

Article

New Method to Correct Vegetation Bias in a Copernicus Digital Elevation Model to Improve Flow Path Delineation

Gabriel Thomé Brochado ^{1,2,*}  and Camilo Daleles Rennó ² 

¹ Brazilian Army Geographic Service (DSG), Quartel General do Exército, Setor Militar Urbano, Brasília 70630-901, DF, Brazil

² Earth Observation and Geoinformatics Division, National Institute for Space Research (INPE), Avenida dos Astronautas, 1758, Jardim da Granja, São José dos Campos 12227-010, SP, Brazil; camilo.renno@inpe.br

* Correspondence: gabriel.brochado@eb.mil.br

Abstract: Digital elevation models (DEM) are widely used in many hydrologic applications, providing key information about the topography, which is a major driver of water flow in a landscape. Several open access DEMs with near-global coverage are currently available, however, they represent the elevation of the earth's surface including all its elements, such as vegetation cover and buildings. These features introduce a positive elevation bias that can skew the water flow paths, impacting the extraction of hydrological features and the accuracy of hydrodynamic models. Many attempts have been made to reduce the effects of this bias over the years, leading to the generation of improved datasets based on the original global DEMs, such as MERIT DEM and, more recently, FABDEM. However, even after these corrections, the remaining bias still affects flow path delineation in a significant way. Aiming to improve on this aspect, a new vegetation bias correction method is proposed in this work. The method consists of subtracting from the Copernicus DEM elevations their respective forest height but adjusted by correction factors to compensate for the partial penetration of the SAR pulses into the vegetation cover during the Copernicus DEM acquisition process. These factors were calculated by a new approach where the slope around the pixels at the borders of each vegetation patch were analyzed. The forest height was obtained from a global dataset developed for the year 2019. Moreover, to avoid temporal vegetation cover mismatch between the DEM and the forest height dataset, we introduced a process where the latter is automatically adjusted to best match the Copernicus acquisition year. The correction method was applied for regions with different forest cover percentages and topographic characteristics, and the result was compared to the original Copernicus DEM and FABDEM, which was used as a benchmark for vegetation bias correction. The comparison method was hydrology-based, using drainage networks obtained from topographic maps as reference. The new corrected DEM showed significant improvements over both the Copernicus DEM and FABDEM in all tested scenarios. Moreover, a qualitative comparison of these DEMs was also performed through exhaustive visual analysis, corroborating these findings. These results suggest that the use of this new vegetation bias correction method has the potential to improve DEM-based hydrological applications worldwide.

Keywords: Copernicus; digital elevation model; vegetation bias; correction; bare-earth; flow path; delineation



Citation: Brochado, G.T.; Rennó, C.D. New Method to Correct Vegetation Bias in a Copernicus Digital Elevation Model to Improve Flow Path Delineation. *Remote Sens.* **2024**, *16*, 4332. <https://doi.org/10.3390/rs16224332>

Academic Editor: Magaly Koch

Received: 13 September 2024

Revised: 6 November 2024

Accepted: 17 November 2024

Published: 20 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Digital elevation models (DEMs) have been an important source of information for a variety of hydrologic applications, given that topography has a major influence on how the water flows through the landscape [1–3]. The emergence of open access quasi-global DEMs (GDEMs), like Shuttle Radar Topographic Mission (SRTM) [4], facilitated study cases and uses involving larger contiguous areas, and increased the interest of researchers in this data source [5], despite their spatial resolutions (no better than 30 m) and uncertainties being limiting factors for some applications [2,6,7].

After the release of SRTM in 2003, other open-access GDEMs have become available, such as the Spaceborne Thermal Emission and Reflection Radiometer (ASTER) GDEM [8], Advance Land Observing Satellite (ALOS) World 3D (AW3D30) DEM [9], NASADEM [10], TanDEM-X [11], and more recently Copernicus DEM [12]. However, all these DEMs share the drawback of being subject to a positive elevation bias in vegetated areas, because the sensors used in their data acquisition process are unable to fully penetrate the forest cover [13,14]. This bias is especially detrimental to hydrologic applications because the elevated vegetated areas act as dams, blocking the flow paths [15]. This can cause, for example, distortions in hydrological features extracted from the DEM, such as drainage networks [16,17], and hinder the accuracy of hydrodynamic models [18].

Throughout the years researchers have attempted to mitigate this problem by estimating the vegetation bias and removing it from a target DEM. In general terms, the correction methods differ on how the bias estimation is achieved and what input datasets are used. Earlier studies estimated the bias by interpolating elevation differences sampled at the borders of the forested areas [16,19–21] or considered it equal to a percentage of the forest height, obtained from a preexisting dataset [18,22]. Later on, the studies focused on the application of regression-based methods to model the bias as a function of variables such as vegetation cover, density, and height, that were obtained from the datasets available at the time, and used airborne and/or orbital LiDAR measurements as reference, because they can penetrate the tree cover and retrieve ground elevations [23–28]. This approach evolved towards the use of more complex machine learning methods, such as random forest [15,29], gradient tree boost [30], artificial neural networks [31–33] and ensemble methods combining different algorithms of this type [34,35]. Some recent studies did not follow this trend and estimated the vegetation bias by interpolating ICESat-2 forest height data points [36] or adjusting preexisting forest height maps to match with a target DEM [37]. Among these studies, only a few resulted in the distribution of quasi-global corrected DEMs. O’Loughlin et al. [25] developed the first global “bare-earth” DEM correcting the 3 arc second SRTM, Yamazaki et al. [26] combined the 3 arc second SRTM and AW3D30 to produce the Multi-Error-Removed Improved-Terrain DEM (MERIT DEM), Zhao et al. [27] corrected the 1-arc second SRTM resulting in the first worldwide product with this resolution, and Hawker et al. [15] reduced the bias from vegetation and buildings on the Copernicus DEM, resulting in the FABDEM product.

Aside from the baseline DEM, the main input datasets used on all methods mentioned above are airborne LiDAR surveys, vegetation cover products, forest height maps, and orbital LiDAR data, which also received improvements over time along with the correction methods. Products such as Vegetation Continuous Fields (VCF) derived from MODIS [38], Landsat [39] and Global Forest Change data [40] provide information on the vegetation extent and canopy density (as the percentage of tree cover per pixel), and were used by most authors. The potential use of higher resolution land use/land cover datasets to characterize vegetation extent, such as the currently available 10 m ESA World Cover [41] and ESRI Land Cover [42], was not properly explored yet. The global forest height map developed by Simard et al. [43], with spatial resolution of 1 km, was used in the majority of the studies up to 2018, and recently a new dataset, with 30 m resolution, was developed for the year 2019 [44], and used in newer correction methods [15,35,37]. Forest height maps, with higher spatial resolutions, have been released already [45,46], however studies employing them for vegetation bias correction are still lacking. Regarding orbital LiDAR measurements, ICESat-1/GLAS data were used in the studies until 2020 and then replaced by ICESat-2/ATLAS and Global Ecosystem Dynamics Investigation (GEDI) observations. Moreover, the area covered by airborne LiDAR surveys increased throughout the years as they became more accessible, enabling researchers to use reference elevation data from a greater variety of regions.

Regarding the base DEM, most studies apply vegetation bias correction to SRTM. However, recent research indicates that, among the uncorrected GDEMs, Copernicus DEM provides the best representation of the terrain [47–49] and depicts more detailed

topography [14]. Additionally, Moges et al. [50] evaluated the impact of DEM selection on hydrological modeling and pointed out that Copernicus DEM and AW3D30 were the ones that delivered a better performance overall. However, from all studies mentioned above, only two targeted the Copernicus DEM [15,30], hence its use as baseline DEM still needs further exploration. Moreover, both studies employed machine learning algorithms to correct the bias in large areas, and while these non-parametric methods can model complex relationships between the variables better than the parametric ones, they are still highly dependent on the quality and size of the training data, as well as being susceptible to overfitting [51]. As orbital LiDAR measurements are sparse, representing only a small portion of the area they cover, and airborne LiDAR availability is limited, especially in developing countries, compiling enough training data to represent a wide variety of areas, with different topographic characteristics and vegetation types, becomes challenging.

With this obstacle in mind, a new deterministic method to correct the vegetation bias in the Copernicus DEM is proposed in this study, which is independent of training data, requires only two input datasets, no feature engineering, and at the same time is adaptive to regional characteristics. The method showed improvement over Copernicus DEM and FABDEM on the hydrology-based evaluation we performed. In the remainder of this article, we introduce the targeted study areas, describe the input datasets, the correction and evaluation methods (Section 2), present the obtained results (Section 3), discuss the advantages and limitations of the method (Section 4), and convey our concluding remarks (Section 5). After a few enhancements, we expect to be able to gradually apply this method to each continent, generating a new open access quasi-global “bare-earth” DEM.

2. Materials and Methods

This section describes the study areas and input datasets used in this work, as well as the methods used to perform the vegetation bias correction on Copernicus DEM and to compare the obtained results with the original and benchmark DEMs. The main innovative aspects of the correction method are the use of local adjustment factors to compensate for errors in the subtraction between the DEM elevations and the forest height data, and also the automatic adjustment of this dataset to match the DEM acquisition year in a target area. The focus of the comparison method was on hydrology, assessing the proximity of the stream flow paths delineated from these DEMs to a reference dataset. All processes described in this study were performed by custom scripts (python version 3.12), and all input and output datasets are available in the Supplementary Materials.

2.1. Study Areas

Since the goal was to develop a correction method that could be applied worldwide, we had to evaluate its performance in different conditions of forest cover and topographical relief. To achieve this, one study area was selected to represent each of the following scenarios: (1) high forest cover and flatter terrain, (2) low forest cover and flatter terrain, (3) high forest cover and rougher terrain, and (4) low forest cover and rougher terrain. These study areas were selected within Brazilian territory, due to the availability of an extensive high-quality drainage network reference dataset, which enabled the selection of study areas that better represented the described scenarios. As shown in Figure 1, areas 1 and 2 are located in the north region of the country, 3 in the southeast region, and 4 in the south.

The differences in forest cover and overall relief between the study areas can be observed in Figure 2, which presents their natural color Sentinel-2 mean composite for the year 2020, and a color representation of the 1 arc second Copernicus DEM elevations. Moreover, the minimum, maximum, and average elevation, forest cover percentage, and average forest height of each study area are presented in Table 1. The forest cover percentage and the average forest height were calculated using the canopy height map produced by Potapov et al. [44], which is detailed in Section 2.2.2.

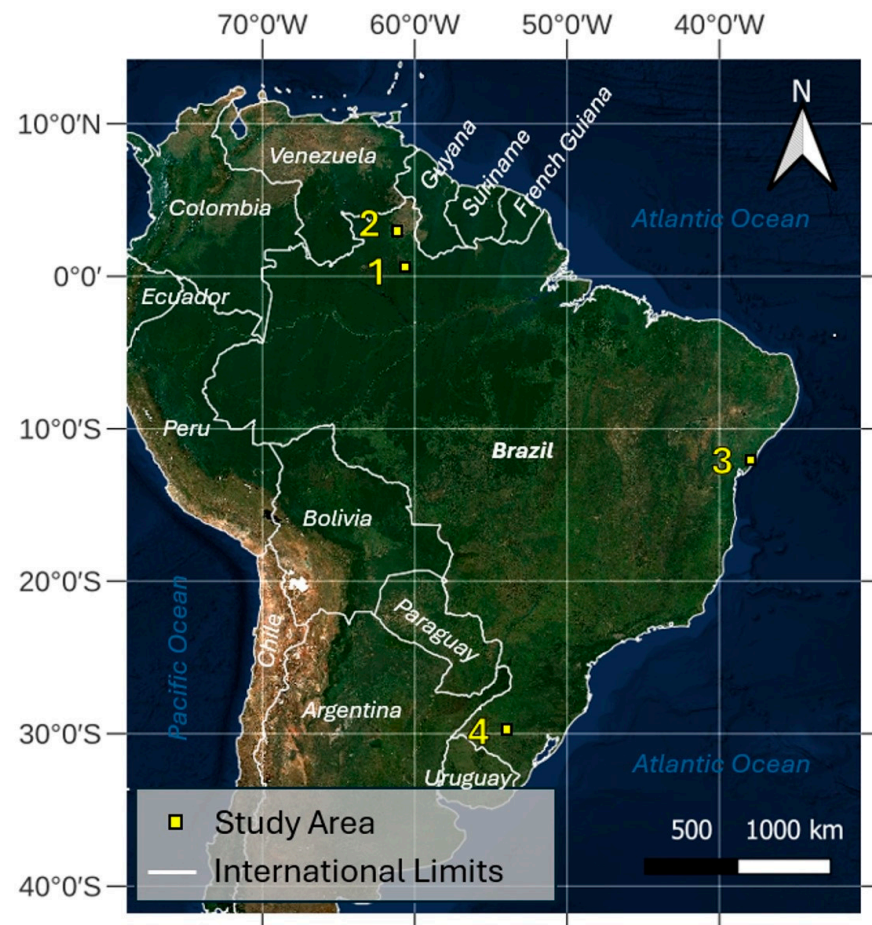


Figure 1. Position of study areas overlaid to a natural color Sentinel-2 cloud free composite of the year 2020 of South America.

