Insights on the working principles of a CNN for forest height regression from single-pass InSAR data

Daniel Carcereri, Luca Dell'Amore, Stefano Tebaldini, *Senior Member, IEEE*, and Paola Rizzoli, *Member, IEEE*

***Abstract*—The increasing use of Artificial Intelligence (AI) models in Earth Observation (EO) applications, such as forest height estimation, has led to a growing need for explainable AI (XAI) methods. Despite their high accuracy, AI models are often criticized for their "black-box" nature, making it difficult to understand the inner decision-making process. In this study, we propose a multifaceted approach to XAI for a convolutional neural network (CNN)-based model that estimates forest height from TanDEM-X single-pass InSAR data. By combining domain knowledge, saliency maps, and feature importance analysis through exhaustive model permutations, we provide a comprehensive investigation of the network working principles. Our results suggests that the proposed model is implicitly capable of recognizing and compensating for the SAR acquisition geometry-related distortions. We find that the mean phase center height and its local variability represents the most informative predictor. We also find evidence that the interferometric coherence and the backscatter maps capture complementary but equally relevant views of the vegetation. This work contributes to advance the understanding of the model's inner workings, and targets the development of more transparent and trustworthy AI for EO applications, ultimately leading to improved accuracy and reliability in the estimation of forest parameters.**

***Index Terms*—Synthetic Aperture Radar, SAR Interferometry, TanDEM-X, Forest Height, Deep Learning, Convolutional Neural Network, Explainable AI**

I. Introduction

Evidence from recent studies suggests that data-driven approaches for forest parameter estimation from Earth Observation (EO) data may achieve accuracies comparable to, and in some cases exceeding, that of physical-based models, even in the absence of privileged data sources [1] [2] [3]. The reason behind these improvements remains fundamentally unclear as machine learning models, especially those based on Deep Learning (DL), remain notoriously difficult to explain. This lack of interpretability is clearly connected to the large number of model parameters (typically in the order of millions) and the use of multiple nonlinear operations, rendering the representation spaces created by these models often incomprehensible. This aspect is arguably made more difficult by the fundamental challenge posed by data leakage, in which overoptimistic results might lead to incorrect interpretations [4].

D. Carcereri (corresponding author), L. Dell'Amore and P. Rizzoli are with the Microwaves and Radar Institute, German Aerospace Center, Münchener Str. 20, 82234 Wessling, Germany. E-mail: Daniel.Carcereri@dlr.de, Luca.Dellamore@dlr.de, Paola.Rizzoli@dlr.de. S. Tebaldini is with the Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Via Ponzio 34/5, 20133 Milan, Italy. E-mail: Stefano.Tebaldini@polimi.it.

The so-called "black-box" nature of AI models increasingly represents a practical limit to their deployment, as interests of scientific, economic, and security nature call for more explainable models. On the one hand, human-interpretable explanations for models' decisions become a requirement in contexts where prediction-guided decision processes hold intrinsic risks and therefore require knowledge about the trustworthiness of the prediction. On the other hand, an improved understanding of the decision-making process can lead to better accuracy, since, in response, developers can learn to inform the model in a more effective way [5] [6].

Within the broader context of Earth observation (EO), explainable AI (XAI) efforts constitute a limited, although growing, area of research [7], which is predominantly concentrated on classification tasks [6] [8]. Possible strategies to improve interpretability include the definition of inherently interpretable models (e.g., linear regression, random forest, Bayesian deep learning) [5] [9], the incorporation of domain knowledge [10], the integration of physical-based models [11], and the use of post hoc explanation methods [7] [12]. A variety of these strategies, first proposed in the context of computer vision tasks, have found their way into EO applications [6]. Much of the existing research efforts are limited to the adoption of a single explainability approach [13] [8]. The most commonly used ones carry out feature importance analyses [7], which consist of the selective removal of input features to evaluate their impact on model performance [14]. Due the large number of model parameters, such analyses remain prohibitive to realize in practice and, as a consequence, they are commonly approximated using perturbation-based methods such as occlusions [15], LIME [16], and SHAP [17]. Unfortunately, these typically fail to consider the relative scaling of features within the models, or tend to confuse feature relevance with out-of-distribution (OOD) performance [18]. Also commonly used are gradient-based methods for the generation of Saliency maps [7]. These approaches include, but are not limited to, Guided Backpropagation [19], Integrated Gradient [18] and Grad-CAM [20]. These are intended to provide human-intelligible explanations, but have also been shown to perform image recovery rather than relevance attribution, leading to unreliable interpretations of a model's decision-making process [21].

In this study, we offer a detailed analysis on the inner working principles of a CNN-based model for the regression of forest height measurement from TanDEM-X InSAR data. We achieve this by relying on a set of XAI methods and cross-

checking their results. First, the proposed CNN is designed to include domain knowledge. Second, we rely on the use of saliency maps to assess, for each predictor (i.e., the input features to the model), the relevance of their spatial context in the forest height inversion problem. Afterwards, we perform a detailed feature importance analysis by training, validating and testing over 100 individual model-permutations on different input feature subsets. We explicitly capture intra-run variance by independently repeating each model permutation three times, allowing for better isolating the feature relevance from the intrinsic fluctuations of the training procedure. By design, we avoid the pitfalls of data leakage by geographically separating each subset. We adopt a physics-aware stance to the analysis of these experiments, being particularly careful at considering the acquisition processes and the co-dependencies between the input features [22], and taking into account the insights offered by physical-based models [23].

The paper is structured as follows. Sec. II introduces the satellite acquisitions and the reference data used for the experiments. Sec. III describes the model architecture, the training and testing strategies, as well as the XAI methods. Sec. IV presents the collection of experiments and their results. Sec. V discusses and combines the findings across different experiments, and offers further analyses to complete the interpretation of the results. Finally, Sec. VI presents a summary of the work, the conclusions and possible future developments.

II. Dataset

A. Study Area and Reference Data

The study area considered in this work coincides with that covered by the LiDAR-based forest height measurements performed during the 2016 AfriSAR campaign [24] [25]. It consists of five heterogeneous areas of interest (shown in Fig. 1), covering different tropical forests and a Savannah region situated within the Gabonese Republic. The reference forest height estimates are derived from full-waveform LiDAR measurements, acquired using the NASA airborne Land Vegetation and Ice Sensor (LVIS) instrument [26] in February and March 2016. Each of the densely sampled laser footprints is used to compute forest height values, expressed as relative height (RH), and aggregated to generate raster maps with a sampling resolution of 25 m. The RH99 statistic is used as reference measurement for this work, which is the height corresponding to the 99^{th} percentile of the returned energy. Furthermore, also the LVIS-derived and rasterized digital terrain model (DTM) is used as a reference for the real topographic elevation, measured at the bottom of the vegetation.

B. TanDEM-X Data

The TanDEM-X mission is a spaceborne interferometric SAR system, consisting of two twin satellites acquiring data at X-band and orbiting in a varying close orbit formation, which allows for the generation of single-pass DEM products [27], [28].

In this work, we consider only bistatic TanDEM-X data acquired in stripmap mode in HH polarization. Due to the limited acquisition resources of the mission, a long period from December 2010 to the end of 2022 is considered for the products, allowing to adequately capture a diversified combination of perpendicular baselines and incidence angles across each of the study sites. The potential effects of the temporal lag between the LiDAR and TanDEM-X measurements has been previously addressed in [3] and has been found to have a negligible impact on the results. Therefore, we do not further consider this aspect in the current work.

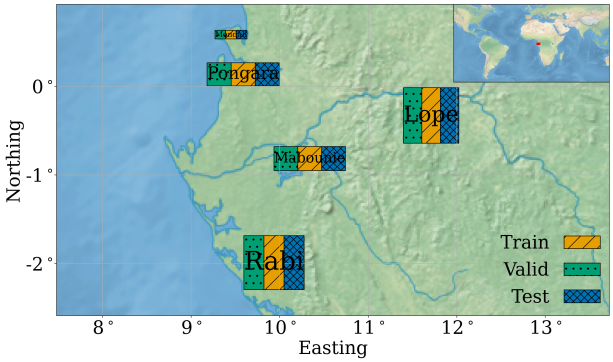


Fig. 1: The proposed geographic sub-setting of the AfriSAR campaign study areas into training, validation and testing

Each product, consisting of a primary (s_1) and a secondary (s_2) acquisition, has been focused and co-registered using the operational TanDEM-X processor (ITP) [29]. Starting from this co-registered single-look complex (CoSSC) format, different higher level features are computed or estimated. The intensity image for the primary acquisition is calibrated to compute the radar brightness β^0, which is in turn used to compute the backscattering coefficient σ^0 as:

$$\sigma^0 = \beta^0 \sin(\theta_{\text{inc}}), \tag{1}$$

where θ_{inc} is the incidence angle estimated using the TanDEM-X Global Edited DEM product at 30 m resolution [30]. The application of further terrain flattening and terrain correction steps (γ^0) are specifically avoided, as over forested areas the quality of the X-band DEM is impacted by the presence of vegetation.

The interferometric image pair s_1 and s_2 is used to estimate the normalized complex correlation coefficient γ as

$$\gamma \approx \frac{\sum_N^{i=1} \left\{ s_{1,i} s_{2,i}^* \right\}}{\sqrt{\sum_N^{i=1} |s_{1,i}|^2} \sqrt{\sum_N^{i=1} |s_{2,i}|^2}}, \tag{2}$$

where $N = 121$ is the number of samples within an 11×11 window (as used in standard TanDEM-X InSAR products [31]) for the asymptotically unbiased coherence estimation [32], and s_2^* is the complex conjugate of the secondary acquisition.

The argument of this quantity ($\angle\gamma$) forms the interferogram, which is used for the generation of the InSAR DEM, here defined as *raw DEM* $h_{\text{dem}}^{\text{acq}}$ to distinguish it from the other DEM products considered in this study. It is relevant to this study to recall that an InSAR-derived DEM measures the topographic elevation corresponding to the location of the mean phase center within a given resolution cell. Furthermore,

the SAR signal at X-band has a modest penetration capability into volumetric targets [23]. These factors result in the InSAR DEM being affected by a positive elevation bias (i.e., compared to the elevation of the ground), which depends on the height and structure of the vegetation, as well as the InSAR acquisition geometry [33].

The modulus of the normalized complex correlation coefficient ($|\gamma|$) is called interferometric coherence γ_{tot} and it assesses the amount of noise affecting the interferogram [32].

It is possible to determine the sources and magnitudes of individual noise terms by factorizing γ_{tot} as [34]

$$\gamma_{\mathrm{tot}} = \gamma_{\mathrm{SNR}} \cdot \gamma_{\mathrm{quant}} \cdot \gamma_{\mathrm{amb}} \cdot \gamma_{\mathrm{rg}} \cdot \gamma_{\mathrm{az}} \cdot \gamma_{\mathrm{temp}} \cdot \gamma_{\mathrm{vol}}, \quad (3)$$

where each term on the right-hand side identifies a unique decorrelation contribution, namely those due to a limited signal-to-noise ratio (γ_{SNR}), quantization errors (γ_{quant}), ambiguities (γ_{amb}), baseline separation (γ_{rg}), relative shift of the Doppler spectra (γ_{az}), temporal baseline (γ_{temp}) and the presence of volumetric scatterers (γ_{vol}). The latter is of particular interest to this study as it quantifies the amount of interferometric decorrelation resulting from interactions with volumetric targets, such as forests. Given the single-pass acquisition capability of the TanDEM-X mission, it is possible to ignore temporal decorrelation events (i.e., $\gamma_{\mathrm{temp}} = 1$). This enables one to directly estimate γ_{vol} by dividing γ_{tot} by the product of the remaining decorrelation terms. In practice, we estimate (γ_{quant}) using error profiles pre-computed for each quantization setting, we estimate (γ_{snr}) using noise equivalent sigma zero profiles pre-computed for each TanDEM-X beam, and we account an additional decorrelation factor of 0.98 for the remaining terms in eq. (3). The comprehensive derivation of these profiles and their motivation is detailed in [34].

Furthermore, we compute two additional geomaps aimed at correctly informing the inversion model about the SAR and InSAR acquisitions geometries, respectively. First, we use the annotated information of the position of the primary satellite and the topographic height measured by $h_{\mathrm{dem}}^{\mathrm{acq}}$ to generate the local incidence angle map $\theta_{\mathrm{lia}}^{\mathrm{acq}}$. Second, we use information on the satellites' positions to generate a height of ambiguity h_{amb} map, which is defined as the vertical change in height corresponding to a complete phase cycle in the interferogram. For the bistatic case it is defined as [27]

$$h_{\mathrm{amb}} = \frac{\lambda \cdot r \cdot \sin\theta_{\mathrm{inc}}}{B_{\perp}}, \quad (4)$$

where $B_{\perp}$ is the baseline component orthogonal to the line of sight, λ is the electromagnetic wavelength, and r is the slant-range distance.

