Estimating Canopy Height at Scale

Jan Pauls [1]   **Max Zimmer** [2]   **Una M. Kelly** [1]   **Martin Schwartz** [3]   **Sassan Saatchi** [4]   **Philippe Ciais** [3]
Sebastian Pokutta [2]   **Martin Brandt** [5]   **Fabian Gieseke** [1 6]

Abstract

We propose a framework for global-scale canopy height estimation based on satellite data. Our model leverages advanced data preprocessing techniques, resorts to a novel loss function designed to counter geolocation inaccuracies inherent in the ground-truth height measurements, and employs data from the Shuttle Radar Topography Mission to effectively filter out erroneous labels in mountainous regions, enhancing the reliability of our predictions in those areas. A comparison between predictions and ground-truth labels yields an MAE / RMSE of 2.43 / 4.73 (meters) overall and 4.45 / 6.72 (meters) for trees taller than five meters, which depicts a substantial improvement compared to existing global-scale maps. The resulting height map as well as the underlying framework will facilitate and enhance ecological analyses at a global scale, including, but not limited to, large-scale forest and biomass monitoring.

1. Introduction

Managing and conserving forest ecosystems worldwide is an indispensable component of climate adaptation and climate change mitigation strategies. Precise and up-to-date information about the health and the carbon balance of forests are, hence, critical to assess the current state of forests, to trigger appropriate countermeasures against forest loss, and to develop improved management strategies. In light of growing carbon emissions and the necessity to comply with the Paris Agreement on climate, understanding all factors

[1]Department of Information Systems, University of Münster, Germany [2]Department for AI in Society, Science, and Technology, Zuse Institute Berlin, Germany [3]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, France [4]Jet Propulsion Laboratory (JPL), California Institute of Technology, USA [5]Department of Geosciences and Natural Resource Management, University of Copenhagen, Denmark [6]Department of Computer Science, University of Copenhagen, Denmark. Correspondence to: Jan Pauls <jan.pauls@uni-muenster.de>.
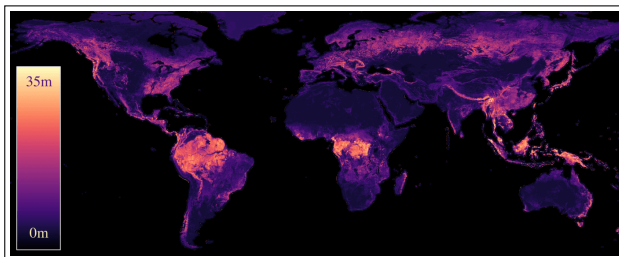
Figure 1. A global canopy height map at a 10 m resolution.

affecting our climate is crucial. This includes an accurate quantification of carbon sinks to monitor carbon distribution, gains, and losses. Despite decades of interest, this quantification is still inadequate for detailed policymaking.

Traditional methods such as National Forest Inventory (NFI) measurements, which are based on tracking and measuring individual trees, have been fundamental in estimating forest growth and loss, but are limited by their costly nature and lack of global reach. This problem is worsened by considerably varying forest monitoring efforts/techniques between nations with different financial resources (Sloan & Sayer, 2015). Advances in both Earth observation and machine learning have paved the way for the automation of forest monitoring using satellite data, including optical, radar, and LiDAR measurements, and nowadays enable more comprehensive global forest assessments (Hu et al., 2020).

A common way to assess the state of forests is to measure or to estimate their height, resulting in so-called canopy height maps. Such height estimates are then used to approximate the biomass and, thus, carbon, stored in the trees. For this reason, accurate high-resolution canopy height maps provide insights into forest cycles and dynamics and are vital for forest management and climate change mitigation. Recent research has addressed canopy height prediction with classical machine learning approaches (Potapov et al., 2021; Kacic et al., 2023) and deep neural networks (Schwartz et al., 2024; Fayad et al., 2023; Lang et al., 2023). While both global and regional canopy height maps are available, there is a notable disparity in their quality.

We address this by providing a framework for generating global-scale height maps based on satellite imagery. The overall pipeline is based on a (fully) convolutional neural

network that is trained using a novel loss function to counter the label noise present in the ground-truth measurements. The framework is also based on several simple yet crucial preprocessing steps. We showcase the effectiveness of our approach by providing a detailed global-scale canopy height map with a resolution of 10 m, see Figure 1.

2. Background

Current canopy height estimation approaches often resort to satellite data of the Sentinel-1, Sentinel-2 (European Space Agency, 2024a;b), or the mission (Williams et al., 2006) as input. As ground truth data, height measurements provided by the Global Ecosystem Dynamics Investigation (GEDI) mission (Dubayah et al., 2020) are commonly used.

2.1. Satellite Imagery

Satellite imagery for canopy height prediction primarily comes from three missions: Landsat, Sentinel-1, and Sentinel-2. The Landsat (Williams et al., 2006) program is based on a series of Earth observation satellites—operated by the NASA since 1972—and provides optical multi-spectral image data with the latest version yielding images with a pixel-resolution of 30 m. The current revisit time is 16 days, i.e., for a region on Earth, new data are collected every 16 days. The satellites belonging to the Sentinel-1 (European Space Agency, 2024a) and the Sentinel-2 (European Space Agency, 2024b) missions—operated by the European Space Agency (ESA) since 2014—offer different capabilities. The Sentinel-1 satellites feature a synthetic-aperture radar sensor, while the satellites of the Sentinel-2 mission are equipped with a multi-spectral sensor. Both types of satellites yield imagery at a resolution of 10 m and capture images every 6 days at a global scale.

The aforementioned satellite missions provide new image data on a regular basis and at a global scale. Single satellite images, however, are typically not directly used for canopy height prediction due to various challenges. Radar satellites, such as the ones of the Sentinel-1 mission, can be significantly affected by heavy rain and noise in their backscatter. On the other hand, multi-spectral sensors, such as the ones of the Sentinel-2 program, often struggle with cloud and cirrus penetration. To overcome these limitations, temporal composites are usually generated, which are based on aggregating images over a specific time frame and on calculating the per-pixel median (see, e.g., Figure 4; top left). Such composites usually help to mitigate issues related to weather and atmospheric conditions, thus ensuring more reliable and consistent data for canopy height analyses.[1]

Figure 2. An RGB image (based on multi-spectral image data collected by one of the Sentinel 2 satellites) along with GEDI height measurements (red/yellow dots) are shown for an area in France.

2.2. Sparse Height Measurements

The GEDI mission (Dubayah et al., 2020), operated by the NASA, is based on a light detection and ranging (Li-DAR) module, which is installed on the International Space Station (ISS) and which is designed to measure global vegetation height. The system features three lasers: two "power beams" scanning two tracks each and one "coverage beam" scanning four tracks. These lasers produce measurements with a 25 m footprint diameter, spaced 60 m apart. The geolocation of each measurement is estimated by combining GPS and star tracker data for ISS positioning. The data acquired via these sensors essentially yield (above-ground) height measurements, i.e., for each location, the signals returned in a diameter of about 25 m can be combined to obtain a single height value/estimate. Figure 2 shows a satellite image along with such (processed) GEDI height measurements. Note that ground-truth GEDI height measurements are only available for a fraction of the pixels.[2]

2.3. Generating Height Maps

The goal of a height estimation model is to provide a height estimate for each location/pixel on Earth/in a given satellite image, see again Figure 2. The output of three height models is shown in Figure 3. In all three cases, satellite imagery was used as input data. Potapov et al. (2021) proposed the first high-resolution (30 m) global canopy height

---

[1]It is worth noting that airborne data are often provided at a much higher resolution (e.g., up to resolution of 10 cm). However, these data are generally only available at a country-level, often with limited public access. This is the reason why the satellite
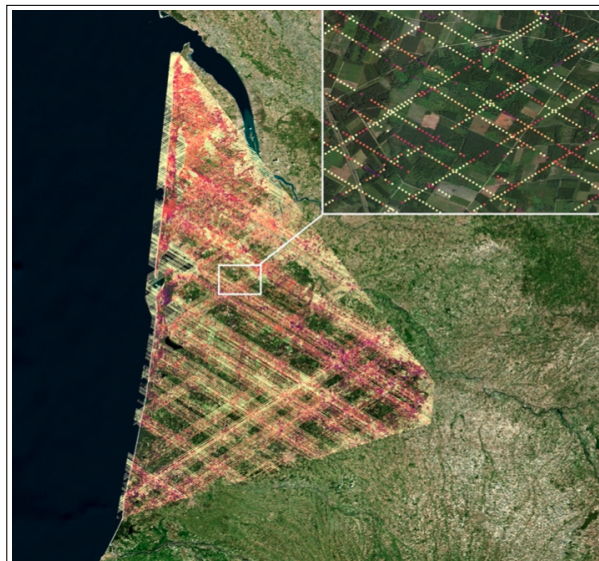
images generally form the basis for global-scale height maps.

