Multi-Layer Modeling of Dense Vegetation from Aerial LiDAR Scans

Ekaterina Kalinicheva[1, 2]   Loic Landrieu[1]   Clément Mallet[1]   Nesrine Chehata[1,3]

[1]Université Gustave Eiffel, ENSG, IGN, LASTIG, F-77454 Marne-la-Vallee, France
[2]INRAE, UMR 1202 BIOGECO, Université de Bordeaux, France
[3]Bordeaux INP, Université Bordeaux Montaigne, France

ekaterina.kalinicheva@ign.fr

Abstract

The analysis of the multi-layer structure of wild forests is an important challenge of automated large-scale forestry. While modern aerial LiDARs offer geometric information across all vegetation layers, most datasets and methods focus only on the segmentation and reconstruction of the top of canopy. We release WildForest3D, which consists of 29 study plots and over 2000 individual trees across 47 000m^2 with dense 3D annotation, along with occupancy and height maps for 3 vegetation layers: ground vegetation, understory, and overstory. We propose a 3D deep network architecture predicting for the first time both 3D pointwise labels and high-resolution layer occupancy rasters simultaneously. This allows us to produce a precise estimation of the thickness of each vegetation layer as well as the corresponding watertight meshes, therefore meeting most forestry purposes. Both the dataset and the model are released in open access: https://github.com/ekalinicheva/multi_layer_vegetation.
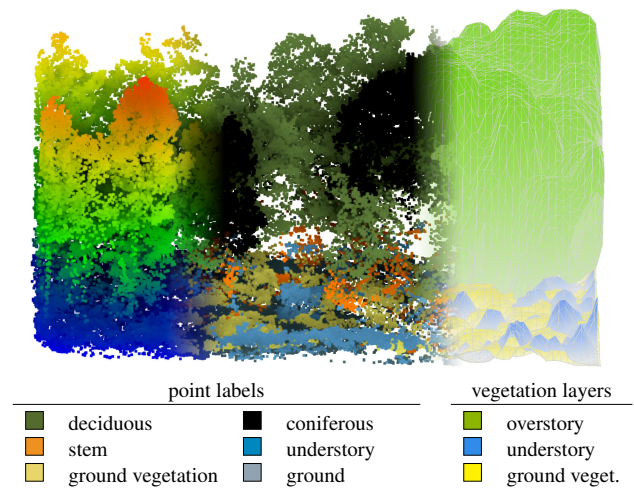
Figure 1. **Overview.** We propose WildForest3D, a novel dataset of annotated UAV-LS point clouds of dense forest (left), and a new model for the multi-layer analysis of vegetation. Our network performs 3D semantic segmentation (middle), and produces height maps and watertight meshes for three vegetation layers (right).

1. Introduction

The retrieval and analysis of the multiple layers of dense vegetation is an essential step for many forestry and ecology applications, such as forest management [20], biodiversity and habitat analysis [14, 31], biomass estimation [11, 29], or forest fire modeling [30, 32, 35]. Thanks to the constant advances in remote sensing technology, we can now gather large amounts of precise geometric and radiometric data on vast forested environments. However, standard satellite or aerial images are only suitable for the analysis of the top canopy layer [37]. Conversely, terrestrial laser scanning (TLS) captures detailed information on the bottom layers with slightly decreasing point density toward the upper vegetation layers. Unfortunately, TLS data acquisition over large areas (> 1 ha) is unpractical due to the high amount of resource involved [7]. In contrast, Airborne Laser Scanners

(ALS), and especially Unmanned Aerial Vehicles equipped with Laser Scanners (UAV-LS), can capture 3D point clouds over larger areas with sufficient point density [6, 7, 27].

Occlusion of the overstory tree cover leads to a limited 3D sampling of the bottom vegetation layers: their analysis from an airborne LiDAR perspective remains challenging, and most vegetation analysis algorithms only explicitly model the visible vegetation layers. Moreover, given the difficulty of annotating the lower strata of dense forests, very few annotated and public datasets exist for the training and evaluation of approaches modeling the multi-layer structure of dense forests. Large-scale assessment is also compromised.

In this paper, we propose a first-of-its-kind dataset of 29 plots of dense forest, 7 million 3D points and 2.1 million individual labels. This corresponds to over 2000 tree instances which were individually located and classified by *in*

situ observations from forestry experts. The labels indicate the nature of the vegetation structure containing the points: deciduous canopy, coniferous canopy, understory, stems, and ground. We use these annotations to produce high resolution maps of the occupancy and thickness of three vegetation layers: ground vegetation, understory, and overstory.

We also introduce a deep learning-based model for the automated analysis of multi-layer vegetation from aerial laser scans, see Figure 1. Our network operates directly on the 3D points to perform semantic segmentation of the point clouds and generate layer occupancy rasters that can be supervised end-to-end. We can then produce height maps for all layers, which we transform into watertight meshes. These outputs are useful for downstream applications such as biomass, carbon stock, fire fuel estimation [9, 13], soil illumination [19], or vegetation parameters extraction for the forest inventory [4]. The contributions of this paper are as follows:

- We present WildForest3D, the first 3D forest dataset with dense annotations, with over 2000 trees.

- We propose a deep learning-based method to model the multi-layer structure of dense forest from UAV-LS 3D point. Our method can be used to generate high resolution meshes of three different vegetation layers.

2. Related Work

The machine learning paradigm is still relatively new for forest data analysis. Most algorithms are based on the regression between point cloud features and *in situ* metrics, which typically do not generalize well outside of a given study area. We first present an overview of the existing open datasets, then the traditional approaches for structured vegetation analysis, and finally the recent deep learning advances.

Existing Forestry Datasets. Detailed annotation of ALS forestry data is a labor-intensive process, requiring expert knowledge and often *in situ* observations. For this reason, ALS annotations are often aggregated at the plot-level [22], but restricted for forest inventory purposes. They have never been exploited for fine-grained multi-layer semantic segmentation. Numerous countries now provide publicly available low resolution ALS scans at large scale [1–3] but annotations are still lacking. This data can be combined with the open forest inventory EFISCEN [36], containing annotations aggregated from plot to regions for 32 countries.

