

Article

The Transferability of Random Forest in Canopy Height Estimation from Multi-Source Remote Sensing Data

Shichao Jin ^{1,2,†}, Yanjun Su ^{1,2,3,†}, Shang Gao ^{1,2}, Tianyu Hu ¹, Jin Liu ¹ and Qinghua Guo ^{1,2,3,*}

¹ State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; jinshichao1993@gmail.com (S.J.); suyanjun1987@gmail.com (Y.S.); gaoshang@ibcas.ac.cn (S.G.); safarihu@gmail.com (T.H.); liujing1030@ibcas.ac.cn (J.L.)

² University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

³ Sierra Nevada Research Institute, School of Engineering, University of California Merced, Merced, CA 95343, USA

* Correspondence: guo.qinghua@gmail.com; Tel.: +86-010-6283-6157

† These authors contributed equally to this work.

Received: 15 May 2018; Accepted: 24 July 2018; Published: 26 July 2018



Abstract: Canopy height is an important forest structure parameter for understanding forest ecosystem and improving global carbon stock quantification accuracy. Light detection and ranging (LiDAR) can provide accurate canopy height measurements, but its application at large scales is limited. Using LiDAR-derived canopy height as ground truth to train the Random Forest (RF) algorithm and therefore predict canopy height from other remotely sensed datasets in areas without LiDAR coverage has been one of the most commonly used method in large-scale canopy height mapping. However, how variances in location, vegetation type, and spatial scale of study sites influence the RF modelling results is still a question that needs to be addressed. In this study, we selected 16 study sites (100 km² each) with full airborne LiDAR coverage across the United States, and used the LiDAR-derived canopy height along with optical imagery, topographic data, and climate surfaces to evaluate the transferability of the RF-based canopy height prediction method. The results show a series of findings from general to complex. The RF model trained at a certain location or vegetation type cannot be transferred to other locations or vegetation types. However, by training the RF algorithm using samples from all sites with various vegetation types, a universal model can be achieved for predicting canopy height at different locations and different vegetation types with self-predicted R^2 higher than 0.6 and RMSE lower than 6 m. Moreover, the influence of spatial scales on the RF prediction accuracy is noticeable when spatial extent of the study site is less than 50 km² or the spatial resolution of the training pixel is finer than 500 m. The canopy height prediction accuracy increases with the spatial extent and the targeted spatial resolution.

Keywords: canopy height; Random Forest; LiDAR; multi-source; vegetation type; location; scale

1. Introduction

Accurate prediction of forest characteristics is vital for forest ecosystem management, which serves a range of functions to the Earth's life supporting system [1]. Moreover, as a carbon sink, forest dynamics are closely related to the global carbon cycle and climate changes [2–4]. Canopy height is an important forest structure parameter for understanding the forest ecosystem, and has been used to estimate forest aboveground biomass and therefore model the global carbon stock and carbon dynamics [5–8].

Traditionally, canopy height can be accurately measured in the field using range finders. However, taking field measurements is highly labor-intensive and time-consuming, which constrained its use in generating spatial continuous canopy height products at large scales. The emergence of remote sensing technology can largely help to solve this problem since it can significantly increase the efficiency and reduce the cost of obtaining large-scale land surface observations. Optical passive remote sensing and radar have been used to estimate forest canopy height at different locations and scales [4,9,10]. However, these derived canopy height products are usually fraught with “saturation” effects since optical and radar signals cannot penetrate forest canopies, which may result in canopy height being underestimated in areas dominated by tall trees [11,12].

Light detection and ranging (LiDAR), an active remote sensing technology, provides an alternative way to measure forest canopy height [13,14]. Through the use of a focused short-wavelength laser pulse, LiDAR can penetrate forest canopy effectively and therefore measure the forest three-dimensional structures [15]. There have been many studies showing that LiDAR can be used to estimate canopy height accurately at different locations and scales. For example, Su et al. estimated the wall-to-wall Lorey’s tree height distribution at 70 m resolution in Sierra Nevada, US with in situ tree height measurement, airborne LiDAR data, Geoscience Laser Altimeter System (GLAS) data, optical imagery, and other ancillary datasets. [14]. M. Clark, D. Clark, and Roberts [16] estimated sub-canopy tree height of a rainforest by using small-footprint LiDAR data; Alexander et al. [17] mapped the canopy height distribution of Denmark using airborne LiDAR data. However, the availability of LiDAR data in regional to global scales is limited. As one of the most frequently used LiDAR data, airborne LiDAR is only available in certain areas due to the high flight mission cost. The Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat), retired in 2009, is the only available spaceborne LiDAR data with global coverage. However, its footprints are about 170 m away from each other along the track and tens of kilometers away across tracks, which are too sparse to be used alone for large-scale canopy height mapping. The proposed Global Ecosystem Dynamics Investigation (GEDI) LiDAR system is designed for observing the 3D structure of the earth at fine spatial resolution (i.e., at the 500×500 m pixel level), which is not fine enough and still on the way (<https://gedi.umd.edu/>). How to integrate LiDAR to improve the accuracy of large-scale canopy height mapping is still an active field of research.

