Enhancing forest attribute prediction by considering terrain and scan angles from Lidar point clouds: a neural network approach

Karun Dayal ¹, Sylvie Durrieu ², Kamel Lahssini ², Dino Ienco ², and Jean-Matthieu Monnet ²

¹INRAE
²Affiliation not available

October 30, 2023

Abstract

Sensitivity of lidar metrics to scan angle can affect the robustness of area-based approach (ABA) models, and modelling the interplay of scan geometry and terrain properties can be complex. The study hypothesises that neural networks can manage the interplay of lidar acquisition parameters, terrain properties, and vegetation characteristics to improve ABA models. The study area is in Massif des Bauges Natural Regional Park, eastern France, comprising 291 field plots in a mountainous environment with broadleaf, coniferous, and mixed forest types. Field plots were scanned with a high overlap from multiple flight lines and the corresponding point clouds were considered independently to expand the standard ABA dataset (291 observations) to create a dataset containing 1095 independent observations. Computation of lidar, terrain and scan metrics for each point cloud associated each observation in the expanded dataset with the scan information in addition to the lidar and terrain information. A multilayer perceptron (MLP) was used to model basal area and total volume to compare the predictions resulting from standard and expanded ABA datasets. With expanded datasets containing lidar, terrain and scan information, the $R^2$ for the median predictions per plot were higher ($R^2$ of 0.83 and 0.85 for $BA$ and $V_{tot}$) than predictions with standard datasets($R^2$ of 0.66($BA$) and 0.71($V_{tot}$)) containing only lidar metrics. It also outperformed an MLP model for a dataset with lidar and terrain information ($R^2$ of 0.77 ($BA$ and $V_{tot}$)). The MLP performed better than RF regression, which could not sufficiently exploit additional terrain and scan information.
Enhancing forest attribute prediction by considering terrain and scan angles from Lidar point clouds: a neural network approach

Karun R. Dayal*, Sylvie Durrieu¹, Kamel Lahssini², Dino Ienco³, Jean-Matthieu Monnet⁴

Abstract—Sensitivity of lidar metrics to scan angle can affect the robustness of area-based approach (ABA) models, and modelling the interplay of scan geometry and terrain properties can be complex. The study hypothesises that neural networks can manage the interplay of lidar acquisition parameters, terrain properties, and vegetation characteristics to improve ABA models. The study area is in Massif des Bauges Natural Regional Park, eastern France, comprising 291 field plots in a mountainous environment with broadleaf, coniferous, and mixed forest types. Field plots were scanned with a high overlap from multiple flight lines and the corresponding point clouds were considered independently to expand the standard ABA dataset (291 observations) to create a dataset containing 1095 independent observations. Computation of lidar, terrain and scan metrics for each point cloud associated each observation in the expanded dataset with the scan information in addition to the lidar and terrain information. A multilayer perceptron (MLP) was used to model basal area and total volume to compare the predictions resulting from standard and expanded ABA datasets. With expanded datasets containing lidar, terrain and scan information, the R² for the median predictions per plot were higher (R² of 0.83 and 0.85 for BA and Vtot) than predictions with standard datasets (R² of 0.66(BA) and 0.71(Vtot)) containing only lidar metrics. It also outperformed an MLP model for a dataset with lidar and terrain information (R² of 0.77 (BA and Vtot)). The MLP performed better than RF regression, which could not sufficiently exploit additional terrain and scan information.

Index Terms—ABA, ANN, forest attribute, lidar, RF, topography

I. INTRODUCTION

The ability of lidar technology to create dense three-dimensional representations of vegetation has been widely used to extract useful information to characterise forest properties[1]. With airborne lidar systems (ALS), it is possible to cover large areas to generate accurately measured three-dimensional point clouds. The lidar system involves a scanning mechanism that emits lidar pulses at a range of incident angles, i.e. vertical to inclined. An ideal lidar acquisition would involve a comprehensive scanning of the forest from multiple angles to obtain an accurate and sufficiently dense representation in the form of three-dimensional point clouds. Area-based approaches (ABA) utilise such point clouds wherein statistical descriptors (lidar metrics) of point clouds of representative forest plots are statistically linked with field measurements of desired forest attributes. In the interest of reliable predictions using lidar metrics, lidar metrics should be stable under different acquisition characteristics.

An essential requirement in lidar remote sensing of forests is detecting the ground surface beneath the canopy, enabling accurate measurement of vegetation heights. As a result, the lidar scan angle, or the (half) field of view, has been limited to 20° to ensure most lidar pulses reach the ground [2]. So far, most studies involving lidar remote sensing for forestry applications have followed this convention [3]–[9]. Recently, some studies [10]–[13] have tried to assess the impact of scan angles greater than 20°, and many studies involving UAV-based lidar data routinely used much higher scan angles [14], [15]. It may be difficult for highly inclined lidar pulses to reach the ground surface owing to the increased occlusions. Nonetheless, it is also true that probing lidar canopies with inclined pulses may also lend newer insights or different perspectives [16]. A related study [13] observed that datasets comprising nadir point clouds did not always result in better ABA models, thereby emphasising that forest canopies are not a homogenous medium, and the lidar-derived information (lidar metrics) depends on how the lidar pulses sample the canopy. Furthermore, two lidar acquisitions may not have identical properties, and the lidar metrics could be affected by the overall acquisition geometry as characterised by the acquisition properties (sensor properties, scan angle, scan azimuth, flying height), terrain properties, and vegetation structural characteristics.