[2]Over its intended two-year lifespan, GEDI was expected to scan 4% of the Earth's (vegetation) height, making it the most relied-upon source for canopy height measurements due to its global coverage.
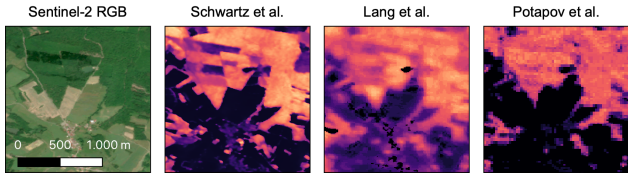
Figure 3. Visual comparison of a regional map (Schwartz et al., 2023) and two global maps (Lang et al., 2023; Potapov et al., 2021) (heights from 0 m to 35 m; see Figure 1 for the colormap).

map in 2011 using satellite imagery of the Landsat program (Williams et al., 2006). To obtain a height estimate per pixel, a random forest model was trained using GEDI height measurements. More recently, fully convolutional neural networks and vision transformers have been used to generate height maps at a country-level using Sentinel-1/-2 data, such as the maps provided by Schwartz et al. (2023) and by Fayad et al. (2023), which both exhibit a resolution of 10 m.[3] Using similar techniques, Liu et al. (2023) developed a canopy height map for Europe using 3m resolution commercial satellite imagery and so-called airborne laser scanning (ALS) data as ground truth. The most recent wall-to-wall (i.e. with global coverage) height map was provided by Lang et al. (2023), who also used Sentinel-1/-2 data with a resolution of 10 m and GEDI labels for training.

Hence, several regional (Schwartz et al., 2023; Fayad et al., 2023; Liu et al., 2023) as well as two wall-to-wall height maps (Potapov et al., 2021; Lang et al., 2023) have been proposed in the recent past. Note that the quality of regional maps is generally higher compared to one of the global maps. In this work, we close this quality gap by providing a wall-to-wall canopy height map with a resolution of 10 m, whose quality is comparable with the ones of the local products, while providing height estimates at a global scale.

3. Approach

We introduce a methodology designed to produce high-quality, high-resolution global forest canopy height maps, addressing the previously mentioned challenges. Selecting appropriate input and label data is crucial. Next, we detail the corresponding selection and preprocessing process, our model architecture, and the training approach. The overall pipeline is provided in Figure 4 and comprises the following steps: (A) data collection and preprocessing, (B) model training, and (C) global-scale inference.

3.1. Data Collection and Preprocessing (A)

For the Sentinel-1 data, we adopt a methodology similar to that used by Schwartz et al. (2024), i.e., we calculate the per-

pixel median for images taken during the summer leaf-on season (April to October 2020 in the northern hemisphere and October 2019 to April 2020 in the southern hemisphere). For the Sentinel-2 data, a distinct method is necessary due to significant cloud coverage in some areas, especially in rainforest regions. Here, we resort to a cloud reduction algorithm adapted from Braaten (2024) to minimize cloud artifacts, cloud shadows, and cirrus in the data. In total, we consider 14 channels/bands (4 from Sentinel-1 and 10 from Sentinel-2). We refer the reader to Appendix A.2 for the implementation details and to Appendix D for examples. As ground-truth labels, we resort to the data provided by the GEDI mission. In particular, we make use of the so-called RH100 metric, which measures the height at which 100% of the returned signals are registered. To enhance the quality of these labels, we consider various filtering steps, such as distinguishing between night and daytime shots to avoid solar radiation disturbing the height measurements. The total volumes of the preprocessed Sentinel-1 and Sentinel-2 images and the GEDI data are 45 TB and 1 TB, respectively.

Typically, in the context of GEDI, the canopy height is calculated as the difference between the first and last contact point of the signal (see Figure 5). In areas with steep slopes, however, this approach often inaccurately treats unforested slopes as high canopy or leads to a significant overestimation of tree height, see Figure 5 for an illustration of the latter issue. To mitigate the impact induced by incorrect canopy height values in mountainous areas, we resort to data from the Shuttle Radar Topography Mission (SRTM), which has mapped the Earth's surface topography with approximately 1 arc second resolution (about 30 m). It is important to note the distinction between the surface return, which is the initial contact point for measurements (such as tree canopies or building surfaces), and the terrain return, which is the actual ground level. The SRTM data, capturing the surface return, show that forest edges reflect a height change equal to the tree height, thereby complicating slope measurement filtering. Hence, for training, we exclude all GEDI measurements where the corresponding SRTM slope exceeds 20°.[4]

Overall, we generate data samples by selecting satellite image tiles uniformly at random. For each image, we extract (the same number of) patches of size 512×512 pixels and for each such patch, we extract the associated GEDI height measurements as labels, resulting in about 10 to 400 height measurements per patch. Overall, 100 000 patches are extracted, out of which 80% are used for training, 10% for validation, and 10% for testing. Patches extracted from an image tile belong to only one of these three sets, ensuring no overlap. More details are provided in the appendix.

---

[3]Such data products are often based on manual preprocessing steps that are tailored to the specific needs of the region at hand.

[4]As we do not want to discard any valid measurement, we apply a 20° filter threshold (a forest edges with trees up to 40 m in height already corresponds to a slope of 15°).
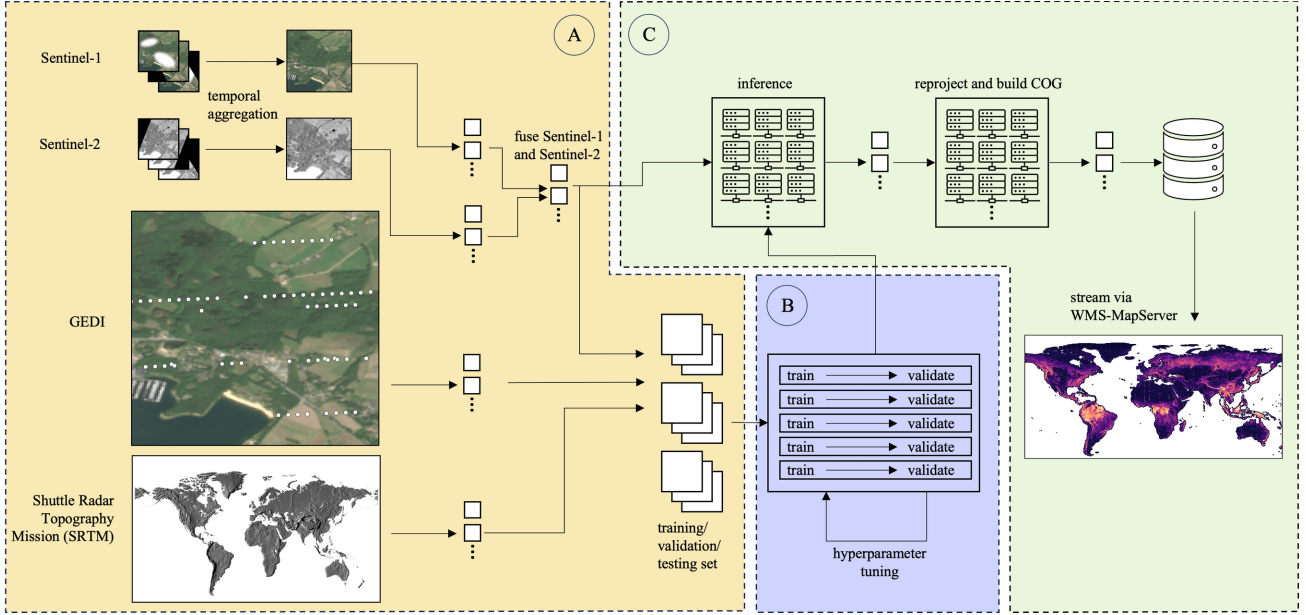
Figure 4. Our approach to generate height maps at a global scale: (A) data collection and preprocessing, (B) model training, and (C) global-scale inference. Initially, image composites based on Sentinel-1 and Sentinel-2 imagery are constructed and corresponding GEDI measurements as well as SRTM data are collected. After training the model using hyperparameter grid search, it is applied in a distributed manner, with the canopy height estimates being subsequently reprojected and made available through streaming services.
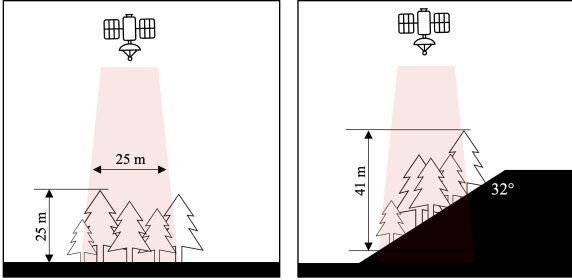

Figure 5. The GEDI instrument records the signals returned from the ground within a 25 m diameter. The height is then essentially computed based on the differences of the first and last signals. While satisfying height values are obtained for most areas, this often leads to inaccurate height measurements for slopes.

3.2. Model Training (B)

Next, we describe the model architecture and the technical details of the training and hyperparameter tuning process.

3.2.1. MODEL ARCHITECTURE & PARAMETERS

Let $\mathcal{X} = \mathbb{R}^{n_c \times w \times h}$ be the input space containing images with n_c channels (in our case $w = 512$, $h = 512$, and $n_c = 14$) and let $\mathcal{Y} = \mathbb{R}^{w \times h}$ be the output space. The goal is to train a height estimation model $f : \mathcal{X} \to \mathcal{Y}$ that assigns one height estimate to each of the input pixels. Given the global application scope, achieving a balance between prediction accuracy and inference time is essential.