Recently, the integration of LiDAR data with other remotely sensed data for large-scale canopy height mapping has been increasingly investigated [14,18–21]. In these data fusion schemes, LiDAR data are usually used as the ground truth to train machine learning algorithms, and therefore predict canopy height from other remotely sensed variables. The large amount of training samples provided by LiDAR data can significantly improve the accuracy of canopy height mapping using machine learning algorithms [15]. For example, support vector machine (SVM) has been applied to estimated canopy height by fusing LiDAR and imagery data, which is powerful to handle multi-dimensional data and has effective generalization capabilities [22,23]. However, results of SVMs are not robust due to the parameter assignment issues [22]. Artificial neural network (ANN) is another kind of machine learning method, which has also been widely used in canopy height and biomass estimations [24]. Although ANN is able to handle non-linearity and non-normality issues, it is easy to be over-fitted when trained with too much data [25]. By contrast, Random Forest (RF) is robust to over-fitting, and has been used successfully in regression problems even with dozens of samples [26,27]. Moreover, RF is (1) not limited by the distribution of covariables; (2) it is not sensitive to outliers and noises; and (3) it is a nonlinear method, which can deal with high-dimensional data [28,29]. In all, RF has been proven to outperform other machine learning algorithms in mapping canopy height through the fusion of multi-source remotely sensed dataset [30].

Although the RF-based data fusion technique has been widely used, the influence of various vegetation types, locations and spatial scales of different study areas on the model transferability is rarely studied. A single regression tree model is usually built regardless the size and location of the study area, which assumes that the ecological process is not scale-dependent and shows

the same or similar patterns in different circumstances. However, that is often violated in real ecological processes [31,32]. Modelling accuracy of these ecological processes are directly affected by heterogeneity in forests [33], and usually in a nonlinear way [34]. Locations, vegetation types, and spatial scales all affect the ecological modelling accuracy. For example, Simard et al. [19] has proven that the accuracy of their canopy height product was lower in protected areas with homogenized low trees and flat terrains; Turner et al. [35] found that the extent and grain of a landscape influenced ecological measurements and landscape patterns.

Therefore, the problem that needs to be solved in this study is to analyze the transferability of RF on mapping canopy height under different vegetation types, site locations, and spatial scales (in both extent and resolution). Ultimately, this study wants to answer the questions whether we can use a universal RF model to predict canopy height at different locations with various vegetation types, and how do spatial scales affect a RF canopy height regression model. The results of this study can provide guidance for generating canopy height products from local to global scales through the fusion of LiDAR data and optical imagery.

2. Data

2.1. Study Area

This study chose 16 study sites across the United States (Figure 1a) based on the following four rules: (1) should be entirely covered by publicly available LiDAR datasets; (2) should be larger than 100 km²; (3) should be dominated (>50%) by one of the four selected vegetation types, which are Evergreen Needleleaf Forest (ENF), Evergreen Broadleaf Forest (EBF), Deciduous Broadleaf Forest (DBF), and Mixed Forest (MF); and (4) four sites of the each vegetation type should be distributed as far as possible to reduce spatial autocorrelation and cover more heterogeneity (e.g., elevation, slope, and canopy cover). Based on these rules, four sites for each vegetation type were selected and each site has an area of 100 km². The four ENF sites are concentrated in high-elevation (1391 m on average) mountainous areas, the four DBF sites are located in mid-elevation (327 m on average) hilly areas, and the MF and EBF sites are in low-elevation (107 m and 41 m on average, respectively) flat areas. The four study sites of each vegetation type have various vegetation conditions (from open to closed and from young to mature) to ensure the representativeness of the analysis results (Table 1).

2.2. Airborne LiDAR Data

Airborne LiDAR data can be used to estimate canopy height accurately at different spatial scales. In this study, we collected over 20 TB airborne LiDAR data from public data sources, such as Open Topography and U.S. Geological Survey, to determine the appropriate study sites. Within the 16 selected study sites, the LiDAR data were acquired by various sensors, such as Leica ALS series (Leica Geosystems AG, Switzerland), Riegl LMS-Q680 (Riegl Inc., USA), and Optech ALTM 213 (Teledyne Optech Inc., Canada). These sensors were mounted on airplanes (such as multiple Cessna Caravans 208B) flown at a height from 900 to 2438 m above the ground. Most of these data were acquired during the growing seasons between 2006 and 2012, with some data covering evergreen forests acquired in leaf-off seasons (Table 2). The pulse and scanning rates range from 38 to 115.6 kHz and from 46.7 to 115.6 Hz, respectively, and the mean point density ranges from 0.32 to 23.62 pts/m².

