UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

Deep convolutional regression modelling for forest parameter retrieval

Sara Björk

A dissertation for the degree of Philosophiae Doctor - September 2023

Abstract

Accurate forest monitoring is crucial as forests are major global carbon sinks. Additionally, accurate prediction of forest parameters, such as forest biomass and stem volume (SV), has economic importance. Therefore, the development of regression models for forest parameter retrieval is essential.

Existing forest parameter estimation methods use regression models that establish pixel-wise relationships between ground reference data and corresponding pixels in remote sensing (RS) images. However, these models often overlook spatial contextual relationships among neighbouring pixels, limiting the potential for improved forest monitoring. The emergence of deep convolutional neural networks (CNNs) provides opportunities for enhanced forest parameter retrieval through their convolutional filters that allow for contextual modelling. However, utilising deep CNNs for regression presents its challenges. One significant challenge is that the training of CNNs typically requires continuous data layers for both predictor and response variables. While RS data is continuous, the ground reference data is sparse and scattered across large areas due to the challenges and costs associated with *in situ* data collection.

This thesis tackles challenges related to using CNNs for regression by introducing novel deep learning-based solutions across diverse forest types and parameters. To address the sparsity of available reference data, RS-derived prediction maps can be used as auxiliary data to train the CNN-based regression models. This is addressed through two different approaches.

Although these prediction maps offer greater spatial coverage than the original ground reference data, they do not ensure spatially continuous prediction target data. This work proposes a novel methodology that enables CNN-based regression models to handle this diversity. Efficient CNN architectures for the regression task are developed by investigating relevant learning objectives, including a new frequency-aware one. To enable large-scale and cost-effective regression modelling of forests, this thesis suggests utilising C-band synthetic aperture radar (SAR) data as regressor input. Results demonstrate the substantial potential of C-band SAR-based convolutional regression models for forest parameter retrieval.

Acknowledgements

Upon completion of this Ph.D. project, I take a moment to reflect on the scientific and personal growth I have experienced over the past few years. This journey would not have been possible without the support, discussions, and knowledge imparted by numerous individuals. For those, I would like to express my sincere gratitude.

First and foremost, I would like to extend my deepest thanks to my supervisor Professor Stian N. Anfinsen. Your guidance, support, and our discussions have been invaluable. Allow me to be personal: without your optimism and your belief in this Ph.D. project, I would probably not have completed it. However, I am glad to say that I did! Stian, through the years, I have also learned that a Ph.D. project is not only a scientific journey, but also a personal one. Although I am a person of words, meaning that I rather write too much than too little, I struggle to find the words to describe how grateful I am to have embarked on this personal journey with your presence on the sideline. Thank you!

I also want to thank my co-supervisor Professor Robert Jenssen for your excellent lectures in Pattern Recognition, which during my master's opened my eyes to machine learning. Thank you for fruitful discussions and valuable feedback during my Ph.D. project.

To Associate Professor Michael Kampffmeyer, thank you for your valuable input and for sharing your knowledge. You really possess the ability to find the needle in a haystack! Many thanks to Associate Professor Benjamin Ricaud and Dr. Ahcene Boubekki for valuable input and beneficial discussions.

I would also express my gratitude to Professor Erik Næsset and Professor Terje Gobakken for sharing their domain knowledge in forestry and for ALS data. Special thanks for your assistance in enabling access to the forest and ALS data used through this project.

Without access to ground reference measurements of forest parameters or ALS-derived predictions maps, this Ph.D. project would not have been possible. I gratefully acknowledge the Norwegian University of Life Sciences, the Tanzania

Forest Services Agency, Professor Eliakimu Zahabu and co-workers at Sokoine University of Agriculture, Viken Skog and the Swedish University of Agricultural Sciences for participation in field work and provision of in situ measurements, ALS-derived AGB and SV products. Special thanks to Dr. Lennart Noordermeer for providing access to the ALS-derived stem volume products, Professor Håkan Olsson for providing ALS data acquired by SLU and to Mr. Svein Dypsund at Viken Skog for providing in situ measurements in Norway.

To the members of Team Satellite: Jørgen and Luigi, what would this project be without you? Without peach pits throwing and discussions about MacBook? We did not make it to Hawaii, but we continue to orbit!

To Jonas and Thomas, thank you for the fruitful discussions that lead to the Fourier Odyssey.

Special thanks to everyone in the UiT Machine Learning group for the discussions, support, ideas and the enjoyable moments we have shared. It has been a pleasure to spend these years with you. I will never forget the experience of sharing a box of fermented herring, on a cold, snowy day in the fall. You are amazing!

I would like to thank my committee members for taking the time to read my thesis and attending the defence.

At this crossroads, which finishing this thesis marks, I am very much looking forward to bringing my scientific and personal experiences with me into the future. I would like to express some final words of thanks.

To KSAT Kongsberg Satellite Services, thank you for supporting me in completing this Ph.D. project. I look forward to the exciting projects that lie ahead for us and Team Object Detection!

To my friends, thank you for your support throughout these years. I look forward to spending more time with you skiing, hiking, sharing more books, knitting, talking, and to bring my sewing machine with me on new adventures! Last, but not least, to my family and to Lars-Eirik, thank you for being with me throughout these years, thank you for the unwavering support and patience. I could not have done this without you, thank you!

Sara Maria Björk Tromsø, June 2023

Contents

List of Figures

List of Tables

List of Abbreviations

AE Autoencoder

AGB Aboveground Biomass

AI Artificial Intelligence

ALS Airborne Laser Scanning

AOI Area of Interest

AzI Azimuthal Integral

BGB Belowground Biomass

BN Batch Normalisation

cGAN Conditional Generative Adversarial Network

CNN Convolutional Neural Network

CV Cross validation

DEM Digital Elevation Model

DL Deep Learning

GAN Generative Adversarial Network

I2I Image-to-Image

IPCC Intergovernmental Panel on Climate Change

LiDAR Light Detection and Raging

lr Learning Rate

MAE Mean Absolute Error

ML Machine Learning

MLP Multilayer perceptron

MSE Mean Squared Error

NFI National Forest Inventory

NN Neural Network

RAR Real Aperture Radar

ReLU Rectified Linear Unit

ResNet Residual Network

RMSE Root Mean Squared Error

Rose-L Radar Observing System for Europe in L-band

RS Remote Sensing

SAR Synthetic Aperture Radar

SGD Stochastic Gradient Decent

SV Stem Volume

tanh Hyperbolic tangent

VAE Variational Autoencoder

XAI Explainable Artificial Intelligence

1 Introduction

Since the first aerial photographs were taken in 1909 [1, 2], there has been a tremendous increase in new RS missions for earth observation, including unmanned aerial vehicles, airborne and spaceborne sensors. These have further led to various surveillance and monitoring applications that rely heavily on RS images. The forestry sector, in particular, extensively utilises RS sensors due to their ability to efficiently map vast land areas and remote regions. The applications of RS in forestry encompass diverse areas, including but not limited to change detection, wildfire mapping, storm damage monitoring and forest parameter retrieval [2–6].

Due to climate change, accurate monitoring of aboveground forest biomass (AGB) becomes essential. Forests can, for example, store more carbon than the atmosphere, making them one of the largest global carbon sinks [7, 8]. Furthermore, from an economic perspective, accurately measuring, monitoring, and predicting AGB is essential to estimate factors like the availability of raw materials and the potential for bioenergy [9, 10]. Considering that the forest SV constitutes a significant proportion of the biomass of each tree, typically ranging from 65% to 80% [11–13], both overall forest AGB and SV are two essential forest parameters to monitor and predict.

For forest parameter retrieval, a small collection of ground reference data is commonly associated with RS data through regression models [14–16]. The purpose is to establish a relationship between measured RS backscatter and biophysical forest parameters such as AGB or commercial SV. Traditional linear

statistical models, such as multiple linear regression, have conventionally been adopted for this task. The evolution of machine learning methods in the last decades has introduced new algorithms for forest parameter retrieval. Especially popular for this task are random forests, support vector machines for regression, or fully connected neural networks (NNs) [15,17]. These methods can capture complex nonlinear relationships between individual points of ground reference data and corresponding RS image pixels. However, a drawback of machine learning-based methods is that these generally do not incorporate spatial contextual relationships from neighbouring image pixels in the learning process [17].

The field of computer vision and pattern recognition has undergone a significant transformation in recent years, thanks to advancements in computational power, the availability of large labelled image databases, and the emergence of deep CNNs [18]. Unlike traditional machine learning methods, the power of CNNs lies in their use of convolutional filters, which allow the image data to be processed in blocks. This enables the model to incorporate the spatial neighbourhood of each pixel in the learning process [3,4,18]. Thus, CNN-based regression models offer great potential for exploiting and modelling the complex relationships between RS data and forest parameters.

Training of CNNs typically requires the predictor and response variables to be continuous data layers sampled on matching grids. However, this requirement poses a challenge in the context of forest remote sensing. While RS data are continuous, the available ground reference data for forests is limited to a sparse set of spatially scattered measurements due to the challenges and costs associated with *in situ* data collection [15,19]. Consequently, it becomes difficult to effectively utilise regression models that rely on convolutional filters when the ground reference data is spatially discrete and cannot be easily aligned with the input data.

The focus of this thesis is to develop methodologies for advancing forest parameter retrieval through the use of deep convolutional regression models. Challenges related to this objective are briefly outlined in the following section and further addressed in the included papers of this thesis.

1.1 Key challenges

The main objective of this thesis is to address three key challenges associated with the application of deep learning methods in the retrieval of forest parameters: (1) the lack of reference data to use as prediction targets, (2) the diversity in spatial coverage of RS-derived prediction maps, and (3) the applicability of

forest regression models. These challenges will be explained and addressed in the following. Detailed discussion on these challenges will be presented in Chapter 3 and Chapter 6.

Lack of reference data to use as prediction targets: Central to the development of accurate regression models for forestry is the need for a sufficient amount of regressor (input) and regressand (target) data. By utilising RS data as input to the model, one ensures that the availability of regressor data is not a limiting factor during model training. However, in forestry, the collection of ground reference data for parameters such as AGB poses limitations due to the labour-intensive, time-consuming, and expensive nature of field inventory campaigns [4, 15]. Consequently, the target data comprising ground reference measurements are sparse. This sparsity has a dual aspect, referring to both the limited number of collected samples and the spatially scattered distribution of ground reference measurements within the area of interest (AOI).

In traditional statistical or machine learning-based regression models, having limited access to prediction targets is typically not an issue, as these models commonly establish a direct relationship at the pixel level. However, as previously mentioned, training of CNNs requires continuous data layers for both predictor and response variables. Consequently, the absence of continuous distributed ground reference measurements poses a significant challenge to the employment of CNN-based models for forest parameter retrieval [4].

Diversity in spatial coverage of RS-derived prediction maps: In this work, RS-derived prediction maps are used as auxiliary data to remedy the sparsity of the available reference data. For instance, due to the strong correlation between airborne laser scanning (ALS) data and e.g. forest height, thereby also to AGB and SV [2, 20], ALS-derived prediction maps can serve as a reliable alternative or complement to a sparse set of ground reference data[1]. Consequently, by training regression models to relate RS data to e.g. ALS-derived forest products, one obtains a denser dataset of prediction targets and facilitates CNN-based regression models.

The acquisition cost of ALS data is high [7, 21, 22], which means that economic considerations often limit the feasibility of obtaining wall-to-wall coverage of ALS data in large regions. One particular example is the study of Ene *et al.* [22] in the Liwale district of Tanzania, where full-coverage ALS mapping was economically infeasible for a study related to the national field inventory (NFI) campaign. Instead, ALS data were acquired as 32 strips, each with a swath width of 1350 m and a spatial separation of 5 km, as shown in Figure 1.1. The swath of the ALS instrument determined the width of the data stripes, and the

1. This is further discussed in Section 3.3.1 and Section 6.3.2.

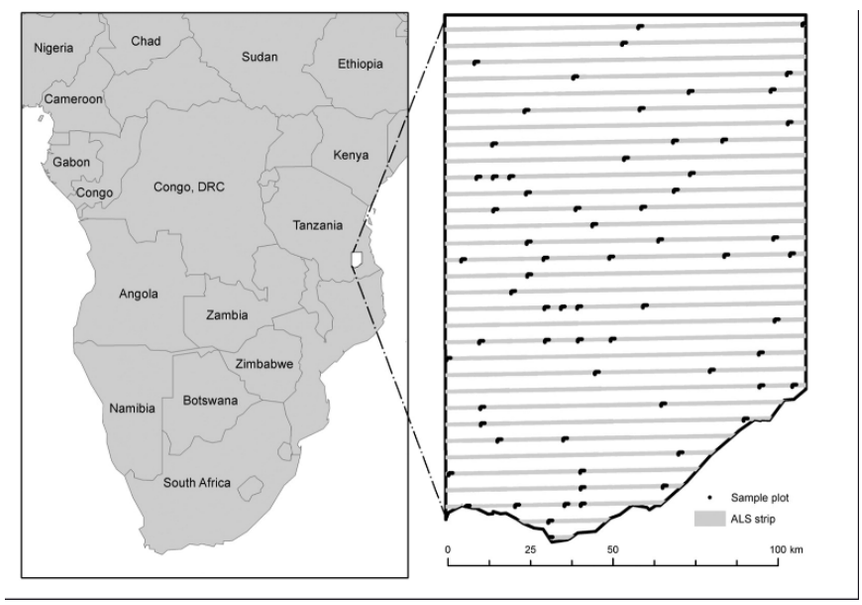


Figure 1.1: Left: Illustration of the AOI in the administrative district of Liwale, Tanzania. **Right:** Distribution of the 32 parallel strips, from which ALS data were acquired, and the associated field plots, where ground reference data were acquired. Image retrieved from [22].

gaps were as narrow as the data budget could allow, given that stripes of data should be captured across the entire district of Liwale to ensure representative coverage of the forest [22]. Without extrapolating, the resulting ALS-derived AGB prediction maps will only cover the same 32 strips of the AOI.

Another example is the work by Noordermeer *et al.* [23], who developed regression models for SV in managed Norwegian boreal forests using ground reference and ALS data. After model fitting, the generation of ALS-derived SV prediction maps was limited to areas where the forest height exceeded 8-9 meters. As further discussed in Section 3.4.2, the height constraint resulted in spatially disjoint ALS-derived SV prediction maps, as shown in Figure 1.2. Thus, although RS-derived prediction maps offer greater spatial coverage than the original ground reference data, and can be used as a complementary source of training data, they do not ensure spatially continuous prediction target data. Instead, these products introduce challenges related to partly continuous target data, which must be considered before utilising this data to train CNN-based regression models in forestry.

Applicability of forest regression models: RS-based regression models provide significant advantages to the forestry sector by enabling the measurement, monitoring, and prediction of forest parameters, including AGB, on a large

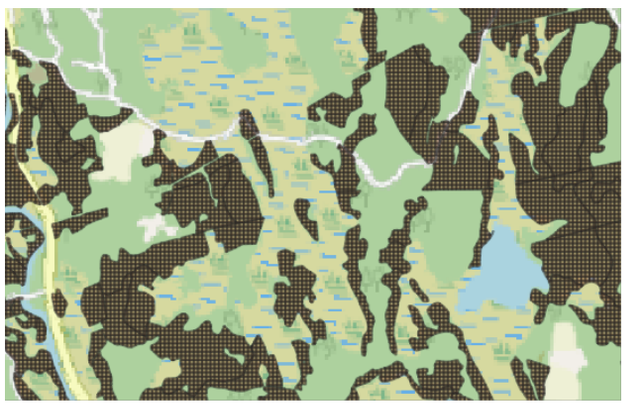


Figure 1.2: A small section of the disjoint ALS-derived SV prediction maps. The dataset is retrieved from the northern parts of Nordre Land.

scale [7,15,21]. However, for these models to be of practical use in operational settings, they must provide accurate predictions. While the convolutional filters in the CNN offer the potential to capture spatial contextual relationships within the data, the model's performance is influenced by several factors, including the suitability of the input and target data and choosing a suitable learning objective. In the context of this thesis, the CNN model has to learn from images outside the natural image domain, such as RS images. While RS image data hold much higher frequencies than natural images [24], the learning process may be challenging, as studies have shown that NNs have a bias against learning the high-frequency content of images [25,26]. Thus, choosing a suitable learning objective implies that CNN-based regression models have to capture the full dynamic range of RS image data to perform accurately in terms of mean absolute error (MAE) or root mean squared error (RMSE), which are the relevant metrics for the regression task.

The suitability of the input and target data is both related to the informativeness of the data, and whether the data meet region-specific or application-specific constraints. While ALS-based regression models are known to be more accurate than models that use radar or optical data [21,27], the associated acquisition cost can restrict its usefulness [19]. In initiatives like the REDD+ program[2], the economic aspect becomes even more crucial, favouring the use of freely available data. Furthermore, the input data should be regularly updated over the AOI to ensure the model's operational usefulness and provide accurate predictions. In certain cases, such as tropical and Arctic regions, it is advantageous if RS data can be acquired regardless of the weather or lighting conditions.

2. The official name of the REDD+ program is "Reducing emissions from deforestation and forest degradation, plus the sustainable management of forests, and the conservation and enhancement of forest carbon stocks".

1.2 Key objectives

To address the aforementioned challenges, this thesis proposes novel methodologies for deep convolutional regression modelling for forest parameter retrieval. The key objectives of this thesis are summarised as follows:

1. Develop methodologies that facilitate the utilisation of CNN-based regression models in forestry by incorporating spatially continuous or partially continuous sensor-derived forest products.

2. Develop methodology that enables CNN-based regression models to learn effectively from a sparse set of ground reference data.

3. Propose a novel loss function to mitigate the spectral bias of CNNs, to improve their learning and generation of the high-frequency content in images.

The contributions of this thesis lie in achieving these key objectives, either alone or in conjunction with one another. These objectives have implications in various ways. For instance, the third objective focuses on enhancing CNN-based regression models to effectively learn from image data, particularly RS data, which is known for its high-frequency information content.

Moreover, the contributions of this thesis extend beyond the forestry sector. The techniques and methodologies developed can be applied to convolutional regression of biophysical parameters in diverse fields since the challenge with limited and spatially scattered ground reference data also exists outside the forestry sector. The subsequent sections elaborate on how the three papers address the key objectives.

1.3 Key solutions

This thesis explores the potential for combining convolutional deep learning methodologies with RS data to advance present methods for forest parameter retrieval. The three papers included in the thesis address the outlined key objectives in different ways.

The first key objective is addressed in Paper I, which focuses on AGB prediction and proposes to utilise sequential modelling to bypass the limited amount of reference data, which are referred to in the following as the true prediction targets. Specifically, Paper I develops the second regression model in the sequence that utilises a wall-to-wall map of ALS-derived AGB predictions as a surrogate

for the true prediction targets, and spatially extensive RS data as regressor data. This is possible because, unlike, for example [22], the AOI studied in Paper I is restricted to a smaller region in the Liwale district of Tanzania for which wall-to-wall coverage of ALS data is available. The smaller AOI ensures that the surrogate prediction target is spatially continuous and enables the use of convolutional regression models based on the conditional generative adversarial network (CGAN). The models are trained to learn the relationship between regressor data from the Sentinel-1 sensor and the wall-to-wall maps of ALS-derived AGB predictions.

Paper III frames its proposed algorithm as a semi-supervised deep learning approach to train deep convolutional regression models on either continuous or partially continuous ALS-derived prediction maps, which are referred to as pseudo-targets. By imputing the pseudo-targets into the sparse set of true prediction targets, i.e. the ground reference data, Paper III address both key objective 1 and 2. The proposed semi-supervised imputation strategy enables the use of convolutional regression models for forest parameter retrieval.

The third key challenge is addressed in different ways in the three papers. In Paper II, a novel frequency-aware loss function is proposed. The frequency-aware loss function is complementary to other loss functions, implying that regression models can be trained to focus on learning both the high-frequency content of the image and other characteristics important for the regression task. While Paper II only evaluates the loss function on images from the natural image domain, the frequency-aware loss function was further evaluated in Paper III on RS images.

Both Papers I and III leverage openly accessible data from the Sentinel-1 sensor as regressor data. The Sentinel-1 sensor is an active C-band SAR sensor that offers dependable acquisition modes and schedules, making it a reliable choice as it can acquire data both at night and under cloudy conditions. As a result, SAR data obtained from the Sentinel-1 sensors facilitate large-scale and cost-effective regression modelling of forests. Moreover, once the regression models are optimised, recently collected Sentinel-1 data can be utilised to provide up-to-date predictions for AGB or SV.

As shown in [3], deep convolutional regression in vegetation applications, such as forestry, is limited in research. This is probably due to challenges related to using deep convolutional regression models in applications when data is sparse and spatially scattered [4]. By training convolutional regression models on ALS-derived forest parameter prediction maps, this thesis contributes to bridging the gap between the sparse and spatially scattered ground reference data and the use of convolutional regression models.

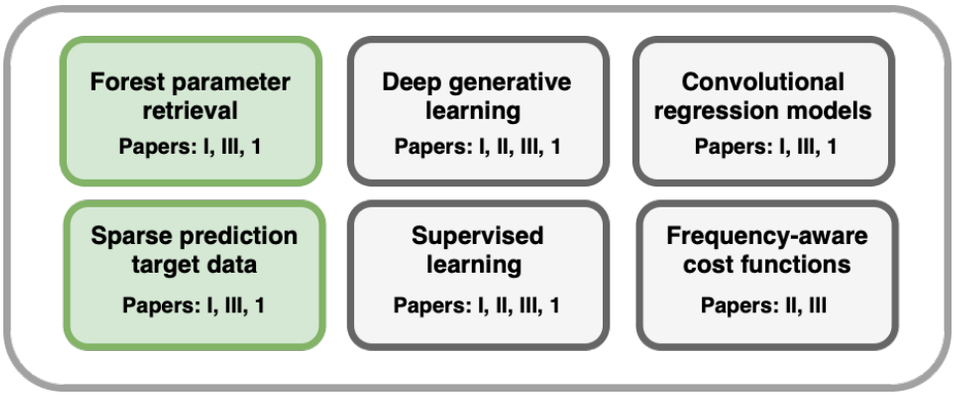


Figure 1.3: Overview of the topics that the various papers address, the Roman numbers refers to each of the papers included in this thesis, i.e. Papers I, II and III. The Arabic numeral refers to the paper listed under Section 1.5. The colours on the boxes refer to themes within forestry or within applications explored in this thesis.

1.4 Brief summary of included papers

This section presents a list of the papers included in this thesis, along with a brief summary of each paper. An extended summary of the listed papers included in the thesis can be found in Sections 7.1, 7.2 and 7.3. Section 1.5 lists additional academic work published during this Ph.D. project.

Figure 1.3 provides an overview of the topics covered in this thesis, where Papers I, II and III are referenced through their Roman number. The reference to Paper 1 in Figure 1.3 refers to the work listed under Section 1.5. The green coloured boxes in Figure 1.3 signify themes associated with forestry, while grey-coloured boxes represent various thematic applications explored in this thesis.

I. Sara Björk, Stian Normann Anfinsen, Erik Næsset, Terje Gobakken, and Eliakimu Zahabu. "**On the Potential of Sequential and Nonsequential Regression Models for Sentinel-1-Based Biomass Prediction in Tanzanian Miombo Forests**", in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4612-4639, 2022.

II. Sara Björk, Jonas N. Myhre, and Thomas Haugland Johansen. "**Simpler is Better: Spectral Regularization and Up-sampling Techniques for Variational Autoencoders**", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3778-3782, 2022.

III. Sara Björk, Stian N. Anfinsen, Michael Kampffmeyer, Erik Næsset, Terje Gobakken, and Lennart Noordermeer. "**Forest Parameter Prediction by Mul-**

tiobjective Deep Learning of Regression Models Trained With Pseudo-Target Imputation", submitted to *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

Paper I: This paper uses a sequential regression modelling approach to explore the potential of utilising CNN-based regression models for forest AGB prediction. In the sequence, the first regression model has linked *in situ* AGB data to ALS data and produced the ALS-derived AGB prediction map. Paper I focuses on developing methods for the second regression model in the sequence. It proposes training cGANs in a supervised setting to translate false colour image patches of Sentinel-1 backscatter into synthetic ALS-derived AGB prediction patches that closely resemble true ALS-derived AGB prediction maps. This approach enables the regression model to exploit the spatial context of the regressor and regressand data during the learning process. The proposed cGAN-based regression models are evaluated against parametric sequential and nonsequential Sentinel-1-based regression models, also proposed in Paper I. Additionally, all the models proposed in the paper are compared with other nonsequential sensor-based regression models previously developed for the AOI [28]. The empirical results demonstrate the potential of utilising C-band Sentinel-1 data for forest monitoring. Furthermore, the contextual cGAN-based regression models seem to capture the dynamic range and local variability of AGB.

Paper II: Proposes a novel frequency-aware objective function that can be incorporated with standard objective functions, such as pixel-aware or adversarial losses, in the learning of deep generative models. The purpose of the frequency-aware objective function is to enforce the model to also focus on achieving agreement of the overall spectral content of the data. The impact of the proposed objective function was evaluated by training generative variational autoencoder (VAE) [29] networks on benchmark datasets of natural images. Empirical results demonstrate that generative VAE models trained with the proposed objective function achieve results equal to, or better than, the current state-of-the-art in frequency-aware losses for generative models.

Paper III: This paper extends the research proposed in Paper I and Paper II. Specifically, Paper III proposes a novel methodology that leverages the available ALS-derived prediction maps and the limited amount of available ground reference measurements of AGB or SV to improve the performance of CNN-based regression models for forest parameter prediction. This is achieved by treating the ALS-derived maps as pseudo-targets and the ground reference measurements as true predictive targets. The methodology employs a semi-supervised imputation strategy where the sparse dataset of true targets is imputed with pseudo-targets, which provides a partially continuous target dataset. Note that the models are trained in a supervised fashion, as in Paper I. By utilising forest

masks, CNN-based regression models for forest parameter retrieval are enabled, as the CNN models are trained only in areas where ALS-derived predictions are available. Paper III further proposes to incorporate different learning objectives in the optimisation process, and especially the new frequency-aware objective function proposed in Paper II to improve learning from RS data that are characterised by high-frequency information content. Empirical results demonstrate that models developed with the proposed pseudo-target imputation strategy achieve state-of-the-art performance that surpasses traditional ALS-based regression models. Results are consistent for experiments on AGB prediction in Tanzania and SV prediction in Norway, which shows the robustness of the proposed methodology.

1.5 Additional work

1. Sara Björk, Stian Normann Anfinsen, Erik Næsset, Terje Gobakken, and Eliakimu Zahabu. "**Generation of Lidar-Predicted Forest Biomass Maps from Radar Backscatter with Conditional Generative Adversarial Networks**", in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020.

1.6 Reading guide

The remainder of this thesis is organised into three parts, *I) Methodology and context, II) Summary of research and concluding remarks,* and *III) Included papers*.

The first part, *Methodology and context* comprises five chapters organised as follows: Chapter 2 provides an overview of the remote sensing background, emphasising the SAR sensor and SAR data. ALS data for forest applications are briefly described here, as this thesis relies on using ALS-derived forest parameter prediction maps. Chapter 3 offers a brief introduction to traditional methods for forest parameter prediction, including methods for retrieval of ground reference data, focusing on AGB prediction. This chapter also introduces the study areas and the datasets relevant to Papers I and III, including the ALS-derived forest parameter prediction maps. Chapter 4 provides the basic machine learning concepts, while Chapter 5 covers the fundamental aspects of deep learning, with an emphasis of CNNs. Chapter 6 combines the basic concepts of Chapter 4 and Chapter 5 and introduces concepts for utilising deep learning regression models for forestry, including the semi-supervised imputation strategy with pseudo-targets.

The second part, *Summary of research and concluding remarks*, consist of four chapters, where Chapter 7-9 provides a brief overview of the three included papers, their scientific contributions and the author's main contributions to the works. Additionally, Chapter 8 includes some concluding remarks and discusses the limitations and potential future work in the field of deep convolutional regression modelling for forestry.

Lastly, the section *Included papers* includes the three research papers that form the basis of this work.

Part I

Methodology and context

/2

Remote sensing background

Within earth observation, the term remote sensing was introduced in 1960-1970 and refers to the acquisition of information about objects in the atmosphere, on the Earth's land or water surfaces without being in physical contact with it [1, 2]. The work presented in this thesis focuses on using data acquired from active remote sensing systems as these can acquire data regardless of the weather and lightning conditions [15, 30]. Unlike passive systems, which rely on naturally occurring energy sources, active systems themselves transmit signals that interact with the Earth. The active sensors then detect and measure reflected backscatter from the ground to acquire information about the objects of interest [1, 6]. Among the most widely used active remote sensing systems are microwave (radar) instruments and light detection and ranging (LiDAR) instruments.

Much of the groundwork and development of radar instruments occurred during World War II, which led to the development of the first real aperture radar (RAR) and SAR during the 1950s. NASA's first SAR sensor, which provided public-domain data, was launched in 1978 [6]. While SAR sensors transmit microwaves, the LiDAR sensor, developed in 1960, instead transmits laser light, often using wavelengths in the visible or near-infrared parts of the electromagnetic spectrum, see Figure 2.1 [1, 6]. An overview of state-of-the-art remote sensing platforms and sensors, including SAR and Lidar sensors, can be found

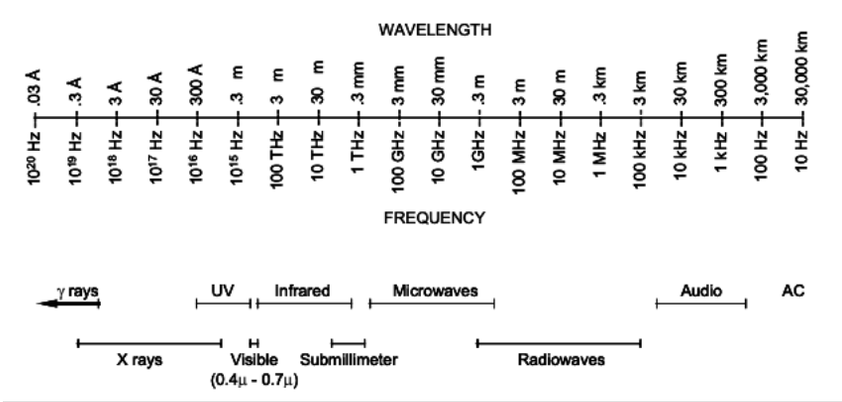


Figure 2.1: The electromagnetic spectrum sectioned by frequency or wavelength. Image retrieved from [1].

in the survey [31].

This thesis, and especially Paper I and III, focuses on using spaceborne SAR data from the Sentinel-1 sensor and remote sensing-derived products from ALS systems for forest parameter prediction. Thus, this chapter commences by establishing the definitions of temporal and spatial resolution in Section 2.1, followed by Section 2.2, which briefly introduces the most important concepts and characteristics of spaceborne SAR data, which can be used to train parametric and nonparametric regression models for forest parameter prediction. Interested readers are directed to [1, 6] for a more comprehensive technical background on SAR and SAR data. The chapter concludes with Section 2.3, where we offer conceptual insight into using ALS data for forest parameter prediction. We refer interested readers to [16] for a comprehensive overview of ALS for forestry applications.

2.1 Temporal and spatial resolution

Within remote sensing, the term *resolution* is characterised into many types, where we focus on the *temporal resolution* and *spatial resolution* and refer to [1, 2] for additional definitions.

Spaceborne SAR sensors map the Earth's surface in a predefined temporal scanning pattern. Based on the satellite's revisit period, this results in frequent revisits of the same point on Earth. Depending on the satellite, the revisit time may vary between days up to several weeks [15]. Thus, the **temporal resolution** reflects the revisiting time, where a fine temporal resolution implies a short

revisit time, while a course temporal resolution reflects a long revisit time [2]. The temporal resolution is not the most important aspect for forest AGB or SV prediction as, except for shedding leaves, changes in AGB or SV are not expected to occur rapidly during a season. However, the temporal resolution can be important to secure data coverage in dry seasons, when the radar cross-section has a larger dynamic range [1,15,30]. This makes it more sensitive to changes in forest parameters and better suited for prediction tasks, including regression. LiDAR systems, such as the ALS, are not operating following a predefined temporal scanning pattern compared to SAR satellites. Instead, they are explicitly employed during so-called flight campaigns for mapping the forest of a specific region. Thus, the term temporal resolution is commonly not employed for ALS systems. Alternatively, the term "multi-temporal ALS datasets" can describe ALS datasets collected at different times [16].

On the other hand, the **spatial resolution** relates to the minimum distance between two points on the surface that allows the two points to be separated in a remote sensing image. A sensor with high spatial resolution reflects a sensor that can discriminate between spatially close objects on the ground. The spatial resolution should not be confused with the pixel size, which for SAR is the measured pixel spacing in azimuthal or range direction after processing of the remote sensing data [15].

2.2 SAR characteristics

Spaceborne SAR sensors are active instruments that enable all-day and all-weather operational capabilities, implying that they can be utilised for regular mapping of regions that are affected by heavy cloud coverage, extended dark winter and persistent rain periods [15]. Consequently, SAR data are especially popular in systems that monitor, measure and predict forest parameters in regions that experience several yearly rain periods, such as Tanzania.

Compared to optical data acquired from standard digital cameras, SAR data is complex to process and requires domain knowledge about how microwave energy interacts with objects in the terrain. For example, SAR imagery has a grainy salt-and-pepper appearance known as *speckle*. The multiplicative speckle phenomenon is an inherent property of narrow-banded coherent imaging systems such as SAR, which result in a coherent addition of many scattering echoes from separate, but adjacent scatterers within a resolution cell [1,2,6,15]. Thus, speckle noise is unavoidable, but its effects in SAR imagery can be reduced through different filtering techniques, see [2,15].

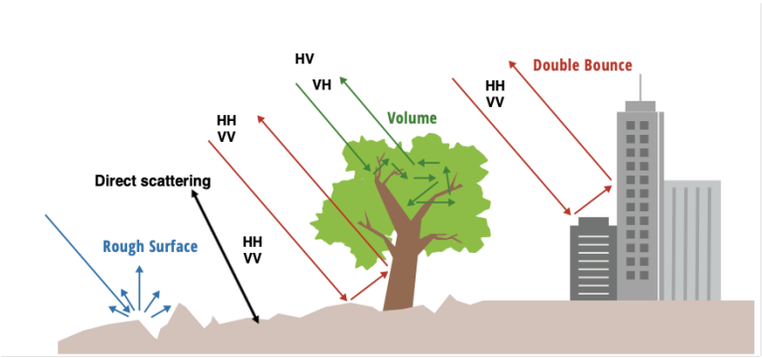


Figure 2.2: Different scattering types for SAR data. Image adapted from [15].

2.2.1 Scattering

The current operational SAR sensors operate at a single frequency band within the microwave region of the electromagnetic spectrum. Conceptually, SAR operates by transmitting long-wavelength microwave energy and measuring the amount of microwave energy backscattered from the terrain [6]. Physical relationships between the backscattered signal and objects present in the terrain can be developed based on knowledge of the surface and terrain being mapped, such as whether buildings or forests cover it. Figure 2.2 depicts the main scattering types for SAR data, which depicts how much of the incoming signal is reflected back to the sensor. For example, the amount of direct and isotropic scattering depends, among other things, on the smoothness or roughness of the surface. A surface's roughness is itself relative to and dependent on the wavelength (frequency) at which the radar operates.

The signal is directly reflected back to the sensor in *surface* (direct) scattering. This direct reflection can be part of diffuse or *isotropic scattering*, where large parts of the incoming signal are scattered away from the sensor [1, 2]. *Double-bounce* scattering implies that the orientation of different objects, such as the vertical structure of e.g. the tree stem (or a building) and the horizontal ground, deflects the incoming signal back to the SAR sensor. In contrast, *volume scattering* implies that the signal bounces multiple times within the vegetation before a proportion of the signal is reflected to the sensor. Depending on the forest type and the signal frequency, double-bounce, volume, and surface scattering are among the most common scattering types in forests [15, 32].

2.2.2 Polarisation

Scattering properties and the scattered signal's strength depend on many factors, such as the *polarisation* of the transmitted and received electromagnetic

Table 2.1: The most commonly employed microwave bands [6,15].

Frequency band	**X**	**C**	**L**	**P**
Frequency [GHz]	12.5-8.0	8.0-4.0	2.0-1.0	1.0-0.3
Wavelength λ [cm]	2.4-3.8	3.9-7.5	15.0-30.0	30.0-100

wave. Another factor is the dielectric properties of the medium, which in turn is affected by *moisture*[1] and material properties [1].

Most SAR sensors transmit either horizontally (H) or vertically (V) polarised microwave pulses and, depending on the radar system, receive either horizontally or vertically polarised energy scattered from the ground or both polarisations (dual-pol). Most single-polarisation (single-pol) SAR systems transmit and record like-polarised signals (e.g. HH or VV). However, there exists a few single-pol SAR systems that are able to measure the cross-polarised signal (e.g. HV or VH). Dual-pol SAR sensors transmit one polarisation and record the like-polarised and cross-polarised signals (e.g. HH and HV). Quadrature-polarimetric SAR sensors are the most refined system, which transmits and receive both polarisation [15].

2.2.3 Wavelength and penetration depth

Table 2.1 outlines the frequency range and corresponding wavelength of the most widely used frequency bands operated by SAR. There are currently no operating spaceborne P-band SAR sensors, which limits their use for large-scale national-level forest mapping. The first planned spaceborne P-band mission to provide global measurements of vegetation and forest biomass is the BIOMASS satellite, which is scheduled for a launch in 2024 [15]. It will be particularly important for biomass estimation in high and dense forests, such as rainforests, which require low frequency to penetrate the whole forest volume.

As shown in Figure 2.3, the *penetration depth* of the microwave signal into forests depends on the signal's wavelength. Furthermore, scattering occurs when the particles are on the same scale as the radar wavelength, causing the X-band radar to mainly interact with the upper layer of the canopy, through surface and volume scattering, and C-band with the crown volume. On the other hand, both L- and P-band radar can penetrate deeper into the forest [15, 32].

1. See Section 2.2.4.

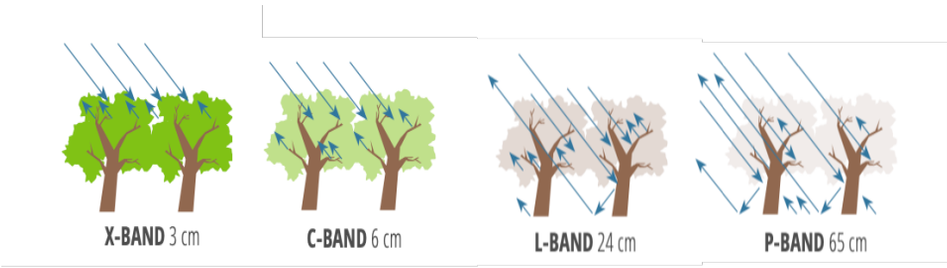


Figure 2.3: Differences in penetration depth, which is related to the wavelength. Illustration from [15].

2.2.4 Moisture and other factors

Other factors also impact the penetration depth and strength of backscattered signals besides those mentioned earlier. For instance, the angle at which the sensor views the terrain, the terrain's topography and environmental conditions, such as soil moisture and vegetation phenology or moisture or moisture in genera [1, 8, 15, 33]. While geometric distortions caused by topography effects can be mitigated through proper geocoding of the SAR data [1, 2], the moisture effects in tropical regions can be limited by choosing SAR imagery outside the periodic rain periods or data from a SAR sensor that operates with longer wavelengths [30].

2.2.5 Saturation

Radar backscatter correlates with biomass, indicating that biomass increases with the magnitude of the backscatter signal. However, the correlation tends to saturate at a level, implying that a further biomass increase generally cannot be inferred from the SAR data. The saturation level is dependent on the wavelength of the radar, with longer wavelengths (such as L-band and P-band SAR) typically having higher saturation levels [8, 21, 32, 34]. Therefore, these longer wavelength SAR data are often preferred for developing SAR-based AGB regression models.

2.3 LiDAR: a conceptual overview

Since the papers included in this thesis solely focus on utilising ALS-derived prediction maps of either AGB or SV, they do not directly address the ALS data itself. Therefore, this overview aims to provide a condensed conceptual understanding of ALS data and to highlight the difference between SAR and

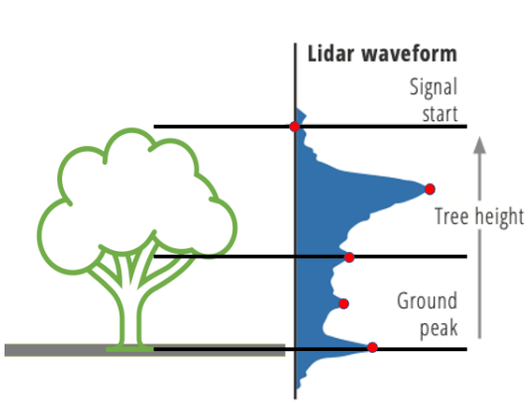


Figure 2.4: Principle of an ALS imaging system, showing both the whole reflected waveform and discrete returns as the red circles. Image adapted from [15].

ALS remote sensing of forests. It is important to note that activities related to performing ALS flight campaigns, processing ALS data, and deriving ALS-based forest parameter prediction models are outside the scope of this thesis.

Similarly to the SAR, LiDAR is also an active remote sensing sensor, however commonly operating in the visible or near-infrared part of the electromagnetic spectrum when utilised for vegetation purposes [20]. For ALS systems, the LiDAR instrument is mounted on airborne platforms. It operates by transmitting laser pulses to target objects on the ground and by measuring the round-trip time for a pulse from the sensor and the target. As each laser pulse interacts with different components of the canopy, vegetation, stem and ground surface, the reflected light pulse creates a waveform of returned energy, shown in Figure 2.4, with different peaks representing the different components. Some ALS sensors record and digitalise the whole waveform, known as *full-waveform* ALS, while others, known as *discrete return* ALS, only record the time and the reflected pulse energy of one to five echoes for each transmitted pulse [2, 16, 20]. These discrete echoes are shown as the red circles in Figure 2.4.

By knowing the returning pulse's position, direction and angle, the distance to the target can be measured [20, 35]. Therefore, echoes from repetitive ALS pulses provide an elevation profile underneath the platform. By having access to a digital elevation model (DEM) representing the terrain surface ground, tree heights can be inferred directly from the ALS data by subtracting the ground elevation from the dataset [2, 20]. ALS data, therefore, provides an accurate and direct measure of the tree height which, by use of *allometric equations*[2], can be related to AGB.

2. See Section 3.2.

3

Traditional methods for forest parameter prediction

Forests play a crucial role in mitigating climate change through their ability to absorb and store carbon dioxide in the vegetation biomass, which is a larger global storage of carbon than the atmosphere [7, 8, 36]. Furthermore, sustainable forest management and estimation of available raw materials or the potential for bioenergy are additional crucial factors that require accurate mapping and monitoring of forests and forest biomass [9, 10, 15, 37–39]. The primary difficulty in monitoring AGB is to obtain field measurements in various parts of the world due to various factors, such as geographical remoteness, lack of capacity, data scarcity and armed conflicts. Thus, the most cost-effective technology to overcome this challenge is the combination of remote sensing and ground measurements for AGB monitoring to obtain current information on forest coverage and carbon stocks at various scales [7,14,15,21,36,39,40].

This chapter provides an overview of forest monitoring, conventional methods of forest parameter prediction and the motivation behind the deep learning methods proposed in Paper I and Paper III, which focus on using deep convolutional regression models for forest parameter prediction. We start this chapter by briefly introducing general concepts related to field data and field inventory campaigns in Section 3.1. Section 3.2 generally introduces allometric equations,

while Section 3.3 describes standard conventional remote sensing methods for forest parameter prediction and more advanced methods based on machine learning and deep learning. Lastly, Section 3.4 presents the study areas and data studied in Paper I and Paper III, which includes the ground reference datasets, the ALS-derived forest prediction maps, the Sentinel-1 data and challenges related to the use of these datasets to train deep convolutional regression models.

For simplicity, Sections 3.1, 3.2 and 3.3 concentrate on methodologies for obtaining AGB ground reference data and performing AGB prediction. Nevertheless, the methodologies for obtaining ground reference data or predicting other forest parameters, such as SV, are similar to those presented.

3.1 Ground reference data

Living forest biomass is classified into two classes; *AGB*, which includes stems, stumps, branches, bark, seeds, and foliage, and *below-ground biomass (BGB)*, which includes all living roots with a diameter larger than 2 mm in diameter [8, 9, 21, 39]. In this thesis, only AGB is considered, and the terms biomass and AGB will be used interchangeably to refer to living forest above-ground biomass. As the forest stem volume, SV, accounts for the highest proportion of the biomass in a tree, approximately 65-80 % [11–13], AGB monitoring commonly focuses on either estimating the total amount of AGB or SV [12, 23, 28, 38]. While Paper I solely focuses on developing methodologies for AGB prediction, Paper III addresses both AGB and SV prediction.

Ground reference field data is essential to support any remote sensing application that aims to monitor and map forest parameters locally, regionally or nationally. *In situ* AGB data are acquired through NFI campaigns or representative case studies of forests using either *destructive* or *nondestructive* sampling [8]. The destructive method estimates available biomass from the weight of dried plants and trees. In contrast, the nondestructive method implies that tree parameters such as tree height and stem diameter are measured for predefined small sites, so-called *sample plots*. These measurements are later related to AGB or SV using *allometric equations* that are determined through statistical regression methods [8, 21, 39]. These equations have typically been developed for a specific forest stand, or AOI [8]. Although destructive sampling is the most accurate for small regions, it is typically avoided due to its high cost and the time-consuming process, resulting in the nondestructive sampling technique being the most commonly used method for obtaining ground reference data.

3.1.1 Field inventory campaigns

Although fieldwork was not conducted as part of this thesis, this section points out some important aspects of field work practices to provide a basic understanding of the methods described in Section 3.3 and the data presented in Section 3.4. The interested reader can consult [15] for a summary of guidelines for field plot or sampling design or references to suitable documentation on choices related to field inventory sampling.

Field inventories are conducted at the local, regional or national level to obtain representative data for a specific application. Various decisions must be made before carrying out a field inventory campaign, such as deciding on the number of plots and plot shape. The number of plots should be large enough to decrease the estimated parameter's uncertainty below some targeted level and to account for local variations in an AOI [2,15]. At the same time, they should not be too numerous as this implies a labour-intensive, time-demanding and costly field inventory campaign [15]. As a result, the cost assessment and error estimate of the intended field inventory are commonly employed as constraints to determine the most cost-effective design [19].

The sample plot shape is usually circular, rectangular or squared. Small circular plots are popular as they are less time-demanding to implement since there is no need to mark corners. In contrast, large circular plots are known to be more difficult to define on the ground, while rectangular plots are a better choice if the sample plots should be associated with remote sensing data. For LiDAR-biomass models are, for example, square plots recommended for most forest types [15]. With some exceptions, [19,28], rectangular plots have been used in tropical forests, while circular plots have commonly been used in boreal and temperate forests [19].

The field plot's size and orientation depend on the application and the topography, but should be large enough to ensure that the forest is correctly represented within them. In SAR-assisted studies, depending on the pixel size, a plot size > 0.25ha (2500m^2) or $>$ 1ha (10000m^2) is recommended [15].

Campbell and Wynne [2] categorise the minimum of information that has to be obtained from each field plot into the following three categories, *attributes*, *location* and *time*. The attribute category includes descriptions of the ground condition at the place, such as identification of tree species, soil moisture content, and size of the trees. For all tree diameters above a threshold, metrics, such as the tree height and crown size, are sampled. The location category includes, for example, elevation information and location reference from a GPS to correctly link the attributes to image data. Lastly, the measurements must also be described in terms of date and time [2,15].

3.2 Allometric equations

Allometric equations are used to relate forest inventory data into estimates of AGB where, in decreasing order, stem diameter, wood specific gravity, total height, and forest type are the most important predictors of AGB for a tree [41]. Based on field measurements of the stem diameter D (cm), total tree height H (m) and wood specific gravity ρ (gcm^{-3}), Chave *et al.* [42] used the following log-log model

$$ln(AGB) = \alpha + \beta ln(\rho \times D^2 \times H) + \epsilon, \tag{3.1}$$

to relate field measurements to AGB, where α and β are model coefficients and ϵ is an error term. Sometimes, allometric equations that only depend on measured trunk diameter are preferred. See for example [41, 42] for some examples. However, allometric equations, independent of the tree height are deemed less accurate, implying that these equations typically are developed for a specific area and forest type [8,42]. Similar to Eq. (3.1), ALS-based allometric equations exist to convert ALS measurements of forest tree height into AGB, these may also be site-dependent [15].

3.3 Remote sensing-assisted methods for forest parameter prediction

Although allometric equations are the most accurate method for inferring AGB from a small set of field measurements, this approach is not practical for large-scale mapping or monitoring of biomass. Instead, LiDAR and radar remote sensing techniques are recognised as better alternatives [15]. Typically, remote sensing data are used to fit parametric or nonparametric models to the small set of ground reference data. Following model fitting and based on the temporal resolution of the sensor, the model can be used to create frequently updated AGB prediction maps over the AOI during inference.

Since tree height can be inferred directly from ALS measurements, ALS-based AGB models are significantly more accurate than corresponding models developed using SAR data [15]. However, ALS data is associated with a high acquisition cost, especially for mapping large areas, implying that regular acquisition of ALS data is limited [21,27]. In contrast, spaceborne SAR can provide frequently updated data with extensive spatial coverage. Nevertheless, there are many limitations to using SAR data for AGB estimation, such as the known saturation of the backscatter signal at frequency-dependent levels of AGB, and

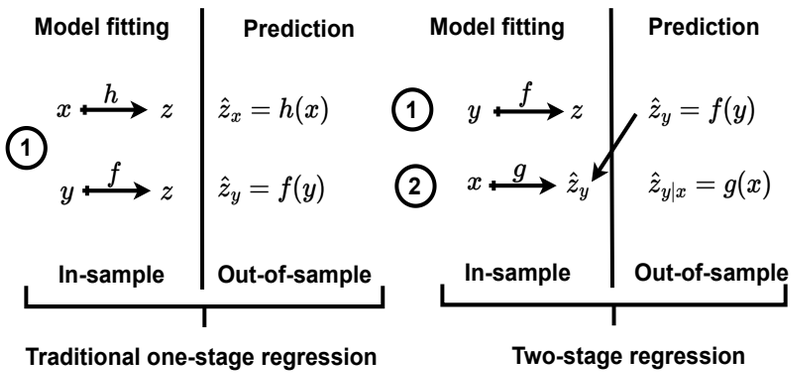


Figure 3.1: Illustration of the difference between a traditional nonsequential and a sequential regression model. Here, x denote data from a SAR sensor, y denote ALS data and z denote AGB ground reference data. Regression models are represented by f, g and h, where f is a regression model between y data and z data. The regression model between x data and z data is denoted h, while g is a regression model between x data and ALS-derived AGB predictions denoted $\widehat{z}_y$. Additionally, $\widehat{z}_x$ denote SAR-derived AGB predictions from a traditional non-sequential regression model. In the sequential setting on the right hand side, $\widehat{z}_{y|x}$ denote the outcome from the second part of the two subsequent regression models, i.e. a generated synthetic ALS-derived AGB predictions retrieved from x data (SAR). Image retrieved from [17].

the variation of the SAR backscatter signal with the moisture of the soil and vegetation [1,15,32]. The latter effects can be limited by using a temporal stack of SAR images over the AOI. This results in a regression model that can be fitted on AGB data from field inventories, using features extracted from the multitemporal SAR data as regressors. Numerous studies have investigated the potential of using ALS or SAR data, or a combination of both, for developing models to map and monitor AGB at various scales accurately [7,16,21,23,28,32,34].

3.3.1 Sequential and nonsequential modelling

In this thesis, we refer to a modelling approach that establishes a direct relationship between a limited set of ground reference data and remote sensing data from sensors like SAR or ALS as *nonsequential*. We also refer to this as a one-stage regression model in this context. This approach is depicted in the left side of Figure 3.1, where x denotes data from a SAR sensor, y denotes ALS data, and z denotes *in situ* AGB ground reference data. In the model fitting stage, the two regression functions h and f aim to approximate the respective relationships between the dependent variable z and the predictor x, and between z and the predictor y. After model fitting, h and f can, in the prediction

phase, be used to create SAR-derived AGB predictions, $\widehat{z}_x$, or ALS-derived AGB predictions, $\widehat{z}_y$.

The ability to infer tree heights directly from ALS data implies that ALS data is highly suitable for accurate up-scaling of measurements from forest inventories to regional and global scales. Consequently, ALS data can be effectively utilised to calibrate and enhance the precision of SAR-based AGB models [7, 15, 21]. In this thesis, we refer to this approach as *sequential* modelling, which implies that two regression models are used in a chain to obtain more training data for AGB prediction [17]. We also refer to this as a two-stage regression model. The two stages are depicted in the right-hand side of Figure 3.1. In the first stage, a regression model f is trained to establish the relationship between *in situ* AGB data, denoted z, and data derived from a single RS data source, such as ALS y, which exhibits a strong correlation with z but possesses limited geographical coverage. Following the model fitting process, f can be employed to generate an accurate ALS-derived AGB prediction map, with each prediction denoted as $\widehat{z}_y$. The prediction map is then utilised in the second regression model, g, as a surrogate for ground reference data to regress on data from an additional RS sensor with a larger spatial extent, denoted with x. The adoption of the sequential modelling approach in Paper I facilitates the utilisation of deep learning-based regression models. This is because the prediction map $\widehat{\boldsymbol{Z}}_y$ is wall-to-wall[1], meaning that it is spatially continuous and can be divided into image patches to be used as training data for a CNN.

