

Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR



Stefan Oehmcke ^{a,*}, Lei Li ^{a,*}, Katerina Trepekli ^b, Jaime C. Revenga ^b, Thomas Nord-Larsen ^b, Fabian Gieseke ^{c,a}, Christian Igel ^a

^a University of Copenhagen, Department of Computer Science, Universitetsparken 1, Copenhagen 2100, Denmark

^b University of Copenhagen, Department of Geosciences and Natural Resource Management, Øster Voldgade 10, Copenhagen 1350, Denmark

^c University of Münster, Department of Information Systems, Leonardo-Campus 3, Münster 48149, Germany

ARTICLE INFO

Edited by Marie Weiss

Keywords:

Climate change
Datasets
Neural networks
Forest biomass
LiDAR

ABSTRACT

Quantifying forest biomass stocks and their dynamics is important for implementing effective climate change mitigation measures by aiding local forest management, studying processes driving af-, re-, and deforestation, and improving the accuracy of carbon accounting. Owing to the 3-dimensional nature of forest structure, remote sensing using airborne LiDAR can be used to perform these measurements of vegetation structure at large scale. Harnessing the full dimensionality of the data, we present deep learning systems predicting wood volume and above ground biomass (AGB) directly from the full LiDAR point cloud and compare results to state-of-the-art approaches operating on basic statistics of the point clouds. For this purpose, we devise different neural network architectures for point cloud regression and evaluate them on remote sensing data of areas for which AGB estimates have been obtained from field measurements in the Danish national forest inventory. Our adaptation of Minkowski convolutional neural networks for regression give the best results. The deep neural networks produce significantly more accurate wood volume, AGB, and carbon stock estimates compared to state-of-the-art approaches. In contrast to other methods, the proposed deep learning approach does not require a digital terrain model and is robust to artifacts along the boundaries of the evaluated areas, which we demonstrate for the case where trees protrude into the area from the outside. We expect this finding to have a strong impact on LiDAR-based analyses of biomass dynamics.

1. Introduction

Robust quantification of forest carbon stocks and their dynamics is important for climate change mitigation and adaptation strategies (FAO and UNEP, 2020). Supporting this notion, the Paris Agreement (United Nations / Framework Convention on Climate Change, 2015) and the IPCC (Shukla et al., 2019) acknowledge that climate change mitigation goals cannot be achieved without a substantial contribution from forests. However, currently many countries do not have nationally specific forest carbon accumulation rates but rather rely on default rates from the IPCC 2018 (Masson-Delmotte et al., 2019; Requena Suarez et al., 2019), without accounting for finer-scale variations of carbon stocks (Cook-Patton et al., 2020).

Spatial details in the carbon budget of forests are often necessary to encourage transformational actions towards a sustainable forest sector

(Harris et al., 2012, 2021). Nevertheless, precise spatio-temporal monitoring of forest carbon dynamics at large scales has proven to be challenging (Erb et al., 2018; Griscom et al., 2017). This is due to the complex structure of forests, topographic features, and land management practices (Tubielo et al., 2021; Lewis et al., 2019). Technological developments in remote sensing and the concurrent increased availability of field-based measurements have led to an improvement in estimating carbon stocks using remote sensing observations of forest attributes that serve as proxy for above-ground biomass (AGB) (Knapp et al., 2018; Bouvier et al., 2015; Pan et al., 2013).

Currently, three remote sensing techniques are applied to collect data for AGB estimates: i) passive optical imagery, ii) synthetic aperture radar (SAR), and iii) light detection and ranging (LiDAR). LiDAR, as an active sensor, can penetrate dense forest canopies regardless of illumination conditions. In comparison, SAR observations are sensitive to

* Corresponding authors.

E-mail addresses: stefan.oehmcke@di.ku.dk (S. Oehmcke), llei@di.ku.dk (L. Li), atr@ign.ku.dk (K. Trepekli), jar@ign.ku.dk (J.C. Revenga), tnl@ign.ku.dk (T. Nord-Larsen), fabian.gieseke@wi.uni-muenster.de (F. Gieseke), igel@di.ku.dk (C. Igel).

topographical variations (e.g., steep slopes or cliffs), while imaging spectroradiometer products are highly dependent on the absence of clouds (Sinha et al., 2015; Treuhaft et al., 2014). Moreover, compared to LiDAR, both optical and radar-based data tend to saturate with increasing AGB exceeding 100 Mg ha⁻¹ (Mermoz et al., 2015; Sinha et al., 2015; Mitchard et al., 2012). Consequently, both alternatives to LiDAR may not be sufficient to detect variations in AGB in the region of interest (Goetz and Dubayah, 2011), leaving LiDAR as the better tool to accurately characterize the fine-scale spatial variability in forest structure and terrestrial carbon stocks (Zhang et al., 2019; Zolkos et al., 2013) across different spatial scales, depending on the platform used (i.e., airborne, drone-based, mobile, and terrestrial laser scanning). Recent developments in spaceborne LiDAR, such as the GEDI mission (Coyle et al., 2019), can be useful for identifying areas with high AGB (Lang et al., 2021). However, the coarse spatial resolution of these sensors (sparse 25 m footprints in case of GEDI Hancock et al., 2019) makes it impossible to derive fine structural information and to produce local as well as sufficiently accurate AGB estimates (Lang et al., 2022; Li et al., 2020; Zhang et al., 2019). In comparison, LiDAR data obtained from UAVs and terrestrial platforms have a very high dimensionality and provide even more detailed information about forest structure and the characteristics of individual trees. Due to the limited mobility of these platforms, however, they are merely suitable for local estimates of forest growing stock (Dalla Corte et al., 2020; Liu et al., 2018; Calders et al., 2018).

The state-of-the-art for predicting forest biomass from LiDAR point clouds is to voxelize the data and compute summarizing statistics of the point distribution along the vertical axis, such as mean heights, relative height quantiles, or metrics of heterogeneity (Magnussen et al., 2018; Knapp et al., 2018; d'Oliveira et al., 2012; Stark et al., 2012; Mascaro et al., 2011). Together with forest attributes obtained from inventory plots, these simple statistical features then serve as inputs to prediction models (Fassnacht et al., 2014) such as linear and non-linear regression, random forests, support vector machines, or neural networks, aimed at predicting the forest characteristics of interest. Based on this approach, several studies are focused on generalizing AGB estimations across different forest types describing different aspects of forest structure at different spatial scales (Bouvier et al., 2015; Asner and Mascaro, 2014; Lefsky et al., 2002). For instance, Knapp et al. (2020) assessed a variety of LiDAR metrics and auxiliary information (e.g., mean canopy height, maximum possible stand density, maximum possible tree height, vertical canopy heterogeneity, and average wood density) to detect structural forest attributes that may explain stand AGB across different regions. However, in this line of research, assessing the optimal combination of LiDAR metrics that can provide accurate AGB predictions has proven to be a challenging task.

An alternative to manually defining informative features based on summarizing point clouds statistics, which loose relevant structural information contained in the 3D data, is to use deep learning (LeCun et al., 2015) approaches that directly work on point clouds (Guo et al., 2020). Compared to deep learning on imagery, which as been been thoroughly explored within remote sensing for classification (Li et al., 2023; Tucker et al., 2023; Yuan et al., 2020; Brandt et al., 2018; Kussul et al., 2017) and regression (Liu et al., 2023; Lang et al., 2022; Zhang et al., 2019), deep learning on point clouds for regression is an almost unexplored learning setting, except for the estimation of geometric properties from point clouds (e.g., hand pose-estimation Ge et al., 2018; Firintepe et al., 2021) and bounding box prediction in general detection tasks (Zhou and Tuzel, 2018; Qi et al., 2019). In addition, ALS, a technology with peculiarities (Næsset, 2009; Goodwin et al., 2006), produces more challenging point clouds than most standard deep learning benchmark data sets for point cloud tasks, since they are measured from the top instead of from the ground, and the corresponding point densities are usually lower compared to other LiDAR-based technologies (e.g., terrestrial- or drone-based LiDAR). The applicability of a deep learning system to estimate crop biomass has also been tested more recently using point

clouds from LiDAR mounted on a roving vehicle (Pan et al., 2022) in a small scale experiment. When point density is high enough to make individual trees distinguishable, we generally expect models that use point clouds directly to give better AGB prediction results than models based on summary statistics, because the latter may be similar for areas with different AGBs.

This study brings forward deep learning systems to improve forest biomass and wood volume quantification based on airborne LiDAR data with high spatial resolution per sample and highly accurate estimates from on-site measurements. Specifically, this is the first work (the first idea of which was presented as a poster by Oehmcke et al., 2022) to apply deep neural networks directly to the point clouds to utilize structural information instead of relying on models being based on simple statistical features to derive forest characteristics. We hypothesize that maintaining the full dimensions of the point cloud, and hereby preventing the information loss associated with reducing dimensionality into a few statistical features strengthens predictions. The experimental evaluation shows that predictions using deep learning are considerably more accurate than currently employed methods on individual and aggregated plot level. Our approach requires neither a digital terrain model (DTM) nor a CHM. We also demonstrate that our chosen method is robust to border artifacts caused by trees reaching into plots, which is an essential when using high-resolution predictions for spatial aggregations, a feature not present in currently employed methods.

2. Methods

In the following, we first introduce the data used in this study in Section 2.1, which combines field-based biomass and wood volume measurements with ALS point clouds. Afterwards, in Section 2.2 the assessed deep learning architectures are presented along with their required adjustments for point cloud regression. The established baseline methods, which rely on statics derived from the point clouds are discussed in Section 2.3. Finally, we give details on why and how to correct the bias for least-square regression tasks in Section 2.4.

2.1. Forest data

This section outlines two data sources and their processing: first, the inventory data for AGB and wood volume and then the corresponding LiDAR data.

2.1.1. Forest inventory

The volume (in m³ ha⁻¹) and biomass (Mg (metric ton) ha⁻¹) estimates used to train and evaluate the models stem from the ground measurements of the Danish National Forest Inventory (NFI), see Fig. 2a for histograms of the above-ground forest biomass, carbon stock, and wood volume distributions. The resulting NFI statistics are used for formation and assessment of national or regional forest programs, sustainability assessments, investment calculations for forest industry, strategic-level planning in general, and reporting to international conventions. The Danish NFI is based on a grid with cells of size 2 × 2 covering the entire land surface of the country (Nord-Larsen and Johannsen, 2016). A plot is composed of four circular subplots with a radius of 15 m. Each subplot is located in the corners of a 200 × 200 square, which is randomly placed within each grid cell (Fig. 1a). The full set of plots is geographically partitioned into five spatially balanced interpenetrating panels (Kish, 1998; McDonald, 2003; Olsen et al., 1999; Zhang et al., 2003). Each year within a 5-year cycle, a different panel is measured on-site, representing one fifth of the total set but representative of the entire country. These labeled data used in our study range from 2013 to 2017.

In the Danish NFI, each subplot is composed of three concentric circles with radii of 3.5 m, 10 m, and 15 m, respectively. Following standard procedures, a single caliper measurement of diameter is made at breast height (d_{bh}) (i.e., 1.3 m from the ground) for all trees in the 3.5

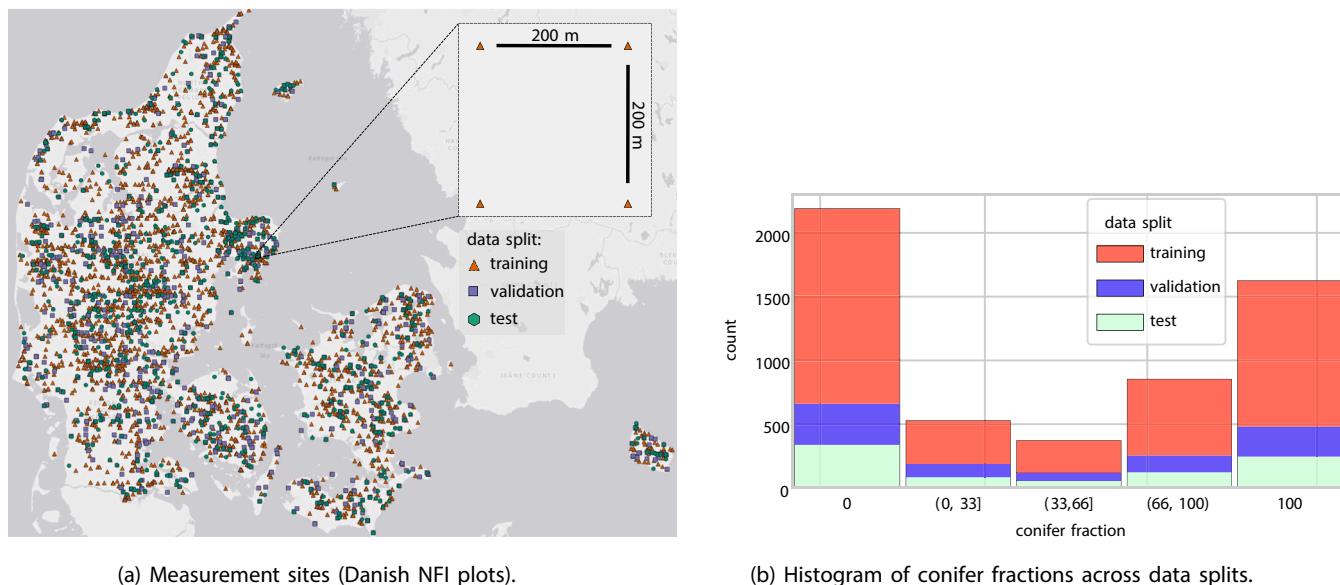


Fig. 1. The left plot (a) shows the spatial distribution of Danish NFI plots (map background based on (ESRI, 2023)). Colour indicates how the data were split into training, validation, and testing. The right plot (b) shows the species distribution by grouping the conifer fraction into 5 categories. The histograms for the different data sets are stacked.

m circle. Trees with d_{bh} larger than 10 cm are measured in the 10 m circle, and only trees with d_{bh} larger than 40 cm are recorded in the 15 circle (see Fig. 2b). For a random sub-sample, further measurements of the total height (h), crown base height (h_c), age, and diameter at stump height (d_{st}) of two to six trees within each subplot are also obtained.

The height of trees for which height was not measured is estimated from models developed for each species and growth region based on the sub-sample of trees for which h and d_{bh} were both measured. These models include the observed mean height and mean diameter within each subplot for creating localized regressions using the approach suggested by Sloboda et al. (1993). For subplots where no height measurements were available, generalized regressions were developed based on a modified Näslund-equation (Näslund, 1936; Johannsen, 1999). Subsequently, individual tree volume and biomass were estimated using species-specific models (Nord-Larsen et al., 2017a) (see Fig. 1b for an overview of the species distribution). Since trees were measured in different concentric circles depending on their diameter, the volume of growing stock and biomass in each subplot were calculated by scaling the estimated AGB and wood volume of each tree according to the circular area in which the tree was measured. Carbon stocks can be considered as a linear function of the forest biomass, thus, in this study was not used as a separate regression target.

2.1.2. Point cloud data

Airborne LiDAR point cloud data covering Denmark from 2014 to 2018 are publicly available from the Danish Agency for Data Supply and Efficiency (<https://dataforsyningen.dk>). Specifically, a sampling campaign in 2014 and 2015 covered the entire country, while a fifth of the country was again scanned in 2018 (in eastern Jutland and Funen). The vertical and horizontal point accuracy is reported to be 5 and 15 (given as root mean squared error (RMSE)), respectively. We extracted the point clouds corresponding to the subplots studied in the NFI (see Fig. 3). The average number of points per subplot are 11,684 with a standard deviation of 6743 points, which corresponds to an average of 16.7 points per m².

