Supplementary material:

BioMassters: A Benchmark Dataset for Forest Biomass Estimation using Multi-modal Satellite Time-series

Dataset structure

The goal of this dataset is to test deep learning algorithms that predict yearly Above Ground Biomass (AGB) for Finnish forests using satellite imagery.

- **Feature data:** Satellite imagery from the European Space Agency and European Commission's joint Sentinel-1 and Sentinel-2 satellite missions, designed to collect a rich array of Earth observation data
- **Label data:** Reference label AGB measurements collected using LiDAR (Light Detection and Ranging) calibrated with in-situ measurements. LiDAR is able to generate high-quality AGB maps, but is more time consuming and intensive to collect than satellite imagery.

The following directory structure is used:

```
|-- features_metadata.csv
|-- train_features
|   |__<satellite files>
|-- test_features
|   |__ <satellite files>
|-- train_agbm_metadata.csv
|-- train_agbm
    |__ <LiDAR files>
```

The satellite feature data files are named `{chip_id}_{satellite}_{month}.tif`, where `month` represents the number of months starting from September (00 is September, 01 is October, 02 is November, and so on). The LiDAR AGBM files are named `{chip_id}_agbm.tif`.

dataset	# files	size
train_features	189078	215.9GB
test_features	63348	73.0GB
train_agbm	8689	2.1GB

Dataset & Baseline Download

Data folders can be downloaded from both HuggingFace Platform or AWS:

1. The dataset can be downloaded from:
   https://huggingface.co/datasets/nascetti-a/BioMassters

2. Data folders can be downloaded from the following AWS s3 bucket links:
   - - training set features: s3://drivendata-competition-biomassters-public-us/train_features/
   - - test set features: s3://drivendata-competition-biomassters-public-us/test_features/
   - - training set AGBM: s3://drivendata-competition-biomassters-public-us/train_agbm

Link to the data challenge website hosted by DrivenData:
https://www.drivendata.org/competitions/99/biomass-estimation/page/760/

Link to the github repository with the top-performing models:
https://github.com/drivendataorg/the-biomassters

Link to the paper repository that will be updated with more content for the camera-ready version:
https://nascetti-a.github.io/BioMasster/

Feature data description

The feature data for this dataset is imagery collected by the Sentinel-1 and Sentinel-2 satellite missions for nearly 13,000 patches of forest in Finland. Each patch (also called a "chip") represents a different 2,560 by 2,560 meter area of forest. The data were collected over a period of 5 years between 2016 and 2021.

Each label in this challenge represents a specific chip, or a distinct area of forest. LiDAR measurements are used to generate the biomass label for each pixel in the chip. For each chip, a full year's worth of monthly satellite images for that area are provided, from the previous September to the most recent August. For example, for a LiDAR-based reference label chip from 2020, monthly satellite data is provided from September 2019 to August 2020.

All of the satellite images have been geometrically and radiometrically corrected and resized to 10 meter resolution. Each resulting image is 256 by 256 pixels, and each pixel represents 10 square meters. Images represent monthly aggregations and are provided as GeoTIFFs with any associated geolocation data removed.

You only need to generate one biomass prediction per chip, but can use as many of the chip's multi-temporal (different months) or multi-modal (Sentinel-1 or Sentinel-2) satellite images as you like. Predictions should include a yearly peak AGB value for each 10 by 10 pixel in the chip.

Information about each satellite image, including its corresponding patch, satellite, and the month in which it was captured, is recorded in `features_metadata.csv`

The fields in `features_metadata.csv` are:

- `chip_id`: A unique identifier for a single patch, or area of forest
- `filename`: The filename of the corresponding image, which follows the naming convention `{chip_id}_{satellite}_{month_number}.tif`. (`month_number` corresponds to the number of months since September of the year previous to when the reference labels were captured, so `00` would represent September, `01` October, and so on, until `12`, which represents August of the same year)
- `satellite`: The satellite the image was captured by (`S1` for Sentinel-1 and `S2` for Sentinel-2)
- `split`: Whether the image is a part of the training data or test data
- `month`: The name of the month in which the image was collected
- `size`: The file size in bytes
- `cksum`: A checksum value to make sure the data was transmitted correctly. For more details on how to use the `cksum`, see the `biomassters_download_instructions.txt` file on the data download page.
- `s3path_us`: The file location of the image in the public s3 bucket in the US East (N. Virginia) region
- `s3path_eu`: The file location of the image in the public s3 bucket in the Europe (Frankfurt) region
- `s3path_as`: The file location of the image in the public s3 bucket in the Asia Pacific (Singapore)
- `corresponding_agbm`: The filename of the tif that contains the AGBM values for the chip_id

Sentinel-1

The provided Sentinel-1 data include two bands "VV" and "VH" for both ascending and descending orbits for a total of four bands. These bands are captured from the sensor transmitting vertically polarized signal (represented by the first "V") and receiving either vertically (V) or horizontally (H) polarized signal.