3.3.2 Approaches to forest parameter prediction

To this date, traditional parametric statistical regression models are popular choices for AGB estimation and prediction[2] using a nonsequential or sequential modelling approach [17, 43]. Among these, variations of traditional linear regression are most common, i.e. simple linear regression, multiple linear regression and step-wise multiple regression, see e.g., [30, 44–52]. While the traditional statistical regression models have their merits, they often face challenges in handling complex and high-dimensional data, and modelling nonlinear relationships between the regressand and the regressor. As a result, non-parametric regression models such as machine learning-based[3] models have

1. Note that $\widehat{\boldsymbol{Z}}_y$ denotes a $N \times M$ prediction map, while $\widehat{z}_y$ denotes single predictions, i.e. $\widehat{z}_y \in \widehat{\boldsymbol{Z}}_y$. See Section 4.1 for the notation used in this thesis.
2. We are aware that the terms estimation and prediction are sometimes used interchangeably and that their usage is subject to discussion. In this thesis, we consistently use the term *prediction* instead of estimation as we consider our approach as model-based inference. According to our knowledge, this is in line with the usage in statistical inference in forestry.
3. See Chapter 4 for an introduction to machine learning and the major differences between traditional statistical regression models and machine learning-based models.

introduced many alternatives to conventional regression models. For example, nonparametric models are more suitable in large-scale geospatial regression [15]. Within machine learning-based models utilised for nonsequential and sequential modelling, random forests, support vector machines and fully connected NNs, such as the multilayer perception (MLP), are the most popular choices see e.g., [15,17,46–48,53–67].

Both conventional statistical regression models and the machine learning methods, mentioned above for AGB prediction, have traditionally operated on an individual pixel level. Thus, these models are limited to establishing relationships between single observations of ground reference measurements and corresponding pixels from the RS data source, without considering the spatial context of neighbouring pixels within the RS dataset. Additionally, the extraction of meaningful features, known as feature engineering, from the RS data requires domain knowledge [68]. This process involves identifying and utilising domain-specific attributes from the data to train powerful regression models that can effectively relate RS data to ground reference measurements [3].

Deep learning-based approaches: Deep learning approaches, and particularly CNNs[4] [18, 69], offer the capability to incorporate information from a spatial neighbourhood surrounding individual pixels through the utilisation of convolutional filters. As a result, the prediction of each pixel is influenced by regressors derived from the spatial neighbourhood around it [3,17]. In contrast to conventional statistical models and machine learning methods, the CNN itself is able to learn from the data and extract relevant features, while also learning the relationship between the regressors and the regressand, which may be the main asset of deep learning-based regression models [3, 4].

While many studies have shown the potential of deep learning and especially CNNs in applications related to the utilisation of RS data for vegetation applications, only a minority of the studies reviewed in [3, 4] focus on forest data and regression tasks. In fact, most existing studies focus on classification and applications related to agriculture [3], which probably is due to the labour-intensive, time-consuming, and costly nature of conducting forest field inventory campaigns [4,15], which are required to obtain the continuous target data required for training of such contextual regression models. As a result, to fully utilise deep learning and the potential of CNNs for regression modelling of forest parameters, the sparse set of target data has to be utilised cleverly and combined with additional data sources. In Papers I and III, we propose utilising ALS-derived prediction maps to address this challenge.

4. See Chapter 5 for an introduction to deep learning and CNNs.

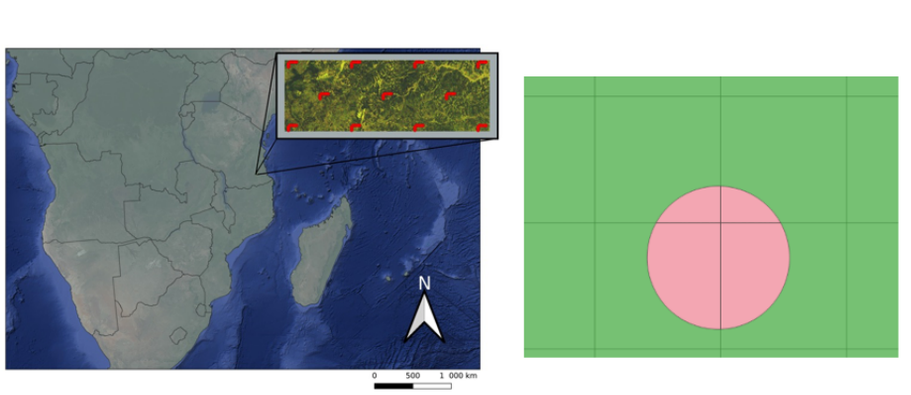


Figure 3.2: Left: The location of the Tanzanian datasets, a section of the Sentinel-1A scene covering the AOI and ground reference data shown as red L-shaped forms in the Sentinel-1A scene. Image retrieved from [17]. **Right:** A part of the ALS-derived AGB prediction map in green and one circular field plot shown in pink.

3.4 Study areas and datasets

In this thesis, two distinct collections of datasets are employed to predict forest parameters. The first collection consists of AGB ground reference data gathered from a dry tropical forest in Tanzania and corresponding ALS-derived AGB prediction maps. This dataset is utilised in both Paper I and Paper III. The second collection comprises SV ground reference data obtained from three managed boreal forests in Norway, accompanied by ALS-derived SV prediction maps from the same three regions. The collection of SV data is explicitly used in Paper III. A brief description of each dataset is provided in the subsequent sections. We refer to [28] for the original source of the datasets from Tanzania, while [23] can be consulted for the original source of the datasets from the Norwegian regions. In addition, RS data from the C-band SAR Sentinel-1 sensor are utilised for both Paper I and Paper III as Sentinel-1 data are freely available. Furthermore, the Sentinel-1 sensor can acquire data both at night and in cloudy conditions, it offers short revisit time and good coverage for the areas of interest.

3.4.1 The Tanzanian datasets

The field work was performed in January-February 2014 within a rectangular region of size 11.25×32.50 km (WGS 84/UTM zone 36S) located in the Liwale district in the southeast of Tanzania (9°52'-9°58'S, 38°19'-38°36'E). Ground reference data were collected from 88 circular field plots, each with an area of size 707 m^2. The field plots were distributed as L-shaped clusters of 11 plots

each within the AOI in Tanzania, as shown on the left side of Figure 3.2. We refer to [22,28,70] for a description of how data from the field work were used to develop large-scale AGB models.

The ALS data used to derive the wall-to-wall map of ALS-derived AGB predictions was acquired in March 2014. See [28] for details on the ALS flight campaign, ALS data processing, and the match-up of ALS data with ground reference AGB data from the field plots. The wall-to-wall map is represented as raster data with square pixels of size 707 m^2. A small part of the ALS wall-to-wall map is shown as green square pixels in the right part of Figure 3.2 together with the outline of one field plot in pink. By sampling the SAR data to the same pixel spacing, and applying the same map projection as the ALS-derived AGB prediction map, the SAR data and AGB prediction map can be used to train deep convolutional regression models.

3.4.2 The Norwegian datasets

The Norwegian datasets encompass data from three regions in the southeast of Norway, shown in the left part of Figure 3.3, and referred to as Nordre Land (A), Tyristrand (B) and Hole (C). The field inventory campaign took place during the summer and fall of 2017, where field measurements were obtained from circular field plots with an area of 250 m^2. For further details on the sampling design, related data properties and how SV was predicted from the field measurements, see [23]. Out of all field plots, SV ground reference data from 264 plots were included in Paper III. Among these, 136 plots were located within the Nordre Land region, while 77 and 51 plots were distributed within the Tyristrand and Hole regions, respectively.

The ALS flight campaigns for all three regions took place in 2016. For further details on how the ALS data were processed, the formulation of the prediction models and the match-up of ALS-derived predictions with SV ground reference data to create SV prediction maps over the three regions, we refer to Noordermeer *et al.* [23]. Compared to the Tanzanian ALS-derived prediction maps, the SV prediction maps were limited to forest areas where the forest height exceeded 8-9 meters, resulting in ALS-derived prediction maps consisting of spatially disjoint polygons. A small part of the spatially disjoint prediction maps in Nordre Land can be seen as the purple lattice in the right part of Figure 3.3, together with the outline of two different field plots in pink. The disjoint polygons are seen in Figure 3.4, where the brown areas indicate where SV predictions are available, whereas the background (other colours) is retrieved from OpenStreetMap [71].

To utilise the spatially disjoint datasets of ALS-derived SV prediction maps

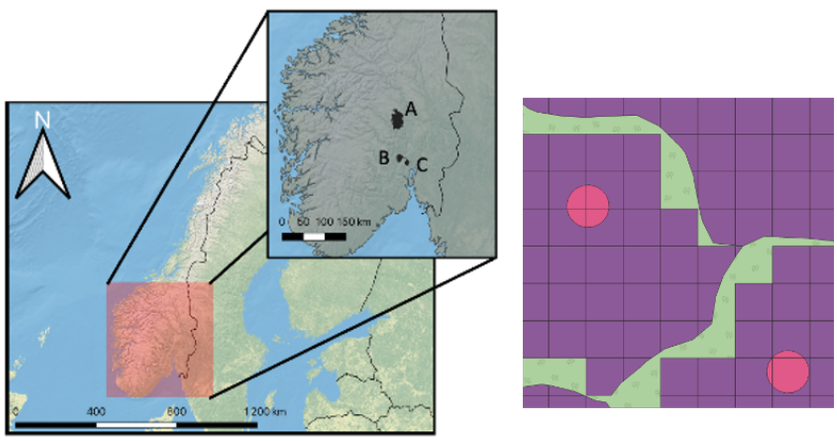


Figure 3.3: Left: The location of the three regions in Norway: Nordre Land (A), Tyristrand (B) and Hole (B). **Right:** A part of the ALS-derived SV prediction map for Nordre Land in purple with the outline of two field plots shown in pink.

as target data for training deep convolutional regression models, we refer to Paper III, see Section 11 for a comprehensive description of the required processing steps. In summary, the processing steps aim to rasterise the spatially disjoint datasets to align them with the pixel grid of the SAR predictor data. Consequently, the resulting SV prediction maps will contain areas with one SV prediction for each pixel and regions with no available data. Further details on how the deep convolutional regression models were adapted to handle this specific type of data can be found in Section 6.3.2.

3.4.3 Comments of the plot shape and size

Compared to [15], the size of the field plots in the Tanzanian and Norwegian regions are smaller than generally recommended. However, the plot size is unlikely to pose a challenge for the Norwegian regions, as the field data were obtained from trees of commercially managed monodominant boreal forests with a minimum forest height of 8-9 meters [23]. This indicates that the forest structure and type within a field plot are similar to the surrounding area. On the other hand, the small plot size may present challenges for the Tanzanian dataset due to the species variability and sparse distribution of large trees in this kind of Tanzanian forest, known as dry miombo woodland [19]. As a result, large trees can be located near the plot periphery of a sample plot, implying that most of the stem is outside the field plot, while the crown is within the field plot

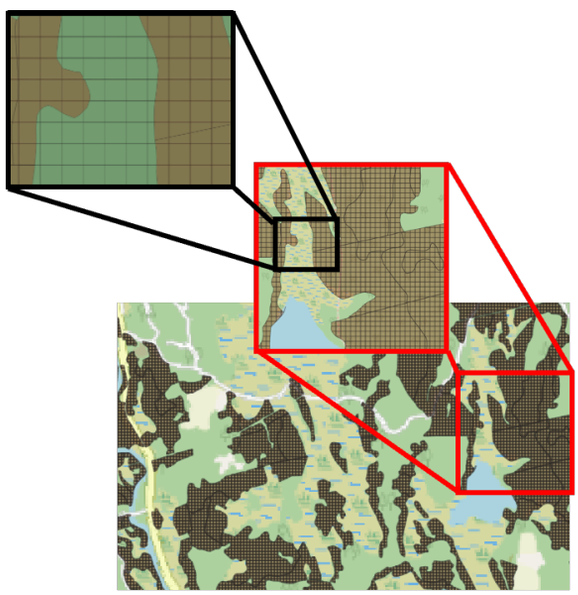


Figure 3.4: A small section of the ALS-derived SV prediction dataset from the northern parts of Nordre Land, where SV predictions are made in the brown areas.

(or vice versa). This can potentially impact the recorded ALS echoes [28] and how a regression model associates ground reference data to RS data. As deep convolutional regression models exploit the spatial relationship in the RS data, employing this type of regression model may be more beneficial compared to traditional statistical regression models and machine learning models that only infer relationships on the pixel level.

Moreover, the circular shape of the field plots in the Tanzanian and Norwegian regions is suboptimal for studies like Paper I and Paper III, where the ground reference data are associated with RS data, represented with rasters of square pixels. The challenge is depicted in the right part of Figure 3.2 and similarly in Figure 3.3, showing that ground reference plots (pink) intersect with three to four different pixels from the raster data of ALS-derived predictions. Since the Sentinel-1 data were processed to align with the wall-to-wall map and share the same map projection, each ground reference plot will also overlap with multiple Sentinel-1 pixels. Consequently, to obtain AGB or SV predictions for each field plot using the proposed regression models, it is necessary to compute the area-weighted mean AGB and SV by considering the neighbouring pixels that intersect with the field plot. Compared to using a rectangular field plot shape, the circular plot shape may introduce inaccuracies in the reported results of Papers I and III. However, it should be noted that decisions about the plot shape were beyond the scope of the thesis.

4

Machine learning basics

Although there are similarities between conventional statistical methods and machine learning techniques, they differ in various ways. Statistical learning has roots in mathematics and statistics and has been used for centuries, if not longer. In contrast, machine learning, which emerged from computer science [72], is a relatively recent field that builds on statistical principles [72,73]. Both statistical and machine learning methods aim to learn patterns or relationships from data by estimating functions. However, statistical models tend to be simpler than machine learning models, and statistical learning emphasises providing confidence intervals to these functions. In contrast, machine learning models can contain from a few to millions of parameters optimised in the iterative learning process, and their algorithms seldom focus on providing confidence intervals. Furthermore, machine learning involves models and algorithms capable of intelligently learning relationships from data without being explicitly told about them [18]. As a result, machine learning algorithms are today used almost everywhere for e.g. data analysis, feature extraction, data transformation, classification and regression purposes [74].

This chapter briefly overviews the key concepts of machine learning, Section 4.1 introduces the notation used throughout this, followed by the task of machine learning in Section 4.2 and prevalent training paradigms in Section 4.3. Lastly, Section 4.4 briefly describes how machine learning algorithms can be optimised using training, test and validation datasets.

4.1 Terminology and Notation

In statistical learning, when we predict $\boldsymbol{y}$ from $\boldsymbol{x}$, $\boldsymbol{x}$ is typically called *independent variables* or *predictors*, and $\boldsymbol{y}$ is referred to as *response* or *dependent variables*. In contrast, in machine learning, $\boldsymbol{x}$ is referred to as *features* or *input data*, while $\boldsymbol{y}$ commonly are known as *output data*, *labels*, or the *prediction target*. The term "label" is commonly used in classification tasks, while the other two terms are used interchangeably for classification and regression purposes. In this thesis, we adopt the notation and terminology from machine learning and refer to [18] and especially to [73, 75] for differences in terminology between these two fields.

This thesis uses x, x_i, y or y_i to refer to single one-dimensional data points, where the subscript i indicates a specific one-dimensional data point. Bold lower case letters denote D-dimensional vectors, such as $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, ...x_{i,d}..., x_{i,D}]^T$. $\boldsymbol{X}$ and $\boldsymbol{Y}$ generally represents $N \times M \times D$ images, i.e. tensors consisting of observations or pixels $x_{i,j,d}$ and $x_{i,j,d}$, with $i = 1, ..., N$, $j = 1, ..., M$, $d = 1, ..., D$. For square images, $N = M$ denotes the spatial dimensions. As with lower case letters, $\boldsymbol{X}$ denotes data from the input domain. In contrast $\boldsymbol{Y}$ denotes data from the output domain where each $y_{i,j,d}$, in the regression setting, represents a prediction target associated with $x_{i,j,d}$.

Each dimension d of $\boldsymbol{x}_i$ or $\boldsymbol{x}_{i,j}$ represents a so-called *feature* that characterises the input data. These features can correspond to various remote sensing measurements. For instance, if the input data comprises of dual-polarisation Sentinel-1 sensor data, then $\boldsymbol{x}_{i,j}$ would be two-dimensional and include backscatter measurements from the VH and HH polarisations. Different data points, $\boldsymbol{x}_{i,j}$, then represent different geographical positions in the acquired Sentinel-1 scene. To simplify the notation, $x_{i,j,d}$ is sometimes referred to as $\boldsymbol{x}_{i,j}$ or just $\boldsymbol{x}$ when discussing general multidimensional data points. The same simplified notation is also applied for $y_{i,j,d}$.

4.2 Machine learning tasks

Machine learning algorithms aim to learn the relationship between the input data $\boldsymbol{x} \in \mathcal{X}$ and corresponding output data $\boldsymbol{y} \in \mathcal{Y}$ in a given training dataset $\mathcal{D}_{tr} = (\boldsymbol{x}, \boldsymbol{y})$. Here, $\mathcal{X}$ and $\mathcal{Y}$ represent the input and target domain, respectively. In a simple classification setting, $\boldsymbol{x}$ could represent image data of animals, while $y_i \in \boldsymbol{y}$ represents animal classes such as "cat" or "cow". The algorithm aims to find a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that approximates the relationship between the two domains properly and that can generalise well to new, unseen samples from a test dataset. This learning process involves an iterative process aiming to find

the optimal model parameters $\boldsymbol{\theta}^*$ that minimise a loss function $\mathcal{L}$, such that the prediction of $f(\boldsymbol{x}; \boldsymbol{\theta})$ is as close as possible to the true prediction target y. Generally, the loss function quantifies how close a prediction $f(\boldsymbol{x})$ is to y [18]. The mapping function, f, is in the literature interchangeably referred to as the *learning function*, as a general *machine learning algorithm* or a *machine learning model* [18]. We will use these terms interchangeably in this thesis.

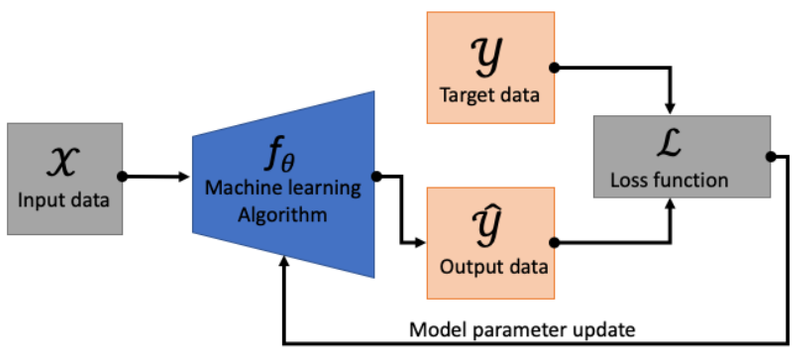


Figure 4.1: The iterative machine learning process in the supervised setting. By comparing the machine learning algorithm's beliefs for the relationship between input and output data to available target data, the model is updated through its model parameters, $\boldsymbol{\theta}$. The process repeats iteratively until convergence, which ultimately implies $\hat{\boldsymbol{y}} \approx \boldsymbol{y}$. No target data is available in the learning process in the unsupervised setting. Moreover, an alternative loss function must be formulated.

A simplified illustration of the machine learning process can be found in Figure 4.1. In short, the loss function computes the difference between predicted targets, i.e. $\hat{\boldsymbol{y}}$, and true prediction targets $\boldsymbol{y}$, and gives feedback to f. The model parameters are updated according to the size of the computed loss and their contribution. This learning process alternates between providing $\hat{\boldsymbol{y}}$ and updating model parameters continuously until some convergence criteria are met, such as a stable loss below a threshold value. Depending on whether the response variable is discrete or continuous, the prediction is either categorical or continuous, the latter representing regression. For the purposes of this thesis, we focus solely on regression and do not cover methods for categorical prediction, which, for example, relates to classification problems and segmentation tasks [72, 73, 75]. Readers interested in categorical prediction can refer to statistical textbooks, such as [75, 76], for an overview.

4.3 Machine learning paradigms

The iterative process described so far and illustrated in Figure 4.1 is defined as *supervised* learning problem and implies that both input and target data are available during training, i.e. $\mathcal{D}_{tr} = \{(\boldsymbol{x}_i, y_i)|i = 1, ..., N_{tr}\}$ [72,74]. While supervised learning is the optimal approach when target data is available, this is not always the case. For instance, when developing regression models for the forestry domain, large amounts of data are often available from the input domain, but only a limited amount of labelled ground reference target data. This is often due to the extensive resources required to collect ground reference data from field plots, which can be limited by seasonal, geographical, economical and labour constraints.

Unsupervised learning is another machine learning paradigm in which the algorithm is trained on unlabelled input data only, i.e. $\mathcal{D}_u = \{\boldsymbol{x}_i|i = 1, ..., N_\}$. The unsupervised learning process is still iterative, as shown in the supervised process shown in Figure 4.1, however, it uses different algorithms that do not require labelled target data. Typically, unsupervised learning involves clustering techniques, which group similar data points together. This thesis does not cover unsupervised learning; readers interested in this topic can refer to [18,72,74].

The third common approach is *semi-supervised* learning. As described in [18,74], the dataset consists of both labelled data, denoted $\mathcal{D}_l = \{(\boldsymbol{x}_i, y_i)|i = 1, ..., N_l\}$, and unlabelled data, denoted $\mathcal{D}_u = \{\boldsymbol{x}_i|i = N_l + 1, ..., N_l + N_u\}$ typically with $N_l << N_u$. The total number of data points is denoted $N_l + N_u = N$. In semi-supervised learning, the small number of labelled target data can assist the unsupervised learning process towards faster convergence and better model performance.

4.4 The machine learning approach

After deciding upon f, the learning process is generally performed by first dividing the available data into training, test and potentially also a validation dataset, i.e. $\mathcal{D}_{tr}, \mathcal{D}_{te}$ and $\mathcal{D}_{val}$. Each dataset consists of a finite set of data points from $\mathcal{X}$ and possibly also from $\mathcal{Y}$, i.e. $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)|i = 1, ..., N, d = 1, ..., D\}$. As previously described, the machine learning algorithm is trained to find the optimal model parameters from $\mathcal{D}_{tr}$, which enables f to generalise well to new unseen data points in the test dataset $\mathcal{D}_{te}$. Viewing linear regression from a machine learning perspective, the model parameters are referred to as the *weights*, $\boldsymbol{w}$, and the *bias*, b. Linear regression for a feature vector $\boldsymbol{x}_i$, can thus

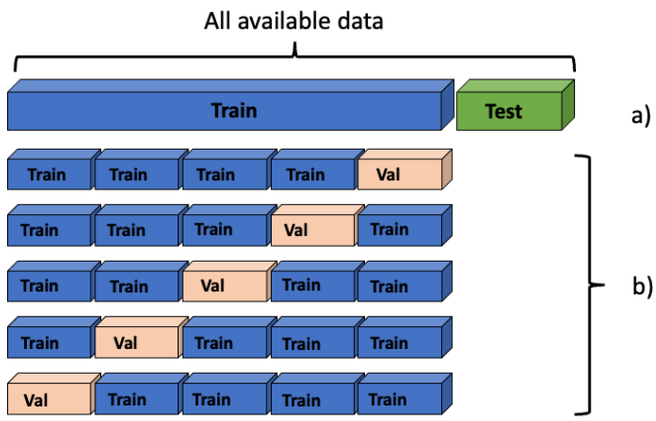


Figure 4.2: 5-fold CV. a) The available data is initially divided into a training and a test set. The test set is held outside during CV. b) The training set is divided into $k = 5$ different sets, that rotationally are used for training and validation sets. The model is iteratively trained on the $k - 1$ training sets and evaluated on the remaining k-th validation set.

be formalised as:

$$\hat{y}_i = \boldsymbol{w}^T \boldsymbol{x}_i + b = w_1 x_{i,1} + w_2 x_{i,2} + \cdots + w_D x_{i,D} + b \tag{4.1}$$

Cross-validation (CV) is a widely used evaluation technique for more complex machine learning problems, which may involve setting a vast number of model parameters along with deciding upon model architecture and hyperparameters [18, 77]. In this thesis, we refer to hyperparameters as parameters that, in contrast to model parameters, are set before the model is trained [18]. These can, for example, include learning rate, model architecture, the number of hidden layers or the number of neurons in a MLP. See Chapter 5 for descriptions of these terms.

Hold-out Validation, *Leave-one-out CV* and *k-fold CV* are three common types of CV techniques, where the latter is the most common [18, 74, 77]. Figure 4.2 illustrates *k-fold cross validation* with $k = 5$. Firstly, the available dataset is divided into a separate train and test dataset, i.e. $\mathcal{D}_{tr}$ and $\mathcal{D}_{te}$. $\mathcal{D}_{tr}$ is then divided into k nonoverlapping subsets. For each trial, $k - 1$ of the subsets are combined into a new CV training dataset fold, $\mathcal{D}_{CV_{tr}}$, while the remaining *k-th* set is used as the validation dataset $\mathcal{D}_{CV_{val}}$. By training k models on each $\mathcal{D}_{CV_{tr}}$ fold, $\mathcal{D}_{CV_{val}}$ can be used to compute a validation error using, for example, the RMSE. By averaging this error over all folds, the performance of different models or the impact of different hyperparameters can be compared. Eventually, after deciding on the optimal hyperparameters, the model is trained on the complete set of training data and evaluated on $\mathcal{D}_{te}$.

5

Deep learning basics

This chapter briefly introduces the basic deep learning (DL) theory that is the fundamentals of Papers I, II and III. Firstly, Section 5.1 focuses on giving an overview of the MLP, which holds the foundations of more advanced DL methods. Section 5.2 transitions from discussing the MLP, and introduces the foundational principles of the CNN, which serve as an example of a DL architecture specifically designed to process image data effectively [18, 69]. Lastly, Section 5.3 briefly introduces some common CNN-based DL architectures, which form the basis of the architectures used in Papers I, II and III.

Artificial intelligence (AI) refers to everything that relates to incorporating human intelligence into computer systems, that aims to learn, solve or perform tasks that generally require human intelligence [78, 79]. Machine learning (ML) is together with DL subfields of AI, where DL also is a subfield of ML [18, 68, 80]. Figure 5.1 illustrates the relation between AI, ML and DL, as well as the main differences between ML and DL: While conventional ML, like DL, focuses on detecting and extrapolating patterns from data in the learning process, ML techniques are limited by the need for domain expertise to extract representative features from the data, from which the learning system could learn. By stacking different processing layers into one model, DL models are, on the other hand, able to intelligently learn feature representations directly from the data in an end-to-end fashion [68]. Here, end-to-end learning refers to a NN that performs feature extraction and prediction simultaneously. Since all network layers are differentiable, they are also optimised simultaneously [81]. The stack of processing layers in the DL model can be thought of as a stack of

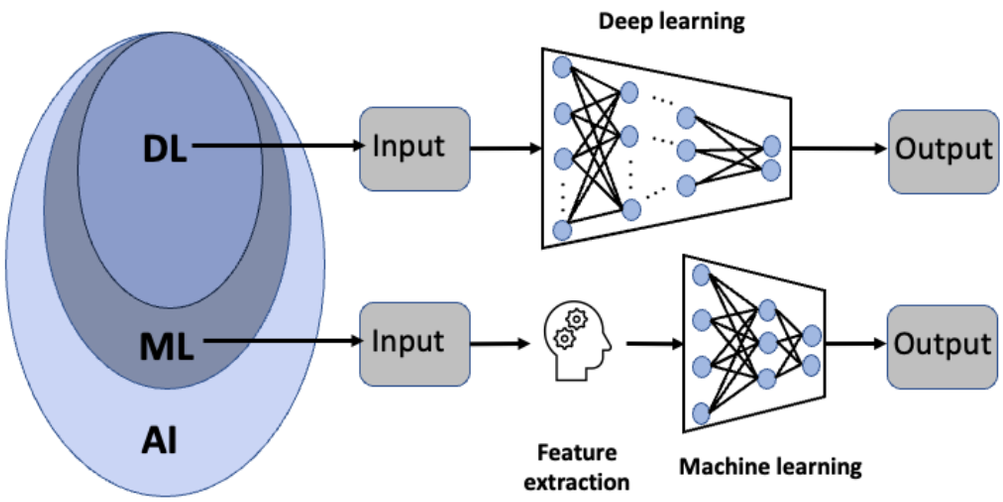


Figure 5.1: A diagram showing the relationship between AI, ML and DL, and the difference between ML and DL. While ML models necessitate a distinct feature extraction phase before they can be trained, DL training incorporates feature extraction as an integral part of the process.

successive and different data transformations that enable the model to learn different levels of information from the raw input data. In each iteration of the learning process, data is propagated layer-wise through the network, which forces the learned feature representation to be updated and refined by each iteration. The specific parts of the learning process of a DL network, as well as different CNNs covered by Papers I, II and III, will be described further in the coming sections. For clarity purposes, we will focus the discussion of this chapter on DL in the setting of supervised learning and regression. The main references for this chapter are [18, 72, 74]. When no other references are explicitly cited, we kindly refer to these works for more details.

5.1 Multilayer perceptrons

Multilayer perceptrons, also known as fully connected NNs or deep feedforward networks, are the basis for deep learning architectures. They are constructed by stacking layers of mathematical functions sequentially to successively transform the input data into the desired representation of the output data. Each layer of the MLP consists of a number of units, commonly referred to as neurons or nodes, where the mathematical mapping of the data occurs.

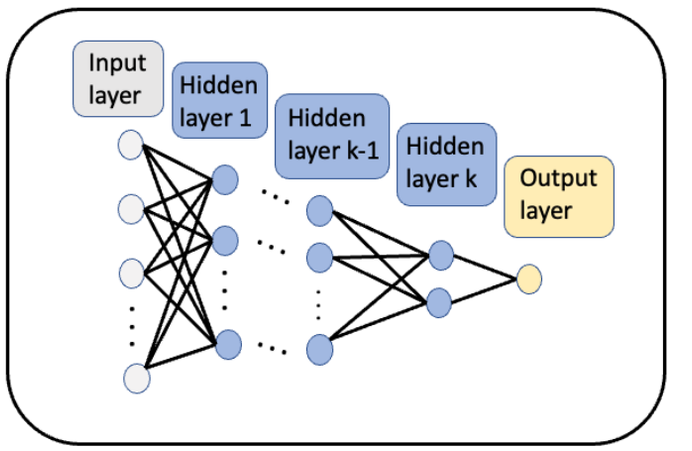


Figure 5.2: A simplified illustration of a MLP with a single output, where network neurons are symbolised with filled circles.

An illustration of a MLP is shown in Figure 5.2, where the filled circles represent the neurons, and each column of neurons represents a layer in the MLP. The layers between the input and output layers are referred to as hidden layers, where the total number of layers represents the depth of the network. Since all neurons in neighbouring layers of the MLP are fully connected, the number of parameters that must be optimised increases sharply with the number of neurons per layer and the depth of the network. Consequently, MLPs are typically only a few layers deep in contrast to e.g. CNNs.

Mathematically, the learning of a MLP can be described as approximating the function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that defines the desired mapping $y = f(\boldsymbol{x}; \boldsymbol{\theta})$ and by learning the optimal model parameters $\boldsymbol{\theta}^*$ that best approximate f. A mapping is performed within each node by first computing the linear combination between the input to the node, and the set of weights and a bias term, followed by a nonlinear activation function, i.e.

$$f_j^{(l)}(\boldsymbol{x}; \boldsymbol{\theta}) = f_j^{(l)}(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{b}) = a_j^{(l)}\big(f_j^{(l-1)}(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{b})^T \boldsymbol{W}^{(l)} + \boldsymbol{b}^{(l)}\big), \tag{5.1}$$

for $l = 1, ..., L$, where L denotes the total number of layers in the MLP, $\boldsymbol{W}^{(l)}$ denotes the set of weights, $\boldsymbol{b}^{(l)}$ the bias terms, both at layer l, while $f_j^{(l-1)}(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{b})$ denotes the output of the jth node from the previous layer. Thus, the output after Lth transformations can be represented as

$$f(\boldsymbol{x}; \boldsymbol{\theta}) = f_j^{(L)}(f_j^{(L-1)}(\cdots(f_j^{(1)}(\boldsymbol{x}; \boldsymbol{\theta})))).$$

Activation function: The activation function accounts for the nonlinearity in the MLP transformation of the input data. One of the most common activation functions for deep networks is the rectified linear unit (ReLU) [82]

$$a_{ReLU}(x) = max(0, x), \tag{5.2}$$

i.e. the element-wise maximum between 0 and the input to the activation function, here represented as x. By definition, a network having the ReLU as the final activation function will never output negative values. Therefore, both Paper I and Paper III employed it to ensure that the deep regression models never provide negative predictions for the nonnegative forest parameters considered.

Model optimisation and loss function: Generally, the optimisation algorithm and the loss function determine the learning of the optimal $\boldsymbol{\theta}^*$. The loss function determines how well the outcome from the MLP, $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{\theta})$, corresponds to the desired output, $\boldsymbol{y}$, and should be chosen based on the learning task. For regression purposes, the $\mathcal{L}_1$ loss

$$\mathcal{L}_1 = \sum_{i=1}^{N} ||\boldsymbol{y}_i - f(\boldsymbol{x}_i; \boldsymbol{\theta})|| = \sum_{i=1}^{N} ||\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i||, \tag{5.3}$$

or the $\mathcal{L}_2$ loss

$$\mathcal{L}_2 = \sum_{i=1}^{N} ||\boldsymbol{y}_i - f(\boldsymbol{x}_i; \boldsymbol{\theta})||^2 = \sum_{i=1}^{N} ||\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i||^2, \tag{5.4}$$

are two popular choices, where $||\cdot||$ denotes the absolute value and $||\cdot||^2$ denotes the squared Euclidean norm. To compute the loss, a mini-batch comprising of k samples is typically chosen from the entire training dataset. The mini-batch is denoted as $(\boldsymbol{x}_l, \boldsymbol{y}_l)|l = 1, ..., k$, with k being significantly smaller than the total number of training samples N. The total cost is then determined by computing the average loss over this mini-batch

$$J(\boldsymbol{\theta}) = \frac{1}{k} \sum_{l=1}^{k} \mathcal{L}(\boldsymbol{y}, f(\boldsymbol{x}; \boldsymbol{\theta})). \tag{5.5}$$

The chosen optimisation algorithm decides how the model parameters should be updated to reduce $J(\boldsymbol{\theta})$ to optimise $\boldsymbol{\theta}$. For MLP, the optimisation is commonly conducted through iterative backpropagation and some gradient-based learning algorithms such as the stochastic gradient descent (SGD), or the more sophisticated ADAM optimiser [83]. Regardless of the optimisation algorithm, it comes with a set of hyperparameters, such as the learning rate (lr), that must

be defined correctly to improve the optimisation process in terms of speed and performance. See [18, 72, 74] for more details on the backpropagation algorithm, gradient-based learning and common hyperparameters for different optimisation algorithms.

5.2 Convolutional neural networks

While building on the principles of MLP, CNNs are specially designed to efficiently process grid-like data, such as 2-D or 3-D image data, through mathematical operations called convolutions. The response from a 3-D discrete convolution between an input image $\boldsymbol{X} \in \mathbb{R}^{H \times W \times C}$ and a convolutional filter $\boldsymbol{K} \in \mathbb{R}^{h \times w \times c}$ with $h << H, w << W$ and $c << C$ is defined by

$$\boldsymbol{Z} = (\boldsymbol{X} \star \boldsymbol{K})(i, j, d) = \sum_{l=-s}^{s} \sum_{m=-t}^{t} \sum_{n=-u}^{u} X_{i-l,j-m,d-n} K_{i,j,k}, \tag{5.6}$$

with $i = 1, ..., H-h+1$, $j = 1, ..., W-w+1$ and $d = 1, ..., C-c+1$ together with $s = \frac{h-1}{2}$, $t = \frac{w-1}{2}$ and $u = \frac{c-1}{2}$. Here, H and h represent the image and kernel window height, W and w the image and kernel window width, C and c the number of channels in the image or the kernel window, while $\star$ represents the convolution operation. By sliding the filter $\boldsymbol{K}$ across the input image $\boldsymbol{X}$ using a predefined *stride*, which is typically a constant value set to one or two, the output, $\boldsymbol{Z}$, known as a *feature map*, is obtained. Multiple convolutional filters are usually learned simultaneously in each stage of the network, resulting in a stack of feature maps, as illustrated in Figure 5.3.

CNNs are designed to utilise spatial relationships in data through convolutional filters, also known as kernels, in their convolutional layers. This approach enables CNNs to effectively process grid-like data, such as images, and extract useful features that may be subsequently used for the prediction task. In contrast, MLPs process 2-D or 3-D images by considering each pixel individually or vectorising pixels before processing them. Therefore, they do not leverage the spatial context that describes e.g. the presence of corners or edges. The *receptive field* is the neighbourhood in the input image that influences the features of a given pixel in a subsequent layer, shown as blue regions in Figure 5.3. The receptive field grows from layer to layer according to the kernel size of the convolutional filters, meaning that an increasing amount of contextual information is considered as the processing reaches deeper layers. As a result, CNNs are contextual models. Thus, contrary to MLPs, the use of convolutional layers in CNN-based regression models enables the models to learn spatial contextual relations.

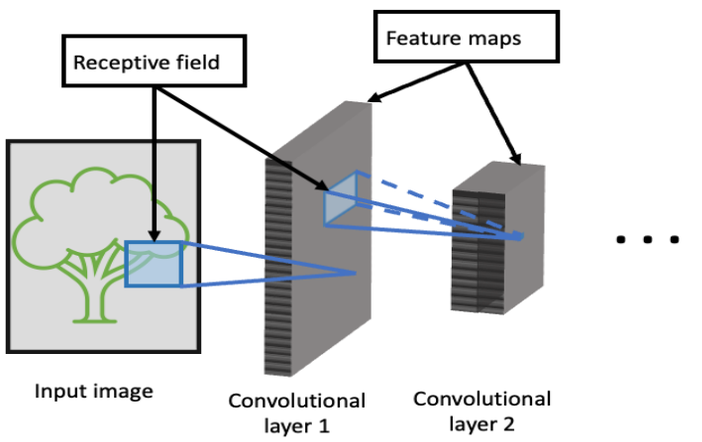


Figure 5.3: A simplified example of a CNN architecture. Input images are processed by convolutional layers using kernels to generate feature maps. The convolutional layers commonly include pooling, normalisation and activation (e.g. ReLU) layers.

The initial components of the CNN architecture are depicted in Figure 5.3. To simplify the diagram, the individual processing layers are not shown individually, but are included within each convolutional layer in the diagram. We will in the following briefly introduce some of the most commonly used layers included in most CNN architectures.

Activation and normalisation layers: Like MLPs, CNNs also use a nonlinear activation function to process the outputs of the convolutional operation. One popular choice is the ReLU activation function, as research has shown that deep CNNs with ReLU activation can train faster than those with hyperbolic tangent (tanh) activation [69]. Moreover, normalisation layers are frequently utilised before or after the activation layer to improve model performance and stability during training. Batch normalisation (BN) [84] is one of the most common normalisation techniques, in which the output from the convolutional filters or the activation layer is normalised by using the training mini-batch statistics. Other common normalisation techniques are layer normalisation [85] or instance normalisation [86].

Pooling layer: The pooling operation in the CNN replaces the output of a layer at a certain location with a summary of statistics of nearby pixels. Compared to the normalisation layers, the pooling operation reduces the spatial dimension and makes the feature representation invariant to small translations in the input. The most common pooling operation is the *max pooling* operation, which replaces the values in an $n \times n$ neighbourhood with its maximum value. Like the convolutional operation, the pooling filter is applied across the entire feature

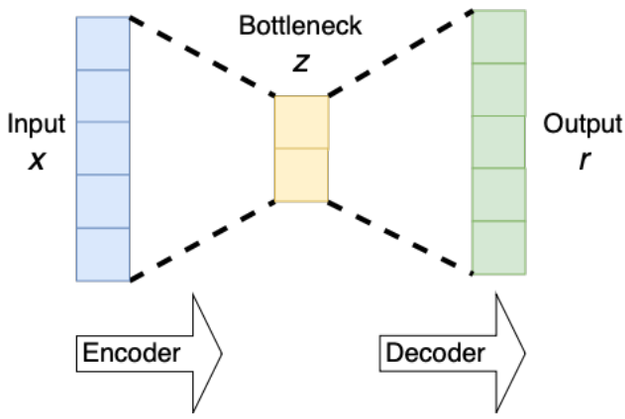


Figure 5.4: A simplified illustration of a traditional autoencoder. The AE utilises the encoder part of the network to map an input $\boldsymbol{x}$ to the code space, shown as the bottleneck in the figure. The decoder aims to reconstruct $\boldsymbol{x}$ from the latent representation of it, i.e. $\boldsymbol{z}$.

map.

5.2.1 Dataset augmentation

Dataset augmentation is a common regularisation technique for both classification and regression purposes, applied to improve model generalisation on new unseen data and increase the training dataset's size. Different augmentation techniques exist, with image rotation and flipping being two common approaches. Image rotation implies that the position of the image data in an image patch is rotated by increments of e.g. $[0, 90, 180, 270]$ degrees, while flipping implies that image data is horizontally or vertically flipped around the image centre.

5.3 CNN architectures

5.3.1 Traditional autoencoders

An autoencoder (AE) is a NN, that consist of two sequential mapping functions, known as the encoder-decoder pair (f, g). Mathematically, the AE aims to optimise

$$f : \mathcal{X} \rightarrow \mathcal{Z} \tag{5.7}$$
$$g : \mathcal{Z} \rightarrow \mathcal{X}, \tag{5.8}$$

where the input domain data is represented by $\mathcal{X}$, and the *code space* or *latent space* data is represented by $\mathcal{Z}$. Generally, the optimisation is performed by minimising a loss function, such as the mean squared error (MSE), computed between the input and the reconstruction, i.e. $\mathcal{L}(\boldsymbol{x}, g(f(\boldsymbol{x}; \boldsymbol{\theta}))) = \mathcal{L}(\boldsymbol{x}, \boldsymbol{r})$, where $\boldsymbol{r} \approx \boldsymbol{x}$ represents the *reconstruction* of $\boldsymbol{x}$.

Merely training the AE to learn two identity mappings, i.e. $\boldsymbol{x} = f(\boldsymbol{x}; \boldsymbol{\theta}) = g(f(\boldsymbol{x}; \boldsymbol{\theta}))$, for the training dataset is not sufficient for the AE to acquire significant features of the input data. Consequently, the AE would not generalise well on new, unseen test data. To overcome this limitation, many AEs constrain the latent representation $\boldsymbol{z}$ to have a smaller dimension than the input $\boldsymbol{x}$ [18], resulting in an AE network with a bottleneck, as shown in Figure 5.4. Thus, the encoder network aims to compress input data $\boldsymbol{x}$ to high-level feature representations, denoted by the latent representation $\boldsymbol{z}$. Following the compression phase, the decoder network is trained to up-sample $\boldsymbol{z}$ to achieve the reconstructed version of $\boldsymbol{x}$. AEs have traditionally been utilised for dimensional reduction or feature learning tasks for e.g. image, sound or video data.

Up-sampling techniques: The encoder down-sampling process generally involves variations of the convolutional layers discussed in Section 5.2. Different methods exist for the up-sampling process performed by the decoder network, where the two most common methods are *up-convolution by interpolation* and *transposed convolution* [87]. Figure 5.5 depicts the up-sampling of a 2×2 low-resolution image to a 5×5 output image (green) through up-convolution by interpolation (left) and transposed convolution (right).

Both up-sampling methods involve two stages; In up-sampling by interpolation, the image resolution is firstly increased through the nearest neighbour or bilinear interpolation. In up-sampling through transposed convolution, the image resolution is first increased by inserting zero-value pixels between the image's original pixel values. Secondly, the interpolated or zero-imputed image is convolved with a standard convolutional filter to achieve the final output image [87, 88].

5.3.2 ResNet

Residual networks (ResNets) [89] are a popular family of CNN architectures, which employ *skip connections*, also referred to as *short-cut connections*, to overcome issues with vanishing gradients. These ease the training of very deep CNN networks. The popular ResNet-34 architecture is depicted in the upper part of Figure 5.6 without the output layer. It consists of 34 convolutional layers distributed among the input layer, four convolutional blocks (denoted by "Layer X") and the fully-connected output layer. By removing the fully-connected out-

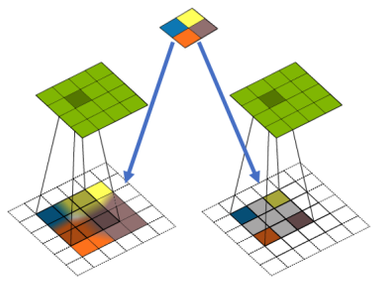
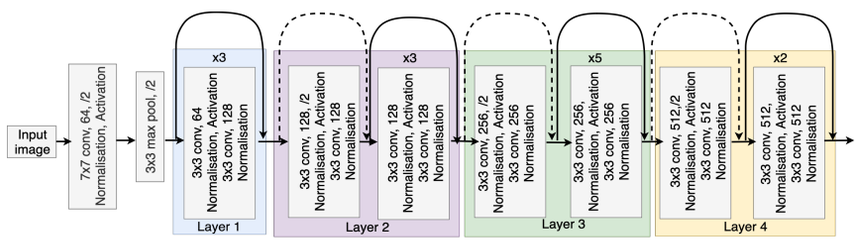


Figure 5.5: Overview of two common up-sampling techniques. **Above**: a 2×2 low-resolution image, **Left**: up-convolution by interpolation and **Rigth**: transposed convolution. The figure is a modified version of a similar one presented in [87].

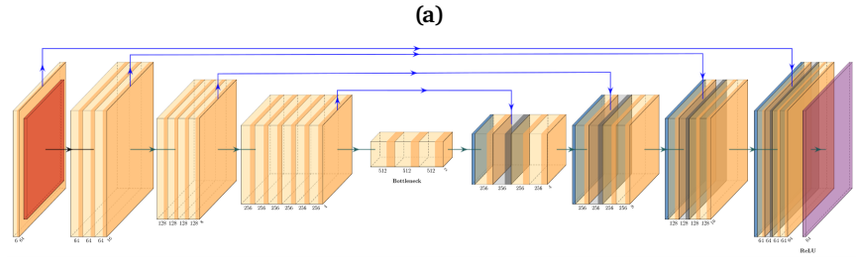
put layer, responsible for mapping the input image to a class prediction, the ResNet-34 architecture can serve as the encoder network in an AE. The solid and dashed lines in the upper part of Figure 5.6 represent the short-cut connections of the ResNet, i.e. the input of a coloured residual block is added to the output of the residual block. At the beginning of Layers 2, 3 and 4, the dimensions of the feature maps are increased. In contrast, the spatial dimension of the filters decreases. Thus, to perform the short-cut connection, the input to the block is firstly convolved with a 1×1 filter using a stride of 2 to achieve the same spatial dimension reduction. This is shown as the dashed short-cut connections in the figure.

5.3.3 U-Net

The U-Net [90], initially invented for semantic segmentation tasks, uses a symmetric u-shaped encoder-decoder structure. A flattened horizontal representation of a U-Net is shown in the lower part of Figure 5.6. Different versions are employed for the encoder network, with the previously introduced ResNet being one example. Compared to the ResNet, the U-Net uses skip-connections between symmetric encoder-decoder blocks, illustrated with blue arrows in the figure. This enables the network to learn to reconstruct and unravel data structures from both low-level feature maps from the encoder and high-level feature maps from the decoder. In contrast to the ResNet, the combination of feature maps through the U-Net skip-connections is performed through concatenation. By replacing the segmentation head, originally employed by the U-Net to map the input image to a segmentation map, with a ReLU activation layer, the U-Net can be employed for regression purposes, and nonnegative output predictions will be ensured.



(a)



(b)

Figure 5.6: (a) ResNet-34 network without the fully-connected output layer. Grey boxes are residual blocks, "3 × 3 *conv 64,* \2" is a 3 × 3 convolutional layer yielding a feature map of 64 filters and halved spatial dimension, *Activation* is e.g. a ReLU function and *Normalisation* is e.g. a BN layer. **(b)** An encoder-decoder U-Net. Orange blocks represent convolutional layers (e.g. residual blocks), while blue arrows are U-Net skip-connections.

/6

Deep learning regression models for forestry

Despite the considerable progress made by DL in processing complex and high-dimensional data across diverse fields and applications, its utilisation for remote sensing approaches in forestry is still in the early stages, although some work has emerged [3, 4, 17]. A recent review by Kattenborn *et al.* [3] examines the use of CNN in remote sensing of vegetation, including studies within agriculture, conservation and forestry. It finds that about 91% of the reviewed studies focus on classification tasks, and that only a minority of all studies reviewed addressed research questions specific to forestry, such as forest biomass prediction. Despite the potential in utilising CNN as contextual regression models, research on and applications of deep convolutional regression models for forest parameter retrieval remain limited. Hamedianfar *et al.* [4] emphasise the challenge of acquiring massive amounts of target data, such as ground reference measurements representing AGB or SV, as these require labour-intensive, time-consuming, and costly field inventory campaigns [4, 15]. This challenge poses a critical factor that has limited the development of robust DL models for forest parameter retrieval. This thesis aims to address some of the obstacles encountered, including the limited availability of ground reference data and the lack of CNN architectures designed specifically for the prediction of biophysical or geophysical parameters from RS image data. To this end, it proposes novel methodologies for deep convolutional regression models for large-scale and low-cost forest parameter retrieval.

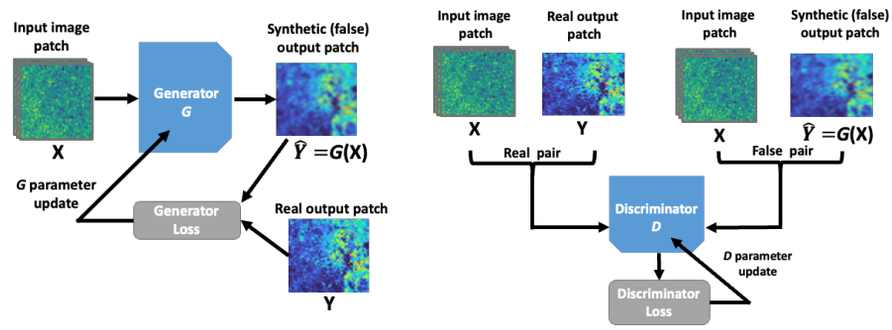
Building upon the foundational knowledge of DL provided in Chapter 5, this chapter provides an overview of DL methods and theories included in the three papers that form the core of this thesis. Section 6.1 provides the theoretical background of some popular convolutional generative models and relates generative modelling to image-to-image (I2I) translation. We focus on how these architectures can be used in the regression setting, particularly emphasising the deep generative CNN regression models employed in Papers I, II and III. Section 6.2 describes how pixel- or frequency-aware convolutional regression models can be trained using concepts from I2I. This section also discusses the theoretical background of frequency-aware training and provides the frequency-aware objective function proposed in Paper II. Finally, Section 6.3 further describes how deep convolutional regression models are employed by the deep convolutional regression models proposed in Papers I and III.

6.1 Generative models

Generative models are extensively employed in deep learning for cross-modal translation of image data from one distribution to another, using a learned data distribution to generate new data points with desired characteristics [18, 91]. Although the group of generative models includes numerous types and architectures, we limit the following background theory of generative models to those used in Papers I, II and III: the generative adversarial network (GAN) [92], the cGAN and the VAE.For comprehensive overviews of generative models and other popular architectures, we refer to [18, 72, 91].

6.1.1 Generative adversarial networks

In 2014, Goodfellow *et al.* introduced the GAN, which has since become one of the most widely used generative models [92]. Its most basic form consists of a generator (G) and a discriminator (D) network. The primary objective of G is to learn the optimal mapping from a random noise vector to a target image that can fool D. On the other hand, the discriminator network aims to correctly distinguish between image samples generated from G and samples of the true target image. The iterative process of adversarial training involves optimising two conflicting objective functions for G and D, which should result in simultaneous improvement of both networks, eventually leading to model convergence. A popular extension of the basic GAN is the cGAN, which distinguishes itself from the GAN by conditioning the mapping of G on images from the input domain instead of a noise vector. Hence, the generator performs a mapping $G : \boldsymbol{X} \rightarrow \boldsymbol{Y}$ that can be utilised for regression purposes. Similarly, the D network is extended to learn to distinguish between a real pair of input



(a) *G* training and model update. **(b)** *D* training and model update.

Figure 6.1: Illustration of a cGAN with its two model components, the generator network (G) and the discriminator network (D). G is trained to translate image patches from the input domain into realistic-looking image patches that resemble data from the target domain. D is trained to detect the difference between a real pair of input and target image patches, $(\boldsymbol{X}, \boldsymbol{Y})$, and a false pair, $(\boldsymbol{X}, G(\boldsymbol{X}))$.

and target images, $(\boldsymbol{X}, \boldsymbol{Y})$ and a false pair, $(\boldsymbol{X}, G(\boldsymbol{X}))$, where $G(\boldsymbol{X})$ refers to images generated by G. The whole training process of G and D of the cGAN is illustrated in Figure 6.1.

6.1.2 Image-to-image translation

Generative models are today used in many image analysis tasks, such as face generation [93, 94], cross-modal image translation, style transfer and image-to-image translation [95–97]. One of the earliest and possibly also most famous methods for I2I translation is the *Pix2Pix* model designed by Isola *et al.* [95], which utilises a cGANs to translate images from one domain or representation to another, e.g translation of greyscale images to corresponding RGB images or converting sketches into paintings.

6.1.3 Training a cGAN for image-to-image translation

To train a cGAN for I2I translation, different suggestions for the G network exist. The Pix2Pix model provides two options: a U-Net with a tanh activation function in the output, or an encoder-decoder network with ResNet blocks in the bottleneck and a tanh output activation function. Similarly, different variations of the D network are proposed by adjusting the patch size N of the discriminator's receptive fields, ranging from a 1×1 PixelGAN to an $N \times N$ PatchGAN [95]. The D network applies convolutional processing to the pair of input image patches to produce multiple classification responses, which are

then averaged to determine whether the processed pair of image patches is a real or false pair.

