



Article

Advancing Forest Degradation and Regeneration Assessment Through Light Detection and Ranging and Hyperspectral Imaging Integration

Catherine Torres de Almeida ^{1,2,*}, Lênio Soares Galvão ², Jean Pierre H. B. Ometto ², Aline Daniele Jacon ², Francisca Rocha de Souza Pereira ², Luciane Yumie Sato ², Celso Henrique Leite Silva-Junior ^{3,4}, Pedro H. S. Brancalion ⁵ and Luiz Eduardo Oliveira e Cruz de Aragão ^{2,6}

¹ Faculty of Agricultural Sciences of Vale do Ribeira, São Paulo State University—UNESP, Registro 11900-000, SP, Brazil

² National Institute for Space Research—INPE, Caixa Postal 515, São José dos Campos 12227-010, SP, Brazil; lenio.galvao@inpe.br (L.S.G.); jean.ometto@inpe.br (J.P.H.B.O.); aline.jacon@inpe.br (A.D.J.); franrspereira@gmail.com (F.R.d.S.P.); luciane.inpe@inpe.br (L.Y.S.); luiz.aragao@inpe.br (L.E.O.e.C.d.A.)

³ Instituto de Pesquisa Ambiental da Amazônia (IPAM), SCN 211, Bloco B, Sala 201, Brasília 70836-520, GO, Brazil; celso.junior@ipam.org.br

⁴ Graduate Program in Biodiversity Conservation, Federal University of Maranhão, São Luís 65080-805, MA, Brazil

⁵ Department of Forest Sciences, “Luiz de Queiroz” College of Agriculture, University of São Paulo, Piracicaba 13418-900, SP, Brazil; pedrob@usp.br

⁶ College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK

* Correspondence: catherine.almeida@unesp.br



Citation: Almeida, C.T.d.; Galvão, L.S.; Ometto, J.P.H.B.; Jacon, A.D.; Pereira, F.R.d.S.; Sato, L.Y.; Silva-Junior, C.H.L.; Brancalion, P.H.S.; Aragão, L.E.O.e.C.d. Advancing Forest Degradation and Regeneration Assessment Through Light Detection and Ranging and Hyperspectral Imaging Integration. *Remote Sens.* **2024**, *16*, 3935. <https://doi.org/10.3390/rs16213935>

Academic Editors: Michael Sprintsin and Luis A. Ruiz

Received: 31 July 2024

Revised: 23 September 2024

Accepted: 21 October 2024

Published: 22 October 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/>).

Abstract: Integrating Light Detection And Ranging (LiDAR) and Hyperspectral Imaging (HSI) enhances the assessment of tropical forest degradation and regeneration, which is crucial for conservation and climate mitigation strategies. This study optimized procedures using combined airborne LiDAR, HSI data, and machine learning algorithms across 12 sites in the Brazilian Amazon, covering various environmental and anthropogenic conditions. Four forest classes (undisturbed, degraded, and two stages of second-growth) were identified using Landsat time series (1984–2017) and auxiliary data. Metrics from 600 samples were analyzed with three classifiers: Random Forest, Stochastic Gradient Boosting, and Support Vector Machine. The combination of LiDAR and HSI data improved classification accuracy by up to 12% compared with single data sources. The most decisive metrics were LiDAR-based upper canopy cover and HSI-based absorption bands in the near-infrared and shortwave infrared. LiDAR produced significantly fewer errors for discriminating second-growth from old-growth forests, while HSI had better performance to discriminate degraded from undisturbed forests. HSI-only models performed similarly to LiDAR-only models (mean F1 of about 75% for both data sources). The results highlight the potential of integrating LiDAR and HSI data to improve our understanding of forest dynamics in the context of nature-based solutions to mitigate climate change impacts.

Keywords: forest disturbance; forest recovery; successional stages; airborne laser scanning (ALS); hyperspectral remote sensing; multisensor analysis; machine learning

1. Introduction

Reducing deforestation and forest degradation and promoting their large-scale restoration are promising nature-based solutions to combat climate change [1]. Several governments, private companies, non-governmental organizations (NGOs), and multilateral organizations have made ambitious pledges to promote forest protection and restoration at unprecedented scales, putting these activities at the core of United Nations’ Sustainable Development Goals and Decade on Ecosystem Restoration (2021–2030) [2]. One major

challenge for integrating deforestation, forest degradation, and restoration as part of climate mitigation initiatives, such as those based on payments for ecosystem services (PESs), is to accurately differentiate forests under degradation or regeneration processes and relatively intact forests. Depending on the stage of degradation and regeneration processes, degraded and secondary forests can be structurally similar but play different ecological functions, such as their potential to store carbon [3]. Thus, the discrimination of forests under anthropogenic influences is essential for reliable accountability of climate mitigation interventions, reducing the uncertainties of carbon stock estimates [4]. One of the most promising ways to better characterize land use and land cover (LULC) is through advanced remote sensing approaches.

Remote sensing is an effective source of information on forest traits on the landscape scale. Specifically, data from passive multispectral sensors have been commonly used to differentiate LULC types [5,6]. Despite the potential of such data in LULC studies, the great heterogeneity and complexity of tropical forests pose a challenge in obtaining accurate information on vegetation disturbance through conventional approaches. Thus, advanced remote sensing technologies, such as Hyperspectral Imaging (HSI) and Light Detection And Ranging (LiDAR), provide new opportunities to answer complex ecological questions in tropical forests experiencing degradation and regeneration processes.

HSI systems acquire data in many narrow and contiguous spectral bands, generating high-resolution reflectance spectra per pixel [7]. The ability of HSI to extract more accurate and detailed information compared with other passive remote technologies makes it suitable for a wide variety of applications. Examples include the classification of tree species or LULC classes [8–11], the identification of physiological responses to stress [12], the estimation of biochemical attributes [13], the detection of burned areas [14], and the study of the canopy phenology [15,16]. According to Thenkabail et al. [17], there are advantages of using hyperspectral data over multispectral imagery for improving the classification accuracy of complex rainforests such as those from southern Cameroon.

Active LiDAR sensors produce three-dimensional measurements of forests, allowing the quantification of important structural attributes such as canopy height, Leaf Area Density, and biomass [18]. That accurate structural information has allowed the discrimination of successional stages in tropical forests [19,20]. However, LiDAR systems currently capture limited spectral information, which creates difficulties in distinguishing structurally similar forests with distinct species composition or under stress conditions. Thus, the spectral information from HSI can complement the structural information provided by LiDAR. As reported by Almeida et al. [21], the combination of LiDAR and hyperspectral data is useful to increase the accuracy of aboveground biomass (AGB) estimates over heterogeneous human-modified landscapes of the Brazilian Amazon. In this context, the integration of LiDAR and hyperspectral data can potentially improve the characterization of the degradation and regeneration status of tropical forests. Despite the potential of this synergism, few studies have tested the combined use of LiDAR and HSI data for the classification of land cover types and successional stages in tropical ecosystems. For instance, in the tropical dry forests of Costa Rica, Sun et al. [22] used different airborne remote sensing data (waveform LiDAR, HSI, and their combination) and machine learning classifiers to map secondary forest age. The best result was found with the combination of LiDAR and HSI data (overall accuracy of 83%).

The Brazilian Amazon concentrates a large share of Reducing Emissions from Deforestation and Forest Degradation (REDD) initiatives and is broadly recognized for its key role in providing critical ecosystem services to local, regional, and global socio-ecological systems [23]. The deforestation of native vegetation in the Brazilian Amazon has been monitored over the past decades using remote sensing approaches [24]. In contrast, assessing the status of secondary vegetation (regrowth after complete clearing) and more subtle forest disturbances (e.g., selective logging and fire) is more challenging [25,26]. Advancing this characterization is critically needed for improving conservation, management, and restoration strategies.

In this study, we evaluated the potential of integrating airborne LiDAR and HSI data acquired over tropical forests of the Amazon to distinguish four classes of vegetation degradation and regeneration. The classes were previously defined with the use of historical Landsat data and auxiliary information: undisturbed old-growth forests (UFs), degraded old-growth forests (DFs), younger second-growth forests (SF_{1-15yr}), and older second-growth forests ($SF_{16-32yr}$). To achieve this goal, several LiDAR and HSI metrics related to structural and functional characteristics were calculated and submitted to three machine learning classifiers: Random Forest (RF), Stochastic Gradient Boosting (SGB), and Support Vector Machine (SVM). Finally, we investigated the effect of the data source and classifiers to better characterize these forest classes, as well as the ability to transfer the predictions of the best models to new sites. With this approach, we aimed to address the following research questions:

1. What LiDAR and HSI metrics are most effective for distinguishing the forest classes, and how do these metrics vary among them?
2. How do the choice of remote sensing data sources (single LiDAR, single HSI, or their combination) and machine learning classifiers impact the accuracy of classifying forest degradation and regeneration stages in tropical forests?
3. What is the capacity of regional models to generalize and accurately predict forest classes in new sites?

To the best of our knowledge, this study is the first to discriminate forest degradation and regeneration classes in the Brazilian Amazon by using a large set of LiDAR and HSI metrics. Although extensive airborne LiDAR datasets have been utilized to support significant research on forest structure in the Amazon [27,28], no previous studies have combined these data with hyperspectral imagery across such a broad spatial scale to specifically enhance the assessment of degradation and regeneration processes. By integrating both structural and compositional vegetation information, this study sought to advance the current understanding of Amazonian forest dynamics, providing new insights into remote sensing methodologies for tropical forest monitoring.

2. Materials and Methods

2.1. Study Area

Twelve sites distributed throughout the Brazilian Amazon were considered in this study (Table 1). Each site is represented by one (most sites) or two (three sites) transects of approximately 12.5 km by 0.3 km, where the airborne remote sensing data were collected (Figure 1). The sites encompass a wide variety of anthropogenic, climatic, geological, and edaphic conditions.

Table 1. Description of the study sites. Altitudes are from the LiDAR Digital Terrain Model. MAT (Mean Annual Temperature) and MAP (Mean Annual Precipitation) are from WorldClim version 2 [29].

Site	Description	Latitude	Longitude	Altitude	MAT	MAP
		(°)	(°)	(m)	(°C)	(mm·yr ⁻¹)
MAM	Mamirauá Sustainable Development Reserve	−2.76	−65.10	36.7	26.7	3406
ZF2	Cuieiras Biological Reserve	−2.60	−60.21	61.6	26.4	2356
DUC	Adolpho Ducke Reserve	−2.95	−59.94	86.2	26.5	2308
AUT	Autazes municipality	−3.51	−59.26	25.7	27.0	2293
TAP	Tapajós National Forest	−3.12	−54.95	123.0	25.8	1848
PAR	Paragominas municipality	−3.28	−47.52	128.8	25.9	1915
JAM	Jamari National Forest	−9.12	−63.01	93.6	25.2	2388
ALF	Alta Floresta/Novo Mundo municipality	−9.58	−55.90	254.0	26.6	2216
SFX1	São Félix do Xingu municipality	−6.43	−52.11	205.3	24.9	1981
SFX2	São Félix do Xingu municipality	−6.56	−51.81	289.1	24.6	1964
FN1	Feliz Natal municipality	−12.00	−54.20	320.0	24.7	1815
FN2	Feliz Natal municipality	−12.26	−55.10	338.7	24.7	1807

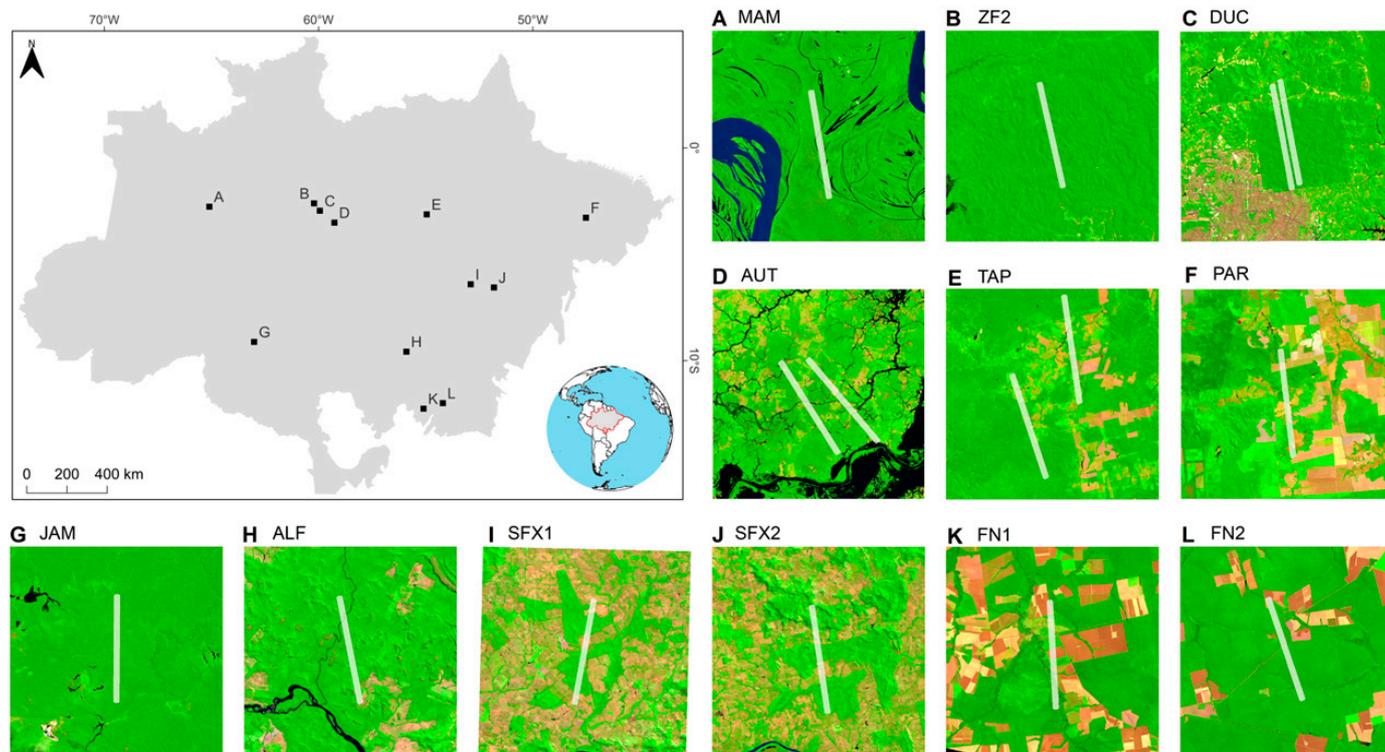


Figure 1. Distribution of the 12 sites in the Brazilian Amazon. The images obtained in 2017 are OLI/Landsat-8 color composites with bands 6 (red), 5 (green), and 4 (blue). The transects in white (12.5×0.3 km) represent the HSI and LiDAR superposed flight lines. The acronyms represent the site names (see Table 1 for details).

