A new approach to small area estimation: improving forest management unit estimates with advanced preprocessing in a multivariate Fay–Herriot model

Aristeidis Georgakis [1],*, Vasileios E. Papageorgiou[2], Georgios Stamatellos[1]

[1]School of Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[2]Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
*Corresponding author. School of Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece. E-mail: arisgeorg@for.auth.gr

Abstract

Forest inventories are essential for informing sustainable forest management decisions, and small area estimation (SAE) techniques aim to enhance the precision of these inventories, particularly when sample sizes are limited. This study presents a novel approach to SAE by integrating trivariate empirical best linear unbiased prediction Fay–Herriot (FH) models with advanced preprocessing techniques. By employing multivariate Fay–Herriot (MFH) models, the methodology utilizes clustering analysis, variable selection, and outlier treatment to improve the precision of estimates for small areas. A comparative analysis with traditional univariate Fay–Herriot (UFH) models demonstrates that MFH outperforms UFH in estimating key forest attributes such as forest growing stock volume, basal area, and Lorey's mean tree Height, even in areas with limited sample sizes. The use of auxiliary variables derived from remote sensing data and past censuses proved critical, with remote sensing playing a dual role: aiding in clustering forest management units into larger small areas of interest and serving as covariates in the FH models. The results highlight the effectiveness of MFH1 (assuming independent and identically distributed random effects), which consistently produced estimates with <5% coefficient of variation, indicating high precision. Across all response variables, MFH1 led to reductions in standard errors compared to UFH, with median percentage gains in precision of 17.22% for volume, 13.91% for basal area, and 3.95% for mean height. Mean precision gains were even higher, at 18.27%, 16.51%, and 10.87%, respectively. This study advances SAE methodologies by providing a robust framework for accurately estimating critical forest attributes in challenging scenarios, including geolocation errors, limited sample sizes, and the smallest applicable small areas for area-level models. It highlights the contribution of the correlation between multiple response variables to improving the precision of estimates. The proposed methodology has significant implications for enhancing the accuracy of forest inventories and supporting informed forest management decisions.

Keywords: linear mixed models; empirical best linear unbiased predictors; remote sensing data; census data; clustering analysis; variable selection

Introduction

Small area estimation (SAE) methodologies have become indispensable tools across various scientific disciplines, facilitating informed decision-making by providing precise estimates for small geographic regions or domains. From mapping poverty in economic studies and official statistics to estimating crop yields in agriculture, SAE techniques have demonstrated their versatility and effectiveness (Pratesi 2016; Rao and Molina 2015; Young and Chen 2022). However, one of the most significant applications of SAE is found within the field of forestry, where the accurate assessment of biometric attributes is crucial for sustainable forest management (Georgakis 2019; Guldin 2021; Dettmann et al. 2022).

Forest inventories (FIs) are crucial for collecting and estimating biometric variables such as growing stock volume, above-ground tree biomass, and tree density, which serve as the foundation for effective forest management. However, relying solely on direct estimates from traditional sampling surveys may result in imprecise estimates due to limited sample size and focus on population means rather than small areas or domains (Georgakis and Stamatellos 2020). Moreover, direct estimators are unable to provide estimates for areas without samples. To overcome these challenges, modern approaches rely on advanced statistical modelling techniques like SAE, which incorporate field and auxiliary data such as census or remote sensing data.

The two primary types of FIs are national forest inventories (NFI) and forest management inventories (FMI). NFIs focus on estimating forest attributes at national and regional scales, supporting policy decisions and fulfilling conservation reporting obligations. In contrast, FMIs aim to provide detailed, spatially explicit estimates for smaller areas, such as forest stands or compartments, which are essential for operational planning and sustainable forest management. However, accurate estimation for small areas like forest management units (FMUs) remains a challenge due to limited field data. The term "small area" refers more to the limited sample size rather than the geographic scale. Recent advancements integrate remote sensing with statistical models

to improve the efficiency and precision of these small area estimates, offering new techniques to address this ongoing difficulty.

Traditional design-based approaches rely on probability sampling to ensure unbiased estimates. However, when sample sizes are too small or non-existent, such as in remote areas, design-based methods struggle to provide reliable direct estimates (McRoberts 2010). In such cases, model-assisted or model-based inferences become valuable alternatives. Design-based methods, including direct and model-assisted estimators, are robust at a national or regional scale, where sufficient data are available to guarantee design-unbiased results. However, model-based estimates prove more suited to smaller scales, such as forest stands and pixel-level analyses (Kangas et al. 2019).

When sampling data are sufficient, model-assisted estimators can be used for small areas like forest stands (Magnussen and Breidenbach 2017). These estimators improve precision through auxiliary data but do not inherently "borrow strength" across areas as model-based methods do. They lack random effects to account for variability across small areas, making them less reliable when dealing with limited sample sizes. A key example of a model-assisted estimator is the generalized regression estimator (GREG) (Särndal et al. 1992), which becomes less reliable when sample sizes are very small because it primarily enhances precision rather than addressing the lack of sample size.

Model-based inference is especially useful when sample sizes are too small for reliable design-based inference, as it is not dependent on probability samples (McRoberts et al. 2014). As the scale becomes smaller, like in FMIs, model-based approaches become increasingly important (Kangas et al. 2019). These estimators can address the lack of precision in areas with small sample sizes but may also introduce biases and computational challenges in variance estimation (McRoberts et al. 2013).

The area-based approach (ABA) is commonly used in the remote sensing community at the unit level of sample plots, regardless of whether the forest inventory is NFI or FMI (Fassnacht et al. 2024). The ABA method uses a synthetic estimator that effectively borrows strength across areas and incorporates auxiliary data such as remote sensing information, which enhances the estimation of forest attributes for FMIs. However, ABA does not consider random area effects between small areas, unlike SAE linear mixed-effects models, which consider both synthetic estimates and random area effects to account for unobserved heterogeneity and random variation among small areas.

A composite estimator, which is a weighted sum or mean of a direct estimator and an indirect synthetic estimator for small areas of interest, can combine the strengths of both direct and indirect estimation (Goerndt et al. 2011). The two main types of composite SAE models that consider synthetic estimation via regression estimators and incorporate variability across small areas through random area effects are the Fay–Herriot (FH) (area-level model) and the Battese-Harter-Fuller (unit-level model). These SAE linear mixed models provide more reliable and realistic small area estimates than ABA approaches because they do not rely as heavily on assumptions of strong correlations and consider the variability of small areas through random area effects. Moreover, these models outperform model-assisted methods like GREG and traditional design-based approaches.

The selection between unit-level and area-level models depends on the availability of auxiliary variables (Rao and Molina 2015; Dettmann et al. 2022). Unit-level or nested-error regression models utilize variables available at the individual sample unit level (Battese et al. 1988), which in the context of FIs correspond to field sample plots representing portions of forested land. On the other hand, area-level models establish a relationship between direct averages or totals of domains and aggregated, area-specific auxiliary variables within the specific small area of interest (Breidenbach and Astrup 2012; Goerndt et al. 2013; Georgakis 2021).

The area-level or FH model, introduced in 1979, is an extension of linear mixed-effects models primarily used for univariate Fay–Herriot (UFH) estimates (Fay and Herriot 1979). FH models are among the most used model-based methods in SAE literature (Rao and Molina 2015). Their operational effectiveness is facilitated by developed frameworks, particularly in R packages, and they have been implemented with various extensions, including back-transformations, spatial and robust extensions, adjusted variance estimation methods, and accounting for measurement errors (Harmening, et al. 2023). FH models have demonstrated notable effectiveness in forestry, particularly in cases where the geographic coordinates of sample plot centers are unavailable, or when positioning errors lead to weak correlations (Magnussen et al. 2017; Breidenbach et al. 2018; Mauro et al. 2017). Moreover, many FMIs rely on Variable Radius Plot sampling methods, which necessitate the use of area-level models (Temesgen et al. 2021).

The multivariate Fay–Herriot (MFH) model extends the UFH model by incorporating the correlation among the response variables. Generally, MFH models exhibit smaller mean square errors (MSEs) compared to univariate models, as indicated by previous studies (Nurizza and Ubaidillah 2019; Desiyanti et al. 2023), and also demonstrate narrower confidence intervals (CIs) (Benavent and Morales 2016; Saegusa et al. 2022). An advantage of MFH models is their ability to eliminate the need for covariates. The key factor contributing to large variance reduction is the correlation between response variables (Datta et al. 2002; Franco and Bell 2022). Furthermore, for non-sampled areas, MFH models show significant improvements compared to FH estimates (Guha and Chandra 2022) by utilizing synthetic-type estimators (Chandra et al. 2011). Moreover, a new empirical best predictor (EBP) has been proposed for estimating unsampled domains (Burgard et al. 2022).

Even in the absence of covariates, MFH models can achieve substantial variance reductions when strong relationships exist between the target variables (Franco and Bell 2022). Similarly, if one response variable in the MFH model lacks strong covariates, it can still benefit from the interrelationship of the response variables. A clear advantage of MFH is its ability to utilize repeated surveys or FIs by leveraging linearly related response variables, thereby enhancing estimates for the latest data (Ngaruye et al. 2017; Franco and Maitra 2023, Georgakis et al. 2024). Most of the above literature used empirical best linear unbiased predictors (EBLUPs) for predicting model parameters in the context of MFH models.

In forestry, research efforts in SAE have focused on the use of univariate approaches, whether they are area-level or unit-level. However, there is a significant research gap regarding multivariate SAE of forest biometric parameters. In the research of FIs, two notable MFH applications have been identified. One involves the application of the bivariate Fay–Herriot model, and the corresponding comparisons in pairs of variables, for estimates of growing stock volume, basal area, stem density, mean height, and mean diameter, at the forest stand level (Frank 2020 Ch. 4, Manuscript III). The second research effort concerns the estimates of four forest parameters: mean biomass, mean basal area, mean volume, and mean tree density at the forest stand level (Ver Planck 2018 Ch. 3). However, the application was based

on the Bayesian approach and used hierarchical Bayesian (HB) models, instead of the classical frequentist approach with EBLUP estimators, as attempted in this study. Generally, apart from these studies, MFH methods are under development in the field of forestry.

In this article, we propose a comprehensive approach to forest inventory estimation by introducing MFH models and advanced preprocessing stages. While traditional UFH models offer valuable estimates, they are limited in their ability to account for correlations among multiple response variables. To address this limitation, we develop the MFH-EBLUP model, which extends the UFH framework to incorporate inter-variable correlations. We explore three variations of the MFH model based on the structure of the variance–covariance matrix of random effects and evaluate the effectiveness of two of them (Benavent and Morales 2016). The first variation (MFH1) assumes independent and identically distributed random effects, while the variance–covariance matrix remains diagonal as in the UFH model. The second variation of the model (MFH2) assumes autocorrelation of random effects and is considered autoregressive (AR(1)). The third variation of the multivariate model (MFH3) assumes both heteroskedasticity and autocorrelation in the random effects (heteroskedastic autoregressive, HAR(1)). Leveraging these interrelationships improves predictive capabilities and precision in estimating forest inventory attributes such as volume, basal area, and Lorey's mean tree height that were explored.

The primary objective of this study is to introduce a novel statistical methodology using variations of the MFH model to generate high-precision small area estimates for FMIs. The study addresses significant challenges related to limited sample sizes within FMUs, geolocation errors, and the complexity of multi-layered, uneven-aged forest ecosystems. To overcome these challenges, we propose a state-of-the-art preprocessing pipeline that includes clustering analysis to optimize correlations between response and auxiliary variables, and additional steps such as outlier treatment, variance smoothing, and variable selection. Clustering, used to define small areas, is crucial for datasets with limited sample sizes, such as in this study, but may be unnecessary for larger NFIs or FMIs with ample data.

We conduct an extensive comparison of the MFH model with traditional UFH models and direct estimates. This demonstrates the superior reliability of the MFH approach, which provides more precise estimates for multiple forest inventory attributes while reducing MSE, and the coefficient of variation (CV), and improving CIs. Finally, we discuss the challenges and constraints faced in the study, the optionality of clustering for larger datasets or simpler forest types, and the potential benefits of advanced remote sensing data like aerial laser scanning (ALS). This article highlights the advancement of the MFH approach in forestry, particularly for SAE under data limitations.

Material and methods
Study area and data

This study was conducted in the University Forest of Pertouli, located in Central Greece (Fig. 1). The forest ecosystem is characterized by uneven-aged forest, primarily dominated by the hybrid fir species, *Abies borisii-regis* Mattf. The forested area spans 2260 hectares, divided into 174 FMUs. After excluding unmanaged FMUs and those lacking adequate auxiliary data, 160 FMUs were included in the analysis. Field data collection in 2018 comprised 239 sample plots (units), each measuring 0.1 hectares, with sampling intensity, approximately 1%, based on systematic sampling.

Approximately half of the FMUs contain a single sample plot, while the remainder have two plots. Most FMUs contained 1–2 sample plots, with only a few consisting of three plots or none at all (UFAMF 2018). In the clustering analysis (detailed below), was used to define 'new small areas' by aggregating FMUs, enabling the application of FH models.