2.1.3. Preprocessing

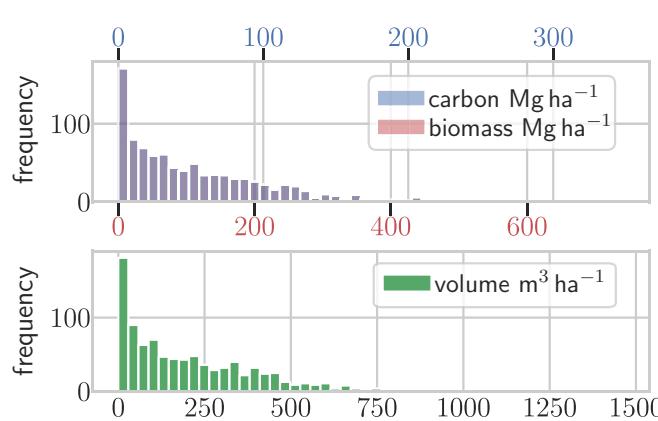
Several preprocessing steps were conducted for both the deep neural networks (operating on the point clouds) as well as for the baseline models that operate on simple statistical features.

For the plot-specific point clouds, the NFI identified four types of errors. In the first type, trees that are visible in the point cloud do not belong to the field-based measurements but reach into the plot from the outside, while in the second type non-forest objects, such as land-lines or buildings, are part of the point cloud. We chosen not to exclude these data since a method that directly works with the point cloud should detect irrelevant parts by itself, meaning that some previously excluded data should now be usable. The third error type occurs when the plot had been harvested in-between field measurements and airborne point cloud acquisition. Finally, the fourth error type consisted of unreasonable values. Since the third and fourth error flags are intractable, the samples were excluded from the model training. In addition, all samples with no points above 1.3 m were removed, as this was the minimal height considered for the biomass measurements.¹

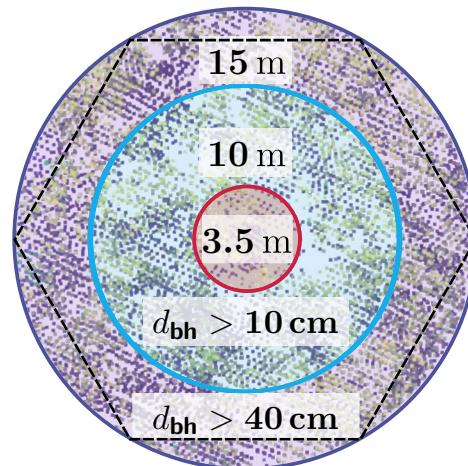
The remaining dataset consists of $n = 6099$ individual point clouds, each being labeled with a corresponding biomass and wood volume measurement as described above. The biomass and wood volume were measured at different times than the collection of point clouds. To ensure that both data sources matched, we only considered point clouds observed within one year of the biomass measurement for our validation and test set. For training, we took gaps larger than one year (up to nine years) into account.

Finally, to train, validate, and test all the models, we split the data as follows: the training set contains 2636 samples with a time interval of no more than one year and 1634 samples with a time interval longer than a year (i.e., 4270 training samples in total, in 3938 unique subplot location). The validation set consists of 918 and the test sets of 911 samples at 903 and 901 unique subplot locations, respectively, with a maximum distance of one year between paired point cloud and biomass measurements. Since some subplots appeared twice due to a subplot being measured again on the ground or via ALS, we ensured that measurements from the same site stayed together in either training, validation, or test datasets (i.e., no site/plot considered during training or for validation occurred in the test set).

¹ Keeping these samples would simply reduce the errors, because they could be handled by a simple rule that predicts zero if the maximum height is below 1.3 m.



(a) Histogram of above-ground forest biomass, carbon stock, and wood volume.



(b) Subplot measurement circle.

Fig. 2. The left plot (a) shows the distribution of target variables, whereby on the y-axis is the frequency of NFI plots within specific value bins on the x-axis. Above-ground biomass and carbon stocks overlap completely since the carbon stock estimate is computed as a linear function of the biomass. As expected, the distributions appear to be log-normally distributed. The right plot (b) shows exemplary input point cloud (top view of a single subplot) and a representation of the measurement circles. The dashed hexagon encloses the points used by the deep learning point cloud methods.

2.2. Deep regression for point clouds

From a mathematical perspective, a single point cloud can be represented via a set $\mathcal{P} = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_{\mathcal{P}}}\} \subset \mathbb{R}^N$, where $n_{\mathcal{P}}$ is the number and N the dimensionality of the points (typically, $N = 3$). Besides these coordinates, additional information is often provided for each point, such as colour or intensity. These features paired with the points define the set

$$\mathcal{P}_{\mathcal{F}} = \{(\mathbf{x}_1, \mathbf{f}_1), \dots, (\mathbf{x}_{n_{\mathcal{P}}}, \mathbf{f}_{n_{\mathcal{P}}})\} \subset \mathbb{R}^N \times \mathbb{R}^D \quad (1)$$

with $\mathbf{f}_i \in \mathbb{R}^D$ containing the additional features. Such a point cloud is given for each instance (e.g., one subplot corresponds to one point cloud, see Fig. 3). The point cloud sets typically have different cardinalities and are unordered. Accordingly, the output of a model should not depend on the order of the provided points, a property referred to as permutation invariance. Furthermore, the spatial structure of such point clouds can be highly irregular and the density of points is typically very sparse (such characteristics usually vary among different regions as well).

Compared to deep convolutional neural networks (CNNs) for standard 2D and 3D images, the research field of deep learning for point cloud data has yet to mature. There are several approaches for adapting neural networks for point cloud processing, differing among others in how they deal with the key questions of how to extend the concept of spatial convolutions to sparse point clouds. Most deep learning systems for point clouds are developed for segmentation and other classification tasks (Hackel et al., 2017; Chang et al., 2015; Wu et al., 2015). In this study, we focused on regression tasks. We adapted and compared three widely-used and conceptually different point cloud classification approaches for regression: the classic PointNet (Qi et al., 2017), the kernel point convolution (KPConv) (Thomas et al., 2019), and a Minkowski CNN (Choy et al., 2019).

We extended the open-source library Torch-Points3D (Chaton et al., 2020) to support regression and will contribute our networks and anonymized dataset to the library upon acceptance of this manuscript. Two modifications were applied to all networks. First, we replaced Rectified Linear Unit (ReLU) activation functions with Gaussian Error Linear Units (GELU) (Hendrycks and Gimpel, 2016). Secondly, we replaced the last global pooling operation (PointNet: max, KPConv and Minkowski: average) with global sum pooling to better reflect the task of aggregating AGB or wood volume from different areas of the point cloud scene.

2.2.1. PointNet

The seminal PointNet is one of the first deep learning architectures developed for point clouds (Qi et al., 2017). To compute new representations $\mathbf{x}'_i \in \mathbb{R}^{D_{\text{out}}}$ for $i \in \{1, \dots, n_{\mathcal{P}}\}$, PointNet-type networks apply the same linear map $\mathbf{x}'_i = W\mathbf{x}_i$ with $W \in \mathbb{R}^{D_{\text{out}} \times D_{\text{in}}}$ to each input $\mathbf{x}_i \in \mathbb{R}^{D_{\text{in}}}$ individually without any interaction between inputs. The original architecture is based on the use of multiple such neural network blocks. In the first layer, the position and additional features are concatenated ($D_{\text{in}} = N + D$), which does not support translation invariance. In between these blocks, input/feature transformations are used to align the data via learnable transformation matrices. At the end, to facilitate a variable number of points and to achieve permutation invariance, a symmetric set function is applied, which usually is a global max-pooling operation. Note that there is only a global interaction between points, local contexts where introduced by follow up works using hierarchical structures (e.g., see KPConv below).

Two PointNet variants have been proposed in the original work to address classification and segmentation scenarios, respectively. To adapt the architecture for regression, we mostly adopted the classification PointNet architecture but removed the learnable transformation matrices since they did not improve the validation scores. Also, we used batch normalization after each hidden layer.

2.2.2. KPConv

The kernel point convolution (Thomas et al., 2019) extends standard discrete convolution to the domain of point clouds. Kernel point convolution as well as discrete convolution on images use a kernel to extract local information. However, KPConv uses a set of *kernel points* instead of a discrete kernel mask to define the convolution kernel, which allows for a representation in Euclidean space instead of a rasterized space. More precisely, if $\mathcal{P}_{\mathcal{F}}$ is convolved by a kernel g , the result at a point $\mathbf{x} \in \mathbb{R}^N$ is defined as

$$(\mathcal{P}_{\mathcal{F}} * g)(\mathbf{x}) = \sum_{(\mathbf{x}_i, \mathbf{f}_i) \in \mathcal{N}_{\mathbf{x}}} g(\mathbf{x}_i - \mathbf{x}) \mathbf{f}_i. \quad (2)$$

Here, the local neighborhood is $\mathcal{N}_{\mathbf{x}} = \{(\mathbf{x}_i, \mathbf{f}_i) \in \mathcal{P}_{\mathcal{F}} \mid \|\mathbf{x}_i - \mathbf{x}\| \leq r\}$, where the radius $r \in \mathbb{R}^+$ is a hyper-parameter defining the domain $\mathcal{B}_r^N = \{\mathbf{y} \in \mathbb{R}^N \mid \|\mathbf{y}\| \leq r\}$ of the kernel g , which is similar to the kernel size in discrete convolution. The kernel g is based on a set $\{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_K\} \subset \mathcal{B}_r^N$ of K kernel points, where each kernel point $\tilde{\mathbf{x}}_k$ has an associated weight

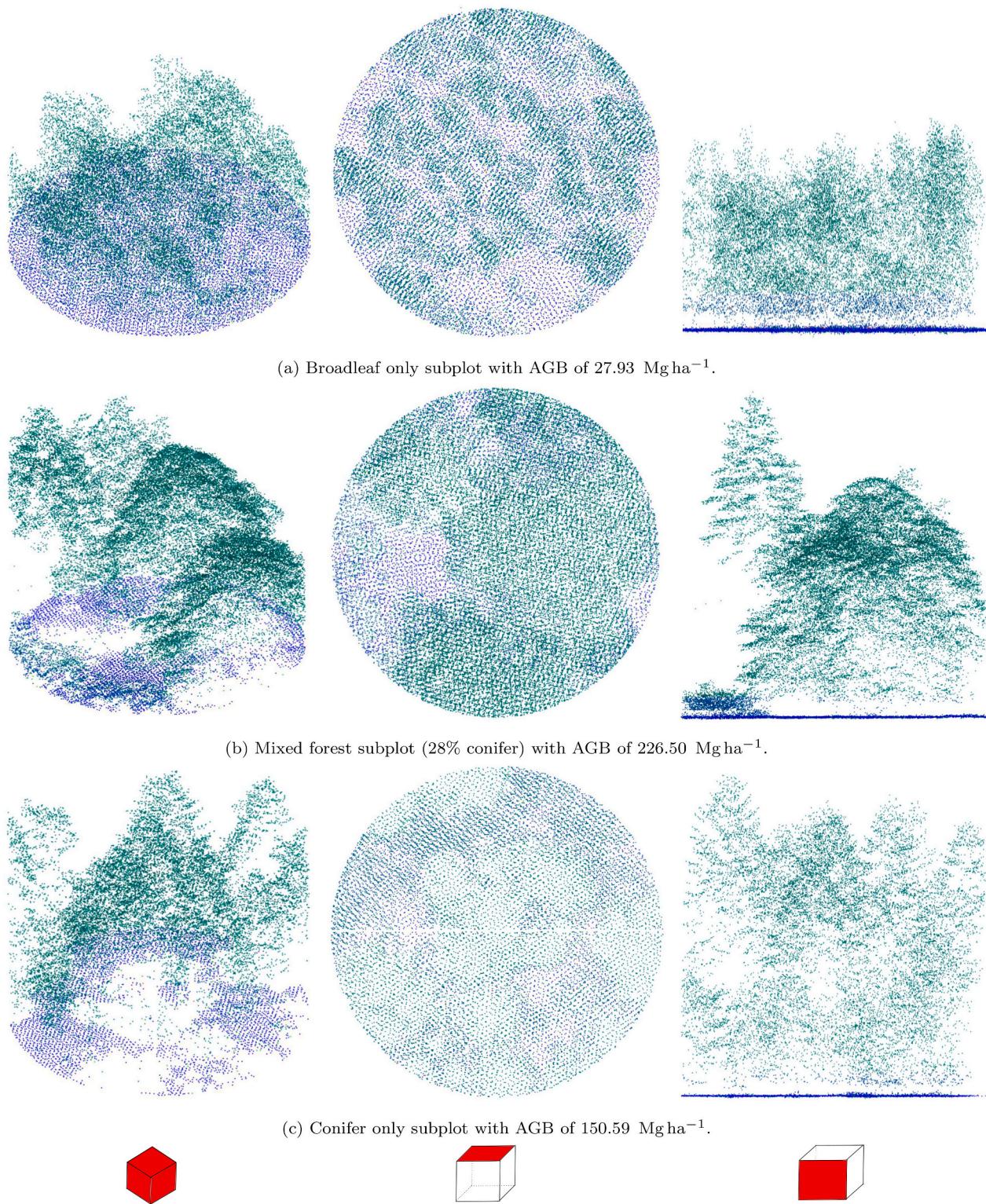


Fig. 3. Examples of subplots with different fractions of broadleaf and conifer trees. Green and blue indicate points classified belonging to trees and ground, respectively. We show three perspectives: isometric front, top, and side view. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

matrix $W_k \in \mathbb{R}^{D_{\text{in}} \times D_{\text{out}}}$, which maps an input feature vector $\mathbf{f} \in \mathbb{R}^{D_{\text{in}}}$ to a new feature vector $\mathbf{f}' \in \mathbb{R}^{D_{\text{out}}}$ (for the first such convolution layer, we have $D_{\text{in}} = D$). These kernel points are conceptually similar to a kernel filter matrix in a CNN. In KPConv the kernel $g : \mathcal{B}_r^N \rightarrow \mathbb{R}^{D_{\text{in}} \times D_{\text{out}}}$ is defined as

$$g(\mathbf{y}) = \sum_{k=1}^K h(\mathbf{y}, \tilde{\mathbf{x}}_k) W_k, \quad (3)$$

where $h(\mathbf{y}, \tilde{\mathbf{x}}_k) = \max\left(0, 1 - \frac{\|\mathbf{y} - \tilde{\mathbf{x}}_k\|}{\sigma}\right)$ is an inverted distance function and

$\sigma \in \mathbb{R}^+$ is a user-defined model parameter (Thomas et al., 2019). Thus, the kernel yields a weighted sum of weight matrices depending on the distance of a kernel point \tilde{x}_k to an input point y . Each kernel is parameterized by its own kernel points and their corresponding weight matrices. The weight matrices are learnt similarly to the coefficients of convolution operators in standard CNNs. The position of kernel points can be either learned (“deformable”) or be systematically fixed (“rigid”) around the center of \mathcal{B}_r^N (Thomas et al., 2019). Pooling operations are realized via creating a 3D grid in which all points in a cell are aggregated (see Section 2.2.3).

We adapted the KPConv architecture proposed for classification to utilize it for regression. This architecture follows the structure of ResNet, thus modifying it for regression was straightforward and was achieved by changing the output layer to have the identity as activation function. Fixed kernel points were preferred over deformable ones, since the latter consistently performed worse on our validation dataset.

2.2.3. Minkowski convolutional neural network

A straightforward approach to process point clouds is to voxelize the points to a 3D image (with D channels) and to apply a standard 3D convolutional neural network (CNN) to the resulting image. This requires spatial binning, but while a coarse binning may discard important information, a fine-grained binning can render the computations intractable. However, the points are typically very sparse in 3D space. This lends to keeping a sparse representation and applying spatially sparse 3D convolutions, which basically operate only in areas where

points exist (Graham, 2015; Graham et al., 2018). A state-of-the-art implementation of sparse convolutions for point clouds is given by the Minkowski Engine, an auto-differentiation library, which allows to efficiently implement standard architectures, such as ResNet with 3D convolutions (Choy et al., 2019).

Both the Minkowski and KPConv architectures progressively downsample the point cloud based on a 3D grid via pooling operations. The initial grid size is only bound to the horizontal and vertical accuracy of the point data, meaning that a too small grid size cannot be processed efficiently. Therefore, it is crucial to set the initial grid size according to the data and task. The subsequent pooling steps are then based on this initial size and work similarly to standard 2D pooling operations. Once a grid is created, a voxel-based approach, such as Minkowski CNNs, aggregates all voxels within a grid cell to a new voxel. In contrast, KPConv takes the mean of the position of all points within a cell to form a new point. By doubling the size of the grid cells in each subsampling/pooling layer, the number of points/voxel is gradually reduced, corresponding to an increase in the size of the receptive field. The features associated with each new point can be obtained, for example, via max-pooling (applied to the feature vectors of the pooled points) or by applying the chosen convolution operator.

