A CNN-based approach for the estimation of canopy heights and wood volume from GEDI waveforms

Ibrahim Fayad, Dino Ienco, Nicolas Baghdadi, Raffaele Gaetano, Clayton Alcarde Alvares, Jose Luiz Stape, Henrique Ferraço Scolforo, Guerric Le Maire

Remote Sensing of Environment 265 (2021) 112652

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A CNN-based approach for the estimation of canopy heights and wood volume from GEDI waveforms

Ibrahim Fayad [a,*], Dino Ienco [a], Nicolas Baghdadi [a], Raffaele Gaetano [a], Clayton Alcarde Alvares [b,c], Jose Luiz Stape [b], Henrique Ferraço Scolforo [c], Guerric Le Maire [d,e]

[a] French National Research Institute for Agriculture, Food and the Environment (INRAE), CIRAD, CNRS, TETIS, Univ Montpellier, AgroParisTech, 34093 Montpellier CEDEX 5, France
[b] Unesp, Faculdade de Ciências Agronômicas, 18610-034, Botucatu-SP, Brazil
[c] Suzano SA, Estrada Limeira, 391, 13465-970, Limeira, SP, Brazil
[d] CIRAD, UMR Eco&Sols, F-34398, Montpellier, France
[e] Eco&Sols, Univ Montpellier, CIRAD, INRAE, IRD, Montpellier SupAgro, Montpellier, France

ARTICLE INFO

Edited by xxx

Keywords:
Lidar
GEDI
Dominant height
Wood volume
Eucalyptus
Brazil
Convolutional neural networks

ABSTRACT

Full waveform (FW) LiDAR systems have proven their effectiveness to map forest biophysical variables in the last two decades, owing to their ability of measuring, with high accuracy, forest vertical structures. The Global Ecosystem Dynamics Investigation (GEDI) system on board the International Space Station (ISS) is the latest FW spaceborne LiDAR instrument for the continuous observation of Earth's forests. FW systems rely on very sophisticated pre-processing steps to generate a priori metrics in order to leverage their capabilities for the accurate estimation of the aforementioned forest characteristics. The ever-expanding volume of acquired GEDI data, which to date comprises more than 25 billion acquired unfiltered shots, and along with the pre-processed data, amounting to more than 90 TB of data, raises new challenges in terms of adapted preprocessing methods for the suitable exploitation of such a huge and complex amount of LiDAR data. To overcome the issues related to the generation of relevant metrics from GEDI data, we propose a new metric-free approach to estimate canopy dominant heights (H_{dom}) and wood volume (V) of Eucalyptus plantations over five different regions in Brazil. To avoid metric computation, we leverage deep learning techniques and, more in detail, convolutional neural networks with the aim to analyze the GEDI Level 1B geolocated waveforms. Performance comparisons were conducted between four convolutional neural network (CNN) variants using GEDI waveform data (either untouched, or subsetted) and a metric based Random Forest regressor (RF). Additionally, we tested if our framework can improve the generalization of the models to different distant regions. First, the models were trained using data from all the study regions. Cross validated results showed that the CNN based models compared well against their RF counterpart for both H_{dom} and V. The RMSE on the estimation of H_{dom} from the CNN based models varied between 1.54 and 1.94 m with a coefficient of determination (R^2) between 0.86 and 0.91, while the RF model produced an accuracy on H_{dom} estimates of 1.45 m (R^2 = 0.92). For V, CNN based estimations ranged from 27.76 to 33.33 $m^3.ha^{-1}$ (R^2 between 0.82 and 0.88), while for RF, the RMSE was 27.61 $m^3.ha^{-1}$ (R^2 = 0.88). Next, model generalization was assessed by means of a spatial transfer experiment. For H_{dom}, both the CNN and RF approaches showed similar performances to a global model, however, the CNN based approach showed higher variability on the estimation accuracy, and the variability was related to the forest structure between the trained and tested data (similar tree heights yield better accuracies). For the estimation of V, considering both approaches, the accuracy was dependent on the allometric relationship between H_{dom} and V in the training and testing regions while lower accuracies on V were obtained when the testing and training regions exhibited a different allometric relationship.

* Corresponding author.
E-mail addresses: ibrahim.fayad@inrae.fr (I. Fayad), dino.ienco@inrae.fr (D. Ienco), nicolas.baghdadi@inrae.fr (N. Baghdadi), raffaele.gaetano@cirad.fr (R. Gaetano), calcarde@suzano.com.br (C.A. Alvares), jlstape@gmail.com (J.L. Stape), hscolforo@suzano.com.br (H. Ferraço Scolforo), guerric.le_maire@cirad.fr (G. Le Maire).

https://doi.org/10.1016/j.rse.2021.112652
Received 28 January 2021; Received in revised form 30 June 2021; Accepted 12 August 2021
Available online 31 August 2021

1. Introduction

Forest structural characteristics, as both a product and driver of ecosystem processes, are key factors in determining forests' carbon sequestration capacity, biomass allocation, and carbon storage (Lefsky et al., 2002; Simard et al., 2011). Moreover, canopy heights are strongly correlated with above ground biomass (AGB) (Lefsky et al., 2005), and in the context of climate change, the accurate estimation of carbon components, such as AGB from forests has become an important step in the United Nations plan for the reduction of carbon emissions due to Deforestation and forest Degradation (REDD).

In the last couple of decades, remote sensing, due to its accurate earth observation capabilities, has increasingly been used for the estimation, on local and global scales, of forest biophysical characteristics, namely forest heights and AGB. Of particular interest are the LiDAR systems. LiDAR measures the vertical structure of objects by emitting laser pulses and measuring the time difference between the transmitted emission and its echoed return. LiDAR systems can be classified into two broad groups as either discrete return or full waveform systems. Discrete LiDAR systems or multi-return LiDAR systems usually record the first and last returned echoes of targets within the travel path of the emitted light as series of x, y, and z points known as LiDAR point clouds. Some discrete LiDAR systems can also record intermediate points while the newer ones record the intensity as well (Wagner et al., 2004). The accuracy of discrete return LiDAR systems is dependent on their density within a footprint, as low-density systems can miss small targets such as treetops and therefore under-estimate forest heights (Anderson et al., 2016). Moreover, discrete LiDAR systems over densely vegetated areas may not reach the ground, leading to errors in the estimation of forest characteristics (Anderson et al., 2016). Full waveform (FW) systems on the other hand, acquire a time-varying distribution of backscattered radiation from the different targets within the illuminated surface. Therefore, FW LiDARs provide much richer information about the spatial arrangement of structures within their waveforms (Alexander et al., 2010), which can then be used to derive forest biophysical characteristics such as canopy heights and AGB (Fieber et al., 2013).

Almost two decades ago, the first spaceborne LiDAR system, the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud, and land Elevation Satellite (ICESat) became operational. ICESat\ GLAS (hereafter referred to as ICESat-1) which was originally designed to measure ice sheets was successfully used in a large body of studies for the estimation of canopy heights, and AGB (Lefsky et al., 2005; Harding, 2005; Boudreau et al., 2008; Chen, 2010; Baghdadi et al., 2014; Fayad et al., 2016; El Hajj et al., 2017; Rajab Pourrahmati et al., 2018). During its operational life from 2003 until 2009, ICESat-1 which was equipped with two laser systems, operating at the infrared (1064 nm) and visible green (532 nm) light spectrums (Schutz et al., 2005), acquired FW over vertical structures within ~60 m footprints at a temporal resolution of 1 ns (15 cm) (Schutz et al., 2005). Each acquired waveform by ICESat-1, like all FW systems, consisted of a series of multiple connected temporal modes, or peaks, representing the different reflections from an object (e.g. top of canopy cover) or different objects close together (e.g. understory and ground) (Sumnall et al., 2016). Despite the accuracies of FW LiDAR for the estimation of forest characteristics, one major challenge is the precise identification of the different features such as the top of the canopy peak, and the ground peak. Therefore, given the complexity of data in the waveform, a series of pre-processing steps are required in order to generate the variables, or metrics from the waveforms (Chauve et al., 2009). First, the connected echoes inside the waveform need to be identified and separated (e.g. to distinguish between top of canopy and ground returns). For ICESat-1, the echoes were identified using Gaussian fitting, where the waveform was decomposed into a series of Gaussian curves (Lefsky et al., 2005). This method is generally well suited for waveforms acquired over vegetation with Gaussian shaped pulses (Hancock et al., 2015). After the identification of the relevant peaks in the waveform, metrics characterizing the vegetation and terrain could then be generated, e.g. percentile relative heights (the height above ground at which the nth % of waveform energy falls below) (Næsset, 2002), waveform extent (Lefsky et al., 2002), and leading and trailing edge extents (Lefsky et al., 2005; Harding, 2005). These metrics have proven to be reliable in regression models for the estimation of both canopy heights and AGB (Baghdadi et al., 2014). Nonetheless, the success of the different metric-based methodologies to derive forest characteristics relies on the accurate estimation of these metrics.

The main source of uncertainty associated to metric values accuracy is the determination of the useful part of the waveform, usually referred to as signal start (top of the canopy) and signal end. The threshold ($n \cdot \sigma + \mu$, where μ and σ are the average and standard deviation of the background noise respectively) used to determine these two variables is not constant. Indeed, this threshold varies based on several factors, such as the laser transmit power and noise which affect the returned waveform levels, the target application, site characteristics, forest types, acquisition time, etc. For ICESat-1, several thresholds were used, and ranged from $3 \cdot \sigma$ (Sun et al., 2008) to $4.5 \cdot \sigma$ (Lefsky et al., 2007).

Recently, the commissioning of GEDI on board the ISS, offers a new frontier for the characterization of forest dynamics. The GEDI mission constitutes an unprecedented opportunity to generate new forest related products compared to the ones derived from previous remote sensors. GEDI has a much higher firing frequency than the previous ICESat-1, and therefore has a much higher sampling density (Dubayah et al., 2020). For reference, in its first 18 months of operation, GEDI managed to acquire over 25 billion unfiltered shots globally, requiring a storage space of more than 90 TB. Moreover, since GEDI is a FW LiDAR system, sophisticated processing of the waveforms is still required for the generation of the useful metrics. The Land Processes Distributed Active Archive Center (LP DAAC) offers main GEDI data as two distinct datasets (Dubayah and Luthcke, 2020a,b). The L1B data product contains the geolocated waveforms, while the L2A data product contains the extracted metrics.

Moreover, given the aforementioned challenges regarding waveform metrics extraction, GEDI waveform metric values are currently proposed using six different configurations for signal smoothing widths and thresholds on noise to suit a wide variety of applications (Dubayah et al., 2020). Nonetheless, given the unprecedented high density of data, the iterative, time-consuming, and error prone approach used to determine the best metrics for a particular waveform acquired over a given study site could prove to be a major setback. Therefore, an approach that can automatically extract, with high precision, feature rich information from the waveforms could be of great interest.

With the advancement of modern computational capabilities, machine learning based methodologies such as neural networks or, more broadly, deep learning (DL) approaches are becoming an important tool in the geoscience field (Reichstein et al., 2019). Beyond the consolidated success of deep learning in the general field of computer vision (LeCun et al., 2015) such approaches are showing high performances in different applications related to the analysis of remote sensing data (Yuan et al., 2020) in the fields of land cover classification (Ienco et al., 2019), agricultural yield prediction (Kim et al., 2019) and estimation of vegetation parameters (Hosseini et al., 2019).

The deep learning methods can be clustered into two main families: the first one involves methods especially tailored for spatial learning while the second family proposes solutions for sequential learning (LeCun et al., 2015). However, there is a recent growing interest in blending these two perspectives. This is particularly suitable for geoscience applications where the phenomenon to characterize is featured by both spatial and temporal correlations (Reichstein et al., 2019). Furthermore, an important and worthy of interest point related to deep learning methods is their ability to work directly with the original signal information without requiring computational-demanding pre-processing stages (Zhu et al., 2017). In contrast to standard machine learning methods that exploit handcrafted features (e.g. metrics extracted from LiDAR waveforms such as canopy and ground positions, canopy height

profiles, etc.), DL approaches are especially suitable to directly manage the original signal (e.g. unprocessed LiDAR waveforms) avoiding most of the complex pre-processing stages.