Auxiliary variables

Effective model-based SAE relies heavily on the availability of strong covariates. In this study, remote sensing and historical census data were used to enhance small area estimates. Auxiliary variables included domain-level height (h) percentiles (e.g. h50, h75) in meters, height shape metrics (kurtosis and skewness) and census-based data on tree density and volume (e.g. ForestDensity97, DenFir88, ForestGSV97) (Table 1). The remote sensing metrics were derived from a 2-meter canopy height model (CHM) created by subtracting a generated LiDAR-based digital terrain model (DTM) from a digital surface model (DSM) data obtained from the WorldView-2 satellite stereo imagery (CHM = DSM − DTM) (Georgakis et al. 2023). To improve model fitting and robustness, all auxiliary variables were transformed into logarithmic and squared format. These transformations provided alternative representations of the data, which improved the model fitting by meeting the assumptions of normal residuals and random area effects, while also avoiding computational errors.

Geolocation accuracy, limited samples, and their impact on model selection

The study initially aimed to employ a multivariate unit-level (MUL) model, which was expected to provide highly reliable small area estimates. However, weak correlations between response and auxiliary variables raised concerns about potential errors, with geolocation inaccuracies identified as a significant factor. The sample plots were initially geolocated using handheld GPS devices, where positional accuracy was not a priority for forest practitioners. During validation of 39 new sample plots, substantial positional errors were detected, significantly affecting Pearson's correlation coefficient (ρ) between key variables. Specifically, the correlation between the mean vegetation height and the woody volume was only $\rho = 0.27$, indicating that the geolocation inaccuracies of the original GPS data hindered reliable model predictions.

To verify these errors, 39 new sample plots were geolocated using post-processed GNSS observations with the network real-time kinematic technique, which provided much higher positional accuracy. This improvement resulted in much stronger correlation of $\rho = 0.67$ (see Γεωργάκης 2024, Ch. 6, Fig. 6.2, p. 91), reinforcing the importance of accurate plot positioning for the success of unit-level models and ABA approaches. Given the validation of geolocation inaccuracies and the weak correlations, the study shifted from attempting to apply the unit-level methodology to utilizing a multivariate area-level approach. This transition was necessary due to the real-world challenges faced with the existing FMI dataset.

However, the application of this methodology in this dataset was still not straightforward. Issues of weak correlations from limited sample sizes—typically one or two plots per FMU—and inadequate estimates of uncertainty at the FMU-level further complicated the problem. To overcome these challenges, we conducted a clustering analysis, as detailed in the first step of the preprocessing pipeline.
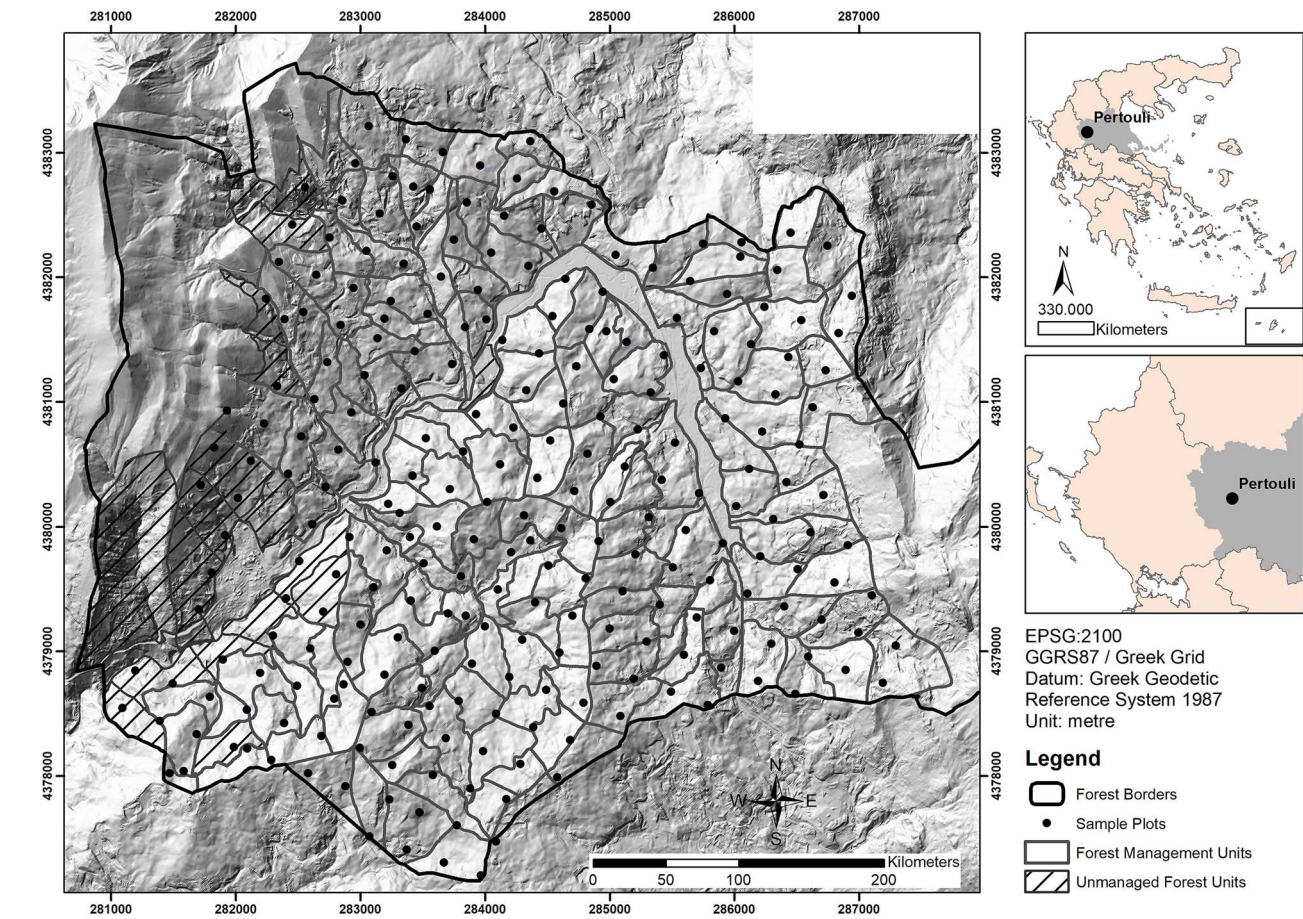
Figure 1. The Pertouli Forest study area and sampling design.

Table 1. Height and census auxiliary variables.

Data type	Data group	Descriptive statistics	Variable abbreviations	Unit metric	Year
Height	A	Quantiles	h25; h50; h75; h90	Meters (m)	2022
Height	A	Central tendency	hmean; hmode	m	2022
Height	B	Dispersion metrics	hsd; hcv	m; dimensionless	2022
Height	B	Shape distribution; L-moments	hLskew, L4	Dimensionless	2022
Census	B	Central tendency	ForestDensity97	Trees/hectare (ha)	1997
Census	B	Central tendency	ForestGSV97	m^3/ha	1997
Census	B	Central tendency	DenFir88	Trees/ha	1988

Advanced preprocessing pipeline

In this study, we introduce an advanced methodology for the pre-processing steps in SAE aimed at improving model performance and reliability. Our approach consists of several key processes designed to enhance the accuracy and robustness of the MFH model. These processes include: (i) Clustering FMUs when field data are sparse to enhance correlations between predictor and response variables, and define "new small areas", which could be avoided in richer datasets like NFIs, (ii) Smoothing sampling variances to stabilize estimates and reduce variability across small areas, (iii) Detecting and addressing outliers within MFH models to ensure robustness and estimating unsampled (outlier) domains by incorporating cluster information to account for random effects and provide more reliable estimates, and (iv) A proposed variable selection method to avoid multicollinearity, ensuring model parsimony and interpretability. Each step is described in detail below.

Clustering forest management units when field data are sparse

In this study, clustering analysis was deemed necessary as a prerequisite for applying FH models. Due to the limited sample size of 1–3 sample units per FMU and the complexity of multilayered forest. Initially, weak correlations between the target variables and auxiliary covariates at the FMU level hindered the use of direct modelling at this spatial level. By clustering similar FMUs into aggregated units, we were able to form "new small areas" that allowed for successful application of area-level modelling. This clustering process ensured the necessary assumptions—correlations between the target variables and auxiliary covariates—were met, providing an optimal solution for the current FMI data.

The objective was to group similar small areas—specifically FMUs—into clusters that would establish strong correlations between direct estimates derived from the sample units and

aggregated auxiliary covariates. For this purpose, the Euclidean distance metric and the Single-Link Hierarchical Clustering algorithm were applied, with the aggregated domain mean height serving as the clustering variable. The goal was to determine the optimal number of clusters, using the Friedman and Rubin index (Friedman and Rubin 1967) as a guide. An index value of 6240.10 was recorded. To explore a wide range of possibilities, the study considered cluster numbers from 8 to 50 and ultimately determined that 32 clusters (or "new small areas") provided the best solution.

The objective was to group similar initial small areas—specifically FMUs—into clusters that would establish strong correlations between direct estimates derived from the sample units and aggregated auxiliary covariates. The Euclidean distance metric and the Single-Link Hierarchical Clustering algorithm were applied with the aggregated domain mean height served as the clustering variable. The goal was to determine the optimal number of clusters ($n_{clusters}$), for SAE purposes, using the Friedman and Rubin index (Friedman and Rubin 1967) as a guide, with a recorded a value of 6240.1. To explore a wide range of possibilities, the study considered $n_{clusters}$ ranging from 8 to 50, ultimately determined that 32 clusters provided the best solution.

Among these clusters, eight consisted of only one plot each and were treated as outliers. The remaining 24 clusters had an average of 9.6 sample plots per cluster. Additionally, after outlier detection (as described in outlier detection and treatment), five more clusters—each containing only two or three sample units—were identified as outliers. These outlier clusters were treated as non-sampled domains, requiring a different estimation methodology in the analysis. Finally, the FH models applied in 19 small areas after clustering, averaging 100.0 hectares and 11.6 sample plots from 7.6 FMUs. The remaining 13 outliers considered as unsampled domains had an additional preprocessing step for estimation.

To evaluate the quality of clustering solutions, we used metrics such as the mean relative standard errors (RSEs) of direct small area estimates for volume means and 90th percentile of RSEs. Clustering results were carefully examined to ensure meaningful cluster formations and adequate variation in predictor variables within each cluster. The mean RSEs of small area volume means were found to be 13.44%, with the 90th percentile of RSEs at 24.31%. This step enhanced both the correlations between predictors and response variables for model-based estimates and improved the accuracy of direct small area estimates.

Although this preprocessing step is optional for FH models, it was mandatory for our study due to the limited sample size. When sample sizes are larger, such as those found in NFIs, this step may be unnecessary as correlations between response and predictor variables are generally strong across small areas. However, in this case, clustering FMUs to optimize correlations between response and auxiliary variables was essential for the successful application of FH models. This methodology is part of a comprehensive study that explores various algorithms, distance metrics, clustering variables, and methods for determining the optimal number of clusters using parameter initializations and indices for SAE purposes (Georgakis et al. 2023).

Smoothing sampling variances

One basic but also unrealistic assumption in real-world FH applications is the knowledge of sampling variance. Attempting to utilize sampling variance and covariances based solely on a few samples per small area often leads to instability, computational issues, zero random area effects, and decreased performance of the

multivariate estimator. To address these challenges, smoothing the sampling variances under the generalized variance function (GVF) can be employed (Wolter 2007), particularly for domains with small sample sizes. In forestry applications, a specific type of smoothing is commonly implemented, considering a weighted mean variance based on the size of small areas (hectares, ha) relative to the total forest population size (Goerndt et al. 2011). Similar smoothing applications have been validated empirically, demonstrating the effectiveness of the variance-smoothing procedure in improving the stability and performance of area-level FH models (Magnussen et al. 2017; Ver Planck et al. 2018; Coulston et al. 2021).

In this study, we implemented variance smoothing to enhance the stability and reliability of the estimators, extending to both the variables of interest and their covariances. The smoothed sampling variances $\widetilde{\sigma}_{\varepsilon d}^2$ were computed using a formula based on the constant mean of variance from the population V_ε, divided by the number of units n_d, within each domain d

$$\widetilde{\sigma}_{\varepsilon d}^2 = \frac{V_\varepsilon}{n_d}, \tag{1}$$

where

$$V_\varepsilon = \frac{\sum_{d=1}^{D} A_d \hat{\sigma}_{\varepsilon d}^2}{\sum_{d=1}^{D} A_d}. \tag{2}$$

Quantity A_d denotes the total area within the domain d and D is the number of domains (small areas). The term $\hat{\sigma}_{\varepsilon d}^2$ calculated directly from the sample plots, is utilized to estimate the original sampling variances (σ_d^2), with $\widetilde{\sigma}_d^2$ replacing these values. The formula $\hat{\sigma}_{\varepsilon d}^2$ for the unbiased sample variance under simple random sampling (SRS) without the finite population correction factor, due to small sampling intensity, is given by Equation (3)

$$\hat{\sigma}_{\varepsilon d}^2 = \sum_{j=1}^{n_d} \left(y_{dj} - \bar{y}_d \right)^2 / (n_d - 1), \tag{3}$$

where y_{dj} are the observed average values of the units j (sample plots) and $\bar{y}_d$ their grand mean within each domain. From now on we note that y_{dj} represent average values of each characteristic (i.e. growing stock volume, basal area, mean height) per unit. The same interpretation remains for the rest of the methodology and results section where we describe the structure of the univariate and MFH models.