Using the Minkowski Engine library, we built two 3D versions of SENet (SENet14 and SENet50) for regression (Hu et al., 2018), see Fig. 4. To that end, we replaced 2D convolutions with their 3D equivalent. The stride of the initial convolution of SENet was changed from 2 to 1 in order to avoid early loss of information. Finally, the output layer was

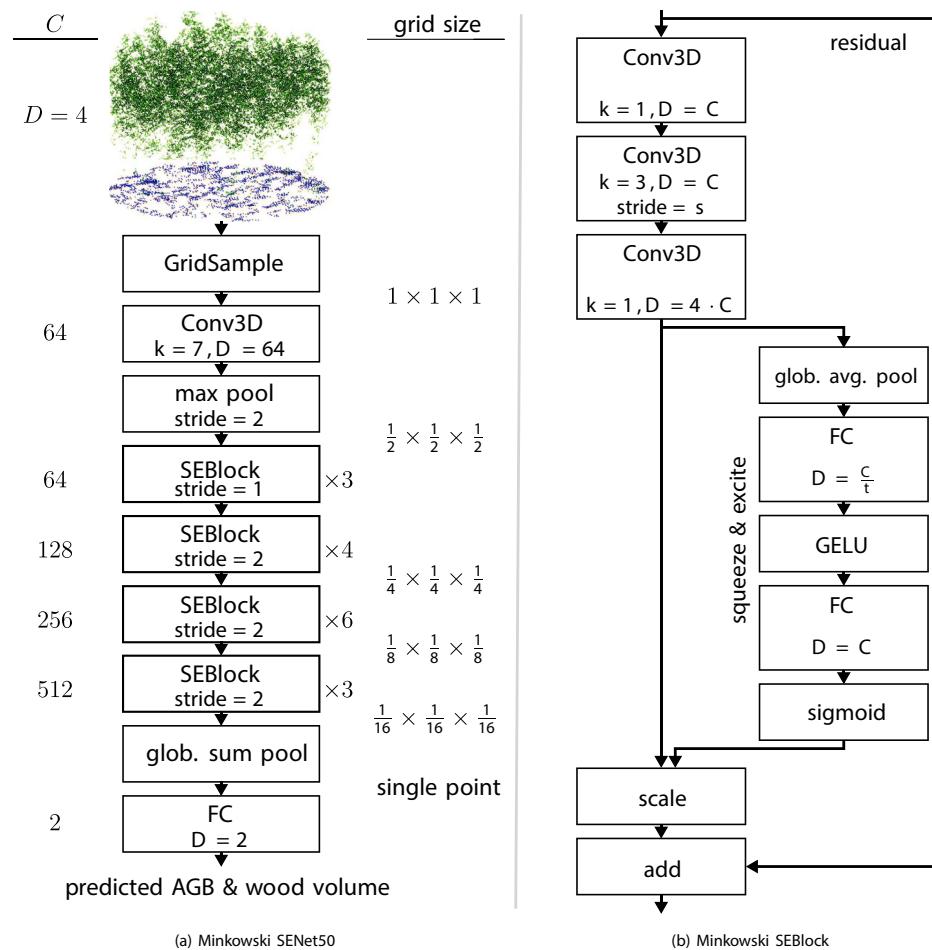


Fig. 4. (a) Implemented Minkowski SENet50 (MSENet50) architecture. The grid sizes are relative to the initial grid size. Conv3D refers to sparse 3D convolution with D channels followed by GELU activation and batch normalization. The kernel size k and stride are given as single numbers but extend to all three dimensions (e.g., $3 = 3 \times 3 \times 3$). FC layers are fully connected layers of dimensionality D . (b) Bottleneck block, parameterized by its input size C , stride s , and reduction rate t (defaults to 16).

changed to fit the number of target variables and does not apply a non-linear activation. The two network versions differ in the number of trainable parameters as shown in [Table 1](#) and are referred to as MSENet14 and MSENet50, respectively.

2.2.4. Experimental setup for forest attribute regression

Three different neural network architectures were applied: the proposed regression versions of PointNet, KPConv, and two 3D version of SENet (MSENet14 and MSENet50) built with the Minkowski Engine. All code was written in Python and PyTorch ([Paszke et al., 2019](#)) and trained on A100 GPUs. The positions of the input points were given in the projected coordinate system EPSG:25832. We centered the x -, y -, and z -coordinates for each subplot individually to get a local reference frame. Scaling was applied by dividing by 15, 15, 20 m, respectively, so that the coordinates lie within $[-1, 1]$ for the x - and y -coordinates. The z -scaling of 20 m was chosen as it represents the expected value range for the trees we were studying, with the highest trees being around 40 m. We do not need to calibrate our model using a DTM, as is required by state-of-the-art methods based on point cloud statistics.

We tuned the initial grid size of KPConv and the MSENet models by gradually decreasing it (i.e., increasing the spatial resolution) until no improvement could be seen on the validation set. The idea behind this is that larger grids can be processed faster but at the cost of losing structural details and finer grids offer more structural details but require longer to process. By gradually reducing the grid size, we can ensure good results without unnecessary computations. Both architectures worked best with the initial grid size set to 0.025. The number of trainable model parameters and runtimes are given in [Table 1](#).

To increase diversity in the training set, several augmentations were considered. Augmentation generates distinct training examples from the given training data set by applying random transformations to the input data that with high probability do not affect the dependent variable ([Zhang et al., 2023](#)). This helps to avoid overfitting because exactly the same input is never presented to the learning system several times. Augmentation improves generalization because it forces the system to learn general features of the input which are invariant under the applied transformations. First, a random rotation around the z -axis was applied since the predictions should be invariant to such rotations. We did not consider rotating the x - and y -axis since a tilt would result in unrealistic samples (e.g., change of growth direction). Other transformations, such as scaling or shearing, were also not applied, as they may invalidate the target label. Thereafter, random dropout of points with a dropout rate of 20% in 50% of all samples was applied, which increases variety and simulates changes in point density. As a last step, each dimension of the point position was shifted by adding a small random value drawn from a zero-mean truncated Gaussian distribution with variance 0.001 and support $[-0.05, 0.05]$ to simulate slight positional errors and shifts. Note that these augmentations are computationally inexpensive and did not add noticeably more time to the training procedure. We only considered augmentations for the neural network models, since the summarizing statistics (which typically are, e.g., rotation invariant by definition) would not change drastically and the baseline methods are not prone to overfitting.

Trees reaching into subplots can change the results of the calculated statistics. For that reason, these kind of plots are often excluded from the

Table 1

Number of trainable parameters and average runtime (elapsed real time using a single GPU) for one pass through the test set ($n = 911$) as well as the total training time for the point cloud models.

| Model | number of parameters | test time | train time |
|----------|----------------------|-----------|------------|
| PointNet | 799,426 | 3 s | 4 h |
| KPConv | 6,105,698 | 28 s | 20 h |
| MSENet14 | 14,447,870 | 3 s | 4 h |
| MSENet50 | 48,760,626 | 6 s | 9 h |

model phase (see type 1 errors in [Section 2.1.3](#)). We decided to keep them as our structure-aware model should be resilient to these border effects. To demonstrate this resilience, we artificially changed the point clouds to simulate border effects where trees would reach into subplots. We used 1482 single ALS tree scans from Germany ([Weiser et al., 2022](#)), where each scan contains the point cloud of an isolated tree without other trees or ground points. We estimated the stem location of each of these trees by calculating the center of points below 2.5 m. If the positional deviation of these points from the estimated center was larger than 5 m, indicating a high uncertainty of the stem location, the tree was discarded. This step removed 130 trees from the dataset, leaving 1353 for the subsequent steps. We added up to 10 randomly chosen trees by positioning each tree center somewhere outside of the subplot (15.1 m – 20.1 m away from the center), which we refer to as *tree-adding augmentation*. During the evaluation phase two runs were conducted, one without and one with these augmentations to each subplot and the results are shown separately.

One goal was to create precise wood volume and above-ground biomass maps over large areas. Therefore, instead of using the entire circle in which the measurements were taken, we resorted to a hexagonal shape in the x - and y -coordinate, see [Fig. 2b](#). This approach facilitates an easy tiling of large areas without gaps and overlaps. The hexagonal shape covers almost all of the points, in particular those in the inner circles (in control experiments, we verified that considering the whole subplot does not lead to a significant performance increase).

We employed the AdaBelief ([Zhuang et al., 2020](#)) optimizer with weight decay set to 0.01 to update the weights of the neural networks. The initial learning rate was 0.005, which was adapted according to the cosine annealing schedule with warm restarts ($T_0 = 10$, $T_{\text{mult}} = 2$) ([Loshchilov and Hutter, 2017](#)) for 310 epochs. The batch size was 32 and the smooth L1 function [Girshick \(2015\)](#) was used as training loss. These hyper-parameter configurations were chosen based on the highest R^2 score on the validation set. After choosing the network weights with the highest R^2 score on the validation set, we only adjusted the running average and variance of the batch normalization layers for another 10 epochs to stabilize training without altering the network weights. Three input channels were given per point: the distance to subplot center and the distance to the lowest point to encode the relative position as well as a channel filled with ones to encode the presence of a point.

2.3. Baseline methods

The baseline models considered in this study for estimating above-ground biomass, rely on precomputed features from the normalized height distribution in accordance with [Nord-Larsen et al. \(2017b\)](#): mean height, height standard deviation, coefficient of variation, skewness, kurtosis, percentiles of the height distribution (5th, 10th, 25th, 50th, 75th, 90th, 95th, 99th), and the interception ratio (IR), which is the fraction of points above one meter compared to all first returns. For each point cloud sample, these features except the IR were computed twice, once for all points and once only for the points 1 m above ground. To determine the height of a point relative to the ground, we calibrated it semi-automatically, using two DTMs with different spatial resolutions (25 m² and 5 m²), and the first return of each light pulse from the airborne LiDAR, following the study of [Magnussen et al. \(2018\)](#). In preliminary analysis, it was observed that this approach led to better results than using all the recorded returns from each LiDAR-derived light pulse. In contrast, the deep learning methods simply use all the information included in the point clouds without the need to filter out the point cloud datasets according to the number of returns. In total, $d = 28$ statistical features were extracted for each point cloud sample.

2.3.1. Linear and power regression

The first two baseline methods were linear regression and power regression models. For the linear regression models, all the d features

and the number of years between LiDAR and NFI measurement were used as explanatory variables. The multivariate power regression can be regarded as the state-of-the-art for predicting AGB, above-ground carbon stocks, and tree volume from the point cloud features (Magnussen et al., 2018). Following Magnussen et al. (2018), we fitted models of the form

$$y = w_1 \cdot z_{\text{meanAG}}^{w_2} \cdot z_{95^{\text{th}} \text{ AG}}^{w_3} \cdot \text{IR}^{w_4} \quad (4)$$

for each quantity of interest. Here, z_{meanAG} is the mean height of the points one meter above ground, $z_{95^{\text{th}} \text{ AG}}$ is the 95th percentile of the height distribution of point one meter above ground, and IR is the inception ratio. The power regression performed worse when including the training data with gaps larger than a year between LiDAR and NFI measurements, which is why only the model using data from within a year is used. For each model, the parameters $w_1, \dots, w_4 \in \mathbb{R}$ were set to the optimal least-squares estimates.

2.3.2. Random forest

We trained random forest (RF) models with 5000 trees (Breiman, 2001) and resorted to the popular Scikit-Learn implementation (Pedregosa et al., 2011) using the same feature set as the linear regression. The out-of-bag (OOB) error was used to tune the RF hyper-parameters. We considered $\{0.1, 0.2, \dots, 1\}$ for the ratio of features to consider at every split, $\{0.1, 0.2, \dots, 1\}$ for the ratio of samples to consider for each tree, $\{5, 6, \dots, 20, \infty\}$ for the maximum tree depth, and $\{1, 2, 4, \dots, 16\}$ for the minimum number of samples required at each leaf node. The best model used 80% of the features at every split, 10% of the training data for each tree, had a maximum depth of 11, and required a minimum number of 4 samples at each leaf node.

2.4. Bias correction

Least-squares regression with deep learning models typically lead to biased models in the sense that the sum of residuals on the training data set is not zero (Igel and Oehmcke, 2023). Consider a regression model $\mathbb{R}^d \rightarrow \mathbb{R}$ of the form

$$f(\mathbf{x}) = \mathbf{a}^T h_\theta(\mathbf{x}) + b \quad (5)$$

with parameters (θ, \mathbf{a}, b) . Here $h_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^m$ for some positive integer m , $b \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^d$. This model can be a deep neural network, where all layers but the final layer are represented by the function h_θ and the parameters θ comprise all weights except those in the final layer. Training a deep neural network for regression using iterative optimization with mini-batches and stopping after a fixed training time or based on performance on a hold-out data set cannot be expected to choose the parameter b in a way that the residuals on the training data sum to zero. That is, on the training data $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ we have $\sum_{i=1}^N (y_i - (\mathbf{a}^T h_\theta(\mathbf{x}_i) + b)) \neq 0$. This can introduce a systematic error that accumulates if we are interested in the total aggregated performance over many data points. We therefore used the method proposed by Igel and Oehmcke (2023) to correct this bias on the training and validation set for all non-linear methods (all except linear regression) by replacing b by

$$b^* = \frac{\sum_{i=1}^N (y_i - \mathbf{a}^T h_\theta(\mathbf{x}_i))}{N}. \quad (6)$$

This bias correction can also be applied to other non-linear models h_θ (e.g., h_θ could be a random forest); h_θ can be wrapped according to (5) with scalar $a = 1$ and b set to b^* computed by (6).

2.5. Evaluation metrics

We compared methods for predicting AGB (and thereby above-ground carbon stocks) and wood volume using different metrics, the root mean square error

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2}, \quad (7)$$

the coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2}{\sum_{i=1}^N \left(y_i - \frac{\sum_{j=1}^N y_j}{N} \right)^2}, \quad (8)$$

and the mean absolute percentage error

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - f(\mathbf{x}_i)}{y_i} \right|. \quad (9)$$

The MAPE is asymmetrical and by definition weights errors of the same magnitude higher when the target values are smaller, but we report it in accordance with the related literature. We also study the mean bias

$$\text{MB} = \frac{1}{N} \sum_{i=1}^N y_i - f(\mathbf{x}_i) \quad (10)$$

to emphasize differences over aggregated predictions.

3. Results

An overview of the results is given in Table 2 showing the R^2 , RMSE, and MAPE for the different algorithms on the test set, which was not used for model development. Overall, the MSENet50 model closely followed by the MSENet14 model achieved the best results with an R^2 of 0.835 for biomass (and carbon stocks) and an R^2 of 0.831 for wood volume. The corresponding RMSEs were 41.10 Mg ha^{-1} and $78.38 \text{ m}^3 \text{ ha}^{-1}$, respectively.

The RF performed the worst. The better performance of the linear regression with an R^2 of 0.762 for biomass and 0.766 for wood volume, indicates that the calculated features are to a certain extend linearly correlated with the target variables. The power regression model based on Magnussen et al. (2018) reached R^2 values of 0.761 and 0.763 for biomass and tree volume, respectively, that is, the deep learning approach clearly improved over the state-of-the-art methods. All more recent point cloud based methods (MSENet and KPConv) outperformed the models based on the precomputed features, justifying that the networks indeed utilized additional information from the point clouds. PointNet only achieved statistically different performance for wood volume. The MAPE, which was not an optimization criterion, showed the largest differences between the methods, which are mainly attributed to errors related to small target values.

The national forest inventory data offers information about the fraction of conifer and broadleaf for each site, which we used to analyze the performance at different fractions in Fig. 5. Again, the high MAPE for the baseline methods that rely on the precomputed features and PointNet was noticeable, in particular when monocultures were analyzed. In general, monocultures turned out to be more difficult than mixed forests w.r.t. MAPE but not w.r.t. RMSE and R^2 (e.g., forests with conifer trees only exhibited the lowest RMSE and R^2). The decrease in RMSE with increasing proportion of conifers supports the intuition that the more structurally complex broadleaf trees are more difficult to predict than conifers. In Fig. 6, we show the distributions of the predictions by the power model and the MSENet model as well as the ground truth NFI data. The power model predictions consistently have lower variance, while the MSENet14 model better matches the actual distribution.

