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.

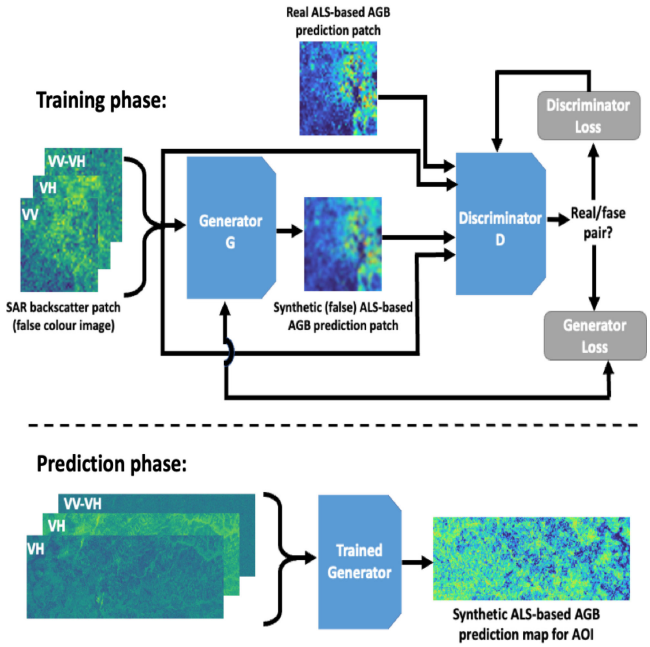

Fig. 1. Flowchart over the proposed cGAN-based sequential modeling approach. The generator network is trained to translate false color Sentinel-1 backscatter patches (consisting of the VV and VH band and their difference, i.e., VV-VH) into realistic-looking synthetic ALS-based AGB prediction patches. 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

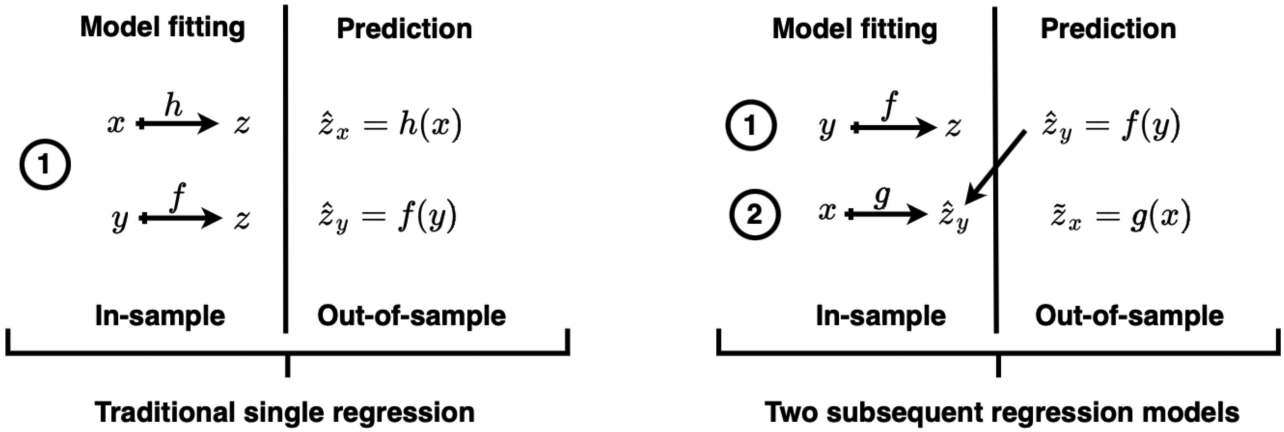

Fig. 2. Illustration of the difference between a traditional nonsequential regression model and the proposed sequential regression models. We let x denote data from a SAR sensor, y denote ALS data, and z denote AGB ground reference data. Regression models are represented by *f*, *g*, and *h*, where *f* 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.

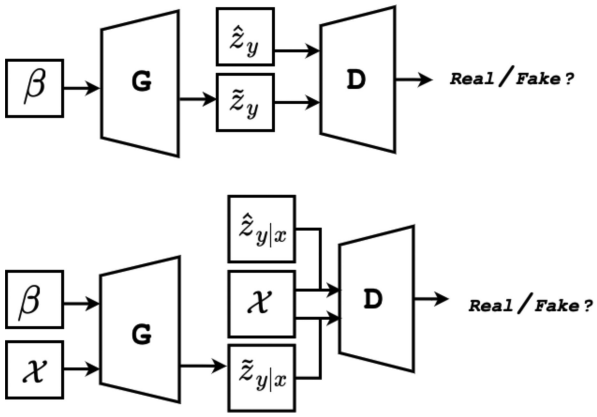

Fig. 3. Illustration of a GAN (upper) and cGAN (lower) model. G and D denote the generator and discriminator networks. $\boldsymbol{x}$ represents images of SAR backscatter intensity from the input domain $\mathcal{X}$. ALS-based AGB predictions in both models are denoted as $\widehat{z_y}$. The subscript 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].

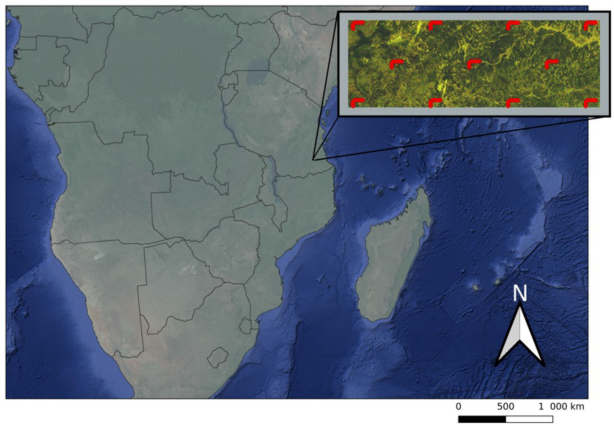

Fig. 4. Location of a subset of the Sentinel-1 scene, as well as the location of the ground reference plots (in red) in the country of Tanzania.

