Combining LiDAR Metrics and Sentinel-2 Imagery to Estimate Basal Area and Wood Volume in Complex Forest Environment via Neural Networks

Kamel Lahssini, Florian Teste, Karun Reuel Dayal, Sylvie Durrieu, Dino Ienco, Jean-Matthieu Monnet

Combining LiDAR Metrics and Sentinel-2 Imagery to Estimate Basal Area and Wood Volume in Complex Forest Environment via Neural Networks

Kamel Lahssini, Florian Teste, Karun Reuel Dayal, Sylvie Durrieu, Dino Ienco, *Member, IEEE*, and Jean-Matthieu Monnet

***Abstract*—Forest ecosystems play a fundamental role in natural balances and climate mechanisms through their contribution to global carbon storage. Their sustainable management and conservation is crucial in the current context of global warming and biodiversity conservation. To tackle such challenges, earth observation data have been identified as a valuable source of information. While earth observation data constitute an unprecedented opportunity to monitor forest ecosystems, its effective exploitation still poses serious challenges since multimodal information needs to be combined to describe complex natural phenomena. To deal with this particular issue in the context of structure and biophysical variables estimation for forest characterization, we propose a new deep learning-based fusion strategy to combine together high density three-dimensional (3-D) point clouds acquired by airborne laser scanning with high-resolution optical imagery. In order to manage and fully exploit the available multimodal information, we implement a two-branch late fusion deep learning architecture taking advantage of the specificity of each modality. On the one hand, a 2-D CNN branch is devoted to the analysis of Sentinel-2 time series data, and on the other hand, a multilayer perceptron branch is dedicated to the processing of LiDAR-derived information. The performance of our framework is evaluated on two forest variables of interest: total volume and basal area at stand level. The obtained results underline that the availability of multimodal remote sensing data is not a direct synonym of performance improvements but, the way in which they are combined together is of paramount importance.**

***Index Terms*—Convolutional neural networks (CNNs), forest monitoring, multiscale remote sensing, multisensor data fusion, structure and biophysical variables estimation.**

I. Introduction

FOREST ecosystems play a fundamental role in natural balances and climate mechanisms through their contribution to global carbon storage. More than 40% of the global terrestrial carbon stock is contained in these natural carbon sinks [1]. Moreover, they are home to a rich, but also very fragile, animal and plant biodiversity [2]. Their sustainable management and conservation is therefore a crucial challenge in the current context of global warming.

To address these issues and provide stakeholders with valuable inputs to support the decision-making process, the effective exploitation of available and up-to-date earth observation data is a major challenge. Nowadays, ongoing space missions and airborne acquisition campaigns allow to collect an increasing amount of remote-sensing data [3] that provide complementary information acquired via various modalities and at different spatial and temporal resolutions [4].

These remote-sensing data supply valuable information to monitor and characterize forest ecosystems [5]–[7].

For instance, light detection and ranging (LiDAR) technology is well suited for the characterization of structure and biophysical variables like the amount of wood resource and biomass [8], [9], due to the fact that the laser signal can penetrate through forest canopies. Airborne laser scanning (ALS) systems are LiDAR-based systems which provide 3-D multiecho point clouds describing the structure of the vegetation from the top of the canopy to the ground. The 3-D spatial distribution of the points resulting from the laser-environment interaction, thus, can be directly linked with the 3-D spatial distribution of the vegetation [10]. For regional-scale monitoring, the analysis of ALS data mainly relies on area-based approaches (ABA) involving statistical analysis of the spatial distribution of the points at stand level, i.e., around 300 to 700 m^2 [11]. The joint use of ground truth (GT) information at plot level and metrics derived from the ALS point clouds enables to develop empirical models to estimate structure and biophysical forest variables. Models are then applied to the whole area of study to produce forest resource maps [12], [13]. ABA models have exhibited robustness and effectiveness in predicting inventory variables for a wide range of homogeneous forests [14]. However, they exhibit limitations when implemented in more complex and heterogeneous environments requiring specific calibration to identify the most relevant variables that best describe a given study area through ALS data [15].

When analyzing ALS point clouds, research studies [16] indicate that topography, amongst other factors, greatly influences ALS-based models' performances. This is mainly due to the

Manuscript received December 26, 2021; revised April 5, 2022 and May 9, 2022; accepted May 11, 2022. Date of publication May 19, 2022; date of current version June 8, 2022. This work was supported in part by the French Agency for Ecological Transition (ADEME) through the PROTEST project under Grant 1703C0069, in part by GRAINE program, and in part by the French Region of Occitanie with the funding of part of K. Dayal's Ph.D. scholarship. *(Corresponding author: Dino Ienco.)*

Kamel Lahssini, Florian Teste, Karun Reuel Dayal, Sylvie Durrieu, and Dino Ienco are with the INRAE, UMR TETIS, University of Montpellier, 34000 Montpellier, France (e-mail: kamel.lahssini@inrae.fr; fteste96@protonmail.com; karun.dayal@inrae.fr; sylvie.durrieu@inrae.fr; dino.ienco@inrae.fr).

Jean-Matthieu Monnet is with the University Grenoble Alpes, INRAE, LESSEM, 38400 Grenoble, France (e-mail: jean-matthieu.monnet@inrae.fr).

Digital Object Identifier 10.1109/JSTARS.2022.3175609

fact that ALS point clouds are usually normalized in height to take into account ground elevation. The normalization process induces a distortion of the tree structure and directly affects the data distribution within each tree, influencing the metrics derived from the point clouds [17].

Sentinel-2 optical images provide radiometric information that can be useful for the characterization of stand composition [18], [19] and, thus, bring complementary information to ALS geometric data. The authors of [20] investigated the usability of Sentinel-2 2-D images in forest inventories against other remote-sensing source-providing 2.5- or 3-D-data-like ALS, elevation model induced by high-resolution optical satellite WorldView-2 images, and syntheticaperture radar stereo data from TerraSAR-X. The results confirmed that the higher spatial resolution input data, in the case of 3-D products, correlate with more accurate forest inventory parameter predictions, which is in line with other results presented in literature. However, the authors also highlight that, despite not having outstanding performances, Sentinel-2 imagery has the advantage, compared to other remote-sensing data sources, to being free of charge and regularly updated. The authors of [21] evaluated a particular machine learning algorithm Bayesian additive regression trees (*BART*) to analyze Sentinel-2 images and topographic variables to estimate the forest stand characteristics. The results indicate that the combination of radiometric and topographic variables (derived from PALSAR data) improved the estimation of the forest attributes compared to the use of only Sentinel-2 information.

Several studies demonstrated the value of combining LiDAR data with radiometric information from high spatial resolution multispectral satellite sensors regarding the characterization of both complex forest environments [22], [23] and urban vegetation [24]. However, enriching ALS-derived metrics with additional radiometric information from optical images and topography descriptors is challenging in commonly used ABA models based on linear regression. Indeed, such models do not allow to fully exploit the complex interplay between all the available input modalities since all the information is treated equally.

In recent years, deep learning (DL) approaches [25] are getting more and more attention in the field of remote sensing [26], since they are demonstrating compelling performances in the analysis of multimodal (i.e., multisource, multitemporal,and multiscale) earth-observation data [27]. Neural networks, through the stacking of nonlinear functions, allow representation and modeling of complex relationships between a set of input and output variables in an end-to-end manner. In a multisource remote-sensing scenario, two main strategies, namely 1) *early fusion* and 2) *late fusion,* are mainly adopted in order to combine heterogeneous and complementary information [27]. While the former strategy firstly combines data together and then feeds a standard neural network architecture, the latter one firstly analyzes each source by means of a dedicated branch (encoder) and then, terminates the processing of the data, combining the intermediate per-branch information via additional neural network layers. In both strategies, the set of network parameters are optimized end-to-end. The majority of the proposed approaches were introduced to cope with the issue of multimodal remote-sensing analysis for classification purposes [28]–[30]. Regarding the specific combination of spectral optical and LiDAR remote-sensing data, several works exist that deal with the general task of land cover mapping by employing both early- and late-fusion strategy to combine hyperspectral and LiDAR data [31]–[33], multispectral (three or four spectral bands) with LiDAR information [34], or hyperspectral, multispectral, and LiDAR data together [35].

However, there is a lack of methodological investigation to leverage the potential of DL-based strategy for regression tasks in the context of multimodal remote sensing analysis. Only recently, the computer science and signal processing communities have started to rigorously investigate the potential of such strategies for monosource regression tasks [36], while some attempts related to the characterization of forest properties are emerging in the remote-sensing community [37]. In [37], the authors proposed a stacked sparse autoencoder network from LiDAR metrics and optical indices (from Landsat 8 imagery) with the aim to estimate forest aboveground biomass. In this approach, all the information (LiDAR metrics, optical indices, and combined optical-LiDAR indices) are fed to a fully-connected neural network model, following an early-fusion strategy, in order to estimate the biomass quantity.

