Forest Height Estimation with TanDEM-X SAR and InSAR Features using Deep Learning

Ragini Bal Mahesh and Ronny Hänsch, *Senior Member, IEEE*

***Abstract*—Accurate forest height estimates lead to improved accuracy of biomass estimation and are crucial for monitoring and conservation efforts. InSAR (Interferometric Synthetic Aperture Radar) techniques use two SAR images to measure the interferometric coherence that includes the volumetric decorrelation which is known to be related to the forest canopy height. Several approximations and assumptions are made in the different steps to compute the volumetric decorrelation and to invert it to forest canopy height using physical models. Data-driven approaches overcome the potential bias introduced by these assumptions by directly estimating forest canopy height. However, the question of optimal representation and level of processing of the input data is often neglected. We address this gap comparing different SAR and InSAR input features such as Single-Look-Complex images (SLCs), backscatter, coherence, and volumetric decorrelation. The resulting best model has an RMSE of 6.12 m with volumetric decorrelation as primary input feature. It is followed by using coherence as primary input with an RMSE of 6.30 m.**

***Index Terms*—Deep Learning, Synthetic Aperture Radar (SAR), TanDEM-X, LVIS, Forest Canopy Height**

I. Introduction

Forest height estimation plays an important role in a wide range of ecological, environmental, and conservation applications. It supports advancing the understanding of forest ecosystems and promotes sustainable practices, e.g. via improved accuracy of forest biomass and carbon stocks estimates [1]. This information is fundamental for comprehending the vital role of forests in carbon sequestration and aligns with global initiatives like REDD+ (Reducing Emissions from Deforestation and Forest Degradation) that aim to reduce emissions from deforestation and forest degradation [2]. Beyond carbon estimation, forest height also emerges as a critical parameter for monitoring ecosystem dynamics. Leveraging frequent and high-resolution observations, it facilitates tracking of forest growth, disturbances (such as logging or natural events), and regrowth. Such insights are important while assessing the overall health and resilience of forest ecosystems, providing essential data for informed conservation and management decisions [3].

Several remote sensing technologies are utilized in the measurement of forest height such as optical imaging sensors [4], Light Detection and Ranging (LiDAR) [5], and Synthetic Aperture Radar (SAR) [6]. Optical sensors capture visible and near-infrared light, offering information on forest health, structure, and indirectly forest height. LiDAR, deployed on both space- and airborne platforms rely on the detection of laser beams from the top of the canopy to create highly detailed three-dimensional models of the forest, allowing for precise height measurements. SAR systems emit microwave signals towards the Earth's surface and record the backscattered echoes. The interaction between the radar waves and the forest canopy provides information about the structure and height of the vegetation. SAR has emerged as a suitable technology for forest height measurements over optical and LiDAR sensors due to several benefits such as robustness regarding weather conditions, day-and-night imaging capabilities, and the ability to partially penetrate the forest canopy.

The conventional approach for forest height estimation using SAR leverages Interferometric SAR (InSAR) data [5], [6] and includes establishing an empirical relationship with the interferometric coherence, a correlation variable derived from two SAR images. Coherence can be decomposed into various decorrelation factors, with volume decorrelation being one of the most important contributions [7]. This particular factor is intricately linked to the volumetric scattering of the forest canopy. It quantifies the level of decorrelation caused by the three-dimensional nature of the vegetation structure. Despite its pivotal role, its estimation remains a challenging task. Often a physical model such as Random Volume over Ground (RVoG) [5], [6] is used in inverting the forest height from the complex volumetric scattering. However, such models have certain limitations in practical applications including potential inaccuracies in representing complex forest structures, sensitivity to assumptions about the distribution of scatterers, and challenges in handling heterogeneous landscapes.

In recent years, Deep Learning (DL) models have shown the capacity to learn complex relationships and patterns from data. They are able to adapt to intricate features within forest structures, to capture nonlinear dependencies that might be challenging for physics-based models, and to accurately represent the forest height structure [8]–[11]. Prior work has focused primarily on the advancement of different DL architectures and training methodology. However, the influence of the nature and representation of the input data on the final accuracy of the height regression has been neglected so far. We close this gap by using a relatively standard DL model to estimate forest canopy height but evaluate the influence of different SAR and InSAR features on the accuracy and robustness of the trained models, providing insights into the key features driving the data-based estimation process.

II. Related Work

Forest height estimation has seen recent advancements, leveraging the integration of data-driven analysis and advanced

Manuscript received May 14, 2024; revised August 11, 2024.

remote sensing technologies. Popular machine learning algorithms used for this task include Random Forests [12], [13] and Support Vector Machines [12], [14]. With the increasing availability of large-scale remote sensing data and the advantage offered by DL in handling complex tasks, many recent studies have focused on DL for estimating forest height and combined different data sources such as LiDAR, SAR, and optical imagery [15]–[18].

