

UNIVERSITÀ
DI TRENTO
Department of
Information Engineering and Computer Science

Doctoral School in
Information and Communication Technology

A Deep Learning Study on the Retrieval of Forest Parameters from Spaceborne Earth Observation Sensors

Daniel Carcereri

Advisor
Prof. Lorenzo Bruzzone
Università di Trento

Co-Advisor
Dr. Paola Rizzoli
Deutsches Zentrum für Luft- und Raumfahrt

July 25th, 2024

Abstract

The efficient and timely monitoring of forest dynamics is of paramount importance and requires accurate, high-resolution and time-tagged predictions at global scale. Despite numerous methodologies have been proposed in the literature, existing approaches often compromise on accuracy, resolution, temporal fidelity or coverage.

To tackle these challenges and limitations, the main objective of this doctoral thesis is the investigation of the potential of artificial intelligence (AI) for the regression of bio-physical forest parameters from spaceborne Earth Observation (EO) data. This work explores for the first time the combined use of TanDEM-X single-pass interferometric products and convolutional neural networks for canopy height estimation at country scale. To achieve this, a novel deep learning framework is proposed, leveraging the capability of deep neural networks to effectively capture the complex spatial relationships between forest properties and satellite data, as well as ensuring the adaptability to different environmental conditions. The design and the understanding of the model is driven by explainable AI principles and by considerations on large-scale forest dynamics, with a great emphasis set on the challenges related to the variable acquisition geometry of the TanDEM-X mission, and by relying on the use of LVIS-derived LiDAR measurements as reference data. Moreover, several investigations are conducted on the adaptability of the developed framework for transferring knowledge to related domains, such as digital terrain model regression and above-ground biomass density estimation. Finally, the capability of the proposed approach to be extended to the use of other EO sensors is also evaluated, with a particular emphasis on the ESA Sentinel-1 and Sentinel-2 missions.

The developed deep learning framework sets a solid groundwork for the generation of large-scale products of bio-physical forest parameters from spaceborne EO data. The approach achieves cutting-edge performance, significantly advancing the current state of forest assessment and monitoring technologies.

Keywords

Forest Parameter Regression, Canopy Height, Above-Ground Biomass, Earth Observation, Remote Sensing, Deep Learning, InSAR, Multi-Spectral, LiDAR, TanDEM-X, Sentinel-1, Sentinel-2, LVIS, GEDI

Contents

1 Introduction **1**
- 1.1 Importance of Forest Monitoring . . . 1
- 1.2 Research Gap and Work Objectives . . . 2
- 1.3 Original Contributions of the Thesis . . . 5
- 1.4 Thesis Structure . . . 6

2 Theoretical Background and Principles of Earth Observation Sensors **9**
- 2.1 Principles of Radar . . . 9
  - 2.1.1 Range Resolution . . . 10
  - 2.1.2 Radar Backscattering . . . 12
  - 2.1.3 Point Scatterer . . . 13
  - 2.1.4 Distributed Scatterer . . . 14
- 2.2 Principles of SAR Imaging and Interferometry . . . 16
  - 2.2.1 SAR Acquisition Geometry . . . 16
  - 2.2.2 SAR Image Formation . . . 18
  - 2.2.3 Principles of SAR Interferometry . . . 20
- 2.3 Principles of LiDAR . . . 29
  - 2.3.1 Transmission Equation . . . 29
  - 2.3.2 Point Cloud LiDAR . . . 30
  - 2.3.3 Full Waveform LiDAR . . . 32
- 2.4 Principles of Optical Sensors . . . 32
  - 2.4.1 Radiation Theory . . . 33
  - 2.4.2 Characterization of Objects on the Earth . . . 34
  - 2.4.3 Acquisition Principle and Resolution . . . 35

3 Deep Learning for Regression Tasks **39**
- 3.1 Regression . . . 39
  - 3.1.1 Linear Regression . . . 39
  - 3.1.2 Nonlinear Regression . . . 43
  - 3.1.3 Optimization Problem . . . 44
- 3.2 Feedforward Neural Networks . . . 46
  - 3.2.1 Perceptron . . . 46
  - 3.2.2 Multi-Layer Perceptron . . . 48
  - 3.2.3 Activation Functions . . . 50

3.2.4 The Backpropagation Algorithm . . . 51
3.2.5 Model Training, Validation and Testing . . . 52
3.3 Regularization . . . 55
3.3.1 L2-Norm . . . 55
3.3.2 L1-Norm . . . 56
3.3.3 Stochastic Gradient Descent . . . 57
3.3.4 Dropout . . . 57
3.3.5 Batch Normalization . . . 58
3.3.6 Early Stopping . . . 59
3.3.7 Data Augmentation . . . 59
3.4 Convolutional Neural Networks . . . 59
3.4.1 The Convolutional Layer . . . 60
3.4.2 Receptive Field . . . 62

4 Forest Parameter Estimation: State of the Art Overview 65
4.1 Physical-Based Models . . . 67
4.2 Data-Driven Models . . . 72

5 EO Missions, Datasets and Pre-Processing 77
5.1 TanDEM-X . . . 77
5.1.1 The TerraSAR-X and TanDEM-X Satellites . . . 77
5.1.2 The TanDEM-X Mission . . . 78
5.1.3 TanDEM-X Processing Chain . . . 79
5.2 LVIS . . . 80
5.2.1 AfriSAR Campaign . . . 81
5.3 GEDI . . . 83
5.4 Sentinel-1 . . . 85
5.5 Sentinel-2 . . . 86
5.6 ESA WorldCover Map . . . 88
5.7 The Dataset Pre-processing Pipeline . . . 89
5.7.1 Aggregation . . . 89
5.7.2 Common Grid Interpolation . . . 90
5.7.3 Alignmnent . . . 92
5.7.4 Sampling . . . 92

6 A Novel Deep Learning Framework for Canopy Height Regression 95
6.1 Training, Validation and Testing Strategies . . . 96
6.1.1 Dataset Composition . . . 96
6.1.2 Geographic Subsetting and Dataset Processing . . . 97
6.1.3 Model Training and Validation . . . 99
6.1.4 Model Testing . . . 99
6.2 The Proposed Deep Learning Model . . . 100
6.2.1 Hyperparameter Tuning . . . 100
6.3 Performance Baseline . . . 106
6.3.1 Overall Performance . . . 108

6.3.2 Individual Areas . . . 109
6.3.3 The Impact of Topography . . . 111
6.4 Spatial Transfer Analysis . . . 113
6.5 Comparison with the RVoG model . . . 116
6.6 Feature Importance Analysis . . . 119
6.6.1 Impact of DEM and local incidence angle . . . 120
6.6.2 SAR and bistatic InSAR feature comparison . . . 121
6.7 Discussion . . . 123
6.8 Conclusion . . . 125

7 Generalization of the Framework to Country-Scale 127
7.1 Country-Scale Inference and Map of Applicability . . . 127
7.1.1 Country-Scale Inference . . . 128
7.1.2 Map of Applicability . . . 128
7.2 The Challenge of Missing Predictor Representations . . . 130
7.2.1 The Updated Training Strategy . . . 130
7.2.2 Results . . . 131
7.2.3 Discussion . . . 132
7.3 Research asset 1: suitability of the GEDI dataset . . . 135
7.3.1 Comparison against LVIS Measurements . . . 135
7.3.2 The Impact on the Proposed Framework . . . 139
7.3.3 Discussion . . . 139
7.4 Research asset 2: modification of the predictors and the training set . . 140
7.4.1 Height of Ambiguity Analysis . . . 141
7.4.2 DEM Analysis . . . 142
7.4.3 Discussion . . . 146
7.5 Research asset 3: Height of Ambiguity Alignment Model . . . 151
7.5.1 Pre-Processing Task . . . 152
7.5.2 Downstream Task . . . 156
7.5.3 Discussion . . . 159
7.6 Conclusions . . . 160

8 Generalization of the Framework to Other Forest-Related Variables and Sensors 163
8.1 Generalization of the Proposed Framework to the Regression of Other Forest-Related Parameters . . . 163
8.1.1 Estimation of the Digital Terrain Model . . . 164
8.1.2 Estimation of the Above-Ground Biomass Density . . . 166
8.1.3 Discussion . . . 168
8.2 Generalization of the proposed Framework to Sentinel-1 Data . . . 171
8.2.1 Sentinel-1 Dataset . . . 171
8.2.2 Results . . . 172
8.2.3 Discussion . . . 175
8.3 Generalization of the proposed Framework to Sentinel-2 Data . . . 176

8.3.1 Sentinel-2 Dataset 176
8.3.2 Results 177
8.3.3 Discussion 180
8.4 Conclusions 181

9 Conclusions and Future Work **183**
9.1 Summary and Conclusions 183
9.2 Outlook 186

10 Bibliography **191**

A SAR Theory **215**
A.1 Azimuth Resolution 215
A.2 Azimuth Compression 217
A.3 The Speckle Model 219
A.4 Flat-Earth Removal 220
A.5 RVoG Dual- and Single-Pol Inversion 222

B Estimation Theory **225**
B.1 Maximum Likelihood Estimator 225
B.2 Backpropagation Example 226
B.3 Sources of Error 227

List of Tables

3.1 The truth table for the XOR function, defined by the inputs x_1 and x_2, and the output y. 48

4.1 A summary of the different state-of-the-art forest parameter regression approaches, with an emphasis on the major advantages and drawbacks for each approach-family. 76

5.1 Summary of the main TerraSAR-X and TanDEM-X system parameters. 78

5.2 Summary of the main system parameters for the ESA Sentinel-1 constellation. 85

5.3 Acquisition modes and products offered for the ESA Sentinel-1 constellation. 85

6.1 Results for the baseline scenario presented in Sec. 6.3 and for all cross-validation permutations of the spatial transfer analysis. Each metric is evaluated for each test site individually and overall. 110

6.2 Overall prediction performance for all the analyzed test case scenarios. Number of considered acquisitions = 12, Number of pixels = 1465947. . 121

7.1 (a) Performance metrics computed for the model trained using TanDEM-X data acquired in 2015-2016 only, for each AfriSAR test site, separately, and overall. (b) Performance metrics computed for the cross-validation experiment, for each AfriSAR test site permutation, separately, and overall. 131

7.2 Performance metrics of the rasterized GEDI shots, compared to those acquired with the LVIS sensor. On the one hand, the different RH metrics display a progressive negative increase in the estimation bias. 135

7.3 (a) Performance metrics computed for the model trained using the extended set of TanDEM-X data acquired in between 2010 and 2021, for each AfriSAR test site, separately, and overall. (b) Performance metrics computed for the corresponding cross-validation experiment, for each AfriSAR test site permutation, separately, and overall. 142

7.4 Performance metrics for the DEM analysis experiment. Each row identify the performance of the model derived from the extended dataset presented in Sec. 7.4.1 (**Baseline**), the model trained without the DEM as predictor (**w/o DEM**), the model trained with the DEM derivatives (∇) and the model trained with the high-pass filtered version of the DEM (**HPF**). 145
7.5 (a) Performance metrics computed for the final model trained using the estimate of the spatial DEM gradient as a replacement for the DEM itself, shown for each AfriSAR test site separately and overall. (b) Performance metrics computed for the corresponding cross-validation experiment, shown for each AfriSAR test site permutation separately and overall. 146
7.6 The performance results for the proposed HAM experiment. 154
7.7 Performance results for the downstream experiments to the proposed HAM solution. The metrics are presented both with and without the HAM pre-processing for better comparison. 157

8.1 DTM estimation: test site-depending and overall prediction accuracy. 164
8.2 AGBD estimation: test site-dependent and overall prediction accuracy. . . . 169
8.3 Performance summary for the estimation of the canopy height using Sentinel-1 data 173
8.4 Performance summary for the canopy height estimation using Sentinel-2 data. 178
8.5 Performance summary for the canopy height estimation using Sentinel-2 data. 179

List of Figures

2.1 Examples of two radar returns after transmitting gated sine signals. (a) shows the signals (left) and their cross-correlation separation for two targets placed at the resolution limit (right), while (b) shows how the peaks of the targets become visible at a slightly larger separation. Note that the separations in the signal domains appear indistinguishable from each other. 11

2.2 Examples of two radar returns after transmitting chirp signals. (a) shows the signals and their cross-correlation separation for two targets placed at the gated sine resolution limit, while (b) shows how the peaks of the targets become indistinguishable at the chirp resolution limit. 12

2.3 Reference geometry for the backscattering normalization areas A_β, A_σ and A_γ. 15

2.4 Representation of the SAR acquisition geometry. 17

2.5 Schematic representation of the SAR focusing process. The star symbol represent the convolution operation. It is possible to observe that the raw data domain is unintelligible, and requires focusing for the different targets to become recognizable. This is due to the large integration-windows across which each signal is spread. 18

2.6 Schematic representation of possible geometric distortions introduced by the side-looking geometry of SAR. Note that when a slope is facing towards the antenna, more targets get compressed into the same resolution cell, thus making the return appear brighter. The opposite is true for slopes facing in the opposite direction. 20

2.7 Schematic overwiew of the acquisition geometry for across-track SAR interferometry. Note that the view in the top-right shows the equivalent simplified geometry for the far-field approximation. 22

2.8 Simplified representation of the two-way fractional phase for each point in the normal plane to the flight direction, as transmitted and measured by the first sensor (a), the second sensor (b) and as measured by the interferometer (c). Note that the horizontal axis represents the flat Earth, while the black line shows an example of topography. It is then possible to compare the relative change in measured interferometric phase between a given point on the horizontal axis and it corresponding point on the topographic curve. . . 23

2.9 Representation of a simulated InSAR scene, showing (a) a synthetic DEM h_0, (b) the corresponding interferometric phase ϕ_{if}, (c) its flat-earth term ϕ_{fe} and (d) the topographic term ϕ_{topo}. 25

2.10 Impact of different levels of coherence on the measured interferogram. As the coherence decreases, the phase noise increases, leading to less clearly defined fringes and more severe phase unwrapping errors. 27

2.11 Relationship between the volume decorrelation coefficient γ_{vol} and the acquisition height of ambiguity h_{amb} for different tree height ranges, sampled using TanDEM-X derived measurements acquired across the tropical forests of Gabon (Africa). 28

2.12 Example of ALS-derived point cloud acquired over forested area in the Hainich National Park, Germany. The $60 \times 60\,\text{m}^2$ scene contains $\approx 10^5$ samples, for an average sample density of $\approx 27\,\text{pts/m}^2$. 31

2.13 Elevation products derived from the point cloud in Fig. 2.12. These consist in the (a) DSM, (b) DTM and (c) CHM, rasterized at 1 m ground sampling distance. 31

2.14 A simulated full-waveform LiDAR signal over a forested area. The image on the left shows the received signal amplitude (in blue) with respect to both the time of flight and the corresponding absolute elevation. The image on the right shows the detected positions (in red) for the top of the canopy and the ground, and the estimated relative height trend (in orange). 32

2.15 Black-body radiation model applied to the Earth and the Sun. Notably, the power radiated by the Earth is completely dwarfed by that emitted by the Sun. With an average surface temperature of 228 K, Earth's emission peak is at $\approx 10\,\mu m$ (i.e. thermal infrared), while with 5778 K the Sun has its emission peak at $\approx 0.5\,\mu m$ (i.e. visible light). 34

2.16 Example of spectral signatures for different classes of objects acquired by the ESA Sentinel-2 multi-spectral sensor over the country of Gabon, Africa. Notably, it is possible to observe the characteristic *red-edge* behaviour for the vegetation classes, characterized by a strong absorption at red wavelengths ($664.6\,\mu$m) and a strong reflection at near-infrared wavelengths ($864.7\,\mu$m). . 35

3.1 Example of a one-dimensional linear regression problem. The data points (in black) follow the linear trend captured by $y = \beta + w\,x$ (blue line). The offset parameter β of the line represents the intercept, while the weighting parameter w represents the slope. 40

3.2 Example of residuals for a 1D linear regression problem. For each of the data points (in black) it is possible to compute the corresponding residual (red dotted lines) w.r.t. the linear regression line (in blue). The objective of the least squares approach is to find the linear model parameters that minimize the sum of the squared residuals. 42

3.3 (a) example of a loss landscape characterized by the presence of multiple local minima. (b) Walk direction found by the GD method for a random starting location corresponding to the set of parameters $\boldsymbol{\theta}_0$. It is possible to note that the unfavourable starting position of $\boldsymbol{\theta}_0$ has forced the solver into a bad local minimum. 45

3.4 The schematic representation of the Rosenblatt Perceptron. All inputs x_j are individually multiplied by their corresponding weighting coefficient w_j, and subsequently added together with the bias term β. The resulting pre-activation signal α is then passed to the Heaviside activation function $H(\alpha)$ to generate the final output y. 47

3.5 The schematic representation of a single-output MLP. Each unit p in layer l has its own set of weights $\boldsymbol{\theta}_{l,p}$. Note that each node is connected to all nodes in the preceding and the succeeding layers, meaning that by design all inputs at a given layer are equally considered. This layer topology is commonly referred to as a *dense* or a *fully-connected* layer. 49

3.6 Examples of different activation functions. From left to right, the *Heaviside* step-function, the *sigmoid*, the *tanh* and the *ReLU*. 50

3.7 A typical example of training and validation loss co-evolution. Both losses decrease together up to a certain point, after which the validation loss starts worsening while the training loss keeps improving. The point of lowest validation loss corresponds to the model with best generalization properties. In the epochs before, the model is said to underfit the data, while after it the model is said to overfit the data. 54

3.8 Impact of the $l2$-norm on the loss landscape of Fig. 3.3 (a). Using the GD method now leads to an improved local minima. 56

3.9 Impact of the $l1$-norm on the loss landscape of Fig. 3.3 (a). Using the GD method now leads to an improved local minimum. 56

3.10 Graphic representation of the 2D convolution operation typically used in CNNs. The output value (in red) is the result of the convolution between the input matrix (in blue) and the *high-pass* kernel function (in yellow). Note that the border values in the output matrix are greyed out to indicate that it is not possible to compute the result for these values. 61

3.11 Example of a 1×1 convolution of a RGB input tensor (left) with a kernel function (center) to produce a new output tensor (right). Note that in a 2D-convolution, the kernel has the same depth as the input tensor, while the output is a tensor of unitary depth (i.e. a single feature). 62

3.12 The output pixel dependency on the pixels in the previous layers of a CNN using 3×3 kernel filters. The pixels in dark blue highlight the expansion of the model's receptive field. 63

3.13 Comparative example of a plain convolutional block (left), a basic residual block (center), and a pre-activated residual block (right). 63

5.1 Hillshaded snippet from the TanDEM-X Global DEM product showing (from left to right) Lake Garda, Monte Baldo, the Val d'Adige and the western Venetian Prealps. 79

5.2 Map of LVIS's platform flight trajectories (a) and of the corresponding ROIs (b) for the NASA/ESA 2016 AfriSAR campaign. 82

5.3 Example of a GEDI transect over the Redwood National Park, USA. The plot shows the individual full-waveform measurements from one of the full-power beams, and, superimposed, the corresponding vegetation metrics extracted from the L2A product. In particular, the bottom line (brown) shows the estimated ground return, the top line (green) shows the estimated return from the top of the canopy, while the center dotted line (red) shows (for reference) the location of the mean phase center measured by the TanDEM-X global DEM product. 84

5.4 Example of change in the Sentinel-1 repeat-pass coherence for increasingly larger revisit-times of (a) 12 days, (b) 24 days and (c) 36 days. The imaged location corresponds to a small region of tropical forest within the Pongara National Park in Gabon, Africa. It is possible to observe the reduction in coherence over time, especially for the mangrove forest along the northern coastline. The primary forest located more inland displays less change, confirming that at 12-days the dense forest is already completely decorrelated. . 87

5.5 Example of a Sentinel-2 acquisition showing a comparison between (a) the raw RGB composite, (b) the *Sen2Cor*-masked result and (c) the *s2cloudless*-masked result. 88

5.6 Examples of intersections between the ROIs (in colors) and (a) TanDEM-X footprints, (b) Global TanDEM-X Edited DEM product and (c) WorldCover Map products (in white). 90

5.7 Example of an alignment input and output state. The input entry (top) consists of one of more products for each source. After inflation, a collection of unique entries is created (bottom), each defined by a single product for each source. 91

5.8 Example of the sampling process. (a) represents the input classification map highlighting the NAN (white) and vegetation (green) pixels. (b) represents the patch size-dilated NAN mask. (c) represents the final patch-center sampling mask. 93

6.1 Input feature collection example, consisting of: (a) backscattering coefficient σ^0_{HH}, (b) local incidence angle θ_{lia}, (c) TanDEM-X global edited DEM, (d) interferometric coherence estimate γ_{tot}, (e) volumetric decorrelation coefficient estimate γ_{vol}, (f) height of ambiguity h_{amb}. 98

6.2 Geographical subsetting of the AfriSAR 2016 campaign test site for the generation of the training (orange), validation (green) and testing (blue) data sets. Discarded areas are identified in yellow, while TanDEM-X acquisition footprints are depicted in white. 98

6.3 Comparison of the considered *plain* (left) and *residual* (right) building blocks for the model hyperparmeter tuning. It is possible to note that, while the number of convolutional layers is the same, the order of the layers differs. As explained in Sec. 3.4.2, this is due to the residual connections requiring the operation order to be changed with respect to the plain architecture. 102

6.4 The hyperparameter tuning results for different neural architectures. Each boxplot represents the choice of two hyperameters, and captures the performance variability associated with the remaining parameter options, as well as the intrinsic randomness due to the weight initialization and patch sampling processes. The black lines show the mean metric values, the boxes contain 50% of all cases, while the whiskers represent the best and worst performing permutations. 103
6.5 The hyperparameter training tuning results for different regularization techniques. Each boxplot represents the choice of two hyperameters, and captures the performance variability associated with the remaining parameter options, as well as the intrinsic randomness due to the weight initialization and patch sampling processes. 105
6.6 Analysis for the definition of the optimal patch size. The plots depict the loss (a) and MAE (b) computed on the validation set with respect to the patch size. 106
6.7 The proposed fully convolutional deep learning model. In blue and yellow the 2D-Convolutional layers with 1×1 and 3×3 kernels, respectively, where the subscript numbers indicate the number of kernel filters. In green the batch normalization layers. The red and blue layers represent the ReLU activation functions and the Linear activation function (i.e. bypass), respectively. . . . 107
6.8 The reference CHM value distributions for the three training, validation and test subsets, as defined for the baseline scenario. 108
6.9 Training and validation results for the baseline model. (a) shows the intra-run validation variance obtained from five independent trainings, expressed in terms of RMSE for both the individual study areas and overall (Total). (b) shows the training and validation loss evolutions for the training run with the highest validation score. 109
6.10 Visual representations of prediction performance for the baseline scenario. (a) scatterplot in logarithmic scale of the predicted and reference canopy height values over all four test sites in Gabon, Africa, together with the side distributions of both the reference and the prediction for the baseline scenario. (b) Training sample distribution (blue histogram in the background - left axis) compared to the ME, which represents the regression bias (red lines in the foreground - right axis). The boxplots display the regression bias spread (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles, respectively) across the CHM range. . . 111
6.11 Prediction examples over the four considered test sites for the Baseline scenario. For each area, the prediction, the reference, the estimation error and sample vertical transect (identified by the red line), are presented from left to right, respectively. The transects in the rightmost column compare the reference CHM (in green) with the predictions (in blue), on top of ALS-derived DEM (in brown). The corresponding TanDEM-X Global Edited DEM is plotted (in red) for visual comparison, highlighting how the mean phase center lies within the canopy. 112
6.12 Baseline scenario: scatterplots of the reference CHM versus the predicted CHM for each of the four considered test-sites in logarithmic scale. 113

6.13 Analysis of the Pongara test site, which is characterized by a region of heavily underestimated tree heights: (a) CHM prediction, (b) prediction error, (c) binary mask used for the identification of the problematic area and (d) TanDEM-X Global Edited DEM. 114

6.14 The test sample distribution compared to the MAE. The boxplots display the mean absolute error spread ((5^th, 25^th, 50^th, 75^th and 95^th percentiles, respectively) across the terrain slope range. 115

6.15 The geographical composition of training, validation and test sets, for each of the four cross-validation permutations implemented for the spatial generalization analysis. 116

6.16 Spatial transfer analysis: cross-validation scatterplots for each of the four test-sites in logarithmic scale. 117

6.17 Example of underestimation over Pongara: (a) CHM prediction, (b) reference Lidar CHM, (c) prediction error and (d) corresponding land cover map. The areas covered by tall mangroves are particularly affected by a degraded performance. 118

6.18 Prediction comparison for the same area in Rabi, resulting from two different TanDEM-X acquisitions. The prediction on the left presents a significant region of underestimation (a), well aligned with a region of low backscatter values (c). The image on the right does not present either pattern (b)(d). 119

6.19 Scatterplot in logarithmic scale for the RVoG model, comparing the predicted and reference canopy height values over all four test sites in Gabon, Africa, together with the side distributions of both the reference and the prediction for the baseline scenario. 120

6.20 (a) Variation of the coherence γ_{tot} and volume decorrelation factor γ_{vol} with respect to the height of ambiguity h_{amb} for different values of CHM. (b) Distribution of h_{amb} for the acquisitions esed for training the proposed DL model. 123

6.21 Scatterplots representing the relationship between CH and σ_0 (a) and between CH and γ_{tot} (b). 125

7.1 (a) Geographic sub-setting of the AfriSAR campaign study areas into training, validation and testing, (b) Exemplary geographic sub-setting for the spatial cross-validation, considering Rabi a testing area only. 130

7.2 (a) The country-scale CHM generated using the baseline model trained using TanDEM-X from 2015-2016 only, (b) the corresponding reliability score map and (c) the binary MoA. 132

7.3 Relative sample frequency for each predictor (indicated below each image) of the model presented in Sec. 7.2, trained with TanDEM-X data acquired in 2015/2016 only. 133

7.4 (a) Height of ambiguity histogram for the training, validation and testing datasets corresponding to the 2015-2016 TanDEM-X acquisitions, (b) height of ambiguity histogram for the 2010-2011 TanDEM-X acquisitions used for the country-scale inference. (c) DEM histogram for the training, validation and testing datasets corresponding to the 2015-2016 TanDEM-X acquisitions, (d) DEM histogram for the 2010-2011 TanDEM-X acquisitions used for the country-scale inference. 134

7.5 Comparative scatterplots between LVIS- and GEDI-derived height statistics.: the RH99 metric (top-left), the DTM (top-right), the DTM and RH99 estimation errors (bottom-left), and the DSM computed from the respective DTM and RH100 metrics (bottom-left). 136

7.6 Performance boxplots between the GEDI- and LVIS-estimated RH99 metrics based on the acquisition date of the GEDI samples. 137

7.7 Performance boxplots between the GEDI- and LVIS-estimated RH99 metrics based on the relative centroid distance (RCD) between the GEDI samples and their associated pixel location. A schematic representations of possible configuration between a GEDI sample (circle), its associated pixel (square) and their RCD (red line) is shown for reference. 138

7.8 Factorial experiment assessing the impact of GEDI measurements on the proposed data-driven approach. The left column shows the result obtained after pre-training the model using the AfriSAR LVIS data, while the right one shows those obtained after fine-tuning on the GEDI measurements. The first row shows the test results evaluated on the AfriSAR LVIS dataset, while the second row shows those obtained from the GEDI measurements. 140

7.9 Height of ambiguity histogram for the extended training, validation and testing dataset, comprising TanDEM-X acquisitions covering the AfriSAR test sites from December 2010 up to the end of 2021. 141

7.10 (a) Country-scale CHM generated with the model trained using the extended dataset comprising TanDEM-X from 2010 up to 2021 (Sec. 7.4.1), (b) corresponding reliability score map and (c) binary MoA. 143

7.11 (Top row) Reliability score maps for the three different DEM-related solutions: (a) removal of the DEM as predictor, (b) substitution of the DEM with its spatial derivatives (local slope), (c) substitution of the DEM with its high-pass filtered version. (Bottom row) Corresponding MoAs. 144

7.12 Comparative estimation performance plots between the selected model for country-scale inference and the reference LVIS dataset. The scatterplot (a) displays the linear prediction agreement. The boxplot sequence (b) captures the estimation bias and spread for different reference tree height sub-ranges; the whisker contain 90% of the samples, the boxes 50%, while the black line represents the median value. The background histograms depict the relative samples distributions of the training (red) and test sets (blue), respectively. . 147

7.13 Estimation error bias and spread versus the acquisition DOY (a), the Terrain Slope (b), the h_{amb} (c) and the incidence angle (d) features, represented as a discrete sequences of boxplots. Each boxplot covers a feature sub-range, and is described by its whiskers (containing 90 percent of the samples), its box (containing 50 percent of the samples) and black median line. In the backgrounds, the respective feature distributions on the test set (in blue). . 148

7.14 (a) Country-scale mosaic of Gabon representing the CHM, generated using TanDEM-X acquisitions from the first global covered of the mission (Dec. 2010 - end of 2011). (b) Zoom-ins of the three regions included in the red boxes of the country-scale CHM mosaic. Invalid values, caused by either shadow and layover or by the unreliability of the model, are depicted in white. 149

7.15 Schematic representation of the proposed HAM workflow. The HAM acts as a pre-processing step to a generic downstream task, harmonizing the TanDEM-X single-pass γ_{tot} and γ_{vol} features. This is done to mitigate the effects of missing InSAR geometry representations during the training of the downstream task. 151

7.16 (a) Location and (b) height of ambiguity distribution for the training, validation and test sets used for the proposed HAM model. 153

7.17 Alignment examples for four acquisitions with h_{amb}, from left to right, of 28 m, 48 m, 72 m and 105 m. The model is expected to re-project the input γ_{vol} (top) to the that acquired with a target $h_{\text{amb,ref}} = 80\,\text{m}$ (bottom). As can be observed, all predictions converge to a very similar representation. 155

7.18 The scatterplots for the HAM test results over the Amazon Rainforest. In particualr, (a) shows the result for interferometric coherence γ_{tot}, and (b) shows that for the volumetric decorrelation factor γ_{vol}. 155

7.19 The h_{amb}-distributions for the (a) **Split** and (b) **Inverse Split** cases. For each case, the histograms for training, validation and testing are presented separately. 157

7.20 Scatterplots for different experiments, showing the impact of the proposed HAM solution in comparison with the reference scenario. In particular, all experiments show the results for the **split** (left) and **inverse split** (right) scenarios. (a)-(b) are the results for the reference scenario, (c)-(d) for the reference without the h_{amb} features, and (e)-(f) for the proposed HAM scenario. 158

8.1 Scatterplots of the reference DTM versus the predicted DTM for each of the five considered test-sites and overall, expressed in logarithmic scale. 165

8.2 Large-scale DTM map of the State of Gabon. The map is produced using TanDEM-X acquisitions from the 2010 and 2011 global coverage with a GSD of 25 m. 166

8.3 (a) AfriSAR AGBD dataset subsetting results and (b) scatterplot for the test set. It can be noted that for AGBD significantly less reference samples are available with respect to the CHM case. Despite this, the test results show a good agreement between prediction and reference values. 167

8.4 Scatterplots of the reference AGBD versus the predicted AGBD for each of the four considered test-sites in logarithmic scale. 168

8.5 Country-scale AGBD map of the State of Gabon. The map uses TanDEM-X images acquired between the end of 2010 and 2011, and has a 50 m GSD. While the overall results appear well harmonized, some striping artefacts remain visible in the north-east (top-right) and south-west (bottom-left) parts of the map. These are probably due to missing TanDEM-X geometries in the training set. 170

8.6 Sentinel-1-derived SAR and InSAR features acquired over the test site of Pongara, Gabon. (a) and (b) show the co- and cross-polarized backscattering maps, respectively. (c) shows the local incidence angle of the satellite computed with respect to the topography represented by the TanDEM-X Global Edited DEM in (d). (e) and (f) represent the co-polarized 12-day and 24-day repeat-pass coherences, while (g) and (h) represent the cross-polarized ones, respectively. 172

8.7 Scatterplot of the predictions for (a) the **Baseline** case and (b) **DEM Only** case. The histograms on the side report the individual distributions for predictions (right side) and for the reference (top), for the test set (blue) and training set (red), respectively. 174

8.8 Examples of Sentinel-2 bands acquired over the test site of Pongara, Gabon. All bands represent the bottom-of-atmosphere corrected surface reflectance. 177

8.9 Scatterplot for the prediction of the canopy height using Sentinel-2 data. . . 178

A.1 Schematic representation of the SAR acquisition geometry, consisting of a platform at elevation z, moving at speed v_s in the azimuth direction, and of a target at distance R from the platform. 216

A.2 Schematic representation of the SAR ambiguities for a given pulse delay time (red circle) and Doppler shift (green hyperbola). Their intersection defines four possible solutions, which can be reduced to one thanks to the SAR side-looking geometry and the sign of the Doppler shift. 216

A.3 Schematic representation of the range distance between the target (in red) and the SAR antenna (in grey) in time. Note that with respect to the slant range direction, the apparent range position R over time forms an hyperbole, whose peak coincides with the position of closest approach R_0. 218

A.4 An example of a SAR image of an agricultural area. (a) shows the characteristic speckle content related ti distributed targets in SAR imagery, while (b) shows the same area after applying a temporal multi-looking filter on a stack of 32 images. As can be seen, the *noise-like* effects of speckle are strongly reduced after the multi-looking process. 221

A.5 Plots comparing the prediction outcomes of the RVoG model for: (blue) the full model using the exponential profile and a non-zero ground-to-volume amplitude ratio, (orange) the model using the uniform vertical profile resulting from the assumption of a null extinction coefficient, (green) the sinc model resulting from the additional assumption of a null ground-to-volume amplitude ratio and (red) the linear approximation of the sinc model. (a) shows the coherence magnitude versus the height of ambiguity normalized vegetation height, while (b) shows the rotation of the complex coherence vector for increasing vegetation heights. 222

B.1 Visualization of the backpropagation example in Sec. 3.2.4. Starting from the output, the loss gradient is iteratively computed with respect to each intermediate product using the chain rule. The gradient at each intermediate product is subsequently used to compute the loss gradient with respect to the corresponding weights and biases. 227

Chapter 1

Introduction

1.1 Importance of Forest Monitoring

According to the 2020 Forest Resources Assessment of the United Nations [1], forests represent the most relevant terrestrial ecosystem by area, estimated to cover an unequally distributed 30.8% of all the Earth's land surface. They are a fundamental habitat for biodiversity, comprising an estimated 80% of all known terrestrial flora and fauna [2]. All forests combined contain more that 60'000 known tree species and are home to 80% of amphibian, 75% of bird and 68% of mammal species. Among these, mangroves effectively constitute the habitat for a plurality of marine-based organisms, including fishes, shellfishes, sea grasses and coral reefs [3]. Across the globe, it is estimated that 880 million people directly benefit from forests as a source of fuel and food, with 90% of all people living in extreme poverty directly depending on forest resources for their livelihoods [1]. Forests also significantly affect the water cycle, as the transpiration process effectively redistributes ground-water into the atmosphere, also impacting the micro-climate in the process [4] [5]. By intercepting and promoting infiltration into the ground, the presence of canopies is known to limit the intensity of water runoff during extended rainfalls [6], helping mitigating flooding events.

The state of forests constitutes a major proxy indicator for the net sinking or sourcing of carbon-based compounds, as healthy and growing plants capture carbon dioxide from the atmosphere to synthesize new carbon-based composites, which get fixed as part of the structure of plants themselves [7], but act as sources if they start to be degraded [3]. Since 1990, 9.4% of all forested areas (≈420 million hectares) has been lost due to logging activities and re-conversion processes into alternative uses [8]. While the deforestation rate has been steadily declining over the course of the last three decades,

from 2015 to 2020 it was still estimated to involve about 10 million hectares of land per year [1]. Locally, this has also reached tipping points, with forests becoming net carbon sources, rather than sinks [9]. Overall, deforestation alone is estimated to account for 11% of all green-house gas emissions, more than the entire transportation sector [10]. Additionally, forests are also affected by climate change, as it is directly linked to the frequency and duration of droughts, as well as to the extent and severity of wildfires [7]. Similarly, the increasing intensity of insect outbreaks is also observed to involve larger forested areas [7]. Globally, in 2015 1.4% of the total forested areas of the reporting countries was invaded by insects [1], with a further 0.4% of the considered forests being affected by diseases.

In this challenging scenario, a proper assessment and monitoring of forest resources worldwide becomes of paramount importance for actively informing stakeholders and decision makers, allowing for the implementation of effective mitigation strategies, such as forest conservation and restoration.

1.2 Research Gap and Work Objectives

Forest monitoring is commonly achieved by observing and tracking the evolution of proxy variables, which capture one of multiple aspects of the state of individual trees or the forest overall. Such forest parameters include, e.g., the *canopy height* (CH) in meters, the *above-ground biomass* (AGB) in megagrams of woody dry mass, the *leaf area index* (LAI) as the one-sided leaf area per unit of ground area and the *canopy cover fraction* (CCF) as the percentage of ground area covered by canopy.

In order to estimate these parameters, the most accurate way consists in directly measuring them on-site. This process can be either *invasive*, when the individual trees are cut down to assess their precise properties, or *non-invasive*, when these properties are measured or inferred without destroying the plants [11]. In both cases the measurement process is very time consuming (especially for remote areas), which makes it economically and logistically viable only for extremely limited surveys [12]. Therefore, in order to characterize forests on regional or national scales it becomes necessary to introduce statistical acquisition strategies that approximate the area of interest with representative sampling grids [13].

In this context, Earth Observation (EO) sensors represent a powerful alternative as they can typically cover significantly larger regions of interest at a given time, with spaceborne systems in particular being capable of mapping the entire surface of the

Earth. The challenge of using these systems lies in the difficulty of estimating the aforementioned forestry parameters starting from the measured quantities, for which a large pool of sensors as well as different techniques are available [14][15][16][17].

LiDAR systems, in particular airborne ones, represent the most accurate alternative to on-ground measurement campaigns, as they are capable of precisely and densely capturing the vertical forest structure, while offering superior spatial coverages [18]. Unfortunately, airborne systems remain impractical beyond the scale of regional surveys, as a trade-off between temporal and spatial coverage must be struck. Conversely, only a handful of spaceborne LiDAR and Laser Altimeter missions exist [19][20] or have existed [21], which improve the coverage offered by airborne systems at the cost of sparse measurements, commonly at significantly reduced spatial resolutions.

A further alternative is represented by spaceborne imaging systems, such Synthetic Aperture Radars (SAR) [22][23] and multi-spectral optical systems [24], which typically offer global or semi-global coverages, short revisit times and high resolutions. In all these cases, the underlying relationship between the acquired data and the desired forest parameters is typically non-linear. These systems measure bascattering/reflectance resulting from the interaction between an electromagnetic wave and the observed forest. The resulting signal is then directly dependent on the properties of the forest, such as its structure and the dielectric coefficients of its composing materials, as well as on external factors, such as the considered wavelength, the temperature and the humidity of the scene [25][26]. The intrinsic complexity of the problem limits the possibility of effectively relating the measured quantities to the forest parameters through theoretical modelling, and thus often results in lower estimation accuracies compared to LiDAR or on-ground measurements. Here, the definition of allometric equations is one of the most commonly used and straightforward approaches to forest parameter inversion from EO data [27]. It consists in fitting a parametric regression model on a sample of input and reference measurements. The resulting model is typically extremely simplistic (e.g. linear or exponential relationships) and is therefore highly tuned to the specific conditions and to the geographic location that have been selected for calibration [16]. This normally results in poor generalization capabilities and limited estimation performance. More sophisticated approaches aim at modelling the physical interaction between the measured quantity and the properties of the scene. The additional complexity introduced by these methodologies can result in more accurate estimates, but often requires large amounts of data [14], ancillary sources of information [28], or the introduction of simplifying assumptions [29] in order for the model to be inverted.

In this scenario, data-driven approaches represent a promising alternative, leverag-

ing the extensive availability of EO sensors and associated datasets to directly discern the intrinsic relationships pertaining to the desired scene properties from the data itself. The complexity of these approaches lies in the creation of a proper dataset for training, validation and testing of the model, which requires to effectively represent all the relevant and impacting aspects of the inversion problem. Conversely, the resulting frameworks are notoriously difficult to interpret and require extensive analyses for deriving meaningful insights. Recently, deep learning (DL)-based approaches have been constantly gaining attention, having experienced increasing usage and proved to be very successful when applied to remote sensing (RS) data [30][31]. Their working principle is based on repeated and iterative processing of the input data to generate increasingly complex higher levels of abstractions, which typically possess a larger descriptive and discriminative content than both the original raw input data or hand-crafted features. In the particular context of imagery, convolutional neural networks (CNN) have gained much of the attention, as an efficient tool for incorporating the spatial statistics of pixels in the close vicinity, finding particular success for classification problems [32] [33].

While in the field of computer vision deep regression techniques have already been applied with great success, with even general-purpose architectures achieving results in line with the state-of-the-art [34], comparatively little work exists regarding the regression of physical and biophysical parameters from RS data [35][36][37][38]. The reasons for this are presumably to be found in the limited availability of large quantities of reference data required for the supervised training of these models and in the complexity of the problem itself.

In light of this context, the main goal of this PhD work has been focused on the investigation of the potential of deep learning for the regression of bio-physical forest parameters from spaceborne RS data, aiming at the definition of an innovative framework that overcomes the aforementioned limitations and compromises. To this end, different EO imagery sensors, including SAR and optical datasets, have been considered and studied, with a particular emphasis on the interferometric SAR (InSAR) products generated using the DLR's bistastic TanDEM-X mission [39]. The proposed forest parameter regression framework aims at simultaneously achieving the following goals:

- *High resolution and estimation accuracy*: the generated estimates should have a fine spatial resolution for precise sampling of the forest properties, and achieve state of the art accuracy outperforming the current paradigm of physical-based modelling.

- *Timely generation of estimates*: the pre-processing and prediction pipelines should allow for the generation of time-tagged estimates, possibly requiring only a single EO product as input.

- *Large-scale deployment*: the final model should be capable of delivering consistent predictions at large-scale, allowing for the operational-oriented generation of forest parameter geomaps.

1.3 Original Contributions of the Thesis

In order to achieve these goals, a deep learning-based solution is proposed, consisting of a fully-supervised training approach for calibrating the model parameters to regress bio-physical forest properties starting from individual high resolution EO products. On top of this, the focus lies in exploring for the first time and in depth the potential synergies between deep learning methods and TanDEM-X interferometric SAR products. These, thanks to the unique single-pass capabilities of the TanDEM-X mission, are fundamentally unaffected by temporal decorrelation and are thus particularly suited for the direct estimation of the volumetric scattering effects caused by the interaction between radar waves and forests [23].

In summary, this PhD work is composed of the following original contributions:

- The definition of a novel regression framework consisting of a flexible and optimized dataset pre-processing pipeline, an original deep learning model architecture and accompanying training, validation and testing strategies, a feature importance analysis and a spatial transfer analysis.

- The framework generalization at country-scale level, consisting in the identification of the practical challenges and of three possible solutions: the use of spaceborne LiDAR dataset, the introduction of a novel alignment pre-processing step, and the re-definition of the training strategy, culminating in the generation of time-tagged country-scale products.

- An inquiry into the versatility of the framework, including the estimation of above-ground biomass density and digital terrain model, as well as the adaptability to the estimation of forest height parameters from Sentinel-1 SAR and repeat-pass InSAR features and from Sentinel-2 multi-spectral bands.

1.4 Thesis Structure

The work presented in this thesis is structured as follows:

- Ch. 2 discusses the theoretical background on the working principles of the EO sensors considered in this work. First, an overview on the Radar, SAR and InSAR technologies is given, followed by an introduction on LiDAR altimeters and, subsequently, on optical passive sensors.

- Ch. 3 presents the theoretical background regarding the considered deep learning-based family of regression techniques. Starting from the principles of estimation theory and regression problems, it introduces the generic concepts of feedforward neural networks, and finally focuses on the specific theory behind convolutional neural networks.

- Ch. 4 establishes the context of the forest parameter regression problem in the form of a short review of the state-of-the-art techniques, exploring both physical-based and data-driven modelling approaches.

- Ch. 5 first presents the primary datasets and EO missions used in this thesis. This includes the details of the DLR TanDEM-X mission, an overview of the NASA LVIS LiDAR instrument and of the associated AfriSAR campaign, a summary of the NASA GEDI mission, a presentation of the ESA Sentinel-1 and Sentinel-2 missions, as well as of the ESA WorldCover Map product. Afterwards, a high-level overview on the developed automated and performance-optimized pre-processing pipeline is presented, which is used to match and pre-process the different EO datasets for the implemented deep learning experiments.

- Ch. 6 introduces the core methodology and experiments related to the proposed forest parameter regression framework. First, the dataset as well as the training, validation and testing strategies are presented. Then the novel deep learning model is defined and hyperparameter tuning strategies are introduced. Finally, a series of experiments are presented, discussing in great detail the prediction performance of the model, as well as its generalization capabilities and the impact of the individual input features on the regression task.

- Ch. 7 initially discusses the challenges related to the practical deployment of the previously defined framework on a country-scale, and subsequently investigates in

details three potential solutions to the problem. The analysis culminates in the definition of a time-tagged, country-scale canopy height map product.

- Ch. 8 assesses the flexibility of the developed framework in the frame of similar tasks. The use of TanDEM-X data is further explored, covering the potential for the estimation of country-scale above-ground biomass density and digital terrain model products. The subsequent study examines the potential of Sentinel-1 derived SAR and repeat-pass InSAR products for the estimation of forest height, while in the final proof-of-concept study Sentinel-2 multi-spectral data is considered.

- Finally, Ch. 9 summarizes the activity of this thesis and presents the key takeaways from the experiments, together with an outlook on the future improvements and activities to be carried out.

Chapter 2

Theoretical Background and Principles of Earth Observation Sensors

In this chapter, the basic working principles of the Earth Observation sensors considered in this work are presented. Consistent with the efforts of this thesis, major attention is given to the principles of *synthetic aperture radar* (SAR) and *SAR interferometry* (InSAR). Additionally, an overview of LiDAR and Optical systems is also discussed. This overview is meant to provide the reader with the necessary theoretical concepts for understanding the research work developed in this PhD thesis.

2.1 Principles of Radar

The *radio detection and ranging* (*radar*) is a radiolocation technology that was originally introduced for military applications, famously playing a crucial role during the Battle of Britain in 1940. The working principle of the radar instrument consists in the transmission of a known electromagnetic (EM) pulsed signal and the reception of the echo that is reflected back by an illuminated target. The detection of an echo and the relative delay between its transmission and reception allow for locating the position of the target at a certain range distance R, which can be simply computed as:

$$R = \frac{c\Delta t}{2} \tag{2.1}$$

where c is the speed of light in the medium and Δt is the two-way travel time from the transmitter to the target and back to the receiver.

In practice, a radar instrument works by continuously alternating a transmission and a reception phase. A full cycle determines the *pulse repetion interval* (PRI), and it can loosely be defined as the sum of the transmitted *pulse duration* τ and of the subsequent *listening time*. The latter defines the maximum effective range of the system, as echoes received in subsequent cycles form the so-called *ambiguities*. The inverse of the PRI identifies the cycle repetition rate, which is called the *pulse repetition frequency* (PRF).

2.1.1 Range Resolution

A radar instrument can be characterized by its *range resolution* δ_{rg}, defined as the minimum separable distance between two targets along the *range direction* (i.e. the direction defined by the antenna boresight). The possible simplest signal $s(t)$ used in transmission consists in a *gated* or *truncated sinusoid*, which can be expressed as:

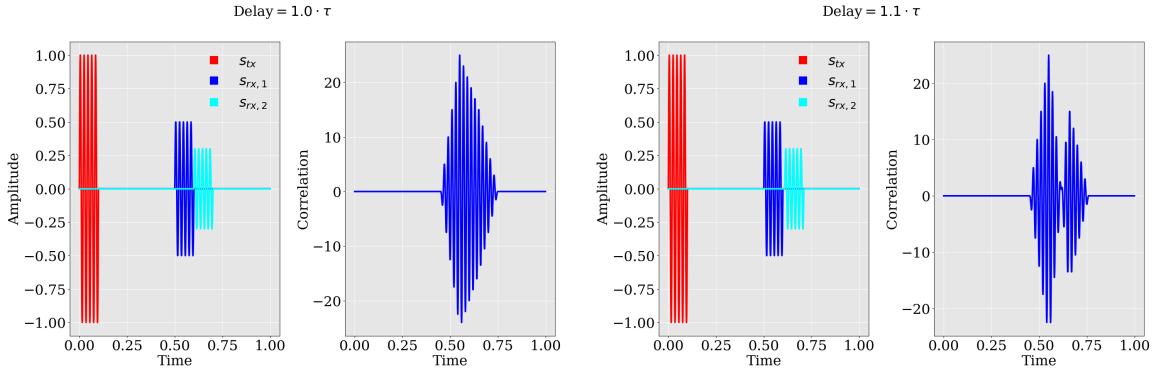
$$s(t) = e^{i2\pi f_0 t} \prod \left(\frac{t - \frac{\tau}{2}}{\tau} \right) \tag{2.2}$$

where f_0 is the frequency of the transmitted sinusoid, $\prod(\cdot)$ represents the rectangle function of duration τ and centered in $t = \frac{\tau}{2}$. For this case, the minimum detectable range separation between two targets is achieved when the beginning of the echo of the second target does not overlap with the end of the echo of the first target. The resulting range resolution δ_{rg} can thus be formalized as:

$$\delta_{rg} = \frac{c\tau}{2} \tag{2.3}$$

It follows that in order to obtain a finer resolution, the pulse duration must be reduced. This is in conflict with the practical requirement of a sufficient signal-to-noise ratio at the receiver, which ensures that the echoes are detectable. In practice, the detection of a target is performed using a *matched filter*, the optimal linear time invariant (LTI) filter for the maximization of the signal-to-noise ratio, which consists in computing the cross-correlation between the received and the transmitted signals (i.e. the convolution between the received signal and time-reversed complex conjugate of the transmitted signal).

Note that the convolution between two rectangle functions of duration τ results in a triangle function of duration 2τ, offering an alternative interpretation for Eq. (2.3).



(a) (b)

Figure 2.1: Examples of two radar returns after transmitting gated sine signals. (a) shows the signals (left) and their cross-correlation separation for two targets placed at the resolution limit (right), while (b) shows how the peaks of the targets become visible at a slightly larger separation. Note that the separations in the signal domains appear indistinguishable from each other.

The impact of this, at the resolution limit and for a larger separation, is shown in the examples in Fig. 2.1.

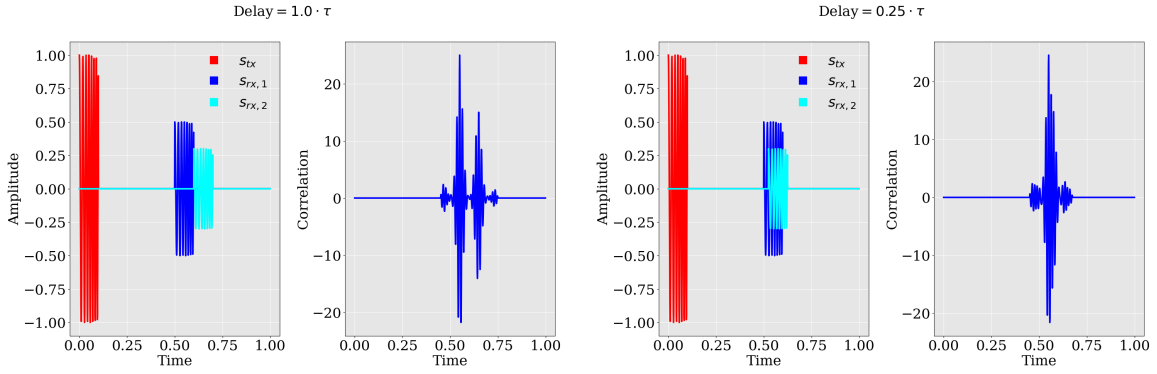
In order to improve the range resolution while achieving a desirable transmission power, it is possible to adopt a *pulse compression* technique. It consists in the manipulation of the frequency content of the transmitted signal in order to achieve a larger frequency bandwidth for the same pulse duration time. A common approach consists in transmitting a *frequency* or *phase modulated* pulse, known as a *chirp*, which can be expressed as:

$$s(t) = e^{i2\pi(f_0 t + \kappa_r t^2)} \prod \left(\frac{t - \frac{\tau}{2}}{\tau} \right) \tag{2.4}$$

where $\kappa_r = \frac{\Delta f}{\tau}$ represents the so-called *chirp rate* and Δf is the linear frequency change of the signal and corresponds to its *bandwidth* B. Following the previous case, it is possible to demonstrate that the cross-correlation between a chirp signal and a time-delayed and attenuated copy of it behaves as a sinc function with $\tau_{3dB} \approx \frac{1}{\Delta f}$, having $\tau_{3dB} < \tau$ [40]. Thus, the combination of the chirp signal and the matched filtering allows one to obtain a narrower signal peak, which leads to the following range resolution equation:

$$\delta_{rg} = \frac{c}{2B} \tag{2.5}$$

The example in Fig. 2.2 (a) highlights how, for the same pulse duration and target separation seen in Fig. 2.1 (a), the use of the chirp signal allows one to better separate the two targets after the matched filter, while Fig. 2.2 (b) shows that at the resolution



(a) (b)

Figure 2.2: Examples of two radar returns after transmitting chirp signals. (a) shows the signals and their cross-correlation separation for two targets placed at the gated sine resolution limit, while (b) shows how the peaks of the targets become indistinguishable at the chirp resolution limit.

limit the two signals become indistinguishable.

2.1.2 Radar Backscattering

Let us consider an EM plane wave illuminating an object. The impigent wave then induces vector-currents whose magnitude and orientation directly depend on the properties of the illuminated object. Following Maxwell's equations, these currents themselves produce a new electromagnetic field that is radiated through space, with amplitude and direction that differ from those of the incident wave. The described mechanism is known as *scattering*, while the specific signal fraction that is redirected back from the object toward the antenna is known as *backscattering*.

In particular, considering an EM wave that is linearly polarized in both vertical (v) and horizontal (h) directions, the transmitted and received electric fields $\boldsymbol{E}^{tx}$ and $\boldsymbol{E}^{rx}$, respectively, can be expressed as:

$$\boldsymbol{E}^{tx} = \hat{\boldsymbol{v}}_{tx} E_v^{tx} + \hat{\boldsymbol{h}}_{tx} E_h^{tx} \tag{2.6}$$

$$\boldsymbol{E}^{rx} = \hat{\boldsymbol{v}}_{rx} E_v^{rx} + \hat{\boldsymbol{h}}_{rx} E_h^{rx} \tag{2.7}$$

where $\hat{\boldsymbol{v}}_{tx}$, $\hat{\boldsymbol{v}}_{rx}$, $\hat{\boldsymbol{h}}_{tx}$ and $\hat{\boldsymbol{h}}_{rx}$ are the unit vectors that define the orientation of propagation for each of the transmitted and received polarization components E_v^{tx}, E_v^{rx}, E_h^{tx} and E_h^{rx}.

It is then possible to map the transmitted field $\boldsymbol{E}^{tx}$ to the received field $\boldsymbol{E}^{rx}$ through

the interaction with a scatterer as:

$$\boldsymbol{E}^{rx} = \begin{bmatrix} E_v^{rx} \\ E_h^{rx} \end{bmatrix} = \frac{e^{-ikR_{rx}}}{R_{rx}} \boldsymbol{S}\boldsymbol{E}^{tx} = \frac{e^{-ikR_{rx}}}{R_{rx}} \begin{bmatrix} s_{vv} & s_{vh} \\ s_{hv} & s_{hh} \end{bmatrix} \begin{bmatrix} E_v^{tx} \\ E_h^{tx} \end{bmatrix} \tag{2.8}$$

where R_{rx} is the radial distance of the receiver from the scatterer, $\frac{e^{-ikR_{rx}}}{R_{rx}}$ is the spherical propagation factor, and $\boldsymbol{S}$ is the so-called *scattering matrix*:

$$\boldsymbol{S} = \begin{bmatrix} s_{vv} & s_{vh} \\ s_{hv} & s_{hh} \end{bmatrix} \tag{2.9}$$

with the cross-polarization components $s_{vh} = s_{hv}$, given the principle of reciprocity.

2.1.3 Point Scatterer

The simplest type of scatterer is a *point target*, which, without loss of generality, can be defined as an object of physical dimensions significantly smaller than the angular resolution of the receiver.

Let us then consider a *bistatic* radar configuration, consisting of two different radar antennas. This scenario can be described by an m-polarized transmitter antenna, a point target placed at distance R_{tx} from the antenna, and an n-polarized receiver antenna placed at distance R_{rx} from the scatterer. Under the assumption of a non-dispersive medium, it is then possible to describe the power of the n-polarized received signal P_n^{rx} as:

$$P_n^{rx} = P_m^{tx} \frac{G_{tx}}{4\pi R_{tx}^2} \frac{\sigma_{nm}}{4\pi R_{rx}^2} \frac{G_{rx}\lambda^2}{4\pi}, \tag{2.10}$$

where P_m^{tx} is the power of the m-polarized transmitted signal, G_{tx} is the transmitter antenna gain, G_{rx} is the receiver antenna gain, λ is the wavelength of the transmitted EM wave and σ_{nm} is the *radar cross-section* (RCS). Eq. (2.10) is the so-called *radar equation* for the bistatic point target case. The equation can be re-written for the *monostatic* case, where the transmitter antenna coincides with receiver one, as:

$$P_n^{rx} = P_m^{tx} \frac{G^2}{(4\pi)^3 R^4} \sigma_{nm}, \tag{2.11}$$

with $G = G_{tx} = G_{rx}$ and $R = R_{tx} = R_{rx}$.

In more detail, the antenna gain captures the effective redistribution of the trans-

mitted or received power in space and can be written as:

$$G = \frac{4\pi A K_a}{\lambda^2} = \frac{4\pi A_{eff}}{\lambda^2} \tag{2.12}$$

where A is the *geometric antenna aperture*, K_a is the antenna's *efficiency* and $A_{eff} = A \cdot K_a$ is the *effective antenna aperture*. Note that for an *isotropic* source $G = 1$, meaning that the signal is equally well transmitted in (or received from) all directions.

The radar cross-section σ_{nm} of an objected determines the amount of useful impingent power that is reflected back to the receiver, and it can be defined as:

$$\sigma_{nm} = \lim_{R \to \infty} \left(4\pi R_{rx}^2 \frac{S_n^{scatt}}{S_m^{inc}} \right) \tag{2.13}$$

where

$$S_m^{inc} = \frac{P_m^{tx} G}{4\pi R^2} \tag{2.14}$$

is the power density of the incident wave at the target and

$$S_n^{scatt} = \frac{P_n^{rx}}{A_{eff}} \tag{2.15}$$

is the power density of the backscattered wave at the receiver. The limit operation is used to indicate that the backscattered signal is measured in the far-field region and can thus be approximated by a plane wave.

The RCS σ_{nm} characterizes the target in the form of the *effective area* of an isotropic source that intercepts the impingent power density and radiates it back to the receiver. In practice, the RCS of a scatterer tends to strongly vary depending on the relative orientation between the scatterer and the radar system. In this context, the RCS can be interpreted as a projected area equivalent to an ideal metal sphere that links the transmitted signal power density to the received one. Finally, it is possible to directly express the RCS as:

$$\sigma_{nm} = 4\pi |s_{nm}|^2 \tag{2.16}$$

where s_{nm} is the scattering amplitude.

2.1.4 Distributed Scatterer

In practice, it is often the case that multiple targets contribute to the received signal, in the form of a coherent sum of the individual backscattered signals within a single resolution cell [41]. Under the assumption of no predominant scattering mechanism,

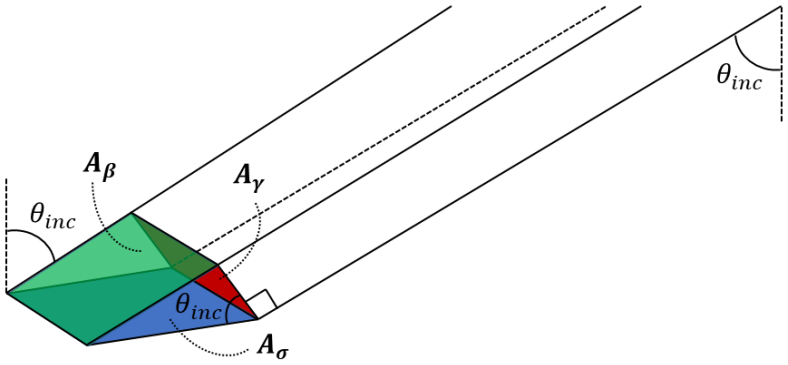


Figure 2.3: Reference geometry for the backscattering normalization areas A_β, A_σ and A_γ.

Eq. (2.11) can be generalized for an illuminated area as:

$$P_n^{rx} = \iint_{A_\sigma} P_m^{tx} \frac{G^2}{(4\pi)^3 R^4} \sigma_{nm}^0 dA_\sigma, \tag{2.17}$$

where A_σ is the area illuminated by the transmitted signal and σ_{nm}^0 is the *backscattering coefficient*. The latter describes the RCS per unit area and can thus be formalized as:

$$\sigma_{nm}^0 = \frac{\sigma_{nm}}{A_\sigma} \tag{2.18}$$

which represents the RCS of a distributed target of area A_σ, normalized by the area itself.

For reference, Fig. 2.3 shows the relationship between different backscattering normalization areas. It is often convenient to express the backscattering coefficient in terms of *acquisition incidence angle* θ_{inc} between the incidence direction of the transmitted signal (known as *slant range* direction) and the normal to the illuminated surface. It is pointed out that this notation assumes the simpler case of a flat surface (known as *flat Earth*), but more in general it becomes useful defining also the *local incidence angle* θ_{lia} in the presence of topography.

The normalization of the radar backscatter with respect to the projection of the illuminated area onto the slant range direction A_β results in a quantity known as the *radar brightness* β_{nm}^0 that can be expressed as:

$$\beta_{nm}^0 = \frac{\beta_{nm}}{A_\beta} = \frac{\sigma_{nm}^0}{\sin(\theta_{\text{inc}})} \tag{2.19}$$

where β_{nm} represents the ratio between the scattered signal power and the incident signal power at ground level.

Finally, a common alternative representation consists in normalizing the backscattering coefficient with respect to the illuminated area projected onto the perpendicular plane to the slant range A_γ. The resulting quantity γ^0_{nm} can be expressed as:

$$\gamma^0_{nm} = \frac{\beta_{nm}}{A_\gamma} = \beta^0_{nm} \tan(\theta_{\text{inc}}) = \frac{\sigma^0_{nm}}{\cos(\theta_{\text{inc}})} \tag{2.20}$$

2.2 Principles of SAR Imaging and Interferometry

The focus of this section is the *synthetic aperture radar* (SAR), an imaging radar system which relies on the movement of the antenna to sample an area on ground. In particular, a SAR system is characterized by two distinctive features. The first is the *side-looking* geometry of the sensor, which is required to avoid the range ambiguities otherwise intrinsic of a *nadir-looking* system (i.e. *left-right ambiguity*). The second aspect consists in the exploitation of the platform movement to synthesize a larger antenna aperture with respect to a *real aperture radar* (RAR) (e.g. *side-looking airborne radar*, SLAR), achieving in this way a higher resolution.

Note that it is assumed that the power of the individual scattered signals is significantly lower than that of the incident wave, thus allowing to neglect the interaction between targets and to treat them independently (*Born approximation*). For the sake of simplicity, the discussed concepts will be elaborated in terms of individual point targets, but they can be extended to complex targets as well.

Furthermore, while multiple SAR acquisition modes exist, all presented concepts are based on the consideration of a *stripmap* acquisition mode.

2.2.1 SAR Acquisition Geometry

Similarly to other radar systems, a SAR instrument works by repeatably alternating the transmission of an EM wave and the collection of the echoes backscattered by the targets. Crucially, each consecutive cycle covers a different portion of the scene because of the movement of the platform. The corresponding acquisition geometry for a SAR system can be seen in Fig. 2.4, and it is defined by the platform movement direction (*along-track* or *azimuth* direction), by the look direction of the radar (*across-track* or *slant-range* direction), and by its projection on ground (*ground-range* direction).

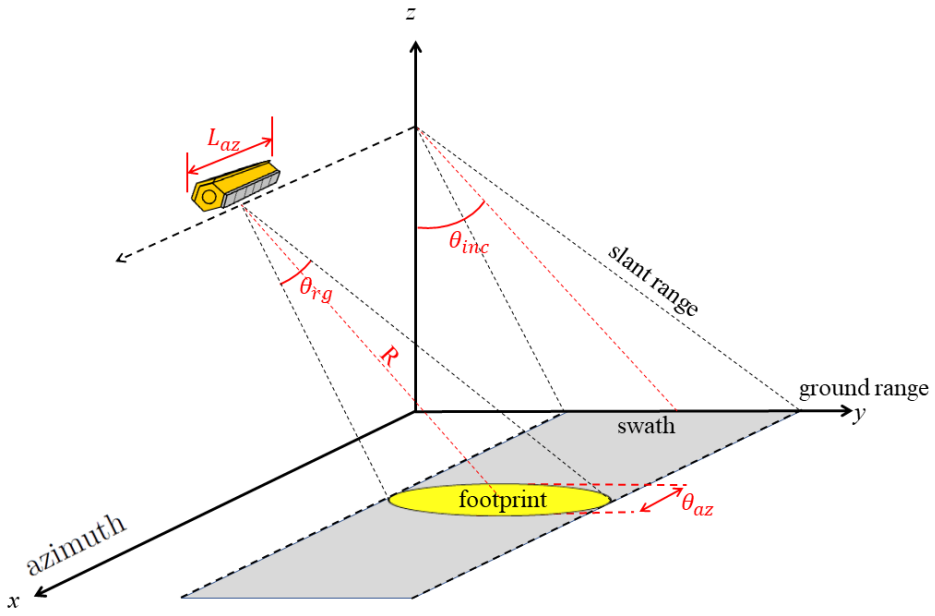


Figure 2.4: Representation of the SAR acquisition geometry.

Range Resolution

The range resolution δ_{rg} defined for a chirp signal in Eq. (2.5) corresponds to the SAR slant-range resolution δ_{sr}, while the corresponding ground-range resolution δ_{gr} can be inferred from the trigonometric properties of the acquisition geometry as:

$$\delta_{rg} = \frac{c}{2B\sin(\theta_{\text{inc}})} \tag{2.21}$$

Azimuth Resolution

When considering a real aperture radar antenna, two targets can be resolved in the azimuth direction if they do not simultaneously fall within the antenna footprint. Thus, the azimuth resolution limit is given by:

$$\delta_{az} = R\theta_{az} = R\frac{\lambda}{L_{az}} \tag{2.22}$$

where $\theta_{az} = \lambda/L_{az}$ is the azimuth beam width angle for an antenna with azimuth size L_{az}.

To overcome this limitation, SAR systems exploit the *Doppler effect* caused by the relative motion between the platform and the target. It can be then shown (see

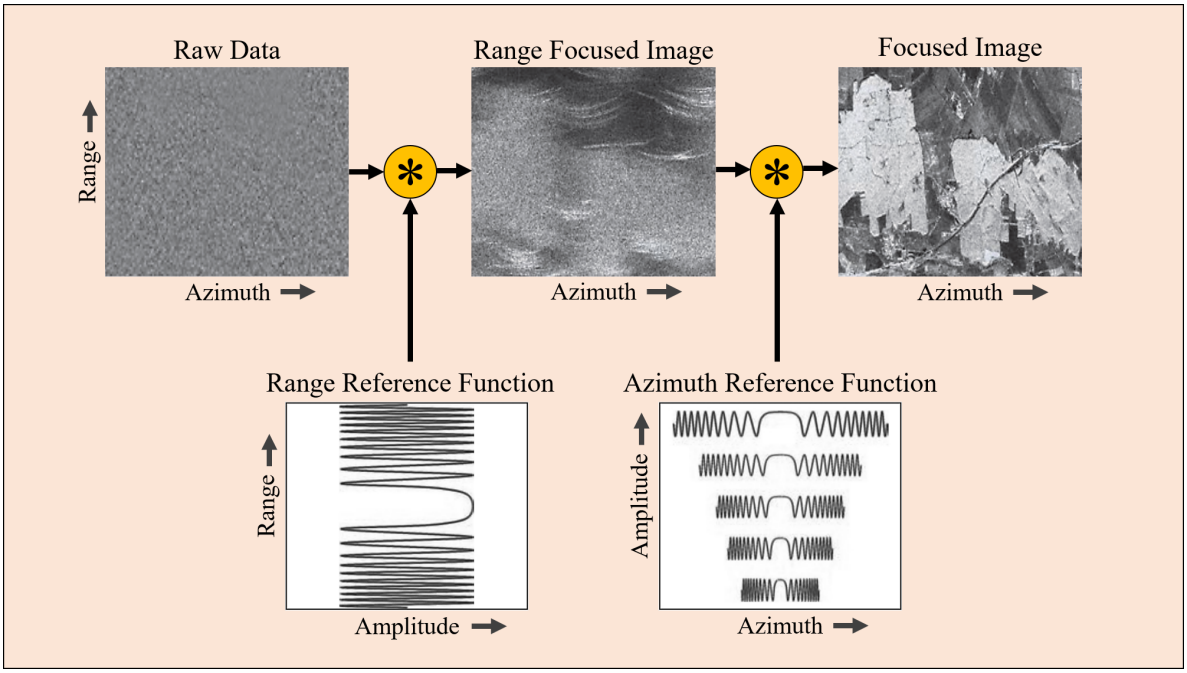


Figure 2.5: Schematic representation of the SAR focusing process. The star symbol represent the convolution operation. It is possible to observe that the raw data domain is unintelligible, and requires focusing for the different targets to become recognizable. This is due to the large integration-windows across which each signal is spread.

Appendix A.1) that the azimuth resolution becomes:

$$\delta_{az} = \frac{L_{az}}{2} \tag{2.23}$$

with the notable consequence that the azimuth resolution does not depend anymore on the range distance, but only on the antenna's physical dimension.

2.2.2 SAR Image Formation

The imaging process of a SAR system consists in the cyclic transmission and reception of individual pulses, the returns of which are digitized and stored on board in the *signal memory*. The resulting 2D array is typically referred to as *raw data*, and it effectively represents the superposition of the smeared returns of individual targets. In order to generate an intelligible image, it is necessary to *focus* the energy of each target into the resolution cell corresponding to its range and azimuth location. In the following, the presented processing steps will refer to the classical *Range-Doppler Algorithm* (RDA), one of the most used SAR processing algorithms, which is schematized in Fig. 2.5.