To cope with the estimation of structure and biophysical forest variables from multimodal remote-sensing data, we propose a DL-based model capable to fully exploit the interplay among multimodal information coming from Sentinel-2 images and derived from ALS point clouds. Conversely to previous attempts in the literature, the proposed framework, named multimodal forest variables estimation based on DL framework (*MMFVE*), implements łate-fusion strategy deploying a two-branch DL architecture to deal with multimodal remote-sensing data: Information derived from ALS point clouds on one hand and time series of Sentinel-2 optical imagery on the other hand. We focus on the estimation of two forest variables, 1) *total volume* and 2) *basal area*, in the study area of *Massif des Bauges Natural Regional Park* (France). The *total volume* of a tree is defined as the volume of all the wood contained in it (stem and branches). When coupled with additional information, it is useful for the computation of biomass. The *basal area* of a tree is the cross-sectional area of the tree at breast height (usually defined as 1.3 m above ground). When computed for all the trees in a plot (m^2/ha), it is the measure of the cross-sectional area occupied by the trees and is a stand density indicator commonly used by forest managers in silvicultural planning. The main objective of our study is to assess the benefit of combining ALS and optical satellite image time series (i.e., Sentinel-2) data for the downstream task of forest structure and biophysical variables estimation, in order to understand if multimodal information can improve the estimation of these physical quantities in a complex forest environment. While the former source of information, ALS, is widely used for the characterization of forest properties, the latter, Sentinel-2, brings knowledge about stand composition. Additionally, we also integrate in our analysis topographic information, which impacts (as confirmed by recent studies [21] and empirically observed in the present experimental evaluation) the quality of both ALS structure information and Sentinel-2 radiometric signal.

To summarize, the main contributions of our research study are as follows:

1) the design of an end-to-end DL framework devoted to forest variables estimation via a multisource late-fusion approach;
2) a first study, to the best of our literature survey, that assesses the combination of topographic information, LiDAR metrics, and Sentinel-2 multispectral and multitemporal data for *total volume* and *basal area* estimation in a complex forest environment.

The rest of this article is organized as follows. Section II introduces the data available on the study site. Section III describes the proposed framework. The experimental settings and the results are reported and discussed in Sections IV and V, respectively. Finally, Section VI concludes this article.

II. Data

A. Study Site

The study site is the Regional Natural Park of the *"Massif des Bauges,"* located in France in the Alps mountains, between the two administrative departments of Savoie and Haute-Savoie (Fig. 1 and Fig. 2). The site is characterized by a steep and irregular topography, and an elevation ranging from 256 to 2217 m. The forest covers an area of 51 136 ha, which accounts for 60% of the total area of this inhabited park. The forest is, thus, a major component of the park's landscape and provides several ecosystem services to local inhabitants and neighboring cities: Wood production, outdoor activities, and biodiversity conservation. The forest stands consist of both deciduous and coniferous species, with a dominance of silver fir (*Abies alba*), Norway spruce (*Picea abies*), and common beech (*Fagus sylvatica*). Managed forests are mostly in an uneven-aged system.

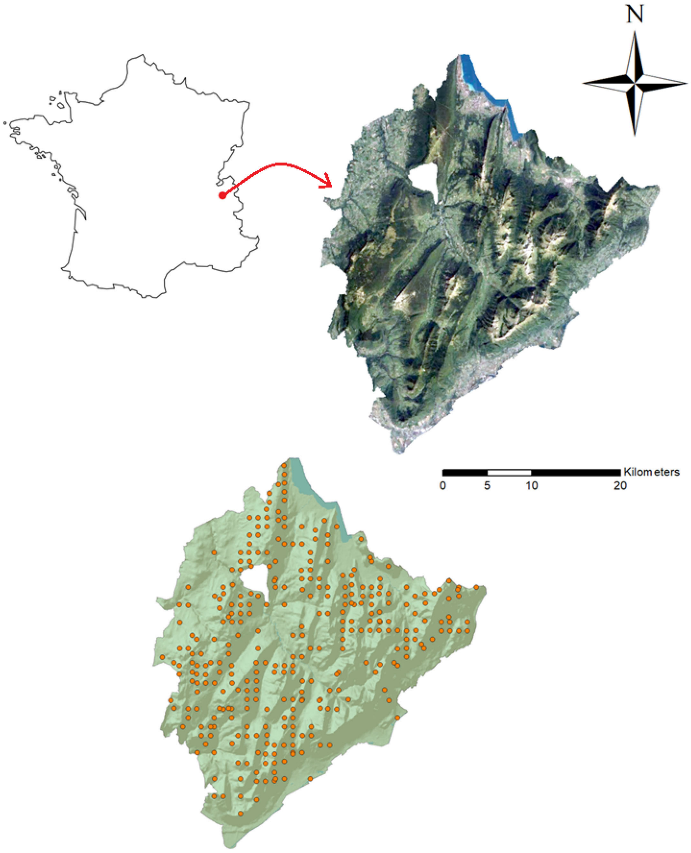


Fig. 1. Location of the study area in France and RGB composite from SENTINEL-2 imagery (top); Extent of the Massif des Bauges Natural Regional Park and location of inventory plots (Bottom).

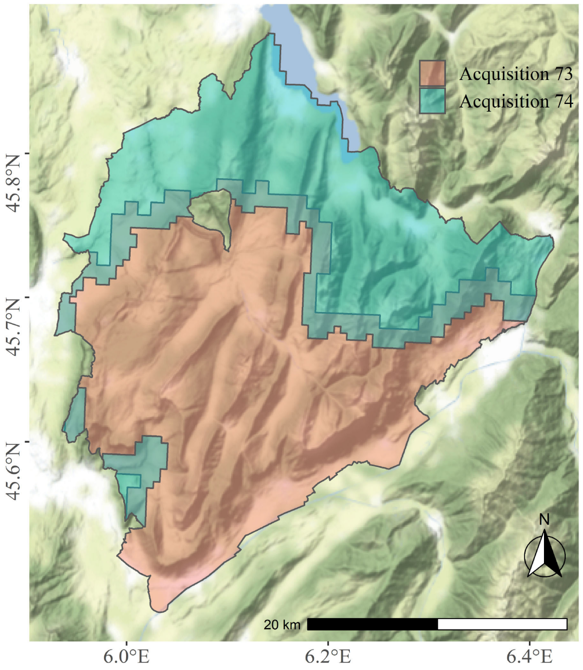
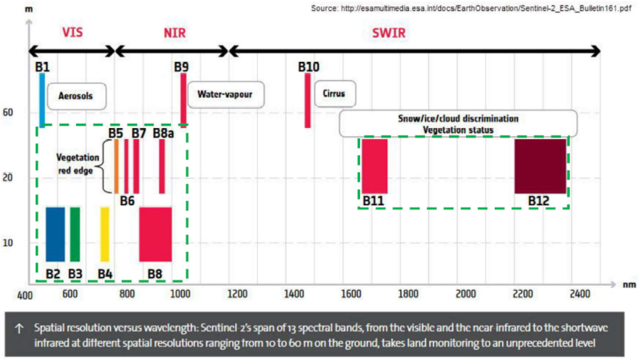


Fig. 2. Location of the study area in France and RGB composite from SENTINEL-2 imagery (top); Extent of the Massif des Bauges Natural Regional Park and location of inventory plots (Bottom).

B. Remote-Sensing Data

Airborne LiDAR data used in this study were not acquired under similar conditions over the whole area. Two surveys were conducted, each one covering a specific area. 1) The first acquisition (Acquisition 73) was conducted between June and September 2016, by the French National Institute of Geographic and Forest Information (IGN), and results from a collaboration between IGN and the governance and management board of Savoie Mont Blanc in order to provide the first LiDAR coverage over a whole French department in a mountainous area. On our study site, it covers an area of 53 600 ha in the southern part of the park, located in the Savoie department. Mean LiDAR point density is 4 points/m^2. Raw data preprocessing, including geolocation, point clouds alignment, and data tiling was performed by a private contractor. 2) The second acquisition (Acquisition 74) was conducted in September 2018, by a private data provider in order to complete the coverage of the park area. It covers an area of 37 350 ha in the northern part of the park, located in the Haute-Savoie department. LiDAR point density is on average



↑ Spatial resolution versus wavelength: Sentinel-2's span of 13 spectral bands, from the visible and the near infrared to the shortwave infrared at different spatial resolutions ranging from 10 to 60 m on the ground, takes land monitoring to an unprecedented level

Fig. 3. S2 spectral bands used in this study are surrounded by a dashed green line.

