

Deep learning tree and forest biomass from sub-meter resolution images

Sizhuo Li (✉ sizli@ign.ku.dk)

University of Copenhagen <https://orcid.org/0000-0001-9583-7230>

Martin Brandt

University of Copenhagen

Xiaoye Tong

University of Copenhagen

Stefan Oehmcke

University of Copenhagen

Christian Igel

University of Copenhagen

Fabian Gieseke

University of Münster

Thomas Nord-Larsen

University of Copenhagen

Rasmus Fensholt

University of Copenhagen

Jerome Chave

CNRS, UPS, IRD, Université Paul Sabatier

Philippe Ciais

Laboratoire des Sciences du Climat et de l'Environnement <https://orcid.org/0000-0001-8560-4943>

Article

Keywords:

Posted Date: September 14th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3335298/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Deep learning tree and forest biomass from sub-meter resolution images

Sizhuo Li^{*1}, Martin Brandt^{*2}, Xiaoye Tong³, Stefan Oehmcke⁴, Christian Igel⁵, Fabian Gieseke⁶, Thomas Nord-Larsen⁷, Rasmus Fensholt⁸, Jerome Chave⁹, Philippe Ciais¹⁰

¹Department of Geosciences and Natural Resource Management, University of Copenhagen, Université Paris-Saclay

^{2,3,7,8}Department of Geosciences and Natural Resource Management, University of Copenhagen

^{4,5}Department of Computer Science, University of Copenhagen

⁶Department of Information Systems, University of Münster

⁹Laboratoire Evolution et Diversité Biologique, CNRS, UPS, IRD, Université Paul Sabatier

¹⁰Laboratoire des Sciences du Climat et de l'Environnement, CEA, CNRS, UVSQ, Université Paris-Saclay

*Sizhuo Li or Martin Brandt

Emails: sizli@ign.ku.dk or mabr@ign.ku.dk

Abstract

Computational visual intelligence has been shown to be able to comprehend the content of images, which has been widely used to foster a digitized society, but is often underutilized in applications related to the green transition and climate change mitigation. Here, we evaluate the capacity of convolutional neural networks (CNN) to interpret spatial semantic patterns in optical RGB images to directly estimate forest biomass, an essential climate parameter previously assessed from structural measures of trees. Trained with forest inventory plots, the CNN model demonstrates its learning via interpreting the composition of biomass at tree level, differing from traditional approaches reliant on conversions of aggregated parameters without explanatory rationale. The CNN approach yields consistently low bias across wide biomass ranges, whereas traditional models show insufficiency without information on tree height. Visually interpretable models link advanced computational tools with the power of data, facilitating the sustainable management of resources for a carbon-neutral society.

1 Introduction

Forests represent a major sink of atmospheric carbon dioxide [1]. To mitigate climate change, sustainable forest management strategies aiming to maintain or increase forest carbon storage have been proposed [2, 3, 4], supported by international conventions such as the United Nations Framework Convention on Climate Change and the Sustainable Development Goals [5]. These programs require monitoring, reporting, and verification of tree and forest carbon stocks to ensure that countries are meeting their emission reduction commitments from forest conservation and restoration efforts [6, 7]. The quantification of forest biomass is also crucial for the implementation of effective management strategies related to the green transition and the maintenance or enhancement of ecosystem services, such as forest productivity and biodiversity [8, 9]. Traditional ground-based surveys and field measurements for forest biomass reporting are accurate yet remain time-consuming and labor intensive [10, 11], and thus effective, rapid, and scalable approaches are needed to complement field-based inventories [12].

Forests vary in regards to tree species composition, tree size, density, and age, all of which can influence the carbon storage capacity of a forest [8, 13, 14]. Remote sensing technologies can only partially capture the spatial variations in forest above-ground biomass (AGB) due to the trade-offs between sensor characteristics, costs, and data availability [15, 16, 17, 18]. High-precision ground-based sensors such as terrestrial laser scanning instruments provide accurate estimates of wood volume, a proxy for AGB, by measuring explicitly the 3D architecture of each tree with millimeter-level of details [17, 19]. However, these approaches are less amenable to upscaling because they rely on fieldwork by trained operators. Airborne Light Detection and Ranging (LiDAR) provides a reliable alternative to upscale biomass measurements directly over larger areas [20], or derive biomass from allometric relationships with tree height [21]. Data from public or commercial optical sensors trained with airborne or spaceborne LiDAR enable height estimation from optical images, which is considered a promising way to estimate biomass over large areas without the need for additional LiDAR data [22, 23, 24]. Yet, most previous approaches are based on conversions of parameters, such as height, reflectance, or radar backscatter, which are typically aggregated to match field plots before relationships between remotely sensed parameters and field-measured biomass are established at the plot level. Results typically have a moderate spatial resolution, inadequately representing the spatial variations in biomass within forests, and often cannot detect trees outside

of forests, such as trees in urban areas or farmlands, which can contribute up to a quarter of national tree biomass for many countries [23].

Recently, high-resolution optical images with spatial resolutions higher than 3 m and advancements in deep learning have made it feasible to detect individual tree crowns and estimate their height for both forested and non-forested landscapes at various scales [25, 26, 27, 28]. These tree-level variables can be converted into AGB using allometric equations established from field measurements [11, 29, 30, 31]. However, the allometric coefficients depend highly on species-specific parameters invisible from airborne sensors, such as wood density [32], and the optimal mathematical formulas reflecting the true relationships between airborne tree variables and AGB are hard to verify [33, 34]. Uncertainties may further arise when upscaling allometric equations to remotely detected tree features, which may deviate substantially from on-ground measurements that align with the established allometric relationships [35]. Besides, optical imagery-based tree-level methods only capture trees visible from an aerial perspective, ignoring all understory that can contribute a considerable part to the total tree count [25]. Consequently, errors at the tree or plot level can be high. Moreover, the availability of LiDAR is not always given, and the reliability of using tree height estimated from optical images as an indicator of biomass remains unclear.

Considering that aerial and satellite images with sub-meter resolution are increasingly available and the image quality has reached a visually interpretable level in relation to information on tree structure, a direct estimation of AGB from the image itself without the intermediate parameter extraction may lead to higher performance by eliminating the information loss caused by the transformations. Convolutional Neural Networks (CNNs) are well-known for their ability to parse spatial context information using convolutional kernels [36, 37], and have been applied to similar real-world applications such as estimating age from human face images [38], food ripeness from photos [39], and house prices from street view and satellite images [40]. However, multi-step feature- or tree-segmentation-based approaches have so far dominated the AGB estimation research at plot or tree scales in the remote sensing field and a direct mapping from images to forest biomass has rarely been tested.

The overall aim of the study is to utilize CNNs to infer AGB, normally assessed from complex field measurements, directly from sub-meter resolution optical images, without any domain-knowledge-based modeling. We train the model using field plots from the Danish National Forest Inventory (NFI) with aggregated plot-level AGB calculated from non-destructive in-situ tree measurements [41], and visualize model learning to infer biomass from the optical red-green-blue (RGB) images. We compare the CNN models with a conventional Random Forest (RF) model, which estimates AGB using handcrafted plot-level features. Besides, we assess the performance variations when incorporating LiDAR-derived height or height estimated from optical images as additional inputs. Furthermore, results are compared with tree-level approaches that use allometric relationships to convert remotely detected crown diameter and height to biomass. Tree-level approaches are relevant in cases where information on single trees is required, such as in restoration areas or regions with sparse tree cover. We apply generalized allometric equations from the literature [29] and additionally propose an equation-free method based on a smooth monotonic neural network (SMNN) [42], which correlates airborne-derived individual tree crown segments directly with plot-based field-measured AGB. Here the aim is to develop a trainable allometric model that provides more flexibility than conventional fixed-form allometric equations [29], thereby eliminating the need for a mathematical formulation of allometric relationships.

2 Results

2.1 Convolutional Neural Networks Infer AGB from Optical Images

In this study, we employed CNNs to infer AGB from high-resolution optical images. We utilized a total of 2,423 Danish NFI field plots (15 m radius) [41] with plot-aggregated AGB as the reference data, which were split into training, validation, and a standalone test set with a ratio of 8:1:1. Optical aerial images captured during the summer of 2018 at a spatial resolution of 20 cm were used as the inputs. Three efficient CNN architectures were selected as spatial feature extractors, which were subsequently connected with two different designs of regression heads (Figure 1.a, see Methods). Regression head A performed a series of non-linear transformations to produce the plot-level AGB, while regression head B generated an activation map reflecting the localized distribution of AGB (Figure 2). In the latter case, the model learned to infer the localized pixel contribution of AGB by plot-level supervision. We trained the CNNs using the training and validation sets with the plot-level AGB labels from NFI, and we evaluated the models using the test data.

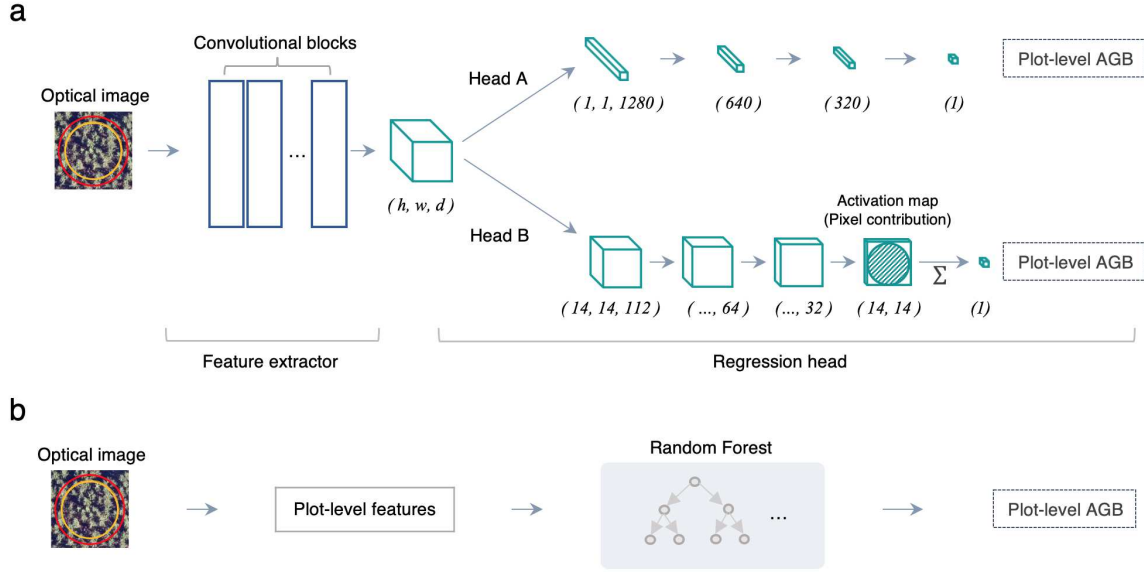


Figure 1: Overview of the plot-level approaches including our proposed Convolutional Neural Network (CNN) and a standard Random Forest (RF) model. a. Architecture of the CNN for above-ground biomass (AGB) estimation from optical images. The model takes optical images as inputs and outputs plot-level AGB estimates. The border of the field plot is demonstrated with the yellow circle (15 m radius) in the input image and the red circle displays the buffered (3 m) area which is seen by the model for enhanced neighborhood context. Two different designs of regression heads are tested. In design head B, the activation map reflects the estimated pixel contribution of AGB. b. Demonstration of the RF approach. The hand-crafted plot-level features are used for the RF regression.

