



Special issue: Complex remote sensing-assisted forest surveys

Sara Franceschi¹, Caterina Pisani¹, Lorenzo Fattorini¹ and Piermaria Corona²

Statistical considerations for enhanced forest resource mapping

Franceschi S., Pisani C., Fattorini L., Corona P. (2025). Statistical considerations for enhanced forest resource mapping. Silva Fennica vol. 59 no. 2 article id 24063. 11 p. <https://doi.org/10.14214/sf.24063>

Abstract

This paper examines forest resource mapping from a statistical perspective, highlighting the opportunity to use a design-based approach to ensure inferential congruency with the estimation of averages and totals of forest attributes. Traditionally, in forest surveys estimates of averages and totals are obtained using design-unbiased estimators, with known variance expressions that can be easily estimated using standard sampling methodologies. The paper emphasizes the prominent role of k NN and Random Forest techniques in forest mapping while addressing the methodological limitations identified over more than thirty years of forest literature in efforts to estimate map precision. The critical importance of design-based map consistency, often overlooked in forest literature, is discussed and clarified, demonstrating that it allows for the development of design-based estimators of map precision through bootstrap resampling from the estimated maps.

Keywords consistency; design-based inference; k NN mapping; pseudo-population bootstrap; Random Forest mapping

Addresses ¹Department of Economics and Statistics, University of Siena, Via Banchi di Sotto 55, 53100 Siena, Italy; ²CREA, Research Centre for Forestry and Wood, Viale Santa Margherita 80, 52100 Arezzo, Italy

E-mail franceschi2@unisi.it

Received 13 November 2024 **Revised** 24 June 2025 **Accepted** 30 June 2025

1 Introduction

Forest surveys, particularly large-scale surveys such as National Forest Inventories (NFIs), are crucial for conducting international and national forest monitoring and assessment programs. Forest data are essential for various purposes, including reporting on forest resources (FAO-FRA 2020), monitoring biodiversity (Corona et al. 2011; Forest Europe 2020), tracking deforestation and forest degradation (Corona et al. 2023), as well as improving decision-making processes, silvicultural practices, harvesting, and conservation activities (Chirici et al. 2020).

Traditionally, forest surveys are based on probabilistic sampling and aim to provide estimates of averages and totals of forest attributes such as forest extents, growing stock volumes and their increments at national and regional level (Tomppo et al. 2010). Commonly, the estimation of these quantities is approached from a design-based perspective in which the values of the attributes of interest are considered as fixed, thus avoiding any model assumption on them, and the properties of the adopted estimators derive only from the sampling schemes actually performed in the survey (Mandallaz 2008). In contrast, in model-based approaches those values are random variables that are assumed to be generated by a super-population probability model, while sampling is of purposive or ignorable in nature (Rubin 1976), so that the properties of the adopted predictors derive from the assumed model (Chambers and Clark 2012).

Here, we do not intend to discuss the advantages and disadvantages of the design-based versus model-based approaches, as these are well delineated in the statistical literature (Smith 1994, 2001; Gregoire 1998; Thompson 2002, chapter 10; Little 2004). Our aim is simply to highlight that design-based inference is the traditional approach used in most forest surveys, including NFIs, and that: i) this approach ensures design-unbiased estimators of averages and totals of forest attributes, with analytical expressions of their sampling variances, which can be estimated using standard survey sampling methods (Gregoire and Valentine 2008); ii) the design-based properties of these estimators can be considered objective because they derive directly from the sampling scheme actually used in the survey, rather than being derived from an assumed model (Särndal et al. 1992, page 21).

Furthermore, the design-based approach enables the use of large datasets obtained from remote sensing (e.g., satellite multispectral data), which are available either for free or at low cost for the entire survey area, thanks to the increasing capabilities of remote sensing technologies (Barrett et al. 2016; Di Biase et al. 2018). In particular, remotely sensed data can be utilized at the design level, for example, in the initial stratification of sampling units based on major land uses (McRoberts and Tomppo 2007), or at the estimation level, within the framework of model-assisted inference. In this approach, models that incorporate remotely sensed data as covariates are used to build improved estimators, while still retaining the properties derived from the sampling scheme used in the field. Specifically, model-assisted inference ensures approximate design-unbiasedness for the resulting estimators of averages and totals, as well as approximate design-based variance expressions, with subsequent variance estimators (Opsomer et al. 2007; Breidt and Opsomer 2017).

