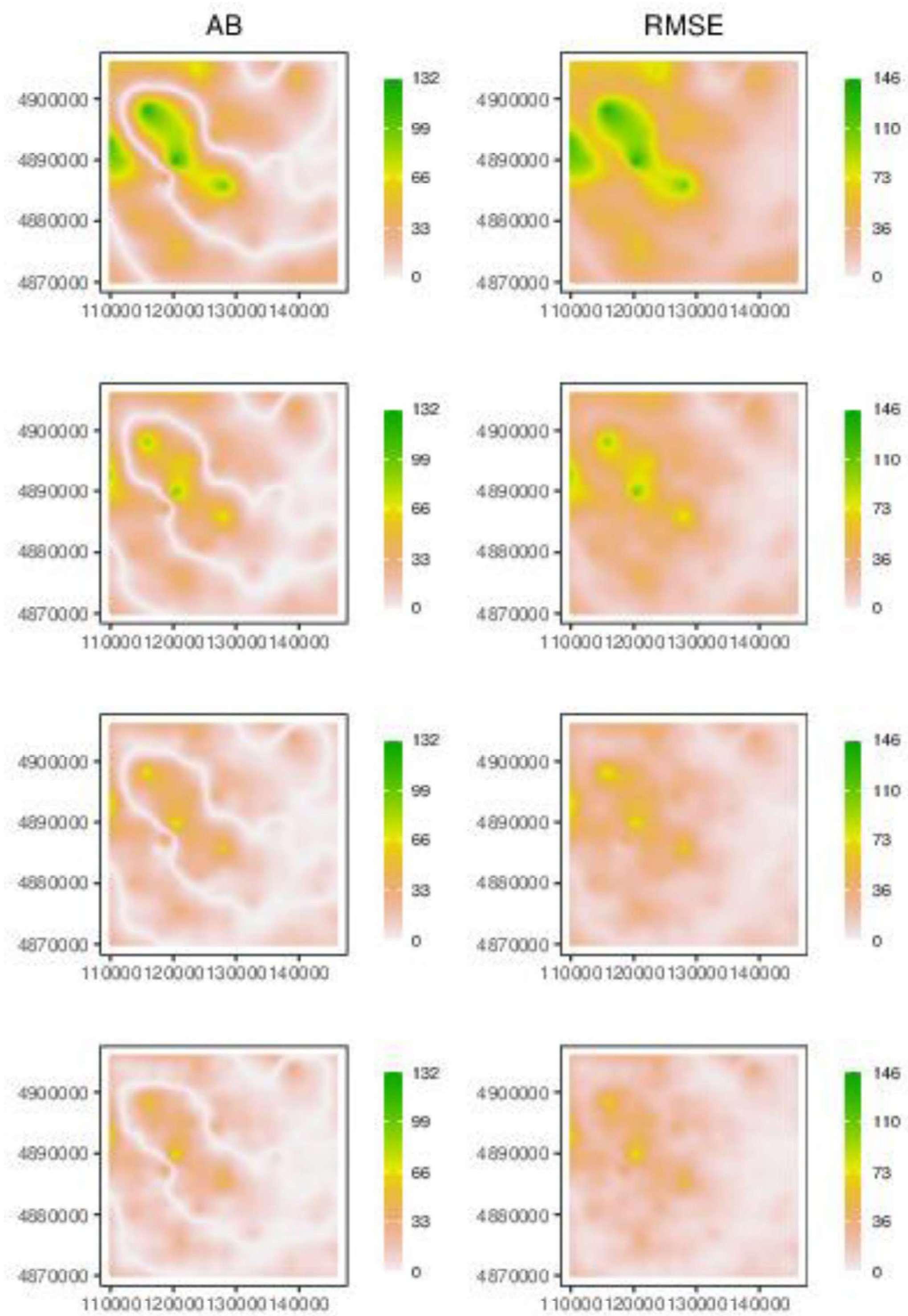


**Figure 13.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of precipitations (expressed in mm) in a quadrat region located in North-Western Tuscany, under TSS and sample size $n = 16, 36, 64,100$.

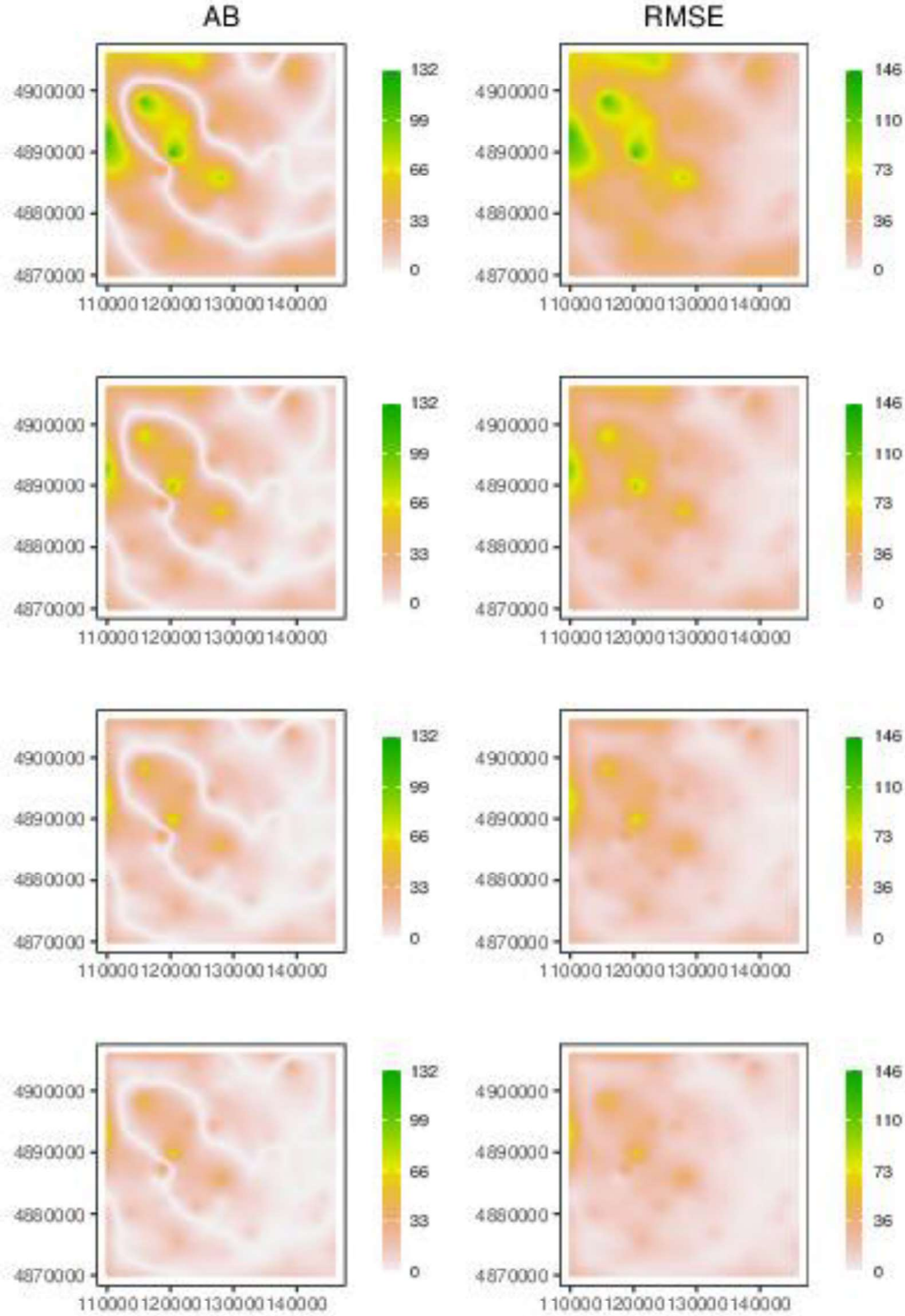


**Figure 14.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of precipitations (expressed in mm) in a quadrat region located in North-Western Tuscany, under SGS and sample size $n = 16, 36, 64, 100$.

**Table 5.** Minima, averages, and maxima of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 1 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | AB | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max |
| SRSWOR | 100 | 0.01 | 0.49 | 1.74 | 0.46 | 0.87 | 2.13 |
| | 400 | 0.00 | 0.21 | 1.06 | 0.19 | 0.39 | 1.33 |
| | 1,600 | 0.00 | 0.09 | 0.64 | 0.09 | 0.17 | 0.81 |
| | 6,400 | 0.00 | 0.04 | 0.37 | 0.04 | 0.08 | 0.47 |
| | 25,600 | 0.00 | 0.02 | 0.21 | 0.02 | 0.04 | 0.27 |
| OPSS | 100 | 0.00 | 0.37 | 1.48 | 0.31 | 0.68 | 1.77 |
| | 400 | 0.00 | 0.15 | 0.86 | 0.12 | 0.29 | 1.04 |
| | 1,600 | 0.00 | 0.07 | 0.49 | 0.05 | 0.13 | 0.60 |
| | 6,400 | 0.00 | 0.03 | 0.27 | 0.03 | 0.06 | 0.34 |
| | 25,600 | 0.00 | 0.01 | 0.15 | 0.01 | 0.03 | 0.19 |
| SYS | 100 | 0.00 | 0.55 | 1.88 | 0.39 | 0.95 | 2.28 |
| | 400 | 0.00 | 0.20 | 1.09 | 0.16 | 0.37 | 1.34 |
| | 1,600 | 0.00 | 0.08 | 0.62 | 0.08 | 0.16 | 0.78 |
| | 6,400 | 0.00 | 0.03 | 0.36 | 0.04 | 0.07 | 0.45 |
| | 25,600 | 0.00 | 0.02 | 0.20 | 0.02 | 0.03 | 0.25 |

**Table 6.** Minima, averages, and maxima of correction factor (RATIO), absolute bias ($AB^*$) and root mean squared error ($RMSE^*$) of the harmonized interpolator (with the overall total estimate) achieved for surface 1 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | RATIO | | | $AB^*$ | | | $RMSE^*$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 100 | 0.76 | 1.00 | 1.20 | 0.01 | 0.49 | 1.76 | 0.64 | 1.03 | 2.27 |
| | 400 | 0.84 | 1.00 | 1.16 | 0.00 | 0.21 | 1.07 | 0.31 | 0.50 | 1.42 |
| | 1,600 | 0.91 | 1.00 | 1.10 | 0.00 | 0.09 | 0.64 | 0.15 | 0.24 | 0.85 |
| | 6,400 | 0.95 | 1.00 | 1.05 | 0.00 | 0.04 | 0.37 | 0.07 | 0.11 | 0.50 |
| | 25,600 | 0.97 | 1.00 | 1.02 | 0.00 | 0.02 | 0.21 | 0.04 | 0.06 | 0.28 |
| OPSS | 100 | 0.90 | 1.00 | 1.11 | 0.01 | 0.37 | 1.46 | 0.31 | 0.72 | 1.82 |
| | 400 | 0.96 | 1.00 | 1.04 | 0.00 | 0.15 | 0.85 | 0.12 | 0.29 | 1.05 |
| | 1,600 | 0.99 | 1.00 | 1.01 | 0.00 | 0.07 | 0.49 | 0.05 | 0.13 | 0.60 |
| | 6,400 | 1.00 | 1.00 | 1.00 | 0.00 | 0.03 | 0.27 | 0.03 | 0.06 | 0.34 |
| | 25,600 | 1.00 | 1.00 | 1.00 | 0.00 | 0.01 | 0.15 | 0.01 | 0.03 | 0.19 |
| SYS | 100 | 0.88 | 1.00 | 1.11 | 0.00 | 0.55 | 1.84 | 0.30 | 1.17 | 2.58 |
| | 400 | 0.91 | 1.00 | 1.09 | 0.00 | 0.20 | 1.04 | 0.16 | 0.54 | 1.64 |
| | 1,600 | 0.95 | 1.00 | 1.05 | 0.00 | 0.08 | 0.60 | 0.09 | 0.26 | 0.98 |
| | 6,400 | 0.97 | 1.00 | 1.03 | 0.00 | 0.03 | 0.35 | 0.04 | 0.13 | 0.56 |
| | 25,600 | 0.99 | 1.00 | 1.02 | 0.00 | 0.02 | 0.20 | 0.02 | 0.06 | 0.32 |

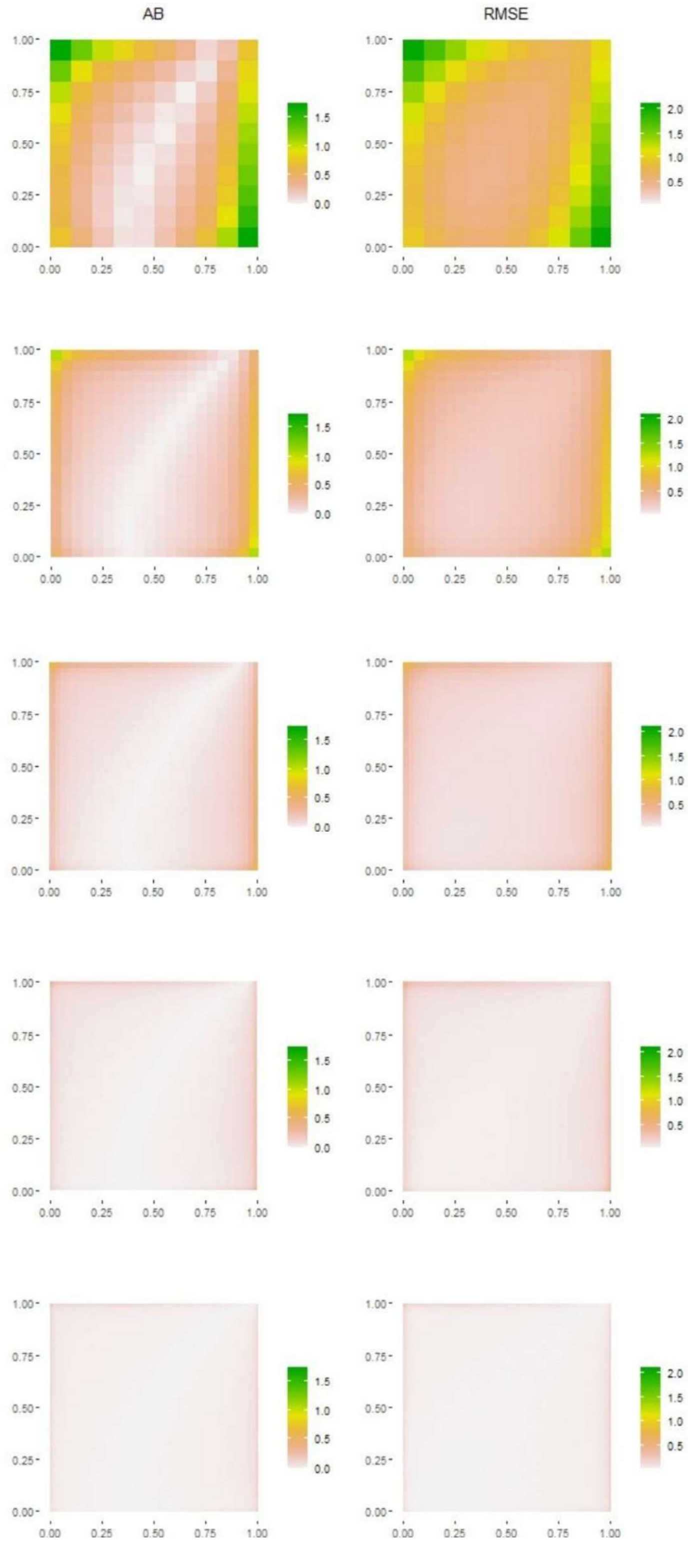


**Figure 15.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 1 partitioned into five populations of areas, under SRSWOR with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