Range Compression

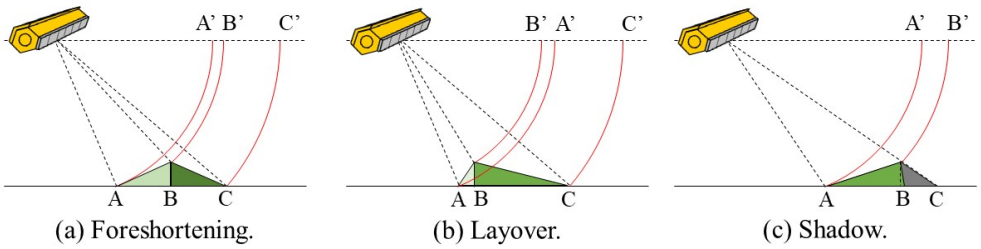
The *range compression* or *range focusing* operation aims at undoing the signal dispersion in the range direction caused by the transmission of a chirp signal of a given duration. As discussed in Sec. 2.1, the chirp compression operation is performed using a matched filter, which computes the cross-correlation function between the received waveform and the transmitted chirp signal. This operation is performed independently on each raw data column, as effects of azimuth cell migration are assumed to be negligible [41].

Azimuth Compression

As shown in Eq. (2.22), the signal return of a particular target is spread across the entire antenna beamwidth in the azimuth direction. Furthermore, in order to exploit the full Doppler bandwidth for the generation of a synthetic aperture, it is necessary to sample the entire azimuth antenna beamwidth with at least two measurements (i.e. the *Nyquist frequency*). This sampling frequency is directly determined by the PRF of the system, which thus defines both the swath width of the acquisition and the maximum Doppler bandwidth. As a consequence, the power corresponding to each target is effectively spread across multiple memory cells in the azimuth direction. It is therefore necessary to compress the signal in the azimuth direction in order to reconstruct the reflectivity of each target at the correct position. It can be shown (see A.2) that the relative motion between the platform and the target produces a range-dependent linear FM (i.e. a chirp signal). Similarly to range focusing then, azimuth focusing is ultimately achieved by computing the cross-correlation between each azimuth raw-data line and the expected azimuth chirp, uniquely defined for a given range R.

Absolute SAR Calibration

After focusing, the values of each pixel in the resulting SAR image are identified by the so-called *digital number* DN. This quantity represents the amplitude of the signal but it is affected, among others, by distortions caused by the range and azimuth compression, the antenna pattern, the system electronics and the effects of propagation [41]. In order to properly generate estimates of physical properties, such as the target backscattering coefficient, it necessary to compensate the measured amplitudes for any remaining effects. For this *calibration* process, the DN is compared against that of reference targets of known properties. Commonly used are passive targets, such as *corner reflectors*, and active targets, known as *transponders*. Comparing the measured intensity DN^2 of these objects against their known values, leads to the determination



(a) Foreshortening. (b) Layover. (c) Shadow.

Figure 2.6: Schematic representation of possible geometric distortions introduced by the side-looking geometry of SAR. Note that when a slope is facing towards the antenna, more targets get compressed into the same resolution cell, thus making the return appear brighter. The opposite is true for slopes facing in the opposite direction.

of an absolute calibration constant, such that:

$$\beta_{0,nm}^{ref} = k_{nm}^{abs} DN^2 \tag{2.24}$$

where $\beta_{0,nm}$ is the estimated radar brightness in nm-polarization and k_{nm}^{abs} is the corresponding calibration constant.

Geometric Distortions

A consequence of the SAR side-looking acquisition geometry is the introduction of geometrical distortions, particularly in the presence of strong topography, as it can be seen in Fig. 2.6.

In presence of a slope oriented towards the sensor, the relative distance of the targets appears to be compressed within the same resolution cells, a phenomenon known as *foreshortening*.

If the slope inclination is larger than the incident angle of the sensor, the relative position between points on top and at the bottom of the slope appears inverted, resulting in what is known as *layover*.

Finally, targets on a slope facing away from the sensor with an inclination larger than the incident angle of the sensor are not illuminated, resulting in *shadows*.

2.2.3 Principles of SAR Interferometry

The goal of *interferometric SAR* (InSAR) is to gain information about the vertical position of a target by acquiring measurements from two different sensor locations (in space and/or in time). Crucially, these acquisitions can be performed simultaneously

in the case of *single-pass* interferometry, or in two different time intervals in the case of *repeat-pass* interferometry. Furthermore, in presence of simultaneous acquisitions, it is possible to distinguish between a *monostatic* configuration with both sensors transmitting and acquiring independently from each other, and a *bistatic* setup with only one sensor transmitting and both receiving.

Let us consider the corresponding signals received by two distinct single-channel SAR measurements, which can be presented as:

$$\begin{aligned} s_1 &= A_1 e^{-i\frac{4\pi}{\lambda}R_1+\phi_{obj,1}} \\ s_2 &= A_2 e^{-i\frac{4\pi}{\lambda}R_2+\phi_{obj,2}} \end{aligned} \tag{2.25}$$

where the first phase term $4\pi R/\lambda$ is related to the two-way travel distance between the SAR antenna an the target, while the second term ϕ_{obj} captures the scattering process of the target itself. Assuming that for the same target $\phi_{obj,1} = \phi_{obj,2}$, it is possible to cancel the scattering phase terms by multiplying the first signal s_1 by the complex conjugate of the second signal s_2, such that:

$$s_1 s_2^* = A e^{-i\frac{2m\pi}{\lambda}\Delta R} \tag{2.26}$$

where $m = 1$ for a bistatic configuration and $m = 2$ for a monostatic one. The key observable of the previous expression is defined by the *interferometric phase* term:

$$\phi_{if} = \angle s_1 s_2^* = -\frac{2m\pi}{\lambda}\Delta R + 2\pi k, \quad k \in \mathbb{Z} \tag{2.27}$$

Here, the second phase term $2\pi k$ is added to capture the 2π ambiguity of the measured phase. From Eq. (2.27), it follows that ϕ_{if} repeats itself for every range difference change $\Delta R = \lambda/2$, allowing for multiple solutions. This is known as the *phase unwrapping problem* [42].

Note that the observation of the target from two different points corresponds to some element of change in ΔR that can be reversed from the particular acquisition geometry to infer knowledge about the position of the target. This separation between the imaging points is known as the *interferometric baseline* and, depending on the application, it can be either *spatial*, *temporal* or both.

Depending on the chosen acquisition geometry, it is therefore possible to distinguish between three different types of InSAR: *across-track*, *along-track* and *differential* interferometry [42]. In the context of this thesis, the focus will be on across-track

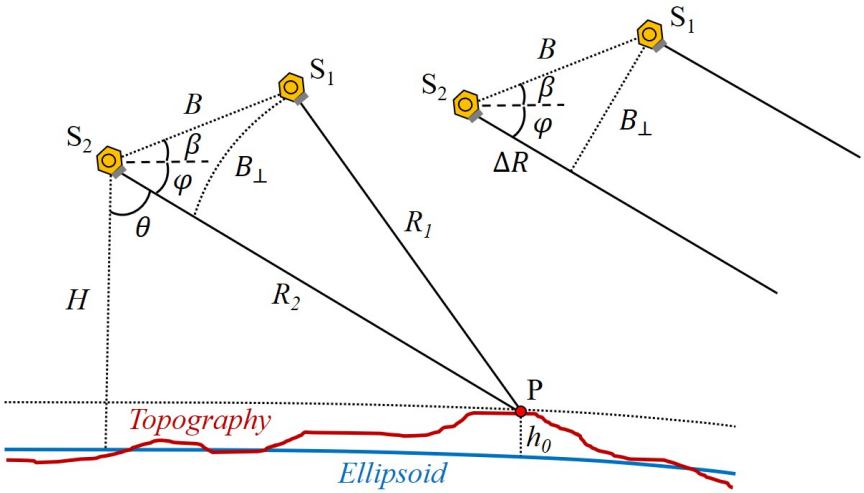


Figure 2.7: Schematic overwiew of the acquisition geometry for across-track SAR interferometry. Note that the view in the top-right shows the equivalent simplified geometry for the far-field approximation.

interferometry only.

Across-Track Interferometry

The measurement of the vertical position of a target with respect to the reference Earth ellipsoid can be achieved by observing the target with a spatial baseline defined on the normal plane to the flight direction (i.e. the plane defined by a fixed azimuth value). This corresponds to the characteristic acquisition geometry for *across-track interferometry* shown in Fig. 2.7. In particular, it is possible to define the elevation h_0 of a target P over the reference ellipsoid as:

$$h_0 = H - R_2 \cos\theta_{\text{inc}} \tag{2.28}$$

where H is the known elevation of the sensor S_2, R_2 is the known distance between the sensor and the target and θ_{inc} is the unknown incidence angle of the sensor. In order to estimate the elevation of the target it is thus necessary to first estimate the incidence angle. This can be obtained from triangulation, starting from the triangle of vertices P, S_1 and S_2, where $R_1 = \overline{PS_1}$ and $R_2 = \overline{PS_2}$ are the slant-range distances between the target and the respective sensor positions and $B = \overline{S_2S_1}$ is the interferometric baseline. The parallel and orthogonal projections onto the slant-range dimension define



Figure 2.8: Simplified representation of the two-way fractional phase for each point in the normal plane to the flight direction, as transmitted and measured by the first sensor (a), the second sensor (b) and as measured by the interferometer (c). Note that the horizontal axis represents the flat Earth, while the black line shows an example of topography. It is then possible to compare the relative change in measured interferometric phase between a given point on the horizontal axis and it corresponding point on the topographic curve.

the *parallel baseline* $B_{||}$ and the *orthogonal baseline* $B_{\perp}$, respectively. We can also define the angle of vertical separation between the two sensors from the local parallel to the ground as β, and the angle of vertical separation between the target and the sensor at position S_3 as $\varphi = \pi/2 - \theta_{\text{inc}}$.

By applying the law of cosines, it is possible to establish that:

$$R_1^2 = R_2^2 + B^2 + 2BR_2\cos(\varphi + \beta) \tag{2.29}$$

From the acquisition geometry, we can define $R_2 - \Delta R = R_1$ and $\varphi = \pi/2 - \theta_{\text{inc}}$, which leads to:

$$\sin(\theta_{\text{inc}} - \beta) = \Delta R + \frac{B^2}{2BR_2} - \frac{\Delta R^2}{2BR_2} \tag{2.30}$$

By observing that in practice $\Delta R << R_1$ and $B << R_1$, it is possible to simplify the geometry using the parallel-ray approximation. This leads to:

$$\sin(\theta_{\text{inc}} - \beta) \approx \Delta R = \frac{\phi_{int}\lambda}{2m\pi} \tag{2.31}$$

where ΔR can be directly estimated from the measured interferometric phase by inverting Eq. (2.27). Thus it becomes possible to compute the elevation h_0 of the target at P, by inverting Eq. (2.31) and estimating θ_{inc}, and by subsequently inserting it in

Eq. (2.28). With an alternative interpretation, the determination of the solution is given by the intersection of the circle defined by the range distance R (as defined by the a single SAR image) with the hyperbole defined by the baseline and the range difference ΔR (as defined by the interferometric operation). Fig. 2.8 shows a representation of the two-way phase propagation from and to each sensor and the corresponding hyperboles of equal interferometric phase.

A final crucial element left to address is the determination of the vertical resolution of the SAR interferometer. It can be shown [43] that the vertical sensitivity, also known as the *vertical wavenumber* κ_z, can be expressed as:

$$\kappa_z = \frac{\partial \phi}{\partial h} = \frac{2m\pi\Delta\theta}{\lambda \sin\theta_{\text{inc}}} \tag{2.32}$$

where $\Delta\theta$ is the angle defined by the sensor ranges. Applying a first order Taylor approximation for small angles (i.e. $B << R$), leads to the alternative formulation:

$$\kappa_z \approx \frac{2m\pi B_\perp}{\lambda R \sin\theta_{\text{inc}}} \tag{2.33}$$

which expresses the sensitivity in terms of of the perpendicular baseline component $B_\perp$. Crucially, another interpretation of *Eq.* (2.32) can be offered by the *height of ambiguity* h_{amb}, which expresses the vertical sensitivity in terms of height measurement corresponding to a full 2π phase cycle and is defined as:

$$h_{amb} = \frac{2\pi}{\kappa_z} = \frac{\lambda \sin\theta_{\text{inc}}}{m\Delta\theta} \approx \frac{\lambda R \sin\theta_{\text{inc}}}{mB_\perp} \tag{2.34}$$

where, for $B << R$, it follows that the sensitivity h_{amb} of the interferometer is coarser than the wavelength λ.

The interferogram

So far, the interferometric inversion process detailed in Sec. 2.2.3 has not analyzed the image formation process of the *interferogram*. It is then possible to generate the interferogram, which consists in the two-dimensional image of the interferometric phase ϕ_{if}, computed from Eq. (2.26) and defined as in Eq. (2.27). It can be shown that the measured interferometric phase is composed of two terms:

$$\phi_{if} = \phi_{topo} + \phi_{fe} \tag{2.35}$$

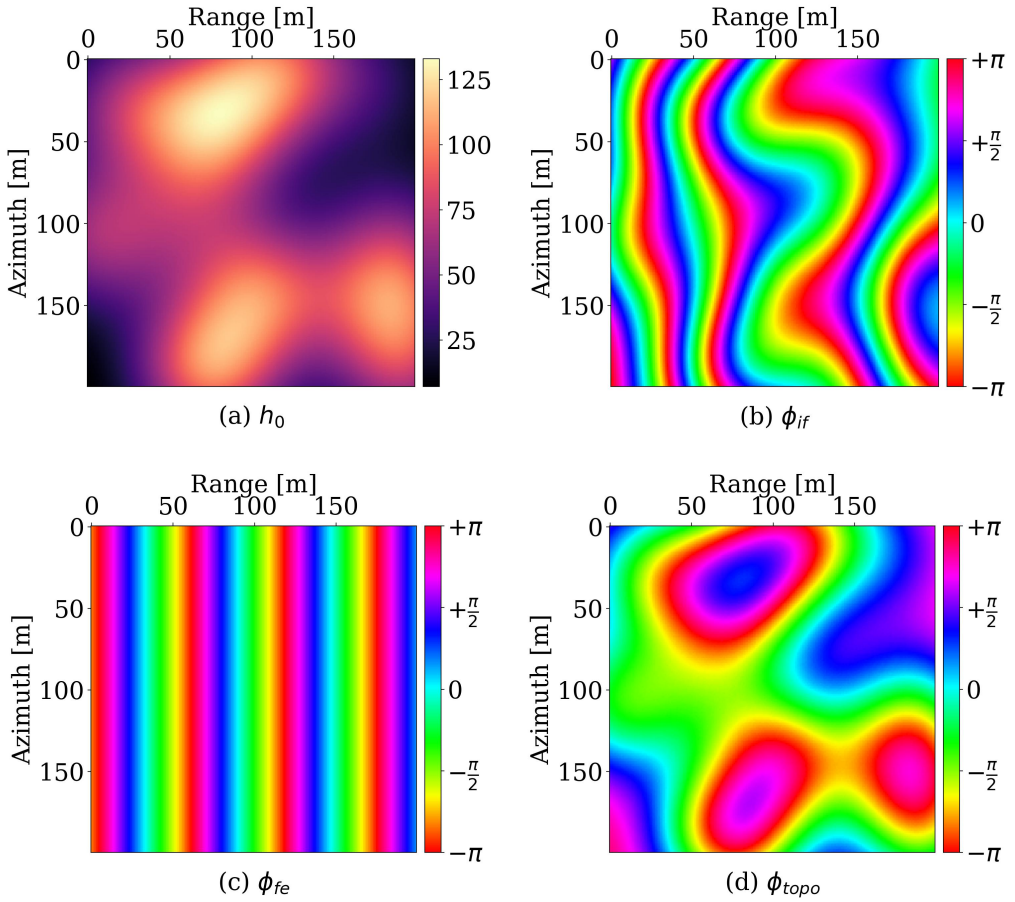


Figure 2.9: Representation of a simulated InSAR scene, showing (a) a synthetic DEM h_0, (b) the corresponding interferometric phase ϕ_{if}, (c) its flat-earth term ϕ_{fe} and (d) the topographic term ϕ_{topo}.

where ϕ_{topo} is the contribution of the topography, and ϕ_{fe} is the *flat-Earth* component. The latter accounts for the interferometric phase increase along the slant-range direction due to the off-nadir look-angle between the sensor pair and the reference ellipsoid, regardless of the presence of topography. Fig. 2.9 shows an example of a simulated DEM, the measured interferogram and its flat-Earth and topographic phase components.

After flat-Earth removal (see A.4), it is necessary to solve the phase ambiguity expressed in Eq. (2.27). To this end, a *phase unwrapping* algorithm is used to estimated the absolute phase term ϕ_{abs} from the wrapped interferometric phase ϕ_{if}. The first robust solution to the phase unwrapping problem was proposed by *Goldstein et al.* in 1988 [44]. It consisted in determining closed loops of phase terms which belonged to the same phase cycle, essentially building phase isolevels which could then be used as limits for the cumulative integration of the phase terms. Multiple alternative algorithms have

been proposed over the years, a selection of which can be found in [45], [46], [47], [48] and [49].

The resulting unwrapped phase estimate ϕ_{abs} can then be used in the process detailed in Sec. 2.2.3 to generate the final *digital elevation model* (DEM), representing the vertical elevation coordinate of each resolution cell above the reference height.

As a final remark, there is a further distinction to be made between the *digital terrain model* (DTM), which is the particular case of a DEM which measures the elevation at the level of the terrain, and the *digital surface model* (DSM), which instead measures the elevation on top of local non-topographic elements, such as trees or snow. It is therefore the case that DSM $\geq$ DTM, with DEM representing the umbrella term.

The Interferometric Coherence

Up to this point, it was assumed that the scattering phase terms $\phi_{obj,1}$ and $\phi_{obj,2}$ of the SAR signals in Eq. (2.25) perfectly described the same process, thus leading to the interferometric phase being only dependent on the topography. In practice, this assumption does not hold and consequently introduces noise in the measured interferogram. Similarly, non-ideal aspects of the acquisition process also introduce their own noise contributions into the interferogram.

The metric used to quantify the quality of an interferometric process is the *interferometric coherence* γ [50], which consists in the normalized cross-correlation coefficient of the complex SAR image pair s_1 and s_2, and can thus be defined as:

$$\gamma = \frac{E[s_1 s_2^*]}{\sqrt{E[|s_1|^2]E[|s_2|^2]}}, \tag{2.36}$$

where the phase of the complex coherence $\angle\gamma = \phi_{if}$ is the interferometric phase, while the magnitude $|\gamma|$ is the degree of coherence and it quantifies the level of correlation between the imagine pair.

Note that the stochastic nature of the SAR signal (as described in Sec. A.3) requires γ to be estimated rather than determined. Under the assumption of stationary ergodic processes, γ is commonly estimated by the *sample coherence* $\hat{\gamma}$ [51]. Given a number L of independent samples, the sample coherence is defined as:

$$\hat{\gamma} = \frac{\sum_W s_1(i,j)s_2^*(i,j)}{\sqrt{\sum_W |s_1(i,j)|^2 \sum_W |s_2^*(i,j)|^2}}, \tag{2.37}$$

which leads to an asymptotically unbiased estimate as the size L of the ensemble gets

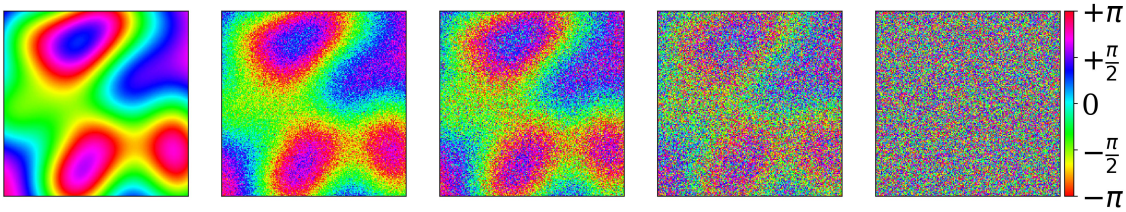


Figure 2.10: Impact of different levels of coherence on the measured interferogram. As the coherence decreases, the phase noise increases, leading to less clearly defined fringes and more severe phase unwrapping errors.

larger. Fig. 2.10 shows the impact of different levels of coherence on the measured interferogram.

The reasons leading to a decrease in the interferometric coherence are now briefly introduced. Broadly speaking, the decorrelation effects can be grouped into three overarching categories: system-related γ_{sys}, scattering-related γ_{scatt} and temporal-related γ_{temp} effects, leading to the following:

$$\gamma = \gamma_{\text{sys}} \cdot \gamma_{\text{temp}} \cdot \gamma_{\text{scatt}} \tag{2.38}$$

The system-related term γ_{sys} captures the effects linked to the non-ideal SAR system and processing implementations, such as the thermal noise of the receiver electronics affecting the SNR (γ_{SNR}), the noise introduced by the analog-to-digital (ADC) conversion and quantization process (γ_{quant}), or the impact of azimuth and range ambiguities (γ_{amb}).

The scattering-related term γ_{scatt} captures the the impact of the InSAR acquisitions geometry on the phase stability of the scatterers. It comprises the effects in the range direction (γ_{rg}), such as the perpendicular baseline induced spectral shift and coregistration errors in range, as well as in the azimuth direction (γ_{az}), such as the shifts in the Doppler spectra and residual coregistration errors in azimuth, and the effects due to vertically distributed volume scatterers (γ_{vol}).

The time-related term γ_{temp} captures geometric and dielectric changes of the target that occurred in the interval between the the acquisition of the interferometric image pair.

It is therefore possible to expand Eq. (2.38) to the following relationship:

$$\gamma = \gamma_{\text{SNR}} \cdot \gamma_{\text{quant}} \cdot \gamma_{\text{amb}} \cdot \gamma_{\text{rg}} \cdot \gamma_{\text{az}} \cdot \gamma_{\text{vol}} \cdot \gamma_{\text{temp}} \tag{2.39}$$

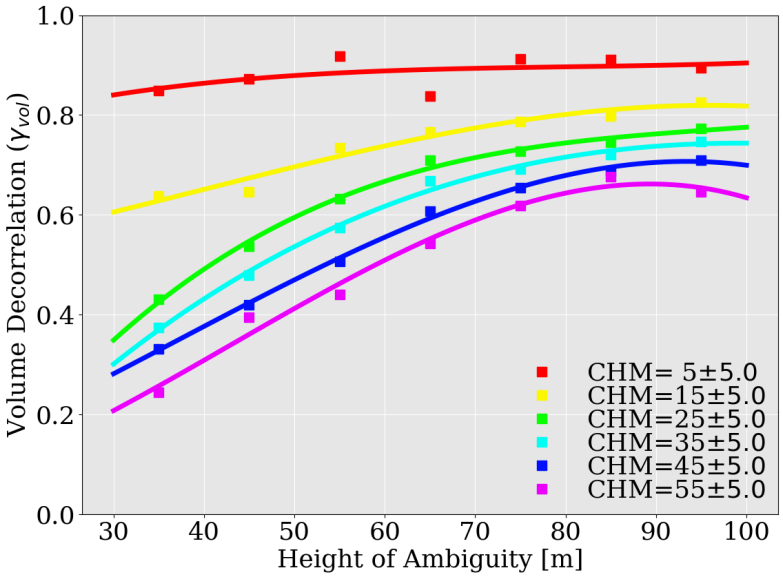


Figure 2.11: Relationship between the volume decorrelation coefficient γ_{vol} and the acquisition height of ambiguity h_{amb} for different tree height ranges, sampled using TanDEM-X derived measurements acquired across the tropical forests of Gabon (Africa).

Crucially, it is possible to estimate the individual factors and thus to compensate their effects on the total coherence γ [52].

Volume Decorrelation and Height Inversion

Of particular interest to this work is the derivation of the *volumetric decorrelation* coefficient γ_{vol}, consisting in a vertical distribution of scatterers, which, due to the orthogonal baseline separation, causes a difference in the coherent sum of the individual scatters. This effect can be caused, for example, by structures such as sand, snow, ice and vegetation. In particular, assuming that the scatterer power variation in the vertical direction z is captured by a generic *vertical reflectivity function* $F(z)$, it is possible to express the received interferometric signal as [43]:

$$s_1 s_2^* = \int_{z_0}^{z_0+h_v} F(z) e^{j\kappa_z z} dz = e^{j\beta_z z_0} \int_0^{h_v} F(z') e^{j\kappa_z z'} dz' \tag{2.40}$$

where z_0 is the elevation of the ground, h_0 is the height of the volumetric scattering profile and $z' = z - h_0$. The volume decorrelation factor can then be obtained from [43]:

$$\hat{\gamma}_{vol}(\vec{w}) = e^{i\kappa_z z_0} \frac{\int_0^{h_v} F(z', \vec{w}) e^{i\kappa_z z'} dz'}{\int_0^{h_v} F(z', \vec{w}) dz'} \tag{2.41}$$

Note that the resulting complex volumetric coherence is directly linked to the height h_0 of the vertical scattering profile. In principle, it is then possible: first, to estimate the volume decorrelation coefficient from the measured sample coherence by compensating the latter for the remaining factors detailed in Eq. (2.39); secondly, to derive the volume height by inverting Eq. (2.41). This requires the formalization of the vertical reflectivity profile $F(z)$.

2.3 Principles of LiDAR

The *Light Detection and Ranging* (*LiDAR*) is an active remote sensing technology used, among other applications, for the determination of target ranges starting from the two-way time of flight of a transmitted electromagnetic laser signal, which is reflected by the target and subsequently collected at the receiver. LiDAR and radar systems primarily distinguish themselves through the EM spectrum in which they operate. In particular, LiDAR covers the range from near infrared ($\lambda \approx 1\,\mu m$) up to ultraviolet light ($\lambda \approx 1\,nm$). The transmitter of a LiDAR system is formed by a *laser* (*light amplification by stimulated emission radiation*). It consists of a *gain medium*, a resonance cavity device formed by specif materials that, when excited by an external source, emit photons of a particular energy. An optical reflector system then forces the photons to repeatedly travel through the medium, triggering the emission of even more photons and coherently adding them up in the process, until a signal of a characteristic dominant frequency is obtained.

2.3.1 Transmission Equation

Similar to radar systems, also for LiDAR it is possible to define a transmission equation, such that:

$$P_{rx} = \frac{P_{tx}}{r\pi R^2} \frac{4\pi}{(\pi/4)\vartheta_{bw}^2} \sigma \frac{1}{2\pi R^2} A O_{eff} \tag{2.42}$$

where P_{rx} is the received power, P_{tx} is the transmitted power, R is the range distance, ϑ_{bw} is the laser beam width, σ is the laser cross-section, A is the aperture of the receiver optics and O_{eff} is the efficiency of receiver optics. Note that the laser beam width is given by:

$$\vartheta_{bw} = \frac{\lambda}{D} = \lambda \sqrt{\frac{\pi}{4A}} \tag{2.43}$$

where D is the diameter of the optical lens and λ is the wavelength of the transmitted signal. The LiDAR equation can thus be re-written as:

$$P_{rx} = \frac{8P_{tx}A^2O_{eff}\sigma}{\pi^3 R^4 \lambda^2} \tag{2.44}$$

Similar to radar, also for LiDAR it is possible to distinguish between *pulsed* and *continuous wave* LiDAR, but in the following we will only focus on the former. The pulsed LiDAR performs ranging measurements by repeatedly alternating between the transmission of a gated signal waveform of duration τ and a listening period used to receive the echoes with a given PRF. Accordingly, it is possible to define the range distance R of a target from its two-way travel time Δt as:

$$R = \frac{c\Delta t}{2} \tag{2.45}$$

Furthermore, the range resolution can be computed as:

$$\delta_{rg} = \frac{c\tau}{2} \tag{2.46}$$

which is commonly in the order of tens of centimeters or better.

2.3.2 Point Cloud LiDAR

In the context of remote sensing, *scanning* LiDAR systems are most commonly used for imaging purposes, which, in the particular case of an airborne mounted sensor, are also known as *airborne laser scanning* (ALS). The nadir-looking acquisition process relies on either a single beam steered across the flight direction, or on multiple, static, across-track oriented beams. The signal resulting from a single transmitter pulse is characterized by one or multiple signal *returns*. In presence of vegetation, each peak corresponds to the interaction with a different part of a plant. The *first return* then typically corresponds to the top of the canopy, while the *last return* corresponds to the position of the ground. An imaging LiDAR system can thus be used to produce DTMs (i.e. the last return image), DSMs (i.e. the first return image), or CHMs (i.e. the difference between the first and last return images). In particular, each recorded return corresponds to a specific three-dimensional (3D) location, which is accurately determined from the range time and the position of the acquisition platform through the combination of inertial sensors and a GNSS system. This typically results in a 3D *point cloud* that irregularly samples the scene, often quantified in terms of samples per unit

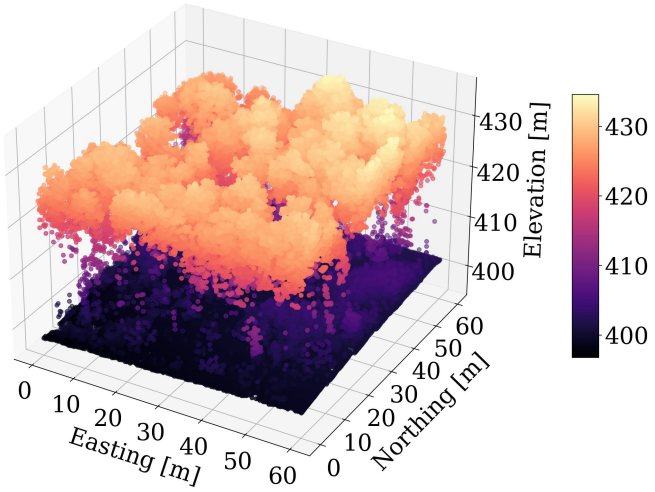


Figure 2.12: Example of ALS-derived point cloud acquired over forested area in the Hainich National Park, Germany. The $60 \times 60\,\mathrm{m}^2$ scene contains $\approx 10^5$ samples, for an average sample density of $\approx 27\,\mathrm{pts/m}^2$.

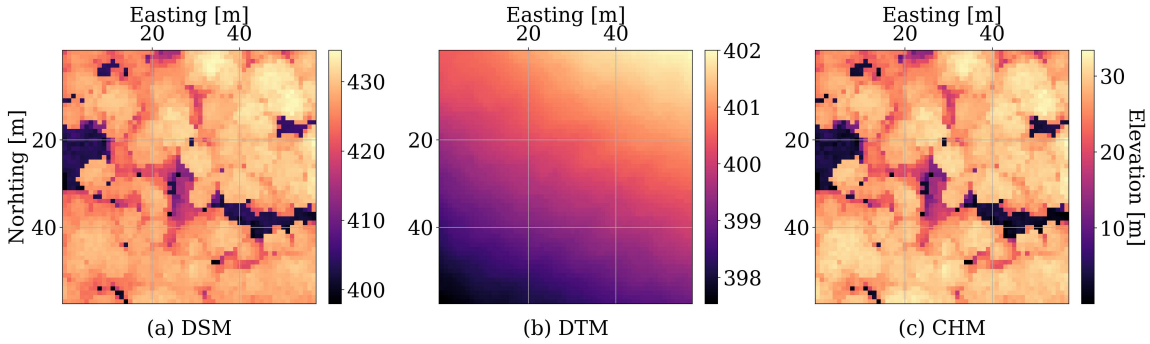


Figure 2.13: Elevation products derived from the point cloud in Fig. 2.12. These consist in the (a) DSM, (b) DTM and (c) CHM, rasterized at 1 m ground sampling distance.

area (i.e. $\mathrm{pts/m}^2$). An example of an ALS-derived point cloud can be seen in Fig. 2.12. The collected data can be converted into a wall-to-wall imagery through a *rasterization* process, consisting in the definition of a regular sampling grid of adequate resolution (typically coarser than that of the LiDAR systems), the sorting or assignment of the individual samples to their corresponding cells, and a final aggregation step to merge multiple samples within the same cell. For example, starting from the point cloud in

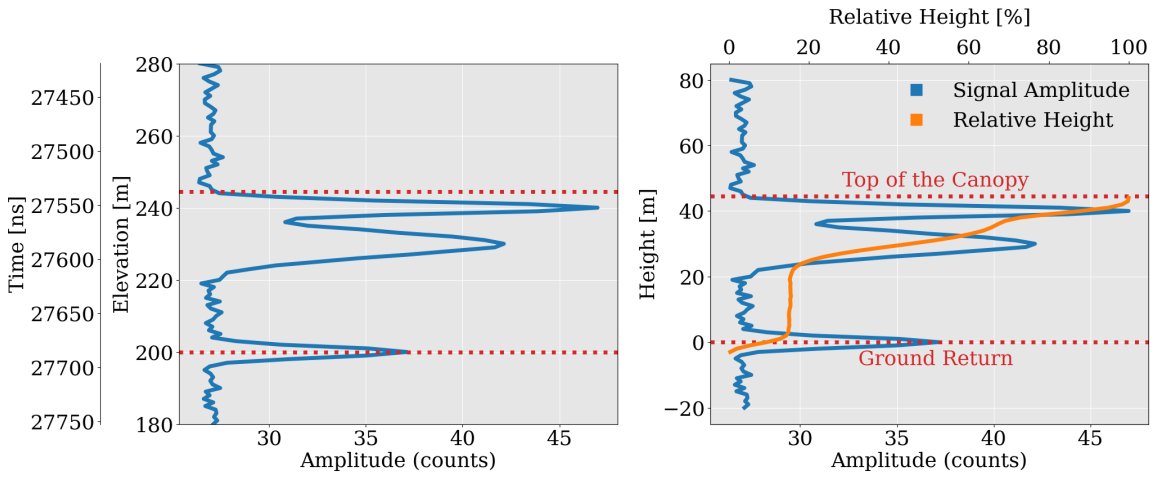


Figure 2.14: A simulated full-waveform LiDAR signal over a forested area. The image on the left shows the received signal amplitude (in blue) with respect to both the time of flight and the corresponding absolute elevation. The image on the right shows the detected positions (in red) for the top of the canopy and the ground, and the estimated relative height trend (in orange).

Fig. 2.12, it is possible to obtain the rasterized elevation products shown in Fig. 2.13.

2.3.3 Full Waveform LiDAR

In many modern systems, the recorded LiDAR signal consists in a so-called *full-waveform* return signal, which must be first pre-processed in order to determine the location of the signal returns [53][20]. Full-waveform signals also allow for the derivation of more complex metrics of the sampled target. For vegetated areas in particular, it is common to compute the *relative height* (RH) metrics, defined as the canopy elevations corresponding to certain percentages of cumulatively returned energy, with *RH100* representing the height at the top of the canopy and *RH0* representing the height at the lowest detected ground point. Fig. 2.14 shows an example of a full-waveform LiDAR signal (a) and the correspondingly extracted elevation statistics (b).

2.4 Principles of Optical Sensors

In this section, the working principles of passive optical systems for imaging are briefly discussed. Starting from the definition, these systems primarily distinguish themselves from active sensors, such as Radar or LiDAR, by the fact that the measured radiation is either emitted by the investigated object, or originated from another source (e.g. the

Sun) and back-reflected by the object itself.

2.4.1 Radiation Theory

Every natural occurring body, at a temperature T greater than absolute zero (i.e. $0\,\mathrm{K} = -273.15\,^{\circ}\mathrm{C}$), both emits EM radiation, which depends on T and the properties of the object, and reflects, absorbs and conducts impinging EM waves. In particular, the *ideal* behaviour of emission of a body in thermal equilibrium is defined by the so-called *black-body radiation*, which dictates that the body emits EM energy at a continuous spectrum of light and with an intensity which directly depends on the temperature. In particular, the emission behaviour of a black-body is defined by the *Planck's Law* as:

$$B(\lambda, T) = \frac{2\pi c^2 h \lambda^{-5}}{\exp(\frac{ch}{\lambda K T}) - 1}\,[\mathrm{Wm}^{-2}\mu\mathrm{m}^{-1}] \tag{2.47}$$

where $B(\lambda, T)$ is the spectral power density (i.e. power per unit volume per unit frequency), λ is the considered wavelength, c is the speed of light in a vacuum, h is the Planck constant and K is the Boltzmann constant.

Notably, the wavelength at the maximum emission power can be estimated by *Wien's Law* as:

$$\lambda_{\max} = \frac{2998}{T}\,[\mu\mathrm{m}] \tag{2.48}$$

and it shows that the peak of emission is inversely proportional to the temperature of the black-body. Therefore, the black-body model is that of an ideal object, which perfectly absorbs impinging radiation and perfectly emits it on the entire EM spectrum according to Planck's Law. As already mentioned, real objects do not present this ideal behaviour, but emit and also reflect, transmit and absorb the impinging radiation on the basis of the object's properties, and are thus referred to as *gray-bodies*.

It is then possible to define the following quantities:

$$\rho(\lambda) = \frac{E_r(\lambda)}{E_i(\lambda)} \qquad \tau(\lambda) = \frac{E_t(\lambda)}{E_i(\lambda)} \qquad \alpha(\lambda) = \frac{E_a(\lambda)}{E_i(\lambda)} \tag{2.49}$$

where $\rho(\lambda)$ is the coefficient of reflection, $\tau(\lambda)$ is the coefficient of transmission and $\alpha(\lambda)$ is the coefficient of absorption, with E_i, E_r, E_t and E_a representing the incident, reflected, transmitted and absorbed energy, respectively.

It follows that Planck's Law for a black-body represents an upper limit of emission, and that the relationship between the ideal black-body and the real gray-body defines

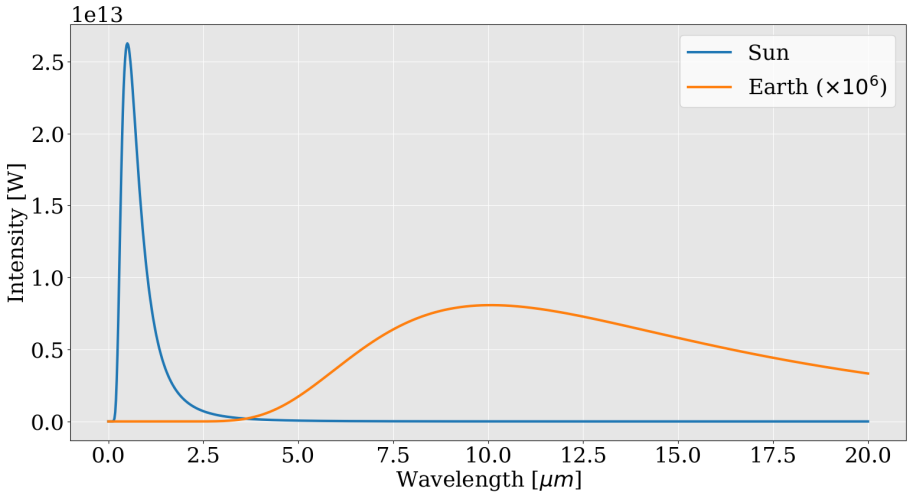


Figure 2.15: Black-body radiation model applied to the Earth and the Sun. Notably, the power radiated by the Earth is completely dwarfed by that emitted by the Sun. With an average surface temperature of 228 K, Earth's emission peak is at $\approx 10\,\mu m$ (i.e. thermal infrared), while with 5778 K the Sun has its emission peak at $\approx 0.5\,\mu m$ (i.e. visible light).

a so-called *coefficient of emissivity* as:

$$\epsilon(\lambda) = \frac{B_{real}(\lambda, T)}{B_{ideal}(\lambda, T)} \tag{2.50}$$

In particular, it was demonstrated by Kirchhoff that a good emitter is also a good absorber, i.e. $\alpha(\lambda) \approx \epsilon(\lambda)$. Finally, it follows from the *law of conservation of energy* that:

$$\rho(\lambda) + \tau(\lambda) + \epsilon(\lambda) = 1 \tag{2.51}$$

2.4.2 Characterization of Objects on the Earth

Two major sources of energy for the passive imaging task are represented by the Sun and the Earth. Modelling each of them as an ideal black-body yields the relationship shown in Fig. 2.15, with the Sun having its emission peak at 0.5 μm (i.e. visible light) and Earth at 9.66 μm (i.e. thermal infrared).

Earth's atmosphere is not completely transparent being characterized by a non-unitary transmission coefficient that depends on its molecular composition. Molecules such as carbon dioxide and water give rise to regions in the EM spectrum in which the atmospheric transmittance is low. Further distortions are caused by the so-called Rayleigh and Mie scattering phenomena, which affect EM waves based on the relative

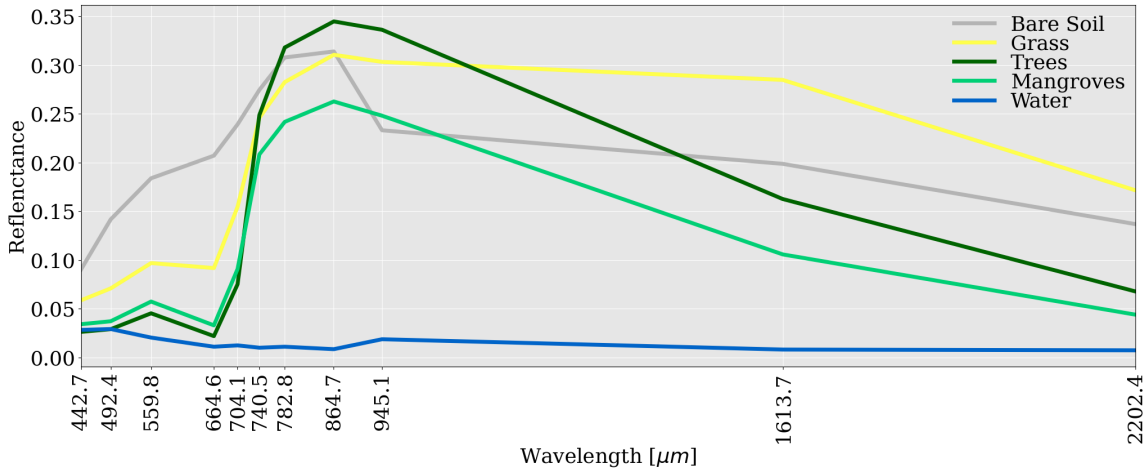


Figure 2.16: Example of spectral signatures for different classes of objects acquired by the ESA Sentinel-2 multi-spectral sensor over the country of Gabon, Africa. Notably, it is possible to observe the characteristic *red-edge* behaviour for the vegetation classes, characterized by a strong absorption at red wavelengths (664.6 μm) and a strong reflection at near-infrared wavelengths (864.7 μm).

size of the encountered particles, leading to further scattering within the atmosphere. These elements need to be considered in order to compensate for the contribution of undesired effects and observe only the properties of the desired target. The collection of corrections that need to be performed to compensate for atmospheric-induced effects is often referred to as *atmospheric correction*. They result in the distinction between the observations before the correction, known as *top of atmosphere* (TOA), and after the atmospheric correction, known as *bottom of atmosphere* (BOA).

Finally, the solar radiation interacts with Earth's surface, resulting in a characteristic reflected energy profile, which depends on the considered wavelength and, crucially, on the properties of the scene. This is known as the *spectral signature* of the target. Ultimately, it is this characteristic curve that is observed by the optical sensor, sampled according to discrete spectral intervals (called *bands* or *channels*), which are defined by the sensor's hardware. Fig. 2.16 shows an example of spectral signatures for different classes of objects, as recorded by the multi-spectral sensor on-board the ESA Sentinel-2 multi-spectral satellite.

2.4.3 Acquisition Principle and Resolution

Optical systems primarily operate by scanning a scene over time to generate an image. The simplest way to achieve this is through opto-mechanical scanning systems, which

repeatably acquire one (i.e. *line Scanner*) or multiple (i.e. *whiskbroom Scanner*) parallel across-track segments. For example, this has been the operational acquisition mode of the Landsat 7 spaceborne mission [54]. More recently, technological developments have allowed for the introduction of static systems made up of large arrays of sensors, which simultaneously acquire in the along-track direction. Such systems are known as *pushbroom Scanners*. An example are the *Multi-Spectral Instrument*s (MSI) sensors mounted on board of the ESA Sentienl-2 satellites [24].

The performance of an optical sensor is defined by its geometric, radiometric and spectral resolutions.

Geometric Resolution

The *geometric resolution* is typically defined in terms of *instantaneous field of view* (IFOV), defined as:

$$\text{IFOV} = 2\arctan\left(\frac{w}{2f}\right) \approx \frac{2}{f} \tag{2.52}$$

where w is the width of the sensor and f is the focal distance of the optical system. The IFOV is subsequently projected onto the ground to obtain the *ground projected IFOV* (GIFOV) as:

$$\text{GIFOV} = 2H\tan\left(\frac{\text{IFOV}}{2}\right) \approx \frac{wH}{F} \tag{2.53}$$

where H is the height of the acquisition platform (i.e. the distance between the sensor and the imaged scene).

Notably, the acquisition process includes an analogue-to-digital (ADC) conversion process, which discretizes both in time and in amplitude the acquired imaging lines, resulting in the creation of individual *pixels*. In particular, the on-ground distance between two consecutive pixels is defined as the *ground sampling distance*, and it is commonly chosen to match the GIFOV of the sensor. The amplitude of each pixel is linearly quantized to produce a so-called *digital number* (DN), which gets typically converted back into a physical quantity, such as the *received energy* or the *reflectance coefficient*, through the knowledge of the specif system parameters.

Spectral Resolution

Crucially, the recorded signal is also the result of the integration over a third dimension, defined by the finite EM bandwidth or spectrum of the sensor. Indeed, it is common for optical imaging sensors to simultaneously acquire multiple images in different EM

spectra. Such sensors are commonly referred to as a *multi-spectral* scanners. Moreover, systems which are characterized by an extremely elevated amount of spectral channels, typically in the order of a few hundreds, are known as *hyperspectral* scanners. It is therefore possible to formulate the concept of *spectral resolution* $\delta\lambda$, defined as the spectral bandwidth that a sensor is able to sample.

Radiometric Resolution

Ultimately, it is possible to see how the amplitude of a single pixel is the result of the signal integration over a discrete window defined by the spatial resolution (δx and δy, i.e. the GIFOV), and the spectral resolution ($\delta\lambda$). It follows that finer spatial and spectral resolutions equate to a smaller energy integration window and are thus coupled together by the concept of *radiometric resolution*, which determines the minimum amount of energy that needs to reach the transducer in order to produce a measurable electrical signal. For these reasons a larger number of narrow-band channels is easier to achieve on airborne rather than spaceborne platforms, as the lower elevation allows the sensor to intercept a larger amount of backradiated energy, counteracting the reduced integration window for each band. Similarly, systems with small revisit times orbit the Earth at higher altitudes, and thus are required to compromise in spatial resolution in order to be able to record sufficient signal power.

Chapter 3

Deep Learning for Regression Tasks

The family of well-known Deep Learning (DL) machine learning algorithms sets its roots in estimation theory. In particular, the optimization mechanism which drives the training process of a neural network is built upon a solid mathematical formalism based on established concepts in statistics and signal processing.

This chapter presents the basic notions of supervised deep learning with respect to regression tasks, aiming at providing the reader with the necessary fundamental basics to fully comprehend the work developed in this PhD dissertation. Whenever possible, a direct link to classical estimation theory fundamentals is provided, but for a more comprehensive review on the topic, the reader can refer to the works published in [55][56].

3.1 Regression

Theoretically, any data can be modeled through a certain set of parameters. The estimation problem reduces therefore to finding (or predicting) the best values of such parameters that can describe the data in the most accurate way possible. In particular, regression is a statistical approach used for the prediction of a real-valued scalar output $y \in \mathbb{R}$ (the so-called dependent variable), from an m-dimensional column-vector of real-valued input observations $\boldsymbol{x} \in \mathbb{R}^m$ (the independent variables).

3.1.1 Linear Regression

In the specific case that the expected value of the output can be modelled as a linear function of the input, the technique takes the name of *linear regression* and is described

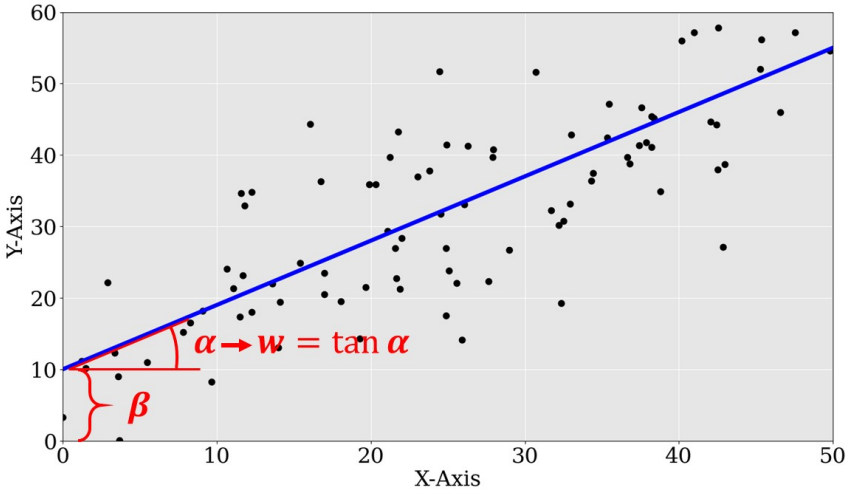


Figure 3.1: Example of a one-dimensional linear regression problem. The data points (in black) follow the linear trend captured by $y = \beta + w\,x$ (blue line). The offset parameter β of the line represents the intercept, while the weighting parameter w represents the slope.

by the following relationship:

$$y_i := f(\boldsymbol{x}_i, \boldsymbol{\theta}) = \beta + w_1 x_{i,1} + w_2 x_{i,2} + \cdots + w_m x_{i,m} \tag{3.1}$$

where $\boldsymbol{\theta} = (\beta, w_1, \ldots, w_m)$ are the $m + 1$ parameters of the liner model, with β representing the *intercept* or *bias* of the function, and the remaining parameters representing the expected change in output given an unit change of the corresponding input variable. The relationship can be equivalently expressed in vector form:

$$y_i = \beta + \boldsymbol{x}_i^T \boldsymbol{w} \tag{3.2}$$

Let us now consider a dataset of n observations $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$, then the previous relationship can be written in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} \tag{3.3}$$

where, in order to accommodate $\boldsymbol{\theta}$ into the matrix notation $\boldsymbol{X}$, the first element of every row is commonly set as a unitary regressor (i.e. $x_{i,1} = 1$, for $i = 1, \ldots, n$) and the remaining values of each row are given by the corresponding row vectors $\boldsymbol{x}_i^T$. An example of a one-dimensional linear regression problem is shown in Fig. 3.1.

In practice, it is often the case that $n > m$, or, in other words, that the system is *overdetermined* and therefore usually has no exact solution. Instead of finding an exact

solution, it then becomes necessary to find the parameters $\boldsymbol{\theta}$ that *best fit* the data, i.e. that minimize an *objective function* $\mathcal{L}$:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{3.4}$$

where $\hat{\boldsymbol{\theta}}$ represents the vector of estimated parameters. The process of *fitting* the parameters $\boldsymbol{\theta}$ to the data $\mathcal{D}$ is at the core of *machine learning* and is often referred to as *training* of a model. The overarching principle is that of *empirical risk minimization* (ERM), consisting in the optimization of the model on the available data (the empirical risk) while not knowing the true distribution of the data itself and its related risk.

Ordinary Least Squares

The Ordinary Least Squares (OLS) is a commonly used approach to estimate the parameters of a linear regression model, and consists in finding the parameters $\boldsymbol{\theta}$ that minimize the *residual sum of squares* (RSS):

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(y_i - \left[\sum_{j=1}^{m} \{x_{i,j}\, w_j\} + \beta \right] \right)^2 = ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}||_2^2 \tag{3.5}$$

Fig. 3.2 shows a visual interpretation of the residuals of for a 1D linear regression task. Equating the partial derivative with respect to $\boldsymbol{\theta}$ to zero yields the solution:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{3.6}$$

In practice, it is more common to equivalently express the minimization problem in terms of mean squared error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.7}$$

where $\hat{y}_i = f(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}})$ is the predicted response of the model. The MSE has the practical advantage of being independent of the dataset size and is thus better comparable, making it often preferable.

Maximum Likelihood Estimator

In practice, perfectly predicting the output variable is unachievable due to two sources of uncertainty. The first one, the so-called *aleatoric* or *stochastic* uncertainty, captures

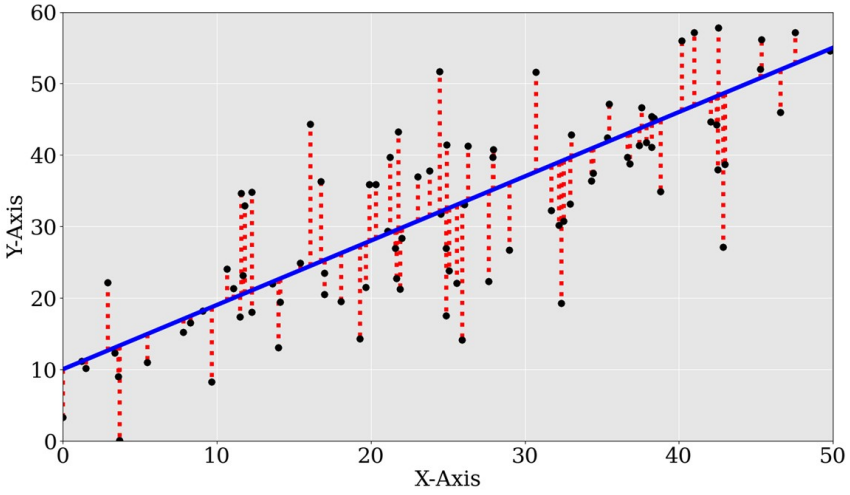


Figure 3.2: Example of residuals for a 1D linear regression problem. For each of the data points (in black) it is possible to compute the corresponding residual (red dotted lines) w.r.t. the linear regression line (in blue). The objective of the least squares approach is to find the linear model parameters that minimize the sum of the squared residuals.

the effects of intrinsic randomness in the mapping process (i.e. in the data generation process itself), which is irreducible and thus can not be modelled. The second, the so-called *epistemic* or *model* uncertainty, describes the effects caused by lack of knowledge or simplifications in the model, which could in theory be overcome by using more data, better measurements, or a more precise modelling, but which in practice remains unaccounted for. It is often convenient to capture the prediction uncertainty explicitly by considering the conditional probability distribution of the output variable.

In statistical terms, a reasonable strategy for estimating $\boldsymbol{\theta}$ coincides with choosing the estimate $\hat{\boldsymbol{\theta}}$ that assigns the highest probability to the training data. Thus we can define the *maximum likelihood estimator* (MLE) as:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta}) \tag{3.8}$$

Under the assumptions of independent and Normally distributed samples with constant variance, it is possible to show (see Appendix B.1) that the MLE solution coincides with the OLS solution:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 \right\} \tag{3.9}$$

3.1.2 Nonlinear Regression

In the case of an underlying nonlinear relationship, finding a good fit to the data cannot always be achieved by numerical means. As an example, let us consider the following exponential relationship:

$$y_i = w_2 \exp(w_1 x_{i,1}) \tag{3.10}$$

Applying the least squares approach leads to a system of nonlinear equations for which no exact closed-form solution exists. In this particular case, it is still possible to find a solution by *linearizing* the problem to the logarithm of the output data, whereas in general such a solution might not always exist or be trivial to find.

Feature Extraction

A potential solution to a nonlinear regression problem consists in performing a polynomial transformation Φ of the independent variable $\boldsymbol{x}_i$, and replacing it with $\Phi(\boldsymbol{x}_i)$ in the regression model:

$$f(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Phi(\boldsymbol{x}_i)^T \boldsymbol{\theta} \tag{3.11}$$

Such a transformation constitutes an operation of *feature extraction* from the raw input and it results in a model which remains linear with respect to $\Phi(\boldsymbol{x}_i)$. An important consequence then arises by also parameterizing Φ with $\boldsymbol{\theta}_\Phi$, as to obtain:

$$f(\boldsymbol{x}_i, \boldsymbol{\Theta}) = \Phi(\boldsymbol{x}_i, \boldsymbol{\theta}_\Phi)^T \boldsymbol{\theta} \tag{3.12}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\theta}_\Phi)$ is the set of all model parameters.

In practice, a single feature transformation operation might not suffice to properly linearize the problem. A potential solution then might reside in chaining a finite sequence L of these parameterized operations, such as to obtain:

$$f(\boldsymbol{x}_i, \boldsymbol{\Theta}) = \Phi_1(\Phi_2(\cdots(\Phi_L(\boldsymbol{x}_i, \boldsymbol{\theta}_L))\cdots, \boldsymbol{\theta}_2), \boldsymbol{\theta}_1) \tag{3.13}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_L)$ is the collection of all model parameters. This is the fundamental idea behind *feedforward neural networks*, which can be seen as extremely articulated feature extractors, whose representative power and modelling capacity arises from the recursive chaining of learnable linear and nonlinear functions. The resulting model remains fundamentally nonlinear, and thus ultimately requires heuristics to find a solution, commonly consisting of iterative optimization approaches.

3.1.3 Optimization Problem

As previously mentioned, finding the model parameters $\boldsymbol{\theta}$ that best fit the data can be expressed through the optimization problem in Eq. (3.4), in which the best solution is the one which minimizes (or conversely maximizes) an objective function. The objective function quantifies the quality of the fit or the distance between predicted and expected values, and takes the specific name of *loss function* or *cost function* if it has to be minimized, or of *score function* or *reward function* if it has to be maximized.

The solution that strictly satisfies Eq. (3.4) corresponds to the *global optimum*, but finding it is often computationally intractable and thus limits one to find a satisfactory *local optimum* $\boldsymbol{\theta}^*$, such that:

$$\forall \boldsymbol{\theta} \in B, \exists \delta > 0 \in \mathbb{R} : |\boldsymbol{\theta} - \boldsymbol{\theta}^*| < \delta, \mathcal{L}(\boldsymbol{\theta}^*) < \mathcal{L}(\boldsymbol{\theta}), \text{ with } \boldsymbol{\theta}^* \in B \tag{3.14}$$

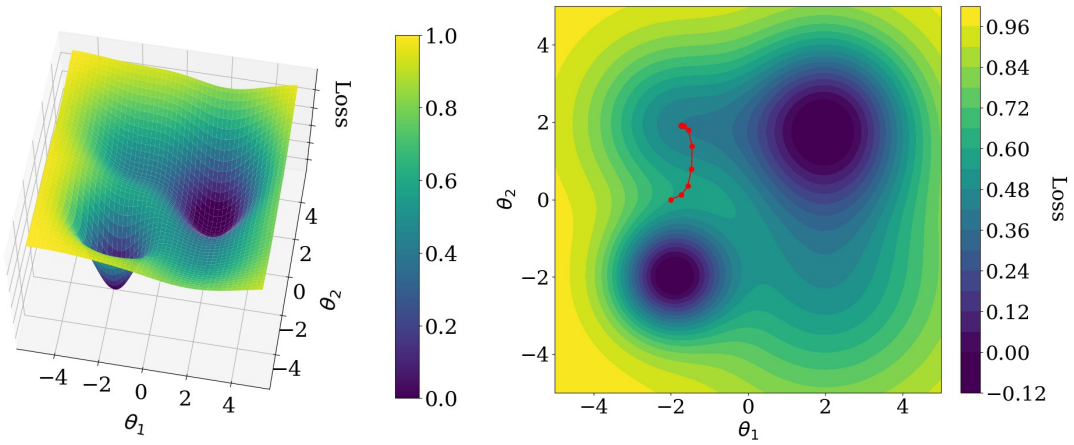
where $B \subseteq \mathbb{R}^d$ is the solution space created by the d trainable parameters of the model. An example of a solution landscape defined by the parameters of the model and the chosen cost function can be seen in Fig. 3.3 (a).

Gradient Descent

Assuming that the objective function $\mathcal{L}(\boldsymbol{\theta})$ is differentiable, *gradient descent* (GD) or *first order* methods use this information to compute the direction of downward slope of the solution landscape. Given an initial set of parameters $\boldsymbol{\theta}_0$, which corresponds to a certain position in the solution landscape, these methods adopt an iterative approach to incrementally update the solution at a generic k^{th}-iteration according to the following general formula:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + r_k \boldsymbol{d}_k \tag{3.15}$$

where r_k is the *learning rate* or *step size* of the update, and $\boldsymbol{d}_k$ is the update direction. As already mentioned, in first order methods the direction of descent is related to the local gradient of the objective function at the current iteration $\nabla\mathcal{L}(\boldsymbol{\theta}_k)$. In the particular case that $\boldsymbol{d}_k = -\nabla\mathcal{L}(\boldsymbol{\theta}_k)$ we talk about the *steepest descent* direction, and following it might not always correspond to the best strategy for rapid convergence to a good solution. As an example, applying the GD method to the solution landscape shown in Fig. 3.3 (a) leads to the solution seen in Fig. 3.3 (b), which does not correspond to the global minimum of the objective function.



(a) (b)

Figure 3.3: (a) example of a loss landscape characterized by the presence of multiple local minima. (b) Walk direction found by the GD method for a random starting location corresponding to the set of parameters $\boldsymbol{\theta}_0$. It is possible to note that the unfavourable starting position of $\boldsymbol{\theta}_0$ has forced the solver into a bad local minimum.

Stochastic Gradient Descent

In the context of an empirical risk minimization-based fitting procedure, computing the empirical mean derivative of the objective function across the entire dataset (e.g. gradient descent) rapidly becomes unfeasible for problems of very large dimensionality ($d >> 1$). A practical alternative consists in building a subset, called *minibatch*, $\mathcal{D}_k \subset \mathcal{D}$ of $n_k << n$ randomly selected dataset samples for each k^{th} iteration step. This is then used to compute an unbiased approximation of the empirical mean derivative for the current iteration:

$$\nabla \mathcal{L}(\boldsymbol{\theta}_k) = \frac{1}{n} \sum_{i \in \mathcal{D}} \nabla_{\boldsymbol{\theta}_k} \mathcal{L}_i(\boldsymbol{\theta}_k) \approx \frac{1}{n_k} \sum_{i \in \mathcal{D}_k} \nabla_{\boldsymbol{\theta}_k} \mathcal{L}_i(\boldsymbol{\theta}_k) \tag{3.16}$$

This method is called *stochastic gradient descent* (SGD). Compared to the GD method, the reduced sample count introduces noise in the approximation, which causes a larger number of iterations to be required to reach convergence. In practice, the significantly lower time-per-iteration of the SGD approach substantially improves the rate of convergence.

A concrete example of an SGD-based solver is the *adaptive moment estimation* (ADAM). Each update step is computed as:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - r_k \frac{1}{\sqrt{\boldsymbol{s}_k} + \epsilon} \boldsymbol{m}_k \tag{3.17}$$

with:

$$\boldsymbol{m}_k = \alpha_1 \boldsymbol{m}_{k-1} + (1 - \alpha_1)\nabla\mathcal{L}(\boldsymbol{\theta}_k) \tag{3.18}$$

$$\boldsymbol{s}_k = \alpha_2 \boldsymbol{s}_{k-1} + (1 - \alpha_2)\nabla\mathcal{L}^2(\boldsymbol{\theta}_k) \tag{3.19}$$

where the default values are $\alpha_1 = 0.9$, $\alpha_2 = 0.999$, $\epsilon = 10^{-8}$, and the initial learning rate value $r_k = r_0$ is fixed. Eq. (3.18) and Eq. (3.19) implement the concept of *momentum*, i.e. of an exponentially weighted moving average of their past states, where the extreme case of $\alpha_{1,2} = 0$ corresponds to a lack of memory or momentum.

3.2 Feedforward Neural Networks

3.2.1 Perceptron

Arguably the first realisation of a biologically-inspired artificial neuron was the *Perceptron*, built by Frank Rosenblatt in 1957 in the form of the *Mark I Perceptron* machine [57]. It was based on an earlier work by Warren McCulloch and Walter Pitts [58], who in 1943 suggested that the human brain could be viewed as a computing device, consisting of *neural nets*.

In their analyses they first introduced the notion of what is now known as the *McCulloch–Pitts neuron* or the Perceptron. In a biological neuron, the signals coming from multiple connections, known as *dendrites*, are aggregated and compared against a certain threshold value. When the total signal exceeds this threshold, the neuron *fires* a signal of its own, known as the *action potential*. At its core, the Perceptron is a deterministic binary classifier which takes a weighted sum of multiple inputs and delivers a binary output described by the following relationship:

$$a(\boldsymbol{x}_i, \boldsymbol{\theta}) = \mathrm{H}(f(\boldsymbol{x}_i, \boldsymbol{\theta})) = \mathrm{H}(\boldsymbol{x}_i^T \boldsymbol{\theta}) \tag{3.20}$$

with a being the *activation*, $\alpha = f(\boldsymbol{x}_i, \boldsymbol{\theta})$ the *pre-activation* and $\mathrm{H}(\alpha)$ being the Heaviside step function defined as:

$$\mathrm{H}(\alpha) := \begin{cases} 1, & \text{if } \alpha \geq 0 \\ 0, & \text{elsewhere} \end{cases} \tag{3.21}$$

Fig. 3.4 shows a schematic representation of the Rosenblatt Perceptron. The innovation of this work consisted in the introduction of a learning algorithm, which allows one to iteratively update the weighting coefficients of the Perceptron on the basis of labelled

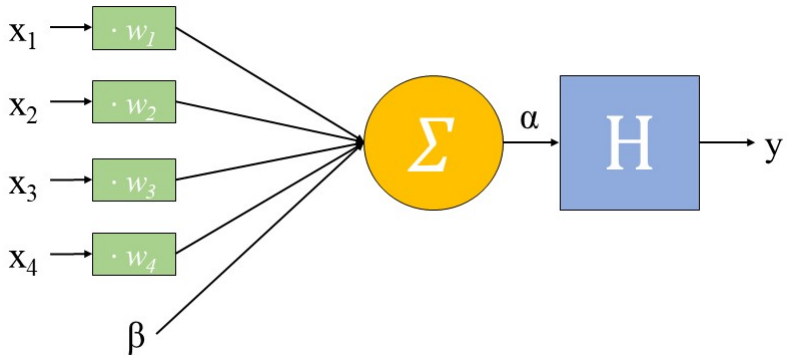


Figure 3.4: The schematic representation of the Rosenblatt Perceptron. All inputs x_j are individually multiplied by their corresponding weighting coefficient w_j, and subsequently added together with the bias term β. The resulting pre-activation signal α is then passed to the Heaviside activation function $H(\alpha)$ to generate the final output y.

training data. The Heaviside step function is non-differentiable, which implies that the update rule could not take advantage of first or higher order derivatives. Instead, the Perceptron update rule was defined as:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - r_k(\hat{y}_i - y_i)\boldsymbol{x}_i \tag{3.22}$$

with the notable consequence that no update is made whenever the prediction is correct.

The XOR Problem

While computationally light, the Perceptron algorithm is not capable of converging for non-linearly separable data. Arguably the most famous example is that of the *XOR problem*, referring to the Perceptron's impossibility to implement the *exclusive OR* (XOR) function described by the following expression:

$$\text{XOR} := (x_1 \text{ OR } x_2) \text{ AND NOT } (x_1 \text{ AND } x_2) \tag{3.23}$$

and shown in Tab. 3.1. The intuitive solution to this nonlinear problem was found with the introduction of the *multilayer perceptron* (MLP), which consists in cascading multiple Perceptrons on top of each other. Research on the matter had begun in the mid '60s [59, 60], but did not see a re-surge until 1986, with the definitive adoption of the *backpropagation* algorithm [61] combined with differentiable nonlinear activation function, leading to an efficient way to train deep multi-layer models.

A last important element for the successful spread of interest for neural networks

XOR		
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Table 3.1: The truth table for the XOR function, defined by the inputs x_1 and x_2, and the output y.

started in the late 2000's, with the increasing availability of cheap and powerful *graphics processing units* (GPU), hardware computing devices optimized for the parallelized matrix-vector computation required by neural networks.

The definitive watershed moment thus arrived in 2012, when a deep neural network managed to drastically improve the classification performance in the *ImageNet* benchmark [62], significantly outpacing the previous year-on-year improvements obtained by more conventional strategies.

3.2.2 Multi-Layer Perceptron

A *multi-layer perceptron* (MLP) is a specific type of *artificial neural network* (ANN) that consists of multiple layers of interconnected nodes, or neurons. It is an example of *feedforward neural network*, meaning that information flows in one direction only: starting from the input nodes, through the hidden nodes and up to the output nodes, without any feedback loop.

Let us now consider a generic MLP model with L-layers. The generic neuron $p = 1, \dots, m_l$ in layer $l = 1, \dots, L$ is defined by its set of parameters $\boldsymbol{\theta}_{l,p} = (\beta_{l,p}, \boldsymbol{w}_{l,p})$. In particular, $\boldsymbol{w}_{l,p} = (w_{l,p,1}, \dots w_{l,p,m_{l-1}})$ are the m_{l-1} weighting coefficients of the linear combination of the set of inputs values (i.e. the set of outputs from the previous layer $l-1$), whereas $\beta_{l,p}$ is an additional bias coefficient. The result is followed by a differentiable *activation function* $\Phi(\cdot)$ such that, for each data sample $i = 1, \dots, n$, it is possible to define the scalar output of the neuron as:

$$a_{l,p,i} = \Phi(\alpha_{l,p,i}) = \Phi\left(\sum_{j=1}^{m_{l-1}} \{w_{l,p,j}\, a_{l-1,j,i}\} + \beta_{l,p}\right) \tag{3.24}$$

where $\boldsymbol{a}_{l-1,i} = (a_{l-1,1,i}, \dots, a_{l-1,m_{l-1},i})$ are the outputs of all neurons in the previous layer, with the special cases of $\boldsymbol{a}_{1,i} = \boldsymbol{x}_i$ being the set of inputs and $\boldsymbol{a}_{L,i} = \boldsymbol{y}_i$ being the

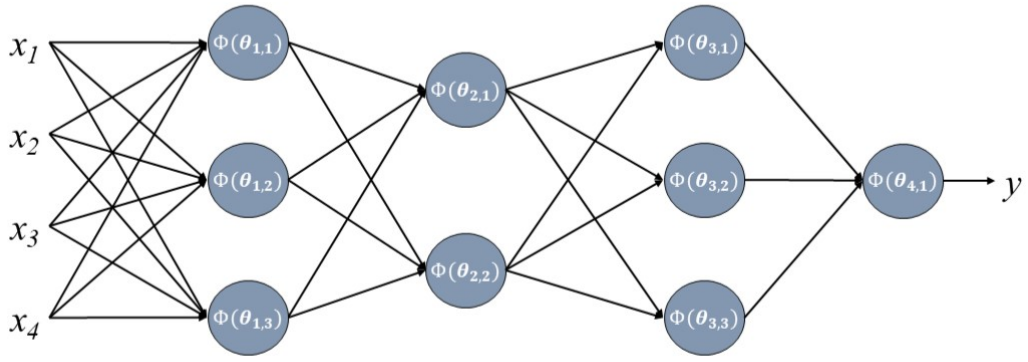


Figure 3.5: The schematic representation of a single-output MLP. Each unit p in layer l has its own set of weights $\boldsymbol{\theta}_{l,p}$. Note that each node is connected to all nodes in the preceding and the succeeding layers, meaning that by design all inputs at a given layer are equally considered. This layer topology is commonly referred to as a *dense* or a *fully-connected* layer.

set of outputs of the overall MLP model. A simplified representation of a single-output MLP network can be seen in Fig. 3.5.