From the sites used in this study, MAM, ZF2, DUC, AUT, and TAP are part of the so-called Central Amazonia region, comprising old sedimentary substrates and low soil fertility [30]. On the other hand, the sites PAR, JAM, ALF, SFX1, SFX2, FN1, and FN2 are located over the Brazilian Shield composed of Pre-Cambrian rocks with related high-fertility soils. The predominant soil types are Acrisols and Ferralsols, with Gleysols occurring in the seasonal floodplains of the MAM site [30]. From a topographic point of view, all sites are considered lowlands, having altitudes lower than 500 m. The AUT and MAM sites present the lowest altitude (<50 m), while the southeastern sites (SFX1, SFX2, ALF, FN1, and FN2) have the highest values (200 to 500 m).

The rainfall gradient ranges from wetter conditions on the MAM, ZF2, DUC, and AUT sites to drier conditions on the PAR, SFX1, SFX2, ALF, FN1, and FN2 sites. The long-term (1973–2013) annual rainfall reported for the Brazilian Legal Amazon (BLA) is approximately 2100 mm [31]. In the studied sites, annual rainfall ranges from 1800 mm at the FN2 site to more than 3000 mm at the MAM site. The mean annual temperature over the BLA is 26.5 °C, varying in the sites from 24.6 °C at SFX2 to 27.0 °C at AUT.

The forest types over the studied sites encompass seasonally flooded ombrophilous forests (MAM site), terra firme (unflooded) transitional forests (ecotones between ombrophilous and seasonal forests in the FN1 and FN2 sites), and terra firme ombrophilous forests (other sites). Undisturbed forests are mainly located in protected areas (sites MAM, ZF2, DUC, TAP, and JAM), with few relatively intact forests on other sites (AUT, ALF, and FN2). Secondary forests usually occur in small areas that were previously cleared near highways or rivers. They were also commonly found around small communities adjacent to conservation units. Specifically, they were observed near the northern borders of the Adolpho Ducke Reserve (DUC site) and in São Jorge, a community located between the boundaries of the Tapajós National Forest (TAP site). These areas are characterized by a mix

of residential settlements and small-scale agricultural activities, contributing to localized deforestation pressures along the conservation boundaries.

Much of the studied forests have been degraded by fire, selective logging, and/or fragmentation. Understory fires were responsible for the degradation in sites SFX1, SFX2, and ALF, especially in fragmented areas. The Central Amazonian forests of the sites AUT [32] and TAP were affected by extensive fires under the effect of the El Niño Southern Oscillation (ENSO) in 1998/99, 2010, and 2015/16. Major fires were also common in previously logged areas, such as the TAP, PAR, FN1, and FN2 sites. Conventional operations of selective logging were also observed without the occurrence of fires in the sites FN1 and FN2. At the JAM site, reduced-impact logging was authorized by forest concession [33].

2.2. Forest Classes Identification and Sampling

To properly assess tropical forest dynamics, it is important to differentiate between deforestation and degradation. Lapola et al. [34] proposed a framework in which deforestation is defined as the conversion of forest to non-forest land cover, often accompanied by changes in land use, such as agriculture or pasture. Additionally, deforestation can result from successive or severe disturbances that reduce forest cover below a critical threshold, even if land-use change does not occur. In contrast, degradation involves a deterioration in forest condition, such as reduced carbon storage or biodiversity, but without a change in land cover (i.e., the forest remains a forest). This study focused on four key human-induced drivers of tropical forest degradation: habitat fragmentation, timber extraction, forest fires, and extreme droughts.

Following this framework, we categorized forest dynamics of degradation/regeneration into four distinct classes (Figure 2A): undisturbed old-growth forests (UFs), degraded old-growth forests (DFs), younger second-growth forests (SF_{1-15yr}), and older second-growth forests ($SF_{16-32yr}$). Old-growth classes are those where no deforestation was observed between 1984 (the first year from which we tracked the historical Landsat time series) and 2017, while second-growth classes were considered as forests regenerating after deforestation occurred at some point during this period. Undisturbed forests were defined as old-growth forests that showed no evidence of disturbance by fire or selective logging and that were at least 300 m away from forest edges. In contrast, degraded old-growth forests presented at least one of these types of disturbance.

Second-growth forests in the Amazon are commonly separated into three successional stages based on the stand age [8,35]: initial (<5 years), intermediate (5–15 years), and advanced (>15 years) stages. Here, due to the limited spatial coverage of the initial stages across the study sites, we grouped the initial and intermediate successional stages into a broader class of younger second-growth forests. This choice was made to ensure sufficient sample representation and statistical robustness, as the initial stage alone did not occupy a large enough area for reliable analysis. Therefore, we considered two stages of second-growth forests: the class SF_{1-15yr} consisted of areas where the last deforestation event occurred between 2002 and 2016, while the class $SF_{16-32yr}$ included areas where such an event was observed between 1984 and 2001.

To reconstruct the history of forest degradation and regeneration over the sites, we conducted a visual inspection of Landsat images from 1984 to 2017 (TM/Landsat-5, ETM+/Landsat-7, and OLI/Landsat-8) on the Google Earth Engine (GEE) platform. To facilitate our visual interpretation, we plotted the time series of two key spectral indices: the Normalized Difference Vegetation Index (NDVI [36]) and the Normalized Burn Ratio (NBR [37]). By examining fluctuations in these indices over time, we were able to visually identify periods of degradation (e.g., sudden drops in NDVI and NBR following fire events) and regeneration (e.g., gradual increases in NDVI and NBR after a deforestation event). These trends, combined with the spatial patterns visible in the imagery, enabled the classification of forest areas into the four defined categories of degradation and regeneration.

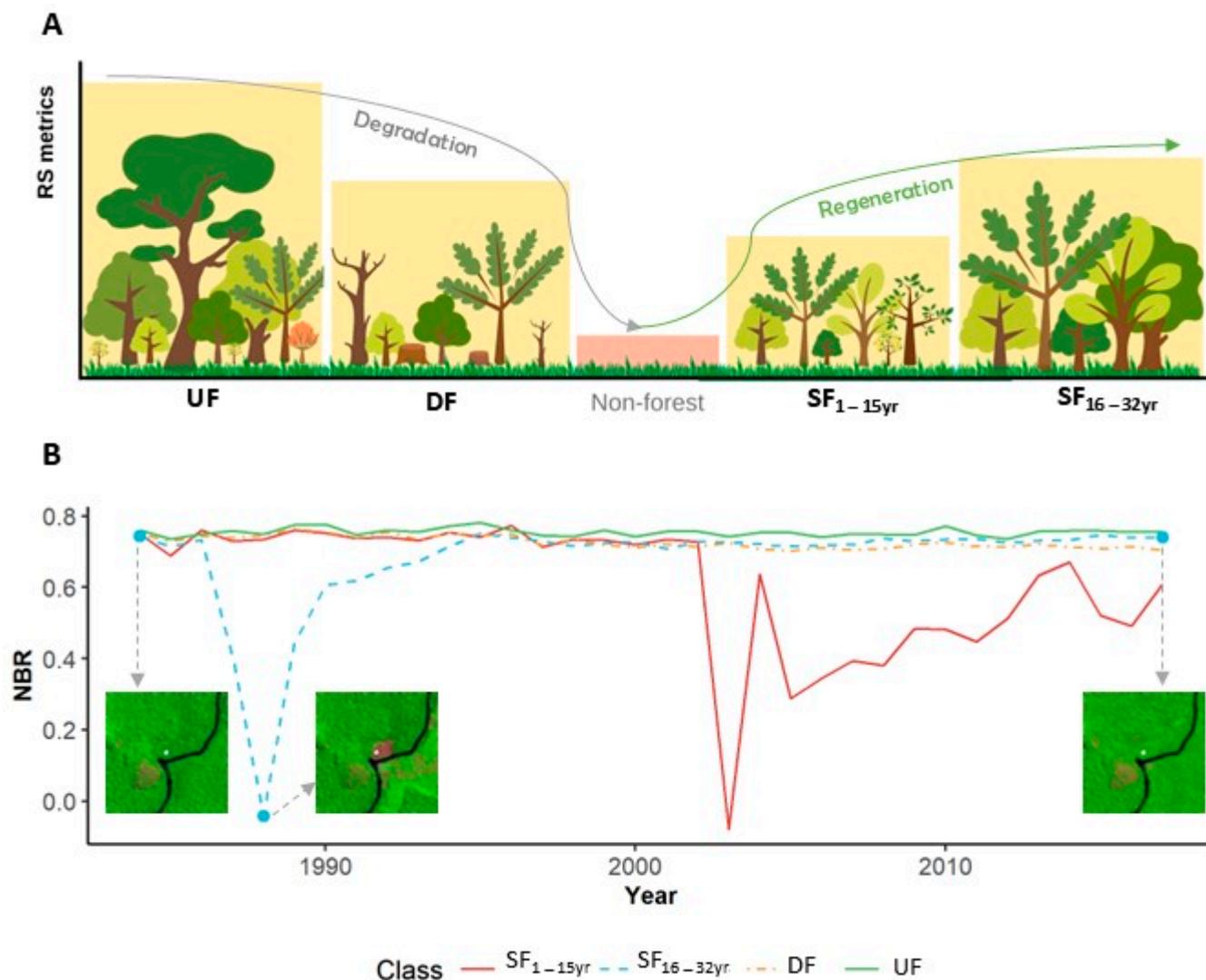


Figure 2. (A) Schematic representation of the variation of a hypothetical remote sensing (RS) metric from LiDAR and HSI in the four classes of vegetation degradation and regeneration considered in this study: undisturbed old-growth forests (UFs); degraded old-growth forests (DFs); younger second-growth forests (SF_{1-15yr}); and older second-growth forests (SF_{16-32yr}). We highlight that some RS metrics may present different patterns of variation; for example, albedo presents an inverse variation of that shown in the figure. (B) Example of time series of the Normalized Burn Ratio (NBR) calculated from Landsat data (1984–2017), indicating clearing events in 1988 and 2003, followed by vegetation recovery in the subsequent years. The inset images are TM/Landsat-5 and OLI/Landsat-8 color composites with bands 6 (red), 5 (green), and 4 (blue). The Landsat band number was standardized to the Landsat-8 pattern for consistent comparison between the sensors.

Specific visual patterns were considered for each of the four forest classes. Undisturbed old-growth forests (UFs) exhibited consistently high NDVI and stable NBR values throughout the time series. Degraded old-growth forests (DFs), on the other hand, showed declines in these indices in response to events such as selective logging or fire. In some instances, index changes were subtle or absent, but visual inspection of spatial context, such as proximity to fragment edges (within 300 m), revealed degradation pressures. Second-growth forests displayed gradual recovery, marked by increasing NDVI and NBR values following deforestation events identified by initially low values in both indices. The age of second-growth forests was determined based on the year when deforestation was last observed. For instance, in Figure 2B, two events of vegetation clearing corresponded

to strong decreases in NBR values in 1988 and 2003, followed by subsequent vegetation regrowth. We also accounted for potential confounding factors, such as index drops caused by residual cloud cover. To confirm possible small-scale deforestation/degradation events, we also checked historical high-resolution images from Google Earth Pro when available. For the JAM site, we obtained geospatial data from the Annual Operative Plans for selective logging [33] to discriminate areas of degraded forest.

Furthermore, to assist in the visual identification of the forest disturbance classes and provide additional proof of the quality of our training samples, we inspected auxiliary data from three other disturbance-related maps: i. Primary humid tropical forest map for the year 2001, from Turubanova et al. [38]; ii. Brazilian secondary forest age map for the year 2017, from Silva Junior et al. [39]; iii. Map of global forest loss due to fire and other disturbance drivers for 2001–2019, from Tyukavina et al. [40]. Due to uncertainties and occasional discordant results from these data sources, the visual interpretation of Landsat images was used as the primary reference for determining the class of a sample.

To collect data for training and testing the separability of the forest classes using the machine learning classifiers, we allocated 50 samples per each of the 12 sites, totaling 600 samples in the Amazon biome. At each site, the 50 samples were randomly distributed along the remote sensing flight line, with a minimum distance of 100 m from each other, to avoid spatial autocorrelation. To capture the spatial variation of forest canopies within a stand, the sample unit chosen was a square plot of 0.25 ha (50 m × 50 m). Plots of 0.25 ha have an adequate size to represent the structural variability of tropical forests, as shown in previous studies [41,42]. Using the identification from the visual interpretation of Landsat time-series supported by auxiliary data, we detected 53 samples in the SF_{1–15yr} class, 41 in the SF_{16–32yr} class, 317 in the DF class, and 189 in the UF class (Table 2). The distribution of samples over the sites, as well as the information that supported the forest classes identification for the training and testing samples, can be consulted on the interactive map at <https://catherine-almeida.github.io/forestmap/>.

Table 2. Distribution of samples (0.25 ha) per site and disturbance class (50 samples per site × 12 sites = 600 samples) for the inspection of Landsat time series (1984–2017) and auxiliary data. The disturbance classes are younger second-growth forests (SF_{1–15yr}), older second-growth forests (SF_{16–32yr}), degraded old-growth forests (DFs), and undisturbed old-growth forests (UFs).

Site	SF _{1–15yr}	SF _{16–32yr}	DF	UF	Total
MAM	0	0	0	50	50
ZF2	0	0	0	50	50
DUC	3	18	0	29	50
AUT	18	4	21	7	50
TAP	10	4	31	5	50
SFX1	3	0	47	0	50
SFX2	2	0	48	0	50
PAR	8	10	32	0	50
JAM	0	0	35	15	50
ALF	1	1	18	30	50
FN1	8	1	41	0	50
FN2	0	3	44	3	50
Total	53	41	317	189	600

2.3. Airborne LiDAR Data

Airborne LiDAR data were obtained by the Trimble HARRIER 68i system between January 2016 and April 2017 as part of the project EBA (Estimation of Biomass in Amazon) [43]. The LiDAR sensor recorded multiple discrete returns with a small footprint (nearly 30 cm) and a minimum point density of four points.m⁻². This minimum point density is considered sufficient for capturing essential forest attributes, such as canopy height and biomass [44,45].

The raw point cloud was denoised with the lasnoise function of the LAStools software, version 171030 [46]. For the subsequent preprocessing steps, we used the functions GroundFilter, TINSurfaceCreate, Clipdata, and PolyClipData of the FUSION/LDV software [47]. In short, the ground points were filtered from the denoised point cloud and interpolated into a 1 m Digital Terrain Model (DTM), which was subtracted from point elevations to obtain the height above the ground of each point. The resulting normalized point clouds were clipped to the 600 samples to calculate the LiDAR metrics of each sample.

