Using Structural Class Pairing to Address the Spatial Mismatch Between GEDI Measurements and NFI Plots

Nikola Besic, Sylvie Durrieu, Anouk Schleich, and Cédric Vega

***Abstract*—The Global Ecosystem Dynamics Investigation (GEDI) mission can significantly enhance multisource national forest inventories (MSNFI) by improving the spatio-temporal resolution of forest attributes while preserving the statistical relevance of the design-based inference approach. The main challenge is the lack of systematic spatial alignment between GEDI footprints and National Forest Inventory (NFI) plots, which is necessary to accurately link in situ forest measurements with GEDI data. In this study, we aim to tackle the aforementioned issue by introducing a methodology for interpolating GEDI measurements to NFI plots, enabling the calibration of GEDI data using localized NFI estimates. Our proposed method incorporates clustering, classification, and regression techniques, and utilizes GEDI and NFI data, along with Sentinel-2 images, land-use information, and topographic data. Beginning with the prediction of profile structural classes and shapes on NFI plots, the proposed method ultimately projects actual measurements onto the NFI plot sites through profile pairing within the predicted structural classes. The method is conceived and validated using the data acquired across the mountainous area of ∼500 kha, covered by >500 NFI plots. Our validation framework shows that the method is able to project relative height profiles at NFI plots, allowing to partly interpolate the lower part of the profile and not only the canopy top height. This enables the construction of models that efficiently relate GEDI profiles and wood volume, demonstrating the importance of incorporating lower relative height values when linking forest attributes and lidar measurements ($R^2 = 0.65$, $MBE = 2.31$ m^3/ha).**

***Index Terms*—Global Ecosystem Dynamics Investigation (GEDI), machine learning, modeling, National Forest Inventory (NFI), Sentinel-2, wood volume.**

I. Introduction

A COMPREHENSIVE and systematic observation of forests is of utmost importance for a variety of reasons, ranging from ecological to economic and societal concerns, especially in light of the rapidly evolving climate emergency [1]. The scope and speed of the climate crisis, along with its impacts on public policies, make it undeniably necessary to monitor forest attributes on a large scale with higher spatial accuracy and temporal frequency.

The design-based inferential approach is by far the most established framework for forest observation and serves as the backbone of National Forest Inventories (NFIs) [2]. In its conventional form this approach principally relies on the ground measurements, i.e., the inventory plots, and allows to infer without bias both quantitative and qualitative forest attributes. The inference is performed following a beforehand defined stratification pattern across a particular territory, and in the time frame spanning typically over several years. The classical NFI approach indeed represents a way to ensure a comprehensive and systematic forest observation at a broad scale, but it does not respond to the previously mentioned and increasingly important requirement concerning the higher spatial accuracy. The latter can be addressed only by involving remote sensing data [3], [4], [5], [6].

When ground measurements are enhanced with auxiliary remote sensing data within an inferential framework, this approach is referred to as the multisource national forest inventory (MSNFI) [7]. There are several ways of introducing remote sensing data into a design-based inferential approach, with all of them requiring to first establish a model able to link ground measurements and remote sensing data [8]. The former is true whether we deal with the model-assisted methods, where ground measurements are still driving the estimation, or with the model-based methods, where the role of ground measurements ends with establishing the model and the estimation is somehow principally driven by remote sensing data.

To construct a model that links ground measurements with remote sensing data, the remote sensing data must have the physical capability to replicate some of the variance observed in the in situ measured or locally estimated forest attributes. The optimal remote sensing dataset is chosen based on the complexity of the forest, considering both its composition and structure. Optical and radar images cover repeatedly large surfaces at high to medium spatial resolution, but do not contain an elaborated vertically resolved information about the forest stand, making them less suitable to face a more complex forested environment. The airborne lidar and to a degree photogrammetric sensors [9],

Manuscript received 27 March 2024; revised 17 June 2024; accepted 4 July 2024. Date of publication 9 July 2024; date of current version 24 July 2024. This work was supported in part by the SLIM project, in part by the TOSCA Continental Surface Program of the Centre National d'Etudes Spatiales (CNES) under Grant 4500066524, in part by the SylvoSanSat project, in part by the TOSCA Continental Surface Program of the Centre National d'Etudes Spatiales (CNES) under Grant 4800001230, and in part by the grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). The work of Anouk Schleich was supported in part by INRAE and in part by IGN. *(Corresponding author: Nikola Besic.)*

Nikola Besic and Cédric Vega are with the IGN, ENSG, Laboratoire d'inventaire forestier, 54042 Nancy, France (e-mail: nikola.besic@ign.fr, n.m.besic@gmail.com).

Sylvie Durrieu and Anouk Schleich are with the INRAE, UMR Territoires, Environnement, Télédétection et Information Spatiale, 34000 Montpellier, France.

Digital Object Identifier 10.1109/JSTARS.2024.3425431

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

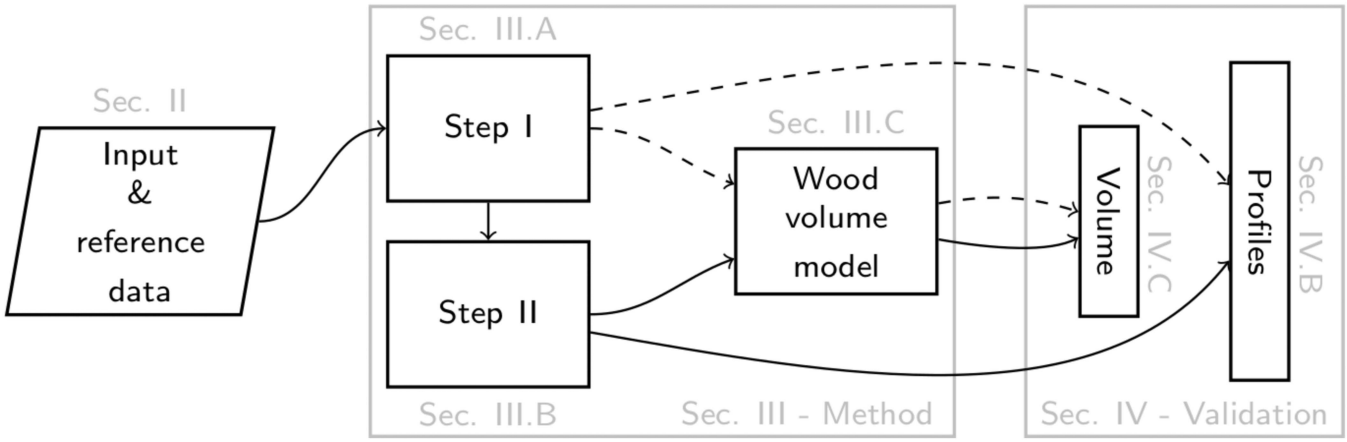

Fig. 1. Diagram presenting the organization of the central part of the manuscript.

[10] acquire a vertically resolved information and have thus the potential to tell on quite a bit of parameters describing a more complex environment, but do not cover large areas on a regular basis. The high energy spaceborne lidar data, as the one acquired by the Global Ecosystem Dynamics Investigation (GEDI) mission, represents a sort of compromise. It namely offers broad and repeated coverage, coupled with slightly less vertical resolution compared to airborne lidar, but lack the spatial continuity characterizing optical and radar imagers.

From aboard the International Space Station (ISS), the GEDI mission [11] is densely sampling, through the acquisition of full-waveform lidar data, a big part of world's forests, and has therefore the remarkable potential to contribute achieving MSNFI ambitions of countries situated between 51.6°N and 51.6°S latitudes. However, despite the high resolution and the dense sampling strategy, as suggested above, GEDI remains a nonimaging sensor, and thus, does not allow to provide the wall-to-wall coverage of entire forest areas. There has been hence an intense development of techniques aiming at interpolating/extrapolating spatially and temporally forest attributes derived from GEDI measurements [12]. Most of these address the interpolation or extrapolation of the canopy height [13], [14], [15], [16], [17], [18], without necessarily involving the field information.

Since the primary goal of the GEDI mission is aboveground biomass (AGB) estimation [19], it is essential to have at least some collocated lidar footprint acquisitions and in situ measurements [20] to parameterize the GEDI waveform-AGB models [21]. The same condition must be fulfilled to allow the integration of GEDI waveforms into the MSNFI framework. Nearly all the methods allowing to respond to this requirement assume the simulation of the GEDI relative height profile at the plot starting from the coinciding airborne lidar point cloud [22], [23], [24]. The exception is the method recently proposed by [25], which is based on a spatial model for waveform prediction using the principal components of the GEDI relative height (RH) metrics space. Adhering to a similar approach and building upon the matching strategies introduced by [26], [27], we propose an alternative method for the interpolation of the GEDI measurements to the French NFI plots. This method also does not require simulation from locally acquired airborne lidar data, and aims to interpolate the entire relative height profile to the NFI plots and not only the canopy height.

The method presented in this article relies on using Sentinel-2, the forest stand type map and the digital terrain model as auxiliary input data. It combines the machine learning routines of time series K-means clustering, multilayer perceptron (MLP) classification and MLP regression in a way that allows to project at every considered NFI plot, through the profile pairing effectuated by structural classes, GEDI RH vegetation cumulative energy profiles issued from the corresponding GEDI full-waveforms. The proposed pairing routine stands out for its focus on balancing geographical space (distance between NFI plots and GEDI footprints) and feature space (similarity between GEDI RH vegetation cumulative energy profiles predicted at NFI plots and GEDI footprint locations). A model, based on the random forest regressor, is then used to link projected GEDI RH vegetation cumulative energy profiles and locally estimated wood volume stocks. The originality of the approach relies also on its scalability and flexibility: while classification alone could be used for poststratification purposes, regression and pairing could be used for downscaling further estimates though model assisted estimation.

The rest of this article is organized as follows. In Section II, we describe the study area, introduce the employed data, and describe the transformation of GEDI waveforms leading to the vegetation cumulative energy profiles used in the method development. In Section III, we provide a detailed explanation of the main steps of the method (see Fig. 2): clustering, classification, and regression (collectively referred to as Step I), along with the concept of profile pairing (referred to as Step II) and the volume model. Section IV provides the description of the validation framework, and the results of the validation both in terms of the profile matching and the wood volume estimation. Section V discusses the content. Finally, Section VI concludes this article.