Let us then assume the following example of a single output, two layer model:

$$y = \beta_0 + w_1\Phi(\beta_1 + w_1 x) + w_2\Phi(\beta_2 + w_2 x) \tag{3.25}$$

where, for simplicity, Φ is a piecewise linear activation function, such as the *rectified linear unit* (ReLU) function shown in Fig. 3.6 and described in Sec. 3.2.3. It follows then that each individual neuron is capable of describing a half-plane, meaning that the combination of these neurons can itself describe up to three linear regions. The generalization of this result is that, given enough units, a two layer MLP can approximate any continuous function with an arbitrary level of precision. This relevant consequence is formalized by the *universal approximation theorem*.

In practice, the number of required units (i.e. the *width* of the model) might become impractical. It is thus often more efficient to increase the number of layers instead (i.e. the *depth* of the model), as the patterns learned at the previous layer can be re-used to create more complex patterns in the following ones. Therefore, the resulting *deep networks* typically perform better than *shallow networks* with a comparable number of trainable parameters (or, conversely, are more parameter efficient for the same level of performance) [56]. Overall, the total number of hidden units and their combination (i.e. depth versus width) describes the *capacity* of a neural network, or intuitively its

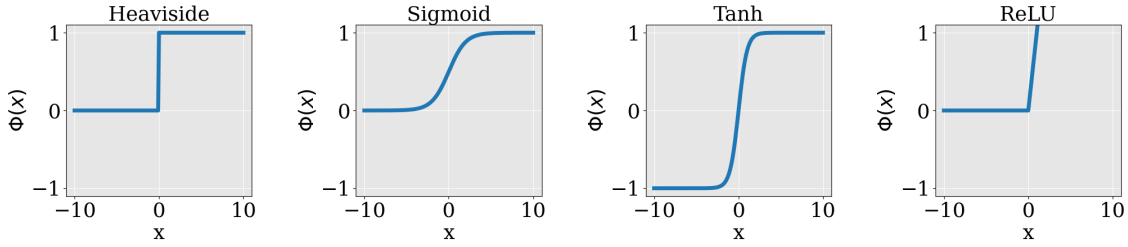


Figure 3.6: Examples of different activation functions. From left to right, the *Heaviside* step-function, the *sigmoid*, the *tanh* and the *ReLU*.

potential ability of modelling complex relationships.

3.2.3 Activation Functions

The role of the activation function is pivotal for the model's capacity to represent nonlinear relationships and thus must itself be nonlinear. As discussed in Sec.3.2.1, in its first years the work on neural networks was limited to the use of the non-differentiable Heaviside step function described in Eq. (3.21), which severely limited the possibility to train more complex architectures. With the definitive adoption of the backpropagation algorithm (discussed in more detail in Sec. 3.2.4), activation functions, such as the *sigmoid* and the *tanh*, became popular options. These are described by the following equations:

$$\sigma(\alpha) = \frac{1}{1 + e^{-\alpha}} \tag{3.26}$$

$$\tanh(\alpha) = \frac{e^{2\alpha} - 1}{e^{2\alpha} + 1} \tag{3.27}$$

Notably, both functions saturate for both large positive and negative inputs, which causes their derivatives to become extremely small for these input regions and consequently impedes the backpropagation of the loss gradient to preceding layers. This phenomenon is often referred to as the *vanishing gradient problem*, and it must be avoided as it implies that the model effectively stops learning. For completeness, the opposite problem also exist; it is know as the *exploding gradient problem* and it occurs when the gradient blows up during the backpropagation steps, making the training process increasingly unstable up to the point of variable overflow.

Nowadays the aforementioned activation functions have most commonly been re-

placed by the *rectified linear unit* (ReLU) [63][64], defined as:

$$\mathrm{ReLu}(\alpha) := \max(\alpha, 0) = \begin{cases} \alpha, & \text{if } \alpha > 0 \\ 0, & \text{elsewhere} \end{cases} \tag{3.28}$$

which blocks the propagation of negative signals, while leaving positive ones unaltered.

In agreement with Eq. (3.11), the output layer for a regression task should implement a linear relationship and is thus not followed by any activation function.

3.2.4 The Backpropagation Algorithm

A powerful advantage of neural networks is that it is possible to train them to fit some data from the available data itself. In order to update the model parameters, it is then reasonable to expect that each individual trainable weight is updated according to its impact on the overall quality of the fitting, as described by a cost function. As discussed in Sec. 3.1.3, the commonly used solvers belonging to the stochastic gradient descent family implement this concept, by updating each parameter by a quantity that is proportional to the partial derivative of the cost function with respect to itself. Then, the motivation behind the backpropagation algorithm becomes to offer a computationally efficient approach to the calculation of the loss function gradient with respect to the model weights, allowing for the training of deep multilayered architectures.

In general, the algorithms can be used for every model that fulfills the following conditions:

- The model can be represented as a *directed acyclic graph* (DAG).
- Each node of the graph corresponds to a fully differentiable function.

In particular, a DAG is a type of directed graph for which no directed cycles exist, an intuitive requirements for the computation of the loss gradient in a finite amount of steps. Under these conditions it becomes possible to deterministically compute each partial derivative using the *chain rule* of calculus, given by:

$$\frac{df}{dx} = \frac{df}{du}\frac{du}{dx} \tag{3.29}$$

The backpropagation then consists of two steps: a forward pass, in which the intermediate products are computed and stored, and a backward pass, in which the derivatives for all parameters are computed. In particular, the intermediate products are used to

derive the partial derivatives with respect to the individual parameters. This can be generalized for the generic l^{th}-layer by the following relationships:

$$\frac{\partial \mathcal{L}_i}{\partial \beta_l} = \left[\frac{\partial \mathcal{L}_i}{\partial f_l}\right] \frac{\partial f_l}{\partial \beta_l} = \left[\frac{\partial \mathcal{L}_i}{\partial f_l}\right] \frac{\partial}{\partial \beta_l}(\beta_l + \boldsymbol{w}_l \boldsymbol{a}_l) = \left[\frac{\partial \mathcal{L}_i}{\partial f_l}\right] \tag{3.30}$$

$$\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{w}_l} = \left[\frac{\partial \mathcal{L}_i}{\partial f_l}\right] \frac{\partial f_k}{\partial \boldsymbol{w}_l} = \left[\frac{\partial \mathcal{L}_i}{\partial f_l}\right] \frac{\partial}{\partial \boldsymbol{w}_l}(\beta_l + \boldsymbol{w}_l \boldsymbol{a}_l) = \left[\frac{\partial \mathcal{L}_i}{\partial f_l}\right] \boldsymbol{a}_l^T \tag{3.31}$$

where $\frac{\partial \mathcal{L}_i}{\partial f_l}$ has already been computed according to the backpropagation flow. The calculation of the loss gradient is extremely computationally efficient, as all computations are done iteratively and by reusing the previously computed results. On the other hand, this puts larger requirements on the memory, as these intermediate results need to be stored for later use, which in practice can limit the overall model size.

3.2.5 Model Training, Validation and Testing

In this subsection, an introduction to the concepts of training, validation and testing phases is presented.

Training

Given the previously explained elements, it is now possible to build a complete training procedure for a deep learning architecture. Given a training set consisting of n_{train} samples, the minibatch approach introduced in Sec. 3.1.3 is used to break up the training procedure into

$$n_{\text{steps}} = \left\lfloor \frac{n_{\text{train}}}{BS} \right\rfloor \tag{3.32}$$

training iterations, with BS being the number of samples that constitute a single minibatch (also called batch size). Thus an *epoch* is defined as a full run across all possible n_{steps} batches. At the end of each epoch, the order of the training samples is randomly shuffled in order to produce a new sequence of minibatches for the next epoch. Following the procedures detailed in Sec. 3.2.4 and Sec. 3.1.3, each training iteration consists of a forward step, the computation of the loss function, the backpropagation step and the final weight update step. In the forward step an output is predicted for the given input, taking care to store all intermediate products. The generated output is then compared with the expected result by computing their pre-determined cost function. In the backpropagation step, the scalar cost associated to the current minibatch is iteratively differentiated with respect to each of the model's trainable parameters, where

the previously stored intermediate products are now used to make the computation more efficient. Finally, a pre-determined optimization algorithm, chosen from the family of SGD-based solvers, is used to efficiently update each of the weights based on their current output derivative and on their evolution history. This procedure is repeated for a pre-determined number of times or until a certain convergence criteria is met.

Validation

Up to this point the determination of the specific model characteristics has not been addressed. The collection of all model-defining parameters, such as the optimal dimensions of the model (e.g., the amount of layers and their respective widths), the type of used layers, the optimization solver and the cost function, fall under the umbrella term *hyperparameters* and the corresponding selection process is known as *hyperparameter tuning*. An evaluation of different hyperparameter values is thus required in order to determine the best performing model setup. A dedicated set-aside dataset is solely used for this purpose and it is known as the *validation set*. Crucially, the hyperparameter tuning process must remain completely independent to avoid biasing the final performance evaluation of the model, which is detailed in the next paragraph.

Furthermore, the validation set is also used to monitor the state of the model training, as the model performance tends to improve up to a certain number of training iterations, after which further training yield worse generalization results. This behaviour can be monitored by tracking the evolution of the loss function for both the training set and the independent validation set. The expected behavior is represented in Fig. 3.7: as the *training loss* typically monotonically decreases, the *validation loss* tends to also improve up to certain amount of iterations, after which it stagnates or gets worse. Training beyond this minimum only increases the model's bias towards the training set, a phenomenon known as *overfitting*. This aspect is strictly related to the relative sizes of the training set and the model capacity, with larger models in particular being complex enough to completely learn the training data. Conversely, arresting the training before the validation loss minimum has been reached implies that the input and output relationship has not been fully captured yet, a condition know as *underfitting*.

Testing

Training a model does not imply that it *generalizes* well across unseen data. It is therefore common to assess the performance of a trained neural network on a dedicated set-aside *test set*, such as to avoid biasing the performance metrics with previously seen

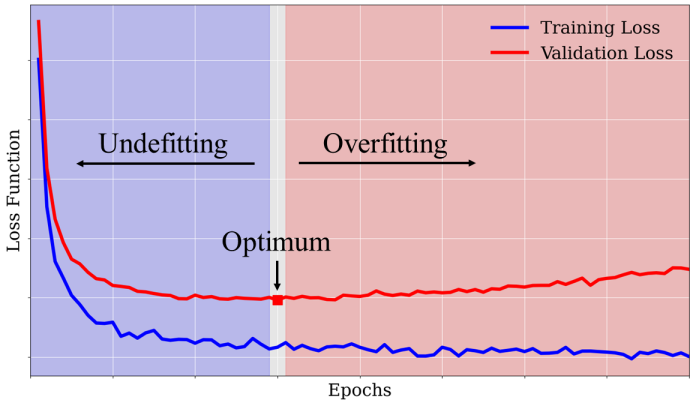


Figure 3.7: A typical example of training and validation loss co-evolution. Both losses decrease together up to a certain point, after which the validation loss starts worsening while the training loss keeps improving. The point of lowest validation loss corresponds to the model with best generalization properties. In the epochs before, the model is said to underfit the data, while after it the model is said to overfit the data.

examples.

Different metrics can be used to evaluate the performance of a model, and in general a single metric should not be used alone, but rather in a set of metrics, each used to capture some particular aspect of the overall modelling problem. In particular, this work is centered around the regression of bio-physical forest parameters, and thus the prediction accuracy is typically evaluated using 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:

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{3.33}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{3.34}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{3.35}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.36}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \tag{3.37}$$

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

mean reference value.

3.3 Regularization

So far, the *conditioning* of the matrix inversion for the least squares solution in Eq. (3.6) has been ignored. In practice, both *underdetermined* and *overdetermined* matrices can lead to *ill-conditioned* matrix inversions. In the context of neural networks, this corresponds to overfitting the model to the training data. Similarly to other inverse problems, also for neural networks it is possible to apply *regularization* strategies, which improve the conditioning. Intuitively, regularization is a process that changes the solution landscape, essentially restricting the space of possible solutions. In general, these techniques are used to improve the generalization capability of a model, and can be formally divided between *explicit* and *implicit* regularization.

3.3.1 L2-Norm

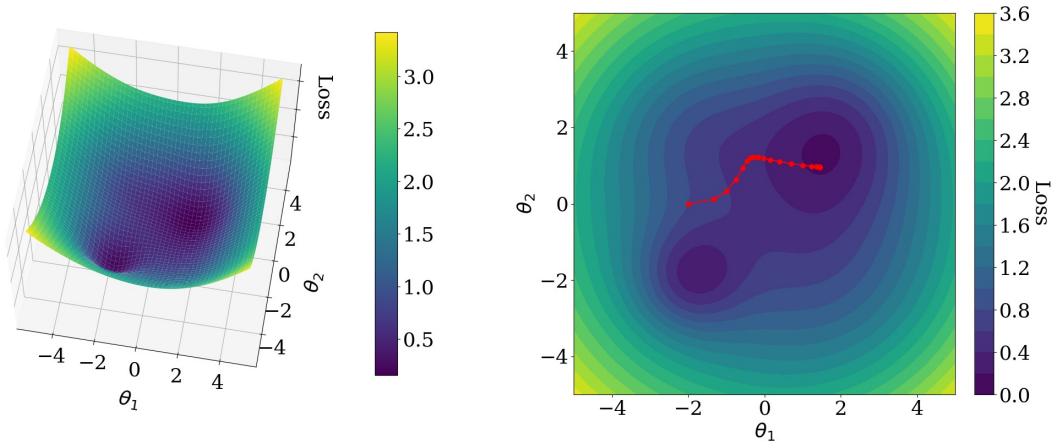
A simple way to reduce overfitting is to add an extra term to the loss function, thus biasing the minimization towards solutions that jointly minimize both the original prediction loss and this additional explicit regularization term.

A popular choice is the $l2$-norm of the model parameters $\boldsymbol{\theta}$, such that:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 + \lambda \sum_j \omega_j^2 \right\} \tag{3.38}$$

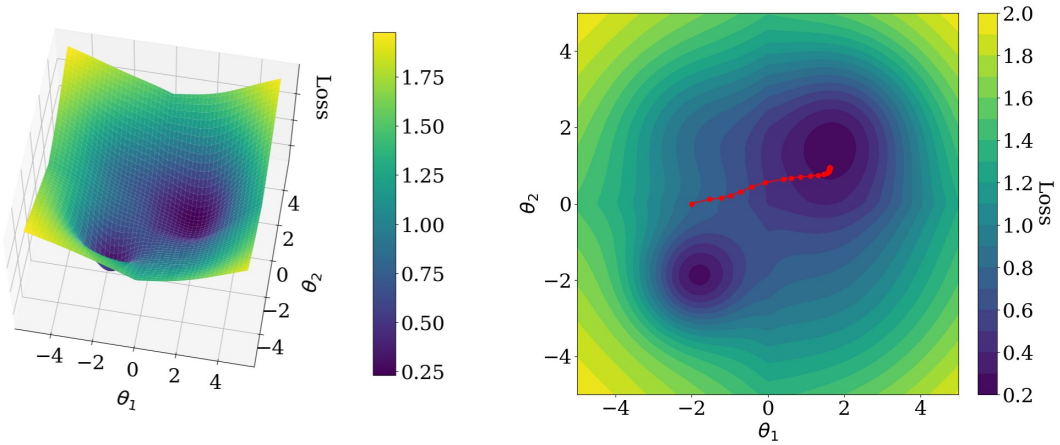
where ω_j is the j^{th} model parameter and λ is a pre-determined weighting coefficient used to balance the impact of the two terms. In practice, it is common to exclude the bias-coefficients from regularization, thus this technique is often referred to as *weight decay* [56].

As an example, applying the $l2$-norm to the GD example shown in Fig. 3.3 can lead to a better solution, as shown in Fig. 3.8.



(a) (b)

Figure 3.8: Impact of the $l2$-norm on the loss landscape of Fig. 3.3 (a). Using the GD method now leads to an improved local minima.



(a) (b)

Figure 3.9: Impact of the $l1$-norm on the loss landscape of Fig. 3.3 (a). Using the GD method now leads to an improved local minimum.

3.3.2 L1-Norm

Another regularization method consists in applying the $l1$-norm of the model parameters $\boldsymbol{\theta}$, such that:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 + \lambda \sum_j |\omega_j| \right\} \tag{3.39}$$

As an example, applying the $l1$-norm to the GD example shown in Fig. 3.3 can also lead to a better solution, as shown in Fig. 3.9.

3.3.3 Stochastic Gradient Descent

Conveniently, it can be shown that the GD-based optimizers offer some *implicit regularization* [65]. The authors in [66] demonstrate that the impact of SGD leads to the following relationship:

$$\mathcal{L}_{\text{SGD}}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \frac{l_0}{4}\left|\left|\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}\right|\right|^2 + \frac{l_0}{4B}\sum_{b=1}^{B}\left|\left|\frac{\partial \mathcal{L}_b}{\partial \boldsymbol{\theta}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}\right|\right|^2 \tag{3.40}$$

where $\mathcal{L}(\boldsymbol{\theta})$ is the original loss, l_0 is the learning rate, B is the batch-size and $\mathcal{L}_b$ is the loss for the b^{th}-minibatch. Note that the first term penalizes solutions with large gradients, corresponding to large curvatures in the loss landscape, which can lead to overfitting. The second term instead penalizes solutions that lead to large variance of the minibatch gradients, which intuitively should lead to more generalized solutions.

3.3.4 Dropout

The *dropout* operation consists in randomly deactivating neurons with a certain probability p, thus removing their contribution to the subsequent layers. This promotes smaller weights, in order to minimize the impact of loosing some of the connections. Furthermore, it reduces the amount of co-adaptation between the units, encouraging each unit to individually better contribute to the output.

Applying dropout is equivalent to applying multiplicative *Bernoulli noise* to the weights:

$$\hat{w}_{lij} = w_{lij}\epsilon_{li} \tag{3.41}$$

where w_{lij} is the weight of j^{th}-unit in the l^{th}-layer taking the i^{th}-output from the previous layer, $\epsilon_{li} \sim \text{Ber}(1-p)$ is the Bernoulli noise term with probability p, and $\hat{w}_{lij}$ is the resulting noisy weight.

During inference, the noise is typically deactivated. But, since now all neurons are active, it is common to multiply each weight by $(1-p)$, such as to keep their expectations equal during testing:

$$w_{lij} = \hat{w}_{lij}\mathbb{E}[\epsilon_{li}] = \hat{w}_{lij}(1-p) \tag{3.42}$$

This step is known as the *weight scaling inference rule*.

3.3.5 Batch Normalization

As discussed in Sec. 3.2.3, especially deep models can be affected by problems of vanishing or exploding gradients. A common way to mitigate this problem is to standardize the statistics of the output of each layer, such as to guarantee a stable backpropagation of the loss gradient. The most popular of these normalization layers is the so-called *batch normalization* [67]. Similar to the dropout layer, the batch normalization layer has also a different behaviour for the training and inference phases.

During the training phase, first the current mininbatch mean and variance are computed:

$$\boldsymbol{\mu}_{\mathcal{B}} = \frac{1}{n_{\mathcal{B}}} \sum_{\boldsymbol{x} \in \mathcal{B}} \boldsymbol{x} \tag{3.43}$$

$$\boldsymbol{\sigma}^2_{\mathcal{B}} = \frac{1}{n_{\mathcal{B}}} \sum_{\boldsymbol{x} \in \mathcal{B}} (\boldsymbol{x} - \boldsymbol{\mu}_{\mathcal{B}})^2 \tag{3.44}$$

where $n_{\mathcal{B}}$ is the size of the minibatch $\mathcal{B}$. The obtained values are subsequently used to standardize the minibatch values to zero mean and unitary variance:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} - \boldsymbol{\mu}_{\mathcal{B}}}{\sqrt{\boldsymbol{\sigma}^2_{\mathcal{B}} + \epsilon}} \tag{3.45}$$

where $\epsilon > 0$ is a small constant to avoid dividing by zero. The two trainable parameters $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ are then used to rescale the minibatch to mean $\boldsymbol{\theta}$ and variance $\boldsymbol{\gamma}$:

$$\tilde{\boldsymbol{x}} = \boldsymbol{\gamma} \odot \hat{\boldsymbol{x}} + \boldsymbol{\theta} \tag{3.46}$$

where $\odot$ is the pointwise multiplication operator. Finally, $\boldsymbol{\mu}_{\mathcal{B}}$ and $\boldsymbol{\sigma}^2_{\mathcal{B}}$ are used to update the exponential moving averages of the mean $\boldsymbol{\mu}_i$ and of the variance $\boldsymbol{\sigma}^2_i$, respectively:

$$\boldsymbol{\mu}_i = \alpha \boldsymbol{\mu}_{i-1} + (1 - \alpha) \boldsymbol{\mu}_{\mathcal{B}} \tag{3.47}$$

$$\boldsymbol{\sigma}^2_i = \alpha \boldsymbol{\sigma}^2_{i-1} + (1 - \alpha) \boldsymbol{\sigma}^2_{\mathcal{B}} \tag{3.48}$$

where α is a pre-selected memory hyperparameters. Note that, during training, a continuous update of the batch normalization statistics and parameters is necessary as these are required to co-evolve with the other changing parameters of the model, although it is not fully understood whether rescaling is strictly necessary [68].

During the inference phase, all layer parameters are frozen, and thus the standardization step in Eq. (3.45) is performed using the moving average results from Eq. (3.47)

and Eq. (3.48).

An empirical consequence of the use of batch normalization layers is that these allow for the use of higher learning rates, which improve the overall training speed.

3.3.6 Early Stopping

Another heuristic can be derived from the validation problem described in Sec. 3.2.5. Instead of letting the model train for defined amount of time, risking that the resulting model is either underfitting or overfitting, it is possible to actively track the progress of the validation loss and stop the training once it starts increasing. In the end, the model weights corresponding the lowest validation loss are selected. This practice is called *early-stopping* and it typically has one single variable, known as *patience*, which determines after how many consecutive validation runs without improvements the training should be stopped.

3.3.7 Data Augmentation

Another popular strategy is that of *data augmentation*, which consists in creating new training examples from the existing ones by applying some kind of manipulation. Commonly used ones are flipping, rotations, cropping, but also the addition of noise and other signal alterations. The resulting samples should encourage the model to become invariant to these types of transformations.

3.4 Convolutional Neural Networks

When dealing with imagery, the data is *structured*, and each of its values, known as a *pixel*, is statistically related to its neighbours. Furthermore images can have vastly different dimensions.

Sec. 3.2.2 introduced the concept of universal function approximation and of deep learning in the context of MLPs. An important characteristic of this neural network topology is that it maps *unstructured* input data to its corresponding output values, meaning that, when applied to images, an MLP has no implicit notion of spatial correlation, but instead treats all input values equally. A further consideration comes from the fact that each input or intermediate value is connected to all the neurons in the following layers. This has two fundamental consequences:

- The input images would always need to have the same size, since the input dimensionality is hard-coded by the dense layer's dimension. The only alternative solution for allowing multiple images sizes consists in training a dedicated model for each possible input shape, an extremely unactracive solution.

- The number of model parameters is enormous. For example, considering a small 256×256 pixel RGB input image, a single Perceptron would required $\approx 2 \cdot 10^5$ weights. It naturally follows that a shallow learner with a number of units proportional to the input dimensionality, would already reach $\approx 4 \cdot 10^{10}$ weights with just one layer! This leads to proportionally large requirements in terms of dataset size, memory capacity and computation resources.

It then becomes apparent how MLPs are ill-suited for dealing with imagery.

The solution to these problems is a special type of artificial neural network (ANN) known as *convolutional neural network* (CNN). In more specific terms, a CNN is a network topology that replaces dense layers with so-called *convolutional layers*.

3.4.1 The Convolutional Layer

A convolutional layer computes the *convolution* between an input feature map and a *kernel-function*, by sliding the kernel across the input and computing the sum of the products between the overlapping kernel- and input-elements. The elements in these kernels then constitute the learnable weights of the model, and are thus inferred from the training set, as seen for the MLP case.

It is important to clarify that in reality what is computed is the mathematical *cross-correlation* between the feature and the kernel. This is done because in practice there is no strict need to formalize an order inversion in the learnable kernel weights, and thus the word "convolution" is used to loosely describe the cross-correlation operation.

For 3D-imagery, consisting of two spatial dimensions and a further channel dimension, it is most common to use 2D-convolutional layers, which compute the cross-correlation over a 3D tensor $\boldsymbol{X}$, by sliding a 3D kernel $\boldsymbol{\Omega}$ along two selected dimensions (e.g. the spatial dimensions). The 2D-convolutional layer computation of a single output value $y(i,j)$ can thus be described as:

$$y(i,j) = \beta + \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} \sum_{c=0}^{C-1} \{ w_{(u,v,c)}\, x_{(i+u,j+v,c)} \} \tag{3.49}$$

where (i,j) is the row-column pixel position in the image, C is the number of features

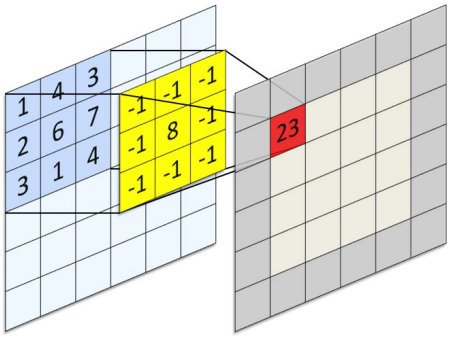


Figure 3.10: Graphic representation of the 2D convolution operation typically used in CNNs. The output value (in red) is the result of the convolution between the input matrix (in blue) and the *high-pass* kernel function (in yellow). Note that the border values in the output matrix are greyed out to indicate that it is not possible to compute the result for these values.

in the image, and H and W are respectively the height and width of the kernel. Note that the kernel function $\boldsymbol{\Omega}$ is in fact itself 3D, has a total of $H \cdot W \cdot C$ parameters, and is iteratively moved across the two spatial dimensions of the input image $\boldsymbol{X}$. On the other hand, the resulting output is a single feature and is thus 2D. In practice, this operation is performed on the same input map using multiple unique kernel functions, and thus the channel count after a convolutional layer is determined by the number of kernel filters used.

It should also be noted that the shape of the output might be also smaller than that of the input. The reason is that the convolution can only be performed where the kernel fully covers an equally sized portion of the input. Thus, for kernels with $H > 1$ and $W > 1$ this effectively shrinks the output size by not being able to compute the output for the input pixels at the border. The resulting operation is thus called *valid convolution*, and an example can be seen in the 3×3 convolution in Fig. 3.10. In general, the size reduction can be computed to be $H - 1$ and $W - 1$, with odd-sized kernels leading to a symmetric border loss, and vice-versa for even-sized ones. An alternative approach consists in compensating ahead, by *padding* the input image. This can be done by simply adding a constant value (e.g. zero), by replicating the border values or by mirroring them. Subsequently, by performing the same convolution as before the input size is now preserved. This operation is therefore called *same convolution*.

A further consequence of Eq. (3.49) is that, for $H > 1$ and $W > 1$, each pixel value in the output is influenced not only by the corresponding pixel value in the input, but also by its neighbouring pixels. Therefore, for each layer, the kernel-size $(H; W)$ defines the size of the considered pixel neighbourhood.

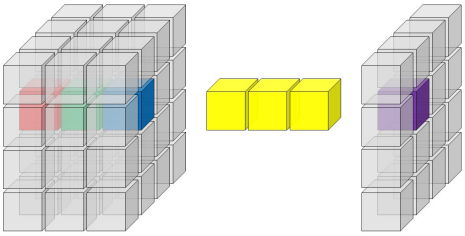


Figure 3.11: Example of a 1×1 convolution of a RGB input tensor (left) with a kernel function (center) to produce a new output tensor (right). Note that in a 2D-convolution, the kernel has the same depth as the input tensor, while the output is a tensor of unitary depth (i.e. a single feature).

Pointwise Convolutions

A particular intuition arises from Eq. (3.49), when $H = 1$ and $W = 1$. In this case, only the individual pixel values themselves are considered when computing the output values, and thus the equation can be simplified to:

$$y(i,j) = \beta + \sum_{c=0}^{C-1} \{w_c \, x_{(i,j,c)}\} \tag{3.50}$$

which is equivalent to a single layer MLP, applied to the feature column at pixel position (i, j). In other words, the pointwise convolutions simply computes a weighted combination of the input features for a given position. An example of a pointwise operation is shown in Fig. 3.11. In practice, this is often use to efficiently increase or decrease the dimensionality of the problem.

3.4.2 Receptive Field

An important consequence of the convolutional layer is the enlargement of the spatial context that is taken into consideration during the computation of an output point. The subsequent use of these layers increasingly extends the input spatial context considered for the output. The overall resulting spatial context of a CNN is known as the *receptive field* (RF) of the model, as a parallelism with the receptive field of biological vision system. Fig. 3.12 shows an example of a receptive field for a three layered CNN using 3×3 kernel filters.

Depending on the application, it is often desirable to increase the RF of a CNN in order to better allow it to contextualize the input for a given output. The way this is usually achieved is by stacking multiple convolutional layers, each using small kernels. In practice, this is preferable over large-sized kernels because it reduces the complexity

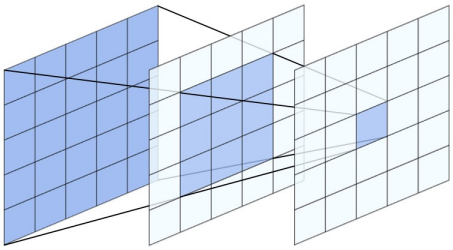


Figure 3.12: The output pixel dependency on the pixels in the previous layers of a CNN using 3×3 kernel filters. The pixels in dark blue highlight the expansion of the model's receptive field.

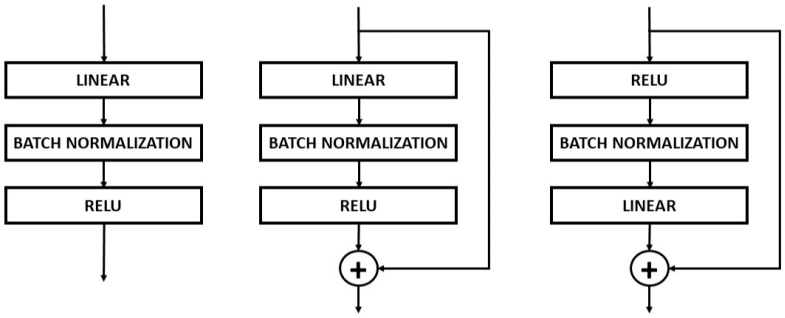


Figure 3.13: Comparative example of a plain convolutional block (left), a basic residual block (center), and a pre-activated residual block (right).

of the model. As an example, the RF covered by a single 7×7 convolution, can be achieved by a sequence of three 3×3 convolutions, but while the former requires 49 weights, the latter option only requires 27.

In fact, the search for parameter-efficient strategies to increase the RF has led to unique solutions. In the following, some of the most popular ones are briefly discussed, together with other design choices that might impact the receptive field of a CNN.

Residual Connections

Arguably one of the more crucial contemporary developments in deep learning is that of the *residual* or *skip connection*. Prior to its introduction in [69], it was observed that adding layers to a network beyond a certain limit would see both training and testing performance start to deteriorate. This phenomenon is known as the *degradation* problem and is still not fully understood. The leading hypothesis is that, as models get deeper, the sensitivity of the output to the input and earlier layers increases, causing the loss landscape to become increasingly erratic and harder to optimize for [70].

The introduction of the residual connection in [69] solved this issue by introducing

parallel signal paths which branch off from the main model sequence, offering a bypass to a later part of the model. Here, the merging is classically performed by simply adding the corresponding signal together, such that:

$$a_i = \mathcal{F}(a_{i-1}) + a_{i-1} \tag{3.51}$$

where a_{i-1} is the bypassed signal, $\mathcal{F}(\cdot)$ is the transformation performed by the bypassed model section, and a_i is the resulting outcome of the operation.

By re-arranging Eq. (3.51), it is possible to see that the bypassed model section essentially acts as a subnetwork that learns the residual between the output and input signal. Assuming that multiples of the *residual blocks* are used, it is possible to write the output at layer L with respect to the input at layer l as:

$$a_L = a_l + \sum_{i=l}^{L-1} \mathcal{F}(a_i) \tag{3.52}$$

As a consequence, the gradient backpropagation to an arbitrary layer l can thus be computed as:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial a_l} &= \frac{\partial \mathcal{L}}{\partial a_L} \frac{\partial a_L}{\partial a_l} \\ &= \frac{\partial \mathcal{L}}{\partial a_L} \frac{\partial}{\partial a_l} \left(a_l + \sum_{i=l}^{L-1} \mathcal{F}(a_i) \right) \\ &= \frac{\partial \mathcal{L}}{\partial a_L} + \frac{\partial \mathcal{L}}{\partial a_L} \sum_{i=l}^{L-1} \frac{\partial}{\partial a_l} \left(\mathcal{F}(a_i) \right) \end{aligned} \tag{3.53}$$

where the first term suggests that the gradient at the l^{th}-layer directly depends from the gradient at the model output.

Another consequence of Eq. (3.52) is that the output at layer L is the ensemble of $L - l$ smaller networks, with each of them being itself an iteratively smaller ensemble. Consequenlty, Eq. (3.53) suggests that during training the model can independently adjust each parallel path, determining its own effective depth.

As a final remark, the order of operation in a residual block is important, as the conventional order of linear and ReLU non-linear operations would limit each residual subnetwork to only learn positive residuals. It is thus common to invert the order and place a linear layer before the first residual block, such as to avoid negative inputs to be clipped to zero by the ReLU, leading to the so-called *pre-activated residual block*. A comparion of the three discussed types of residual blocks is presented in Fig. 3.13.

Chapter 4

Forest Parameter Estimation: State of the Art Overview

In this chapter, an overview of the existing state of the art techniques for forest parameter estimation is presented. Consistently with the core activity of this thesis, the focus of the literature review is predominately set on the retrieval of canopy height model (CHM) and above-ground biomass (AGB) estimates. A high-level summary of the major approach families can be found in Tab. 4.1.

As already mentioned in Sec. 1.2, the most accurate way of measuring tree parameters is achieved through direct in-situ measurement campaigns [27] [71] [18]. In particular, some properties such as tree height, diameter at breast height or crown width, can be obtained using non-invasive approaches, while others, such as the above-ground biomass, have to be acquired either invasively (i.e. by cutting down and measuring trees) or through indirect modelling [12].

As an example, the work in [72] developed a pantropical allometric model for the estimation of tree-level above-ground biomass from the tree's diameter, height and wood density. Fitting of the model was achieved by selecting only data from destructively-acquired tree parameters, collected across 58 sites and with a trunk diameter $\geq$5 cm, resulting in a total of 4004 samples. Compared to the invasively acquired AGB measurements, the resulting pantropical model achieved a mean coefficient of variation (CV) of 56.5% and a mean relative bias of 5.31%, while local models, individually tuned to their respective sites, achieved a similar CV of 47.4% but a better mean relative bias of 0.38%. With a tree-level AGB uncertainty of $\approx$50% the authors reflected on the limitations of their model, which instead performed much better on a plot-level, obtaining an uncertainty of $\approx$10-15% at $50^2\,m^2$ plots, which further dropped to $\approx$5-10%

for $100^2\,m^2$ plots. Interestingly, the authors evidenced that most of the spread is explained by intra-site variations, rather than inter-site ones, highlighting the limitations in generalization capability. It was also noted that the exclusion of the wood specific gravity from the equation would lead to significantly worse results.

In [11] the use of terrestrial laser scanning (TLS) equipment was tested to improve the collection time and precision of AGB estimates over traditional non-destructive measurement approaches. Compared to destructively-sampled quantities, the TLS-based estimates achieved a correlation coefficient of 0.98, whereas the ones derived from allometric equations stopped at a lower score between 0.68 and 0.78. In terms of bias, the authors found a positive bias of 9.68% for the LiDAR derived predictions and an underestimation of -29.85% to -36.57% for the allometry-derived ones. Evidently, field surveys provide the best accuracy on a tree or a plot level, but also remain too expensive and time consuming for large-scale surveys.

Up to country-level scales, estimates can be effectively obtained from airborne LiDAR (ALS) measurement campaigns, such as those acquired with the Laser Vegetation Imaging Sensor (LVIS) [53][73], which are capable of achieving a high vertical accuracy in the order of centimeters [18]. Crucially, vertical structure parameters (e.g. forest height) are straightforward to be derived from the measured LiDAR signals, while more stuctural parameters (e.g. AGB) require the use of more complex models.

As an example, the authors in [27] analyzed a series of allometric equations for estimating tree-level stem diameter and AGB from ALS data. Their AGB estimates obtained an RMSE of 1.78 Mg and a bias of 30.1%, while the on-ground derived estimates achieved a better RMSE of 0.86 Mg and a similar mean bias of 27.7%. Crucially, the authors found a strong direct correlation between the height of the trees and their AGB values, thus hypothesizing that the latter can be effectively obtained from RS data. Furthermore, the authors indicate that the AGB scaling relationships may also systematically vary among different biomes, and that accounting for these differences could further improve the estimation accuracy of the AGB models.

Coverage of larger geographic extensions can be obtained with spaceborne LiDAR systems, such as the Global Ecosystem Dynamics Investigation (GEDI) [20], or the Ice, Cloud and land Elevation Satellite 2 (ICESat-2) [19], which sparsely sample Earth's surface with a sequence of parallel laser beams, each resulting in an on-ground diameter in the order of few tens of meters. Besides the scattered acquisition patterns, the currently existing spaceborne systems are limited by their subpar performance. On the one hand, they suffer from relatively poor horizontal geo-location accuracies: for example, the latest generation of GEDI products achieves a 1σ error of 10.2 m [74], compared

to the 1 m horizontal accuracy of LVIS [18]. On the other hand, the transmitted signal power appears to be insufficient to consistently reach the ground, especially in the presence of dense vegetation [75].

In order to achieve large-scale, wall-to-wall estimates, it becomes fundamentally necessary to derive these parameters from satellite-derived imagery products, by relating the bio-physical forest parameters to the acquired spaceborne feature-maps. This can be achieved by either physical-based approaches, which model the interaction of the measured electromagnetic signal with the forest, or by relying on data-driven approaches, which directly learn the underlying relationship from large-amounts of informative case samples.

4.1 Physical-Based Models

In this scenario, Synthetic Aperture Radar (SAR) sensors have received great attention from the remote sensing community, as the interaction between the electromagnetic waves and the imaged scatterers strongly depends on the geometrical structure and the dielectric properties of the target, i.e. on the characteristics of the vegetation. In [76], the Water Cloud Model (WCM) was proposed to relate the measured SAR backscatter over canopies to the humidity of the ground, as well as to that of the canopy itself, modelled as a cloud of water droplets held in place by the canopy structure. More recently, the focus on forest height estimation has shifted towards the use of SAR polarimetry (PolSAR), interferometry (InSAR), or a combination of the two (PolInSAR) [77]. Here, the Random Volume over Ground (RVoG) model represents one of the most studied and best understood physical interpretations of the InSAR microwave interaction with the forest structure[78].

Notably, the full model requires either multiple acquisitions with different InSAR baselines or PolInSAR measurements in order to be inverted and to deliver a vegetation height estimate. In [78] and in [79], the authors elaborated on the possibility of inverting the model using only a single-baseline fully-polarimetric (i.e. quad-pol) acquisition. Modern spacerbone SAR systems, such as the German TanDEM-X mission [23][80], are indeed also capable of acquiring fully-polarimetric InSAR products, but these do not represent the operational acquisition mode and can instead only be acquired as experimental products over limited test sites. Therefore, much effort has gone into the definition of effective strategies to invert the model using less informative features.

The work in [14] evaluated three different InSAR models: the interferometric water

cloud model (IWCM), the random volume over ground model (RVoG) and a simpler penetration depth (PD)-based model. The parameters of all models were obtained by least-squares regression with the LiDAR-based AGB reference estimates. These where applied to 18 TanDEM-X single-pol InSAR acquisitions over the hemiboreal site of Remningstorp (Sweden) with the objective of estimating AGB. All three models performed very similarly on average, with the IWCM achieving an average RMSE of 24.3 m, the RVoG of 27.0 m and the PD of 24.6 m, respectively. By inversely weighting and combining all of the 18 IWCM estimates according to the square of their respective acquisition height of ambiguities, resulted in an improved relative RMSE of 17.3 Mg/ha and a coefficient of determination of 0.93, with reference to forest stands larger than 1 ha.

In [28], the authors assessed the potential of TanDEM-X dual- and single-pol acquisition for forest height estimation using the RVoG model. In particular, dual-pol inversion was obtained by assuming that the ground contributions of one of the two polarizations is zero, while for single-pol also an external DTM was used to determine the topographic InSAR phase corresponding to the ground. The inversion performance over the boreal site of Krycklan (Sweden) saw dual-pol achieve an RMSE of 2.02 m and a r^2 of 0.86, while single-pol an RMSE of 1.58 m and a r^2 of 0.91, respectively. Testing across the temperate Traunstein test site (Germany), resulted in the dual-pol inversion obtaining an RMSE of 2.8 m and a r^2 of 0.77, and single-pol a peak RMSE of 3.4 m and a r^2 of 0.86. Finally, testing over the tropical forest area of Mawas (Indonesia) resulted in a variable RMSE of 5.7-6.3 m and a r^2 of 0.54-0.69 for the dual-pol inversion, and an RMSE of 1.8-3.2 m and r^2 of 0.97-0.98 for the single-pol inversion, respectively.

A series of assumptions, compatible with the X-band radar frequency of the considered TanDEM-X sensor, were introduced in [81] to simplify the model down to a sinc-relationship and to obtain forest height estimates directly from the approximated sinc-function inversion, requiring only one single-baseline coherence amplitude. Validation of the model against LiDAR measurements acquired over the test site in Alberta (Canada) resulted in an RMSE of 2.9 m and a r^2 score of 0.75 at an output resolution of 56 m. Similarly, in [29] the authors investigated the impact of further parametrization of the zero extinction coefficient, sinc and linear model simplifications. They found that the resulting semi-empirical model offered improved robustness with respect to their original vanilla counterparts for the tested hemiboreal study area.

More recently, research attention has shifted towards sparse fusion strategies, aiming at exploiting the availability of modern dedicated spaceborne LiDAR-based missions (e.g. GEDI, ICESat-2) to retrieve the model parameters which are necessary for for-

est height inversion. In order to improve multi-baseline quad-polarization inversion accuracy, in [82] it was proposed to train a support vector machine (SVM) algorithm with sparse LiDAR samples in order to select the optimal baseline for the RVoG model inversion. The analysis was performed on data acquired with NASA's UAVSAR (L-band) and LVIS instruments in occasion of the 2016 AfriSAR campaign in Gabon, West Central Africa [73][18]. Performance results, validated against LVIS measurements for the study areas of Pongara, Akanda/Mondah and Lopé, saw the proposed baseline selection method improve the RMSE from 7.47 m to 5.64 m, from 9.39 m to 4.99 m, and from 9.60 m to 5.99 m, respectively, with respect to the traditional methods of baseline selection based on PolInSAR data alone.

In [83], the authors proposed to avoid the parametrization of a simplified vertical profile function and to instead estimate it directly from LiDAR waveforms. This occurs under the assumption that, despite the different wavelengths, the reflectivity profiles of TanDEM-X and of the LiDAR sensor are considered "similar enough" to build a robust approximation, with the strongest discrepancy deriving from the different acquisition geometries (side-looking for TanDEM-X and nadir-looking for LiDAR). A representative vertical scatterer profile was estimated by aggregating the normalized LiDAR profiles into a matrix, building its diagonalized covariance matrix and retrieving its first eingenfunction (i.e. zero order). Subsequently, the forest height was inverted from TanDEM-X coherence samples acquired over the AfriSAR test site of Lopé, and filtered for estimates falling below the interferometric phase center. This resulted in the removal of 10.69% of the estimates, achieving an RMSE of 8.16 m and a r^2 coefficient of 0.16. Finally, by generating two separate profiles for vegetation below 25 m and above 40 m, respectively, and interpolating the profiles in the transitional range, the authors achieved a RMSE of 8.62 m and a r^2 of 0.40 (after dropping 14.17% of the underestimated samples).

Similarly, in [84] a mean TanDEM-X vertical reflectivity profile was estimated using the zero-order eigenvector of diagonalized profile covariance matrix, derived from GEDI LiDAR waveforms. Using this approach, the authors generated a continuous 25 m forest height map of the island of Tanzania (Australia). In comparison with reference LiDAR measurements, their proposed approach achieved varying RMSE values between 6.6 m and 7.2 m, and r^2 values between 0.40 and 0.42, respectively, depending on the considered orbit direction and orthogonal baseline.

All of the works discusses up to this point have either used single-pass InSAR data [28][84] or have ignored the effects of temporal decorrelation in the case of repeat-pass interferometry [82]. In [85] an alternative to the RVoG model was proposed,

which builds upon the exponential temporal decorrelation model introduced in [50], by assuming that the variance of the Gaussian motion of the scatterers has a vertical dependency. The resulting model assumes zero across-track baseline, effectively ignoring volume decorrelation effects and instead assuming a purely temporal effect that depends on the height and the polarization-dependent properties of the canopy. In a subsequent work in [86], the authors combined this temporal decorrelation model with the RVoG to propose a novel Random Motion over Ground (RMoG) model. Furthermore, they presented a parameter estimation strategy from single-baseline quad-pol InSAR data. Validation of the model with L-band UAVSAR data acquired over the Harvard Forest, USA resulted in an RMSE of 3.6 m, superior to the 11.7 m obtained with the RVoG model.

Alternatively, it is also possible to exploit multiple SAR acquisitions to directly image the vertical structure of the vegetation. This technique, known as *SAR tomography* (TomoSAR), consists in acquiring a series of SAR images with different vertical baselines, leading to the composition of the so-called tomographic stack. Subsequently, this diversity allows for focusing the signal in elevation, in addition to the conventional SAR range and azimuth directions, yielding the reconstruction of the vertical scattering profile of the illuminated volumetric targets [87].

For example, in [88] TomoSAR processing is applied to L-band and P-band stacks of 10 images each, acquired using DLR's F-SAR system over the Lopé, Mabounié and Rabi sites of the AfriSAR campaign (Gabon). When compared to LVIS LiDAR measurements, the ground elevation estimates resulted in a variable RMSE of 2-5 m, whereas the canopy height ones were consistently underestimated by 5-10 m for L-band, and by about 10 m for P-band, respectively.

The work in [89] also used the P-band F-SAR measurements to estimate the ground topography across the AfriSAR sites. The performance analysis yielded very similar results, obtaining an RMSE of 2.8 m for Lopé, of 2.4 m for Mondah and of 5.6 m for Mabounié, respectively.

In [90], the authors assessed and compared the capabilities of airborne P-band TomoSAR and GEDI spaceborne LiDAR measurements to estimate the canopy and ground heights, using airborne LiDAR measurements acquired over the tropical forest site at Paracou (French Guiana) as reference. The canopy height results saw the TomoSAR achieve an RMSE of 3.51 m and GEDI of 4.70 m, respectively, while the ground elevation estimates yielded an RMSE of 1.90 m and of 3.34 m, respectively.

The biggest limitation of the TomoSAR approach remains the considerable number of acquisitions required in order to keep the ambiguities manageable. Currently no ac-

tive spaceborne system is capable of operationally delivering the required measurements at the available frequencies (L-, C- and X-band); the existing revisit times and baselines are the cause of significant decorrelation, making the collection of a tomographic stack extremely challenging. First attempts were demonstrated using TanDEM-X bistatic time-series, thanks to the varying orbit formation [91].

The upcoming ESA BIOMASS mission [92] will for the first time offer a dedicated acquisition phase capable of generating tomographic products at a near-global scale. Due to bandwidth constraints at P-band, it follows that the spatial resolution will be significantly lower compared to the existing spaceborne systems capabilities, reducing the attractiveness for operational product generation.

Interestingly, the work in [93] suggested an alternative to overcome the need for a tomographic stack. Starting from related works in [94] and [95], the authors proposed to locally cluster the unwrapped single-baseline InSAR phase information into a vertical sample distribution, known as the *phase histogram* (PH). Taking advantage of the fact that, at single- or few-look resolution, the heterogeneously distributed vegetation presents gaps that can be penetrated by the SAR signals, the proposed process resulted in the creation of tomography-like 3D-stacks, which correlate with the vertical structure of the canopy. In particular, different filtering and detection algorithms were applied to UAVSAR-derived L-band InSAR products acquired over Lopé, Gabon. Comparison against the LVIS measurements from the AfriSAR campaign, showed that the PH allowed for estimating the DST with an RMSE of 4.7 m, and the DTM with an RMSE of 3.1 m, respectively. Regarding the CHM, an RMSE of 5.1 m and a R^2 score of 0.90 where obtained, outperforming the Quad-pol InSAR approach applied to the same data in [82]. Furthermore, a qualitative comparison with the TomoSAR-based works in [88] and [89] yielded similar or better results for the DTM estimation, but arguably worse results for the canopy height estimation.

The comparison between the PH and TomoSAR methods is further explored in [96] and [97], based on stacks of 30 airborne images acquired at L-band over two sites in the Eifel Park (Germany). Canopy height retrieval estimates, evaluated against reference LiDAR measurements, saw the TomoSAR approach obtaining an RMSE of 2.63 m and of 1.84 m, respectively, whereas the PH method archieved an RMSE of 4.35 m and 5.46 m, respectively. It was observed, that the combination of a strong topography and L-band data potentially limited the PH approach, suggesting further investigation into high-resolution, high-frequency data.

Moving now from elevation metrics to AGB, the authors in [98] assessed the extent and the generality of the linear relationship existing between AGB and canopy height

metrics, and the volumetric decorrelation factor computed from a TanDEM-X bistatic acquisition. Validation was performed across different pantropical field plots in Paracou (French Guiana), Mawas (Indonesia), and the Lopé, Mondah and Rabi sites (Gabon) of the AfriSAR campaign. The result highlighted a strong linear relationship of the volume decorrelation factor with LiDAR data, achieving an RMSE of 6.4 m and a R^2 score of 0.72 for with the canopy height, and an RMSE of 88.2 Mg/ha and a R^2 score of 0.59 for the AGB, respectively.

4.2 Data-Driven Models

Due to the lack of significant penetration and, consequently, interaction with the vertical structures of canopy, optical data has remained overwhelmingly limited to the use of data-driven approaches. For example, the authors in [16] tried to overcome the need for allometric AGB models by testing the regression performance of different machine learning techniques: ordinary least-squares regression, ridge regression, support vector regression, random forest, stochastic gradient boosting and the cubist regression model. Testing was applied to ALS and hyperspectral imagery (HSI) data, individually and combined, acquired across 12 sites within the Brazilian Amazon forest. For each source, multiple known features and indexes were manually generated and testes for importance as inputs to the different regression models. These consistently evidenced that the upper canopy cover and tree height percentiles represented the most relevant LiDAR proxies for AGB, while the absorption bands in the NIR and SWIR spectra formed the most relevant features for the HSI data. The experiments confirmed the reliability of LiDAR-derived features in predicting AGB, but also concluded that HSI metrics could achieve similar performance, albeit requiring more input variables. Furthermore, performance analyses resulted in no algorithm consistently outperforming the others, with the choice of the particular input source having a greater effect on the specific algorithms. The best LiDAR model achieved an RMSE of 67.6 Mg/ha and an R^2 score of 0.58, while the HSI model achieved very similar RMSE and R^2 scores of 68.1 Mg/ha and 0.58, respectively. Combining metrics from both datasets sensibly improved the accuracy, obtaining an RMSE of 57.7 Mg/ha and a R^2 score of 0.70.

In [99] the use of a random forest regressor was proposed for the prediction of AGB estimates from both Sentinel-1 and Sentinel-2 time-series. The analysis found that the Sentinel-2 dataset led to better results, achieving an RMSE of 60.6 Mg/ha and a MAE of 19.2 Mg/ha, while the Sentinel-1 acquisitions stopped at 78.6 Mg/ha and 25.6 Mg/ha,

respectively. Interestingly, the most accurate results were obtained by combining both sensors, obtaining an RMSE of 45.4 Mg/ha and MAE of 16.3 Mg/ha, suggesting that the sensors could effectively complement each other.

The work in [100] assessed the single- and multi-sensor performance of Sentinel-1 and Sentinel-2 for forest height estimation. Multiple features were manually generated for both sensors and used as inputs, individually or combined, to three different machine learning algorithms (ordinary least-squares regression, classification and regression trees, and random forest regression) trained using LiDAR samples covering the Atlantic Forest of Paraná (Brazil). The results saw the Sentinel-1 models performing the worst, with the Sentinel-2 ones performing better but with greater variance throughout the year, and the combined models also performing similarly well. Looking only at the best models for each sensor, Sentinel-1 achieved a MAE of 4.29 m and an RMSE of 5.38 m, Sentinel-2 a MAE of 3.88 m and a RMSE of 4.78 m, and Sentinel-1 and Sentinel-2 combined a MAE of 3.69 m and an RMSE of 4.48 m. Of all the considered features, the 20 m Sentinel-2 NIR and SWIR bands had the greatest impact on the regression performance.

A similar conclusion was reached in the analysis presented in [17]. Here, a comparison of the relative performance of Sentinel-2 and Landsat 8 imagery was evaluated for both MLP and regression tree models. The regression performance of various forest variables acquired across Southern Finland found that Sentinel-2 performed better than Landsat 8, with the Red Edge (NIR) and SWIR bands being especially relevant for the regression accuracy.

A random forest regression algorithm was applied for the estimation of forest fractional-cover, height and above-ground biomass density across Australia [101]. As predictions, the authors considered yearly geomedian multi-spectral data acquired using Landsat 5, 6 and 8, as well as HH and HV L-band backscatter maps acquired using ALOS/PALSAR, and sparse ICESAT altimetry height percentiles. Results for the final models saw a MAE of 0.07 for the fractional-cover and of 3.4 m for the forest height. The addition of ancillary temperature maps allowed the AGB model to reach an MAE of 80 Mg/ha.

Also of interest is the work presented in [102], which proposed the use of the k-nearest neighbours (kNN) regression algorithm to model the relation between the Sentinel-2 multi-spectral and TanDEM-X InSAR features, with the AGB and stem volume estimates over the Swedish national forest. It was found that the combination of the sources offered improvements over their individual use, with the TanDEM-X interferometric phase height being key to the volume estimation, while the optical bands were essential for tree species identification. The combined model achieved an overall

accuracy of 18.0 Mg/ha, after a linear error model was used in conjunction with the estimators to correct the sampling error over the reference fields.

Recently, Deep learning algorithms have received an constantly increasing amount of attention, in great part justified by the extraordinary results obtained in the fields of computer vision [69][34][103]. Specifically in the EO domain, these techniques have seen similar levels of success for the broader family of classification tasks [33][30], but (arguably) still remain severely under-researched for regression problems, such as that of bio-physical parameter estimation.

The work in [35] investigated the use of a modified Xception convolutional neural network (CNN), trained to generate forest height estimates at 10 m of resolution, from individual Sentinel-2 acquisitions. The model was trained and validated across two alpine regions in Switzerland using stereophotogrammetric measurements, and over the five AfriSAR sites (Gabon) using LVIS-derived measurements. In Switzerland, the proposed approach achieved a MAE of 2.0 m and an RMSE of 3.9 m, when considering only the estimates from the least clouded acquisitions, and of 1.7 m and 3.4 m, respectively, when combining the estimates using the median height. In Gabon, the authors obtained a MAE of 4.9 m and an RMSE of 6.5 m, using the least clouded acquisitions, respectively 4.3 m and 5.6 m when considering the median height. Of particular relevance is the importance analysis of the textural features, which, when replaced by simpler pixel-level predictors, result in severe performance losses, demonstrating the potential benefits of CNNs over conventional machine learning algorithms.

In [104] a complex-valued deep learning model was proposed for the prediction of forest height estimates from single-baseline L-band PolInSAR measurements, acquired over Traunstein (Germany) using DLR's E-SAR instrument. The reported validation accuracy achieved an RMSE of 2.59 m and a R^2 score of 0.90, outperforming that of the RVoG model, with an RMSE of 4.14 m and a R^2 score of 0.83 m, respectively. Interestingly, the proposed network is relatively shallow, suggesting that very deep architecture, such as the one used in [35], might not be required to achieve state-of-the-art performance.

A Generative Adversarial Network (GAN)-based approach was proposed in [36] to estimate forest height estimates from sparse LiDAR measurements and wall-to-wall PolInSAR imagery. In particular, a pan-sharpening formulation was used to incrementally improve the resolution of the GEDI-like LiDAR measurements by combining it with different polarimetric and interferometric features acquired using the L-band UAVSAR instrument. The framework consists in a generator model, which produces the high-resolution forest height estimates, and two discriminators, one assessing the

spatial similarity between the polarimetric decomposition features and the predicted forest height map and the other comparing the measured repeat-pass coherences with those obtained with the RMoG using the estimated forest heights. The reported model performance beats that of the RVoG-based approach detailed in [82], but the validation is limited to two small hand-picked regions within the main training areas, and no clear information on the training and testing strategies is provided.

In [37], the authors proposed a Bayesian Deep Learning (BDL) approach to overcome the problematic overconfidence of machine learning estimates. Their method consisted in a dual branch CNN network, based on the popular ResNeXt architecture, using both Sentinel-1 dual-pol backscatter maps and 12 Sentinel-2 multi-spectral bands as predictors, and estimating the Normally-distributed probability density functions for five different forest structure parameters. Five models were individually trained to minimize the model's posterior probability, and subsequently combined in an ensemble. The resulting framework, was validated at country-level over Norway, achieving state-of-the art performance. Furthermore, it was demonstrated that the ensemble could generate well calibrated uncertainty estimates, enabling the integration into pre-existing statistical frameworks or the self-determination of the estimates' reliability.

The results of this work and that in [35] were expanded upon in [105], where the authors proposed to train their CNN ensemble using sparse 25 m footprint GEDI samples as ground-truth data. This resulted in the generation of a global-scale canopy height map based on sole Sentinel-2 acquisitions as input. The performance of the proposed approach was evaluated using a mixture of independent LVIS and ALS measurement campaigns, sparsely covering North America, Costa Rica (Central America), Europe and Gabon (Africa), they achieved an RMSE of 7.9 m and a mean error (ME) of 1.7 m. However, the weak supervision from GEDI samples represents a major limitation of this approach, due to the limited 25 m footprint resolution, the sparsity of the samples and the geolocation errors of first-generation products used for this study. Despite the very promising performance results, the claim of 10 m resolution products has to be doubted due to the complete or partial reliance of lower resolution LVIS measurements for their validation.

Finally, the recent work in [38] is also highlighted, which proposed to estimate the canopy height of maritime pines in the Landes de Gascogne (France) using a U-Net as the model, and temporal composites of both Sentinel-1 and Sentinel-2 acquisition as main predictors. The model is trained at 10 m of resolution using GEDI measurements as sparse reference, Validation was performed using GEDI, French national forest inventory data, local forest inventory data and 3D-stereo reconstruction products from

Literature Review Summary			
Approach Family	**Citations**	**Advantages**	**Disadvantages**
In-Situ	[11][12][72]	Best precision	Superior financial cost; worst coverage
Airborne LiDAR	[18][27][53]	Great precision for forest height retrieval; wall-to-wall measurements	Financial cost; limited coverage
Spaceborne LiDAR	[19][20]	Good trade-off between precision and coverage	Sparse sampling; inconsistent precision
Physical-Based Modelling	[14][28][76][78][83][85]	Strong interpretability of the physics driving the problem	Requires quad-pol data, ancillary data or simplifications; simplistic assumptions are required
SAR-Tomography	[87][88][89][90][91]	Precise reconstruction of the vertical backscattering profile	Requires a large stack of acquisitions with different interferometric baselines
Phase Histogram	[93][94][95]	Requires only a single InSAR pair; approximation of vertical backscattering profile	Assumption of dominant scatterer; noisy approximation of Tomography
Machine Learning	[16][99][100][101][102]	Independence from a priori assumptions; modelling capacity	Only relies on pixel-wise information; challenging interpretation
Deep Learning	[35][36][37][38][104][105]	Independence from a priori assumptions; superior modelling capacity	Requires large amounts of data; challenging interpretation

Table 4.1: A summary of the different state-of-the-art forest parameter regression approaches, with an emphasis on the major advantages and drawbacks for each approach-family.

Skysat imagery, and results strongly vary depending on the considered dataset. Interestingly, the authors noted that Sentinel-2 alone performs worse than Sentinel-1 alone, which is in stark contrast with all previous works and could be attributed to the long temporal medians or to issues in the cloud masks.

Chapter 5

Earth Observation Missions, Datasets and Pre-Processing

This chapter presents the different types of EO sensors and datasets used in the context of this thesis. Additionally, the dataset pre-processing pipeline, developed within the scope of the activity, is also described.

5.1 TanDEM-X

This section presents an overview of the TerraSAR-X and TanDEM-X satellites and of the TanDEM-X mission, focusing on the generation and derivation of the products used in the proposed investigations discussed in Ch. 6 and Ch. 7.

5.1.1 The TerraSAR-X and TanDEM-X Satellites

The TerraSAR-X (TSX) and TanDEM-X (TDX) satellites are two German spaceborne SAR systems developed in the frame of a public–private partnership between the German Aerospace Center (DLR) and Airbus Defence and Space (former EADS Astrium GmbH) [106][23].

TerraSAR-X was launched in June 2007 and TanDEM-X in June 2010, both presenting nearly identical components as summarized in Tab. 5.1. The two satellites fly in a dusk-dawn, sun-synchronous orbit, at a nominal elevation of 514 km and with an orbit repeat-cycle of 11 days [80].

Both satellites serve the two homonymous missions:

- The TerraSAR-X mission (TSM) [106], which aims at generating high-quality SAR

TerraSAR-X and TanDEM-X System Parameters		
Orbit	Elevation	514 km
	Inclination	Dusk-Dawn Sun-Synchronous (97.4°)
	Tube	250 m
Antenna Size	Azimuth	4.8 m
	Elevation	0.7 m
Signal	Center Frequency	9.65 GHz
	Bandwidth	≤300 MHz
	PRF	2 KHz - 6.5 KHz
	Polarization	HH, HV, VV, VH (Quad-Pol)
Acquisition	Modes	Spotlight, Stripmap, ScanSAR
	Swath-Width	10 km (Spotlight) - 100 km (ScanSAR)
	Ground Resolution	0.25 m (Spotlight) - 40 m (ScanSAR)
	Look Angle	15° - 60°

Table 5.1: Summary of the main TerraSAR-X and TanDEM-X system parameters.

products for the commercial sector, as well as multimode data for the scientific community.

- The TanDEM-X mission (TDM) [80], with the primary goal of generating a consistent global DEM product of unprecedented accuracy (2 m relative height accuracy and 12 m horizontal resolution [39]). Additionally, it is at the center of a multitude of scientific applications, such as glacier height change estimation, monitoring of volcanoes dynamics, land-cover mapping and forest parameter retrieval [39].

In particular, the TerraSAR-X satellite was already designed and built to be compliant with the subsequent TanDEM-X mission, mounting additional hardware for precise orbit determination, high RF signal stability and inter-satellite synchronization [106].

5.1.2 The TanDEM-X Mission

The TanDEM-X (*TerraSAR-X add-on for Digital Elevation Measurements*) mission was born as an extension to the original TerraSAR-X SAR mission, consisting in the close-formation flying of a second satellite of nearly identical specifications [80]. In particular, since October of 2010 the two TSX and TDX satellites co-orbit in the characteristic *Helix* formation [107][108]. It consists of both vertical and horizontal displacements that evolve in time, typically varying between 100 m and 500 m, without having the individual satellite orbits intersect each other. This unique orbit configuration allows for single-pass interferometry operation, overcoming the intrinsic limitations of repeat-pass strategies, namely the effects of temporal decorrelation and of atmospheric disturbances.

The large quantity of data takes acquired for the generation of the global DEM has been subsequently used for the generation of additional products, such as a global Forest/Non-Forest (FNF) map [109].

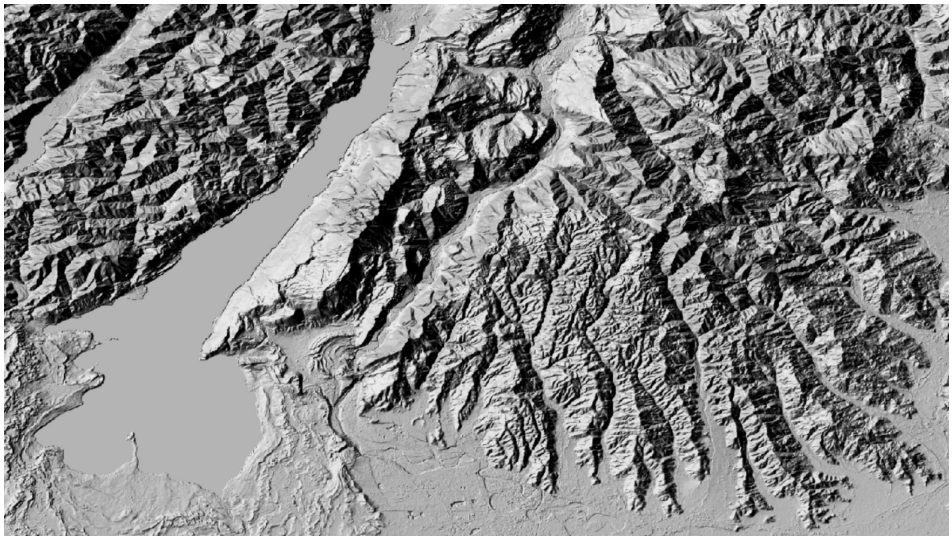


Figure 5.1: Hillshaded snippet from the TanDEM-X Global DEM product showing (from left to right) Lake Garda, Monte Baldo, the Val d'Adige and the western Venetian Prealps.

The definition of a refined post-processing editing solution [110], has first led to the generation of a global Water Body Layer (WBL) product [111], and subsequently to the release of the TanDEM-X Global Edited DEM product at 30 m, with filled gaps and flattened water surfaces. Furthermore, a new global acquisition campaign from 2017 to 2020, was used to generate a follow-up Global DEM and, crucially, for computing its difference with the original global DEM. This culminated in the creation of the global *DEM Change Map* (DCM) product [112].

5.1.3 TanDEM-X Processing Chain

In the following, an overview of the TanDEM-X bistatic product processing chain is given in the context of this work.

Processing starts with the ITP [113], which splits the acquired raw data resulting from a given data take into individual scenes with an extent of roughly 30×50 km in range and azimuth, respectively. Subsequently, the SAR focusing step and the co-registration step between the master and slave image pair results in the creation of the *co-registered single-look complex* (CoSSC) products.

Following this, the interferometric processing and raw DEM generation is performed using the *experimental TanDEM-X interferometric processor* (TAXI) [114], developed at the DLR Microwaves and Radar Institute (HR). In particular, TAXI allows for the flexible selection of different algorithms for the estimation of the interferometric phase and coherence. In the context of this work, the estimation is done by applying Φ-

Net [115], a state-of-the-art residual deep-learning denoising architecture, capable of preserving the spatial resolution (compared to the more commonly used boxcar multi-looking approach). Subsequently, the scene interferogram is further process to generate the raw acquisition DEM. As for the external DEM, the generated raw DEM is then used to compute the local incidence angle map of the interferometer. The apparent redundancy of this feature is justified by the fact that it offers to the user the choice between temporal consistency and better spatial resolution, or a higher level of processing quality than would otherwise be possible from a single-baseline product. With regards to the estimated total coherence, it is used to estimate the volumetric decorrelation coefficient γ_{vol} by inverting the relationship detailed in Eq. (2.39) and compensating it by the remaining decorrelation factors, estimated for the TanDEM-X system according to [52]. Additionally, for each scene a β^0 image is computed starting from the detected and absolutely calibrated amplitude image as recorded by the transmitting satellite only (i.e. the monostatic channel). Starting from the relationship in Eq. (2.19), the backscattering coefficient σ^0 is subsequently derived. Further terrain flattening and terrain correction steps (i.e. γ^0) are purposefully avoided over forested areas, as the external reference DEM (by default the TanDEM-X Global Edited DEM) is itself affected by the presence of vegetation (see Sec. 5.1.2), and might thus introduce errors into the data. Starting from the acquisition's metadata files and the external reference DEM, feature maps of the different elements of the SAR and InSAR acquisition geometry are generated. In particular, the satellite's orbit state vector is used to compute the incidence angle map θ_{inc} with respect to the reference ellipsoid, the local incidence angle map θ_{lia} with respect to the topographic slope angle, as well as a classification map of geometric distortions, such as shadow and layover [116][110]. Furthermore, the information about the interferometric acquisitions geometry, in the form of the height of ambiguity h_{amb}, is mapped starting from the annotated information on the satellites' position and following Eq. (2.34).