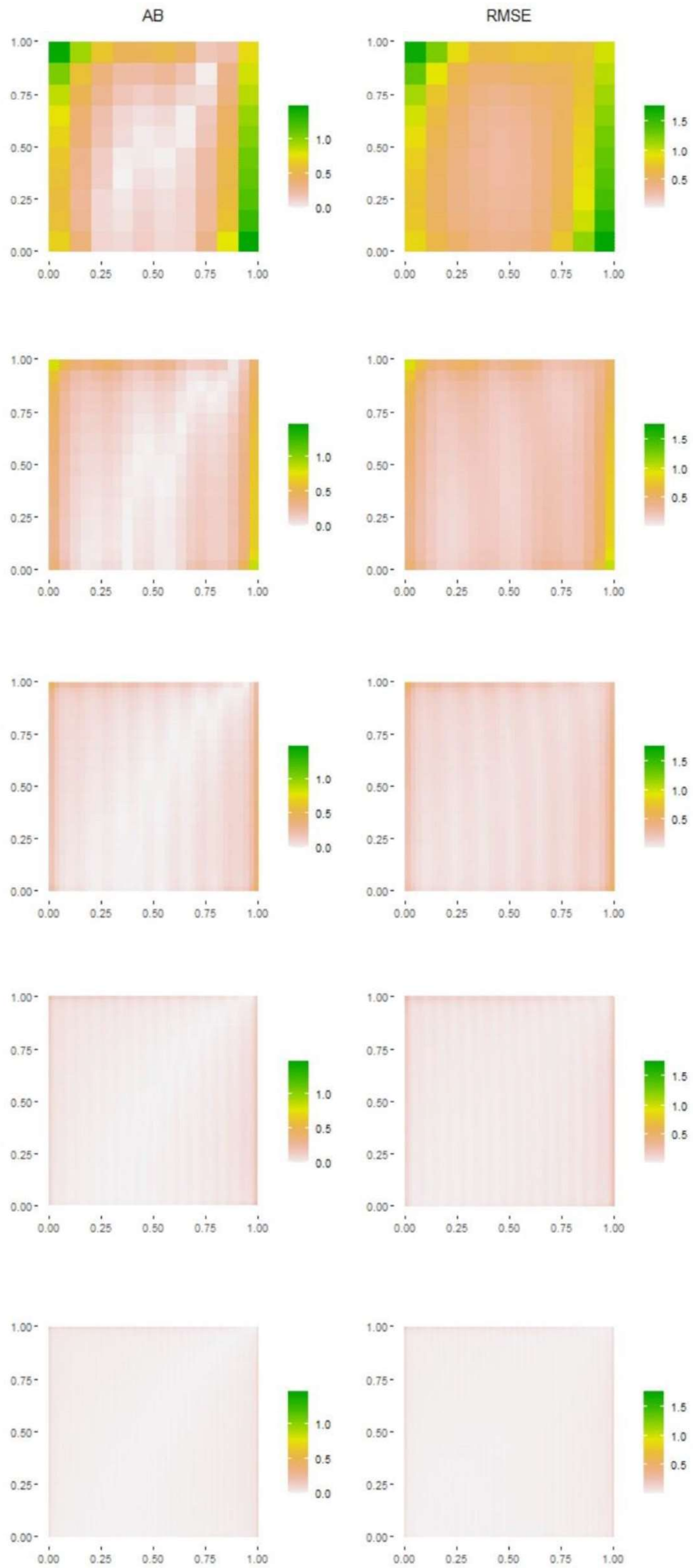


**Figure 16.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 1 partitioned into five populations of areas, under OPSS with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

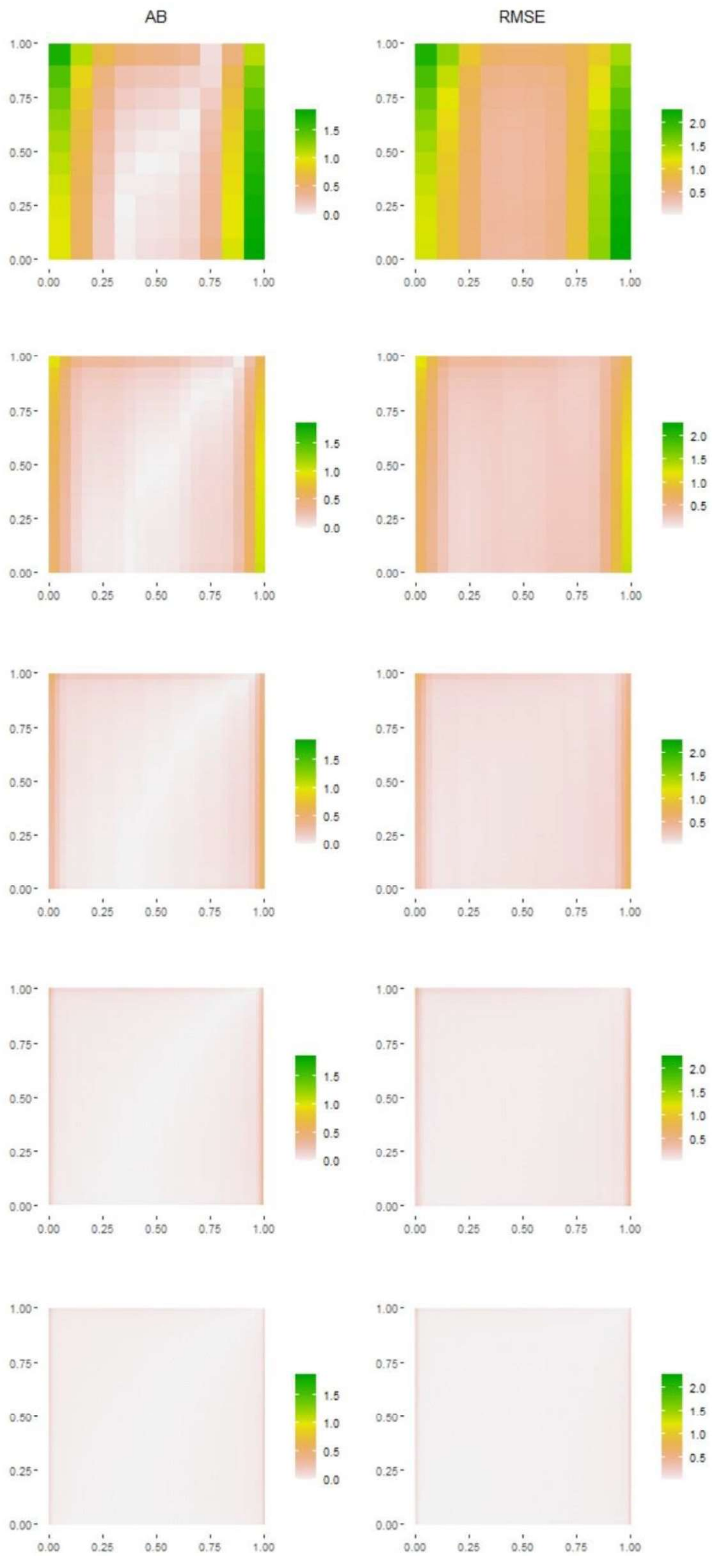


**Figure 17**. Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 1 partitioned into five populations of areas, under SYS with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

**Table 7.** Minima, averages, and maxima of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 2 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | AB | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max |
| SRSWOR | 100 | 0.06 | 1.40 | 2.77 | 1.40 | 2.18 | 3.36 |
| | 400 | 0.01 | 0.73 | 1.72 | 0.67 | 1.17 | 2.06 |
| | 1,600 | 0.00 | 0.35 | 1.25 | 0.30 | 0.57 | 1.51 |
| | 6,400 | 0.00 | 0.16 | 0.81 | 0.12 | 0.27 | 1.00 |
| | 25,600 | 0.00 | 0.08 | 0.51 | 0.05 | 0.13 | 0.64 |
| OPSS | 100 | 0.03 | 1.20 | 2.61 | 1.13 | 1.85 | 3.14 |
| | 400 | 0.00 | 0.58 | 1.68 | 0.46 | 0.90 | 2.01 |
| | 1,600 | 0.00 | 0.27 | 1.14 | 0.17 | 0.42 | 1.37 |
| | 6,400 | 0.00 | 0.13 | 0.70 | 0.06 | 0.20 | 0.85 |
| | 25,600 | 0.00 | 0.06 | 0.42 | 0.03 | 0.10 | 0.51 |
| SYS | 100 | 0.03 | 1.35 | 3.05 | 0.41 | 1.85 | 3.49 |
| | 400 | 0.01 | 0.65 | 2.49 | 0.05 | 0.92 | 2.92 |
| | 1,600 | 0.00 | 0.29 | 1.70 | 0.02 | 0.42 | 2.03 |
| | 6,400 | 0.00 | 0.13 | 1.07 | 0.00 | 0.19 | 1.30 |
| | 25,600 | 0.00 | 0.06 | 0.65 | 0.00 | 0.09 | 0.80 |

**Table 8.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the overall total estimate) achieved for surface 2 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 100 | 0.51 | 0.89 | 1.15 | 0.01 | 1.47 | 3.54 | 1.33 | 2.16 | 4.00 |
| | 400 | 0.70 | 0.95 | 1.17 | 0.00 | 0.79 | 1.95 | 0.69 | 1.25 | 2.35 |
| | 1,600 | 0.81 | 0.98 | 1.16 | 0.00 | 0.38 | 1.23 | 0.32 | 0.63 | 1.49 |
| | 6,400 | 0.89 | 0.99 | 1.10 | 0.00 | 0.17 | 0.81 | 0.12 | 0.31 | 0.99 |
| | 25,600 | 0.94 | 1.00 | 1.05 | 0.00 | 0.08 | 0.51 | 0.05 | 0.15 | 0.64 |
| OPSS | 100 | 0.68 | 0.92 | 1.09 | 0.03 | 1.26 | 3.18 | 0.99 | 1.84 | 3.61 |
| | 400 | 0.89 | 0.97 | 1.05 | 0.01 | 0.62 | 1.62 | 0.43 | 0.92 | 1.94 |
| | 1,600 | 0.95 | 0.99 | 1.02 | 0.00 | 0.29 | 1.12 | 0.16 | 0.43 | 1.36 |
| | 6,400 | 0.99 | 1.00 | 1.01 | 0.00 | 0.13 | 0.70 | 0.06 | 0.20 | 0.85 |
| | 25,600 | 1.00 | 1.00 | 1.00 | 0.00 | 0.06 | 0.42 | 0.03 | 0.10 | 0.51 |
| SYS | 100 | 0.76 | 0.88 | 0.97 | 0.05 | 1.39 | 3.44 | 0.18 | 1.80 | 3.53 |
| | 400 | 0.89 | 0.95 | 0.99 | 0.00 | 0.72 | 2.35 | 0.13 | 0.96 | 2.76 |
| | 1,600 | 0.96 | 0.98 | 1.00 | 0.00 | 0.32 | 1.67 | 0.03 | 0.44 | 2.00 |
| | 6,400 | 0 .99 | 0.99 | 1.00 | 0.00 | 0.15 | 1.07 | 0.01 | 0.20 | 1.30 |
| | 25,600 | 0.99 | 1.00 | 1.00 | 0.00 | 0.07 | 0.65 | 0.00 | 0.10 | 0.80 |

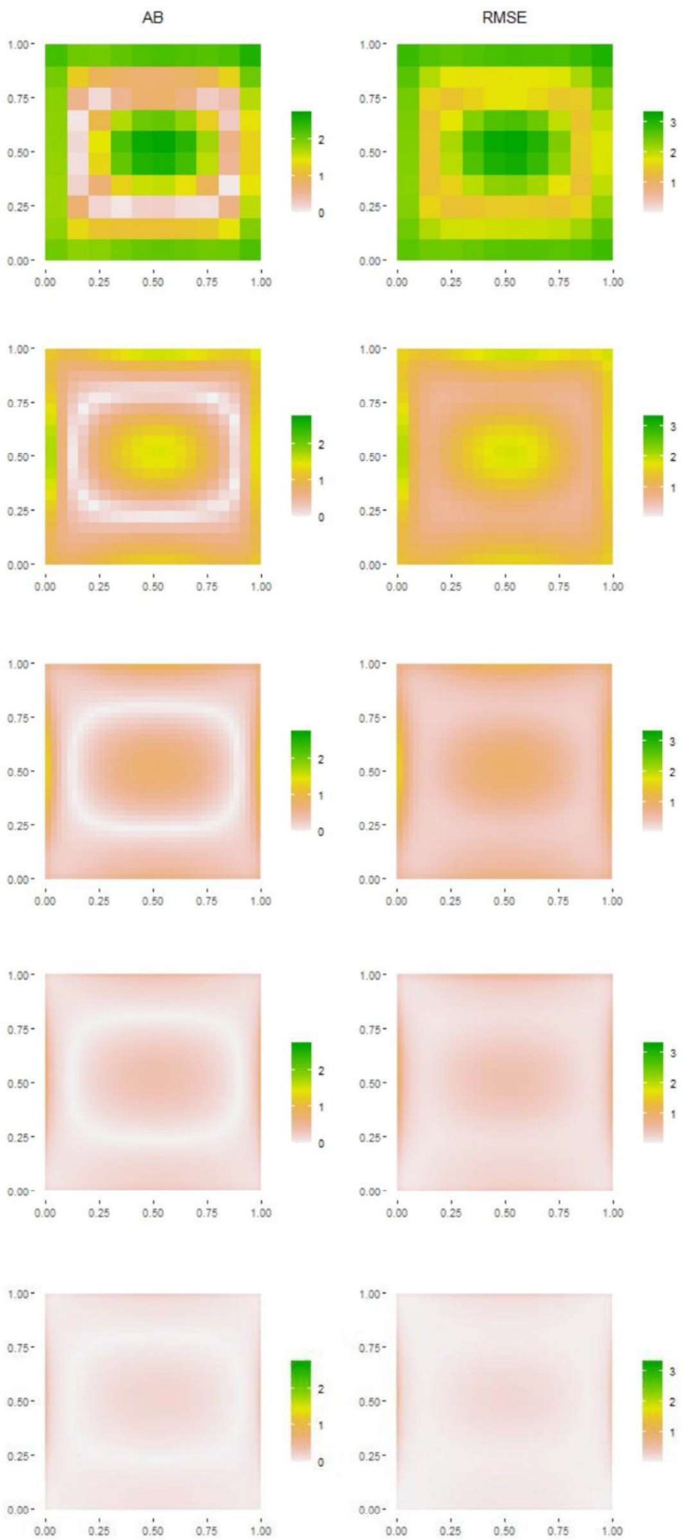


**Figure 18.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 2 partitioned into five populations of areas, under SRSWOR with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

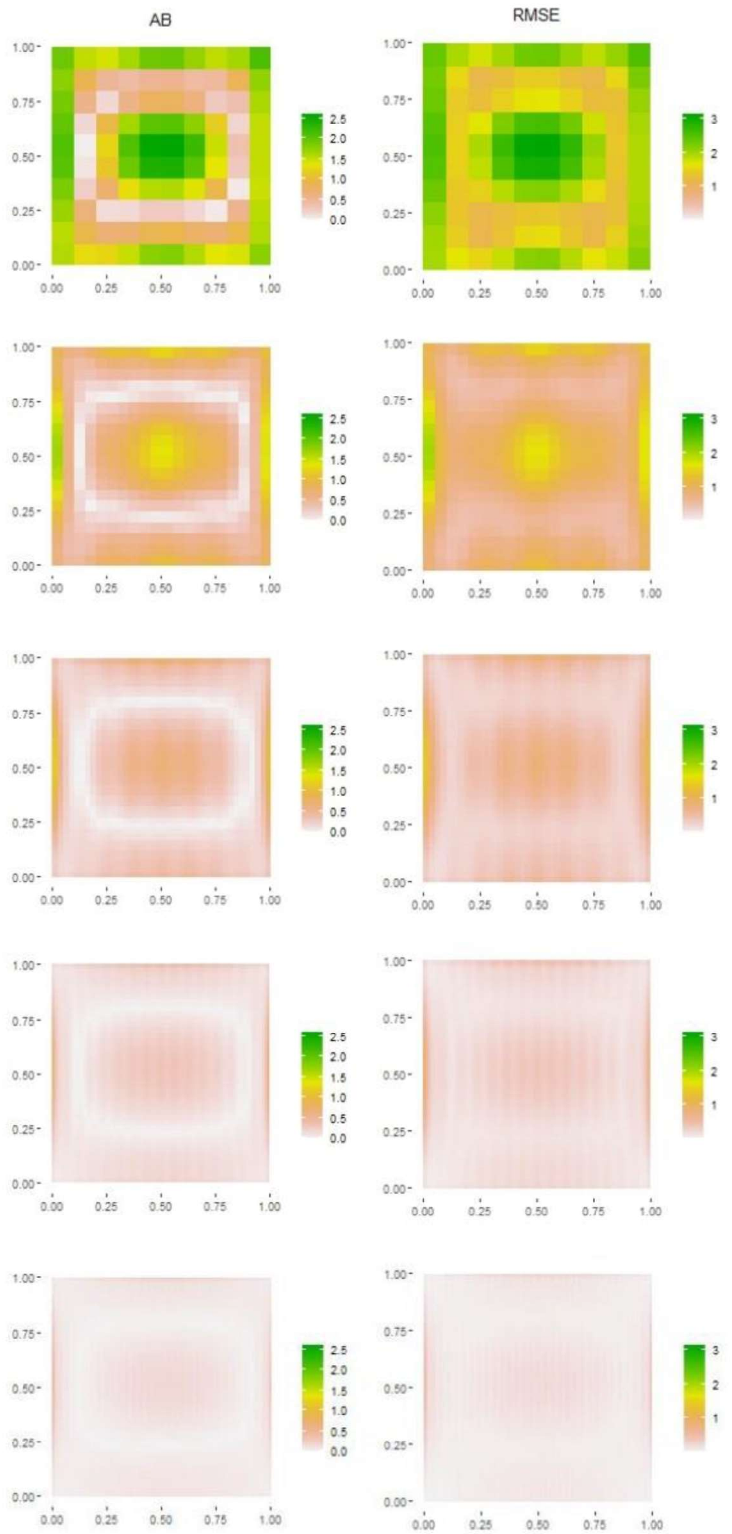


**Figure 19.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 2 partitioned into five populations of areas, under OPSS with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