The aim of this paper is to provide a metric-free approach to deal with biophysical variable estimation from GEDI LiDAR data. Here, we propose two Convolutional Neural Networks (CNN) based approaches to leverage deep learning tools to cope directly with GEDI waveforms while avoiding time-consuming and task-unrelated metric generation. More in detail, we propose two Convolutional Neural Network (CNN) based approaches. The first CNN treats the waveform signal as a sequential data information and it analyzes the signal by means of one dimensional convolutions on the spectral dimension; the second CNN-based method tackles the waveform by reshaping it in a two dimensional format with the aim to increase the internal signal contrast. While the former strategy has been widely adopted in the remote sensing community to deal with 1D signal processing like satellite image time series (Ienco et al., 2020) and hyperspectral data (Audebert et al., 2019), the latter approach (reshaping a 1D signal to a 2D representation and, subsequently employ a two dimensional convolutional neural network) is a less explored and novel research contribution.

The two CNN-based solutions for the GEDI waveforms analysis are deployed to estimate the height and the wood volume of fast-growing Eucalyptus plantations in Brazil. These plantations provide a valuable case study due to the homogeneous canopy cover and the availability of high quality field measurements in comparison to natural forests where *in situ* data are scarce. CNN based approaches are compared against a traditional metric based approach using a random forest regressor and performances are evaluated by means of traditional summary statistics (e.g. coefficient of determination, root mean square error, etc.).

The paper is organized as follows. The data and the study area are introduced in Section 2. The convolutional neural network based methodologies are presented in section 3. Section 4 describes the experimental settings, model parameters, and the results. Finally the discussion and conclusions are presented in Sections 5 and 6, respectively.

2. Study regions and datasets

2.1. Study area

The study area is located in Brazil covering five regions, across a latitudinal gradient (From 2°15′S to 24°43′S, Fig. 1), and covering different climate and soil types. Maranhão (MA) is located in a typical equatorial region with dominant monsoon rainfall (1200 to 2500 mm/year). MA's intricate relief and complex spatial distribution of almost all tropical soil classes, results in a truly "genotype × environment × management" puzzle that foresters must solve for guaranteed wood production in a sustainable way. Bahia (BA) and Espírito Santos (ES) are located in a tropical coastal region with strong rainfall anisotropy (800 to 1500 mm/year), that directly affects wood productivity from near-shore towards hinterland. Mato Grosso do Sul (MS) is located in a tropical region (1200 to 1500 mm/year), with some subtropical variability (rare frost), it is the most environmentally homogeneous among the five study regions, resulting in less variation in wood productivity within region. São Paulo (SP) climate is mainly subtropical with orographic effects (1100 to 2000 mm/year). Heavy frost-days are frequent in the southern part of SP. The complex relief and a wide range of deep and shallow tropical soils results in a large variability of wood productivity across the region.

Across the five regions, clonal seedlings of mainly *E. grandis* (W. Hill) and *E. urophylla* (S.T. Blake) as well as different types of hybrids are planted in rows at a density of 1000–1667 trees/ha. The wood productivity of the plantations has an average of 40 $m^3 \cdot ha^{-1} \cdot year^{-1}$, with 80% of the stands being between 30-50 $m^3 \cdot ha^{-1} \cdot year^{-1}$ and some stands could reach values as high as 60 $m^3 \cdot ha^{-1} \cdot year^{-1}$. At harvest time, stand dominant height is between 20 and 35 m (for 80% of stands) with a stand volume between 180 and 300 $m^3.ha^{-1}$. The plantations are managed locally by stand units (~50 ha), where the same management is applied for each stand: planting, harvesting and weed control, genetic material, soil preparation and fertilization. The plantations are generally characterized by sparse understory and herbaceous strata, due to chemical weeding during the first year, the high competitive strength of Eucalyptus, and the closing of the canopy. Tree height is homogeneous within a stand, with 95% of the trees having heights at ±10.5% around the average tree height in plot inventories. Eucalyptus plantations

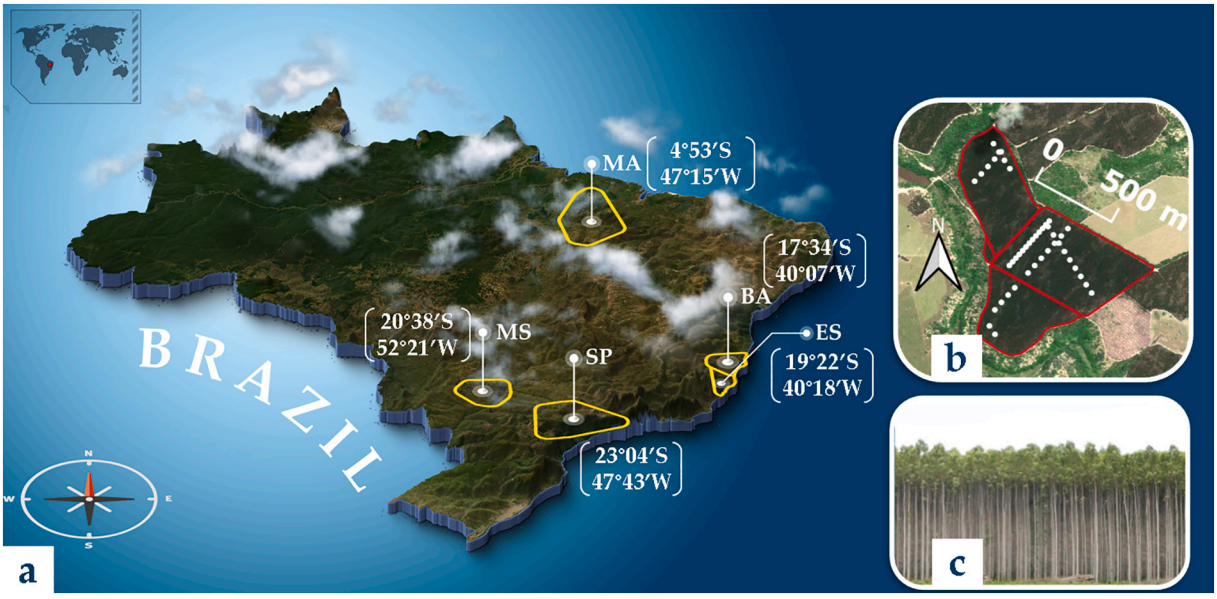

Fig. 1. (a) Location of the five study regions; (b) an example of GEDI tracks (white circles) over some stands (red polygons); (c) Eucalyptus stands during harvest illustrating the clearly separated crown and trunk strata. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

exhibit a simple structure, with a tree crown strata of 3 to 10 m in width above a "trunk strata" with few Eucalyptus leaves and few understories (Fig. 1c). The relative crown depth is age dependent and varies from ~95% (1 year old trees) to ~25% (7 year old trees) (Binkley et al., 2020). Finally, the "soil strata" is mainly constituted of litter accumulation of branches and leaves, with some patches of herbaceous species.

2.2. Inventory measurements

For this study, 566 Eucalyptus stands were selected which correspond to stands where GEDI footprints were acquired between April 20, 2019 and September 4, 2019. An additional 50 m internal buffer strip from the stand borders was used to account for any footprint geolocation errors and to avoid footprints that match the boundary between the stand of interest and the surrounding medium. These 566 Eucalyptus stands were also selected because they had field inventories performed by the Suzano company close to GEDI's acquisition dates (i.e. time difference lower than 2 months) on their 2 to 8 permanent inventory plots. Permanent inventory plots had an area of approximately 400 m^2 and were systematically distributed throughout the stand with a density of one plot per 10 ha. They included 30 to 100 trees with an average of 58 trees. During a field inventory, the diameter at breast height (DBH, 1.3 m above the ground) of each tree in the inventory plot, the height of a central subsample of 10 trees and the height of the largest 10% of trees in terms of DBH (dominant trees) were measured. The mean height of the 10% of the largest trees defined the dominant height of the plot (henceforth referred to as H_{dom}). H_{dom}, basal area and age on the inventory date were then used in local volume equations to estimate the plot's total and merchantable volume (merchantable volume is a tree's volume up to 6 cm stem diameter with bark). The allometric equation used to estimate wood volume is very precise with an accuracy (RMSE) of 0.5 $m^3.ha^{-1}$ assuming 1180 trees on average per hectare (R^2 close to 0.99) (Scolforo et al., 2019).

Finally, as the dates of the inventory measurements were different from GEDI acquisition dates, only data with a date difference between GEDI acquisitions and inventory fewer than 2 months were used. In fact, on these fast growing plantations, a 2-month difference results in an up to 50 cm growth in H_{dom} and 10 $m^3.ha^{-1}$ in V (depending on genetic material, pedoclimatic conditions, season, and age). However, this reasonable compromise allows keeping a large number of stands including a large variability of age and growing conditions. Fig. 2 shows the distribution of field measurements of H_{dom} and wood volume (V).

2.3. GEDI data

2.3.1. Processing of GEDI waveforms

GEDI is equipped with three lasers emitting 1064 nm light pulses, with one of the lasers split into two beams, and therefore GEDI acquires waveforms over eight tracks of data (through dithering of the laser beams) (Dubayah et al., 2020). GEDI beams illuminate a surface on the ground, the so-called footprint, with a 25 m diameter, at a frequency of 242 Hz, over which 3D structures are measured. The footprints are separated by ~60 m (center to center) along the track, and the tracks are separated by ~600 m. GEDI measures vertical structures using a 1064 nm laser pulse, and the echoed waveforms are digitized to a maximum of 1246 bins with a vertical resolution of 1 ns (15 cm), corresponding to a maximum of 186.9 m of height ranges, with a vertical accuracy over relatively flat, non-vegetated surfaces of ~3 cm (Dubayah et al., 2020).

As described in the GEDI Algorithm Theoretical Basis Document (ATBD) (Dubayah and Luthcke, 2020a,b), the received waveforms are first smoothed to reduce the noise in the signal, and thus permitting the determination of the useful part of the waveform within the corresponding footprint. Waveform smoothing is performed by means of a Gaussian filter (*Smooth Width*) with a width of 6.5 ns. After smoothing, two locations in the waveform denoted as *searchstart* and *searchend* are determined. *searchstart* and *searchend* are respectively the first and last positions in the signal where the signal intensity is above the following threshold:

$$\text{threshold} = \text{mean} + \sigma \cdot v \quad (1)$$

where '*mean*' is the mean noise level, 'σ' is the standard deviation of noise of the smoothed waveform, and 'v' is a constant currently set at 4 (Dubayah et al., 2020).

After determining the locations of *searchstart* and *searchend*, the region between them, denoted as the waveform extent, is extended by a predetermined number of sample bins, currently set to 100 bins at both sides. Inside the waveform extent, the highest (*toploc*) and lowest (*botloc*) detectable returns are determined (Fig. 3). *toploc* and *botloc* respectively represent the highest and lowest locations inside the waveform extent were two adjacent intensities are above a threshold. The threshold equation used to determine *toploc* and *botloc* is the same as Eq. (1). In the ATBD, the value of 'v' used to determine *toploc* is named *Front_threshold* and *Back_threshold* for *botloc*.