Training objectives:
Adversarial training of G and D in the most basic cGAN setting results in the following Vanilla GAN (VGAN) min-max objective function [95]:

$$\begin{aligned}\min_G \max_D \mathcal{L}_{VGAN}(D,G) = \mathbb{E}_{X,Y}[\log D(\boldsymbol{X},\boldsymbol{Y})] \\ +\mathbb{E}_{\boldsymbol{X}}[\log(1-D(\boldsymbol{X},G(\boldsymbol{X})))],\end{aligned} \tag{6.1}$$

where $\boldsymbol{X}$ represents an image patch from the input domain, $\boldsymbol{Y}$ is an image patch from the target domain and $G(\boldsymbol{X})$ is a generated image patch. However, many variations to the original Vanilla GAN objective function exist, such as the Wasserstein GAN with gradient penalty (WGAN-GP), which was proposed for increased training stabilisation and high-quality image generation [98], or the Least Squares GAN (LSGAN) [99], which in the conditional setting utilises the following objective functions

$$\begin{aligned}\min_D \mathcal{L}_{LSGAN}(D) =& \frac{1}{2}\mathbb{E}_{X,Y}[(D(\boldsymbol{X},\boldsymbol{Y})-b)^2]+ \\ & \frac{1}{2}\mathbb{E}_{\boldsymbol{X}}[(D(\boldsymbol{X},G(\boldsymbol{X}))-a)^2] \\ \min_G \mathcal{L}_{LSGAN}(G) =& \frac{1}{2}\mathbb{E}_{\boldsymbol{X}}[(D(\boldsymbol{X},G(\boldsymbol{X}))-c)^2],\end{aligned} \tag{6.2}$$

where c denotes a value that G tricks D to believe for false data, while a and b are the labels used by the discriminator for false and real data [99].

6.1.4 Variational autoencoders

Another generative model family is the VAEs that, similarly to the AE, utilise a coupled encoder-decoder network to generate data. However, the differences between the simple AE and the VAE are many. The AE is a deterministic model that aims to obtain the optimal reconstruction of the input data and is, by definition, not a generative model [18]. The VAE, on the other hand, is a probabilistic Bayesian generative model, implying that samples of the data distribution could be generated through its marginal likelihood $p_\theta(\boldsymbol{x}) = \int p_\theta(\boldsymbol{z})p_\theta(\boldsymbol{x}|\boldsymbol{z})d\boldsymbol{z}$. However, as in the usual Bayesian setup, the marginal likelihood or the corresponding true posterior distribution $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ is intractable to compute directly.

The workaround to enable the generation of data from the VAE model is to train an encoder model, $q_\phi(\boldsymbol{z}|\boldsymbol{x})$, also referred to as the *recognition model*, to map input data $\boldsymbol{x}$ to a low-dimensional *latent representations* of the data $\boldsymbol{z}$. Thus, given samples from the distribution of $\boldsymbol{z}$, the probabilistic decoder model, $p_\theta(\boldsymbol{x}|\boldsymbol{z})$,

can be utilised to generate a distribution over possible values of $\boldsymbol{x}$ [18, 29]. Letting ϕ represent the parameters of the encoder network and θ the parameters of the decoder network, the optimisation problem boils down to achieving $q_\phi(\boldsymbol{z}|\boldsymbol{x}) \approx p_\theta(\boldsymbol{z}|\boldsymbol{x})$, where $q_\phi(\cdot)$ and $p_\theta(\cdot)$ can be parameterised using deep NNs or CNNs [100]. Using the reparametrisation trick [29], assuming that the prior $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and that $q_\theta(\boldsymbol{z}|\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ the VAE can be trained by maximising the variational lower bound $\mathcal{L}(\theta, \phi; \boldsymbol{x})$ associated with $\boldsymbol{x}$, formulated as

$$\mathcal{L}(\theta, \phi; \boldsymbol{x}) = \mathbb{E}_{\boldsymbol{z}\sim q_\phi(\boldsymbol{z}|\boldsymbol{x})} log(p_\theta(\boldsymbol{x}|\boldsymbol{z})) - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})), \tag{6.3}$$

where the Kullback Leibler divergence $D_{KL}(\cdot)$ enforce the approximate posterior distribution $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ and the model prior $p_\theta(\boldsymbol{z})$ to approach each other [18]. For full derivations, see [29,100].

6.2 Pixel- and frequency-aware convolutional regression models

By exploiting concepts from I2I translation, simpler nonadversarial convolutional models for regression can be defined. That is, we can train a chosen DL architecture to translate images from one image domain to another without using a GAN-type loss. An example of such an architecture is the U-Net, which was initially designed for image segmentation tasks [90]. When determining the appropriate deep learning architecture, the convolutional regression model is optimised by selecting suitable loss functions and using output activation layers relevant to the regression task. In image segmentation, the softmax activation function is commonly used [90,101]. Alternatively, for other tasks, it can be omitted [102] or replaced with e.g. the ReLU activation function to predict targets that are nonnegative and unbounded from above as we propose for Paper III, see Section 11.

In contrast to adversarial training, which aims to improve the perceptual quality of images, pixel-wise training minimises some pixel-wise error measure between the prediction $\hat{y}_{i,j}$ and the corresponding target $y_{i,j}$ for all pixels (i, j) in the image. Thus, in some applications, a pixel-wise regression U-Net might be better suited to provide prediction maps with high accuracy measured in terms of a low RMSE or MAE. This could imply optimising the pixel-wise regression model on the $\mathcal{L}_1$ (see Eq. (5.3)) or $\mathcal{L}_2$ (see Eq. (5.4)) loss functions, which are closely connected with RMSE or MAE. In Paper III, the definition of [102] is adopted: A U-Net trained to translate RS image patches into forest prediction

maps through optimisation of a pixel-wise objective function is referred to as a *pixel-wise regression U-Net*.

In other applications, pixel-wise losses, however, are known to lead to model-generated images with blurry appearance [103,104]. This implies that the training paradigm and loss functions should be selected wisely. Recent research on convolutional generative models has indicated that challenges related to blurriness and lack of details in CNN-generated images could be due to issues referred to as *Fourier spectrum discrepancy*, *spectral inconsistency*, *frequency bias* or *spectral bias* [24–26, 87, 88, 105–107]. These terms are used interchangeably in the literature and refer to the difficulty that CNNs face in learning to generate high-frequency image components, such as structures, edges and textures, leading to blurry and less detailed images. One suggested reason for this behaviour is that NNs prioritise learning low-frequency image data components first, with the consequence that higher frequencies are learnt later or to a very small extent in the optimisation process [26]. To address this challenge, researchers have proposed to use frequency-aware spectral losses in combination with common pixel-aware or GAN losses to force the model to preserve the frequency content of the image data during training [87,106,107].

In 2020, Durall *et al.* [87] proposed to add a frequency-aware loss to the GAN's generator loss, $\mathcal{L}_G$, to force the G network to also focus on the spectral agreement during training. Thus, the updated G loss can be formulated as

$$\mathcal{L}_{G_{total}} = \mathcal{L}_G + \lambda \mathcal{L}_{AzI}, \tag{6.4}$$

where λ is a hyperparameter that weights the influence of their spectral loss $\mathcal{L}_{AzI}$, where (AZI) represents the azimuthal integral, i.e. a 1-D representation of the Fourier power spectrum. Their spectral loss $\mathcal{L}_{AzI}$ is given by

$$\begin{aligned} \mathcal{L}_{AzI} = \frac{1}{N/2-1} \sum_{i=0}^{N/2-1} & AzI(\boldsymbol{y}_i) \cdot \log AzI(\hat{\boldsymbol{y}}_i) \\ & + (1 - AzI(\boldsymbol{y}_i)) \cdot \log(1 - AzI(\hat{\boldsymbol{y}}_i)) \,, \end{aligned} \tag{6.5}$$

where N represents the image size, $\boldsymbol{y}_i$ is an image pixel from the potentially multidimensional target image, $\hat{\boldsymbol{y}}_i$ is the corresponding predicted image pixel. Mathematically, $\mathcal{L}_{AzI}$, is computed through azimuthal integration over the radial frequencies, ϕ, that are present in the 2-D Fourier transform $\mathcal{F}(I)$ of an input image I with size $N \times N$, i.e.

$$AzI(\omega_k) = \int_0^{2\pi} ||\mathcal{F}(I)(\omega_k \times \cos(\phi), \omega_k \times \sin(\phi))||^2 d\phi, \tag{6.6}$$

for $k = 0, ..., N/2$.1. See [87] for details on the loss and results for models trained with Eq. (6.4).

Another spectral-aware loss was proposed by Czolbe *et al.* [107], who suggest to improve the training of VAEs with a loss function based on a modified version of Watson's perceptual model of the human visual system [108]. While Watsons's original model computes the loss between true and generated images as a weighted distance in the frequency space and using the discrete cosine transform, [107] suggests to replace the discrete cosine transform with the discrete Fourier transform. This results in a loss that can be applied to colour images and to achieve improved robustness to translational shifts [107]. See [108] for details of the original formulation of the loss and [107] for its justification and formulation.

However, simpler and better-performing frequency-aware losses exist, with the FFT-loss proposed in [106] as one example. The FFT-loss, denoted $\mathcal{L}_{FFT}$, is defined as

$$\begin{aligned}\mathcal{L}_{FFT} = &\frac{1}{k}\sum\left(imag[\mathcal{F}(\boldsymbol{Y})] - imag[\mathcal{F}(\hat{\boldsymbol{Y}})]\right)^2 + \\ &\frac{1}{k}\sum\left(real[\mathcal{F}(\boldsymbol{Y})] - real[\mathcal{F}(\hat{\boldsymbol{Y}})]\right)^2,\end{aligned} \tag{6.7}$$

where $\mathcal{F}$ denotes the discrete Fourier transform computed for either target images $\boldsymbol{Y}$ or predicted images $\boldsymbol{Y}$ retreieved from a mini-batch of k images. In practice, this is naturally done with the fast Fourier transform (FFT), hence the name of the loss function. As seen from the definition, $\mathcal{L}_{FFT}$ uses the MSE to enforce alignment of the *real* and *imaginary* parts of the target and predicted image patches in the frequency domain. As it is complementary to existing losses [106], it can, for example, replace $\mathcal{L}_{AI}$ in Eq. (6.4) when training frequency-aware generative CNN models.

6.3 Deep learning approaches to forest parameter retrieval

The forest sector is one domain where access to labelled target data is very limited. This scarcity is evident in Tanzania, where only 88 field plots are available for the entire AOI, but also in the three Norwegian regions, where a total of 264 field plots exist. Without modifications of the target data, the limited target dataset hinders the application of deep CNNs as regression models, since CNNs require spatially continuous training datasets, comprising both input and target data. Papers I and III circumvent the issue of scarcity target data by exploiting accessible large-scale prediction maps acquired from ALS-based forest mapping campaigns. This transforms the problem issue from one where conventional statistical or pixel-based ML regression models are employed as

the first option, into a DL setting where contextual CNNs can be applied to learn effective regression models between Sentinel-1 data and the forest prediction maps.

However, the use of prediction maps acquired from ALS-based forest mapping campaigns does not ensure that these are continuous wall-to-wall maps, which is required to train contextual CNN-based regression models. Both datasets acquired by commercial forest owners and datasets retrieved through NFI campaign can be noncontiguous. As described in Section 1.1, full-coverage ALS mapping of the entire Liwale district of Tanzania was economically infeasible during the original NFI campaign [22]. Utilising this dataset as prediction targets instead of the sparse ground reference dataset would have resulted in noncontinuous ALS-derived AGB prediction maps that resemble the ALS strips shown in Figure 1.1. As introduced in Section 3.4.1, the Tanzanian ALS data used in this thesis are indeed continuous. However, this dataset was recorded at a later time, after additional funding for wall-to-wall mapping of a smaller, limited area was secured. See [28] for a description of this mapping campaign. The shifted focus to the smaller AOI in the Liwale district of Tanzania enables the use of CNN-based regression models without further modifications, see Papers I and III.

For the Norwegian ALS dataset used in Paper III, it is not the data acquisition cost that limits the spatial coverage, but the censoring of data that is not relevant to the task of the project: to build regression models for the commercial part of the forest. In this case, it makes sense to stratify the forest, since more accurate regression models can be obtained within homogeneous areas. This resulted in the spatially disjoint ALS-derived SV predictions presented in Section 3.4.2.

The examples from the campaign in support of the Tanzanian NFI and from the mapping of commercial forests in Norway show that noncontiguous prediction maps with spatially disjoint segments occur for various reasons. To leverage these noncontiguous datasets for forest parameter retrieval, it is relevant to study how CNN-based regression models can handle them. Utilising noncontinuous datasets as prediction targets implies increased computational complexity, as specialised processing is required if this data shall be applied as prediction targets when training a CNN. Paper III suggests using forest masks in the training for the Norwegian regions to confine the loss computation. As a result, the CNN models are only being actively trained in areas where ALS-derived SV predictions are available, whereas convolution operations are not disturbed or complicated. Moreover, compared to Paper I, Paper III proposes combining the small set of ground reference target data and the much denser ALS-derived SV predictions to improve model performance by drawing inspiration from a subfield of semi-supervised learning.

6.3.1 Pseudo-labels for semi-supervised learning

As described in Section 4.3, semi-supervised learning is a training paradigm that allows us to combine an amount of unlabelled data, which is often extensive, with a set of labelled data that is normally sparse. One example of how the small number of labelled target data can be utilised to boost model performance would be to train the algorithm in a supervised fashion using both labelled and unlabelled data simultaneously. Lee *et al.* proposed in 2013 to utilise semi-supervised learning with pseudo-labels to improve on tasks such as image classification [109]. The learning process is performed in two phases. Initially, the model is pretrained on the sparse set of labelled data, before it is used to predict labels on the unlabelled data to generate so-called pseudo-labels. Pseudo-labels with a high prediction confidence are combined with true target labels to augment the labelled set in the subsequent model fine-tuning. In the fine-tuning phase, the model is iteratively retrained, gradually introducing more and more pseudo-labels according to the confidence in the predictions, which can e.g. be measured in terms of distances to class prototypes. This process allows the model to learn simultaneously from true target labels and pseudo-labels, improving its performance. Since then, more recent works have built on the idea of utilising pseudo-labels in a semi-supervised fashion for image classification [110], to train multiple models simultaneously [111] or in cluster-based learning [112]. Common for semi-supervised learning with pseudo-labels is that "new" pseudo-labels iteratively are produced and added to the set of labelled data to improve model performance [109–112].

6.3.2 Regression models with imputed pseudo-targets

In Paper III, we distinguish between the ground reference target data and the ALS-derived SV predictions by defining the former as the true target dataset and the latter as the pseudo-target dataset. Due to the known high correlation between ALS echoes and forest height, the ALS-derived prediction maps are deemed to exhibit a high correlation with AGB and SV. Consequently, unlike the conventional approach of semi-supervised learning with pseudo-labels, we do not employ an iterative process to enhance the quality of these pseudo-targets. This is because we assume the ALS-derived prediction maps to possess a high degree of prediction confidence, but also because it is not trivial how the confidence in prediction targets can be computed in the regression setting.

Moreover, in Paper III, we train the deep convolutional regression models simultaneously on the true predictive targets and the pseudo-targets by imputing the pseudo-targets into the dataset with true targets. Practically, as the number of true targets is notably fewer than the extensive set of pseudo-targets for both Tanzania and the Norwegian regions, this is done by inserting the true

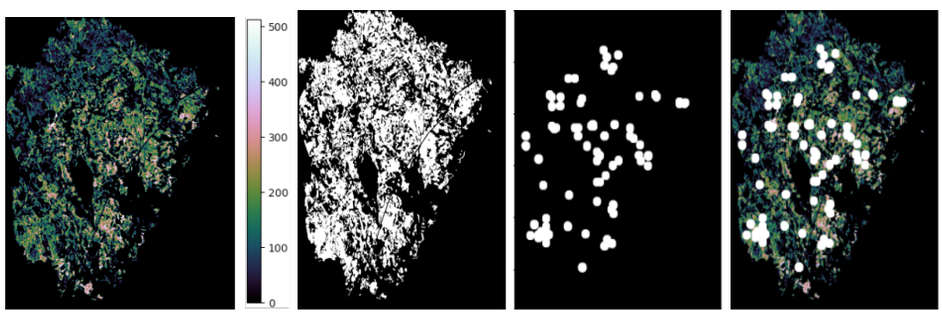


Figure 6.2: **(a)** The discontinuous ALS-derived SV prediction map, where the colour bar indicates the amount of SV. **(b)** The pseudo-target mask. **(c)** The ground reference target mask. **(d)** The ground reference dataset, imputed with pseudo-targets. target mask. All images are from the AOI in Tyristrand. The size of the pixels representing true targets has been magnified for illustrative purposes.

targets into the pseudo-target prediction maps. To enforce the deep convolutional regression models to learn from regions where both pseudo-targets and true targets are available, models are trained with two binary masks. The first is a pseudo-target mask, which for the discontinuous Norwegian datasets holds the position of each pseudo-target. In contrast, for the Tanzanian wall-to-wall map of ALS-derived predictions, the pseudo-target mask only contains ones as the datasets of pseudo-targets cover the whole AOI. The second mask, referred to as the ground reference mask, holds the positions of each target. Figure 6.2 shows the original ALS-derived SV dataset for the AOI in Tyristrand, the two masks and the target dataset imputed with pseudo-targets.

Part II

Summary of research and concluding remarks

7

Summary of research

7.1 Paper I

Sara Björk, Stian Normann Anfinsen, Erik Næsset, Terje Gobakken, and Eliakimu Zahabu. "**On the Potential of Sequential and Nonsequential Regression Models for Sentinel-1-Based Biomass Prediction in Tanzanian Miombo Forests**", in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4612-4639, 2022.

7.1.1 Summary

The majority of current methods used for forest biomass prediction, such as traditional statistical and machine learning regression methods, typically operate on a pixel-by-pixel basis. These methods establish a relationship between a limited number of individual points representing estimated ground reference AGB, and corresponding RS image pixels. The main reason for this is the need for a sufficient amount of ground reference data needed to train deep CNN-based regression models, which through their convolutional filters can exploit contextual spatial information from neighbouring pixels for each prediction. Consequently, the utilisation of DL-based methodologies for applications in

forestry is limited. As a result, most RS-based AGB regression models do not use the spatial contextual information from neighbouring pixels during the learning process.

In this paper, we propose methodologies to train DL-based AGB prediction models by employing sequential regression modelling[1], where the first regression stage has linked *in situ* AGB data to ALS data and produced the ALS-derived AGB prediction map. Formally, we in this paper focus on developing methods for the second regression model in the sequence of two. We propose to train regression models on regressor data from the Sentinel-1 sensor and utilise ALS-derived AGB prediction maps as a surrogate for ground reference data. This dramatically increases the amount of available training data and enables deep CNN-based models to be utilised.

We propose to train cGANs in a supervised setting to translate false colour image patches of Sentinel-1 backscatter into realistic-looking synthetic ALS-derived AGB prediction patches. Figure 7.1 illustrates the proposed method, where the generator network G is responsible for learning relationships between Sentinel-1 data and data from ALS-derived AGB predictions. Simultaneously, the discriminator network D is trained to distinguish between a "real" combination of image data patches from Sentinel-1 and the actual ALS-derived AGB prediction map to a "fake" combination of image data from Sentinel-1 and a generated ALS-derived AGB prediction patch. The cGAN components, the G network and the D network, are trained with the traditional minimax optimisation procedure for GANs. Following the training phase, the production of realistic-looking synthetic ALS-derived AGB prediction patches from corresponding false colour Sentinel-1 data can be achieved by utilising the trained G network in the prediction phase.

In addition to the sequential cGAN-based AGB regression models, two parametric regression models were also implemented in this paper, both utilising Sentinel-1 backscatter data as regressor data. The first model is trained in a non-sequential setting, directly associating Sentinel-1 data with ground reference data of AGB through the regression model. In contrast, the second parametric model is trained using the sequential approach, similar to the sequential cGAN-based AGB regression models. These serve as references to quantify the impact of DL-based models compared to traditional parametric models.

All regression models proposed in this work were evaluated against each other, and additional traditional nonsequential regression models that were developed for AGB prediction in the same AOI in Tanzania, see [28]. The empirical results indicate the benefit of including deep CNN-based regression models

1. See Section 3.3.1 for details.

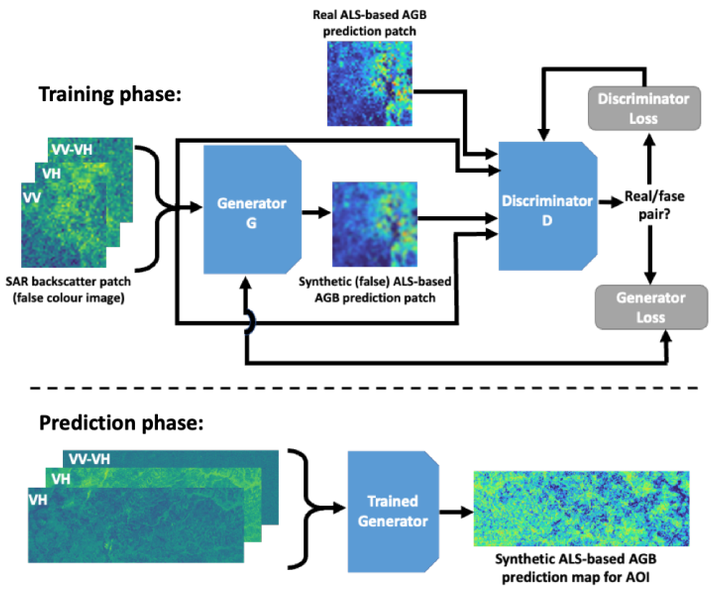


Figure 7.1: Illustration over the proposed cGAN-based sequential modelling approach proposed in Paper I.

for AGB prediction. Although Sentinel-1 data are considered inferior for forest monitoring, empirical results demonstrate the potential of utilising Sentinel-1 data for AGB prediction in Tanzania.

7.1.2 Contributions by the author

- The approach was conceived by me and Prof. Stian N. Anfinsen.
- I was responsible for processing the Sentinel-1 data. I further processed the AGB ground reference data and the ALS-derived AGB prediction maps so that they could be used to train and evaluate the proposed models.
- I made all implementations and conducted all experiments.
- The discussion and analysis were conducted in collaboration with domain experts Prof. Erik Næsset, Prof. Terje Gobakken and Prof. Håkan Olsson.
- I wrote the original draft of the manuscript. The manuscript was further edited in collaboration with Prof. Stian N. Anfinsen, Prof. Erik Næsset and Prof. Terje Gobakken.

7.2 Paper II

Sara Björk, Jonas N. Myhre, and Thomas Haugland Johansen. "**Simpler is Better: Spectral Regularization and Up-sampling Techniques for Variational Autoencoders**", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3778-3782, 2022.

7.2.1 Summary

Many CNN-generated images suffer from varying degrees of blurriness and lack of details, such as sharp transitions from an object to the background. One theory that explains the shortcoming of CNN-based generative models is that these networks tend to prioritise the learning of low-frequency components of image data initially, leading to a bias against learning high-frequency image content such as edges and textures [25,26], see discussion in Section 6.2.

In this paper, we propose a novel frequency-aware objective function to address the shortcomings related to the quality of images generated by convolutional generative models. The frequency-aware objective function denoted $\mathcal{L}_{FFT}$, computes the 2-D Fourier transform of the target and generated image patches and uses the MSE to enforce alignment between these, see Eq. (6.7). As $\mathcal{L}_{FFT}$ is complementary to other objective functions, including it in the training of generative models forces the model to also focus on achieving agreement of the overall spectral content of the data.

The impact of the $\mathcal{L}_{FFT}$ objective function was evaluated by training a generative VAE with a traditional spatial objective function (Vanilla VAE) and comparing it to a VAE that included the $\mathcal{L}_{FFT}$ to the spatial objective function. The performance of the $\mathcal{L}_{FFT}$ objective function was also evaluated against two other recently proposed frequency-aware losses [87,107], see Section 6.2. All experiments were performed on public benchmark datasets such as the CelebA dataset [113]. Empirical results demonstrate that generative VAE models trained with the $\mathcal{L}_{FFT}$ achieve results equal to or better than the current state-of-the-art in frequency-aware losses for generative models. Figure 7.2 illustrates the impact of the $\mathcal{L}_{FFT}$ objective function, where the top row shows images in the spatial domain while the bottom row shows the corresponding Fourier power spectrum of each image. Column (a) shows the target image (top) and its corresponding Fourier power spectrum, the corresponding images reconstructed by use of a Vanilla VAE in column (b) and by use of a VAE with

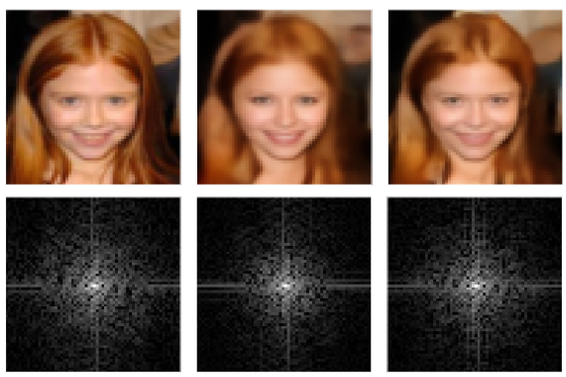


Figure 7.2: Column (a): real image; (b): the image reconstructed with a Vanilla VAE; and (c): the image reconstructed from a VAE optimised by including $\mathcal{L}_{FFT}$.

the $\mathcal{L}_{FFT}$ objective function in (c). Comparing (b) to (c), discrepancies in the highest frequencies of the 2D Fourier spectrum can be seen in (b). Furthermore, when comparing image (b) to image (a) and image (c) to image (a), it is apparent that the Vanilla VAE suffers from a greater lack of details in the spatial representation of the image.

7.2.2 Contributions by the author

- The approach was conceived by me and the co-authors.
- I made most of the implementations and conducted all experiments.
- The discussion and analysis were conducted in collaboration with the co-authors, Prof. Stian N. Anfinsen and Prof. Robert Jenssen.
- I wrote the original draft of the manuscript. The manuscript was further edited in collaboration with the co-authors.

7.3 Paper III

Sara Björk, Stian N. Anfinsen, Michael Kampffmeyer, Erik Næsset, Terje Gobakken, and Lennart Noordermeer. "**Forest Parameter Prediction by Multiobjective Deep Learning of Regression Models Trained With Pseudo-Target Imputation**", submitted to *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

7.3.1 Summary

This paper follows up on the work in Paper I and Paper II by improving the performance of CNN-based regression models for forest parameter prediction. Specifically, this paper proposes a novel methodology that leverages both the limited amount of available ground reference measurements of AGB or SV, and the available ALS-derived prediction maps in the SAR-based prediction. Additionally, taking inspiration from Paper II and the literature on single image super-resolution, see e.g [103, 104], this paper proposes a multiobjective training approach. This approach utilises composite loss functions with varying objectives to train deep CNN-based regression models.

In this paper, we depart from the sequential modelling strategy of Paper I by incorporating the ground reference dataset, interchangeably referred to as true targets, in the training of CNN-based regression models. This is possible by getting inspiration from the related deep learning paradigm, semi-supervised learning with pseudo-labels, see Section 6.3.1 and Section 6.3.2. Thus, for this paper, we propose to train models that benefit from both true targets and ALS-derived prediction maps to improve the CNN model's performance in prediction of forest parameters. The latter dataset is in Paper III referred to as pseudo-targets.

Compared to Paper I, this paper focuses on training CNN-based regression models for AGB prediction in Tanzanian miombo woodlands and for SV prediction in three managed boreal forests in Norway. We follow the training procedure of [104] and divide the training into pretraining and fine-tuning stages. As detailed in Section 5.3.3 and Section 6.2, the CNN-based regression models employ the U-Net architecture, trained to map image patches of Sentinel-1 data into AGB or SV predictions of true targets and pseudo-targets. Two baseline CNN models were trained during pretraining: an $\mathcal{L}_1$-based regression U-Net and a cGAN-based generative U-Net. Subsequently, these baseline models are

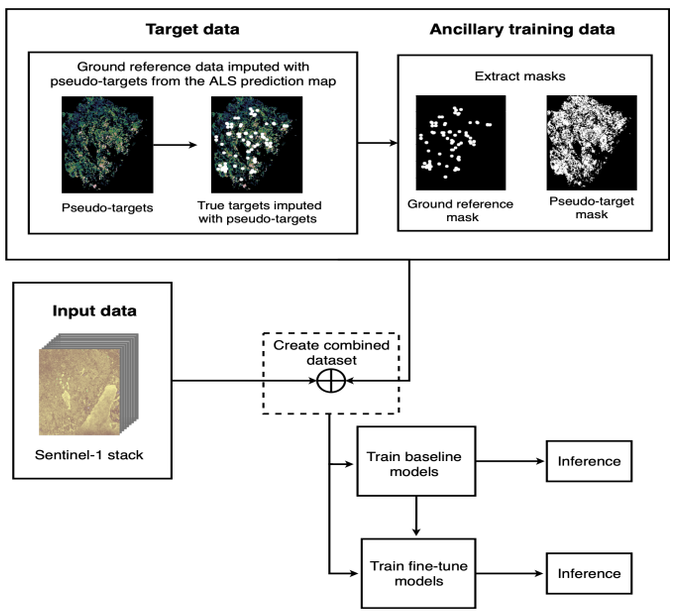


Figure 7.3: The overall workflow, shown for the Tyristrand dataset proposed in Paper III. The workflow includes dataset generation, model training and inference to create prediction maps. Prediction targets, shown as white circles in the figure, have been magnified for illustrative purposes.

further trained using additional loss functions in the fine-tuning stage. Pretraining a model on the $\mathcal{L}_1$ objective function reduces reconstruction errors in terms of MSE, while adversarial pretraining focuses on enhancing the perceptual quality of the generated images [104], which is not important for the forest parameter regression task. Figure 7.3 shows the proposed workflow for Paper III, including dataset generation, model training and inference.

Experiments were conducted on the Tanzanian AGB dataset and the three Norwegian SV datasets, where models were trained and evaluated against the performance of ALS-derived prediction models previously developed for the same regions in [23, 28]. The results demonstrate that models developed with the proposed pseudo-target imputation strategy achieve state-of-the-art performance that surpasses traditional ALS-based regression models. Since the results are consistent for experiments on AGB prediction in Tanzania and SV prediction in Norway, we have shown the robustness of our method for different forest types and the prediction of different forest parameters. Moreover, our results improve on Paper I and further demonstrate the usefulness of CNN-based regression models that utilises freely available C-band Sentinel-1 data.

7.3.2 Contributions by the author

- The approach was conceived by me and Prof. Stian N. Anfinsen. Assoc. Prof. Michael Kampffmeyer provided valuable insight into the domain of semi-supervised learning with pseudo-labels.

- I was responsible for processing the Sentinel-1 data for Tanzania. Additionally, I further processed the AGB and SV ground reference data and the ALS-derived AGB and SV prediction maps so that they could be used to train and evaluate the proposed models.

- I made all implementations and conducted all experiments.

- The discussion and analysis were conducted in collaboration with domain experts Prof. Erik Næsset, Prof. Terje Gobakken and Dr. Lennart Noordermeer.

- I wrote the original draft of the manuscript. The manuscript was further edited in collaboration with the co-authors.

8

Concluding remarks

The aim of this thesis was to develop methodologies for advancing forest parameter retrieval through the use of deep convolutional regression models. In particular, we focused on addressing the following three key challenges that limit the use of CNN-based regression models in forestry: (1) the lack of reference data to use as prediction targets, (2) the diversity in spatial coverage of RS-based prediction maps, and (3) the applicability of forest regression models.

Through the work of this thesis, we have demonstrated that ALS-derived forest prediction maps can serve as a substitute or complement for limited sets of ground reference target data in forest parameter retrieval tasks. In the first scenario, we proposed a two-stage sequential modelling approach for large-scale forest parameter retrieval. In this approach, the second model establishes a relationship between RS data and ALS-derived AGB prediction maps. We developed two types of models for the subsequent stage: a traditional parametric regression model and a convolutional cGAN-based regression model. Both models utilised an accurate wall-to-wall map of ALS-derived AGB predictions as a surrogate for the true prediction targets, and extensive spatial SAR data as regressor data. By employing this sequential modelling approach, we demonstrated how contextual forest regression models can be created without relying on true prediction targets. This contribution is valuable because obtaining ground reference measurements of forest parameters, such as AGB, is often challenging. Moreover, forest ground reference measurements are rarely openly available within forestry.

In the second scenario, we improved the accuracy of deep convolutional regression models by proposing a novel semi-supervised imputation strategy for forest parameter retrieval. In this strategy, we proposed to use ALS-derived forest prediction maps as pseudo-targets. These pseudo-targets were imputed into the sparse dataset that contains the true prediction targets, allowing the contextual CNN-based regression model to leverage both datasets during the learning process. As a result, both scenarios address the first key challenge.

To specifically address the second key challenge, we proposed a method for training deep convolutional regression models that effectively utilise either continuous or partially continuous wall-to-wall maps of ALS-derived forest parameter predictions by employing forest masks. By applying forest masks, the CNN models' loss computation and active training are confined to areas where ALS-derived prediction maps and true prediction targets are available. Combined with the semi-supervised imputation strategy for regression, this contribution can significantly advance the application of deep convolutional regression models in forest parameter retrieval. Moreover, it reduces the requirement for complete wall-to-wall prediction maps, thereby enabling the application of CNN-based regression models in forestry.

We addressed the third key challenge in two ways. Firstly, to enhance the performance of CNN models, we proposed a frequency-aware objective function that complements other commonly used objective functions. This new objective function enforces CNN models to learn both the low-frequency and high-frequency components of image data in generative or I2I translation tasks. As RS image data contain higher frequencies compared to natural images [24], the frequency-aware objective function facilitates better learning from datasets beyond the natural image domain. Secondly, we demonstrated the potential of using C-band SAR data from the Sentinel-1 sensors as input data to enable large-scale and low-cost regression modelling of forests. Despite the general perception of Sentinel-1 data being inferior for forest mapping, primarily due to its limited penetration capabilities in forested areas, we showed that Sentinel-1-based models are a viable alternative for forest parameter retrieval.

8.1 Limitations and outlook

We acknowledge that each research paper has both strengths and limitations. In this section, we discuss the limitations of the papers included in this thesis and suggest potential directions for future research related to deep convolutional regression modelling for forest parameter retrieval.

A general limitation of both Papers I and III is that these rely on the ground

reference measurements of AGB or SV obtained from circular field plots. As both the Sentinel-1 image data and the convolutional regression models proposed in this thesis are represented with, or operate on square pixels, uncertainties arise during model training and evaluation. This is because each circular field plot intersects with multiple square pixels in the RS dataset, leading to inferred uncertainties. One possibility for new inventory campaigns would be to collect ground reference measurements from square field plots, as recommended for many LiDAR-biomass models [15]. However, it is important to note that the field plots used in typical field inventory campaigns often need to be specifically designed to serve as reference data for RS-assisted regression models [19, 28]. As a result, there is a possibility that smaller circular field plots may still be employed due to their practical efficiency [15]. Moreover, the use of square plots does not guarantee a perfect match between the RS data and the field plots in terms of matching grids. This implies that even with the use of square grids, uncertainties may arise when the RS datais resampled to match the grid of the square pixels.

Paper I The methodology proposed in Paper I relies on having access to wall-to-wall maps of ALS-derived prediction maps. However, practical constraints, such as economic limitations or site-specific factors [19, 22, 23], often prevent continuous large-scale ALS mapping of forests. As a result, access to ALS-derived prediction maps for forest parameters is commonly limited. Consequently, the methodology proposed in Paper I primarily applies to smaller sites. Additionally, the regression models proposed in this study were trained without incorporating ground reference measurements of AGB. This omission may partially explain why the proposed cGAN-based models achieved lower prediction accuracy than a conventional ALS-based regression model. While the proposed cGAN-based regression models were evaluated in terms of MAE and RMSE, they were optimised through adversarial training that aims to achieve a high perceptual quality. We hypothesise that the performance of convolutional regression models for forest parameter retrieval could improve by including additional learning objectives that aim to achieve high accuracy in terms of MAE and RMSE, or that also focus on inferring more information from the image data. The latter would, for example, imply that the learning objective enforces the model to regress on both low-frequency and high-frequency content in the image, with high-frequency components representing features like edges or corners in the scene.

Paper II The objective function proposed for Paper II demonstrates that incorporating frequency-aware training enhances the performance of generative models. However, the evaluation of the proposed frequency-aware objective function was limited to generative VAEs using benchmark data from the natural image domain. To truly assess the potential impact of the contribution presented in Paper II, it is necessary to conduct further evaluations of the pro-

posed objective function across other CNN-based models and other real-world datasets. This evaluation could encompass images from outside the natural image domain, such as RS or medical image data.

Paper III This paper addresses the limitations outlined for Papers I and II, leading to the development of CNN-based regression models with improved accuracy. Specifically, for the Tanzanian dataset, these models surpass the performance of a conventional ALS-based regression model previously developed for the same region. Moreover, Paper III demonstrates the versatility of the methodology by successfully handling various forest types and parameters. Despite this, a major drawback of Paper III is the lack of uncertainty estimates for provided AGB or SV predictions. The report on Good Practice Guidance for Land Use, Land-Use Change and Forestry from 2003 [114] and the guidelines from Intergovernmental Panel on Climate Change (IPCC) from 2006 [115], specifically highlight the need for identifying and reporting uncertainties from e.g. AGB estimation from both ground and RS data. Thus, if the aim is to estimate the carbon budget, providing prediction maps for forest parameters is just an intermediate step. While AGB estimates can be retrieved by aggregating individual pixel values from the constructed pixel maps, estimating the uncertainty from DL algorithms and CNN-based regression models is not trivial [17].

8.1.1 Future directions

In this section, we provide our thoughts on some potential research directions for developing DL-based regression models in forestry using remote sensing data.

The first promising research field would be to meet the requirements of IPCC and provide uncertainty estimates for estimated forest parameters. In [116], Abdar *et al.* reviews possible uncertainty quantification methods and their challenges in deep learning, these and the work on uncertainty quantification in forestry by Leonhardt *et al.* [117] could serve as a natural starting points to extend the work of Paper III. Another related research field, referred to as explainable artificial intelligence (XAI) [118,119], focuses on providing explanations on, for example, why a particular prediction was made. Thus, incorporating XAI into the work of Paper III could, for example, provide the model prediction with examples of which features in the dataset that have the greatest importance for accurate forest parameter predictions.

Another promising area of research is transfer learning. As stated in this thesis: retrieving ALS data in large regions is costly. Therefore, it is worth exploring the potential of transfer learning, which refers to utilising a model developed for one region for accurate forest parameter prediction across different regions. How-

ever, for transfer learning to benefit forestry, it is crucial to provide uncertainties and explanations for the model's predictions. Once again, XAI techniques can play a vital role in providing these.

While the proposed methods provide promising results in utilising data from the freely available C-band Sentinel-1 sensor for forest parameter retrieval, both L and P-band microwaves are known to penetrate deeper into the forest [15,32], and thereby providing more sensitivity to forest parameters such as AGB and SV. In the coming years, two new spaceborne SAR missions are planned: the P-band SAR BIOMASS mission, with a scheduled launch in 2024, and the Radar Observing System for Europe in L-band (ROSE-L) mission, which is scheduled for launch in 2028[1]. Both missions are planned to provide global monitoring of, among other things, forest and biomass. Thus, training deep convolutional regression models that utilise RS data from any of these two missions has the potential for improved model performance, especially in dense tropical forests and other forests with high levels of AGB.

Finally, this thesis did not investigate the potential of sensor fusion beyond the combination of SAR data as regressors and ALS-derived prediction maps as pseudo-targets. Sensor fusion could for instance be used to train models that utilise data from several RS sensors as regressors. Another interesting prospect is to train models directly on the ALS data, assuming they are available, instead of using prediction maps resulting from regression models trained on these ALS data. While C-band radar mainly interacts with the crown volume, L- and P-band interact with larger parts of the trees, like large branches and trunks [15]. Thus, by combining data from different SAR sensors, DL-based regression models could potentially learn better features and characteristics from the RS data and thereby achieve higher accuracy in forest parameter retrieval.

1. See *https://www.eoportal.org/satellite-missions/rose-l#space-and-hardware-components* for specifications,

Part III

Included papers

9

Paper I

On the Potential of Sequential and Nonsequential Regression Models for Sentinel-1-Based Biomass Prediction in Tanzanian Miombo Forests

---

Sara Björk, Stian Normann Anfinsen, Erik Næsset, Terje Gobakken, and Eliakimu Zahabu

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4612-4639, 2022

On the Potential of Sequential and Nonsequential Regression Models for Sentinel-1-Based Biomass Prediction in Tanzanian Miombo Forests

Sara Björk, *Associate Member, IEEE*, Stian Normann Anfinsen, *Member, IEEE*, Erik Næsset, Terje Gobakken, and Eliakimu Zahabu

***Abstract*—This study derives regression models for aboveground biomass (AGB) estimation in miombo woodlands of Tanzania that utilize the high availability and low cost of Sentinel-1 data. The limited forest canopy penetration of C-band SAR sensors along with the sparseness of available ground truth restricts their usefulness in traditional AGB regression models. Therefore, we propose to use AGB predictions based on airborne laser scanning (ALS) data as a surrogate response variable for SAR data. This dramatically increases the available training data and opens for flexible regression models that capture fine-scale AGB dynamics. This becomes a sequential modeling approach, where the first regression stage has linked *in situ* data to ALS data and produced the AGB prediction map; we perform the subsequent stage, where this map is related to Sentinel-1 data. We develop a traditional, parametric regression model and alternative nonparametric models for this stage. The latter uses a conditional generative adversarial network (cGAN) to translate Sentinel-1 images into ALS-based AGB prediction maps. The convolution filters in the neural networks make them contextual. We compare the sequential models to traditional, nonsequential regression models, all trained on limited AGB ground reference data. Results show that our newly proposed nonsequential Sentinel-1-based regression model performs better quantitatively than the sequential models, but achieves less sensitivity to fine-scale AGB dynamics. The contextual cGAN-based sequential models best reproduce the distribution of ALS-based AGB predictions. They also reach a lower RMSE against *in situ* AGB data than the parametric sequential model, indicating a potential for further development.**

***Index Terms*—Aboveground biomass (AGB), airborne laser scanning (ALS), conditional adversarial generative network (cGAN), sensor fusion, Sentinel-1, synthetic aperture radar (SAR).**

Manuscript received February 6, 2022; revised May 10, 2022; accepted May 23, 2022. Date of publication June 3, 2022; date of current version June 17, 2022. *(Corresponding author: Sara Björk.)*

Sara Björk is with the Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway, and also with the Applied Deep Learning DevOps Team, KSAT Kongsberg Satellite Services, 9011 Tromsø, Norway (e-mail: sara.bjork@uit.no).

Stian Normann Anfinsen is with the Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway, and also with the Earth Observation Group, Energy and Technology Department, NORCE Norwegian Research Centre, 9019 Tromsø, Norway (e-mail: stia@norceresearch.no).

Erik Næsset and Terje Gobakken are with the Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, 1432 Ås, Norway (e-mail: erik.naesset@nmbu.no; terje.gobakken@nmbu.no).

Eliakimu Zahabu is with the Department of Forest Resources Assessment and Management, Sokoine University of Agriculture, Morogoro 10022, United Republic of Tanzania (e-mail: zahabue@yahoo.com).

Digital Object Identifier 10.1109/JSTARS.2022.3179819

I. Introduction

As a consequence of climate change, there is an increasing need for accurate carbon accounting systems for measuring, reporting, and verification (MRV) on a national level. Through the REDD+ program (officially named "Reducing emissions from deforestation and forest degradation and the role of conservation, sustainable management of forests, and enhancement of forest carbon stocks in developing countries"), developing countries are motivated to implement such an MRV system to monitor the potential reduction of carbon emissions from tropical forests [1]. The documentation of reduced deforestation on a national level could potentially result in a financial reward being released through the program for the countries associated with the REDD+ program [2].

Forests are well known for being one of the major carbon sinks and need to be properly and accurately monitored by the MRV system. This can be achieved by accurately estimating the amount of forest aboveground biomass (AGB), as AGB is a primary variable related to the carbon cycle [3], [4]. To calibrate the MRV system, AGB data over the area of interest (AOI) is needed. It can be collected either through destructive or nondestructive *in situ* sampling. The former implies harvesting, drying, and weighing the plants to estimate the biomass. The latter does not involve harvesting trees but measuring parameters such as tree height and stem diameter. Measured parameters from the nondestructive sampling can be used to predict AGB by allometric models developed for the AOI [4]. Unfortunately, AGB *in situ* measurements of both above categories are costly and time-demanding to collect manually. As a consequence, most research instead focuses on establishing a relationship between a small amount of AGB field data and remote sensing (RS) data using different sensors [2], [5]–[19].

Among different platforms and sensor types, airborne laser scanning (ALS) systems are shown to provide AGB models that are significantly more accurate than models developed using radar or passive optical data [20], [21]. The reason is probably that ALS can provide accurate data describing canopy cover density and canopy height, which is highly correlated with forest AGB [3], [21]. This result was also confirmed in [22], where the ALS-based regression model achieved the highest accuracy of AGB estimates in the miombo woodlands of Tanzania. However, airborne data are associated with high acquisition cost, which

limits the use of ALS data in national MVR systems that require regular acquisitions to keep forest inventories up to date [3], [21].

One of the advantages of employing spaceborne SAR sensors to AGB estimation is that it provides data with extensive spatial coverage that can be acquired with high temporal frequency. SAR data can thus yield frequently updated AGB predictions over large areas. Another advantage is the SAR sensor's ability to penetrate clouds, which makes it effective to monitor regions with a significant amount of cloud coverage. Unfortunately, the use of SAR data for AGB estimation is limited by the saturation level, the property that SAR intensity does not increase with AGB beyond a certain AGB level. This property is dependent on the specific wavelength used by the SAR sensor and implies, in general, that AGB at middle-to-high level cannot be distinguished in the SAR intensity data [3], [23]–[25]. Additionally, SAR data are strongly dependent on the environmental conditions on the ground, where a change in moisture conditions impacts the measured backscatter [23]. The former is a well-known limitation of SAR data that may restrict its use in MRV systems of high precision, and the latter might be circumvented by the use of SAR data acquired at, e.g., dry seasons [24]. The different challenges of SAR and ALS have fostered studies on their combined use for forest AGB estimation. Several of these studies were reviewed in [3] and [21], which conclude that the combination of SAR and ALS may improve AGB estimation, especially when SAR data are used to upscale and extend accurate ALS measurements of forest height to obtain accurate AGB predictions over large areas [3].

Well-known regression models from statistics have traditionally been used to directly relate a small set of ground reference data of AGB to RS data from a single sensor. A popular choice among the conventional regression models is a variation of traditional linear regression: Multiple linear regression and stepwise multiple regression; see, e.g., [11], [14], [15], [17], [19], [26], [27]. The evolution of machine learning (ML) methods has introduced many alternative methods for AGB estimation, with random forests, artificial neural networks (ANNs), and support vector machines for regression as some of the most prominent, see, e.g., [9], [10], [12], [14]–[18], [28]–[31]. Like the traditional statistical regression models, these ML-based models also directly relate ground reference data of AGB to RS data from a single sensor. Due to the limited amount of ground reference AGB data, both traditional statistical regression models and ML-based models are restricted to relate single observations of the ground reference AGB data to single pixels from the RS data source. Thus, the spatial contextual information from neighboring pixels in the RS data source are generally not incorporated in the learning of the regression model. This is likely to inhibit the learning of the AGB dynamics and fine scale variability. The emerging field of deep learning (DL) methods has further opened many new possibilities in the analysis of RS images. Deep neural networks (DNNs) have, among other things, increased the ability to perform accurate regression between different image modalities acquired from different sensors at possibly different times. The combination of multimodal RS images, such as, e.g., SAR and ALS, has been shown to improve AGB estimation results through regression models of increased complexity. Although the different RS images cover the same scene, their pixel measurements represent different domains, like, for example, ALS-derived measurements of heights or SAR-based backscatter intensity data. Transfer learning (TL), domain adaptation (DA) [32]–[34], and image translation [35] are some theoretical frameworks of recent popularity that can be used to handle such challenging and complex problem settings. Also, a challenging regression problem arises when data from different multimodal RS sensors are combined to upscale the extent of an accurate sensor-based AGB prediction map. In the context of such a data fusion task, sequential approaches with two subsequent regression models become relevant as an alternative to the simpler strategy with a single-stage regression model.

In this article, we refer to *sequential modeling* as the process where two regression models are used in a chain to achieve more training data for AGB prediction. Sequential modeling can also be used to upscale the spatial extent of an initial AGB prediction map. In the first stage, one regression model relates ground reference AGB data to a single RS data source with high information content about the target variable, but with limited geographical coverage. The outcome of the first model is an accurate sensor-based AGB prediction map, which is used in the second regression model as a surrogate for ground reference data to regress on data from an additional RS sensor with larger spatial extent. Both traditional regression models, such as simple and multiple linear regression (see, e.g., [36]–[38]), and ML-based models, such as random forest and support vector regression (e.g., [39]–[42]), have previously been applied in a sequential modeling fashion for AGB estimation. In this work, we differentiate between sequential modeling and the traditional approach with a single-stage regression model by referring to the latter as a *nonsequential modeling* approach.

Both sequential and nonsequential regression models for AGB estimation have traditionally operated on an individual pixel level. That is, the prediction at a pixel location is based on regressors exclusively from the same location, without any use of spatial context of neighboring pixels. However, a key feature of DNNs, that partly explains their success in many prediction and regression problems, is their use of convolutional filters. This implies that the prediction of any single pixel is based on regressors from a spatial neighborhood that surrounds it. It also means that the prediction is done by processing blocks of pixels, with image layers of regressor variables in input and a corresponding layer for the response variable in output. This mapping of predictor images to a response variable image is equivalent to the operation known as *image translation* in DL. Isola *et al.* [35] define image translation as follows: *Given sufficient training data, image-to-image translation is defined as the problem of translating one possible representation of a scene into another*. Within DL, the family of generative models is known to enable cross-modal image translation by translating data from one known distribution to another target distribution. Among the generative models are the generative adversarial networks (GANs) [43] particularly popular; see, e.g., [35], [44]–[50]. GANs are trained to capture the data distribution of a target

domain in a minimax optimization procedure. After training, the generator network, G, can be used to map a random noise vector to a target output image. This idea was later extended to the conditional generative adversarial network (cGAN) architecture [51]. In the cGAN setting, the learnt mapping to the target output image distribution is conditioned on the distribution of an input image [35]. Considering the enormous potential of GANs, we wish to address AGB prediction from a DL perspective. However, as a DNN, the cGAN model requires a substantial amount of training data for cross-modal image translation. Therefore, it cannot learn to directly translate between a small set of AGB ground reference data and spatially continuously RS data. Thus, we propose to tackle the regression problem through sequential modeling by applying the cGAN architecture in the second regression model in the sequence. This approach is only possible as we propose to use an AGB prediction map as a surrogate for ground reference data, which makes a large amount of spatially continuous target data available to the regression model. The cGAN's convolutional filters open for the use of spatial contextual information in the predictions. Based on the discussion above, the definition of the research problem in this article is described as follows.

A. Problem Definition

As a developing country and associated with the REDD+ program, Tanzania has the potential to achieve a financial benefit by implementing an MRV system to monitor their forests. Therefore, the primary aim of this work is to develop forest AGB prediction models that could be implemented in an MRV system for Tanzania. For an AGB prediction model to be of practical use in the MRV system of Tanzania, the model should be able to provide frequently updated AGB predictions with extensive spatial coverage, of a high accuracy, and at a low cost. This puts some constraints on the data used.

1) We need to rely on RS data, as large-scale *in situ* sampling will be infeasible.
2) We cannot afford performing frequent ALS campaigns to frequently update a low-cost MRV system.
3) Due to its location, Tanzania experiences rain periods, which constrains the use of passive sensors, as they are not able to penetrate clouds.

The second constraint further limits the use of RS data from sensors that are neither freely available, nor easily accessible. Based on the constraints of this project, we have decided to utilize the Sentinel-1 sensor, as it provides us with freely available and frequently updated data with extensive spatial coverage. However, a simple SAR-based AGB prediction model may limit the precision of the MRV system and consequently the advantage of implementing the system for operational forest monitoring.

Both [3] and [52] advocate the potentials of combining ALS and SAR for large-scale AGB mapping with improved accuracy. Encouraged by this, we restrict the focus of this work to an AOI in the Liwale district in southeast Tanzania. Here, we have access to a small amount of ground reference vector data and continuous raster of ALS data, which has previously been used in combination with four other RS datasets: optical RapidEye and Landsat imagery, interferometric TanDEM-X radar imagery (X-band SAR), and ALOS-PALSAR (hereby PALSAR) radar imagery (L-band SAR), to develop five different traditional nonsequential regression models; see [22]. The ALS-based prediction model of Næsset *et al.* [22] was further used to create a wall-to-wall map of ALS-based forest AGB predictions. Their ground reference dataset and the wall-to-wall map of ALS-based forest AGB predictions were provided to us for this work, and will be used together with Sentinel-1 data to develop low-cost AGB prediction models for the AOI. However, since we aim to contribute with AGB prediction models that can be applied not only in the AOI, but also in extended areas, we put further restrictions on the focus of this work.

1) To develop AGB prediction models of high accuracy and with potentially extensive spatial coverage, we wish to investigate if a sequential modeling approach is better than a traditional nonsequential regression model.
2) By utilizing the wall-to-wall ALS-based AGB prediction map as a surrogate for AGB ground reference data, we are able to implement the second part of the sequential model with a DDN. Thus, in the case of sequential modeling, we additionally investigate the possible benefits of applying a DL-based model instead of a traditional regression model.

