Article

Above-Ground Biomass Prediction for Croplands at a Sub-Meter Resolution Using UAV–LiDAR and Machine Learning Methods

Jaime C. Revenga 1,* Katerina Trepekli 1☯, Stefan Oehmcke 1,2, Rasmus Jensen 1, Lei Li 2, Christian Igel 2☯, Fabian Cristian Gieseke 2,3 and Thomas Friborg 1☯

1 Department of Geosciences and Natural Resources Management (IGN), Copenhagen University, DK-1350 Copenhagen, Denmark
2 Department of Computer Science (DiKU), Copenhagen University, DK-2100 Copenhagen, Denmark
3 Department of Information Systems, University of Münster, 48149 Münster, Germany
* Correspondence: jar@ign.ku.dk

Abstract: Current endeavors to enhance the accuracy of in situ above-ground biomass (AGB) prediction for croplands rely on close-range monitoring surveys that use unstaffed aerial vehicles (UAVs) and mounted sensors. In precision agriculture, light detection and ranging (LiDAR) technologies are currently used to monitor crop growth, plant phenotyping, and biomass dynamics at the ecosystem scale. In this study, we utilized a UAV–LiDAR sensor to monitor two crop fields and a set of machine learning (ML) methods to predict real-time AGB over two consecutive years in the region of Mid-Jutland, Denmark. During each crop growing period, UAV surveys were conducted in parallel with AGB destructive sampling every 7–15 days, the AGB samples from which were used as the ground truth data. We evaluated the ability of the ML models to estimate the real-time values of AGB at a sub-meter resolution (0.17–0.52 m²). An extremely randomized trees (ERT) regressor was selected for the regression analysis, based on its predictive performance for the first year’s growing season. The model was retrained using previously identified hyperparameters to predict the AGB of the crops in the second year. The ERT performed AGB estimation using height and reflectance metrics from LiDAR-derived point cloud data and achieved a prediction performance of $R^2 = 0.48$ at a spatial resolution of 0.35 m². The prediction performance could be improved significantly by aggregating adjacent predictions ($R^2 = 0.71$ and $R^2 = 0.93$ at spatial resolutions of 1 m² and 2 m², respectively) as they ultimately converged to the reference biomass values because any individual errors averaged out. The AGB prediction results were examined as function of predictor type, training set size, sampling resolution, phenology, and canopy density. The results demonstrated that when combined with ML regression methods, the UAV–LiDAR method could be used to provide accurate real-time AGB prediction for crop fields at a high resolution, thereby providing a way to map their biochemical constituents.

Keywords: UAV–LiDAR; machine learning; above-ground biomass; precision agriculture

1. Introduction

Non-destructive crop surveying techniques are currently being developed to aid decision-making in precision agriculture [1]. Developing accurate crop monitoring methods is an important step toward improving climate change adaptation strategies within agroecosystem management and food production systems [2]. It is expected that the results of this study could have particular impacts on the following two research areas: (i) food security assessment [3] in the face of global land use [4] and climate changes [5–7]; (ii) crop yield modeling [8–10] and nutrient cycle studies [11,12].

The expansion of field studies and crop monitoring systems is therefore strongly encouraged by both the modeling community (to test and calibrate models) and the culti-
vation community (to enable adaptive harvesting operations and the real-time monitoring of nutritional constituents) [13].

Above-ground biomass (AGB) is a fundamental agronomic parameter that accounts for the content of dry biomass in the standing part of plants. It is an indicator for the phenological and health status of plants, as well as their nutrient contents [14]. As such, accurate AGB mapping is a crucial component for a wide range of scientific disciplines, including precision agriculture and agroecology [15]. Most established techniques for acquiring remote sensing (RS)-based estimates of AGB primarily rely on (i) passive optical and (ii) synthetic aperture radar (SAR) methods. However, these techniques present limitations. When SAR signals become saturated with increasing AGB [16,17], its efficacy in AGB mapping is sensitive to polarization and the combination of available bands [18]. Furthermore, it is primarily suited to mapping highly vegetated areas, such as woodlands and forests [19–21], and it has been shown to overestimate AGB in areas with low biomass levels [22,23]. Passive optical methods have been extensively applied in precision agriculture [24,25]. These methods are bound to two-dimensional features that are derived from various parameters, such as canopy structure [26], image texture [27,28], and greenness indices [24,29], which also suffer from substantial signal saturation and limitations that are linked to spectral resolution and radiometric sensor calibration [24,25]. In comparison to SAR and passive optical methods, regression models that are derived from light detection and ranging (LiDAR) data consistently achieve better results [30,31] and have become the preferred option when precise spatial AGB distributions are required [21].

LiDAR is an active sensing technique that is based on pulses of light. The reflections of these pulses are detected by the same device, which allows the precise location of intercepted objects to be determined based on the laser pulse travel time [32]. Variations in illumination and other atmospheric conditions do not affect the resulting data as much as the data from passive optical sensors. Progress in LiDAR technology has enabled advanced terrestrial ecosystem research. High-density LiDAR point cloud data (PCD) are currently used to investigate several variables, such as plant-level morphological traits in agroecosystems [33] and forests [34–36], as well as light–ecosystem interactions [37,38]. In fact, the integration of LiDAR systems in ecosystem monitoring at different scales (from the plot to ecosystem or regional level), combined with new modeling techniques [39–41], is broadening the field’s perspectives in an unprecedented manner [32]. The use of LiDAR systems in unstaffed aerial vehicles (UAVs) allows for the collection of PCD at sub-cm resolutions, which has enabled the development of a growing number of applications in agroecosystem monitoring [42–44]. This data collection technique, combined with machine learning (ML) regression methods, has proved to be useful for the precise AGB quantification of croplands [45–48] and forest ecosystems [49].

