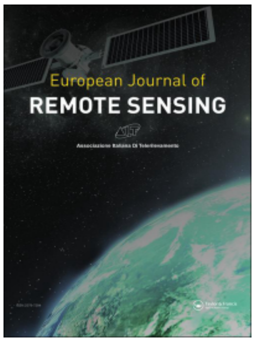

European Journal of Remote Sensing

ISSN: 2279-7254 (Online) Journal homepage: www.tandfonline.com/journals/tejr20

Small area estimation of forest volume using mixed effects random forests and multi-source remote sensing data

Elia Vangi

To cite this article: Elia Vangi (2026) Small area estimation of forest volume using mixed effects random forests and multi-source remote sensing data, European Journal of Remote Sensing, 59:1, 2680926, DOI: 10.1080/22797254.2026.2680926

To link to this article: https://doi.org/10.1080/22797254.2026.2680926

EUROPEAN JOURNAL OF REMOTE SENSING
2026, VOL. 59, NO. 1, 2680926
https://doi.org/10.1080/22797254.2026.2680926

Taylor & Francis
by informa

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Small area estimation of forest volume using mixed effects random forests and multi-source remote sensing data

Elia Vangi

geoLAB - Laboratory of Forest Geomatics, Department of Agriculture, Food, Environment and Forestry, Università degli Studi di Firenze, Firenze, Italy.

ABSTRACT

Accurate estimation of forest growing stock volume (GSV) at fine spatial scales is essential for sustainable management, carbon accounting, and local decision-making. However, traditional inventories often lack sufficient sampling density for small areas. This study evaluates two small-area estimation (SAE) approaches: the Empirical Best Predictor (EBP), based on a nested-error linear regression model, and the Mixed-Effects Random Forest (MERF), using multi-source remote sensing data. The analysis was conducted in the Vallombrosa Nature Reserve (Italy), integrating field measurements from 101 plots with auxiliary variables from Sentinel-2 and airborne LiDAR. Both methods estimated mean and total GSV across 658 forest stands, many lacking direct observations. Performance was assessed via spatial cross-validation, and uncertainty was quantified using root-mean-square error (RMSE). Results show that MERF outperformed EBP in predictive accuracy, achieving higher R^2 (0.67 vs. 0.37) and lower RMSE (151 vs. 202 m^3 ha^{-1}). MERF produced more stable uncertainty estimates with improved coverage. While both methods yielded comparable total GSV, EBP exhibited greater sensitivity to model assumptions. Conversely, MERF effectively captured non-linear relationships and handled multicollinearity, despite reduced interpretability and higher computational demand. Overall, findings highlight the advantages of integrating machine learning with mixed-effects modeling for SAE under sparse sampling.

ARTICLE HISTORY
Received 24 April 2026
Revised 20 May 2026
Accepted 22 May 2026

KEYWORDS
forest inventory; remote sensing; small-area estimation; mixed effects; random forests

Introduction

Forests are a fundamental element of climate change mitigation and biodiversity conservation efforts, and monitoring them is crucial to promoting sustainable management, carbon accounting, and the evaluation of ecosystem services. Traditional national forest inventories (NFIs), such as those implemented in many European countries, typically rely on sparse ground plot networks and purely design-based inference techniques (Tomppo et al., 2008; McRoberts et al., 2009). While these approaches enable unbiased estimates of forest parameters at national or regional levels, they are generally inadequate for generating reliable (i.e. with errors below the maximum tolerance) data at smaller administrative or ecological units, such as municipalities or protected areas, due to limited plot density.

As decision-making processes become more decentralized and spatially detailed, there is an increasing need for forest information at local levels, along with reliable uncertainty estimates. This need arises at two related levels. On the one hand, countries require robust estimates of forest variables (e.g. growing stock volume, carbon stocks) for their international reporting obligations under frameworks such as the UNFCCC, LULUCF, and REDD+. These estimates must explicitly quantify uncertainty, in line with the IPCC's good-practice guidelines. On the other hand, subnational actors, such as municipalities, regions, and forest managers, require reliable, detailed spatial data to support operational decisions and policymaking. Local estimates and maps of forest resources are essential for several forest activities, including management, conservation, risk assessment, and quantifying ecosystem services. Local forest

CONTACT Elia Vangi elia.vangi@unifi.it geoLAB - Laboratory of Forest Geomatics, Dept. of Agriculture, Food, Environment and Forestry, Università degli Studi di Firenze, Via San Bonaventura 13, 50145, Firenze, Italy

Supplemental data for this article can be accessed online at https://doi.org/10.1080/22797254.2026.2680926.

management plays a crucial role in sustainable land-use planning, as it provides the foundation for balancing ecological, economic, and social objectives at the landscape scale. Accurate knowledge of growing stock volume (GSV) and above-ground biomass (AGB) is fundamental in this context, as it directly informs decisions on harvesting intensity, timing, and the spatial allocation of interventions. By quantifying the available forest resources, forest managers can ensure that harvest levels remain within sustainable limits, avoiding overexploitation while optimizing economic returns. Moreover, reliable GSV estimates support the design of silvicultural treatments, the planning of infrastructure such as forest roads, and the assessment of carbon stocks and ecosystem services. Here, aggregated statistics alone are insufficient; spatially explicit, fine-scale data, preferably linked to measures of reliability and confidence, should be the standard. Addressing both needs simultaneously requires estimation frameworks that combine the statistical rigor of forest inventories with the spatial detail and coverage of remote sensing data.

To address the limitations of traditional inventory methods, the concept of Enhanced Forest Inventories (EFI) has been developed, combining ground measurements with comprehensive remote-sensing data using model-based techniques (White et al., 2025). EFI methods have shown their ability to produce detailed forest attribute maps, such as GSV or AGB, by utilizing the predictive capabilities of satellite and airborne data (e.g. Landsat, GEDI, LiDAR), often employing regression or machine learning algorithms (Chirici et al., 2020; Vangi et al., 2021; Saarela et al., 2015).

However, a key challenge remains: while predictive maps provide detailed spatial information, they rarely include reliable uncertainty estimates at aggregated levels (Saarela et al., 2020). Simply aggregating pixel-based predictions across small areas does not yield valid estimates, as it overlooks spatial model bias and fails to account for sampling error. To address this problemSmall-Area Estimation (SAE) provides a rigorous statistical framework. Originating from official statistics (Molina & Rao, 2015), SAE techniques have been adapted for forest inventory applications to generate unbiased and accurate estimates in domains with few or no field plots (Breidenbach & Astrup, 2012; Goerndt et al., 2013; Magnussen & Breidenbach, 2017). SAE methods can operate at the unit or area level. Traditional SAE methods rely heavily on model-based approaches that "borrow strength" across areas by exploiting auxiliary information. Among these, the Empirical Best Predictor (EBP) under unit-level mixed models—most commonly the nested-error linear regression model (Battese et al., 1988)—has long served as a benchmark due to its solid theoretical foundation, design-consistent properties, and relatively straightforward implementation. However, its reliance on strict linearity, normality, and homoscedasticity can limit its efficacy in complex ecosystems (Molina & Marhuenda, 2015). Forestry applications are frequently characterized by high-dimensional remote sensing covariates, non-linear structural interactions, and spatial heterogeneity. Under these conditions, the rigid parametric assumptions of classical EBP risk model misspecification, potentially leading to biased domain estimates and deflated predictive efficiency (Rojas-Perilla et al., 2020).

These limitations have motivated the development of more flexible, data-driven approaches that can better capture complex patterns while retaining the core idea of borrowing strength across areas. In the context of estimating and mapping socio-economic indicators (i.e. poverty, income, inequality; Hobza & Morales, 2016), as well as in the health care field (i.e. care-cost opportunity work) and policy settings, the Mixed-Effects Random Forest (MERF) has emerged as a promising alternative (Hajjem et al., 2014; Krennmair et al., 2022), demonstrating superior performance compared to classical SAE methodology, in both unit and area-level frameworks. MERF combines the strengths of machine learning methods—specifically, random forests' ability to model non-linearities and interactions without explicit specification (Breiman, 2001)—with the hierarchical structure of mixed-effects models, which account for area-level heterogeneity through random effects. By iteratively estimating fixed effects via random forests and random effects via mixed-model techniques, MERF provides a hybrid framework that relaxes parametric assumptions while preserving the capacity to account for hierarchical dependencies. However, despite the immense popularity of standard Random Forests in forestry, their integration into SAE has historically been hindered by their inability to naturally account for such grouped dependencies. This study addresses this gap by presenting the first application of MERF to forest inventory management and introducing a data-driven framework capable of mapping localized forest attributes, even in data-scarce zones. While recent modifications have successfully enhanced MERF's performance—such as using PCA or rotation forests to address multicollinearity (Ananda et al., 2024)—its introduction to the forestry domain opens new methodological and practical avenues for interpretability, uncertainty quantification, and computational efficiency.

These approaches are especially practical when auxiliary variables, such as canopy height models (CHM), GEDI RH metrics, or spectral indices, are strongly correlated with the target variable. Recent studies highlight the importance of generating not just estimates but also uncertainty maps, such as pixel-level RMSE or confidence intervals (Saarela et al., 2020; Foody et al., 2003). This information is crucial for transparent reporting, model validation, and making risk-informed decisions in forestry and land use. Furthermore, although high-quality auxiliary data (such as Landsat, Sentinel-1, and Sentinel-2 composites) are becoming increasingly accessible, they have not yet been fully integrated into design-based SAE at the forest stand level. Few studies have examined SAE applications in the Mediterranean context (e.g. Galidaki et al., 2017; Mesquida et al., 2026), and, to our knowledge, none have compared the EBP with the MERF framework in the forest sector.

This paper contributes to the growing literature on SAE by providing a comparative analysis of EBP and MERF in forest inventory, with particular emphasis on the trade-offs among model interpretability, flexibility, and predictive performance. By highlighting the conditions under which each approach is most appropriate, this study aims to clarify the role of modern machine learning methods within the established SAE paradigm. Furthermore, this study aims to bridge the gap between advanced SAE methodologies and their practical application in forest management. I introduce a reproducible approach to estimate GSV and its associated uncertainty at the stand level using field plots and multi-source remote-sensing-based auxiliary variables.