Several 3D LiDAR-based benchmarks have been released in the two last decades but they were focused on the specific task of individual tree crown delineation, i.e., a segmentation task over the canopy layer [41]. Weinstein *et al*. [42] provide a first-of-its-kind dataset for canopy crown detection and delineation containing co-registered airborne LIDAR point clouds, and RGB and hyperspectral images. In total, 30 975
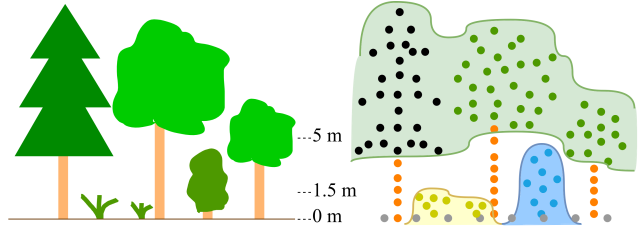

Figure 2. **Multi-Layer Structure.** We distinguish three vegetation layers: ground vegetation, understory, and overstory. Note that the layers are determined by the height of the vegetation instances they belong to, and not the point height: the bottom limit of the overstory layer may be lower than the top limit of the understory.

tree crowns are delimited with 2D bounding boxes. Our proposed WildForest3D contains richer annotations, at the level of the individual 3D points, which required an extensive campaign of *in situ* observations.

Vegetation Structure Analysis from ALS Data. Automated analysis of vegetation structure can be divided into tree-based [18, 34, 38, 39] and area-based [5, 25] methods. The former start by segmenting individual trees and aggregating their parameters over an area. The latter derive vegetation statistics from a study area and directly regress target variables. Both methods tend to overlook small elements of bottom layers due to the lower ALS point density in bottom layers, focusing on the analysis of higher vegetation, e.g., dominant trees.

Despite their limited financial value [17], the lower vegetation layers are important for many ecological applications, such as the study of forest habitats [14] or fire risk management [32]. However, detecting these layers is comparatively more challenging than overstory tree canopy. Hamraz *et al*. [17] report that while 90% of overstory trees are correctly segmented by most methods, lower vegetation detection barely reaches 60% accuracy. The lower vegetation analysis is often performed at an aggregated level over large areas. It consists in vegetation statistics at low-resolution level around $100 - 400m^2$ [21, 31, 40]. Ferraz *et al*. [10] model canopy cover of different vegetation layers using kernel-based estimators coupled with individual tree crown detection. Their approach proves particularly relevant for modeling the overstory, but is affected by the low point density of lower vegetation.

Deep Learning-Based Vegetation Structure Analysis. Bolstered by their wide adoption by the computer vision and remote sensing communities, 3D deep learning methods [15] have started to be used for forestry analysis. Lang *et al*. [24] estimate the canopy height at global scale from satellite LiDAR data. Ayrey *et al*. [4] use a 3D CNN model

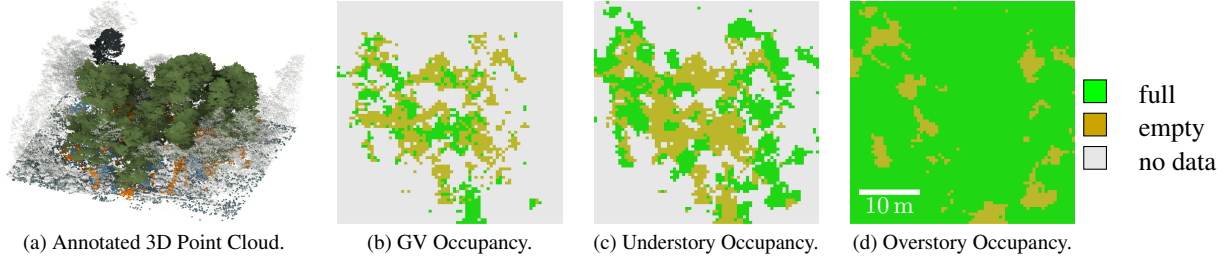| (a) Annotated 3D Point Cloud. | (b) GV Occupancy. | (c) Understory Occupancy. | (d) Overstory Occupancy. |

Figure 3. **Annotations of WildForest3D.** Our dataset contains 29 annotated dense forest plot, each with dense 3D annotations (a), and ground truth occupancy maps at 50cm resolution for the layers of ground vegetation (b), understory (c), and overstory (d). The colormap of (a) is the same as in Figure 1, with unannotated points in grey.

to extract forest inventory indicators at 10m resolution from voxelized ALS point clouds. Liu *et al.* [28] propose a deep-learning based regression model for estimating forest structural parameters from handcrafted 3D features. Kalinicheva *et al.* [23] use plot-level annotations to generate high resolution layer occupancy maps from ALS data. In this paper, we propose alternatively to explore the potential of using fine-grained 3D annotations to model the vegetation multi-layer structure using a deep network.

3. WildForest3D

In this section, we introduce WildForest3D, a novel dataset of aerial LiDAR scans of dense forests with dense 3D annotations and associated layer occupancy maps.

3.1. Vegetation layers

Relevant definitions of vegetation layers depend on the study area and are often subjective. In this paper, we choose vegetation layers deemed fit by forestry experts after the analysis of the study area. However, the number and heights of the layers can easily be changed in our provided dataset generation scripts for specific biomes. In the rest of the paper, we operate with three layers: (i) Ground Vegetation (GV) comprises shrubs, small trees, ferns and other vegetation that are between 0.5 and $1.5m$ of height; (ii) understory refers to small trees or large shrubs vegetation with a height between $1.5m$ and $5m$; (iii) overstory corresponds specifically to the crown of trees with a height above $5m$.

3.2. Characteristics of WildForest3D

The study area is located in the heavily forested French region of Aquitaine, and was scanned using a LiDAR installed on unmanned aerial vehicle with an average of 60 pulses per square meter. Each point is associated with its coordinates in Lambert-93 projection, the intensity value of returned laser signal, and the echo return number. The elevation of the points is given using a digital elevation model, such that the ground points are always at $z = 0m$.

Forestry experts selected 29 representative study plots of

size around $40 \times 40m^2$, for a total of $47\,000m^2$. Airborne LiDAR data of dense forest cannot be annotated without *in situ* data collection, as the branches of trees and shrubs intersect and are too difficult to parse from just 3D point clouds. This makes the annotation extremely costly and error-prone, all the more that the plots are typically not easily accessible. In each plot, an operator takes detailed measurements of the trees in its immediate proximity: position, height, inclination, species, stem circumference, and height of crown base. This data is reported in polar coordinates relative to the plot center, which is not directly conducive to supervising machine learning algorithms operating on 3D points.

