GENERATION OF LIDAR-PREDICTED FOREST BIOMASS MAPS FROM RADAR BACKSCATTER WITH CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

Sara Björk[†], *Stian Normann Anfinsen*[†], *Erik Næsset*[‡], *Terje Gobakken*[‡], *Eliakimu Zahabu*[*]

[†]UiT The Arctic University of Norway, Department of Physics and Technology, Tromsø, Norway
[‡]Norwegian University of Life Sciences, Faculty of Environmental Sciences and Natural Resource Management, Ås, Norway
[*]Sokoine University of Agriculture, Department of Forest Resources Assessment and Management, Morogoro, Tanzania

ABSTRACT

This paper studies the generation of LiDAR-predicted aboveground biomass (AGB) maps from synthetic aperture radar (SAR) intensity images by use of conditional generative adversarial networks (cGANs). The purpose is to improve on traditional regression models based on SAR intensity, which are trained with a limited amount of AGB in situ measurements. Although they are costly to collect, data from airborne laser scanning (ALS) sensors are highly correlated with AGB and can replace in situ measurements as the regression target. Thus, the amount of training data increases dramatically, and we can learn an expressive two-stage regression model for SAR backscatter intensity. We propose to model the regression function between SAR intensity and ALS-predicted AGB with a Pix2Pix convolutional neural network for image translation that uses a ResNet-5-based cGAN architecture with the Wasserstein GAN gradient penalty (WGAN-GP) objective function. The synthesized ALS-predicted AGB maps are evaluated qualitatively and quantitatively against real ALS-predicted AGB maps. Our results show that the proposed architecture manages to capture characteristics of the real data, which suggests further use of the ResNet-5 for a SAR intensity regression model of AGB.

1. INTRODUCTION

The REDD+ program (Reducing Emissions from Deforestation and Forest Degradation) was initiated to reduce carbon emissions from tropical forests. As part of this, developing countries have been motivated to implement an efficient measuring, reporting, and verification (MRV) system. Ideally, if deforestation and carbon emission are kept low and this is documented with the MRV system, a financial reward will be released through the REDD+ program [1]. Since aboveground biomass (AGB) is a primary variable related to the carbon cycle [2], enabling accurate estimates of AGB in large areas is a necessary part of the MRV system. For this system, AGB field data is needed, but it is unfortunately both costly and time-demanding to collect manually. As a consequence, the focus instead lies on regression models that establish a relationship between a small amount of AGB field data and remote sensing (RS) measurements from different sensors. Previous analyses on the use of different RS sensors for this task conclude that among different platforms and sensor types, AGB models based on airborne laser scanning (ALS) are significantly more accurate than models developed using radar or passive optical data [3]. This was confirmed in [4], which also states that in tropical forests, ALS data is highly correlated with AGB. However, the high cost of airborne data acquisition limits the use of ALS data in a national MVR system [2]. AGB estimation with spaceborne synthetic aperture radar (SAR) images has, on the other hand, the advantage of providing data with large spatial coverage acquired with high temporal frequency. Unfortunately, SAR suffers from limited estimation accuracy, which restricts the use of SAR data for the MRV system of high precision. The separate challenges of SAR and ALS have fostered studies on their combined use for forest AGB estimation. Several of these studies were reviewed in [2], which concludes that the combination of SAR and ALS may improve AGB estimation, especially when SAR data is used to upscale accurate ALS AGB predictions to large areas.

Artificial neural networks and deep learning has opened a lot of new possibilities in the analysis of RS images. The ability to perform accurate regression between different image modalities, also known as image translation, is one such example. Cross-modal image translation based on generative adversarial networks (GANs) has drawn considerable attention since the architecture was proposed in 2014 [5]. In the standard GAN setting, the generative model learns a mapping from a random noise vector z to an output image y. This idea was later extended to the conditional GAN (cGAN) architecture, where the learned mapping to the output image y is conditioned on an input image x [6]. In this project, we have access to SAR data covering most of Tanzania, while the ALS data only covers a limited part of the country. Motivated by the conclusions of [2] and the achievements of image-to-image translation with cGANs in [6], we propose to train a cGAN to synthesise ALS-predicted AGB maps from SAR intensity images. The cGAN model becomes the second part