5.2 LVIS

The *Laser Vegetation Imaging Sensor* (LVIS) [53] is an airborne-mounted scanning laser altimeter, a LiDAR instrument used to map the topography of a scene. Designed and developed by NASA, its intended goal is the measurement of forest heights and structures. Each individual measurement consists of a short pulsed laser light that is used to time the distance between the airborne platform and the target. The sensor

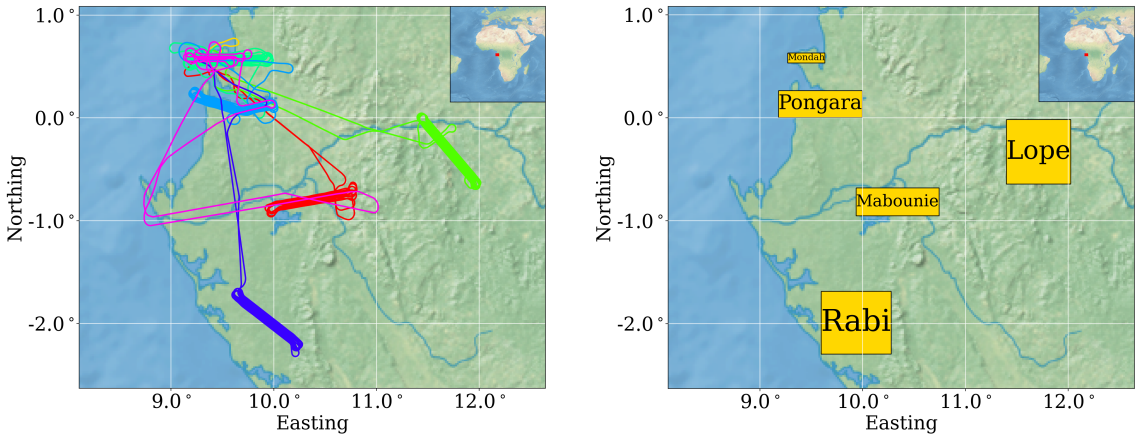
nominally operates at elevations of up to 10 km, which corresponds to individual on-ground measurements covering footprints of up to 25 m. A set of galvanometer-driven scan mirrors is used to steer the instrument, that allows it to have a potential field of view of 7°, corresponding to a maximum 1 km-wide swath.

Each transmitted signal is spatially and temporally Gaussian shaped, with a duration of less than 10 ns, for a total transmission power of 5 mJ. The LVIS instrument subsequently detects and digitally collects the entire full-waveform signal. The sampling begins prior to the expected target location, to allow for the estimation of noise statistics.

The collected data is co-registered with the airplane trajectory, and corrected for atmospheric effects, laser pulse timing errors and the airplane attitude [117]. A detection algorithm is then applied to extract the topographical and vegetation elevations from the received signal. Fig. 2.14 shows an example of a received waveform (a) and the correspondingly extracted elevation statistics (b). Typically, a binary threshold based on the estimated noise standard deviation is used to detect the location of the reflected signal. The signal is subsequently low-pass filtered and fitted with a mixture of Gaussians in order to approximate the statistics of each recorded peak [118]. Subsequently, the lower tail of the first peak is chosen as the beginning of the signal, while the upper tail of the last peak is chosen as its ending. Note that in the absence of vegetation the first and second peaks coincide. The elevation corresponding to the mode of the lowest peak is chosen as the representative ground elevation point, whereas for the vegetation height it is common to compute the RH metrics. For this, the signal selected by the aforementioned limits is corrected for the noise floor, by subtracting the average noise offset. The clean signal is then cumulatively added up, normalized and finally scaled by 100. Note that the smallest RH values correspond to negative heights, as these are quantified with respect to the reference ground elevation, which, as previously stated, coincides with the mode of the last signal peak.

5.2.1 AfriSAR Campaign

A relevant LVIS dataset for this work was acquired during the NASA/ESA AfriSAR campaign [18]. Carried out between 2015 and 2016, the AfriSAR campaign had the goal of simultaneously collecting SAR and LiDAR measurements in support of the NASA *Global Ecosystem Dynamics Initiative* (GEDI) [20], the *NASA-ISRO Synthetic Aperture Radar* (NISAR) [119] and the ESA *BIOMASS* missions [92]. All measurements were acquired over the West-Central African country of Gabon, using NASA's UAVSAR



(a) (b)

Figure 5.2: Map of LVIS's platform flight trajectories (a) and of the corresponding ROIs (b) for the NASA/ESA 2016 AfriSAR campaign.

(L-band SAR) and LVIS (LiDAR) airborne instruments, with the aim of developing processing algorithms for the generation of forest height and above-ground biomass density estimates.

The LiDAR measurements are of particular interest for this work, as they constitute a large collection of precise forest height estimates, which are suitable as reference for the proposed regression framework. The LVIS measurements cover five heterogeneous equatorial areas [73], as depicted in Fig. 5.2:

- The Lopé National Park, consisting of a mixture of seasonal tropical forest and Savannah, both affected by a distinct separation between wet and dry seasons. The area is characterized by strong topography, representing the highest elevation among all the considered regions of interest.

- The Mondah forest, which represents a small protected coastal site, partially flooded and characterized by the presence of both Mangroves and tropical hard-wood forests.

- The Mabounié site, which is a predominantly forested areas, with localized sites mostly affected by anthropogenic degradation.

- The Pongara National Park, located on the southern side of the Gabon River and characterized by the presence of seasonally flooded forests, as well as very tall mangroves stands and some grassy Savannah.

- The Rabi site, characterized by the presence of an onshore oil-drilling location, is largely covered by dense rainforest.

The individual laser shots illuminate a nominal footprint of 18 m of diameter. The associated vertical energy profiles were used to derive multiple forest structure proxy parameters, including the forest height estimates, expressed in terms of relative height. These products are also made available in the form of geomaps, which aggregate, interpolate and rasterize the discrete samples to generate a dense representation with a ground sampling distance of 25 m [120].

In this work, the rasterized RH99 is used as reference tree height metric since it still represents a good proxy for the top of the canopy, while reducing the effects of strong outliers. The rasterized digital terrain model (DTM) is also used as reference "real" topographic information below the dense canopy.

5.3 GEDI

The *Global Ecosystem Dynamics Investigation* (GEDI) is a spaceborne, full-waveform LiDAR instrument specifically designed for the monitoring of forest vertical structures [20]. Launched in late 2018, it is mounted on the International Space Station (ISS) and therefore operates at latitudes between 51.6°N and 51.6°S, covering all of the tropics, most of the temperate forests and, in limited amount, boreal forests as well.

The sensor consists in a multi-beam laser altimeter comprising 2 "full-power" lasers capable of producing 15 mJ/pulse and a third laser split through dedicated optics into two "coverage beams" of 4.5 mJ/pulse each. The 3 lasers are dithered to generate 8 on-ground beams with an across- and along-track separation of 600 m and 60 m, respectively, each illuminating a footprint of 25 m. This results in a nearly global, sparse sampling of the 3D structure of all covered vegetated areas, with no plans for multiple revisits of already sampled footprints within the original 2 year mission lifetime.

The laser power was chosen to detect the ground in presence of canopy covers of up to 98% and 95% with the full power beams and the coverage beams, respectively.

Different products levels are generated and distributed starting from the daily raw measurements. The level 1A (L1A) and level 1B (L1B) products correspond to the raw and to the geolocated waveforms, respectively [121]. The level 2A (L2A) products provide the ground elevation, top of the canopy and relative height metrics, while the level (L2B) products additionally provide the *canopy cover fraction* (CCF), the CCF profile, the *leaf area index* (LAI) and the LAI profile metrics [121]. Fig. 5.3 shows

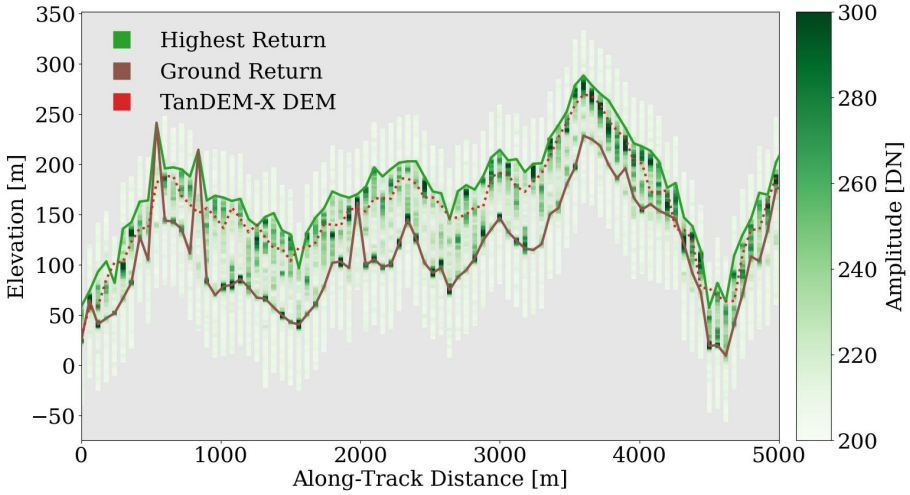


Figure 5.3: Example of a GEDI transect over the Redwood National Park, USA. The plot shows the individual full-waveform measurements from one of the full-power beams, and, superimposed, the corresponding vegetation metrics extracted from the L2A product. In particular, the bottom line (brown) shows the estimated ground return, the top line (green) shows the estimated return from the top of the canopy, while the center dotted line (red) shows (for reference) the location of the mean phase center measured by the TanDEM-X global DEM product.

an example of GEDI transect from one of the full-power beams, with superimposed vegetation metrics extracted from the corresponding L2A product. The level 3 (L3) products contain the 1 km gridded L2A-L2B metrics [122], while the level 4A (L4A) and level 4B (L4B) products contained the footprint-level and 1 km gridded above-ground biomass density (AGBD) metrics [123].

The L2A products are of particular relevance for this work. In their second product release (i.e. V2), they provide an improved mean 1σ horizontal geolocation error of 10.2 m. This has been significantly improved with respect to the 23.8 m of the first release [74][124].

In order to ensure the use of best-quality measurements only, the suggested pre-processing guidelines are applied [74]. First, the provided quality and degradation flags are selected to remove the laser shots that either do not meet quality criteria based on amplitude, energy, sensitivity and consistency, or are affected by orbit degradation periods. Furthermore, all diurnal samples are removed, as their lower SNR might cause errors during the waveform analysis. A larger shot sensitivity of 0.98 is selected since it is related to larger canopy penetration probability, corresponding to a larger probability of reaching the ground.

Sentinel-1 System Parameters		
Orbit	Elevation	693 km
	Inclination	Dusk-Dawn Sun-Synchronous (98.18°)
	Tube	50 m (RMS)
Antenna Size	Azimuth	12.3 m
	Elevation	0.821 m
Signal	Center Frequency	5.405 GHz
	Bandwidth	≤100 MHz
	PRF	1 KHz - 3 KHz
	Polarization	HH, HV, VV, VH (Dual-Pol)
Acquisition	Modes	Stripmap (SM), Interferometric Wide Swath (IW) Extra Wide Swath (EW), Wave (WV)
	Look Angle	20° - 46°

Table 5.2: Summary of the main system parameters for the ESA Sentinel-1 constellation.

Modes	**Range Resolution** [m]	**Azimuth Resolution** [m]	**Swath Width** [km]	**Product Level** L0 (Raw)	L1 (SLC)	L1 (GRD)	L2 (OCN)
SM	[1.7;3.6]	[4.3;4.9]	80	X	X	X	X
IW	[2.7;3.5]	22	250	X	X	X	X
EW	[7.9;15]	43	400	X	X	X	X
WV	[2.0;3.1]	4.8	20		X		X

Table 5.3: Acquisition modes and products offered for the ESA Sentinel-1 constellation.

5.4 Sentinel-1

Sentinel-1 (S1) is a spaceborne SAR EO mission and the first constellation launched by the European Space Agency's (ESA) for the European Copernicus program [22]. At its peak the constellation consisted of two satellites: the first (Sentinel-1A) launched in April 2014 and the second (Sentinel-1B) launched in April 2016. Currently, only Sentinel-1A remains operational, after a critical hardware failure in December 2021 occurred on-board of Sentinel-1B [125]. This has accelerated the construction of a further sensor (Sentinel-1C), which is currently expected to be launched in late 2024.

The system specifications for the SAR sensors of the Sentinel-1 constellation are reported in Tab. 5.2. Both satellites are equipped with a C-band SAR imaging system, working at a central wavelength of ≈5.5 cm. When considering the full constellation of two satellites and a single orbit direction (i.e. either ascending or descending) the nominal interferometric revisit time of Sentinel-1 is of 6 days. In practice, this is reserved only for specific areas of interest (such as Europe), and it is otherwise reduced to a 12-day global revisit time. The system is capable of operating at different acquisition modes: stripmap (SM), interferometric wide swath (IW), extra wide swath (EW) and wave (WV) mode.

For each of the modes, different product processing levels are generated. The Level-0 (L0) products correspond to the quantized unfocused SAR raw data. The Level-1

(L1) products comprise the focused and geo-referenced SAR images, and are available as either single-look complex (SLC) products in slant-range projection, or as ground-range detected (GRD) products consisting of the ground-range projected signal amplitude only. Also Level-2 (L2) oceanography (OCN) products exist for wind, wave and currents applications. A summary of the available acquisition modes, specifications and related products is reported in Tab. 5.3.

All Sentinel-1 products are freely available for scientific and commercial use through the *Copernicus Sentinel Data Access* (CSDA) platform [126].

In the context of this thesis, the most relevant products consist in the IW L1 SLCs. The IW is the main operational acquisition mode over land. It consists in the interleaved acquisition of three sub-swaths using *Terrain Observation with Progressive Scans SAR* (TOPSAR) [127], allowing for illuminating a nominal swath-width of 250 km. The L1 SLC products provide the full-resolution (i.e. single-look) complex samples (i.e. containing both amplitude and phase) images, one for each of the three individual sub-swaths and for each recorded polarization channel (VV and VH).

Using the DLR-developed *TAXI* SAR and InSAR processor [114], the IW SLC product is further processed to generate the backscatter map in terms of σ_0 and the local incidence angle map θ_{lia}. By ingesting a short time-series of Sentinel-1 acquisitions it is also possible to use TAXI to generate a repeat-pass InSAR stack and the corresponding coherence matrices, where each map results from an image pair separated by a temporal baseline corresponding to an integer multiple of the revisit time. Due to the multi-day revisit time and the small wavelength of the considered C-Band signal, temporal decorrelation becomes a major source of noise which significantly impacts the data usage. This makes the use of InSAR features from the Sentinel-1 mission over vegetated areas extremely challenging or unfeasible. As an example, Fig. 5.4 shows three coherence maps over a forested area along the coast of Gabon for Sentinel-1 repeat-pass acquisitions at different temporal baselines.

5.5 Sentinel-2

Sentinel-2 (S2) is a multi-spectral EO mission of the Copernicus program [24], consisting of two twin satellites, launched in June 2015 (Sentinel-2A) and March 2017 (Sentinel-2B), respectively. The two platforms follow a Sun-synchronous orbit at a 786 km mean altitude and with a 180° phase separation, allowing for a global revisit time of 10 days for one satellite and of 5 days when considering the full constellation.

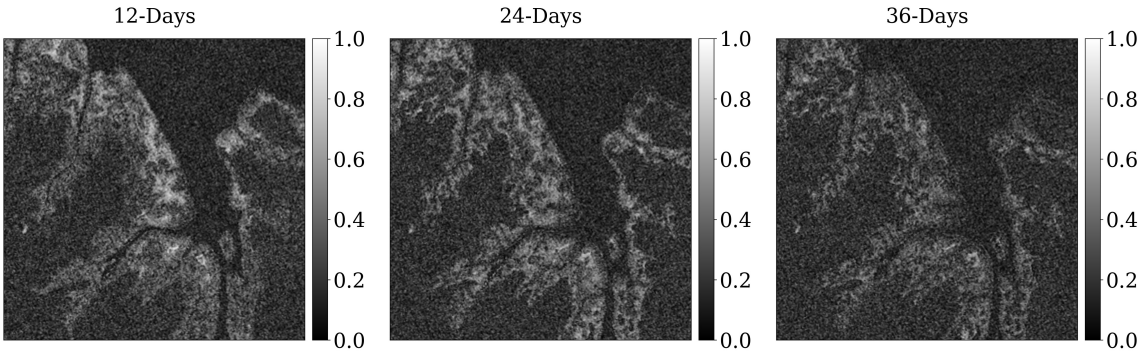


Figure 5.4: Example of change in the Sentinel-1 repeat-pass coherence for increasingly larger revisit-times of (a) 12 days, (b) 24 days and (c) 36 days. The imaged location corresponds to a small region of tropical forest within the Pongara National Park in Gabon, Africa. It is possible to observe the reduction in coherence over time, especially for the mangrove forest along the northern coastline. The primary forest located more inland displays less change, confirming that at 12-days the dense forest is already completely decorrelated.

The satellites mount a single *multi-spectral instrument* (MSI) acquiring a total of 13 bands covering the visible, near infrared and short-wave infrared spectra, with varying spectral and spatial resolutions. Of particular interest is the presence of the Red Edge bands (i.e. B05, B06 and B07), specifically designed to capture the characteristic rapid transition in reflectance from the red to the near infrared spectrum of vegetation.

The Sentinel-2 mission comprises the generation of freely available, UTM/WGS84 projected, ortho-image products covering individual tiles of $110 \times 110\,\mathrm{km}^2$. As for the Sentinel-1 mission, also Sentinel-2 products are freely available for commercial and scientific purposes [126].

The products are available at two processing levels: Level-1C (L1C) *top-of-atmosphere* (TOA) reflectance products; Level-2A (L2A) *bottom-of-atmosphere* (BOA) surface reflectance analysis-ready data (ARD). In particular, the operational L2A products are obtained from the L1C ones using ESA's *Sen2Cor* atmospheric correction tool [128], which since December 2018 has been integrated into the ground segment [129]. Not all L1C products have been converted to L2A yet, but they can be processed by the user using Sen2Cor as a plugin for the *Sentinel Application Platform* (SNAP) [130] or as a standalone tool [131]. Alternative third-party atmospheric correction tools, such as DLR's *Python-Based Atmospheric Correction* (PACO) [132], also exist but are out of the scope of this work.

Finally, a crucial processing step for optical data is the automatic detection and masking of clouds, as the partial or complete occlusion renders the surface reflectance

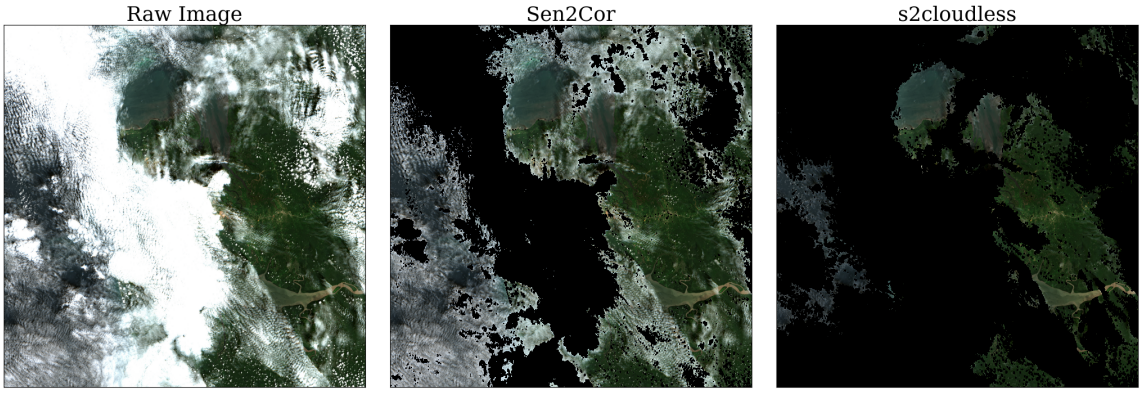


Figure 5.5: Example of a Sentinel-2 acquisition showing a comparison between (a) the raw RGB composite, (b) the *Sen2Cor*-masked result and (c) the *s2cloudless*-masked result.

information unrecoverable. ESA's Sen2Cor toolbox automatically generates a cloud probability mask during the atmospheric correction process, which can be thresholded to generate a binary classification map. Unfortunately, the performance of the resulting product suffers from a high false negatives rate, leaving a large number of cloudy pixels undetected [133][134]. Alternative third-party solutions exist, such as the CNES/CESBIO developed *MAJA* processor [135], the DLR developed *ATCOR* and *PACO* processors [132], or the USGS developed *Fmask* processor [136]. In this work the *s2cloudless* tool is used [137][134], a light single-scene Gradient Boosting Decision Tree (GBDT)-based solution developed by Sinergize [138]. As an example, Fig. 5.5 shows a visual performance comparison between the Sen2Cor-derived cloud mask and the s2cloudless-derived one.

5.6 ESA WorldCover Map

The *ESA WorldCover map* is a 10 m resolution global land-cover product. It has been computed using data from both the ESA Sentinel-1 and Sentinel-2 missions and it is freely accessible [139]. The WorldCover product classifies the scene into the following 11 land cover types: Tree cover, Shrubland, Grassland, Cropland, Built-up, Bare or Sparse vegetation, Snow or Ice, Permanent water bodies, herbaceous wetland, Mangroves, Moss or Lichen.

Two versions of the product exist:

1. The *WorldCover 2020 v100*, which was the first release of the product. It was generated by using Sentinel-1 and Sentinel-2 acquisitions from 2020, achieving an

average accuracy of 74.4%.

2. The *WorldCover 2021 v200*, which is the subsequent improved release. It relies on acquisitions from 2021, reaching an overall classification accuracy of 76.7%.

In this work only the latter product is considered, and it is used as a pre-processing step to mask out canopy height measurements acquired over undesirable land cover type, such as built-up areas or water-bodies.

5.7 The Dataset Pre-processing Pipeline

A significant part of the PhD activity has been spent developing a flexible EO data pre-processing framework, which matches up different data sources, brings them into a common geographic information systems (GIS), builds the necessary data structures for deep learning and stores them for efficient runtime retrieval. The resulting software collection, called *DataSet Creation Tool* (DSCT), has been fully written in python, and is based on the use of different open-source processing and GIS packages, including *Numpy*, *GDAL* and *GeoPandas*, while a significant subset of the heavier processing algorithms have been parallelized using the python *multiprocessing*-package. This allows the entire script to manage multiple hundreds of EO products (i.e. terabytes worth of data) in few hours, enabling the user to rapidly prototype multiple experiments within a single day.

The DSCT has been designed to include a large range of possibilities and detailed processing choices, such as to accommodate a diversified set of experiments. The details of each processing run are user-defined through a dedicated configuration-file, a sort of fingerprint of a particular dataset.

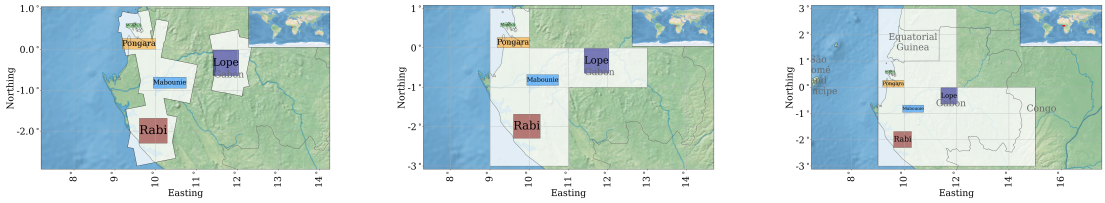
Finally, the DSCT is structured as independent modules that can be combined in order to create a fully automatic processing pipeline, which is briefly presented and discussed in the following sections.

5.7.1 Aggregation

All input EO datasets are typically characterized by different extensions, projections and spatial resolutions. The objective of the *Aggregation* module is then:

1. For each source, to parse the GIS information of products.

2. To determine a set of nominal regions of interest (ROI).



(a) (b) (c)

Figure 5.6: Examples of intersections between the ROIs (in colors) and (a) TanDEM-X footprints, (b) Global TanDEM-X Edited DEM product and (c) WorldCover Map products (in white).

3. To associate each product to all of the intersecting ROIs.

Concretely, the first step iteratively runs through each data-source (e.g. TanDEM-X, Sentinel-2, etc.), parsing the available products for relevant metadata, such as their geographic footprint, their projection system and the acquisition date, and stores this information into geo-referenced data-frame objects.

A user-defined flag subsequently determines whether the ROIs should be inferred from the one of the sources (e.g. the ground-truth dataset), or if these should be computed to accommodate the retrieved products. The choice is also connected to the definition of the output projection system (e.g. EPSG:4326, UTM, etc.), which can be similarly inferred from a reference dataset or be user-defined. Finally, the retrieved ROIs can be split into an arbitrary number of sub-regions.

Once the ROIs are defined, the previously computed data-frames are conveniently used to intersect the products of each source with each of the ROIs, ultimately resulting in a new data-frame which contains for each ROI an entry for each source detailing the list of intersecting products. This object represents the entry point for the next module. Fig. 5.6 shows an example of the intersection between the ROIs and different products.

5.7.2 Common Grid Interpolation

The previously aggregated products still use different projections and resolutions. Accordingly, the objective of the *Common Grid Interpolation* (CGI) module is to process each product as follows:

1. Load the product into memory.
2. Classify potential invalids.
3. Compute auxiliary features.

ROI-ID	Reference	TanDEM-X	Edited DEM	WCM
LOPE1	lvis2_lope_2016022020160308_l3a1_RH99_e32732_v0301_r025m.tif	TDM1_SAR__COS_BIST_SM_S_SRA_20151214T050542_20151214T050550	TDM1_EDEM_10_S01E009_EDEM.tif	ESA_WorldCover_10m_2021_V200_S03E009_Map.tif
		TDM1_SAR__COS_BIST_SM_S_SRA_20151001T173852_20151001T173900		
		TDM1_SAR__COS_BIST_SM_S_SRA_20151001T173849_20151001T173853		

Alignment

ROI-ID	Reference	TanDEM-X	Edited DEM	WCM
LOPE1_0	lvis2_lope_2016022020160308_l3a1_RH99_e32732_v0301_r025m.tif	TDM1_SAR__COS_BIST_SM_S_SRA_20151214T050542_20151214T050550	TDM1_EDEM_10_S01E009_EDEM.tif	ESA_WorldCover_10m_2021_V200_S03E009_Map.tif
LOPE1_1	lvis2_lope_2016022020160308_l3a1_RH99_e32732_v0301_r025m.tif	TDM1_SAR__COS_BIST_SM_S_SRA_20151001T173852_20151001T173900	TDM1_EDEM_10_S01E009_EDEM.tif	ESA_WorldCover_10m_2021_V200_S03E009_Map.tif
LOPE1_2	lvis2_lope_2016022020160308_l3a1_RH99_e32732_v0301_r025m.tif	TDM1_SAR__COS_BIST_SM_S_SRA_20151001T173849_20151001T173853	TDM1_EDEM_10_S01E009_EDEM.tif	ESA_WorldCover_10m_2021_V200_S03E009_Map.tif

Figure 5.7: Example of an alignment input and output state. The input entry (top) consists of one of more products for each source. After inflation, a collection of unique entries is created (bottom), each defined by a single product for each source.

4. Cache the intermediate product.

5. Perform a CGI for each intersecting ROI:

   - Load the cached intermediate product.
   - Reproject and then resample each feature to match the ROI's reference grid.
   - Save the interpolated product with a unique ID.
   - Free the modified intermediate product from memory.

Note that the details of the implementation vary according to the considered source, with some requiring additional pre-processing steps. In practice, each source is processed sequentially, with each source-specific CGI pipeline being run concurrently on multiple CPU cores.

Finally, the resulting products are parsed an collected into a new data-frame, which serves as input for the next module.

5.7.3 Alignmnent

To each input ROI correspond multiple products for each of the considered sources. The objective of the *Alignment* module is to manage these multiple entries to create unique pairs composed of a single product for each source.

This step can be effectively implemented in different ways:

- Inflation: simply consists in taking each product as is, each building a unique dataset.
- Temporal composite: build time-series' and apply a merging criteria, such as the mean or the median, to obtain a single product; this can be used for optical images to mitigate the effects of clouds, or for SAR images to perform temporal despeckling.
- Keep Best: keeps only the products that performs the best on a certain benchmark criteria, such as the number of invalids.

Alternatives of these algorithms also exists, which for example are applied to monthly sub-groups instead of the entire product stack. It is also pointed out that it is common for EO products to be the result of the same data-take, but subdivided into smaller products for better management of the data. For this reason, a same-day mosaicking pre-processing step is commonly selected in order to reduce the amount of products and to fill-in gaps.

The result of this module is an inflated version of the input data-frame, were each entry now corresponds to a unique set of products, one for each data source. An example of this operation can be viewed in Fig. 5.7. The described data-frame serves as input for the final module.

5.7.4 Sampling

Now that all products have been aggregated, interpolated to a common grid and properly aligned, the last step consists in converting and storing the product features into a format which is suitable for the deep learning task. In particular, it is convenient to distinguish between the *datacube* and *patch* formats, the first consisting of the stack of features covering the entire ROI, while the latter being a collection of equally sized crops of the former used for the training and validation of a deep learning model.

In the proposed implementation only datacubes are physically computed and stored, whereas a *memory-map* (mmap) file is created and used to directly read the patch from

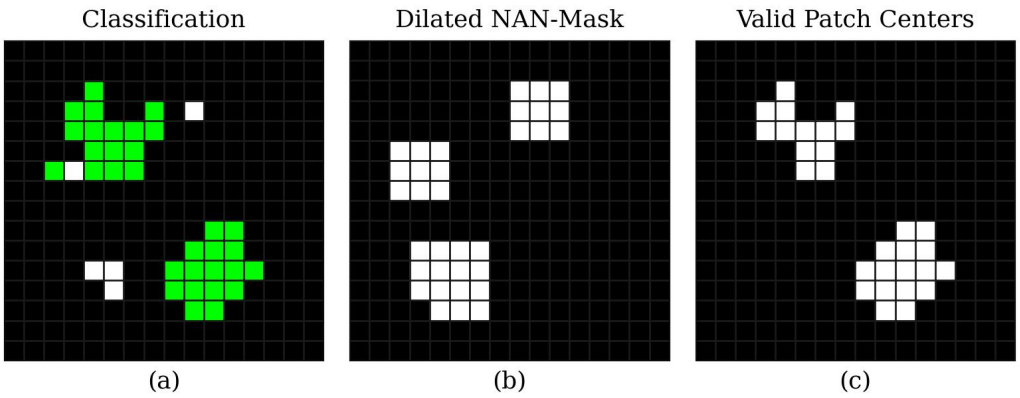


Figure 5.8: Example of the sampling process. (a) represents the input classification map highlighting the NAN (white) and vegetation (green) pixels. (b) represents the patch size-dilated NAN mask. (c) represents the final patch-center sampling mask.

the corresponding datacube without the need for loading the entire datacube first. This is primarily done to save disk resources, as storing the patch samples in a dedicated file would effectively correspond to storing redundant copies of the same data. One should note, that any overhead caused by the mmap-process is effectively compensated for at runtime by concurrently requesting multiple read operations.

The process of generation of the datacube is straightforward, as it involves looping through each ROI and subsequently each source, reading the associated features, stacking them into a datacube and storing it as a dedicated file.

The process for the patch sampling is somewhat more complex. Given a nominal patch-size, the goal is to determine all patch centers that fulfill a series of input and output criteria, such as the partial or total absence of *Not-A-Number* (NANs) values, or being labelled as forested pixels.

The fulfillment of these conditions is determined using the classification maps previously computed in the CGI step, in combination with different morphological filters. An example of such an operation can be seen in Fig. 5.8.

Finally, the resulting location of the computed valid patch-centers is stored in a dedicated file.

Chapter 6

A Novel Deep Learning Framework for Canopy Height Regression

The motivation behind this thesis is to simultaneously generate accurate, high resolution and time-tagged predictions from a single satellite imagery input. These requirements are essential in order to allow for detailed and timely monitoring of forest dynamics, at a low cost and at large scale. As detailed in Ch. 4, these objectives are arguably not yet simultaneously met by any method proposed in the literature, as compromises on accuracy, geometrical or temporal resolution are made to arbitrarily meet a subset of the listed criteria.

In order to accomplish these goals, this work proposes the definition of a deep learning-based approach, which relies on the extremely effective modelling capacity of neural networks (involving the optimization of millions of parameters) for ensuring, on the one hand, the possibility of properly capturing the complex relationship between the forest properties and the satellite data, and, on the other hand, the capability to adapt the model to a large variety of environments and boundary conditions. Moreover, if compared to traditional shallow machine learning approaches, the performance is not limited by the quality of handcrafted features, as these are automatically acquired and optimized from the input data to best solve the inversion problem. Regarding the design of the deep neural model, particular emphasis is given to understanding the regression task and the dataset on a regional scale, following a reasoning driven by explainable AI principles.

Furthermore, currently no study exists that thoroughly investigates the use of TanDEM-X single-pass InSAR products in combination with a purely data-driven deep learning-based modelling scheme for forest height estimation. While physical-based in-

vestigations have already established the potential of X-band interferometric data for forest height inversion [28], the attained results remain modest, especially when considering tropical areas. This work addresses these shortcomings by intentionally investigating the modelling problem within the challenging context offered by the Gabonese tropical environments.

In summary, the chapter focuses on the following aspects and it is structured accordingly:

- The development of proper training and testing strategies, based on considerations about the dataset composition and the working principles of the deep learning architecture.
- The design of the deep learning model, using extensive hyperparameter tuning in order to determine both the underlying architecture of the model and its dimensions, as well as regularization strategies for achieving peak performance.
- The performance evaluation, consisting of detailed performance analyses for the individual study areas and for acquisitions of particular interest.
- A feature importance analysis, aiming at highlighting the impact of the individual input features on the prediction performance and understanding their role.
- A spatial-transfer analysis, investigating the generalization performance of the framework.

Most of the considerations and results shown in this chapter are based on the work published in [140]. Ultimately, this work sets the basis for the further development of the framework for country-scale applications, as it will be later presented in Ch. 7.

6.1 Training, Validation and Testing Strategies

The following section presents the training, validation and testing strategies used for all the experiments discussed throughout this chapter. These also represent the basis for the activities in the subsequent chapters, therefore any potential modification will be presented accordingly.

6.1.1 Dataset Composition

Starting from the dataset, all of the experiments presented in this chapter are performed using the ALS-derived 25 m gridded height products from the 2016 NASA/ESA

AfriSAR campaign (see Sec. 5.2.1). In particular, the RH99 products are selected to represent the reference top of the canopy elevation, while the DTM measurements are used for performance analyses.

The operational acquisition mode for the TanDEM-X mission consists in single-pol (HH), bistatic stripmap acquisitions. A conservative assumption of stationarity is introduced, by which no significant changes have occurred in the scene within a temporal interval of $\pm$ 1 year around the reference data acquisition. Consequently, available TanDEM-X stripmap products, acquired in 2015 and 2016, are selected and preprocessed as described in Sec. 5.1.3, to obtain the following input feature set:

- Backscattering coefficient σ^0_{HH}, in HH polarization.
- Local incidence angle θ_{lia}.
- Total interferometric coherence γ_{tot}, estimated using Φ-Net [115].
- Volume decorrelation factor estimate γ_{vol}.
- Height of ambiguity h_{amb}.
- Global TanDEM-X edited DEM.

Note that θ_{lia} and h_{amb} are both required to correctly inform the model since γ_{vol}, γ_{tot} and σ^0 all depend on the acquisition geometry, as described in Sec. 2.2 and Sec. 2.2.3. Similarly, the DEM is added to inform the model of the local terrain topography, which has an impact on the SAR side-looking geometry. Furthermore, over forested areas the TanDEM-X derived DEM effectively measures the height corresponding to mean phase center (see Sec. 5.1.2), which is itself dependant on the properties of the underlying canopy and might thus provide additional information about the inversion problem.

The collection of input features is presented in Fig. 6.1.

6.1.2 Geographic Subsetting and Dataset Processing

The resulting geographic intersection between the different products described in Sec. 6.1.1 restricts the analysis to the four study areas of Mondah, Mabonié, Pongara and Rabi.

All of the four areas are split into three equally sized sub-regions according to their geographic extent. Each resulting sub-area is subsequently assigned to either the training set, the validation set or the test set. This is done to build three geographically separated datasets, such as to avoid *data leakage*. The resulting geographical subsetting can be seen in Fig. 6.2.

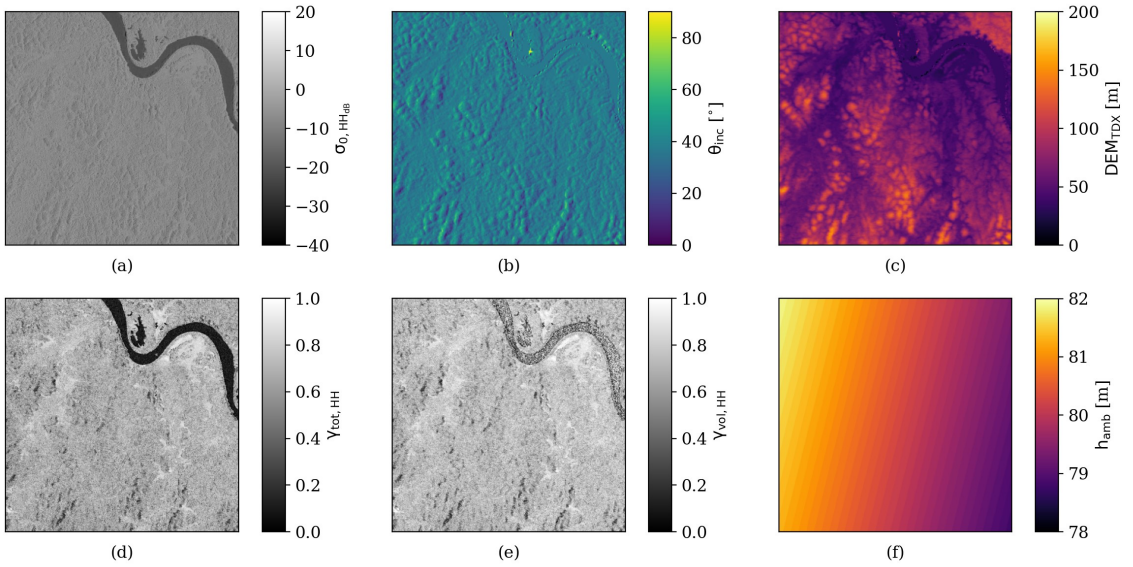


Figure 6.1: Input feature collection example, consisting of: (a) backscattering coefficient σ^0_{HH}, (b) local incidence angle θ_{lia}, (c) TanDEM-X global edited DEM, (d) interferometric coherence estimate γ_{tot}, (e) volumetric decorrelation coefficient estimate γ_{vol}, (f) height of ambiguity h_{amb}.

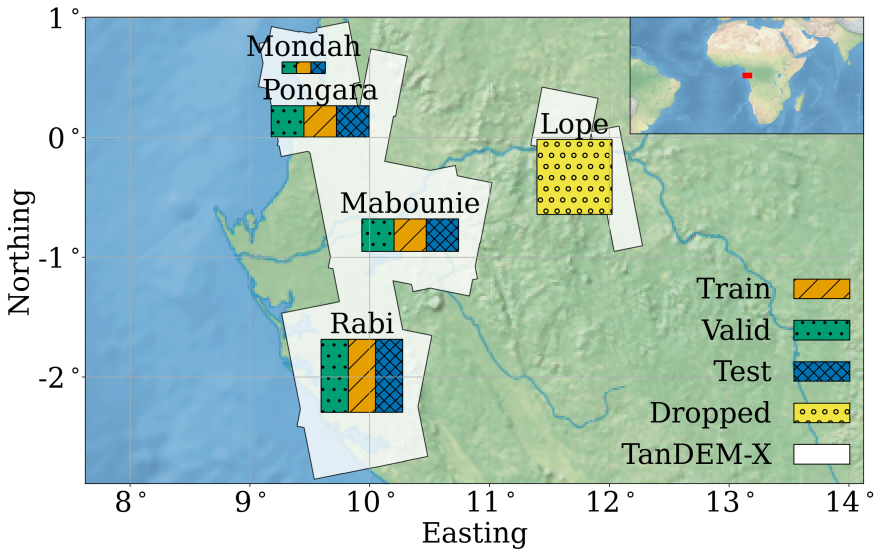


Figure 6.2: Geographical subsetting of the AfriSAR 2016 campaign test site for the generation of the training (orange), validation (green) and testing (blue) data sets. Discarded areas are identified in yellow, while TanDEM-X acquisition footprints are depicted in white.

According to the dataset processing pipeline described in Sec. 5.7, all input features are reprojected, resampled and cropped to match reference grid of the output

features. The resulting input and output datacubes are subsequently used to determine the location for the training and validation patches, corresponding to areas which are simultaneously covered by the reference LVIS measurements and flagged as vegetation areas by the WorldCover Map (see Sec. 5.6). This is done in order to avoid working on classes of pixels which are not strictly related to the regression task, the reasoning being that their inclusion would add complexity in the design and validation of the dataset.

The resulting location indices are stored and used to efficiently load the corresponding patches at runtime.

6.1.3 Model Training and Validation

The training of the model is performed using mini-batches, composed of 250 patches each. At each iteration, a new batch is built from randomly sampled patches and used to perform a forward pass in which the corresponding canopy height values are estimated. At the end of each iteration, the quality of the prediction, i.e. its distance to the reference value, is assessed by computing an MSE-based loss function. The combination of the backpropagation algorithm and the Adam optimizer are then used to update the weights accordingly. At the end of each epoch (defined as a group of 500 batch-iterations) the state of the model is evaluated on the validation set using the same loss function. The relative co-evolution of the training and validation losses is used to determine the completion of the training phase. In particular, a stagnation in the validation loss for 30 consecutive epochs leads to the reduction in the learning rate by a factor 10, while after 35 epochs the training it is stopped altogether.

At the end of the training phase, the weights corresponding to the epoch with the lowest validation loss are selected for the subsequent testing phase.

6.1.4 Model Testing

In order to evaluate the performance of the resulting model, an inference run is carried out for each input acquisition covering the designated test sites. As a pre-processing step, the edges of each scene are symmetrically padded by half the training patch size, and removed again after the inference to preserve the original image size. As a final post-processing step, invalid areas are similarly dilated by half the patch size. These steps guarantee that that only pixels predicted under similar conditions as those seen during training are considered to be reliable.

A suite of performance metrics, consisting in ME, MAE, MAPE, RMSE, R^2 (defined in Sec. 3.2.5), is used to quantitatively evaluate the performance of the trained model.

Furthermore, a diverse set of histograms, scatterplot and boxplots is automatically created for all acquisitions and areas, as well as for the overall test set, to further analyze and understand the model performance.

6.2 The Proposed Deep Learning Model

This section discusses in detail the creation of the proposed regression model. The architecture is created from scratch by relying on both consolidated design principles and an extensive hyperparameter tuning approach.

Since the design of a neural architecture reflects the goal of its deployment, the architecture chosen for this work is first and foremost expected to handle structured multi-modal imagery of comparatively large spatial dimensions. Another of the listed goals is the generation of time-tagged estimates, which ideally do not require multiple acquisitions as input. The analysis of multi-temporal image sequences is therefore not of interest, which removes the need for recurrent neural networks. Fundamentally, this already limits the choice to either MLPs or CNNs.

As discussed in Sec. 3.4, MLP models do not represent a practical choice for handling imaging data, as the input dimensionalities are immutable and must be chosen during the design of the model instance.

Furthermore, under the reasonable assumption of a natural imaging process, neighbouring pixels in remotely sensed data are strongly correlated, while by design the connections of a dense network treat all inputs equally. In this context, CNNs provide a pragmatic alternative to MLP models, as they exploit the spatial autocorrelation of neighbouring image pixels by replacing the dense connections with smaller kernel functions that are shared across the entire image. This hard-coded inductive bias drastically reduces the complexity of the problem, allowing for smaller data requirements and nimbler architectures. These insights ultimately motivate the use of a CNN architecture for this work.

6.2.1 Hyperparameter Tuning

The first step for the creation of the regression framework is the definition of a deep neural model to be trained. An extremely common approach found in the literature consists in relying on *off-the-shelf* solutions and introducing small changes to increase the novelty of the proposed works, without extensive investigation on the impact of neither the choice of the backbone model nor of the introduced modifications [35][37][38].

In this work a different approach was selected, consisting in the creation of an architecture from scratch by means of hyperparameter tuning. The motivation behind this choice is twofold: firstly, this allows one to better understand the impact of every design element on the final regression task; secondly, it allows for converging towards an efficient architecture with minimal training data and computational requirements.

As seen in Ch. 3, neural networks are characterized by a set of hyperparameters, each defining different aspects of the architecture itself, such as the family of networks, the width and depth and the layer composition, and of the training strategy, such as the optimization algorithm, the learning rate, the use and strength of regularizers, the batch size or the number of training epochs.

In order to define these settings, the hyperparameters and their co-dependencies are investigated in small sets, each by means of exhaustive search to find suitable values.

All experiments are evaluated on the set-aside validation set, in order to ensure independent optimization from the test set and related experiments. For this, all available TanDEM-X stripmap images acquired in 2015 and 2016, and covering the four study areas of Mondah, Mabounié, Pongara and Rabi, are considered. The acquisitions are processed according to the description presented in Sec. 6.1. For this analysis, the patch-size is chosen to be equal to the receptive field of each tested model. In order to keep the training set size constant across all experiments, the patches are sampled according to the largest required patch-size and cropped according to the individual model specifications. For comparison sake, the loss function was computed only on the center pixel of each considered patch in order to exclude possible variations due to the differently sized patches and boundary effects.

The Model Architecture

The first set of the tweaking steps involves the definition of the basic network family and its dimensions (i.e. depth and width).

For this test two architecture types are considered: *plain CNNs* and *residual CNNs*. Both models types are set up to be as comparable as possible and can be thought as being composed of three separate stages: the *head* or *input* section, the *backbone* section and the *tail* or *output* section.

The purpose of the input section is to gradually increase the dimensionality of the features from that of the input to that of the hidden layers in the backbone. It consists of two 2D convolutional layers, i.e. linear layers, each using 1×1 kernels, and followed by a batch normalization layer and a ReLU activation function. Note that in the case

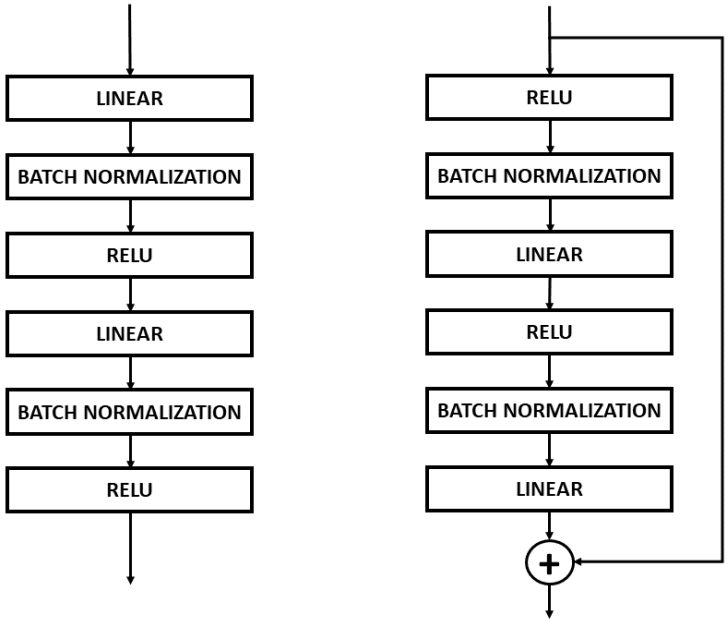


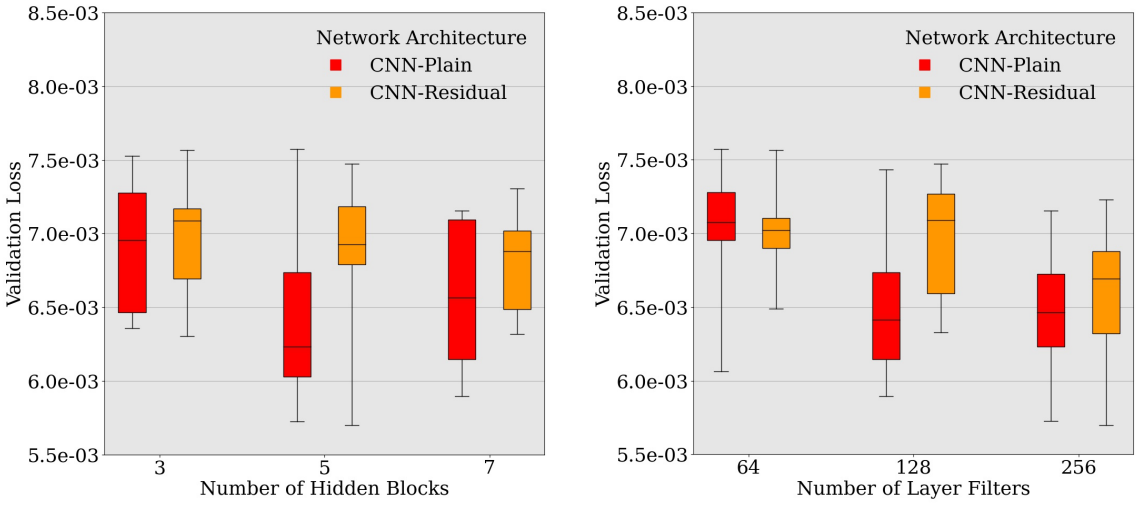
Figure 6.3: Comparison of the considered *plain* (left) and *residual* (right) building blocks for the model hyperparmeter tuning. It is possible to note that, while the number of convolutional layers is the same, the order of the layers differs. As explained in Sec. 3.4.2, this is due to the residual connections requiring the operation order to be changed with respect to the plain architecture.

of the residual CNNs the last ReLU is omitted in order to properly stage the residual blocks in the backbone, according to the explanation given in Sec. 3.4.2.

The backbone forms the core feature extraction module of the architecture. It can be split up into multiple identical blocks, each composed of two 2D-convolutional layers using 3×3 kernels, two batch normalization layers and two ReLU activation functions. The order of these layers differs for the two network families:

- For the plain CNN, as shown in Fig. 6.3 (a), each block consists of two repetitions of a linear layer, followed by the batch normalization layer and the ReLU activation function.

- For the residual CNN, as seen in Fig. 6.3 (b), each block is structured as a *pre-activation* block, with the layers in the block-subnetwork consisting of two repetitions of the a ReLU activation function, followed by the batch normalization layer and the linear layer. The subnetwork is bypassed by an identity mapping that gets added to the output of the subnetwork. As previously stated, in this case the last ReLU in the input section is purposefully omitted.



(a) (b)

Figure 6.4: The hyperparameter tuning results for different neural architectures. Each boxplot represents the choice of two hyperameters, and captures the performance variability associated with the remaining parameter options, as well as the intrinsic randomness due to the weight initialization and patch sampling processes. The black lines show the mean metric values, the boxes contain 50% of all cases, while the whiskers represent the best and worst performing permutations.

The tail consists of two 2D convolutional layers, each using 1×1 kernels, interleaved by a batch normalization layer and a ReLU activation function. No layers follow the second linear layer as it generates the regression estimate, effectively acting as a linear regression model.

Each architecture is tested for three different hidden block counts of 3, 5 and 7, and for three different numbers of hidden layer filters of 64, 128 and 256. This corresponds to a total of 9 possible permutations for each model type, for a combined total of 18. For each permutation, three independent trainings are performed, for a final total of 54 individual trainings that required roughly 7 days of GPU time to complete using a single NVIDIA A100 40GB GPU. This is done to guarantee meaningful results, as both the random initialization and batch shuffling processes inevitably lead to intra-run variance.

The results for the three considered variables are presented in Fig. 6.4. It is possible to observe that the plain CNN outperforms, on average and across all experiments, the residual CNN. For the plain CNN the best performance in achieved with 5 hidden blocks, each using 128 kernel functions, while for the residual CNN it is achieved with 7 blocks and 256 filter size. This implies that the plain CNN lies within the tested value

range, while the residual CNN might still profit from larger dimensions.

For the purpose of this work the nimbler plain CNN architecture is ultimately selected.

Regularization

Another investigated aspect is that of the model regularization. Three different techniques are considered for this purpose: dropout layers, $l1$-norm and $l2$-norm. For the dropout layer three different drop-rates are tested, 0.0 or no dropout, 0.25 as an intermediate drop-rate, and 0.5 as a large drop-rate. For each of the $l1$- and $l2$-norms the same λ weighting coefficient ranges are considered, the first being 0.0 or no regularization, the next being a light 10^{-4}, and the last one being a more aggressive 10^{-3}. In order to avoid further increasing the training times beyond reasonable levels, these tests are performed only on the CNN configuration that performed the best in the previous tests, i.e. the plain CNN. This results in a total of 27 permutations or 81 independent trainings, which took roughly 4 days of GPU time to run.

The results for the regularization experiments are presented in Fig. 6.5. In particular, it becomes immediately apparent from Fig. 6.5 (a) that dropout has (on average) a negative impact on the validation loss. Since batch normalization already regularizes the data, it is possible that the additional noise introduced by the dropout process is detrimental, resulting in worse performance.

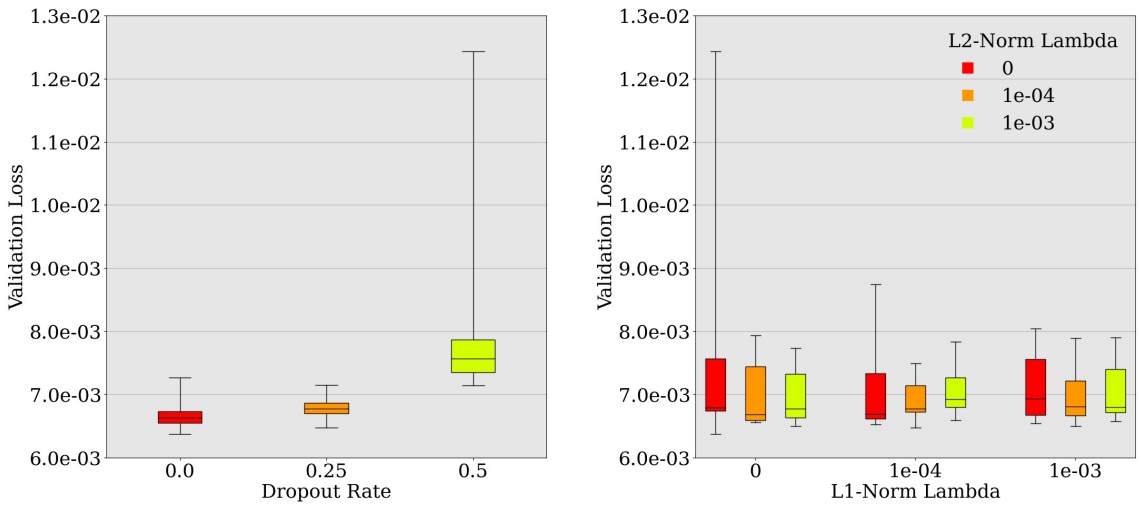
Fig. 6.5 (b) shows that both $l1$-norm and $l2$-norm can potentially improve the validation performance. The $l2$-norm alone (i.e. left group of boxplots) achieves the overall best absolute performance with $\lambda_{l2} = 10^{-4}$, while the $l1$-norm alone (i.e. red boxplots), obtains the second best absolute score with $\lambda_{l1} = 10^{-4}$.

In conclusion, dropout layers are not added to the model, while a $l2$-norm regularization term with a modest $\lambda_{l2} = 10^{-4}$ is added to the loss function during training.

Patch Size

A final set of experiments aims at determining the impact of patch size on the validation performance. The goal is to select the possibly smallest patch size in order to optimize the sampling rate of the reference dataset and also to reduce the computation times.

To this end, a single independent training is performed for a different patch size, obtained by cropping the same larger input patches to the desired output size. The results of this analysis are shown in Fig. 6.6. As can be seen in Fig. 6.6 (a), the spread in the validation loss itself is not necessarily conclusive. Fitting a polynomial to the



(a) (b)

Figure 6.5: The hyperparameter training tuning results for different regularization techniques. Each boxplot represents the choice of two hyperameters, and captures the performance variability associated with the remaining parameter options, as well as the intrinsic randomness due to the weight initialization and patch sampling processes.

data points shows a small positive trend fitting the expectation of smaller patch sizes leading to worse performance than larger ones.

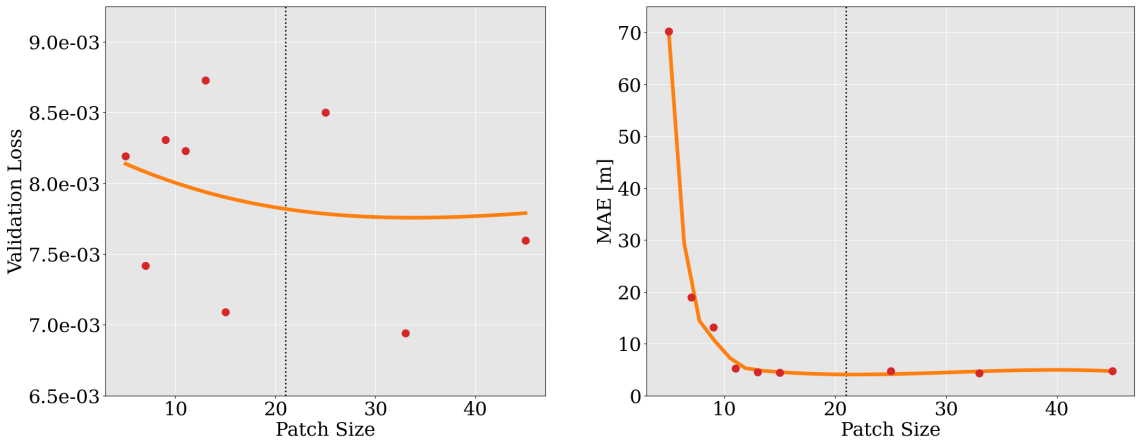
When looking instead at the MAE computed over the validation set (Fig. 6.6 (b)) a clearer trend becomes apparent. Patch sizes larger than 15×15 have no significant impact on performance, while progressively smaller patches exponentially reduce it, suggesting that the model is excessively starved of information. This is a clear result which ultimately leads to the choice of a 15×15 patch size for the selected CNN model.

Final Model

In summary, the final selected model is represented in Fig. 6.7 and it consists of:

- An entry block, with two 1×1 2D convolutional layers.
- Five hidden blocks, each consisting of two 3×3 2D convolutional layers, for a total of ten convolutional layers.
- An exit block, with two 1×1 2D convolutional layers.

All the 2D convolutional layers, but the last one, are followed by a batch normalization layer and a ReLU activation function. The last 2D convolutional layer generates the final output prediction, and thus effectively implements a linear regression operation.



(a) (b)

Figure 6.6: Analysis for the definition of the optimal patch size. The plots depict the loss (a) and MAE (b) computed on the validation set with respect to the patch size.

Consequently, the previous layers can be interpreted as generating the nonlinear feature-representation of the input, and thus form the main feature extractor stage of the model.

The final model has a total of $1.5 \cdot 10^6$ trainable parameters and is trained on minibatches of 15×15 pixel patches, each fully-covered by both the input and reference features. The performance of the model and the update of its weights for the k^{th}-batch is based on the following two-term loss function:

$$\mathcal{L}_k = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 + \lambda \cdot \sum_{j=1}^{m} w_j^2, \tag{6.1}$$

where $\hat{y}_i$ is the output estimate, y_i is its corresponding expected value, w_j are the trainable model weights, and $\lambda = 10^{-4}$ weights the impact of the $l2$-regularization term.

6.3 Performance Baseline

In order to best contextualize the proposed investigations, the initial analysis consists in the definition of a baseline scenario against which to measure all follow-up experiments.

For this purpose the full dataset is considered, consisting of the geographical subsets described in Sec. 6.1.2 and shown in Fig. 6.2, and including all available acquisitions and all six of the input features listed in Sec. 6.1.1.

The histograms of the resulting reference values for training, validation and testing

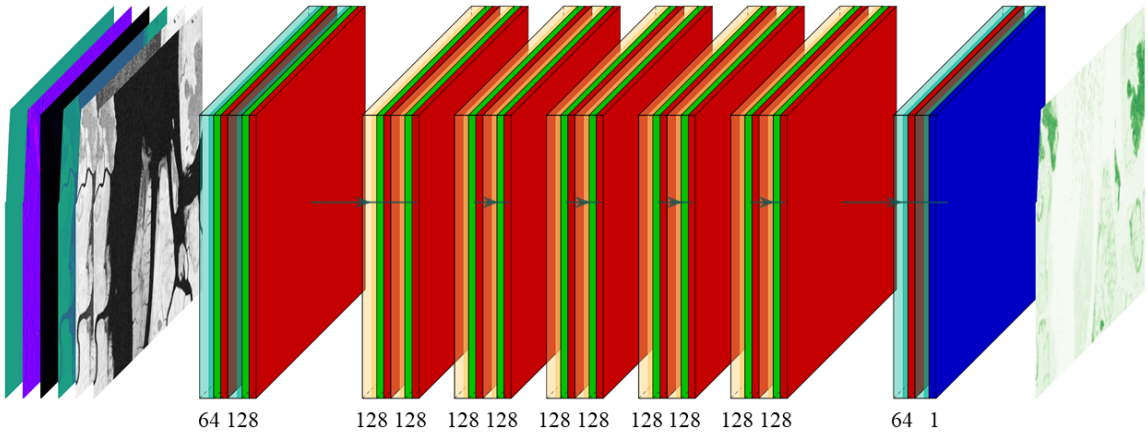


Figure 6.7: The proposed fully convolutional deep learning model. In blue and yellow the 2D-Convolutional layers with 1×1 and 3×3 kernels, respectively, where the subscript numbers indicate the number of kernel filters. In green the batch normalization layers. The red and blue layers represent the ReLU activation functions and the Linear activation function (i.e. bypass), respectively.

are compared in Fig. 6.8. Crucially, all three subsets result in similar bimodal distributions, with a peak corresponding to barren areas or of low vegetation (<10 m) and another peak corresponding to fully developed primary forest stands (30 m<CHM<40 m). This suggests that all three subsets should be equally representative of the regional diversities.

Finally, the impact of the TanDEM-X orbit direction on the performance has been considered, with ascending and descending orbits corresponding to data acquired in the morning and in the afternoon, respectively. Empirical evidence shows that training and testing with either ascending-only or descending-only acquisitions, or with mixed-orbit acquisitions paired with supporting information, such as orbit direction, acquisition time of day or no information, has no significant influence on the overall performance. Therefore, in order to guarantee the best possible temporal coverage, the following results are based on a similar ratio of both ascending and descending orbit acquisitions.

The baseline model is trained following the procedure detailed in Sec. 6.1.3. Fig. 6.9 (a) shows the impact of the random weight initialization and patch sampling on the RMSE performance across different independent training runs. Note that the different test sites show different degrees of variability, corresponding to RSME variations in the order of several decimeters. In an effort to counteract these effects of intra-run variance, for each of the following experiments, the validation step selects the best model out of five independent training runs.

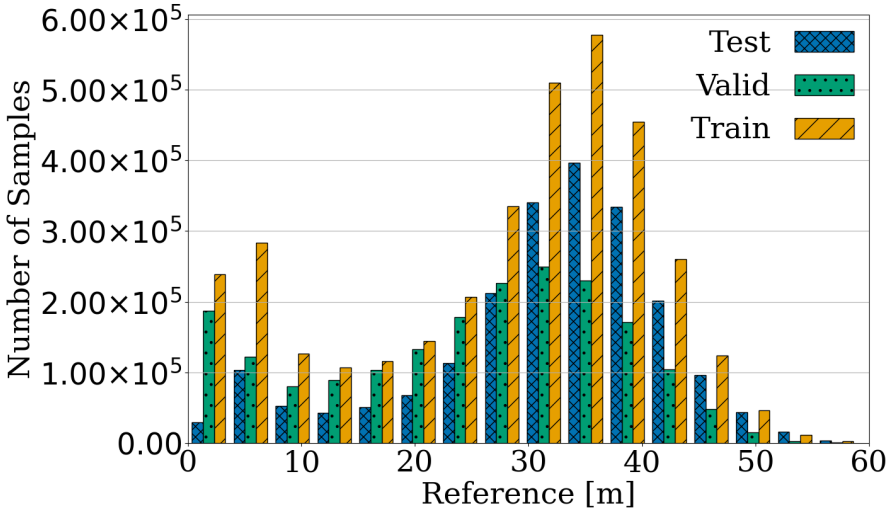


Figure 6.8: The reference CHM value distributions for the three training, validation and test subsets, as defined for the baseline scenario.

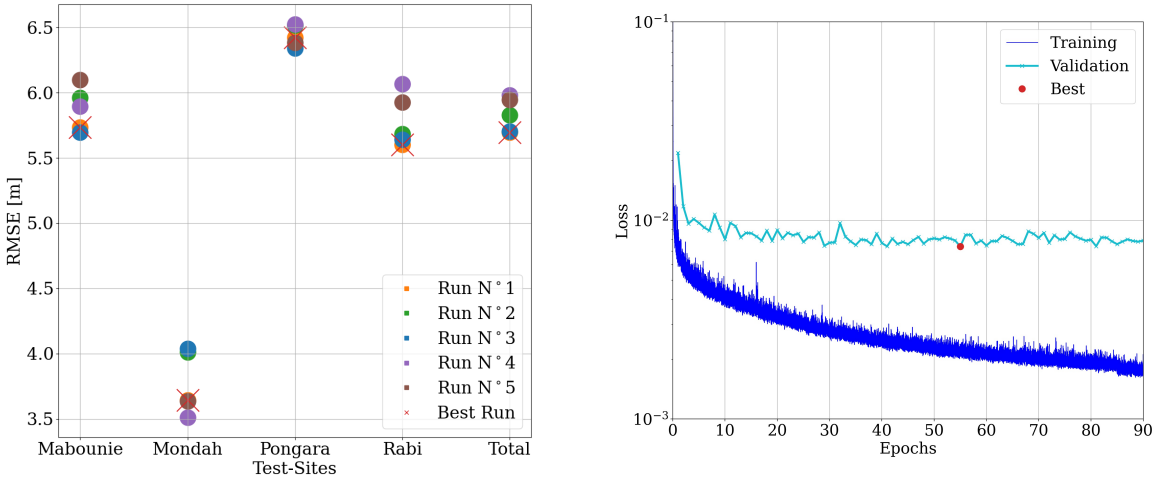
6.3.1 Overall Performance

The loss behaviour for the overall best model can be seen in Fig. 6.9 (b). The test results listed in Tab. 6.1, show the baseline achieving an overall bias of −1.46 m, a MAE of 4.20 m and an R^2 score of 0.73.

The scatterplot shown in Fig. 6.10 (a), highlights an overall good agreement between the prediction and the reference measurements across all test sites, with similar distributions for both the reference CHM and the model predictions. Notably, a small cluster of underestimated samples can be spotted, which will be better addressed later on in this section.

More in detail, Fig. 6.10 (b) shows the relationship between the estimation bias and the canopy heights for different subranges, suggesting that on average the estimation remains unbiased for CHM below 35 m, above which the model starts to increasingly underestimate taller trees. This behaviour might be linked to either input data saturation effects, often documented in the literature [141], or the limited representation in the training set. This aspect will be considered later on when analyzing the Pongara test site in Sec. 6.3.2.

Finally, Fig. 6.11 shows prediction examples for each of the four study areas for the current baseline scenario. Sample transects are plotted in order to compare the predicted tree height line with that of the reference. An overall good agreement for both tall and low vegetation heights, can be observed. As expected, the TanDEM-



(a) (b)

Figure 6.9: Training and validation results for the baseline model. (a) shows the intra-run validation variance obtained from five independent trainings, expressed in terms of RMSE for both the individual study areas and overall (Total). (b) shows the training and validation loss evolutions for the training run with the highest validation score.

X global edited DEM measurements, depicted for comparison purposes, can be seen falling within the canopy. The DEM struggles to match all changes in canopy height and maintains a non-linear relationship with respect to the tree height.

6.3.2 Individual Areas

Looking at the results for the individual test sites offers additional insight into the performance behaviour. The scatterplots for the individual areas are shown in Fig. 6.12, while the corresponding performance metrics can be found in Tab. 6.1.

Mondah The site is characterized by a patchy coverage of tall vegetation including mangrove forests, alternated by open areas without any coverage. The resulting large dynamic range is correspondingly met by a large coefficient of determination of 0.88. The large absence of vegetation, evidenced by the reference distribution in Fig. 6.12 (b), simultaneously leads to the lowest MAE of 2.49 m and the larges MAPE of 27.51% across all areas.

Pongara The site covering the Pongara National Park is characterized by an homogeneous tree height distribution, including a large sample of tall mangroves. Overall,

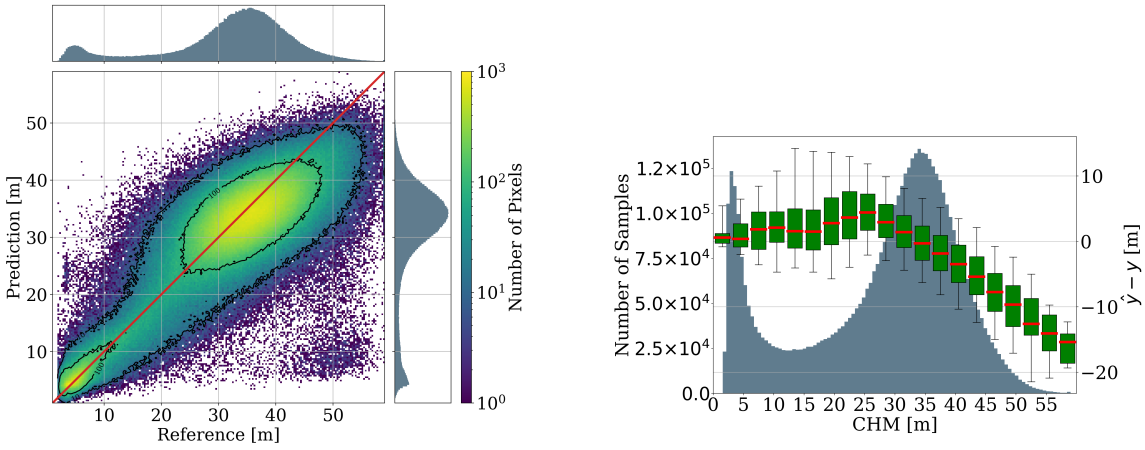
Table 6.1: Results for the baseline scenario presented in Sec. 6.3 and for all cross-validation permutations of the spatial transfer analysis. Each metric is evaluated for each test site individually and overall.

