

HyBiomass: Global Hyperspectral Imagery Benchmark Dataset for Evaluating Geospatial Foundation Models in Forest Aboveground Biomass Estimation

Aaron Banze¹, Timothée Stassin², Nassim Ait Ali Braham³, Ridvan Salih Kuzu⁴, Simon Besnard⁵,
and Michael Schmitt⁶, *Senior Member, IEEE*

Abstract—Comprehensive evaluation of geospatial foundation models (Geo-FMs) requires benchmarking across diverse tasks, sensors, and geographic regions. However, most existing benchmark datasets are limited to segmentation or classification tasks, and focus on specific geographic areas. To address this gap, we introduce a globally distributed dataset for forest aboveground biomass (AGB) estimation, a pixelwise regression task. This benchmark dataset combines co-located hyperspectral imagery (HSI) from the Environmental Mapping and Analysis Program (EnMAP) satellite and predictions of AGB density estimates derived from the global ecosystem dynamics investigation (GED) lidars, covering seven continental regions. Our experimental results on this dataset demonstrate that the evaluated Geo-FMs can match or, in some cases, surpass the performance of a baseline U-Net, especially when fine-tuning the encoder. We also find that the performance difference between the U-Net and Geo-FMs depends on the dataset size for each region and highlight the importance of the token patch size in the Vision Transformer (ViT) backbone for accurate predictions in pixelwise regression tasks. By releasing this globally distributed hyperspectral benchmark dataset, we aim to facilitate the development and evaluation of Geo-FMs for HSI applications. Leveraging this

dataset additionally enables research into geographic bias and the generalization capacity of Geo-FMs.

Index Terms—Aboveground biomass (AGB), Environmental Mapping and Analysis Program (EnMAP), geospatial foundation models (Geo-FMs), global ecosystem dynamics investigation (GED), hyperspectral imagery (HSI), remote sensing.

I. INTRODUCTION

FORESTS play a vital role in the global carbon cycle, making accurate estimation of related parameters such as aboveground biomass (AGB) of major interest. Optical remote sensing, particularly hyperspectral imagery (HSI), provides detailed spectral information that enables the detection of subtle variations in forest biochemical properties, such as chlorophyll, nitrogen, water content, and lignin. These properties are important indicators of vegetation condition and can serve as proxies for AGB estimation. In recent years, the availability of HSI has increased drastically with the launch of satellites such as the Environmental Mapping and Analysis Program (EnMAP) satellite in 2022 [1]. This has enabled the development of geospatial foundation models (Geo-FMs) that are pretrained on HSI and are adapted to the unique requirements of high spectral dimensionality. Initially, Geo-FMs like SatMAE [2] were limited to RGB imagery and multispectral imagery (MSI). More recently, several approaches have been developed to address the higher spectral resolution of HSI. SpectralEarth pretrained models rely on a spectral adapter using 1-D convolutions and pooling to reduce the number of channels before the patch embedding layer to be compatible with classical vision backbones [3]. The dynamic one-for-all (DOFA) model implements a dynamic hypernetwork to adjust its weights to different wavelengths of any given sensor [4]. Lastly, Panopticon utilizes cross-attention to flexibly process varying numbers of spectral bands [5].

These models have been pretrained using different datasets. Specifically, the SpectralEarth models were exclusively pretrained on EnMAP HSI from the SpectralEarth dataset [3]. Panopticon also utilized the SpectralEarth dataset serving as a basis for HSI training, while relying on additional datasets for other sensor modalities. In contrast, DOFA was pretrained on the HySpecNet-11k dataset [6] as a source for HSI, which is substantially smaller, consisting of 11 483 patches extracted from 250 EnMAP tiles, compared to the 538 974 patches in the SpectralEarth dataset.

Received 30 April 2025; revised 18 July 2025 and 10 September 2025; accepted 10 September 2025. Date of publication 15 September 2025; date of current version 25 September 2025. This work was supported in part by the HYPER-AMPLIFAI Project funded by the Helmholtz Association of German Research Centres (HGF) under Contract ZT-I-PF-4-056 and in part by European Space Agency (ESA) as part of the Fostering Advancements in Foundation Models via Unsupervised and Self-Supervised Learning for Downstream Tasks in Earth Observation (FAST-EO) Project under Contract 4000143501/23/I-DT. (*Corresponding author: Aaron Banze.*)

Aaron Banze is with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Wessling, Germany, and also with the Department of Aerospace Engineering, University of the Bundeswehr Munich, 85579 Neubiberg, Germany (e-mail: aaron.banze@dlr.de).