Different model architecture designs led to slightly varying testing performances, with coefficient of determination (R^2) scores, calculated between the estimated and referential AGB, ranging between 0.62 and 0.71, and relative mean absolute errors (rMAE) in the range of 34.71% to 38.79% (Table E1). The activation map visualizes the learned relationship between trees and AGB at a sub-plot level at roughly 2 m ground resolution, allowing for an enhanced interpretation of the deep learning process (Figures 2.a-b). The model demonstrated a tendency to estimate high pixel contributions of AGB for dense or large trees, and lower AGB for gaps, which agreed with human cognition and the expected sub-plot distribution of AGB. The sum of all pixel values for each activation map corresponded to the total AGB estimate for the plot (Figures 2.a and 2.c). CNNs commonly rely on high-resolution input images to interpret spatial patterns [43, 37]. To simulate the model performance induced by coarser input resolutions, we down-sampled the images to different resolutions and established separate models accordingly (Figure 2.c and Table E2). The plot-level R^2 score exhibited a rapid drop, falling under 0.5 when the input pixel size was set above 80 cm (Figure 2.d). Degrading the input resolution resulted in a skewed shift in the estimated AGB distribution, where estimates tended to center around the average AGB (100-200 $\text{Mg}\cdot\text{ha}^{-1}$, Figure 2.e) rather than following a similar distribution as the NFI references. With a coarser input resolution and diminishing spatial contextual details, the activation maps tended to diverge from human interpretation, and the visual saliency faded away, leading to more blurred representations (Figure 2.c).

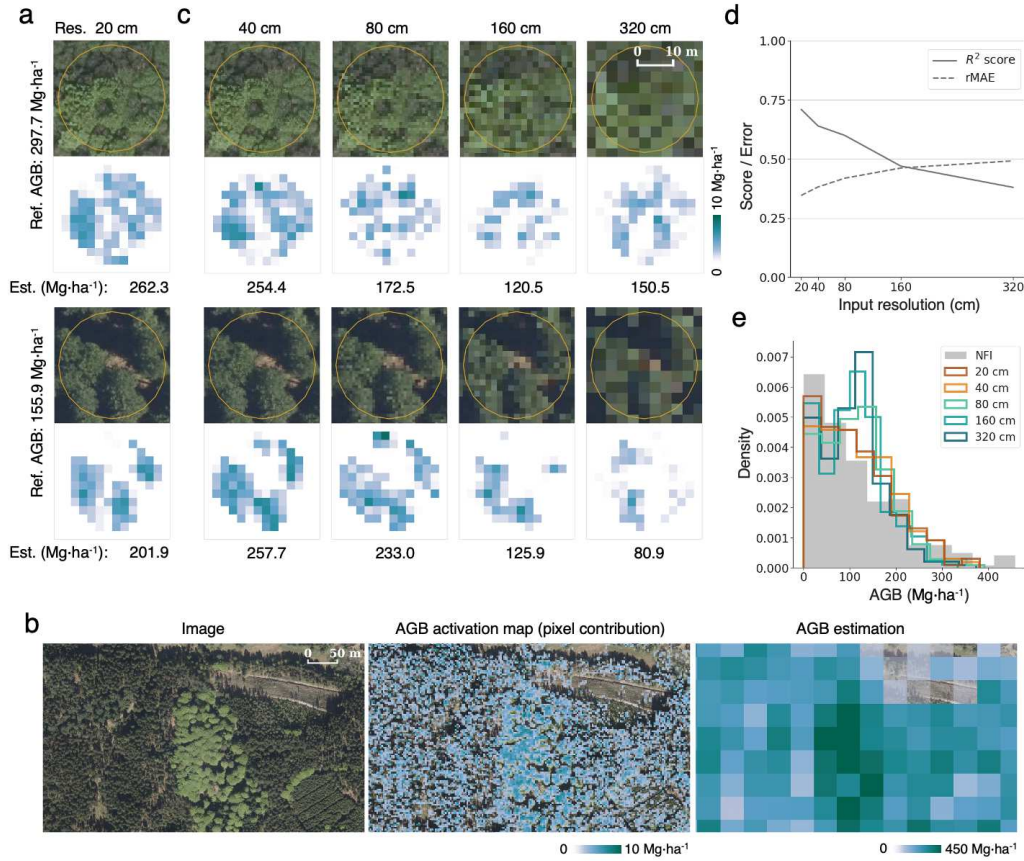


Figure 2: Localized pixel contribution of AGB estimation and performance variation with a decreasing input resolution (Res.). a. Examples of activation maps generated from the CNN model using 20 cm-resolution optical images. The pixel color represents the localized pixel contribution of AGB at roughly 2 m ground resolution. Examples were sampled from the test set. The yellow circles represent the borders of National Forest Inventory (NFI) plots. The reference (Ref.) and estimated (Est.) AGB at the plot level are shown alongside. b. Example of the AGB activation map (second column) and the AGB estimation map (third column) for a mixed-type forest area. The AGB values of the activation map were spatially aggregated to produce the final AGB estimation map at 40 m resolution. c. Activation maps generated from optical images of different pixel resolutions ranging from 40 cm to 320 cm. d. Changes in plot-level R^2 scores and rMAE with an increasing input pixel size evaluated on the test data. e. Histograms of AGB estimates on the test set using separate models with different input resolutions from 20 cm to 320 cm. The histogram of the reference AGB from NFI is shown in gray color.

We used the same data split to compare results from the best-performing CNN design with a Random Forest regression model (Figure 3), which estimated plot-level AGB using hand-crafted features extracted from the optical imagery, including band statistics, vegetation indices, histograms, and texture properties (Figure 1.b and Table.E3). We focused on two major aspects of the model performance: the model's capacity of capturing the spatial variation of AGB, and the systematic bias when aggregated beyond plot scales. Compared with the CNN model, which captured the variation of AGB across landscapes by interpreting the spatial patterns, the RF model generated average values with limited variation, attaining an R^2 score of 0.3 with up to 58.09% rMAE (Figures 3.a and 3.c). Absolute errors of the estimates from CNN were significantly lower than those from RF (Mann-Whitney U test, $n = 258$, $p < 0.001$). To evaluate the systematic bias when aggregated to larger scales, we categorized the test plots into AGB groups with 50 Mg·ha⁻¹ intervals and calculated the bias within each group (Figure 3.d). The CNN model consistently maintained relatively low bias across a wide range of AGB values, achieving lower than 12% bias for AGB values between 100 and 250 Mg·ha⁻¹ and lower than 29% bias for AGB > 50 Mg·ha⁻¹. In contrast, the RF model was remarkably biased (35%-330%) on low- (< 100 Mg·ha⁻¹) or high-value (> 200 Mg·ha⁻¹) AGB samples.

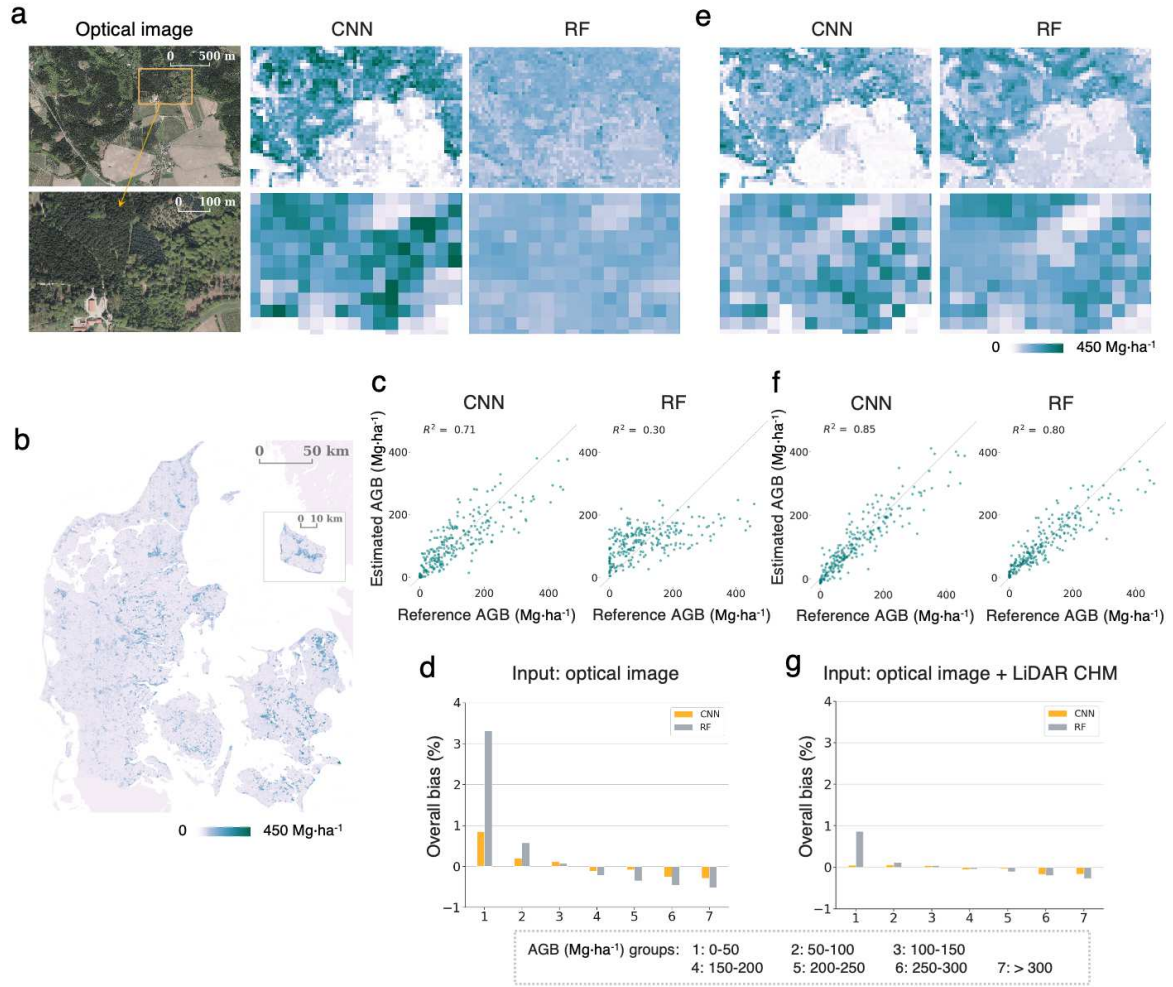


Figure 3: Visual examples and evaluation of AGB estimation using CNN and RF models. Sub-figures a-d demonstrate the results when using optical images as input; sub-figures e-g demonstrate the results when using optical images and LiDAR-derived canopy height maps (CHM) as inputs. a. AGB estimated using CNN and RF models for an example area composed of both coniferous and deciduous trees. The AGB was estimated at 40 m pixel grids, approximately matching the resolution of the NFI plots used for training. b. AGB estimation for the entire country of Denmark using optical images from 2018. c. Plot-level evaluation against the referential AGB using the test plots. d. Overall bias evaluated at 50 Mg-ha⁻¹ interval levels. Positive bias indicates overestimation and negative bias indicates underestimation. e. AGB estimated using CNN and RF models from optical images and LiDAR-derived CHM. f. Plot-level evaluation against the referential AGB. g. Overall bias evaluated within each 50 Mg-ha⁻¹ interval.

