A country-level deep-learning approach for canopy height estimation from TanDEM-X InSAR data

Daniel Carcereri[a,b], Paola Rizzoli[a], Dino Ienco[c], and Lorenzo Bruzzone[b]
[a]Microwaves and Radar Institute, German Aerospace Center (DLR), 82234 Wessling, Germany
[b]Universitá degli studi di Trento (UNITN), 38122 Trento, Italy
[c]National Research Institute for Agriculture, Food and the Environment (INRAE), 34000 Montpellier, France

Abstract

The estimation of forest parameters, such as canopy height and above-ground biomass (AGB), at large scales is of paramount importance for forest disturbance analysis, carbon-cycle modelling, wild fire propagation simulations and resource inventorying. In this work, we propose a fully convolutional deep learning architecture, trained with a dedicated data set of TanDEM-X features, to generate wall-to-wall forest parameters products. We investigate the challenges imposed by the proposed deep learning approach for large-scale applications, concentrating in particular on the design of an effective training dataset on the basis of theoretical requirements on single-pass InSAR theory. We test the regression performance of our approach over the five tropical regions mapped by the ESA/NASA AfriSAR campaign in Gabon, Africa. The obtained CHM estimation accuracies are extremely competitive with those of the state-of-the-art methods, with the advantage to be achieved with only a single input TanDEM-X bistatic pair.

1 Introduction

Forests play a key role in the biosphere, covering approximately 31% of the Earth's global surface [1]. For example, forests majorly affect the local water-cycle, act as carbon sinks and limit water runoff during extended rainfalls [2], [3], [4]. Moreover, an estimated 80% of all known terrestrial flora and fauna lives in forests, making them critical for biodiversity preservation [5]. About 880 million people have their livelihoods depending on forests [1].

The assessment and monitoring of forest parameters becomes therefore of paramount importance for assessing their impact on the global environment and for supporting effective decision-making processes. The most accurate way to quantify forest tree parameters is to measure them manually on-site [6]. The process can be either invasive, when individual trees are cut down to precisely assess their properties, or non-invasive, when a quicker and less intrusive approach is preferred over absolute accuracy [7]. In general, the process is very expensive and time consuming, especially in remote areas. Therefore in-situ measurement campaigns are typically feasible only for small surveys [8].

As a consequence of these downsides, remote sensing (RS) approaches to forest modelling have gained wide interest in the last decades, as a large pool of sensors and techniques are available for direct and indirect parameter estimation [9][10]. In particular, Airborne LiDAR systems represent an accurate alternative to manual measurements, especially for the estimation of the canopy height model of a forest, its vertical structure, and other density properties, as these can all be directly measured [11]. Unfortunately, airborne systems remain impractical for large-scale surveys, as a trade-off between temporal and spatial coverage must be struck. Alternatively, spaceborne earth observation (EO) systems can be used, as these typically offer global coverages and short revisit times, combined with a large availability of products [12][13]. Spaceborne LiDAR systems, such as NASA's GEDI instrument [14], enable measurements on a global scale, but acquire over sparse sampling grids and still have extremely low revisit times.

Modern optical and synthetic aperture radar (SAR) spaceborne systems overcome these limitations by offering global coverage and revisit times in the order of a few days [13]. With these systems the challenge consists in modelling the relationship between the acquired imagery and the on-ground forest parameters, as these cannot be directly measured [15]. The definition of allometric equations is the most commonly used approach to indirectly relate forest parameters with RS data [6]. It typically consists in using precise LiDAR or in-situ measurement data to fit parametric regression models with the available datasets, thus resulting in parameters tuned to the specific conditions and geographic locations that have been chosen for calibration [16]. The relationship between data and forest parameters is typically highly non-linear. For example, when considering radar sensors, the measured back-propagated energy is the result of multiple complex interactions between the incident electromagnetic waves and the geometric and dielectric properties of the target on ground [17][18]. Some of these non-linearities are also introduced by the well documented feature saturation problems [16][10], consisting of regions in the feature space in which very small changes in the backscatter are associated to large changes in the parameters to be estimated. Parametric equations typically struggle to maintain sensitivity over these regions, compounding the overall estimation error. Furthermore, aver-

aging over larger resolution cells is often used to improve the overall prediction uncertainty, resulting in a geometric resolution which tends to be significantly worse than the one obtained by airborne LiDAR or photogrammetric systems.