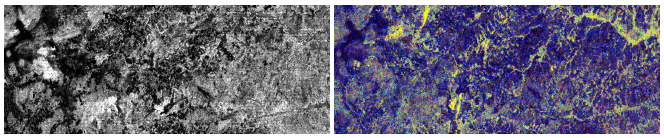


Fig. 1: AGB predictions from ALS data (left) and false-colour SEN1A image of same area (right).

of a two-stage regression model, where the first is the regression model used to produce the ALS-predicted AGB maps from ground reference data, as described in [1, 4]. It is to our knowledge the first time that image-to-image translation has been performed to simulate ALS data from SAR data for AGB regression purposes. SAR image simulation of vehicles was studied in [7]. Our work differs from it in many ways. First of all, we consider different *conditional* GAN networks to simulate ALS data, while [7] considered three ordinary GAN networks to simulate SAR data. Furthermore, [7] performed their study on a benchmark dataset, while we use a real-world dataset. The rest of this paper is organised as follows: Sec. 2 introduces the dataset used in this paper. Sec. 3 describes the preparation of the dataset and the cGAN models used. Results are presented and discussed in Sec. 4, and conclusions are drawn in Sec. 5.

2. DATA AND PREPROCESSING

The SAR data consists of a scene from the Sentinel-1a (SEN1A) sensor, containing two bands in VV and VH polarisation, acquired on 10 September 2015. It was chosen since SEN1A data is freely available and since the scene fulfils our three criteria: 1) it covers our area of interest, 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, this to achieve optimal sensitivity to dynamic AGB levels.

ALS data and ground reference data for 88 field plots were acquired in 2014. A regression model was in [1, 4] trained to predict AGB for pixel cells of size 700m^2, corresponding to a ground resolution of 26.6m. The ALS-predicted AGB maps were made available for this project and will serve as the desired output for the cGAN.

Our SEN1A scene was radiometrically calibrated into σ_0 values, terrain corrected, and resampled to the same pixel size as the ALS-predicted AGB data. These steps and subsetting of the SEN1A scene to match the ALS scene were performed with the ESA SNAP toolbox. Both scenes were co-registered with QGIS. After this process, each pixel in the dataset contains a predicted AGB value and the corresponding measurements in the VV and VH polarisation [dB]. A false-colour image was created from the SEN1A bands with: red = VV, green = VH, and blue = VV-VH. The ALS scene was kept as a greyscale image. Fig. 1 shows the corresponding ALS scene and the false-colour SEN1A scene after preprocessing.

3. METHOD

For the image-to-image translation task, we created training and test sets as follows. Firstly image patches of size 64×64 were extracted from the whole scene in a grid manner with a 10% overlap between the patches. This to obtain as many patches as possible without transferring too much information between patches. Approximately 20% of these patches were extracted to a test set, while the remaining 80% were kept in the training set. Data augmentation, with flipping and rotation, were applied on the separate datasets. This results in 232 patches in the test set and 944 patches in the training set.

3.1. Generation of ALS-predicted AGB image patches

The generation of synthetic ALS-predicted AGB image patches is based on the image-to-image translation framework *Pix2Pix* [6]. In our application, the input domain $\mathcal{X}$ consists of SEN1A patches and the output domain $\mathcal{Y}$ of ALS-based AGB patches. Conditioned on images from the input domain $\mathcal{X}$, the generator network (G) of the cGAN aims to capture the data distribution of the output domain to generate corresponding samples from $\mathcal{Y}$. 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 (real AGB estimates and SEN1A) or fake pair (generated, synthetic AGB estimates and real SEN1A). During adaption of the cGAN, both G and D are trained simultaneously to outperform each other, resulting in the following minmax objective function [5]