Timothée Stassin and Simon Besnard are with the Global Land Monitoring Group, Helmholtz German Research Centre for Geosciences (GFZ), 14473 Potsdam, Germany (e-mail: stassin@gfz-potsdam.de; besnard@gfz-potsdam.de).

Nassim Ait Ali Braham is with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Wessling, Germany, and also with the Chair of Data Science in Earth Observation (SIPEO), Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: nassim.aitalibraham@dlr.de).

Ridvan Salih Kuzu is with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: ridvan.kuzu@dlr.de).

Michael Schmitt is with the Department of Aerospace Engineering, University of the Bundeswehr Munich, 85579 Neubiberg, Germany (e-mail: michael.schmitt@unibw.de).

Data is available on-line at <https://github.com/a-banze/HyBiomass>
Digital Object Identifier 10.1109/LGRS.2025.3610178

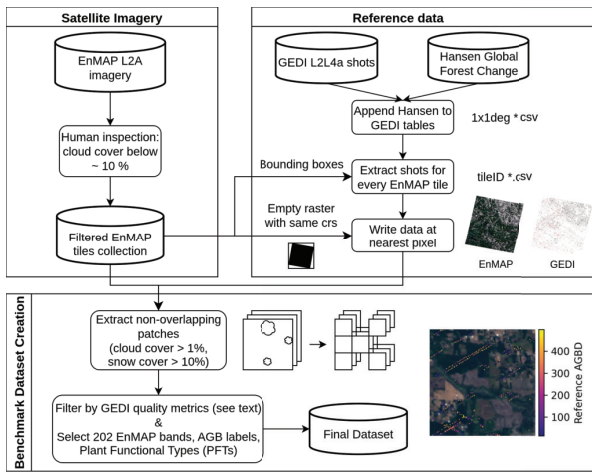


Fig. 1. Overview of the data and workflow for creating the benchmark dataset.

Evaluating the performance of Geo-FMs requires comprehensive benchmarking on a diverse range of tasks and sensors, encompassing all regions globally. However, existing benchmark datasets are predominantly limited to segmentation or classification tasks for specific geographic areas [7]. This letter introduces a globally distributed benchmark dataset for the pixelwise regression task of forest AGB estimation. It consists of co-located HSI from the EnMAP satellite and predictions of AGB density (AGBD) as reference labels derived from the global ecosystem dynamics investigation (GEDI) lidar mission. We present the results of benchmarking the aforementioned Geo-FMs adapted to HSI and a baseline U-Net model, which has been shown to achieve great results for forest AGB estimation [8], using this dataset across seven continental world regions. The regional classification adopted here is consistent with the scheme provided by the GEDI dataset, dividing the Earth into seven continental regions [9].

II. FOREST AGB BENCHMARK DATASET

An overview of the data sources and processing steps used to create the benchmark dataset is provided in Fig. 1. In Sections II-A–II-C, we first introduce the EnMAP satellite and outline the criteria we applied to select the tiles included in this dataset. Next, we introduce the GEDI mission specifications and methodology employed to derive the AGBD estimates, which we use as labels. Finally, we describe the procedure for creating the co-located dataset, which consists of spatially nonoverlapping patches.

A. EnMAP Hyperspectral Spaceborne Imagery

The EnMAP satellite, launched in April 2022, is equipped with a prism-based dual-spectrometer instrument covering the electromagnetic spectrum from visible (420 nm) to short-wave infrared (2450 nm) with 224 spectral bands. It has a 30-m ground sampling distance, and its scenes cover a 30-km swath width with a length of up to 1000 km/orbit and 5000 km/day. The satellite is expected to collect data for a period of five years. In contrast to other missions such as Sentinel-2, EnMAP is a sampling missions, meaning that scenes are acquired according to the user’s request or background scientific mission [1].

