

Review paper

Statistical inferential techniques for approaching forest mapping. A review of methods

Rosa Maria Di Biase^{*1}, Lorenzo Fattorini², Maurizio Marchi³*Received 28/11/2018 - Accepted 29/11/2018 - Published online 24/12/2018*

Abstract - The increasing availability of remote sensing data at no or low costs can be used as ancillary data in order to spatialize and improve the estimation of forest attributes and without increasing the sampling effort and costs. In this review paper, a description of the main statistical inferential techniques for approaching forest mapping is proposed. This article reviews the most used forest mapping methods based on the sole spatial information as well as techniques exploiting auxiliary information from remotely sensed data. The advantages and drawbacks of each method have been described on the basis of several factors, such as the aims of the investigation and the area under examination. Two main groups were here discussed with model-based methods on one side and model-assisted methods on the other, moving the attention from the model used to interpolate surfaces to the sampling scheme. Model-based methods include kriging, locally weighted regression, K-NN, decision trees and neural networks, while the inverse distance weighting interpolator is presented in the model-assisted group.

Reliable and up-to-date information on forest characteristics are mandatory tools for any decisional process. The main input data of such systems are wall-to-wall maps depicting the spatial structures of forests and additional elements. Actually, if the original aim of forest inventories was to estimate harvestable timber amounts, a general interest towards multipurpose surveys is mandatory. Such information must deal with increased costs and more time-consuming procedures.

Keywords - spatial interpolation; forest inventories; model-based inference; design-based inference; remote sensing; wall-to-wall forest attributes

Introduction

Forests are essential for life on the Earth and lots of human activities directly benefit from forest services, including timber, energy and non-woody forest products (Holmgren and Persson 2002, Corona 2016). Forests occupy a central role in a wide range of environmental issues related to biodiversity conservation, water regulation and soil protection, mitigation of climate change impacts and the provision of many ecosystem services (e.g. Holmgren and Persson 2002, Köhl et al. 2006, Maselli et al. 2005).

Forest management and assessment are relevant for forest industry and environmental stakeholders (McRoberts and Tomppo 2007), representing the main requirements for compliance of international agreements such as the FAO Global Forest Resource Assessment (FRA), the United Nations Convention on Biological Diversity (CBD) and the Kyoto protocol (e.g. McRoberts and Tomppo 2007). At the same time, forest management decisions must rely on objective, reliable and geo-referenced information (Corona et al. 2011), as that provided

by forest inventories and remote sensing techniques (Köhl et al. 2006). Forest inventory should allow for statistically-sound estimation of forest attributes in a given area (Corona 2010). Sampling procedures for large-scale forest inventories, such as National Forest Inventories (NFI), were pioneered in the 20th century in North America and Scandinavia (Corona 2000) with the main aim to estimate harvestable timber amount across a determined forest stand. Sweden, for instance, had its first NFI in 1923, when the country feared the beginning of a wood shortage (Holmgren and Persson 2002). Gradually, the interest around NFI estimation changed and the aims moved towards multipurpose surveys (Lund 1998, Corona et al. 2002). In this transformation, variables not directly related to timber assessment and communities constituted by non-traditional objects (e.g. urban forests, woodlots and tree rows) were included within targeted ones (Kleinn 2002). However, an expected drawback of multipurpose forest inventories is the increase in costs and time, unavoidable when working with an increasing number of variables (McRoberts and Tomppo 2007). Fortunately,

(1) Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, Viterbo (Italy)

(2) Department of Economics and Statistics, University of Siena, Siena (Italy)

(3) CREA - Research Centre for Forestry and Wood, Arezzo (Italy)

* corresponding author: rmdibiase@unitus.it

the increasing availability of remote sensing tools, techniques and information, i.e. data acquired from satellites and aircraft-based sensors (Schowengerdt 2006) at no or low costs, can be used as ancillary data to spatialize and improve the estimation of forest attributes, without increasing the sampling effort and costs (e.g. Opsomer et al. 2007, Mattioli et al. 2012).

A further step in the scenario of enlarging NFI goals is to perform spatially-explicit estimation in order to generate forest maps for geographically

y_1, \dots, y_n of Y collected at the n sampled locations $\mathbf{p}_1, \dots, \mathbf{p}_n$. For any pair of locations $\mathbf{p}, \mathbf{q} \in D$, $\|\mathbf{p} - \mathbf{q}\|$ denotes their Euclidean distance. Moreover, X_1, \dots, X_k denote k auxiliary variables and $\mathbf{x}(\mathbf{p}) = [x_1(\mathbf{p}), \dots, x_k(\mathbf{p})]^T$ denotes the k -vector containing the values of these variables at $\mathbf{p} \in D$, while \mathbf{x}_0 and $\mathbf{x}_1, \dots, \mathbf{x}_n$ denotes the k -vectors of these variables at \mathbf{p}_0 and at the n sampled locations $\mathbf{p}_1, \dots, \mathbf{p}_n$, respectively.

Finally, it is worth noting that the mapping meth-

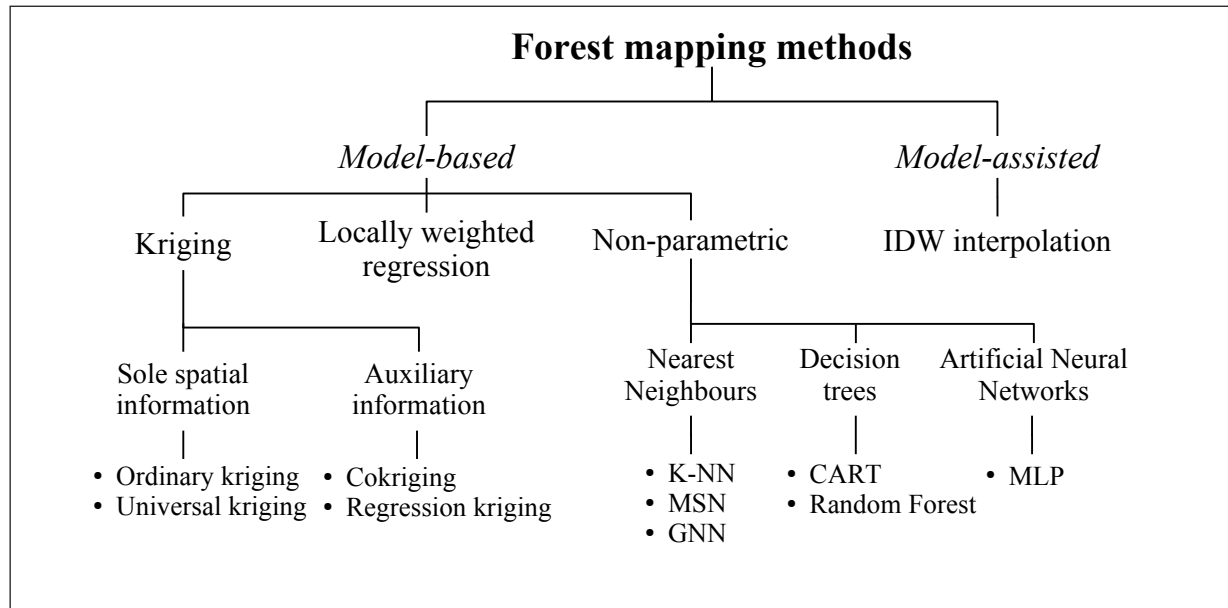


Figure 1 - Classification of literature regarding statistical methods currently applied for forest mapping and discussed in this review paper

depicting forest resource location (McRoberts and Tomppo 2007, Corona 2010). Moreover, intertwining forest inventorying and mapping has the advantage of using the resulting map to achieve a better stratification of the area, as to improve future inventory estimates.

The purpose of this review article is to summarize the main statistical methods of forest mapping in accordance with the scheme reported in Figure 1. The existing literature devoted to model-based methods is reviewed in Section 2, while literature devoted to model-assisted methods is reviewed in Section 3. The review considers both techniques based on the sole spatial information as well as techniques exploiting auxiliary information from remotely sensed data. Finally, Section 4 contains discussion and remarks.