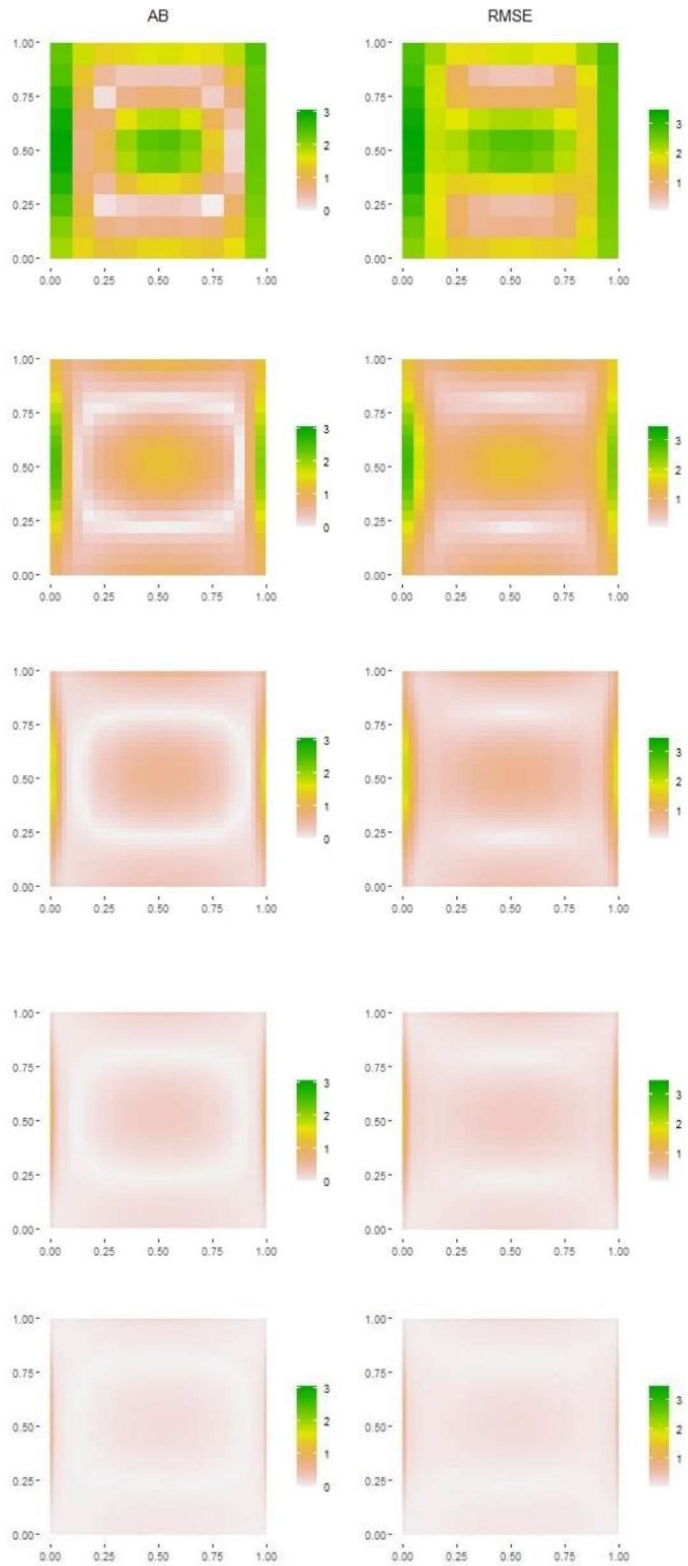


**Figure 20.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 2 partitioned into five populations of areas, under SYS with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

**Table 9.** Minima, averages, and maxima of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 3 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | AB | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max |
| SRSWOR | 100 | 0.28 | 1.01 | 2.97 | 0.50 | 1.54 | 3.53 |
| | 400 | 0.09 | 0.53 | 2.94 | 0.15 | 0.77 | 3.40 |
| | 1,600 | 0.04 | 0.29 | 2.93 | 0.05 | 0.41 | 3.35 |
| | 6,400 | 0.02 | 0.17 | 2.91 | 0.02 | 0.22 | 3.32 |
| | 25,600 | 0.01 | 0.09 | 2.96 | 0.01 | 0.13 | 3.36 |
| OPSS | 100 | 0.15 | 0.78 | 2.87 | 0.21 | 1.09 | 3.37 |
| | 400 | 0.06 | 0.42 | 2.79 | 0.08 | 0.54 | 3.18 |
| | 1,600 | 0.02 | 0.24 | 2.79 | 0.03 | 0.30 | 3.17 |
| | 6,400 | 0.01 | 0.14 | 2.82 | 0.01 | 0.17 | 3.18 |
| | 25,600 | 0.00 | 0.08 | 2.82 | 0.01 | 0.10 | 3.17 |
| SYS | 100 | 0.18 | 0.86 | 2.98 | 0.25 | 1.16 | 3.42 |
| | 400 | 0.06 | 0.46 | 2.95 | 0.08 | 0.61 | 3.36 |
| | 1,600 | 0.03 | 0.26 | 2.96 | 0.03 | 0.33 | 3.34 |
| | 6,400 | 0.01 | 0.14 | 2.96 | 0.01 | 0.18 | 3.33 |
| | 25,600 | 0.00 | 0.08 | 2.96 | 0.01 | 0.10 | 3.33 |

**Table 10.** Minima, averages, and maxima of correction factor (RATIO), absolute bias ($AB^*$) and root mean squared error ($RMSE^*$) of the harmonized interpolator (with the overall total estimate) achieved for surface 3 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | RATIO | | | $AB^*$ | | | $RMSE^*$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 100 | 0.42 | 0.97 | 1.70 | 0.29 | 1.02 | 3.01 | 0.58 | 1.71 | 3.98 |
| | 400 | 0.55 | 0.99 | 1.58 | 0.09 | 0.54 | 2.95 | 0.17 | 0.91 | 3.45 |
| | 1,600 | 0.68 | 1.00 | 1.29 | 0.04 | 0.30 | 2.93 | 0.06 | 0.50 | 3.37 |
| | 6,400 | 0.85 | 1.00 | 1.18 | 0.02 | 0.17 | 2.91 | 0.02 | 0.27 | 3.33 |
| | 25,600 | 0.90 | 1.00 | 1.08 | 0.01 | 0.10 | 2.96 | 0.01 | 0.15 | 3.36 |
| OPSS | 100 | 0.74 | 1.00 | 1.38 | 0.15 | 0.78 | 2.90 | 0.22 | 1.11 | 3.37 |
| | 400 | 0.89 | 1.00 | 1.12 | 0.06 | 0.41 | 2.79 | 0.08 | 0.55 | 3.18 |
| | 1,600 | 0.97 | 1.00 | 1.04 | 0.02 | 0.24 | 2.79 | 0.03 | 0.30 | 3.17 |
| | 6,400 | 0.99 | 1.00 | 1.01 | 0.01 | 0.14 | 2.82 | 0.01 | 0.17 | 3.18 |
| | 25,600 | 1.00 | 1.00 | 1.00 | 0.00 | 0.08 | 2.82 | 0.01 | 0.10 | 3.17 |
| SYS | 100 | 0.70 | 1.00 | 1.37 | 0.17 | 0.86 | 3.20 | 0.22 | 1.29 | 3.47 |
| | 400 | 0.81 | 1.01 | 1.22 | 0.06 | 0.46 | 3.11 | 0.07 | 0.69 | 3.33 |
| | 1,600 | 0.89 | 1.00 | 1.12 | 0.02 | 0.25 | 3.05 | 0.03 | 0.39 | 3.32 |
| | 6,400 | 0.94 | 1.00 | 1.06 | 0.01 | 0.14 | 3.01 | 0.01 | 0.22 | 3.32 |
| | 25,600 | 0.97 | 1.00 | 1.03 | 0.00 | 0.08 | 2.99 | 0.01 | 0.12 | 3.33 |

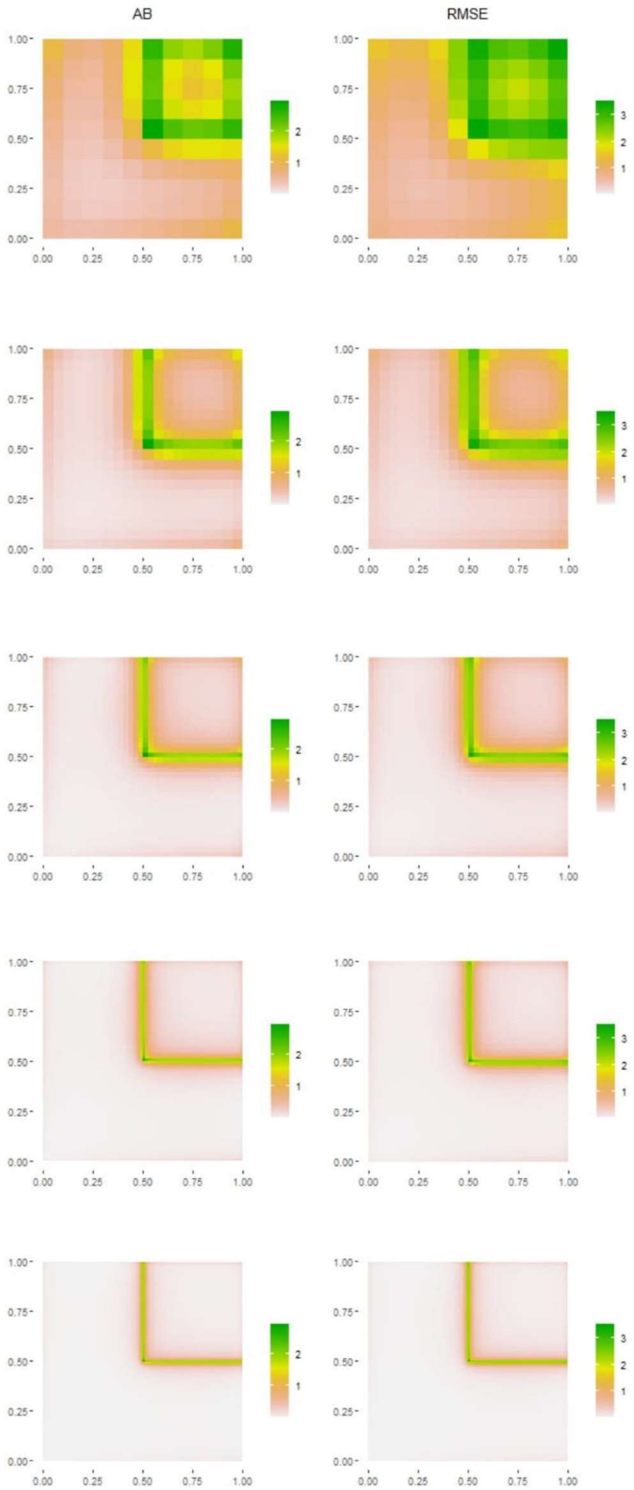


**Figure 21.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 3 partitioned into five populations of areas, under SRSWOR with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

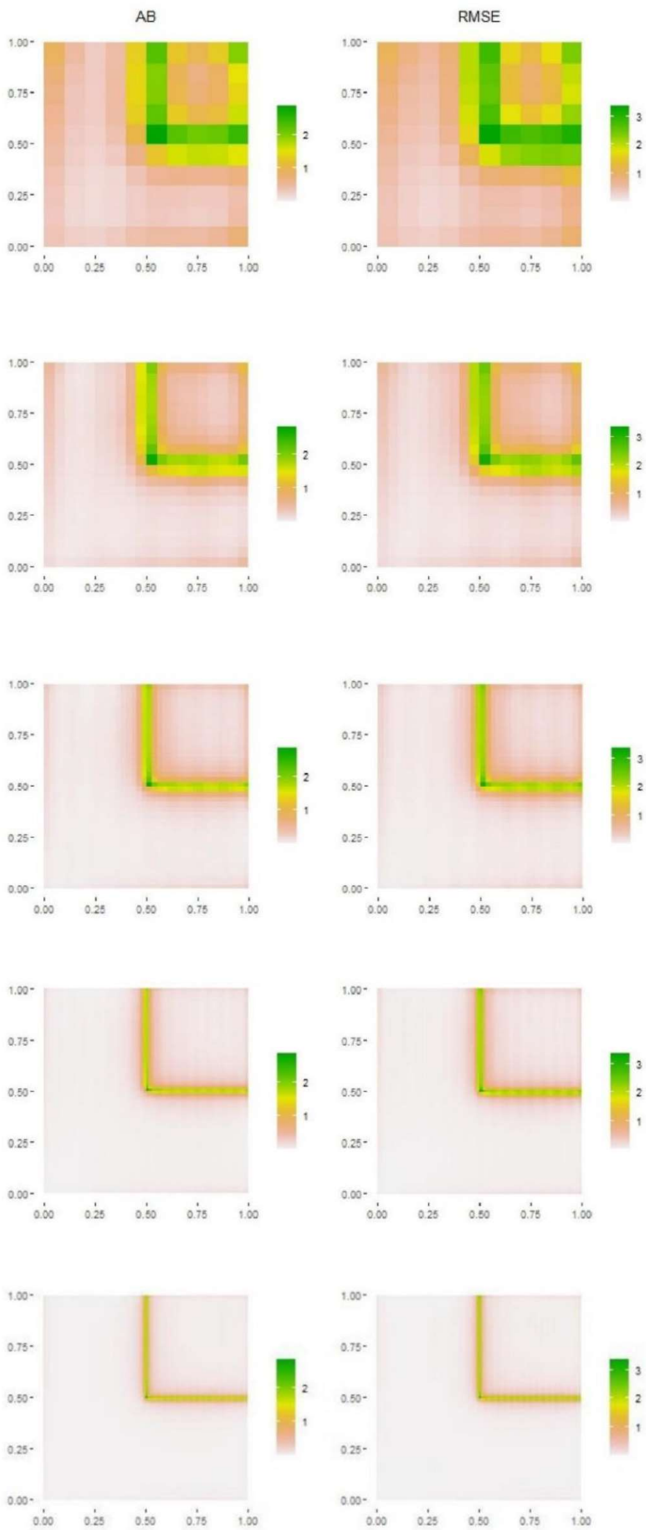


**Figure 22.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 3 partitioned into five populations of areas, under OPSS with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

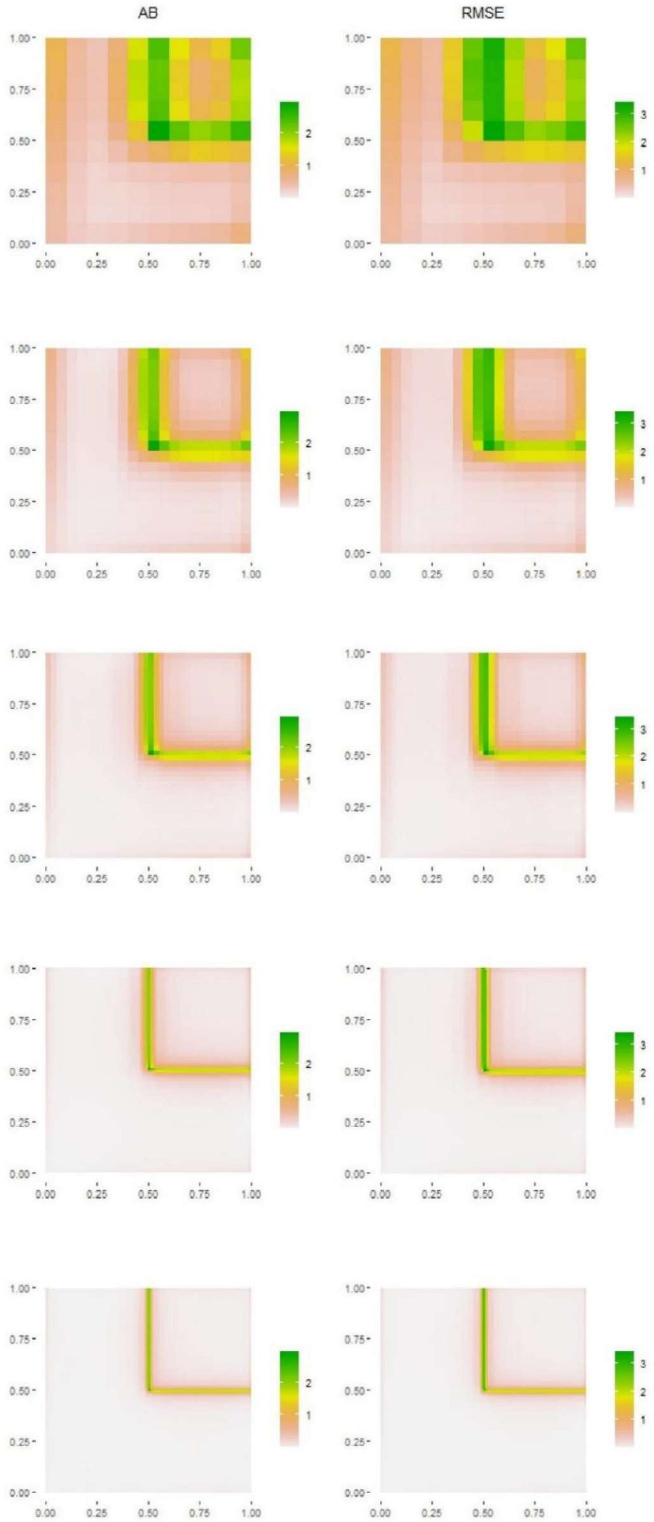


**Figure 23.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator achieved for surface 3 partitioned into five populations of areas, under SYS with sampling fraction of 10% and population size $N = 100, 400, 1{,}600, 6{,}400, 25{,}600$.