III. Methodology

This section introduces the DL-based regression framework, detailing the model architecture, and the training, validation and testing procedures, as well as the XAI approaches.

A. Model Architecture

The model used in this work is based on that of [3] [35]. It consists of the fully-convolutional network (FCN) shown in Fig. 2, which was specifically designed for the task of regressing forest parameters from satellite InSAR data. We define as the reference for the proposed experiments the particular model which makes use of the following set of TanDEM-X-derived predictors:

- The backscattering coefficient in HH polarization, σ_0^{HH},
- The raw acquisition DEM, $h_{\mathrm{dem}}^{\mathrm{acq}}$,
- The estimated total interferometric coherence, γ_{tot},
- The estimated volume decorrelation factor, γ_{vol},
- The local incidence angle computed on $h_{\mathrm{dem}}^{\mathrm{acq}}$, $\theta_{\mathrm{lia}}^{\mathrm{acq}}$,
- The height of ambiguity, h_{amb}.

In general, CNNs operate by performing multiple cross-correlation operations between the input features and multiple kernel filters, which are defined by the training process. The resulting set of features is applied to a non-linear function to generate the final output. These operations are repeated multiple times in sequence according to a flow-through design, taking as input the features from the previous block to generate those for the following one. The last block has the special role of generating the final forest parameter estimate. Crucially, for non-unitary kernel sizes, the relationship between neighboring pixels (colloquially referred to as *texture*) is considered during the computation of the output pixels. The overall image extent that is taken into account by the model, for each predicted pixel, defines its *receptive field*.

B. Model Training, Validation and Testing

For training, validating and testing the model, each of the regions of interest is split into three distinct and non-overlapping sub-regions, one for each dedicated subset (see Fig. 1). This choice represents a trade-off between accurately capturing the heterogeneity found across the five study areas and minimizing data leakage, that is, an excessive geographic correlation between training, validation, and test sets leading to inflated performance estimates. Matching input and output feature-pairs are sourced and built from acquisitions that geographically overlap with the reference data. In particular, the training and validation steps are performed on batches of 15x15 pixel patches that are sampled from their respective subsets.

Training is performed using a fully-supervised approach, consisting in the minimization of the following two-term objective function

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 + \lambda \cdot \sum_{j=1}^{m} w_j^2, \quad (5)$$

where $\hat{y}_i$ is the i^{th} predicted sample, y_i is the corresponding value for the i^{th} reference sample, n is the total number of samples, w_j is the j^{th} model coefficient, m is the total number of weights and $\lambda = 10^{-4}$ is the weighting factor for the l_2-norm term. At each iteration, the training loss value achieved for a given batch is used in combination with the backpropagation algorithm to compute the loss gradient with respect to each trainable model parameter, and it is used in conjunction with the Adam optimizer to update their values [36]. This process is repeated multiple times across every

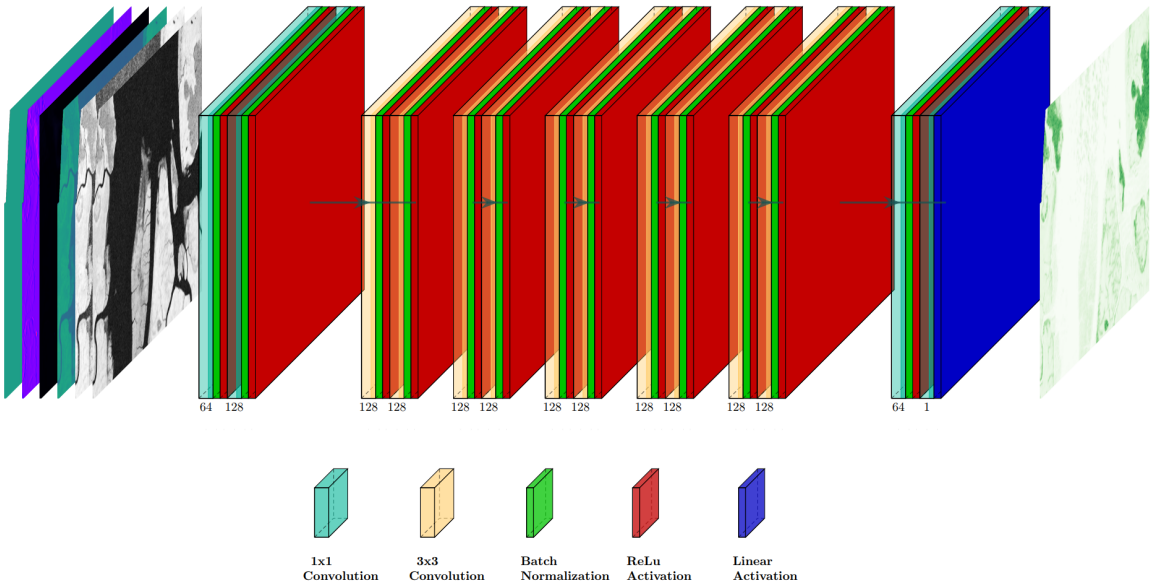


Fig. 2: The proposed fully convolutional deep learning model. The subscript numbers indicate the number of kernel filters.

training sample and monitored through the co-evolution of the independently computed validation loss. An early-stopping heuristic is adopted to determine the stagnation of the training process, allowing for optimizing both training time and model generalization. Overall, $18 \cdot 10^6$ patches are available for training and another $1.2 \cdot 10^6$ are used for validation. Training is performed on a single NVIDIA A100 40GB GPU, lasting from a minimum of 1 up to a maximum of 3 hours to complete, depending on the considered predictor set and on the random parameter initialization.

At inference time, the trained model is applied to each TanDEM-X product in its entirety and is tested on the areas that intersect with the available LVIS data. An example of the TanDEM-X input features and the corresponding deep learning-predicted canopy height map can be seen in Fig. 3. The performance of each trained model is evaluated on a diverse suite of metrics, necessary to correctly highlight potential issues in the heterogeneous testing conditions. This consists of the mean error (ME), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the root mean squared error (RMSE) and the coefficient of determination (R^2), which are defined as follows

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right), \tag{6}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|, \tag{7}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i + \epsilon}\right|, \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}, \tag{9}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2}, \tag{10}$$

where $\epsilon = 10^{-10}$ is added to avoid divisions by zero. Each of the experimental setups under investigation is re-trained for three times from randomly sampled initializations, such as to capture the effects of inter-run variance on performance.

C. Explainable AI Methods

This section introduces the XAI methods used for the investigation into the decision-making process of the model. Let the proposed CNN model f be a function

$$f : \mathbb{R}^{H\times W\times C} \rightarrow \mathbb{R}, \tag{11}$$

with $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_C)$ the set of input features and $\hat{y} = f(\mathbf{X})$ the predicted output feature, where H and W represent the dimensions of the input context and C is the number of input channels or features. Then a (generic) explanation method Ψ_i attributes a measure of contribution or relevance to the generic input feature $\mathbf{x}_i$ for a given prediction $f(\mathbf{X})$. In this work, we consider:

a) Saliency Maps: In order to capture the spatial relationships learned by the model, we propose the generation of saliency maps that quantify the impact of the predictors' pixels on the model output. The most straightforward and interpretable approach is based on the backprogagation of the gradient of the prediction $f(\mathbf{X})$ with respect to the input feature $\mathbf{x}_i$ [37], such that

$$\Psi_{\mathrm{i}}^{\mathrm{grad}} = \nabla_{\mathbf{x}_i} f(\mathbf{X}) = \left(\frac{\partial f(\mathbf{X})}{\partial x_{1,1,i}}, \ldots, \frac{\partial f(\mathbf{X})}{\partial x_{H,W,i}}\right), \tag{12}$$

which highlights the sensitivity of the output value to a (signed) perturbation for each value in the input, as well as the position of these values within the considered spatial context.

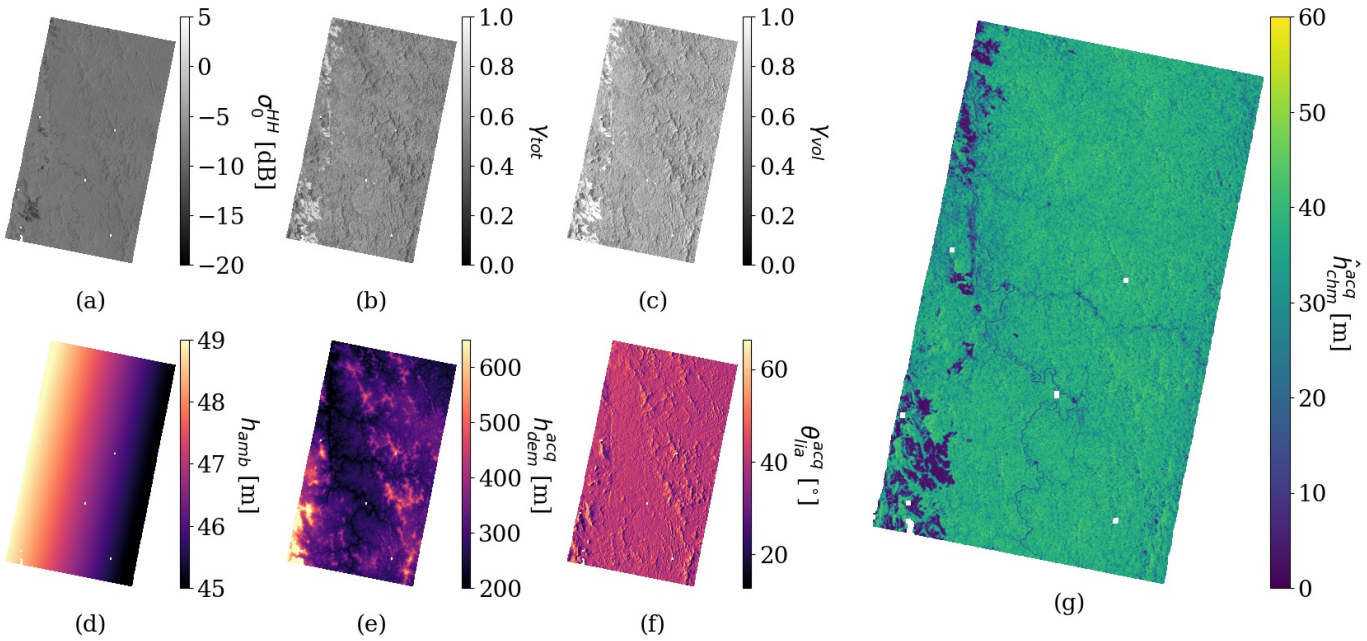


Fig. 3: Example of TanDEM-X-derived input features (a-f) and corresponding deep learning-derived canopy height prediction map (g). The original acquisition has been acquired in stripmap mode in HH polarization over the study area of Lopé, and has Acquisition Item ID (AIID) 1010159. The missing values (patches in white) correspond to invalids that have been masked in the input data and propagated to the prediction.

A natural extension to this idea is represented by the element-wise product with the input itself [38], such that

$$\Psi_{\mathrm{i}}^{\mathrm{grad}\times\mathrm{input}} = \mathbf{x}_i \odot \nabla_{\mathbf{x}_i} f(\mathbf{X}), \tag{13}$$

which effectively forms the linear approximation (minus the bias term) to the model's output. This representation is preferable over the former, as it uniformly expresses all sensitivity maps in the same units and removes the dependency on a given feature's dynamic range. Even though more advanced strategies exist in the literature, such as *Deconvnet* [15], *Guided Backpropagation* [19], and *Grad-CAM* [20], we favor the simpler *grad×input* approach as it holds a more straightforward and potentially less misleading mathematical interpretation [21]. We are also aware of the reported *saturation* problems of the approach [38], given by the fact that the gradient only expresses the slope of the linear approximation to the model, but does not recover its intercept, which limits one to capture the sensitivity to changes to the input, but not its absolute relevance. The *Integrated Gradient* approach [18] promises to solve this shortcoming by approximating the line integral of the derivative, but comes at a much greater computational cost. For this reason, we prefer to rely on the gradient-based approximation for the spatial sensitivity analysis and complement it with a more robust feature importance analysis, which numerically quantifies the contribution of each input feature.

b) Feature Importance: This analysis has the objective of highlighting the contribution of each predictor to the performance of the model. Since it is not possible to selectively remove inputs from a trained DL model, each feature contribution has to be evaluated by training a dedicated model. This process can be computationally unfeasible, leading perturbation-based approximations to be commonly used instead [17] [39]. Unfortunately, these schemes can lead to the generation of "unnatural" samples, forcing the model to operate in OOD regimes [18].