EnMAP tiles included in this dataset are sourced from Braham et al. [3], originally collected for the SpectralEarth dataset. Thereby, tiles from the EnMAP archive from April 2022 to April 2024 were selected based on human visual inspection, keeping the cloud coverage below 10%. We removed 22 spectral bands affected by water vapor absorption, resulting in low radiance values, reducing the total number to 202 spectral bands. The same bands have been excluded for the pretraining of the Geo-FMs evaluated in this letter.

B. GEDI-Derived AGBD Estimates and Quality Filtering

The GEDI mission is designed to provide high-resolution measurements of forest vertical structure using a geodetic-class waveform lidar instrument mounted on the International Space Station (ISS). GEDI employs three lasers to generate eight parallel tracks, each producing footprints of approximately 25 m in diameter, spaced 60 m apart along-track. Since its launch in 2018, GEDI has been collecting data across the globe within a latitudinal range of 51.6°N–51.6°S. The returned lidar waveforms are used to derive surface elevation, canopy height, and relative height energy metrics, which inform key forest structural attributes. AGBD, expressed in $\text{Mg} \times \text{ha}^{-1}$, is estimated at the footprint level using parametric models trained on field-based biomass observations. These models are stratified by plant functional type (PFT) and continental region to account for biome-specific structure–biomass relationships [9].

The footprint-level GEDI data included in this dataset are sourced from the spatially indexed $1 \times 1^\circ$ tables (EPSG 4326 geographic coordinates) compiled by Burns et al. [10]. These data tables were generated by: 1) downloading all GEDI L2A, L2B, and L4A orbit granule files from April 17 2019 to March 16 2023; 2) matching those files based on their unique file names; 3) extracting metrics associated with the default ground-finding algorithm (“a0”); 4) filtering for highest quality GEDI vegetation shots based on ground elevation and vegetation structure metrics; and 5) storing in $1^\circ \times 1^\circ$ tables based on the original shot coordinates.

We applied the L4A high-quality flag developed by Burns et al. [10] to filter for shots suitable for AGBD estimates. This involves setting beam sensitivity thresholds depending on the region and PFT, as well as defining a valid range of detected ground elevation among other criteria. For a complete overview of the quality filtering protocol, we refer to the original publication [10]. We further removed GEDI shots with AGBD estimates outside the range 0–500, as suggested by Carreiras et al. [11]. Additionally, only GEDI shots captured by power beams were retained, while those originating from coverage beams were excluded, following the recommendation by Lahssini et al. [12]. To filter for shots where the PFT corresponds to a forest class, we keep GEDI shots with PFT values in the range of 1–4, following the MODIS MCD12Q1 V006 product Type 5 classification scheme. We filter the GEDI shots based on PFT, as it is utilized in the preceding step, to select the model for AGBD prediction corresponding to each GEDI shot. However, due to the coarse 500-m spatial resolution of the MODIS product, we additionally filtered based on the Hansen Global Forest

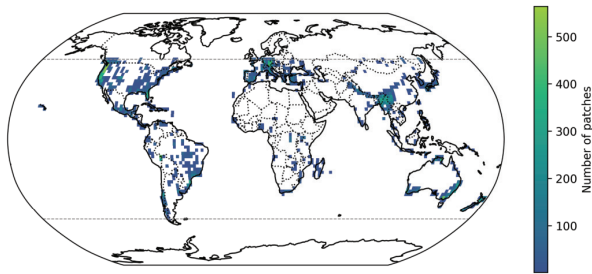


Fig. 2. Geographic distribution of patches in the final dataset. Dotted lines at 51.6°N and 0°S indicate the spatial bounds of the GEDI mission.

Change v1.11 dataset [13], which has a 30-m spatial resolution. Specifically, we utilized the `Landsat_treecover` variable, which is included in the GEDI data and defines the canopy closure of vegetation taller than 5 m in height in the year 2010, expressed as a percentage per grid cell. Only GEDI shots with `Landsat_treecover` $\geq 10\%$ were retained. Finally, we excluded GEDI shots that coincided with gross forest cover loss events occurring between 2010 and 2023 based on the `lossyear` variable from the Hansen Global Forest Change dataset.