More sophisticated model-based approaches have also been extensively studied for the regression of forest parameters. These tend to achieve higher precision in their estimates, while maintaining better resolution. To obtain these results, typically large amounts of acquisitions [9], ground reference samples [15] or strong assumptions are required to initialize the model. This represents a limit when frequent updates become necessary, especially in the context of national- or continent-wide forest disturbance analysis.

With the recent advancements in machine learning and computer vision techniques, and the availability of large dataset collections from EO sensors, new approaches to forest parameter regression have started to be considered. Of particular interest are the deep learning-based ones, which are constantly gaining more attention in the field of RS image analysis [19]. Much of the attention has gone into convolutional neural network-based deep learning architectures, as these analyze the spatial context information to generate higher level abstractions. These features typically possess a larger descriptive and discriminative content than both the simple input imagery and specific handcrafted features [20].

In the field of computer vision, deep regression techniques have already been used with great success, with even general-purpose architectures achieving results close to the state-of-the-art [21]. On the contrary, comparatively little work still exists regarding the regression of physical and biophysical parameters from RS data [19][22], presumably due to the limited availability of large quantities of reference-data required for supervised training.

Aiming at providing large-scale, frequently updated forest parameter metrics, our research effort focuses on overcoming the aforementioned limitations by proposing a CNN-based regression framework, which requires a single TanDEM-X bistatic product for inference. Building on the knowledge gained from our first ablation study at regional level [23], in this work we present and tackle the issues that emerge when moving from the regional to the country-level scales. We focus our efforts on the estimation of forest height over challenging tropical forests in Gabon, Africa. In particular, we investigate the requirements that should be satisfied by a large-scale application. In this context, we analyze how missing or insufficient contextual information in the training-set affects the prediction performance. Finally we present an end-to-end solution to tackle these problems, that results in the generation of a consistent, time-tagged canopy height map (CHM) over the entire State of Gabon using single bistatic TanDEM-X acquisitions.

The paper is structured as follows. In Section 2 we present the proposed framework and the data-sources chosen to address the research problem. Section 3 gives an overview of the obtained preliminary results (which will be replaced by an extensive analysis in the final version of the paper). Finally, Section 4 draws the conclusions and outlooks from the results of our work.

2 Methodology

In this section we present a forest parameters estimation and analysis approach for the generation of accurate, high resolution, and frequently updatable geomaps at large scale. The proposed approach takes advantage of the knowledge of the physical mechanism of volumetric decorrelation in InSAR products caused by multiple scattering within forests.

2.1 Dataset

Our input measurements consist of single-pass interferometric SAR (InSAR) products. Across-track SAR interferometry requires the acquisition of two SAR images over the same area from two slightly different positions, characterized by a certain distance called baseline B. The phase difference between the received pair of SAR signals allows for the retrieval of the mean topographic location of the scatterers within a resolution cell on ground, according to the sensitivity of the acquisition configuration represented by the height of ambiguity h_{amb}. The latter is defined as the topographic height change corresponding to a complete 2π cycle in the interferogram and depends on geometric acquisition parameters, among which the baseline component perpendicular to the line of sight $B_{\perp}$.

Multiple sources of noise affect the interferogram [24][18], thus the normalized cross-correlation between the InSAR pair, i.e. s_1 and s_2, is used to quantify the quality of the resulting inteferometric phase. This feature is called *interferometric coherence* γ_{Vol} and it is defined as:

$$\gamma_{\mathrm{Tot}} = \frac{\mathbb{E}\left[s_1 s_2^*\right]}{\sqrt{\mathbb{E}\left[\left|s_1\right|^2\right]\mathbb{E}\left[\left|s_2\right|^2\right]}}, \tag{1}$$

where $\mathbb{E}\left[\cdot\right]$ represents the expectation operator and s_2^* indicates the complex conjugate of s_2. Assuming the statistical independence of individual error contributions, the interferometric coherence can be decomposed in the following factors [18]:

$$\gamma_{\mathrm{Tot}} = \gamma_{\mathrm{SNR}} \cdot \gamma_{\mathrm{Quant}} \cdot \gamma_{\mathrm{Amb}} \cdot \gamma_{\mathrm{Rg}} \cdot \gamma_{\mathrm{Az}} \cdot \gamma_{\mathrm{Temp}} \cdot \gamma_{\mathrm{Vol}}, \tag{2}$$

representing the decorrelation effects due to limited signal-to-noise ratio (γ_{SNR}), quantization errors (γ_{Quant}), ambiguities (γ_{Amb}), baseline decorrelation (γ_{Rg}), relative shift of the Doppler spectra (γ_{Az}), temporal difference between the acquisitions (γ_{Temp}) and volumetric scattering effects (γ_{Vol}).

