International Journal of Image and Data Fusion

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tidf20

A brief overview and perspective of using airborne Lidar data for forest biomass estimation

Dengsheng Lu & Xiandie Jiang

To cite this article: Dengsheng Lu & Xiandie Jiang (2024) A brief overview and perspective of using airborne Lidar data for forest biomass estimation, International Journal of Image and Data Fusion, 15:1, 1-24, DOI: 10.1080/19479832.2024.2309615

To link to this article: https://doi.org/10.1080/19479832.2024.2309615

Published online: 25 Jan 2024.

INTERNATIONAL JOURNAL OF IMAGE AND DATA FUSION
2024, VOL. 15, NO. 1, 1–24
https://doi.org/10.1080/19479832.2024.2309615

REVIEW ARTICLE

A brief overview and perspective of using airborne Lidar data for forest biomass estimation

Dengsheng Lu[a,b,c] and Xiandie Jiang[a,b]

[a]Key Laboratory for Humid Subtropical Eco-Geographical Processes of the Ministry of Education, Fujian Normal University, Fuzhou, China; [b]Institute of Geography, Fujian Normal University, Fuzhou, China; [c]Fujian Provincial Engineering Research Center for Forest Carbon Metering, Fujian Normal University, Fuzhou, China

ABSTRACT

Lidar data have been regarded as the most important data source for accurate forest biomass estimation. Different platforms such as terrestrial Laser scanning, unmanned aerial vehicle Laser scanning, airborne Laser scanning, and spaceborne Lidar (e.g. ICESat-1/2, GEDI, GF-7 Lidar) provide new opportunities to map forest biomass distribution at different scales. The ground-based Lidar data are mainly used for extracting individual tree parameters such as diameter at breast height (DBH) and tree height, attempting to replace or reduce field work, while spaceborne Lidar data are often used to extract canopy height data at national and global scales, but cannot provide wall-to-wall mapping. The airborne Lidar may be the most frequently used data for forest biomass estimation at local scale. Many studies have been conducted for mapping forest biomass distributions in different climate zones, but current research situations and challenges of using airborne Lidar data have not been fully overviewed. This paper attempts to provide an overview of using airborne Lidar data for forest biomass estimation and discuss current research problems and future directions, which will be valuable for professionals and practitioners to better understand the important role of using airborne Lidar data for forest biomass estimation at the local scale.

ARTICLE HISTORY
Received 26 December 2023
Accepted 21 January 2024

KEYWORDS
Airborne Lidar; forest biomass; modelling; regression; machine learning

1. Introduction

The spatial distribution and dynamic change of forest biomass at regional and global scales are important data sources for examining climate change and ecosystem services (Walker *et al*. 2022), while their spatial patterns and dynamic change at local scale are required for precise forest management and planning (Erb *et al*. 2018, Mo *et al*. 2023). Therefore, accurately mapping forest biomass distribution at timely manner is necessary. Traditional method for calculation of forest biomass is based on field surveys of sample plots using certain sampling technologies such as systematic sampling (Zeng *et al*. 2023). However, this method cannot provide the spatial distribution of forest biomass with different forest types. The advancement of remote sensing technologies makes it possible to quickly and accurately estimate forest biomass in a large area (Lu *et al*. 2016, Araza *et al*. 2022).

CONTACT Dengsheng Lu ludengsheng@fjnu.edu.cn Key Laboratory for Humid Subtropical Eco-Geographical Processes of the Ministry of Education, Fujian Normal University, Fuzhou, China

Optical sensor data such as Landsat in 1990s and 2000s and Sentinel-2 in recent years have been extensively used for forest biomass estimation (Lu 2006, Lu *et al.* 2016, Xiao *et al.* 2019, Ahmad *et al.* 2021). Many previous studies showed that the medium spatial resolution (e.g. 10–30 m) images are important data sources for forest biomass estimation, and combination of spectral responses and textural images can improve modelling performance and estimation accuracy (Lu 2005, Zhao *et al.* 2016a). However, one big problem in using optical sensor data for biomass estimation is the signal saturation problem, that is, when forest aboveground biomass reaches a certain value such as around 120–160 Mg/ha in subtropical forests (Zhao *et al.* 2016a), the Landsat Thematic Mapper (TM) imagery cannot effectively reflect the subtle spectral change, resulting in considerable underestimation of forest biomass. The data saturation values vary, depending on the complexity of forest landscape (e.g. forest types and stand structures) and spectral wavelengths (e.g. visible, near-infrared, or short-wave infrared bands) (Zhao *et al.* 2016a). Figure 1 shows the relationships between forest aboveground biomass and entropy (used to represent complexity of forest stand structure) in the tropical forests in the Brazilian Amazon (Lu *et al.* 2005), indicating that Landsat TM image can be effectively used for biomass estimation of successional forests, but not for primary forest because of the data saturation. This situation is also found in subtropical regions that different forest types have various

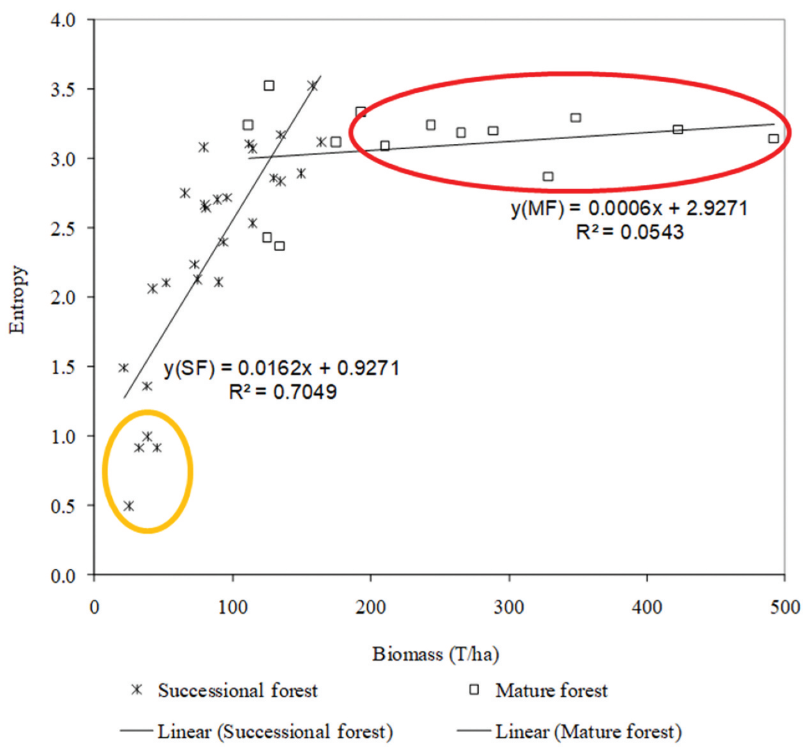

Figure 1. Data saturation problem in landsat imagery based on tropical forests (SF – successional forest; MF – mature forest) in the Brazilian Amazon (this figure is modified from Lu 2005, International Journal of Remote Sensing).

data saturation values (Zhao *et al*. 2016a). Another problem is the high uncertainty when biomass is small because of the impacts of background (e.g. soils, understory) under the canopy on the surface reflectance, resulting in overestimation of forest aboveground biomass (Lu *et al*. 2005, Zhao *et al*. 2016a, Chen *et al*. 2018).

Comparing with optical sensor data that mainly capture land surface features, SAR data such as ALOS PALSAR L-band can penetrate forest canopy to a certain depth to capture more information about forest structure features than optical sensor data, thus SAR data are also used for forest biomass estimation (Berninger *et al*. 2019, Ruusa *et al*. 2022). However, the noise problem and the impacts of terrain and moisture conditions on SAR data, and limited capability in separation of different forest types (Li *et al*. 2012) make SAR data poor performance in forest biomass estimation (Zhao *et al*. 2016b). The SAR data also have serious data saturation problem, similar to optical sensor data. Considering the impacts of external factors such as soil moisture and atmospheric conditions on the surface reflectance of optical sensor data or on the backscatter coefficients of SAR data, the same forest type in different locations may have different radiometric values, resulting in difficulty of model transfer at spatial and temporal scales (Lu 2005, Zhao *et al*. 2016a, b).

Considering the common data saturation problem in optical sensor or SAR data, Lidar data can solve this problem because of its ability to extract tree height which is closely related to forest biomass/volume, as shown in Figure 2. Lidar data can be obtained from different platforms such as ground, mobile, unmanned aerial vehicle (UAV), airborne, and satellite. The ground and mobile Lidar due to their very high point cloud density are often used for extracting single tree parameters such as diameter at breast height (DBH), tree height, and crown size, but they are difficult to use in a large area because of expense in data collection, and huge data volume making processing difficult. The spaceborne Lidar data itself cannot provide wall-to-wall mapping of forest biomass without integration with optical or SAR data (Chi *et al*. 2015, Duncanson *et al*. 2020, Liang *et al*. 2023). The airborne Lidar data have been recognised as the most important data source for forest biomass estimation at local scale, and many studies have been conducted for mapping forest biomass/carbon stock distributions in different climate zones (Li *et al*. 2016, Guo *et al*. 2018, Stereńczak *et al*. 2020), but current research situations and challenges of using airborne Lidar data for biomass estimation have not been fully overviewed. This paper tried to provide a brief overview of forest biomass estimation procedure using airborne

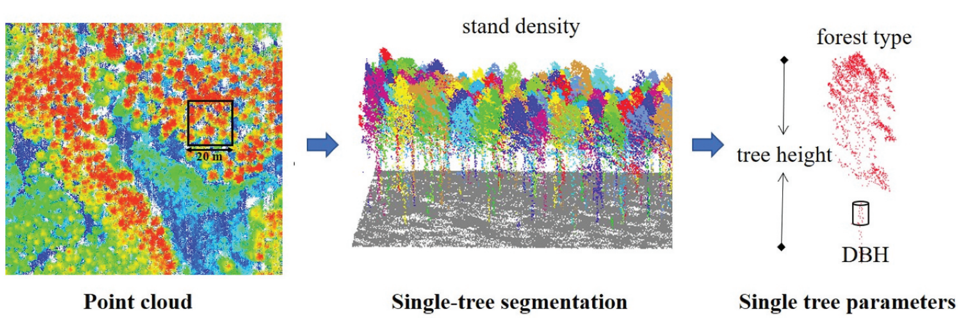

Figure 2. The unique characteristics of Lidar data in extracting individual tree parameters (e.g. DBH, tree height) and canopy features (CHM).