In this work, we take advantage of the limited parameter count of the proposed model, which allows for using independently trained models for each feature permutation. From these, it is possible to numerically assess the impact of including (or omitting) each predictor $\mathbf{x}_i$ from the set of input features $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_C\}$.

Firstly, we define the *single-feature* performance for the ith-feature as

$$\Psi_i^{\mathrm{sf}} = \Omega(f(\mathbf{x}_i), y), \tag{14}$$

where Ω is a generic performance metric for the model prediction $\hat{y} = f(\mathbf{x}_i)$ with respect to the reference value y.

We also assess the *leave-one-feature-out* performance as

$$\Psi_i^{\mathrm{lofo}} = \Omega(f(\mathbf{X}), y) - \Omega(f(\mathbf{X} \setminus \{\mathbf{x}_i\}), y), \tag{15}$$

where $\mathbf{X} \setminus \{\mathbf{x}_i\}$ is the subset of all input features excluding $\mathbf{x}_i$.

Finally, we highlight the mean performance variation between the add-one-in and the leave-one-out cases across all permutations by computing the Shapley values [40] for each feature. Instead of an equally weighted average, the Shapley value of a feature represents the weighted average of that feature's marginal contribution across all possible subsets. The particular weighting scheme follows a combinatorial formula

derived from cooperative game theory, such that

$$\Psi_i^{\text{shapley}} = \sum_{\mathbf{S}\subseteq X\setminus\{\mathbf{x}_i\}} w_{\mathbf{S}} \cdot [\Omega(f(\mathbf{S}\cup\{\mathbf{x}_i\}), y) - \Omega(f(\mathbf{S}), y)]\,, \tag{16}$$

where $w_{\mathbf{S}}$ is the weighting coefficient for the subset $\mathbf{S}$ and is defined as

$$w_{\mathbf{S}} = \frac{|\mathbf{S}|!\cdot(|\mathbf{X}|-|\mathbf{S}|-1)!}{|\mathbf{X}|!}, \tag{17}$$

where $|F|$ for the generic set F indicates its size, and $n!$ is the factorial of n.

IV. Experiments and Results

In this section, we present extensive numerical analyses aimed at evaluating the interpretability of the model defined in III-A, and the role of its six baseline predictors: the backscatter σ_0^{HH}, the real-valued interferometric coherence γ_{tot}, the volumetric decorrelation factor γ_{vol}, the height of ambiguity h_{amb}, the acquisition raw DEM $h_{\text{dem}}^{\text{acq}}$, and the local incidence angle $\theta_{\text{lia}}^{\text{acq}}$ computed using $h_{\text{dem}}^{\text{acq}}$.

This amounts to a total of 64 unique input feature permutations, for each of which three models are independently trained, validated and tested, according to the methodology described in III-B. We also set up additional experiments wherever a more detailed analysis is required.

A. Kernel Size

Arguably one of the most significant theoretical advantages of CNN-based methods is the ability to integrate information about the spatial context into the modeling. The architecture described in Sec. III-A makes use of 2D kernels, which mathematically allow the model to access the values of neighboring pixels. To assess the significance of this capability, we separately train two versions of the proposed models: the baseline model, which in the hidden layers uses kernels of size 3×3 pixels, and a second model, which replaces these with kernels of size 1×1 pixel. The former architecture results in a theoretical receptive field of 21×21 pixels, whereas the latter has a unitary receptive field. The comparative performance results for the two models are reported in Tab. I, and highlight that the omission of contextual information leads to a significant loss in accuracy. This can also be observed in the visual example in Fig. 4, which shows that the 1×1-kernel model fails to predict any substantial spatial pattern in the vegetation height, as it instead predicts the mean elevation height seen during training. In contrast to that, the latter 3×3-kernel model correctly manages to capture the presence of taller vegetation in the valleys and in presence of water surfaces.

B. Saliency Maps

Having established the benefits of a context-aware model, we now assess the spatial co-dependencies between the input features to the model and its outputs, based on the computation of saliency maps according to Eq. 13. In particular, the analysis focuses on the effects of each predictor map on a single output pixel, as the spatial contributions would otherwise overlap and become impossible to separate. This computation serves to emphasize trends in the model decision-making process, as the interpretation of decisions for individual samples remains non-linear.

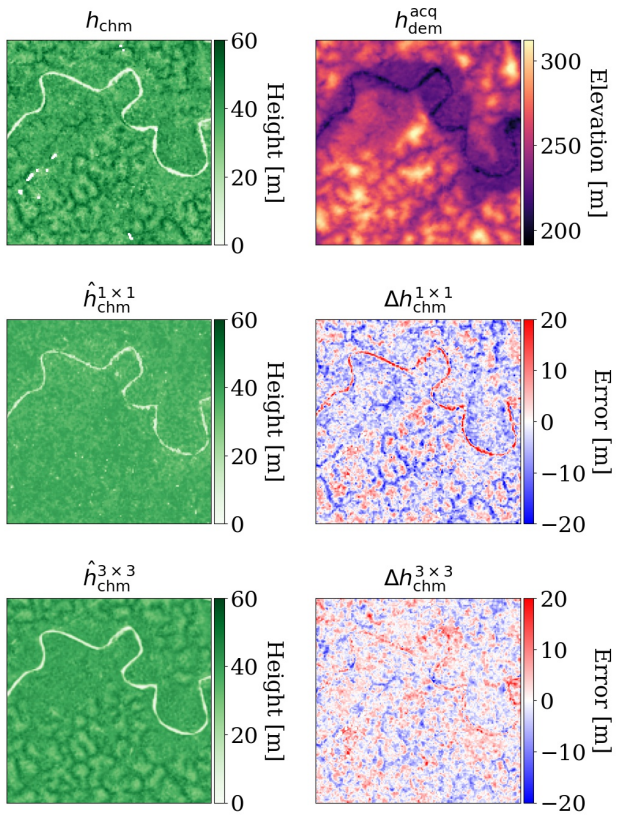


Fig. 4: Example of the impact of the CNN kernel size. The first row shows an example of an LVIS-derived CHM (top-left) and the corresponding TanDEM-X DEM (top-right). The second row shows the CHM predicted by the 1×1-kernel model (center-left) and the corresponding error (center-right). The third row shows the CHM predicted by the 3×3-kernel model (bottom-left) and the corresponding error (bottom-right).

TABLE I: Summary of the mean performance results for the 1×1- and the 3×3-kernel models averaged across three independent trainings.

Individual Feature Performance					
Experiment	**ME**	**MAE**	**MAPE**	**RMSE**	**R²**
	[m]	[m]	[%]	[m]	[-]
1×1	-2.11	5.15	17.31	6.76	0.62
3×3	-0.41	3.75	13.15	4.96	0.80

The results for a collection of validation patches are presented in Fig. 5. For each example, a centered, circular neighborhood of predictor pixels is considered according to differential patterns. Of particular interest is the example in the fifth row, corresponding to a sample of extremely low and isolated vegetation: a reduction in model sensitivity can be observed across all predictors, but in particular for $h_{\text{dem}}^{\text{acq}}$.

To better highlight global trends, the average feature saliency maps and the average of their magnitudes are shown in Fig. 6. Notably, the absolute values assume Gaussian-like distributions centered in the output pixel, which suggest a

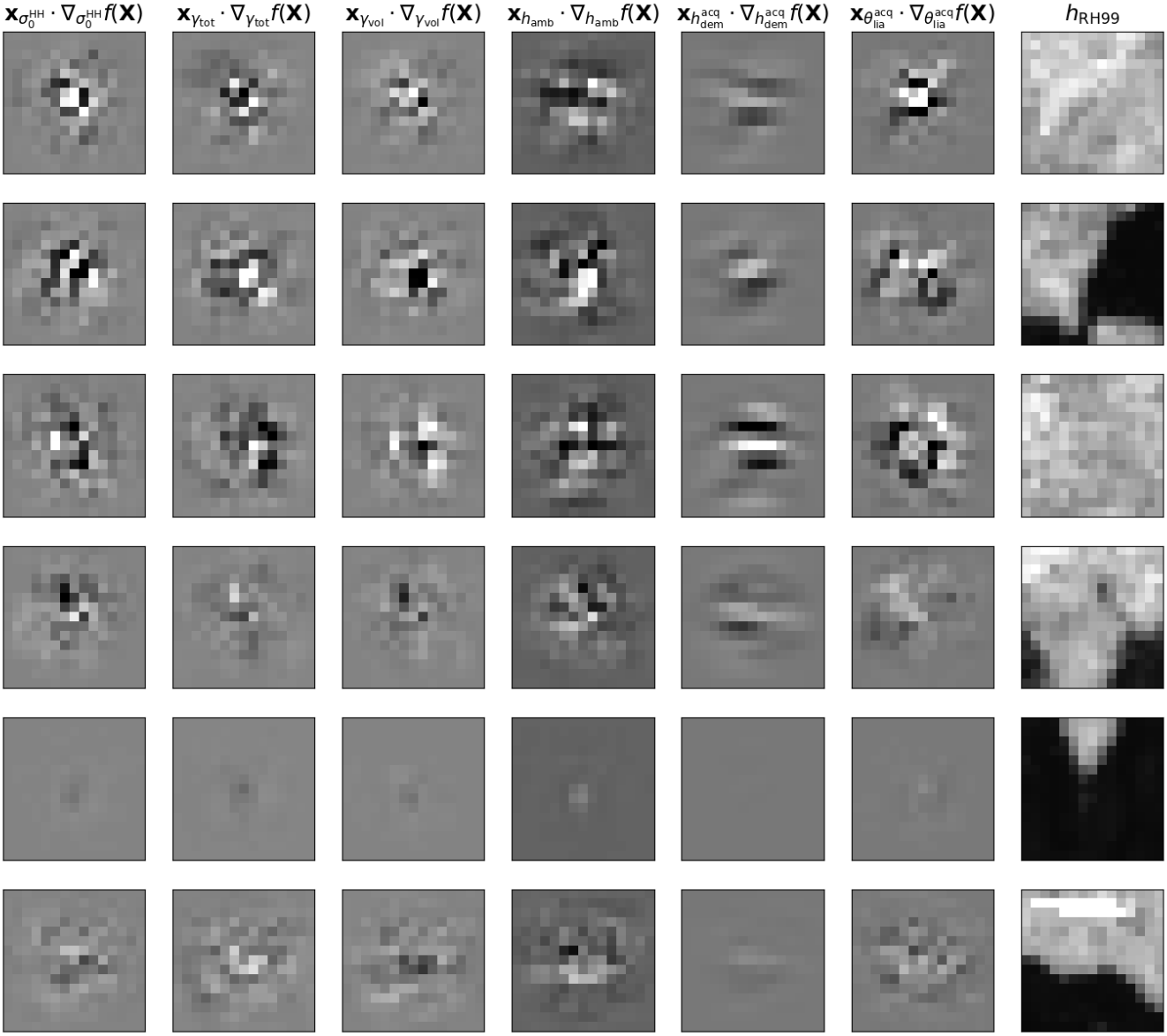
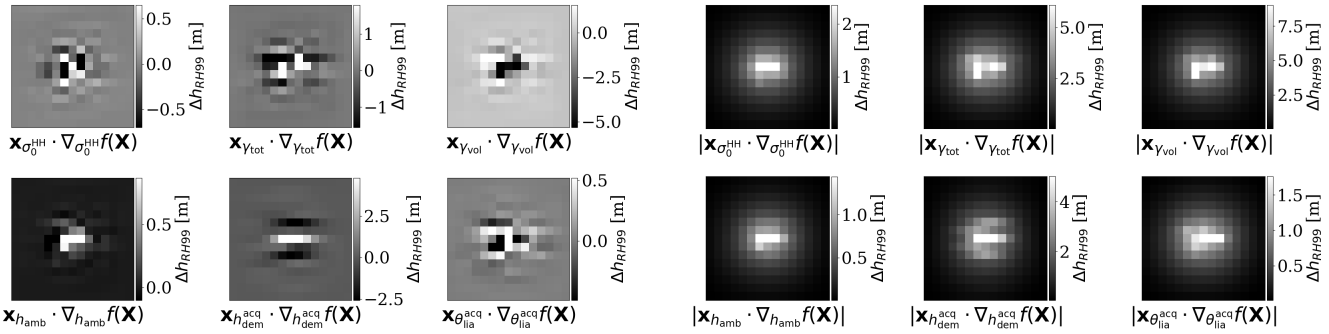


Fig. 5: A collection of saliency-map examples and their corresponding canopy height ground-truths. Each of the saliency maps is product of the partial derivative of the center output pixel, computed with respect to each of the input pixels, and the inputs themselves.