The values in these bands represent the energy that was reflected back to the satellite measured in decibels (dB). Pixel values can range from negative to positive values. A pixel value of -9999 indicates missing data. An advantage of Sentinel-1's use of SAR is that it can acquire data across day or night, under all weather conditions. Clouds or darkness do not impede the ability of Sentinel-1 to collect images.

Finally, Sentinel-1 has a 6-day revisit orbit, which means that it returns to the same area about five times per month. We have provided a single composite image from Sentinel-1 for each calendar month, which is generated by taking the mean across all images acquired by Sentinel-1 for the patch during that time. For more details on how to interpret SAR data, participants might find it helpful to consult NASA's guide to SAR.

Example of a Sentinel-1 image:

`001b0634_S1_00.tif` is an image from Sentinel-1 provided as a part of the training dataset. The filename follows the format `{chip_id}_{satellite}_{month_number}.tif`, so we know that the chip_id is `001b0634` and that the image was captured by Sentinel-1 in September (the month number corresponds to the number of months since September).

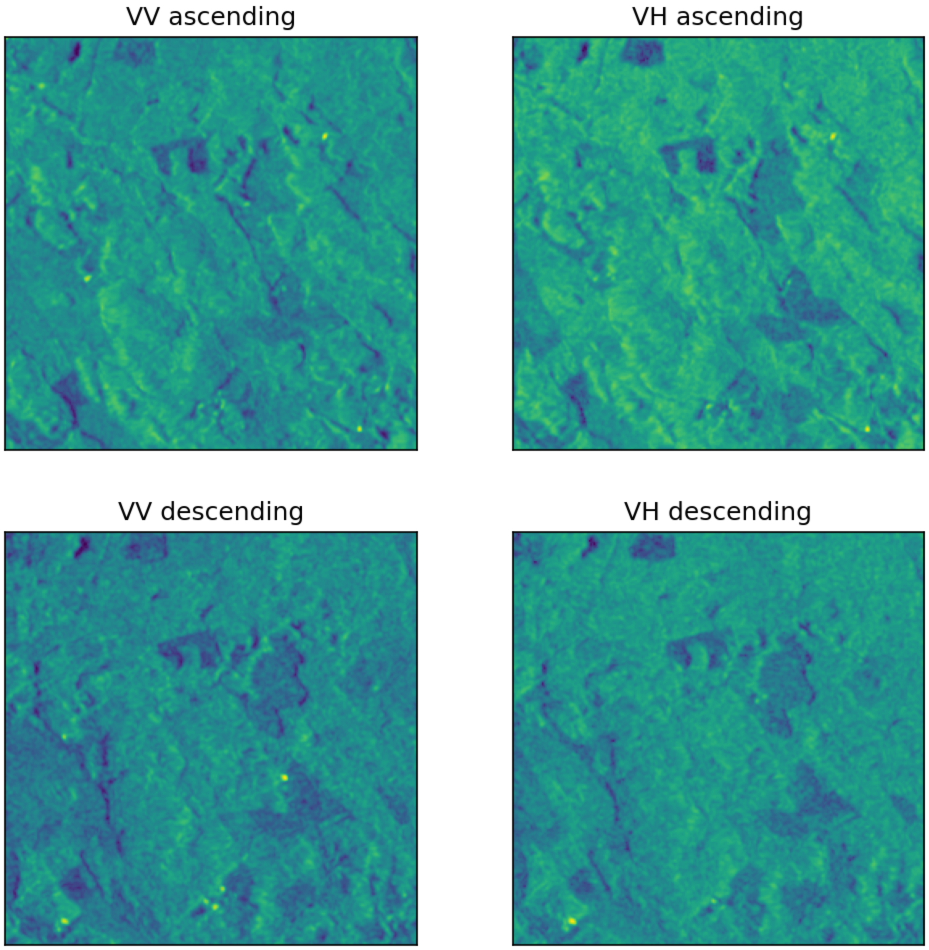


Figure1: Sentinel-1 image, by band

- ## Sentinel-2

Sentinel-2 is a high-resolution imaging mission that monitors vegetation, soil, water cover, inland waterways, and coastal areas. Sentinel-2 satellites have a Multispectral Instrument (MSI) on board that collects data in the visible, near-infrared, and short-wave infrared portions of the electromagnetic spectrum. We have selected the best image for each month from the S2 data, as opposed to taking the mean of all images collected over the month, as is done with S1.