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 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 ($\mathrm{Mg\,ha^{-1}}$).

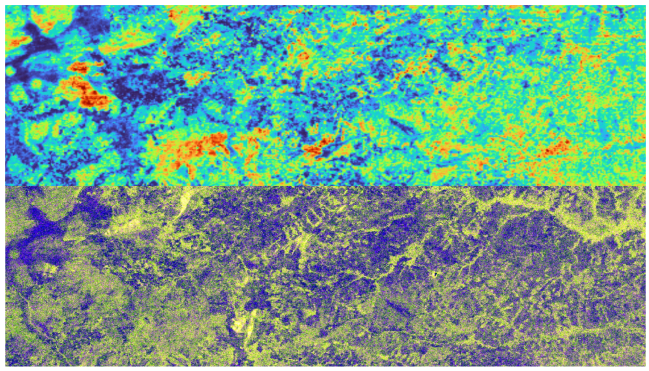

Fig. 5. Top row: ALS-based AGB predictions from [22]. Bottom row: False-color image of the Sentinel-1 dataset.

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

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

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]:

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

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

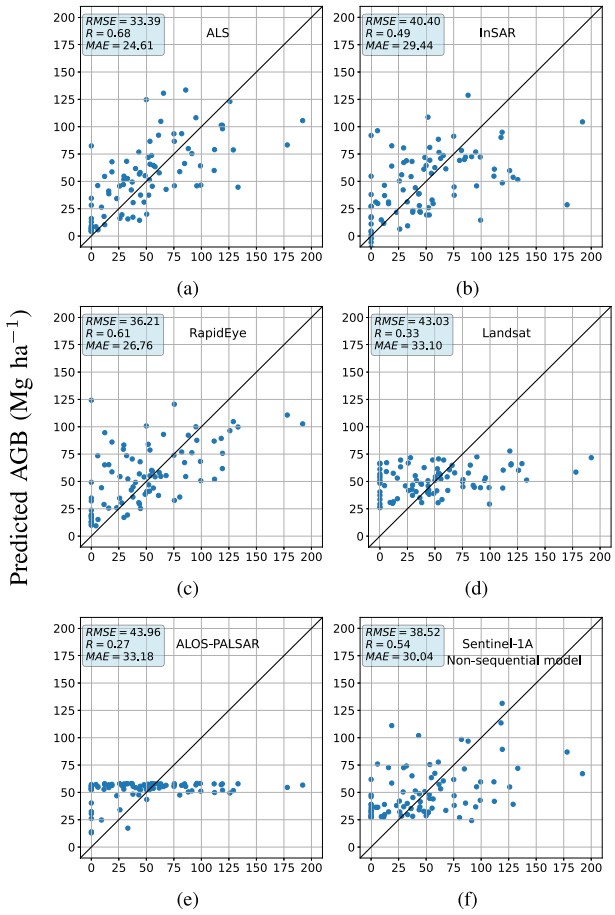

Fig. 6. Scatter plots between ground reference AGB, z, and model-predicted AGB. Model-predicted AGB is retrieved from either (a) the ALS, (b) InSAR, (c) RapidEye, (d) Landsat, (e) PALSAR, or (f) our proposed Sentinel-1-based nonsequential regression model. The black lines are reference lines indicating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