However, even if design-based estimation of averages and totals of forest attributes provides reassuring statistical soundness, the resulting estimates can support forest monitoring and management at large scale only, without providing explicit spatial details. On the other hand, the construction of wall-to-wall maps and the provision of small-area estimates or predictions of forest attributes are becoming essential in most large-scale forest surveys. In particular, wall-to-wall forest mapping is now considered as unavoidable in modern NFIs, usually referred to as Enhanced Forest Inventories (Stinson and White 2018).

The purpose of this discussion paper is to argue how forest attribute mapping should be conducted from a statistically sound perspective, while ensuring consistency with the design-based

inference used for estimating averages and totals. Section 2 focuses on the dominant role of k NN and Random Forest (RF) methods in forest resource mapping. Section 3 outlines the methodological shortcomings of the approaches applied over the past thirty years in forest literature regarding these mapping techniques. Section 4 highlights the necessary steps to achieve methodological soundness. Finally, concluding remarks are presented in Section 5.

2 Forest attributes mapping

For a long time, mapping has typically relied on geostatistical methods. Among these, kriging predictors are probably the most widely used and well-established techniques. In the most general case, known as kriging with external drift or universal kriging, it is assumed that the values to be mapped follow a deterministic trend. This trend is expressed as a function of a set of covariates available for the entire survey region, plus error terms that stem from a centered, intrinsically stationary, and Gaussian random process, characterized by a spatial autocovariance function.

The theoretical appealing of kriging and its large use are due to the fact that, under the assumed model, kriging ensures the best linear unbiased predictors with an analytical variance expression that can be readily estimated from sample observations (Cressie 1993). Alternatively, a theoretically founded, well-experienced mapping criterion is the geographically weighted regression, based on the assumption that the relationship between the interest variable and covariates varies across the survey region and tend to be similar for nearby locations. Therefore, the values to be mapped are predicted by a regression model in which the covariate values at sample locations are weighted in relation to their distances to the location under estimation (Fotheringham et al. 2002). Surprisingly, despite their solid theoretical basis, both the techniques has been scarcely applied for mapping forest resources. Sales et al. (2007) and Wang et al. (2021) constitutes rare, even if not unique, examples.

In contrast, towards the end of the last century, mapping techniques based on the nearest neighbor method became increasingly popular in forest studies. This mapping approach was first introduced in forest research by Tomppo (1990, 1991) for use in the Finnish NFI. Among these techniques, the k NN method has been the most widely applied (Corona et al. 2014). Its popularity is likely due to its simplicity and flexibility. Specifically, k NN predictions are easily generated by linearly combining the k sample observations that are “nearest” to the location where the prediction is being made, according to some distance criterion and in some space, thus avoiding the computational challenges of more complex geostatistical and weighted regression methods. Furthermore, the number of neighbors (k) involved in the predictions determines the smoothness of the resulting map, and various distance metrics (e.g., Euclidean distance, Mahalanobis distance, etc.) can be chosen, offering great flexibility in its implementation. Another advantage of k NN is its ability to efficiently incorporate auxiliary information by determining the k nearest neighbors in the space of the auxiliary variables, rather than in geographical space. After the two seminal papers by Tomppo, hundreds of applications of k NN stemmed from forest studies. Chirici et al. (2016) provide a detailed review of these applications and, among the listed plethora of articles, Landsat imagery is the most frequent source of auxiliary information followed by airborne laser scanning metrics, digital aerial imagery, SPOT imagery and combinations of these sources. At least in this review article, there is no mention of k NN applications directly performed in the geographical space or including spatial coordinates as auxiliary information.