Materials and methods

Study area

The study was conducted in the Nature Reserve of Vallombrosa (43.745 E, 11.562 *N*), located in the Tuscany region, on the north-west side of the Pratomagno Massif (Italy) (Figure 1). The Nature Reserve of

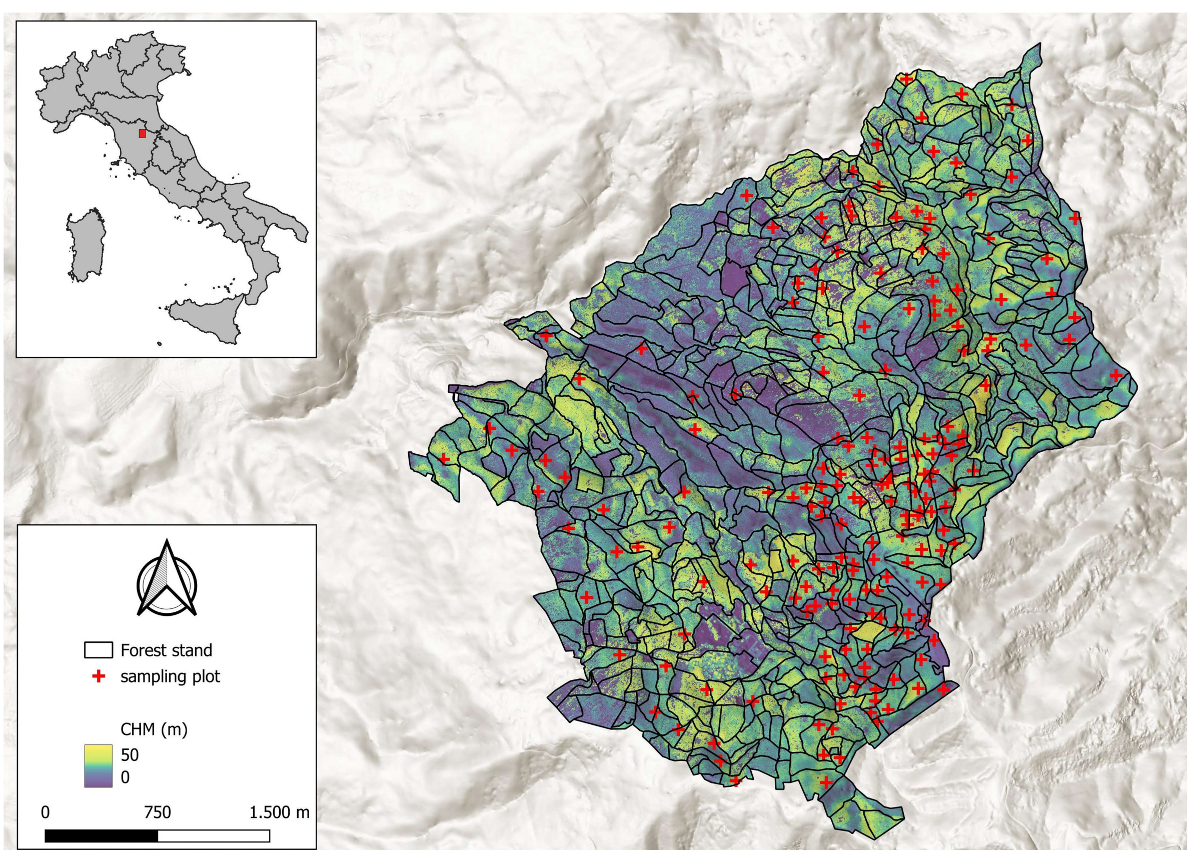

Figure 1. Boundaries of the forest stands and distribution of the sampling field plot. In the background, the CHM (in m).

Vallombrosa covers 1273 ha and ranges in altitude from 470 to 1447 m a.s.l.. The climate is temperate-humid, with Mediterranean-type rainfall and a summer minimum. Mean annual temperature is 9.7 °C and total annual rainfall is 1337 mm, with an average of 71.2 mm in the summer months (June-August). Parent material is Oligocene sandstone of the "Macigno del Chianti" formation. The forest grows on sandy-loamy or loamy soils with rich humus in the surface horizons and generally steep or very steep slopes. The dominant forest species is silver fir (*Abies alba* Mill.) (664 ha), followed by European beech (*Fagus sylvatica* L.). Traditionally, silver fir has been cultivated in pure stands in the Vallombrosa Forest for many centuries, while European beech mainly originated from coppice conversion to high forest (Ciancio, 2009). Other common forest species included *Pinus nigra* J.F. Arnold, *Pseudotsuga menziesii* (Mirb.) Franco, introduced for experimental purposes, and *Castanea sativa* Mill., *Quercus cerris* L., *Quercus pubescens* Willd., *Ostrya carpinifolia* Scop., *and Fraxinus ornus* L. The forest management plan for the period 2006–2025 (Ciancio, 2009) divided the Reserve into 658 forest stands (i.e. management units) homogeneous in species composition and environmental condition. The mean domain area is 1.9 ha, ranging from 0.15 ha to 13 ha. The Reserve Management plans are currently being revised in preparation for the next management cycle (2026–2046). The next plan will guide the harvesting period, the maximum amount of wood that can be removed from each forest stand, and will set the monitoring system for forest resources and biodiversity. For this reason, detailed spatial information on GSV distribution and its uncertainty will inform plan updating and intervention planning.

Field data

In this study, field data from 101 circular sampling plots with a radius of 13 m were used. Field data were collected as part of a systematic, unaligned sampling design with a 1-ha hexagonal grid, stratified on a high-resolution canopy height model acquired in 2015 (see details in paragraph 2.4.2). We selected plots measured in the winter seasons of 2015/2016 to best match the remote sensing data and produce a single-year GSV map. Within each plot, tree species, diameter at breast height (DBH), tree height, and age were measured and aggregated at the plot level. From structural variables (i.e. DBH and height), growing stock volume (GSV) was computed using species-specific allometric equations developed within the Italian NFI field survey framework (Tabacchi et al., 2011). Only 79 of 658 forest stands (~12% of the total) have at least one sample plot, with a median of one and a maximum of four plots. Table 1 reports field measurements, grouped by dominant species. This study used GSV plot-level data to estimate the total and mean GSV indicators for all forest stands in the management plan, along with their uncertainties (mean squared error [MSE] and root mean squared error [RMSE]).

Forest mask

To ensure reliable results in small-area GSV estimation, a forest mask was needed to restrict the estimation to forest areas. The forest mask produced by D'Amico et al. (2021) and further updated for the reference year 2020 (D'Amico et al., 2023) was used, which closely mimics the standard forest definition used by the Italian NFI and the FAO, and is also used by the Vallombrosa Reserve management plan.

Remote sensing variables

SAE relies on a statistical model that links field observations to auxiliary variables available at the unit level, i.e. over the entire area of interest (often referred to as wall-to-wall). RS variables meet this fundamental

Table 1. Average structural information at the dominant species level.

Dominant species	DBH	N	BA	H	GSV	AGE
A.Alba	29.5	835.8	3.4	22.4	817.4	87.6
C.Sempervirens	27.7	1509.4	5.0	20.0	945.1	92.0
F. sylvatica	25.2	882.2	2.6	19.0	604.9	117.5
F. Excelsior	12.3	3264.2	2.9	12.5	422.3	62.0
P. abies	26.9	811.3	3.7	18.7	912.6	102.0
P. nigra	13.4	2811.3	3.0	9.7	355.7	62.0
P. menziesii	47.1	554.7	4.5	29.3	1194.8	89.2

criterion and are usually globally available, offering high spatial and temporal resolution. In this study, a total of 13 RS variables (hereinafter referred to as predictors), mostly derived from the Sentinel-2 mission, were used to fit the nested-error linear regression and random forest models for the EBP and MERF methods, respectively.

Sentinel-2-based predictors

All S2 surface reflectance level 2 images acquired over the study area from the Google Earth Engine archive (Gorelick et al., 2017) from 1st April to 30th September 2015, with less than 50% cloud cover (White et al., 2014), were used. The S2 cloud probability dataset (Shi et al., 2022) was used to mask all pixels with a cloud probability of 65% or higher (Parisi and Vangi, 2023). From each image, six spectral bands were retained (red, green, blue, nir, swir1, and swir2) and five vegetation indices (VIs) were computed (Table 2), resulting in a total of 11bands per image (6 original bands + 5 VIs). A cloud-free summer composite mosaic for each band and VI using the medoid method (Kennedy et al., 2018) was constructed. The Medoid pixel compositing algorithm seeks to construct the final image using pixels whose surface reflectance values are as close as possible to the median derived from the entire image collection. The distance from the median was calculated via the Euclidean distance.

ALS-based predictors

The airborne laser scanning (ALS) dataset was obtained from a survey conducted in May 2015 using a RIEGL LMS-Q680i LiDAR system (Horn, Austria) and a DIGICAM H39 RGB and CIR optical sensor, both installed on an Eurocopter AS350 B3 helicopter. The mission was flown at approximately 1100 m above ground level, at 70 knots, with 30% overlap between adjacent flight lines. The LiDAR system collected full-waveform data with an average point density of 10 points per square meter. Georeferencing of the ALS data was performed in the WGS84 UTM32N coordinate system by refining flight trajectories with Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) data acquired from two Italtopos network base stations (Giannetti et al., 2018). Standard preprocessing steps—such as outlier and noise removal, classification of ground and non-ground returns, and height normalization—were performed using LAStools (version 240220; Gilching, Germany). Subsequently, digital terrain and digital surface models (DTM and DSM) were generated at a 1 m spatial resolution, and the canopy height model (CHM) was derived by subtracting the DTM from the DSM.