Residual errors increased with magnitude of the target variables as seen in Fig. 7 for the power, KPConv, and MSENet14 model. The results also show a symmetric distribution of the residuals around zero,

Table 2

Comparison of methods showing the R^2 score, RMSE, MAPE, and MB on the test set. The R^2 and MAPE for biomass are the same as for carbon stocks. Note that we excluded 0 biomass measurements from MAPE to avoid numerical issues. RMSE and MB are in Mg ha^{-1} for AGB and $\text{m}^3 \text{ha}^{-1}$ for wood volume. The median (med.) and best result (min or max) of 5 trials are shown, the best results are highlighted in bold font. The median for the MB is based on the absolute values, which is more meaningful when combining multiple trials. The min values are also based on the absolute values, however the signs are indicated in the table. To highlight significant differences between power and the other models, we employ the paired Wilcoxon signed-rank test (H_0 : The paired samples in each subplot have a median difference of zero), where a sample is the absolute error for each subplot ($n = 911$). Different p -values are indicated: $p > 0.1$ by 'ns', $p \leq 0.05$ by '**', and $p \leq 0.001$ by ***.

| Target | model | R^2 | | RMSE | | MAPE | | MB | |
|---------------|----------|--------------|--------------|--------------|--------------|--------|--------|-------------|-------------|
| | | med. | max | med. | min | med. | min | med. | min |
| AGB | | | | | | | | | |
| ns | linear | 0.762 | 0.762 | 49.42 | 49.42 | 425.61 | 425.61 | 1.89 | 1.89 |
| | power | 0.761 | 0.761 | 49.51 | 49.51 | 365.34 | 365.34 | 2.03 | 2.03 |
| ns | RF | 0.754 | 0.754 | 50.19 | 50.16 | 625.44 | 616.64 | 1.47 | 1.46 |
| ns | PointNet | 0.770 | 0.773 | 48.55 | 48.27 | 218.52 | 142.12 | 0.54 | 0.12 |
| ** | KPConv | 0.800 | 0.815 | 45.25 | 43.54 | 90.24 | 84.42 | 0.46 | 0.39 |
| ** | MSENet14 | 0.825 | 0.829 | 42.37 | 41.80 | 179.98 | 103.58 | 0.67 | -0.29 |
| ** | MSENet50 | 0.827 | 0.835 | 42.09 | 41.10 | 155.99 | 105.00 | 0.84 | -0.11 |
| volume | | | | | | | | | |
| ns | linear | 0.766 | 0.766 | 92.29 | 92.29 | 171.48 | 171.48 | 4.60 | 4.60 |
| | power | 0.763 | 0.763 | 92.82 | 92.82 | 223.64 | 223.64 | 4.50 | 4.50 |
| ns | RF | 0.757 | 0.757 | 94.09 | 94.07 | 223.65 | 222.60 | 3.98 | 3.96 |
| * | PointNet | 0.777 | 0.782 | 90.16 | 89.16 | 135.46 | 84.64 | 1.99 | 1.37 |
| ** | KPConv | 0.800 | 0.805 | 85.41 | 84.23 | 68.62 | 60.00 | 0.38 | 0.28 |
| ** | MSENet14 | 0.823 | 0.826 | 80.31 | 79.62 | 76.91 | 68.16 | 0.51 | 0.39 |
| ** | MSENet50 | 0.825 | 0.831 | 79.90 | 78.38 | 90.01 | 68.45 | 0.17 | 0.12 |

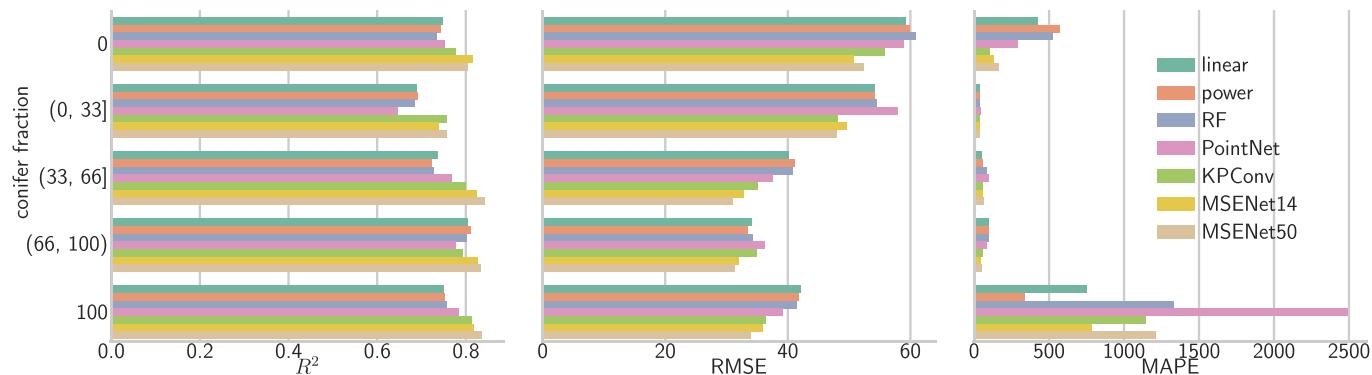


Fig. 5. Errors for different fractions of conifer and broadleaf trees for biomass on the test set. The columns show R^2 (higher is better), RMSE (lower is better), and MAPE (lower is better). The corresponding figure for wood volume is qualitatively the same and can be found in the supplementary material (see Fig. B.10).

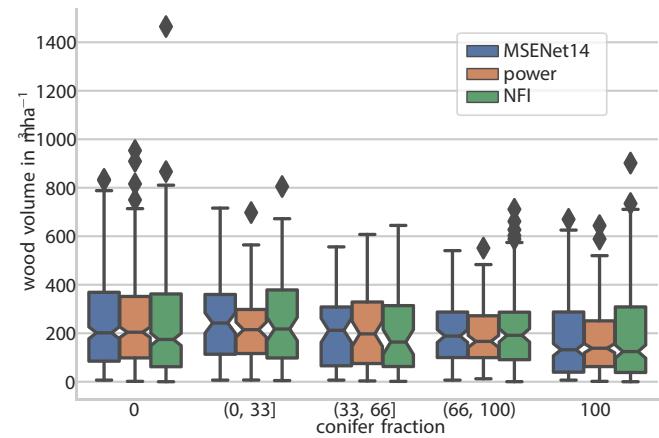
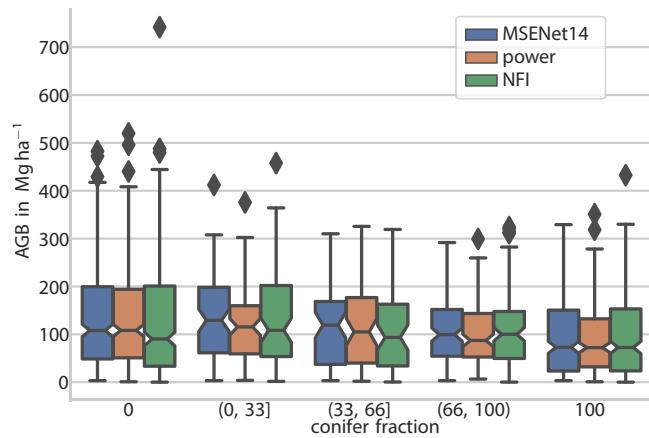


Fig. 6. Comparing value distribution of (a) AGB and (b) wood volume at different conifer fraction intervals with boxplots. The baseline (power) and MSENet14 model predictions are compared to the ground truth (NFI). Each box depicts the median in its center, the first and third quartiles form the boundaries of the box, while the whiskers extend the box by $1\frac{1}{2}$ of the inter-quartile range. Outliers are shown as dots. The notches in the box represent the confidence interval around the median.

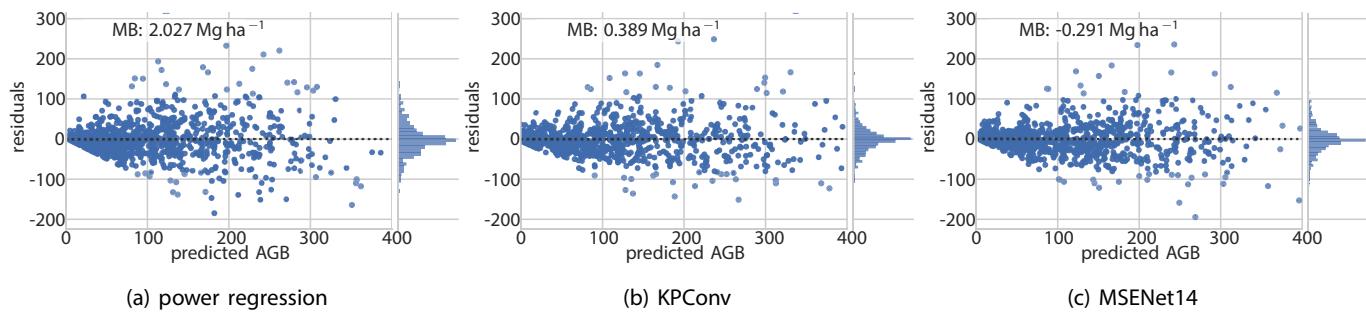


Fig. 7. Biomass residual (left side) and error distribution (right side) plots of the test performance for power regression, KPConv, and MSENNet14. The mean bias is given as well to quantify the observed bias.

suggesting that errors at plot level would partly cancel out when aggregating predictions. Comparing the distribution of errors as well as the mean bias of the assessed models, one can see that the KPConv and Minkowski models had a generally lower spread of errors than the power model. See Fig. B.12 and B.13 in the appendix for the remaining plots.

The results of the tree adding augmentation as described in Section 2.2.4 are shown in Table 3. Our analysis demonstrates that the Minkowski and PointNet models are only marginally affected, while the model performance of the baseline methods and KPConv decreases significantly. This behaviour is expected from the baseline methods since they do not capture structure and the utilized statistics are not robust to these kind of changes (see Fig. B.15 in the appendix for more concrete examples). Even though the changes in error in KPConv are of smaller magnitude than the baseline, the model is affected negatively.

Training deep learning model does not necessarily require vast amounts of data to achieve useful results. To support this, we conducted an additional experiment, where we reduced the original training data ($n = 4270$) used with the MSENNet14 until the performance decreases to the power regression baseline. The first reduction was achieved by removing all labels where LiDAR and NFI measurements are more than one year apart since these labels are considered to be noisy. Then, from these $n = 2636$ we gradually removed data: by 75% ($n = 1977$), 50% ($n = 1318$), 25% ($n = 659$), 12.5% ($n = 330$), 6.25% ($n = 165$). For each setting five repetitions were conducted, whereby each time a different random subset for training was chosen to reduce the impact of selecting a non-representative subset. The results in Fig. 8 show that

even a model trained with only 165 training samples outperforms the power regression baseline on the test data. While larger datasets often provide more robust generalization, we show that even with a comparatively small amount of data, superior performance can be achieved. It is worth noting that pretrained models, such as the ones we created and provide with this work (see Appendix B), will reduce the need for labeled data even further.

An exemplary biomass map is demonstrated in Fig. 9. The figure shows the results of the MSENNet50 model applied to a randomly selected 1 km area in Denmark without any additional processing. The predictions are convincing since heavily forested areas have a higher amount of biomass than areas with sparser tree cover.

4. Discussion

The results give clear evidence that the Minkowski deep learning approach is preferable to the methods working on point cloud statistics on individual subplot level as well as when aggregating plots. Specifically, the RMSE of the power regression is on average 8.41 Mg ha^{-1} or 588.7 per subplot larger than the best MSENNet50 model (see Table 2). When normalizing the RMSE with the mean AGB per subplot, The difference in absolute MB is 1.92 Mg ha^{-1} per 0.07 ha sample and when considering the $640,835 \text{ ha}$ forest area of Denmark in 2021 (Nord-Larsen et al., 2023), the baseline model would accumulate an additional error of 17.6 M tons of AGB. We consider MSENNet to be the method of choice for AGB estimation from ALS, because the models produced were robust

Table 3

Results of border artifact simulation with tree adding augmentations showing R^2 , RMSE, and MB on the augmented test set. RMSE and MB are given in Mg ha^{-1} for AGB and $\text{m}^3 \text{ ha}^{-1}$ for wood volume. The median (med.), the difference to the unaltered test set (diff.), and the best result (min or max) of 5 trials are shown. The best results are highlighted in bold font. To highlight significant differences between having border artifacts (tree augmentations enabled) and using the test data as is, we employ the paired Wilcoxon signed-rank test (H_0 : The paired samples in each subplot have a median difference of zero), where a sample is the absolute error for each subplot ($n = 911$). Different p -values are indicated: $p > 0.1$ by 'ns', $p \leq 0.05$ by '**', and $p \leq 0.001$ by ***'. That is, 'ns' indicates the desired property that the border artifacts do not negatively influence the model predictions.

| target | model | R^2 | | | RMSE | | | MB | | |
|---------------|-----------|--------------|---------------|--------------|--------------|-------------|--------------|-------------|-------------|-------------|
| | | med. | diff. | max | med. | diff. | min | med. | diff. | min |
| AGB | | | | | | | | | | |
| ** | linear | 0.195 | -0.566 | 0.195 | 90.80 | 41.38 | 90.80 | 39.15 | 37.26 | -39.15 |
| ** | power | 0.034 | -0.727 | 0.034 | 99.48 | 49.97 | 99.48 | 57.53 | 55.50 | -57.53 |
| ** | RF | 0.151 | -0.603 | 0.157 | 93.24 | 43.05 | 92.93 | 47.63 | 46.15 | -47.52 |
| ns | PointNet | 0.766 | -0.003 | 0.768 | 48.93 | 0.37 | 48.75 | 2.46 | 1.92 | 1.77 |
| ** | KPConv | 0.780 | -0.020 | 0.803 | 47.53 | 2.26 | 44.98 | 3.66 | 3.20 | -0.71 |
| ns | MSENNet14 | 0.823 | -0.002 | 0.829 | 42.60 | 0.22 | 41.85 | 0.31 | -0.35 | 0.12 |
| ns | MSENNet50 | 0.824 | -0.003 | 0.837 | 42.48 | 0.34 | 40.91 | 0.89 | 0.05 | 0.60 |
| volume | | | | | | | | | | |
| ** | linear | 0.243 | -0.523 | 0.243 | 166.03 | 73.74 | 166.03 | 72.34 | 67.74 | -72.34 |
| ** | power | 0.120 | -0.643 | 0.120 | 178.97 | 86.15 | 178.97 | 101.10 | 96.61 | -101.10 |
| ** | RF | 0.192 | -0.565 | 0.197 | 171.48 | 77.38 | 170.93 | 85.63 | 81.65 | -85.47 |
| ns | PointNet | 0.773 | -0.003 | 0.776 | 90.84 | 0.66 | 90.22 | 5.71 | 3.72 | 4.58 |
| * | KPConv | 0.778 | -0.021 | 0.792 | 89.81 | 4.37 | 87.00 | 7.89 | 7.51 | -1.01 |
| ns | MSENNet14 | 0.821 | -0.002 | 0.825 | 80.75 | 0.44 | 79.72 | 2.58 | 2.06 | 1.83 |
| ns | MSENNet50 | 0.822 | -0.003 | 0.832 | 80.57 | 0.59 | 78.18 | 3.57 | 3.40 | 2.65 |

