A Deep Learning Framework for the Estimation of Forest Height From Bistatic TanDEM-X Data

Daniel Carcereri, Paola Rizzoli, Dino Ienco, *Member, IEEE*, and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—Up-to-date canopy height model (CHM) estimates are of key importance for forest resource monitoring and disturbance analysis. In this article, we present a study on the potential of deep learning (DL) for the regression of forest height from TanDEM-X bistatic interferometric synthetic aperture radar (InSAR) data. We propose a novel fully convolutional neural network framework, trained in a supervised manner using reference CHM measurements derived from the LiDAR LVIS airborne sensor from NASA. The reference measurements were acquired during the joint NASA–ESA 2016 AfriSAR campaign over five sites in Gabon, Africa, characterized by the presence of different kinds of vegetation, spanning from tropical primary forests to mangroves. Together with the DL architecture and training strategy, we present a series of experiments to assess the impact of different input features on the network estimation accuracy (in particular of bistatic InSAR-related ones). When tested on all the considered sites, the proposed DL model achieves an overall performance of 1.46-m mean error, 4.2-m mean absolute error, and 15.06% mean absolute percentage error. Furthermore, we perform a spatial transfer analysis aimed at deriving preliminary insights on the generalization capability of the network when trained and tested on datasets acquired over different locations, combining different kinds of tropical vegetation. The obtained results are promising and already in line with state-of-the-art methods based on both physical-based modeling and data-driven approaches, with the remarkable advantage of requiring only one single TanDEM-X acquisition at inference time.**

***Index Terms*—Bistatic coherence, convolutional neural network (CNN), deep learning, forest height, synthetic aperture radar (SAR), synthetic aperture radar interferometry, TanDEM-X.**

I. Introduction

FORESTS are one of the most relevant ecosystems on the planet. They cover about 31% of the total Earth surface [1], impacting a variety of biophysical processes, such as the carbon and water cycles, as well as weather and local climate [2], [3], [4]. During their natural growth process, plants extract carbon atoms from the atmosphere, in the form of carbon dioxide, to combine them with water molecules and create carbohydrates. Some of these synthesized compounds are stored by plants themselves, resulting in a net buildup of new biomass [5]. About 30% of total vegetated areas are primarily used to gather forest products, with an additional 18% assigned for multiple use, including, e.g., the production of food. Furthermore, roughly 880 million people depend on forests for fuel production from wood, with an estimated 90% of people living in extreme poverty having their livelihoods crucially depending on forests [1]. Human logging activities and forest degradation affect an estimated 10 million ha on a yearly basis [1]. Moreover, natural hazards, such as wildfires and severe weather events, can also impact forest resources as they relate to propagation patterns and damage evaluations [6], [7].

All these aspects highlight the need for forest disturbance analysis, that is, to monitor changes in forests over time, to characterize their causes, and to quantify their impact. To properly assess the state of a forest, different parameters, such as forest cover, canopy height (CH) and above ground biomass (AGB), are typically used. The most accurate way to retrieve these properties is to manually perform in situ measurements [8], [9]. Especially for remote areas, the process is highly expensive, time consuming, and ultimately unfeasible for large-scale mapping or recurrent assessments [10].

As a way to overcome these constraints, remote sensing (RS)-based approaches for forest parameters estimation have gained wide attention, as an ever-growing variety of earth observation (EO) sensors and techniques has been developed and made available throughout the years [11], [12], [13], [14]. In this context, Light Detection and Ranging (LiDAR) systems represent the most straightforward alternative to replace direct approaches, as the height of the canopies can be directly inferred from the time-of-flight of the laser signal returns. Airborne and spaceborne laser scanning systems, such as NASA's Land, Vegetation, and Ice Sensor (LVIS) [15] and the Global Ecosystem Dynamics Investigation (GEDI) mission [16], are particularly attractive, as the current technology is capable of achieving high sampling rates for medium-to-high-resolution products.

Differently, modern spaceborne optical and synthetic aperture radar (SAR) imaging systems offer global continuous coverage and revisit times in the order of a few days [17], [18], overcoming the coverage limitations of LiDAR systems. An effective exploitation of these products poses its own challenges, since forest parameters cannot be directly estimated from the observed quantities. Instead, they require physical model-based

Manuscript received 24 February 2023; revised 26 May 2023 and 13 July 2023; accepted 19 August 2023. Date of publication 30 August 2023; date of current version 18 September 2023. *(Corresponding author: Daniel Carcereri.)*

Daniel Carcereri is with the Microwaves and Radar Institute, German Aerospace Center, 82234 Weßling, Germany, and also with the Department of Information Engineering and Computer Science, University of Trento, 38123 Povo, Italy (e-mail: daniel.carcereri@dlr.de).

Paola Rizzoli is with the Microwaves and Radar Institute, German Aerospace Center, 82234 Weßling, Germany (e-mail: paola.rizzoli@dlr.de).

Dino Ienco is with the INRAE, 34093 Montpellier, France (e-mail: dino.ienco@inrae.fr).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Povo, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

Digital Object Identifier 10.1109/JSTARS.2023.3310209

or data-driven approaches to model the relationship between the imagery and the on-ground forest properties [11], [19].

Particularly, allometric equations are an extensively used approach for the indirect estimation of forest parameters from RS data [20]. The result is the generation of theoretical models, which are normally tuned to the specific conditions and geographic locations that have been chosen for their calibration [13].

More sophisticated approaches to parameters regression consist in physically modeling the interaction of the received signal with the illuminated vegetation. For example, regarding radar sensors, the propagation of the electromagnetic signal through the canopy structure can be theoretically modeled, and the resulting interaction can be expressed in terms of different properties of the forest, such as its density, height, and composition. The extraction of these specific parameters is, thus, performed through the inversion of such models, to be estimated using a sufficient number of observations.

A well-known example of a physical-based model based on SAR data is the random volume over ground (RVoG), first proposed to estimate forest height from single baseline fully polarimetric (full-pol) interferometric SAR (PolInSAR) observations [21], [22]. Despite the relative simplicity of the model, it can achieve a good prediction accuracy using few observations or *a priori* information, establishing the RVoG as one of the main references in the literature. It has been demonstrated that the limited availability of full-pol PolInSAR acquisitions for the inversion of the RVoG model could be overcome by utilizing a larger stack of dual-pol or single-pol interferometric SAR (InSAR) acquisitions instead. For example, when estimating tree heights using X-band SAR (and in particular TanDEM-X, which is the focus of the present study), Kugler et al. [19] found good agreement between their predictions and the reference CH measurements over boreal (they reported $r^2 = 0.86$, where r represents the Pearson correlation coefficient) and temperate ($r^2 = 0.77$) forests. On the other hand, over a tropical site, the denser vegetation was found to significantly reduce the overall performance ($r^2 \simeq 0.50$), mostly due to the limited penetration of radar waves at X-band into dense vegetation. Askne et al. [11] applied an interferometric water cloud model to a TanDEM-X bistatic time series of 18 InSAR acquisitions for the estimation of AGB. Their findings showed that a considerable estimation accuracy could be achieved (root-mean-square error (RMSE) of 16% with coefficient of determination $R^2 = 0.93$ for forest stands larger than 1 ha), but required an ancillary LiDAR-based digital terrain model (DTM) to improve the absolute calibration of the procedure. Furthermore, the authors acknowledged that the use of *a priori* information about the study site poses a limit to the application of the technique over unmanaged or remote forests. Guliaev et al. [23] proposed to reduce the TanDEM-X PolInSAR requirements for RVoG model inversion to one single polarization, by leveraging the availability of sparse full-waveform LiDAR measurements to approximate the vertical radar reflectivity function. A CH estimation accuracy was evaluated over the tropical forest area of the Lopé National Park (Gabon), resulting in a peak performance of 8.62-m RMSE ($r^2 = 0.40$). Choi et al. [24] presented a similar scheme for mapping forest height from TanDEM-X single-polarization data by combining interferometric coherence maps and GEDI waveform measurements over Tasmania. They validated the technique against airborne LiDAR reference data, achieving an RMSE of 7.3 m ($r = 0.66$).

Recently, machine learning (ML) approaches have constantly been gaining attention, as a further valid alternative to allometric and physical-based models. The increasing availability of frequently updated global and high-resolution EO dataset collections has made data-driven techniques particularly attractive. In the context of RS applications, ML techniques have already become a staple in the solution of binary or multiclass classification problems [25]. However, for regression tasks, such as forest parameter estimation, their usage is still limited. In particular, deep neural networks have gathered much of the attention when solving EO-related problems. Here, the information is iteratively processed and extracted to generate feature maps of higher levels of abstraction and descriptive power than those of both input data or hand-crafted features. In this context, convolutional neural networks (CNNs) have found widespread adoption and success when dealing with imagery type products, given their capability to extract information from 2-D spatial patterns [26]. Deep learning (DL) techniques have already been applied with great success to the solution of regression problems in the field of computer vision [27]. However, they have been less commonly used in the regression of biophysical parameters from RS data, as the demand for large quantities of reliable labeled reference data has slowed down their adoption. Lang et al. [28] used a deep CNN model to estimate canopy height model (CHM) at a ground sampling distance (GSD) of 10 m from Sentinel-2 multispectral data over Gabon and Switzerland, demonstrating that spatial contextual information is crucial to obtain accurate estimates. When considering only the prediction from the acquisitions with the lowest cloud coverage probability, the authors achieved a mean absolute error (MAE) of about 2 m in Switzerland and of 4.9 m in Gabon, with an RMSE of 3.9 and 6.5 m, respectively. Aggregating the estimates using the median over all available acquisition dates within a year reduced the MAE to 1.7 and 4.3 m and the RMSE to 3.4 and 5.6 m, respectively. Becker et al. [29] proposed a Bayesian DL approach to estimate forest parameters at a GSD of 10 m, using SAR Sentinel-1 and multispectral Sentinel-2 data across Norway. Their approach allows for the estimation of forest structure variables together with their uncertainty estimates, providing intrinsic assessments on the trustworthiness of their predictions. The estimation of the CH, here quantified as the height corresponding to the 95th percentile of the returned reference LiDAR energy, resulted in an MAE of 1.65 m and an RMSE of 2.30 m. Their analyses showed that among the two considered sensors, Sentinel-2 was the more informative source, achieving an MAE of 1.81 m, compared to 3.05 m of Sentinel-1.

In this context, the application of DL models to SAR, and in particular to InSAR data, still remains at its early stages and needs to be further investigated and understood. Moreover, to the best of our knowledge, no specific work on the use of bistatic InSAR in combination with DL for the retrieval of biophysical parameters has appeared in the literature yet, and a deeper understanding of its potential represents a crucial aspect

not only for a better exploitation of on-going missions, such as TanDEM-X, but also in view of future bistatic or multistatic SAR missions, such as the ESA Harmony Earth Explorer 10 mission proposal [30].

In this study, we propose a novel DL framework for the prediction of forest heights from TanDEM-X bistatic InSAR data. We present an original architecture based on CNNs, together with its training strategy and performance evaluation approach. Initially, we consider a best case scenario, where all possible informative features are used as input to the network, to explore the potential of DL to model highly nonlinear relationships for achieving the best possible performance with respect to state-of-the-art methods. Our goal is to understand if DL is able to properly model the relationship linking SAR-/InSAR-/geometry-related features to biophysical parameters (CH), mitigating the well-known saturation effects caused by the limited penetration capabilities of X-band into the canopy. We train and validate our DL framework over the challenging scenario provided by tropical forests in Gabon, West Africa, characterized by the presence of dense vegetation. Afterward, we concentrate on a feature importance analysis, aiming at understanding the role that different features play in achieving the final performance. We mainly focus on bistatic InSAR features, which represent the added value of TanDEM-X with respect to all other on-going spaceborne SAR missions. In this context, we also aim at verifying the capability of the proposed DL framework to automatically manage the use of different bistatic acquisition geometries, which directly impact the bistatic InSAR features (as explained later in Section II). This would be an important property, making it possible to obtain updated predictions at each TanDEM-X revisit, thus enabling the effective monitoring of forest disturbances. In addition, we present a high-level comparison with the RVoG physical model under the same dataset constraints. Finally, we take advantage of the data-driven nature of our approach to perform a preliminary analysis on the generalization capability of the proposed DL framework in the spatial domain, aiming at deriving first insights for a future extension to large-scale processing.

The rest of this article is structured as follows. In Section II, we introduce the TanDEM-X mission as well as the theoretical background for bistatic InSAR products. In Section III, we describe the utilized datasets and the preprocessing steps that we applied for our experiments. Section IV introduces the concept of CNN and explains the proposed architecture and training strategies as well as the performance evaluation metrics. In Section V, we present a series of experiments designed to assess the impact of each input feature on the overall prediction performance. Section VI introduces a cross-validation scenario, exploring the geographical generalization capabilities of our framework. In Section VII, we discuss our results. Finally, Section VIII concludes this article and provides outlook related to our work.

II. TanDEM-X and the Bistatic Coherence

The German TanDEM-X mission is currently the only spaceborne SAR mission comprising two separate twin spacecraft, namely TerraSAR-X and TanDEM-X, which fly in a synchronized close-orbit formation. It operationally provides single-pass InSAR acquisitions at X-band with variable acquisition geometries and polarizations [31]. The main objective of the mission has been the generation of a global digital elevation model (DEM), which was successfully completed in 2016 [32]. Aimed at generating an updated version of the TanDEM-X Global DEM product, an additional global dataset of bistatic InSAR data has recently been acquired, and a new DEM product is scheduled to be released in the near future [33]. Moreover, several specific mission phases have provided the scientific community with an unprecedented variety of test cases for the development of novel algorithms and applications [34].

