Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 2, 2025 (New Series), pp 291–316
http://www.ssca.org.in/journal

Small Area Estimation Technique in Forestry Sector with Special Emphasis on National Forestry Inventory: A Review

Kamal Pandey[1], Sunil Chandra[1], Girish Chandra[2] and Salman Khan[1]

[1]*Forest Survey of India, Kaulagarh Road, Dehradun, Uttarakhand-248195*
[2]*Department of Statistics, University of Delhi-110007*

Received: 26 September 2024; Revised: 16 April 2025; Accepted: 20 April 2025

Abstract

The National Forest Inventory (NFI) focuses on gathering and analyzing data regarding forest resources, including tree species, sizes, age, density, and distribution, to assess forest conditions and inform management decisions. While traditional forest inventories offer detailed statistics at larger scales, they often fail to provide precise estimates for smaller areas such as forest divisions and ecologically sensitive regions. Small Area Estimation (SAE) techniques bridge this gap by employing statistical methods to enhance estimates for subpopulations with limited sample sizes. These methods incorporate auxiliary data, such as remote sensing, to enhance the precision of estimates. SAE techniques, including regression models and Bayesian methods, have proven effective in forestry and other sectors, providing robust estimates crucial for localized forest management and policy-making. Substantial theoretical work has been carried out in these techniques; however, in comparison, less work is seen in the application aspect. It is seen that SAE is being used in biomass estimation, carbon sequestration estimation, biodiversity assessment, and various other domains of research. This review highlights the application of SAE in forestry, emphasizing its potential to deliver accurate estimates for small areas and supporting informed forest management practices. Different auxiliary variables required for getting higher precision are also discussed in detail.

Key words: Auxiliary information; Domain; Forestry parameters; National Forest Inventory; Small Area Estimation.

1. Introduction

Forest inventory is the process which involves gathering and analyzing data about forests and their resources, such as tree types, growth sizes, age, density, distribution, *etc.* (Vidal *et al.* (2016); Knoke *et al.* (2021)). Forest inventory includes field surveys, followed by validating through remote sensing and other concomitant techniques, and thereafter statistical analysis to ensure accurate and detailed information about forested areas (Wulder *et al.*

Corresponding Author: Kamal Pandey
Email: kamalpandey.iss@gmail.com

(2008); Pandey *et al.* (2019)). It is done systematically to evaluate forest conditions, forest increment, volume estimation, forest dynamics, and guide forest management decisions (Tomppo *et al.* (2008a); Corona *et al.* (2011)). These assessments help in understanding the forest ecosystems, biodiversity change, carbon storage from all five carbon pools (above ground, below ground, litter, deadwood and soil), timber availability and economic evaluation (Heym *et al.* (2021). Forest inventory provides details about the forestry statistics up to the major geographical scale, such as country, state or district level. It does not provide results at regional or small area levels such as forest division, range, compartment, section, ecological sensitive areas, *etc.* Importantly, these small areas are crucial segments of forest management, because they are the preliminary site where the management practices start. To solve this issue, researchers are utilizing different estimation techniques based on inventory and other auxiliary data of the concerned area (Breidenbach and Astrup (2012). One such method, popularly known as the Small Area Estimation (SAE) technique, which has been recently used in various sectors including forestry. In almost all the studies, it has been proven that SAE methods provide statistically precise estimates of the variables in small domains.

SAE technique refers to statistical methods used to estimate parameters for small sub-populations or geographic areas where traditional survey methods may not provide accurate or reliable estimates due to limited or even zero sample sizes (Rao (2003); Rao and Molina (2015); ADB (2020); Sugasawa and Kubokawa (2020)). However, Guldin (2021) mentioned and described Small Domain Estimation (SDE), which is a relatively similar conceptual term as SAE and further, it was suggested that SAE should be more appropriate to use over SDE due to its spatial domain. These techniques are crucial to obtain robust estimates for localized regions or domains within a larger population (Rao (2003)). In this context, SAE methodologies might find utility not only in scenarios characterized by a small geographic or temporal scope of the domain of interest with consequently limited sample sizes but also in circumstances necessitating enhanced precision in estimates across any such domain (Coulston *et al.* (2021)). In the realm of SAE, the specific level being assessed, whether it be stands, districts, or municipalities, and for which the estimation is desired, is termed a small area or domain (Breidenbach *et al.* (2018)). Small area models leverage information from both within and outside the domain of interest, as well as other supporting data, to augment the precision of parameter estimates (Lehtonen *et al.* (2003)). There are two distinct categories of SAE models exist: unit-level and area-level. Unit-level models pertain to estimators formulated at the level of sample units (such as field inventory plots in NFI), where auxiliary information is directly linked to these sample units (Rao and Molina (2015); Coulston *et al.* (2021)). Whereas, Area-level models generate estimates at aggregated geographic levels, such as administrative regions, forest range/beat/compartments *etc.*, where auxiliary information is linked to entire areas rather than individual sample units, making them suitable when unit-level data are limited or unavailable (Rao and Molina (2015)). More precisely, SAE methods typically involve combining information from both survey data and auxiliary information, such as census data or remote sensing data, to improve the precision of estimates for smaller areas (Jeong and Son (2009); Marchetti *et al.* (2015); Jiang and Rao (2020)). Recent focus on remote sensing-assisted forest inventories has shifted towards unit-level estimators for population parameter inference and associated uncertainty assessment. Numerous studies (McRoberts *et al.* (2014); Saarela *et al.* (2016); Chen *et al.* (2016); Mauro *et al.* (2016)) highlight this trend, while area-level estimators are also gaining attention (Goerndt *et al.* (2011); Magnussen *et al.* (2017)). However, only a few research reports are available for comparative analysis between unit- and area-level estimators.

SAE includes model-based estimation methods like regression models and Bayesian methods, as well as design-based methods like stratified random sampling and cluster sampling (Noble *et al.* (2002); Rao and Molina (2015)). The application of SAE, for providing reliable estimates for policy-making and decision support, has been widely seen in almost all the important sectors like public health, economics, forestry and environmental studies, demography analysis, social research, hydrological research, agricultural sector, economic analysis *etc.* (Jiang and Rao (2020); Chandra and Chandra (2015)). Within forest statistics, SAE studies focus on particular environmental data within limited geographic regions. Although national forest inventories (NFIs) provide precise assessments at larger scales, local forest officials encounter difficulties amid evolving market dynamics and financial limitations. SAE investigations act as a conduit between forest biometricians and policymakers, furnishing enhanced estimations tailored to smaller areas, thereby facilitating informed management choices (Guldin (2021)). The review serves as a crucial resource by synthesizing existing knowledge on Small Area Estimation techniques, highlighting their transformative potential in forestry research and management. By bridging the gap between traditional forest inventory methods and localized data requirements, this paper provides a roadmap for enhancing precision in small-area estimates, thereby supporting more informed policy-making, sustainable forest management, and biodiversity conservation. It also suggests the forestry researchers to utilize the SAE for generation of results based on SAE models.

This review paper is structured to provide a comprehensive exploration of SAE techniques and their significance in the forestry sector. It begins by presenting an overview of SAE methods, their classifications, and statistical approaches. It also includes detailed discussion on the diverse applications of SAE in forestry, such as biomass estimation, carbon sequestration, biodiversity assessment, and water resource management. A separate section delves into the integration of SAE techniques into national forest inventories, highlighting key methodologies, advancements, and regional case studies. Additionally, the paper examines the use of auxiliary data sources, including satellite imagery and LiDAR (Light Detection and Ranging), to enhance the precision of SAE estimates and discusses the challenges and limitations associated with these approaches. The review concludes by offering insights into future research opportunities and recommendations for the broader application of SAE in forestry and allied fields.

2. Search stratagem and data extraction

The information quoted in this review on ”Forest Resource Assessment and Use of SAE Technique in Forestry Sectors” was searched and obtained using databases such as online databases subscribed, Access e- Resources like PubMed, Science Direct, Scopus, Research Gate and Google Scholar. All these resources were searched with the keyword “SAE” along with “techniques”, “models”, “auxiliary variables”, “in forestry sectors”, “in national forest inventories”, “in NFI”, “forest resource assessment”, “Indian context” and “uses”. No language restrictions were imposed. Research papers were assessed for the information about the reviews, methodologies and results and discussions on SAE and its application in forestry, of that particular study. To enhance search sensitivity, specific search techniques such as the use of quotation marks, parentheses, and asterisks were employed to identify exact terms or expressions, as well as to capture related search terms and variations. Additionally, a manual screening of reference lists from selected studies and previous reviews was conducted to identify additional relevant publications. All the searched literature was studied thoroughly. The information from the full-text research papers and review articles of appropriate studies were extracted. Among the vast suggestions, published articles listed in the reference section were found in the databases that contain SAE techniques and their application in forestry sectors in general and the national forest inventory (NFI) in particular. The observed information in the included literature have been summarized in the present review.