A total of 34 area-based LiDAR metrics were considered, including metrics related to height distribution (maximum, mean, percentiles of height, standard deviation, coefficient of variation, skewness, and kurtosis), canopy cover (Leaf Area Density in a specific height interval and proportion of first returns respective the number of all returns), structural complexity (Shannon and Simpson diversity indices), and topography (terrain roughness).

To obtain the metrics related to height and LAD (Leaf Area Density), we tested both using all returns and only the first returns. Since the two approaches produced highly correlated metrics ($r > 0.9$), we chose the metrics calculated only from the first returns, which have been considered more stable across different LiDAR acquisition settings [48]. Additionally, the first returns effectively capture the top-of-canopy structure, reducing noise from lower vegetation and enhancing measurement consistency for cross-site comparisons. This method also alleviates computational challenges by lowering data volume while maintaining essential structural attributes for forest assessments [45].

The height and LAD metrics also considered just the points above a 2 m height to avoid low vegetation points. The LAD profile was calculated with the LAD function of the lidR package [49] using a height bin of 2 m and the extinction coefficient k of 0.695 based on the study by Stark et al. [50] in central Amazon.

The HSCI and DSCI metrics are related to canopy structural complexity, based on the commonly used entropy measures of Shannon (H') and Simpson (D) indices, respectively [51]. These metrics were normalized between 0 and 1 by considering a fixed number of 30 height bins.

The terrain roughness was defined as the difference between the highest and lowest altitude in a 3×3 moving window [52]. To mitigate extreme localized roughness values, the altitude data were averaged from a 1 m DTM to a 10 m DTM. The 10 m spatial resolution was selected as it aligned with our objective of capturing local topographic variability within the sample areas. More details on LiDAR data processing and metrics calculation can be found in Almeida et al. [21].

2.4. Airborne HSI Data

Airborne HSI data were collected by the EBA project between September and October 2017 using the AISAfenix sensor (Specim, Spectral Imaging, Ltd., Oulu, Finland). The flight lines were oriented close to the N-S direction to reduce variations in viewing–illumination geometry. In addition, data acquisition was carried out on sunny days, between 10 a.m. and 1 p.m. (local time), with an average solar zenith angle of 30° and a standard deviation of 7° . The AISAfenix sensor provided images at a spatial resolution of 1 m in 361 bands covering the spectral range of 380–2500 nm. Of these, 87 bands in the VNIR (visible and near-infrared) had a bandwidth of approximately 6.8 nm, while the other 274 bands in the SWIR (shortwave infrared) had a bandwidth of about 5.7 nm. We reduced the total number of bands to 232 to remove the noise outside the range of 460–2330 nm and around the spectral intervals of 1400 and 1900 nm (major atmospheric water vapor absorptions). Surface reflectance images were obtained from the at-sensor radiance by using the Atmospheric/Topographic Correction for Airborne Imagery tool (ATCOR-4; version 6.3). The water vapor estimates were based on the 940 nm absorption feature. For geometric correction, we used the data provided by a GNSS (Global Navigation Satellite System) receiver onboard the aircraft.

We considered a total of 278 potential metrics from the HSI data: 232 reflectance bands; 30 vegetation indices (Table 3); 10 continuum-removal absorption band parameters; and

6 sub-pixel metrics derived from linear spectral mixture analysis. The continuum-removed features were characterized by the depth at the absorption center (Dc) and the width at the half depth (Wc) from five fixed absorption bands: chlorophyll band at 495 nm (461–536 nm); chlorophyll band at 670 nm (556–749 nm); leaf water band at 980 nm (893–1074 nm); leaf water band at 1200 nm (1097–1265 nm); and lignin-cellulose band at 2100 nm (2039–2199 nm). To generate the continuum-removed spectrum, the original reflectance was first smoothed for noise reduction using a Savitzky–Golay filter with a window size of five bands and a first polynomial order. Then, the smoothed reflectance values within each absorption band were divided by the values of a continuum line between the band edges [53].

Table 3. Hyperspectral vegetation indices from the AISAfenix surface reflectance data.

Abbr.	Vegetation Index	Equation	Reference
ARI1	Anthocyanin Reflectance Index 1	$(1/R_{549}) - (1/R_{701})$	[54]
ARI2	Anthocyanin Reflectance Index 2	$[(1/R_{549}) - (1/R_{701})] * R_{797}$	[54]
CAI	Cellulose Absorption Index	$0.5 (R_{2039} + R_{2199}) - R_{2100}$	[55]
CRI1	Carotenoid Reflectance Index 1	$(1/R_{515}) - (1/R_{549})$	[54]
CRI2	Carotenoid Reflectance Index 2	$(1/R_{515}) - (1/R_{701})$	[54]
D _{LAI}	Difference for Leaf Area Index	$R_{1724} - R_{969}$	[56]
DWSI1	Disease Water Stress Index 1	R_{797}/R_{1662}	[57]
DWSI2	Disease Water Stress Index 2	R_{1662}/R_{549}	[57]
DWSI3	Disease Water Stress Index 3	R_{1662}/R_{680}	[57]
DWSI4	Disease Water Stress Index 4	R_{549}/R_{680}	[57]
DWSI5	Disease Water Stress Index 5	$(R_{797} + R_{549})/(R_{1662} + R_{680})$	[57]
EVI	Enhanced Vegetation Index	$2.5 (R_{797} - R_{673})/(R_{797} + 6 R_{673} - 7.5 R_{474} + 1)$	[58]
GNDVI	Green Normalized Difference Vegetation Index	$(R_{797} - R_{549})/(R_{797} + R_{549})$	[59]
LWVI1	Leaf Water Vegetation Index 1	$(R_{1096} - R_{983})/(R_{1096} + R_{983})$	[60]
LWVI2	Leaf Water Vegetation Index 2	$(R_{1096} - R_{1204})/(R_{1096} + R_{1204})$	[60]
ND _{leaf}	Normalized Difference for Leaf Biomass	$(R_{2160} - R_{1540})/(R_{2160} + R_{1540})$	[56]
ND _{chl}	Normalized Difference for Leaf Chlorophyll	$(R_{927} - R_{708})/(R_{927} + R_{708})$	[56]
NDLI	Normalized Difference Lignin Index	$[\log(1/R_{1751}) - \log(1/R_{1679})]/[\log(1/R_{1751}) + \log(1/R_{1679})]$	[61]
NDNI	Normalized Difference Nitrogen Index	$[\log(1/R_{1512}) - \log(1/R_{1679})]/[\log(1/R_{1512}) + \log(1/R_{1679})]$	[61]
NDVI	Normalized Difference Vegetation Index	$(R_{797} - R_{680})/(R_{797} + R_{680})$	[36]
NDWI	Normalized Difference Water Index	$(R_{859} - R_{1237})/(R_{859} + R_{1237})$	[62]
PRI	Photochemical Reflectance Index	$(R_{529} - R_{570})/(R_{529} + R_{570})$	[63]
PSRI	Plant Senescence Reflectance Index	$(R_{680} - R_{502})/R_{749}$	[64]
PWI	Plant Water Index	R_{900}/R_{969}	[65]
REP	Red-Edge Position	$700 + 40 [(R_{re} - R_{701})/(R_{742} - R_{701})]$ $R_{re} = (R_{673} + R_{783})/2$	[66]
RVSI	Red-Edge Vegetation Stress Index	$[(R_{714} + R_{749})/2] - R_{735}$	[67]
SR	Simple Ratio	R_{797}/R_{680}	[68]
VI _{green}	Vegetation Index green	$(R_{549} - R_{680})/(R_{549} + R_{680})$	[69]
VOG1	Vogelmann Index 1	R_{742}/R_{721}	[70]
VOG2	Vogelmann Index 2	$(R_{735} - R_{749})/(R_{714} + R_{728})$	[70]

We used the linear spectral mixture analysis from the unmix function of the hsdar R package [71] to calculate the fractions of green vegetation (GV), shade, and non-photosynthetic vegetation/soil (NP) endmembers. NP represents a mixture of bright soils and non-photosynthetic vegetation since these components could not be easily distinguished from each other in the scenes. To select reference endmembers for GV and NP, we applied the minimum noise fraction (MNF) followed by the pixel purity index (PPI) technique in the ENVI software, version 5.3 (Harris Geospatial Solutions, Inc., Boulder, CO, USA). To pick the purest pixels at each site, the endmembers detected by the PPI were projected over an n-dimensional scatterplot. The final GV and NP endmembers were then obtained by averaging the purest pixels of all sites. For the shade endmember, a photometric shade with a uniform reflectance of zero was considered [72].

For each metric (reflectance, vegetation indices, absorption features, and endmember fractions), we averaged the pixel values within each of the 600 samples to calculate the plot-level metrics. For the shade endmember fraction, we also calculated the proportion of pixels with a shade fraction below 30% (S_{0_30}), between 30 and 60% (S_{30_60}), and above 60% (S_{60}).

2.5. Feature Selection and Importance

To maximize the information extracted from LiDAR and HSI data, we initially calculated a large set of metrics. However, this high data dimensionality might cause overfitting in the modeling process. To address this issue, we implemented a feature selection process to reduce the number of metrics and avoid redundancy, simplifying the dataset while preserving its informative value.

First, we eliminated highly correlated metrics using the *findCorrelation* function from the R package *caret* [73]. This function evaluates the absolute values of pairwise Pearson's correlations between metrics, and if two metrics have a correlation greater than 0.95, it removes the one with the largest mean absolute correlation. Even though some remote sensing metrics had skewed distribution, Pearson's correlation is recognized as robust against extreme violations of assumptions of normality [74], effectively fulfilling its purpose of reducing data redundancy.

Next, we inspected the metrics for linear dependencies with the *findLinearCombs* function from the same package. These two steps resulted in 20 LiDAR metrics and 42 HSI metrics (Figure 3 and Table 4) selected from the original set of variables (34 LiDAR metrics and 278 HSI metrics). These refined set of metrics were then used as predictors for the machine learning classification (RF, SGB, and SVM) of the four forest classes (UF, DF, SF_{1-15yr}, and SF_{16-32yr}) by considering three different scenarios of datasets: a LiDAR-only dataset (20 metrics), an HSI-only dataset (42 metrics), and the combination of both data sources (62 metrics).

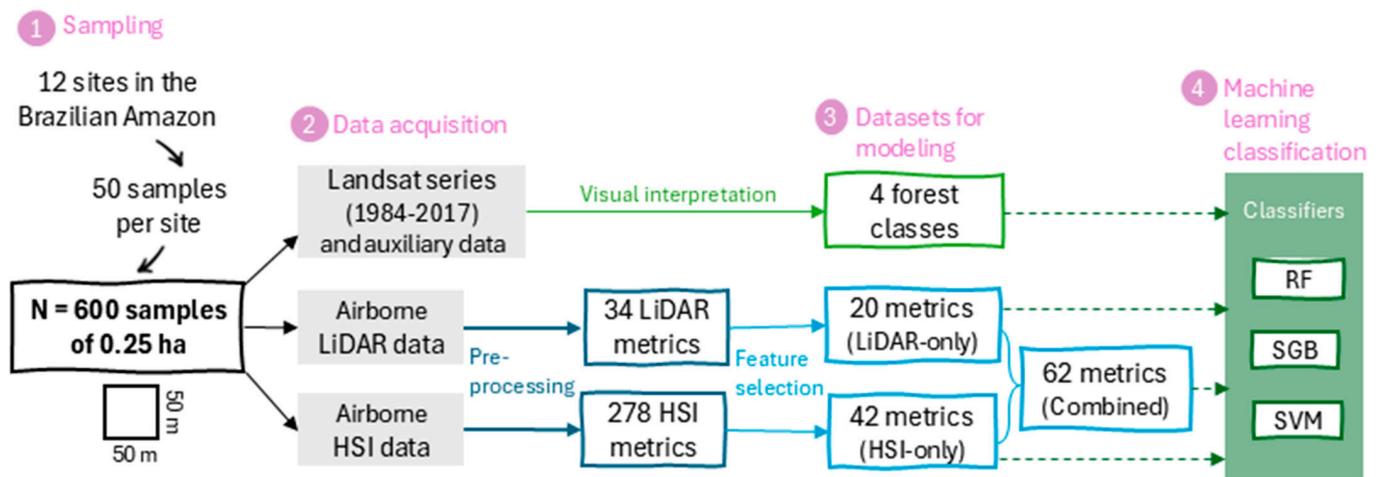


Figure 3. Flowchart representing the steps involved in dataset preparation for machine learning classification using three classifiers: Random Forest (RF), Stochastic Gradient Boosting (SGB), and Support Vector Machine (SVM).

Although the remaining number of metrics still produces a high-complexity model, which may be more difficult to apply in practice, our main purpose here was to understand the full potential of each data source. Mainly, we aimed to identify which metrics related to structural and functional characteristics were most important to differentiate the four classes of forests. Therefore, to further explore the discriminative power of metrics, we applied the Kruskal–Wallis test to evaluate differences in metric values across the forest

classes. We then calculated the eta squared based on the H statistic from the Kruskal–Wallis test (Equation (1)):

$$\eta^2[H] = \frac{H - k + 1}{n - k}, \quad (1)$$

where H is the value obtained in the Kruskal–Wallis test, k is the number of classes, and n is the total number of observations. The non-parametric Kruskal–Wallis test was chosen because some remote sensing metrics had skewed distribution, violating the assumptions of parametric methods. In this context, the eta squared indicates the proportion of total variation in the metric explained by the forest classes, serving as a univariate measure of metric importance. Additionally, we ranked the final set of selected metrics using the importance measures provided by the RF and SGB classifiers, further refining our understanding of which metrics were most informative for the classification task.

Table 4. Description of LiDAR and HSI metrics used as predictors to the machine learning classification models after the feature selection process.