C. Co-Locating EnMAP and GEDI Data

The EnMAP and GEDI data were spatially co-located by: 1) extracting all GEDI shots from the $1^\circ \times 1^\circ$ tables that intersect the footprint for all EnMAP tiles in our collection and 2) appending the tabular data of each GEDI shot as additional bands in the raster data at the nearest pixel value based on the geographic coordinates, with no-data values assigned to pixels lacking GEDI data.

The final dataset was created by extracting nonoverlapping patches of size 128×128 pixels from the EnMAP- and GEDI-tiles. The patchification process utilized local R-Trees to check for overlap with other patches in the same UTM zone, to avoid inaccuracies and distortions that would result from reprojecting to a shared coordinate system. To account for tiles that fall within two UTM zones, a global R-Tree was employed, allowing for a maximum overlap area of 10% between patches to account for inaccuracies. During patchification, the tiles were processed in ascending temporal order of their EnMAP acquisition date to minimize the average temporal deviation from the GEDI shots to the EnMAP tile. Patches with cloud cover exceeding 1% and snow cover above 10% based on the EnMAP quality masks were not included in the final dataset. These threshold values were determined through visual inspection of excluded patches for different threshold values. Considering inherent inaccuracies of the EnMAP quality masks, we found that these thresholds optimize the tradeoff between avoiding the exclusion of false positives and ensuring predominantly cloud- and snow-free patches. Furthermore, patches with less than 1% of pixels containing GEDI shots were also omitted.

The geographic distribution of patches in the dataset is illustrated in Fig. 2, which is primarily determined by the distribution of forest areas and the availability of nearly cloud-free EnMAP tiles. The fact that EnMAP is a German mission based on user’s requests explains, for example, the over-representation of patches over Germany compared to France.

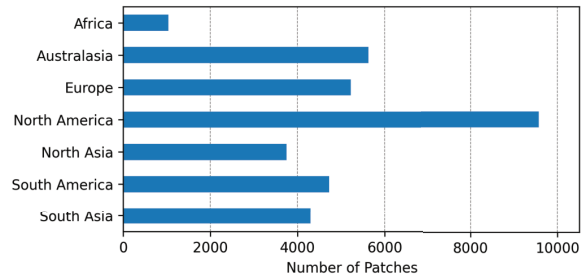


Fig. 3. Number of patches per continental world region.

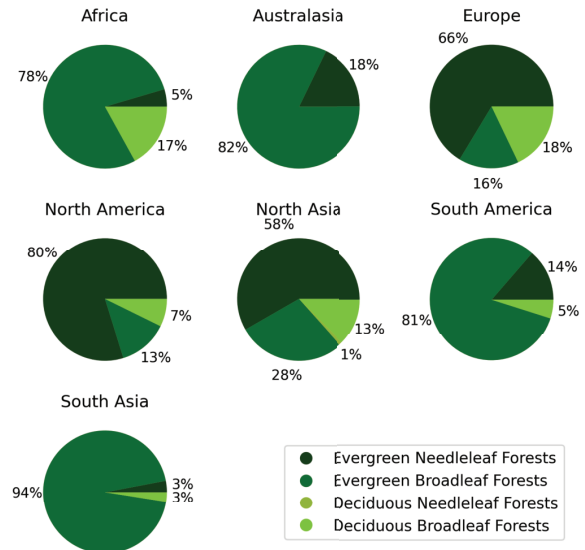


Fig. 4. Distribution of PFTs across the seven continental regions. The percentages are based on the fraction of GEDI shots corresponding to each PFT for the given region.

This results in a relatively even distribution of the number of patches per region with the exception of North America (9570 patches) and Africa (1041 patches), as visualized in Fig. 3. The distribution of PFTs per region is shown in Fig. 4, highlighting the uniqueness of each region and revealing similarities between regions, such as Europe and North America.