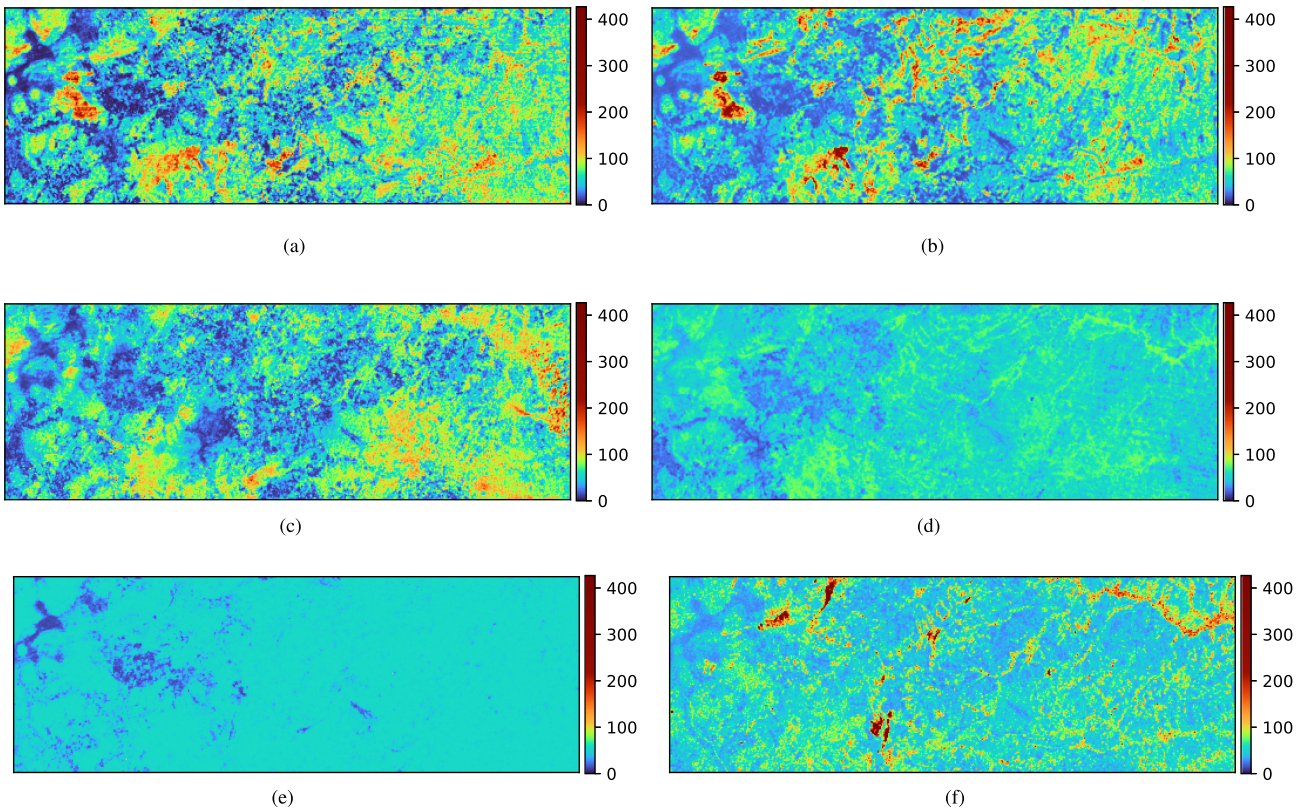

Fig. 7. Aboveground biomass prediction maps (in $\mathrm{Mg\,ha^{-1}}$). (a)–(e) Results of the traditional nonsequential regression models presented in [22]. The AGB biomass map in (f) was constructed from the proposed nonsequential Sentinel-1-based AGB model in Table II. (a) ALS. (b) InSAR. (c) RapidEye.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

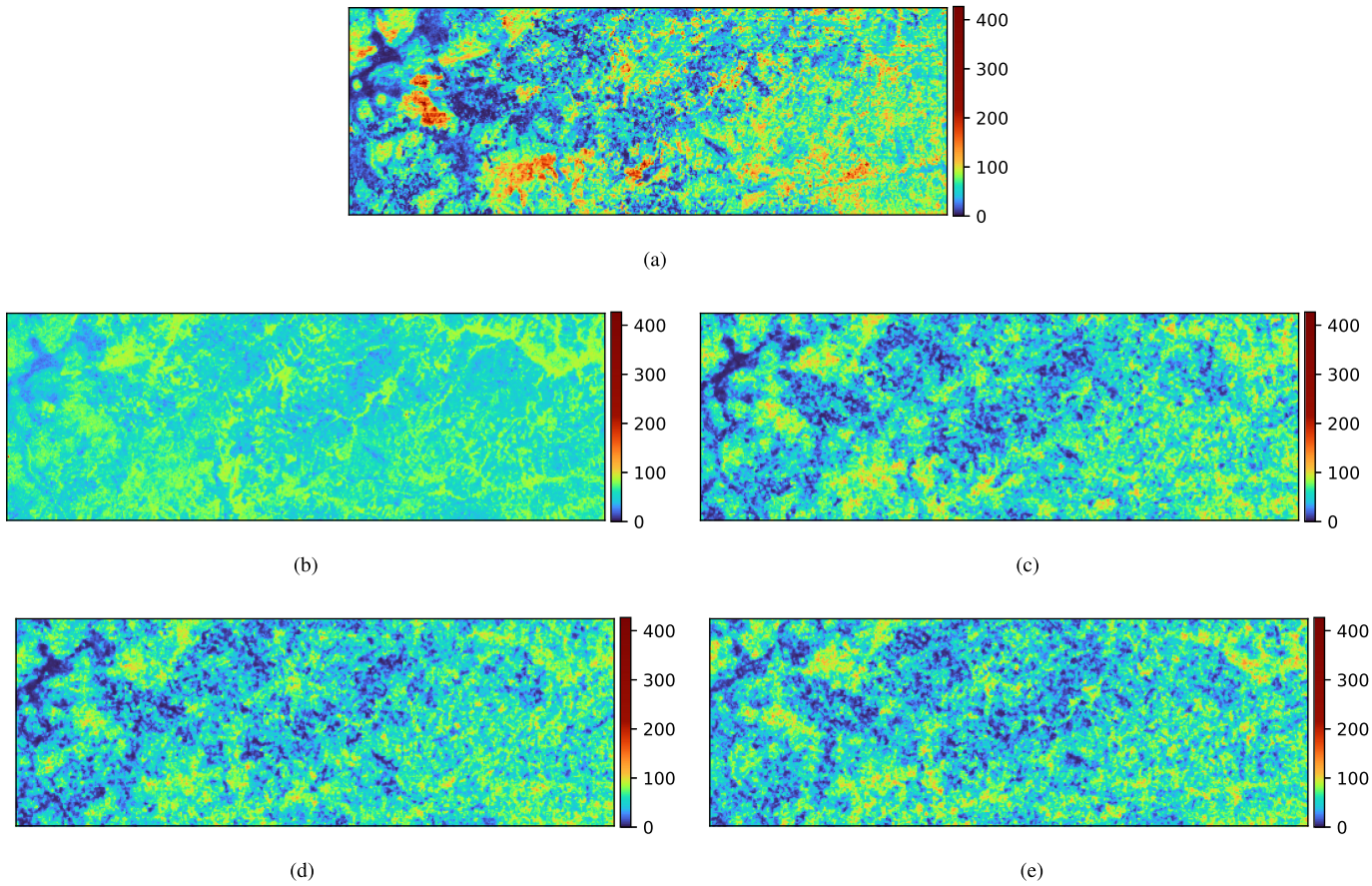

Fig. 8. Generated synthetic ALS-based AGB prediction maps (in $\mathrm{Mg\,ha^{-1}}$) together with the surrogate for ground reference plots, i.e., the ALS-based AGB map shown in (a) [this AGB map is the same as in Fig. 7(a)]. (b) A synthetic ALS-based AGB prediction map generated through the baseline 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.