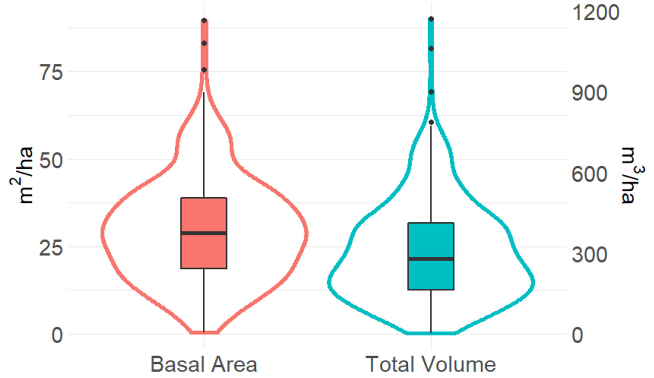


Fig. 4. Distribution of GT data for the two forest variables *basal area* and *total volume*.

TABLE I
ALS METRICS

	Metric name	Metric definition
1	gap_F	Gap fraction
2	rumple_index	Rumple index
3	zmax	Max height
4	zmean	Mean height
5	zsd	Standard deviation of heights
6-24	zq*	Height percentiles (* = 5, 10, ..., 95)
25	zskew_first	Skewness of heights (first echos)
26	zskew_last	Skewness of heights (last echos)
27	zkurt_first	Kurtosis of heights (first echos)
28	zkurt_last	Kurtosis of heights (last echos)
29	zentropy	Entropy of heights
30	pzabovezmean	Percentage of echos above zmean
31	itot	Total intensity
32	imean	Mean intensity
33	imax	Max intensity
34	isd	Standard deviation of intensities
35	iskew	Skewness of intensities
36	ikurt	Kurtosis of intensities
37-45	zpcum*	Cumulative percentage of echoes in layer number * (* = 1, 2, ..., 9)
46-50	ipcumzq*	Percentage of total intensity below zq* height percentile (* = 10, 30, ..., 90)
51-55	p*	Percentage of $*^{th}$ echoes (* = 1, 2, ..., 5)

of 13 points/m^2, with a strong spatial heterogeneity and local densities ranging from 4 to 25 points/m^2 over the whole area of acquisition. A 1-m spatial resolution digital terrain model (DTM) over the whole study site was derived from the LiDAR data. The LiDAR data of both acquisitions were normalized in height, using the DTM, to account for differences in terrain elevation.

Satellite data consists of a Sentinel-2 (S2) time series of five images, acquired between the years 2017 and 2018 at the following dates: October 7, 2017, November 1, 2017, August 28, 2018, September 27, 2018, and October 17, 2018. The images were chosen to ensure a cloud-free and snow-free coverage of the study site. Moreover, they were also selected to match, as much as possible, the dates of field and LiDAR surveys. The S2 images were obtained from the THEIA data platform[1] at level-2A (top of canopy reflectance). Ten spectral bands were considered in this study (Fig. 3) and all the bands were resampled (with the nearest neighbor strategy) to a 10-m spatial resolution. In addition, three topographic variables were computed at a spatial resolution of 10 m. The 1-m DTM was down-scaled to a 10-m raster of elevations, which was further used to compute both the aspect and slope layers, leading to a total of 53 channels describing the study site at a spatial resolution of 10 m.

C. Field Data

The GT data were obtained through a forest inventory campaign conducted by the French National Forest Office and the French National Research Institute for Agriculture, Food, and Environment between June and September 2018. A total of 291 circular plots, regularly distributed over the study site, were inventoried and georeferenced. Measurements were performed within concentric plots of 10-m and 15-m radius. Within the 15-m radius, all the trees with a diameter at breast height (DBH) above 17.5 cm had their DBH measured using a tape and the species were recorded. Within the 10-m radius, the trees between 7.5 and 17.5 cm DBH were classified and counted according to two criteria: 1) the diameter category (7.5 to 12.5 cm or 12.5 to 17.5 cm) and 2) the species category (coniferous or deciduous). Because no heights and not all the diameters were measured, the database from the French National Forest Inventory produced by IGN was used to enrich the forest inventory with additional information to allow the computation of the two forest variables of interest in this study: *Total volume* and *basal area*. The variables were computed at the scale of a 15-m radius plot and successively converted to per hectare values. Fig. 4 shows the data distribution for each of the forest variables over the 291 plots.

D. Remote-Sensing-Based Variables

Inventory plots are formed by 30-m diameter disks. For each plot, the normalized LiDAR point cloud located within the extent of the 15-m radius circle was extracted. A threshold was applied to remove all the points below 5 m corresponding to lower vegetation. As shown in Table I, 55 standard LiDAR metrics were then computed at plot level using the R *lidR* package [38]. Three topographic values were also averaged at plot level using the 1-m DTM, i.e., 1) elevation; 2) slope; and 3) aspect, leading to a total of 58 descriptors characterizing each plot. Regarding

[1] [Online]. Available: https://www.theia-land.fr/en/homepage-en/

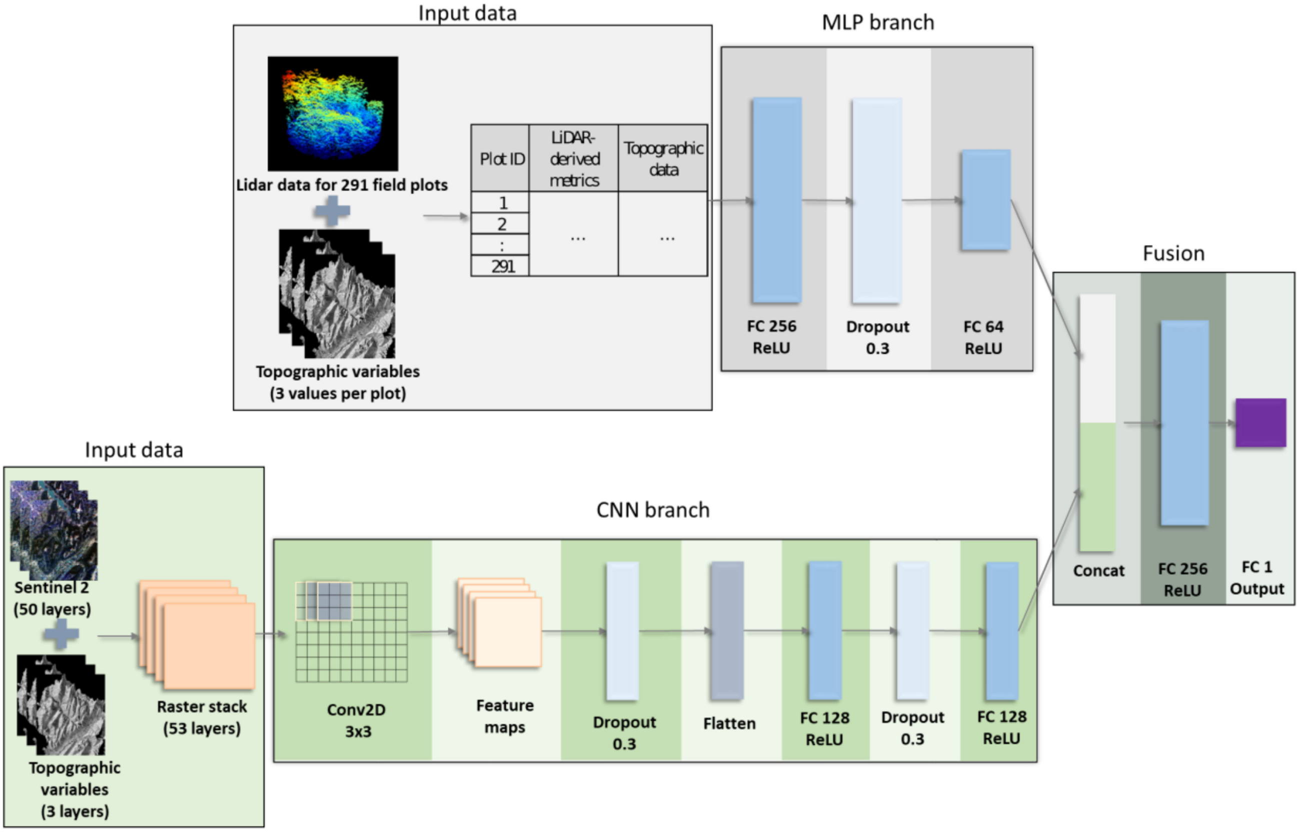


Fig. 5. Overview of *MMFVE* framework. The architecture has two branches, one devoted to the analysis of Sentinel-2 time-series data and one dedicated to the analysis of metrics extracted from LiDAR measurements. Then, the per-source feature representations are aggregated, under late-fusion strategy, by the means of the concatenation operation in order to perform the final forest-variable estimation.

optical imagery, because of the spatial resolution of the resampled S2 images, the spectral information associated to each plot is extracted on a 3 × 3 pixel window that corresponds to a 30-m side square (referred to as a *patch*). The central pixel of a given patch is selected in order to contain the centre of the associated plot. The final 3 × 3 patch is then built around this central pixel. In addition to the 58 descriptors previously mentioned, each plot is also described by a 3 × 3 pixel patch of 53 channels at a 10-m spatial resolution extracted from the S2 time series plus features extracted from the 10-m DTM (see Subsection II-B).

