On the Data Quality of Remotely Sensed Forest Maps

Shadi Ghasemitaheri^{1,*}, Amelia Holcomb², Lukasz Golab¹ and Srinivasan Keshav²

¹University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

²University of Cambridge, The Old Schools, Trinity Lane, Cambridge, CB2 1TN, United Kingdom

Abstract

Accurate forest monitoring data are essential for understanding and conserving forest ecosystems. However, the remoteness of forests and the scarcity of ground truth make it hard to identify data quality issues. We present two state-of-the-art forest monitoring datasets, Annual Forest Change (AFC) and GEDI, and highlight their data quality problems. We then introduce a novel method that leverages GEDI to identify data quality issues in AFC. We show that our approach can identify subsets with three times more errors than a random sample, thus prioritizing expert resources in validating AFC and allowing for more accurate deforestation detection.

Keywords

Data Quality, Data Cleaning, Remote Sensing, Forest Monitoring, GEDI

1. Introduction

Data-intensive models are only as good as their training data. As a result, the past two decades have seen a great deal of research and industry effort toward monitoring and improving data quality. Solutions exist for deduplication, missing data imputation, and identifying and repairing incorrect data [1]. However, as data-intensive systems gain traction in new application areas, new data quality problems arise, complicating the task of identifying incorrect data.

We present a novel approach to finding data errors in one such application area: forest monitoring. Forests have a significant impact on the Earth's climate and biodiversity [2, 3], but they have been severely damaged by deforestation and climate change [4]. To create effective conservation policies, it is crucial to accurately map forest change (e.g., deforestation or degradation) on a global scale. Forest change maps help scientists understand the impacts of deforestation [5, 6] and are used in government policymaking and reporting [7].

Many forest change maps are created from satellite images [8, 9], leading to new data quality problems related to sensor limitations, cloud cover, and low resolution. For instance, these images lack forest height information, which is useful in detecting deforestation [9]. Evaluating the accuracy of these maps is also a complex and costly task due to the limited availability of ground truth data, as collecting forest condition data through field visits is expensive and does not scale. As a result, there is no simple way of identifying errors in forest change maps.

To identify errors in a widely-used forest change map, we use GEDI [10], a recent spaceborne LiDAR dataset that provides information about the 3D structure of forests not available in optical images. Specifically, we identify nonforested areas, either deforested or non-forest vegetation. that are incorrectly labeled as tropical forests. Despite GEDI's limited spatial and historical coverage, we show that GEDI's estimates of canopy height (the height of the top of the forest) can identify parts of the forest change map that are three times more likely to contain errors than a random sample. Our approach can be used to prioritize resources for validating a forest change map and assist in more accurate detection of deforestation.

Our contributions are as follows:

- · We describe data quality issues associated with two datasets: a state-of-the-art forest change map (AFC) [8] (Section 2), and GEDI [10] (Section 3).
- We propose and evaluate a method that leverages GEDI data to identify potential errors in AFC (Section 4, 5).

We review related work in Section 6 and conclude in Section 7.

2. Annual Forest Change Data

The Annual Forest Change (AFC) dataset (Figure 1a) tracks annual changes in tropical moist forests (TMFs) from 1990 to 2021 [8]. It segments different land categories, such as TMFs, water bodies, and grasslands, as well as identifying changes in land cover, such as degradation and deforestation.

Joint Workshops at 49th International Conference on Very Large Data Bases (VLDBW'23) - the 12th International Workshop on Quality in Databases (QDB'23), August 28 - September 1, 2023, Vancouver, Canada

^{*}Corresponding author.

[†]These authors contributed equally.

Shadi.ghasemitaheri@uwaterloo.ca (S. Ghasemitaheri);

ah2174@cam.ac.uk (A. Holcomb); lgolab@uwaterloo.ca (L. Golab); sk818@cam.ac.uk (S. Keshav) © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 40 International (CC BY 40). CEUR Workshop Proceedings (CEUR-WS.org)



December 2021.

(a) An illustration of the Annual Forest Change (AFC) map, (b) A Satellite image. Landsat L9, January 2022. The bottom right corner of the image shows an example of missing data due to cloud coverage.

Figure 1: The AFC map and a satellite image from the same location.