DL-based forest height estimation using InSAR data, formulates the task as either a classification or a regression problem, each offering distinct advantages. Classification frameworks for forest height estimation categorize vegetation into discrete height classes. This approach enables the identification of height ranges corresponding to different forest structures. Wang et al. [9] proposes a supervised technique based on a complex-valued convolutional neural network (CNN) to estimate forest height from complex interferometric coherence derived from polarimetric interferometric SAR data. Reference heights are categorized into eight height classes and resulting models have RMSE values ranging from 2.58 m to 3.74 m and R^2 values from 0.83 to 0.92. The proposed complex-valued CNN allows the use of complex-valued coherence while it adds to the complexity of architectural design with the integration of complex-valued convolutional operations and activation functions. Yang et al. [19] also use a classification approach to estimate forest height by quantizing the LiDAR-based reference heights into 1 m classes. The covariance matrix calculated from the SAR Tomography stack is used as input which contains both polarimetric and interferometric information. The model achieves an RMSE of 2.33 m. Despite this success, the classification approach bears the risk that it may not capture the full range of variation in forest height, as it bins data into discrete categories.

Regression-based approaches treat forest height estimation as a continuous value prediction task and learn the intricate relationships between input features and the continuous forest height variable. Thus, they are better suited for capturing fine-grained variations in forest height. Carcereri et al. [10] present a new framework that uses the unique capabilities of TanDEM-X bistatic SAR data and AfriSAR LVIS measurements for the estimation of forest height over Gabon using a fully convolutional network. The model leverages coherence, volume decorrelation, backscatter, and ancillary information such as a DEM, the height of ambiguity, and the local incidence angle as inputs and achieves an overall accuracy of 5.41 m RMSE and 0.76 R^2. Li et al. [11] similarly uses a CNN architecture to estimate forest height from complex coherence as input over one of the AfriSAR campaign regions (i.e. Lope, Gabon). To overcome the requirement of LiDAR as reference data, the framework uses estimations from two traditional methods, i.e. RVOG and DEM Differencing, and aims to compensate for the over- and under-estimations resulting from the two approaches. The model achieves an RMSE of 10.15 m and R^2 of 0.87. Authors note that the shortcomings of the traditional methods in producing good predictions equivalent to the ground truth also directly affect the results of the DL model.

In a first preliminary study [20] we investigate how forest height estimation depends on various input features, namely coherence, volume decorelation, and complex-valued backscatter, combined with information about the sensor geometry. The current work extends this study by also considering topographic corrected backscatter as well as including a more rigorous analysis of the combination of various input features. The aim of this paper is to understand the influence of different SAR (SLCs, backscatter) and InSAR (coherence, volumetric decorrelation) input features and the effect of their combinations (including the combination with ancillary information) on model performance in determining the best suitable input features for forest height estimation.

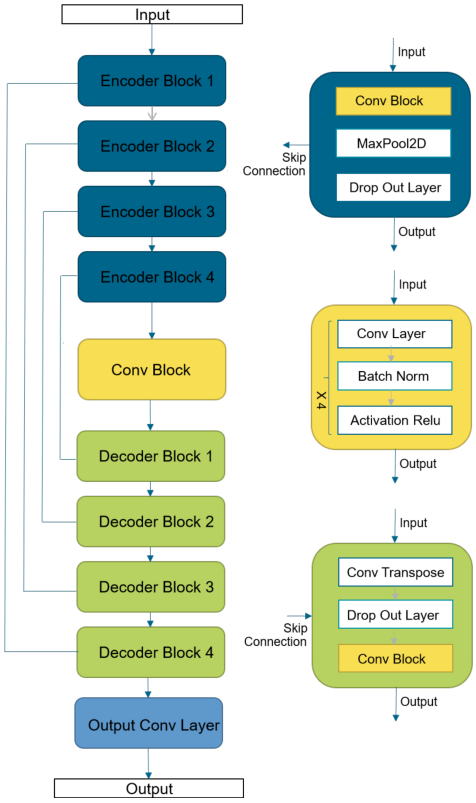

Fig. 1: U-Net model with input block, four encoder blocks, a convolutional block (conv block), four decoder blocks, and a convolutional output layer, along with skip connections.

III. Methodology

For the following experiments (see Section IV), we leverage different features derived from interferometric SAR (InSAR) data acquired by the German TanDEM-X mission representing various amounts of preprocessing shown in Figure 3. InSAR is based on two images acquired from two different positions (spatially and sometimes even temporally separated). Bistatic sensors such as TanDEM-X consist of two satellites flying in a close formation separated by a spatial baseline acquiring SAR data simultaneously from the same area. The two acquired SAR images are focused and spatially aligned before being used for further processing. All of the following features are derived from this data product and used as input for a neural network to derive forest height.

Additionally to the individual SAR images (both master and secondary image), we also use the backscatter coefficient σ_0 [10] (individually for image pair) which describes the local

texture of the scene, providing information about roughness and other surface characteristics.