Performance Metrics					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	[·]
Baseline Scenario					
Mabounié	-0.75	4.41	14.22	5.73	0.46
Mondah	0.31	2.49	27.51	3.65	0.88
Pongara	-2.31	3.92	16.21	6.42	0.82
Rabi	-1.78	4.39	13.60	5.60	0.40
Overall	-1.46	4.20	15.06	5.69	0.73
Spatial Transfer Analysis					
Mabounié	0.18	4.41	15.00	5.68	0.47
Mondah	0.53	2.37	27.31	3.49	0.89
Pongara	-3.97	5.65	23.73	7.96	0.73
Rabi	-2.49	4.82	14.61	6.20	0.27
Overall	-1.87	4.70	17.06	6.28	0.68
RVoG ($\sigma_{ext} = 0, g = 0$)					
Mabounié	0.25	6.05	22.34	7.76	0.01
Mondah	4.99	6.31	89.50	8.20	0.42
Pongara	2.88	9.33	72.24	11.19	0.45
Rabi	-0.05	5.78	20.22	7.45	-0.05
Overall	0.84	6.49	33.87	8.33	0.43

the scatterplot in Fig. 6.12 (c) shows good agreement between the predicted and the expected canopy heights. Indeed, the performance is well balanced across all metrics, achieving a MAPE of only 16.21% and an R^2 of 0.82.

It is also possible to note that the previously observed cluster of outlier samples is located within Pongara, and corresponds to large values of canopy height being strongly underestimated. In order to geographically position these values, a manual threshold of $y < x - 20$ is used to highlight these samples. As can be seen from Fig. 6.13, these are concentrated in a single cluster corresponding to a gap in the TanDEM-X global edited DEM. Indeed, in the absence of TanDEM-X coverages, the automatic editing algorithm of the global DEM product falls back to an ALOS-derived DEM to fill in the gap, resulting in inconsistencies and mutual calibration problems between the two DEM sources. In this particular instance the ALOS DEM falls below the adjacent sea level and is thus raised and flattened by the editing software, producing a smooth topographic surface.

The manual masking of the region and re-evaluation of the performance metrics on the remaining valid areas a loss of 4686 test samples, corresponding to only 1.85% of the Pongara samples, and 0.32% of the overall available test samples. Indeed, the RMSE performance was especially negatively affected by the aforementioned problem,



(a) (b)

Figure 6.10: Visual representations of prediction performance for the baseline scenario. (a) scatterplot in logarithmic scale of the predicted and reference canopy height values over all four test sites in Gabon, Africa, together with the side distributions of both the reference and the prediction for the baseline scenario. (b) Training sample distribution (blue histogram in the background - left axis) compared to the ME, which represents the regression bias (red lines in the foreground - right axis). The boxplots display the regression bias spread (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles, respectively) across the CHM range.

and saw a significant improvement from 6.42 m to 4.83 m, making Pongara the site with the best RMS performance after Mondah. The overall performance shown in Tab. 6.2 shows the MAPE reduce to 14.91% and the R^2 rise to 0.76.

Mabounié and Rabi These sites are predominantly covered by tall forest stands, thus resulting in unimodal tree height distributions centered around 35 m of height. In this scenario, the absolute model uncertainty appears to increase, with performance for both Mabounié and Rabi scoring a below average MAE of 4.41 m and of 4.39 m, respectively. Similarly, the fitting of the observations yields a lower R^2 of 0.46 for Mabounié and of 0.40 for Rabi. Nevertheless, when looking at the relative performance, both sites perform above average, with a MAPE of 14.22% and of 13.60% for Mabounié and Rabi, respectively.

6.3.3 The Impact of Topography

The impact of the SAR side-looking acquisition geometry on the performance is assessed by evaluating the relationship between the prediction error and the local topographic slope. The latter is computed from the LVIS-derived DTM, in order to

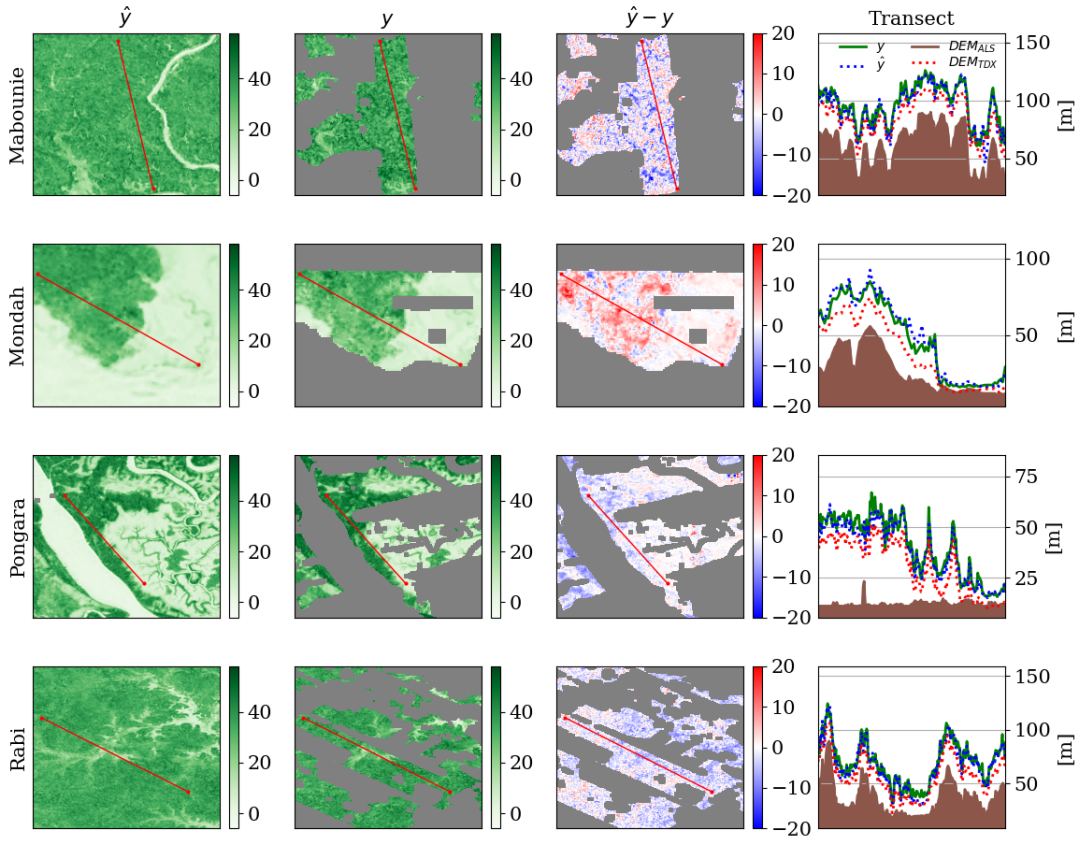
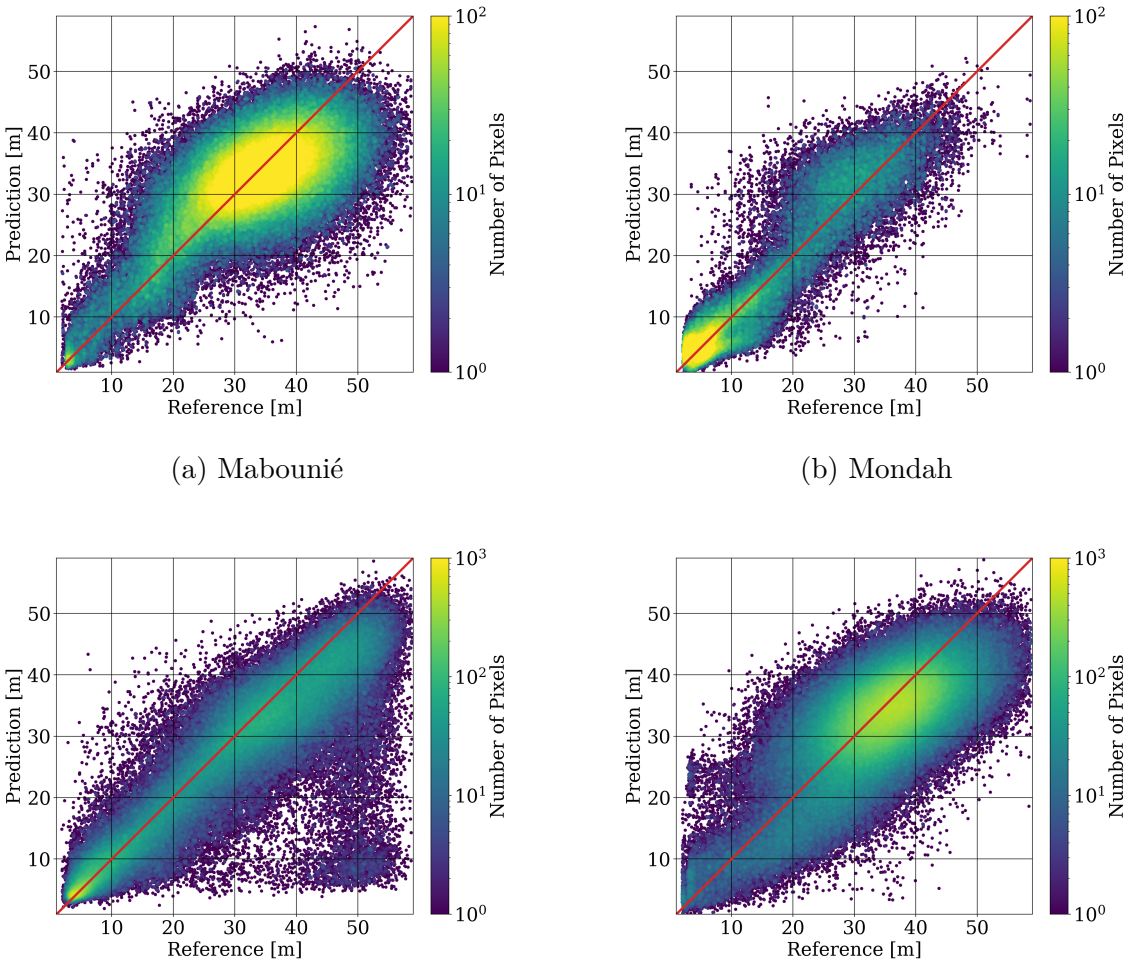


Figure 6.11: Prediction examples over the four considered test sites for the Baseline scenario. For each area, the prediction, the reference, the estimation error and sample vertical transect (identified by the red line), are presented from left to right, respectively. The transects in the rightmost column compare the reference CHM (in green) with the predictions (in blue), on top of ALS-derived DEM (in brown). The corresponding TanDEM-X Global Edited DEM is plotted (in red) for visual comparison, highlighting how the mean phase center lies within the canopy.

obtain a vegetation-independent estimate. Subdividing the slope range into equally sized intervals and plotting for each the absolute prediction error distribution yields the performance relationship shown in Fig. 6.14. For slopes between 0° to 40°, an almost linear relationship can be identified with the MAE. Notably, the mean absolute error remains below the overall average of 4.20 m for slope inclinations up to ≈ 25°, suggesting that the performance remains reasonably consistent even over high-relief terrain, where geometric distortions and geocoding errors in the RS data may become increasingly significant.



(a) Mabounié

(b) Mondah

(c) Pongara

(d) Rabi

Figure 6.12: Baseline scenario: scatterplots of the reference CHM versus the predicted CHM for each of the four considered test-sites in logarithmic scale.

6.4 Spatial Transfer Analysis

As discussed in Sec. 6, one of the goals for the proposed DL framework involves the capability of extending the trained model across larger and unseen scenarios. In order to investigate the generalization capability of the model, a *leave-one-out* cross-validation investigation is performed, by iteratively removing one out of four study-areas from the training and validation phases and setting it aside for the sole purpose of testing. The proposed test approach represents a fair attempt at providing a rich-enough training set to the network, while keeping the test area geographically isolated. Comparison with the performance obtained from the baseline scenario in Sec. 6.3 should then highlight how well generalization is achieved.

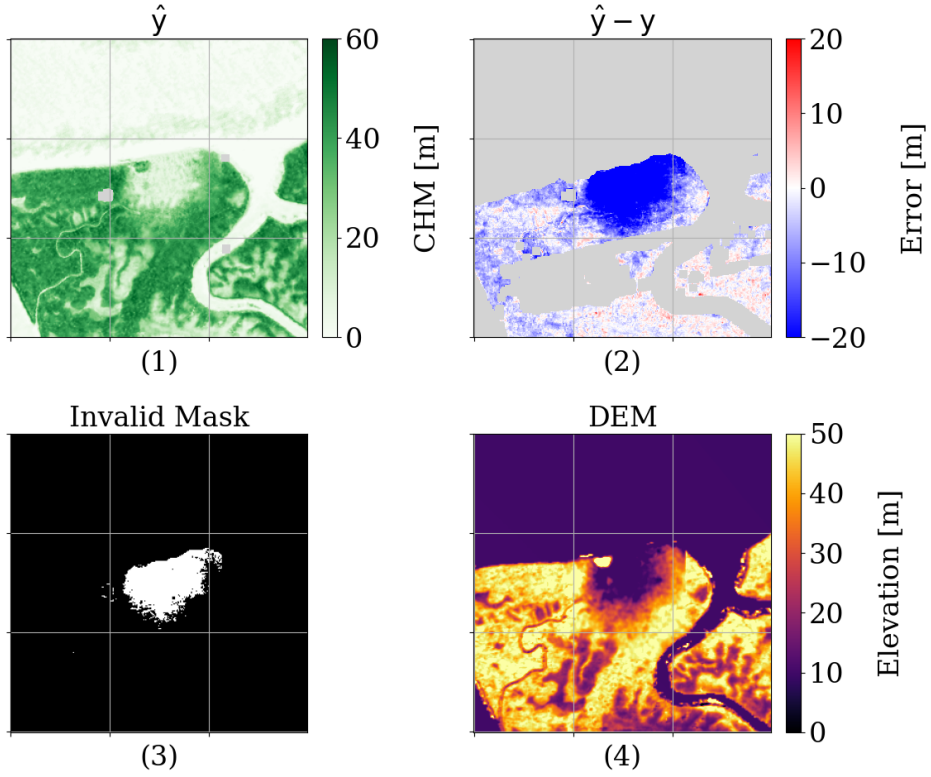


Figure 6.13: Analysis of the Pongara test site, which is characterized by a region of heavily underestimated tree heights: (a) CHM prediction, (b) prediction error, (c) binary mask used for the identification of the problematic area and (d) TanDEM-X Global Edited DEM.

The previous study area assignment strategy (detailed in Sec. 6.1.2) is modified such that, at each iteration one of the four areas is selected as the sole testing area and is completely left out from both training and validation. In order for the results to be comparable with the baseline, for each area, only the originally assigned testing subset is considered, effectively dropping $^2/_3$ of one study area at each iteration. Finally, in order to compensate for the loss of $^1/_4$ of the training areas, the testing subsets of the remaining three areas are reassigned to the training set.

The resulting subset composition obtained for each iteration is shown in Fig. 6.15, while the results for each permutation and for the experiment as a whole are presented in Tab. 6.1, together with the baseline scenario.

Overall, the experiment results in a non-negligible drop in performance compared to the baseline. MAE and MAPE increase by 0.50 m and 2.00%, respectively, while R^2 drops from 0.73 to 0.68. Looking at the individual sites, the results for Mabounié

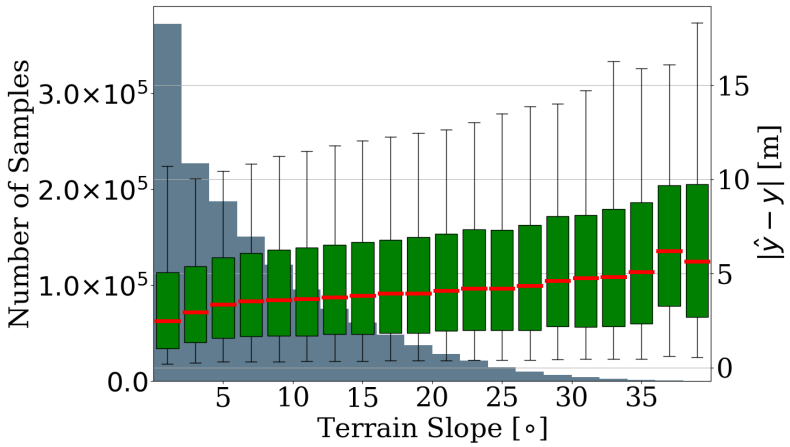


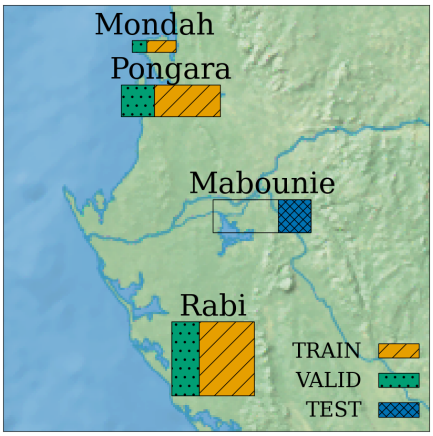
Figure 6.14: The test sample distribution compared to the MAE. The boxplots display the mean absolute error spread ($(5^th$, 25^th, 50^th, 75^th and 95^th percentiles, respectively) across the terrain slope range.

and Mondah see the performance being essentially preserved, indicating that the relative complexities of both areas are well captured by the remaining training sites.

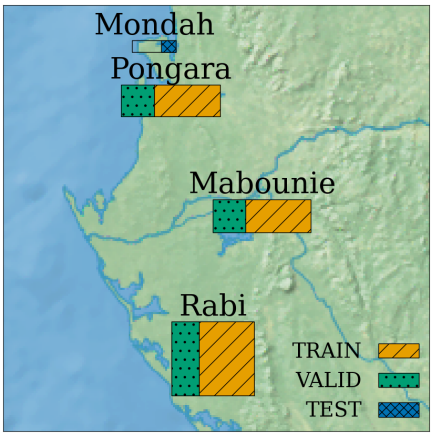
On the other hand, testing on the Pongara area yields considerably degraded performance results. In particular, a direct comparison between the scatterplots in Fig. 6.12 (c) and Fig. 6.16 (c) highlights that the model struggles with the estimation of taller canopies, effectively saturating at 45 m of height.

As described in Sec. 5.2.1, both the sites of Mondah and Pongara are partially covered by mangrove forests, with the trees found in Pongara being distinctly taller than those found in Mondah. Indeed, from Fig. 6.17 it is possible to observe that the underestimation problem is limited to the shorelines that, according to the ESA WorldCover Map, are covered by mangroves. It is therefore possible to assume that the absence of training examples of tall mangroves from the remaining areas severely compromises the regression performance.

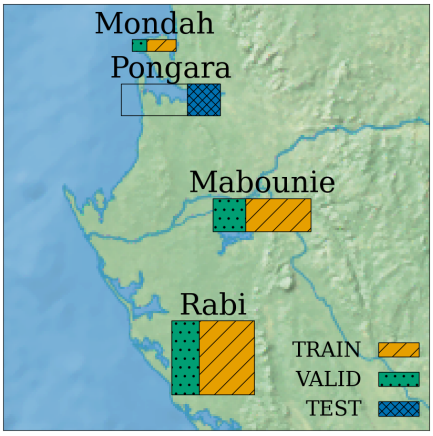
The results for Rabi also highlight a degradation in regression performance compared to the baseline. The analysis of the results does not suggest any systematic issues related to this particular iteration, and instead highlights a problematic related to only one TanDEM-X product, acquired on the 28^{th} of October, 2016. In particular, an area of significant underestimation can be spotted in Fig. 6.18 (a), which is not found in any other prediction (e.g. Fig. 6.18 (b)). Further investigation into the problem reveals that the affected area coincides with an anomalous drop in the input backscatter intensity map of about 2 dB, which can be splotted in Fig. 6.18 (c) (Fig. 6.18 (d) for reference).



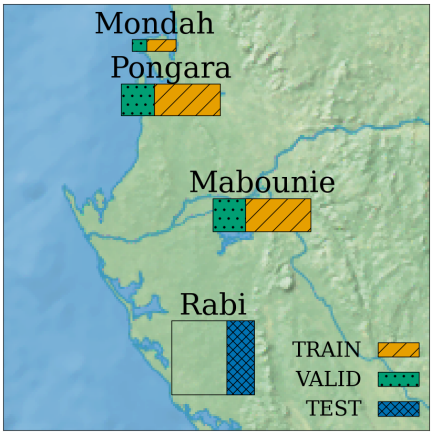
(a) Mabounié



(b) Mondah



(c) Pongara



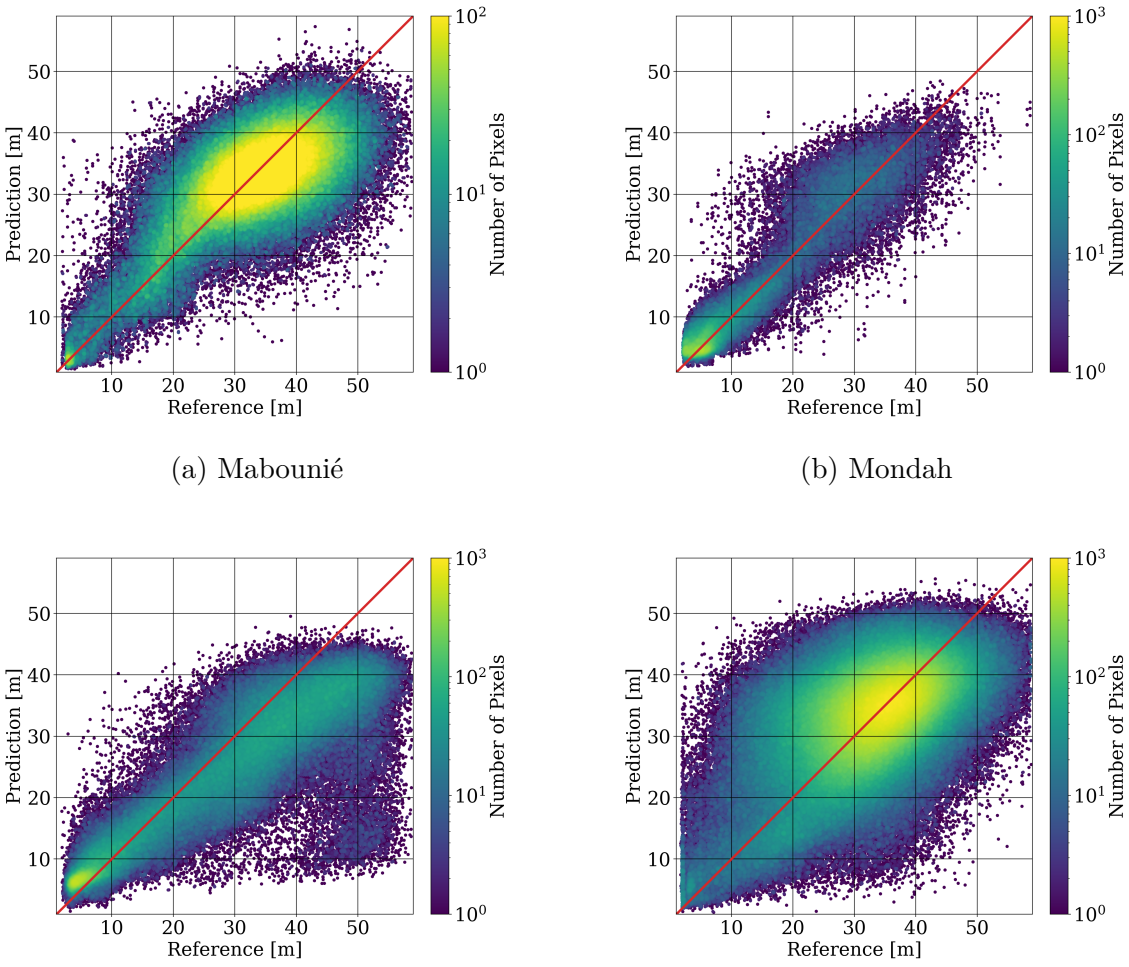
(d) Rabi

Figure 6.15: The geographical composition of training, validation and test sets, for each of the four cross-validation permutations implemented for the spatial generalization analysis.

Since neither the reference dataset nor high resolution optical imagery are able to explain the drop in backscatter intensity, it is hypothesized that this might have been induced by unusual weather conditions at the time of the acquisition. While a rare occurrence, it is known that severe precipitation and thunderstorms are capable of causing strong attenuation in the backscatter returns [142].

6.5 Comparison with the RVoG model

A high level comparison between the baseline scenario and the physical-based RVoG model is introduced to contextualize the performance of the proposed DL model with



(a) Mabounié (b) Mondah

(c) Pongara (d) Rabi

Figure 6.16: Spatial transfer analysis: cross-validation scatterplots for each of the four test-sites in logarithmic scale.

respect to physical-based approaches.

To this end, we impose to the RVoG model the same dataset limitations experienced by the proposed method, consisting of single-baseline, single-pol InSAR products and no external sources of information. These constraints effectively require the introduction of model approximations in order to guarantee a unique solution. Following the description in Sec. A.5, the model simplifies to the use of the inverse sinc approximation shown in Eq. (A.32), which only requires the estimated volume decorrelation factor γ_{vol} and the acquisition height of ambiguity h_{amb}. The resulting estimates are then used to repeat the tests detailed in the baseline scenario, in order to provide a direct comparison between the two approaches.

The performance metrics presented in Tab. 6.1 highlight that the overall performance

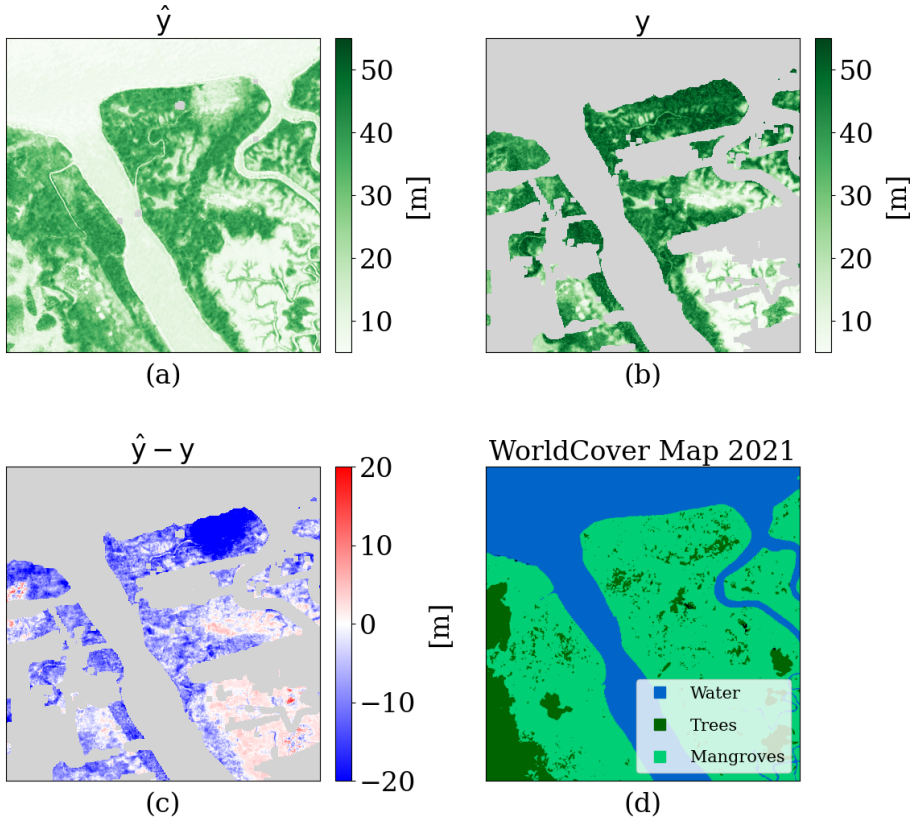


Figure 6.17: Example of underestimation over Pongara: (a) CHM prediction, (b) reference Lidar CHM, (c) prediction error and (d) corresponding land cover map. The areas covered by tall mangroves are particularly affected by a degraded performance.

of the simplified RVoG model is worse than that of the proposed DL model, achieving an overall RMSE of 8.33 m, which is significantly higher than the 5.69 m obtained by the proposed method. Of particular interest is the comparison between Fig. 6.19 and Fig. 6.10 (a), from which it is possible to observe that the RVoG model typically fails to properly predict the height of canopies below 10 m, and tends to overestimate vegetation of up to 30 m of height. This behaviour is in line with the assumption of no ground return component of the sinc-model, which becomes increasingly unrealistic for shorter vegetation. This is also confirmed by the significantly larger MAPE results for the Mondah (89.50%) and Pongara (72.24%) sites, which, due to the significant presence of open areas, heavily penalize the overestimation of the RVoG model in presence of a dominant ground scattering contribution.

For higher trees the bias decreases but the uncertainty of the prediction remains large. This behaviour can be explained by the sinc-model assumption of a uniform

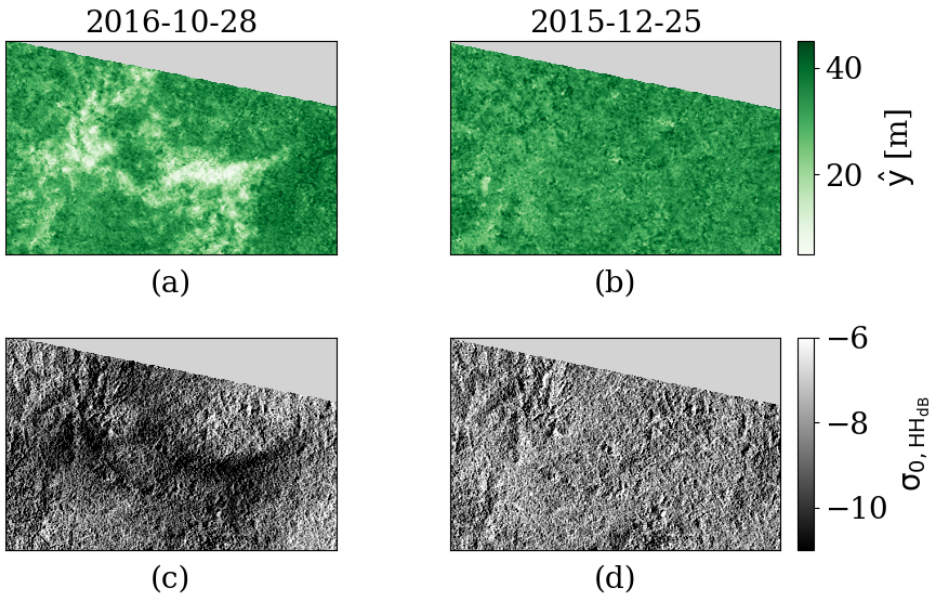


Figure 6.18: Prediction comparison for the same area in Rabi, resulting from two different TanDEM-X acquisitions. The prediction on the left presents a significant region of underestimation (a), well aligned with a region of low backscatter values (c). The image on the right does not present either pattern (b)(d).

vertical scattering profile, as it fails to capture the heterogeneous distribution of different plant species in different locations.

Notably, it is assumed that the lack of proper ground elevation modelling has no significant impact on this particular analysis, as the areas are mostly characterized by flat topography.

6.6 Feature Importance Analysis

This section delivers a comprehensive performance analysis of the proposed model, with particular focus on the contribution of the different input features on the final results.

In more detail, the following aspects are investigated:

- The relative importance of the different input features for the regression problem, concentrating on the complementarity between SAR backscatter and InSAR information.

- The impact of the SAR and InSAR acquisition geometry on the prediction performance, as well as the capability of the ancillary geometric features to contextualize

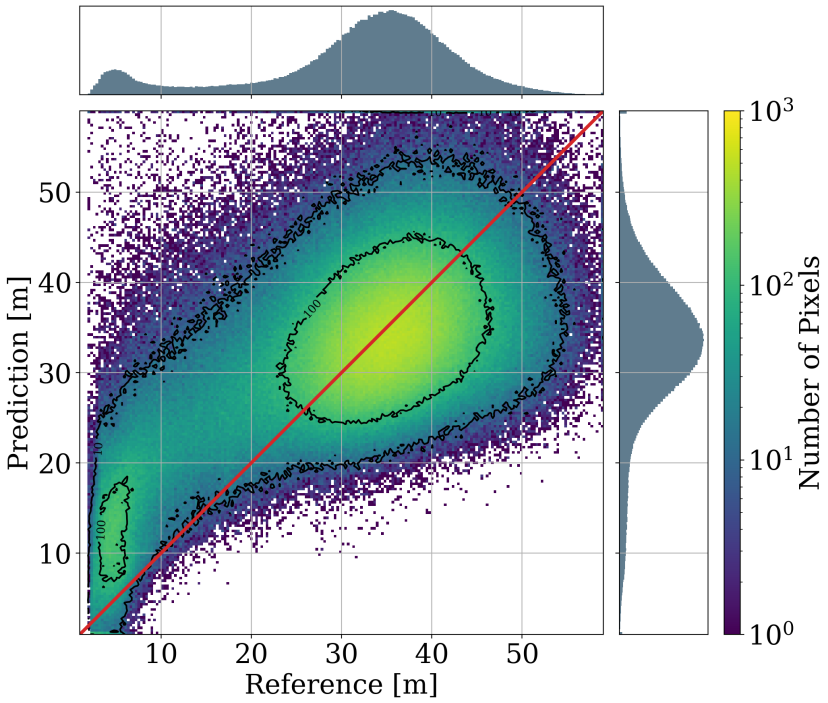


Figure 6.19: Scatterplot in logarithmic scale for the RVoG model, comparing the predicted and reference canopy height values over all four test sites in Gabon, Africa, together with the side distributions of both the reference and the prediction for the baseline scenario.

the peculiarities of the SAR side-looking geometry.

- The potential redundancy of the InSAR input features, namely the total coherence and the volume decorrelation factor.

6.6.1 Impact of DEM and local incidence angle

The SAR side-looking acquisition geometry of TanDEM-X causes the acquisitions to be intrinsically dependent on both the sensor look angle and the local topography. In order for the deep learning network to include the acquisition geometry context into the regression model, ancillary information, in the form of the local incidence angle map (θ_{lia}) and the TanDEM-X DEM, are therefore assumed to constitute necessary input features. To validate this hypothesis, dedicated models are trained without either or both features, and subsequently compared against the baseline.

The overall performance metrics, summarized in Tab. 6.2, see different levels of drops in the estimation accuracy, depending on the considered feature. While removing either features sees a small reduction in performance across all considered metrics, it is possible

Table 6.2: Overall prediction performance for all the analyzed test case scenarios. Number of considered acquisitions = 12, Number of pixels = 1465947.

Scenario	**Input features**						**ME**	**MAE**	**MAPE**	**RMSE**	$\mathbf{R^2}$
	σ^0	γ_{tot}	γ_{vol}	θ_{lia}	DEM	h_{amb}	[m]	[m]	[%]	[m]	[·]
Baseline	x	x	x	x	x	x	-1.46	4.20	15.06	5.69	0.73
Baseline w/ DEM mask	x	x	x	x	x	x	-1.38	4.12	14.91	5.41	0.76
w/o DEM	x	x	x	x	-	x	-1.10	4.48	17.66	5.98	0.71
w/o θ_{lia}	x	x	x	-	x	x	-2.48	4.65	15.85	6.30	0.67
w/o DEM, w/o θ_{lia}	x	x	x	-	-	x	-1.39	4.80	17.96	6.34	0.67
w/o σ^0	-	x	x	x	x	x	-1.51	4.37	15.38	5.95	0.71
w/o InSAR features	x	-	-	x	x	-	-2.10	4.88	17.51	6.62	0.64
w/o γ_{tot}	x	-	x	x	x	x	-1.69	4.31	15.32	5.80	0.73
w/o γ_{vol}	x	x	-	x	x	x	-1.91	4.33	15.16	5.84	0.72
w/o h_{amb}	x	x	x	x	x	-	-1.07	4.19	14.98	5.67	0.74
w/o γ_{vol}**, w/o** h_{amb}	x	x	-	x	x	-	-1.13	4.22	15.10	5.75	0.73
w/o γ_{tot}**, w/o** h_{amb}	x	-	x	x	x	-	-2.07	4.59	15.70	6.42	0.66

to observe that removing the local incidence angle map θ_{lia} especially impacts the ME, MAE, RMSE and R^2 scores (cases **w/o DEM**), while removing the DEM particularly impacts the MAPE metric (case **w/o** θ_{lia}).

The performance analysis in Sec. 6.3.2 has already highlighted a dependency between the DEM and the regression performance. This subsequent finding further adds to the interpretation that the DEM, in the form of its local variation, appears to be especially useful in detecting the absence of vegetation, since a larger MAPE for a lower MAE can be directly tied to a disproportionally larger error for reference values closer to zero (i.e. in absence of vegetation).

On the other hand, the absolute loss in performance associated to the θ_{lia} feature, shows an intrinsic dependency on the slant-range distance, which is directly mirrored in SAR features such as the backscatter.

Finally, simultaneously removing both features (case **w/o DEM, w/o** θ_{lia}) from the input causes the performance to see a further, less significant drop in line with the worst performing metrics obtained by dropping both features. This suggests that neither metric is fully capable of replacing the other one, as these both serve complementary purposes.

6.6.2 SAR and bistatic InSAR feature comparison

One of the unique advantages of the TanDEM-X constellation is the capability to acquire single-pass InSAR information. As seen in Sec. 2.2.3, this allows one to directly isolate the volumetric scattering contribution of forests from other decorrelation effects. In order to assess what impact the interferometric features have over the more readily available SAR backscatter maps, two more scenarios are presented in which either of these feature-subsets is removed. The results for the SAR-only (**w/o InSAR features**)

and the InSAR-only (**w/o** σ^0) cases are presented in Tab. 6.2.

Both cases see a decrease in the overall performance when compared to the baseline scenario. The results for the SAR-only case see the MAPE worsen an additional 2.12% and the RMSE by 0.68 m compared to the InSAR-only case, achieving the overall worst performance across all tested scenarios.

Focusing on InSAR features then, five additional scenarios are considered to determine the relative impact of the interferometric coherence, the volume decorrelation factor and the ancillary height of ambiguity map on the overall results.

In particular, removing either the total coherence (γ_{tot} case) or the volume decorrelation factor (γ_{vol} case) does not yield significant loss in the overall performance. This meets the theoretical expectation that both features should provide similar information, given their direct relationship through other unrelated decorrelation factors.

Moreover, removing the height of ambiguity information (**w/o** h_{amb} case) does not cause any significant performance degradation. This behavior is expected, given the direct relationship between h_{amb} and γ_{vol} (an by extension γ_{tot}) [26].

In order to better contextualize this result, the impact of pairwise dropping h_{amb} and γ_{vol}, or h_{amb} and γ_{tot}, is tested.

The results for the former (**w/o** γ_{vol}**, w/o** h_{amb} case) see no performance degradation, suggesting that γ_{tot} does not require additional ancillary information in the form of the InSAR h_{amb}. On the other hand, the latter scenario (**w/o** γ_{tot}**, w/o** h_{amb} case) sees a significant deterioration in performance, which is in line with the SAR-only case (i.e. **w/o InSAR features**), indicating that the additional contextual information from the h_{amb} is indeed necessary. Finding an explanation for this behaviour is not straightforward. From the theory, it is possible to derive the following two principles about the relationship between γ_{tot}, γ_{vol} and h_{amb}:

- For the same tree heights, γ_{vol} (and therefore also γ_{tot}) is directly proportional to h_{amb} (see Eq. (A.33)).

- From the relationship between γ_{tot} and γ_{vol}, expressed in Eq. (2.39), it follows that $\gamma_{\text{vol}} = \alpha \cdot \gamma_{\text{tot}}$, where $\alpha > 1$ is the inverse of the combined effects of all remaining decorrelation factors.

While both γ_{tot} and γ_{vol} are monotonically increasing proportionally to h_{amb}, γ_{tot} grows slower than γ_{vol} (i.e. their spread increases) and therefore remains more constant throughout the whole h_{amb} range. This behaviour can be seen in Fig. 6.20 (a), which shows the actual relationship between γ_{tot} and γ_{vol} for the considered dataset, in

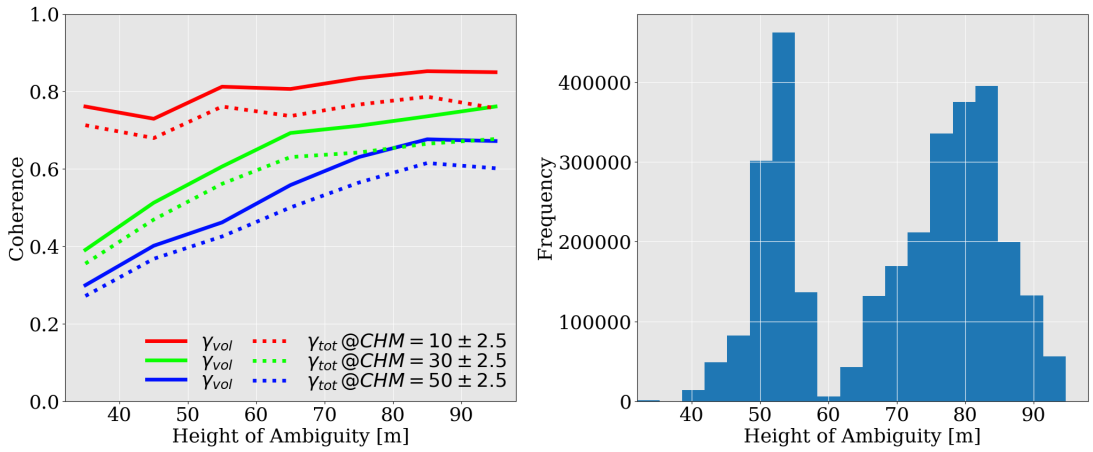


Figure 6.20: (a) Variation of the coherence γ_{tot} and volume decorrelation factor γ_{vol} with respect to the height of ambiguity h_{amb} for different values of CHM. (b) Distribution of h_{amb} for the acquisitions esed for training the proposed DL model.

presence of different tree heights. A possible interpretation for the observed behaviour could therefore be given by the particular h_{amb}-range covered by the considered dataset (Fig. 6.20 (b)), which could be limited enough to not require any additional ancillary information when considering the less sensitive γ_{tot}, as it predominantly covers a relatively constant region of the relationship.

6.7 Discussion

The experimental results presented in this chapter provide an interesting insight into the synergistic potential of TanDEM-X products and deep learning methods for the retrieval of forest height estimates.

The performance baseline results (Sec. 6.3) show a strong agreement between the predictions and the reference samples. The forest height estimates mostly saturate around 50 m of height, and achieve the lowest biases and spreads for the most represented height values (i.e. CHM≈5 m and CHM≈35 m). This simultaneously displays a clear dependency on the dataset composition but no significant dependency on the actual estimated tree height.

The spatial transfer analysis (Sec. 6.4) shows that, in principle, the proposed framework generalizes well across unseen areas, as far as all the possible complexities of that area are also represented by the training set, a condition which can be met with careful planing and analysis of the target area.

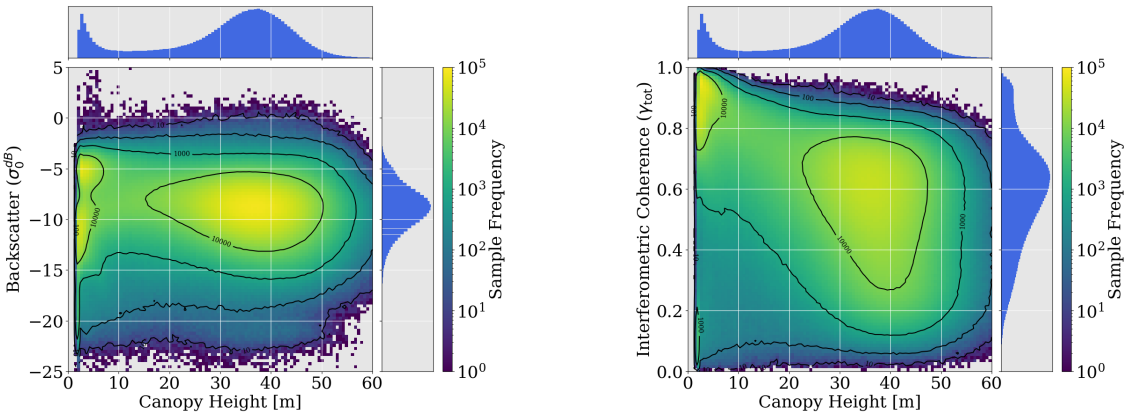
The feature importance analysis (Sec. 6.6) also evidences interesting behaviours for the considered input features. Firstly, it confirms the key role of the local incidence angle map θ_{lia} for the inversion problem, as it directly affects other features, such as the SAR backscatter and the interferometric coherence, as well as describes the interaction with the canopy layer. This meets the theoretical expectation of the SAR acquisition principle and has since long been integrated into the existing physical-based models [76][143], but is surprisingly absent in the existing data-driven approaches which instead rely only on terrain-corrected features [99][102][100][37][38].

Secondly, it also shows that the InSAR-derived DEM plays a crucial role when it comes to correctly detecting the absence of vegetation, which is interestingly achieved by separating the high-frequency contribution from the encoded mean phase center from the low-frequency topography. Again, the inclusion of such a source of information is consistently missing in the data-driven literature [99][102][100][37][38].

In conclusion, neither geometric feature is fully capable of replacing the other, and both are strictly necessary to achieve peak accuracy.

Regarding the impact of SAR and bistatic InSAR features, it is possible to reach the following conclusions:

- Removing the InSAR features has a significantly larger impact than removing the backscatter information.
- Removing either the interferometric coherence or the volume decorrelation factor results in no significant loss in accuracy. This is in line with theoretical expectation, as it confirms that both features hold similar information.
- The ancillary height of ambiguity maps are not strictly necessary if both γ_{tot} and γ_{vol} are used as inputs. This might suggest that information about the baseline might be inferred from the two features alone, but it might be more likely connected with the following aspect.
- Using the interferometric coherence alone incurs in no performance penalties, while using the volume decorrelation factor alone does. This could be explained by the considered dataset, which mostly samples small sub-ranges of the height of ambiguity curve, in which the interferometric coherence presents an almost constant relationship with the height of ambiguity itself. This behaviour would consequently also explain the previous finding, but a dataset covering a more representative portion of the height of ambiguity range is necessary to fully confirm this behaviour.



(a) (b)

Figure 6.21: Scatterplots representing the relationship between CH and σ_0 (a) and between CH and γ_{tot} (b).

In the quantitative performance analysis with the RVoG model (Sec. 6.5), the approximated inverse of the sinc function achieves results in line with the state-of-the-art physical-based approaches proposed in [28], [83] and [84], while in comparison to the proposed deep learning approach, the RVoG estimates perform significantly worse. In practice the physical-based model displays also further disadvantages, such as a hard-cap on the maximum appraisable tree height (due to the acquisition height of ambiguity), as well as the impossibility to reliably estimate the absence of vegetation without the use of external forest classification maps.

Finally, Fig. 6.21 displays the complex and noisy relationship between the canopy height with backscatter (a) and with the interferometric coherence (b). Fitting a curve to each feature individually is insufficient to achieve a reliable model and demands a more complex architecture that can jointly leverage multiple predictors.

6.8 Conclusion

This chapter presented a novel study on the potential of deep learning methods for forest height regression from TanDEM-X bistatic InSAR data. A fully convolutional framework was proposed, which is capable of delivering tree height estimates from a single acquisition, enabling the creation of time-tagged geomaps. The model was trained in a fully-supervised fashion using LVIS data from the 2016 AfriSAR campaign, consisting of ALS-derived canopy height maps covering the tropical forests in Gabon,

West Africa. The framework achieved an overall regression performance of 4.12 m MAE, 14.9% MAPE and 5.41 m RMSE.

The feature-importance analysis showed that the model is primarily driven by InSAR features, with contextual information about the acquisition geometry and the scene topography being crucial in delivering peak performance. The performance comparison between the proposed framework and the RVoG model highlighted a considerable performance advantage of the DL model over the physical-based model.

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

The experiments have also suggested that the achievable accuracy remains strongly connected to the training set characteristics, requiring careful balancing to maintain performance over different challenging scenarios. Saturation problems in the estimates were also encountered, which, on the one hand, could be caused by the limited penetration capabilities of the X-band signal into the canopy, and, on the other hand, by a general lack of labelled data for very tall trees above 40 m.

In order to mitigate these effects, future works should extend the analysis to higher resolutions, as TanDEM-X products with fewer looks have already shown a certain potential to retrieve the location of the ground surface [93][144]. Furthermore, smarter dataset sampling and balancing approaches should be investigated, aiming at homogenizing the network performance. Finally, it is worth pointing out that the proposed spatial transfer analysis was performed on a limited region of interest (the state of Gabon), and thus should be considered as a preliminary assessment of the generalization capability of the network. Therefore, more effort should also be devoted to the generalization of the model for large-scale inference, e.g. by considering a larger span of viewing geometries and InSAR configurations, as well as forest types, local terrain characteristics and spaceborne-derived reference data.

Chapter 7

Generalization of the Framework to Country-Scale

Following the definition of the core regression framework in Ch. 6, this chapter covers its further development for the purpose of a country-scale deployment. The methodology for the generation of large-products and that for their validation is described in Sec. 7.1. The practical challenges related to a naïve large-scale deployment are extensively discussed in Sec. 7.2 and subsequently addressed through the investigation of three pragmatic research assets presented in Sec. 7.3, in Sec. 7.4 and in Sec. 7.5, respectively. Furthermore, the already known framework limitations (as highlighted in Sec. 6.8) are investigated and addressed. The chapter culminates in the final sample demonstration of a country-scale, time-tagged canopy height mosaic for the state of Gabon, Africa. The main content of this chapter, corresponding to the work in Sec. 7.2 and in Sec. 7.4, is based on the work published in [145], whereas the pre-processing algorithm detailed in Sec. 7.5 is currently patent pending [146].

7.1 Country-Scale Inference and Map of Applicability

This section describes the strategy for the generation of the country-scale CH maps and introduces the proposed product validation strategy in form of the novel *Map of Applicability* (MoA) concept.

7.1.1 Country-Scale Inference

For the generation of the final country-scale CHM mosaic, the TanDEM-X products from the first global coverage are considered, which were acquired for the generation of the global DEM product [147]. This choice guarantees an almost complete coverage of the country of Gabon with minimal gaps, given that the TanDEM-X products are otherwise acquired irregularly and depending on the specific acquisition planning. Conversely, some regions are imaged multiple times per year (e.g., the AfriSAR test sites). In order to evaluate the single-baseline quality of the proposed method a simple filtering strategy is implemented. First, the mean h_{amb} value is computed from the overall training distribution. Then, in presence of overlapping acquisitions only the one with the minimum distance from this value is considered. The trained model is then applied to each acquisition individually to generate a corresponding CH estimate. Finally, the resulting list of forest height maps is mosaicked together to generate the country-scale product.

7.1.2 Map of Applicability

The independent validation of the Gabon country-scale product is extremely challenging, since:

- the AfriSAR campaign dataset, while constituting a relevant and varied subset of the Gabonese ecosystems, is inadequate to determine the trustworthiness of a significantly larger product.
- the existing field plots are limited to significantly smaller geographic scales and are linked to the areas covered by the AfriSAR campaign.
- existing spacerborne LiDAR missions, such as the Global Ecosystem Dynamics Investigation (GEDI) [20], are reportedly ill-suited to function as reference data in the presence of tall and dense canopy [75][148][149], an aspect which will be analyzed in detail in Sec. 7.3.

A possible alternative approach consists in assessing the reliability of the regression model itself. For a deep learning approach, this can be achieved by validating that during inference the input predictors fall within the subspace sampled by the training data, as data-driven methods fail to perform reliably on out-of-distribution (OOD) predictor combinations [150][151]. Ideally, this assessment would be performed exhaustively by first determining the minimum distance between the inference and training predictors

in the hyperspace defined by the predictor features, and subsequently determining, on the basis of validation data, where the trustworthiness of the model falls off. Unfortunately, such an exhaustive calculation is computationally intractable for a large number of samples.

In order to overcome this issue, an ad-hoc OOD-detector is proposed, which is based on an approximation of the joint predictor distribution as a proxy for the training set representation in the predictor space. First, histograms are computed for each of the predictors across globally-defined value ranges [1]. Then, the relative sample occurrence $d_{j,i}$ (i.e. the density distribution) is computed for each of the $i = 1, \ldots, N$ bins and the $j = 1, \ldots, J$ predictors as:

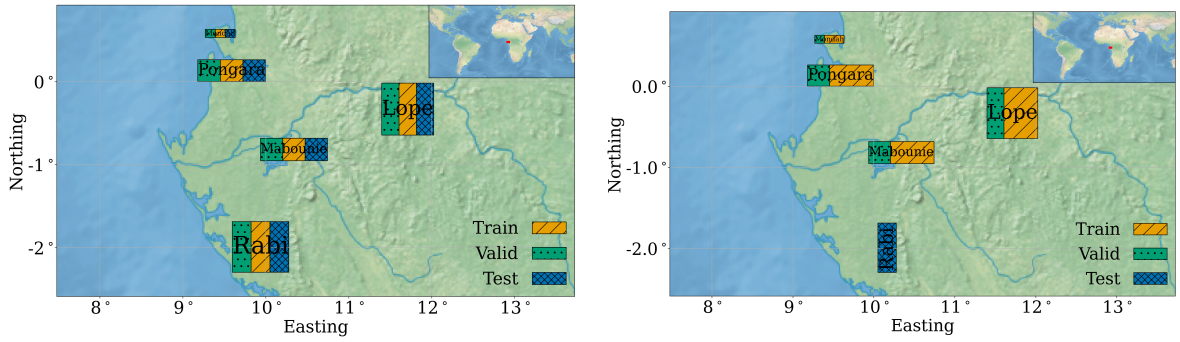
$$d_{j,i} = 100\% \cdot \frac{h_{j,i}}{\sum_{n=1}^{N} h_{j,n}} \tag{7.1}$$

where $h_{j,i}$ is the (absolute) sample occurrence. These pre-computed histograms and their bin-edge values are used at inference to classify each predictor-pixel and associating to it the corresponding relative sample occurrence value $d_{j,i}$. Repeating this process for all pixels leads to the generation of a set of geographic maps representing, for the position (x, y) and the feature j, the relative sample frequency $d_{j,x,y}$ of seen by the model during training. These maps are then aggregated into a single reliability score map $S_{x,y}$ by computing their geometric mean as:

$$S_{x,y} = \sqrt[J]{\prod_{j=1}^{J} d_{j,x,y}} \tag{7.2}$$

The resulting mean density map is directly correlated to the model reliability, as values approaching zero have jointly seen fewer samples in the corresponding training sub-space. Finally, the validation set is used to select the threshold value of $S_{x,y}$ that minimizes the prediction MSE, allowing for the generation of a binary *Map of Applicability* (MoA). In particular, values smaller than the threshold are considered unreliable and can be discarded.

[1] For each predictor, the following ranges are considered: $\sigma_{HH}^{0,dB} \in [-25, 10]$, $\theta_{\text{inc}} \in [0, -\frac{\pi}{2}]\,\text{rad}$, $\gamma_{\text{tot}} \in [0, 1]$, $\gamma_{\text{vol}} \in [0, 1]$, $h_{\text{amb}} \in [15, 120]\,\text{m}$, $h_{\text{DEM}} \in [0, 1100]\,\text{m}$, $h_{\text{DEM}}^{\text{HPF}} \in [-700, 700]\,\text{m}$, $\nabla h_{\text{DEM}} \in [-4, 4]\,\text{m}\,\text{m}^{-1}$.



(a) (b)

Figure 7.1: (a) Geographic sub-setting of the AfriSAR campaign study areas into training, validation and testing, (b) Exemplary geographic sub-setting for the spatial cross-validation, considering Rabi a testing area only.

7.2 The Challenge of Missing Predictor Representations

Starting from the core regression framework defined in Ch. 6, a series of modifications are discussed in Sec. 7.2.1 to address the limitations previously evidenced in Sec. 6.8. These updates affect the geographic and feature composition of the dataset, but preserve the previously defined architecture. The resulting updated framework is subsequently tested and used for the generation of a country-scale CHM prototype in Sec. 7.2.2, which represents the starting point for large-scale inference.

7.2.1 The Updated Training Strategy

Firstly, the fifth and previously missing AfriSAR study area of Lopé is added to the considered pool of test sites, resulting in a larger and more heterogeneous dataset. The updated geographical sub-setting configuration is shown in Fig. 7.1 (a) and the example of the Rabi cross-validation permutation is depicted in Fig. 7.1 (b).

Secondly, the TanDEM-X edited DEM product is replaced by the raw acquisition DEM. This avoids post-processing related editing artefacts and allows all predictors to be temporally consistent.

Finally, patches are now centered and extracted for each existing reference sample. This also allows for the inclusion of non-vegetated pixels in the off-center area of a patch, thus adding a larger variety of scenarios and potentially improving the robustness of the model. In order to accommodate this change, the previously used two-term loss function expressed in Eq. (6.1) is now computed only on patch centers. The updated

2015-2016 Dataset Performance					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Experiment	[m]	[m]	[%]	[m]	[·]
Lopé	-0.42	4.12	11.25	5.34	0.40
Mabounié	-1.03	4.82	15.84	6.23	0.37
Mondah	0.89	2.25	28.57	3.22	0.90
Pongara	-0.05	3.03	17.30	4.28	0.92
Rabi	-2.03	4.18	13.58	5.32	0.53
Overall	-0.96	4.05	14.28	5.31	0.75

(a)

Cross-validation Performance					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Experiment	[m]	[m]	[%]	[m]	[·]
Lopé	-0.24	5.23	15.27	6.75	0.04
Mabounié	-0.25	4.79	16.20	6.21	0.38
Mondah	0.84	2.35	29.50	3.32	0.90
Pongara	-1.60	4.56	23.27	5.97	0.84
Rabi	-0.79	4.03	13.54	5.18	0.56
Overall	-0.52	4.54	16.38	5.94	0.69

(b)

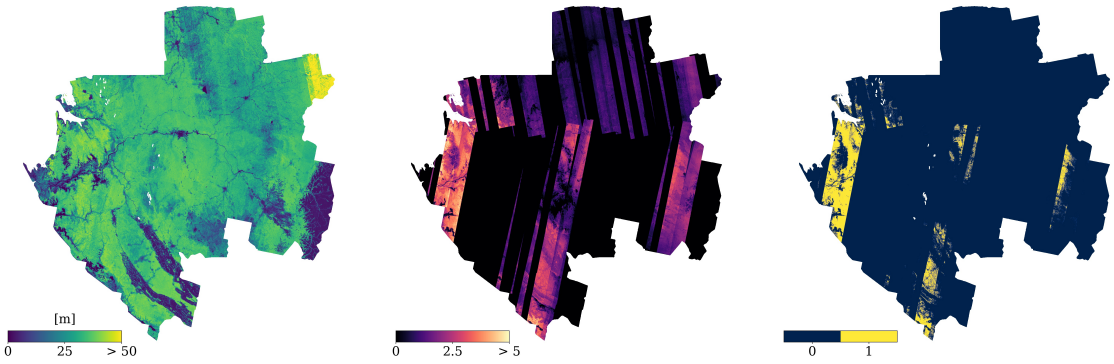
Table 7.1: (a) Performance metrics computed for the model trained using TanDEM-X data acquired in 2015-2016 only, for each AfriSAR test site, separately, and overall. (b) Performance metrics computed for the cross-validation experiment, for each AfriSAR test site permutation, separately, and overall.

sampling strategy results in $13 \cdot 10^6$ unique training samples, which on a single NVIDIA A100 GPU results in a training time of roughly 9 hours for the proposed model.

7.2.2 Results

As already done in Ch. 6, only TanDEM-X data acquired during 2015-2016 are selected for training, in order to be as temporally consistent as possible with the LVIS reference dataset. The resulting regression performance for this baseline model is summarized in Tab. 7.1a for each test site and overall, resulting in a ME of -0.96 m, a MAE of 4.05 m, a MAPE of 14.28%, an RMSE of 5.31 m and an R^2 of 0.75.

As in Sec. 6.4, a spatial transfer analysis (i.e., a geographic cross-validation) is performed to assess the generalization capabilities of the model; the results of the test are summarized in Tab. 7.1b. The performance metrics confirm that the model is robust also when tested on totally independent test sites, as only a small loss in performance (with respect to the baseline case) is detected (ME of -0.52 m, MAE of 4.54 m, MAPE of 16.38%, RMSE of 5.94 m and R^2 of 0.69). This is limited to the Lopé and Pongara test permutations, and can be attributed to the unique phenological and topographical characteristics found at these sites.



(a) (b) (c)

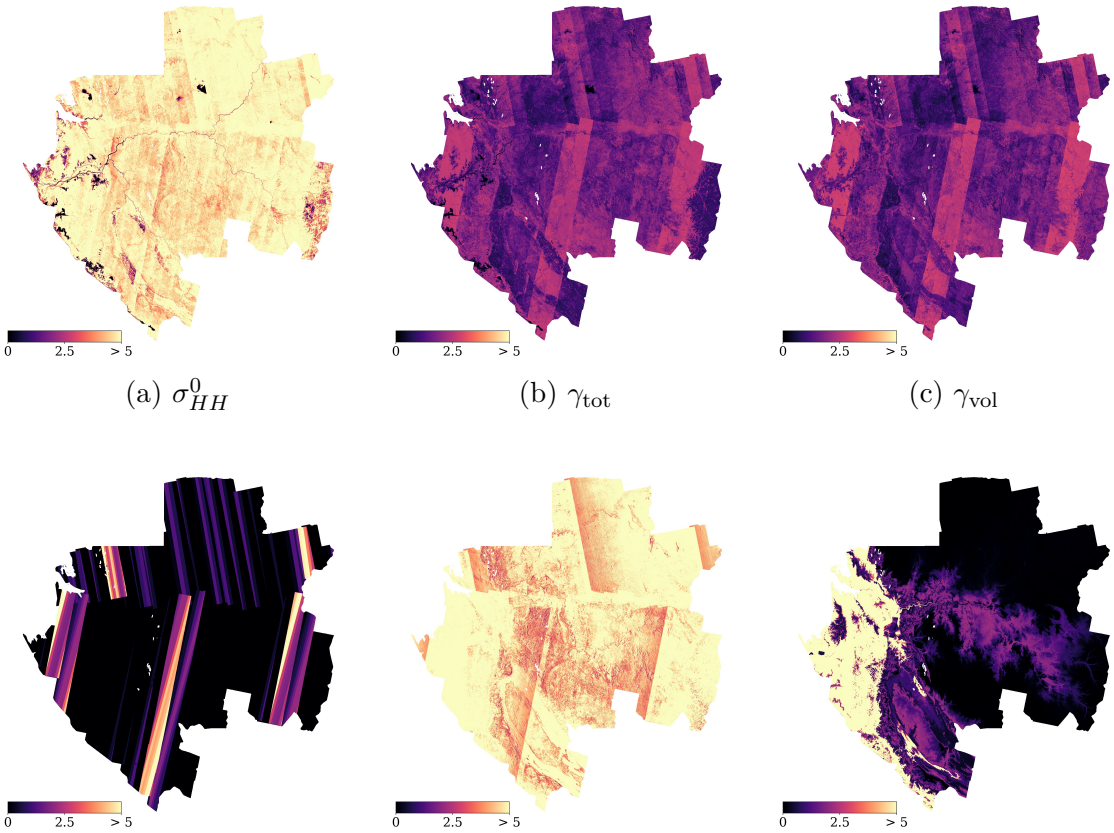
Figure 7.2: (a) The country-scale CHM generated using the baseline model trained using TanDEM-X from 2015-2016 only, (b) the corresponding reliability score map and (c) the binary MoA.

Finally, the baseline model is applied at country-scale by considering all available TanDEM-X acquisitions acquired between December 2010 and the end of 2011 in correspondence of the first global coverage of the TanDEM-X mission. The country-scale CH mosaic, as well as the corresponding reliability score map and MoA are generated as presented in Sec. 7.1 and the results are depicted in Fig. 7.2.

7.2.3 Discussion

The MoA for the country-scale CH product presents extended regions of null values which correspond to entire TanDEM-X data-takes. The actual contribution of each single predictor to the reliability score map is presented in Fig. 7.3 and it reveals that the most critically underrepresented predictor is the height of ambiguity, which is directly related to the InSAR acquisition geometry. Furthermore, also the relative sample frequency map for the raw DEM presents extended null areas, significantly compounding the unreliability of the country-scale CH mosaic. The missing representations of the corresponding value subranges can also be identified in the feature histograms presented in Fig. 7.4. Consistently, severe underestimation of the CH occurs in the country-scale mosaic in Fig. 7.2 (a) in correspondence of zero values of the MoA. Therefore, at the present stage, the initial baseline model cannot be considered to be reliable outside of the AfriSAR test regions.

Achieving naturally harmonized predictions primarily requires to solve the representation mismatch between the training and prediction heights of ambiguity, as already demonstrated in [32] for large-scale forest mapping. To tackle this issue, three research



(a) σ^0_{HH} (b) γ_{tot} (c) γ_{vol}

(d) h_{amb} (e) θ_{inc} (f) Raw DEM

Figure 7.3: Relative sample frequency for each predictor (indicated below each image) of the model presented in Sec. 7.2, trained with TanDEM-X data acquired in 2015/2016 only.

assets are proposed in this chapter:

- The addition of more reference data represents the first and the most straightforward proposed solution as it potentially leads to both a larger pool of matching TanDEM-X acquisitions (and therefore a better representation of the acquisition geometries), and a more complex representation of the considered phenology. In practice, this approach is hindered by the scarce availability of large-scale reference data. In this context, the GEDI mission represents a unique source of information, and its potential for the proposed regression framework is explored in Sec. 7.3.

- The second proposed approach consists in a modification of the predictors and the training set. In particular, a relaxation of the stationarity constraint is introduced in Sec. 6.1.1. This results in a larger compatibility with the existing TanDEM-X data archive, thus leading to a more representative selection of acquisition

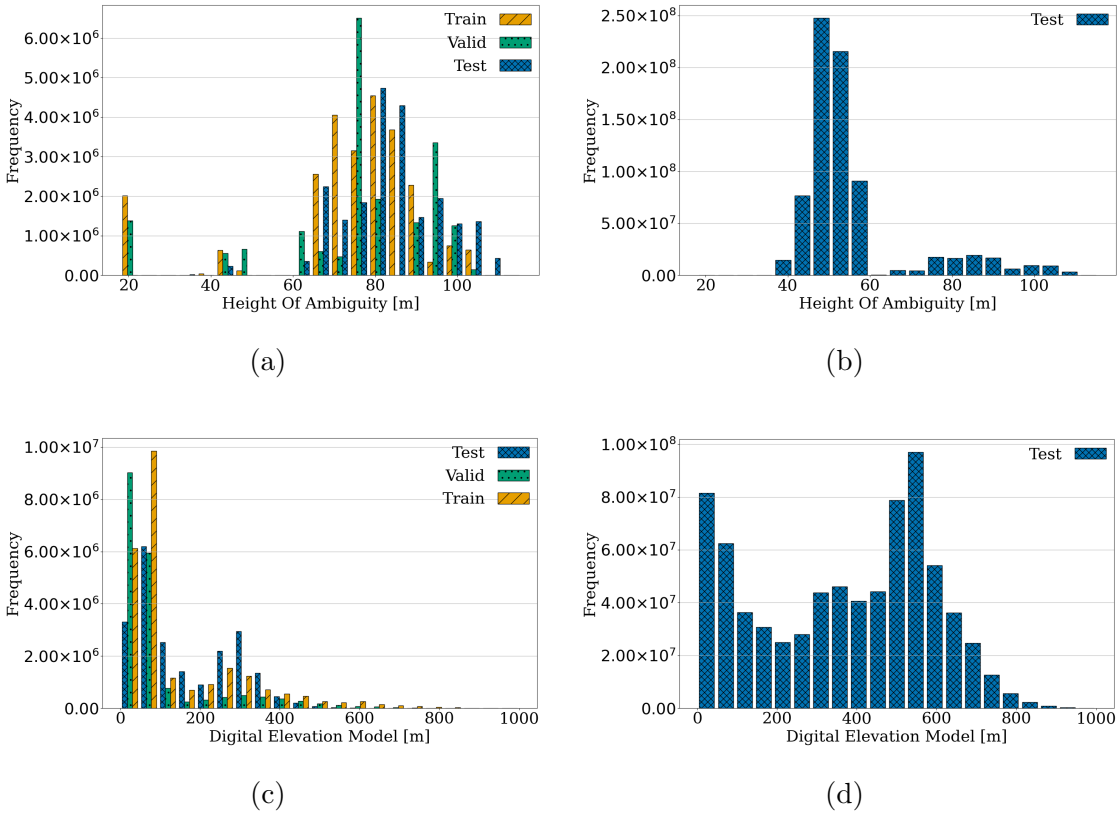


Figure 7.4: (a) Height of ambiguity histogram for the training, validation and testing datasets corresponding to the 2015-2016 TanDEM-X acquisitions, (b) height of ambiguity histogram for the 2010-2011 TanDEM-X acquisitions used for the country-scale inference. (c) DEM histogram for the training, validation and testing datasets corresponding to the 2015-2016 TanDEM-X acquisitions, (d) DEM histogram for the 2010-2011 TanDEM-X acquisitions used for the country-scale inference.

geometries, at the cost of a noise increase in the input-to-output mapping process. Moreover, missing representations in the DEM feature space during training are handled by proposing the introduction of proxy variables describing the local topographic characteristics, such as the DEM gradient and a high-pass filtered version. The the details of this proposed solution are investigated in Sec. 7.4.

- Finally, the third proposed solution consists in the introduction of a deep learning-based pre-processing step, which harmonizes the TanDEM-X acquisitions before these are used as inputs to the proposed regression framework. This particular aspect is currently under investigation, but the principles of the approach, as well as early promising results are discussed in Sec. 7.5.

Table 7.2: Performance metrics of the rasterized GEDI shots, compared to those acquired with the LVIS sensor. On the one hand, the different RH metrics display a progressive negative increase in the estimation bias.

GEDI versus LVIS Performance Metrics					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	[·]
RH95	-0.93	5.76	22.54	8.16	0.469
RH99	-1.16	5.84	19.82	8.33	0.465
RH100	-2.20	6.02	18.20	8.58	0.467
DTM	3.41	5.28	8.29	8.03	0.997
DSM	1.21	3.90	3.47	6.58	0.999

7.3 Research asset 1: suitability of the GEDI dataset

The possibility to add a large quantity of reference data matching the TanDEM-X acquisitions is directly constrained by the very limited availability of canopy height measurements at large scale. The GEDI mission (see Sec. 5.3) represents a unique exception, consisting of sparsely sampled CHM LiDAR measurements covering a significant portion of the Earth's surface, making it an attractive proposition for large-scale supervised applications.