III. EXPERIMENTAL SETUP FOR BENCHMARKING GEO-FMS

We evaluate the performance of all Geo-FMs under two settings: 1) using frozen encoders with pretrained weights to assess the quality of extracted features and 2) fully fine-tuning including the encoder. For all Vision Transformer (ViT)-based GeoFMs, we implement a UPerNet [14] decoder, following the setup of existing benchmarks like PANGAEA [15]. All Geo-FMs included in this benchmark are based on the ViT-B architecture with 12 layers, where we select the 4th, 6th, 8th, and 12th intermediate feature layers to serve as input to the UPerNet. The output is, then, upsampled to the original patch size using bilinear interpolation. For the Spectral-ResNet-50, we append two final convolutional layers as a simple decoder, following the original publication’s setting [3]. While different pretraining strategies have been compared in SpectralEarth [3], we limit our experiments to two variants: Spectral-ResNet-50, pretrained with MoCo-V2 [16], and Spectral-ViT-B, pretrained

TABLE I

BENCHMARK RESULTS ACROSS REGIONS FOR THE BASELINE U-NET AND PRETRAINED FOUNDATION MODELS FOR FROZEN ENCODER AND FINE-TUNED ENCODER. THE BEST SCORE FOR EACH PROTOCOL AND REGION IS HIGHLIGHTED IN BOLD, AND THE SECOND BEST IS UNDERLINED

Region	U-Net		Spectral-ResNet-50		Spectral-ViT-B		DOFA ViT-B		Panopticon	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
	(Supervised Baseline)				Frozen Encoder					
Africa	65.65 ± 2.55	0.56 ± 0.03	<u>61.87 ± 2.43</u>	<u>0.61 ± 0.03</u>	60.83 ± 2.28	0.62 ± 0.03	66.02 ± 1.69	0.55 ± 0.04	63.04 ± 1.8	0.6 ± 0.03
Australasia	87.95 ± 1.39	0.47 ± 0.02	<u>85.76 ± 1.71</u>	<u>0.49 ± 0.01</u>	83.9 ± 1.66	0.51 ± 0.01	90.66 ± 1.43	0.43 ± 0.01	87.01 ± 1.7	0.48 ± 0.01
Europe	<u>74.29 ± 1.15</u>	<u>0.38 ± 0.02</u>	74.37 ± 1.13	0.38 ± 0.02	72.71 ± 1.17	0.41 ± 0.01	78.65 ± 1.29	0.31 ± 0.02	76.59 ± 1.33	0.34 ± 0.01
North America	<u>90.62 ± 0.88</u>	<u>0.41 ± 0.01</u>	90.45 ± 0.96	0.41 ± 0.01	91.19 ± 1.25	0.4 ± 0.02	96.23 ± 1.1	0.33 ± 0.01	91.07 ± 0.87	0.4 ± 0.01
North Asia	94.67 ± 0.85	0.3 ± 0.01	<u>94.01 ± 0.68</u>	<u>0.31 ± 0.01</u>	91.78 ± 1.42	0.34 ± 0.02	97.35 ± 0.7	0.26 ± 0.01	94.36 ± 0.93	0.3 ± 0.01
South America	82.79 ± 1.95	<u>0.49 ± 0.02</u>	81.79 ± 1.93	0.5 ± 0.02	83.26 ± 2.38	0.48 ± 0.03	86.12 ± 1.71	0.45 ± 0.02	83.15 ± 1.94	0.48 ± 0.02
South Asia	97.34 ± 0.84	0.27 ± 0.01	<u>97.95 ± 0.52</u>	<u>0.26 ± 0.01</u>	98.07 ± 1.53	0.26 ± 0.03	100.5 ± 0.48	0.22 ± 0.01	98.83 ± 0.86	0.25 ± 0.01
Average	84.76	0.41	<u>83.74</u>	<u>0.42</u>	83.11	0.43	87.93	0.36	84.86	0.41
	(Supervised Baseline)				Fine-tuned Encoder					
Africa	65.65 ± 2.55	0.56 ± 0.03	59.04 ± 2.35	0.65 ± 0.03	<u>62.34 ± 1.67</u>	<u>0.6 ± 0.03</u>	67.47 ± 1.51	0.54 ± 0.04	65.82 ± 2.1	0.56 ± 0.03
Australasia	87.95 ± 1.39	0.47 ± 0.02	80.08 ± 1.6	0.56 ± 0.01	<u>80.57 ± 2.01</u>	<u>0.55 ± 0.01</u>	88.92 ± 2.27	0.45 ± 0.02	85.2 ± 1.82	0.5 ± 0.01
Europe	74.29 ± 1.15	0.38 ± 0.02	<u>69.6 ± 1.17</u>	<u>0.46 ± 0.01</u>	69.08 ± 1.25	0.46 ± 0.01	77.25 ± 1.28	0.33 ± 0.02	74.98 ± 1.43	0.37 ± 0.01
North America	90.62 ± 0.88	0.41 ± 0.01	<u>82.45 ± 0.86</u>	<u>0.51 ± 0.01</u>	81.94 ± 0.76	0.52 ± 0.01	92.53 ± 0.8	0.38 ± 0.01	88.75 ± 1.0	0.43 ± 0.01
North Asia	94.67 ± 0.85	0.3 ± 0.01	89.23 ± 0.91	0.38 ± 0.01	<u>89.28 ± 1.06</u>	<u>0.38 ± 0.01</u>	99.43 ± 0.62	0.23 ± 0.01	95.5 ± 1.17	0.29 ± 0.01
South America	82.79 ± 1.95	0.49 ± 0.02	77.6 ± 1.99	0.55 ± 0.02	<u>78.77 ± 1.93</u>	<u>0.54 ± 0.02</u>	84.34 ± 1.79	0.47 ± 0.02	82.51 ± 1.91	0.49 ± 0.02
South Asia	97.34 ± 0.84	0.27 ± 0.01	94.03 ± 0.97	0.32 ± 0.01	<u>94.06 ± 0.66</u>	<u>0.32 ± 0.01</u>	99.71 ± 0.79	0.23 ± 0.02	98.58 ± 0.59	0.25 ± 0.01
Average	84.76	0.41	78.86	0.49	<u>79.43</u>	<u>0.48</u>	87.09	0.38	84.48	0.41