Auxiliary predictors

To complement the set of Sentinel-2 + ALS-based predictors, the elevation from the DTM and the position (X and Y coordinates) were added. These predictors are strictly related to the position of each unit (i.e. pixel), and incorporating them allows the model to capture part of the spatial structure through the fixed effects component. Although this approach does not fully model spatial dependences, it provides a simple and effective way to mitigate spatial autocorrelation when more complex spatial models are not adopted.

Each predictor was produced at a 10 × 10 m resolution. Within each sampling plot, the mean values for each predictor in the set were computed.

Methods

The empirical best predictor estimator

The benchmark estimator for the stand-level mean GSV was the EBP, as described by Molina and Rao (2010) and Molina and Rao (2015).

Table 2. List of vegetation indices used as predictors, along with their formulation and source.

VI	Formula	Reference
Normalized difference vegetation index	$NDVI = (NIR-R)/(NIR + R)$	Jönsson et al. (2018)
Normalized burnt ratio	$NBR = (NIR-SWIR)/(NIR + SWIR)$	Roy et al. (2006)
Enhanced vegetation index	$EVI = 2.5(NIR-RED)/(NIR + 6RED-7.5BLU) + 1$	Bajocco et al. 2019
Wetness (TCW)	0.1509 Blue + 0.1973 Green + 0.3279 Red + 0.3406 NIR − 0.7112 SWIR_1 − 0.4572 SWIR_2	Zhang et al., (2002)
Greeness (TCG)	−0.2848 Blue − 0.2435 Green − 0.5436 Red + 0.7243 NIR + 0.0840 SWIR_1 − 0.1800 SWIR_2	Zhang et al. (2002)

The notation used in this study follows that of Kreutzmann et al. (2019), who developed the R package emdi that implements the EBP. The same notation is used by Krennmair (2022), who developed the R package SAEforest, which implements the MERF algorithm. As demonstrated by Molina and Rao (2015), in the case of linear parameters (i.e. any parameter that can be expressed as a linear combination of unit-level responses), the EBP estimator is equal to the Empirical Best Linear Unbiased Predictor (EBLUP) estimator. Since Battese et al., (1988), Molina and Rao (2015), and most R packages that implement EBLUP rely on a simplified version in which unit-level auxiliary data are replaced by their domain means (area-level model), I opted for EBP implemented in the emdi package instead, to make full use of the available unit-level auxiliary information.

The finite population of size N (here the total number of pixel in the study area) is denoted by U and partitioned into D domains (the forest stand in this study) $U_1, U_2, \ldots U_D$ of sizes $N_1, N_2, \ldots N_D$, where $i = 1,\ldots, D$ refers to an i^{th} domain and $j = 1,\ldots, N_j$ to the j^{th} pixel in a domain. From this population, a random sample of size n is drawn. This leads to $n_1,\ldots, n_D$ observations in each domain. If $n_i = 0$, the i^{th} domain is not in the sample. The target variable is denoted y_{ij}.

The EBP model is the nested error linear regression (NER) model (Battese et al., 1988), which allows incorporation of both fixed and random effects to account for variability within and between domains. Additionally, it enables the use of unit-level (i.e. pixel-level) data, unlike area-level models (e.g. the Fay-Herriot model). The model is often used in finite-population models, as in this study (Sugasawa & Kubokawa, 2020). The NER model is expressed as:

$$T(y_{ij}) = x_{ij}{}^{T}\beta + u_i + e_{ij} \tag{1}$$

where $T(y_{ij})$ is the Box-Cox transformation of the dependent variable measured at the unit-level (pixel-level), x_{ij} is a p-vector of known covariates, with p the number of covariates, β is a p-vector of unknown fixed effects (the coefficient of regression), and u_i and e_{ij} are area-specific random effects and unit-level error terms, respectively. The underlying assumptions are that the random effect and the error term are mutually independent and distributed as $u_i \sim \mathrm{N}(0, \sigma_u{}^2)$ and $e_{ij} \sim \mathrm{N}(0, \sigma_e{}^2)$. It is worth noting that, since u_i is a domain-specific random effect and thus common to all the observations in the same domain, its variance $\sigma_u{}^2$ is often referred to as the "between" component of variance, in contrast to $\sigma_e{}^2$ the "within" component.

Since the model assumes normality and homoscedasticity of the random effects and error terms, a transformation can be applied to relax these assumptions. The Box-Cox transformation is among the most widely used in practice, as the logarithmic transformation is a special case. The distribution of GSV volume within a sampling field dataset is already known to be right-skewed, both at the local and national level (Chirici et al., 2020; Vangi et al., 2021), so the data were transformed using the Box-Cox transformation defined by:

$$T(y_{ij}) = \begin{cases} \frac{(y_{ij} + s)^{\lambda}}{\lambda}, & \lambda \neq 0 \\ \log(y_{ij} + s), & \lambda = 0 \end{cases} \tag{2}$$

where λ is an unknown transformation parameter; since the Box-Cox transformation needs positive values, s is a shift parameter that guarantees the non-negativity assumption. Both λ and s are estimated via the restricted Stransformation by inspecting Q-Q plots and using the Shapiro-Wilk (SW) and Kolmogorov-Smirnov (KS) tests (depending on the sample size). At the same time, homoscedasticity was evaluated by examining the skewness and excess kurtosis of residuals. Skewness and excess kurtosis values $< |2|$ are considered acceptable for homoscedasticity (George & Mallery, 2010; Gravetter & Wallnau, 2014).

In model (2), the coefficients β and the variance components $u_i + e_{ij}$ are estimated via the REML estimator. The EBP of small area parameters is computed as:

$$\theta^{EBP}{}_i = E\left[h(Y_i)|y_{is}; \hat{\psi}\right] \tag{3}$$

where $Y_i = (y_1,\ldots,y_n)$ are the unit values (both sampled and non-sampled), y_{js} are s^{th} observed sample values in the i^{th} domain, $\hat{\psi} = (\hat{\beta}, \hat{\sigma_u}^2, \hat{\sigma_e}^2))$ are model parameter estimates, $h(\cdot)$ is the functional form of the target parameter (linear or non-linear). In this study, the parameters used were the mean and the total, respectively defined as:

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$$

$$\tau_i = \sum_{j=1}^{N_i} y_{ij}$$

Monte Carlo approximation can be used to generate synthetic populations for the target variable, both for in-sample and out-of-sample domains, given x_{ij} and $\hat{\psi}$, as explained in Kreutzmann et al. (2019). For more detailed information about the algorithm, refer to Kreutzmann et al. (2019), Molina and Rao (2010), and Molina and Marhuenda (2015). The small-area parameter for the i^{th} domain is estimated by averaging the Monte Carlo replicates' parameter.

The MSE is estimated via a semi-parametric wild bootstrap procedure, described in detail by Kreutzmann et al., (2019, Appendix A) and Rojas-Perilla et al. (2020). Wild bootstrap not only takes into account the uncertainty derived from the transformation, but also protects against "'mild' departures from normality after using transformations" (Rojas-Perilla et al., 2020, *p* 123).

To express the estimator's uncertainty on a meaningful scale, the square root of the MSE was computed, yielding the RMSE. The RMSE was also reported relative to the total and mean GSV for each stand, expressed as a percentage (RMSE%).

The mixed-effect random forests estimator

In Equation (1), the term $x_{ij}{}^T\beta$ models the conditional indicator (i.e. the mean) of y_{ij} given $x_{ij,}$ but can be substituted with the general form $f(x_{ij})$. While f is a linear mixed model in EBP, in MERF, f is a random forest model. The original procedure for estimating $\hat{f}$, $\hat{u}$, $\hat{\sigma}^2{}_u$ and $\hat{\sigma}^2{}_e$ was developed by Hajjem et al. (2014) for random forests and called the Expectation–Maximization (EM) algorithm, which alternates between estimating the non-linear fixed effects and the random effects. The MERF algorithm, adjusted by Krennmair and Schmid (2022), fits Equation 4 by iteratively estimating: 1) the forest function, assuming the correctness of random effects; 2) the random effect part, assuming the correctness of the out-of-bag (OOB) prediction of the random forest. Given for any small area i:

$$y_i = f(X_i) + Z_i u_i + \varepsilon_i \tag{4}$$

where $f(X_i)$ represents any parametric or non-parametric function that expresses the conditional mean of y_i given X_i (random forest in the case of MERF), and $Z_i u_i$ is the linear part that captures the random effects. The iterative algorithm works as follows:

1) Set $b = 0$ and $\hat{u}_{(0)} = 0$
2) Set $b = b + 1$. Update $\hat{f}(X)_{(b)}$ and $\hat{u}_{(b)}$ as follow:
   a. $y_{(b)}{}^* = y - Z\hat{u}_{(b-1)}$
   b. Estimate $\hat{f}()_{(b)}$ with random forest using $y_{(b)}{}^*$ as the dependent variable and X as covariates
   c. Get the OOB prediction from $\hat{f}()_{(b)}$
   d. Fit a linear mixed model with no intercept for $\hat{f}(X)_{(b)}{}^{OOB}$: $y = \hat{f}(X)_{(b)}{}^{OOB} + Z\hat{u}_{(b)} + \varepsilon$
   e. Extract the variance component $\hat{\sigma}^2{}_{\varepsilon,(b)}$ and the variance-covariance matrix $\hat{H}_{(b)}$ of ε and estimate the random effects by: $\hat{u}_{(b)} = \hat{H}_{(b)} Z' \hat{U}_{(b)}{}^{-1}(y - \hat{f}(X)_{(b)}{}^{OOB})$
3) Repeat step 2) until convergence is reached. In this study, the convergence criterion is a marginal change in the generalized log-likelihood (GLL) of less than $1e^{-4,}$ as already used by Krennmair and Schmid (2022) and Ananda et al. (2024)

The estimation of the variance component $\hat{\sigma}^2_\varepsilon$ was naive and thus could not be considered a valid estimator for the variance of the unit-level error . Consequently, this estimator was corrected for bias to obtain the residual variance $\hat{\sigma}^2_\varepsilon$ from the random forest model through the bootstrap approach proposed by Mendez and Lohr (2011).