Throughout the paper, $D \subset \mathbb{R}^d$ denotes a study domain and $\{y(\mathbf{p}) : \mathbf{p} \in D\}$ denotes a surface that describes the behaviour of the forest attribute Y throughout D . Our task is the estimation/prediction of $y_0 = y(\mathbf{p}_0)$, i.e. the realization of the interest variable Y at any non-sampled location $\mathbf{p}_0 \in D$, based on the observed values

ods reviewed in this paper concern survey variables that are defined for each point \mathbf{p} of the study domain D , a situation that is usually referred to as continuous population mapping. However, in some situations the domain is partitioned into a finite set U of N spatial units and the survey variable is the amount of a forest attribute within the units y_1, \dots, y_N . In these cases, a sample S of $n < N$ units is selected from U and the goal is to estimate the amount of forest attribute within any unsampled units on the basis of the amounts recorded within the sampled ones, $y_j, j \in S$, a situation usually referred to as finite population mapping. Even if finite population mapping is not considered in this paper, if the units locations are univocally identified by adequate "centers" $\mathbf{p}_1, \dots, \mathbf{p}_N$, most of the techniques described for continuous population mapping can be readily adopted, *mutatis mutandis*, for mapping in finite populations.

2. Model-based methods

Most statistical methodologies applied for wall-to-wall mapping rely on model-based inference. The crucial reason of this choice is the impossibility to estimate non-sampled values without any assumptions. In opposition to the design-based approach (e.g. Särndal et al. 1992, Thompson 2002, Gregoire and Valentine 2008, for details on the differences between model-based and design-based inference), model-based inference views the surface $\{y(\mathbf{p}) : \mathbf{p} \in D\}$ as the realization of a random process, called super-population, therefore making assumptions about the mechanism generating the super-population and views the n sampled locations $\mathbf{p}_1, \dots, \mathbf{p}_n$ 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).

Model-based methodologies can be classified into three major groups: parametric, semi-parametric and non-parametric. Contrary to parametric methods, non-parametric approaches do not require to fully specify the distribution of the survey variables; a mixed approach using fully specified distributions subsequently corrected by data-driven procedures is referred to as semi-parametric.

Due to the vastness of model-based methodologies adopted for mapping, only few methods, chosen on the basis of their popularity in forest studies, are presented in detail in the following paragraphs. The selected methods (see Fig. 1) include kriging and its variations (Section 2.1) as parametric methods, locally weighted regression (Section 2.2) as a semi-parametric method and nearest neighbour techniques (Section 2.3), decision trees (Section 2.4) and artificial neural networks (Section 2.5) as non-parametric methods.

2.1 Kriging

In geostatistics, kriging denotes a set of procedures for spatial prediction, i.e. the estimation of the value of the interest variable at non-sampled locations. Matheron (1963) named the method after the proposal of the mining engineer Danie Gerhardus Krige, even though the formulation of spatial prediction did not come from Krige's work (Cressie 1993).

Kriging is "a minimum-mean-squared error method of spatial prediction" (Cressie 1993, p. 106) predicting the value of the interest variable at non-sampled locations as linear combinations of the values collected at the sampled locations. However, it should be noticed that, even though kriging is a linear predictor (e.g. Journel and Huijbregts 1978, Krige 1978, Papritz and Stein 1999), nonlinear spatial prediction methods are now part of the "kriging family"

(Cressie 1990) as well. Methods of this family are summarized by Schabenberger and Pierce (2002).

2.1.1 Ordinary kriging

The ordinary kriging is the mathematically simplest version of kriging predictor, which presumes the surface $\{y(\mathbf{p}) : \mathbf{p} \in D\}$ as the realization of an intrinsically stationary spatial process (Cressie 1990, Corona et al. 2014). A process is said to be intrinsically stationary if its expected values do not depend on the spatial locations, i.e. $E[y(\mathbf{p})] = \mu$ and if the variances of increments only depend on the spatial gaps between the two locations, i.e.

$$\text{Var}[y(\mathbf{p}), y(\mathbf{p} + \mathbf{h})] = E[\{y(\mathbf{p}) - y(\mathbf{p} + \mathbf{h})\}^2] = 2\gamma(\mathbf{h})$$

The function $\gamma(\cdot)$ is generally called semi-variogram. From the assumption of intrinsic stationarity, $y(\mathbf{p}) = \mu + \delta(\mathbf{p})$, where $\{\delta(\mathbf{p}) : \mathbf{p} \in D\}$ is a zero-mean intrinsically stationary process with semi-variogram γ .

The ordinary kriging predictor can be expressed as $\hat{y}_0 = \sum \lambda_{0i} y_i$, where the weights λ_{0i} sum to 1 in order to ensure model-based unbiasedness of the predictor and are derived from the Lagrange multiplier method (Cressie 1993) in order to minimize the model-based mean squared error. Therefore, \hat{y}_0 is the best linear unbiased predictor (BLUP). It is important to point out that the semi-variogram is usually unknown and it needs to be estimated from the sample data, using several techniques, such as ordinary least squares or maximum likelihood. For a complete overview of these methods refer to Schabenberger and Gotway (2005).

This technique is rarely used in forest mapping. Forest inventory data rarely show a "pure" spatial autocorrelation trend. Some comparisons were just performed in literature and with the aim of testing the performance of different methods. A valuable example is shown by Freeman and Moisen (2006) who used the ordinary kriging with the aim of improving the point prediction accuracy of the nationwide forest biomass map; in this case, kriging was tested in conjunction with the existing map, developed with nonparametric functions: the main finding was that neither the field biomass nor the residual biomass are proved to be good candidates for ordinary kriging. Biondi et al. (1994) tested the use of variogram analysis to derive indicators on spatial structure of monitoring plots; however, this work can't be included among the "statistical mapping papers" given the aim of characterising the spatial structure of the forest and not the production of wall-to-wall maps.

2.1.2 Universal kriging

Universal kriging can be considered as a generalization of the ordinary kriging (Matheron 1969, Huijbregts and Matheron 1971). The main relevant feature of this method is that the mean is “driven” by an ancillary data available for the whole study region and in addition to the sampled locations. In this sense, this technique can be used supposing that expectations are not constant, but they are linear combinations of $k + 1$ functions of spatial coordinates

$$y(\mathbf{p}) = \sum_{h=0}^k \beta_h f_h(\mathbf{p}) \quad \text{with } f_0(\mathbf{p}) = 1 \text{ for any } \mathbf{p} \in D.$$

The coefficients β_h are usually unknown and they need to be estimated from the data.

Following Wackernagel (2003), the underlying model can be written as $y(\mathbf{p}) = \mu(\mathbf{p}) + \delta(\mathbf{p})$, where $\mu(\mathbf{p})$ can be viewed as the deterministic trend and $\{\delta(\mathbf{p}) : \mathbf{p} \in D\}$ is a zero-mean intrinsically stationary spatial process with semi-variogram γ . Under these assumptions the universal kriging predictor at any unsampled location can once

again be expressed as $\hat{y}_0 = \sum_{i=1}^n \lambda_{0i} y_i$, where in

this case the weights must satisfy the constraints

$$\sum_{i=1}^n \lambda_{0i} f_h(\mathbf{p}_i) = f_h(\mathbf{p}_0) \quad \text{for } h = 0, \dots, k \text{ in order}$$

to ensure model-based unbiasedness. It is worth noting that for $k = 0$, universal kriging reduces to ordinary kriging. Also, when universal kriging is adopted, the semi-variogram is usually unknown and needs to be estimated from the sample data. For instance, Lochhead et al. (2018) recently tested this technique to provide spatial, wall-to-wall information on forest-attributes in order to support management strategies. Main findings of the work were that while fair estimations were found with nonparametric methods (nearest neighbours with $k \leq 2$) the universal kriging was very accurate but computationally more difficult if implemented for a macroscale. Another example is provided by Mandallaz (2000), where universal kriging was tested to compare design-based and kriging techniques for the estimation of spatial averages in the context of double sampling, as used in forest inventory; the Author concluded that geostatistical techniques are useful in the context of two-phase two-stage forest inventory, primarily for local estimation, where they are superior to the classical design-based techniques.