3.3. 3D Point Annotation.

We designed a semi-automatic approach to convert these annotations into point labels across a set $\mathcal{C}$ of 6 classes: *ground*, *ground vegetation* (GV), *understory*, tree *stem*, and *coniferous* and *deciduous* tree crowns. As seen in Figure 2, each point is classified according to the vegetation instance it belongs to, and not its elevation. A point at a height of 3m but belonging to a coniferous with a tree top height of 8m will be classified as coniferous tree crown, and not as understory. Due to its nature, ground vegetation is particularly hard to annotate in 3D and is usually only reported by experts through local coverage ratio [43]. In our dataset, we chose to remove the few ground vegetation label altogether for 3D points, as they proved inconsistent. However, as we show in the next section, we can still recover this important layer through a geometric analysis and using other annotations.

3.4. Layer Occupancy Maps

Since our goal is to model the multi-layer nature of dense forests, we generate binary occupancy maps for the ground vegetation O^{GV}, understory O^{under}, and overstory O^{over}, see Figure 3. Our maps are of size $H \times W$, corresponding to an adjustable resolution of $r = 0.5$ m per pixel. Each pixel of each map is associated with either the value *full*, *empty*, or *no data*. This last value represents the ambiguity induced by partial annotations: not all trees and bushes are reported, and no ground vegetation label is annotated.
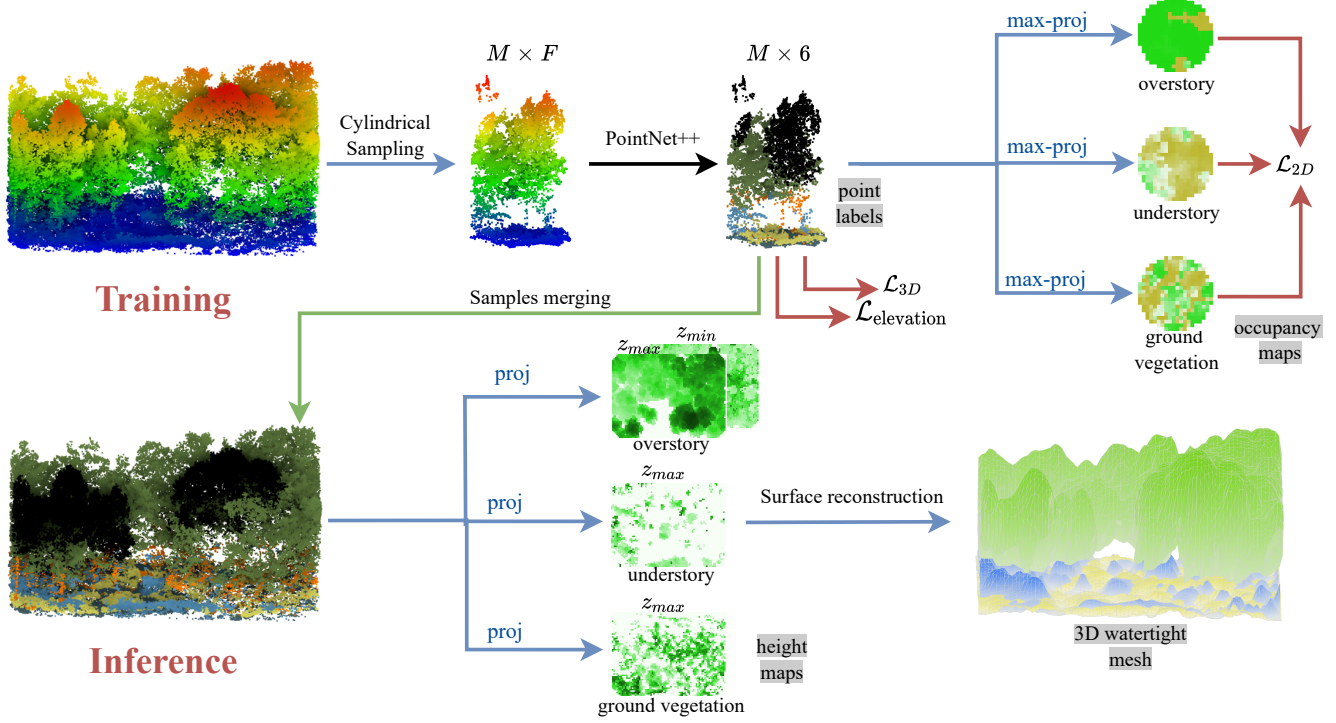
Figure 4. **Pipeline.** During training, our network performs the semantic segmentation of a 3D point cloud sample within 6 different classes. The point probabilities are projected onto rasters to obtain soft occupancy maps for 3 different vegetation layers. The network is supervised using both 2D and 3D predictions. During inference, the predictions are computed for an entire plot and the prediction are used to derive the minimum and maximum elevation of each layer. Finally, we can produce a watertight 3D mesh representing each vegetation layer.

The occupancy maps are obtained by vertically projecting the 3D points to their corresponding stratum. First, we project all points with labels *deciduous* and *coniferous* to the overstory layer to determine the full pixels of this map. We repeat this process for the understory by projecting points with the understory label. We do not have explicit labels for the ground vegetation and the annotations are partial. However, we can use the elevation to determine the ground vegetation layer map and complete the understory and overstory layers. However, using unannotated points yields more ambiguity, which we represent with the no data class.

First, we remove all annotated points. For each pixel, we compute the maximal height of all remaining points in its vertical extent. We then annotate the ground vegetation map and the pixels of the understory and overstory maps *that are not yet marked as full* according to this height:

- under 0.5m: the pixels of all maps are marked as empty;
- between 0.5 and 1.5m: the ground vegetation pixel is annotated as full, and the pixels of both overstory and understory are marked as empty;
- between 1.5 and 5m: the pixels of the overstory are marked as empty, the understory as full, and the ground vegetation as no data.

- over 5m: the pixels of the overstory are full, and the other maps as no data.

This geometric criterion does not overwrite full pixels of the overstory and understory derived from explicit point labels. The canopy layer do not have pixels without label, except if no points fall into the pixel.