Data Source	Metric Type	Selected Metrics	Description
LiDAR	Height statistics	H.max, H.mean, H.p05, H.p95, H.sd, H.cv, H.skew, H.kurt	Statistics (maximum, mean, 5% percentile, 95% percentile, standard deviation, coefficient of variation, skewness, and kurtosis, respectively) of the height distribution (H).
	Canopy cover	PD1st	Number of first returns above 2 m divided by the number of all returns above 2 m.
		LAD _{2_10} , LAD _{10_20} , LAD _{20_30}	LAD (Leaf Area Density in m ² m ⁻³) between the height intervals 2–10, 10–20, or 20–30 m.
		LAD ₂ , LAD ₆ , LAD ₁₄ , LAD ₂₂ , LAD ₂₆	LAD (Leaf Area Density in m ² m ⁻³) above the height 2, 6, 14, 22, or 26.
	Structural complexity indices	HSCI, DSCI	Shannon (H) and Simpson (D) Structural Complexity Indices calculated from the LAD profile.
	Topography	Roughness	Mean terrain roughness from 10 m DTM.
HSI	Reflectance bands	R ₄₆₁ , R ₅₄₉ , R ₆₇₃ , R ₈₅₂ , R ₁₁₈₁ , R ₁₇₃₅ , R ₂₁₄₉ , R ₂₂₆₅	Reflectance (R) at the respective wavelength band center.
	Vegetation indices	ARI1, ARI2, CAI, DLAI, DWSI2, DWSI3, DWSI4, DWSI5, LWVI1, LWVI2, ND _{Bleaf} , ND _{chl} , NDLI, NDNI, NDVI, NDWI, PRI, PSRI, PWI, REP, RVSI, SR	Vegetation indices calculated from the AISAfenix reflectance (see Table 3 for the full names and equations).
	Continuum-removal absorption features	D ₄₉₅ , D ₉₈₀ , D ₁₂₀₀ , D ₂₁₀₀	Depth (D) at the absorption centers of 495, 980, 1200, and 2100 nm, respectively.
		W ₄₉₅ , W ₆₇₀ , W ₉₈₀ , W ₁₂₀₀ , W ₂₁₀₀	Width at the half depth (W) at the absorption centers of 495, 670, 980, 1200, and 2100 nm, respectively.
	Sub-pixel fractions	NP	Fraction of non-photosynthetic vegetation/soil endmember.
	S _{30_60} , S ₆₀	Proportion of pixels with shade fraction (S) between 30–60% and above 60%, respectively.	

2.6. Model Validation, Optimization, and Generalization

The machine learning approach was used to investigate the potential of LiDAR and HSI metrics, used separately and together, for the discrimination of the four forest classes. To validate the machine learning models, we first split the dataset ($n = 600$ samples) into training and test sets, in which the test set was composed of a given site ($n = 50$) and the training set by the remaining 11 sites ($n = 550$). Thus, this procedure was repeated 12 times, always leaving one of the 12 sites out to serve as a test set. The training set results were used

to evaluate the models' performance in a regional context by considering different sites throughout the Amazon. In contrast, the test set results evaluated the model generalization capability, that is, its ability to accurately predict outcomes for a completely new site not included in the training data.

We used the train function of the caret package to train each classifier (RF, SGB, and SVM) in the training sets and calculate an unbiased performance measure via 11-fold cross-validation (each fold had approximately 50 samples for the model validation). For each classifier, we evaluated different tuning parameters to select the optimal model from those parameters. For the RF classifier, the mtry parameter was tuned from the values 2, 4, 6, 8, and 10, and the ntree parameter was set to 1000. For the SGB, tuning parameters were n.trees (50, 100, and 150) and interaction.depth (1, 2, and 3). The parameters shrinkage and n.minobsinnode were set to the default values (0.1 and 10, respectively). For SVM, we used the Radial Basis Function Kernel by tuning the parameters cost (0.5, 1, 2, and 4) and setting the sigma value with the sigest function from the R package kernlab [75].

We considered the overall F1 score (F1 average of the four classes) as the performance measure to select the optimal model. The F1 score combines *precision* (also known as user's accuracy) and *recall* (aka producer's accuracy or sensitivity) by calculating its harmonic mean (Equation (2)), thus providing a single performance measurement for a given class [76]:

$$F1 = \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}} \quad (2)$$

The overall F1 score is known to be better suited for imbalanced data because it gives the same weight to every class. However, to be able to evaluate the performance for each class, we also provided the confusion matrices with the *precision* and *recall* values of each class (see Tables S1–S9 in the Supplementary Material) and calculated the by-class F1. In addition to the overall and by-class F1, the overall accuracy (OA) was also reported. Even though OA tends to undervalue the performance of classifiers on smaller classes, this measure is widely used and may be useful for comparison among other studies.

A two-way analysis of variance (ANOVA), followed by a Tukey test, was used to assess whether there were any differences in performance measures from cross-validation (OA and overall/by-class F1) among the three datasets (LiDAR, HSI, and its combination), three classifiers (RF, SGB, and SVM), and interaction between data and classifiers. To examine the effect size of these factors (data source, classifier, and their interaction) on the overall model performance (OA and F1), we also calculated the eta squared (η^2). From the ANOVA results, the η^2 is the ratio of the sum of the squares of the factor by the total sum of squares and can be considered as a large effect size when greater than 0.14 [77].

Finally, we chose the best model (among data sources and classifiers) for each of the 12 training sets and used it to predict the forest classes on its corresponding test set. Thus, we calculated the overall accuracy and F1 for each test set, allowing us to analyze the ability of the best model to generalize to sites that were not used in its training.

All statistical analyses were performed in the R software, version 4.3.1, and considered a significance level of 0.05. The data processing and analysis were conducted on a computer equipped with an Intel Core i7-1255U processor (10 cores, 4.7 GHz) and 16 GB of RAM (Random Access Memory). For storage, a 4 TB external hard drive was used to store the datasets and results.

3. Results

3.1. Importance of LiDAR and HSI Metrics for Separating the Forest Classes

The LiDAR and HSI metrics that explained most of the forest classes' variability (highest η^2 [H]) are presented in Figure 4. Most of these metrics were also ranked as the most important for the RF and SGB models (Figure 5). Table 5 summarizes the main structural and functional characteristics associated with the LiDAR and HSI metrics useful for discriminating the forest disturbance classes.

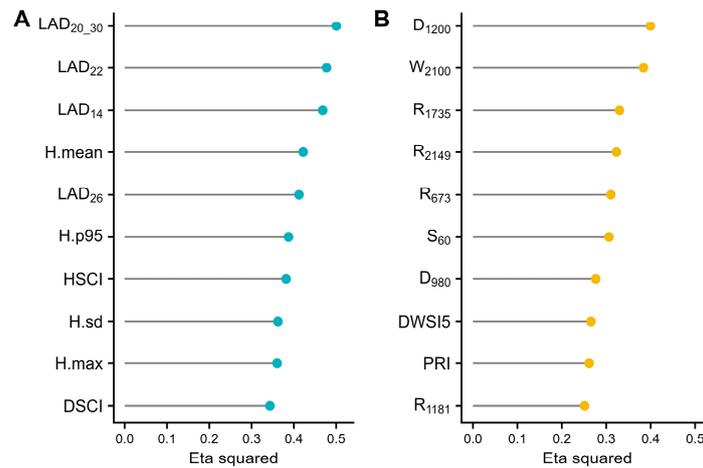


Figure 4. Top 10 LiDAR (A) and HSI (B) metrics ranked according to eta squared ($\eta^2[H]$). See Table 4 for the description of metrics.

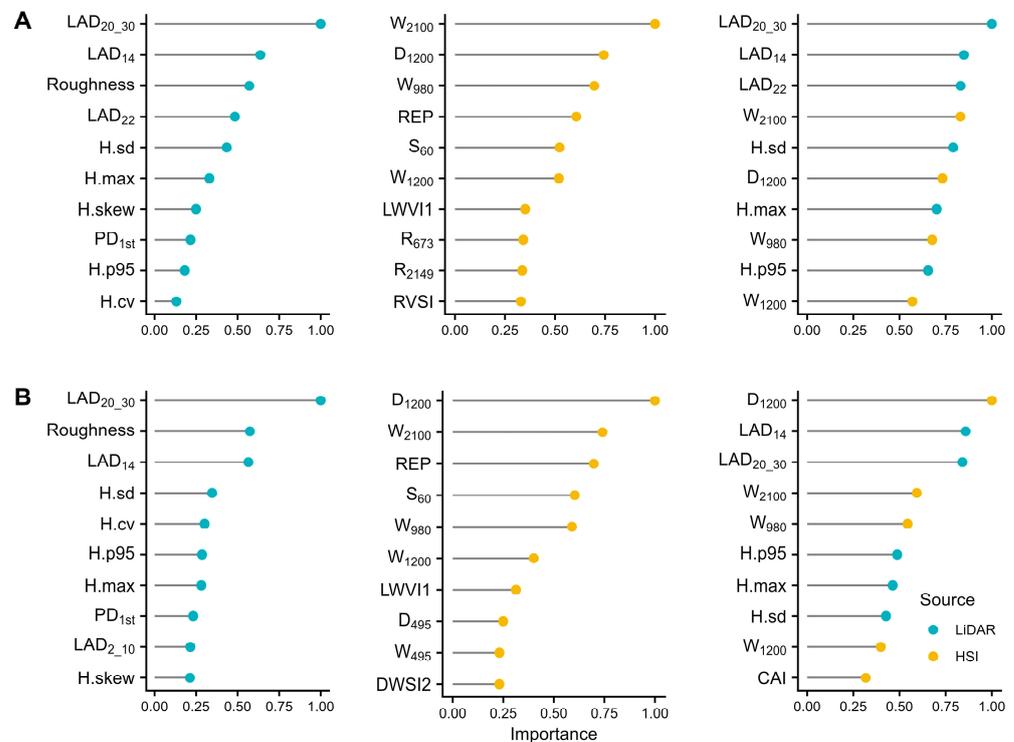


Figure 5. Relative importance of the 10 highest ranked metrics for the (A) RF and (B) SGB classifiers as a function of the LiDAR (left), HSI (center), and hybrid (right) datasets. See Table 4 for the description of metrics.

From the LiDAR dataset, metrics related to the upper canopy density (LAD_{20_30}, LAD₂₂, and LAD₁₄) were the most important for classification. The upper canopy density exhibited a significant increase from the SF_{1–15yr} to the UF (Table 5 and Figure 6). The LAD between the height interval of 20 and 30 m (LAD_{20_30}) presented the highest $\eta^2[H]$ (0.50) and high relative importance for the RF and SGB classifiers, either in single or hybrid models. Such height interval represents the canopy height distribution of old-growth forests (DF and UF), which showed a mean canopy height (H.mean) of around 20 m and a top of canopy height (H.p95) of around 30 m (Table 5). Very tall trees (>30 m) were rare in second-growth forests. When they occurred, such tall trees may be remnants of the native vegetation. Thus, metrics related to the top of canopy height, such as H.max and H.p95, showed good separability between second and old-growth forests. Canopy

structural complexity and heterogeneity, as measured by metrics such as HSCI and H.sd, were also good indicators of vegetation regrowth. Consequently, second-growth forests had homogeneous canopies with little height variation, while old-growth forests had heterogeneous and complex canopies, commonly presenting several vertical strata. The terrain roughness had the lowest $\eta^2[H]$ (0.03), indicating a poor ability to explain alone the variability between the disturbance classes. However, RF and SGB using only LiDAR data ranked this metric among the three most important. This result suggests that terrain roughness, when combined with other LiDAR variables, contributes to improving model performance by accounting for additional variability among forest classes.

Table 5. Comparison of the average structural and functional characteristics derived from LiDAR and HSI data among the forest classes. See Table 4 for the description of metrics. Distinct letters in a row indicate significant differences in the characteristics among classes from a pairwise Wilcoxon test with a Holm correction.

Characteristic	Metric	Unit	Class			
			SF _{1-15yr}	SF _{16-32yr}	DF	UF
LiDAR						
Upper canopy density	LAD _{20_30}	m ² m ⁻³	0.00 a	0.10 b	0.28 c	0.68 d
Mean canopy height	H.mean	m	6.69 a	12.10 b	16.68 c	21.51 d
Top of canopy height	H.p95	m	11.12 a	16.95 b	28.83 c	31.35 d
Structure complexity	HSCI	unitless	0.38 a	0.56 b	0.67 c	0.74 d
Canopy heterogeneity	H.sd	m	2.63 a	3.27 b	7.20 d	6.76 c
HSI						
Canopy moisture/LAI	D ₁₂₀₀	%	17.84 a	18.96 c	18.61 b	20.04 d
Canopy moisture/LAI and non-photosynthetic biochemical	W ₂₁₀₀	nm	51.41 a	59.41 b	52.46 a	61.43 b
Photosynthetic pigments	R ₆₇₃	%	2.71 a	2.03 c	2.23 b	1.77 d
Canopy gaps/emergent trees	S ₆₀	unitless	0.01 a	0.04 b	0.08 c	0.13 d
Health	DWSI5	unitless	1.74 a	2.04 c	1.91 b	2.09 c

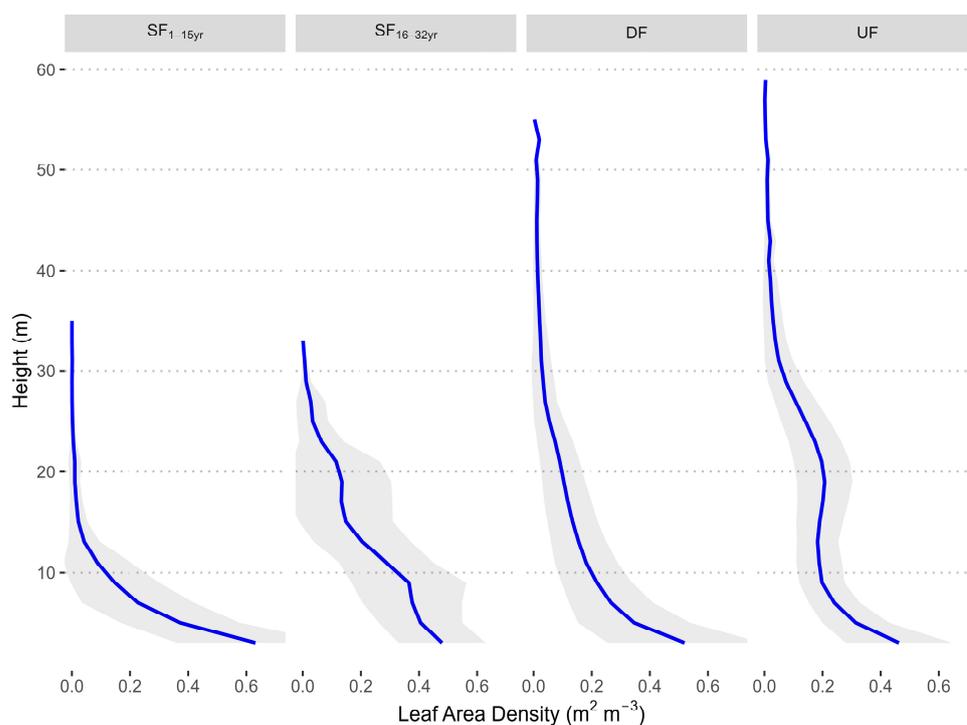


Figure 6. Leaf Area Density profile for each class of vegetation disturbance. Blue lines are mean values from all samples, while gray areas represent the standard deviation.