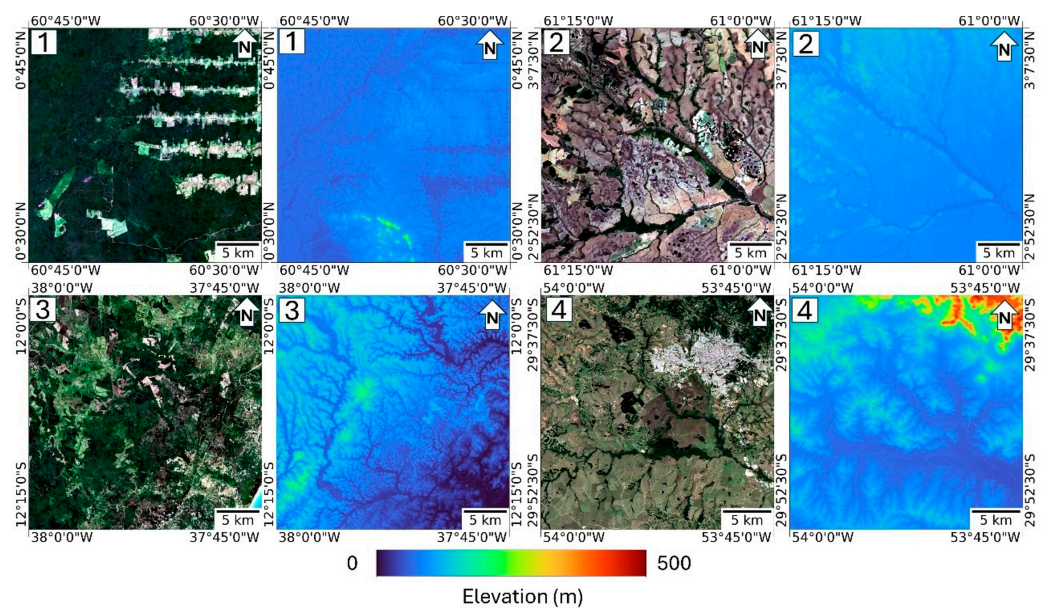


Figure 2. Sentinel-2 cloud free composite of the year 2020 and color representation of Copernicus DEM elevations of the study areas. The numbers in the top left corner of each panel refer to the study area depicted on it.

Table 1. Study areas characteristics.

| Study Area | Min Elevation (m) | Max Elevation (m) | Avg. Elevation (m) | Forest Cover (%) | Avg. Forest Height (m) |
|------------|-------------------|-------------------|--------------------|------------------|------------------------|
| 1 | 43.28 | 296.58 | 76.95 | 91.08 | 23.98 |
| 2 | 67.80 | 141.34 | 96.06 | 16.37 | 13.78 |
| 3 | 0 | 222.51 | 79.68 | 70.93 | 13.08 |
| 4 | 43.50 | 477.39 | 111.94 | 18.54 | 11.06 |

2.2. Input Datasets

The datasets used in the vegetation bias correction method were Copernicus DEM [12], the forest height map elaborated by Potapov et al. [44], and forest loss per year from Global Forest Change [40]. Drainage networks obtained from the Brazilian official topographic maps and FABDEM [15] were used during product evaluation.

2.2.1. Digital Elevation Model

The 1 arc second Copernicus DEM (GLO-30) is a freely available global product derived from data acquired during the TanDEM-X mission, from December 2010 to January 2015. This mission was a partnership between German Aerospace Centre (DLR) and Airbus, which employed two twin X-band SAR sensors (TerraSAR-X and TanDEM-X), flying in tandem orbit, to enable the generation of global DEMs by the means of SAR interferometry. The first product of this mission was the commercial DSM WorldDEM™, with the spatial resolution of 0.4 arc second (~12 m). This dataset was edited to correct coast/shorelines, special terrain features (e.g., airports), and implausible terrain formations, and also to include consistent river flows and flatten the water bodies [12]. Posteriorly, the WorldDEM™ was resampled to generate the Copernicus DEM, with the 1 arc second version being made publicly available in 2020. This product may benefit from the editing processes performed on its predecessor; however, the authors stated that the hydrological consistency cannot be guaranteed after the resampling process.

The validation of the Copernicus DEM was performed by the data producers using ICESat/GLAS reference points and achieved an absolute vertical accuracy lower than 4 m. Independent studies found similar values [47,49] and, as mentioned before, pointed out that, among the freely available 1 arc second global DSMs, Copernicus DEM provides the best representation of the terrain. Moreover, it contains updated data when compared to the other free global DEMs, due to its more recent acquisition period, and it uses the EGM 2008 geoid model [52], which was shown to be superior to the EGM1996 [53] used by the other previously mentioned DEMs [54]. Besides the elevation data, the Copernicus DEM has 6 auxiliary quality layers, including a water bodies mask that is also used in the method proposed in this study.

2.2.2. Forest Height Data

The 2019 global forest canopy height map, with 30 m spatial resolution, developed by Potapov et al. [44], was generated by extrapolating LiDAR forest height measurements from the Global Ecosystem Dynamics Investigation (GEDI) instrument, using multitemporal metrics derived from analysis-ready Landsat data [55] and a regression tree model. The coverage is limited by the availability of GEDI measurements, which were collected between 51.6°N and 51.6°S.

The product was validated using independent GEDI observations (RMSE = 6.6 m, MAE = 4.45 m) and airborne LiDAR data (RMSE = 9.07 m, MAE = 6.36 m), separately. The validation results indicated underestimation of the forest height for both test datasets, especially for shorter (<7 m) and taller forests (>30 m). However, Hawker et al. [15] also reported overestimation in areas of complex topography, caused by issues in the original GEDI data. It is also important to mention that this dataset only maps vegetation with heights of 3 m or above.

Despite these limitations, the boundaries of forest mask derived from these data (forest height > 0) showed a better agreement with the sharp increases in elevation at the forest/non-forest transitions present in Copernicus DEM, when compared to a more recent forest height map, developed by Lang et al. [45]. This dataset was derived from Sentinel-2 images of the year 2020 and GEDI data using a deep learning approach. The above-mentioned match between the forest mask and the DEM is critical for the proposed correction method, as detailed in Section 2.3, hence the dataset elaborated by Potapov et al. [44] was selected. Another promising option was the 1 m global canopy height developed by Tolan et al. [46]. This product was derived from the combination of high-resolution optical imagery (<1 m), acquired between the years 2017 and 2020, and LiDAR forest height measurements from both airborne instruments and GEDI. However, this product was not used because data gaps were identified throughout all study areas. Figure 3 illustrates the differences between these three forest height datasets in a region within Area 1, where it can be observed that our selection has better agreement with the sudden elevation changes present in the DEM.

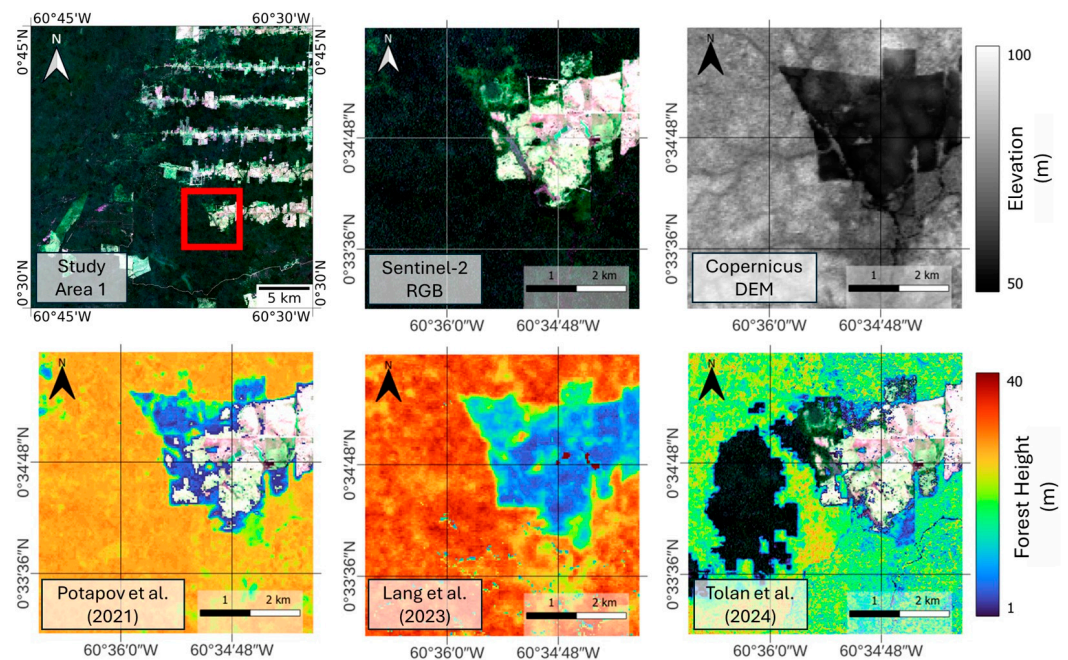


Figure 3. Comparison between the forest height datasets. The figure presents a natural color Sentinel-2 cloud free composite of the year 2020 of the entire Area 1, with the subset area marked by the red rectangle (**top left**); the Sentinel-2 image of the subset area (**top center**); a grayscale representation of Copernicus DEM elevations on the subset area (**top right**); the Sentinel-2 composite overlaid by a color representation of Potapov et al. [44] (**bottom left**), Lang et al. [45] (**bottom center**) and Tolan et al. [46] (**bottom right**) forest height datasets, where heights equal to zero are transparent.

2.2.3. Global Forest Change Data

The Global Forest Change (GFC) dataset is the product originated from the study performed by Hansen et al. [40], which used metrics derived from Landsat time-series and a decision tree model to characterize the forest extent, loss, and gain globally. The authors considered a forest as having only trees with a height of at least 5 m. The data originally covered the period of 2000 to 2012 but has received annual updates since its release, currently covering up to the year 2023. The main layers of this dataset are the global tree cover for the year 2000 (encoded as a percentage of the canopy cover on the pixel), the forest loss per year (from 2001 to 2023), and the forest gain between 2000 and 2012. In the present study, we used only the forest loss per year (lossyear). These data are encoded with pixel values ranging from 1 to 23, referring to the year the forest loss was observed in the region covered by each pixel (e.g., 5 = 2005), and 0 where no forest loss

was detected. This information was used to better adjust the forest height dataset to the Copernicus DEM, as further explained in Section 2.3. The forest gain information was not used for the referred adjustment because it is not discriminated per year and it does not cover the entire data acquisition period of the Copernicus DEM (2010–2015).

2.2.4. Reference Drainage Networks

The reference drainage networks were extracted from the official topographic maps produced by the Brazilian Army Geographic Service (DSG). We selected the most accurate and recent maps available covering the study areas. The cartographic scale and acquisition dates of the topographic maps used for each area are shown in Table 2. The drainage networks on these maps were manually acquired by photogrammetric restitution and vectorization processes, using aerial or orbital images with sub-metric spatial resolutions. In Area 1, instead of photogrammetric restitution, a 5 m spatial resolution DTM, obtained by a p-band SAR interferometry system, capable to penetrate the forest and retrieve ground elevations [56], was used to perform a semi-automatic acquisition process [57]. The drainage networks were in all shapefile format and, according to their official data acquisition specification [58], their lines were traced following the water flow direction (up to downstream) and segmented at every intersection between lines. This feature facilitated the implementation of the comparison method presented in Section 2.4.

