



## Article

# Estimating Brazilian Amazon Canopy Height Using Landsat Reflectance Products in a Random Forest Model with Lidar as Reference Data

Pedro V. C. Oliveira <sup>1,2,\*</sup> , Hankui K. Zhang <sup>2</sup> and Xiaoyang Zhang <sup>2</sup> <sup>1</sup> Imaging Center, Image Processing Laboratory, South Dakota State University, Brookings, SD 57007, USA<sup>2</sup> Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA; hankui.zhang@sdstate.edu (H.K.Z.); xiaoyang.zhang@sdstate.edu (X.Z.)

\* Correspondence: pedro.oliveira@sdstate.edu; Tel.: +1-605-688-6714

**Abstract:** Landsat data have been used to derive forest canopy structure, height, and volume using machine learning models, i.e., giving computers the ability to learn from data and make decisions and predictions without being explicitly programmed, with training data provided by ground measurement or airborne lidar. This study explored the potential use of Landsat reflectance and airborne lidar data as training data to estimate canopy heights in the Brazilian Amazon forest and examined the impacts of Landsat reflectance products at different process levels and sample spatial autocorrelation on random forest modeling. Specifically, this study assessed the accuracy of canopy height predictions from random forest regression models impacted by three different Landsat 8 reflectance product inputs (i.e., USGS level 1 top of atmosphere reflectance, USGS level 2 surface reflectance, and NASA nadir bidirectional reflectance distribution function (BRDF) adjusted reflectance (NBAR)), sample sizes, training/test split strategies, and geographic coordinates. In the establishment of random forest regression models, the dependent variable (i.e., the response variable) was the dominant canopy heights at a 90 m resolution derived from airborne lidar data, while the independent variables (i.e., the predictor variables) were the temporal metrics extracted from each Landsat reflectance product. The results indicated that the choice of Landsat reflectance products had an impact on model accuracy, with NBAR data yielding more trustful results than the other products despite having higher RMSE values. Training and test split strategy also affected the derived model accuracy metrics, with the random sample split (randomly distributed training and test samples) showing inflated accuracy compared to the spatial split (training and test samples spatially set apart). Such inflation was induced by the spatial autocorrelation that existed between training and test data in the random split. The inclusion of geographic coordinates as independent variables improved model accuracy in the random split strategy but not in the spatial split, where training and test samples had different geographic coordinate ranges. The study highlighted the importance of data processing levels and the training and test split methods in random forest modeling of canopy height.

**Keywords:** canopy height; Brazilian Amazon; airborne lidar; time series; Landsat; top-of-atmosphere reflectance; surface reflectance; NBAR; random forest; spatial autocorrelation



**Citation:** Oliveira, P.V.C.; Zhang, H.K.; Zhang, X. Estimating Brazilian Amazon Canopy Height Using Landsat Reflectance Products in a Random Forest Model with Lidar as Reference Data. *Remote Sens.* **2024**, *16*, 2571. <https://doi.org/10.3390/rs16142571>

Academic Editors: Henrik J. Persson, Milto Miltiadou and Rorai Pereira Martins-Neto

Received: 26 April 2024

Revised: 7 July 2024

Accepted: 11 July 2024

Published: 13 July 2024



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

## 1. Introduction

The largest uncertainties in global carbon stocks and changes mainly occur in tropical forests [1,2]. The Amazon forest is the largest tropical moist forest in the world, with about 60% of it located in Brazil. According to data collected at the end of the last decade, the Brazilian Amazon forest is estimated to hold an average of 174 Mg ha<sup>-1</sup> of aboveground biomass (AGB) [3]. AGB varies depending on the vegetation structure (form, growth stage, etc.), floristic composition, and disturbances, including drought, fire, degradation, and deforestation [4–7]. Conventionally, forest AGB has been measured using destructive methods, i.e., by cutting and weighing trees, which is not practical in tropical forests, or

using allometric methods to derive representative statistical relationships between AGB and the tree height or the tree diameter at breast height [8]. Lidar remote sensing, alternatively, provides new capabilities for estimating canopy height and other structural attributes and has the potential to improve or even replace allometric models [9–11]. Lidar sensors on satellites are able to repeatedly produce wall-to-wall mapping. For example, the NASA Global Ecosystem Dynamics Investigation Lidar (GEDI) on the International Space Station provides new opportunities for tropical forest canopy height and AGB estimation [12,13]. Previously, the satellite-borne Geoscience Laser Altimeter System (GLAS) was used to produce global canopy height maps [14–16], although it was not designed for monitoring vegetation. The GLAS was a waveform lidar with an elliptical, approximately 65 m diameter footprint operating on a 33-day period three times per year [17], which was too coarse to capture landscape variability and change/disturbance.

Spaceborne and airborne lidar, acquired under cloud-free conditions, have been combined with time-series Landsat surface reflectance data to derive national and continental-scale maps of forest height through supervised (training-based) statistical modeling [18–21]. However, no studies have yet explored how different Landsat reflectance products (i.e., top-of-atmosphere reflectance, surface reflectance, and nadir bidirectional reflectance distribution function (BRDF) adjusted reflectance (NBAR)) influence the accuracy of estimating tropical forest canopy height and AGB. Commonly, the selection of different products in the existing literature [18–21] is based on their availability of the history Landsat top-of-atmosphere reflectance data or their capability to process the data to a higher level. The surface reflectance (i.e., suppressing the atmospheric condition variations across time) and the NBAR (i.e., suppressing the viewing angle induced reflectance variations across swath) are expected to have better performance in canopy height estimation than the top-of-atmosphere reflectance data. However, their quantitative differences in estimating tropical forest canopy height and AGB are poorly understood.

Moreover, the spatial autocorrelation of data samples from Landsat surface reflectance could have impacts on the canopy height and AGB estimates using machine learning (i.e., giving computers the ability to learn from data and make decisions and predictions without being explicitly programmed) models such as random forests (RFs) because it violates model assumptions of independence of the observations [22–24]. A few studies have paid attention to this issue. For example, geographic position information was included as an independent variable in RF models for the predictions of aboveground carbon density [25] and mean canopy height [26]. Furthermore, altering the sampling scheme, often by ensuring that training and test data are geographically distant from one another, seems to decrease the occurrence of spatial autocorrelation [27], but it is yet to be tested in canopy height modeling.

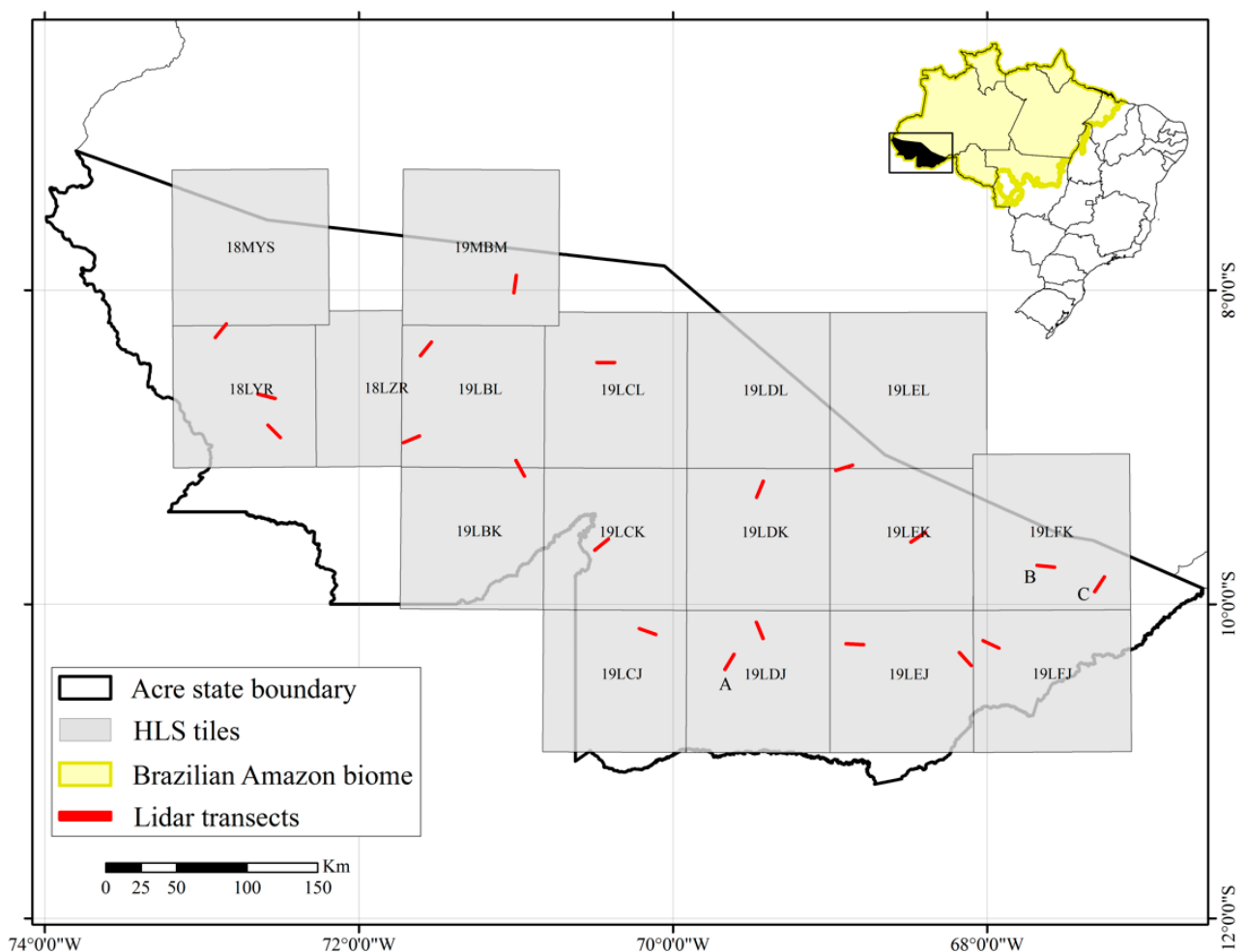
The main goal of this work is to investigate how the different Landsat reflectance products and the sample spatial autocorrelation using different sampling strategies influence RF modeling of canopy heights in the Brazilian Amazon. To do this, we used lidar-derived dominant canopy heights and Landsat time-series data as dependent (i.e., response) and independent (i.e., predictor) variables, respectively, to train RF regression models to predict canopy height. We then examined quantitatively the sensitivity of canopy height predictions to different Landsat products (top-of-atmosphere reflectance, surface reflectance, and NBAR products), sample size, training/test split strategy, and inclusion or not of geographic coordinates. Finally, we discussed the implications of this research.

## 2. Materials and Methods

### 2.1. Study Area and Data

#### 2.1.1. Study Area

The study area consisted of 20 transects in Acre state, located on the southwest side of the Brazilian Amazon biome and bordered by the states of Amazonas to the north, Rondônia to the east, and Bolivia to the south and west (Figure 1). In spite of being under pressure of deforestation and degradation, the study area is mostly covered by pristine old-growth open- and dense-canopy rainforests [28] with a few rivers and fertile lowlands. Acre state presents a tropical monsoon and tropical rainforest climate (Am and Af, respectively) according to the Köppen–Geiger climate classification system. The average annual temperature measured at the state’s capital city weather station is  $\sim 25.5$  °C. Total yearly rainfall is  $\sim 1800$  mm, and the dry season occurs between June and August. During the early dry season, the vegetation greenness in the state increases due to leaf flush, and in the late dry season, it decreases due to the shedding of leaves [29].