In recent years, the use of k NN applications has decreased in favor of machine learning techniques, with RF imputation becoming increasingly popular (e.g., Baccini et al. 2008; Cartus et al. 2014; Chirici et al. 2020). However, Lin and Jeon (2006) demonstrated that RF can be viewed

as an adaptively weighted k NN method, suggesting that it is likely to share similar properties. The similarity between the two techniques is evident from the results of a large simulation study by Fattorini et al. (2024), in which the performances of k NN and RF are found to be virtually identical. Therefore, from this point onward, the two mapping methods will be considered jointly whenever possible.

3 What was done

As already stated, one of the main reasons for the popularity of k NN and RF in forest resource mapping is the computational simplicity of the first with respect to the more challenging computational tasks involved in geostatistical and weighted regression methods, and the wide disposal of software for the automated applications of the second (e.g., package *randomForest*, R Core Team 2021). Unfortunately, this utilitarian choice has not considered the subsequent difficulties in performing a rigorous statistical inference.

In fact, although the properties of k NN and RF predictors have a long tradition in the statistical literature, starting from the seminal works by Stone (1977) and Devroye et al. (1994) to the more recent papers by Breiman (2001), Hall et al. (2008), Samworth (2012) and Gadat et al. (2016), the results are derived by considering the sample data as realizations of independent and identically distributed random variables. This restriction prevents the extension of these results to the mapping of forest attributes in which sample data traditionally arise from a probabilistic sampling of locations, usually performed without replacement (thus precluding independence) and possibly with different inclusion probabilities of locations (thus precluding identical distributions).

Specific studies would have been necessary for determining the properties of k NN and RF predictors in forest mapping. In contrast, while at least to our knowledge no theoretical study regarding RF mapping is present in forest literature, the review by Chirici et al. (2016) found that most papers regarding k NN are of applicative nature (about 80%) while only about 7% are of methodological nature, with proposals invariably supported by practical applications or simulations, with no theoretical investigation able to justify their use. The theoretical weakness of k NN studies in forest literature has been evidenced by Eskelson et al. (2009), who are the unique authors, among the myriads who have dealt with k NN, pointing out that “In general, estimation techniques are chosen based on statistical properties such as unbiasedness, consistency and efficiency. These properties are not well understood for NN imputation”.

The main methodological issues faced – with little success – in forest literature regarding k NN mapping is bias mitigation and estimation of precision. The bias issue has been well recognized by Magnussen et al. (2010b) and Baffetta et al. (2012). Both papers evidence that the largest values of the interest variable tend to be underestimated and the smallest ones to be overestimated and propose empirical machineries to reduce bias. Nevertheless, the performance of the proposed methods is evaluated only by case studies, in the first case, and by a simulation study, in the second case, without any theoretical demonstration of their effectiveness.

Regarding the estimation of precision, Baffetta et al. (2012, section 4) pessimistically argue that any rigorous inference at the pixel level is precluded in k NN mapping. Other authors, however, have been more optimistic. Model-based inference is the approach typically used to estimate the mean squared errors of k NN predictors, reflecting foresters’ skepticism toward the possibility of adopting design-based methods. For instance, Kim and Tomppo (2006, equation 3) ignore the presence of bias and propose a variance estimator based on a variogram model defined in the space of auxiliary variables, estimated using the Matern class of variograms. Three years later, Tomppo and co-authors (Magnussen et al. 2009, equation 4 and appendices A.1–A.5) develop a mean squared

error estimator, again model-based, because in their view, “distributional assumptions are needed for estimating the uncertainty of k NN predictions”. This estimator is derived under the assumption that population values are random variables with expectations and variances depending on a vector of auxiliary variables. Surprisingly, one year later, the same authors (Magnussen et al. 2010a, equation 4), neglecting bias, propose another estimator for the variance of k NN predictions based on modified balanced repeated replications. Once again, the authors’ skepticism toward design-based approaches is evident in their statement that “the nonparametric nature of k NN estimators precludes an analytical design-based k NN variance estimator.” As we will discuss in the next section, this view turns out to be unjustified in light of more recent studies.