Our approach to sequential modeling is to coregister and resample the SAR intensity image data to the same spatial resolution as the available wall-to-wall map of ALS-based AGB predictions, produced with the classical nonsequential regression model presented in [22]. Motivated by the achievements of image-to-image translation, we propose to utilize a cGAN model for the second model in the sequence. We train the cGAN model to synthesize ALS-based AGB maps from false color SAR intensity images. As far as we know, this is the first time contextual DNNs, in the form of cGAN models, have been utilized in a sequential modeling strategy to upscale a limited amount of ground reference data and simulate AGB predictions. We see any modification of the ALS-based regression model as outside the scope of this work. Fig. 1 shows the overall view of the proposed cGAN-based sequential approach used to generate synthetic ALS-based AGB predictions from false color Sentinel-1 image patches. We validate the proposed cGAN-based sequential model against two noncontextual Sentinel-1-based regression models, also proposed for this work: a nonsequential model and a traditional sequential model. The nonsequential regression model relates single pixels of Sentinel-1 data to the small set of AGB ground reference data. For the noncontextual sequential regression model, we trained the second model in the sequence to relate ALS-based AGB predictions to single pixels of Sentinel-1 data. For both noncontextual models, we use the state-of-the-art regression model in the AOI, i.e., a multiple linear regression model with square root transformation of the response variable. This is the same regression model as used by Næsset *et al.* [22].

B. Contribution

To summarize, the contributions of this article are as follows.

1) We extend the work in [22] by developing a similar type of regression model based on Sentinel-1 data.

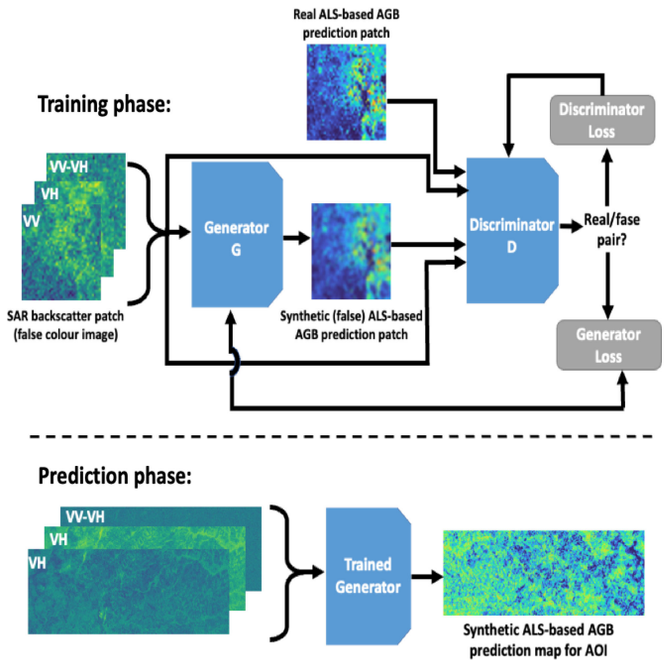


Fig. 1. Flowchart over the proposed cGAN-based sequential modeling approach. The generator network is trained to translate false color Sentinel-1 backscatter patches (consisting of the VV and VH band and their difference, i.e., VV-VH) into realistic-looking synthetic ALS-based AGB prediction patches. The discriminator network is trained to distinguish between a "real" combination of the input patch from Sentinel-1 and the actual AGB prediction patch to a "fake" combination of the input patch from Sentinel-1 and the synthetic AGB prediction patch. The cGAN components, the G network and the D network, are trained in a minimax optimization procedure. After training, the G network can generate realistic-looking synthetic ALS-based AGB prediction patches in an AOI from corresponding false color Sentinel-1 data in the AOI (see prediction phase). Both the individual bands of the false color SAR patch and the ALS-based AGB patches only consist of one channel but are here represented in colors to ease the interpretation.

2) We propose to model forest AGB by a novel sequential modeling approach, in which the second model relates SAR data to ALS-based AGB predictions. We propose two different regression models for the second stage of regression.
   a) One traditional regression model, similar to 1);
   b) one DL-based regression model based on image-to-image translation with a cGAN [35].
3) Since the application of cGANs as AGB regression models is uncommon, we provide a comprehensive study on different hyperparameters, objective functions, and G and D networks.
4) We empirically evaluate the three proposed AGB prediction models against previous results presented in [22] and against each other.
5) We demonstrate the potential of using Sentinel-1 data for AGB predictions and show that our C-band-based models perform better than some of the previously developed models for the AOI.

While we argue for the benefit of using Sentinel-1-based models to extend the spatial coverage of the AGB predictions, the scope for this study is to develop models for the AOI. We therefore see the construction of AGB prediction maps over an extended area as outside the scope of this work.

The remainder of this article is organized as follows. In Section II, we introduce our proposed sequential modeling approach for forest AGB prediction. Section III presents published research in related areas within nonsequential and sequential regression models for AGB prediction through sensor fusion, and related research on image translation through GANs. Section IV presents the datasets, and formally define the proposed nonsequential and sequential regression models. Results are presented and analyzed in Section V, while we discuss our work in Section VI. Finally, Section VII concludes this article. Additional experiments and methodological contributions are collected in the Appendix.

II. Background

In this section, we introduce the proposed sequential modeling approach for forest AGB prediction in both general terms and with a particular emphasize on employing a cGAN for the second part of the sequential model. We continue with a general introduction to the concepts of the cGAN model and how it can be utilized for image-to-image translation in our sequential modeling approach.

A. Non-sequential modeling

As previously introduced, colocated ALS data (y) and AGB ground reference data (z) consisting of 88 field plots were in Næsset *et al.* [22] used to fit a traditional nonsequential regression model $f : y \mapsto z$. The specific regression model from [22], denoted f, uses a square root transformation of the response variable and was trained using ordinary least squares (OLS) regression with stepwise forward selection of the variables. It was used to map spatially continuous ALS measurements into what we refer to as a ALS-based AGB prediction map by

$$\widehat{z}_y = f(y)$$

where $\widehat{z}_y$ denotes each individual ALS-based AGB prediction. The regression coefficients are published in [22] and the resulting prediction map has been made available to us by the authors. The traditional nonsequential approach is illustrated on the left-hand side of Fig. 2, where a single regression model is trained to relate some remotely sensed predictor, such as SAR backscatter intensity (denoted x) or ALS data (y), to a colocated set of sparse AGB ground reference data (z). Here, $\widehat{z}_x$ refers to SAR-based AGB predictions obtained with the traditional nonsequential regression model. The ALS-based biomass prediction map, $\widehat{z}_y$, is of relatively high accuracy compared to maps made from other RS data sources in the same work [22].

B. Sequential modeling

In the modeling strategy with two sequential regression models, we keep the regression model from [22], i.e., f, as the first model in the sequence. We then propose the second regression model in the sequence to relate SAR backscatter intensity data, x, to wall-to-wall maps of ALS-based forest AGB predictions, $\widehat{z}_y$. We thereby utilize $\widehat{z}_y$ as a dense surrogate for z. This gives rise to the second regression model, $g : x \mapsto \widehat{z}_y$, which in the

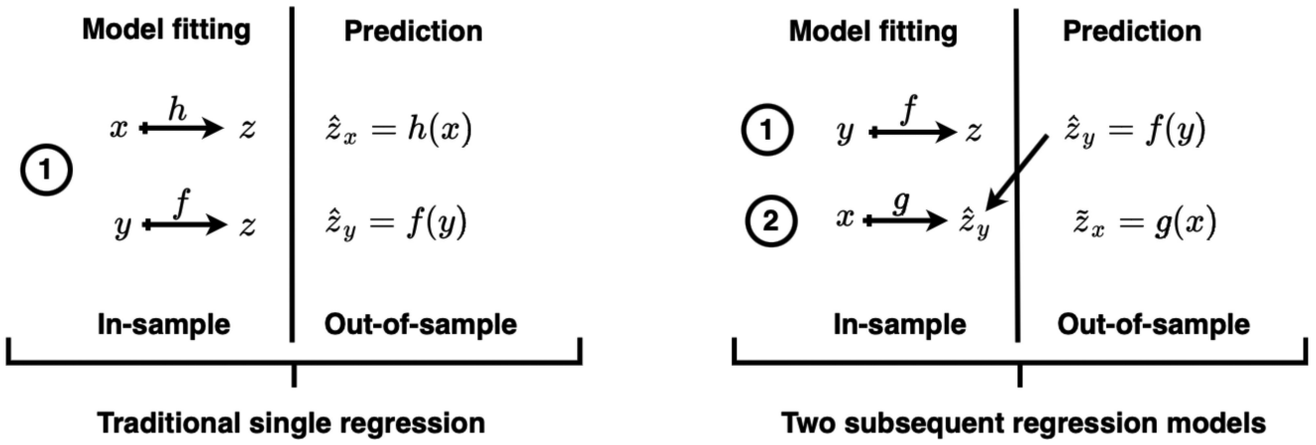


Fig. 2. Illustration of the difference between a traditional nonsequential regression model and the proposed sequential regression models. We let x denote data from a SAR sensor, y denote ALS data, and z denote AGB ground reference data. Regression models are represented by *f*, *g*, and *h*, where *f* is a regression model between y data and z, *h* is a regression model between x data and z, while *g* is a regression model between x data and ALS-based AGB predictions denoted $\widehat{z_y}$. Additionally, $\widehat{z_x}$ denote SAR-based AGB predictions from a traditional nonsequential regression model. In the sequential setting, $\widehat{z_{y|x}}$ denote the outcome from the second part of the two subsequent regression models, i.e., a generated synthetic ALS-based AGB predictions retrieved from x data.

prediction phase can be used to map SAR images, unseen by the model, to generate synthetic ALS-based AGB maps by

$$\widehat{z}_{y|x} = g(x)$$

where $\widehat{z}_{y|x}$ denotes each individual generated synthetic ALS-based AGB prediction. Thus, the two regression models f and g link SAR intensity data to AGB ground reference data in a sequential process. The main benefit of the sequential modeling approach is that the model g can be trained with a large amount of spatially continuous data instead of the few ground reference field plots. Consequently, our sequential modeling approach additionally facilitates for the full exploitation of convolutional DL models for AGB regression as they require access to spatially continuous data. Our proposed sequential modeling approach is shown on the right-hand side of Fig. 2. It should be noted that the described sequential approach is lacking in one respect: The SAR-based prediction, $\widehat{z}_{y|x}$, is regressed against a surrogate regression target $\widehat{z}_y$, which, despite its relatively high accuracy, must necessarily contain some uncertainty. Therefore, the sequential modeling could be followed by a calibration step step where the mean of g is calibrated against the original ground reference data, z. This is discussed in footnote 3.

We propose two different versions for model g: A traditional sequential model and a DL-based sequential model. In the traditional sequential regression setting, we let *g* take the same form as *f*, i.e., a multiple linear regression model with square root transformation of the response variable. In the DL-based sequential regression setting, we instead use a cGAN model as the second regression model. The latter is only possible due to the sequential modeling approach, which allows *g* to be trained on the wall-to-wall map of ALS-based AGB predictions. As the cGAN model utilizes convolutional filtering to exploit the contextual information between neighboring pixels, it carries the potential to capture more information and possibly make better predictions of forest AGB compared to a noncontextual sequential regression model. We let $\widehat{z}_{y|x}$ denote generated synthetic ALS-based AGB predictions from the noncontextual sequential model, while $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ denote generated synthetic ALS-based AGB predictions from the contextual sequential model. The bold font therefore specifies that both the input and the output of g is an image patch (i.e., a subimage from the AOI) and not a single pixel value. For the remaining of this work, we use plain font for variables representing single pixels while a notation in bold font represents a set of pixels.

C. Conditional Generative Adversarial Networks

Cross-modal image translation based on GANs has drawn considerable attention since the architecture was proposed in 2014 [43]. Image translation is achieved through a generative model, referred to as the generator G, that is trained to capture the data distribution of the target domain. Simultaneously, a discriminative model, referred to as the discriminator D, is trained to distinguish between image samples generated by G and images from the actual target domain. The GAN components G and D are trained in a minimax optimization procedure, where they are adapted alternatingly while seeking to optimize conflicting performance criteria. The convergence of both benefits from the battle with the adversary as long as the alternating adaption is appropriately balanced. After training, G can be utilized separately to generate data from the specific distribution.

In the standard GAN setting, the generative model G learns a mapping from a random noise vector to a target output image, while the discriminative model D is trained to distinguish between the generated output image and the corresponding target output image. The whole process, with respect to AGB estimation, is illustrated in Fig. 1 and the upper part of Fig. 3. Here, $\boldsymbol{\beta}$ denotes a random noise vector, the target output image, i.e., ALS-based AGB predictions, is represented by $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$, while the generated synthetic output image is represented by $\widetilde{\boldsymbol{z}}_{\boldsymbol{y}}$. Thus, $\widetilde{\boldsymbol{z}}_{\boldsymbol{y}}$ represent an approximation to $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$, generated from random noise.

In the cGAN setting, the learned mapping to the target output image is conditioned on the distribution of an input image.

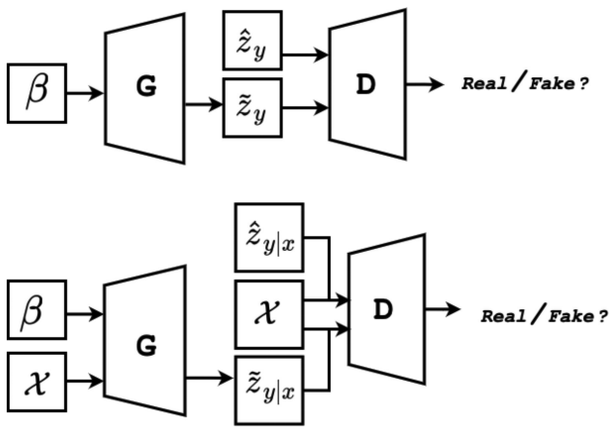


Fig. 3. Illustration of a GAN (upper) and cGAN (lower) model. G and D denote the generator and discriminator networks. $\boldsymbol{x}$ represents images of SAR backscatter intensity from the input domain $\mathcal{X}$. ALS-based AGB predictions in both models are denoted as $\widehat{z_y}$. The subscript y indicates that the AGB predictions are retrieved from a model trained on ALS data. Generated synthetic ALS-based AGB predictions retrieved from $\mathcal{X}$ domain using a cGAN are denoted $\widehat{z_{y|x}}$. $\widetilde{z_y}$ represents generated synthetic ALS-based AGB predictions retrieved from random noise, and $\boldsymbol{\beta}$ in a GAN.

Consequently, the discriminative model, D, instead learns to distinguish between a real pair or false pair of images. The training process of a cGAN, with respect to AGB estimation, is shown in the lower part of Fig. 3. When we let the second part of the sequential model, i.e., g, be represented by a cGAN model, we condition the regression model on a patch of SAR backscatter intensity data, $\boldsymbol{x}$. By the condition on SAR data, the generated synthetic output image of ALS-based AGB predictions is now denoted $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$. In the cGAN setting, the aim of D is to distinguish between $\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}}\}$ and $\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}\}$.

III. Related Work

This section frames our work within related research literature on sensor fusion with a particular emphasis on fusion between ALS and radar, traditional nonsequential regression modeling, sequential regression modeling, and image translation through GANs.

A. Traditional Nonsequential Regression by Sensor Fusion

In this context, we refer to traditional nonsequential regression as the conventional process of relating ground reference data of AGB directly to RS data through a single regression model. This process is illustrated on the left-hand side of Fig. 2. Research on traditional regression models that map SAR backscatter to forest AGB has gained considerable research attention over the years. Two seminal and much-cited works from the year 1992 are the publications of Dobson *et al.* [53] and Le Toan *et al.* [54], which both investigate the dependence between forest AGB and SAR intensity data acquired with different frequencies. Since then, a natural research progression has been to investigate traditional nonsequential regression models by utilizing sensor fusion, i.e., fusion of different RS data sources. Some popular models within traditional regression methods are linear regression, multiple linear regression and stepwise multiple regression [11], [14], [15], [17], [19], [26], [27] for fusion of different radar data sources [27], fusion of radar and optical data [11], [17], [19], [30], or fusion of ALS and optical data [14], [26].

Since [53] and [54] published their classical statistical approaches, the possibilities of using ML and DL models for forest AGB retrieval through sensor fusion have also been investigated widely. Within these fields have fusion of radar and optical data attracted considerable attention [9], [12], [15], [17], [28]–[30], but also fusion of ALS with a multitude of data sources [14], [16], [18], [31] and fusion of different radar data sources [10]. Among the different ML and DL algorithms, random forest-based algorithms are some of the most popular for AGB estimation, see for example [9], [10], [12], [14], [15], [17], [18], [28]–[30], in addition to ANNs (in particular multilayer perceptrons) [12], [16], [18], [28]–[31] and support vector machines for regression [14], [16], [18], [28]–[30]. Research on pure DL methods applied to sensor fusion within traditional nonsequential regression for AGB estimation is still limited. This can probably be explained by the sparsity of ground reference data, which makes it challenging to train DL models. However, one example is found in the work by Zhang *et al.* [14], where ALS data and optical Landsat 8 imagery are integrated to achieve both structural and spectral information predictors for forest AGB estimation. The DL-based model they consider is a stacked sparse autoencoder (SSAE) network, which consists of several sparse autoencoder networks (SAE), each consisting of an encoder and a decoder network. After training each individual SAE, they remove all decoder networks to establish an SSAE by stacking the remaining encoder networks layer-wise. The final SSAE regression network is obtained by adding an unspecified regression model to the end of the SSAE model. While not explicitly mentioned in [14], their SSAE model is a noncontextual model that operates on a single pixel level as it learns to relate RS predictor variables to single AGB measurements, retrieved from a total of 236 field plots. The SSAE network obtains the best performance in comparison with four other traditional regression models and ML models evaluated in [14].

B. Data Fusion With Sequential Regression Models

In this section, we review related research that, like us, applies a modeling strategy with sequential regression models. Characteristic for this review is that it does not focus on the choice of estimation technique. We instead emphasize research on forest AGB estimation through data fusion of different types of RS data sources, which all employs a chain of two models. Common for the research we identified is that the second model exploits predictions from the first model as a dependent variable in the second modeling stage; see right-hand side of Fig. 2. We found that research on AGB estimation applying this particular modeling strategy has been a topic in several studies from year 2008 [55] until today; see, for example, [23], [36]–[42], [56]–[65]. While reviewing earlier research that applies two sequential regression models in their modeling strategy, we noted a variety of terms describing the same concept in the literature. While we choose to refer to this as a sequential regression approach, we additionally found the following use of terminology for

similar, but not necessary identical approaches: *two-step modeling strategy* [40], [57], [65], *two-stage regression* [41], [62], *two-stage up-scaling method* [23], [42], *two-phase estimator* [59], *two-phase (or three-phase) sampling design* [56], [61], *hybrid and hierarchical model-based inference* [60], [64], and *three-phase design* [36]. Additionally, [37]–[39], [55], [58], [63] also apply a modeling approach with two sequential regression models without labeling it by any particular term. Most of the previous research that we identified focuses on relating ground reference data to ALS, and then relates ALS-derived AGB estimates to spaceborne LiDAR data [36], [55], [56], [58], [59], [61] or a combination of different sensors [23], [38], [42], [60], [63]–[65]. Some others relate the ALS-derived AGB estimates to a single sensor, such as Sentinel-2 [39], [41], Landsat [40], [62], GEDI Lidar [65], PALSAR, [57], or SRTM X-band radar [37].

In previous research that adopts a modeling strategy with two sequential regression models, we found traditional regression models to be most common [36]–[38], [55]–[57], [59]–[61], [64], [65], such as, e.g., [38], which focuses on multiple linear regression for upscaling biomass estimates to large areas in the tropical forest of Indonesia. Although Englhart *et al.* [38] included neither ML nor DL, their overall idea has similarities with our modeling strategy. Their work starts by relating collected AGB sample plots to colocated ALS measurements, resulting in a regression model used to predict AGB on the whole ALS dataset. In the final stage, their second regression model relates X- and/or L-band SAR data to ALS-based AGB estimates to extend the AGB estimates to the spatial coverage of the SAR data.

Different ML models have also been applied for AGB estimation that involves data fusion and sequential modeling. As for traditional regression, we find that random forest is one of the most commonly used ML methods, see, e.g., [39]–[42], [63], while, e.g., [23], [63] can be consulted for some additional examples of ML-based methods. In the intersection between traditional regression models and ML models, we also find [58], which applies three different kriging methods [66]: co-kriging, regression kriging, and regression co-kriging, to extend ALS-derived biomass transects to wall-to-wall AGB maps by including L- and C-band data.

Among research that applies a modeling strategy with two sequential regression models, we notice an absence of research using DL models for the regression task. Only one study was identified [63], which similarly to [14] employs an SSAE for the regression task.[1] While [63], like us, uses a sequential modeling approach to establish a relationship between ALS-derived forest biomass predictions and satellite predictors from, e.g., Sentinel-1 data, there are some distinct differences. Although Shao *et al.* consider some contextual predictor variables, their SSAE model is a noncontextual model that only considers single pixels in the training and prediction phase. A novelty of this work is that the cGAN model lets us exploit the contextual information between neighboring pixels through its convolutional filters. Second, [63] adds a nonspecified regression model to the end of the trained SSAE network to perform AGB predictions, as does [14]. In our case, the cGAN model is in itself the regression model and there is no need for additional models to accomplish AGB predictions. Thus, by letting one of our proposed sequential models employ a cGAN model, we contribute with new insight on how DL and RS data can be combined for AGB prediction.

[1] See Section III-A and [14] for a discussion on the SSAE.

C. Image Translation With Generative Adversarial Networks

Image-to-image translation is the task of translating a representation of the imaged scene into another. Examples of this process can, for example, be to translate from a grayscale representation into an RGB image or translating an aerial photo into a map view of the same area [35]. In such a translation process, the G network is commonly conditioned on the first representation, i.e., the input signal or distribution, to achieve better translation. This makes the cGAN and the *Pix2pix* architecture [35], as one specific example, better suited for this task than a generator network conditioned on noise, as the traditional GAN [67]. In this work, we choose to condition the G network on SAR measurements of the backscatter coefficient in the same area, from which we wish to generate ALS-based biomass prediction maps.

Research on RS data simulation through image translation can be found in, e.g., [48], [50], [67]. Li *et al.* [50] focus on change detection (CD) and propose a GAN-based deep translation network for translation between SAR and optical images. By translating images from both sensors into a common feature domain, image characteristics from both images become comparable and can aid the network in the CD task. Ao *et al.* [67] proposed a framework for translation between different SAR sensors. By conditioning their dialectical GAN on urban input images from the low-resolution (LR) Sentinel-1 sensor, they enable generation of corresponding high-resolution TerraSAR-X images. The dialectical GAN uses a modification of the *Pix2pix* cGAN proposed in [35] and combines concepts of both the cGAN and traditional neural networks. Bao *et al.* [48] consider three nonconditional GAN networks to simulate SAR data of vehicles from random noise. While [50] focuses on translating between instruments with different physical measurement principles, does neither of [48], [50] focus on using image translation through GANs for regression purposes as we intend to.

In general, most of today's research on semisupervised learning through GANs focuses on solving a classification task; see e.g., [49], which propose the DLR-GAN architecture to perform LR image classification. To improve classification on this challenging task, they propose to let the G network learn to recover the LR components and the high-frequency components of the LR image. Only a very very few studies were identified that apply their architecture to regression tasks [68]. Within the GAN literature, Rezagholizadeh and Haidar [68] presented one of the first models aimed at regression, named the Reg-GAN. They use two different networks, where one learns data generation while the other predicts continuous labels. It is applied in a computer vision task for self-driving vehicles, where the GAN generates images of a road segment and a regression network predicts the

matching steering angle. Olmschenk *et al.* [69] later proposed the feature contrasting loss function and outperformed [68] on the same semisupervised GAN regression task. Additional examples were also shown in [69] on the combined task of face generation and age prediction as well as on crowd counting. The proposed work in our article differs from earlier related research [68], [69], as we do not perform any additional regression on the image content of the generated synthetic patches. This is possible due to the nature of our proposed modeling strategy with two sequential regression models, which results in a cGAN-based model that is able to make predictions in new unseen areas through the image translation.

IV. Material and Methodology

The related work presented in Section III positions our work with respect to published research in related areas. Based on this literature survey, and previous published research on AGB estimation in the AOI, we make the following methodological contributions.

1) By proposing our Sentinel-1-based nonsequential AGB regression model, we extend the work of Næsset *et al.* [22].
2) The two proposed sequential models extend previous work on sensor fusion in the AOI. Furthermore, by introducing the DL-based sequential model, this work also contributes with novel insight on the possibilities for AGB prediction by using DL models for sensor fusion. These DDNs have convolutional layers that extract contextual spatial information, which has been exploited both in other types of regression problems [70] and also for AGB prediction [71], but not in a sequential regression approach to upscaling and information enhancement.
3) The proposed method applies image translation to truly heterogeneous images and domains in a regression context. Similar image translation has previously been done for general purposes [72] and within image analysis tasks like change detection [73], but is new in the biomass estimation and regression setting.

We accomplish the mentioned novelties in 2) and 3) for the DL-based sequential model by using a modification of the Pix2Pix image translation architecture [35] to generate synthetic yet realistic ALS-based AGB predicted maps with SAR intensity data as input. We refer to the Appendix, i.e., Section A1, for a list over these modifications and their motivation.

We will in the following describe the datasets used in this article, the preprocessing steps applied to the data, and give an overview of the different models we consider.

A. Study Area and Dataset Description

1) Study Area: The AOI is a rectangular area with size 11.25×32.50 km (WGS 84/UTM zone 36S), located in the Liwale district in southeast Tanzania (9°52′–9°58'S, 38°19′–38°36′E). Fig. 4 shows the relative location of the AOI in Tanzania. The Liwale district experiences two rain periods each year: A shorter period from late November to January and a longer period from March to May. Liwale's main dry season occurs between July and October. The miombo woodlands of the Liwale districts is characterized by a large diversity of tree species, with *Brachystegia* sp., *Julbernadia* sp., and *Pterocarpus angolensis* being the most dominant ones [2], [7], [22].

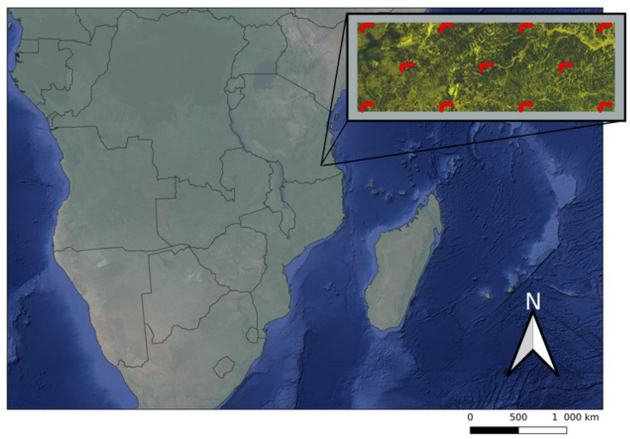


Fig. 4. Location of a subset of the Sentinel-1 scene, as well as the location of the ground reference plots (in red) in the country of Tanzania.

2) Field Data: The field data used in this work, from now on referred to as AGB ground reference data or z, were collected within 88 field plots during January–February 2014 [22]. These field plots were distributed in groups of eight in each of the 11 L-shaped clusters, shown with red dots in the Sentinel-1 scene in Fig. 4. The sample plots are circular, each of size 707m^2, i.e., they have a radius of 15 m. We refer to [74] for a thorough work on the national level sampling design for Tanzania, and to, e.g., [2], [7], [22] for reference work on, e.g., the use of field data in the AOI for large-scale AGB estimation. Measured AGB in the AOI ranged from 0 to 213.4 $\mathrm{Mg\,ha^{-1}}$ [22].

3) ALS-Based AGB Data: The ALS data were acquired in 2014; see [7], [22] for details of this process. Næsset *et al.* [22] trained a regression model on the ALS data to make ALS-based AGB predictions on a grid with square pixels of size 707 m^2. Their model, referred to as f, is the first regression model in our proposed modeling strategy with two sequential regression models. The output from the ALS-based regression model in [22], i.e., ALS-based AGB predictions, $\widehat{z}_y$, was made available for this work. These ALS-based AGB predictions will serve as a surrogate for the AGB ground reference data in the second regression model g, when SAR data is used with either a traditional regression model or a cGAN model for image translation to upscale the ALS-based AGB predictions. See right-hand side of Fig. 2 for an illustration of the sequential modeling strategy with notation.

4) SAR Data: Our SAR data consists of a C-band SAR scene obtained from the Sentinel-1sensor, which provides data in two bands, i.e., the VV and VH polarization. This sensor was chosen since an AGB model trained on data from this sensor meets most of the needs listed in Section I-A; the data is frequently updated, it has extensive spatial coverage, and is freely available. For this article, we choose a Sentinel-1scene acquired on 15 September 2015, as it fulfils three additional criteria: 1) It covers our AOI, 2) it is closest in time to acquisition of the ALS data,

and 3) it was acquired during one of the area's two yearly dry seasons. The latter implies that the scene achieves optimal sensitivity to dynamic AGB levels. We initially aimed to create a multitemporal stack of Sentinel-1scenes, but as only one scene meets all the three additional criteria, we had to settle for this single scene. The SAR data are obtained in a high-resolution Level-1 ground range detected (GRD) format, with a pixel size of 10 m. It was downloaded from Copernicus Sentinel Scientific Data Hub.[2] Fig. 4 visualizes the scene and indicates its relative location in Tanzania.

B. SAR Data Processing and Preparation of Datasets

To process the Sentinel-1 GRD product, we used the ESA SNAP toolbox [75] and followed the workflow suggested in [76] with some modifications. The final processing workflow is summarized as follows:

1) apply orbit file;
2) thermal noise correction;
3) border noise removal;
4) calibration;
5) range Doppler terrain correction (bilinear interpolation);
6) (conversion to dB).

We also experimented with speckle filtering, using a refined Lee filter [77] with the SNAP default window size of 7×7 as an optional additional processing step between step 4) and step 5). However, since models trained on speckle filtered Sentinel-1 data experience higher variations in AGB predictions than models trained without speckle filtered Sentinel-1 data, we decided to omit speckle filtering in the processing workflow. See Section A2 in the Appendix for details. Step 6) was only applied to the cGAN-based sequential regression model. We provide an investigation of the impact that Sentinel-1 data on dB scale or linear scale have on AGB predictions for cGAN-based models in the Appendix, see Section A5. During step 6) for the data used in the cGAN-based regression model or after step 5) for the two other models, we also applied the same map projection as in [22], i.e., WGS 84/UTM zone 36S, to make sure that the Sentinel-1 dataset and the ALS-based AGB prediction dataset are aligned.

After performing the above processing steps, our Sentinel-1 dataset was further processed in QGIS [78]. In QGIS, we first reprojected the Sentinel-1 dataset to the same projection that the ALS-based AGB grid pixel dataset used in [22]. Then, cubic convolution resampling was applied to resample the pixel size of the Sentinel-1 dataset from its original pixel spacing of $10\,\mathrm{m} \times 10$ m to the same pixel size as the grid pixels of the ALS-based AGB predictions, i.e., 26.6 m $\times$ 26.6 m. As a final step, a subset of the entire Sentinel-1 scene corresponding to the extension of the ALS-based AGB data was extracted.

For the image-to-image translation task, i.e., the cGAN-based model g, a false-color image was created from the processed Sentinel-1 dataset. This was done since the chosen cGAN architecture, Pix2Pix, requires three-channel RGB images or grayscale images as input. The false-color image was created as follows: red = VV, green = VH, and blue = VV-VH. The ALS-based AGB prediction dataset was kept as a grayscale image as each grid pixel in the dataset only consists of one feature, i.e., an AGB prediction. Fig. 5 shows the ALS-based AGB prediction dataset and the corresponding false-color Sentinel-1 scene after performing all processing steps with the ESA SNAP toolbox and QGIS. For illustrative purposes, we choose to show the ALS-based AGB prediction dataset of Fig. 5 in pseudo-colors, where dark blue pixels indicate biomass closer to $0\ \mathrm{Mg\,ha^{-1}}$ while green through yellow to red pixels indicate increasing biomass content ($\mathrm{Mg\,ha^{-1}}$).

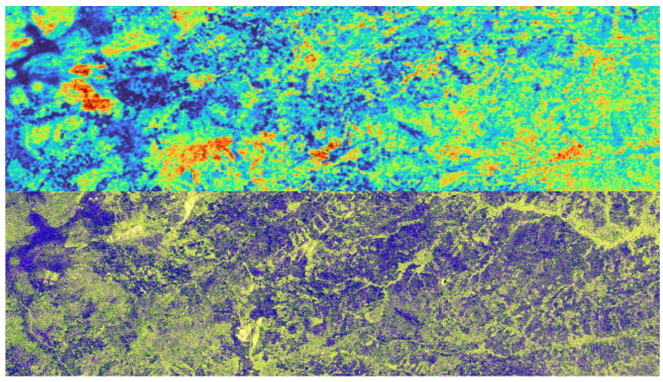


Fig. 5. Top row: ALS-based AGB predictions from [22]. Bottom row: False-color image of the Sentinel-1 dataset.

C. Traditional Sentinel-1A-Based AGB Regression Models

In [22], several models were explored to construct traditional nonsequential regression models for AGB relating different remotely sensed datasets and the 88 field plots. They settled for a model with square root transformation of the response variable for ALS, RapidEye, Landsat, and PALSAR, since this model performed equally well as more complex models and since it always predicts values > 0. Inspired by their findings, we develop a similar baseline nonsequential regression model for AGB between Sentinel-1 and the same 88 field plots according to

$$E\left[\sqrt{AGB}\right] = \alpha_0 + \Sigma_{j=1}^{J} \alpha_j x_j \tag{1}$$

where α_0 is the intercept, i.e., a constant, α_j are regression coefficients, and x_j are explanatory variables. We followed the procedure in [22] and performed OLSs regression with stepwise forward selection of the variables. Our inclusion criteria focus on variables being significant at 5% level using an F-test. For the Sentinel-1 product, VH and VV backscatter coefficients on a linear scale plus square and square root transformations of these variables were subject to the stepwise selection. We follow the procedure from [22] and correct for bias when transforming our model to arithmetic scale in accordance with [79]

$$\widehat{AGB} = \left(\hat{\alpha}_0 + \Sigma_{j=1}^{J} \hat{\alpha}_j x_j\right)^2 + MSE \tag{2}$$

where MSE is the mean square error computed from the fitted model on square root form, i.e., from 1.

[2] See [Online]. Available: https://scihub.copernicus.eu/dhus/#/home

TABLE I
SYMBOLS AND NOTATION INTRODUCED IN SECTION I-A AND USED THROUGHOUT THE ARTICLE FOR THE DIFFERENT DATASETS, IN NONSEQUENTIAL MODELING, SEQUENTIAL MODELING, THE GAN, AND THE CGAN MODEL

$\boldsymbol{\beta}$	Noise vector, input to the G of a GAN/cGAN
D	Discriminator network of a GAN/cGAN
G	Generator network of a GAN/cGAN
$\mathcal{X}$	Represent the input domain, SAR data
$\mathcal{Y}$	Represent the domain of ALS data
$\mathcal{Z}$	Represent the domain of AGB data
$\widehat{z}_x$	SAR-based AGB predictions, $\widehat{z}_x \in \mathcal{Z}$
$\widetilde{\boldsymbol{z}}_{\boldsymbol{y}}$	A patch of generated synthetic ALS-based AGB predictions from a GAN. Retrieved from $\boldsymbol{\beta}$ data, $\widetilde{\boldsymbol{z}}_{\boldsymbol{y}} \in \mathcal{Z}$
$\widehat{z}_y, \widehat{\boldsymbol{z}}_{\boldsymbol{y}}$	ALS-based AGB predictions, $\widehat{z}_y, \widehat{\boldsymbol{z}}_{\boldsymbol{y}} \in \mathcal{Z}$
$\widehat{z}_{y\|x}$	Generated synthetic ALS-based AGB predictions from the baseline sequential regression model, trained with x data (SAR data) as the regressor. $\widehat{z}_{y\|x} \in \mathcal{Z}$
$\widehat{\boldsymbol{z}}_{\boldsymbol{y}\|\boldsymbol{x}}$	A patch of generated synthetic ALS-based AGB predictions from a cGAN. Retrieved from $\boldsymbol{x}$ data (SAR data), $\widehat{\boldsymbol{z}}_{\boldsymbol{y}\|\boldsymbol{x}} \in \mathcal{Z}$
f	A regression function between y data and z
g	A regression function between x data and $\widehat{z}_y$
h	A regression function between x data and z
$x, \boldsymbol{x}$	Data from the SAR sensor, i.e. $x \in \mathcal{X}$
$y, \boldsymbol{y}$	Data from the ALS sensor, $y \in \mathcal{Y}$
z	Ground reference AGB data, $z \in \mathcal{Z}$

Notation in plain font indicates variables represented by single pixels, while notation in bold font indicates variables represented by image patches consisting of a pixel neighborhood.

D. cGAN-Based AGB Regression Models

This section formally introduces some popular choices of objective functions, the generator network, and the discriminator network of a cGAN, with a special focus on the Pix2Pix architecture [35]. We also relate the cGAN framework to model g in our sequential modeling strategy by using the same notation that was introduced in Section I-A. See Table I for a summary of the notation, and Figs. 2 and 3 for illustrations of how the different entities of Table I are used in the sequential modeling approach or in the cGAN network.

In our application, the input domain consists of image patches from the Sentinel-1 scene, and the output domain of corresponding image patches from the ALS-based AGB wall-to-wall map. Thus, conditioned on images from the input domain, $\boldsymbol{x} \in \mathcal{X}$, the generator network G of the cGAN aims to capture the data distribution of the output domain $\widehat{\boldsymbol{z}}_{\boldsymbol{y}} \in \mathcal{Z}$, by generating corresponding synthetic image samples $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}} \in \mathcal{Z}$. Image pairs are then presented to the discriminator network D of the cGAN, which aims to distinguish if it is presented with a real pair of images, $\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}}\}$, or a fake pair, $\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}\}$. The whole training process of a cGAN is illustrated in the lower part of Fig. 3. As G aims to fool D, its ultimate goal is to obtain $\widehat{z}_{y|x} \approx \widehat{z}_y \approx z$, where $\widehat{z}_{y|x}, \widehat{z}_y, z \in \mathcal{Z}$. In other words, at the position of each single AGB ground reference measurement, the generated synthetic ALS-based AGB predictions should resemble both z and the ALS-based AGB predictions well on a pixel basis. During adaption of the cGAN, both G and D are trained simultaneously to outperform each other, resulting in the following minimax objective function [43]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x},\widehat{\boldsymbol{z}}_{\boldsymbol{y}}}[\log D(\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}})] + \mathbb{E}_{\boldsymbol{x}}[\log(1 - D(\boldsymbol{x}, G(\boldsymbol{x}))]. \quad (3)$$

A cGAN network trained with the objective function in 3 is referred to as a Vanilla GAN. The least squares generative adversarial network (LSGAN) was proposed to overcome issues with stability during training of the Vanilla GAN [80]. Its objective functions in a conditional setting are

$$\min_D V_{\text{LSGAN}}(D) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x},\widehat{\boldsymbol{z}}_{\boldsymbol{y}}}\left[(D(\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}}) - b)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{x}}\left[(D(\boldsymbol{x}, G(\boldsymbol{x})) - a)^2\right]$$
$$\min_G V_{\text{LSGAN}}(G) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x}}\left[(D(\boldsymbol{x}, G(\boldsymbol{x})) - c)^2\right] \quad (4)$$

where a and b are labels for fake and real data, while c denotes a value that G tricks D to believe for fake data [80]. Introduced in [81] for further stabilization of training and high-quality image generation, we also consider the Wasserstein GAN with gradient penalty (WGAN-GP). It considers real data, simulated data, and a combination of these in its objective function, which in the conditional setting has the following form [81]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}}[D(\boldsymbol{x}, G(\boldsymbol{x}))] - \mathbb{E}_{\boldsymbol{x},\widehat{\boldsymbol{z}}_{\boldsymbol{y}}}[D(\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}})] + \lambda\mathbb{E}_{\hat{\boldsymbol{z}}}\left[(||\nabla_{\hat{\boldsymbol{z}}} D(\hat{\boldsymbol{z}})||_2 - 1)^2\right] \quad (5)$$

with

$$\hat{\boldsymbol{z}} = \epsilon\widehat{\boldsymbol{z}}_{\boldsymbol{y}} + (1-\epsilon)G(\boldsymbol{x})\,. \quad (6)$$

$\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ in 3, 4, and 5 denotes a real ALS-based AGB image patch from the $\mathcal{Z}$ domain while $G(\boldsymbol{x}) = \widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ represents a generated synthetic image patch.

1) Generator Network: Three different G networks were tested, all based on the ResNet model [82]: ResNet-4, ResNet-5, and ResNet-6. ResNet-6 is a part of the original Pix2Pix architecture [35] and consists of two encoder blocks followed by six residual blocks and two decoder blocks. ResNet-4 and ResNet-5 consist of the same number of encoder–decoder blocks as ResNet-6, but only 4 and 5 residual blocks, respectively. The two smaller networks were proposed as we work with small image patches of 64×64 pixels; see Section V-B2.

2) Discriminator Network: Isola *et al.* [35] evaluate different variations of the neural network discriminator architecture by varying the patch size N of the discriminator receptive fields from a 1×1 *PixelGAN* to an $N \times N$ *PatchGAN*. Since we work with fairly small image patches in number of pixels, we decide to settle for the following three discriminator networks:

- a 1×1 *PixelGAN*;
- a 16×16 *PatchGAN*;
- a 34×34 *PatchGAN*.

The two *PatchGAN* networks were designed by adjusting the depth of the GAN discriminator to obtain a receptive field of 16×16 or 34×34, respectively. In a *PixelGAN*, the discriminator tries to classify each 1×1 pixel in the image patch as real or fake, while for the two *PatchGAN* networks, the discriminator tries to classify each $N \times N$ patch of pixels in the image patch as real or fake. The discriminator network is applied across an image patch in a convolutional matter during the discriminator

TABLE II
PEARSON CORRELATION COEFFICIENT, R, RMSE, LEAVE-ONE-OUT CROSS-VALIDATION RMSE (LOOCV RMSE), AND MEAN ABSOLUTE ERROR (MAE) COMPUTED BETWEEN GROUND REFERENCE PLOTS OF AGB, z, AND AREA-WEIGHTED MEANS OF PREDICTED AGB FROM EITHER THE FIVE NONSEQUENTIAL REGRESSION MODELS [22] OR OUR SENTINEL-1-BASED NONSEQUENTIAL REGRESSION MODEL

Auxiliary data source	Modelling approach	Model	R	RMSE	LOOCV RMSE	MAE
ALS[a]	Non-sequential (traditional)	[c]	0.68	33.39	[c]	24.61
InSAR[a]	Non-sequential (traditional)	[c]	0.49	40.40	[c]	29.44
RapidEye[a]	Non-sequential (traditional)	[c]	0.61	36.21	[c]	26.76
Landsat[a]	Non-sequential (traditional)	[c]	0.33	43.03	[c]	33.10
PALSAR[a]	Non-sequential (traditional)	[c]	0.27	43.96	[c]	33.18
Sentinel-1[b]	Non-sequential (traditional)	$\widehat{AGB} = \left(2.96 + 41.60VV\right)^2 + 10.51$	0.54	38.52	39.6	30.04

All units are in $\mathrm{Mg\,ha^{-1}}$.
[a]Indication of which remote -sensed data source that were used in [22] to train traditional their non-s equential regression models.
[b]The traditional non-sequential regression model developed between Sentinel-1 and AGB reference data.
[c]See [22] for reference to specific models and computed LOOCV RMSE.
.

phase to produce several classification responses. Eventually, all responses are averaged to provide the discriminator output with a real or false decision. Thus, for each image patch pair,$\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}}\}$ or $\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}\}$, D outputs a binary prediction, based on D's belief of the input pair. Optimally, we wish D to predict a fake pair when the image par consists of an image patch from $\boldsymbol{x}$ and another from $G(\boldsymbol{x})$, i.e., $\{\boldsymbol{x}, \widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}\}$.

V. EXPERIMENTS AND RESULTS

In this section, the proposed Sentinel-1-based regression models for AGB prediction are presented: The nonsequential regression model, the baseline sequential regression model, and the cGAN-based sequential regression model. The performance of the proposed models is evaluated by comparing predicted AGB to AGB ground reference data and the constructed AGB prediction maps to each other, and the AGB prediction maps of [22]. Qualitative and quantitative results are provided. We keep the notation introduced in Table I and let z denote ground reference AGB data, $\widehat{z}_x$ denotes AGB predictions from the Sentinel-1-based nonsequential regression model, $\widehat{z}_y$ denotes AGB predictions from the nonsequential ALS-based model [22], and $\widehat{z}_{y|x}$ denotes either generated synthetic ALS-based AGB predictions from the baseline sequential regression model or single predictions from the cGAN-based sequential regression model. In contrast, $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ denotes a patch of predictions from the cGAN-based sequential regression model. We refer to the Sentinel-1-based nonsequential regression model as h, the ALS-based nonsequential regression model from [22] as f, and either of the sequential models, i.e., the baseline traditional sequential regression model or the cGAN-based sequential regression model, as g.

A. A Traditional Nonsequential Regression Model for AGB

We extend the work of [22] by developing a traditional nonsequential regression model, h, for the 88 field plots of AGB ground reference data (z) according to (2). To do so, we laid the circular field plots of z on top of the Sentinel-1 pixel grid. VH and VV backscatter values corresponding to z were found by computing the area-weighted mean of Sentinel-1 pixels intersecting the field plots. Only one explanatory variable, i.e., VV, was selected in the stepwise forward selection procedure. The achieved model, h, for AGB per hectare, is listed in Table II. Since the model was fitted on the whole ground reference dataset z, we follow [22] and perform additional quantitative model assessment analysis through leave-one-out

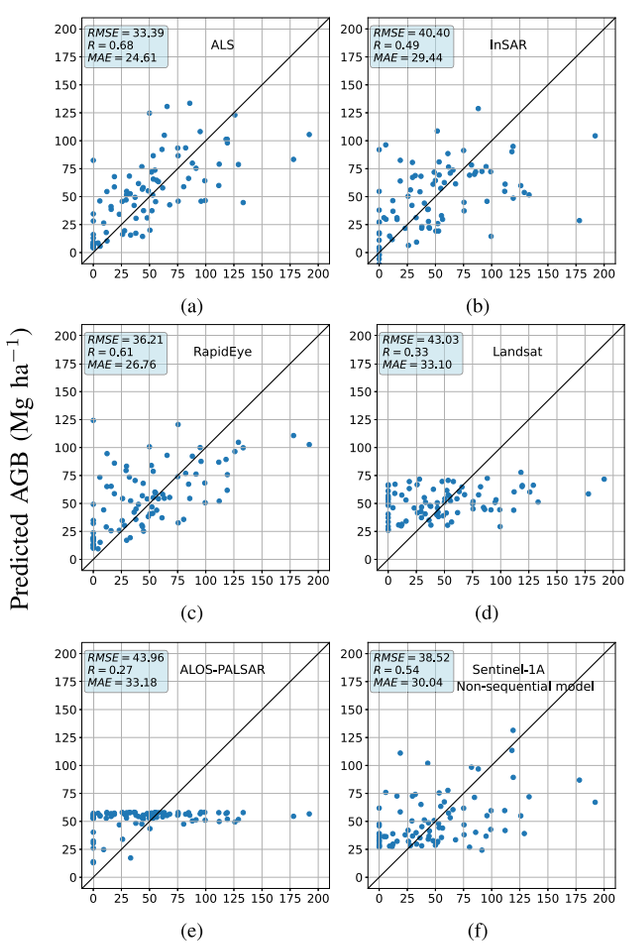


Fig. 6. Scatter plots between ground reference AGB, z, and model-predicted AGB. Model-predicted AGB is retrieved from either (a) the ALS, (b) InSAR, (c) RapidEye, (d) Landsat, (e) PALSAR, or (f) our proposed Sentinel-1-based nonsequential regression model. The black lines are reference lines indicating 100% correlation between z and predictions. Units are in Mg ha^{-1}

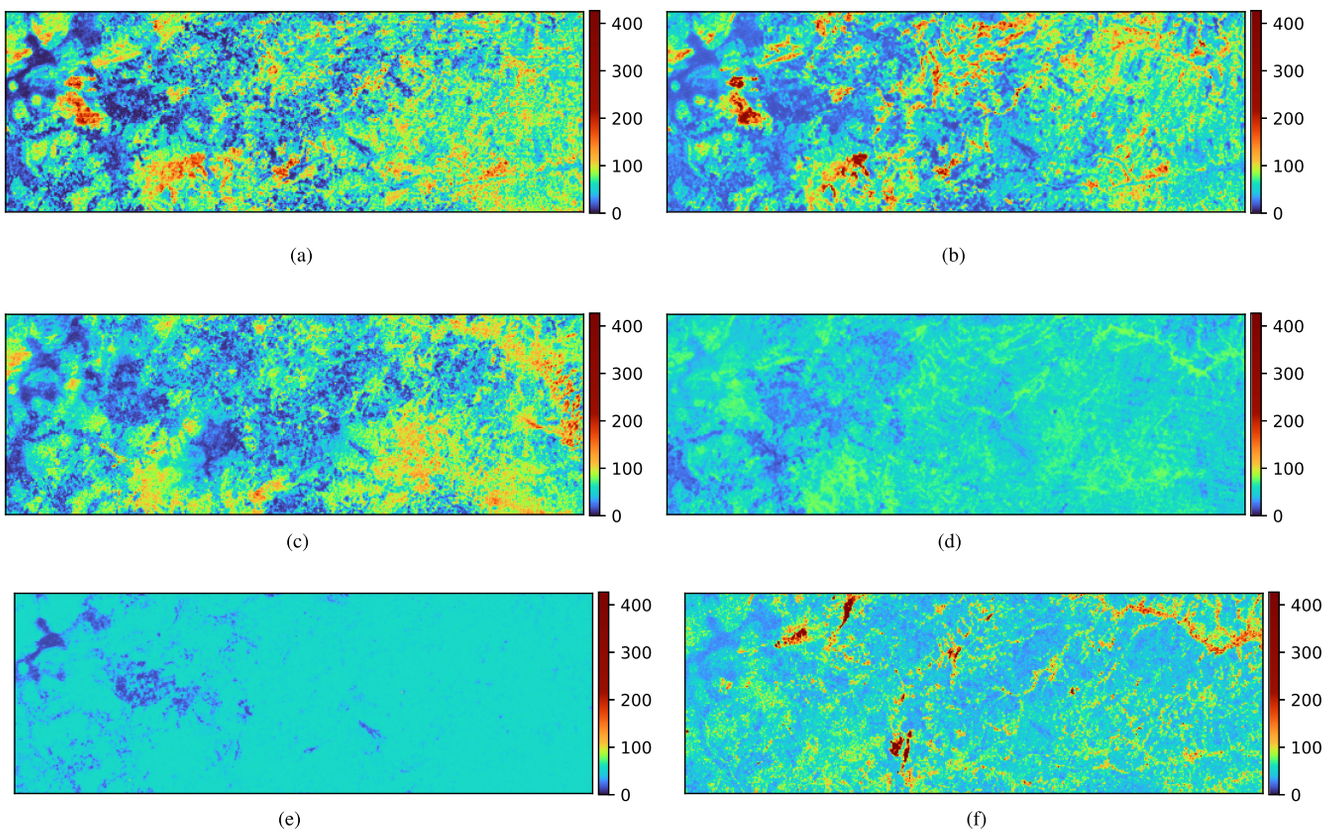


Fig. 7. Aboveground biomass prediction maps (in $\mathrm{Mg\,ha^{-1}}$). (a)–(e) Results of the traditional nonsequential regression models presented in [22]. The AGB biomass map in (f) was constructed from the proposed nonsequential Sentinel-1-based AGB model in Table II. (a) ALS. (b) InSAR. (c) RapidEye. (d) Landsat. (e) PALSAR. (f) Non-sequental Sentinel-1.

cross-validation (LOOCV) to compare the consistency of predicted AGB. We also compute the Pearson correlation coefficient (R), root-mean-squared error (RMSE), and mean absolute error (MAE) between model predicted AGB and z. These metrics are collected in Table II together with computed R and RMSE from the nonsequential regression model developed in [22]. Additionally, we qualitatively assessed our model against those developed in [22] by plotting model-predicted AGB against z in Fig. 6 and by illustrating model-derived AGB wall-to-wall maps in Fig. 7. Minor differences between the scatter plots in Fig. 6(a)–(e) and data reported in Table II, compared to the corresponding figures and table in [22], can be explained by differing pixel grids used in the area-weighting of RS pixel values. Næsset *et al.* [22] developed their traditional nonsequential regression models for InSAR, RapidEye, Landsat, and PALSAR by using the original pixel grid of the satellite data. When reporting metrics, they further used each sensor's original pixel grid to compute the area-weighted average of pixel values within the coverage of each field plot. After preprocessing the Sentinel-1 scene, both the Sentinel-1 dataset and the ALS-based AGB predictions are on the same grid with pixel size 707 m^2, representing an area of 26.6 m $\times$ 26.6 m on the ground. In this work, we did not have access to the original pixel grids of the ALS, InSAR, RapidEye, Landsat, and PALSAR data. Therefore, we chose to use the grid with pixel size 707 m^2 for also these models whenever area-weighted metrics were computed. The resulting differences to [22] must therefore be endured.