The following 11 bands are provided for each S2 image: B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, and CLP (a cloud probability layer). See the Sentinel-2 guide for a description of each band. The CLP band — cloud probability — is provided because S2 cannot penetrate clouds. The cloud probability layer has values from 0-100, indicating the percentage probability of cloud cover for that pixel. In some images, this layer may have a value of `255`, which indicates that the layer has been obscured due to excessive noise.

Example of a Sentinel-2 image:

`001b0634_S2_00.tif` is an image from Sentinel-2 that is part of the training dataset. Its name indicates it is for chip `001b0634` and collected by Sentinel-2 during September.

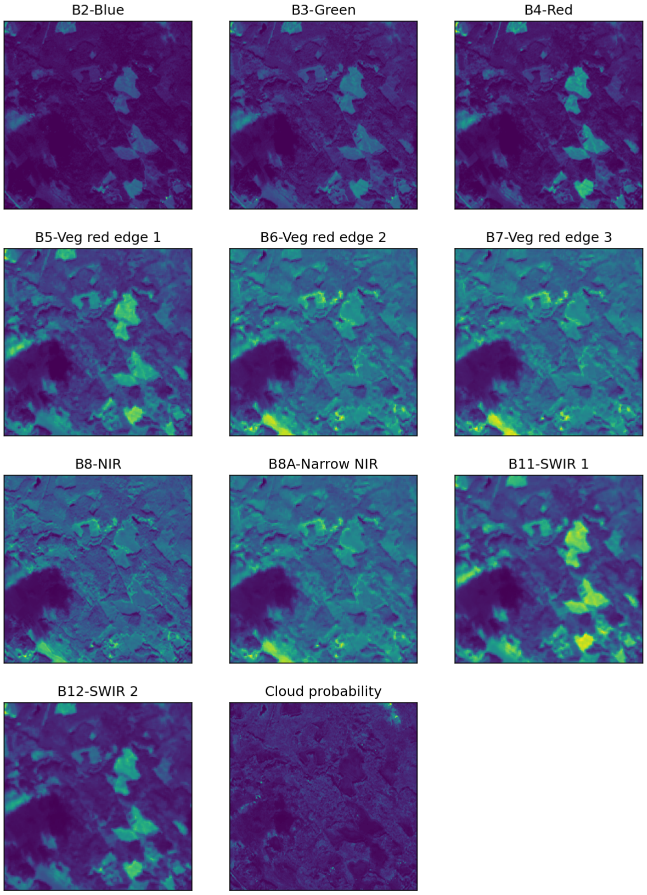


Figure 2: Sentinel-2 image, by band

Label data description

The reference labels for this competition are yearly AGB measured in tonnes. Labels for each patch are derived from LiDAR (Light Detection and Ranging), a remote sensing technology that provides 3D information about the terrain and vegetation. The label for each patch is the peak biomass value measured during the summer.

Similarly to the feature satellite imagery, LiDAR data is provided as images that cover 2,560 meter by 2,560 meter areas at 10 meter resolution, which means they are 256 by 256 pixels in size. For the same chip ID, each pixel in the satellite data corresponds to a pixel in the same position in the LiDAR data. Note that `0` values in this dataset can represent areas with zero biomass or areas where there is missing data. The file `train_agbm_metadata.csv` provides the following information about AGB images:

- `chip_id`: The patch the image corresponds to
- `filename`: The filename the image can be found under. The filename follows the convention `{chip_id}_agbm.tif`
- `size`: The file size in bytes
- `cksum`: A checksum value to make sure the data was transmitted correctly. For more details on how to use the `cksum`, see the `biomassters_download_instructions.txt` file on the data download page.
- `s3path_us`: The file location of the image in the public s3 bucket in the US East (N. Virginia) region
- `s3path_eu`: The file location of the image in the public s3 bucket in the Europe (Frankfurt) region
- `s3path_as`: The file location of the image in the public s3 bucket in the Asia Pacific (Singapore)