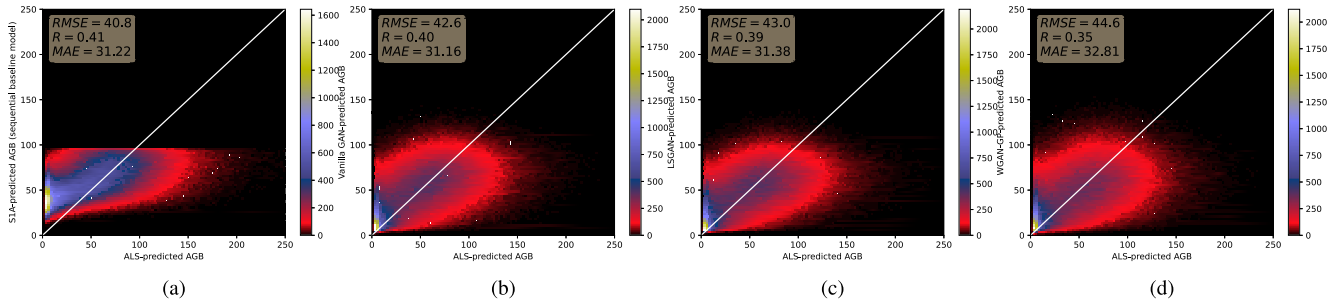

Fig. 9. Density plots between constructed AGB maps and ALS-based AGB biomass predictions, $\widehat{z}_y$, for (a) baseline sequential model, (b) Vanilla GAN, (c) LSGAN, and (d) WGAN-GP models. Reported metrics are the RMSE, Pearson correlation coefficient (R), and the MAE between $\widehat{z}_y$ and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

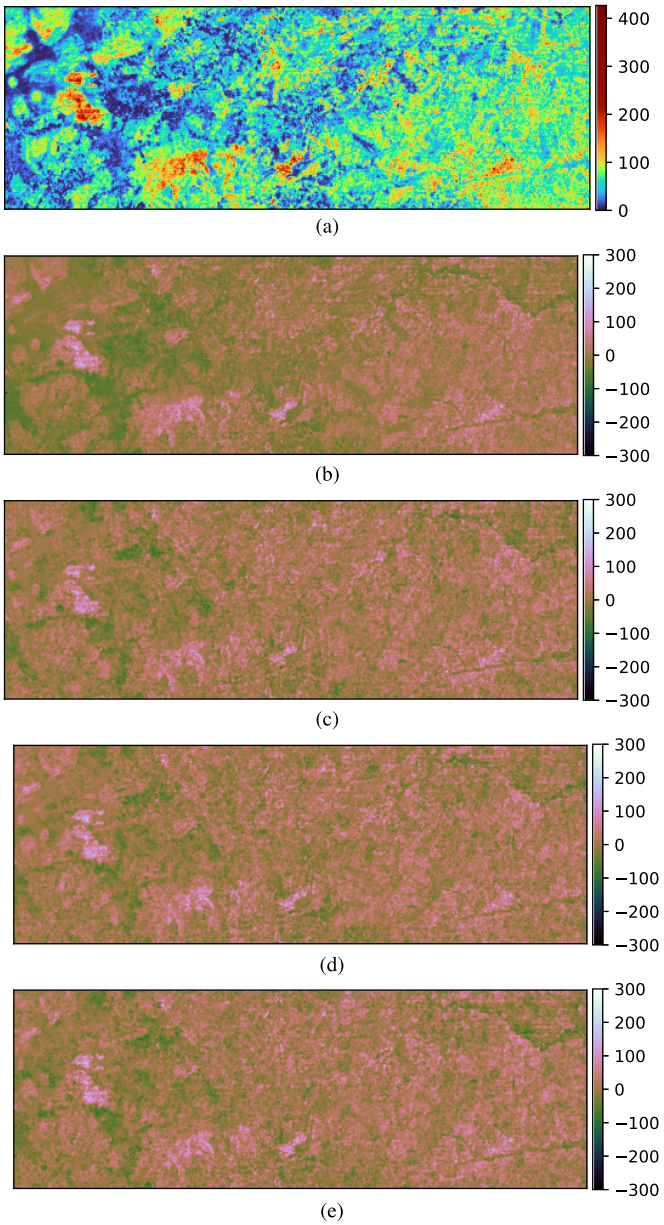

Fig. 10. AGB difference maps (in $\mathrm{Mg\,ha^{-1}}$). Pixel-wise difference between the ALS-based AGB prediction map, shown in **(a)**, and constructed AGB prediction maps achieved form the four sequential models: baseline sequential model **(b)**, Vanilla GAN **(c)**, LSGAN **(d)** and WGAN-GP **(e)**.

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.