Table 2. Scale and acquisition year of the Brazilian official topographic maps used per study area.

| Study Area | Cartographic Scale | Acquisition Year |
|------------|--------------------|------------------|
| 1 | 1:50,000 | 2010 |
| 2 | 1:50,000 | 2021 |
| 3 | 1:25,000 | 2019 |
| 4 | 1:25,000 | 2016 |

2.2.5. FABDEM

The Forest and Building removed DEM (FABDEM) was developed to correct the building and vegetation bias in the Copernicus DEM. The dataset has 1 arc second (~30 m) of spatial resolution and it is available between 60°S and 80°N. According to the authors, the method reduced the mean absolute vertical error by 30% in built-up areas and 44% in vegetated ones [15]. Moreover, independent validations of this dataset found similar results [49,54,59], pointing out that this product is the best free global DEM currently available.

To perform the correction, separate random forest regressor models were trained for the vegetated and built-in areas. The predictor variables used for the forest model were extracted from the 2019 global forest canopy height map [44]; ICESat-2 land and vegetation height data [60], for the areas not covered by the forest height map; and Copernicus Global Land Service canopy cover percentage [61]. For the built-in areas model, several predictor datasets were used, such as night-time lights [62] and urban building footprints [63]. The reference elevation data used to train the model were a set of airborne LiDAR DEMs, selected from 12 different countries to increase the robustness of the product. The latest version of this dataset (FABDEM V1-2), released on January 2023 [64], was used in the current work.

2.3. Vegetation Bias Correction Method

The current method was developed while attempting to correct the canopy effect of the Copernicus DEM by subtracting from its elevations the estimated forest height from the dataset developed by Potapov et al. [44]. It was expected that the sudden increases in elevation, observed on the borders of the forest patches, caused by the presence of vegetation bias, would be reduced, resulting in smoother transitions. However, after the correction, the majority of the vegetated areas became depressions and the sharp increases in elevation on vegetation borders were replaced by abrupt descents, making them easily distinguishable from their surroundings. This was an indicative that, in these areas, the

forest height was overestimated and/or the canopy elevation was underestimated in the DEM. Overestimation of the forest height can be caused by the inaccuracies of the source dataset mentioned in Section 2.2.2, and the underestimation of canopy elevations can be caused by the penetration of the SAR pulses into the vegetation during the data acquisition step of DEM generation, even when shorter wavelengths, such as the X-band, are used in the process [65]. Figure 4 depicts how both of these effects can affect the estimated ground elevation on the corrected DEM, where Δh_1 is the forest height overestimation and Δh_2 is the canopy elevation underestimation.

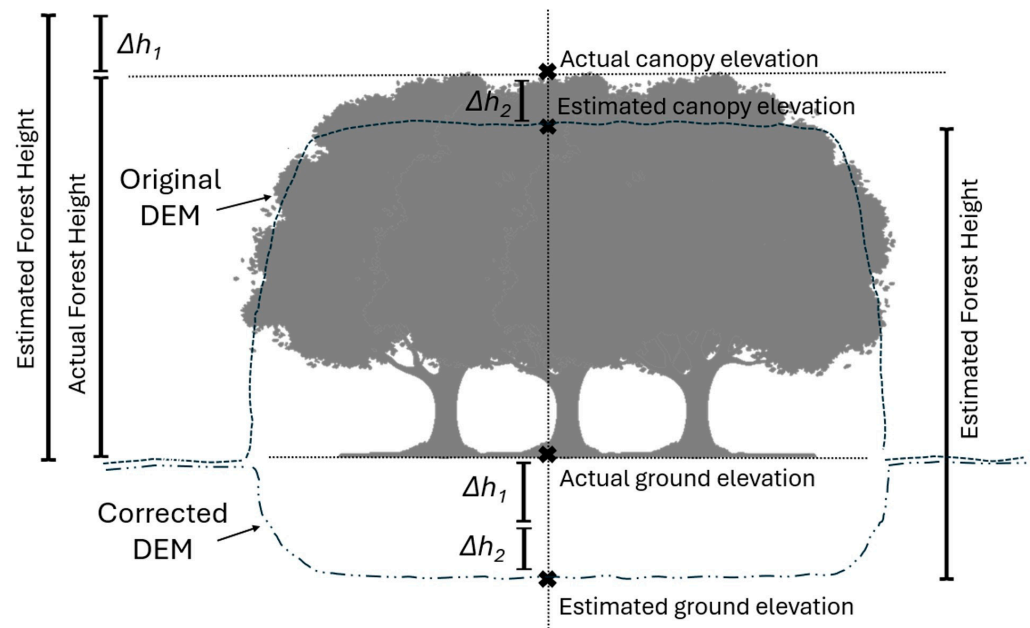


Figure 4. Effect of forest height overestimation and canopy elevation underestimation on the estimated ground elevation. The illustration represents the difference (Δh_1) between the estimated and actual forest heights, the original and corrected DEMs elevation profiles (dotted lines), the differences between the actual and estimated canopy elevations (Δh_2), and ground elevations ($\Delta h_1 + \Delta h_2$).

To improve the correction process, the influence of the above-mentioned effects was mitigated by the application of an adjustment factor (k) to reduce the estimated forest height, making the resulting estimated ground elevations better match the actual ones. However, this factor needed to be tuned locally, and the exact k could only be calculated where the actual ground elevation were known, so it was necessary to estimate it. The strategy devised to do this consisted of finding the values of k that minimizes the elevation spikes around the pixels located at the borders of the forest patches. The local slope was used to quantify this. The optimal k values found at the border pixels of each forest patch were averaged, and the resulting adjustment factor was used for all pixels in that patch. The corrected DEM was obtained by applying this process to all forest patches of a given region and subtracting from the original DEM elevations the forest height adjusted by k value of each patch.

However, since the data acquisition periods of the forest height dataset (2019) and the Copernicus DEM (2010 to 2015) are different, mismatches in the forest cover can occur, mainly due to forest loss, causing errors in the correction process. So, the forest cover loss information from Global Forest Change was used to adjust the forest height to the Copernicus DEM acquisition period. One adjusted forest height dataset was created for each possible acquisition year, and its respective corrected DEMs were obtained. Among these DEMs, the one that resulted in overall smoother transitions between forest and non-forest areas was selected. This was measured by the average slope of the DEMs, since higher slope values appear wherever the transition was not properly corrected. Posteriorly, the final product was obtained by performing post-processing steps, such as adaptive

smoothing, to improve this selected DEM. Figure 5 presents the general workflow of the entire process.

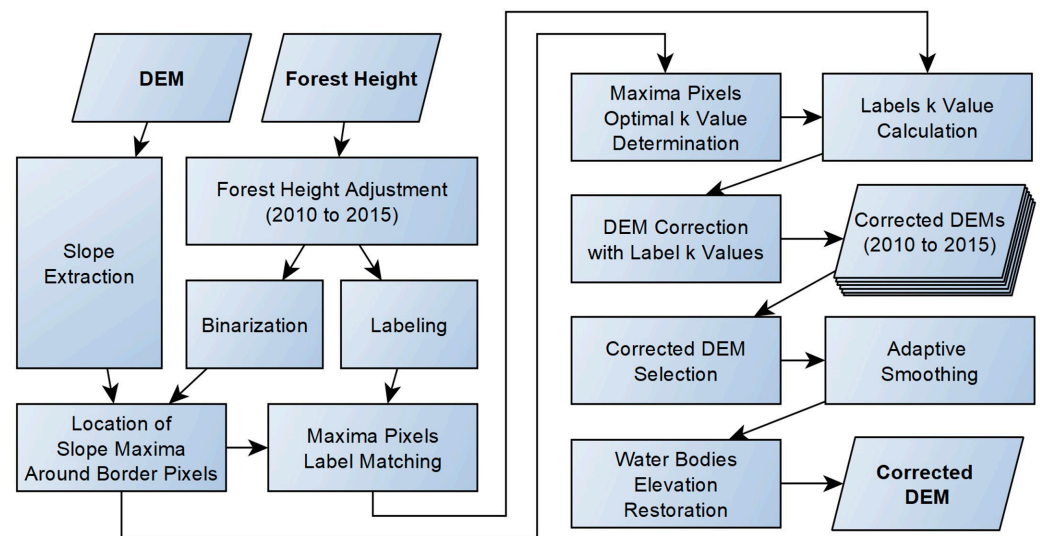


Figure 5. Copernicus DEM vegetation bias correction workflow.

The details of each step of the vegetation bias correction method presented in Figure 5 are described in the remainder of this section.

Forest Height Adjustment: The forest height dataset was adjusted to each year from 2010 to 2015 as mentioned above. This was performed by adding to the forest height dataset the forested areas that were lost between the target year and 2020. The extent of these added areas was obtained from the lossyear layer of the GFC dataset ($\text{lossyear} \geq \text{target year}$), and their canopy height was estimated pixel by pixel using the average of the 128 nearest pixels of the original forest height dataset. For this estimation, only pixels where the lossyear was zero and the forest height was greater than zero were considered. The original forest height was substituted for the new one in the pixels where the added areas overlapped the forest height dataset. The reference years the forest height dataset was automatically adjusted and were 2012, 2013, 2010, and 2013 for study areas 1 to 4, respectively.

Location of Slope Maxima Around Border Pixels: The vegetation height dataset was converted to a binary forest mask ($\text{forest height} > 0$), and, for each pixel at the borders of the forest patches, the position of the local slope maxima was found, considering a 3×3 neighborhood. This approach was used because eventual mismatches between the forest mask and the vegetation in the DEM may occur, and, in this case, the border pixel itself may not represent the elevation spike (slope maxima), caused by the vegetation bias, correspondent to the transition between forest and non-forest areas.

Maxima Pixels Optimal k Value Determination: For each slope maxima pixel, the k value that results in the minimum average slope, for a 3×3 neighborhood around it, was identified. To do this, a set of corrected DEMs was created using global adjustment factors, i.e., the same k value was applied to all forest patches. The selected k values varied from 0 to 1 in steps of 0.05. Then, the slopes of these DEMs were extracted, the average slope within the neighborhood of each maxima pixel was calculated, and the k value correspondent to the minimum result was selected. If there was more than one k value corresponding to the minimum result, the lowest one was selected. Prior to this step, the forest height was smoothed using a 5×5 mean filter, to reduce errors caused by minor mismatches between the forest mask and the vegetation represented in the DEM. If an adjustment factor equal to zero was found, that slope maxima pixel was disregarded, because this suggests that its position does not correspond to a transition between forest and non-forest areas in the original DEM. Additionally, if a maxima pixel was located within or touched any object of the DEM water bodies mask, the pixel was also disregarded, because the elevations of such

water bodies are artificially set to values lower than the surrounding areas, and this could introduce errors in the correction process.

Labeling: Each vegetation patch of the forest height dataset received a unique label. A dilation process was applied to this labeled mask, in order to match its extents with the smoothed forest height used to generate the DEMs corrected with the global adjustment factors, while maintaining the original labels. If the labeling was applied directly to the binarized smoothed forest height, areas that were separated in the original mask could be merged together because of the dilation effect of the smooth filter, receiving the same adjustment factor as a result. This was potentially detrimental to the correction process, because averaging the k value over larger areas could fail to address the individual variability and behavior of the forest patches present in the original mask.

Maxima Pixels Label Matching: A matching process was performed to identify which slope maxima pixels should be considered in the calculation of the k value of each forest patch. Each maxima pixel received the label of the nearest forest patch of the mask that resulted from the labeling process mentioned above. If a specific mask object had zero correspondent maxima pixels, the label of each pixel of the object was substituted with the label of the nearest forest patch that had correspondent maxima pixels.

Label k Value Calculation: The adjustment factor of each forest patch of the labeled mask, resultant from the last step, was calculated by averaging the k values of all maxima slope pixels that received its label. And the combination of the resulting averaged adjustment factors of all forest patches originated a k value map.

DEM Correction with Label k Values: The k value map was multiplied by the smoothed forest height and then subtracted from the original DEM elevations, resulting in the raw corrected DEM, which was then refined as described hereon.

Adaptive Smoothing: Although the last step resulted in a DEM with smoother transitions between forest and non-forest areas, the regions where the correction was performed still presented a texture that was visually rougher than the unmodified ones. Therefore, an edge-preserving smooth filter was applied to the modified areas to further smooth these transitions and achieve a better blending while preserving the local topographic features.