The InSAR-derived features are computed from the interferometric pair of SAR images. Coherence γ is a cross-correlation factor [10] offering insights into the similarity of scattering properties between two images. We apply corrections for flat Earth and topography related phase effects by subtracting their contributions.

The coherence γ can be decomposed into several decorrelation factors, i.e. $\gamma = \gamma_s \cdot \gamma_t \cdot \gamma_q \cdot \gamma_{SNR} \cdot \gamma_{vol}$ [7]. The volume decorrelation γ_{vol} is caused by the presence of volumetric scattering and thus of particular relevance for forest height estimation [5], [6]. It is derived from γ by neglecting other decorrelation factors such as quantization decorrelation (γ_q), temporal decorrelation ($\gamma_t = 1$) due to the bistatic configuration, and decorrelation due to other sensor parameters ($\gamma_s = \gamma_{amb} \cdot \gamma_{rg} \cdot \gamma_{az}$) such SAR ($\gamma_{amb}$) and azimuth ($\gamma_{az}$) ambiguities as they do not depend on the illuminated scene. However, γ is dependent on the baseline unless proper range filtering actions are made resulting in range decorrelation (γ_{rg}). Thus, a rough approximation of $\gamma_{vol} = \gamma \cdot \gamma_{rg} / \gamma_{SNR}$ is computed compensating only for noise (γ_{SNR}) and γ_{rg}.

Additionally to the features above, we use the topographically corrected vertical wavenumber κ_z [6] that quantifies the rate of change in interferometric phase related to terrain height changes and serves as a sensitivity measure that is proportional to the baseline between SAR satellites.

All features are geocoded to the coordinate system of the reference LVIS measurements, i.e. EPSG:4326 - WSG84. We performed Range-Doppler terrain correction using Copernicus' 30 m Global DEM, and resampled the data to 20 m at which better forest height mapping is obtained with TanDEM-X data.

We use a standard U-Net [21] (shown in Figure 1) consisting of an input block delivering information to a sequence of four encoder blocks, four decoder blocks, and an output block together with dropout and batch normalization layers. As loss function we use Root-Mean-Squared Error (RMSE) with L2 regularization which is optimized via ADAM.

IV. Experiments

A. Experimental Setup

We use TanDEM-X (DLR) CoSSC data acquired in 2016 over Mabounie, Gabon. As reference forest height serves "AfriSAR: Gridded Forest Biomass and Canopy Metrics Derived from LVIS, Gabon, 2016" containing geolocated forest canopy cover derived from full-waveform LiDAR data obtained during the 2016 NASA-ESA AfriSAR campaign, recorded by LVIS, i.e. the readily processed Level 3 (l3a1) product RH98 (the mean relative height at which 98% of the waveform energy occurs) with a spatial resolution of 25 m. We match the interferometric products by resampling to 20 m and apply a Forest/Non-Forest map.

The available data is partitioned into spatially separated training (five image pairs) and validation (one image pair) sets. Attention was paid while choosing the validation set that it is representative of the forest height distribution found in the training set. Further visualization of the study area with splits is provided in the Appendix. The images are divided into patches of size 64×64 with a stride of 32. The U-Net is trained for 100 epochs with a batch size of 8 and a learning rate varying between 10^{-4} and 10^{-3} based on the combinations of available input features shown in Table I.

	1	2	3	4	5	6	7	8	9	10	11	12
SLC	X	X	X	X								
σ_0					X	X	X	X				
γ									X	X		
γ_{vol}			X	X			X	X			X	X
κ_z		X		X		X		X		X		X

TABLE I: Combination of input features used in the experiments representing different levels of preprocessing.

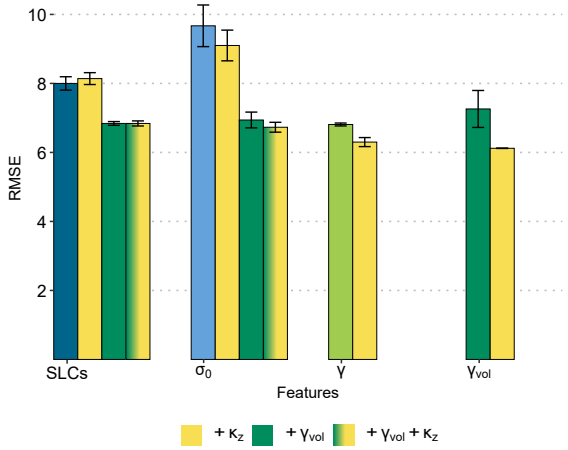

Fig. 2: Mean RMSE and standard deviation of trained model for each input feature and their combinations $+\kappa_z$, $+\gamma_{vol}$, and $+\gamma_{vol} + \kappa_z$ on validation scene.

We use RMSE and the coefficient of determination (R^2) as quantitative performance metrics. All experiments are repeated three times and performance averaged.