2.1.3 Regression kriging

Similarly to universal kriging, regression kriging (Odeh et al. 1995) is a hybrid method combin-

ing a regression model with kriging. Indeed, many authors (see e.g. Deutsch and Journel 1992, Hengl et al. 2003, Wackernagel, 2003) use the term universal kriging when the deterministic trend is supposed to be a function of the coordinates only. Conversely, the term “regression” is used when the trend is supposed to be a function of some auxiliary variable. Then, the underlying model can

be written as $y(\mathbf{p}) = \boldsymbol{\beta}^T \mathbf{x}(\mathbf{p}) + \delta(\mathbf{p})$, where

$\boldsymbol{\beta} = [\beta_1, \dots, \beta_k]^T$ is an unknown vector of parameters and $\{\delta(\mathbf{p}) : \mathbf{p} \in D\}$ is once again a zero-mean intrinsically stationary spatial process with semi-variogram γ . In regression kriging, trend and residuals are usually estimated separately and summed afterwards (e.g., Ahmed and de Marsily 1987, Odeh et al. 1995). Therefore, the regression kriging predictor at

any unsampled location is $\hat{y}_0 = \hat{\boldsymbol{\beta}}^T \mathbf{x}_0 + \sum_{i=1}^n \lambda_{0i} e_i$

where $\hat{\boldsymbol{\beta}}$ is an estimate of the parameters $\boldsymbol{\beta}$ obtained from the sample, $e_i = y_i - \hat{\boldsymbol{\beta}}^T \mathbf{x}_i$ are the regression residuals and λ_{0i} are the ordinary kriging weights. Practically speaking, the residuals are predicted by means of ordinary kriging and then added to the estimated spatial trend (Goovaerts 1997). Once again, the semi-variogram is not known and need to be estimated from the data.

An interesting comparison between ordinary kriging, universal kriging, cokriging and regression kriging was published by Meng et al. (2009) for mapping the basal area as the response variable and using Landsat ETM+ images as auxiliary data in Loblolly pine (*Pinus taeda*) and Slash pine (*Pinus eliotii*) stands. In this case, regression kriging resulted in the smallest errors and the highest R-squared indicating the best geostatistical method for spatial predictions. In a similar paper on *Pinus pinaster* stands, Viana et al. (2012) mapped the above ground biomass (AGB) using remotely sensed data as ancillary variables. Tested methods included regression kriging, ordinary kriging, universal kriging and other deterministic methods such as inverse distance weighting and Thiessen polygons estimations. The considered forest variables showed low spatial autocorrelation; in this situation, kriging methods were unsuitable for the aimed purposes and regression kriging did not increase the accuracy of estimates developed by the direct radiometric relationships. Authors concluded that denser sampling schemes and different auxiliary variables should be explored.

Another example of regression kriging implementation is provided by Brus et al. (2012), where a statistical mapping of forest tree species across Eu-

rope was performed. In this paper, the ICP-Forests Level I plots were extended with the NFI plot data of eighteen countries. A soil map, a biogeographical map and bioindicators derived from temperature and precipitation data were used as predictors and estimated overall accuracy was 43%. Moreover, in areas with NFI plot data, overall accuracy was higher and around 57%. Once again, this gain was mainly attributable to the much denser plot data, less to the prediction method.

2.1.4 Cokriging

Cokriging methods constitute alternative methodologies to regression kriging for exploiting auxiliary information $\mathbf{x}(\mathbf{p})$ known for each $\mathbf{p} \in D$.

In cokriging, auxiliary information $\{\mathbf{x}(\mathbf{p}), \mathbf{p} \in D\}$

and interest variable $\{y(\mathbf{p}), \mathbf{p} \in D\}$ are both presumed to be intrinsically stationary random processes with known semi-variograms and cross-variograms, in such a way that the prediction at any unsampled location can be expressed

$$\text{as } \hat{y}_0 = \sum_{i=1}^n \lambda_{0i} y_i + \sum_{h=1}^k \sum_{i=1}^n \delta_{0hi} x_{hi}, \text{ where } \lambda_{0i}$$

are the ordinary kriging weights, summing to 1, attached to the values of the survey variable Y at sampled locations. The cokriging weights (summing to 0

for each $h = 1, \dots, k$) are denoted as δ_{0hi} and are attached to the values of the auxiliary variables at the same locations in order to ensure model-based unbiasedness of the predictor. As in the ordinary kriging, the weights are derived from the Lagrange multiplier method in order to minimize the model-based mean squared error. Therefore, \hat{y}_0 is the best linear unbiased predictor (BLUP).

However, similarly to ordinary and universal kriging, semi-variograms and cross-variograms are usually not known and they need to be estimated from the data (Corona et al. 2014).

This technique has been successfully tested by Hudak et al. (2002) in comparison with regressive models and other kriging versions: according to their results, this methods produced less biased results than regression but poorly reproduced vegetation patterns, especially at the sparser (2000 and 1000 m) sampling frequencies. Cokriging, using the Landsat panchromatic band as ancillary variable, produced slightly more accurate predictions than ordinary kriging. Cokriging was also successfully used to map the spatial variability of plant diversity. Hernandez Stefanoni and Ponce-Hernandez (2005)

mapped the number of species, the exponent Shannon and the reciprocal Simpson indices from 141 sites sampled in a tropical forest. Several spatial interpolation techniques were compared and used to prepare a map of plant diversity. Results were quite unsatisfactory for cokriging. Indeed, this method performed among the poorest interpolators due to the poor correlation between the plant diversity variables and vegetation indices computed by remote sensing data.

2.2 Locally weighted regression

Locally weighted regression (LWR) or geographically weighted regression (GWR), often referred to as “loess model”, was introduced by Cleveland and Davlin (1988). This algorithm is based on the semiparametric model $y_i = g(\mathbf{x}_i) + \varepsilon_i$ that is supposed to generate the data. As usual, ε_i s are supposed to be independent normal variables with mean zero and variance σ^2 , while g is a smooth function of the covariates. Following a notation akin to Corona et al. (2014), for any unsampled location, locally weighted regression provides a prediction

$\hat{y}_0 = \hat{g}(\mathbf{x}_0)$ achieved exploiting a neighbourhood set Q_0 of $1 \leq m \leq n$ sampled locations whose vectors of auxiliary variables are closest to \mathbf{x}_0 . Each point in Q_0 is weighted accordingly to its distance from \mathbf{x}_0 , in such a way that closer points have higher weights. Successively, these weights are used in a weighted least squares regression for fitting a linear function of the covariates of type $y = \boldsymbol{\beta}_0^T \mathbf{x}$, in such a way that the locally weighted regression predictor is $\hat{y}_0 = \hat{\boldsymbol{\beta}}_0^T \mathbf{x}_0$.

The use of locally weighted regression predictor is particularly recommended when the spatial non-stationarity of the relationships between the interest variable and the covariates reduces the efficiency of kriging methods (see e.g. Brunson et al. 1996, Maselli 2002, Foody 2003, Maselli 2014).

It is worth noting that the number m of observation in the neighbourhood set Q_0 plays an important role for the prediction. However, as suggested by Cleveland and Devlin (1988), rather than m , one must focus on the fraction $f = m/n$ of points in the neighbourhood, commonly referred to as span value. As f increases, $\hat{g}(\mathbf{x}_0)$ becomes smoother, but a too large span causes an over-smoothed function that may not fit the data well. On the other hand, a too small span produces prediction affected by much noise (Corona et al. 2014).

According to forest literature, this technique has been often used to study spatial trends or to derive insights on single tree growth trends. For example, Wang et al. (2005) successfully implemented this

technique to obtain a net primary production regression model and including spatial non-stationary in the parameters estimated for forest co-systems in China. Elevation and climatic variables were included as covariates and in addition to the time-integrated normalized difference vegetation index and this technique successfully overcame the more classic least squares regression. Zhang and Shi (2004) investigated the spatial heterogeneity of multivariate relationships between tree growth and diameter at breast height. The Authors attempted to capture spatial variation by calibrating a multiple regression model fitted at each tree in a sample plot, weighting all neighboring trees by a function of distance from the subject tree. Similarly, Subedi et al. (2017) applied a Bayesian approach to GWR to model the relationship between tree crown and diameter at breast height. Observed tree data and simulated data were used to investigate model fitting and performance in order to overcome some limitations of GWR. Also in this case, model fitting was used as diagnostic tool to map spatial heterogeneity across a study plot. Chave et al. (2005) worked to convert inventory data into an estimate of aboveground biomass in tropical forests. Finally, drivers of forest transition in a province of Northern Vietnam between 1993 and 2000 were studied by Clement et al. (2009) by GWR of remotely sensed and field data. This technique has been used more to map the spatial variation of a target variable across a study area more than derive interpolated surfaces (i.e. the spatial distribution of basal area).