Following Schwartz et al. (2024), we resort to the well-known U-Net architecture (Ronneberger et al., 2015), which is a semantic segmentation model. We set the number of output classes to one and make use of a linear activation for the final layer in order to estimate the GEDI heights. We also replace the original architecture's backbone by a ResNet50 (He et al., 2016) backbone. We optimize the model weights using the AdamW (Loshchilov & Hutter, 2017) optimizer with a weight decay of 0.001, a batch size of 32, and an initial learning rate of 0.001. We also resort to a linear learning rate warm-up for the first 10% of the iterations and a linear learning rate scheduler for the remaining 90% of the iterations. To address the skewed label distribution (also see Section 4.2) in our training set, a weighted sampler is used. The validation set was used for hyperparameter selection. We refer to the appendix for more details related to the hyperparameter search and the training process.

3.2.2. SHIFT-RESILIENT LOSS

When inspecting GEDI measurements, it can be observed that their geolocation is not always precise, i.e., a systematic pattern in geolocation errors is often given in the data, characterized by consistent shifts in each track of measurements, see Figure 6. These shifts are problematic since correct predictions made by a model might be considered to be wrong by standard loss functions. A recent approach employs a high-resolution (1m) digital terrain model (DTM)
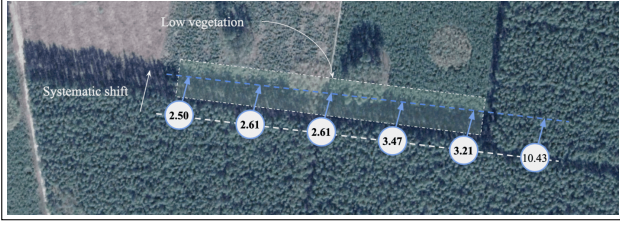
Figure 6. The geolocations of GEDI measurements often exhibit a systematic shift, leading to faulty ground-truth labels (each circle shows a height measurement; background image: Google Maps).

to individually correct GEDI measurements based on their ground return (Schleich et al., 2023). While this method enhances the geolocation accuracy of GEDI measurements, it is not feasible for global-scale applications due to a lack of up-to-date DTMs at a global scale.

We propose a shift-resilient loss function tailored to this learning scenario that can be added on top of any other regression loss function. The basic idea is that we allow each track to be shifted in any direction by a magnitude smaller than a radius $r > 0$ and select the shift with the lowest associated loss. More specifically, for a given input instance $I \in \mathcal{X}$, let $T_I = \{t_1, \ldots, t_n\}$ be the corresponding set of GEDI tracks containing the (potentially shifted) ground-truth height measurements. Here, each of the tracks $t \in T_I$ is composed of a set $t = \{m_1, \ldots, m_{n_t}\}$ of ground-truth measurements and each such measurement $m \in t$ is represented by $m = (m_x, m_y, m_l)$, where m_x and m_y depict the pixel coordinates and m_l the height value. Note that the measurements belonging to one track are on a "line". For example, the "GEDI" input image patch on the left hand side of Figure 4 contains five tracks, where each white dot on a track represents one measurement.

For a track $t \in T_I$ in a set of tracks T_I associated with an input image $I \in \mathcal{X}$, we define $t_I \in \mathcal{Y}$ as

$$t_I(x, y) \coloneqq \begin{cases} m_l & \text{if } \exists\, m \in t \text{ s.t. } (x, y) = (m_x, m_y), \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

For each such t_I, we also define a $t_I^0 \in \{0, 1\}^{w \times h}$ as

$$t_I^0(x, y) \coloneqq \begin{cases} 1 & \text{if } t_I(x, y) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

For an $I \in \mathcal{X}$ with corresponding height estimates $f(I)$, the (non-shifted) loss $\mathcal{L}_{NS}$ can be computed as

$$\mathcal{L}_{NS}\left(f(I), T_I\right) \coloneqq \frac{1}{N} \sum_{t \in T_I} \mathcal{L}\left(f(I) \odot t_I^0, t_I\right), \quad (3)$$

where $N = \sum_{t \in T_I} |t|$, $\odot$ the element-wise Hadamard product, and $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$ a standard (pixel-wise) loss for regression (e.g., L_2 loss).

As mentioned above, the geolocations of the GEDI measurements are often not precise. Hence, the loss function $\mathcal{L}_{NS}$ is not an ideal choice to assess the quality of $f(I)$ since $f(I)$ would be compared with ground-truth values inaccurate in their geolocation, potentially leading to a high loss although the model output might be correct. Note, however, that the shift is (relatively) consistent across all the GEDI measurements belonging to the *same* track. Hence, to compensate for possible (unknown) systematic geolocation shifts $\delta = (\delta_x, \delta_y) \in \mathbb{Z}^2$ of the GEDI measurements belonging to the same track, we define the following shifted ground-truth targets $t_{I,\delta}(x, y) \coloneqq t_I(x - \delta_x, y - \delta_y)$ (measurements close to any of the borders are handled appropriately). Then, we consider the following (shifted) loss function:

$$\mathcal{L}_S\left(f(I), T_I\right) \coloneqq \frac{1}{N} \sum_{t \in T_I} \min_{\delta} \mathcal{L}\left(f(I) \odot t_{I,\delta}^0, t_{I,\delta}\right) \quad (4)$$

$$\text{s.t.} \quad \sqrt{\delta_x^2 + \delta_y^2} \leq r.$$

Thus, all shifts meeting the constraint are considered, and the one with the lowest pixel-wise loss is chosen.[5] In our case, we choose the pixel-wise Huber loss for $\mathcal{L}$, as it is more robust to outliers than the L_2 loss. The variable r is a user-defined parameter specifying the allowed shift radius. For the GEDI measurements, it makes sense to restrict r to about $\sqrt{2}$ as 80.8% of the GEDI tracks exhibit a mean geolocation error below 10 meters (Tang et al., 2023).

3.3. Global-Scale Inference (C)

After training, the model is deployed globally by partitioning the Earth into patches of size 312×312 pixels. For predictions, we consider image patches of size 512×512 pixels, i.e., we include 100 pixels at each border for context (the predictions within these borders are ignored for the final map/mosaic). The computational process consumes approximately $1\,500$ GPU hours on a GPU cluster with various GPU devices (e.g., Nvidia RTX3090s and A100s).

For an efficient visualization of the map, we resort to several postprocessing steps. Initially, we standardize the map projection across all predictions to the EPSG:3857 (Web Mercator) projection. Subsequently, we convert predictions into the cloud optimized GeoTiff (COG) format. This conversion step involves, among other things, data compression and the creation of high-level overviews. These steps are computationally very demanding, with reprojection and COG creation requiring about $81\,000$ CPU hours.

Finally, we resort to a web map service (WMS) to stream the canopy height map to any geographic information system software. We visualize the map using a magma color ramp,

---

[5] In case there are fewer than ten measurements available in a track t_I, we do not allow shifting for that track, as it is hard to estimate the systematic shift with only a few measurements.

Table 1. Comparison of our height map with two publicly available global-scale height maps (Potapov et al., 2021; Lang et al., 2023).

FILTER	METRIC	UNIT	LANG	POTAPOV	OURS
NO	MAE	m	6.47	6.92	2.43
	MSE	m^2	74.70	85.58	22.41
	RMSE	m	8.62	9.25	4.73
	RRMSE		1.39	2.38	0.53
	MAPE		1.01	0.97	0.22
$m_l > 5$	MAE	m	8.80	10.01	4.45
	MSE	m^2	121.51	154.45	45.12
	RMSE	m	11.02	12.43	6.72
	RRMSE		1.02	1.77	0.49
	MAPE		0.76	0.77	0.28

a common choice for depicting canopy heights. A high-level overview of the final map is presented in Figure 1 (a more detailed version can be found in the appendix).

3.4. Source Code & Canopy Height Map

Our source code, as well as detailed documentation, are publicly available on GitHub.[6] The induced global canopy height map is accessible through the Google Earth Engine (see Appendix F).

4. Results

We provide a comparison of our height map with two other global height map products commonly used in the field (Potapov et al., 2021; Lang et al., 2023) and with one regional height map that has recently been produced for France (Schwartz et al., 2023). We also analyze the errors still made by our model and the impact of key filtering steps.

4.1. Comparison of Height Maps

We resort to several performance metrics as well as to visual examples in order to compare the quality of the height maps.

4.1.1. METRIC-BASED COMPARISON

We consider the mean absolute error (MAE), the mean squared error (MSE), the root mean squared error, the relative root mean squared error (RRMSE), and the mean absolute percentage error (MAPE) to asses the quality of the different height map products (all maps are compared using the same non-shifted GEDI data). All numbers the provided are obtained using the test area (see Appendix B). No test instances have been used for training and model selection.

The results are shown in Table 1. As Lang et al. (2023) and Potapov et al. (2021) both use a forest mask, the metrics are shown with no filters applied as well as only for GEDI mea-

surements larger than 5 m. The top part of the table presents the results without any additional filter being applied to the GEDI data. It can be seen that our height map yields a significantly smaller MAE, standing at 2.43 m, compared to the maps of Lang et al. (2023) and Potapov et al. (2021), which result in an MAE of 6.47 m and 6.92 m, respectively. Similar improvements are given for the other metrics. Note that these values do not account for the fact that a GEDI label of 3 m corresponds to the ground, but most height maps resort to forest masks in order to set the corresponding height values to 0 m. This results in a consistent 3 m error for all accurately classified ground predictions. To address this, we also provide a comparison using only the GEDI height measurements larger than 5 m ($m_l > 5$). For these filtered data, a similar quality gain of our map compared to the other two existing height map products can be observed.