Structure AFC maps the annual boundaries and status of TMFs. An AFC map is a 2D grid of pixels, each corresponding to a 30 m \times 30 m (0.09 ha) area on the Earth's surface at the equator. It classifies each pixel into one of six categories: Undisturbed TMF, Degraded TMF, Deforested land, TMF Regrowth, Water, and Other Land Covers.

Source AFC maps are derived from optical satellite imagery of the Landsat satellites [11] (Figure 1b). A Landsat satellite captures images of the Earth's surface from 705 kilometers above, revisiting each location every 16 days. These images are taken by cameras onboard the satellites that capture different wavelengths of light including, visible light, infrared, and other wavelengths. Vegetation types are recognized by how they absorb or reflect light.

Methodology The AFC dataset is based on per-pixel classifications of Landsat images. Each pixel is classified using expert rules as either potential moist forest, potential non-forest, or invalid (cloud, shadows, noise). Each pixel is then assigned a final class (from the above six categories) based on the changes over time.

Accuracy AFC is reported to be 91.4% accurate but underestimates forest disturbance by 11.8% [8]. This corresponds to over 38 million hectares of land, which is a significant area [8].

Data Quality Challenges Forest maps, including AFC, face the following challenges.

Missing Data: There could be gaps in satellite observations for several reasons, including cloud cover and atmospheric conditions (Figure 1b), sensor limitations or failures, and intentional pauses in data collection.

- Noisy Data: Satellite imagery is prone to sensor noise, miscalibration, and atmospheric noise.
- Spatial and Temporal Resolution: The resolution of a forest map is determined by the resolution of the source data. For instance, AFC cannot tell the precise location of disruptions or changes smaller than 0.09 hectares.
- Spectral Mixing: Satellite images often have mixed pixels containing different land cover types (e.g., half forest and half deforested). This issue occurs frequently in complex vegetation covers or at the boundaries between different land cover types.
- Spectral Confusion: This occurs when different types of land cover have similar appearance when viewed from space. For instance, Figure 2 shows how a cocoa agroforest looks similar to a forest in optical satellite imagery [12].
- Lack of 3D Information: Optical satellite images lack 3D information such as forest height, limiting their ability to distinguish between some land cover types. For example, height information can distinguish forests from grasslands.
- Limited Ground Truth: Collecting data by visiting forests ranges from expensive to impossible (for remote and inaccessible locations). As a result, experts rely on remote sensing to create a reference dataset.

We introduce a method that directs experts' attention toward a subset of samples that are more likely to contain errors. To achieve this goal, we use a new forest height

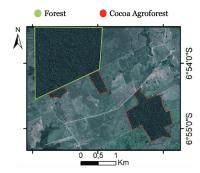


Figure 2: Spectral confusion [12]. The cocoa agroforest looks identical to a forest in optical satellite imagery.

dataset, described below, with its own unique data quality issues and challenges.

3. Forest Height Data

Global Ecosystem Dynamics Investigation (GEDI) is a LiDAR (Light Detection and Ranging) instrument that collects data about the Earth's forests from space [10]. LiDAR emits laser beams and measures the time it takes for the light to return to the sensor. GEDI LiDAR is designed to penetrate the canopy, allowing scientists to study the 3D structure of forests. GEDI operated on the International Space Station (ISS) from 2019 to 2023. Figure 3 illustrates a GEDI return waveform.

Structure A GEDI observation corresponds to a fragment of the Earth's surface with a 25 m diameter called a footprint (Figure 3). GEDI was estimated to collect over 10 billion cloud-free observations in two years [10].

Data Products Raw GEDI waveforms are processed into higher-level data products that describe the 3D features of forests. For example, GEDI Level 2A data include measurements of ground elevation and relative height (RH) [13]. RH is the height above ground at which a certain quantile of cumulative energy was returned (Figure 3), and the *RH95* (95% quantile) has been shown to estimate canopy height (height of the top of the forest) [14]. These products are used to create 1 km × 1 km gridded maps.

Accuracy GEDI was calibrated and validated using a ground truth dataset in which the evergreen broadleaf forests of South America were well represented [10]. Studies show that GEDI can accurately estimate RH95 with RMSE of 2.08 m [15].

Data Quality Challenges Similar to other data collected from space, GEDI data has the following data quality issues:

 Noisy Data: As a laser-based technology, GEDI is sensitive to atmospheric conditions, including

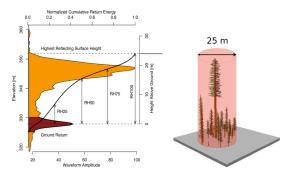


Figure 3: GEDI return waveform [10]. The waveform (left) captures the reflected energy at different elevations from the 25 m diameter footprint (right).

cloud cover and moisture. Sensor noise and miscalibration also contribute to errors in the data.