To summarize, the labelled data used in this study consist of 291 plots described by two forest variables (*total volume* and *basal area*), 55 LiDAR-derived metrics, 3 topographic values at plot level, and 50 Sentinel-2 spectral bands plus 3 topographic bands at a 10-m spatial resolution. The details of all LiDAR-derived metrics and Sentinel-2 information are listed in Table I.

III. Method

In this section we describe our framework, named *MMFVE*, to cope with the fusion of multimodal data for the estimation of structure and biophysical variables. Firstly, we supply a general overview of our two-branch architecture and we detail the end-to-end learning strategy we have adopted to learn the model parameters. Secondly, we introduce the per-source encoders we have designed to take into account the specificity of both LiDAR-derived metrics and Sentinel-2 time-series data.

A. Multimodal DL Architecture

Fig. 5 depicts the proposed framework, *MMFVE*. In our scenario, each geospatial location is described by means of two different kinds of information: 1) metrics extracted from LiDAR acquisition and 2) spatial patches extracted from Sentinel-2 time series.

To manage and exploit the available multimodal information, *MMFVE* has two input branches, one for each of the sources to process. Each branch is associated to an encoder network that extracts a source specific representation: R_{S2} and R_{LIDAR}. Subsequently, the different per-source representations are aggregated together considering a late-fusion schema [27] by concatenation of the per-source representations, with the aim to obtain a multimodal representation (R_{MM}), gathering together the multifaceted information describing the specific geospatial location, i.e., a 0.07 ha circular plot. Finally, the multimodal representation is fed through a hidden layer and a final output layer with only one neuron. The output neuron has no activation function associated to it and it only performs a linear combination of the input signals coming from the previous connected layer as commonly done in regression-based DL models [36]. The whole process is performed end-to-end.

The detailed description of the different components on which *MMFVE* is built on is reported in Fig. 5.

Regarding the general training process, since we are dealing with a regression task, we employ the Huber loss (HL) [36] as cost function. The HL is defined as follows:

$$HLoss(Y, Y') = \begin{cases} \frac{1}{2}(Y - Y')^2 & if|Y - Y'| < \Delta \\ \Delta|Y - Y'| - \frac{1}{2}\Delta^2 & otherwise \end{cases} \tag{1}$$

where Y and Y' are the true and the estimated values, respectively. Basically, the HL behaves as the mean squared error (MSE) loss when the error is lower than a threshold Δ, while it mimics the mean absolute error (MAE) loss otherwise. In addition, it is also differentiable at 0. The HL combines good properties from both MAE and MSE and avoids their limitations. Conversely to MAE, it bypasses large gradient backpropagation when the estimated quantity is getting closer to the real value and, differently from the MSE loss, it is more robust to outliers. In our scenario we set Δ to 1 for the HL.

B. Sentinel-2 and LiDAR Encoders

Regarding the S2 time-series data, we exploit a 2-D convolutional neural network (2-D-CNN) with the goal to take into account the spatial context describing the geospatial location associated to the forest variable to estimate. To this end, the final raster cube with 53 bands (50 Sentinel-2 bands and 3 bands containing topographic characteristics, as explained in Section II-D) is used as input. The encoder architecture has one 2-D convolutional layer with a kernel of size 3×3 and 256 filters, followed by two fully connected dense layers with 128 neurons each. The 2-D convolutional layer and the first fully-connected layers are followed by a dropout layer with a drop rate equal to 0.3.

Concerning the LiDAR information, describing the geospatial location of interest, we analyze the unstructured set of 55 metrics together with the additional three topographic variables (as described in Section II-D) by means of two fully-connected dense layers with 256 and 64 neurons. The first fully-connected layer is followed by a dropout layer with a drop rate equal to 0.3.

Once the encoder extracts per-source features (R_{S2} and R_{LiDAR}), they are concatenated (R_{MM}) and successively processed via a fully-connected dense layer with 256 neurons and ReLU activation function that, finally, feeds the output layer constituted by only one neuron.

IV. Experimental Evaluation

In this section, we present the experimental settings and the results obtained on the study area introduced in Section II.

Regarding both forest variables (*total volume* and *basal area*), we firstly validate the importance of considering topographic information in addition to LiDAR-derived metrics for the estimation tasks. Then, we assess the use of only S2 and S2 plus topographic information. Successively, we evaluate the performance of our multimodal framework, *MMFVE*, to leverage the interplay between the different modalities we have as inputs (LiDAR, topographic, and S2 data) with respect to both machine and DL competitors. Finally, we report some information related to the training time associated to the different multimodal approaches.

As additional evaluations, we also assess the sensitivity of our framework to the loss function employed to optimize the internal network parameters (by leveraging standard loss functions for regression task [36]) and we supply a summary of the training time required by each of the employed competing methods, considering the full multisource setting (LiDAR metrics, topographic information, and S2 imagery).

A. Experimental Settings

Considering all the available input modalities in our multimodal setting, our first goal is to perform several preliminary experiments in order to better understand the contribution of each data source. We then evaluate the performances of our two-branch DL framework, *MMFVE*, with respect to its competitors.

Regarding LiDAR-derived metrics, as reported in Section II, each plot is described by 55 metrics plus three additional topographic variables (aspect, slope, and elevation). Concerning the S2 time series, enriched with the same three topographic variables, we consider input patches of size 3×3 ($30\ \mathrm{m} \times 30\ \mathrm{m}$) that approximately correspond to the extent of a plot.

As competing approaches, we use standard and recent machine and DL techniques commonly employed to estimate forest variables at stand level from remote-sensing data [21], [37], [39], [40]. More in detail, we adopt a K nearest neighbors regressor (KNN) based on Euclidean distance, a random forest (RF) based regressor method that leverages ensemble of trees via a bagging strategy to estimate the target variable, and the $BART$ model recently evaluated in [21]. In addition, we include as competing strateg,y the sparse stacked autoencoder (*SSAE*) model introduced in [37]. This model is based on an early-fusion strategy to aggregate all the different information sources before the analysis step. It also provides an implicit comparison between early- and late-fusion strategies. According to [37], the parameters of the *SSAE* model are learnt with MSE as loss function.

When only LiDAR-based (and topographic) information are considered, we include a multilayer perceptron (MLP) with the same architecture as the encoder branch of *MMFVE* devoted to manage the LiDAR-derived metrics. Similarly, when only S2 imagery (and topographic) information are considered, we include a 2-D-CNN with the same architecture as the encoder branch of *MMFVE* devoted to manage the Sentinel-2 information.