Additionally, covariances were computed by taking the square root of the smoothed variances and multiplying them by the correlation coefficient of the variables (Georgakis et al. 2024). Namely

$$\widetilde{\sigma}_{\varepsilon d \mathbf{kl}} = \rho \cdot \sqrt{\widetilde{\sigma}_{\varepsilon d \mathbf{k}}^2 \cdot \widetilde{\sigma}_{\varepsilon d l}^2} = \rho \cdot \widetilde{\sigma}_{\varepsilon d k} \cdot \widetilde{\sigma}_{\varepsilon d l}, \tag{4}$$

for $l, k = 1, \ldots, K$, $l \neq k$, where K represents the number of examined forest characteristics. Also, ρ denotes the Pearson correlation of the two examined forest characteristics. Specifically, the $\widetilde{\sigma}_{\varepsilon d k}$ represents the smoothed variance of one of the examined response variables. This is based on SRS, which gives equal inclusion probabilities to all sampling units within a specific area. When different probabilities are given to each group based on the area, we are led to a weighted sampling variance. Weighted variances are usually smaller, but the smoothed variances exhibit greater homogeneity. Details regarding the estimation of the weighted variance are provided in (Särndal et al. 1992). To illustrate the substantial variability of

the direct sampling variances and the necessity for smoothing, we provide a comparison of the direct variances with the smoothed variance estimates in Figure S.1 of the Supplementary Material.

Outlier detection and treatment

Before proceeding with the MFH model, we conducted essential data curation, which involved correlation analysis and outlier detection. From the clustering analysis it was observed that few small clusters or single FMUs—FMUs with only two sample plots—were more prone to outlier detection, primarily because of their extreme direct estimates, variances, and failure to meet model assumptions. The outliers were limited to these small clusters with isolated FMUs (observations) that were difficult to merge into larger, more homogeneous groups. Through outlier detection, we identified three distinct types of outliers, each with negative impacts on model performance:

i) Extreme response variable outliers: These outliers were characterized by extreme values in the response variables, which weakened the Pearson correlation between the response and auxiliary variables. They predominantly occurred in clusters (referred to as "small areas") with only two sample plots, leading to extreme direct estimates that negatively influenced the correlation of response variables with area-level covariates.

ii) Convergence and computational issue outliers: Outliers in this category caused convergence or computational difficulties due to extreme variances and covariances. This often led to zero model variances or issues in estimating random area effects. Similar to the first category, these outliers were mainly found in domains with only two sample plots, where extreme (smoothed) variances and covariances were observed.

iii) Outliers not meeting model assumptions: These outliers were characterized by non-normal distributions of the residuals or the random area effects, violating model assumptions. Again, these outliers were typically present in small areas with only two sample plots.

Addressing these outliers was crucial for ensuring robust modelling that adhered to the necessary assumptions and provided reliable small area estimates. Their presence could lead to biased estimates, reduced correlation strength, and computational challenges, underscoring the importance of proper outlier treatment in the modelling process.

Estimating unsampled domains with cluster information

In this study, we are clustering one more time the clusters to provide more reliable estimates for unsampled and outlier small areas. When addressing unsampled domains or outliers, traditional FH is not applicable, resulting in synthetic estimation. In such cases, predictions based solely on auxiliary information and regression models, ignore the within-stand (small area) variation due to the absence of direct sample-based estimates. Among the 32 initial domains, we identified 13 outliers with minimal sample sizes: eight domains with one sample unit, four domains with two units, and one domain with three units (refer to Figure S.2 of the Supplementary Material). To address this, we applied a modified FH model that integrates cluster information to estimate areas with zero sample size or very small sample sizes. This method allowed us to incorporate random effects into

non-sampled domains, assuming that these domains share similar variability with the larger clusters or domains.

This modified approach entails adding the average random area effect to the synthetic estimation of non-sampled areas within the same cluster, enabling the estimation of MSE in unsampled areas (Desiyanti et al. 2023; Torkashvand et al. 2017; Haris and Ubaidillah 2020). To execute this, we utilize cluster analysis employing Partitioning Around Medoids (PAM) (Saligkaras and Papageorgiou 2023; Saligkaras and Papageorgiou 2022), with "hmean" serving as the clustering variable. Through analysis of the silhouette graph (Rousseeuw 1987), we identified four large domains/clusters, ensuring that small areas were grouped into clusters with adequate sample sizes (Figure S.3 of Supplementary Material). EBLUPs are then computed for non-sampled areas based on the MFH model. This second clustering step is independent of the earlier "*Clustering Forest Management Units when Field Data are Sparse*" and is optional. It could be applied in other studies involving small areas, such as forest stands with limited or no sample plots, where the FH model cannot be directly applied. Additionally, the large domains from this second clustering are used for calibration diagnostics.

Variable selection in the absence of multicollinearity

Variable selection is a crucial aspect of SAE aimed at improving interpretability and model performance. Our study proposes a systematic approach to variable selection, prioritizing the utilization of original variables derived from Principal Component Analysis (PCA) due to their interpretive clarity. Initially, we organize the variables into two distinct groups: Group A, consisting of height metrics, and Group B, comprising height and census variables. Group A displays very strong inter-variable correlations, characterized by coefficients exceeding $|r| \geq 0.9$, while Group B demonstrates comparatively weaker correlations. This classification is based on consistent multicollinearity observed among variables in group A and visual examination of the first three principal components (PCs) in PCA.

To mitigate multicollinearity within Group A, we adopt a sequential selection strategy. We begin by selecting one variable from Group A and incorporating all remaining variables from Group B in a stepwise EBLUP-FH selection process. This process entails evaluating the Akaïke information criterion (AIC) to guide model selection (Marhuenda et al. 2014). Subsequently, we include the next variable from Group A and repeat the stepwise selection procedure, continuing until all variables from Group A have been considered, and corresponding AIC values have been generated. In Fig. 2, the left panel illustrates the variable contribution in the first three PCs, with the group of variables A exhibiting close concentration towards the end of PC1 and demonstrating the largest loadings. The right panel of Fig. 2 presents highly correlated variables highlighted in red within the correlation heatmap.

The stepwise EBLUP-FH is implemented through the "saeBest" package in R (Ubaidillah and Aziz 2021). Comparative analysis between stepwise selection for simple regression models and the EBLUP-FH method, accounting for random area effects, demonstrates the latter's superior performance, as evidenced by lower MFH model variance. To ensure the robustness of the final model, we conducted a multicollinearity assessment, focusing on variance inflation factor (VIF) values. A threshold of VIF < 9.5 is applied to address multicollinearity concerns, thereby enhancing the stability and reliability of the model. After the selection of the variables, we validated that VIF was smaller than 5.5.
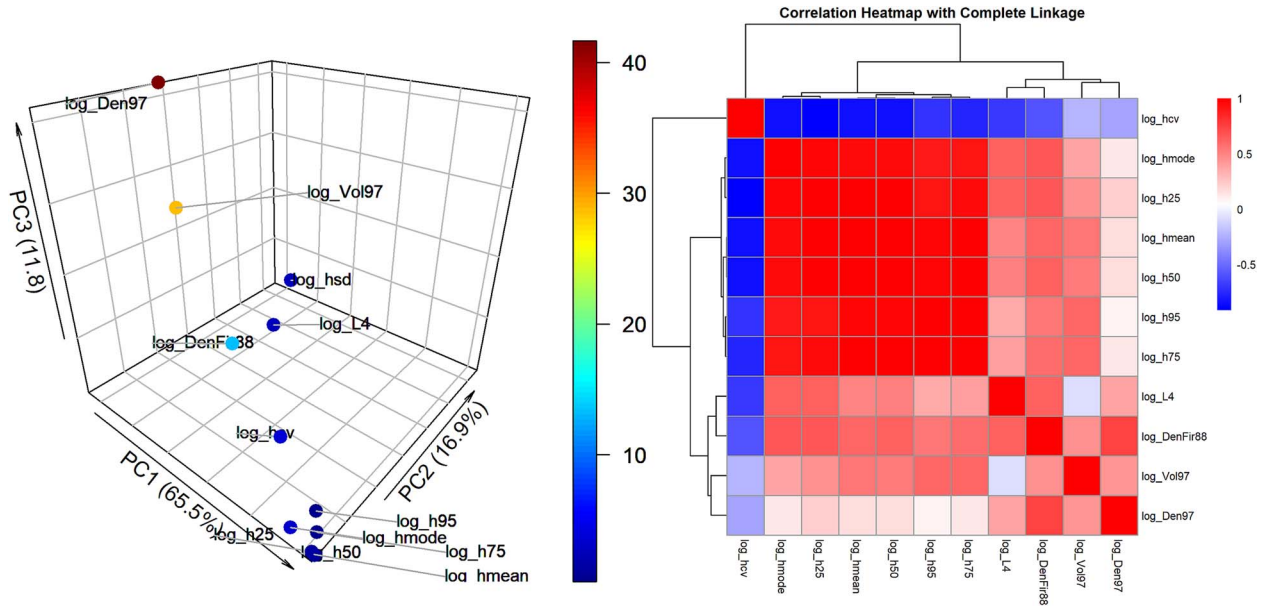
Figure 2. Contribution of auxiliary variables to the three most dominant PCs (left panel) and correlation matrix of auxiliary variables (right panel).

Our stepwise variable selection approach utilizes covariates from Groups A and B in their original form or one of the two transformed forms, without mixing them. Specifically, for the volume response variable, log transformation positively influenced both correlation and model assumptions. Consequently, for Volume, only Group A variables were represented by log-transformed forms. This procedure was similarly applied to all response variables.

Multivariate Fay–Herriot models

This section introduces the modelling methodology based on MFH and specifically for a trivariate FH approach. Initially, we define N as a finite population comprising a total of D domains $N_1, \ldots, N_D$. Each domain d is characterized by a vector of three features of interest, denoted by $\boldsymbol{m}_d = (m_{d1}, m_{d2}, m_{d3})^T$. Corresponding to each feature vector $\boldsymbol{m}_d$, we have a series of direct estimators, $\boldsymbol{y}_d = (y_{d1}, y_{d2}, y_{d3})^T$. Hence, we formulate the (linear) sampling model as

$$\begin{pmatrix} y_{d1} \\ y_{d2} \\ y_{d3} \end{pmatrix} = \begin{pmatrix} m_{d1} \\ m_{d2} \\ m_{d3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{d1} \\ \varepsilon_{d2} \\ \varepsilon_{d3} \end{pmatrix}, d \in \{1, \ldots, D\}, \quad (5)$$

where $\boldsymbol{\varepsilon}_d$ represents a vector of white noises with $\boldsymbol{\varepsilon}_d \sim N(\boldsymbol{0}, \mathbf{V}_{\varepsilon d})$. Here, $\mathbf{V}_{\varepsilon d}$ denotes an assumed known covariance matrix of dimensions 3×3.

Furthermore, the features of interest m_{dk} for $\{d, k\} \in \{1, \ldots, D\} \times \{1, \ldots, 3\}$, are correlated (linearly) with a series of p_k explanatory/independent variables, where $\mathbf{x}_{dk} = (x_{dk1}, x_{dk2}, \ldots, x_{dkp_k})$ is a row vector containing the respective explanatory variables. We define $p = p_1 + p_2 + p_3$ as the total number of explanatory variables related to the elements of the $\boldsymbol{m}_d$ vector. Since, m_{dk} and $\mathbf{x}_{dk}$ are linearly related, we lead to the regression formula:

$$\begin{pmatrix} y_{d1} \\ y_{d2} \\ y_{d3} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{d1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{d2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}_{d3} \end{pmatrix} \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \boldsymbol{b}_3 \end{pmatrix} + \begin{pmatrix} u_{d1} \\ u_{d2} \\ u_{d3} \end{pmatrix}, d \in \{1, \ldots, D\}$$

or

$$\begin{pmatrix} y_{d1} \\ y_{d2} \\ y_{d3} \end{pmatrix} = \begin{pmatrix} x_{d11} & \cdots & x_{d1p_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & x_{d21} & \cdots & x_{d2p_2} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & x_{d31} & \cdots & x_{d3p_3} \end{pmatrix}_{3 \times p}$$

$$\begin{pmatrix} b_{11} \\ \vdots \\ b_{1p_1} \\ b_{21} \\ \vdots \\ b_{2p_2} \\ b_{31} \\ \vdots \\ b_{3p_3} \end{pmatrix}_{p \times 1} + \begin{pmatrix} u_{d1} \\ u_{d2} \\ u_{d3} \end{pmatrix}, d = 1, \ldots, D. \quad (6)$$

We observe that $\boldsymbol{b}_k^T$, $k = 1, \ldots, 3$, is a row vector containing the regression parameters for the feature m_{dk}. Additionally, $\boldsymbol{u}_d$ is a vector of white noises with $\boldsymbol{u}_d \sim N(\boldsymbol{0}, \mathbf{V}_{ud})$, where $\mathbf{V}_{ud}$ is a covariance matrix of dimension 3×3. Matrices $\mathbf{V}_{ud}$ depend on n unknown parameters $(\theta_1, \ldots, \theta_n)$, where $1 \leq n \leq 3 + \binom{3}{2}$ (Benavent and Morales 2016). Typically, the elements of the white noise vectors $\boldsymbol{u}_d$ and $\boldsymbol{\varepsilon}_d$ are assumed to be independent. The linear model of Equation (6) is referred to as linking model, while the MFH model is defined in two stages, based on the formulations of Equations (5) and (6). To consolidate the aforementioned quantities, we concatenate them to create column vectors, such as $\boldsymbol{\varepsilon} = \underset{1 \leq d \leq D}{col} \left(\underset{1 \leq k \leq 3}{col} (\varepsilon_{dk}) \right)$ and $\boldsymbol{u} = \underset{1 \leq d \leq D}{col} \left(\underset{1 \leq k \leq 3}{col} (u_{dk}) \right)$, which denoted the vectors of random errors and effects, respectively. In addition, we take $\mathbf{X} = \underset{1 \leq d \leq D}{col} (\mathbf{x}_d)$, $\mathbf{V}_u = \underset{1 \leq d \leq D}{diag} (\mathbf{V}_{ud})$, $\mathbf{V}_\varepsilon = \underset{1 \leq d \leq D}{diag} (\mathbf{V}_{\varepsilon d})$ and $\mathbf{Z}_D = \underset{1 \leq r \leq D}{col} (\delta_{rd}\mathbf{I}_3)$, $\mathbf{Z} = \underset{1 \leq d \leq D}{row} (\mathbf{Z}_d) = \mathbf{I}_{|3D|}$.