- Spatial and Temporal Resolution: GEDI footprints cover a limited portion of the Earth (around 4% in two years of operation) [10], and gridded maps have a relatively coarse resolution of 1 km. Additionally, operating from 2018 to 2023, GEDI does not offer extensive historical information. Finally, there are no guaranteed revisits of the same location, making it difficult to monitor changes.
- **Terrain:** Sloped or complex terrain introduces additional errors in GEDI data [16].
- Geolocation Error: Slight geolocation uncertainties (0 m mean, 10 m standard deviation exist in the reported coordinates. This uncertainty can significantly affect RH metrics in mixed canopies and forest edges [17].

GEDI's spatiotemporal limitations prevent scientists from creating high-resolution forest change maps based solely on GEDI data. Nevertheless, GEDI can help address the *lack of 3D information, spectral confusion*, and *limited ground truth* problems in AFC. GEDI offers 3D information for remote unreachable forests. Therefore, GEDI data (like canopy height) can help distinguish forest covers that may look the same in optical satellite imagery. While the spatial limitations of GEDI prevent us from evaluating the entire AFC dataset, we present a novel method to identify data quality issues in AFC more efficiently than random sampling while accounting for geolocation error and noise in GEDI data.

4. Approach

Forests commonly consist of tall green trees: the formal definition of a forest requires canopies to be at least 5 metres tall [18]. Therefore, areas with shorter canopies

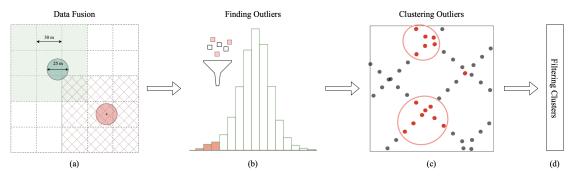


Figure 4: An illustration of dataflow. (a) *Data Fusion*. The grid is the AFC map, and the circles are GEDI shots. The nearest 3×3 windows are highlighted for each shot. (b) *Finding Outliers*. Samples with *RH*95 < *h* are selected. (c) *Clustering Outliers*. Nearby outliers are clustered. (d) *Filtering Clusters*. Smaller clusters are filtered to increase reliability.

are more likely to be instances of *deforestation* or *other land covers* such as grasslands.

We define areas with an *undisturbed* label in AFC and short canopy height in GEDI as conflicts or outliers. We suggest these conflicts can be more effective in identifying errors in the AFC dataset than randomly selected samples. Note that these conflicts represent potential errors that could arise from noise in either the GEDI or AFC data. Thus, several challenges need to be addressed:

- Integrating the two datasets, the AFC map and GEDI footprint data, while accounting for geolocation errors.
- Accounting for the noise in GEDI data and finding samples that are more trusted.
- Prioritizing some outliers when dealing with thousands of conflicts, as manually examining all of them is too time-consuming.

We propose a four-step process to utilize GEDI canopy heights (RH95) to identify forests labeled as undisturbed but having conflicting (short) canopy heights. Figure 4 shows the dataflow.

Step 1: Data Fusion. We identify the nine nearest AFC pixels to each GEDI shot. These pixels form a 3 × 3 window on the AFC map, with the center pixel containing the GEDI shot center (Figure 4). We only consider GEDI shots within homogeneous windows. This accounts for potential geolocation errors in the GEDI shot.

Step 2: Finding Outliers. We select GEDI shots with RH95 < h, where h is a tuneable parameter. These shots are *undisturbed* TMFs with an abnormally short canopy. The parameter h can be selected based on expert knowledge or the RH95 distribution.

Step 3: Clustering Outliers. We merge nearby outliers into clusters using hierarchical clustering for two reasons:

• Reducing data noise. Several nearby conflicting

observations are more trusted than a single outlier.

 Conflicts occurring close together can belong to the same area and cover type, corresponding to spatially correlated errors [19]. For instance, two consecutive GEDI shots are only 60 m apart, and both may be from a grassland misclassified as an *undisturbed* TMF in AFC. Using clustering, we avoid reporting these points separately.