The values of the different sources were normalized in the interval [0, 1]. The dataset was split into training, validation, and test set with 191, 50, and 50 plots, respectively. The evaluated models were optimized via training/validation procedure [41]. For the KNN approach, the training/validation procedure allows to choose the best value of K in the set {1, 3, 5, 7, 9}, while for the RF regressor the optimization procedure permits to choose the best number of estimators (trees) in the set {200, 300, 400, 500, 600, 700, 800}, as well as the number of attributes for each subset of samples in the bagging procedure in the range

TABLE II
Performances of the Different Competing Approaches (KNN, RF, $BART$, and MLP) in Terms of MAE, RMSE, and R^2 Evaluation Metrics When LiDAR-Based Information is Considered Alone or Coupled With Topographic Information to Estimate Total Volume

Method	**MAE**	**RMSE**	$\boldsymbol{R^2}$
KNN(L)	0.082 ± 0.011	0.109 ± 0.017	0.550 ± 0.109
RF(L)	0.079 ± 0.008	0.107 ± 0.014	0.572 ± 0.080
BART(L)	0.078 ± 0.008	0.105 ± 0.012	0.564 ± 0.105
MLP(L)	**0.069** ± 0.009	**0.092** ± 0.015	**0.668** ± 0.078
KNN(L, T)	0.082 ± 0.010	0.109 ± 0.015	0.553 ± 0.093
RF(L, T)	0.076 ± 0.008	0.103 ± 0.013	0.604 ± 0.068
BART(L, T)	0.074 ± 0.007	0.098 ± 0.011	0.621 ± 0.091
MLP(L, T)	**0.062** ± 0.008	**0.082** ± 0.015	**0.737** ± 0.063

TABLE III
Performances of the Different Competing Approaches (KNN, RF, $BART$, and MLP) in Terms of MAE, RMSE, and R^2 Evaluation Metrics When LiDAR-Based Information is Considered Alone or Coupled with Topographic Information to Estimate Basal Area

Method	**MAE**	**RMSE**	$\boldsymbol{R^2}$
KNN(L)	0.094 ± 0.012	0.121 ± 0.016	0.492 ± 0.110
RF(L)	0.088 ± 0.008	0.117 ± 0.013	0.529 ± 0.069
BART(L)	0.086 ± 0.009	0.113 ± 0.011	0.533 ± 0.095
MLP(L)	**0.081** ± 0.009	**0.104** ± 0.013	**0.606** ± 0.072
KNN(L, T)	0.094 ± 0.011	0.122 ± 0.014	0.486 ± 0.093
RF(L, T)	0.085 ± 0.008	0.113 ± 0.012	0.559 ± 0.065
BART(L, T)	0.082 ± 0.009	0.108 ± 0.010	0.572 ± 0.093
MLP(L, T)	**0.071** ± 0.008	**0.092** ± 0.012	**0.693** ± 0.062

[3, 10]. For the BART method, we adopt the parameter setting employed in [21].

Concerning *MMFVE* and the SSAE competitor, the learning stage was conducted over 400 epochs and the Adam optimizer [42] was used to learn trainable parameters with a learning rate of 10^{-5}. Dropout rate was set to 0.3 and batch size was fixed to 1.

The assessment of the models' performances was done considering the test set and the following metrics: *MAE*, root mean squared error (*rmse*), and R^2 (*R*-squared, the coefficient of determination). The performance metrics are defined as follows:

$$\text{MAE}(Y, Y') = \frac{1}{|Y|} \sum_{i=1}^{|Y|} |Y_i - Y'_i| \tag{2}$$

$$rmse(Y, Y') = \sqrt{\frac{1}{|Y|} \sum_{i=1}^{|Y|} (Y_i - Y'_i)^2} \tag{3}$$

$$R^2(Y, Y') = 1 - \frac{\sum_{i=1}^{|Y|} (Y_i - Y'_i)^2}{\sum_{i=1}^{|Y|} (Y_i - \bar{Y})^2} \tag{4}$$

where Y contains the original values of the forest variable to estimate and Y' is the set of values estimated by a particular approach.

Since model performances may vary depending on data splitting due to simpler or more complex samples involved in the different partitions, all metrics were averaged over 30 random splits of the dataset following the strategy mentioned above. The different neural network architectures were implemented using the Python Tensorflow library.

B. Results

Here, we provide the experimental evaluation according to different combinations of the input sources as well as the complete multimodal setting with LiDAR metrics, topographic information, and S2 imagery.

1) Assessing LiDAR Metrics With and Without Topographic Information: Tables II and III report the performances of the different competing approaches when LiDAR-derived metrics (L) and topographic information (T) are considered to estimate *total volume* and *basal area*, respectively. Source information is specified as input of the method. For instance, $RF(L)$ indicates an *RF* taking as input only LiDAR-derived metrics, while $RF(L, T)$ indicates the same method with LiDAR-derived metrics and corresponding topographic variables as inputs.

Concerning the estimation of *total volume* (Table II), we can observe that the use of topographic information (in addition to LiDAR-derived metrics) generally ameliorates the results of the estimation algorithms. The only method that does not seem capable to exploit such additional information is the KNN algorithm. This can be explained by the fact that this approach relies on the computation of Euclidean distance and, to compute such a measure, all the input variables have the same importance. Since the number of topographic variables is much smaller than the number of LiDAR-derived features (3 versus 55), the contribution of the former is largely dominated by the contribution of the latter.

Conversely, RF, MLP, and $BART$ are capable to leverage the useful information carried out by the topographic variables providing better performances when topographic information is considered (lower values of *MAE* and *rmse* and higher values of R^2). While RF and $BART$ slightly improve their behavior, MLP clearly supplies the best performances gain regarding all the employed evaluation metrics. This latter method also provides the best absolute results outperforming all the other competing approaches when LiDAR-derived metrics alone and both LiDAR-derived and topographic variables are involved.

Regarding the estimation of *basal area* (Table III), we can observe a similar trend as the one depicted for the estimation of *total volume*. Also in this case, the KNN algorithm is unable of taking advantage of the extra information supplied by topographic variables, while the MLP approach exhibits the best performances with and without topographic variables (low *MAE* and *rmse* and high R^2), still demonstrating its ability to fully leverage the interplay between LiDAR-derived metrics and topographic information.

2) Assessing S2 Imagery With and Without Topographic Information: In order to understand the contribution of S2 time series in the estimation task, we have also assessed the performances of both forest variables (*total volume* and *basal area*) estimation considering only Sentinel-2 data with and without topographic information. The results are reported in Tables IV and V for the *total volume* and *basal area* forest variables, respectively. The achieved results are far inferior to

TABLE IV
Performances of the Competing Approaches in Terms of MAE, RMSE, and R^2 Evaluation Metrics When Sentinel-2 Data are Considered Alone or Coupled With Topographic Information to Estimate Total Volume

Method	MAE	RMSE	R^2
KNN(S2)	0.124 ± 0.011	0.156 ± 0.013	0.034 ± 0.189
RF(S2)	0.108 ± 0.010	0.134 ± 0.012	0.285 ± 0.149
BART(S2)	0.107 ± 0.009	0.133 ± 0.012	0.299 ± 0.124
2D-CNN(S2)	**0.091** ± 0.013	**0.115** ± 0.015	**0.478** ± 0.125
KNN(S2, T)	0.126 ± 0.011	0.158 ± 0.012	0.010 ± 0.183
RF(S2, T)	0.108 ± 0.009	0.134 ± 0.012	0.284 ± 0.146
BART(S2, T)	0.107 ± 0.009	0.133 ± 0.011	0.306 ± 0.112
2D-CNN(S2, T)	**0.086** ± 0.011	**0.108** ± 0.014	**0.530** ± 0.094

TABLE V
Performances of the Competing Approaches in Terms of MAE, RMSE, and R^2 Evaluation Metrics When Sentinel-2 Data are Considered Alone or Coupled With Topographic Information to Estimate Basal Area

Method	MAE	RMSE	R^2
KNN(S2)	0.135 ± 0.011	0.167 ± 0.013	0.029 ± 0.206
RF(S2)	0.115 ± 0.009	0.141 ± 0.011	0.269 ± 0.137
BART(S2)	0.114 ± 0.011	0.139 ± 0.012	0.293 ± 0.145
2D-CNN(S2)	**0.099** ± 0.012	**0.121** ± 0.013	**0.462** ± 0.131
KNN(S2, T)	0.136 ± 0.011	0.169 ± 0.012	0.050 ± 0.206
RF(S2, T)	0.115 ± 0.009	0.141 ± 0.011	0.272 ± 0.136
BART(S2, T)	0.114 ± 0.011	0.139 ± 0.012	0.289 ± 0.132
2D-CNN(S2, T)	**0.098** ± 0.010	**0.120** ± 0.012	**0.471** ± 0.097

TABLE VI
Performances of the Different Competing Approaches in Terms of MAE, RMSE, and R^2 Evaluation Metrics When LiDAR-Based Information is Jointly Exploited With Sentinel-2 Data and Topographic Information to Estimate Total Volume

Method	MAE	RMSE	R^2
MLP(L, T)	0.062 ± 0.008	0.082 ± 0.015	0.737 ± 0.063
KNN(L, T, S2)	0.124 ± 0.014	0.157 ± 0.019	0.100 ± 0.099
RF(L, T, S2)	0.078 ± 0.008	0.102 ± 0.012	0.643 ± 0.076
BART(L, T, S2)	0.072 ± 0.008	0.093 ± 0.011	0.657 ± 0.076
SSAE(L, T, S2)	0.055 ± 0.009	0.070 ± 0.011	0.803 ± 0.064
MMFVE	**0.046** ± 0.010	**0.061** ± 0.013	**0.850** ± 0.072

The $MLP(L,T)$ method is also reported to assess the added value to also consider Sentinel-2 data.