A non-parametric random-effect block (REB) bootstrap is used to estimate the area-level MSE, following the formulation proposed by Chambers and Chandra (2013). The REB bootstrap ensures that the variability of residuals within each area is captured. The iterative algorithm for fitting equation (4) and the bootstrap MSE estimator are described in detail in Krennmair and Schmid (2022).

MERF was shown to be robust to model misspecification under EBP, particularly in the presence of non-Gaussian error structure and complex interactions among covariates.

Note that, since RF does not assume normality and normally distributed residuals, Equation (4) was fitted without applying the Box-Cox transformation used in the EBP approach.

The SAEforest R package was used to fit the MERF algorithm, and the number of bootstrap iterations for MSE estimation was set to 200, while keeping the default values for the RF hyperparameters (ntree = 500 and mtry = 3).

Validation

Both the NER and the RF model were evaluated using a buffer leave-one-out cross-validation (BLOOCV) scheme with the R package blockCV (Valavi et al., 2018), which creates evenly spatially distributed training and test sets (Figure S2), considering a buffer of a specified distance (ideally, the range of spatial autocorrelation) around each observation. Each fold is generated by excluding observations within the specified distance of each testing point, ensuring robust validation in the presence of possible spatial autocorrelation. The buffer distance (the range over which observations are independent) was determined by constructing the empirical variogram, a fundamental geostatistical tool for measuring spatial autocorrelation. The empirical variogram models the structure of spatial autocorrelation by measuring variability between all possible pairs of points (O'Sullivan & Unwin, 2010).

Models were evaluated using the mean R^2, RMSE, and RMSE% across the BLOOCV iterations:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{5}$$

$$RMSE\% = \frac{RMSE}{\bar{y}_i}100 \tag{6}$$

where y_i, $\hat{y}_i$, $\bar{y}_i$ are the ith observation of the dependent variable, the ith prediction, and the mean of the observations, respectively; n represents the number of observations.

Results

Model performance

The final NER model yielded a marginal R^2 of 0.52 and a conditional R^2 of 0.78 (including the random effect) on the training set. After back-transforming the data to the original range, the BLOOCV yielded an R^2 of 0.37 and an RMSE of 202 m^3 ha^{-1} (RMSE% = 28.9). The optimal λ and shift parameters for the Box-Cox transformation, estimated via the REML approach, resulted in $\lambda = 0.45$ and shift = 0, respectively. Normality of residuals, both for fixed and random effects, was assessed using the SW and KS tests, and by inspecting the QQ-plot (Figure S1). All tests indicated normality ($p < 0.05$). Homoscedasticity was evaluated by calculating the skewness and excess kurtosis of fixed and random residuals. Since skewness was 0.01 and 0.2 and excess kurtosis was 2.65 and 2.4 for fixed and random residuals, respectively, homoscedasticity was confirmed.

The MERF model achieved an R^2 of 0.7 on the OOB observations. On the BLOOCV, the model achieved an R^2 of 0.67 and an RMSE of 151 m^3 ha^{-1}, corresponding to 21.7% of the mean observed GSV. Figure 2 reports the results of the BLOOCV for both models, comparing predicted versus observed values.

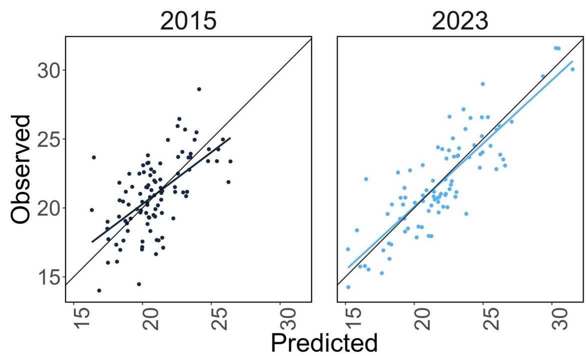

Figure 2. Scatterplots of predicted versus observed values (in m^3 ha^{-1}) for EBP (left panel) and MERF (right panel) resulting from the BLOOCV.

Small area estimation

EBP estimation of mean GSV at the stand level ranged between 297 and 1821 m^3 ha^{-1} with an average value of 767 m^3 ha^{-1}, while the mean GSV for MERF was between 350 and 1261 m^3 ha^{-1} with an average of 761 m^3 ha^{-1}. The total GSV ranged from 181 and 219 m^3 (minimum) to 8695 and 9995 m^3 (maximum) for EBP and MERF, respectively. The average total GSV was 1768 and 1783 for EBP and MERF, respectively.

The RMSE for the mean GSV varies from 54 to 1768 with a mean of 198 m^3 ha^{-1} (EBP) and from 51 to 197 with a mean of 155 m^3 ha^{-1} (MERF).

To evaluate the model reliability across individual management units, the empirical 95% confidence intervals for each domain prediction was constructed (i.e. 1.96*RMSE). The observed GSV mean of the plots in each sampled domains fell within these corresponding 95% prediction intervals for 66 out of 79 domains when utilizing the EBP framework, and for 77 out of 79 domains when utilizing the MERF framework (Figure 3).

Table 3 summarizes the results for the mean (in m^3 ha^{-1}) and total (in m^3) GSV, along with their respective RMSEs, for both EBP and MERF.

Figures 4 and 5 show the mean GSV value at the forest stand level (left panel) and its standard error (right panel) for EBP and MERF, respectively (Figures S3 and S4 in the Appendix show the total and their respective RMSE for EBP and MERF), while Figure 6 shows the distribution of mean and total GSV, along with their RMSE and RMSE%.

The RMSE of the mean GSV decreases with increasing plot count within the domain, but not linearly (Fig. S4), for both methods. In contrast, the RMSE of the total GSV varies with the domain dimension rather than with the number of sampling points. Overall, the RMSE of MERF for mean and total shows a narrower distribution around the mean than EBP.

A positive relationship between GSV and its RMSE is observed in mean and total estimates for both methods, but with significant differences between the two approaches (Figure 7). EBP shows a higher standard deviation and a steeper regression line. Furthermore, several outliers are observed for high GSV values (over ~1200 m^3 ha^{-1}), with RMSE% approachin 100%. In contrast, MERF shows a more compact point distribution and a much shallower slope. The RMSE remains relatively stable across the entire GSV range, with lower variability and fewer extreme outliers than observed for EBP.

Discussion

Direct estimates of mean and total GSV at the forest stand level are limited to sampled domains and, therefore, unavailable for all areas of interest. The use of model-based SAE methods that incorporate unit-level auxiliary information enables extending predictions to non-sampled domains while improving overall estimate precision compared to direct estimators or even to area-level mixed models, such as the Fay–Herriot model (Fay III & Herriot, 1979). Numerous forestry studies have demonstrated the advantage

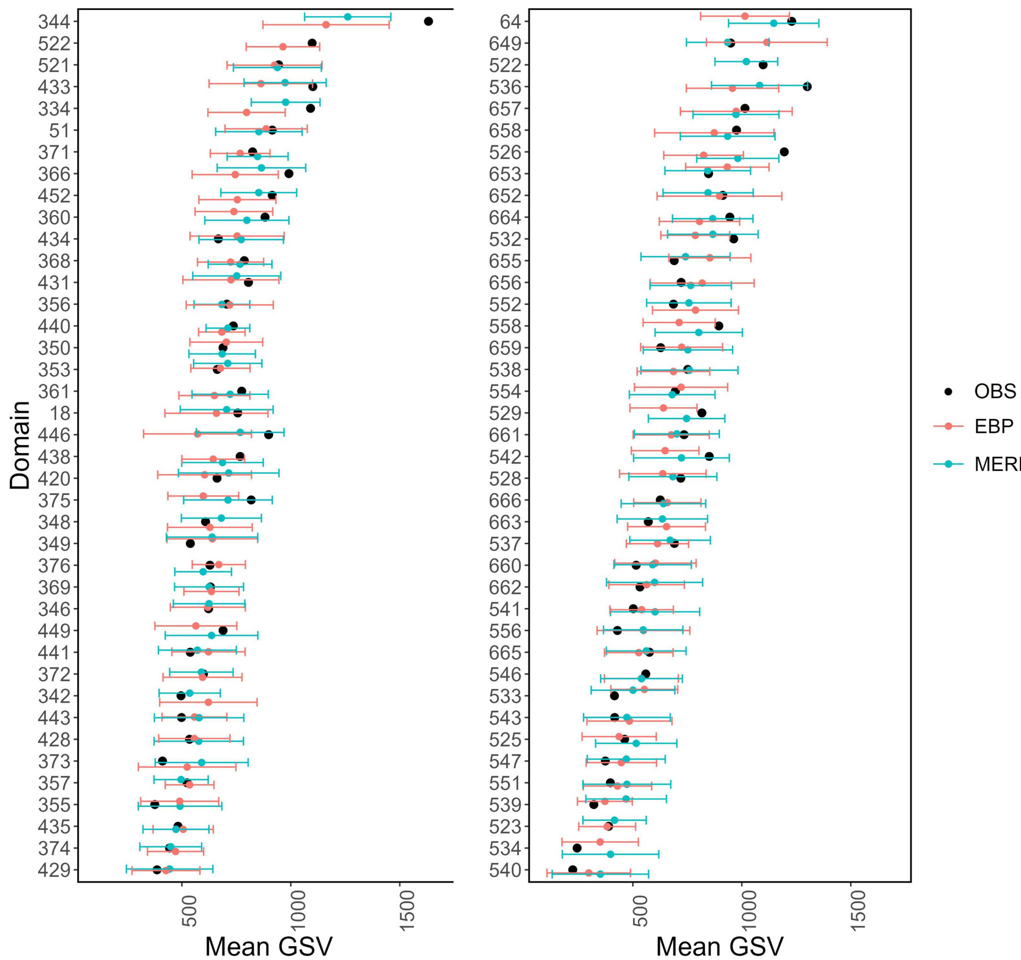

Figure 3. Estimate of mean GSV (in m^3 ha^{-1}) and its RMSE at the domain level obtained with EBP and MERF. The black dot represents the mean GSV from the observed sampling point.