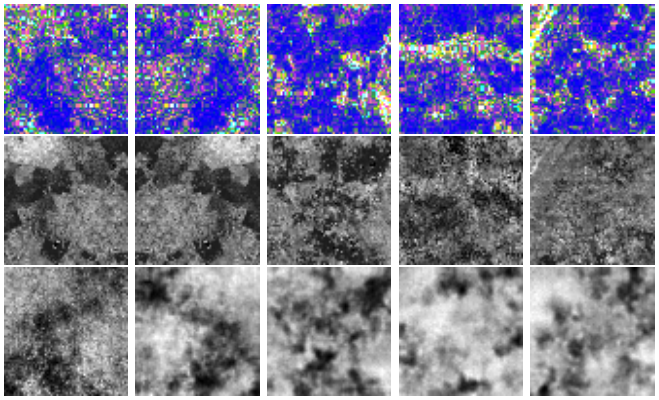
$$\begin{aligned}\min_G \max_D V(D,G) =& \mathbb{E}_{\mathbf{x},\mathbf{y}}[\log D(\mathbf{x},\mathbf{y})] + \\ & \mathbb{E}_{\mathbf{x}}[\log(1 - D(\mathbf{x}, G(\mathbf{x}))].\end{aligned} \tag{1}$$

A cGAN network trained with the objective function in Eq. (1) 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 [8]. Its objective functions in a conditional setting are

$$\begin{aligned}\min_D V_{LSGAN}(D) =& \frac{1}{2}\mathbb{E}_{\mathbf{x},\mathbf{y}}[(D(\mathbf{x},\mathbf{y}) - b)^2] + \\ & \frac{1}{2}\mathbb{E}_{\mathbf{x}}[(D(\mathbf{x}, G(\mathbf{x})) - a)^2] \\ \min_G V_{LSGAN}(G) =& \frac{1}{2}\mathbb{E}_{\mathbf{x}}[(D(\mathbf{x}, G(\mathbf{x})) - c)^2],\end{aligned} \tag{2}$$

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 [8]. Introduced by [9] for further stabilisation of training and high quality image generation, we also consider the WGAN-GP. It consider real data, simulated data and a combination of these in its objective function, which in the conditional setting has the following form [9]

$$\begin{aligned}\min_G \max_D V(D,G) =& \mathbb{E}_{\mathbf{x}}[D(\mathbf{x}, G(\mathbf{x}))] - \\ & \mathbb{E}_{\mathbf{x},\mathbf{y}}[D(\mathbf{x},\mathbf{y})] + \lambda \mathbb{E}_{\hat{\mathbf{y}}}[(||\nabla_{\hat{\mathbf{y}}} D(\hat{\mathbf{y}})||_2 - 1)^2]\end{aligned} \tag{3}$$



(a) (b) (c) (d) (e)

Fig. 2: **First row:** SEN1A patches. **Second row:** Real ALS patches. **Third row:** Generated ALS patches. Column (a) Vanilla GAN, ResNet-5, (b) LSGAN, ResNet-5, (c) WGAN-GP, ResNet-4, (d) WGAN-GP, ResNet-5, (e) WGAN-GP, ResNet-6.

with

$$\hat{\mathbf{y}} = \epsilon \mathbf{y} + (1 - \epsilon)\tilde{\mathbf{y}}\,, \tag{4}$$

where $\mathbf{y}$ is a real image patch of $\mathcal{Y}$ while $\tilde{\mathbf{y}}$ is a generated image patch from $G(\mathbf{x})$.

Generator network: Three different G networks were tested, all based on the ResNet model, i.e. ResNet-4, ResNet-5, and ResNet-6. ResNet-6 is a part of [6] and consists of 2 encoding blocks followed by 6 residual blocks and 2 decoding blocks. ResNet-4 and ResNet-5 consist of the same number of encoder-decoder blocks as ResNet-6 but only 4 or 5 residual blocks. The two smaller networks were proposed as we work with small image patches of 64×64 pixels.