Throughout the entire mission duration, the bistatic interferometric coherence has represented the key parameter for monitoring the global interferometric performance, as shown in [32], [35], and [36]. The coherence γ_{tot} is defined as the normalized cross-correlation coefficient between the interferometric image pair, composed by a master (u_1) and slave (u_2) images

$$\gamma_{\text{Tot}} = \frac{|E[u_1 \cdot u_2^*]|}{\sqrt{E[|u_1|^2] \cdot E[|u_2|^2]}} \tag{1}$$

where $E[\cdot]$ represents the statistical expectation, $*$ the complex conjugate operator, and $|\cdot|$ the absolute value. High values of γ_{tot} are associated with low levels of noise in the interferogram and *vice versa*. γ_{tot} is commonly estimated from InSAR data by applying the maximum likelihood estimator originally introduced in [37], which requires the application of a moving boxcar window. More recently, advanced algorithms have been proposed, based on the nonlocal paradigm, which allows for a significant denoising of the InSAR signal while preserving a high spatial resolution [38], [39].

Following the factorization presented in [31] and [40], one can express γ_{tot} as

$$\gamma_{\text{tot}} = \gamma_{\text{rg}} \cdot \gamma_{\text{amb}} \cdot \gamma_{\text{az}} \cdot \gamma_{\text{quant}} \cdot \gamma_{\text{SNR}} \cdot \gamma_{\text{temp}} \cdot \gamma_{\text{vol}} \tag{2}$$

where the terms on the right-hand side are hereby called *decorrelation factors* and account for different error contributions. The first five terms are related to the system properties and quantify the decorrelation caused by coregistration errors in range and baseline estimation (γ_{rg}), SAR ambiguities (γ_{amb}), misregistration in azimuth and relative shift of the Doppler spectra (γ_{az}), quantization (γ_{quant}), and thermal noise (γ_{SNR}). γ_{temp} quantifies the decorrelation caused by changes on ground occurring between the acquisition of the master and slave images. The uniqueness of TanDEM-X data resides in the bistatic nature of the system, which allows for the simultaneous acquisition of single-pass InSAR data unaffected by temporal decorrelation, which leads to $\gamma_{\text{temp}} = 1$.

The last term γ_{vol} is called *volume decorrelation* factor, and it can be derived from the interferometric coherence by inverting (2) and compensating for all other decorrelation sources, as presented in [41] for nominal TanDEM-X bistatic acquisitions. It quantifies the amount of decorrelation caused by scattering from a volumetric target. This effect occurs, e.g., when radar waves penetrate into forest canopies or ice- and snow-covered regions.

It depends on several factors, such as the radar frequency, the acquisition geometry and the intrinsic properties of the illuminated target. In particular, the lower the operational sensor frequency, the deeper radar waves can penetrate into volumetric targets. The volume decorrelation factor can, thus, be modeled as [21], [42]

$$\gamma_{\text{vol}} = \frac{\int_0^{h_v} F(z) \cdot \exp\left(j \frac{2\pi}{h_{\text{amb}}} z\right) \cdot dz}{\int_0^{h_v} F(z) \cdot dz} \tag{3}$$

where h_v denotes the vegetation height and $F(z)$ represents the vertical backscattering profile, typically modeled as an exponential decay factor that depends on the so-called one-way extinction coefficient (i.e., the coefficient of attenuation), the vegetation height, and the incidence angle.

Regarding the dependence on the bistatic InSAR acquisition geometry, γ_{vol} is closely related to the height of ambiguity h_{amb} [42], which is defined as the topographic height corresponding to a complete 2π cycle in the interferogram and, for the single-pass case, can be expressed as

$$h_{\text{amb}} = \frac{\lambda \cdot r \cdot \sin\theta}{B_\perp} \tag{4}$$

where $B_\perp$ is the orthogonal baseline, λ is the wavelength, r is the slant range, and θ is the incidence angle. Finally, γ_{vol} is also linked to the intrinsic properties of the illuminated target. For example, when considering vegetated areas, the amount of volume decorrelation is impacted by vegetation characteristics, such as tree height and density, presence of gaps, and tree species.

Many scientific works in the field of biosphere-related applications have been developed exploiting the availability of the TanDEM-X bistatic coherence information. For example, it is widely used for model-based tree height [19], [23], [43], [44], [45], [46], [47], [48] and above-ground biomass estimation [11], [49], [50], [51], [52]. It represents a useful input feature for the retrieval of vegetation height in agricultural areas [53], [54] as well as for crop-type mapping [55]. Moreover, the added value of the bistatic coherence has also been demonstrated for land cover classification and forest mapping purposes [56], [57], [58], [59].

III. Dataset Description and Preprocessing

A. AfriSAR Campaign and the Reference LiDAR Data

To train our model in a supervised fashion, we use ground-truth labels from the AfriSAR campaign [60]. The AfriSAR campaign was a joint NASA–ESA campaign, which took place in 2016 and aimed at collecting a combination of field and airborne LiDAR and radar measurements of tropical forests located in the West African country of Gabon [15]. Related to our work, the mission included a series of full-waveform LiDAR measurements acquired by NASA's airborne-mounted LVIS instrument [61]. This particular acquisition campaign was performed between February and March 2016 and covered the five study sites of Mabouniè, Mondah, Lope, Pongara, and Rabi, which can be seen in Fig. 1.

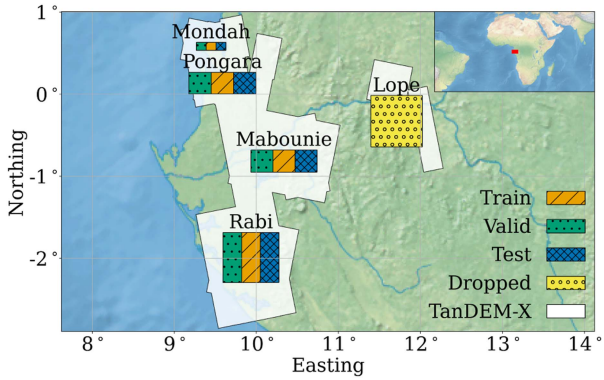

Fig. 1. Study areas in Gabon (Africa) and their subdivision into three subsets for training, validation, and testing of the proposed DL architecture. The TanDEM-X footprints are identified in white.

During the acquisition phase, each area was sampled by a sequence of regularly spaced laser beams, each one resulting in a nominal footprint diameter of 18 m, given the acquisition flight path. It is, thus, possible to use the signal statistics of each individual full-waveform LiDAR return to derive forest parameters, such as the canopy cover, the plant area index, the foliage height diversity, the AGB density, and, of particular interest for our case, the CHM. For our study, we choose to use the available gridded products, which contain completely preprocessed CHM estimates, aggregated and sampled at a GSD of 25 m [60]. As a reference metric, we use the CH value corresponding to the 99th percentile of the total backpropagated laser energy. This statistic takes the name of relative height and is selected in order to improve robustness against noise and outliers.

B. TanDEM-X Bistatic Products and Derived Features

Between 2015 and 2016, a series of dedicated TanDEM-X bistatic acquisitions were commanded over Gabon, in order to cover the same test sites of the AfriSAR campaign. The illuminated footprints mainly overlap with four of the five AfriSAR test sites, as depicted in Fig. 1. Thus, the test site of Lope is not considered for this investigation.

All the acquisitions were commanded in StripMap, single polarization mode (HH channel), extending by about 30 km in range and with different incidence angles and bistatic geometries.

In the present work, we consider as TanDEM-X input data the CoSSC products, corresponding to coregistered single-look complex bistatic SAR data products at full resolution. Focusing and coregistration are performed by the operational TanDEM-X processor [62]. The complete list of the utilized TanDEM-X products and their main acquisition parameters is presented in Table IV in the Appendix.

For each acquisition, we compute the backscattering coefficient σ^0 from the monostatic channel only, as recorded by the transmitting satellite. σ^0 is derived from the absolutely calibrated intensity β^0 (i.e., the radar brightness) and the local

TABLE I
OVERALL PREDICTION PERFORMANCE FOR ALL THE ANALYZED TEST CASE SCENARIOS

Scenario	**Input features**						**ME**	**MAE**	**MAPE**	**RMSE**	**R²**
	σ^0	γ_{tot}	γ_{vol}	θ_{inc}	DEM	h_{amb}	[m]	[m]	[%]	[m]	[·]
Baseline	x	x	x	x	x	x	-1.46	4.20	15.06	5.69	0.73
Baseline w/ DEM mask	x	x	x	x	x	x	-1.38	4.12	14.91	5.41	0.76
w/o DEM	x	x	x	x	-	x	-1.10	4.48	17.66	5.98	0.71
w/o θ_{inc}	x	x	x	-	x	x	-2.48	4.65	15.85	6.30	0.67
w/o DEM, w/o θ_{inc}	x	x	x	-	-	x	-1.39	4.80	17.96	6.34	0.67
w/o σ^0	-	x	x	x	x	x	-1.51	4.37	15.38	5.95	0.71
w/o InSAR features	x	-	-	x	x	-	-2.10	4.88	17.51	6.62	0.64
w/o γ_{tot}	x	-	x	x	x	x	-1.69	4.31	15.32	5.80	0.73
w/o γ_{vol}	x	x	-	x	x	x	-1.91	4.33	15.16	5.84	0.72
w/o h_{amb}	x	x	x	x	x	-	-1.07	4.19	14.98	5.67	0.74
w/o γ_{vol}**, w/o** h_{amb}	x	x	-	x	x	-	-1.13	4.22	15.10	5.75	0.73
w/o γ_{tot}**, w/o** h_{amb}	x	-	x	x	x	-	-2.07	4.59	15.70	6.42	0.66

Number of considered acquisitions = 12; number of pixels = 1 465 947.

incidence angle θ_{inc} as

$$\sigma^0 = \beta^0 \sin(\theta_{\text{inc}}) \tag{5}$$

where θ_{inc} is computed by considering the satellite orbit position and the underlying global TanDEM-X edited DEM product (see Section III-C). For the estimation of the total interferometric coherence γ_{tot}, we apply Φ-Net [63], a novel residual DL architecture for the joint estimation of the InSAR phase and coherence, which has been shown to achieve state-of-the-art denoising performance by preserving the spatial resolution. Moreover, from the coherence, we also derive the volume decorrelation factor γ_{vol} by following the estimation procedure presented in [41]. Finally, a 2-D map of the height of ambiguity h_{amb} is generated by considering the annotated information on the satellites position and by applying (4).

C. Ancillary Data

As additional ancillary data, we make use of the following two products.

1) The *global TanDEM-X edited DEM:* This is an edited version of the original global TanDEM-X DEM product, processed at a lower resolution of 30 m by DLR. The applied editing algorithm is a fully automatic procedure, which takes care of filling voids and flattening water surfaces [64]. No artificial correction was applied over forested areas. Therefore, in the presence of vegetation, the TanDEM-X DEM height does not correspond to the height of the top of the canopy, but, given the radar penetration effects explained in Section II, it is located close to or below the canopy surface, depending on the characteristics of the vegetation [42], [57], [65].
2) The *ESA 2021 WorldCover map:* This is a 10-m-resolution global land cover product that refers to 2021. It was generated using both ESA's Sentinel-1 and Sentinel-2 constellation, and it is freely accessible. The WorldCover map categorizes different land cover types among 11 classes. The product was independently validated with an overall global accuracy of 76.7% [66]. We take advantage of the information it provides to coarsely mask out nonforested areas from our dataset.

D. Common-Grid Interpolation

At the end of the preprocessing steps, all available reference datasets and input feature maps are projected onto a common pixel grid. First, all products are reprojected to the target coordinate reference system by means of bilinear interpolation. Second, the products are resampled to match the reference extent and GSD. Operationally, we align all features onto the grid used by the AfriSAR reference data. For continuous features, this is achieved by means of averaging when downscaling and of bilinear interpolation otherwise. For features with discrete value ranges, the mode of the samples is used instead. After this processing step, all features are aligned at pixel level.

IV. PROPOSED DL FRAMEWORK

We propose the use of a DL framework based on a CNN for the estimation of the CHM. At its core, a CNN consists in the computation of the cross-correlations between a stack of input features with a set of kernel functions. A nonlinear activation function is then applied to the results to generate a new stack of output features. By stacking multiple of these sequences of operations, the complexity of the network can be increased, with each subsequent layer yielding a higher level of representation, which results in progressively more expressive features with respect to the task the network has to deal with. The kernel functions are empirically determined during the training phase of the network, consisting in a forward pass in which the network is used to generate a prediction from a certain input, followed by a backpropagation phase in which the weights of the model are updated according to their impact on the prediction's error. By repeating the process over multiple heterogeneous examples, the network will iteratively converge toward a set of kernel filters well suited for the downstream task.

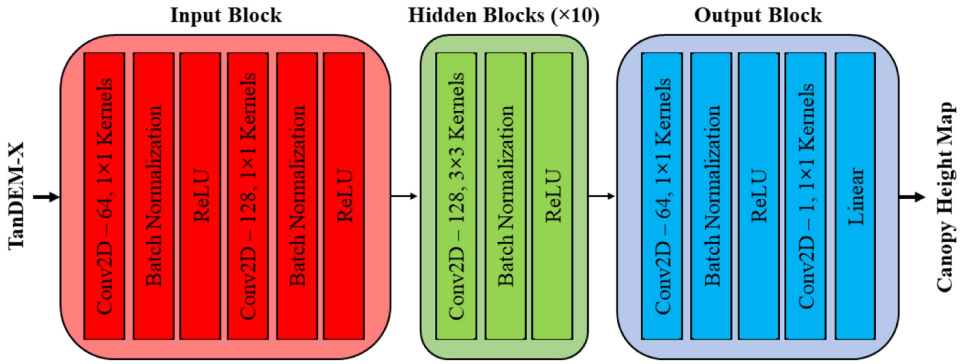

Fig. 2. Proposed CNN architecture for CHM regression. It can be functionally split into one input block, ten hidden blocks, and an output block.