The HSI dataset provided better discrimination between DF and UF, as degraded forests were generally more spectrally similar to the older second-growth forests (Figure 7A). From the HSI reflectance bands, the most relevant spectral interval for discriminating vegetation classes ($\eta^2[H]$ values > 0.3) was the SWIR (1500–2260 nm), especially the bands located around 1735 nm and 2149 nm (Figure 7B). Red reflectance bands (close to 670 nm), related to chlorophyll absorption, were also of great importance ($\eta^2[H] \sim 0.3$) in characterizing forest degradation/regeneration. They were followed by leaf-water absorption bands in the NIR at 980 nm and, especially at 1200 nm and by photosynthetic pigment features in the blue-to-green spectral region (~ 500 nm). In contrast, the spectral transition from the red edge to the NIR (740 to 890 nm) showed the lowest $\eta^2[H]$ values. The greater relevance of the spectral regions associated with absorption by biochemical constituents was evidenced by the use of the continuum-removal technique. This approach enhances the vegetation absorption features of interest while reducing the interference of other factors, such as the effects of soil background and view-illumination geometry. Absorption features at 1200 nm (D_{1200}) and 2100 nm (W_{2100}) were the most relevant metrics, displaying the highest $\eta^2[H]$ values (0.40 and 0.38, respectively) and relative importance values for the RF and SGB classifiers among all HSI metrics.

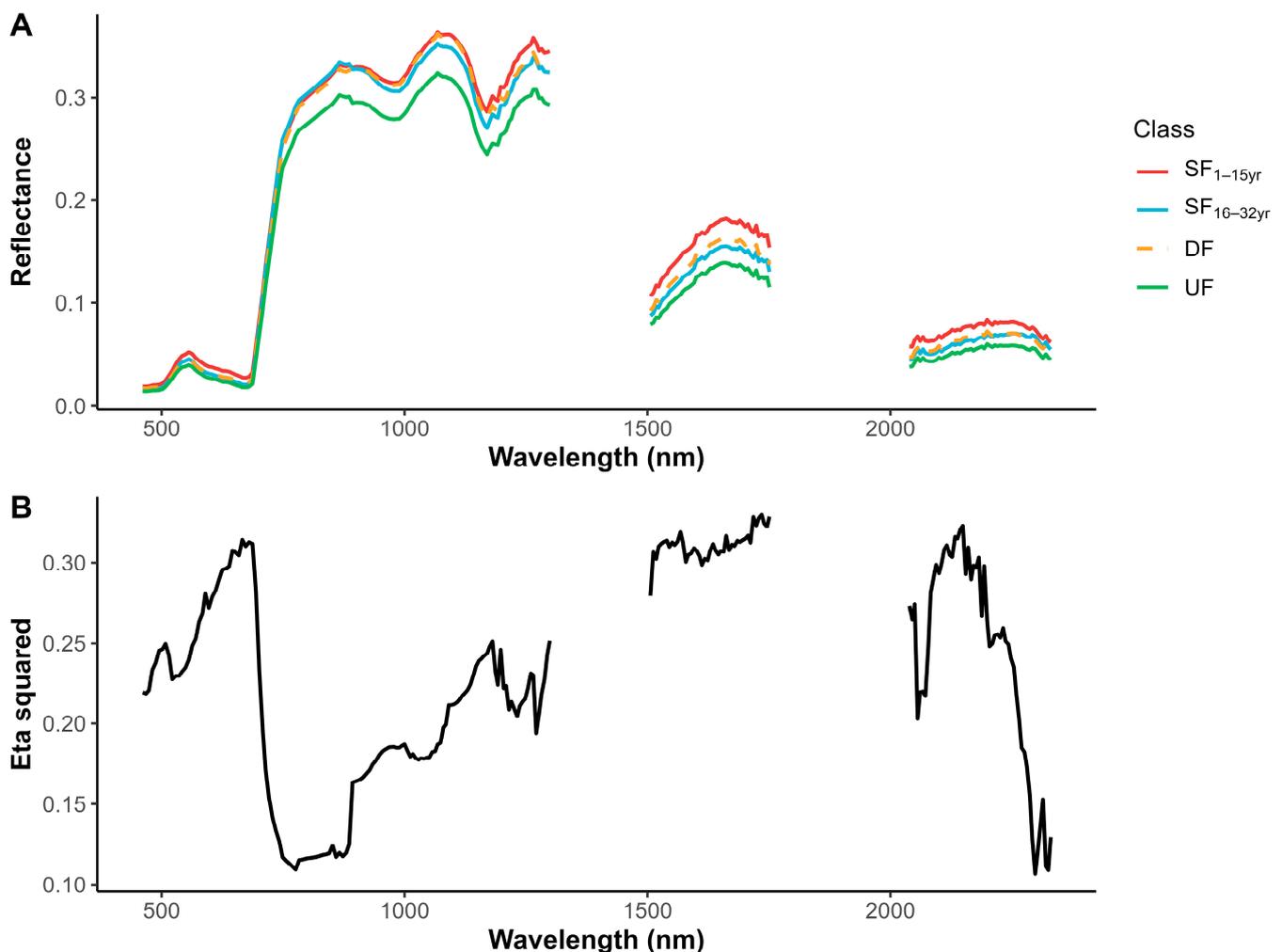


Figure 7. (A) Average AISA Fenix surface reflectance spectra of each studied class and (B) eta squared ($\eta^2[H]$) for all reflectance bands. Missing data around 1400 nm and 1900 nm refer to spectral intervals of strong atmospheric water vapor absorption.

In addition to the reflectance bands and absorption features, other HSI metrics were also relevant to distinguish forest classes. From the sub-pixel metrics, the proportion of

pixels with a shade fraction above 60% (S_{60}) had great importance, presenting a significantly different average among the four classes (Table 5). Deep shaded areas occur due to canopy gaps and shadowing of emergent trees, thus serving as a measure of canopy complexity. Furthermore, the increase in shade proportion (S_{60}), as well as in canopy moisture (evidenced by D_{1200}), from second- to old-growth forests explains the decrease in NIR and SWIR reflectance in the complex forests (Figure 7A). Among the vegetation indices, those with higher $\eta^2[H]$ (>0.2) were DWSI5, PRI, PSRI, ARI1, and DWSI4. REP ($\eta^2[H] = 0.20$) was also very important for the performance of the RF and SGB models using HSI data. The older second-growth forests were not significantly different from the undisturbed forests according to some HSI metrics, such as W_{2100} and DWSI5 (Table 5).

Differences in the variability of LiDAR and HSI metrics in relation to the discrimination of classes are useful to better characterize forests that are placed in the intersection between degradation and regeneration processes, such as degraded forests (class DF). LiDAR metrics of degraded old-growth forests generally had intermediate values between second-growth and undisturbed old-growth forests. Meanwhile, for some HSI metrics, degraded forests displayed intermediate values between younger and older second-growth forests (Figure 8).

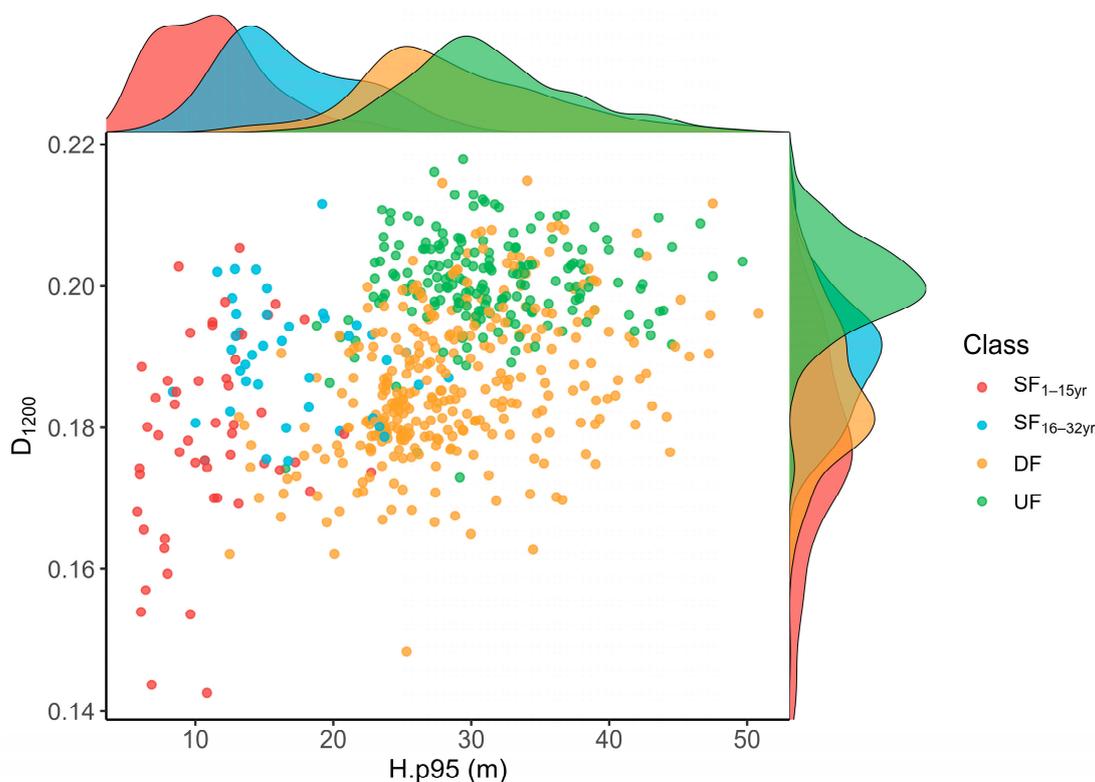


Figure 8. Scatterplot relating an HSI metric (D_{1200} = depth of the 1200 nm absorption band) to a LiDAR metric ($H.p95$ = 95% percentile of height). Points are colored according to the vegetation class. The density plot of each metric is also shown on the respective axis.

3.2. Effect of Remote Sensing Data Source and Machine Learning Classifiers

Figure 9 shows the results of the cross-validation (mean overall accuracy and overall F1) for each data source and classifier. Irrespective of the classifier used, the best performances were achieved with the use of multisource data for both evaluation metrics (accuracy mean cross-validated of 0.88–0.89 and F1 mean cross-validated of 0.82–0.83). The ANOVA results confirmed that the classification performance was mostly affected by the data source, which explained 85% of the OA variation and 81% of the F1 variation (Table 6). The effect of the classifiers was non-significant (p -value > 0.05), indicating that no single classifier consistently outperformed the others across all types of data sources used in the analysis. However, the interaction of data sources and classifiers was signif-

icant (p -value < 0.05) for the OA due to the poor performance of SVM when used with LiDAR data. Furthermore, the performance of the SVM was slightly better in discriminating the SF_{16–32yr} class (Table 7). This is an advantage of SVM, considering that the older second-growth forest was the most difficult class to be distinguished from the others (lower F1 score).

HSI-only models presented higher OA than LiDAR-only models due to the better performance of HSI data in the prevailing classes of DF and UF. In terms of overall F1, a more suitable metric for imbalanced datasets, the HSI-only models generally performed similarly to LiDAR-only models, except for the RF classifier. RF with LiDAR data displayed a significantly greater overall F1 than the RF with HSI data. LiDAR data generally produced better discrimination of second-growth forests compared with HSI data, especially for the SF_{1–15yr} class. The combination of the LiDAR's ability to better discriminate the two successional stages with the HSI's ability to better discriminate the disturbed and undisturbed old-growth forests increased the overall F1 of hybrid models by up to 12.2% of the best single model.

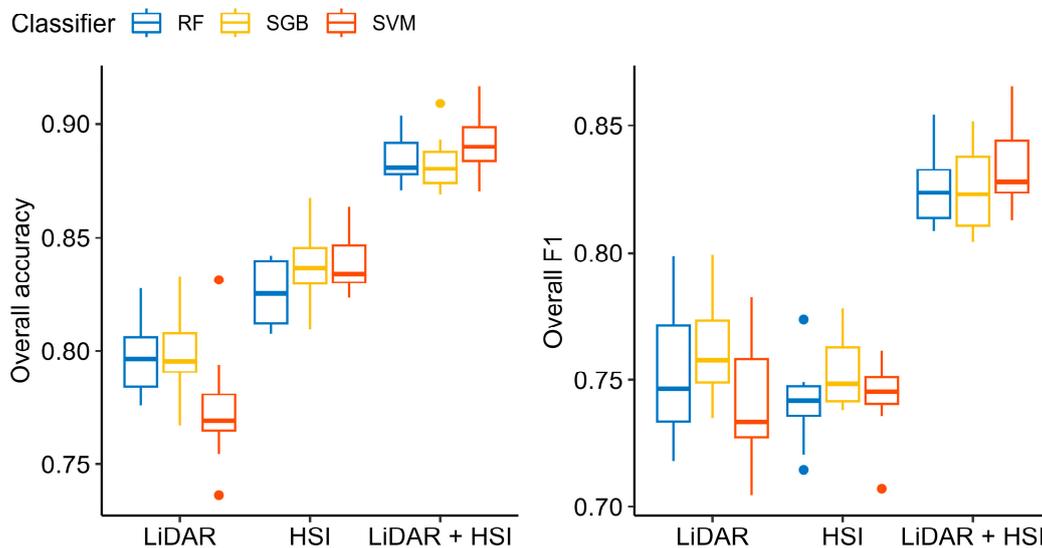


Figure 9. Variations in overall accuracy and F1 to discriminate the vegetation classes using different machine learning classifiers (RF = Random Forest, SGB = Stochastic Gradient Boosting, and SVM = Support Vector Machine) and data sources (LiDAR, HSI, and hybrid dataset).

Table 6. ANOVA results for the assessment of differences in performance measures to discriminate the vegetation classes.

Factor	Degree of Freedom	Sum of Squares	Mean Square	F Value	p -Value	η^2
Overall Accuracy						
Data	2	0.17	0.08	359.49	0.00	0.85
Classifier	2	0.00	0.00	1.14	0.32	0.00
Data:Classifier	4	0.01	0.00	6.31	0.00	0.03
Residuals	99	0.02	0.00			
Overall F1						
Data	2	0.15	0.08	233.75	0.00	0.81
Classifier	2	0.00	0.00	1.87	0.16	0.01
Data:Classifier	4	0.00	0.00	1.89	0.12	0.01
Residuals	99	0.03	0.00			

Table 7. Cross-validated overall and by-class performance for each data source and classifier. Distinct letters in a column indicate significant differences in performance from the Tukey test.