Hierarchical clustering has two parameters: linkage and distance threshold. Linkage determines how the distance between two clusters is calculated; e.g., single-linkage uses the minimum distance between members of the two clusters. The distance threshold determines if clusters should be combined, merging only those closer than the threshold.

Step 4: Filtering Clusters. Clusters with few conflicts are less likely to represent areas with AFC errors than ones with many conflicts. Hence, we prioritize clusters larger than a certain threshold, *c*. Additionally, clusters containing GEDI shots from multiple satellite orbits are more reliable and less susceptible to systematic errors. This is because consecutive shots within a single orbit could all be incorrect due to atmospheric conditions or sensor issues.

There are three parameters in this method: height threshold (h), clustering distance (d), and minimum cluster size (c). We discuss the effects and trade-offs of these parameters in Section 5.

5. Evaluation

We used our method to find data quality issues in the 2021 AFC map of the Brazilian Amazon region. We used RH95 from quality-filtered GEDI shots, as detailed in Burns et al. [20], collected during the second half of 2021. In this section, we describe two results. The first relied on ground truth from a validated forest cover map, while the second was from our own visual interpretation of high-resolution satellite images.

Study Region We focus on the Brazilian Amazon rainforest, which plays a vital role in global climate stability, and is home to unique plant and animal species. The availability of numerous forest maps and freely accessible satellite data makes the Brazilian Amazon an ideal region for our studies.

Parameters *Height Threshold* (h) is a tradeoff between precision and recall. A lower threshold reduces the number of outliers, which can reduce false positives but may affect recall. A lower *Clustering Distance Threshold* (d) leads to smaller clusters that may represent correlated errors. A higher distance threshold can merge unrelated clusters or create clusters of multiple noisy samples. *Minimum Cluster Size* (c) also impacts precision and recall. Although small clusters are more likely to be false positives, choosing a large *c* affects the recall of small-scale errors.

Based on empirical fine-tuning, we selected h = 3.44 meters to mark 0.3% of the GEDI shots in *undisturbed* TMFs as outliers. A lower threshold (e.g., 2 meters) eliminated some evident AFC errors, whereas a higher threshold (e.g., 4 meters) included many shots that were ambiguous as to whether they were AFC errors. We apply single-linkage hierarchical clustering with a distance of d = 700 meters to group outliers that are from the same GEDI orbit. We also filter clusters with fewer than c = 8 shots.

MapBiomas Evaluation MapBiomas [21, 22] is an annual dataset of Brazil's land cover maps from 1985 to 2021 at a 30-metre resolution. It uses a hierarchical classification system with four levels to categorize land covers. At Level 1, land covers are classified into six categories: forest, non-forest, farming, non-vegetated, water, and not observed. Level 2 expands Level 1 classes into 22 subcategories. MapBiomas is created from Landsat images, primarily using a Random Forest classifier, and it is validated annually on over 75,000 samples. Level 1 and 2 classification error is estimated to be 7.5% and 9.3%, respectively [21].

In this analysis, we use MapBiomas as a ground truth dataset. Specifically, we assume that if an outlier shot is labeled as *undisturbed* TMF in the AFC map but classified as non-forest in MapBiomas, then we consider MapBiomas to be correct, meaning that the outlier is an error in the AFC map. We report two validation metrics: (1) the percentage of outliers with non-forest MapBiomas labels and (2) the percentage of clusters with at least one such outlier.

Visual Interpretation After finding outlier clusters, we randomly select one GEDI shot per cluster. Then, we determine if this represents an AFC error by analyzing the 3×3 surrounding AFC pixels in a cloud-free image.

We use higher-resolution satellite images with approximately 4 m per pixel resolution from the Planet NICFI data program [23, 24]. Specifically, we used the last cloudfree Planet images of 2021. Each cluster is assigned one of three validation labels: *Ambiguous* (if no cloud-free observations are available or if it is unclear whether the area is an AFC error), *AFC Error*, or *False Positive* (if the pixels are correctly classified in AFC). Analyzing 3 × 3 windows of the map is similar to AFC's validation method [8].

Results We identified 23,306 conflicts (i.e., marked *undisturbed* forest in AFC with *RH*95 < 3.44 m) in Step 2. After filtering clusters in Step 4, 5,740 samples remain, of which 1.88% are labeled non-forest in MapBiomas. This gives 240 clusters, 12.08% of which have at least one outlier with a non-forest MapBiomas label. Since manual evaluation is time-consuming, we evaluate 100 random clusters out of the 240 clusters using Planet images. Out of the 100 clusters, 33 were found to be AFC errors, 63 were Ambiguous, and 4 were False Positives (see Figure 5 for examples). Assuming that all Ambiguous cases are False Positives, the precision of our method is at least 33%, which is almost three times greater than the precision of random sampling reported by [8].