4. RESULTS AND DISCUSSION

For the generation of synthetic ALS-predicted AGB patches from SEN1A data, we train nine different cGAN architectures (combining the three ResNet networks and the three objective functions from Sec. 3) and compare their performance. In our experiments, we apply batch normalisation (BN) for Eq. 1 and Eq. 2, while for Eq. 3 we apply layer normalisation (LN) for D and BN for the G network, as suggested in [9]. We experiment with batch sizes (BS) between 1 and 4 and keep the D network as recommended in [6]. We train the different architectures for 200 epochs, with a learning rate of 2×10^{-4}. After training, the performance of the different models was evaluated by generating synthetic ALS-based AGB using SEN1A data from the test set and comparing it to the real AGB predictions. We evaluate the result by computing the root mean square error (RMSE) averaged over the 232 patches in the test set, and refer to it as the average RMSE.

To evaluate the risk of overfitting while training the cGAN, we performed two separate training regimes: one where the model train on the whole training set during each epoch and another where the model utilises a random sample containing 60% of the whole training set in each epoch. By evaluating the average RMSE on the test set for the two regimes we found that the average RMSE decreased while using only 60% of the whole training set per each epoch. Thus, all results presented in this section follow the second training regime.

Average RMSE [Mgha^{-1}]		
Vanilla GAN (BN,BS=3,ResNet-5)	LSGAN (BN,BS=3,ResNet-5)	WGAN-GP (BN,LN,BS=4,ResNet-5)
68.1	57.9	**56.7**
WGAN-GP (BN, LN, BS=4)		
ResNet-4	ResNet-5	ResNet-6
57.0	**56.7**	56.8

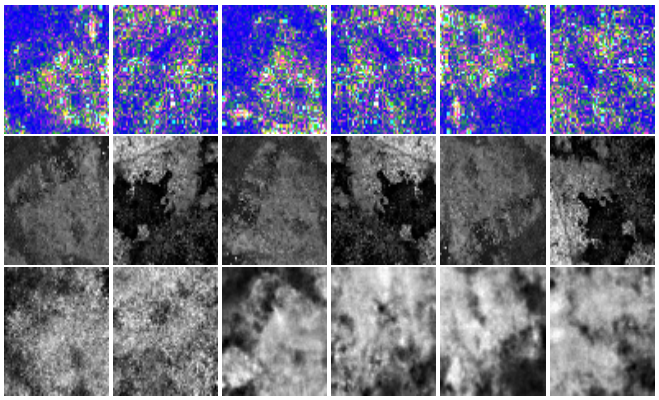
Table 1: Average RMSE on the test set in Mgha^{-1}. Results are given for all the tested GAN architectures, specifying the configuration that produces the lowest average RMSE.

Among all possible models we found that the WGAN-GP trained on a ResNet-5 network with BN on the G, LN on the D, and using a BS of 4 yielded the lowest average RMSE of all models: 56.7 Mgha^{-1}. We kept ResNet-5 fixed and compared models trained with Eq. 1 and Eq. 2 against the best WGAN-GP. We also kept WGAN-GP fixed while training the different ResNet-4, 5 and 6 networks, to evaluate the difference between them. Tab. 1 summarises the results with model specifications. All three ResNet networks trained on the specific WGAN-GP perform similarly, although ResNet-5 yields slightly lower average RMSE. Tab. 1 shows that the choice of objective function has the largest impact on the average RMSE, where the WGAN-GP is clearly better than the other two for our dataset. Fig. 2 shows a qualitative comparison of the five different models from Tab. 1, where each column correspond to the patch closest to the average RMSE shown in Tab. 1. The first row of Fig. 2 visualises patches from the (input) SEN1A domain, the middle row from the real (output) ALS domain and the third row from the generated synthetic ALS patches. Among the generated patches in Fig. 2, the Vanilla GAN differs from the other models by generating more crispy looking patches, while the other models generate more blurry looking patches.