3. Overview of SAE methods

SAE is an important technique for estimating the resources for smaller areas to reduce the standard error and coefficient of variations of the estimators under interest. The direct estimation of a small area yields large standard error and this eventually decreases the quality and accuracy of the estimates (Ghosh and Rao (1994)). The important factor and basis for the SAE technique is auxiliary variables. Breidenbach and Astrup (2012) mentioned that SAE techniques offer the solution to the statistical problem (low precision in estimation) if the correlated auxiliary variables are available. SAE methodologies, as delineated by Rao (2003), are typically classified into three categories: direct domain estimation, indirect domain estimation, and small area model-based domain estimation (Figure 1). Direct domain estimation focuses solely on sample data within a small area, particularly for estimating variance or mean squared error (MSE). Whereas, the indirect domain estimation objective is to enhance MSE by utilizing the estimates of the external domain through statistical modelling (Breidenbach *et al.* (2018)). Moreover, small area model-based domain estimation incorporates random effects to address unexplained variation between small areas, distinguishing between area-level and unit-level models. Area-level models, where auxiliary information pertains only to the small area level, can be advantageous in forest inventory contexts, particularly when sample plot locations are not geographically recorded (Goerndt *et al.* (2011); Breidenbach and Astrup (2012)). Auxiliary variables are the backbone of SAE and improve its precision and accuracy. In forestry, auxiliary variables can be either satellite (remote sensing) data, LiDAR (Airborne Laser Scanning) or other forestry parameters. Researchers have identified various auxiliary variables and have utilized them in estimating the characteristics of small domains. Like, Coulston *et al.* (2021) for their forest inventory and analysis program, used country-level timber products output and tree cover loss data along with SAE. Tree cover loss data was collected from global forest watch and Landsat imagery. Such timber product output summaries could be used for country and survey unit-level forest health and stand-level forest health models, aiding in understanding forest dynamics and management. Earlier, Breidenbach and Astrup (2012) and Breidenbach *et al.* (2018) in their experiment, considered the canopy height from photogrammetric models as an important auxiliary variable, revealing a strong linear correlation with forest biomass. The use of remote sensing data as an auxiliary variable is also mentioned by Goerndt *et al.* (2013) who compared different SAE models using 16 LANDSAT variables along with field data (tree density, basal area, cubic volume, quadratic mean diameter, and the average height of the top 100 trees/ha), land cover classification, tree cover, and elevation as the supportive variables. Thereafter, Breidenbach *et al.* (2018) used digital remote sensing data as an auxiliary variable in the area-level and unit-level estimate models under heteroscedasticity conditions. Airborne laser scanning is also proven auxiliary information, which enhances the quality and eventually leads to achieving greater precision in SAE. For instance, Green *et al.* (2020) have identified and mentioned the use of LiDAR (Light Detection and Ranging) height percentile and stand thinning status as auxiliary variables in SAE of Loblolly Pine-dominated areas in the Southern US. They informed that Area-level SAE models that incorporated both LiDAR height percentiles and stand-thinning status demonstrated substantial improvements in precision, outperforming models that relied solely on LiDAR data. Conclusively, Georgakis (2019) have also mentioned the importance and utilization of auxiliary variables in SAE. Thus, it is identified that the LiDAR canopy height model and Landsat imagery are the two most used auxiliary variables in SAE for national forest inventories.

3.1. Direct estimation methods

Direct estimate utilizes summary statistics derived from a domain or small area, to predict a particular attribute of interest associated with that area. It also includes sample means, sample proportions, variance, MSEs, and the products of sample means and the size of the domain's population, assuming the population size is known (Jiang and Rao (2020)). Direct estimates do not incorporate information from other domains or external sources when making predictions for a specific area using model-based or design-based estimation (Figure 1). According to Breidenbach *et al.* (2018), direct estimate (population mean of timber volume) is derived from the sample unit of each domain of a small population unit. Considering simple random sampling within each domain, the estimator was calculated using the Horvitz-Thompson estimation method (survey sampling method). In this method, the weights are assigned to sampled units based on their inclusion probabilities to estimate population parameters, say population total (Berger (1998)). These weights ensure that each unit's contribution to the estimate reflects its representation in the population, useful for complex sampling designs.

3.2. Indirect estimation methods

The indirect estimation approach is divided into indirect domain estimation and SAE Models (Figure 1). In the indirect domain estimation approach, the researcher (knowingly or unknowingly) uses the survey data to generate estimates using design-based or model-based estimations without keeping checks on optimum sample size. Thus, the generated estimates are not precise enough because of small or zero sample sizes. Therefore, to improve the precision of estimates, researcher may use additional information from related domains or external auxiliary data sources which is possible through Small Area Estimation (SAE) models. SAE models, under the indirect estimation, are applied when there is insufficient or no direct data for a specific domain, such as a small geographic area or demographic group. It integrates domain-specific data with auxiliary sources or utilizes information from similar domains to enhance accuracy. Common methods include synthetic estimation, SAE model-based approaches, and composite estimation techniques. For, example, in estimating health statistics for a district one may merge local survey data with broader regional or national trends.

3.2.1. SAE model-based estimation methods

As already discussed, Rao and Molina (2015) and Coulston *et al.* (2021) have mentioned that there are two types of SAE models, Unit-level and Area level (Figure 1). However, according to Jiang and Lahiri (2006) and and Jiang and Rao (2020), SAE models could be of three different types: area level model (Fay III and Herriot (1979)), unit level model (Battese *et al.* (1988)), and mixed logistic model (Jiang and Lahiri (2001)). The difference in classification arises due to variations in methodological perspectives and specific applications. The mixed logistic model, as mentioned by Jiang and Lahiri (2006)), is often considered a specialized extension of unit-level models, particularly for categorical data and binary response variables. In contrast, studies that classify SAE models into only two types typically focus on continuous variable estimation, where unit-level and area-level models sufficiently cover most applications. Thus, the distinction is not contradictory but rather a reflection of different contexts and modeling approaches used in SAE research. In the SAE model, mathematical equations are used to predict values for different small areas having survey data. These predictions are based on some factors that can influence the outcome. The model also includes random factors that account for differences that can't be explained with predictors. Sometimes, predictions contain errors which are independent of random factors. Although exact values for all errors are not available, they can be estimated accurately. These estimations help to improve the reliability and precision of the prediction (Jiang and Rao (2020)). The assumptions of these models are considered robust as they completely validate the data distribution. This enables inference methods like maximum likelihood (ML) or restricted maximum likelihood (REML) (Jiang and Nguyen (2021)). However, when these assumptions are violated, REML may not be reliable. Alternatives to ML or REML, such as consistent estimators of variance components (Prasad and Rao (1990)), can be used for computing the empirical best linear unbiased predictor (EBLUP). Nonetheless, Jiang and Rao (2020) mentioned that measures of uncertainty are more affected by distributional assumptions. Breidenbach and Astrup (2012) presented theoretical and empirical evidence supporting the superiority of the EBLUP estimator over simple random and generalized regression models. They mentioned that the EBLUP estimator adjusts the bias correction factor based on model variance and the number of observations within the domain, enhancing its accuracy compared to the generalized regression estimator. Also, the EBLUP estimator consistently exhibits smaller mean square errors than both the simple random sampling and generalized regression estimators, especially in domains with few sample plots. However, the validity of the EBLUP estimator more rely on the robust mixed-effects model.

Various researchers in forestry have worked on SAE and predicted the different parameters. They utilized different models supported with the required auxiliary variables and forest inventory data. For example, in Mexico, Reich and Aguirre-Bravo (2009) estimated the reliability and accuracy of synthetic and regression estimators by SAE estimation of forest stand. For their experiment, the modelling process involved using multiple linear regressions to capture large-scale spatial variation, while a tree-based stratified design was adopted to address small-scale variability linked to site-specific differences in forest stand structure. Various independent variables, including Landsat-7 *ETM+* bands, and climatic, and topographic data, were considered for the models. Later, the use of model-based approaches along with NFI by different countries rises following 2016 for smaller spatial estimates (provinces, forest compartments and cities) of forestry parameters (Guldin (2021)). As depicted by Breidenbach *et al.* (2018), the nested-error linking model was used for unit-level estimation and the mixed-effects model for area-level estimation. In their study, for unit-level modelling, the mixed effects model focusing timber volume per hectare calculated on sample plots. The model included the mean height observed on a sample plot and its square as explanatory variables, resulting in three different parameters. Also, REML was employed for parameter estimations, to depict the variations. The selection of the constant was optimized using Akaike's information criterion and Breusch-Pagan tests (Breusch and Pagan (1979)). For the area-level model, the response variable was the direct mean timber volume estimate of sample plots, while the sole explanatory variable was the stand-level mean of aerial photogrammetry mean height observed on all grid cells having parameter estimation using REML (Breidenbach *et al.* (2018)). They found that both the estimators (unit level and area level EBLUP) were indicating similar estimates with some differences in their standard errors.