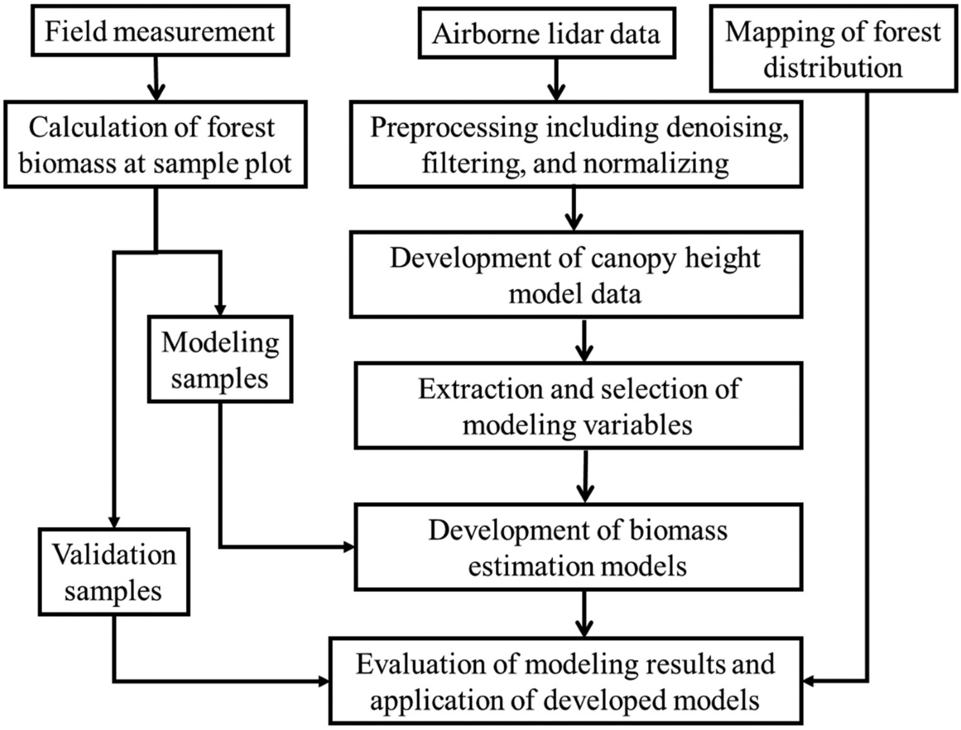

Figure 3. General procedure of using airborne Lidar data for forest biomass estimation.

Lidar data. Emphasis is on the discussion of potential challenge and future directions of using airborne Lidar data for improving forest biomass estimation.

2. General procedure of using airborne Lidar data for forest biomass estimation

A general procedure of using airborne Lidar data for forest biomass estimation is illustrated in Figure 3, including data preparation (e.g. collection of samples, airborne Lidar data), extraction and selection of proper variables for modelling, selection of a suitable modelling algorithm, evaluation of modelling results and application of the model to entire study area.

2.1. Data preparation

Collection of sufficient number of representative sample plots is required for conducting forest biomass estimation research. As summarised by Lu *et al*. (2016), three methods – destructive sampling, allometric models, and conversion from stocking volume to biomass – can be used to obtain biomass reference data. Allometric models are the most common approach when DBH and/or tree height at individual tree level are available. Thus, accurate field measurements and properly selected allometric equations are two crucial factors for providing high-quality biomass reference data

(Liao *et al*. 2022). This usually involves the following considerations: (1) when tree height data is difficult to acquire during field work, calculation of tree biomass in a large area has high uncertainty if DBH is applied alone (Liao *et al*. 2022). This is because the same DBH may have very different tree height caused by soil conditions and topographic factors; (2) some tree species do not have allometric equations to calculate single tree biomass, improper use of an allometric equation may result in high calculation uncertainty (Yang *et al*. 2023).

Traditional methods using telescopic poles or altimeters cannot accurately measure tree height in densely forested lands. The advances of Lidar technology (e.g. ground or mobile laser scanning, UAV) provide new opportunities to accurately extract tree height. In particular, integration of ground or handheld laser scanner data and UAV Lidar data can effectively measure DBH and tree height (Liao *et al*. 2022). Major process of the Lidar point clouds included denoising, filtering, and normalising (Dong and Chen 2018, Guo *et al*. 2018). Denoising included removal of low points, air noise, and isolated points. Filtering is used to remove non-ground Lidar points so that bare-earth digital elevation models can be created from the remaining ground Lidar points. An improved progressive triangulated irregular network (TIN) densification (IPTD) filtering algorithm was frequently used for the filtering purpose in forest biomass estimation research because this algorithm can deal with various forest landscapes, especially forested areas with complex terrain and environment (Zhao *et al*. 2016). The ground returns were interpolated into a digital terrain model (DTM), and the remaining point clouds were then normalised against the ground surface height. Canopy height model (CHM) is then produced by differencing DTM from Digital Surface Model (DSM). These point cloud data processing functions can be implemented in commercial software like lidar360 (https://greenvalleyintl.com/software/lidar360/) or free software packages like Fusion and LAStools (McGaughey 2009).

2.2. Extraction and selection of modeling variables at the grid-level or object level

Some previous studies had summarised the commonly used variables from Lidar data for biomass estimation (e.g. Chen 2013, Lu *et al*. 2016, Dong and Chen 2018, Guo *et al*. 2018). CHM may be the most frequently used data source for extraction of variables such as mean height, standard deviation, and height percentile (10th-100th) (Dong and Chen 2018, Guo *et al*. 2018). The CHM-based variables are insensitive to flight conditions and sensor settings (Roussel *et al*. 2017), and more efficient than point cloud-derived variables by focusing solely on canopy surface heights, but may miss the structural variations within canopy. Some studies have indicated that various methods of variable generation have limited effects on the predictive performance of forest attribute estimation (Lu *et al*. 2012). In addition to height-related variables, point-based variables that capture canopy cover measurements, canopy return density measurements (Cao *et al*. 2019), and canopy geometric volume (Chen *et al*. 2007) have also been used for biomass estimation modelling. Laser return intensity (Xie *et al*. 2023), textural features from CHM data using the grey level co-occurrence matrix (GLCM) (Li *et al*. 2019, Jiang *et al*. 2020b) combined with height variables have been proven effective in improving biomass modelling performance compared to using height variables alone.

Lidar-based biomass prediction is generally conducted based on extracted variables at the plot level (the plot size can be from 10 m × 10 m to 100 m × 100 m; and shapes can be square, rectangle, or circle) (Frazer *et al*. 2011). The plot size is often designed as 1 mu (25.82 m × 25.82 m, square) in forest inventory in China, and 20 m × 20 m in different research purposes (Bouvier *et al*. 2019, Zeng *et al*. 2023). Previous research has confirmed that sample plot size of 20 m × 20 m is optimal considering spatially continuous forest distribution in mountainous regions (Jiang *et al*. 2020b, Sousa *et al*. 2023). However, for forest distribution with sparse coverage and narrow strips in plain regions, variables are extracted at segmentation units (referred to as object-based approach) provide better modelling prediction performance than based on squared plot size (Li *et al*. 2019, Silveira *et al*. 2019, Tamiminia *et al*. 2022, Wang *et al*. 2023). Whether grid or object level is used to extract variables from Lidar data or other remotely sensed data depends on the characteristics of the forest landscape under investigation. A general rule may be the selection of grid level for spatially continuous forest distribution in mountainous regions and of object level for the complex and small patch size of forest distribution in plain regions (Wang *et al*. 2023).

2.3. Biomass estimation modeling methods

The biomass estimation models can be generally grouped into parametric and non-parametric algorithms. Traditional biomass estimation models based on Lidar data rely on the parametric regression methods including linear, logistic, exponential and Richards that has a straightforward interpretability (Xie *et al*. 2023). It is interesting to note that linear regression is the most used approach for plantations with relatively low tree densities (Nie *et al*. 2017, Jiang *et al*. 2020b, Gao *et al*. 2022), but for mature or over-mature forests like tropical rainforests, nonlinear transform using logarithms and square roots are often used as response or explanatory variables (Longo *et al*. 2016, Phua *et al*. 2016, Feng *et al*. 2017). Some researchers have employed machine learning methods to establish forest biomass estimation based on Lidar data (Rex *et al*. 2020, Marchesan *et al*. 2023, Xie *et al*. 2023), but there is an ongoing debate about the effectiveness of machine learning versus parametric models. It is evident that there are two primary reasons for the preference of machine learning algorithms: (1) Considering the heterogeneity and complexity of vegetation composition and the impact of environmental conditions (e.g. soil and topography) on tree growth, combining Lidar data with machine learning methods could be appropriate for predicating biomass (Jiang *et al*. 2020b, Xie *et al*. 2023). For a forest type with complex stand structure or tree species composition such as mixed coniferous and broadleaf forest, a machine learning algorithm with strong ability to find the potential regularity between biomass and variables is more suitable than the linear model (Jiang *et al*. 2020b); (2) Machine learning is favoured when multiple source data such as optical, SAR, and terrain data are used for forest biomass modelling (Gao *et al*. 2022, Marchesan *et al*. 2023). In such cases, the input data may not adhere to certain distribution requirements of parametric regression, and the relationships tend to be intricate. The machine learning techniques have the advantages of dealing with multiple source data without considering the requirements of input data and the relationships between biomass and input variables, thus, can be used to estimate biomass using a wide range of variables.