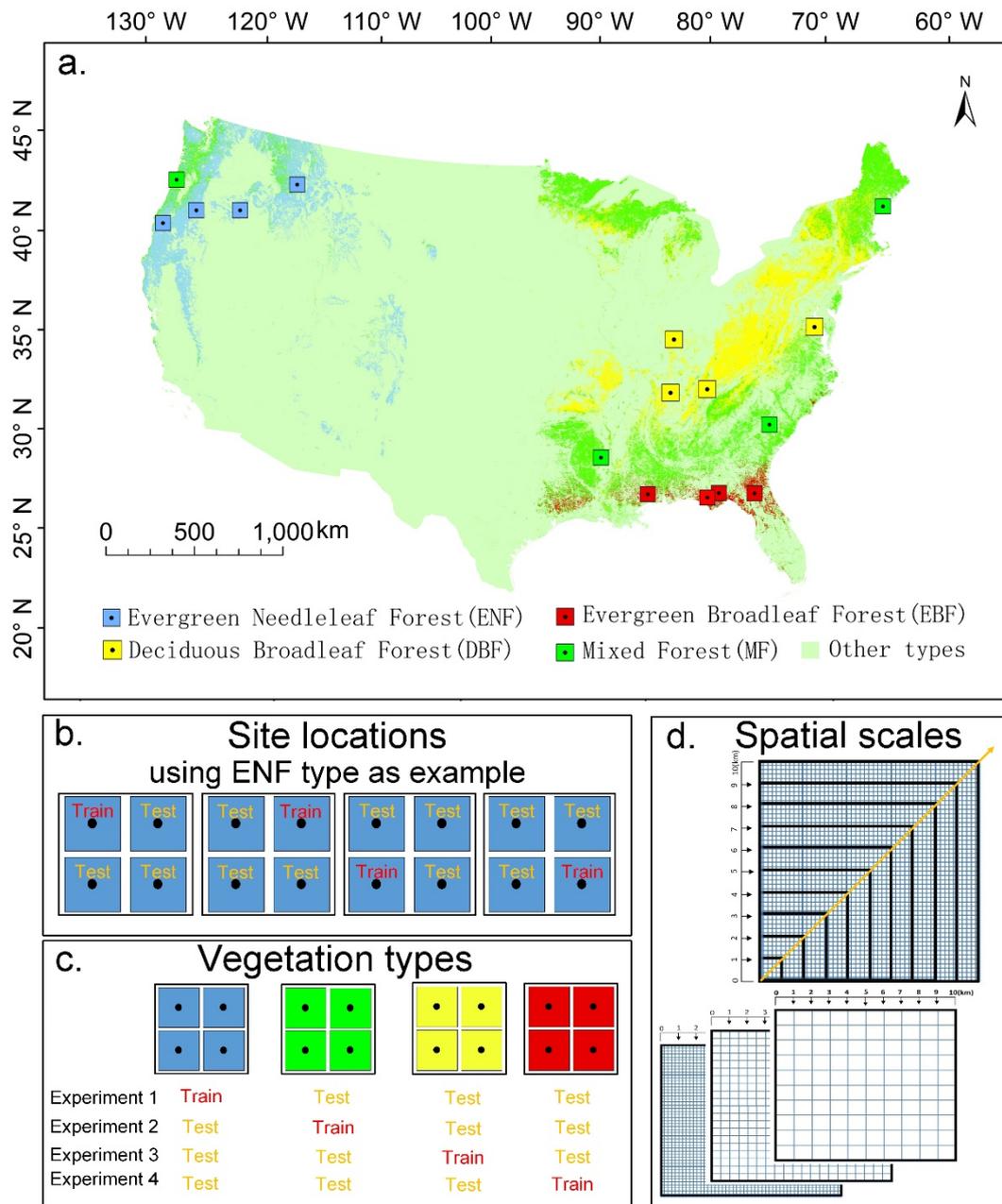


Figure 1. The study area and illustration of Random Forest (RF) model transferability evaluation among various geolocations, vegetation types, and spatial scales. (a) The geolocation of the 16 sites; (b) The RF model transferability evaluation among different study sites within the same vegetation type; (c) The RF model transferability evaluation among different vegetation types; (d) The RF model transferability evaluation among different spatial extents and spatial resolutions.

Table 1. The geolocations, terrain information, and vegetation information of the 16 study sites in this study. ENF1–4 represent the four Evergreen Needleleaf Forest sites, EBF1–4 represent the four Evergreen Broadleaf Forest sites, DBF1–4 represent the four Deciduous Broadleaf Forest sites, and MF1–4 represent the four Mixed Forest sites.

Study Site	Longitude (°)	Latitude (°)	Elevation (m)	Slope (°)	Area (km ²)	Percentage * (%)	Mean Tree Height (m)	Mean Canopy Cover	Forest Integrity
ENF1	−121.68	43.89	1631.69	6.49	96.87	100.00	6.11	0.67	unmanaged
ENF2	−118.51	44.54	1600.22	11.12	99.00	99.22	1.50	0.28	unmanaged
ENF3	−123.75	42.69	683.47	21.07	98.00	100.00	18.77	0.89	unmanaged
ENF4	−114.71	46.62	1789.64	15.56	100.00	88.95	4.23	0.38	unmanaged
EBF1	−84.60	30.54	51.24	3.37	90.12	75.84	13.54	0.86	unmanaged
EBF2	−85.34	30.38	28.81	1.99	100.93	89.58	16.36	0.95	unmanaged
EBF3	−88.97	30.88	39.16	2.50	95.66	72.75	15.78	0.86	unknown
EBF4	−82.43	30.25	45.36	2.44	100.00	99.52	6.77	0.57	unknown
DBF1	−76.54	38.53	47.38	5.02	100.00	80.70	13.46	0.62	unmanaged
DBF2	−86.92	36.25	217.52	11.08	100.00	100.00	26.09	0.91	unmanaged
DBF3	−86.29	39.13	593.27	17.69	98.27	100.00	10.86	0.82	unmanaged
DBF4	−84.46	36.21	242.08	8.36	100.00	100.00	10.08	0.91	unknown
MF1	−91.77	33.02	47.64	2.55	99.00	96.35	38.78	0.95	managed
MF2	−80.78	33.79	52.61	3.17	101.00	100.00	16.68	0.74	unmanaged
MF3	−69.51	43.93	37.69	4.05	100.00	91.77	7.21	0.73	unknown
MF4	−123.81	45.16	364.89	15.95	100.00	67.58	19.78	0.89	unmanaged

* Percentage means the area of target vegetation type divides the area of all vegetation types in each site.