Area-level and unit-level SAE models using relative error ratios were compared by Green *et al.* (2020) to enhance the accuracy of forest characteristic assessments in pine plantations. They discovered that area-level models incorporated with LiDAR height percentiles and stand thinning status were showing significant precision gains, especially when both types of auxiliary information were used. Whereas, unit-level models offer higher precision but have few limitations due to the complexity and the lack of insufficiency of precise spatial data. Furthermore, Guldin (2021) observed that in Switzerland's NFI, the design-based model-assisted approach is being employed for SAE, relying on probability samples for validity. Usually, a model-assisted method involves implementing a model to support estimation following after probability sampling. This approach ensures unbiased estimators of population parameters. Apart from the model-assisted approach, researchers compared different inferences in SAE. In particular, Brewer (2013) has compared design-based inference with randomization-based inference, contrasting it with model-based inference, which was associated with prediction-based inference. Gregoire *et al.* (2016) exemplified this with a forest inventory scenario, demonstrating how statistical regression can link airborne LiDAR height measurements to estimates of forest biomass. However, Pulkkinen and Zell (2019) mentioned that the Swiss NFI prefer the design-based and model-assisted approach as a method in SAE. Subsequently, Coulston *et al.* (2021) in their experiment, they examined different variables in forest health and stand-level forest health models, to estimate removal volumes categorized by total, species group, and merchantability class. The models integrated relevant connections between auxiliary data and the parameters under the experiment. Their experiment revealed that simpler models encountered fewer convergence challenges, with certain models relying on either a single explanatory variable or a combination of two variables and their interaction. It was also reported that their methodologies resulted in the development of 168 unique model parameterizations spanning diverse variables, methodologies, time frames, and spatial scopes in SAE. Conclusively it was advised that SAE techniques can meaningfully increase the temporal and spatial resolution of removal estimates, benefiting broad-scale inventory programs.

In model-based approaches, the regression models are the most used comprised of linear, generalized linear models, *etc.* Many researchers have contributed to developing regression models to be utilized in forestry. Figure 1 depicts different approaches of small area estimation and various tools and models for forestry application. Some other models are also paved their utilization is SAE such as the logistic mixed model which can be useful to improve the accuracy of small area proportion estimation. Authors like, Jeong and Son

(2009) introduced and described the logistics mixed model for the estimation of the small area. They considered the best linear unbiased predictor (BLUP) for the experiment. Parametric bootstrap and linear approximation were equated using the Monte Carlo approach to study the random variation across small areas. They found that linear approximation overestimated the Mean Squared Error (MSE) in comparison to bootstrap. Additionally, Reich and Aguirre-Bravo (2009) assessed the use of synthetic and regression methods in small-area analyses for estimating forest stand characteristics by taking forest inventory as auxiliary data in Jalisco, Mexico. Also, it was reported that the regression method, leveraging spatial models, showed better performance, with the accuracy of estimates heavily dependent on the spatial resolution and correlation of the models with the variable of interest. Moreover, Breidenbach and Astrup (2012) employed SAE in the forestry sector (NFI) in Norway by utilizing the photogrammetric model as an important auxiliary variable. They utilized the generalized regression estimation model for SAE and compared the outcome with simple random sampling and EBLUP estimator models. Also, the goodness-of-fit for the models was assessed through the root mean squared deviation.

Many researchers like Guldin (2021) provided a detailed description of the application of different models in the forestry sector of different countries and found that where there was not enough information from field surveys, countries relied on imputation models which combined various spatial datasets with the available field data, to create maps and calculate local estimates. It was also reported that few countries use a variety of techniques for these predictions. Some used supervised methods like maximum likelihood, discriminant analysis, and various types of regression. Others used nonparametric methods like k-NN and unsupervised approaches such as neural networks or clustering methods.

Researchers have also provided different models for SAE estimation due to its dependence on probability samples and statistical models for unbiased estimation. For example, the sampling in the Switzerland forest inventory program, had two phases: randomly selecting sample points and collecting auxiliary information, then using a simple random sample from these points to locate field plots for data collection. Further, the Swiss NFI follows this design-based Monte Carlo approach for sampling, with the fifth cycle continuing systematic measurements across the country (Pulkkinen and Zell (2019)). Similarly, Guldin (2021) hinted at new design-based SAE model-assisted estimators developed by incorporating non-exhaustive auxiliary data and different linear models. Thus, these estimators can compute the averages of model predictions using auxiliary data points and residuals from field plots within the area. In Norway, most of the research on SAE has been done with a special focus on national forest inventory *e.g.* Breidenbach and Astrup (2012) employed simple random sampling for SAE of sub-population in NFI of Norway. Conclusively, design-based models are important for the SAE point of view because simple random sampling is a type of direct estimator and utilizes the data of sample plots falling within the area of interest. Despite, various researchers having carried out growing stock estimation using remote sensing and GIS, still application of SAE in forestry research is found limited. The efforts are made to identify and provide a detailed review of various research on applications of the SAE technique in the forestry sector in general and NFI in particular.

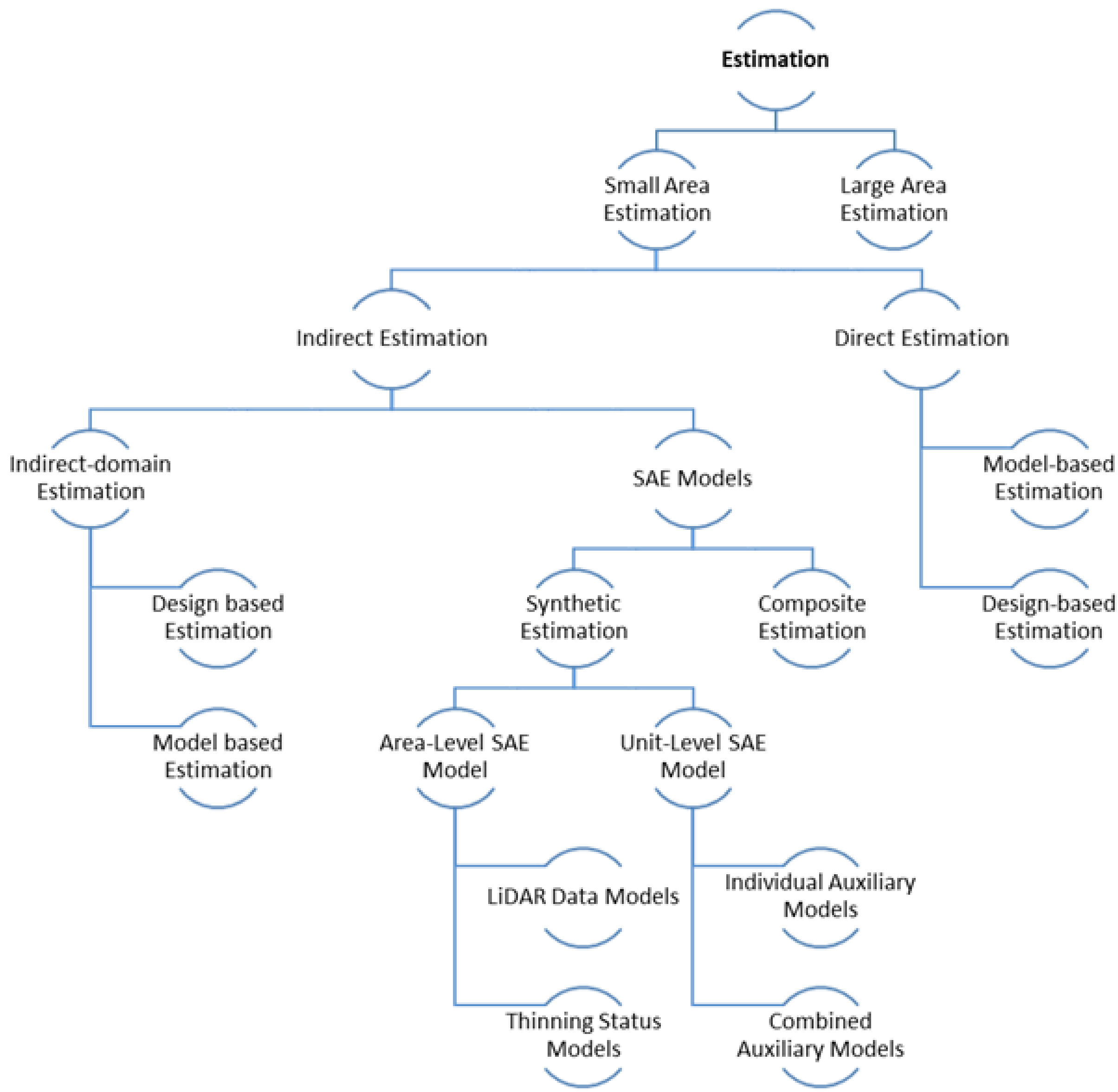

Figure 1: Small area estimation approaches and their utilization in forestry

4. Utilization of SAE techniques

Following section provides detail on application of small area estimation approaches in various sector including forestry, natural resource estimation (biomass, carbon sequestration, biodiversity assessment, water resource assessment), national forest inventories and others.

4.1. Forestry sector

SAE techniques in forestry utilize a combination of data sources, including satellite imagery, remote sensing data, and ground-based surveys, to produce detailed estimates of forest attributes such as tree density, biomass, and carbon sequestration rates at finer spatial resolutions (McRoberts (2012); Chandra and Chandra (2015)). These data sources are applied in the unit-level or area-level SAE models as auxiliary variables or additional information to generate precise estimates (Figure 1).

SAE is being used in various domains but in forestry, it has great importance as it reduces the cost and time. Guldin (2021) discussed that despite, forest inventories providing credible estimates at the national level; their spatial sampling often lacks the precision required for smaller scales or in areas with sparse or unevenly distributed forest resources (Chandra and Chandra (2015)). Worldwide, forest officials resort to auxiliary datasets like digital aerial photography to enhance inventory data accuracy through imputation models. These models employ parametric and non-parametric approaches and assign values to pixels corresponding to forest attributes. Subsequently, these imputed values are aggregated within defined small areas, employing algorithms to generate SAE estimates with acceptable accuracy (Goerndt *et al.* (2011); Breidenbach *et al.* (2018)). This bridging of statistical methods with practical forestry aims to facilitate sustainable forest management practices, support biodiversity conservation efforts, and assess the impacts of climate change on forest ecosystems. Additionally, SAE enables policymakers and stakeholders to make evidence-based decisions by providing accurate information on forest resources and their spatial distribution (Guldin (2021)).