The bilateral filter [66] was selected for this purpose. This filter replaces the intensity of each pixel with a nonlinear combination of the intensities of the nearby pixels. However, this combination is based on both the spatial distances and the intensity differences (range) between the central and neighboring pixels. Higher weights are assigned to closer and more similar pixels, following gaussian distributions for both the distance and range components. This allows sharp edges to be preserved, because pixels with large intensity differences will have relatively small contributions to each other [67]. Due to this property, this filter was also used as a post-processing step in the methodology used to generate the FABDEM [15], that was used in our study as the benchmark for vegetation bias correction.

The smoothing was performed using the Whitebox Tools [68] python package (version 2.3.5) implementation of the bilateral filter, where the spatial distance parameter (σ_{dist}) controls the extent of the neighborhood considered in the process, while the range parameter (σ_{int}) controls how much the intensity differences influences it. These parameters were manually tuned to preserve the sharp edges throughout all study areas, resulting in $\sigma_{\text{dist}} = 3$ and $\sigma_{\text{int}} = 5$.

Water Bodies Elevation Restoration: As a result of all steps performed up to this point, the elevations of some pixels belonging to the water bodies present in the original DEM were modified. In these cases, the original water body elevations were restored, and the final corrected DEM was obtained.

2.4. Stream Flow Paths Delineation Comparison

This section describes how the original Copernicus DEM, our product, and FABDEM were compared quantitatively regarding how close stream flow paths derived from them were to those obtained from the reference drainage network. Figure 6 presents the general workflow of this method.

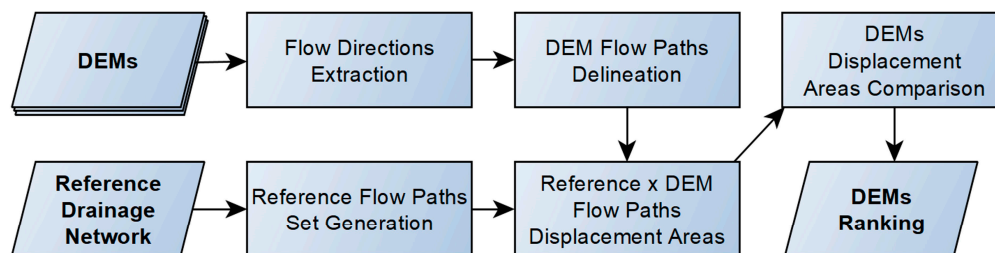


Figure 6. Stream flow paths comparison workflow.

The comparison among DEMs was specifically designed to not depend on the selection of arbitrary accumulation area thresholds to extract the flow paths from the DEMs, because that would limit the scope of the analysis. Moreover, it is important to mention that, prior to all steps described here, some inconsistencies found in the reference drainage networks were manually corrected. All comparison method steps presented on Figure 6 are detailed in the remainder of this section.

Reference Flow Paths Set Generation: In this step, we generated a set of non-overlapping reference flow paths by repeating the process of selecting one random vertex from the vector lines in the topographic map drainage network and tracing its flow path, following the downstream direction, until the distance between the start and end points was equal to a predetermined radius. New flow paths that intersected with previous ones were discarded. This process was repeated until we failed 500 consecutive times to add a new flow path to the set. As a result, a set of flow paths evenly distributed throughout the drainage network is generated. We used three different radii (1000, 2000, and 3000 m), which were selected in order to yield a set size greater than 50 flow paths each. Larger radii, such as 4000 m, were not able to fulfill this prerequisite, since the set size decreases as the radius increases.

Flow Directions Extraction: Parallel to the previous step, the DEMs were hydrologically conditioned to resolve the flat areas and remove the pits. The goal of this process is to obtain a DEM that, on every pixel, there is at least one neighbor with a lower elevation, allowing the flow directions to be extracted properly. To achieve this, first the flat areas were carved in a “v” shape, the pits that can be filled without creating other new pits were filled, and the remaining pits were removed by using the Priority First Search (PFS) algorithm [69] to carve a path to an outlet cell. Finally, the flow directions were extracted from the conditioned DEM using the D8 method [70]. All these processes were performed following the methods presented by Jardim [71] and are implemented in the software Terrahidro [17] version 5.2.0. It is important to emphasize that these hydrologically conditioned DEMs were exclusively used for the purpose of extracting the flow directions.

DEM Flow Paths Delineation: For each initial point of the reference flow paths we obtained, the flow path of each DEM was delineated following its own D8 flow directions, until the distance between its starting and ending points was equal to the same radius that was used to determine the respective reference flow path.

Displacement Areas Calculation: Figure 7 illustrates the process used to calculate the displacement area of each flow path. The displacement area is defined here as the sum of the regions formed between each DEM flow path and its respective reference flow path. The figure shows the drainage network overlaid by an initial point from where the reference and DEM-extracted flow paths are traced, until they reach the circle with radius, r , forming the displacement area between them. In this step, the displacement area of all flow paths derived from each DEM are calculated, resulting in three separated sets of this metric.

Displacement Areas Comparison: Since the lower the displacement areas, the closer the flow path is to the reference, this metric can be used to evaluate the performance of candidate DEMs regarding flow path delineation. This was achieved using the Wilcoxon signed-rank test [72]. This non-parametric test evaluates whether the differences between paired observations of two variables can have a median equal to zero. In the context of the present work, the test was applied to displacement area observations of two different

DEMs, paired by their respective reference flow path. When the test was significant, it indicated that one of the DEMs produced displacement areas that were significantly larger than the other, making it less suitable for flow path delineation.

The observations we used in the tests were obtained from the set of displacement areas of the DEMs. Eventually, due to problems in the original DEM, none of the products could be adequate when compared to the reference. In this case, differences in the displacement areas would be irrelevant to indicate the best product. Therefore, it was decided to select a subset of 50 reference flow paths in order to prioritize those that indicated that at least one of the evaluated products would be better than the others. For this selection, the flow paths were ordered according to their best performance (lowest displacement area among the 3 evaluated DEMs). Since it was expected that the differences between the products would occur over the vegetation areas, it was decided to stratify the selection of the 50 flow paths in predominantly vegetated and non-vegetated ones, according to their location relative to the adjusted forest mask, using the proportion observed in the flow paths set. Thereby, the flow paths with the best performance were selected for each stratum. Figure 8 illustrates this process by showing the reference drainage network for Area 1, the set of flow paths extracted from it using a 2000 m radius, and the flow paths selected from the latter.

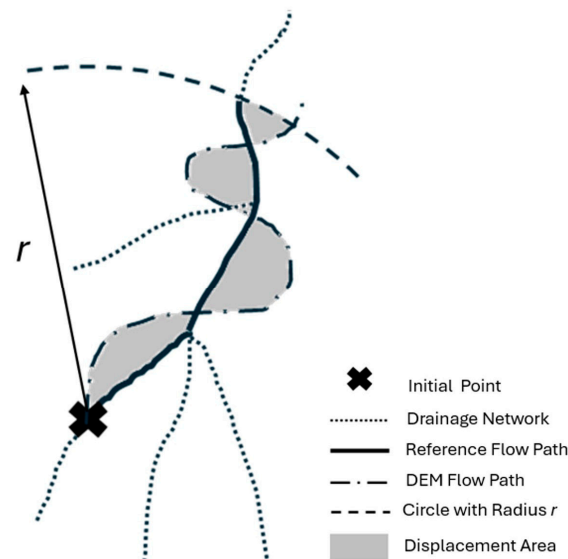


Figure 7. Flow path displacement area calculation. The illustration shows the drainage network overlaid by an initial point from where the reference and DEM-extracted flow paths are traced, until they reach the circle with radius r centered around the point. The flow path displacement area, highlighted in gray, is the sum of the areas located between these lines.

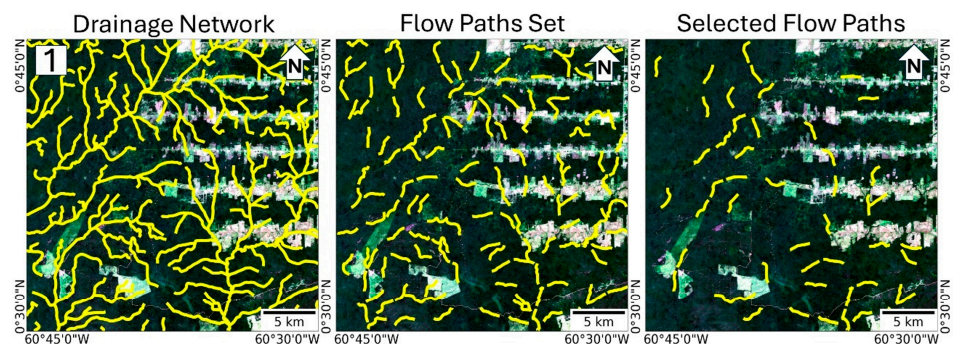


Figure 8. Example of flow paths selection. The panels present the reference drainage network for Area 1 (left), the set of flow paths extracted from it using a 2000 m radius (center), and the flow paths selected from the latter (right). The lines are represented in yellow color in all panels, with the Sentinel-2 composite of the year 2020 in the background.

3. Results

The correction and comparison methods described in the sections above were applied to all study areas, and the p -values obtained by performing the Wilcoxon signed-rank test to the DEM pairs were compiled in Table 3.

Table 3. Resultant p -values obtained for each study area and radius when performing the Wilcoxon signed-rank test featuring the new corrected DEM against Copernicus DEM and FABDEM and also FABDEM against Copernicus DEM. Significant differences at 5% are highlighted in gray. The letter indicates which product was significantly better in each test: C—original Copernicus DEM, F—FABDEM, and N—New corrected DEM.

| Area | Radius (m) | New x Copernicus | New x FABDEM | FABDEM x Copernicus |
|------|------------|------------------|--------------|---------------------|
| 1 | 1000 | 0.3115 | 0.0036 (N) | 0.1433 |
| | 2000 | 0.0129 (N) | 0.0186 (N) | 0.3438 |
| | 3000 | 0.0186 (N) | 0.0048 (N) | 0.4962 |
| 2 | 1000 | 0.0029 (N) | 0.0417 (N) | 0.0013 (F) |
| | 2000 | 0.0004 (N) | 0.4695 | $\cong 0$ (F) |
| | 3000 | 0.0004 (N) | 0.3474 | $\cong 0$ (F) |
| 3 | 1000 | 0.2408 | 0.0103 (N) | 0.0076 (C) |
| | 2000 | 0.1327 | 0.0009 (N) | 0.0001 (C) |
| | 3000 | 0.1306 | 0.2543 | 0.4429 |
| 4 | 1000 | 0.2358 | 0.3976 | 0.3039 |
| | 2000 | 0.0016 (N) | 0.0703 | 0.0231 (F) |
| | 3000 | 0.0001 (N) | 0.0032 (N) | 0.0181 (F) |

The results point out that our product was significantly better than Copernicus DEM and FABDEM in the majority of the tests we performed. On the other hand, the comparison between Copernicus DEM and FABDEM showed mixed results, with the latter not being significantly superior in most cases, including two where it was outperformed by Copernicus (Area 3, $r = 1000$ and 2000 m).

Analyzing the results per study area, we found that in Area 1 our product was superior to Copernicus DEM in all but one of the tested radii ($r = 1000$ m), and outperformed FABDEM in all of them, whereas these two DEMs were tied in all cases. In Area 2, our method produced similar results to FABDEM, both showing improvement over Copernicus DEM in all radii, but in the direct comparison our product was better when the radius was 1000 m, presenting a p -value slightly below the significance level (0.0417). In Area 3, Copernicus DEM and our product yielded comparable results, being tied for all radii, and also being superior to FABDEM when the radius was 1000 or 2000 m. In Area 4, FABDEM and the new corrected DEM had similar results, both being better than Copernicus DEM in all radii except 1000 m, but our product outperformed FABDEM for the 3000 m radius.

These results suggest that, in areas with lower vegetation cover (such as 2 and 4), aside from the slight advantage our product presented over FABDEM, both correction methods have comparable results in improving the original DEM, regardless of the local topography. This outcome can be explained because the limited vegetation present in these areas tends to be mostly located along water courses, which were the main focus of the comparison we performed. Consequently, the benefits of the correction methods in enhancing the Copernicus DEM were emphasized. However, due to the limited extent of the vegetation cover, differences between the correction methods were not as pronounced.

In areas where the vegetation cover is high (1 and 3), the overall relief has a greater influence, since the new correction method only showed improvement over Copernicus DEM in flatter terrain (Area 1). This can be explained because in rougher terrain, where elevation gradients are much greater than the average forest height, they tend to dominate the variations in Copernicus DEM elevation, and the influence of the vegetation cover

becomes less important. So, the application of the correction methods in these areas yields little to no improvement or can actually introduce errors that will lead to lower performances, as we observed in the comparison between FABDEM and Copernicus DEM in Area 3.