Table 2. Flight information of the airborne LiDAR surveys and statistics of the collected airborne LiDAR data for each study site. Please note that the flight information of each study site was provided by the project report from the data provider.

Study Area	Year	Month	Accuracy (m)	Ground Density (pts/m ²)	Flight Height (m)	Sensor Type	Pulse Rate (kHz)	Scan Rate (Hz)	Data Source
ENF1	2009–2010	Jan, Feb, Mar, Apr, Sep, Oct	0.05	3.20	900–1300	LeicaALS50II, ALS60	105.00	52.00	Oregon Department of Geology and Mineral Industries
ENF2	2008	Aug	0.05	8.00	900	LeicaALS50II	105.00	52.20	Oregon Department of Geology and Mineral Industries
ENF3	2012	Aug	0.05	8.00	900–1300	LeicaALS50, ALS60, ALS70	52.2@900 m, 46.7@1300 m	NA	Oregon Department of Geology and Mineral Industries
ENF4	2011	Aug	0.04	4.00	1200	LeicaALS60	88.00	NA	United States Geological Survey
EBF1	2007–2008	Mar	0.08	1.42	2286	LeicaALS50	52.50	24.00	Northwest Florida Water Management district
EBF2	2007	Feb, Mar	0.01	2.73	800	LeicaALS50	55.00	36.00	Northwest Florida Water Management district
EBF3	2006	Mar, Apr	0.18	0.33	2438	LeicaALS50	38.00	20.00	Mississippi department of environment quality
EBF4	2010	Mar, Apr	0.12	1.00	1371	RieglLMS-Q680, LMS-Q680i	100.00	NA	United States Geological Survey
DBF1	2011	Mar	0.10	1.22	2174	Leica ALS50II	96.80	39.80	Maryland Department of Information Technology
DBF2	2011	Mar	0.18	1.45	1524	Optech3100	70.00	35.00	United States Geological Survey
DBF3	2011	Apr	0.13	1.30	1981	LeicaALS50II, ALS60 OptechALTM Gemini	115.60	41.80	United States Geological Survey
DBF4	2011	Mar, Apr	0.06	2.77	1981	Leica ALS50II	115.60	46.80	United States Geological Survey
MF1	2011–2012	Jul	0.23	2.00	2286	OptechALTM213	50.00	26.00	United States Geological Survey
MF2	2010	Mar	0.23	2.37	NA	NA	NA	NA	United States Geological Survey
MF3	2010	Sep	0.15	2.40	NA	NA	NA	NA	United States Geological Survey
MF4	2010	Apr	0.04	8.00	900–1300	LeicaALS50,ALS60	105.00	52.00	Department of Geology and Mineral Industries

Note: NA represents that there is no description about the information in the flight report.

2.3. Ancillary Datasets

In this study, four ancillary datasets were collected to generate independent covariables for the RF-based canopy height prediction procedure, including Landsat5 Thematic Mapper (TM) data, Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data, topographic data, and climate surfaces (Table 3). Landsat5 TM surface-reflectance images during the growing season (from May to September) of the same year as the corresponding LiDAR data of each study site were collected from the Google Earth Engine (GEE) platform (<https://earthengine.google.com/>). These surface-reflectance data are high-level products that have been through atmospheric correction using the Satellite Signal in the Solar Spectrum Radiative Transfer model [36]. The collected data were manually examined to ensure that the coverage of cloud and snow was lower than 10%. Land cover classification map was obtained from 500-m MODIS land cover data (MCD12Q1) [37], which has been used in many studies to stratify vegetation types [38,39]. The terrain information was represented by the 30-m Shuttle Radar Topography Mission (SRTM) digital terrain model (DTM) data collected from the United States Geological Survey. Although LiDAR data can be used to generate highly accurate terrain elevation data, they are usually unavailable in large-scale canopy height mapping practice, and SRTM is one of the most frequently used dataset in this kind of practice across the globe. Annual mean temperature and annual total precipitation with 800 m resolution between 1981 and 2010 were obtained from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) climate group (<http://www.prism.oregonstate.edu/>).

Table 3. Ancillary parameters used as independent variables for building Random Forest (RF) models.

Variable	Year	Resolution (m)	Data Source
Land cover map	2001–2010	500	MODIS
Landsat TM images	2006–2012	30	Land surface reflectance product
NDVI	2006–2012	30	Land surface reflectance product
Brightness calculated from Landsat TM images	2006–2012	30	Land surface reflectance product
Greenness calculated from Landsat TM images	2006–2012	30	Land surface reflectance product
Wetness calculated from Landsat TM images	2006–2012	30	Land surface reflectance product
Elevation	2000	30	SRTM
Slope	2000	30	SRTM
Aspect	2000	30	SRTM
Annual mean temperature	1981–2010	800	PRISM
Annual mean precipitation	1981–2010	800	PRISM

3. Methods

To evaluate the transferability of the RF-based canopy height mapping method for different vegetation types, locations and spatial scales, we first preprocessed all collected LiDAR and ancillary datasets and generate independent and dependent variable layers to feed into the RF algorithm (Figure 2). Then, 22 RF regression models were built by varying vegetation types and study sites (Figure 1b,c), and 124 RF regression models were built by varying spatial scales (40 models for extent and 84 models for resolution) (Figure 1d). These models are all build with the all the independent variables (i.e., CHM ~ Elevation + Slope + Aspect + Brightness + Greenness + Wetness + NDVI + Temperature + Precipitation). Two statistical parameters, coefficient of determination (R^2) and root-mean-squared error (RMSE), were used to evaluate the influence of vegetation type, location and spatial scale on the canopy height prediction accuracy.