2.2 Performance Gain from LiDAR Height but Not from Estimated Height

We then tested the impacts of adding two different types of canopy height maps (CHM) as additional input sources reflecting the structural information of tree height. Firstly, we evaluated the CNN and RF models trained with both optical images and LiDAR-derived CHM (Figures 3.e-g and Table E4). As expected, we noticed a performance improvement for both models, particularly for RF with a rise in R^2 score from 0.3 to 0.8 and significantly lower plot-level estimation errors (Mann-Whitney U-test, $n = 258$, p -value < 0.001). Yet, the CNN models still outperformed RF, reaching an R^2 score of 0.85 ($rMAE = 24.15\%$, Figure 3.f). Visual examples illustrate that both models exhibited comparable distributions of AGB estimations with broad variations (Figure 3.e). The CNN model attained the minimum bias across a wide range of AGB values, falling consistently below 17% (Figure 3.g).

Secondly, estimated heights were utilized in place of the LiDAR heights to simulate the performance when potentially erroneous height approximates were used due to the unavailability of LiDAR data, which has become a recent trend for large-scale biomass estimation [22, 24, 44]. We used canopy heights estimated from the optical images using a deep regression model [25], which reported a tree-level MAE of 2.9 m in an independent test. Switching from LiDAR heights to estimated heights degraded the performance of all CNN and RF models (Table E1 and Table E5). Notably, the inclusion of estimated heights for CNNs resulted in slightly lower R^2 scores than using only optical images (Mann-Whitney U-test, $n = 8$, p -value < 0.001). To analyze the underlying reasons for the marginally worsened performance in CNNs, we calculated the Pearson correlation between the MAE of the estimated heights against the LiDAR heights and the absolute error of the estimated AGB against the referential AGB. The Pearson r was 0.47 (p -value < 0.01), indicating a positive correlation between high errors in the AGB estimation and high errors in the height estimation, and that AGB estimation could thereby be skewed by incorrect heights. Yet, integrating the estimated heights resulted in stronger RF models than

those trained with only optical features, as evidenced by an increase in R^2 score from 0.3 to 0.58 with significantly lower estimation errors (Mann–Whitney U-test, $n = 258$, $p\text{-value} < 0.001$).

2.3 Neural Network Enables Auto-learning of Tree-level Allometric Relationships

Finally, we analyzed how the plot-level CNN approach compares with tree-level approaches that utilize crown area and tree height to derive biomass for trees visible from an aerial perspective [30, 31, 45]. Tree-level estimators commonly involve a preliminary step of single tree identification from the input imagery. Here, we used individual tree crown segments predicted from the same aerial data in a previous study [25], which was also used as inputs for the abovementioned CNNs and RFs. According to an independent test, the tree crown segmentation model achieved an F1 score of 0.77 (recall = 0.69, precision = 0.96) [25]. Tree heights were attributed to tree crowns by using the 95th percentile values from two different CHMs, including the LiDAR-derived CHM and the estimated CHM [25]. We tested two approaches for tree-level AGB estimation (Figure 4.a). Firstly, we applied generalized allometric equations derived from literature [29] directly on the extracted tree-level features, including crown diameter and height. Generalized equations were utilized since field data and local allometric equations involving crown features were not available. Secondly, to overcome the drawbacks of generalized allometric equations, we introduced a neural network-based model that could learn to automatically establish a correlation between tree features and the total AGB by interpreting large amounts of data. The model estimated AGB at individual tree level with the ultimate goal of approaching the plot-level NFI references when aggregated (Figures 4.b-c). The bias from the missing understory was thereby calibrated internally. The problem can be formulated as shown in Eq. (1):

$$\min_{\theta_T} \frac{1}{N} \sum_{i=1}^N D \left(\sum_{t=1}^{C_i} f_{\theta_T}(x_1^t, x_2^t, \dots, x_d^t), y_i \right) \quad (1)$$

Specifically, given N training plots, we aim at minimizing the average distance (D) between the reference AGB (y) and the aggregated tree-level AGB estimated from an allometric model (f_{θ_T}) which is forest type specific (deciduous, coniferous, and non-forest). For plot i with C_i segmented tree crowns, the AGB of tree t is calculated by f_{θ_T} using tree-level features ($x_1^t, x_2^t, \dots, x_d^t$) where d denotes the number of features involved. The proposed model, namely SMNN, was implemented based on a smoothed version of the monotonic neural network (Figure 4.b) [42], which enforced monotonicity constraints between the input features and the target labels. This was intuitive as the AGB of trees should increase monotonically along both crown diameter and tree height, and enforcing monotonicity given prior knowledge has been shown to be effective in many applications [42, 46].

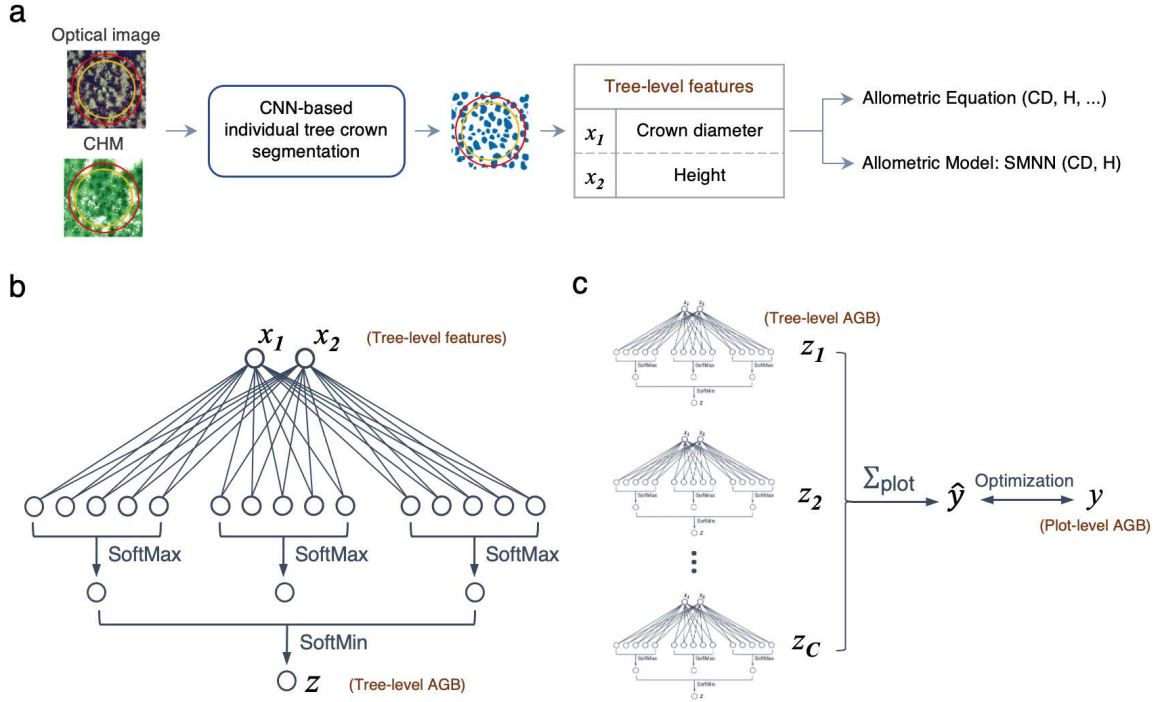


Figure 4: Demonstration of the tree-level approaches. a. Overall scheme for estimating tree-level AGB from aerial imagery. The tree-level features are extracted from optical images and canopy height maps using a CNN-based individual tree crown segmentation model [25]. The extracted features (crown diameter (CD) and height (H)) are converted into AGB using allometric equations or the proposed allometric model. b. Architecture of the learnable allometric model based on smooth monotonic neural network (SMNN) [42]. The network takes tree-level features as inputs and outputs tree-level AGB estimates. c. Illustration of the optimization. The AGB per tree (z) is estimated from SMNN. The individual estimates ($z_1, \dots, z_C$) are aggregated at the plot level ($\hat{y}$) and optimized against the reference (y).

SMNN models were trained with tree-level features, allowing them to reveal the hidden relationship between query variables automatically, without knowledge-based formulation of allometric relationships. We visualize the learned relationship between tree crown diameter (CD), height (H), and estimated AGB in a simple manner (Figure 5.a). When given both crown diameter and LiDAR height, the SMNN model established an increasing correlation towards AGB, with considerable variability as reflected by the broad scattering. High AGB could arise from high crown diameter, large height, or a combination of both. With the auto-learned allometric relationship, the estimated tree-level AGB yielded an aggregated plot-level R^2 score of 0.68 (Figure 5.c and Table E6). Removal of LiDAR heights in SMNN degraded both plot- and tree-level correlation. Particularly, when crown diameter was employed as the only attribute, a limited correlation was established towards AGB, resulting in horizontal gradient lines and a plot-level R^2 of 0.37 (Figure 5.c). The weak performance of SMNN with crown diameter implied that the monotonicity constraints might be violated, which was logically expected as tall trees with smaller crowns could hold higher AGB than short trees with larger crowns. In contrast to SMNN, the tree-level AGB derived from generalized allometric equations had a low plot-level R^2 score of -0.26 and an underestimation tendency, indicating a poor fit to the local scenario (Figure 5).