The operators $col(.)$ and $row(.)$ are used to denote the concatenation by columns and rows, correspondingly. Matrix $\mathbf{I}_{|3D|}$ denotes

the identity matrix of order $3D$ and δ_{rd} is the Kronecker function taking the value 1 in case $r = d$. As a result, the MFH model-1 (MFH1 hereafter) that derives from (5) and (6), can be written in matrix form as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \boldsymbol{\varepsilon} = \mathbf{Xb} + \mathbf{u} + \boldsymbol{\varepsilon} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \ldots + \mathbf{Z}_D\mathbf{u}_D + \boldsymbol{\varepsilon}. \tag{7}$$

According to the above comments, we understand that quantities $\mathbf{u}_1, \ldots, \mathbf{u}_D$ and $\boldsymbol{\varepsilon}$ are considered to be independent with $\boldsymbol{\varepsilon}_d \sim N(\mathbf{0}, \mathbf{V}_{\varepsilon d})$, $\mathbf{u}_d \sim N(\mathbf{0}, \mathbf{V}_{ud})$, for $d \in \{1, \ldots, D\}$ and consequently $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V}_\varepsilon)$, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u)$. The elements of the $\boldsymbol{\varepsilon}$ and $\mathbf{u}$ represent the sampling errors and random errors effects of the FH model. We emphasize that matrix $\mathbf{V}_\varepsilon$ is not necessarily diagonal resulting in correlated random errors. On the other hand,

$$\mathbf{V}_{ud} = \operatorname*{diag}_{1 \leq k \leq 3}\left(\sigma_{uk}^2\right), d \in \{1, \ldots, D\}, \tag{8}$$

where $n = 3$ and $\theta_k = \sigma_{uk}^2, k \in \{1, \ldots, 3\}$. In case that $\mathbf{V}_\varepsilon$ is also diagonal, we lead to the description of the UFH approach. Other multivariate approaches are associated with several specificities of the random effects, where autocorrelation (Esteban et al. 2011) or heteroscedasticity (González-Manteiga et al. 2010; Herrador et al. 2011) might exist.

The former model is known as autoregressive MFH AR(1) model-2 (MFH2 hereafter), which assumes the existence of first-order autocorrelations between the random effects of the model. In this case, the matrix $\mathbf{V}_{ud}$ is now non-diagonal and takes the form:

$$\mathbf{V}_{ud} = \sigma_u^2 \mathbf{W}_d = \frac{\sigma_u^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}, d \in \{1, \ldots, D\}. \tag{9}$$

Expression (14) requires the estimation of $n = 2$ parameters, $\theta_1 = \rho$ and $\theta_2 = \sigma_u^2$. The latter considers the existence of both autocorrelated and heteroscedastic effects (HAR(1)) in the MFH model-3 (MFH3 hereafter). On this occasion the random effects u_{d1}, u_{d2}, u_{d3} satisfy the formula

$$u_{dk} = \rho u_{dk-1} + a_{dk}, \tag{10}$$

where $u_{d0} \sim N\left(0, \sigma_0^2\right)$ and $a_{dk} \sim N\left(0, \sigma_k^2\right)$ for $k = 1, \ldots, 3$. Therefore, $n = 4$ parameters have to be estimated, namely $\theta_1 = \sigma_1^2, \theta_2 = \sigma_2^2, \theta_3 = \sigma_3^2$ and $\theta_4 = \rho$. The elements $\sigma_{udij}, i, j = 1, \ldots, 3$ of $\mathbf{V}_{ud}$ are $\sigma_{udii} = \sum_{k=0}^{i} \rho^{2k}\sigma_{i-k}^2$ and $\sum_{k=0}^{|j-i|} \rho^{2k+|j-i|}\sigma_{|j-i|-k}^2$ for $i \neq j$.

MFH3 was not presented in this paper due to computational infeasibility for this dataset, even with different auxiliary variables, unlike bivariate estimates that yielded results.

EBLUP

Based on the previously described model structure that characterizes the MFH methodology, we take that $E(\mathbf{y}) = \mathbf{Xb}$ and $\mathbf{V} = \operatorname{var}(\mathbf{y}) = \mathbf{Z}^T\mathbf{V}_u\mathbf{Z} + \mathbf{V}_\varepsilon = \mathbf{V}_u + \mathbf{V}_\varepsilon$. Thus, these quantities need to be estimated for each domain $d \in \{1, \ldots, D\}$. The best linear unbiased estimator for the coefficient vector $\mathbf{b}$ is

$$\hat{\mathbf{b}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}, \tag{11}$$

while

$$\hat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}^T\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\right), \tag{12}$$

and

$$\hat{\mathbf{m}} = \mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}}, \tag{13}$$

represent the BLUP. The residual maximum likelihood method (REML) maximized the joint probability distribution of a vector of $3D - p$ independent contrasts $\boldsymbol{\omega} = \boldsymbol{\Omega}^T\mathbf{y}$, where matrix $\boldsymbol{\Omega}$ consists of linearly independent columns. Moreover, $\boldsymbol{\Omega}$ is of dimension $3D \times (3D - p)$, where $\boldsymbol{\Omega}^T\mathbf{X} = \mathbf{0}$ and $\boldsymbol{\Omega}^T\boldsymbol{\Omega} = \mathbf{I}_{|3D-p|}$. The joint probability distribution of $\boldsymbol{\omega}$ corresponds to the REML likelihood. Hence, the REML log-likelihood of model the MFH model (11) is:

$$l(\boldsymbol{\theta}) = -\frac{3D - p}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{V}| + \frac{1}{2}\log|\mathbf{X}^T\mathbf{X}|$$
$$-\frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^T\mathbf{P}\mathbf{y}, \tag{14}$$

where $= (\theta_1, \theta_2, \ldots, \theta_n)$, $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$, $\mathbf{PX} = \mathbf{0}$ and $\mathbf{PVP} = \mathbf{P}$. After taking the partial derivatives of (13) with respect to the elements of the parametric vector $\boldsymbol{\theta}$, we lead to the score vector $\mathbf{s}(\boldsymbol{\theta}) = (s_1(\boldsymbol{\theta}), \ldots, s_n(\boldsymbol{\theta}))^T$ with $s_i(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i}$. These scores participate in the REML updating equation of the Fisher-scoring algorithm leading to the estimation of the θ_i parameters, $\hat{\theta}_i$ for $i \in \{1, \ldots, n\}$ (Desiyanti et al. 2023). Taking advantage of this series of parameters we culminate in

$$\hat{\mathbf{V}} = \hat{\mathbf{V}}_u + \mathbf{V}_\varepsilon = \mathbf{V}_u\left(\hat{\boldsymbol{\theta}}\right) + \mathbf{V}_\varepsilon, \tag{15}$$

while the EBLUP of $\mathbf{m}$ is given by $\hat{\mathbf{m}}_E = \mathbf{X}\hat{\mathbf{b}}_E + \mathbf{Z}\hat{\mathbf{u}}_E$, where

$$\hat{\mathbf{b}}_E = \left(\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}, \tag{16}$$

and

$$\hat{\mathbf{u}}_E = \hat{\mathbf{V}}_u\mathbf{Z}^T\hat{\mathbf{V}}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\right). \tag{17}$$

We note that the E subscript included in quantities $\hat{\mathbf{m}}_E, \hat{\mathbf{b}}_E$ and $\hat{\mathbf{u}}_E$ is employed to distinguish EBLUP from the BLUP estimations. More details regarding the mean squared crossed errors matrices (MSCE) of EBLUP estimates are included in the Supplementary Material.

Model assumptions validation

This study conducts a thorough validation of model assumptions for MFH models in forest inventory estimation. The methodology involves various diagnostic analyses to evaluate the reliability, accuracy, and precision of the forestry estimation models. These analyses involve testing normality assumptions, conducting homoscedasticity analysis, performing bias analysis, and applying calibration diagnostics. Additionally, the study compares direct estimates with SAE CIs, and assesses coverage diagnostics to evaluate model performance (Brown et al. 2001; Fasulo 2022). More specifically we conduct:

i) Validation of normality assumptions in FH residuals and random effects: We evaluated the normality of standardized residuals and random area effects, crucial for validating model assumptions. This involved graphical inspections of Q-Q plots and the Shapiro–Wilk statistical tests.

ii) Homoscedasticity analysis: The homoscedasticity assumption was investigated using the Goldfeld–Quandt test for the response variables within the MFH models. Visual inspection of standardized residuals was also performed to detect

any patterns suggestive of violations of the homoscedasticity assumption.

iii) Bias analysis: To assess the bias introduced by the MFH models, a Bias Diagnostic Analysis was conducted. This involved comparing MFH estimates with direct estimates for the three response variables using diagnostic scatter plot analysis. The alignment of the regression line with the diagonal bisector line (y = x) was evaluated to determine unbiasedness, with high R-squared values indicating the absence of bias. Additionally, the goodness of fit diagnostic was employed to evaluate how close the model-based estimates are to the direct estimates when they are accurate, in $\alpha = 0.05$ significance level. Squared differences between model and direct estimates are computed, weighted by their variances, and then summed. Comparisons between this sum and critical values from the quantiles of Chi-squared distribution are made to determine the lack of bias of the model estimates. Finally, a Wald goodness of fit statistic is calculated to further assess the goodness of fit of the model.

iv) Calibration diagnostic: Furthermore, to assess the reliability and detect potential bias in (M)FH estimates relative to direct estimates (DIR) across higher aggregation levels or larger small areas, the calibration ratio (CR) was applied. Given the probability sampling design, the study maintains design-unbiased properties, enabling the comparison between the unbiased direct estimates and the model-based small area estimates at larger domains (see: "Estimating Unsampled Domains with Cluster Information"). CR was calculated for each variable in large domains, comparing the mean MFH estimates with the mean direct estimates. Within the entire population, comprising a total of 220 sample units, and across each of four large domains, each encompassing an average of 55 units, the following formula was utilized: $CR_{dk} = \frac{\left(\overline{y}_{dk}^{FH} - \overline{y}_{dk}^{DIR}\right)}{\overline{y}_{dk}^{DIR}} * 100$. Here $\overline{y}_{dk}^{FH}$ represents the mean FH estimates and $\overline{y}_{dk}^{DIR}$ represents with the mean direct estimates for each variable k in large domain d. A CR of 0% indicates equality between MFH and direct estimates, while positive/negative values suggest overestimation/underestimation by MFH relative to direct estimates.

v) Comparison of direct estimates and SAE CIs: The number of direct estimates falling within the SAE CIs for all MFH response variables was counted. Additionally, the overlap of CIs across different SAE methods and variables was assessed.

vi) Coverage diagnostics: A coverage diagnostic was conducted to compare the 95% adjusted CIs of model-based and direct estimates. Adjusted CIs were calculated for both types of estimates, and statistical tests were performed to quantify the degree of overlap between the intervals, providing insights into the reliability and accuracy of the model-based estimates relative to direct estimates.

Evaluation metrics for assessing FH precision

To assess the precision of our estimates, we employed several evaluation metrics. First, we evaluated the estimates based on the CV, a commonly used measure for assessing estimator error in management forestry inventories. CV values below 10 or 15 are generally considered reliable (Mauro et al. 2016), akin to the RSE in SRS, indicating lower variability and higher precision of the estimators.

Moreover, we evaluated the precision of direct estimates from SRS and model-based estimators by comparing their standard error ratios (SER_d) for each small area to that of EBLUP for domain d. We compare the standard error of direct estimates, $SE(\overline{Y})^{DIR} = \left[V\left(\overline{Y}\right)^{DIR}\right]^{0.5}$, with the root MSE (RMSE) of EBLUP estimates $RMSE(\overline{Y})^{EBLUP} = \left[MSE\left(\overline{Y}\right)^{EBLUP}\right]^{0.5}$. This comparison allowed us to gauge the precision of our model-based estimates relative to direct estimates (Coulston et al. 2021). The SER was examined for FH models, where $SER_d = \frac{RMSE(\tilde{Y}_d)}{SE(\tilde{Y}_d)}$. A SER_d value <1 signifies that the EBLUP estimates exhibit enhanced precision compared to the direct estimates. It is important to note that while EBLUP may not be guaranteed unbiased, it can be considered when the model assumptions are met (after model validation).