Table 3. Results for mean and total GSV and their respective RMSE for EBP and MERF models.

		Mean GSV		Total GSV	
		SAE	**RMSE**	**SAE**	**RMSE**
EBP	min	297	54	1814	43
	mean	767	198	1768	470
	max	1821	1768	8695	4119
	sd	225	126	1237	463
MERF	min	351	51	219	48
	mean	761	155	1783	366
	max	1261	197	9995	2736
	sd	110	25	1257	275

of using indirect or composite estimators over direct, design-based estimators (e.g. Breidenbach et al., 2018; Ver Planck et al., 2018; Mauro et al., 2017; Frank, 2020; Georgakis et al., 2025; see also the review by Dettmann et al., 2022). The resulting maps reveal gradients and clusters of forest volume that would not be observable using sampled data alone, highlighting the practical value of SAE methods in forest inventory and management. Given the skewed distribution of GSV, model misspecification was a potential concern when using standard linear mixed models. For this reason, both the EBP approach with dependent variable transformation and the MERF model were considered appropriate alternatives for handling non-normal data, as confirmed by the results of the assumption tests.

The comparison between EBP and MERF highlights important differences in predictive performance, robustness, and suitability for SAE of GSV. While both approaches enable the extension of estimates to

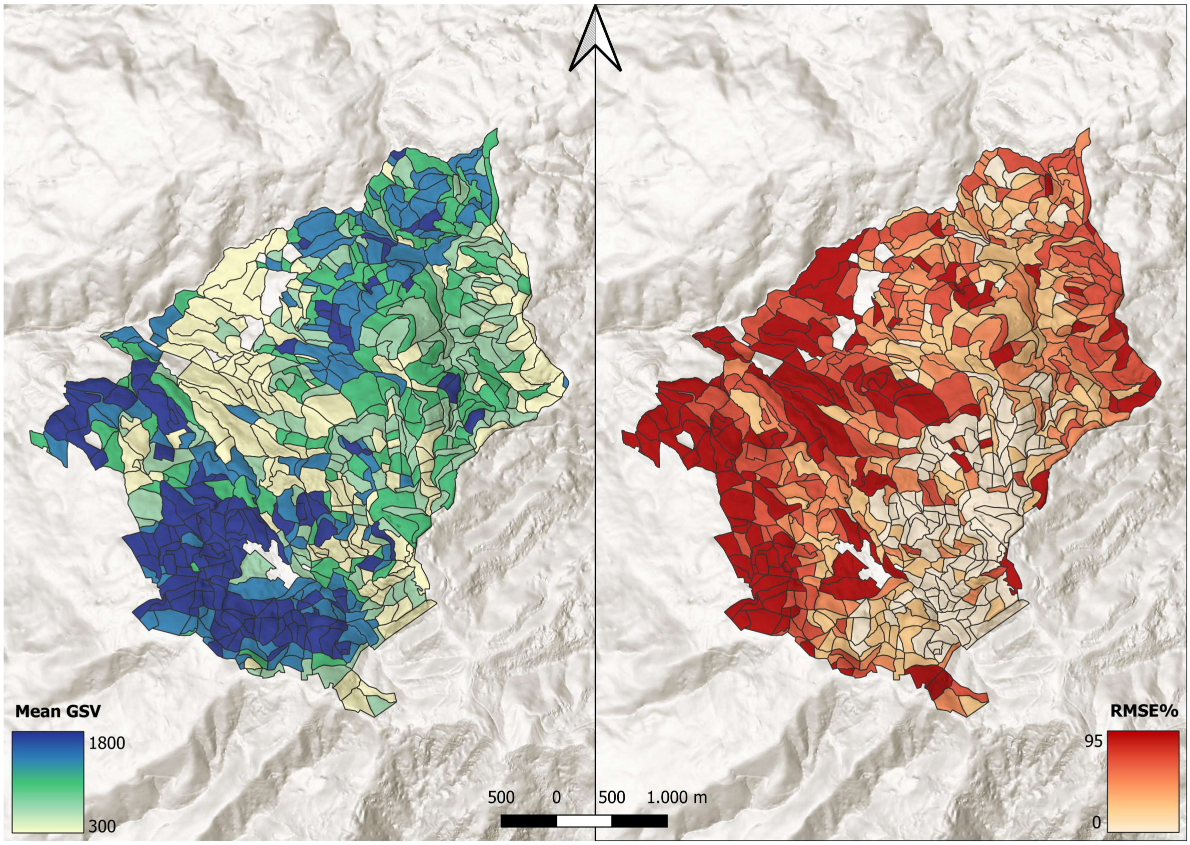

Figure 4. Map of mean GSV (left panel, in m^3 ha^{-1}) and its RMSE% (right panel) using the EBP estimator at the forest stand level.

non-sampled forest stands and provide comparable estimates of total GSV, their behavior differs substantially in terms of accuracy, variability, and stability across domains.

The NER model showed a moderate explanatory power on the training data, with a marginal R^2 of 0.52 and a conditional R^2 of 0.78, confirming the importance of random effects in capturing hierarchical variability. However, this improvement did not fully translate into predictive performance, as indicated by the lower BLOOCV R^2 (0.37) and relatively high RMSE (202 m^3 ha^{-1}; 28.9%). Although the Box–Cox transformation ensured compliance with model assumptions, these statistical properties did not yield strong generalization performance. This suggests that, despite being well-specified, the NER model is unable to capture complex, non-linear relationships under heterogeneous forest conditions(May et al., 2024). Moreover, multicollinearity in the training dataset may inflate the variance of the regression coefficients, thereby affecting unit-level predictions and the variance component of the MSE estimator (Molina & Rao, 2015).

In contrast, the MERF model demonstrated consistently greater predictive performance, achieving R^2 values of 0.70 (OOB) and 0.67 under BLOOCV, along with lower RMSE (151 m^3 ha^{-1}; 21.7%). These results indicate stronger generalization capability, likely due to MERF's ability to model non-linear relationships and interactions among predictors while accounting for hierarchical structure through random effects. Furthermore, MERF is more robust to multicollinearity due to its underlying RF model, which prevents variance inflation and ultimately yields more stable MSE estimates. Anand et al. (2024) estimated per capita expenditure in Indonesia, demonstrating that MERF effectively handles multicollinearity, especially in very small domains. This advantage becomes particularly relevant in forest ecosystems, where structural complexity and environmental variability often violate the assumptions of parametric models, and multicollinearity among remote sensing variables is common. Partial dependence plot in Appendix (Fig. S5) reveals the complex, non-monotonic, non-linear relations between GSV and almost all covariates, except for the CHM.

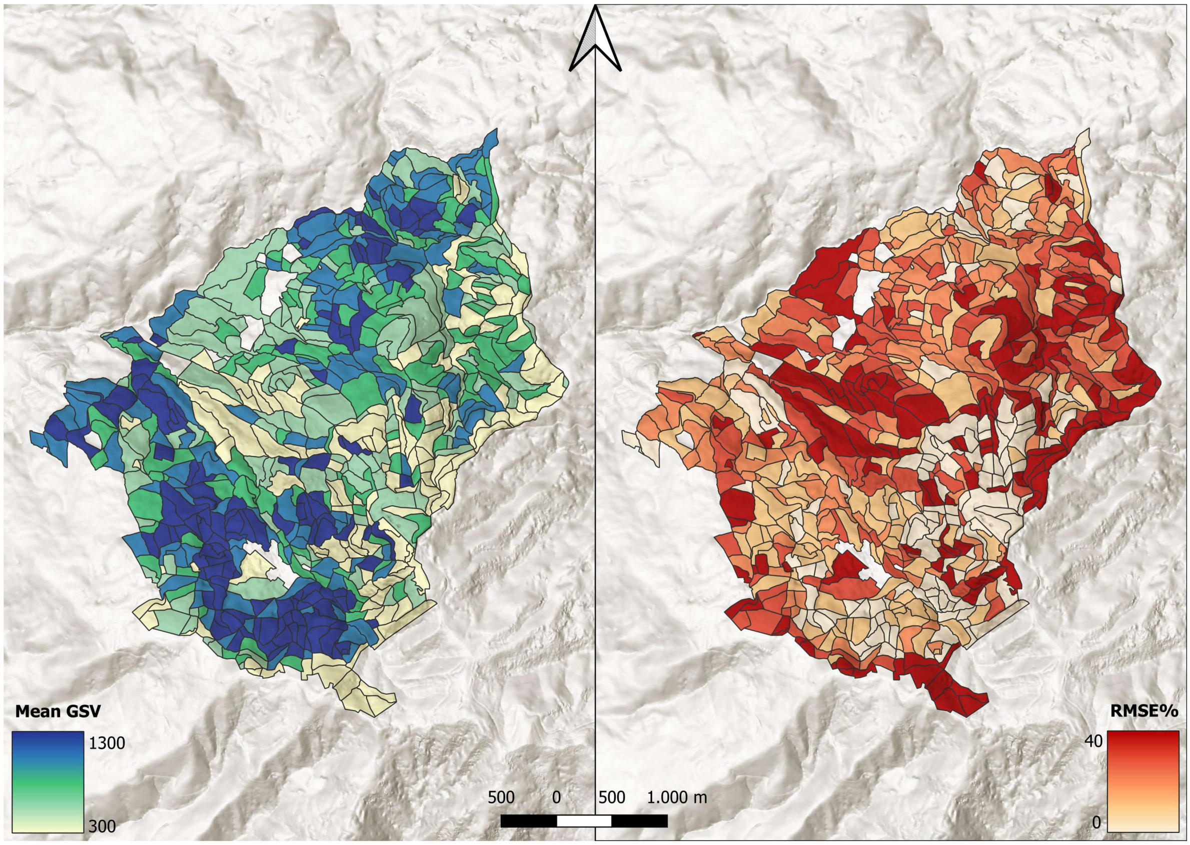

Figure 5. Map of mean GSV (left panel, in m^3 ha^{-1}) and its RMSE% (right panel) using the MERF estimator at the forest stand level.