B. Results and Discussions

We quantitatively evaluate 12 variations of input modalities listed in Table I. Figure 2 shows the forest height accuracy of input features and their combinations. The RMSE mean is plotted against all input features with confidence intervals representing the standard deviation calculated over three repetitions. Features are grouped by their primary input modality and ordered according to the level of preprocessing, i.e. starting with SLCs and ending with volumetric decorrelation.

Among the evaluated input modalities, the best-performing model is the configuration $\gamma_{vol} + \kappa_z$ with an RMSE of 6.12 m and R^2 of 0.69. This is expected since the geometric distribution of scatterers in a canopy is the main cause of this type of decorrelation. Accordingly, it plays a significant role in physical models to estimate forest height from interferometric γ. This configuration $\gamma_{vol} + \kappa_z$ also aligns with practical knowledge of the physical modeling of forest height from $\gamma_{vol}(\kappa_z, w)$ as a function of the vertical distribution (z) of the scatterers at a given spatial baseline κ_z and single polarization w. The appropriate choice of κ_z values impacts the accuracy of physical models. Too large values saturate the sensitivity to forest height with underestimated heights for tall trees whereas too small values result in residual noise introducing large height errors due to the wrongful scaling of γ_{vol} to forest

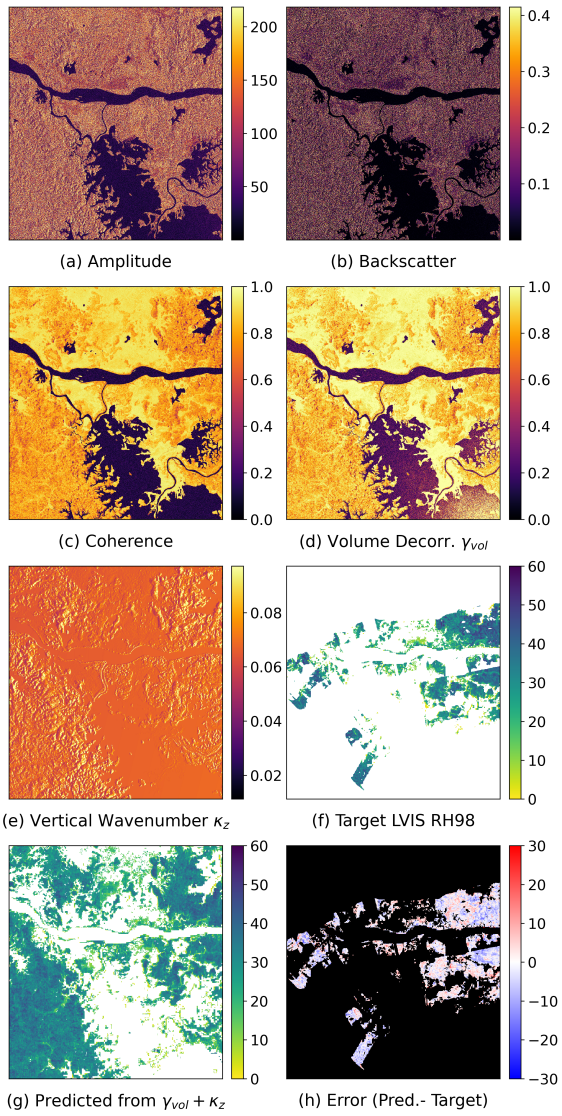

Fig. 3: Input data, reference target (white represent regions without data), and predicted forest height (white represent non-forest regions) for the test scene over Mabounie, Gabon.

heights. The dependency of γ_{vol} on κ_z is also captured by the DL model. Performance increases when using κ_z in addition to γ_{vol}, i.e. an RMSE of 6.12 m (R^2 = 0.69) versus not using κ_z with an RMSE of 7.26 m (R^2 = 0.56). Similar to physical models, providing the network with κ_z values gives a reference scaling factor for forest height inversion and helps improve estimates. A good agreement is obtained between the predicted and reference LVIS RH98 height for this model both qualitatively as shown in Figure 3 via predicted, target, and error maps and quantitatively shown by the density scatter plot with the respective marginal distributions in Figure 4. The error map is computed by subtracting target height from prediction and predominantly shows blue colors indicating an underestimation. Areas with the negative errors are observed to be located in very dense forest canopy regions. Typically tropical rainforests in Gabon can have canopy heights ranging from 20 m to 50 m which is observed in both the target and predicted heights resulting in higher density regions of the scatter plot given by red color gradients. Data points on the plot have a positive slope and are clustered around the regression line indicating that the model fits the data well and captures the relationship between the two variables.

The next best-performing model is the coherence model configuration $\gamma+\kappa_z$ with slightly higher RMSE than $\gamma_{vol}+\kappa_z$. The mean RMSE of this model is 6.30 m and R^2 is 0.67. Performance increases with the addition of κ_z as for the previous γ_{vol} model, i.e. without κ_z the RMSE increases to 6.81 m (R^2 = 0.61). The similarly good performance of the γ model compared to γ_{vol} can be attributed to the automatic denoising capability of the DL network when mapping γ to forest heights without the need to compensate for noise from other other decorrelation factors. Both InSAR input features γ and γ_{vol} with the addition of κ_z perform well overall and show robustness in mapping forest height estimates.