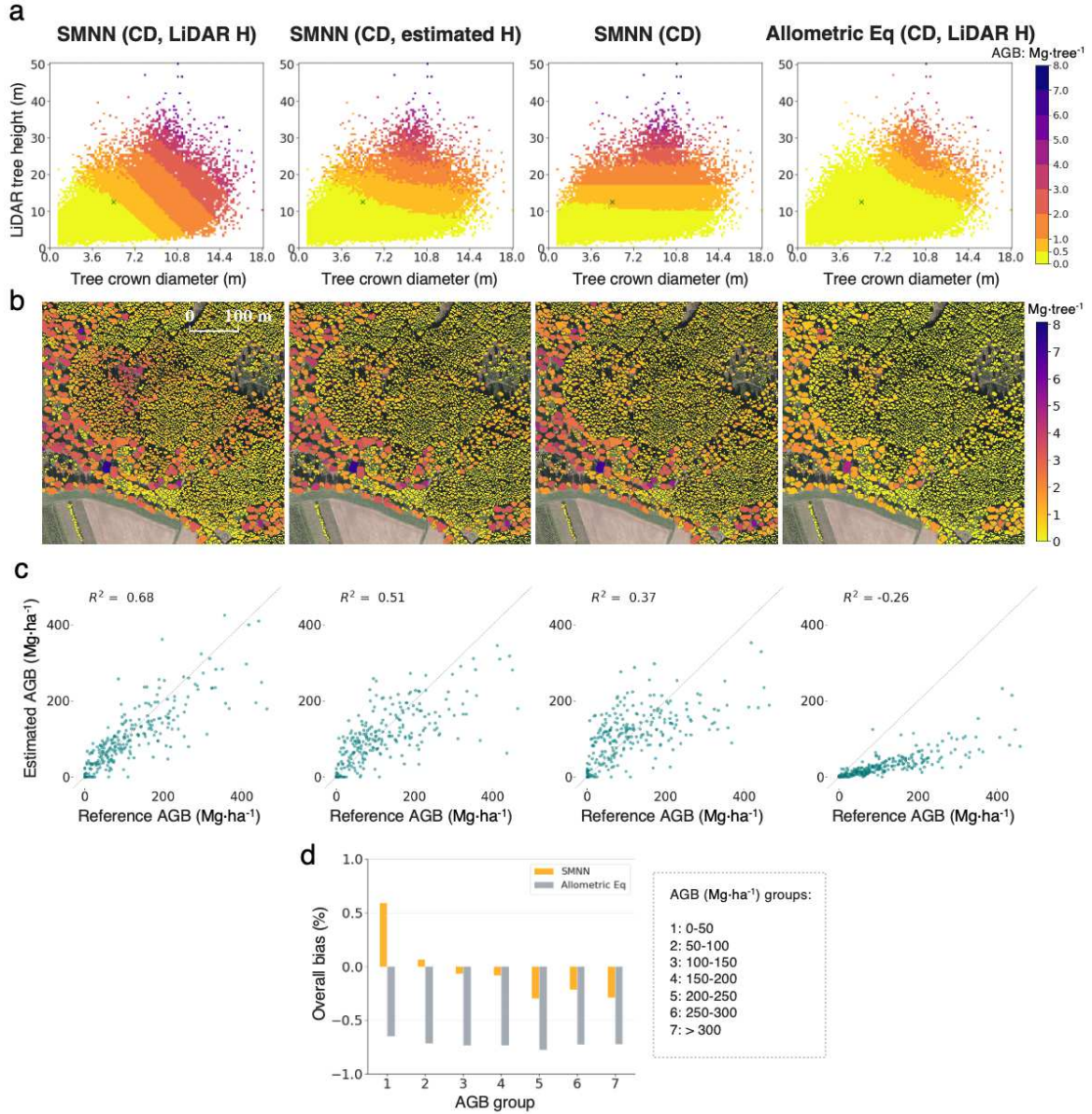


Figure 5: Tree-level AGB estimation and plot-level evaluation when aggregated. The comparison involves results from allometric equations (Eq) and SMNN models associated with different input sources including crown diameter (CD), LiDAR-derived height (LiDAR H), and estimated height (estimated H). a. 2D histograms displaying the relationship uncovered between tree crown diameter (x-axis), tree height from LiDAR (y-axis), and AGB estimation (color). Each pixel represents a specific detected tree with a specific crown diameter and LiDAR height, colored by AGB estimations derived from separate models. For instance, the green cross markers represent a specific tree featuring a 5.4 m crown diameter and 12.5 m height with an AGB estimation of 0.47, 0.48, 0.69, and 0.13 Mg derived from each model, respectively. A total of 105024 trees covering an area of 680 hectares are included in the histogram. b. Visual examples showing the tree-level AGB estimates. c. Plot-level evaluation against the reference AGB by aggregating tree-level estimations in each test plot. d. Overall bias of the SMNN model and allometric equations using CD and LiDAR H as input features, evaluated at 50 Mg-ha⁻¹ interval levels.

Overall, the plot-level CNNs reached higher R^2 scores over RF and SMNN models regardless of input sources, particularly when only optical RGB data was provided without the structural information of height (Figure 6). When LiDAR height was provided, the tree-level SMNN model achieved less than 10% bias on low-to-medium (50-200 Mg-ha⁻¹) AGB samples, with no statistically significant differences in estimation errors compared with the best-performing CNN model (Mann-Whitney U test, $n = 130$, $p > 0.05$). However, an increase in errors was observed for larger AGB values, implying a saturation in performance towards high AGB scenarios (Figure 5.d). In contrast, plot-level CNNs, which analyzed the original input images directly, did not show a notable decrease in performance for high biomass cases (Figure 3.g), implying that these models were capable of inferring biomass in complex dense forests without auxiliary information on individual trees.

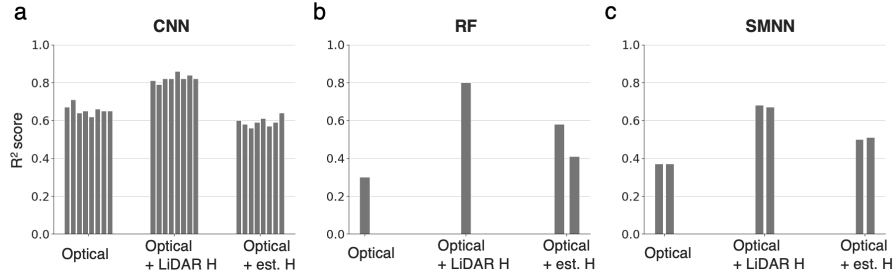


Figure 6: Performance comparison in relation to varying input sources, including optical images, optical images combined with LiDAR-derived heights (LiDAR H), and optical images combined with estimated heights (est. H). a. R^2 scores of CNN models calculated between estimated and referential AGB on the test set. Variation in R^2 is a result of different model architecture designs (Figure 1.a). b. R^2 scores of RF models. Two sources of estimated heights were tested: one estimated from the optical images [25] and another from an existing global vegetation elevation map [22]. c. R^2 scores of the tree-level SMNN models. Variation in R^2 scores is a result of different tree-level features involved in the estimation (Table E6).

3 Discussion

In recent years, extensive research has explored the application of deep learning in the field of remote sensing analysis for improved understanding of environmental resources, yet existing methods often fall short of fully utilizing the learning capacity of interpreting spatial patterns from images to derive information that goes beyond the mapping of visible features. Here, we proposed end-to-end trainable CNN models for forest AGB estimation directly from optical RGB images, without intermediate estimation of tree crown size or height. Given optical images, CNN models established a transition from coarse-resolution labels (i.e., plot-level AGB references) to a higher-resolution approximation, presenting a localized distribution of AGB that conformed to human visual cognition (Figure 2). To quantify forest or tree biomass, studies have traditionally utilized vegetation indices [47], spectral statistics [48, 49] or texture features [50, 51, 52], which were handcrafted from optical images. Our model proved its capacity of automatic spatial context interpretation and thus simplified feature extraction and multi-step workflow designs [45]. Automatic learning of spatial context turned out to be feasible only with convolutional kernel-based models, whereas a traditional RF model was ineffective, which was expected considering that CNNs are particularly strong in parsing spatial context information [53, 54]. We demonstrated the feasibility of such a method applied across various AGB gradients and found that optical images with a spatial resolution higher than approximately 1 m are a prerequisite. While our study only covered Danish landscapes, NFI measurements and high-resolution optical images are available in many European countries, supporting efficient AGB estimation at national scales (Figure 3.b). Satellite images with sub-meter resolution, broad coverage, and rapid acquisition such as Skysat [55] facilitate the implementation of timely and larger-scale analysis. Overall, the fast emergence of high-quality sub-meter resolution imagery and advanced deep learning networks are propelling future research in this domain, and while other studies concentrate on estimating height from images [22, 24, 44], we provide an alternative approach to biomass estimation.

Tree height is an essential parameter for AGB estimation [56, 57], yet the contribution of height turns out to be contingent upon its accuracy. Our analysis indicated that when considering CNNs with only optical images as the baseline, including LiDAR-derived heights increased the performance, whereas including estimated heights did not increase the performance (Figure 6.a), suggesting that estimated heights would not replace the reliable LiDAR heights for accurate AGB estimation. This could be explained by the fact that estimated heights were derived from the optical images. Consequently, the inclusion of estimated heights did not introduce new independent inputs but rather propagated its errors. This has been further proved by the positive correlation between height estimation errors and AGB estimation errors, and raises concerns about the use of estimated regional or global height models for AGB estimation [58, 59, 60, 61]. We argue that the error accumulations might be higher than expected through intermediate steps. However, integrating estimated heights dramatically improved the RF and tree-level models that were based only on optical data or crown diameters. This indicated the crucial role of heights for AGB estimation using non-texture interpretable models, consistent with common findings [62, 63]. Here, the estimated heights proved to be an effective input, strengthening the underperforming models even in the presence of noise.

Many studies routinely utilize high-resolution data to derive individual tree attributes, which are subsequently converted into AGB using allometric equations on crown variables and height [11, 29], or even only the crown variables [30, 31]. However, the limited number of adjustable coefficients of allometric equations may raise uncertainties and limitations when adapting to new datasets [34, 35]. This was demonstrated in this study by the weak performance using generalized allometric equations on crown diameter and height. In contrast, the proposed SMNN approach does not assume any fixed relationship between remotely detected features and the field-measured AGB, allowing for diverse combinations of these features through a learnable system. One possible explanation for the flexible combinations outperforming the one-fixed combination of allometric equations is that the relationship between remotely derived tree crowns, heights, and field-collected AGB may exhibit a more complex functional form than assumed. Through an end-to-end optimization from

individual tree attributes to the referential plot-level AGB labels, SMNN leverages the scalability of airborne attributes and the exactness of the field-measured AGB without explicitly establishing fixed allometric relationships. Finally, while we do not recommend an estimation of AGB from crown diameter only, we show that SMNN delivers acceptable performance for this scenario.

Overall, plot-level CNNs outperformed the multi-step tree-level approaches by maintaining relatively low bias across a broad range of AGB values. Tree-level AGB estimation necessitated a prior step of individual tree crown segmentation, which has become popular in recent years [26, 30, 64, 65]. Yet, these models were commonly found to be inefficient for dense or ambiguously structured canopies [25, 66, 67]. Crown segmentation models that were limited to top-layer canopies tended to be prone to performance saturation when applied to cases of high biomass ($> 200 \text{ Mg}\cdot\text{ha}^{-1}$), including particularly dense canopy crown cover of similar characteristics or complex multi-layer structures. Such high bias could lead to severe underestimations of the total biomass when aggregated for large areas. Concerns about the error propagation from the crown delineation and its complex relationship with AGB estimation errors have also been raised by similar studies [68]. Despite this, the low bias ($< 10\%$) of the tree-level SMNN for AGB ranging between $50\text{-}200 \text{ Mg}\cdot\text{ha}^{-1}$ suggested that tree-based approaches could be reliably applied for low-to-medium dense forests or trees outside forests. Moreover, tree-level approaches enable biomass tracking of small individual trees, which may be part of tree planting activities relevant to restoration and carbon crediting. Whereas plot-level models may ignore these trees, as they contribute little to the total biomass. Different environmental settings and precision of estimation are favored by plot-level and tree-level approaches and the optimal solutions depend on specific applications (Table 1).

Table 1: Comparison of the plot-level and tree-level approaches involved in this study.

Level	Approach	Advantages	Disadvantages
Plot	CNN	Operates without height Estimates localized AGB from plot-level labels Highest plot-level performance Readily applicable to similar landscapes Identifiable species-specific features are taken into account	Requires high spatial input resolution Requires large amounts of data for training Large model size Black box - limited transparency and explainability
Plot	RF	High performance with accurate heights Fast training and technically not challenging Conversion of tree height to AGB has been well studied	Requires manual feature extraction Uncertainties from species-specific parameters Black box - limited transparency and explainability
Tree	Allometric Eq	Easy fitting Broadly applicable, particularly for scattered trees or difficult terrain	Requires effective allometric equations, ideally landscape- and species-specific Accumulated bias from tree crown segmentation (undetectable or understory trees) Largely ignores species-specific parameters (wood density)
Tree	SMNN	Trainable with local plot-level field data No allometric relationship modeling Broadly applicable, particularly for scattered trees or difficult terrain Accounts for undetectable or understory trees via training Reasonable performance with crown diameter as only input Provides AGB for all segmented trees, ideal for restoration and plantation areas	Reduced aggregated plot-level performance Accumulated errors from tree crown segmentation Black box - limited transparency and explainability

Field-measured AGB from the Danish NFI served as the foundation for this study. Apart from the limitation of unavailable tree-level AGB measurements to conduct tree-level estimation, the allometry-based conversion of field-measured parameters to plot-level AGB and the temporal and spatial alignment of NFI plots and the input images could induce considerable labeling noise. The cost of collecting coordinated field data and airborne data across large areas could be high [69], thus limiting the evaluation of precise AGB estimation approaches to typically small scales [11, 68, 70]. Some countries conduct NFI and aerial scanning campaigns with narrow time intervals and rigorous designs [71, 72, 73, 74], making it a potentially huge database for relevant studies. Notably, NFI plots usually cover a rather small area with a radius varying from 4 m to 15 m [17, 71, 75], and the inclusion of trees standing close to plot edges tends to differ from what can be recognized from aerial images. For instance, the Danish NFI includes all trees with stems inside the plot borders [41], while stems are typically shaded by tree crowns when seen from above (Figure 2.a). This implies a potential bias between the airborne input data and the corresponding field references. The GPS positioning errors might further result in bias. However, the significance of the bias is not easy to verify or quantify, considering the complexities of situations and temporal discrepancies.