Researchers started utilizing SAE in the estimation of forestry parameters. The majority of applications have concentrated on unit-level models (Coulston *et al.* (2021)). For instance, McRoberts (2012) employed unit-level methodologies to refine the precision of volume per acre estimates, leveraging Landsat Thematic Mapper imagery as auxiliary data. On a similar pattern, Mauro *et al.* (2017) incorporated LiDAR-based auxiliary information in their SAE approach, where LiDAR-derived metrics such as mean canopy height and biomass density were aggregated at the regional level. These variables served as key auxiliary data in area-level models, aligning with the classification in Figure (1), which places "LiDAR Data Models" under area-level SAE approaches. Similarly, Goerndt *et al.* (2013) utilized Landsat-derived spatial variables, such as land cover and tree cover, which, when aggregated over predefined areas, enhance area-level SAE predictions.

Likewise, Breidenbach and Astrup (2012) investigated the utilization of photogrammetric canopy heights within a unit-level approach to enhance the precision level of mean canopy height estimations. Their study compared unit-level and area-level SAE approaches, proposing methodologies to handle the inherent heteroscedasticity in forest parameters using photogrammetric methods.

On a similar trend, Reich and Aguirre-Bravo (2009) explored the efficacy of synthetic and regression methods for estimating forest stand characteristics in small geographic regions within Jalisco, Mexico. They interpret that despite, the synthetic method being user-friendly, its accuracy is contingent on meeting specific model assumptions. Whereas, the regression method employs spatial models, and offers improved accuracy and precision, though reliant on model quality and spatial resolution. They highlighted the importance of valid and precise data for informed decision-making in ecosystem sustainability. They concluded that the regression-based approach demonstrates notable improvements over synthetic methods, offering consistent estimates across spatial scales and aiding decision-making processes for resource management. Moreover, Goerndt *et al.* (2013) compared several SAE estimators in 12 counties of the northern Oregon Coast range. Multiple linear regression (MLR), gradient nearest neighbor imputation (GNN), most similar neighbor imputation (MSN), and five composite estimators combining MLR, MSN, and GNN with county-level direct estimates were tested and compared. The supportive data related to forest inventory such as tree density, basal area, cubic volume, quadratic mean diameter, and the average height of the top one hundred trees per hectare belonging to about 680 forest inventory plots were considered for testing with the SAE model. Besides, the tree characteristics, 16 Landsat variables, land cover classification, tree cover, and elevation were also utilized as auxiliary variables. They found that the composite estimators were better than any other method, by offering a high level of precision and minimum bias.

Thumaty *et al.* (2016) estimated Above Ground Biomass (AGB) for central Indian deciduous forests in Madhya Pradesh, India using 2010 ALOS-PALSAR L-band data and field-based AGB estimates through empirical models. Data from a survey of 415 sampling plots (0.1 ha each) collected in 2009-10 were used as auxiliary variables. They found that plot-level AGB estimates were modelled with PALSAR backscatter data, showing the HV backscatter relation ($R^2 = 0.51$) with field-based AGB estimates. Their study estimated the total AGB (AGB) for Madhya Pradesh forests to be substantial, with validation results showing a low root mean square error (RMSE), indicating robust model accuracy. Though, their experiment did not fully satisfy the SAE approach they have a similar model which is equally important. Aardt *et al.* (2006) worked on SAE for predicting timber volume under heteroscedasticity conditions using satellite imagery. Reich and Aguirre-Bravo (2009) investigated the reliability and accuracy of synthetic and regression estimators in small-area analyses, using forestry data from Jalisco's state of Western Mexico. Chandra and Chandra (2020) estimated total basal cover of trees, herbs and shrubs using SAE methods.

4.2. SAE in natural resource estimation

Effective management and planning of natural resources often necessitate precise estimates for small geographic regions. SAE methods have become increasingly important for delivering reliable estimates in these contexts, especially when direct survey estimates are inadequate due to limited sample sizes. SAE techniques have been effectively utilized in multiple areas of natural resource estimation such as:

4.2.1. Biomass estimation

Biomass is an integral part of any natural system including forests, oceans *etc.* Researchers have tried to estimate the biomass in forestry using SAE and successfully assessed the precise and accurate biomass. The choice of SAE model depends on whether auxiliary information is available at the unit level (*e.g.*, individual sample plots) or the area level (*e.g.*, forest compartments, districts). Unit-Level Model (Nested-Error Regression Model or EBLUP) or Area-Level Model (Fay III and Herriot (1979)) may be utilized for biomass estimation (Figure 1). For illustration, Breidenbach *et al.* (2018) estimated the forest biomass of different areas in Norway. For this, they considered the average height of the tree canopy (acquired from the canopy height model) as an auxiliary variable. Further, the same was compared with three different SAE techniques, *viz.* simple random sample mean, generalized regression and unit level EBLUP. They concluded that the SAE estimation of biomass through the SAE technique by unit-level EBLUP estimator was significantly more precise than the other three estimators. Similarly, Green *et al.* (2020) estimated the biomass of Loblolly pine in USA using the SAE model at both area level and unit level and found that auxiliary variable LiDAR and stand thinning status can generate precise estimates of the forest biomass and volume. In the USA, Gaines III and Affleck (2021) suggested the SAE model for postfire forest regeneration. However, the MSE estimates exhibited frequent negativity and variability, which suggests the limited operational utility of this model.

4.2.2. Carbon sequestration estimation

Utilization of SAE in carbon sequestration was reported by Guldin (2021) who reviewed and informed that SAE application in forestry at the global level has increased since 2010. Carbon sequestration can also be expressed in terms of AGB (Above Ground Biomass). It was reported that the estimation of AGB volume has gained significant attention in SAEs, surpassing the earlier focus on timber volume estimations. This rise in SAE application in carbon assessment can be attributed to recent global policies like greenhouse gas reporting and REDD+ (Reducing Emission from Deforestation and Degradation+), which emphasize the importance of monitoring forest carbon stocks and fluxes (Bhattarai *et al.* (2015)). Additionally, there is growing interest from both public and governmental sectors in evaluating the capacity of forests to sequester carbon at both national and regional scales. In SAE models, information such as soil samples, land use data, and environmental variables are incorporated to estimate the soil carbon content for small areas of interest (Prisley *et al.* (2021) Stanke *et al.* (2022)). As described in Figure (1) both unit-level and area-level model may be employed for carbon sequestration estimation through LiDAR models or auxiliary models (Battese *et al.* (1988)).

4.2.3. Biodiversity assessment

Biodiversity includes the variety of life forms within the ecosystems, which is very essential for sustainable ecosystem functioning and resilience. SAE methods overcome the limitation of a small area by integrating information from various origins and accounting for spatial variability, thus allowing for more accurate biodiversity estimates at smaller resolutions. Researchers such as Packalén and Maltamo (2007) mentioned the use of different canopy height models for the identification of different forest species through satellite imagery; Breidenbach *et al.* (2018) highlighted that these imageries can be utilized in biodiversity assessment in small areas of a survey unit. Prisley *et al.* (2021) highlighted that SAE techniques such as machine learning, spatial interpolation and hierarchical Bayesian models are useful in various fields of biodiversity such as species distributional modelling, community composition estimations, rare species and endemic assessments. Both Fay-Herriot Model (unit-level) and Nested-error regression model (area-level) may be utilized for biodiversity assessment if species richness, abundance, frequencies and other variables are available as auxiliary variable (Figure 1).

4.2.4. Water resource assessment

SAE can also be utilized in estimating groundwater levels and surface water availability by integrating survey data with hydrological models and remote sensing data (Ferreira *et al.* (2022)). By leveraging the Fay-Herriot model, their study provided district-specific estimates for indicators such as household hunger, access to medical supplies, and piped water availability. Earlier, Opsomer *et al.* (2003) had applied small area estimation (SAE) techniques to evaluate soil erosion within the Rathbun Lake Watershed in Iowa, USA. They utilized survey data with auxiliary information to enhance the precision of erosion estimates at the sub-watershed level. The methodology involved integrating direct survey estimates with model-based predictions, accounting for spatial correlations and measurement errors. This approach yielded more reliable and detailed erosion assessments, facilitating targeted soil conservation efforts and informed watershed management decisions.

Later, Nguyen (2023) reported the use of SAE in water hygiene and sanitation coverage in Vietnam. Study integrates survey data with auxiliary information to produce disaggregated estimates of water, sanitation, and hygiene (WASH) coverage at provincial and district levels in Vietnam. This approach involved combining data from national household surveys with additional sources, such as census data and administrative records through regression, to enhance the precision of estimates in smaller geographic areas. By applying small area estimation techniques, the study was able to construct detailed geographic maps of WASH coverage, revealing significant disparities, particularly in poorer provinces and districts, which exhibited notably lower access to safely managed sanitation and water services.