Data	Classifier	Overall Accuracy	Overall F1	By-Class F1			
				SF _{1–15yr}	SF _{16–32yr}	DF	UF
LiDAR	RF	0.80 a	0.75 a	0.77 ab	0.62 ab	0.84 a	0.76 a
	SGB	0.80 a	0.76 a	0.79 ab	0.65 abc	0.84 a	0.76 a
	SVM	0.77 b	0.74 a	0.75 a	0.67 acd	0.82 a	0.71 b
HSI	RF	0.82 c	0.74 a	0.62 c	0.61 b	0.87 b	0.83 c
	SGB	0.84 c	0.75 a	0.64 c	0.62 ab	0.88 b	0.85 cd
	SVM	0.84 c	0.74 a	0.61 c	0.63 ab	0.88 b	0.86 cde
LiDAR + HSI	RF	0.88 d	0.82 b	0.80 b	0.69 cd	0.91 c	0.89 de
	SGB	0.88 d	0.83 b	0.79 ab	0.71 de	0.91 c	0.88 de
	SVM	0.89 d	0.83 b	0.76 ab	0.75 e	0.92 c	0.90 e

3.3. Generalization of Predictions

For each training set (12 replicates, each without a given site), we selected the models with the highest cross-validated overall F1 (Table 8). In all cases, the best data were the combination of LiDAR + HSI. For the classifier, the SVM appeared with the highest F1 for most cases, although the performance difference between the classifiers was not significant. When we used these best models to predict the classes of the site that was left out, we found that the OA in the test set ranged from 0.32 to 0.92, and the F1 ranged from 0.39 to 0.93. The results indicate that, for most sites, models trained in other locations were able to generalize to this new environment, maintaining high classification performance. More specifically, the sites DUC, FN1, FN2, PAR, SFX1, SFX2, and ZF2 had a test set OA ≥ 0.78 and a test set F1 ≥ 0.7 . However, some sites (especially JAM and MAM) performed poorly in the test set, evidencing that the models were not able to generalize to these sites.

Table 8. Overall performance of the best models in training and test sets. The test set only contains data for a specific site, which was not used in model training.

Test Set Site	Best Classifier	Best Data	Training Set Performance		Test Set Performance	
			Overall Accuracy	Overall F1	Overall Accuracy	Overall F1
ALF	SVM	LiDAR + HSI	0.90	0.84	0.56	0.67
AUT	RF	LiDAR + HSI	0.89	0.84	0.64	0.59
DUC	SGB	LiDAR + HSI	0.88	0.85	0.80	0.70
FN1	RF	LiDAR + HSI	0.88	0.82	0.92	0.86
FN2	SGB	LiDAR + HSI	0.89	0.84	0.86	0.93
JAM	SVM	LiDAR + HSI	0.92	0.87	0.40	0.39
MAM	SVM	LiDAR + HSI	0.89	0.83	0.32	0.48
PAR	RF	LiDAR + HSI	0.90	0.85	0.80	0.82
SFX1	SVM	LiDAR + HSI	0.90	0.86	0.88	0.72
SFX2	SVM	LiDAR + HSI	0.88	0.82	0.90	0.81
TAP	SVM	LiDAR + HSI	0.90	0.84	0.72	0.65
ZF2	SVM	LiDAR + HSI	0.88	0.83	0.78	0.88

4. Discussion

The present study brings relevant insights into how advanced remote sensing technologies can be used together to improve our understanding of forest dynamics concerning anthropogenic disturbances and natural regeneration in tropical forests of the Amazon region. It goes beyond simpler evaluations of forest cover loss and advances toward the assessment of forest quality and ecosystem services in heterogeneous human-modified tropical landscapes. Despite the high complementary potential of LiDAR and HSI data for improving the characterization and quantification of forest characteristics, previous studies displayed divergent results on the synergistic use of both technologies. Some investigations showed no or slight information gain when combining structural (LiDAR) and

spectral (HSI) data [72,78,79], while others demonstrated a significant effect on increasing the performance of models [22,80,81].

In our study, the complementary information of LiDAR and HSI data considerably improved the characterization of tropical forest degradation and regeneration. While LiDAR performed well in classifying successional stages by differentiating them from old-growth forests, HSI was effective in distinguishing degraded forests from undisturbed forests. Canopy structural characteristics related to LiDAR metrics, such as height, basal area, and biomass, have been used to characterize successional stages [82]. For instance, the most important LiDAR metrics found here were similar to the ones used to estimate aboveground carbon density in Borneo's tropical forests [83]. Examples include canopy cover at 20 m aboveground (Cover20), which was similar to the metrics LAD_{20_30} or LAD₂₂ used here, and top of canopy height (TCH), which was related to the metrics H.max or H.p95 used in this study. Additionally, the HSI and LiDAR metrics that had the greatest influence on the classification of forest degradation and regeneration were also the most important for estimating AGB in a previous study in the Amazon biome [21]. This is because AGB integrates important forest structural and functional information associated with forest disturbance and regrowth, such as tree height, basal area, number of trees per area, and wood density.

The main errors related to our LiDAR-only models expressed the confusion between degraded and undisturbed old-growth forests. This challenge arises because some degraded areas, while structurally similar to undisturbed forests, may experience changes in species composition. For instance, disturbances like timber harvesting and forest fires may reduce the abundance of certain tree species and favor the regeneration of early-successional species in forest gaps and edges [84], which have contrasting functional and spectral features [5,85]. While LiDAR effectively captures canopy structure and height, it lacks the ability to detect changes in species composition, which can be a crucial indicator of forest health and disturbance. Thus, approaches based solely on structural characteristics limit the characterization of a broad spectrum of forest disturbance conditions.

In contrast, the most important HSI metrics found in this study were related to functional characteristics, especially canopy moisture, biochemical components, and health. The SWIR spectral region, especially the absorption feature around 2100 nm, was very relevant for characterizing forest degradation/regeneration. In consonance with our results, other studies have indicated that SWIR bands contain most of the relevant information to distinguish forest regeneration [5,86]. This fact can be explained by the increased canopy complexity, shadowing, and moisture along the succession, which decrease the SWIR reflectance. Furthermore, absorption features around 1700 nm and 2100 nm have been related to non-pigment biochemical components, such as lignin and cellulose [13], indicating the occurrence of dead or senescent vegetation. Water absorption bands, especially at 1200 nm, were also very important. Asner et al. [87], using EO-1 Hyperion data in the central Amazon, showed that the canopy water metrics (SWAM, spectroscopic water-absorption metric) and pigment metrics (PRI and ARI) were a proxy for physiological and biochemical changes from chronic water stress. In agreement with our results, Thenkabail et al. [17] found that EO-1 Hyperion bands related to absorption by biochemical constituents (spectral intervals in the 1300–1900 nm, 1100–1300 nm, 1900–2350 nm, and 600–700 nm) were the most important to characterize African tropical forests following anthropogenic disturbance of different magnitudes. Thus, the importance of HSI metrics in identifying tropical disturbance suggests a greater susceptibility to canopy stress in degraded forests, facilitating their distinction from healthy forests with similar structural characteristics.

It is important to note that among the tested HSI metrics, the most important were the absorption features obtained through the continuum-removal method. These features, unlike reflectance values and some vegetation indices, cannot be derived from multispectral data due to the need for high spectral resolution to detect specific absorption bands associated with biochemical and physiological properties, such as pigments and water content in leaves. Consequently, our study points to a clear advantage of hyperspectral data over

multispectral data in this context. This finding aligns with those reported by Thenkabail et al. [17], who compared three broadband sensors (IKONOS, ALI, and ETM+) to the narrowband hyperspectral Hyperion data for classifying complex rainforest vegetation. Their results indicated that forest classifications using hyperspectral data achieved overall accuracies 45–52% higher than those using multispectral data. Similarly, for non-forest habitats (meadows, grasslands, heaths, and mires), Jarocińska et al. [88] found that HSI outperformed multispectral Sentinel-2 imagery, with greater improvements in classification accuracy observed in areas of high α or β -diversity.

Although the combination of LiDAR and HSI offers significant advantages for characterizing the complex dynamics of regeneration and degradation in tropical forests, future studies should assess the cost-effectiveness of this approach using airborne data. To expand the analytical scope in larger areas, researchers should consider recently launched hyperspectral missions such as the Environmental Mapping and Analysis Program (EnMAP) and the possibility of integrating their observations with data acquired by existing Synthetic Aperture Radar (SAR) or future LiDAR missions. This strategy would facilitate the development of approaches combining structural and compositional information for large-scale monitoring initiatives.

Most studies on forest ecosystems based on LiDAR and HSI data integration have been performed locally on single study sites [89–92]. Although it is important to consider different spatial scales in the study of multisensor data integration, the full realization of its potential as a source of forest information requires an ability to generalize in different environmental conditions and human-induced degradation and regeneration dynamics. Even considering that our best model cannot be generalized to all evaluated sites, it was satisfactorily transferred to most sites that represent the distinct environmental and anthropogenic conditions of the Amazon. Thus, the integrated use of LiDAR and HSI data can also help to understand the dynamics of complex Amazonian forests from a regional perspective.

Second-growth and degraded forests are an integral part of tropical landscapes. However, they present different compositions and structures, leading to divergent functioning patterns. Therefore, their accurate characterization and discrimination from the remaining undisturbed forests are essential for establishing conservation and management priorities. The distinct structural and functional characteristics of undisturbed forests suggest that some of their ecosystem services cannot be replaced by degraded or secondary forests [93]. Therefore, it is necessary to conserve forests that are still relatively intact, preventing new areas from degradation or deforestation. The remote sensing approaches described here may play a key role in planning interventions under REDD, intact forest landscapes, and the Alliance for the Restoration in the Amazon [94,95].

Previous studies [96–98] have reported a “secondarization” of degraded forests, described as a process that transforms closed-canopy primary forests into more open forests dominated by short-lived pioneer species due to recurrent anthropogenic disturbances. In fact, older second-growth forests and degraded old-growth forests may represent forests with similar structural and compositional characteristics positioned at the intersection of degradation and regeneration trajectories. However, depending on the degradation intensity and recurrence, degraded forests can retain important structural characteristics of the former primary forests, as well as a generally heterogeneous species composition [99]. Likewise, the older second-growth forests had distinct characteristics from the younger ones and, according to some HSI metrics, were more similar to the undisturbed forests. Thus, both degraded and secondary forests have great potential to provide significant environmental benefits, as well as to contribute to poverty alleviation through products and services of socio-economic importance. However, avoiding recurrent degradation is essential to ensure the continued functioning of forests.

Finally, as far as we know, this is the first study that evaluates the potential of combining LiDAR and hyperspectral remote sensing to discriminate different levels of forest degradation and regeneration over the Brazilian Amazon. This knowledge is important for

future large-swath orbital missions of both instruments, even considering the constraints of data upscaling from the airborne to the orbital level of data acquisition. Understanding the variability of LiDAR and HSI metrics in different degradation and regeneration gradients from a wide spatial sampling over the Amazon forests can help to build models based on equivalent metrics derived from orbital platforms in an upscaling exercise. Thus, the role played by data sources, metrics, and models described in this study represents the first step toward the production of large-scale maps to be further validated with detailed field information in the Amazon.

5. Conclusions

We concluded that the combined use of LiDAR and HSI data was more effective than the type of machine learning classifier in improving discrimination between tropical forests with different degradation and regeneration status. Models based on a single remote sensing data presented a reasonable overall performance (F1 of 0.74–0.76 for LiDAR models and 0.74–0.75 for HSI models) but displayed superior accuracy in specific classes. While LiDAR produced significantly fewer errors for discriminating second-growth from old-growth forests, HSI performed significantly better than LiDAR for separating degraded from undisturbed old-growth forests. Combining the strengths of each data source significantly improved the classification performance, increasing the overall F1 by up to 12.2% relative to the best single model. When evaluating the best models in the test set, composed of a site left out in the training process, we found that the models were able to generalize to most sites, although they performed poorly in some specific environments. These results underscore the importance of multisource remote sensing for tropical forest characterization, advancing the assessment of forest degradation and regeneration processes in the Amazon. Future research should continue to explore the integration of advanced remote sensing technologies to further enhance conservation, sustainable management, and restoration policies.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16213935/s1>, Table S1: Cross-validated (11 folds, repeated 12 times) confusion matrix for the RF classifier with LiDAR data (entries are aggregated counts across resamples and repetitions), Table S2: Cross-validated (11 folds, repeated 12 times) confusion matrix for the RF classifier with HSI data (entries are aggregated counts across resamples and repetitions), Table S3: Cross-validated (11 folds, repeated 12 times) confusion matrix for the RF classifier with both LiDAR and HSI data (entries are aggregated counts across resamples and repetitions), Table S4: Cross-validated (11 folds, repeated 12 times) confusion matrix for the SGB classifier with LiDAR data (entries are aggregated counts across resamples and repetitions), Table S5: Cross-validated (11 folds, repeated 12 times) confusion matrix for the SGB classifier with HSI data (entries are aggregated counts across resamples and repetitions), Table S6: Cross-validated (11 folds, repeated 12 times) confusion matrix for the SGB classifier with both LiDAR and HSI data (entries are aggregated counts across resamples and repetitions), Table S7: Cross-validated (11 folds, repeated 12 times) confusion matrix for the SVM classifier with LiDAR data (entries are aggregated counts across resamples and repetitions), Table S8: Cross-validated (11 folds, repeated 12 times) confusion matrix for the SVM classifier with HSI data (entries are aggregated counts across resamples and repetitions), Table S9: Cross-validated (11 folds, repeated 12 times) confusion matrix for the SVM classifier with both LiDAR and HSI data (entries are aggregated counts across resamples and repetitions).

Author Contributions: Conceptualization, C.T.d.A., L.S.G. and L.E.O.e.C.d.A.; methodology, C.T.d.A., L.S.G., C.H.L.S.-J. and L.E.O.e.C.d.A.; software, C.T.d.A. and C.H.L.S.-J.; validation, C.T.d.A.; formal analysis, C.T.d.A.; investigation, C.T.d.A., L.S.G. and L.E.O.e.C.d.A.; resources, L.S.G., L.E.O.e.C.d.A. and P.H.S.B.; data curation, C.T.d.A., L.S.G., L.E.O.e.C.d.A., A.D.J., F.R.d.S.P., L.Y.S. and J.P.H.B.O.; writing—original draft preparation, C.T.d.A.; visualization, C.T.d.A.; supervision, L.S.G., L.E.O.e.C.d.A. and P.H.S.B.; project administration, L.S.G., L.E.O.e.C.d.A. and J.P.H.B.O.; funding acquisition, L.S.G., L.E.O.e.C.d.A. and J.P.H.B.O. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) [grant numbers 140502/2016-5, 301486/2017-4, and 141035/2021-8], FAPESP

(Fundação de Amparo à Pesquisa do Estado de São Paulo) [grant number 2017/22269-2, 2018/18416-2, and 2020/06734-0], and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil) [Finance Code 001]. C.H.L.S.-J was supported by the University of Manchester through the “Forest fragmentation mapping of Amazon and its vulnerable margin Amazon-Cerrado transition forests” project, the National Council for Scientific and Technological Development—CNPq (grant 401741/2023-0). L.E.O.C.A (grant numbers 305054/2016-3 and 314416/2020-0) and L.S.G. (307792/2021-8) were supported by CNPq. L.E.O.C.A is also grateful to the FAPESP ARBOLES project [grant number 18/15001-6]. The collection of airborne LiDAR and hyperspectral data was financed by the project Environmental Satellite Monitoring in the Amazon Biome (MSA-BNDES)—Activity 7 “Improvement of biomass estimation methods and emission estimation models for change of land use”, funded by Amazon Fund (process #14209291).

