Learning UAV-Based Above-Ground Biomass Regression Models in Sparse Training Data Environments

F. Kröber, G. Fernandez Garcia, F. Guiotte, F. Delerue, T. Corpetti, S. Lefèvre

LEARNING UAV-BASED ABOVE-GROUND BIOMASS REGRESSION MODELS IN SPARSE TRAINING DATA ENVIRONMENTS

F. Kröber [1], G. Fernandez Garcia [2], F. Guiotte [3], F. Delerue [4], T. Corpetti [2], S. Lefèvre [1]

[1] Université Bretagne Sud, IRISA UMR 6074, Vannes, F-56000, France
[2] CNRS, LETG UMR 6554 Rennes, F-35000, France
[3] L'Avion Jaune, Saint-Clément-de-Rivière, F-34980, France
[4] Université Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, Pessac, France

ABSTRACT

This study aims at recovering above-ground biomass information from ultra-high resolution UAV RGB-NIR orthophotos. We focus on a realistic scenario where a limited number of training samples for a landscape with heterogeneous herbaceous vegetation is given. Consequently, we explore different machine learning methods explicitly addressing the limitations of small training samples and compare their predictions quantitatively and qualitatively. Our results show that random forest models perform similarly well to deep learning models. While simpler machine learning models may, therefore, still be preferable, our study also points the way to promising architectures and regularisation techniques for deep learning approaches. Beyond vegetation cover, accurate regression of other variables, including vegetation height, volume and biomass remains a difficult task regardless of the model choice.

***Index Terms*—** vegetation biomass regression, random forest, deep learning, semi-supervised learning, transfer learning, UAV

1. INTRODUCTION

Precise estimation of the amount of vegetation present in a given ecosystem is vital for understanding its functioning, managing its resources and predicting its evolution under different constraints including global change. Therefore, many ecological studies measured different indicators of vegetation development (like vegetation cover or vegetation volume), the most precise one being vegetation biomass. However, traditional field campaigns for measuring biomass are time-consuming, costly and destructive, making remote sensing an attractive alternative. Unmanned Aerial Vehicles (UAVs) have proven to be suitable platforms for performing data collection at the field scale with high spatial resolutions in an efficient and non-destructive manner. From a modelling perspective, UAV-based estimations of biomass-related variables can be framed as a regression task between orthophotos and a set of field measurements. The often limited availability of the latter may explain the prevailing use of robust models based on traditional machine learning. While Convolutional Neural Networks (CNNs) have been applied to UAV data for various tasks, their use for solving regression problems [1] and especially biomass-related ones [2] is still underrepresented. In this paper, we explore three possibilities for recovering biomass information from ultra-high resolution (i.e., millimetric resolution) UAV RGB-NIR orthophotos. We aim to compare different machine and deep learning methods which are explicitly addressing the limitations of small training samples: Random Forest (RF) (Model A), transfer learning (Model B) and data augmentation (model C). In addition to the above-ground biomass (AGB) itself, vegetation cover, height and volume as related variables are also targeted. It should be noted that in this study we focus on a particularly complex case: our data campaign concerns post-mining stressed ecosystems, with small, heterogeneous, herbaceous plant communities, characterised by low values of AGB. This constraint makes a precise estimation of biomass-related variables more challenging, especially from images with a millimetric resolution.

This work has been supported by the SIXP project, funded by the French National Research Agency (ANR-19-CE02-0013-01).

2. RELATED WORKS

Today, most studies to estimate biomass still focus on classical parametric learning methods such as simple or multiple linear regressions [2, 3, 4]. While the authors of [5] have pointed out that on-site measurements are crucial for satellite-borne biomass missions, the lack of in situ measurements also hinders the training of more complex deep learning-driven models when the data is collected by UAVs. So far, preprocessing, including the creation of 3D models to derive structural information, is frequently carried out to allow the usage of simpler models based on a few samples. For example, in [6], the authors first calculate a crop surface model from UAV data, then derive the plant height, and finally use regression models from plant height to biomass. Also, multi-temporal data is used to improve the quality of estimations [7], along with other complementary information such as meteorological features [8]. Meanwhile, the problem of estimating biomass from monotemporal UAV orthophotos without extensive preprocessing and/or auxiliary data remains an open problem. Our paper is a contribution in this direction.