using the masked autoencoder (MAE) approach [17]. Whereas DOFA has also been trained using the MAE methodology, Panopticon is built on the DINOv2 framework [18]. To provide a baseline for comparison, we additionally train a fully supervised U-Net [19], which has demonstrated impressive results in many tasks, including AGB estimation [8]. We limit hyperparameter tuning to the learning rate of the AdamW optimizer and train all models for 100 epochs using mean squared error loss with early stopping. We train all models using five seeds, resulting in distinct dataset splits, with a 70:20:10 split ratio for training, validation, and testing in the regional dataset. Since DOFA and Panopticon were pretrained on 224×224 pixel patches, we resample the input data and labels using bilinear interpolation to match the pretraining dimensions of these models.

IV. BENCHMARKING RESULTS

The benchmarking results are presented in Table I, where the mean and standard deviation are calculated based on five random dataset splits for each region. We provide RMSE in Mg/ha and R^2 . In the frozen encoder setting, the average R^2 values across all regions range from 0.36 for DOFA ViT-B to 0.43 for Spectral-ViT-B. The U-Net's performance falls in between, varying by dataset size per region. Specifically, for North America, the U-Net's performance is comparable to that of the top-performing Geo-FM, whereas for Africa, which has the smallest dataset, the gap is significantly larger. The performance achievable in each region varies significantly, likely due to the differing complexities of estimating AGBD in each region and the accuracy of AGBD estimates used as labels. In the fine-tuned encoder setting, Spectral-ResNet50 and Spectral-ViT-B exhibit a substantial performance improvement, outperforming the U-Net across all regions. Interestingly, both DOFA and Panopticon exhibit superior performance in the fine-tuned encoder setting than in the frozen encoder setting across all regions, except Africa and North Asia. This deviation may be attributed to

increased data requirements associated with fine-tuning these models. DOFA's average performance improvement can also be explained by the limited amount of HSI data in the pretraining dataset, making the benefits of fine-tuning the encoder more pronounced.