Different machine learning algorithms such as K-nearest neighbour, artificial neural network, random forest (RF), support vector machine (SVM), and Maximum Entropy can be used to establish forest biomass estimation models and their performances may be different (Lu *et al*. 2016, Gao *et al*. 2018). Based on Lidar data, Marchesan *et al*. (2023) and Xie *et al*. (2023) compared the performance of these algorithms in biomass estimation and found that SVM and RF performed better than other algorithms. This may be the fact that SVM and RF can work with relatively small number of samples and have a good generalisation ability (Marchesan *et al*. 2023, Xie *et al*. 2023). The SVM method searches for the optimal hyperplane to construct a maximum margin separator, and suitably generalises a given model with limited training samples. The RF algorithm is an ensemble learning approach using many decision trees, which reduces overfitting problems, and RF is less sensitive to data noise and outliers. In recent years, the boosting algorithms that convert weak learners into strong learners, such as extreme gradient boosting (XGB) and light gradient boosting machine (LGBM) (Tang *et al*. 2022, Xie *et al*. 2023) have achieved good performance in biomass estimation using multiple data sources (Fang *et al*. 2022, Huang *et al*. 2022). However, such methods are easily influenced by the setting of super parameters, and cautions should be taken when only limited number of samples are available.

2.4. Evaluation of modeling performance and application of the developed models

Evaluation of modelling performance is often required for understanding the robustness and reliability of the models and whether the developed model has transferability or not (Lu *et al*. 2016). The selection of reference data is critical for modelling evaluation. In general, three strategies of reference data selection are used: leave-one-out cross-validation (LOOCV), random separation of samples from one population into modelling and validation ones, and completely independent samples from different study areas. Considering the small number of samples available, LOOCV is often used for accuracy assessment (Longo *et al*. 2016, Nie *et al*. 2017, Rex *et al*. 2020). LOOCV leaves one sample out as a testing set and uses the remaining samples as a training set to construct biomass models. This process is repeated until all samples are used as the testing set once (Jiang *et al*. 2020b). Although LOOCV provides an unbiased estimate of model performance, it is sensitive to outliers and may not be representative of the model's performance, and the estimation accuracy may be over optimistic. If sufficient samples are available, we can split all samples into distinct training and validation sets in advance and use the validation samples to evaluate the modelling performance. An alternative is to use completely different samples from other sites that are different from the modelling site. This validation may be optimal but requires much workload to collect the samples.

When validation samples are available, we can use different quantitative measures such as coefficient of determination (R^2), root mean square error (RMSE), and relative RMSE (RMSEr) to conduct the modelling evaluation (Lu *et al*. 2016). In addition to the traditional accuracy assessment measures, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are also used to measure the complexity and goodness of fit of the machine learning models (Vaglio Laurin *et al*. 2014, Huang *et al*. 2022). In general, multiple measures are often combined to evaluate the modelling performance for better providing the quantitative assessment. For example, a higher R^2 and lower RMSE and RMSEr indicate better modelling performance. A higher R^2 cannot guarantee a lower RMSE if system error exists.

One objective to develop a model is to apply it to entire study area for mapping forest biomass distribution or to the same study area but using different years of Lidar data for examining forest biomass dynamics (Lu *et al*. 2016). For grid-based modelling approach, the grid cell size for model prediction (i.e. minimal mapping unit) should be equivalent to the size of modelling sample plots, while for the object-based modelling approach, selection of minimum segmentation unit for model prediction is needed. Sometimes a good modelling performance cannot guarantee a good prediction result because of the complexity of land surface features and the representativeness of the modelling samples. Therefore, it is important to conduct the evaluation of modelling prediction using independent samples. Evaluation of modelling prediction results can be further used to indicate the modelling performance and provide possible improvement of optimising the modelling procedure, including collection of sample plots and selection of modelling variables. If possible, conducting uncertainty analysis will be valuable to identifying the major factors influencing modelling results, so that emphasis can be focused on the improvement of the key variables for biomass modelling.

3. Special concerns on the Lidar-based forest biomass estimation

3.1. The needs to obtain accurate sample plot data through use of lidar data

Collection of sufficient number of representative sample plots are prerequisite for developing robust biomass estimation models. Previous studies have shown that the allometric equations using both DBH and tree height provided better calculation accuracy than the one using DBH alone (Liao *et al*. 2022). However, one challenge is to obtain accurate tree height in forest sites because traditional methods using altimeter or telescopic pole cannot effectively measure tree height, thus, most of biomass calculation at plot level is based on DBH alone, resulting in high uncertainty (Phalla *et al*. 2017, Liao *et al*. 2022). UAV Lidar data can provide accurate tree height measurement, as shown in Figure 4 that comparing with the true values based on tree height measurement of felled Masson pine and eucalyptus, the UAV Lidar data provide more accurate tree height values than traditional measurement using telescopic poles. In another study of Baisha Forest Farm in Shanghang County, Fujian Province, we found that fusion of handheld laser scanner data and UAV Lidar data can further improve the retrieval accuracy of tree height, especially for dense and tall plantations (unpublished results).

Terrestrial or handheld laser scanning instrument captures data through the down-to-up scanning way with very high density of point clouds, thus, this data can be used to distinguish stem, branches and leaves within a tree, and can effectively segment individual trees (Liang *et al*. 2018, Chung *et al*. 2019). The airborne laser scanning instrument captures data with the up-to-down scanning way with relatively lower density of point clouds than handheld laser instrument, thus, it cannot provide so detailed tree structure features for separation of stem, branches and leaves, but airborne Lidar data can provide more accurate tree height than handheld laser scanning data, especially when canopy density is relatively high (Wang *et al*. 2019a, b, Jurjević *et al*. 2020, Liao *et al*. 2022). Since use of airborne Lidar data alone cannot accurately extract single tree parameters such as tree segmentation and diameter at breast height, combination of handheld and airborne Lidar data can take their advantages into the fused data for accurate extraction of tree

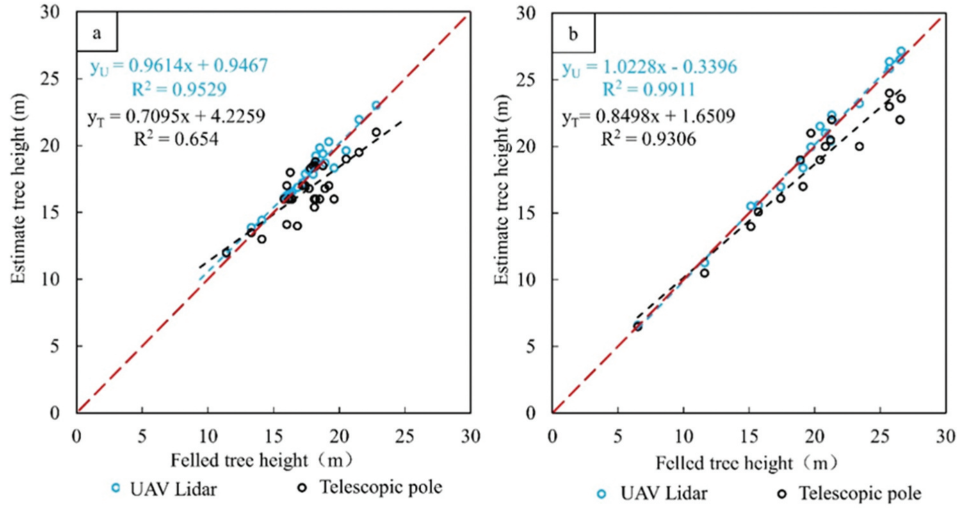

Figure 4. Relationship between reference data and estimated tree heights using UAV lidar and the telescopic pole for (a) Masson pine in Baisha Forest Farm in Shanghang County and (b) eucalyptus in Yuanling Forest Farm in Yunxiao County, Fujian Province (this figure is from Liao *et al*. 2022. Remote sensing).

parameters such as DBH and tree height (Dai *et al*. 2019, Zhao *et al*. 2023). These experiments imply that effective use of Lidar data can improve the quality of biomass sample data comparing with using traditional field measurements, in addition to reducing the field work loads.

3.2. The importance of using stratification-based modeling strategy

The stratification-based biomass estimation approach has been proven to be effective for improving estimation accuracy either using optical sensor of Lidar data (Zhao *et al*. 2016a, b, Heiskanen *et al*. 2019, Lin *et al*. 2022). Previous research indicated that the differences in forest types, environmental conditions (e.g. terrain features, soil types, and climate conditions), and study areas under investigation lead to different relationships between forest biomass and Lidar-derived variables (Patrick *et al*. 2018, Ewijk *et al*. 2020, Jiang *et al*. 2020b). Stratification of forest types is a common approach in biomass modelling (Cao *et al*. 2014, Heiskanen *et al*. 2019), considering that a single tree biomass is mainly determined by a specific tree species, DBH, and tree height (Dong and Chen 2018). In a south subtropical region of China, Jiang *et al*. (2020b) confirmed that biomass modelling based on stratification of forest types reduced the RMSEr from 38.01% for non-stratification to 20.05% for Masson pine, 18.73% for Chinese fir, and 24.82% for Eucalyptus.