(a) Mean Saliency Maps (b) Mean Absolute Saliency Maps

Fig. 6: The mean (a) and mean absolute (b) feature saliency maps computed across all validation samples.

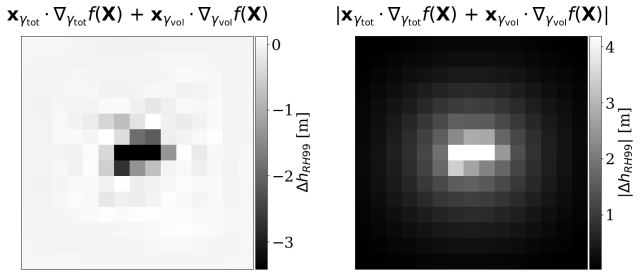


Fig. 7: The mean (top) and corresponding mean absolute (bottom) saliency maps for pixel-wise sum of the γ_{tot} and γ_{vol} sensitivity maps.

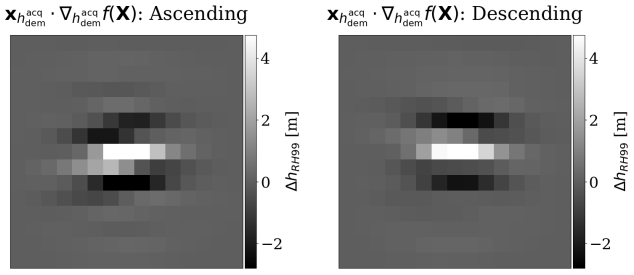
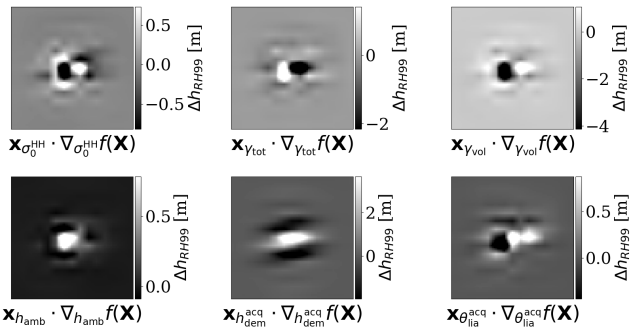
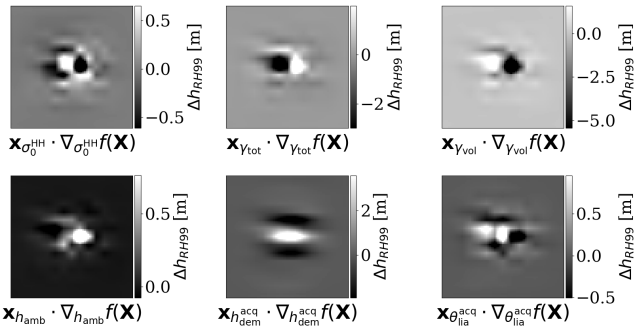


Fig. 8: The mean saliency maps for the DEM predictor for ascending (top) and descending (bottom) acquisitions.



(a) Ascending



(b) Descending

Fig. 9: The mean (a) and mean absolute (b) feature saliency maps computed across all validation samples. The results have been linearly upsampled to better highlight their relationships.

rapidly decaying relevance of the predictors as the radial distance to the estimated pixel increases. This central area of highest sensitivity defines the *effective receptive field* of the model [41], namely the spatial context that the model considers for its prediction. The saliency maps for γ_{tot} and γ_{vol} show similar sensitivity patterns, but with opposite signs. This is made more evident by their pixel-wise summation: Fig. 7 shows the averaged result across all validation samples and highlights how the energy of the resulting saliency map is focused on the location of the predicted output pixel, with a net inversely proportional relationship. This complementarity between the saliency maps for γ_{tot} and γ_{vol} suggests that the model might be relying on a direct comparison between the two features to produce a more robust prediction. The saliency map for $h_{\text{dem}}^{\text{acq}}$ (Fig. 6) consists of parallel segments of increased sensitivity, arranged in an alternating pattern reminiscent of 2D Gabor filters, which are typically used in image processing to detect directional frequency contents and are hypothesized to be at the heart of the mammalian vision system [42]. This suggests a directional sensitivity to vertically occurring changes in $h_{\text{dem}}^{\text{acq}}$. Given the small context window, any detected variation in the DEM is necessarily a high-frequency spatial component, thus excluding any slow-varying topographic features. Such highly varying patterns might instead be connected to transition areas between forested and non-forested areas, or to significant variations in mean-phase center height over forested areas.

Interestingly, the preference for the North-South direction is found across all three independent training runs (i.e. all models find the same relationship), suggesting a deeper connection with the input data. We evaluate a potential dependency on the near-polar orbit direction and the side-looking acquisition geometry of the TanDEM-X mission by computing the saliency maps for ascending and descending acquisitions, separately. Indeed, the results in Fig. 8 highlight rotations in the patterns - counterclockwise for the ascending pattern, and clockwise for the descending one - that are compatible with the estimated off-axis rotations of $-11.3°$ and $+11.3°$ for the TanDEM-X ascending and descending orbits, respectively.

A left-right symmetry can also be found in the remaining saliency maps, as it can be observed particularly clearly in the bilinearly upscaled patterns shown in Fig. 9.

C. Baseline Feature Importance Analysis

To capture the relevance of each predictor, three performance evaluations are presented according to the importance analysis strategies discussed in Sec. III-C. All evaluation results are numerically detailed in the appendix (Sec. VI).

a) Single-Feature Importance: The first experiment, reported in Fig. 10, evaluates the information content independently carried by each single feature.

It is possible to observe that γ_{tot} and γ_{vol} alone perform sensibly better than σ_0^{HH}: the scatterplots in Fig. 11 suggest that the coherence-related features are more effective at separating vegetated from non-vegetated samples, while σ_0^{HH} better manages to distinguish between different canopy height ranges.

$h_{\text{dem}}^{\text{acq}}$ significantly outperforms all other predictors, while the local incidence angle map $\theta_{\text{lia}}^{\text{acq}}$ performs the second best,

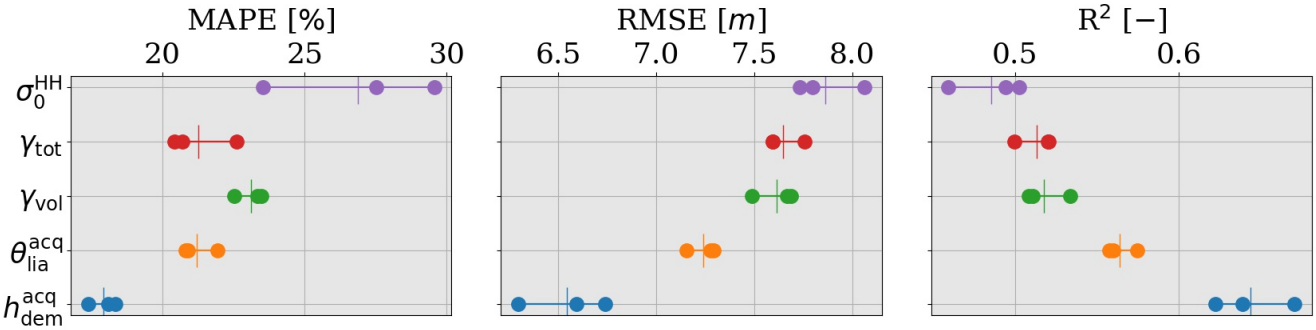
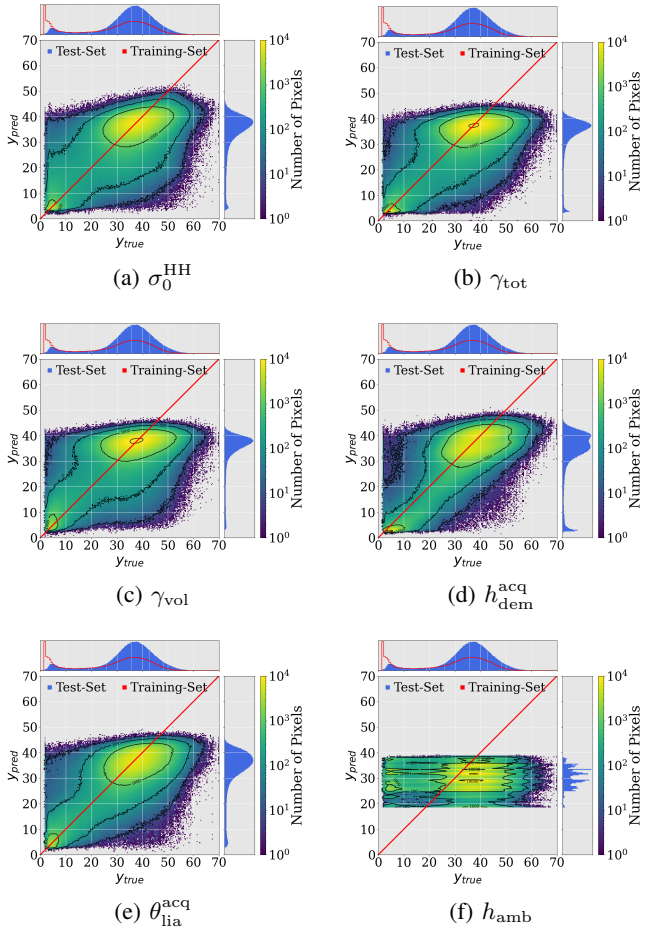


Fig. 10: Distribution of the performance metrics for each model trained on only one of the six baseline predictors.



(a) σ_0^{HH} (b) γ_{tot}

(c) γ_{vol} (d) h_{dem}^{acq}

(e) θ_{lia}^{acq} (f) h_{amb}

Fig. 11: Scatterplots for each of the models trained on only one of the six baseline predictors.

possibly due to its direct dependence on h_{dem}^{acq}. As mentioned, h_{dem}^{acq} is positively biased and results in spatially varying patterns in the presence of vegetation, which could be good indicators of canopy height.

Finally, h_{amb} performs the worst, given that it does not hold any information about the target: Fig. 11 (f) shows that the model learns to predict only a narrow range of forest height values corresponding to the largest modality found in the training set. This is expected, as predicting the most occurring value is the best option to minimize training loss in the absence of target information.

TABLE II: The shapley values for each of the model's predictors, expressed for different performance metrics.

Shapley Values					
Experiment	**ΔME**	**ΔMAE**	**ΔMAPE**	**ΔRMSE**	**ΔR²**
	[m]	[m]	[%]	[m]	[-]
σ_0^{HH}	1.67	-1.52	-6.90	-1.60	0.24
γ_{tot}	1.32	-1.50	-8.37	-1.60	0.24
γ_{vol}	1.51	-1.51	-7.99	-1.61	0.24
h_{amb}	0.63	-0.50	-1.09	-0.43	0.06
h_{dem}^{acq}	1.16	-1.89	-11.91	-2.09	0.29
θ_{lia}^{acq}	1.44	-1.49	-7.79	-1.59	0.24

b) Leave-One-Feature-Out Performance Loss: The contribution of individual features to the baseline model is assessed through a leave-one-feature-out analysis according to Eq. 15. The resulting performance losses are reported in Fig. 12.

The h_{dem}^{acq} remains by far the most impactful feature, as evidenced by the significant performance loss across all metrics and especially for the MAPE one.

In contrast, the impact of each coherence feature is shown to become negligible and suggests a fundamental redundancy in the information provided by γ_{tot} and γ_{vol}. We remark that γ_{vol} is estimated from γ_{tot} by compensating the latter for additional decorrelation sources, such as γ_{quant} and γ_{snr}. γ_{tot} still provides a sensible improvement in terms of MAPE, which corresponds to an improvement in the estimation of shorter tree stands.