Data Availability Statement: Airborne LiDAR data from the EBA project are available at: <https://zenodo.org/records/4552699>, <https://zenodo.org/records/7689909>, and <https://zenodo.org/records/7636454>. The distribution of samples over the sites as well as the information that supported the forest classes identification for the training and testing samples can be consulted at <https://catherine-almeida.github.io/forestmap/>.

Acknowledgments: We greatly appreciate the time and effort that the academic editor and reviewers have invested in providing constructive feedback on this manuscript.

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

References

- Griscom, B.W.; Adams, J.; Ellis, P.W.; Houghton, R.A.; Lomax, G.; Miteva, D.A.; Schlesinger, W.H.; Shoch, D.; Siikamäki, J.V.; Smith, P.; et al. Natural Climate Solutions. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11645–11650. [CrossRef] [PubMed]
- Brancalion, P.H.S.; Holl, K.D. Guidance for Successful Tree Planting Initiatives. *J. Appl. Ecol.* **2020**, *57*, 2349–2361. [CrossRef]
- Wilson, S.J.; Schelhas, J.; Grau, R.; Nanni, A.S.; Sloan, S. Forest Ecosystem-Service Transitions: The Ecological Dimensions of the Forest Transition. *Ecol. Soc.* **2017**, *22*, 38. [CrossRef]
- Frolking, S.; Palace, M.W.; Clark, D.B.; Chambers, J.Q.; Shugart, H.H.; Hurtt, G.C. Forest Disturbance and Recovery: A General Review in the Context of Spaceborne Remote Sensing of Impacts on Aboveground Biomass and Canopy Structure. *J. Geophys. Res. Biogeosci.* **2009**, *114*, G00E02. [CrossRef]
- Vieira, I.C.G.; De Almeida, A.S.; Davidson, E.A.; Stone, T.A.; Reis De Carvalho, C.J.; Guerrero, J.B. Classifying Successional Forests Using Landsat Spectral Properties and Ecological Characteristics in Eastern Amazônia. *Remote Sens. Environ.* **2003**, *87*, 470–481. [CrossRef]
- Da Silva, R.D.; Galvão, L.S.; Dos Santos, J.R.; Camila, C.V.; De Moura, Y.M. Spectral/Textural Attributes from ALI/EO-1 for Mapping Primary and Secondary Tropical Forests and Studying the Relationships with Biophysical Parameters. *GIScience Remote Sens.* **2014**, *51*, 677–694. [CrossRef]
- Goetz, A.F.H.; Vane, G.; Solomon, J.E.; Rock, B.N. Imaging Spectrometry for Earth Remote Sensing. *Science* **1985**, *228*, 1147–1153. [CrossRef]
- Galvão, L.S.; Ponzoni, F.J.; Liesenberg, V.; Santos, J.R. dos Possibilities of Discriminating Tropical Secondary Succession in Amazônia Using Hyperspectral and Multiangular CHRIS/PROBA Data. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 8–14. [CrossRef]
- Féret, J.B.; Asner, G.P. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 73–84. [CrossRef]
- Baldeck, C.A.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Knapp, D.E.; Kellner, J.R.; Wright, S.J. Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS ONE* **2015**, *10*, e0118403. [CrossRef]
- Clark, M.L.; Kilham, N.E. Mapping of Land Cover in Northern California with Simulated Hyperspectral Satellite Imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 228–245. [CrossRef]
- Sanches, I.D.A.; Souza Filho, C.R.; Kokaly, R.F. Spectroscopic Remote Sensing of Plant Stress at Leaf and Canopy Levels Using the Chlorophyll 680nm Absorption Feature with Continuum Removal. *ISPRS J. Photogramm. Remote Sens.* **2014**, *97*, 111–122. [CrossRef]
- Kokaly, R.F.; Asner, G.P.; Ollinger, S.V.; Martin, M.E.; Wessman, C.A. Characterizing Canopy Biochemistry from Imaging Spectroscopy and Its Application to Ecosystem Studies. *Remote Sens. Environ.* **2009**, *113*, S78–S91. [CrossRef]
- Schepers, L.; Haest, B.; Veraverbeke, S.; Spanhove, T.; Borre, J.V.; Goossens, R. Burned Area Detection and Burn Severity Assessment of a Heathland Fire in Belgium Using Airborne Imaging Spectroscopy (APEX). *Remote Sens.* **2014**, *6*, 1803–1826. [CrossRef]
- de Moura, Y.M.; Galvão, L.S.; Hilker, T.; Wu, J.; Saleska, S.; do Amaral, C.H.; Nelson, B.W.; Lopes, A.P.; Wiedeman, K.K.; Prohaska, N.; et al. Spectral Analysis of Amazon Canopy Phenology during the Dry Season Using a Tower Hyperspectral Camera and Modis Observations. *ISPRS J. Photogramm. Remote Sens.* **2017**, *131*, 52–64. [CrossRef]

16. Galvão, L.S.; dos Santos, J.R.; Roberts, D.A.; Breunig, F.M.; Toomey, M.; de Moura, Y.M. On Intra-Annual EVI Variability in the Dry Season of Tropical Forest: A Case Study with MODIS and Hyperspectral Data. *Remote Sens. Environ.* **2011**, *115*, 2350–2359. [CrossRef]
17. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; Legg, C.; De Dieu, M.J. Hyperion, IKONOS, ALI, and ETM+ Sensors in the Study of African Rainforests. *Remote Sens. Environ.* **2004**, *90*, 23–43. [CrossRef]
18. Lefsky, M.A.; Cohen, W.B.; Harding, D.J.; Parker, G.G.; Acker, S.A.; Gower, S.T. Lidar Remote Sensing of Above-Ground Biomass in Three Biomes. *Glob. Ecol. Biogeogr.* **2002**, *11*, 393–399. [CrossRef]
19. Bispo, P.D.C.; Pardini, M.; Papathanassiou, K.P.; Kugler, F.; Balzter, H.; Rains, D.; dos Santos, J.R.; Rizaev, I.G.; Tansey, K.; dos Santos, M.N.; et al. Mapping Forest Successional Stages in the Brazilian Amazon Using Forest Heights Derived from TanDEM-X SAR Interferometry. *Remote Sens. Environ.* **2019**, *232*, 111194. [CrossRef]
20. Castillo, M.; Rivard, B.; Sánchez-Azofeifa, A.; Calvo-Alvarado, J.; Dubayah, R. LIDAR Remote Sensing for Secondary Tropical Dry Forest Identification. *Remote Sens. Environ.* **2012**, *121*, 132–143. [CrossRef]
21. Almeida, C.T.; Galvão, L.S.; Ometto, J.P.H.B.; Jacon, A.D.; de Souza Pereira, F.R.; Sato, L.Y.; Lopes, A.P.; Graça, P.M.L.d.A.; Silva, C.V.d.J.; Ferreira-Ferreira, J.; et al. Combining LiDAR and Hyperspectral Data for Aboveground Biomass Modeling in the Brazilian Amazon Using Different Regression Algorithms. *Remote Sens. Environ.* **2019**, *232*, 111323. [CrossRef]
22. Sun, C.; Cao, S.; Sanchez-Azofeifa, G.A. Mapping Tropical Dry Forest Age Using Airborne Waveform LiDAR and Hyperspectral Metrics. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *83*, 101908. [CrossRef]
23. Strand, J.; Soares-Filho, B.; Costa, M.H.; Oliveira, U.; Ribeiro, S.C.; Pires, G.F.; Oliveira, A.; Rajão, R.; May, P.; van der Hoff, R.; et al. Spatially Explicit Valuation of the Brazilian Amazon Forest's Ecosystem Services. *Nat. Sustain.* **2018**, *1*, 657–664. [CrossRef]
24. Alencar, A.; Shimbo, J.Z.; Lenti, F.; Marques, C.B.; Zimbres, B.; Rosa, M.; Arruda, V.; Castro, I.; Fernandes, M.; Alencar, I.; et al. Mapping Three Decades of Changes in the Brazilian Savanna Native Vegetation Using Landsat Data Processed in the Google Earth Engine Platform. *Remote Sens.* **2020**, *12*, 924. [CrossRef]
25. Asner, G.P.; Rudel, T.K.; Aide, T.M.; Defries, R.; Emerson, R. A Contemporary Assessment of Change in Humid Tropical Forests. *Conserv. Biol.* **2009**, *23*, 1386–1395. [CrossRef]
26. Tyukavina, A.; Hansen, M.C.; Potapov, P.V.; Krylov, A.M.; Goetz, S.J. Pan-Tropical Hinterland Forests: Mapping Minimally Disturbed Forests. *Glob. Ecol. Biogeogr.* **2016**, *25*, 151–163. [CrossRef]
27. Jacon, A.D.; Galvão, L.S.; Martins-Neto, R.P.; Crespo-Peremarch, P.; Aragão, L.E.O.C.; Ometto, J.P.; Anderson, L.O.; Vedovato, L.B.; Silva-Junior, C.H.L.; Lopes, A.P.; et al. Characterizing Canopy Structure Variability in Amazonian Secondary Successions with Full-Waveform Airborne LiDAR. *Remote Sens.* **2024**, *16*, 2085. [CrossRef]
28. Longo, M.; Keller, M.; dos-Santos, M.N.; Leitold, V.; Pinagé, E.R.; Baccini, A.; Saatchi, S.; Nogueira, E.M.; Batistella, M.; Morton, D.C. Aboveground Biomass Variability across Intact and Degraded Forests in the Brazilian Amazon. *Glob. Biogeochem. Cycles* **2016**, *30*, 1639–1660. [CrossRef]
29. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
30. Quesada, C.A.; Lloyd, J.; Anderson, L.O.; Fyllas, N.M.; Schwarz, M.; Czimczik, C.I. Soils of Amazonia with Particular Reference to the RAINFOR Sites. *Biogeosciences* **2011**, *8*, 1415–1440. [CrossRef]
31. Almeida, C.T.; Oliveira-Júnior, J.F.; Delgado, R.C.; Cubo, P.; Ramos, M.C. Spatiotemporal Rainfall and Temperature Trends throughout the Brazilian Legal Amazon, 1973–2013. *Int. J. Climatol.* **2017**, *37*, 2013–2026. [CrossRef]
32. Pontes-Lopes, A.; Silva, C.V.J.; Barlow, J.; Rincón, L.M.; Campanharo, W.A.; Nunes, C.A.; de Almeida, C.T.; Silva Júnior, C.H.L.; Cassol, H.L.G.; Dalagnol, R.; et al. Drought-Driven Wildfire Impacts on Structure and Dynamics in a Wet Central Amazonian Forest. *Proc. R. Soc. B Biol. Sci.* **2021**, *288*, 20210094. [CrossRef] [PubMed]
33. SFB—Serviço Florestal Brasileiro Madeflona Industrial Madeireira: Execução Financeira e Técnica Da Concessão (Jamari—UMF I). Available online: <https://www.gov.br/florestal/pt-br/assuntos/concessoes-e-monitoramento/concessoes-florestais-em-andamento/floresta-nacional-do-jamari-ro-2/madeflona-industrial-madeireira-execucao-tecnica-da-concessao-jamari-umf-i-2> (accessed on 20 October 2024).
34. Lapola, D.M.; Pinho, P.; Barlow, J.; Aragão, L.E.O.C.; Berenguer, E.; Carmenta, R.; Liddy, H.M.; Seixas, H.; Silva, C.V.J.; Silva-Junior, C.H.L.; et al. The Drivers and Impacts of Amazon Forest Degradation. *Science* **2023**, *379*, eabp8622. [CrossRef] [PubMed]
35. Moran, E.F.; Brondizio, E.S.; Tucker, J.M.; Silva-Forsberg, M.C.d.; Mccracken, S.; Falesi, I. Effects of Soil Fertility and Land-Use on Forest Succession in Amazônia. *For. Ecol. Manag.* **2000**, *139*, 93–108. [CrossRef]
36. Rouse, J.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the ERTS-1 SYMPOSIUM, Washington, DC, USA, 10–14 December 1973; NASA: Washington, DC, USA, 1973; pp. 309–317.
37. Key, C.H.; Zhu, Z.; Ohlen, D.; Howard, S.; McKinley, R.; Benson, N. The Normalized Burn Ratio and Relationships to Burn Severity: Ecology, Remote Sensing and Implementation. In Proceedings of the Ninth Forest Service Remote Sensing Applications Conference, San Diego, CA, USA, 8–12 April 2002; American Society for Photogrammetry and Remote Sensing: San Diego, CA, USA, 2002.
38. Turubanova, S.; Potapov, P.V.; Tyukavina, A.; Hansen, M.C. Ongoing Primary Forest Loss in Brazil, Democratic Republic of the Congo, and Indonesia. *Environ. Res. Lett.* **2018**, *13*, 074028. [CrossRef]