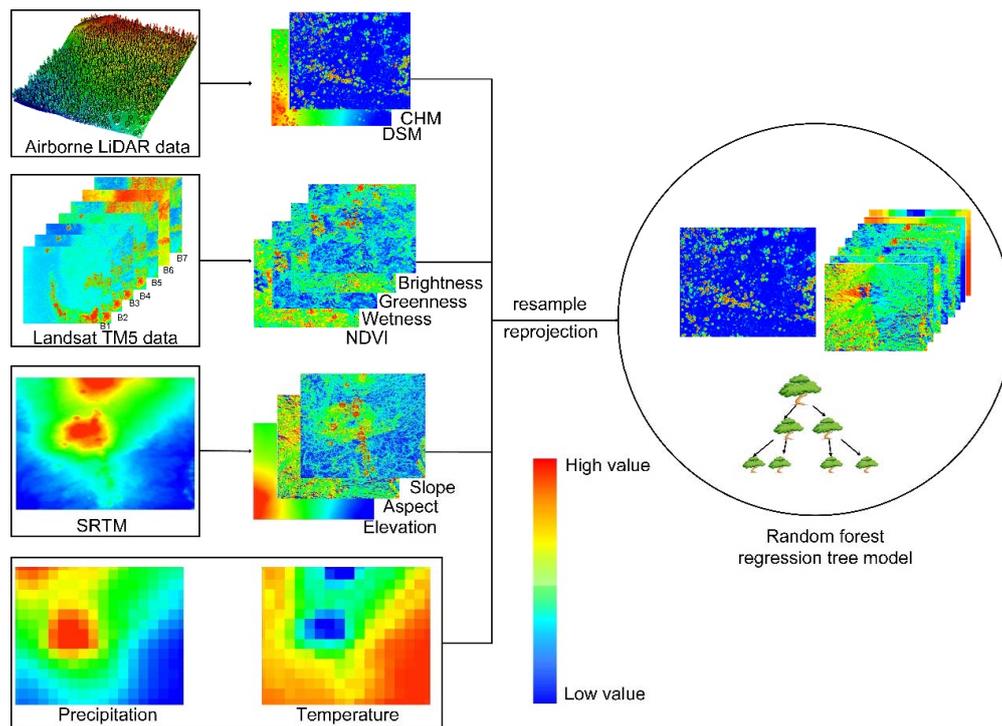


Figure 2. The flow chart of data preprocessing steps as well as the procedure of building the RF model for each individual run.

3.1. Data Preprocessing

3.1.1. Airborne LiDAR Data

To generate canopy height from the collected airborne LiDAR data, we first filtered the raw point clouds to ground and non-ground points using an improved progressive triangulated irregular network densification filtering algorithm [40], which is integrated in the Green Valley International[®] LiDAR360 software. Then, the ground returns were interpolated into 30 m resolution digital terrain model (DTM) using the ordinary kriging algorithm, since it has been proven that can be used to generate more accurate terrain surfaces from LiDAR points than other interpolation algorithms [16,41]. The digital surface model (DSM) in 30 m resolution was generated from the LiDAR first returns using the same procedure. Finally, the canopy height model (CHM) in 30 m resolution was derived as the difference between DSM and DTM. Although the collected LiDAR data in this study has a range of point density, it has been proven that a point density from 0.1 to 1 pt/m² is enough to estimate vegetation parameters such as canopy height model (CHM) and canopy cover [42–44]. Moreover, there have been studies showing that LiDAR-derived canopy height is highly accurate and comparable to field measurements [45,46]. Therefore, these airborne LiDAR derived CHM were used as “ground truth” to train and validate the RF-based method.

3.1.2. Ancillary Datasets

Time-series Landsat images for each study site were first visually examined to remove pixels with cloud and snow coverage. Then, a maximum-value composite (MVC) method was used to generate an image composite for each study site. The MVC method examined the time-series images pixel by pixel, and used the value from the image with the highest normalized difference vegetation index (NDVI) at each pixel location. This method was selected since it has been proven that it can minimize the cloud contamination and atmospheric attenuation from single date-imagery [14,47]. The generated MVC composites were processed in the GEE platform to calculate NDVI and obtain brightness, greenness

and wetness index through the Tassels Cap transformation. The MODIS MCD12Q1 land cover data was used to recognize the vegetation type for each study site. The terrain slope and aspect of each study site were derived from the SRTM data using the ESRI® ArcMap 10.1. The PRISM climate surfaces were resampled to 30 m resolution using the bilinear interpolation method so that they can match the resolution of other collected datasets