We observe from Table II that only two of the previously developed models in [22], i.e., the ALS-based (f) and the RapidEye-based models, experience lower RMSE and a higher Pearson correlation coefficient with respect to z than our model h. Surprisingly, the respective InSAR and PALSAR-based models perform worse than the proposed model h in terms of R and RMSE. The InSAR-based AGB model, used in [22] and developed by [83], uses mean InSAR heights as the only explanatory variable. As canopy heights are highly correlated with AGB [3], [21], this model was expected to correlate better with z than our model h. However, Næsset *et al.* [22] highlight the temporal mismatch between the acquisition of the InSAR data (2012) and the acquisition of the field work (2014) as a probable explanation for the model's low performance. In one case, for example, they identified that a field plot recently had been harvested in 2014, while the InSAR data from 2012 identified biomass in the same area [22]. In theory, we expect a model based on the L-band ALOS PALSAR data to perform better than our C-band based Sentinel-1 model, as C-band data is known to saturate at a lower biomass level than L-band data [5], [53], [54]. However, Table II shows that this is not the case. As the PALSAR data used in [22] consist of a mosaic of yearly scenes, the mosaic does not achieve optimal sensitivity to dynamic AGB levels as scenes from wet and dry seasons are mixed. The low dynamic range of the PALSAR-based and the Landsat-based models is also shown in Fig. 6 and the wall-to-wall maps in Fig. 7. Although most Sentinel-1 predictions on the ground

reference AGB dataset are bounded between 25 and 75 $\mathrm{Mg\,ha^{-1}}$ [see Fig. 6(f)], the model as a whole is able to predict AGB up to around 200 $\mathrm{Mg\,ha^{-1}}$; see Fig. 7(f). The upper limit of the $\widehat{z}_x$-based predictions [Fig. 6(f)] can probably be explained by the low saturation limit of C-band data. Nevertheless, our upper limit of C-band-based AGB predictions is still remarkable, compared to previous studies on biomass retrieval from C-band data, e.g., Imhoff [84] who showed that C-band data saturates around 20 $\mathrm{Mg\,ha^{-1}}$ in the tropical forests of Hawaii. We wish to highlight the fact that the proposed model h is not able to predict biomass close to 0 $\mathrm{Mg\,ha^{-1}}$[see Figs. 6(f) and 7(f). This is probably due to the square root transform in 1 and the correction of bias in 2, the latter applied to achieve correct AGB predictions on arithmetic form, i.e., back-transformation from the $\sqrt{\mathrm{AGB}}$ domain. The InSAR-based model, on the other hand, is able to predict AGB levels close to 0 $\mathrm{Mg\,ha^{-1}}$ [see Table II and Fig. 6(b)] and also achieves lower MAE than the proposed model h.

B. Sequential Regression Models for AGB

This section presents the two alternatives for g, the second model in the sequential modeling approach, i.e., the traditional baseline sequential model and the cGAN-based model. Since the regression model f achieves the highest correlation to z, see [22], we train our two versions of g to use the ALS-based AGB predictions (on pixel-wise form: $\widehat{z}_y$, or patch-wise form: $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$) as a surrogate for z. Each AGB prediction, i.e., $\widehat{z}_y$, represents a square pixel of size 26.6 m $\times$ 26.6 m on the ground. Qualitative and quantitative results from both models are presented and discussed in Section V-B3.

1) Baseline Sequential Regression Model: The proposed baseline sequential regression modeling strategy utilizes the traditional regression model in (2) for both stages in the sequence. In Section V-A, the small size of the z dataset constrained us to use all available data during both model fitting and evaluation. Reusing all available data for both model fitting and evaluation is not optimal, which also Table II shows, i.e., the RMSE computed for model h is lower than the corresponding LOOCV RMSE. In contrast to the situation in Section V-A, the sequential model setting provides access to 516 906 AGB predictions to be used as surrogate response variables. Thus, the dataset size enables us to fit and evaluate model g on different parts of the dataset.

We adopt a dataset split of 20% validation data and 80% test data. We use the validation data to select the models's explanatory variables through stepwise forward selection. Contrary to the nonsequential model h, which only selects VV as a regressor, all six explanatory variables are included in the baseline model g by this method. The final baseline sequential regression model is shown in Table IV. The test dataset was divided into $k = 5$ subsets for k-fold cross-validation (CV). The chosen test metric is CV RMSE (CV-RMSE), which is reported in addition to the Pearson correlation coefficient and the RMSE in Table IV. The latter two metrics are computed on the entire dataset. All reported metrics are computed between the surrogate, i.e., $\widehat{z}_y$, and AGB predictions achieved from the baseline sequential model, i.e., $\widehat{z}_{y|x}$.

TABLE III
THREE OPTIMAL CGAN-BASED MODELS APPLIED FOR THE SECOND PART OF THE SEQUENTIAL MODELING APPROACH

Model reference	Trained with:
Vanilla GAN	ResNet-6, BN, BS = 3 and *PixelGAN* discriminator
LSGAN	ResNet-6, BN, BS = 3 and *PixelGAN* discriminator
WGAN-GP	ResNet-6, BN, BS = 3 and *PixelGAN* discriminator

They were identified from experiments reported in the Appendix; see Sections A2 and A3. Vanilla GAN, LSGAN, and WGAN-GP refer to specific objective functions. BN denotes batch normalization and BS denotes batch size.

2) cGAN-Based Sequential Regression Models: Finally, we approach the sequential modeling strategy from a DL perspective by applying a cGAN for the second regression model, g. The cGAN-based model utilizes convolutional filters in both the G and the D network. Therefore, the image-to-image translation requires the data we condition on, and the output data, to be represented by image patches instead of individual image pixels. Image patches were created from the input data, i.e., the processed Sentinel-1 image, and the output dataset of 516 906 ALS-based AGB predictions, i.e., $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$, similarly and simultaneously. For simplicity, we only describe the process for the Sentinel-1 data. First, nonoverlapping image patches of size 64×64 pixels were extracted in a grid manner from the Sentinel-1 scene in Fig. 5. Each patch corresponds to an area of approximately 289.6 ha on the ground. These nonoverlapping image patches were randomly divided into five disjoint sets for five-fold CV. For each of the five folds, one of the disjoint sets was considered the test set, while the remaining four folds were combined into a training set. To increase the number of image patches further, we extracted additional training patches in each training set by allowing a 50% overlap between adjacent patches. Finally, we applied data augmentation with flipping and rotation to the training image patches. Since we do not allow overlap between test and training image patches, it implies that the final five training sets, after data augmentation, range between 2264 and 2424 patches. Each test set consists of 22 image patches since no data augmentation was applied to the test sets.

By condition on Sentinel-1 image patches, we trained different cGAN-based models to generate realistic-looking synthetic ALS-based AGB prediction image patches, $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$, of size 64×64 pixels. Optimal translation would imply $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}} = \widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ or at least $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}} \approx \widehat{\boldsymbol{z}}_{\boldsymbol{y}}$. All models were trained for 200 epochs with a learning rate of 2×10^{-4}. We refer to Sections A2 and A3 in the Appendix for an extensive evaluation of the impact that the choice of hyperparameters, objective function, and/or discriminator network have on the performance of the different cGAN models. For the remaining of this article, we only report results for the three optimal cGAN-based models listed in Table III, which were identified from the extensive evaluation. Despite the selected objective function, these three models were trained with identical generator architecture, discriminator architecture, and hyperparameters. We therefore refer to them by their objective function, i.e., as the Vanilla GAN, LSGAN, or WGAN-GP model.

As the input and output to each of the optimal cGAN-based sequential models are of size 64×64 pixels, we created synthetic

TABLE IV
PEARSON CORRELATION COEFFICIENTS, R, RMSE, AND CV-RMSE COMPUTED BETWEEN ALS-PREDICTED AGB, $\widehat{z}_y$, AND MODEL-PREDICTED AGB, $\widehat{z}_{y|x}$, ACHIEVED FROM OUR SEQUENTIAL MODELING APPROACH

Auxiliary data source	Modelling approach	Model	R	RMSE	CV-RMSE
Sentinel-1	Baseline sequential[a]	$\widehat{AGB} = \big(-1.61 - 150.51VH - 29.92VV + 53.58\sqrt{VH} + 25.64\sqrt{VV} + 271.36VH^2 + 9.64VV^2\big)^2 + 6.94$	0.41	40.8	40.6
Sentinel-1	Sequential[b]	Vanilla GAN	0.40	42.6	43.6
Sentinel-1	Sequential[b]	LSGAN	0.39	43.0	43.7
Sentinel-1	Sequential[b]	WGAN-GP	0.35	44.6	44.1

All metrics are in units of $\mathrm{Mg\,ha^{-1}}$.
[a]Baseline sequential model, see Sec. V-B1.
[b]cGAN-based sequential models, see Sec. V-B2.

ALS-based AGB prediction maps from the Sentinel-1 scene as follows: The whole AOI was first partitioned into 64×64 image patches with 50% overlap. For each of the optimal models, these patches were then fed into the trained G network to generate synthetic image patches with 50% overlap. The generated synthetic image patches were then merged to construct a $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ prediction map. Due to the overlap between the generated synthetic image patches, most pixels in this intermediate prediction map constitute of a weighted average of pixels from neighboring image patches. Therefore, as a last step to the final $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ prediction map, we apply mosaicking through linear image blending, using the p-norm with a heuristic value of $p = 5$. Different norms were also considered; however, we conclude that the specific choice of the norm has little impact on the blended result.

After training, we evaluated the performance of the Vanilla GAN, LSGAN, and WGAN-GP models against each other and the baseline sequential regression model defined in Section V-B1. We qualitatively and quantitatively compared $\widehat{z}_{y|x}$ generated from the cGAN-based models against the 88 ground reference AGB plots, z, and the surrogate wall-to-wall map of AGB predictions, i.e., $\widehat{z}_y$.

3) Sequential Model Evaluation: Here, we present results and evaluate the two subsequent models, g, that were proposed in Sections V-B1 and V-B2. Note that the performance assessment in Table IV and Fig. 9 is performed with respect to the ALS-predicted $\hat{z}_y$, which in the sequential modelling strategy replaces ground reference z.

Computed metrics between $\widehat{z}_{y|x}$ and $\widehat{z}_y$, i.e., the Pearson correlation coefficient (R), RMSE, and CV-RMSE, for all four sequential models are collected in Table IV. Results in Table IV indicate that the baseline sequential model is preferred to the three cGAN-based models as it experiences both a smaller RMSE and CV-RMSE, and a higher R with respect to $\widehat{z}_y$. Among the cGAN-based models, the Vanilla GAN is preferred as it achieves the highest correlation and the lowest RMSE to $\widehat{z}_y$. However, the Vanilla GAN model also experiences the largest difference between RMSE and CV-RMSE, implying that AGB predictions retrieved from this model are less consistent.

Generated synthetic AGB prediction maps for the proposed sequential models are shown in Fig. 8. The prediction map from the baseline sequential model is shown in Fig. 8(b), while Fig. 8(c)–(e) shows corresponding prediction maps constructed from the cGAN-based models, i.e., the Vanilla GAN, LSGAN, and WGAN-GP model. The ultimate goal of the sequential model g is to achieve AGB prediction maps that resemble the $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ prediction map in Fig. 8(a). Although the computed metrics for the baseline sequential regression model indicate that this model is preferred to the cGAN-based models, it is unable to capture the dynamic range of ALS-based AGB predictions; see Fig. 8(b). The model's inability to predict near-zero biomass is particularly severe, which, similar to model model h, can be explained by the square root transform and the bias correction applied. The cGAN-based models are, however, able to predict zero biomass. Their constructed biomass maps also exhibit a higher dynamic range in levels of predicted biomass. All sequential AGB models are generally underpredicting $\widehat{z}_y$.

In Fig. 9, we visualize density plots between $\widehat{z}_y$ and predicted AGB from the proposed sequential AGB regression models. The white lines indicate a reference line for 100% correlation between $\widehat{z}_y$ and $\widehat{z}_{y|x}$. While the baseline model achieves better RMSE and R, the Vanilla GAN model achieves the lowest MAE. We note that all four sequential models struggle to predict $\widehat{z}_y$ correctly at low AGB levels. They are generally biased toward overpredicting at low $\widehat{z}_y$. While the cGAN-based models manage to predict zero biomass, the baseline model cannot. Since the baseline model only predicts AGB over 100 $\mathrm{Mg\,ha^{-1}}$ occasionally, it consequently underpredicts high $\widehat{z}_y$. The density plots of the three cGAN-based models indicate that they also underpredict high levels of $\widehat{z}_y$, but not to the same extent as the baseline sequential model.

We also compute the pixel-wise difference between $\widehat{z}_y$ and $\widehat{z}_{y|x}$, i.e., $\widehat{z}_y - \widehat{z}_{y|x}$, for each proposed sequential models. The pixel-wise differences are visualized in Fig. 10, where Fig. 10(b) is the difference for the baseline model, Fig. 10(c) for the Vanilla GAN model, Fig. 10(d) for the LSGAN model, and Fig. 10(e) for the WGAN-GP model. By comparing the AGB difference maps in Fig. 10 with the actual $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ prediction maps in Fig. 8, we again show that all sequential models underpredict AGB in areas with high levels of $\widehat{z}_y$ [shown as pink or blue in (b)–(e)]. We also highlight that at all sequential models overpredict AGB areas with low levels of $\widehat{z}_y$ [shown as green in (b)–(e)]. The baseline sequential model's inability to predict zero or low levels of biomass can probably explain the larger extent of green regions in Fig. 10(b), compared to Fig. 10(c)–(e).

For further comparison, we provide sequential modeling results for the few ground reference AGB measurement we have available. We argue that achieving large-scale AGB maps that reflect the dynamic range of $\widehat{z}_y$ is one desired goal, but more important is the ability of the AGB predictions to match z values.

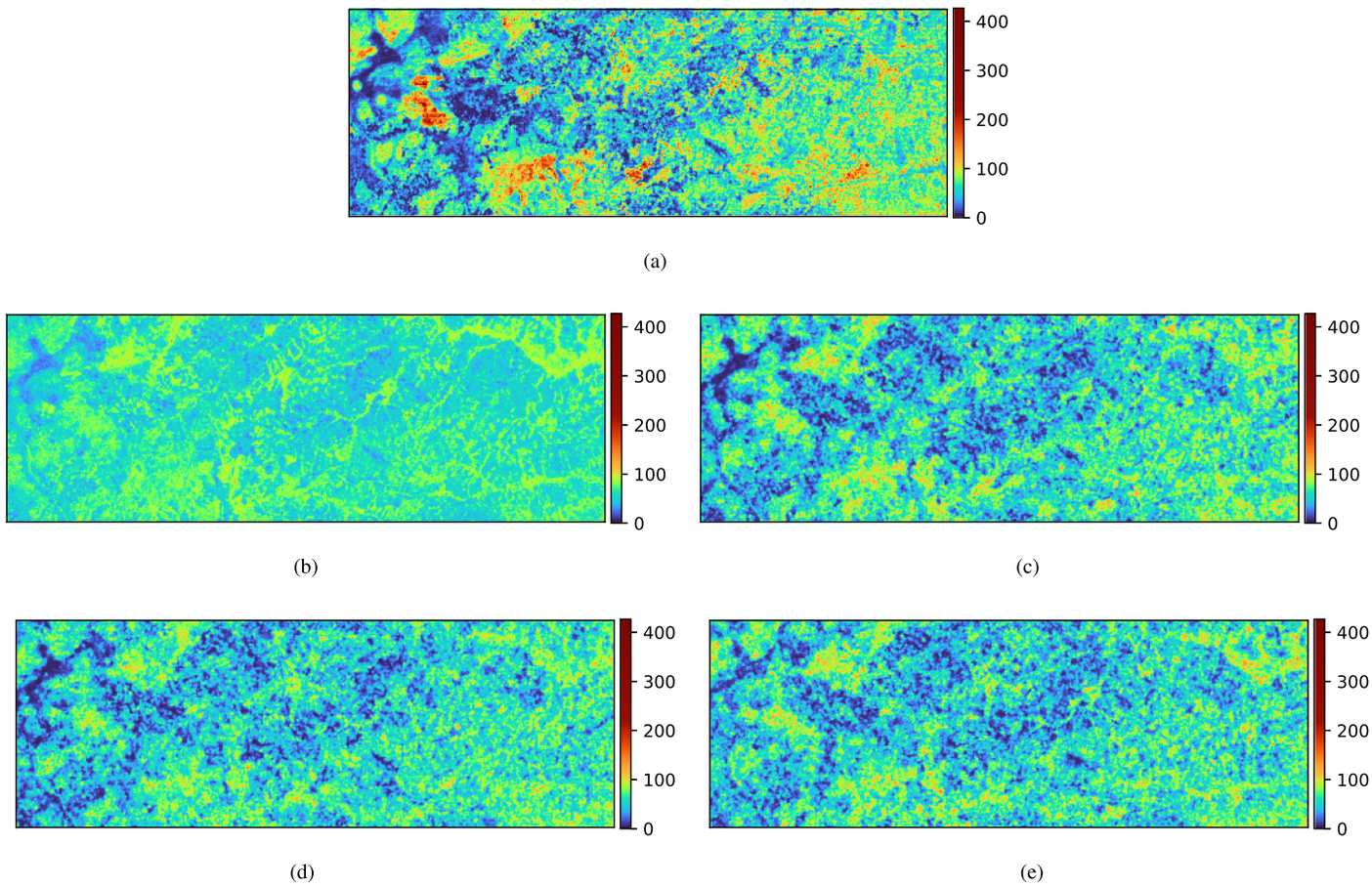


Fig. 8. Generated synthetic ALS-based AGB prediction maps (in $\mathrm{Mg\,ha^{-1}}$) together with the surrogate for ground reference plots, i.e., the ALS-based AGB map shown in (a) [this AGB map is the same as in Fig. 7(a)]. (b) A synthetic ALS-based AGB prediction map generated through the baseline sequential Sentinel-1-model [see (2)]. (b)–(e) Generated synthetic ALS-based AGB prediction maps generated through our proposed sequential regression models using (c) Vanilla GAN, (d) LSGAN, and (e) WGAN-GP.

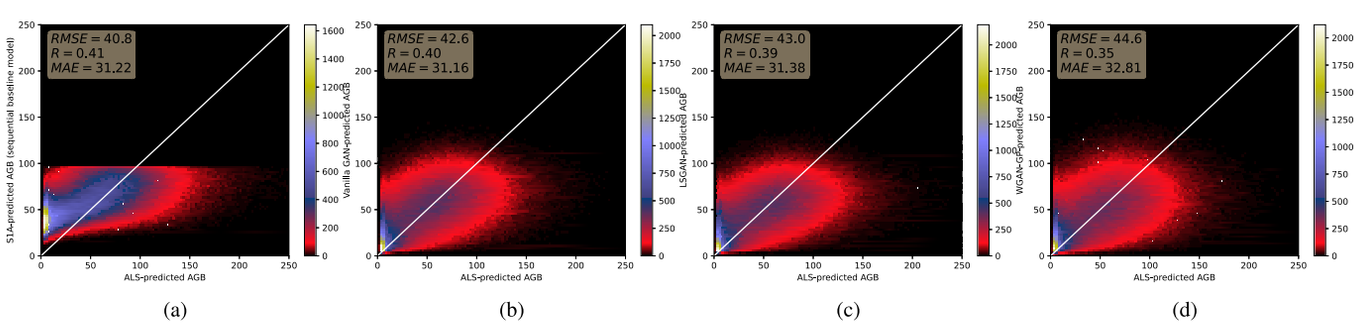


Fig. 9. Density plots between constructed AGB maps and ALS-based AGB biomass predictions, $\widehat{z}_y$, for (a) baseline sequential model, (b) Vanilla GAN, (c) LSGAN, and (d) WGAN-GP models. Reported metrics are the RMSE, Pearson correlation coefficient (R), and the MAE between $\widehat{z}_y$ and the sequential model-based AGB predictions. The white lines are reference lines indicating 100% correlation between $\widehat{z}_y$ and predictions.

Thus, we computed the correlation between AGB predictions obtained with the proposed sequential modeling strategy and the 88 ground reference plots, shown with red markers in Fig. 4. Since the physical area of each ground reference plot could intersect with several of the grids with pixel size 707 m^2, we calculated the area-weighted mean of grid pixels intersecting with each separate ground reference plot. Fig. 11 shows scatter plots of the correlation between z and model-predicted AGB, retrieved from the sequential models, together with computed metrics: i.e., RMSE, R, and MAE. Quantitative results derived from Fig. 11 are also summarized in Table V together with computed metrics for model f. Similar to the scatter plot for model h, Fig. 11(a) also indicates that AGB predictions from the baseline sequential model are bounded between 25 and 75 $\mathrm{Mg\,ha^{-1}}$. Table V shows that neither of the proposed sequential models achieves as high correlation or low RMSE and MAE with respect to z that model f achieves. Nevertheless, it should be noted that f [22] was fitted against the available z. The sequential

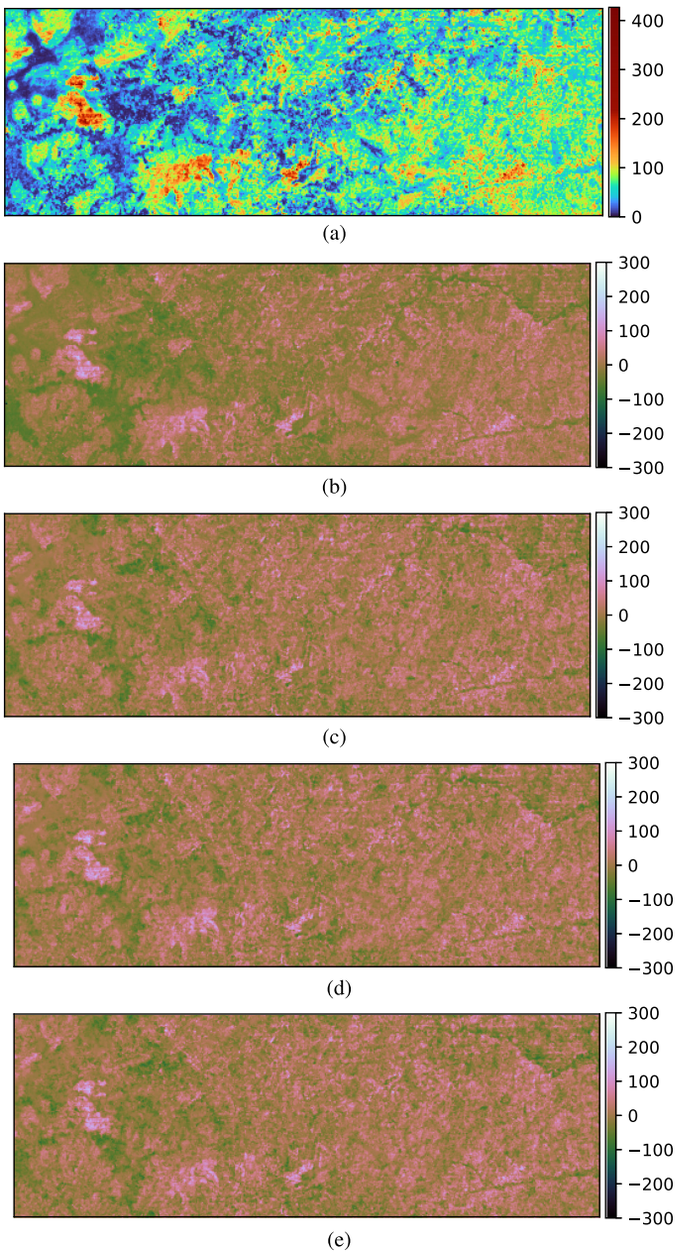


Fig. 10. AGB difference maps (in $\mathrm{Mg\,ha^{-1}}$). Pixel-wise difference between the ALS-based AGB prediction map, shown in **(a)**, and constructed AGB prediction maps achieved form the four sequential models: baseline sequential model **(b)**, Vanilla GAN **(c)**, LSGAN **(d)** and WGAN-GP **(e)**.

TABLE V
COMPUTED PEARSON CORRELATION COEFFICIENTS (R), RMSE, AND MAE BETWEEN AREA-WEIGHTED MEANS RETRIEVED FROM REGRESSION MODELS AND GROUND REFERENCE PLOTS OF AGB ($\mathrm{Mg\,ha^{-1}}$)

AGB prediction models based on:	R	RMSE	MAE
ALS[a]	0.68	33.39	24.61
Baseline sequential [b]	0.43	41.88	33.36
Vanilla GAN[c]	0.46	41.33	32.82
LSGAN[c]	0.50	39.84	31.46
WGAN-GP[c]	0.42	42.03	32.97

[a]The non-sequential ALS-based regression model proposed in [22].
[b]The baseline sequential regression model, proposed in Sec. V-B1.
[c]The cGAN-based sequential regression models, proposed in Sec. V-B2.

TABLE VI
OVERALL RMSE AND RMSE COMPUTED FOR EACH QUARTILE, I.E., RMSE($Q_{0,1}$), RMSE ($Q_{1,2}$), RMSE($Q_{2,3}$), AND RMSE ($Q_{3,4}$) (LOWER IS BETTER)

Model	RMSE	RMSE ($Q_{0,1}$)	RMSE ($Q_{1,2}$)	RMSE ($Q_{2,3}$)	RMSE ($Q_{3,4}$)
Non-sequential	84.5	67.0	59.8	86.3	114.3
Baseline sequential	40.8	40.0	27.2	19.0	63.0
Vanilla GAN	42.6	32.4	29.0	27.6	67.7
LSGAN	43.0	31.7	27.6	27.3	69.8
WGAN-GP	44.6	34.0	30.7	30.4	70.3

The RMSE metrics are computed between AGB prediction maps constructed in this work and the ALS-based AGB prediction map. All metrics are in units of $\mathrm{Mg\,ha^{-1}}$.

models, on the other hand, were optimized to achieve $\widehat{z}_{y|x} \approx \widehat{z}_y$ as they were fitted against $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$.[3] While Table IV indicates that the baseline sequential regression model predicts $\widehat{z}_y$ best, Table V indicates that both the LSGAN model and the Vanilla GAN model perform better than the baseline sequential model on all three metrics. Additionally, all cGAN-based models obtain lower MAE with respect to z than the baseline sequential model achieves. Among them, the LSGAN model performs best in predicting z. Additionally, all cGAN-based models obtain lower MAE with respect to z than the baseline sequential model achieves. Interestingly, by comparing Table II with Table V, we identify the LSGAN model, in terms of R and RMSE, to perform better in predicting z than the InSAR model. We therefore argue that the LSGAN and the Vanilla GAN model should be the first and second choice if one aims to achieve a model that reflects the dynamic range of the true AGB best.

C. Nonsequential and Sequential Modeling

To broaden the discussion, evaluate the suitability of the Sentinel-1 sensor as a data source for AGB regression models and enable further comparison of the nonsequential and sequential modeling strategies, we provide three additional results: Fig. 12 and Tables VI and VII.

In Fig. 12(d), we show histogram plots over predicted AGB values derived from the ALS-based regression model f together with AGB predictions from models proposed in this work: The nonsequential Sentinel-1 model [Fig. 12(b)], the baseline sequential model [Fig. 12(c)], the Vanilla GAN model [Fig. 12(e)], the LSGAN model [Fig. 12(f)], and the WGAN-GP model [Fig. 12(g)]. We also show a histogram of measured ground reference AGB, z, in Fig. 12(a) overlaid with a nonparametric estimate of the underlying probability density function. Note the similarities between the distributions of z and $\widehat{z}_y$ [22] in Fig. 12(b). Besides not being able to predict low AGB values [see Fig. 12(b) and (c)], both the nonsequential Sentinel-1 model and

[3]In Section A6 in the Appendix, we experiment with an additional calibration step to further calibrate model g against z. Results indicate that post-calibration of the output from g with either gamma or linear calibration increases the accuracy and the correlation by a small amount. Nevertheless, the possible improvement is modest and we omit this additional step as the nonsequential Sentinel-1-based model still outperforms the post-calibrated sequential models on computed RMSE, MAE and R.

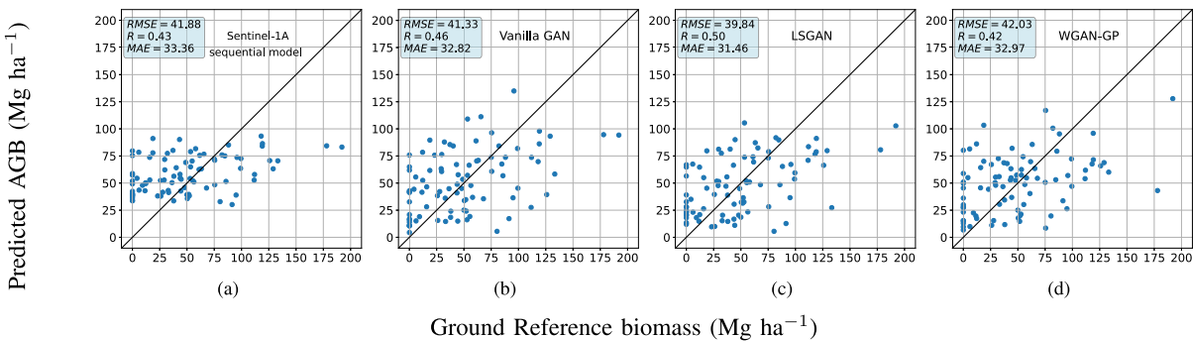


Fig. 11. Scatter plots between ground reference AGB, z, and model-predicted AGB. Model-predicted AGB values are retrieved from the (a) baseline sequential regression model; see Section V-B1, or the proposed cGAN-based sequential models; (b) Vanilla GAN, (c) LSGAN, and (d) WGAN-GP. See Section V-B2 for details on the cGAN-based methods. The black lines are reference lines indicating 100% correlation between z and predictions.

TABLE VII
PEARSON CORRELATION COEFFICIENT COMPUTED BETWEEN PIXEL-WISE PREDICTIONS FOR PAIRS OF MAPS VISUALIZED IN FIG. 7 (UPPER TABLE) AND FIG. 8

	InSAR	RapidEye	Landsat	PALSAR	Sentinel-1 (non-sequential)
ALS	0.65	0.48	0.42	0.28	0.16
InSAR		0.38	0.57	0.38	0.15
RapidEye			0.31	0.22	0.15
Landsat				0.34	0.13
PALSAR					0.02

	Vanilla GAN	LSGAN	WGAN-GP	Sentinel-1 (sequential)
ALS	0.4	0.39	0.35	0.41
Vanilla GAN		0.59	0.55	0.62
LGAN			0.53	0.60
WGAN-GP				0.59

(Lower Table). Models referred to as ALS, InSAR, RapidEye, Landsat, and PALSAR are retrieved from [22]. Remaining models are developed for this work.

the baseline sequential model predict some extreme AGB values of 15 640 $\mathrm{Mg\,ha^{-1}}$ in Fig. 12(b) and 1751 $\mathrm{Mg\,ha^{-1}}$ in Fig. 12(b), which neither of the cGAN-based models do. Instead, the maximum predicted AGB from the three cGAN-based AGB models are rather close to the maximum measured AGB in the field plots, i.e., 213.4 $\mathrm{Mg\,ha^{-1}}$ [22]. Also, all cGAN-based models behave more similar to z and f for middle-to-high levels of AGB; see Fig. 12(e), (f), and (g) compared to Fig. 12(a) and (d). This could indicate that the more complex cGAN-based models have learned AGB dynamics of z and f better in middle-to-high levels of AGB, than the simpler nonsequential and baseline sequential model manages.

To emphasize where the proposed models are more or less consistent with the ALS-based AGB prediction map, we evaluate AGB predictions from the five models against $\widehat{z}_{y|x}$ in terms of overall RMSE and RMSE computed for each quartile. Results provided in Table VI clearly show that AGB predictions from the nonsequential Sentinel-1 model deviate most from $\widehat{z}_{y|x}$, both overall and in each quartile. The baseline sequential model is most similar to $\widehat{z}_{y|x}$ in the second and third quartile and achieves the smallest RMSE among all five proposed models in the fourth quartile. As expected from the histograms in Fig. 12 and the constructed AGB prediction maps in Fig. 8, Table VI shows that all cGAN-based models produce low RMSE in the first quarter quartile, with the LSGAN model being better than the Vanilla GAN model. Among the cGAN-based models, the Vanilla GAN model only receives the smallest RMSE in the fourth quartile. Once again, it is shown in Table VI that the WGAN-GP model is the worst among the cGAN-based models.

Table VII shows the Pearson correlation coefficient computed between pixel-wise AGB predictions for pairs of maps from either Fig. 7 or 8. Correlations computed between AGB predictions retrieved from the nonsequential models are listed in the upper part of Table VII, while correlations computed between AGB predictions from the sequential models and the surrogate regression target, i.e., $\widehat{z}_y$, are combined in the lower part of the table. Previous results from [22] identified the ALS-based AGB prediction map and the InSAR-based AGB map to have the greatest correlation with each other (see Table VII), and with z (see Table II). AGB predictions from the Landsat- and PALSAR-based models achieved the smallest correlation with z; see Table II. The proposed nonsequential model h achieves by far the lowest correlations with any of the other five nonsequential AGB models; see Table VII. This is probably a consequence of the Sentinel-1-based model's inability to predict low biomass levels. For example, the left part of the AOI (see Fig. 7), the ALS-, InSAR-, and the RapidEye-based AGB models predict AGB around 0 $\mathrm{Mg\,ha^{-1}}$ in approximately the same areas, while predicted AGB levels retrieved from the nonsequential Sentinel-1-based model deviates highly in the same areas. Note that all sequential models achieve a much higher correlation with model f than what model h achieves. Logically, this could be explained by the fact that all sequential models were fitted against f. While the nonsequential Sentinel-1-based model h, the InSAR model, and the ALS model f achieve the highest correlations and lowest RMSE with respect to z, the surprisingly low correlation between h and f is notable. It could imply that model h is overconfident on the small set of z measurements. Among the sequential models, Tables IV and VII show that the proposed baseline model achieves the lowest RMSE and highest correlation coefficient with respect to $\widehat{z}_y$. Furthermore, the cGAN-based model trained with the WGAN-GP objective function achieves the smallest correlation with $\widehat{z}_y$; see Table VII. Overall, the correlations between the sequential models and the ALS-based model f are all higher than the corresponding correlation between AGB predictions from f and the PALSAR model,

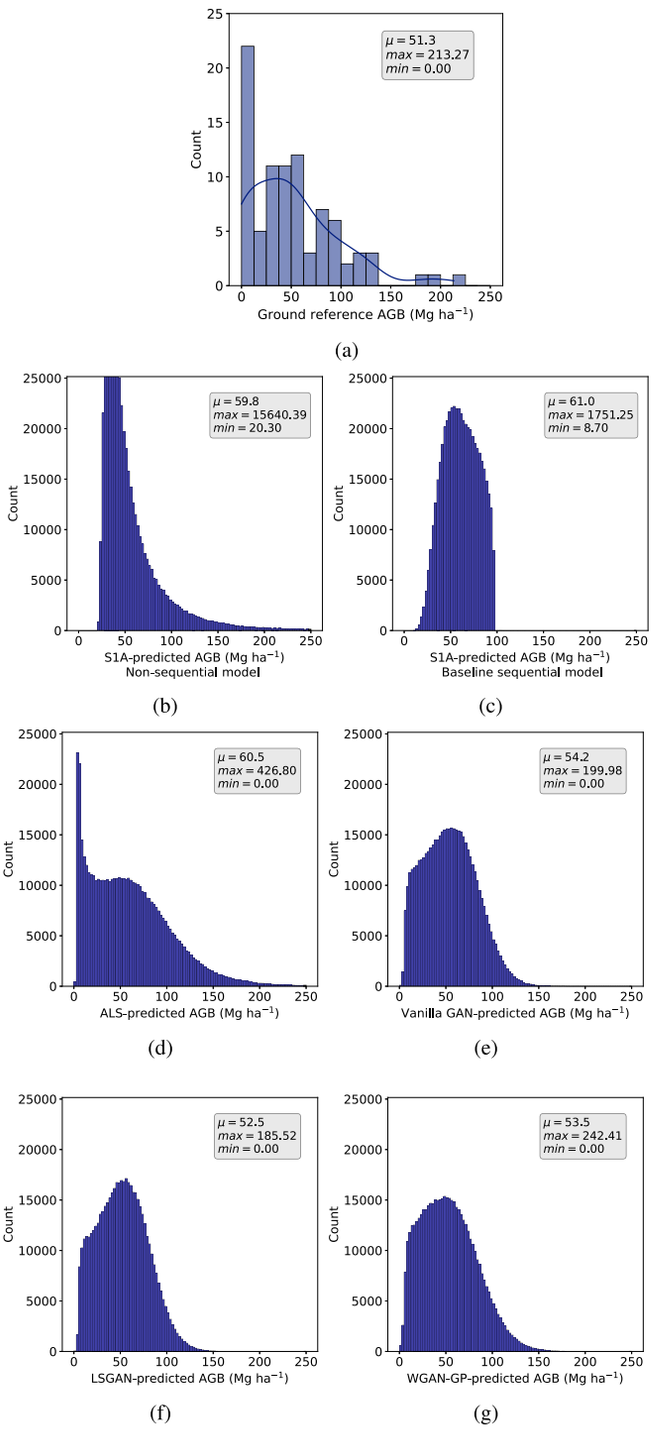


Fig. 12. Histograms of AGB predictions from the proposed AGB models. (b) Nonsequential Sentinel-1. (c) Baseline sequential. (e) Vanilla GAN. (f) LSGAN. (g) WGAN-GP. A histogram over the collected ground reference AGB is shown in (a), while (d) shows a histogram over ALS-based AGB predictions. Reported metrics are the sample mean, μ, median, and maximum and minimum predicted AGB (Mg ha^{-1}).

and similar to the correlation between AGB predictions from f and the Landsat model. In addition to the discovery that the LSGAN model performs better than the nonsequential InSAR model in predicting z, these results suggest that the cGAN-based sequential modeling approach and the use of Sentinel-1 data for AGB prediction are worth pursuing further.

VI. Discussion

The focus of this work has been to develop nonsequential and sequential regression models based on C-band SAR for AGB prediction in Tanzania. One main advantage of utilizing Sentinel-1 data as regressors is that it enables frequent and affordable updates of an AGB map with extensive coverage. This approach has a low cost compared to keeping the most accurate prediction model from [22] up-to-date by repeated acquisition of ALS data. Our results show that the proposed nonsequential Sentinel-1-based model h and the sequential LSGAN model best provide AGB predictions close to measured ground reference AGB, z. Only the ALS and the RapidEye-based model in [22] perform better on this task. Noteworthy, in terms of R and RMSE, both the model h and the sequential LSGAN model were identified to be more accurate than the InSAR-based nonsequential model on the same task. Since the InSAR-based model provides estimates of canopy height that are highly correlated with AGB [3], [21], we expected it to be superior in predicting z. We emphasize that we are training all our models using C-band SAR intensity data, which have previously been shown to suffer from much lower saturation levels than, e.g., the L-band ALOS PALSAR sensor. As C-band data neither penetrates as deeply into the forest volume as L-band data, nor can it compete with the accuracy of AGB estimates produced from optical data [20], [52]–[54], it has traditionally been considered an inferior information source for AGB estimation. Thus, we have in this work demonstrated the potential of using Sentinel-1 data for AGB prediction and suggest further research on Sentinel-1-based models for AGB retrieval.

Formally, the proposed models were assessed in terms of their relative accuracy on AGB prediction with respect to model f, [22], and available AGB *in situ* measurements. However, whenever a certain methodology is implemented for operational purposes in an MRV system, the ultimate goal is to produce estimates of carbon stocks and changes. Among these, estimates for the AGB pool are essential. Further, the Intergovernmental Panel on Climate Change specifies that results should be reported as inferences in the form of confidence intervals [85] (p. 1.10). Thus, although the maps themselves can be useful, for example, to identify critical areas of carbon loss, the prediction map is just an intermediate product on the way to estimating the carbon budget. AGB can easily be estimated from the prediction maps constructed by the current methods by aggregating individual pixel values. Estimating the uncertainty of AGB estimates in the form of variances or confidence intervals for nonparametric methods such as ANNs, support vector machines, random forest regression, and other techniques is a current research issue. To provide such estimates was beyond the scope of the current study. Recent applications of, e.g., bootstrap resampling for random forest-based prediction models demonstrate that such variance estimators may easily be adopted for ANN models as well; see e.g., [86]. However, the computational burden will be substantial.

By approaching AGB prediction through sequential modeling with ALS-based predictions as a surrogate for z, deep contextual models could be utilized for the regression task. As far as we know, this is the first time that contextual cGAN models

have been used to simulate ALS-based AGB prediction maps from Sentinel-1 data. A natural question is whether DL-based approaches for AGB predictions are worth further investigation, especially since they are more complex to train than traditional statistical regression models. We would argue that more research is needed in utilizing contextual DL models to retrieve biophysical parameters from RS data. We have shown that the LSGAN model performs well and reproduces dynamic AGB levels more realistically than simpler noncontextual models. Despite this, the cGAN-based models fall behind the traditional sequential and nonsequential models on RMSE with respect to ground reference data. The trade-off between perceptual quality and reconstruction accuracy is known from the research field of single image superresolution (SR), [87]–[91]. SR in RS data has been studied in, e.g., [92]–[94]. For future work on AGB prediction by DL regression, it appears relevant to incorporate ideas from the field of SR and investigate additional architectures and balancing of GAN losses against traditional L_1 and L_2 losses for reconstruction. The purpose would be to obtain a model that focuses on the reconstruction loss, yet produces AGB prediction maps that maintain local dynamics.

A. Error Discussion

The accuracy of the proposed models is influenced by many factors, such as the radiometric accuracy of radar images, time lag, and error propagation through the model sequence. The latter was also pointed out in [38]. We refer to the time lag as the time difference between collecting the field inventory data in January–February 2014, the acquirement of ALS measurements in 2014, and the acquisition of the Sentinel-1 scene in September 2015. Possible inaccuracies may propagate, first when the ALS model upscales the field inventory data to a $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ map, and, second, when the sequential models are trained. Additional factors that may affect the overall accuracy is resampling of the Sentinel-1 scene to the same grid as the ALS-based AGB prediction map or the image blending process which is applied to construct the full cGAN-based AGB prediction maps from a set of patches. Despite this, the advantage of using a sequential modeling approach on Sentinel-1 data is the ability to achieve biomass prediction maps with high update frequency on a national level. Our sequential approach also has potential use in biomass change detection, where the relative change of biomass from one time to another is of higher interest than the absolute AGB values.

As previously mentioned, z was collected within circular sample plots, the most common plot shape in boreal and temperate forest sampling [74]. However, all remotely sensed datasets used in [22] and this work are represented by square pixels. Therefore, using circular field plots is suboptimal, as each model's correspondence to z needs to be computed by an area-weighted mean of neighboring pixels. The sequential models are not directly related to the circular plots, but through the $\widehat{z}_y$, which was trained against z. Nevertheless, when computing the correspondence between the sequential model's AGB predictions and z, the above challenge arises when the area-weighted mean between square pixels intersecting a circular pixel is computed. In the end, this will influence the overall accuracy of the models. Note that the sampling design in [74] was optimized for field-based estimation of AGB given a limited budget for inventories, not for upscaling supported by RS, in which case the species diversity and spatial variability of AGB in the miombo woodlands imply that larger sample plots should be used. We sustain [95], which concludes that decisions regarding the sample plot size, and thereby its shape, is one of many parameters that have to be considered in future field-based surveys if one aims to enhance estimation through the use of remotely sensed data.

VII. Conclusion

The focus of this work was to investigate the suitability of Sentinel-1-based models for AGB prediction in an MRV system for miombo forests in Tanzania. Previously, Næsset *et al.* [22] developed traditional nonsequential regression models for AGB in a Tanzanian AOI using either ALS, TanDEM-X InSAR, RapidEye, Landsat, or PALSAR data with a limited amount of ground reference AGB data. The ALS-based AGB predictions achieved the highest accuracy, but the cost and infrequent update of ALS data prevent this model from being of practical use in an MRV system. Therefore, we turned to freely available and easily accessible Sentinel-1 data for this work and developed three different models for AGB prediction from this source: A traditional nonsequential model, a baseline sequential model, and a DL-based sequential model. We compared each model's accuracy on the AGB prediction task. Additionally, maps of AGB predictions were compared and evaluated with respect to their ability to recreate realistic biomass dynamics. The model performances and most important results are summarized below.

1) Nonsequential Sentinel-1 Model: This model was, as the models in [22], trained against the limited ground reference data. Its performance can, therefore, be directly compared to the results in [22]. Among all models proposed for this work, this model achieves the lowest RMSE and highest correlation coefficient (R) against ground reference data. Although this model cannot predict AGB levels between 0 and 20 $\mathrm{Mg\,ha^{-1}}$, it performs better than the InSAR-based model in terms of R and RMSE. It is only beaten by the ALS-based and the RapidEye-based models [22]. However, the nonsequential Sentinel-1 model achieves the highest RMSE in a pixel-by-pixel comparison with the ALS-predicted AGB map. Hence, we conclude that it sacrifices a more realistic prediction of the dynamic range and local variability of AGB values to meet the goal of producing a low RMSE against ground reference data.

2) Sequential Models: These were developed to enable AGB prediction on a larger scale through a modeling strategy with two subsequent regression models. We propose to employ the ALS-based model [22] as the first model. The second model in the chain is trained to relate SAR backscatter images to ALS-based AGB prediction maps, which are used as a surrogate for ground reference data. The baseline sequential model applies a traditional statistical regression model also in the second stage. The alternative sequential model instead uses a DDN for cross-modal image-to-image translation, i.e., the Pix2Pix cGAN architecture [35] with some modifications warranted by the application. This cGAN architecture generates synthesized

ALS-based AGB predictions during model fitting by conditioning on SAR backscatter data. In contrast to the other models, it uses contextual information from pixel neighborhoods in its predictions. The baseline sequential model, followed by the Vanilla GAN model, achieves the highest R and lowest RMSE against the ALS-based AGB predictions. Conversely, the LS-GAN model is the best among the sequential models at reproducing ground reference data, and is only beaten by the ALS-based and RapidEye-based models from [22]. In this respect, the LSGAN model achieves slightly higher RMSE and lower R than the nonsequential Sentinel-1-based model. However, the LSGAN model can predict AGB levels around 0 $\mathrm{Mg\,ha^{-1}}$ and also achieves higher correlation with the ALS-based predictions. Thus, the contextual cGAN-based models seem to better capture the dynamic range and local variability of AGB.

We have in this research demonstrated the potential of utilizing Sentinel-1 data for AGB prediction in Tanzania. Although C-band Sentinel-1 data traditionally have been considered an inferior information source for AGB estimation due to low penetration of the canopy, our results show that Sentinel-1-based models are a viable alternative for forest AGB retrieval, especially considering that the data are freely available.

APPENDIX

This appendix includes a specification of the modifications done to the *Pix2Pix* architecture [35] to make it suitable for generation of synthetic ALS-based AGB image patches in our sequential modeling strategy. It also provides additional experiments and results that were conducted for this work.

A. Modified Pix2Pix Architecture

The cGAN-based sequential model used for generation of synthetic ALS-based AGB image patches, $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$, is based on the image-to-image translation framework *Pix2Pix* [35]. To meet our needs, we modified it in the following ways.

1) We enable the use of calibrated pixel values read from image files in GeoTIFF format. This is necessary since we work with images with pixel values that carry information about physical entities and represent either calibrated σ_0 values (backscatter coefficients) or AGB predictions measured in $\mathrm{Mg\,ha^{-1}}$.
2) We change the activation function in the output layer from a hyperbolic tangent (tanh) function used in [35] to a rectified linear unit (ReLU) activation function. In an earlier phase of this work [96], we noticed that the tanh activation function we used in the output layer generated AGB values that overestimated the ALS-based AGB predictions from [22], and particularly failed to predict AGB values close to zero. An essential criterion for our cGAN regression model is that it should be able to predict zero biomass to correlate well with AGB ground reference data, z, in non-vegetated areas. The overprediction observed in [96] can be explained by the nature of the tanh activation function. As the range of the tanh function is [0.0, 1.0], it implies that all data introduced to the cGAN need to be normalized to the same range. The tanh function must output exactly zero to predict zero biomass, which only happens when the action potential goes to $-\infty$. This explains why prediction with the tanh function seems to clip the AGB values at a level higher than zero.

In conclusion, by substituting the tanh activation function with a ReLU function in the output layer and allowing the regression target to be calibrated AGB values in $\mathrm{Mg\,ha^{-1}}$ units, instead of being normalized to [0.0, 1.0], our modified Pix2Pix architecture no longer overestimates AGB that should be close to zero.

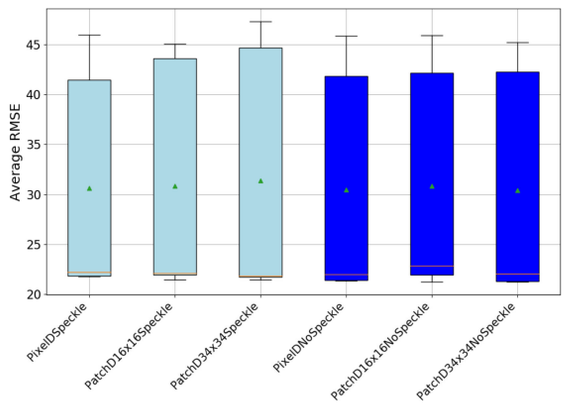


Fig. 13. Boxplot comparison between models trained with different types of D on datasets produced with or without speckle filtering. Green triangles indicate the mean value computed over the five folds, while orange horizontal lines indicate the median.

B. Experiment 1: A Study of the Impact of Speckle Filtering and Choice of Discriminator Network

A common preprocessing step for SAR products is speckle filtering. Speckle filters reduce the effects of the inherent speckle phenomenon on the product and smooths the pixel values. In this experiment, we evaluate if speckle filtering of the Sentinel-1product affects the accuracy and the quality of cGAN-generated AGB predictions. To this end, we created two different datasets from the Sentinel-1 GRD product: The first was produced by following the SAR processing workflow defined in Section IV-B; for the second dataset, we used the refined Lee filter [77] with SNAP's default window size of 7×7 to apply speckle filtering between steps 4) and 5) in the same workflow. We refer to them as the Sentinel-1 dataset with and without speckle filtering. A separate cGAN network was trained on each.

Additionally, we evaluated the three discriminator networks D presented in Section IV-D2 against each other to assess their effect on cGAN performance for data generation. For all experiments in this section, we trained the cGAN for 200 epochs using a ResNet-6 network, WGAN-GP objective function, batch size (BS) of 2, layer normalization (LN) for D, and batch normalization (BN) for G. These settings were determined by the model validation results presented in [96].

Results: A boxplot of average RMSE, computed between $\widehat{z}_y$ and $\widehat{z}_{y|x}$ for the different models trained with five-fold CV, is shown in Fig. 13. Light blue bars indicate results obtained with models trained on speckle filtered data, while dark blue bars represent models trained on unfiltered data. Within a specific color, the left, middle, and right-most bar represent models trained with a *PixelGAN*, a 16×16 *PatchGAN*, and a 34×34

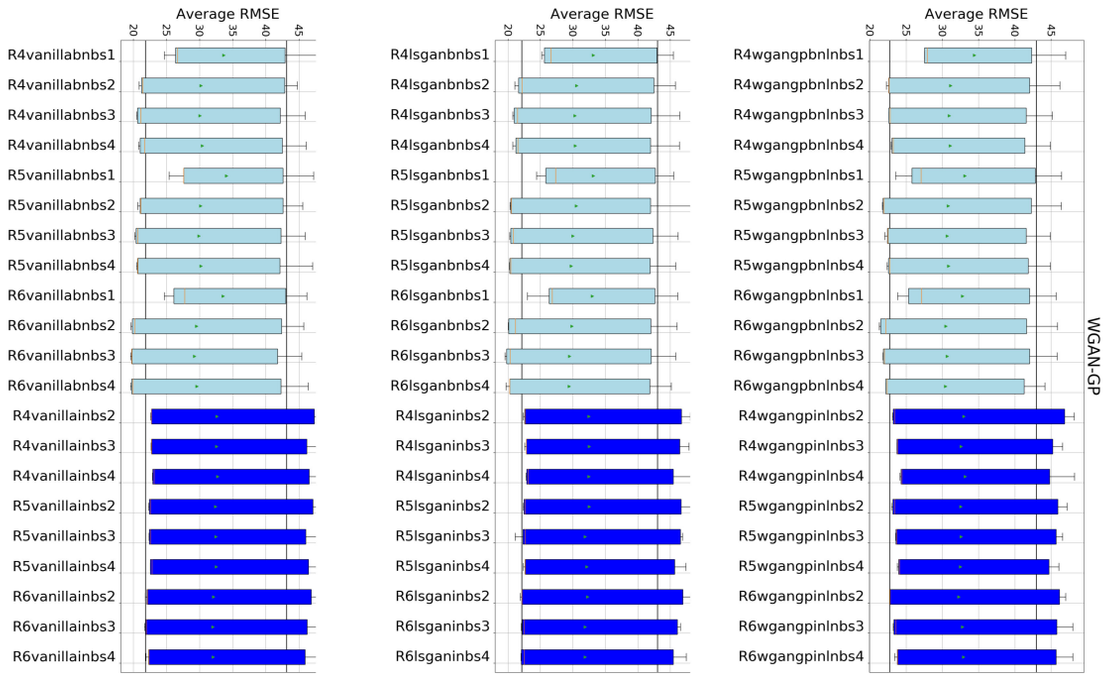


Fig. 14. Impact of normalization method, BN or IN, on model performance. Light blue bins and dark blue bins represent models trained with BS and IN, respectively. Green triangles are mean values computed over the five folds, while orange vertical lines are medians. The two vertical black lines in each column are arbitrary reference lines to easen visual comparison. *Columns:* Models trained with Vanilla GAN (left), LSGAN (middle), and WGAN-GP (right).

PatchGAN, respectively. Overall, Fig. 13 shows less spread and tighter boxes for models trained on the dataset where speckle filtering was omitted. Thus, during preprocessing of the Sentinel-1 product, speckle filtering should be skipped to achieve slightly smaller RMSE between $\widehat{z}_y$ and $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$. In general, Fig. 13 also shows that the specific type of discriminator has little impact on the average RMSE for the dataset without speckle filtering. As the *PixelGAN* discriminator produces slightly less spread than the two other discriminator networks, we applied it to all remaining experiments in this work and choose to omit speckle filtering in the processing of the Sentinel-1 product.

C. Experiment 2: A Comparison of Model Architectures, Normalization Methods, and Objective Functions

Here, we investigated if any combination of model architecture, normalization method, and cGAN objective function improves the accuracy of $\widehat{z}_{y|x}$ with respect to $\widehat{z}_y$. Based on the results in Section A2, we kept the dataset fixed, i.e., we used the Sentinel-1product processed without speckle filtering and applied the 1×1 *PixelGAN* discriminator for all models trained in this section. Nine different cGAN generator architectures G were trained by combining the three ResNet networks and the three objective functions from Section IV. We also applied BN or instance normalization (IN) for Vanilla GAN and LSGAN, while for WGAN-GP, we applied LN for D and either BN or IN the G network, as suggested in [81]. We additionally experimented with a BS between 1 and 4. For each model, we applied 5-fold CV, and trained it for 200 epochs. We evaluate the different models on the 5-fold CV test sets by visualizing boxplots of average RMSE computed between $\widehat{z}_y$ and $\widehat{z}_{y|x}$.