4. Methodology

Our objective is to automatically derive from an aerial 3D LiDAR scans the occupancy and depth of different vegetation layers: ground vegetation, understory, and overstory. As illustrated in Figure 4, we compute 3D point classification and 2D occupancy maps, which are both used for supervision. Once trained, our approach can be used to derive the layers' thickness, and even to produce watertight meshes for each layer of such large scans.

4.1. 3D Point Classification

During training, we split the data into batches comprising N cylindrical samples of radius R and infinite vertical extent, randomly sampled on a grid with step size $R/5$. Each point of the cylinders is associated with a vector of F features containing their position, radiometric information, and the

number of echoes. The point's horizontal coordinates are given relative to the center of the cylinder. For the rest of the section, we consider a cylinder of M points and denote its point feature as $x \in \mathbf{R}^{M \times F}$.

This cylinder is randomly subsampled to a fixed number of points S, which are then classified using the PointNet++ (PTN++) network [33] among the class set $\mathcal{C}$ defined in Section 3.2. The prediction is then upsampled to the full resolution point cloud with nearest neighbor interpolation to obtain a predicted probability vector $p \in [0,1]^{M \times 6}$:

$$p = \underset{S \mapsto M}{\text{upsample}} \circ \text{PTN++} \circ \underset{M \mapsto S}{\text{downsample}}(x) \ . \quad (1)$$

The point-wise predictions are supervised with the cross-entropy loss $\mathcal{L}_{\text{3D}}$, using the point annotations.

4.2. Layer Occupancy Prediction

We now predict occupancy maps of size $[H, W]$ and pixel size s for each layer l in {GV, understory, overstory}. For $i, j \in [1 \cdots H, 1 \cdots W]$, we denote by $\text{proj}(i,j)$ the set of 3D points of the considered cylinder whose vertical projection falls into the pixel (i,j). We define the occupancy $o_{i,j}^l$ as the maximum probability associated with the points of the relevant class:

$$o_{i,j}^{\text{GV}} = \max_{k \in \text{proj}(i,j)} p_k^{\text{GV}} \ , \quad (2)$$

$$o_{i,j}^{\text{understory}} = \max_{k \in \text{proj}(i,j)} p_k^{\text{understory}} \ , \quad (3)$$

$$o_{i,j}^{\text{overstory}} = \max_{k \in \text{proj}(i,j)} \left(p_k^{\text{deciduous}} + p_k^{\text{coniferous}} \right) \ . \quad (4)$$

Note that neither the stem class nor the ground class are projected onto rasters, as they do not participate in the vegetation coverage. The predicted occupancy maps are supervised with the ground truth occupancy maps described in Section 3.4, and using the binary cross-entropy loss $\mathcal{L}_{\text{2D}}$ applied pixel-wise.

4.3. Unsupervised Elevation Modeling

To overcome the problem of partial annotations, we propose to explicitly model the elevation of points within two aggregated strata as suggested by Kalinicheva *et al.* [23]. We model the distribution of the point elevations as a mixture of two Gamma distributions: the lower (ground and ground vegetation) and higher strata. This mixture can be approximated using the Expectation–Conditional–Maximization algorithm [44] with manual initialization but without explicit stratum supervision.

The learned distributions Γ^{lower} and Γ^{higher} of the lower and higher strata allow us to define an additional loss as the average negative log-likelihood of the elevation of all points:

$$\mathcal{L}_{\text{elevation}} = -\frac{1}{M} \sum_m \log \left((p_m^{\text{soil}} + p_m^{\text{GV}}) \Gamma^{\text{lower}}(z_m) \right.$$
$$\left. + (p_m^{\text{under}} + p_m^{\text{decid}} + p_m^{\text{conif}} + p_m^{\text{stem}}) \Gamma^{\text{higher}}(z_m) \right) \ . \quad (5)$$

This loss encourages points with low elevation (high Γ^{lower}) to be predicted as ground or GV, and conversely for Γ^{higher}.

4.4. Model supervision

We supervise our model by combining three different losses: point-wise, pixel-wise, and elevation:

$$\mathcal{L} = \mathcal{L}_{\text{3D}} + \lambda \mathcal{L}_{\text{2D}} + \mu \mathcal{L}_{\text{elevation}} \ , \quad (6)$$

with λ, μ hyperparameters, set to 1 and 0.1. We train all models for 100 epochs of $N = 1000$ cylinder with a batch size of 5, a weight decay of 10^{-3}, and a learning rate of $5 \cdot 10^{-4}$, divided by 2 every 20 epochs. We use data augmentation by randomly rotating cylinder point clouds around the z-axis and adding a random Gaussian noise $\mathcal{N}(0, 0.01)$ clipped to $[-0.03, 0.03]$ to the normalized intensity values.

4.5. Inference

Once fully trained, our model can be used to model the vegetation layers on entire scanned areas. We sample cylinders on a grid of size R and compute the prediction for each point. When cylinders overlap, we average the predicted probabilities of points appearing in several cylinders. We associate to each point a hard assignment $l \in \mathcal{C}^M$ as the class with the highest predicted probability. This allows to define for each layer a binary occupancy map $\hat{o}^l$: a pixel is full if at least one point of the corresponding classes falls in its vertical extent. We can also define the minimal and maximal heights of each layer by considering the points of $\text{proj}(i,j)$ with corresponding labels. We set the lower boundary of the GV and understory layers as 0.

Using the occupancy, the minimal and maximal height maps, we can reconstruct a watertight triangular mesh for each layer by connecting adjacent non-empty pixels in a 4-neighborhood graph and using their corresponding height. Such side-product mesh can be used for downstream visualisation, computational or simulation tasks such as biomass estimation, forest fire prevention, ecological studies or ground illumination modeling [8, 16].

5. Numerical Experiments

In this section, we present quantitative and qualitative evaluations of our approach. To the best of our knowledge, no method exists with available implementation to derive fine-grained height maps across different vegetation layers from airborne LiDAR point clouds. To encourage reproducibility, both our dataset and method are released in open-source.