In particular, volume scattering effects arise from the fact that corresponding resolution cells are representative of different sets of vertically distributed scatterers, caused by the slight satellites offset between the acquired SAR image pair. The resulting degree of volume decorrelation depends on the properties of the canopy, the radar frequency, as well as the height of ambiguity: the lower the height of ambiguity, the higher the volume decorrelation.

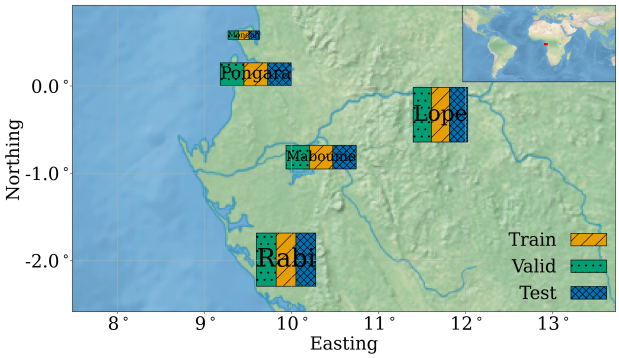

Figure 1 The five study-areas of Lopé, Mabounié, Mondah, Pongara and Rabi. These have been subdivided into three sub-regions, each assigned to either training, validating or testing the proposed DL model.

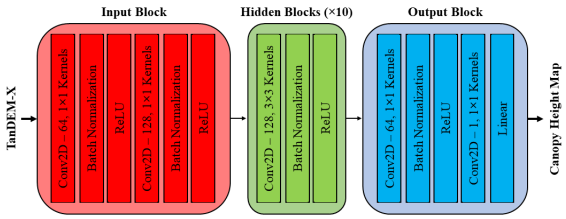

Figure 2 Proposed CNN architecture for forest parameters regression.

As source for our InSAR products we choose the bistatic acquisitions imaged using DLR's TanDEM-X constellation. This is because the temporal decorrelation is intrinsically connected to repeat-pass interferometry applications, meaning that for bistatic interferometry this term can be neglected as both images are acquired simultaneously. The resulting six input feature collection consists of the interferometric coherence γ_{Tot}, the volume decorrelation factor γ_{Vol}, the backscattering coefficient σ^0_{HH} in HH polarization, the acquisition DEM h_{DEM}, the local incidence angle θ_{inc} and the acquisition height of ambiguity h_{amb}.

As reference data for our proposed method, we consider the airborne LiDAR measurements acquired using NASA's LVIS instrument in the context of the NASA/ESA 2016 AfriSAR campaign [11]. These consist of discrete full-waveform measurements, each covering a nominal footprint of 18 m of diameter. For this study, we use the already rasterized products with a ground-sampling distance (GSD) of 25 m, and spanning across the five heterogeneous test sites represented in figure 1.

2.2 Proposed CNN Architecture

The vast majority of the state-of-the-art approaches model the relationship between RS data an forest parameters by introducing simplifying assumptions on the boundary conditions and on the electromagnetic interactions with the vegetation, such as no ground interactions or constant extinction coefficients. While these assumptions might hold for relatively small test areas, when the context is changed the accuracy of the estimates drops. To overcome this limitations we propose a data-driven approach to properly learn the missing contextual information and the underlying regression models from the large pool of available RS data.

The architecture of our model is depicted in figure 2, and it was obtained by means of extensive hyperparameter tuning using the training and the validation sets [23]. The architecture is fully convolutional and includes three types of functional blocks: an input block, which incrementally increases the feature count; a sequence of 10 hidden blocks, which progressively extract higher level abstractions of the input features; and an output block, which reduces the dimensionality of the hidden representation and generates the final prediction. Each hidden block consists of the following three layers:

- A two-dimensional spatial convolutional layer (Conv2D) that uses 128 3×3 pixel kernels.
- A batch normalization layer, which is used to alleviate problems of exploding and vanishing gradients.
- A non-linear activation function layer that consists of a rectified linear unit function (ReLU).