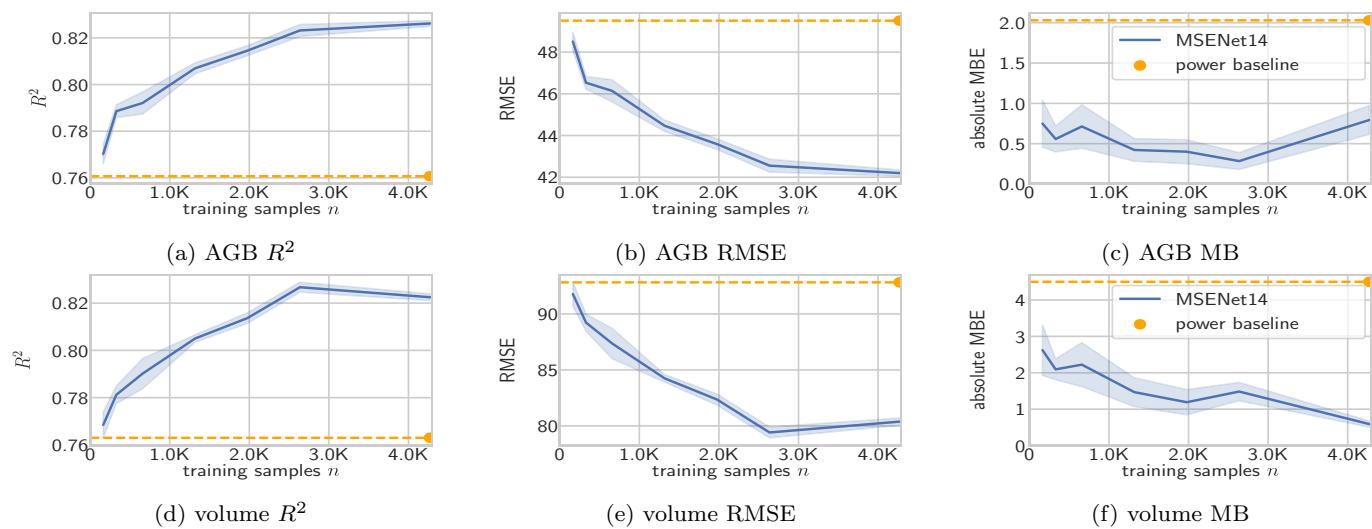


Fig. 8. Ablation of training set size to see effect on R^2 -score, RMSE, and absolute MB (columns) on AGB and wood volume (rows). Averaged runs of the MSENet14 model are shown in blue, while the power regression baseline is given for comparison in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

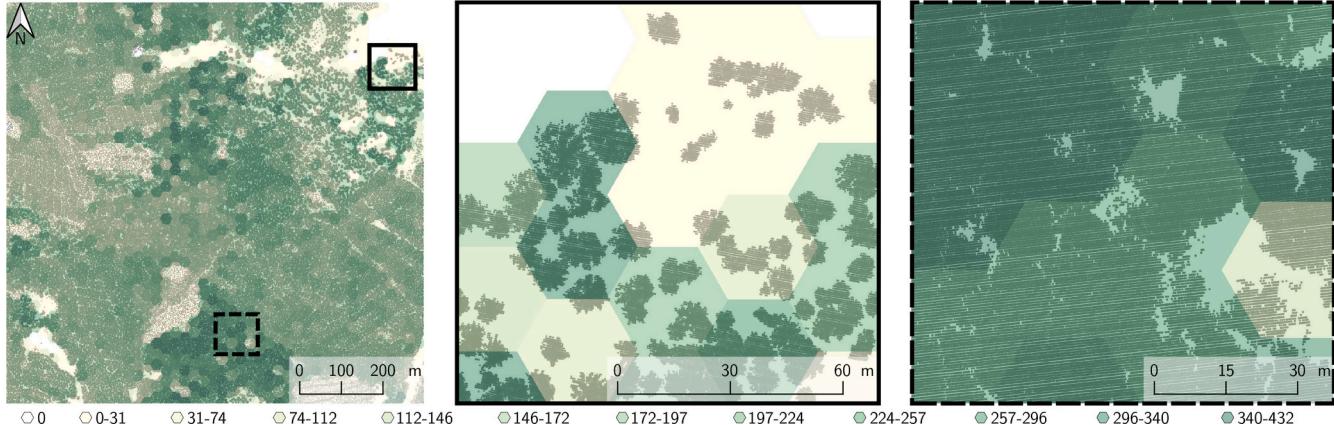


Fig. 9. Exemplary map showing predictions at different zoom levels. Points of the point cloud colored grey (only ≈ 1 above the ground) and overlaid with the estimated biomass in Mg ha^{-1} , where a deeper green indicates more biomass. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to border artifacts, achieved the high predictive performance, and do not require much computation time.

As pointed out by, for example, Xu et al. (2021) and Lu (2006), it is necessary to reduce the current uncertainty in estimates of carbon sequestration over regional scale. The classical statistical regression methods do not effectively describe the complex nonlinear relationship between forest AGB and remote sensing data (Li et al., 2020), which also becomes apparent in our results (e.g., see Fig. 6). Our proposed method goes beyond summarizing statistics by learning to utilize structural information and can be used to quantify forest biomass more accurately, which has been identified as a key component of understanding carbon budget responses to human intervention and climate (Harris et al., 2021). In addition, automatic generation of high-resolution AGB maps can also improve monitoring of AGB variations over time, resulting from land use or land use changes, and thereby increase the effectiveness of climate policies for adaptation and mitigation.

4.1. Quantitative comparison to other studies

Caution is warranted when comparing our prediction accuracy with that of other studies. Reported errors of ALS-based prediction models for

AGB range from 17 to 40 Mg ha^{-1} in the tropics (Asner et al., 2011), and this error is comparable to estimates reported in other regions (Lefsky et al., 2002) as well as previous studies with the Danish NFI (Nord-Larsen and Schumacher, 2012). Reassuringly, we achieve error rates at the lower end of this spectrum. It is important to emphasize, however, that it is not uncommon that the indicators documented in these studies measure the error on a training data set as a goodness of fit index (e.g., in the study by Lefsky et al., 1999 the R^2 for AGB drops from 80% down to 33% when the model is applied to a different data set). Similarly but not as severe, earlier studies of the Danish NFI with fewer data also reported a 0.6% increase in RMSE when using hold-out sets (Nord-Larsen and Schumacher, 2012). Conversely, the training errors of our models are similar to the test set results in Table 2 (see Table B.4 in the supplementary material), indicating that the models are not overfitted.

The error of the overall carbon stock prediction has been shown to decrease with increasing area, which is partly explained by error cancellation (Asner et al., 2010; Mascaro et al., 2011; Dalponte and Coomes, 2016; Ferraz et al., 2016; Magnussen et al., 2018). This is an important property when it comes to country-wide quantification of carbon. The residuals in Fig. 7 indeed suggest that our errors average out if several (disconnected) subplots are combined.

4.2. Applicability and limitation

The results are evaluated on a previously unseen and non-overlapping test set, meaning that the model should exhibit a similar performance on other arbitrarily chosen forest subplots from the study area (e.g., Denmark). For new study areas that are dissimilar to ours (e.g., different climate zone, tree species, or topographical features), reevaluation and potential fine-tuning of the models would be needed. However, since little preprocessing is required, this verification and adjustment step is comparatively minor.

The question naturally arises whether the method can be transferred to predict other forest characteristics, such as stem size distribution, species classification, dead wood detection, tree quality estimation, tree diameter at breast height, crown size, or stand age. The employed method would be applicable to extract such characteristics on subplot level without any modification. As shown in our experiments (see Fig. 8), even a reduced dataset gave better results than methods based on point statistics using the full dataset. In any case, it is required that the input point cloud data needs to be dense enough to give structural information about the target (e.g., if branch structure is to be extracted, singular branches need to be identifiable in the point cloud). Since our method does not require manual feature engineering, it is possible to solve a variety of tasks for different settings. For example, one would expect the relevant features to be different between temperate and tropical forests and therefore a data-driven model would learn different representations automatically. Also, even though our current model is bound to the subplot size of the Danish NFI, one could easily adjust our method to support plots of a different size.

Our method allows for dense, non-interpolated biomass mappings with unprecedented accuracy at large scale, for different types of forest (Fig. 9), and at a resolution below 30 m. The deep learning models were trained on existing NFI and ALS data, emphasizing that such models are possible today. That said, utilizing more data from different countries could lead to further improvements. Regarding runtime, our approach is also feasible to compute on a larger scale. For example, based on Table 1, the estimated prediction time for the Danish forest area of 640,835 ha amounts to 8.4 h for MSENet14 or 16.8 h for MSENet50, which could be further reduced (e.g., by using multiple GPUs). This ignores pre-processing and saving steps that need to be applied for all methods (e.g., creating point cloud cutouts), but these are reasonably fast if modern data structures, such as cloud optimized point clouds (Bell et al., 2021), are employed. Compared to the precomputation of features and the subsequent application of the baseline models, our deep-learning approach currently incurs a modest amount of additional computational overhead, but we consider this to be appropriate for achieving higher quality predictions.

The lack of interpretability of deep learning models has been perceived as a potential disadvantage compared to linear regression models based on manually selected statistical features that may offer a sense of understanding. However, combinations of a large number of such features often provide little insight owing to complex internal correlations and because they may fail to capture the relevant properties of forests (Li et al., 2020). Therefore, insights gained from these models may be limited, particularly when attempting to understand the complex relationships between vegetation and its environment. When utilizing other machine learning methods, such as random forests based on these statistical features, the interpretability is further reduced. Because our deep learning method learns to utilize structural information directly from the data, which could be abstract features about individual trees, such as differences in height, DBH, crown area, species composition, and other characteristics, the resulting model can provide valuable insights into complex relationships in the data that are not captured by standard metrics. To understand these insights, one could leverage ongoing research (Samek et al., 2019; Montavon et al., 2018; Sundararajan et al., 2017) on techniques for interpreting and visualizing deep learning models.

When plot sizes are small, border artifacts can influence the results drastically (Knapp et al., 2021). However, our results with tree-adding augmentations in Table 3 show that our PointNet and Minkowski models are not negatively affected by trees reaching into the subplot, thus allowing for the aggregation of high-resolution plots. Previously, this kind of aggregation was only possible by individual tree detection methods (Ferraz et al., 2016), but this kind of data is difficult to acquire for larger areas. The main reason why the methods based on point cloud statistics are susceptible to border artifacts is that the underlying statistics are canopy focused, even though the ground truth is stem-localized (Mascaro et al., 2011). In future research, it would be of interest to increase the model robustness even further. Specifically, robustness to non-tree objects, such as electric poles or radio towers, could be added.

5. Conclusions

Precise quantification of the spatial distribution of forest biomass and carbon content is a valuable tool to understand the processes driving af-, re-, and deforestation, and the variations in carbon cycle from the ecosystem to the regional and global scale (Zolkos et al., 2013).

This study brings forward deep learning systems for predicting above-ground biomass, wood volume, and carbon stocks in forests efficiently and directly from airborne LiDAR point clouds. Three conceptually different neural network architectures for classification (i.e., PointNet, KPConv, and Minkowski CNNs) were modified to perform regression, and were evaluated using a unique data set combining field measurements and LiDAR data.

Our adaptation of Minkowski CNNs outperformed the other deep learning approaches as well as the baseline methods. Specifically, the coefficient of determination R^2 for above-ground biomass and stored carbon exceeded 0.82 at 0.07 on a test set, which was not considered in the modelling process. The main advantage of the method proposed is the efficient implementation of 3D convolutions for sparse data, as it relies on the same operations that CNNs use for standard image processing.

The assessed deep learning models achieved high accuracies, leading to a considerably better performance in predicting AGB of forest areas, compared to the state-of-the-art approach operating on basic statistics of the point clouds. Our approach simplifies the production of AGB and carbon content maps at high resolution. It does neither require digital elevation models (e.g., DTMs or canopy height models) nor forest canopies with structural heterogeneity. Finally, our model is robust to border artifacts caused by trees reaching into a subplot, which allows for aggregation of small subplots that was not possible with the previous methods.

These encouraging results may be the first step towards utilizing airborne LiDAR data for more accurate quantification and analysis of forest carbon dynamics, which is a key component for achieving the United Nations Strategic Plan for Forests under the Paris agreement and the Sustainable Development Goals (SDGs) (United Nations Department of Economic and Social Affairs, 2017, 2021) in accordance with the IPCC 2018 (Masson-Delmotte et al., 2019).

Funding

This work was supported by the research grant DeReEco (grant number 34306) from Villum Foundations, the Independent Research Fund Denmark through the grant *Monitoring Changes in Big Satellite Data via Massively-Parallel Artificial Intelligence* (grant number 9131-00110B), a Villum Experiment grant by the Velux Foundations (MapCland project, project number 00028314), the DeepCrop project (UCPH Strategic plan 2023 Data+ Pool), the PerformLCA project (UCPH Strategic plan 2023 Data+ Pool), the EU LIFE Forests Fit for Future project (grant number: LIFE19ENV/DK/000013), and the Pioneer Centre for AI (DNRF grant number P1).

CRediT authorship contribution statement

Stefan Oehmcke: Conceptualization, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Lei Li:** Software, Visualization, Writing – review & editing. **Katerina Trepekli:** Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. **Jaime C. Revenga:** Writing – review & editing. **Thomas Nord-Larsen:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **Fabian Gieseke:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. **Christian Igel:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing interests. Furthermore, the authors confirm that the material is not under consideration for publication elsewhere. A short poster abstract with preliminary results has been published (Oehmcke et al., 2022), which does not contain most of the content of the submitted manuscript.

Appendix A. Additional results

Table B.4 and Table B.5 presents the results on the training and validation set, respectively. Note that the models using extracted features (linear regression, power regression, and RF) were trained on the union of training and validation set, while our point cloud regression models were not. Still, the more recent models directly working on the point cloud data consistently produced better results. Fig. B.12 and B.13 illustrate residual plots and error distribution of biomass and wood volume, respectively, which completes Fig. 7. Fig. B.11 shows AGB and wood volume distributions for the different data splits. We also give a visualization of the plots and subplots in Fig. B.16 (as described in Section 2.1).

In Fig. B.14 we show more subplots with particularly low or high values for above-ground biomass. For example, the first two subplots have low AGB with 1.79 and 1.89 Mg ha⁻¹, which is reflected by the sparse vegetation in the point cloud. This is difficult to capture with commonly used statistics derived from point cloud data, and therefore models based on these features (e.g., RF, linear, and power model) overestimated. Interestingly, PointNet also overestimate values due to the employed global aggregation without intermediate downsampling steps (similar to global feature extraction). The third example in Fig. B.14 exhibits a high AGB value with 417.60 Mg ha⁻¹ and the point cloud shows a densely populated and natural broadleaf forest. Again, this is difficult to estimate since the derived statistics do not offer information about individual biomass contributions and thus most methods underestimate them. The fourth subplots biomass is 245.89 Mg ha⁻¹ from a conifer forest and the predictions align well with the target.

Four examples of subplots augmented with additional trees are given in Fig. B.15. The trees are always added outside of the subplot, so no change in AGB or wood volume is expected. The first subplot in Fig. B.15a contains large trees and the augmentation adds small trees, which results in the baseline methods starting to underpredict. In the second subplot in Fig. B.15b small trees are augmented with large trees and this results in overprediction of the baseline methods. In the third example in Fig. B.15c, the baseline method applied to the original point cloud is already overpredicting. Plots with empty space are not well captured by the used statistics and the results further decrease when adding border trees. In Fig. B.15d, an originally empty plot is drastically overpredicted by the baseline when trees are visible on the border. Overall, only minor changes are observed for the deep learning methods. The statistics that the power is based on are mostly affected when larger trees are added but also show a decrease in predictive power when smaller changes occur. Further, subplot predictions from power that show little biomass suffer most from border artifact errors.

If the ground is not visible (e.g., due to dense forest stands), extracting the DTM is difficult, which makes the extraction of height-related statistics error-prone. In contrast, our method is still applicable. To demonstrate this, we removed the ground points from the test data and re-ran our models again without retraining them. Some loss in predictive power is to be expected simply because missing ground points were not part of the training set. The results shown in Table B.6 confirm this.

Appendix B. Reproducibility

We uploaded the deep learning repository, trained weights, data, and evaluation scripts to an anonymous Git repository (<https://github.com/SteOe/DPCR-AGB>). We removed geo-coordinates and information not directly required for the regression task. Upon acceptance, the content of this repository will be made publicly available.