Also of interest is the insignificant impact on the performance of h_{amb}, as it appears to become redundant in the presence of other predictors.

c) Shapley Values: To estimate the overall importance of each feature, we compute their Shapley values according to Eq. 16. Each resulting value represents the fractional contribution of that specific predictor to the total performance when going from a model based on an empty set of features to that with the complete set of features. The corresponding performance results for each feature are summarized in Tab. II. As it can be seen, h_{dem}^{acq} remains by far the most relevant

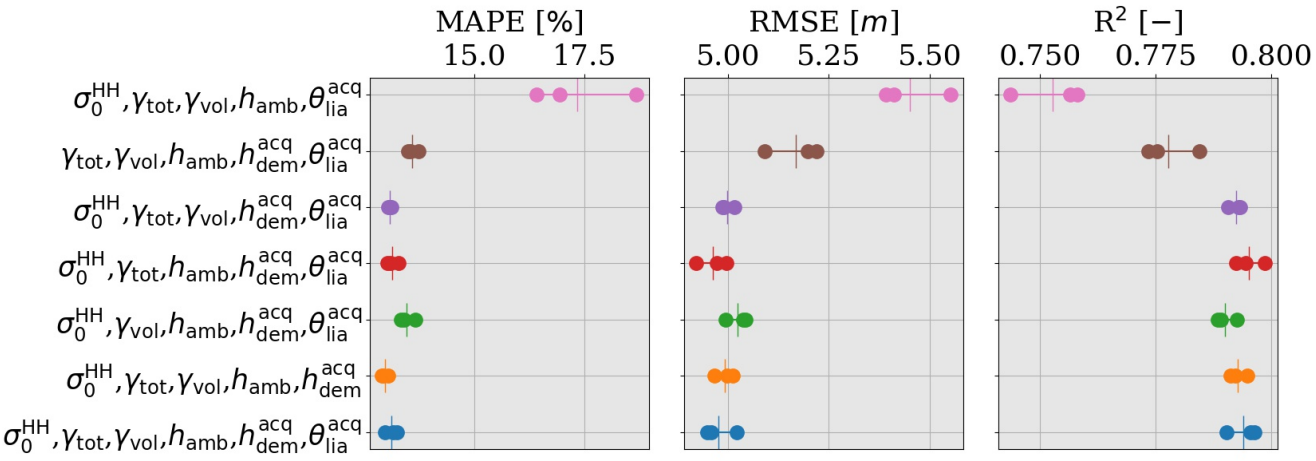


Fig. 12: The performance metrics for the leave-one-feature-out analysis. Each model is trained on all-but-one of the six baseline predictors.

feature to explain the model performance, with σ_0^{HH}, γ_{tot}, γ_{vol} and $\theta_{\mathrm{lia}}^{\mathrm{acq}}$ achieving very similar relevance with respect to each other. The only metric with an interesting behavior is the MAPE, which sees σ_0^{HH} performing worse and γ_{tot} performing sensibly better.

On the basis of these results, we note that the assumption of independence between all predictors is not satisfied. Therefore, more detailed analyses are required to understand their specific role.

D. Interferometric Phase Importance Analysis

Given the particular relevance of $h_{\mathrm{dem}}^{\mathrm{acq}}$, this section presents more experimental results focused on the role of the acquisition DEM and its precursor, the interferometric phase ϕ_{acq}. As discussed in Sec. II-B, the use of the raw acquisition DEM stems from previous analyses in [35] and [3], which highlighted the importance of temporal consistency between all predictors.

a) Phase Unwrapping: An interesting aspect to consider is the impact of the phase unwrapping process on the regression problem. It is reasonable to assume that the conversion of a relative (i.e., ambiguous) height measurement into an absolute one might provide relevant contextual information to the model (e.g., about different climatological and elevation-dependent changes in the phenology).

To test this assumption, we first convert $h_{\mathrm{dem}}^{\mathrm{acq}}$ to the absolute phase and then wrap it to once again represent the topography in terms of an ambiguous phase term. For this, we consider two different strategies: wrapping it with a period of h_{amb} and wrapping it with a globally defined constant period (i.e., $h_{\mathrm{amb}}^{const} = 100\,\mathrm{m}$). The first strategy allows one to revert the DEM back to the geocoded interferometric phase ϕ_{acq}^{amb} (excluding potential changes introduced by the unwrapping process itself), while the second one is selected to potentially decouple the model's regression performance from the acquisition geometry through the definition of ϕ_{acq}^{const}. An example for $h_{\mathrm{dem}}^{\mathrm{acq}}$ and its corresponding ϕ_{acq}^{amb} and ϕ_{acq}^{const} representations is shown in Fig. 13.

It is noted that, by definition, phase jumps are encountered when moving between two adjacent phase cycles. Therefore, each of the aforementioned wrapped phase features will be also expressed in terms of a $\{\sin\phi_{acq}, \cos\phi_{acq}\}$ pair, which removes these jumps and preserves the unambiguous direction of phase growth. The results for each of the four experiments are reported in Fig. 14.

Compared to the baseline result, a small performance degradation can be observed for the $\{\sin\phi_{acq}^{const}, \cos\phi_{acq}^{const}\}$ pair. This outcome meets the expectation that the wrapped phase should not contain additional information compared to $h_{\mathrm{dem}}^{\mathrm{acq}}$, but loses the information about the absolute topographic elevation.

Interestingly, this outcome is not repeated for ϕ_{acq}^{amb}, as sensible performance gains can be observed over the baseline, especially for the $\{\sin\phi_{acq}^{amb}, \cos\phi_{acq}^{amb}\}$ pair. A possible explanation for the improved performance could lie in the fact that a period-correct wrapping of $h_{\mathrm{dem}}^{\mathrm{acq}}$ would, by definition, eliminate any phase unwrapping errors, which would survive for ϕ_{acq}^{const}. The comparatively poor performance for $\{\sin\phi_{acq}^{const}, \cos\phi_{acq}^{const}\}$ could also support this thesis: sudden *jumps* in the predictors are not expected, and when phase unwrapping errors occur, these could lead to unexpected predictions. This might not be an issue for the ϕ_{acq}^{const} case, which naturally includes phase jumps between contiguous phase cycles, and might thus offer some intrinsic robustness. Finally, we remark that, while the $\{\sin\phi_{acq}^{amb}, \cos\phi_{acq}^{amb}\}$ pair represents a robust choice against processing errors, it does not provide additional information on the regression problem.

b) Mean Phase Center Height: The performance results in Sec. IV-C have evidenced the primary role of the interferometric phase information, especially in terms of MAPE. It has also been shown that this role is preserved even when the interferometric phase information is presented in its wrapped state, as the absolute elevation appears to not hold any particular value over the AfriSAR sites. This behavior suggests the crucial role of the phase information in correctly distinguishing between forested and non-forested areas. A

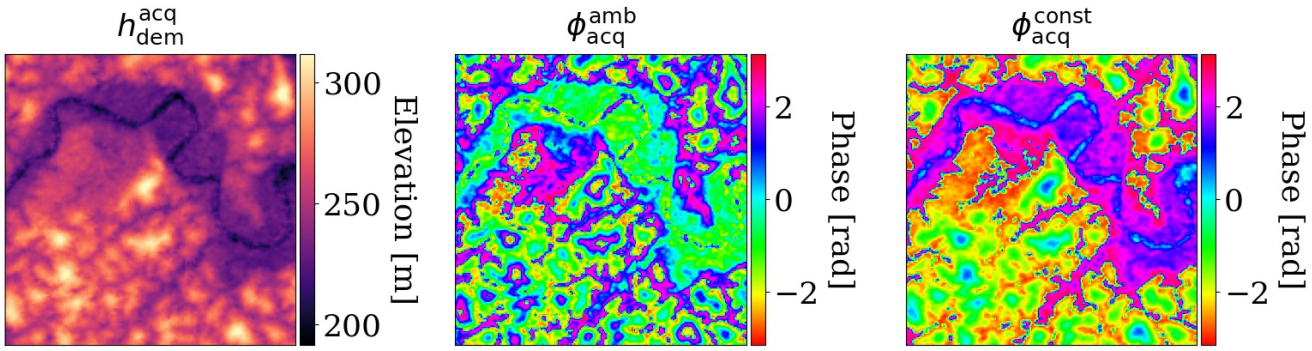


Fig. 13: An example of a TanDEM-X-derived DEM (left), and the corresponding phases obtain by wrapping the DEM with an acquisition dependent period of h_{amb} (center), and with an acquisition independent period of $h_{\text{amb}}^{const} = 100$ (right).

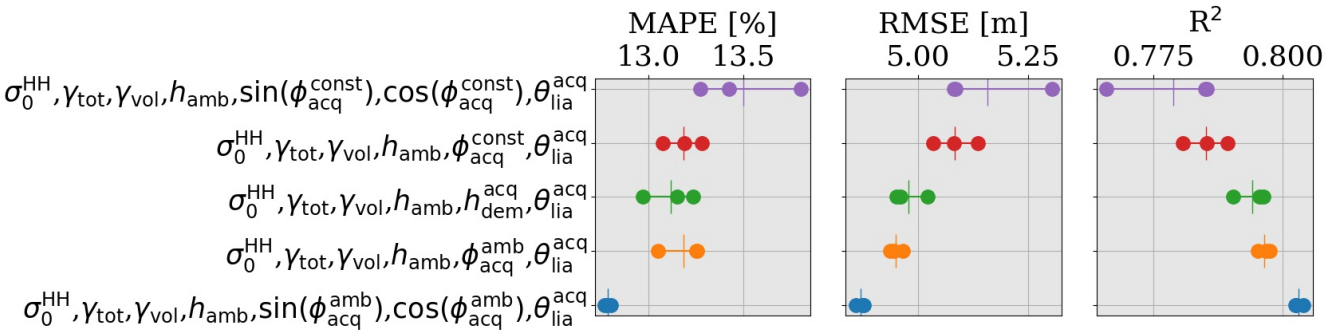


Fig. 14: The comparison of the performance metrics for the interferometric phase experiment. Each model uses all of the baseline predictors, but evaluates different means of representing the interferometric phase information.

possible explanation for this behavior might reside in the significant local variability of the mean phase center height over forested areas, especially compared to the relatively slow varying behavior of the underlying topography. To test this idea, the LiDAR-derived DTM h_{dtm}, produced in the context of the 2016 AfriSAR campaign, is investigated as a possible drop-in replacement for $h_{\text{dem}}^{\text{acq}}$. Furthermore, also the $h_{\text{dem}}^{\text{acq}}$-derived local incidence angle map ($\theta_{\text{lia}}^{\text{acq}}$) is compared to that derived from h_{dtm} ($\theta_{\text{lia}}^{\text{dtm}}$), as to completely remove any dependencies on the mean phase center variability. As it can be seen in Fig. 15, there are clear differences between the distributions for $h_{\text{dem}}^{\text{acq}}$ and $\theta_{\text{lia}}^{\text{acq}}$, with those for h_{dtm} and $\theta_{\text{lia}}^{\text{dtm}}$. It is also possible to visually observe the high local variability in $h_{\text{dem}}^{\text{acq}}$, which effectively manifests itself as a noise-like behavior that is propagated to $\theta_{\text{lia}}^{\text{acq}}$. Finally, the distributions for h_{dsm} and $h_{\text{dem}}^{\text{acq}}$ highlight a relative elevation offset, which corresponds to the penetration bias of the X-band radar wave into the canopy.

Looking at the performance metrics reported in Fig. 16, the results for the individual predictors confirm that no information about forest height can be recovered from h_{dtm} or $\theta_{\text{lia}}^{\text{dtm}}$, which perform significantly worse on their own than their InSAR counterparts. This is also confirmed by the results for the complete predictor set, for which the models replacing $h_{\text{dem}}^{\text{acq}}$ with h_{dtm} maintain similar results in terms of RMSE and R^2, but perform measurably worse in terms of MAPE. This behavior reflects a greater difficulty in estimating short vegetation, providing another clear indication of the relevance of the interferometric phase (and by extension $h_{\text{dem}}^{\text{acq}}$) in providing a means to separate short vegetation from open terrain.

Interestingly, replacing $\theta_{\text{lia}}^{\text{acq}}$ with $\theta_{\text{lia}}^{\text{dtm}}$ leads to a significant improvement in performance over the baseline. As can be seen in Fig. 15, relying on $h_{\text{dem}}^{\text{acq}}$ for the local incidence angle computation introduces a lot of noise that is related to the incorrect estimation of the normal to the surface. This evidence suggests that selecting the correct DEM is critical for optimal SAR pre-processing (e.g., geometric and radiometric terrain correction, shadow/layover estimation, local incidence angle computation) [43].

Finally, by combining the information provided by $h_{\text{dem}}^{\text{acq}}$ and h_{dtm} it is possible to further simplify the inversion problem, effectively providing a (measured) lower boundary to the estimated canopy height.