SAR input features have lower performance with higher error in comparison to InSAR input features. Models with SLCs and σ_0 as input have an RMSE of 8.00 m and 9.67 m with R^2 of 0.47 and 0.21, respectively. Addition of κ_z does not show consistent improvement across SAR features. While it reduces the RMSE of the $\sigma_0 + \kappa_z$ model to 9.1 m, the error of the $SLC + \kappa_z$ model increases slightly to 8.14 m but stays well within the range of the standard deviation of the model without κ_z. Texture information contained in σ_0 has its limitation in improving forest height estimation when used as primary input. Meanwhile, the network shows promising results in directly estimating forest height from SLCs but is not as efficient as estimating from InSAR features. Although networks can learn complex features and spatial relationships in the data, it seems challenging for the current network to learn the multiplication operations required for InSAR processing to estimate representations closer to γ_{vol}.

If the input modalities combine SAR with InSAR features, i.e. γ_{vol} and $\gamma_{vol} + \kappa_z$, the network potentially learns to disregard σ_0 in the presence of more straightforward input modalities reducing the RMSE error to 6.84 m for using SLCs and 6.94 m and 6.73 m for backscatter σ_0, respectively, i.e. a model performance that is nearly equivalent to using InSAR features as the primary input. However, adding any other feature to the best performing input of $\gamma_{vol}+\kappa_z$ leads to a slight decrease of performance. Adding a partially redundant yet less informative feature increases the problem complexity as the model has to learn to ignore this feature. Thus, input dimensionality does not yield an increased performance.

Results demonstrate that it is advantageous to provide pre-computed features to the network if they are known to be beneficial for the prediction task rather than relying on the general ability of neural networks to learn directly from data.

V. Conclusion

In conclusion, this study uses DL models (i.e. a U-Net) to assess the impact of various input features including SAR (SLCs and σ_0) and InSAR features (γ and γ_{vol}) on forest height estimation from TanDEM-X scenes with LVIS RH98 reference targets over Mabounie, Gabon. Different combinations of these features are evaluated using changes in RMSE and R^2 across diverse processing scenarios. Among all models, the configuration with $\gamma_{vol} + \kappa_z$ emerges as the best-

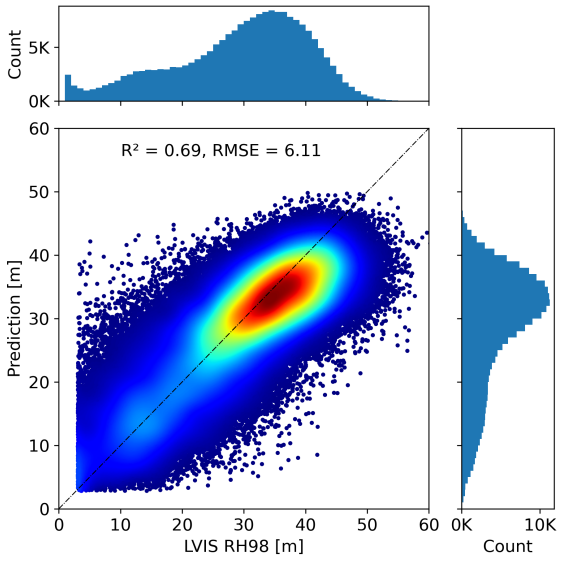

Fig. 4: Density Scatter Plots of the predicted forest [m] height vs. LVIS RH98 [m] for the validation scene for the best performing model with input combination of $\gamma_{vol} + \kappa_z$.

performing approach aligning with theoretical and practical forest height modeling principles. Coherence coupled with κ_z also shows very good performance eliminating the need to denoise data to obtain volumetric decorrelation from coherence by compensating for other decorrelation factors. In contrast to the InSAR features, the models with SAR input features showed weaker performance. σ_0 doesn't contribute any useful information in the presence of InSAR features highlighting the significance of strategic feature selection in optimizing DL models for accurate forest height estimation.

Future work will focus on specific adjustments to the model architecture to tailor it for tasks that involve the multiplication of input features as in interferogram computation from SLCs to capture complex information. These adaptations include modifying the network architecture to an adequately complex model, a sufficient number of training samples, and enough computational resources (e.g. number of epochs). Additionally, the complexity of directly mapping from SLCs can be addressed by novel training strategies or loss functions that introduce knowledge from physical inversion models to guide the network and perform a more streamlined forest height mapping. While the current results clearly show the benefits of InSAR derived features, these modifications have the potential to yield comparable SLC-based outcomes. Such an approach would not rely on assumptions and processing steps necessary to compute InSAR features (e.g. window size to compute coherence, coherence decomposition, etc.).