II. Study Area and Data

The area chosen for the development and the validation of the method principally corresponds to the French sylvo-ecological

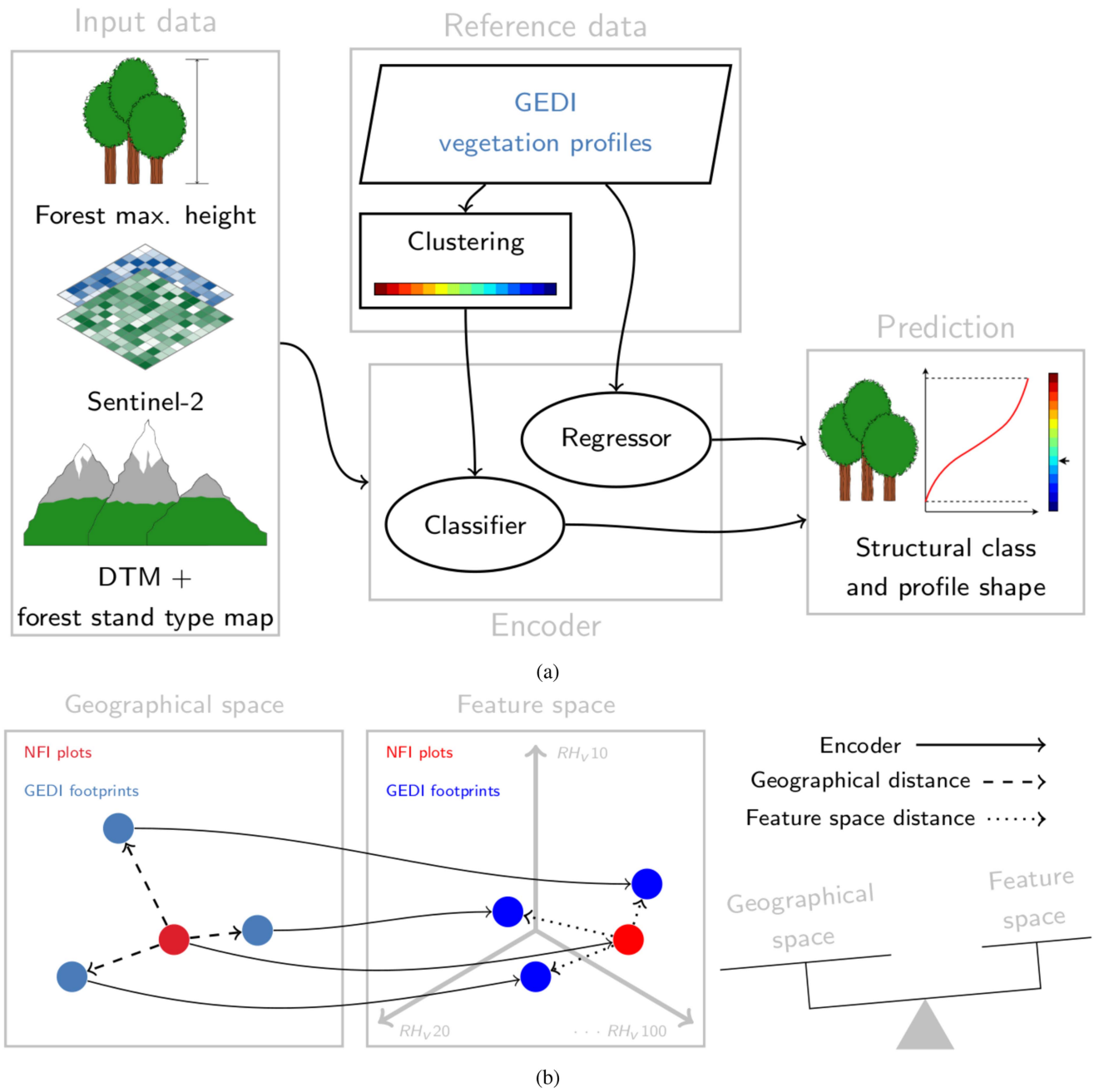

Fig. 2. Schemes presenting (a) Step I and (b) Step II of the proposed method.

region named "Central Vosges Mountains," located in the eastern part of the continental metropolitan France. Up to 75% of this mountainous area is covered by forest, which is broad-leaved below 500 m of altitude—mostly the European beech (*Fagus sylvatica*), toward the geographical limits of the study area; and coniferous above 500 m—mostly silver fir (*Abies alba*) and the European spruce (*Picea abies*), toward the center of the depicted study area. The area is known for the important wood industry and trade, which have been seriously affected by the bark-beetle epidemic driven by climate change [28], making this region a representative example of the need for forest monitoring at higher spatio-temporal resolution, in the light of the climate crisis.

A. NFI Data

The French NFI is continuous in both space and time, employing a two-phase stratified sampling design [29], [30]. This design uses a 1 km grid defined for a 10-year period, with $1/10$ of the grid surveyed each year. In the first phase, one point is randomly selected from the yearly grid sample (approximately 100 000 points total). The land use and land cover around these points are photo-interpreted using aerial photographs to estimate forest area. The second phase involves a subsample of the first phase's forest points (around 7000 points), which are surveyed in the field to estimate forest attributes. Field measurements are conducted in four circular concentric plots with radii of 6, 9, 15,

and 25 m, allowing to derive among a number of NFI attributes the two parameters used in this study.

1) The maximum height (H_{NFI})—defined as the maximum height at the NFI plot. Since height is measured for only a sample of trees (one per diameter class and species), missing values were imputed using a random forest method (MissForest [31]). This method was applied separately for each species and sylvo-ecological region, utilizing species, diameter at breast height, height, and plot-level variables (density, basal area, and volume). The imputation was validated using older data and resulted in a relative root-mean-square error (RMSE) of 13%.
2) Wood volume per hectare (V_{NFI})—estimated from the circumference at breast height, the height, the timber height, and the tree inclusion probability, by accounting for trees having a circumference at breast height greater than 7.5 cm.

The fundamental NFI measurements, such as wood volume per hectare, are subsequently poststratified based on an external criterion (the forest stand type map). This approach enables the estimation of wood volume per hectare values for specific areas over a five-year period. As discussed in the introduction, our overall goal is to reduce both this surface and the 5-year time span, without compromising the estimation accuracy, by incorporating remote sensing measurements. This article specifically focuses on the objective of projecting GEDI profiles to the NFI plots, an essential step in order to be able to utilize GEDI measurements to achieve this goal.

In the presented study, we therefore use H_{NFI} and V_{NFI} estimations from 529 NFI plots across the study region, acquired between 2017 and 2020.

B. GEDI Data

GEDI contains three lasers, two operating in the full power mode (one beam per laser) and one operating in the coverage mode (split into two beams) [11]. These four beams, operating at 1064 nm wavelength, produce footprints averaging 25 m in diameter on the ground, which are separated by 600 m across track and by 60 m along track.

We are using GEDI level 2 A products, providing: ground elevation, canopy top height, and relative height metrics. Relative height metrics indicate the height at which a specific percentile of returned energy is reached relative to the detected ground, specifically from the center of the ground peak. These metrics describe the shape of the normalized cumulative return energy, starting from the bottom of the ground return (with the center of the ground peak normalized to zero) to the top of the canopy (normalized to one).

The original GEDI relative height (RH) profiles, apart from the vegetation segment, also include the ground echo component. Since the latter was deemed irrelevant to the objectives of this article, we opted to remove it and base our method on profiles intended to solely capture the above-ground vegetation portion. To remove bad quality data both quality and degrade flags available in GEDI products were used [11]. Footprints with a quality flag of one (according to the L2A criterion, although the stricter L4A threshold could be applied if necessary, which was not the case for the ecosystems under study) and a degrade flag of zero were selected. GEDI RH profiles were further transformed into vegetation cumulative energy profiles, addressed as vegetation profiles across this manuscript (RH_v). The latter represent the relative height metrics at 1% intervals for the vegetation component of the backscattered signal. To that aim, a simplified waveform was reconstructed from RH values and the ground component of the signal was removed by adjusting a Gaussian function to the ground return before subtracting this function from the waveform. Finally, the resulting vegetation waveform was back transformed into an RH type of profile. An assessment of the cover rate was obtained through the comparison of original GEDI RH profiles and corresponding vegetation profiles.

In this study, after filtering, we have access to 185 725 GEDI footprints, each accompanied by vegetation profiles, obtained from areas designated as forest according to the forest stand type map [32]. 100 750 out of these are issued from "Full Power" beams and 84 975 are acquired in the "Coverage" mode, in the period ranging from 2019 to 2021.

C. Other Data

Aside from NFI and GEDI data, the study uses multispectral Sentinel-2 summer and winter acquisitions (year 2021), as well as the digital terrain model (DTM) mean elevation, slope and aspect. Regarding Sentinel-2 data, we used Theia level 3 A products. These products, generated using the weighted average synthesis processor (WASP) [33], offer monthly cloud-free syntheses of the Sentinel level 2 A product. The employed DTM is generated at a 1 m resolution using a triangulated irregular network algorithm applied on the airborne laser scanning (ALS) data acquired by the French National Mapping Agency over the study area in 2014 [34].

As previously mentioned, we also included the vector forest stand type map, which is used in both the development of the method (partially relying on polygon surfaces) and the validation framework (utilizing polygon classes) [32].

All three mentioned auxiliary datasets cover both NFI plots and GEDI footprints.

III. Method

When a GEDI footprint that corresponds geographically to an NFI plot location is absent, it becomes necessary to identify an alternative representative GEDI footprint. According to the criteria discussed in Section IV-A, this applies to 94% of the NFI data used in this study. This task is achieved by training an encoder model to predict a structural class and to map predictor variables into a 10-D feature space consisting of RH metrics [see Fig. 2(a)]. Within this feature space and a specific class, samples demonstrating similarity (i.e., forests with similar structures) are positioned close to each other [see Fig. 2(b)]. Subsequently, we can encode the predictor data at the NFI plot location into the feature space and conduct a search for nearby GEDI samples based on both feature similarity and geographical proximity. This methodology has been developed in two steps.