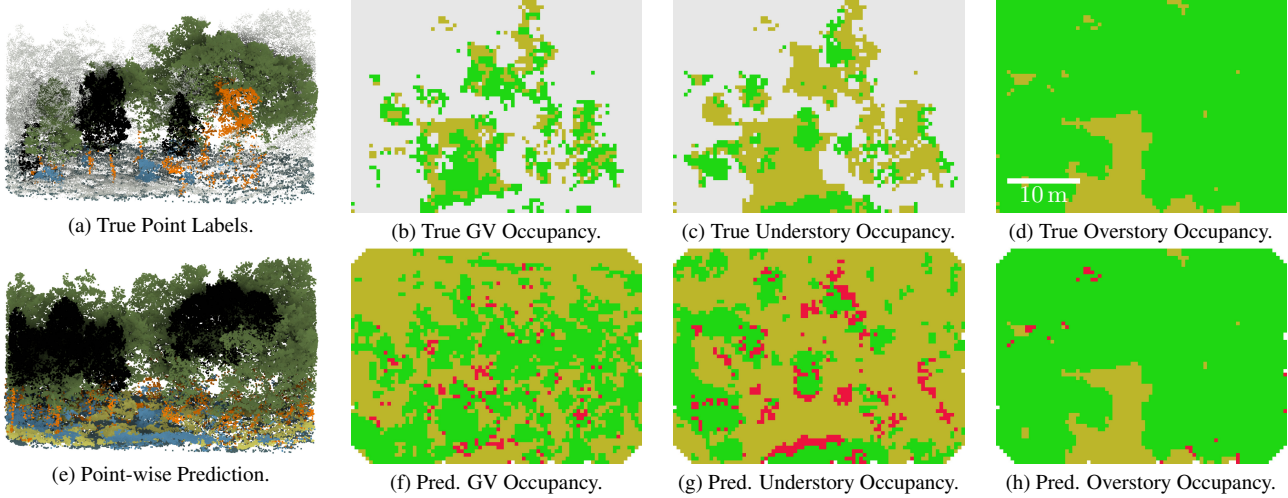| (a) True Point Labels. | (b) True GV Occupancy. | (c) True Understory Occupancy. | (d) True Overstory Occupancy. |
| (e) Point-wise Prediction. | (f) Pred. GV Occupancy. | (g) Pred. Understory Occupancy. | (h) Pred. Overstory Occupancy. |

Figure 5. **Qualitative Results.** We report the ground truth (top row) and prediction (bottom row) for the point labels and layer occupancy maps. The prediction errors are in red.

5.1. Experimental settings

We select the hyperparameters of our model to retain a fast inference and high precision on a NVIDIA RTX 3060 GPU. We use a pixel size of 0.5 m for the occupancy and height raster, and a sample cylinder radius of $R = 5$ m. To handle the varying density of point clouds and to facilitate batch processing, we randomly sample $S = 2^{14} = 16\,384$ points in each cylinder.

We base our implementation on PyTorch Geometric's [12] version of PointNet++ [33]. Our model configuration has a total of $147\,638$ parameters, see our repository for details. Our model can be trained from scratch in 7 hours, and can process $75\,\mathrm{m}^2$/second in inference on a standard workstation.

5.2. Evaluation Metrics and Baselines

We evaluate the quality of the point prediction, occupancy maps, and height prediction with the following metrics:

- **3D Prediction:** We use the Overall Accuracy (OA) and the Intersection-over-Union (IoU) of the point-wise predictions for all classes, except the ground vegetation class, which is not annotated in the 3D supervision. mIoU3D refers to the class-wise average over the classes without weight.

- **2D Raster Prediction:** We measure the performance with the OA and IoU of the vegetation binary occupancy prediction for all vegetation layers. mIoU2D is averaged across all three layers.

- **Height Prediction:** We use the Mean Absolute Error (MAE) and Mean Relative Error (MRE) of the layer height predictions, independently computed for each layer and only for pixels with full occupancy. The MRE

is defined as the ratio between the absolute value of the error and the true height of the considered pixel.

In order to put our results into perspective, we implement four simple baselines in the spirit of regression-based methods [21, 26, 40]. For each pixel, we compute handcrafted descriptors of the points above (maximum, minimum, mean, and standard deviation of z values, mean intensity value, mean return number, and number of points in 10 irregular bins of elevation between 0 and 30 m). We then train a logistic regression and a Random Forest (RF) classifier to predict 2D occupancy maps, and a linear regression and RF regression to predict height maps. Both RF models were set with maximum tree depth of 4 and 100 estimators.

5.3. Analysis

We select three plots for evaluation for their overall representativeness and the quality of their annotation. We remove them from the training set. Please refer to Figure 5 for visualisation of the results for one of those plots.

In Table 1, we report the quantitative performance for 3D point classification and 2D binary vegetation occupancy maps. While we observe lower score for stems, this class is particularly hard to annotate and may not be as accurately labeled as the other classes. We also notice the poor IoU score for the coniferous trees. This can be explained by the fact that the coniferous trees are more rare in our dataset that the deciduous ones and are taller in general. Hence, some tall deciduous trees are mistakenly classified as coniferous and vice versa, see Figure 6.

The binary occupancy rasters are overall well predicted, with a lower score for the challenging understory layer. The logistic regression and Random Forest baselines show good results for the prediction of the occupancy maps. However,

Table 1. **2D and 3D Prediction Performances:** Intersection-over-Union and Overall Accuracy of the prediction of 3D point (3D) and 2D occupancy maps regression (2D) on the evaluation set. We evaluate our approach for the 3D task, as well as a regression and random forest-based baselines for the 2D tasks. Our method performs both tasks simultaneously.

Task	Method	IoU, %							OA, %
		Ground	Ground veg.	Understory	Decid.	Conif.	Stem	Mean	
3D	Ours	95.1	-	43.3	90.0	23.5	15.5	53.5	90.5
2D	Ours	-	81.5	61.0	99.3		-	80.6	92.3
2D	Logistic Regression	-	62.2	40.5	99.5		-	67.4	87.3
2D	RF Classifier	-	89.3	62.5	99.6		-	83.8	94.2

Table 2. **Height Regression Performance.** We report the mean and standard deviation of the height of the top and bottom of different vegetation layers of our validation set. We compare the Mean Average Error (m) and Mean Error Ratio (%) of the predictions of our approach and two baseline regression models.