Example of an AGB reference image:

`001b0634_agbm.tif` is an image from LiDAR data for the training dataset. From the name, we can tell it is for chip `001b0634`:

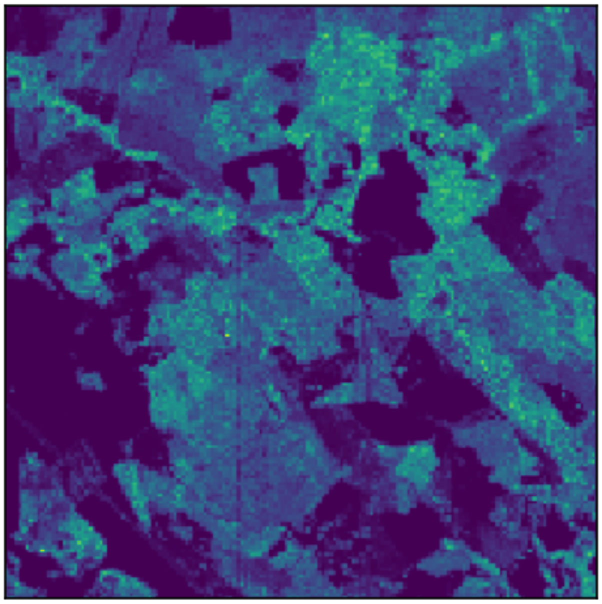
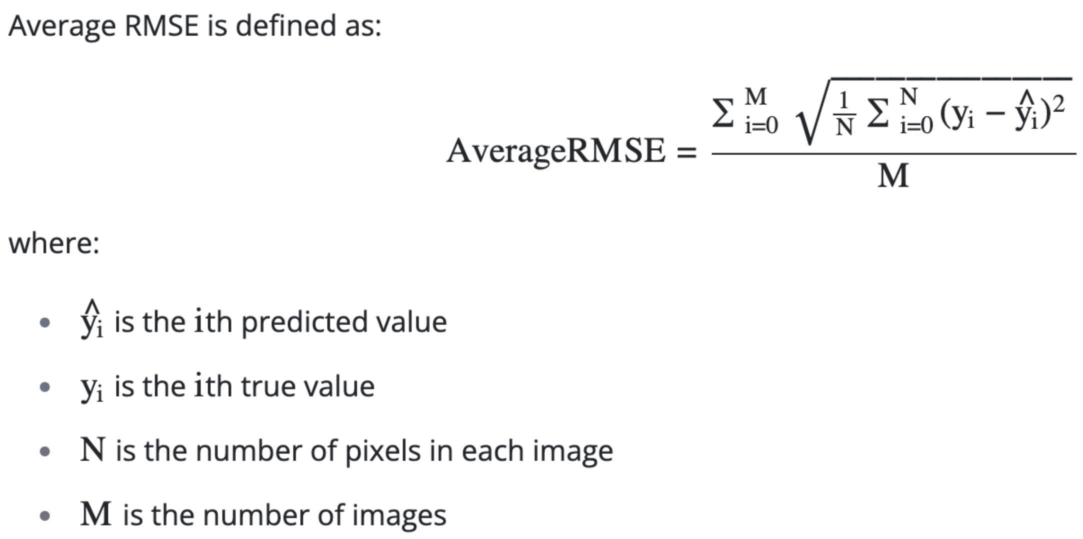


Figure 3: AGB reference image

Evaluation metric

To measure your model’s performance, we’ll use a metric called average Root Mean Square Error (RMSE). RMSE is the square root of the mean of squared differences between estimated and observed values. RMSE will be calculated on a per-pixel basis (i.e., each pixel in your submitted `tif` for a patch will be compared to the corresponding pixel in the reference label `tif` for the patch). RMSE will be calculated for each image, and then averaged over all images in the test set. This is an error metric, so a lower value is better. Note: There are some outliers in this dataset, and they are included in the scoring. Pixels with a value of zero are not included in the scoring.