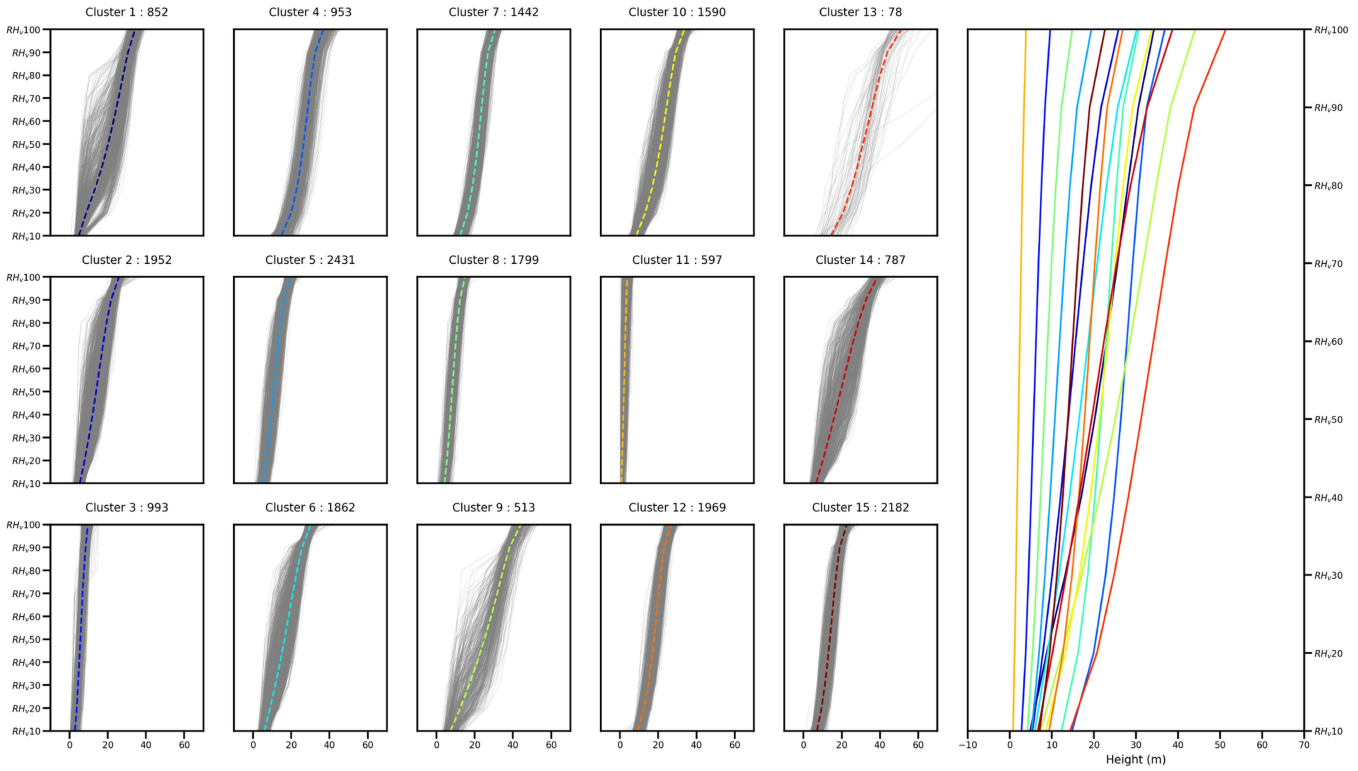

Fig. 3. Defining the structural class of the vegetation profile through the clustering of its subsample. The vegetation profiles are depicted in gray, while the mean vegetation profiles representing different clusters/structural classes are traced in nongray nuances.

A. Step I

This part of the method, depicted in Fig. 2(a), starts by defining vegetation profile structural classes (Section III-A1). It further relies on the auxiliary datasets introduced in the previous section to predict the class among the beforehand defined ones (Section III-A2) at locations corresponding to every NFI plot in the area, as well as at locations corresponding to every footprint of a subsample of vegetation profiles. Aside from the structural class, we as well predict at the very same locations the shape of the vegetation profile (Section III-A3)—consisting of 10 RH values ($\mathrm{RH}_v = [\mathrm{RH}_v10,\ \mathrm{RH}_v20,\ \cdots \mathrm{RH}v100]$).

1) Clustering: Defining Profile Structural Classes: The vegetation profile structural classes are defined through the unsupervised clustering of randomly selected vegetation profiles—the subsample *"training n°1"* (see Fig. 3). The method we apply is the time series k-means clustering based on the dynamic time warping (DTW) principle [35], [36], [37]. We basically treat the vegetation profile composed of ten relative height values as the time series, with the idea of the DTW being to align profiles such that their Euclidean distance ($||\cdot||$) is minimal.

If two considered vegetation profiles are annotated as

$$\mathrm{RH}_v^{(1)} = \left[\mathrm{RH}_v^{(1)}10,\ \mathrm{RH}_v^{(1)}20,\ \cdots \mathrm{RH}v^{(1)}100\right] \quad \text{with } [10,\ 20,\ \cdots 100] \text{ being } n_1, \text{ and}$$
$$\mathrm{RH}_v^{(2)} = \left[\mathrm{RH}_v^{(2)}10,\ \mathrm{RH}_v^{(2)}20,\ \cdots \mathrm{RH}v^{(2)}100\right] \quad \text{with } [10,\ 20,\ \cdots 100] \text{ being } n_2 \tag{1}$$

their DTW distance ($D(100, 100)$) is calculated recursively, using the formula

$$D(n_1, n_2) = \left|\left|\mathrm{RH}_v^{(1)}n_1 - \mathrm{RH}_v^{(2)}n_2\right|\right| + \min\left\{\begin{matrix} D(n_1 - 1, n_2) \\ D(n_1 - 1, n_2 - 1) \\ D(n_1, n_2 - 1) \end{matrix}\right\} \tag{2}$$

with the initial condition being $D(10, 10) = ||\mathrm{RH}_v^{(1)}10 - \mathrm{RH}_v^{(2)}10||$.

The DTW k-means clustering is, unlike the approach based on the Euclidean distance metric, therefore less sensitive to shifts between profiles, meaning that it should be able to assemble similar shapes of lidar returns even if the canopy height (RH_v100) varies among them. In practice, as shown in Fig. 3, the canopy height still plays a significant role in defining the clusters, although it is logically less important than it would be with the conventional k-means Euclidean approach.

The number of clusters ($N = 15$) is determined using the elbow technique, which comes down to identifying the elbow in the curve depicting the within-cluster sum of square (WCSS) as a function of number of clusters [38].

The parallel effort to propose supervised model-based clustering, which involves defining a sigmoid function with two degrees of freedom—the translation for the canopy height and the slope for the shape of the lidar return—resulted in fairly good performance when clustering original GEDI RH profiles. However, it proved to be less effective than the previously

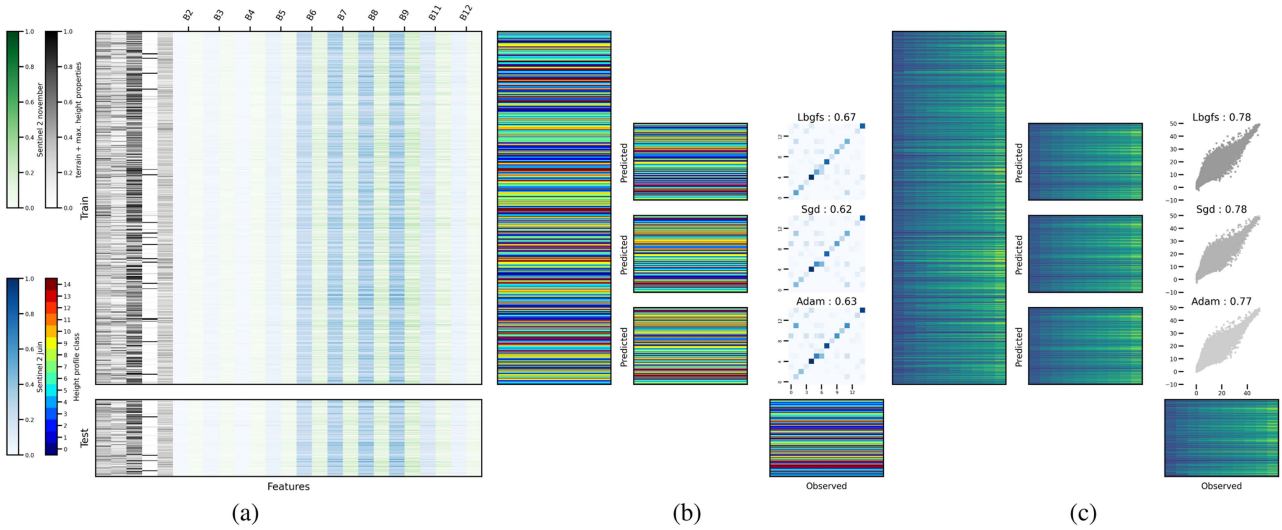

Fig. 4. MLP classifier and regressor. (a) Input data. (b) Classifier training and test data, as well as the confusion matrices for three different solvers. (c) Regressor training and test data, as well as the point clouds for three different solvers.

described unsupervised approach for the vegetation profiles used in this work.

2) Classification: Predicting Profile Structural Classes: The classification step assumes predicting the vegetation profile structural class using the auxiliary data available both at the locations of GEDI footprints and at NFI field plots. The method we apply is the multilayer perceptron classifier [39], [40], which serves as the first part of the encoder, as shown in Fig. 2(a). The reference data consist of clustering labels (i.e., structural classes) from subsample *"training n°1."* The input data include [see Fig. 4(a)] the following.

1) The canopy height indicator: RH_v98 is used during the training and test when only vegetation profiles are employed, and H_{NFI} as defined in Section II-A when predicting the structural classes at NFI plots.
2) Sentinel-2 frequency bands (B2, B3, B4, B5, B6, B7, B8, B9, B11, and B12) corresponding to the data acquired in the summer (June 2021) as well as the data acquired in winter (November 2021) [41].
3) Size of the polygon in the forest stand type map that contains the point of interest (whether it is GEDI the footprint or the NFI plot) [32].
4) The attributes of the DTM, which include: mean elevation, slope, and aspect.