3.2. Evaluation of RF Transferability on Canopy Height Prediction

The regression tree method of RF consists of three main steps: (1) Draw samples from original data with bootstrap method (i.e., sampling from a population uniformly with replacement); (2) Build unpruned regression tree for the bootstrap samples, and chose the best splits from the randomly sampled variables at each node; (3) Aggregating predictions of all regression trees to get the final prediction result. In this study, 70% of the LiDAR-derived canopy height pixels of each study site were randomly selected as the ground truth to train the RF model, and 10 commonly used independent variables (including NDVI, elevation, slope, aspect, greenness, brightness, wetness, vegetation type, temperature, precipitation) were used to feed the RF algorithm and predict canopy height. The RF training process was performed in the “randomForest” R package [48]. The “randomForest” regression functions have two primary parameters needed to be defined, i.e., the number of trees and the number of variables tried at each split node. In this study, these two parameters were determined using grid search method. The best number of trees and number of variables tried at each split were determined by the values that produced the minimum mean-squared error. Here, 500 “RF trees” were included and 4 variables were tried at each split for each model run based on grid search. In total, 144 RF models were built to evaluate the influence of vegetation type, locations, and spatial scales on the canopy height prediction results. In this study, each model was run 100 times by choosing 100 replicate samples randomly, and chose the average self-prediction result to ensure stability.

3.2.1. The Influence of Locations

To evaluate the influence of different study site locations on the RF model performance, we fixed the vegetation type and spatial scale of the input layers for comparison. Overall, a single RF canopy height prediction model was built for each study site and it was used to predict canopy height of the study sites with the same vegetation type at a constant resolution of 30 m. Taking ENF vegetation type as an example, we have four 100 km² ENF sites, namely ENF1, ENF2, ENF3, and ENF4 (Figure 1b). For each ENF site, 70% of the CHM pixels were used to obtain a canopy height prediction model following the above-mentioned parameter set. The self-predicted R^2 and RMSE values for each prediction run were calculated through the comparison between the predicted canopy height values and the remaining 30% LiDAR-derived validation pixels at each ENF site. The cross-predicted R^2 and RMSE were evaluated by using each build model to predict canopy height of other different sites, which was compared with the self-predicted results.

Besides, we evaluated whether a universal model can be achieved at different study sites with the same vegetation type. To accomplish this, we mixed the training samples from the study sites of one vegetation type and randomly picked one-fourth of them to train a RF regression model for each vegetation type. Then, the mixed RF model of each vegetation type (namely Model_{ENFm}, Model_{EBFm}, Model_{DBFm}, and Model_{MFm}) was used to predict the canopy height at each study site with the corresponding vegetation type. The prediction results at each study site were compared with its validation samples, and the R^2 and RMSE were calculated correspondingly.

3.2.2. The Influence of Vegetation Types

Vegetation type is another factor that may have significant influence on the transferability of RF canopy height prediction model. Because the 16 study sites are from different locations, we cannot fully remove the influence of location on the canopy height prediction results. Nevertheless, we tried to minimize this effect by using the RF regression models built from the mixed training samples

of all four sites of each vegetation type (i.e., Model_{ENFm}, Model_{EBFm}, Model_{DBFm}, and Model_{MFm}). Within each vegetation type, 70% of the CHM pixels of each study site were randomly selected and grouped together to train the RF canopy height prediction model. The self-predicted R^2 and RMSE values for each prediction run were calculated through the comparison between the predicted canopy height values and the remaining 30% LiDAR-derived validation pixels at each vegetation type. The cross predicted R^2 and RMSE were evaluated by using each built model to predict canopy height of other different vegetation types with mixed site samples, which was compared with the self-predicted results.

Moreover, we further examined whether a universal RF canopy height prediction model can be achieved across different vegetation types. Around 6% of the training samples from each study site were randomly picked and used as the training samples to build two RF canopy height prediction models. One model (Model_{TV}) included vegetation type as a dummy variable in the training process, and the other one (Model_{TNV}) did not. The canopy height prediction results from both models were evaluated by the independent variables at each study site.

3.2.3. The Influence of Spatial Scales

We examined the influence of spatial scales on RF canopy height prediction results in two ways, the spatial resolution of the targeted canopy height product and the spatial extent of the study site. To get comprehensive results in each vegetation type, four study sites with the highest self-prediction accuracy (i.e., the highest R^2 and the lowest RMSE) of each vegetation type were chosen. To evaluate the influence of spatial resolution, we resampled the independent input variables to 50 m firstly using the bilinear method in ArcMap10.1; then, we manually altered the spatial resolution of these input layers from 50 m to 1000 m with a step of 50 m also using bilinear method. These input layers with different spatial resolutions were used to build 84 RF regression models and predict canopy height at different spatial resolutions. Similar to previous steps, 70% of the resampled pixels were randomly selected as the training samples, and the remaining 30% pixels were used as the validation samples.

To evaluate the influence of spatial extent on the canopy height prediction results, we manually decreased the study extent from 10 km \times 10 km to 1 km \times 1 km with a step of 1 km \times 1 km. In total, 40 RF-based canopy height prediction processes were conducted in this step. For each run, 70% of the pixels within the corresponding spatial extents were selected as the training samples, and the remaining 30% were used as the validation samples.