In the following, the GEDI samples are evaluated for their consistency with reference LVIS data and are subsequently employed in a proof-of-concept study for the training and testing of the TanDEM-X regression framework.

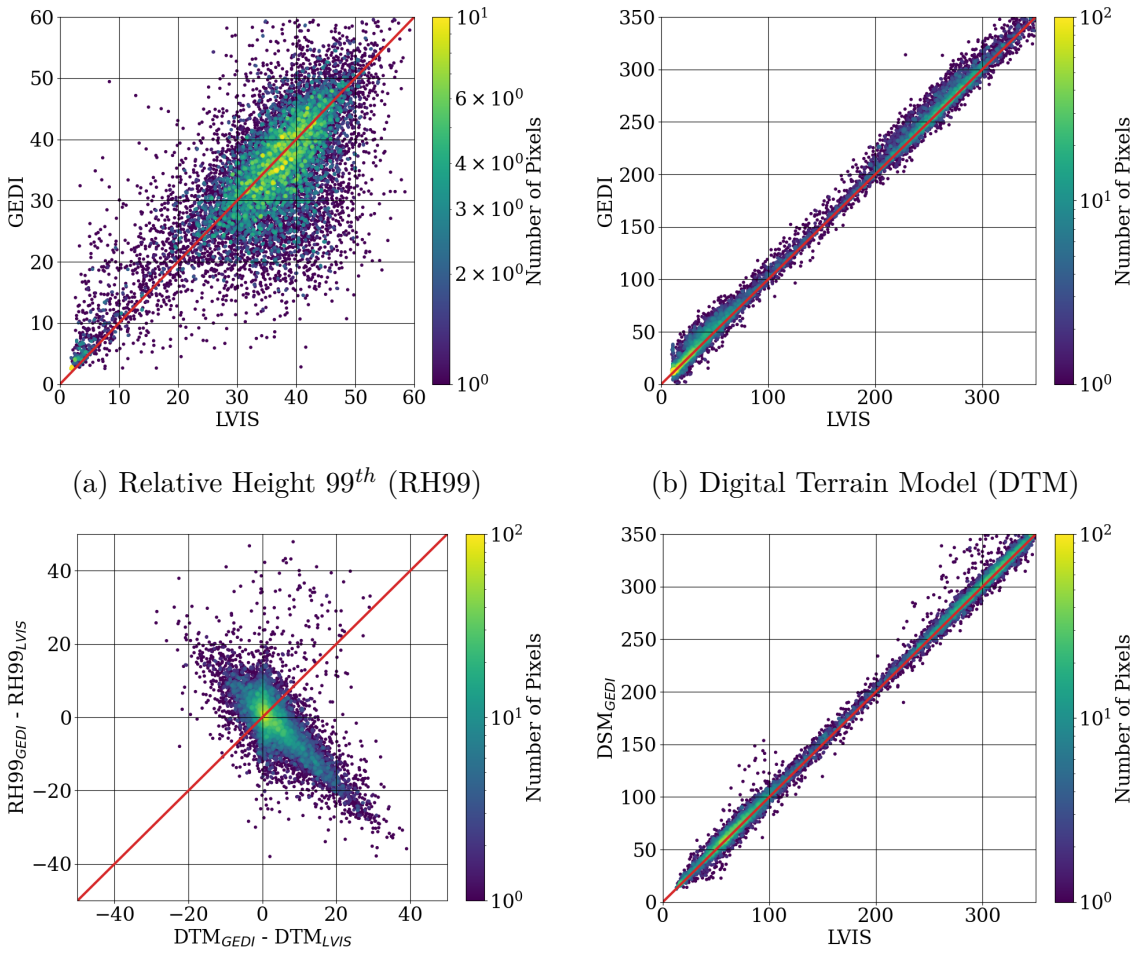
7.3.1 Comparison against LVIS Measurements

In order to determine with confidence the suitability of the GEDI products for the considered regression task, in the following the calibration of the GEDI measurements is evaluated against the airborne LVIS measurements of the AfriSAR campaign.

Downloading the entire available L2A dataset for the five AfriSAR ROIs results in an initial sample count of $8.8 \cdot 10^6$ footprints, which reduces to $1.2 \cdot 10^4$ samples (or 13.4% of the initial count) after applying the filtering steps described in Sec. 5.3.

Following this, the samples are rasterized on top of the AfriSAR grid, such as to generate spatially consistent maps of the DTM, RH95, RH99 and RH100 statistics.

The results of the comparative analysis for all metrics are reported in Tab. 7.2. When looking at the overall performance for the RH99 metric, GEDI presents an average bias of -1.16 m, a MAPE of 19.82% and an R^2 of 0.465. The scatterplot in Fig. 7.5 (a) displays a good agreement between a subset of the GEDI samples and the LVIS measurements. On the other hand, it also highlights a small subset of the samples in strong disagree-



(a) Relative Height 99^{th} (RH99)

(b) Digital Terrain Model (DTM)

(c) DTM and RH99 Estimation Errors

(d) Digital Surface Model (DSM)

Figure 7.5: Comparative scatterplots between LVIS- and GEDI-derived height statistics.: the RH99 metric (top-left), the DTM (top-right), the DTM and RH99 estimation errors (bottom-left), and the DSM computed from the respective DTM and RH100 metrics (bottom-left).

ment, with a large spread from the identity line. This behaviour could be attributed to the well documented geo-location errors of the GEDI shots [74]. Finally, it is also possible to note a third phenomenon, with samples at large tree heights that are, on average, underestimated by GEDI, which requires more investigation.

Interestingly, in Fig. 7.5 (b) it is possible to appreciate an inverted trend for the DTM samples, with terrain elevation estimates being slightly overestimated by GEDI, with an overall positive bias of 3.41 m.

Plotting the relative estimation errors between the RH99 and DTM statistics reveals a significant linear trend in Fig. 7.5 (c), with the DTM underestimation being linked to the CH overestimation and vice versa. The resulting relationship appears to be

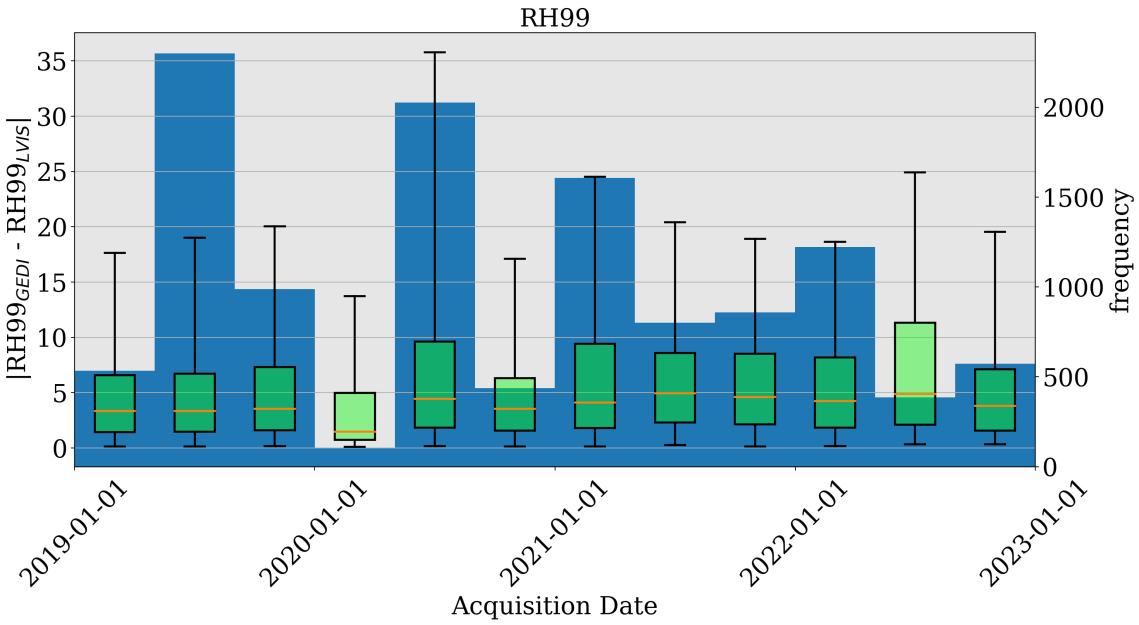


Figure 7.6: Performance boxplots between the GEDI- and LVIS-estimated RH99 metrics based on the acquisition date of the GEDI samples.

preferentially skewed towards an underestimation of the canopy height (CH). Given that the CH metrics are estimated from the full-waveform signal by detecting the ground and top-of-the-canopy returns [74], it is hypothesized at this stage that the issue consists in an erroneous detection of the ground return, as the likelihood of the transmitted signal to reach the true ground and to be reflected back to the sensor could fall below that of the noise floor for large and dense canopies. It is possible to confirm this by comparing the DSM measured using GEDI with that of LVIS, as the overall elevation measurement at the top of the canopy is independent from the probability of reaching the ground, and should only be affected by geo-location errors or by changes that occurred in the scene between the compared measurements. Indeed, Fig. 7.5 (d) displays a significant reduction in the estimation bias compared to the DTM, from 3.41 m to 1.21 m, respectively.

The role of the temporal distance between the GEDI and the LVIS samples is addressed in Fig. 7.6. This shows the absolute height error distribution of the prediction based on the acquisition date of the GEDI samples, with more contemporary measurements being expected to potentially result in a worse agreement with the 2016-acquired LVIS measurements. Differently, the plot shows no clear trends, with the randomness of the considered GEDI samples dominating the absolute error variation.

Finally, the impact of the rasterization processes is represented in Fig. 7.7, which

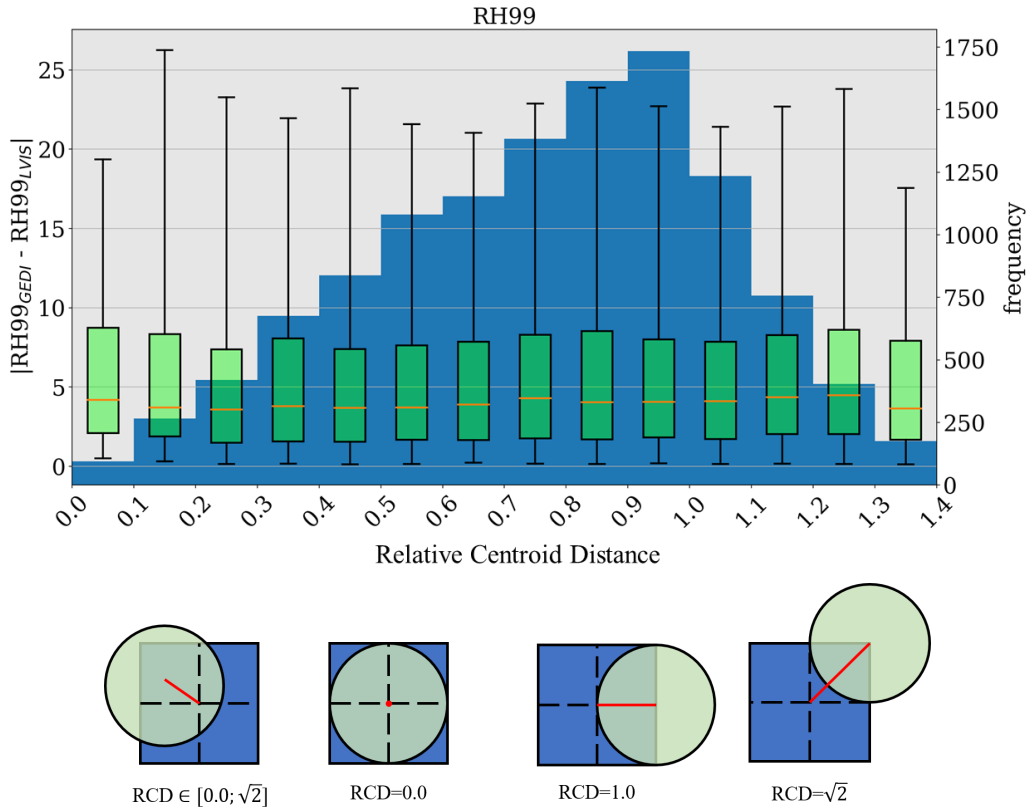


Figure 7.7: Performance boxplots between the GEDI- and LVIS-estimated RH99 metrics based on the relative centroid distance (RCD) between the GEDI samples and their associated pixel location. A schematic representations of possible configuration between a GEDI sample (circle), its associated pixel (square) and their RCD (red line) is shown for reference.

shows the impact of the relative centroid-distance between the pixel and the GEDI footprint centers. It is possible to observe a small linear increase in the median absolute error for larger distances, but the overall impact remains below that of the overall MAE.

The detected issue could thus be traced back to the insufficient SNR of the GEDI spaceborne sensor to consistently reach the ground in presence of tall and dense canopies, as the errors become even more evident after reintroducing the previously discarded coverage-beams, characterized by a lower per-beam transmission energy with respect to the full-power ones. In this last case, the metrics for the RH99 statistic see the bias increase to -2.00 m, the MAE to 6.15 m and the RMSE to 8.75 m. Furthermore, the use of simple threshold-based signal detection algorithms, presumably, also limits the chances of detecting the ground returns in low SNR regimes.

7.3.2 The Impact on the Proposed Framework

It is also possible to directly assess the impact of the GEDI inconsistencies on the proposed data-driven approach. To this end a simple factorial experiment is proposed, consisting of two possible training solutions as well as two possible test sets, for a total of four scenarios.

The first training solution corresponds to the model proposed in Sec. 6.2.1 and is completely trained on the LVIS dataset, while the second one corresponds to a fine-tuning of the former using the sparse GEDI samples. Similarly, the first test set corresponds to the LVIS samples acquired for the AfriSAR campaign, while the second test set is composed of the GEDI footprints sampled across the rest of Gabon.

The four resulting scatterplots are shown in Fig. 7.8. It is possible to note that the almost linear relationship obtained after pre-training on the AfriSAR dataset is completely lost after a fine-tuning step using the GEDI samples. Also, neither model achieves any reasonable agreement with the GEDI test set, with both saturating at a maximum canopy height of about 40 m.

7.3.3 Discussion

In this section the quality of the GEDI-derived canopy height estimates of the L2A products have been assessed. Unfortunately, the comparison results with the LVIS-derived CHM measurements, reported in Tab. 7.2, shows that the estimates for the two sources significantly diverge. Furthermore, it is possible to note that by re-combining the DTM and RH100 estimates from both sensors to produce DSM estimates, the performance noticeably improves. This suggests that a major cause of the performance disagreement lies in the inability of GEDI to reliably reach the ground in the presence of very dense and tall canopies. The authors in [75] and [148] report a similar behaviour, observed in their study cases over the tropical forests of French Guiana and Gabon, respectively, with both GEDI full power and coverage beams frequently underestimating tall vegetation when compared to LVIS measurements. Similarly, in [152] the authors detect biases also over mixed broadleaf and needleleaf forests in Germany, reporting performance improvements for broadleafs during the leaf-off season but declines for the needleleafs.

Finally, the impact of these inconsistencies on the proposed data-driven approach has also been empirically verified in Sec. 7.3.2.

These inconsistencies in the height estimates make the current L2A GEDI products ill-suited as a reference training source for data-driven approaches in the considered

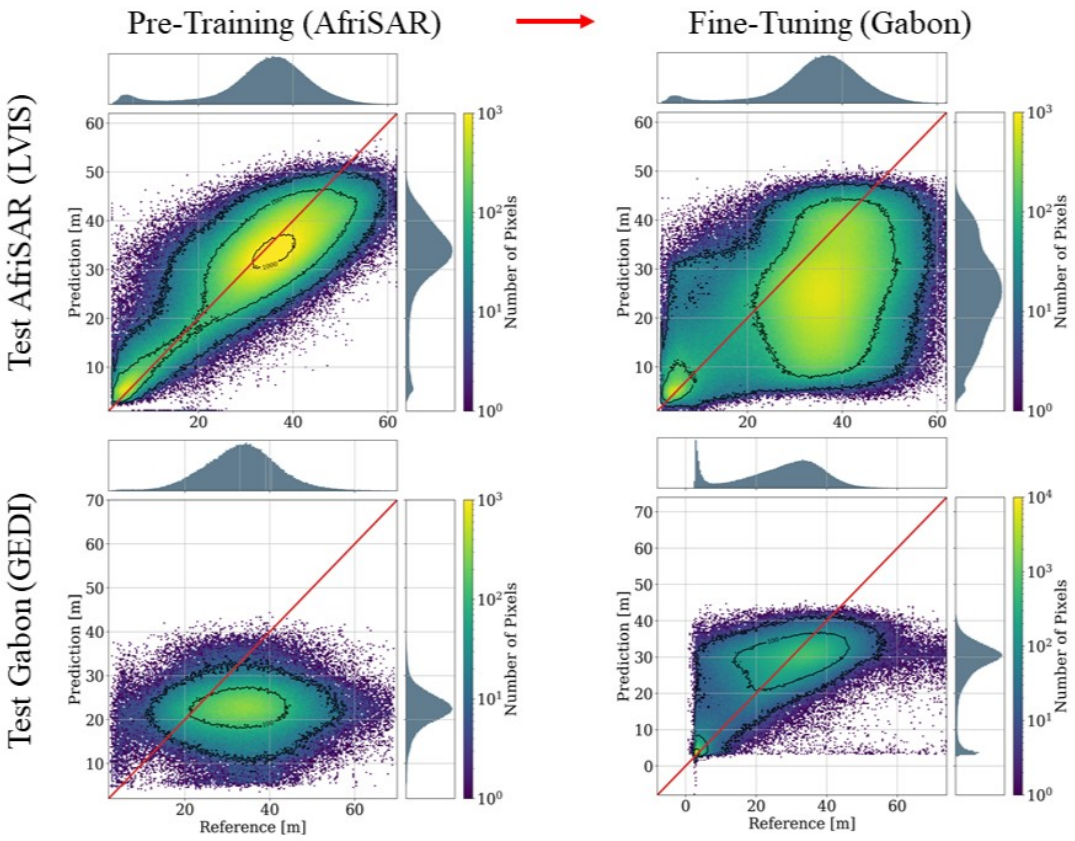


Figure 7.8: Factorial experiment assessing the impact of GEDI measurements on the proposed data-driven approach. The left column shows the result obtained after pre-training the model using the AfriSAR LVIS data, while the right one shows those obtained after fine-tuning on the GEDI measurements. The first row shows the test results evaluated on the AfriSAR LVIS dataset, while the second row shows those obtained from the GEDI measurements.

tropical scenario, as well as a reference height estimate against which the model performance could be tested. These findings motivate the choice to limit further analyses to comparisons with the LVIS measurements only.

7.4 Research asset 2: modification of the predictors and the training set

In this section, an alternative strategy to the generalization problem of the reference dataset discussed in Sec. 7.2.3 is presented, concentrating on the issues of missing representations of both the height of ambiguity and the DEM in Sec. 7.4.1 and Sec. 7.4.2, respectively.

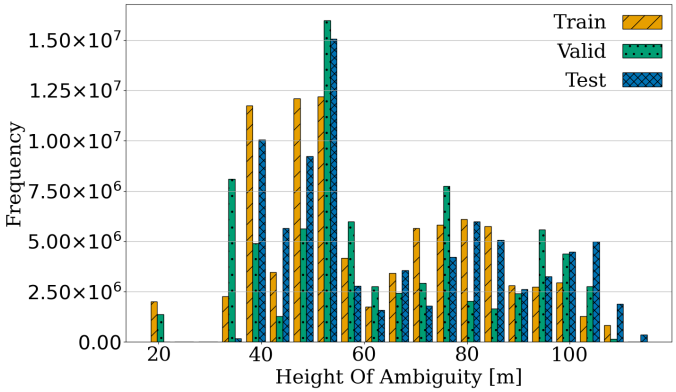


Figure 7.9: Height of ambiguity histogram for the extended training, validation and testing dataset, comprising TanDEM-X acquisitions covering the AfriSAR test sites from December 2010 up to the end of 2021.

7.4.1 Height of Ambiguity Analysis

In order to tackle the challenge of missing representation in the input heights of ambiguity, the relaxation of the temporal stationarity constraint is proposed. This can be achieved by considering TanDEM-X products covering the AfriSAR test sites and acquired over a time span of about 11 years, starting from the end of 2010 (i.e., the beginning of the bistatic TanDEM-X mission) up to 2021. This allows for the generation of a complete dataset characterized by the distribution of h_{amb} presented in Fig. 7.9. The distribution of the h_{amb} used for the country-scale inference depicted in Fig. 7.4 (b) is much better represented with respect to the initial 2015/16 case (Sec. 7.2). By considering multiple acquisitions the model is also able to learn a more robust relationship between the acquisition conditions of the input imagery and the reference canopy height. This assumes that the temporal misalignment between the input and the reference data results only in minor forest height inconsistencies due to natural phenomena, such as growth and tree replacement. These can be characterized as an additive noise contribution, which in turn can be interpreted as a data augmentation process. Conversely, drastic logging, fire or afforestation events (if present) are assumed to be limited in scope and are considered as outliers whose effects are mitigated by the availability of a large number of training samples.

In order to be as consistent as possible with the LVIS reference data and with the settings of the previous experiment in Sec. 7.2, the model performance is tested only on TanDEM-X data acquired between 2015 and 2016. The resulting model performance, summarized in Tab. 7.3 (a) (ME of -0.54 m, MAE of 3.78 m, MAPE of 13.08%, RMSE

Extended Dataset Performance					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Experiment	[m]	[m]	[%]	[m]	[·]
Lopé	-0.01	3.88	10.68	5.02	0.47
Mabounié	0.02	4.56	15.49	5.91	0.44
Mondah	-0.15	2.15	22.81	3.12	0.91
Pongara	-0.90	2.84	14.91	4.00	0.93
Rabi	-1.29	3.81	12.36	4.93	0.60
Overall	-0.54	3.78	13.08	4.98	0.78

(a)

Cross-validation Performance					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Experiment	[m]	[m]	[%]	[m]	[·]
Lope	-0.22	4.29	12.15	5.52	0.36
Mabounie	1.58	4.86	17.19	6.31	0.36
Mondah	0.87	2.19	26.26	3.17	0.91
Pongara	-1.38	4.95	26.64	6.37	0.81
Rabi	-0.80	3.84	12.80	4.97	0.59
Overall	-0.18	4.20	15.43	5.49	0.73

(b)

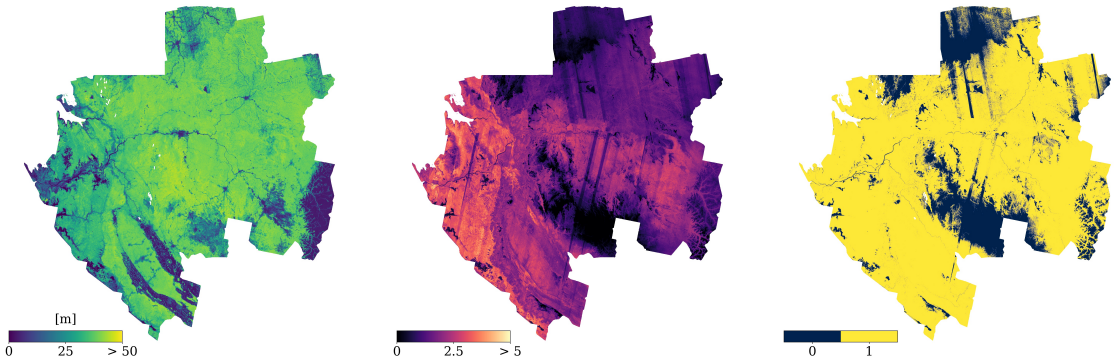
Table 7.3: (a) Performance metrics computed for the model trained using the extended set of TanDEM-X data acquired in between 2010 and 2021, for each AfriSAR test site, separately, and overall. (b) Performance metrics computed for the corresponding cross-validation experiment, for each AfriSAR test site permutation, separately, and overall.

of 4.98 m and R^2 of 0.78), shows an overall improvement with respect to the initial baseline scenario. The cross-validation results are summarized in Tab. 7.3 (b) and again confirm the capability of the model to generalize to completely independent areas.

The updated country-scale CH mosaic, the corresponding reliability score map and the MoA are depicted in Fig. 7.10. The MoA now presents significantly less null values with respect to the initial model presented in Fig. 7.2 (b), while the remaining critical areas can be linked to local topographic effects only. Moreover, the areas of severe CH underestimation shown in the CH mosaic of Fig. 7.2 (a) are not present anymore in the new mosaic of Fig. 7.10 (a).

7.4.2 DEM Analysis

When considering the issue of missing DEM representations in the training set it is not possible to follow the same strategy proposed for the height of ambiguity in Sec. 7.4.1, since the inclusion of new acquisitions over the same initial regions of interest would not allow the model to see a larger variety of topographies during training. In particular, when comparing the histograms in Fig. 7.4 (c) and (d), it is possible to observe that only elevations up to about 400 m are well represented within the AfriSAR test sites. At



(a) (b) (c)

Figure 7.10: (a) Country-scale CHM generated with the model trained using the extended dataset comprising TanDEM-X from 2010 up to 2021 (Sec. 7.4.1), (b) corresponding reliability score map and (c) binary MoA.

inference, this results in a majority of the elevation samples being poorly represented during the training phase.

To address this issue, it is possible to either completely remove the DEM as predictor or to substitute it with some proxy variables which describe only local topographic variations instead of the absolute elevation of the scene. Regarding the former solution, some performance will inevitably be lost with respect to the results presented in Sec. 7.4.1 and traded in favor of a more robust model. Regarding the second solution, the use of two different DEM-derived variables is investigated: the estimates for the set of spatial partial derivatives ∇h_{DEM}, which correspond to an estimation of the local terrain slope, and a high-pass filtered version of the DEM $h_{\mathrm{DEM}}^{\mathrm{HPF}}$, which removes the mean elevation of the TanDEM-X scene, highlighting only local high-frequency variations of the topography. One should note that ∇h_{DEM} is computed as:

$$\nabla h_{\mathrm{DEM}}(x,y) = \left(\frac{\partial h_{\mathrm{DEM}}(x,y)}{\partial x}, \frac{\partial h_{\mathrm{DEM}}(x,y)}{\partial y} \right) \tag{7.3}$$

where x and y are the horizontal and vertical coordinates, respectively. This leads to the use of two separate input features to the network.

By considering the same extended dataset presented in Sec. 7.4.1 for training (TanDEM-X acquisitions from 2010 up to 2021), the performance results for all listed test cases are summarized in Tab. 7.4, together with the performance of the model derived from the height of ambiguity analysis in Sec. 7.4.1 for comparison purposes (**Baseline** case). As expected, the complete removal of the DEM from the set of predictors (**w/o DEM** case) causes an overall loss in performance, which is partly mitigated by the

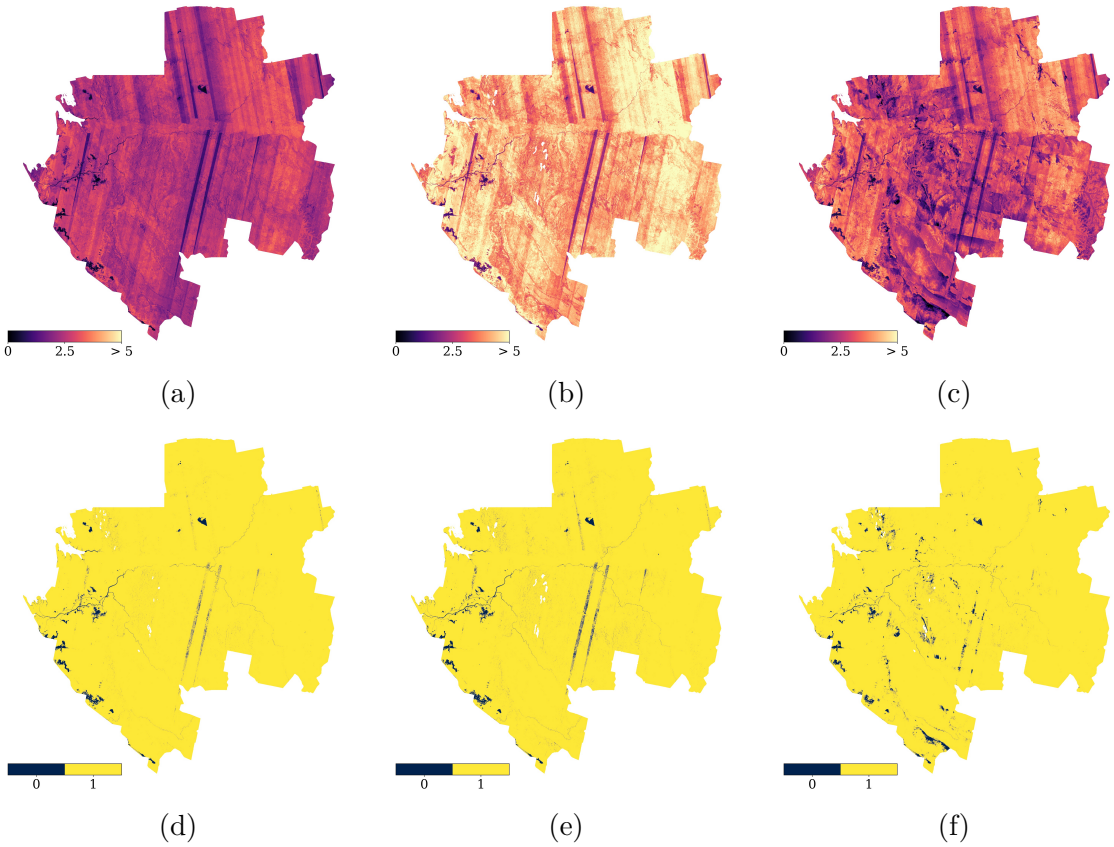


Figure 7.11: (Top row) Reliability score maps for the three different DEM-related solutions: (a) removal of the DEM as predictor, (b) substitution of the DEM with its spatial derivatives (local slope), (c) substitution of the DEM with its high-pass filtered version. (Bottom row) Corresponding MoAs.

use of the DEM spatial derivatives (∇ case) or the high-pass filtered version (**HPF** case), with the former achieving the overall peak performance (ME of 0.12 m, MAE of 3.90 m, MAPE of 15.38%, RMSE of 5.08 m and R^2 of 0.77). On the other hand, the robustness of model for country-scale inference significantly improves, as it can be seen from the MoAs for all three considered cases in Fig. 7.11. Each of the MoAs shows an almost wall-to-wall coverage of valid values, with the ∇ case (Fig. 7.11 (b) and (e)) being characterized by overall higher values in terms of reliability score. The majority of the remaining invalids is located in correspondence of water surfaces, which were not seen during training.

Final Country-scale Model and Inference

In light of the knowledge gained from the previous experiments, the final model for the generation of the country-scale CH map of Gabon relies on the extended training

Table 7.4: Performance metrics for the DEM analysis experiment. Each row identify the performance of the model derived from the extended dataset presented in Sec. 7.4.1 (**Baseline**), the model trained without the DEM as predictor (**w/o DEM**), the model trained with the DEM derivatives (∇) and the model trained with the high-pass filtered version of the DEM (**HPF**).

DEM Analysis Performance					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Experiment	[m]	[m]	[%]	[m]	[·]
Baseline	-0.54	3.78	13.08	4.98	0.78
w/o DEM	-0.58	4.13	15.93	5.35	0.75
∇	0.12	3.90	15.38	5.08	0.77
HPF	0.11	4.00	15.26	5.23	0.76

dataset presented in Sec. 7.4.1 and replaces the input DEM predictor with the estimate of its spatial derivatives, as proposed in Sec. 7.4.2.

The detailed performance metrics are reported in Tab. 7.5 (a) (ME of 0.12 m, MAE of 3.90 m, MAPE of 15.38%, RMSE of 5.08 m and R^2 of 0.77). Compared to the reference model presented in Sec. 7.4.1 only a minor degradation in performance can be observed and it is predominantly limited to a slight overestimation of very short vegetation, as can be seen in the scatterplot in Fig. 7.12 (a). The estimation bias (ME) with respect to different reference tree height sub-ranges is shown in Fig. 7.12 (b) together with the overall reference RH99 distribution of test samples. Notably, measurements are on average slightly overestimated for vegetation heights below 15 m, are unbiased between 30 m and 40 m, with a tendency to more strongly overestimate forest heights in the 15 m to 30 m range and to underestimate for values above 40 m. The results for the geographic cross-validation experiments are reported in Tab. 7.5 (b) and show a similar degree of spatial independence as seen for the previous results.

To further analyze the model's bias with respect to parameters characterizing both the illuminated areas and the acquisition geometry, the test dataset is extended to TanDEM-X acquisitions from 2010 up to 2021 overlapping the AfriSAR test sites and not used during training. In this way, an extensive representation of the analyzed parameters can be considered. Fig. 7.13 (a) displays the relative dependency of the estimation error on the acquisition Day Of the Year (DOY), highlighting a comparably unbiased relationship across the value range. Fig. 7.13 (b) relates the error to the local terrain slope, which is computed using the LVIS-derived DTM estimates. It is possible to note that the median absolute error shows almost no dependency on the local slope of the underlying topography. Additionally, the performance dependency on the TanDEM-X acquisition geometry is evaluated. It is possible to observe that for both the h_{amb} (Fig. 7.13 (c)) and the incidence angle ranges (Fig. 7.13 (d)) the median estimation

∇ Performance					
	ME	**MAE**	**MAPE**	**RMSE**	$\mathbf{R^2}$
Experiment	[m]	[m]	[%]	[m]	[·]
Lopé	-0.02	3.78	10.41	4.88	0.50
Mabounié	0.44	4.45	15.24	5.77	0.46
Mondah	1.98	2.95	39.69	4.00	0.85
Pongara	1.52	4.21	28.27	5.45	0.86
Rabi	-0.59	3.79	13.01	4.93	0.60
Overall	0.12	3.90	15.38	5.08	0.77

(a)

Cross-validation Performance					
	ME	**MAE**	**MAPE**	**RMSE**	$\mathbf{R^2}$
Experiment	[m]	[m]	[%]	[m]	[·]
Lopé	-3.82	5.27	13.36	6.65	0.07
Mabounié	1.50	4.65	16.19	6.07	0.41
Mondah	3.25	4.12	62.94	5.71	0.69
Pongara	-1.17	5.37	28.91	7.01	0.77
Rabi	-1.18	3.85	12.84	4.96	0.59
Overall	-1.43	4.64	17.69	6.03	0.68

(b)

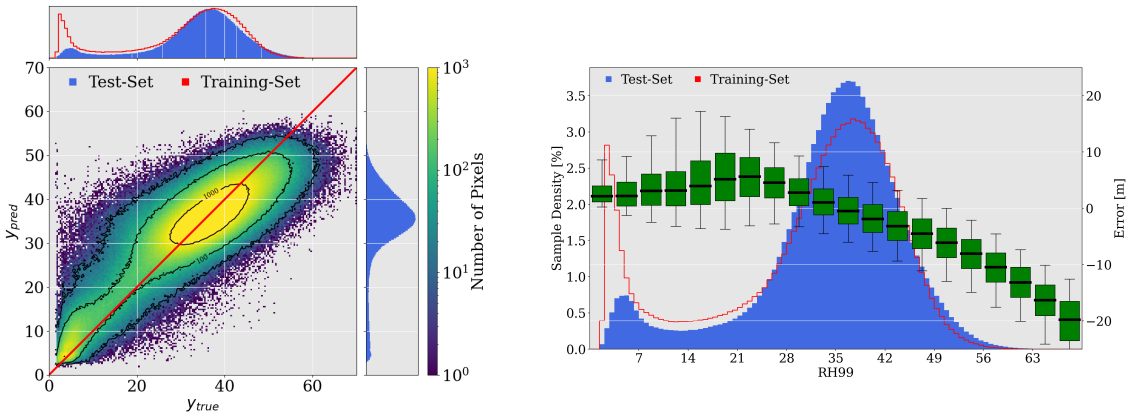
Table 7.5: (a) Performance metrics computed for the final model trained using the estimate of the spatial DEM gradient as a replacement for the DEM itself, shown for each AfriSAR test site separately and overall. (b) Performance metrics computed for the corresponding cross-validation experiment, shown for each AfriSAR test site permutation separately and overall.

errors suggest an essentially unbiased estimation.

Finally, the country-scale mosaic of the CHM over Gabon at 25 m resolution, inferred for the TanDEM-X acquisitions of 2010/2011 is depicted in Fig. 7.14a. Further details at a higher resolution, corresponding to the areas included in the red squares, are presented in Fig. 7.14b. The first zoom-in (left) shows the presence of primary tropical forest, a complex system of rivers and antropogenic activities. In the second one (center), it is possible to observe the presence of tall mangroves along the shores of the Gabon estuary, with peak canopy heights above 45 m. Finally, the third one (right) shows a further example of a dense mangrove forest along the coast.

7.4.3 Discussion

The results detailed in Sec. 7.4.1 and in Sec. 7.4.2 provide a complete overview on the challenges and solutions related to the country-scale application of the proposed method. In particular, moving from the confined study areas of the 2016 AfriSAR campaign to the country-scale inference poses the natural challenge of validating the final product, especially in the absence of complementary reference measurements.

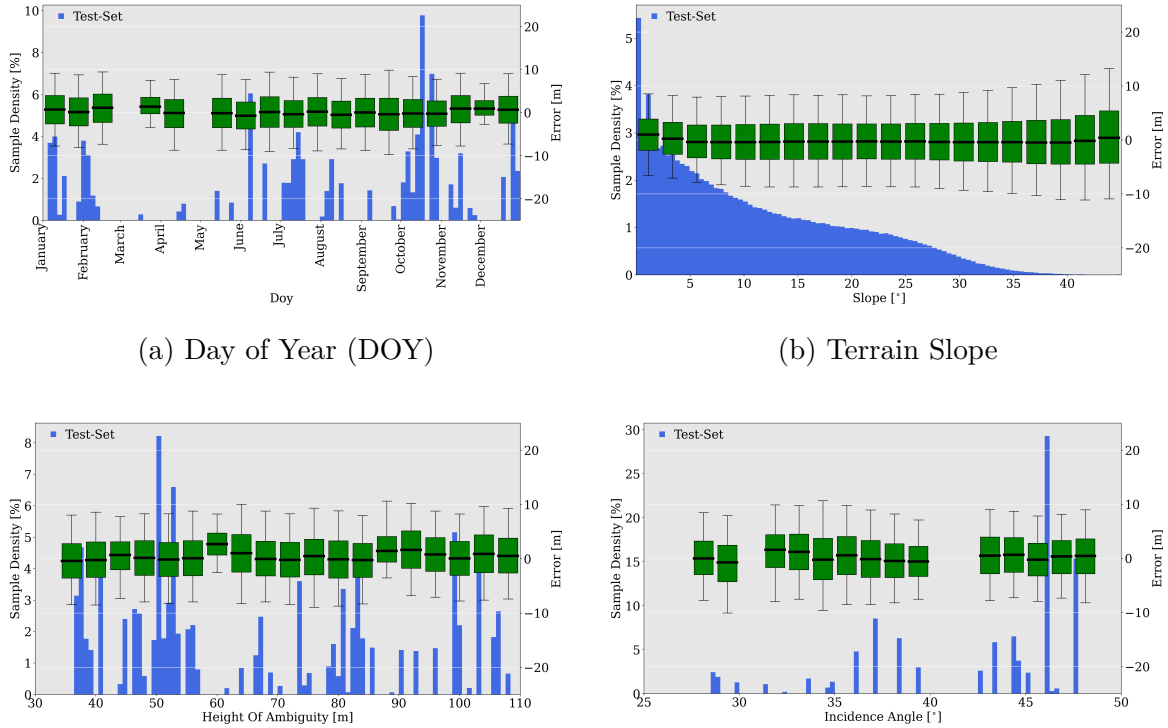


(a) Prediction Scatterplot

(b) Estimation Bias

Figure 7.12: Comparative estimation performance plots between the selected model for country-scale inference and the reference LVIS dataset. The scatterplot (a) displays the linear prediction agreement. The boxplot sequence (b) captures the estimation bias and spread for different reference tree height sub-ranges; the whisker contain 90% of the samples, the boxes 50%, while the black line represents the median value. The background histograms depict the relative samples distributions of the training (red) and test sets (blue), respectively.

Te proposed solution to the problem follows two different paths. On the one hand, the reliability of the achieved performance metrics was examined by verifying the absence of spatial correlation between the training and test sets. To do so, geographically-independent cross-validations were carried out, which highlighted consistently small deviations in performance with respect to the standard testing strategy. In particular, these are limited to the Lopé and Pongara test permutations, being the former characterized by high-relief terrain and the latter by the presence of tall mangroves. These observations allow us to confirm the soundness of the achieved model performance. On the other hand, the applicability of the model was assessed at country scale, where the estimates cannot be otherwise validated. Based on the assumption of a unique bijective relationship between predictors and forest height, the proposed approach is meant to identify those predictors positioned within the subspace sampled by the training dataset. Rather than providing a pixel-wise validation, this approach allows for assessing the trustworthiness of the CH estimates. This means that it is possible to identify whether the model accuracy falls within the performance boundaries defined during the test phase. Clearly, should the underlying assumption of a bijective relationship not hold anymore (i.e., by missing a necessary discriminative feature), also the MoA would fail to detect unreliable estimates.



(a) Day of Year (DOY)

(b) Terrain Slope

(c) Height of Ambiguity h_{amb}

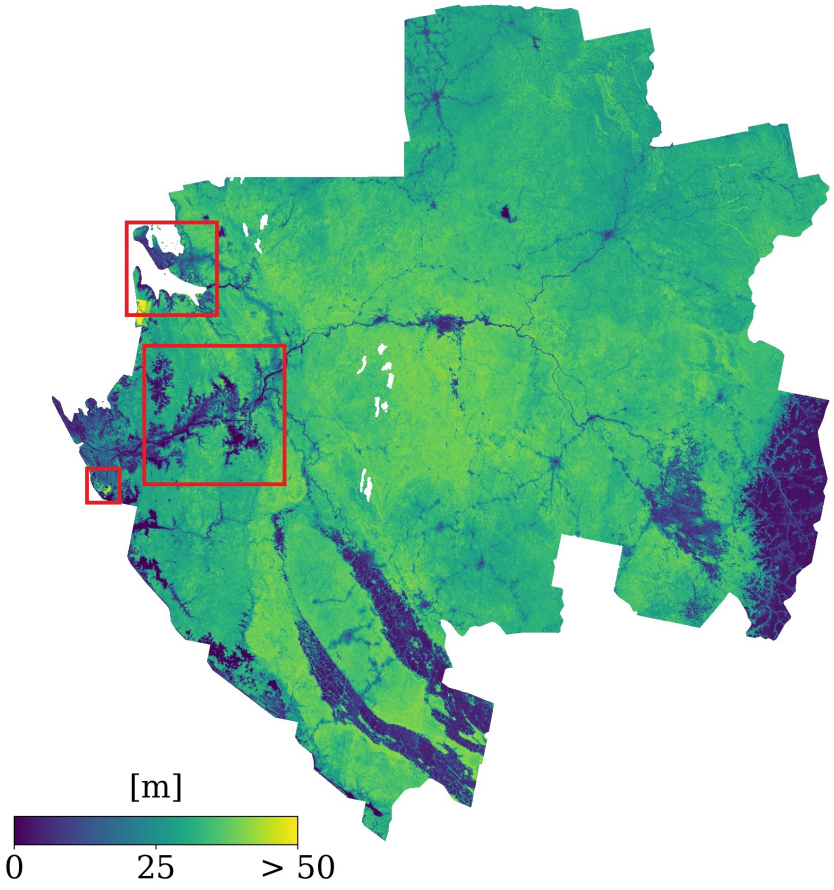
(d) Incidence Angle θ_{inc}

Figure 7.13: Estimation error bias and spread versus the acquisition DOY (a), the Terrain Slope (b), the h_{amb} (c) and the incidence angle (d) features, represented as a discrete sequences of boxplots. Each boxplot covers a feature sub-range, and is described by its whiskers (containing 90 percent of the samples), its box (containing 50 percent of the samples) and black median line. In the backgrounds, the respective feature distributions on the test set (in blue).

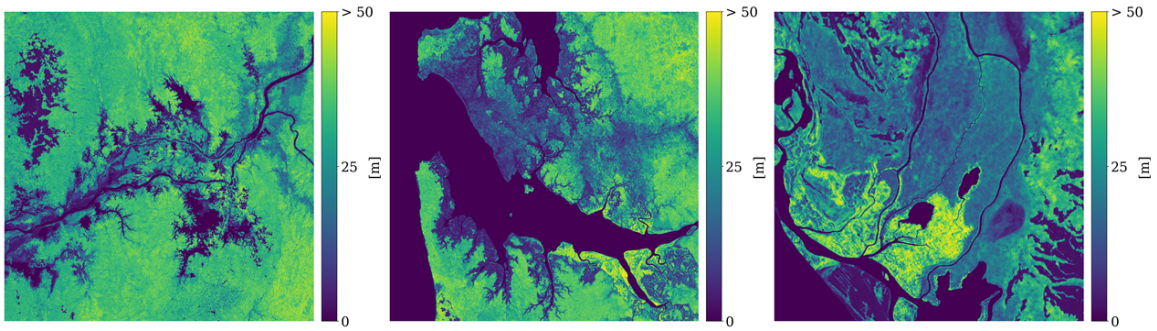
In practice, in the analyses in Sec. 7.2.2 and Sec. 7.4.1, the MoAs have allowed for detecting missing representations in the training data, which directly match with strongly underestimated forest heights. Furthermore, the proposed modifications of the training data set in Sec. 7.4.1 and the new input predictors defined in Sec. 7.4.2 have led to the definition of a robust model for country-scale inference.

The final model achieves overall state of the art performance, with significant underestimation bias only for canopy heights above 45 m. Beyond canopy heights of 55 m, the model tends to saturate, a behaviour often observed in the literature [35][153][38]. This effect might be linked to different aspects. On the one hand, it could be related to the disproportionately low frequency of high tree samples occurring in the training set or, on the other hand, to the limited capability of radar waves at X-band to penetrate into dense forests.

Regarding the fact that the performance remains stable with respect to the DOY,



(a)



(b)

Figure 7.14: (a) Country-scale mosaic of Gabon representing the CHM, generated using TanDEM-X acquisitions from the first global covered of the mission (Dec. 2010 - end of 2011). (b) Zoom-ins of the three regions included in the red boxes of the country-scale CHM mosaic. Invalid values, caused by either shadow and layover or by the unreliability of the model, are depicted in white.

it is reasonable to assume that this remains valid for tropical forests only. For example, it can be expected that temperate forests are characterized by more complex changes throughout the year, possibly requiring additional input information to the model, such as the DOY or the time of acquisition, to remain unbiased.

Similarly, the unbiased estimation with respect to the height of ambiguity and the local incidence angle is extremely relevant for large-scale applications using single-pass InSAR, as it suggests that the model is capable of delivering spatially consistent estimates, independently of the SAR and InSAR geometries.

A performance comparison of the proposed method with respect to the state of the art in the literature is also of interest. Compared to the preliminary work published in [140], the newly presented approach yields an overall improvement across all considered metrics. The total bias has improved by 1.36 m, the MAE by 0.30 m, and the RSME by 0.61 m. When comparing the methodology to the physical-based models, the RVoG model represents the most investigated approach [78][79][83][81][29]. Given the considered predictors, the performance is compared with the inverted sinc-approximation of the RVoG model [29], which also only requires information about the acquisition geometry (i.e., θ_{inc} and h_{amb}) and the volumetric decorrelation coefficient γ_{vol} from a single-pol, single-baseline acquisition. The RVoG achieves an overall ME of -2.24 m, a MAE of 8.60 m and an RMSE of 10.85 m. The RVoG inversion scheme in [83], which relies on a combination of TanDEM-X imagery and LiDAR profiles, achieved an RMSE of 8.16 m and an r^2 value of 0.16 over the site of Lopé. For comparison, the proposed DL approach achieves an RMSE of 4.88 m and an r^2 of 0.50. In [82], the RVoG is inverted using multi-baseline, quad-pol acquisitions and selecting the optimal baseline using a support vector machine (SVM) trained on sparse LiDAR measurements. The experiments led to an RMSE of 5.64 m over Pongara, of 4.99 m over Mondah, and of 5.99 m over Lopé, respectively. Using the approach proposed in this work, it is possible to achieve an RMSE of 5.45 m, 4.00 m and 4.88 m, respectively. Finally, in [35] the authors propose a deep learning approach, which estimates the CHM values from Sentinel-2 multi-spectral data. The analyses over the AfriSAR Campaign test sites achieve a MAE of 4.9 m and an RMSE of 6.5 m when considering the yearly least cloudy acquisitions, and a MAE of 4.3 m and an RMSE of 5.6 m when applying a temporal median filter across a one year inference stack. In this context, the DL framework presented in this work achieves extremely competitive results, with the great advantage of requiring only a single TanDEM-X acquisition as input.

In light of these results, the method shows a significant potential for generating multi-temporal, time-tagged products and for monitoring forest height changes in time.

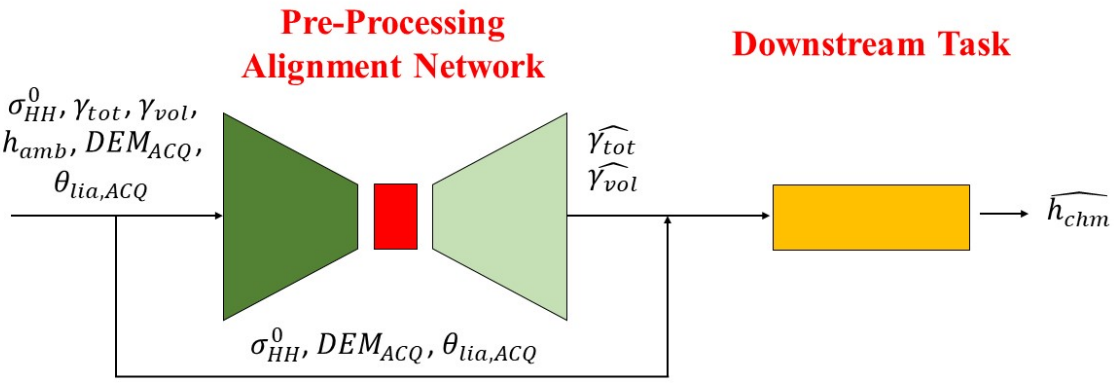


Figure 7.15: Schematic representation of the proposed HAM workflow. The HAM acts as a pre-processing step to a generic downstream task, harmonizing the TanDEM-X single-pass γ_{tot} and γ_{vol} features. This is done to mitigate the effects of missing InSAR geometry representations during the training of the downstream task.

On the one hand, clear cuts and afforestation can be easily identified, since these represent abrupt changes. On the other hand, the challenge is to monitor forests dynamics, whose variations lie within the current uncertainty boundaries of the model. To this aim, further validation is required to assess the reliability of the derived model with respect to reference data acquired at different times.

7.5 Research asset 3: Height of Ambiguity Alignment Model

In this section, a third possible solution to the generalization problem of the reference dataset discussed in Sec. 7.2.3 is presented. It consists in the introduction of a pre-processing step to the regression problem, for the harmonization of the TanDEM-X products with respect to their acquisition geometry. In this way, performance losses caused by unseen geometries during training are expected to be significantly mitigated, leading to a more robust model. The proposed concept involves the definition of an ad-hoc developed neural network for the alignment of the TanDEM-X InSAR features of a give acquisition to those acquired using a reference InSAR geometry. This reference geometry is defined in terms of height of ambiguity $h_{\text{amb,ref}}$, leading to the creation of the proposed *Height of ambiguity Alignment Model* (HAM). The complete workflow of the HAM, including its contextual positioning with respect to a generic downstream task, can be seen in Fig. 7.15.

The following subsections deal with the definition of a training and testing strategy for the proposed HAM model, as well as offering an analysis of the impact of the aligned features on the forest height regression problem, core to this thesis.

7.5.1 Pre-Processing Task

As already mentioned, the focus of this ablation study lies in the definition of a neural network-based architecture, which is tasked with converting the InSAR features of a TanDEM-X product, acquired with a certain InSAR geometry, to the corresponding representation given by a reference InSAR geometry.

This first analysis focuses on the InSAR features used for the forest height regression framework defined in Ch. 6, consisting of the interferometric coherence γ_{tot}, the volume decorrelation factor γ_{vol} and the ancillary height of ambiguity h_{amb} information.

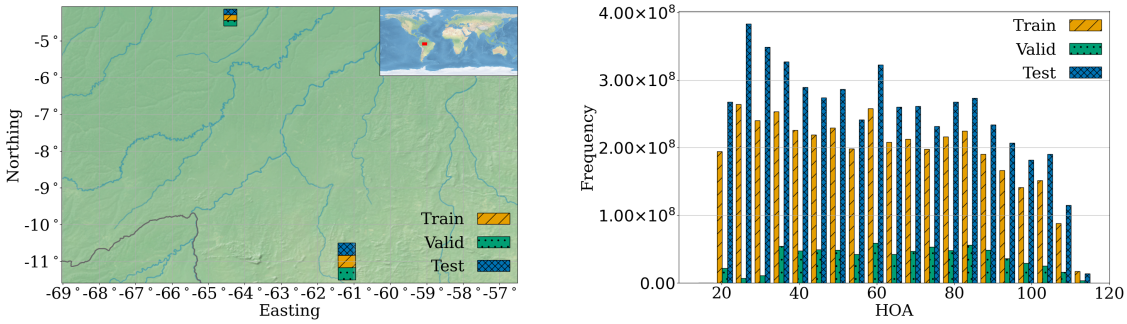
Model, Dataset and Training Strategy

The alignment task is investigated at a resolution of 6 m, specifically selected in order to be compatible with the operative features resulting from the TAXI processor, discussed in Sec. 5.1.3.

At this stage, no particular considerations are made for the definition of the model architecture, and the plain convolutional network defined for the regression task in Sec. 6.2.1 is used for this investigation as well. Similarly, the same six predictors, previously selected for the forest height regression framework in Sec. 7.2.1, are also considered; these consisting in:

- Backscattering coefficient σ^0_{HH}, in HH polarization.
- TanDEM-X raw acquisition DEM.
- Local incidence angle θ_{inc}.
- Total interferometric coherence γ_{tot}, estimated using Φ-Net [115].
- Volume decorrelation factor estimate γ_{vol}.
- Height of ambiguity h_{amb}.

The proposed model is trained to simultaneously generate the corrected estimates for both γ_{tot} and γ_{vol}.



(a) (b)

Figure 7.16: (a) Location and (b) height of ambiguity distribution for the training, validation and test sets used for the proposed HAM model.

In order to learn the alignment problem, a dataset collection consisting of acquisitions with a large variety of different geometries for the same illuminated scene is necessary. To this end, two TanDEM-X long time-series acquired over the Amazon forest are selected, comprising more than 10 years of regular bistatic acquisitions with variable geometries. Their intended use is that of monitoring the calibration of the TanDEM-X backscatter in time and they have been selected to cover two stable regions of primary tropical forest. This aspect is extremely favourable for the proposed analysis, as minimal changes in the scene can be expected between the different acquisitions, such that all changes in the InSAR features can be reduced to changes in the InSAR acquisition geometry. Moreover, the use of acquisitions over the Amazon rainforest was made necessary by the unavailability of similar time-series over Gabon. Consequently, for this proof-of-concept it is assumed that no significant differences in the EM response between Brazilian and African tropical forests exist. Obviously, this assumption would not hold anymore if very different types of forests were considered.

A geographic subsetting is used to split each of the two considered regions of interest into three sub-regions: the first one assigned to training, the second to validation and the third to testing. The resulting geographic locations of these subsets can be viewed in Fig. 7.16 (a). It is possible to observe from the corresponding h_{amb} distributions shown in Fig. 7.16 (b), that all three subsets are similarly distributed in their representation of the considered $h_{\mathrm{amb}} \in [20; 115]\,\mathrm{m}$.

The following step consists in the definition of the reference height of ambiguity value, which for this first analysis is chosen to be $h_{\mathrm{amb,ref}} = 80\,\mathrm{m}$ as it surpasses the tallest considered vegetation in both the current pre-processing task and in the downstream regression task.

Table 7.6: The performance results for the proposed HAM experiment.

HAM Performance Metrics					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
γ_{tot}	-0.007	0.064	10.17	0.083	0.41
γ_{vol}	-0.008	0.061	8.57	0.079	0.51

The creation of the final dataset then starts with the selection of the reference or target representations of γ_{tot} and γ_{vol}, chosen to have a nominal acquisition h_{amb} that falls within the $h_{amb,ref} = 80 \pm 0.5\,m$ neighbourhood. The resulting target reference acquisitions are subsequently used to build input/output image pairs with all available acquisitions, including the reference acquisitions themselves, since the model must also be trained to bypass products which were already acquired using the target geometry.

The patches extracted from the resulting image pairs are subsequently used to train and validate the HAM-model using the same minibatch strategy detailed in Sec. 6.1.3. In this case, the loss function for the k^{th}-batch is expressed by the following three-term function:

$$\mathcal{L}_k = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\gamma}_{tot,i} - \gamma_{tot,i}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\gamma}_{vol,i} - \gamma_{vol,i}\right)^2 + \lambda \cdot \sum_{j=1}^{m} w_j^2, \tag{7.4}$$

where $\hat{\gamma}_{tot,i}$ and $\hat{\gamma}_{vol,i}$ are the γ_{tot}- and γ_{vol}-estimates, $\gamma_{tot,i}$ and $\gamma_{vol,i}$ are the expected γ_{tot} and γ_{vol} values, w_j are the trainable model weights, and $\lambda = 10^{-4}$ weights the impact of the $l2$-regularization term.

Results

Fig. 7.18 shows the overall alignment results for both γ_{tot} and γ_{vol}, which in both cases see a good agreement between a significant portion of the prediction and the reference values. Nevertheless, the relationship is also characterized by a significant spread for the remaining samples. It is possible to observe that for γ_{tot} (Fig. 7.18 (a)) the estimates fully cover the expected [0;1] range, while for γ_{vol} (Fig. 7.18 (b)) the lower boundary of the predictions saturates at a value of $\gamma_{vol} \approx 0.2$. The reason for this behaviour is not immediately clear and might be related to inconsistencies for the smaller values in the reference sample γ_{vol} that could lead to the aforementioned saturation effects. On the other hand, it can be observed that the bimodal test distribution (horizontal gray histogram in Fig. 7.18 (b)) is correctly regressed by the model (vertical gray histogram in Fig. 7.18 (b)), despite having apparently been trained on an different unimodal distribution (horizontal red histogram in Fig. 7.18 (b)).

When looking at the performance metrics reported in Tab. 7.6 one can observe that

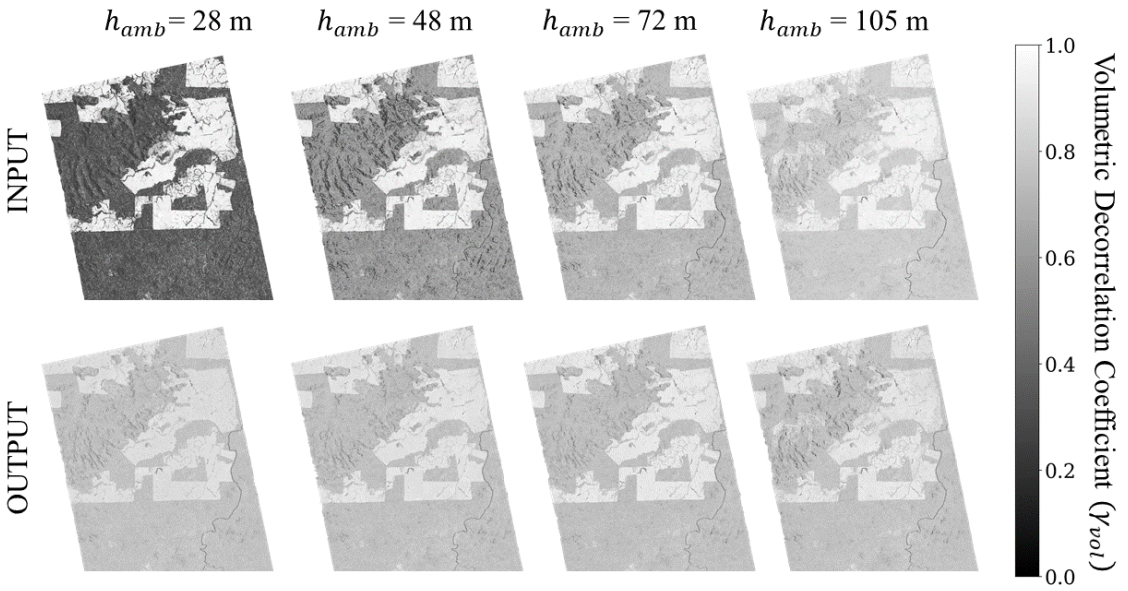
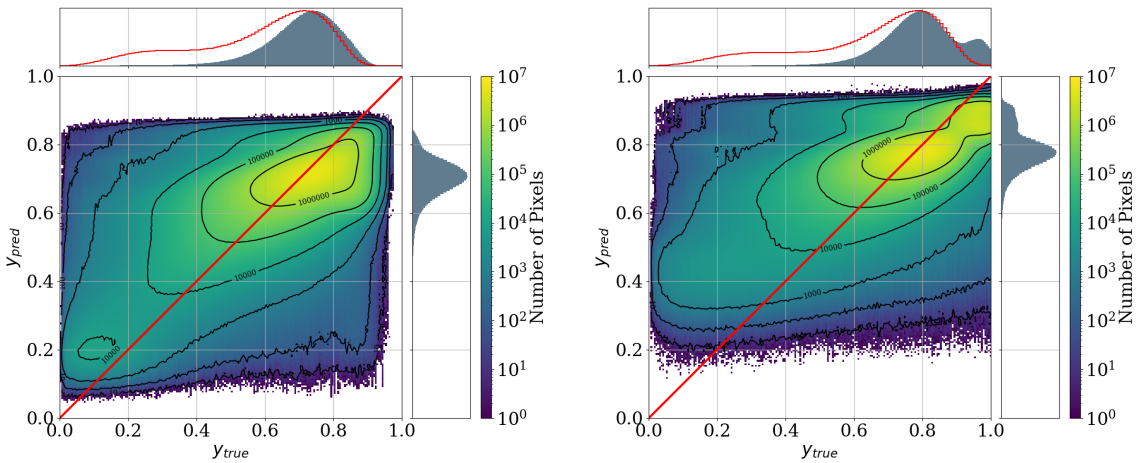


Figure 7.17: Alignment examples for four acquisitions with h_{amb}, from left to right, of 28 m, 48 m, 72 m and 105 m. The model is expected to re-project the input γ_{vol} (top) to the that acquired with a target $h_{\text{amb,ref}} = 80$ m (bottom). As can be observed, all predictions converge to a very similar representation.



(a) γ_{tot} (b) γ_{vol}

Figure 7.18: The scatterplots for the HAM test results over the Amazon Rainforest. In particualr, (a) shows the result for interferometric coherence γ_{tot}, and (b) shows that for the volumetric decorrelation factor γ_{vol}.

the γ_{vol}-predictions perform slightly better than the γ_{tot}-ones, despite the reported saturation effects. In particular, the γ_{tot} and γ_{vol} estimates achieve a low bias of -0.007 and -0.008, a MAE of 0.064 and of 0.061, and a MAPE of 10.17% and 8.57%, respectively. Finally, Fig. 7.17 displays four qualitative examples of the alignment task for different input baselines, yielding very similar target representations.

7.5.2 Downstream Task

The theoretical alignment results obtained for the HAM model already constitute a promising first insight into the potential of the proposed approach, but in order to fully determine its practical usefulness, the impact on the downstream regression task has also to be evaluated. In the following section the HAM is inserted into the existing dataset pre-processing pipeline in order to generate the two additional aligned γ_{tot} and γ_{vol} features, as presented in Fig. 7.15.

The downstream task itself follows the dataset description and the training, validation and testing strategies reported in Sec. 7.2.1. The only changes concern the input set composition, which sees the ordinary acquisition γ_{tot} and γ_{vol} features being replaced by those pre-processed using the HAM. Consequently, the h_{amb} feature becomes unnecessary and it is removed altogether, leading to a total of only 5 predictors.

Results

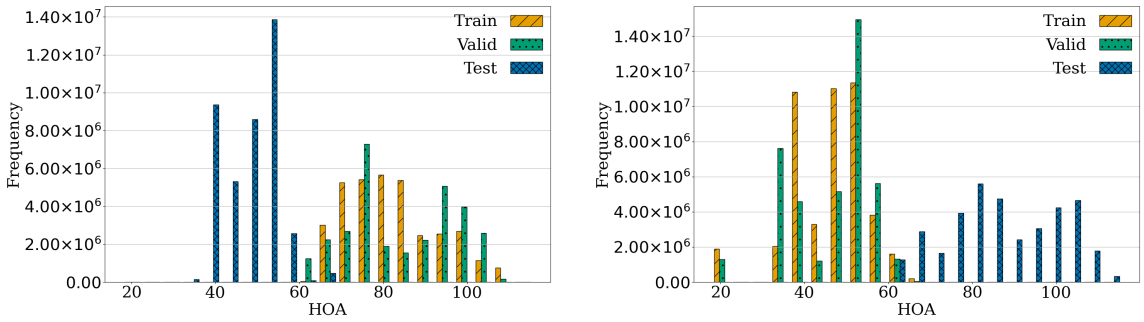
In order to evaluate the potential benefits of the HAM pre-processing, all experiments are compared against the reference performance of the country-scale framework defined in Sec. 7.4.

The first evaluation consists in assessing how much the performance of a controlled scenario (defined as the **baseline** case) diverges between the original and the HAM frameworks. The results in Tab. 7.7 report that the HAM-framework suffers a minor reduction in performance over the reference one, with the MAE increasing from 3.87 m to 4.06 m, the MAPE from 13.31% to 14.31%, and the R^2 reducing from 0.75 to 0.73. Removing the h_{amb} from the input set of the reference framework (i.e. **baseline w/o** h_{amb} case) for a consistent feature comparison sees the gap reduce even further. In the context of the previous analyses presented in Sec. 7.4.2, the resulting changes are extremely limited in scope, suggesting that in a best case scenario the impact of the HAM is remarkably neutral.

In order to test the potential benefits of the HAM procedure in a more real-case scenario, characterized by the absence of certain geometries during training, the nominal

Table 7.7: Performance results for the downstream experiments to the proposed HAM solution. The metrics are presented both with and without the HAM pre-processing for better comparison.

Downstream Task Performance Metrics					
	ME	**MAE**	**MAPE**	**RMSE**	**R^2**
Experiment	[m]	[m]	[%]	[m]	[·]
Reference					
Baseline	-0.32	3.87	13.31	5.09	0.75
Baseline w/o h_{amb}	-0.16	3.90	13.52	5.16	0.74
Split	2.34	7.37	24.61	9.32	0.15
Inverse Split	-18.38	19.37	54.04	23.66	-4.40
Split w/o h_{amb}	1.72	5.45	20.46	6.97	0.52
Inv.-Split w/o h_{amb}	-2.97	5.47	16.94	7.22	0.51
HAM					
Baseline	-0.11	4.06	14.31	5.33	0.73
Split	-0.47	4.60	16.09	6.10	0.64
Inverse Split	-0.05	4.35	15.11	5.67	0.70
Split w/o γ_{vol}	0.15	4.36	15.64	5.78	0.67
Inverse Split w/o γ_{vol}	-0.09	4.34	15.31	5.67	0.70
Split w/o γ_{tot}	-0.22	4.44	15.62	5.89	0.66
Inverse Split w/o γ_{tot}	-0.16	4.38	15.30	5.72	0.69



(a) Split (b) Inverse Split

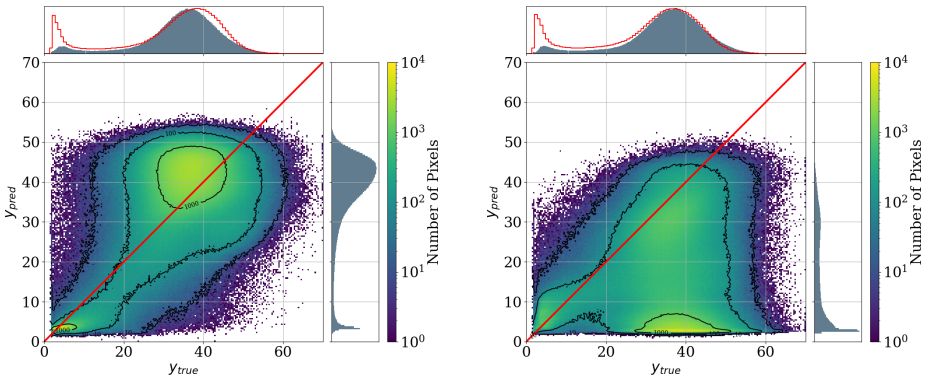
Figure 7.19: The h_{amb}-distributions for the (a) **Split** and (b) **Inverse Split** cases. For each case, the histograms for training, validation and testing are presented separately.

acquisition range of $h_{\text{amb}} \in [20; 110]$ is divided into two sub-ranges. This results in the following two cases:

- a **split** scenario, where the model is trained and validated on $h_{\text{amb}}^{tr,va} \in [65; 110]$ and tested on $h_{\text{amb}}^{te} \in [20; 65]$,
- an **inverted split** scenario, where the model is trained and validated on $h_{\text{amb}}^{tr,va} \in [20; 65]$ and tested on $h_{\text{amb}}^{te} \in [65; 110]$.

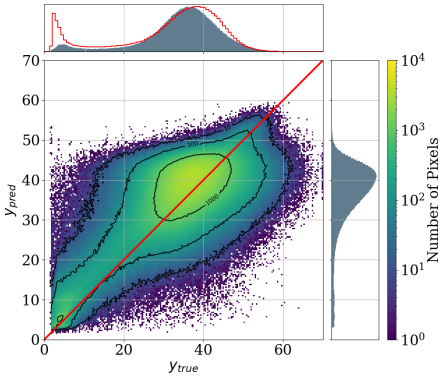
The corresponding h_{amb} distributions are presented in Fig. 7.19.