In area-based approaches (ABA), there are generally numerous lidar metrics to choose from, and new metrics are constantly being developed to comprehensively summarise the vegetation structural information. Standard metrics used over the years include statistical descriptors such as the mean, standard deviation, variance, entropy, and percentiles of the height or intensity values, cover rate metrics, density metrics, and gap-fraction. A stepwise selection procedure is often employed to identify metrics useful in predicting forest attributes using multiple linear regression [17]. However, the final set of metrics may vary depending on the forest type or

Karun Reuel Dayal, Sylvie Durrieu, Dino Ienco and Kamel Lahssini are with the INRAE, UMR TETIS, University of Montpellier, 34000Montpellier, France (e-mail: karun.dayal@inrae.fr; sylvie.durrieu@inrae.fr; dino.ienco@inrae.fr; kamel.lahssini@inrae.fr).

Jean-Matthieu Monnet is with the University Grenoble Alpes, INRAE, LESSEM, 38400 Grenoble, France (e-mail: jean-matthieu.monnet@inrae.fr).
lidar acquisition parameters. Another approach is to use expert knowledge to define and select a short list of metrics that could explain most, if not all, of the variance of the dependent variables [18]. Still, assessing the influence of scan angle on selected metrics and, subsequently, on the forest attribute predictions may not always be practical.

Furthermore, the influence of lidar metrics may be site-specific, and it is advisable to assess the effects of scan angle before further analysis on a case-by-case basis [19], [20]. Traditionally, the modelling of forest attributes is done using parametric and non-parametric models. Due to their simplicity, parametric methods such as ordinary least squares (OLS) regression have been widely used by studies to model forest attributes [17], [21], [22]. However, parametric methods such as OLS use only a few metrics to avoid overfitting and the use of correlated variables. In non-parametric methods, there are no such limitations as these methods do not depend on any assumptions regarding the data. They can accommodate nonlinear relationships between the dependent and independent variables [23]. Therefore, they are suited for modelling complex interactions between several lidar variables, acquisition geometry, and vegetation properties. Both KNN and RF are among the most commonly used non-parametric approaches in ABA [24]. However, RF was found to have a higher level of transferability to new areas than KNN [21]. Artificial neural networks (ANN) have also become a popular non-parametric method to address inherent non-linearity in datasets [25], [26]. The feed-forward back-propagation multi-layered perceptron (MLP) is often used with remote sensing data [25]. It consists of a network of several interconnected layers of neurons designed to mimic human brain capabilities, such as generalisation and understanding complex patterns. Among the various non-parametric methods, the MLP has been demonstrated to have better generalisation capabilities [27], [28].

However, MLP methods depend on the volume of data, which generally comprises large datasets with several thousand samples. In ABA approaches, which involve collecting labour-intensive field measurements in often-complex terrains, it is impossible to measure many field plots (samples) as field measurements make up a significant part of the costs. The number of field plots in ABA models typically ranges from a few tens to a few hundred. In addition, only a few field plots describe particular stand types. Generally, lidar acquisitions for forests are planned with multiple overlaps to thoroughly sample each forest area (or field plot) from multiple locations, and the point clouds acquired from each location may be considered independent observations. In addition, owing to the heterogeneous nature of the vegetation, the point clouds for any field plot retain differences and be used independently to increase the number of samples since each point cloud results from the interaction of the physical lidar signal with the natural vegetation. In other words, a point cloud obtained from a flight line is defined by the interplay of acquisition parameters, terrain properties, and vegetation characteristics, making it possible to consider it a unique and independent observation in ABA models. The combination of acquisition properties and (virtually limitless) lidar metrics for each observation pose challenges that are better handled with non-parametric methods.

A previous study over mountainous terrain [29] demonstrated that neural networks are well-suited to exploit terrain-related information to improve standard ABA predictions. In another study over the same area [13], single flight line datasets resulted in variable ABA predictions with parametric models comprising widely used metrics sensitive to scan angle. The objective of this study was to evaluate the benefit of using individual point clouds obtained from multiple flight lines independently to a) increase the number of observations and b) retain acquisition properties of each flight line to improve ABA predictions. We demonstrate i) the benefits of expanding lidar datasets based on flight lines to build ABA models and ii) the capacity of multilayer perceptron to model complex interactions between lidar signal and acquisition properties.

II. Materials and methods

A. Study area and field measurements

The study site is the Massif des Bauges Natural Regional Park in the French Alps. It is located between the two administrative departments of Savoie and Haute-Savoie and covers an area of approximately 850 km². The terrain is hilly (plot altitudes range from 420 m to 1760 m). The most common tree species comprise silver fir (Abies alba), Norway spruce (Picea abies), and common beech (Fagus sylvatica). Field inventory involved measurements of 291 15 m radius circular plots during the spring and fall of 2018. Plot centre locations were measured using differential GNSS (DGNSS, Trimble, USA). Field inventory protocol involved measuring tree Diameter at Breast Height (DBH, measured 1.3 m above ground) of trees with DBH greater than 17.5 cm. Small trees (7.5 cm ≤ DBH < 17.5 cm) were counted within a plot radius of 10 m and classified as either coniferous or broadleaf.

Since DBH and height measurements were unavailable for all the trees with DBH greater than 7.5 cm, computation of basal area, stem, and total volumes at plot level required estimations for unmeasured trees. Firstly, the number of small trees was extrapolated from the number of trees in 10 m radius plots to 15 m radius plots. Secondly, the nationwide tree inventory database (NFI) generated by IGN (Institut National de l’Information Géographique et Forestière) containing measurements of trees with DBHs in the 7.5 cm to 17.5 cm range was used to extrapolate DBH and height values for non-measured trees. All NFI plots in the ecoregion that includes the study site were selected to have forest plots with similar climatic and growing conditions to those measured on the study site. For trees with DBH ranging from 7.5 cm to 17.5 cm, the median DBH value in the NFI database is 11.1 cm. This value was used to compute the basal area of the trees with DBHs lower than 17.5 cm. Using NFI measurements, allometric relationships were established for each species (or group of species when the number of trees was not high enough) to estimate the heights of all the trees when there were no available height measurements. Volumes were then computed using the allometric
equations available in [29], [30] and [13] followed the same protocol. A summary of field measurements is given in Table I.