TABLE VII
Performances of the Different Competing Approaches in Terms of MAE, RMSE, and R^2 Evaluation Metrics When LiDAR-Based Information is Jointly Exploited With Sentinel-2 Data and Topographic Information to Estimate Basal area

Method	MAE	RMSE	R^2
MLP(L, T)	0.071 ± 0.008	0.092 ± 0.012	0.693 ± 0.062
KNN(L, T, S2)	0.131 ± 0.014	0.164 ± 0.018	0.088 ± 0.097
RF(L, T, S2)	0.086 ± 0.008	0.110 ± 0.011	0.626 ± 0.080
BART(L, T, S2)	0.079 ± 0.007	0.100 ± 0.008	0.632 ± 0.089
SSAE(L, T, S2)	0.053 ± 0.008	0.067 ± 0.010	0.831 ± 0.062
MMFVE	**0.050** ± 0.009	**0.066** ± 0.011	**0.836** ± 0.065

The $MLP(L,T)$ method is also reported to assess the added value to also consider Sentinel-2 data.

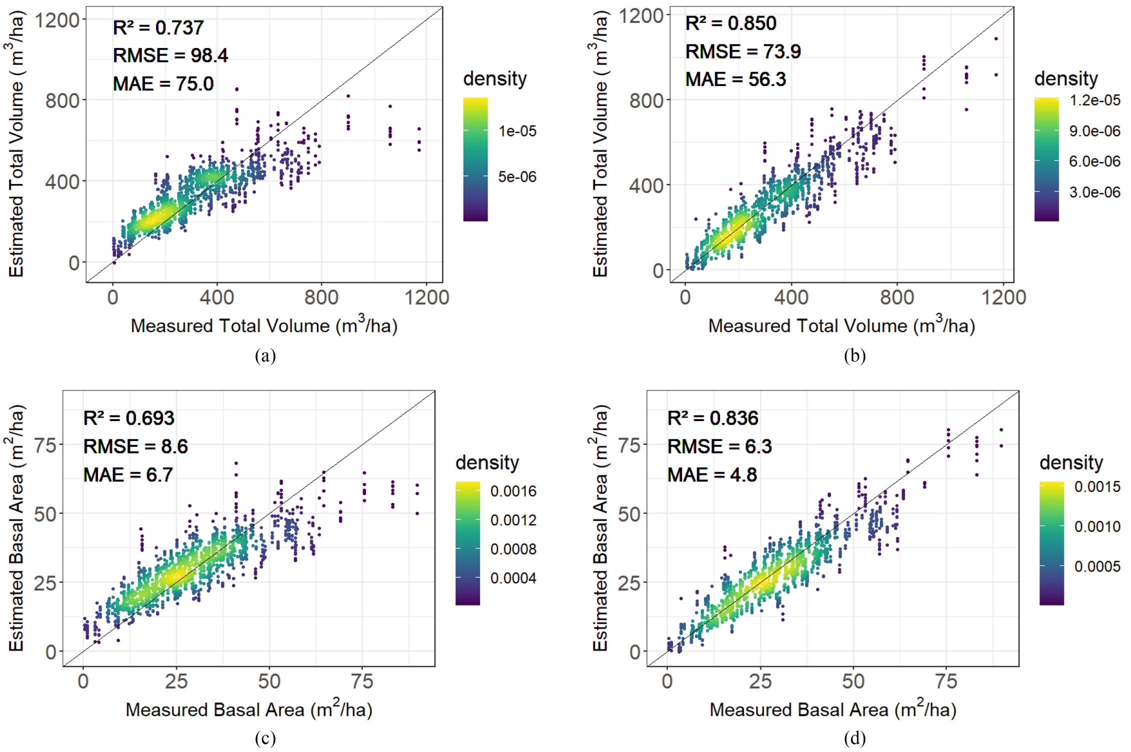
the results previously reported, with the best results achieved when Sentinel-2 data are jointly exploited with topographic information via the 2-D-CNN model.

3) Multimodal Setting: Tables VI and VII summarize the performances of the different competing methods when all the multimodal information, describing the study area, are employed for the estimation of *total volume* and *basal area*, respectively. In this case, given a method (i.e., RF), we indicate with {L,T,S2} the joint use of all the available information [i.e., $RF(L,T,S2)$]. In this comparison, as baseline to evaluate the added value of the Sentinel-2 time-series data, we consider the best performing method when only LiDAR-derived metrics and topographic information are used [$MLP(L,T)$].

For the case of *total volume* estimation (Table VI), we observe that *MMFVE* achieves the best results for all the evaluation metrics and the $SSAE$ approach achieves the second best estimation. A direct comparison between our multimodal framework and $MLP(L,T)$ (that can be considered as an ablation of *MMFVE*) underlines that the proposed neural network architecture is well suited to leverage the complementarity of the different input modalities. About the RF and the $BART$ approaches, we can note that also in this case the use of Sentinel-2 data provides some improvement with respect to the results reported in Table II. But, we can also note that, in spite of feeding all the available data to the *RF* classifier, its performances are still lower than the performances achieved by the $MLP(L,T)$. Unlike the other methods, the use of all the available modalities negatively influenced the behavior of the KNN regression, thus, resulting in a systematic degradation of the evaluation metrics.

Concerning the estimation of the *basal area* variable, when all the available modalities are used as input, we observe similar behaviors as the ones exhibited for the estimation of *total volume* (Table VII). *MMFVE* outperforms all the other approaches demonstrating its ability to combine together multiple input modalities with $SSAE$, based on an early-fusion strategy, that achieves performances that are comparable to our framework. Both, the RF and the $BART$ approaches, achieve marginal improvements with respect to its counterpart when Sentinel-2 input modality is missing. Table III highlights again their limited ability to fully exploit the interplay between LiDAR-derived metric, topographic information, and Sentinel-2 data. Also in this case, the KNN regressor is negatively affected by the large and heterogeneous amount of input information, emphasizing the inadequacy of this approach in a multimodal context.

Fig. 6 depicts the scatter plots of the estimated versus measured values for the two forest variables for two DL-based frameworks we have evaluated (*MMFVE* and *MLP*). More in detail, Fig. 6(a) and (b) represents the distributions related to *total volume* with only LiDAR and topographic information and with all the available sources (LiDAR, topographic, and Sentinel-2 data), respectively. We can observe that the latter configuration has higher estimation performances, and approximates the diagonal line, drawing the ideal scenario better than the former approach. This is also confirmed by the reported values for all the three evaluation metrics. Fig. 6(c) and (d) depicts the estimation of *basal area* when only LiDAR and topographic information are considered and when all the available sources (LiDAR, topographic, and Sentinel-2 data) are considered, respectively. Also in this case, we can see that the latter configuration, *MMFVE*, allows to achieve higher estimation performances. Similarly to what happens for the other forest variable, the use of all the available modalities permits to reach the best performances and



(a)

(b)

(c)

(d)

Fig. 6. Estimated verus measured values for the *total volume* and *basal area* biophysical forest variables when only LiDAR and topographic information or all the available sources (LiDAR, topographic, and Sentinel-2 data) are considered. (a) Estimated versus measured total volume using LiDAR and topographic data by $MLP(L, T)$. (b) Estimated versus measured total volume using LiDAR, topographic, and S2 data by *MMFVE*. (c) Estimated versus measured basal area using LiDAR and topographic data by $MLP(L, T)$. (d) Estimated versus measured basal area using LiDAR, topographic, and S2 data by *MMFVE*.

TABLE VIII
PERFORMANCES OF THE PROPOSED *MMFVE* FRAMEWORK IN TERMS OF MAE, RMSE, AND R^2 EVALUATION METRICS WHEN DIFFERENT LOSS FUNCTIONS ARE USED IN THE TRAINING PROCESS TO ESTIMATE TOTAL VOLUME

Loss	MAE	RMSE	R^2
MSE	0.048 ± 0.007	0.064 ± 0.009	0.836 ± 0.049
MAE	0.052 ± 0.009	0.072 ± 0.012	0.794 ± 0.069
HL	0.046 ± 0.010	0.061 ± 0.013	0.850 ± 0.072

TABLE IX
PERFORMANCES OF THE PROPOSED *MMFVE* FRAMEWORK IN TERMS OF MAE, RMSE, AND R^2 EVALUATION METRICS WHEN DIFFERENT LOSS FUNCTIONS ARE USED IN THE TRAINING PROCESS TO ESTIMATE BASAL AREA

Loss	MAE	RMSE	R^2
MSE	0.054 ± 0.007	0.071 ± 0.009	0.816 ± 0.048
MAE	0.052 ± 0.010	0.071 ± 0.014	0.810 ± 0.084
HL	0.050 ± 0.009	0.066 ± 0.011	0.836 ± 0.065

TABLE X
RUNNING TIMES FOR THE TRAINING PHASE OF *MMFVE* FRAMEWORK AND ITS COMPETITORS WHEN ALL THE AVAILABLE MULTIMODAL INFORMATION IS USED TO ESTIMATE ONE OF THE CONSIDERED FOREST VARIABLE

Method	Running time (seconds)
KNN	< 10
RF	15
BART	540
SSAE	440
MMFVE	360

to make a step closer to the ideal scenario represented by the diagonal line portrayed in the scatter plot.