For a qualitative evaluation of the results, we present the predictions of each model for a sample patch from North America and South America in Fig. 5. The EnMAP patch is visualized as an infrared false color composite and as RGB, with the reference AGBD values from the GEDI shots overlaid to highlight the sparse nature of the labels. The predictions are generated by Geo-FMs with a fully fine-tuned encoder for the corresponding region. Visually, the U-Net, Spectral-ResNet-50, and Spectral-ViT-B produce the highest quality results. In contrast, predictions by DOFA and Panopticon appear blurry, which can be explained by their token patch sizes of 16 and 14, respectively. To address this, previous studies have implemented pixel shuffle layers instead of bilinear interpolation of the UPerNet output for a learned pixelwise reconstruction [20]. To improve the pixel-level reconstruction quality in the evaluation of Prithvi 2.0 on pixelwise AGB regression, the architecture was modified to Prithvi-UNet, which integrates the pretrained Prithvi encoder and decoder as the bottleneck in a U-Net architecture, tuning all model parameters [21]. It is important to note that the varying token patch sizes in the ViT-based Geo-FMs hinder direct comparisons on their approaches to handling HSI, as this has a significant impact on the performance. The scatterplots in Fig. 5 visualize the relationship between the labels (x -axis) and AGB predictions (y -axis) by each model for all the GEDI shots in the patch. Across all models, an overestimation of low reference AGB values and an underestimation of high reference AGB values are observed. This bias has been consistently demonstrated in AGB maps produced by remote sensing models and can be attributed to signal saturation, leading to an underestimation of high AGB values, among other sources of error [22].

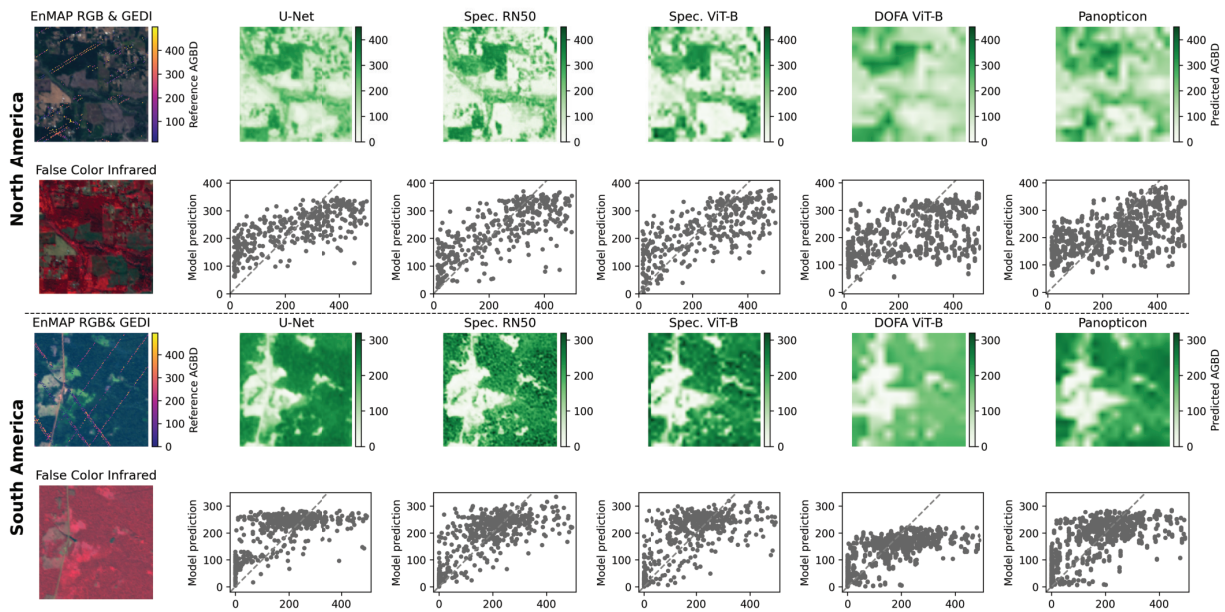


Fig. 5. (Top two rows) Model predictions for one patch from North America and (bottom two rows) another patch from South America. Each EnMAP patch is visualized in RGB with overlaid GEDI reference AGBD and as an infrared false color composite. The scatterplots show the relationship between reference AGBD values and predictions for each model. Predictions are generated using Geo-FMs with a fine-tuned encoder.

V. CONCLUSION

This letter presents a novel globally distributed hyperspectral benchmark dataset for forest AGB estimation. Our benchmarking results demonstrate that existing Geo-FMs achieve comparable performance to the baseline U-Net when the encoder is frozen, and significant improvements are observed when fully fine-tuning the encoder, particularly for Spectral-ResNet-50 and Spectral-ViT, highlighting their potential for hyperspectral biomass mapping. We also demonstrate that the token patch size is a key design factor in ViT backbones for pixelwise regression tasks. By releasing this dataset, we aim to facilitate the development and evaluation of Geo-FMs for HSI, and enable systematic study of the geographic bias, generalization capacity, and the design of pretraining strategies to improve transferability in global remote sensing applications.