In the input and output blocks the 3×3 kernels have been replaced by 1×1 ones. In the final block, the batch normalization layer and the ReLU activation function have been dropped in favor of a linear operator.

To train and validate our model, we sample 15×15 pixel patches from the respective subsets. The sampling process is done with overlap but without repetition, by extracting a patch centered around each vegetated pixel. We train the model using a mini-batch gradient descent approach to optimize the following two-term loss function:

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 + \lambda \cdot \sum_{j=1}^{m} w_j^2, \qquad (3)$$

where $\hat{y}_i$ is the canopy height values estimate, y_i is its corresponding reference value, w_j are the trainable model weights, and λ weights the impact of the *l2*-normalization term. Finally, we evaluate the model over the independent test set, and measure the estimation performance compared to the expected values by computing the mean error (ME), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the root mean-squared error (RMSE) and the coefficient of determination (R^2).

The most crucial aspect in the design of the training data set is to show the network all possible variations of the input features. For example, given the varying close orbit formation of TanDEM-X, one of the most challenging aspects is how to make the network robust with respect to the possible bistatic acquisition geometries, which are needed for achieving an almost gap-free coverage when inferring at large-scale. To tackle this issue, we use all available TanDEM-X acquisitions acquired between 2010 and 2021 over the AfriSAR test sites to fully cover the entire range of possible heights of ambiguity. This results in roughly 10

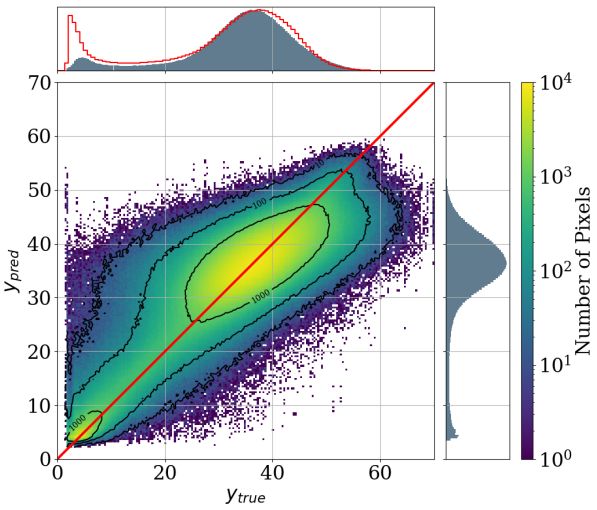

Figure 3 Scatterplot comparison between predicted and reference CHM values in Gabon, Africa.

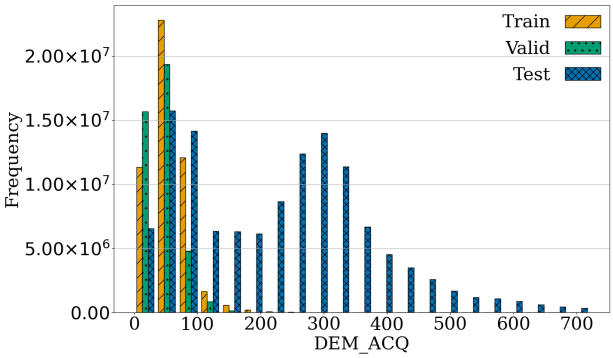

Figure 4 The relative DEM sample distributions across the training, validation, and test sets when Lopé is assigned in its entirety to the test set.

to 20 acquisitions for each of the covered regions of interest. Training, validation and testing of the model are then performed on independent sub-regions, as shown in figure 1.

3 Experimental Results

The set up proposed in section 2 defines a baseline performance, which achieves a bias of -0.40 m, a MAE of 3.83 m, and a R^2 of 0.76. The estimation agreement with the reference values is represented in the scatterplot in figure 3, while the performance for each experiment is reported in table 1.

Test-Sites	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	[·]
Baseline	-0.40	3.83	13.23	5.06	0.76
w/o Lopé	-2.26	5.01	15.97	7.21	0.50
w/o DEM, w/o Lopé	-1.38	4.50	15.75	5.88	0.67
w/o DEM	-0.47	4.25	15.13	5.54	0.71

Table 1 CH estimation: test sites and overall prediction accuracies.