4.3. SAE in national forest inventories

National Forest Inventory (NFI) is the systematic program designed to collect, analyze, and report data on forest resources within a whole area of any specific region, state or country to meet international reporting requirements for agreements such as the Kyoto Protocol and the United Nation's Food and Agriculture Organization. NFIs also fulfil intergovernmental mandates, as seen in initiatives like the Montreal Process, the Ministerial Conference on the Protection of Forests in Europe, and COST Action *E43* (Tomppo *et al.* (2008b); McRoberts (2010); Georgakis (2019)). This structure ensures that NFIs offer critical data for domestic management and international compliance, aiding sustainable forest management and adherence to global environmental commitments. As NFI is large-scale work in which different forestry parameters such as the composition of tree species, their size and age distribution, density, volume, biomass, carbon sequestration and biodiversity are measured and calculated to monitor the health and changes in forest conditions and to evaluate forest dynamics, there is greater cost and time involved in it. So, to overcome this, SAE, become a helpful technique which is an economical and time-saving method and also provides greater accuracy and precision in NFI estimates for small domains. For NFI estimates, Unit-level combined auxiliary model (if data is available at individual plots) or composite models (if sample size is very low) or area level model (if field data is at the area level like range/compartment/beat/others) can be employed (Figure 1). Various researchers belonging to the forest services and universities worldwide have developed precise estimates for NFI of the concerned regional forests area using SAE methodology.

In Norway, Breidenbach and Astrup (2012) estimated the mean above-ground forest biomass of the sub-population using SAE techniques by employing auxiliary variables, specifically canopy height from a photogrammetric model, correlated with other variables of interest. They recommended that the unit-level EBLUP yielded the most stable and accurate estimates.

Likewise, Næsset (2014) highlighted that operational forest management inventories now often use individual tree crown approaches and area-based approaches because both the variable of interest and the explanatory variables are accessible at the level of individual population units, such as geographically located trees or field sample plots. On the other hand, domain-level approaches have predominantly found application in research investigations, as demonstrated by studies conducted by Aardt *et al.* (2006) and Goerndt *et al.* (2011). Chandra and Chandra (2015) obtained the small area estimates for the three forestry parameters frequency, density, and total basal cover of trees, shrubs, and herbs for the state of Maharashtra in India. The auxiliary data, the percentage of forest cover at the small area level, was used for the purpose. They showed that the forest type-wise estimates of all three parameters are reliable as compared to direct estimates. Moreover, Mauro *et al.* (2017) conducted a comparison in an airborne laser scanning supported forest inventory. These findings underscore the importance of incorporating diverse auxiliary data sources in forest inventories to enhance the reliability of estimates.

Similarly, Barrett *et al.* (2016) provided an overview of the practical application of remotely sensed data in NFIs, drawing on insights from experts representing 45 countries, which collectively cover 65 percent of global forest areas. Their analysis revealed the extensive use of remotely sensed data from various sensors to refine the estimates across numerous forestry parameters in smaller areas. However, they highlighted the need for estimation through more effective integration of remotely sensed data with field inventories dataset. This comprehensive approach ensures that forest inventories remain accurate and reflective of current forest conditions. Various researchers in Spain have experimented with combining NFI and LiDAR data, yielding significant results. Condés and McRoberts (2017) focused on updating NFI-based estimates by using models to predict annual plot-level volume change and assess uncertainties. Novo-Fernández *et al.* (2019) researched to generate a fine-resolution database of forest yield variables. They described estimation methods that combined Spanish NFI data with Airborne Laser Scanning (ALS) data to predict the volume of growing stock for three main commercial tree species found in northwestern Spain. Additionally, Durante *et al.* (2019) studied the vast expanse of approx. 2.8 million acres in southwestern Spain. They integrated field plot data from the Spanish NFI, high-precision ALS, and bio-geophysical spectral variables sourced from MODIS. They explored a two-stage upscaling approach for biomass estimation, leveraging ALS data calibrated with NFI field plots to create regional biomass maps. Esteban *et al.* (2019) described a model-assisted inference approach using random forests, comparing different bootstrap estimators and constructing change maps. In a similar experimental format during 2012, Breidenbach and Astrup (2012) continued working on NFI and reported that Norway employs permanent sample plots for the estimation of forestry parameters, yet precise estimates for small sub-populations could not be achieved due to very limited data. In their study, they evaluated simple random sampling, generalized regression, and unit-level EBLUP for estimating above-ground forest biomass across various locations. They discovered that EBLUP produced more stable estimates in areas with numerous sample plots. Ultimately, they emphasized that increasing the sample size enhances estimation precision.

Similarly, Breidenbach *et al.* (2018) carried out a forest survey in parts of Vestfold county in southeastern Norway, utilizing complete digital aerial photogrammetry data. The survey, based on Norway's NFI methodology, aimed to estimate the mean timber volume per hectare for stands sampled using fixed-area plots. Trees were recorded, and volume was predicted from diameter and height using standard models. Stand samples were collected from areas with available forest management inventories in Norway, with explanatory variables including aerial photogrammetry heights calculated for sample plots and grid cells.

Germany is also not lagging in developing their NFI using SAE techniques. Like, Hill *et al.* (2018) mentioned a double-sampling extension for the German NFI to create design-based SAEs at the forest district level. It was suggested to utilize airborne laser scan (LiDAR) based derived canopy height models and tree species classification maps as auxiliary data with regression models to predict timber volumes (Mandallaz *et al.* (2013)). In addition to this, Wagner *et al.* (2017) conducted a study in Germany aimed at developing cost-efficient SAE methods to address the limitations of traditional forest inventories in a dense forest which require forest management information at various administrative levels. They investigated and tested different design-based regression estimators, considering laser data of varying temporal resolution. Subsequently, they found a reduction in the significant error of estimate at the district level and suggested more research is required for sub-district-level estimations. Further, the SAE method expanded to other German states for their NFI data estimates.

In Nordic countries, Kangas *et al.* (2018) reviewed the scientific advancements in NFI, focusing on the integration of remotely sensed data to boost the accuracy of NFI and Forest Management Inventory (FMI) datasets and to minimize uncertainties in parameter estimations at both national and small area levels. These FMIs are important sources to enhance the estimate precision thus, various other researchers started to utilize FMIs in their experiments. For example, Jiang and Rao (2020) explored the utilization of remotely sensed data to complement NFI and FMI datasets, aiming to improve the precision of parameter estimates at a small area level.

Furthermore, Georgakis (2019) provided a detailed review of SAE in forestry and explained that SAE methods, including synthetic estimation and the Fay-Herriot model, addressed the scarcity of terrestrial sample data and improved accuracy in the NFI. His review highlighted how these statistical techniques are crucial for enhancing the reliability of forest resource estimates, especially in regions where terrestrial data is limited.

In their study of the application of SAE in the NFI of Switzerland, Pulkkinen and Zell (2019) explored how the Swiss forest inventory utilizes a design-based Monte Carlo approach for sampling. This method involves sampling an infinite population of points within a designated region of interest to estimate spatial means of tree population totals. Field data collection employed a simple random technique. Estimation was carried out using a specific model incorporating auxiliary data and a linear model based on probability sampling, ensuring an unbiased nature in the estimation process. Whereas, Breidenbach *et al.* (2021) utilized Sentinel-2 mosaics in conjunction with NFI data to develop models and cartographic representations of various Norwegian conifer types. These models were then leveraged to generate species-specific distribution maps tailored to smaller geographic areas, such as municipalities. Further, Astrup *et al.* (2019) detailed the integration of photogrammetric point cloud data with NFI point cloud data to produce a raster map featuring selected modelled attributes applicable to forest management inventories. The pair of studies by Haakana *et al.* (2020) investigated post-stratification as an alternative method for utilizing auxiliary information to estimate parameters for municipalities based on Finland's NFI data.

In India, Chandra and Chandra (2020) experimented with the SAE approach to generate estimates of total basal cover *(m2/ha)* for trees, shrubs, and herbs in small areas of Maharashtra, India. Seven different forest types were considered as small areas for the study. Survey data (nested quadrate of $10\ m \times 10\ m$, $3\ m \times 3\ m$, and $1\ m \times 1\ m$ for tree, shrub, and herb layers, respectively) of the year 2011-12 were utilized as one of the important variables for the estimation. India's State of Forest Report (ISFR) depicting NFI and remote sensing-based forest cover and forest change data were utilized as auxiliary information. Their results indicated that estimates of total basal cover for trees, shrubs, and herbs, generated using the SAE approach, were reliable and precise compared to direct survey estimates.

In the Southeastern US, Coulston *et al.* (2021) in their experiment, utilized SAE methods in the NFI, considering Landsat-based tree cover change information and mill survey data as auxiliary variables. In their study, more than 35,000 permanent sample plots across forest and non-forest areas were assessed. The result suggested that improvements in the precision of estimates can be achieved when NFI data is measured at considerably fine spatial and temporal scales. They observed that the precision of estimates in tree species groups and size classes improves when SAE methods are used compared to design-based assessments. They utilized the Forest Inventory and Analysis (FIA) program's rotating panel design for data collection and followed the recommendation of Bechtold and Patterson (2005), where removal estimates were calculated based on midpoint assumptions within the re-measurement period. Parameters of interest included total removals, hardwood and softwood removals, and removals based on tree size categories.