Finally, we foresee a couple of potential directions for further improving the current study. Firstly, enhancing the interpretability of the black-box CNN models by integrating prior knowledge or ordinal relationships [76] would help us to better understand or guide the training process. Secondly, integrating tree crown segmentation models with subsequent tree-level AGB estimation models to establish an end-to-end trainable system directly from input imagery to tree-level AGB estimates would mitigate the error propagation from intermediate steps. Thirdly, collecting detailed tree-level

data including remotely detectable features such as crown measurements during field inventories would lead to more interpretable and precise modeling or evaluation. Lastly, harmonizing the forest inventories with the aerial imagery campaigns would reduce the uncertainties related to the temporal and spatial alignment of the input and target output data for the model training.

4 Methods

4.1 Data and Pre-processing

The study area spans across Denmark, utilizing two primary sources of airborne data as input: aerial photographs featuring red, green, blue (RGB), and near-infrared (NIR) bands captured during the summer of 2018 at 20 cm resolution, and canopy height maps (CHM) derived from Light Detection and Ranging (LiDAR) scanning data collected during leaf-on seasons from 2014 to 2018 at 40 cm resolution. The average density of the LiDAR data is 4.5 points per m^2 , and the vertical and horizontal root mean squared errors are approximately 5 cm and 15 cm. We also used individual tree crown segmentation and height estimation products derived from the abovementioned airborne data [25], both at a spatial resolution of 20 cm. Specifically, individual tree crowns were predicted from the airborne data by training with manual tree crown delineations, and canopy height maps were estimated from the optical aerial images by training with LiDAR-derived CHM [25]. Lastly, a global vegetation height estimation map (ETH height) [22] for the year 2020 at a 10 m resolution was utilized for comparison.

The ground truth reference data for this study were obtained from the Danish National Forest Inventory (NFI) [41]. The NFI database comprised field plots with a radius of 15 m, where individual representative trees were measured. The above-ground biomass (AGB) of the field-measured trees was calculated using local AGB functions associated with tree species, stem diameter, and height [41]. We focused on the total AGB at the plot level, which was calculated by aggregating individual measurements with a scaling factor that accounted for the distribution of tree sizes within the plot, including all trees with stem diameter > 0 mm [41]. To mitigate the temporal inconsistency between the ground truth references and the input imagery, only plots measured between 2015 and 2018 were used.

The high-resolution input imagery, including the optical RGB images, LiDAR-derived CHM, and estimated CHM, were cropped into image patches matching the locations of the NFI field plots, with an outward buffer of 3 m to include the neighborhood context. We tested different buffer sizes and found 3 m to be appropriate for including larger crowns on the edges without introducing too much neighboring noise. The coarser-resolution ETH height maps were cropped to cover the plot areas. Regarding the crown segmentation masks, we kept tree crowns with centroids falling inside the plot circles to align approximately with the NFI scheme where all trees with stems standing inside the plots were included in the AGB calculation.

The dataset was randomly split into training, validation, and test sets, with a ratio of 8:1:1. The training set was used to train or fit the models, the validation set was used for model selection or hyper-parameter tuning, and the test set was reserved exclusively for the final evaluation presented in Results. Through visual assessment, we noticed occasional misalignment between the input imagery and the target AGB labels primarily due to three reasons: clearcut or growing of trees due to discrepancies in dates between field campaigns (2015-2018) and image acquisition (summer 2018), variation in the inclusion of trees at plot edges between field inventory schemes and image inspections, and uncertainties associated with GPS coordinates, particularly given the small plot sizes (15 m radius). Firstly, we removed samples of obvious mismatch from all sets. Then, we assigned sample weights ($w \in \{0.3, 0.5, 0.7, 1\}$) to the training and validation samples based on visual pairwise assessments of the matching degree between the optical images and AGB references (higher w indicates a better match). The weights assigned to the validation samples were neutral, while the weights for the training samples were less stringent (Figure E1). A higher tolerance was granted to the training samples due to the model's expected robustness to a certain level of noise, with the goal of maximizing the number of effective samples collected. All testing samples were assessed equally without sample weighting. Notably, the sample weights only concern the training and do not concern the application for areas without AGB references. The resulting dataset consisted of 1897 training samples, 268 validation samples, and 258 test samples (Table E7).

4.2 Plot-level Convolutional Neural Network Regression

We tested three common CNN backbones as feature extractors, namely Efficientnet-B0, Efficientnet-B1 [77], and RegnetY800MF [78]. Efficientnet-B0 combined the inverted residual blocks [79] with squeeze-and-excitation blocks. Efficientnet-B1 expanded Efficientnet-B0 by uniformly scaling up the width, depth, and resolution of the network. RegnetY800MF was proposed by Radosavovic *et al.* for flexible and efficient model design, which involved a linear parameterization of block widths and depths [78]. The backbones were initialized with weights pre-trained on Imagenet. We designed two types of regression heads on top of the feature extractors to produce the desired output. Regression head A consisted of a global average pooling across each feature map, followed by three dense layers combined with indeterminate dropout layers with a dropout rate of 0.2. The dense layers gradually reduced the feature dimensionality to one, which corresponded to the plot-level total AGB value we targeted for. Regression head B extended the feature extractor with a sequence of (1×1) pointwise convolutional layers to ultimately generate a one-channel feature activation map (Figure 1.a). Here, the sum over all pixels of this last activation map corresponded to the total AGB value we targeted for. With four dimension reduction branches of the feature extractor, the spatial resolution of the input imagery sequentially reduced from (224×224) pixels to (14×14) pixels, resulting in a feature activation map at roughly 2 m ground resolution. A circular mask

was applied to align the activation map with the referential circular plots (15 m radius), and the summation only involved the masked values, avoiding potential artifacts caused by applying convolutional kernels on the circular edges. Notably, the circular masks were only considered in the training and evaluation phases to compare with the NFI references. For applications beyond plot scales, no masking was applied.

During training, the input images were augmented by random rotation, flipping, and color jittering including alteration of brightness, contrast, saturation, and hue. The pixel intensity values were rescaled to the range of (0, 1) and normalized based on the mean and standard deviation of the ImageNet dataset. The models were trained in a fully supervised manner with plot-level NFI reference labels. We used the mean absolute error (MAE) as loss function, which was weighted w.r.t. the sample weights. We used the mini-batch training strategy [80] with a batch size of 32. The mini-batch-level loss for gradient descent can be formulated as in Eq. (2), where y_i denotes the reference AGB, $\hat{y}_i$ the estimated AGB, w_i the sample weight, and n the batch size:

$$wMAE = \frac{\sum_{i=1}^n w_i \cdot |\hat{y}_i - y_i|}{\sum_{i=1}^n w_i} \quad (2)$$

We re-weighted the loss function calculated for the entire batch using the total sample weights for that batch instead of the total number of samples to avoid reaching lower losses simply by drawing random samples with lower sample weights by chance. The sample weights could be seen as prior knowledge or a warm-up model for the training. The final model was determined as the one achieving the highest R^2 score on the validation set. After training, we further adjusted the bias parameter in the final layer while fixing all other parameters to mitigate the systematic bias potentially caused by the iterative training [81]. Given a CNN model of the form $f(x) = a^T h_w(x) + b$ where a and b (unadjusted bias) denote the parameters in the final layer and h_w denotes all other parameters, the adjusted bias b^* was calculated using the combined training and validation data by Eq. (3), where N denotes the total number of samples and y denotes the reference AGB [81]:

$$b^* = \frac{\sum_i^N (y_i - a^T h_w(x_i))}{N} = \frac{\sum_i^N y_i - \sum_i^N f(x_i)}{N} + b \quad (3)$$

4.3 Plot-level Pixel Intensity and Texture-based Random Forest Regression

Features extracted from the optical RGB images included statistics of each color band, common vegetation indices [82], histogram intensities, and statistical texture properties calculated from the gray level co-occurrence matrix [83] (Table E3). Features extracted from the LiDAR-derived CHM, and estimated CHM included statistics and histogram intensities of the maps (Table E3). Random Forest (RF) regressors were fitted using these hand-crafted features, with total feature counts ranging between 99-124 depending on whether heights were included. We conducted a grid search to determine the optimal hyper-parameters including the number of trees, the maximum tree depth, and the number of features for node splitting using the combined training and validation data. The fitting samples were weighted by the sample weights. The final model was selected as the one reaching the lowest out-of-bag error [84]. After training, we further adjusted the bias parameter using Eq. (3) by setting $a = 1$ and the initial $b = 0$ for RF regression [81].

4.4 Tree-level Allometric Equation-based Estimation

Here, the tree-level AGB was estimated in two steps. Firstly, individual tree crowns were segmented from the airborne data, and tree-level features were extracted based on the crown segments. Secondly, the AGB for each segmented tree was derived using the generalized allometric equations [29].

4.4.1 Individual Tree Crown Segmentation

Recently, a relevant study has used an attention-UNet-based model for predicting individual tree crowns across a wide range of environmental landscapes in Denmark [25]. Here, we utilized the predicted individual tree crowns as inputs to derive AGB at the tree level. The crown diameter, tree height, and forest type were attributed to each tree. Specifically, the crown diameter was calculated by assuming a circular shape of the crown. The tree height was derived by overlaying CHM with the segmentation masks and taking the 95th percentile height within the crown area. The forest type was obtained by overlaying the 10 m resolution Copernicus forest-type map (2018) [85] with the segmentation masks.

4.4.2 AGB Estimation from Generalized Allometric Equations

We derived generalized allometric equations following the formula and field data from Jucker *et al.* [29] (Eq. (4), Eq. (5), and Figure E2). These equations were fitted using public in situ measurements in global temperate forest zones, and the

measuring of crown diameters allowed for a direct correlation between remotely detected tree crowns and field-collected AGB [29]. Allometric equations involving crown diameters could not be established using the Danish NFI data due to the unavailability of crown measurements. The fitting incorporated log-transformed crown diameter and height as regression variables and the bias of logarithmic regression was corrected [86]. Two allometric equations were fitted between crown diameter (CD, in meters), height (H, in meters), and AGB (in kilograms), one using 223 tree measurements in temperate coniferous forest for the segmented coniferous trees, and the other using 440 tree measurements in temperate mixed forest for the deciduous and non-forest tree segmentations:

$$AGB_{Deciduous} = e^{-2.811} \times (CD \times H)^{1.807} \times e^{(0.0215^2/2)} \quad (4)$$

$$AGB_{Coniferous} = e^{-0.951} \times (CD \times H)^{1.367} \times e^{(0.0297^2/2)} \quad (5)$$

4.5 Tree-level Smooth Monotonic Neural Network Regression

Here, AGB was estimated in two steps. Firstly, individual tree crowns were predicted from the airborne data (section 4.4.1). Secondly, tree-level features extracted from the crown segments were used to estimate tree-level AGB using an equation-free data-driven allometric model named SMNN.