Results: Fig. 14 visualizes models trained on the three different objective functions in separate columns, i.e., Vanilla GAN in the left column, LSGAN in the middle column, and WGAN-GP in the right column. We show models trained with BN in light blue color, while models trained with IN are shown in dark blue color. For all three objective functions, models trained with BN achieve a smaller average RMSE. Additionally, Fig. 14 shows that most models trained with BN also experience a smaller spread in average RMSE over the 5-fold CV dataset. Thus, we conclude from Fig. 14 that applying BN is preferable to produce $\widehat{z}_{y|x}$ predictions with smaller average RMSE.

In Fig. 15, we compare models trained with different ResNet architectures and BS values to each other. In the left column, the models are first sorted by objective function, then by ascending BS, and finally by ascending ResNet model order. The grouping by BS is indicated with colors. In the middle column, models are again first sorted by objective function, but then by ascending ResNet model order (color-coded groups), and finally by ascending BS. In the right column, models are first sorted by ascending ResNet model order, then by ascending BS (color-coded groups), and, finally, by objective function. Overall, Fig. 15 shows that the choice of objective function has little influence on the average RMSE, as the group of bins for the different objective functions look very similar to each other. Neither does the choice of ResNet model order have a significant impact on the average RMSE, although the positions of the green triangles in the left column of Fig. 15 indicate that ResNet-6 has a slightly smaller mean value than ResNet-5 and ResNet-4. What influences the average RMSE the most is the choice of BS. All columns show that BS $= 1$ yields a smaller spread of average RMSE, but also a higher mean value. Models trained on BS $=$ 2, 3, or 4 achieve a similar spread of average RMSE for all three

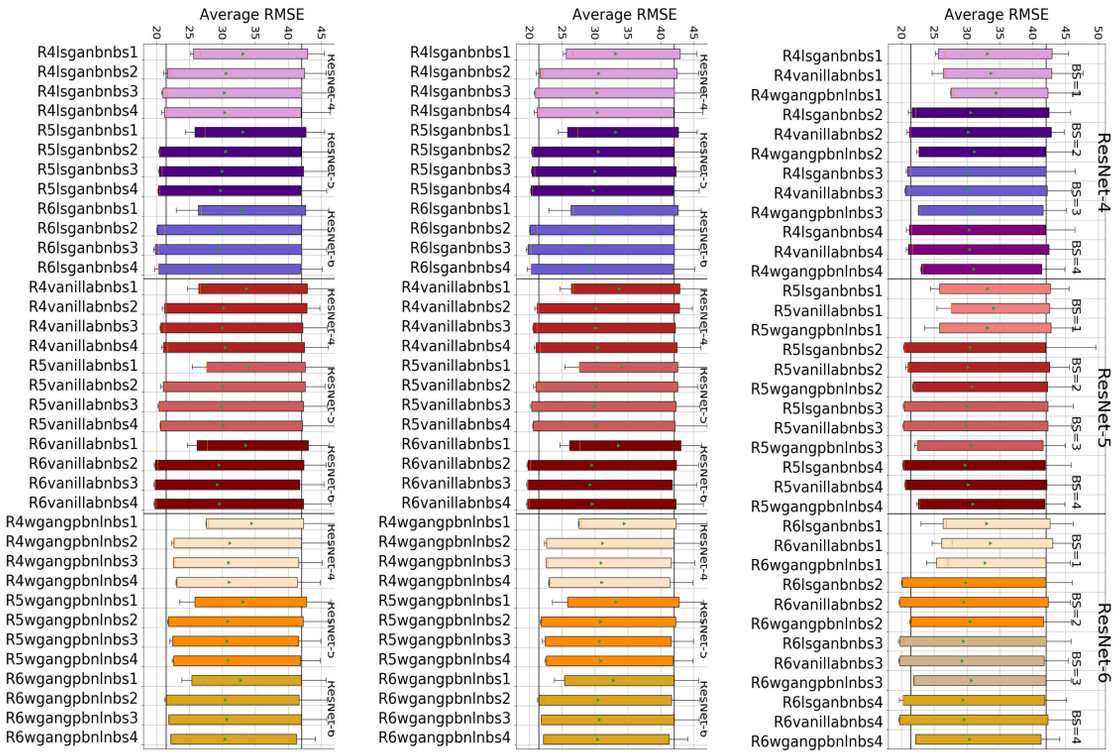


Fig. 15. Boxplot of average RMSE for models trained with all three objective functions, ResNet-4, 5, or 6, with BS varying between 1 and 4 and BN only. Green triangles indicate the mean value computed over the five folds, while orange vertical lines indicate the median. The two vertical black lines are arbitrary chosen reference lines to easen visual comparison. *Left column:* Grouped by BS in ascending order (top to bottom). *Middle column:* Grouped by ResNet in ascending order (top to bottom) in addition to BS in ascending order (top to bottom). *Right row:* Grouped by objective functions together with similar hyperparameters, sorted by ResNet and BS in ascending order (top to bottom).

objective functions, although the WGAN-GP shows slightly less spread. To summarize, the choice of normalization method and batch size has the largest impact on the RMSE, compared to other hyperparameters and the objective functions for the G or D networks. We recommend that BN should be chosen instead of IN, and that BS $= 1$ should be avoided.

D. Image Patch Generation

In this experiment, we evaluated the correspondence between generated image patches $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ to $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$. Table III lists the implementational choices for each cGAN variant used in this experiment; these are based on the validation described above; see Sections A2 and A3. Each cGAN variants were trained on a training set, while the test sets were kept aside. After training, we allowed the trained G network of each model to generate $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ image patches from Sentinel-1 image patches. These Sentinel-1 image patches were from the test set, and had therefore not been seen by the network during training. Since the test set also contains the corresponding target, i.e., $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ image patches, these were used to evaluate the generator's performance quantitatively and qualitatively.

Results: For each of the models in Table III, we select test patches, i.e., $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ and corresponding $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$, having the smallest and greatest RMSE ($\mathrm{Mg\,ha^{-1}}$) to investigate the worst and best case scenarios. The RMSE is computed over all pixels within the image test patch. Fig. 16 shows a qualitative comparison of the identified test patch with the smallest and greatest RMSE for the three models. The first row of Fig. 16 visualizes patches from the input domain, i.e., Sentinel-1, the middle row from the target domain, i.e., $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$, and the third row from the generated domain, i.e., $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$. Columns with caption *Min* indicate an image patch with the smallest RMSE for a specific model, while caption *Max* instead indicates an image patch with the largest RMSE. Columns (a) and (b) correspond to patches from the optimal Vanilla GAN, (c) and (d) from the optimal LSGAN, while (e) and (f) are from the optimal WGAN-GP. Quantitative comparisons of RMSE for the patches in Fig. 16 are shown in Table VIII.

As the same patch was identified as the easiest to translate by both the Vanilla GAN and the LSGAN models, these two cGAN variants must have learned similar translation dynamics between the input and output domains. See columns (a) and (c) of Fig. 16. The results provided in Section V-B3 also point to the same direction; overall, the Vanilla GAN and the LSGAN

TABLE VIII
LIST OF MINIMUM AND MAXIMUM RMSE FOR THE TEST IMAGE PATCHES SHOWN IN FIG. 16

Model	Min [$\mathrm{Mg\,ha^{-1}}$]	Max [$\mathrm{Mg\,ha^{-1}}$]
Vanilla GAN; ResNet-6 BN, BS=3	11.03	27.04
LSGAN; ResNet-6 BN, BS=3	10.92	27.34
WGAN-GP; ResNet-6 BN, BS=3	22.23	57.27

The listed models are from Section A4 and only differ from each other by the objective functions.

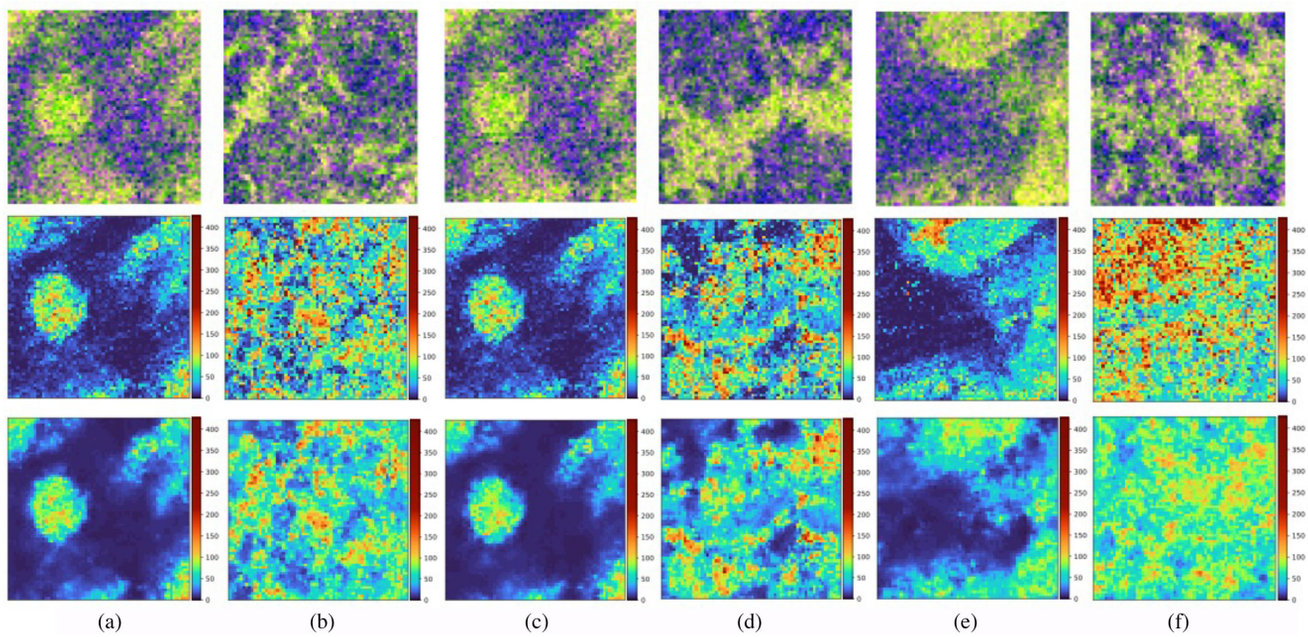


Fig. 16. First row: Sentinel-1 patches. Second row: Target image patches, i.e., ALS-based AGB predictions $\widehat{z}_{\boldsymbol{y}}$. Third row: Generated synthetic image patches, i.e., $\widehat{z}_{\boldsymbol{y}|\boldsymbol{x}}$. Columns (a) and (b): Vanilla GAN; (c) and (d): LSGAN; (e) and (f): WGAN-GP. Columns with caption *Min* and *Max*, respectively, refer to an image patch within the test set that achieves minimum and maximum RMSE, computed over all 64×64 pixels in the test patch ($\mathrm{Mg\,ha^{-1}}$).

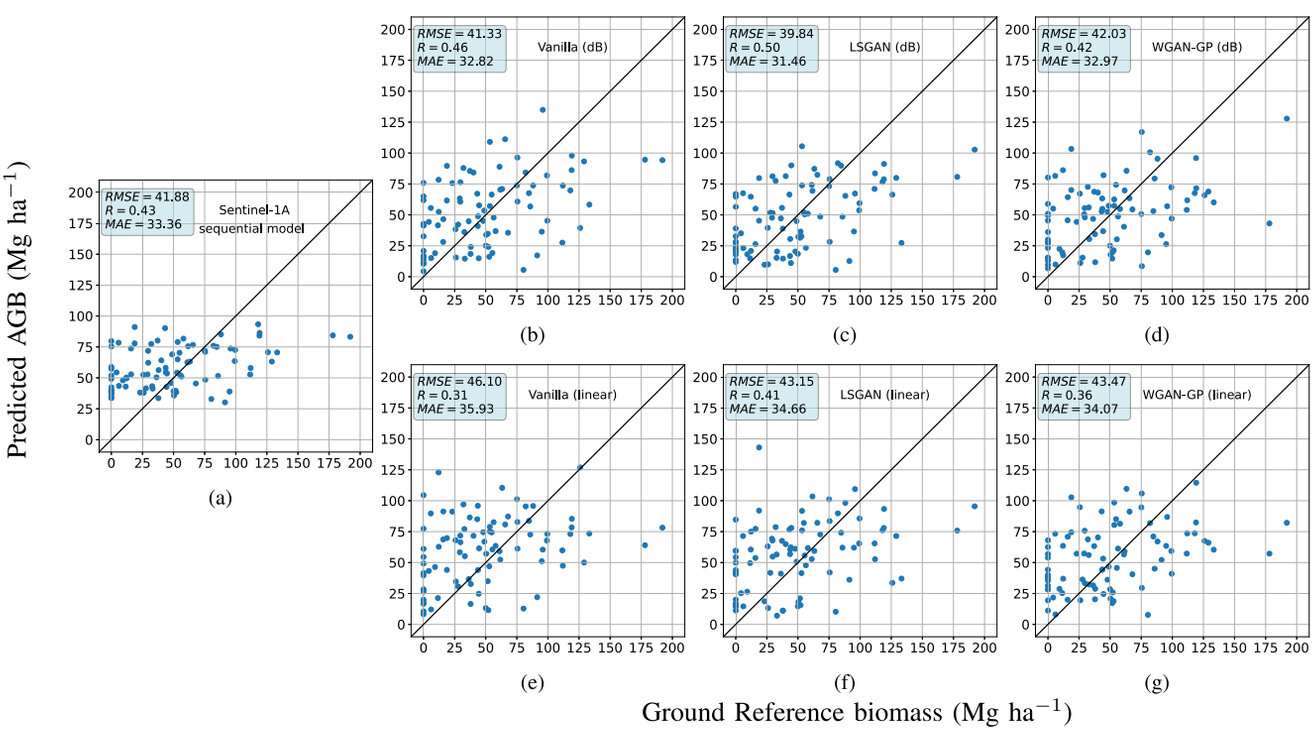


Fig. 17. Scatter plots between AGB ground reference data, z, and model-predicted AGB. Upper row: models trained with Sentinel-1 data on dB scale, i.e., (b) the proposed Vanilla GAN, (c) LSGAN, and (d) WGAN-GP models. Lower row, same models as above, but trained with Sentinel-1 on linear scale. (a) Model-predicted AGB values from the baseline sequential regression model given in (2) trained with Sentinel-1 data on linear scale. The black lines are reference lines indicating 100% correlation between z and AGB predictions. Units are in Mg ha^{-1}.

models perform more similar to each other and achieve higher accuracy than the WGAN-GP model. Table VIII clearly shows that the WGAN-GP model is the worse among the three, having an RMSE which is almost twice as high as for the Vanilla GAN or LSGAN. From Fig. 16, it can be noted that all three objective functions seem to be approximately equally appropriate for translating from $\mathcal{X}$ to $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}} \in \mathcal{Z}$ when patches from the two domains have similar appearance, but struggle when the $\mathcal{X}$ and $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}} \in \mathcal{Z}$ domains deviate from each other in appearance. Visually, all three objective functions generate synthetic patches which are somewhat more blurry than $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$ predictions. Blurriness is a known weakness with generative models, such as GANs [97], [98]. Several possible explanations to it exists; for example, that blurriness can be related to the transposed convolution upsampling method used in the second part of the G network. These upsampling methods affect the model's ability to correctly reproduce the spectral distribution in images, or to generate new images with sharp high-frequent components such as edges [98].

E. Comparison of Linear or dB-Scale SAR Input

In the Sentinel-1 processing workflow, we settled for, see Section IV-B, conversion to dB scale was only applied if the Sentinel-1 scene was used by the cGAN-based sequential models. The use of dB scale on the Sentinel-1 data for these models was decided by the results of the experiments provided in this section. We evaluated the impact of keeping the Sentinel-1 input data on linear scale versus to transform it to a logarithmic decibel (dB) scale. This was done by creating two versions of the Sentinel-1 dataset, where conversion to dB was applied to one of these. Except for this step, both Sentinel-1 datasets, referred to as Sentinel-1 linear or Sentinel-1 dB, were identically processed. For each of the optimal model implementations listed in Table III, we trained one model on the Sentinel-1 linear dataset and another on the Sentinel-1 dB dataset. This yielded six different possibilities to generate $\widehat{z}_{y|x}$, i.e., three different linear cGAN-based models and three different dB cGAN-based models. From each of these six models, we extracted $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$ predictions corresponding to the position of each AGB ground reference measurement z.

Results: We provide scatter plots of $\widehat{z}_{y|x}$ predictions and z in Fig. 17, where Fig. 17(b)–(d) represents results from the cGAN models trained on linear scale, while Fig. 17(e)–(g) represent corresponding results from the cGAN models trained on dB scale. For comparison with the baseline sequential Sentinel-1 model, we also show a corresponding scatter plot of it in Fig. 17(a) [it is the same figure as in Fig. 11(a)]. We also provide computed RMSE, R, and MAE in each scatter plot. Overall, Fig. 17 shows that R decreases while both RMSE and MAE increase if any of the cGAN models are trained on linear scale as compared to dB scale. We conclude that the conversion of calibrated σ_0 values to dB scale, which increases the dynamic range of the pixel values in the image, is advantageous for achieving more accurate image-to-image translation through the cGAN architecture.

F. Postcalibration of Sequential Models

Although the nonsequential Sentinel-1 model cannot predict AGB between 0 and 20.3 $\mathrm{Mg\,ha^{-1}}$, it still achieves a higher correlation coefficient R and a lower RMSE/MAE with respect to z than any of the proposed sequential models. One explanation can be that the nonsequential model had access to the ground reference data z during model fitting. By contrast, the sequential models were only using $\widehat{z}_y$ during model fitting and have therefore not been calibrated against z. In this experiment, we investigated if the accuracy of the sequential regression models could improve if we, after constructing the synthetic AGB prediction maps, calibrated them against z. As the original LSGAN model achieved the highest correlation with z, we focus the experiments in this section on this model and the baseline sequential Sentinel-1 model. Furthermore, for the LSGAN, we considered both Sentinel-1 data on linear scale and dB scale. Overall, we investigated five common calibration methods, i.e., *linear, exponential, gamma, nth-root,* and *logarithmic* calibration. Among these, we choose to show gamma and linear calibration results, as we obtained the best results with these methods.

Results: Fig. 18 shows results from the experiment with postcalibration of $\widehat{z}_{y|x}$, i.e., scatter plots between z and calibrated model-predicted AGB. To ease the comparison, we have provided some reference images, which are retrieved from the results presented in Section V, i.e., scatter plots for the ALS-based model [Fig. 18(a)], the nonsequential Sentinel-1-based model [Fig. 18(b)], LSGAN on dB scale Fig. 18(c)], LSGAN on linear scale [Fig. 18(f)], and the baseline sequential Sentinel-1 model [Fig. 18(i)]. We show results for the calibrated LSGAN model on dB scale using gamma calibration in Fig. 18(d) and linear calibration in Fig. 18(e). Furthermore, we show results for the calibrated LSGAN model on linear scale using gamma calibration in Fig. 18(g) and linear calibration in Fig. 18(h). Fig. 18(j) and (k) shows the results for the calibrated baseline sequential Sentinel-1 model on linear scale using gamma calibration [Fig. 18(j)] and linear calibration [Fig. 18(k)].

We note from the figure that the gamma and linearly calibrated models yield slightly lower or lower RMSE/MAE for all models included in the evaluation. For the LSGAN models, the gamma calibration reduces R slightly, while the correlation coefficient is unchanged for the linear calibration. For the baseline sequential model, R is unchanged for both the gamma and the linear calibration. Unfortunately, neither of the models achieve as high R and low RMSE/MAE as the nonsequential Sentinel-1-based model, nor the nonsequential ALS-based model. However, the LSGAN models, with or without calibration, can still predict 0 AGB, while neither of the baseline sequential models, with or without calibration, can produce such low AGB predictions. We conclude from this experiment that postcalibrating sequential AGB predictions against z can yield some modest improvements to higher accuracy. However, as these possible modest improvements come with the cost of applying an extra step to the prediction process, we choose to omit it in the results provided in Section V-B3.

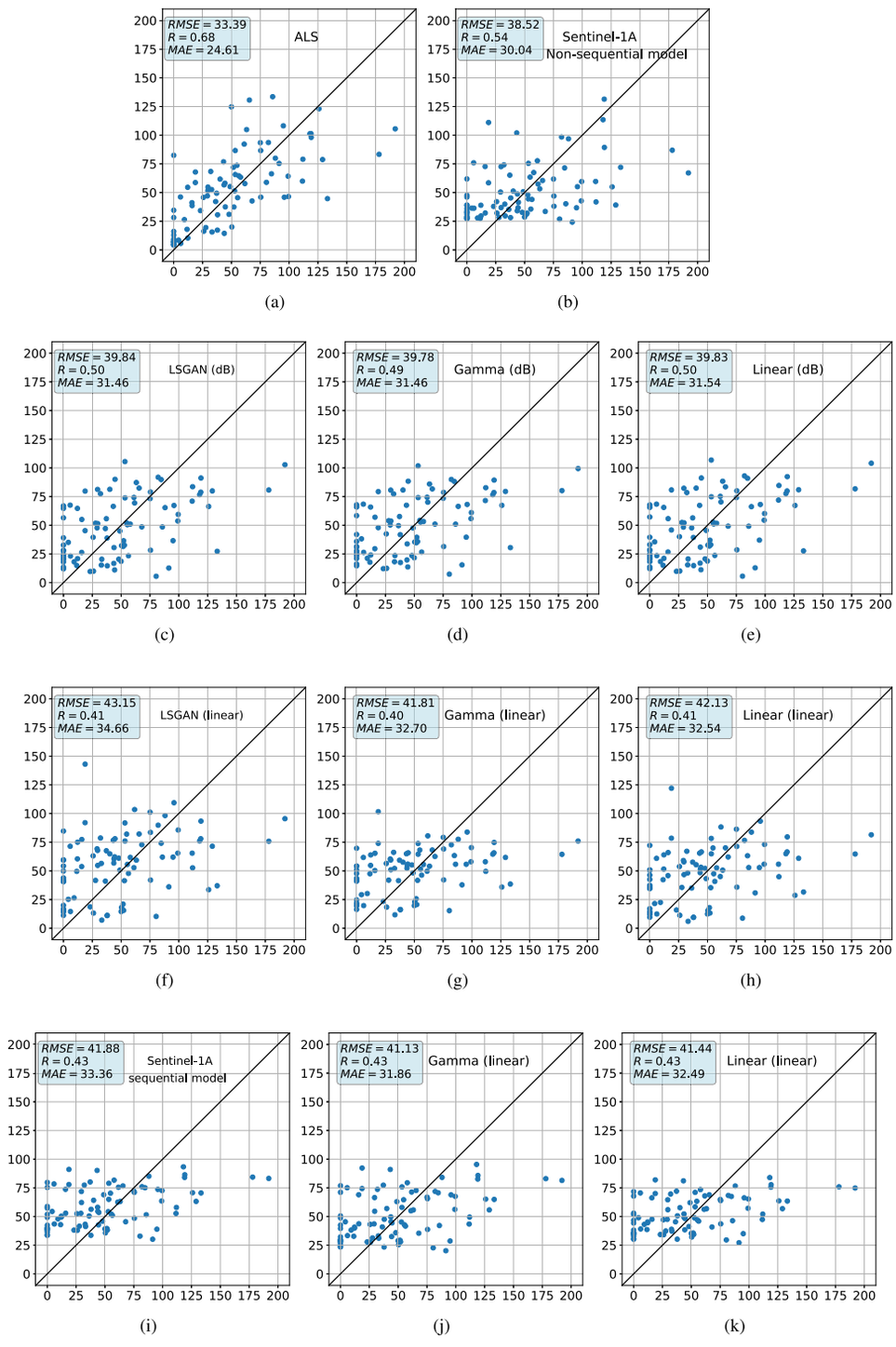


Fig. 18. Scatter plots between predicted AGB and ground reference AGB data, z. (a)–(c), (f), and (i) Reference images, corresponding to AGB predictions from the ALS-based regression model, the nonsequential Sentinel-1 model, the LSGAN model trained with dataset on dB scale, the LSGAN model trained with dataset on linear scale, and the baseline sequential Sentinel-1 model trained with dataset on linear scale. (d), (g), and (j) AGB predictions from respective model after calibration with gamma transform. (e), (h), and (k) Corresponding results after calibration with a linear transform. The black lines are reference lines indicating 100% correlation between z and predictions. Units are in Mg ha^{-1}.

ACKNOWLEDGMENT

We gratefully acknowledge employees of the Tanzania Forest Services Agency, Sokoine University of Agriculture, Norwegian University of Life Sciences, and the Swedish University of Agricultural Sciences (SLU) for participation in field work and provision of *in situ* measurements, RS data and derived AGB products. Special thanks to professor Håkan Olsson for providing access to ALS data acquired by SLU and for comments on the manuscript. S. Björk and S.N. Anfinsen further acknowledge discussions with and input from colleagues at the UiT Machine Learning Group.

REFERENCES

[1] Conference of the Parties, United Nations Framework Convention on Climate Change, Report of the Conference of the Parties on its sixteenth session, held 1788 in Cancun from 29 November to 10 December 2010 - Addendum - Part two: Action taken by the Conference of the Parties at its sixteenth session. UN Doc FCCC/CP/2010/7/Add.1 (15 March 2011) Decision 1/CP.16 ("The Cancun Agreements: Outcome of the work of the Ad Hoc Working Group on Long-term Cooperative Action under the Convention"). [Online]. Available: https://unfccc.int/documents/652

[2] L. T. Ene, E. Næsset, T. Gobakken, O. M. Bollandsaas, E. W. Mauya, and E. Zahabu, "Large-scale estimation of change in aboveground biomass in Miombo woodlands using airborne laser scanning and national forest inventory data," *Remote Sens. Environ.*, vol. 188, pp. 106–117, Jan. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425716304254

[3] S. Kaasalainen *et al.*, "Combining Lidar and synthetic aperture radar data to estimate forest biomass: Status and prospects," *Forests*, vol. 6, no. 12, pp. 252–270, Jan. 2015. [Online]. Available: http://www.mdpi.com/1999-4907/6/1/252

[4] A. Bombelli *et al.*, *Biomass–Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables*, Rome, Italy: FAO, 2009, pp. 1–18.

[5] T. Le Toan, G. Picard, J.-M. Martinez, P. Melon, and M. Davidson, "On the relationships between radar measurements and forest structure and biomass," *ESASP*, vol. 475, pp. 3–12, 2002.

[6] R. Hall, R. Skakun, E. Arsenault, and B. Case, "Modeling forest stand structure attributes using Landsat ETM data: Application to mapping of aboveground biomass and stand volume," *Forest Ecol. Manage.*, vol. 225, no. 1–3, pp. 378–390, Apr. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378112706000235

[7] L. T. Ene *et al.*, "Large-scale estimation of aboveground biomass in Miombo woodlands using airborne laser scanning and national forest inventory data," *Remote Sens. Environ.*, vol. 186, pp. 626–636, Dec. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425716303455

[8] E. Santi *et al.*, "The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas," *Remote Sens. Environ.*, vol. 200, no. 19, pp. 63–73, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003442571730353X

[9] S. M. Ghosh and M. D. Behera, "Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest," *Appl. Geography*, vol. 96, no. 1, pp. 29–40, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143622818303114

[10] M. A. Stelmaszczuk-Górska, M. Urbazaev, C. Schmullius, and C. Thiel, "Estimation of above-ground biomass over boreal forests in siberia using updated in situ, ALOS-2 PALSAR-2, and RADARSAT-2 data," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1550. [Online]. Available: https://www.mdpi.com/2072-4292/10/10/1550

[11] S. Sinha *et al.*, "Multi-sensor approach integrating optical and multi-frequency synthetic aperture radar for carbon stock estimation over a tropical deciduous forest in India," *Carbon Manage.*, vol. 11, no. 1, pp. 39–55, 2020. [Online]. Available: https://doi.org/10.1080/17583004.2019.1686931

[12] A. Debastiani, C. Sanquetta, A. Corte, N. Pinto, and F. Rex, "Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest," *Ann. Forest Res.*, vol. 62, no. 2, pp. 109–122, 2019. [Online]. Available: http://afrjournal.org/index.php/afr/article/view/1267

[13] L. L. Narine, S. C. Popescu, and L. Malambo, "Synergy of ICESat-2 and Landsat for mapping forest aboveground biomass with deep learning," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1503. [Online]. Available: https://www.mdpi.com/2072-4292/11/12/1503

[14] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1459. [Online]. Available: https://www.mdpi.com/2072-4292/11/12/1459

[15] L. Chen, Y. Wang, C. Ren, B. Zhang, and Z. Wang, "Optimal combination of predictors and algorithms for forest above-ground biomass mapping from Sentinel and SRTM data," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 414. [Online]. Available: https://www.mdpi.com/2072-4292/11/4/414

[16] E. Santi *et al.*, "Machine-learning applications for the retrieval of forest biomass from airborne p-band SAR data," *Remote Sens.*, vol. 12, no. 5, p. 804, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/5/804

[17] Y. Li, M. Li, C. Li, and Z. Liu, "Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms," *Sci. Rep.*, vol. 10, no. 1, Jun. 2020, Art. no. 9952, doi: 10.1038/s41598-020-67024-3.

[18] Y. Zhang, J. Ma, S. Liang, X. Li, and M. Li, "An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products," *Remote Sens.*, vol. 12, no. 24, 2020, Art. no. 4015. [Online]. Available: https://www.mdpi.com/2072-4292/12/24/4015

[19] N. Nuthammachot, A. Askar, D. Stratoulias, and P. Wicaksono, "Combined use of Sentinel-1 and Sentinel-2 data for improving above-ground biomass estimation," *Geocarto Int.*, vol. 37, no. 2, pp. 366–376, 2022, doi: 10.1080/10106049.2020.1726507.

[20] S. Zolkos, S. Goetz, and R. Dubayah, "A meta-analysis of terrestrial aboveground biomass estimation using Lidar remote sensing," *Remote Sens. Environ.*, vol. 128, pp. 289–298, Jan. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425712004051

[21] G. Galidaki *et al.*, "Vegetation biomass estimation with remote sensing: Focus on forest and other wooded land over the Mediterranean ecosystem," *Int. J. Remote Sens.*, vol. 38, no. 7, pp. 1940–1966, Apr. 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/01431161.2016.1266113

[22] E. Næsset *et al.*, "Mapping and estimating forest area and aboveground biomass in Miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision," *Remote Sens. Environ.*, vol. 175, no. 15, pp. 282–300, Mar. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425716300062

[23] M. Urbazaev *et al.*, "Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico," *Carbon Balance Manage.*, vol. 13, no. 1, Feb. 2018, Art. no. 5. doi: 10.1186/s13021-018-0093-5.

[24] M. A. Tanase, M. Santoro, J. de la Riva, F. Pé rez-Cabello, and T. Le Toan, "Sensitivity of X-, C-, and L-band SAR backscatter to burn severity in mediterranean pine forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3663–3675, Oct. 2010.

[25] M. Tanase, J. de la Riva, M. Santoro, F. Pérez-Cabello, and E. Kasischke, "Sensitivity of SAR data to post-fire forest regrowth in Mediterranean and boreal forests," *Remote Sens. Environ.*, vol. 115, no. 8, pp. 2075–2085, Aug. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425711001192

[26] M.-H. Phua *et al.*, "Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests," *Forest Ecol. Manage.*, vol. 406, pp. 163–171, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378112717307247

[27] S. Sinha, "Assessment of vegetation vigor using integrated synthetic aperture radars," in *Remote Sensing and GIScience*. Berlin, Germany: Springer, 2021, pp. 35–58.

[28] L. Chen, C. Ren, B. Zhang, Z. Wang, and Y. Xi, "Estimation of forest above-ground biomass by geographically weighted regression and machine learning with Sentinel imagery," *Forests*, vol. 9, no. 10, 2018, Art. no. 582. [Online]. Available: https://www.mdpi.com/1999-4907/9/10/582

[29] S. Vafaei *et al.*, "Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: A case study of the Hyrcanian forest area (Iran)," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 172. [Online]. Available: https://www.mdpi.com/2072-4292/10/2/172

[30] L. Chen, Y. Wang, C. Ren, B. Zhang, and Z. Wang, "Assessment of multi-wavelength SAR and multispectral instrument data for forest aboveground biomass mapping using random forest kriging," *Forest Ecol. Manage.*, vol. 447, pp. 12–25, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378112719304736

[31] L. Yang, S. Liang, and Y. Zhang, "A new method for generating a global forest aboveground biomass map from multiple high-level satellite products and ancillary information," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2587–2597, 2020.

[32] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[33] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, Dec. 2016, Art. no. 9. [Online]. Available: http://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6

[34] J. Zhang, W. Li, P. Ogunbona, and D. Xu, "Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1-38, Feb. 2019, doi: 10.1145/3291124.

[35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.

[36] C. S. Neigh *et al.*, "Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR," *Remote Sens. Environ.*, vol. 137, pp. 274–287, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425713002125

[37] S. Solberg, R. Astrup, T. Gobakken, E. Næsset, and D. J. Weydahl, "Estimating spruce and pine biomass with interferometric X-band SAR," *Remote Sens. Environ.*, vol. 114, no. 10, pp. 2353–2360, Oct. 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425710001513

[38] S. Englhart, V. Keuck, and F. Siegert, "Aboveground biomass retrieval in tropical forests – The potential of combined X- and L-band SAR data use," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1260–1271, May 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425711000216

[39] D. Wang *et al.*, "Estimating aboveground biomass of the mangrove forests on northeast Hainan Island in China using an upscaling method from field plots, UAV-LiDAR data and Sentinel-2 imagery," *Int. J. Appl. Earth Observation Geoinformation*, vol. 85, 2020, Art. no. 101986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243419306440

[40] A. T. Hudak *et al.*, "A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA," *Environ. Res. Lett.*, vol. 15, no. 9, Aug. 2020, Art. no. 095003, doi: 10.1088/1748-9326/ab93f9.

[41] D. Wang, B. Wan, P. Qiu, Z. Zuo, R. Wang, and X. Wu, "Mapping height and aboveground biomass of mangrove forests on Hainan island using UAV-LiDAR sampling," *Remote Sens.*, vol. 11, no. 18, 2019, Art. no. 2156. [Online]. Available: https://www.mdpi.com/2072-4292/11/18/2156

[42] O. Cartus, J. Kellndorfer, M. Rombach, and W. Walker, "Mapping canopy height and growing stock volume using airborne Lidar, ALOS PALSAR and landsat ETM," *Remote Sens.*, vol. 4, no. 11, pp. 3320–3345, 2012. [Online]. Available: https://www.mdpi.com/2072-4292/4/11/3320

[43] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.* – vol. 2, ser. NIPS'14. Cambridge, MA, USA:MIT Press, 2014, pp. 2672–2680.

[44] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse image synthesis for multiple domains," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8188–8197.

[45] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8107–8116.

[46] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4396–4405.

[47] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[48] X. Bao, Z. Pan, L. Liu, and B. Lei, "SAR image simulation by generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9995–9998. [Online]. Available: https://ieeexplore.ieee.org/document/8899286/

[49] Y. Xi *et al.*, "DRL-GAN: Dual-stream representation learning GAN for low-resolution image classification in UAV applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1705–1716, 2021.

[50] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 179, pp. 14–34, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271621001842

[51] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[52] D. Lu, Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran, "A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems," *Int. J. Digit. Earth*, vol. 9, no. 1, pp. 63–105, Jan. 2016. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/17538947.2014.990526

[53] M. C. Dobson, F. T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke, and N. Christensen, "Dependence of radar backscatter on coniferous forest biomass," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 412–415, Mar. 1992.

[54] T. Le Toan, A. Beaudoin, J. Riom, and D. Guyon, "Relating forest biomass to SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 403–411, Mar. 1992.

[55] J. Boudreau, R. F. Nelson, H. A. Margolis, A. Beaudoin, L. Guindon, and D. S. Kimes, "Regional aboveground forest biomass using airborne and spaceborne LiDAR in Qué bec," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3876–3890, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425708001995

[56] R. Nelson *et al.*, "Estimating Quebec provincial forest resources using ICESat/GLAS," *Can. J. Forest Res.*, vol. 39, no. 4, pp. 862–881, 2009.

[57] G. Sun, K. J. Ranson, Z. Guo, Z. Zhang, P. Montesano, and D. Kimes, "Forest biomass mapping from Lidar and radar synergies," *Remote Sens. Environ.*, vol. 115, no. 11, pp. 2906–2916, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425711001386

[58] O. W. Tsui, N. C. Coops, M. A. Wulder, and P. L. Marshall, "Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass," *Remote Sens. Environ.*, vol. 139, pp. 340–352, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425713002708

[59] H. A. Margolis *et al.*, "Combining satellite Lidar, airborne Lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America," *Can. J. Forest Res.*, vol. 45, no. 7, pp. 838–855, 2015. [Online]. Available: https://doi.org/10.1139/cjfr-2015-0006

[60] S. Saarela *et al.*, "Hierarchical model-based inference for forest inventory utilizing three sources of information," *Ann. Forest Sci.*, vol. 73, no. 4, pp. 895–910, 2016.

[61] S. Holm, R. Nelson, and G. Staahl, "Hybrid three-phase estimators for large-area forest inventory using ground plots, airborne Lidar, and space Lidar," *Remote Sens. Environ.*, vol. 197, pp. 85–97, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717301542

[62] T. Kauranne *et al.*, "LiDAR-assisted multi-source program (LAMP) for measuring above ground biomass and forest carbon," *Remote Sens.*, vol. 9, no. 2, 2017, Art. no. 154. [Online]. Available: https://www.mdpi.com/2072-4292/9/2/154

[63] Z. Shao, L. Zhang, and L. Wang, "Stacked sparse autoencoder modeling using the synergy of airborne LiDAR and satellite optical and SAR data to map forest above-ground biomass," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5569–5582, Dec. 2017.

[64] S. Saarela *et al.*, "Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and Landsat data," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1832.

[65] W. Qi, S. Saarela, J. Armston, G. Staahl, and R. Dubayah, "Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data," *Remote Sens. Environ.*, vol. 232, 2019, Art. no. 111283. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719303025

[66] D. Krige, "Two-dimensional weighted moving average trend surfaces for ore-evaluation," *J. South Afr. Inst. Mining Metall.*, vol. 66, pp. 13–38, 1966.

[67] D. Ao, C. O. Dumitru, G. Schwarz, and M. Datcu, "Dialectical GAN for SAR image translation: From Sentinel-1 to TerraSAR-X," *Remote Sens.*, vol. 10, no. 10, Oct. 2018, Art. no. 1597. [Online]. Available: http://dx.doi.org/10.3390/rs10101597

[68] M. Rezagholiradeh and M. A. Haidar, "REG-GAN: Semi-supervised learning based on generative adversarial networks for regression," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 2806–2810.

[69] G. Olmschenk, Z. Zhu, and H. Tang, "Generalizing semi-supervised generative adversarial networks to regression using feature contrasting," *Comput. Vis. Image Understanding*, vol. 186, no. C., pp. 1-12, Sep. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1077314219300955

[70] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[71] L. Dong *et al.*, "Application of convolutional neural network on Lei bamboo above-ground-biomass (AGB) estimation using WorldView-2," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 958. [Online]. Available: https://www.mdpi.com/2072-4292/12/6/958

[72] M. F. Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "SAR-to-Optical image translation based on conditional generative adversarial networks–Optimization, opportunities and limits," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2067 . [Online]. Available: https://www.mdpi.com/2072-4292/11/17/2067

[73] L. T. Luppino *et al.*, "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, 2021.
[74] E. Tomppo *et al.*, "A sampling design for a large area forest inventory: Case Tanzania," *Can. J. Forest Res.*, vol. 44, no. 8, pp. 931–948, 2014.
[75] ESA sentinel application platform (SNAP), computer software, version 8.0, 2021. [Online]. Available: https://step.esa.int/main/download/snap-download/
[76] F. Filipponi, "Sentinel-1 GRD preprocessing workflow," *Proceedings*, vol. 18, no. 1, 2019, Art. no. 11. [Online]. Available: https://www.mdpi.com/2504-3900/18/1/11
[77] J.-S. Lee, "Refined filtering of image noise using local statistics," *Comput. Graph. Image Process.*, vol. 15, no. 4, pp. 380–389, 1981.
[78] QGIS Development Team, "QGIS geographic information system. Open source geospatial foundation project," 2019. [Online]. Available: http://qgis.osgeo.org
[79] T. G. Gregoire, Q. F. Lin, J. Boudreau, and R. Nelson, "Regression estimation following the square-root transformation of the response," *Forest Sci.*, vol. 54, no. 6, pp. 597–606, 2008.
[80] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2813–2821.
[81] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[82] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[83] S. Solberg *et al.*, "Monitoring forest carbon in a Tanzanian woodland using interferometric SAR: A novel methodology for REDD," *Carbon Balance Manage.*, vol. 10, no. 1, pp. 1–14, Jun. 2015, doi: 10.1186/s13021-015-0023-8.
[84] M. L. Imhoff, "Radar backscatter and biomass saturation: Ramifications for global biomass inventory," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 511–518, Mar. 1995.
[85] J. Penman *et al.*, "Good practice guidance for land use, land-use change and forestry," in *Good Practice Guidance for Land Use, Land-Use Change and Forestry*. Hayama, Japan: IGES, 2003.
[86] J. Esteban, R. E. McRoberts, A. Fernández-Landa, J. L. Tomé, and E. Næsset, "Estimating forest volume and biomass and their changes using random forests and remotely sensed data," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1944.
[87] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 334–355.
[88] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019.
[89] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 63–79.
[90] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.
[91] J. W. Soh, G. Y. Park, J. Jo, and N. I. Cho, "Natural and realistic single image super-resolution with explicit natural manifold discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8114–8123.
[92] H. Liu, Y. Qian, X. Zhong, L. Chen, and G. Yang, "Research on super-resolution reconstruction of remote sensing images: A comprehensive review," *Opt. Eng.*, vol. 60, no. 10, 2021, Art. no. 100901.
[93] Y. Chang and B. Luo, "Bidirectional convolutional LSTM neural network for remote sensing image super-resolution," *Remote Sens.*, vol. 11, no. 20, 2019, Art. no. 2333. [Online]. Available: https://www.mdpi.com/2072-4292/11/20/2333
[94] W. Ma, Z. Pan, F. Yuan, and B. Lei, "Super-resolution of remote sensing images via a dense residual generative adversarial network," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2578. [Online]. Available: https://www.mdpi.com/2072-4292/11/21/2578
[95] E. Næsset, O. M. Bollandsaas, T. Gobakken, S. Solberg, and R. E. McRoberts, "The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data," *Remote Sens. Environ.*, vol. 168, no. 100, pp. 252–264, 2015.
[96] S. Björk, S. N. Anfinsen, E. Næsset, T. Gobakken, and E. Zahabu, "Generation of lidar-predicted forest biomass maps from radar backscatter with conditional generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 4327–4330. [Online]. Available: https://ieeexplore.ieee.org/document/9324296/
[97] M. Khayatkhoei and A. Elgammal, "Spatial frequency bias in convolutional generative adversarial networks," 2020, *arXiv:2010.01473*.
[98] R. Durall, M. Keuper, and J. Keuper, "Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7890–7899.

Sara Björk (Associate Member, IEEE) received the M.Sc. degree in applied physics and mathematics, in 2016, from UiT The Arctic University of Norway, Tromsø, Norway, where she is currently working toward the Ph.D. degree in science with Machine Learning Group, Department of Physics and Technology.

Since May 2022, she has been a System Developer with DevOps Team Applied Deep Learning, KSAT Kongsberg Satellite Services. Her research interests include image processing, machine learning, deep learning, and generative methods, with emphasis on information extraction from remote sensing data.

Stian Normann Anfinsen (Member, IEEE) received the M.Sc. degree in communications, control, and digital signal processing from the University of Strathclyde, Glasgow, U.K., in 1998, and the Cand. mag. and Cand. scient. degrees in physics and the Ph.D. degree in science from UiT The Arctic University of Norway, Tromsø, Norway, in 1997, 2000, and 2010, respectively.

Since 2014, he has been a Faculty Member with the Department of Physics and Technology, UiT, formerly with the Earth Observation Group and currently as a Professor with the Machine Learning Group. Since August 2021, he has been a Senior Researcher with NORCE Norwegian Research Centre. His research interests include statistical modeling, pattern recognition, and machine learning algorithms for image, graph, and time-series analysis in earth observation and energy analytics.

Erik Næsset received the M.Sc. degree in forestry and the Ph.D. degree in forest inventory from the Agricultural University of Norway, Ås, Norway, in 1983 and 1992, respectively.

He has played a major role in developing and implementing airborne LiDAR in operational forest inventory. He has been the Leader and Coordinator of more than 60 research programs funded by the Research Council of Norway, the European Union, and private forest industry. He has authored or coauthored around 250 papers in international peer-reviewed journals. His teaching includes lectures and courses in forest inventory, remote sensing, forest planning, and sampling techniques. His research interests include forest inventory and remote sensing, with particular focus on operational management inventories, sample surveys, photogrammetry, and airborne LiDAR.

Terje Gobakken received the M.Sc. degree in forestry and the Ph.D. degree in science from the Agricultural University of Norway, Ås, Norway, in 1995 and 2001 respectively. He is currently a Professor in forest planning with the Norwegian University of Life Sciences, Ås, Norway. He was with Norwegian National Forest Inventory, and has participated in compiling reports of emissions and removals of greenhouse gases from land use, land-use change, and forestry in Norway. He has coordinated and participated in a number of externally funded projects—including international projects funded by, for example, NASA and EU (FP6 and FP7), and has broad practical and research-based experience with development of big data and information infrastructures for forest inventory, planning, and decision support. He has authored or coauthored more than 190 peer-reviewed scientific articles related to forest inventory and planning in international journals.

10

Paper II

Simpler is better: spectral regularization and up-sampling techniques for variational autoencoders

Sara Björk, Jonas N. Myhre, and Thomas Haugland Johansen

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3778-3782, 2022

11

Paper III

Forest Parameter Prediction by Multiobjective Deep Learning of Regression Models Trained With Pseudo-Target Imputation

---

Sara Björk, Stian N. Anfinsen, Michael Kampffmeyer, Erik Næsset, Terje Gobakken, and Lennart Noordermeer

Submitted to IEEE Transactions on Geoscience and Remote Sensing, 2023

Forest Parameter Prediction by Multiobjective Deep Learning of Regression Models Trained With Pseudo-Target Imputation

Sara Björk, Stian N. Anfinsen, Michael Kampffmeyer, Erik Næsset, Terje Gobakken, and Lennart Noordermeer

***Abstract*—In prediction of forest parameters with data from remote sensing (RS), regression models have traditionally been trained on a small sample of ground reference data. This paper proposes to impute this sample of true prediction targets with data from an existing RS-based prediction map that we consider as pseudo-targets. This substantially increases the amount of target training data and leverages the use of deep learning (DL) for semi-supervised regression modelling. We use prediction maps constructed from airborne laser scanning (ALS) data to provide accurate pseudo-targets and free data from Sentinel-1's C-band synthetic aperture radar (SAR) as regressors. A modified U-Net architecture is adapted with a selection of different training objectives. We demonstrate that when a judicious combination of loss functions is used, the semi-supervised imputation strategy produces results that surpass traditional ALS-based regression models, even though Sentinel-1 data are considered as inferior for forest monitoring. These results are consistent for experiments on above-ground biomass prediction in Tanzania and stem volume prediction in Norway, representing a diversity in parameters and forest types that emphasises the robustness of the approach.**

***Index Terms*—Forest remote sensing, above-ground biomass (AGB), stem volume, synthetic aperture radar (SAR), Sentinel-1, airborne laser scanning (ALS), deep neural networks, regression modelling, U-Net, composite loss function, semi-supervised learning, pseudo-targets, imputation.**

I. Introduction

ACCURATE monitoring of forest above-ground biomass (AGB) is essential to better understand the carbon cycle. Vegetation biomass is, for example, a larger global storage of carbon than the atmosphere [1], [2]. Additionally, to monitor, measure and predict the amount of available AGB correctly is important for economic aspects, e.g. to estimate available raw materials or the potential for bioenergy [3], [4].

As the stem volume (SV) accounts for the highest proportion of biomass in each tree, typically 65-80% [5]–[7], AGB monitoring often focuses on the available SV. In other applications, the total amount of available biomass is of interest, which comprises stems, stumps, branches, bark, seeds and foliage [2], [3], [8]. Today, remote sensing (RS) data from radar, optical or airborne laser scanning systems (ALS) are commonly used together with a sparse sample of collected ground reference forest measurements to develop prediction models over larger areas and regions [9]–[11].

Satellite and airborne RS have become an important source of information about these forest parameters and others. Traditionally, AGB and SV prediction models use relatively simple statistical regression algorithms, such as multiple linear regression, or machine learning regression models like random forests or multilayer perceptrons (MLPs) [12]. These models are usually noncontextual, as they restrict the regressor information to the pixel that is being predicted and do not combine regressor and regressand information from neighbouring pixels, known as spatial context.

Remote sensing is commonly used to infer forest parameters on spatial scales that are coarser than the pixel size, for instance on stand level. Hence, there is no formal reason to avoid the use of contextual information and one should select the method that provides the highest accuracy on the desired scale. This motivates the use of deep learning (DL) and convolutional neural networks (CNNs), whose popularity hinges on their efficient use of spatial context and the inference accuracy obtained by these highly flexible function approximators. The ability of CNNs to exploit spatial patterns was also pointed out in a recent review [13] as an explanation as to why CNN are particularly suitable for RS of vegetation.

A recent review [14] of DL methods applied to forestry concludes that these are in an early phase, although some work has emerged. We build our proposed method on Björk *et al*'s sequential approach to forest biomass prediction [12], which uses a conditional generative adversarial network (cGAN) to generate AGB prediction maps by using synthetic aperture radar (SAR) as regressors and AGB predictions from ALS as the regressand. Their regression approach consists of two models that operate in sequence to provide more target data for training the model that regresses on SAR data. This implies that the first regression model learns the mapping between a small set of ground reference data and RS data from a sensor known to provide a high correlation with the response variable. ALS data are suitable for this purpose [8], [15], but are expensive to acquire. Hence, the second model in the sequence establishes a relationship between the ALS-derived prediction map, as a surrogate for the ground reference data, and RS data from a sensor that offers large data amounts at

Manuscript received ; revised .

S. Björk is with the Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway and the Earth Observation Team, Kongsberg Satellite Services, 9011 Tromsø, Norway (e-mail: sara.bjork@ksat.no).

S. N. Anfinsen is with the Earth Observation Group, NORCE Norwegian Research Institute, 9019 Tromsø, Norway and the Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

Michael Kampffmeyer is with the Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

E. Næsset, T. Gobakken and L. Noordermeer are with Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, 1432 Ås, Norway.

low cost, namely the Sentinel-1 SAR sensors.

This paper preserves some of the principal ideas from [12]: The first is to train the regression model on an ALS-derived prediction map of the target forest parameter to increase the amount of training data. The motivation is that the small amount of ground reference data used to train conventional models limits their ability to capture the dynamics of the response variable, as demonstrated in [12]. The second is to carry forward the use of CNNs to leverage their exploitation of contextual information, their flexibility as regression functions, and their demonstrated performance in other applications.

At the same time, we make several new design choices to improve on the previous approach and remedy its weaknesses: Firstly, the sequential model is replaced by an approach where ground reference data are imputed with data from the ALS-derived prediction map. In practice, this is done by inserting the sparse set of true targets into the dense map of pseudo-targets. By letting these data sources together form the prediction target, the SAR-based prediction model can be trained simultaneously on ground reference data and the ALS-derived prediction map in a problem setting that we frame as semi-supervised learning; A second improvement is that we replace or combine the generative adversarial network (GAN) loss used in [12] with a pixel-wise error loss and a frequency-aware spectral loss. This modification is motivated by an emerging awareness that the GAN loss used by the Pix2Pix [16] model employed in [12] may be well suited to preserve perceptual quality and photo-realism, which is required in many image-to-image translation tasks, but is less appropriate for the regression task that we address.

This paper has a stronger technical and methodological focus than [12] and emphasises the method's ability to handle different tasks and cases: It demonstrates the proposed regression framework both on AGB prediction in dry tropical forests in Tanzania and on SV prediction in boreal forests in Norway, representing different parameters and very different forest types. Another difference is that the ALS-derived SV predictions used as pseudo-targets in the Norwegian dataset cover spatially non-contiguous forest stands, and is not a wall-to-wall prediction map. We have adapted the CNN-based regression algorithm for use with such data by implementing masked computation of the loss functions.

In summary, we make the following contributions:

1) We develop a method that enables us to train contextual deep learning models to predict forest parameters from C-band SAR data from the Sentinel-1 satellite.
2) We enable the CNN-based regression model to use target data that consist of spatially disjoint polygons, thereby showing that it can be trained on complex datasets that arise in operational forest inventories.
3) By testing the method on AGB prediction in Tanzania and SV prediction in Norway, we demonstrate that it can handle different forest parameters and forest types.
4) We investigate an established consensus from the image super-resolution (SR) field about the trade-off between reconstruction accuracy and perceptual quality. For this purpose, we perform an ablation study of composite cost functions, including the GAN loss, a pixel-wise loss, and a recently proposed frequency loss.
5) We demonstrate state-of-the-art prediction performance on datasets from Tanzania and Norway. Notably, we show that a deep learning model with C-band SAR data as input supercedes a conventional ALS-based prediction model after it has been trained on ground reference data imputed with ALS-derived predictions of the forest parameters.

The remainder of this paper is organised as follows: Section II reviews published research on related topics in deep learning applied to forest parameter prediction and other topics relevant to the proposed method. Section III presents the datasets used in this work. Section IV details the proposed approach and describes how we facilitate the imputation of pseudo-targets for regression modelling, enabling the CNN model to learn from continuous and discontinuous target data using a variety of loss functions. Experimental results are provided in Section V and discussed in Section VI. Finally, Section VII concludes the paper.