| TABLE I
| SUMMARY OF THE BASAL AREA (BA) AND TOTAL VOLUME (V_TOT) FOR THE 291 INVENTORY PLOTS |
|-----------------|-----------------|-----------------|-----------------|
| Basal area (m²/ha) | Total volume (m³/ha) |    |    |
| Min | Mean | Max | Min | Mean | Max |
| 0.36 | 30.2 | 89.7 | 2.52 | 312.1 | 1172 |

B. Lidar data

Lidar data acquisition was carried out in two missions. The first mission (summer 2016) covering areas of department 73 resulted in a dataset of 4–5 points/m² point density on average and the second mission (summer 2018) covering areas in department 74 resulted in a dataset of approximately 14 points/m² point density on average. Lidar acquisitions were carried out with multiple overlaps to scan each field plot from several locations with different azimuths and scan angles. Fig 1 shows the locations of the field plots in the study area along with aircraft locations while scanning respective field plots. The acquisition parameters for the two missions are given in Table II.

| Table II
| ACQUISITION PARAMETERS FOR THE TWO MISSIONS (CALCULATED FROM DATA) |
|-----------------|-----------------|-----------------|
| Date of acquisition | Sensor | Wavelength (nm) | Scan angle (deg) | Beam divergence (mrad) | Ground speed (m/s) | Point density (pts/m²) | Flight height (AGL) (m) |
| September 2016 | Leica ALS70-HP | 1064 | 46° (+23°/-23°) | 0.15 | 85 | 4 | 1500* |
| September 2018 | Riegl LMSQ780 | 1064 | 60° (+30°/-30°) | ≤0.25 | 45 | 14 | 1050 |

C. Splitting of point clouds based on flight lines

Point clouds corresponding to the field plots were clipped from the lidar data using the coordinates of the plot centres and plot radius (15 m). Due to flight line overlaps, the point cloud for a given plot is typically a composite of point clouds acquired with different scanning configurations. The point cloud was split for each plot based on the constituent flight lines. Each resulting constituent point cloud was characterised by the mean of the scan angles (MSA) with which it was scanned. We did not consider those point clouds acquired with MSA greater than 30° as they were most likely acquired when the aircraft made turns, and there were few such instances. The fundamental ‘unit’ in our experiments is the point cloud for a plot acquired from a single flight line. We assessed pulse densities for each point cloud, and 90% of the constituent point clouds had a pulse density greater than one pulse per m². We computed the area covered by each constituent point cloud by fitting a two-dimensional hull to the points projected onto a horizontal plane. Then, an area threshold was used to drop any constituent point cloud that covered less than 90% of the total plot area (Fig 2).

Fig. 1. Location of the study site, distribution of field plots, and coverage of lidar missions. Field plots are depicted in larger dots with the colours corresponding to the number of unique flight lines from which the plots were scanned completely. The black dots depict the approximate (average) location of the aircraft when it scanned a field plot.

Fig. 2. Flight line that partially covers a plot

The flight trajectory data was used to extract the locations of the aircraft while scanning respective field plots, and the average location of the aircraft was computed. The azimuth of
the same splits of scan geometry, i.e. mean scan angle, azimuth, and distance. Therefore, two kinds of datasets were considered. In the first kind of dataset, point clouds were not separated based on the flight lines. This dataset was called the standard dataset and contained as many point clouds as field plots in the study (291). Point clouds were separated based on the flight line information for the second dataset, called the expanded dataset. In the expanded dataset, there was one to eight point clouds per field plot resulting in 1095 point clouds and corresponding to a mean of 3.8 point clouds per plot (Fig. 1).

D. Lidar metrics

Lidar metrics were computed for each point cloud in both datasets after normalising the point clouds in height, i.e., transforming the point elevation into height above the ground using the lidar-derived DTM. All points below a height threshold of five meters were considered lower vegetation and filtered out. Fifty-five metrics related to heights, intensities, and canopy were computed. The height-based metrics are the statistical distributions calculated for the Z values of the point cloud. The intensity metrics comprise statistical descriptors of the intensity values. Canopy metrics consist of gap fraction [31] and rumple index [32]. The gap fraction was computed as the ratio of the number of returns below the 5m threshold to the total number of returns. Rumple index is the ratio of the 3D surface area of the canopy to the surface area of the ground. Gap fraction and rumple index were found to be very sensitive to the scan angle [33]. The summary of these metrics is given in Table III.

Point clouds can change according to the local topography and viewing configuration for a given forest plot. Depending on the slope, orientation, elevation, scan angle, and aircraft position, there could be several cases. In Fig 3, the illustrations depict two possibilities of lidar scanning a plot on a slope with similar scan angles, a) scanning along the slope and b) scanning against the slope. An example of a point cloud scanned from different directions is shown in 3c, wherein two point clouds with similar mean scan angles can have different properties due to the interaction between terrain properties and scanning parameters. Information on scan geometry and terrain properties were thus added as six additional variables, resulting in a total of 61 variables (Table III). The three previously defined variables defined scan geometry, i.e. mean scan angle, scanning azimuth, and scanning distance (see section C), differed for the same plot location according to the flight line. Terrain information was computed by generating digital terrain models (DTM) of a resolution of 1m. The DTM of each plot was used to generate slope and aspect maps. The average slope, aspect, and elevation values were computed from the slope and aspect maps and DTMs, respectively and were replicated for all the point clouds acquired from different flight lines over a given plot. As the overall point cloud geometry is defined by the acquisition parameters, terrain properties and vegetation properties, point clouds obtained from different flight lines were used independently to compute all the metrics mentioned earlier to retain the heterogeneities and homogeneities.