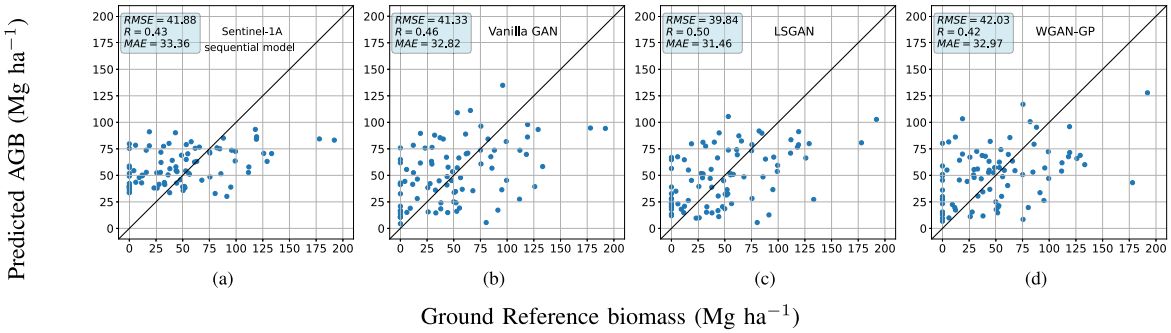

Fig. 11. Scatter plots between ground reference AGB, z, and model-predicted AGB. Model-predicted AGB values are retrieved from the (a) baseline sequential regression model; see Section V-B1, or the proposed cGAN-based sequential models; (b) Vanilla GAN, (c) LSGAN, and (d) WGAN-GP. See Section V-B2 for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

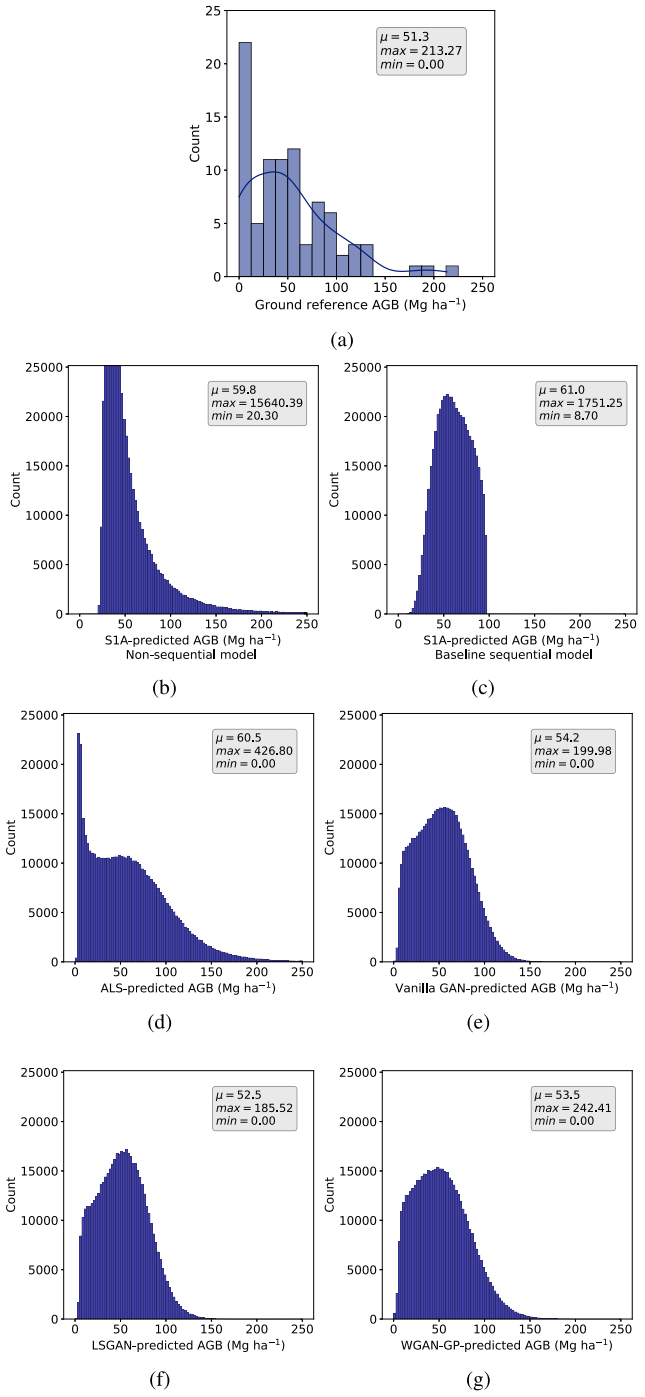

Fig. 12. Histograms of AGB predictions from the proposed AGB models. (b) Nonsequential Sentinel-1. (c) Baseline sequential. (e) Vanilla GAN. (f) LSGAN. (g) WGAN-GP. A histogram over the collected ground reference AGB is shown in (a), while (d) shows a histogram over ALS-based AGB predictions. Reported 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 LSGAN 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 nonvegetated 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.

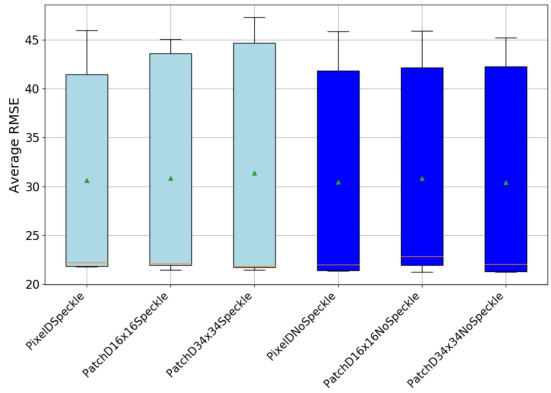

Fig. 13. Boxplot comparison between models trained with different types of D on datasets produced with or without speckle filtering. Green triangles indicate the mean value computed over the five folds, while orange horizontal lines indicate the median.