**Table 11.** Minima, averages, and maxima of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of growing stock volume densities in a rectangular region located in Calabria and partitioned into three populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | AB | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max |
| SRSWOR | 250 | 0.40 | 38.05 | 111.64 | 16.72 | 56.54 | 128.46 |
| | 1,000 | 0.03 | 26.27 | 90.90 | 4.55 | 38.42 | 113.06 |
| | 4,000 | 0.00 | 16.05 | 65.17 | 2.22 | 23.73 | 83.47 |
| OPSS | 250 | 0.16 | 34.79 | 107.69 | 8.89 | 50.20 | 117.75 |
| | 1,000 | 0.01 | 23.38 | 87.11 | 3.69 | 33.64 | 103.23 |
| | 4,000 | 0.01 | 14.06 | 60.17 | 1.57 | 20.28 | 73.66 |
| SYS | 250 | 0.03 | 37.07 | 105.86 | 7.77 | 53.15 | 121.68 |
| | 1,000 | 0.00 | 24.64 | 92.84 | 1.86 | 34.89 | 102.63 |
| | 4,000 | 0.01 | 14.37 | 72.68 | 0.86 | 20.18 | 83.81 |

**Table 12.** Minima, averages, and maxima of correction factor (RATIO), absolute bias ($AB^*$) and root mean squared error ($RMSE^*$) of the harmonized interpolator (with the overall total estimate) of growing stock volume densities in a rectangular region located in Calabria and partitioned into three populations of areas, for each combination of sampling scheme with sampling fraction of 10% and population size ($N$).

| Scheme | $N$ | RATIO | | | $AB^*$ | | | $RMSE^*$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 250 | 0.87 | 1.00 | 1.10 | 0.00 | 38.10 | 112.67 | 18.94 | 57.80 | 130.04 |
| | 1,000 | 0.93 | 1.00 | 1.08 | 0.02 | 26.28 | 90.83 | 6.64 | 39.31 | 113.62 |
| | 4,000 | 0.95 | 1.00 | 1.04 | 0.01 | 16.05 | 65.15 | 3.51 | 24.22 | 83.70 |
| OPSS | 250 | 0.95 | 1.00 | 1.03 | 0.06 | 34.83 | 107.25 | 9.06 | 50.39 | 117.53 |
| | 1,000 | 0.98 | 1.00 | 1.02 | 0.01 | 23.38 | 87.08 | 3.89 | 33.71 | 103.29 |
| | 4,000 | 0.99 | 1.00 | 1.01 | 0.01 | 14.06 | 60.16 | 1.63 | 20.30 | 73.68 |
| SYS | 250 | 0.96 | 0.99 | 1.02 | 0.16 | 37.08 | 107.62 | 7.33 | 54.19 | 121.53 |
| | 1,000 | 0.99 | 1.00 | 1.01 | 0.00 | 24.65 | 92.66 | 1.81 | 35.00 | 103.26 |
| | 4,000 | 1.00 | 1.00 | 1.00 | 0.00 | 14.36 | 72.50 | 0.83 | 20.21 | 84.11 |

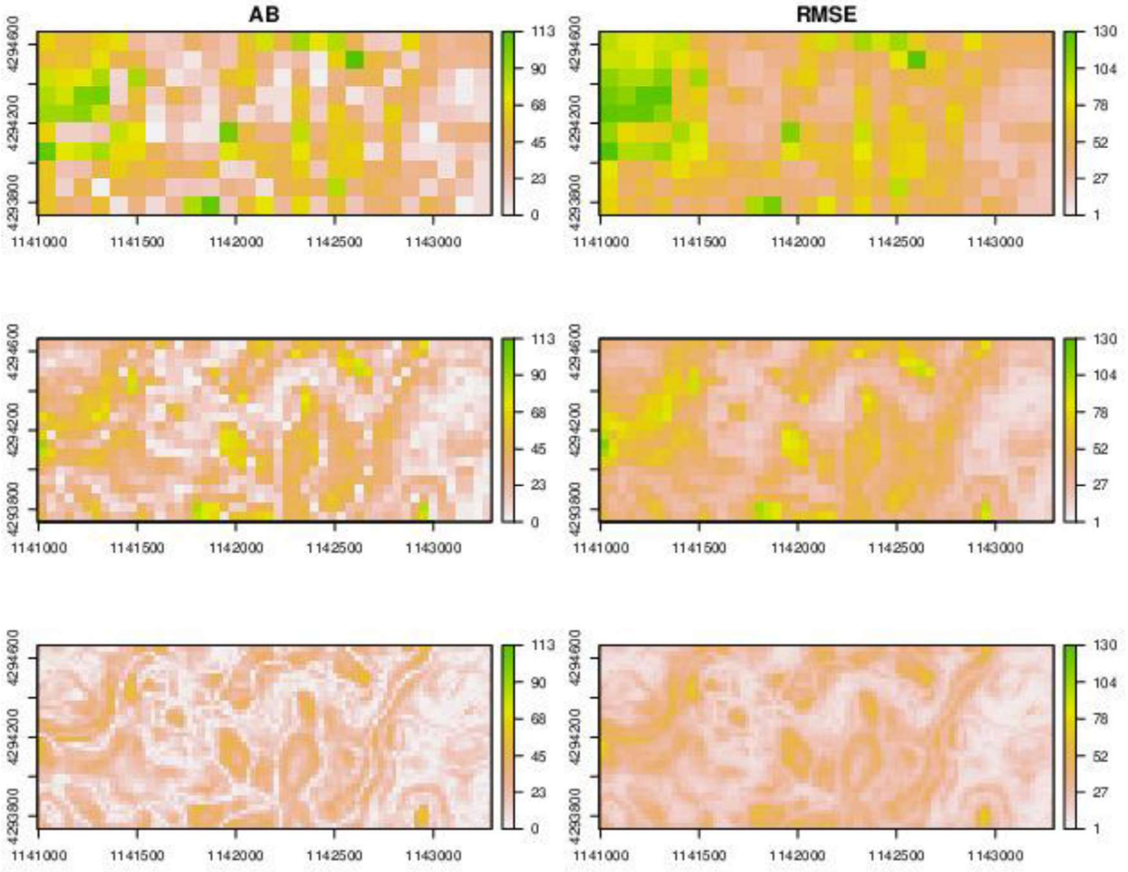


**Figure 24.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of growing stock volume densities in a rectangular region located in Calabria and partitioned into three populations of areas, under SRSWOR with sampling fraction of 10% and population size $N = 250, 1{,}000, 4{,}000$.

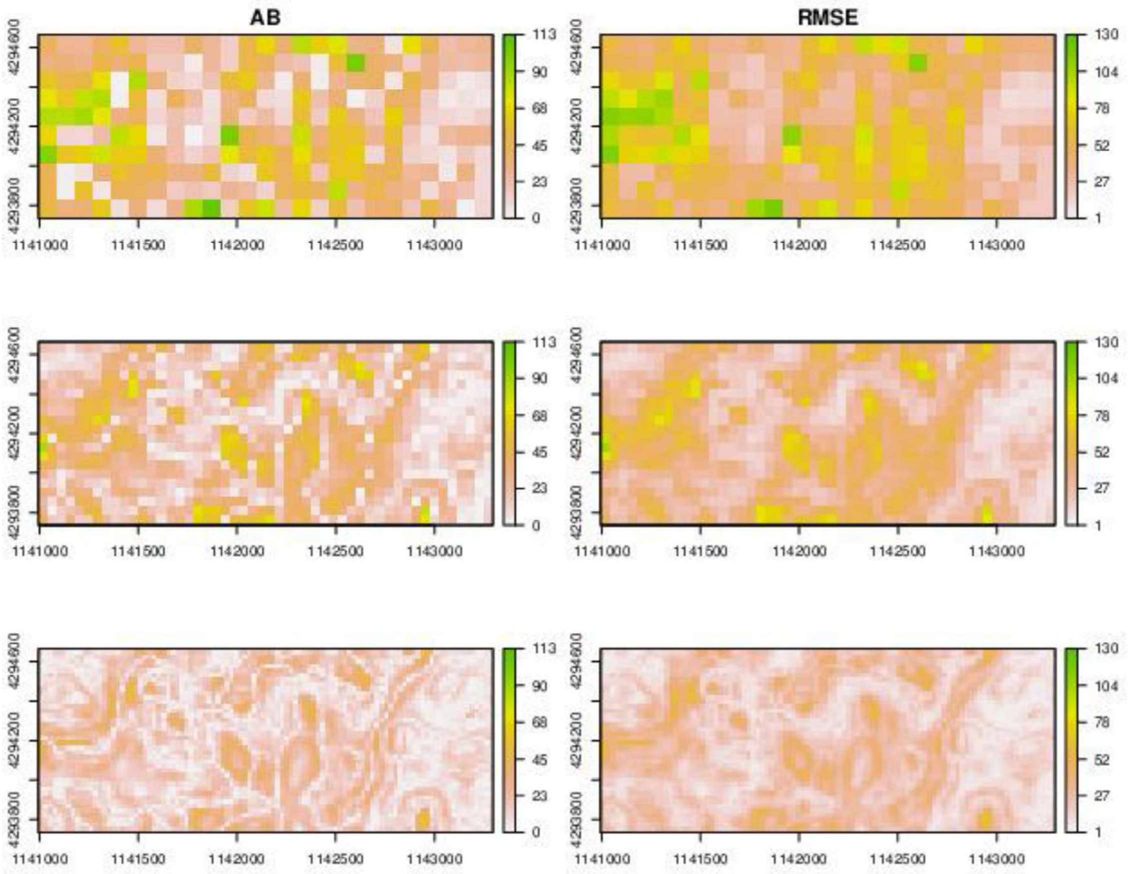


**Figure 25.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of growing stock volume densities in a rectangular region located in Calabria and partitioned into three populations of areas, under OPSS with sampling fraction of 10% and population size $N = 250, 1{,}000, 4{,}000$.

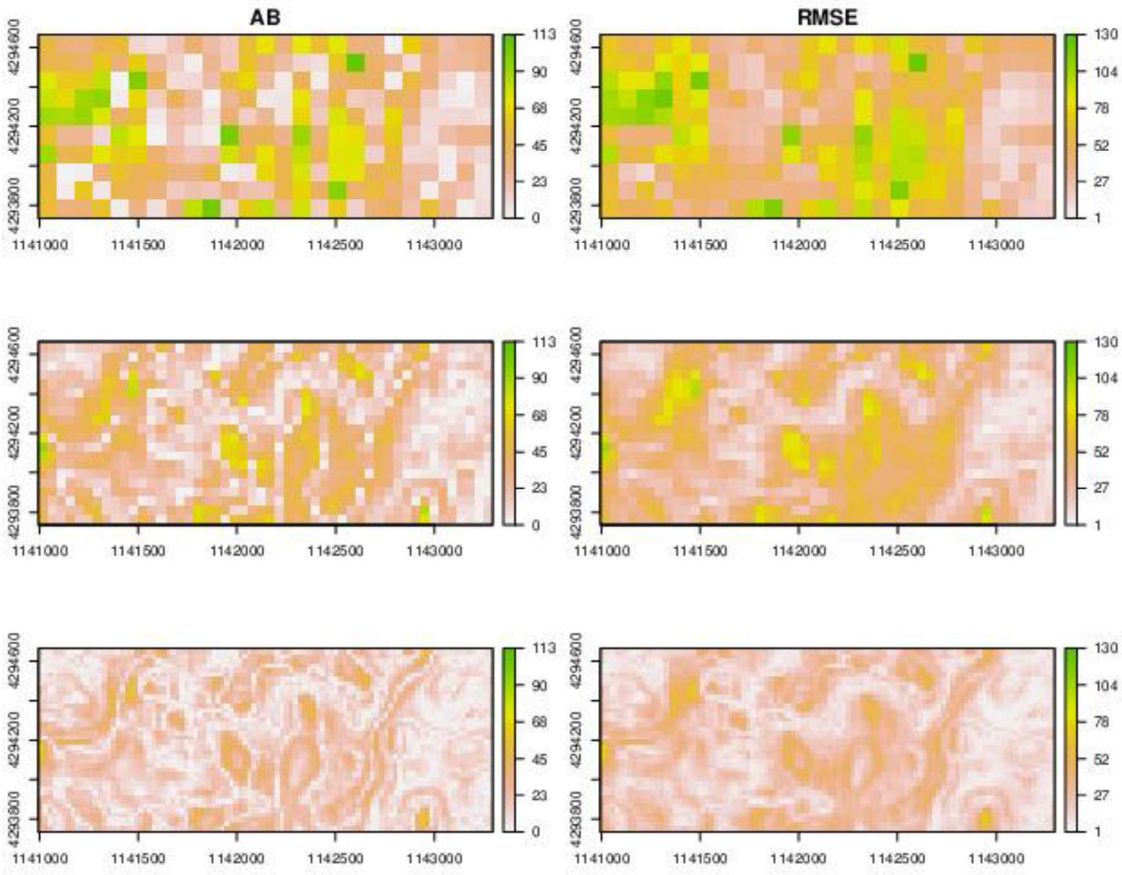


**Figure 26.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of growing stock volume densities in a rectangular region located in Calabria and partitioned into three populations of areas, under SYS with sampling fraction of 10% and population size $N = 250, 1{,}000, 4{,}000$.

**Table 13.** Minima, averages, and maxima of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of wood volumes of three nested populations of trees settled in a forested region of Val di Sella, under 3P sampling with sampling fraction of about 10%, for each population size ($N$) and maximum prediction error ($r$). $R^2$ is the r-square between tree volumes and predictions and $E(n)$ is the expected sample size.

| $N$ | $r$ | $R^2$ | $E(n)$ | AB | | | RMSE | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | min | mean | max | min | mean | max |
| 219 | 0.15 | 0.95 | 23.0 | 0.00 | 0.31 | 1.16 | 0.15 | 0.44 | 1.41 |
| 365 | | 0.96 | 37.1 | 0.00 | 0.31 | 1.15 | 0.13 | 0.43 | 1.40 |
| 608 | | 0.96 | 61.0 | 0.00 | 0.29 | 1.17 | 0.10 | 0.40 | 1.44 |
| 219 | 0.10 | 0.98 | 23.0 | 0.00 | 0.21 | 0.75 | 0.10 | 0.30 | 0.93 |
| 365 | | 0.98 | 37.1 | 0.00 | 0.21 | 0.75 | 0.09 | 0.29 | 0.92 |
| 608 | | 0.98 | 61.0 | 0.00 | 0.20 | 0.76 | 0.07 | 0.27 | 0.95 |
| 219 | 0.05 | 0.99 | 23.0 | 0.00 | 0.11 | 0.37 | 0.05 | 0.15 | 0.46 |
| 365 | | 1.00 | 37.1 | 0.00 | 0.11 | 0.36 | 0.05 | 0.14 | 0.45 |
| 608 | | 1.00 | 61.0 | 0.00 | 0.10 | 0.37 | 0.03 | 0.13 | 0.47 |

**Table 14.** Minima, averages, and maxima of correction factor (RATIO), absolute bias ($AB^*$) and root mean squared error ($RMSE^*$) of the harmonized interpolator (with the overall total estimate) of wood volumes of three nested populations of trees settled in a forested region of Val di Sella, under 3P sampling with sampling fraction of about 10%, for each population size ($N$) and maximum prediction error ($r$). $R^2$ is the r-square between tree volumes and predictions and $E(n)$ is the expected sample size.

| $N$ | $r$ | $R^2$ | $E(n)$ | RATIO | | | $AB^*$ | | | $RMSE^*$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | min | mean | max | min | mean | max | min | mean | max |
| 219 | 0.15 | 0.95 | 23.0 | 0.82 | 0.96 | 1.03 | 0.00 | 0.31 | 1.68 | 0.17 | 0.43 | 1.83 |
| 365 | | 0.96 | 37.1 | 0.88 | 0.96 | 1.00 | 0.00 | 0.30 | 1.68 | 0.12 | 0.42 | 1.84 |
| 608 | | 0.96 | 61.0 | 0.91 | 0.96 | 0.99 | 0.00 | 0.29 | 1.66 | 0.12 | 0.39 | 1.84 |
| 219 | 0.10 | 0.98 | 23.0 | 0.87 | 0.97 | 1.02 | 0.00 | 0.21 | 1.14 | 0.11 | 0.29 | 1.24 |
| 365 | | 0.98 | 37.1 | 0.92 | 0.97 | 1.00 | 0.00 | 0.21 | 1.14 | 0.08 | 0.28 | 1.25 |
| 608 | | 0.98 | 61.0 | 0.93 | 0.97 | 0.99 | 0.00 | 0.20 | 1.12 | 0.08 | 0.27 | 1.24 |
| 219 | 0.05 | 0.99 | 23.0 | 0.93 | 0.98 | 1.01 | 0.00 | 0.11 | 0.58 | 0.06 | 0.15 | 0.63 |
| 365 | | 1.00 | 37.1 | 0.96 | 0.98 | 1.00 | 0.00 | 0.11 | 0.58 | 0.04 | 0.15 | 0.63 |
| 608 | | 1.00 | 61.0 | 0.97 | 0.98 | 1.00 | 0.00 | 0.10 | 0.57 | 0.04 | 0.14 | 0.63 |