To validate the generalization performance of our proposed framework across unseen regions, we repeat the baseline experiment by removing Lopé from the training and validation steps (**w/o Lopé** case in table 1). The results show that the MAE drastically decreases to 6.40 m, and the R^2 to 0.50. Analyzing the individual contributions of the regions of interest highlights that the results are completely driven by the poor accuracy over the Lopé test area. Further investigations suggest that this reduction in performance might be due to missing representations of the terrain elevations in the training data set. As one can see in figure 4, the test site of Lopé is significantly more mountainous than the remaining four training sites. To test this assumption, we repeat the previous experiments while completely removing the DEM from the set of input features (**w/o DEM, w/o Lopé** case in table 1). From the previous work in [23], we know that topographic information provided by the DEM has a minor impact on overall performance, therefore we expect a slight decrease in accuracy in favor of a potentially more robust and consistent dataset. Indeed, the results confirm this intuition, as training and validating without Lopé, results in a significant performance improvement to an MAE of 4.50 m and an R^2 of 0.67. As one might expect, re-introducing Lopé sees a further increase to a MAE of 4.25 m and a R^2 of 0.71, and represents the pure performance impact of the DEM over the baseline performance (**w/o DEM** case in table 1).

These results confirm a basic notion of data-driven approaches, namely that the application to unseen areas requires a training set that covers the expected range of values of all input features. In this scenario, the the inclusion of the Lopé sub-region into the training pool becomes crucial for the generation of the country-scale canopy height map. Furthermore, we conclude that dropping the DEM feature from the input becomes a necessity, since the presence of taller elevations outside of the range covered by the training set would affect the performance.

Moreover, the developed training strategy assured a reliable performance of the model for all different TanDEM-X geometries. An example is shown in figure 5, which depicts the model performance with respect to the height of ambiguity.

To generate a large-scale CHM of Gabon, we consider all available TanDEM-X data acquired in 2010 and 2011 during the first TanDEM-X global coverage, dropping completely overlapping acquisitions to avoid averaging our estimates. Finally, we individually apply our trained model to each remaining product, and mosaic the estimates without further harmonization steps to generate the large-scale CHM map shown in figure 6. Qualitatively, the map shows no evident strippings due to the different incidence angles and interferometric baselines of the images, suggesting that the model delivers robust estimates with respect to all the different acquisition geometries. Figure 7 depicts the full-

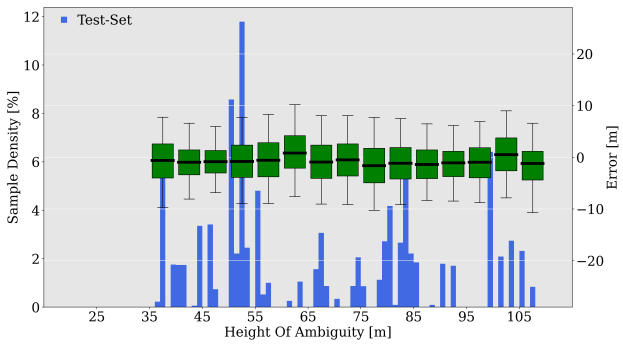

Figure 5 Performance of the proposed model versus the height of ambiguity (black lines identify the median error, green boxes delimit the 25^{th}-75^{th} percentiles and the error boxes the 5^{th}-95^{th} percentiles. The blue histogram shows the height of ambiguity distribution of the test data set).

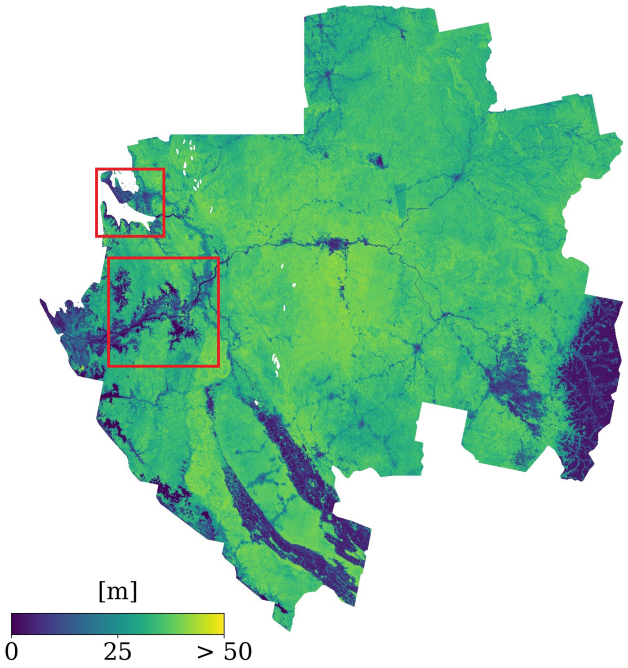

Figure 6 Country-scale CHM estimate for the year 2011 over Gabon. The product is obtained by arbitratily removing overlapping acquisitions, and simply moisacking the estimates obtained from the remaining images. The red squares highlight the zoom-ins shown in figure 7.