**Figure 1.** Location of the 20 airborne lidar transects (red bars) within Acre state in the Brazilian Amazon biome and the  $17,110 \times 110$  km HLS tiles intersecting these transects. Transects A, B, and C are evaluated in detail in the Results section.

#### 2.1.2. Airborne Lidar Transect Data

From 2016 to 2018, the Brazilian National Institute for Space Research (INPE) collected full-waveform lidar data over 901 transects to support the estimation of forest biomass and carbon stocks within the Brazilian Amazon biome (<http://www.ccst.inpe.br/projetos/eba-estimativa-de-biomassa-na-amazonia/>, accessed on 25 June 2024). The data were

captured using a Harrier 68i lidar system operating at 400 kHz with a beam divergence of  $\leq 0.5$  mrad and a pulse diameter of 15–30 cm. The lidar scans were conducted from a Cessna 206 aircraft flying at an altitude of 600 m above the ground. The criteria for each transect included a minimum length of 12.5 km, a minimum width of 0.3 km, horizontal and vertical location accuracy of  $\leq 1$  m and  $\leq 0.5$  m, respectively, and a density of returns of  $\geq 4$  returns per square meter [30]. In this study, we used 20 lidar transects (Figure 1) collected over the Acre (AC) state in the southwestern part of the Brazilian Amazon forest biome.

### 2.1.3. Landsat Reflectance Products

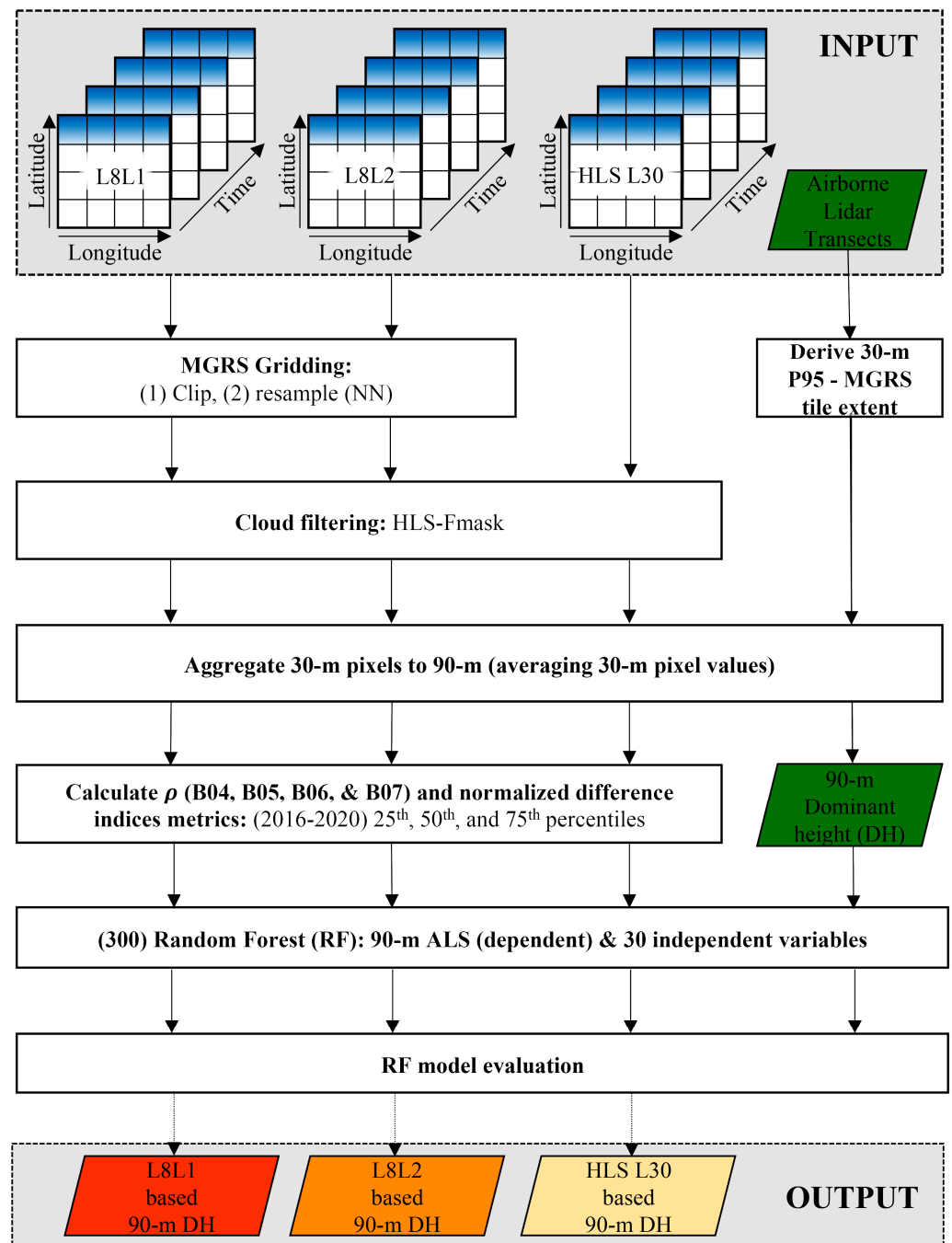
This study used Landsat 8 Operational Land Imager (OLI) image time series acquired between 2016 and 2020, intersecting the airborne lidar transects (Figure 1). The image time series increased the chances of finding multiple cloud-free measurements in the study area, which is regularly contaminated by cloud cover throughout most of the year [31]. OLI has eight 30 m spectral bands, including coastal aerosol (0.43–0.45  $\mu\text{m}$ ), blue (0.45–0.51  $\mu\text{m}$ ), green (0.53–0.59  $\mu\text{m}$ ), red (0.64–0.67  $\mu\text{m}$ ), near-infrared (0.85–0.88  $\mu\text{m}$ ), shortwave infrared (SWIR) 1 (1.57–1.65  $\mu\text{m}$ ), SWIR 2 (2.11–2.29  $\mu\text{m}$ ), and cirrus (1.36–1.38  $\mu\text{m}$ ) and one 15 m panchromatic (0.50–0.68  $\mu\text{m}$ ) spectral band.

Three Landsat OLI reflectance products are available, which are from Landsat 8 Collection 2 Level 1 (L8L1), Collection 2 Level 2 (L8L2), and HLS 30 m Landsat product (L30) V2.0 images. L8L1 and L8L2 images use the Worldwide Reference System (WRS) path and row coordinates. Each image covers an area of approximately 185 km by 180 km and is provided in the Universal Transverse Mercator (UTM) projection. L8L1 top-of-atmosphere (TOA) reflectance is obtained by applying radiometric rescaling values found in the metadata. These values are employed to convert the digital numbers (DNs) of pixels in each spectral band into TOA reflectance. L8L2 surface reflectance is derived using the Landsat Surface Reflectance Code (LaSRC), which is based on the Second Simulation of a Satellite Signal in the Solar Spectrum Vector (6SV) radiative transfer code [32]. A comprehensive description of the L8L1 and L8L2 products can be found in the work by Crawford et al. [33].

The Landsat reflectance product from HLS, HLS L30, is derived from L8L1 data using atmospheric correction and BRDF correction. The atmospheric correction is still undertaken to speed up the HLS generation process to meet near real-time application requirements despite the USGS Collection-2 surface reflectance being available. In addition, HLS L30 BRDF correction uses the c-factor method [34] and derives NBAR for a modeled solar zenith that is very close to the observed solar zenith [35]. The tile format of HLS L30 imagery is produced following the Sentinel-2 tiling system, which is a variation of the Military Grid Reference System (MGRS) and is available in the UTM projection. The Sentinel-2 tiling system divides the Earth's surface into 6-degree longitudinal zones and 8-degree latitudinal bands, and each 6- by 8-degree grid is further subdivided into 109.8 km by 109.8 km tiles.

## 2.2. Methods

We pre-processed the time series L8L1, L8L2, and HLS L30 image data before using their spectral bands as independent variables in RF models. We then sampled training and test data and used the training data to build the RF models for dominant canopy heights (DHs). We further conducted a statistical analysis using the test data to evaluate model performance. The flowchart (Figure 2) shows the main procedures adopted in this study.



**Figure 2.** The main (pre-)processes to model dominant canopy height using L8L1, L8L2, and HLS L30 as independent variables. B04, B05, B06, and B07 are Landsat 8 red, NIR, SWIR 1, and SWIR 2 spectral bands.

### 2.2.1. Image Data Pre-Processing

The image pre-processing consisted of bringing all the image data to a common reference system (i.e., MGRS), resampling the pixels of the three Landsat products to the same positions in the MGRS, cloud masking, and spatially aggregating the images to a coarser spatial resolution. Specifically, the L8L1 and the L8L2 image data were cropped to fit the HLS tiles. We resampled the pixels into the HLS tiles using the nearest neighbor algorithm to accommodate the disparity in pixel positioning between L8L1 and L8L2 with HLS L30. L8L1 and L8L2 scenes that intersected tiles from another UTM zone were reprojected before cropping.

A cloud mask included in the Quality Assessment (QA) layer provided by the HLS product was employed for the L8L1, L8L2, and HLS L30 data. According to Masek et al. [36], the QA's cloud mask was derived using the Fmask version 4.2 [37], which is an updated version of the algorithm designed for Landsat Collection 2 data. All cloud-affected pixels (i.e., cloud, adjacent to cloud/shadow, and shadow) were removed from the time-series analysis.

Finally, the 30 m pixels in the L8L1, L8L2, and HLS L30 time series were then aggregated into 90 m to accelerate the data processing and to mitigate the impact of any re-projection and registration errors. Within each HLS tile, the 90 m pixels were assigned the average value based on the nine 30 m pixels encompassed by them.

### 2.2.2. Temporal Metric Extraction for Multi-Year Observations

We derived temporal metrics of Landsat reflectance for the 2016–2020 period and used them for large-area modeling of canopy heights. Temporal metrics are generally insensitive to spatial phenological differences and to missing data [19,38,39]. Following a previous study [39], the temporal metrics included the 25th percentile (P25), 50th percentile (P50), and 75th percentile (P75) for the red, NIR, and two SWIR bands. Additionally, we also calculated the same percentile values for their normalized difference (ND) band ratios, which are defined as:

$$ND_{xy} = \frac{B_x - B_y}{B_x + B_y} \quad (1)$$

where  $B_x$  and  $B_y$  represent all possible combinations using the red, NIR, and two SWIR bands. This process resulted in a total of 30 (10 bands and band ratios  $\times$  3 percentiles per band and band ratio) temporal metrics for each 90 m tile pixel.