ACKNOWLEDGEMENTS

We would like to express our gratitude to Islam Mansour, German Aerospace Center (DLR) for providing the dataset.

REFERENCES

[1] Changhyun Choi, Matteo Pardini, Michael Heym, and Konstantinos P. Papathanassiou, "Improving forest height-to-biomass allometry with structure information: A TanDEM-X study," *IEEE JSTARS*, vol. 14, pp. 10415–10427, 2021.

[2] Mucahid Mustafa Bayrak and Lawal Mohammed Marafa, "Ten years of REDD+: A critical review of the impact of REDD+ on forest-dependent communities," *Sustainability*, vol. 8, no. 7, pp. 620, 2016.

[3] Cornelius Senf, "Seeing the system from above: The use and potential of remote sensing for studying ecosystem dynamics," *Ecosystems*, vol. 25, no. 8, pp. 1719–1737, 2022.

[4] Nico Lang, Konrad Schindler, and Jan Dirk Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sensing of Environment*, vol. 233, pp. 111347, 2019.

[5] Roman Guliaev, Victor Cazcarra-Bes, Matteo Pardini, and Konstantinos Papathanassiou, "Forest height estimation by means of TanDEM-X InSAR and waveform lidar data," *IEEE JSTARS*, vol. 14, pp. 3084–3094, 2021.

[6] Florian Kugler, Daniel Schulze, Irena Hajnsek, Hans Pretzsch, and Konstantinos P Papathanassiou, "TanDEM-X Pol-InSAR performance for forest height estimation," *IEEE TGRS*, vol. 52, no. 10, pp. 6404–6422, 2014.

[7] Paola Rizzoli, Luca Dell'Amore, José-Luis Bueso-Bello, Nicola Gollin, Daniel Carcereri, and Michele Martone, "On the derivation of volume decorrelation from tandem-x bistatic coherence," *IEEE JSTARS*, vol. 15, pp. 3504–3518, 2022.

[8] Islam Mansour, Konstantinos Papathanassiou, Ronny Haensch, and Irena Hajnsek, "Towards a symbiosis of model-based and machine learning forest height estimation based on TanDEM-X InSAR," in *EUSAR 2022*. VDE, 2022, pp. 1–4.

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

[10] Daniel Carcereri, Paola Rizzoli, Dino Ienco, and Lorenzo Bruzzone, "A deep learning framework for the estimation of forest height from bistatic TanDEM-X data," *IEEE JSTARS*, 2023.

[11] Dandan Li, Hailiang Lu, Chao Li, Linda Mohaisen, and Weipeng Jing, "Forest height inversion by convolutional neural networks based on L-band PolInSAR data without prior knowledge dependency," *IEEE JSTARS*, 2023.

[12] Maryam Pourshamsi, Junshi Xia, Naoto Yokoya, Mariano Garcia, Marco Lavalle, Eric Pottier, and Heiko Balzter, "Tropical forest canopy height estimation from combined polarimetric SAR and LiDAR using machine-learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 172, pp. 79–94, 2021.

[13] Xinyi Liu, Li He, Zhengwei He, and Yun Wei, "Estimation of broadleaf tree canopy height of wolong nature reserve based on InSAR and machine learning methods," *Forests*, vol. 13, no. 8, pp. 1282, 2022.

[14] Maryam Pourshamsi, Mariano Garcia, Marco Lavalle, and Heiko Balzter, "A machine-learning approach to PolInSAR and LiDAR data fusion for improved tropical forest canopy height estimation using NASA AfriSAR campaign data," *IEEE JSTARS*, vol. 11, no. 10, pp. 3453–3463, 2018.

[15] Shaojia Ge, Hong Gu, Weimin Su, Jaan Praks, and Oleg Antropov, "Improved semisupervised unet deep learning model for forest height mapping with satellite SAR and optical data," *IEEE JSTARS*, vol. 15, pp. 5776–5787, 2022.

[16] Shaojia Ge, Oleg Antropov, Tuomas Häme, Ronald E McRoberts, and Jukka Miettinen, "Deep learning model transfer in forest mapping using multi-source satellite SAR and optical images," *Remote Sensing*, vol. 15, no. 21, pp. 5152, 2023.

[17] Alexander Becker, Stefania Russo, Stefano Puliti, Nico Lang, Konrad Schindler, and Jan Dirk Wegner, "Country-wide retrieval of forest structure from optical and SAR satellite imagery with deep ensembles," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 269–286, 2023.

[18] Wang Li, Zheng Niu, Rong Shang, Yuchu Qin, Li Wang, and Hanyue Chen, "High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 92, pp. 102163, 2020.

[19] Wenyu Yang, Sergio Vitale, Hossein Aghababaei, Giampaolo Ferraioli, Vito Pascazio, and Gilda Schirinzi, "A deep learning solution for height estimation on a forested area based on Pol-TomoSAR data," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[20] Ragini Bal Mahesh and Ronny Hänsch, "Deep learning for forest canopy height estimation from SAR," in *IGARSS 2023*, 2023, pp. 5672–5675.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015, Munich, Germany, October 5-9, 2015, part III 18*. Springer, 2015, pp. 234–241.