At some point, we tested a series of spectral indices at the input instead of directly introducing the values of different spectral bands, but this did not produce any significant difference in terms of classifying performances.

The employed MLP classifier was parameterized separately, through the optimization procedure, for each of the three solvers [see Fig. 4(b)]: limited-memory BFGS optimizer (LBFGS), stochastic gradient descent (SGD), and stochastic gradient-based optimizer (Adam). This is done using vegetation profiles structural classes (without NFI plots), more precisely using 80% of the data subsample n°1 (i.e., 16 000 profiles). When applied on 20% of the used subsample, the first solver achieved a test accuracy of 67% (percentage of well predicted classes), by slightly outperforming the others, and was therefore applied further on in the method. The retained classifier is therefore based on the LBFGS solver, has three hidden layers, the rectified linear unit function as an activation function, and uses the adaptive learning rate.

The classifier trained and tested using vegetation profile classes is then applied at locations corresponding to NFI plots [see Fig. 5(a)] and GEDI footprints [see Fig. 5(b)]. It is important to clarify here that the classifier is not applied to the subsample used for its training (and testing), but rather to another randomly selected subsample, referred to as the *"application"* sample, of the same size (20 000 footprints).

3) Regression: Predicting Profile Shapes: The regression step consists of predicting the vegetation profile shape, i.e., ten relative height values ($\mathrm{RH}_v = [\mathrm{RH}_v10, \mathrm{RH}_v20, \cdots \mathrm{RH}v100]$), using the auxiliary data available both at locations of GEDI footprints and at NFI field plots. The method we apply is the multilayer perceptron regressor [39], [40], designated as the second part of the encoder in Fig. 2(a). It uses a random subsample of vegetation profiles (subsample *"training n°2"*) as the reference data and shares the same composition of input data as the previously described classifier [see Fig. 4(a)].

In the nearly equivalent procedure as the one described in the previous section, the regressor was parameterized simultaneously for three solvers [see Fig. 4(b)]: LBFGS, SGD, and Adam. The first one achieved the best performances—model explaining 78% of the variance ($R^2 = 0.78$) when applied on the test dataset (i.e., 20% of the dataset). The retained regressor is thus based on the LBFGS solver, has three hidden layers, with the hyperbolic tangent activation function, and uses the constant learning rate.

The regressor trained and tested using vegetation profiles is then applied at locations corresponding to NFI plots [see

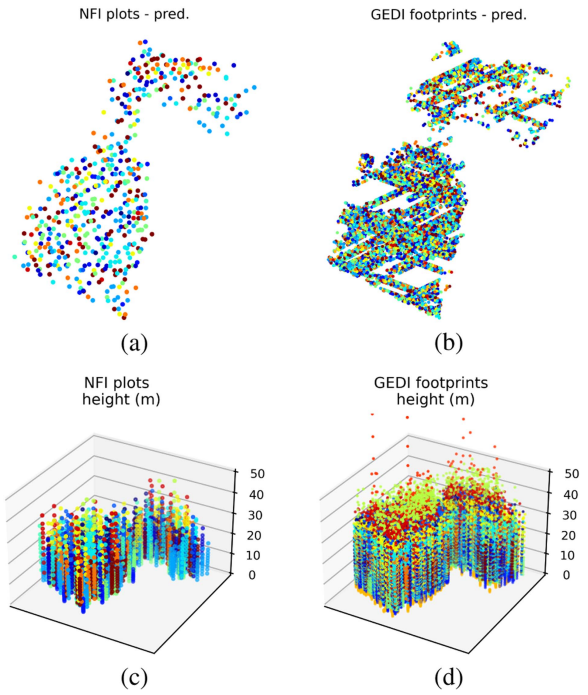

Fig. 5. Prediction of (a) vegetation profile structural classes at NFI plots, (b) vegetation profile structural classes at locations corresponding to GEDI footprints, (c) vegetation profile relative heights (RH_v) at NFI plots, and (d) RH_v at locations corresponding to GEDI footprints.

Fig. 5(c)] and GEDI fooprints [see Fig. 5(d)], which were already characterized by a vegetation profile class (subsample *"application"* in the classification step).

B. Step II

Once we have the predicted class and shape for each plot and footprint, we proceed to the pairing procedure. This is performed within the given structural class, considering both the geographical distance and the similarity between the predicted shapes (at the NFI plot versus the GEDI footprint). This process results in projecting real vegetation profiles onto NFI plots, using the predicted ones as a proxy [see Fig. 2(b)].

Previously described predicting of the vegetation profile class and shape at locations where we have actual GEDI measurements (subsample *"application"*), could seem counter-intuitive. It is presumed to be necessary given that the core part of the proposed method refers to the pairing of the predicted profiles within the common class. The assumption is therefore that it is more suitable to perform pairing among the synthesized, predicted profiles than between the synthesized and the real ones. This means that the predictions of both class and shape serve only as an intermediary product, acting as a proxy to associate a real vegetation profile with each NFI plot. The predicted structural class could find its utility in the poststratification phase of the NFI inference phase.

So, what would be the most suitable way of pairing the predicted profiles within one structural class?

1) Should one be very confident in the classification step and therefore only consider the closest profile in terms of geographical distance (nearest neighbor), assuming that the classification sufficiently homogenized the profiles [see Fig. 6(a)]? Perhaps, even though we acknowledge that neither the clustering (e.g., class 13 in Fig. 3) nor the classification (with a test accuracy score of 67%) are flawless in terms of performance.
2) Alternatively, should one choose to entirely disregard the intuitive proximity criterion and instead rely solely on the similarity between predicted profiles, i.e., the nearest profile in the feature space, regardless of the geographical distance separating them within the specified region and class [see Fig. 6(b)]? Perhaps, even if we know that the regression (test $R^2 = 0.78$) is not perfect either, and that the vicinity nevertheless can be a strong indicator of dealing with the very similar forest stand.

Unable to resolve this dilemma, we decided to take both into account [see Fig. 2(b)]. Namely, we opted to explore the tradeoff between the geographical distance (d_g) and the feature space distance (d_{fs}) by defining the weighted distance

$$d = w \cdot d_{\text{fs}} + (100 - w) \cdot d_g \tag{3}$$

with w being the weight ranging from 0 to 100. The geographical distance is the Euclidean distance between two coordinates, while the feature space distance uses the DTW principle introduced in Section III-A1. The latter implies that the d_{fs} is defined in the same way as D with the only difference that we now compare the predicted rather than the original vegetation profiles. The distance d is considered within a class as predicted in Step I, meaning we calculate the distance and perform pairing only between the NFI plots and the GEDI footprints belonging to the same class. Due to the difference in magnitude, both d_g and d_{fs} are normalized with respect to their maximum values within a class.

On the example illustrated in Fig. 6(c) we can indeed see that as we move from $w = 0$ ($d = d_g$) to $w = 100$ ($d = d_{\text{fs}}$), we gradually release the geographical proximity constraint and give more space to the similarity in the feature space.

As for the choice of using the DTW rather than the Euclidean distance, we aim at reproducing the vegetation profile and therefore prefer not to penalize the pairing of profiles very similar in shape due to a potentially minor offset in terms of total height.

Given the significance of geographical distance and the geolocations of NFI plots and GEDI footprints, it is crucial to note that the horizontal geolocation error of GEDI is estimated at 10.2 m [42]. In contrast, the relative mean error for the French NFI plots is 3.7 m, obtained by comparing the original position estimate with the one determined during routine quality control between 2008 and 2017. This estimate is based on a combination of photo-interpretation, chainage, and GPS, with the latter being particularly imprecise in forestry environments [43], [44].

To effectively utilize both geographical distance and feature space distance using the weighted distance defined in (3), we must identify the optimal value of the weight w. This basically involves determining the extent to which we should depend on geographical distance versus feature space distance, and inherently provides valuable information about the homogeneity of the forest stand in the area of application.