3. MATERIALS AND METHODS

3.1. Data Description

UAV and in situ measurements were taken for four sites in different altitude zones (1850 to 2100 m) in the Pyrenees, France, from spring to summer 2021 (Fig. 1). Due to a difficult environment for plant growth in the surveyed areas, corresponding plant communities were herbaceous and particularly small (<20 cm). UAVs were equipped with R, G, B, NIR sensors and the obtained images post-processed to derive orthophotos with a spatial resolution of 3 mm. In situ measurements were taken for four variables: fractional vegetation coverage, vegetation volume, vegetation height and AGB. About 30-35 measurements were taken per site leading to 160 measurements. For

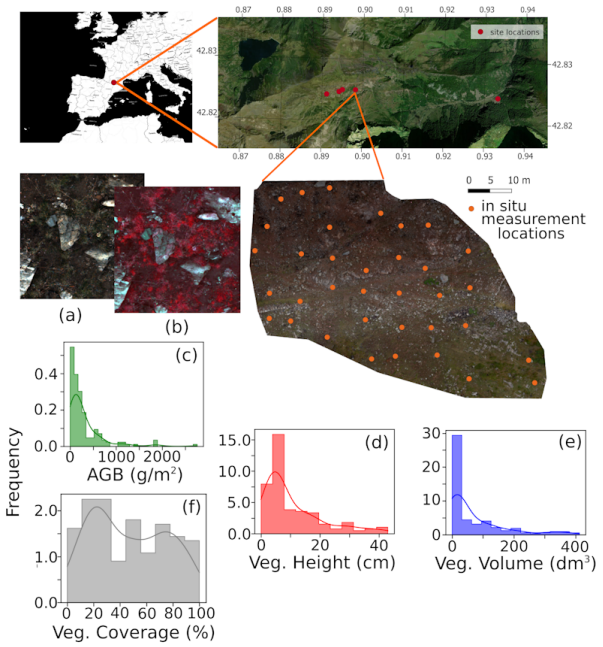

Fig. 1. Site localisation, an example of UAV images along with the in situ measurements locations and data description: (a) UAV tile in RGB; (b) UAV tile in false-color; (c) distribution of AGB values (g/m^2) in the labelled dataset (the green line is the kernel density estimation of the distribution); (d) vegetation height (cm); (e) vegetation volume (dm^3); (f) vegetation coverage (%).

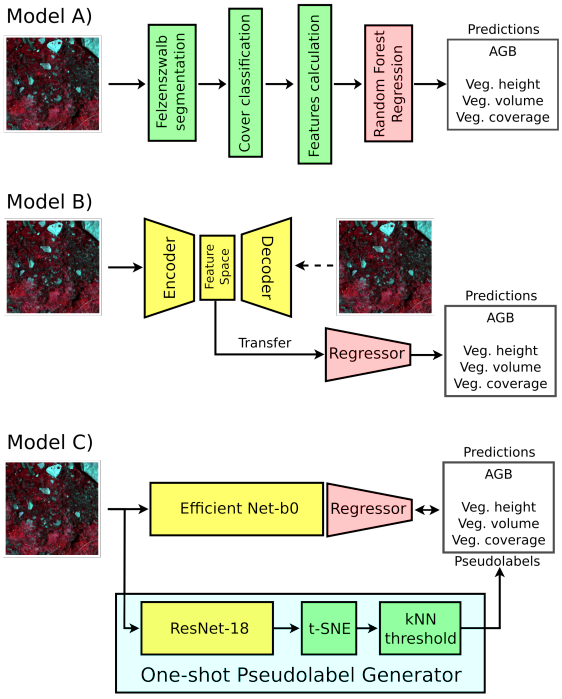

Fig. 2. Schematic representation of the three models.

each point, all variables were recorded based on a 25x25 cm square representative of the 1x1 m area centered around the point. Two datasets have been generated from these data: a labelled dataset and an unlabelled one. For the former, the UAV images were cropped to the in situ measurement geolocations generating tiles with a 1x1 m extent centered around each point measurement, resulting in 160 couples of tiles and in situ measurements. 124 of them were splitted in a spatially stratified random manner into train (75), validation (24) and test (25) sets. Another 36 data points, all drawn from a single site, are used as a spatially independent control site (CMB test site). For the unlabelled dataset, the UAV images have been tiled at 1x1 m with an overlap of 50%, resulting in 18,085 tiles of 256x256 pixels, splitted in train (70%), validation (15%) and test (15%) sets.