Strîmbu *et al.* (2021) addressed inconsistency in aggregating parameter estimates for SAEs in Norway. They proposed a method to align model-based and model-assisted estimators, ensuring consistency in AGB estimates. The evaluation was conducted through simulated sampling in a 50 *km2* forest area, adjusting estimates within confidence intervals using heuristic optimization. Artificial forest populations were generated from airborne laser scanning data. Comparing adjusted AGB estimators with unadjusted ones showed minimal bias introduction, with decreased RMSE for stand-level and property-level estimators, showcasing the potential for consistency in complex estimator systems.

For NFI estimates in France, Vega *et al.* (2021) experimented with and introduced a distinct estimation algorithm aiming to balance the estimate precision in the Oak-dominated areas. The algorithm relies on NFI field data along with auxiliary information such as forest cover maps and canopy height models to achieve the desired precision in forest characteristics estimates. However, Guldin (2021) mentioned that NFI in France is based upon a two-phase sampling strategy: Photo-interpretation of the plot for land use and land cover identification followed by drawing of a sample based on land cover. Previously, Fortin (2020) investigated the challenge of annual sampling of NFI plots and its impact on variance over the period. To address this, older plot measurements were updated by a forest growth model. However, that introduced uncertainty from both sample design and growth model and suggested to use of a forest growth model to improve the estimate precision in NFI. Likewise, Irulappa-Pillai-Vijayakumar *et al.* (2019) utilized 3D variables from photogrammetric estimated canopy height models and auxiliary attributes (forest type maps, variables related to vegetation and characteristics, satellite maps, *etc.*) to enhance French NFI data precision in the forested region. Their experiment demonstrated a significant reduction in errors, specifically in volume estimates, suggesting potential for improved forest attribute estimation in smaller areas using SAE. Dettmann *et al.* (2022) and and Wilson *et al.* (2023) reviewed the extensive use of SAE in forest inventory and management, including the terminology, methods, concerns, data sources, research findings, challenges, and future opportunities. They also informed about the various methodologies such as direct, indirect, and composite estimation within design-based and model-based frameworks of SAE and are supplemented with remote sensing and geospatial data, to enhance precision in small domains, avoiding instability from small sample sizes.

All these research studies constantly show improved precision compared to direct estimates based solely on field data, highlighting the potential of SAE in national forest inventory for future research and development.

5. Data sources and challenges-remote sensing data

In SAE, the supported data and auxiliary data are important factors which determine the precision and accuracy of the model. In the case of NFI estimation, these data can be satellite imagery, airborne laser scanning (LiDAR) data or any other data. The description of these are as follows:

5.1. Satellite imagery

Satellite-borne sensor data such as LANDSAT (Land Satellite) is an important approach for integration with SAE for utilization in the small domain forestry estimation. Various other spatial data like DEM (Digital Elevation Model), soil maps, NDVI (Normalized difference Vegetation Index), watershed maps, forestry maps, geology maps, species maps, forest boundary maps, *etc.* are being considered in forestry research including SAE. These spatial data are used as auxiliary variable in SAE models to generate estimates for various theme such as biodiversity, carbon, forest volume/growing stock *etc.* (Figure 1). Pioneer work is done by Tomppo and Katila (1991) by combining satellite data with NFI data. Furthermore, Reich and Aguirre-Bravo (2009) for their experiment in Mexico's NFI utilized Landsat-7 imagery having *ETM+* along with climatic and topographical information. Different bands of satellite data were identified for specific uses *viz.* Goerndt *et al.* (2013) utilized Landsat 5 TM (Thematic Mapper) data focusing on six bands *(B1-B5, B7)* of *30x30 m* spatial resolution. Normalized Difference Vegetation Index (NDVI), calculated from bands 3 and 4, and various band ratios *(B4/B3, B5/B4, B7/B5)* were used to describe forest characteristics. After identification, the imagery was processed for geometric rectification, radiometric correction and several transformations, including tasselled-cap (TC) transformation, optimizing brightness, greenness, and wetness. These transformations aid in visualizing the forest structure and health in a better way. However, researchers like Hill *et al.* (2018) had taken tree species maps as auxiliary variables for their estimation in Germany's NFI. Other satellite imagery like Moderate Resolution Imaging Spectroradiometer (MODIS) is also commonly used imagery in NFI. In particular, Durante *et al.* (2019) reported the use of MODIS data for bio-geophysical spectral attributes. Likewise, Astrup *et al.* (2019) utilized photogrammetric point cloud data along with NFI data to generate raster maps depicting forestry parameters/attributes. High-resolution satellite data was also suggested to be utilized by Breidenbach *et al.* (2021)). hey employed Sentinel-2 mosaics in conjunction with NFI for modelling conifers in Norway. The use of satellite data in forestry estimation was also suggested by Guldin (2021)). They mentioned the worldwide practice of utilising satellite data or aerial photography as an auxiliary variable for the estimation of timber volume, tree cover and land use change identification and classification.

5.2. LiDAR (Light Detection and Ranging) data

Earlier, Airborne Laser Scanning (ALS) data such as LiDAR was found to be strongly useful in forestry but due to higher cost and availability of free satellite data, it is less commonly used nowadays. Various LiDAR models can be utilized for SAE based estimation along with other auxiliary variable (Figure 1). In the review, Guldin (2021) informed that Switzerland is working on SAE in the forestry sector for their country and utilizing the LiDAR-based terrain model. They are utilizing the vegetation height and other forestry parameters as auxiliary variables. This system estimates various parameters for cantons, forest districts, and municipalities in Switzerland. Also, Gregoire *et al.* (2016) defined the process of regression modelling using the point cloud data of airborne LiDAR-based height measurements to estimate forest biomass. Various researchers have supported the argument that the airborne laser scans integrated with NFI data are more useful and economical when compared with the traditional inventory (Haakana *et al.* (2020); Rahlf *et al.* (2021); Strîmbu *et al.* (2021)). Other important researchers include Mandallaz (2013) who tested the SAE approach based on the NFI and auxiliary data such as LiDAR and floral information. Also, Wagner *et al.* (2017) applied SAE methods along with laser data for estimating the timber in Germany's selective forest. Similar to this, Hill *et al.* (2018) utilized airborne laser scan data-based canopy height for SAE modelling. Furthermore, Durante *et al.* (2019) constructed an experiment to measure biomass through the Laser data and created a model based on it. Additionally, Novo-Fernández *et al.* (2019) estimated the yield of the plantation forest using laser data in Spain. Alongside, Breidenbach *et al.* (2018) highlighted the SAE status in Norway and focused on LiDAR-based SAE estimation. Apart from this, Green *et al.* (2020) mentioned that LiDAR data are very important as an auxiliary variable despite their age and utilized 4-5-year-old LiDAR data and projected this with current year forest inventory data. However, if the LiDAR data are of the same year of forest inventory then the result would be more precise as the chances of variation decrease significantly. Additionally, Geogarkis informed that SAE using LiDAR data shows superior results compared to those based on photogrammetry. For extensive areas like whole counties, remote sensing data from sources like Landsat, which are freely available, can provide comprehensive coverage. However, due to the high costs and sometimes limited availability of LiDAR data, its use is often restricted to smaller areas of interest, such as forest districts, where its higher accuracy can be better justified (Goerndt (2010)). This highlights the trade-off between cost and precision when choosing appropriate data sources for SAE in different spatial contexts.

6. Some work in other sectors and future recommendations

SAE is already being considered in various sectors including agriculture, fisheries, medical, big-data analytics and others. According to Chandra and Chandra (2015), early developments in SAE for crop yield began in the late 1960s with Panse *et al.* (1966) who used a double sampling approach at the Block level. Later, Srivastava *et al.* (1999) and Sisodia and Chandra (2012) applied a synthetic method for Block-level crop estimation, utilizing extensive auxiliary information from crop-cutting experiments. Singh and Goel (2000) further refined crop yield estimation using remote sensing data. In the context of India's National Agricultural Insurance Scheme, alternative approaches for scaling down estimates to the Gram Panchayat level were explored by Sharma *et al.* (2004) and Sud *et al.* (2001)). In contrast, Sud *et al.* (2011) also applied a random effect model to account for small area dissimilarities. Beyond forestry, Srivastava *et al.* (2007) applied SAE techniques to estimate district-level economic parameters, such as loan amounts and poverty measures, in Uttar Pradesh using survey data. These studies demonstrate the precision and representativeness of SAE in various agricultural and socio-economic contexts. Additionally, Marchetti *et al.* (2015) mentioned the SAE-based estimator using big data sources. This is also validated by Kordos (2016) who advocated the utilization of big data which provides opportunities to understand complex socioeconomic phenomena such as poverty and resource distribution. When combined with SAE methods, big data offers faster access to auxiliary variables compared to traditional data sources. Despite advantages, challenges arise in ensuring data quality and representativeness, especially concerning privacy and self-selection biases.

Methodological advancements are necessary to integrate this with survey data effectively, revolutionizing scientific research and policymaking by providing precise, continuous insights into societal well-being (Rao and Molina (2015)). Using SAE, Pramanik *et al.* (2015) predicted 2011 vaccination coverage rates for 26 states not covered by the 2010-11 Annual Health Survey (AHS) in India. The model-based estimates provided almost similar results as to AHS results which ensures the reliability of the SAE method. Similarly, Jiang and Rao (2020) discussed the SAE models being utilized in the field of medical science. SAE is also being utilized in the happiness index through social media. For example, Aziz and Ubaidillah (2021) studied SAE models, EBLUP Fay-Herriot and Error Measurement, utilizing auxiliary variables from Twitter data. Conclusively, the SAE technique is an important statistical method which helps in estimating precise and accurate information for small areas and eventually reduces time and cost and increases reliability. Various models can be utilized for SAE depending upon the need and budgets. Auxiliary variables are also very important for generating precise estimates. SAE has shown enormous uses in the fields of forestry, medical science, and agriculture and its application is still under use in various other fields.