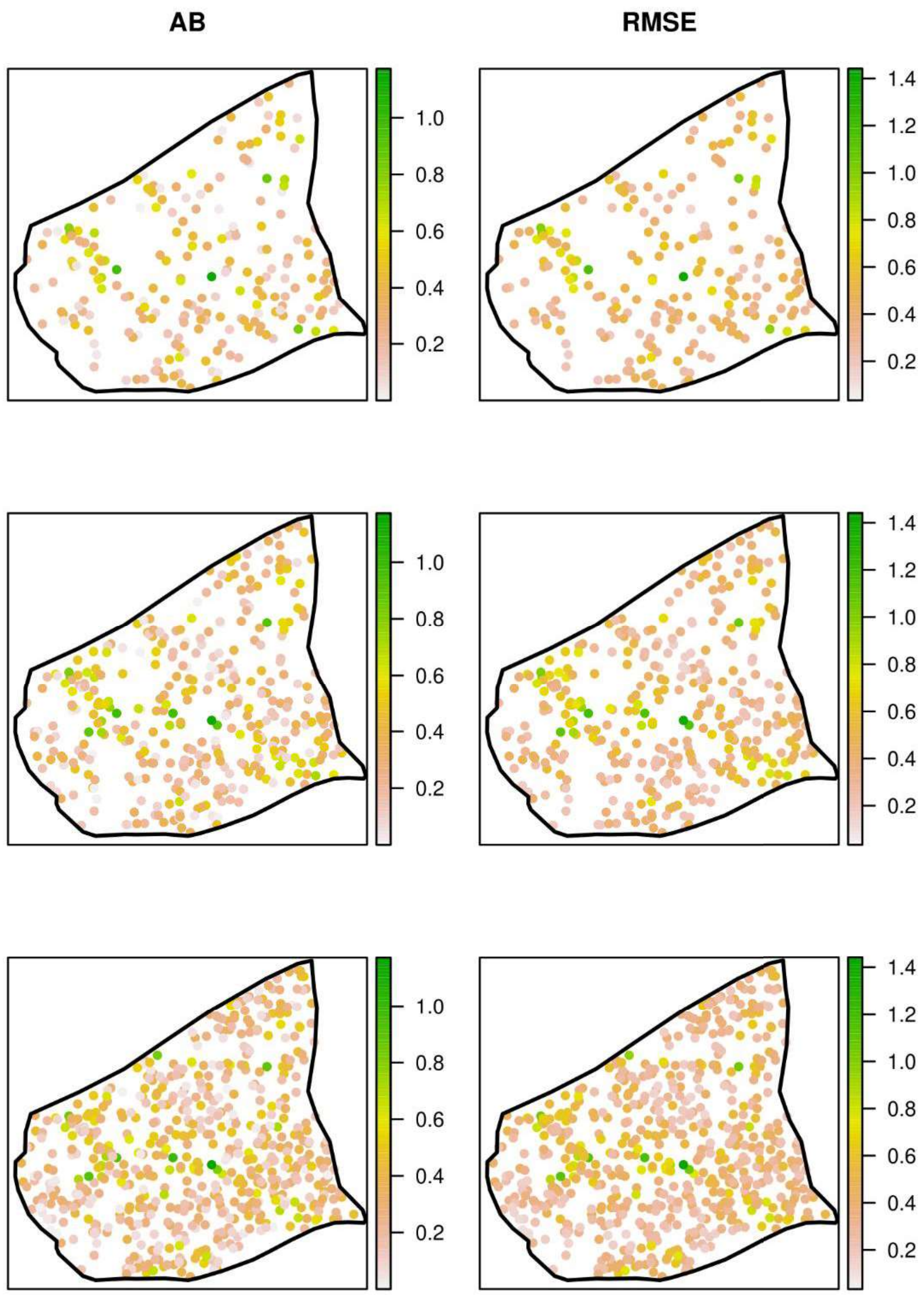


**Figure 27.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of wood volumes of three nested populations of trees settled in a forested region of Val di Sella, under 3P sampling with sampling fraction of about 10%, maximum prediction error $r = 0.15$ and population size $N = 219, 365, 608$.

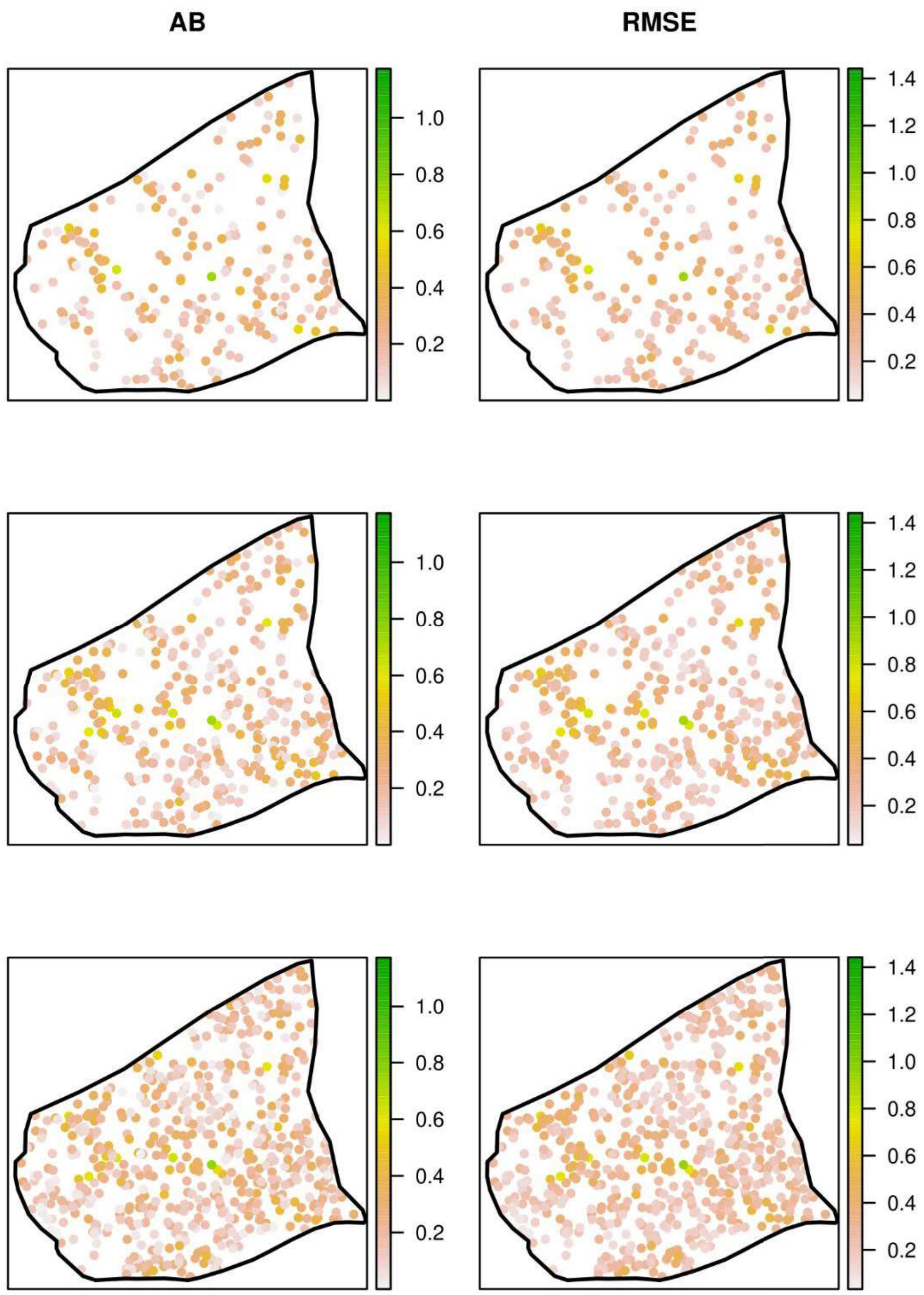


**Figure 28.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of wood volumes of three nested populations of trees settled in a forested region of Val di Sella, under 3P sampling with sampling fraction of about 10%, maximum prediction error $r = 0.10$ and population size $N = 219, 365, 608$.

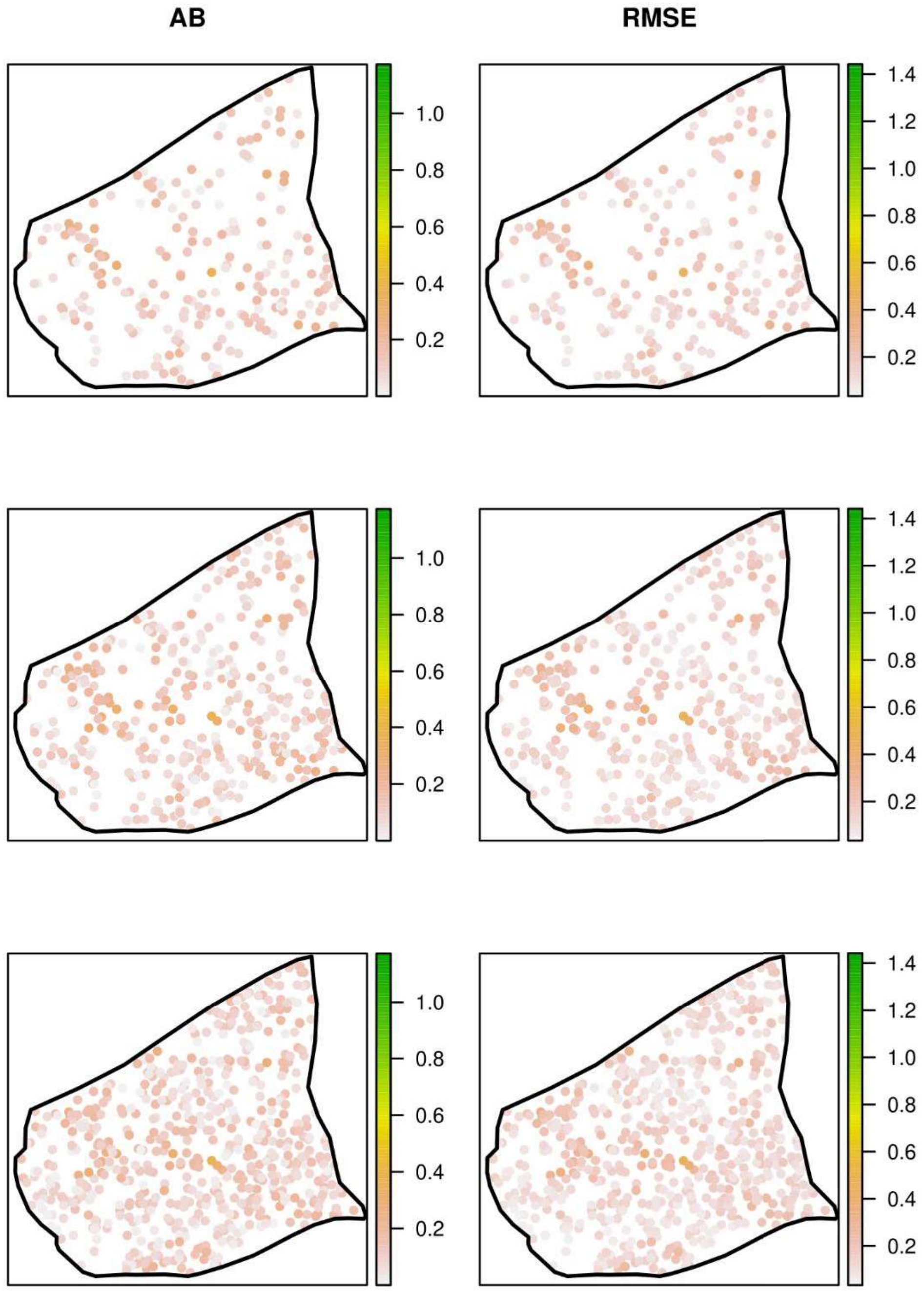


**Figure 29.** Maps of absolute bias (AB) and root mean squared error (RMSE) of the IDW interpolator of wood volumes of three nested populations of trees settled in a forested region of Val di Sella, under 3P sampling with sampling fraction of about 10%, maximum prediction error $r = 0.05$ and population size $N = 219, 365, 608$.

**Table 15.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the domain average estimates) of precipitations (expressed in mm) in a quadrat region located in North-Western Tuscany, for each combination of sampling scheme, sample size ($n$) and number of domains ($D$).

| Scheme | $D$ | $n$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | min | mean | max | min | mean | max | min | mean | max |
| URS | 2 | 16 | 0.70 | 0.98 | 1.24 | 0.00 | 30.69 | 136.12 | 20.79 | 52.82 | 152.51 |
| | | 36 | 0.80 | 0.99 | 1.15 | 0.00 | 21.82 | 112.19 | 13.15 | 37.78 | 126.61 |
| | | 64 | 0.87 | 0.99 | 1.12 | 0.00 | 16.95 | 95.37 | 9.45 | 29.48 | 108.71 |
| | | 100 | 0.90 | 1.00 | 1.10 | 0.00 | 13.74 | 83.92 | 7.754 | 24.07 | 96.13 |
| TSS | | 16 | 0.87 | 0.99 | 1.14 | 0.00 | 26.29 | 125.28 | 11.62 | 42.54 | 138.20 |
| | | 36 | 0.93 | 0.99 | 1.09 | 0.00 | 17.91 | 99.31 | 6.48 | 28.37 | 110.88 |
| | | 64 | 0.95 | 1.00 | 1.06 | 0.00 | 13.38 | 78.99 | 4.44 | 20.90 | 86.66 |
| | | 100 | 0.96 | 1.00 | 1.04 | 0.00 | 10.74 | 75.01 | 3.18 | 16.64 | 84.08 |
| SGS | | 16 | 0.94 | 0.99 | 1.04 | 0.00 | 24.02 | 114.83 | 5.21 | 36.71 | 124.40 |
| | | 36 | 0.97 | 1.00 | 1.03 | 0.00 | 15.70 | 93.74 | 2.95 | 23.85 | 105.42 |
| | | 64 | 0.98 | 1.00 | 1.02 | 0.00 | 11.59 | 78.96 | 2.47 | 17.54 | 89.40 |
| | | 100 | 0.98 | 1.00 | 1.02 | 0.00 | 9.22 | 68.22 | 2.07 | 13.92 | 77.55 |
| URS | 4 | 16 | 0.52 | 0.98 | 1.59 | 0.01 | 29.66 | 125.11 | 25.94 | 59.29 | 152.34 |
| | | 36 | 0.60 | 0.99 | 1.40 | 0.00 | 21.06 | 104.36 | 16.25 | 41.04 | 124.10 |
| | | 64 | 0.72 | 0.99 | 1.23 | 0.00 | 16.17 | 89.85 | 11.41 | 31.53 | 106.94 |
| | | 100 | 0.80 | 0.99 | 1.19 | 0.00 | 13.04 | 79.55 | 8.71 | 25.64 | 57.84 |
| TSS | | 16 | 0.64 | 0.98 | 1.26 | 0.00 | 25.14 | 120.36 | 16.60 | 47.22 | 139.09 |
| | | 36 | 0.81 | 0.99 | 1.16 | 0.00 | 17.09 | 94.30 | 9.28 | 30.44 | 108.94 |
| | | 64 | 0.85 | 1.00 | 1.11 | 0.00 | 12.71 | 76.44 | 5.77 | 22.04 | 84.69 |
| | | 100 | 0.89 | 1.00 | 1.09 | 0.00 | 10.11 | 72.31 | 4.20 | 17.30 | 82.34 |
| SGS | | 16 | 0.74 | 0.99 | 1.22 | 0.00 | 23.10 | 120.77 | 11.44 | 40.20 | 136.51 |
| | | 36 | 0.85 | 1.00 | 1.14 | 0.00 | 15.07 | 98.68 | 6.47 | 26.01 | 113.40 |
| | | 64 | 0.90 | 1.00 | 1.10 | 0.00 | 11.03 | 81.65 | 4.84 | 18.80 | 94.02 |
| | | 100 | 0.91 | 1.00 | 1.08 | 0.00 | 8.68 | 70.07 | 3.71 | 14.90 | 81.20 |
| URS | 8 | 16 | 0.52 | 0.98 | 1.82 | 0.00 | 27.32 | 121.95 | 21.75 | 61.57 | 144.86 |
| | | 36 | 0.56 | 0.98 | 1.68 | 0.00 | 20.47 | 101.55 | 14.47 | 44.28 | 119.25 |
| | | 64 | 0.61 | 0.99 | 1.53 | 0.00 | 16.23 | 86.49 | 10.06 | 33.72 | 104.68 |
| | | 100 | 0.63 | 0.99 | 1.40 | 0.00 | 13.54 | 79.25 | 7.72 | 27.43 | 94.67 |
| TSS | | 16 | 0.49 | 0.97 | 1.40 | 0.00 | 24.58 | 116.30 | 16.29 | 51.55 | 140.46 |
| | | 36 | 0.74 | 0.99 | 1.24 | 0.00 | 17.19 | 92.24 | 9.95 | 32.67 | 108.49 |
| | | 64 | 0.80 | 0.99 | 1.17 | 0.00 | 13.23 | 76.27 | 6.91 | 23.58 | 84.80 |
| | | 100 | 0.86 | 1.00 | 1.13 | 0.00 | 10.99 | 74.24 | 5.24 | 18.65 | 84.38 |
| SGS | | 16 | 0.66 | 0.97 | 1.35 | 0.00 | 22.87 | 132.15 | 8.16 | 44.69 | 152.24 |
| | | 36 | 0.81 | 0.99 | 1.21 | 0.00 | 15.35 | 100.70 | 5.24 | 28.33 | 116.58 |
| | | 64 | 0.85 | 0.99 | 1.15 | 0.00 | 11.79 | 80.88 | 4.48 | 20.79 | 93.80 |
| | | 100 | 0.87 | 1.00 | 1.13 | 0.00 | 9.76 | 68.57 | 3.52 | 16.85 | 80.54 |