2.3 Nearest neighbour techniques

To account for spatial heterogeneity, non-parametric models have been proposed as an alternative to the parametric prediction provided by universal and regression kriging. Among non-parametric approaches, the popularity of nearest neighbour (NN) techniques has quickly increased among researchers, especially in forest applications (McRoberts et al. 2010b, Baffetta et al. 2012, Mattioli et al. 2012). NN techniques were first introduced by Fix and Hodges (1951) and then applied in forest inventories for the first time in the Finnish NFI (Tomppo 1991). This pioneering work contributed to the increasing success of the NN techniques in the inventory framework (e.g. Gjertsen et al. 1999, Katila and Tomppo 2001, Chirici et al. 2008).

In practice, NN methods predict unsampled values using a linear combination of observations that are nearest or most similar to the location to be predicted in the space of the auxiliary variables. Main advantages of NN methodologies include: i) prediction in both univariate and multivariate cases and ii) no assumptions on the distribution of the variables

(McRoberts et al. 2010a).

2.3.1 k-NN

The most general NN technique is the k-NN method that includes most NN techniques adopted in forest studies (McRoberts 2012). Consider a distance metric, such as Euclidean distance or Mahalanobis distance, in the space of the auxiliary variables. Then, adopting a notation similar to Baffetta et al. (2009), for any unsampled location \mathbf{P}_0 , let $H(0, j)$ be the label of the sampled elements whose distance from \mathbf{x}_0 has rank j in the sequence of the n distances. Thus, the k-NN predictor of y_0 can be written

as $\hat{y}_0 = \sum_{j=1}^k w_j y_{H(0,j)}$ where $k < n$ is the number

of adopted neighbours and w_j is the weight of the

j -th neighbour, so that $\sum_{j=1}^k w_j = 1$. Weights can be

constant (e.g. $w_j = 1/k$ for each $j = 1, \dots, k$), but usually they are inversely proportional to the distance between the j -th neighbours and \mathbf{x}_0 , (e.g.

Baffetta et al. 2009), i.e. $w_j \propto \|\mathbf{x}_0 - \mathbf{x}_{H(0,k)}\|^{-t}$, where

t is a positive real number. Commonly $t = 0, 1, 2$, but no studies have compared the effects of different values of t (McRoberts 2012). Regarding the number of neighbours, usually a small k is preferred. However, as McRoberts (2012) points out, other criteria may be more suitable and multiple optimization criteria are possible as well. It is worth noting that there are no theoretical studies on the model-based properties of k-NN (Corona et al. 2014). The unique theoretical attempt is due to McRoberts et al. (2007) but involve severe, restrictive assumptions, e.g. the realizations of the interest variable Y are random variables with the same expectation and variance and their covariances have a parametric structure (Corona et al. 2014).

k-NN is probably the most used technique in forest mapping from NFI plots (McRoberts and Tomppo 2007). Franco-Lopez et al. (2001) tested several k-NN setups and including: distance metric, weighting function, feature weighting parameters and number of neighbours. Ancillary information and image enhancement techniques were also tested and the Euclidean distance, a three date 18-band composite image, and feature weighting parameters were the best options to build maps of basal area, volume, and cover type. The Authors addressed the simplicity of this method and its role in post stratification. Forest variables were also investigated by Reese et al. (2003) in Sweden where k-NN was used to provide a synoptic coverage using a consistent method and data source over all of Sweden. Ohmann et al. (2011) investigated the use of k-NN to map gradi-

ents of community composition. Main findings were that community composition gradients were strongly associated with climate and elevation, and less so with topography and soil.

2.3.2 MSN & GNN

Most similar neighbour (MSN) (Moeur and Stage 1995) and gradient nearest neighbour (GNN) (Ohmann and Gregory 2002) are two variations of the k-NN technique, both using a single neighbour, i.e. $k = 1$, to impute the value of the interest variable at an unsampled location.

With MSN, the most similar neighbour is chosen based on a similarity function, which is derived from canonical correlation analysis (Hotelling 1936). On the other hand, the GNN procedure models the relationship between the survey and the auxiliary variables with direct gradient analysis (Gauch 1982) using stepwise canonical correspondence analysis (CCA) (terBraak 1986). Moeur and Stage (1995) used MSN to model ecosystem functioning for landscape design. Actually, the Authors worked in a different way, simply choosing the most similar plot from the whole dataset instead of estimating design attributes element-by-element in a traditional sense for each first-phase observation. In this particular case, the canonical correlation analysis was used to derive a similarity function for this procedure a MSN inference.

2.4 Decision trees

A decision tree is a non-parametric method for discriminating among classes of objects (Carbonell et al. 1983). In practice, decision trees recursively partition the space of auxiliary variables into classes, which are determined from the sampled values (McRoberts et al. 2010b).

Decision trees are widely used for prediction of forest attributes (e.g. Helmer et al. 2010) because they can accommodate non-linear responses, continuous and categorical auxiliary variables, missing data and collinear variables (Urban 2002, Brosofske et al. 2014). Furthermore, their graphic representation can be easily interpreted, even for complex cases (Young et al. 2009). A shortcoming of this group of algorithms is that the procedure is not completely known and they are generally described as a “black box” where input and output data are known but nothing is given about the actual structure of the calculation process.

2.4.1 CART

Classification and Regression trees (CARTs) (Breiman et al. 1984) are single decision tree models predicting categorical (classification tree) or contin-

uous (regression tree) variables. The most important steps in building a CART are splitting, stopping and pruning (Song and Lu 2015).

The splitting rule allows for the partition of the observations into two nodes, with the observations going into the left child-node only if the splitting condition is true. The splitting conditions are determined by the most important auxiliary variables, chosen accordingly to some characteristics related to the degree of “purity” of the resulting child-node (Song and Lu 2015). For CARTs, this degree of purity is measured via the Gini index and the twoing criteria (Breiman et al. 1984). Other decision tree methods using different criteria are reported in the table by Song and Lu (2015).

The splitting procedure ends when the stopping criteria (e.g. minimum number of observations in an end node and maximum number of steps) are met. Those should prevent the data to be overfitted, i.e. the criteria become so complex that the number of observations in the end nodes is insufficient for a reliable prediction.

An alternative to the stopping rule is the pruning procedure, which consists in growing a large tree and then pruning, i.e. removing, the nodes providing few information (Hastie et al. 2009). Alongside this kind of pruning, referred to as post-pruning, is pre-pruning, which prevents the creation of non-significant branches.

An interesting case study was provided by Torresan et al. (2016) where metrics extracted from an airborne LiDAR sensor could be exploited to predict different forest structure types by means of classification trees: while the model has provided moderately satisfactory results in term of classification performance, Authors foresee substantial room for improvement by multi- or hyperspectral imaging that allow detailed characterization of the spectral behaviour of the forest structure types.

2.4.2 Random forest

Among the methods based on the construction of a multitude of decision trees, usually referred to as ensemble classifiers, random forest (Breiman 2001) is probably the most used. The random forest algorithm (RF) uses unpruned CARTs constructed from bootstrap samples, while a random set of auxiliary variables determines the splitting rules at each node of a tree (McInerney and Nieuwenhuis 2009). It should be noted that unpruned, fully grown trees reduce the chances of overly fitted trees (Breiman 2001), whereas the random set of auxiliary variables used for each tree decreases the correlation between trees (Brosofske et al. 2014). Each fully-grown tree is used to predict the out-of-bag data, i.e. the data not in the bootstrap sample.

Other decision tree methodologies, such as bagging (or bootstrap aggregation, Breiman 1996) and boosting (Schapire et al. 1998), are not considered in this review because they are not common in forest applications (Brosofske et al. 2014). Furthermore, random forest produces similar or better results than those achieved with these techniques. Indeed, Breiman (2001) demonstrated that random forest is as accurate as AdaBoost, which is a boosting technique introduced by Freund and Schapire (1996), faster than bagging and less sensitive to noise than other boosting techniques (Schapire et al. 1998). Random Forest was successfully used by Hudak et al. (2008) comparing several approaches for imputing the basal area and tree density aggregated at the plot-scale and species-level. Topographic variables and canopy structure were used as predictors and derived from discrete-return airborne LiDAR data. Main findings were that RF produced the best results overall, especially after reducing the number of response variables to the most important species in each RF demonstrated to be very suitable and flexible to predict canopy structure and topographic metrics derived from LiDAR surveys can be very useful for species-level imputation (Chirici et al. 2013).