Average RMSE is defined as:

$$\text{AverageRMSE} = \frac{\Sigma_{i=0}^{M} \sqrt{\frac{1}{N} \Sigma_{i=0}^{N} (y_i - \hat{y}_i)^2}}{M}$$

where:

- $\hat{y}_i$ is the ith predicted value
- y_i is the ith true value
- N is the number of pixels in each image
- M is the number of images

Datasheet for BioMasster dataset

B.1 Motivation

Q1 *For what purpose was the dataset created? Was there a specific task in mind? Was there a particular gap that needed to be filled? Please provide a description.*

The goal of this dataset is to test deep learning algorithms that estimate yearly Above Ground Biomass (AGB) for Finnish forests using satellite multi-modal time series .

Q2 *Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?*

The dataset has been created by the authors through a collaboration between the University of Liege (Belgium), the KTH Royal Institute of Technology (Sweden) and Driven Data (US).

Q3 *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

Core university funds. The prize for the associate data competition hosted by Driven Data was sponsored by Matworks.

Q4 *Any other comments?*

B.2 Composition

Q5 *What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?*

An instance is a region of Finland forests that covers 2560 x 2560 square meters. The corresponding AGB map and satellite data is provided at 10m spatial resolution resulting in patches of 256 x 256 pixels.

Q6 How many instances are there in total (of each type, if appropriate)?

We provide nearly 13000 patches covering the entire Finland.

Q7 Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

The instances are sampled from the Finnish free and open forestry inventory data (see the manuscript for the details).

Q8 What data does each instance consist of?

Each instance consists of satellite image time series from the Sentinel-1 and Sentinel-2 constellations with the corresponding reference AGB map.

Q9 Is there a label or target associated with each instance?

Yes, we provide the corresponding AGB map derived by airborne LiDAR surveys.

Q10 Is any information missing from individual instances?

Yes, for some instances it is possible that we have some gaps in the satellite time series

Q11 Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?

[No]

Q12 Are there recommended data splits (e.g., training, development/validation, testing)?

- Yes, we provide data splits for reproducing the results of the top-performing models.

Q13 Are there any errors, sources of noise, or redundancies in the dataset?

The reference AGB maps (labels) are estimated using airborne LiDAR data and they are affected by measurement errors. However, the level of these errors could be considered negligible considering that we are using medium resolution satellite multi-spectral and SAR data to estimate the AGB (see the manuscript for more details)..

Q14 Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?

This dataset is self-contained and is stored and distributed using the HuggingFace platform (https://huggingface.co/datasets/nascetti-a/BioMassters). The dataset is under the CC-BY-4.0 License.

Q15 Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals' non-public communications)?
[No]

Q16 Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.
[No]

Q17 Does the dataset relate to people?
[No]

Q18 Does the dataset identify any subpopulations (e.g., by age, gender)?
[No]

Q19 Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?
[No]

Q20 Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?
[No]

Q21 Any other comments?
[No]

B.3 Collection Process

Q22 How was the data associated with each instance acquired?

- The Sentinel-1 and Sentinel-2 time series were pre-processed and downloaded using the Google Earth Engine platform (see the manuscript for more information)

- The reference AGB labels were computed using custom python script and the Finland Open Forest Database

Q23 What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?

We use python scripts exploiting the capabilities of the Google Earth Engine platform

Q24 If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

Yes, we describe in detail the sampling strategy in the manuscript.

Q25 Who was involved in the data collection process (e.g., students, crowdworkers, contrac- tors) and how were they compensated (e.g., how much were crowdworkers paid)?

The authors

Q26 Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)?

The collection of satellite imagery spanned from 2017 to 2021, which coincides with the duration required for covering the Finland forested areas with an aerial survey

Q27 Were any ethical review processes conducted (e.g., by an institutional review board)?
[No]

Q28 Does the dataset relate to people?
[No]

Q29 Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?
[N/A]

Q30 Were the individuals in question notified about the data collection?
[N/A]

Q31 Did the individuals in question consent to the collection and use of their data?
[N/A]

Q32 If consent was obtained, were the consenting individuals provided with a mechanism
to revoke their consent in the future or for certain uses?
[N/A]

Q33 Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?
[No]

Q34 Any other comments?
[No]

B.4 Preprocessing, Cleaning, and/or Labeling

Q35 Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?

Yes, the satellite images are pre-processed, see the manuscript for the details.