Table B.4

Comparison of methods showing the R^2 score, RMSE, and MAPE on the training set. The models with best validation set performances were chosen. The R^2 and MAPE for biomass are the same as for carbon stocks. Note that we excluded 0 biomass measurements from MAPE to avoid numerical issues. Best results are highlighted.

| target | model | R^2 | | RMSE | | MAPE | |
|--------|--------|-------|-------|-------|-------|--------|--------|
| | | med. | max | med. | min | med. | min |
| AGB | linear | 0.706 | 0.706 | 56.09 | 56.09 | 802.65 | 802.65 |
| | power | 0.700 | 0.700 | 56.63 | 56.63 | 690.59 | 690.59 |

(continued on next page)

Table B.4 (continued)

| target | model | R^2 | | RMSE | | MAPE | |
|--------|----------|--------------|--------------|--------------|--------------|---------------|---------------|
| | | med. | max | med. | min | med. | min |
| volume | RF | 0.723 | 0.723 | 54.45 | 54.42 | 797.90 | 779.69 |
| | PointNet | 0.714 | 0.726 | 55.34 | 54.19 | 516.35 | 287.00 |
| | KPConv | 0.800 | 0.828 | 46.27 | 42.89 | 173.58 | 135.66 |
| | MSENet14 | 0.804 | 0.841 | 45.85 | 41.26 | 240.54 | 148.11 |
| | MSENet50 | 0.796 | 0.804 | 46.68 | 45.85 | 249.11 | 151.39 |
| | linear | 0.708 | 0.708 | 104.92 | 104.92 | 185.27 | 185.27 |
| | power | 0.703 | 0.703 | 105.83 | 105.83 | 197.08 | 197.08 |
| | RF | 0.728 | 0.728 | 101.31 | 101.26 | 204.57 | 202.69 |
| | PointNet | 0.721 | 0.731 | 102.52 | 100.67 | 144.41 | 119.17 |
| | KPConv | 0.795 | 0.824 | 87.95 | 81.47 | 82.28 | 68.53 |
| volume | MSENet14 | 0.800 | 0.838 | 86.85 | 78.15 | 94.70 | 67.94 |
| | MSENet50 | 0.795 | 0.803 | 87.97 | 86.19 | 94.24 | 71.30 |

Table B.5

Comparison of methods showing the R^2 score, RMSE, and MAPE on the validation set. The R^2 and MAPE for biomass are the same as for carbon stocks. Note that we excluded 0 biomass measurements from MAPE to avoid numerical issues. Best results are highlighted. Note that the deep learning models did not train on this data directly but the feature extraction methods did.

| target | model | R^2 | | RMSE | | MAPE | |
|--------|----------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | med. | max | med. | min | med. | min |
| AGB | linear | 0.739 | 0.739 | 54.02 | 54.02 | 283.79 | 283.79 |
| | power | 0.739 | 0.739 | 53.99 | 53.99 | 384.63 | 384.63 |
| | RF | 0.758 | 0.758 | 52.08 | 52.01 | 445.04 | 429.35 |
| | PointNet | 0.711 | 0.713 | 56.93 | 56.69 | 163.98 | 134.58 |
| | KPConv | 0.790 | 0.802 | 48.46 | 47.09 | 74.90 | 62.09 |
| | MSENet14 | 0.808 | 0.811 | 46.36 | 45.99 | 165.69 | 66.96 |
| | MSENet50 | 0.805 | 0.810 | 46.72 | 46.16 | 183.55 | 68.73 |
| volume | linear | 0.766 | 0.766 | 94.65 | 94.65 | 98.58 | 98.58 |
| | power | 0.766 | 0.766 | 94.61 | 94.61 | 118.03 | 118.03 |
| | RF | 0.784 | 0.785 | 90.93 | 90.81 | 110.56 | 109.67 |
| | PointNet | 0.746 | 0.750 | 98.79 | 97.84 | 160.81 | 149.83 |
| | KPConv | 0.808 | 0.817 | 85.76 | 83.72 | 65.94 | 58.08 |
| | MSENet14 | 0.821 | 0.826 | 82.75 | 81.70 | 68.49 | 61.62 |
| | MSENet50 | 0.824 | 0.826 | 82.08 | 81.68 | 64.84 | 57.12 |

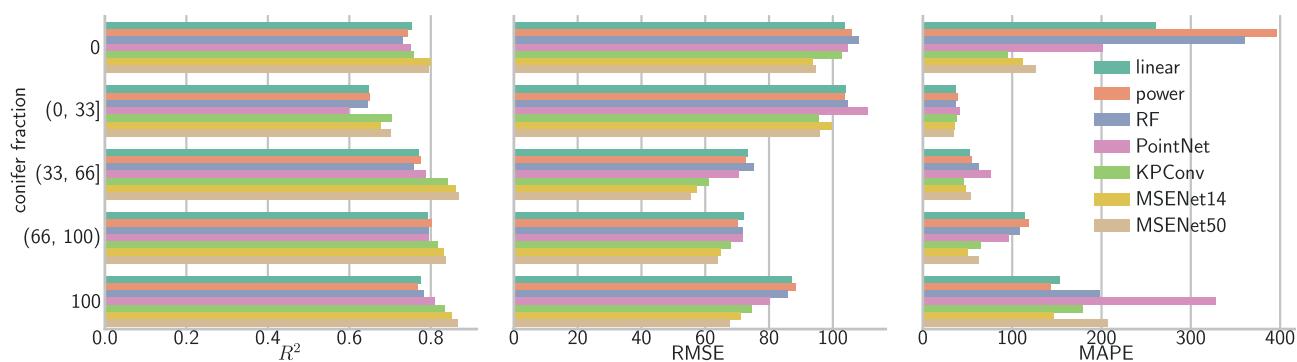


Fig. B.10. Errors for different fractions of conifer and broadleaf trees for wood volume on the test set. The columns show R^2 (higher is better), RMSE (lower is better), and MAPE (lower is better).

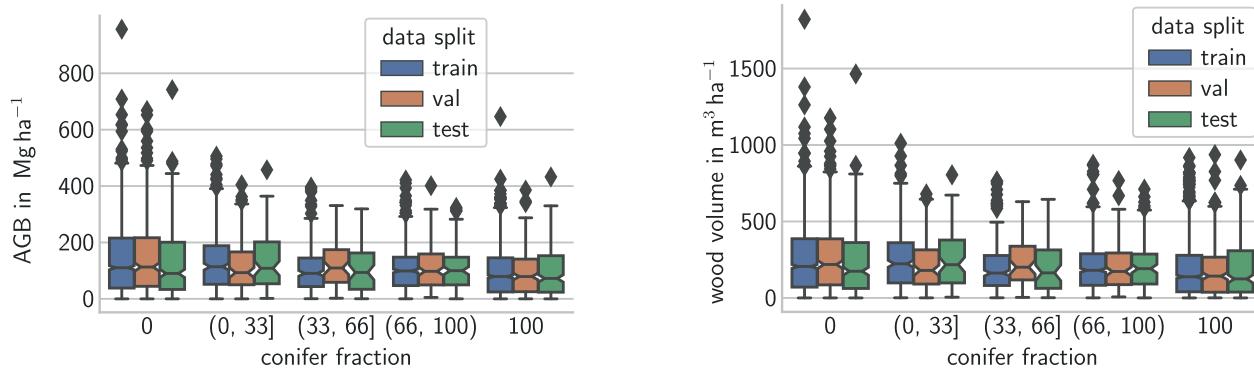


Fig. B.11. Boxplots of value distributions over considered conifer fractions. Left side shows AGB and right side shows wood volume.

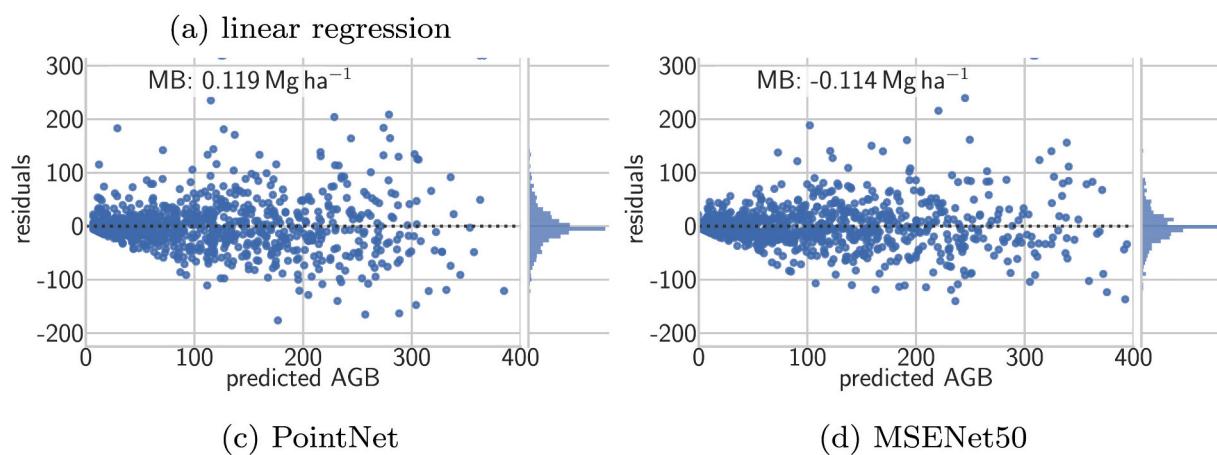
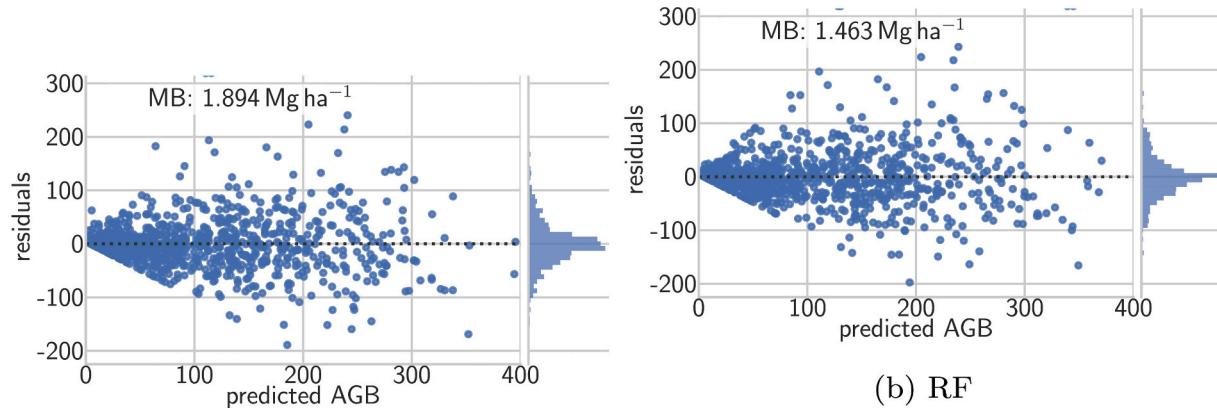


Fig. B.12. Biomass residual (left side) and error distribution (right side) plots of the test performance.

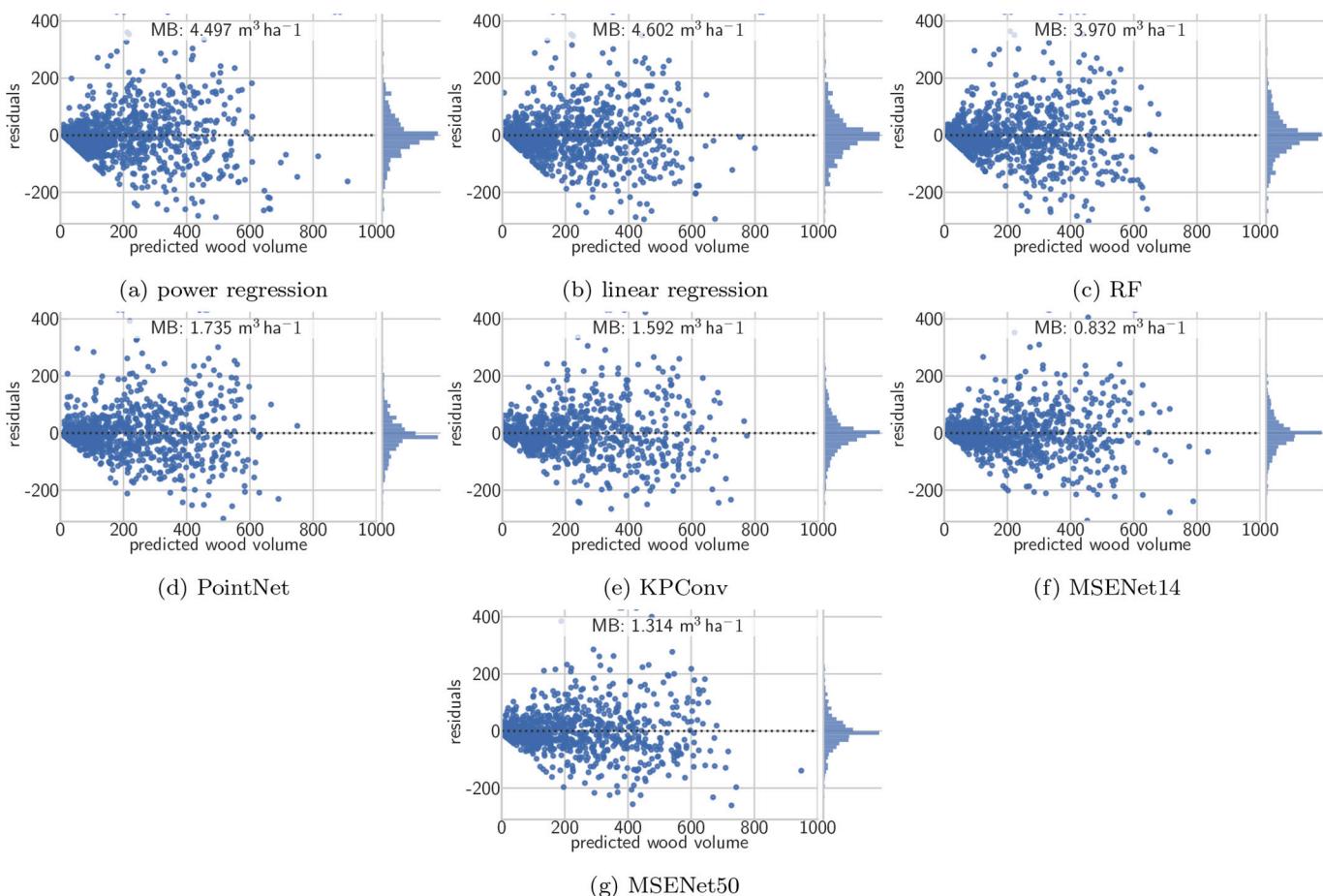
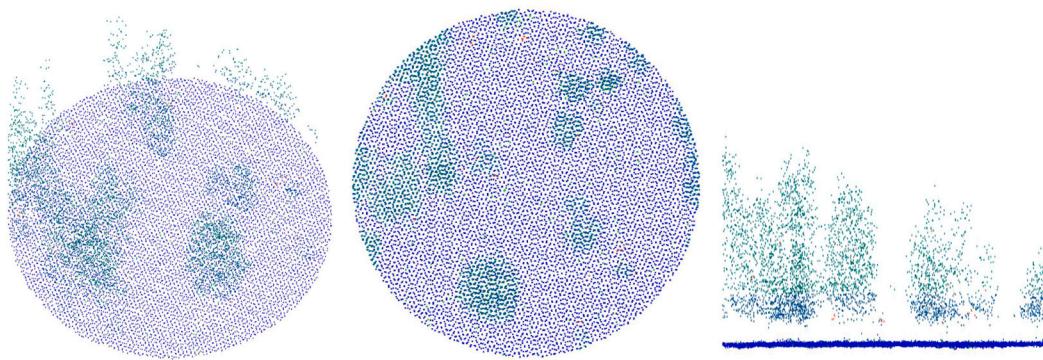
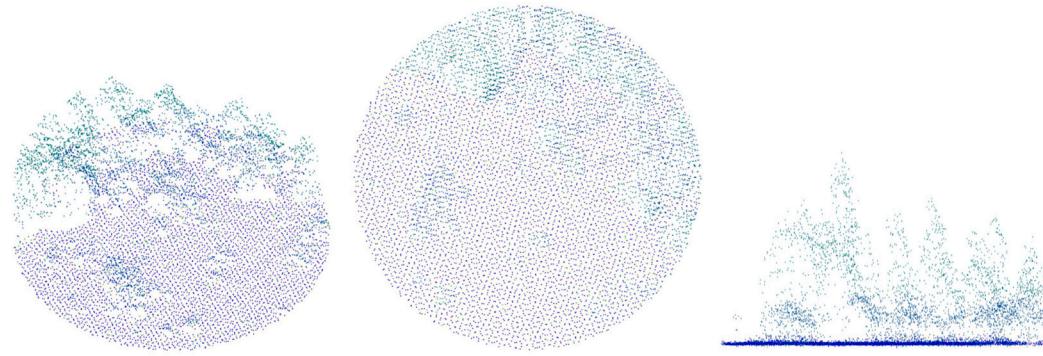


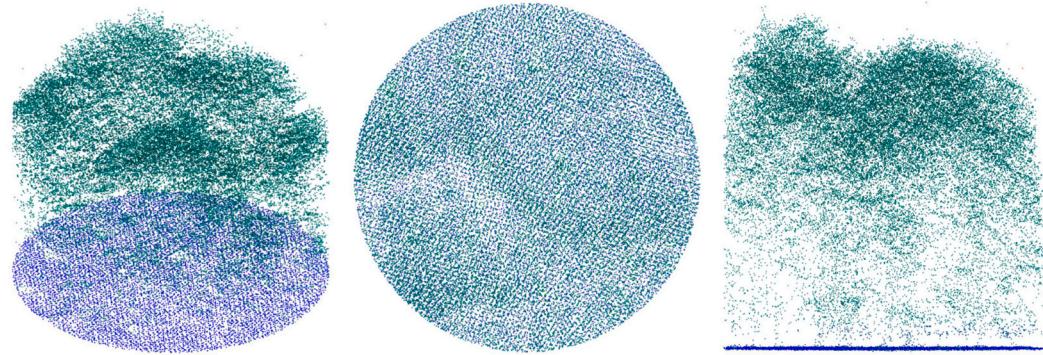
Fig. B.13. Wood volume residual (left side) and error distribution (right side) plots of the test performance.