Advanced ML methods can attain better performances in regression tasks than the methods that are commonly used in AGB research (e.g., linear and power regression models that take the architectural and ancillary traits of plants as their inputs [50–52]). This is due to the ability of ML methods to account for more complex non-linear relationships between predictors and AGB. There is a growing interest in applying these recently developed ML methods to AGB prediction for land ecosystems. Recent state-of-the-art studies have combined ML methods and UAV-sensed data [53–55]. For example, Ma et al. [56] estimated the AGB of croplands using convolutional neural networks and close-range imagery, while Pan et al. [48] used neural networks (with so-called attention-based modules) and rover-borne LiDAR PCD as direct inputs instead of statistics. There have been many studies on AGB prediction using ML methods, ranging from sub-meter resolutions [48] to coarser resolutions [53,54,56]. However, the spatial resolution and spatial distributions of AGB predictions [55] are important factors that have strong influences on the obtained results. The effects of these two factors are usually not considered, which makes it difficult to compare the results that have been reported in the literature.

This study aimed to test the following hypotheses: discrete return UAV–LiDAR surveys are a valid technique for providing accurate estimates of crop structures in dense agricultural crops that allow for AGB regression analysis; the distribution of LiDAR PCD
features empirically relates to the patterns of AGB distribution across crop fields; the empirical relationship between (i) the patterns in the remotely sensed data and (ii) those in the AGB field-based measurements can therefore be formulated as a supervised regression problem, which can then be learnt by an ML model to predict AGB; and, therefore, it is possible to automate the accurate estimation of AGB distributions in dense crops based on UAV–LiDAR surveys and ML regression methods at high spatial resolutions.

We proposed a method for estimating the spatial distributions of AGB in crop fields using UAV–LiDAR for data acquisition and ML methods for regression analysis. Our results showed that this method allowed real-time estimates of AGB in crops to be obtained at sub-meter resolutions and was effective for two different species (barley and winter wheat) and contrasting crop structures.

2. Materials and Methods

This section presents the study area, instrumentation, field measurements, datasets, and methods that were employed in this study.

2.1. Study Area

The study area (Figure 1) was a conventionally managed cropland site, which is located near to an Integrated Carbon Observation System (ICOS) class-1 ecosystem station at Voulundgaard (DK-Vng) in Mid-Jutland, Denmark (56.037476N, 9.160709E). It has an area of ca. 13 ha and is located on the eastern part of the Skjern River catchment. The cropland site is a relatively flat plain at an altitude of 64–68 m above sea level. The terrain has smooth undulations and a slight slope to the northwest.

![Figure 1](image-url) The location of the study site (*) in Mid-Jutland (DK). The inset shows a top-down view of the field site and the surrounding area. Source: www.icos-cp.eu (accessed on: 8 August 2021) and Google Earth Engine.

The crops that were investigated were barley (*Hordeum vulgare* L.) in 2020 and 2021 and winter wheat (*Triticum aestivum* L.) in 2021. The growing period of the barley lasted from the end of April 2020 (shoot emergence) to the end of August 2020 (harvest) and followed a similar cycle in 2021. In 2021, the growing period of the winter wheat extended from January 2021 (shoot emergence) to the end of August 2021 (harvest). The conventional agricultural practice at the site included the application of fertilizers in the form of slurry (according to local regulations [57]), pesticides, and fungicides throughout the growing season, as well as sufficient irrigation to prevent water stress [58]. The region has a humid temperate climate that is characterized by a mean annual precipitation of 961.0 mm, a mean annual temperature of 8.1 °C, and an overcast or scattered cloud cover (with a mean annual incoming shortwave radiation of 108 W/m²). The ploughing layer (30 cm deep) sits on a sandy loam with pebble inclusions that are around 3–5 cm in diameter. The water table depth lies at 5.5 ± 1 m below ground. A visual documentation of the two crops is shown in Figure 2c,d.
2.2. Instrumentation, Flight Parameters, and Field Measurements

The UAV–LiDAR system that was utilized (i.e., LS Nano M8, LidarSwiss GmbH, CH) was equipped with three built-in devices: (i) a Global Navigation Satellite System (GNSS) receiver; (ii) an inertial measurement unit (IMU) with a mono-antenna navigation system; and (iii) a discrete LiDAR scanner (M8 sensor, Quanergy Systems, Inc., Sunnyvale, CA, USA). The overall setup allowed for synchronization between the three data streams so as to produce a georeferenced point cloud dataset output after processing, as described in [59]. The system was mounted to a commercial DJI Matrice 600 Pro payload (Figure 3) at a 90° pitch angle and the same heading and roll angles as the UAV platform. The laser-firing system consisted of a stack of eight lasers, which emitted discrete near-infrared pulses (at 905 nm) in sequence at a constant rate of 53 kHz, with an individual beam spacing (along the direction of the UAV movement) of 3.20°. The beam divergence was 3 mrad, which provided a footprint diameter increase of 3 cm every 10 m along the propagation direction. However, the beam energy was spatially Gaussian distributed [60], which resulted in smaller active laser footprints. During data collection, a differential Global Positioning System (dGPS, Trimble R8) was set up as the receiver base station in real-time kinematic (RTK) positioning mode, which logged real-time corrected satellite coverage data during the UAV–LiDAR survey.
After running a number of tests with this instrumentation and considering our scanner specifications with reference to the available literature [59], the following LiDAR flight parameters were considered for this study: a flight height of 40 m above ground level; a horizontal speed of 5 m/s; a vertical field of view (FoV) of 21.5°; a horizontal FoV of 180°; a range gate of 2–200 m; LiDAR scan swaths of 50 m wide; a side overlap between individual LiDAR swaths of 20%; and a scan angle of 0–32°. These parameters provided a balanced trade-off between LiDAR penetration through the canopy (i.e., PCD porosity), gap fraction, point cloud density (ca. 200 pp/m$^2$), and field coverage. The UAV–LiDAR surveys followed the calendar and frequency of the AGB sampling campaigns (Figure 2a,b), which resulted in a maximum time difference between the UAV–LiDAR surveys and AGB collection of 48 h in case of adverse weather conditions.