4) Evaluating Different Loss Functions for MMFVE*:* In addition, we compare the performances of our framework, *MMFVE*, coupled with standard loss functions commonly employed for regression tasks: MSE and MAE. We indicate with *MMFVE*(MSE), *MMFVE*(MAE), and *MMFVE*(HL) the *MMFVE* framework coupled with MSE, MAE, and HL functions, respectively. Tables VIII and IX depict the obtained results for *total volume* and *basal area* variables, respectively. The obtained results underline that the HL function allows some kind of improvement compared to both MSE and MAE regression loss functions. This is probably due to the fact that the HL combines good properties from both MAE and MSE and avoids their limitations: It bypasses large gradient backpropagation when the estimated quantity is getting closer to the real value, and it is more robust to outliers than MSE loss.

5) Time Performance Analysis for the Multimodal Setting: Finally, Table X summarizes the performances, in terms of training time, of all the different competing approaches in the case of multisource analysis (LiDAR metrics, topographic information, and Sentinel-2 imagery). We remind that all the experiments were performed on the Google Cloud Platform, Colab [43]. We can observe that the training times of all the methods are quite reasonable and affordable (spanning from a few seconds to 10 min.), with the highest running time exhibited by the $BART$ approach, that takes around 9 min. to train a regression model.

V. Discussion

To summarize, our research study proposes a novel multisource DL framework that adopts late- instead of early-fusion [37] strategy for the estimation of two forest variables with the aim to assess the combination of LiDAR metric and multispectral/multitemporal Sentinel-2 data. In addition, to the best of our literature survey, this is the first attempt to combine such sources of information via modern neural network approaches for the estimation of *total volume* and *basal area* variables in a complex forest environment. The proposed DL framework exhibits convincing performances in an operational scenario characterized by multimodal information as well as a realistic amount of available training samples.

Firstly, we have noted that the joint use of topographic information, LiDAR-derived metrics, and Sentinel-2 time series improved the estimation of both *total volume* and *basal area* with respect to the case in which only LiDAR metrics or only Sentinel-2 data are employed. For LiDAR-based *total volume* models, MAE, and *rmse* were reduced by –3.8% and –10.1%, for the *RF* and *MLP* models, respectively, and R^2 was increased by 5.6% and 10.3% for the same models when topographic data were introduced. Improvements were even slightly greater for *basal area* models. This improvement might be due to the fact that topographic information provides a kind of correction effect that compensates for the distortion of tree architecture induced by height normalization of the point cloud. This distortion increases with slope and can be considered as a source of signal artefacts. Trees growing in mountainous areas can also have curved trunks due to snow and wind. The spatial distribution of LiDAR 3-D point clouds and the derived metrics can thus change according to both slope and aspect for the same kind of stand type. Using topographic information is helpful to deal with such differences in point clouds. In the same spirit, we noticed that, despite the radiometric correction of topographic effects applied to Sentinel-2 images, differences in reflectances due to changes in exposition are still visible. Adding slope and aspect as input data in the model gives an opportunity to the neural network to better cope with these residual differences.

Secondly, Sentinel-2 data bring information mainly related to stand composition. Sentinel-2 data alone are not the best source of information to predict forest variables that are primarily driven by tree and stand structure, as also underlined in [20]. For example, when comparing the best Sentinel-2 based models to the best LiDAR-based models, including topographic data in both cases, MAE and *rmse* increased by 38.7% and 31.7%, respectively, and R^2 decreased by 28.1% for *total volume*.

However, adding Sentinel-2 data to LiDAR information led to a significant improvement in the prediction with the proposed DL framework. R^2 increased by 13.4% and by 17.7% for *total volume* and *basal area*, respectively and both MAE and *rmse* were reduced by more than 20% for the two variables. These results underline the complementarity between this information, as also underlined in the recent study proposed in [21], where the combination of the Sentinel-2 and POLSAR-derived topographic information improved the estimation of forest variables in temperate forests. At plot level, for a given volume or *basal area*, the 3-D distribution of LiDAR points and the derived metrics depend on tree species that are present in the plot. Therefore, the relationship between LiDAR-derived metrics and the targeted forest variables is dependant on stand composition. Results seem to demonstrate that this dependence can be addressed using Sentinel-2 data within an appropriate modeling framework. As a result, the quality of the predictions is unexpectedly high with regards to the complexity of the forest in the study area.

Thirdly, from a methodological point of view, when comparing competing approaches, we have observed that the *MLP* model exhibits the best behavior, in terms of evaluation metrics, for the analysis of both LiDAR-derived metrics and topographic information. We can explain this fact by the intrinsic ability of this approach to manage the different input information permitting to extract effective knowledge for the downstream regression task.

Fourthly, we have demonstrated the quality of our DL-based framework, based on late-fusion strategy, for the task of structure and biophysical variables estimation from heterogeneous (multimodal) remote-sensing data. The comparison with the strategy proposed in [37], based on an early-fusion option, advocates in favour of our framework, underlying the appropriateness of a multibranch architecture for our multimodal task. This result is completely in line with recent studies [27] that pinpoint the benefit of late-fusion approaches over early-fusion ones for multimodal remote-sensing analysis. More in detail, the obtained findings highlight that the availability of complementary remote-sensing data is not a direct synonym of performance improvements, but the way in which the diverse remote-sensing sources are combined together is crucial to fully exploit the interplay among the different input sources. This point is underlined by the empirical results we have reported in Section IV that pinpoint the ability of *MMFVE* to get the most out of the different information we can access.

Finally, we remind that our task is characterized by operational/realistic constraints, thus, resulting in a limited amount of GT information from which the relationship between multimodal remote-sensing data and the target forest variables is learnt. Despite such a data paucity setting, our approach clearly outperforms standard (i.e., *KNN* and *RF* regressor) and recent (i.e., *BART* and *SSAE*) competing strategies, demonstrating the quality of modern DL-based methods to provide competitive results also in scenarios where the amount of labelled samples is scarce. Dealing with data from multiple sources can also bring up additional issues due to the specificity of each available modality. Indeed, in this study, LiDAR and Sentinel-2 data were acquired through airborne and spaceborne configurations, respectively, which results in differences in terms of scale and resolution between these data sources. Moreover, the data georeferencing is of paramount importance in remote-sensing applications that rely on GT and reference data. In the context of multimodal data fusion, the input data on which models are trained are derived from multiple datasets, each acquired under specific conditions and preprocessed independently, and thus, misregistration among all the available modalities can negatively impact the final results. In this scenario, our proposed framework dealt with each modality independently through dedicated branches before performing a late-fusion step. This architecture allowed to take into account the peculiarities of each input data source and to process them in an adequate manner considering their differences in scale and resolution, before leveraging the interplay between all the sources in the internal-fusion step.

VI. Conclusion

In this work, we have evaluated the complementarity of LiDAR, topographic, and Sentinel-2 information for the estimation of two forest variables (*total volume* and *basal area*) to characterize temperate forest properties in a multimodal scenario. To this end, we have designed and deployed a strategy named *MMFVE*, based on a DL framework, to cope with the richness and complexity of the available multimodal input data.

MMFVE is based on a two-branch architecture with dedicated per-source encoders and a latefusion step based on the concatenation of the per-source features. Thanks to an end-to-end learning strategy, our framework is capable to effectively manage the available multimodal information associated to a particular spatial location.

The experimental evaluation with competing approaches, usually employed to cope with structural and biophysical variables estimation, highlights that the use of multimodal information alone does not guarantee an improvement of the final performances, but an effective exploitation of multimodal information tightly depends on the way the data-fusion process is conducted. More in detail, the results we have obtained on the estimation of *total volume* and *basal area* forest variables, in terms of all the considered evaluation metrics, underline the adequateness and the added value of our neural network based approach to fuse together LiDAR, topographic, and Sentinel-2 data.