Additionally, to compare the improvements in small area estimates between MFH and UFH models, we calculate the relative efficiency, expressed as the percentage gain (PG): $\left(1 - \frac{RMSE(\tilde{Y}_d)^{MFH1}}{RMSE(\tilde{Y}_d)^{UFH}}\right)$%. Here, $RMSE(\tilde{Y}_d)^{MFH1}$ represents the RMSE for the small area estimates obtained from the MFH1 and UFH models, respectively. The PG indicates the relative improvement in precision achieved by the MFH model over the UFH model, with higher values indicating greater enhancement. Finally, we present maps to visually depict the geographic distribution of CV and volume estimates obtained from the MFH1 model, including small areas with no sample or outliers.

All computational analyses and graphical visualizations were conducted using the R statistical language (R Core Team 2023). The implementation of MFH models was facilitated by the MSAE package (Permatasari and Ubaidillah 2021), while for estimating non-sampled areas using computed cluster information, EBLUP-FH estimates were obtained using the "msaefhns" function of the "msaeDB" package (Perwira and Ubaidillah 2021).

Results
Model selection

After conducting variable selection guided by the EBLUP-FH approach, minimizing the AIC criterion, and confirming the absence of multicollinearity, we derived the trivariate representation (Equation 18)

$$\begin{pmatrix} \widehat{Volume} \\ \widehat{BasalArea} \\ \widehat{MeanHeight} \end{pmatrix} \sim \begin{pmatrix} \log(h50) + \log(hcv) + \log(Den97) + \log(DenFir88) \\ Vol97^2 \\ hmean^2 + \log(DenFir88) \end{pmatrix}$$

(18)

This equation, formulated with logarithmic transformations and powers, aims to optimize model performance while ensuring adherence to the normality assumption for FH residuals and random effects. In Figure S.4 of Supplementary Material we present the correlation values after the application of the aforementioned preprocessing pipeline, while in Table S.1 of Supplementary Material we present the beta coefficients, standard errors, t-statistics, and p-values originating from the fitting of the UFH, MFH1 and MFH2 models.

1) Validating normality assumptions in FH residuals and random effects: Before evaluating the estimates of FH models, we present the validated model assumptions. Firstly, it was confirmed that both the standardized residuals and random area effects in the FH models adhered to normality assumptions. This validation was supported by graphical inspections of Q-Q plots, and the Shapiro–Wilk statistical tests.

2) Heteroscedasticity—homoscedasticity: The Goldfeld–Quandt test (Table S.2 of Supplementary Material) results indicated homoscedasticity for all variables of interest across the FH models, except for MFH2 for Height. Specifically, while the null hypothesis was accepted for UFH, MFH1 and MFH2, indicating homoscedasticity, it was rejected only for MFH2 and Height, suggesting potential heteroscedasticity. Figure S.5 of Supplementary Material provides a visual inspection of standardized residuals, which in almost all cases did not reveal any discernible patterns, supporting the homoscedasticity of residuals.

3) Bias and R-squared: The bias diagnostic scatter plots (Fig. 3) demonstrated a close alignment between the blue regression line and the black diagonal bisector line across all three examined forestry characteristics. This alignment indicated minimal bias in UVF and MFH1 estimates, whereas MFH2 exhibited biases in volume and basal area, supported by R-square values of 0.71 and 0.67 for volume and BA, respectively, compared to MFH1 with 0.90 and 0.86 or UFH with 0.96 and 0.94. Generally, UFH exhibited the lowest bias for volume and basal area variables, followed by MFH1. Notably, mean height estimates showed negligible bias across all FH models, attributed to the strong correlation between height auxiliary variables and response variables. More details can be found in Table S.3 of Supplementary Material.

Additionally, the goodness of fit diagnostic consistently supports the null hypothesis across all variables and methods, with the Wald statistic (W) consistently indicating that the null hypothesis cannot be rejected. The Chi-squared statistic (χ^2) reinforces this, with *P*-values close to or equal to 1 in all cases. Hence, in all cases the *P*-values declare that there is no strong evidence against the null hypothesis (Table S.4 of Supplementary Material). These results could be considered as an additional indicator of the low or zero levels of bias, increasing the trustworthiness of the produced generated estimates.

1) Calibration—BIAS from Grand Mean—Large Domains and Quantiles: Table 2 presents percentage CR diagnostic results for UFH, MFH1, and MFH2 models, evaluating Volume, basal area, and Mean Height estimates across various domains and the overall population. Overall, the models demonstrate satisfactory performance, with minor deviations observed between model and direct estimates. Most models exhibit small discrepancies across different domains, indicating effective calibration. Only a few instances show deviations exceeding CR > 1%, suggesting generally accurate estimation. Notably, MFH2 had slight deviations with a CR of −1.56% for volume in the large domain.4 and 1.18% for basal area in the large domain.2, while UFH for large domain.3 and basal area had a CR of 1.06%. The CR for the entire population (grand means) reflects minimal differences between direct and model-based estimates, <0.5%, indicating minimal bias across the entire population.

2) Analysis of the direct estimates with respect to the SAE CI: The examination of direct estimates and SAE CIs showed that UFH and MFH1 methods generally aligned well, with most direct estimates falling within the range of SAE CIs across various variables. However, significant deviations were observed with the MFH2 method. For volume, basal area, and mean height variables, a substantial proportion of direct estimates (58%, 58%, and 26%, respectively) fell outside the corresponding SAE CIs under the MFH2 method, indicating potential bias. Despite these deviations, there was

a consistent overlap of CIs across all response variables, regardless of whether they were derived directly from sampling or the FH models.

3) Coverage diagnostics—tests the validity between the 95% adjusted CIs: In our study, we conducted coverage diagnostic to evaluate the alignment between the 95% adjusted CIs of model-based estimates and those of direct estimates. Across all variables and SAE methods tested, no instances of non-coverage or non-overlap were observed. The *P*-values, all being 1, indicate strong support for accepting the null hypothesis that the overlap between CIs is 95%. These findings assure reliable coverage of estimated parameters, ensuring consistency between model-based and direct estimates across various variables and SAE methods.

Evaluation of estimates

1) CV: In the evaluation of precision estimates, both MFH1 and MFH2 models consistently demonstrated CV values below 5%, indicating high precision across all variables (Table 3). However, the UFH model showed CV values exceeding 5% in certain domains, particularly for volume and basal area. Given the observed bias in the MFH2 model, the reliability of forestry estimates from MFH1 is deemed suitable for effective forest management decisions. Specifically, MFH1 performed exceptionally well, with 0% of estimates exceeding CV > 5%, contrasting with UFH's 10.5%. Furthermore, both MFH1 and UFH exhibited 0% for CV > 10%, while direct volume estimates exceeded 89.5%.

2) SER—effectiveness: In analyzing the distribution of SERs across the FH methods, discernible patterns emerge. Violin plots (Fig. 4) visually represent the SER distribution for FH estimates, where lower SER values indicate more precise model estimates compared to SRS. Notably, UFH tends to exhibit higher SERs, implying greater variability in its estimates. In contrast, both MFH1 and MFH2 consistently display lower SERs, suggesting more accurate small area statistics. Specifically, MFH2 consistently demonstrates the lowest SERs, indicating superior precision. However, it is crucial to interpret these findings in light of bias considerations identified in previous diagnostics. Overall, MFH1 exhibits higher efficiency relative to all FH models, considering both precision and bias. In violin plots we excluded one SER value exceeding 3 in mean height, which had only three sample plots and a 0.559% RSE.

3) PGs: The MFH1 method consistently exhibits enhanced precision compared to UFH across all the response variables, as indicated by positive PGs. Median PGs were 17.22% for volume, 13.91% for basal area, and 3.95% for mean height, with mean PGs slightly higher at 18.27%, 16.51%, and 10.87%, respectively. These findings underscore MFH1's efficacy in improving UFH estimate precision, particularly when targeting desired levels such as CV < 5%. A graphical representation via a boxplot (Fig. 5) further illustrates these PGs.

4) CIs: In Fig. 6, the points representing direct estimates (in gray), UFH estimates (in black), and MFH1 (in red) model estimates are depicted, along with their corresponding 95% CIs for growing stock volume. This illustrates that direct estimates do not significantly differ from the FH estimates. Regarding CIs, the MFH1 model shows relatively narrower intervals compared to those of the UFH model, while the wide intervals of the direct estimates are not shown. This observation underscores the necessity of applying SAE
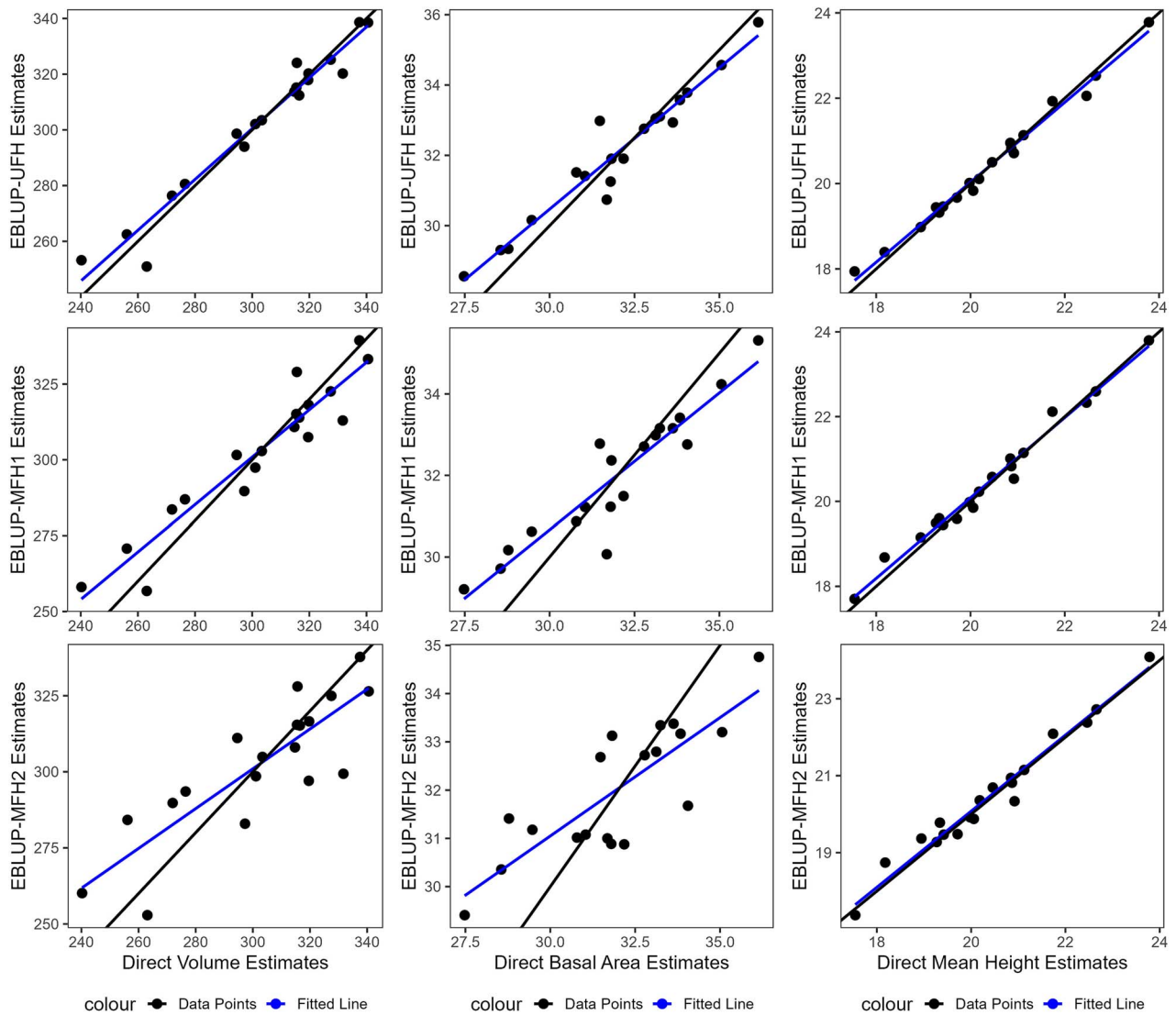
Figure 3. Bias diagnostic plot comparing MFH and UFH estimates with direct estimates for volume, basal area, and mean height. The black diagonal line represents the y = x bisector, and the blue regression line deviates from it, indicating systematic bias.

Table 2. Percentage CR in population and four large domains.

Variable		Volume			Basal area			Mean height		
Area \ Method	Sample Size	UFH	MFH1	MFH2	UFH	MFH1	MFH2	UFH	MFH1	MFH2
Domain.1	111	0.22	0.05	0.53	−0.25	−0.11	0.21	0.08	0.13	0.20
Domain.2	71	0.07	0.39	0.79	0.41	0.67	1.18	−0.11	0.19	0.20
Domain.3	27	0.07	0.03	−0.32	1.06	0.11	−0.51	−0.21	0.33	0.66
Domain.4	11	−0.10	−0.07	−1.56	−0.37	−0.75	−0.65	0.70	0.83	0.39
Population	220	0.09	0.13	0.07	0.28	0.09	0.18	0.03	0.31	0.36

techniques such as FH models, as very wide CIs indicate high uncertainty in the estimated direct values.