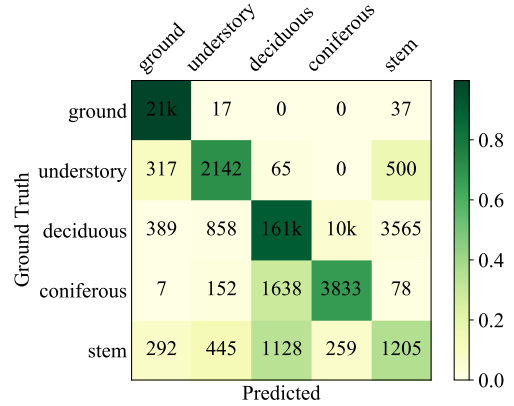|		Gr. veg. top	Under. top	Over. base	Over. top
	Height, m	0.9 ±0.3	2.7 ± 1.1	11.3 ± 5.2	17.3 ± 4.0
MAE, m	Ours	**0.03**	**0.3**	**1.1**	**0.1**
	Lin. Reg.	0.8	2.2	6.9	3.5
	RF Reg.	0.3	2.0	7.1	3.3
MRE, %	Ours	**2.9**	**22.0**	**26.5**	**0.7**
	Lin. Reg.	81.6	111.7	61.5	24.3
	RF Reg	35.8	85.0	64.4	22.7

Figure 6. **3D Prediction Confusion Matrix.** Cells are colored according to the proportion of points with the corresponding class. We observe a strong class imbalance, and a concentration of errors between the deciduous, coniferous, and stem classes.

note that our approach performs both 3D prediction and 2D regression simultaneously, which is necessary to recover the ground vegetation. Furthermore, the 2D occupancy rasters can not be used to predict the layers heights, but are used purely as a complementary supervision to the point labels.

In Table 2, we report the precision of the reconstructed height maps for all layers. The estimation is almost perfect for the top of the overstory, which is expected as the top of this layer can be reconstructed using the highest points. The other values are harder to estimate, particularly the top of the understory layer. Nonetheless, the MAE is overall below $1\,\mathrm{m}$ ($\sim 25\%$ MRE), which constitutes a solid baseline for such a complicated regression task. Both regression algorithms yield poor performance on height estimation, especially for the lower layers. We interpret this by the fact that the baselines operate pixel by pixel while our network has a large receptive field of several meters. We conclude that global spatial features are key for the structured reconstruction of vegetation layers in dense forests.

5.4. Ablation Study

We now evaluate the impact of our model design choices with quantitative results in Table 3 and qualitative illustra-

tions in Figure 7. We compare our main model with the following configurations.

a) No 2D Supervision: we set $\lambda = 0$ in Eq 6. Since there is no 3D annotation of the ground vegetation, the prediction for this layer cannot be performed, significantly impacting the prediction of the understory as well. Direct supervision of the occupancy raster proves to be necessary.

b) Elevation Modeling: we remove $\mathcal{L}_{el}$ by setting $\mu = 0$ in Eq 6. The performance is not significantly affected. However, a close visual estimation (Figures 7a and 7b) show an increasing number of confusions for ground and ground vegetation points, which are erroneously classified as stems. In contrast, when setting $\mu = 0.5$ instead of 0.1, the height precision of the estimation of both ground vegetation and understory decreases.

c) Sparser Point Sampling: we lower the number of points per cylinder to $S = 2^{13} = 8192$. Due to the lower point density of close-to-the-ground vegetation, especially under tall trees, sparser point sampling results in errors in the ground vegetation and understory layers, and tree stems are poorly classified, as seen in Figure 7c).

d) Size of Cylinders: we set the sampled cylinders radius

Table 3. **Quantitative Ablation Study.** Quantitative impact of some design and parameter choices.

Model	Param.	MRE, %				mIoU, %	
		Ground veg.	Understory	Overstory base	Overstory top	2D	3D
Our Parameterization	-	2.9	22.0	26.5	0.7	80.0	53.9
a) No 2D Supervision	$\lambda = 0$	-	32.6	25.1	0.7	44.4	56.7
b) No Elevation Modeling	$\mu = 0$	3.1	25.1	21.7	0.7	79.2	56.7
b) Higher Elevation Modeling	$\mu = 0.5$	4.0	28.8	26.4	0.7	80.6	51.7
b) Sparser Point Sampling	$S = 8192$	2.6	26.6	31.3	0.7	79.3	51.0
d) Smaller Cylinders	$R = 2$ m	3.1	26.0	28.7	0.6	82.7	55.4
d) Larger Cylinders	$R = 10$ m	3.9	29.5	30.2	0.7	77.6	52.8
e) Finer Raster Resolution	$r = 0.25$ m	1.1	15.5	25.3	0.8	79.7	56.9
e) Coarser Raster Resolution	$r = 1$ m	20.0	64.5	30.3	6.8	70.1	59.5
e) Coarser Raster Resolution	$r = 2$ m	37.7	99.5	47.5	19.8	59.4	56.4

(a) Ours. (b) No elevation modeling. (c) Sparser sampling. (d) Ours. (e) Small Cylinders.
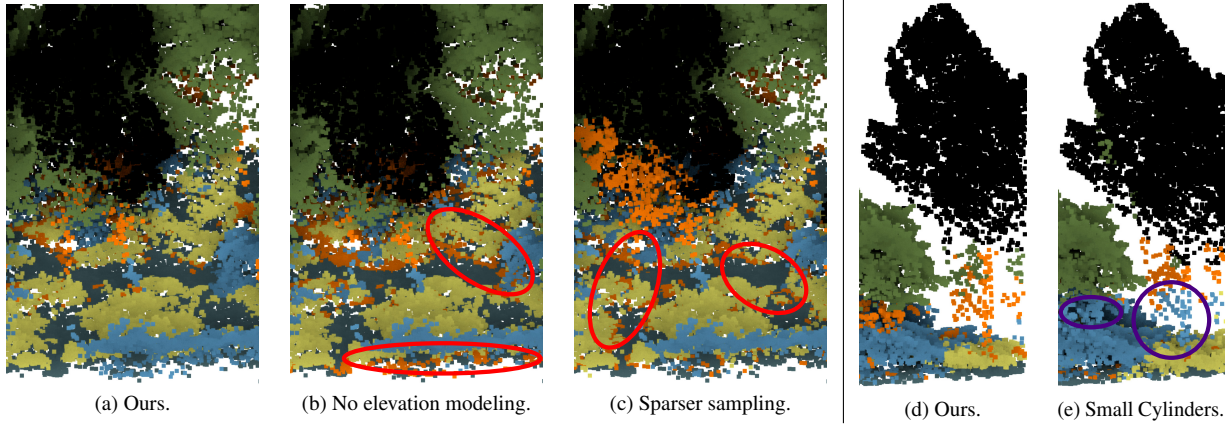
Figure 7. **Qualitative Ablation Study.** Illustration of the impact of the absence of elevation modeling (b), sparser sampling (c) and smaller cylinders (e). Among other inaccuracies, we observe in both (b) and (c) that many ground vegetation points are erroneously classified as stems ○. In (e), we observe that smaller cylinder size ($R = 2$m) leads to misclassify tree stems as understory ○.