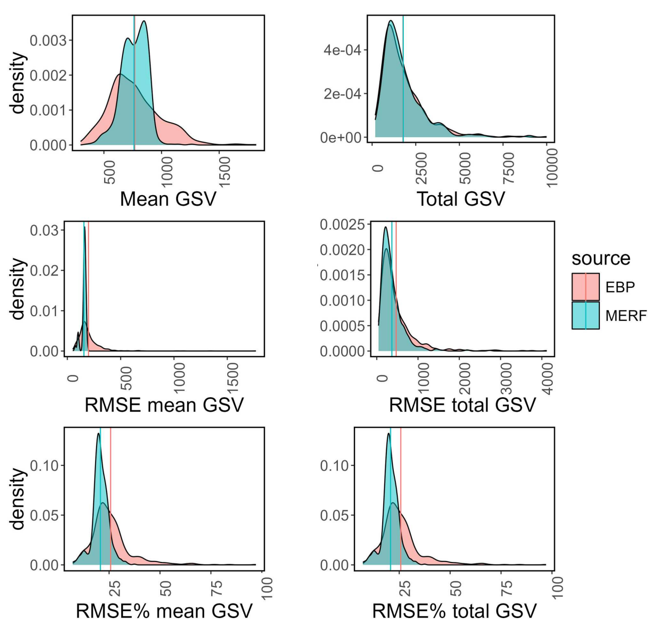

Figure 6. Distribution of mean (in m^3 ha^{-1}) and total (in m^3) SAE and their respective RMSE and RMSE% for EBP (red) and MERF (blue). The vertical lines represent the mean values.

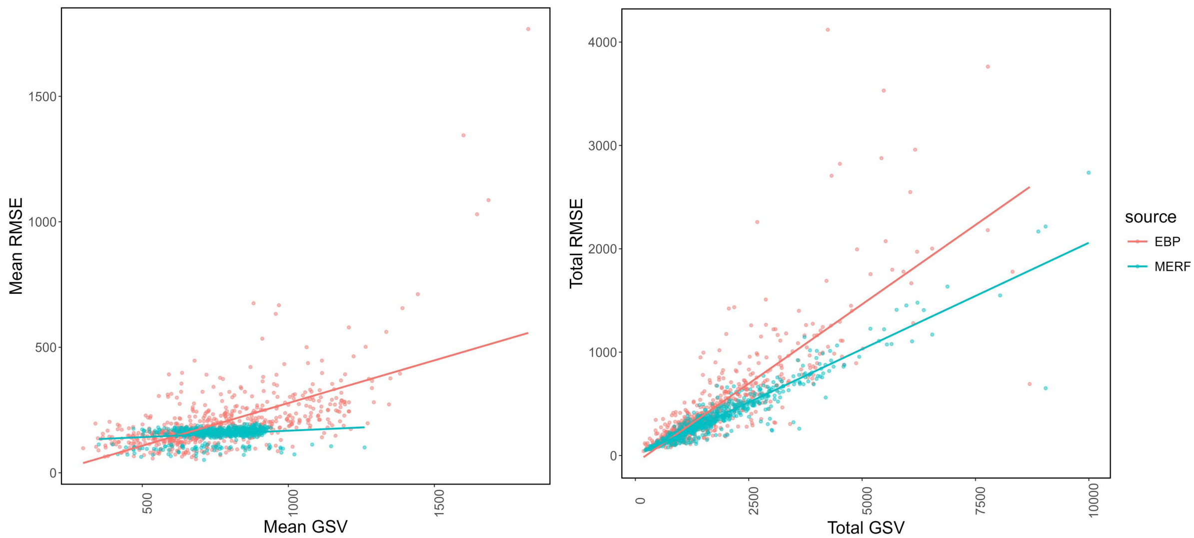

Figure 7. Scatterplot of mean (left panel, in m^3 ha^{-1}) and total (right panel, in m^3) GSV against their respective RMSE. Each dot represents a single domain (forest stand).

At the small-area level, both methods produced slightly overestimated overall mean GSV estimates (767 m^3 ha^{-1} for EBP and 761 m^3 ha^{-1} for MERF against an observed mean of 707 m^3 ha^{-1}), but with different distributions of the domain's means. In particular, the RF predictions tend to cluster around the observed mean, a phenomenon known as shrinkage towards the mean. This behavior is consistent with the properties of random forest, which introduce implicit regularization through variance reduction via ensemble averaging (Breiman, 2001; Hastie et al., 2009). Also, Krennmair and Schmid (2022) observed the same effect when estimating total household per capita income in Mexico, using the MERF model. The same phenomenon had already been observed in biomass and carbon mapping, confirming that the shrinkage effect is not an emergent feature of this study. For example, Zhang et al. (2024) found that regularized RF and quantile RF significantly reduce overestimation and underestimation compared with vanilla RF. Mascaro et al. (2014) found that adding spatial context to RF improved tropical forest carbon mapping. The reduction in prediction dispersion lowers the overall RMSE, but at the cost of reduced ability to accurately represent extreme values. The shrinkage effect implies that the model underestimates high GSV values and overestimates low ones, as suggested by the weak correlation between RMSE and GSV. This is not a MERF-specific feature, despite being amplified by the RF implicit regularization; in Kubokawa (2009) and Sugasawa's and Kubokawa's (2020) reviews on SAE, the authors point out that among the desirable properties of the EBLUP are the shrinkage function and the pooling effects, which arise from the setup of random effects and common parameters in LMM, indicating the intrinsic shrinkage behavior of the SAE framework.

The same effect is not observed in the total GSV estimate, which instead shows strong agreement with the EBP method. This apparent discrepancy can be explained by the scale-dependent behavior of the bias–variance trade-off. At the unit level, MERF reduces variance through ensemble averaging, resulting in smoother predictions and a tendency to shrink extreme values toward the mean. When unit-level predictions are summed to obtain area-level totals, individual over- and under-estimations tend to compensate, and the variance reduction achieved by shrinkage dominates over the bias (Molina and Rao, 2015), leading to more consistent total estimates. Future studies should focus on mitigating the shrinkage effect of RF by exploring, e.g. post-hoc bias correction, hyperparameterization, and estimating the conditional distribution rather than the mean, e.g. with quantile random forest regression.

More pronounced differences emerge when considering uncertainty measures. The RMSE associated with MERF predictions is consistently lower and exhibits a narrower distribution (mean RMSE = 155 m^3 ha^{-1}) compared to EBP (198 m^3 ha^{-1}), indicating greater precision and stability. Additionally, the coverage of observed values within the SAE confidence intervals is higher for MERF (77 out of 79 domains) than for EBP (66 domains), suggesting improved reliability in uncertainty quantification. However, it is worth noting that the size of small areas of interest plays a fundamental role in determining the precision of estimates within the SAE framework. Although SAE techniques are designed to produce reliable estimates at increasingly fine spatial resolutions, the number of sample plots available within each area (n_d) remains a critical factor. Specifically, the sampling variance of direct estimators is inversely proportional to n_d, implying that small sample sizes lead to high variability and unstable estimates. SAE methods mitigate this limitation by leveraging auxiliary information; however, this introduces a trade-off between variance reduction and potential model-induced bias. (Molina & Rao, 2015; Georgakis et al., 2025). In particular, in unit-level methods such as NER models under EBP and RF under MERF, the MSE comprises sampling variance, model variance, and parameter uncertainty. When n_d is small, estimates rely more heavily on model predictions, leading to stronger shrinkage and MSE dominated by model assumptions. As n_d increases, the influence of direct observations grows, reducing shrinkage and lowering MSE. Our study was characterized by low coverage of the sampling domain (79 out of 658) and low sampling intensity within domains, with a mean of 1 sampling point per domain and a maximum of 4, suggesting a strong dependence of the estimate MSE on the model prediction. In turn, this dependency confirms the goodness-of-fit of both models, given the relatively low RMSE obtained over the domains. Rahlf et al. (2014) obtained similar performance (RMSE% = 18.2) in estimating timber volume within small forest stands (1-3 ha) using the NER model, but with a higher sampling intensity of 5–7 plots per domain.

The behavior of RMSE in relation to domain characteristics provides further insight. For both methods, the RMSE of mean GSV decreases with increasing sample size, although not linearly, reflecting diminishing returns from increased sampling intensity. In contrast, the RMSE of total GSV is more strongly influenced by domain size than by the number of sampling points, emphasizing the role of spatial extent in error propagation. A positive relationship between total GSV and RMSE is observed for both approaches; however, this relationship is not observed for the mean estimate in the MERF approach, while it still exists for EBP, which shows higher variability and several extreme outliers, particularly for high GSV values (above ~1200 m^3 ha^{-1}), where RMSE% approaches 100%. This difference in behavior between EBP and MERF can be attributed to their underlying model structures: In the NER model, the back transformation inherently links the variance of predictions to the magnitude of the estimated variable, leading to a proportional increase in RMSE with both the mean and total estimates (Molina & Rao, 2010). In contrast, MERF does not impose a parametric relationship between the mean and variance, as prediction uncertainty is driven by the residual structure of the random forest and random effects. Consequently, the RMSE of the mean does not necessarily increase with the magnitude of the estimate, although the RMSE of totals still scales with the population size (domain's area) due to aggregation. It should be noted, however, that EBP has a mathematically proven estimator of MSE, firmly grounded in statistical theory. At the same time, the non-parametric bootstrap scheme of MERF remains an active area of statistical research regarding its theoretical consistency in small domains (Krennmair & Schmid, 2022).

From a modeling perspective, a key distinction lies in how predictors are treated. The EBP approach requires explicit model specification and variable selection, whereas MERF benefits from the implicit variable selection of random forests, enabling it to identify relevant predictors, capture complex interactions, and reduce preprocessing steps. While this improves predictive performance, it also introduces a trade-off in interpretability (Krennmair and Schmid, 2022). Nonetheless, interpretability is often less important than estimate precision in the SAE framework. Interpretability can be partially recovered through diagnostic tools such as variable importance and partial dependence analysis, as well as explainable AI methods such as SHapley Additive exPlanations (SHAP) (Strumbelj et al., 2010). However, it must be noted that the computational time required to fit the MERF model was more than six times that of the EBP, especially due to the complex non-parametric bootstrap scheme for RMSE estimation and the bootstrap-based adjustment of the residual variance in RF. For large-scale applications, computational intensity can become an issue, which could be partially mitigated by parallelizing the bootstrap scheme.