resolution details of the areas within the red boxes in figure 6. In figure 7 (a) it is possible to note the presence of tall mangroves along the shores of the Gabon estuary, with peak canopy heights above 45 m. The area in figure 7 (b) is characterized by the presence of primary tropical forest, a complex system of rivers and antropogenic activities.

4 Conclusions and Outlook

In this work, we proposed a deep learning-based approach to the creation of operational predictions of canopy height

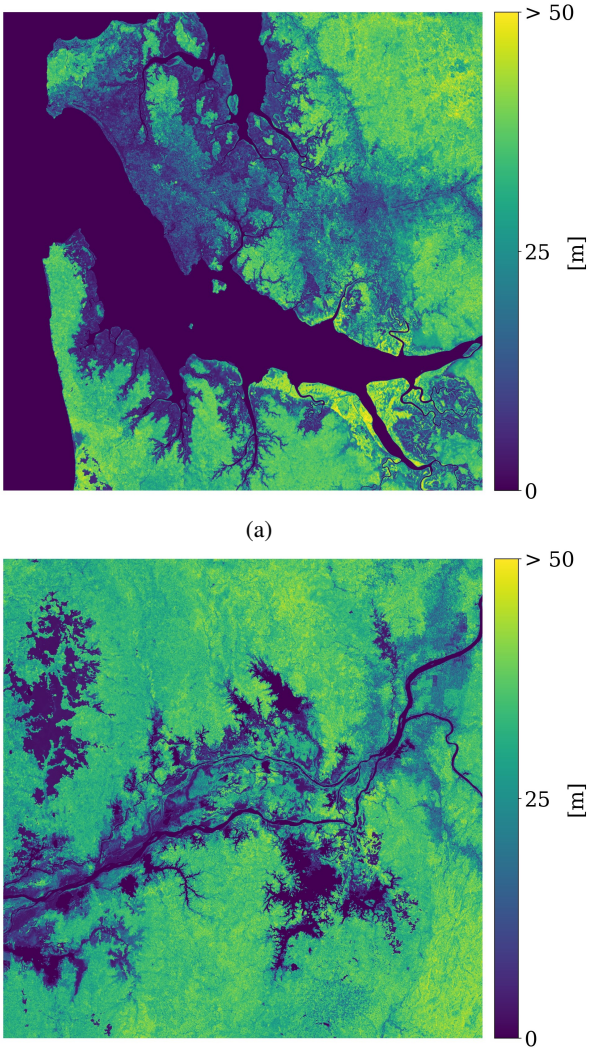

Figure 7 Two full-resolution zoom-in examples from the country-scale mosaic of figure 6. (a) includes the test sites of Mondah and Pongara, while (b) the test site of Mabounié.

models targeting the tropical forests of Gabon. As input to our framework, we used DLR's TanDEM-X bistatic data, exploiting the relationship between the volumetric decorrelation loss and the biophysical properties of the canopy to estimate its height. We presented some of the challenges inherent to data-driven approaches, highlighting how missing value-range representations of the input features in the training-set can negatively affect the prediction accuracy. In particular, we demonstrated that the missing DEM elevation coverage can be successfully overcome by dropping the related feature, without a significant performance loss. Subsequently, we deployed the trained model to generate a country-scale CHM of Gabon at 25 m of resolution, for the year 2011. By arbitrarily discarding overlapping acquisitions, we demonstrated that our approach can achieve robust and consistent wall-to-wall predictions without the need of averaging practices, and despite the challenging variability in acquisition-geometries. Ultimately, the results we obtained in Gabon using TanDEM-X interferometric SAR products are extremely compelling for large-

scale applications of forest biophysical parameters retrieval. While the TanDEM-X mission possesses worse and more inconsistent revisit times than those achieved by modern spaceborne SAR missions, its unique bistatic configuration and its almost 13 year long operational time represent an unique resource for forest disturbance monitoring. In our future works, we plan to extend our approach to larger scales and different types of forest. Furthermore, we aim at exploiting the existing database to monitor changes in time.