7. Conclusion

To conclude, Small Area Estimation (SAE) techniques offer significant potential to transform forestry practices by overcoming the challenges associated with traditional forest inventories. SAE provides accurate and dependable estimates at finer spatial scales, such as forest divisions or ecologically sensitive regions, which are essential for informed decision-making in forest management and conservation. Its flexibility is evident through diverse applications, ranging from estimating biomass and carbon storage to assessing biodiversity and water resources. By incorporating advanced tools like satellite imagery, LiDAR, and remote sensing, SAE not only improves the precision of forest assessments but also offers a practical, cost-effective solution. This review brings together key developments, methodologies, and examples, shedding light on how SAE can advance sustainable forestry. By connecting theoretical advancements with real-world applications, the paper emphasizes the value of SAE as a vital tool for enhancing forest management strategies on a global scale.

Acknowledgement

The authors express their gratitude to the Director General of the Forest Survey of India (FSI), Dehradun, Uttarakhand, India, for his invaluable support and guidance. Special appreciation is extended to the Director, Forest Research Institute, Dehradun, for providing the necessary facilities and permissions during this work. The authors are also deeply indebted to Sh. S.P. Singh, IFS (Retd.), Senior Consultant at FSI Dehradun, for his unwavering support and continuous encouragement throughout the writing process.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Aardt, J. A. V., Wynne, R. H., and Oderwald, R. G. (2006). Forest volume and biomass estimation using small-footprint LiDAR-distributional parameters on a per-segment basis. *Forest Science*, **52**, 636–649.

ADB (2020). *Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistics Offices*. Asian Development Bank.

Astrup, R., Rahlf, J., Bjørkelo, K., Debella-Gilo, M., Gjertsen, A. K., and Breidenbach, J. (2019). Forest information at multiple scales: development, evaluation and application of the norwegian forest resources map sr16. *Scandinavian Journal of Forest Research*, **34**, 484–496.

Aziz, S. D. and Ubaidillah, A. (2021). Big data for small area estimation: Happiness index with twitter data. In *Proceedings: The International Conference on Data Science and Official Statistics*.

Barrett, F., McRoberts, R. E., Tomppo, E., Cienciala, E., and Waser, L. T. (2016). A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. *Remote Sensing of Environment*, **174**, 279–289.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

Bechtold, W. B. and Patterson, P. L. (2005). *The Enhanced Forest Inventory and Analysis Program: National Sampling Design and Estimation Procedures*. U.S. Department of Agriculture, Forest Service, Southern Research Station.

Berger, Y. G. (1998). Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference*, **67**, 209–226.

Bhattarai, T., Skutsch, M., Midmore, D., and Shrestha, H. L. (2015). Carbon measurement: an overview of forest carbon estimation methods and the role of geographical information system and remote sensing techniques for REDD+ implementation. *Journal of Forest and Livelihood*, **13**, 69–86.

Breidenbach, J. and Astrup, R. (2012). Small area estimation of forest attributes in the Norwegian national forest inventory. *European Journal of Forest Research*, **131**, 1255–1267.

Breidenbach, J., Magnussen, S., Rahlf, J., and Astrup, R. (2018). Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data. *Remote Sensing of Environment*, **212**, 199–211.

Breidenbach, J., Waser, L. T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S., and Astrup, R. (2021). National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. *Canadian Journal of Forest Research*, **51**, 365–379.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294.

Brewer, K. (2013). Three controversies in the history of survey sampling. *Survey Methodology*, **39**, 249–263.

Chandra, H. and Chandra, G. (2015). An overview of small area estimation techniques. In Chandra, G., Nautiyal, R., Chandra, H., Roychoudhury, N., and Mohammad, N., editors, *Statistics in Forestry: Methods and Applications*. TFRI Jabalpur, India.

Chandra, H. and Chandra, G. (2020). Small area estimation for total basal cover in the state of Maharashtra in India. In G. Chandra, R. N. and Chandra, H., editors, *Statistical Methods and Applications in Forestry and Environmental Sciences*. Springer.

Chen, Q., McRoberts, R. E., Wang, C., and Radtke, P. J. (2016). Forest aboveground biomass mapping and estimation across multiple spatial scales using model-based inference. *Remote Sensing of Environment*, **184**, 350–360.

Condés, S. and McRoberts, R. E. (2017). Updating national forest inventory estimates of growing stock volume using hybrid inference. *Forest Ecology and Management*, **400**, 48–57.

Corona, P., Chirici., G., McRoberts, R. E., Winter, S., and A., B. (2011). Contribution of large-scale forest inventories to biodiversity assessment and monitoring. *Forest Ecology and Management*, **262**, 2061–2069.

Coulston, J. W., Green, P. C., Radtke, P. J., Prisley, S. P., Brooks, E. B., Thomas, V. A., Wynne, R. H., and Burkhart, H. E. (2021). Enhancing the precision of broad-scale forestland removals estimates with small area estimation techniques. *Forestry*, **94**, 427–441.

Dettmann, G. T., Radtke, P. J., Coulston, J. W., Green, P. C., Wilson, B. T., and Moisen, G. G. (2022). Review and synthesis of estimation strategies to meet small area needs in forest inventory. *Frontiers in Forests and Global Change*, **5**, 1–16.

Durante, P., Martín-Alcón, S., Gil-Tena, A., Algeet, N., Tomé, J. L., Recuero, L., Palacios-Orueta, A., and Oyonarte, C. (2019). Improving aboveground forest biomass maps: from high-resolution to national scale. *Remote Sensing*, **11**, 1–20.

Esteban, J., McRoberts, R. E., Fernández-Landa, A., Tomé, J. L., and Næsset, E. (2019). Estimating forest volume and biomass and their changes using random forests and remotely sensed data. *Remote Sensing*, **11**, 1–21.

Fay III, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Ferreira, T., Stone, W., Vercuil, E., Lourens, M., Made, N., and Madonsela, T. (2022). Small area estimation for south african resource distribution and policy impacts during COVID-19. *Open Research Africa*, **5**, 1–15.

Fortin, M. (2020). Updating plots to improve the precision of small-area estimates: the example of the Lorraine region, France. *Canadian Journal of Forest Research*, **50**, 648–658.

Gaines III, G. C. and Affleck, D. L. (2021). Small area estimation of postfire tree density using continuous forest inventory data. *Frontiers in Forests and Global Change*, **4**, 761509.

Georgakis, A. (2019). Small area estimation in forest inventories. In *Proceedings of the Seventh International Conference on Environmental Management, Engineering, Planning & Economics*.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55–76.

Goerndt, M. E. (2010). *Comparison and analysis of small area estimation methods for improving estimates of selected forest attributes*. PhD thesis, Oregon State University. Doctoral Dissertation, Doctor of Philosophy in Forest Resources.

Goerndt, M. E., Monleon, V. J., and Temesgen, H. (2011). A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. *Canadian Journal of Forest Research*, **41**, 1189–1201.

Goerndt, M. E., Monleon, V. J., and Temesgen, H. (2013). Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information. *Forest Science*, **59**, 536–548.

Green, P. C., Burkhart, H. E., Coulston, J. W., and Radtke, P. J. (2020). A novel application of small area estimation in loblolly pine forest inventory. *Forestry: An International Journal of Forest Research*, **93**, 444–457.

Gregoire, T. G., Næsset, E., McRoberts, R. E., Ståhl, G., Andersen, H. E., Gobakken, T., Ene, L., and Nelson, R. (2016). Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sensing of Environment*, **173**, 98–108.

Guldin, R. W. (2021). A systematic review of small domain estimation research in forestry during the twenty-first century from outside the United States. *Frontiers in Forests and Global Change*, **4**, 1–15.

Haakana, H., Heikkinen, J., Katila, M., and Kangas, A. (2020). Precision of exogenous post-stratification in small-area estimation based on a continuous national forest inventory. *Canadian Journal of Forest Research*, **50**, 359–370.

Heym, M., Uhl, E., Moshammer, R., Dieler, J., Stimm, K., and Pretzsch, H. (2021). Utilising forest inventory data for biodiversity assessment. *Ecological Indicators*, **121**, 107196.

Hill, A., Mandallaz, D., and Langshausen, J. (2018). A double-sampling extension of the German national forest inventory for design-based small area estimation on forest district levels. *Remote Sensing*, **10**, 1–26.

Irulappa-Pillai-Vijayakumar, D. B., Renaud, J. P., Morneau, F., McRoberts, R. E., and Vega, C. (2019). Increasing precision for french forest inventory estimates using the k-NN technique with optical and photogrammetric data and model-assisted estimators. *Remote Sensing*, **11**, 1–19.

Jeong, K. M. and Son, J. (2009). Estimation of small area proportions based on logistic mixed model. *The Korean Journal of Applied Statistics*, **22**, 153–161.

Jiang, J. and Lahiri, P. (2001). Empirical best prediction for small area inference with binary data. *Annals of the Institute of Statistical Mathematics*, **53**, 217–243.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1–96.

Jiang, J. and Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications (Vol. 1)*. Springer.