Finally, the locations of the distinctive peaks or modes in the waveform, such as the ground peak, or top of canopy peaks are determined using a second Gaussian filtering of the waveform section between *toploc* and *botloc*, and then finding all the zero crossings of the first

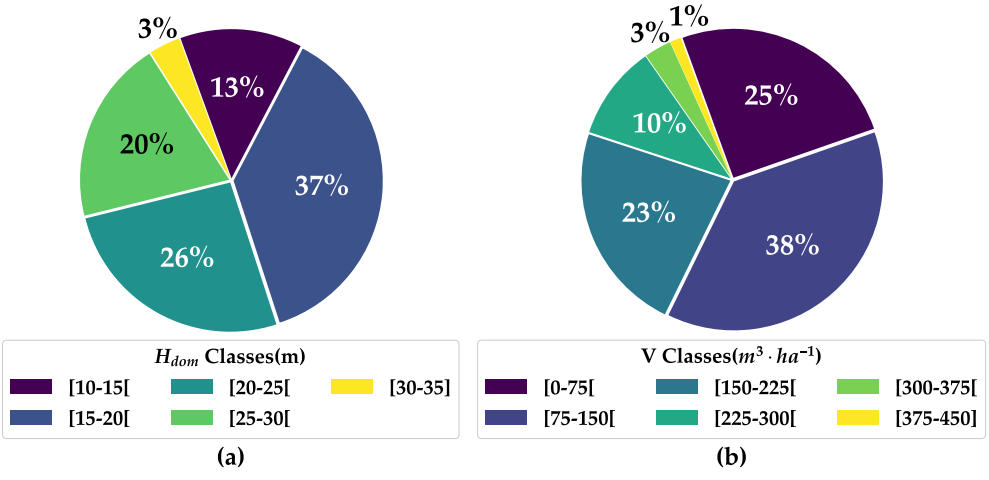

Fig. 2. Distribution of dominant canopy heights (H_{dom}, a) and wood volume (V, b) densities from field inventories of the 566 Eucalyptus stands.

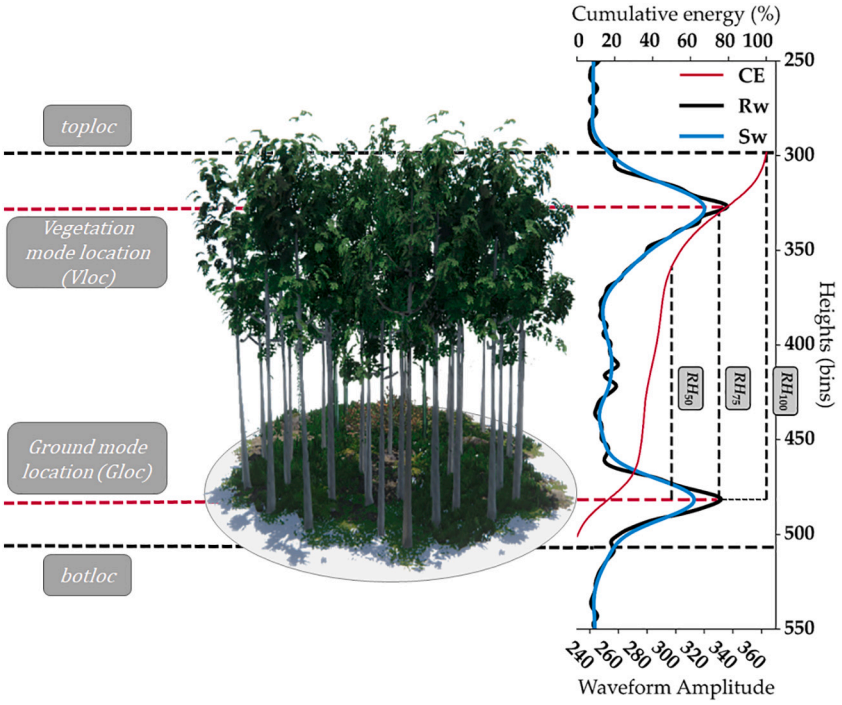

Fig. 3. Example of an acquired GEDI waveform (Rw) over an Eucalyptus stand (H_{dom} = 25.9m; V= 230.7 $m^3.ha^{-1}$), its smoothing (Sw) and corresponding waveform metrics. (red curve) The cumulative energy of the waveform (CE) between *botloc* and *toploc* and the corresponding relative heights (RH_n) at different percentages 'n' for the same waveform. 1 bin = 1 ns and corresponds to 15 cm sampling distance in the waveform. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

derivative of the filtered waveform (Fig. 3). The width of the second Gaussian filter is denoted *Smoothwidth_zcross*. Finally, the position of the ground return within the waveform is determined using the position of the last detected mode.

Waveform metric values are extracted using thresholds on Smoothwidth_zcross, Front_threshold, and Back_threshold. Currently there are six configurations (henceforth referred to as algorithms) with different thresholds on these variables, which are used to extract waveform metrics for a variety of acquisition scenarios. The thresholds for each of the six algorithms (denoted a1 to a6) are presented in Appendix A. The use of each of the six different algorithms could lead to different estimates of the waveform metrics, and could in turn lead to six different canopy height estimates.

Over forest stands, the recorded waveforms are multi-modal in shape, with each mode representing a reflection from a distinct surface height. Fig. 3 shows a typical waveform over an Eucalyptus forest stand on relatively flat terrain. Over flat terrain, the first Gaussian corresponds to a reflection from the top of the canopy while the last Gaussian mostly refers to the lowest point in the footprint, i.e. the ground surface.

In this study, several machine learning regression models were tested in order to estimate the stand dominant height H_{dom} (m) and stand merchantable wood volume V ($m^3.ha^{-1}$) from GEDI data. These Models could be classified into two categories. (1) Models that rely solely on the waveform signals, and (2) models that rely on metrics extracted from the waveforms. The GEDI metrics used in this study are published by the Land Processes Distributed Active Archive Center (LP DAAC). For each acquisition, the received waveforms, their geolocation (longitude, and latitude), as well as their acquisition times were extracted from the L1B product. Next, we extracted from L2A for each beam the following metrics pre-processed with algorithms a1 through a6: (1) the position within the waveform of *toploc* and *botloc*, and (2) The Relative height metrics at 10% intervals from the determined ground position (0%) to *toploc* (100%) (RH_n, $10\% \leq n \leq 100\%$, *step*10%). RH_n represents the height between the ground position and the location at n% of cumulative energy (Fig. 3). In theory RH_{100} should correspond to the real dominant canopy height (H_{dom}). Nonetheless, noise, uncertainties on the position of the detected ground return, and vegetation and ground variability, affect the accuracy of this metric (Comparison between field-measured H_{dom} and different RH_n values extracted using algorithm a1 (RH_{90} to RH_{100}) can be found in Appendix B). Variables used for the estimation of the stand dominant height H_{dom} (m) and stand merchantable wood volume V ($m^3.ha^{-1}$) are summarized in Table 1.

2.3.2. Filtering of GEDI waveforms

GEDI acquisitions are sensitive to atmospheric conditions (e.g. clouds) which render their usability (Fayad et al., 2021). Therefore, waveforms that meet any of the following criteria were removed:

- waveforms with reported elevations that are significantly higher than the corresponding elevations in the SRTM DEM (Baghdadi et al., 2014). In essence, we removed all waveforms where the absolute difference is higher than 100 m.
- waveforms with a difference between waveform extent (height difference between *botloc* and *toploc*,W_{ext}) and (*Gloc-Vloc*) is higher than 400 bins (60 m)

In total, 6166 footprints were acquired over the 566 reference stands between April 2019 and September 2019 with the majority of these footprints (5682) providing exploitable waveforms.

3. Methodology

3.1. Convolutional neural network based methodology

In this section we introduce the two different approaches that we have designed in order to cope with the estimation of tree heights and

Table 1
List of all variables to be used for the canopy height and wood volume estimation models.

Variable	Gedi Dataset
Footprint geolocation	L1B
Waveform samples	L1B
Search start	L2A
Search end	L2A
Relative canopy height (botloc (0%) to toploc (100%))	L2A

wood volumes from the GEDI data. Both approaches are based on Convolutional Neural Networks (CNN) (LeCun et al., 2015). CNN strategies have largely demonstrated their value and effectiveness in the general signal processing domain (LeCun et al., 2015) dealing with both one dimensional (e.g. time series) as well as two dimensional (image) signals (Huang et al., 2017). The basic block of such neural network architectures is implemented by the convolutional operation (Zhu et al., 2017) that allows to deal with the underlying (spatial or temporal) auto-correlation that intrinsically characterizes many types of signal data. Due to the fact that GEDI information consists of a univariate waveform signal, traveling from the atmosphere to the surface, we can naturally deploy one dimensional (1D) CNN with the aim to exploit the autocorrelation existing on the wave dimension. Such approach is presented in Section 3.1.1. Another, less common way, to manage the GEDI signal is to reorganize the waveform into the shape of a two dimensional structure. Once the signal is reorganized in such way, a (2D) CNN method is used to estimate the biophysical parameters. Despite the fact that reorganizing a waveform signal as a 2D information seems inappropriate, in the case of GEDI information, due to the high sparsity that such signals exhibit, this encoding strategy retains the structural information carried out by the signal and it highlights internal signal contrast. Such approach is described in Section 3.1.2. Regarding both strategies, we get inspiration by previous works in convolutional neural network devoted to design simple but effective architectures for both 1D (Ienco et al., 2020) and 2D (Springenberg et al., 2015) signals. For instance, we adopt small kernel (size equal to 3) to: i) capture more localised complex features in the input signal; ii) extract a vast amount of features which can be useful in later layers; and iii) permit to go further in the depth of the neural network architecture allowing to learn more complex and non-linear relationships.

Another interesting point related to the use of neural network based solutions, in the context of regression tasks, is their ability to deal simultaneously with multiple output variables to regress starting from the same input data, thus, working in a multi-output regression setting. This means that, in our analysis, both approaches (1D-CNN and 2D-CNN) are able to estimate both height and wood volume biophysical parameters with only one model, avoiding the need to use specific models for each estimated variable. Finally, in Section 3.1.3 we detail the way in which the neural networks are learnt. We adopt the same training procedure for both models.

3.1.1. One dimensional CNN

When the GEDI signal is considered as a wave signal, one dimensional convolution (1D-CNN) can be directly applied on the raw data with the goal to leverage the information similarly to what is done for time series analysis (Fawaz et al., 2019). Table 2 and Fig. 4 report the architecture of the 1D-CNN we have conceived to deal with the biophysical parameters estimation from the GEDI waveform. We follow general principles applied in the design of Convolutional Neural Networks (Radosavovic et al., 2020), where the number of filters along the network structure grows and the convolutional operations are followed by a non linear activation function (Rectifier Linear Unit in our case, ReLU) and Dropout. Our 1D-CNN has ten blocks where the first eight involve parameters associated to convolution operations. We adopt filters with a kernel size of 3. The seventh block only flattens the features extracted by the convolutional kernels while blocks 8 and 9 combine all the features together by means of fully connected (dense) layers. Finally, block 10 performs multi-output regression with two output neurons, one for each biophysical variable. We underline that, following common practices for neural network based regression, no activation function is associated to the output layer of the network (Lathuilière et al., 2020).

Table 2
Architectures of the one dimensional Convolutional Neural Network (1D-CNN) where *nf* are the number of filters, *k* is the one dimensional kernel size, *s* is the value of the stride while *act* is the activation function.

	CNN1D
Block 1	Conv(nf=96, k=3, s=1, act=ReLU) DropOut()
Block 2	Conv(nf=96, k=3, s=1, act=ReLU) DropOut()
Block 3	Conv(nf=96, k=3, s=2, act=ReLU) DropOut()
Block 4	Conv(nf=192, k=3, s=1, act=ReLU) DropOut()
Block 5	Conv(nf=192, k=3, act=ReLU) DropOut()
Block 6	Conv(nf=192, k=3, s=2, act=ReLU) DropOut()
Block 7	Flatten()
Block 8	Dense(n=128, act=ReLU)
Block 9	Dense(n=128, act=ReLU)
Block 10	Dense(n=2, act=None)

3.1.2. Two dimensional CNN

In order to analyze the GEDI waveforms by means of a two dimensional CNN (2D-CNN), as a first step each waveform was first encoded in a two dimensional (2D) format, which is then fed as input for the 2D-CNN.