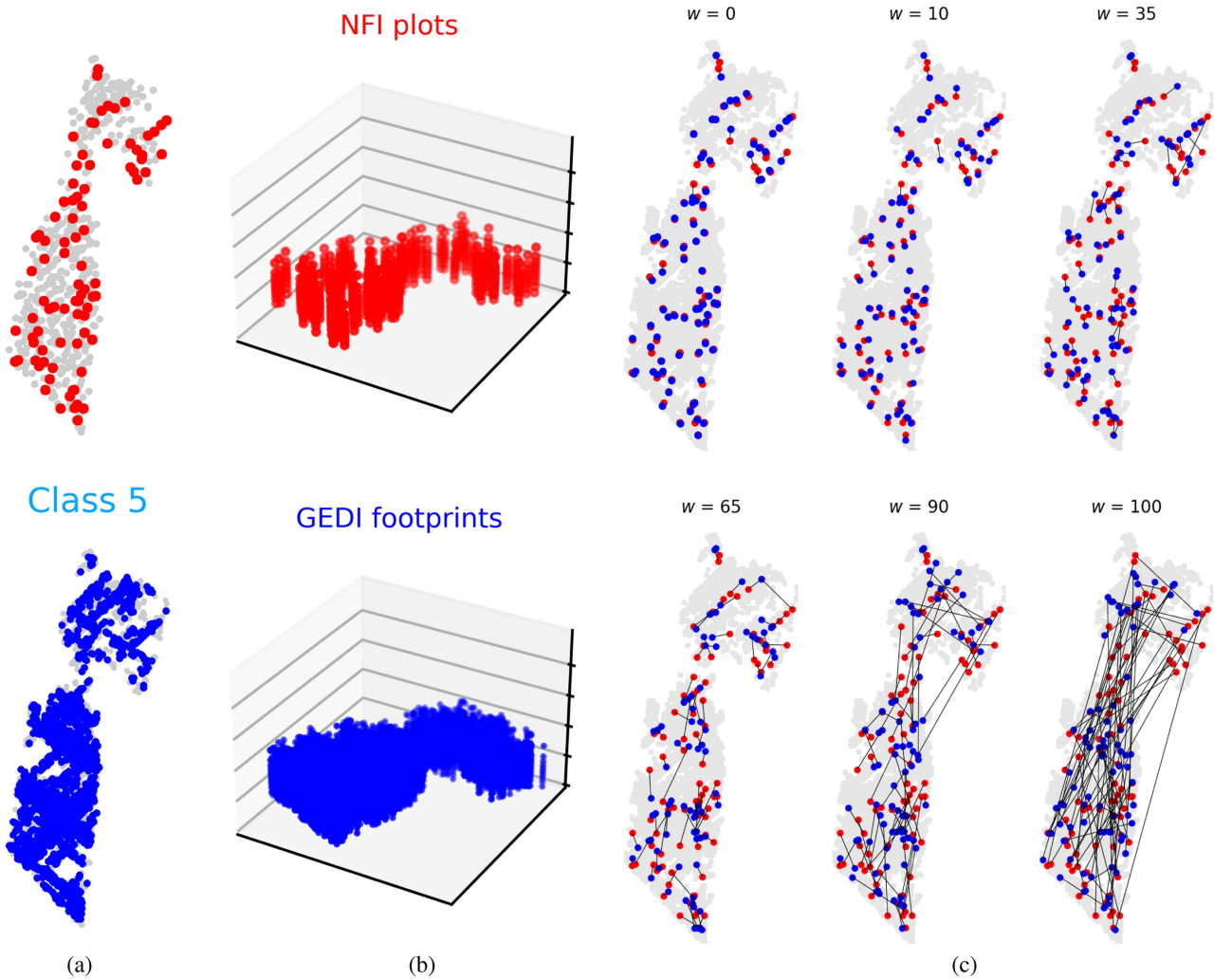

Fig. 6. Example of the GEDI-NFI pairing (within class 1). (a) Locations of NFI plots (red) and GEDI footprints (blue). (b) Predicted shape of vegetation profiles (RH_v) at locations corresponding to NFI plots (red) and GEDI footprints (blue). (c) Six ways of pairing profiles corresponding to six different values of w.

Since Step II aims to enhance the performance of directly estimated profiles from Step I, the optimization of w is conducted within the validation framework (see Section IV-A) used to evaluate the method's performance. Essentially, this involves using the pairing mechanism to improve the baseline results achieved by the machine learning method in Step I.

The decision to base the method constitution phase (Steps I and II) on a randomly selected sample of 20 000 footprints, rather than using all available data (185 725 footprints), was made primarily for time efficiency. Additionally, this approach aims to demonstrate the potential of the proposed method by highlighting its effectiveness with a sparser GEDI footprint coverage.

At the end of this step we have a projection of the actual vegetation profiles for each of the NFI plots in the area of application.

C. Wood Volume Modeling

Once we have obtained projections of vegetation profiles at NFI plots, we can proceed to establishing a link between the remotely sensed "measurement" and various forest attributes measured or estimated at the plot. As already mentioned in Sections I and II-A, the forest attributes we are particularly interested in are the wood volume and the AGB. While waiting for the formalization of the novel biomass estimation protocol at French NFI plots, we settled upon the wood volume model for illustrating the benefits of the proposed method.

The method we apply in order to link the projected profiles and the locally estimated wood volume (V_{NFI}) is the random forest regressor [39], [45]. The method is optimized for two different variants, i.e., relating the in situ estimated wood volume to:

1) GEDI-issued vegetation profiles—RH_v (10 deciles);
2) GEDI profiles top value—$\mathrm{RH}_v 100$.

Principal component analysis (PCA) is applied beforehand to address the multicollinearity of RH_v metrics, retaining 7 out of 10 components. The final regressor uses 400 trees, with nodes expanding until all leaves are pure or contain fewer than two samples, and bootstrap samples are used in building the trees.

For every variant, the resulting models are applied to all available GEDI footprints in the considered area.

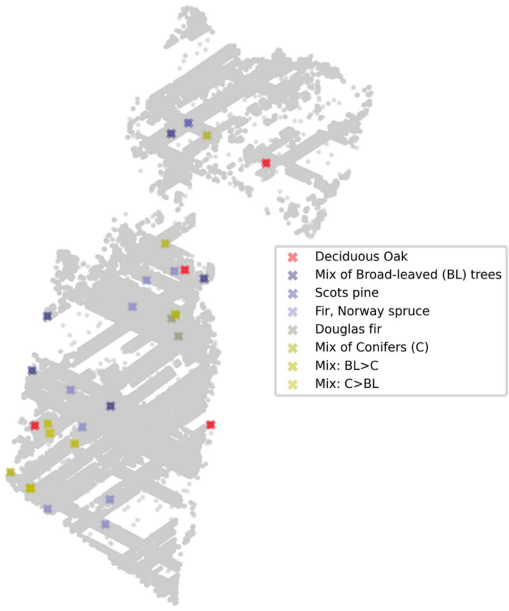

Fig. 7. Validation framework: Almost coinciding NFI plots and GEDI footprints, as well as the description of the forest stand in question [32].

IV. Results

The results presented in this section pertain to the efficacy demonstrated by the proposed method in interpolating accurate vegetation profiles to the NFI plots, as well as the performance of wood volume interpolation from the NFI plots to the GEDI footprints. Both assessments are enabled by the validation framework introduced in Section IV-A.

In Section IV-B, we verify the performance of predicted vegetation profiles, as detailed in Step I of the method presented in Section III (dashed line trajectory in Fig. 1). Section IV-C describes the optimization of the weight w necessary for the application of Step II of the method, and finally presents the results of validating the final projections, derived at the conclusion of Step II of the method (full line trajectory in Fig. 1).

A. Validation Framework

Validation framework is made possible by the fortuitous spatio-temporal quasi-coincidence between certain NFI plots and the portion of GEDI footprints. This allowed us to identify 32 NFI plots (6%), addressed further as the validation plots, where we can assume with a relative certainty that we know what the projected vegetation profile should look like (see Fig. 7). This identification is done by respecting the following criteria establishing the coincidence between 32 plots and corresponding footprints.

1) *Distance:* The geographical distance separating the plot and the footprint should not exceed 40 m, which we found to be a compromise between what would be the expected forest stand spatial auto-correlation length and the size of the validation sample. That is to say, more than 40 m would increase the risk of not dealing with the extremely similar part of the stand, and less would simply leave us with insufficient number of pairs in the sample to allow a proper statistical analysis. The analysis carried out in the forest of Sologne (Central France) and presented in [27] shows that the semivariogram of GEDI measurements reached the horizontal asymptote at the distance corresponding to 500 m, meaning that the 40 m maximum distance represents a fairly rigorous choice.
2) *Forest stand type class:* Both the plot and the footprint should belong to the same polygon of the vector forest stand type data base [32].
3) *Height check:* Due to the geo-localization issues which can characterize GEDI footprints [42], [46], we included an extra verification which refers to the height difference $|H_{\text{NFI}} - \text{RH}_v 98|$ which should not exceed 4 m.

Validation of profiles consists of observing the concordance between the ensemble of vegetation profiles predicted (Section IV-B) or projected (Section IV-C) at the NFI validation plots (RH_v predicted/projected) and the reference profiles coming from the validation footprints (RH_v "observed"). We evaluate this concordance globally, by comparing all relative height values at all validation plots at once, and in the more stratified manner, by comparing one relative height ($\text{RH}_v n$) at a time at all validation plots.

Validation of estimated wood volume refers to the comparison of the wood volume estimation at GEDI footprints and the wood volume estimated locally at NFI plots for the 32 validation plot-footprint pairs. The GEDI footprints—NFI wood volume pairs corresponding to the validation plots were evidently kept out of the construction of the wood volume models.

B. Validation of Predicted Profiles (Step I)

As what it concerns the validation of predicted profiles, i.e., before the pairing [see Fig. 8(a)], the global evaluation demonstrates the coefficient of determination scores (R^2) rising up to the value of 0.88. It is the "stratified" evaluation that we found particularly interesting. Namely, the coefficients of correlation ($r_{\text{RH}_v n}$) for the lower part of the considered relative heights (RH10 $-$ RH50) do not drop below 0.5. Although there is some dependence between the metrics, a slight decrease in correlation is observed as we move down the profile. This is expected, considering the auxiliary data used does not provide direct information about the lower part of the stand.

The results for the so-called two variants of the final wood volume estimation, introduced and enumerated from a to b in Section III-C are presented in two panels of Fig. 8 ranging from b to c. Depicted point clouds, as well as the corresponding coefficients of determination (R^2), normalized relative mean square error (NRMSE), and mean bias error (MBE), allow us to deduce the following.

1) Step I of the method allows building the GEDI profiles-based model explaining 53% of the wood volume variance ($R^2 = 0.53$), with 44.99% of NRMSE and 21.43 m^3/ha of MBE. One could rather say at least, knowing that the employed random forest model was not subject to an extensive optimization given its somewhat auxiliary role as the demonstrator of the GEDI measurements interpolation efficiency, the latter being the "raison d'être" of this

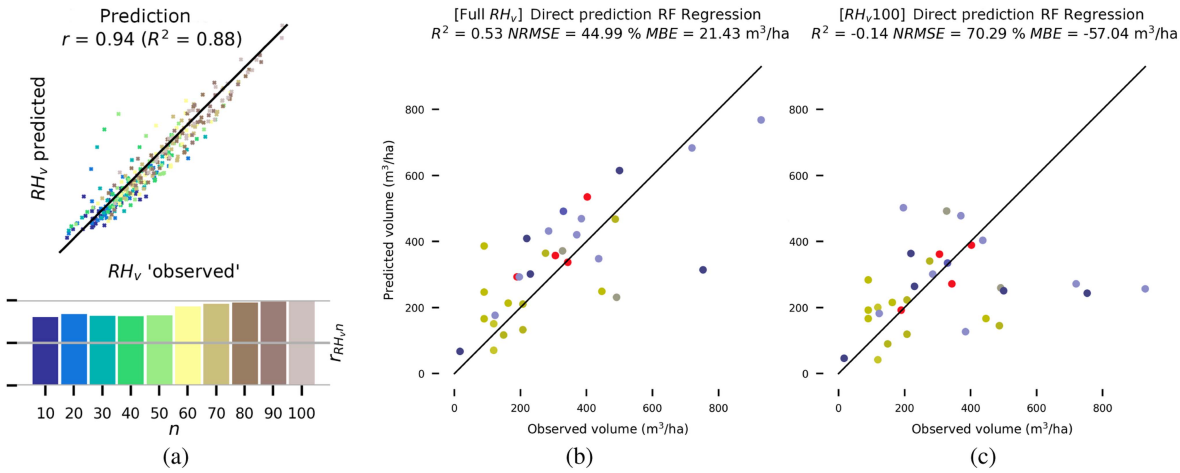

Fig. 8. Validation of predicted profiles (Step I). (a) Validation of predicted profiles, (b) validation of estimated wood volume using predicted profiles, and (c) validation of estimated wood volume using only profile top values. Different point colors in the upper part of subfigure (a) correspond to various metrics, as indicated by the bar plot in the lower part.

article. However, these results are far from satisfactory, particularly due to the notably high value of the mean bias estimate.