II. Related work

Björk *et al.* showed in a precursor of this paper [12] that the popular cGAN architecture *Pix2pix* [16] can be used in the forestry sector to predict AGB from Sentinel-1 data by training it on ALS-derived prediction maps. Their work inspired [17] to also exploit ALS-derived AGB prediction maps and cGANs to predict AGB from multispectral and radar imagery and to quantify aleatoric and epistemic uncertainty. Despite apparent similarities, the current paper distinguishes itself from both [12] and [17] in many ways. The differences from [12] are discussed in Section I when listing the contributions of the paper. Just like [12], Leonhardt *et al.* [17] train their regression network with adversarial learning through a cGAN architecture, but pretrain the generator with a mean square error (MSE) loss to find a proper initialisation. Notably, their final goal is not point prediction in the MSE sense or according to similar metrics, but to develop probabilistic methods for AGB prediction that quantify uncertainty.

Another example of deep learning applied to AGB prediction is Pascarella *et al.* [18], who show that a traditional U-Net [19] trained with a pixel-wise error loss can be used as a regression model to predict AGB from image patches of optical Sentinel-2 data. Compared to [18], we focus on utilising data from the Sentinel-1 radar sensor that, as opposed to the optical Sentinel-2 sensor, can acquire data both at night and under cloudy conditions and is therefore a more reliable source of data.

Besides these examples, the literature on deep learning for regression modelling of forest parameters is sparse. This is also pointed out in the review of the use of CNNs in vegetation RS conducted by Kattenborn *et al.* [13]. It found that only 9% of the studies surveyed focused on regression modelling and only 8% were specifically related to forestry and forest parameter retrieval, such as biomass prediction. A recently published review by Hamedianfar *et al.* [14] attributes this literature gap to the challenge of acquiring the large amounts of target data needed to train accurate contextual CNN models

for forest. This has been a main motivation for using pseudo-targets from existing prediction maps to train our SAR-based prediction models. For further inspiration, we have had to look to alternative topics in the literature.

Another image processing task that has inspired us to consider alternative loss functions and combinations of these is image super-resolution (SR). Single-image SR techniques are trained in a similar fashion as regression models: A full-resolution image is often used as the prediction target and a reduced resolution version of it as predictor data (see e.g. [20]), which renders the problem a prediction task that resembles the one in regression. Both the regression and the single-image SR task can be solved with generative models, but it is noteworthy that the literature identifies the SR task as an attempt to achieve two conflicting goals: It should produce images with high perceptual quality, meaning that they should appear natural and realistic. At the same time, it should reconstruct the underlying truth, that is, the high-resolution version of the input image, as closely as possible [20]–[24].

The SR literature associates GAN losses and adversarial training with the perceptual quality criterion, as these enforce realistic fidelity and crispness in the generated image. This is achieved at the expense of accurate reconstruction in the MSE sense, since the generator module of the GAN effectively learns to hallucinate the kind of spatial pixel configurations that fools the discriminator module, but does not consider pixel-wise reconstruction. On the other hand, pixel-wise losses such as error measures based on the L_1 and L_2 norm naturally reduce the reconstruction error, but lead to a blurry appearance of the generated image that is not realistic [20], [21].

This has made us realise that although the Pix2Pix model has established itself as a preferred standard model in image-to-image translation, its GAN loss and adversarial learning approach may be better suited for generative tasks where the result must be visually credible. This is not a concern in the regression of biophysical parameters, where regression performance in terms of root mean square error (RMSE), mean absolute error (MAE) or similar metrics is used to evaluate and rank methods. When training such models, one should therefore consider other loss functions or composite loss functions that support the relevant aspects of the regression task. The SR literature exemplifies ways of combining different loss functions, both regarding which losses to select and how they should interact [20], [21]. For instance, different losses can be used sequentially in pretraining and fine-tuning, or they can be used simultaneously as a composite loss.

Although perceptual quality is not of the essence for prediction maps of forest parameters, it may still be worth including loss functions that promote sharpness and visual information fidelity as part of a composite loss. One particular class of loss functions we find interesting to investigate is frequency-aware losses. Their aim is to preserve the high-frequency content of the image, which can e.g. be related to forest boundaries, structure and texture. These have not previously been utilised in forest applications, and to a limited extent in SR, but relevant work is found in the more general computer vision literature, where issues referred to as Fourier spectrum discrepancy, spectral inconsistency, frequency bias or spectral bias have gained a lot of attention [25]–[31]. These terms relate to CNN-based generative models' lack of ability to capture the image distribution's high-frequency components, leading to blurriness and low perceptual quality.

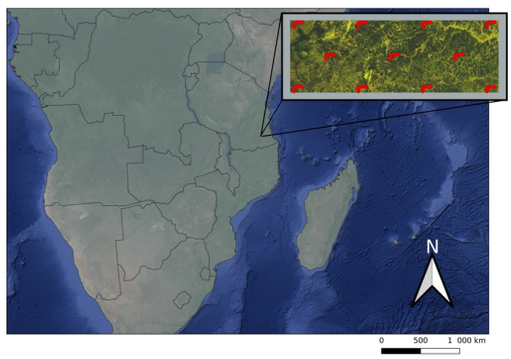


Fig. 1. The location of the Tanzanian dataset, represented by Sentinel-1A image data covering the AOI overlaid with ground reference data shown as red L-shaped clusters of ground plots. Figure from [12].

Some claim that the spectral bias is caused by the up-sampling method, e.g. transposed convolutions, used by the generator network [26], [27], [31]. Thus, changing the up-sampling method in the last layer of the generator network has been suggested [27]. However, Björk *et al.* [29] claim that changing the up-sampling procedure in the last layer from transposed convolution to e.g. nearest-neighbour interpolation followed by standard convolution gives ambiguous results. Chen *et al.* [25] argue that the down-sampling modules in the discriminator network of the GAN are the issue, resulting in a generator network that lacks an incentive from the discriminator to learn high-frequency information of the data. However, more recent work [28] proves that the frequency bias must be rooted in the GAN's generator and not the discriminator. Hence, there has been a focus on modifying the generative training objective by incorporating a spectral or frequency-aware loss with the traditional spatial loss during training [26], [29], [30].

The observations and lessons from the precursor paper [12] and from the literature on SR and generic generative models prompts us to investigate if model accuracy improves when we combine loss functions and whether pretraining of the model is enough or if we can increase model performance with a fine-tuning phase. Among the loss functions we combine is a newly proposed frequency-aware loss: the simple but promising FFT loss [29]. It has been shown to perform better than other more complex frequency-aware losses [26], [30] on experiments where it was used to train a generative variational autoencoder (VAE) [32]. As the FFT loss has previously only been evaluated on VAEs with images from common benchmark datasets [29], we contribute with new insight into its behaviour when employed for other models and tasks.

III. Study areas and datasets

This section introduces the datasets used throughout this work, i.e. the ground reference target data, the ALS-derived

prediction maps of AGB and SV, and the SAR data from the Sentinel-1 sensors. The ALS-derived prediction maps will interchangeably be referred to as the pseudo-target datasets, while the ground reference data are also referred to as field data, data from the field plots, or true prediction targets. The AGB dataset comes from the Liwale district in Tanzania. The SV datasets are from three regions in the southeast of Norway: Nordre Land, Tyristrand and Hole.

For Tanzania, both the field data and the ALS data were acquired in 2014, as described in [9] and Section III-A1. The Sentinel-1A satellite was launched in April 2014 and only one single Sentinel-1A scene acquired in September 2015 was found to comply with our requirements, meaning that it covers one of Liwale's two yearly dry seasons and is close enough in time to the field inventory and the ALS campaigns in Tanzania. For Norway, the acquisition of the ALS data in 2016 and the field inventory in 2017 (see [10] and Section III-A3) implies that more Sentinel-1 data are available. Thus, the models we develop for the Norwegian test sites utilise a temporal stack of Sentinel-1A and Sentinel-1B scenes from July 2017.

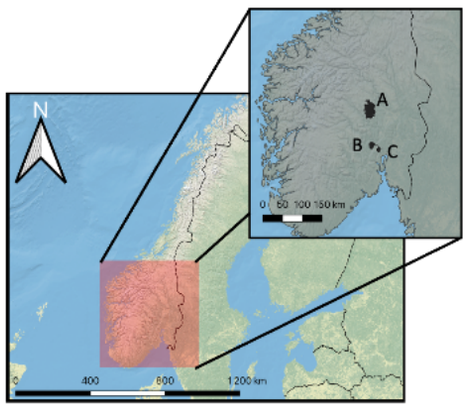


Fig. 2. Location of the regions Nordre Land (A), Tyristrand (B) and Hole (C) in the Norwegian dataset.

A. Study area and dataset description

This section briefly describes the Tanzanian and Norwegian study areas, including the ground reference data and related ALS-derived prediction maps. The interested reader is referred to [9] and [10], respectively, for in-depth descriptions of the ground reference data and the ALS-derived prediction maps.

1) Tanzanian study area: This work focuses on the same study area as [12], i.e. the Liwale district in the southeast of Tanzania (9°52'-9°58'S, 38°19'-38°36'E). The area of interest (AOI) is a rectangular region with a size of 11.25×32.50 km (WGS 84/UTM zone 36S). Fig. 1 shows the location of the AOI in Tanzania and the distribution of the 88 associated field plots. These field plots were collected within 11 L-shaped clusters, each containing eight plots, as seen in Fig. 1.

The field work was performed in January-February 2014, and a circular area of size 707 m^2 represents each sample plot on the ground, i.e. they have a radius of 15 m. We refer to [33] for a description of the national level sample design in Tanzania, while [9], [34], [35] explain how data from the field work are used to develop large-scale AGB models. Generally, the miombo woodlands of the AOI are characterised by a large diversity of tree species. Measured AGB from the field work ranged from 0 to 213.4 $\mathrm{Mg\,ha^{-1}}$ [9] with a mean and standard deviation of $\mu = 51.3$ and $\sigma = 45.6 \mathrm{Mg\,ha^{-1}}$.

2) Tanzanian ALS-predicted AGB data: We follow [12] and use the same ALS data from the Liwale AOI, which was acquired in 2014. We refer to [9] for details on the ALS flight campaign, ALS data processing, and the match-up of ALS data with ground reference AGB data from the field plots. After model fitting, the ALS-based AGB model was in [9] used to infer a wall-to-wall prediction map for the whole AOI in Liwale. The wall-to-wall map is represented as a grid with square pixels of size 707 m^2. We have gained access to this prediction map and will use it as pseudo-targets to train contextual CNN models for AGB predictions based on Sentinel-1 SAR data.

3) Norwegian study area: The Norwegian study area consists of three regions shown in Fig. 2 and referred to as Nordre Land (A), Tyristrand (B) and Hole (C). All field work was performed during the summer and fall of 2017, initially resulting in 386 circular field plots of shape 250 m^2 distributed over the three regions. We refer to [10] for a description of the sampling design and related data properties.

Of the original 386 field plots used for modelling stem volume, a total of 122 plots were not located within polygons of forest stands delineated in the inventories, and thus fell outside the spatial extent of the ALS-predicted SV datasets. We therefore excluded these plots from the analysis. In Table I, the column *No. of plots (after filtering)* indicates the number of field plots included in the current study. The remaining entities of Table I, such as geographical coordinates, inventory size, field inventory information and distribution of the dominant tree species in each region, are sourced from [10].

In Nordre Land, ground reference values of SV ranged from 33.7 to 659.2 $\mathrm{m^3\,ha^{-1}}$ with a mean and standard deviation of $\mu = 252.7$ and $\sigma = 145.5 \mathrm{m^3\,ha^{-1}}$. In Tyristrand it ranged from 56.1 to 513.3 $\mathrm{m^3\,ha^{-1}}$ with $\mu = 212.6$ and $\sigma = 96.9 \mathrm{m^3\,ha^{-1}}$, while in Hole it ranged from 29.5 to 563.9 $\mathrm{m^3\,ha^{-1}}$ with $\mu = 253.4$ and $\sigma = 125.8 \mathrm{m^3\,ha^{-1}}$.

4) Norwegian ALS-predicted SV data: The ALS flight campaigns were performed in 2016 for all three regions of Norway. We refer to [10] for a description of how the ALS data were processed, the formulation of the nonlinear local prediction models and the match-up of ALS-derived predictions with ground reference data. After model fitting, maps of SV predictions were generated for all three regions, limited to areas where the forest height exceeded 8-9 meters. We refer to these as the ALS-derived SV prediction maps. In all regions, predictions were made for square pixels of size 250 m^2, i.e. $15.8\ \mathrm{m} \times 15.8\ \mathrm{m}$ on the ground. The ALS-derived SV is given in units of $\mathrm{m^3\,ha^{-1}}$.

TABLE I
CHARACTERISTICS OF EACH OF THE THREE NORWEGIAN REGIONS INCLUDED IN THIS WORK. ALL ENTITIES, EXCEPT FOR COLUMN *No. of plots (after filtering)*, WHICH REFERS TO THE FIELD PLOTS THAT ARE USED FOR THIS WORK, ARE BORROWED FROM [10].

Region	Name	Geographical coordinates	Inventory size (km^2)	No. of plots (after filtering)	Proportion spruce	Proportion pine	Proportion deciduous
A	Nordre Land	60°50´N, 10°85´E	490	136	74%	23%	3%
B	Tyristrand	60°6´N, 10°20´E	60	77	15%	80%	5%
C	Hole	60°1´N, 10°20´E	45	51	89%	4%	7%

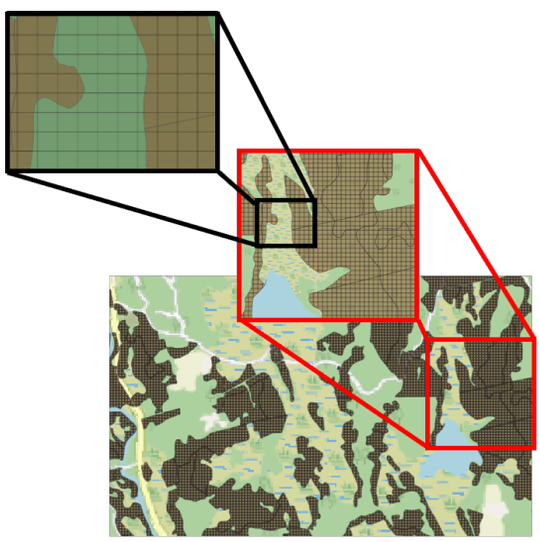


Fig. 3. Small section of the ALS-derived SV prediction map from Nordre Land. SV has been predicted in the brown areas. The lattice represents the common pixel grid of the SAR predictor data and the rasterised SV prediction map. The original prediction map is obtained in vector format, with one SV prediction per polygon and multiple polygons per grid cell. The rasterisation process with merging of polygons is described in detail in the text.

B. Postprocessing of the ALS-derived prediction maps

The ALS-derived prediction maps have been obtained as vector data in polygon format stored as shapefiles. These must be converted to raster data in order to be used as training data for CNN models. This conversion is straightforward for the Tanzania datasets, where all polygons are square and have the same areal coverage. Hence, we map project and sample the SAR data such that the SAR pixels coincide with the polygons of the AGB prediction map.

The process for the Norway dataset is more complicated. Fig. 3 shows a section of the ALS-derived SV prediction map retrieved in the Nordre Land municipality. Brown areas show where SV predictions are available, whereas the background (other colours) is retrieved from OpenStreetMap [36]. An overlaid lattice of square grid cells can be seen at all zoom levels of Fig. 3. This lattice represents two things: Firstly, it contributes to the delineation of the polygons in the SV prediction map. In this dataset, SV has been predicted for polygons of varying size and shape, that are delimited by: 1) the grid cells of the lattice, as mentioned above; 2) the commercial forest boundaries that enclose the brown areas; and 3) curves within the brown areas that mark internal forest boundaries and subdivide different forest areas. These are seen at all zoom levels of the figure. Secondly, the lattice coincides with the map grid of the SAR data, since we have map projected and resampled the SAR images to align their map grid with the lattice of the SV polygons. Hence, the lattice grid is identical to the pixel grid we want for our training dataset.

In summary, SV is only predicted in brown areas. Each prediction is associated with a polygon, which can be square if it is only delimited by the lattice and coincides with a lattice grid cell. It can also be of irregular shape and size, if a forest perimeter or an internal forest boundary delimits it. Each polygon is assigned a stem volume, V, and an areal coverage, A. Some of the square lattice cells are fully covered by one or more polygons, while others are only partly covered. Some lattice cells contain one polygon, while others contain two or more. We refer to this as a multipolygon format, as every lattice grid cell potentially contains multiple polygons.

The multipolygon dataset must be rasterised into a target dataset with the same pixel grid as the SAR predictor data. This means that all polygons within a lattice grid cell must be merged, and the grid cell must be assigned a single SV value and the associated areal coverage. The predicted SV contributed by all intersecting multipolygons is computed as

$$V_{merged} = \sum_{i=1}^{n} V_{mp(i)}, \tag{1}$$

where *mp(i)* indicates multipolygon number i and n is the number of multipolygons in a grid cell. Simultaneously, the total areal coverage is computed as:

$$A_{merged} = \sum_{i=1}^{n} A_{mp(i)}. \tag{2}$$

The described merging process guarantees that each grid cell is assigned a unique SV, but this value does not necessarily represent a full grid cell of 250 m^2. To quality assure the SV dataset, we remove all SV predictions with less than 40% areal grid cell coverage. This threshold is chosen heuristically to accommodate all three regions, as this removes less than 12% of the Nordre Land and Tyristrand dataset and less than 10% of the Hole dataset. The remaining SV prediction dataset is deemed suitable for the training of CNN regression models. All postprocessing steps are applied using QGIS [37].

C. SAR data

Low data cost can sometimes be crucial for developing forest parameter monitoring systems suitable for commercial

or operational use. This paper utilises SAR data from the freely available Sentinel-1 sensors, which also offer short revisit time and good coverage for the areas of interest. The SAR images are dual-polarisation (VV and VH) C-band scenes acquired in a high-resolution Level-1 ground range detected (GRD) format with a 10 m pixel size. The SAR data was downloaded from Copernicus Sentinel Scientific Data Hub[1].

For the AOI in Tanzania, we use a single scene acquired on 15 September 2015, as this is the only available Sentinel-1 product that covers the AOI at a time close to the acquisition of the ALS data and during one of Liwale's two yearly dry seasons. The latter criterion implies that the radar signal achieves sufficient sensitivity to dynamic AGB levels.

We utilise data from the Sentinel-1A and -1B satellites for the three Norwegian regions. Since the field work for the three Norwegian regions was performed during the summer and fall of 2017, we decided to create temporal stacks of Sentinel-1 -scenes from July 2017 for each of the three regions.

D. SAR data processing and preparation of datasets

The Sentinel-1 GRD product in the Tanzanian dataset was processed with the ESA SNAP toolbox [38] following the workflow described in [12].

The Sentinel-1 GRD products in the Norway dataset have been processed with the GDAR SAR processing software at NORCE Norwegian Research Institute. They are geocoded with a $10\ \mathrm{m} \times 10\ \mathrm{m}$ digital elevation model to the same map projection as the ALS-derived SV prediction map and resampled to a pixel resolution of $15.8\ \mathrm{m}$ to match the $250\ \mathrm{m}^2$ grid cells of the prediction map. Since [12] showed that it is more advantageous to train CNN-based prediction models with Sentinel-1 intensity data on decibel (dB) scale, the stacks of Sentinel-1 scenes for the Norwegian regions are converted to dB format. The final Sentinel-1 products for the Norwegian regions contain nine features that were extracted from the Sentinel-1 time series: NDI, mean(VV), mean(VH), min(VV), min(VH), max(VV), max(VH), median(VV), median(VV). NDI denotes the normalised difference index feature, a normalised measure of how much the measured backscatter differs in VV and VH. It is computed as

$$NDI = (VV - VH)/(VV + VH). \tag{3}$$

IV. Methodology

This section describes the proposed methodology to train contextual CNN models for forest parameter prediction. We describe the semi-supervised approach and how training, test and validation datasets are created for each region. In general terms, we introduce the CNN models we use in our work and describe the changes proposed to improve on the performance obtained in [12]. This section focuses on a semi-supervised learning strategy where we impute the sparse reference data with data from ALS-derived prediction maps to increase the amount of training data and to create a dataset that allows us to train CNN models. It also explains the multiobjective training approach, which exploits composite loss functions with varying objectives in the pretraining and fine-tuning stages.

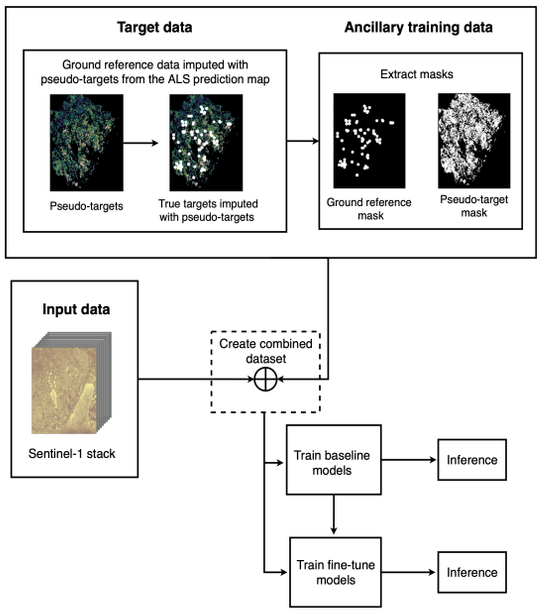


Fig. 4. Overall workflow for dataset generation, model training and inference to create prediction maps, displayed with image data from the Tyristrand dataset. True targets (white circles) have been magnified for illustrative purposes.

1) Overview: The framework of the proposed method is illustrated in Fig. 4. Initially, the ground reference data, also known as the true prediction targets, are imputed with the ALS-derived prediction map, also called the pseudo-target dataset. Then two binary masks are created, one indicating the pixel positions of the true targets and the other indicating pixels where pseudo-target data are available. The two masks are referred to as ancillary training data. They enable the CNN to learn from discontinuous pseudo-target data and boost learning in regions where ground reference data are available. When the pseudo-target data are spatially continuous and have the same extent as the predictor data, the pseudo-target mask will have a constant value of one. The imputed target dataset and the two masks are combined with regressor data from the Sentinel-1 sensor. See Section IV-B and Fig. 5 for details. Fig. 4 shows that baseline models are pretrained as an initial training step. Following the pretraining stage, fine-tuning may be applied to the baseline CNN models with a composition of different losses. Inference, i.e. production of SAR-derived prediction maps, is done with the resulting models[2].

A. Imputing ground reference data with pseudo-targets

The cGAN-based models developed in [12] for SAR-based regression trained on ALS-derived prediction maps could not compete with the conventional ALS-based regression model in terms of prediction accuracy. We argue that this is because the cGAN model is not trained on the true prediction targets and therefore inherits too much of the uncertainty in the ALS-based prediction maps. By contrast, the conventional ALS

[1]See https://scihub.copernicus.eu/dhus/#/home

[2]Code will be available from https://github.com/sbj028/DeepConvolutionalForestParameterRegression

model was fitted directly to all the true prediction targets. To address this shortcoming and improve the performance of CNN models, we propose to impute pseudo-targets from the ALS-derived prediction maps into the dataset of true prediction targets, so that the CNN model is trained on the complete set of available targets. Since the ground reference dataset is much smaller than the prediction maps, this is in practice done by inserting true targets into the pseudo-target prediction maps. Following the imputation process, the Tanzanian dataset comprises less than 0.08% of target values originating from the ground reference data. For the Norwegian datasets, the ground reference data represents less than 0.04%, 0.11%, and 0.13% of the pixels in the respective Nordre Land, Tyristrand, and Hole datasets after the imputation process.

We would generally use all available ground reference data for model training and hyperparameter tuning. However, for model evaluation, we report the performance after cross-validation (CV), where we have trained models on a target dataset that only contains 80% of the true target labels. The remaining 20% are reserved for validation. Results obtained with CV are referred to as CV-RMSE in the result section.

B. Preparing the datasets for contextual learning

To create training, test and validation datasets for the Norwegian regions, all true target labels from the field inventory were first inserted into the ALS prediction maps of pseudo-targets. Two binary masks were additionally created; the pseudo-target mask, denoted $\boldsymbol{\mathcal{M}}_{pt}$ indicates the positions of available ALS-derived predictions. It is needed for masked computation of the loss functions, which are restricted to pixels where prediction targets are available. The ground reference mask, denoted $\boldsymbol{\mathcal{M}}_{gr}$, holds the positions of the true prediction targets. It is also used in the loss computation, where we weight the loss for the true prediction targets higher than the pseudo-targets.

After having produced the imputed target dataset and the two masks, we follow the workflow shown in Fig. 5 to create datasets with training, test and validation image patches. The figure illustrates the process for Tyristrand, but it is identical for all three Norwegian regions. Firstly, all available data are combined into a stack, including the Sentinel-1 mosaic of nine feature bands, the imputed target map, and the two masks. Then the entire scene is divided into superpatches by splitting it into blocks with no overlap. A superpatch is defined as a block of pixels that is larger than the image patches we use for training, testing and validation. See Table II for an overview of the total number of pixels in each region, the corresponding size of each superpatch and the number of possible superpatches that can be extracted for that region. $\boldsymbol{\mathcal{M}}_{pt}$ was used to remove superpatches with no overlap with pseudo-targets. Among all available superpatches, those with at least 10% overlap with $\boldsymbol{\mathcal{M}}_{pt}$ were identified as candidates for the test dataset. Fulfilling this criterion, approximately 15% of all available superpatches were randomly selected as test superpatches. These were further split into test patches of 64×64 pixels without overlap. Test patches having no overlap with $\boldsymbol{\mathcal{M}}_{pt}$ were discarded. Table II shows each region's final number of test patches.

TABLE II
DESCRIPTION OF THE DIFFERENT REGIONS, INFORMATION RELATED TO CREATING TRAINING AND TEST PATCHES AND THE FINAL NUMBER OF TEST AND TRAINING PATCHES PER REGION. THE LISTED NUMBER OF TRAINING PATCHES ARE AFTER DATA AUGMENTATION. SEE SECTION IV-B FOR DETAILS

Region Name	Region shape (pixels)	Superpatch size	Number of superpatches	Number of test patches (64×64)	Number of training patches (64×64)
Nordre Land	3136×1984	224×248	128	87	18072
Tyristrand	1152×896	128×128	63	25	2776
Hole	768×768	128×128	12	14	1384
Liwale	423×1222	–	–	14	2784

The remaining superpatches were initially used for hyperparameter tuning. See Appendix A for details. After this, all patches not used for testing were combined into training sets for the Norwegian models by splitting superpatches into training image patches of 64×64 pixels using 50% overlap and data augmentation with flipping and rotation. Patches with no overlap with $\boldsymbol{\mathcal{M}}_{pt}$ were discarded. Table II lists the number of training image patches per region after hyperparameter tuning.

The training, test and validation datasets for Tanzania were created by similar use of superpatches. Since the Tanzanian ALS-derived prediction map covers the whole AOI without any discontinuities, there is no need to check if image patches overlap with pseudo-targets. See Table II for details on region sizes in pixels and the number of test and training patches.

To evaluate the models, we also created CV target datasets where we used only 80% of the true target labels from the field inventory. We compute the model's performance both when it is trained with all true target labels and also a CV performance for the case when 20% of the true target labels are held out and used for testing. When comparing these results, one must recall that in the former case, the model has seen the test data during training. Moreover, the models are in the CV case trained with less true prediction targets.

C. Backbone U-Net Implementation

The CNN we use for SAR-based prediction of forest parameters is a modified version of the U-Net architecture in [19], a fully convolutional encoder-decoder network originally developed for biomedical image segmentation. The U-Net consists of a contraction part and a symmetric extraction path, with skip connections between each encoder block and the associated decoder block. The skip connections imply that low-level feature maps from the contraction part are concatenated with high-level feature maps from the extraction part to improve the learning in each level of the network.

Fig. 6 illustrates the U-Net generator network we use with an encoder-decoder depth of 4. This is the depth used by the Norwegian models, determined by hyperparameter tuning, while the Tanzanian models use a depth of 5. In both cases, we use ResNet34 [39] as backbone for the convolutional encoder network and refer to the whole model as a regression U-Net.

The regression U-Net is trained to perform image-to-image translation. I.e., given Sentinel-1 image patches from the input domain, the model translates these into prediction maps of AGB or SV maps for the same area, guided by the imputed target data. For the Norwegian datasets, we have modified the

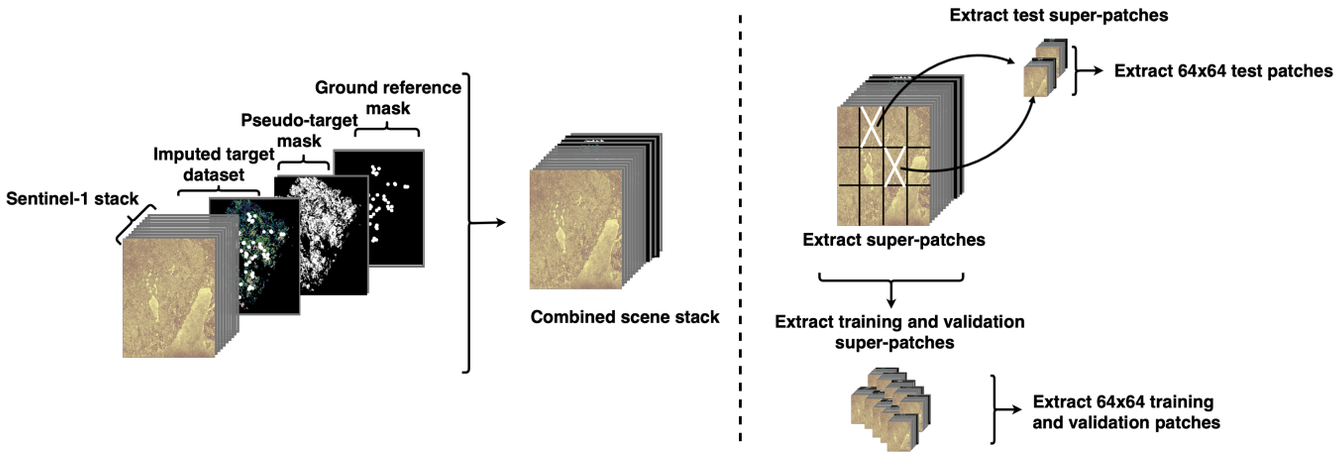


Fig. 5. Illustration of how training and test image patches are extracted from the stack of Sentinel-1 dataset, imputed target dataset, pseudo-target mask and ground reference mask. The datasets shown are retrieved from the AOI in Tyristrand. However, the process is identical for all Norwegian regions and representative of how the Tanzanian dataset is prepared. See Section IV-B for details.

first layer of the encoder to enable nine-channel inputs, i.e. input tensors of dimension $9 \times 64 \times 64$. The Tanzanian models take three-channel inputs with a shape of $3 \times 64 \times 64$. Additionally, the segmentation head was removed from the original U-Net architecture, as our work concerns a regression task and not a segmentation task. Finally, the softmax activation function in the final layer was replaced with a ReLU activation function to ensure non-negative AGB and SV predictions.

The initial layer of the encoder network uses a 7×7 convolution kernel with a stride of 2, followed by a normalisation layer, ReLU activation and a max-pooling operation. This implies that the number of feature channels is increased to 64, while the image dimension is decreased to 16×16 pixels. The following layer combines residual basic blocks, each using a 3×3 convolution, followed by a normalisation layer, ReLU activation, 3×3 convolution and a normalisation layer. The following encoder layers' residual blocks additionally employ down-sampling layers, which double the feature channels and half the spatial resolution of the image. In addition to the skip connections previously mentioned, each residual block uses common short connections [39].

Each block in the extraction part uses upsampling through nearest-neighbour interpolation and combines feature maps from the skip connection. It further processes the feature maps through two identical transformations, each including 3×3 convolutional filtering followed by a normalisation layer, ReLU activation and identity mapping. The upsampling procedure halves the number of feature maps while doubling the spatial resolution. We use the Pytorch implementation of the U-Net model from [41] for our regression U-Net, with the above-mentioned modifications.

D. Pretraining Stage

We follow the training procedure proposed for ESRGAN, a super-resolution model trained with multiple objectives in [20], and divide the training of the U-Net architecture into two stages: pretraining and fine-tuning.

In the pretraining stage, we train two baseline CNN models: an $\mathcal{L}_1$-based regression U-Net and a cGAN-based generative U-Net. In the fine-tuning stage, described in Section IV-E, we continue to train the baseline models with additional losses.

1) Pixel-aware Regression U-Net: We refer to a regression-type U-Net model optimised on a pixel-wise loss computed between model-inferred predictions and target predictions as a pixel-aware regression U-Net (PAR U-Net). In this work, the PAR U-Net is optimised on the $\mathcal{L}_1$ loss similar to [20], i.e.

$$\mathcal{L}_1 = \sum_k ||\boldsymbol{Y} - F(\boldsymbol{X})|| = \sum_k ||\boldsymbol{Y} - \hat{\boldsymbol{Y}}||, \tag{4}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ represent a corresponding pair of input and target image patches from the training dataset, $\hat{\boldsymbol{Y}} = F(\boldsymbol{X})$ is the image patch predicted by a CNN model $F(\cdot)$, and k is the total number of image patches.

2) cGAN U-Net: In addition to training the modified U-Net with a $\mathcal{L}_1$ loss, we also train it as a cGAN, like in [12]. Formally conditioned on image patches from the Sentinel-1 input domain, the cGANs generator (G) network is trained to learn the optimal mapping $G : \mathcal{X} \rightarrow \mathcal{Y}$ to generate realistic-looking image patches from the target domain. The G network also uses the regression U-Net architecture in Fig. 6.

Simultaneously as the G network aims to improve the generation task, the adversarially trained discriminator network (D) is trained to distinguish between real or false pairs of image patches successfully. A real pair of image patches corresponds to one Sentinel-1 image patch and the corresponding target ALS-derived prediction map. On the other hand, a false pair corresponds to one Sentinel-1 image patch and the corresponding prediction map generated by G. Adversarial training of G and D results from optimising the minimax loss function of the so-called Vanilla GAN (VGAN) [42]:

$$\begin{aligned}\min_G \max_D \mathcal{L}_{VGAN}(D, G) = \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[\log\, D(\boldsymbol{X}, \boldsymbol{Y})] \\ +\mathbb{E}_{\boldsymbol{X}}[\log(1 - D(\boldsymbol{X}, G(\boldsymbol{X})))].\end{aligned} \tag{5}$$

However, to address stability issues during training of the VGAN, the least squares GAN (LSGAN) was introduced [43].

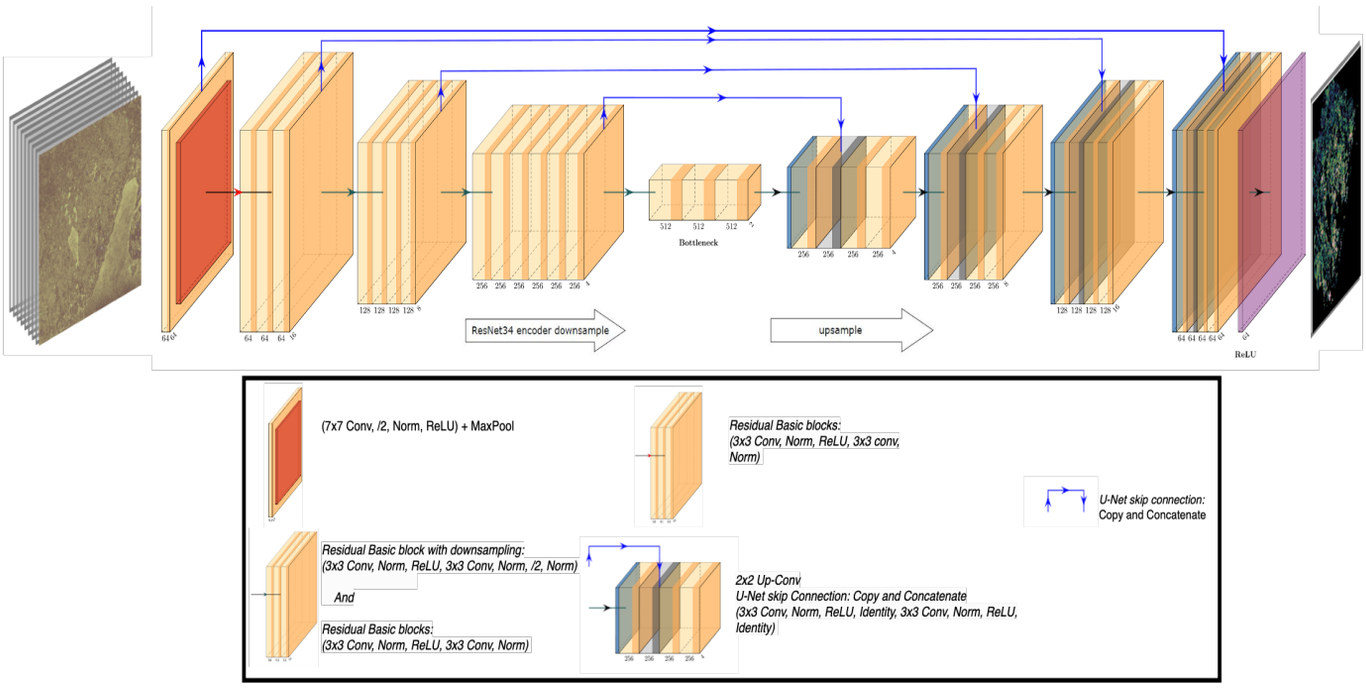


Fig. 6. Above: Regression U-Net architecture used for image-to-image translation. Below: modules of the regression U-Net. Modification of figure in [40].

In a conditional setting, it optimises the objective functions

$$\begin{aligned}\min_D \mathcal{L}_{LSGAN}(D) =& \frac{1}{2}\mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[(D(\boldsymbol{X},\boldsymbol{Y})-b)^2]+ \\ & \frac{1}{2}\mathbb{E}_{\boldsymbol{X}}[(D(\boldsymbol{X},G(\boldsymbol{X}))-a)^2]\,,\end{aligned} \tag{6}$$

$$\min_G \mathcal{L}_{LSGAN}(G) = \frac{1}{2}\mathbb{E}_{\boldsymbol{X}}[(D(\boldsymbol{X},G(\boldsymbol{X}))-c)^2], \tag{7}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are image patches from the input and the target domain, a and b are labels for false and real data, while c denotes a value that G tricks D to believe for false data [43].

Isola *et al.* [16] suggest to combine a GAN loss with an $\mathcal{L}_1$ loss to reduce visual artefacts in the generated images. The contribution of the $\mathcal{L}_1$ loss to the overall objective function is weighted by a regularisation parameter α, which is determined by hyperparameter tuning. In [12], the LSGAN model was found to outperform a VGAN and a Wasserstein GAN [44]. We therefore replace (7) with the following objective function for the generator of the baseline cGAN U-Net:

$$\mathcal{L}_{cGAN}(G) = \mathcal{L}_{L1} + \alpha\mathcal{L}_{LSGAN}(G), \quad \alpha \in [0,1]. \tag{8}$$

We find an optimal value of $\alpha = 0.01$, as in [16]. Similar to [16], we do not change the objective function of D for the baseline cGAN U-Net.

In [16], different architectures were evaluated for the discriminator D by altering the patch size N of the receptive fields, ranging from a 1×1 PixelGAN to an $N\times N$ PatchGAN. The D network applies convolutional processing to the pair of image patches to produce several classification responses, which are then averaged to determine whether the pair of image patches is real or false. In a PixelGAN, the discriminator attempts to classify each 1×1 pixel within the image patch as either real or false. In contrast, for the two PatchGAN networks, the discriminator tries to differentiate each $N \times N$ patch of pixels in the image patch as real or false.

E. Fine-tuning Stage

This section describes the loss functions used to fine-tune the PAR U-Net model and the cGAN U-Net model.

1) Pixel- and frequency-aware regression U-Net: To enforce the regression U-Net to focus on the alignment of the image frequency components during training, we propose to add a frequency-aware loss to the pixel-aware regression model or the adversarial cGAN model. We choose to employ the FFT loss from [29], which has shown promising results and is complementary to existing spatial losses. It is formulated as

$$\begin{aligned}\mathcal{L}_{FFT} =& \frac{1}{k}\sum\left(imag[\mathcal{F}(\boldsymbol{Y})] - imag[\mathcal{F}(\hat{\boldsymbol{Y}})]\right)^2 + \\ & \frac{1}{k}\sum\left(real[\mathcal{F}(\boldsymbol{Y})] - real[\mathcal{F}(\hat{\boldsymbol{Y}})]\right)^2,\end{aligned} \tag{9}$$

where $\mathcal{F}$ denotes the fast Fourier transform. The $\mathcal{L}_{FFT}$ uses MSE to enforce alignment of the *real* and *imaginary* parts of target and generated image patches in the frequency domain.

The total composite loss function becomes:

$$\mathcal{L}_{tot} = \mathcal{L}_1 + \alpha\mathcal{L}_{LSGAN} + \gamma\mathcal{L}_{FFT}, \quad \alpha,\gamma \in [0,1]. \tag{10}$$

A regularisation parameter γ is associated with $\mathcal{L}_{FFT}$ to adjust its influence on $\mathcal{L}_{tot}$. The α is still associated with $\mathcal{L}_{cGAN}$. All objective functions we use in the fine-tuning can be formulated with $\mathcal{L}_{tot}$, as we can ablate it by setting $\alpha = 0$ or $\gamma = 0$.

The baseline PAR U-Net model is either fine-tuned on the $\mathcal{L}_{cGAN}$ loss ($\mathcal{L}_{tot}$ with $\gamma = 0$), the combined $\mathcal{L}_1$ and $\mathcal{L}_{FFT}$ loss ($\mathcal{L}_{tot}$ with $\alpha = 0$), or the $\mathcal{L}_{tot}$ loss. The baseline cGAN U-Net model is fine-tuned on $\mathcal{L}_{tot}$. We refer to Appendix A for an extensive evaluation of model settings and hyperparameters used in the pretraining and fine-tuning phase.

F. Masked Loss Computation on Discontinuous Data

Due to the discontinuity of the ALS-derived SV prediction maps from Norway, there is not a target pixel for every pixel of the continuous SAR predictor dataset. To remedy this, we introduce masked loss computation. In this way, the convolutional processing of the predictor data creates a wall-to-wall map of model predictions, but in the comparison with the target dataset, pixels without prediction targets are masked out and excluded from the learning process. In addition to masking the pseudo-targets, we want to boost learning for pixels and patches with true prediction targets, hence reducing the impact of pseudo-targets relative to true targets.

As shown in Fig. 4, the training dataset contains two binary masks of the same size as the input and target data patches: the ground reference mask, $\boldsymbol{\mathcal{M}}_{gr}$, and the pseudo-target mask, $\boldsymbol{\mathcal{M}}_{pt}$, which for the Tanzanian AOI contains only ones. Masked losses are computed through simple Hadamard products, i.e. element-wise multiplication, denoted $\odot$. For instance, the masked $\mathcal{L}_1$ loss becomes:

$$\begin{aligned}\mathcal{L}_1^{\mathcal{M}} &= \mathcal{L}(\boldsymbol{\mathcal{M}} \odot \boldsymbol{Y}, \boldsymbol{\mathcal{M}} \odot \hat{\boldsymbol{Y}}) \\ &= \frac{1}{N \times N} \sum_{i,j} (\boldsymbol{\mathcal{M}} \times |y_{i,j} - \hat{y}_{i,j}|),\end{aligned} \tag{11}$$

where $\boldsymbol{\mathcal{M}}$ can be $\boldsymbol{\mathcal{M}}_{pt}$ or $\boldsymbol{\mathcal{M}}_{gr}$, $y_{i,j}$ and $\hat{y}_{i,j}$ are pixels of the target patch $\mathbf{Y}$ and the predicted target patch $\hat{\mathbf{Y}}$, whose size is $N \times N$. Similarly, $\mathcal{L}_{FFT}$ can be computed on $\mathcal{F}(\boldsymbol{\mathcal{M}} \odot \boldsymbol{\mathcal{Y}})$ and $\mathcal{F}(\boldsymbol{\mathcal{M}} \odot \hat{\boldsymbol{\mathcal{Y}}})$. Also the discriminator *D* can be fed with masked patches, either the real pair $(\boldsymbol{\mathcal{M}} \odot \boldsymbol{X}, \boldsymbol{\mathcal{M}} \odot \boldsymbol{Y})$ or the fake pair $(\boldsymbol{\mathcal{M}} \odot \boldsymbol{X}, \boldsymbol{\mathcal{M}} \odot G(\boldsymbol{X}))$. With this input to *D*, the $\mathcal{L}_{LSGAN}$ losses in (6) and (7) generalise to the masked case.

Let loss functions masked with $\boldsymbol{\mathcal{M}}_{pt}$ and $\boldsymbol{\mathcal{M}}_{gr}$ be denoted $\mathcal{L}^{\mathcal{M}_{pt}}$ and $\mathcal{L}^{\mathcal{M}_{gr}}$, respectively. To weight the true targets and the pseudo-targets differently, the total loss is decomposed as:

$$\begin{aligned}\mathcal{L}_{tot} &= \delta \mathcal{L}_{tot}^{\mathcal{M}_{gr}} + \mathcal{L}_{tot}^{\mathcal{M}_{pt}} \\ &= \delta \mathcal{L}_1^{\mathcal{M}_{gr}} + \mathcal{L}_1^{\mathcal{M}_{pt}} + \gamma \left(\delta \mathcal{L}_{FFT}^{\mathcal{M}_{gr}} + \mathcal{L}_{FFT}^{\mathcal{M}_{pt}} \right) \\ &\quad + \alpha \left(\delta \mathcal{L}_{LSGAN}^{\mathcal{M}_{gr}} + \mathcal{L}_{LSGAN}^{\mathcal{M}_{pt}} \right),\end{aligned} \tag{12}$$

with $\alpha, \gamma \in [0, 1]$ and true target weighting parameter $\delta \gg 1$, found from hyperparameter tuning (see Appendix A). A masked loss decreases when the mask has many zeros, which is as intended, since the amount of true or pseudo-targets contained in a patch should determine its impact. This is inspired by pseudo-labelling [45], a related semi-supervised learning algorithm for categorical prediction. It recommends to balance the losses computed over pseudo-labels (the categorical equivalent to the pseudo-targets in the regression task) and true labels, as there are generally much more pseudo-labels than true labels. In our training paradigm, this translates to boosting the masked loss computed over the true targets.

V. Experimental Results

This section presents experimental results of the prediction models trained on the Tanzanian and Norwegian datasets. We provide results on both regional and pan-regional models for the Norwegian datasets. The pan-regional models have been trained on all available training datasets from Nordre Land, Tyristrand and Hole. The regional models were trained on datasets from either Nordre Land, Tyristrand or Hole, and evaluated on the test data from the same region it was trained on. Appendix A provides details on hyperparameter tuning and settings used during model training.

Results are given both for the pretraining stage, i.e. the baseline models, and the fine-tuning stage as root mean square error (RMSE) and mean absolute error (MAE). Models with a low RMSE and MAE are preferred, as indicated by the symbol $\downarrow$ in the tables. Models have been trained to in two ways: (i) using all true target imputed with pseudo-targets; (ii) in cross-validation (CV) mode by rotationally imputing 80% of the target labels with the available pseudo-targets. For the latter case, a CV-RMSE is reported as μ (mean) $\pm\sigma$ (standard deviation). In the evaluation, we report model performance on the true targets and unseen test dataset. Since the CNN models work on image patches, model predictions are inferred by processing the AOI as 64×64 Sentinel-1 image patches with 50% overlap. A wall-to-wall prediction map is created by mosaicking patches through linear image blending, using the p-norm with a heuristic value of $p=5$, as proposed in [12].

A. Results: Tanzania models

The Tanzanian test set consists of 14 patches of pseudo-target AGB predictions and true targets from the 88 field plots. Quantitative results in terms of model performance on both the pseudo-target dataset and on the true targets are given in Table III. Metrics for the original ALS-derived AGB model, see [9], [12], and the best sequential cGAN model from [12] are also provided. Note that the best cGAN model from [12] was trained only on pseudo-targets, without access to true targets. We do not report the performance of the original ALS-derived AGB model and the sequential cGAN model on the test dataset, or the $\mu\pm\sigma$ CV-RMSE on the true targets, as these metrics were not provided in [9], [12]. All units in Table III are of $\mathrm{Mg\,ha^{-1}}$. Numbers in boldface indicate the best-performing model per column, while (•) indicates that a model performs better than the baseline ALS model.

B. Results: Norwegian models

The Norwegian models have all been trained to translate Sentinel-1 data into ALS-derived SV predictions for commercial forests. Table II shows that data from Nordre Land is over-represented in the Norwegian dataset. I.e., approximately 80% of the training image patches are from Nordre Land, while only 6% are from the Hole region. Four types of Norwegian models were developed: one pan-regional model that represents all three regions and separate regional models for Nordre Land, Tyristrand and Hole. The pan-regional models were trained on pooled training data from all regions, but evaluated separately on each region's pseudo-target data and true target data. The three regional models were both trained and evaluated on data from each separate region.

Since Nordre Land is over-represented in the dataset, we wish to investigate if the pan-regional models evaluated on Nordre Land perform similarly to the corresponding regional

TABLE III

QUANTITATIVE EVALUATION OF MODELS TRAINED ON THE TANZANIAN DATASET. METRICS OF $RMSE$ AND MAE ARE MEASURED WITH RESPECT TO THE PSEUDO-TARGET DATA (LEFT SIDE OF THE TABLE) AND GROUND REFERENCE DATA (RIGHT SIDE OF THE TABLE). CV-RMSE IS GIVEN AS MEAN ± STANDARD DEVIATION. NUMBERS IN BOLDFACE INDICATE THE BEST-PERFORMING MODEL PER COLUMN. ALL UNITS ARE IN $\mathrm{Mg\,ha^{-1}}$

Models		**RMSE ↓**	**CV-RMSE ↓**	**MAE ↓**	**RMSE ↓**	**CV-RMSE ↓**	**MAE ↓**
Pretraining:	**Fine-tuning:**	w.r.t. pseudo-target dataset			w.r.t. ground reference data		
Baseline ALS[a]	–	–	–	–	33.39	—	24.61
Sequential cGAN[b]	–	–	–	–	39.84	—	31.46
PAR U-Net ($\mathcal{L}_1$)	–	29.82	29.47 ± 0.15	20.05	34.24	35.64 ± 7.31	25.82
cGAN U-Net ($\mathcal{L}_{cGAN}$)	–	32.02	31.19 ± 0.21	22.46	37.22	38.78 ± 5.9	28.91
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	29.31	29.81 ± 0.51	19.92	**32.91**•	**35.53 ± 4.14**	**25.53**
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_1$ + $\mathcal{L}_{FFT}$	**26.10**	**26.17 ± 0.05**	**18.11**	34.40	36.13 ± 3.09	26.54
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	32.75	33.58 ± 0.54	24.21	36.19	37.46 ± 5.36	28.24
cGAN U-Net ($\mathcal{L}_{cGAN}$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	29.40	30.73 ± 0.26	20.85	37.65	38.49 ± 7.17	29.18

[a] The conventional ALS-based statistical regression model proposed in [9]. Metrics are retrieved from [9] and [12].
[b] The optimal cGAN-based sequential regression models proposed in [12]. Provided metrics are from the same source.
• Indicates that a model performs better than the Baseline ALS model.

TABLE IV

QUANTITATIVE EVALUATION OF PAN-REGIONAL NORWEGIAN MODELS. METRICS FOR EACH REGION ARE PROVIDED WITH RESPECT TO PSEUDO-TARGET DATA (LEFT SIDE OF THE TABLE) AND GROUND REFERENCE DATA (RIGHT SIDE OF THE TABLE) AS RMSE, MAE AND CV-RMSE, THE LATTER GIVEN AS MEAN ± STANDARD DEVIATION. NUMBERS IN BOLDFACE INDICATE THE BEST-PERFORMING MODEL PER COLUMN. ALL UNITS ARE IN $m^3 ha^{-1}$

Models		**RMSE ↓**	**CV-RMSE ↓**	**MAE ↓**	**RMSE ↓**	**CV-RMSE ↓**	**MAE ↓**
Pretraining:	**Fine-tuning:**	w.r.t. pseudo-target data			w.r.t. ground reference data		
Region: Nordre Land							
Baseline ALS	–	–	–	–	83.54	–	63.29
PAR U-Net ($\mathcal{L}_1$)	–	**68.08**	**68.67 ± 0.29**	**31.89**	73.72•	92.63 ± 3.74	38.32•
cGAN U-Net ($\mathcal{L}_{cGAN}$)	–	77.66	77.22 ± 3.30	34.31	88.30	120.6 ± 27.35	48.03•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	68.94	69.17 ± 0.26	32.90	73.93•	**79.36± 3.37**	**32.86**•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_1$ + $\mathcal{L}_{FFT}$	72.38	71.96 ± 0.48	35.32	70.92•	79.66 ± 4.01	33.66•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	71.57	72.84 ± 2.34	34.03	**68.72**•	99.67 ± 36.23	37.71•
cGAN U-Net ($\mathcal{L}_{cGAN}$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	78.48	77.22 ± 1.19	37.46	93.09	92.55 ± 33.16	53.51•
Region: Tyristrand							
Baseline ALS	–	–	–	–	75.62	–	59.17
PAR U-Net ($\mathcal{L}_1$)	–	**62.31**	**63.01 ± 1.37**	**24.83**	43.77•	58.96 ± 6.20	**28.30**•
cGAN U-Net ($\mathcal{L}_{cGAN}$)	–	84.96	79.43 ± 6.44	30.17	76.59	98.98 ± 14.35	55.68•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	64.22	66.51 ± 0.93	25.77	**40.75**•	**45.78 ± 4.61**	28.56•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_1$ + $\mathcal{L}_{FFT}$	76.86	75.63 ± 1.41	28.94	42.80•	49.17 ± 2.16	28.28•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	65.50	65.99 ± 1.72	26.74	55.04•	74.89 ± 22.09	35.74•
cGAN U-Net ($\mathcal{L}_{cGAN}$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	84.84	78.30 ± 5.95	29.88	101.55	98.54 ± 14.62	65.96
Region: Hole							
Baseline ALS	–	–	–	–	60.94	–	50.06
PAR U-Net ($\mathcal{L}_1$)	–	**113.82**	**116.18 ± 1.00**	**57.68**	72.47	82.43 ± 7.37	41.39•
cGAN U-Net ($\mathcal{L}_{cGAN}$)	–	136.03	129.57 ± 3.62	65.24	95.61	126.87 ± 12.53	64.37
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	121.14	124.71 ± 1.15	59.90	67.54	72.24± 5.49	**37.05**•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_1$ + $\mathcal{L}_{FFT}$	124.13	123.51 ± 0.54	61.09	**64.69**	**69.89 ± 3.13**	39.22•
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	114.59	116.10 ± 1.78	59.13	68.83	94.44 ± 27.35	43.06•
cGAN U-Net ($\mathcal{L}_{cGAN}$)	$\mathcal{L}_{cGAN}$ + $\mathcal{L}_{FFT}$	123.98	125.40 ± 2.39	60.99	94.10	122.90 ± 17.08	59.66

• Indicates that a model performs better than the Baseline ALS model.

models developed for Nordre Land. On the other hand, as the available data from both Hole and Tyristrand are limited, we wish to compare the respective regional models to the pan-regional model. The aim is to identify and quantify any difference in performance and, if possible, to draw conclusions about transferability and impacts of dataset size.