Immediately after the UAV–LiDAR surveys, the AGB sampling was carried out. At the study site, scattered georeferenced locations were selected to conduct the AGB collection. Several sampling campaigns were conducted each year (six in 2020 and seven in 2021). With regard to the spatial distribution of the AGB sampling locations (Figure 4), the two major campaigns (i.e., on 22 June 2020 and on 14 July 2021) utilized randomized sampling, while the rest (i.e., minor campaigns) were conducted in areas that showed greater variability. After the AGB collection, each sample was oven-dried for 72 h at 65 °C and weighed according to the reference protocols [61]. This procedure was replicated for every sampling campaign over both growing seasons. While the procedure that was followed in 2020 was used as the baseline reference, this AGB collection method was adapted for the 2021 AGB campaigns in order to enlarge the AGB reference labeled datasets (Table 1). Further details about the datasets are presented in Section 2.4.
2.3. LiDAR Data Processing

The steps in the laser data processing pipeline could be divided into two parts: (i) point cloud scene generation and (ii) point cloud scene processing. Lastly, a post-processing step was included to add the phenological stage information from each of the datasets.

2.3.1. Point Cloud Scene Generation

The raw positioning file (which was logged by the internal IMU and GNSS during the survey) was coupled with the satellite coverage data to produce a processed trajectory file using NAVsolve software (Oxford Technical Solutions, Bicester, UK). We stored the 3D position and UAV attitude data (i.e., longitude, latitude, ellipsoidal height, and orientation angles) at a 10 Hz sampling resolution. Then, in order to filter out LiDAR data that were compromised due to adverse circumstances (e.g., heading deviations that were too large or disturbances in cruise speed that were caused by wind gusts), the UAV trajectory was used to mask out the raw laser files following a quality criterion ($\Delta$heading $> 25^\circ$ was taken as the threshold for the rejection of raw laser data). Finally, the processed laser files were output in parallel swaths, according to the UAV–LiDAR survey trajectory plan. In this way, several parallel swaths produced the full PCD scene (Figure 5).

![Figure 5](image)

**Figure 5.** The point cloud data (PCD) scenes (the crops are shown at the maturity stage and the PCD scenes are colored by elevation): (a) the barley field in 2020; (b) the wheat field in 2021. In both (a,b), the upper panels show the cross-sectional views of the PCD, with a buffer depth of 0.5 m. The x (−), y (−), and z (−) axes indicate easting, northing, and elevation, respectively. It can be seen that there was a higher PCD porosity in (b) than in (a).

2.3.2. Point Cloud Scene Processing

At each location within the PCD scene, a PCD clipping was created that covered the same area as each individual AGB sample (i.e., 1 m $\times$ 0.35 m in 2020 and 0.5 m $\times$ 0.35 m in 2021). These clipped areas were considered separately (as shown in Figure 6a). Then, for each individual PCD clipping, the PCD height value (i.e., ellipsoidal height) was normalized by subtracting the terrain height. Finally, we extracted the statistical metrics of height and reflectance intensity from each individual PCD clipping, which were used as prediction features.

In order to select the prediction features, a set of candidate predictors was initially short-listed based on the Pearson coefficient between them and the target AGB labels. Then, iteratively, we removed the considered predictors one by one until we observed a drop in the regression performance using the validation set. Following this process, the PCD features that were finally selected as the predictors were the following:

- The metrics of height were the mean, median, standard deviation, variance, skewness, and kurtosis;
• The metrics of reflectance were the mean and standard deviation.

Figure 6. The generation of the augmented datasets: (a) the partitioning of the individual LiDAR samples into three adjacent parts; (b) the three original adjacent AGB samples (individual size: 0.175 m²); (c) the shaded area represents each augmented sample (size: 0.35 m²–0.52 m²). The two augmented datasets were produced by sampling (with replacements) the individual original samples and adding either two or three of them together to produce a new instance. Each of the instances that were produced in this way were attributed the mean value of AGB of the individual samples while the LiDAR-derived features were re-calculated by considering all of the LiDAR returns contained in the individual samples.