3.2. Model A: Random Forest Model

Given our low amount of data points, an RF classifier [9, 10] with manually engineered features has been used as a robust baseline architecture. First, all tiles of the labelled dataset were segmented with a Felzenszwalb segmentation [11] (Figure 2). The derived objects were then evaluated with a simple automated tile cover classification, distinguishing between vegetation, bare ground and shadowed areas. To this aim, segmented tiles with mean NDVI values above a given threshold were classified as vegetation. Shadowed areas classification was based on consistently low reflectance values across the R, G and B channels. Given these classified tiles, two types of features were engineered. First, spectral features and their derivates, i.e. the share of vegetation, bare ground and shadow, were used. The shares of vegetation were calculated for three different NDVI thresholds (0.3, 0.4, 0.5) to let the RF algorithm choose the most informative variable. Second, the spatial arrangement of these classes was also considered: the size and shape properties of the segmented objects were used. Specifically, the mean value, standard deviation of the size and shape index of all objects classified as vegetation were taken into account as features. In order to represent multi-scale information, the features were calculated for two segmentation scale levels (20, 200). This made a set of 40 features derived from the images fed into the RF for the regression of vegetation parameters.

3.3. Model B: Transfer Learning

The underlying hypothesis of model B is that an Autoencoder (AE) [12], trained on the unlabelled dataset, can generate a feature space that can be easily adapted to the regression task on the labelled dataset. To this end, the encoding part of an AE model was designed to have a latent representation of size 32x42x42. This dimensionality of the latent feature space allowed to create a simple regression CNN with less than 10 K parameters in the second step. The regression CNN starts with a 1D convolutional layer intended to reduce the number of feature maps to the ones needed for the task at hand. Subsequently, three 2D convolutional layers are used again. Average pooling is applied to the output, the features are flattened, fed into a linear layer and sigmoid is applied to force the output to take values between 0 and 1. Rescaling (100x) represents the final step to get the common value range of the variables of interest.

3.4. Model C: Data augmentation/Semi-supervised approach

Another possibility to exploit the great number of unlabelled tiles is to consider their image similarity to the labelled ones, following the semi-supervised setting from [13]. Therefore, model C (Figure

2) uses a one-shot pixel-wise pseudo-label generator for unlabelled tiles. With a ResNet-18 [14], pre-trained on ImageNet and without the last layer, 512 features were calculated for each UAV tile. The proximity of two vectors in this feature space is assumed to indicate the structural similarity of images and can be exploited to assign the pseudo-labels. To circumvent the issues arising from the high dimensionality of the feature space, vectors were mapped to a 2D space with t-Distributed Stochastic Neighbour Embedding (t-SNE) [15], before performing the assignment. In detail, the dataset was divided into small batches of 150 tiles, containing an equal number of labelled training tiles and unlabelled tiles. The mean nearest neighbour (NN) distance between all points was calculated and used as a distance threshold for assigning values to unlabelled tiles. If the distance between an unlabelled point and one or multiple labelled ones was lower than the mean NN distance, a pseudo-value was calculated for the unlabelled one using the weighted average of the values of all the labelled tiles being considered. This procedure was repeated for all unlabelled tiles. To handle t-SNE inherent stochasticity, multiple trials were conducted for each batch. The one with minimum Kullback-Leibler divergence and root mean squared error (RMSE) calculated between labelled point pairs less than the mean NN distance apart was chosen to be the best one used for pseudo-value assignment. Afterwards, the datasets were balanced, ensuring a uniform quantile spacing for each variable using a random over-sampling technique. Finally, the number of generated pseudo-values was capped at 5000 to limit the amount of data and to speed up the subsequent training. The regression was then performed by placing a regression block (as described in Section 3.3) in place of the last layer of a pre-trained EfficientNet-b0 [16].

3.5. Validation strategy

For all models, hyperparameter tuning has been performed by measuring the model performance on the validation set. For model A, the number of features considered when looking for the best split at a given tree node and the minimum number of samples required to be at a leaf node were tuned as hyperparameters. They were optimised using the RMSE. For models B and C, the learning rate, optimiser (Adam vs. SGD) and loss (L1 vs. L2) were considered as relevant hyperparameters and chosen based on the mean absolute error (MAE). The stratified sampling splits as described in Section 3.1 were created ten times to perform cross-validation of the results. Only for the AE training within model B, cross-validation was not necessary due to the large number of unlabelled tiles.

4. RESULTS AND DISCUSSION

The quantitative comparison between the models using MAE, RMSE and Person's correlation coefficient (r) is presented in Table 1. Across all metrics, the three models have similar accuracy. For AGB, model A performs slightly better, while for all other variables model C produced the most accurate results. For each singular variable, the r coefficient is coherent between models. It varies considerably between variables: while predictions and in situ measurements are strongly correlated for coverage, volume and height exhibit intermediate values. For AGB lowest correlations are observed. This reflects the increased difficulty of estimating more complex structural-morphological variables based on 2D UAV data. Interestingly though, the increased complexity of the regression tasks for these structural variables is not handled better by the deep learning models compared to the RF model. A correlation between the models' capacities, the quality of their predictions and the complexity of the task is not apparent.