In order to better comprehend the results of the quantitative comparison, a qualitative analysis was also carried on. We performed visual interpretation of the DEMs, manual inspection of elevation profiles, and analysis of drainage networks. Among the inspected profiles, four were selected to represent the results obtained in each study area, and the drainage networks derived from the DEMs were compared to the reference one. As mentioned before, the comparison based on drainage networks generated with arbitrary accumulation area thresholds is of limited scope; however, as a complementary qualitative analysis tool, it can still provide further insight about the DEMs suitability for hydrological applications.

Figure 9 presents Copernicus DEM, FABDEM, and the new corrected DEM elevations in Area 1, accompanied with the selected profile lines overlaying a Sentinel-2 cloud free composition for the year 2020 of the same region, and charts containing the observed elevations along them. The background of the charts are highlighted in a gray color where vegetation is present, according to the adjusted forest height data obtained for each study area.

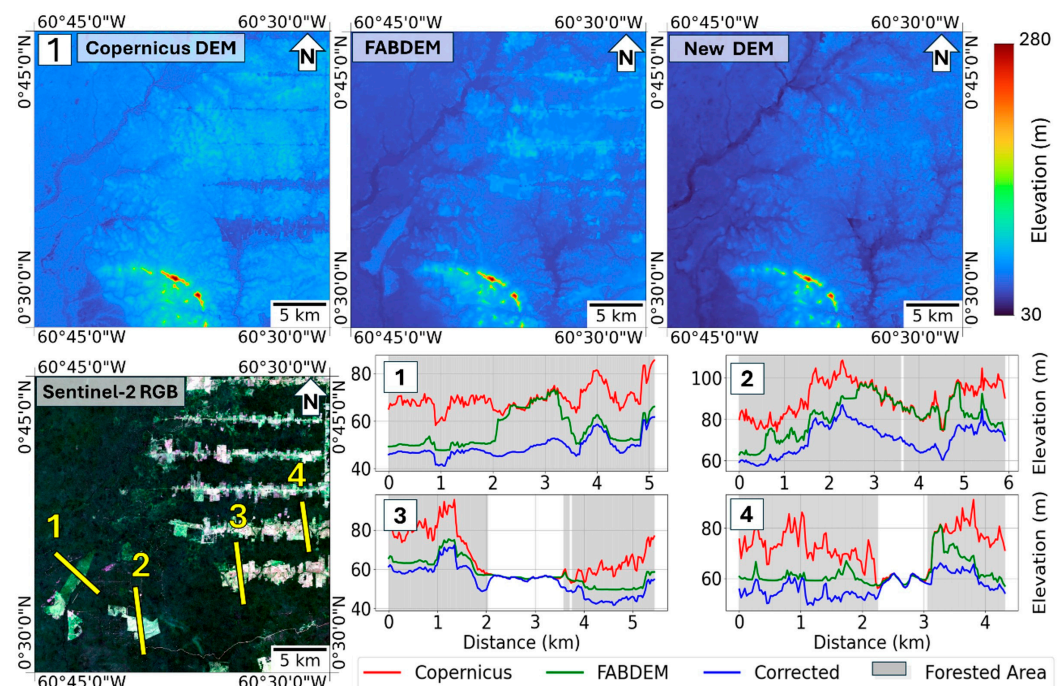


Figure 9. Comparison between DEMs and vertical profiles in Area 1. The figure presents color representations of Copernicus, FABDEM, and the new corrected DEM elevation data over the study area (**top**); its natural color Sentinel-2 cloud-free composite of the year 2020, overlaid by the elevation profile lines identified by their respective numbers (**bottom left**); and charts showing the observed DEM elevations along the profile lines, with the background colored gray in areas covered by vegetation, according to the adjusted forest height obtained for the area (**bottom right**).

It can be observed that the apparent abrupt depressions visible in the Copernicus DEM, caused by the presence of deforestation in the east side of the study area, became less distinguishable in our product, due to the reduction in the elevation of the surrounding vegetated areas. In FABDEM, however, the correction process introduced elevation artifacts in the areas that were deforested after the acquisition period of the Copernicus DEM (2010–2015). Since the forest height dataset used by the authors was acquired in 2019 [15], in the areas where forest loss took place between the acquisition periods, no vegetation

bias reduction was performed, and as a result, they appear as regions of higher elevation, because the elevations of the surrounding, not deforested, areas were reduced. This effect certainly contributed to FABDEM not being significantly better than Copernicus DEM in this area in the quantitative analyses we performed. Our method solved this issue by automatically adjusting the forest height dataset to a reference year that best fitted the original DEM, as exposed in Section 2.3. Analyzing the elevation profiles, we can identify in excerpts 1, 2, and 4 areas that are vegetated in the DEM, according to the adjusted forest height data (year 2012), but were not corrected by FABDEM, due to the above-mentioned acquisition period mismatch. In areas where this problem was not observed, the correction performed by both methods were similar, although our product generally resulted in greater elevation reduction than FABDEM.

Additionally, a blurring effect was observed in vegetated areas of both corrected DEMs, which degraded the finer topographic features that, despite the forest cover, were visible in the original DEM. This was caused by the post-processing steps of the correction methods, where the DEMs were smoothed by filters to reduce the noise in the final product. In our method, a single bilateral filter was applied, whereas in FABDEM a series of filtering steps were performed [15]. As a result, the blur in FABDEM was greater than the one observed in the new corrected DEM. Figure 10 shows the blurring effect in a region of Area 1, where topographic features related to the local drainage network were lost in FABDEM but remained preserved in our product. This problem may also have negatively contributed to its performance in Area 1.

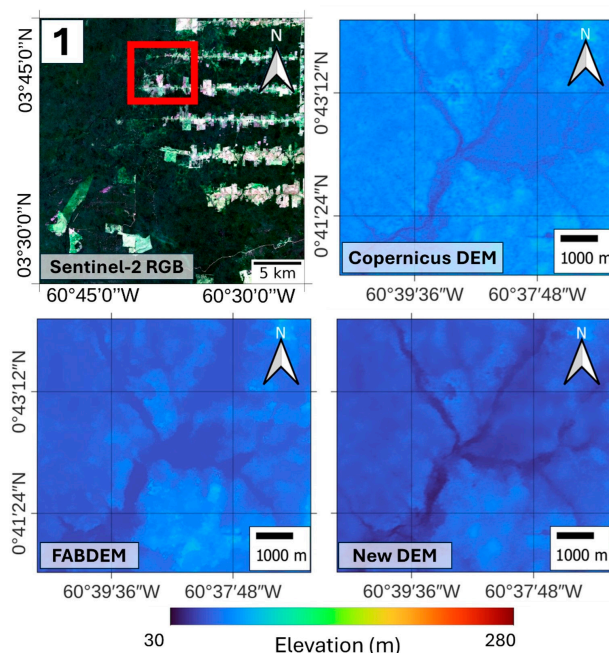


Figure 10. Example of a region within Area 1 where the blurring effect was identified. The figure presents a natural color Sentinel-2 cloud-free composite of the year 2020 of the study area, overlaid by red rectangle highlighting the region featured in the other panels (top left); a color representation of the elevations of Copernicus DEM (top right), FABDEM (bottom left) and the new corrected DEM (bottom right), showing the different level of degradation of the finer topographic features visible in the original DEM.

Figure 11 presents the results obtained for Area 2, where it can be observed that the bias, caused by the riparian vegetation existing along the main rivers, was reduced by both methods, but its remnants are more evident in FABDEM than they are in the new corrected DEM. This can be observed in profiles 1, 2, and 3, where the bias appears as plateaus centered around the river, which are not perceptible in our product. However, in areas where the terrain is more rugged, river valleys are narrower and the riparian vegetation

less prominent, such as the one represented by profile 4, this effect is less noticeable, and we could not identify which method performed better bias reduction, based only on the information provided by their elevation profiles. It can also be noticed that our method resulted in a pronounced reduction in the elevation of the valley bottom located in this profile, but this kind of effect does not affect the delineation of flow paths. Moreover, no blurring effect was detected in this area, due to its reduced forest cover in comparison to Area 1.

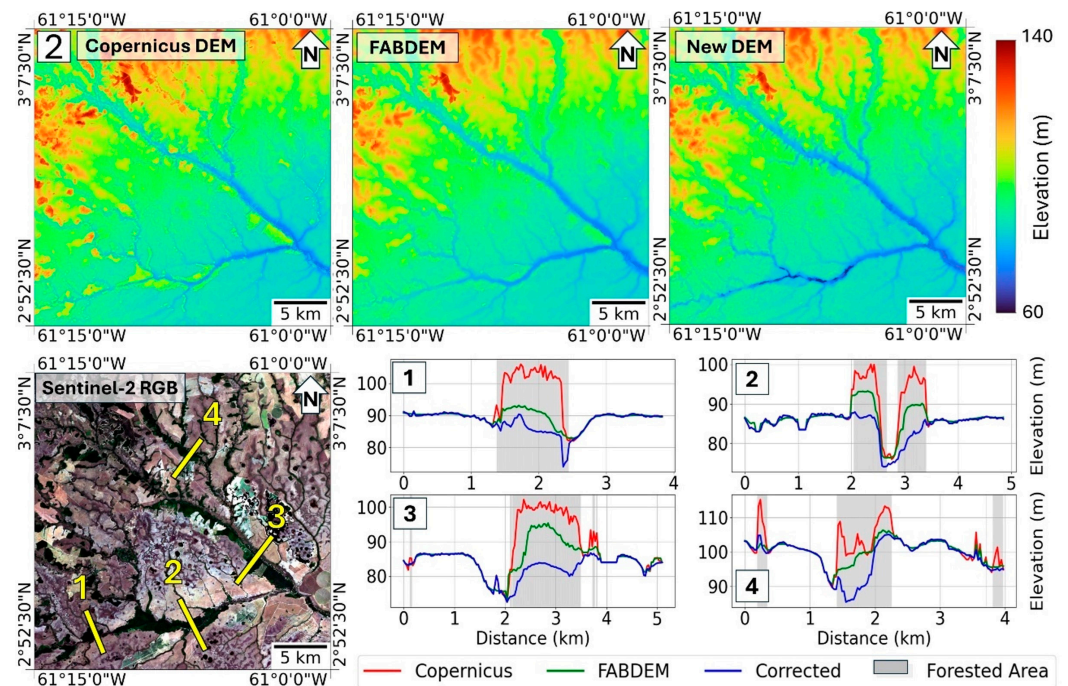


Figure 11. Comparison between DEMs and vertical profiles in Area 2. The figure presents color representations of Copernicus, FABDEM, and the new corrected DEM elevation data over the study area (**top**); its natural color Sentinel-2 cloud-free composite of the year 2020, overlaid by the elevation profile lines identified by their respective numbers (**bottom left**); and charts showing the observed DEM elevations along the profile lines, with the background colored gray in areas covered by vegetation, according to the adjusted forest height obtained for the area (**bottom right**).

Figure 12 presents the results obtained for Area 3. In this area, because of its rougher terrain and relatively greater elevation gradients, compared to other study areas and to the local average forest height, the topography is the main factor contributing to the variations in Copernicus DEM elevation throughout the region. As a result, the difference among all DEMs, in terms of vegetation bias presence, is hardly noticeable by visual interpretation. The elevation profile analysis showed that, except for few regions where the characteristic plateau shape was identified in the original data and FABDEM (represented by profile 2), the result of both correction methods was also similar, not providing enough indication of which one had a better performance.

The results of the visual analysis are not supported by the findings of the statistical comparison in this area, since they pointed out that both our product and Copernicus DEM were significantly better than FABDEM, when the radius was 1000 and 2000 m. However, it can be seen in Figure 13 that a blurring effect was also detected in the vegetated regions of Area 3, which was more intense in FABDEM than in the new corrected DEM. Since, aside from this effect, the DEMs are visually similar in this area, and this is indicative that the increased blurring present in FABDEM contributed to its lower performance in this area.