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

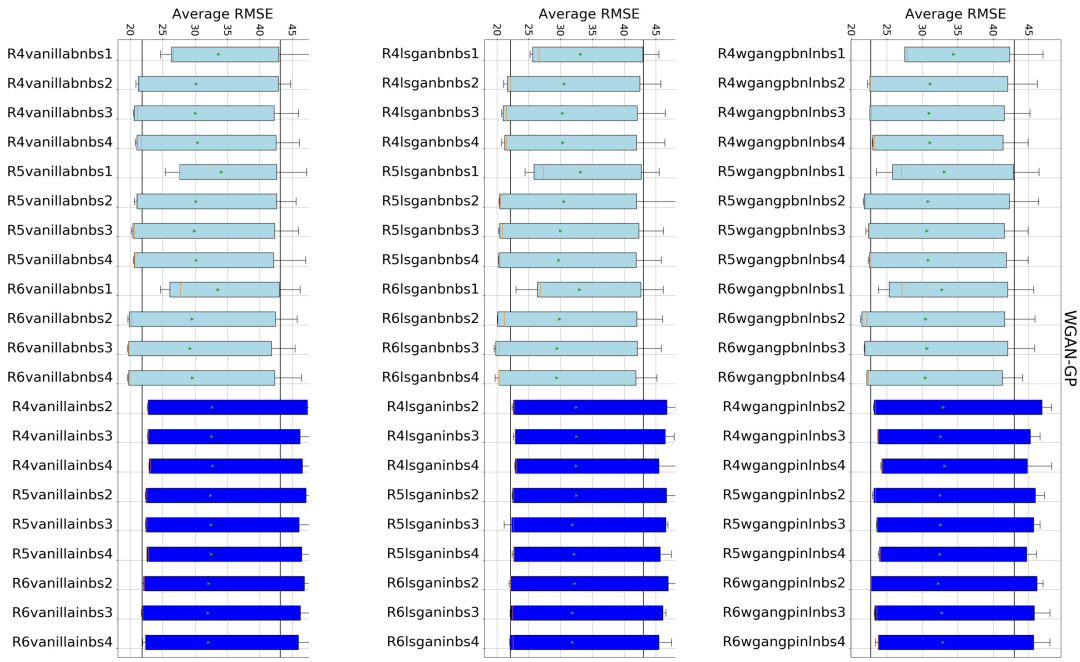

Fig. 14. Impact of normalization method, BN or IN, on model performance. Light blue bins and dark blue bins represent models trained with BS and IN, respectively. Green triangles are mean values computed over the five folds, while orange vertical lines are medians. The two vertical black lines in each 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

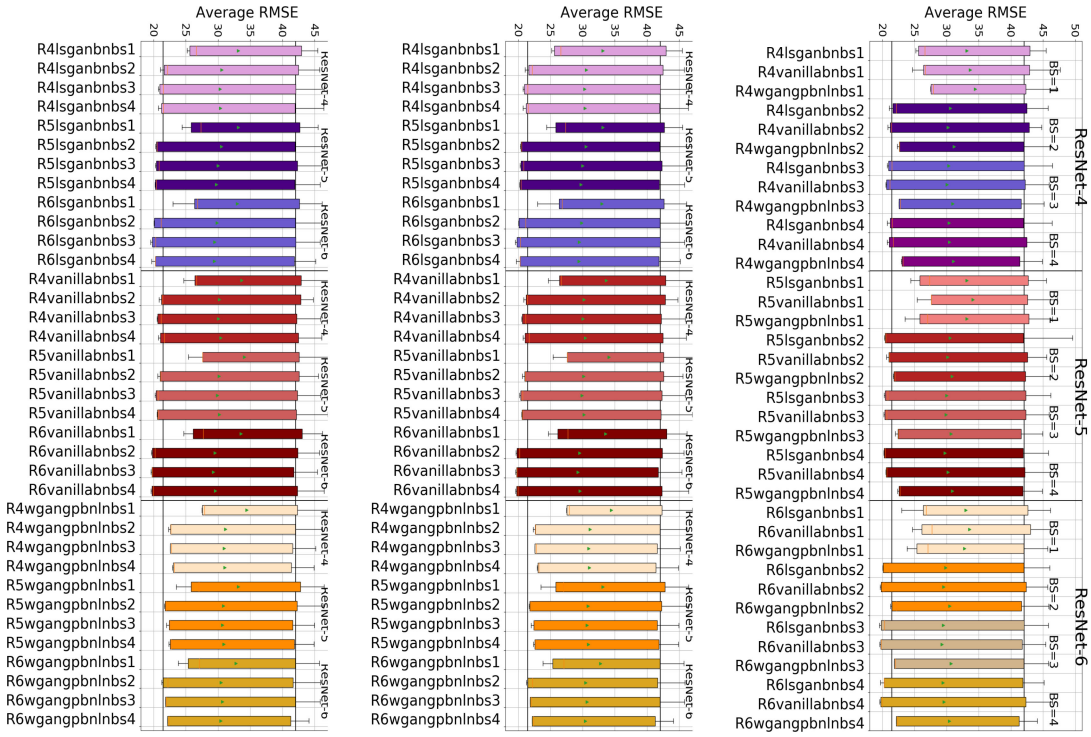

Fig. 15. Boxplot of average RMSE for models trained with all three objective functions, ResNet-4, 5, or 6, with BS varying between 1 and 4 and BN only. Green triangles indicate the mean value computed over the five folds, while orange vertical lines indicate the median. The two vertical black lines are 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.