For the expanded dataset data set, the values of the dependent variables, i.e., $BA$ and $V_{tot}$ were also replicated for each plot depending on the corresponding number of flight lines or point clouds. All the values were scaled between 0 and 1. All lidar metrics were computed using the lidR package in R [34].

E. Experiments and cross-validation scheme

When the point clouds were considered per plot, i.e. the standard dataset, there were 291 samples. The same splits of data as those used for cross-validation of the models in [30] were used to compare results from our study directly. The standard dataset containing the field plot measurements and

---

<table>
<thead>
<tr>
<th>Name</th>
<th>Nos</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{nth}$ $</td>
<td>n \in {1,2,3,4,5}$</td>
<td>5</td>
<td>% of nth echos</td>
</tr>
<tr>
<td>kurtosis_first</td>
<td>1</td>
<td>Kurtosis of first returns</td>
<td>Height-based</td>
</tr>
<tr>
<td>kurtosis_last</td>
<td>1</td>
<td>Kurtosis of intensity</td>
<td></td>
</tr>
<tr>
<td>pzabovzmean</td>
<td>1</td>
<td>Percentage of points above mean</td>
<td></td>
</tr>
<tr>
<td>skewness_first</td>
<td>1</td>
<td>Skewness of first returns</td>
<td></td>
</tr>
<tr>
<td>skewness_last</td>
<td>1</td>
<td>Skewness of intensity</td>
<td></td>
</tr>
<tr>
<td>zentropy</td>
<td>1</td>
<td>Entropy</td>
<td></td>
</tr>
<tr>
<td>zmax</td>
<td>1</td>
<td>Maximum height value</td>
<td></td>
</tr>
<tr>
<td>zmean</td>
<td>1</td>
<td>Maximum height value</td>
<td></td>
</tr>
<tr>
<td>$z_{pcumn}$ $</td>
<td>n \in {1,2...9}$</td>
<td>9</td>
<td>Density</td>
</tr>
<tr>
<td>$z_{qn}$ $</td>
<td>n \in {10, 15...95}$</td>
<td>19</td>
<td>Percentile</td>
</tr>
<tr>
<td>$zsd$</td>
<td>1</td>
<td>Standard deviation height value</td>
<td></td>
</tr>
<tr>
<td>ikurt</td>
<td>1</td>
<td>Kurtosis of intensity</td>
<td></td>
</tr>
<tr>
<td>imax</td>
<td>1</td>
<td>Maximum intensity</td>
<td></td>
</tr>
<tr>
<td>imean</td>
<td>1</td>
<td>Mean intensity</td>
<td></td>
</tr>
<tr>
<td>$ipcumzqn$ $</td>
<td>n \in {10, 30...90}$</td>
<td>5</td>
<td>Cumulative intensity proportion</td>
</tr>
<tr>
<td>$isd$</td>
<td>1</td>
<td>Standard deviation of intensity</td>
<td></td>
</tr>
<tr>
<td>iskew</td>
<td>1</td>
<td>Skewness of intensity</td>
<td></td>
</tr>
<tr>
<td>itot</td>
<td>1</td>
<td>Total intensity</td>
<td></td>
</tr>
<tr>
<td>gap_F</td>
<td>1</td>
<td>Gap fraction</td>
<td></td>
</tr>
<tr>
<td>rumple_index</td>
<td>1</td>
<td>Rumple index</td>
<td></td>
</tr>
<tr>
<td>asp</td>
<td>1</td>
<td>Aspect of plot</td>
<td>Acquisition geometry</td>
</tr>
<tr>
<td>az</td>
<td>1</td>
<td>Azimuth of acquisition</td>
<td></td>
</tr>
<tr>
<td>dist</td>
<td>1</td>
<td>Distance of scanner</td>
<td></td>
</tr>
<tr>
<td>ele</td>
<td>1</td>
<td>Elevation of plot</td>
<td></td>
</tr>
<tr>
<td>meanang</td>
<td>1</td>
<td>Mean scan angle</td>
<td></td>
</tr>
<tr>
<td>slope</td>
<td>1</td>
<td>Slope of plot</td>
<td></td>
</tr>
</tbody>
</table>

**Total** 61
corresponding lidar metrics were split into training and test sets. The training set was further subdivided into training (191 field plots or samples) and validation set (50 field plots or samples), with roughly an 80:20 ratio. The test set (50 field plots or samples) was completely blind to the training and validation sets. When the point clouds were considered per plot and flight line, there were 1095 samples in total in this expanded dataset, while the number of field plots was still the same (291). We used the field plot ids from the test set of the standard dataset to create the corresponding test (and training) set(s) of the expanded datasets to ensure that point clouds for the same sample field plots were not present in both the test and training sets. However, the training set was randomly divided into an 80:20 ratio. The cross-validation scheme is illustrated in Fig 4. Thirty different splits containing training, validation, and test data were generated for all the datasets.

The benefit of the data expansion strategy was tested via three experiments:

1. The standard dataset (std) comprising only the lidar metrics was used to build a model and then compared with a model built with the expanded dataset (exp) comprising only the lidar metrics.
2. The standard dataset was appended with terrain variables (slope, azimuth, and elevation) (std_terrain) to build a model and compared to a model built with the

---

**Fig. 3.** Illustration of lidar scanning along the slope and against the slope. (a) top view; (b) side view; (c) example point clouds with similar mean scan angles 26° (yellow) and 22° (green); (d) example scan geometry with relevant parameters.

**Fig. 4.** Illustration of the cross-validation scheme for standard and expanded datasets. The process was repeated 30 times (30 splits).
expanded dataset also appended with terrain variables (exp\_terrain).

3. The expanded dataset was appended with both terrain and scan geometry variables (exp\_terrain+scan) and compared to both the standard dataset appended with terrain variables (scan geometry variables not applicable for this data set) and the expanded dataset appended with only terrain variables (exp\_terrain).