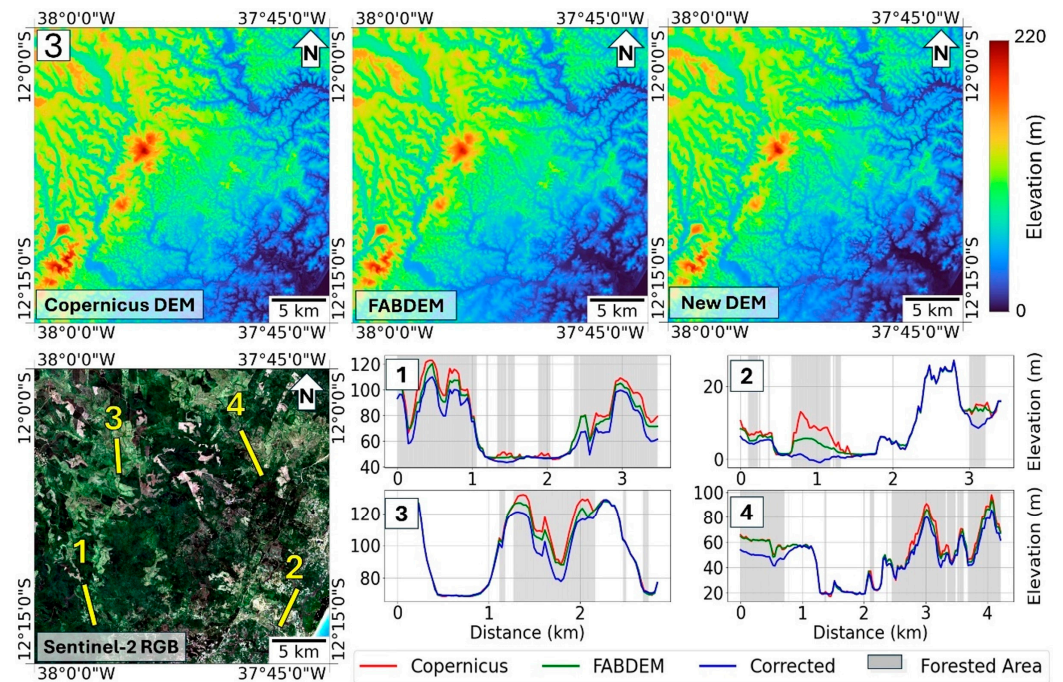


Figure 12. Comparison between DEMs and vertical profiles in Area 3. The figure presents color representations of Copernicus, FABDEM, and the new corrected DEM elevation data over the study area (top); its natural color Sentinel-2 cloud-free composite of the year 2020, overlaid by the elevation profile lines identified by their respective numbers (bottom left); and charts showing the observed DEM elevations along the profile lines, with the background colored gray in areas covered by vegetation, according to the adjusted forest height obtained for the area (bottom right).

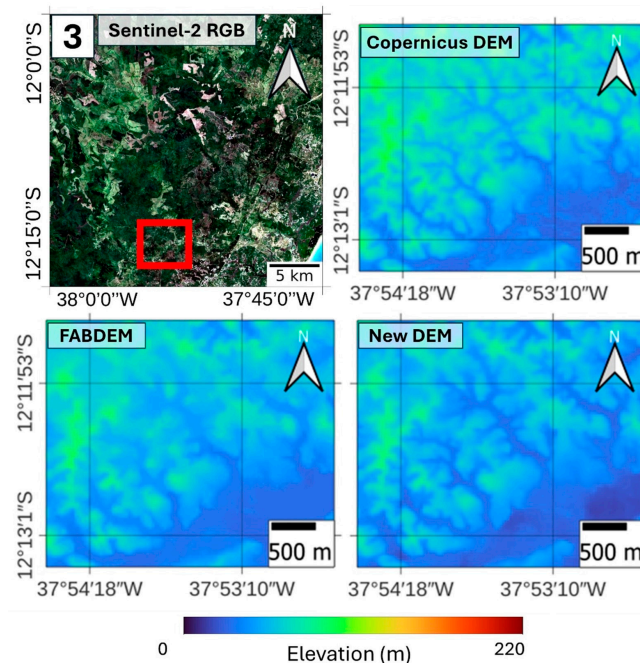


Figure 13. Example of a region within Area 3 where the blurring effect was identified. The figure presents a natural color Sentinel-2 cloud-free composite of the year 2020 of the study area, overlaid by red rectangle highlighting the region featured in the other panels (top left); a color representation of the elevations of Copernicus DEM (top right), FABDEM (bottom left) and the new corrected DEM (bottom right), showing the different level of degradation of the finer topographic features visible in the original DEM.

Figure 14 presents the results for Area 4, and similarly to Area 2 the bias caused by the riparian vegetation surrounding the main rivers was reduced by both methods, but our corrected DEM presented smoother forest/non-forest transitions than FABDEM, especially in the valley of the main river that crosses the area from west to east. Profiles 1 to 3 are placed along this region and show that the plateau-shaped elevation increases caused by the vegetation bias was further reduced by our corrected DEM. Area 4 also contains regions with rougher terrain, where just as in Area 3, the local topography variations mask the vegetation bias, making it difficult to perceive the effect of its correction by visual interpretation alone. Profile 4 represents one of such regions, where the bias correction result of each method was similar. Moreover, no blurring effect was detected in Area 4 as well, because it also has a reduced forest cover compared to Areas 1 and 3.

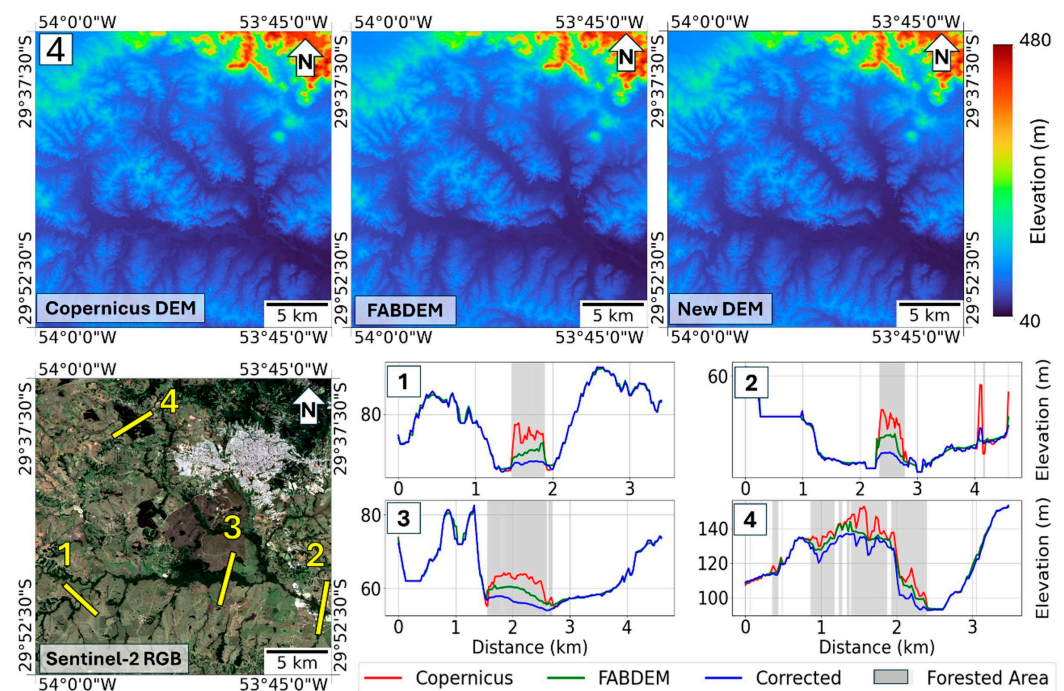


Figure 14. Comparison between DEMs and vertical profiles in Area 4. The figure presents color representations of Copernicus, FABDEM and our corrected DEM elevation data over the study area (**top**); its natural color Sentinel-2 cloud-free composite of the year 2020, overlaid by the elevation profile lines identified by their respective numbers (**bottom left**); and charts showing the observed DEM elevations along the profile lines, with the background colored gray in areas covered by vegetation, according to the adjusted forest height obtained for the area (**bottom right**).

The drainage line visual evaluation was performed on drainage networks extracted from the DEMs using arbitrary accumulation area thresholds that resulted in a drainage density on par with the reference drainage networks. For areas 1 and 2, which were obtained from topographic maps at the scale of 1:50,000, the selected threshold was equivalent to the area of 1000 pixels, and for areas 3 and 4, obtained from 1:25,000 maps, the selected threshold was equivalent to 100 pixels, considering the spatial resolution of 1 arc second. During the visual inspection of the drainage networks, minor deviations from the reference, in terms of line tracing, were not considered; however, other errors were found. The errors detected in a particular location were present either on the drainage lines derived from all DEMs, or only on those extracted from Copernicus DEM and/or FABDEM. We did not find cases where the drainage lines obtained from the new corrected DEM did not match the reference when the ones from any of the other DEMs matched it. Figure 15 illustrates the differences between the drainage networks, where it can be observed that the new corrected DEM was superior to Copernicus DEM in Areas 2 and 4, and superior to FABDEM in Areas 1, 2, and 4. In Area 3, the result of all DEMs was similar to the reference.

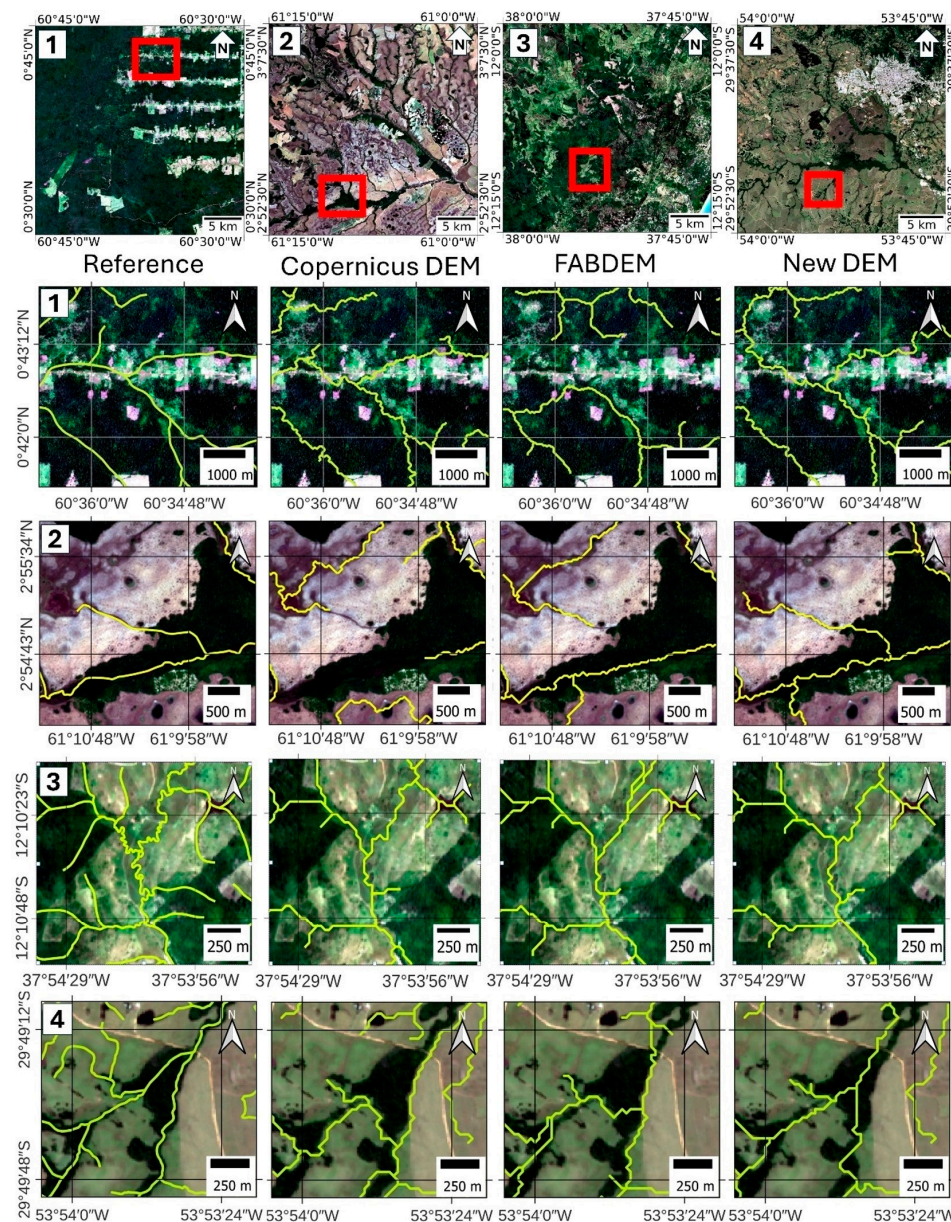


Figure 15. Comparison of drainage networks extracted from the DEMs. The figure is composed of the natural color Sentinel-2 cloud-free composite of the year 2020 of the study areas overlaid by a red rectangle/highlighting the regions featured in the panels below (first row); Sentinel-2 composite of the highlighted regions, overlaid by the reference drainage lines and the ones extracted from Copernicus DEM, FABDEM, and the new corrected DEM, all in yellow color and placed side by side, organized in rows per study area.