### 2.2.3. Lidar Data Pre-Processing and 90 m Dominant Canopy Height Estimation

We pre-processed the airborne lidar data before estimating dominant canopy height (DH) using the publicly available FUSION lidar processing software version 4.50 [40]. Within this pre-process, we separated the ground returns from the non-ground returns based on statistical criteria for each transect point cloud [41] using FUSION's 'GroundFilter' tool. The resulting dataset was then utilized to generate a digital terrain model (DTM) at a 2 m resolution using FUSION's 'GridSurfaceCreate' tool. Pixel elevations were determined by either calculating the average of all ground points within the pixel or by interpolating neighboring elevations in cases where ground returns were not available. Finally, we used the first returns in the point clouds to compute canopy height metrics in relation to the elevations extracted from the DTM through FUSION's 'GridMetrics' tool, utilizing 30 m resolution cells that were aligned with the 30 m HLS L30 pixel locations.

We adopted the 95th percentile (P95) canopy height metric as the representation of the top of the canopy to prevent the inclusion of heights above the canopies. Then, we created a P95 raster and transformed it into the DH [42]. This conversion involved aggregating the P95 30 m pixels into 90 m pixels aligned with the 90 m temporal metrics and assigning them the corresponding average value. Finally, 90 m pixels in the edge of the lidar transects were removed to avoid the use of discrepant DH values.

### 2.2.4. Random Forest DH Prediction Experiments

Training and test data should optimally be collected by random sampling [43]. Conventional sampling methods (e.g., simple random or stratified random sampling) do not prevent the selected training and test data from being spatially autocorrelated. A standard approach to reduce the incidence of spatial autocorrelation is to manipulate the sampling scheme, typically by ensuring that the training and test data are not spatially close to each other. However, this may result in smaller training and test data sets, particularly over airborne lidar transects such as the ones in this study that are narrow. It is well established that model prediction and accuracy reliability are reduced with small training and test sample sizes, respectively [44–46]. In this study, two DH prediction sensitivity analyses

were examined to evaluate spatial dependence between training and test sets. The first analysis used a random split of training and test and aimed to evaluate the model accuracy by varying sample sizes (i.e., 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%), with each size using the randomly sampled 80% data to train and 20% to test the RF (random split 80%-20%). The second analysis used a spatial split of training and test and aimed to evaluate the model accuracy by also varying sample sizes from 1% to 50% but using all the sampled observations from 16 lidar transects to train and 4 lidar transects to test the RF (spatial split 16-4).

The RF DH prediction sensitivity analysis to the size of the training and test samples was examined without (lat/lon: no) and with (lat/lon: yes) geolocation that was the 90 m pixels center latitude and longitude as independent variables in the RF. Including the geolocation in models was assumed to reduce the spatial autocorrelation effect [25,26].

RF regression is an ensemble of a number of trees, with each tree trained by only randomly selected samples from all the training samples [47]. In this study, following the recommendations in Liaw and Wiener [48], we used 500 trees (Ntrees), with each tree derived independently using 63.2% of the training data selected randomly without replacement and 33.33% (=10 in this study) of the independent variables also selected randomly. The RF prediction accuracy may be evaluated using the remaining ‘out-of-bag’ 32.8% (1 – 63.2%) training data samples for each tree. In the out-of-bag evaluation, each sample is predicted using all the trees that have never used the sample for training, and the mean of those estimates is considered the out-of-bag predicted value [47]. The RF prediction accuracy can also be evaluated using test data that are not used to generate the RF, which is the case of this study. Both approaches, particularly the latter, require a large amount of training/test data that may be expensive or difficult to obtain. For this reason, we ran the RF 300 times (each with 500 trees), which also allowed us to create a distribution of RF DH predictions. The accuracy of each RF DH prediction was assessed using the root mean squared error (RMSE) and mean error (ME), and they are defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2)$$

$$ME = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (3)$$

where  $\hat{y}_i$  is the RF predicted DH,  $y_i$  is the lidar-derived 90 m DH at a particular location, and  $n$  is the total number of 90 m test pixel observations.

#### 2.2.5. Random Forest DH Prediction Spatial Autocorrelation Quantification Experiments

The spatial autocorrelation was quantified using the Moran’s  $I$  correlogram [22,49,50]. It is the correlation between pairs of observations (points, polygons, or pixels) separated by a given distance. Moran’s  $I$  is a multi-directional spatial autocorrelation measure bounded from  $-1$  to  $1$ , with negative values indicating that neighbor pairs of observations are more dissimilar to each other and positive values indicating that they are more similar. A Moran’s  $I$  equals the expected value ( $E_I$ ), which is a negative value that approaches zero as the number of observations ( $n$ ) that has at least one neighbor increases, indicates no spatial autocorrelation (perfect randomness), and is defined as:

$$E_I = \frac{-1}{(n - 1)} \quad (4)$$

The Moran's  $I$  is defined in two spatial dimensions [22] as:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

where  $x_i$  is the variable observation at a particular location;  $x_j$  is the neighbor variable observation;  $\bar{x}$  is the mean of all observations in the data being considered; and  $w_{ij}$  is the  $ij$  element in the row-standardized spatial weight derived from a neighbor matrix. In this study, we derived the neighbor matrix based on the "Queen" contiguity, where neighboring observations received a value of 1, and 0 otherwise.

We evaluated spatial autocorrelation on independent variables and on the RF model's residuals (i.e., RF-predicted DH—ground-truth airborne lidar-derived DH) using the Moran's  $I$  correlogram with lags sampled every 90 m (e.g., up to 12 lags = up to 1080 m). The correlogram enables the analysis of spatial patterns within a region. In two dimensions, the correlogram can be used to show Moran's  $I$  estimates as a function of lags (i.e., discrete classes of distance) derived using different  $w$  row-standardized spatial weight matrices, each one derived from consecutive order-based or distance-based neighbor matrices [49,50].

### 3. Results

Table 1 presents the overall statistics for the DH of the 20 airborne lidar transects, comprising a total of 17,212 DH observations. The average DH of all transects is 31.10 m with a standard deviation of 5.92 m, and DHs range from 1.50 m to 45.30 m. Half the transects present minimum DH values < 10 m, ensuring that the training data likely cover a wide range of DH values. Table 2 shows the number of observations considering the different sample sizes (i.e., 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%), including those used for training and test RF models.

**Table 1.** Overall statistics for the DH of the 20 airborne lidar transects used in this study. Mean, standard deviation (std), minimum (min), maximum (max), and 25th (P25), 50th (P50), and 75th (P75) percentile units are meters.

Transect	Count	Mean	Std	Min	P25	P50	P75	Max
1	877	36.30	2.61	27.95	34.59	36.20	37.97	45.30
2	872	35.45	2.44	25.00	33.91	35.42	37.08	43.17
3	781	33.39	4.25	4.03	32.41	34.13	35.58	40.68
4	888	30.99	2.83	19.74	29.47	31.11	32.89	38.47
5 *	849	30.95	2.82	14.19	29.28	30.96	32.77	40.08
6	890	31.42	3.09	22.30	29.48	31.52	33.60	40.56
7	856	30.53	4.53	3.69	29.06	31.24	33.16	39.20
8	892	30.87	2.66	22.15	29.16	30.76	32.66	39.48
9	902	30.48	3.12	21.00	28.43	30.62	32.65	39.22
10	902	32.17	4.18	7.70	30.38	32.78	34.86	41.28
11	864	24.90	10.77	1.79	16.46	29.07	33.46	40.47
12	855	31.32	2.75	20.84	29.52	31.28	33.19	39.09
13	837	27.48	8.86	1.64	23.46	30.78	33.83	39.67
14 **	819	22.42	9.71	1.51	14.88	25.97	30.19	37.63
15 ***	641	28.64	5.99	2.88	26.57	30.03	32.29	38.46
16	900	31.11	3.23	9.23	29.26	31.42	33.24	39.26
17	893	30.46	4.18	5.31	28.91	30.92	32.98	40.78
18	919	33.44	2.44	18.44	31.93	33.50	35.02	40.86
19	848	32.65	2.65	25.58	30.85	32.57	34.36	43.56
20	918	35.58	4.43	8.09	33.94	36.14	38.20	44.45
All	12,212	31.10	5.93	1.51	29.45	32.05	34.47	45.30

Note: \*, \*\*, and \*\*\* are Figure 1 transects A, B, and C.



**Table 2.** Sample size and total number of DH observations ( $N_{\text{Total}}$ ) used to train ( $N_{\text{Train}}$ ) and test ( $N_{\text{Test}}$ ) RF models.

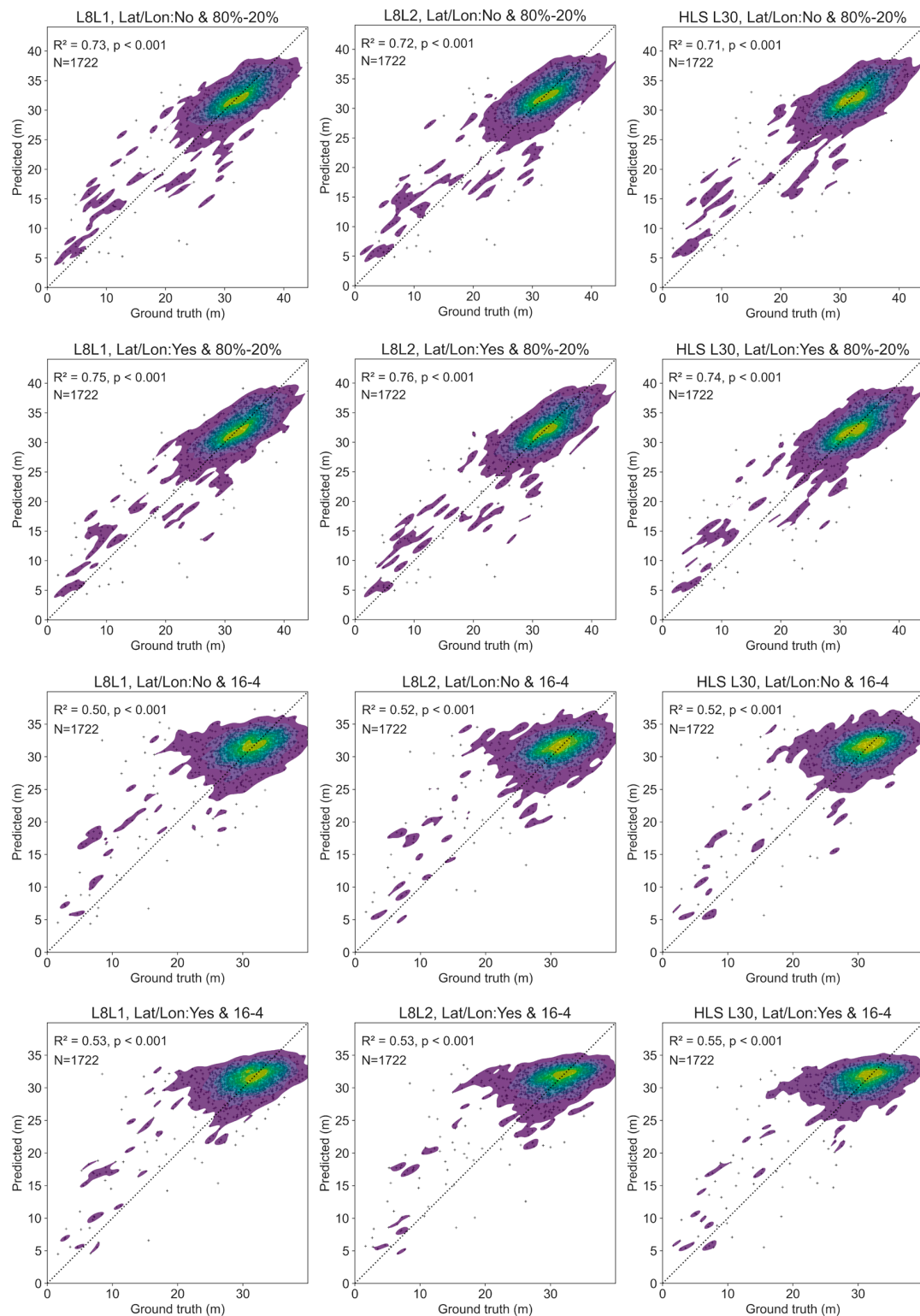
Sample Size	$N_{\text{Total}}$	$N_{\text{Train}}$	$N_{\text{Test}}$
1%	172	137	35
5%	861	688	173
10%	1721	1376	345
15%	2582	2065	517
20%	3442	2753	689
25%	4303	3442	861
30%	5164	4131	1033
35%	6024	4819	1205
40%	6885	5508	1377
45%	7745	6196	1549
50%	8606	6884	1772