4.1.2. VISUAL COMPARISON

Next, we compare the maps using various visual examples. Here, visual quality refers to a map's appearance, particularly at forest edges, patches, and in terms of "resolution". Some examples are given in Figure 7. In addition to the two global height maps (Lang et al., 2023; Potapov et al., 2021), we also consider the regional map for France recently proposed by Schwartz et al. (2023).

We start with a comparison of the two global map products and our map. The map by Lang et al. (2023) generally yields a smoother texture compared to our map, leading to some loss of detail. For instance, it sometimes fails to accurately depict forest structure (as it can be seen, e.g., in the first column). The map also yields the highest predictions, which sometimes leads to fewer details in low-height areas (visible in, e.g., the fourth column). The map of Potapov et al. (2021) is derived from Landsat data with a resolution of 30 m. Hence, naturally, it appears to be more coarse compared to the other maps (which are based on Sentinel-1/Sentinel 2 imagery with a resolution of 10 m). It also struggles with capturing forest variances. In particular, it often renders forest areas using an almost uniform height and, thus, does not capture smaller forest patches and fine structural details. This is, for instance, visible in the fifth and the last column of the table. In contrast, our map often more accurately identifies fine structural details, such as the pathways and small forest patches (see, e.g., the first and the fifth column).

Comparing our height map with the regional map for France provided by Schwartz et al. (2023), we observe notable similarities between both maps. Compared to the other two global maps, the most significant distinction lies in the ability of these two maps to identify fine structures such as forest gaps, roads within forests, and adjacent forest areas. For instance, in the fifth column, our map successfully identifies multiple gaps and roads, a detail both global maps

---

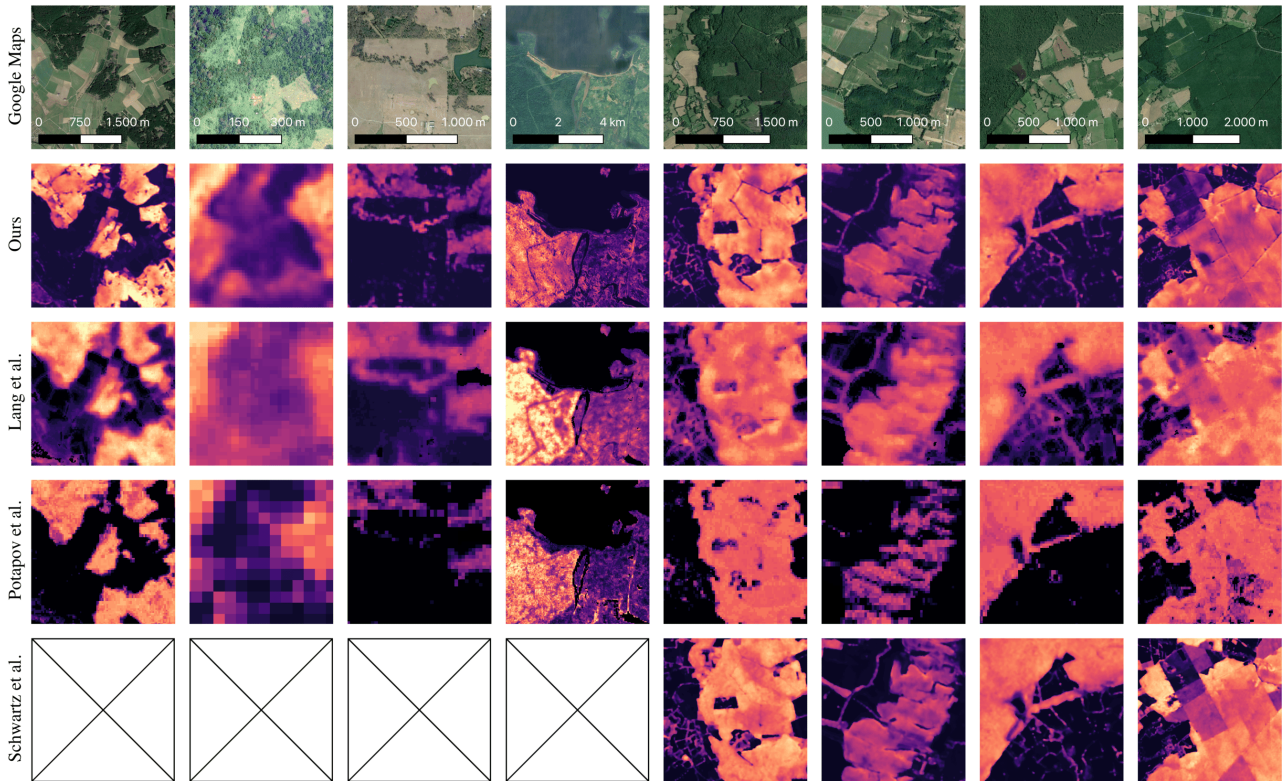[6]https://github.com/AI4Forest/Global-Canopy-Height-Map

Figure 7. Visual comparison of our map with two global height maps (first four columns) (Lang et al., 2023; Potapov et al., 2021) and a regional map for France (Schwartz et al., 2023) (last four columns). The first row provides high-resolution Google Maps imagery for visual context. Our canopy height map shows an improvement compared to existing global canopy height maps in terms of clarity and detail for all eight examples. The visual quality of our map for the four examples in France is close to the one of the regional map (heights from 0 m to 35 m; see Figure 1 for the colormap).

do no capture. Notably, even the regional map fails to accurately depict the gap line in the upper half of the image.[7]

4.2. Error Analysis & Filtering Steps

Next, we analyze the remaining errors in more detail as well as the effect of the SRTM filtering step.

4.2.1. UNDERESTIMATION OF HIGH HEIGHTS

In Figure 8, a scatter plot comparing GEDI ground-truth measurements and our height estimates is provided. The logarithmic scale of density in the scatter plot illustrates the label imbalance present in the GEDI data, with a mean height of 6.33 m and a standard deviation of 7.17 m. Measurements at 40 m are approximately 423 times rarer than those at 3 m, and 90% of the measurements are below 14 m, showing a pronounced bias towards lower height measurements. The use of the shifted loss function significantly reduces completely incorrect predictions (i.e., errors "located" close to the x- or y-axis), which are typically prevalent in canopy height maps. However, our model still underestimates the ground-truth measurements, which is especially the case for high heights (e.g., $m_l > 30$ m). Note that this is a common problem across the available height maps.

Figure 9 presents boxplots for all three global maps, where the errors are sorted according to the ground-truth GEDI height measurements in bins of 10 m. Our map shows significantly lower error rates for vegetation heights under 20 m compared to the other two maps. For heights up to 10 m, the error rate for the map of Potapov et al. (2021) is comparable to that of the map of Lang et al. (2023). However, for heights greater than 10 m, it deteriorates, generally falling 5–10 m short of both other maps. While our map shows an improved accuracy for heights up to 20 m, this advantage does not extend to greater heights, where our errors align more closely with those of Lang et al. (2023).[8] Across all

---

[7]This enhanced detail in our map could be attributed to the shifted loss function, which makes the model more resilient against the label noise present in the ground-truth GEDI measurements. This seems to effectively reduce smoothing at ambiguous locations.

[8]The approach of Lang et al. (2023) strongly aimed at addressing the underestimation of high heights/tall trees, which results in the overestimation of smaller heights, as indicated by the +20 m whiskers in the 10 m to 30 m height ranges.
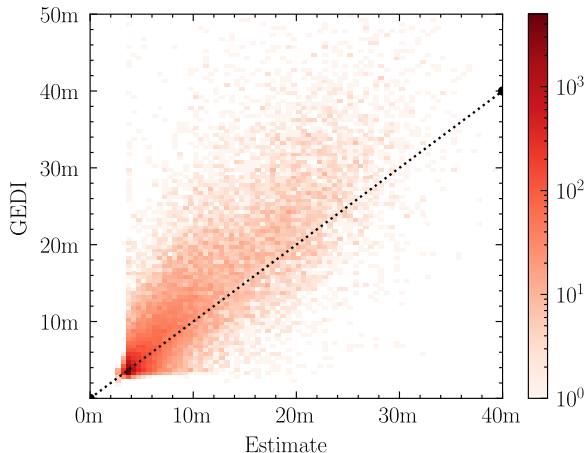
Figure 8. Scatter plot showing height estimates versus ground-truth GEDI height measurements (R^2: 0.72; log scale).
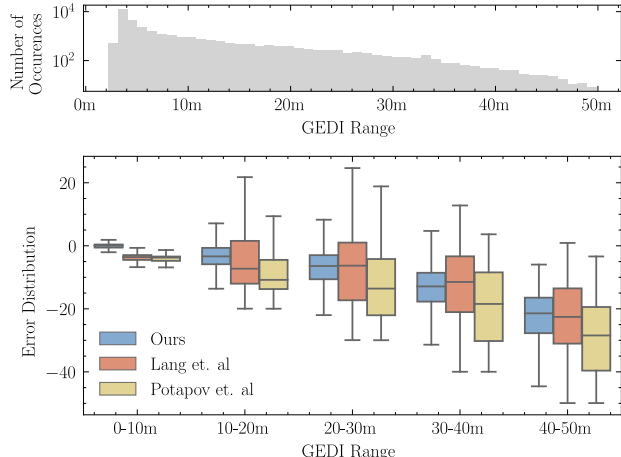

Figure 9. Top: Distribution of GEDI measurements (log scale). Bottom: Evaluation of the errors given different height ranges from 0 m to 60 m. Here, negative errors indicate that the estimates are smaller than the ground-truth GEDI measurements.

height bins, our map consistently shows the lowest variance, particularly notable in trees up to 30 m, where the reduction in variance is significantly prominent.