**Table 16.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the domain total estimates) achieved for surface 1 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10%, population size ($N$) and number of domains ($D$).

| Scheme | $D$ | $N$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 2 | 100 | 0.05 | 0.99 | 2.11 | 0.02 | 0.45 | 1.57 | 1.32 | 2.03 | 3.37 |
| | | 400 | 0.31 | 1.00 | 1.69 | 0.00 | 0.20 | 1.01 | 0.59 | 1.02 | 1.96 |
| | | 1,600 | 0.63 | 1.00 | 1.36 | 0.00 | 0.09 | 0.63 | 0.27 | 0.51 | 1.11 |
| | | 6,400 | 0.84 | 1.00 | 1.17 | 0.00 | 0.04 | 0.36 | 0.12 | 0.25 | 0.62 |
| | | 25,600 | 0.92 | 1.00 | 1.08 | 0.00 | 0.02 | 0.21 | 0.06 | 0.12 | 0.33 |
| OPSS | | 100 | 0.57 | 1.00 | 1.38 | 0.00 | 0.36 | 1.39 | 0.76 | 1.16 | 2.09 |
| | | 400 | 0.93 | 1.00 | 1.06 | 0.00 | 0.15 | 0.80 | 0.14 | 0.30 | 0.99 |
| | | 1,600 | 0.98 | 1.00 | 1.02 | 0.00 | 0.06 | 0.46 | 0.06 | 0.13 | 0.57 |
| | | 6,400 | 0.99 | 1.00 | 1.01 | 0.00 | 0.03 | 0.27 | 0.03 | 0.06 | 0.32 |
| | | 25,600 | 1.00 | 1.00 | 1.00 | 0.00 | 0.01 | 0.15 | 0.01 | 0.03 | 0.18 |
| SYS | | 100 | 0.67 | 1.00 | 1.33 | 0.00 | 0.54 | 1.76 | 0.74 | 1.66 | 2.93 |
| | | 400 | 0.88 | 1.00 | 1.09 | 0.00 | 0.19 | 1.02 | 0.15 | 0.54 | 1.56 |
| | | 1,600 | 0.93 | 1.00 | 1.06 | 0.00 | 0.08 | 0.59 | 0.08 | 0.26 | 0.94 |
| | | 6,400 | 0.96 | 1.00 | 1.03 | 0.00 | 0.03 | 0.34 | 0.04 | 0.13 | 0.55 |
| | | 25,600 | 0.98 | 1.00 | 1.02 | 0.00 | 0.01 | 0.20 | 0.02 | 0.06 | 0.31 |
| SRSWOR | 4 | 100 | 0.11 | 0.98 | 3.11 | 0.01 | 0.33 | 1.42 | 1.57 | 2.72 | 4.38 |
| | | 400 | 0.09 | 0.99 | 2.11 | 0.00 | 0.15 | 0.95 | 0.76 | 1.54 | 2.69 |
| | | 1,600 | 0.49 | 0.99 | 1.57 | 0.00 | 0.07 | 0.61 | 0.33 | 0.77 | 1.42 |
| | | 6,400 | 0.69 | 1.00 | 1.28 | 0.00 | 0.03 | 0.36 | 0.15 | 0.38 | 0.79 |
| | | 25,600 | 0.86 | 1.00 | 1.14 | 0.00 | 0.01 | 0.21 | 0.07 | 0.19 | 0.43 |
| OPSS | | 100 | 0.12 | 0.99 | 2.20 | 0.01 | 0.28 | 1.35 | 1.07 | 2.12 | 3.56 |
| | | 400 | 0.48 | 0.99 | 1.53 | 0.00 | 0.12 | 0.79 | 0.37 | 0.80 | 1.49 |
| | | 1,600 | 0.76 | 1.00 | 1.19 | 0.00 | 0.05 | 0.45 | 0.14 | 0.31 | 0.73 |
| | | 6,400 | 0.92 | 1.00 | 1.07 | 0.00 | 0.02 | 0.26 | 0.06 | 0.11 | 0.35 |
| | | 25,600 | 0.97 | 1.00 | 1.03 | 0.00 | 0.01 | 0.15 | 0.02 | 0.04 | 0.19 |
| SYS | | 100 | 0.39 | 0.99 | 1.74 | 0.01 | 0.39 | 1.41 | 0.58 | 2.00 | 4.87 |
| | | 400 | 0.56 | 0.99 | 1.49 | 0.00 | 0.14 | 0.99 | 0.27 | 1.14 | 3.46 |
| | | 1,600 | 0.79 | 1.00 | 1.27 | 0.00 | 0.06 | 0.55 | 0.14 | 0.61 | 2.08 |
| | | 6,400 | 0.90 | 1.00 | 1.12 | 0.00 | 0.03 | 0.34 | 0.07 | 0.28 | 1.03 |
| | | 25,600 | 0.95 | 1.00 | 1.06 | 0.00 | 0.01 | 0.19 | 0.03 | 0.14 | 0.55 |
| SRSWOR | 8 | 100 | 0.31 | 0.98 | 2.87 | 0.00 | 0.31 | 1.3 | 1.32 | 2.32 | 3.82 |
| | | 400 | 0.06 | 0.99 | 3.36 | 0.00 | 0.14 | 0.85 | 0.98 | 2.24 | 3.82 |
| | | 1,600 | 0.17 | 0.99 | 2.13 | 0.00 | 0.06 | 0.58 | 0.44 | 1.13 | 2.01 |
| | | 6,400 | 0.57 | 1.00 | 1.43 | 0.00 | 0.03 | 0.35 | 0.19 | 0.56 | 1.06 |
| | | 25,600 | 0.80 | 1.00 | 1.21 | 0.00 | 0.01 | 0.2 | 0.09 | 0.28 | 0.56 |
| OPSS | | 100 | 0.32 | 0.98 | 1.94 | 0.00 | 0.46 | 1.68 | 1.34 | 2.15 | 3.63 |
| | | 400 | 0.30 | 0.99 | 1.69 | 0.00 | 0.11 | 0.76 | 0.44 | 1.09 | 2.05 |
| | | 1,600 | 0.72 | 1.00 | 1.25 | 0.00 | 0.04 | 0.43 | 0.16 | 0.40 | 0.92 |
| | | 6,400 | 0.89 | 1.00 | 1.11 | 0.00 | 0.02 | 0.25 | 0.06 | 0.14 | 0.37 |
| | | 25,600 | 0.96 | 1.00 | 1.04 | 0.00 | 0.01 | 0.15 | 0.03 | 0.05 | 0.19 |
| SYS | | 100 | 0.47 | 0.98 | 1.77 | 0.02 | 0.59 | 1.58 | 0.89 | 2.07 | 3.70 |
| | | 400 | 0.32 | 0.99 | 1.71 | 0.00 | 0.15 | 1.11 | 0.48 | 1.84 | 3.82 |
| | | 1,600 | 0.67 | 1.00 | 1.32 | 0.00 | 0.06 | 0.61 | 0.18 | 0.89 | 2.19 |
| | | 6,400 | 0.84 | 1.00 | 1.15 | 0.00 | 0.02 | 0.34 | 0.08 | 0.42 | 1.07 |
| | | 25,600 | 0.92 | 1.00 | 1.07 | 0.00 | 0.01 | 0.19 | 0.04 | 0.21 | 0.57 |

**Table 17.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the domain total estimates) achieved for surface 2 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10%, population size ($N$) and number of domains ($D$).

| Scheme | $D$ | $N$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 2 | 100 | 0.00 | 0.89 | 2.08 | 0.09 | 1.43 | 3.47 | 1.62 | 2.55 | 4.66 |
| | | 400 | 0.27 | 0.95 | 1.74 | 0.00 | 0.78 | 1.97 | 0.78 | 1.45 | 2.85 |
| | | 1,600 | 0.58 | 0.98 | 1.38 | 0.00 | 0.37 | 1.23 | 0.33 | 0.74 | 1.55 |
| | | 6,400 | 0.78 | 0.99 | 1.22 | 0.00 | 0.17 | 0.81 | 0.12 | 0.36 | 1.00 |
| | | 25,600 | 0.89 | 1.00 | 1.11 | 0.00 | 0.08 | 0.51 | 0.04 | 0.18 | 0.64 |
| OPSS | | 100 | 0.17 | 0.91 | 1.51 | 0.06 | 1.26 | 3.18 | 1.13 | 2.16 | 4.13 |
| | | 400 | 0.80 | 0.97 | 1.12 | 0.00 | 0.62 | 1.66 | 0.42 | 0.94 | 1.97 |
| | | 1,600 | 0.93 | 0.99 | 1.04 | 0.00 | 0.29 | 1.13 | 0.16 | 0.43 | 1.36 |
| | | 6,400 | 0.98 | 1.00 | 1.01 | 0.00 | 0.13 | 0.70 | 0.06 | 0.20 | 0.85 |
| | | 25,600 | 0.99 | 1.00 | 1.00 | 0.00 | 0.06 | 0.42 | 0.02 | 0.10 | 0.51 |
| SYS | | 100 | 0.57 | 0.88 | 1.20 | 0.02 | 1.35 | 3.25 | 0.62 | 1.95 | 3.74 |
| | | 400 | 0.80 | 0.95 | 1.07 | 0.00 | 0.73 | 2.37 | 0.26 | 1.03 | 2.80 |
| | | 1,600 | 0.91 | 0.98 | 1.04 | 0.00 | 0.32 | 1.68 | 0.14 | 0.49 | 2.01 |
| | | 6,400 | 0.96 | 0.99 | 1.02 | 0.00 | 0.14 | 1.07 | 0.06 | 0.23 | 1.30 |
| | | 25,600 | 0.98 | 1.00 | 1.01 | 0.00 | 0.07 | 0.65 | 0.03 | 0.11 | 0.80 |
| SRSWOR | 4 | 100 | 0.00 | 0.87 | 3.28 | 0.02 | 1.38 | 4.03 | 1.93 | 3.06 | 6.05 |
| | | 400 | 0.00 | 0.94 | 2.38 | 0.00 | 0.76 | 2.40 | 0.87 | 1.75 | 3.95 |
| | | 1,600 | 0.29 | 0.98 | 1.69 | 0.00 | 0.37 | 1.23 | 0.34 | 0.90 | 2.20 |
| | | 6,400 | 0.63 | 0.99 | 1.37 | 0.00 | 0.17 | 0.81 | 0.12 | 0.44 | 1.15 |
| | | 25,600 | 0.80 | 1.00 | 1.18 | 0.00 | 0.08 | 0.51 | 0.04 | 0.22 | 0.64 |
| OPSS | | 100 | 0.00 | 0.89 | 2.89 | 0.00 | 1.24 | 3.72 | 1.41 | 2.61 | 5.92 |
| | | 400 | 0.22 | 0.96 | 1.82 | 0.01 | 0.61 | 1.98 | 0.45 | 1.24 | 3.18 |
| | | 1,600 | 0.60 | 0.99 | 1.37 | 0.00 | 0.28 | 1.13 | 0.16 | 0.53 | 1.49 |
| | | 6,400 | 0.85 | 1.00 | 1.12 | 0.00 | 0.13 | 0.70 | 0.06 | 0.23 | 0.85 |
| | | 25,600 | 0.95 | 1.00 | 1.04 | 0.00 | 0.06 | 0.42 | 0.02 | 0.10 | 0.51 |
| SYS | | 100 | 0.38 | 0.88 | 1.46 | 0.04 | 1.53 | 4.63 | 0.44 | 2.09 | 5.09 |
| | | 400 | 0.61 | 0.95 | 1.22 | 0.00 | 0.68 | 2.16 | 0.12 | 0.99 | 2.47 |
| | | 1,600 | 0.86 | 0.98 | 1.08 | 0.00 | 0.31 | 1.61 | 0.05 | 0.47 | 1.89 |
| | | 6,400 | 0.96 | 0.99 | 1.03 | 0.00 | 0.14 | 1.05 | 0.01 | 0.21 | 1.27 |
| | | 25,600 | 0.98 | 1.00 | 1.01 | 0.00 | 0.07 | 0.65 | 0.00 | 0.10 | 0.79 |
| SRSWOR | 8 | 100 | 0.01 | 0.83 | 3.94 | 0.00 | 1.33 | 3.92 | 1.72 | 3.24 | 5.78 |
| | | 400 | 0.00 | 0.93 | 4.31 | 0.00 | 0.72 | 2.46 | 0.87 | 2.20 | 4.90 |
| | | 1,600 | 0.12 | 0.97 | 2.34 | 0.00 | 0.36 | 1.26 | 0.34 | 1.14 | 2.90 |
| | | 6,400 | 0.44 | 0.99 | 1.67 | 0.00 | 0.17 | 0.81 | 0.12 | 0.57 | 1.54 |
| | | 25,600 | 0.69 | 1.00 | 1.29 | 0.00 | 0.08 | 0.51 | 0.04 | 0.28 | 0.80 |
| OPSS | | 100 | 0.01 | 0.82 | 2.86 | 0.00 | 1.26 | 4.65 | 1.36 | 3.19 | 6.32 |
| | | 400 | 0.08 | 0.95 | 2.20 | 0.00 | 0.60 | 2.02 | 0.47 | 1.44 | 3.95 |
| | | 1,600 | 0.54 | 0.98 | 1.45 | 0.00 | 0.28 | 1.14 | 0.16 | 0.60 | 1.76 |
| | | 6,400 | 0.81 | 0.99 | 1.18 | 0.00 | 0.13 | 0.70 | 0.06 | 0.25 | 0.85 |
| | | 25,600 | 0.93 | 1.00 | 1.06 | 0.00 | 0.06 | 0.42 | 0.02 | 0.11 | 0.52 |
| SYS | | 100 | 0.05 | 0.86 | 1.68 | 0.01 | 1.59 | 4.95 | 0.40 | 2.32 | 5.84 |
| | | 400 | 0.18 | 0.94 | 1.95 | 0.00 | 0.69 | 2.24 | 0.41 | 1.65 | 5.16 |
| | | 1,600 | 0.58 | 0.98 | 1.37 | 0.00 | 0.31 | 1.63 | 0.13 | 0.77 | 2.81 |
| | | 6,400 | 0.79 | 0.99 | 1.18 | 0.00 | 0.14 | 1.06 | 0.06 | 0.37 | 1.48 |
| | | 25,600 | 0.90 | 1.00 | 1.09 | 0.00 | 0.07 | 0.65 | 0.02 | 0.18 | 0.79 |

**Table 18.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the domain total estimates) achieved for surface 3 partitioned into five populations of areas, for each combination of sampling scheme with sampling fraction of 10%, population size ($N$) and number of domains ($D$).