Moreover, through Table S.5 of Supplementary Material, we show another disadvantage of the model with the autoregressive random effects (MFH2), which leads to the selection of the more robust MFH1 model. An important characteristic that validates the efficiency and unbiasedness of a FH model, is the inclusion

of direct estimates in the CIs of the model-based estimations. We observe that in the case of MFH2, only 8, 8, and 14 out of the 19 estimates are contained in the CIs for the forest growing stock volume, basal area, and mean height, correspondingly. On the other hand, the direct estimates of only two and one domain for the growing stock volume and basal area, respectively, are excluded from the CIs of the MFH1 model, increasing the trustworthiness of its application.

Table 3. Distribution of CV and RSE across forestry variables and method/model.

| Variable	CV Threshold	RSE DIR	CV UFH	CV MFH1	CV MFH2
Volume	[0.00–2.00%]	0 (0.00%)	2 (10.53%)	3 (15.79%)	16 (84.21%)
	[2.10–5.00%]	2 (10.53%)	15 (78.95%)	16 (84.21%)	3 (15.79%)
	[5.10–10.00%]	10 (52.63%)	2 (10.53%)	**0 (0.00%)**	**0 (0.00%)**
	[10.10–15.00%]	2 (10.53%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	>15.00%	5 (26.32%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Basal area	[0.00–2.00%]	0 (0.00%)	3 (15.79%)	5 (26.32%)	16 (84.21%)
	[2.10–5.00%]	3 (15.79%)	15 (78.95%)	14 (73.68%)	3 (15.79%)
	[5.10–10.00%]	10 (52.63%)	1 (5.26%)	**0 (0.00%)**	**0 (0.00%)**
	[10.10–15.00%]	3 (15.79%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	>15.00%	3 (15.79%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Height	[0.00–2.00%]	1 (5.26%)	14 (73.68%)	18 (94.74%)	18 (94.74%)
	[2.10–5.00%]	12 (63.16%)	5 (26.32%)	1 (5.26%)	1 (5.26%)
	[5.10–10.00%]	5 (26.32%)	**0 (0.00%)**	**0 (0.00%)**	**0 (0.00%)**
	[10.10–15.00%]	1 (5.26%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	>15.00%	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
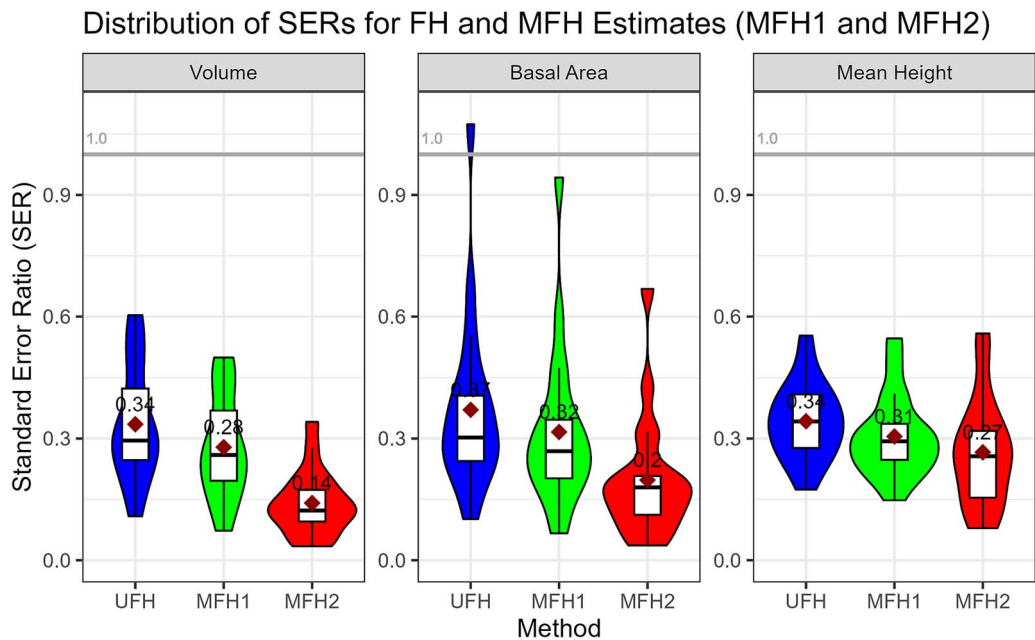
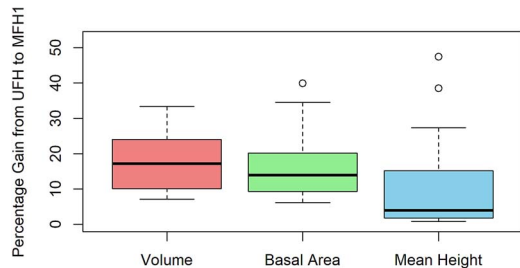**Figure 4.** Distribution of SERs for UFH, MFH1, and MFH2.

Figure 5. Distribution of PGs of MFH1 compared to UFH across parameters.

1) Maps—geographic distribution of volume predictions and CV: Finally, Fig. 7 illustrates the geographic distribution of volume predictions (left panel) and the corresponding CV% estimates (right panel) obtained from the MFH1 model. These maps provide a visual depiction of the forest inventory estimates, including predictions for small areas with no sample or outliers. The produced estimations derived from
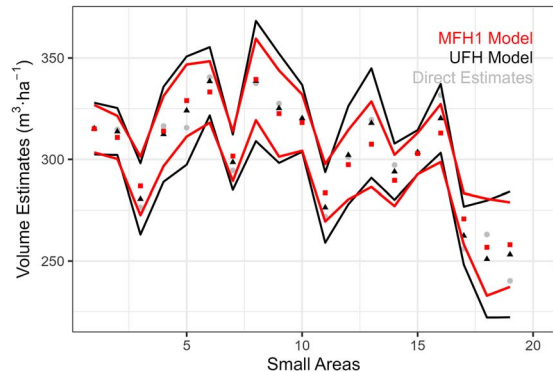

Figure 6. Direct, UFH, and MFH1 point estimates of the forest growing stock volume are shown as points, triangles, and rectangles, respectively. CIs for the univariate UFH model (black line) are wider than those for the multivariate MFH1 model (red line).

MFH1 model, accompanied by the respective MSE values, are included in Table S.6 of Supplementary Material. In
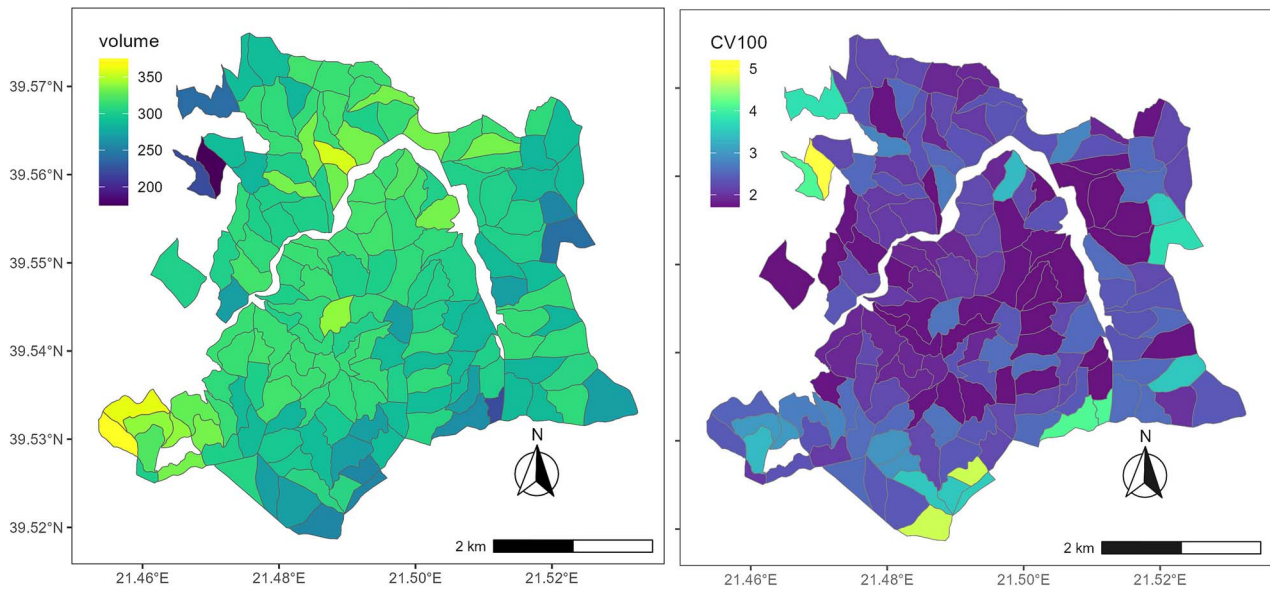
Figure 7. The left panel displays the map of volume predictions obtained from the MFH1, while the right panel shows the corresponding CV% estimates.

addition, Table S.7 of the Supplementary Material provides a comparison between the Direct, UFH, and MFH1 estimates of forest growing stock Volume, alongside their corresponding RSE% and CV% for sampled small areas (excluding outliers). On the left side of Table S.7, the distribution of sample size, number of FMUs, and the total area in hectares for each small area or cluster is displayed. This table emphasizes the precision improvements achieved through the UFH and MFH1 models, with significantly reduced CV% compared to the higher variability found in direct estimates, particularly in areas with small sample sizes.

Discussion

In the present study, a comparison and evaluation of the reliability of forest biometric estimates produced by the MFH was conducted against those of the UFH model and direct estimates, which are exclusively based on the sampling design. The goal of two (MFH1 and MFH2) out of the three variations of the MFH model explored was to reduce the MSE, the CV, and CIs compared to the UFH model and direct estimates. The application of MFH methods recently began to appear in the forestry sector by Ver Planck (2018 Ch. 3) and Frank (2020 Ch. 4, Manuscript III), aiming for SAE of forest biometric parameters. Here, a larger number of variables of interest than those discussed by Frank (2020 Ch. 4, Manuscript III) are explored, and parameter estimation is carried out using the EBLUP estimators instead of HB models by Ver Planck (2018 Ch. 3).

The MFH methodology enables the simultaneous prediction of multiple response variables by considering their interrelationships and integrating auxiliary data sources, such as satellite imagery and field measurements. To further support model applicability, the preprocessing pipeline identifies small areas with stronger linear correlations between target and auxiliary variables through clustering, making area-level models feasible even in challenging scenarios. By utilizing correlations between forest attributes, such as biomass, volume, basal area, tree height, stem density, and tree diameter distribution, the multivariate approach produces more reliable and accurate estimates compared to univariate models.

Among the MFH variations tested, MFH1 with homogeneous random effects yielded the best estimation efficiency, maintaining all CV values below 5%. MFH2, with heteroscedastic random effects, yielded smaller CVs but exhibited relatively larger bias, while MFH3, accounting for both heteroscedasticity and autocorrelation random effects, was not computationally feasible for this dataset even with different auxiliary variables. Notably, mean height estimates showed the minimum variance reduction or efficiency gain compared to volume and basal area estimates across all FH models, attributed to highly correlated height remote sensing auxiliary variables. This observation aligns with another study that confirms that "benefits from borrowing strength from another survey tend to diminish when strongly predictive covariates are available" (Franco and Maitra 2023 p. 4).

Our methodology estimates unsampled or outlier-removed areas by aggregating small areas with similar patterns through cluster-based random effects, offering an efficient alternative to synthetic estimation (Desiyanti et al. 2023, Torkashvand et al. 2017). While UFH models rely on the strong linear relationship of response variables with covariates, MFH models prioritize the strong relationship between response variables themselves. When these correlations are less pronounced, both UFH and MFH models may yield similar results (Guha and Chandra 2022). Notably, for bivariate FH models, response correlations below 0.5 result in relatively low variance reduction (<20%), irrespective of the ratio of random area effects and sampling variances (Franco and Bell 2022).

Another critical assumption in MFH modelling is the estimation of sampling variances, which are practically derived from sample data. However, using direct sampling variances based on very small sample sizes can lead to significant discrepancies in the estimated variances and covariances, resulting in computational issues due to an ill-posed inversion of the covariance matrix $V_{\varepsilon d}$. To mitigate discrepancies caused by direct sampling variances from small samples, our pipeline includes a variance smoothing step, which aids convergence. In this case study, we applied a straightforward smoothing method (Goerndt et al. 2011; Coulston

et al. 2021). Nonetheless, utilizing more advanced GVFs may lead to more accurate smoothing and, consequently, improved model performance (Guha and Chandra 2022).

Although frequentist (EBLUP) FH models are more commonly applied than Bayesian alternatives, they still face estimation challenges. For instance, maximum likelihood algorithms (ML, RELM) in FH-EBPLUP model fitting can result in negative or zero estimates for of random area effects variances. When encountering zero model variances, an adjusted form of maximum likelihood (Angkunsit and Suntornchost 2020; Angkunsit and Suntornchost 2022) or the expectation–maximization algorithm (Ávila-Valdez et al. 2020) can be effective, as they refine variance component estimates for FH models of small area means. From a Bayesian perspective, Hierarchical and Empirical Bayes approaches have been applied to MFH models (Ghosh et al. 1991), though studies in median income estimation indicate these methods may not offer clear advantages over simpler univariate models (Datta et al. 2002).

The reliability of small area estimates depends on key statistical assumptions, such as the normality and homoscedasticity of residuals, typical of regression models. To address these assumptions, we implemented preprocessing steps that included outlier detection and removal, smoothing of sampling variances and covariances, and applying logarithmic transformations, as shown in Equation (18). When these preprocessing steps are insufficient, alternative FH models, such as hierarchical FH, may offer a viable solution. These models are advantageous due to their reduced reliance on statistical assumptions, including the ability to model unknown sampling variances, and are particularly useful in small areas with limited data. However, the comparison of MFH approaches remains largely unexplored in forestry.