### 3.1. Sensitivity of Random Forest DH Prediction to the Three Landsat Reflectance Products and Sampling Strategies

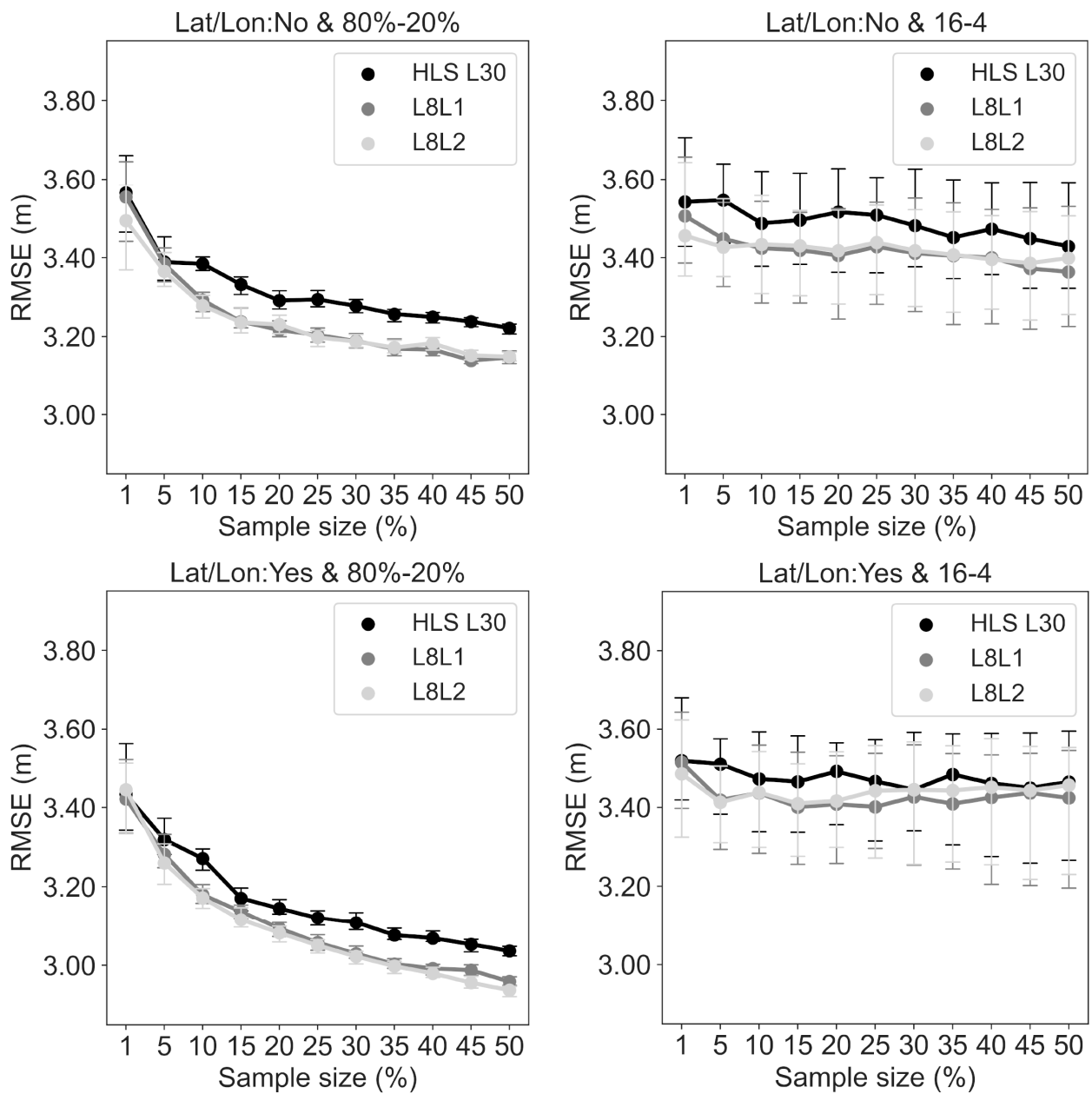
Figures 3–5 show that the RFs and their RMSE and ME vary with independent variables and RF model establishment schemes. Specifically, the independent variables were obtained, respectively, from the three Landsat reflectance products (i.e., L8L1, L8L2, and HLS L30) and geolocations, and RF models were established by considering the different sample sizes (Table 2), as well as different training and test split strategies (i.e., random 80%-20% and spatial 16-4).

Observing one of the 300 RF runs with a 50% sample size for each case mentioned above (Figure 3), a meaningful difference in the RF result was observed only when the training and test split strategy changed. RFs trained using 80%-20% random split produced DH estimates with  $R^2$  ranging from 0.71 to 0.76, with the best results when including latitude and longitude as independent variables. The use of different Landsat products marginally impacted the modeling of DHs. A similar pattern was found for the RFs trained with a 16-4 spatial split, although  $R^2$  varied between 0.50 and 0.55.

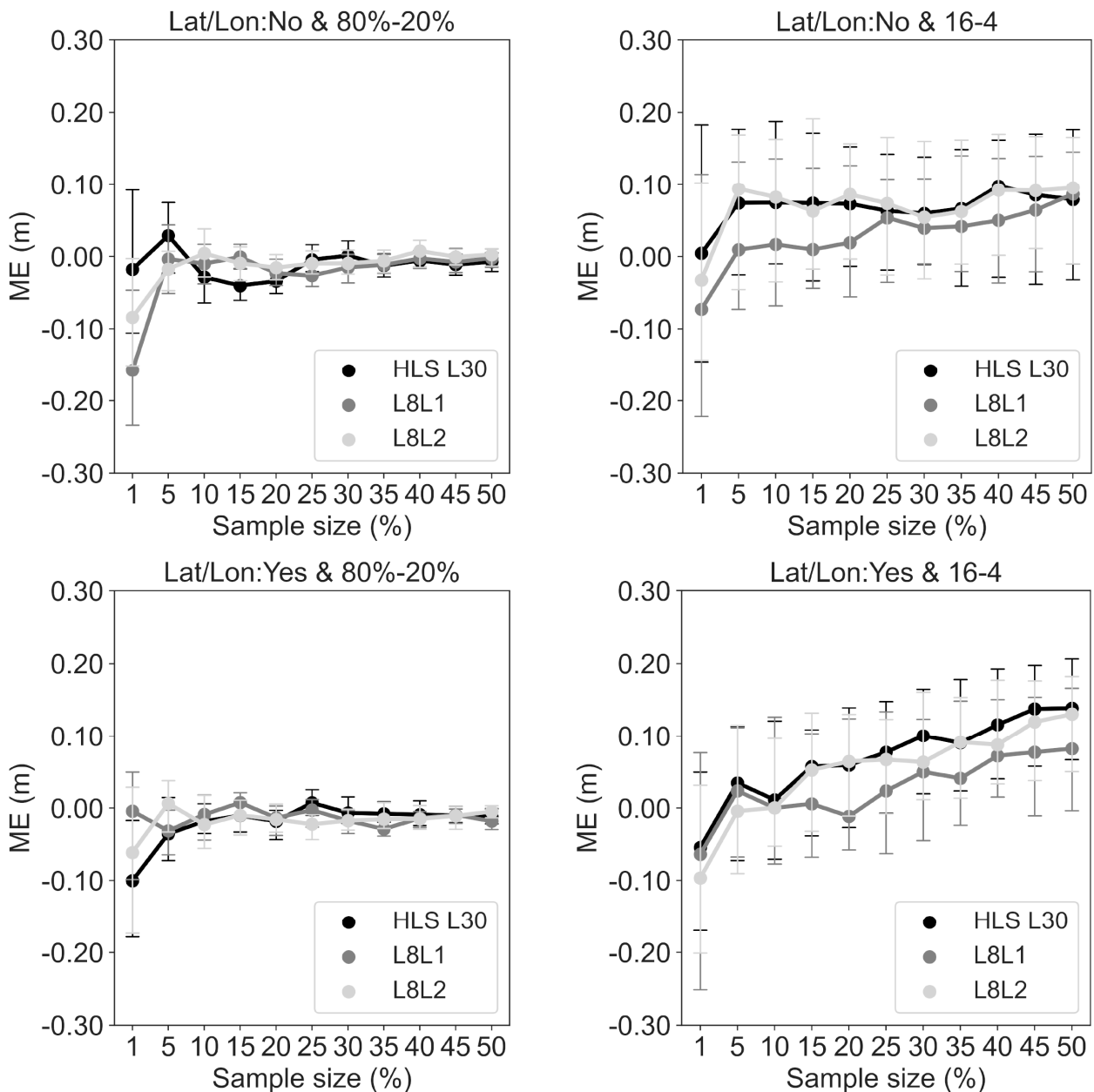
However, when the results of 300 RFs are combined for each sample size and sampling strategy, model errors (i.e., RMSE and ME) highlighted a variable effect among the different Landsat products. In the case of RFs that used an 80%-20% random split, the RMSE values (Figure 4) decreased with increasing sample sizes, as expected. RMSEs remained unchanged, considering a 95% confidence interval, for the RFs that used a 16-4 spatial split. The L8L1 and L8L2 products produced the same RMSE, considering the dispersion along the mean value in all the RF modeling experiments. The HLS L30 product had a similar RMSE to the L8L1 and L8L2 products when the RFs were modeled using spatial 16-4 training and test split. However, it generally had higher RMSEs when employing an 80%-20% random split and sample size  $\geq 10\%$ . The MEs in Figure 5 indicated no bias when RFs used an 80%-20% random split and sample size  $\geq 5\%$ . In RFs that used 16-4 spatial split and included latitude and longitude, the MEs of L8L2 and HLS L30 were positively biased (i.e., RF predicted DH mostly greater than observed DH) at sample sizes  $\geq 30\%$ . Without including latitude and longitude in RFs, all Landsat products were unbiased, considering a 95% confidence interval.