2) Even so, results are far better when using the ten relative height values (Full RH_v) than when using only the canopy height (RH_v100), where the employed random forest regressor fails to do better than the mean estimate—negative R^2 [see Fig. 8(c)].

C. Validation of Projected Profiles (Step I + Step II)

The method (Step I + Step II) is iteratively run for every value of w ranging from 0 to 100, collecting scores based on Pearson's correlation coefficient between "observed" and projected RH_v, as well as the coefficient of determination (R^2), NRMSE, and MBE of the resulting wood volume model. Displayed in Fig. 9, the scores indicate that the optimal values of the wood volume R^2 and NRMSE for the considered subsample *"application"* are located either in the midpoint of the possible range of w values, halfway between the nearest neighbor and the nearest profile, or toward the end of the range, at the nearest profile. However, when we add the wood volume MSE to the equation, we restrain the selection to the former subrange, ultimately choosing $w = 47$.

After performing the pairing with the optimal weight selected in the previous section ($w = 47$), we repeat the same analysis as in Section IV-B, but with actual vegetation profiles. It is important to note that the matching of NFI plot and GEDI footprint acquisition years did not significantly impact the pairing procedure (e.g., out of 123 NFI plots from 2019, only 40 (33%) were paired with the GEDI shots from the same year).

In the part concerning the validation of projected profiles [see Fig. 10(a)], we observe results similar to those obtained with the predicted profiles. Specifically, the global evaluation coefficient of determination (R^2) rises to 0.85, and the "stratified" evaluation shows that the correlation coefficients ($r_{\mathrm{RH}_v n}$) for the lower range of the considered relative heights ($\mathrm{RH}10 - \mathrm{RH}50$) do not drop below 0.5.

The results for the two variants of the final wood volume estimation are presented in two panels of Fig. 10 ranging from b to c, allowing us to deduce the following.

1) Step I + Step II of the method allows building the GEDI profiles-based model explaining up to 65% of the wood volume variance ($R^2 = 0.65$), with 39.10% of NRMSE and only 2.31 m^3/ha of MBE. These results are far better than the ones obtained after the direct predictions (Section IV-B), especially in terms of the mean bias estimate, which is particularly relevant when dealing with the forest resources estimation.
2) Results are significantly better when using the ten relative height values (Full RH_v) compared to using only the canopy height RH_v100 [see Fig. 10(c)].

V. Discussion

Despite the loss of correlation in the lower RH values, our approach has the key advantage of proposing a match for the entire set of NFI plots. With our strict approach in building the validation framework (Section IV-A), based on distance (i.e., 40 m, ±4 m, same forest polygon), only 32 matches were obtained, representing 6% of the plots available in the area. Using a 200 m distance and a 2 m height thresholds, [26] matched 69% of the NFI plots in their area. Using distance thresholds of 100, 300, and 500 m following a semivariogram analysis, [27] were able to match respectively 22%, 52%, and 57% of the NFI plots. Also, such result did not guarantee that the matched data are representative of the distribution of forest resources [27], with possible impact on models and associated inferences. In this prospect, our approach could be considered as a milestone for the development of models of forest attributes using GEDI and NFI data.

The capability to predict field attributes at locations corresponding to the entire set of GEDI footprints over an area of interest, highlights the potential of the approach for post-stratification purposes [47]. It is also of interest for grid-level (i.e., 41 km^2 hybrid-inference approaches [19], [48], in order to

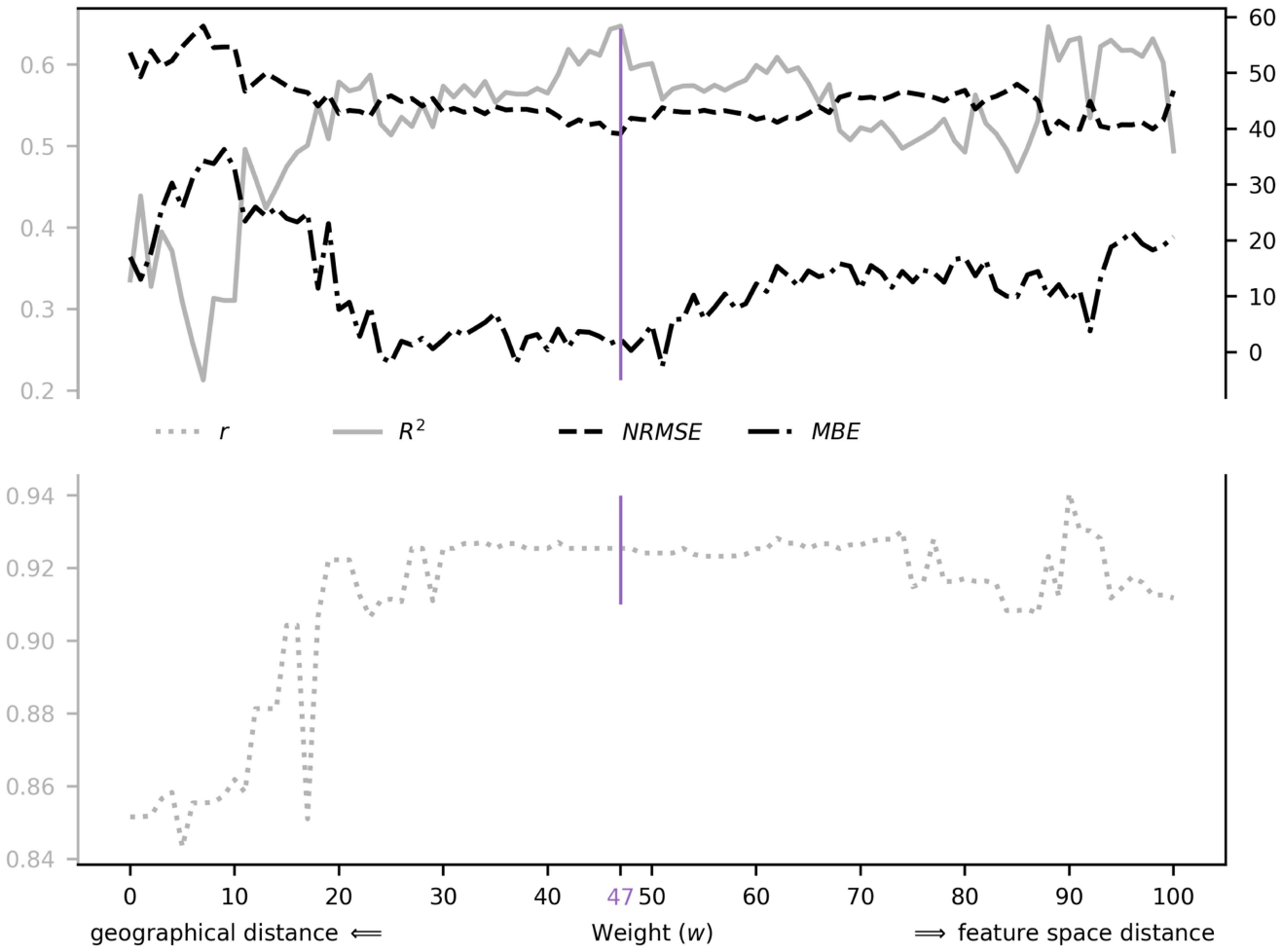

Fig. 9. Weight w optimization (Step II), conducted using validation performance criteria: R^2, NRMSE, and MBE, representing the coefficient of determination, NRMSE, and MBE of the resulting wood volume model, respectively. r denotes the Pearson's correlation coefficient between "observed" and predicted RH_v.

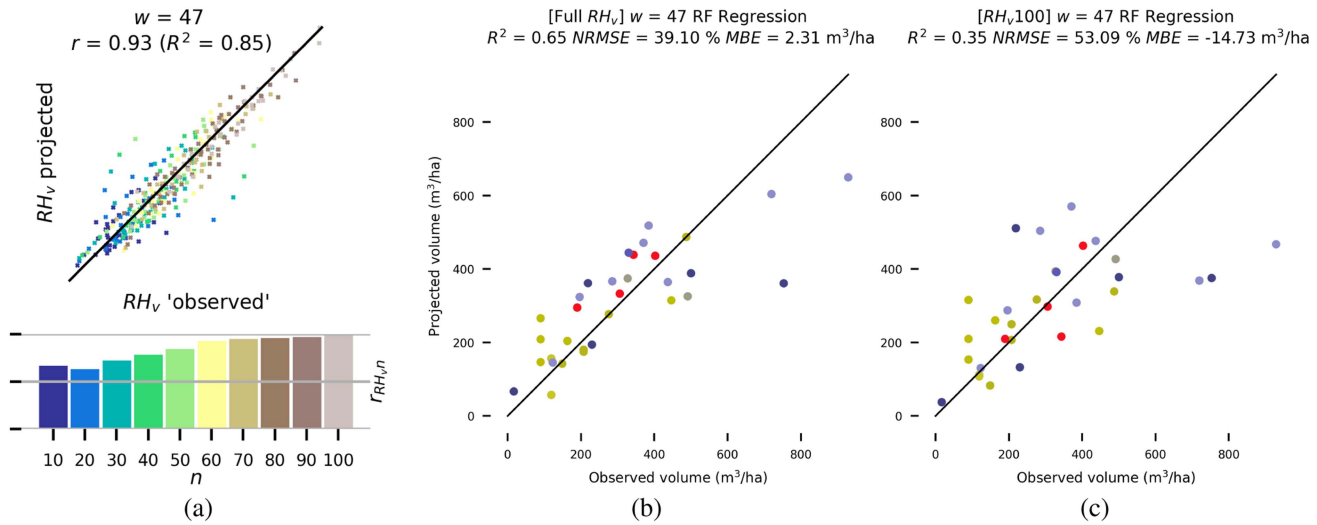

Fig. 10. Validation of the projected profiles (Step I + Step II). (a) Validation of projected profile shapes, (b) validation of estimated wood volume using projected profiles, and (c) validation of estimated wood volume using only profile top values. Different point colors in the upper part of subfigure (a) correspond to various metrics, as indicated by the bar plot in the lower part.