4.2.2. SRTM FILTERING FOR MOUNTAINOUS REGIONS

We resort to a filtering step for the GEDI height measurements for mountainous regions using SRTM data, see Section 3.1. While the SRTM provides valuable surface topography data, it is important to note that it captures surface elevation rather than pure terrain elevation. Our use of SRTM data is, due to this limitation, not a perfect solution for improving height estimation in mountainous regions, see Figure 10. However, incorporating SRTM data still represents an improvement over not addressing the topographical challenges at all. Hence, the implementation of a filter based on SRTM data predominantly ensures the exclusion of unrealistically high canopy predictions in steep terrains.

5. Conclusion

We propose a comprehensive framework for generating precise height maps at a global scale, utilizing both Sentinel and GEDI data for training. Our framework is built upon fully convolutional neural networks and utilizes cloud-free satellite imagery composites, a novel loss function designed to mitigate GEDI geolocation inaccuracies, and several preprocessing steps. It enables the generation of high-resolution (10 m) global-scale height maps that stand on par with or even surpass the quality of specialized regional height map products. Such height maps play a crucial role in understanding Earth's carbon dynamics and are instrumental in addressing climate change mitigation efforts.
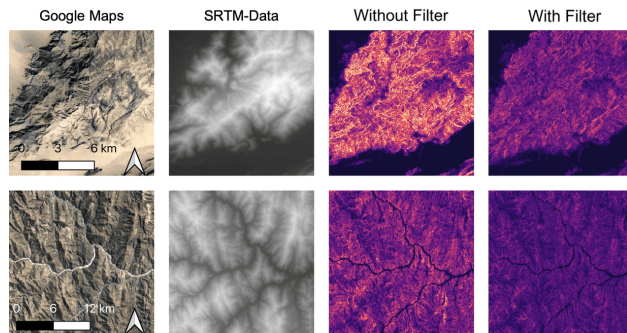
Figure 10. Visual illustration of the impact of the SRTM filtering step. The SRTM filtering drastically reduces the overestimation of high heights in mountainous areas (heights from 0 m to 35 m; see Figure 1 for the colormap).

Acknowledgements

Impact Statement

This work aims to enhance the accuracy of forest monitoring through machine learning techniques. Currently, forests absorb approximately half of the CO_2 emissions generated by human activities and play a crucial role in mitigating global warming (Friedlingstein et al., 2022). However, they are increasingly vulnerable to the impacts of climate change and direct land use changes, such as deforestation and degradation (Anderegg et al., 2022). In line with the UN's Sustainable Development Goals (SDGs)[9], the Bonn Challenge on forests[10], and the Glasgow Declaration of halting forest loss, managing and conserving forests is an indispensable component of climate adaptation and mitigation strategies. Moreover, numerous private entities engage in carbon credit investments, financing projects for forest growth and tree planting within the voluntary carbon market. Accurate monitoring of these projects is essential to ensure their additional value and permanence, thereby mitigating potential leakage effects.[11]

Forest conservation, regeneration, afforestation, and reforestation are key ingredients of climate mitigation policies in future scenarios that meet the goals of the Paris Agreement on climate.[12] Up to now, the means to assess forest carbon stocks and their changes over time has been through labor-intensive field inventories. Although inventories provide robust estimates of carbon stocks at national scale, their sampling is too sparse to allow monitoring of regional and local changes, and their infrequent revisit cycle does allow to capture shocks causing abrupt forest loss like fires, insects attacks, windthrown events. Moreover, most tropical forested countries do not have an inventory and use default growth rates and simple methods to estimate their forest carbon changes (Grassi et al., 2022).

Precise and up-to-date wall-to-wall high-resolution information about the structure, health, and carbon stocks of forests in the world is critical to assess the current state of forest carbon sinks, predict their future state, and trigger appropriate measures against forest loss. The framework developed in this work aids in providing a consistent and global height map utilizing publicly available satellite data from optical, radar, and GEDI spaceborne sensors. This map can be used as a reference to derive more accurate biomass carbon stocks, as a baseline for policymakers to make more accurate decisions about forest management, and further related policies.

---

[9]https://sdgs.un.org/2030agenda

[10]https://www.bonnchallenge.org

[11]https://www.theguardian.com/environment/2023/jan/18/revealed-forest-carbon-offsets-biggest-provider-worthless-verra-aoe

[12]https://www.ipcc.ch/report/sixth-assessment-report-workinggroup-3/

References

Anderegg, W. R., Wu, C., Acil, N., Carvalhais, N., Pugh, T. A., Sadler, J. P., and Seidl, R. A climate risk analysis of Earth's forests in the 21st century. *Science*, 377(6610): 1099–1103, 2022.

Braaten, J. Sentinel-2 Cloud Masking with s2cloudless. https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless, 2024. Accessed 2024-05-15.

Chaurasia, A. and Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4. IEEE, 2017.

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818, 2018.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1:100002, 2020.

European Space Agency. Sentinel-1 – radar vision for Copernicus. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1, 2024a. Accessed: 2024-05-15.

European Space Agency. Sentinel-2 – colour vision for Copernicus. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2, 2024b. Accessed: 2024-05-15.

Fan, T., Wang, G., Li, Y., and Wang, H. Ma-Net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020.

Fayad, I., Ciais, P., Schwartz, M., Wigneron, J.-P., Baghdadi, N., de Truchis, A., d'Aspremont, A., Frappart, F., Saatchi, S., Pellissier-Tanon, A., and Bazzi, H. Vision transformers, a new approach for high-resolution and large-scale mapping of canopy heights. *arXiv preprint arXiv:2304.11487*, 2023.

Fort, S., Hu, H., and Lakshminarayanan, B. Deep ensembles: A loss landscape perspective. *arXiv preprint arXiv:1912.02757*, 2019.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L. M., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, O., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G. R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J. Global carbon budget 2021. *Earth System Science Data*, 14(4):1917–2005, 2022.

Ganaie, M., Hu, M., Malik, A., Tanveer, M., and Suganthan, P. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, 2022.

Grassi, G., Conchedda, G., Federici, S., Abad Viñas, R., Korosuo, A., Melo, J., Rossi, S., Sandker, M., Somogyi, Z., Vizzarri, M., and Tubiello, F. N. Carbon fluxes from land 2000–2020: bringing clarity to countries' reporting. *Earth System Science Data*, 14(10):4643–4666, 2022.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

Hu, T., Zhang, Y., Su, Y., Zheng, Y., Lin, G., and Guo, Q. Mapping the global mangrove forest aboveground biomass using multisource remote sensing data. *Remote Sensing*, 12(10), 2020.

Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D., and Wilson, A. G. Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407*, 2018.

Kacic, P., Thonfeld, F., Gessner, U., and Kuenzer, C. Forest structure characterization in Germany: Novel products and analysis based on GEDI, Sentinel-1 and Sentinel-2 data. *Remote Sensing*, 15(8), 2023.

Lang, N., Jetz, W., Schindler, K., and Wegner, J. D. A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*, 7(11):1778–1789, 2023.

Li, H., Xiong, P., An, J., and Wang, L. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*, 2018.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recogfnition*, pp. 2117–2125, 2017.

Liu, S., Brandt, M., Nord-Larsen, T., Chave, J., Reiner, F., Lang, N., Tong, X., Ciais, P., Igel, C., Pascual, A., Guerra-hernandez, J., Li, S., Mugabowindekwe, M., Saatchi, S., Yue, Y., Chen, Z., and Fensholt, R. The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. *Science Advances*, 9(37), 2023.

Loshchilov, I. and Hutter, F. Fixing weight decay regularization in Adam. *CoRR*, abs/1711.05101, 2017.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., and Hofton, M. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253: 112165, 2021.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.

Schleich, A., Durrieu, S., Soma, M., and Vega, C. Improving GEDI footprint geolocation using a high-resolution digital elevation model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 7718–7732, 2023.

Schwartz, M., Ciais, P., De Truchis, A., Chave, J., Ottlé, C., Vega, C., Wigneron, J.-P., Nicolas, M., Jouaber, S., Liu, S., Brandt, M., and Fayad, I. FORMS: Forest multiple source height, wood volume, and biomass maps in France at 10 to 30 m resolution based on Sentinel-1, Sentinel-2, and Global Ecosystem Dynamics Investigation (GEDI) data with a deep learning approach. *Earth System Science Data*, 15(11):4927–4945, 2023.