**Figure 3.** RF-predicted versus airborne lidar ground-truth DHs for Landsat 8 level 1 (L8L1), Landsat 8 level 2 (L8L2), and HLS L30 products. In the first and second rows, 80% of the randomly sampled observations from the 20 airborne lidar transects were used for training and 20% for testing the RF (80%-20%). In the third and fourth rows, the sampled observations from 16 transects were used for training and from 4 for testing the RF (16-4 spatial split). Latitude and longitude were not included (Lat/Lon: No) in the first and third rows and were included (Lat/Lon: Yes) in the second and fourth rows. All the RFs used a 50% sample size. The color gradient from dark blue to yellow shows increasing concentration of data points. Dotted line is 1:1 line.



**Figure 4.** Sensitivity analysis of RF-dominant canopy height prediction root mean squared error (RMSE) sensitivity analysis to sample size for Landsat 8 level 1 (L8L1), Landsat 8 level 2 (L8L2), and HLS L30 products. In the left column, 80% of the randomly sampled observations from the 20 airborne lidar transects were used for training and 20% for testing the RF (80%-20%). In the right column, the sampled observations from 16 transects were used for training and 4 for testing the RF (16-4 spatial split). Latitude and longitude were not included (Lat/Lon: No) in the first row and were included (Lat/Lon: Yes) in the second row. Each dot is the median root mean squared error (RMSE), and error bars are 95% confidence interval.

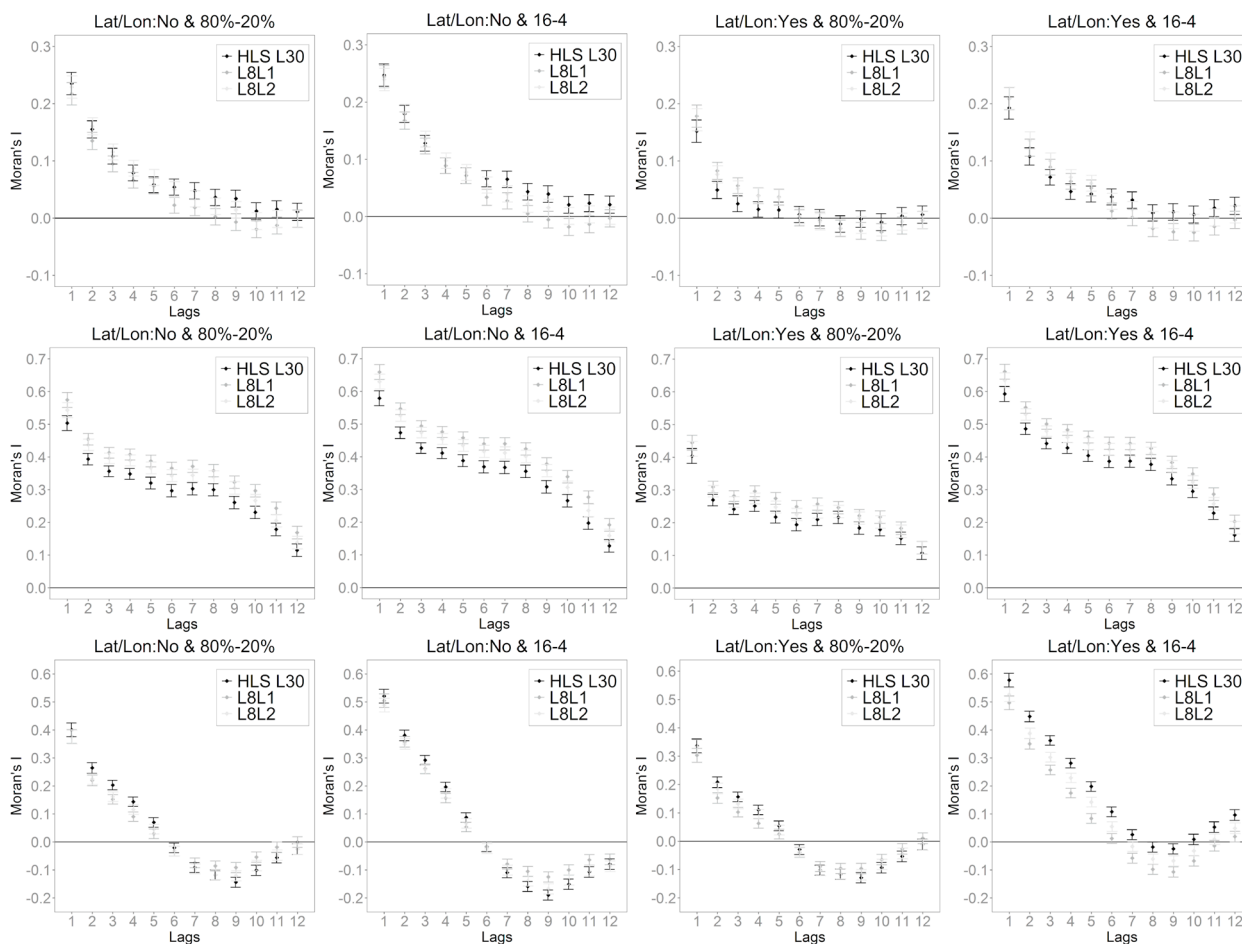


**Figure 5.** Sensitivity analysis of RF-dominant canopy height prediction to sample size for L8L1, L2L1, and HLS L30 products. In the left column, 80% of the sampled observations from the 20 airborne lidar transects were used for training and 20% for testing the RF (random split 80%-20%). In the right column, the sampled observations from 16 transects were used for training and 4 for testing the RF (spatial split 16-4). Latitude and longitude were not included (Lat/Lon: No) in the first row and were included (Lat/Lon: Yes) in the second row. Each dot is the median mean error (ME), and error bars are 95% confidence interval.

### 3.2. Impacts of Sample Spatial Autocorrelation on Random Forest DH Predictions

The residuals of RF models were not independent among the samples in most cases, according to the correlograms in Figure 6. About half the airborne lidar transects (e.g., Transect A in the first row of Figure 6) presented low positive spatial autocorrelation on residuals (Moran's  $I < 0.2$ ) when pixels were closer to each other (i.e., short lags), and occasionally they were perfectly random (Moran's  $I = 0$ ) when pixels were far away from each other. Other airborne lidar transects (e.g., Transect B, second row of Figure 6) presented a decreasing positive spatial autocorrelation (Moran's  $I$  varying from 0.7 to 0.1) and up to lag 12

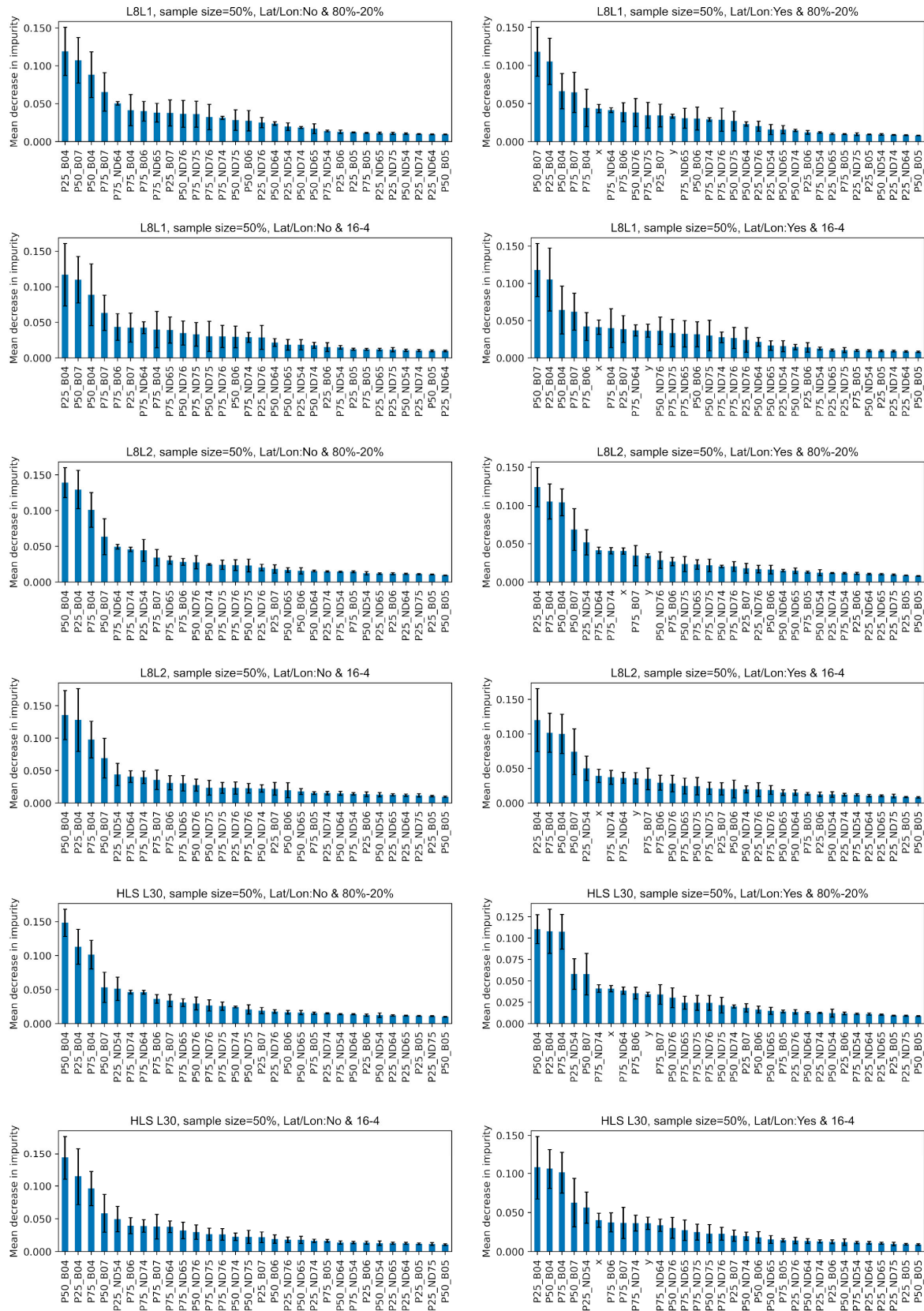
(i.e., 1080 m); residuals never reached perfect randomness. A few airborne lidar transects (e.g., Transect C, third row of Figure 6) presented a decreasing positive spatial autocorrelation, reached spatial randomness at lag 6, and from lag 7 onwards, presented negative spatial autocorrelation. Overall, spatial autocorrelation in residuals was reduced when RFs were modeled using an 80%-20% random split and included latitude and longitude. The inclusion of geographic coordinates tended to reduce the Moran's  $I$  at any given lag when RFs were modeled using an 80%-20% random split; however, no improvement has been identified when modeling using a 16-4 spatial split. Note that Transects A, B, and C are labeled in Figure 1.



**Figure 6.** Spatial autocorrelation on the average residuals of 300 RF models using L8L1, L8L2, and HLS L30 products at Transects A (top row), B (middle row), and C (bottom row) (see Figure 1). The sample size of the RF models was 50%. Results are displayed in terms of training/test strategy and inclusion or not of latitude and longitude as independent variables. Latitude and longitude were not included (Lat/Lon: No) in the first and second rows and were included (Lat/Lon: Yes) in the third and fourth rows. Each lag indicated a distance corresponding to a 90 m pixel times the lag value (i.e., 90–1080 m range).