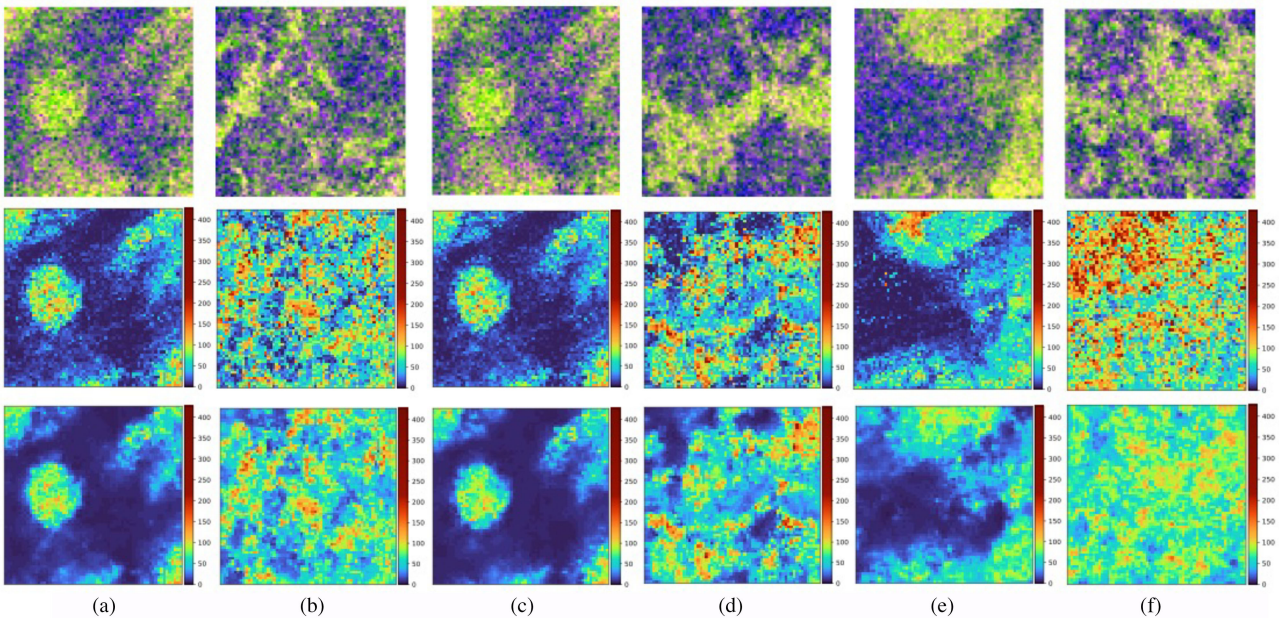

Fig. 16. First row: Sentinel-1 patches. Second row: Target image patches, i.e., ALS-based AGB predictions $\widehat{\boldsymbol{z}}_{\boldsymbol{y}}$. Third row: Generated synthetic image patches, i.e., $\widehat{\boldsymbol{z}}_{\boldsymbol{y}|\boldsymbol{x}}$. Columns (a) and (b): Vanilla GAN; (c) 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 ($\mathrm{Mg\,ha^{-1}}$).