In addition to specific tree species group, other forest groups including deciduous and non-deciduous forests (Ryan and Paul 2013), agroforestry and secondary forests (Feng *et al*. 2017), and the on-year and off-year bamboo forests (Chen *et al*. 2018) have also been used for stratification. However, improper stratification of forest types may reduce the biomass modelling performance (Jiang *et al*. 2020b). As shown in Figure 5, although both Chinese fir and Masson pine belong to coniferous forest, they have

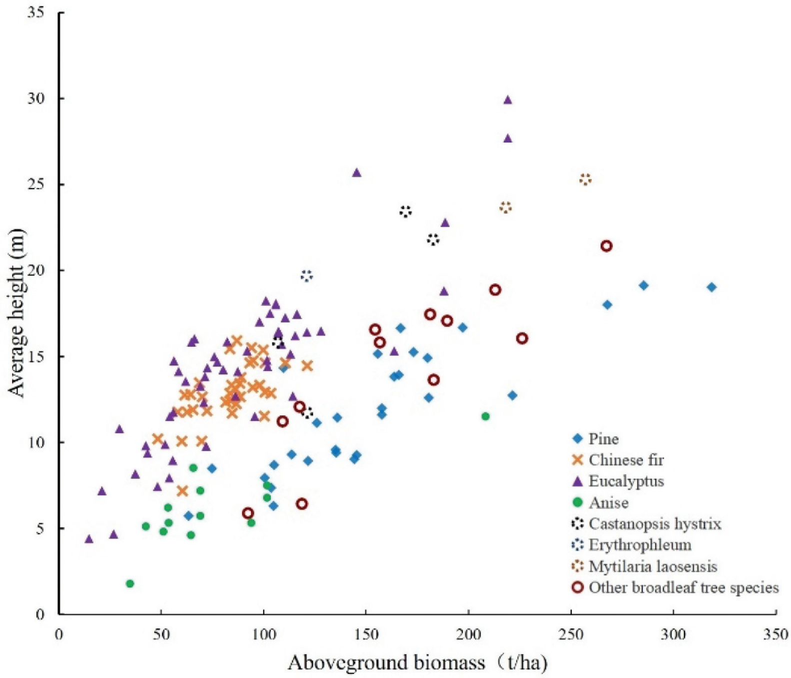

Figure 5. A comparison of forest biomass and canopy height relationships among different tree species in s subtropical region – Gaofeng forest Farm of Guangxi (this figure was modified from Jiang *et al*. 2020a, remote sensing).

different relationships between average height and biomass in a southern tropical region (Jiang *et al*. 2020b). This is because they have different crown sizes and shapes, and forest stand structures, that is, Chinese fir is often in a pure plantation with homogenous stand structure and age, while Masson pine is often mixed with some broadleaf tree species with uneven ages. Thus, biomass modelling for Chinese fir and Masson pine separately has better performance than modelling for coniferous forest (Jiang *et al*. 2020b). Understanding the relationships between tree height and biomass for different tree species is essential for determining stratification strategy, as Jiang *et al*. (2020b) found that stratification based on the analysis of forest stand structure relationship can improve modelling performance while reducing the number of modelling samples.

Forest type is often used for stratification, an alternative is based on environmental factors such as topographic factors and soil conditions because different site conditions can affect the tree growth, thus affecting the relationships between tree height and biomass. The stratification based on topographic position such as valley, slope and ridges can lower vertical heterogeneity and highlight the horizontal variations, especially when optical sensor data are used (Molina *et al*. 2016, Zhao *et al*. 2016a). However, airborne Lidar data with high density of point clouds is less affected by terrain conditions, thus, the stratification based on topographic factors has limited effects on Lidar-based biomass estimation modelling performance (Lin *et al*. 2022). Due to the limited extent of Lidar data and the similar environmental condition, stratification based on environmental factors has less effects than forest types on modelling performance (Tamga *et al*. 2023). Some previous research explored regional stratification based on the difference in climate, topography, soil conditions and even management practices when Lidar data are available in different regions (Ewijk *et al*. 2020). As an example, Naesset and Gobakken (2008) developed Lidar-based OLS (Ordinary Least Squares) regression models for biomass estimation in ten different study sites in Norway and the model predictions varied markedly among the geographical regions. As shown in Figure 6, even for the same tree species (Chinese fir here), the relationship between Lidar variables (H_{mean}) and biomass varies in different regions, implying the necessity of implementing stratification based on different regions for establishing biomass estimation models if sufficient number of modelling samples are available for each study area. In reality, the number of sample plots for each forest type is a key to determine the stratification strategy. Too many stratified groups require a large number of sample plots, which is often a challenging task to collect many sample plots considering the workload and cost.

Since limited number of field plots for each forest type make it difficult to develop reliable models for different strata, the universal models such as mixed-effects model and hierarchical Bayesian approach (HBA) provide new ways to address this problem (Chen *et al*. 2012, Lin *et al*. 2022, Xie *et al*. 2023). The mixed-effects model, whose regression coefficients are regarded as random Gaussian variables rather than constants and these random Gaussian variables vary with stratification factors (e.g. forest type, geographical location), is used to work with sparse samples (Feng *et al*. 2021). Compared to the use of Lidar data alone in multiplicative models, incorporation of vegetation types via mixed-effects models increased the R^2 from 0.77 to 0.83 and reduced RMSE by 10% (from 80.8 to

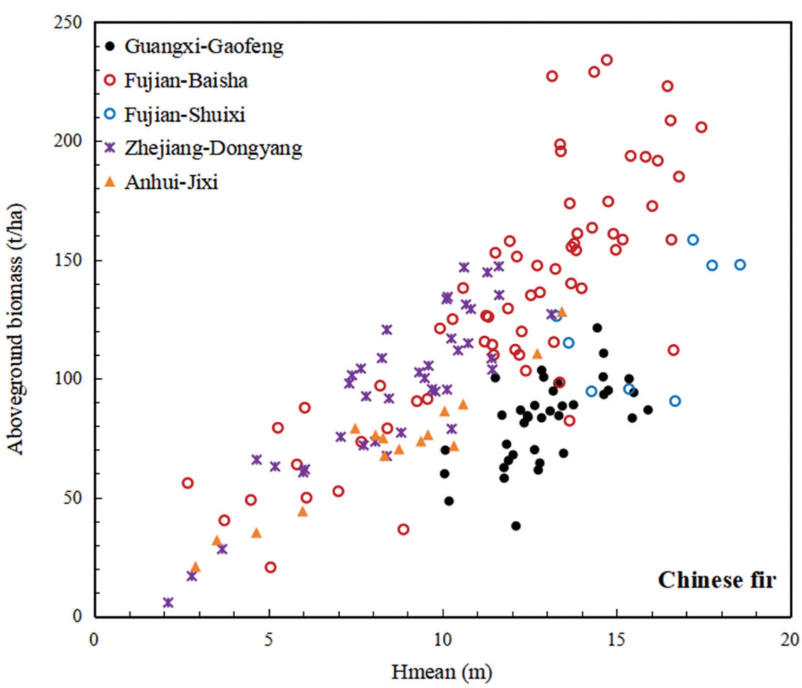

Figure 6. A comparison of forest biomass and canopy height relationships for Chinese fir forest from different study areas of subtropical regions.

72.2 Mg/ha) (Chen *et al*. 2012). The HBA allows for the incorporation of prior knowledge into the model, which can help mitigate issues related to small sample sizes (Wang *et al*. 2019a). Furthermore, HBA provides probability distributions for parameters, aiding in quantifying the uncertainty of parameter estimation, which is crucial for small sample situations (Wang and Blei 2018). The double-strata HBA (e.g. forest type and slope aspect) provides better forest volume estimation than single-stratum HBA (e.g. forest type) and solves the modelling problem due to limited sample sizes for forest types in a north subtropical region (Lin *et al*. 2022). These experiments imply that mixed-effects model or HBA can provide another way to establish biomass estimation models when other modelling approaches such as regression or machine learning algorithms cannot be effectively used to develop biomass estimation models for each forest type because of the limited number of sample plots.

3.3. The importance of combining multiple data sources in a modeling procedure

Among various types of sensors, Lidar has been recognised as the best data source for biomass estimation due to its ability to capture three-dimensional forest structure features, thus, effectively solving the data saturation problem existing in optical and radar data (Lu *et al*. 2016, Dong and Chen 2018, Guo *et al*. 2018). However, despite its advantages, Lidar data alone have limited ability to accurately detect biomass variations caused by different tree species composition and environmental stress (Lu *et al*. 2016, Catherine *et al*. 2019). Researchers have sought ways to integrate both Lidar and high spatial resolution optical sensor data to enhance biomass estimation modelling, taking their distinct capabilities in representing surface features into account. One option is that Lidar-based biomass estimation modelling performance can be improved by stratifying forest types which are developed from optical sensor data and terrain conditions (Lin *et al*. 2022). Another way is to directly establish the biomass estimation models using both Lidar-derived variables and optical imagery-based features using machine learning algorithms (Luo *et al*. 2017, Catherine *et al*. 2019).

Previous studies have obtained different conclusions on the combined use of Lidar and optical sensor data, depending on the selection of different data sources, algorithms, and characteristics of different land surfaces under investigation (Vaglio Laurin *et al*. 2014, Xu *et al*. 2015, Feng *et al*. 2017, Luo *et al*. 2017, Catherine *et al*. 2019, Gao *et al*. 2022). Compared to the multispectral images with the disadvantages of limited spectral bands, hyperspectral imaging (HSI) has ability in describing the physiological and ecological characteristics of forests (Catherine *et al*. 2019, Gao *et al*. 2022). Thus, the combination of HSI and Lidar data has the potential to improve forest biomass estimation (Luo *et al*. 2017, Catherine *et al*. 2019). For example, Anderson *et al*. (2008) found the improvements of 8–9% for biomass estimation across all forest conditions and 25% or more for the unmanaged forest using the fused Lidar and HSI data. Similar conclusion was obtained for fused Lidar and HSI data to improve biomass estimation in northwest China by reducing RMSE of 7.9% compared with the use of Lidar variables alone (Luo *et al*. 2017). HSI may capture ecosystem information relating to plant functional types that could in turn affect forest stand structures, thus improve biomass estimation (Vaglio Laurin *et al*. 2014). However, some research showed that incorporation of airborne Lidar with multispectral (Feng *et al*. 2017) or hyperspectral (Clark *et al*. 2011, Fassnacht *et al*. 2014) imagery did not much improve the biomass model performance. More research is needed to examine which modelling variables can be extracted from the multi-source data and which modelling approach should be used corresponding to these variables.