E. Interferometric Coherence Importance Analysis

This section focuses on the relevance of γ_{tot}, γ_{vol} and h_{amb}, on their co-dependencies and potential redundancies. The performance results presented in Fig. 17 depict the relationship

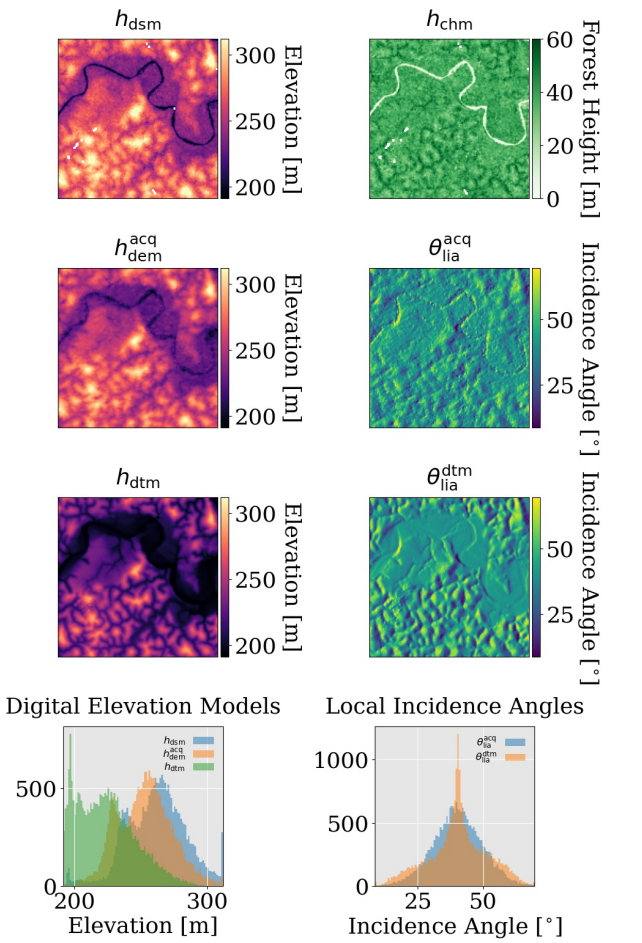


Fig. 15: The first row shows an example of LVIS-derived DSM (left) and CHM (right). The second row shows the TanDEM-X-derived DEM (left) and the corresponding LIA (right). The third row shows the LVIS-derived DTM (left) and the corresponding LIA (right). The fourth row shows the distributions for the LVIS-derived DSM and DTM, and the TanDEM-X derived DEM (left), and the distribution for the LIA derived from the DTM and the DEM (right).

between these three features in absence of other predictors. It can be seen that γ_{tot} and γ_{vol} lead to similar results, with γ_{tot} leading to measurably better MAPE scores, while models using γ_{vol} perform sensibly better in terms of RMSE and R^2. As expected, the addition of h_{amb} provides the necessary contextual information, helping both features to better separate ambiguous vegetation contributions from those caused by the interferometric baseline. The combination of γ_{tot} and γ_{vol} leads to an even more notable improvement, which is further enhanced by the addition of h_{amb}. From the results in Fig. 12, it is possible to observe that the information provided by either γ_{tot}, γ_{vol} and h_{amb} becomes redundant in the presence of the remaining predictors.

V. Discussion

The experiments presented in Sec. IV offer detailed insights into the inner workings of the proposed deep learning-based framework for the regression of forest height from spaceborne InSAR data. In particular, the developed fully-convolutional baseline architecture models the canopy height from a set of SAR, InSAR and ancillary features, all derived from only one single-pass TanDEM-X SAR acquisition. The relationship between these physics-aware predictors and the forest height variable is inferred from the labeled training data, by teaching the model to sequentially capture increasingly complex patterns.

By design, the prediction for a single pixel is based on the information context that spans across all predictors and over a local spatial neighborhood of pixels. The experiments on the CNN's kernel size, presented in Sec. IV-A, measured the contribution offered by this spatial context. The results for the 3×3-kernel model showcased improvements over the 1×1-kernel model of 80.6% in ME, of 36.3% in RMSE, of 24.0% in MAPE, and of 29.0% in R^2, respectively, emphasizing the key relevance of the spatial context.

In Sec. IV-B we assessed the nature of these spatial patterns through the computation of gradient-based saliency maps, indicating the sensitivity of the model's prediction to each input pixel. The results show that the model's attention to the input has a radial symmetry that rapidly decays to zero beyond a distance of 100 m from the predicted pixel, suggesting that a relatively small contextual window might be sufficient to correctly inform the model. This could represent a limit to the usefulness of larger models (e.g., U-Nets and ResNets) or of general purpose architectures (e.g., multi-layer perceptrons and visual transformers). We remark that this observation depends on the specific data source, the observed phenology and the pixel spacing used in this study, and it might not apply for different contexts.

In the analysis in Sec. IV-B, we observed that the saliency patterns for the inputs to the baseline model are related to the SAR acquisition geometry, since left-right symmetries emerged for the saliency maps obtained for ascending and descending acquisitions, separately. Furthermore, the saliency map for $h^{\mathrm{acq}}_{\mathrm{dem}}$ shows a clear sensitivity to changes occurring along the sensor flight direction. The model is arguably capable of correctly inferring the differences in acquisition geometries related to ascending and descending orbits and it learns to model the spatial relationships within the predictors accordingly. The independence of deep learning models from the SAR orbit direction was already observed in [35], but no empirical evidence had been found up to this point.

In Sec. IV-C we presented an extensive feature importance analysis to infer the direct influence of each predictor, as well as of any potential co-dependencies, on the canopy height regression performance.

Across all tested input permutations, the interferometric features (γ_{tot}, γ_{vol}, $h^{\mathrm{acq}}_{\mathrm{dem}}$) form the most informative set of predictors, with σ^{HH}_0 and $\theta^{\mathrm{acq}}_{\mathrm{lia}}$ remaining extremely relevant, and h_{amb} resulting in only limited gains. More in detail, the differences in feature importance found between the single-feature and the leave-one-feature-out analyses suggest that some of the features might offer relevant information to the model when the number of predictors is limited, but become increasingly obsolete when complementary sources of information are added. A relevant example is the apparent redundancy introduced by the two coherence terms γ_{tot} and

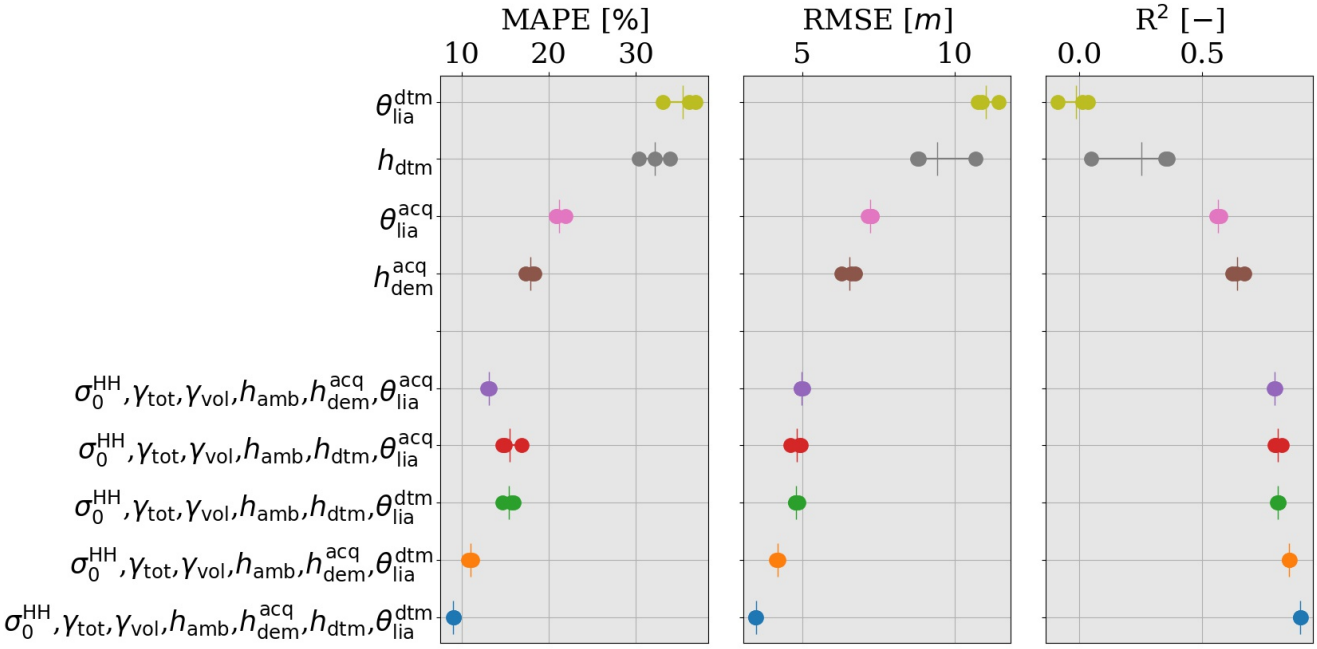


Fig. 16: The comparison of the performance metrics for the DTM experiment. The information content of the LVIS-derived h_{dtm} and its corresponding $\theta_{\text{lia}}^{\text{dtm}}$, is compared against that of the TanDEM-X-derived $h_{\text{dem}}^{\text{acq}}$ and its corresponding $\theta_{\text{lia}}^{\text{acq}}$.

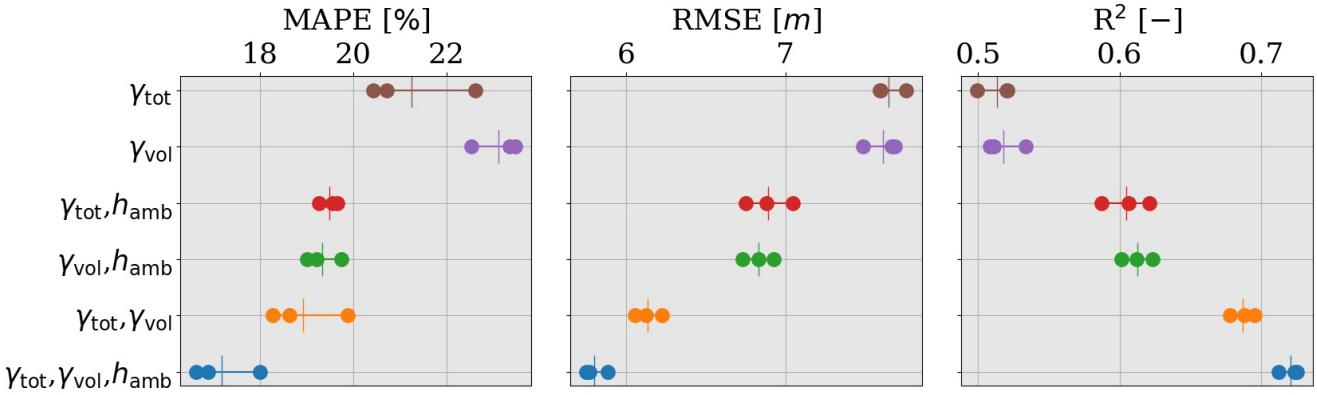


Fig. 17: The comparison of the performance metrics for the InSAR coherence experiment. The analysis highlights the co-dependencies between γ_{tot}, γ_{vol} and h_{amb}.

γ_{vol}, since it follows from Eq. 3 that γ_{vol} is a factor of γ_{tot}. The results in Fig. 12 appear to imply that either could be removed without significant alterations to the overall performance, while the results reported in Tab. III suggest that their combination offers improvements in either a feature individually whenever one among σ_0^{HH}, $h_{\text{dem}}^{\text{acq}}$, or $\theta_{\text{lia}}^{\text{acq}}$ is missing. Surprisingly, this observation holds even if h_{amb} is present, but it falls short when σ_0^{HH} is introduced alongside h_{amb}. In general, σ_0^{HH} appears to provide a great complementary source of information for each of the interferometry-derived features, so that their combination might lead to a better separation of different acquisition characteristics and phenological contexts. As reported in Tab. IV, we additionally evaluated the role of γ_{snr} as a potential replacement for γ_{vol} (in the presence of γ_{tot}), and as an explicit contextualization for the range-dependent antenna pattern. We found that in practice γ_{snr} fails to provide any useful information to the model, and rather leads to a slight reduction in performance.