Q36 Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

Sentinel 1 and Sentinel 2 satellite images are public available on the ESA sentinel hub (https://scihub.copernicus.eu/)

Q37 Is the software used to preprocess/clean/label the instances available?
Not yet, we will clean the GEE scripts that we developed and include them in the Dataset repository

Q38 Any other comments?
[No]

B.5 Uses

Q39 Has the dataset been used for any tasks already?
[No]

Q40 Is there a repository that links to any or all papers or systems that use the dataset?
[No]

Q41 What (other) tasks could the dataset be used for?

We encourage future researchers to use BioMasster dataset for other tasks. In Particular, we see applications in developing multimodal regression models capable of uncertain estimation. Due to the data size, it also offers an opportunity for pre-training of models for other geospatial analysis tasks.

Q42 Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

This dataset is geographically limited to Finland. It could be difficult to use it as is for regions with different type of climate (e.g. tropical forest)

Q43 Are there tasks for which the dataset should not be used?
[No].

Q44 Any other comments?
[No].

B.6 Distribution

Q45 Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?

[Yes] the dataset will be open-source.

Q46 How will the dataset be distributed (e.g., tarball on website, API, GitHub)?

The data will be available through .zip files available on the HuggingFace platform (https://huggingface.co/datasets/nascetti-a/BioMassters). We plan to implement also dataviewer and dataloader using the API of the platform.

Q47 When will the dataset be distributed?

- All data with the exception of the test split labels are accessible. The entire dataset, including the test split labels, will be released for the camera ready paper .

Q48 Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

[Yes] . CC-BY-4.0

Q49 Have any third parties imposed IP-based or other restrictions on the data associated with the instances?
[No]

Q50 Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?
[No]

Q51 Any other comments?

[No]

B.7 Maintenance

Q52 Who will be supporting/hosting/maintaining the dataset?
The University of Liege and Driven Data

Q53 How can the owner/curator/manager of the dataset be contacted (e.g., email address)?
andrea.nascetti@uliege.be or rituy@kth.se

Q54 Is there an erratum?
Not yet

Q55 Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?
We plan to extend it to other regions and provide all the geographical information of the image tiles.

Q56 If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)?
- N/A

Q57 Will older versions of the dataset continue to be supported/hosted/maintained?
- [Yes] .

Q58 If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

Yes, the HuggingFace platform where the dataset is hosted enable the contributions from other users upon the approval of the owner.

Q59 Any other comments?
- [No].

BioMassters: A Benchmark Dataset for Forest Biomass Estimation using Multi-modal Satellite Time-series

1 Appendix A: Top-performing models details

1.1 U-TAE Model details

We adapted U-TAE Model [1]. We consider input as an image time sequence X, organized into a four-dimensional tensor of shape $T \times C \times H \times W$, with T the length of the sequence, C the number of channels, and $H \times W$ the spatial extent.

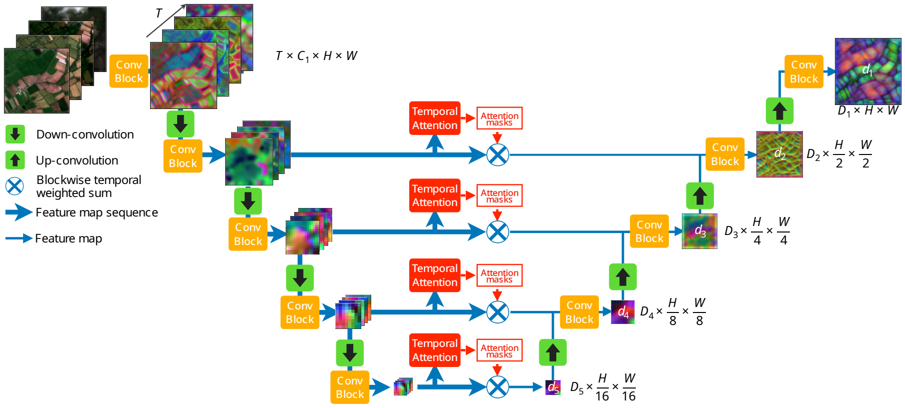

T
Conv Block
$T \times C_1 \times H \times W$
Down-convolution
Up-convolution
Blockwise temporal weighted sum
Feature map sequence
Feature map
Temporal Attention
Attention masks
d_1
$D_1 \times H \times W$
d_2
$D_2 \times \frac{H}{2} \times \frac{W}{2}$
d_3
$D_3 \times \frac{H}{4} \times \frac{W}{4}$
d_4
$D_4 \times \frac{H}{8} \times \frac{W}{8}$
d_5
$D_5 \times \frac{H}{16} \times \frac{W}{16}$

Figure 1: U-TAE Model Architecture (edited from [1])

Spatio-Temporal Encoding The model encodes a sequence X in three steps: (a) each image in the sequence is embedded simultaneously and independently by a shared multi-level spatial convolutional encoder, (b) a temporal attention encoder collapses the temporal dimension of the resulting sequence of feature maps into a single map for each level, (c) a spatial convolutional decoder produces a single feature map with the same resolution as the input images, see Figure. 1.