### 3.3. Important Independent Variables in Random Forest

The independent variables most important to the RF models are illustrated in Figure 7. In all scenarios, such as the 80%-20% random split with and without latitude and longitude and the 16-4 spatial split with and without latitude and longitude, different percentiles of B04 and B07 became the most important variables for DH modeling. In the case of RF models using L8L1, B07 presents greater importance than RF models using L8L2 and HLS L30. Latitude and longitude, when included, demonstrated a moderate level of importance, falling between the 5th and 11th positions.



**Figure 7.** Variable importance for RF models using L8L1, L8L2, and HLS L30. The error bars are the standard deviation of each variable in the 300 RF models. The sample size was 50%. Results were divided by split strategy (i.e., 80%–20% random or 16–4 spatial). When included as independent variables in the RF, latitude is represented by *y* and longitude by *x*.

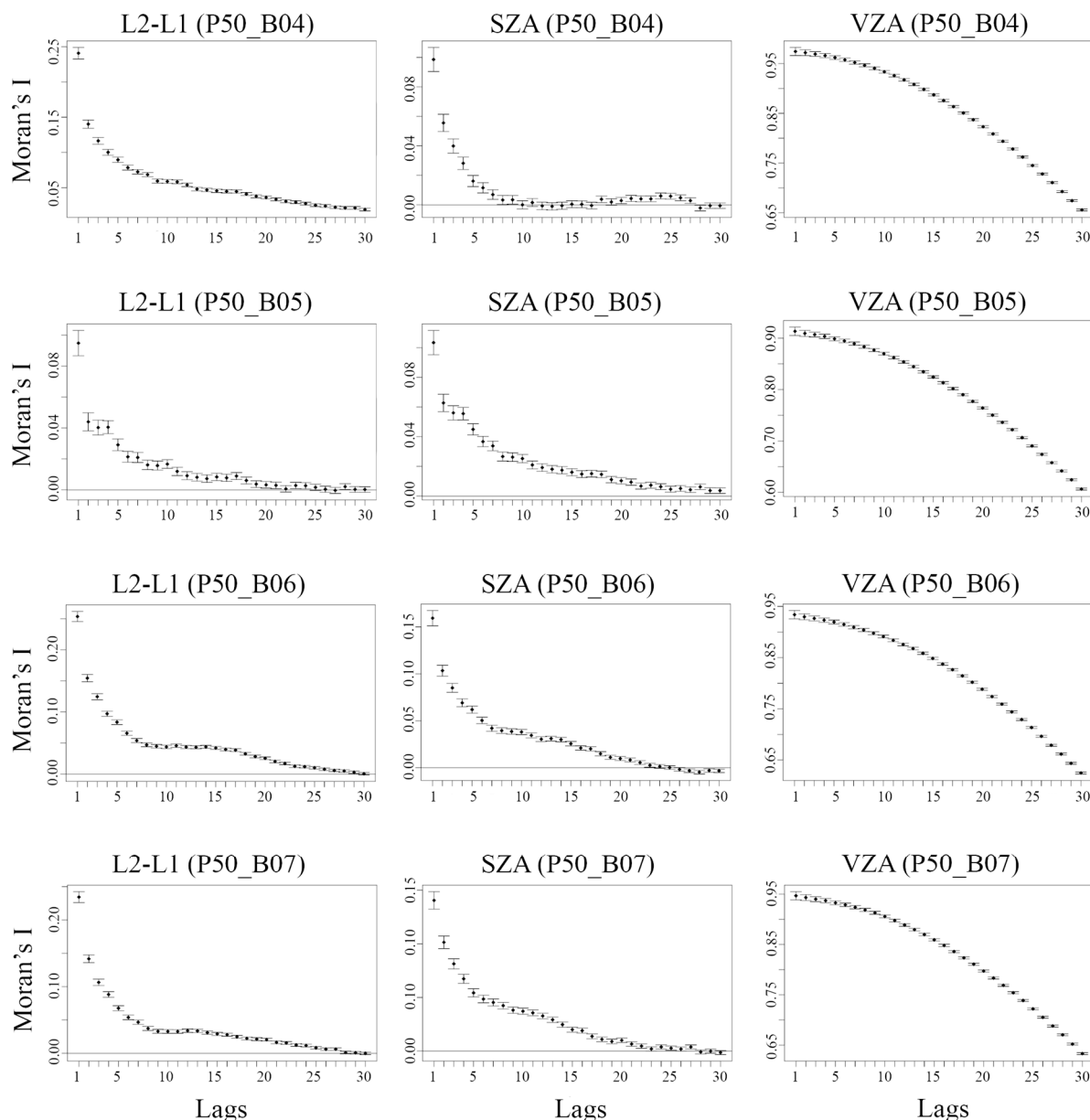
#### 4. Discussion

A series of experiments were undertaken to predict the DH at 90 m spatial resolution for 20 airborne lidar transects in Acre state, western Brazilian Amazon, using the L8L1, L8L2, and HLS L30 images and an RF machine learning model. The lidar-derived DH data were used as dependent variables, and temporal metrics derived from the three Landsat reference products (level 1 top-of-atmosphere reflectance, level 2 surface reflectance, and HLS NBAR) from 2016 to 2020 were used as independent variables. The experiments allowed not only the evaluation of different Landsat products but also the investigation of differences between RFs using different training and test split strategies, different sample sizes, and the impact of geolocation as independent variables in the RF models. The RF model performance was evaluated based on RMSE and ME values.

Two training/test split strategies, i.e., the random split (80%-20%) and the spatial split (16-4), were adopted before modeling DH. The  $R^2$  in models using an 80%-20% random split was higher than those using a 16-4 spatial split. In addition, RMSE and RMSE variability were lower in the models using an 80%-20% random split compared to the models using a 16-4 spatial split, showing an agreement with previous studies [27,51,52]. However, this higher accuracy was attributed to the presence of spatial autocorrelation between training and test observations in the 80%-20% random split. Because the airborne lidar transects were not able to provide spatially well-distributed training samples in canopy height estimation, the spatial split (16-4 split) was recommended since it reflected the accuracy in practical applications.

The geographic coordinates (i.e., latitude and longitude) were included as independent variables to evaluate whether they would incorporate spatial dependence in the RF models. The  $R^2$  increased in all experiments that included them as independent variables. The RMSE also decreased for all the Landsat products when using the 80%-20% random split strategy but remained the same when using the 16-4 spatial split. These results indicated that including geolocation did not significantly improve the accuracy in practical applications when the training data were far away from the data to be tested as the 16-4 spatial split. This was because the geographic coordinates had different data ranges between training and test data, i.e., to predict DH for data with different distributions from the training. Unless the training samples were randomly and well distributed in the study area, the analysis supports the idea that the spatial split reflects more reliable model accuracy in practical applications.

The ME results in DH predictions were also impacted by spatial autocorrelation. While MEs showed mostly no bias in predictions for any sample sizes in RFs using 80%-20% random split and using 16-4 spatial split without geographic coordinates, they did show biased predictions in RFs using sample sizes  $\geq 30\%$  and 16-4 spatial split with geographic coordinates for L8L2 and HLS L30 products. One possible explanation was that both atmospheric correction and BRDF correction reduced some degree of spatial autocorrelation intrinsic to the temporal metrics of the independent variables. For example, Figure 8 illustrates the spatial autocorrelation in the temporal metrics of L8L1 product red, NIR, and SWIR bands, which were not BRDF-corrected nor atmospherically corrected, for the Transect A area. All the Moran's  $I$  correlograms presented some degree of spatial autocorrelation. As the lag increased, both the solar zenith angle (SZA) and the difference between the L8L2 and L8L1 (L2-L1), which was assumed to account for the atmospheric effects in reflectance values, tended to reach spatial randomness (Moran's  $I = 0$ ) in all the bands. This implied an absence of spatial autocorrelation at those lags/distances. The VZA, on the other hand, was highly spatially autocorrelated and may only reach spatial randomness at very far lags/distances not represented in the correlograms. The correlogram patterns were similar for the 25th and 75th percentiles and at the areas around the other airborne lidar transects.



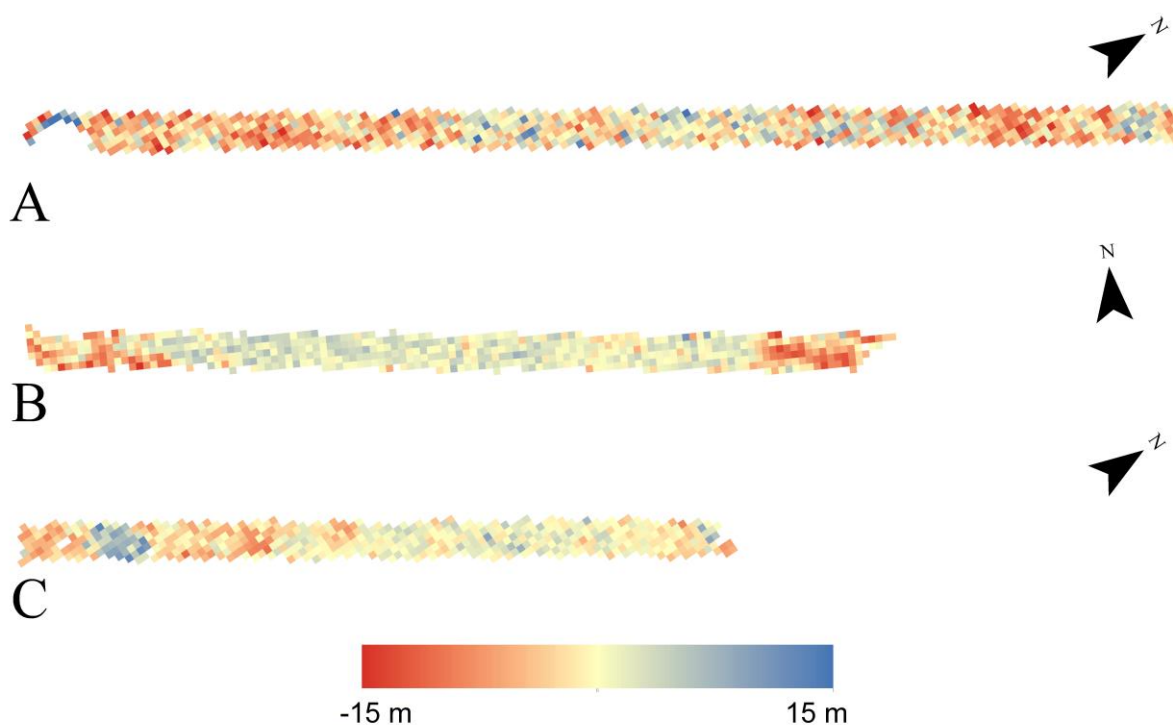
**Figure 8.** Spatial autocorrelation intrinsic to L8L1 temporal metrics product. In the left column are the Moran's  $I$  correlograms showing the difference between the L8L2 and L8L1 (L2-L1) red, NIR, and SWIR bands. In the middle column is the solar zenith angle (SZA) and in the right column is the view zenith angle (VZA) red, NIR, and SWIR band correlograms. Here, only correlograms representing the 50th percentile (P50) of temporal metrics for Transect A are included. Each lag indicates a distance corresponding to a 90 m pixel times the lag value (i.e., 90–2700 m range).