Accurately identifying and managing outliers was critical, as they may occur in correlation analysis, in residuals of fixed and random effects, or, uniquely in MFH models, within covariance matrices—adding an additional layer of complexity. Outlier treatment was integral to the preprocessing pipeline to ensure reliable estimates. In forestry, synthetic estimators are commonly used for small areas with fewer than two plots (Magnussen et al. 2017). In this study, we incorporated random area effects variability in unsampled and outliers areas by employing a linear mixed-effects model, minimizing potential bias of synthetic regression (Magnussen and Breidenbach 2017). To prevent the underestimation of variance for the 13 outliers among a total 32 small areas, we calculated a proxy for random area-level effects by re-clustering all areas into four larger clusters.

While the MFH model offers improved reliability over UFH, it introduces significant computational challenges. Limiting the number of response variables is essential to managing complexity, as each variable added requires estimating an increasing set of variance and covariance components. For a MFH model with K response variables, up to $K + \binom{K}{2}$ components must be estimated—K for the means and $\binom{K}{2}$ for the variance and covariance terms. This high dimensionality can lead to convergence challenges, especially when dealing with a relatively small number of small areas.

The integration of wall-to-wall remote sensing data has significantly advanced FIs, transitioning FMIs from subjective visual assessments to more objective, model-based approaches. This shift has narrowed the gap between FMIs and NFIs, enhancing data-driven decision-making. Despite these advances, the challenge of obtaining reliable small area estimates remains significant. Effective SAE methods rely on the availability of robust auxiliary data, careful model selection, and adequate field data. A

major constraint in this study was geolocation errors, which precluded the use of more reliable (multivariate) unit-level models or ABA that require accurate plot georeferencing (Maltamo et al. 2020). Correcting these inaccuracies or collecting new samples was impractical due to cost and labor demands, prompting our adoption of area-level approach.

Implementing the area-level approach for our heterogeneous forest data introduced challenges due to sparse field data, which initially weakened correlations with remote sensing auxiliary data. Initially, within FMUs, the assumptions of the area-level model were unmet due to weak correlations between field data and remote-sensing-derived auxiliary data. To address these correlation challenges, we conducted cluster analysis to define the smallest possible geographic small areas for area-level model application. This clustering enabled more effective use of remote sensing data, both as a tool for segmentation and as auxiliary data for FH models, reinforcing model accuracy where direct field data were limited (Georgakis et al. 2023). Without clustering, the small sample sizes would have weakened the area-level model's effectiveness. A similar strategy was employed by aggregating forest stands of the same age class and species group into 84 larger FMUs (small areas), with 58 containing at least two plots (Mauro, et al. 2017).

Although spatial MFH models would be ideal for spatially connected clusters, in our study, non-adjacent FMU clusters were more effective in maintaining strong correlation between auxiliary and target variables (Georgakis et al. 2023). Cluster analysis remains an optional tool, useful when the assumptions of the FH model are not met, as in our case study. Moreover, this methodology can be extended to include spatial random effects in scenarios where FMUs or small areas are spatially connected.

The size of small areas of interest is crucial for the precision of estimates in SAE. While SAE aims to downscale estimates to the finest geographic levels, ensuring a minimum sample size per small area is essential for reliable results. Research on SAE models, whether unit-level or area-level, has primarily utilized NFI data to deliver attributes for strategic decision-making at broader spatial scales, such as large regions, counties, municipalities, forest districts, or large FMUs. In contrast, SAE research has also focused on using local FMI data to provide estimates for smaller forest stand areas, supporting sustainable forest management and operational decision-making (Georgakis 2019).

A major challenge, and ongoing research question, is how to downscale NFI estimates to produce information comparable to FMIs at the stand level using SAE techniques. Replacing FMIs with NFI data is difficult and remains an ongoing effort. While NFI-based maps (e.g. SR16) can be useful for assessing unproductive forests or as baseline data for FMIs (Astrup et al. 2019), there is ongoing research into their application in managed forest types (Rahlf et al. 2021). Table 4 compares Rahlf et al. (2021) in their use of NFI-based maps as substitutes for FMIs to estimate timber volume in mature spruce stands with our research (Georgakis et al., this study).

This comparison highlights the advancements and contributions of our study within FMI and SAE under specific constraints. The work of Rahlf et al. (2021) pioneered the use of NFI-based predictions with ALS data, estimating timber volume in simpler forest types. In contrast, our research addressed challenges such as small sample sizes and geolocation errors by introducing cluster analysis as a preprocessing step. This step ensures strong correlations between predictive and auxiliary variables, critical for the SAE model's robustness. We employed an MFH model to estimate multiple forest attributes simultaneously using

Table 4. Comparison of national and FMI for small area estimates.

Aspect	Georgakis et al. (this study)	Rahlf et al. (2021)
Purpose	Improved small area statistics based on FMI data (without NFI data), addressing sample size limitations.	First SAE analysis of NFI-based map predictions for FMI use.
Methodology-Model	Multivariate area-level linear mixed model, accounting for small area variation. Unit-level was not applicable due to geolocation errors, limiting the full potential of 3D remote sensing data.	Compared synthetic regression or ABA from FMI and linear mixed-effects models with random effects from NFI, using 3D remote sensing data.
Field data	Sample plots had significant geolocation errors.	Sample plots had precise geolocations.
Auxiliary data	Used satellite-based WorldView-derived canopy height metrics, with little expected improvement from ALS data due to area-level model (Green et al. 2019). Strong correlation at area-level.	Used ALS metrics with strong correlations at unit-level.
Target variables	Estimated multiple (three) forest attributes.	Estimated timber volume (single attribute).
Forest type	Multi-layer, highly structured uneven-aged fir forest.	Simple, mature spruce stands; other forest types unexplored.
Additional preprocessing	Applied cluster analysis on FMUs to address sparse plots and improve correlation.	N/A
Bias validation	Validated on four large domains (aggregated clusters) with an average of 55 plots, enhancing direct and bias estimates.	Validated on six forest stands with an average of 10 plots each.
Area extent of predictions	Estimated at larger aggregated FMUs, with higher precision in MSE and CV.	Provided small area estimates at forest stand level, with larger errors.
Model accuracy	Demonstrated the MFH model's efficiency compared to UFH and direct estimates for FMU groups.	Similar accuracy for NFI and FMI models at plot and stand levels.

satellite-derived 3D data in an area-level context. This approach improved precision across larger FMUs and adapted to complex forest structures.

Our study can also be compared to Rahlf et al. (2014), who used 3D remote sensing data—such as ALS, stereo aerial photogrammetry (AP), satellite interferometric synthetic aperture radar (InSAR) from the TanDEM-X mission, and satellite radar-grammetry—for timber volume estimation. While Rahlf et al. (2014) employed unit-level models, our study applied area-level models. Rahlf et al. linear mixed-effects models with vegetation height and density metrics, reporting stand-level RMSEs of 12% for ALS and 13% for AP, with leave-one-out cross-validation RMSE for AP reaching 18.12%. Their focus was on small forest stands of 1–3 hectares, with 5–7 sample plots per stand (Rahlf et al. 2014).

In contrast, we achieved higher precision with CVs below 5% for the define small areas. This increased accuracy is likely due to our clustering approach, which aggregated FMUs into larger areas (average 100 hectares, 11.6 sample plots per small area) through clustering. While our approach led to improved precision, the use of cluster analysis did result in some loss of spatial resolution, producing less detailed estimates. However, this impact was minimized due to the homogeneity within each cluster. Clustering was optimized for within-cluster homogeneity, reducing the risk of losing important fine-scale details. Lastly, our method proved more precise and informative than traditional stratification based solely on species composition, social stage, or forest age.

The challenge of obtaining precise small area estimates remains an open problem, especially for detailed forest attributes at the forest stand or FMU level. FMIs and NFIs complement each other to provide a holistic approach to forest conservation and climate change mitigation. While NFIs target broader spatial scales, small area estimates at the management level ensure that individual forest stands effectively contribute to carbon absorption. Managing small forest areas can optimize local carbon

sequestration, ensuring forests remain productive and serve as effective carbon sinks. Monitoring and managing growth rates at the stand level is critical for maximizing carbon sequestration rates, thereby enhancing forests' capacity to mitigate climate change while maintaining ecosystem services.

This study proposes a methodology to address common forest inventory challenges, including sparse data and geolocation errors. The proposed methodology is expected to yield increasingly reliable estimates as the number of small areas increases, thereby improving the efficiency of SAE (Rao and Molina 2015). The application of the MFH model in complex, uneven-aged forests demonstrates its potential adaptability to simpler, even-aged forests and smaller spatial units, offering high resolution estimates. While this study used WorldView-derived CHM data, which are less accurate than LiDAR or ALS data, future research could explore the integration of LiDAR to assess their impact on MFH models (Green et al. 2020). Additionally, extending these methods to diverse forest types, testing various remote sensing technologies, and refining small area estimates using available forest inventory data.

The integration of advanced preprocessing techniques further enhances the flexibility of MFH models allowing them to be adapted to various forest ecosystem types, including both even-aged and uneven-aged forests. These models have been successfully applied to complex, multilayered, uneven-aged forests, demonstrating their versatility across a range of forest types and dataset. This adaptability makes them suitable for applications in both small-scale FMUs and large-scale NFIs.

The methodology is particularly effective in scenarios with very limited sample sizes (0–3 plots per small area). When sufficient sample sizes (7–12 plots per 100-hectare area in uneven aged forests) are available, MFH models reduce the need for preprocessing steps like clustering. For larger datasets, such as NFIs with more sample plots per area, the MFH approach becomes more

straightforward to apply, maintaining precision and reliability without the complexity of intensive preprocessing.

In this study, we explored the application of MFH models in forestry, demonstrating their advantages over UFH models. Previous studies, including those by Datta (1999) and Moretti (2018), have shown that MUL models provide more efficient and precise estimates than univariate unit-level (UUL) models. Drawing on findings from comparisons of MUL and UUL in agricultural datasets (e.g. Battese et al. 1988), we anticipate similar improvements in the efficiency and precision of forestry applications using MUL models.

A key advantage of unit-level models is their ability to use individual-level data and for MUL models, the additionally benefit of using correlations between target variables. Thia approach captures more detailed relationships than area-level models, which rely on aggregated data. MUL models reduce the need for data aggregation or clustering, often required in UFH or MFH models with small sample sizes, and aim to provide estimates at the original small areas (e.g. FMUs) without introducing uncertainty from clustering. However, a significant limitation of MUL models is the requirement for detailed plot-level data, which are often unavailable in FIs. Furthermore, they are computationally demanding. When only aggregated data or unit-level data with errors are available, area-level models are more practical.

Future research could provide valuable insights into the relative strengths of MUL and MFH models. While MFH models outperformed UFH models in this study, direct comparisons between MUL and MFH models remain unexplored across disciplines. Although MUL models have shown superior performance over UUL models in agricultural datasets (Datta et al. 1999; Moretti et al. 2018), similar comparisons in forestry are lacking. Exploring the MUL approach may offer more efficient estimates compared to MFH models, particularly when covariates are available at the sample unit level. Ongoing SAE research aims to further refine this methodology, exploring how NFIs can achieve the precision of FMIs and deliver the detailed stand-level estimates essential for effective forest management.

Conclusion

This study introduces a novel approach to SAE in forestry by integrating trivariate EBLUP-FH models with advanced preprocessing techniques. The application of clustering analysis, variable selection, and outlier treatment has proven essential for the effective use of MFH models, enabling precise small area estimates. One of the main advantages of the MFH model is its ability to consider the correlation among response variables and make multiple predictions without the need for multiple model fittings. Despite its added complexity and computational challenges, the MFH model led to increased precision of small area estimates compared to the traditional UFH models.

The findings demonstrate that auxiliary variables from remote sensors and past censuses are effective covariates for the MFH models. MFH achieved small area estimates with CVs below 5% for attributes such as forest growing stock volume, basal area, and Lorey's mean tree height—providing highly precise estimates for forest management. However, when response variables are highly correlated, MFH models shows slightly diminished gains in precision, particularly for mean height.

The adaptability of the proposed methodology is anticipated to be significant in other SAE contexts. It was applied effectively in one of the most challenging forest inventory scenarios, with sparse data in uneven-aged forests, proving its utility for simpler forest structures, more homogeneous forest types, and larger datasets (with more domains and sample sizes per domain). As sample sizes increase, the methodology simplifies, reducing the need for extensive preprocessing steps like clustering and outlier treatment. This flexibility makes the approach valuable for both small and large datasets, applicable to both NFIs and FMIs. Overall, the MFH methodology is highly generalizable and adaptable, providing a robust framework for SAE in forestry with similar data constraints and potentially extending to other inventory datasets. Future research should explore its application in different forest types to further refine and validate its effectiveness.

Acknowledgements

Author contributions

Aristeidis Georgakis (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing), Vasileios Papageorgiou (Formal analysis, Investigation, Methodology, Validation, Writing—original draft, Writing—review & editing), and Georgios Stamatellos (Conceptualization, Project administration, Resources, Supervision)

Supplementary data

Supplementary data are available at *Forestry* online.

Conflict of interest

The authors declare that there are no conflicts of interest.