As for the Tanzania, different CNN models were evaluated against each other by comparing their performance on unseen test patches of pseudo-target data and on true targets of field measured SV. The number of field plots, i.e. true targets, in each region, can be found in Table I. The Hole test set consists of 14 patches of pseudo-target data, Tyristrand of 25 and Nordre Land of 87 test patches, each of 64×64 pixels.

Quantitative results from the evaluation of the pan-regional Norwegian models are listed in Table IV while Table V lists results for the regional models. For the regional models, only results for the baseline PAR U-Net model and the model pretrained on $\mathcal{L}_1$ and fine-tuned with the $\mathcal{L}_{cGAN}$ loss are given, as these have proven to be robust on both the Tanzanian data and the pan-regional Norwegian dataset. Metrics obtained with the original ALS-derived SV model have been computed for each region by extracting the area-weighted mean of ALS-derived SV predictions at the location of each field plot. The CV-RMSE for the original ALS-derived SV models were not provided to us for this work and are therefore not given in Table IV or Table V. All metrics in both tables are in units of $\mathrm{m^3\,ha^{-1}}$. Boldface numbers in a column of Table III indicate the model that performs best. A (•) symbol indicates that a model performs better than the baseline ALS model.

TABLE V
QUANTITATIVE EVALUATION OF REGIONAL NORWEGIAN MODELS. METRICS FOR EACH REGION ARE PROVIDED WITH RESPECT TO PSEUDO-TARGET DATA (LEFT SIDE OF THE TABLE) AND GROUND REFERENCE DATA (RIGHT SIDE OF THE TABLE) AS RMSE, MAE AND CV-RMSE, THE LATTER GIVEN AS MEAN ± STANDARD DEVIATION. NUMBERS IN BOLDFACE INDICATE THE BEST-PERFORMING MODEL PER COLUMN. ALL UNITS ARE OF m^3ha^{-1}.

Models		**RMSE ↓**	**CV-RMSE ↓**	**MAE ↓**	**RMSE ↓**	**CV-RMSE ↓**	**MAE ↓**
Pretraining:	**Fine-tuning:**	w.r.t. pseudo-target data			w.r.t. ground reference data		
Region - Nordre Land:							
Baseline ALS	–	–	–	–	83.54	–	63.29
PAR U-Net ($\mathcal{L}_1$)	–	**69.65**	**69.48± 0.23**	**31.89**	75.53$^\bullet$	92.36 ± 2.06	36.82$^\bullet$
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	70.45	70.51 ± 0.15	32.94	**69.41**$^\bullet$	**82.40 ± 3.68**	**32.70**$^\bullet$
Region - Tyristrand:							
Baseline ALS	–	–	–	–	75.62	–	59.17
PAR U-Net ($\mathcal{L}_1$)	–	**66.77**	**67.96 ± 0.59**	**27.79**	48.04$^\bullet$	63.77± 8.22	32.5$^\bullet$
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	68.80	69.36 ± 0.42	28.96	**45.22**$^\bullet$	**60.73 ± 4.80**	**27.35**$^\bullet$
Region - Hole:							
Baseline ALS	–	–	–	–	60.94	–	50.06
PAR U-Net ($\mathcal{L}_1$)	–	136.19	138.82 ± 1.13	71.88	74.60	99.48 ± 10.47	42.55$^\bullet$
PAR U-Net ($\mathcal{L}_1$)	$\mathcal{L}_{cGAN}$	**132.47**	**132.58 ± 0.76**	**70.73**	**65.47**	**77.38± 5.37**	**38.22**$^\bullet$

$^\bullet$ Indicates that a model performs better than the Baseline ALS model.

VI. DISCUSSION

Six new CNN-based regression models (two baseline and four fine-tuned ones) have been developed to improve earlier work on the Tanzanian dataset using the semi-supervised imputation strategy proposed herein. Above all, Table III shows that the model pretrained on the $\mathcal{L}_1$ loss and fine-tuned on the $\mathcal{L}_{cGAN}$ loss performs better than the conventional statistical ALS-based AGB model proposed in [9], and all other Tanzanian models on the field data. The CNN model that most accurately recreates the AGB pseudo-target data is pretrained on the $\mathcal{L}_1$ loss and fine-tuned on the combined $\mathcal{L}_1$ and $\mathcal{L}_{FFT}$ loss, see Table III. The results on the Tanzanian dataset show the potential of a two-stage training paradigm and of frequency-aware training to reduce the impact of spectral bias. Furthermore, the results in Table III show that the baseline PAR U-Net model performs better than the baseline cGAN U-Net model on both the pseudo-target and the true target data. These findings align with existing knowledge in the field of image super-resolution: It is disadvantageous to adopt a purely adversarial training strategy on tasks that require high reconstruction accuracy in terms of RMSE. In this case, employing a simpler pixel-wise regression U-Net is better. The proposed baseline cGAN U-Net model is most similar to the sequential cGAN model proposed in [12]. Table III shows that the proposed semi-supervised imputation strategy improves the CNN models' performance in AGB prediction.

Several new CNN models are also proposed for SV prediction on the Norwegian datasets. Our approach is to train pan-regional models by combining data from all three Norwegian regions, Nordre Land, Tyristrand and Hole, followed by evaluation of test and field data from each individual region. The purpose of the pan-regional models is to develop models that generalise well to more than one region, which is particularly advantageous for regions with little training data. As a result, these models hold the potential for substantial cost-savings if field work can be reduced during operational inventories.

According to Table IV, the baseline PAR U-Net model outperforms the other models in accurately recreating the pseudo-target SV data. We advise to avoid the baseline cGAN U-Net model or the pan-regional model that was pretrained on the $\mathcal{L}_{cGAN}$ loss, followed by fine-tuning on the combined $\mathcal{L}_{cGAN}$ and $\mathcal{L}_{FFT}$ losses, when training CNN models for SV prediction. As the models are evaluated on RMSE and not perceptual quality, the results suggest that adversarial training should be avoided in the initial training phase. As demonstrated in Table IV, fine-tuning and the composition of losses generally improve model performance with respect to field data, with few exceptions. Moreover, all pan-regional fine-tuned models perform better than the conventional statistical ALS-based models derived for either Nordre Land, Tyristrand, or Hole. Based on the CV-RMSE, we recommend using fine-tuned models that are pretrained on $\mathcal{L}_1$ and fine-tuned on either the combined $\mathcal{L}_1$ and $\mathcal{L}_{FFT}$ loss or on $\mathcal{L}_{cGAN}$. For instance, the model fine-tuned on the combined $\mathcal{L}_1$ and $\mathcal{L}_{FFT}$ loss performs best on the Hole field data, whereas the model fine-tuned on $\mathcal{L}_{cGAN}$ performs best on Tyristrand field data.

In addition to the pan-regional models trained on the whole Norwegian dataset, regional models were developed for this work. Unlike the pan-regional models, these were only trained and evaluated on a specific region. Table II shows a significant difference in the amount of available training data among the regions. The Hole region has the least data, followed by Tyristrand, while Nordre Land has the most data. Consequently, it implies that the regional Nordre Land model has been trained on almost the same training data as the pan-regional model. For Hole (and Tyristrand), the regional models are trained on only a fraction of the training data available for the pan-regional model, which could impact their relative performance. Based on the discussion above, we train the following two models: a regional PAR U-Net model and a regional model pretrained on $\mathcal{L}_1$ and fine-tuned on the $\mathcal{L}_{cGAN}$ loss. The fine-tuned model was chosen among the other three, as it has proven to be robust on both the Tanzanian data and the pan-regional Norwegian dataset.

In general, comparing the results of the pan-regional models in Table IV to the regional models in Table V, we observe that the pan-regional Norwegian models perform better than all regional models with one exception. The regional Nordre Land model pretrained on the $\mathcal{L}_1$ objective and fine-tuned on the $\mathcal{L}_{cGAN}$ objective performs better than the pan-regional

model on the corresponding regional model on field data. These results show the potential of training regional models that utilises all available data from nearby regions.

To our knowledge, it is the first time that the $\mathcal{L}_{FFT}$ loss has been evaluated outside the natural image domain, e.g. on remote sensing images. Our results from both the Tanzanian and Norwegian models show that the simple $\mathcal{L}_{FFT}$ objective function efficiently reduces the impact of spectral bias and thereby improves the performance of the CNN model.

VII. Conclusion

Through the use of a semi-supervised imputation strategy, we demonstrate the ability of contextual generative CNN models to accurately map Sentinel-1 C-band data to target data consisting of spatially disjoint polygons of ALS-derived prediction maps. The generalisation ability of our modelling approach was evaluated for AGB prediction in the Tanzanian miombo woodlands and for SV prediction in three managed boreal forests in Norway. Our results show that the models developed using the imputation strategy achieve state-of-the-art performance compared to previous studies, suggesting that the contextual C-band SAR-based models outperform conventional statistical ALS-based models in accurately predicting the target labels of ground reference data. Furthermore, we demonstrate that a two-phased learning strategy, which includes pretraining with a pure pixel-wise regression U-Net followed by either a regression cGAN model or a pixel- and frequency-aware regression U-Net in the fine-tuning phase, improves model performance. We argue that pixel-aware pretraining enforces the model to focus on pixel-to-pixel relationships before learning general relationships.

Acknowledgments

We gratefully acknowledge the Norwegian University of Life Sciences, the Tanzania Forest Services Agency, Prof. Eliakimu Zahabu and coworkers at Sokoine University of Agriculture, Viken Skog and the Swedish University of Agricultural Sciences for participation in field work and provision of in situ measurements, ALS-derived AGB and SV products. Special thanks to Prof. Håkan Olsson for providing ALS data acquired by SLU and to Mr. Svein Dypsund at Viken Skog for providing in situ measurements in Norway. Many thanks to Assoc. Prof. Benjamin Ricaud for valuable input on relevant experience from the field of single-image super-resolution.

Sara Björk received the M.Sc. degree in Applied Physics and Mathematics from UiT The Arctic University of Norway, in 2016, where she is currently pursuing the Ph.D. degree in physics. Since 2022, she has been working as a system developer in the Earth Observation Team at KSAT Kongsberg Satellite Services. Her research interests include computer vision, image processing, and deep learning, with a particular focus on developing methodologies that leverage deep learning techniques and remote sensing data for forest parameter retrieval.

Stian Normann Anfinsen received the M.Sc. degree in communications, control and digital signal processing from the University of Strathclyde, Glasgow, UK (1998) and the Cand.scient. (2000) and Ph.D. degrees (2010) in physics from UiT The Arctic University of Norway (UiT), Tromsø, Norway. He is a faculty member at the Dept. of Physics and Technology at UiT since 2014, currently as adjunct professor in machine learning. Since 2021 he is a senior researcher with NORCE Norwegian Research Centre in Tromsø. His research interests are in statistical modelling and machine learning for image and time series analysis.

Michael Kampffmeyer is an associate professor and head of the Machine Learning Group at UiT The Arctic University of Norway. He is also an adjunct senior research scientist at the Norwegian Computing Center in Oslo. His research interests include explainable AI and learning from limited labels (e.g. clustering, few/zero-shot learning, domain adaptation and self-supervised learning). Kampffmeyer received the Ph.D. degree from UiT in 2018. He has had long-term research stays in the Machine Learning Department at Carnegie Mellon University and Berlin Center for Machine Learning at the Technical University of Berlin. He is general chair of the annual Northern Lights Deep Learning Conference.

Erik Næsset received the M.Sc. degree in forestry and the Ph.D. degree in forest inventory from the Agricultural University of Norway, Ås, Norway, in 1983 and 1992, respectively. His major field of research is forest inventory and remote sensing, with particular focus on operational management inventories, sample surveys, photogrammetry, and airborne LiDAR. He has played a major role in developing and implementing airborne LiDAR in operational forest inventory. He has been the leader and coordinator of more than 60 research programs funded by the Research Council of Norway, the European Union, and private forest industry. He has published around 250 papers in international peer-reviewed journals. His teaching includes lectures and courses in forest inventory, remote sensing, forest planning, and sampling techniques.

Terje Gobakken is professor in forest planning and has published more than 190 peer-reviewed scientific articles related to forest inventory and planning in international journals. He has been working at the Norwegian National Forest Inventory and participated in compiling reports of emissions and removals of greenhouse gases from land use, land-use change and forestry in Norway. He has coordinated and participated in a number of externally funded projects, including international projects funded by for example NASA and EU, and has broad practical and research-based experience with development of big data and information infrastructures for forest inventory, planning and decision support.

Lennart Noordermeer received the M.Sc. degree in forestry and the Ph.D. degree in forest inventory from the Norwegian University of Life Sciences (NMBU) in 2017 and 2020, respectively. He currently has a researcher position in the Forest Inventory Group at the Faculty of Environmental Sciences and Natural Resource Management, NMBU. His research focuses on operational forest inventory, with emphasis on the use of data from forest harvesters as well as the use of multitemporal remotely sensed data for forest productivity estimation.

REFERENCES

[1] S. Kaasalainen et al., "Combining lidar and synthetic aperture radar data to estimate forest biomass: Status and prospects," *Forests*, vol. 6, no. 12, pp. 252–270, Jan. 2015.

[2] A. Bombelli et al., *Biomass: Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables*. Rome, Italy: GTOS Secretariat, Food and Agriculture Organization of the United Nations (FAO), 2009, no. GTOS 67.

[3] T. Johansson, "Biomass production of Norway spruce (Picea abies (L.) Karst.) growing on abandoned farmland," *Silva Fennica*, vol. 33, no. 4, pp. 261–280, 1999.

[4] K. Ericsson, S. Huttunen, L. J. Nilsson, and P. Svenningsson, "Bioenergy policy and market development in Finland and Sweden," *Energy Policy*, vol. 32, no. 15, pp. 1707–1721, 2004.

[5] M. Segura and M. Kanninen, "Allometric models for tree volume and total aboveground biomass in a tropical humid forest in Costa Rica," *Biotropica*, vol. 37, no. 1, pp. 2–8, 2005.

[6] J. Urban, J. Čermák, and R. Ceulemans, "Above-and below-ground biomass, surface and volume, and stored water in a mature Scots pine stand," *Eur. J. Forest Research*, vol. 134, pp. 61–74, 2015.

[7] L. G. Marklund, "Biomass functions for pine, spruce and birch in Sweden," Department of Forest Survey., Swedish University of Agricultural Sciences, Umeå, Sweden, Tech. Rep. 45, 1988.

[8] G. Galidaki et al., "Vegetation biomass estimation with remote sensing: Focus on forest and other wooded land over the Mediterranean ecosystem," *Int. J. Remote Sens.*, vol. 38, no. 7, pp. 1940–1966, Apr. 2017.

[9] E. Næsset et al., "Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision," *Remote Sens. Environ.*, vol. 175, no. 15, pp. 282–300, Mar. 2016.

[10] L. Noordermeer, O. M. Bollandsås, H. O. Ørka, E. Næsset, and T. Gobakken, "Comparing the accuracies of forest attributes predicted from airborne laser scanning and digital aerial photogrammetry in operational forest inventories," *Remote Sens. Environ.*, vol. 226, pp. 26–37, 2019.

[11] S. Solberg, R. Astrup, T. Gobakken, E. Næsset, and D. J. Weydahl, "Estimating spruce and pine biomass with interferometric X-band SAR," *Remote Sens. Environ.*, vol. 114, no. 10, pp. 2353–2360, Oct. 2010.

[12] S. Björk, S. N. Anfinsen, E. Næsset, T. Gobakken, and E. Zahabu, "On the potential of sequential and nonsequential regression models for Sentinel-1-based biomass prediction in Tanzanian miombo forests," *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4612–4639, 2022.

[13] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogram. Remote Sens.*, vol. 173, pp. 24–49, 2021.

[14] A. Hamedianfar, C. Mohamedou, A. Kangas, and J. Vauhkonen, "Deep learning for forest inventory and planning: A critical review on the remote sensing approaches so far and prospects for further applications," *Forestry*, vol. 95, no. 4, pp. 451–465, Feb. 2022.

[15] S. Zolkos, S. Goetz, and R. Dubayah, "A meta-analysis of terrestrial aboveground biomass estimation using Lidar remote sensing," *Remote Sens. Environ.*, vol. 128, pp. 289–298, Jan. 2013.

[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1125–1134.

[17] J. Leonhardt, L. Drees, P. Jung, and R. Roscher, "Probabilistic biomass estimation with conditional generative adversarial networks," in *Proc. DAGM German Conf. Pattern Recognit. (GCPR)*, Konstanz, Germany, 2022, pp. 479–494.

[18] A. E. Pascarella, G. Giacco, M. Rigiroli, S. Marrone, and C. Sansone, "ReUse: REgressive Unet for carbon storage and above-ground biomass estimation," *J. Imaging*, vol. 9, no. 3, 2023.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Part III 18*, Munich, Germany, 2015, pp. 234–241.

[20] X. Wang et al., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *Eur. Conf. Comput. Vis.*, 2018, pp. 63–79.

[21] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 334–355.

[22] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.

[23] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 105–114.

[24] J. W. Soh, G. Y. Park, J. Jo, and N. I. Cho, "Natural and Realistic Single Image Super-Resolution With Explicit Natural Manifold Discrimination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 8114–8123.

[25] Y. Chen, G. Li, C. Jin, S. Liu, and T. Li, "SSD-GAN: Measuring the Realness in the Spatial and Spectral Domains," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 1105–1112.

[26] R. Durall, M. Keuper, and J. Keuper, "Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7890–7899.

[27] K. Chandrasegaran, N.-T. Tran, and N.-M. Cheung, "A closer look at Fourier spectrum discrepancies for CNN-generated images detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 7200–7209.

[28] M. Khayatkhoei and A. Elgammal, "Spatial Frequency Bias in Convolutional Generative Adversarial Networks," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 7, pp. 7152–7159, Jun. 2022.

[29] S. Björk, J. N. Myhre, and T. H. Johansen, "Simpler is better: Spectral regularization and up-sampling techniques for variational autoencoders," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2022, pp. 3778–3782.

[30] S. Czolbe, O. Krause, I. Cox, and C. Igel, "A loss function for generative neural networks based on Watson's perceptual model," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020.

[31] Y. Wang, L. Cai, D. Zhang, and S. Huang, "The frequency discrepancy between real and generated images," *IEEE Access*, vol. 9, pp. 115 205–115 216, 2021.

[32] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learning Representations (ICLR)*, Banff, Canada, 2014.

[33] E. Tomppo et al., "A sampling design for a large area forest inventory: Case Tanzania," *Can. J. Forest Res.*, vol. 44, no. 8, pp. 931–948, 2014.

[34] L. T. Ene, E. Næsset, T. Gobakken, O. M. Bollandsås, E. W. Mauya, and E. Zahabu, "Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data," *Remote Sens. Environ.*, vol. 188, pp. 106–117, Jan. 2017.

[35] L. T. Ene et al., "Large-scale estimation of aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data," *Remote Sens. Environ.*, vol. 186, pp. 626–636, Dec. 2016.

[36] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org," 2017.

[37] "QGIS Development Team (2019). QGIS Geographic Information System. Open Source Geospatial Foundation Project."

[38] "SNAP - ESA Sentinel Application Platform v.8.0."

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[40] H. Iqbal, "HarisIqbal88/PlotNeuralNet v1.0.0," Dec. 2018.

[41] P. Iakubovskii, "Segmentation Models Pytorch," 2019.

[42] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.

[43] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least Squares Generative Adversarial Networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2813–2821.

[44] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5767–5777.

[45] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on Challenges in Representation Learning, ICML*, vol. 3, 2013, p. 896.

[46] L. Biewald, "Experiment Tracking with Weights and Biases," https://www.wandb.com/, 2020.
[47] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 136–144.
[48] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3883–3891.

APPENDIX

A. Hyperparameter tuning for model selection

Extensive hyperparameter tuning was performed on the training and validation datasets for the Tanzanian and Norwegian models. Experiment tracking with Weights & Biases sweeps [46] was used, employing grid search during both pretraining and fine-tuning phases. Unlike most studies in forestry deep learning research [14], the Adam optimiser was used for all proposed models. Hyperparameter tuning focused on finding optimal batch size (BS), β_1 value for the Adam optimiser, learning rate (lr), encoder network, encoder/decoder depth, discriminator network, cGAN loss, and number of epochs. Three discriminator networks were evaluated: 1×1 PixelGAN, 16×16 PatchGAN, and 34×34 PatchGAN. The two PatchGAN networks were created following [12] by modifying discriminator depth to achieve receptive field sizes of 16×16 or 34×34. Hyperparameter tuning also evaluated normalisation layers, two different weight initialisation methods, and selection of objective functions for the pretraining and fine-tuning stages. 5-fold cross-validation (CV) was used for hyperparameter tuning of the Tanzanian and Norwegian models. For Norwegian models, training and validation datasets used for CV consisted of image patches from all three Norwegian regions. Superpatches not used for test sets were divided into training and validation splits with 80% for training and 20% for testing in 5-fold CV. Superpatches were further divided into training (or validation) image patches of 64×64 pixels with 50% overlap allowed for training data. Data augmentation with flipping and rotation was applied to increase training data. Validation loss, based on mean and median RMSE, was used to identify optimal hyperparameters and model settings for both the Tanzanian and Norwegian models. Table VI lists evaluated hyperparameters and search ranges for the Tanzanian and Norwegian models, where normalisation "None" refers to no normalisation layers.

1) Summary of findings from hyperparameter tuning: We observed that the three ResNet networks used as convolutional encoder had similar performance, but ResNet34 was the most accurate and was therefore selected. We found that cGAN-based models optimised on the $\mathcal{L}_{cGAN}$ objective were more accurate than those optimised on the VGAN objective, and the $\mathcal{L}_{cGAN}$ loss was therefore selected. The evaluation of different *D* networks showed that the Tanzanian adversarial models preferred the PixelGAN, while the 16×16 PatchGAN was preferred by the Norwegian adversarial models.

We also investigated the impact of the network's normalisation layers on the model performance. Previous work has argued that using BN in the network might harm the inherent range flexibility of the features [20], [47], [48]. They suggest to remove the BN layers from the model architecture to increase performance and reduce computational complexity for reconstruction tasks that optimise e.g. the RMSE. Motivated by [20], [47], [48], we compared batch normalisation (BN) layers to instance normalisation (IN) layers and no normalisation. Our experiments did not confirm that normalisation, and BN in particular, should be avoided. On the contrary, most models preferred BN or IN.

The potential of transfer learning was investigated by initialising the Tanzanian or Norwegian networks with or without ImageNet weights. Use of pretrained ImageNet weights requires that the input image patches from Sentinel-1 must be scaled to the range [0, 1] and normalised with ImageNet mean and standard deviation. This implies that the Sentinel-1 data no longer are in dB form. However, experiments showed that randomly initialised weights gave better performance than pretrained ImageNet weights. This confirms the claims from [12] that avoiding normalisation and keeping the input data on dB form resulted in improved prediction of ALS-derived AGB maps. Thus, no models in this study employ pretrained ImageNet weights.

In [16], the regularisation weight α was applied on the $\mathcal{L}_1$ part of the generator loss function and evaluated for $\alpha \in [0, 100]$. As explained in Section IV-D2 and shown in Eq. (8), we apply α on $\mathcal{L}_{LSGAN}$and combine it with $\mathcal{L}_1$, to form $\mathcal{L}_{cGAN}$. We evaluated α = [0.01, 0.1, 1]. In accordance with [16], we found that models trained with $\alpha = 0.01$ performed best.

Initial experiments showed that the $\mathcal{L}_1$ loss magnitude varies around 3×10^1, while the $\mathcal{L}_{GAN}$ loss approximates 1×10^0. In contrast, the $\mathcal{L}_{FFT}$ loss assumes magnitudes around 3×10^8. To balance the impact of the $\mathcal{L}_{FFT}$ loss with the other loss functions, we evaluated the following range for its regularisation weight: γ = [1e-8, 3e-8, 5e-8, 7e-8, 9e-8, 1e-7]. Our experiments show that γ = 9e-8 or 1e-7 is best.

The true target weight δ, used in Eq. (12), must be large to compensate for the strong imbalance between the numbers of true targets and pseudo-targets. We experimented with $\delta = [100, 200, 300, 400, 500]$ for the Tanzanian models and $\delta = [200, 300, 400, 500, 600, 700]$ for the Norwegian models.

The selected hyperparameters for the models pretrained on the Tanzania dataset and the Norway dataset are shown in Table VII. The resulting hyperparameters for the fine-tuned Tanzania models and the fine-tunes pan-regional Norwegian models are shown in Table VIII. The hyperparameters selected for the regional Norwegian models are similar to those of their pan-regional counterparts. The exceptions are the regional PAR U-Net baseline model, which was trained for 250 epochs instead of 200, and the regional model pretrained on the $\mathcal{L}_1$ loss and fine-tuned on the $\mathcal{L}_{cGAN}$ loss, which was fine-tuned for 250 epochs instead of 100. The tables report the number of epochs as $E_p + E_f$, denoting E_p epochs of pretraining and E_f in fine-tuning.

TABLE VI
SEARCH RANGE FOR HYPERPARAMETERS AND MODEL SETTINGS USED IN TANZANIAN AND NORWEGIAN MODELS.

Hyperparameters	Tanzanian dataset	Norwegian dataset
Batch size (BS)	[2, 4, 6]	[8, 32, 64, 128]
$\beta 1$ (Adam optimiser)	[0.4, 0.5, 0.6, 0.7, 0.8, 0.9]	[0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
Learning rate (lr)	[1e-2, 1e-3, 2e-3, 1e-4, 2e-4, 2e-5, 1e-5]	[1e-2, 1e-3, 2e-3, 1e-4, 2e-4, 2e-5, 1e-5]
Encoder Network	[ResNet18, ResNet34, ResNet50]	[ResNet18, ResNet34, ResNet50]
Encoder, Decoder depth	[4, 5]	[4, 5]
Discriminator network	[PixelGAN, PatchGAN(16), PatchGAN(34)]	[PixelGAN, PatchGAN(16), PatchGAN(34)]
cGAN objective	[VGAN, LSGAN]	[VGAN, LSGAN]
Normalisation	[Instance, Batch, None]	[Instance, Batch, None]
α	[0.01, 0.1, 1]	[0.01, 0.1, 1]
γ	[1e-8, 3e-8, 5e-8, 7e-8, 9e-8, 1e-7]	[1e-8, 3e-8, 5e-8, 7e-8, 9e-8, 1e-7]
True target weight δ	[100, 200, 300, 400, 500]	[200, 300, 400, 500, 600, 700]

TABLE VII
SELECTED HYPERPARAMETERS AND MODEL SETTINGS FOR PRETRAINED MODELS.

Hyperparameters	Tanzanian models		Norwegian models	
	PAR U-Net	cGAN U-NET	PAR U-Net	cGAN U-NET
Batch size (BS)	2	2	8	8
$\beta 1$ (Adam)	0.7	0.7	0.8	0.8
Learning rate (lr)	0.0001	0.001	0.0001	0.001
Encoder Network	ResNet34	ResNet34	ResNet34	ResNet34
Encoder/decoder depth	5	5	4	4
Discriminator network	—	PixelGAN	—	PatchGAN(16)
Normalisation	None	Instance	Instance	Batch
α	0	0.01	0	0.001
True target weight δ	500	300	200	400
Epochs	150	150	200	200

TABLE VIII
SELECTED HYPERPARAMETERS AND MODEL SETTINGS FOR FINE-TUNED MODELS.

	Fine-tuned Tanzanian models				Fine-tuned pan-regional Norwegian models			
Pretraining: / Fine-tuning: / Hyperpar.	PAR U-Net $\mathcal{L}_{cGAN}$	PAR U-Net $\mathcal{L}_1+\mathcal{L}_{FFT}$	PAR U-Net $\mathcal{L}_{cGAN}+\mathcal{L}_{FFT}$	cGAN U-Net $\mathcal{L}_{cGAN}+\mathcal{L}_{FFT}$	PAR U-Net $\mathcal{L}_{cGAN}$	PAR U-Net $\mathcal{L}_1+\mathcal{L}_{FFT}$	PAR U-Net $\mathcal{L}_{cGAN}+\mathcal{L}_{FFT}$	cGAN U-Net $\mathcal{L}_{cGAN}+\mathcal{L}_{FFT}$
Batch size (BS)	2	2	2	2	8	8	8	8
$\beta 1$ (Adam)	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.8
Learning rate (lr)	0.0001	0.0001	0.0001	0.001	0.0001	0.0001	0.0001	0.001
Encoder Network	ResNet34	ResNet34	ResNet34	ResNet34	ResNet34	ResNet34	ResNet34	ResNet34
Encoder/decoder depth	5	5	5	5	4	4	4	4
Discriminator network	PixelGAN	—	PixelGAN	PixelGAN	PatchGAN(16)	—	PatchGAN(16)	PatchGAN(16)
Normalisation	Instance	None	Instance	Instance	Instance	Instance	Instance	Batch
α	0.01	0	0.01	0.01	0.01	0	0.01	0.01
γ	0	9e-8	1e-7	1e-7	0	9e-8	1e-7	1e-7
True target weight δ	400	200	500	400	700	700	200	700
Epochs	150+250	150+100	150+50	150+100	300+100	300+100	200+150	300+50

Bibliography

[1] C. Elachi and J. J. Van Zyl, *Introduction to the physics and techniques of remote sensing*. John Wiley & Sons, 2nd ed., 2006.

[2] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*. Guilford Press, 5th ed., 2011.

[3] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 173, pp. 24–49, 2021. Publisher: Elsevier.

[4] A. Hamedianfar, C. Mohamedou, A. Kangas, and J. Vauhkonen, "Deep learning for forest inventory and planning: a critical review on the remote sensing approaches so far and prospects for further applications," *Forestry: An International Journal of Forest Research*, vol. 95, pp. 451–465, Feb. 2022. https://academic.oup.com/forestry/article-pdf/95/4/451/45293980/cpac002.pdf.

[5] L. T. Luppino, M. Kampffmeyer, F. M. Bianchi, G. Moser, S. B. Serpico, R. Jenssen, and S. N. Anfinsen, "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–22, 2021.

[6] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice Hall, 2000.

[7] S. Kaasalainen, M. Holopainen, M. Karjalainen, M. Vastaranta, V. Kankare, K. Karila, and B. Osmanoglu, "Combining Lidar and Synthetic Aperture Radar Data to Estimate Forest Biomass: Status and Prospects," *Forests*, vol. 6, pp. 252–270, Jan. 2015.

[8] A. Bombelli, V. Avitabile, H. Balzter, L. Belelli Marchesini, M. Bernoux, M. Brady, R. Hall, M. Hansen, M. Henry, M. Herold, A. Janetos, B. Law, R. Manlay, L. Marklund, H. Olsson, D. Pandey, M. Saket, C. Schmullius,

R. Sessa, and M. Wulder, *BIOMASS Assessment of the status of the development of the standards for the Terrestrial Essential Climate Variables*. Jan. 2009.

[9] T. Johansson, "Biomass production of Norway spruce (Picea abies (L.) Karst.) growing on abandoned farmland," *Silva Fennica*, vol. 33, no. 4, pp. 261–280, 1999. Publisher: The Finnish Society of Forest Science.

[10] K. Ericsson, S. Huttunen, L. J. Nilsson, and P. Svenningsson, "Bioenergy policy and market development in Finland and Sweden," *Energy Policy*, vol. 32, no. 15, pp. 1707–1721, 2004.

[11] M. Segura and M. Kanninen, "Allometric models for tree volume and total aboveground biomass in a tropical humid forest in Costa Rica 1," *Biotropica: The Journal of Biology and Conservation*, vol. 37, no. 1, pp. 2–8, 2005. Publisher: Wiley Online Library.

[12] J. Urban, J. Čermák, and R. Ceulemans, "Above-and below-ground biomass, surface and volume, and stored water in a mature Scots pine stand," *European journal of forest research*, vol. 134, pp. 61–74, 2015. Publisher: Springer.

[13] L. G. Marklund, "Biomass functions for pine, spruce and birch in Sweden," Tech. Rep. 45, Department of Forest Survey., Swedish University of Agricultural Sciences, Umeå, Sweden, 1988.

[14] D. Lu, Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran, "A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems," *International Journal of Digital Earth*, vol. 9, pp. 63–105, Jan. 2016.

[15] A. I. Flores-Anderson, K. E. Herndon, R. B. Thapa, and E. Cherrington, "The SAR handbook: comprehensive methodologies for forest monitoring and biomass estimation," tech. rep., 2019.

[16] M. Maltamo, E. Næsset, and J. Vauhkonen, "Forestry applications of airborne laser scanning," *Concepts and case studies. Manag For Ecosys*, vol. 27, p. 460, 2014. Publisher: Springer.

[17] S. Björk, S. N. Anfinsen, E. Næsset, T. Gobakken, and E. Zahabu, "On the Potential of Sequential and Nonsequential Regression Models for Sentinel-1-Based Biomass Prediction in Tanzanian Miombo Forests," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4612–4639, 2022.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[19] E. Tomppo, R. Malimbwi, M. Katila, K. Mäkisara, H. M. Henttonen, N. Chamuya, E. Zahabu, and J. Otieno, "A sampling design for a large area forest inventory: case Tanzania," *Canadian Journal of Forest Research*, vol. 44, no. 8, pp. 931–948, 2014. Publisher: NRC Research Press.

[20] J. B. Drake, R. O. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, R. L. Chazdon, J. F. Weishampel, and S. Prince, "Estimation of tropical forest structural characteristics using large-footprint lidar," *Remote Sensing of Environment*, vol. 79, no. 2, pp. 305–319, 2002.

[21] G. Galidaki, D. Zianis, I. Gitas, K. Radoglou, V. Karathanassi, M. Tsakiri–Strati, I. Woodhouse, and G. Mallinis, "Vegetation biomass estimation with remote sensing: focus on forest and other wooded land over the Mediterranean ecosystem," *International Journal of Remote Sensing*, vol. 38, pp. 1940–1966, Apr. 2017.

[22] L. T. Ene, E. Næsset, T. Gobakken, E. W. Mauya, O. M. Bollandsås, T. G. Gregoire, G. Ståhl, and E. Zahabu, "Large-scale estimation of aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data," *Remote Sensing of Environment*, vol. 186, pp. 626–636, Dec. 2016.

[23] L. Noordermeer, O. M. Bollandsås, H. O. Ørka, E. Næsset, and T. Gobakken, "Comparing the accuracies of forest attributes predicted from airborne laser scanning and digital aerial photogrammetry in operational forest inventories," *Remote Sensing of Environment*, vol. 226, pp. 26–37, 2019.

[24] Y. Wang, L. Cai, D. Zhang, and S. Huang, "The frequency discrepancy between real and generated images," *IEEE Access: Practical Innovations, Open Solutions*, vol. 9, pp. 115205–115216, 2021. Publisher: IEEE.

[25] M. Khayatkhoei and A. Elgammal, "Spatial Frequency Bias in Convolutional Generative Adversarial Networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 7152–7159, June 2022.

[26] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks," in *International Conference on Machine Learning*, pp. 5301–5310, PMLR, 2019.

[27] S. Zolkos, S. Goetz, and R. Dubayah, "A meta-analysis of terrestrial above-

ground biomass estimation using Lidar remote sensing," *Remote Sensing of Environment*, vol. 128, pp. 289–298, Jan. 2013.

[28] E. Næsset, H. O. Ørka, S. Solberg, O. M. Bollandsås, E. H. Hansen, E. Mauya, E. Zahabu, R. Malimbwi, N. Chamuya, H. Olsson, and T. Gobakken, "Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision," *Remote Sensing of Environment*, vol. 175, pp. 282–300, Mar. 2016.

[29] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2014.

[30] S. Englhart, V. Keuck, and F. Siegert, "Aboveground biomass retrieval in tropical forests — The potential of combined X- and L-band SAR data use," *Remote Sensing of Environment*, vol. 115, pp. 1260–1271, May 2011.

[31] C. Toth and G. Jóźków, "Remote sensing platforms and sensors: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 22–36, 2016.

[32] T. Le Toan, A. Beaudoin, J. Riom, and D. Guyon, "Relating forest biomass to SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 2, pp. 403–411, 1992. Publisher: IEEE.

[33] T. Le Toan, G. Picard, J.-M. Martinez, P. Melon, and M. Davidson, "On the relationships between radar measurements and forest structure and biomass," *ESASP*, vol. 475, pp. 3–12, 2002.

[34] M. C. Dobson, F. T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke, and N. Christensen, "Dependence of radar backscatter on coniferous forest biomass," *IEEE Transactions on Geoscience and remote Sensing*, vol. 30, no. 2, pp. 412–415, 1992. Publisher: IEEE.

[35] J. Awange, J. Kiema, J. Awange, and J. Kiema, "Light Detection And Ranging (LiDAR)," *Environmental Geoinformatics: Extreme Hydro-Climatic and Food Security Challenges: Exploiting the Big Data*, pp. 291–306, 2019. Publisher: Springer.

[36] P. Rodríguez-Veiga, J. Wheeler, V. Louis, K. Tansey, and H. Balzter, "Quantifying forest biomass carbon stocks from space," *Current Forestry Reports*,

vol. 3, pp. 1–18, 2017. Publisher: Springer.

[37] Y. Zhang, T. Liu, M. Long, and M. Jordan, "Bridging theory and algorithm for domain adaptation," in *International conference on machine learning*, pp. 7404–7413, 2019.

[38] M. Urbazaev, C. Thiel, F. Cremer, R. Dubayah, M. Migliavacca, M. Reichstein, and C. Schmullius, "Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico," *Carbon Balance and Management*, vol. 13, p. 5, Feb. 2018.

[39] N. Ghasemi, M. R. Sahebi, and A. Mohammadzadeh, "A review on biomass estimation methods using synthetic aperture radar data," *International Journal of Geomatics and Geosciences*, vol. 1, no. 4, pp. 776–788, 2011. Publisher: Integrated Publishing Association.

[40] S. Abbas, M. S. Wong, J. Wu, N. Shahzad, and S. Muhammad Irteza, "Approaches of Satellite Remote Sensing for the Assessment of Above-Ground Biomass across Tropical Forests: Pan-tropical to National Scales," *Remote Sensing*, vol. 12, no. 20, 2020.

[41] J. Chave, C. Andalo, S. Brown, M. A. Cairns, J. Q. Chambers, D. Eamus, H. Fölster, F. Fromard, N. Higuchi, T. Kira, and others, "Tree allometry and improved estimation of carbon stocks and balance in tropical forests," *Oecologia*, vol. 145, pp. 87–99, 2005. Publisher: Springer.

[42] 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, and others, "Improved allometric models to estimate the aboveground biomass of tropical trees," *Global change biology*, vol. 20, no. 10, pp. 3177–3190, 2014. Publisher: Wiley Online Library.

[43] J. García-Gutiérrez, E. González-Ferreiro, D. Mateos-García, and J. C. Riquelme-Santos, "A preliminary study of the suitability of deep learning to improve LiDAR-derived biomass estimation," in *Hybrid artificial intelligent systems: 11th international conference, HAIS 2016, seville, spain, april 18-20, 2016, proceedings 11*, pp. 588–596, Springer, 2016.

[44] M.-H. Phua, S. A. Johari, O. C. Wong, K. Ioki, M. Mahali, R. Nilus, D. A. Coomes, C. R. Maycock, and M. Hashim, "Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests," *Forest Ecology and Management*, vol. 406, pp. 163–171, 2017.

[45] S. Sinha, S. Mohan, A. K. Das, L. K. Sharma, C. Jeganathan, A. Santra, S. S. Mitra, and M. S. Nathawat, "Multi-sensor approach integrating optical and multi-frequency synthetic aperture radar for carbon stock estimation over a tropical deciduous forest in India," *Carbon Management*, vol. 11, no. 1, pp. 39–55, 2020. Publisher: Taylor & Francis tex.eprint: https://doi.org/10.1080/17583004.2019.1686931.

[46] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data," *Remote Sensing*, vol. 11, no. 12, p. 1459, 2019.

[47] L. Chen, Y. Wang, C. Ren, B. Zhang, and Z. Wang, "Optimal combination of predictors and algorithms for forest above-ground biomass mapping from Sentinel and SRTM data," *Remote Sensing*, vol. 11, no. 4, p. 414, 2019.

[48] Y. Li, M. Li, C. Li, and Z. Liu, "Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms," *Scientific Reports*, vol. 10, p. 9952, June 2020.

[49] N. Nuthammachot, A. Askar, D. Stratoulias, and P. Wicaksono, "Combined use of Sentinel-1 and Sentinel-2 data for improving above-ground biomass estimation," *Geocarto International*, vol. 37, no. 2, pp. 366–376, 2022. doi:10.1080/10106049.2020.1726507 Publisher: Taylor & Francis.

[50] S. Sinha, "Assessment of vegetation vigor using integrated synthetic aperture radars," in *Remote sensing and GIScience*, pp. 35–58, Springer, 2021.

[51] C. S. Neigh, R. F. Nelson, K. J. Ranson, H. A. Margolis, P. M. Montesano, G. Sun, V. Kharuk, E. Næsset, M. A. Wulder, and H.-E. Andersen, "Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR," *Remote Sensing of Environment*, vol. 137, pp. 274–287, 2013.

[52] S. Solberg, R. Astrup, T. Gobakken, E. Næsset, and D. J. Weydahl, "Estimating spruce and pine biomass with interferometric X-band SAR," *Remote Sensing of Environment*, vol. 114, pp. 2353–2360, Oct. 2010.

[53] R. Gupta and L. K. Sharma, "Mixed tropical forests canopy height mapping from spaceborne LiDAR GEDI and multisensor imagery using machine learning models," *Remote Sensing Applications: Society and Environment*, vol. 27, p. 100817, 2022.

[54] D. Wang, B. Wan, J. Liu, Y. Su, Q. Guo, P. Qiu, and X. Wu, "Estimating

aboveground biomass of the mangrove forests on northeast Hainan Island in China using an upscaling method from field plots, UAV-LiDAR data and Sentinel-2 imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 85, p. 101986, 2020.

[55] A. T. Hudak, P. A. Fekety, V. R. Kane, R. E. Kennedy, S. K. Filippelli, M. J. Falkowski, W. T. Tinkham, A. M. S. Smith, N. L. Crookston, G. M. Domke, M. V. Corrao, B. C. Bright, D. J. Churchill, P. J. Gould, R. J. McGaughey, J. T. Kane, and J. Dong, "A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA," *Environmental Research Letters*, vol. 15, p. 095003, Aug. 2020. Publisher: IOP Publishing.

[56] D. Wang, B. Wan, P. Qiu, Z. Zuo, R. Wang, and X. Wu, "Mapping height and aboveground biomass of mangrove forests on Hainan island using UAV-LiDAR sampling," *Remote Sensing*, vol. 11, no. 18, p. 2156, 2019.

[57] O. Cartus, J. Kellndorfer, M. Rombach, and W. Walker, "Mapping canopy height and growing stock volume using airborne Lidar, ALOS PALSAR and landsat ETM+," *Remote Sensing*, vol. 4, no. 11, pp. 3320–3345, 2012.

[58] P. M. López-Serrano, C. A. López-Sánchez, J. G. Álvarez González, and J. García-Gutiérrez, "A Comparison of Machine Learning Techniques Applied to Landsat-5 TM Spectral Data for Biomass Estimation," *Canadian Journal of Remote Sensing*, vol. 42, pp. 690–705, Nov. 2016. Publisher: Taylor & Francis.

[59] S. M. Ghosh and M. D. Behera, "Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest," *Applied Geography*, vol. 96, no. 1, pp. 29–40, 2018.

[60] M. A. Stelmaszczuk-Górska, M. Urbazaev, C. Schmullius, and C. Thiel, "Estimation of above-ground biomass over boreal forests in siberia using updated in situ, ALOS-2 PALSAR-2, and RADARSAT-2 data," *Remote Sensing*, vol. 10, no. 10, p. 1550, 2018.

[61] A. Debastiani, C. Sanquetta, A. Corte, N. Pinto, and F. Rex, "Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest," *Annals of Forest Research*, vol. 62, no. 2, pp. 109–122, 2019.

[62] E. Santi, S. Paloscia, S. Pettinato, G. Cuozzo, A. Padovano, C. Notarnicola, and C. Albinet, "Machine-learning applications for the retrieval of forest biomass from airborne p-band SAR data," *Remote Sensing*, vol. 12, no. 5,

p. 804, 2020.

[63] Y. Zhang, J. Ma, S. Liang, X. Li, and M. Li, "An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products," *Remote Sensing*, vol. 12, no. 24, p. 4015, 2020.

[64] L. Chen, C. Ren, B. Zhang, Z. Wang, and Y. Xi, "Estimation of forest aboveground biomass by geographically weighted regression and machine learning with Sentinel imagery," *Forests*, vol. 9, no. 10, p. 582, 2018.

[65] S. Vafaei, J. Soosani, K. Adeli, H. Fadaei, H. Naghavi, T. D. Pham, and D. Tien Bui, "Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: A case study of the Hyrcanian forest area (Iran)," *Remote Sensing*, vol. 10, no. 2, p. 172, 2018.

[66] L. Chen, Y. Wang, C. Ren, B. Zhang, and Z. Wang, "Assessment of multi-wavelength SAR and multispectral instrument data for forest aboveground biomass mapping using random forest kriging," *Forest Ecology and Management*, vol. 447, pp. 12–25, 2019.

[67] L. Yang, S. Liang, and Y. Zhang, "A new method for generating a global forest aboveground biomass map from multiple high-level satellite products and ancillary information," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2587–2597, 2020.

[68] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. Publisher: Nature Publishing Group UK London.

[69] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. Publisher: AcM New York, NY, USA.

[70] L. T. Ene, E. Næsset, T. Gobakken, O. M. Bollandsås, E. W. Mauya, and E. Zahabu, "Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data," *Remote Sensing of Environment*, vol. 188, pp. 106–117, Jan. 2017.

[71] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org," 2017. Published: https://www.openstreetmap.org.

[72] C. M. Bishop, *Pattern recognition and machine learning*, vol. 4. Springer, 2006.

[73] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[74] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*. USA: Academic Press, Inc., 4th ed., 2008.

[75] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: With applications in R*, vol. 103 of *Springer texts in statistics*. New York: Springer, 2013. ISSN: 1431-875X Publication Title: An introduction to statistical learning.

[76] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY: Springer New York, 2013.

[77] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation.," in *Encyclopedia of Database Systems* (L. Liu and M. T. Özsu, eds.), pp. 532–538, Springer US, 2009.

[78] D. Jakhar and I. Kaur, "Artificial intelligence, machine learning and deep learning: definitions and differences," *Clinical and experimental dermatology*, vol. 45, no. 1, pp. 131–132, 2020. Publisher: Blackwell Publishing Ltd Oxford, UK.

[79] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *2017 15th international conference on ICT and knowledge engineering (ICT&KE)*, pp. 1–6, IEEE, 2017.

[80] S. J. Russell, *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

[81] T. Glasmachers, "Limits of End-to-End Learning," in *Proceedings of the Ninth Asian Conference on Machine Learning* (M.-L. Zhang and Y.-K. Noh, eds.), vol. 77 of *Proceedings of Machine Learning Research*, (Yonsei University, Seoul, Republic of Korea), pp. 17–32, PMLR, Nov. 2017.

[82] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 315–323, JMLR Workshop and Conference Proceedings, 2011.

[83] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[84] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, pmlr, 2015.

[85] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[86] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[87] R. Durall, M. Keuper, and J. Keuper, "Watch your Up-Convolution: CNN based Generative Deep Neural Networks are Failing to Reproduce Spectral Distributions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7890–7899, 2020.

[88] K. Chandrasegaran, N.-T. Tran, and N.-M. Cheung, "A Closer Look at Fourier Spectrum Discrepancies for CNN-generated Images Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7200–7209, 2021.

[89] 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)*, (Los Alamitos, CA, USA), pp. 770–778, IEEE Computer Society, June 2016. ISSN: 1063-6919.

[90] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[91] J. M. Tomczak, *Deep generative modeling*. Springer, 2022.

[92] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th international conference on neural information processing systems - volume 2*, NIPS'14, (Cambridge, MA, USA), pp. 2672–2680, MIT Press, 2014. 10.5555/2969033.2969125 Number of pages: 9

Place: Montreal, Canada.

[93] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *Proc. of IEEE CVPR*, vol. 2019-June, pp. 4396–4405, 2019. ISBN: 9781728132938 _eprint: 1812.04948.

[94] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," *Proc. of IEEE CVPR*, pp. 8107–8116, 2020. _eprint: 1912.04958.

[95] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, 2017.

[96] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse Image Synthesis for Multiple Domains," in *Proc. of IEEE CVPR*, pp. 8188–8197, 2020.

[97] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proc. of IEEE ICCV*, pp. 2223–2232, 2017.

[98] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved Training of Wasserstein GANs," in *Advances in neural information processing systems*, pp. 5767–5777, 2017.

[99] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least Squares Generative Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, IEEE, 2017.

[100] D. P. Kingma and M. Welling, "An Introduction to Variational Autoencoders," *Foundations and trends in machine learning*, vol. 12, no. 4, pp. 307–392, 2019. Place: Boston - Delft Publisher: Now Publishers.

[101] T. R. Andersson, J. S. Hosking, M. Pérez-Ortiz, B. Paige, A. Elliott, C. Russell, S. Law, D. C. Jones, J. Wilkinson, T. Phillips, J. Byrne, S. Tietsche, B. B. Sarojini, E. Blanchard-Wrigglesworth, Y. Aksenov, R. Downie, and E. Shuckburgh, "Seasonal Arctic sea ice forecasting with probabilistic deep learning," *Nature Communications*, vol. 12, p. 5124, Aug. 2021.

[102] A. E. Pascarella, G. Giacco, M. Rigiroli, S. Marrone, and C. Sansone, "ReUse: REgressive Unet for Carbon Storage and Above-Ground Biomass Estimation," *Journal of Imaging*, vol. 9, no. 3, 2023.

[103] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM Challenge on Perceptual Image Super-Resolution," in *European Conference on Computer Vision*, pp. 334–355, Springer, 2018.

[104] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *European Conference on Computer Vision*, pp. 63–79, Springer, 2018.

[105] Y. Chen, G. Li, C. Jin, S. Liu, and T. Li, "SSD-GAN: Measuring the Realness in the Spatial and Spectral Domains," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 1105–1112, 2021. Issue: 2.

[106] S. Björk, J. N. Myhre, and T. Haugland Johansen, "Simpler is Better: Spectral Regularization and Up-Sampling Techniques for Variational Autoencoders," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3778–3782, May 2022. Journal Abbreviation: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[107] S. Czolbe, O. Krause, I. Cox, and C. Igel, "A Loss Function for Generative Neural Networks Based on Watson's Perceptual Model," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[108] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Human vision, visual processing, and digital display IV*, vol. 1913, pp. 202–216, SPIE, 1993.

[109] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, p. 896, 2013. Issue: 2.

[110] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *2020 international joint conference on neural networks (IJCNN)*, pp. 1–8, IEEE, 2020.

[111] S. Ruder and B. Plank, "Strong baselines for neural semi-supervised learning under domain shift," *arXiv preprint arXiv:1804.09530*, 2018.

[112] X. Zhang, Y. Ge, Y. Qiao, and H. Li, "Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3436–3445, 2021.

[113] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of international conference on computer vision (ICCV)*, Dec. 2015.

[114] J. Penman, M. Gytarsky, T. Hiraishi, T. Krug, D. Kruger, R. Pipatti, L. Buendia, K. Miwa, T. Ngara, K. Tanabe, and others, "Good practice guidance for land use, land-use change and forestry.," *Good practice guidance for land use, land-use change and forestry.*, 2003. Publisher: Institute for Global Environmental Strategies.

[115] IPCC, "IPCC Guidelines for national greenhouse gas inventories," tech. rep., IGES, Japan, 2006. Publisher: Prepared by the National Greenhouse Gas Inventories Programme.

[116] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, and others, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021. Publisher: Elsevier.

[117] J. Leonhardt, L. Drees, P. Jung, and R. Roscher, "Probabilistic Biomass Estimation with Conditional Generative Adversarial Networks," in *Pattern Recognition: 44th DAGM German Conference, DAGM GCPR 2022, Konstanz, Germany, September 27–30, 2022, Proceedings*, pp. 479–494, Springer, 2022.

[118] A. Holzinger, P. Biecek, and W. Samek, "Explainable AI methods-A brief overview," in *XxAI-Beyond explainable AI: International workshop, held in conjunction with ICML 2020, july 18, 2020, vienna, austria, revised and extended papers*, vol. 13200, p. 13, Springer Nature, 2022.

[119] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller, "Explaining deep neural networks and beyond: A review of methods and applications," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 247–278, 2021. Publisher: IEEE.