Importantly, a large uncertainty is attached to all metrics introduced by the small training dataset. To investigate the statistical significance of the results, a Bayesian correlated t-test [17] was conducted. The advantage of this framework is the ability to calculate a posterior distribution of performance differences between the models, which can be interpreted as an actual probability distribution. This allows to evaluate if the observed differences exceed a level for which we would consider two algorithms to obtain practically equivalent results. The regions of practical equivalence (aka ROPEs) were set to 0.05 times the standard deviation of the respective variable of interest. As Figure 3 shows, most of the model comparisons obtain non-significant differences and the models are more likely to differ only for vegetation volume.

To evaluate the results qualitatively, predictions for each model were created for the independent CMB test site using a dense sampling interval of 0.5m. Subsequently, the predicted values were linearly interpolated to obtain spatially continuous representations (Fig. 4). As can be seen in these maps, there is a high correlation between the predictions of the models and the different variables. For all variables and models, the predicted distributions generally match the spatial patterns as emergent from the true and false-color visualisations of the site. Differences between the individual models concern a stronger tendency of model C towards high-frequent noise and less accurate predictions of intermediate values especially for vegetation cover and height. At the same time, the predictions across variables seem to be better correlated for model C. While spots may appear to be densely vegetated based on predicted vegetation cover and height, they may counter-intuitively have low biomass and vegetation volume values at these locations for models A and B.

Variable	Model	MAE	RMSE	r
AGB	A	**160.83 (± 26.99)**	**220.99 (± 50.36)**	0.58 (± 0.17)
	B	176.94 (± 33.29)	234.54 (± 55.87)	0.55 (± 0.18)
	C	164.21 (± 29.09)	235.43 (± 47.00)	**0.61 (± 0.17)**
coverage	A	8.60 (± 1.43)	11.07 (± 2.06)	0.93 (± 0.03)
	B	9.36 (± 1.13)	11.47 (± 1.36)	0.93 (± 0.03)
	C	**8.07 (± 1.84)**	**10.45 (± 2.94)**	**0.94 (± 0.04)**
height	A	5.93 (± 0.97)	8.00 (± 1.40)	0.68 (± 0.07)
	B	5.56 (± 1.08)	7.74 (± 1.36)	0.70 (± 0.09)
	C	**5.46 (± 1.00)**	**7.32 (± 1.34)**	**0.76 (± 0.09)**
volume	A	42.76 (± 7.63)	65.94 (± 14.84)	0.77 (± 0.08)
	B	37.71 (± 9.84)	62.60 (± 16.23)	0.79 (± 0.09)
	C	**36.66 (± 8.12)**	**58.15 (± 12.16)**	**0.83 (± 0.06)**

Table 1. Test accuracy measured for the 25 samples of the test set. Mean values and standard deviations across 10-folds are specified. Best models for each variable and metric are highlighted.

5. CONCLUSIONS

In this paper, we proposed learning-based methods to derive biophysical parameters (AGB, vegetation height, volume and coverage) from ultra-high resolution UAV imaging (i.e., millimetric resolution). We explored approaches that are explicitly addressing the limitation of sparse training data: the first one relied on a robust machine learning algorithm, the second one exploited the large quantity of unlabelled UAV data and the last one additionally considered the statistical distribution of the in situ measurements (pseudo-label gen-

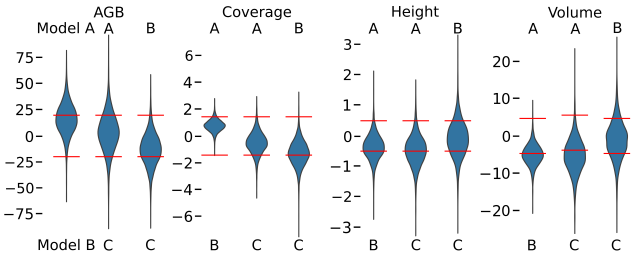

Fig. 3. Posterior distributions for mean absolute differences between models. Red bars enclose the ROPEs.