2.3.3. Data Post-Processing

The data were split into training and validation sets (70% and 30%, respectively). Both datasets were then subjected to stratification [62] to produce separate subgroups of instances. This processing step was inspired by the three main phenological growth stages of crops: (i) “early season”, from shoot emergence to the start of stem elongation; (ii) “rapid development”, from when crops start growing to when they reach their peak leaf area index; and (iii) “maturity”, when nutrients are allocated to grains and senescence occurs [63,64]. Both the training and validation sets were stratified according to the mean AGB value from the training distribution using K-means clustering [65,66] (Figure 7a).
Figure 7. The data processing pipeline, model training, and the evaluation of the predictions: (a) the 2020 barley dataset was split into training and validation sets. The instances in both sets were stratified according to the mean AGB values from the training distribution. A cross-validated grid search was conducted to optimize the hyperparameters for model selection; (b) the ML model with the best performance was fitted to a new dataset (i.e., 70% of either the 2021 barley or wheat datasets) and the final prediction performance was evaluated using the remaining test set (i.e., 30% of the 2021 datasets).

2.4. Datasets

Following the aforementioned procedure, three original datasets were produced, which contained AGB- and LiDAR-derived prediction features. There was one dataset of original samples that corresponded to the year 2020 (barley) and two original datasets that corresponded to the year 2021 (wheat and barley). Data were also collected during 2021 so as to allow for the production of two additional datasets of augmented samples (Figure 6c).

The datasets that were composed of the augmented samples were produced as follows. During the 2021 AGB collection campaigns, samples were located and collected from adjacent plots (three samples per location) along the planting line (Figure 6b). In this way, by adding either two or three original samples, a new augmented instance was obtained, whose AGB value was determined by averaging the respective AGB values from the original samples. With regard to the LiDAR-derived metrics of the augmented samples, they were re-calculated by considering all of the LiDAR returns that were contained in the original samples. The numbers of instances that were included in each dataset, as well as the sample size, are summarized in Table 1.

The augmented datasets were used to test the model’s ability to generalize predictions because (i) the augmented data samples provided a better representation of the vegetation structures of the target plots than the original datasets (i.e., as the augmented samples covered a larger area, there was a higher count of LiDAR returns per sample) and (ii) the augmented datasets had a lower sample variance among the AGB labels (because the standard error decreased when averaged out over a larger area). These two observations were consistent with the fact that the augmented datasets had fewer noisy samples and higher correlations between the AGB and the assessed predictors.
2.5. ML Model Training and Evaluation

Three different regression models were considered. Then, according to their performance using the validation set, the most suitable model was selected for testing. The Huber regressor [67,68] was selected as a representative of the linear methods with regularization. This model is a common choice in AGB research due to its robustness to outliers [69]. Furthermore, we employed extremely randomized trees (ERT) [70], which is a tree-based ensemble method. The XGBoost regressor [71] was selected as a representative of the boosting methods.

In contrast to the model parameters (e.g., the weights that were fitted during the training phase), the hyperparameters (e.g., the definition of the specific loss function to be minimized) were external to the learning model in question and were pre-defined to control the learning process [72]. Here, the hyperparameter optimization [73] of the three ML models (Table 2) was conducted using cross-validated grid searches (10 folds) in the training set. The barley20 dataset was used for training and validation. The remaining four datasets from the 2021 growing season were used for refitting (70%) and testing (30%).

Table 2. A description of the models that were evaluated. The implementations were standardized Python modules.

<table>
<thead>
<tr>
<th>Regression Model</th>
<th>Family</th>
<th>Description</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Randomized Trees (ERT)</td>
<td>Tree-Based Ensemble</td>
<td>Ensemble of decision trees (parallel setup) [70] in which the output is the average of individual predictions</td>
<td>scikit-learn</td>
</tr>
<tr>
<td>XGBoost</td>
<td>Boosting</td>
<td>Gradient boosting method that is based on stage-wise additive expansions [74,75]</td>
<td>xgboost</td>
</tr>
<tr>
<td>Huber</td>
<td>Linear</td>
<td>Regularized linear regression that is robust to outliers [67,68]</td>
<td>scikit-learn</td>
</tr>
<tr>
<td>Linear Regression (Baseline)</td>
<td>Linear</td>
<td>Ordinary least squares linear regression</td>
<td>scikit-learn</td>
</tr>
</tbody>
</table>

After hyperparameter tuning (Table 3), the model selection was conducted by comparing the regression performance of the three models using the 2020 validation set (Figure 7a), all of which were compared to a linear model (the results are shown in Section 3.1). The selected model was then re-fitted to the 70% of unseen data from 2021. The final performance evaluation was conducted using the remaining 30% of the 2021 datasets. The metrics that were used to assess the quality of predictions were the coefficient of determination ($R^2$), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE). When the reporting of the scores as average values was considered to resolve any doubts about outliers, the scores are reported with an overline (e.g., $\overline{MAPE}$), which indicated that the result is a mean value over 10 repetitions. Figure 7 shows the overall workflow of the applied method.

Table 3. The models that were evaluated and the considered hyperparameters.