The statistical weaknesses of these proposals are evident from the fact that they have been all tested on real or artificial data without any theoretical investigation into their properties, and without even being subjected to simulations. Furthermore, at least to our knowledge, none of these proposals have applied in the subsequent literature. In practice, they have been only used in the papers in which they were introduced.

In conclusion, drawing on the extensive forest literature on k NN and RF mapping, it is clear that assessments of map precision are, in most cases, based on highly empirical cross-validation or leave-one-out techniques. While these methods may be useful as global indices of fitting performance, they do not reflect the actual precision of k NN and RF predictions at any specific location within the survey area. In other words, the forest literature has yet to provide reliable tools for estimating the true precision of k NN and RF maps.

4 What needs to be done

As inference on averages and totals of forest attributes, from their estimation to the estimation of their precision, is traditionally performed in a design-based framework, the estimation of the precision of forest maps should be congruent and should be performed in the same design-based framework without resorting to models. As we shall see in this section, the track toward a design-based estimation of the precision of forest maps, even if essential to ensure inferential congruence, is much more complex than the fast tracks that can be travelled by supposing and exploiting more or less realistic model assumptions, in which the resulting estimates of precision crucially depend on the validity of these assumptions.

In this context, it is important to note that the k NN mapping criterion, as well as similar – albeit much less commonly used – criteria such as the inverse distance weighting (IDW) and nearest neighbour (NN) mapping, are invariably based on the naïve assumption that locations that are close (in some sense) tend to be more similar than those that are farther apart. This is typically referred to as Tobler’s Law (Tobler 1970). Based on this assumption, estimation at unsampled locations is naturally performed by linear combinations of the sample values observed at locations that are neighbour (in some sense) to the locations under estimation. Regarding the NN mapping, note that it can be viewed as a variation of k NN when a single neighbour (the nearest) is exploited instead of k , or a variation of IDW mapping when the smoothing parameter adopted to weight distances approaches infinity (Fattorini et al. 2022).

Obviously, when applying these criteria to estimate the value at a single location an error invariably occurs, but nothing ensures that the error has a design-based distribution centred and clumped at the true value. In practice, for finite samples, the design-based finite sample properties of these mapping estimators are not clearly delineated. Little is known about their design-based bias and mean squared errors and how these are related to the sampling effort. Accordingly, the sole way to render statistically sound these mapping techniques is to determine conditions ensur-

ing the design-based consistency of the corresponding estimators. In addition, if the variables to be mapped are bounded, as always happens in practical applications, consistency also involves asymptotic design-based unbiasedness.

Surprisingly, consistency has been invariably neglected in forest literature, probably for the fact that it holds as the number of sampled locations approaches infinity, so that consistency is viewed as a pure theoretical property never attainable in practice, as in practice the number of sample locations is always finite. On the other hand, as pointed out by Sarndal et al. (1992, section 5.3), design-based consistency has a very practical importance, in the sense that, if consistency holds, then we can trust in the resulting estimates because the sampling distribution of the estimators of the population values can be considered tightly concentrated around the true values when the number of sample locations is large enough. Moreover, owing to the asymptotic unbiasedness, these sampling distributions can be considered centred around the true values.

Fattorini et al. (2018a,b, 2022) demonstrate the design-based consistency of IDW and NN mapping in geographical space under mild continuity conditions on the surfaces to be mapped, highlighting the crucial role of the sampling schemes used in the surveys. In particular, they prove that consistency holds under random selection of locations, as well as under tessellated schemes widely applied in forest surveys and NFIs, such as tessellation-stratified sampling and systematic grid sampling for continuous populations, and their counterparts for finite populations of areas (pixels).

Furthermore, for the first time in over thirty years of k NN applications, Fattorini et al. (2024) prove the consistency of k NN mapping (and presumably RF mapping) under the same mild conditions on the surfaces to be mapped, and the same random and tessellated sampling schemes. These results may be surprising to forest scientists who have traditionally applied k NN and RF in the space of auxiliary variables, often neglecting geographical coordinates, as in such cases, the two mapping criteria may still lack consistency. These findings emphasize the critical need to incorporate geographic coordinates into the distance calculations, alongside the auxiliary variables.