A. CNN Architecture

The DL model that we conceived for this article consists of a fully CNN. This choice was made considering the strong capabilities of such architectures to handle 2-D data, such as SAR images and related features [67]. We converged to the final architecture by means of hyperparameter tuning of different layer typologies, regularization techniques, and model dimensions. This was implemented by evaluating and comparing the potential impact of these choices using the validation set detailed in Section V-A. Notably, we empirically verified that our model was not benefiting from the introduction of skip connections between the blocks. We suppose that, due to its relatively shallow depth, our model is likely to be less affected by problems of vanishing gradients.

The overall architecture of the resulting CNN is shown in Fig. 2. It can be broken down into three types of functional blocks: an input block, a sequence of ten hidden blocks, and an output block. We built our network to process as input SAR and InSAR features computed from a single TanDEM-X acquisition pair. To this end, we begin each block structure with a 2-D convolution operation, which simultaneously applies a 3-D kernel across all input features while sliding along the spatial dimensions of the datacube. In order to preserve the shape of the input, the feature stack is padded before each convolution along its spatial dimensions. In the input and output blocks, two 1×1 convolutional operations are applied in sequence to gradually increase and decrease the number of input features. Each convolutional layer is followed by a batch normalization operation, which shifts and scales the data to be approximately zero mean with unitary variance. This operation was found to lead to better and faster convergence, as it allows for using higher learning rates as well as reducing the sensitivity to the initialization of the model weights [68]. For all but the last layer, we apply the rectified linear unit as a nonlinear activation function, which clips negative values to zero and preserves positive ones. In the last block, the second convolutional operation is directly followed by a linear activation function to deliver the final regression output.

We purposefully avoid to change the spatial resolution of the features in all layers (e.g., by applying pooling layers) as we want to limit any unnecessary loss in geometric resolution.

The network performance was assessed over a dedicated validation set, sampled from areas geographically disconnected from those of the training and test sets, respectively. More details on these aspects are addressed later on in Section V.

B. Input Features

As input feature maps to the network, we selected the following quantities, derived from TanDEM-X geocoded and coregistered bistatic products:

1) backscattering coefficient σ^0 in HH polarization, computed according to (5);
2) associated local incidence angle θ_{inc};
3) total interferometric coherence γ_{tot}, derived by Φ-Net [63];
4) volume decorrelation factor γ_{vol}, computed from (2);
5) height of ambiguity h_{amb}, computed from (4);
6) global TanDEM-X edited DEM.

θ_{inc} and h_{amb} are both required to correctly inform the model of the dependence of γ_{vol}, γ_{tot}, and σ^0 on such geometry-related features, as described in (2), (3), and (5). The DEM is used as input feature to exploit the capability of CNNs to recognize spatial patterns in the data and, thus, to better contextualize the radar-related input features with respect to the observed terrain characteristics.

An example of input feature maps is presented in Fig. 3. From the resulting input and output datacubes, we derived patches of 15×15 pixels, corresponding to an extension on ground of $375 \times 375\,\text{m}^2$ or, equivalently, to an area of approximately 14 ha. The patches are sampled at runtime, from areas which contain forested pixels (i.e., covered by the reference LVIS dataset) and which are fully covered by the input sources to the DL architecture.

C. Training Strategy

To train the model, we use a mini-batch strategy, consisting in the constant update of the model weights based on the iterative prediction over individual batches of training samples. We use batches composed of 250 patches each.

At each training iteration, we build a new batch of randomly sampled patches, and we then perform a forward pass in which the CH for the current batch is estimated. The quality of the prediction is then assessed by computing a loss function. This

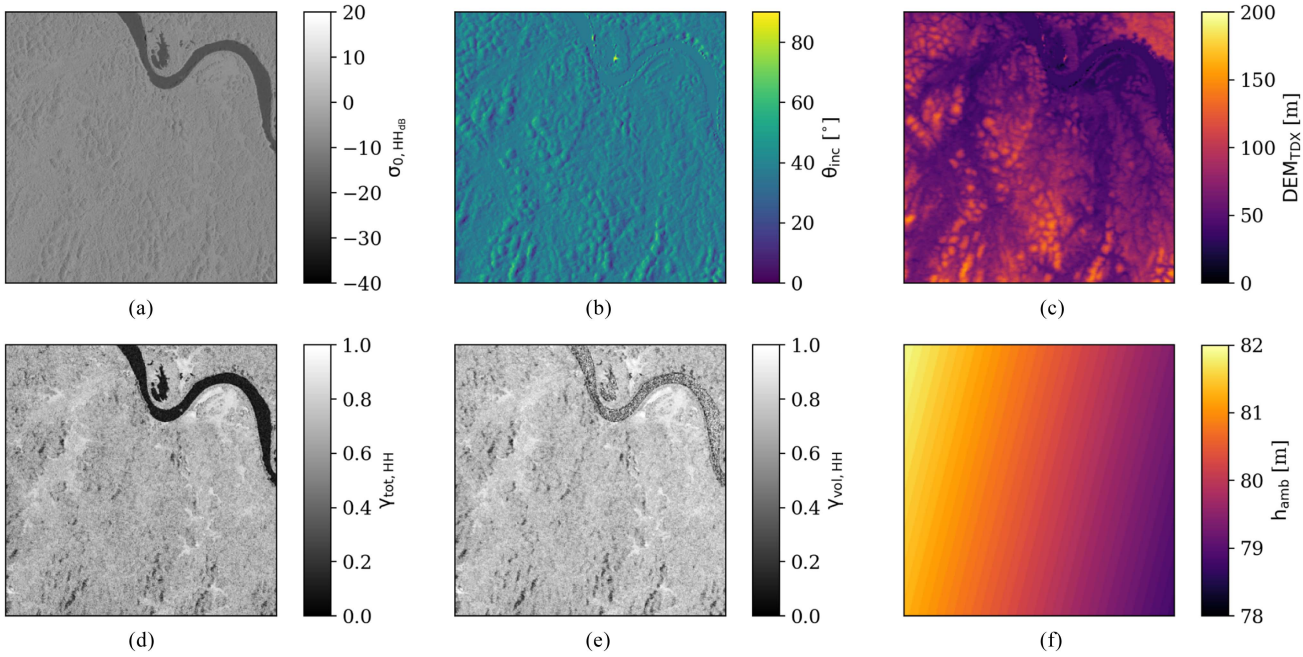

Fig. 3. Example of input feature maps. (a) Backscattering coefficient σ^0. (b) Local incidence angle θ_{inc}. (c) TanDEM-X edited DEM. (d) Interferometric coherence γ_{tot}. (e) Volume decorrelation factor γ_{vol}. (f) Height of ambiguity h_{amb}.

function quantifies the prediction error by computing the mean squared error between the prediction and its associated ground truth. It also includes a regularization term in the form of the l_2-norm (i.e., the Euclidean distance) among the model weights. We can, therefore, express the training objective as the solution to the minimization of the following loss function:

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 + \lambda \cdot \sum_{j=1}^{m} w_j^2 \tag{6}$$

where $\hat{y}_i$ is the ith CH sample predicted by the model, y_i is the corresponding ith ground-truth sample, w_j is the jth weight of the model, n is the total number of samples, m is the total number of weights, and λ is a factor that scales the impact that the l_2-norm has on the overall loss.

The impact of each model weight on the loss is computed as the gradient of the loss with respect to that weight (i.e., the partial derivative). By applying the chain rule, the local gradient is iteratively backpropagated from the output layers to the input of the network. Given the resulting set of local gradients, the corresponding weights are then updated following a stochastic-gradient-descent-based strategy; for our work, we apply the commonly used Adam optimization algorithm [69]. The initial learning rate of the network is set to 10^{-4}.

To monitor the training status of the model, we group the process into epochs of 500 batches each. At the end of each epoch, the training state of the model is evaluated on a dedicated independent validation set using the loss function in (6).

The relative evolution of two losses, one computed on the training set (i.e., the training loss) and the other computed on the validation set (i.e., the validation loss), is used to determine the completion of the training phase. If the network does not improve in terms of validation loss for 30 consecutive epochs, the learning rate is reduced by an order of magnitude. In order to avoid overfitting of our network on the training samples, we adopt an early stopping approach to cease the training if the network has not improved for a total of 35 consecutive epochs. Finally, at the end of the training phase, we select the weights corresponding to the epoch with the lowest validation loss.

D. Performance Evaluation Metrics

In order to test the performance of our trained model, we perform an inference run for all input acquisitions that cover the previously designated test sites. The prediction accuracy is evaluated using the mean error (ME), which identifies the estimator's bias, the MAE, the mean absolute percentage error (MAPE), the RMSE, and the coefficient of determination (R^2), which are defined as follows:

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{7}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{8}$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{11}$$

where $\hat{y}_i$ is the predicted CH value, y_i is the corresponding reference value, and $\bar{y}_i$ is the mean reference value.

V. Feature Analysis

In this section, we aim at obtaining a comprehensive overview of the impact that different input settings have on the final performance. In more detail, we are interested in better understanding the following key aspects:

1) the relevance that the different input features have in terms of regression performance, with particular interest on the SAR backscatter and InSAR information complementarity;
2) the impact that the SAR acquisition geometry has on the prediction accuracy, and the role that ancillary geometric features play in the contextualization of the side-looking geometry peculiarity of SAR systems.

We also performed an analysis of the impact of ascending and descending orbit directions on the final performance.[1] We observed that training and testing with either ascending-only or descending-only acquisitions, as well as with mixed-orbit acquisitions paired with supplementary orbit direction, acquisition time of day, or no information, showed negligible influence on the overall performance. Thus, to guarantee the best possible temporal coverage, our results are based on a similar ratio of both ascending and descending orbit acquisitions, as reported in Table IV.

A. Baseline Scenario

In order to properly conduct each of the proposed analyses, we start by setting up a baseline scenario against which the subsequent experiments will be compared. To this end, we aim at building three homogeneous but not overlapping subsets for a best case scenario, one for each of the training, validation, and test phases. To do so, we geographically split each of the four considered study areas into three subareas of equal dimensions. From West to East, each resulting subarea is assigned to the validation, training, and test sets, respectively. This concept is depicted in Fig. 1. The histograms of the reference CHM for the training, validation, and test sets are presented in Fig. 4. All the subsets show similar distribution trends, with most of the vegetation heights concentrated below 10 m (short vegetation), and between 30 and 40 m (which are typical of mature primary tropical forest).

During the training phase, the model weights are randomly initialized. Similarly, all training and validation patches are randomly shuffled after each one has been used once. These elements of randomness potentially cause the model to converge toward different local minima during the training phase, each time the process is repeated. In an effort to find the best possible local solution, we train five independent models. For each of them, we follow the training and validation procedure detailed in Section IV. Fig. 5 shows, in practice, the impact of random weights initialization on the RMSE performance of four test images, one for each test site. The different test sites show different degrees of variability. When looking at the overall variability, the RMSE spread is in the order of several decimeters.

[1]The considered TanDEM-X data were acquired in the morning in ascending orbit direction and in the afternoon in descending orbit direction.

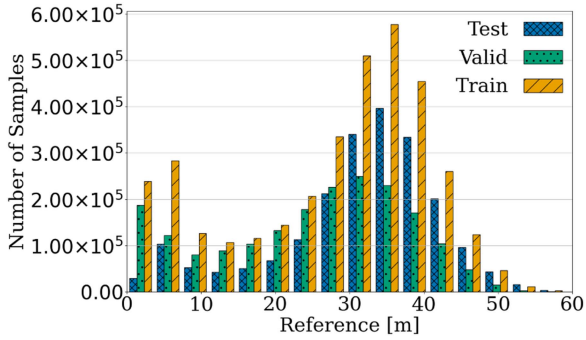

Fig. 4. Reference CHM value distributions for the three subsets, training, validation, and test, as defined for the baseline scenario.

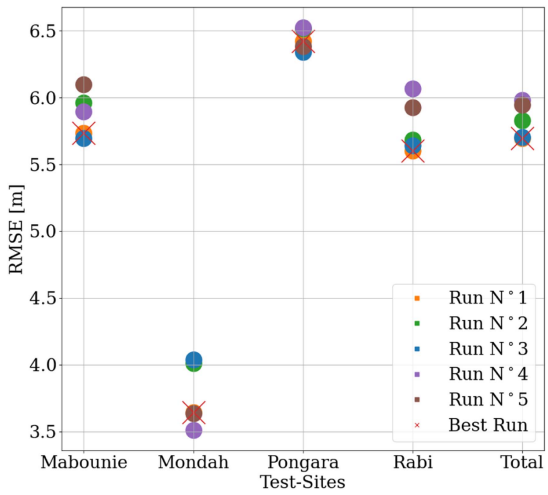

Fig. 5. RMSE values obtained for each independent training run, computed on each test site individually and overall (Total). The best run is automatically selected based on the validation loss.