4. Discussion

The results presented above show that our product outperformed FABDEM, currently the best global DEM available at 1 arc second resolution [49,54,59], at stream flow path delineation on the targeted study areas. However, even though the method was tested in areas presenting different characteristics, only the forest cover percentage and the overall topography of the region were considered in their selection. Further testing considering the influence of different vegetation and landform types can be performed to complement the results presented in this work. Additionally, the evaluation of our product using other DEM comparison methods, not specific to stream flow path delineation, can be performed to assess if the results translate to other applications.

Regarding limitations, the method can be affected by the errors and inaccuracies present on the forest height dataset, described in Section 2.2.2, although their effects can be mitigated by the adaptive way the adjustment factors are calculated. Even though the automatic adjustment of the forest height data to a reference year is an advantage our method has over FABDEM, especially on areas where the deforestation was more dynamic in the last decade, the process can also be affected by omission and commission errors in the GFC forest loss data.

Another characteristic of our method that can become a limitation in certain cases is the use of a single adjustment factor per forested area to perform the bias correction. Although this was not observed in the selected study areas, we need to assume that distinct regions within a single vegetated area can require different adjustment factors to be successfully corrected. However, since our method calculates it based only on the borders of a given forested area, the result may not be adequate for all regions inside of it, introducing errors in the correction process. However, the application of a single k value per area also has the advantage of better preserving the local topographic features that are visible, despite the forest cover, in areas where the vegetation is homogeneous. Conversely, if multiple adjustment factors would be applied to a single area, these features could be distorted or even lost. So, this matter is more a design choice than a flaw in the method.

While these limitations need to be taken into consideration when correcting Copernicus DEM to perform stream flow path delineation of specific areas, they were outweighed by the advantages our method provided in the target study areas, as indicated by the results presented in this work. Additionally, the limitations can also be mitigated by replacing the input datasets (such as base DEM, forest height, and loss data) with more accurate ones as they become available.

5. Conclusions

This study presented a new method to correct the vegetation bias on Copernicus DEM, which depending on regional characteristics, can restrict the use of these data for certain applications. The method is deterministic, not requiring training data collection, presenting an important advantage over the recent trend of vegetation bias correction solutions based on machine learning. These solutions are highly dependent on the quality and size of the training data, a factor that can be limiting, since compiling enough samples to represent a wide variety of areas is challenging, especially on continental or global scales. Moreover, the new method requires only two input datasets, no feature engineering, and at the same time is regionally adaptive.

The bias was corrected by subtracting from the DEM elevations the forest height, obtained from an independent dataset, modified by adjustment factors, which were locally determined through slope analysis of the borders of every vegetation patch. Moreover, the forest height dataset was automatically adjusted to match the Copernicus DEM acquisition year in each study area, to prevent vegetation cover temporal mismatch.

The resulting product was compared to FABDEM, which is the best free global DEM currently available. The comparison focus was on hydrology, more specifically in DEM-based stream flow path delineation. In this application, the presence of vegetation bias prevents proper estimation of downstream paths, because the spurious elevations on vegetated areas can block the flow. The results pointed out that our product showed significant improvement in stream flow path delineation over FABDEM and Copernicus DEM in the majority of the cases tested, and it was not significantly inferior to them in any of the tests. A visual analysis was performed corroborating the results and showing that our method prevented the creation of elevation artifacts caused by the mismatch between the acquisition periods of Copernicus DEM and the forest height dataset. Additionally, this analysis also indicated that our method resulted in lower degradation of the finer topographic features on forested areas caused by noise-reducing steps during the vegetation bias correction processes.

The results presented in this work suggest that the use of this new vegetation bias correction method has the potential to improve DEM-based hydrological applications worldwide. The findings can be complemented by testing the method in more study areas, to evaluate the influence of different vegetation and landform types in the product, and different comparison methods can be employed to assess if the results translate to other applications besides stream flow path delineation. Regarding data distribution, our next goal is to use the method to generate a new open access corrected DEM for the South American continent and later for the entire globe.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16224332/s1>, Table S1: Correction method input and output files; Table S2: Databases generated from the displacement areas of all study areas; Table S3: Geopackage files containing the flow paths set for each radius and their respective displacement areas for each DEM; Table S4: Geopackage files containing the selected samples from the flow paths set for each radius and their respective displacement areas for each DEM.

Author Contributions: Conceptualization, G.T.B. and C.D.R.; Methodology, G.T.B. and C.D.R.; Software, G.T.B.; Validation, C.D.R.; Formal Analysis, G.T.B.; Investigation, G.T.B. and C.D.R.; Resources, G.T.B.; Data Curation, G.T.B.; Writing—Original Draft Preparation, G.T.B.; Writing—Review and Editing, C.D.R.; Visualization, G.T.B.; Supervision, C.D.R.; Project Administration, C.D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article Supplementary Materials, and the scripts of the correction and comparison methods can be provided upon request; further enquiries can be directed to the corresponding author. Copernicus DEM is publicly available at <https://sentinels.copernicus.eu/web/sentinel/view-data-product> (accessed on 10 September 2024). The forest height data [44] and the GFC forest loss per year [40] were obtained from via Google Earth Engine, respectively, at <https://glad.umd.edu/dataset/GLCLUC2020> (accessed on 10 September 2024) and https://developers.google.com/earth-engine/datasets/catalog/UMD_hansen_global_forest_change_2023_v1_11 (accessed on 10 September 2024). FABDEM is publicly distributed by the University of Bristol and can be accessed at <https://data.bris.ac.uk/data/dataset/s5hqmjcdj8yo2ibzi9b4ew3sn> (accessed on 10 September 2024). The reference drainage networks can be accessed at the Brazilian Army Geographic Databases (BDGEx—<https://bdgex.eb.mil.br/>) (accessed on 10 September 2024), but high clearance level is required to download the vectorial data from the topographic maps. Alternatively, these data can be made available upon request by the authors, if authorized by the Brazilian Army Geographic Service (DSG).

Acknowledgments: The authors are thankful for the support provided by the Brazilian Army Science and Technology Department (DCT), the Brazilian Army Geographic Service (DSG), and the National Institute of Space Research (INPE).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications. *Hydrol. Process.* **1991**, *5*, 3–30. [CrossRef]
2. Wechsler, S.P. Uncertainties Associated with Digital Elevation Models for Hydrologic Applications: A Review. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1481–1500. [CrossRef]
3. Zaidi, S.M.; Akbari, A.; Gisen, J.I.; Kazmi, J.H.; Gul, A.; Fhong, N.Z. Utilization of Satellite-Based Digital Elevation Model (DEM) for Hydrologic Applications: A Review. *J. Geol. Soc. India* **2018**, *92*, 329–336. [CrossRef]
4. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [CrossRef]
5. Ali, M.H.; Popescu, I.; Jonoski, A.; Solomatine, D.P. Remote Sensed and/or Global Datasets for Distributed Hydrological Modelling: A Review. *Remote Sens.* **2023**, *15*, 1642. [CrossRef]
6. Schumann, G.J.-P.; Bates, P.D. The Need for a High-Accuracy, Open-Access Global DEM. *Front. Earth Sci.* **2018**, *6*, 225. [CrossRef]
7. Tavares Da Costa, R.; Mazzoli, P.; Bagli, S. Limitations Posed by Free DEMs in Watershed Studies: The Case of River Tanaro in Italy. *Front. Earth Sci.* **2019**, *7*, 141. [CrossRef]

8. Tachikawa, T.; Hato, M.; Kaku, M.; Iwasaki, A. Characteristics of ASTER GDEM Version 2. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 3657–3660.
9. Tadono, T.; Nagai, H.; Ishida, H.; Oda, F.; Naito, S.; Minakawa, K.; Iwamoto, H. GENERATION OF THE 30 M-MESH GLOBAL DIGITAL SURFACE MODEL BY ALOS PRISM. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B4*, 157–162. [[CrossRef](#)]
10. Crippen, R.; Buckley, S.; Agram, P.; Belz, E.; Gurrola, E.; Hensley, S.; Kobrick, M.; Lavallo, M.; Martin, J.; Neumann, M.; et al. Nasadem global elevation model: Methods and progress. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B4*, 125–128. [[CrossRef](#)]
11. Rizzoli, P.; Martone, M.; Gonzalez, C.; Wecklich, C.; Borla Tridon, D.; Bräutigam, B.; Bachmann, M.; Schulze, D.; Fritz, T.; Huber, M.; et al. Generation and Performance Assessment of the Global TanDEM-X Digital Elevation Model. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 119–139. [[CrossRef](#)]
12. Fahrland, E.; Paschko, H.; Jacob, P.; Kahabka, D. *Copernicus DEM Product Handbook*; Airbus Defence and Space—Intelligence: Potsdam, Germany, 2020; Volume 2.
13. Carabajal, C.C.; Harding, D.J. SRTM C-Band and ICESat Laser Altimetry Elevation Comparisons as a Function of Tree Cover and Relief. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 287–298. [[CrossRef](#)]
14. Li, H.; Zhao, J.; Yan, B.; Yue, L.; Wang, L. Global DEMs Vary from One to Another: An Evaluation of Newly Released Copernicus, NASA and AW3D30 DEM on Selected Terrains of China Using ICESat-2 Altimetry Data. *Int. J. Digit. Earth* **2022**, *15*, 1149–1168. [[CrossRef](#)]
15. Hawker, L.; Uhe, P.; Paulo, L.; Sosa, J.; Savage, J.; Sampson, C.; Neal, J. A 30 m Global Map of Elevation with Forests and Buildings Removed. *Environ. Res. Lett.* **2022**, *17*, 024016. [[CrossRef](#)]
16. Gallant, J.C.; Read, A.M.; Dowling, T.I. Removal of tree offsets from SRTM and other digital surface models. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *XXXIX-B4*, 275–280. [[CrossRef](#)]
17. Rosim, S.; Oliveira, J.d.F.; Jardim, A.C.; Namikawa, L.M.; Rennó, C.D. TerraHidro: A Distributed Hydrology Modelling System with High Quality Drainage Extraction. In Proceedings of the International Conference on Advanced Geographic Information Systems, Applications, and Services, Nice, France, 24 February–1 March 2013; Volume 5, pp. 161–167.
18. Baugh, C.A.; Bates, P.D.; Schumann, G.; Trigg, M.A. SRTM Vegetation Removal and Hydrodynamic Modeling Accuracy: SRTM vegetation error removal. *Water Resour. Res.* **2013**, *49*, 5276–5289. [[CrossRef](#)]
19. Köthe, R.; Bock, M. Preprocessing of Digital Elevation Models—Derived from Laser Scanning and Radar Interferometry—for Terrain Analysis in Geosciences. In Proceedings of the Geomorphometry 2009, Zurich, Switzerland, 31 August–2 September 2009; pp. 155–161.
20. Brochado, G.T. Atenuação do Efeito do Desflorestamento em Dados SRTM por Meio de Diferentes Técnicas de Interpolação. Master’s Thesis, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, 2015.
21. Gallant, J.C.; Read, A.M. A near-global bare-earth dem from srtm. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B4*, 137–141. [[CrossRef](#)]
22. Wilson, M.; Bates, P.; Alsdorf, D.; Forsberg, B.; Horritt, M.; Melack, J.; Frappart, F.; Famiglietti, J. Modeling Large-Scale Inundation of Amazonian Seasonally Flooded Wetlands. *Geophys. Res. Lett.* **2007**, *34*. [[CrossRef](#)]
23. Su, Y.; Guo, Q. A Practical Method for SRTM DEM Correction over Vegetated Mountain Areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 216–228. [[CrossRef](#)]
24. Su, Y.; Guo, Q.; Ma, Q.; Li, W. SRTM DEM Correction in Vegetated Mountain Areas through the Integration of Spaceborne LiDAR, Airborne LiDAR, and Optical Imagery. *Remote Sens.* **2015**, *7*, 11202–11225. [[CrossRef](#)]
25. O’Loughlin, F.E.; Paiva, R.C.D.; Durand, M.; Alsdorf, D.E.; Bates, P.D. A Multi-Sensor Approach towards a Global Vegetation Corrected SRTM DEM Product. *Remote Sens. Environ.* **2016**, *182*, 49–59. [[CrossRef](#)]
26. Yamazaki, D.; Ikeshima, D.; Tawatari, R.; Yamaguchi, T.; O’Loughlin, F.; Neal, J.C.; Sampson, C.C.; Kanae, S.; Bates, P.D. A High-Accuracy Map of Global Terrain Elevations. *Geophys. Res. Lett.* **2017**, *44*, 5844–5853. [[CrossRef](#)]
27. Zhao, X.; Su, Y.; Hu, T.; Chen, L.; Gao, S.; Wang, R.; Jin, S.; Guo, Q. A Global Corrected SRTM DEM Product for Vegetated Areas. *Remote Sens. Lett.* **2018**, *9*, 393–402. [[CrossRef](#)]
28. Magruder, L.; Neunschwander, A.; Klotz, B. Digital Terrain Model Elevation Corrections Using Space-Based Imagery and ICESat-2 Laser Altimetry. *Remote Sens. Environ.* **2021**, *264*, 112621. [[CrossRef](#)]
29. Chen, C.; Yang, S.; Li, Y. Accuracy Assessment and Correction of SRTM DEM Using ICESat/GLAS Data under Data Coregistration. *Remote Sens.* **2020**, *12*, 3435. [[CrossRef](#)]
30. Laipelt, L.; Comini De Andrade, B.; Collischonn, W.; De Amorim Teixeira, A.; Paiva, R.C.D.D.; Ruhoff, A. ANADEM: A Digital Terrain Model for South America. *Remote Sens.* **2024**, *16*, 2321. [[CrossRef](#)]
31. Wendi, D.; Liang, S.; Sun, Y.; Doan, C.D. An Innovative Approach to Improve SRTM DEM Using Multispectral Imagery and Artificial Neural Network. *J. Adv. Model. Earth Syst.* **2016**, *8*, 691–702. [[CrossRef](#)]
32. Kulp, S.A.; Strauss, B.H. CoastalDEM: A Global Coastal Digital Elevation Model Improved from SRTM Using a Neural Network. *Remote Sens. Environ.* **2018**, *206*, 231–239. [[CrossRef](#)]
33. Li, Y.; Li, L.; Chen, C.; Liu, Y. Correction of Global Digital Elevation Models in Forested Areas Using an Artificial Neural Network-Based Method with the Consideration of Spatial Autocorrelation. *Int. J. Digit. Earth* **2023**, *16*, 1568–1588. [[CrossRef](#)]