Indeed, scan geometry features are not available along with standard lidar metrics as, in this case, the point cloud results from a merging of scanning configurations. The workflow used in the study is illustrated in Fig 5.

**F. Regression models**

We used the TensorFlow (2.6.0) library in Python (3.9.7) for the fully connected multilayer perceptron (MLP) [35]. The MLP network consisted of two hidden layers. Each neuron in a layer is fully connected (FC) to all the neurons in the following layer. The components of the designed MLP include the input layer, two hidden layers, and an output layer. The rectified linear units (ReLU) function was used as the activation function. It defines how the input values it receives are output to the next neuron. A dropout rate of 0.3 was used to regularise the network to prevent overfitting. The adaptive moment estimation (ADAM) optimiser was used to optimise the network. The number of trees built was 500, and mtry value was set to default, i.e. the number of independent variables divided by 3. The model was implemented using the randomForest package in R [37].

**G. Model accuracy assessment**

The goodness-of-fit of the MLP and RF models was assessed using the determination coefficient (R²), the Root Mean Squared Error (RMSE), the relative Root Mean Squared Error (rRMSE), and the Mean Percentage Error (MPE). The formulae for these measures are as follows:

\[
R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} 
\]

\[
\text{RMSE} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} 
\]

\[
r\text{RMSE} = \frac{\text{RMSE}}{\bar{y}} \times 100 
\]

\[
\text{MAE} = \frac{\sum|y_i - \hat{y}_i|}{n} 
\]

Depending on the number of flight lines that scanned a given plot, there could be multiple predictions per plot for models built with the three kinds of expanded datasets (exp, exp\_terrain, and exp\_terrain+scan). The median value was considered for computing the goodness-of-fit criteria.

**H. Variable importance**

Notwithstanding the benefits of ANNs, one of the challenges is that they are considered ‘black boxes’ with no clear indication of how they use the data provided to them to result in predictions. ANN methods are not ideal when interpreting the models is required. Lundberg and Lee (2017) proposed the SHAP (Shapely Additive exPlanations) values to identify the feature importance in predictions based on the Shapely values.
SHAP values are calculated by measuring the impact of a given variable in various combinations with other dataset variables. It is based on the game theory wherein the marginal contribution of one variable (player) in the presence of other variables (players) is estimated. SHAP values were computed for the variables for each split of the data to understand how various lidar metrics contributed to our predictions. We also present the variable importance output (ΔincMSE) provided by the Random Forest models. The mean values across thirty splits were reported.

### III. Results

#### A. Hyperparameter tuning

Tuning for each dataset resulted in different hyperparameters (Table IV). For the standard datasets, there were 256 neurons, while for the expanded datasets, the tuning resulted in 1024 neurons in the first hidden layer. For the second hidden layer, the number of neurons varied between 32, 64, and 128. The learning rate was either 0.01 (std) or 0.001 (exp).

#### B. Model performances

The $R^2$, MAE, RMSE, and rRMSE are presented in Table V for both MLP and RF models. The observed and predicted values for $BA$ and $V_{tot}$ are shown in Fig 6 and Fig 7, comparing the results of the MLP ABA models built with a) std and exp datasets, b) std$_{terrain}$ and both exp$_{terrain}$ and exp$_{terrain+scan}$, respectively.

In Fig 6 and Fig 7, the regression lines reveal biases in the predictions with the MLP, to different degrees for all the datasets. The plots with higher values of BA and $V_{tot}$ have underestimated predictions, and those with lower values have slight overestimations, especially for BA. In all the cases, the MLP systematically outperformed RF. The RF $R^2$ was lower by 19% ($BA$ and $V_{tot}$, std). All three error measurements, i.e., MPE, RMSE and rRMSE%, were higher for RF than MLP. For example, the rRMSE values with exp$_{terrain+scan}$ datasets were higher for RF by approximately 60% for both $BA$ and $V_{tot}$.

#### TABLE IV

Summary of the tuned hyperparameters for different experiments (neurons in the first hidden layer, neurons in the second hidden layer, learning rate). Note: ‘Metrics’ refers to lidar metrics only

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Basal area</th>
<th>Total volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>std</td>
<td>(256, 32, 0.01)</td>
<td>(256, 64, 0.01)</td>
</tr>
<tr>
<td>std$_{terrain}$</td>
<td>(256, 64, 0.01)</td>
<td>(256, 32, 0.01)</td>
</tr>
<tr>
<td>exp</td>
<td>(1024, 32, 0.001)</td>
<td>(1024, 64, 0.001)</td>
</tr>
<tr>
<td>exp$_{terrain}$</td>
<td>(1024, 128, 0.001)</td>
<td>(1024, 128, 0.001)</td>
</tr>
<tr>
<td>exp$_{terrain+scan}$</td>
<td>(1024, 128, 0.001)</td>
<td>(1024, 128, 0.001)</td>
</tr>
</tbody>
</table>

Regarding the data sets used, the trends were broadly similar for MLP and RF, but the rate of improvement with additional variables was higher for the MLP. The lowest model accuracies were observed for the model built with the std datasets ($BA$: $R^2$=0.66 and 0.53, rRMSE=30.5% and 35.8%; $V_{tot}$: $R^2$=0.71 and 0.57, rRMSE=34.4% and 35.8%, for MLP and RF, respectively). The exp datasets demonstrated relative improvements for the MLP ($BA$: $R^2$=0.76; rRMSE=26%; $V_{tot}$: $R^2$=0.78, rRMSE=30.5%) with 15% and 10% increase in $R^2$ and 15% and 11% percentage points reduction in the rRMSE for BA and $V_{tot}$, respectively. Incorporating terrain properties (std$_{terrain}$) resulted in better models with both MLP and RF. However, in contrast to the MLP, RF models only marginally benefited from the data expansion (exp datasets). BA predictions improved marginally ($R^2$=0.54 for exp vs 0.53 for std), while $V_{tot}$ predictions deteriorated ($R^2$=0.55 for exp vs 0.57 for std). RF models with exp$_{terrain}$ were better than those with std$_{terrain}$ for both $BA$ ($R^2$=0.58 for std$_{terrain}$ vs 0.61 for exp$_{terrain}$) and $V_{tot}$ ($R^2$=0.62 for std$_{terrain}$ vs 0.64 for exp$_{terrain}$) with increases of 5% and 3% in $R^2$ values, respectively. Their errors were reduced in the range of 3%-5%.