The interferometric phase (expressed in terms of $h_{\text{dem}}^{\text{acq}}$) represents by far the most informative feature, leading to the best performance scores, both individually and in combination with other predictors. In particular, there is strong evidence that $h_{\text{dem}}^{\text{acq}}$ is especially relevant for separating between the absence and the presence of low vegetation, as evidenced by the large gains in the MAPE as well as the scatterplots in Fig. 11. The specific mechanism by which this is achieved is made evident by the corresponding saliency map shown in Fig. 8: The model appears to be sensitive to (orbit-aligned) differential

patterns in the $h_{\mathrm{dem}}^{\mathrm{acq}}$, which are related to the high variability in the measured mean phase center height found for volumetric scatterers. This is indirectly also confirmed by the results presented in Fig. 16, where a drop-in replacement of $h_{\mathrm{dem}}^{\mathrm{acq}}$ with h_{dtm} (and consequently of $\theta_{\mathrm{lia}}^{\mathrm{acq}}$ with $\theta_{\mathrm{lia}}^{\mathrm{dtm}}$) produces similar RMSE and R^2 scores, but results in a significantly worse MAPE, implying that the prediction of low vegetation is disproportionally affected. The results presented in Fig. 16 also showed that replacing $\theta_{\mathrm{lia}}^{\mathrm{acq}}$ with $\theta_{\mathrm{lia}}^{\mathrm{dtm}}$ (while keeping $h_{\mathrm{dem}}^{\mathrm{acq}}$) can measurably improve performance. This aspect can again be explained by the large variability in mean phase center heights measured over forested areas, as these lead to an incorrect estimation of the normal to the topographic surface and, ultimately, to an incorrect computation of the local incidence angle value. Hence, adopting the ALS-derived h_{dtm} as a proxy for the real topography better captures the real acquisition geometry of the systems, leading to a less ambiguous interpretation of the measured features. Finally, we also showed that the combination of $h_{\mathrm{dem}}^{\mathrm{acq}}$, h_{dtm} and $\theta_{\mathrm{lia}}^{\mathrm{dtm}}$ represents the upper performance limit of what is achievable (with the proposed regression framework) with an injection of privileged information. More so, the remaining prediction errors are evidence of further inconsistencies in the inversion problem, hinting at residual modeling ambiguities (i.e. missing information) and residual noise contributions in the satellite and the reference data measurements.

VI. Conclusions

In this work, we presented an XAI perspective on the inner workings of a CNN-based model for forest height regression from TanDEM-X single-pass InSAR data. The proposed analyses focused on the role of the model's predictors and their spatial context. We computed gradient-based saliency maps to assess the relationship between each input feature and their spatial context with the predicted outputs, finding strong evidence in favor of the model being capable of recognizing the orbit and look direction of the InSAR instrument. A detailed feature importance analysis, comprising more than 300 independently trained models, allowed us to determine the role of each predictor and to discover new insights into the physics of the problem. We found the interferometric features to be the main performance drivers of the model, with the interferometric phase of particular relevance, especially for the correct discrimination of short or very sparse vegetation. The use of the X-band TanDEM-X raw DEM led to errors in the calculation of local incidence angle maps, since the measured mean phase center height is highly variable over forested areas, yielding an incorrect estimation of the surface normal direction and of the returned energy. Additionally, we showed that the joint use of the interferometric coherence and the volume decorrelation factor improves performance in the absence of further predictors, but becomes redundant when the full feature set is considered. The backscatter information also becomes relevant in combination with the interferometric features, offering a wider degree of separability between the forest height ranges.

In future work, we would like to further analyze the hidden projection spaces found by the model, relying on linear or distance-preserving approximations to help improve their interpretations. We are also interested in better understanding the impact of distribution shifts or out-of-distribution samples on the model's predictions, improving the understanding of the generalization capabilities of these algorithms. Of interest is also the inclusion of ancillary environmental factors into the analyses, such as time of day, day of year, and meteorological conditions, to assess if it is possible to infer these aspects from the predictors alone or whether the information has to be provided in addition for a correct estimate.

Appendix: Additional Tables

The following section presents the performance results for each predictor permutation considered in this work, each representing the average across three independent training, validation and test runs. In particular, Tab III summarizes the results for the feature importance analysis for the six baseline predictors, and Tab IV lists the results for the interferometric features.

References

[1] M. Denbina, M. Simard, and B. Hawkins, "Forest Height Estimation Using Multibaseline PolInSAR and Sparse Lidar Data Fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3415–3433, Oct. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8386434/

[2] N. Lang, K. Schindler, and J. D. Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sensing of Environment*, vol. 233, p. 111347, Nov. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719303669

[3] D. Carcereri, P. Rizzoli, L. DellÁmore, J.-L. Bueso-Bello, D. Ienco, and L. Bruzzone, "Generation of country-scale canopy height maps over Gabon using deep learning and TanDEM-X InSAR data," *Remote Sensing of Environment*, vol. 311, p. 114270, Sep. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425724002888

[4] P. Ploton, F. Mortier, M. Réjou-Méchain, N. Barbier, N. Picard, V. Rossi, C. Dormann, G. Cornu, G. Viennois, N. Bayol, A. Lyapustin, S. Gourlet-Fleury, and R. Pélissier, "Spatial validation reveals poor predictive performance of large-scale ecological mapping models," *Nature Communications*, vol. 11, no. 1, p. 4540, Sep. 2020. [Online]. Available: https://www.nature.com/articles/s41467-020-18321-y

[5] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, May 2019. [Online]. Available: https://www.nature.com/articles/s42256-019-0048-x

[6] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 103, p. 102520, Dec. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0303243421002270

[7] C. M. Gevaert, "Explainable AI for earth observation: A review including societal and regulatory perspectives," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, p. 102869, Aug. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1569843222000711

[8] G. Costa, A. Virgilio Monti Guarnieri, A. Parizzi, and P. Rizzoli, "Dani-net: A physics-aware deep learning framework for change detection using repeat-pass insar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–27, 2025.

[9] A. G. Wilson, "The Case for Bayesian Deep Learning," Jan. 2020, arXiv:2001.10995 [cs]. [Online]. Available: http://arxiv.org/abs/2001.10995

[10] M. Datcu, Z. Huang, A. Anghel, J. Zhao, and R. Cacoveanu, "Explainable, Physics-Aware, Trustworthy Artificial Intelligence: A paradigm shift for synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 11, no. 1, pp. 8–25, Mar. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10035918/

TABLE III: The performance results for the feature importance analysis, consisting of all 64 model permutations obtainable from the baseline predictor set.

Appendix					
Experiment	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
	[m]	[m]	[%]	[m]	[-]
$\emptyset$	-8.17	12.18	57.17	13.90	-0.53
σ_0^{HH}	-0.86	5.92	26.87	7.86	0.49
γ_{tot}	-1.85	5.80	21.26	7.65	0.51
γ_{vol}	-1.22	5.80	23.12	7.61	0.52
h_{amb}	-4.27	9.89	55.04	12.24	-0.24
$h_{\mathrm{dem}}^{\mathrm{acq}}$	-1.34	5.06	17.94	6.54	0.64
$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-1.50	5.51	21.22	7.24	0.56
σ_0^{HH},γ_{tot}	-0.37	4.73	19.86	6.14	0.69
σ_0^{HH},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.65	4.30	14.93	5.64	0.74
σ_0^{HH},h_{amb}	-1.42	5.92	24.80	7.90	0.48
σ_0^{HH},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.03	4.80	19.50	6.28	0.67
σ_0^{HH},γ_{vol}	-0.29	4.76	19.88	6.19	0.68
γ_{tot},γ_{vol}	0.04	4.71	18.93	6.13	0.69
γ_{tot},h_{amb}	-1.15	5.24	19.49	6.89	0.60
γ_{tot},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-1.02	4.67	16.01	6.15	0.69
γ_{tot},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.35	4.77	18.47	6.22	0.68
γ_{vol},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.26	4.80	19.33	6.26	0.67
γ_{vol},h_{amb}	-0.64	5.19	19.33	6.83	0.61
γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.85	4.53	15.60	5.96	0.70
h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.99	4.78	16.84	6.26	0.67
h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-1.64	5.42	20.87	7.10	0.58
$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.77	4.63	16.63	6.07	0.69
σ_0^{HH},γ_{vol},h_{amb}	-0.33	4.41	16.95	5.73	0.73
σ_0^{HH},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.38	4.29	14.99	5.64	0.74
σ_0^{HH},γ_{vol},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.17	4.50	19.70	5.81	0.72
σ_0^{HH},γ_{tot},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.36	4.45	19.64	5.73	0.73
σ_0^{HH},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.32	3.97	13.83	5.23	0.77
σ_0^{HH},γ_{tot},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.61	4.06	14.01	5.31	0.77
σ_0^{HH},γ_{tot},γ_{vol}	-0.07	4.56	19.51	5.92	0.71
σ_0^{HH},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.37	4.14	14.48	5.44	0.75
σ_0^{HH},γ_{tot},h_{amb}	0.01	4.40	17.43	5.70	0.73
σ_0^{HH},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.16	4.85	22.23	6.30	0.67
γ_{tot},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.37	4.37	15.40	5.70	0.73
γ_{tot},γ_{vol},h_{amb}	-0.21	4.45	17.17	5.80	0.72
γ_{tot},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.62	4.67	18.24	6.08	0.69
γ_{tot},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-1.38	4.60	15.64	6.00	0.70
γ_{tot},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.42	4.06	14.00	5.36	0.76
γ_{tot},γ_{vol},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.02	4.49	18.94	5.80	0.72
γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.50	4.30	15.06	5.66	0.73
γ_{vol},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.61	4.61	17.55	6.02	0.70
γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.99	4.41	15.28	5.80	0.72
h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.75	4.65	16.39	6.09	0.69
σ_0^{HH},γ_{tot},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.09	4.29	17.34	5.54	0.75
σ_0^{HH},γ_{tot},γ_{vol},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.60	4.45	20.39	5.75	0.72
σ_0^{HH},γ_{tot},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.43	3.84	13.16	5.06	0.79
σ_0^{HH},γ_{vol},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.02	4.25	17.34	5.50	0.75
σ_0^{HH},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.30	3.87	13.31	5.10	0.78
σ_0^{HH},γ_{tot},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.32	3.90	13.45	5.14	0.78
σ_0^{HH},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.13	4.04	14.33	5.31	0.77
σ_0^{HH},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.39	3.82	13.09	5.02	0.79
σ_0^{HH},γ_{tot},γ_{vol},h_{amb}	-0.23	4.28	16.46	5.57	0.74
σ_0^{HH},γ_{tot},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.12	3.85	13.43	5.07	0.79
γ_{tot},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.21	3.98	13.68	5.26	0.77
γ_{tot},γ_{vol},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.26	4.27	16.67	5.54	0.74
γ_{tot},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.61	4.20	14.73	5.52	0.75
γ_{tot},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.34	3.98	13.72	5.24	0.77
γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.61	4.24	14.78	5.58	0.74
σ_0^{HH},γ_{tot},γ_{vol},h_{amb},$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	0.05	4.20	17.34	5.45	0.75
σ_0^{HH},γ_{tot},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$	-0.27	3.78	12.96	4.99	0.79
σ_0^{HH},γ_{tot},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.14	3.77	13.13	4.96	0.80
σ_0^{HH},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.04	3.81	13.46	5.02	0.79
σ_0^{HH},γ_{tot},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.22	3.80	13.08	5.00	0.79
γ_{tot},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.30	3.93	13.59	5.17	0.78
σ_0^{HH},γ_{tot},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.43	3.77	13.12	4.98	0.79

TABLE IV: The performance results for the coherence-focused feature importance analysis.