Discussion Visual interpretation revealed cases where both AFC and MapBiomas were inaccurate. This can be because of MapBiomas limitations due to the lack of 3D information in Landsat images. While MapBiomas has the advantage of evaluating every pixel in AFC, GEDI, despite its limited coverage, uncovers errors that Map-Biomas may not detect. Additionally, there are some vegetation types that should be classified as non-forest covers in AFC but are considered forests in MapBiomas. For instance, MapBiomas assigns wooded savannah and tropical evergreens to the same class, while AFC refers to the former as *other land covers*. Therefore, per-pixel evaluation cannot identify AFC errors in wooded savannahs, but using canopy height can.

It is useful to study where AFC made errors. We found many outlier clusters in the Brazilian Amazon's northwest region, with a vegetation cover known as *campinarana* that can be difficult to distinguish in satellite images [25]. This region is remote and challenging to access, making it difficult to obtain field data. Various types of campinarana include forest, wood, shrub, and grass on sandy infertile soil [26]. Forest campinaranas are up to 20 meters tall, whereas canopy height in the non-forest classes does not exceed 4 meters [26]. Shrub campinarana appears green in satellite images which might lead to AFC errors.

Some *False Positives* were located near shores and water that could potentially be covered by mangroves (Figure 5d). Mangroves have a distinct structure that differs from evergreen or semi-evergreen forests. However, all three types are categorized as TMFs in AFC. Excluding such areas from the analysis could improve precision.

We also attempted to identify TMFs misclassified as deforested by filtering deforested samples with RH95 > 20 metres. However, we were unsuccessful for several reasons. First, this approach does not reflect the complex nature of forests. Seeing a few square meters of trees does not indicate the presence of a forest. Second, the lack of historical height data prevents us from analyzing changes to distinguish primary forests from replacing tree plantations. Furthermore, RH95 is prone to obstacle noise, such as from a flock of birds.

6. Related Work

Holcomb et al. [27] recently discovered discrepancies between GEDI biomass (amount of organic matter in a forest) data and AFC label. They observed instances where areas with near-zero biomass were classified as regrowing forests (6+ years old), and other regions with substantial (> 200 Mg/ha) biomass, were classified as young (< 3 years old) regrowing forests after deforestation.

In this study, we used canopy height to find data quality issues in a forest change map. A recent study explored the use of canopy height to distinguish forested and unforested tropical wetlands [28]. They used a global canopy height map with a 30 metre resolution, created by combining GEDI RH metrics and Landsat images [14]. In contrast, our approach relies solely on raw GEDI height measurements.

In addition to creating a forest change map that estimates the year of forest loss, Hansen et al. [9] studied the relationship between loss year and canopy height using an older spaceborne LiDAR dataset. They observed that samples from undisturbed forests, on average, had greater canopy height than disturbed forests. This is consistent with our work.

Assessing the pixel-level agreement of two forest change maps is another way to find errors. However, this can be challenging due to variations in the map legends and differences in resolution. Moreover, two maps based on Landsat optical images can be subject to the same data quality problems. Cohen et al. compared seven forest change maps based on Landsat at pixel level, revealing a low level of agreement [29]. On the other hand, GEDI allows us to leverage 3D information that does not exist in Landsat.

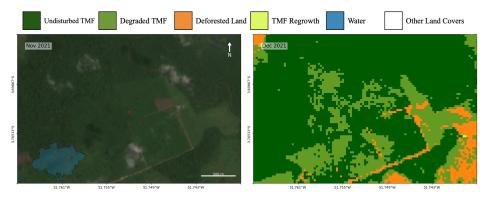
7. Conclusions

We described AFC and GEDI, two important forest monitoring datasets, and their data quality issues. Although GEDI alone cannot be used to create a forest change map, it provides 3D information about forests missing from optical satellite imagery. We proposed a novel approach to find data quality issues in AFC using GEDI data, specifically, areas marked as TMF in the AFC map but with low RH95. Our findings show that this information can be used to create more accurate forest change maps.