#### TABLE V

Compilation of the goodness-of-fit criteria for all the experiments. The best results for each model, i.e., MLP and RF, are highlighted in bold. The results of the best models are underlined.

<table>
<thead>
<tr>
<th>Forest attribute</th>
<th>Dataset</th>
<th>$R^2$ (MP)</th>
<th>MPE (MP)</th>
<th>RMSE (MP)</th>
<th>rRMSE (MP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLP</td>
<td>RF</td>
<td>MLP</td>
<td>RF</td>
<td>MLP</td>
</tr>
<tr>
<td>Basal area</td>
<td>std</td>
<td>0.66</td>
<td>0.53</td>
<td>6.7</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>std$_{terrain}$</td>
<td>0.77</td>
<td>0.58</td>
<td>5.4</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>exp</td>
<td>0.76</td>
<td>0.54</td>
<td>5.7</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>exp$_{terrain}$</td>
<td>0.81</td>
<td>0.61</td>
<td>5.0</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>exp$_{terrain+scan}$</td>
<td><strong>0.83</strong></td>
<td><strong>0.60</strong></td>
<td><strong>4.7</strong></td>
<td><strong>7.2</strong></td>
</tr>
<tr>
<td>Total volume</td>
<td>std</td>
<td>0.71</td>
<td>0.57</td>
<td>78.2</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td>std$_{terrain}$</td>
<td>0.77</td>
<td>0.62</td>
<td>71.7</td>
<td>89.2</td>
</tr>
<tr>
<td></td>
<td>exp</td>
<td>0.78</td>
<td>0.55</td>
<td>68.2</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td>exp$_{terrain}$</td>
<td>0.83</td>
<td>0.64</td>
<td>57.6</td>
<td>86.5</td>
</tr>
<tr>
<td></td>
<td>exp$_{terrain+scan}$</td>
<td><strong>0.85</strong></td>
<td><strong>0.64</strong></td>
<td><strong>54.6</strong></td>
<td><strong>86.2</strong></td>
</tr>
</tbody>
</table>
The MLP models built with std\textsubscript{terrain} resulted in better goodness-of-fit values (BA: $R^2=0.77$, rRMSE=25%; $V_{tot}$: $R^2=0.77$, rRMSE=32.2%) than those built with std datasets (BA: $R^2=0.66$, rRMSE=30.5%; $V_{tot}$: $R^2=0.71$, rRMSE=42.2%). The goodness-of-fit of models built with exp\textsubscript{terrain} datasets were better for both BA and $V_{tot}$ (BA: $R^2=0.81$, rRMSE=25%; $V_{tot}$: $R^2=0.83$, rRMSE=32.2%) with increases of 5% and 8% respectively. The three error goodness-of-fit criteria (MAE, RMSE, and rRMSE) reduced in the 7% to 15% range and around 19% for BA and $V_{tot}$, respectively. Incorporating additional information about the scan geometry (exp\textsubscript{terrain}+scan) resulted in slightly better MLP models with 3% higher $R^2$ values and lower errors in the 5%–8% range for both BA and $V_{tot}$. The RF models with exp\textsubscript{terrain}+scan datasets did not result in any improvements (less than a 1% reduction in errors). In addition, in Fig. 7, the saturation problem of underestimating large values is evident. It appears to be well-handled by the MLPs when combined with terrain and scan information (Fig. 7).

C. Variable importance

In MLP predictions, gap fraction and elevation were the two most important variables in predicting BA and $V_{tot}$. The distance metric was also ranked third in both instances, indicating its influence in the prediction with exp\textsubscript{terrain} datasets. The slope of the plot was the other acquisition geometry variable that was present among the top ten variables. The intensity metrics itot and imean were also present among the top ten variables in BA prediction while only itot was present among the top ten variables for $V_{tot}$ prediction. Several height-based metrics such as the density metrics (zpcum3, zpcum5 and zpcum7 and zpcum6) were present among the top fifteen variables. The second canopy metric, i.e. rumple index, was ranked only at the 25th and 22nd place for Basal area and Total volume predictions, respectively and the mean scan angle was ranked at the 31th and 30th place.

In comparison, for the RF models of BA and $V_{tot}$, slope and aspect were ranked within the top five variables. Gap fraction and plot elevation were also highly ranked. Compared to the MLP variable importance, the distance variable was less
Fig. 7. Multilayer perceptron scatterplots of predicted and observed values for models built with \textit{std} and \textit{exp} datasets with lidar metrics and terrain properties (left and middle panels) and with expanded datasets with lidar metrics, terrain and scan properties (right panels) for \textit{BA} (top panels) and \textit{Vtot} (bottom panels). \textit{Std} datasets comprise metrics computed from standard lidar point clouds per plot obtained from one or many flight lines. \textit{Exp} datasets comprise metrics computed by considering lidar point clouds obtained from each flight line as independent observations. Multiple point clouds per plot in expanded datasets are not to be confused with multiple \textit{Y} values for the same \textit{X} values in this figure as this indicates the repeated occurrence of a plot in the tests sets in different splits of data.

critical, and the rumple index (the second canopy variable) and the gap fraction were ranked better. The height-based density metrics were relatively lower ranked, with only zpcum8 and zpcum9 in the top ten variables. Some metrics found important in the MLP predictions, such as itot, imean and p2th were of lower importance in RF predictions. The traditional metrics such as zmax, zmean or zq95 were found to be more important metrics.