| Scheme | $D$ | $N$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 2 | 100 | 0.00 | 0.86 | 2.24 | 0.01 | 0.63 | 2.36 | 0.24 | 1.58 | 4.23 |
| | | 400 | 0.08 | 0.88 | 1.89 | 0.00 | 0.35 | 2.67 | 0.10 | 0.87 | 3.36 |
| | | 1,600 | 0.27 | 0.92 | 1.41 | 0.00 | 0.22 | 2.80 | 0.04 | 0.49 | 3.30 |
| | | 6,400 | 0.64 | 0.95 | 1.26 | 0.00 | 0.14 | 2.84 | 0.02 | 0.28 | 3.28 |
| | | 25,600 | 0.79 | 0.97 | 1.10 | 0.00 | 0.08 | 2.92 | 0.01 | 0.16 | 3.33 |
| OPSS | | 100 | 0.11 | 0.87 | 1.68 | 0.01 | 0.47 | 2.39 | 0.14 | 0.98 | 3.21 |
| | | 400 | 0.58 | 0.89 | 1.19 | 0.00 | 0.29 | 2.59 | 0.04 | 0.47 | 3.04 |
| | | 1,600 | 0.77 | 0.93 | 1.07 | 0.00 | 0.18 | 2.68 | 0.02 | 0.27 | 3.09 |
| | | 6,400 | 0.87 | 0.96 | 1.03 | 0.00 | 0.12 | 2.76 | 0.01 | 0.16 | 3.14 |
| | | 25,600 | 0.93 | 0.97 | 1.01 | 0.00 | 0.07 | 2.79 | 0.00 | 0.09 | 3.15 |
| SYS | | 100 | 0.22 | 0.87 | 1.66 | 0.00 | 0.58 | 2.78 | 0.10 | 1.20 | 3.70 |
| | | 400 | 0.48 | 0.90 | 1.27 | 0.00 | 0.34 | 2.91 | 0.04 | 0.62 | 3.18 |
| | | 1,600 | 0.67 | 0.93 | 1.14 | 0.00 | 0.20 | 2.94 | 0.02 | 0.36 | 3.25 |
| | | 6,400 | 0.81 | 0.96 | 1.07 | 0.00 | 0.12 | 2.95 | 0.01 | 0.21 | 3.28 |
| | | 25,600 | 0.89 | 0.97 | 1.04 | 0.00 | 0.07 | 2.96 | 0.00 | 0.12 | 3.30 |
| SRSWOR | 4 | 100 | 0.00 | 0.81 | 3.58 | 0.00 | 0.71 | 2.35 | 0.22 | 1.94 | 5.39 |
| | | 400 | 0.01 | 0.86 | 2.31 | 0.00 | 0.42 | 2.71 | 0.1 | 1.12 | 3.51 |
| | | 1,600 | 0.24 | 0.92 | 1.65 | 0.00 | 0.25 | 2.84 | 0.04 | 0.61 | 3.36 |
| | | 6,400 | 0.58 | 0.96 | 1.33 | 0.00 | 0.14 | 2.87 | 0.02 | 0.34 | 3.31 |
| | | 25,600 | 0.79 | 0.98 | 1.18 | 0.01 | 0.08 | 2.95 | 0.01 | 0.18 | 3.35 |
| OPSS | | 100 | 0.00 | 0.82 | 2.39 | 0.00 | 0.57 | 2.54 | 0.14 | 1.56 | 4.50 |
| | | 400 | 0.15 | 0.88 | 1.56 | 0.00 | 0.33 | 2.63 | 0.05 | 0.69 | 3.13 |
| | | 1,600 | 0.54 | 0.93 | 1.24 | 0.00 | 0.20 | 2.72 | 0.02 | 0.35 | 3.13 |
| | | 6,400 | 0.80 | 0.97 | 1.09 | 0.00 | 0.12 | 2.79 | 0.01 | 0.18 | 3.16 |
| | | 25,600 | 0.92 | 0.98 | 1.03 | 0.00 | 0.07 | 2.81 | 0.00 | 0.10 | 3.16 |
| SYS | | 100 | 0.21 | 0.81 | 1.91 | 0.01 | 0.58 | 2.98 | 0.08 | 1.41 | 4.82 |
| | | 400 | 0.43 | 0.87 | 1.51 | 0.00 | 0.37 | 3.16 | 0.02 | 0.85 | 3.50 |
| | | 1,600 | 0.72 | 0.93 | 1.25 | 0.00 | 0.22 | 3.08 | 0.01 | 0.46 | 3.29 |
| | | 6,400 | 0.88 | 0.96 | 1.12 | 0.00 | 0.13 | 3.03 | 0.01 | 0.25 | 3.30 |
| | | 25,600 | 0.95 | 0.98 | 1.06 | 0.00 | 0.07 | 3.00 | 0.00 | 0.14 | 3.31 |
| SRSWOR | 8 | 100 | 0.01 | 0.76 | 3.21 | 0.00 | 0.30 | 2.05 | 0.20 | 1.28 | 4.31 |
| | | 400 | 0.00 | 0.82 | 3.15 | 0.00 | 0.24 | 2.32 | 0.11 | 1.15 | 3.89 |
| | | 1,600 | 0.06 | 0.88 | 2.13 | 0.00 | 0.19 | 2.64 | 0.05 | 0.64 | 3.29 |
| | | 6,400 | 0.43 | 0.93 | 1.45 | 0.00 | 0.14 | 2.75 | 0.02 | 0.36 | 3.25 |
| | | 25,600 | 0.63 | 0.96 | 1.24 | 0.00 | 0.09 | 2.88 | 0.01 | 0.20 | 3.31 |
| OPSS | | 100 | 0.01 | 0.77 | 2.42 | 0.01 | 0.33 | 1.99 | 0.17 | 1.12 | 3.53 |
| | | 400 | 0.04 | 0.83 | 1.87 | 0.00 | 0.24 | 2.37 | 0.05 | 0.67 | 3.00 |
| | | 1,600 | 0.47 | 0.89 | 1.35 | 0.00 | 0.18 | 2.56 | 0.02 | 0.33 | 3.02 |
| | | 6,400 | 0.77 | 0.94 | 1.14 | 0.00 | 0.12 | 2.70 | 0.01 | 0.18 | 3.10 |
| | | 25,600 | 0.89 | 0.97 | 1.05 | 0.00 | 0.08 | 2.75 | 0.00 | 0.10 | 3.12 |
| SYS | | 100 | 0.07 | 0.77 | 1.77 | 0.00 | 0.35 | 2.15 | 0.03 | 1.05 | 3.67 |
| | | 400 | 0.16 | 0.83 | 1.84 | 0.00 | 0.25 | 2.92 | 0.05 | 0.94 | 4.04 |
| | | 1,600 | 0.40 | 0.89 | 1.34 | 0.00 | 0.18 | 2.92 | 0.02 | 0.49 | 3.18 |
| | | 6,400 | 0.60 | 0.94 | 1.17 | 0.00 | 0.12 | 2.93 | 0.01 | 0.27 | 3.24 |
| | | 25,600 | 0.76 | 0.97 | 1.08 | 0.00 | 0.08 | 2.94 | 0.00 | 0.16 | 3.28 |

**Table 19.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the domain total estimates) of growing stock volume densities in a rectangular region located in Calabria and partitioned into three populations of areas, for each combination of sampling scheme with sampling fraction of 10%, population size ($N$) and number of domains ($D$).

| Scheme | $D$ | $N$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | min | mean | max | min | mean | max | min | mean | max |
| SRSWOR | 2 | 250 | 0.24 | 1.00 | 1.77 | 0.09 | 36.44 | 109.75 | 45.06 | 84.15 | 144.11 |
| | | 1,000 | 0.61 | 1.00 | 1.40 | 0.01 | 25.81 | 88.24 | 21.88 | 50.37 | 116.40 |
| | | 4,000 | 0.78 | 1.00 | 1.22 | 0.00 | 15.94 | 64.33 | 9.34 | 29.13 | 84.37 |
| OPSS | | 250 | 0.69 | 1.00 | 1.29 | 0.08 | 33.85 | 106.55 | 23.96 | 59.93 | 121.40 |
| | | 1,000 | 0.86 | 1.00 | 1.12 | 0.02 | 23.14 | 84.19 | 8.61 | 35.79 | 101.26 |
| | | 4,000 | 0.94 | 1.00 | 1.05 | 0.01 | 14.00 | 60.38 | 3.37 | 20.80 | 72.64 |
| SYS | | 250 | 0.72 | 1.00 | 1.25 | 0.01 | 36.10 | 105.74 | 40.98 | 84.16 | 138.32 |
| | | 1,000 | 0.86 | 1.00 | 1.12 | 0.02 | 24.36 | 88.68 | 15.53 | 47.80 | 110.02 |
| | | 4,000 | 0.93 | 1.00 | 1.06 | 0.02 | 14.30 | 70.54 | 5.49 | 26.12 | 86.79 |
| SRSWOR | 4 | 250 | 0.07 | 0.99 | 2.39 | 0.27 | 35.71 | 103.78 | 71.41 | 118.51 | 183.91 |
| | | 1,000 | 0.36 | 1.18 | 1.71 | 0.00 | 25.79 | 89.31 | 33.23 | 65.89 | 130.61 |
| | | 4,000 | 0.67 | 1.00 | 1.34 | 0.01 | 15.92 | 64.34 | 13.66 | 36.37 | 89.09 |
| OPSS | | 250 | 0.49 | 1.00 | 1.56 | 0.01 | 33.63 | 103.13 | 40.38 | 75.71 | 132.23 |
| | | 1,000 | 0.75 | 1.00 | 1.23 | 0.04 | 23.19 | 85.96 | 12.56 | 40.12 | 105.18 |
| | | 4,000 | 0.90 | 1.00 | 1.09 | 0.00 | 14.01 | 59.87 | 4.03 | 21.92 | 74.06 |
| SYS | | 250 | 0.56 | 0.99 | 1.46 | 0.00 | 35.92 | 101.57 | 46.15 | 100.45 | 157.18 |
| | | 1,000 | 0.77 | 1.00 | 1.21 | 0.03 | 24.29 | 88.54 | 13.29 | 54.44 | 120.29 |
| | | 4,000 | 0.89 | 1.00 | 1.11 | 0.00 | 14.28 | 70.45 | 3.71 | 29.56 | 91.89 |
| SRSWOR | 8 | 250 | 0.12 | 0.98 | 3.66 | 0.51 | 33.45 | 126.96 | 96.05 | 155.73 | 219.21 |
| | | 1,000 | 0.06 | 0.99 | 2.21 | 0.01 | 24.60 | 96.24 | 41.09 | 87.88 | 146.36 |
| | | 4,000 | 0.49 | 1.00 | 1.61 | 0.00 | 15.57 | 68.72 | 19.07 | 46.78 | 95.32 |
| OPSS | | 250 | 0.12 | 0.99 | 2.16 | 0.08 | 32.45 | 119.55 | 70.70 | 116.92 | 179.80 |
| | | 1,000 | 0.65 | 1.00 | 1.26 | 0.01 | 22.61 | 88.80 | 15.86 | 45.05 | 107.09 |
| | | 4,000 | 0.84 | 1.00 | 1.12 | 0.00 | 13.79 | 63.39 | 5.08 | 23.18 | 72.46 |
| SYS | | 250 | 0.26 | 0.99 | 1.94 | 0.08 | 34.82 | 125.75 | 60.30 | 125.50 | 193.35 |
| | | 1,000 | 0.68 | 1.00 | 1.26 | 0.12 | 23.73 | 98.20 | 24.95 | 70.55 | 133.87 |
| | | 4,000 | 0.83 | 1.00 | 1.13 | 0.02 | 14.04 | 66.25 | 12.38 | 37.27 | 94.40 |

**Table 20.** Minima, averages, and maxima of correction factor (RATIO), absolute bias (AB*) and root mean squared error (RMSE*) of the harmonized interpolator (with the domain total estimates) of wood volumes of three nested populations of trees settled in a forested region of Val di Sella, under 3P sampling with sampling fraction of about 10%, for each population size ($N$), maximum prediction error ($r$) and number of domains ($D$).

| $D$ | $r$ | $N$ | RATIO | | | AB* | | | RMSE* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | min | mean | max | min | mean | max | min | mean | max |
| 2 | 0.15 | 219 | 0.81 | 0.96 | 1.05 | 0.00 | 0.31 | 1.67 | 0.17 | 0.43 | 1.82 |
| | | 365 | 0.83 | 0.96 | 1.03 | 0.00 | 0.30 | 1.67 | 0.13 | 0.42 | 1.82 |
| | | 608 | 0.89 | 0.96 | 1.02 | 0.00 | 0.29 | 1.68 | 0.12 | 0.39 | 1.85 |
| | 0.10 | 219 | 0.87 | 0.97 | 1.03 | 0.00 | 0.21 | 1.13 | 0.11 | 0.29 | 1.24 |
| | | 365 | 0.88 | 0.97 | 1.02 | 0.00 | 0.21 | 1.13 | 0.09 | 0.28 | 1.23 |
| | | 608 | 0.92 | 0.97 | 1.01 | 0.00 | 0.20 | 1.14 | 0.08 | 0.27 | 1.25 |
| | 0.05 | 219 | 0.93 | 0.98 | 1.02 | 0.00 | 0.11 | 0.58 | 0.06 | 0.15 | 0.63 |
| | | 365 | 0.94 | 0.98 | 1.01 | 0.00 | 0.11 | 0.57 | 0.04 | 0.15 | 0.63 |
| | | 608 | 0.96 | 0.98 | 1.01 | 0.00 | 0.10 | 0.58 | 0.04 | 0.14 | 0.63 |
| 4 | 0.15 | 219 | 0.78 | 0.96 | 1.14 | 0.00 | 0.31 | 1.67 | 0.17 | 0.43 | 1.84 |
| | | 365 | 0.80 | 0.96 | 1.07 | 0.00 | 0.30 | 1.60 | 0.14 | 0.41 | 1.77 |
| | | 608 | 0.85 | 0.96 | 1.04 | 0.00 | 0.29 | 1.67 | 0.12 | 0.39 | 1.84 |
| | 0.10 | 219 | 0.84 | 0.97 | 1.09 | 0.01 | 0.21 | 1.14 | 0.11 | 0.29 | 1.25 |
| | | 365 | 0.86 | 0.97 | 1.05 | 0.00 | 0.21 | 1.08 | 0.09 | 0.28 | 1.20 |
| | | 608 | 0.90 | 0.97 | 1.03 | 0.00 | 0.20 | 1.14 | 0.09 | 0.27 | 1.25 |
| | 0.05 | 219 | 0.92 | 0.98 | 1.05 | 0.00 | 0.11 | 0.58 | 0.06 | 0.15 | 0.64 |
| | | 365 | 0.92 | 0.98 | 1.02 | 0.00 | 0.11 | 0.55 | 0.05 | 0.14 | 0.61 |
| | | 608 | 0.94 | 0.99 | 1.01 | 0.00 | 0.10 | 0.58 | 0.04 | 0.14 | 0.63 |
| 8 | 0.15 | 219 | 0.75 | 0.96 | 1.27 | 0.01 | 0.31 | 1.69 | 0.18 | 0.44 | 1.89 |
| | | 365 | 0.77 | 0.96 | 1.21 | 0.00 | 0.30 | 1.59 | 0.14 | 0.42 | 1.78 |
| | | 608 | 0.82 | 0.96 | 1.13 | 0.00 | 0.29 | 1.73 | 0.12 | 0.39 | 1.89 |
| | 0.10 | 219 | 0.83 | 0.97 | 1.18 | 0.00 | 0.21 | 1.15 | 0.12 | 0.30 | 1.28 |
| | | 365 | 0.84 | 0.97 | 1.14 | 0.00 | 0.21 | 1.07 | 0.10 | 0.28 | 1.20 |
| | | 608 | 0.88 | 0.97 | 1.09 | 0.00 | 0.20 | 1.18 | 0.08 | 0.27 | 1.29 |
| | 0.05 | 219 | 0.91 | 0.99 | 1.09 | 0.00 | 0.11 | 0.59 | 0.06 | 0.15 | 0.66 |
| | | 365 | 0.91 | 0.98 | 1.07 | 0.00 | 0.11 | 0.55 | 0.05 | 0.15 | 0.61 |
| | | 608 | 0.93 | 0.99 | 1.04 | 0.00 | 0.10 | 0.60 | 0.04 | 0.14 | 0.66 |

### 6.2 Case study results

The total living wood volume estimate achieved by means of the HT estimator (13) is equal to 133,603 $m^3$, whereas the total estimate arising from the map (14) is equal to 135,054 $m^3$. Therefore, there is a discrepancy between the two results of 1,451 $m^3$. The correction factor adopted to achieve the rescaled estimates results to be approximately 0.989. Once multiplied the IDW values of each population unit by the correction factor, the total volume estimate of the whole study area arising from the harmonized map coincides with that achieved by means of the HT criterion. Since the correction factor is very close to 1, harmonized estimates of wood volume density and the corresponding RMSEs estimates turn out to be quite similar to the non-harmonized ones. Harmonized estimates of the wood volume density range from 78 to 1,042 $m^3ha^{-1}$ with a mean value of 494 $m^3ha^{-1}$, and the corresponding RMSEs estimates range from 0 to 560 $m^3ha^{-1}$ with a mean value of 59 $m^3ha^{-1}$. Figure 30 reports the harmonized map of living wood volume densities together with the map of the RMSEs estimates.