4.5.1 AGB Estimation from Tree-level Features via SMNN

The smooth monotonic neural network had a three-layer structure (Figure 4.b). The first layer contained several groups of neurons, each computing a linear transformation of the input features (tree crown diameter and height). The second layer extracted the maximum units within each group. And the third layer extracted the global minimum units across all groups. Increasing monotonicity on all features was enforced by constraining all weights in the linear transformation layer to be positive [87]. We applied additional smoothing/softening by replacing the nonsmooth maximum operations with the scaled Log-Sum-Exp (LSE, Eq. (6) [42]), a smooth approximation to the maximum function:

$$LSE = \frac{1}{\beta} \log \sum_{n=1}^N \exp(\beta x_n) \quad (6)$$

Likewise, a smooth minimum operation was implemented by the negation of LSE. An additional non-negative constraint was applied to the tree-level AGB estimations to avoid estimating negative AGB values. The model architecture was rather flexible and involved two tunable hyper-parameters: the number of groups in the linear layer (K) and the number of neurons per group (K_n). Empirically, we set $K = K_n$ and conducted grid-search to determine the optimal choices of K using the training and validation data. We also tested different features including crown diameter (CD), height (H), and their log-transformed counterparts ($\log(CD)$ and $\log(H)$).

We trained three separate models for deciduous trees, coniferous trees, and non-forest trees for two primary reasons. First, allometric relationships are commonly known as species or forest-type specific [32, 88]. Second, monotonicity does not hold for the forest type as a feature with respect to AGB. During training, the three models were trained independently using the corresponding subsets of data determined by the forest type. The resilient backpropagation (Rprop) optimizer was used with a learning rate of 0.1 [89]. The loss function was MAE weighted by the assigned sample weights:

$$wMAE_2 = \frac{\sum_{i=1}^N w_i \cdot |\hat{y}_i - y_i|}{N} \quad (7)$$

Here, y_i and $\hat{y}_i$ denote the reference and the estimated AGB, and N denotes the total number of samples. We used only reliable samples with weight ≥ 0.5 in the training phase. The optimal weights for the three-model ensemble were determined as the ones that achieved the highest R^2 score on the whole validation set. All models were trained for 1000 iterations with batch gradient descent [80].

4.6 Performance Evaluation

We evaluated the model performance on the test data using the following metrics. The coefficient of determination (R^2) score measured how well the reference data could be replicated by the model, also known as a goodness-of-fit metric. Here, y_i , $\hat{y}_i$, and $\bar{y}$ denote the reference value, the estimated value, and the mean reference value, respectively:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (8)$$

The MAE between the reference and the estimated values was calculated as Eq. (9):

$$MAE = \frac{1}{N} \sum_i^N |\hat{y}_i - y_i| \quad (9)$$

The MAE relative to the mean reference value (rMAE) was calculated as Eq. (10):

$$rMAE(\%) = \frac{1}{N * \bar{y}} \sum_i^N |\hat{y}_i - y_i| \times 100 \quad (10)$$

The above metrics quantified errors at the plot level, which could be highly affected by the location noise inherent in the alignment of field data with airborne data. Therefore, we grouped the plots according to the reference AGB values with a 50 Mg·ha⁻¹ interval and calculated the overall bias within each group, which reflected the aggregated performance beyond plot scales:

$$bias(\%) = \frac{\sum_i (\hat{y}_i - y_i)}{\sum_i y_i} \times 100 \quad (11)$$

5 Extended Data - Figures

418

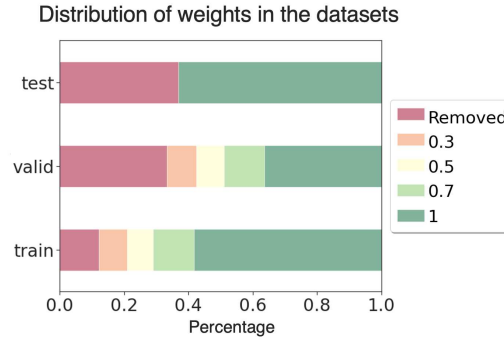


Figure E1: Distribution of removed samples and sample weights assigned to the training and validation sets. The varying distribution of weights in the training data and that in the validation and testing data is a result of higher tolerance given to doubtful samples in the training set than those in the validation or testing sets. See details in Methods.

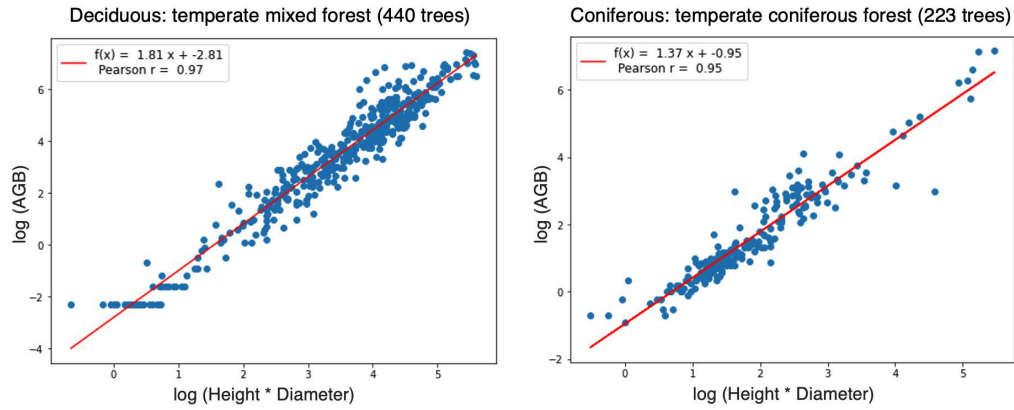


Figure E2: Generalized allometric equations fitted between height (meters), crown diameter (meters), and AGB (kilograms) using data from the literature [29].

6 Extended Data - Tables

419

Table E1: Performance of models trained using optical images as inputs, evaluated plot-by-plot using the test dataset. The CNN models featured two types of regression head designs, where -A represented the dense layer design and -B represented the activation map design (Figure 1.a). The best-performing model of each approach is highlighted in bold text. MAEs are shown in $\text{Mg} \cdot \text{ha}^{-1}$.

Approach	Model design	Input	MAE (rMAE)	R^2
CNN	Efficientnetb0-A	Optical images	41.04 (36.45%)	0.66
	Efficientnetb0-B	Optical images	39.08 (34.71%)	0.71
	Efficientnetb1-A	Optical images	40.30 (35.79%)	0.66
	Efficientnetb1-B	Optical images	43.67 (38.79%)	0.62
	Efficientnetb2-A	Optical images	42.41 (37.67%)	0.64
	Efficientnetb2-B	Optical images	41.53 (36.89%)	0.66
	RegnetY800MF-A	Optical images	41.71 (37.05%)	0.64
	RegnetY800MF-B	Optical images	41.78 (37.11%)	0.65
RF		Optical images	65.40 (58.09%)	0.30

Table E2: Performance of models trained using optical images of different resolutions, evaluated plot-by-plot using the test dataset. MAEs are shown in $\text{Mg} \cdot \text{ha}^{-1}$.

Model design	Input	Input resolution	MAE (rMAE)	R^2
Efficientnetb0-B	Optical images	40 cm	42.81 (38.02%)	0.64
		80 cm	46.71 (41.49%)	0.59
		160 cm	52.16 (46.33%)	0.47
		320 cm	55.55 (49.34%)	0.38

Table E3: Hand-crafted features for Random Forest regression. R, G, B, and NIR stands for red, green, blue, and near-infrared, respectively.

Type	Features	Formula
Band statistics	Mean, Median, Variance, Percentile (5th and 95th)	
Histogram intensities		
Vegetation index	Normalized Difference Vegetation Index (NDVI) [90]	$(\text{NIR} - \text{R}) / (\text{NIR} + \text{R})$
	Excess Green Index (ExGI) [91]	$2 * \text{G} - (\text{R} + \text{B})$
	Green Chromatic Coordinate (GCC) [92]	$\text{G} / (\text{R} + \text{G} + \text{B})$
	Ratio Vegetation Index (RVI) [93]	R / NIR
	Green Leaf Index (GLI) [94]	$(\text{G} - \text{R} + \text{G} - \text{B}) / (2 * \text{G} + \text{B} + \text{R})$
	Green Normalized Difference Vegetation Index (GNDVI) [95]	$(\text{NIR} - \text{G}) / (\text{NIR} + \text{G})$
Texture [83]	Contrast	
	Dissimilarity	
	Homogeneity	
	Energy	
	Correlation	
	Angular second moment	

Table E4: Performance of models trained using optical images and LiDAR-derived CHMs as inputs, evaluated plot-by-plot using the test dataset. The CNN models featured two types of regression head designs, where -A represented the dense layer design and -B represented the activation map design (Figure 1.a). The best-performing model of each approach is highlighted in bold text. MAEs are shown in $\text{Mg} \cdot \text{ha}^{-1}$.

Approach	Model design	Input	MAE (rMAE)	R^2
CNN	Efficientnetb0-A	Optical images + CHM	27.19 (24.15%)	0.85
	Efficientnetb0-B	Optical images + CHM	30.76 (27.32%)	0.80
	Efficientnetb1-A	Optical images + CHM	30.23 (26.85%)	0.81
	Efficientnetb1-B	Optical images + CHM	28.63 (25.43%)	0.82
	Efficientnetb2-A	Optical images + CHM	30.85 (27.40%)	0.82
	Efficientnetb2-B	Optical images + CHM	29.02 (25.77%)	0.82
	RegnetY800MF-A	Optical images + CHM	28.08 (24.94%)	0.83
	RegnetY800MF-B	Optical images + CHM	30.17 (26.80%)	0.82
RF		Optical images + CHM	30.19 (26.81%)	0.80

Table E5: Performance of models trained using optical images and estimated CHMs as inputs, evaluated plot-by-plot using the test dataset. The CNN models featured two types of regression head designs, where -A represented the dense layer design and -B represented the activation map design (Figure 1.a). The best-performing model of each approach is highlighted in bold text. MAEs are shown in $\text{Mg} \cdot \text{ha}^{-1}$.

Approach	Model design	Input	MAE (rMAE)	R^2
CNN	Efficientnetb0-A	Optical images + estimated CHM	44.46 (39.49%)	0.57
	Efficientnetb0-B	Optical images + estimated CHM	43.63 (38.75%)	0.58
	Efficientnetb1-A	Optical images + estimated CHM	43.82 (38.92%)	0.59
	Efficientnetb1-B	Optical images + estimated CHM	43.81 (38.91%)	0.59
	Efficientnetb2-A	Optical images + estimated CHM	43.84 (38.94%)	0.61
	Efficientnetb2-B	Optical images + estimated CHM	46.68 (41.46%)	0.57
	RegnetY800MF-A	Optical images + estimated CHM	44.75 (39.75%)	0.59
	RegnetY800MF-B	Optical images + estimated CHM	42.17 (37.45%)	0.64
RF		Optical images + estimated CHM	45.01 (39.98%)	0.58
		Optical images + ETH height (10 m) [22]	58.78 (52.21%)	0.41

Table E6: Performance of SMNN models trained using different tree-level features, including crown diameter (CD), height (H), and their log-transformed counterparts. MAEs are shown in $\text{Mg} \cdot \text{ha}^{-1}$.