Jiang, J. and Rao, J. S. (2020). Robust small area estimation: An overview. *Annual Review of Statistics and Its Application*, **7**, 337–360.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K. T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., and Olsson, H. (2018). Remote sensing and forest inventories in Nordic countries–roadmap for the future. *Scandinavian Journal of Forest Research*, **33**, 397–412.

Knoke, T., Kindu, M., Schneider, T., and Gobakken, T. (2021). Inventory of forest attributes to support the integration of non-provisioning ecosystem services and biodiversity into forest planning-from collecting data to providing information. *Current Forestry Reports*, **7**, 38–58.

Kordos, J. (2016). Development of small area estimation in official statistics. *Statistics in Transition. New Series*, **17**, 105–132.

Lehtonen, R., Särndal, C. E., and Veijanen, A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, **29**, 33–44.

Magnussen, S., Mauro, F., Breidenbach, J., Lanz, A., and Kändler, G. (2017). Area-level analysis of forest inventory variables. *European Journal of Forest Research*, **136**, 839–855.

Mandallaz, D. (2013). Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research*, **43**, 441–449.

Mandallaz, D., Breschan, J., and Hill, A. (2013). New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, **43**, 1023–1031.

Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Pappalardo, L., and Gabrielli, L. (2015). Small area model-based estimators using big data sources. *Journal of Official Statistics*, **31**, 263–281.

Mauro, F., Molina, I., García-Abril, A., Valbuena, R., and Ayuga-Téllez, E. (2016). Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. *Environmetrics*, **27**, 225–238.

Mauro, F., Monleon, V. J., Temesgen, H., and Ford, K. R. (2017). Analysis of area level and unit level models for small area estimation in forest inventories assisted with LiDAR auxiliary information. *PloS One*, **12**, e0189401.

McRoberts, R. E. (2010). Probability-and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment*, **114**, 1017–1025.

McRoberts, R. E. (2012). Estimating forest attribute parameters for small areas using nearest neighbors techniques. *Forest Ecology and Management*, **272**, 3–12.

McRoberts, R. E., Andersen, H. E., and Næsset, E. (2014). Using airborne laser scanning data to support forest sample surveys. In Maltamo, M., Næsset, E., and Vauhkonen, J., editors, *Forestry Applications of Airborne Laser Scanning. Managing Forest Ecosystems*. Springer, Dordrecht.

Nguyen, C. V. (2023). Coverage of safely managed water, sanitation and hygiene: An application of small area estimation. *Science of The Total Environment*, **896**, 165277.

Noble, A., Haslett, S., and Arnold, G. (2002). Small area estimation via generalized linear models. *Journal of Official Statistics*, **18**, 45–60.

Novo-Fernández, A., Barrio-Anta, M., Recondo, C., Cámara-Obregón, A., and López-Sánchez, C. A. (2019). Integration of national forest inventory and nationwide airborne laser scanning data to improve forest yield predictions in north-western spain. *Remote Sensing*, **11**, 1693.

Næsset, E. (2014). Area-based inventory in Norway – from innovation to an operational reality. In Maltamo, M., Næsset, E., and Vauhkonen, J., editors, *Forestry Applications of Airborne Laser Scanning. Managing Forest Ecosystems*. Springer, Dordrecht.

Opsomer, J. D., Botts, C., and Kim, J. Y. (2003). Small area estimation in a watershed erosion assessment survey. *Journal of Agricultural, Biological, and Environmental Statistics*, **8**, 139–152.

Packalén, P. and Maltamo, M. (2007). The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, **109**, 328–341.

Pandey, P. C., Srivastava, P. K., Chetri, T., Choudhary, B. K., and Kumar, P. (2019). Forest biomass estimation using remote sensing and field inventory: a case study of Tripura, India. *Environmental Monitoring and Assessment*, **191**, 1–15.

Panse, V. G., Rajagopalan, M., and Pillai, S. S. (1966). Estimation of crop yields for small areas. *Biometrics*, **66**, 374–388.

Pramanik, S., Muthusamy, N., Gera, R., and Laxminarayan, R. (2015). Vaccination coverage in India: a small area estimation approach. *Vaccine*, **33**, 1731–1738.

Prasad, N. N. and Rao, J. N. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163–171.

Prisley, S., Bradley, J., Clutter, M., Friedman, S., Kempka, D., Rakestraw, J., and Sonne Hall, E. (2021). Needs for small area estimation: Perspectives from the US private forest sector. *Frontiers in Forests and Global Change*, **4**, 746439.

Pulkkinen, M. and Zell, J. (2019). *Overview of Research Carried out on Small-Area Estimation Around Swiss NFI Since 2010.* Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL), Birmensdorf.

Rahlf, J., Hauglin, M., Astrup, R., and Breidenbach, J. (2021). Timber volume estimation based on airborne laser scanning-comparing the use of national forest inventory and forest management inventory data. *Annals of Forest Science*, **78**, 1–14.

Rao, J. N. K. (2003). Some new developments in small area estimation. *Journal of the Iranian Statistical Society*, **2**, 145–169.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. John Wiley and Sons Inc, Hoboken, NJ, USA, 2nd edition.

Reich, R. M. and Aguirre-Bravo, C. (2009). Small-area estimation of forest stand structure in Jalisco, Mexico. *Journal of Forestry Research*, **20**, 285–292.

Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T. G., Nelson, R. F., and Ståhl, G. (2016). Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science*, **73**, 895–910.

Sharma, S. D., Srivastava, A. K., and Sud, U. C. (2004). Small area crop estimation for crop yield estimates at Gram Panchayat level. *Journal of the Indian Society of Agricultural Statistics*, **57**, 26–37.

Singh, R. and Goel, R. C. (2000). Use of remote sensing satellite data in crop surveys. Technical report, Indian Agricultural Statistics Research Institute, New Delhi.

Sisodia, B. V. S. and Chandra, H. (2012). Estimation of crop production at smaller geographical level in India. *Journal of the Indian Society of Agricultural Statistics*, **66**, 313–319.

Srivastava, A. K., Ahuja, D. L., Bhatia, D. K., and Mathur, D. C. (1999). Small area estimation techniques in agriculture. In *Souvenir on Advances in Agricultural Statistics and Computer Application*, New Delhi. Indian Agricultural Statistics Research Institute.

Srivastava, A. K., Sud, U. C., and Chandra, H. (2007). Small area estimation - an application to national sample survey data. *Journal of the Indian Society of Agricultural Statistics*, **61**, 249–254.

Stanke, H., Finley, A. O., and Domke, G. M. (2022). Simplifying small area estimation with rFIA: a demonstration of tools and techniques. *Frontiers in Forests and Global Change*, **5**, 745874.

Strîmbu, V. F., Ørka, H. O., and Næsset, E. (2021). Consistent forest biomass stock and change estimation across stand, property, and landscape levels. *Canadian Journal of Forest Research*, **51**, 848–858.

Sud, U. C., Chandra, H., and Srivastava, A. K. (2011). Crop yield estimation at district level using improvement of crop statistics scheme data - an application of small area estimation technique. *Journal of the Indian Society of Agricultural Statistics*, **66**, 321–326.

Sud, U. C., Mathur, D. C., Srivastava, A. K., Bathla, H. V. L., and Jha, G. K. (2001). Crop yield estimation at small area level using farmers' estimates. Technical report, Indian Agricultural Statistics Research Institute, New Delhi.

Sugasawa, S. and Kubokawa, T. (2020). Small area estimation with mixed models: a review. *Japanese Journal of Statistics and Data Science*, **3**, 693–720.

Thumaty, K. C., Fararoda, R., Middinti, S., Gopalakrishnan, R., Jha, C. S., and Dadhwal, V. K. (2016). Estimation of above ground biomass for central indian deciduous forests using ALOS PALSAR L-band data. *Journal of the Indian Society of Remote Sensing*, **44**, 31–39.

Tomppo, E., Haakana, M., Katila, M., and Peräsaari, J. (2008a). *Multi-source National Forest Inventory: Methods and applications.* Springer Science & Business Media.

Tomppo, E. and Katila, M. (1991). Satellite image-based national forest inventory of Finland for publication in the IGARSS 91 digest. In *Proceedings IGARSS'91 Remote Sensing: Global Monitoring for Earth Management.* IEEE.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., and Katila, M. (2008b). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, **112**, 1982–1999.

Vega, C., Renaud, J. P., Sagar, A., and Bouriaud, O. (2021). A new small area estimation algorithm to balance between statistical precision and scale. *International Journal of Applied Earth Observation and Geoinformation*, **97**, 1–11.

Vidal, C., Alberdi, I., Hernández, L., and Redmond, J. J. (2016). *National forest inventories*, volume 10. Springer Science+ Business Media.

Wagner, J., Münnich, R., Hill, J., Stoffels, J., and Udelhoven, T. (2017). Non-parametric small area models using shape-constrained penalized b-splines. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **180**, 1089–1109.

Wilson, B., Coulston, J., Prisley, S., and Radtke, P. (2023). *Small area estimation in forest inventories: New needs, methods, and tools.* Frontiers in Forests and Global Change.

Wulder, M. A., White, J. C., Fournier, R. A., Luther, J. E., and Magnussen, S. (2008). Spatially explicit large area biomass estimation: three approaches using forest inventory and remotely sensed imagery in a GIS. *Sensors*, **8**, 529–560.