2.5 Artificial neural networks

Artificial neural networks (ANNs) are prediction techniques inspired by the biological neural network of animal brains and its processing information system (McRoberts et al. 2010b). The first attempt of ANN was indeed a simple model to explain how neurons in the brain might work (McCulloch and Pitts 1943). ANNs are particularly useful for complex and non-linear problems (Ingram et al. 2005).

Typically, an ANN is formed by a collection of single processing units, the so-called artificial neurons, linked by the neural structure, i.e. an assemblage of weighted connections (Agatonovic-Kustrin and Beresford 2000), and a learning rule (Baret 1995). An ANN gathers its knowledge by detecting patterns and relationships in the training data. This means that an ANN learn, or is trained, through experience with appropriate learning exemplars (Agatonovic-Kustrin and Beresford 2000), in a way similar to human brains. As learning proceeds, the weighted connections are iteratively adjusted and once all patterns and relationships in the training data are learned, the ANN can be used to predict unknown values (Carvalho 2001). A valuable comparison between regression methods and machine learning techniques in forest mapping was described by Garcia-Gutierrez et al. (2014): the Authors presented a comparison between the classic

multiple linear regression-based methodology and regression techniques in machine learning and including neural networks; the main aim was to estimate many variables and including single-tree stem biomass, crown biomass, total volume, basal area, dominant height, mean height at stand level. Also, LiDAR metrics were exploited and included in the model as predictors and the main findings were that classic multiple linear regression performed lower than machine learning technique. Gørgens et al. (2015) compared the performance of three machine learning tools (neural network, random forest and support vector regression) for predicting stand volume of fast-growing forest plantations. LiDAR metrics were used and proven to be more effective for the estimation of stand volume: in this particular case study, the previously described random forest algorithm had the best RMSE compared to neural network.

2.5.1 Multi Layer Perceptron (MLP)

Among the several types of ANNs, Multi Layer Perceptron (MLP) is the most widely applied in remote sensing studies (Carvalho 2001). This algorithm has at least three layers of artificial neurons that, with the exception of the input ones, are activated by non-linear functions. It should be noted that too many layers may decrease the predictive ability (Blackard and Dean 1999). As for the learning rule, MLP uses the backpropagation, which is the best learning principle for non-linear relationships (Jensen et al. 1999).

Valuable examples of MLP algorithm are represented by Foody et al. (2001) and Mas et al. (2004). In the first paper, the biomass dynamics have been modelled in tropical forests of Malaysia from remotely sensed data. In the second example, deforestation processes were analysed in a GIS environment aiming at predicting the spatial distribution of tropical deforestation. In this case, an MLR was trained in order to estimate the propensity to deforestation as a function of the explanatory variables and was used to develop deforestation risk assessment maps. The model performance was quite high and able to classify correctly 69% of the grid cells. This study strengthened the knowledge that artificial neural networks and derived methods such as MLP have a great potential to predict land cover changes, mainly due to their flexibility and the possibility to develop complex, non-linear models.

3. Design-based methods

Design-based inference views the surface $\{y(\mathbf{p}) : \mathbf{p} \in D\}$ as fixed, therefore making no assumptions about the mechanism generating it and

views the n sampled locations $\mathbf{p}_1, \dots, \mathbf{p}_n$ as random, being the outcome of a random selection generated by a probabilistic sampling scheme. As stated in Section 2, pure design-based methods cannot be used for constructing wall-to-wall maps, owing to the impossibility of estimating non-sampled values without any assumptions. Specifically, from a design-based point of view, either a location is sampled, and then there is no need for estimation, or it is unsampled and then there is no information to perform a design-based estimation. Thus, as pointed out by Fattorini et al. (2018a), the sole way to recover information for the unsampled locations is to use an assisting model.

While use of an assisting model is a widely adopted and effective way to estimate totals and averages of finite populations, that is not true for estimating single population values. Indeed, when estimating totals and averages the assisting model is used to predict each population value, but the total of errors performed by the model is estimated from the sample and then adopted to correct the total or average achieved from those predictions in accordance with the criterion provided by the difference estimator (Särndal et al. 1992). In this way, we achieve design-unbiasedness of the resulting estimators, or approximate unbiasedness up to the first term of approximation, as well as exact or approximate design-based variance expressions and suitable variance estimators, as happens for the well-known generalized regression and ratio estimators. On the other hand, when estimating the value at a single location, there is no way to correct the error invariably provided by the model prediction. Thus, any model-assisted estimator at a single location and the subsequent map are destined to be design biased. Accordingly, as pointed out by Fattorini et al. (2018a), any map arising from a model-assisted criterion can achieve statistical soundness only if it is proven to be design-based asymptotically unbiased and consistent (DBAU&C). These issues have precluded the use of model-assisted inference in forest mapping.

A recent work (Fattorini et al. 2018b) deals with the condition ensuring DBAU&C for model-assisted maps achieved by using an inverse distance weighting (IDW) interpolator (Shepard 1968) to estimate unsampled locations. The assisting model, i.e. the criterion leading to the IDW interpolator, is the simple Tobler's first law of geography (Tobler 1970), asserting that a spatial location is more similar to the nearby locations than to those further apart. In accordance with this principle, the prediction of the interest variable at the unsampled units is obtained as a weighted sum of the sampled values with weights decreasing with the distance to the point to

be predicted.

3.2 Inverse Distance Weighting interpolation

Following Fattorini et al. (2018b), let $\mathbf{p}_1, \dots, \mathbf{p}_n$ be n sample locations selected by means of a sampling scheme that induces a sampling design. The sampling design is a probability measure defined on the sample space, which ensures the existence of a joint probability density function for any n -tuple $\mathbf{p}_1, \dots, \mathbf{p}_n$. In this case the IDW in-

terpolator is
$$\hat{y}_0 = \frac{\sum_{i=1}^n y_i \varphi(\|\mathbf{p}_0 - \mathbf{p}_i\|)}{\sum_{i=1}^n \varphi(\|\mathbf{p}_0 - \mathbf{p}_i\|)}$$
 where

$\varphi: [0, \infty) \rightarrow \mathbb{R}^+$ is a non-increasing function

with $\varphi(0) = 0$ and $\lim_{d \rightarrow 0^+} \varphi(d) = \infty$. Fattorini et al. (2018b) determined the conditions ensuring DBAU&C of \hat{y}_0 . These conditions concern the spatial pattern of the interest variable in the area, i.e. the behaviour of the surface $\{y(\mathbf{p}) : \mathbf{p} \in D\}$, the sampling design adopted to select the n sample locations and the distance function φ .

The first condition is the continuity of the surface $\{y(\mathbf{p}) : \mathbf{p} \in D\}$ with discontinuities admitted only over regions of zero measure. The assumption is quite reasonable in many real situations. Indeed, surfaces usually change smoothly, well approaching the continuity assumption, while abrupt variations usually occur along borders delineating sudden changes in the area. Therefore, those borders can be assimilated by curves that satisfy the theoretical condition of being a set of zero measure. Regarding the sampling design, it should be asymptotically spatially balanced, i.e. able to spread out the sample locations, so that as the sample size n increases any location of the domain is likely to have nearby locations sampled. Fattorini et al. (2018b) proved that the most widely sampling schemes adopted in forest surveys, such as tessellation stratified sampling and systematic grid sampling (Barabesi and Franceschi 2011, Barabesi et al. 2012) ensure the required spatial balance. Finally, the distance function φ should

satisfy
$$\lim_{d \rightarrow 0^+} d^2 \varphi(d) = \infty$$
, that can be readily ensured by $\varphi(d) = d^{-\alpha}$ for any $\alpha > 2$.

Actually, in real situations, there may be cases in which the surface shows many discontinuities, deteriorating the properties of the interpolator. However, Fattorini et al. (2018b) have demonstrated the consistency of the whole map if these discontinuities occur for a set of zero measure.

As a first application (Fattorini et al. 2018b), the IDW interpolator was adopted to estimate the cov-

erage of holly oak in the Montagnola Senese (Central Italy). Based on the tessellation stratified sampling scheme previously adopted in the last Italian NFI (Fattorini et al. 2006), quadrats of size 10x10m were centred at 106 inventory points and the percentage of plot covered by holly oak was recorded (Chiarucci et al. 2008).