Different factors such as the regression method used, the chosen input metrics, the specific forest type being studied, and quality of the field measurements may result in different modelling performances (Catherine *et al*. 2019). Hyperspectral data have an indirect connection to biophysical properties, incorporation of these features may require more complex models than use of Lidar-based variables alone. Thus, different modelling methods may be used corresponding to specific tree species and associate their own structural features. Previous research showed that the best feature for eucalyptus biomass estimation is the height extracted from the point cloud, and the most effective feature for other broadleaf trees is the texture variables obtained through wavelet transform. These experiments indicate that the optimal features for tree species differ due to specific vertical and canopy structures.

The Lidar-based forest biomass estimation is based on the extracted CHM metrics, which tree height has strong relationship with forest biomass. From an individual tree level, once forest age exceeds a certain threshold, tree height no longer shows obvious variation (Cao *et al*. 2016). This could lead to increased uncertainty when estimating biomass in mature or over-mature forests based on Lidar data. In such cases, biomass estimates may heavily depend on forest age and DBH data. The increasing availability of time-series remote sensing data offers a potential solution to this problem (Jiang *et al*. 2023). However, it is worth noting that Landsat data, for instance, dates back to as early as 1985, and can monitor forests less than 40 years old (Crawford *et al*. 2023). This falls short of capturing the stable biomass growth phase of some tree species. This is why currently there is limited research combining Lidar data with age data. Nevertheless, addressing the challenge of high biomass uncertainty still requires further exploration and investigation.

An important challenge in integrating multi-source datasets is how to select the most appropriate metrics for estimating biomass (Catherine *et al*. 2019, Gao *et al*. 2022). When Lidar data alone are used, the selection of modelling variables is often to use stepwise regression. However, when multi-source data are used, stepwise regression may be not a proper choice for selection of modelling variables because different types of variables such as height-based variables, and terrain or soil-related variables cannot meet the linear relationship assumption between biomass and these variables (Lu *et al*. 2016). In this case, some algorithms that can handle nonlinear relationship will be valuable. Recursive Feature Elimination (RFE) is a feature selection method commonly used in machine learning-based modelling (Catherine *et al*. 2019). Its primary objective is to reduce model complexity, improve model performance, decrease the risk of overfitting, and enhance model interpretability by progressively eliminating unimportant features (Feng *et al*. 2017, Jiang *et al*. 2020b). In contrast, Gao *et al*. (2022) adopted three-level feature screening strategy, which can effectively avoid the problem of feature redundancy and reduce irrelevant features in the case of few measured samples. The Lidar and hyperspectral features underwent individual screening, followed by the screening of fused features from two data sources. The purpose of these variable selection methods is to considerably reduce the number of metrics used as input to produce parsimonious models for practical applications without losing much accuracy in biomass estimation (Catherine *et al*. 2019). More research is needed to use proper selection approach for identification of key modelling variables from multi-source data and to select proper modelling approach corresponding to these variables considering the modelling performance and applicability.

3.4 The potential of using deep learning approach for Lidar-based biomass estimation

Machine learning algorithms are often used to conduct biomass estimation based on multi-source data, but the common algorithms such as RF and SVM cannot provide satisfactory modelling results (Lu *et al*. 2016). An alternative is to use deep learning because of its powerful data mining ability for multiple data sources (Zhang *et al*. 2019, Ghosh and Behera 2021, Tian *et al*. 2023). Compared to the machine learning algorithms (e.g. K-nearest Neighbor, RF and SVM), the Stacked Sparse Autoencoder network (SSAE) model with inputs of combined optical and Lidar variables provided the best performance in forest biomass estimation (Zhang *et al*. 2019). Deep learning models generally capture and interpret different complexity levels within data through using multiple layers to improve biomass estimation accuracy (Ghosh and Behera 2021). Another advantage of deep learning model is that it can automatically extract abstract features which is different from artificial features (Tian *et al*. 2023). Some studies found that the convolutional neural networks (CNN) model, which can interpret spatial data using a series of trainable moving windows was much superior to other classic machine learning algorithms and multiple linear regression (Du *et al*. 2021, Somayeh and Ali 2022). However, the applications of deep learning algorithms in predicting forest biomass using Lidar data alone are still not common. The main reason is the high data requirements including samples and Lidar data. The collection of reference data could be one of the most labour-intensive and time-consuming components of a biomass mapping project.

Analysis of existing literature utilising airborne Lidar data for deep learning-based forest biomass estimation indicates that human-derived features and two-dimensional data, rather than three-dimensional images, have been applied to deep learning models (Zhang *et al*. 2019, Du *et al*. 2021). This results in the loss of the advantage of automatic feature extraction and selection that deep learning models offer. In situations where training data is limited, the ability of deep learning models to fit the data may not necessarily surpass that of other machine learning methods (Somayeh and Ali 2022). Additionally, if point cloud data is directly used as input, considerations such as data volume, processing speed, model complexity, and generalisability need to be addressed (Tian *et al*. 2023). Therefore, it is crucial to carefully assess the strengths and weaknesses of deep learning methods and consider them within the specific context. The difficulty in collecting sufficient number of biomass samples may be the key issue constraining the extensive application of deep learning in forest biomass estimation. Nevertheless, the deep learning model using multi-source data will become a powerful tool for applications in precision forest biomass monitoring in the future.

3.5. The importance to develop robust and transferable biomass estimation models

The importance of developing a forest biomass estimation model having transferability at temporal and spatial scale has been recognised and much research has been conducted in the past decade (Navarro *et al*. 2020, Passel *et al*. 2020). In general, the optical sensors or SAR based modelling approaches have weak transferability because land surface features captured by a sensor are strongly affected by external factors such as atmospheric and environmental conditions (Lu 2005, Zhao *et al*. 2016b). These limitations make the developed models not transferable in different study areas or even the same study area but different data acquisition dates (Lu 2005, Jiang *et al*. 2020a). In contrast, Lidar-based modelling approaches have the potential for model transferability because Lidar-derived canopy height variables have strong and reliable relationships with biomass (Görgens *et al*. 2015, Domingo *et al*. 2019). Some previous research has confirmed the temporal transferability of Lidar data through multi-temporal Lidar analysis (Fekety *et al*. 2015, Domingo *et al*. 2019, Navarro *et al*. 2020). For example, the biomass model using high density Lidar data which were acquired in 2016 was extrapolated to the data in 2011, indicating that transferable model is a successful way to reduce fieldwork efforts and accurately estimate forest attributes in two different dates (Domingo *et al*. 2019). However, considering the expense of using Lidar data, more research is needed to take other 3D data sources such as digital aerial photogrammetry for forest estimation modelling and model transferability (Li *et al*. 2019, Navarro *et al*. 2020, Lin *et al*. 2022).

Unlike the temporal transferability properties of Lidar-based estimation models, whether a developed model can be transferred to different regions at spatial scale remains unclear. Ewijk *et al*. (2020) and Tompalski *et al*. (2019) examined different transferability scenarios among study sites using different modelling methods (OLS, RF and k-NN), and concluded that prediction accuracy varied with modelling approaches and sites. In general, complex algorithms with more variables used increase the risk of overfitting to local conditions and hence lose generality when models need to be transferred to new environments (Werkowska *et al*. 2017). Moreover, models developed in one study area may be transferred successfully to cross-scene data if biophysical environments are similar (Werkowska *et al*. 2017, Fekety *et al*. 2018). In recent years, much research has been emphasised on developing Lidar-based universal models to predict forest attributes (Feng *et al*. 2021, Lin *et al*. 2022). These models attempted to improve estimation results by incorporating many sample plots from various environmental conditions and reducing the dependence on model coefficients for individual locations. In addition, incorporating sample data from extension areas into the calibration set along with the base data enhanced the transferability of biomass predictions of all modelling approaches and at each site (Ewijk *et al*. 2020). Due to the difficulty in obtaining field data, it is difficult to adjust the model with many samples from the extrapolation areas. Therefore, effectively using existing biomass models and limited samples from an extrapolation area to establish a new model will be valuable but has not been fully examined.

Compared to the area-based methods using statistical models to link biomass to Lidar metrics calculated at plot level, a new approach integrating LBI (Lidar Biomass Index) with an individual tree crown segmentation algorithm for estimating tree-level biomass using airborne Lidar point cloud data has been shown to have high potential in model's transferability (Du *et al*. 2023). The LBI-based approach provided a measure of crown volume weighted by height may better represent a tree than simple Lidar-based crown metrics, and it strengthens spatial extrapolation in biomass estimation. Du *et al*. (2023) found that the LBI-based method was not affected by those differences of different study areas and its predictions across three larch forest farms did not have obvious bias. This method has an obvious advantage in that it requires small fractions of the trees needed to calibrate area-based models. It may provide a cost-effective Lidar technology to monitor forest biomass for a broad range of tree species in different regions.

Although empirical models are convenient to use, easy to implement and often give accurate results, they have strong dependency on sample data, resulting in limited transferability (Christina *et al*. 2012). Physical models built on ecological and physiological mechanisms behind tree growth and biomass accumulation might allow for more robust and accurate biomass estimation than empirical approaches. The coupling of physical models and machine learning models has been regarded as an important focus in multiple interdisciplinary fields including earth sciences, and obtained increasingly attention in recent years (Shen and Zhang 2023). Combining two methods can effectively alleviate the 'bias' of the physical model and avoid the 'arrogance' of the machine learning model. In general, physical models often require numerous input parameters which are often not available at fine scale, while machine learning method may be employed to acquire accurate model parameters, thus providing improved initial conditions for subsequent model computations. As an example, Zhang *et al*. (2023) optimised and calibrated the physiological principles in predicting growth (3-PG) model using survey and UAV Lidar data firstly, and then add high-resolution forest age data derived from time series Landsat images to improve forest biomass estimation. This kind of research is still in infancy, and more research is needed in near future to explore how to effectively couple the physical-based models with machine learning algorithms based on multi-source data.