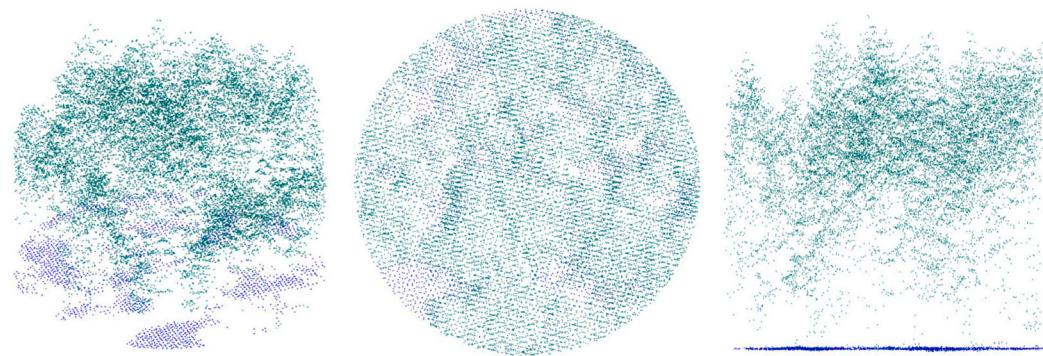
(a) NFI: 1.79, linear: 10.67, RF: 11.73, KPConv: 1.02, PointNet: 8.25, power: 12.04, MSENNet14: 2.42, MSENNet50: 3.69



(b) NFI: 0.86, linear: 7.67, RF: 6.01, KPConv: 0.95, PointNet: 6.94, power: 10.67, MSENNet14: 0.89, MSENNet50: 2.07



(c) NFI: 417.6, linear: 275.4, RF: 266.2, KPConv: 407.5, PointNet: 297.0, power: 271.7, MSENNet14: 418.1, MSENNet50: 431.3



(d) NFI: 245.9, linear: 214.3, RF: 210.4, KPConv: 238.8, PointNet: 232.2, power: 218.6, MSENNet14: 229.0, MSENNet50: 249.7



Fig. B.14. More examples of subplots in each row with three perspectives: isometric front, top, and side view. Each subplot caption contains the measured (named NFI) and estimated AGB in Mg ha⁻¹.

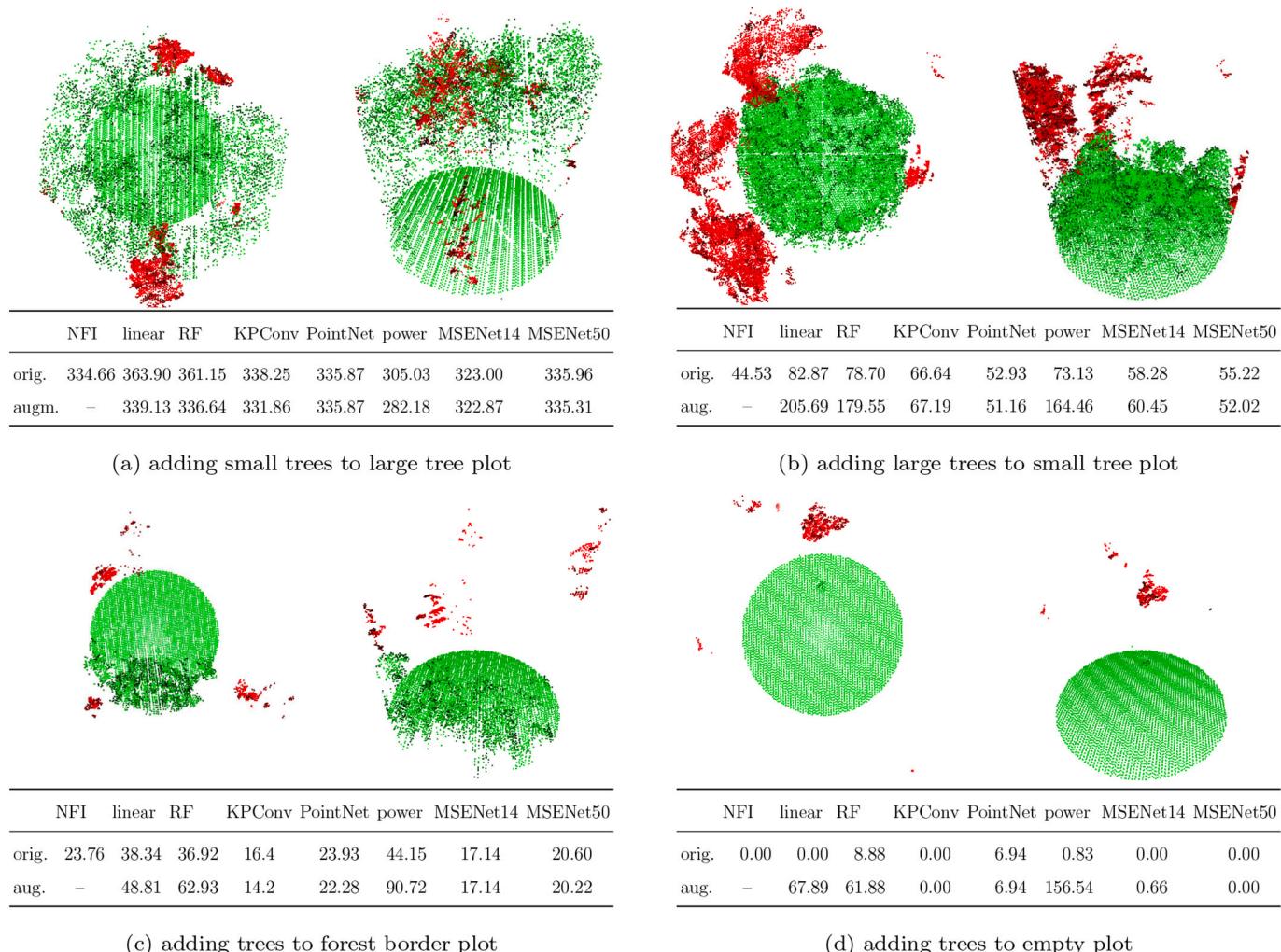


Fig. B.15. Subplot examples illustrating the tree augmentations. Green points are coming from the original point cloud and red points are augmented tree points. Two views are shown: a top view and an angled side view. The tables report the measured (named NFI) and estimated AGB in Mg ha^{-1} with and without tree augmentations, which we refer to as orig. and aug., respectively. AGB is not expected to change by the augmentation, as the tree stems of the added trees are outside the subplot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

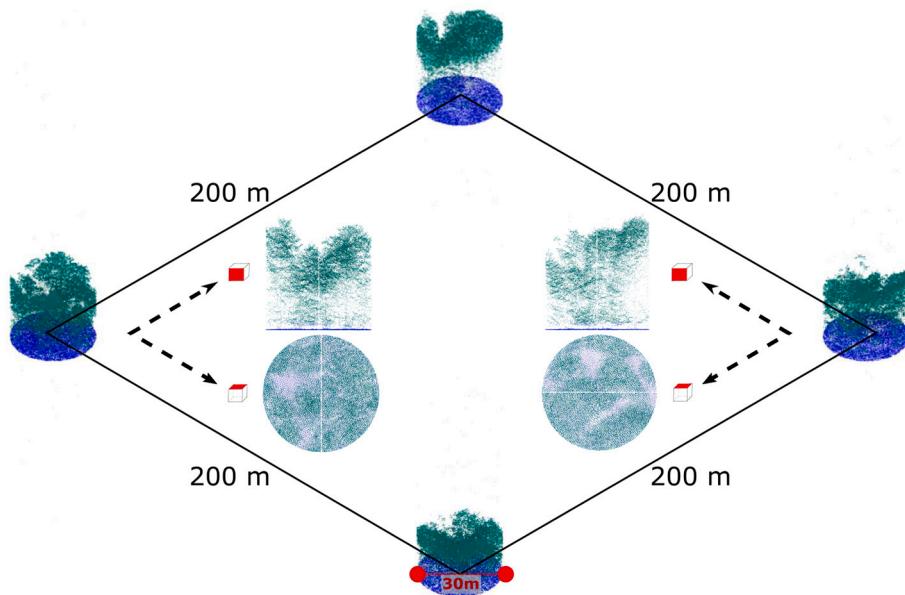


Fig. B.16. One plot sample from a $2 \times 2 \text{ km}$ area as described in Section 2.1. Each plot usually contains four subplots in a $200 \times 200 \text{ m}$ square. All LiDAR subplots have a diameter of 30 m and their own biomass as well as wood volume measurements.

Table B.6

Comparison of methods when ground points are removed. The given metrics are the coefficient of determination (R^2) score, root mean square error (RMSE), mean absolute percentage error (MAPE), and mean bias (MB) on the test set. Note that we excluded 0 biomass measurements from MAPE to avoid numerical issues. RMSE and MB are in Mg ha⁻¹ for AGB and m³ ha⁻¹ for wood volume. The median (med.) and best result (min or max) of 5 trials are shown, the best results are highlighted in bold font. The median for the MB is based on the absolute values, which is more meaningful when combining multiple trials. The min values are also based on the absolute values, however the signs are indicated in the table.

| target | method | no ground | R^2 | | RMSE | | MAPE | | MB | |
|--------|-----------|-----------|-------|-------|--------|--------|---------|---------|------|-------|
| | | | med. | max | med. | min | med. | min | med. | min |
| | | | | | | | | | | |
| AGB | MSENNet50 | no | 0.825 | 0.835 | 42.30 | 41.08 | 443.89 | 174.25 | 0.90 | -0.11 |
| | MSENNet50 | yes | 0.801 | 0.811 | 45.10 | 43.98 | 258.68 | 171.81 | 0.40 | -0.04 |
| | power | no | 0.761 | 0.761 | 49.51 | 49.51 | 365.61 | 365.61 | 2.03 | 2.03 |
| | power | yes | 0.686 | 0.686 | 56.74 | 56.74 | 1552.48 | 1552.48 | 2.81 | 2.81 |
| | MSENNet50 | no | 0.824 | 0.831 | 80.13 | 78.34 | 127.64 | 72.38 | 0.16 | -0.04 |
| | MSENNet50 | yes | 0.797 | 0.798 | 86.03 | 85.85 | 132.57 | 111.81 | 1.84 | 0.18 |
| | power | no | 0.763 | 0.763 | 92.82 | 92.82 | 223.65 | 223.65 | 4.50 | 4.50 |
| | power | yes | 0.688 | 0.688 | 106.50 | 106.50 | 394.09 | 394.09 | 6.11 | 6.11 |

Appendix C. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2023.113968>.