The adapted model has two major differences then the U-TAE Model [1]. (a) Spatial Encoding: Unlike [1] we do not use group normalization in encoder because we do not see any improvements over batch normalization. (b) Temporal Encoding : Unlike [1] we use a simplified attention-based scheme without grouping strategy which processes the temporal dimension at each feature map resolution. For each resolution map, we apply shared attention weights independently at each pixel of e^l, the feature map sequence at the level resolution l. This generates a temporal attention mask a^l for each pixel. The masks a^l at level l of the encoder are then used as weights to aggregate e^l on the temporal dimension resulting f^l map:

$$f^l = \sum_{t=1}^{T} a_t^l \odot e_t^l, \tag{1}$$

with $\odot$ term-wise multiplication with channel broadcasting.

Training details We take `tf_efficientnetv2_l_in21k`encoder from `timm` framework. The inputs to the encoder are 15-band ($C = 15$) images with a resolution of $W \times H = 256 \times 256$ from joint Sentinel-1 and Sentinel-2 satellite missions. The encoder is shared for all $T = 12$ months. We directly optimize RMSE loss for 900 epochs using AdamW optimizer with learning rate 10^{-3} and CosineAnnelingLR scheduler. We don't compute loss for high AGB values over 400. We use vertical flips, rotations, and random month dropout as augmentations. Month dropout removes images.

1.2 Swin UNETR Model details

Data Preprocessing The 0.1st and 99.9th percentiles are obtained as the lower and upper bound of inliers of each feature and AGB. Outliers are replaced by lower or upper limitations. The missing features of a specific month or modality (Sentinel-1 and Sentinel-2) are substituted by a zero array. Features are normalized using the Z-score method, and AGB is rescaled to range [0, 1]. After that, for each training sample, 4 Sentinel-1 features and 11 Sentinel-2 features of 12 months are concatenated to 4D tensor in shape [15, 12, 256, 256], AGB is in shape [1, 256, 256] as training target.

Model and Losses The adapted Swin UNEet TRansformer (Swin UNETR) is applied as the spatial-temporal regression model. The original Swin UNETR is designed for semantic segmentation of 3D medical images [2]. It has a UNet-based architecture with Swin Transformer V1 [3] as an encoder to extract multi-scale features using self-attention in an efficient shifted window partitioning scheme. We replace the attention layer of V1 block with the attention proposed in Swin Transformer V2 [4] to improve the training stability.

The adapted Swin UNETR contains a 4-stage encoder to learn multi-scale features from input, and then a 5-stage decoder upsamples feature maps to the same spatial-temporal size as input. Feature maps are averaged on the

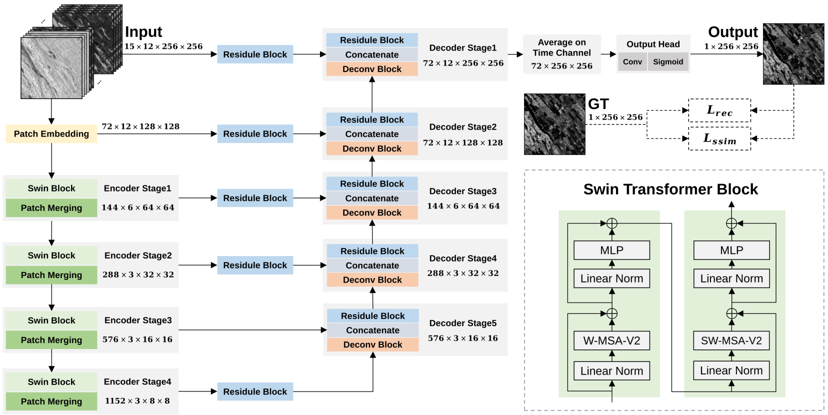


Figure 2: Swin UNETR model architecture

time channel and fed into the output head to generate AGB prediction. We apply mean absolute error as the reconstruction loss L_{rec} to measure content consistency between AGB ground truth I and prediction $\hat{I}_i$ as

$$L_{rec} = \frac{1}{n} \sum_{i=1}^{n} (I_i - \hat{I}_i) \tag{2}$$

where n is the number of pixels in AGBM.