<table>
<thead>
<tr>
<th>Regression Model</th>
<th>Hyperparameters</th>
<th>Total b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Randomized Trees (ERT)</td>
<td>Criterion [mae; mse], max. depth (None; 1, . . . , 9), bootstrap [True; False], max. features [log2; sqrt]</td>
<td>17</td>
</tr>
<tr>
<td>XGBoost</td>
<td>Booster [gbtree; glinear; dart], step size shrinkage (0.1, . . . , 0.5), learning rate (0.01, . . . , 0.1), L1 regularization (0, . . . , 0.5)</td>
<td>29</td>
</tr>
<tr>
<td>Huber</td>
<td>Epsilon (1.1, . . . , 1.75), alpha (5 \times 10^{-3}, . . . , 10^{-3}), fit intercept [True; False], tolerance (10^{-9}, . . . , 10^{-4})</td>
<td>6</td>
</tr>
<tr>
<td>Linear Regression (Baseline)</td>
<td>Fit intercept [True; False]</td>
<td>1</td>
</tr>
</tbody>
</table>

a The hyperparameters that were included in the cross-validation grid search for parameter selection (values in curled brackets show parameter sets and those in round brackets show the ranges of the search); b the tunable hyperparameters that were considered for each model in scikit-learn or xgboost Python libraries.
2.6. Description of the Selected ML Model

ERT is an ensemble learning technique that aggregates the results of multiple individually created decision trees to produce regression results [70]. It was originally derived from the random forest (RF) model [76]. Every individual predictor, i.e., binary decision tree, in an ERT is constructed using the whole training set. At each node, a single tree decides which split of a random subset of feature splits would reduce the reconstruction error (e.g., MAE or MSE) the most. The random sampling of features and the random splits within the feature ranges lead to diverse and less correlated decision trees. Each tree is considered to be a “weak” regressor performance-wise but in combination, they create an ensemble that can outperform individual regressors. As the final prediction (AGB in our case), the average of the individual predictions of all of the decision trees in the forest is used.

2.7. Generation of AGB Prediction Maps

The PCD scenes were subjected to a binary classification, which separated the ground from vegetation LiDAR returns. For this purpose, several filtering algorithms were evaluated [77–79]. The cloth simulation filtering algorithm [78] (implemented in R software [80]) and the slope-based filter [77] (implemented in Lidar360 software [81]) produced the best results, in contrast to morphological filters [79], which performed poorly. However, in order to ensure optimal accuracy, the segmentation was completed manually. Next, the ground returns were rasterized using the inverse distance weighting algorithm [82,83] and then extended to the whole region of interest through the interpolation of the natural neighbors [84] in order to produce a digital terrain model (DTM). The DTM was used to normalize the vegetation returns to the ground returns by subtracting the DTM height from the initial PCD scene. Based on the spatial distributions of the height and reflectance of the normalized vegetation returns, a series of LiDAR-derived prediction features were mapped at a 1 m² resolution. These prediction feature maps were used as the input for the trained regression model in order to produce the AGB prediction maps (Figure 8).

![Figure 8](image-url)
2.8. Uncertainty of AGB Field-Based Measurements

The collection of AGB data is often prone to estimation errors [85]. Therefore, accounting for the uncertainty that is introduced into AGB field-based measurements is a necessary step to avoid the underestimation of errors in AGB prediction mapping products [86]. In this study, an additional experiment was dedicated to account for the uncertainty that was introduced by the AGB sampling technique and the effects of crop sparsity on the AGB labels, which is shown in Appendix A.

3. Results
3.1. Model Selection

The extremely randomized trees (ERT), XGboost, and Huber regression models (Table 2) were trained using an optimized set of model-specific hyperparameters (Table 3) and were evaluated based on their regression performance using the validation set, according to the MAE, MAPE, RMSE, and $R^2$ metrics. The comparison of their performances using the validation set indicated that the ERT model provided the most accurate predictions across all four metrics. Figure 9 shows the predictive performances of the three considered models using both the training and validation sets. The criterion for model selection was the model performance using the validation set, for which ERT achieved the highest $R^2$ value and the lowest RMSE, MAPE, and MAE values (note that the Huber model consistently achieved the same performance as it was a deterministic linear model, hence the flat boxplot in the figure). In all four subplots in Figure 9, the horizontal blue and green lines show the performance of the linear model that was employed as the comparative baseline using the training and validation sets, respectively.

**Figure 9.** A comparison of the regression performances of the considered models using the training and validation datasets: (a) $R^2$; (b) RMSE; (c) MAPE; (d) MAE. The blue and green horizontal lines represent the performance of the linear regression baseline model using the training and validation sets, respectively. The overlined scores represent the mean values of 10 randomized executions.
Therefore, ERT was selected for the final evaluation using the testing sets. The optimized parameters that were selected for the ERT regression model were: mean square error, which was used as the criterion to evaluate the quality of a split; the maximum depth that was allowed for any individual tree was limited to seven; the sampling was without bootstrapping; and the maximum number of features to consider per individual tree was limited to $\log_2(\text{number of features})$. Finally, the number of trees was set to 1000 [70].

3.2. AGB Prediction at a Sub-Meter Resolution

The AGB predictions at a 0.35 m² spatial resolution that were produced by the ERT model using the barley and wheat testing datasets are shown in Figures 10a and 11a, respectively. In both cases, the results were compared to the linear regression model, which was fitted to the same datasets (i.e., fitting to the 70% and predicting the remaining 30% of the 2021 data). The predictive performances of this model are shown in Figures 10b and 11b for comparison.

![Figure 10](image1.png)

**Figure 10.** The AGB predictions for the barley crops at a 0.35 m² resolution that were produced by the ERT model compared to those that were produced by the linear model (baseline): (a) the regression performance of the ERT model using the testing dataset ($R^2 = 0.48$; RMSE = 207 g/m²; MAE = 162 g/m²; MAPE = 42%); (b) the regression performance of the linear model using the testing dataset ($R^2 = 0.1$; RMSE = 302 g/m²; MAE = 247 g/m²; MAPE = 34%).