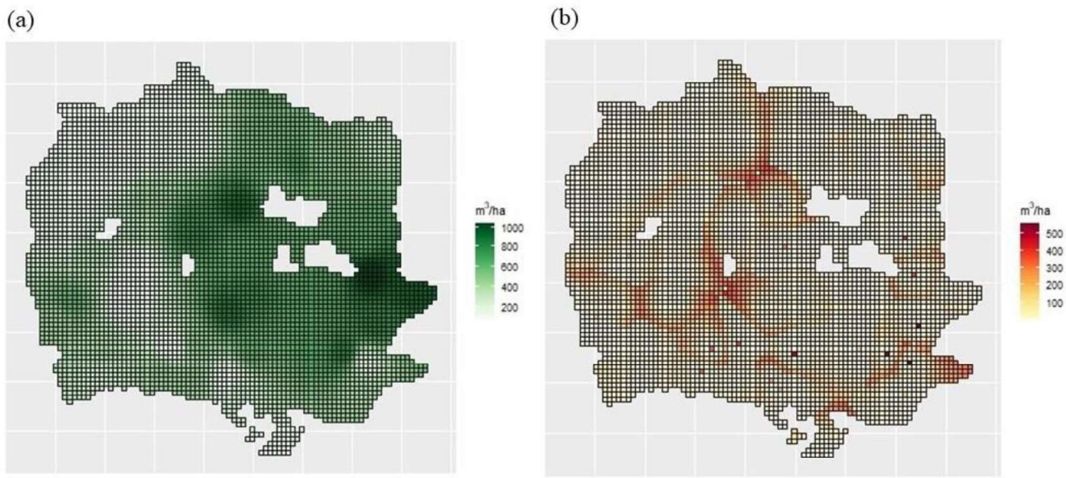


**Figure 30.** Harmonized map of living wood volume densities in Rincine (a) and map of the corresponding RMSEs estimates (b).

# Chapter 7

## Conclusions and Future Perspectives

The recent papers by Fattorini *et al.* (2018a, 2018b and 2019) proposed a novel approach for mapping continuous populations or finite populations of areas and points in a design-based framework, avoiding the complex task of modelling the surfaces to be mapped. However, in the design-based approach an unresolved problem takes rise.

Since totals and averages throughout the whole survey region or within domains are traditionally achieved by the HT or related criteria, these estimates invariably differ from the estimates achieved by the resulting maps. For avoiding these unsuitable discrepancies, that would appear awkward especially in a reporting phase, we here propose to rescale the resulting maps in such a way that the estimates arising from maps will match the traditional estimates. We also prove the asymptotic convergence of the two maps, i.e., the convergence to one of the rescaling factor. This is unambiguously confirmed by the simulation study, as the expectations of the rescaling factor quickly approach one as populations and/or sampled sizes increase and even for moderately small domains. Therefore, if we simply look at those expected values, we may argue that harmonization is a useless procedure because IDW and harmonized maps are very similar.

However, if we look at the minima and maxima of the rescaling constant reported in the previous section, it is apparent that in some situations the constant may vary from about 0.4 to 1.7, and the issue is even exacerbated when the estimation of domain totals is involved. In these cases, the rescaling constant may vary from about 0.0 to 4.3.

On the other hand, the deterioration of precision of harmonized maps when domain size decreases ought to warn against an acritical use of harmonization. Obviously, as the number of domains increases and their extents become small, the number of sample units available to perform estimation becomes small and in some case may be 0. This fact will necessarily introduce an increase in the variance of the estimator based on the HT or related criteria. As we cannot trust in these estimates, harmonization with respect to small domains should be avoided. These complications are well known in literature as small area estimation problems, for which an extensive literature has emerged, invariably of model-dependent nature (see e.g., Rao & Molina, 2015 and references therein). However, our design-based approach to mapping offers an alternative model-free solution to small area problems: we can simply exploit maps and estimate totals and averages from the interpolated values within. The resulting estimators are proven to be consistent under the most familiar sampling schemes of environmental surveys.

Finally, it should be noticed that we deliberately neglect the issue of selecting the $\alpha$ value in the distance function $d^{-\alpha}$. Actually, $\alpha$ is a smoothing parameter and it should be chosen by a data-driven procedure. However, because consistency of the IDW mapping is ensured for any $\alpha > 2$, including $\alpha = \infty$, in which case the IDW interpolation coincides with the nearest-neighbour interpolation (Fattorini *et al.*, 2021), the $\alpha$-choice should be irrelevant in investigating the effects of harmonization. Similarly, in simulation studies we neglected the problem of estimating the precision of the harmonized maps because it can be performed, *mutatis mutandis*, by the naïve procedure proposed by Fattorini *et al.* (2018a, 2018b, 2019) for the three types of populations.

## Acknowledgements

I gratefully acknowledge the Research and Innovation Centre Fondazione Edmund Mach (San Michele all' Adige, TN) for supporting this work and providing the funding received towards my Ph.D.

## References


Ahmed S, de Marsily G (1987) Comparison of Geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resour Res*. **23**:1717-1737.

Baffetta F, Corona P, Fattorini L (2012) A matching procedure to improve k-nn estimation of forest attribute maps. *Forest Ecol Manag.* **272**:35-50.

Barabesi L (2003) A Monte Carlo integration approach to Horvitz-Thompson estimation in replicated environmental designs. *Metron*. LXI, 355-374.

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.

Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr Anal*. **28**:281-298.

Brus DJ, Hengeveld GM, Walvoort DJJ, Goedhart PW, Heidema AH, Nabuurs GJ, Gunia K (2012) Statistical mapping of tree species over Europe. *Eur J For Res.* **131**:145-157.

Chirici G, Barbati A, Corona P, Marchetti M, Travaglini D, Maselli F, Bertini R (2008) Nonparametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sens Environ*. **112**:2686-2700.

Chirici G, Giannetti F, McRoberts RE, Travaglini D, Pecchi M, Maselli F, Chiesi M, Corona P (2020) Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data. *Int J Appl Earth Obs Geoinf*. **84**: online first.

Cleveland WS (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. *J Am Stat Assoc.* **74**:829-836.

Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc.* **83**:596-610.

Cordy CB (1993) An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Stat Probabil Lett*. **18**:353-362.

Corona P (2010) Integration of forest mapping and inventory to support forest management. *iForest* **3**:59-64.

Corona P (2016) Consolidating new paradigms in large-scale monitoring and assessment of forest ecosystems. *Environ Res*. **144**:8-14.

Corona P, Marchetti M (2007) Outlining multi-purpose forest inventories to assess the ecosystem approach in forestry. *Plant Biosyst*. **141**:243–251.

Corona P, Chirici G, McRoberts RE, Winter S, Barbati A (2011) Contribution of large-scale forest inventories to biodiversity assessment and monitoring. *For Ecol Manag*. **262**:2061-2069.

Corona P, Fattorini L, Franceschi S, Chirici G, Maselli F, Secondi L (2014) Mapping by spatial predictors exploiting remotely sensed and ground data: A comparative design-based perspective. *Remote Sens Environ*. **152**:29-37.

Cressie N (1993) Statistics for spatial data. New York: Wiley.

Deutsch C, Journel A (1992) *Geostatistical Software Library and User's Guide*, New York: Oxford University Press.

Di Biase RM, Fattorini L, Marchi M (2018) Statistical inferential techniques for approaching forest mapping. A review of methods. *Ann Silvic Res*. **42(2)**:46-58.

Food and Agriculture Organization of the United Nations (2010) Global Forest Resources Assessment 2010: Main Report, Rome: Food and Agriculture Organization of the United Nations.

Fattorini L, Marcheselli M, Pisani C (2006) A three-phase sampling strategy for large-scale multiresource forest inventories. *J Agric Biol Environ Stat*. **11**:296-316.

Fattorini L, Marcheselli M, Pisani C, Pratelli L (2018a) Design-based maps for continuous spatial populations. *Biometrika*. **105**:419-429.

Fattorini L, Marcheselli M, Pratelli L (2018b) Design-based maps for finite populations of spatial units. *J Am Stat Assoc*. **113**:686-697.

Fattorini L, Marcheselli M, Pisani C, Pratelli L (2019) Design-based mapping for finite population of marked points. *Electron J Stat*. **13**:2121-2149.

Fattorini L, Franceschi S, Corona P (2020a) Design-based mapping of tree attributes by 3P sampling. *Biom J*. **62**:1810-1825.

Fattorini L, Marcheselli M, Pisani C, Pratelli L (2020b) Design-based consistency of the Horvitz-Thompson estimator under spatial sampling with applications to environmental surveys. *Spat Stat*. **35**.

Fattorini L, Marcheselli M, Pisani C, Pratelli L (2021) Design-based properties of the nearest neighbour spatial interpolator and its bootstrap mean squared error estimator. *Biometrics*, 1-10.

Fix E, Hodges JL (1951) Discriminatory analysis. Nonparametric discrimination: consistency properties. Report Number 4, Project 21-49-004. USAF School of Aviation Medicine. Randolph Field, Texas.

Foody GM (2003) Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship. *Remote Sens Environ*. **88**:283-293.

Franco-Lopez H, Ek AR, Bauer ME (2001) Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens Environ*. **77**:251-274.

Freeman EA, Moisen GG (2007) Evaluating kriging as a Tool to Improve to Improve Moderate Resolution Maps of Forest Biomass. *Environ. Monit. Assess.* **128**: 395-410.

Gauch HG (1982) *Multivariate analysis in community ecology*, New York: Cambridge University Press.

Goovaerts P (1997) Geostatistics for natural resources evaluation. New York: Oxford university press.

Greaver TL, Clark CM, Compton JE, Vallano D, Talhem AF, Weaver CP, Band LE, Baron, JS, Davidson EA, Tague CL, Felker-Quinn E, Lynch JA, Herrick JD, Liu L, Goodale CL, Novak KJ, Haeuber RA (2021) Key ecological responses to nitrogen are related by climate change. *Nat Clim Change*. **6**:836-843.

Gregoire TG (1998) Design-Based and Model-Based Inference in Survey Sampling: Appreciating the Difference. *Can J For Res*. **28**:1429-1447.

Gregoire TG, Valentine H T (2008) Sampling Strategies for Natural Resources and the Environment. Boca Raton: Chapman and Hall.

Grosenbaugh LR (1964) Some suggestions for better sample-tree measurement. In Proceedings of the Society of American Foresters Meeting, 20-23 October 1963, Boston, Mass. Bethesda, Md: Society of American Foresters, pp. 36-42.

Hengl T, Heuvelink GBM, Stein A (2003) Comparison of kriging with external drift and regression kriging. Technical report. International Institute for Geo-information Science and Earth Observation (ITC), Enschede.

Henley S (1981) Nonparametric Geostatistics. London: Applied Science Publishers.

Holmgren P, Persson R (2002) Evolution and prospects of global forest assessments. *Unasylva* **53**:3-9.

Hotelling H (1936) Relations between two sets of variates. *Biometrika* **28**:321–377.

Hudak AT, Lefsky MA, Cohen WB, Berterretche M (2002) Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sens Environ* **82**:397-416.

Iles K (2003) A sampler of inventory topics: a practical discussion for resource samplers, concentrating on forest inventory techniques. Nanaimo, B.C.: Kim Iles & Associates Ltd.

Kelley HW (1990) Keeping the land alive, FAO Soils bulletin 50, Soil erosion – its causes and cures. Food and Agriculture Organization of the United Nations, Rome, 1990.

Krieger DJ (2001) The Economic Value of Forest Ecosystem Services: A Review. The Wilderness Society, Washington DC.

Köhl M, Magnussen SS, Marchetti M (2006) Sampling methods, remote sensing and GIS multiresource forest inventory. Heidelberg: Springer.

Little RJ (2004) To model or not to model? Competing modes of inference for finite population sampling. *J Am Stat Assoc*. **99**:546-556.

Lochhead K, LeMay V, Bull G, Schwab O, Halperin J (2018) Multivariate estimation for accurate and logically consistent forest-attributes maps at macroscales. *Can J For Res.* **48**:345-359.

Lund G (1998) A comparison of multipurpose resource inventories (MRIs) throughout the world. *EFI working paper* **14**, Joensuu.

Maselli F, Chiesi M, Corona P (2014) Use of geographically weighted regression to enhance the spatial features of forest attribute maps. *J Appl Remote Sens*. **8**.

Matheron G (1969) *Le Krigeage Universel*. Fontainebleau: Cahiers du Centre de Morphologie Mathématique.

McRoberts RE (2012) Estimating forest attribute parameters for small areas using nearest neighbors techniques. *Forest Ecol Manag.* **272**:3-12.

McRoberts RE, Tomppo EO, Finley AO, Heikkinen J (2007) Estimating areal means and variances of forest attributes using the k-nearest neighbours technique and satellite imagery. *Remote Sens Environ*. **111**:466-480.

McRoberts RE, Tomppo EO, Næsset E (2010a) Advances and emerging issues in national forest inventories. *Scand J For Res.* **25**:368-381.

McRoberts RE, Cohen WB, Næsset E, Stehman SV, Tomppo EO (2010b) Using remotely sensed data to construct and assess forest attribute maps and related spatial products. *Scand J For Res.* **25**:340-367.

Meng Q, Cieszewski C, Madden M (2009) Large area forest inventory using Landsat ETM+: A geostatistical approach. *ISPRS J Photogramm Remote Sens.* **64**:27-36.

Moeur M, Stage AR (1995) Most similar neighbour: an improved sampling inference procedure for natural resource planning. *For Sci*. **41**:337-359.

Odeh IOA, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*. **67**:215-226.

Ohmann JL, Gregory MJ (2002) Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, USA. *Can J For Res.* **32**:725-741.

Ohmann JL, Gregory MJ, Henderson EB, Roberts HM (2011) Mapping gradients of community composition with nearest-neighbour imputation: extending plot data for landscape analysis. *J Veg Sci*. **22**:660-676.

Rao JNK, Molina I (2015) Small Area Estimation. New York: Wiley.

Reese H, Nilsson M, Granqvist PT, Hagner O, Joyce S, Tingelöf U, Egberth M, Olsson H (2003) Countrywide Estimates of Forest Variables Using Satellite Data and Field Data from the National Forest Inventory. *Ambio.* **32**:542-548.

Särndal CE, Swensson B, Wretman J (1992) Model Assisted Survey Sampling. New York: Springer.

Schreuder HT, Gregoire TG, Wood GB (1993) Sampling methods for multiresource forest inventory. New York: Wiley.

Smith TMF (1994) Sample Surveys 1975-1990; An Age of Reconciliation? *Int Stat Rev*. **62**:5-34.

Smith TMF (2001) Biometrica Centenary: Sample Surveys. *Biometrika*. **88**:67-134.

Stevens DL (1997) Variable density grid-based sampling designs for continuous spatial populations. *Environmentrics*. **8**:167-195.

ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*. **67**:1167-1179.

Tabacchi G, Di Cosmo L, Gasparini P (2011) Aboveground tree volume and phytomass prediction equations for forest species in Italy. *Eur J For Res*. **130**:911-934.

Thompson SK (2002) *Sampling*, 2nd ed, New York: Wiley.

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr*. **46**:234-240.

Tomppo EO (1991) Satellite image-based national forest inventory of Finland. *ISPRS Archives*. **28**:419-424.

Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (Eds.) (2010) National Forest Inventories. Pathways for Common Reporting. Dordrecht: Springer.

Wackernagel H (2003) Multivariate Geostatistics: an introduction with applications 3rd ed. Berlin: Springer-Verlag.

Wang Q, Ni J, Tenhunen J (2005) Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Glob Ecol Biogeogr.* **14**:379-393.

Zhang L, Shi H (2004) Local Modeling of Tree Growth by Geographically Weighted Regression. *For Sci*. **50**:225-244.

## List of papers

Marcelli A, Corona P, Fattorini L (2019) Design-based estimation of mark variograms in forest ecosystem surveys. Spat. Stat., 30: 27-38. doi: 10.1016/j.spasta.2019.02.002

Grotti M, Puletti N, Chianucci F, Mattioli W, Floris A, Clementel F, Torresan C, Marchi M, Gentile A, Pisante M, Marcelli A, Corona P (2019) A plot sampling strategy for estimating the area of olive tree crops and olive tree abundance in a mediterranean environment. Trends in Earth Observation, 1:11-14.

Marcelli A, Mattioli W, Puletti N, Chianucci F, Gianelle D, Grotti M, Chirici G, D'Amico G, Francini S, Travaglini D, Fattorini L, Corona P (2020) Large-scale two-phase estimation of wood production by poplar plantations exploiting Sentinel-2 data as auxiliary information. Silva Fenn., 54(2). doi: 10.14214/sf.10247

Corona P, Chianucci F, Marcelli A, Gianelle D, Fattorini L, Grotti M, Puletti N, Mattioli W (2020) Probabilistic sampling and estimation for large-scale assessment of poplar plantations in Northern Italy. Eur. J. For. Res. doi: 10.1007/s10342-020-01300-9

Bartolini A, Di Biase RM, Fattorini L, Franceschi S, Marcelli A (2021) Design-based mapping of plant species presence, association and richness by nearest-neighbor interpolation. Working Paper n.854. Quaderni DEPS, University of Siena.

Marcelli A, Fattorini L, Franceschi S, Harmonization of design-based mapping for spatial populations [submitted]