In literature, there are several research studies that address the transformation of one dimensional (1D) signal to 2D representations (Xu et al., 2020; Zhang et al., 2019; Dias et al., 2020). The majority of these studies leverage point-wise distance matrices in order to model interactions between subsequent portions of the 1D signal (short interactions) as well as interactions between portions of the 1D signal that are far away from each other (longer interactions). Unfortunately, the results of such methods is a new representation with a size that is, at least, quadratic w.r.t. the size of the original signal. These transformations not only increase the data size, but also make the use of convolutional neural networks challenging due to the fact that the input data will be represented by an image with millions of pixels (for instance, in the case of GEDI signal, they will provide 2D representations with approximately 1 200 $\times$ 1 200 pixels). With the objective to provide a 2D representation with a reasonable size that can also be suitable as input for standard (two dimensional) convolutional neural networks, we propose a simple, yet effective, 2D transformation for GEDI data. The preprocessing procedure is illustrated in Fig. 5.

Given the GEDI signal (in Fig. 5 the signal has a length of 1 225 bins), we compute the square root of the waveform length (in this case 35) and then, rearrange the elements of the 1D signal in a two dimensional matrix (in the example, the two dimensional matrix has a size of 35 $\times$ 35). The rearrangement process is performed by splitting the GEDI signal in vectors having the same size (35 bins) and then, vertically stacking together such information. We can observe in Fig. 5 the result of the transformation (2D representation block). If the squared root of the wave length is not a whole number, a zero padding (red colored values) is performed on the original data to achieve a signal length that provides an integer value when square rooted. This 2D representation leverages the fact that GEDI waveforms are characterized by a high degree of sparsity resulting in a stable signal with a limited number of peaks. When the waveform is arranged as a 2D representation (as opposed to the one dimensional one), the contrast between an information peak (e.g. vegetation or ground peaks) and its surrounding (e.g. tree trunks that don't reflect much light due to their relatively smaller surface area compared to canopy cover or the ground) is exacerbated due to the signal stationarity. Thus, the gradient around that specific information peak (considering the 2D representation) is generally higher than the gradient around the same information peak when the original one dimensional representation is considered.

At this point, similarly to what we have done for the 1D-CNN, we conceive and design a 2D-CNN model according to the general principles proposed by Radosavovic et al. (2020). Table 3 and Fig. 5 summarize the

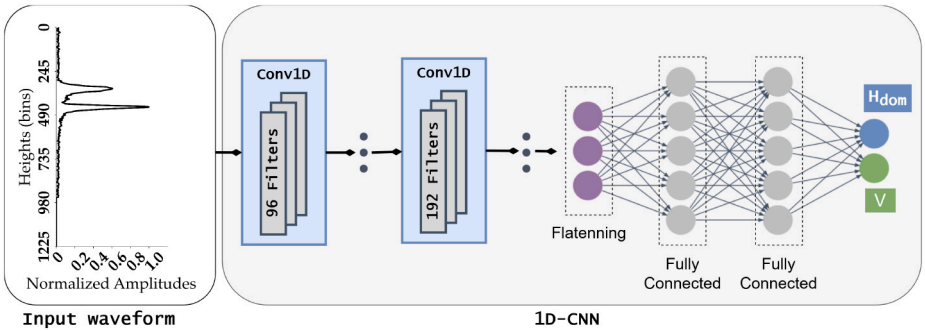

Fig. 4. Graphical illustration of the 1D-CNN architecture.

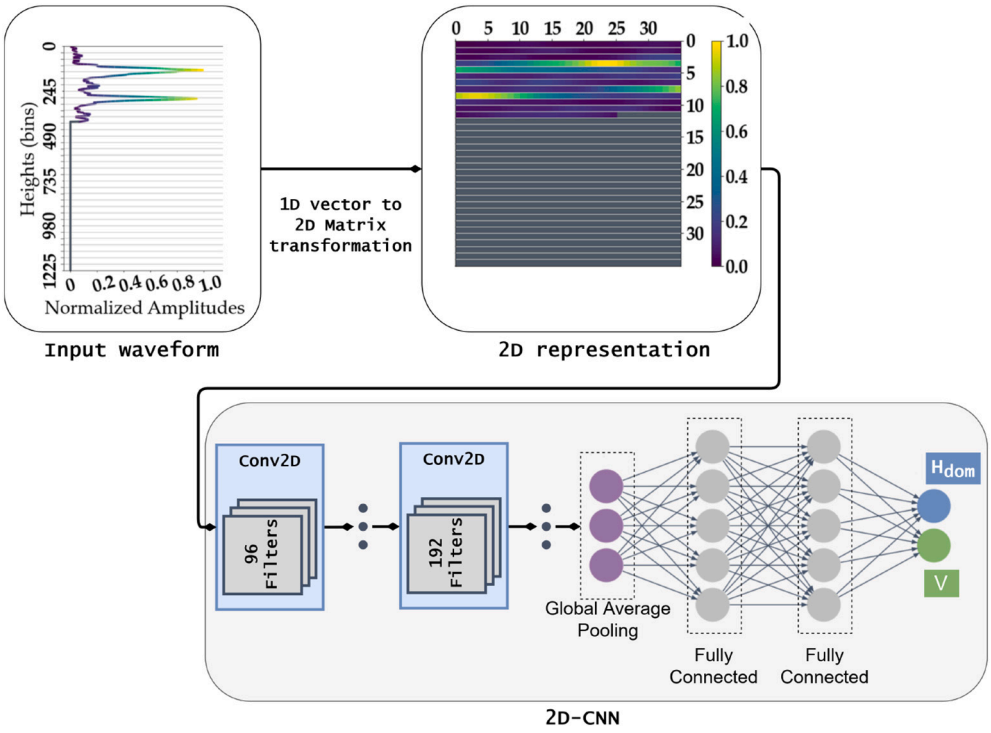

Fig. 5. Graphical illustration of the 2D-CNN architecture.

2D-CNN architecture.

Our 2D-CNN has ten blocks where the first six involve parameters associated to convolution operations as well as batch normalization (Ioffe and Szegedy, 2015). We adopt filters with a kernel size equal to 3. The seventh block performs a global average pooling (GAP) over the feature map dimension while block 8 and block 9 combine all the features together by means of fully connected (dense) layers. Finally, similarly to the case of 1D-CNN, block 10 performs multi-output regressions with two output neurons, one for each biophysical variable. We adopt batch normalization and global average pooling layers since, for the 2D-CNN architecture, they empirically ameliorate the estimation performances regarding the corresponding architectural choices we have made for the 1D-CNN model.

3.1.3. *Training procedure for multi-output regression*

The training stage for both CNN-based models is the same. Since we are dealing with a regression task, to compute the cost function associated to the optimization of the neural network parameters, we adopt the Mean Absolute Error (MAE) (Lathuilière et al., 2020) metric defined as follows:

$$\mathrm{MAE}(Y, Y') = \frac{1}{|Y|}\sum_{i=1}^{|Y|} ||Y_i - Y'_i||_1 \tag{2}$$

where Y is the vector of the original values to estimate and Y' is the values estimated by a particular method. $||\cdot||_1$ is the L1-norm (equivalent to the MAE) that considers the absolute value of the difference between the two terms Y_i and Y'_i. Since we are coping with multi-output regression, we define the loss function for our CNN-based models as follows:

$$\mathrm{Loss}(H, V, H', V') = \mathrm{MAE}(H, H') + \mathrm{MAE}(V, V') \tag{3}$$

where H (resp. V) is the original value for the biophysical variable Height (resp. Volume) while H' (resp. V') are the values estimated by any of our CNN-based models (1D-CNN or 2D-CNN). The multiple output tasks are simply managed by summing the individual errors on each of the biophysical estimated variables. We remind that the network parameters are learnt end-to-end via common gradient descend strategy.

Table 3
Architectures of the two dimensional Convolutional Neural Network (2D-CNN) where *nf* are the number of filters, *k* is the two dimensional kernel size while *s* is the value of the stride.

CNN2D	
Block 1	Conv(nf=96, k=3, s=1) BatchNormalization() ReLU() DropOut()
Block 2	Conv(nf=96, k=3, s=1) BatchNormalization() ReLU() DropOut()
Block 3	Conv(nf=96, k=3, s=2) BatchNormalization() ReLU() DropOut()
Block 4	Conv(nf=192, k=3, s=1) BatchNormalization() ReLU() DropOut()
Block 5	Conv(nf=192, k=3) BatchNormalization() ReLU() DropOut()
Block 6	Conv(nf=192, k=3, s=2) BatchNormalization() ReLU() DropOut()
Block 7	GlobalAveragePooling()
Block 8	Dense(n=128, act=ReLU)
Block 9	Dense(n=128, act=ReLU)
Block 10	Dense(n=2, act=None)

4. Experimental evaluation

In this section we evaluate the two convolutional neural networks (CNN) based methods we have previously introduced for the estimation of H_{dom} and V through GEDI waveforms data. For comparison purposes, a random forest (RF) model is also considered as standard competitor according to the recent work proposed in Fayad et al. (2021). With the aim to assess the behavior of the proposed CNN based methods, we perform different tests. First, we evaluate the single-output regression version of the competing approaches in which a model for each estimated variable is considered. Second, we evaluate the behavior of the multiple-output approaches as introduced in Section 3.1.3 compared with the multiple-output random forest regressor model. Third, we inspect the estimation results obtained by our proposals as well as the RF method at regional level to understand how a global model (learnt from samples coming from different regions) behaves locally. Finally, we design a spatial transfer experiment in which we calibrate the considered models on a particular study region and we deploy them considering (spatially) different ones (Fig. 1a). In addition, regarding the first three experiments, we also differentiate the analysis between CNN based approaches that take as input the full GEDI waveforms or the subset GEDI waveforms data.

4.1. Model performance evaluation

To assess how the different models perform, a 5-fold cross validation was used. Moreover, we imposed that footprints belonging to the same stand were assigned exclusively to one of the data partitions (training or test) with the aim to avoid possible spatial bias in the evaluation procedure. On average, we employ a training (resp. test) set with a number of footprints equal to 4 452 (resp. 1 065) GEDI waveforms. Finally, model performances were assessed using the coefficient of determination (R^2), the root mean square error (RMSE), and the root mean squared percentage error (RMSPE). R^2, RMSE, and RMSPE are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \quad (4)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \quad (5)$$

$$\mathrm{RMSPE} = 100\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(\frac{y_i - \widehat{y_i}}{y_i}\right)^2} \quad (6)$$

where y_i is the observed value, $\widehat{y_i}$ the estimated value, $\overline{y_i}$ is the mean of all the observed values, and n is the sample size.

The RF based competitor was built using a set of 500 trees (higher tree count slightly increased model accuracy), with a tree depth equal to the square root of the number of available factors. The RF model estimates H_{dom} and V using the relative height metrics (RH_n, $10\% \leq n \leq 100\%$, step 10%) as presented in section 2.3.1. In addition, we will use the RH_n metric values that were extracted using algorithm a1 (Smooth Wdith: 6.5 ns, Smoothwidth_zcross: 6.5 ns, Front_threshold: 3 ns, Back_threshold: 6 ns) since it produced the metrics with the highest accuracy to *in situ* H_{dom} (Fayad et al., 2021). The comparison of the relative height values to *in situ* H_{dom} was made using the RH_{100} metric values extracted from the six processing algorithms (a1 to a6, Appendix C).

Finally, both CNN-based models were optimized via a standard gradient descent procedure through the Adam optimizer (Kingma and Ba, 2015) with a learning rate of 1×10^{-4}, a batch size of 64 and a dropout rate of 0.2.