estimate various forest attributes of interest for NFIs. Another interesting application is related to the high resolution mapping of forest attributes using deep learning approach. Indeed, by generalizing the NFI information to the GEDI footprints, it is possible to use deep learning approaches to directly map NFI attributes using optical or radar data, instead of height, such as in, e.g., [13], [16], and [17]. Such approach might be more efficient than those attempting to predict either volume or biomass from GEDI-based height maps [49]. Though this seems obvious, we find it important to state clearly the necessity for the community to focus less on the canopy height interpolation or extrapolation which alone represents only a fraction of the GEDI data potential. The presented results demonstrate this clearly: as shown in Fig. 10, applying a similar random forest regressor reveals that the use of 10 decile vegetation profiles (Full RH_v) significantly outperforms using only the RH_v100. This is evidenced by a gain of 0.3 in R^2, a 14% reduction in NRMSE, and an improvement of over 12 m^3/ha in MBE.

One could notice that the method is indeed somehow conditioned by the geographical limits of the study. This is consistent with our ambition to expand the application of the proposed work all across the metropolitan France, which is split into 91 sylvo-eco regions (out of which 86 nonalluvial), meaning that the method would be independently parameterized, i.e., locally adjusted in each of them.

As it makes sense to redo the weight optimization in different sylvo-ecological regions, it is also sensible to redo it after updating the GEDI dataset with upcoming measurements. The new measurements could influence the observed balance between geographical and feature space. The same principle applies when using a different sample size or the entire set of available GEDI footprints. Specifically, the curve used to determine the value of w (as shown in Fig. 9) is influenced by the sample size used in the study (subsample *"application"*). Employing a smaller or larger sample would lead to different tradeoffs between geographical and feature space.

Throughout the manuscript, we refer to the approach as an interpolation of GEDI measurements, given that the NFI plots are spatially situated between GEDI footprints and the datasets roughly coincide temporally. However, this method could also be used for extrapolation, especially in the temporal sense, which is particularly relevant for its application in the MSNFI framework.

Finally, we attempted to develop a model that does not require the canopy height indicator at the input. Currently, the inclusion of this indicator restricts the interpolation to NFI plots. However, the auxiliary data used proved absolutely insufficient for the model to reproduce the variance in vegetation classes without this information, i.e., without the canopy height indicator the employed classifier for example achieves the test accuracy of only 18%. This inherent limitation of the proposed method can be overcome by employing AI methods: either by enhancing the complexity of the proposed encoder (Step I), or by incorporating a spatially continuous height indicator such as the combination of height maps mentioned throughout the manuscript [50], or GEDI gridded RH100 metrics. The latter therefore should not eclipse the added methodological value of this work which dominantly consists in successfully merging the feature and geographical space when interpolating GEDI measurements. Additionally, the method offers scalability and flexibility, meaning that while the whole method can be used do downscale the estimates of forest attributes, one could also use classification alone for poststratification purposes.

VI. Conclusion

In this article, we sought to respond to the issue of the spatial mismatch between GEDI footprints and NFI plots, representing a major obstacle for the deeper integration of GEDI data into the MSNFI frameworks of countries whose forest ecosystems are covered by the mission. This was achieved by proposing a method for interpolating GEDI measurements to the NFI plots, which relies on the sequential use of clustering, classification, and regression machine learning routines, all integrated within a framework of profile pairing by structural class. The latter means that we were able to project, through the pretty much methodologically transparent pairing process, real non-synthesized GEDI-issued vegetation profiles to each NFI plot in the considered area. The fact that the method is capable of associating GEDI measurements to each NFI plot in the considered area makes it already distinctive with respect to most of the state-of-the-art approaches in dealing with the mismatch without simulating GEDI profiles from the local ALS data. This association allows further on to link GEDI data and the wood volume, or any other forest attribute of interest, by proposing GEDI-issued vegetation profile $\leftrightarrow$ wood volume models. The proposed method proves to be able to explain 85% of the variance of the projected profiles, allowing notably to project reasonably well the lower part of the vegetation profile. The resulting GEDI-wood volume ensemble model is capable to reproduce at least 65% of the wood-volume variance with a MBE of only 2.31 m^3/ha, illustrating the clear modeling benefit of using ten relative height values rather than the canopy height only.

More generally, the proposed method could be used to improve the calibration of GEDI-forest attribute formulas, particularly the GEDI-AGB ones. The step of the proposed method introducing the notion of the profile structural class could as well find its utility in the poststratification step of the (MS)NFI inference procedure.

Scanning the Earth surface from more than 400 km above the ground is challenging, and therefore, the principal asset of the GEDI mission is not the level of stability and precision usually required from national airborne lidar campaigns [51], [52], [53], [54], but its capacity to repeatedly cover extremely large areas. Therefore, the most immediate short-term goal of this work is to establish a framework that allows the simultaneous and, to some extent, perpetual application of the method in the other 85 nonalluvial sylvo-ecological regions of France (in addition to the Vosges mountains, which were used for the method's conception and presentation in this article), and its integration into the MSNFI framework. We are simultaneously planning to pursue the approach of reinforcing the validation step by running the GEDI simulator [22] over a subset of NFI plots with ALS data. The medium-term perspective would be to work around the key limitation of the presented method

which currently projects profiles only to the NFI plots, and by doing so to spatio-temporally densify the interpolation allowing the forest attributes high-resolution mapping application, which could further more improve the MSNFI spatial resolution. We plan to achieve this by: 1) enhancing the complexity of the proposed encoder; and 2) integrating a spatially continuous height indicator from an external source.

ACKNOWLEDGMENT

Above all, the authors would like to thank two anonymous reviewers whose comments and suggestions have been incredibly valuable in enhancing both the conceptualization and presentation of this study.

The authors would also like to thank IGN colleagues who collected, processed, and organized the NFI field data, as well as the GEDI mission science team for the public availability of GEDI data and their continuous extraordinary work on GEDI.

The English wording refinement was partly carried out using the publicly available GenAI (ChatGPT).

REFERENCES

[1] IPCC, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, U.K. and New York, NY, USA: Cambridge Univ. Press, 2021, vol. In Press.

[2] E. Tomppo, T. Gschwantner, and M. Lawrence, Eds., *National Forest Inventories*. Amsterdam: The Netherlands: Springer, 2010.

[3] R. E. McRoberts and E. O. Tomppo, "Remote sensing support for national forest inventories," *Remote Sens. Environ.*, vol. 110, no. 4, pp. 412–419, 2007, forestSAT Special Issue.

[4] Y. Yu et al., "Making the us national forest inventory spatially contiguous and temporally consistent," *Environ. Res. Lett.*, vol. 17, no. 6, May 2022, Art. no. 065002.

[5] N. C. Coops, P. Tompalski, T. R. H. Goodbody, A. Achim, and C. Mulverhill, "Framework for near real-time forest inventory using multi source remote sensing data," *Forestry: Int. J. Forest Res.*, vol. 96, no. 1, pp. 1–19, 2022.

[6] F. E. Fassnacht, J. C. White, M. A. Wulder, and E. Næsset, "Remote sensing in forestry: Current challenges, considerations and directions," *Forestry: Int. J. Forest Res.*, vol. 97, no. 1, pp. 11–37, 2023.

[7] E. Tomppo, M. Haakana, M. Katila, and J. Peräsaari, *Multi-Source National Forest Inventory: Methods and Applications.* New York, NY, USA: Springer Dordrecht, 2008.

[8] S. Durrieu, C. Véga, M. Bouvier, F. Gosselin, J.-P. Renaud, and L. Saint-André, "Optical remote sensing of tree and stand heights," in *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*. Boca Raton, FL, USA: Taylor & Francis, 2016, pp. 449–485.

[9] D. B. Irulappa-Pillai-Vijayakumar, J.-P. Renaud, F. Morneau, R. E. McRoberts, and C. Vega, "Increasing precision for french forest inventory estimates using the k-NN technique with optical and photogrammetric data and model-assisted estimators," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 991.

[10] C. Vega, J.-P. Renaud, A. Sagar, and O. Bouriaud, "A new small area estimation algorithm to balance between statistical precision and scale," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 97, 2021, Art. no. 102303.

[11] R. Dubayah et al., "The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography," *Sci. Remote Sens.*, vol. 1, 2020, Art. no. 100002.

[12] N. C. Coops et al., "Modelling Lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends," *Remote Sens. Environ.*, vol. 260, 2021, Art. no. 112477.

[13] P. Potapov et al., "Mapping global forest canopy height through integration of GEDI and landsat data," *Remote Sens. Environ.*, vol. 253, 2021, Art. no. 112165.

[14] D. Morin et al., "Improving heterogeneous forest height maps by integrating GEDI-based forest height information in a multi-sensor mapping process," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2079.

[15] S. Ge, H. Gu, W. Su, J. Praks, and O. Antropov, "Improved semisupervised UNet deep learning model for forest height mapping with satellite SAR and optical data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5776–5787, 2022.

[16] M. Schwartz et al., "High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 128, 2024, Art. no. 103711.

[17] N. Lang, W. Jetz, K. Schindler, and J. D. Wegner, "A high-resolution canopy height model of the earth," *Nature Ecol. Evol.*, vol. 7, no. 11, pp. 1778–1789, Nov. 2023.