3.6. The concerns of modeling performance and prediction accuracy

Evaluation of forest biomass estimation is mainly based on two broad aspects: modelling performance and prediction accuracy (Lu *et al*. 2016). Most studies tend to prioritise modelling performance while not taking predictive ability into account. However, a good modelling performance cannot guarantee a good prediction accuracy in a large area because of the representativeness of modelling samples and complex characteristics of land surface features in the study area. Modelling performance depends on the use of a sufficient number of representative samples, selection of proper variables and modelling algorithms (Lu *et al*. 2016). If sufficient samples are available for each forest type, the models developed for each type can provide better modelling performance than for a mixture of different forest types (Jiang *et al*. 2020b). However, in a large study area, a classification result for detailed forest types may be not available or the classification accuracy cannot meet the application purpose (Schlund *et al*. 2017, Jiang 2021). Accurate identification of forest types ahead of using the developed models to estimate forest biomass is needed to get an accurate unbiased estimation (Schlund *et al*. 2017). The forest classification accuracy may have considerable influence on the model prediction accuracy when the biomass models have different estimation performance for each forest type (Jiang 2021).

Sample plots having small biomass range and standard deviation often lead to poor modelling performance (Jiang *et al.* 2020b) and use of many variables generally results in a higher modelling R^2, indicating better model performance, especially when machine learning methods are used (Feng *et al.* 2017), but the prediction accuracy may be low because of the overfitting problem. Incorporation of additional variables with weak relationship with forest biomass may lead to the complexity of modelling (Vaglio Laurin *et al.* 2014, Huang *et al.* 2022) and poor modelling application. Feng *et al.* (2017) validated the prediction accuracy by iteratively removing low-importance variables and the results showed that as the number of variables decreased, the prediction accuracy remains stable until critical variables were removed. It is worth noting that machine learning algorithms, due to their susceptibility to overfitting, often exhibit higher modelling R^2 than linear regression models (Rex *et al.* 2020, Lin *et al.* 2022), but their predictive accuracy does not match their modelling performance. Furthermore, machine learning algorithms can be heavily influenced by the range of input samples, particularly in the case of random forests, where predictions are constrained within the range of input samples (Jiang *et al.* 2020b). Therefore, it is essential to strike a balance between modelling performance and predictive capabilities. In general, a model which is developed using a large number of samples with sufficient representativeness in a study area has both good modelling performance and prediction accuracy if detailed forest types for modelling are also accurately available for the study area.

3.7. The potential of using airborne lidar data in facilitating the integration of ground-level observations with satellite-based data

Airborne Lidar data with its high spatial resolution and ability to capture fine-scale forest stand structures (Dong and Chen 2018, Guo *et al.* 2018, Poorazimy *et al.* 2020) have been used as a bridge connecting ground-based sample plots with coarse-resolution remotely sensed data (Shin *et al.* 2016, Bouvier *et al.* 2019). The most common approach is to utilise airborne Lidar-based biomass map as both modelling and validation samples, addressing the spatial sampling limitations inherent in ground surveys (Puliti *et al.* 2018, Coops *et al.* 2021). Furthermore, this type of samples by allowing for adjustments in both size (e.g. 100–10,000 m^2) and shape (square, rectangle, circle) eliminates issues related to mismatches with image pixels (Jiang *et al.* 2020a, Michael *et al.* 2021). For instance, Jiang *et al.* (2020a) aggregated aboveground carbon density estimates of Lidar data to 500 m resolution cells (the same as the MODIS data) as reference and developed a regional-scale aboveground carbon density estimation model based on references and variables from MODIS datasets. In addition, considering spaceborne Lidar provide increased spatial and temporal coverage, some studies have linked local-scale airborne Lidar data to spaceborne Lidar data firstly to enable regional-scale biomass estimation (Michael *et al.* 2021). For active Lidar data, apart from directly utilising airborne Lidar-based biomass map as samples, an alternative involves using airborne Lidar data to simulate variables derived from spaceborne Lidar data for calibrating biomass models to reduce prediction error due to geolocational uncertainty (Laura *et al.* 2022). In general, use of the Lidar data has two major objectives: solve the large geometric differences between ground observations and coarse spatial resolution satellite data and increase the size and representativeness of samples in a large study area. More research is needed in the future on how to effectively use the Lidar-derived biomass data or directly use Lidar data as a bridge to link ground observations to satellite data.

4. Summary and perspectives

The needs of understanding carbon source or sink and the carbon sequestration require accurate biomass estimation in a timely manner. Airborne Lidar is one of the major data sources for accurate estimation of forest biomass at the local scale, thus, many studies have been conducted to explore the effective selection of modelling methods. This review summarises the general procedure of using airborne Lidar data for forest biomass estimation: data preparation, extraction and selection of modelling variables, modelling algorithms, and evaluation of modelling performance. Special emphasis is on the discussion of some important concerns related to improving biomass estimation, including:

(1) improving the quality of biomass samples by effective use of airborne Lidar data;
(2) forest biomass modelling based on proper stratification strategies and use of universal models such as mixed-effects models and HBA;
(3) effective incorporation of multiple source data in a biomass modelling procedure;
(4) the role of deep learning approach in improving forest biomass estimation;
(5) the need to develop models with good transferability at temporal and spatial scales;
(6) the need for a balance between modelling performance and predictive accuracy;
(7) the potential roles of using Lidar as a bridge between ground observations and coarse resolution satellite data.

More research is needed to incorporate Lidar and other data sources for developing advanced biomass estimation models, explore the integration of physical-based model and machine learning methods to improve biomass estimation, and to identify the major factors influencing forest biomass estimation accuracy through uncertainty analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Natural Science Foundation of China [32271870], Fujian Provincial Science and Technology Department [2021R1002008] and China Scholarship Council [202108350072].

References

Ahmad, A., Gilani, H., and Ahmad, S.R., 2021. Forest aboveground biomass estimation and mapping through high-resolution optical satellite imagery—A literature review. *Forests*, 12 (7), 914. doi:10.3390/f12070914.

Anderson, J., *et al.*, 2008. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, 112 (4), 1856–1870. doi:10.1016/j.rse.2007.09.009.

Araza, A., *et al*., 2022. A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps. *Remote Sensing of Environment*, 272, 112917. doi:10.1016/j.rse.2022.112917

Berninger, A., *et al*., 2019. Canopy height and above-ground biomass retrieval in tropical forests using multi-pass X- and C-band pol-InSAR data. *Remote Sensing*, 11 (18), 2105. doi:10.3390/rs11182105.

Bouvier, M., *et al*. 2019. Influence of sampling design parameters on biomass predictions derived from airborne lidar data. *Canadian Journal of Remote Sensing*, 45 (5), 650–672. doi:10.1080/07038992.2019.1669013

Cao, L., *et al*., 2014. Using small-footprint discrete and full-waveform airborne lidar metrics to estimate total biomass and biomass components in subtropical forests. *Remote Sensing*, 6 (8), 7110–7135. doi:10.3390/rs6087110.

Cao, L., *et al*., 2016. Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. *Remote Sensing of Environment*, 178, 158–171. doi:10.1016/j.rse.2016.03.012

Cao, L., *et al*., 2019. Estimating canopy structure and biomass in bamboo forests using airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 114–129. doi:10.1016/j.isprsjprs.2018.12.006

Catherine, T., *et al*., 2019. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian amazon using different regression algorithms. *Remote Sensing of Environment*, 232, 111323. doi:10.1016/j.rse.2019.111323.

Chen, Q., *et al*., 2007. Estimating basal area and stem volume for individual trees from Lidar data. *Photogrammetric Engineering and Remote Sensing*, 73 (12), 1355–1365. doi:10.14358/PERS.73.12.1355.

Chen, Q., *et al*., 2012. Integration of airborne Lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. *Remote Sensing of Environment*, 121, 108–117. doi:10.1016/j.rse.2012.01.021

Chen, Q., 2013. Lidar remote sensing of vegetation biomass. *In*: Q. Weng and G. Wang, eds. *Remote sensing of natural resources*. Boca Raton: CRC Press: Taylor & Francis Group, pp. 399–420.

Chen, Y., *et al*. 2018. Exploring bamboo forest aboveground biomass estimation using sentinel-2 data. *Remote Sensing*, 11 (1), 7. doi:10.3390/rs11010007

Chi, H., *et al*. 2015. National forest aboveground biomass mapping from ICESat/GLAS data and MODIS imagery in China. *Remote Sensing*, 7 (5), 5534–5564. doi:10.3390/rs70505534

Christina, E., Claudia, K., and Stefan, D., 2012. Derivation of biomass information for semi-arid areas using remote-sensing data. *International Journal of Remote Sensing*, 33 (9), 2937–2984. doi:10.1080/01431161.2011.620034.

Chung, C., *et al*., 2019. Comparison of forest canopy height profiles in a mountainous region of Taiwan derived from airborne lidar and unmanned aerial vehicle imagery. *GIScience & Remote Sensing*, 56 (8), 1289–1304. doi:10.1080/15481603.2019.1627044.

Clark, M.L., *et al*., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*, 115 (11), 2931–2942. doi:10.1016/j.rse.2010.08.029.