![Figure 11](image2.png)

**Figure 11.** The AGB predictions for the wheat crops at a 0.35 m² resolution that were produced by the ERT model compared to those that were produced by the linear model (baseline): (a) the regression performance of the ERT model using the testing dataset ($R^2 = 0.20$; RMSE = 288 g/m²; MAE = 216 g/m²; MAPE = 23%); (b) the regression performance of the linear model (baseline) using the testing dataset ($R^2 = 0.14$; RMSE = 304 g/m²; MAE = 254 g/m²; MAPE = 33%).

3.3. Aggregated AGB Predictions

The prediction results using the testing set were subjected to a resampling technique (i.e., random sampling with replacement) and aggregated into $i \in \{1, 2, \ldots, N\}$ samples, where $i$ is the number of pairs (AGB prediction vs. AGB target value) that were aggregated at each step. As shown in Figures 12a and 13a, the mean value of the residual distribution of the barley and wheat testing datasets approached zero. There was a systematic overestimation of 2% in the wheat testing dataset (barley: 2.2 g/m²; wheat: −18.5 g/m²). The residuals were independent and distributed around zero, so they were partially canceled
out when averaged over a larger area, which caused the $R^2$ score to converge to 1 as the number of aggregate predictions increased (Figures 12b and 13b). Thus, the prediction could be improved by coarsening the spatial resolution (Table 4). This technique is known as the spatial averaging of errors [87].

By only aggregating two predictions that were randomly sampled for each testing set, the AGB results at a 0.7 m$^2$ resolution for both the barley and wheat datasets improved significantly compared to the respective AGB predictions at a 0.35 m$^2$ resolution (Table 4 and Figures 12b and 13b).

The $R^2$ score of the aggregate predictions presented a turning point between 1 m$^2$ and 2 m$^2$ of spatial resolution (i.e., three and six aggregate samples, respectively), reaching optimal predictions ($R^2 \approx 0.95$) from 3 m$^2$ onward for the barley crops (Figure 12b) and from 4 m$^2$ onward for the wheat crops (Figure 13b). At 2 m$^2$ of spatial resolution, the $R^2$ reached 0.93 and 0.89 for the barley and wheat testing datasets, respectively (Table 4 and Figures 12b.1 and 13b.1).

The spatial distribution of the AGB regression results is shown as an AGB prediction map in Figure 8c at a 1 m$^2$ spatial resolution.

Table 4. The prediction results as a function of spatial resolution for the barley and wheat crops.

<table>
<thead>
<tr>
<th>Testing Dataset</th>
<th>Spatial Resolution (m$^2$)</th>
<th>$R^2$</th>
<th>RMSE (g/m$^2$)</th>
<th>MAE (g/m$^2$)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barley Testing Dataset</td>
<td>0.35</td>
<td>0.48</td>
<td>207</td>
<td>162</td>
<td>42.0</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.71</td>
<td>232</td>
<td>214</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.93</td>
<td>300</td>
<td>266</td>
<td>13.0</td>
</tr>
<tr>
<td>Wheat Testing Dataset</td>
<td>0.35</td>
<td>0.20</td>
<td>288</td>
<td>216</td>
<td>23.7</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.58</td>
<td>284</td>
<td>264</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.89</td>
<td>400</td>
<td>351</td>
<td>16.0</td>
</tr>
</tbody>
</table>

Figure 12. Cont.
**Figure 12.** An analysis of the aggregated predictions using the barley testing dataset: (a) the residual distribution had a mean value approaching zero (i.e., 2.2 g/m² in the testing set, where N = 57); (b) the $R^2$ score converged to 1 as the number of aggregated samples increased. At every step along the $x$-axis, the data series took the mean (green solid line) of 100 repetitions (at a 1 m² spatial resolution, where $R^2 = 0.71$). The light gray line shows the worst performance in each iteration, while the shaded area covers the confidence interval (i.e., ± the standard deviation); (b.1) a scatter plot of the predicted AGB values vs. the AGB field measurements at a 1 m² spatial resolution.

**Figure 13.** Cont.
**Figure 13.** An analysis of the aggregated predictions using the wheat testing dataset: (a) the residual distribution had a mean value approaching zero with a slight overestimation (i.e., $-18.5 \text{ g/m}^2$ in the testing set, where $N = 183$), which represented a systematic error of 2% for the average wheat sample weight; (b) the $R^2$ score converged to 1 as the number of aggregated samples increased. At every step along the x-axis, the data series took the mean (green solid line) of 100 executions (at a 1 m$^2$ spatial resolution, where $R^2 = 0.58$). The light gray line shows the worst performance in each iteration, while the shaded area covers the confidence interval (i.e., $\pm$ the standard deviation); (b.1) a scatter plot of the predicted AGB values vs. the AGB field measurements at a 1 m$^2$ spatial resolution.

4. Discussion

4.1. AGB Prediction

This study explored a new method to acquire real-time estimates of the AGB distribution in croplands at a sub-meter resolution using UAV–LiDAR data and ML regression models. The results indicated that local variations in crop morphology (e.g., crop structure and height development) could be captured by a close-range UAV–LiDAR survey, thereby enabling a regression analysis of the AGB. The accuracy that was achieved using the ERT model was on a par with the currently established methods, i.e., multivariate linear regression and power regression models, while attaining resolutions that were one to two orders of magnitude higher than those in the reference studies, e.g., Han et al. (2.4 m$^2$) [53], Ma et al. (12 m$^2$) [56] and Zha et al. (160–300 m$^2$) [54].