[18] I. Fayad et al., "Hy-tec: A hybrid vision transformer model for high-resolution and large-scale mapping of canopy height," *Remote Sens. Environ.*, vol. 302, 2024, Art. no. 113945.

[19] R. Dubayah et al., "Gedi launches a new era of biomass inference from space," *Environ. Res. Lett.*, vol. 17, no. 9, Aug. 2022, Art. no. 095001.

[20] N. Labrière et al., "In situ reference datasets from the TropiSAR and AfriSAR campaigns in support of upcoming spaceborne biomass missions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3617–3627, Oct. 2018.

[21] L. Duncanson et al., "Aboveground biomass density models for Nasa's global ecosystem dynamics investigation (GEDI) lidar mission," *Remote Sens. Environ.*, vol. 270, 2022, Art. no. 112845.

[22] S. Hancock et al., "The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions," *Earth Space Sci.*, vol. 6, no. 2, pp. 294–310, 2019.

[23] A. Pascual et al., "Assessing the performance of Nasa's GEDI l4a footprint aboveground biomass density models using national forest inventory and airborne laser scanning data in mediterranean forest ecosystems," *Forest Ecol. Manage.*, vol. 538, 2023, Art. no. 120975.

[24] E. Aragoneses, M. García, P. Ruiz-Benito, and E. Chuvieco, "Mapping forest canopy fuel parameters at European scale using spaceborne lidar and satellite data," *Remote Sens. Environ.*, vol. 303, 2024, Art. no. 114005.

[25] P. M. May, R. O. Dubayah, J. M. Bruening, and G. C. Gaines, "Connecting spaceborne lidar with NFI networks: A method for improved estimation of forest structure and biomass," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 129, 2024, Art. no. 103797.

[26] E. L. Bullock et al., "Estimating aboveground biomass density using hybrid statistical inference with GEDI lidar data and Paraguay's national forest inventory," *Environ. Res. Lett.*, vol. 18, no. 8, Jul. 2023, Art. no. 085001.

[27] S. Zhang et al., "Modelling forest volume with small area estimation of forest inventory using GEDI footprints as auxiliary information," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 114, 2022, Art. no. 103072.

[28] A. Gilles, J. Lisein, J. Cansell, N. Latte, C. Piedallu, and H. Claessens, "Spatial and remote sensing monitoring shows the end of the bark beetle outbreak on Belgian and north-eastern France Norway spruce (Picea abies) stands," *Environ. Monit. Assessment*, vol. 196, no. 3, Feb. 2024, Art. no. 226.

[29] N. Robert, C. Vidal, A. Colin, J. Hervé, N. Hamza, and C. Cluzeau, "French national forest inventory," in *National Forest Inventories*, E. Tomppo, T. Gschwantner, and M. Lawrence, Eds. Amsterdam, The Netherlands: Springer, 2010, ch. 12.

[30] J.-C. Hervé, S. Wurpillot, C. Vidal, and B. Roman-Amat, "L'inventaire des ressources forestières en France : Un nouveau regard sur de nouvelles forêts," *Revue Forestiére Française*, vol. 66, no. 3, pp. 247–260, 2014.

[31] D. J. Stekhoven and P. Bühlmann, "Missforest non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, Jan. 2012.

[32] Institut national de l'information géographique et forestière, "Bd forêt v2," 2018. [Online]. Available: https://geoservices.ign.fr/bdforet

[33] O. Hagolle, D. Morin, and M. Kadiri, "Detailed processing model for the weighted average synthesis processor (WASP) for Sentinel-2," Zenodo, Aug. 21, 2018, doi: 10.5281/zenodo.1401360.

[34] A. Sagar, C. Vega, O. Bouriaud, C. Piedallu, and J.-P. Renaud, "Multi-source forest inventories: A model-based approach using k-NN to reconcile forest attributes statistics and map products," *ISPRS J. Photogrammetry Remote Sens.*, vol. 192, pp. 175–188, 2022.

[35] H. Sakoe and S. Chiba, "A dynamic programming approach to continuous speech recognition," in *Proc. 7th Int. Congr. Acoust.*, 1971, pp. 65–69.

[36] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 43–49, Jan. 1978.

[37] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, 2011.

[38] R. L. Thorndike, "Who belongs in the family?," *Psychometrika*, vol. 18, no. 4, pp. 267–276, Dec. 1953.
[39] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[40] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1, pp. 503–528, Aug. 1989.
[41] F. Spoto et al., "Overview of sentinel-2," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 1707–1710.
[42] A. Schleich, S. Durrieu, M. Soma, and C. Vega, "Improving GEDI footprint geolocation using a high-resolution digital elevation model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7718–7732, 2023.
[43] C. Piedallu and J.-C. Gégout, "Effects of forest environment and survey protocol on GPS accuracy," *Photogrammetric Eng. Remote Sens.*, vol. 71, no. 9, pp. 1071–1078, 2005.
[44] C. O. Galán, J. R. Rodríguez-Pérez, J. M. Torres, and P. J. G. Nieto, "Analysis of the influence of forest environments on the accuracy of GPS measurements by using genetic algorithms," *Math. Comput. Model.*, vol. 54, no. 7, pp. 1829–1834, 2011.
[45] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[46] H. Tang et al., "Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI," *Remote Sens. Environ.*, vol. 291, 2023, Art. no. 113571.
[47] H. Haakana, J. Heikkinen, M. Katila, and A. Kangas, "Efficiency of post-stratification for a large-scale forest inventory–case Finnish NFI," *Ann. Forest Sci.*, vol. 76, no. 1, Jan. 2019, Art. no. 9.
[48] S. Saarela et al., "Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and landsat data," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1832.
[49] M. Schwartz et al., "Forms: Forest multiple source height, wood volume, and biomass maps in France at 10 to 30 m resolution based on Sentinel-1, Sentinel-2, and global ecosystem dynamics investigation (GEDI) data with a deep learning approach," *Earth Syst. Sci. Data*, vol. 15, no. 11, pp. 4927–4945, 2023.
[50] N. Besic et al., "Remote sensing-based high-resolution mapping of the forest canopy height: Some models are useful, but might they be even more if combined?," *Geoscientific Model Develop. Discus.*, vol. 2024, pp. 1–26, 2024. [Online]. Available: https://gmd.copernicus.org/preprints/gmd-2024-95/
[51] M. Nilsson et al., "A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the national forest inventory," *Remote Sens. Environ.*, vol. 194, pp. 447–454, 2017.
[52] M. Lang et al., "Remote-sensing support for the estonian national forest inventory, facilitating the construction of maps for forest height, standing-wood volume, and tree species composition," *Forestry Stud.*, vol. 73, no. 1, pp. 77–97, 2020.
[53] M. Hauglin, J. Rahlf, J. Schumacher, R. Astrup, and J. Breidenbach, "Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and national forest inventory data," *Forest Ecosystems*, vol. 8, 2021, Art. no. 65.
[54] M. Urbazaev et al., "Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne Lidar, and SAR and optical satellite data in Mexico," *Carbon Balance Manage.*, vol. 13, no. 1, Feb. 2018, Art. no. 5.

Nikola Besic received the master's degree in optics and RF engineering from Grenoble INP, Grenoble, France, in 2011 and the Ph.D. degree in remote sensing from Grenoble-Alpes University, Grenoble, France, in 2014.

He subsequently served as a Postdoctoral Researcher with the Environmental Remote Sensing Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, and collaborated with the Radar, Satellites, and Nowcasting Group, MeteoSwiss, Locarno, Switzerland. His academic journey also includes positions with the Centre for Radar Meteorology (Météo-France), Toulouse, France, and with AgroParisTech, Nancy, France, prior to his current role with the French National Mapping Agency (IGN), Saint-Mande, France. His scholarly pursuits encompass statistical signal and image processing, electromagnetism, and the application of machine learning and data mining techniques, with a particular emphasis on Earth sciences and the critical issue of climate change. His research interests include the remote sensing and modeling of the cryosphere, atmosphere, and biosphere.

Sylvie Durrieu received the M.Sc. degree in forest and plant biology from the University of Nancy, Nancy, France, in 1989, and an engineer degree in forestry in 1989 and the Ph.D. degree in remote sensing in 1994 from the French Institute of Forestry, Agricultural and Environmental Engineering (ENGREF-AgroParisTech), Palaiseau Cedex, France.

She obtained an accreditation to supervise research in Earth and Water Sciences from the University of Montpellier in 2018. From 1994 to 1999, she was with the French National Forest Inventory (IFN) and carried out R&D in the field of remote sensing applied to forest inventory and mapping, focusing on optical imagery technology. Since 1999, she has been with INRAE, Paris, France. She has developed methods for characterizing and monitoring forest resources and biodiversity from 3-D LiDAR data. She was PI of LEAF (LiDAR for Earth and Forests), a project of space LiDAR mission submitted to ESA in 2008, and supported by CNES from 2013 to 2020 as a potential midterm space mission. Her research interests include LiDAR technology for forest ecosystem monitoring.

Anouk Schleich received the M.Sc. degree in geomatics engineering from École Nationale des Sciences Géographiques (ENSG), Marne-la-Vallée, France, in 2020, and the Ph.D. degree in geomatics from INRAE Montpellier, Montpellier, France, in 2024.

Her research interests include LiDAR and forest inventories.

Cédric Vega received the master's degree in ecology from Paul Sabatier University, Toulouse, France, in 2000 and the Ph.D. degree in environmental sciences from UQAM, Montreal, Canada, in 2006.

He obtained the accreditation to supervise research (HDR) in ecology from Paul Sabatier University in 2017. From 2010 to 2014, he headed the Laboratory of Applied Informatics and Geomatics, Institut Français de Pondichéry in India. He is currently a Senior Researcher with ENSG-Geomatics, Institute of Geographic and Forest Information (IGN), Nancy, France. He is particularly focusing on developing methods for both quantifying forest structures and dynamics, and estimating forest biophysical parameters. His research interests include remote sensing of forest ecosystems, particularly using LiDAR, photogrammetry, and high resolution optical imagery.