It is worth noting that RFs trained with both L8L1 and L8L2 temporal metrics misleadingly performed better than the ones trained with HLS L30 when using an 80%-20% random split strategy (Figures 3 and 4). VZA's strong spatial autocorrelation effect (Figure 7) likely improved the  $R^2$  and deflated the RMSEs of the RFs trained with L8L1 and L8L2 temporal metrics. The accuracy of RFs trained with HLS L30 data is, therefore, more credible in practical applications.

The residuals of RF models were independent of one another only at a few specific spatial lags and at a few airborne lidar transects. The shape of the Moran's  $I$  correlograms was strongly influenced by the RF-predicted DHs, which tended to be overestimated (blue colors in Figure 9) in areas where ground-truth DH was low, such as short forest canopies



or deforested areas, and underestimated (red colors in Figure 9) in areas with tall trees where ground-truth DH was high. Transects with spatially well-distributed residuals (i.e., predicted–ground truth), like Transect A, tended to present smoother correlograms, with Moran's  $I$  regularly lower than 0.3 and gently decreasing with an increasing lag. The presence of clusters of low or high residuals in the transects, like those in Transect B and Transect C, led to greater Moran's  $I$  at short lags, and it was less likely that their residuals would become spatially random at any lag. Including latitude and longitude as independent variables was incapable of handling the spatial autocorrelation in the residuals, confirming that geographic coordinates poorly contributed to incorporating spatial dependence into models [25,53]. Furthermore, the correlograms showed that the 16-4 spatial split strategy tended to present stronger spatial autocorrelation in the residuals in almost all the lags compared with the 80%-20% random split. This result was likely attributed to the fact that our RF models only incorporated the reflectance of red, NIR, and SWIR bands, along with a combination of their ND, as independent variables. The omission of crucial environmental independent variables that could contribute to explaining the variability of the dependent variable, coupled with the failure to account for spatial dependence, likely contributed to the stronger residuals' spatial autocorrelation in the 16-4 spatial split strategy.



**Figure 9.** Spatial distribution of the residuals (i.e., predicted–ground truth) of the RF models within Transect A, Transect B, and Transect C. Transects have a 90 m pixel size.

This work has some limitations, such as sample distribution and representativeness, as well as mismatched acquisition time periods between lidar data and Landsat temporal metrics. Specifically, this study used data from 20 airborne lidar transects, each approximately 12.5 km by 300 m in size. The RF results in the areas without airborne lidar data could potentially cause under- or overestimations of DH, particularly if very short and very tall canopies were misrepresented. Additionally, mismatched acquisition time periods between lidar and Landsat data could impact the RF results, especially if lidar data were collected over forested areas that were subsequently deforested or over deforested areas that later regenerated. Despite these limitations, this study successfully identified how the different Landsat reflectance products and the sample spatial autocorrelation using different sampling strategies influence the RF modeling of canopy heights in the Brazilian Amazon.

For future research endeavors focused on addressing spatial autocorrelation in modeling canopy heights through remotely sensed data, we suggest considering the use of geostatistical or autoregressive models. In addition, we suggest the use of the Getis statistic [54], as demonstrated in previous studies [55,56]. This statistic not only offers a measure of spatial dependence for each pixel but also indicates the relative magnitudes of the digital numbers in the pixel's neighborhood. Alternatively, for those interested in utilizing RF for canopy height modeling, Moran's eigenvector maps [49,57] can be considered, which account for the spatial structure of ecosystems. An analysis like the one conducted in this study is also advised for those who are interested in using deep learning algorithms [58]. The use of deep learning algorithms in forest applications has been thoroughly reviewed by [59]. Furthermore, we recommend the inclusion of additional variables, such as number of cloud-free days, soil clay content, elevation, annual precipitation, and precipitation seasonality [60], in forthcoming investigations.

## 5. Conclusions

Canopy heights derived from airborne discrete return lidar and time-series Landsat reflectance products using machine learning models are impacted by data pre-processing and spatial autocorrelation. This study examined the differences between predicted dominant canopy heights (DHs) that were modeled with Landsat top-of-atmosphere reflectance (L8L1), surface reflectance (L8L2), and HLS NBAR as independent variables. The study also examined the sensitivity to training and test split strategy and the number of training samples. Results indicated that the use of a large training sample size had the anticipated effect of improving RF DH prediction accuracy when using the 80%-20% random split, but no improvement was found using a 16-4 spatial split. RF models using L8L1 and L8L2 temporal metrics tended to yield lower errors compared to models using HLS L30 in the 80%-20% random split, but their difference was subtle in the RF models trained using the 16-4 spatial split strategy. The model differences among different input Landsat data in the 80%-20% random split were attributed to the training and test data correlation, e.g., the strong spatial autocorrelation of VZA found in both L8L1 and L8L2 products. It was also observed that RF models consistently exhibited lower errors and maintained unbiased performance when employing the 80%-20% random split. Similarly, this outcome was likely attributed to the presence of spatial autocorrelation between the training and test data. In contrast, when a 16-4 spatial split was used, higher errors were evident, and bias tended to occur when geographic coordinates were included, presumably due to the reduction or absence of spatial autocorrelation between training and test data. The inclusion of geographic coordinates also improved the accuracy of RF models using an 80%-20% random split. However, no such improvement was found in models using a 16-4 spatial split. It is essential to note that the inclusion of latitude and longitude did not effectively handle spatial autocorrelation in the RF model residuals. Furthermore, the spatial autocorrelation of RF model residuals was likely influenced by the presence of short forest canopies, deforested areas, or patches of tall trees within regions or transects. In summary, this study provided evidence for favoring NBAR data over TOA or surface reflectance products when modeling canopy heights in the Brazilian Amazon biome. It also underscores the influence of spatial autocorrelation on RF models, and we do not recommend the use of an 80%-20% random split for small study areas.

**Author Contributions:** Conceptualization, P.V.C.O.; methodology, P.V.C.O. and H.K.Z.; formal analysis, P.V.C.O., H.K.Z. and X.Z.; writing—original draft, P.V.C.O., H.K.Z. and X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the U.S. Geological Survey Earth Resources Observation and Science Center, grant number G19AS00001 and the NASA grant number 80NSSC21K1962.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

**Acknowledgments:** The authors thank NASA’s Land Processes Distributed Active Archive Center (LP DAAC) for providing the Landsat and HLS data. The authors also thank Jean P. H. B. Ometto and the Amazon Biomass Estimate (EBA) project for providing the airborne lidar dataset.

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

## References

- Gibbs, H.K.; Brown, S.; Niles, J.O.; Foley, J.A. Monitoring and Estimating Tropical Forest Carbon Stocks: Making REDD a Reality. *Environ. Res. Lett.* **2007**, *2*, 045023. [\[CrossRef\]](#)
- Ometto, J.P.; Aguiar, A.P.; Assis, T.; Soler, L.; Valle, P.; Tejada, G.; Lapola, D.M.; Meir, P. Amazon Forest Biomass Density Maps: Tackling the Uncertainty in Carbon Emission Estimates. *Clim. Chang.* **2014**, *124*, 545–560. [\[CrossRef\]](#)
- Ometto, J.P.; Gorgens, E.B.; de Souza Pereira, F.R.; Sato, L.; de Assis, M.L.R.; Cantinho, R.; Longo, M.; Jacon, A.D.; Keller, M. A biomass map of the Brazilian Amazon from multisource remote sensing. *Sci. Data* **2023**, *10*, 668. [\[CrossRef\]](#) [\[PubMed\]](#)
- Aragão, L.E.O.C.; Anderson, L.O.; Fonseca, M.G.; Rosan, T.M.; Vedovato, L.B.; Wagner, F.H.; Silva, C.V.J.; Silva Junior, C.H.L.; Arai, E.; Aguiar, A.P.; et al. 21st Century Drought-Related Fires Counteract the Decline of Amazon Deforestation Carbon Emissions. *Nat. Commun.* **2018**, *9*, 536. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fearnside, P.M. Greenhouse gases from deforestation in Brazilian Amazonia: Net committed emissions. *Clim. Chang.* **1997**, *35*, 321–360. [\[CrossRef\]](#)
- Foley, J.A.; Asner, G.P.; Costa, M.H.; Coe, M.T.; DeFries, R.; Gibbs, H.K.; Howard, E.A.; Olson, S.; Patz, J.; Ramankutty, N.; et al. Amazonia revealed: Forest degradation and loss of ecosystem goods and services in the Amazon Basin. *Front. Ecol. Environ.* **2007**, *5*, 25–32. [\[CrossRef\]](#)
- Nogueira, E.M.; Myanai, A.; Fonseca, F.O.R.; Fearnside, P.M. Carbon Stock Loss from Deforestation through 2013 in Brazilian Amazonia. *Glob. Chang. Biol.* **2015**, *21*, 1271–1292. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M.S.; Delitti, W.B.C.; Duque, A.; Eid, T.; Fearnside, P.M.; Goodman, R.C.; et al. Improved Allometric Models to Estimate the Aboveground Biomass of Tropical Trees. *Glob. Chang. Biol.* **2014**, *20*, 3177–3190. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jucker, T.; Caspersen, J.; Chave, J.; Antin, C.; Barbier, N.; Bongers, F.; Dalponte, M.; van Ewijk, K.Y.; Forrester, D.I.; Haeni, M.; et al. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Glob. Chang. Biol.* **2017**, *23*, 177–190. [\[CrossRef\]](#)
- Disney, M.; Boni Vicari, M.; Burt, A.; Calders, K.; Lewis, S.L.; Raunonen, P.; Wilkes, P. Weighing trees with lasers: Advances, challenges and opportunities. *Interface Focus* **2018**, *8*, 20170048. [\[CrossRef\]](#)
- Duncanson, L.; Kellner, J.R.; Armston, J.; Dubayah, R.; Minor, D.M.; Hancock, S.; Healey, S.P.; Patterson, P.L.; Saarela, S.; Marselis, S.; et al. Aboveground Biomass Density Models for NASA’s Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sens. Environ.* **2022**, *270*, 112845. [\[CrossRef\]](#)
- Silva, C.A.; Duncanson, L.; Hancock, S.; Neuenschwander, A.; Thomas, N.; Hofton, M.; Fatoyinbo, L.; Simard, M.; Marshak, C.Z.; Armston, J.; et al. Fusing Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping. *Remote Sens. Environ.* **2021**, *253*, 112234. [\[CrossRef\]](#)
- Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global Canopy Height Regression and Uncertainty Estimation from GEDI LIDAR Waveforms with Deep Ensembles. *Remote Sens. Environ.* **2022**, *268*, 112760. [\[CrossRef\]](#)
- Lefsky, M.A. A Global Forest Canopy Height Map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophys. Res. Lett.* **2010**, *37*, L15401. [\[CrossRef\]](#)
- Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping Forest Canopy Height Globally with Spaceborne Lidar. *J. Geophys. Res. Biogeosci.* **2011**, *116*, G04021. [\[CrossRef\]](#)
- Los, S.O.; Rosette, J.A.B.; Kljun, N.; North, P.R.J.; Chasmer, L.; Suárez, J.C.; Hopkinson, C.; Hill, R.A.; Van Gorsel, E.; Mahoney, C.; et al. Vegetation Height and Cover Fraction between 60° S and 60° N from ICESat GLAS Data. *Geosci. Model Dev.* **2012**, *5*, 413–432. [\[CrossRef\]](#)
- Schutz, B.E.; Zwally, H.J.; Shuman, C.A.; Hancock, D.; DiMarzio, J.P. Overview of the ICESat Mission. *Geophys. Res. Lett.* **2005**, *32*, L21S01. [\[CrossRef\]](#)
- Ahmed, O.S.; Franklin, S.E.; Wulder, M.A.; White, J.C. Characterizing Stand-Level Forest Canopy Cover and Height Using Landsat Time Series, Samples of Airborne LiDAR, and the Random Forest Algorithm. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 89–101. [\[CrossRef\]](#)
- Hansen, M.C.; Potapov, P.V.; Goetz, S.J.; Turubanova, S.; Tyukavina, A.; Krylov, A.; Kommareddy, A.; Egorov, A. Mapping Tree Height Distributions in Sub-Saharan Africa Using Landsat 7 and 8 Data. *Remote Sens. Environ.* **2016**, *185*, 221–232. [\[CrossRef\]](#)
- Ota, T.; Ahmed, O.S.; Franklin, S.E.; Wulder, M.A.; Kajisa, T.; Mizoue, N.; Yoshida, S.; Takao, G.; Hirata, Y.; Furuya, N.; et al. Estimation of Airborne Lidar-Derived Tropical Forest Canopy Height Using Landsat Time Series in Cambodia. *Remote Sens.* **2014**, *6*, 10750–10772. [\[CrossRef\]](#)
- Potapov, P.V.; Turubanova, S.A.; Hansen, M.C.; Adusei, B.; Broich, M.; Altstatt, A.; Mane, L.; Justice, C.O. Quantifying Forest Cover Loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ Data. *Remote Sens. Environ.* **2012**, *122*, 106–116. [\[CrossRef\]](#)