REFERENCES

- [1] L. Guanter et al., “The EnMAP spaceborne imaging spectroscopy mission for Earth observation,” *Remote Sens.*, vol. 7, no. 7, pp. 8830–8857, Jul. 2015.
- [2] Y. Cong et al., “SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 197–211.
- [3] N. A. A. Braham, C. M. Albrecht, J. Mairal, J. Chanusot, Y. Wang, and X. X. Zhu, “SpectralEarth: Training hyperspectral foundation models at scale,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 16780–16797, 2025.
- [4] Z. Xiong et al., “Neural plasticity-inspired multimodal foundation model for Earth observation,” 2024, *arXiv:2403.15356*.
- [5] L. Waldmann et al., “Panopticon: Advancing any-sensor foundation models for Earth observation,” in *Proc. Comput. Vis. Pattern Recognit. Conf.*, Jun. 2025, pp. 2204–2214.
- [6] M. H. P. Fuchs and B. Demir, “HySpecNet-11k: A large-scale hyperspectral dataset for benchmarking learning-based hyperspectral image compression methods,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 1779–1782.
- [7] M. Schmitt et al., “There are no data like more data: Datasets for deep learning in Earth observation,” *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 3, pp. 63–97, Sep. 2023.
- [8] A. E. Pascarella, G. Giacco, M. Rigioli, S. Marrone, and C. Sansone, “ReUse: REgressive unet for carbon storage and above-ground biomass estimation,” *J. Imag.*, vol. 9, no. 3, p. 61, Mar. 2023.
- [9] R. Dubayah et al., “The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth’s forests and topography,” *Sci. Remote Sens.*, vol. 1, Jun. 2020, Art. no. 100002.
- [10] P. Burns, C. R. Hakkenberg, and S. J. Goetz, “Multi-resolution gridded maps of vegetation structure from GEDI,” *Sci. Data*, vol. 11, no. 1, p. 881, Aug. 2024.
- [11] J. M. B. Carreiras et al., “Coverage of high biomass forests by the ESA BIOMASS mission under defense restrictions,” *Remote Sens. Environ.*, vol. 196, pp. 154–162, Jul. 2017.
- [12] K. Lahssini, N. Baghdadi, G. L. Maire, and I. Fayad, “Influence of GEDI acquisition and processing parameters on canopy height estimates over tropical forests,” *Remote Sens.*, vol. 14, no. 24, p. 6264, Dec. 2022.
- [13] M. C. Hansen et al., “High-resolution global maps of 21st-century forest cover change,” *Science*, vol. 342, no. 6160, pp. 850–853, Nov. 2013.
- [14] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, “Unified perceptual parsing for scene understanding,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 418–434.
- [15] V. Marsocci et al., “PANGAEA: A global and inclusive benchmark for geospatial foundation models,” 2024, *arXiv:2412.04204*.
- [16] X. Chen, H. Fan, R. Girshick, and K. He, “Improved baselines with momentum contrastive learning,” 2020, *arXiv:2003.04297*.
- [17] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, “Masked autoencoders are scalable vision learners,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16000–16009.
- [18] M. Oquab et al., “DINOv2: Learning robust visual features without supervision,” *Trans. Mach. Learn. Res.*, pp. 1–32, Jan. 2024. [Online]. Available: <https://openreview.net/forum?id=a68SUt6zFt>
- [19] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [20] M. Muszynski et al., “Fine-tuning of geospatial foundation models for aboveground biomass estimation,” 2024, *arXiv:2406.19888*.
- [21] D. Szwarcman et al., “Prithvi-EO-2.0: A versatile multi-temporal foundation model for Earth observation applications,” 2024, *arXiv:2412.02732*.
- [22] M. Réjou-Méchain et al., “Upscaling forest biomass from field to satellite measurements: Sources of errors and ways to reduce them,” *Surv. Geophys.*, vol. 40, no. 4, pp. 881–911, Jul. 2019.