Schwartz, M., Ciais, P., Ottlé, C., de Truchis, A., Vega, C., Fayad, I., Brandt, M., Fensholt, R., Baghdadi, N., Morneau, F., Morin, D., Guyon, D., Dayau, S., and Wigneron, J.-P. High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach. *International Journal of Applied Earth Observation and Geoinformation*, 128:103711, 2024.

Sloan, S. and Sayer, J. A. Forest resources assessment of 2015 shows positive global trends but forest loss and degradation persist in poor tropical countries. *Forest Ecology and Management*, 352:134–145, 2015.

Smith, L. N. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472. IEEE, 2017.

Tang, H., Stoker, J., Luthcke, S., Armston, J., Lee, K., Blair, B., and Hofton, M. Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI. *Remote Sensing of Environment*, 291:113571, 2023.

Williams, D. L., Goward, S., and Arvidson, T. Landsat. *Photogrammetric Engineering & Remote Sensing*, 72 (10):1171–1178, 2006.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890, 2017.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. UNet++: Redesigning skip connections to exploit multi-scale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2019.

Zimmer, M., Spiegel, C., and Pokutta, S. How I Learned To Stop Worrying And Love Retraining. In *International Conference on Learning Representations*, 2023.

Zimmer, M., Spiegel, C., and Pokutta, S. Sparse model soups: A recipe for improved pruning via model averaging. In *International Conference on Learning Representations*, 2024.

A. Data

A.1. Sentinel-1

For Sentinel-1, we use all available images from the Sentinel-1 Ground Range Detected (GRD) dataset, taken between April and October 2020 for the northern hemisphere, and between October 2019 and April 2020 for the southern hemisphere (European Space Agency, 2024a). This difference in time frame accounts for the varying seasonality. Summer leaf-on images are particularly useful for estimating canopy height, so we aim to obtain a summer leaf-on composite for each hemisphere.

We use Sentinel-1 satellite images acquired in both VV (Vertical Transmit, Vertical Receive) and VH (Vertical Transmit, Horizontal Receive) polarization. Given the substantial differences in the data acquired during ascending and descending satellite orbits, we categorically separate the images from different orbits. Although Sentinel-1 is mostly unaffected by clouds and general weather, using single images as input data for the model is insufficient. Sentinel-1 images often encounter issues with noise (e.g., speckle or thermal noise), and heavy rain can disrupt image quality. We follow a common approach for addressing these issues (which we follow) is to use the temporal per-pixel median, which effectively eliminates most of the noise problems. By following this approach, we generate four distinct channels: VV from ascending orbits, VV from descending orbits, VH from ascending orbits, and VH from descending orbits.

A.2. Sentinel-2

For Sentinel-2, there are different data products available. We use all Sentinel-2 bottom-of-the-atmosphere (BOA) images from the same time frame as the Sentinel-1 data (European Space Agency, 2024b). Typically, one resorts to a similar temporal composition approach for Sentinel-2 as for Sentinel-1 data. However, such temporal per-pixel median composites are often insufficient for Sentinel-2 data, since the corresponding optical sensor cannot penetrate clouds. Hence, while taking the temporal per-pixel median might yield cloud-free image composites for most regions, such an approach is generally not suited for rainforest regions, where the cloud coverage is often persistent throughout the year. These regions are crucial though for carbon monitoring as they store a significant portion of global carbon. Therefore, we deviate from the common procedure and make use of the following steps to generate Sentinel-2 image composites:

1. Instead of directly taking the temporal per-pixel median, we first filter out clouds in each image before aggregating the remaining pixels. We do this by adapting an approach from Braaten (2024). More precisely, each Sentinel-2 image comes with a so-called cloud mask that contains the (estimated) cloud cover/probability for each pixel in that image. We make use of this mask to filter out images with less than 10% cloud-free pixels.

2. We then employ multiple steps to identify and remove disturbed pixels. First, we identify and remove pixels with a cloud probability larger than 30%. Knowing the sun angle at the time of image capture, we examine each cloud pixel up to 1 km in the (opposite) direction of the sun and search for potential shadow/dark pixels. To identify those dark pixels, we make use of the near-infrared band B8 and the scene-classification band SCL. We then remove, for each pixel identified as cloud or shadow, all neighbored pixels within a 300 m radius.

3. Finally, we aggregate all the remaining pixels for each image using a temporal per-pixel median.

The approach described above generally yields reasonable image composites containing no "gaps". For very few areas, no valid pixels might remain, thus leading to "empty" regions in the input data, see the last row of Figure 12.

A.3. GEDI

To enhance the quality of the GEDI measurements, we apply several filters:

1. $beam\_int > 5$: The GEDI instrument captures data via 8 tracks, 4 of those are measured by only one beam, the so-called "coverage" beam. This beam has lower power, making its measurements more susceptible to noise. For this reason, we discard those tracks and only resort to the measurements of the remaining so-called "power" beams.

2. $quality\_flag = 1$: This flag filters out the GEDI measurements with a poor quality, i.e., those measurements that were significantly disturbed by clouds or adverse weather conditions.

3. $solar\_elevation < 0$: To minimize the impact of solar radiation on GEDI measurements, we exclude all daytime measurements, retaining only those taken during nighttime.
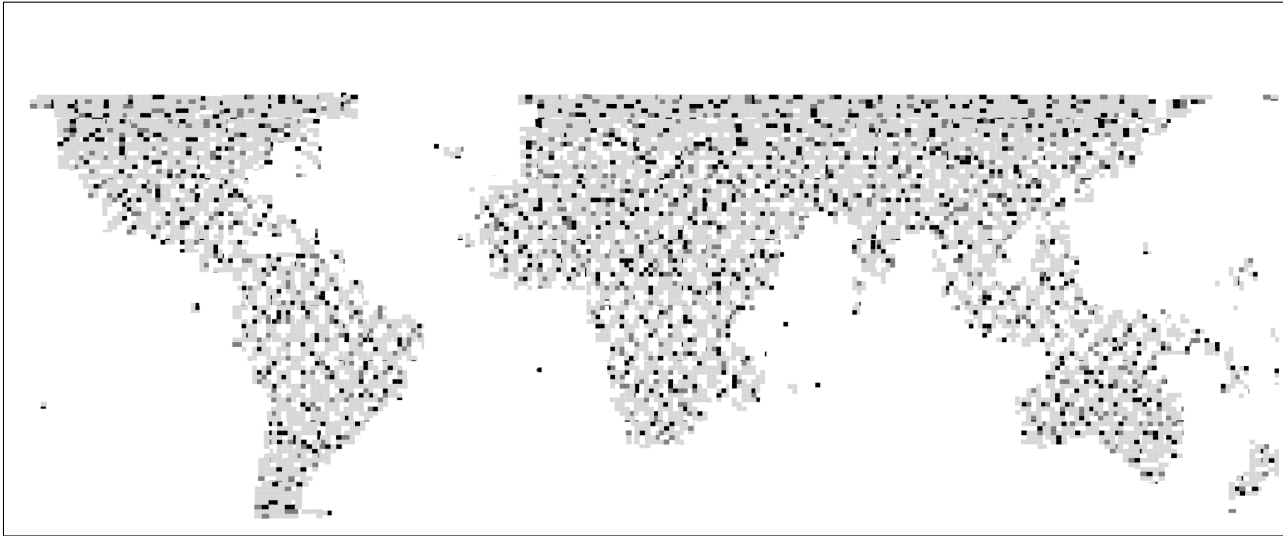
Figure 11. Training (light grey), validation (dark grey), and test (black) regions.

A.4. SRTM

GEDI measurements might be incorrect for mountainous regions (cf. Figure 5). GEDI determines (vegetation) height by calculating the time difference between the first and the last of the returned signals. For example, by multiplying a time difference of 112 ns between the first and the last signal with the speed of light (15 cm/ns) results in a height of about 16.8 m. On slopes, the first and last signals may not originate from the same object/tree, but from nearby objects/trees of *different* heights, leading to an overestimation of the canopy height. Additionally, even in the absence of any object/vegetation, the first and last signals may be reflected at different heights, causing GEDI to register a positive height for these areas as well.

Ideally, the exact slope at each measurement location should be determined using a high-resolution digital terrain model (DTM), which would allow to identify and exclude these faulty measurements. However, such DTMs are not publicly available at a global scale. Therefore, we resort to data of the SRTM mission, which mapped the Earth's surface height with a resolution of about 30 m. Although surface height data are not perfect in this context (since forest borders can also appear as slopes), they currently depict the best publicly available data source for the identification of the critical areas. To improve the label quality, we filter out any measurement with a SRTM slope of $20°$ or larger. We calculate the slope by taking the height difference within an area of size 5×5 pixels (corresponding to 150×150 m). To ensure no valid GEDI measurement is discarded, we set the filter threshold above the possible slope at forest borders. More precisely, if a forest border has trees of height 40 m, the slope is approximately $15°$. Hence, we set the threshold to $20°$, allowing for a height change (excluding the forest border) of up to 15 m.

B. Training/Validation/Test

We split the data into geographically non-overlapping regions for training, validation, and testing. An illustration of the distribution of training, validation, and test instances can be seen in Figure 11. Each Universal Transverse Mercator (UTM) zone has, relative to its size, the same proportion of training, validation, and test instances. Note that the northern regions are not covered due to the ISS orbit not covering this area.