22. Cliff, A.D.; Ord, J.K. *Spatial Processes: Models & Applications*; Pion: London, UK, 1981.
23. Legendre, P. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* **1993**, *74*, 1659–1673. [[CrossRef](#)]
24. Lennon, J.J. Red-Shifts and Red Herrings in Geographical Ecology. *Ecography* **2000**, *23*, 101–113. [[CrossRef](#)]
25. Mascaro, J.; Asner, G.P.; Knapp, D.E.; Kennedy-Bowdoin, T.; Martin, R.E.; Anderson, C.; Higgins, M.; Chadwick, K.D. A Tale of Two “Forests”: Random Forest Machine Learning Aids Tropical Forest Carbon Mapping. *PLoS ONE* **2014**, *9*, e85993. [[CrossRef](#)]
26. Xu, L.; Saatchi, S.S.; Yang, Y.; Yu, Y.; White, L. Performance of Non-Parametric Algorithms for Spatial Mapping of Tropical Forest Structure. *Carbon Balance Manag.* **2016**, *11*, 18–20. [[CrossRef](#)] [[PubMed](#)]
27. Karasiak, N.; Dejoux, J.F.; Monteil, C.; Sheeren, D. Spatial Dependence between Training and Test Sets: Another Pitfall of Classification Accuracy Assessment in Remote Sensing. *Mach. Learn.* **2022**, *111*, 2715–2740. [[CrossRef](#)]
28. IBGE. *Manual Técnico da Vegetação Brasileira*, 2nd ed.; IBGE: Rio de Janeiro, Brazil, 2012; ISBN 9788524042720.
29. Saleska, S.R.; Wu, J.; Guan, K.Y.; Araujo, A.C.; Huete, A.; Nobre, A.D.; Restrepo-Coupe, N. Dry-season greening of Amazon forests. *Nature* **2016**, *531*, E4–E5. [[CrossRef](#)] [[PubMed](#)]
30. Ometto, J.P.; Gorgens, B.G.; Assis, M.; Cantinho, R.Z.; Pereira, F.R.d.S.; Sato, L.Y. L3A—Summary of Airborne LiDAR Transects Collected by EBA in the Brazilian Amazon (Version 20210616) [Data Set]. *Zenodo* 2021. Available online: <https://zenodo.org/records/4968706> (accessed on 25 June 2024). [[CrossRef](#)]
31. Kotchenova, S.Y.; Vermote, E.F.; Matarrese, R.; Klemm, F.J. Validation of a Vector Version of the 6S Radiative Transfer Code for Atmospheric Correction of Satellite Data. Part I: Path Radiance. *Appl. Opt.* **2006**, *45*, 6762–6774. [[CrossRef](#)] [[PubMed](#)]
32. Asner, G.P. Cloud cover in Landsat observations of the Brazilian Amazon. *Int. J. Remote Sens.* **2001**, *22*, 3855–3862. [[CrossRef](#)]
33. Crawford, C.J.; Roy, D.P.; Arab, S.; Barnes, C.; Vermote, E.; Hulley, G.; Gerace, A.; Choate, M.; Engebretson, C.; Micijevic, E.; et al. The 50-Year Landsat Collection 2 Archive. *Sci. Remote Sens.* **2023**, *8*, 100103. [[CrossRef](#)]
34. Roy, D.P.; Zhang, H.K.; Ju, J.; Gomez-Dans, J.L.; Lewis, P.E.; Schaaf, C.B.; Sun, Q.; Li, J.; Huang, H.; Kovalsky, V. A General Method to Normalize Landsat Reflectance Data to Nadir BRDF Adjusted Reflectance. *Remote Sens. Environ.* **2016**, *176*, 255–271. [[CrossRef](#)]
35. Li, Z.; Zhang, H.K.; Roy, D.P. Investigation of Sentinel-2 bidirectional reflectance hot-spot sensing conditions. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 3591–3598. [[CrossRef](#)]
36. Masek, J.; Ju, J.; Roger, J.; Skakun, S.; Vermote, E.; Claverie, M.; Dungan, J.; Yin, Z.; Freitag, B.; Justice, C. HLS Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m (v2.0) [Data Set]. NASA EOSDIS Land Processes DAAC 2021. Available online: <https://lpdaac.usgs.gov/products/hls30v002/> (accessed on 25 June 2024). [[CrossRef](#)]
37. Qiu, S.; Zhu, Z.; He, B. Fmask 4.0: Improved Cloud and Cloud Shadow Detection in Landsats 4–8 and Sentinel-2 Imagery. *Remote Sens. Environ.* **2019**, *231*, 111205. [[CrossRef](#)]
38. DeFries, R.; Hansen, M.; Townshend, J. Global Discrimination of Land Cover Types from Metrics Derived from AVHRR Pathfinder Data. *Remote Sens. Environ.* **1995**, *54*, 209–222. [[CrossRef](#)]
39. Zhang, H.K.; Roy, D.P. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sens. Environ.* **2017**, *197*, 15–34. [[CrossRef](#)]
40. McGaughey, R.J. *FUSION/LDV: Software for LIDAR Data Analysis and Visualization v. 4.50*; United States Department of Agriculture: Seattle, WA, USA, 2023.
41. Kraus, K.; Pfeifer, N. Determination of Terrain Models in Wooded Areas with Airborne Laser Scanner Data. *ISPRS J. Photogramm. Remote Sens.* **1998**, *53*, 193–203. [[CrossRef](#)]
42. Coops, N.C.; Hilker, T.; Wulder, M.A.; St-Onge, B.; Newnham, G.; Siggins, A.; Trofymow, J.A. Estimating Canopy Structure of Douglas-Fir Forest Stands from Discrete-Return LiDAR. *Trees - Struct. Funct.* **2007**, *21*, 295–310. [[CrossRef](#)]
43. Cochran, W.G. *Sampling Techniques*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1977.
44. Foody, G.M. Status of Land Cover Classification Accuracy Assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [[CrossRef](#)]
45. Rogan, J.; Franklin, J.; Stow, D.; Miller, J.; Woodcock, C.; Roberts, D. Mapping Land-Cover Modifications over Large Areas: A Comparison of Machine Learning Algorithms. *Remote Sens. Environ.* **2008**, *112*, 2272–2283. [[CrossRef](#)]
46. Yan, L.; Roy, D.P. Improved Time Series Land Cover Classification by Missing-Observation-Adaptive Nonlinear Dimensionality Reduction. *Remote Sens. Environ.* **2015**, *158*, 478–491. [[CrossRef](#)]
47. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
48. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
49. Legendre, P.; Legendre, L. *Numerical Ecology*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 1998.
50. Diniz-Filho, J.A.F.; Bini, L.M.; Hawkins, B.A. Spatial Autocorrelation and Red Herrings in Geographical Ecology. *Glob. Ecol. Biogeogr.* **2003**, *12*, 53–64. [[CrossRef](#)]
51. Lyons, M.B.; Keith, D.A.; Phinn, S.R.; Mason, T.J.; Elith, J. A Comparison of Resampling Methods for Remote Sensing Classification and Accuracy Assessment. *Remote Sens. Environ.* **2018**, *208*, 145–153. [[CrossRef](#)]
52. Schratz, P.; Muenchow, J.; Iturrutxa, E.; Richter, J.; Brenning, A. Hyperparameter Tuning and Performance Assessment of Statistical and Machine-Learning Algorithms Using Spatial Data. *Ecol. Modell.* **2019**, *406*, 109–120. [[CrossRef](#)]
53. Miller, J.; Franklin, J.; Aspinall, R. Incorporating Spatial Dependence in Predictive Vegetation Models. *Ecol. Modell.* **2007**, *202*, 225–242. [[CrossRef](#)]
54. Getis, A.; Ord, J.K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* **1992**, *24*, 189–206. [[CrossRef](#)]

55. Wulder, M.; Boots, B. Local Spatial Autocorrelation Characteristics of Remotely Sensed Imagery Assessed with the Getis Statistic. *Int. J. Remote Sens.* **1998**, *19*, 2223–2231. [[CrossRef](#)]
56. Ghimire, B.; Rogan, J.; Miller, J. Contextual Land-Cover Classification: Incorporating Spatial Dependence in Land-Cover Classification Models Using Random Forests and the Getis Statistic. *Remote Sens. Lett.* **2010**, *1*, 45–54. [[CrossRef](#)]
57. Borcard, D.; Legendre, P. All-Scale Spatial Analysis of Ecological Data by Means of Principal Coordinates of Neighbour Matrices. *Ecol. Modell.* **2002**, *153*, 51–68. [[CrossRef](#)]
58. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
59. Yun, T.; Li, J.; Ma, L.; Zhou, J.; Wang, R.; Eichhorn, M.P.; Zhang, H. Status, advancements and prospects of deep learning methods applied in forest studies. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103938. [[CrossRef](#)]
60. Gorgens, E.B.; Nunes, M.H.; Jackson, T.; Coomes, D.; Keller, M.; Reis, C.R.; Valbuena, R.; Rosette, J.; de Almeida, D.R.A.; Gimenez, B.; et al. Resource Availability and Disturbance Shape Maximum Tree Height across the Amazon. *Glob. Chang. Biol.* **2021**, *27*, 177–189. [[CrossRef](#)] [[PubMed](#)]

**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.