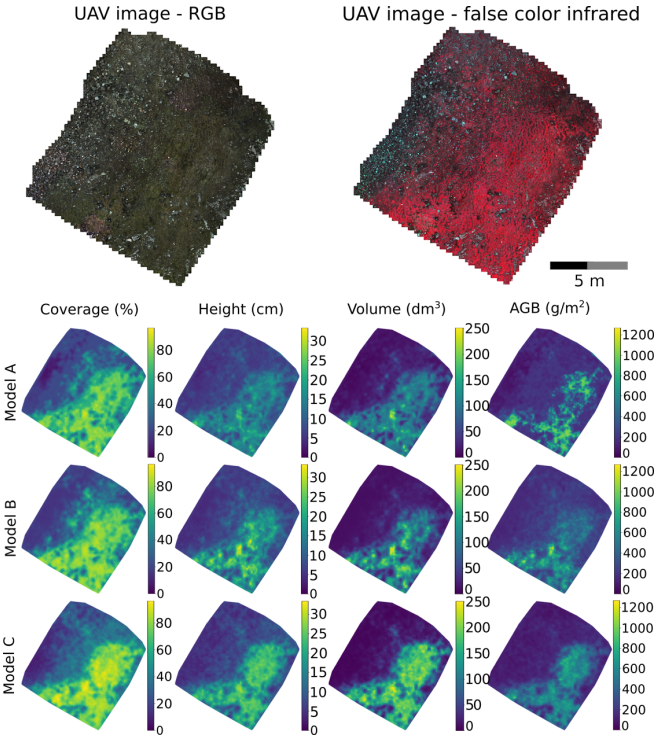

Fig. 4. Mapped predictions for CMB test site.

eration). Strikingly, the considerable uncertainty in predictions did not allow any model to be identified unequivocally as superior or inferior to another. This is another consequence of the lack of in situ measurements but also due to their distribution and intrinsic measurement precision. We note that the study of the uncertainties (via the t-test in our study) is a crucial step to compare the models in such cases, even if a cross-validation approach is already used. Despite the existing uncertainties, it has to be noted that the two deep learning models were not significantly worse than the RF model. The performances of model C were even promising. Even if the literature on AGB's regression from UAV data is still focused mainly on machine learning approaches, this exploratory study was able to pave the way for novel deep learning methods to study biophysical parameters in the case of sparse training data. A transfer to and test with other data sets such as GrowliFlower [18] remains to be done in the future. Finally, from an ecological perspective, the most interesting parameters to study degradated ecosystems are AGB (also the most challenging with height) and vegetation cover. With this in mind, we can affirm that the good performances on the latter variable puts this study in an interesting position for pratical applications.

6. REFERENCES

[1] L.P. Osco et al., "A review on deep learning in UAV remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, pp. 102456, Oct. 2021.

[2] T. Wang et al., "Applications of UAS in crop biomass monitoring: A review," *Frontiers in Plant Science*, vol. 12, pp. 616689, Apr. 2021.

[3] T.G. Morais et al., "The use of machine learning methods to estimate aboveground biomass of grasslands: A review," *Ecological Indicators*, vol. 130, pp. 108081, Nov. 2021.

[4] L.G. Poley and G.J. McDermid, "A systematic review of the factors influencing the estimation of vegetation aboveground biomass using unmanned aerial systems," *Remote Sensing*, vol. 12, no. 7, pp. 1052, Mar. 2020.

[5] J. Chave et al., "Ground data are essential for biomass remote sensing missions," *Surveys in Geop.*, vol. 40, pp. 863–880, 2019.

[6] J. Bendig et al., "Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging," *Remote sensing*, vol. 6, no. 11, pp. 10395–10412, 2014.

[7] I. Sinde-González et al., "Biomass estimation of pasture plots with multitemporal UAV-based photogrammetric surveys," *International Journal of Applied Earth Observation and Geoinformation*, vol. 101, pp. 102355, 2021.

[8] Q. Jiang et al., "UAV-based biomass estimation for rice-combining spectral, TIN-based structural and meteorological features," *Remote Sensing*, vol. 11, no. 7, pp. 890, 2019.

[9] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[10] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. of Photo. and Rem. Sens.*, vol. 114, pp. 24–31, 2016.

[11] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int J. of Comp. Vis.*, vol. 59, no. 2, pp. 167–181, Sep 2004.

[12] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *CoRR*, vol. abs/2003.05991, 2020.

[13] J. Castillo-Navarro et al., "Semi-supervised semantic segmentation in earth observation: The minifrance suite, dataset analysis and multi-task network study," *Machine Learning*, vol. 111, pp. 3125–3160, 2022.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[15] G.E. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Advances in Neural Information Processing Systems*, 2002, vol. 15.

[16] M. Tan and Q.V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *ArXiv*, vol. abs/1905.11946, 2019.

[17] A. Benavoli et al., "Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis," *Journal of Machine Learning Research*, vol. 18, no. 77, pp. 1–36, 2017.

[18] Jana Kierdorf et al., "Growliflower: An image time-series dataset for growth analysis of cauliflower," *Journal of Field Robotics*, 2022.