In addition, structure similarity loss L_{ssim} is used in training to produce more visually pleasing AGBM predictions. Structure similarity comprehensively measures the differences between images in brightness, contrast, and structure, and it correlates more with human perception of image quality. L_{ssim} is calculated as

$$L_{ssim} = 1 - SSIM(I, \hat{I}) \tag{3}$$

where SSIM is implemented as [5].

The total loss L_{total} is the weighted combination of L_{rec} and L_ssim where λ_1 is the weight of L_{rec}, λ_2 is the weight of L_{ssim}.

$$L_{total} = \lambda_1 L_{rec} + \lambda_2 L_{ssim} \tag{4}$$

Training details The adapted Swin UNETR is trained for 100 epochs using AdamW optimizer with constant betas (0.9, 0.99) and weight decay 0.01. The learning rate increases linearly from 0.0 to 0.001 in the first 10 epochs, then it anneals in a cosine schedule to 0.0 in the last 90 epochs. The batch size is set to 4 to take full advantage of GPU A100-40G. In each training step, Volumentations-3D [6] carries out the 3D data augmentation on features and AGB targets, including vertical flipping, horizontal flipping and randomly rotating in 90 degrees with the probability of 0.1 for each operation. In the loss function, λ_1 is set to 1.0, and λ_2 is set to 0.2. The training samples are split into 5 folds to train 5 models. For each testing sample, the average of 5 outputs is the final AGB prediction.

1.3 UNET++ Model details

Data Preprocessing Sentinel-1 (S1) and Sentinel-2 (S2) imagery were preprocessed into six cloud-free median composites (Table 1) to reduce data dimensionality while preserving the maximum amount of information. S1 imagery was reduced to seasonal median composites and then stacked. Similarly, S2 imagery was first cloud masked using a 50% threshold of the cloud probability layer, and then also reduced to seasonal median composites and stacked. For two composites (i.e. 2SI and 4SI) multiple vegetation (e.g. NDVI for S2) and spectral indices (e.g. VV/VH ratio for S1) were also generated.

Training Details The data consisting of 8692 stacked images were divided into the train (98.9%, n=8596) and validation (1.1%, n=96) datasets. This was done in a stratified manner by binning AGB values into four 25th-percentile bins. S1, S2, and AGBD images were also standardized using mean and standard deviation calculated on the train set. Then 15 models were trained using a UNet++ architecture in combination with various encoders and attention blocks (e.g. scse) and median composites (Figure 1). The pre-trained on imagenet dataset models were further trained with multiple augmentations (e.g. flips and rotations), batch size of 32, AdamW optimizer with 0.001 initial learning rate, weight decay of 0.0001, and a ReduceLROnPlateau scheduler. UNet++ models were optimized using a Huber loss to reduce the effect of outliers in the data for 200 epochs. To improve the performance of each UNet++ model they were further fine-tuned (after freezing pre-trained encoder weights and removing augmentations) for another 100 epochs. For each UNet++ model, the average of the two best predictions was used for further ensembling and evaluation using a root-mean-square error (RMSE). The ensemble of all 15 models using a weighted average was used for the final evaluation of the test set (n=2773). The training of the ensemble model took approximately 360 hours (i.e. 24 hours/model) on a single NVIDIA GeForce RTX 3090 (24 GB VRAM).

References

[1] Vivien Sainte Fare Garnot and Loic Landrieu. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *ICCV*, 2021.

[2] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger Roth, and Daguang Xu. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images, January 2022. arXiv:2201.01266 [cs, eess].

[3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, Montreal, QC, Canada, October 2021. IEEE.

[4] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin Transformer V2: Scaling Up Capacity and Resolution.

[5] Loss Functions for Image Restoration With Neural Networks | IEEE Journals & Magazine | IEEE Xplore.

[6] Roman Solovyev, Alexandr A. Kalinin, and Tatiana Gabruseva. 3D convolutional neural networks for stalled brain capillary detection. *Computers in Biology and Medicine*, 141:105089, February 2022.