Given that the model was trained using 15×15 pixels patches, the edges of each scene were symmetrically padded by half of the patch size as a preprocessing step prior to the inference. These borders were then removed to preserve the original size. Furthermore, we enlarged internal areas with invalid pixels by half the patch size as a final postprocessing step. These steps guarantee that the model is evaluated only over areas with similar conditions with respect to the training phase.

The loss behavior of the overall best performing model can be seen in Fig. 6. The results of the performance analysis on such a model are shown in the first row (Baseline case) of Table I. Here, the performance is computed for the joint combination of all available predictions. The detailed values for each test site and experiments can be found in Tables III and V (in the Appendix).

Overall, a good agreement between the prediction and the reference measurements can be seen from the scatterplot in Fig. 7, which comprises all test sites. Similar distributions of the reference CHM and of the prediction can be identified.

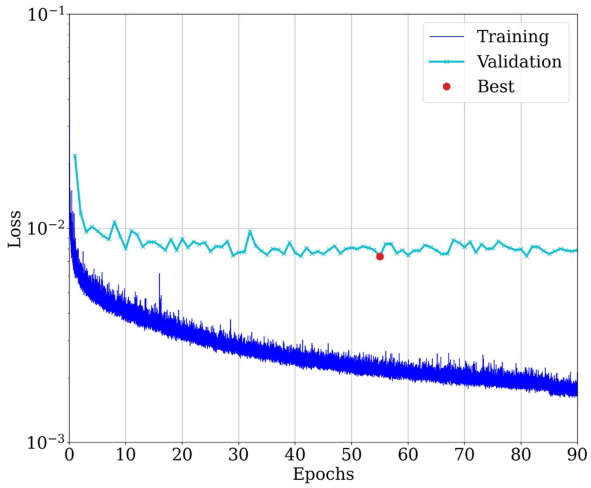

Fig. 6. Loss evolution computed on the training and validation sets for the baseline scenario.

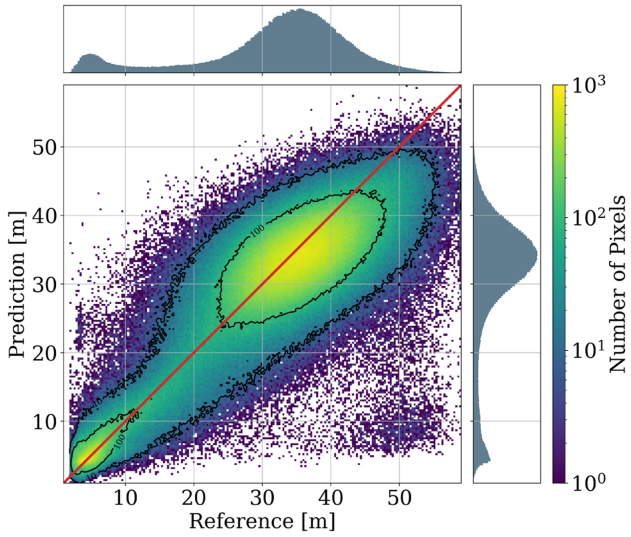

Fig. 7. Scatterplot in logarithmic scale of the predicted and reference CH values over all four test sites in Gabon, Africa, together with the side distributions of both the reference and the prediction for the baseline scenario.

TABLE II
OVERALL PREDICTION PERFORMANCE FOR THE DL BASELINE CASE COMPARED TO THE RVOG PREDICTIONS

Scenario	ME	MAE	MAPE	RMSE	R^2
	[m]	[m]	[%]	[m]	[·]
Baseline	−1.46	4.20	15.06	5.69	0.73
RVoG	0.84	6.49	33.87	8.33	0.43

Number of considered acquisitions = 12; number of pixels = 1 465 947.

TABLE III
RESULTS FOR THE BASELINE SCENARIO PRESENTED IN SECTION V AND FOR ALL CROSS-VALIDATION PERMUTATIONS OF THE SPATIAL TRANSFER ANALYSIS

	Performance Metrics						
	No. Acq.	**ME**	**MAE**	**MAPE**	**RMSE**	R^2	**No. pixels**
Unit		[m]	[m]	[%]	[m]	[·]	
Baseline Scenario							
Mabounié	4	−0.75	4.41	14.22	5.73	0.46	410 018
Mondah	4	0.31	2.49	27.51	3.65	0.88	88 371
Pongara	2	−2.31	3.92	16.21	6.42	0.82	253 394
Rabi	4	−1.78	4.39	13.60	5.60	0.40	714 164
Overall	12	−1.46	4.20	15.06	5.69	0.73	1 465 947
Spatial Transfer Analysis							
Mabounié	4	0.18	4.41	15.00	5.68	0.47	410 018
Mondah	4	0.53	2.37	27.31	3.49	0.89	88 371
Pongara	2	−3.97	5.65	23.73	7.96	0.73	253 394
Rabi	4	−2.49	4.82	14.61	6.20	0.27	714 164
Overall	12	−1.87	4.70	17.06	6.28	0.68	1 465 947

Each metric is evaluated for each test site individually and overall.

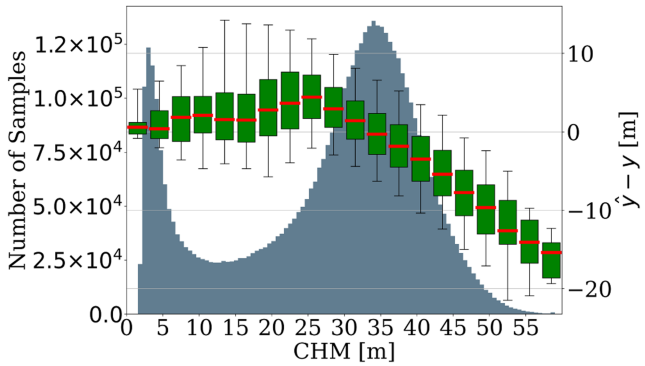

Fig. 8. Training sample distribution (blue histogram in the background—left axis) compared to the ME, which represents the regression bias (red lines in the foreground—right axis). The boxplots display the regression bias spread (5th, 25th, 50th, 75th, and 95th percentiles, respectively) across the CHM range.

When assessing the relationship between the estimation bias and the CHs, Fig. 8 shows that the estimation bias remains on average quite stable and close to 0 m for CHM below 35 m, above which the model starts to increasingly underestimate taller trees.

When looking at the performance of the individual test areas in Table III and Fig. 9, as expected, a clear relation with the distribution of the reference CH can be assessed. The Mondah forest is predominantly characterized by large extents of low vegetation, with isolated patches of tall forests. This reflects on the MAE and MAPE metrics, which are the smallest and the largest among all test sites, respectively. With lower tree heights, the absolute error scales accordingly, but the relative estimation error has a bigger impact on average. Differently, the site corresponding to the Pongara National Park is characterized by a more homogeneous tree height population. At first glance, the scatterplot in Fig. 9 shows a good agreement between the predicted and the expected CHs. Indeed, the performance is well balanced for all metrics, and it notably achieves an MAPE of only 16.21%, while obtaining a R^2 value of 0.82. A small cluster of samples is spotted outside of the main distribution (in the low-right corner of the image), corresponding to large values of CH being strongly underestimated (such samples are also clearly visible in the overall scatterplot in Fig. 7). This error can be attributed to inconsistencies in the TanDEM-X edited DEM, as represented in Fig. 10. Further investigations revealed how the area lacked of TanDEM-X data during the first two global coverages used for the generation of the global DEM, requiring the automatic editing algorithm to rely on an ALOS-based DEM to fill the gap. Unfortunately, also the ALOS-based DEM showed some inconsistencies in the same area, resulting in mutual calibration problems between the two different DEM

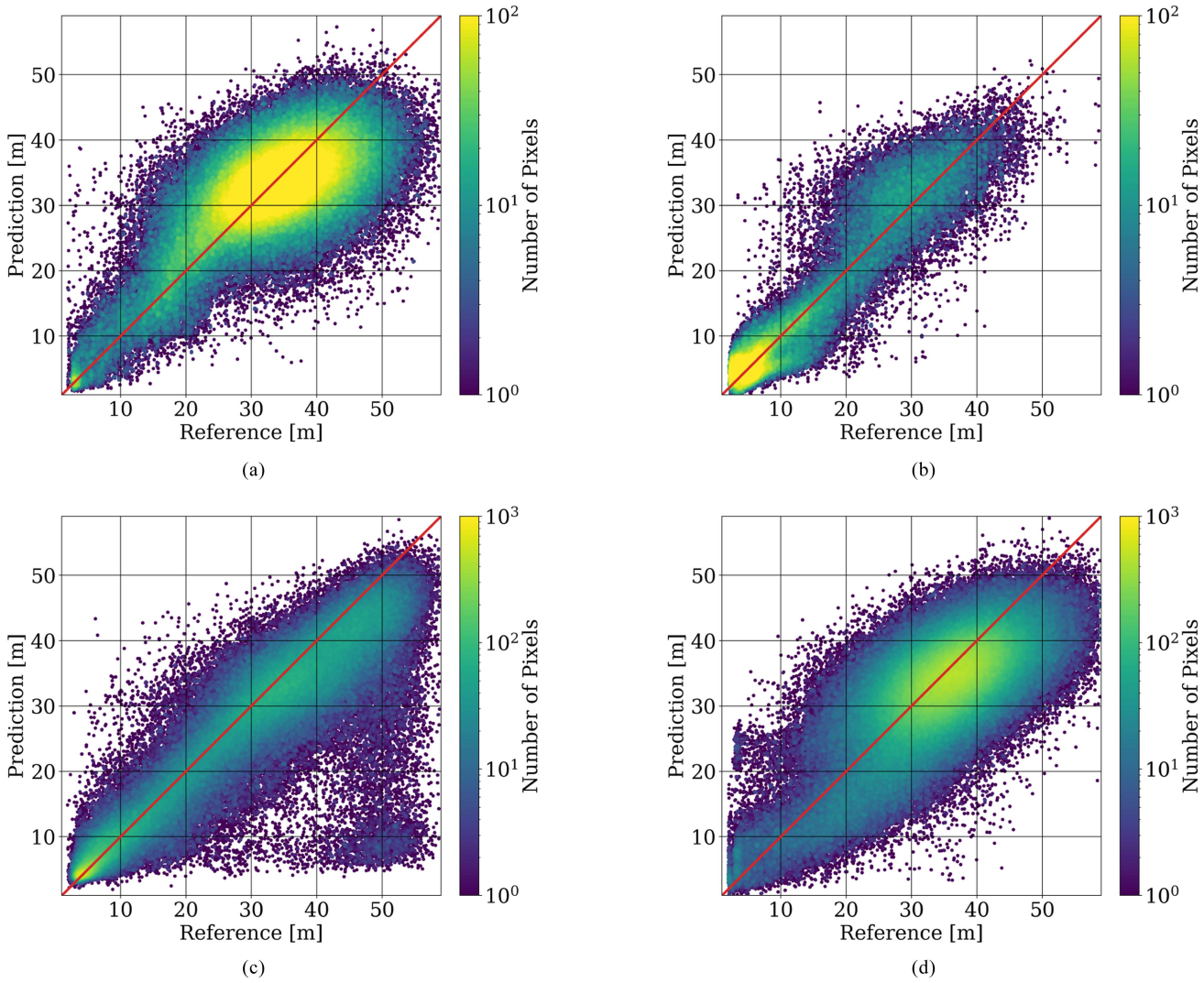

Fig. 9. Baseline scenario: scatterplots of the reference CHM versus the predicted CHM for each of the four considered test sites in logarithmic scale. (a) Mabounié. (b) Mondah. (c) Pongara. (d) Rabi.

sources. This led to an underestimated terrain height which was clipped to that of the adjacent sea level.

To quantify the impact of this nonnegligible inconsistency in the edited DEM, we manually masked the region (as depicted in Fig. 10) and reevaluated the performance metrics on the remaining valid areas. This resulted in a loss of 4686 test samples, corresponding to 1.85% of the Pongara samples, and 0.32% of the overall available test samples. The results in Table V (*Baseline w/ DEM mask* case) show a significant improvement across all metrics by removing the critical region. Indeed, the RMSE performance was especially negatively affected by the aforementioned problem and saw an improvement from 6.42 to 4.83 m, making Pongara the site with the best performance after Mondah. The overall performance in Table I shows the MAPE fall to 14.91% and the R^2 rise to 0.76.

The sites of Mabounié and Rabi exhibit similar behaviors, as they are predominantly covered by tall forest structures in the range of 25–45 m of height. In this scenario, the model uncertainty appears to increase, with the R^2 metric in particular being penalized by the very clustered concentration of tree heights. The MAE for both sites also worsens to about 4.4 m, while the MAPE improves as the estimation error for taller tree stand population has a lower relative impact on average.

Next, we assess the impact of the local topography in combination with the side-looking geometry of the instrument, by evaluating the relationship between the local terrain slope and the performance degradation. In order to compute a vegetation-independent slope estimate, we use the DTM derived from the LVIS measurements acquired during the AfriSAR campaign. We then subdivide the slope range into equally sized intervals, plotting for each of them the MAE of all associated predicted CHs (see Fig. 11). For slopes from 0° to 40°, an almost linear relationship can be identified between slope inclination and MAE. Notably, the MAE remains below the global average of 4.20 m up to slope inclinations of 25°, suggesting that the performance remains reasonably consistent even over high-relief terrain, where geometric distortions and geocoding errors in the SAR imagery and LiDAR measurements may become increasingly significant.