39. Silva Junior, C.H.L.; Heinrich, V.H.A.; Freire, A.T.G.; Broggio, I.S.; Rosan, T.M.; Doblaz, J.; Anderson, L.O.; Rousseau, G.X.; Shimabukuro, Y.E.; Silva, C.A.; et al. Benchmark Maps of 33 Years of Secondary Forest Age for Brazil. *Sci. Data* **2020**, *7*, 269. [CrossRef]
40. Tyukavina, A.; Potapov, P.; Hansen, M.C.; Pickens, A.H.; Stehman, S.V.; Turubanova, S.; Parker, D.; Zalles, V.; Lima, A.; Kommareddy, I.; et al. Global Trends of Forest Loss Due to Fire From 2001 to 2019. *Front. Remote Sens.* **2022**, *3*, 825190. [CrossRef]
41. Grussu, G.; Testolin, R.; Saulei, S.; Farcomeni, A.; Yosi, C.K.; De Sanctis, M.; Attorre, F. Optimum Plot and Sample Sizes for Carbon Stock and Biodiversity Estimation in the Lowland Tropical Forests of Papua New Guinea. *Forestry* **2016**, *89*, 150–158. [CrossRef]
42. Zolkos, S.G.; Goetz, S.J.; Dubayah, R. A Meta-Analysis of Terrestrial Aboveground Biomass Estimation Using Lidar Remote Sensing. *Remote Sens. Environ.* **2013**, *128*, 289–298. [CrossRef]
43. Ometto, J.P.; Gorgens, B.G.; Assis, M.; Cantinho, R.Z.; Pereira, F.R.d.S.; Sato, L.Y. Summary of the Airborne LiDAR Transects Collected by EBA in the Brazilian Amazon (Version 20210219) [Data Set]. Zenodo. Available online: <https://zenodo.org/records/4552699> (accessed on 20 October 2024).
44. LaRue, E.A.; Fahey, R.; Fuson, T.L.; Foster, J.R.; Matthes, J.H.; Krause, K.; Hardiman, B.S. Evaluating the Sensitivity of Forest Structural Diversity Characterization to LiDAR Point Density. *Ecosphere* **2022**, *13*, e4209. [CrossRef]
45. Singh, K.K.; Chen, G.; Vogler, J.B.; Meentemeyer, R.K. When Big Data Are Too Much: Effects of LiDAR Returns and Point Density on Estimation of Forest Biomass. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3210–3218. [CrossRef]
46. Isenburg, M. LAStools—Efficient LiDAR Processing Software (Version 171030, Unlicensed). 2018. Available online: <https://rapidlasso.de/downloads/> (accessed on 5 August 2021).
47. McGaughey, R.J. FUSION/LDV LIDAR Analysis and Visualization Software (Version 4.61). 2022. Available online: http://forsys.sefs.uw.edu/fusion/fusion_overview.html (accessed on 12 July 2022).
48. Næsset, E. Effects of Different Sensors, Flying Altitudes, and Pulse Repetition Frequencies on Forest Canopy Metrics and Biophysical Stand Properties Derived from Small-Footprint Airborne Laser Data. *Remote Sens. Environ.* **2009**, *113*, 148–159. [CrossRef]
49. Roussel, J.R.; Auty, D. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R Package Version 1.6.1. 2018. Available online: <https://cran.r-project.org/web/packages/lidR/index.html> (accessed on 5 August 2021).
50. Stark, S.C.; Leitold, V.; Wu, J.L.; Hunter, M.O.; de Castilho, C.V.; Costa, F.R.C.; McMahon, S.M.; Parker, G.G.; Shimabukuro, M.T.; Lefsky, M.A.; et al. Amazon Forest Carbon Dynamics Predicted by Profiles of Canopy Leaf Area and Light Environment. *Ecol. Lett.* **2012**, *15*, 1406–1414. [CrossRef]
51. Magurran, A.E. *Measuring Biological Diversity*; Blackwell Science, Ltd.: Oxford, UK, 2004.
52. Wilson, M.F.J.; O’Connell, B.; Brown, C.; Guinan, J.C.; Grehan, A.J. Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Mar. Geod.* **2007**, *30*, 3–35. [CrossRef]
53. Clark, R.N.; Roush, T.L. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *J. Geophys. Res.* **1984**, *89*, 6329–6340. [CrossRef]
54. Gitelson, A.A.; Keydan, G.P.; Merzlyak, M.N. Three-Band Model for Noninvasive Estimation of Chlorophyll, Carotenoids, and Anthocyanin Contents in Higher Plant Leaves. *Geophys. Res. Lett.* **2006**, *33*, 2–6. [CrossRef]
55. Nagler, P.L.; Daughtry, C.S.T.; Goward, S.N. Plant Litter and Soil Reflectance. *Remote Sens. Environ.* **2000**, *71*, 207–215. [CrossRef]
56. le Maire, G.; François, C.; Soudani, K.; Berveiller, D.; Pontailier, J.Y.; Bréda, N.; Genet, H.; Davi, H.; Dufréne, E. Calibration and Validation of Hyperspectral Indices for the Estimation of Broadleaved Forest Leaf Chlorophyll Content, Leaf Mass per Area, Leaf Area Index and Leaf Canopy Biomass. *Remote Sens. Environ.* **2008**, *112*, 3846–3864. [CrossRef]
57. Apan, A.; Held, A.; Phinn, S.; Markley, J. Detecting Sugarcane “orange Rust” Disease Using EO-1 Hyperion Hyperspectral Imagery. *Int. J. Remote Sens.* **2004**, *25*, 489–498. [CrossRef]
58. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
59. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
60. Galvão, L.S.; Formaggio, A.R.; Tisot, D.A. Discrimination of Sugarcane Varieties in Southeastern Brazil with EO-1 Hyperion Data. *Remote Sens. Environ.* **2005**, *94*, 523–534. [CrossRef]
61. Serrano, L.; Peñuelas, J.; Ustin, S.L. Remote Sensing of Nitrogen and Lignin in Mediterranean Vegetation from AVIRIS Data: Decomposing Biochemical from Structural Signals. *Remote Sens. Environ.* **2002**, *81*, 355. [CrossRef]
62. Gao, B.C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
63. Gamon, J.A.; Peñuelas, J.; Field, C.B. A Narrow-Waveband Spectral Index That Tracks Diurnal Changes in Photosynthetic Efficiency. *Remote Sens. Environ.* **1992**, *41*, 35–44. [CrossRef]
64. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-Destructive Optical Detection of Pigment Changes during Leaf Senescence and Fruit Ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]
65. Peñuelas, J.; Piñol, J.; Ogaya, R.; Filella, I. Estimation of Plant Water Concentration by the Reflectance Water Index WI (R900/R970). *Int. J. Remote Sens.* **1997**, *18*, 2869–2875. [CrossRef]

66. Guyot, G.; Baret, F. Utilisation de La Haute Resolution Spectrale Pour Suivre l'état Des Couverts Vegetaux. In *Proceedings of the Spectral Signatures of Objects in Remote Sensing*; Guyenne, T.D., Hunt, J.J., Eds.; ESA Special Publication: Aussois, France, 1988; Volume 287, pp. 279–286.
67. Merton, R.N. Monitoring Community Hysteresis Using Spectral Shift Analysis and the Red-Edge Vegetation Stress Index. In *Proceedings of the Seventh Annual JPL Airborne Earth Science Workshop*, Pasadena, CA, USA, 12–16 January 1998; NASA: Pasadena, CA, USA, 1998.
68. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [[CrossRef](#)]
69. Gitelson, A.A.; Zur, Y.; Chivkunova, O.B.; Merzlyak, M.N. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochem. Photobiol.* **2002**, *75*, 272. [[CrossRef](#)]
70. Vogelmann, J.E.; Rock, B.N.; Moss, D.M. Red Edge Spectral Measurements from Sugar Maple Leaves. *Int. J. Remote Sens.* **1993**, *14*, 1563–1575. [[CrossRef](#)]
71. Lehnert, L.W.; Meyer, H.; Bendix, J. Hsdar: Manage, Analyse and Simulate Hyperspectral Data in R. R Package Version 0.7.1. 2018. Available online: <http://cran.nexr.com/web/packages/hsdar/index.html> (accessed on 5 August 2021).
72. Clark, M.L.; Roberts, D.A.; Ewel, J.J.; Clark, D.B. Estimation of Tropical Rain Forest Aboveground Biomass with Small-Footprint Lidar and Hyperspectral Sensors. *Remote Sens. Environ.* **2011**, *115*, 2931–2942. [[CrossRef](#)]
73. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
74. Havlicek, L.L.; Peterson, N.L. Robustness of the Pearson Correlation against Violations of Assumptions. *Percept Mot Ski.* **1976**, *43*, 1319–1334. [[CrossRef](#)]
75. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab—An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20. [[CrossRef](#)]
76. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
77. Cohen, J. *Statistical Power Analysis for Behavioural Sciences*, 2nd ed.; Erlbaum: Hillsdale, MI, USA, 1988.
78. Dalponte, M.; Member, S.S.; Bruzzone, L.; Member, S.S.; Gianelle, D.; Member, S.S.; Bruzzone, L.; Member, S.S.; Gianelle, D. Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *Geosci. Remote Sens. IEEE Trans.* **2008**, *46*, 1416–1427. [[CrossRef](#)]
79. Fassnacht, F.E.; Hartig, F.; Lati, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B.; Latifi, H.; Berger, C.; Hernández, J.; et al. Importance of Sample Size, Data Type and Prediction Method for Remote Sensing-Based Estimations of Aboveground Forest Biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114. [[CrossRef](#)]
80. Dalponte, M.; Bruzzone, L.; Gianelle, D. Tree Species Classification in the Southern Alps Based on the Fusion of Very High Geometrical Resolution Multispectral/Hyperspectral Images and LiDAR Data. *Remote Sens. Environ.* **2012**, *123*, 258–270. [[CrossRef](#)]
81. Shoot, C.; Andersen, H.; Moskal, L.M.; Babcock, C.; Cook, B.D.; Morton, D.C. Classifying Forest Type in the National Forest Inventory Context with Airborne Hyperspectral and Lidar Data. *Remote Sens.* **2021**, *13*, 1863. [[CrossRef](#)]
82. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Classification of Successional Forest Stages in the Brazilian Amazon Basin. *For. Ecol. Manag.* **2003**, *181*, 301–312. [[CrossRef](#)]
83. Jucker, T.; Asner, G.P.; Dalponte, M.; Lewis, S.L.; Teh, Y.A.; Malhi, Y.; Deere, N.J.; Nilus, R.; Svátek, M.; Kvasnica, J.; et al. Estimating Aboveground Carbon Density and Its Uncertainty in Borneo's Structurally Complex Tropical Forests Using Airborne Laser Scanning. *Biogeosciences* **2018**, *15*, 3811–3830. [[CrossRef](#)]
84. Prestes, N.C.C.d.S.; Massi, K.G.; Silva, E.A.; Nogueira, D.S.; de Oliveira, E.A.; Freitag, R.; Marimon, B.S.; Marimon-Junior, B.H.; Keller, M.; Feldpausch, T.R. Fire Effects on Understory Forest Regeneration in Southern Amazonia. *Front. For. Glob. Chang.* **2020**, *3*, 10. [[CrossRef](#)]
85. Guariguata, M.R.; Ostertag, R. Neotropical Secondary Forest Succession: Changes in Structural and Functional Characteristics. *For. Ecol. Manag.* **2001**, *148*, 185–206. [[CrossRef](#)]
86. Wang, Y.; Ziv, G.; Adami, M.; Mitchard, E.; Batterman, S.A.; Buermann, W.; Marimon, B.S.; Junior, B.H.M.; Reis, S.M.; Rodrigues, D.; et al. Mapping Tropical Disturbed Forests Using Multi-Decadal 30 m Optical Satellite Imagery. *Remote Sens. Environ.* **2019**, *221*, 474–488. [[CrossRef](#)]
87. Asner, G.P.; Nepstad, D.; Cardinot, G.; Ray, D. Drought Stress and Carbon Uptake in an Amazon Forest Measured with Spaceborne Imaging Spectroscopy. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6039–6044. [[CrossRef](#)]
88. Jaročníšková, A.; Kopeček, D.; Niedzielko, J.; Wylazłowska, J.; Halladin-Dąbrowska, A.; Charyton, J.; Piernik, A.; Kamiński, D. The Utility of Airborne Hyperspectral and Satellite Multispectral Images in Identifying Natura 2000 Non-Forest Habitats for Conservation Purposes. *Sci. Rep.* **2023**, *13*, 4549. [[CrossRef](#)]
89. Anderson, J.E.; Plourde, L.C.; Martin, M.E.; Braswell, B.H.; Smith, M.L.; Dubayah, R.O.; Hofton, M.A.; Blair, J.B. Integrating Waveform Lidar with Hyperspectral Imagery for Inventory of a Northern Temperate Forest. *Remote Sens. Environ.* **2008**, *112*, 1856–1870. [[CrossRef](#)]
90. Jones, T.G.; Coops, N.C.; Sharma, T. Assessing the Utility of Airborne Hyperspectral and LiDAR Data for Species Distribution Mapping in the Coastal Pacific Northwest, Canada. *Remote Sens. Environ.* **2010**, *114*, 2841–2852. [[CrossRef](#)]
91. Thomas, V.; Treitz, P.; McCaughey, J.H.; Noland, T.; Rich, L. Canopy Chlorophyll Concentration Estimation Using Hyperspectral and Lidar Data for a Boreal Mixedwood Forest in Northern Ontario, Canada. *Int. J. Remote Sens.* **2008**, *29*, 1029–1052. [[CrossRef](#)]

92. Almeida, D.R.A.d.; Broadbent, E.N.; Ferreira, M.P.; Meli, P.; Zambrano, A.M.A.; Gorgens, E.B.; Resende, A.F.; de Almeida, C.T.; do Amaral, C.H.; Corte, A.P.D.; et al. Monitoring Restored Tropical Forest Diversity and Structure through UAV-Borne Hyperspectral and Lidar Fusion. *Remote Sens. Environ.* **2021**, *264*, 112582. [[CrossRef](#)]
93. Watson, J.E.M.; Evans, T.; Venter, O.; Williams, B.; Tulloch, A.; Stewart, C.; Thompson, I.; Ray, J.C.; Murray, K.; Salazar, A.; et al. The Exceptional Value of Intact Forest Ecosystems. *Nat. Ecol. Evol.* **2018**, *2*, 599–610. [[CrossRef](#)]
94. Alliance for Restoration in the Amazon Forest. Landscape Restoration in the Amazon—Overview and Paths to Follow. Position Paper: 16p ISBN 978-65-00-12760-7 2020. Available online: https://aliancaamazonia.org.br/wp-content/uploads/2021/06/PAPER_ALIANCA_EN_2020_FINAL.pdf (accessed on 20 October 2024).
95. Potapov, P.; Hansen, M.C.; Laestadius, L.; Turubanova, S.; Yaroshenko, A.; Thies, C.; Smith, W.; Zhuravleva, I.; Komarova, A.; Minnemeyer, S.; et al. The Last Frontiers of Wilderness: Tracking Loss of Intact Forest Landscapes from 2000 to 2013. *Sci. Adv.* **2017**, *3*, e1600821. [[CrossRef](#)] [[PubMed](#)]
96. Barlow, J.; Peres, C.A. Fire-Mediated Dieback and Compositional Cascade in an Amazonian Forest. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 1787–1794. [[CrossRef](#)] [[PubMed](#)]
97. Berenguer, E.; Ferreira, J.; Gardner, T.A.; Aragão, L.E.O.C.; De Camargo, P.B.; Cerri, C.E.; Durigan, M.; De Oliveira, R.C.; Vieira, I.C.G.; Barlow, J. A Large-Scale Field Assessment of Carbon Stocks in Human-Modified Tropical Forests. *Glob. Chang. Biol.* **2014**, *20*, 3713–3726. [[CrossRef](#)] [[PubMed](#)]
98. Xaud, H.A.M.; Martins, F.d.S.R.V.; Dos Santos, J.R. Tropical Forest Degradation by Mega-Fires in the Northern Brazilian Amazon. *For. Ecol. Manag.* **2013**, *294*, 97–106. [[CrossRef](#)]
99. ITTO—International Tropical Timber Organization. *ITTO Guidelines for the Restoration, Management and Rehabilitation of Degraded and Secondary Tropical Forests*; International Tropical Timber Organization: Yokohama, Japan, 2002.

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.