R to 10 m while subsampling the same amount of points. A lower density leads to a similar performance decrease than the previous experiment. However, increasing the subsampling size for $R = 10$ would require a significant increase in GPU memory consumption.

We set $R = 2$ m and adjust the number of subsampled points to to $S = 2^{12} = 4096$ as well as corresponding parameters in the network. The resulting cylinders are tall and thin, which do not match well with the spherical feature aggregation scheme of PointNet++, and yield to overall lower performance (Figure 7e)).

e) Raster Resolution: we try different pixel sizes for the occupancy and height rasters: $r = 0.25$, 1, and 2 m. Small pixel sizes give good scores, but the ground truth has a higher proportion of no label pixels. In practice, this results in occupancy and height maps that are less easily visually interpretable. Larger pixels lead to lower precision as their size exceeds the scale of some individual instances.

6. Conclusion

In this paper, we introduced WildForest3D a novel airborne LiDAR dataset of dense forest with fine-grained 3D annotations. We also presented a deep learning algorithm for the structured analysis of vegetation layers from large 3D LiDAR point clouds. Contrary to existing approaches, our algorithm is able to provide high precision vegetation occupancy maps for different layers along with their corresponding height. We hope that by releasing our data and code in open source we will encourage the computer vision community to consider the challenging and impactful problem of automated forest data analysis at a large scale.

In future works we will modify our model to operate on complementary data sources such as high resolution aerial images, multispectral satellite image time series, and contextual information.

References

[1] Dataforsyningen. https://dataforsyningen.dk/data. 2

[2] Lantmateriet, Laser data Download, forest. https://www.lantmateriet.se/en/maps-and-geographic-information/geodataprodukter/produktlista/laserdata-nedladdning-skog/. 2

[3] National Land Survey of Finland. https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/laser-scanning-data. 2

[4] Elias Ayrey, Daniel J. Hayes, John B. Kilbride, Shawn Fraver, John A. Kershaw, Bruce D. Cook, and Aaron R. Weiskittel. Synthesizing disparate LiDARand satellite datasets through deep learning to generate wall-to-wall regional inventories for the complex, mixed-species forests of the Eastern United States. *Remote Sensing*, 2021. 2

[5] Marc Bouvier, Sylvie Durrieu, Richard Fournier, and Jean-Pierre Renaud. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sensing of Environment*, 2015. 2

[6] Benjamin Brede, Alvaro Lau, Harm M. Bartholomeus, and Lammert Kooistra. Comparing RIEGL RiCOPTER UAV LiDAR derived canopy height and DBH with terrestrial LiDAR. *Sensors*, 2017. 1

[7] Kim Calders, Jennifer Adams, John Armston, Harm Bartholomeus, Sebastien Bauwens, Lisa Patrick Bentley, Jerome Chave, F. Mark Danson, Miro Demol, Mathias Disney, Rachel Gaulton, Sruthi M. Krishna Moorthy, Shaun R. Levick, Ninni Saarinen, Crystal Schaaf, Atticus Stovall, Louise Terryn, Phil Wilkes, and Hans Verbeeck. Terrestrial laser scanning in forest ecology: Expanding the horizon. *Remote Sensing of Environment*, 2020. 1

[8] Frederick Draper, Flavia Costa, Gabriel Arellano, and et al. Amazon tree dominance across forest strata. *Nature Ecology & Evolution*, 2021. 5

[9] António Ferraz, Frédéric Bretar, Stéphane Jacquemoud, Gil Gonçalves, Luisa Pereira, Margarida Tomé, and Paula Soares. 3D mapping of a multi-layered Mediterranean forest using ALS data. *Remote Sensing of Environment*, 2012. 2

[10] António Ferraz, Clément Mallet, Stéphane Jacquemoud, Gil Gonçalves, Margarida Tomé, Paula Soares, Luísa Pereira, and Frédéric Bretar. Canopy density model: A new als-derived product to generate multilayer crown cover maps. *Transactions on Geoscience and Remote Sensing*, 2015. 2

[11] António Ferraz, Sassan Saatchi, Clément Mallet, Stéphane Jacquemoud, Gil Gonçalves, Carlos Alberto Silva, Paula Soares, Margarida Tomé, and Luisa Pereira. Airborne LiDAR estimation of aboveground forest biomass in the absence of field inventory. *Remote Sensing*, 2016. 1

[12] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. 6

[13] Matthew G. Gale, Geoffrey J. Cary, Albert I.J.M. Van Dijk, and Marta Yebra. Forest fire fuel through the lens of remote sensing: Review of approaches, challenges and future directions in the remote sensing of biotic determinants of fire behaviour. *Remote Sensing of Environment*, 2021. 2

[14] Etienne Gaujour, Bernard Amiaud, Catherine Mignolet, and Sylvain Plantureux. Factors and processes affecting plant biodiversity in permanent grasslands. a review. *Agronomy for sustainable development*, 2012. 1, 2

[15] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3D point clouds: A survey. *Transactions on pattern analysis and machine intelligence*, 2020. 2

[16] Torsten Hädrich, Daniel T. Banuti, Wojtek Pałubicki, Sören Pirk, and Dominik L. Michels. Fire in paradise: Mesoscale simulation of wildfires. *Transactions on Graphics*, 2021. 5

[17] Hamid Hamraz, Marco Contreras, and Jun Zhang. Forest understory trees can be segmented accurately within sufficiently dense airborne laser scanning point clouds. *Scientific Reports*, 7, 2017. 2

[18] Hamid Hamraz, Marco A. Contreras, and Jun Zhang. A robust approach for tree segmentation in deciduous forests using small-footprint airborne LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 2016. 2

[19] Werner Härdtle, Goddert von Oheimb, and Christina Westphal. The effects of light and soil conditions on the species richness of the ground vegetation of deciduous forests in northern germany (schleswig-holstein). *Forest ecology and management*, 2003. 2