Coops, N., *et al*., 2021. Modelling lidar-derived estimates of forest attributes over space and time: a review of approaches and future trends. *Remote Sensing of Environment*, 260, 112477. doi:10.1016/j.rse.2021.112477

Crawford, C.J., *et al*., 2023. The 50-year landsat collection 2 archive. *Science of Remote Sensing*, 8, 100103. doi:10.1016/j.srs.2023.100103

Dai, W., *et al*., 2019. Automated fusion of forest airborne and terrestrial point clouds through canopy density analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 156, 94–107. doi:10.1016/j.isprsjprs.2019.08.008

Domingo, D., *et al*. 2019. Temporal transferability of pine forest attributes modeling using low-density airborne laser scanning data. *Remote Sensing*, 11 (3), 261. doi:10.3390/rs11030261

Dong, P. and Chen, Q., 2018. *LiDAR remote sensing and applications*. New York, USA: CRC Press, Taylor & Francis Group.

Du, C., *et al*., 2021. The effect of synergistic approaches of features and ensemble learning algorithms on aboveground biomass estimation of natural secondary forests based on ALS and Landsat8. *Sensors*, 21 (17), 5974. doi:10.3390/s21175974.

Du, L., *et al*., 2023. A LiDAR biomass index-based approach for tree- and plot-level biomass mapping over forest farms using 3D point clouds. *Remote Sensing of Environment*, 290, 113543. doi:10.1016/j.rse.2023.113543

Duncanson, L., *et al*., 2020. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sensing of Environment*, 242, 111779. doi:10.1016/j.rse.2020.111779

Erb, K.H., *et al*., 2018. Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature*, 553 (7686), 73–76. doi:10.1038/nature25138.

Ewijk, K., *et al*. 2020. Transferability of ALS-derived forest resource inventory attributes between an eastern and western Canadian boreal forest mixed wood site. *Canadian Journal of Remote Sensing*, 46 (2), 214–236. doi:10.1080/07038992.2020.1769470

Fang, G., *et al*., 2022. Comparison of variable selection methods among dominant tree species in different regions on forest stock volume estimation. *Forests*, 13 (5), 787. doi:10.3390/f13050787.

Fassnacht, F.E., *et al*., 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment*, 154, 102–114. doi:10.1016/j.rse.2014.07.028

Fekety, P.A., *et al*., 2018. Transferability of Lidar-derived basal area and stem density models within a Northern Idaho ecoregion. *Canadian Journal of Remote Sensing*, 44 (2), 131–143. 1461557. doi:10.1080/07038992.2018.1461557

Fekety, P., Falkowski, M., and Hudak, A., 2015. Temporal transferability of Lidar-based imputation of forest inventory attributes. *Canadian Journal of Forest Research*, 45 (4), 422–435. doi:10.1139/cjfr-2014-0405.

Feng, H., *et al*. 2021. Estimation of forest aboveground biomass by using mixed-effects model. *International Journal of Remote Sensing*, 42 (22), 8675–8690. doi:10.1080/01431161.2021.1984611

Feng, Y., *et al*. 2017. Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian amazon. *International Journal of Digital Earth*, 10 (10), 996–1016. doi:10.1080/17538947.2017.1301581

Frazer, G.W., *et al*. 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment*, 115 (2), 636–649. doi:10.1016/j.rse.2010.10.008

Gao, Y., *et al*., 2018. Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. *Remote Sensing*, 10 (4), 627. doi:10.3390/rs10040627.

Gao, L., Chai, G., and Zhang, X., 2022. Above-ground biomass estimation of plantation with different tree species using airborne lidar and hyperspectral data. *Remote Sensing*, 14 (11), 2568. doi:10.3390/rs14112568.

Ghosh, S.M. and Behera, M.D., 2021. Aboveground biomass estimates of tropical mangrove forest using sentinel-1 SAR coherence data - the superiority of deep learning over a semi-empirical model. *Computers & Geosciences*, 150, 104737. doi:10.1016/j.cageo.2021.104737

Görgens, E., *et al*., 2015. Stand volume models based on stable metrics as from multiple ALS acquisitions in eucalyptus plantations. *Annals of Forest Science*, 72 (4), 489–498. doi:10.1007/s13595-015-0457-x.

Guo, Q., *et al*., 2018. *Lidar principles, processing and applications in forest ecology*. ISBN 9787040493016In Chinese. Beijing, China: Higher Education Press.

Heiskanen, J., *et al*., 2019. Do airborne laser scanning biomass prediction models benefit from landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes? *International Journal of Applied Earth Observation and Geoinformation*, 81, 176–185. doi:10.1016/j.jag.2019.05.017

Huang, Z., *et al*. 2022. Estimating mangrove above-ground biomass at Maowei Sea, Beibu Gulf of China using machine learning algorithm with sentinel-1 and sentinel-2 data. *Geocarto International*, 37 (27), 15778–15805. doi:10.1080/10106049.2022.2102226

Jiang, X., *et al*., 2020a. Modeling forest aboveground carbon density in the Brazilian amazon with integration of MODIS and airborne lidar data. *Remote Sensing*, 12 (20), 3330. doi:10.3390/rs12203330.

Jiang, X., *et al*., 2020b. Stratification-based forest aboveground biomass estimation in a subtropical region using airborne lidar data. *Remote Sensing*, 12 (7), 1101. doi:10.3390/rs12071101.

Jiang, X., 2021. Forest classification and biomass estimation based on multi-source data in a typical subtropical region [D]. *Zhejiang Agriculture & Forest University*. doi:10.27756/d.cnki.gzjlx.2020.000274.

Jiang, X., *et al*., 2023. Eucalyptus carbon stock estimation in subtropical regions with the modeling strategy of sample plots – airborne LiDAR – landsat time series data. *Forest Ecosystems*, 10, 100149. doi:10.1016/j.fecs.2023.100149

Jurjević, L., *et al*., 2020. Is field-measured tree height as reliable as believed – part II, a comparison study of tree height estimates from conventional field measurement and low-cost close-range remote sensing in a deciduous forest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 227–241. doi:10.1016/j.isprsjprs.2020.09.014

Laura, D., *et al*., 2022. Aboveground biomass density models for nasa's global ecosystem dynamics investigation (GEDI) lidar mission. *Remote Sensing of Environment*, 270, 112845. doi:10.1016/j.rse.2021.112845.

Li, G., *et al*., 2012. A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, 26–38. doi:10.1016/j.isprsjprs.2012.03.010

Li, G., *et al*., 2019. Integration of ZiYuan-3 multispectral and stereo data for modeling aboveground biomass of larch plantations in North China. *Remote Sensing*, 11 (19), 2328. doi:10.3390/rs11192328.

Liang, M., *et al*., 2023. Quantifying aboveground biomass dynamics from charcoal degradation in Mozambique using GEDI lidar and landsat. *Remote Sensing of Environment*, 284, 113367. doi:10.1016/j.rse.2022.113367

Liang, X., *et al*., 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, 137–179. doi:10.1016/j.isprsjprs.2018.06.021

Liao, K., *et al*., 2022. Examining the role of UAV Lidar data in improving tree volume calculation accuracy. *Remote Sensing*, 14 (17), 4410. doi:10.3390/rs14174410.

Lin, W., *et al*., 2022. A comparative analysis of modeling approaches and canopy height-based data sources for mapping forest growing stock volume in a northern subtropical ecosystem of China. *GIScience & Remote Sensing*, 59 (1), 568–589. doi:10.1080/15481603.2022.2044139.

Li, Z., Qingwang, L., and Yong, P., 2016. Review on forest parameters inversion using LiDAR. *National Remote Sensing Bulletin*, 20 (5), 1138–1150. doi:10.11834/jrs.20165130.

Longo, M., *et al*., 2016. Aboveground biomass variability across intact and degraded forests in the Brazilian amazon. *Global Biogeochemical Cycles*, 30 (11), 1639–1660. doi:10.1002/2016GB005465.

Lu, D., 2005. Aboveground biomass estimation using landsat TM data in the Brazilianamazon. *International Journal of Remote Sensing*, 26 (12), 2509–2525. doi:10.1080/01431160500142145.

Lu, D., 2006. The potential and challenge of remote sensing–based biomass estimation. *International Journal of Remote Sensing*, 27 (7), 1297–1328. doi:10.1080/01431160500486732.

Lu, D., *et al*., 2012. Aboveground forest biomass estimation with landsat and LiDAR data and uncertainty analysis of the estimates. *International Journal of Forestry Research*, 2012, 1–16. doi:10.1155/2012/436537.

Lu, D., *et al*. 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, 9 (1), 63–105. doi:10.1080/17538947.2014.990526

Lu, D., Batistella, M., and Moran, E., 2005. Satellite estimation of aboveground biomass and impacts of forest stand structure. *Photogrammetric Engineering and Remote Sensing*, 71 (8), 967–974. doi:10.14358/PERS.71.8.967.

Luo, S., *et al*., 2017. Fusion of airborne LiDAR data and hyperspectral imagery for aboveground and belowground forest biomass estimation. *Ecological Indicators*, 73, 378–387. doi:10.1016/j.ecolind.2016.10.001

Marchesan, J., *et al*., 2023. Aboveground biomass stock and change estimation in amazon rainforest using airborne light detection and ranging, multispectral data, and machine learning algorithms. *Journal of Applied Remote Sensing*, 17 (2). doi:10.1117/1.JRS.17.024509.

McGaughey, R.J., 2009. *FUSION/LDV: software for LIDAR data analysis and visualization*. Seattle, WA: USDA Forest Service. Pacific Northwest Research Station.

Michael, J., *et al*., 2021. Scaled biomass estimation in woodland ecosystems: testing the individual and combined capacities of satellite multispectral and lidar data. *Remote Sensing of Environment*, 262, 112511. doi:10.1016/j.rse.2021.112511.

Mo, L., *et al*., 2023. Integrated global assessment of the natural forest carbon potential. *Nature*, 624 (7990), 92–101. doi:10.1038/s41586-023-06723-z.

Molina, P.X., *et al*., 2016. Spatially-explicit testing of a general aboveground carbon density estimation model in a western amazonian forest using airborne lidar. *Remote Sensing*, 8 (1), 9. doi:10.3390/rs8010009.