4.2. Experimental settings

In sections 3.1.1 and 3.1.2, two CNN based approaches are presented. For both approaches, the full and subset waveforms were independently tested. The full waveform represents the returned energy for each time bin as recorded by GEDI. In essence, the full waveform contains the interaction of LiDAR and the surface within the footprint as well as noise. The subset waveform represents a section of the full waveform between *searchstart* and *searchend* (Fig. 3), where most of the noise has been cropped. Due to the fact that GEDI waveforms can have different lengths, we perform zero padding to uniform all the waveforms resulting in a signal with a length of 1 444 and 1 225 samples for the full and subset GEDI waveforms, respectively. A list of the CNN models as well as their input data are summarized in Table 4.

To evaluate if the CNN based approach can be (spatially) transferred from a study region to another one, we propose two scenarios on which H_{dom} and V are estimated. In the first scenario, the best CNN based model, and the best RF model are trained with GEDI waveforms data and field measurements coming from the MS region (region with the highest GEDI footprint count) and deployed on the other study regions (Fig. 1a). In the second scenario, the two best models (best CNN and best RF) are trained with GEDI data from both MS and SP regions (Fig. 1a). The choice of adding SP and not any other region was motivated by the fact that SP has the second largest count of GEDI footprints after MS.

The trained model performance was assessed using RMSE, RMSPE, and R^2. A randomized t.test was also used to compare the estimation of the full model (calibrated on all regions), and the models calibrated on

Table 4
List of the tested CNN models.

Model designation	Data used
$CNN1D_s$	subset waveforms in original 1D representation
$CNN1D_f$	full wavefroms in original 1D representation
$CNN2D_s$	subset wavefroms in 2D representation
$CNN2D_f$	full waveforms in 2D reprentation

datasets from one or two regions (van der Voet, 1994). The rand.t test is a permutation test that tests the equal distribution of errors based on the mean difference between two competing model's mean squared errors. This test will allow to evaluate if the behavior of the model changes with respect to the calibration region. In this paper, we used 199 permutations with $\alpha = 0.05$.

4.3. Results

4.3.1. Single-output based estimation

The results presented in Fig. 6a show that the estimation of H_{dom} using the single-output RF model produces estimates with an accuracy (RMSE) on H_{dom} of 1.45 m and a coefficient of determination (R^2) of 0.92. In contrast, the 1D-CNN (Fig. 6b and c) are less accurate with an RMSE of 1.73 m (R^2 of 0.89) when using the subset waveforms ($CNN1D_s$), and the accuracy decreased to an RMSE of 1.94 m (R^2 of 0.86) when using the full waveforms ($CNN1D_f$). Nonetheless, the encoding of GEDI waveforms as 2D matrices (Fig. 6d and e) allows to reach performances comparable to the RF approach. Indeed, for the 2D-CNN model the estimation of H_{dom} is 1.54 m (R^2 of 0.90) for the subset waveforms ($CNN2D_s$) and 1.61 m (R^2 of 0.90) when using the full GEDI waveforms $CNN2D_f$.

The accuracy of the RF approach in estimating the wood volume (V) (Fig. 7a) is 27.61 $m^3.ha^{-1}$ with an R^2 of 0.88. Moreover, the performance of the convolutional neural networks approach are in-line with the results obtained for the estimation of H_{dom}. In essence, the 1D-CNN methods (Fig. 7b and c) produce the least accurate results with an RMSE of 30.08 $m^3.ha^{-1}$ ($R^2 = 0.85$) when the subset GEDI waveforms is considered ($CNN1D_s$) and an RMSE of 33.33 $m^3.ha^{-1}$ ($R^2 = 0.82$) for the full waveforms ($CNN1D_f$). The 2D-CNN models showed behaviours similar to the RF based approach with an RMSE of 28.42 $m^3.ha^{-1}$ ($R^2 = 0.87$) (Fig. 7d, $CNN2D_s$) when using the subset GEDI waveforms, and a slightly better RMSE (27.76 $m^3.ha^{-1}$, $R^2 = 0.88$) for the full GEDI waveforms (Fig. 7e, $CNN2D_f$).

4.3.2. Multiple-output based estimation

As described at the beginning of the experimental evaluation, we also estimate H_{dom} and V simultaneously using RF and CNN considering a multiple-output regression scenario. Regarding the CNN based approaches, these models are as described in Sections 3.1.1 and 3.1.2. Figs. 8 and 9 show that the simultaneous estimation of H_{dom} and V using RF gave an RMSE of respectively 1.47 m ($R^2 = 0.92$) and 27.60 $m^3.ha^{-1}$ ($R^2 = 0.88$) (respectively Fig. 8a, and Fig. 9a).

The proposed CNN based methods capacity in estimating H_{dom} ranged from 1.53 m (RMSE) ($CNN2D_s$, Fig. 8d) to 2.02 m ($CNN1D_f$, Fig. 8c). RMSE values for wood volumes ranged from 26.83 $m^3.ha^{-1}$ ($CNN2D_s$, Fig. 9d) to 34.83 $m^3.ha^{-1}$ ($CNN1D_f$, Fig. 9c). These results underline that the multiple-output regressions strategy, using either the RF or the CNN-based approaches, achieves identical performances to the ones obtained using the single-output regression frameworks. Nonetheless, the multiple-out regression strategy is advantageous as it reduces model training (only one model has to be trained instead of two). Such approach is also computationally less expensive, with faster deployment time (only one model is deployed to produce the estimation of both biophysical variables).

4.3.3. Intra-region inspection with a global estimation model

The results presented in Table 5 show that the intra-region results for the ES, MA, SP, and MP study regions (Fig. 1a) are very close, with a root mean squared percentage error (RMSPE) ranging between 6.8 to 7.9 % on the estimation of H_{dom} with the CNN based approach and from 6.8 to 8.2 % with the RF based approach.

Moreover, for the MA study region, the CNN model is slightly more accurate on the estimation of H_{dom} with an RMSE of 1.48 m against 1.67 m obtained with the RF model (Table 5), while for the MS study region, the RF model showed 6.9% better accuracy than CNN (RMSE of 1.35 m with RF against 1.45 m with the CNN-based approach). The estimation on H_{dom} obtained for the BA study region is the least accurate with either RF or CNN, and exhibiting an RMSE of 1.79 m (RMSPE=11.4 %, $R^2 = 0.70$) with CNN and an RMSE of 1.71 m (RMSPE=11.7 %, $R^2 = 0.67$) with RF (Table 5).

The intra-region behavior on the estimation of V showed more variability than H_{dom}, and was observed using both methods (CNNs and RF) (Table 6). The estimation of V over the ES, SP, and MS study regions is the highest with an RMSPE between 15.3 and 21.8 % with the CNN-based strategy and between 16.1 and 22.3 % with RF. Wood volume estimation accuracy over BA is a bit less precise compared to the ES, SP, and MS regions using both the CNN and RF approaches, with an RMSPE of 26.4 % ($R^2 = 0.71$) and 26.0 % ($R^2 = 0.73$), respectively (Table 6). Wood volume estimation performance for MA (Fig. 1a) is the lowest one with respect to the other regions using both methods. For MA, the CNN approach exhibits better wood volume estimation accuracy than RF,

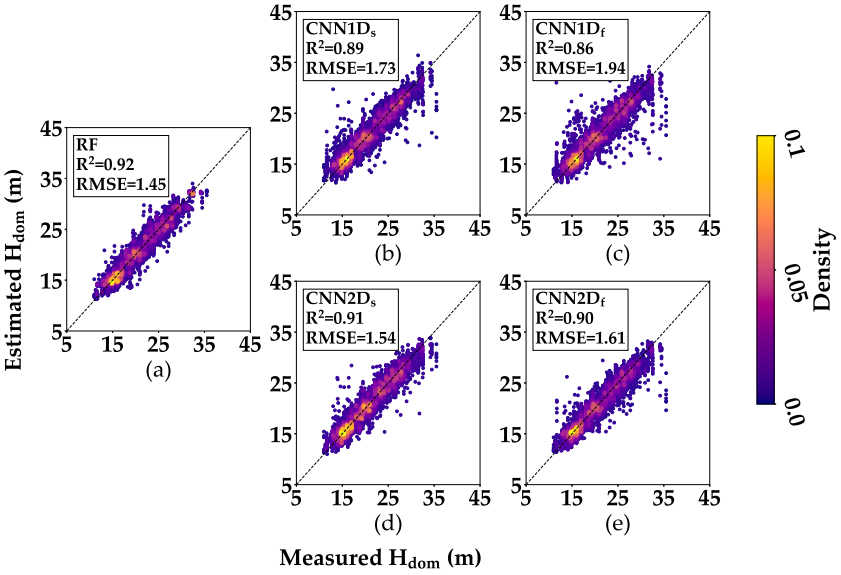

Fig. 6. Comparison of measured vs. estimated H_{dom} using single-output RF and CNN based models. (a) random forest. (b and c) 1D-CNN using subset and full GEDI waveforms respectively. (d and e) 2D-CNN using subset and full GEDI waveforms respectively.

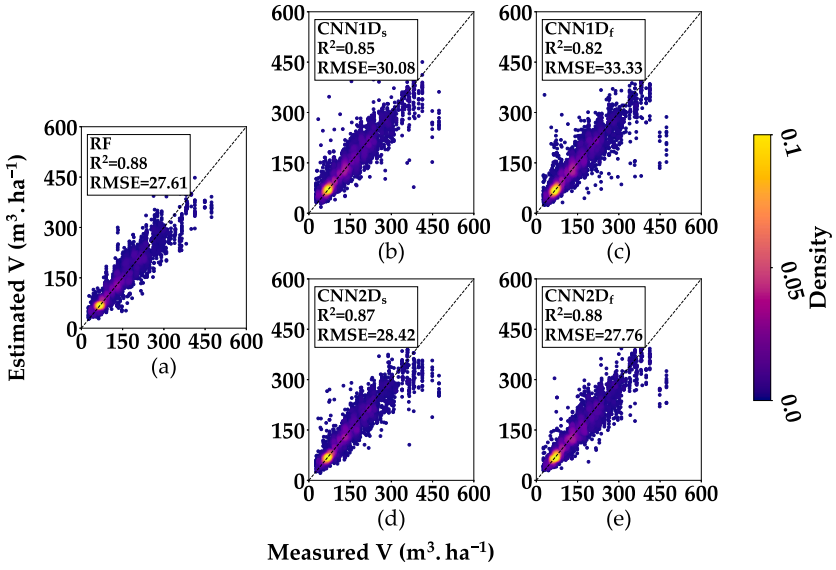

Fig. 7. Comparison of measured vs. estimated wood volume (V) using single-output RF and CNN based models. (a) random forest. (b and c) 1D-CNN using subset and full GEDI waveforms respectively. (d and e) 2D-CNN using subset and full GEDI waveforms respectively.

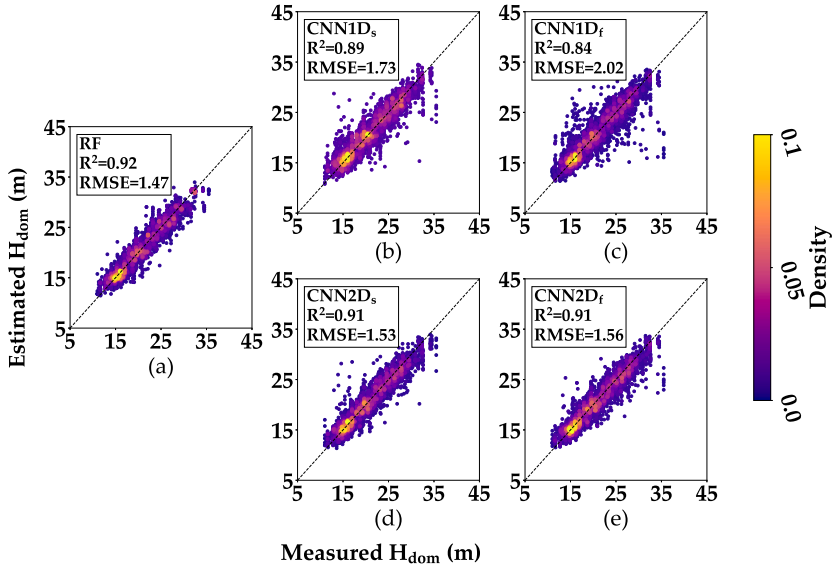

Fig. 8. Comparison of measured vs. estimated H_{dom} using multiple-output RF and CNN based models. (a) random forest. (b and c) 1D-CNN using subset and full GEDI waveforms respectively. (d and e) 2D-CNN using subset and full GEDI waveforms respectively.

with an RMSPE of 24.4 % (RMSE=30.32 $m^3.ha^{-1}$, $R^2 = 0.65$) against an RMSPE of 37.6 % (RMSE=35.92 $m^3.ha^{-1}$, $R^2 = 0.50$) for the RF model.