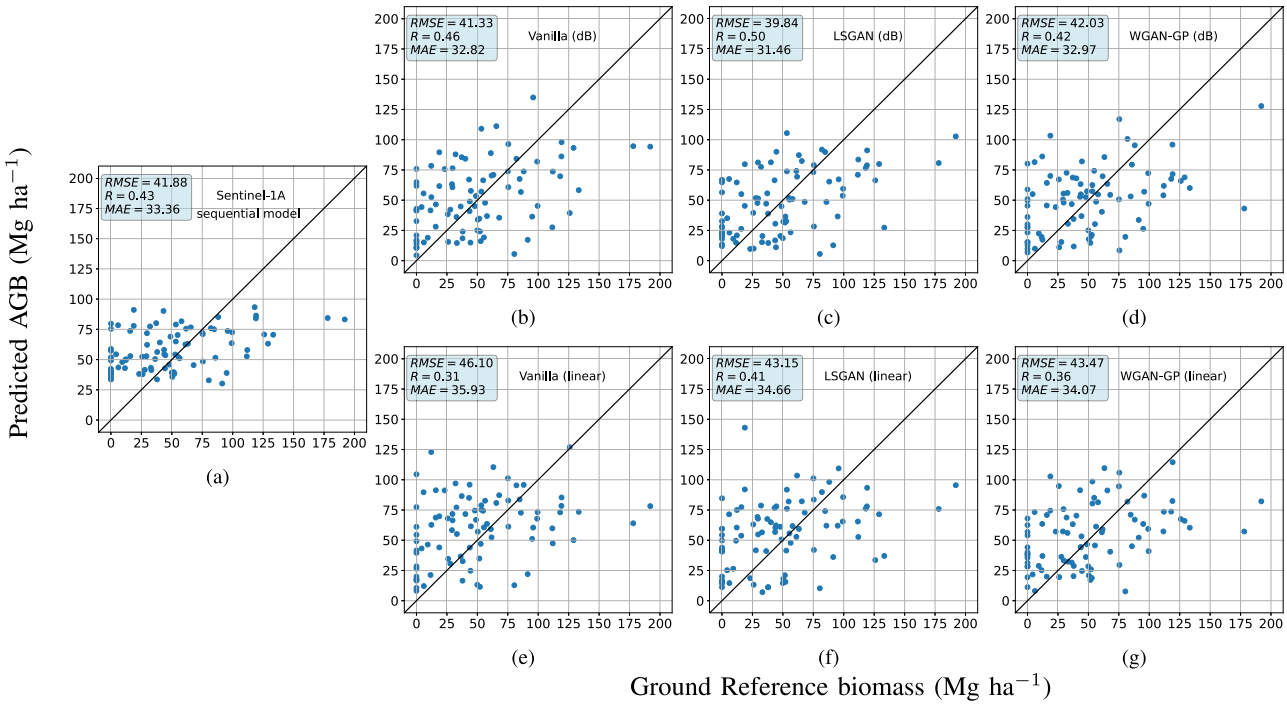

Fig. 17. Scatter plots between AGB ground reference data, z, and model-predicted AGB. Upper row: models trained with Sentinel-1 data on dB scale, i.e., (b) the proposed Vanilla GAN, (c) LSGAN, and (d) WGAN-GP models. Lower row, same models as above, but trained with Sentinel-1 on linear scale. (a) Model-predicted 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.

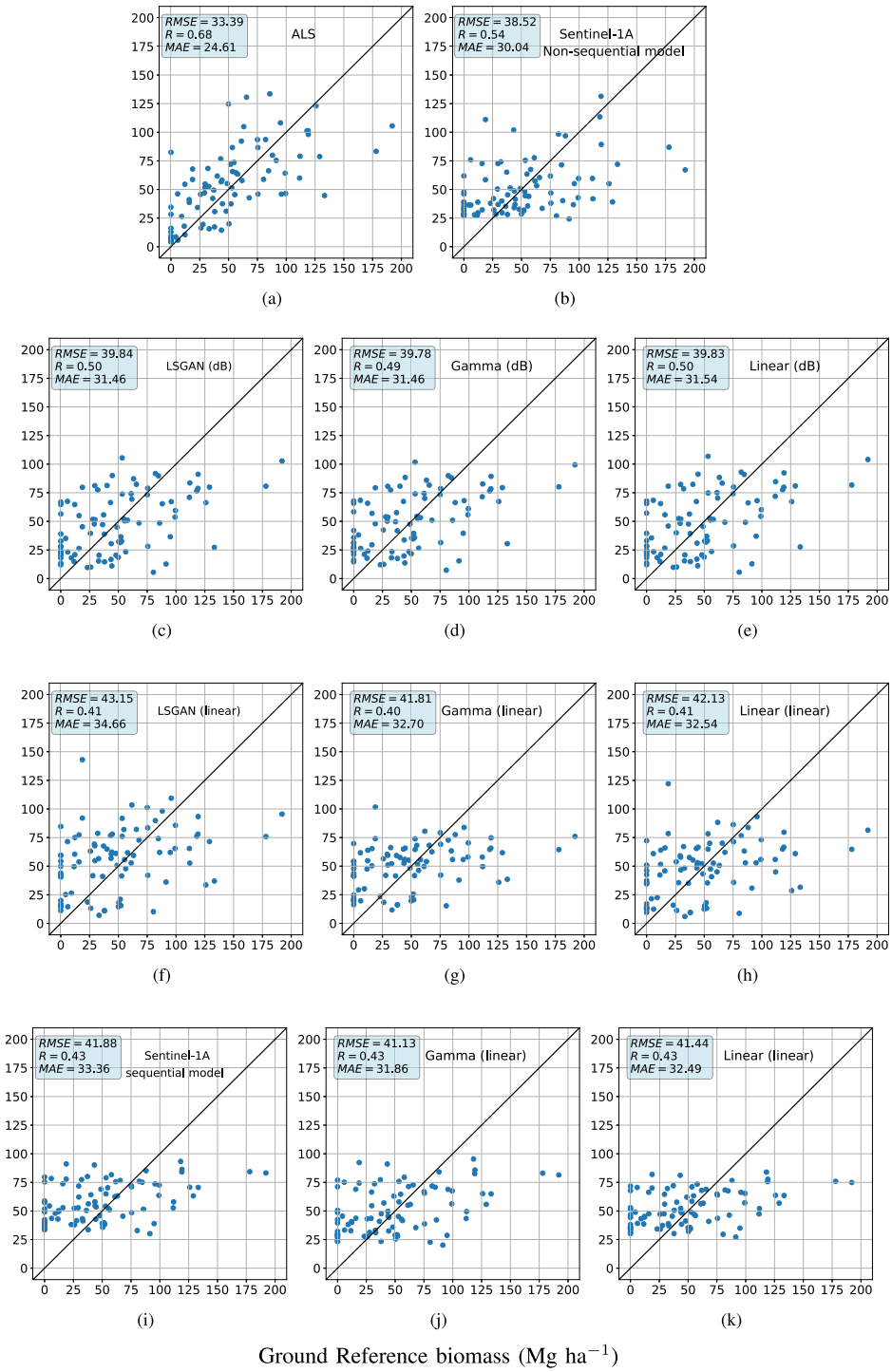

Fig. 18. Scatter plots between predicted AGB and ground reference AGB data, z. (a)–(c), (f), and (i) Reference images, corresponding to AGB predictions from the ALS-based regression model, the nonsequential Sentinel-1 model, the LSGAN model trained with dataset on dB scale, the LSGAN model trained with 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.