4. Results

4.1. Variable Importance for RF-Based Canopy Height Prediction

The importance of covariables for the RF-based canopy height prediction method was evaluated by the percentage increase in the mean-squared error (%IncMSE). The larger the %IncMSE of a variable is, the more important the variable is [48]. As can be seen in Figure 3, overall, topography-related variables (i.e., elevation, slope, and aspect), temperature, precipitation and NDVI (or vegetation greenness) were important variables in the RF-based canopy height prediction method, but the variable importance varied with sites. ENF sites and DBF sites were more controlled by topography-related variables, EBF sites were more determined by elevation and precipitation, and MF sites were more influenced by NDVI, temperature and precipitation. Although different study sites depended on different variable inputs on canopy height prediction, the prediction results all had significant correlations with independent LiDAR-derived canopy height (Table 4). However, the canopy height prediction accuracy varied significantly with sites. The R^2 ranged from 0.1 to 0.94, and the RMSE ranged from 1.18 m to 9.83 m. Nevertheless, the RF-based canopy height prediction method showed the capability to estimate canopy height across different study sites with different vegetation types. The mean R^2 of the ENF, EBF, DBF, and MF sites were 0.22, 0.56, 0.36 and 0.54, respectively.

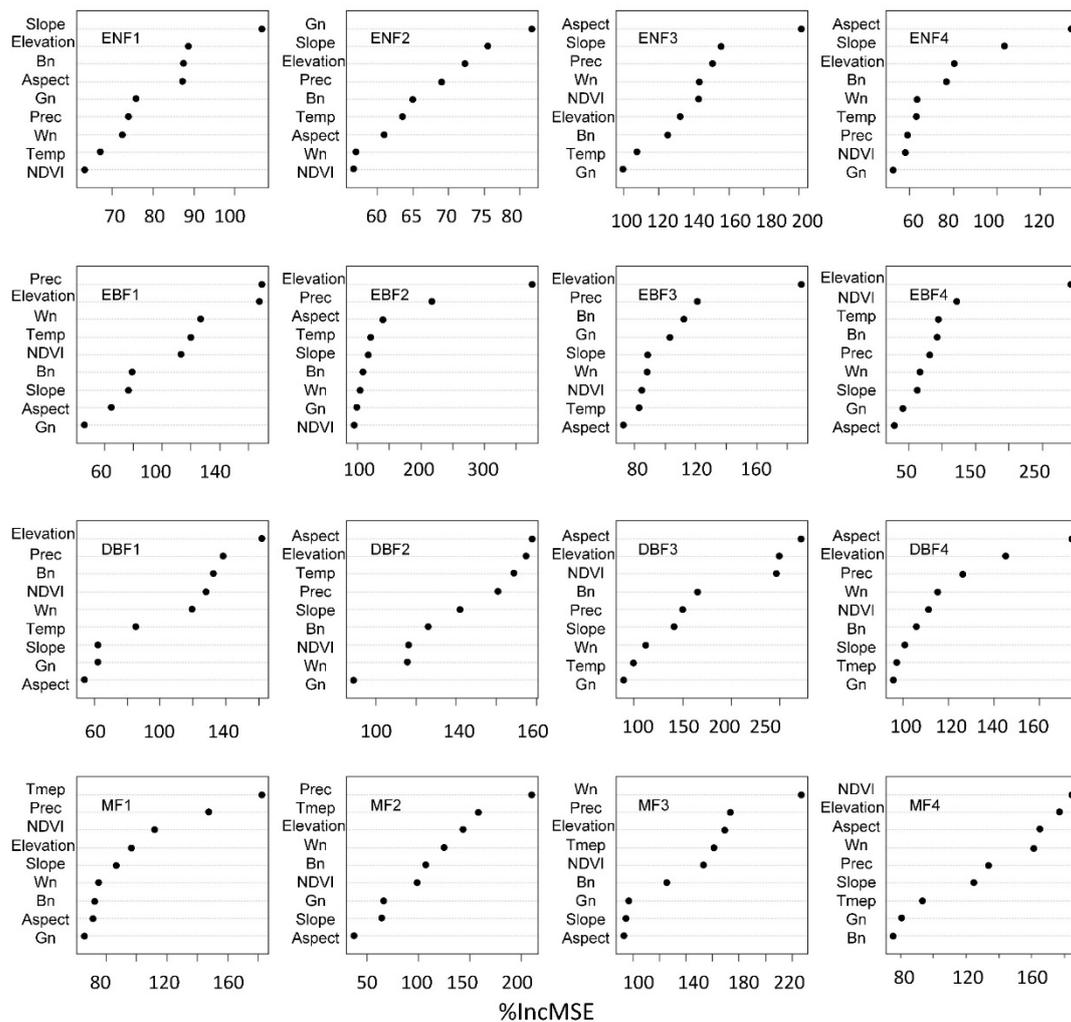


Figure 3. Importance of variables, denoted by percentage increase of mean-squared error (%IncMSE), for the RF tree height prediction models. Bn, Gn, Wn, NDVI, Prec, Temp, and Elevation represent the brightness, greenness, wetness, normalized difference vegetation index, precipitation, temperature, and SRTM elevation, respectively.

Table 4. Accuracy assessment of site-specific RF-based tree height prediction results. The coefficient of determination (R^2) and root-mean-squared error (RMSE) of each study site were calculated by comparing the site-specific self-prediction result with independent validation pixels. The corresponding R^2 value are all significant at a 99.9% confident level.