[20] Juha Jantunen, Kimmo Saarinen, and Olli Marttila. Effects of forest management on field layer vegetation: a comparison between finnish and russian karelia. *Scandinavian Journal of Forest Research*, 2001. 1

[21] Lukas R. Jarron, Nicholas C. Coops, William H. MacKenzie, Piotr Tompalski, and Pamela Dykstra. Detection of sub-canopy forest structure using airborne LiDAR. *Remote Sensing of Environment*, 2020. 2, 6

[22] Ekaterina Kalinicheva. Vegetation Stratum Occupancy Prediction from Airborne LiDAR 3D Point Clouds, Zenodo, Oct. 2021. 2

[23] Ekaterina Kalinicheva, Loic Landrieu, Clément Mallet, and Nesrine Chehata. Predicting vegetation stratum occupancy from airborne LiDAR data with deep learning, arXiv 2201.08051, 2022. 3, 5

[24] Nico Lang, Nikolai Kalischek, John Armston, Konrad Schindler, Ralph Dubayah, and Jan Dirk Wegner. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sensing of Environment*, 2022. 2

[25] Hooman Latifi, Marco Heurich, Florian Hartig, Jörg Müller, Peter Krzystek, Hans Jehl, and Stefan Dech. Estimating over- and understorey canopy density of temperate mixed stands by airborne LiDAR data. *Forestry*, 2016. 2

[26] Hooman Latifi, Marco Heurich, Florian Hartig, Jörg Müller, Peter Krzystek, Hans Jehl, and Stefan Dech. Estimating over- and understorey canopy density of temperate mixed stands by airborne LiDAR data. *Forestry: An International Journal of Forest Research*, 08 2015. 6

[27] Xinlian Liang, Yunsheng Wang, Jiri Pyörälä, Matti Lehtomäki, Xiaowei Yu, Harri Kaartinen, Antero Kukko, Eija Honkavaara, Aimad EI Issaoui, Olli Nevalainen, et al. Forest in situ observations using unmanned aerial vehicle as an alternative of terrestrial measurements. *Forest ecosystems*, 2019. 1

[28] Hao Liu, Xin Shen, Lin Cao, Ting Yun, Zhengnan Zhang, Xiaoyao Fu, Xinxin Chen, and Fangzhou Liu. Deep learning in forest structural parameter estimation using airborne lidar data. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021. 3

[29] Dengsheng Lu. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, 2006. 1

[30] David A MacLean. Forest management strategies to reduce spruce budworm damage in the fundy model forest. *The Forestry Chronicle*, 1996. 1

[31] Sebastian Martinuzzi, Lee Vierling, William Gould, Michael Falkowski, Jeffrey Evans, Andrew Hudak, and Kerri Vierling. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, 2009. 1, 2

[32] Donald McKenzie and Crystal Raymond. Modeling understory vegetation and its response to fire. *Models for Planning Wildlife Conservation in Large Landscapes*, 2011. 1, 2

[33] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *NeurIPS*, 06 2017. 5, 6

[34] J. Reitberger, Cl. Schnörr, P. Krzystek, and U. Stilla. 3d segmentation of single trees exploiting full waveform lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), 2009. 2

[35] David V. Sandberg, Roger D. Ottmar, and Geoffrey H. Cushon. Characterizing fuels in the 21st century. *International Journal of Wildland Fire*, 2001. 1

[36] Mart-Jan Schelhaas, Simo Varis, Andreas Schuck, and Gert-Jan Nabuurs. EFISCEN Inventory Database, 2006. `http://www.efi.int/portal/virtual_library/databases/efiscen/`. 2

[37] Mao-Ning Tuanmu, Andrés Viña, Scott Bearer, Weihua Xu, Zhiyun Ouyang, Hemin Zhang, and Jianguo Liu. Mapping understory vegetation using phenological characteristics derived from remotely sensed data. *Remote Sensing of Environment*. 1

[38] Jari Vauhkonen, Liviu Ene, Sandeep Gupta, Johannes Heinzel, Johan Holmgren, Juho Pitkänen, Svein Solberg, Yunsheng Wang, Holger Weinacker, K. Marius Hauglin, Vegard Lien, Petteri Packalén, Terje Gobakken, Barbara Koch, Erik Næsset, Timo Tokola, and Matti Maltamo. Comparative testing of single-tree detection algorithms under different types of forest. *Forestry*. 2

[39] Cédric Vega, Ahmed Hamrouni, Salima El Mokhtari, Jules Morel, Jérôme. Bock, Jean-Pierre Renaud, Marc Bouvier, and Sylvie Durrieu. PTrees: A point-based approach to forest tree extraction from LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 2014. 2

[40] Lisa A. Venier, Tom Swystun, Marc J. Mazerolle, David P. Kreutzweiser, Kerrie L. Wainio-Keizer, Ken A. McIlwrick, Murray E. Woods, and Xianli Wang. Modelling vegetation understory cover using LiDAR metrics. *PLOS ONE*, 2019. 2, 6

[41] Yunsheng Wang, Juha Hyyppä, Xinlian Liang, Harri Kaartinen, Xiaowei Yu, Eva Lindberg, Johan Holmgren, Yuchu Qin, Clément Mallet, António Ferraz, Hossein Torabzadeh, Felix Morsdorf, Lingli Zhu, Jingbin Liu, and Petteri Alho. International benchmarking of the individual tree detection methods for modeling 3-d canopy structure for silviculture and forest ecology using airborne laser scanning. *Transactions on Geoscience and Remote Sensing*, 2016. 2

[42] Ben Weinstein, Sergio Marconi, and Ethan White. Data for the Neon Tree evaluation benchmark, Zenodo, Jan. 2022. 2

[43] Brian M. Wing, Martin W. Ritchie, Kevin Boston, Warren B. Cohen, Alix Gitelman, and Michael J. Olsen. Prediction of understory vegetation cover with airborne LiDAR in an interior ponderosa pine forest. *Remote Sensing of Environment*, 2012. 3

[44] Derek S. Young, Xi Chen, Dilrukshi C. Hewage, and Ricardo Nilo-Poyanco. Finite mixture-of-Gamma distributions: estimation, inference, and model-based clustering. *Advances in Data Analysis and Classification*, 2019. 5