These findings are especially relevant because design-based consistency is also crucial for the design-based estimation of map uncertainty. The core idea, first introduced by Fattorini et al. (2022) in NN mapping and later applied to IDW mapping with the smoothing parameter selected from the sample data using leave-one-out procedures (Fattorini et al. 2023), should be easily extendable to k NN, where the smoothing parameter is the number k of neighbors used in the interpolation. The approach involves treating the estimated map as a pseudo-population (Quatemberg 2016), from which bootstrap samples are drawn using the same spatial scheme adopted for the original sample. Then, all the steps performed on the original sample to estimate the map are repeated in each bootstrap sample. The key intuition behind this proposal is that, under consistency, as the estimated maps converge to the true maps, the bootstrap distributions arising from these maps should also converge to the true distributions of the estimators, thus providing consistent estimators of their mean squared errors. Franceschi et al. (2022) empirically investigate the effectiveness of NN maps as pseudo-populations, showing that they are likely to be good representations of the true maps.

In this framework, the pseudo-population bootstrap constitutes a very effective tool that allows to consider the design-based uncertainty entailed not only by the sampling scheme, but also by any estimation step performed on the sample data, from the choice of the “best” smoothing parameter to the choice of the “best” auxiliary variables to be exploited, irrespective of the criterion adopted to define what is “best”. That has been experienced in a model-assisted approach in which the auxiliary variables are exploited as covariates in a linear prediction model and the estimated values are obtained from the predictions plus an error component achieved from the IDW mapping of the sample errors in the geographic space (Di Biase et al. 2022). The strategy has been subsequently automated in the R programming language by adopting Sentinel-2 remote

sensing data as auxiliary information downloaded from the Google Earth Engine cloud computing platform (Francini et al. 2024).

A further issue that may sound awkward in the reporting phase is that the estimates of averages and totals of forest attributes achieved by the familiar design-based estimators for the whole study area and for domains partitioning the area (e.g., the administrative regions of a country) will invariably differ from the total estimates achieved from the resulting map as the sum of the estimated values. To obtain non-discrepant results, Marcelli et al. (2022) suggest a harmonization procedure in which the values estimated by means of the IDW technique are rescaled, so that their totals coincide with those achieved by traditional estimation. The capacity of the harmonization procedure to maintain consistency of the rescaled maps is proven providing that the traditional estimators are themselves consistent, a feature that has been proved to hold under the above mentioned random and tessellated sampling schemes (Fattorini et al. 2020). Therefore, under IDW mapping and tessellated sampling schemes, the uncertainty involved by harmonization can be included in the pseudo-population bootstrap, viz. by resampling from the harmonized maps and including the harmonization as a further step to be performed in each bootstrap sample.

Obviously, as the number of domains increases and their size decreases, the number of sample locations available for traditional estimation within each domain becomes smaller. This inevitably leads to an increase in the variance of traditional estimators, making the resulting estimates unreliable. Therefore, harmonization with excessively small domains should be avoided. However, the sum of interpolated values within these domains can still be used as appropriate design-based small-area estimates, thus bypassing the complex modeling typically employed in small-area estimation (Rao and Molina 2015). The design-based consistency of these small-area estimators has been proven under IDW mapping and tessellated sampling schemes (Fattorini et al. 2018a, appendix B), while their mean squared errors can once again be estimated within the framework of pseudo-population bootstrap.

5 Conclusions

The integration of sample survey inventory data and mapping information has become a key issue in the development and implementation of surveys designed to monitor and assess forests and forest ecosystems. The ever-growing availability of remote sensing data provides valuable ancillary information for spatially estimating forest attributes, i.e., for forest resource mapping, such as wall-to-wall mapping, which is now considered essential in modern NFIs.

From this perspective, this paper has explored the opportunity to use a design-based approach to ensure inferential congruency between mapping and the estimation of averages and totals of forest attributes, while also addressing the methodological challenges identified over more than thirty years of forest literature in attempts to estimate map precision.