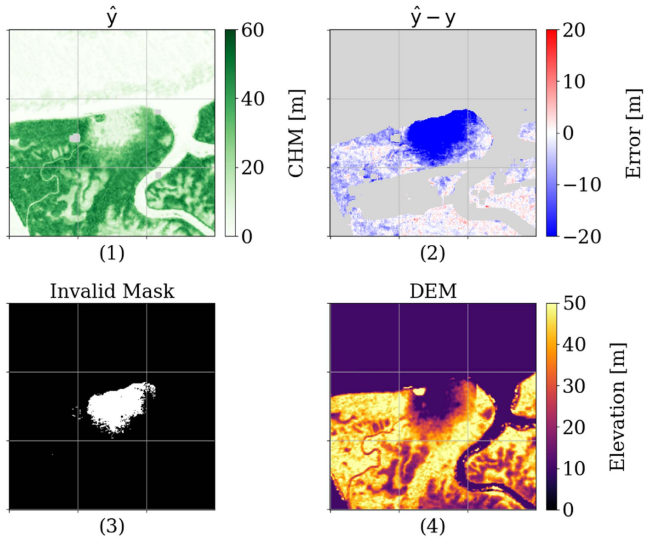

Fig. 10. Analysis of the Pongara test site, which is characterized by a region of heavily underestimated tree heights. (a) CHM prediction. (b) Prediction error. (c) Binary mask used for the identification of the problematic area. (d) TanDEM-X Global Edited DEM.

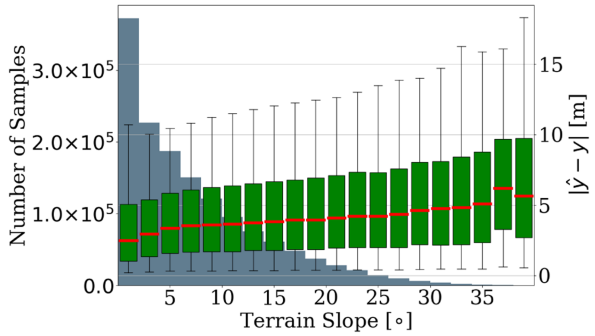

Fig. 11. Test sample distribution compared to the MAE. The boxplots display the MAE spread (5th, 25th, 50th, 75th, and 95th percentiles, respectively) across the terrain slope range.

Finally, Fig. 12 shows prediction samples for each of the four study areas for the Baseline scenario. In particular, sample transects are plotted following the red solid lines, in order to compare the estimated tree line with that of the reference. This shows an overall good agreement for both tall and low vegetation heights. Also, the TanDEM-X global edited DEM is depicted for comparison purposes, indeed highlighting how the mean phase center falls within the canopy. The DEM struggles to match all changes in canopy height, and, in general, it maintains a nonlinear relationship with the tree height.

B. Impact of DEM and Local Incidence Angle

The side-looking acquisition geometry characterizing SAR causes the TanDEM-X acquisitions to be intrinsically dependent on the local incidence angle, which is related to the acquisition geometry and the local topography. In order to allow the network to better contextualize the variations in the backscattered signal, ancillary information in the form of the local incidence angle (θ_{inc}) and the TanDEM-X DEM were assumed to be necessary additions to the input feature pool. To test this assumption against the baseline scenario performance, we repeat the tests in Section V-A, while removing θ_{inc}, the DEM, or both features.

The overall performance metrics are summarized in Table I (cases: *w/o DEM*, *w/o* θ_{inc} and *w/o DEM, w/o* θ_{inc}). A significant drop in estimation accuracy is associated with the removal of one of the two features.

Moreover, Table I shows indeed that removing both features from the input causes the performance to drop significantly, but somewhat in line with the *w/o* θ_{inc} scenario.

C. SAR and Bistatic InSAR Feature Comparison

One of the main advantages of the TanDEM-X constellation is the capability to acquire bistatic InSAR information. As described in [35] and [42], the quality of the interferometric product correlates with the composition of forests through the volume scattering mechanism. To properly assess the impact that the additional bistatic interferometric feature set shows with respect to the available single-polarization backscatter map, we present two more scenarios in which we selectively remove one of the two feature sets and validate them against the Baseline scenario. For both test scenarios, we keep the ancillary θ_{inc} and DEM features.

The accuracy for the SAR-only (*w/o InSAR features* case) and InSAR-only (*w/o* σ^0 case) predictions are presented in Table I. Both the scenarios see a decrease in the overall performance when compared to the Baseline scenario. In particular, testing the results for the SAR-only scenario sees the MAPE worsen by 2.12% and the RMSE increase by 0.68 m when compared to the InSAR-only scenario.

In the previous experiments, we used as InSAR-related input features both the interferometric coherence (γ_{tot}) and the volume decorrelation factor (γ_{vol}), derived from the original coherence, as well as the ancillary height of ambiguity (h_{amb}) information, to help the network to better contextualize the relationship between the volumetric decorrelation effects and the acquisition geometry, as detailed in Section II. In order to fully understand the importance that each single InSAR-related feature holds for solving the regression problem, we alternatively remove each of these from the input set, achieving the regression performance detailed in the last five rows of Table I. Notably, removing either the coherence (*w/o* γ_{tot} case) or the volume decorrelation factor (*w/o* γ_{vol} case) results in a very consistent and minor loss in the overall prediction accuracy.

Removing h_{amb} from the input (*w/o* h_{amb} case) sees the regression performance fall in line with the baseline scenario. This is an interesting result given that both γ_{tot} and the γ_{vol} over vegetated areas are strictly related to h_{amb} [42]. To better contextualize this result, we set up two further scenarios in which we never use h_{amb}, and we alternate γ_{tot} and γ_{vol} as sole InSAR-based input features. This refers to the *w/o* γ_{vol}, *w/o* h_{amb} and *w/o* γ_{tot}, *w/o* h_{amb} cases in Table I, respectively.

The results in Table I suggest that γ_{tot} does not require information about the acquisition h_{amb} in order to maximize the prediction accuracy, while the addition of γ_{vol} helps to further improve the performance bringing it in line with the baseline scenario. Contrary to γ_{tot}, using γ_{vol} also requires the additional information about h_{amb} in order to achieve comparable results.

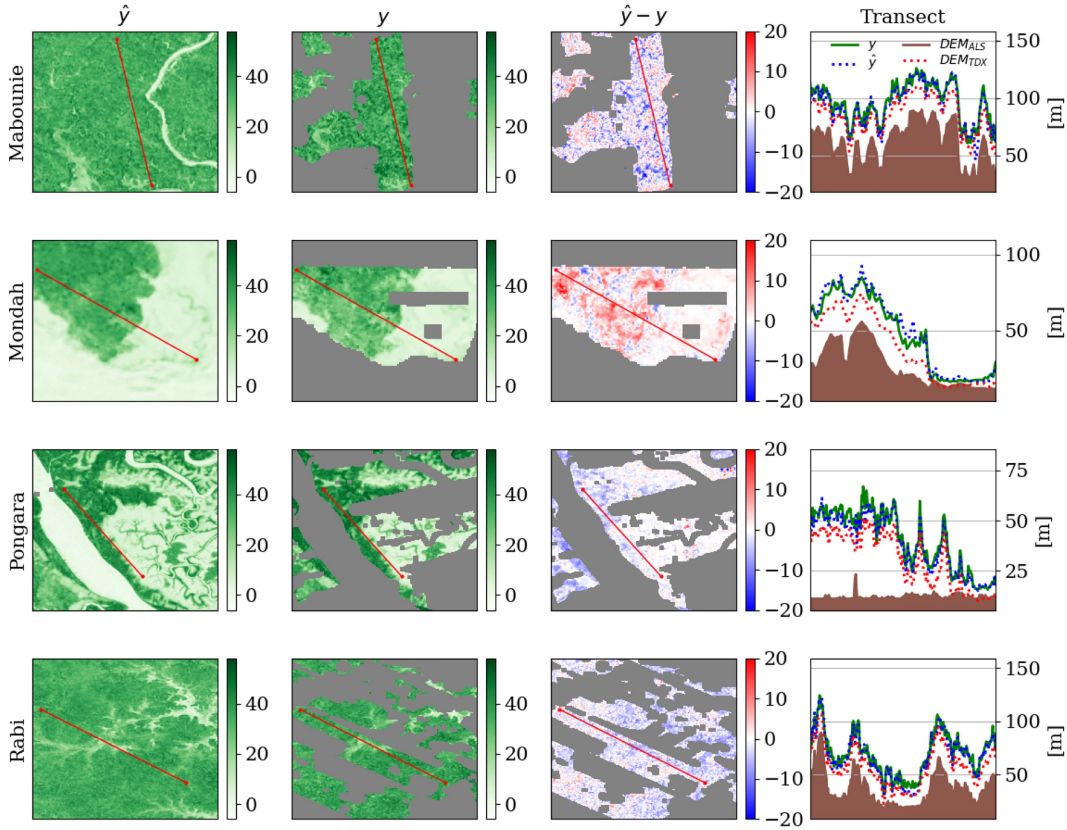

Fig. 12. Prediction examples over the four considered test sites for the Baseline scenario. For each area, the prediction, the reference, the estimation error, and sample vertical transect (identified by the red line) are presented from left to right, respectively. The transects in the rightmost column compare the reference CHM (in green) with our predictions (in blue), on top of ALS-derived DEM (in brown). The corresponding TanDEM-X Global Edited DEM is plotted (in red) for visual comparison, highlighting how the mean phase center lies within the canopy.

D. Comparison With the RVoG Model

In order to help better contextualize the performance of the proposed DL model, we implemented a high level comparison of the baseline scenario with the RVoG model, which is often used in the literature [23], [24], [44], [45]. Given the relationship expressed in (3), the RVoG model parameterizes the vertical reflectivity function $F(z)$ as a two-layer model consisting of a Dirac-like ground component and a vegetation volume component, which is modeled as a continuously extended volume layer of randomly oriented scatterers [21]. This results in the complex volume coherence γ_{vol} to depend on multiple model parameters, requiring fully polarimetric acquisitions or external reference data in order to allow for the model inversion. By following the strategies detailed in [44] and [45], it is possible to reduce the number of unknowns required for the inversion, ultimately allowing the CH h_v to be directly estimated from single-pol acquisitions according to the following relationship:

$$h_v = h_{\text{amb}} \left(1 - \frac{2}{\pi} \sin^{-1}(|\gamma_{\text{vol}}|)\right). \tag{12}$$

The resulting equation is independent of external sources and only requires information about the height of ambiguity h_{amb} and the estimated volume decorrelation γ_{vol}, making it feature compatible with our proposed method. Using this model approximation, we repeat the tests detailed in the baseline scenario to provide a direct comparison between the two approaches.

The resulting performance metrics presented in Table II show that the overall performance of the approximated RVoG model is worse than the one of the proposed DL model, with an RMSE of 8.33 m that is sharply higher than 5.69 m obtained in the proposed baseline scenario. Of particular interest is the comparison between Figs. 7 and 13. It is possible to note that the RVoG model typically fails to properly predict the height of canopies below 10 m and tends to overestimate up to about 30 m of height. For higher tree heights, the bias decreases, but the uncertainty of the prediction increases. This behavior can partially be explained by the approximations and assumptions used by the RVoG model, as well as by the lack of proper ground topography information to precisely estimate the height of ambiguity.

VI. Analysis of the Generalization Capability in the Spatial Domain

As highlighted in Section I, the goals for our proposed DL framework include the capability to extend the work on both larger scales and across different scenarios. In order to investigate the generalization capabilities of the network, we choose to

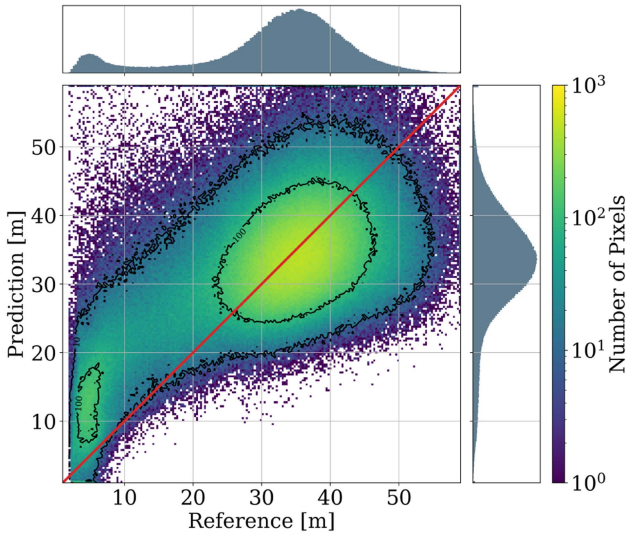

Fig. 13. Scatter plot in logarithmic scale for the RVoG model, comparing the predicted and reference CH values over all four test sites in Gabon, Africa, together with the side distributions of both the reference and the prediction for the baseline scenario.