C. Model

To fine-tune our model, we conduct a comprehensive hyperparameter search, see Table 2. The final model was trained from scratch using the AdamW optimizer with a weight decay of 0.001, a batch size of 32, and an initial learning rate of 0.001. We implemented a linear learning rate warm-up for the first 10% of the total iterations, followed by a linear learning rate scheduler for the remaining 90%. Additionally, gradient clipping was applied to prevent gradient explosion. Finally, the shifted Huber loss was used during training, i.e., $\mathcal{L}_S$ with the Huber loss as pixel-wise loss function $\mathcal{L}$ in Equation (4).[13]

---

[13]Using sample weights for the training instances led to a worse performance. Hence, no sample weights have been used for training.

Table 2. Model hyperparameters for fine-tuning.

HYPERPARAMETER	VALUES	DESCRIPTION
weight decay	$0, 0.01, 0.001$	Weight decay of the AdamW optimizer
shifted loss	$\mathcal{L}_S, \mathcal{L}_{NS}$	We consider both the standard non-shifted loss function $\mathcal{L}_{NS}$ defined in Equation (3) as well as the shifted variant $\mathcal{L}_S$ defined in Equation (4).
pixel-wise loss	L_1, L_2, Huber	Both $\mathcal{L}_{NS}$ and $\mathcal{L}_S$ are based on a pixel-wise loss $\mathcal{L}$. Here, we consider the L_1 (mean absolute error), L_2 (mean squared error), and the Huber loss as loss functions. For the Huber loss, we set the cutoff parameter δ to 3 (meters).
model weights	None, ImageNet	We use the ResNet-50 (He et al., 2016) backbone and train the weights either from scratch (None) or resort to pre-trained weights (that stem from the ImageNet (Deng et al., 2009) dataset).

Table 3. Ablation study on different loss function: L_1, L_2 and Huber loss both with and without $\mathcal{L}_S$. Each row represents the loss function used for training the model, while each column represents the loss function used for evaluating the model's performance.

VALIDATION LOSS TRAINING LOSS	L_1	L_2	Huber	$\mathcal{L}_S + L_1$	$\mathcal{L}_S + L_2$	$\mathcal{L}_S +$ Huber
L_1	1.68 ± 0.03	11.19 ± 0.25	1.04 ± 0.02	1.65 ± 0.03	10.82 ± 0.27	1.01 ± 0.02
L_2	1.76 ± 0.01	11.12 ± 0.12	1.05 ± 0.00	1.73 ± 0.01	10.76 ± 0.13	1.02 ± 0.00
Huber	1.73 ± 0.01	11.11 ± 0.18	1.04 ± 0.01	1.70 ± 0.01	10.73 ± 0.20	1.01 ± 0.01
$\mathcal{L}_S + L_1$	1.68 ± 0.03	11.15 ± 0.21	1.04 ± 0.02	1.64 ± 0.03	10.63 ± 0.19	1.00 ± 0.02
$\mathcal{L}_S + L_2$	1.75 ± 0.02	11.07 ± 0.08	1.05 ± 0.01	1.71 ± 0.02	10.57 ± 0.08	1.01 ± 0.01
$\mathcal{L}_S +$ Huber	1.70 ± 0.01	10.92 ± 0.16	1.03 ± 0.01	1.66 ± 0.01	10.42 ± 0.17	0.99 ± 0.01

D. Examples

Figure 12 shows Sentinel input data as well as the height estimates induced by our model for some more examples. Note that the last row sketches the (rare) situation in case no input pixels are considered to be valid (see above), leading to an empty area in the composite image. Here, some parts of the Sentinel-2 image were masked out due to a lack of non-cloud pixels. Yet, our model estimates the height reasonably well using the spatial context given in the Sentinel-2 composite as well as the Sentinel-1 data.

E. Ablation Studies

We conduct several ablation studies to evaluate the impact of the involved model components.

E.1. Loss Function

First, we start by assessing effectiveness of the shifted loss function. Our objective is to ascertain whether the use of $\mathcal{L}_S$ (see Equation (4)) enhances the model performance and to identify the most suitable pixel-wise loss function. We compare three commonly used loss functions (L_1, $L2$, Huber loss) and conduct experiments using both $\mathcal{L}_S$ and $\mathcal{L}_{NS}$. The results are given in Table 3. All models are trained with the same set of hyperparameters, except for the choice of loss function used during training (first column). We present the results for all metrics (remaining columns) based on the validation set. One can see that using $\mathcal{L}_S$ generally leads to equal or better results compared to $\mathcal{L}_{NS}$, which was expected since $\mathcal{L}_{NS}$ is a special case of $\mathcal{L}_S$ with $\delta_x = 0$ and $\delta_y = 0$.

E.2. SRTM-Filtering

Next, we assess the effectiveness of the SRTM filtering approach. For this, we resort to Airbone Laser Scanning (ALS) data from the Needle Mountains in the USA (covering an area of about 51.83 km^2). The ALS data depict (more) precise height
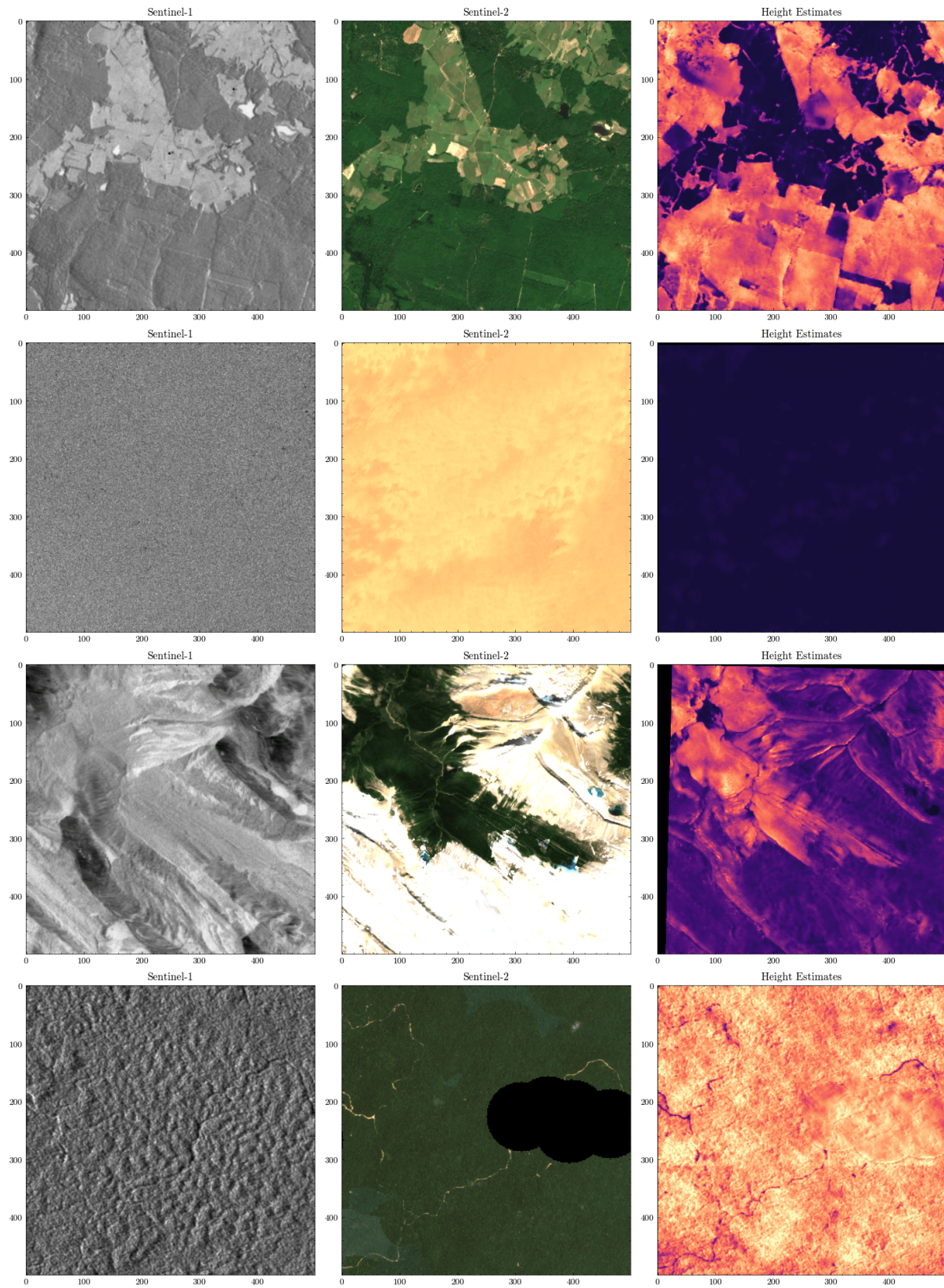
Figure 12. Sentinel-1 and Sentinel-2 images (RGB channels) and our prediction in France (first row), in the Sahara (second row), in Canada (third row), and in the Congo Basin (fourth row; heights from 0 m to 35 m; see Figure 1 for the colormap)
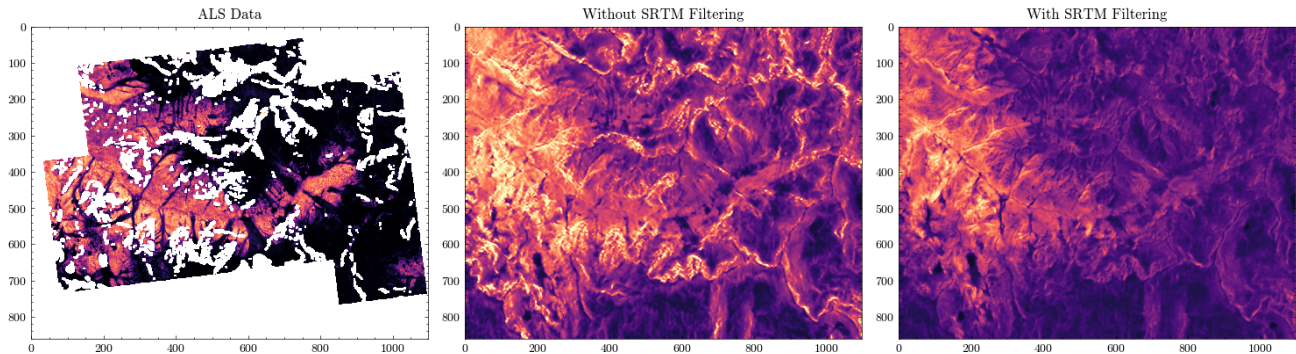
Figure 13. ALS data and both predictions for the ablation area.