Features	MAE (RMAE)	R^2
CD, H, log(CD), log(H)	37.78 (33.56%)	0.68
CD, H	38.41 (34.11%)	0.67
CD, estimated H, log(CD), log(estimated H)	48.77 (43.32%)	0.50
CD, estimated H	48.34 (42.93%)	0.51
CD, log(CD)	58.35 (51.83%)	0.37
CD	58.24 (51.73%)	0.37

Table E7: Summary of the AGB references ($\text{Mg} \cdot \text{ha}^{-1}$) in each dataset.

Dataset	No. plots	AGB mean	AGB 99th percentile	Forest type [41] fraction		
				Deciduous	Coniferous	Mixed
Training	1897	118.42	440.77	47.9%	42.1%	9.9%
Validation	268	110.28	415.05	36.9%	48.6%	14.4%
Testing	258	112.59	433.45	39.8%	50.6%	9.5%

References

- [1] Yude Pan, Richard A Birdsey, Jingyun Fang, Richard Houghton, Pekka E Kauppi, Werner A Kurz, Oliver L Phillips, Anatoly Shvidenko, Simon L Lewis, Josep G Canadell, et al. A large and persistent carbon sink in the world's forests. *Science*, 333(6045):988–993, 2011.
- [2] Rupert Seidl, Mart-Jan Schelhaas, Werner Rammer, and Pieter Johannes Verkerk. Increasing forest disturbances in Europe and their impact on carbon storage. *Nature Climate Change*, 4(9):806–810, 2014.
- [3] Josep G Canadell and Michael R Raupach. Managing forests for climate change mitigation. *Science*, 320(5882):1456–1457, 2008.
- [4] Valentin Bellassen and Sebastiaan Luyssaert. Carbon sequestration: Managing forests in uncertain times. *Nature*, 506(7487):153–155, 2014.
- [5] Jeffrey D Sachs, Guido Schmidt-Traub, Mariana Mazzucato, Dirk Messner, Nebojsa Nakicenovic, and Johan Rockström. Six transformations to achieve the sustainable development goals. *Nature Sustainability*, 2(9):805–814, 2019.
- [6] Martin Herold and Margaret Skutsch. Monitoring, reporting and verification for national redd+ programmes: two proposals. *Environmental Research Letters*, 6(1):014002, 2011.
- [7] Aarti Gupta, Eva Löybrand, Esther Turnhout, and Marjanneke J Vijge. In pursuit of carbon accountability: the politics of REDD+ measuring, reporting and verification systems. *Current Opinion in Environmental Sustainability*, 4(6):726–731, 2012.
- [8] Lars Gamfeldt, Tord Snäll, Robert Bagchi, Micael Jonsson, Lena Gustafsson, Petter Kjellander, María C Ruiz-Jaen, Mats Fröberg, Johan Stendahl, Christopher D Philipson, et al. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, 4(1):1340, 2013.
- [9] Gert-Jan Nabuurs, Marcus Lindner, Pieter J Verkerk, Katja Gunia, Paola Deda, Roman Michalak, and Giacomo Grassi. First signs of carbon sink saturation in European forest biomass. *Nature Climate Change*, 3(9):792–796, 2013.
- [10] Jerome Chave, Stuart J Davies, Oliver L Phillips, Simon L Lewis, Plinio Sist, Dmitry Schepaschenko, John Armston, Tim R Baker, David Coomes, Mathias Disney, et al. Ground data are essential for biomass remote sensing missions. *Surveys in Geophysics*, 40:863–880, 2019.
- [11] Alejandro Navarro, Mary Young, Blake Allan, Paul Carnell, Peter Macreadie, and Daniel Ierodiaconou. The application of Unmanned Aerial Vehicles (UAVs) to estimate above-ground biomass of mangrove ecosystems. *Remote Sensing of Environment*, 242:111747, 2020.
- [12] Ronald E McRoberts and Erkki O Tomppo. Remote sensing support for national forest inventories. *Remote Sensing of Environment*, 110(4):412–419, 2007.
- [13] Nathan L Stephenson, AJ Das, R Condit, SE Russo, PJ Baker, Noelle G Beckman, DA Coomes, ER Lines, WK Morris, Nadja Rüger, et al. Rate of tree carbon accumulation increases continuously with tree size. *Nature*, 507(7490):90–93, 2014.
- [14] Yude Pan, Richard A Birdsey, Oliver L Phillips, and Robert B Jackson. The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics*, 44:593–622, 2013.
- [15] S.G. Zolkos, S.J. Goetz, and R. Dubayah. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 128:289–298, 2013.
- [16] Nancy L Harris, David A Gibbs, Alessandro Baccini, Richard A Birdsey, Sytze De Bruin, Mary Farina, Lola Fatoyinbo, Matthew C Hansen, Martin Herold, Richard A Houghton, et al. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change*, 11(3):234–240, 2021.
- [17] Xinlian Liang, Ville Kankare, Juha Hyyppä, Yunsheng Wang, Antero Kukko, Henrik Haggrén, Xiaowei Yu, Harri Kaartinen, Anttoni Jaakkola, Fengying Guan, et al. Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:63–77, 2016.
- [18] Jonathan P. Dandois and Erle C. Ellis. High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment*, 136:259–276, 2013.
- [19] Kim Calders, Glenn Newnham, Andrew Burt, Simon Murphy, Pasi Raunonen, Martin Herold, Darius Culvenor, Valerio Avitabile, Mathias Disney, John Armston, et al. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution*, 6(2):198–208, 2015.
- [20] Stefan Oehmcke, Lei Li, Jaime C Revenga, Thomas Nord-Larsen, Katerina Trepekli, Fabian Gieseke, and Christian Igel. Deep learning based 3D point cloud regression for estimating forest biomass. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, pages 1–4, 2022.
- [21] Feng Zhao, Qinghua Guo, and Maggi Kelly. Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA. *Agricultural and Forest Meteorology*, 165:64–72, 2012.
- [22] Nico Lang, Walter Jetz, Konrad Schindler, and Jan Dirk Wegner. A high-resolution canopy height model of the Earth. *arXiv preprint arXiv:2204.08322*, 2022.
- [23] Siyu Liu, Martin Brandt, Thomas Nord-Larsen, Jerome Chave, Florian Reiner, Nico Lang, Xiaoye Tong, Philippe Ciais, Christian Igel, Sizhuo Li, et al. The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. 2023.
- [24] Jamie Tolan, Hung-I Yang, Ben Nosarzewski, Guillaume Couairon, Huy Vo, John Brandt, Justine Spore, Sayantan Majumdar, Daniel Haziza, Janaki Vamaraju, et al. Sub-meter resolution canopy height maps using self-supervised learning and a vision transformer trained on Aerial and GEDI Lidar. *arXiv preprint arXiv:2304.07213*, 2023.

- [25] Sizhuo Li, Martin Brandt, Rasmus Fensholt, Ankit Kariryaa, Christian Igel, Fabian Gieseke, Thomas Nord-Larsen, Stefan Oehmcke, Ask Holm Carlsen, Samuli Junttila, et al. Deep learning enables image-based tree counting, crown segmentation, and height prediction at national scale. *PNAS nexus*, 2(4), 2023. 472
473
474
- [26] Martin Brandt, Compton J Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, et al. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature*, 587(7832):78–82, 2020. 475
476
- [27] Florian Reiner, Martin Brandt, Xiaoye Tong, David Skole, Ankit Kariryaa, Philippe Ciais, Andrew Davies, Pierre Hiernaux, Jerome Chave, Maurice Mugabowindekwe, et al. More than one quarter of Africa’s tree cover is found outside areas previously classified as forest. *Nature Communications*, 14(1):2258, 2023. 477
478
479
- [28] Riccardo Dainelli, Piero Toscano, Salvatore Filippo Di Gennaro, and Alessandro Matese. Recent advances in unmanned aerial vehicle forest remote sensing—A systematic review. part I: A general framework. *Forests*, 12(3):327, 2021. 480
481
- [29] Tommaso Jucker, John Caspersen, Jerome Chave, Cecile Antin, Nicolas Barbier, Frans Bongers, Michele Dalponte, Karin Y van Ewijk, David I Forrester, Matthias Haeni, et al. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Global Change Biology*, 23(1):177–190, 2017. 482
483
484
- [30] Maurice Mugabowindekwe, Martin Brandt, Jerome Chave, Florian Reiner, David L Skole, Ankit Kariryaa, Christian Igel, Pierre Hiernaux, Philippe Ciais, Ole Mertz, et al. Nation-wide mapping of tree-level aboveground carbon stocks in Rwanda. *Nature Climate Change*, 13(1):91–97, 2023. 485
486
487
- [31] Compton Tucker, Martin Brandt, Pierre Hiernaux, Ankit Kariryaa, Kjeld Rasmussen, Jennifer Small, Christian Igel, Florian Reiner, Katherine Melocik, Jesse Meyer, et al. Sub-continental-scale carbon stocks of individual trees in African drylands. *Nature*, 615(7950):80–86, 2023. 488
489
- [32] TM Basuki, PE Van Laake, AK Skidmore, and YA Hussin. Allometric equations for estimating the above-ground biomass in tropical lowland dipterocarp forests. *Forest Ecology and Management*, 257(8):1684–1694, 2009. 490
491
- [33] Gudeta W Sileshi. A critical review of forest biomass estimation models, common mistakes and corrective measures. *Forest Ecology and Management*, 329:237–254, 2014. 492
493
- [34] Michiel Van Breugel, Johannes Ransijn, Dylan Craven, Frans Bongers, and Jefferson S Hall. Estimating carbon stock in secondary forests: decisions and uncertainties associated with allometric biomass models. *Forest Ecology and Management*, 262(8):1648–1657, 2011. 494
495
- [35] Razi Ahmed, Paul Siqueira, Scott Hensley, and Kathleen Bergen. Uncertainty of forest biomass estimates in north temperate forests due to allometry: Implications for remote sensing. *Remote Sensing*, 5(6):3007–3036, 2013. 496
497
- [36] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. IEEE, 2017. 498
499
- [37] Teja Kattenborn, Jens Leitloff, Felix Schiefer, and Stefan Hinz. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:24–49, 2021. 500
501
- [38] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018. 502
503
- [39] Michael Halstead, Christopher McCool, Simon Denman, Tristan Perez, and Clinton Fookes. Fruit quantity and ripeness estimation using a robotic vision system. *IEEE Robotics and Automation Letters*, 3(4):2995–3002, 2018. 504
505
- [40] Stephen Law, Brooks Paige, and Chris Russell. Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5):1–19, 2019. 506
507
- [41] Thomas Nord-Larsen and Vivian Kvist Johannsen. Danish national forest inventory. *Design and calculations.(Management DoGaNR, trans.) IGN report*, page 33, 2016. 508
509
- [42] Christian Igel. Smooth monotonic networks. *arXiv preprint arXiv:2306.01147*, 2023. 510
- [43] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307, 2020. 511
512
- [44] Fabien H Wagner, Sophia Roberts, Alison L Ritz, Griffin Carter, Ricardo Dalagnol, Samuel Favrichon, Mayumi Hirye, Martin Brandt, Philippe Ciais, and Sassan Saatchi. Sub-meter tree height mapping of California using aerial images and LiDAR-informed U-Net model. *arXiv preprint arXiv:2306.01936*, 2023. 513
514
515
- [45] Dengsheng Lu, Qi Chen, Guangxing Wang, Lijuan Liu, Guiying Li, and Emilio Moran. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, 9(1):63–105, 2016. 516
517
- [46] Hoang Tuy. Monotonic optimization: Problems and solution approaches. *SIAM Journal on Optimization*, 11(2):464–494, 2000. 518
- [47] Fabrício L Macedo, Adélia MO Sousa, Ana Cristina Gonçalves, Jose R Marques da Silva, Paulo A Mesquita, and Ricardo AF Rodrigues. Above-ground biomass estimation for quercus rotundifolia using vegetation indices derived from high spatial resolution satellite images. *European Journal of Remote Sensing*, 51(1):932–944, 2018. 519
520
521
- [48] Daud Jones Kachamba, Hans Ole Ørka, Terje Gobakken, Tron Eid, and Weston Mwase. Biomass estimation using 3D data from unmanned aerial vehicle imagery in a tropical woodland. *Remote Sensing*, 8(11):968, 2016. 522
523
- [49] Yuying Liang, Weili Kou, Hongyan Lai, Juan Wang, Qiuhua Wang, Weiheng Xu, Huan Wang, and Ning Lu. Improved estimation of aboveground biomass in rubber plantations by fusing spectral and textural information from UAV-based RGB imagery. *Ecological Indicators*, 142:109286, 2022. 524
525
526