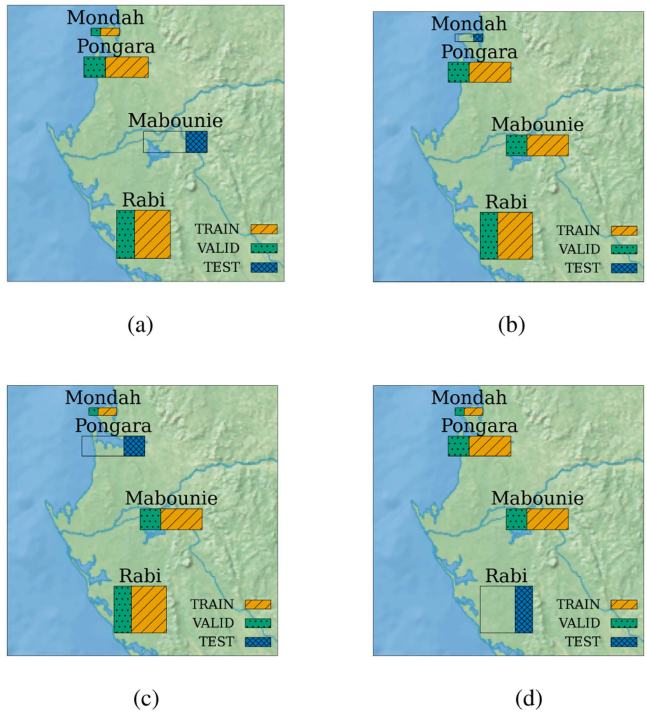

Fig. 14. Geographical composition of training, validation, and test sets, for each of the four cross-validation permutations implemented for the spatial generalization analysis. (a) Mabounié. (b) Mondah. (c) Pongara. (d) Rabi.

adopt a leave-one-site-out cross-validation approach, by which we iteratively set aside one of the four study areas to be solely used to test the model performance, leaving the other three sites for training and validation. Within each experiment, we start from the study area assignment strategy detailed in Section V-A and depicted in Fig. 1, but we reassign the rightmost third area of each of the three training and validation sites to be part of the training set instead. This choice guarantees that we can dispose of enough training samples to replace those which are lost by dedicating a whole site to testing alone. Furthermore, given the implicit assumption that data-driven approaches are fundamentally limited in their application scope by the comprehensiveness and representativeness of the training set composition, the proposed test approach represents a fair attempt to provide a rich-enough training pool to the network, while keeping the test area geographically isolated. The subdivision of the different sites for training, validation, and testing is presented in Fig. 14 for each of the four cross-validation permutations.

The results for each testing permutation are presented in Table III, together with those of the baseline scenario for each test site for comparison purposes. Overall, the results show an appreciable loss in regression accuracy when compared to the baseline scenario. The MAE and MAPE increase from 4.2 m and 15.1% to 4.7 m and 17.1%, respectively, while R^2 drops from 0.73 to 0.68. The test results for the Mabounié and Mondah sites show a better agreement with the baseline performance, indicating that the complexities of both areas were equivalently well captured by the remaining training sites.

In the case of the Pongara area, the performance is considerably degraded. A comparison of the scatterplot in Fig. 15 with that of Fig. 9 suggests that in the cross-validation scenario, the model struggles to regress higher canopies, essentially saturating at around 45 m of height. As described in Section III-A, the sites of Mondah and Pongara are partially covered by Mangroves. In particular, Pongara presents taller canopies than those found in Mondah. Indeed, Fig. 16 highlights how the underestimation problem is evident close to the shorelines, where the WorldCover Map predicts mangrove coverage. We conclude that removing Pongara from the training set severely limits the amount of tall mangrove examples, compromising the regression performance.

Finally, the analysis of the Rabi test site also highlights a performance lower than expected. In particular, the prediction associated with the TanDEM-X imagery acquired on October 28, 2016 displays significant underestimation issues. A specific investigation into the problem revealed that the underestimated area coincides with that of an anomalous drop in backscatter intensity. Fig. 17 shows a comparison between the examined acquisition and a reference one, which does not display the same issues. Neither the reference dataset nor high resolution optical satellite imagery gives a potential explanation for the drop in backscatter intensity. We suppose that this could have been induced by severe weather conditions, as thunderstorms are known to be capable of causing strong attenuation of the backscatter returns [70].

VII. Discussion

The results of the experiments related to the feature analysis presented in Section V provide interesting insights on the capabilities of TanDEM-X data in combination with DL for forest height retrieval, which are extremely useful for understanding the impact and the relationship of different input features.

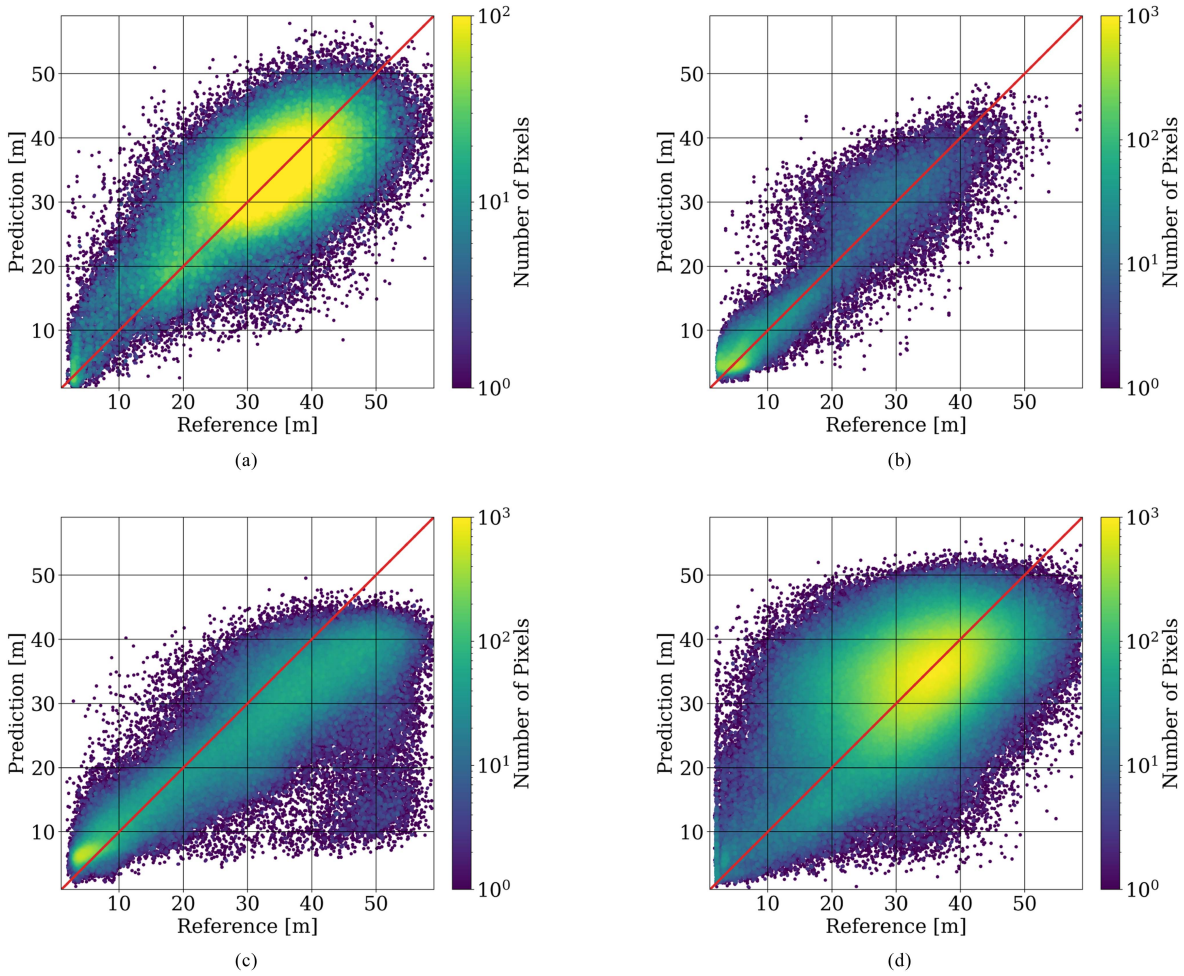

Fig. 15. Spatial transfer analysis: cross-validation scatterplots for each of the four test sites in logarithmic scale. (a) Mabounié. (b) Mondah. (c) Pongara. (d) Rabi.

From the baseline scenario in Section V-A, where all input features are considered, we observe a sudden increase in the estimation bias for trees taller than 35 m, resulting in underestimation. On the one hand, this behavior might be linked to the poor availability of training samples for such heights, as it can be inferred from the corresponding histogram. On the other hand, it could be due to the limited penetration capability of X-band, which leads to saturation effects in the data when the radar wave is not able to penetrate deep enough into extremely tall vegetation [65]. Moreover, by separately analyzing the different test sites, one can note that none of the considered performance metrics is on its own sufficient to conclusively evaluate the prediction performance of the model. This happens as the different scenarios, and in particular their CH distributions, have different impacts on the accuracy metrics.

By analyzing the impact of the DEM and the local incidence angle as input features, one can notice that removing the DEM-derived θ_{inc} map has a bigger impact than removing the DEM itself. This is probably caused by the fact that θ_{inc} shows an intrinsic dependence on the slant-range distance, which is directly mirrored in SAR features such as the backscatter. The performance analysis in the baseline scenario (see Section V-A) already demonstrated the potential impact of inconsistencies in the input DEM, pointing out the overall relevance that the DEM holds in terms of inference performance. These two insights suggest that the model is able to partially substitute the information contained in the DEM with that of θ_{inc}, but not *vice versa*. As expected, both features are required to achieve peak accuracy, with neither of them being able to completely supersede the other one.

Regarding the impact of SAR and bistatic InSAR features, the five considered test scenarios resulted in the following intuitions.

1) We noted that removing the backscatter information has a significantly smaller impact than removing the InSAR features.
2) The removal of either the coherence or the volume decorrelation factor causes a minor loss in the overall accuracy. This suggests that, while not completely redundant, both features hold very similar information. This is in line with the theoretical expectations of their respective definitions, as the volume decorrelation factor quantifies a constituent decorrelation source of the total interferometric coherence, as detailed in (2).
3) The use of the total coherence γ_{tot} does not strictly require information about the height of ambiguity h_{amb}, whereas the use of the volume decorrelation factor γ_{vol} requires

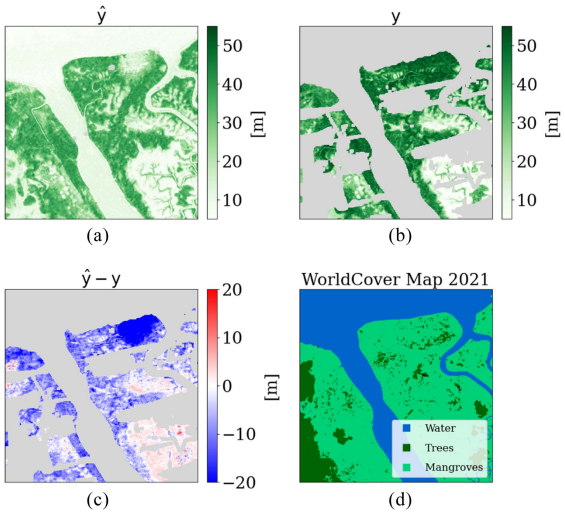

Fig. 16. Example of underestimation over Pongara. (a) CHM prediction. (b) Reference LiDAR CHM. (c) Prediction error. (d) Corresponding land cover map. The areas covered by tall mangroves are particularly affected by error.

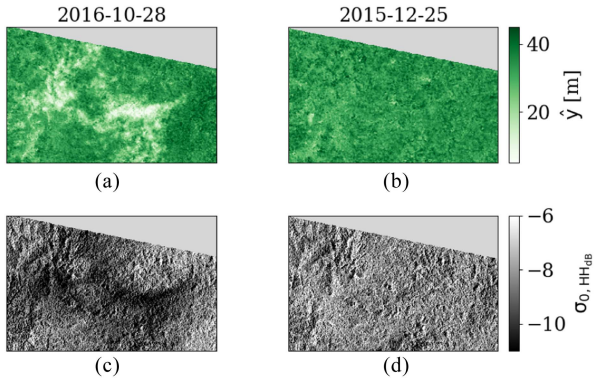

Fig. 17. Prediction comparison for the same area in Rabi, resulting from two different TanDEM-X acquisitions. The prediction on the left presents a significant region of underestimation (a), well aligned with a region of low backscatter values (c). (b) and (d) Image on the right does not present either pattern.

information about h_{amb}. The use of either γ_{tot} or γ_{vol} (and h_{amb}) achieves similar performance, while using both matches the baseline scenario. Differently, the use of both γ_{tot} and γ_{vol} drops the requirement for the presence of h_{amb}. We think that these inconsistencies between the γ_{tot} and γ_{vol} cases might come from the fact that γ_{vol} is potentially more sensitive to changes in h_{amb} than γ_{tot}, as the latter still contains the other decorrelation factors, as described in (2).

Furthermore, the composition of the utilized TanDEM-X dataset presented in the Appendix (see Table IV) shows that the vast majority of data were acquired with h_{amb} in the 75–90 m range. Only three products have h_{amb} between 67 and 69 m, and other three have h_{amb} between 44 and 50 m. For h_{amb} values above 60 m, the γ_{vol} decay over forested areas has been shown to be relatively small [42], whereas it rapidly increases when h_{amb} falls below 50 m. This suggests that our test set is not representative enough across the h_{amb} range to require a complete complex modeling of the relationship between γ_{vol} and h_{amb}.

The additional comparison of the proposed baseline scenario with the RVoG model confirms the potential of the proposed DL framework for CH estimation, with a considerable performance improvement of almost 3 m in terms of RMSE.

Finally, the preliminary analysis on the spatial generalization capability of the neural network presented in Section VI highlights that this is able to correctly regress forest height when considering previously unseen areas. Nevertheless, given the strong dependence of the achievable performance on the distribution of the training dataset, we are aware that the currently designed architecture still needs to be further optimized in order to reach a level of robustness and generalization, which is suitable for a reliable large-scale inference. Indeed, given the focus on the AfriSAR 2016 Campaign test sites, only a limited amount of both TanDEM-X acquisition geometries and kinds of vegetation were considered to train the network.