The results achieved for IDW and NN mapping in geographical spaces, as presented and discussed in this paper, can be easily extended, with minimal adjustments, to k NN and RF mapping in the space of auxiliary variables, as long as geographical coordinates are included to ensure consistency. This approach would ultimately enable statistically sound, design-based estimation of map precision, at least within the tessellated sampling schemes commonly used in forest surveys, particularly large-scale ones such as NFIs.

References

- Baccini A, Laporte N, Goetz SJ, Sun M, Huang D (2008) A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environ Res Lett* 3, article id 045011. <https://doi.org/10.1088/1748-9326/3/4/045011>.
- 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. <https://doi.org/10.1016/j.foreco.2011.06.037>.
- Barrett F, McRoberts RE, Tomppo E, Cienciala E, Waser LT (2016) A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. *Remote Sens Environ* 174: 279–289. <https://doi.org/10.1016/j.rse.2015.08.029>.
- Breidt FJ, Opsomer JD (2017) Model-assisted survey estimation with modern prediction techniques. *Stat Sci* 32, 190–205. <https://doi.org/10.1214/16-STS589>.
- Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Cartus O, Kellndorfer J, Walker W, Franco C, Bishop J, Santos L, Fuentes JMM (2014) A national, detailed map of forest aboveground carbon stocks in Mexico. *Remote Sensing* 6: 5559–5588. <https://doi.org/10.3390/rs6065559>.
- Chambers RL, Clark RG (2012) An introduction to model-based survey sampling with applications. Oxford University Press, New York. <https://doi.org/10.1093/acprof:oso/9780198566625.001.0001>.
- Chirici G, Mura M, McInerney D, Py N, Tomppo E, Waser LT, Travaglini D, McRoberts RE (2016) A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sens Environ* 176: 282–294. <https://doi.org/10.1016/j.rse.2016.02.001>.
- 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 of Appl Earth Obs* 84, article id 101959. <https://doi.org/10.1016/j.jag.2019.101959>.
- Corona P, Chirici G, McRoberts RE, Winter S, Barbati A (2011) Contribution of large-scale forest inventories to biodiversity assessment and monitoring. *Forest Ecol Manag* 262: 2061–2069. <https://doi.org/10.1016/j.foreco.2011.08.044>.
- 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. <https://doi.org/10.1016/j.rse.2014.05.011>.
- Corona P, Di Stefano V, Mariano A (2023) Knowledge gaps and research opportunities in the light of the European Union Regulation on deforestation-free products. *Ann Silv Res* 48: 87–89. <https://doi.org/10.12899/asr-2445>.
- Cressie NAC (1993) Statistics for spatial data. Wiley, New York. <https://doi.org/10.1002/9781119115151>.
- Devroye L, Györfi L, Krzyżak A, Lugosi G (1994) On the strong universal consistency of nearest neighbor regression function estimates. *Ann Stat* 22: 1371–1385. <https://doi.org/10.1214/aos/1176325633>.
- Di Biase RM, Fattorini L, Marchi M (2018) Statistical inferential techniques for approaching forest mapping. A review of methods. *Ann Silv Res* 42: 46–58. <https://doi.org/10.12899/asr-1738>.
- Di Biase RM, Fattorini L, Franceschi S, Grotti M, Puletti N, Corona P (2022) From model selection to maps: a completely design-based data-driven inference for mapping forest resources. *Environmetrics* 33, article id e2750. <https://doi.org/10.1002/env.2750>.
- Eskelson BN, Temesgen H, Lemay V, Barrett TM, Crookston NL, Hudak AT (2009) The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring data-