5 Literature

[1] Food and Agriculture Organization. *Global Forest Resources Assessment 2020*. FAO, 2020.

[2] Natural Resources Canada. Forest Carbon. `https://www.nrcan.gc.ca/climate-change/impacts-adaptations/climate-change-impacts-forests/forest-carbon/13085`.

[3] Phoebe G. Aron, Christopher J. Poulsen, Richard P. Fiorella, and Ashley M. Matheny. Stable water isotopes reveal effects of intermediate disturbance and canopy structure on forest water cycling. *Journal of Geophysical Research: Biogeosciences*, 124(10):2958–2975, 2019.

[4] David Ellison. Forests and water: background analytical study. Thirteenth session of the united nations forum on forests, United Nations, 2018.

[5] World Wide Fund for Nature. Forest Habitat. `https://www.worldwildlife.org/habitats/forest-habitat`.

[6] Tommaso Jucker et al. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Global Change Biology*, 23(1):177–190, 2017.

[7] Kim Calders et al. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution*, 6(2):198–208, 2015.

[8] Picard, Saint-André, Laurent, Nicolas, Henry, and Matieu. *Manual for building tree volume and biomass allometric equations from filed measurement to prediction*. Food and Agriculture Organization of the United Nations (FA0), 2012. OCLC: 931325352.

[9] Jan Askne, Johan Fransson, Maurizio Santoro, Maciej Soja, and Lars Ulander. Model-Based Biomass Estimation of a Hemi-Boreal Forest from Multitemporal TanDEM-X Acquisitions. *Remote Sensing*, 5(11):5574–5597, October 2013.

[10] Heikki Astola et al. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment*, 223:257–273, 2019.

[11] Temilola Fatoyinbo et al. The NASA AfriSAR campaign: Airborne SAR and lidar measurements of tropical forest structure and biomass in support of current and future space missions. *Remote Sensing of Environment*, 264:112533, 2021.

[12] Pierre Potin, Betlem Rosich, Nuno Miranda, and Patrick Grimont. Sentinel-1 mission status. *Procedia Computer Science*, 100:1297–1304, 2016.

[13] Matthias Drusch et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120:25–36, 2012.

[14] Ralph Dubayah et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. *Science of Remote Sensing*, 1:100002, 2020.

[15] Florian Kugler et al. TanDEM-X Pol-InSAR Performance for Forest Height Estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6404–6422, October 2014.

[16] Catherine Torres de Almeida et al. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the brazilian amazon using different regression algorithms. *Remote Sensing of Environment*, 232:111323, 2019.

[17] Robert N. Treuhaft, Søren N. Madsen, Mahta Moghaddam, and Jakob J. van Zyl. Vegetation characteristics and underlying topography from interferometric radar. *Radio Science*, 31(6):1449–1485, 1996.

[18] Michele Martone, Paola Rizzoli, and Gerhard Krieger. Volume decorrelation effects in TanDEM-x interferometric SAR data. *IEEE Geoscience and Remote Sensing Letters*, 13(12):1812–1816, 2016.

[19] Nico Lang, Konrad Schindler, and Jan Dirk Wegner. Country-wide high-resolution vegetation height mapping with sentinel-2. *Remote Sensing of Environment*, 233:111347, 2019.

[20] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657, 2017.

[21] Stéphane Lathuilière, Pablo Mesejo, Xavier Alameda-Pineda, and Radu Horaud. A comprehensive analysis of deep regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9):2065–2081, 2020.

[22] Xiao Wang and Haipeng Wang. Forest height mapping using complex-valued convolutional neural network. *IEEE Access*, 7:126334–126343, 2019.

[23] Daniel Carcereri, Paola Rizzoli, Dino Ienco, and Lorenzo Bruzzone. A Deep Learning Framework for the Estimation of Forest Height From Bistatic TanDEM-X Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:8334–8352, 2023.

[24] H.A. Zebker and J. Villasenor. Decorrelation in interferometric radar echoes. *IEEE Transactions on Geoscience and Remote Sensing*, 30(5):950–959, 1992.