VIII. Conclusion

In this article, we presented a novel study on the potential of DL for the regression of forest height from TanDEM-X bistatic InSAR data. We proposed a fully convolutional framework, which is capable of delivering tree height estimates from a single TanDEM-X acquisition, enabling the creation of large-scale and up-to-date geomaps. We trained the model in a supervised fashion using LVIS data from the 2016 AfriSAR campaign, consisting of ALS-derived CH maps covering tropical forests in Gabon, West Africa. We achieved an overall regression performance of 4.12-m MAE, 14.9% MAPE, and 5.41-m RMSE, for a best case scenario (baseline).

Our feature analysis showed that the regression performance is primarily driven by InSAR features and that ancillary information about the acquisition geometry as well as scene topography is crucial to deliver peak performance. We also performed a high-level comparison of the baseline scenario settings with the single-polarization approximation of the RVoG model. The obtained performance is in line with or above the performance of state-of-the-art methods for CH estimation using TanDEM-X data, presented in [19], [23], and [24] and using Sentinel-2 and Sentinel-1 data as in [28] and [29].

Our spatial transfer analysis demonstrated that under suitable conditions, the network is capable of correctly transferring knowledge acquired during training onto previously unseen regions of the same geographical area.

Our experiments have also highlighted that the achievable accuracy remains strongly connected with the characteristics of the training set, requiring a careful balancing in order to preserve performance over different challenging scenarios. We also encountered saturation problems for very tall canopies, which we attribute to two separate phenomena: on the one hand, the limited penetration capabilities of X-band over forested

TABLE IV
LIST OF THE UTILIZED TANDEM-X COSSC PRODUCTS AND MAIN ACQUISITION PARAMETERS

List	Acquisition Date	AIID	E[θ_i]	h_{amb}	Orbit Direction
1	2015-10-12	1318689	47.7	91.4	A
2	2015-10-12	1318689	47.7	90.6	A
3	2015-10-12	1316435	47.7	88.6	A
4	2015-10-23	1322791	46.3	80.9	A
5	2015-10-23	1322791	46.2	81.4	A
6	2015-10-23	1322791	46.2	84.4	A
7	2015-11-03	1322908	44.5	78.8	A
8	2015-11-03	1322908	44.5	76.3	A
9	2015-11-03	1322908	44.5	76.8	A
10	2015-11-03	1322908	44.5	79.3	A
11	2015-11-14	1323042	47.7	82.1	A
12	2015-11-22	1323134	42.7	79.6	D
13	2015-11-22	1323134	42.6	77.5	D
14	2015-12-08	1323317	31.3	49.6	D
15	2015-12-14	1323371	40.6	76.0	D
16	2015-12-14	1323371	40.6	75.0	D
17	2015-12-25	1323439	38.5	68.3	D
18	2015-12-25	1323439	38.4	68.5	D
19	2015-12-25	1323439	38.4	67.0	D
20	2015-12-28	1325427	46.2	74.2	A
21	2015-12-28	1325427	46.2	74.8	A
22	2015-12-28	1325427	46.2	74.6	A
23	2015-12-30	1325426	28.8	44.3	D
24	2015-12-30	1325426	28.8	44.5	D
25	2016-10-17	1381828	37.1	80.9	D
26	2016-10-17	1381828	37.1	80.3	D
27	2016-10-17	1381828	37.1	78.5	D
28	2016-10-28	1382273	39.3	85.9	D
29	2016-10-28	1382273	39.3	85.3	D
30	2016-10-28	1382273	39.3	87.3	D

AIID is the unique acquisition item identifier, E[θi] is the mean incidence angle at beam center, hamb is the height of ambiguity in meters, and the orbit direction is characterized as A for ascending and D for descending.

areas and, on the other hand, a general lack of labeled data for very tall trees above 40 m.

In order to mitigate these effects, in future works, we will extend the analysis to higher resolutions, as TanDEM-X products with fewer looks have already shown a certain potential to retrieve the location of the ground surface [71]. Furthermore, we will investigate smarter sampling and balancing approaches for our training dataset, aiming at homogenizing the network performance. As a final remark, we would like to point out that the proposed spatial transfer analysis was performed on a limited region of interest (the state of Gabon) and, thus, should be considered as a preliminary assessment of the generalization capability of the network. Therefore, more effort will also be devoted to the generalization of the model for large-scale inference, e.g., by considering a larger span of viewing geometries and InSAR configurations, as well as forest types, local terrain characteristics, and spaceborne-derived reference data.

APPENDIX
ADDITIONAL TABLES

The complete list of utilized TanDEM-X CoSSC (single-look coregistered L1b bistatic) products for this work is presented in Table IV. The performance for all analyzed scenarios as in Section V and for each test site, separately, is summarized in Table V. Note that the performance of the baseline scenario can already be found in Table III.

TABLE V
PERFORMANCE SUMMARY OF ALL ANALYZED SCENARIOS AS IN SECTION V AND FOR EACH TEST SITE, SEPARATELY

Performance Metrics							
	No. Acq.	ME	MAE	MAPE	RMSE	R^2	No. pixels
Unit		[m]	[m]	[%]	[m]	[·]	
Baseline w/ DEM mask							
Mabounié	4	-0.75	4.41	14.22	5.73	0.46	410 018
Mondah	4	0.31	2.49	27.51	3.65	0.88	88 371
Pongara	2	-1.80	3.45	15.37	4.83	0.90	248 708
Rabi	4	-1.78	4.39	13.60	5.60	0.40	714 164
Overall	12	-1.37	4.12	14.92	5.42	0.76	1 461 261
w/o DEM							
Mabounié	4	-1.46	4.86	15.65	6.43	0.32	410 018
Mondah	4	0.60	2.93	35.98	3.97	0.86	88 371
Pongara	2	-0.90	4.89	25.71	6.57	0.82	253 394
Rabi	4	-1.18	4.31	13.70	5.70	0.38	714 164
Overall	12	-1.10	4.48	17.67	5.98	0.71	1 465 947
w/o θ_{inc}							
Mabounié	4	-1.51	4.56	14.46	5.91	0.43	410 018
Mondah	4	0.31	2.18	24.00	3.31	0.91	88 371
Pongara	2	-4.11	5.17	19.16	8.02	0.72	253 394
Rabi	4	-2.80	4.82	14.47	6.09	0.29	714 164
Overall	12	-2.48	4.65	15.85	6.30	0.67	1 465 947
w/o DEM, w/o θ_{inc}							
Mabounié	4	-1.24	5.03	16.50	6.55	0.30	410 018
Mondah	4	-0.08	2.54	25.80	3.77	0.88	88 371
Pongara	2	-2.54	5.64	25.36	7.56	0.76	253 394
Rabi	4	-1.23	4.65	15.21	5.98	0.32	714 164
Overall	12	-1.39	4.80	17.96	6.33	0.67	1 465 947
w/o σ^0							
Mabounié	4	-1.34	4.94	15.71	6.50	0.31	410 018
Mondah	4	0.10	2.23	22.14	3.48	0.89	88 371
Pongara	2	-2.34	3.83	16.56	6.28	0.83	253 394
Rabi	4	-1.51	4.50	13.94	5.74	0.37	714 164
Overall	12	-1.51	4.37	15.38	5.95	0.71	1 465 947
w/o InSAR features							
Mabounié	4	-0.35	4.81	15.85	6.25	0.36	410 018
Mondah	4	0.14	2.58	26.84	3.91	0.87	88 371
Pongara	2	-2.74	4.66	21.56	7.45	0.76	253 394
Rabi	4	-3.16	5.29	15.87	6.79	0.12	714 164
Overall	12	-2.10	4.88	17.51	6.62	0.64	1 465 947
w/o γ_{tot}							
Mabounié	4	-0.63	4.44	14.56	5.70	0.47	410 018
Mondah	4	-0.64	2.43	23.05	3.69	0.88	88 371
Pongara	2	-3.14	4.33	18.64	6.74	0.81	253 394
Rabi	4	-1.92	4.46	13.63	5.69	0.39	714 164
Overall	12	-1.69	4.31	15.32	5.79	0.72	1 465 947
w/o γ_{vol}							
Mabounié	4	-0.49	4.36	14.32	5.59	0.49	410 018
Mondah	4	-0.38	2.28	22.72	3.46	0.90	88 371
Pongara	2	-2.67	4.04	16.85	6.58	0.81	253 394
Rabi	4	-2.65	4.67	14.11	5.91	0.34	714 164
Overall	12	-1.91	4.33	15.16	5.83	0.72	1 465 947
w/o h_{amb}							
Mabounié	4	0.47	4.51	15.04	5.80	0.45	410 018
Mondah	4	0.03	2.36	24.67	3.49	0.89	88 371
Pongara	2	-2.57	4.05	16.73	6.54	0.82	253 394
Rabi	4	-1.56	4.28	13.12	5.47	0.43	714 164
Overall	12	-1.07	4.19	14.98	5.67	0.74	1 465 947
w/o γ_{vol}, w/o h_{amb}							
Mabounié	4	0.03	4.40	14.56	5.66	0.47	410 018
Mondah	4	0.96	2.36	26.11	3.66	0.88	88 371
Pongara	2	-2.23	3.90	15.92	6.58	0.81	253 394
Rabi	4	-1.67	4.47	13.77	5.70	0.38	714 164
Overall	12	-1.13	4.22	15.10	5.75	0.73	1 465 947
w/o γ_{tot}, w/o h_{amb}							
Mabounié	4	-1.71	5.52	17.23	7.56	0.06	410 018
Mondah	4	0.14	2.14	22.21	3.28	0.91	88 371
Pongara	2	-2.61	4.04	16.61	6.64	0.81	253 394
Rabi	4	-2.36	4.56	13.69	5.91	0.34	714 164
Overall	12	-2.07	4.59	15.70	6.42	0.66	1 465 947

The performance of the baseline scenario can be found in Table III.

REFERENCES

[1] Food and Agriculture Organization, *Global Forest Resources Assessment 2020*. Rome, Italy: Food and Agriculture Organization, 2020. [Online]. Available: http://www.fao.org/documents/card/en/c/ca9825en

[2] P. G. Aron, C. J. Poulsen, R. P. Fiorella, and A. M. Matheny, "Stable water isotopes reveal effects of intermediate disturbance and canopy structure on forest water cycling," *J. Geophys. Res.: Biogeosci.*, vol. 124, no. 10, pp. 2958–2975, 2019, doi: 10.1029/2019JG005118.

[3] M. Wolosin and N. Harris, "Tropical forests and climate change: The latest science," World Resources Institute, Washington, DC, USA, 2018, p. 14. [Online]. Available: https://www.wri.org/research/ending-tropical-deforestation-tropical-forests-and-climate-change-latest

[4] D. Ellison, "Forests and water: Background analytical study," United Nations, Thirteenth session of the United Nations Forum on Forests, 2018. [Online]. Available: https://www.un.org/esa/forests/wp-content/uploads/2018/04/UNFF13_BkgdStudy_ForestsWater.pdf

[5] *Forest Carbon*, Natural Resources Canada. Accessed. Aug. 22, 2023. [Online]. Available: https://www.nrcan.gc.ca/climate-change/impacts-adaptations/climate-change-impacts-forests/forest-carbon/13085

[6] J. K. Agee, "The influence of forest structure on fire behavior," in *Proc. 17th Forest Vegetation Manage. Conf.*, 1996, pp. 52–68.

[7] H.-S. Song and S.-H. Lee, "Effects of wind and tree density on forest fire patterns in a mixed-tree species forest," *Forest Sci. Technol.*, vol. 13, no. 1, pp. 9–16, 2017, doi: 10.1080/21580103.2016.1262793.

[8] K. Calders et al., "Nondestructive estimates of above-ground biomass using terrestrial laser scanning," *Methods Ecol. Evol.*, vol. 6, no. 2, pp. 198–208, 2015, doi: 10.1111/2041-210X.12301.

[9] *Surveying the Forest*, Bundeswaldinventur. Accessed. Aug. 22, 2023. [Online]. Available: https://www.bundeswaldinventur.de/en/third-national-forest-inventory/surveying-the-forest/

[10] N. Picard, L. Saint-André, and M. Henry, "Manual for building tree volume and biomass allometric equations from field measurement to prediction," Food and Agriculture Organization of the United Nations (FA0), Quebec City, QC, Canada, 2012. [Online]. Available: http://www.fao.org/docrep/018/i3058e/i3058e.pdf

[11] J. Askne, J. Fransson, M. Santoro, M. Soja, and L. Ulander, "Model-based biomass estimation of a hemi-boreal forest from multitemporal TanDEM-X acquisitions," *Remote Sens.*, vol. 5, no. 11, pp. 5574–5597, Oct. 2013.

[12] A. Alboabidallah, J. Martin, S. Lavender, and V. Abbott, "Using Landsat-8 and Sentinel-1 data for above ground biomass assessment in the Tamar valley and Dartmoor," in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2017, pp. 1–7.

[13] C. T. Almeida et al., "Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms," *Remote Sens. Environ.*, vol. 232, 2019, Art. no. 111323.

[14] H. Astola, T. Hame, L. Sirro, M. Molinier, and J. Kilpi, "Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in Boreal region," *Remote Sens. Environ.*, vol. 223, pp. 257–273, 2019.

[15] T. 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 Sens. Environ.*, vol. 264, 2021, Art. no. 112533.

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

[17] P. Potin, B. Rosich, N. Miranda, and P. Grimont, "Sentinel-1 mission status," *Procedia Comput. Sci.*, vol. 100, pp. 1297–1304, 2016.