- bases. *Scand J Forest Res* 24: 235–246. <https://doi.org/10.1080/02827580902870490>.
- FAO-FRA (2020) Global forest resources assessment 2020 – key findings. FAO, Rome. <https://doi.org/10.4060/ca8753en>.
- Fattorini L, Marcheselli M, Pratelli L (2018a) Design-based maps for finite populations of areal units. *J Am Stat Assoc* 113: 686–697. <https://doi.org/10.1080/01621459.2016.1278174>.
- Fattorini L, Marcheselli M, Pisani C, Pratelli L (2018b) Design-based maps for continuous spatial populations. *Biometrika* 105: 419–429. <https://doi.org/10.1093/biomet/asy012>.
- Fattorini L, Marcheselli M, Pisani C, Pratelli L (2020) Design-based consistency of the Horvitz–Thompson estimator under spatial sampling with applications to environmental surveys. *Spat Stat* 35, article id 100404. <https://doi.org/10.1016/j.spasta.2019.100404>.
- Fattorini L, Marcheselli M, Pisani C, Pratelli L (2022) Design-based properties of the nearest neighbor spatial interpolator and its bootstrap mean squared error estimator. *Biometrics* 78: 1454–1463. <https://doi.org/10.1111/biom.13505>.
- Fattorini L, Franceschi S, Marcheselli M, Pisani C, Pratelli L (2023) Design-based spatial interpolation with data driven selection of the smoothing parameter. *Environ Ecol Stat* 30: 103–129. <https://doi.org/10.1007/s10651-023-00555-w>.
- Fattorini L, Franceschi S, Pisani C (2024) Design-based consistent strategies exploiting auxiliary information in environmental mapping. *J Agric Biol Envir St*. <https://doi.org/10.1007/s13253-024-00664-4>.
- Forest Europe (2020) State of Europe’s forests 2020. Ministerial Conference on the Protection of Forests in Europe - FOREST EUROPE, Liaison Unit Bratislava. https://foresteurope.org/wp-content/uploads/2016/08/SoEF_2020.pdf.
- Fotheringham AS, Brunson C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, New York.
- Franceschi S, Di Biase RM, Marcelli A, Fattorini L (2022) Some empirical results on nearest-neighbour pseudo-populations for resampling from spatial populations. *Stats* 5: 385–400. <https://doi.org/10.3390/stats5020022>.
- Francini S, Marcelli A, Chirici G, Di Biase RM, Fattorini L, Corona P (2024) Per-pixel forest attribute mapping and error estimation: the Google Earth engine and R *dataDriven* tool. *Sensors* 24, article id 3947. <https://doi.org/10.3390/s24123947>.
- Gadat S, Klein T, Marteau C (2016) Classification in general finite dimensional spaces with the k -nearest neighbor rule. *Ann Stat* 44: 982–1009. <https://doi.org/10.1214/15-AOS1395>.
- Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. *Can J Forest Res* 28: 1429–1447. <https://doi.org/10.1139/x98-166>.
- Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. New York, Chapman & Hall. <https://doi.org/10.1201/9780203498880>.
- Hall P, Park BU, Samworth RJ (2008) Choice of neighbor order in nearest-neighbor classification. *Ann Stat* 36: 2135–2152. <https://doi.org/10.1214/07-AOS537>.
- Kim HJ, Tomppo E (2006) Model-based prediction error uncertainty estimation for k -nn method. *Remote Sens Environ* 104: 257–263. <https://doi.org/10.1016/j.rse.2006.04.009>.
- Lin Y, Jeon Y (2006) Random forests and adaptive nearest neighbors. *J Am Stat Assoc* 101: 578–590. <https://doi.org/10.1198/016214505000001230>.
- Little RJ (2004) To model or not to model? Competing modes of inference for finite population sampling. *J Am Stat Assoc* 99: 546–556. <https://doi.org/10.1198/016214504000000467>.
- Magnussen S, Mc Roberts RE, Tomppo E (2009) Model-based mean square error estimators for k -nearest neighbour predictions and applications using remotely sensed data for forest inventories. *Remote Sens Environ* 113: 476–488. <https://doi.org/10.1016/j.rse.2008.04.018>.
- Magnussen S, Mc Roberts RE, Tomppo E (2010a) A resampling variance estimator for the k near-