APPENDIX A
ADDITIONAL RESULTS TABLE

TABLE II: Mean performance metrics MAE (m), RMSE (m), and R^2 for all combinations of input features over three iterations.

Model	Mean MAE	Mean RMSE	RMSE STD	Mean R^2
SLCs	6.31	8.00	0.1936	0.47
SLCs + κ_z	6.41	8.14	0.1721	0.45
SLCs + γ_{vol}	5.39	6.84	0.0544	0.61
SLCs + κ_z + γ_{vol}	5.33	6.84	0.0735	0.61
σ_0	7.56	9.67	0.6034	0.21
σ_0 + κ_z	7.06	9.1	0.4450	0.31
σ_0 + γ_{vol}	5.40	6.94	0.2290	0.6
σ_0 + κ_z + γ_{vol}	5.31	6.73	0.1429	0.62
γ	5.25	6.81	0.0424	0.61
γ + κ_z	4.89	6.30	0.1310	0.67
γ_{vol}	5.75	7.26	0.5354	0.56
γ_{vol}+ κ_z	4.74	6.12	0.0047	0.69

APPENDIX B
DATASET PREPROCESSING

All the extracted input features are geocoded from slant range geometry to an Earth coordinate reference system using the SNAP toolbox. Geometric distortions (e.g. foreshortening) present in the SAR data due to the slant range geometry are removed by performing Range Doppler Terrain Correction using Copernicus 30 m Global DEM, where all pixels are shifted to their correct locations according to the input DEM using bi-linear interpolation. At the end of the processing, all input features and the reference measurements are resampled to 20 m spatial resolution.

Figure 5 shows the selected study area along with reference data and the split into train and test set. Figure 6 shows the distribution of the vertical wavenumber κ_z within these scenes.

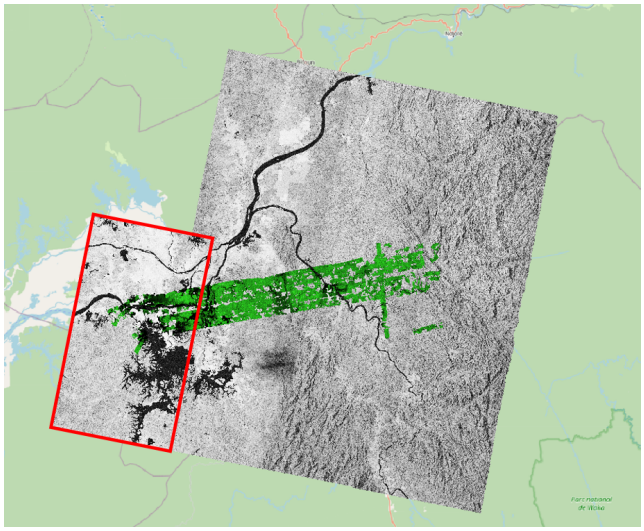

Fig. 5: Study area in Mabounie, Gabon consisting of training (5 images) and validation (1 image) split. Dark green color here represents the Lidar measurements overlaid on top of Tandem-X Coherence Images and red boundary is the validation region.

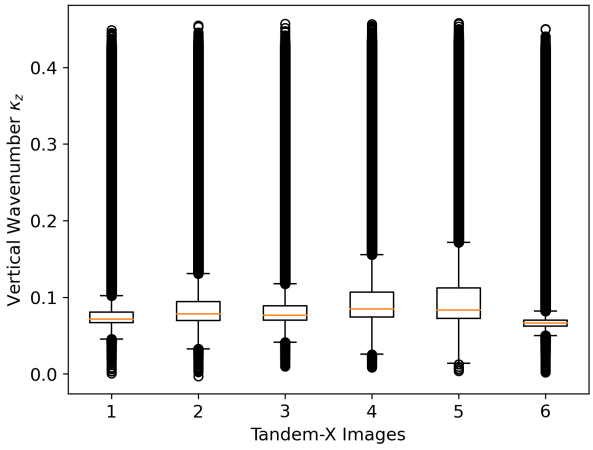

Fig. 6: Vertical Wavenumber κ_z for training images (1-5) and validation image (6).

APPENDIX C
LEARNING CURVES

Figure 7 shows that both training and validation loss converge relatively quickly and stable for the best performing model. While the training loss continues to decrease, the validation loss does not improve further.