Coherence Performance					
Experiment	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
	[m]	[m]	[%]	[m]	[-]
γ_{tot}	-1.85	5.80	21.26	7.65	0.51
γ_{vol}	-1.22	5.80	23.12	7.61	0.52
γ_{snr}	-0.51	6.07	26.74	8.14	0.44
γ_{tot},h_{amb}	-1.15	5.24	19.49	6.89	0.60
γ_{vol},h_{amb}	-0.64	5.19	19.33	6.83	0.61
γ_{snr},h_{amb}	-0.96	6.04	26.11	8.16	0.43
γ_{tot},γ_{vol}	0.04	4.71	18.93	6.13	0.69
γ_{snr},γ_{tot}	-0.63	4.95	20.07	6.57	0.63
γ_{snr},γ_{vol}	-0.52	5.00	21.43	6.62	0.63
γ_{tot},γ_{vol},h_{amb}	-0.21	4.45	17.17	5.80	0.72
γ_{snr},γ_{tot},h_{amb}	0.05	4.62	18.43	6.07	0.69
γ_{snr},γ_{vol},h_{amb}	-0.56	4.65	18.07	6.13	0.68
γ_{snr},γ_{tot},γ_{vol}	-0.14	4.90	21.59	6.44	0.65
γ_{snr},γ_{tot},γ_{vol},h_{amb}	-0.11	4.56	18.32	5.98	0.70
σ_0^{HH},γ_{tot},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.43	3.84	13.16	5.06	0.79
σ_0^{HH},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.39	3.82	13.09	5.02	0.79
σ_0^{HH},γ_{snr},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.32	4.22	15.45	5.58	0.74
σ_0^{HH},γ_{tot},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.14	3.77	13.13	4.96	0.80
σ_0^{HH},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.04	3.81	13.46	5.02	0.79
σ_0^{HH},γ_{snr},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.20	4.17	15.23	5.51	0.74
σ_0^{HH},γ_{tot},γ_{vol},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.22	3.80	13.08	5.00	0.79
σ_0^{HH},γ_{tot},γ_{snr},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.17	3.95	14.23	5.23	0.77
σ_0^{HH},γ_{vol},γ_{snr},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.26	3.92	14.16	5.20	0.77
σ_0^{HH},γ_{tot},γ_{vol},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.43	3.77	13.12	4.98	0.79
σ_0^{HH},γ_{tot},γ_{snr},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.23	3.87	13.90	5.14	0.77
σ_0^{HH},γ_{vol},γ_{snr},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.38	3.92	14.06	5.21	0.77
σ_0^{HH},γ_{tot},γ_{vol},γ_{snr},h_{amb},$h_{\mathrm{dem}}^{\mathrm{acq}}$,$\theta_{\mathrm{lia}}^{\mathrm{acq}}$	-0.15	3.85	13.89	5.13	0.78

[11] R. B. Mahesh and R. Hänsch, "Better coherence, better height: Fusing physical models and deep learning for forest height estimation from interferometric sar data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2025, pp. 2329–2338.

[12] A. Venugopal, M. Farnaghi, and R. Zurita-Milla, "Comparative evaluation of xai methods for transparent crop yield estimation using cnn," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2024, pp. 7478–7482.

[13] Y. Duan, Y. Bo, X. Yao, G. Chen, K. Liu, S. Wang, B. Yang, and X. Li, "A deep learning framework for long-term soil moisture-based drought assessment across the major basins in china," *Remote Sensing*, vol. 17, no. 6, 2025. [Online]. Available: https://www.mdpi.com/2072-4292/17/6/1000

[14] R. B. Mahesh and R. Hänsch, "Forest height estimation with tandem-x sar and insar features using deep learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[15] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision - ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, vol. 8689, pp. 818–833, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-319-10590-1_53

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM, Aug. 2016, pp. 1135–1144. [Online]. Available: https://dl.acm.org/doi/10.1145/2939672.2939778

[17] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017, arXiv:1705.07874 [cs]. [Online]. Available: http://arxiv.org/abs/1705.07874

[18] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," 2017.

[19] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for Simplicity: The All Convolutional Net," Apr. 2015, arXiv:1412.6806 [cs]. [Online]. Available: http://arxiv.org/abs/1412.6806

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, Oct. 2017, pp. 618–626. [Online]. Available: http://ieeexplore.ieee.org/document/8237336/

[21] W. Nie, Y. Zhang, and A. B. Patel, "A Theoretical Explanation for Perplexing Behaviors of Backpropagation-based Visualizations," 2018.

[22] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," 1997.

[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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3084–3094, 2021,

publisher: Institute of Electrical and Electronics Engineers (IEEE). [Online]. Available: https://ieeexplore.ieee.org/document/9353204/

[24] J. Armston, H. Tang, S. Hancock, S. Marselis, L. Duncanson, J. Kellner, M. Hofton, J. Blair, T. Fatoyinbo, and R. Dubayah, "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

[25] T. Fatoyinbo, J. Armston, M. Simard, S. Saatchi, M. Denbina, M. Lavalle, M. Hofton, H. Tang, S. Marselis, N. Pinto, S. Hancock, B. Hawkins, L. Duncanson, B. Blair, C. Hansen, Y. Lou, R. Dubayah, S. Hensley, C. Silva, J. R. Poulsen, N. Labriére, N. Barbier, K. Jeffery, D. Kenfack, M. Herve, P. Bissiengou, A. Alonso, G. Moussavou, L. T. White, S. Lewis, and K. Hibbard, "The NASA AfriSAR campaign: Airborne SAR and lidar measurements of tropical forest structure and biomass in support of current and future space missions," *Remote Sensing of Environment*, vol. 264, p. 112533, 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425721002534

[26] 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 Journal of Photogrammetry and Remote Sensing*, vol. 54, no. 2-3, pp. 115–122, Jul. 1999. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0924271699000027

[27] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3317–3341, Oct 2007.

[28] M. Zink, A. Moreira, I. Hajnsek, P. Rizzoli, M. Bachmann, R. Kahle, T. Fritz, M. Huber, G. Krieger, M. Lachaise, M. Martone, E. Maurer, and B. Wessel, "TanDEM-X: 10 years of formation flying bistatic SAR interferometry," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3546–3565, Feb 2021.

[29] T. Fritz, H. Breit, C. Rossi, U. Balss, M. Lachaise, and S. Duque, "Interferometric Processing and Products of the TanDEM-X Mission," in *Proceedings of the IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2012, pp. 1904–1907.

[30] C. Gonzalez, M. Bachmann, J. Bueso-Bello, P. Rizzoli, and M. Zink, "A fully automatic algorithm for editing the TanDEM-X global DEM," *Remote Sensing*, vol. 12, no. 23, p. 3961, Dec 2020.

[31] P. Rizzoli, M. Martone, C. Gonzalez, C. Wecklich, D. Borla Tridon, B. Braeutigam, M. Bachmann, D. Schulze, T. Fritz, M. Huber, B. Wessel, G. Krieger, M. Zink, and A. Moreira, "Generation and performance assessment of the global TanDEM-X digital elevation model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 132, pp. 119–139, Oct. 2017, publisher: Elsevier BV. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S092427161730093X

[32] R. Touzi, A. Lopes, J. Bruniquel, and P. Vachon, "Coherence estimation for SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 135–149, 1999, publisher: Institute of Electrical and Electronics Engineers (IEEE). [Online]. Available: http://ieeexplore.ieee.org/document/739146/

[33] K. Gdulová, J. Marešová, and V. Moudrỳ, "Accuracy assessment of the global tandem-x digital elevation model in a mountain environment," *Remote Sensing of Environment*, vol. 241, p. 111724, 2020.

[34] P. Rizzoli, L. Dell'Amore, J. Bueso-Bello, N. Gollin, D. Carcereri, and M. Martone, "On the derivation of volume decorrelation from TanDEM-X bistatic coherence," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3504–3518, Apr 2022.

[35] D. Carcereri, P. Rizzoli, D. Ienco, and L. Bruzzone, "A Deep Learning Framework for the Estimation of Forest Height From Bistatic TanDEM-X Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8334–8352, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10234548/

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[37] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," Apr. 2014, arXiv:1312.6034 [cs]. [Online]. Available: http://arxiv.org/abs/1312.6034

[38] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features Through Propagating Activation Differences," 2017.

[39] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Muller, "Evaluating the Visualization of What a Deep Neural Network Has Learned," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017, publisher: Institute of Electrical and Electronics Engineers (IEEE). [Online]. Available: https://ieeexplore.ieee.org/document/7552539/

[40] L. Shapley, *7. A Value for n-Person Games. Contributions to the Theory of Games II (1953) 307-317*. Princeton: Princeton University Press, 1997, pp. 69–79. [Online]. Available: https://doi.org/10.1515/9781400829156-012

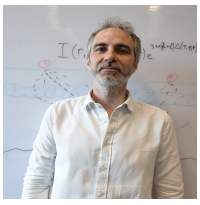
[41] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.

[42] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996. [Online]. Available: https://doi.org/10.1038/381607a0

[43] M. Dalponte, D. Marinelli, and Y. T. Solano-Correa, "Effect of dem used for terrain correction on forest windthrow detection using cosmo skymed data," *Remote Sensing*, vol. 16, no. 22, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/22/4309

Daniel Carcereri received the B.Sc. degree in Electronics and Telecommunications Engineering in 2018, and the M.Sc. degree in Information and Communication Engineering, with a specialization in Signal Processing and Understanding, in 2020, from the University of Trento, Italy. He pursued his master's thesis on quantization error reduction in synthetic aperture radar (SAR) systems, at the Microwaves and Radar Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany. In 2020, he was awarded a German Academic Exchange Service (DAAD) doctoral scholarship to conduct research within the Radar Science group at the Microwaves and Radar Institute, DLR. In 2023 he visited the Radar Science and Engineering Section at the NASA Jet Propulsion Laboratory, Pasadena, California (USA), to work on L-band repeat-pass InSAR data. In 2024 he received his Ph.D. degree (summa cum laude) in Information and Communication Technology from the University of Trento, Italy, with a thesis on deep learning for the retrieval of forest parameters from spaceborne Earth observation sensors. Since April of 2024, he joined the Microwaves and Radar Institute at the German Aerospace Center (DLR), as research scientist. Currently, his research interests involve signal processing, estimation theory, SAR interferometry, the development of novel deep-learning-based algorithms applied to bio-physical parameter estimation and analysis, and the generation of large-scale products.

Luca Dell'Amore Luca Dell'Amore received the B.Sc. degree in Electronics and Telecommunication Engineering in 2016 and the M.Sc. degree in Information and Communication Engineering, with a specialization in Signal Processing and Understanding, in 2019, both from the Università degli Studi di Trento, Trento, Italy. He carried out his master's thesis at the Microwaves and Radar Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany, where he investigated novel waveform-encoded Synthetic Aperture Radar (SAR) concepts. In 2019, he was awarded a German Academic Exchange Service (DAAD) doctoral scholarship to conduct research within the Radar Concepts Department at the Microwaves and Radar Institute, DLR. His research interests during this period included distributed SAR systems, with particular emphasis on Multiple-Input Multiple-Output (MIMO) SAR. Since 2021, he has been with the Radar Science Group, Satellite SAR Systems Department, as a Research Scientist and Project Engineer. His current work focuses on performance assessment for the TerraSAR-X and TanDEM-X missions, encompassing signal processing, SAR interferometry, the development of novel deep-learning-based algorithms, and the generation of large-scale products. His research activities also include the development of innovative methods for performance monitoring using Sentinel-1 data. Mr. Dell'Amore was the recipient of the Young Scientist Awards at the International Radar Symposium and Kleinheubacher Tagung, both in 2019.



Stefano Tebaldini Stefano Tebaldini (Senior Member, IEEE) received the M.Sc. degree in telecommunication engineering and the Ph.D. degree from Politecnico di Milano, Milan, Italy, in 2005 and 2009, respectively.,Since 2005, he has been with the Digital Signal Processing Research Group, Politecnico di Milano, where he has been an Associate Professor with the Dipartimento di Elettronica, Informazione e Bioingegneria, since 2018. He has authored over 60 articles published in international peer-reviewed journals. His research is mostly focused on imaging radars. His research activities include spaceborne, airborne, automotive, and ground-based synthetic aperture radar (SAR) processing, multistatic and distributed radars' processing, calibration, accurate platform localization, synchronization recovery, and ionospheric correction.,Prof. Tebaldini is a member of the BIOMASS Mission Advisory Group at ESA, tasked with providing scientific and technical advice on the BIOMASS Mission. His work on FDM MIMO SAR received the Best Paper Award at EuSAR 2024

Paola Rizzoli received the B.Sc. (2003) and M.Sc. (2006) degrees in telecommunication engineering from the Politecnico di Milano University (Polimi), Milan, Italy, and the Ph.D. degree (2018, with honors) in electrical engineering and information technology from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. From 2006 to 2008, she was with the Politecnico di Milano and Aresys s.r.l., a PoliMi spin-off company, as Scientific Researcher and Project Engineer. At the end of 2008, she joined the Microwaves and Radar Institute at the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, as Project Engineer, where she has been 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. Since 2016 she has been leading the Radar Science Group of the Satellite SAR Systems Department, 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. 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, the Best Paper Award at the German Microwave Conference in 2019 and the Best Poster Award at the ESA Worldcover Conference in 2017. She regularly serves as a reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.