To investigate the worst and best case scenario of each the three models in the upper part of Tab. 1, we plot the patches from the test set that yield minimum and maximum RMSE in Fig. 3. Tab. 2 shows the corresponding minimum and maximum RMSE for these patches. From Fig. 3 it can be noted that the same two patches, but with different flipping/rotation, are selected as the easiest and the hardest one for the image-to-image translation. Once again, the Vanilla GAN achieves the crispiest look, but with the largest deviation between the real and generated ALS patch, compare column (b) with col-

Model	Min [Mgha^{-1}]	Max [Mgha^{-1}]
Vanilla GAN, ResNet-5	49.9	85.6
LSGAN, ResNet-5	40.0	**73.7**
WGAN-GP, ResNet-5	**37.0**	75.4

Table 2: Computed minimum and maximum RMSE in the test set when generating ALS patches using the ResNet-5 network and the three different objective functions.



(a) Min (b) Max (c) Min (d) Max (e) Min (f) Max

Fig. 3: **First row:** SEN1A patches. **Second row:** Real ALS patches. **Third row:** Generated ALS patches. Column (a), (b) Vanilla GAN, ResNet-5 (c), (d) LSGAN, ResNet-5 and (e) , (f) WGAN-GP, ResNet-5. Columns with caption *min* indicate patch with min RMSE within the test set, while caption *max* indicate patch with max RMSE.

umn (d) and (f). From the same figure, it can be noted that all three objective functions seem to be approximately equally appropriate for translating from $\mathcal{X}$ to $\mathcal{Y}$ when patches from the two domains have similar appearance, but struggle when the $\mathcal{X}$ and $\mathcal{Y}$ domains deviate from each other in appearance.

5. CONCLUSION

So far, we have shown that the cGAN architectures are suitable for generating synthesised images of ALS-based AGB estimates from SEN1A scenes. Three different objective functions and three different deep neural networks of different dimensions were optimised and applied in the generation. Overall we found that the models manage to capture the ALS specific structure when trained to perform image-to-image translation from the SEN1A domain. Results show that the G networks perform similarly, but the ResNet-5 might be slightly better suited for the task. In general, the Vanilla GAN produces sharper synthesised images than applying the LSGAN or the WGAN-GP. Despite the visual results, the Vanilla GAN also receives the highest RMSE. As pointed out in [10], it is a delicate issue to evaluate the performance of GANs, which suggests that the evaluation needs to match the target application. Therefore, selecting one model in favour of another boils down to selecting a model that fulfils the aim of this study best. Our interest lies in synthesising ALS-based AGB estimates from corresponding SEN1a data, not to generate data that visually appear similar to the real ALS data. We therefore suggest further work to focus on applying a ResNet-5 generator network, a WGAN-GP objective function trained with BN and LN, and a BS of 4, as this model achieves the lowest average RMSE over all considered models. We aim to finalise this work by performing statistical characterisation of the obtained AGB estimates and compare to ground reference data.

6. ACKNOWLEDGEMENTS

We gratefully acknowledge employees of the Tanzania Forest Services (TFS) Agency, Sokoine University of Agriculture, Norwegian University of Life Sciences, and the Swedish University of Agricultural Sciences for participation in field work and provision of in situ measurements, remotely sensed data and derived products.

7. REFERENCES

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

[2] S. Kaasalainen, M. Holopainen, M. Karjalainen, M. Vastaranta, V. Kankare, K. Karila, and B. Osmanoglu, "Combining lidar and synthetic aperture radar data to estimate forest biomass: Status and prospects," *Forests*, vol. 6, no. 12, pp. 252–270, 2015.

[3] 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, 2013.

[4] E. Næsset, H. O. Ørka, S. Solberg, O. M. Bollandsås, E. H. Hansen, E. Mauya, E. Zahabu, R. Malimbwi, N. Chamuya, H. Olsson, and T. Gobakken, "Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision," *Remote Sens. Environ.*, vol. 175, pp. 282–300, 2016.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.

[6] 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 Recogn. (CVPR)*, 2017, pp. 1125–1134.

[7] X. Bao, Z. Pan, L. Liu, and B. Lei, "SAR image simulation by generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2019, pp. 9995–9998.

[8] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," *arXiv preprint arxiv:1611.04076*, 2017.

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

[10] L. Theis, A. van den Oord, and M. Bethge, "A note on the evaluation of generative models," *arXiv preprint arXiv:1511.01844*, 2016.