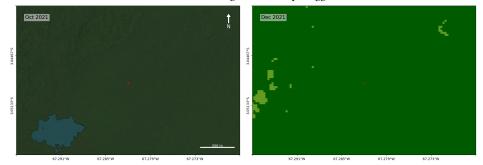
Since no ground truth data were available, we used ancillary data in our evaluation [26, 22]. This process was limited by the interpretation of a single evaluator, and future studies could benefit from using a voting technique and involving experts. Our method could be used to identify errors in other land cover classification maps by finding inconsistencies between GEDI data and the target class. For instance, we can apply this method to find errors in forest/non-forest maps. Furthermore, GEDI data products provide additional measurements, such as various RH metrics and Leaf Area Index (LAI). Exploring these metrics in future work can reveal other data quality errors in existing datasets.

References

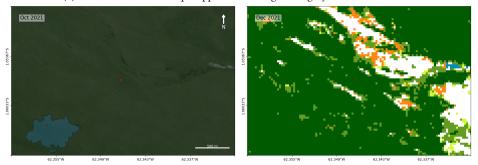
- I. F. Ilyas, X. Chu, Data Cleaning, Association for Computing Machinery, New York, NY, USA, 2019.
- [2] Y. Pan, R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D. McGuire, S. Piao, A. Rautiainen, S. Sitch, D. Hayes, A large and persistent carbon sink in the world's forests, Science 333 (2011) 988–993. doi:10.1126/science.1201609.
- [3] L. Gibson, T. M. Lee, L. P. Koh, B. W. Brook, T. A. Gardner, J. Barlow, C. A. Peres, C. J. Bradshaw, W. F. Laurance, T. E. Lovejoy, et al., Primary forests are irreplaceable for sustaining tropical biodiversity, Nature 478 (2011) 378–381.
- [4] UNEP, UNEP Annual Report 2022, Technical Report, UN Environment Programme (UNEP), 2023. URL: https://www.unep.org/resources/ annual-report-2022.
- [5] V. H. Heinrich, C. Vancutsem, R. Dalagnol, T. M. Rosan, D. Fawcett, C. H. Silva-Junior, H. L. Cassol, F. Achard, T. Jucker, C. A. Silva, et al., The carbon sink of secondary and degraded humid tropical forests, Nature 615 (2023) 436–442.
- [6] L. Zhu, W. Li, P. Ciais, J. He, A. Cescatti, M. Santoro, K. Tanaka, O. Cartus, Z. Zhao, Y. Xu, et al., Comparable biophysical and biogeochemical feedbacks on warming from tropical moist forest degradation, Nature Geoscience 16 (2023) 244–249. doi:10.1038/s41561-023-01137-y.
- [7] E. Commission, J. R. Centre, R. Beuchle, F. Achard, C. Bourgoin, C. Vancutsem, Deforestation and forest degradation in the Amazon : status and trends



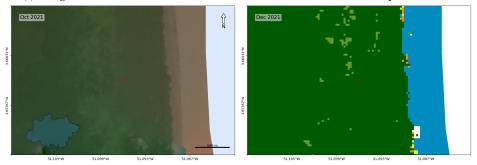
(a) AFC Error. The colour difference, texture difference, and geometric shape suggest that this area is not an undisturbed TMF.



(b) *False Positive*. This sample appears to belong to a highly dense forest cover.



(c) Ambiguous. The available context is insufficient to determine whether this sample is an error.



(d) False Positive. This area could potentially be covered by mangroves, which are classified as TMF.

Figure 5: Four instances of evaluation decisions. The images on the left are higher-resolution Planet imagery. Images on the right visualize the AFC map in the same location.

up to year 2021, Publications Office of the European Union, Luxembourg, 2022. doi:10.2760/398695.