34. Hengl, T.; Leal Parente, L.; Krizan, J.; Bonannella, C. Continental Europe Digital Terrain Model at 30 m Resolution Based on GEDI, ICESat-2, AW3D, GLO-30, EUDEM, MERIT DEM and Background Layers. Available online: <https://zenodo.org/records/4724549> (accessed on 10 September 2024).
35. Ouyang, Z.; Zhou, C.; Xie, J.; Zhu, J.; Zhang, G.; Ao, M. SRTM DEM Correction Using Ensemble Machine Learning Algorithm. *Remote Sens.* **2023**, *15*, 3946. [[CrossRef](#)]
36. Li, Y.; Fu, H.; Zhu, J.; Wu, K.; Yang, P.; Wang, L.; Gao, S. A Method for SRTM DEM Elevation Error Correction in Forested Areas Using ICESat-2 Data and Vegetation Classification Data. *Remote Sens.* **2022**, *14*, 3380. [[CrossRef](#)]
37. Pimenova, O.; Roberts, C.; Rizos, C. Regional “Bare-Earth” Digital Terrain Model for Costa Rica Based on NASADEM Corrected for Vegetation Bias. *Remote Sens.* **2022**, *14*, 2421. [[CrossRef](#)]
38. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Carroll, M.; Dimiceli, C.; Sohlberg, R.A. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. *Earth Interact.* **2003**, *7*, 1–15. [[CrossRef](#)]
39. Sexton, J.O.; Song, X.-P.; Feng, M.; Noojipady, P.; Anand, A.; Huang, C.; Kim, D.-H.; Collins, K.M.; Channan, S.; DiMiceli, C.; et al. Global, 30-m Resolution Continuous Fields of Tree Cover: Landsat-Based Rescaling of MODIS Vegetation Continuous Fields with Lidar-Based Estimates of Error. *Int. J. Digit. Earth* **2013**, *6*, 427–448. [[CrossRef](#)]
40. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [[CrossRef](#)]
41. Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. ESA WorldCover 10 m 2021 V200. Available online: <https://zenodo.org/record/7254221> (accessed on 10 September 2024).
42. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global Land Use/Land Cover with Sentinel 2 and Deep Learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4704–4707.
43. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping Forest Canopy Height Globally with Spaceborne Lidar. *J. Geophys. Res.* **2011**, *116*, G04021. [[CrossRef](#)]
44. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sens. Environ.* **2021**, *253*, 112165. [[CrossRef](#)]
45. Lang, N.; Jetz, W.; Schindler, K.; Wegner, J.D. A High-Resolution Canopy Height Model of the Earth. *Nat. Ecol. Evol.* **2023**, *7*, 1778–1789. [[CrossRef](#)]
46. Tolan, J.; Yang, H.-I.; Nosarzewski, B.; Couairon, G.; Vo, H.V.; Brandt, J.; Spore, J.; Majumdar, S.; Haziza, D.; Vamaraju, J.; et al. Very High Resolution Canopy Height Maps from RGB Imagery Using Self-Supervised Vision Transformer and Convolutional Decoder Trained on Aerial Lidar. *Remote Sens. Environ.* **2024**, *300*, 113888. [[CrossRef](#)]
47. Guth, P.L.; Geoffroy, T.M. LiDAR Point Cloud and ICESat-2 Evaluation of 1 Second Global Digital Elevation Models: Copernicus Wins. *Trans. GIS* **2021**, *25*, 2245–2261. [[CrossRef](#)]
48. Purinton, B.; Bookhagen, B. Beyond Vertical Point Accuracy: Assessing Inter-Pixel Consistency in 30 m Global DEMs for the Arid Central Andes. *Front. Earth Sci.* **2021**, *9*, 758606. [[CrossRef](#)]
49. Bielski, C.; López-Vázquez, C.; Grohmann, C.H.; Guth, P.L.; Hawker, L.; Gesch, D.; Trevisani, S.; Herrera-Cruz, V.; Riazanoff, S.; Corseaux, A.; et al. Novel Approach for Ranking DEMs: Copernicus DEM Improves One Arc Second Open Global Topography. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–22. [[CrossRef](#)]
50. Moges, D.M.; Virro, H.; Kmoch, A.; Cibin, R.; Rohith, A.N.; Martínez-Salvador, A.; Conesa-García, C.; Uuemaa, E. How Does the Choice of DEMs Affect Catchment Hydrological Modeling? *Sci. Total Environ.* **2023**, *892*, 164627. [[CrossRef](#)] [[PubMed](#)]
51. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [[CrossRef](#)]
52. Pavlis, N.K.; Holmes, S.A.; Kenyon, S.C.; Factor, J.K. The Development and Evaluation of the Earth Gravitational Model 2008 (EGM2008). *J. Geophys. Res.* **2012**, *117*, 2011JB008916. [[CrossRef](#)]
53. Goyal, R.; Featherstone, W.E.; Dikshit, O.; Balasubramania, N. Comparison and Validation of Satellite-Derived Digital Surface/Elevation Models over India. *J. Indian Soc. Remote Sens.* **2021**, *49*, 971–986. [[CrossRef](#)]
54. Marsh, C.B.; Harder, P.; Pomeroy, J.W. Validation of FABDEM, a Global Bare-Earth Elevation Model, against UAV-Lidar Derived Elevation in a Complex Forested Mountain Catchment. *Environ. Res. Commun.* **2023**, *5*, 031009. [[CrossRef](#)]
55. Potapov, P.; Hansen, M.C.; Kommareddy, I.; Kommareddy, A.; Turubanova, S.; Pickens, A.; Adusei, B.; Tyukavina, A.; Ying, Q. Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sens.* **2020**, *12*, 426. [[CrossRef](#)]
56. Dutra, L.V.; Elmiro, M.T.; Soares, B.S.; Mura, F.J.C.; Santos, J.R.; Freitas, C.C.; Araujo, L.S.; De Albuquerque, P.C.G.; Vieira, P.R.; Gama, F.F. Assessment of digital elevation models obtained in Brazilian Amazon based on P and X band airborne interferometric data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; IEEE: Toronto, ON, Canada, 2002; Volume 6, pp. 3617–3619.
57. Correia, A.; Vieira, P. Topographic and Forestry Mapping from Brazilian Amazon Using Synthetic Aperture Radar Images. In Proceedings of the 25th International Cartographic Conference, Paris, France, 3–8 July 2011.

58. DSG—Diretoria de Serviço Geográfico. Norma Da Especificação Técnica Para Aquisição de Dados Geoespaciais Vetoriais 2018. Available online: https://geoportal.eb.mil.br/portal/images/Documentos/2024/ET-ADGV_3.0_211218.pdf (accessed on 10 September 2024).
59. Meadows, M.; Jones, S.; Reinke, K. Vertical Accuracy Assessment of Freely Available Global DEMs (FABDEM, Copernicus DEM, NASADEM, AW3D30 and SRTM) in Flood-Prone Environments. *Int. J. Digit. Earth* **2024**, *17*, 2308734. [[CrossRef](#)]
60. Neuenschwander, A.; Pitts, K.; Jelley, B.; Robbins, J.; Klotz, B.; Popescu, S.; Nelson, R.; Harding, D.; Pederson, D.; Sheridan, R. ATLAS/ICESat-2 L3A Land and Vegetation Height. *Version* **2020**, *4*, 17.
61. Buchhorn, M.; Smets, B.; Bertels, L.; De Roo, B.; Lesiv, M.; Tsendbazar, N.-E.; Herold, M.; Fritz, S. Copernicus Global Land Service: Land Cover 100m: Collection 3 Epoch 2015, Globe V3.0.1. Available online: <https://zenodo.org/records/3939038> (accessed on 10 September 2024).
62. NASA Earth at Night: Flat Maps. Available online: <https://earthobservatory.nasa.gov/features/NightLights/page3.php> (accessed on 11 July 2024).
63. Marconcini, M.; Metz-Marconcini, A.; Üreyen, S.; Palacios-Lopez, D.; Hanke, W.; Bachofer, F.; Zeidler, J.; Esch, T.; Gorelick, N.; Kakarla, A.; et al. Outlining Where Humans Live, the World Settlement Footprint 2015. *Sci. Data* **2020**, *7*, 242. [[CrossRef](#)]
64. Neal, J.; Hawker, L. *Fabdem V1-2*; University of Bristol: Bristol, UK, 2023. [[CrossRef](#)]
65. Schlund, M.; Baron, D.; Magdon, P.; Erasmi, S. Canopy Penetration Depth Estimation with TanDEM-X and Its Compensation in Temperate Forests. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 232–241. [[CrossRef](#)]
66. Tomasi, C.; Manduchi, R. Bilateral Filtering for Gray and Color Images. In Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 4–7 January 1998; Narosa Publishing House: Bombay, India, 1998; pp. 839–846.
67. Durand, F.; Dorsey, J. Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, San Antonio, TX, USA, 23–26 July 2002; pp. 257–266.
68. Lindsay, J.B. Whitebox GAT: A Case Study in Geomorphometric Analysis. *Comput. Geosci.* **2016**, *95*, 75–84. [[CrossRef](#)]
69. Sedgewick, R. *Algorithms in C++*; Addison-Wesley Publishing Company: Reading, PA, USA, 1992; ISBN 978-0-201-51059-1.
70. O’Callaghan, J.F.; Mark, D.M. The Extraction of Drainage Networks from Digital Elevation Data. *Comput. Vis. Graph. Image Process.* **1984**, *28*, 323–344. [[CrossRef](#)]
71. Jardim, A.C. Direções de Fluxo em Modelos Digitais de Elevação: Um Método com Foco na Qualidade da Estimativa e Processamento de Grande Volume de Dados. Ph.D. Thesis, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, 2017.
72. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80–83. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.