4.4. Assessment of the spatial transfer behavior of the competing approaches

Table 7 reports the results achieved by the different methods considering the spatial transfer experiment. We can observe that the accuracy on the estimation of the H_{dom} variable degraded considering both approaches when only samples from MS were considered at the training stage. For the CNN based model, an increase in RMSE between 5 (ES region) and 44 cm (MA region) is observed, and between 5 (BA region) and 28 cm (ES region) with the RF model. Conversely, we can observe that the RF model had an increase in the estimation performance of H_{dom} over SP when this method is trained only with the MS data. Indeed, for SP, we observed a decrease in RMSE from 1.71 m (Table 5) to 1.56 m (Table 7). Leveraging an additional region to enrich the training datasets (GEDI waveforms data from MS and SP) does not seem to increase the estimation performance of H_{dom} using either the 2D-CNN or the RF approaches. For the CNN-based approach, while the RMSE related to the estimation of H_{dom} increased by 10 and 25 cm for BA and MA, respectively, it decreased by 23 cm for the ES (Table 7). The RF behaved almost the same, considering the estimation of H_{dom}, when data from MS or MS and SP regions are considered as training sets (Table 5).

Finally, despite the difference in performances on H_{dom} obtained by the 2D-CNN and the RF models when the training data involves waveforms coming from all the study regions, and those obtained with the same models but trained over MS or MS plus SP, the estimations are only statistically significant for MA with the 2D-CNN approach, and ES with the RF approach (Table 7). For MA with the 2D-CNN approach, the results of the rand.t test for the two scenarios (MS and MS+SP) showed a p-value ≤ 0.05, meaning that the null hypothesis of equal distribution of

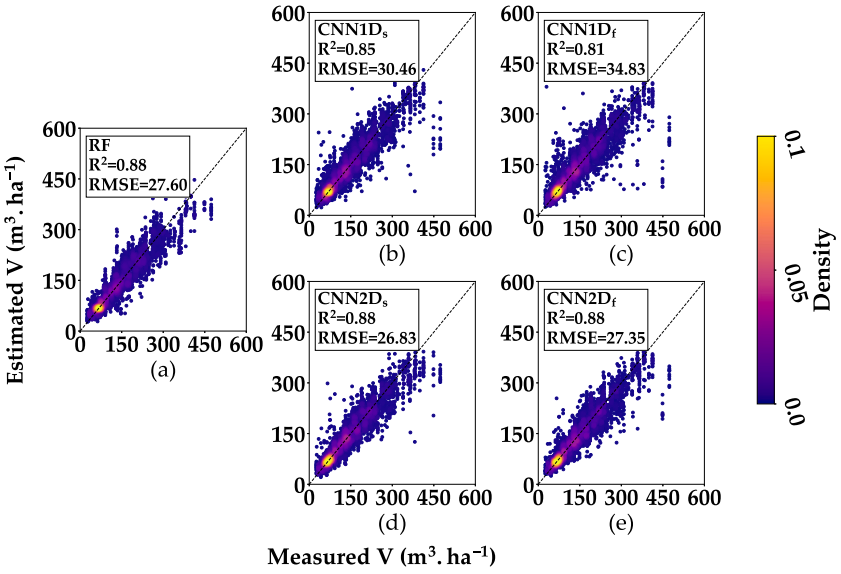

Fig. 9. Comparison of measured vs. estimated wood volume (V) using multiple-output RF and CNN based models. (a) random forest. (b and c) 1D-CNN using subset and full GEDI waveforms respectively. (d and e) 2D-CNN using subset and full GEDI waveforms respectively.

Table 5
intra-region accuracy on the estimation of H_{dom} using the 5-fold cross validated results from the 2D full waveform CNN ($CNN2D_f$) and the random forest model using GEDI metrcis extracted with algorithm a1 (RF). Both models were calibrated on data from all regions. RMSE is expressed in m, while RMSPE is the root mean squared percentage error. Region locations are shown in Fig. 1a

Region	Gedi footprint count	$CNN2D_f$			RF		
		RMSE	RMSPE	R^2	RMSE	RMSPE	R^2
BA	502	1.79	9.7	0.82	1.71	9.3	0.83
ES	727	1.43	7.9	0.80	1.48	8.1	0.79
MA	613	1.48	6.8	0.87	1.67	8.2	0.84
SP	1068	1.76	7.2	0.90	1.71	6.8	0.90
MS	2586	1.45	7.5	0.92	1.35	7.2	0.93

Table 6
intra-region accuracy on the estimation of V using the 5-fold cross validated results from the 2D full waveform CNN ($CNN2D_f$) and the random forest model using GEDI metrcis extracted with algorithm a1 (RF). Both models were calibrated on data from all regions. RMSE is expressed in $m^3.ha^{-1}$, while RMSPE is the root mean squared percentage error.

Region	Gedi footprint count	$CNN2D_f$			RF		
		RMSE	RMSPE	R^2	RMSE	RMSPE	R^2
BA	502	30.61	26.4	0.71	29.47	26.0	0.73
ES	727	18.76	15.7	0.77	17.57	16.6	0.80
MA	613	30.32	24.4	0.65	35.92	37.6	0.50
SP	1068	36.90	15.3	0.84	32.81	16.1	0.87
MS	2586	22.61	21.8	0.92	25.27	22.3	0.90

errors is rejected. For the RF approach, the null hypothesis is rejected for ES considering both settings (MS and MS+SP as training regions) (Table 7).

Table 8 summarizes the results in terms of RMSE regarding the estimation of V. For the 2D-CNN model trained only on MS (first scenario), it achieves similar performances on BA, ES and MA regions compared to the results achieved considering waveforms from all the study regions with a maximum difference of 2.66 $m^3.ha^{-1}$ in terms of RMSE.

Moreover, for these three regions (BA, ES, and MA) the rand.t test underlines that the obtained estimation of the variable V is statistically close to that obtained considering previous experiments in which the model is trained on waveforms coming from the whole set of study regions.

Generally, the estimation of V with the RF model showed more variability in the first scenario (Table 8). Here, the RMSE on the estimation of V varied between 1.53 and 4.29 $m^3.ha^{-1}$ (Table 8) in comparison to the RF trained considering waveform metrics coming from different study regions. Moreover, the rand.t test results indicate that the estimation of V over the ES, MA, and SP regions with the first scenario are statistically significant compared to those obtained with the RF trained on waveforms coming from different study regions (p-values $\leq$ 0.05, Table 8) indicating a different distribution of errors. Similarly to the results obtained when MS and SP are employed as training regions for the estimation of H_{dom}, the additional information carried out by a different study region is not guaranteed to yield better performance for the estimation of V. For example, the 2D-CNN model performances, over ES and MA, decreased by 4.70 and 8.52 $m^3.ha^{-1}$, respectively, in terms of RMSE when MS and SP are considered as a training set against considering only data coming from MS (Table 8) and the estimation of V becomes statistically significant w.r.t. the one achieved by the 2D-CNN model when waveforms from all the study regions are contemplated (p-value $\leq$ 0.05, Table 8).

5. Discussion

The results presented in this study show that a CNN based approach using GEDI waveforms can be used to obtain high accuracies of canopy heights and wood volume estimates, with a root mean square percentage error (RMSPE) varying between 7.96 and 9.95 % on canopy heights and between 21.45 and 32.02 % on wood volume. These varying accuracies depend on the CNN variant (1D CNN or 2D CNN trained with full or subset GEDI waveforms) that was used. Nonetheless, the results achieved by the CNN-based approaches with the waveforms on both the H_{dom} and V biophysical variables are comparable to those obtained using conventional estimation models such as random forests (RF) which uses a computationally intensive approach for the generation of GEDI metrics from the waveforms to estimate forest characteristics. Indeed, with RF, the RMSPE on H_{dom} and V is 7.45 and 22.46 % respectively. These results confirm that the CNN methods can automatically extract useful information from LiDAR waveforms in order to estimate H_{dom} and V on par with classical methodologies based on metrics that require expert knowledge to be produced and selected.

Table 7
inter-region accuracy on the estimation of H_{dom} using two training datasets (MS, and MS+SP) from the 2D full waveform CNN (CNN) and the random forest model using GEDI metrcis extracted with algorithm a1 (RF). RMSE is expressed in meters (m), RMSPE is expressed in %, while rand.t p-value is the randomization t-test p-value.

Scenario	CNN			rand.t	RF			rand.t
	RMSE	RMSPE	R^2	p-value	RMSE	RMSPE	R^2	p-value
MS → BA	2.02	9.59	0.76	0.0850	1.76	8.99	0.82	0.6250
MS → ES	1.48	8.25	0.79	0.5850	1.76	10.06	0.70	0.0050
MS → MA	1.92	8.26	0.79	0.0050	1.71	8.33	0.83	0.2050
MS → SP	2.07	8.38	0.86	0.0650	1.56	7.55	0.92	0.0700
MS + SP → BA	1.92	9.12	0.79	0.2800	1.79	8.92	0.81	0.3350
MS + SP → ES	1.66	9.03	0.73	0.2150	1.77	10.09	0.70	0.0050
MS + SP → MA	1.67	7.61	0.84	0.0050	1.64	8.28	0.84	0.9400

Table 8
inter-region accuracy on the estimation of V using two training datasets (MS, and MS+SP) from the 2D full waveform CNN (CNN) and the random forest model using GEDI metrcis extracted with algorithm a1 (RF). RMSE is expressed in $m^3.ha^{-1}$, RMSPE is expressed in %, while rand.t p-value is the randomization t-test p-value.

Scenario	CNN			rand.t	RF			rand.t
	RMSE	RMSPE	R^2	p-value	RMSE	RMSPE	R^2	p-value
MS → BA	33.27	20.96	0.65	0.2800	27.94	19.39	0.75	0.5300
MS → ES	19.38	17.18	0.76	0.9300	21.86	20.73	0.69	0.0050
MS → MA	31.07	26.02	0.63	0.2400	38.07	34.31	0.44	0.0050
MS → SP	44.31	17.52	0.77	0.0050	36.33	17.88	0.84	0.0050
MS + SP → BA	31.10	20.24	0.69	0.8800	27.79	19.82	0.76	0.5500
MS + SP → ES	24.08	20.54	0.63	0.1125	25.15	24.85	0.59	0.0162
MS + SP → MA	39.59	30.51	0.40	0.0015	42.18	39.03	0.31	0.0756

From a user's point of view, CNN based methods usually require a more sophisticated approach for the trained model procedure and subsequent optimization of the hyperparameters in comparison to random forests. CNN models' training times can also be higher than the one exhibited by random forests. For example, over our study area, the full RF model (trained over all regions) required around three minutes to complete a 5-fold training validation combination while the CNN based model took around twelve minutes. Nonetheless, these disadvantages are overshadowed by other factors. First, metric based methodologies require that waveform metrics be as accurately estimated as possible, whereas, similarly to ICESat-1, the thresholds used to extract these GEDI waveform metrics are tightly related to the study site and the typology of the studied forests (Fayad et al., 2014) and often these metrics are obtained via multiple trial and error tests. Currently, the LP DAAC provides six different values of each GEDI waveform metric which are issued from six different processing algorithms. Secondly, GEDI data are provided via two main different products (c.f. Section 2.3.1), the process to explore which particular information related to GEDI waveform metrics is necessary to support the study of forest characteristics means the access to additional data, and requires more computational resources in terms of both storage and\ or pre-processing. For instance, for the period between April 18, 2019 and September 02, 2020, the GEDI sensor produced more than 90 TB of L1B and L2A data, with L1B, the data product containing the waveforms, representing more than 54 TB. Therefore, while CNN required more time for model training (12 minutes vs 3 minutes), metric based methodologies require additional storage capacity as well as additional time to perform the data pre-processing stage to extract the required metrics, which becomes a hindrance especially when large study areas are considered (Computational performance for both approaches can be found in D). For these reasons, metric-free models such as CNN-based solutions are worthy of interest.