- [8] C. Vancutsem, F. Achard, J.-F. Pekel, G. Vieilledent, S. Carboni, D. Simonetti, J. Gallego, L. E. Aragao, R. Nasi, Long-term (1990–2019) monitoring of forest cover changes in the humid tropics, Science Advances 7 (2021) eabe1603.
- [9] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, et al., Highresolution global maps of 21st-century forest cover change, science 342 (2013) 850–853.
- [10] R. Dubayah, J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, et al., The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography, Science of remote sensing 1 (2020) 100002.
- [11] M. A. Wulder, D. P. Roy, V. C. Radeloff, T. R. Loveland, M. C. Anderson, D. M. Johnson, S. Healey, Z. Zhu, T. A. Scambos, N. Pahlevan, et al., Fifty years of landsat science and impacts, Remote Sensing of Environment 280 (2022) 113195.
- [12] J. E. Batista, N. M. Rodrigues, A. I. Cabral, M. J. Vasconcelos, A. Venturieri, L. G. Silva, S. Silva, Optical time series for the separation of land cover types with similar spectral signatures: cocoa agroforest and forest, International Journal of Remote Sensing 43 (2022) 3298–3319.
- [13] R. Dubayah, M. Hofton, J. Blair, J. Armston, H. Tang, S. Luthcke, Gedi l2a elevation and height metrics data global footprint level v002, 2021. URL: https: //lpdaac.usgs.gov/products/gedi02_av002/. doi:10. 5067/GEDI/GEDI02_A.002.
- [14] P. Potapov, X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C. E. Silva, et al., Mapping global forest canopy height through integration of gedi and landsat data, Remote Sensing of Environment 253 (2021) 112165.
- [15] C. Wang, A. J. Elmore, I. Numata, M. A. Cochrane, L. Shaogang, J. Huang, Y. Zhao, Y. Li, Factors affecting relative height and ground elevation estimations of gedi among forest types across the conterminous usa, GIScience & Remote Sensing 59 (2022) 975–999.
- [16] P. V. Oliveira, X. Zhang, B. Peterson, J. P. Ometto, Using simulated gedi waveforms to evaluate the effects of beam sensitivity and terrain slope on gedi l2a relative height metrics over the brazilian amazon forest, Science of Remote Sensing 7 (2023) 100083. doi:https://doi.org/10.1016/j. srs.2023.100083.
- [17] D. P. Roy, H. B. Kashongwe, J. Armston, The impact of geolocation uncertainty on gedi tropical forest

canopy height estimation and change monitoring, Science of Remote Sensing 4 (2021) 100024.

- [18] A. Pekkarinen, Global forest resources assessment 2020, 2020. URL: https://www.fao.org/3/I8661EN/ i8661en.pdf.
- [19] P. Ploton, F. Mortier, M. Réjou-Méchain, N. Barbier, N. Picard, V. Rossi, C. Dormann, G. Cornu, G. Viennois, N. Bayol, A. Lyapustin, S. Gourlet-Fleury, R. Pélissier, Spatial validation reveals poor predictive performance of large-scale ecological mapping models, Nature Communications 11 (2020) 4540. URL: https://doi.org/10.1038/s41467-020-18321-y. doi:10.1038/s41467-020-18321-y.
- [20] P. Burns, C. Hakkenberg, S. Goetz, Gedi l2ab + l4a global processing from april 2019 to december 2022, 2023. Manuscript in Preparation.
- [21] MapBiomas, Algorithm theoretical basis document (atbd): Collection 7, 2022. https://mapbiomas-br-site.s3.amazonaws.com/ ATBD_Collection_7_v2.pdf.
- [22] M. A. Project, Collection 7.0 of annual land cover and land use maps, 2021. Accessed on 05.2023 via link http://amazonia.mapbiomas.org.
- [23] P. L. PBC, Planet application program interface: In space for life on earth, 2018–. URL: https://api. planet.com.
- [24] NICFI, Norway's international climate and forest initiative, 2022. URL: https://www.nicfi.no/.
- [25] C. L. d. O. Cordeiro, D. d. F. Rossetti, Mapping vegetation in a late quaternary landform of the amazonian wetlands using object-based image analysis and decision tree classification, International Journal of Remote Sensing 36 (2015) 3397–3422.
- [26] IBGE, Manual técnico da vegetação brasileira, 2012.
- [27] A. Holcomb, S. Mathis, S. Keshav, D. A. Coomes, Computational tools for assessing forest recovery with gedi and forest change maps., 2023. In Review.
- [28] K. Verhelst, Y. Gou, M. Herold, J. Reiche, Improving forest baseline maps in tropical wetlands using gedi-based forest height information and sentinel-1, Forests 12 (2021) 1374.
- [29] W. B. Cohen, S. P. Healey, Z. Yang, S. V. Stehman, C. K. Brewer, E. B. Brooks, N. Gorelick, C. Huang, M. J. Hughes, R. E. Kennedy, et al., How similar are forest disturbance maps derived from different landsat time series algorithms?, Forests 8 (2017) 98.