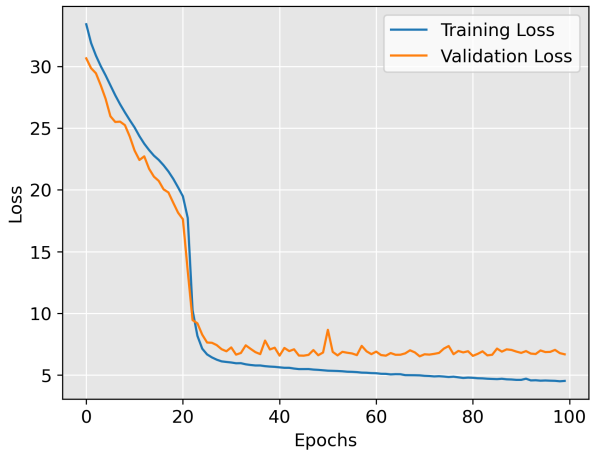

Fig. 7: Training and validation loss (RMSE) curves for the best performing model γ_{vol}+ κ_z.

APPENDIX D
ADDITIONAL QUALITATIVE RESULTS

Figure 8 shows predictions and error maps for all features combinations.

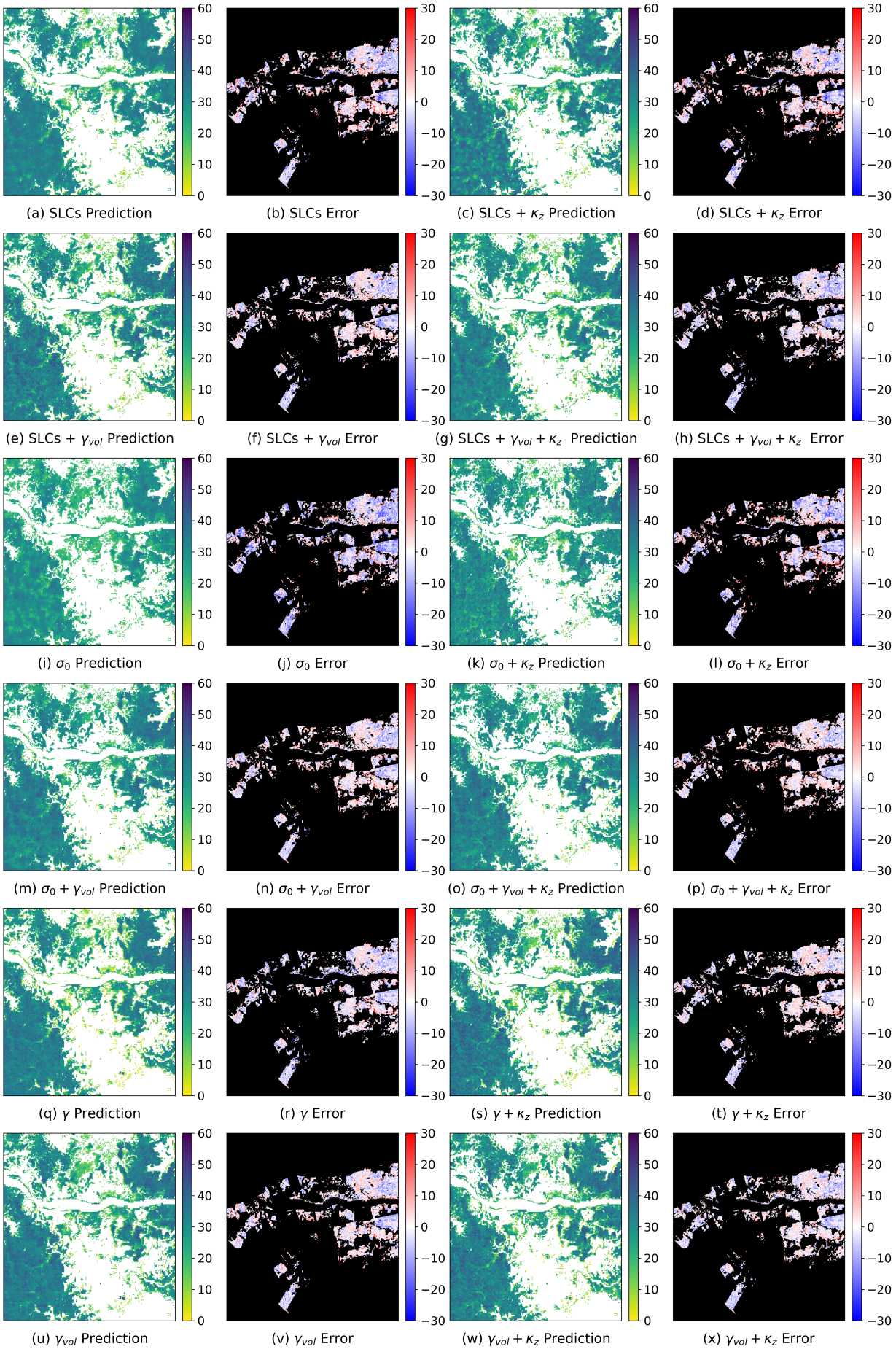

Fig. 8: Prediction and error (prediction - target) maps for all input feature combinations in this study. Predictions are masked with forest and non-forest masks (white regions).