Four variants of CNNs for the estimation of H_{dom} and V were tested in this work. These models differed in their representation of GEDI waveforms. Two models used the 1D representation of the waveform, while the remaining two used the waveform as 2D encoded matrices. The results showed that the 1D representation produced less accurate results than its 2D counterpart did, especially when the full waveforms were used, and this for both H_{dom} and V estimates.

The better performance of the 2D-CNN models is tightly related to the representation we adopted for the waveform signal. As shown in Fig. 5, typically, the GEDI waveform is featured by a stationary signal with a limited number of peaks. This signal characteristic influences the design of the neural network architecture where max pooling operations are preferred w.r.t. average pooling ones. When the signal is arranged in a 2D shape, the contrast between an information peak (e.g. vegetation or ground peaks) and its surroundings (e.g. tree trunks that don't reflect much light due to their relatively smaller surface area compared to canopy cover our the ground) is exacerbated in the 2D matrix due to signal stationarity. Conversely, in the case when the original (sequential) waveform is considered, the contrast between a peak and its spatial context (surroundings) is smoothed and this fact does not allow the model to focus on the informative portions of the signal. Because of the 2D waveforms representation, the filters of the 2D-CNN model are well adapted to recognize useful portions of the signal since they can exploit a more contrasted representation helpful to discard between high and low informative content. In addition, it is worth noting that the 2D-CNN involves around 1M parameters while the 1D-CNN contains around 7M parameters.

For multiple-output 2D CNNs, the difference in accuracy between the model using the full waveforms and the one using the subset waveforms was imperceptible. For H_{dom}, the multiple-output 2D-CNN model showed less than 3 cm difference in terms of RMSE between the model using the full waveforms and the one using the subset waveforms, and less than 1 $m^3.ha^{-1}$ on the estimation of V. However, this difference in accuracy on the estimation of both H_{dom} and V is negligible considering that the model exploiting the full waveforms does not require the pre-processing step related to the determination of the useful part of the waveform. Thus, eliminating another degree of freedom that demands additional computational effort as well as a choice among possible methods to perform the selection of the subset waveforms.

To study the viability of model transferability across the study regions, first we analyzed the accuracy on H_{dom} and V from the global models (models trained using data from all the study regions) over each study region.

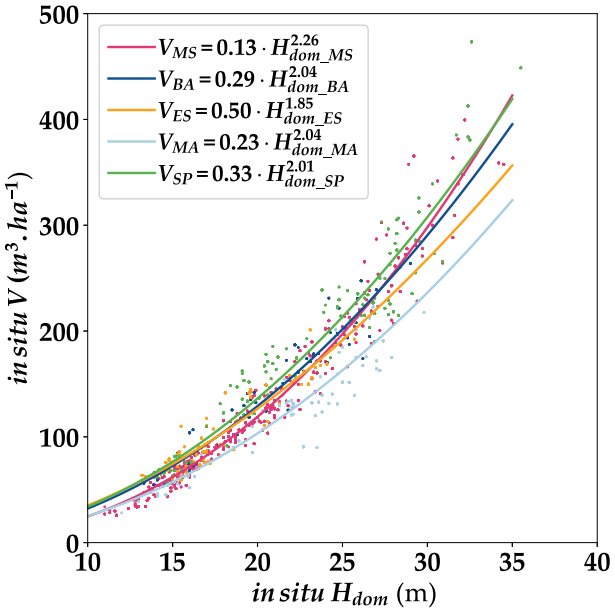

Fig. 10. *In situ* wood volume (V) as a function of *in situ* stand dominant heights (H_{dom}) for the five study regions. Region locations are shown in Fig. 1a

Regarding H_{dom}, both the multiple-output 2D-CNN and RF were able to estimate canopy heights with good precision, with more or less the same accuracies for each region. However, for V, the accuracies showed more variability across the regions, with V estimates for MA showing the worse accuracies with an RMSE of 30.32 $m^3.ha^{-1}$ ($R^2 = 0.65$) with the 2D-CNN model, and 35.92 $m^3.ha^{-1}$ ($R^2 = 0.50$) with the RF model. This was not unexpected given that GEDI, as all waveform based LiDAR sensors, measures only the vertical structure of objects within the footprint, and thus the echoed waveforms do not contain enough information about the trunk volume of the measured canopies. The direct link between GEDI FW information and H_{dom} leads to high R^2 even for a general equation across regions. The wood volume estimation from GEDI is therefore due to the existing relationship between H_{dom} and V, as observed in Fig. 10. However, such allometric relationship changes from one region to another as can be seen in Fig. 10, where MA had the lowest V/H_{dom} ratio, and therefore, this explains why a globally trained model did not produce high accuracies in this study region. Nonetheless, the 2D-CNN approach was able to generalize better on the wood volume variable than the competing random forest model (Table 6). Indeed, the RF model could not estimate accurately the wood volume over MA, while the CNN based model can improve significantly the RMSE, RMSPE and R^2 on that region. These results suggest that deep learning algorithms extract and exploit more information from the LiDAR waveform linked to local/regional stand characteristics than the information included in the GEDI metrics, which have improved the use of a general common model across regions for V.

Regarding the transferability of our proposed approach, both the 2D-CNN, and the RF models were able to successfully estimate H_{dom} using a model trained on a different study region. However, the RF model seemed to perform slightly better (similar accuracies across the four regions). For the 2D-CNN approach, training a model using data from MS, produced the highest accuracies over ES, and lower accuracies for BA, MA, and SP (Table 7). This is due to the fact that ES, as MS, had on average lower tree heights ($\overline{H_{\mathrm{dom}}} = 19.5m$) compared to BA, MA, and SP ($\overline{H_{\mathrm{dom}}} = 22.3m$), indicating that the neural network when trained over MS, was not trained with enough waveforms acquired over higher trees. This also explains the higher accuracies obtained over BA, and MA, when data from SP ($\overline{H_{\mathrm{dom}}} = 23.1m$) were added to the training model. In contrast, the RF was able to better generalize H_{dom} even when trained over lower trees. Indeed, the RF model trained with data from either MS, or MS+SP, produced the same accuracies on H_{dom}. For the estimation of V, both the 2D-CNN and the RF approaches behaved similarly. For both approaches, the successful transfer of the model, as stated previously, relies on the constancy of the allometric relationship of H_{dom} and V between GEDI data from the training and the testing sites. Moreover, the use of data from an additional study region to the training model does not guarantee the increase of accuracy on the estimation of V, as underlined by the (MS+SP) scenario. Indeed, the RMSE on the estimation of V for MA (lowest V to H_{dom} ratio) increased from 31.07 $m^3.ha^{-1}$ with scenario (MS) to 39.59 $m^3.ha^{-1}$ when adding GEDI data from SP (highest V to H_{dom} ratio). Additional experiments regarding the transferability can be found in Appendix E

6. Conclusions

The results presented in this study have proven that a CNN based on waveforms could be the next step in the estimation of forest biophysical parameters such as canopy heights and above ground biomass with accuracies similar to traditional approaches.

Traditional approaches rely on an extracted set of metrics from the waveforms for the intended application. While these metrics could be accurate in representing different canopy features (e.g. location of canopy and ground modes), they can be affected by possible effects on the acquired waveforms related to instrumental biases and atmospheric or environmental factors. Thus, the extraction of these metrics relies on several trial and error iterations as well as a priori knowledge related to a particular study site. Conversely, our proposed CNN based framework has shown the ability to directly process the full LiDAR waveforms avoiding the storage and computation of ad-hoc metrics that may, or may not, be suitable for a specific case study.

Finally, the spatial transfer assessments have also highlighted the ability of our framework to cope with inter-region variability. These results confirm that the CNN methods can automatically extract useful information from the LiDAR waveform analysis to predict H_{dom} and V on par with classical GEDI metrics that require expert knowledge to be produced and selected.

Nonetheless, while the CNN approach has the ability to better spatially generalize both H_{dom} and V, the generalization of V was more challenging, as V can only be accurately estimated in regions having similar allometric relationships between H_{dom} and V. This finding is not a limitation of our approach, but rather an inherit limitation of the full waveform LiDAR sensors and their inadequacy to directly measure the volume of the target entities. Therefore, future works can be devoted to data fusion approaches with the aim to overcome such limitations. For instance, jointly exploit both LiDAR based data and data from other sensors such as the one from the upcoming P-band BIOMASS mission.

Funding

This research received funding from the French Space Study Center (CNES, TOSCA 2021 project) and the National Research Institute for Agriculture, Food and the Environment (INRAE). Suzano SA Company supported the forest-field data collection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the GEDI team and the NASA LPDAAC (Land Processes Distributed Active Archive Center) for providing GEDI data. The authors acknowledge Suzano's researchers

Italo Ramos Cegatta, Renan Tarenta Meirelles Brasil and Carla Foster Feria for their technnical support and the CIRAD 2020 Suzano project.

Appendix A. Threshold and smoothing parameters

Table A.9
The different parameters used in each of the six algorithms for the anaylsis of the received waveforms.

Algorithm	Smooth width	Smoothwidth_zcross	Front_threshold	Back_threshold
a1	6.5	6.5	3	6
a2	6.5	3.5	3	3
a3	6.5	3.5	3	6
a4	6.5	6.5	6	6
a5	6.5	3.5	3	2
a6	6.5	3.5	3	4

Appendix B. Comparison between H_{dom} and RH_n

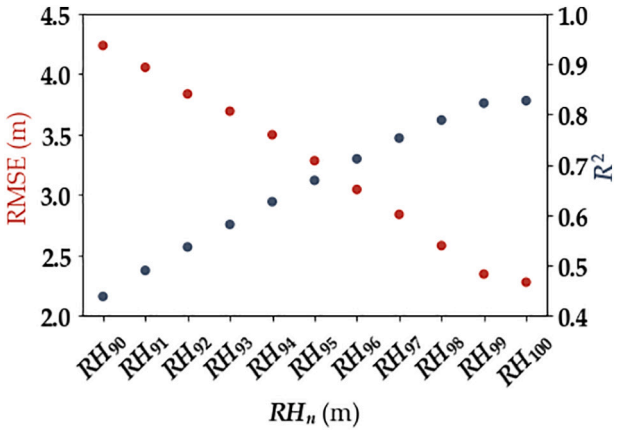

Fig. B.11. Accuracy comparison on the estimation of H_{dom} using different relative height metric (RH) values.

Appendix C. Accuracy of RH_{100} values extracted using a1 to a6

Table C.10
Accuracy of RH_{100} on the estimation of H_{dom}. a1 to a6 correspond to the configuration used to extract the RH_{100}.

	a1	a2	a3	a4	a5	a6
RMSE (m)	**2.30**	4.14	2.34	2.31	7.81	5.25
R^2	**0.83**	0.34	0.78	0.82	0.21	0.27

Appendix D. Models' computation times

The computation times reported in Table D.11 are based on an Intel® Xeon® W-2133 CPU @ 3.60GHz and an NVIDIA® GeForce® 1080Ti GTX GPU. Regarding the 2D CNN model, the pre-processing stage comprises the normalization of the waveforms and the transformation into the 2D representation. This process is done on CPU using a single core, but it can be parallelized to decrease the pre-processing times, especially for larger datasets. The training of the 2D CNN model is done on the GPU, and this too can take advantage of parallel GPU computation. Regarding the competing RF model, training is done on 10 CPU cores. And while the training times of the competing RF model is less than the CNN based model, the bottleneck of metric based models is the pre-processing stage which takes several hours and requires human intervention in order to choose the best metric values.