4. Conclusions

Forest monitoring and assessment are rapidly evolving as new information needs arise and new techniques and tools become available. However, the exploitation of the latter, as well as their implementation within operative forest management processes, should be evidence-based (Corona 2018). Distinctively, the advancement of remote sensing imagery and statistical methods has eased the process of forest mapping, providing easily accessible and affordable sources of information.

The most popular methods exploiting the sole spatial information as well as those requiring auxiliary variables have been here reviewed. Despite their popularity and numerous forest applications, few investigations about the properties of applied statistical methods have been conducted and there are even less comparative studies on their performances. For instance, McInerney and Nieuwenhuis (2009) found that k-nn is more effective than random forest when estimating volume and basal area in Ireland and a study of Maselli and Chiesi (2006) showed that locally weighted regression, kriging and k-nn perform similarly when estimating standing volume in Central Italy.

It is clear that the advantages and disadvantages of each method depend on the aims of the investigation, the variables involved, the type of remote sensing data and the study area, all elements to take in consideration when choosing an appropriate technique (Brosofske et al. 2014). As Corona et al. (2014, page 30) pointed out, “the matter is still controversial” and further investigation on the statistical drawbacks and benefits of such methods are needed. A good dataset with reliable and unbiased data is the main starting point. Actually, the complexity of a model is generally unable to solve deficiencies of the sampling methods.

References

- Agatonovic-Kustrin S., Beresford R. 2000 - *Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research*. Journal of Pharmaceutical and Biomedical Analysis 22 (5): 717-727. doi: 10.1016/S0731-7085(99)00272-1
- Ahmed S., de Marsily G. 1987- *Comparison of Geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity*. Water Resources Research 23 (9):1717-1737. doi: 10.1029/WR023i009p01717
- Baffetta F. Fattorini L., Franceschi S., Corona P. 2009 - *Design-based, approach to k-nearest neighbors technique for coupling field and remotely sensed data in forest surveys*. Remote Sensing of Environment 113 (3): 463-475. doi: 10.1016/j.rse.2008.06.014
- Baffetta F., Corona P., Fattorini L. 2012 - *A matching procedure to improve k-nn estimation of forest attribute maps*. Forest Ecology and Management 272: 35-50. doi: 10.1016/j.foreco.2011.06.037
- Barabesi L., Franceschi S. 2011 - *Sampling properties of spatial total estimators under tessellation stratified designs*. Environmetrics 22 (3): 271-278. doi: 10.1002/env.1046
- Barabesi L., Franceschi S., Marcheselli M. 2012 - *Properties of design-based estimation under stratified spatial sampling*. The Annals of Applied Statistics 6 (1): 210-228. doi: 10.1214/11-AOAS509
- Baret F. 1995 - *Use of spectral reflectance variation to retrieve canopy biophysical characteristics*. In: “Advances in environmental remote sensing”. Danson F.M., Plummer S.E. Eds, Wiley, Chichester: 33-51.
- Biondi F., Myers D.E., Avery C.C. 1994 - *Geostatistically modeling stem size and increment in an old-growth forest*. Canadian Journal of Forest Research 24 (7): 1354-1368. doi: 10.1139/x94-176
- Blackard J. A., Dean D.J. 1999 - *Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables*. Computers and Electronics in Agriculture 24 (3):131-151. doi: 10.1016/S0168-1699(99)00046-0
- Breiman L. 1996 - *Bagging predictors*. Machine Learning 24 (2): 123-140. doi: 10.1023/A:1018054314350
- Breiman L. 2001 - *Random forests*. Machine Learning 45 (1): 5-32. doi: 10.1023/A:1010933404324
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 1984 - *Classification and Regression Trees*. Wadsworth, Belmont. 368 p.
- Brosofske K.D., Froese R.E., Falkowski M.J., Banskota A. 2014 - *A review of methods for mapping and prediction of inventory attributes for operational forest management*. Forest Science 60 (4): 733-756. doi: 10.5849/forsci.12-134
- Brunsdon C., Fotheringham A.S., Charlton M.E. 1996 - *Geographically weighted regression: A method for exploring spatial nonstationarity*. Geographical Analysis 28 (4): 281-298. doi: 10.1111/j.1538-4632.1996.tb00936.x
- Brus D.J., Hengeveld G.M., Walvoort D.J.J., Goedhart P.W., Heidema A.H., Nabuurs G.J., Gunia K. 2012 - *Statistical mapping of tree species over Europe*. European Journal of Forest Research 131 (1): 145-157. doi: 10.1007/s10342-011-0513-5
- Carbonell J.G., Michalski R.S., Mitchell T.M. 1983 - *An Overview of Machine Learning*. In: “Machine Learning”. Symbolic Computation, Michalski R.S., Carbonell J.G., Mitchell T.M. Eds., Springer, Berlin. 3-23. doi: 10.1007/978-3-662-12405-5_1
- Carvalho L.M.T. 2001 - *Mapping and monitoring forest remnants: A multiscale analysis of spatio-temporal data*.