Starting from the results for the reference framework, for both the **split** and the **inverted split** scenarios, the metrics reported in Tab. 7.7 and the scatterplots shown

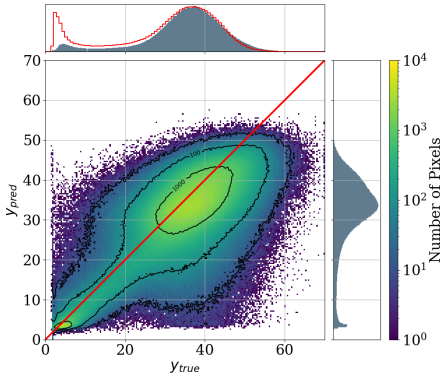


(a) Reference (Split)

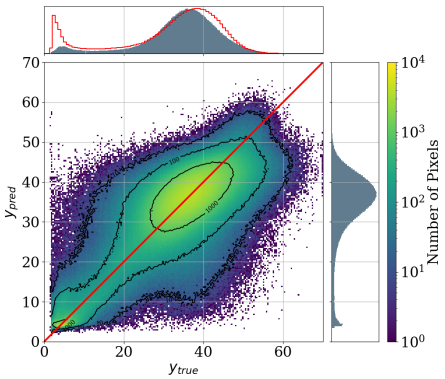
(b) Reference (Inverse)



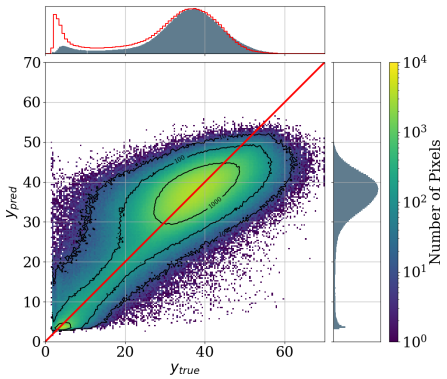
(c) Reference w/o h_{amb} (Split)



(d) Reference w/o h_{amb} (Inverse)



(e) HAM (Split)



(f) HAM (Inverse)

Figure 7.20: Scatterplots for different experiments, showing the impact of the proposed HAM solution in comparison with the reference scenario. In particular, all experiments show the results for the **split** (left) and **inverse split** (right) scenarios. (a)-(b) are the results for the reference scenario, (c)-(d) for the reference without the h_{amb} features, and (e)-(f) for the proposed HAM scenario.

in Fig. 7.20 display a significant reduction in performance. The **inverse split** scenario performs the worst; the neural network might not understand what to predict for input value ranges outside those seen during training, destroying the relationship built with the remaining predictors. This prompts the addition of a second set of experiments, in which the h_{amb} feature is removed altogether from the inputs of the model. Indeed the results for the corresponding **split w/o** h_{amb} and the **inverse split w/o** h_{amb} scenarios see the performance of the reference framework deteriorate less.

Crucial for this ablation study is the performance obtained by the HAM-based framework. The results for **split** and the **inverve split** scenarios reported for the HAM-case in Tab. 7.7 show the performance to be substantially preserved. This is also evidenced in the scatterplots depicted in Fig. 7.20, which suggest a high degree of linear correlation between the predictions and the reference data.

7.5.3 Discussion

Section 7.5 introduced the concept of a third potential solution to the missing representation of InSAR acquisition geometries in the forest parameter regression task. It consists in the use of a deep learning-based pre-processing model, trained ad-hoc to align the InSAR features of a TanDEM-X acquisition with those acquired with a pre-defined reference geometry, potentially removing the need to guarantee a full representation of the InSAR geometries in the forest parameter regression task. To investigate the potential of the proposed solution, the analysis was split into two stages: the first one evaluating the performance of the pre-processing step itself and the second one evaluating the impact of this pre-processing step onto the downstream forest parameter regression task. Looking at the pre-processing alignment task, the results for the proposed ablation study already yield promising insights, achieving an overall good agreement between the predicted and the target samples at the extremely challenging resolution of 6 m. From the performance results it becomes clear that the investigated alignment framework is indeed succeeding in compensating for the relative difference in h_{amb}, achieving both a low estimation bias and a satisfactory absolute error. The parameters selected for this analysis, comprising the model, the reference height of ambiguity and the limited area, all constitute obvious limitations to the approach and will require a deeper investigation and optimization. Crucially, the analysis on the impact on the downstream regression task has helped contextualizing the performance obtained from the alignment problem. Already in its current state, the impact of the HAM remains overall neutral in presence of a complete dataset (i.e. it does not degrade

the performance), but it strongly helps mitigating the negative effects of a sub-optimal training set composition. Notably, the investigated experiments neglected the impact of training the HAM over the Amazonas and applying it over Gabon. It is to be expected that differences between the two regions might have also negatively affected the alignment process and, consequently, the downstream task performance.

7.6 Conclusions

This chapter addressed the challenges related to the country-scale extension of the proposed regression framework first defined in Ch. 6. In the context of missing validation data, the definition of a novel OOD-detector, named the Map of Applicability, was proposed in Sec. 7.1.2. This instrument was used to drive the definition of a robust dataset for training, concentrating on the role of the height of ambiguity and of the raw DEM as input predictors. More in detail, Sec. 7.2 discussed the challenges arising from both the variable close-orbit formation of the bistatic TanDEM-X mission and the missing DEM representation of the existing topographic elevation values. Following that, three solutions were proposed and investigated.

The first one, discussed in Sec. 7.3, investigated the inclusion of additional reference data in order to expand the pool of compatible TanDEM-X acquisitions. In particular, the potential of GEDI L2A products was considered for this purpose. The analyses showed that over the tropical region of Gabon the GEDI canopy height metrics are often in disagreement with the reference LVIS measurements. This result was attributed to an insufficient SNR, not allowing the GEDI shots to consistently reach the ground. Furthermore, it has been shown that these inconsistencies lead to severe performance issues with the data-driven approach when used for either training or testing. These findings ultimately motivate the dismissal of the L2A GEDI products for the investigated supervised regression task.

A second more promising solution has been discussed in Sec. 7.4, consisting in both the relaxation of the stationarity assumption between the reference and the TanDEM-X data, and in the definition of DEM alternatives. This has lead to significantly expansion of the acquisition pool, leading to greater robustness across a larger spectrum of different InSAR geometry configurations. Furthermore, the set of input features was improved by replacing the global TanDEM-X edited DEM product with the acquisition raw DEM, leading to an improved temporal consistency between all the predictors. Also, the training logic was updated to work on pixel-wise samples and validated

patches, including more heterogeneous spatial contexts in the process. Finally, the proposed analyses were extended to the previously uncovered AfriSAR study area of Lopé and found that the overall performance noticeably improved. A performance analysis with respect to different acquisition and scene parameters concluded that the model delivers accurate height estimates, which are fundamentally unbiased by the considered day of the year, the interferometric baseline and the acquisition incidence angle. It was also shown that accuracy fall-offs can be avoided before operative deployment by comparing the feature distributions in the targeted area of interest with those of the training set. Finally, the proposed model was deployed to map the CH for the entirety of Gabon at 25 m of resolution. The resulting geomap was produced by moisaicking a single coverage of TanDEM-X acquisitions (i.e. considering only a single baseline for each pixel).

A third and final solution has been discussed in Sec. 7.5, consisting in the introduction of a deep learning-based pre-processing step, tasked with harmonizing the TanDEM-X InSAR products to a nominal reference acquisition geometry. The ablation study relied on the use of two long time series' acquired over the Amazon rainforest to training and validate the alignment model at a 6 m resolution, which resulted in a satisfactory alignment accuracy. The trained model was subsequently benchmarked by pre-processing and replacing the InSAR features of the TanDEM-X acquisitions over Gabon, and using these as predictors for the forest height regression task. The comparison with the reference highlighted a small performance drop using the full training set, but also suggested performance was much better preserved when acquisition geometry ranges were removed from the training set. This first proof-of-concept suggested the potential for this approach to effectively decouple the InSAR geometry representation problem from that of the downstream forest parameter regression task. Such a pre-processing step would simplify the creation of suitable datasets, especially when considering locations for which no full interferometric baseline-representation can be achieved.

In conclusion, out of the three proposed solutions to the representation problem, only two yielded satisfactory results. The second one in particular, has uniquely led to the generation of a harmonized and consistent country-scale product, whereas the third one has resulted in promising early results which motivate the intention for further investigations.

Chapter 8

Generalization of the Framework to Other Forest-Related Variables and Sensors

The main objective of this chapter is to investigate the flexibility of the developed framework for the adaptation to similar tasks. The focus is on the generalization to both different forestry-related variables, namely the regression of the digital terrain model and of the above-ground biomass density, and to different input sensors, concentrating on Sentinel-1 repeat-pass InSAR and Sentinel-2 multi-spectral optical data. The presented analyses are self-contained, limited in scope and based on the regression framework defined in Ch. 6 and improved in Ch. 7.

8.1 Generalization of the Proposed Framework to the Regression of Other Forest-Related Parameters

A significant property of for the proposed regression framework is the possibility of potentially generalizing it to any forestry-related parameter. This represents one of the biggest advantages of working with deep learning architectures since, once they have properly learned the model that represents the problem in its whole complexity, they are typically capable of transferring "knowledge" to similar tasks in a very effective manner.

This section presents early findings on the extension of the proposed framework to

Table 8.1: DTM estimation: test site-depending and overall prediction accuracy.

DTM Estimation Performance Metrics					
Test-Sites	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	
Lopé	1.86	5.16	1.90	6.83	0.985
Mabounié	0.79	5.27	7.43	6.89	0.983
Mondah	1.32	3.11	21.37	9.11	-0.974
Pongara	-0.89	3.68	25.83	10.02	-0.417
Rabi	0.33	0.33	10.92	4.94	0.947
Overall	0.82	4.45	9.29	6.80	0.997

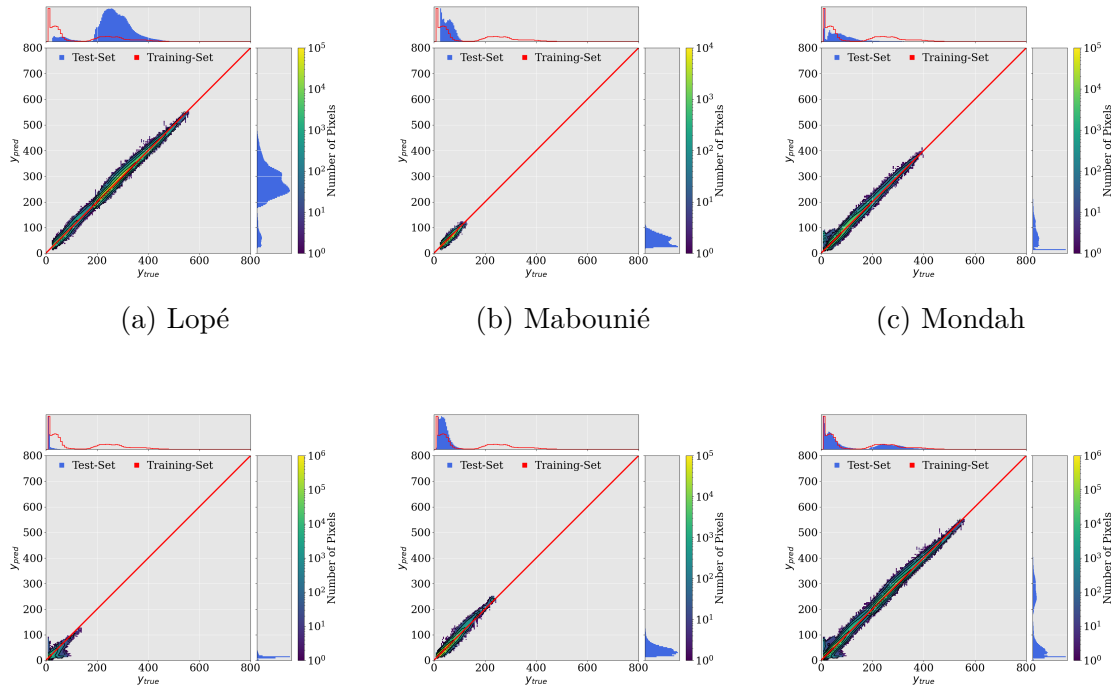
the retrieval of both the digital terrain model (DTM) and the above-ground biomass density (AGBD) estimates from TanDEM-X data. These parameters are clearly related to the investigated task of forest height estimation, being both linked to the height and vertical structure of vegetation, which reflect into the properties of both the SAR backscatter and the interferometric coherence.

Starting from the pre-trained large-scale model (**w/o DEM** case) described in Sec. 7.4. The proposed approach consists in taking advantage of the existing AGBD and DTM reference estimates generated in the context of the AfriSAR campaign using LVIS data, to fine-tune the model for the specific task.

8.1.1 Estimation of the Digital Terrain Model

The process for DTM estimation has a lot in common with the canopy height regression framework detailed in Ch. 6 and Ch. 7. At its core, the idea is to teach the model the penetration bias relationship between the X-band signal of TanDEM-X and the canopy, and to compensate the DEM for the resulting offset.

The reference AfriSAR dataset also offers LiDAR-derived DTM estimates matching those of the other relative height metrics. As input, it is apparent how the inclusion of the DEM among the predictors becomes mandatory, as it acts as a reference height for the model for estimating the DTM offset. For this study the raw DEM was selected, effectively matching the input feature set of the baseline scenario described in Sec. 7.2. Notably, the raw DEM is not calibrated against any reference height measurement, meaning that its inclusion potentially introduces inconsistencies between the mean phase center height and the topography elevation. Considering the already good absolute vertical accuracy of TanDEM-X raw DEMs, which is in the order of few meters, this aspect is assumed to be negligible for a first performance investigation. The remaining details regarding the dataset, the training and the testing procedures, can be found in Sec. 7.2.1.



(a) Lopé (b) Mabounié (c) Mondah

(d) Pongara (e) Rabi (f) Overall

Figure 8.1: Scatterplots of the reference DTM versus the predicted DTM for each of the five considered test-sites and overall, expressed in logarithmic scale.

Tab. 8.1 reports the performance results across the individual test sites and for the dataset as a whole. In particular, the estimation bias is well balanced across all sites, resulting in a small overall ME of 0.82 m, and achieving a MAE of 4.45 m and an RMSE of 6.80 m. Looking at the sites in more detail, the MAPE and R^2 metrics highlight issues with the performance over Mondah and Pongara, which cause the relatively large estimation errors seen for very low elevations in the scatterplots in Fig. 8.1. These errors are consistently found on the shorelines, implying issues related to the mutual calibration between DEM and DMT.

Finally, the use of the large-scale approach detailed in Sec. 7.1 allows for the generation of a country-scale DTM estimate of Gabon. In particular, the output for the 2010/11 TanDEM-X coverage is shown in Fig. 8.2, which results in a wall-to-wall estimate with a resolution of 25 m. Notably, the product is naturally calibrated without the need of additional post-processing steps thanks to the theory-driven training set definition. Localized artefacts remain due to the areas of shadow and layover induced by the SAR geometry, and could be conceivably resolved by combining estimates from both ascending and descending acquisitions for areas with strong topography.

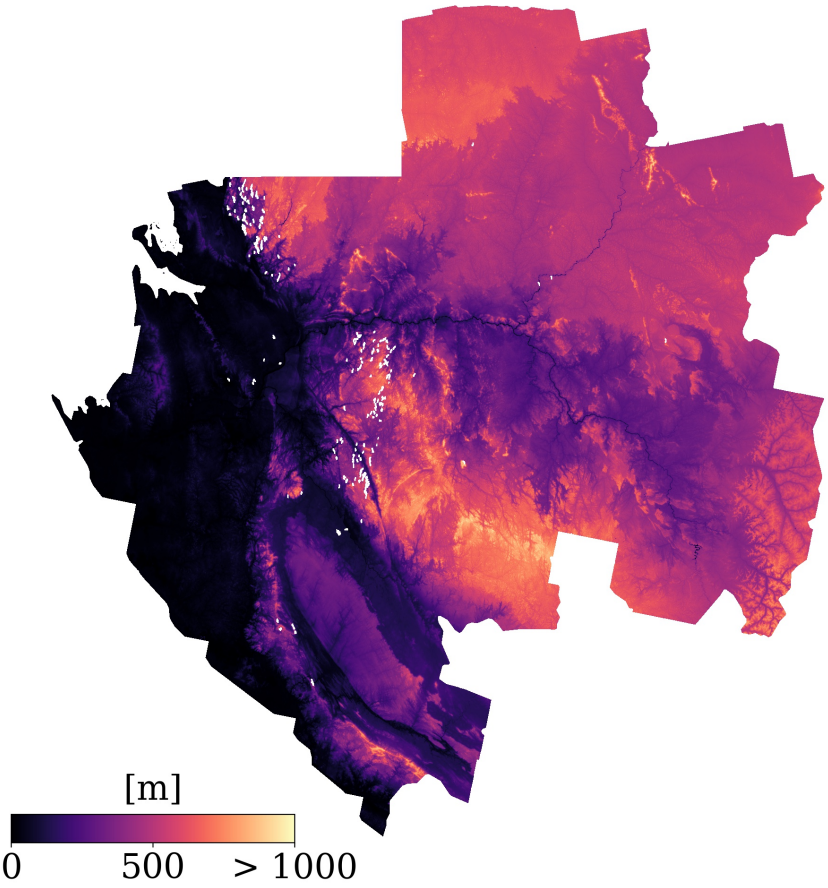
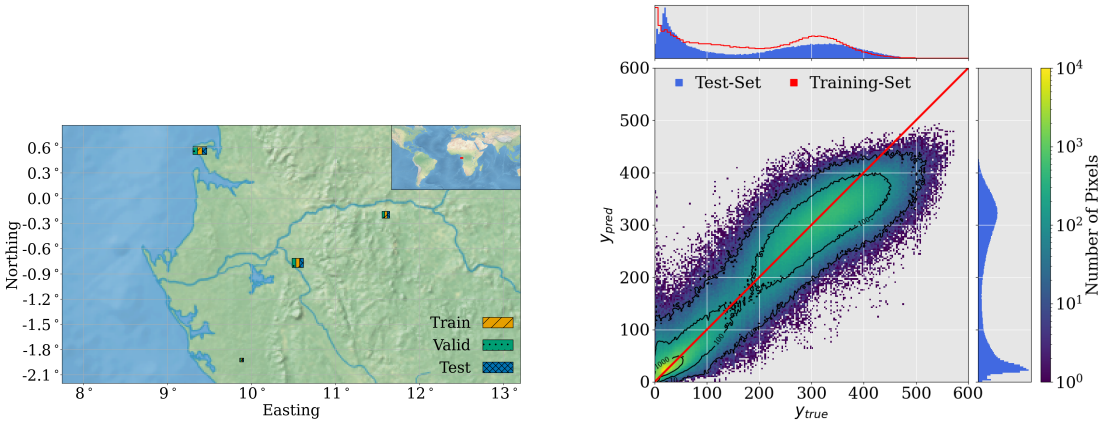


Figure 8.2: Large-scale DTM map of the State of Gabon. The map is produced using TanDEM-X acquisitions from the 2010 and 2011 global coverage with a GSD of 25 m.

8.1.2 Estimation of the Above-Ground Biomass Density

It is also possible to adapt the existing regression framework to the estimation of other phenological parameters, such as the AGBD. In particular, the AfriSAR campaign has also allowed for the generation of AGBD raster products, which are characterized by a significantly smaller extension with respect to the CHM maps, covering only four study-areas (Lopé, Mondah, Mabounié and Rabi) and each consisting only of a small fraction of their original extent. This results in the geographic subsetting shown in Fig. 8.3 (a). It is pointed out that the AGBD estimates were generated with a coarser spatial resolution of 50 m, and were obtained using allometries from LiDAR measurements and field-plots.

Overall, the amount of available training samples is significantly reduced and thus requires the development of a different training strategy. This study investigates the opportunity of re-using the model resulting from Sec. 7.4.2 as a basis for a fine-tuning



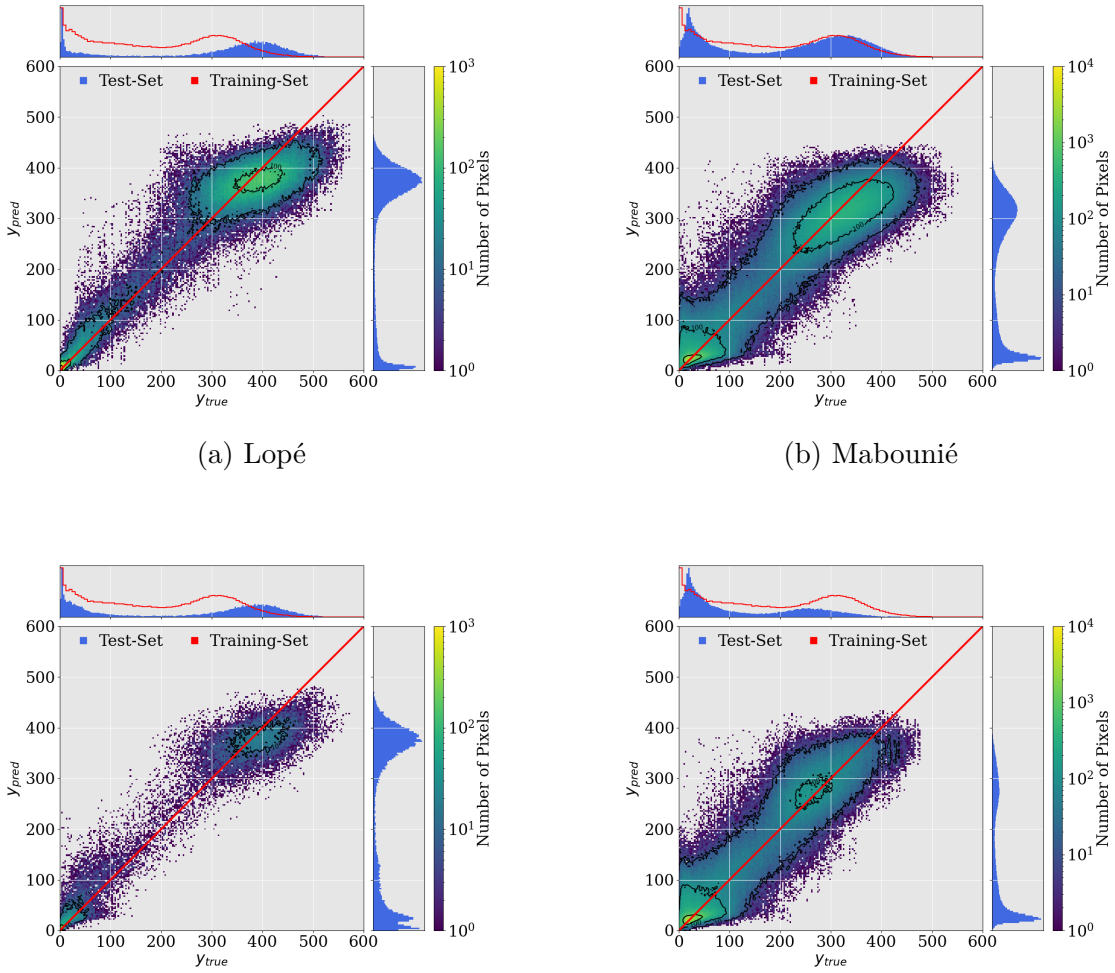
(a) (b)

Figure 8.3: (a) AfriSAR AGBD dataset subsetting results and (b) scatterplot for the test set. It can be noted that for AGBD significantly less reference samples are available with respect to the CHM case. Despite this, the test results show a good agreement between prediction and reference values.

approach over the more limited AGBD dataset. In practical terms, this is achieved by initializing the model with the weights estimated for the CHM regression and then re-training it for the AGBD regression with a reduced initial learning-rate of 10^{-5} until convergence is reached. The resulting fine-tuned model achieves the performance results shown in Tab. 8.2.

In particular, the model yields an extremely competitive MAE of 31.28 Mg/ha corresponding to a low estimation bias of −1.89 Mg/ha. The coefficient of determination reflects a large degree of liner correlation with the reference, resulting in a score of 0.92. This can also be visually assessed from the scatterplot depicted in Fig. 8.3 (b). The only unsatisfactory metric is the MAPE, in particular the result of 100.67% obtained over Mondah, whereas for the remaining sites the score varies between more reasonable 12.53% and 22.50%. Fig. 8.4 reports the scatterplots for the individual test sites. In particular, the scatterplot for Mondah in Fig. 8.4 (c) sees a subset of null vegetation values being overestimated by the model. These could explain the significant peak reported in the MAPE metric, but the reason for this outlier is not yet clear and requires further investigation.

Similarly to the CHM and DTM cases analyzed before, also in this case it is possible to deploy the resulting trained model at country scale, yielding the wall-to-wall AGBD map of Gabon for the year 2010/11 shown in Fig. 8.5. As one can see, the generated map largely results in a high level of consistency. Nevertheless, it is also possible to



(a) Lopé

(b) Mabounié

(c) Mondah

(d) Rabi

Figure 8.4: Scatterplots of the reference AGBD versus the predicted AGBD for each of the four considered test-sites in logarithmic scale.

spot the characteristic "jumps" caused by changes in the coherence in the north-east and the south-west parts of the country.

8.1.3 Discussion

This section has looked into first insights for the definition of an adaptation strategy of the existing canopy height regression framework for the estimation of other forestry-related parameters.

The investigation presented in Sec. 8.1.1 has demonstrated the feasibility of generalizing the existing canopy height regression model to the estimation of the DTM, already

Table 8.2: AGBD estimation: test site-dependent and overall prediction accuracy.

AGBD Estimation Performance Metrics					
Test-Sites	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[Mg/ha]	[Mg/ha]	[%]	[Mg/ha]	[·]
Lopé	-9.03	34.96	22.50	47.12	0.91
Mabounié	-18.5	38.98	12.53	48.99	0.38
Mondah	6.16	26.65	100.67	36.59	0.87
Rabi	9.15	26.29	13.98	46.38	0.49
Overall	-1.89	31.28	64.22	41.95	0.92

yielding extremely promising results. For comparison, this first analysis already resulted in better performance than that obtained by the GEDI spaceborne LiDAR system (see Sec. 7.3.1), which achieved a large positive bias of 3.41 m and a RMSE of 8.03 m, compared to the 0.82 m and 6.80 m, respectively, of the proposed approach. With respect to the literature, the work in [93] estimated the DTM in Lopé using the locally computed phase histogram from single-look UAVSAR data. Validation against LVIS measurements resulted in an RMSE 4.7-8.3 m, depending on the considered InSAR baseline and the detection algorithm, which places the proposed DL algorithm within the considered interval, with the notable advantages of only requiring a single-polarization, of operating on a significantly larger spectrum of interferometric baselines and of working at a much coarser input resolution of 25 m. A more in depth analysis will be necessary to fully comprehend the issues reported for Mondah and Pongara along the coastal regions. In order to achieve peak performance, the DEM calibration issues also need to be addressed as they could significantly affect the training process and therefore the test results. Finally, the dataset considerations discussed in Sec. 7.4.3 remain valid also in this case, and warrant a deeper analysis for this specific scenario as the DEM is an essential input source for the estimation of the DTM, and thus cannot be simply replaced.

Similarly, the analyses presented in Sec. 8.1.2 confirmed that the proposed proof-of-concept regarding the re-adaptation of the existing CHM regression model to the estimation of AGBD is capable of achieving extremely promising results with minimal modifications, despite a significant limitation in the reference data availability. This behavior is probably made possible by the fact that the initial CHM regression model is already robust and well generalized, allowing for an effective knowledge transfer to a similar domain even with few reference data. Qualitatively, it is possible to observe some calibration issues in the final country-scale product. These are probably related to the reduced amount of TanDEM-X geometries that were shown during the training phase, a consequence of both the naive fine-tuning approach and of the significantly reduced AGBD dataset. It is conceivable to assume that these effects could be avoided by

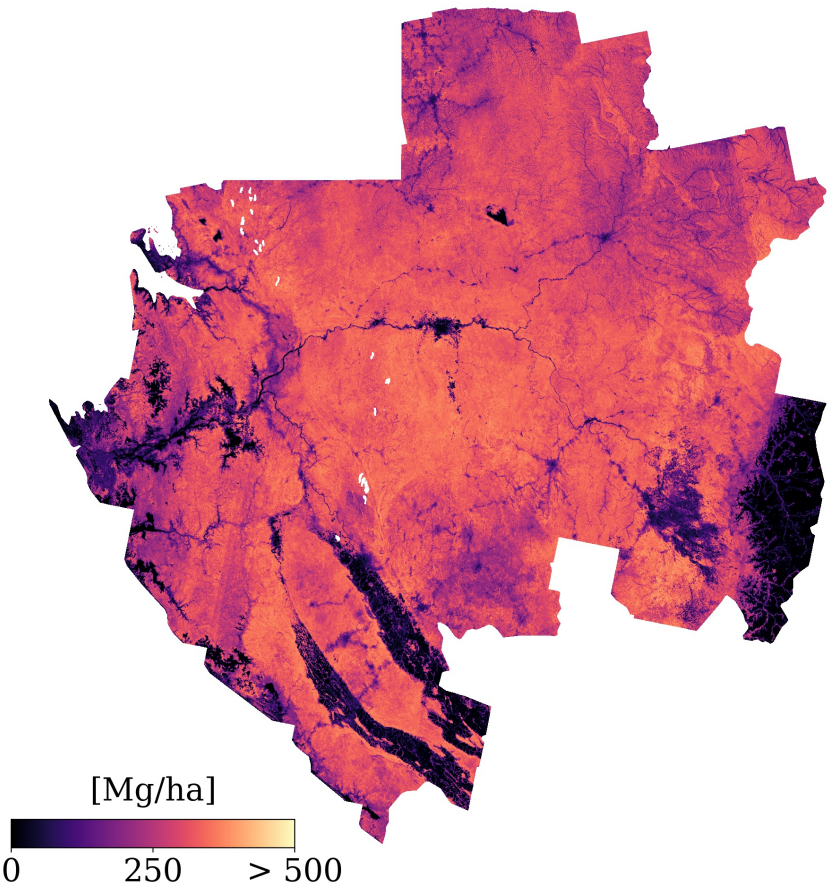


Figure 8.5: Country-scale AGBD map of the State of Gabon. The map uses TanDEM-X images acquired between the end of 2010 and 2011, and has a 50 m GSD. While the overall results appear well harmonized, some striping artefacts remain visible in the north-east (top-right) and south-west (bottom-left) parts of the map. These are probably due to missing TanDEM-X geometries in the training set.

either freezing the earlier model-layers, or by employing the height of ambiguity alignment model proposed in Sec. 7.5. Looking at the literature, the ablation study in [154] evaluated the AGBD regression performance of different machine learning algorithms using quad-pol SAR backscatter maps as predictors. The analysis considered the Lopé site, achieving an RMSE of 72.69 Mg/ha and an R^2 of 0.83, whereas the proposed DL model obtains an RMSE of 47.12 Mg/ha and an R^2 of 0.91, respectively.

Overall, the obtained results consolidate the hypothesis that other bio/physical parameters could also be estimated using a similar approach, confirming the extreme flexibility of the proposed regression framework.

8.2 Generalization of the proposed Framework to Sentinel-1 Data

This section investigates the generalization capability of the proposed framework to use Sentinel-1 data for forest parameter regression. Up to this point only features derived from the TanDEM-X mission have been considered because of their positive characteristics residing in the single-pass nature of the mission. One particular constraint affecting the TanDEM-X mission lies in its limited acquisition resources, resulting in irregular temporal coverages. Differently, Sentinel-1 (and the Copernicus missions more in general) offers regular InSAR repeat-pass acquisitions resulting in multiple yearly global coverages. This renders Sentinel-1 an attractive information source worth exploring in view of large-scale monitoring purposes.

In order to asses its potential, a purely supervised study is performed, aiming at testing the potential of C-band SAR and repeat-pass InSAR features for the estimation of forest heights in the tropical context of the AfriSAR campaign. To this end, the regression framework, previously defined for TanDEM-X in Ch. 6 and Ch. 7, is modified to allow for the use of Sentinel-1 features instead of TanDEM-X ones.

8.2.1 Sentinel-1 Dataset

To build the Sentinel-1 dataset, all available IW SLC products acquired between 2015 and 2016, and intersecting the five regions of interest of the AfriSAR campaign have been queried, downloaded and pre-processed using TAXI [114]. In particular, since this analysis is interested in evaluating the impact of the repeat-pass coherence for the regression task, InSAR processing has been applied. The products are first split according to their absolute orbit number and subsequently organized in contiguous short time-series from which the interferometric features at 12-days temporal baselines can be estimated. This results in a total of 292 individual acquisitions and 5.4 TB of processed data.

For each of the resulting Sentinel-1 datacubes, the following input features are considered:

- Co-pol backscatter map in VV polarization: σ_0^{VV}.
- Cross-pol backscatter map in VH polarization: σ_0^{VH}.
- Cross-pol 12-day repeat-pass coherence: γ^{VH1}.

[1]The co-pol coherence was excluded in the context of related experiments performed during the

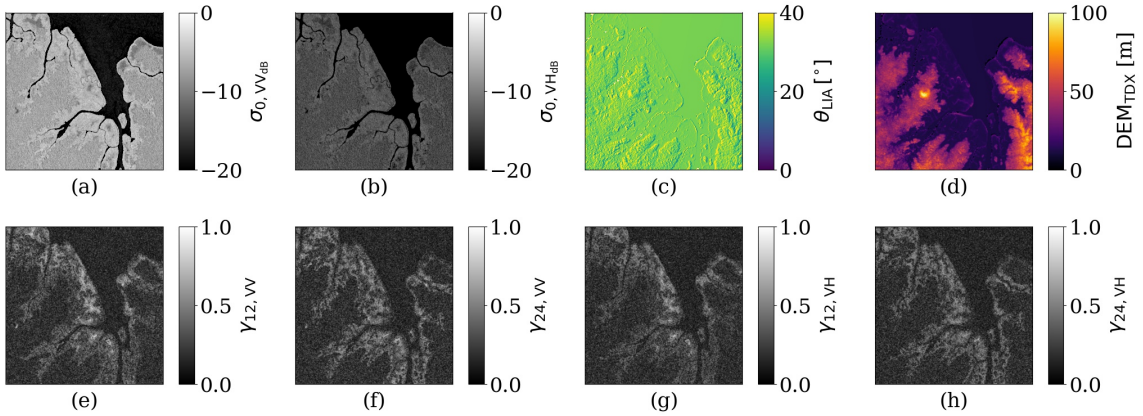


Figure 8.6: Sentinel-1-derived SAR and InSAR features acquired over the test site of Pongara, Gabon. (a) and (b) show the co- and cross-polarized backscattering maps, respectively. (c) shows the local incidence angle of the satellite computed with respect to the topography represented by the TanDEM-X Global Edited DEM in (d). (e) and (f) represent the co-polarized 12-day and 24-day repeat-pass coherences, while (g) and (h) represent the cross-polarized ones, respectively.

- Local Incidence Angle: θ_{lia}.
- TanDEM-X Global Edited DEM: DEM_{ext}.

Compared to the TanDEM-X case, it is not possible to separate the contribution of volumetric decorrelation from the remaining decorrelation sources due to the unknown impact of temporal decorrelation [52]. Furthermore, given the stable orbital tube of the Sentinel-1 mission, it is assumed that the normal baseline for two repeat-pass acquisitions can be neglected, leading to only a very small contribution from volumetric decorrelation effects.

8.2.2 Results

As in the previous experiments, the analysis starts with the definition of a baseline scenario, which uses all of the available input features, Sentinel-1 acquisitions and regions of interest.

Testing the resulting model on the dedicated set-aside dataset yield the performance metrics shown in Tab. 8.3. In particular, for the baseline scenario (**Baseline** case) the model achieves a low bias of $-0.64\,\text{m}$, a MAE of $4.92\,\text{m}$ and a R^2 score of 0.64. For reference, these performance metrics are worse than those obtained using TanDEM-X,

visiting period at JPL.

Table 8.3: Performance summary for the estimation of the canopy height using Sentinel-1 data

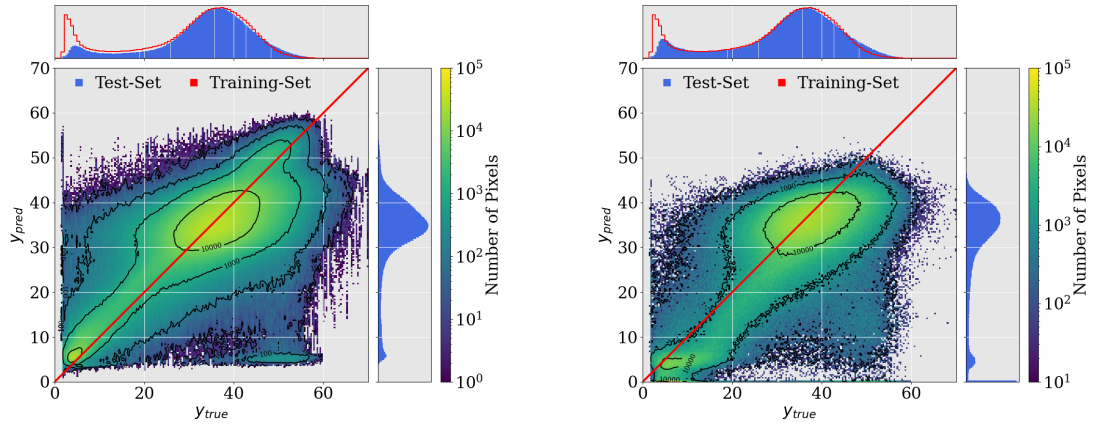
Sentinel-1 CHM Performance Metrics					
Test-Sites	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	[·]
Baseline	-0.64	4.92	18.82	6.67	0.64
SAR	-1.48	5.03	18.95	6.82	0.62
w/o DEM	-1.40	5.83	22.73	7.65	0.52
DEM Only	-3.15	6.30	27.31	8.74	0.48

which under similar conditions obtained a ME of -0.40 m, a MAE of 3.83 m and a R^2 score of 0.76.

Fig. 8.7 shows the scatterplot between the predictions and the reference values, and indicates that the estimates are in overall good linear agreement with the reference. The relationship saturates at 60 m of tree heights, a slightly higher value than that obtained with TanDEM-X. Finally, it is also possible to spot a small cluster of samples being underestimated, which corresponds to the DEM editing issues over Pongara discussed in Sec. 6.4.

Similarly to Sec. 6.6.2, the importance of the Sentinel-1 coherence is evaluated by removing it from the set of input features and training a dedicated model from scratch. The results reported in Tab. 8.3 (**SAR** case) see a small reduction in performance across all considered metrics, which suggests that the role of the coherence map is somewhat negligible. An hypothesis is that the 12-day revisit time is already too long to preserve any amount of correlation between the considered interferometric pair over forested areas, effectively yielding an estimate which is completely biased and therefore uninformative. Crucially, some amount of correlation remains over non-forested areas and mangroves, which suggests that the coherence maps could play a role in better determining the predictions for these classes.

The importance of the external DEM is similarly evaluated, by removing it from the set of input features. Notably, the performance metrics reported in Tab. 8.3 (**w/o DEM** case) see a significant deterioration in the estimation performance, with a MAE of 5.83 m and an R^2 score of 0.52. This is in stark contrast with the TanDEM-X results seen in Sec. 6.6.1, which saw the DEM have only a modest impact on performance. This result suggests the apparent relevance of the TanDEM-X-derived DEM for the Sentinel-1 case, and motivates the decision to evaluate the regression framework using the Global Edited DEM as the sole predictor, i.e. assessing how well it is possible to estimate the forest height directly from the global TanDEM-X DEM product. The results reported in Tab. 8.3 (**DEM Only** case) are significant. The performance consists of a low ME



(a) (b)

Figure 8.7: Scatterplot of the predictions for (a) the **Baseline** case and (b) **DEM Only** case. The histograms on the side report the individual distributions for predictions (right side) and for the reference (top), for the test set (blue) and training set (red), respectively.

of −3.15 m, but a MAE of 6.30 m and an R^2 score of 0.48, which are not distant from the previously reported **w/o DEM** case results. This appears to indicate that the baseline performance for the Sentinel-1 analysis is primarily sustained by the external TanDEM-X DEM product, and suggests that these results are effectively based on an implicit fusion between Sentinel-1 and TanDEM-X derived measurements.

Crucially, the scatterplot in Fig. 8.7 (b) gives more context behind the performance results of the **DEM Only** case. The predictions follow three main distributions, one for tall vegetation, another one for low vegetation and the last one for zero-height vegetation. In particular, the zero-height distribution is indicative of a significant estimation error as, on the one hand, the full-waveform LiDAR-measurements do not allow for null heights do to the width of the transmitted impulse, on the other hand, these estimates correspond to a wide spectrum of different reference tree heights. Looking at the issue in more detail reveals how the this peak of zero-height estimates is found only for the Mondah and Pongara sites along the coastlines. Given that the TanDEM-X global DEM is the only considered input feature, it suggests that the water body flattening errors of the DEM editing processor might cause inconsistencies which cannot be resolved by relying on the other features. The large variation in topography throughout the five study areas suggests that the model mostly relies on local variations in the DEM (i.e. spatial texture features) to estimate the canopy height: strong variations are caused by the penetration into the canopy, and are thus a clear indication of the presence of

trees, while low-frequency variations are more indicative of the absence of vegetation. It appears reasonable to assume that the model mostly focuses on the binary distinction between these to macro-classes, and predominantly associates to each of them the mode of the corresponding distribution seen during training. This aspect will require further analysis, but it could strongly imply that with the TanDEM-X DEM alone the model would fail to generalize to forests with different statistics.

8.2.3 Discussion

This section has introduced a first investigation into the potential use of Sentinel-1-derived SAR and InSAR features for the estimation of forest height using deep learning.

The baseline scenario suggests a reasonable level of regression performance, which in absolute terms performs worse than TanDEM-X under similar conditions, but also manages to improve over TanDEM-X in terms of saturation effects. Similarly to the analysis in Sec. 6.4, the Sentinel-1 baseline performance is negatively impacted by the DEM editing issues in Pongara.

The feature importance analysis showed that the repeat-pass coherence feature has a minor impact on the performance. This is likely caused by the 12-day revisit time of Sentinel-1 over the considered study area, which results in a complete decorrelation over forested areas.

Moreover, the analysis of the impact of the external DEM source has evidenced a strict dependency on it. In particular, using only the TanDEM-X DEM has shown surprisingly similar results to the case Sentinel-1 case without the DEM. This suggest how the baseline scenario effectively represents a naive sensor fusion experiment between Sentinel-1- and TanDEM-X-derived features.

The work in [100], which used different machine learning algorithms for forest height estimation in Brazil, evidenced that Sentinel-1-derived predictors achieved noticeable worse performance than Sentinel-2 ones, and had little impact when both sensors were combined. The DL-based experiment in [37] similarly evidenced that for forest parameter regression in Norway, the dual-pol Sentinel-1 backscatter maps alone perform significantly worse than the Sentinel-2 optical counterpart, and had little contribution when a sensor fusion approach was investigated. The authors in [99] also evidenced a similar behaviour for AGB estimation in West Africa, with Sentinel-1 performing decidedly worse than Sentinel-2. Interestingly thought, in this study multi-source scaling was arguably better, achieving sensible performance improvements over the single-sensor approach. These analyses evidence similar trends, suggesting that Sentinel-1 alone is

effectively ill-suited for accurately mapping forest-related parameters. This is not too surprising, as all experiments were limited to the backscatter maps alone. In particular, the current Sentinel-1 analysis and that for TanDEM-X in Sec. 6.6.2 also converge to similar levels of performance when the TanDEM-X InSAR features are removed, suggesting how at these shorter wavelengths interferometry represents the only viable information source. Given the long global revisit times of 12-days, this effectively limits the capability of the Sentinel-1 mission to accurately map forest parameters.

A more significant investigation is required over areas for which 6-day repeat-pass acquisitions are available, and therefore allowing for a more complete understanding of the impact of temporal decorrelation on the regression problem. Furthermore, the effective combination of the Sentinel-1 features with ancillary information should be further explored, as the synergies with TanDEM-X have shown early promising results hampered by pre-processing artefacts.

8.3 Generalization of the proposed Framework to Sentinel-2 Data

The following section explores the potential of optical data for the regression of forest height estimates. In particular, the analysis investigates the possibility of extending the forest parameter regression framework, originally developed for TanDEM-X InSAR data (see Ch. 6), to Sentinel-2 multi-spectral data.

8.3.1 Sentinel-2 Dataset

Similarly to the previous experiments, this investigation focuses on the training of a supervised regression architecture using the AfriSAR dataset as reference. As input, the Sentinel-2 atmospherically corrected BOA products are selected. Following the introduction in Sec. 5.5, the L2A products are obtained through an ad-hoc developed automatic processing algorithm, which queries and downloads the top-of-atmoshpere L1C products from the ESA CDSE platform[126], atmospherically corrects them using the ESA Sen2Cor processor [129], and applies the s2cloudless algorithm for the generation of a refined cloud-mask.

As discussed in Sec. 5.5, Sentinel-2A was launched in 2015, while Sentinel-2B was only launched two years later in 2017, which limits the availability of optical products around the 2016 acquisition period of the LVIS LiDAR measurements. This aspect is

severely compounded by a heavy presence of cloud cover, which is particularly problematic in tropical regions such as that of Gabon. In order to obtain a sufficiently large number of Sentinel-2 products, all available products acquired between 2015 and 2018 are queried and pre-processed, for a total of 556 products and 431 GB of data. All these acquisitions are further processed following the pre-processing pipeline described in Sec. 5.7.

For the experiments, all 12 bands available in the L2A products are considered, as well as the TanDEM-X derived Global Edited DEM product as an ancillary source of information. As an example, Fig. 8.8 depicts the different Sentinel-2 bands acquired over the test site of Pongara, Gabon. It is possible to observe that almost all of the visible light is absorbed by the photosynthetic process on land (a)-(d), while the harmful NIR radiation is mostly reflected (e)-(i).

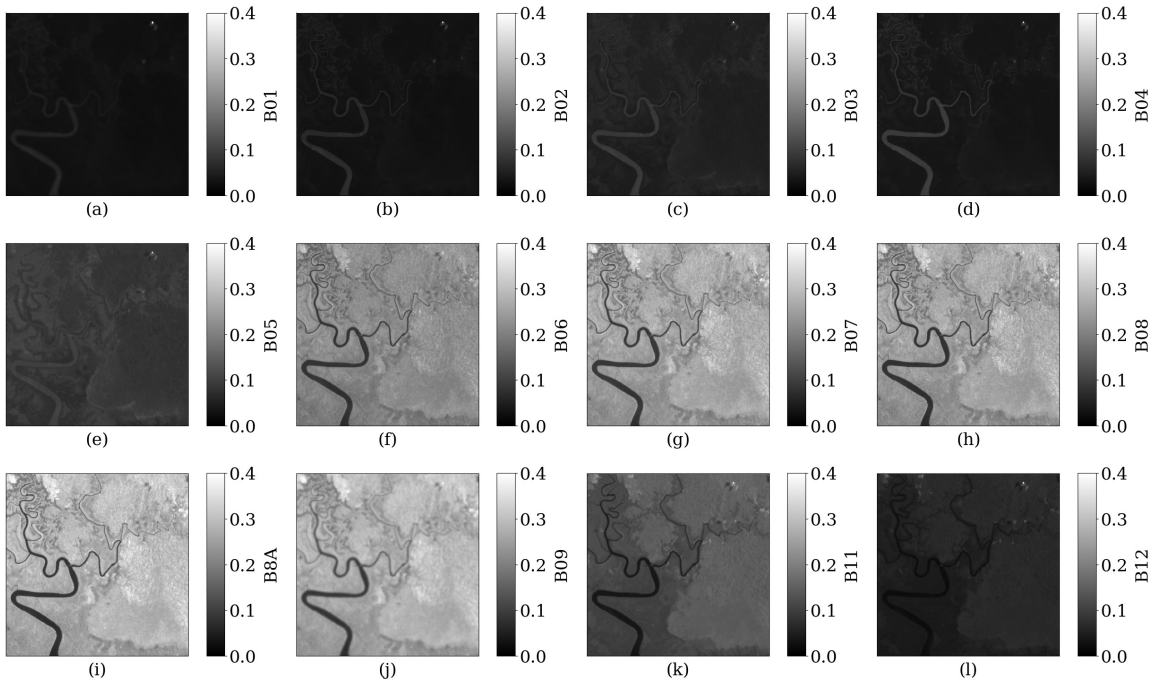


Figure 8.8: Examples of Sentinel-2 bands acquired over the test site of Pongara, Gabon. All bands represent the bottom-of-atmosphere corrected surface reflectance.

8.3.2 Results

Similarly to the previous experiments, also for the Sentinel-2 case a baseline scenario is defined first, which takes advantage of all 12 available multi-spectral bands as its input. The scatterplot for the resulting forest height estimates is shown in Fig. 8.9 and suggests

a high degree of linear correlation with the reference samples. The performance metrics reported in Tab. 8.4 display an overall good R^2 score of 0.73, a low bias of −0.94 m and a MAE of 4.57 m.

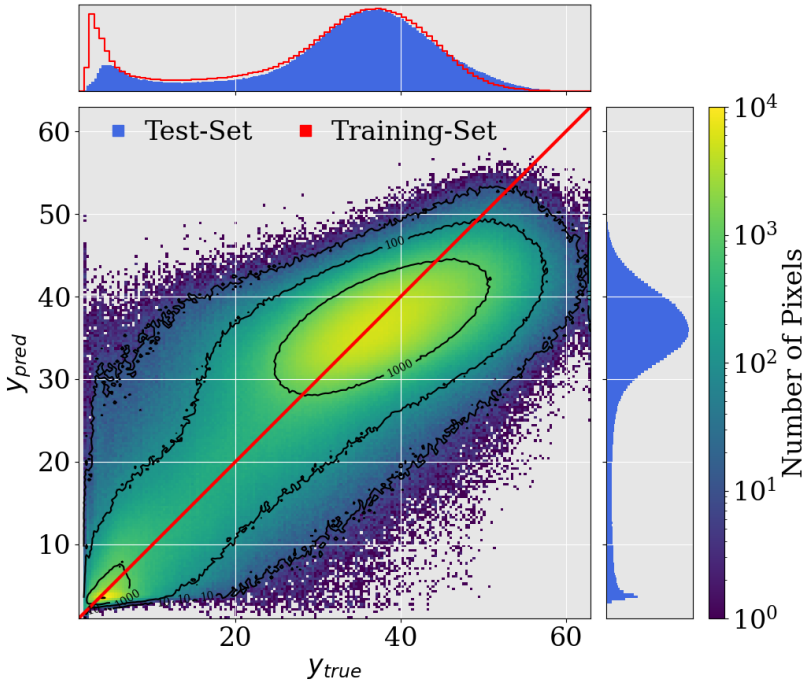


Figure 8.9: Scatterplot for the prediction of the canopy height using Sentinel-2 data.

Table 8.4: Performance summary for the canopy height estimation using Sentinel-2 data.

Sentinel-2 CHM Performance Metrics					
Test-Sites	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	[·]
Lopé	-1.03	4.34	11.79	5.58	0.37
Mabounié	0.55	4.78	17.27	6.18	0.41
Mondah	1.29	3.22	36.86	4.50	0.81
Pongara	-2.27	5.07	22.95	6.76	0.80
Rabi	-1.10	4.58	17.15	5.98	0.43
Overall	-0.94	4.57	17.61	6.01	0.73

In order to assess the impact of the different Sentinel-2 bands on the overall regression performance, a feature importance analysis is performed, consisting in the re-training of the deep learning model using only a subset of the features and assessing its performance compared to the baseline scenario. The addition of an external DEM source, in the form of the TanDEM-X Global Edited DEM product, is also investigated.

Table 8.5: Performance summary for the canopy height estimation using Sentinel-2 data.

Sentinel-2 CHM Performance Metrics					
Test-Sites	**ME**	**MAE**	**MAPE**	**RMSE**	**R^2**
Unit	[m]	[m]	[%]	[m]	[·]
Baseline	-0.94	4.57	17.61	6.01	0.73
Baseline w/ DEM	-0.94	4.30	16.01	5.92	0.74
RGB	-1.64	5.26	19.53	6.93	0.65
RGBN	-1.18	4.92	18.58	6.47	0.69
IR	-0.71	4.73	19.10	6.21	0.72
RGB + IR	-0.76	4.53	17.66	5.96	0.74

Tab. 8.5 reports the performance results for all the investigated scenarios.

First, the addition of the external DEM source to the baseline scenario is considered, as the ancillary information on the local topography, as well as the relationship between the encoded mean phase center and volumetric scatterers is potentially a valuable source of information. Indeed the results improve over the baseline scenario, albeit a modest one. This could suggest that the topographic information is not as crucial for the nadir-looking Sentinel-2 sensors as it was for to the side-looking geometry of SAR sensors, or that the provided information is somewhat redundant with respect to the multi-spectral bands.

Looking only at Sentinel-2 features, the impact of the RGB bands (i.e. B04, B03 and B02) is considered first. It is possible to see from Tab. 8.5 that the performance significantly suffers from the removal of the NIR and SWIR bands, reaching a bias of -1.64 m, a MAE of 5.26 m and an R^2 score of 0.65.

The addition of the NIR band B08 (i.e. **RGBN** case) results in a significant improvement with respect to the previous case (i.e. **RGB** case). Phenological studies commonly rely on the NDVI index [155], which for Sentinel-2 is computed using bands B04 and B08. It is therefore reasonable to assume that capturing the vegetation red edges is crucial for optimal performance using multi-spectral sensors.

Interestingly, when considering only the 20 m IR bands (i.e. **IR** case) the obtained performance is even higher. It is not clear whether this purely depends on the inclusion of the red edge bands (i.e. B05, B06 and B07) for the previously cited motivation, or on the inclusion of the SWIR bands (i.e B11 and B12) that are known to characterize the presence of lignin [156], or on the combination of the two. This aspect requires a more detailed experimental analysis that investigates the role of individual bands and their combination for an optimal feature selection.

Finally, when removing only the 60 m bands from the input (i.e. **RGB + IR** case) the performance slightly outperforms that of the baseline, but it can arguably be motivated by intra-run training variance. Instead, it suggests that the information

included in the dropped 60 m bands might be irrelevant for the forest height regression problem, or that the coarser resolution limits its impact.

8.3.3 Discussion

In this section a proof-of-concept study has been presented, assessing the potential of Sentinel-2 for canopy height regression over a tropical scenario.

The baseline model shows extremely compelling results, with a bias of -0.94 m, a MAE of 4.57 m, an RSME of 6.01 m and a R^2 of 0.73. These results are competitive with those obtained by the TanDEM-X large-scale framework discussed in Sec. 7.4.2, which instead achieved a ME of -0.51 m, a MAE of 4.21 m and an R^2 score of 0.71.

Similar conclusions can be reached when looking at state of the art results reported in the literature. A direct comparison with the study in [35] is possible, as the same reference area and the same input source were considered. Crucially, it is noted that their performance was evaluated at a nominal GSD of 10 m, which is obtained after naively upsampling the original sparse LVIS measurements, each covering a nominal 18-20 m of footprint. Notably, the authors achieve a MAE of 4.9 m and a RMSE of 6.5 m when the prediction on the least clouded acquisition is considered, which reduced to a MAE of 6.0 m and 7.9 m, respectively, when the overall performance across all individual acquisition is evaluated. In both cases, the performance scores are outperformed by the proposed baseline scenario. Finally, the median of the acquisition predictions was also considered, which significantly increases their performance to a MAE of 4.3 m and an RMSE of 5.6 m.

The feature importance analysis evidenced the relevance of the 20 m resolution NIR and SWIR bands, and the complete irrelevance of the 60 m bands for the considered forest height regression problem. Notably, similar observations were also made in [17] and [100], consolidating the findings of this study.

In summary, the considered study proves the efficacy with which the proposed framework can be extended to the used of Sentinel-2 data. The results highlight the relevance of the approach for further large-scale investigation, especially considering both the global coverage and the small-revisit time capabilities achieved by the Sentinel-2 constellation, as well as the planed long-term support of the Copernicus mission. A potential investigation of the synergies between the complementary TanDEM-X and Sentinel-2 missions is also of interest, as early investigations have already suggested a significant potential [157].

8.4 Conclusions

This chapter presented the results of several different studies aimed at understanding the generalization capability of the TanDEM-X-based forest height regression framework previously defined in Ch. 6 and Ch. 7.

The experiments in Sec. 8.1.1 and Sec. 8.1.2 investigated the generalization capability of the model to new forest-related variables, by re-purposing it for either the estimation of the DTM or the AGBD maps. Both experiments resulted in extremely compelling results, achieving good agreement with the LiDAR-derived reference estimates, matching or beating the state-of-the-art methods, and, at the same time, allowing for the country-scale generation of products.

The flexibility of the proposed framework was further analyzed in Sec. 8.2, where the single-pass X-band TanDEM-X data was replaced by 12-day repeat-pass C-band Sentinel-1 data. Interestingly, the results showed that the considered 12-day repeat-pass coherence could not aid the regression performance, yielding results in line with those achieved by solely using the backscatter information of either Sentinel-1 or TanDEM-X.

Finally, the use of multi-spectral Sentinel-2 data was investigated in Sec. 8.3. The experimental results yielded an overall extremely promising regression performance, in line with that achieved using TanDEM-X without the DEM as predictor. An early insight into the relative feature importance also revealed that the NIR and SWIR bands constituted the main performance drivers, whereas the 10 m resolution visible bands only marginally contributed to the regression performance.

In conclusion, the collection of these investigations further cements the flexibility of the proposed framework for the estimation of bio-physical parameters from EO data as sets the basis for further developments.

Chapter 9

Conclusions and Future Work

9.1 Summary and Conclusions

This doctoral thesis presented a novel deep learning framework for the estimation of forest parameters from EO sensors. The intended goal of the work consisted in the precise, flexible and timely prediction of such parameters, resulting in the technical requirements of prompt, accurate and high-resolution estimates at large-scale.

For the first time, the activity primarily focused on the use of X-band InSAR data acquired by the TanDEM-X mission, taking advantage of its high resolution, long operational lifetime and, crucially, the unique single-pass nature of the interferometric mission. In this context, the single-pass acquisition capability of TanDEM-X represents a unique opportunity, as it is not affected by the effects of temporal decorrelation, typically found in repeat-pass configurations, and it allows for a precise characterization of the volumetric contribution of forests instead [28][109].

In summary, the characterization of the proposed framework begun in Ch. 6 with the definition of a deep regression architecture based on the use of interferometric TanDEM-X data as the main predictor. This started with the definition in Sec. 6.1 of dedicated training, validation and testing strategies, based on the considered AfriSAR campaign dataset and on the rigorous notion of geographic sub-setting. Subsequently, Sec. 6.2 focused on the definition of a novel regression model from scratch based on CNNs, by relying on well-documented design principles, theoretical considerations on the input EO sensors, as well as on a thorough hyperparameter tuning analysis, leading to an extremely nimble, fully convolutional model with only 1.4 million trainable parameters. The resulting model performance was first evaluated in Sec. 6.3, which quantified the overall performance as well as the one for the individual AfriSAR sites. Afterwards, a

geographic cross-validation experiment was used to assess the generalization capability of the model, as documented in Sec. 6.4, while Sec. 6.6 presented a series of feature importance analyses, designed as a first approach for the interpretation and explainability of the model's inner workings. The resulting framework already positioned ahead of the state-of-the-art physical-based models when trained and tested on the AfriSAR campaign test sites and was in line with available methodologies based on deep learning. On the other hand, some issues related to the predictor choice remained unresolved, while early investigations revealed suboptimal scaling to country-scales.

These aspects were subsequently further elaborated in Ch. 7, with the analysis in Sec. 7.2 assessing the issues related to the use of TanDEM-X data in combination with data-driven approaches for large-scale inference, with particular emphasis on the importance of a proper representation of the interferometric acquisition geometry in the training data. Of great relevance was the definition of a novel out-of-distribution detector, the Map of Applicability, designed to detect inference sample occurring outside the distributions seen during the training of the model. The resulting tool enabled the design of the necessary dataset requirements for the operational generation of country-scale CH maps.

To address these requirements, three possible solutions were proposed and discussed.

The first research asset, presented in Sec. 7.3, considered the possibility of using the spaceborne LiDAR products from the GEDI mission to improve the compatibility with the TanDEM-X mission. However, SNR-related calibration caused inconsistencies with respect to the LVIS reference measurements and led to subpar training and testing results of the deep learning model.

The second approach, discussed in Sec. 7.4, made use of the MoA-tool to systematically address the dataset requirements. Firstly, the relaxation of the previously defined stationarity constraint between the TanDEM-X products and the reference dataset allowed for a significantly larger selection of TanDEM-X acquisitions and, consequently, InSAR geometries. Secondly, the DEM predictor was replaced by its partial derivatives, removing the dependency of the model from the absolute topographic elevation of a scene. This updated framework was successfully employed to map the entire country of Gabon, producing a time-tagged canopy height map at a resolution of 25 m. In comparison with the state of the art, the proposed approach yielded even better results, effectively outperforming the best methods in the literature while relying only on single-pol, single-baseline acquisitions and being capable of scaling to country-scales.

A third promising approach was also investigated in Sec. 7.5, consisting in a deep-learning based pre-processing step, aimed at pre-aligning the TanDEM-X InSAR fea-

tures to those acquired with a reference geometry. The proposed model was trained using two long time-series acquired over the Amazon forest and it was subsequently used as a pre-processing step for the forest height estimation downstream task. In the absence of adequate input geometries, the proposed alignment model allowed for achieving a significantly better performance than that of the reference framework, making it an extremely promising approach, especially for regions without adequate availability of TanDEM-X and reference data coverages for a proper training of the model.

With the main regression framework in place, the work in Ch. 8 focused on the exploration of the framework flexibility to be adapted to both the estimation of different forest-related parameters and the use of Sentinel-1 and Sentinel-2 data. The analyses in Sec. 8.1 considered extending the model to regress other parameters via fine-tuning, resulting in the generation of country-scale digital terrain model and above-ground biomass products. The proof-of-concept study presented in Sec. 8.2 looked into replacing single-pass TanDEM-X InSAR features with repeat-pass Sentinel-1 ones. The performance results highlighted how the available 12-day repeat-pass products were ineffective predictors for the considered tropical scenario. Similarly, the work in Sec. 8.3 attempted at replacing the TanDEM-X predictors with Sentinel-2 ones and found that these were surprisingly effective, achieving a performance very similar to that of TanDEM-X itself and outperforming similar studies.

The combined results obtained in the presented studies have not only contributed to the definition of the final regression framework, but have also led to significant insights into the inner working principles of the developed deep learning model. In particular, they highlighted the importance of a correct characterization of each input feature, both in terms of its individual contribution to the forest height inversion task and in terms of a complete representation of its value range inside the model. Specifically, the height of ambiguity plays a key role, as the extension from the regional application in Ch. 6 to the large-scale deployment in Ch. 7 exposed the inadequacy of the training set to properly characterize the varying TanDEM-X single-pass InSAR acquisition geometry. Moreover, the analysis in Sec. 7.5.2 demonstrated how an incomplete value range representation could lead to a more severe performance degradation than removing the contextual information altogether.

The use of the TanDEM-X-derived DEM as an ancillary source of information also received particular attention. The different analyses in this work revealed how the primary utility of the DEM lies in the correct distinction between forested and non-forested areas. In particular, the distinct height difference between vegetation and ground and the high-frequency patterns characterizing forested areas were shown to be

extremely relevant properties, in contrast to the low-pass behaviour of the underlying topography. The study results discussed in Sec. 8.2.2, showed that the DEM alone could be used to give a coarse estimation of the forest height, which would primarily differentiate between the absence of canopy and the canopy height mode of the training set. A detailed perspective on the individual sites in Sec. 6.3 revealed how editing errors in the TanDEM-X Global DEM, leading to an artificial flattening of the DEM, would result in strong canopy height underestimation, while the feature importance analysis in Sec. 6.6 suggested that the the DEM is particularly relevant for lower canopies, as removing it from the input feature pool disproportionately affected the lower canopy height ranges.

Regarding the generalization of the output domain, the spatial transfer analysis presented in Sec. 6.4 showed that the proposed regression framework scaled well across unseen areas and displayed only minimal performance loss. Only testing over the test site of Pongara evidenced output-related difficulties in estimating tall mangrove-canopies, which were completely absent from the complementary training set.

Finally, the feature importance analysis in Sec. 6.6 revealed that the InSAR-derived features, namely the interferometric coherence and the volumetric decorrelation coefficient, were the key performance drivers for the proposed model. Interestingly, it was shown that the two measurements were non-redundant and displayed different dependencies on the ancillary height of ambiguity information source. Furthermore, the information about the SAR incidence angle was demonstrated to be essential for achieving absolute regression performance, as it directly affected all SAR and InSAR measurements.

9.2 Outlook

The proposed framework represents a robust basis toward large-scale regression of biophysical parameters from spaceborne EO data. In future activities, the method should be further consolidated by addressing the remaining issues regarding the role of the DEM inside the deep learning model, either by replacing it with proxy variables in order to preserve its rich textural information, or by tweaking the scope of the training set to include a larger range of topographic excursions. The former represents a more flexible method for fixing the issue, but it implicitly assumes that the absolute height of the topography plays no role whatsoever in correctly contextualizing the regression problem (e.g. one might think about the difference between broadleaf and needleleaf

trees found at different altitudes in the Alpine ecosystems). The latter represents the more complete approach, but it strongly depends on the availability of suitable reference data. Alternatively, hybrid solutions, consisting in the unsupervised pre-training of the model followed by a conventional supervised fine-tuning operation, could effectively decouple the issue of the input DEM representation from the final output regression task.

Another aspect to be further investigated is the role of high spatial resolution in the regression problem, since in this work the analyses have been restricted to the resolution of the available reference data at 25 m. An increase in the input InSAR imagery resolution would allow to capture finer textural details, as well as increasing the likelihood of penetrating through the gaps in the canopy [93], yielding a richer description of the properties of the observed vegetation.

Of great interest is also the expansion of the geographic context for the model, with the pantropics representing a favourable first candidate, both for the importance of such an ecosystem and for the similarities with the analyzed tropical forests of Gabon. A first valuable approach would consist in a simple spatial transfer analysis of the existing framework to similar ecosystems on different continents, eventually followed by dedicated fine-tuning strategies. Moreover, the ablation study for the Height of ambiguity Alignment Model (HAM) already showed first promising signs of interoperability between the TanDEM-X features acquired over the Amazon and those in Gabon, which shall be further explored in follow-up studies.

Ultimately, the envisioned goal is that of global-scale deployment, which will inevitably require the framework to incorporate knowledge about a multitude of different biomes and ecoregions, from subtropical to temperate to boreal. Beyond the inherent diversity of the considered scenarios, this would also add new layers of complexity to the problem, such as the inclusion of the concept of seasonality, or of long lasting weather events such as snowfall, floods or droughts, or the need for efficiently operating in biomes where the tree cover is sparse or in presence of high-relief topography. Also, as the generalization requirements and complexity of the regression problem increase, this will make an overall revision of the existing model and training strategy inevitable. Therefore, a further development of the core architecture is similarly relevant, and should aim at investigating and integrating the latest cross-field advancements in artificial intelligence into the proposed regression framework. In particular, large language models (LLM) and vision transformers (ViT) [103] have found enormous success in generative and computer vision tasks but remain strongly underresearched in the field of EO. Some early, purely computer vision-driven investigations have achieved encourag-

ing results [158], but still miss the integration of the physics governing the acquisition process of RS data.

Another important area of interest is the definition of a robust and efficient fine-tuning and adaptation strategy of the existing model aimed at the estimation of additional bio-physical parameters. The preliminary study regarding the estimation of above-ground biomass density has already shown a great potential, but still requires some refinement in order to generate large-scale estimates with the same level of quality as the proposed canopy height maps. Crucially, the currently considered naive fine-tuning approach destroys much of the relationship built between the TanDEM-X products and the canopy height estimates, by re-adjusting it to a smaller subsets of possible TanDEM-X interferometric geometries. An improved approach could consist in freezing the relationship learned within the earlier stages of the model and to more effectively train only the regression head instead.

Similarly, the study on the digital terrain model estimation also showed promising early results which demand follow-up investigations to more concisely frame the regression problem at large scale. For this purpose, a subsequent study should take care of mutually pre-calibrating the input TanDEM-X raw DEMs and the reference DTM, reducing the systematic bias in the regression process. Furthermore, re-formulating the problem as a penetration bias compensation problem could potentially simplify it, as the model would only have to learn to predict a vegetation-dependant offset correction, rather than estimating the absolute topographic height. This could potentially eliminate the necessity to show the model all possible topographic heights as input, as well as allowing it to be applied directly to uncalibrated DEMs, but it would still require to build reference penetration bias estimates from calibrated DEMs and DTMs for the supervised training process.

The proposed study related to the multi-spectral Sentinel-2 mission revealed a great potential for further investigations into the generation of operational large-scale products. A possible synergistic use of the TanDEM-X and Sentinel-2 features, as these offer complementary sets of information about the investigated scene, is also of particular interest, potentially enabling a sharp improvement in the regression performance. The challenge with such a setup lies in the proper fusion of the two streams of information, which requires further investigation into the specifics of the combined model architecture. Differently, the presented Sentinel-1 study revealed a negligible contribution of the 12-day C-band repeat-pass coherence to the forest height regression problem in the considered tropical scenario. Further attention should be given to assess the potential of the 6-day InSAR products over the available regions of interest, such as Europe, as the

Sentinel-1 mission still offers great capabilities in terms of the available data volumes and regular revisit times.

Moreover, further developments of the framework will also be beneficial in view of the future ESA Harmory mission [159], which will consist of two add-on receiver satellites to the Sentinel-1 mission, with a planed single-pass across-track interferometric acquisition phase. This would completely remove the effects of temporal decorrelation, allowing for the isolation of the volumetric decorrelation effects at C-band. As already seen with TanDEM-X, such a configuration would have a large benefit for the regression of forestry parameters, with the added value of short revisit times and the regular coverage of the Planet.

Finally, with the upcoming launches of the L-band NASA/ISRO NISAR [119] and P-band ESA Biomass [92] missions, there is also a particular interest in exploring the synergy between the proposed deep learning approach and the interferometric repeat-pass capabilities offered by longer wavelengths. These missions primarily tend to interact with the larger structural elements of the trees, such as the truck and the branches, and therefore potentially offer better measurements of the above-ground biomass, while being less affected by temporal decorrelation effects than shorter wavelengths, such as C- and X-band.

In light of the work presented in this doctoral thesis, as well as of the state of the art in the literature, the use of data-driven approaches will represent a giant leap in the near future for spaceborne remote sensing applications that will offer remarkable societal benefits, by effectively supporting stakeholders in decision-making processes.

Chapter 10

Bibliography

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

[2] W. W. F. for Nature, "Forest Habitat," https://www.worldwildlife.org/habitats/forest-habitat.