[18] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[19] F. Kugler, D. Schulze, I. Hajnsek, H. Pretzsch, and K. P. Papathanassiou, "TanDEM-X Pol-InSAR performance for forest height estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6404–6422, Oct. 2014.

[20] T. Jucker et al., "Allometric equations for integrating remote sensing imagery into forest monitoring programmes," *Glob. Change Biol.*, vol. 23, no. 1, pp. 177–190, 2017.

[21] K. Papathanassiou and S. Cloude, "Single-baseline polarimetric SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2352–2363, Nov. 2001.

[22] S. R. Cloude and K. Papathanassiou, "Three-stage inversion process for polarimetric SAR interferometry," *Proc. Inst. Elect. Eng.—Radar, Sonar Navigat.*, vol. 150, no. 3, pp. 125–134, Jun. 2003.

[23] R. Guliaev, V. Cazcarra-Bes, M. Pardini, and K. Papathanassiou, "Forest height estimation by means of TanDEM-X InSAR and waveform Lidar data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3084–3094, 2021.

[24] C. Choi et al., "Large scale forest height mapping by combining TanDEM-X and GEDI data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2374–2385, 2023.

[25] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[26] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

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

[28] N. Lang, K. Schindler, and J. D. Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111347.

[29] A. Becker, S. Russo, S. Puliti, N. Lang, K. Schindler, and J. D. Wegner, "Country-wide retrieval of forest structure from optical and SAR satellite imagery with Bayesian deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 195, pp. 269–286, 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0924271622003045

[30] P. López-Dekker et al., "The Harmony mission: End of phase-0 science overview," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 7752–7755.

[31] G. Krieger et al., "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317–3341, Nov. 2007.

[32] P. Rizzoli et al., "Generation and performance assessment of the global TanDEM-X digital elevation model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 132, pp. 119–139, Oct. 2017.

[33] M. Lachaise, M. Bachmann, T. Fritz, M. Huber, B. Schweisshelm, and B. Wessel, "The TanDEM-X Change DEM: The new temporal DEM of the TanDEM-X mission," in *Proc. 13th Eur. Conf. Synth. Aperture Radar*, 2021, pp. 1–6.

[34] M. Zink et al., "TanDEM-X: 10 years of formation flying bistatic SAR interferometry," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3546–3565, 2021.

[35] M. Martone, B. Braeutigam, P. Rizzoli, C. Gonzalez, M. Bachmann, and G. Krieger, "Coherence evaluation of TanDEM-X interferometric data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 73, pp. 21–29, Sep. 2012.

[36] P. Rizzoli, M. Martone, and B. Braeutigam, "Global interferometric coherence maps from TanDEM-X quicklook data," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1861–1865, Nov. 2014.

[37] R. Touzi, A. Lopes, J. Bruniquel, and P. W. Vachon, "Coherence estimation for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 135–149, Jan. 1999.

[38] C.-A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1441–1452, Apr. 2011.

[39] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, "InSAR-BM3D: A nonlocal filter for SAR interferometric phase restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3456–3467, Jun. 2018.

[40] H. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 950–959, Sep. 1992.

[41] P. Rizzoli, L. Dell'Amore, J. Bueso-Bello, N. Gollin, D. Carcereri, and M. Martone, "On the derivation of volume decorrelation from Tan-DEM-X bistatic coherence," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3504–3518, 2022.

[42] M. Martone, P. Rizzoli, and G. Krieger, "Volume decorrelation effects in TanDEM-X interferometric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1812–1816, Dec. 2016.

[43] M. J. Soja, H. Persson, and L. Ulander, "Estimation of forest height and canopy density from a single InSAR correlation coefficient," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 646–650, Mar. 2015.

[44] H. Chen, S. Cloude, and D. G. Goodenough, "Forest canopy height estimation using TanDEM-X coherence data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3177–3188, Jul. 2016.

[45] A. Olesk, J. Praks, O. Antropov, K. Zalite, T. Arumae, and K. Voormansik, "Interferometric SAR coherence models for characterization of hemiboreal forests using TanDEM-X data," *Remote Sens.*, vol. 8, no. 700, pp. 1–23, Aug. 2016.

[46] H. Chen, S. Cloude, D. G. Goodenough, D. A. Hill, and A. Nesdoly, "Radar forest height estimation in mountainous terrain using TanDEM-X coherence data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3443–3452, Oct. 2018.

[47] M. Nannini et al., "Coherence-based SAR tomography for space-borne applications," *Remote Sens. Environ.*, vol. 225, pp. 107–114, Jun. 2019.
[48] M. Schlund, P. Magdon, B. Eaton, C. Aumann, and S. Erasmi, "Canopy height estimation with TanDEM-X in temperate and boreal forests," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 82, Oct. 2019, Art. no. 101904.
[49] R. Treuhaft et al., "Tropical-forest biomass estimation at X-band from the spaceborne TanDEM-X interferometer," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 239–243, Feb. 2015.
[50] W. Qi and R. O. Dubayah, "Combining TanDEM-X InSAR and simulated GEDI Lidar observations for forest structure mapping," *Remote Sens. Environ.*, vol. 187, pp. 253–266, Dec. 2016.
[51] J. Askne, M. Soja, and L. Ulander, "Biomass estimation in a boreal forest from TanDEM-X data, LiDAR DTM, and the interferometric water cloud model," *Remote Sens. Environ.*, vol. 196, pp. 265–278, Jul. 2017.
[52] H. Persson, M. Soja, J. Fransson, and M. Ulander, "National forest biomass mapping using the two-level model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6391–6400, 2020.
[53] E. Erten, C. Rossi, and O. Yüzügüllü, "Polarization impact in TanDEM-X data over vertical-oriented vegetation: The paddy-rice case study," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1501–1505, Jul. 2015.
[54] J. Lopez-Sanchez, F. Vicente-Guijalba, E. Erten, M. Campos-Taberner, and F. Garcia-Haro, "Retrieval of vegetation height in rice fields using polarimetric SAR interferometry with TanDEM-X data," *Remote Sens. Environ.*, vol. 192, pp. 30–44, Apr. 2017.
[55] M. Busquier, J. Lopez-Sanchez, A. Mestre-Quereda, E. Navarro, M. Gonzalez-Dugo, and L. Mateos, "Exploring TanDEM-X interferometric products for crop-type mapping," *Remote Sens.*, vol. 12, no. 11, Jun. 2020, Art. no. 1774.
[56] M. Schlund, F. von Poncet, D. H. Hoeckman, S. Kuntz, and C. Schmullius, "Importance of bistatic SAR features from TanDEM-X for forest mapping and monitoring," *Remote Sens. Environ.*, vol. 151, pp. 16–26, Aug. 2014.
[57] M. Martone et al., "The global forest/non-forest map from TanDEM-X interferometric SAR data," *Remote Sens. Environ.*, vol. 205, pp. 352–373, Feb. 2018.
[58] M. Martone, F. Sica, C. Gonzalez, J.-L. Bueso-Bello, P. Valdo, and P. Rizzoli, "High-resolution forest mapping from TanDEM-X interferometric data exploiting nonlocal filtering," *Remote Sens.*, vol. 10, no. 9, Sep. 2018, Art. no. 1477.
[59] A. Mazza, F. Sica, P. Rizzoli, and G. Scarpa, "TanDEM-X forest mapping using convolutional neural networks," *Remote Sens.*, vol. 11, 2019, Art. no. 2980.
[60] J. Armston et al., "AfriSAR: Gridded forest biomass and canopy metrics derived from LVIS, Gabon, 2016," 2020. [Online]. Available: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1775
[61] J. Blair, D. L. Rabine, and M. A. Hofton, "The laser vegetation imaging sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography," *ISPRS J. Photogrammetry Remote Sens.*, vol. 54, no. 2-3, pp. 115–122, Jul. 1999.
[62] T. Fritz, H. Breit, C. Rossi, U. Balss, M. Lachaise, and S. Duque, "Interferometric processing and products of the TanDEM-X mission," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2012, pp. 1904–1907.
[63] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, "ϕ-Net: Deep residual learning for INSAR parameters estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3917–3941, May 2021.
[64] C. Gonzalez, M. Bachmann, J. Bueso-Bello, P. Rizzoli, and M. Zink, "A fully automatic algorithm for editing the TanDEM-X global DEM," *Remote Sens.*, vol. 12, no. 23, Dec. 2020, Art. no. 3961.
[65] C. Choi, M. Pardini, J. Armstrong, K. Papathanassiou, and R. Dubayah, "Relating TanDEM-X local InSAR phase center variations to LiDAR full waveforms over forest scenarios," in *Proc. Eur. Conf. Synth. Aperture Radar*, 2018, pp. 684–687.
[66] D. Zanaga et al., *ESA WorldCover 10m 2021 v200*. Paris, France: Eur. Space Agency, 2022.
[67] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.
[68] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html
[69] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2014, pp. 1–15.
[70] A. Danklmayer and M. Chandra, "Precipitation induced signatures in SAR images," in *Proc. Eur. Conf. Antennas Propag.*, 2009, Art. no. 5.
[71] Y. Lei, R. Treuhaft, and F. Goncalves, "Automated estimation of forest height and underlying topography over a Brazilian tropical forest with single-baseline single-polarization TanDEM-X SAR interferometry," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112132.

Daniel Carcereri received the B.Sc. degree in electronics and telecommunications engineering and the M.Sc. degree in information and communication engineering from the University of Trento, Povo, Italy, in 2018 and 2020, respectively. He is currently working toward the Ph.D. degree in information and communication technology with the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento, and the Microwaves and Radar Institute, German Aerospace Center, Weßling, Germany.

He completed his M.Sc. thesis on quantization error reduction in synthetic aperture radar (SAR) systems with the Microwaves and Radar Institute, German Aerospace Center. In 2023, he visited the Radar Science and Engineering Section, NASA Jet Propulsion Laboratory, Pasadena, CA, USA, to work on *L*-band repeat-pass interferometric SAR data. His research interests include signal processing, estimation theory, and deep learning algorithms applied to forest parameter estimation and analysis.

Paola Rizzoli received the B.Sc. and M.Sc. degrees in telecommunication engineering from the Politecnico di Milano (Polimi), Milan, Italy, in 2003 and 2006, respectively, and the Ph.D. degree (*summa cum laude*) in electrical engineering and information technology from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2018.

From 2006 to 2008, she was a Scientific Researcher and Project Engineer with the Politecnico di Milano and Aresys s.r.l., a Polimi spin-off company. At the end of 2008, she joined the Microwaves and Radar Institute, German Aerospace Center (DLR), Weßling, Germany, as Project Engineer, where she is involved in the development and optimization of the TerraSAR-X and TanDEM-X missions, concentrating, in particular, on the generation of the TanDEM-X global digital elevation model. From 2016 to 2020, she led the System Performance Research Group, Satellite Synthetic Aperture Radar (SAR) Systems Department, DLR, being responsible for the final performance assessment of the global TanDEM-X DEM and the generation of the global TanDEM-X Forest/Non-Forest map. Since 2020, she has been leading the Radar Science Research Group, DLR. Her main research interests include SAR systems design, data reduction techniques, estimation theory, signal processing, and artificial intelligence algorithms.

Dr. Rizzoli received the DLR Science Award in 2018 and the Best Paper Award at German Microwave Conference in 2019. She regularly serves as a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Dino Ienco (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Torino, Torino, Italy, in 2006 and 2010, respectively.

In 2011, he joined the TETIS Laboratory, IRSTEA, Montpellier, France, as a Junior Researcher. He is currently with INRAE, Montpellier. His main research interests include machine learning, data science, graph databases, social media analysis, information retrieval, and spatiotemporal data analysis with a particular emphasis on remote sensing data and earth observation data fusion.

Dr. Ienco served in the Program Committee of many international conferences on data mining, machine learning, and database including IEEE International Conference on Data Mining, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Asian Conference on Machine Learning, and International Joint Conference on Artificial Intelligence as well as served as a Reviewer for many international journal in the general field of data science and remote sensing.

Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor of Telecommunications with the University of Trento, Povo, Italy, where he teaches remote sensing, radar, and digital communications. He is the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He promotes and supervises research within the framework of many national and international projects. He is the Principal Investigator of many research projects. He is currently the Principal Investigator of the Radar for Icy Moon Exploration instrument in the framework of the Jupiter Icy Moons Explorer mission of the European Space Agency (ESA) and the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA. He is the author (or coauthor) of more than 360 scientific publications in referred international journals, more than 350 papers in conference proceedings, and 22 book chapters. He is editor/coeditor of 18 books/conference proceedings and one scientific book. His papers are highly cited, as proven from the total number of citations (more than 47 000) and the value of the H-index (103) (source: Google Scholar). He was invited as keynote speaker in more than 40 international conferences and workshops. His current research interests include remote sensing, radar and synthetic aperture radar, signal processing, machine learning, and pattern recognition.

Dr. Bruzzone has been a Member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2009. He has been the Vice-President for Professional Activities of the IEEE GRSS since 2019. He ranked first place in the Student Prize Paper Competition of 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, WA, USA, in July 1998. He is recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards, and the 2019 WHISPER Outstanding Paper Award. He was a Guest Co-Editor of many special issues of international journals. He is the co-founder of IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images series and is a Member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He is the founder of *IEEE Geoscience and Remote Sensing Magazine*, for which he was an Editor-in-Chief between 2013 and 2017. He is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE GRSS between 2012 and 2016.