References

- Asner, G.P., Mascaro, J., 2014. Mapping tropical forest carbon: calibrating plot estimates to a simple lidar metric. *Remote Sens. Environ.* 140, 614–624.
- Asner, G.P., Powell, G.V., Mascaro, J., Knapp, D.E., Clark, J.K., Jacobson, J., Kennedy-Bowdoin, T., Balaji, A., Paez-Acosta, G., Victoria, E., et al., 2010. High-resolution forest carbon stocks and emissions in the Amazon. *Proc. Natl. Acad. Sci.* 107, 16738–16742.
- Asner, G.P., Hughes, R.F., Mascaro, J., Uowolo, A.L., Knapp, D.E., Jacobson, J., Kennedy-Bowdoin, T., Clark, J.K., 2011. High-resolution carbon mapping on the million-hectare island of Hawaii. *Front. Ecol. Environ.* 9, 434–439.
- Bell, A., Butler, H., Manning, C., 2021. Cloud Optimized Point Cloud (COPC). URL <https://github.com/copcio/copcio.github.io>.
- Bouvier, M., Durrieu, S., Fournier, R.A., Renaud, J.P., 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne lidar data. *Remote Sens. Environ.* 156, 322–334.
- Brandt, M., Rasmussen, K., Hieraux, P., Hermann, S., Tucker, C.J., Tong, X., Tian, F., Mertz, O., Kergoat, L., Mbwo, C., et al., 2018. Reduction of tree cover in west African woodlands and promotion in semi-arid farmlands. *Nat. Geosci.* 11, 328–333.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Calders, K., Origo, N., Burt, A., Disney, M., Nightingale, J., Raumonen, P., Åkerblom, M., Malhi, Y., Lewis, P., 2018. Realistic forest stand reconstruction from terrestrial lidar for radiative transfer modelling. *Remote Sens.* 10, 933.
- Chang, A.X., Funkhouser, T.A., Guibas, L.J., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F., 2015. Shapenet: an information-rich 3D model repository. *CoRR abs/1512.03012*. arXiv:1512.03012.
- Chaton, T., Nicolas, C., Horache, S., Landrieu, L., 2020. Torch-Points3D: a modular multi-task framework for reproducible deep learning on 3D point clouds. In: *3D Vision (3DV)*. IEEE, pp. 1–10.
- Choy, C., Gwak, J., Savarese, S., 2019. 4D spatio-temporal convnets: Minkowski convolutional neural networks. In: *IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 3075–3084.
- Cook-Patton, S.C., Leavitt, S.M., Gibbs, D., Harris, N.L., Lister, K., Anderson-Teixeira, K.J., Briggs, R.D., Chazdon, R.L., Crowther, T.W., Ellis, P.W., et al., 2020. Mapping carbon accumulation potential from global natural forest regrowth. *Nature* 585, 545–550.
- Coyle, D.B., Stysley, P.R., Chirag, F.L., Frese, E., Poulios, D., 2019. The global ecosystem dynamics investigation (GEDI) LiDAR laser transmitter. In: *Infrared Remote Sensing and Instrumentation XXVII*. Int. Society for Optics and Photonics, pp. 112–125.
- Dalla Corte, A.P., Souza, D.V., Rex, F.E., Sanquette, C.R., Mohan, M., Silva, C.A., Zambrano, A.M.A., Prata, G., de Almeida, D.R.A., Trautenmüller, J.W., et al., 2020. Forest inventory with high-density uav-lidar: machine learning approaches for predicting individual tree attributes. *Comput. Electron. Agric.* 179, 105815.
- Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods Ecol. Evol.* 7, 1236–1245.
- d’Oliveira, M.V., Reutebuch, S.E., McGaughey, R.J., Andersen, H.E., 2012. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in amitary state forest, acre state, western brazilian amazon. *Remote Sens. Environ.* 124, 479–491.
- Erb, K.H., Kastner, T., Plutzar, C., Bais, A.L.S., Carvalhais, N., Fetzel, T., Gingrich, S., Haberl, H., Lauk, C., Niedertscheider, M., Pongratz, J., Thurner, M., Luyssaert, S., 2018. Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* 553, 73–76.
- ESRI, 2023. Map background. <https://www.esri.com>.
- FAO, UNEP, 2020. *The State of the World’s Forests 2020*. Forests, biodiversity and people.
- Fassnacht, F., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., Koch, B., 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sens. Environ.* 154, 102–114.
- Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016. Lidar detection of individual tree size in tropical forests. *Remote Sens. Environ.* 183, 318–333.
- Firintepé, A., Vey, C., Asteriadis, S., Pagani, A., Stricker, D., 2021. From IR images to point clouds to pose: point cloud-based AR glasses pose estimation. *J. Imaging* 7.
- Ge, L., Ren, Z., Yuan, J., 2018. Point-to-point regression PointNet for 3D hand pose estimation. *Eur. Conf. Comput. Vis.* 475–491.
- Girshick, R., 2015. Fast R-CNN. In: *International Conference on Computer Vision (ICCV)*, pp. 1440–1448.
- Goetz, S., Dubayah, R., 2011. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Manage.* 2, 231–244.
- Goodwin, N.R., Coops, N.C., Culvenor, D.S., 2006. Assessment of forest structure with airborne lidar and the effects of platform altitude. *Remote Sens. Environ.* 103, 140–152.
- Graham, B., 2015. Sparse 3D convolutional neural networks. In: *BMVC*, pp. 150.1–150.9.
- Graham, B., Engelcke, M., Van Der Maaten, L., 2018. 3D semantic segmentation with submanifold sparse convolutional networks. In: *IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 9224–9232.
- Griscom, B.W., Adams, J., Ellis, P.W., Houghton, R.A., Lomax, G., Miteva, D.A., Schlesinger, W.H., Shoch, D., Siikamäki, J.V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campani, J., Conant, R.T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M.R., Herrero, M., Kiesecker, J., Landis, E., Laestadius, L., Leavitt, S.M., Minnemeyer, S., Polasky, S., Potapov, P., Putz, F.E., Sanderman, J., Silvius, M., Wollenberg, E., Fargione, J., 2017. Natural climate solutions. *Proc. Natl. Acad. Sci.* 114, 11645–11650.
- Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2020. Deep learning for 3D point clouds: a survey. In: *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 43, pp. 4338–4364.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. SEMANTIC3D.NET: a new large-scale point cloud classification benchmark. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 91–98.
- Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., Dubayah, R., 2019. The gedi simulator: a large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth Space Sci.* 6, 294–310.
- Harris, N.L., Brown, S., Hagen, S.C., Saatchi, S.S., Petrova, S., Salas, W., Hansen, M.C., Potapov, P.V., Lotsch, A., 2012. Baseline map of carbon emissions from deforestation in tropical regions. *Science* 336, 1573–1576.
- Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A., et al., 2021. Global maps of twenty-first century forest carbon fluxes. *Nat. Clim. Chang.* 11, 234–240.
- Hendrycks, D., Gimpel, K., 2016. Gaussian error linear units (gelus) *arXiv preprint arXiv: 1606.08415*.
- Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: *IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 7132–7141.

- Igel, C., Oehmcke, S., 2023. Remember to correct the bias when using deep learning for regression! *Künstl. Intell.* <https://doi.org/10.1007/s13218-023-00801-0>.
- Johannsen, V.K., 1999. A Growth Model for Oak in Denmark. Ph.D. thesis. Royal Veterinary and Agricultural University, Copenhagen, Denmark.
- Kish, L., 1998. Space/time variations and rolling samples. *J. Off. Stat.* 14, 31.
- Knapp, N., Fischer, R., Huth, A., 2018. Linking lidar and forest modeling to assess biomass estimation across scales and disturbance states. *Remote Sens. Environ.* 205, 199–209.
- Knapp, N., Fischer, R., Cazcarra-Bes, V., Huth, A., 2020. Structure metrics to generalize biomass estimation from lidar across forest types from different continents. *Remote Sens. Environ.* 237, 111597.
- Knapp, N., Huth, A., Fischer, R., 2021. Tree crowns cause border effects in area-based biomass estimations from remote sensing. *Remote Sens.* 13 <https://doi.org/10.3390/rs13081592>. URL: <https://www.mdpi.com/2072-4292/13/8/1592>.
- Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* 14, 778–782.
- Lang, N., Schindler, K., Wegner, J.D., 2021. High carbon stock mapping at large scale with optical satellite imagery and spaceborne LiDAR. *CoRR arXiv:2107.07431*.
- Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., Wegner, J.D., 2022. Global canopy height regression and uncertainty estimation from GEDI lidar waveforms with deep ensembles. *Remote Sens. Environ.* 268, 112760.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., Shugart, H.H., 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sens. Environ.* 67, 83–98.
- Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., Gower, S.T., 2002. Lidar remote sensing of above-ground biomass in three biomes. *Glob. Ecol. Biogeogr.* 11, 393–399.
- Lewis, S.L., Wheeler, C.E., Mitchard, E.T., Koch, A., 2019. Regenerate natural forests to store carbon. *Nature* 568, 25–28.
- Li, Y., Li, M., Li, C., Liu, Z., 2020. Forest aboveground biomass estimation using landsat 8 and sentinel-1a data with machine learning algorithms. *Sci. Rep.* 10, 1–12.
- Li, S., Brandt, M., Fensholt, R., Karirya, A., Igel, C., Gieseke, F., Nord-Larsen, T., Oehmcke, S., Carlsen, A.H., Junnila, S., Tong, X., d'Aspremont, A., Ciais, P., 2023. Deep Learning Enables Image-Based Tree Counting, Crown Segmentation and Height Prediction at National Scale. *PNAS Nexus*.
- Liu, K., Shen, X., Cao, L., Wang, G., Cao, F., 2018. Estimating forest structural attributes using uav-lidar data in ginkgo plantations. *ISPRS J. Photogramm. Remote Sens.* 146, 465–482.
- Liu, S., Brandt, M., Nord-Larsen, T., Chave, J., Reiner, F., Lang, N., Tong, X., Ciais, P., Ristian Igel, C., Pascual, A., Guerra-Hernandez, J., Li, S., Mugabowindekwe, M., Saatchi, S., Yue, Y., Chen, Z., Fensholt, R., 2023. The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. *Sci. Adv.* 9.
- Loshchilov, I., Hutter, F., 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. In: International Conference on Learning Representation (ICLR). OpenReview.net.
- Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* 27, 1297–1328.
- Magnussen, S., Nord-Larsen, T., Riis-Nielsen, T., 2018. Lidar supported estimators of wood volume and aboveground biomass from the Danish national forest inventory (2012–2016). *Remote Sens. Environ.* 211, 146–153.
- Mascaro, J., Dettò, M., Asner, G.P., Muller-Landau, H.C., 2011. Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sens. Environ.* 115, 3770–3774.
- Masson-Delmotte, V., Zhai, P., Pörtner, H.O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., Waterfield, T., 2019. IPCC, 2018: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Cambridge University Press.
- McDonald, T.L., 2003. Review of environmental monitoring methods: survey designs. *Environ. Monit. Assess.* 85, 277–292.
- Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., Gourlet-Fleury, S., 2015. Decrease of 1-band Sar backscatter with biomass of dense forests. *Remote Sens. Environ.* 159, 307–317.
- Mitchard, E.T.A., Saatchi, S.S., White, L.J.T., Abernethy, K.A., Jeffery, K.J., Lewis, S.L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H., Meir, P., 2012. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences* 9, 179–191.
- Montavon, G., Samek, W., Müller, K.R., 2018. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* 73, 1–15.
- Næsset, E., 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sens. Environ.* 113, 148–159.
- Näslund, M., 1936. Skogsförskönsanstalten gallringsförsök i tallskog. *Meddelanden Statens Skogsförskönsanstalt* 29, 1–169.
- Nord-Larsen, T., Johannsen, V.K., 2016. Danish National Forest Inventory: Design and Calculations. IGN Report. Department of Geosciences and Natural Resource Management, University of Copenhagen.
- Nord-Larsen, T., Schumacher, J., 2012. Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. *Remote Sens. Environ.* 119, 148–157. <https://doi.org/10.1016/j.rse.2011.12.022>.
- Nord-Larsen, T., Meilby, H., Skovsgaard, J.P., 2017a. Simultaneous estimation of biomass models for 13 tree species: effects of compatible additivity requirements. *Can. J. For. Res.* 47, 765–776.
- Nord-Larsen, T., Riis-Nielsen, T., Ottosen, M.B., 2017b. Forest Resource Map of Denmark: Mapping of Danish Forest Resource Using ALS from 2014–2015. IGN Report. Department of Geosciences and Natural Resource Management, University of Copenhagen.
- Nord-Larsen, T., Johannsen, V., Riis-Nielsen, T., Thomsen, I., Bentsen, N., Jørgensen, B., 2023. Skovstatistik 2021. Institut for Geovidenskab og Naturforvaltning, Københavns Universitet. URL: https://static-curis.ku.dk/portal/files/338858104/Rapport_Skovstatistik_2021_web.pdf.
- Oehmcke, S., Li, L., Revenga, J., Nord-Larsen, T., Trepeki, K., Gieseke, F., Igel, C., 2022. Deep learning based 3D point cloud regression for estimating forest biomass. In: International Conference on Advances in Geographic Information Systems (SIGSPATIAL). ACM.
- Olsen, A.R., Sedransk, J., Edwards, D., Gotway, C.A., Liggett, W., Rathbun, S., Reckhow, K.H., Yyoung, L.J., 1999. Statistical issues for monitoring ecological and natural resources in the United States. *Environ. Monit. Assess.* 54, 1–45.
- Pan, Y., Birdsey, R.A., Phillips, O.L., Jackson, R.B., 2013. The structure, distribution, and biomass of the world's forests. *Annu. Rev. Ecol. Evol. Syst.* 44, 593–622.
- Pan, L., Liu, L., Condon, A.G., Estavillo, G.M., Coe, R.A., Bull, G., Stone, E.A., Petersson, L., Rolland, V., 2022. Biomass prediction with 3d point clouds from LiDAR. In: Winter Conference on Applications of Computer Vision (WACV), pp. 1330–1340.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: an imperative style, high-performance deep learning library. *NeurIPS* 32, 8026–8037.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: deep learning on point sets for 3D classification and segmentation. In: IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), pp. 652–660.
- Qi, C.R., Litany, O., He, K., Guibas, L.J., 2019. Deep hough voting for 3d object detection in point clouds. In: International Conference on Computer Vision (ICCV). IEEE, pp. 9276–9285.
- Requena Suarez, D., Rozendaal, D.M., De Sy, V., Phillips, O.L., Alvarez-Dávila, E., Anderson-Teixeira, K., Araujo-Murakami, A., Arroyo, L., Baker, T.R., Bongers, F., et al., 2019. Estimating aboveground net biomass change for tropical and subtropical forests: refinement of IPCC default rates using forest plot data. *Glob. Chang. Biol.* 25, 3609–3624.
- Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.R., 2019. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, vol. 11700. Springer Nature.
- Shukla, P.R., Skeg, J., Buendia, E.C., Masson-Delmotte, V., Pörtner, H.O., Roberts, D., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Pereira, J.P., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J. (Eds.), 2019. IPCC, 2019: Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. Cambridge University Press.
- Sinha, S., Jeganathan, C., Sharma, L.K., Nathawat, M.S., 2015. A review of radar remote sensing for biomass estimation. *Int. J. Environ. Sci. Technol.* 12, 1779–1792.
- Sloboda, B., Gaffrey, D., Matsumura, N., 1993. Regionale und lokale Systeme von Höhenkurven für gleichalte Waldbestände. *Allgemeine Forst Jagdzeitung* 164, 225–228.
- Stark, S.C., Leitold, V., Wu, J.L., Hunter, M.O., de Castillo, C.V., Costa, F.R., McMahon, S.M., Parker, G.G., Shimabukuro, M.T., Lefsky, M.A., et al., 2012. Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. *Ecol. Lett.* 15, 1406–1414.
- Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. In: International Conference on Machine Learning, PMLR, pp. 3319–3328.
- Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. KPConv: flexible and deformable convolution for point clouds. In: International Conference on Computer Vision (ICCV), pp. 6411–6420.
- Treuhhaft, R., Gonçalves, F., dos Santos, J.R., Keller, M., Palace, M., Madsen, S.N., Sullivan, F., Graça, P.M., 2014. Tropical forest biomass estimation at x-band from the spaceborne tandem-x interferometer. *IEEE Geosci. Remote Sens. Lett.* 12, 239–243.
- Tubiello, F.N., Conchedda, G., Wanner, N., Federici, S., Rossi, S., Grassi, G., 2021. Carbon emissions and removals from forests: new estimates, 1990–2020. *Earth Syst. Sci. Data* 13, 1681–1691.
- Tucker, C., Brandt, M., Hiernaux, P., Karirya, A., Rasmussen, K., Small, J., Igel, C., Reiner, F., Melocik, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Fitts, Y., Morin, A., Pinzon, J., McClain, D., Morin, P., Porter, C., Loeffle, S., Kergoat, L., Issoufou, B.A., Savadogo, P., Wigneron, J.P., Poulet, B., Ciais, P., Kaufmann, R., Myneni, R., Saatchi, S., Fensholt, R., 2023. Sub-continental scale carbon stocks of individual trees in African drylands. *Nature* 615, 80–86.
- United Nations / Framework Convention on Climate Change, 2015. Paris agreement. In: Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change (21st Session, 2015: Paris).
- United Nations Department of Economic and Social Affairs, 2017. The sustainable development goals report. <https://doi.org/10.18356/4d038e1-en>.
- United Nations Department of Economic and Social Affairs, 2021. The global forest goals report 2021. <https://doi.org/10.18356/9789214030515>.

- Weiser, H., Schäfer, J., Winiwarter, L., Krašovec, N., Fassnacht, F.E., Höfle, B., 2022. Individual tree point clouds and tree measurements from multi-platform laser scanning in german forests. *Earth Syst. Sci. Data* 14, 2989–3012. URL: <https://essd.copernicus.org/articles/14/2989/2022/> <https://doi.org/10.5194/essd-14-2989-2022>.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3d shapenets: A deep representation for volumetric shapes. In: IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR). IEEE Computer Society, pp. 1912–1920.
- Xu, L., Saatchi, S.S., Yang, Y., Yu, Y., Pongratz, J., Bloom, A.A., Bowman, K., Worden, J., Liu, J., Yin, Y., et al., 2021. Changes in global terrestrial live biomass over the 21st century. *Sci. Adv.* 7 eabe9829.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al., 2020. Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241, 111716.
- Zhang, W., Yao, Q., Tong, H., Stenseth, N., 2003. Smoothing for spatiotemporal models and its application to modeling muskrat-mink interaction. *Biometrics* 59, 813–821.
- Zhang, L., Shao, Z., Liu, J., Cheng, Q., 2019. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and landsat 8 data. *Remote Sens.* 11, 1459.
- Zhang, A., Lipton, Z.C., Li, M., Smola, A.J., 2023. Dive into Deep Learning. Cambridge University Press.
- Zhou, Y., Tuzel, O., 2018. Voxelnet: end-to-end learning for point cloud based 3d object detection. In: IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), pp. 4490–4499.
- Zhuang, J., Tang, T., Ding, Y., Tatikonda, S.C., Dvornek, N., Papademetris, X., Duncan, J., 2020. Adabelief optimizer: adapting stepsizes by the belief in observed gradients. *NeurIPS* 33, 18795–18806.
- Zolkos, S.G., Goetz, S.J., Dubayah, R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sens. Environ.* 128, 289–298.