Figures 10 and 11 show how much the ERT model predictions diverged from the ideal case (i.e., the line of equality) at a 0.35 m$^2$ resolution when predicting AGB for the same crop species that was used for training and when predicting AGB for a different crop species, respectively. The ERT regression model achieved valid predictions in both cases and the ERT model predictions outperformed those of the linear regression model that was used as the comparative baseline (Figures 10b and 11b). However, the prediction results for different crop species (i.e., training the model with the barley dataset and predicting the AGB for wheat crops) were not as accurate as those for the same species, which was expected. The higher accuracy of the predictions for the same crop species also meant that the aggregated prediction required fewer aggregated samples to converge in the barley set (Figure 12b) than in the wheat set (Figure 13b). The performance reduction when training and predicting for different species could be attributed to (i) the different plant-level AGB distributions between the two crop species, as captured by the AGB labels (Figure 2), and (ii) the different morphology in the canopy structures, as portrayed by the LiDAR-derived predictors. The differences between both crops were retained in the datasets, which caused a dataset shift effect [88]. This shift challenged the accuracy of the AGB predictions since the joint distributions of the predictors and the target AGB values were different for the training and testing phases.
Regarding the importance assessment of the selected predictors (i.e., independent variables), care had to be taken when considering predictors that had high levels of correlation between them, as occurs in LiDAR-derived height metrics. It has been observed empirically [89,90] and analytically [91] that high correlation across prediction features compromises the importance assessment in tree-based methods.

The predictive features that showed the highest correlations with AGB were those that were derived from the height metrics and showed a Pearson correlation of between 0.65–0.71. In contrast, the intensity metrics did not correlate consistently with AGB. This was probably due to the low cardinality of the intensity range, which prevented the production of sufficiently descriptive features.

With regard to phenology, the barley dataset that was collected over an extended time period in 2020 covered different crop development phases and presented higher correlation coefficients between height-related predictors and AGB compared to the barley dataset that was collected in 2021, which did not cover all of the phases of vegetation growth. The results showed that it was especially important to use datasets that covered different phases of vegetation development to train the models for AGB predictions in order for the regression models to learn the relationships between the predictive features and the target values. However, once a model was adequately trained, it could predict AGB using datasets that were collected during “snapshot surveys” (i.e., data acquisition campaigns that were conducted in one single day), such as the barley21 dataset.

When considering vegetation structure, it was observed that open canopies and heterogeneous crop patterns were morphological characteristics that compromised the use of LiDAR-derived height metrics as predictors of AGB. Plots that were surrounded by sparse vegetation were commonly reached by LiDAR beams at lower heights than plots in denser areas, thus distorting the relationships between the AGB and height metrics.

Additionally, increased variations in the AGB labels could enlarge the training domain, which favored the regression model’s ability to interpolate predictions using the testing set [92]. Likewise, more variability in the prediction features could make them more descriptive and, therefore, better at capturing the morphological traits of the crops (e.g., attaining a better portrayal of stem density and canopy structure).

4.2. Aggregation of AGB Predictions

It was notable that AGB predictions with significantly higher accuracies could be achieved using slightly coarser spatial resolutions (Figures 12 and 13). The aggregation improved the model’s predictions considerably by only adding two instances (i.e., at a spatial resolution of 0.7 m$^2$), while the 1 m$^2$ resolution was found to be the optimal trade-off between spatial resolution and prediction performance for capturing local variations in AGB (Table 4 and Figures 12b.1 and 13b.1). In terms of the sampling resolution, the datasets that were composed of smaller samples (i.e., 0.5 m × 0.34 m) showed lower correlations between the height-related predictors and AGB than the same datasets after the aggregation of two or three samples (1 m × 0.34 m and 1.5 m × 0.34 m, respectively). The AGB samples with a very high spatial resolution (e.g., 0.175 m$^2$) could suffer from poor PCD representation (i.e., low counts of LiDAR returns), thereby compromising the reliability of the statistics that were extracted. Additionally, the datasets that had small sample sizes presented higher variance in the AGB ground truth values (i.e., the measurements at a 0.175 m$^2$ resolution produced noisier labels).

4.3. Applied ML Methods

Although we considered different ML methods, it was not our intention to systematically compare the performances of a series of ML models to find the lowest possible error rate since the experimental design of this study did not allow for such comparisons. Only relatively few training instances were available and the target AGB values had a high variance (Figure 2a,b). Instead, we aimed to test an ML regression method that utilized features that were derived from point cloud data that represented the structures of crop canopies to predict AGB, taking into account the aforementioned limitations. The suggested
method consisted of selecting one representative model per family of methods (Table 2) and the set of parameters (Table 3) that performed the best using the validation set, followed by re-fitting the selected model (i.e., ERT) to an unseen dataset that was collected the following year (Figure 7). As such, the selection of a specific ML model was not relevant in this case. Indeed, by exploring extended parameter sets, one model could slightly outperform the others for individual predictions or a different model could perform better using datasets from subsequent years or another crop type. Achieving greater accuracy and predicting AGB at higher spatial resolutions would require more training data that are representative of different growing conditions and collected in a protocolized manner using both AGB labels and PCD scenes in order to achieve AGB predictions that are robust to inter-annual and inter-species variations (e.g., crop structure, weather conditions, different crops, etc.).