- [50] Christophe Proisy, Pierre Couteron, and François Fromard. Predicting and mapping mangrove biomass from canopy grain analysis using fourier-based textural ordination of ikonos images. *Remote Sensing of Environment*, 109(3):379–392, 2007. 527 528
- [51] Pierre Ploton, Raphaël Pélissier, Christophe Proisy, Théo Flavenot, Nicolas Barbier, SN Rai, and Pierre Couteron. Assessing aboveground tropical forest biomass using Google Earth canopy images. *Ecological Applications*, 22(3):993–1003, 2012. 529 530
- [52] Jean-François Bastin, Nicolas Barbier, Pierre Couteron, Benoît Adams, Aurélie Shapiro, Jan Bogaert, and Charles De Cannière. Aboveground biomass mapping of African forest mosaics using canopy texture analysis: toward a regional approach. *Ecological Applications*, 24(8):1984–2001, 2014. 531 532 533
- [53] Masoud Mahdianpari, Bahram Salehi, Mohammad Rezaee, Fariba Mohammadimanesh, and Yun Zhang. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sensing*, 10(7):1119, 2018. 534 535
- [54] Cheolhee Yoo, Daehyeon Han, Jungho Im, and Benjamin Bechtel. Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 157:155–170, 2019. 536 537
- [55] Kiran Murthy, Michael Shearn, Byron D Smiley, Alexandra H Chau, Josh Levine, and M Dirk Robinson. SkySat-1: very high-resolution imagery from a small satellite. In *Sensors, Systems, and Next-generation Satellites XVIII*, volume 9241, pages 367–378. SPIE, 2014. 538 539
- [56] Ted R Feldpausch, Jon Lloyd, Simon L Lewis, Roel JW Brien, Manuel Gloor, Abel Monteagudo Mendoza, Gabriela Lopez-Gonzalez, Lindsay Banin, Kamariah Abu Salim, Kofi Affum-Baffoe, et al. Tree height integrated into pantropical forest biomass estimates. *Biogeosciences*, 9(8):3381–3403, 2012. 540 541 542
- [57] Ville Kankare, Markus Holopainen, Mikko Vastaranta, Eetu Puttonen, Xiaowei Yu, Juha Hyypä, Matti Vaaja, Hannu Hyypä, and Petteri Alho. Individual tree biomass estimation using terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 75:64–75, 2013. 543 544
- [58] Sassan S Saatchi, Nancy L Harris, Sandra Brown, Michael Lefsky, Edward TA Mitchard, William Salas, Brian R Zutta, Wolfgang Buermann, Simon L Lewis, Stephen Hagen, et al. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108(24):9899–9904, 2011. 545 546 547
- [59] Peter Potapov, Xinyuan Li, Andres Hernandez-Serna, Alexandra Tyukavina, Matthew C Hansen, Anil Komareddy, Amy Pickens, Svetlana Turubanova, Hao Tang, Carlos Edibaldo Silva, et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253:112165, 2021. 548 549 550
- [60] Michael A Lefsky. A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system. *Geophysical Research Letters*, 37(15), 2010. 551 552
- [61] Thuy Le Toan, S Quegan, MWJ Davidson, Heiko Balzter, Philippe Paillou, Konstantinos Papathanassiou, S Plummer, F Rocca, S Saatchi, H Shugart, et al. The biomass mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115(11):2850–2860, 2011. 553 554 555
- [62] Jerome Chave, Maxime Rejou-Mechain, Alberto Burquez, Emmanuel Chidumayo, Matthew S Colgan, Wellington BC Delitti, Alvaro Duque, Tron Eid, Philip M Fearnside, Rosa C Goodman, et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20(10):3177–3190, 2014. 556 557 558
- [63] Jerome Chave, Christophe Andalo, Sandra Brown, Michael A Cairns, Jeffrey Q Chambers, Derek Eamus, Horst Fölster, François Fromard, Niro Higuchi, Tatuo Kira, et al. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145:87–99, 2005. 559 560
- [64] Josselin Aval, Jean Demuynck, Emmanuel Zenou, Sophie Fabre, David Sheeren, Mathieu Fauvel, Karine Adeline, and Xavier Briottet. Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:197–210, 2018. 561 562 563
- [65] Mélaine Aubry-Kientz, Raphaël Dutrieux, Antonio Ferraz, Sassan Saatchi, Hamid Hamraz, Jonathan Williams, David Coomes, Alexandre Piboule, and Grégoire Vincent. A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, 11(9):1086, 2019. 564 565 566
- [66] Guillaume Lassalle, Matheus Pinheiro Ferreira, Laura Elena Cué La Rosa, and Carlos Roberto de Souza Filho. Deep learning-based individual tree crown delineation in mangrove forests using very-high-resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 189:220–235, 2022. 567 568 569
- [67] Mojdeh Miraki, Hormoz Sohrabi, Parviz Fatehi, and Mathias Kneubuehler. Individual tree crown delineation from high-resolution UAV images in broadleaf forest. *Ecological Informatics*, 61:101207, 2021. 570 571
- [68] Colin J Gleason and Jungho Im. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment*, 125:80–91, 2012. 572 573
- [69] Glenn J Newnham, John D Armston, Kim Calders, Mathias I Disney, Jenny L Lovell, Crystal B Schaaf, Alan H Strahler, and F Mark Danson. Terrestrial laser scanning for plot-scale forest measurement. *Current Forestry Reports*, 1:239–251, 2015. 574 575
- [70] Xinlian Liang, Juha Hyypä, Harri Kaartinen, Matti Lehtomäki, Jiri Pyörälä, Norbert Pfeifer, Markus Holopainen, Gábor Brolly, Pirotti Francesco, Jan Hackenberg, et al. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:137–179, 2018. 576 577 578
- [71] Erkki Tomppo, Thomas Gschwantner, Mark Lawrence, Ronald E McRoberts, K Gabler, K Schadauer, C Vidal, A Lanz, G Ståhl, E Cienciala, et al. National forest inventories. *Pathways for Common Reporting. European Science Foundation*, 1:541–553, 2010. 579 580
- [72] Ronald E McRoberts, Erkki O Tomppo, and Erik Næsset. Advances and emerging issues in national forest inventories. *Scandinavian Journal of Forest Research*, 25(4):368–381, 2010. 581 582

- [73] Claude Vidal, Iciar Alberdi, John Redmond, Martin Vestman, Adrian Lanz, and Klemens Schadauer. The role of European National Forest Inventories for international forestry reporting. *Annals of Forest Science*, 73(4):793–806, 2016. 583 584
- [74] Annika Kangas, Rasmus Astrup, Johannes Breidenbach, Jonas Fridman, Terje Gobakken, Kari T Korhonen, Matti Maltamo, Mats Nilsson, Thomas Nord-Larsen, Erik Næsset, et al. Remote sensing and forest inventories in Nordic countries—roadmap for the future. *Scandinavian Journal of Forest Research*, 33(4):397–412, 2018. 585 586 587
- [75] Erkki Tomppo, Håkan Olsson, Göran Ståhl, Mats Nilsson, Olle Hagner, and Matti Katila. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, 112(5):1982–1999, 2008. 588 589
- [76] Axel Berg, Magnus Oskarsson, and Mark O’Connor. Deep ordinal regression with label diversity. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2740–2747, 2021. 590 591
- [77] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019. 592 593
- [78] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollar. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 594 595
- [79] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018. 596 597
- [80] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016. 598
- [81] Christian Igel and Stefan Oehmcke. Remember to correct the bias when using deep learning for regression! *KI-Künstliche Intelligenz*, pages 1–8, 2023. 599 600
- [82] Abdou Bannari, Daniel Morin, F Bonn, and AjRsr Huete. A review of vegetation indices. *Remote Sensing Reviews*, 13(1-2):95–120, 1995. 601
- [83] Robert M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979. 602
- [84] Adele Cutler, D Richard Cutler, and John R Stevens. Random forests. *Ensemble machine learning: Methods and applications*, pages 157–175, 2012. 603 604
- [85] European Union. Copernicus land monitoring service 2018. European Environment Agency (EEA). 605
- [86] GL Baskerville. Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research*, 2(1):49–53, 1972. 606
- [87] Joseph Sill. Monotonic networks. *Advances in Neural Information Processing Systems*, 10, 1997. 607
- [88] Akira Komiyama, Jin Eong Ong, and Sasitorn Pongpam. Allometry, biomass, and productivity of mangrove forests: A review. *Aquatic Botany*, 89(2):128–137, 2008. 608 609
- [89] Martin Riedmiller and Heinrich Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *IEEE International Conference on Neural Networks*, pages 586–591. IEEE, 1993. 610 611
- [90] Compton J Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150, 1979. 612 613
- [91] Yaxiao Niu, Liyuan Zhang, Huihui Zhang, Wenting Han, and Xingshuo Peng. Estimating above-ground biomass of maize using features derived from UAV-based RGB imagery. *Remote Sensing*, 11(11):1261, 2019. 614 615
- [92] Oliver Sonnentag, Koen Hufkens, Cory Teshera-Sterne, Adam M Young, Mark Friedl, Bobby H Braswell, Thomas Milliman, John O’Keefe, and Andrew D Richardson. Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, 152:159–177, 2012. 616 617 618
- [93] Nikolaos G Silleos, Thomas K Alexandridis, Ioannis Z Gitas, and Konstantinos Perakis. Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto International*, 21(4):21–28, 2006. 619 620
- [94] E Raymond Hunt Jr, CST Daughtry, Jan UH Eitel, and Dan S Long. Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, 103(4):1090–1099, 2011. 621 622
- [95] Anatoly A Gitelson, Yoram J Kaufman, and Mark N Merzlyak. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3):289–298, 1996. 623 624