- PhD thesis. Wageningen University, 138 p.
- Chave J., Andalo C., Brown S., Cairns M.A., Chambers J.Q., Eamus D., Fölster H., Fromard F., Higuchi N., Kira T., Lescuré J.P., Nelson B.W., Ogawa H., Puig H., Riéra B., Yamakura T. 2005 - *Tree allometry and improved estimation of carbon stocks and balance in tropical forests*. *Oecologia* 145 (1): 87-99. doi: 10.1007/s00442-005-0100-x
- 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 Sensing of Environment* 112 (5): 2686-2700. doi: 10.1016/j.rse.2008.01.002
- Chirici G., Scotti R., Montagni A., Barbati A., Cartisano R., Lopez G., Marchetti M., McRoberts R.E., Olsson H., Corona P. 2013 - *Stochastic gradient boosting classification trees for forest fuel types mapping through airborne laser scanning and IRS LISS-III imagery*. *International Journal of Applied Earth Observation and Geoinformation* 25: 87-97
- Clement F., Orange D., Williams M., Mulley C., Epprecht 2009 - *Drivers of afforestation in Northern Vietnam: Assessing local variations using geographically weighted regression*. *Applied Geography* 29 (4): 561-576. doi: 10.1016/j.apgeog.2009.01.003
- Cleveland W.S., Devlin S.J. 1988 - *Locally weighted regression: an approach to regression analysis by local fitting*. *Journal of the American statistical association* 83 (403): 596-610. doi: 10.1080/01621459.1988.10478639
- Corona P. 2000 - *Introduzione al rilevamento campionario delle risorse forestali*. EdizioniCUSL, Firenze. 283 p.
- Corona P. 2010 - *Integration of forest mapping and inventory to support forest management*. *iForest* 3: 59-64. doi: 10.3832/ifor0531-003
- Corona P. 2016 - *Consolidating new paradigms in large-scale monitoring and assessment of forest ecosystems*. *Environmental Research* 144 (Pt B): 8-14. doi: 10.1016/j.envres.2015.10.017.
- Corona P. 2018 - *Communicating facts, findings and thinking to support evidence-based strategies and decisions*. *Annals of Silvicultural Research* 42: 1-2. doi: 10.12899/ASR-1617.
- Corona P., Chirici G., Marchetti M. 2002 - *Forest ecosystem inventory and monitoring as a framework for terrestrial natural renewable resource survey programmes*. *Plant Biosystems* 136 (1): 69-82. doi: 10.1080/11263500212331358531
- Corona P., Chirici G., McRoberts R.E., Winter S., Barbati A. 2011 - *Contribution of large-scale forest inventories to biodiversity assessment and monitoring*. *Forest ecology and management* 262 (11): 2061-2069. doi: 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 sensing of Environment* 152: 29-37. doi: 10.1016/j.rse.2014.05.011
- Cressie N. 1990 - *The origins of kriging*. *Mathematical Geology* 22 (3): 239-252. doi: 10.1007/BF00889887
- Cressie N. 1993 - *Statistics for spatial data*. Wiley, New York. 900 p. doi: 10.1002/9781119115151
- Deutsch C., Journel A. 1992 - *Geostatistical Software Library and User's Guide*. Oxford University Press, New York. 340 p.
- Fattorini L., Marcheselli M., Pisani C. 2006 - *A three-phase sampling strategy for large-scale multiresource forest inventories*. *Journal of Agricultural, Biological, and Environmental Statistics* 11 (3): 296-316. doi: 10.1198/108571106X130548
- Fattorini L., Marcheselli M., Pratelli L. 2018a - *Design-based maps for finite populations of spatial units*. *Journal of the American Statistical Association* 113 (522): 686-697. doi: 10.1080/01621459.2016.1278174
- Fattorini L., Marcheselli M., Pisani C., Pratelli L. 2018b - *Design-based maps for continuous spatial populations*. *Biometrika* 105 (2): 419-429. doi: 10.1093/biomet/asy012
- Fix E., Hodges J.L. 1951 - *Discriminatory analysis. Nonparametric discrimination: consistency properties*. Report Number 4, Project 21-49-004. USAF School of Aviation Medicine. Randolph Field, Texas.
- Foody G.M. 2003 - *Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship*. *Remote Sensing of Environment* 88 (3): 283-293. doi: 10.1016/j.rse.2003.08.004
- Foody G.M., Cutler M.E., McMorro J., Pelz D., Tangki H., Boyd D.S., Douglas I. 2001 - *Mapping the Biomass of Bornean Tropical Rain Forest from Remotely Sensed Data*. *Global Ecology and Biogeography* 10 (4): 379-387.
- Franco-Lopez H., Ek A.R., Bauer M.E. 2001 - *Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method*. *Remote Sensing of Environment* 77 (3): 251-274. doi: 10.1016/S0034-4257(01)00209-7
- Freeman E.A., Moisen G.G. 2007 - *Evaluating Kriging as a Tool to Improve Moderate Resolution Maps of Forest Biomass*. *Environmental Monitoring and Assessment* 128 (1-3): 395-410. doi:10.1007/s10661-006-9322-6
- Freund Y., Schapire R. 1996 - *Experiments with a new boosting algorithm*. *Proceedings of the 13th International Conference on Machine Learning*, Bari: 148-156.
- Garcia-Gutierrez J., Martínez-Álvarez F., Troncoso A., Riquelme J.C. 2014 - *(2014) A Comparative Study of Machine Learning Regression Methods on LiDAR Data: A Case Study*. In: "International Joint Conference SOCO'13-CISIS'13-ICEUTE'13. *Advances in Intelligent Systems and Computing*, vol. 239", Herrero Á. et al. Eds., Springer-Cham. 249-258. doi: 10.1007/978-3-319-01854-6_26
- Gauch H.G. 1982 - *Multivariate analysis in community ecology*. Cambridge University Press, New York. 298 p. doi: 10.1017/CBO9780511623332
- Gjertsen A.K., Tomppo E.O., Tomter S. 1999 - *National forest inventory in Norway: using sample plots, digital maps, and satellite images*. *IEEE 1999 International Geoscience and Remote Sensing Symposium*, Hamburg. doi: 10.1109/IGARSS.1999.774421.
- Goovaerts P. 1997 - *Geostatistics for natural resources evaluation*. Oxford university press, New York. 483 p.
- Görgens E.B., Montagni A., Rodriguez L.C.E. 2015 - *A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics*. *Computers and Electronics in Agriculture* 116: 221-227. doi: 10.1016/j.compag.2015.07.004
- Gregoire T.G., Valentine H.T. 2008 - *Sampling strategies for natural resources and the environment*. Chapman & Hall, New York. 494 p.
- Hastie T.J., Tibshirani R.J., Friedman J.H. 2009 - *The elements of statistical learning: data mining, inference and prediction* 2nd ed. Springer, New York. 745 p. doi: 10.1007/978-0-387-84858-7
- Helmer E.H., Ruzyccki T.S., Wunderle J.M., Vogesser S., Ruefenacht B., Kwit C., Brandeis T.J., Ewert D.N. 2010 - *Mapping tropical dry forest height, foliage height profiles and disturbance type and age with a time series of cloud-cleared Landsat and ALI image mosaics to characterize avian habitat*. *Remote Sensing of Environment* 114 (11): 2457-2473. doi: 10.1016/j.rse.2010.05.021

- Hengl T., Heuvelink G.B.M., 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.
- Hernandez-Stefanoni J.L., Ponce-Hernandez R. 2006 - *Mapping the Spatial Variability of Plant Diversity in a Tropical Forest: Comparison of Spatial Interpolation Methods*. Environmental Monitoring and Assessment 117 (1-3): 307-334. doi: 10.1007/s10661-006-0885-z
- Holmgren P., Persson R. 2002 - *Evolution and prospects of global forest assessments*. Unasylva 53 (210): 3-9. url: <http://www.fao.org/docrep/005/y4001e/y4001e02.htm>
- Hotelling H. 1936 - *Relations between two sets of variates*. Biometrika 28 (3): 321-377. doi:10.1093/biomet/28.3-4.321
- Huijbregts C.J., Matheron G. 1971 - *Universal kriging (an optimal method for estimating and contouring in trend surface analysis)*. Proceedings of ninth international symposium on techniques for decision-making in the mineral industry, The Canadian institute of mining and metallurgy, 12: 159-169.
- Hudak A.T., Lefsky M.A., Cohen W.B., Berterretche M. 2002 - *Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height*. Remote Sensing of Environment 82 (2-3): 397-416. doi: 10.1016/S0034-4257(02)00056-1
- Hudak A.T., Crookston N.L., Evans J.S., Hall D.E., Falkowski M.J. 2008 - *Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data*. Remote Sensing of Environment 112 (5): 2232-2245. doi: 10.1016/j.rse.2007.10.009
- Ingram J.C., Dawson T.P., Whittaker R.J. 2005 - *Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks*. Remote Sensing of Environment 94 (4): 491-507. doi: 10.1016/j.rse.2004.12.001
- Jensen J.R., Qiu F., Ji M.H. 1999 - *Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data*. International Journal of Remote Sensing 20: 2805-2822. doi: 10.1080/014311699211804
- Journel A.G., Huijbregts C.J. 1978 - *Mining Geostatistics*. Academic Press, London. 600 p.
- Kangas A., Maltamo M. 2007 - *Forest inventory. Methodology and applications*. Springer, Dordrecht. 362 p. doi: 10.1007/1-4020-4381-3
- Katila M., Tomppo E.O. 2001 - *Selecting estimation parameters for the Finnish multisource national forest inventory*. Remote Sensing of Environment, 76 (1): 16-32. doi: 10.1016/S0034-4257(00)00188-7
- Kleinn C. 2002 - *New technologies and methodologies for national forest inventory*. Unasylva 53 (210): 10-15. url: <http://www.fao.org/docrep/005/y4001e/Y4001E03.htm>
- Köhl M., Magnussen S.S., Marchetti M. 2006 - *Sampling methods, remote sensing and GIS multiresource forest inventory*. Springer, Heidelberg. 373 p. doi: 10.1007/978-3-540-32572-7
- Krige D.G. 1978 - *Lognormal-de Wijsiangeostatistics for ore evaluation*. South African Institute of Mining and Metallurgy Monograph Series, Johannesburg. 60 p.
- Lund G. 1998 - *A comparison of multipurpose resource inventories (MRIs) throughout the world*. EFI working paper 14, Joensuu.
- Lochhead K., LeMay V., Bull G., Schwab O., Halperin J. 2018 - *Multivariate estimation for accurate and logically consistent forest-attributes maps at macroscales*. Canadian Journal of Forest Research 48 (4): 345-359. doi: 10.1139/cjfr-2017-0221
- Mandallaz D. 2000 - *Estimation of the spatial covariance in Universal Kriging: Application to forest inventory*. Environmental and Ecological Statistics 7 (3): 263-284. doi: 10.1023/A:1009619117138
- Mas J.F., Puig H., Palacio J.L., Sosa-López A. 2004 - *Modelling deforestation using GIS and artificial neural networks*. Environmental Modelling & Software 19 (5): 461-471. doi: 10.1016/S1364-8152(03)00161-0
- Maselli F. 2002 - *Improved estimation of environmental parameters through locally calibrated multivariate regression analyses*. Photogrammetric engineering and Remote Sensing 68 (11): 1163-1171.
- Maselli F., Chiesi M. 2006 - *Evaluation of statistical methods to estimate forest volume in a Mediterranean region*. IEEE Transactions on Geoscience and Remote Sensing 44 (8): 2239-2250. doi: 10.1109/TGRS.2006.872074
- Maselli F., Chirici G., Bottai L., Corona P., Marchetti M. 2005 - *Estimation of Mediterranean forest attributes by the application of k NN procedures to multitemporal Landsat ETM+ images*. International Journal of Remote Sensing 26 (17): 3781-3796. doi: 10.1080/01431160500166433
- Maselli F., Chiesi M., Corona P. 2014 - *Use of geographically weighted regression to enhance the spatial features of forest attribute maps*. Journal of Applied Remote Sensing 8 (1): 083533 doi: 10.1117/1.JRS.8.083533
- Matheron G. 1963 - *Principles of geostatistics*. Economic Geology 58 (8): 1246-1266. doi: 10.2113/gsecongeo.58.8.1246
- Matheron G. 1969 - *Le Krigeage Universel*. Cahiers du Centre de Morphologie Mathématique. Fontainebleau.
- Mattioli W., Quatrini V., Di Paolo S., Di Santo D., Giuliarelli D., Angelini A., Portoghesi P., Corona P. 2012 - *Experimenting the design-based k-NN approach for mapping and estimation under forest management planning*. iForest 5 (1): 26-30. doi: 10.3832/ifor0604-009
- McCulloch W., Pitts W. 1943 - *A logical calculus of ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics 5 (4): 115-133. doi:10.1007/BF02478259
- McInerney D.O., Nieuwenhuis M. 2009 - *A comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory*. International Journal of Remote Sensing 30 (19): 4937-4955. doi: 10.1080/01431160903022936
- McRoberts R.E. 2012 - *Estimating forest attribute parameters for small areas using nearest neighbors techniques*. Forest Ecology and Management 272: 3-12. doi: 10.1016/j.foreco.2011.06.039
- McRoberts R.E., Tomppo E.O. 2007 - *Remote sensing support for national forest inventories*. Remote sensing of environment 110 (4): 412-419. doi: 10.1016/j.rse.2006.09.034
- McRoberts R.E., Tomppo E.O., Finley A.O., Heikkinen J. 2007 - *Estimating areal means and variances of forest attributes using the k-nearest neighbours technique and satellite imagery*. Remote Sensing of Environment 111 (4):466-480. doi: 10.1016/j.rse.2007.04.002
- McRoberts R.E., Tomppo E.O., Næsset E. 2010a - *Advances and emerging issues in national forest inventories*. Scandinavian Journal of Forest Research 25 (4): 368-381. doi: 10.1080/02827581.2010.496739
- McRoberts R.E., Cohen W.B., Næsset E., Stehman S.V., Tomppo E.O. 2010b - *Using remotely sensed data to construct and assess forest attribute maps and related spatial products*. Scandinavian Journal of Forest Research 25 (4): 340-367. doi: 10.1080/02827581.2010.497496
- Meng Q., Cieszewski C., Madden M. 2009 - *Large area forest inventory using Landsat ETM+: A geostatistical approach*. ISPRS Journal of Photogrammetry and Remote Sensing 64 (1): 27-36. doi: 10.1016/j.isprsjprs.2008.06.006
- Moeur M., Stage A.R. 1995 - *Most similar neighbour: an improved sampling inference procedure for natural resource planning*. Forest Science 41 (2): 337-359. doi:

- 10.1093/forests/41.2.337
- Odeh I.O.A., McBratney A.B., Chittleborough D.J. 1995 - *Further results on prediction of soil properties from terrain attributes: Heterotopic kriging and regression-kriging*. Geoderma 67 (3): 215-226. doi: 10.1016/0016-7061(95)00007-B
- Ohmann J.L., Gregory M.J. 2002 - *Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, USA*. Canadian Journal of Forest Research 32 (4): 725-741. doi: 10.1139/x02-011
- Ohmann J.L., Gregory M.J., Henderson E.B., Roberts H.M. 2011 - *Mapping gradients of community composition with nearest neighbour imputation: extending plot data for landscape analysis*. Journal of Vegetation Science 22 (4): 660-676. doi: 10.1111/j.1654-1103.2010.01244.x
- Opsomer J.D., Breidt J., Moisen G.G., Kauermann G. 2007 - *Model-assisted estimation of forest resources with Generalized Additive Models*. Journal of the American Statistical Association 102 (478): 400-416. doi: 10.1198/016214506000001491
- Papritz A., Stein A. 1999 - *Spatial prediction by linear kriging*. In: "Spatial Statistics for Remote Sensing". Remote Sensing and Digital Image Processing 1, Stein A., Van der Meer F., Gorte B. Eds. Springer, Dordrecht: 83-113. doi: 10.1007/0-306-47647-9_6
- Reese H., Nilsson M., Granqvist P.T., 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*. A Journal of the Human Environment 32(8):542-548. doi: 10.1579/0044-7447-32.8.542
- Särndal C.E., Swensson B., Wretman J. 1992 - *Model assisted survey sampling*. Springer-Verlag, New York. 695 p.
- Schabenberger O., Pierce F. 2002 - *Contemporary statistical models for the plant and soil sciences*. CRC Press, Boca Raton. 760 p.
- Schabenberger O., Gotway C.A. 2005 - *Statistical methods for spatial data analysis*. CRC Press, Boca Raton. 512 p.
- Schapire R., Freund Y., Bartlett P., Lee W. 1998 - *Boosting the margin: A new explanation for the effectiveness of voting methods*. Annals of Statistics 26 (5): 1651-1686. doi: 10.1214/aos/1024691352
- Schowengerdt R.A. 2006 - *Remote Sensing: Models and Methods for Image Processing*^{3rd} ed. Academic Press, Burlington. 560 p.
- Shepard D. 1968 - *A two-dimensional interpolation function for irregularly-spaced data*. Proceedings of the 1968 ACM National Conference, 517-524. doi:10.1145/800186.810616
- Song Y., Lu Y. 2015 - *Decision tree methods: applications for classification and prediction*. Shanghai archives of Psychiatry 27 (2): 130-135. doi: 10.11919/j.issn.1002-0829.215044
- Subedi N., Zhang L., Zhen Z. 2018 - *Bayesian geographically weighted regression and its application for local modeling of relationships between tree variables*. iForest 11: 542-552. doi: 10.3832/ifer2574-011
- ter Braak C.J.F. 1986 - *Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis*. Ecology 67 (5): 1167-1179. doi: 10.2307/1938672
- Tobler W.R. 1970 - *A computer movie simulating urban growth in the Detroit region*. Economic Geography 46: 234-240.
- Thompson S.K. 2002 - *Sampling* 2nd ed. Wiley, New York. 400 p.
- Tomppo E.O. 1991 - *Satellite image-based national forest inventory of Finland*. International Archives of Photogrammetry and Remote Sensing 28 (7): 419-424.
- Torresan C., Corona P., Scrinzi G., Valls Marsal J. 2016 - *Using classification trees to predict forest structure types from LiDAR data*. Annals of Forest Research 59 (2): 281-298. doi: 10.15287/afr.2016.423
- Urban D.L. 2002 - *Classification and regression trees*. In: "Analysis of ecological communities". MjMSoftware Design, McCune B., Grace J.B. Eds., Gleneden Beach: 222-232.
- Viana H., Aranha J., Lopes D., Cohen W.B. 2012 - *Estimation of crown biomass of Pinuspinaster stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models*. Ecological Modelling 226: 22-35. doi: 10.1016/j.ecolmodel.2011.11.027
- Wackernagel H. 2003 - *Multivariate Geostatistics: an introduction with applications* 3rd ed. Springer-Verlag, Berlin. 388 p. doi: 10.1007/978-3-662-05294-5
- Wang Q., Ni J., Tenhunen J. 2005 - *Application of a geographically weighted regression analysis to estimate net primary production of Chinese forest ecosystems*. Global Ecology and Biogeography 14 (4): 379-393. doi: 10.1111/j.1466-822X.2005.00153.x
- Young T.M., Wang Y., Hodges D.G., Guess F.M. 2009 - *Decision tree applications for forestry and forest products manufacturers*. Proceedings of the Southern forest economics workers annual meeting. University of Georgia Athens.
- Zhang L., Shi H. 2004 - *Local Modeling of Tree Growth by Geographically Weighted Regression*. Forest Science 50(2): 225-244.