IV. DISCUSSION

A. Model tuning

Hyparameter tuning was a crucial step in this study. We rebuilt the models with \textit{std} and \textit{std\_terrain} \cite{30} and observed that the goodness-of-fit criteria were higher for models with both \textit{std} and \textit{std\_terrain} (\textit{BA}: $R^2$ of 0.66 and 0.71; \textit{Vtot} : 0.71 and 0.77) compared to \cite{30} (\textit{BA}: $R^2$ of 0.61 and 0.69; \textit{Vtot}: 0.67 and 0.74). This could be due to variations in hyperparameter tuning resulting in better models and underlining the importance of the tuning process to obtain better models.

The random initialisations of the algorithms used in the models may yield varying hyperparameters. In our experiments, we observed the learning rate as an essential parameter and often tuned to 0.001 for the expanded datasets. In contrast, the learning rate was set to 10-5 based on expert knowledge and was not included in the parameter optimisation step in \cite{30}, which could explain the differences in model performances.

B. Multiangle perspectives of the same vegetation

A prevalent problem regarding saturation was also observed in this study with \textit{std} datasets for both basal area (\textit{BA}) and total volume (\textit{Vtot}). The saturation issue was handled well by a deep-learning-based fusion strategy using lidar and optical (Sentinel-2) datasets \cite{30}. In this study, however, the saturation effects appear less apparent for models built with expanded datasets consisting of terrain properties (\textit{exp\_terrain} and \textit{exp\_terrain+scan} datasets). The changes in lidar point cloud properties due to different acquisitions may convey information on species composition at the plot level, as done by the Sentinel-2 time series \cite{30}.

The models with expanded datasets consistently outperformed those with corresponding standard datasets. Between \textit{exp\_terrain} and \textit{exp\_terrain+scan}, the $R^2$ values improved for MLP models from 0.81 to 0.83 and 0.83 to 0.85 for \textit{BA} and \textit{Vtot}, respectively. It appears that the improvement was not of a large magnitude. However, it is worth noting that all error measurements decreased (rRMSE reduced by 1.4 and 2 percentage points for \textit{BA} and \textit{Vtot}) and that the saturation effect
was better handled with acquisition geometry variables. On the other hand, RF could not provide comparable results (Fig 9). The goodness of fit criteria for RF did not change between the two data sets, which could be attributable to the fact that there may be some redundancy in the information offered by the point clouds that were considered independent observations. Figs 3a and 3b illustrate the differences in point clouds due to slope, even if the scan parameters are nearly similar due to the steep slope. Irrespective of the variations due to differences in scan angle, the slope directly affects the point cloud and the resulting lidar metrics (see Fig 3c). The MLP could learn this complex and nonlinear relationship in this instance since, in the expanded datasets, we retained both heterogeneity (metrics sensitive to acquisition properties) and homogeneity (metrics not sensitive to acquisition properties) in lidar metrics.

As the addition of scan geometry demonstrated improvements with the MLP, a qualitative assessment of the scatterplots in Fig 7 reveals that a model built with the exp_terrain+scan dataset could deal with the issue of saturation commonly observed with large values. The scatterplots are comparable to those obtained by [30] after implementing a fusion of lidar and optical information, which creates interesting possibilities for future studies. Furthermore, while the MLP outperformed the RF models, some of the relative improvements across the datasets, though marginal, were apparent even in the RF models.

C. Interpreting SHAP values

Regarding the SHAP values (Fig. 8), the improvement in the ABA predictions corresponds to introducing new variables across the datasets. ANNs benefitted from including terrain variables (slope, aspect and elevation), yielding improved results [30]. The SHAP values for the elevation metric, which were among the highest, confirm that the terrain-related properties are crucial, especially in highly varying terrain. This is supported by the fact that slope and aspect also figured in the top thirty variables. Similarly, the distance of the scanner (considered an acquisition-related metric in this study) was among the top three metrics. Based on its importance, it is apparent that it was mainly responsible for resulting in improvements between exp_terrain and exp_terrain+scan datasets. The mean scan angle was ranked lower at around 30 (out of 61), whereas the azimuth of the acquisition was of lower importance. It is, however, crucial to note the improvements in ABA predictions on account of the data expansion. The rRMSE values for the exp_terrain datasets were lower by 9% for BA and 20% for Vtot compared to the std_terrain dataset. Metrics, such as the gap fraction, provided unique and diverse perspectives of the same point cloud to the NN. Although in this study we did not implement range normalisation of the intensity, variables such as itot and imean were ranked high based on their SHAP values (Fig 8). In ABA models, virtually infinite metrics are being (or, have been) developed to capture vegetation properties comprehensively. SHAP values could be useful to inform this process.

On comparing with the ranking of the variables used in RF models (exp_terrain+scan) (Fig. 9), it is apparent that metrics such as elevation, gap-fraction and terrain properties figure among the top-ranked variables (the distance, i.e. dist, of the scanner was ranked lower). However, this did not translate into better predictions, thereby demonstrating the capability of neural networks to utilise the additional geometry information provided via the expansion of ABA datasets or additional metrics or both.