Possible future works related to our research study can be devoted to a in-depth analysis of the interplay between LiDAR metrics and Sentinel-2 imagery on one hand, and topographic information on the other hand. For instance, the intensity channel derived from the laser wavelength of the LiDAR acquisition can be considered as an additional spectral band with respect to the original Sentinel-2 spectral bands due to the fact that it registers a different wavelength of the spectrum. Furthermore, evaluating which is the more appropriate way to integrate topographic information (at raster or metric level) as an additional and/or independent source of information (i.e., considering a three-branch architecture) can be another follow-up of the proposed research work. Finally, another possible avenue of research could be related to the analysis of the importance of the different input information in order to establish which LiDAR metric or spectral band the neural network retains most useful following a similar approach to the one proposed in [44].

References

[1] Y. Pan *et al.*, “A large and persistent carbon sink in the world’s forests,” *Science*, vol. 333, no. 6045, pp. 988–993, 2011.

[2] D. Simberloff, “The role of science in the preservation of forest biodiversity,” *Forest Ecol. Manage.*, vol. 115, no. 2/3, pp. 101–111, 1999.

[3] D. S. Boyd and F. M. Danson, “Satellite remote sensing of forest resources: Three decades of research development,” *Prog. Phys. Geography Earth Environ.*, vol. 29, no. 1, pp. 1–26, 2005.

[4] L. L. F. Janssen, *Principles of Remote Sensing: An Introductory Textbook*. Enschede, The Netherlands: ITC, 2004.

[5] R. DeFries, A. Hansen, B. L. Turner, R. Reid, and J. Liu, “Land use change around protected areas: Management to balance human needs and ecological function,” *Ecological. Appl.*, vol. 17, no. 4, pp. 1031–1038, Jun. 2007.

[6] J. Cabello *et al.*, “The ecosystem functioning dimension in conservation: Insights from remote sensing,” *Biodiversity Conservation*, vol. 21, no. 13, pp. 3287–3305, Dec. 2012.

[7] H. Le Goff, L. De Grandpré, D. Kneeshaw, and P. Bernier, “Sustainable management of old-growth boreal forests: Myths, possible solutions and challenges,” *Forestry Chronicle*, vol. 86, no. 1, pp. 70–76, Feb. 2010.

[8] J. F. Franklin *et al.*, “Ecological characteristics of old-growth Douglas-fir forests,” U.S. Dept. Agriculture, Forest Serv., Pacific Northwest Forest and Range Exp. Station, Portland, OR, USA, Tech. Rep. PNW-GTR-118, 1981.

[9] J. Chave *et al.*, “Tree allometry and improved estimation of carbon stocks and balance in tropical forests,” *Oecologia*, vol. 145, no. 1, pp. 87–99, Aug. 2005.

[10] S. Magnussen and P. Boudewyn, “Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators,” *Can. J. Forest Res.*, vol. 28, no. 7, pp. 1016–1031, Jul. 1998.

[11] J. Hyyppä, H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo, “Review of methods of small–footprint airborne laser scanning for extracting forest inventory data in boreal forests,” *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1339–1366, Mar. 2008.

[12] M. Bouvier, S. Durrieu, R. A. Fournier, and J.-P. Renaud, “Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data,” *Remote Sens. Environ.*, vol. 156, pp. 322–334, Jan. 2015.

[13] C. Véga, J.-P. Renaud, S. Durrieu, and M. Bouvier, “On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters,” *Remote Sens. Environ.*, vol. 175, pp. 32–42, Mar. 2016.

[14] E. Nasset, “Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser,” *Remote Sens. Environ.*, vol. 91, no. 2, pp. 243–255, May 2004.

[15] G. Vincent *et al.*, “Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure,” *Remote Sens. Environ.*, vol. 125, pp. 23–33, Oct. 2012.

[16] S. Zolkos, S. Goetz, and R. Dubayah, “A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing,” *Remote Sens. Environ.*, vol. 128, pp. 289–298, Jan. 2013.

[17] C. Vega *et al.*, “PTrees: A point-based approach to forest tree extraction from lidar data,” *Int. J. Appl. Earth Obser. Geoinf.*, vol. 33, pp. 98–108, Dec. 2014.

[18] N. Karasiak, D. Sheeren, M. Fauvel, J. Willm, J.-F. Dejoux, and C. Monteil, “Mapping tree species of forests in southwest France using Sentinel-2 image time series,” in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2017, pp. 1–4.

[19] E. Grabska, P. Hostert, D. Pflugmacher, and K. Ostapowicz, “Forest stand species mapping using the Sentinel-2 time series,” *Remote Sens.*, vol. 11, no. 10, May 2019, Art. no. 1197.

[20] S. Wittke, X. Yu, M. Karjalainen, J. Hyyppä, and E. Puttonen, “Comparison of two-dimensional multitemporal Sentinel-2 data with three-dimensional remote sensing data sources for forest inventory parameter estimation over a boreal forest,” *Int. J. Appl. Earth Obser. Geoinf.*, vol. 76, pp. 167–178, 2019.

[21] K. Ahmadi, B. Kalantar, V. Saeidi, E. K. G. Harandi, S. Janizadeh, and N. Ueda, “Comparison of machine learning methods for mapping the stand characteristics of temperate forests using multi-spectral sentinel-2 data,” *Remote. Sens.*, vol. 12, no. 18, 2020, Art. no. 3019.

[22] J.-M. Monnet, “Estimation de paramétres forestiers par Lidar aéroporté et imagerie satellitaire RapidEye: étude de sensibilité,” *Revue Française de Photogrammétrie et de Télédétection*, vol. 1, no. 211–212, pp. 71–79, Dec. 2015.

[23] Z. Wu, D. Dye, J. Vogel, and B. Middleton, “Estimating forest and woodland aboveground biomass using active and passive remote sensing,” *Photogrammetric Eng. Remote Sens.*, vol. 82, no. 4, pp. 271–281, Apr. 2016.

[24] Y. Zhang and Z. Shao, “Assessing of urban vegetation biomass in combination with LiDAR and high-resolution remote sensing images,” *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 964–985, 2021.

[25] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[26] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[27] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[28] R. Interdonato, D. Ienco, R. Gaetano, and K. Ose, "DuPLO: A dual view point deep learning architecture for time series classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 91–104, Mar. 2019.

[29] D. Ienco, R. Interdonato, R. Gaetano, and D. Ho Tong Minh, "Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 11–22, Dec. 2019.

[30] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, and S. Dupuy, "M^{3}Fusion: A. deep learning architecture for multiscale multimodal multitemporal satellite data fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4939–4949, Dec. 2018.

[31] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens*., vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[32] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote. Sens*., vol. 56, no. 2, pp. 937–949, Feb. 2018.

[33] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, "Classification of hyperspectral and LiDAR data using coupled CNNs," *IEEE Trans. Geosci. Remote. Sens*., vol. 58, no. 7, pp. 4939–4950, Jul. 2020.

[34] Y. Sun, X. Zhang, X. Zhao, and Q. Xin, "Extracting building boundaries from high resolution optical images and lidar data by integrating the convolutional neural network and the active contour model," *Remote. Sens*., vol. 10, no. 9, 2018, Art. no. 1459.

[35] B. Bigdeli, P. Pahlavani, and H. A. Amirkolaee, "An ensemble deep learning method as data fusion system for remote sensing multisensor classification," *Appl. Soft Comput*., vol. 110, 2021, Art. no. 107563.

[36] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A comprehensive analysis of deep regression," *IEEE Transaction Pattern Anal. Mach. Intell.*, vol. 42, no. 9, pp. 2065–2081, 2020.

[37] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined LiDAR and landsat 8 data," *Remote. Sens*., vol. 11, no. 12, 2019, Art. no. 1459.

[38] J.-R. Roussel *et al.*, "LiDR: An R package for analysis of airborne laser scanning (ALS) data," *Remote Sens. Environ.*, vol. 251, 2020, Art. no. 112061.

[39] D. N. Cosenza *et al.*, "Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock," *Forestry Int. J. Forest Res.*, vol. 94, no. 2, pp. 311–323, Mar. 2021.

[40] 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, Jul. 2021, Art. no. 112477.

[41] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, 2017.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[43] E. Bisong, *Google Colaboratory*. Berkeley, CA, USA: Apress, 2019, pp. 59–64.

[44] M. Campos-Taberner *et al.*, "Understanding deep learning in land use classification based on Sentinel-2 time series," *Sci. Rep.*, vol. 10, pp. 2045–2322, 2020.