- est neighbours technique. *Can J Forest Res* 40: 648–658. <https://doi.org/10.1139/X10-020>.
- Magnussen S, Tomppo E, Mc Roberts RE (2010b) A model-assisted k-nearest neighbour approach to remove extrapolation bias. *Scand J Forest Res* 25: 174–184. <https://doi.org/10.1080/02827581003667348>.
- Mandallaz D (2008) Sampling techniques for forest inventories. Chapman & Hall, Boca Raton. <https://doi.org/10.1201/9781584889779>.
- Marcelli A, Fattorini L, Franceschi S (2022) Harmonization of design-based mapping for spatial populations. *Stoch Env Res Risk A* 36: 3171–3182. <https://doi.org/10.1007/s00477-022-02186-2>.
- McRoberts RE, Tomppo EO (2007) Remote sensing support for national forest inventories. *Remote Sens Environ* 110: 412–419. <https://doi.org/10.1016/j.rse.2006.09.034>.
- Opsomer JD, Breidt FJ, Moisen GG, Kauermann G (2007) Model-assisted estimation of forest resources with generalized additive models. *J Am Stat Assoc* 102: 400–416. <https://doi.org/10.1198/016214506000001491>.
- Quatemberg A (2016) Pseudo-populations. A basic concept in statistical surveys. Springer, Berlin. <https://doi.org/10.1007/978-3-319-11785-0>.
- R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rao JNK, Molina I (2015) Small area estimation. Wiley, New York. <https://doi.org/10.1002/9781118735855>.
- Rubin DB (1976) Inference and missing data. *Biometrika* 53: 581–592. <https://doi.org/10.1093/biomet/63.3.581>.
- Sales MH, Carlos Souza M, Kyriakidis PC, Roberts DA, Vidal E (2007) Improving spatial distribution estimation of forest biomass with geostatistics: a case study for Rondônia, Brazil. *Ecol Model* 205: 221–230. <https://doi.org/10.1016/j.ecolmodel.2007.02.033>.
- Samworth RJ (2012) Optimal weighted nearest neighbour classifiers. *Ann Stat* 40: 2733–2763. <https://doi.org/10.1214/12-AOS1049>.
- Särndal CE, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer, New York.
- Smith TMF (1994) Sample surveys 1975–1990; an age of reconciliation? *Int Stat Rev* 62: 5–34. <https://doi.org/10.2307/1403539>.
- Smith TMF (2001) Biometrika centenary: sample surveys. *Biometrika* 88: 67–134. <https://doi.org/10.1093/biomet/88.1.167>.
- Stinson G, White J (2018) What’s the difference between EFI and NFI? Demystifying current acronyms in forest inventory in Canada. *BC Forest Professional*, Jan-Feb 2018: 10–11.
- Stone CJ (1977) Consistent nonparametric regression. *Ann Stat* 5: 595–620. <https://doi.org/10.1214/aos/1176343886>.
- Thompson SK (2002) Sampling, 2nd edition. Wiley, New York.
- Tobler WR (1970) A computer movie simulating urban growth in the Detroit Region. *Econ Geogr*, 46: 234–240. <https://doi.org/10.2307/143141>.
- Tomppo E (1990) Designing a satellite image-aided national forest survey in Finland. Proceedings of “The usability of remote sensing for forest inventory and planning”. Swedish University of Agricultural Sciences, Remote Sensing Laboratory, Report 4: 43–47.
- Tomppo E (1991) Satellite image-based national forest inventory of Finland. Proceedings of the Symposium on Global and Environmental Monitoring, Techniques, and Impacts. 17–21 September 1990. Victoria, British Columbia, Canada. *International Archives of Photogrammetry and Remote Sensing* 28: 419–424.
- Tomppo LM, Gschwantner RE, McRoberts RE (2010) National forest inventories: pathways for common reporting. Springer, Heidelberg. <https://doi.org/10.1007/978-90-481-3233-1>.

Wang J, Du H, Li X, Mao F, Zhang M, Liu E, Ji J, Kang F (2021) Remote sensing estimation of bamboo forest aboveground biomass based on geographically weighted regression. Remote Sens 13, article id 2962. <https://doi.org/10.3390/rs13152962>.

Total of 59 references.