D. Potential of different modelling strategies

Modelling strategies certainly influence the results. A few studies have explored different deep-learning methods to predict forest attributes from lidar data. [40] used an MLP.
architecture with the principal components of a set of metrics similar to our study. They observed an rRMSE of 22.5% for the predictions of BA in heterogeneous tropical forests, but with gentle or no relief. In our study, despite the mountainous relief, which is known to add issues in ABA modelling, the best-performing model was the exp terrain+scan dataset with an rRMSE of 19.9%. However, [40] did not consider metrics such as the gap fraction with proven explanatory power for forest structure characterisation. Additionally, the gap fraction and the rumple index (used in this study) are metrics sensitive to lidar scan angle [33], [41]. The data expansion strategy may have benefited from additional information from these two metrics, among other sensitive metrics. In contrast, RF models were unable to utilise the additional information as is apparent in the saturation of large values with exp terrain+scan datasets (Fig. 10).

The choice of lidar metrics is also a part of the modelling strategy. We used the intensity information provided by the data provider without implementing a normalisation step to enhance it, as demonstrated in different studies [40], [42], [43], [44]. [13] showed that intensity information is more effective than height-based metrics in discriminating tree species. The forests in PNR Bauges comprise forest plots of broadleaved, coniferous, and mixed types of forests with different tree species. Calibrated intensity information could further help in improving the accuracies of the models. In addition, the intensity information is also known to be affected by the scan angle [40], thus potentially providing multiple perspectives of the vegetation.

[27] observed lower rRMSE values of 14.5% in volume predictions in predominantly Eucalyptus, and Chinese-fir-dominated stands. In comparison, an rRMSE of 24% was observed in this study. Even if they used a more advanced modelling framework that combines a fully-connected neural network and an optimised radial basis neural network, the result difference is also likely to be linked to relatively simpler forest stands under study. In addition, studies using other modelling methods, such as OLS or RF, reported rRMSE values of basal area (BA) and volume ($V_{\text{tot}}$) predictions were in the range of 23% to 29% and 22% to 34%, respectively [19]. In this study, we observed an rRMSE for BA and $V_{\text{tot}}$ of 19.9% and 24% for a model built with exp terrain+scan datasets.

The data expansion strategy may have benefited from additional information from these two metrics, among other sensitive metrics. In contrast, RF models were unable to utilise the additional information as is apparent in the saturation of large values with exp terrain+scan datasets (Fig. 10). The choice of lidar metrics is also a part of the modelling strategy. We used the intensity information provided by the data provider without implementing a normalisation step to enhance it, as demonstrated in different studies [40], [42], [43]. [44] showed that intensity information is more effective than height-based metrics in discriminating tree species. The forests in PNR Bauges comprise forest plots of broadleaved, coniferous, and mixed types of forests with different tree species. Calibrated intensity information could further help in improving the accuracies of the models. In addition, the intensity information is also known to be affected by the scan angle [40], thus potentially providing multiple perspectives of the vegetation.

[13] demonstrated the benefits of using voxel-based metrics. Voxel-based metrics contributed by reducing rRMSE values for ordinary least squares regression (OLS) ABA models based on forest type (riparian, broadleaf, coniferous, and mixed). For OLS ABA models, using voxel-based metrics improved the predictions by reducing the rRMSE by 10%, 4%, and 14% for riparian, broadleaf, and mixed types, respectively. The rRMSE values for different forest types and with only four lidar metrics (including voxel-based metrics) ranged between 30%–40%, comparable to those observed with MLP models built with std, std terrain, and exp datasets in this study. Therefore, including normalised intensity information along with stand-type characteristics and voxel-based metrics could be beneficial for building more accurate models using deep-learning approaches and a possible future area of exploration.
E. Impact of data sample characteristics

The distribution of BA and $V_{tot}$ across different test sets reveals that the dataset was not balanced, as there are very few plot measurements with very high values. Many of the cross-validation splits used in the study suffered from a lack of balanced training and testing samples. As a result, for all models, the impact of a few plots with high amounts of wood resource, i.e. $BA$ greater than 70 m²/ha and $V_{tot}$ greater than 800 m³/ha, can be observed. Such stand types are scarce, and the predictions for those plots might be outside the range of the training set values for some models and explain the saturation effect. MLP models are known for their capacity to generalise [27], [28], which could also explain why they performed better than RF. We believe that field plot measurement representing diverse forest stands will further help build robust models. In this study, the sampling strategy used to collect field measurements followed a systematic sampling scheme to establish sites for periodic monitoring. A stratified sampling scheme would be more suitable for building models. Moreover, some differences may arise because small trees (DBH<7.5cm) were not measured as per the field measurements protocol. The contribution of such trees in estimating the signal could be significant.

In this study, we opted to describe our datasets based on the flight lines as ‘expanded datasets.’ The common practice combines all these observations in lidar ABA models to generate the ‘standard’ datasets. Essentially, each observation in the ‘expanded’ dataset represents an independent physical observation (or lidar scanning) of a given field plot. As the vegetation in the field plots is rarely identical when viewed from multiple directions, it may be argued that the resulting differences from different scans are comparable to the data augmentation procedure that is commonly used to increase the number of samples when dealing with images [45]–[48]. To avoid confusion with commonly practised data augmentation strategies, we referred to our modified datasets as ‘expanded’ datasets. Nonetheless, we would like to emphasise the similarities.

V. CONCLUSION

This study demonstrated that considering point clouds from different flight lines as independent observations in non-parametric models can improve ABA predictions for forest attributes. By considering the point clouds as independent observations, we retained the heterogeneity in the lidar metrics due to variations in the acquisition geometry in the form of an expanded dataset with a significantly higher number of observations than a standard dataset. A multilayer perceptron (MLP) could harness the expanded information to predict forest attributes in a complex forest environment with higher accuracies than a Random Forest model, commonly used in ABA approaches. The present MLP model also demonstrated the potential to result in predictions comparable to methods involving optical and lidar data fusion. Optical data may be incorporated to improve further the results observed in this study.

REFERENCES

An earth observation methodology to estimate age, basal area, and wood volume using LiDAR data in heterogeneous forest stands in Central Europe.  

Denis Cosenza et al., 2021.  