Table 4. Ablation study comparing different model architectures showing the shifted Huber loss.

ARCHITECTURE	DEEPLABV3	DEEPLABV3+	FPN	LINKNET	MANET	PAN	PSPNET	UNET	UNET++
PRE-TRAINED	1.22	1.10	1.14	1.06	1.04	1.13	1.13	1.10	1.10
NOT PRE-TRAINED	1.15	1.08	1.08	1.04	1.03	1.07	1.12	1.01	1.04

measurements and can be considered as ground-truth for this ablation study. We train two models using the same data and the same parameters, except for SRTM filtering. The MAE is 9.77 m without SRTM filtering and improves to 7.33 m with SRTM filtering. An illustration of the ALS data as well as both results is provided in Figure 13.

E.3. Architectures

To assess the effectiveness of various deep learning architectures for canopy height prediction, we conducted comparative analyses on several models, including DeepLabV3 (Chen et al., 2017) and V3+ (Chen et al., 2018), FPN (Lin et al., 2017), LinkNet (Chaurasia & Culurciello, 2017), MANet (Fan et al., 2020), PAN (Li et al., 2018), PSPNet (Zhao et al., 2017), and UNet++ (Zhou et al., 2019). These models were evaluated in both configurations: with and without pre-trained backbones based on ImageNet (Deng et al., 2009). The comparative analysis, provided in Table 4, suggests a strong disparity between different model architectures. Specifically for canopy height prediction, the standard U-Net model (Ronneberger et al., 2015), despite its simplicity, outperformed more complex architectures like DeepLabV3 (Chen et al., 2017) or PSPNet (Zhao et al., 2017). This outcome underscores the unique demands of canopy height prediction tasks, which may not align with the strengths of architectures designed for broader or different types of image segmentation tasks. Moreover, the experiment revealed a counter-intuitive finding regarding the use of pre-trained backbones. Contrary to expectations, employing a pre-trained backbone resulted in a decrease in performance across all examined architectures. This decrease was consistent but varied in magnitude among the different models, indicating a potential mismatch between the generalized features learned from ImageNet (Deng et al., 2009) and the specific features of satellite images relevant to canopy height estimation.

E.4. Learning Rate Scheduler

Throughout our experiments, we relied on a linear learning rate scheduler with tuned initial value. The use of cyclical scheduler has previously been found to aid generalization (Smith, 2017; Zimmer et al., 2023). To investi-

Table 5. Learning rate schedulers with and without amplitude decay (AD).

MODE	LINEAR	1 CYCLE	2 CYCLES	3 CYCLES	4 CYCLES
WITHOUT AD	–	1.022	1.012	1.010	1.004
WITH AD	1.014	1.016	1.060	1.026	1.053

gate the impact in our setting, we compare two different cyclical schedulers (with and without overall decay of the cycle amplitude) in Table 5 for different numbers of cycles, using 0.001 as the learning rate bound. These two modes correspond to `triangular` and `triangular2` in PyTorch's `CyclicLR` implementation. We report the shifted Huber loss on the validation set.
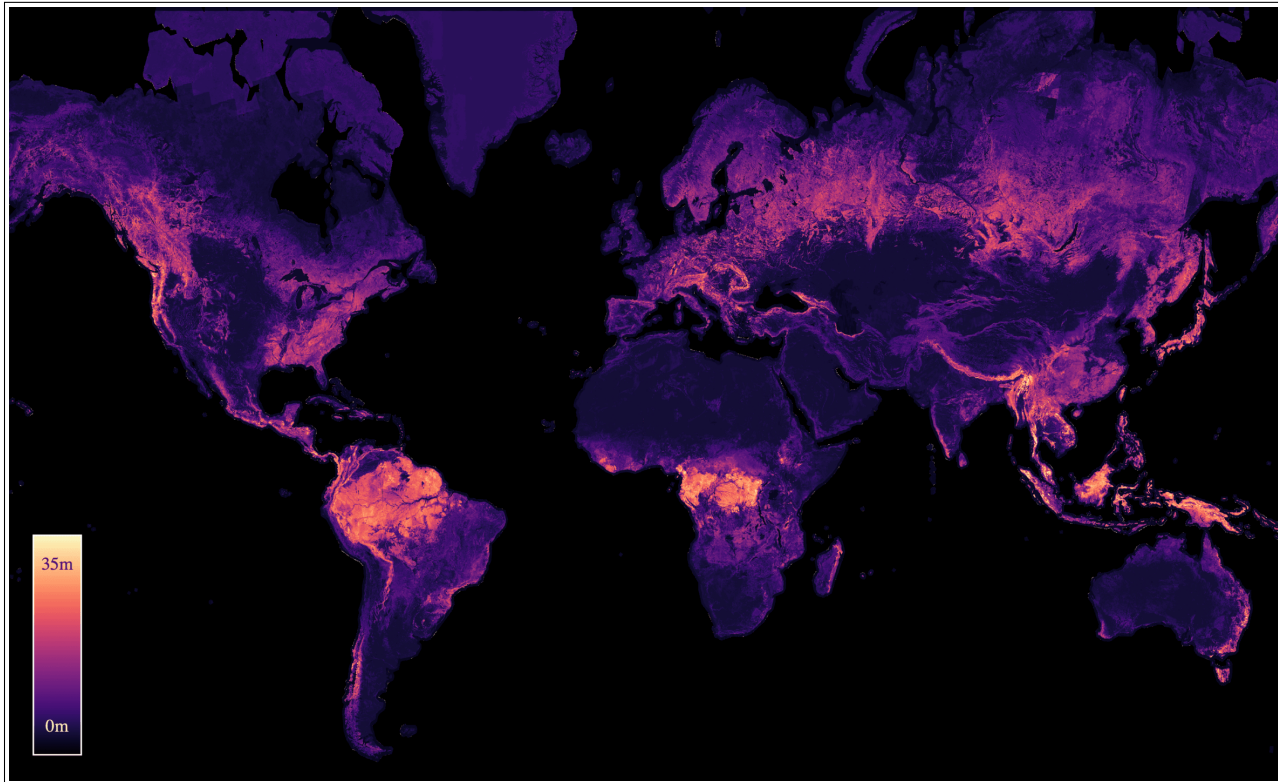
Figure 14. Global canopy height map (enlarged version of Figure 1).

E.5. Ensembles

Previous work has shown that the performance of models can be improved by building ensembles (Fort et al., 2019; Ganaie et al., 2022) or to parameter averages (Izmailov et al., 2018; Zimmer et al., 2024)) from multiple models. While param-

Table 6. Single versus ensemble model.

MODEL	MAE	MSE	RMSE	MAPE	RRMSE
SINGLE	4.03	36.57	6.05	0.29	0.50
ENSEMBLE	3.94	35.07	5.92	0.28	0.49

eter averages do not increase the inference complexity, more effort is required to find models suitable for averaging (hence, parameter averages are not suited for our case). In Table 6, we evaluate the single model against an ensemble of six models. Here, the ensemble aggregates the estimates of multiple models and, hence, increases the performance in all metrics, albeit at the price of having to evaluate multiple models for inference.

E.6. Backbone

In order to validate our backbone choice, we do an ablation study testing different versions of our backbone and compare them with each other regarding the number of parameters and the shifted Huber loss (on the validation set) for both pre-

Table 7. Different backbones.

	RESNET18	RESNET50	RESNET101
PARAMETERS	14 362 705	32 555 601	51 547 729
PRE-TRAINED	1.117	1.086	1.095
NOT PRE-TRAINED	1.051	1.017	1.020

trained and not pre-trained backbones. The results can be seen in Table 7. As already mentioned above, training the backbones from scratch seems to be beneficial in this context, and the ResNet50 backbone seems to be a reasonable choice.

F. Canopy Height Map

The final canopy height map is shown in Figure 14. For an interactive map, we refer the reader to the Google Earth Engine, which allows to inspect any location on the map in more detail.[14]

---

[14]https://worldwidemap.projects.earthengine.app/view/canopy-height-2020