[3] Food and Agriculture Organization, *The State of the World's Forests 2020: Forests, biodiversity and people*, ser. The State of the World's Forests (SOFO). FAO, 2020. [Online]. Available: https://www.fao.org/documents/card/en?details=ca8642en

[4] P. G. Aron, C. J. Poulsen, R. P. Fiorella, and A. M. Matheny, "Stable water isotopes reveal effects of intermediate disturbance and canopy structure on forest water cycling," *Journal of Geophysical Research: Biogeosciences*, vol. 124, no. 10, pp. 2958–2975, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1029/2019JG005118

[5] M. Wolosin and N. Harris, "TROPICAL FORESTS AND CLIMATE CHANGE: THE LATEST SCIENCE," *World Resources Institute*, p. 14, 2018.

[6] David Ellison, "Forests and water: background analytical study," United Nations, Thirteenth session of the United Nations Forum on Forests, 2018.

[7] N. R. Canada, "Forest Carbon," https://www.nrcan.gc.ca/climate-change/impacts-adaptations/climate-change-impacts-forests/forest-carbon/13085.

[8] E. Commission, "Illegal Logging," https://ec.europa.eu/environment/forests/illegal_logging.htm.

[9] A. Baccini, W. Walker, L. Carvalho, M. Farina, D. Sulla-Menashe, and R. A. Houghton, "Tropical forests are a net carbon source based on aboveground measurements of gain and loss," *Science*, vol. 358, no. 6360, pp. 230–234, 2017. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aam5962

[10] P. R. Shukla, J. Skeg, E. C. Buendia, V. Masson-Delmotte, H. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, S. van Diemen, M. Ferrat, E. Haughey, S. Luz, M. Pathak, J. Petzold, J. P. Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, and J. Malley, Eds., *Climate Change and Land: An Ipcc Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*, 2019.

[11] K. Calders, G. Newnham, A. Burt, S. Murphy, P. Raumonen, M. Herold, D. Culvenor, V. Avitabile, M. Disney, J. Armston, and M. Kaasalainen, "Nondestructive estimates of above-ground biomass using terrestrial laser scanning," *Methods in Ecology and Evolution*, vol. 6, no. 2, pp. 198–208, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12301

[12] Picard, Saint-André, Laurent, Nicolas, Henry, and Matieu, *Manual for building tree volume and biomass allometric equations from filed measurement to prediction.* Food and Agriculture Organization of the United Nations (FA0), 2012, OCLC: 931325352. [Online]. Available: http://www.fao.org/docrep/018/i3058e/i3058e.pdf

[13] Bundeswaldinventur, "Surveying the Forest," https://www.bundeswaldinventur.de/en/third-national-forest-inventory/surveying-the-forest/.

[14] J. Askne, J. Fransson, M. Santoro, M. Soja, and L. Ulander, "Model-Based Biomass Estimation of a Hemi-Boreal Forest from Multitemporal TanDEM-X Acquisitions," *Remote Sensing*, vol. 5, no. 11, pp. 5574–5597, Oct. 2013. [Online]. Available: http://www.mdpi.com/2072-4292/5/11/5574

[15] A. Alboabidallah, J. Martin, S. Lavender, and V. Abbott, "Using landsat-8 and sentinel-1 data for above ground biomass assessment in the tamar valley and

dartmoor," in *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 2017, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/8035247/

[16] C. T. d. Almeida, L. S. Galvão, L. E. d. O. C. e. Aragão, J. P. H. B. Ometto, A. D. Jacon, F. R. d. S. Pereira, L. Y. Sato, A. P. Lopes, P. M. L. d. A. Graça, C. V. d. J. Silva, J. Ferreira-Ferreira, and M. Longo, "Combining LiDAR and hyperspectral data for aboveground biomass modeling in the brazilian amazon using different regression algorithms," *Remote Sensing of Environment*, vol. 232, p. 111323, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719303426

[17] H. Astola, T. Häme, L. Sirro, M. Molinier, and J. Kilpi, "Comparison of sentinel-2 and landsat 8 imagery for forest variable prediction in boreal region," *Remote Sensing of Environment*, vol. 223, pp. 257–273, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719300252

[18] 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. Labrière, 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, Oct. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425721002534

[19] T. Markus, T. Neumann, A. Martino, W. Abdalati, K. Brunt, B. Csatho, S. Farrell, H. Fricker, A. Gardner, D. Harding, M. Jasinski, R. Kwok, L. Magruder, D. Lubin, S. Luthcke, J. Morison, R. Nelson, A. Neuenschwander, S. Palm, S. Popescu, C. Shum, B. E. Schutz, B. Smith, Y. Yang, and J. Zwally, "The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation," *Remote Sensing of Environment*, vol. 190, pp. 260–273, Mar. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425716305089

[20] R. Dubayah, J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, J. Armston, H. Tang, L. Duncanson, S. Hancock,

P. Jantz, S. Marselis, P. L. Patterson, W. Qi, and C. Silva, "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography," *Science of Remote Sensing*, vol. 1, p. 100002, Jun. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2666017220300018

[21] B. E. Schutz, H. J. Zwally, C. A. Shuman, D. Hancock, and J. P. DiMarzio, "Overview of the ICESat Mission," *Geophysical Research Letters*, vol. 32, no. 21, p. 2005GL024009, Nov. 2005. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2005GL024009

[22] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan, "GMES Sentinel-1 mission," *Remote Sensing of Environment*, vol. 120, pp. 9–24, May 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425712000600

[23] 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, Nov. 2007. [Online]. Available: http://ieeexplore.ieee.org/document/4373373/

[24] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, May 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425712000636

[25] R. N. Treuhaft, S. N. Madsen, M. Moghaddam, and J. J. van Zyl, "Vegetation characteristics and underlying topography from interferometric radar," *Radio Science*, vol. 31, no. 6, pp. 1449–1485, Nov. 1996. [Online]. Available: http://doi.wiley.com/10.1029/96RS01763

[26] M. Martone, P. Rizzoli, and G. Krieger, "Volume decorrelation effects in TanDEM-x interferometric SAR data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1812–1816, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7590078/

[27] T. Jucker, J. Caspersen, J. Chave, C. Antin, N. Barbier, F. Bongers, M. Dalponte, K. Y. van Ewijk, D. I. Forrester, M. Haeni, S. I. Higgins, R. J. Holdaway, Y. Iida, C. Lorimer, P. L. Marshall, S. Momo, G. R. Moncrieff, P. Ploton, L. Poorter, K. A. Rahman, M. Schlund, B. Sonké, F. J. Sterck, A. T. Trugman, V. A. Usoltsev, M. C. Vanderwel, P. Waldner, B. M. M. Wedeux, C. Wirth, H. Wöll, M. Woods, W. Xiang, N. E. Zimmermann, and D. A. Coomes, "Allometric equations for integrating remote sensing imagery into forest monitoring programmes," *Global Change Biology*, vol. 23, no. 1, pp. 177–190, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/gcb.13388

[28] F. Kugler, D. Schulze, I. Hajnsek, H. Pretzsch, and K. P. Papathanassiou, "TanDEM-X Pol-InSAR Performance for Forest Height Estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6404–6422, Oct. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6716039/

[29] A. Olesk, J. Praks, O. Antropov, K. Zalite, T. Arumäe, and K. Voormansik, "Interferometric SAR Coherence Models for Characterization of Hemiboreal Forests Using TanDEM-X Data," *Remote Sensing*, vol. 8, no. 9, p. 700, Aug. 2016. [Online]. Available: http://www.mdpi.com/2072-4292/8/9/700

[30] E. Donini, M. Amico, L. Bruzzone, and F. Bovolo, "Unsupervised semantic segmentation of radar sounder data using contrastive learning," in *Image and Signal Processing for Remote Sensing XXVIII*, N. Pierdicca, L. Bruzzone, and F. Bovolo, Eds. Berlin, Germany: SPIE, Oct. 2022, p. 23. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12267/2636437/Unsupervised-semantic-segmentation-of-radar-sounder-data-using-contrastive-learning/10.1117/12.2636437.full

[31] D. Ienco, R. Gaetano, and R. Interdonato, "A constrastive semi-supervised deep learning framework for land cover classification of satellite time series with limited labels," *Neurocomputing*, vol. 567, p. 127031, Jan. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231223011542

[32] J.-L. Bueso-Bello, D. Carcereri, M. Martone, M. Martone, C. González, P. Posovszky, and P. Rizzoli, "Deep Learning for Mapping Tropical Forests with TanDEM-X Bistatic InSAR Data," *Remote Sensing*, 2022.

[33] L. Bergamasco, F. Bovolo, and L. Bruzzone, "A Dual-Branch Deep Learning Architecture for Multisensor and Multitemporal Remote Sensing Semantic

Segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2147–2162, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10041791/

[34] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A comprehensive analysis of deep regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2065–2081, 2020. [Online]. Available: http://arxiv.org/abs/1803.08450

[35] 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

[36] Q. Zhang, L. Ge, S. Hensley, G. Isabel Metternicht, C. Liu, and R. Zhang, "PolGAN: A deep-learning-based unsupervised forest height estimation based on the synergy of PolInSAR and LiDAR data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 186, pp. 123–139, Apr. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0924271622000429

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

[38] M. Schwartz, P. Ciais, C. Ottlé, A. De Truchis, C. Vega, I. Fayad, M. Brandt, R. Fensholt, N. Baghdadi, F. Morneau, D. Morin, D. Guyon, S. Dayau, and J.-P. Wigneron, "High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach," *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103711, Apr. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1569843224000657

[39] 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, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9364685/

[40] A. Hein, *Processing of SAR Data: Fundamentals, Signal Processing, Interferometry*, ser. Signals and Communication Technology. Springer Berlin Heidelberg, 2013. [Online]. Available: https://books.google.de/books?id=LCvrCAAAQBAJ

[41] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, ser. SciTech radar and defense series. SciTech Publ., 2004. [Online]. Available: https://books.google.de/books?id=IeGKe40S77AC

[42] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, pp. R1 – R54, 1998.

[43] S. Cloude, *Polarisation: Applications in Remote Sensing*. OUP Oxford, 2010. [Online]. Available: https://books.google.de/books?id=krAVDAAAQBAJ

[44] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Science*, vol. 23, pp. 713–720, 1988. [Online]. Available: https://api.semanticscholar.org/CorpusID:56427653

[45] G. Fornaro, G. Franceschetti, R. Lanari, and E. Sansosti, "Robust phase-unwrapping techniques: a comparison," *Journal of the Optical Society of America A*, vol. 13, no. 12, p. 2355, Dec. 1996. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josaa-13-12-2355

[46] M. Costantini, "A novel phase unwrapping method based on network programming," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 813–821, May 1998. [Online]. Available: http://ieeexplore.ieee.org/document/673674/

[47] C. Chen and H. Zebker, "Phase unwrapping for large SAR interferograms: statistical segmentation and generalized network models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 8, pp. 1709–1719, Aug. 2002. [Online]. Available: http://ieeexplore.ieee.org/document/1036000/

[48] T. Fritz, U. Balss, R. Bamler, and M. Eineder, "Phase unwrapping correction with dual-baseline data for the TanDEM-X mission," in *2012 IEEE International Geoscience and Remote Sensing Symposium*. Munich,

Germany: IEEE, Jul. 2012, pp. 5566–5569. [Online]. Available: http://ieeexplore.ieee.org/document/6352344/

[49] F. Sica, F. Calvanese, G. Scarpa, and P. Rizzoli, "A CNN-Based Coherence-Driven Approach for InSAR Phase Unwrapping," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9234534/

[50] H. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 5, pp. 950–959, Sep. 1992. [Online]. Available: http://ieeexplore.ieee.org/document/175330/

[51] 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, Jan. 1999. [Online]. Available: http://ieeexplore.ieee.org/document/739146/

[52] P. Rizzoli, L. Dell'Amore, J.-L. 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, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9763376/

[53] 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

[54] S. N. Goward, J. G. Masek, D. L. Williams, J. R. Irons, and R. J. Thompson, "The Landsat 7 mission: Terrestrial research and applications for the 21st century," *Remote Sensing of Environment*, vol. 78, no. 1, pp. 3–12, Oct. 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425701002620

[55] K. Murphy, *Probabilistic Machine Learning: An Introduction*, ser. Adaptive Computation and Machine Learning series. MIT Press, 2022.

[56] S. Prince, *Understanding Deep Learning*. MIT Press, 2023.

[57] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042519

[58] W. S. McCulloch and W. H. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. Volume 5, 1943.

[59] A. G. Ivakhnenko, V. G. Lapa, and R. N. McDonough, *Cybernetics and forecasting techniques*. New York: American Elsevier Pub. Co. New York, 1967, section: xxvii, 168 pages illustrations 24 cm.

[60] S. Amari, "A Theory of Adaptive Pattern Classifiers," *IEEE Transactions on Electronic Computers*, vol. EC-16, no. 3, pp. 299–307, Jun. 1967. [Online]. Available: http://ieeexplore.ieee.org/document/4039068/

[61] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," in *Readings in Cognitive Science*, A. Collins and E. E. Smith, Eds. Morgan Kaufmann, 1988, pp. 399–421. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781483214467500352

[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[63] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *ICML 2010*, 2010, pp. 807–814.

[64] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks." in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, USA, April 11-13, 2011*, 2011, pp. 315–323. [Online]. Available: http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf

[65] D. G. T. Barrett and B. Dherin, "Implicit Gradient Regularization," in *ICLR Proceedings*, 2021.

[66] S. L. Smith, B. Dherin, D. G. T. Barrett, and S. De, "On the Origin of Implicit Regularization in Stochastic Gradient Descent," in *ICLR Proceedings*, 2021.

[67] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, 2015, pp. 448–456. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html

[68] Y. Wang, X. Wu, Y. Chang, S. Zhang, Q. Zhou, and J. Yan, "Batch Normalization: Is Learning An Adaptive Gain and Bias Necessary?" in *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*. Macau China: ACM, Feb. 2018, pp. 36–40. [Online]. Available: https://dl.acm.org/doi/10.1145/3195106.3195150

[69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778. [Online]. Available: http://ieeexplore.ieee.org/document/7780459/

[70] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the Loss Landscape of Neural Nets," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/a41b3bb3e6b050b6c9067c67f663b915-Paper.pdf

[71] H. J. Persson, M. J. Soja, J. E. S. Fransson, and L. M. H. Ulander, "National forest biomass mapping using the two-level model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6391–6400, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9222480/

[72] J. Chave, M. Réjou-Méchain, A. Búrquez, E. Chidumayo, M. S. Colgan, W. B. Delitti, A. Duque, T. Eid, P. M. Fearnside, R. C. Goodman, M. Henry, A. Martínez-Yrízar, W. A. Mugasha, H. C. Muller-Landau, M. Mencuccini, B. W. Nelson, A. Ngomanda, E. M. Nogueira, E. Ortiz-Malavassi, R. Pélissier, P. Ploton, C. M. Ryan, J. G. Saldarriaga, and G. Vieilledent, "Improved allometric models to estimate the aboveground biomass of tropical trees," *Global Change Biology*, vol. 20, no. 10, pp. 3177–3190, 2014, _eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.12629. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12629

[73] S. Saatchi, J. Chave, N. Labriere, N. Barbier, M. Réjou-Méchain, A. Ferraz, and S. Tao, "Afrisar: Aboveground biomass for lope, mabounie, mondah, and rabi sites, gabon," 2019. [Online]. Available: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1681

[74] R. Dubayah, M. Hofton, J. Blair, J. Armston, H. Tang, and S. Luthcke, "Gedi l2a elevation and height metrics data global footprint level v002. 2021," 2021. [Online]. Available: https://doi.org/10.5067/GEDI/GEDI02_A.002.Accessed2023-11-28

[75] I. Fayad, N. Baghdadi, and K. Lahssini, "An Assessment of the GEDI Lasers' Capabilities in Detecting Canopy Tops and Their Penetration in a Densely Vegetated, Tropical Area," *Remote Sensing*, vol. 14, no. 13, p. 2969, Jun. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/13/2969

[76] E. P. W. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio Science*, vol. 13, no. 2, pp. 357–364, Mar. 1978. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/RS013i002p00357

[77] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6–43, Mar. 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6504845/

[78] K. Papathanassiou and S. Cloude, "Single-baseline polarimetric SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2352–2363, Nov. 2001. [Online]. Available: http://ieeexplore.ieee.org/document/964971/

[79] S. Cloude and K. Papathanassiou, "Three-stage inversion process for polarimetric SAR interferometry," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 150, no. 3, p. 125, 2003. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/ip-rsn_20030449

[80] G. Krieger, M. Zink, M. Bachmann, B. Bräutigam, D. Schulze, M. Martone, P. Rizzoli, U. Steinbrecher, J. Walter Antony, F. De Zan, I. Hajnsek,

K. Papathanassiou, F. Kugler, M. Rodriguez Cassola, M. Younis, S. Baumgartner, P. López-Dekker, P. Prats, and A. Moreira, "TanDEM-X: A radar interferometer with two formation-flying satellites," *Acta Astronautica*, vol. 89, pp. 83–98, Aug. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0094576513000805

[81] H. Chen, S. R. Cloude, and D. G. Goodenough, "Forest Canopy Height Estimation Using Tandem-X Coherence Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 3177–3188, Jul. 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7533503/

[82] 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/

[83] 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. [Online]. Available: https://ieeexplore.ieee.org/document/9353204/

[84] C. Choi, V. Cazcarra-Bes, R. Guliaev, M. Pardini, K. P. Papathanassiou, W. Qi, J. Armston, and R. Dubayah, "Large Scale Forest Height Mapping by Combining TanDEM-X and GEDI Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–14, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10049501/

[85] M. Lavalle, M. Simard, and S. Hensley, "A Temporal Decorrelation Model for Polarimetric Radar Interferometers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2880–2888, Jul. 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6096407/

[86] M. Lavalle and S. Hensley, "Extraction of Structural and Dynamic Properties of Forests From Polarimetric-Interferometric SAR Data Affected by Temporal Decorrelation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4752–4767, Sep. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7084639/

[87] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000. [Online]. Available: http://ieeexplore.ieee.org/document/868873/

[88] M. Pardini, M. Tello, V. Cazcarra-Bes, K. P. Papathanassiou, and I. Hajnsek, "L- and P-Band 3-D SAR Reflectivity Profiles Versus Lidar Waveforms: The AfriSAR Case," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3386–3401, Oct. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8424006/

[89] M. Mariotti D'Alessandro and S. Tebaldini, "Digital Terrain Model Retrieval in Tropical Forests Through P-Band SAR Tomography," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6774–6781, Sep. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8701518/

[90] Y.-N. Ngo, Y. Huang, D. H. T. Minh, L. Ferro-Famil, I. Fayad, and N. Baghdadi, "Tropical Forest Vertical Structure Characterization: From GEDI to P-Band SAR Tomography," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, conference Name: IEEE Geoscience and Remote Sensing Letters. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9900356

[91] M. Nannini, M. Martone, P. Rizzoli, P. Prats-Iraola, M. Rodriguez-Cassola, A. Reigber, and A. Moreira, "Coherence-based SAR tomography for spaceborne applications," *Remote Sensing of Environment*, vol. 225, pp. 107–114, May 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719300446

[92] S. Quegan, T. Le Toan, J. Chave, J. Dall, J.-F. Exbrayat, D. H. T. Minh, M. Lomas, M. M. D'Alessandro, P. Paillou, K. Papathanassiou, F. Rocca, S. Saatchi, K. Scipal, H. Shugart, T. L. Smallman, M. J. Soja, S. Tebaldini, L. Ulander, L. Villard, and M. Williams, "The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space," *Remote Sensing of Environment*, vol. 227, pp. 44–60, Jun. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719301233

[93] G. H. X. Shiroma and M. Lavalle, "Digital Terrain, Surface, and Canopy Height Models From InSAR Backscatter-Height Histograms," *IEEE Transactions on*

Geoscience and Remote Sensing, vol. 58, no. 6, pp. 3754–3777, Jun. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8961153/

[94] R. N. Treuhaft, B. D. Chapman, J. R. dos Santos, F. G. Gonçalves, L. V. Dutra, P. M. L. A. Graça, and J. B. Drake, "Vegetation profiles in tropical forests from multibaseline interferometric synthetic aperture radar, field, and lidar measurements," *Journal of Geophysical Research: Atmospheres*, vol. 114, no. D23, 2009, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD011674. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1029/2008JD011674

[95] R. Treuhaft and F. Goncalves, "Exploring Vegetation Profiles from TanDEM-X Phase, Lidar, and Field Measurements in Tropical Forests," in *EUSAR 2014 - 2014 European Conference on Synthetic Aperture Radar*, 2014.

[96] C. Wu, S. Tebaldini, M. M. d'Alessandro, L. Zhang, and M. Liao, "Sar Tomography And Phase Histogram Techniques For Remote Sensing Of Forested Areas: An Experimental Study Based On Tomosense Data," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*. Pasadena, CA, USA: IEEE, Jul. 2023, pp. 1883–1886. [Online]. Available: https://ieeexplore.ieee.org/document/10282119/

[97] C. Wu, S. Tebaldini, Y. Yu, M. Manzoni, and M. Liao, "A Comparison Between SAR Tomography and the Phase Histogram Technique for Remote Sensing of Forested Areas," Oct. 2023. [Online]. Available: https://www.researchsquare.com/article/rs-3427467/v1

[98] M. Schlund and H. D. V. Boehm, "Assessment of linear relationships between TanDEM-x coherence and canopy height as well as aboveground biomass in tropical forests," *International Journal of Remote Sensing*, vol. 42, no. 9, pp. 3405–3425, 2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/01431161.2020.1871101

[99] G. Forkuor, J.-B. Benewinde Zoungrana, K. Dimobe, B. Ouattara, K. P. Vadrevu, and J. E. Tondoh, "Above-ground biomass mapping in west african dryland forest using sentinel-1 and 2 datasets - a case study," *Remote Sensing of Environment*, vol. 236, p. 111496, 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719305152

[100] C. Torres de Almeida, J. Gerente, J. Rodrigo dos Prazeres Campos, F. Caruso Gomes Junior, L. A. Providelo, G. Marchiori, and X. Chen, "Canopy Height Mapping by Sentinel 1 and 2 Satellite Images, Airborne LiDAR Data, and Machine Learning," *Remote Sensing*, vol. 14, no. 16, p. 4112, Aug. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/16/4112

[101] Z. Liao, A. I. Van Dijk, B. He, P. R. Larraondo, and P. F. Scarth, "Woody vegetation cover, height and biomass at 25-m resolution across australia derived from multiple site, airborne and satellite observations," *International Journal of Applied Earth Observation and Geoinformation*, vol. 93, p. 102209, 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0303243420304189

[102] H. J. Persson, J. Jonzén, and M. Nilsson, "Combining TanDEM-x and sentinel-2 for large-area species-wise prediction of forest biomass and volume," *International Journal of Applied Earth Observation and Geoinformation*, vol. 96, p. 102275, 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0303243420309181

[103] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *ICLR 2021 Proceedings*, Jun. 2021.

[104] X. Wang and H. Wang, "Forest height mapping using complex-valued convolutional neural network," *IEEE Access*, vol. 7, pp. 126 334–126 343, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8822431/

[105] N. Lang, W. Jetz, K. Schindler, and J. D. Wegner, "A high-resolution canopy height model of the Earth," *Nature Ecology & Evolution*, Sep. 2023. [Online]. Available: https://www.nature.com/articles/s41559-023-02206-6

[106] R. Werninghaus and S. Buckreuss, "The TerraSAR-X Mission and System Design," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 606–614, Feb. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5339240/

[107] A. Moreira, G. Krieger, and J. Mittermayer, "Satellitenkonfiguration zur interferometrischen und/oder tomographischen abbildung der erdoberfläche mittels radar mit synthetischer apertur," 1 2003, patent no. EP 1273518 A2.

[Online]. Available: https://data.epo.org/publication-server/document?iDocId=2280642&iFormat=0

[108] ——, "Satellite configuration for interferometric and/or tomographic remote sensing by means of synthetic aperture radar (sar)," 1 2003, patent no. US 6,677,884 B2.

[109] M. Martone, P. Rizzoli, C. Wecklich, C. González, J.-L. Bueso-Bello, P. Valdo, D. Schulze, M. Zink, G. Krieger, and A. Moreira, "The global forest/non-forest map from TanDEM-X interferometric SAR data," *Remote Sensing of Environment*, vol. 205, pp. 352–373, Feb. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425717305795

[110] C. González, M. Bachmann, J.-L. 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. [Online]. Available: https://www.mdpi.com/2072-4292/12/23/3961

[111] J.-L. Bueso-Bello, M. Martone, C. González, F. Sica, P. Valdo, P. Posovszky, A. Pulella, and P. Rizzoli, "The Global Water Body Layer from TanDEM-X Interferometric SAR Data," *Remote Sensing*, vol. 13, no. 24, p. 5069, Dec. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/24/5069

[112] M. Lachaise, C. Gonzalez, P. Rizzoli, B. Schweibhelm, and M. Zink, "The New Tandem-X DEM Change Maps Product," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. Kuala Lumpur, Malaysia: IEEE, Jul. 2022, pp. 5432–5435. [Online]. Available: https://ieeexplore.ieee.org/document/9883612/

[113] T. Fritz, H. Breit, C. Rossi, U. Balss, M. Lachaise, and S. Duque, "Interferometric processing and products of the TanDEM-X mission," in *2012 IEEE International Geoscience and Remote Sensing Symposium*. Munich, Germany: IEEE, Jul. 2012, pp. 1904–1907. [Online]. Available: http://ieeexplore.ieee.org/document/6351133/

[114] P. Prats, M. Rodriguez-Cassola, L. Marotti, M. Naninni, S. Wollstadt, D. Schulze, N. Tous-Ramon, M. Younis, G. Krieger, and A. Reigber, "Taxi: A versatile processing chain for experimental TanDEM-X product evaluation," in *2010 IEEE International Geoscience and Remote Sensing Symposium*.

Honolulu, HI, USA: IEEE, Jul. 2010, pp. 4059–4062. [Online]. Available: http://ieeexplore.ieee.org/document/5651002/

[115] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, "Φ-Net: Deep Residual Learning for InSAR Parameters Estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 3917–3941, May 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9198099/

[116] W. Kropatsch and D. Strobl, "The generation of SAR layover and shadow maps from digital elevation models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 1, pp. 98–107, Jan. 1990. [Online]. Available: http://ieeexplore.ieee.org/document/45752/

[117] J. B. Blair and M. Hofton, "AfriSAR LVIS L1B Geolocated Return Energy Waveforms, Version 1," 2017. [Online]. Available: https://nsidc.org/data/aflvis1b/versions/1

[118] M. Hofton, J. Minster, and J. Blair, "Decomposition of laser altimeter waveforms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 4, pp. 1989–1996, Jul. 2000. [Online]. Available: http://ieeexplore.ieee.org/document/851780/

[119] P. A. Rosen and R. Kumar, "NASA-ISRO SAR (NISAR) Mission Status," in *2021 IEEE Radar Conference (RadarConf21)*. Atlanta, GA, USA: IEEE, May 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9455211/

[120] 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, publisher: ORNL Distributed Active Archive Center. [Online]. Available: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1775

[121] M. Hofton and J. B. Blair, "Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products," University of Maryland, Tech. Rep., Dec. 2019.

[122] S. B. Luthcke, T. J. Sabaka, J. Nicholas, S. Preaux, and M. Hofton, "Algorithm Theoretical Basis Document (ATBD) for GEDI L3 Gridded Land Surface Metrics," University of Maryland, Tech. Rep., Mar. 2021.

[123] J. Kellner, J. Armston, and L. Duncanson, "Algorithm theoretical basis document for GEDI footprint aboveground biomass density," University of Maryland, preprint, Aug. 2022. [Online]. Available: http://eartharxiv.org/repository/view/3514/

[124] H. Tang, J. Stoker, S. Luthcke, J. Armston, K. Lee, B. Blair, and M. Hofton, "Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI," *Remote Sensing of Environment*, vol. 291, p. 113571, Jun. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425723001220

[125] E. S. Agency, "Mission ends for Copernicus Sentinel-1B satellite," https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Mission_ends_for_Copernicus_Sentinel-1B_satellite.

[126] ——, "Copernicus Data Space Ecosystem," https://dataspace.copernicus.eu/.

[127] F. De Zan and A. Monti Guarnieri, "TOPSAR: Terrain Observation by Progressive Scans," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2352–2360, Sep. 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1677745/

[128] E. S. Agency, "Sentinel-2 level-2a algorithm theoretical basis document(atbd), s2-pdgs-mpc-atbd-l2a," https://step.esa.int/thirdparties/sen2cor/2.10.0/docs/S2-PDGS-MPC-L2A-ATBD-V2.10.0.pdf.

[129] J. Louis, B. Pflug, M. Main-Knorn, V. Debaecker, U. Mueller-Wilm, R. Q. Iannone, E. Giuseppe Cadau, V. Boccia, and F. Gascon, "Sentinel-2 Global Surface Reflectance Level-2a Product Generated with Sen2Cor," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. Yokohama, Japan: IEEE, Jul. 2019, pp. 8522–8525. [Online]. Available: https://ieeexplore.ieee.org/document/8898540/

[130] E. S. Agency, "ESA Sentinel Application Platform," https://step.esa.int/.

[131] ——, "sen2cor," https://step.esa.int/main/snap-supported-plugins/sen2cor/.

[132] R. de los Reyes, M. Langheinrich, P. Schwind, R. Richter, B. Pflug, M. Bachmann, R. Müller, E. Carmona, V. Zekoll, and P. Reinartz, "PACO:

Python-Based Atmospheric Correction," *Sensors*, vol. 20, no. 5, p. 1428, Mar. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/5/1428

[133] K. Tarrio, X. Tang, J. G. Masek, M. Claverie, J. Ju, S. Qiu, Z. Zhu, and C. E. Woodcock, "Comparison of cloud detection algorithms for Sentinel-2 imagery," *Science of Remote Sensing*, vol. 2, p. 100010, Dec. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2666017220300092

[134] S. Skakun, J. Wevers, C. Brockmann, G. Doxani, M. Aleksandrov, M. Batič, D. Frantz, F. Gascon, L. Gómez-Chova, O. Hagolle, D. López-Puigdollers, J. Louis, M. Lubej, G. Mateo-García, J. Osman, D. Peressutti, B. Pflug, J. Puc, R. Richter, J.-C. Roger, P. Scaramuzza, E. Vermote, N. Vesel, A. Zupanc, and L. Žust, "Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2," *Remote Sensing of Environment*, vol. 274, p. 112990, Jun. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425722001043

[135] O. Hagolle, M. Huc, C. Desjardins, S. Auer, and R. Richter, "MAJA Algorithm Theoretical Basis Document," Dec. 2017, publisher: Zenodo. [Online]. Available: https://zenodo.org/records/1209633

[136] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images," *Remote Sensing of Environment*, vol. 159, pp. 269–277, Mar. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425714005069

[137] Sinergize, "s2cloudless: Sentinel hub's cloud detector for sentinel-2 imagery," https://github.com/sentinel-hub/sentinel2-cloud-detector.

[138] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

[139] D. Zanaga, R. Van De Kerchove, D. Daems, W. De Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N. Tsendbazar, P. Xu, F. Ramoino, and O. Arino, "Esa worldcover 10 m 2021 v200," 2022.

[140] 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/

[141] C. Choi, M. Pardini, J. Armston, K. Papathanassiou, and R. Dubayah, "Relating TanDEM-X Local InSAR Phase Center Variations to Li- dar Full Waveforms over Forest Scenarios," in *EUSAR 2018 - 2018 European Conference on Synthetic Aperture Radar*, 2018.

[142] A. Danklmayer and M. Chandra, "Precipitation induced signatures in SAR images," *European Conference on Antennas and Propagation*, Mar. 2009.

[143] S. Cloude and K. Papathanassiou, "Polarimetric SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1551–1565, Sep. 1998. [Online]. Available: http://ieeexplore.ieee.org/document/718859/

[144] Y. Lei, R. Treuhaft, and F. Goncalves, "Automated estimation of forest height and underlying topography over a Brazilian tropical forest with single-baseline single-polarization TanDEM-X SAR interferometry," *Remote Sensing of Environment*, vol. 252, p. 112132, Jan. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425720305058

[145] D. Carcereri, P. Rizzoli, L. Dell'Amore, 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

[146] D. Carcereri, P. Rizzoli, and L. Bruzzone, "A computer-implemented method for generating a data driven model based on information from a plurality of single-pass interferometric sar acquisitions," patentde 2 024 022 013 564 100DE.

[147] P. Rizzoli, M. Martone, C. Gonzalez, C. Wecklich, D. Borla Tridon, B. Bräutigam, 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. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S092427161730093X

[148] K. Lahssini, N. Baghdadi, G. Le Maire, and I. Fayad, "Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests," *Remote Sensing*, vol. 14, no. 24, p. 6264, Dec. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/24/6264

[149] D. Morin, M. Planells, N. Baghdadi, A. Bouvet, I. Fayad, T. Le Toan, S. Mermoz, and L. Villard, "Improving heterogeneous forest height maps by integrating GEDI-based forest height information in a multi-sensor mapping process," *Remote Sensing*, vol. 14, p. 2079, 2022. [Online]. Available: www.mdpi.com/2072-4292/14/9/2079

[150] W. Liu, X. Wang, J. D. Owens, and Y. Li, "Energy-based out-of-distribution detection," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020.

[151] H. Meyer and E. Pebesma, "Predicting into unknown space? Estimating the area of applicability of spatial prediction models," *Methods in Ecology and Evolution*, vol. 12, no. 9, pp. 1620–1633, Sep. 2021. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13650

[152] M. Rajab Pourrahmati, N. Baghdadi, and I. Fayad, "Comparison of GEDI LiDAR Data Capability for Forest Canopy Height Estimation over Broadleaf and Needleleaf Forests," *Remote Sensing*, vol. 15, no. 6, p. 1522, Mar. 2023. [Online]. Available: https://www.mdpi.coma/2072-4292/15/6/1522

[153] F. H. Wagner, S. Roberts, A. L. Ritz, G. Carter, R. Dalagnol, S. Favrichon, M. C. Hirye, M. Brandt, P. Ciais, and S. Saatchi, "Sub-meter tree height mapping of California using aerial images and LiDAR-informed U-Net model," *Remote Sensing of Environment*, vol. 305, p. 114099, May 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S003442572400110X

[154] N. Ramachandran and O. Dikshit, "Forest Aboveground Biomass Estimation from Airborne L-Band SAR Data Using Machine Learning," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium.*

Kuala Lumpur, Malaysia: IEEE, Jul. 2022, pp. 6403–6405. [Online]. Available: https://ieeexplore.ieee.org/document/9884839/

[155] Y. T. Solano-Correa, D. Carcereri, F. Bovolo, and L. Bruzzone, "Large-Scale Precise Mapping of Agricultural Fields in Sentinel-2 Satellite Image Time Series," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa, HI, USA: IEEE, Sep. 2020, pp. 2288–2291. [Online]. Available: https://ieeexplore.ieee.org/document/9323150/

[156] ——, "Identification of non-photosynthetic vegetation areas in Sentinel-2 satellite image time series," in *Image and Signal Processing for Remote Sensing XXV*, L. Bruzzone, F. Bovolo, and J. A. Benediktsson, Eds. Strasbourg, France: SPIE, Oct. 2019, p. 32. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11155/2533761/Identification-of-non-photosynthetic-vegetation-areas-in-Sentinel-2-satellite/10.1117/12.2533761.full

[157] D. Carcereri, P. Rizzoli, D. Ienco, J.-L. Bueso-Bello, C. Gonzalez, S. Puliti, and L. Brurzone, "Large Scale Forest Parameter Estimation Through a Deep Learning-Based Fusion of Sentinel-2 and Tandem-X Data," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. Kuala Lumpur, Malaysia: IEEE, Jul. 2022, pp. 5773–5776. [Online]. Available: https://ieeexplore.ieee.org/document/9884872/

[158] J. Tolan, H.-I. Yang, B. Nosarzewski, G. Couairon, H. V. Vo, J. Brandt, J. Spore, S. Majumdar, D. Haziza, J. Vamaraju, T. Moutakanni, P. Bojanowski, T. Johns, B. White, T. Tiecke, and C. Couprie, "Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar," *Remote Sensing of Environment*, vol. 300, p. 113888, Jan. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S003442572300439X

[159] P. Lopez-Dekker, J. Biggs, B. Chapron, A. Hooper, A. Kaab, S. Masina, J. Mouginot, B. B. Nardelli, C. Pasquero, P. Prats-Iraola, P. Rampal, J. Stroeve, and B. Rommen, "The Harmony Mission: End of Phase-0 Science Overview," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. Brussels, Belgium: IEEE, Jul. 2021, pp. 7752–7755. [Online]. Available: https://ieeexplore.ieee.org/document/9554896/

[160] J. Curlander and R. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, ser. Wiley Series in Remote Sensing and Image Processing. Wiley, 1992. [Online]. Available: https://books.google.de/books?id=nM98QgAACAAJ

Appendix A

SAR Theory

A.1 Azimuth Resolution

Based on the acquisition geometry shown in Fig. A.1, it is possible to describe the range distance between a target and the sensor using Pythagoras's theorem:

$$R^2(t) = (x - v_s t)^2 + y^2 + z^2 \tag{A.1}$$

where v_s is the platform travel speed in the azimuth direction, x is the azimuth coordinate of the target, y is its ground-range coordinate and z is the elevation coordinate of the platform.

The two-way pulse delay between the satellite and the target is given by:

$$\tau = \frac{2R(t)}{c} \tag{A.2}$$

while the Doppler shift for the target is given by:

$$f_D = \frac{2\dot{R}(t)}{\lambda} = \frac{2v_s \sin\theta}{\lambda} \approx \frac{2v_s x}{\lambda R} \tag{A.3}$$

where θ is the off-broadside angle between $R(t)$ and $R(0)$. For $t = 0$, it is possible to show that substituting Eq. (A.2) and Eq. (A.3) in Eq. (A.1), respectively, leads to:

$$(\frac{c\tau_0}{2})^2 = x^2 + y^2 + z^2 \tag{A.4}$$

$$(\frac{2vx}{\lambda f_D})^2 = x^2 + y^2 + z^2 \tag{A.5}$$

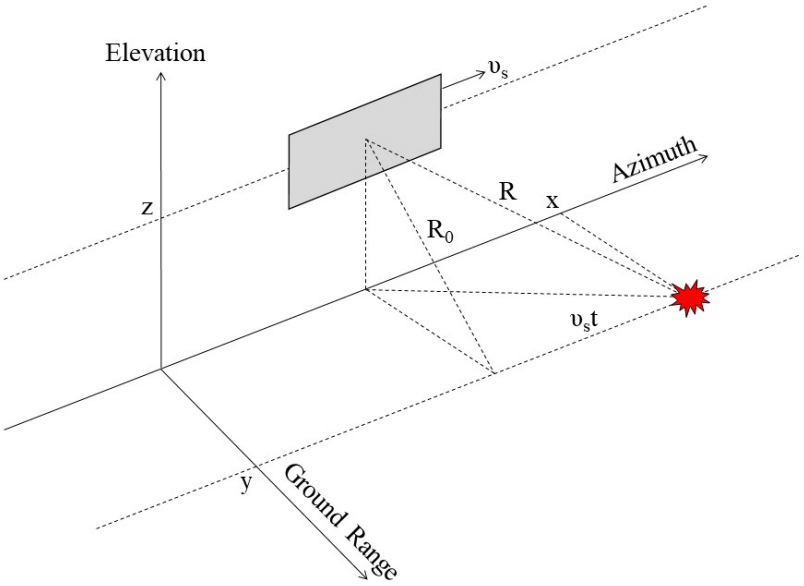


Figure A.1: Schematic representation of the SAR acquisition geometry, consisting of a platform at elevation z, moving at speed v_s in the azimuth direction, and of a target at distance R from the platform.

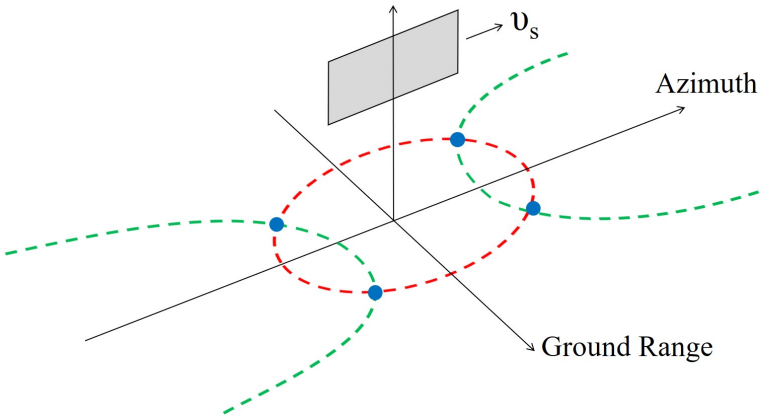


Figure A.2: Schematic representation of the SAR ambiguities for a given pulse delay time (red circle) and Doppler shift (green hyperbola). Their intersection defines four possible solutions, which can be reduced to one thanks to the SAR side-looking geometry and the sign of the Doppler shift.

where τ_0 is the pulse delay for $R(0)$. After rearranging, these lead to:

$$x^2 + y^2 = (\frac{c\tau_0}{2})^2 - z^2 \tag{A.6}$$

$$x^2[(\frac{2v_s}{\lambda f_D})^2 - 1] - y^2 = z^2 \tag{A.7}$$

where the first equation is that of the circle defined by all points with a corresponding pulse delay of τ_0, while the second equation is that of an hyperbola defined by the all points with Doppler shift f_D. As it can be seen in Fig. A.2, their intersection leads to four possible solutions, one placed in each quadrant defined by the azimuth and ground-range coordinate systems. With respect to the flight direction, the left-right ambiguity is resolved thanks to the side-looking geometry of the SAR system, while the front-back ambiguity is resolved by the sign of the Doppler shift, leading to a unique solution.

By using this Doppler analysis on the targets, it is possible to define their azimuth resolution in terms of their Doppler bandwidth δf_D, which in turn can be defined as:

$$\delta f_D = \frac{1}{T_e} \tag{A.8}$$

where T_e is the target exposure time, defined as the time interval within which the target is illuminated by the antenna footprint. It thus follows that:

$$T_e v = R\theta_{az} = R\frac{\lambda}{L_{az}} \tag{A.9}$$

Following Eq. (A.3) and Eq. (A.9), the azimuth resolution becomes:

$$\delta x = \frac{\lambda R}{2v_s}\frac{L_{az}v_s}{\lambda R} = \frac{L_{az}}{2} \tag{A.10}$$

with the notable consequence that the azimuth resolution does not depend anymore on the range distance, but only on the antenna physical dimension.

A.2 Azimuth Compression

Assuming that the antenna pointing direction is normal to the travel direction (i.e. zero squint angle), the distance between a point target and the sensor in the slant-range domain, as shown in Fig. A.3, is defined by the following hyperbole:

$$R(t) = \sqrt{R_0^2 + (v_s t)^2} \tag{A.11}$$

where t is the azimuth time, v is the speed of the platform and R_0 is the position of closest approach and corresponds to the zero Doppler position of the target. The consequence of this relationship is that, in practice, the power of a point target is not

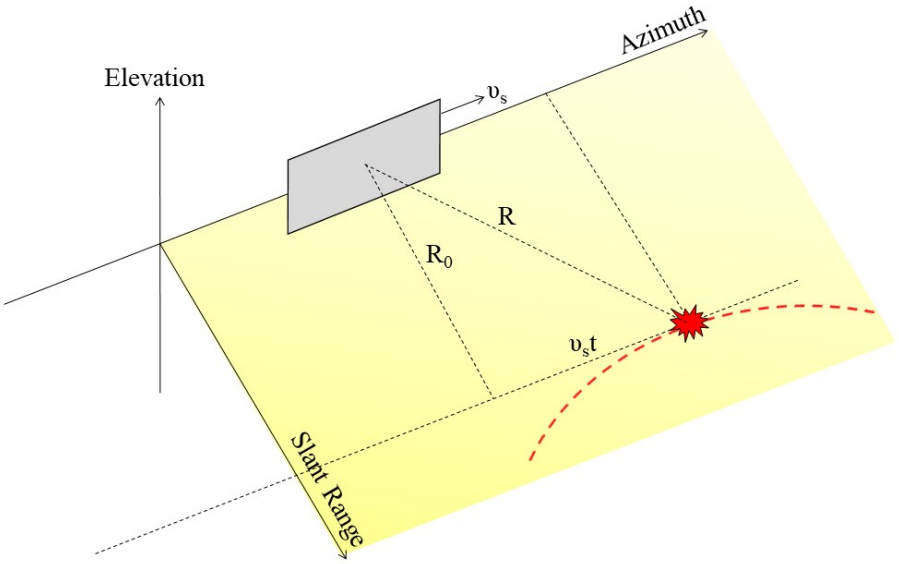


Figure A.3: Schematic representation of the range distance between the target (in red) and the SAR antenna (in grey) in time. Note that with respect to the slant range direction, the apparent range position R over time forms an hyperbole, whose peak coincides with the position of closest approach R_0.

only spread across the azimuth direction but also in the range direction. The resulting phenomenon is known as *range cell migration*, and it must be corrected before the signal can be focused in the azimuth direction. An example of range cell migration can be seen in the range focused image in Fig. 2.5.

By using a second order Taylor approximation, Eq. (A.11) can be simplified to the following parabola:

$$R(t) \approx R_0 + \frac{(v_s t)^2}{2R_0} \tag{A.12}$$

This allows one to estimate the two-way phase delay as:

$$\phi(t) = \frac{4\pi}{\lambda}\left(R_0 + \frac{(v_s t)^2}{2R_0}\right) \tag{A.13}$$

Thus, the received signal due to the motion in azimuth of the antenna can be expressed as:

$$s(t) = e^{-i\phi(t)} = e^{-i\frac{4\pi}{\lambda}R_0} e^{-i\frac{4\pi}{\lambda}\frac{(v_s t)^2}{2R_0}} \tag{A.14}$$

where the first part is a complex constant capturing the round trip delay at radius R_0, while the second part is a chirp signal. In other words, the relative motion between

the platform and the target produces a linear FM *instantaneous* or *Doppler frequency* $f_D(t)$

$$f_D(t) = -\frac{1}{2\pi}\frac{\partial\phi(t)}{\partial t} = -\frac{2v_s^2 t}{\lambda R_0} \tag{A.15}$$

and with a chirp rate defined by the *Doppler rate* κ_D

$$\kappa_D(t) = \frac{\partial f_D(t)}{\partial t} = -\frac{2v_s^2}{\lambda R_0} \tag{A.16}$$

A.3 The Speckle Model

As discussed in Sec. 2.1.4, in presence of distributed scatterers the observed power is the result of the coherent sum of N elementary scatterers within the resolution cell. Consisting in the constructive and destructive interference of the individual back-radiated EM waves, the resulting process is extremely sensitive to small changes in the individual scatterers. This results in the characteristic noise-like "salt and pepper" look of radar images, which is known as *speckle*. Notably, the measured value is the result of the specific geometrical arrangement of the individual scatterers, and is thus deterministic and (in principle) repeatable. The backradiated complex signal, modulated by the distributed scatterers, can be described by the following relationship:

$$\sqrt{\sigma}e^{i\phi} = \sum_{k=1}^{N}\sqrt{\sigma_k}e^{i\phi_k} \tag{A.17}$$

where $\sqrt{\sigma}$ is the amplitude of the complex signal and ϕ is its phase.

The speckle model gives a statistical interpretation to the process: assuming that each resolution cell (i.e. pixel) is characterized by a specific RCS value σ, the measured intensity $\hat{\sigma}$ represents a realization of a random variable parameterized by the true σ. The mean intensity of the resolution cell is given by:

$$\mathrm{E}[\hat{\sigma}] = \sum_{j=1}^{N}\sum_{k=1}^{N}\mathrm{E}\left[\sqrt{\sigma_j\sigma_k}e^{i(\phi_k-\phi_k)}\right] \tag{A.18}$$

Assuming that the amplitudes are independent from the phases, and that the phases of the elementary scatterers are independent from each other:

$$\mathrm{E}[\hat{\sigma}] = \sum_{k=1}^{N}\mathrm{E}[\sigma_k] = \sigma \tag{A.19}$$

meaning that the the mean intensity σ is the result of the incoherent sum of the mean intensities of the elementary scatterers. Dropping the assumption of phase and amplitude independence, the observed mean intensity is the result of a random walk in the complex plane. In practice, the observed in-phase and quadrature components of the observed signals are independent and identically distributed (i.i.d.) random variables with a zero mean Gaussian distribution with variance 0.5σ [160][41]:

$$\sqrt{\sigma}\cos(\phi) \sim \mathcal{N}(0, \sigma/2) \tag{A.20}$$

$$\sqrt{\sigma}\sin(\phi) \sim \mathcal{N}(0, \sigma/2) \tag{A.21}$$

As a consequence, this leads to the phase term being uniformly distributed in $[-\pi; \pi]$. This is an expected consequence, as in practice the range resolution dimension equates to multiple phase cycles (e.g. ≈ 16 for a TerraSAR-X acquisition with a 9.65 GHz center frequency and a 300 MHz bandwidth), and the random sampling of it values intuitively approximates a uniform distribution. On the other hand, the amplitude presents a Rayleigh distribution, while the intensity shows the following negative exponential distribution:

$$p(\hat{\sigma}) = \frac{1}{\sigma} e^{-\frac{\hat{\sigma}}{\sigma}} \tag{A.22}$$

which leads to the multiplicative noise interpretation of speckle.

In order to reduce the effects of speckle and, conversely, to improve the estimation of σ, it is possible to incoherently average multiple realizations sampled from the same distribution. This process takes the name of *multi-looking*, and the number of aggregated realizations N_L defines the number of considered *looks*. An example is presented in Fig. A.4. The use of N_L looks leads to a reduction in the speckle standard deviation of a factor $1/\sqrt{N_L}$. In practice, multi-looking can be implemented through either spatial or temporal averaging, the choice between which is typically driven by pragmatic considerations, such as the availability of multiple acquisitions or the target application.

A.4 Flat-Earth Removal

It is possible to estimate the flat-Earth contribution by considering two adjacent targets separated by incidence angle variation $\Delta\theta$ and placed at the same elevation. From



(a) (b)

Figure A.4: An example of a SAR image of an agricultural area. (a) shows the characteristic speckle content related ti distributed targets in SAR imagery, while (b) shows the same area after applying a temporal multi-looking filter on a stack of 32 images. As can be seen, the *noise-like* effects of speckle are strongly reduced after the multi-looking process.

Eq. (2.31), it is possible to write:

$$\phi_{if}^{1} = -\frac{2m\pi}{\lambda} B \sin(\theta_{\text{inc}} - \beta) \tag{A.23}$$

$$\phi_{if}^{2} = -\frac{2m\pi}{\lambda} B \sin(\theta_{\text{inc}} - \beta + \Delta\theta) \tag{A.24}$$

Computing the difference then leads to:

$$\phi_{if}^{2} - \phi_{if}^{1} = -\frac{2m\pi}{\lambda} B \left[\sin(\theta_{\text{inc}} - \beta + \Delta\theta) - \sin(\theta_{\text{inc}} - \beta)\right] \tag{A.25}$$

Given that both points are at the same elevation, their topographic contributions cancel out leaving behind a residual component, which coincides with the flat-Earth contribution ϕ_{fe}. Re-writing of previous expression finally leads to:

$$\phi_{fe} = \Delta\phi = \frac{2m\pi}{\lambda} \Delta\theta B \cos(\theta_{\text{inc}} - \beta) \approx \frac{2m\pi B_{\perp} \Delta R_{sr}}{\lambda R \tan(\theta_{\text{inc}})} \tag{A.26}$$

where ΔR_{sr} is the slant range separation between the points, and $B_{\perp} = B \cos(\theta_{\text{inc}} - \beta)$ is the perpendicular baseline.

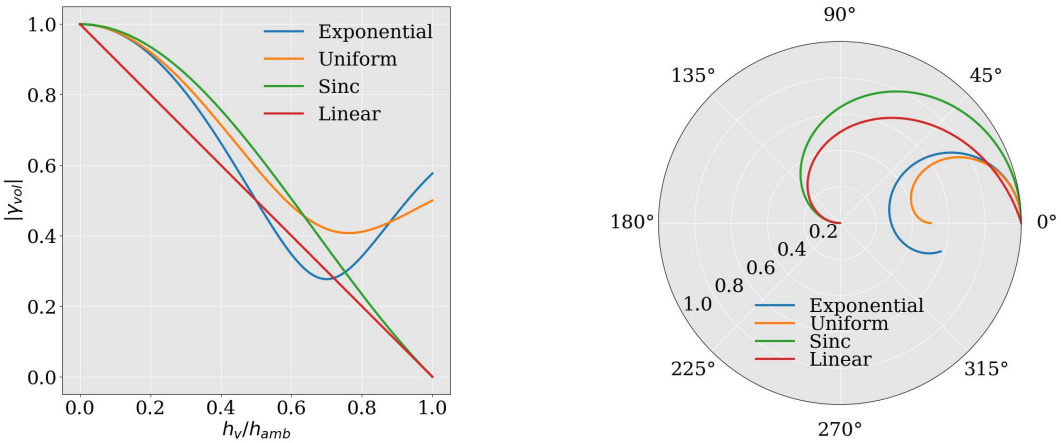


Figure A.5: Plots comparing the prediction outcomes of the RVoG model for: (blue) the full model using the exponential profile and a non-zero ground-to-volume amplitude ratio, (orange) the model using the uniform vertical profile resulting from the assumption of a null extinction coefficient, (green) the sinc model resulting from the additional assumption of a null ground-to-volume amplitude ratio and (red) the linear approximation of the sinc model. (a) shows the coherence magnitude versus the height of ambiguity normalized vegetation height, while (b) shows the rotation of the complex coherence vector for increasing vegetation heights.

A.5 RVoG Dual- and Single-Pol Inversion

In the case of dual-pol data, it is sufficient to assume to have zero ground-to-volume amplitude ratio (i.e. $m(\vec{w}) = 0$) in at least one polarization $\vec{w}$ [28]. Notably, this assumption is not always realistic [78], especially for longer wavelengths such as L-band and P-band, and thus requires to pre-determine the polarization $\vec{w}_{min}$ corresponding to the lowest $m(\vec{w}_{min})$ in order to minimize the error introduced by this assumption. The solution to the dual-pol inversion problem can then be found by solving the following system:

$$\begin{bmatrix} \hat{h}_v \\ \hat{\sigma}_{ext} \\ \hat{\phi}_0 \\ \hat{m}(\vec{w}_{max}) \end{bmatrix} = \underset{h_v,\sigma_{ext},\phi_0,m(\vec{w}_{max})}{\arg\min} \left|\left| \begin{bmatrix} \hat{\gamma}(\vec{w}_{max})e^{-i\phi_0} \\ \hat{\gamma}(\vec{w}_{min})e^{-i\phi_0} \end{bmatrix} - \begin{bmatrix} \hat{\gamma}_V(h_v,\sigma_{ext},m(\vec{w}_{max})) \\ \hat{\gamma}_V(h_v,\sigma_{ext},0) \end{bmatrix} \right|\right| \quad \text{(A.27)}$$

In the case of single-pol data, the assumption of zero ground-to-volume amplitude ratio is necessary but not sufficient for getting a balanced inversion problem. It is then possible to either use ancillary data to estimate one of the parameters, at the cost of flexibility due to the additional data requirement, or to introduce a further assumption, thus further reducing the capacity of the model and its predictive accuracy [28].

The former case is commonly implemented by introducing an external (LiDAR-based) *Digital Terrain Model* (DTM), in order to estimate the ground topographic phase by converting height to phase though as:

$$e^{i\phi_{DTM}} = e^{ih_{DTM}\kappa_k} \tag{A.28}$$

Note that a residual phase offset ϕ_Δ between ϕ_{DTM} and $\angle\hat{\gamma}$ must be compensated for through mutual calibration, which is commonly performed over bare soil (i.e. points with high coherence, e.g. $|\hat{\gamma}| \geq 0.98$) [28][83]. The solution to the single-pol inversion problem can then be found by solving the following:

$$\begin{bmatrix} \hat{h}_v \\ \hat{\sigma}_{ext} \end{bmatrix} = \underset{h_v,\sigma_{ext}}{\arg\min} \left|\left| \hat{\gamma}(\vec{w})e^{-i\phi_0} - \hat{\gamma}_V(h_v, \sigma_{ext}) \right|\right| \tag{A.29}$$

As mentioned above, it is alternatively possible to fix another of the remaining unknowns. A common choice for this is to set the extinction coefficient to be constant (i.e. $\sigma_{ext} = \sigma_{ext,0}$.), which equates to fixing the volume scattering profile $f_v(z, \vec{w})$ [83]. The alternative solution to the single-pol inversion problem can then be found by solving:

$$\begin{bmatrix} \hat{h}_v \\ \hat{\phi}_0 \end{bmatrix} = \underset{h_v,\phi_0}{\arg\min} \left|\left| \hat{\gamma}(\vec{w})e^{-i\phi_0} - \hat{\gamma}_V(h_v, \sigma_{ext,0}) \right|\right| \tag{A.30}$$

Note that the choice of setting $\sigma_{ext} = 0$ leads to the particular solution for a constant, instead of exponential, volume scattering profile [29]. In particular, this allows to define:

$$\hat{\gamma}(\vec{w}) = \exp(i\kappa_z z_0 + i\frac{h_v\kappa_z}{2})\,\mathrm{sinc}(\frac{h_v\kappa_z}{2}) \tag{A.31}$$

with the notable consequence that the magnitude is solely defined by the real-valued sinc function. It is therefore possible to estimate the forest height from the inversion of the sinc function, using only the sample coherence magnitude as [81]:

$$\hat{h}_v = \frac{\eta}{k_z}\left(\pi - 2\sin^{-1}(|\hat{\gamma}|^{0.8})\right) \tag{A.32}$$

where $\eta = 2$ determines the ratio between the forest height and the penetration depth [81]. Thus, the sinc model in Eq. (A.31) characterizes the penetration depth into the canopy by assuming that the scattering phase center lies halfway between the top of the canopy and the ground. This assumption is clearly sub-optimal and leads to varying levels of accuracy depending on the properties of the considered forest. It is therefore

common to calibrate the value of η from some reference height measurements, in order to best capture the local properties (e.g. $\eta = 2.5$ in [43]).

Finally, it is possible to simplify the RVoG even more by using a linear approximating of the sinc function, thus leading to [29]:

$$\hat{\gamma}(\vec{w}) = \exp(i\kappa_z z_0 + i\frac{h_v\kappa_z}{2})(1 - \frac{h_v\kappa_z}{2\pi}) \tag{A.33}$$

which can be inverted to obtain:

$$\hat{h}_v = \frac{2\pi}{\kappa_z}(1 - |\hat{\gamma}|) = h_{amb}(1 - |\hat{\gamma}|) \tag{A.34}$$

Fig. A.5 shows a comparison of the relative prediction deviations for the RVoG model under different assumptions. Taking the full exponential model for reference, it is possible to observe from Fig. A.5 (a) that all models achieve good agreement for the coherence magnitude values corresponding to normalized vegetation heights of up to $h_v/h_{amb} \approx 0.7$, beyond which the impact of the ground-component causes the predictions to diverge.

Appendix B

Estimation Theory

B.1 Maximum Likelihood Estimator

The MLE is defined as:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} := \arg\max_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta}) \tag{B.1}$$

By assuming that the data samples are independent and identically distributed (i.e. *iid* assumption), the MLE becomes:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max_{\boldsymbol{\theta}} \left\{ \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}) \right\} \tag{B.2}$$

It is common to express the relationship in terms of *negative log likelihood* (NLL):

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\min_{\boldsymbol{\theta}} \{\text{NLL}(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \left\{ -\sum_{i=1}^{n} log\, p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}) \right\} \tag{B.3}$$

If the output's conditional probability distribution can be modeled as a *Gaussian* random variable, such that:

$$p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}) \sim \mathcal{N}(y_i|f(\boldsymbol{x}_i, \boldsymbol{\theta}), \sigma^2) \tag{B.4}$$

under the assumption of homoscedasticity (i.e. $\sigma^2 = const.$), the NLL becomes:

$$\text{NLL}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} log \left[\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2}{2\sigma^2} \right) \right] \tag{B.5}$$

$$= -\frac{n}{2} log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 \tag{B.6}$$

where minimizing the NLL with respect to $\boldsymbol{\theta}$ is independent of the first constant term and the scaling factor of the second term, ultimately yielding:

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 \right\} \tag{B.7}$$

B.2 Backpropagation Example

Let us consider a single input, single output MLP model $F(x_i, \boldsymbol{\Theta})$ of four layers, each consisting of a single neuron with all but the last one using a ReLU activation function Φ, and let us use a least squares loss function. The first step then consists in computing the intermediate products at each step, to obtain the following:

$$\begin{aligned} f_0 &= \theta_0 x_i + \beta_0 \\ a_1 &= \Phi(f_0) \\ f_1 &= \theta_1 a_1 + \beta_1 \\ a_2 &= \Phi(f_1) \\ f_2 &= \theta_2 a_2 + \beta_2 \\ a_3 &= \Phi(f_2) \\ f_3 &= \theta_3 a_3 + \beta_3 \end{aligned}$$

$$\mathcal{L}_i(\boldsymbol{\Theta}) = (f_3 - y_i)^2 \tag{B.8}$$

with x_i and y_i being the i^{th} training input and output samples, respectively.

In the second step, first the partial derivatives of the sample loss $\mathcal{L}_i$ are computed

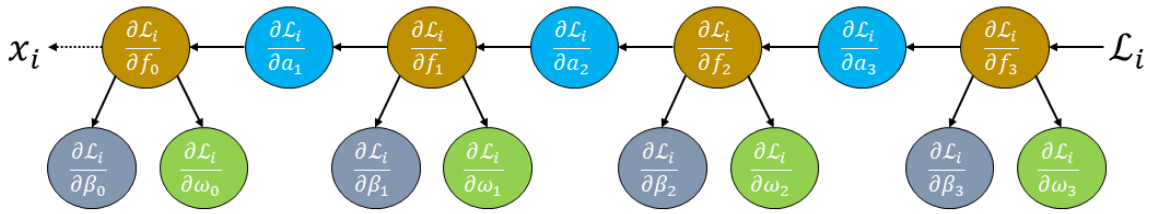


Figure B.1: Visualization of the backpropagation example in Sec. 3.2.4. Starting from the output, the loss gradient is iteratively computed with respect to each intermediate product using the chain rule. The gradient at each intermediate product is subsequently used to compute the loss gradient with respect to the corresponding weights and biases.

with respect to the intermediate products in (B.8) in inverse order, to obtain:

$$\begin{aligned}
\frac{\partial \mathcal{L}_i}{\partial f_3} &= 2(f_3 - y_i) \\
\frac{\partial \mathcal{L}_i}{\partial a_3} &= \left[\frac{\partial \mathcal{L}_i}{\partial f_3}\right] \frac{\partial f_3}{\partial a_3} \\
\frac{\partial \mathcal{L}_i}{\partial f_2} &= \left[\frac{\partial \mathcal{L}_i}{\partial f_3}\frac{\partial f_3}{\partial a_3}\right] \frac{\partial a_3}{\partial f_2} \\
\frac{\partial \mathcal{L}_i}{\partial a_2} &= \left[\frac{\partial \mathcal{L}_i}{\partial f_3}\frac{\partial f_3}{\partial a_3}\frac{\partial a_3}{\partial f_2}\right] \frac{\partial f_2}{\partial a_2} \\
\frac{\partial \mathcal{L}_i}{\partial f_1} &= \left[\frac{\partial \mathcal{L}_i}{\partial f_3}\frac{\partial f_3}{\partial a_3}\frac{\partial a_3}{\partial f_2}\frac{\partial f_2}{\partial a_2}\right] \frac{\partial a_2}{\partial f_1} \\
\frac{\partial \mathcal{L}_i}{\partial a_1} &= \left[\frac{\partial \mathcal{L}_i}{\partial f_3}\frac{\partial f_3}{\partial a_3}\frac{\partial a_3}{\partial f_2}\frac{\partial f_2}{\partial a_2}\frac{\partial a_2}{\partial f_1}\right] \frac{\partial f_1}{\partial a_1} \\
\frac{\partial \mathcal{L}_i}{\partial f_0} &= \left[\frac{\partial \mathcal{L}_i}{\partial f_3}\frac{\partial f_3}{\partial a_3}\frac{\partial a_3}{\partial f_2}\frac{\partial f_2}{\partial a_2}\frac{\partial a_2}{\partial f_1}\frac{\partial f_1}{\partial a_1}\right] \frac{\partial a_1}{\partial f_0}
\end{aligned} \tag{B.9}$$

where the previously computed results are put into square brackets. The visual example for the implementation of the backpropagation algorithm is depicted in Fig. B.1.

B.3 Sources of Error

It is now necessary to elaborate on the possible sources of error that can affect the performance of a model and what these depend on.

As described in Sec. 3.1.1, in every data generation process there is some intrinsic amount of randomness (called stochastic uncertainty) that fundamentally prevents a model from outputting the exact prediction for a given input. Assuming that this

source of *noise* is additive and has zero mean and variance σ^2, the output can be modelled as an aleatoric variable with probability density function $p(y|x)$, where its mean value $\mu_y(\boldsymbol{x})$ and variance σ_y^2 are defined as:

$$\mu_y(\boldsymbol{x}) = \mathbb{E}_y[y(\boldsymbol{x})] = \int y(\boldsymbol{x}) p(y|\boldsymbol{x}) dy \tag{B.10}$$

$$\sigma_y^2 = \mathbb{E}_y[(\mu_y(\boldsymbol{x}) - y(\boldsymbol{x}))^2] = \int (\mu_y(\boldsymbol{x}) - y(\boldsymbol{x}))^2 p(y|\boldsymbol{x}) dy \tag{B.11}$$

having, $\sigma_y^2 = \sigma^2$. Let us now consider a linear regression model $f(\boldsymbol{x}, \boldsymbol{\theta})$ and train the parameters $\boldsymbol{\theta}$ using the least squares criteria. Given the stochastic nature of y, one can compute the expected loss value as:

$$\begin{aligned}
\mathbb{E}_y[\mathcal{L}(\boldsymbol{x})] &= \mathbb{E}_y[(f(\boldsymbol{x}, \boldsymbol{\theta}) - y(\boldsymbol{x}))^2] \\
&= \mathbb{E}_y[(f(\boldsymbol{x}, \boldsymbol{\theta}) - y(\boldsymbol{x}) + \mu_y(\boldsymbol{x}) - \mu_y(\boldsymbol{x}))^2] \\
&= \mathbb{E}_y[(f(\boldsymbol{x}, \boldsymbol{\theta}) - \mu_y(\boldsymbol{x}))^2 + 2(f(\boldsymbol{x}, \boldsymbol{\theta}) - \mu_y(\boldsymbol{x}))(\mu_y(\boldsymbol{x}) - y(\boldsymbol{x})) + (\mu_y(\boldsymbol{x}) - y(\boldsymbol{x}))^2] \\
&= (f(\boldsymbol{x}, \boldsymbol{\theta}) - \mu_y(\boldsymbol{x}))^2 + 2(f(\boldsymbol{x}, \boldsymbol{\theta}) - \mu_y(\boldsymbol{x}))(\mu_y(\boldsymbol{x}) - \mathbb{E}_y[y(\boldsymbol{x})]) + \mathbb{E}_y[(\mu_y(\boldsymbol{x}) - y(\boldsymbol{x}))^2] \\
&= (f(\boldsymbol{x}, \boldsymbol{\theta}) - \mu_y(\boldsymbol{x}))^2 + \sigma^2
\end{aligned} \tag{B.12}$$

Therefore, the expected loss value, given the aleatoric nature of the output, can be expressed as the sum of the squared deviation between the model prediction and the expect output value, and of the intrinsic noise itself.

In practice, the model parameters depend on the training set $\mathcal{D}$, which is itself a random sample of the underlying data generation process. In other terms, the parameters found by solving the optimization problem change if $\mathcal{D}$ changes. Thus, taking into account the training set, the expected model output can be written as:

$$\mu_f(\boldsymbol{x}) = \mathbb{E}_{\mathcal{D}}[f(\boldsymbol{x}, \boldsymbol{\theta}|_{\mathcal{D}})] \tag{B.13}$$

As a consequence, the expected value of the squared model deviation is defined as:

$$\mathbb{E}_{\mathcal{D}}[(f(\boldsymbol{x}, \boldsymbol{\theta}|_{\mathcal{D}}) - \mu_y(\boldsymbol{x}))^2] = \mathbb{E}_{\mathcal{D}}[(f(\boldsymbol{x}, \boldsymbol{\theta}|_{\mathcal{D}}) - \mu_f(\boldsymbol{x}))^2] + (\mu_f(\boldsymbol{x}) - \mu_y(\boldsymbol{x}))^2 \tag{B.14}$$

where $\mathbb{E}[(f(\boldsymbol{x}, \boldsymbol{\theta}|_{\mathcal{D}}) - \mu_f(\boldsymbol{x}))^2]$ is the model variance and $(\mu_f(\boldsymbol{x}) - \mu_y(\boldsymbol{x}))$ is the bias of the model. Therefore, Eq. (B.12) can be re-written as:

$$\mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_y[\mathcal{L}(\boldsymbol{x})]\right] = \mathbb{E}_{\mathcal{D}}[(f(\boldsymbol{x}, \boldsymbol{\theta}|_{\mathcal{D}}) - \mu_f(\boldsymbol{x}))^2] + (\mu_f(\boldsymbol{x}) - \mu_y(\boldsymbol{x}))^2 + \sigma^2 \tag{B.15}$$

This means that, after having considered the uncertainties of both the training and test sets, the expected error of a linear regression model fitted using the least squares criteria can be expressed as the sum of the following three terms:

- The model variance, which captures the uncertainty of the model, trained with a certain training set.
- The model bias, which captures the systematic offset of the trained model with respect to the true underlying function.
- The noise power, which captures the inherent uncertainty of the true mapping process from input to output.

It follows that the variance of the model can be reduced by using a larger and more representative training set, since the noise caused by the sampling process reduces as the training set increases and better represents the feature space.

Similarly, the bias of the model can also be improved. In this case, a large bias implies that the model struggles to fit the underling relationship. Thus, it is possible to increase the size of the network, to improve its modelling capacity.

Intuitively, increasing the size of a network equates to an increase of its trainable parameters, thus a larger dataset is required for its training. In practice, the size of the available training set is limited, which in turn also limits the maximum model size that can be trained. This co-dependency between model complexity and training set size, is at the core of the *bias-variance trade-off*.