Data availability

Sample survey data used in the study are available upon request by the University Forest Administration and Management Fund at Aristotle University of Thessaloniki (https://www.auth.gr/en/university_unit/tameio-uniforest-en/ and https://uniforest.auth.gr/en/, accessed on 01 May 2024). The digital Canopy Height Model is available from the authors upon request.

References

Angkunsit A, Suntornchost J. Adjusted maximum likelihood method for multivariate Fay–Herriot model *J Stat Plan Infer*. 2022;**219**: 231–49.

Angkunsit A, Suntornchost J. Bivariate Fay–Herriot models with application to Thai socio-economic data *Naresuan Univ J Sci Technol*. 2020;**29**:36–49.

Astrup R, Rahlf J, Bjørkelo K *et al*. Forest information at multiple scales: development, evaluation and application of the Norwegian Forest resources map SR16 *Scand J For Res*. 2019;**34**:484–96.

Ávila-Valdez JL, Huerta M, Leiva V *et al*. The Fay–Herriot model in small area estimation: EM algorithm and application to official data *REVSTAT-Stat J*. 2020;**18**:613–35.

Battese GE, Harter RM, Fuller WA. An error-components model for prediction of county crop areas using survey and satellite data *J Am Stat Assoc*. 1988;**83**:28–36.

Benavent R, Morales D. Multivariate Fay–Herriot models for small area estimation *Comput Stat Data Anal*. 2016;**94**:372–90.

Breidenbach J, Astrup R. Small area estimation of forest attributes in the Norwegian National Forest Inventory *Eur J For Res*. 2012;**131**: 1255–67.

Breidenbach J, Magnussen S, Rahlf J *et al*. Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data *Remote Sens Environ*. 2018;**212**: 199–211.

Brown G, Chambers R, Heady P *et al*. Evaluation of Small Area Estimation Methods—An Application to Unemployment Estimates from the UK LFS. In: *Proceedings of Statistics Canada Symposium 2001. Achieving Data Quality in a Statistical Agency: A Methodological Perspective*. Statistics Canada. Ottawa, Canada, 2001, 1–10.

Burgard JP, Morales D, Wölwer A-L. Small area estimation of socioeconomic indicators for sampled and unsampled domains *AStA Adv Stat Anal*. 2022;**106**:287–314.

Chandra H, Salvati N, Sud UC. Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India: an application of small-area estimation technique *J Appl Stat*. 2011;**38**: 2413–32.

Coulston JW, Green PC, Radtke PJ *et al*. Enhancing the precision of broad-scale forestland removals estimates with small area estimation techniques *Forestry: Int J For Res*. 2021;**94**:427–41.

Datta GS, Day B, Basawa I. Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation *J Stat Plan Infer*. 1999;**75**:269–79.

Datta GS, Lahiri P, Maiti T. Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data *J Stat Plan Infer*. 2002;**102**:83–97.

Desiyanti A, Ginanjar I, Toharudin T. Application of an Empirical Best Linear Unbiased Prediction Fay–Herriot (EBLUP-FH) Multivariate Method with Cluster Information to Estimate Average Household Expenditure. *Mathematics*. 2023;**11**:135.

Dettmann GT, Radtke PJ, Coulston JW *et al*. Review and synthesis of estimation strategies to meet small area needs in Forest inventory *Front For Glob Change*. 2022;**5**:1–15.

Esteban MD, Morales D, Pérez A *et al*. Two area-level time models for estimating small area poverty indicators *J Indian Soc Agric Stat*. 2011;**66**:75–89.

Fassnacht FE, White JC, Wulder MA *et al*. Remote sensing in forestry: Current challenges, considerations and directions *Forestry: Int J For Res*. 2024;**97**:11–37.

Fasulo A. *R Package SAEval: Small Area Estimation Evaluation*, 2022.

Fay RE, Herriot RA. Estimates of income for small places: an application of James-stein procedures to census data *J Am Stat Assoc*. 1979;**74**:269–77.

Franco C, Bell WR. Using American community survey data to improve estimates from smaller U.S. surveys through bivariate small area estimation models *J Surv Stat Methodol*. 2022;**10**: 225–47.

Franco C, Maitra P. Combining surveys in small area estimation using area-level models *WIREs Comput Stat*. 2023;**15**:18.

Frank BM. *Aerial Laser Scanning for Forest Inventories: Estimation and Uncertainty at Multiple Scales*. PhD diss: Oregon State University, 2020.

Friedman HP, Rubin J. On some invariant criteria for grouping data *J Am Stat Assoc*. 1967;**62**:1159–78.

Georgakis A. Small area estimation in forest inventories. In: *Seventh International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 2019) and SECOTOX Conference*. Greece: Mykonos island, 2019, 769–80.

Georgakis A. Further improvements of growing stock volume estimations at stratum-level with the application of Fay-Herriot model. In: *33rd PanHellenic Statistics Conference. Statistics in the Economy and Administration, Greek Statistical Institute and the Departments of Business Administration and of Economics*. Larissa, Greece: University of Thessaly, 2021, 248–61.

Georgakis A, Gatziolis D, Stamatellos G. A primer on clustering of Forest management units for reliable design-based direct estimates and model-based small area estimation *Forests*. 2023;**14**: 1994.

Georgakis A, Papageorgiou VE, Gatziolis D *et al*. Temporal-like bivariate Fay-Herriot model: Leveraging past responses and advanced preprocessing for enhanced small area estimation of growing stock volume *Oper Res Forum*. 2024;**5**:9.

Georgakis A, Stamatellos G. Sampling design contribution to small area estimation procedure in Forest inventories *Mod Concep Dev Agron*. 2020;**7**:694–7.

Ghosh M, Datta GS, Fay RE. Hierarchical and empirical multivariate bayes analysis in small area estimation. In: *Proceedings of the 7th Annual Research Conference Bureau of the Census*, 1991, 63–79.

Goerndt ME, Monleon VJ, Temesgen H. A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables *Can J For Res*. 2011;**41**: 1189–201.

Goerndt ME, Monleon VJ, Temesgen H. Small-area estimation of county-level Forest attributes using ground data and remote sensed auxiliary information *For Sci*. 2013;**59**:536–48.

González-Manteiga W, Lombarda MJ, Molina I *et al*. Small area estimation under Fay–Herriot models with non-parametric estimation of heteroscedasticity *Stat Model*. 2010;**10**:215–39.

Green PC, Burkhart HE, Coulston JW *et al*. A novel application of small area estimation in loblolly pine forest inventory *Forestry: Int J For Res*. 2019;**93**:444–57.

Green PC, Burkhart HE, Coulston JW *et al*. Auxiliary information resolution effects on small area estimation in plantation forest inventory *Forestry: Int J For Res*. 2020;**93**:685–93.

Guha S, Chandra H. Measuring and mapping micro level earning inequality towards addressing the sustainable development goals – a multivariate small area modelling approach *J Off Stat*. 2022;**38**:823–45.

Guldin RW. A systematic review of small domain estimation research in forestry during the twenty-first century from outside the United States *Front For Glob Change*. 2021;**4**:1–16.

Haris F, Ubaidillah A. Mean Square Error of Non-Sampled Area in Small Area Estimation. In: *Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019*. Bogor, Indonesia, 2020, 383–394.

Harmening S, Kreutzmann A-K, Schmidt S *et al*. A framework for producing small area estimates based on area-level models in R *R J*. 2023;**15**:316–41.

Herrador M, Esteban MD, Hobza T *et al*. A Fay–Herriot model with different random effect variances *Commun Stat Theory Methods*. 2011;**40**:785–97.

Kangas A, Räty M, Korhonen KT *et al*. Catering information needs from global to local scales—Potential and challenges with National Forest Inventories *Forests*. 2019;**10**:800.

Magnussen S, Breidenbach J. Model-dependent forest stand-level inference with and without estimates of stand-effects *Forestry: Int J For Res*. 2017;**90**:675–85.

Magnussen S, Mauro F, Breidenbach J *et al*. Area-level analysis of forest inventory variables *Eur J For Res*. 2017;**136**:839–55.

Maltamo M, Packalen P, Kangas A. From comprehensive field inventories to remotely sensed wall-to-wall stand attribute data — a

brief history of management inventories in the Nordic countries *Can J For Res.* 2020;**51**:257–66.

Marhuenda Y, Morales D, del Carmen Pardo M. Information criteria for Fay–Herriot model selection *Comput Stat Data Anal.* 2014;**70**: 268–80.

Mauro F, Molina I, García-Abril A *et al.* Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels *Environ.* 2016;**27**:225–38.

Mauro F, Monleon VJ, Temesgen H *et al.* Analysis of area level and unit level models for small area estimation in forest inventories assisted with LiDAR auxiliary information *PloS One.* 2017;**12**:14.

McRoberts RE. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data *Remote Sens Environ.* 2010;**114**:1017–25.

McRoberts RE, Næsset E, Gobakken T. Inference for lidar-assisted estimation of forest growing stock volume *Remote Sens Environ.* 2013;**128**:268–75.

McRoberts RE, Næsset E, Gobakken T. Estimation for inaccessible and non-sampled forest areas using model-based inference and remotely sensed auxiliary information *Remote Sens Environ.* 2014;**154**:226–33.

Moretti A, Shlomo N, Sakshaug JW. Parametric bootstrap mean squared error of a small area multivariate EBLUP *Commun Stat Simul Comput.* 2018;**49**:1474–86.

Ngaruye I, Nzabanita J, Rosen, D.v. and Singull, M. Small area estimation under a multivariate linear model for repeated measures data *Commun Stat Theory Methods.* 2017;**46**:10835–50.

Nurizza WA, Ubaidillah A. A comparative study of multivariate Fay-Herriot model for small area estimation in various sample sizes *IOP Conf Ser Earth Environ Sci.* 2019;**299**:012027.

Permatasari N, Ubaidillah A. MSAE: Multivariate Fay Herriot models for small area estimation. In: *R Package Version 0.1.4*, 2021.

Perwira ZY, Ubaidillah A. msaeDB: Difference benchmarking for multivariate small area estimation. In: *R Package Version 0.2.1*, 2021.

Pratesi M. *Analysis of Poverty Data by Small Area Estimation.* United Kingdom: John Wiley & Sons, Ltd, 2016.

R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing, 2023.

Rahlf J, Breidenbach J, Solberg S *et al.* Comparison of four types of 3D data for timber volume estimation *Remote Sens Environ.* 2014;**155**: 325–33.

Rahlf J, Hauglin M, Astrup R *et al.* Timber volume estimation based on airborne laser scanning—Comparing the use of national forest inventory and forest management inventory data *Ann For Sci.* 2021;**78**:49.

Rao JN, Molina I. *Small Area Estimation.* New Jersey, USA and Canada: John Wiley & Sons, Inc, 2015.

Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis *J Comput Appl Math.* 1987;**20**:53–65.

Saegusa T, Sugasawa S, Lahiri P. Parametric bootstrap confidence intervals for the multivariate Fay–Herriot model *J Surv Stat Methodol.* 2022;**10**:115–30.

Saligkaras D, Papageorgiou VE. On the detection of patterns in electricity prices across European countries: an unsupervised machine learning approach *AIMS Energy.* 2022;**10**:1146–64.

Saligkaras D, Papageorgiou VE. *Seeking the Truth beyond the Data. An Unsupervised Machine Learning Approach.* 1st edition. AIP Publishing, 2023.

Särndal CE, Swensson B, Wretman JH. *Model Assisted Survey Sampling.* Stockholm, Sweden: Springer-Verlag, 1992, 1–694.

Temesgen H, Mauro F, Hudak AT *et al.* Using Fay–Herriot models and variable radius plot data to develop a stand-level inventory and update a prior inventory in the Western cascades, OR, United States *Front For Glob Change.* 2021;**4**:17.

Torkashvand E, Jozani MJ, Torabi M. Clustering in small area estimation with area level linear mixed models *J R Stat Soc A Stat Soc.* 2017;**180**:1253–79.

Ubaidillah A, Aziz SD. *saeBest: Selecting Auxiliary Variables in Small Area Estimation (SAE) Model*, 2021.

UFAMF. *Pertouli University Forest Management Plan 2019–2028.* Thessaloniki, Greece: University Forest Administration and Management Fund (UFAMF), 2018.

Ver Planck NR. *Hierarchical Bayesian Models for Small Area Estimation of Biophysical and Social Forestry Variables Ph.D.* Ann Arbor, United States: Michigan State University, 2018.

Ver Planck NR, Finley AO, Kershaw JA *et al.* Hierarchical Bayesian models for small area estimation of forest variables using LiDAR *Remote Sens Environ.* 2018;**204**:287–95.

Wolter KM. Generalized variance functions. In: Wolter KM, (ed.). *Introduction to Variance Estimation.* New York, NY: Springer New York, 2007, 272–97.

Young LJ, Chen L. *Using Small Area Estimation to Produce Official Statistics.* Stats, 2022.

Γεωργάκης Α. *Εκτιμήσεις μικρής έκτασης σε ανομήλικα δασικά οικοσυστήματα, με ομαδοποίηση των δασικών διαχειριστικών μονάδων και πολυμεταβλητά Fay-Herriot μοντέλα [Small Area Estimation in Uneven-Aged Forest Ecosystems, by Clustering Forest Management Units and Multivariate Fay-Herriot Models]Διδακτορική διατριβή.* [PhD Thesis],. Thessaloniki, Greece: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης [Aristotle University of Thessaloniki], 2024.