Table D.11
Comparison of computational times for the best performing CNN based model and the competing RF model.

Model	Data used	Pre-processing time	Training time	Inference time
2D CNN	full waveforms	< 2s (~3340 waveforms/s)	~12 minutes	< 1m
RF	waveform extracted metrics	Several hours	~3 minutes	< 1m

Appendix E. Additional assessment of the spatial transfer behavior

Table E.12
Inter-region accuracy on the estimation of H_{dom} using as training data the combination MS+MA via the 2D full waveform CNN (CNN) and the random forest model using GEDI metrics extracted with algorithm a1 (RF). RMSE is expressed in meters (m), RMSPE is expressed in %, while rand.t p-value is the randomization t-test p-value.

Scenario	CNN				RF			
	RMSE	RMSPE	R^2	rand.t p-value	RMSE	RMSPE	R^2	rand.t p-value
MS + MA → BA	1.60	8.17	0.85	0.3950	1.40	7.18	0.89	0.7850
MS + MA → ES	1.65	9.44	0.73	0.2000	1.62	8.97	0.75	0.0050
MS + MA → SP	2.01	8.77	0.87	0.0050	1.67	8.29	0.90	0.6000

Table E.13
Inter-region accuracy on the estimation of V using as training data the combination MS+MA via the 2D full waveform CNN (CNN) and the random forest model using GEDI metrics extracted with algorithm a1 (RF). RMSE is expressed in $m^3.ha^{-1}$, RMSPE is expressed in %, while rand.t p-value is the randomization t-test p-value.

Scenario	CNN				RF			
	RMSE	RMSPE	R^2	rand.t p-value	RMSE	RMSPE	R^2	rand.t p-value
MS + MA → BA	27.71	18.14	0.76	0.5900	26.37	17.22	0.78	0.0350
MS + MA → ES	18.60	17.61	0.78	0.5750	18.91	17.07	0.77	0.0100
MS + MA → SP	37.61	17.77	0.83	0.0050	35.41	17.03	0.85	0.0050

References

Alexander, C., Tansey, K., Kaduk, J., Holland, D., Tate, N.J., 2010. Backscatter coefficient as an attribute for the classification of full-waveform airborne laser scanning data in urban areas. ISPRS J. Photogramm. Remote Sens. 65, 423–432. https://doi.org/10.1016/j.isprsjprs.2010.05.002.

Anderson, K., Hancock, S., Disney, M., Gaston, K.J., 2016. Is waveform worth it?. a comparison of lidar approaches for vegetation and landscape characterization. Remote Sens. Ecol. Conserv. 2, 5–15. https://doi.org/10.1002/rse2.8.

Audebert, N., Saux, B.L., Lefèvre, S., 2019. Deep learning for classification of hyperspectral data: A comparative review. IEEE Geosci. Remote Sens. Mag. 7.

Baghdadi, N., le Maire, G., Fayad, I., Bailly, J.S., Nouvellon, Y., Lemos, C., Hakamada, R., 2014. Testing different methods of forest height and aboveground biomass estimations from icesat/glas data in eucalyptus plantations in brazil. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 290–299.

Binkley, D., Campoe, O.C., Alvares, C.A., Carneiro, R.L., Stape, J.L., 2020. Variation in whole-rotation yield among eucalyptus genotypes in response to water and heat stresses: The techs project. Forest Ecol. Manage. 462, 117953. https://doi.org/10.1016/j.foreco.2020.117953. URL: https://www.sciencedirect.com/science/article/pii/S0378112719324843.

Boudreau, J., Nelson, R., Margolis, H., Beaudoin, A., Guindon, L., Kimes, D., 2008. Regional aboveground forest biomass using airborne and spaceborne lidar in québec. Remote Sens. Environ. 112, 3876–3890. https://doi.org/10.1016/j.rse.2008.06.003.

Chauve, A., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Pierrot Deseilligny, M., Puech, W., 2009. Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest. Int. J. Remote Sens. 30, 5211–5228. https://doi.org/10.1080/01431160903023009.

Chen, Q., 2010. Retrieving vegetation height of forests and woodlands over mountainous areas in the pacific coast region using satellite laser altimetry. Remote Sens. Environ. 114, 1610–1627. https://doi.org/10.1016/j.rse.2010.02.016.

Dias, D., da Silva Pinto, A., Dias, U., Lamparelli, R.A.C., le Maire, G., da Silva Torres, R., 2020. A multirepresentational fusion of time series for pixelwise classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 4399–4409.

Dubayah, R., Luthcke, S., 2020a. Gedi l1b geolocated waveform data global footprint level v001.

Dubayah, R., Luthcke, S., 2020b. Gedi l2a elevation and height metrics data global footprint level v001.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., et al., 2020. The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. Sci. Remote Sens. 1, 100002.

El Hajj, M., Baghdadi, N., Fayad, I., Vieilledent, G., Bailly, J.-S., Minh, D.H.T., 2017. Integration of spaceborne lidar data to improve the forest biomass map in madagascar. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 5786–5789. https://doi.org/10.1109/IGARSS.2017.8128323. URL: https://ieeexplore.ieee.org/document/8128323/.

Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P., 2019. Deep learning for time series classification: a review. Data Min. Knowl. Discov. 33, 917–963.

Fayad, I., Baghdadi, N., Bailly, J.-S., Barbier, N., Gond, V., Hajj, M., Fabre, F., Bourgine, B., 2014. Canopy height estimation in french guiana with lidar icesat/glas data using principal component analysis and random forest regressions. Remote Sens. 6, 11883–11914. https://doi.org/10.3390/rs61211883.

Fayad, I., Baghdadi, N., Guitet, S., Bailly, J.-S., Hérault, B., Gond, V., El Hajj, M., Tong Minh, D.H., 2016. Aboveground biomass mapping in french guiana by combining remote sensing, forest inventories and environmental data. Int. J. Appl. Earth Obs. Geoinform. 52, 502–514. https://doi.org/10.1016/j.jag.2016.07.015.

Fayad, I., Baghdadi, N.N., Alvares, C.A., Stape, J.L., Bailly, J.S., Scolforo, H.F., Zribi, M., Maire, G.L., 2021. Assessment of gedi's lidar data for the estimation of canopy heights and wood volume of eucalyptus plantations in brazil. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7095–7110. https://doi.org/10.1109/JSTARS.2021.3092836.

Fayad, I., Baghdadi, N., Riedi, J., 2021. Quality assessment of acquired gedi waveforms: Case study over france, tunisia and french guiana. Remote Sen. 13 https://doi.org/10.3390/rs13163144. URL: https://www.mdpi.com/2072-4292/13/16/3144.

Fieber, K.D., Davenport, I.J., Ferryman, J.M., Gurney, R.J., Walker, J.P., Hacker, J.M., 2013. Analysis of full-waveform lidar data for classification of an orange orchard scene. ISPRS J. Photogramm. Remote Sens. 82, 63–82. https://doi.org/10.1016/j.isprsjprs.2013.05.002.

Hancock, S., Armston, J., Li, Z., Gaulton, R., Lewis, P., Disney, M., Mark Danson, F., Strahler, A., Schaaf, C., Anderson, K., et al., 2015. Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy. Remote Sens. Environ. 164, 208–224. https://doi.org/10.1016/j.rse.2015.04.013.

Harding, D.J., 2005. Icesat waveform measurements of within-footprint topographic relief and vegetation vertical structure. Geophys. Res. Lett. 32, L21S10. https://doi.org/10.1029/2005GL023471.

Hosseini, M., McNairn, H., Mitchell, S.W., Robertson, L.D., Davidson, A.A., Homayouni, S., 2019. Synthetic aperture radar and optical satellite data for estimating the biomass of corn. Int. J. Appl. Earth Obs. Geoinform. 83.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. in: CVPR, pp. 2261–2269.

Ienco, D., Interdonato, R., Gaetano, R., Minh, D.H.T., 2019. 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. 158, 11–22.

Ienco, D., Gbodjo, Y.J.E., Gaetano, R., Interdonato, R., 2020. Weakly supervised learning for land cover mapping of satellite image time series via attention-based CNN. IEEE Access 8, 179547–179560.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML, volume 37, pp. 448–456.

Kim, N., Ha, K., Park, N., Cho, J., Hong, S., Lee, Y., 2019. A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern united states, 2006-2015. ISPRS Int. J. Geo Inf. 8, 240.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.

Lathuilière, S., Mesejo, P., Alameda-Pineda, X., Horaud, R., 2020. A comprehensive analysis of deep regression. IEEE Trans. Pattern Anal. Mach. Intell. 42, 2065–2081.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. BioScience 52, 19 doi: 10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., de Oliveira, R., 2005. Estimates of forest canopy height and aboveground biomass using icesat: Icesat estimates of canopy height. Geophys. Res. Lett. 32 https://doi.org/10.1029/2005GL023971 n/a-n/a.

Lefsky, M.A., Keller, M., Pang, Y., De Camargo, P.B., Hunter, M.O., 2007. Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms. J. Appl. Remote Sens. 1, 013537.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.

Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P., 2020. Designing network design spaces. CoRR abs/14.

Rajab Pourrahmati, M., Baghdadi, N., Darvishsefat, A.A., Namiranian, M., Gond, V., Bailly, J.-S., Zargham, N., 2018. Mapping lorey's height over hyrcanian forests of iran using synergy of icesat/glas and optical images. Eur. J. Remote Sens. 51, 100–115. https://doi.org/10.1080/22797254.2017.1405717.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., 2019. Prabhat, Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204.

Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D., DiMarzio, J.P., 2005. Overview of the icesat mission. Geophys. Res. Lett. 32, L21S01. https://doi.org/10.1029/2005GL024009.

Scolforo, H.F., McTague, J.P., Burkhart, H., Roise, J., Carneiro, R.L., Stape, J.L., 2019. Generalized stem taper and tree volume equations applied to eucalyptus of varying genetics in brazil. Can. J. Forest Res. 49, 447–462. https://doi.org/10.1139/cjfr-2018-0276.

Simard, M., Pinto, N., Fisher, J.B., Baccini, A., 2011. Mapping forest canopy height globally with spaceborne lidar. J. Geophys. Res.: Biogeosci. 116.

Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.A., 2015. Striving for simplicity: The all convolutional net. in: ICLR.

Sumnall, M.J., Hill, R.A., Hinsley, S.A., 2016. Comparison of small-footprint discrete return and full waveform airborne lidar data for estimating multiple forest variables. Remote Sens. Environ. 173, 214–223. https://doi.org/10.1016/j.rse.2015.07.027.

Sun, G., Ranson, K., Kimes, D., Blair, J., Kovacs, K., 2008. Forest vertical structure from glas: An evaluation using lvis and srtm data. Remote Sens. Environ. 112, 107–117. https://doi.org/10.1016/j.rse.2006.09.036.

van der Voet, H., 1994. Comparing the predictive accuracy of models using a simple randomization test. Chemom. Intell. Lab. Syst. 25, 313–323. https://doi.org/10.1016/0169-7439(94)85050-X.

Wagner, W., Ullrich, A., Melzer, T., Briese, C., Kraus, K., 2004. From single-pulse to full-waveform airborne laser scanners: potential and practical challenges volume 35, na.

Xu, H., Li, J., Yuan, H., Liu, Q., Fan, S., Li, T., Sun, X., 2020. Human activity recognition based on gramian angular field and deep convolutional neural network. IEEE Access 8, 199393–199405.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716.

Zhang, G., Si, Y., Wang, D., Yang, W., Sun, Y., 2019. Automated detection of myocardial infarction using a gramian angular field and principal component analysis network. IEEE Access 7, 171570–171583.

Zhu, X., Tuia, D., Mou, L., Zhang, G.X.L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.