Overall, the results confirm that model-based SAE methods substantially enhance the estimation of forest attributes at fine spatial scales. While the EBP framework remains theoretically robust and interpretable, its predictive performance is limited in complex and highly variable forest environments. The MERF model, by combining the strengths of machine learning and mixed-effects modeling, provides a more accurate, stable, and reliable alternative for GSV estimation. These findings support the adoption of machine learning-based approaches in forest inventory, particularly for handling non-linear relationships, heterogeneous conditions, and low-intensity sampling coverage.

Limitation and future direction

Forest variables are inherently spatial. While SAE models include area-level random effects, traditional RF and NER models do not explicitly model continuous spatial autocorrelation. To mitigate spatial autocorrelation, we included the coordinates and altitude of the sampling points as predictors and evaluated the model using a spatial block cross-validation approach. However, future studies should focus on integrating explicit spatial models into SAE, for example, comparing the spatial empirical best linear unbiased predictor (SEBLUP, Petrucci & Salvati, 2004a, 2004b) or the spatial empirical Bayes predictor (SEBP, Handayani et al., 2018) and including Geographically Weighted RF or Spatial RF into MERF. Emick et al. (2023) proposed a Geostatistical model-based estimator to account for spatial autocorrelation and to correct for bias, based on summaries of Monte Carlo Markov Chain samples from Bayesian posterior distributions. Another possibility to relax the NER model's assumptions, particularly in the presence of multicollinearity or outliers, is to consider robust estimation techniques, such as the estimators developed by Chambers et al. (2014) and further extended by Schmid et al. (2016) to account for spatial correlation. On the other hand, Ananda et al. (2024) have already used Principal Component Analysis and rotation forests in MERF to mitigate multicollinearity.

Regarding the indicators, while the mean is the gold standard in almost all SAE research, totals and other indicators are often understudied, especially indirect indicators (indicators that are not in a linear relationship with the unit-level predictions), such as the Gini index or quantiles, which may be interesting parameters in biodiversity conservation studies.

Finally, an inherent limitation, and, at the same time, a notable feature of the study, was the large proportion of unsampled domains and the small sampling size within domains ($n_d \leq 4$), which limited the accuracy of the estimation in terms of RMSE, but highlighted the necessity of such approaches to deal with real-world datasets, often with similar charateristics in the field of local forest management.

Conclusion

Small-area estimation has become an increasingly important research focus in forest inventory, primarily because it enables model-based inference in small domains where limited sample data prevent reliable estimates based solely on design-based methods, and secondly because of its inherent ability to provide an assessment of the estimate via standard error or MSE estimation. While the EBP approach remains theoretically sound and interpretable, its performance is constrained by complex, non-linear relationships and multicollinearity in the training dataset. In contrast, the MERF model offers advantages in model specification, affecting its predictive accuracy, robustness, and stability, providing more reliable uncertainty estimates, albeit at the cost of increased computational demand and reduced interpretability. Both methods exhibit inherent shrinkage effects that influence domain-level estimates but tend to balance out at aggregate levels, resulting in consistent total GSV estimates. Ultimately, the findings highlight a trade-off between theoretical rigor and predictive performance, supporting the integration of machine-learning-based approaches into forest inventory practices and local management plans, especially in contexts characterized by low sampling intensity and high environmental heterogeneity, thereby contributing to the ongoing development of more robust SAE methods.

Author contributions

CRediT: **Elia Vangi:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Disclosure statement

The author declares that he has no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding

E.V. acknowledges NextGenCarbon H2020 project funded by the European Commission, number 101184989 call HORIZON-CL5-2024-D1-01-07 and the Space It Up! project funded by the Italian Space Agency, ASI, and the Ministry of University and Research, MUR, under contract n. 2024-5-E.0 - CUP n. I53D24000060005.

Data availability statement

Sentinel-2 data are freely available at the original source. All other data and codes will be shared upon request to the author.

References

Ananda, R., Notodiputro, K. A., & Aidi, M. N. (2024). Modified mixed effects random forest in small area estimation using PCA and rotation forest with correlated auxiliary variables. *Scientific Journal of Informatics*, *11*(3), 705–720. https://doi.org/10.15294/sji.v11i3.10633

Bajocco, S., Ferrara, C., Alivernini, A., Bascietto, M., & Ricotta, C. (2019). Remotely-sensed phenology of Italian forests: Going beyond the species. *International Journal of Applied Earth Observation and Geoinformation*, *74*, 314–321. https://doi.org/10.1016/j.jag.2018.10.003

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-components model for prediction of county crop areas suing survey and satellite data. *Journal of the American Statistical Association*, *83*(401), 28–36. https://doi.org/10.1080/01621459.1988.10478561

Breidenbach, J., & Astrup, R. (2012). Small area estimation of forest attributes in the Norwegian national forest inventory. *European Journal of Forest Research*, *131*(4), 1255–1267. https://doi.org/10.1007/s10342-012-0596-7

Breidenbach, J., Magnussen, S., Rahlf, J., & Astrup, R. (2018). Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data. *Remote Sensing of Environment*, *212*, 199–211. https://doi.org/10.1016/j.rse.2018.04.028

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chambers, R., & Chandra, H. (2013). A random effect block bootstrap for clustered data. *Journal of Computational and Graphical Statistics*, *22*(2), 452–470. https://doi.org/10.1080/10618600.2012.681216

Chambers, R., Chandra, H., Salvati, N., & Tzavidis, N. (2014). Outliner robust small area estimation. *Journal of the Royal Statistical Society: Series B*, *76*, 47–69. https://doi.org/10.1111/rssb.12019

Chirici, G., Giannetti, F., McRoberts, R. E., 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. International Journal of Applied Earth Observation and Geoinformation, *84*. 101959. https://doi.org/10.1016/J.JAG.2019.101959

Ciancio, O., Riserva Naturale Statale Biogenetica di Vallombrosa. (2009). Piano di gestione e silvomuseo 2006–2025, *Corpo Forestale dello Stato, Ufficio Territoriale per la Biodiversità di Vallombrosa* (pp. 113–134). Florence, Italy: Reggello (FI).

D'Amico, G., Francini, S., Giannetti, F., Vangi, E., Travaglini, D., Chianucci, F., Mattioli, W., Grotti, M., Puletti, N., Corona, P., & Chirici, G. (2021). A deep learning approach for automatic mapping of poplar plantations using S2 imagery. *Giscience & Remote Sensing*, *58*(8), 1352–1368. https://doi.org/10.1080/15481603.2021.1988427

D'Amico, G., Chirici, G., Corona, P., Romano, R., Di Domenico, G., Giannetti, F., & Mattioli, W. (2023). Differenze locali e prospettive globali per le foreste italiane: La definizione di bosco nel prossimo sistema informativo forestale nazionale. *L'Italia forestale e montana*, *78*(1), 15–29. https://doi.org/10.36253/ifm-1094

Dettmann, G. T., Radtke, P. J., Coulston, J. W., Green, P. C., Wilson, B. T., & Moisen, G. G. (2022). Review and synthesis of estimation strategies to meet small area needs in forest inventory. *Frontiers in Forests and Global Change*, *5*. 813569. https://doi.org/10.3389/ffgc.2022.813569

Emick, E., Babcock, C., White, G. W., Hudak, A. T., Domke, G. M., & Finley, A. O. (2023). An approach to estimating forest biomass while quantifying estimate uncertainty and correcting bias in machine learning maps. *Remote Sensing of Environment*, *295*. 113678. https://doi.org/10.1016/j.rse.2023.113678

Fay Iii, R. E., & Herriot, R. A. (1979). Estimates of income for small places: An application of James-stein procedures to census data. *Journal of the American Statistical Association*, *74*(366a), 269–277. https://doi.org/10.1080/01621459.1979.10482505

Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from landsat TM data and their transferability between regions. *Remote Sensing of Environment*, *85*, 463–474. https://doi.org/10.1016/S0034-4257(03)00039-7

Frank, B. M. (2020). *Aerial laser scanning for forest inventories: estimation and uncertainty at multiple scales* (Vol. 2020). PhD diss Oregon State University.

Galidaki, G., Zianis, D., Gitas, I., Radoglou, K., Karathanassi, V., Tsakiri–Strati, M., Mallinis, G., & Woodhouse, I. (2017). Vegetation biomass estimation with remote sensing: Focus on forest and other wooded land over the

Mediterranean ecosystem. *International Journal of Remote Sensing*, *38*(7), 1940–1966. https://doi.org/10.1080/01431161.2016.1266113

Georgakis, A., Papageorgiou, V. E., & Stamatellos, G. (2025). A new approach to small area estimation: Improving forest management unit estimates with advanced preprocessing in a multivariate Fay–Herriot model. *Forestry: An International Journal of Forest Research*, *98*(4), 605–622. https://doi.org/10.1093/forestry/cpae061

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference* (Vol. *17*, 10a ed.). Boston: Pearson.

Giannetti, F., Chirici, G., Gobakken, T., Naesset, E., Travaglini, D., & Puliti, S. (2018). A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. *Remote Sensing of Environment*, *213*, 195–205. https://doi.org/10.1016/j.rse.2018.05.016

Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2013). Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information. *Forest Science*, *59*(5), 536–548. https://doi.org/10.5849/forsci.12-073

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environmen*t, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gravetter, F., & Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth.