5. Conclusions

In this study, we developed a method that combines UAV–LiDAR surveys and ML techniques to predict the above-ground biomass of standing vegetation for cereal crops at sub-meter resolutions. This method was developed and tested for two crop species at the same field site over two consecutive growing seasons: winter wheat (Triticum aestivum L.) and barley (Hordeum vulgare L.). The ERT model performed real-time AGB estimation, taking as its input the height and intensity metrics that were derived from the point cloud data scene, and achieved a prediction performance of $R^2 = 0.48$, RMSE = 207 g, MAE = 162 g, and MAPE = 42% at a spatial resolution of 0.35 m$^2$. However, by aggregating the individual predictions, it was observed that the prediction performance could be increased significantly by coarsening the spatial resolution as the predictions were statistically independent and uncorrelated. At a spatial resolution of 2 m$^2$, the regression performance achieved $R^2 = 0.93$, RMSE = 300 g/m$^2$, MAE = 266 g/m$^2$, and MAPE = 13% when the training and testing datasets corresponded to the same crop species. The aggregated AGB predictions also achieved reasonable results at a 2 m$^2$ spatial resolution when the model was trained on one crop type and tested on another: $R^2 = 0.89$, RMSE = 400 g/m$^2$, MAE = 351 g/m$^2$, and MAPE = 16%. This slight reduction in the prediction performance was explained by the differences between (i) the canopy structures and (ii) the plant-level biomass distributions of the two crop species under consideration. We encourage the continuation of protocolized field-based data collection campaigns, as well as UAV–LiDAR surveying, as a means to gain valuable data that are representative of crops under different environmental conditions in order to develop AGB regression models that are capable of generalizing predictions that are even more robust to inter-annual and inter-species variations.

More precise AGB estimates allow for real-time evaluations of crop yields, thereby enabling adaptive management practices (such as adjusting harvest dates according to mill capacity) and estimated production, as well as improving early warning systems for food security in climate-vulnerable regions. The method that was introduced here for acquiring real-time AGB estimates for croplands using UAV–LiDAR surveys and ML supervised regression models could pave the way for future studies on predicting the spatial distributions of AGB-related biochemical constituents at sub-meter resolutions by applying this method throughout vegetation growth periods for precision agriculture and agroecological applications. This quantitative method could also assist in management decision-making. Finally, this line of research could help to narrow down the margins of uncertainty that are still present in the estimates of current above-ground carbon stocks and C-turnover values at the ecosystem level, for example.

Funding: This project received funding support from the Talent Program Horizon 2020/Marie Skłodowska-Curie Actions; a Villum Experiment grant from the Velux Foundations, the Drone-Borne LiDAR and Artificial Intelligence for Assessing Carbon Storage (MapCland) project (grant number: 00028314); the Deep Learning for Accurate Quantification of Carbon Stocks in Cropland and Forest Areas (DeepCrop) project (UCPH Strategic plan 2023 Data + Pool); as well as a UAS ability infrastructure grant from the Danish Agency for Science, Technology, and Innovation. The authors also acknowledge the financial support from the Independent Research Fund, Denmark, through the Monitoring Changes in Big Satellite Data via Massively-Parallel Artificial Intelligence project (grant number: 9131-00110B) and the Villum Fonden through the Deep Learning and Remote Sensing for Unlocking Global Ecosystem Resource Dynamics (DeReEco) project (grant number: 34306).

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the contributions of René Lee, Lars Rasmussen, Rune Skov Maigoord, Binsheng Gao, and Alek Wieckowski, who supported the task of field data acquisition and contributed to this study as fieldwork and laboratory assistants.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations
The following abbreviations are used in this manuscript:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGB</td>
<td>Above-Ground Biomass</td>
</tr>
<tr>
<td>dGPS</td>
<td>Differential Global Positioning System</td>
</tr>
<tr>
<td>DTM</td>
<td>Digital Terrain Model</td>
</tr>
<tr>
<td>ERT</td>
<td>Extremely Randomized Trees</td>
</tr>
<tr>
<td>FoV</td>
<td>Field of View</td>
</tr>
<tr>
<td>ICOS</td>
<td>Integrated Carbon Observation System</td>
</tr>
<tr>
<td>LiDAR</td>
<td>Light Detection and Ranging</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>PCD</td>
<td>Point Cloud Data</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RS</td>
<td>Remote Sensing</td>
</tr>
<tr>
<td>RTK</td>
<td>Real-Time Kinematic</td>
</tr>
<tr>
<td>SAR</td>
<td>Synthetic Aperture Radar</td>
</tr>
<tr>
<td>UAV</td>
<td>Unstaffed Aerial Vehicle</td>
</tr>
</tbody>
</table>

Appendix A
In order to quantify the effects of crop sparsity on the uncertainty that was contained in the AGB labels, we collected 180 samples from 10 different locations of 1 m² each (see Figure A1). This extra study contrasted with the main field study, for which sub-meter samples were taken exclusively. To calculate the variance, each of the 18 sub-meter samples was upscaled to g/m² and then subtracted from the total AGB in the 1 m² that was measured at each location. Due to oven capacity limitations, this analysis was conducted with wet AGB, assuming that the variations in water content within 1 m² would be negligible. A higher variance was found in the datasets that were composed of samples with higher spatial resolutions (i.e., the 2021 datasets) and within those, it was substantially higher for crops with higher sparsity (i.e., wheat) than for the more homogeneous crops (i.e., barley), as was expected.
**References**