Naesset, E. and Gobakken, T., 2008. Estimation of above and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112 (6), 3079–3090. doi:10. 1016/j.rse.2008.03004.

Navarro, J., *et al*., 2020. Assessing the transferability of airborne laser scanning and digital aerial photogrammetry derived growing stock volume models. *International Journal of Applied Earth Observation and Geoinformation*, 91, 102135. doi:10.1016/j.jag.2020.102135

Nie, S., *et al*., 2017. Above-ground biomass estimation using airborne discrete-return and full-waveform LiDAR data in a coniferous forest. *Ecological Indicators*, 78, 221–228. doi:10.1016/j.ecolind.2017.02.045

Passel, J.V., Keersmaecker, W.D., and Somers, B., 2020. Monitoring woody cover dynamics in tropical dry forest ecosystems using sentinel-2 satellite imagery. *Remote Sensing*, 12 (8), 1276. doi:10.3390/rs12081276.

Patrick, A., *et al*., 2018. Transferability of lidar-derived basal area and stem density models within a northern Idaho ecoregion. *Canadian Journal of Remote Sensing*, 44 (2), 131–143. doi:10.1080/07038992.2018.1461557.

Phalla, T., *et al*., 2017. The importance of tree height in estimating individual tree biomass while considering errors in measurements and allometric models. *AGRIVITA Journal of Agricultural Science*, 40 (1), 131–140. doi:10.17503/agrivita.v40i1.1730.

Phua, M., *et al*., 2016. Estimating logged-over lowland rainforest aboveground biomass in Sabah, Malaysia using airborne lidar data. *Terrestrial Atmospheric and Oceanic Sciences*, 27 (4), 481–489. doi:10.3319/TAO.2016.01.06.02.

Poorazimy, M., *et al*., 2020. Integrating airborne laser scanning data, space-borne radar data and digital aerial imagery to estimate aboveground carbon stock in hyrcanian forests, Iran. *Remote Sensing of Environment*, 240, 111669. doi:10.1016/j.rse.2020.111669

Puliti, S., *et al*., 2018. Combining UAV and sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. *Remote Sensing of Environment*, 204, 485–497. doi:10.1016/j.rse.2017.10.007

Rex, F.E., *et al*., 2020. Comparison of statistical modelling approaches for estimating tropical forest aboveground biomass stock and reporting their changes in low-intensity logging areas using multi-temporal lidar data. *Remote Sensing*, 12 (9), 1498. 1498. doi:10.3390/rs12091498

Roussel, J.R., *et al*., 2017. Removing bias from LiDAR-based estimates of canopy height: accounting for the effects of pulse density and footprint size. *Remote Sensing of Environment*, 198, 1–16. doi:10.1016/j.rse.2017.05.032

Ruusa, M., Rosser, N.J., and Donoghue, D.N.M., 2022. Improving above ground biomass estimates of Southern Africa dryland forests by combining sentinel-1 SAR and sentinel-2 multispectral imagery. *Remote Sensing of Environment*, 282, 113232. doi:10.1016/j.rse.2022.113232

Ryan, S. and Paul, V., 2013. Estimating aboveground biomass and average annual wood biomass increment with airborne leaf-on and leaf-off lidar in great lakes forest types. *Northern Journal of Applied Forestry*, 30 (1), 16–22. doi:10.5849/njaf.12-015.

Schlund, M., Scipal, K., and Davidson, M., 2017. Forest classification and impact of BIOMASS resolution on forest area and aboveground biomass estimation. *International Journal of Applied Earth Observation and Geoinformation*, 56, 65–76. doi:10.1016/j.jag.2016.12.001

Shen, H. and Zhang, L., 2023. Earth surface feature parameter inversion and simulation: a mechanistic-learning coupling paradigm. *China Science: Earth Science*, 53 (3), 546–560. doi:10.1360/SSTe-2022-0089.

Shin, J., *et al*. 2016. Comparing modeling methods for predicting forest attributes using LiDAR metrics and ground measurements. *Canadian Journal of Remote Sensing*, 42 (6), 739–765. doi:10.1080/07038992.2016.1252908

Silveira, E., *et al*., 2019. Object-based random forest modelling of aboveground forest biomass outperforms a pixel-based approach in a heterogeneous and mountain tropical environment. *International Journal of Applied Earth Observation and Geoinformation*, 78, 175–188. doi:10.1016/j.jag.2019.02.004

Somayeh, T. and Ali, S., 2022. Global-scale biomass estimation based on machine learning and deep learning methods. *Remote Sensing Applications: Society & Environment*, 28, 100868. doi:10.1016/j.rsase.2022.100868

Sousa, C., *et al*., 2023. Size of sampling units appropriate for management level forest inventory in Eastern Amazon, Brazil. *Revista em Agronegócio e Meio Ambiente*, 16 (2), 1–13. doi:10.17765/2176-9168.2023v16n2e10938.

Stereńczak, K., *et al*., 2020. Global airborne laser scanning data providers database (GlobALS)-A new tool for monitoring ecosystems and biodiversity. *Remote Sensing*, 12 (11), 1877. doi:10.3390/rs12111877.

Tamga, K., *et al*., 2023. Estimation of aboveground biomass in agroforestry systems over three climatic regions in West Africa using sentinel-1, sentinel-2, ALOS, and GEDI data. *Sensors*, 23 (1), 349. doi:10.3390/s23010349.

Tamiminia, H., *et al*., 2022. Evaluating pixel-based and object-based approaches for forest aboveground biomass estimation using a combination of optical, SAR, and an extreme gradient boosting model. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 485–492. doi:10.5194/isprs-annals-V-3-2022-485-2022.

Tang, Z., *et al*., 2022. Estimation of national forest aboveground biomass from multi-source remotely sensed dataset with machine learning algorithms in China. *Remote Sensing*, 14 (21), 5487. doi:10.3390/rs14215487.

Tian, L., *et al*. 2023. Review of remote sensing-based methods for forest aboveground biomass estimation: progress, challenges, and prospects. *Forests*, 14 (6), 1086. doi:10.3390/f14061086

Tompalski, P., *et al*. 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. *Remote Sensing of Environment*, 227, 110–124. doi:10.1016/j.rse.2019.04.006 [006].

Vaglio Laurin, G., *et al*., 2014. Above ground biomass estimation in an African tropical forest with Lidar and hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 89, 49–58. doi:10.1016/j.isprsjprs.2014.01.001

Walker, W.S., *et al*., 2022. The global potential for increased storage of carbon on land. *PNAS*, 119 (23), e2111312119. doi:10.1073/pnas.2111312119.

Wang, M., *et al*. 2019a. Airborne Lidar-derived aboveground biomass estimates using a hierarchical Bayesian approach. *Remote Sensing*, 11 (9), 1050. doi:10.3390/rs11091050

Wang, Y., *et al*., 2019b. Is field-measured tree height as reliable as believed – a comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 132–145. doi:10.1016/j.isprsjprs.2018.11.008

Wang, R., *et al*., 2023. A comparative analysis of grid-based and object-based modeling approaches for poplar forest growing stock volume estimation in plain regions using airborne LIDAR data. *Geo-Spatial Information Science*, 1–19. doi:10.1080/10095020.2023.2169199.

Wang, C. and Blei, D.M., 2018. A general method for robust Bayesian modeling. *Bayesian Analysis*, 13 (4), 1163–1191. doi:10.1214/17-BA1090.

Werkowska, W., *et al*. 2017. A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25 (1), 127–133. doi:10.1139/er2016-0045

Xiao, J., *et al*., 2019. Remote sensing of the terrestrial carbon cycle: a review of advances over 50 years. *Remote Sensing of Environment*, 233, 111383. doi:10.1016/j.rse.2019.111383

Xie, D., *et al*., 2023. Aboveground biomass prediction of arid shrub-dominated community based on airborne Lidar through parametric and nonparametric methods. *Remote Sensing*, 15 (13), 3344. doi:10.3390/rs15133344.

Xu, T., *et al*., 2015. Estimates of subtropical forest biomass based on airborne LiDAR and landsat 8 OLI data. *Chinese Journal of Plant Ecology*, 39 (4), 309–321. doi:10.17521/cjpe.2015.0030.

Yang, M., *et al*., 2023. Developing allometric equations to estimate forest biomass for tree species categories based on phylogenetic relationships. *Forest Ecosystems*, 10, 100130. doi:10.1016/j.fecs.2023.100130

Zeng, W., Chen, X., and Yang, X., 2023. Estimating changes of forest carbon storage in China for 70 years (1949–2018). *Scientific Reports*, 13 (1), 16864. doi:10.1038/s41598-023-44097-4.

Zhang, L., *et al*. 2019. Deep learning based retrieval of forest aboveground biomass from combined Lidar and landsat 8 data. *Remote Sensing*, 11 (12), 1459. doi:10.3390/rs11121459

Zhang, Y., *et al*., 2023. Forest structure simulation of eucalyptus plantation using remote-sensing-based forest age data and 3-PG model. *Remote Sensing*, 15 (1), 183. doi:10.3390/rs15010183.

Zhao, P., *et al*., 2016a. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sensing*, 8 (6), 469. doi:10.3390/rs8060469.

Zhao, P., *et al*., 2016b. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *International Journal of Applied Earth Observation and Geoinformation*, 53, 1–15. doi:10.1016/j.jag.2016.08.007

Zhao, X., *et al*., 2016. Improved progressive TIN densification filtering algorithm for airborne LiDAR data in forested areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 79–91. doi:10.1016/j.isprsjprs.2016.03.016

Zhao, Y., *et al*., 2023. Towards accurate individual tree parameters estimation in dense forest: optimized coarse-to-fine algorithms for registering UAV and terrestrial LiDAR data. *GIScience & Remote Sensing*, 60 (1), 2197281. doi:10.1080/15481603.2023.2197281.