Hajjem, A., Bellavance, F., & Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, *84*(6), 1313–1328. https://doi.org/10.1080/00949655.2012.741599

Handayani, D., Folmer, H., Kurnia, A., & Notodiputro, K. A. (2018). The spatial empirical Bayes predictor of the small area mean for a lognormal variable of interest and spatially correlated random effects. *Empirical Economics*, *55*, 147–167. https://doi.org/10.1007/s00181-018-1452-5

Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome (2009). The Elements of Statistical Learning, *Springer Series in Statistics*. 10.1007/978-0-387-84858-7

Hobza, T., & Morales, D. (2016). Empirical best prediction under unit-level logit mixed models. *Journal of official statistics*, *32*(3), 661–692. https://doi.org/10.1515/jos-2016-0034

Jönsson, P., Cai, Z., Melaas, E., Friedl, M. A., & Eklundh, L. (2018). A method for robust estimation of vegetation seasonality from landsat and Sentinel-2 time series data. *Remote Sensing*, *10*(4), 635. https://doi.org/10.3390/rs10040635

Kennedy, R. E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W. B., & Healey, S. (2018). Implementation of the landtrendr algorithm on google earth engine. *Remote Sensing*, *10*, 1–10. https://doi.org/10.3390/rs10050691

Krennmair, P. (2022). Tree-based. *Machine Learning in Small Area Estimation*, *10*, 36.

Krennmair, P., & Schmid, T. (2022). Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *12600*(0035-9254), 1467–9876. http://arxiv.org/abs/2201.10933

Krennmair, P., W¨urz, N., & Schmid, T. (2022). Analysing Opportunity Cost of Care Work using Mixed Effects Random Forests under Aggregated Census Data Apr. 22, 2022. arXiv: 2204.10736[stat]. http://arxiv.org/abs/2204.10736

Kreutzmann, A. K., Pannier, S., Rojas-Perilla, N., Schmid, T., Templ, M., & Tzavidis, N. (2019). emdi: Estimating and Mapping Disaggregated Indicators. R package version 1.1.6, https://CRAN.R-project.org/package=emdi

Kubokawa, T. (2009). A review of linear mixed models and small area estimation. *Journal of Statistical Research*, *44*(1), 31 .

Magnussen, S., & Breidenbach, J. (2017). Model-dependent forest stand-level inference with and without estimates of stand-effects. *Forestry: An International Journal of Forest Research*, *90*(5), 675–685. https://doi.org/10.1093/forestry/cpx023

Mascaro, J., Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Martin, R. E., Anderson, C., Chadwick, K. D., Higgins, M., & Bond-Lamberty, B. (2014). A tale of two "forests": Random forest machine learning aids tropical forest carbon mapping. *PLoS One*, *9*(1), e85993. https://doi.org/10.1371/journal.pone.0085993

May, Paul B., Schlund, Michael, Armston, John, Kotowska, Martyna M., Brambach, Fabian, Wenzel, Arne, & Erasmi, Stefan. (2024). Mapping aboveground biomass in Indonesian lowland forests using GEDI and hierarchical models. *Remote Sensing of Environment*, *313*, 114384. https://doi.org/10.1016/j.rse.2024.114384

Mauro, F., Monleon, V. J., Temesgen, H., & Ford, K. R. (2017). Analysis of area level and unit level models for small area estimation in forest inventories assisted with LiDAR auxiliary information. *PLoS One*, *12*(12), e0189401. https://doi.org/10.1371/journal.pone.0189401

McRoberts, R., Tomppo, E., Schadauer, K., Vidal, C., Stahl, G., Chirici, G., Lanz, A., Cienciala, E., Winter, S., & Smith, B. (2009). Harmonizing national forest inventories. *Journal of Forestry*, *107*, 179–187. https://doi.org/10.1093/jof/107.4.179

Mendez, G., & Lohr, S. (2011). Estimating residual variance in random forest regression. *Computational statistics & data analysis*, *vol. 55*(11), 2937–2950. https://doi.org/10.1016/j.csda.2011.04.022

Mesquida, B. G., Pascual, A., Rodriguez-Puerta, F., Guerra-Hernández, J., Perroy, R. L., García-Gómez, R., & Mauro, F. (2026). Assessment of spaceborne and airborne lidar metrics using fay-herriot models to support forest biomass estimation. *Forest Ecology and Management*, *601*. 123369. https://doi.org/10.1016/j.foreco.2025.123369

Molina, I., & Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, *38*(3), 369–385. https://doi.org/10.1002/cjs.10051

Molina, I., & Marhuenda, Y. (2015). Sae: An R package for small area estimation. *The R Journal*, *7*(1), 81–98. https://doi.org/10.32614/rj-2015-007

O'Sullivan, D., & Unwin, D. J. (2010). Geographic Information Analysis (2nd ed.). John Wiley & Sons.

Parisi, F., Vangi, E., Francini, S., D'Amico, G., Chirici, G., Marchetti, M., Lombardi, F., Travaglini, D., Ravera, S., De Santis, E., & Tognetti, R. (2023). S2 time series analysis for monitoring multi-taxon biodiversity in mountain beech forests. *Front. Forests and Global Change*, *6*, 1020477. https://doi.org/10.3389/ffgc.2023.1020477

Petrucci, A., & Salvati, N. (2004a). Small area estimation using spatial information, the rathbun lake watershed case study. Working paper no 2004/02G.Parent Department of Statistics Department of Statistics viale Morgagni, *59*, 50134. University of Florence

Petrucci, A., & Salvati, N. (2004b). Small area estimation considering spatially correlated errors: The unit level random effects model. Working paper no 2004/10. In G. Parenti (Ed.), *Department of Statistics*. University of Florence.

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., & Astrup, R. (2014). Comparison of four types of 3D data for timber volume estimation. *Remote Sensing of Environment*, *155*, 325–333. https://doi.org/10.1016/j.rse.2014.08.036

Molina, I., & Rao, J. N. K. (2015). Small area estimation, *Wiley Series in Survey Methodology* (2nd ed.). New Jersey: Wiley.

Rojas-Perilla, N., Pannier, S., Schmid, T., & Tzavidis, N. (2020). Data-driven transformations in small area estimation. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *183*(1), 121–148. https://doi.org/10.1111/rssa.12488

Roy, D. P., Boschetti, L., & Trigg, S. N. (2006). Remote sensing of fire severity: Assessing the performance of the normalized burn ratio. *IEEE Geoscience and remote sensing letters*, *3*(1), 112–116. https://doi.org/10.1109/LGRS.2005.858485

Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., Nordkvist, K., & Hyyppä, J. (2015). Model-assisted estimation of growing stock volume using different combinations of LiDAR and landsat data as auxiliary information. *Remote Sensing of Environment*, *158*, 431–440. https://doi.org/10.1016/j.rse.2014.11.020

Saarela, S., Wästlund, A., Holmström, E., Mensah, A. A., Holm, S., Nilsson, M., Ståhl, G., & Fridman, J. (2020). Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. *Forest Ecosystems*, *7*(1), 43. https://doi.org/10.1186/s40663-020-00245-0

Schmid, T., Tzavidis, N., Münnich, R., & Chambers, R. (2016). Outlier robust small-area estimation under spatial correlation. *Scandinavian Journal of Statistics*, *43*, 806–826. https://doi.org/10.1111/sjos.12205

Shi, Y., Ke, G., Chen, Z., Zheng, S., & Liu, T.-Y. (2022). Quantized training of gradient boosting decision trees. In I. S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in Neural Information Processing Systems* (Vol. *35*, pp. 18822–18833). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/aper/2022/file/77911ed9e6e864ca1a3d165b2c3cb258-Paper-Conference.pdf.

Strumbelj, E., & Kononenko, I. (2010). Journal of Machine Learning Research, *11*, 1–18.

Sugasawa, S., & Kubokawa, T. (2020). Small area estimation with mixed models: A review. *Japanese Journal of Statistics and Data Science*, *3*, 693–720. https://doi.org/10.1007/s42081-020-00076-x

Tabacchi, G., Di Cosmo, L., Gasparini, P., & Morelli, S. (2011). Stima del volume e della fitomassa delle principali specie forestali italiane. Equazioni di previsione, tavole del volume e tavole della fitomassa arborea epigea (Vol. Trento: Consiglio per la Ricerca e la sperimentazione in Agricoltura, Unità di Ricerca per il Monitoraggio e la Pianificazione Forestale.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., & Katila, M. (2008). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, *112*, 1982–1999. https://doi.org/10.1016/j.rse.2007.03.032

Valavi, Roozbeh, Elith, Jane, Lahoz-Monfort, José J., & Guillera-Arroita, Gurutzeta. (2018). Block CV : An r package for generating spatially or environmentally separated folds for *k* -fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*, 225–232. https://doi.org/10.1111/2041-210x.13107

Vangi, E., D'Amico, G., Francini, S., Giannetti, F., Lasserre, B., Marchetti, M., McRoberts, R. E., & Chirici, G. (2021). The effect of forest mask quality in the wall-to-wall estimation of growing stock volume. *Remote Sensing*, *13*, 1038. https://doi.org/10.3390/rs13051038

Ver Planck, N. R., Finley, A. O., JrKershaw, J. A., Weiskittel, A. R., & Kress, M. C. (2018). Hierarchical Bayesian models for small area estimation of forest variables using LiDAR. *Remote Sensing of Environment*, *204*, 287–295. https://doi.org/10.1016/j.rse.2017.10.024

White, J. C., Wulder, M. A., Hobart, G. W., Luther, J. E., Hermosilla, T., Griffiths, P., Coops, N. C., Hall, R. J., Hostert, P., Dyk, A., & Guindon, L. (2014). Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Canadian Journal of Remote Sensing*, *40*, 192–212. https://doi.org/10.1080/07038992.2014.945827

White, J. C., Tompalski, P., Bater, C. W., Wulder, M. A., Fortin, M., Hennigar, C..., & White, R. (2025). Enhanced forest inventories in Canada: Implementation, status, and research needs. *Canadian Journal of Forest Research*, *55*, 1–37.

Zhang, X., Schaaf, C. B., Friedl, M. A., Strahler, A. H., Gao, F., & Hodges, J. C. (2002). MODIS tasseled cap transformation and its utility, *IEEE International Geoscience and Remote Sensing Symposium* (Vol. 2, pp. 1063–1065). IEEE.

Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., & Ou, G. (2024). Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery. *Ecological Indicators*, *159*, 111752.