	Site 1		Site 2		Site 3		Site 4	
	R^2	RMSE (m)						
ENF	0.15	5.26	0.10	1.48	0.32	9.83	0.30	4.82
EBF	0.43	4.83	0.59	2.91	0.47	4.40	0.75	8.78
DBF	0.37	6.15	0.35	7.68	0.48	4.84	0.23	3.91
MF	0.94	1.18	0.51	5.26	0.36	3.72	0.34	9.96

4.2. The Transferability of RF across Different Locations

It can be concluded that a RF canopy height prediction model trained at one study site cannot be transferred to other study sites with the same vegetation type. As shown in Figure 4, the canopy height prediction accuracy of each study site of one vegetation type was the highest (i.e., the largest R^2 and the lowest RMSE) when it was estimated by the model built with its own training samples.

When the model built for each study site was used to predict canopy height of other study sites with the same vegetation type, the prediction accuracy dropped drastically. The mean R^2 value of self-prediction for each vegetation type was around 2 to 13 times higher than cross-site predictions, and the mean RMSE was around 400% to 800% lower than cross-site predictions. However, for each vegetation type, a location-transferable RF canopy height prediction model could be achieved by selecting training samples from all study sites. The four mixed models (i.e., $\text{Model}_{\text{ENF}_m}$, $\text{Model}_{\text{EBF}_m}$, $\text{Model}_{\text{DBF}_m}$, and $\text{Model}_{\text{MF}_m}$) could generate slightly higher canopy height prediction accuracy than site-specific self-prediction models in 15 out of the 16 study sites. The R^2 was around 0.06 higher, and the RMSE was around 0.33 m lower (Table 5). The MF1 site was the only exception that had a slightly lower canopy height prediction accuracy from the mixed model $\text{Model}_{\text{MF}_m}$.

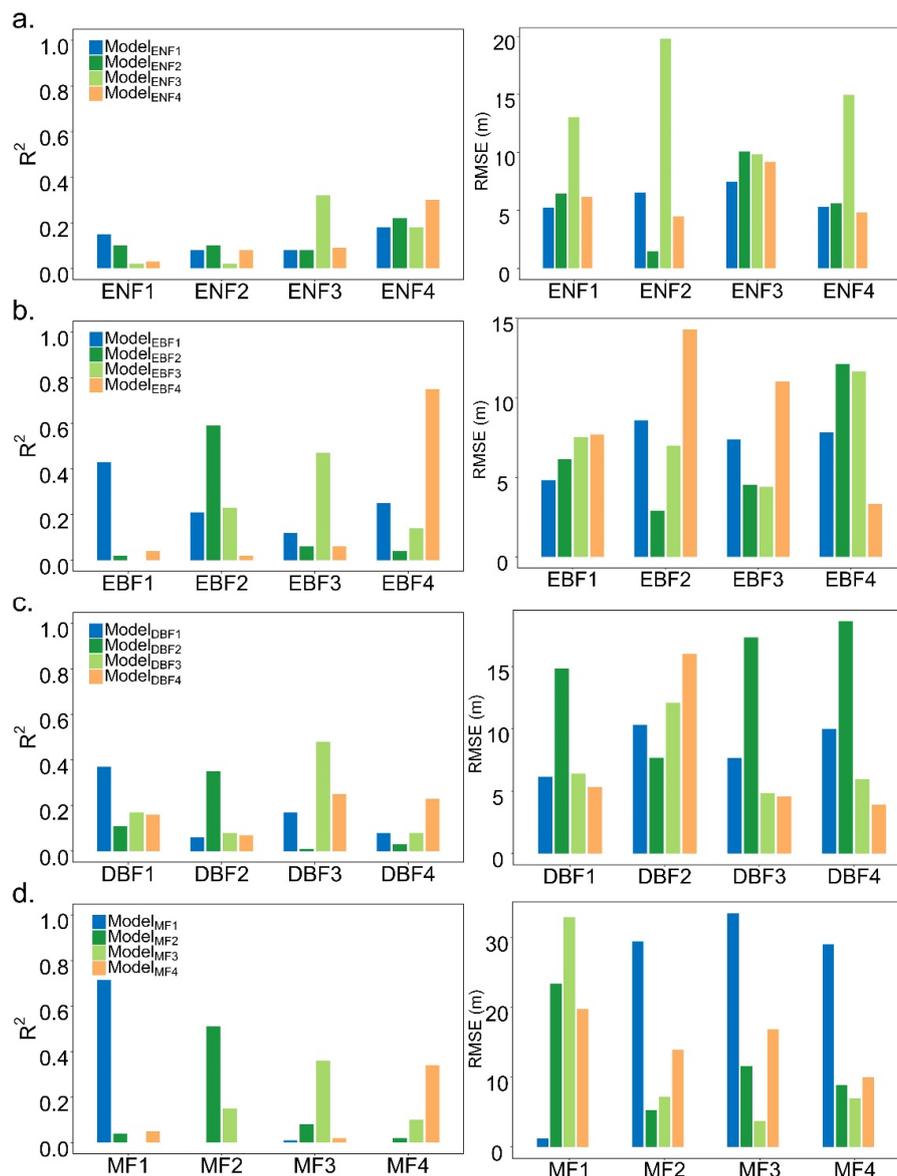


Figure 4. The RF model transferability among different study sites of each vegetation type. The x-axis represents different study sites of a certain vegetation type, and the y-axis is either the R^2 or the RMSE calculated by directly comparing the prediction results with the independent LiDAR-derived validation pixels within the corresponding study site. (a–d) represent the evaluation results among different sites of vegetation type ENF, EBF, DBF, and MF, respectively.

Table 5. The accuracy differences between site-specific tree height prediction results and tree height prediction results using mixed training samples from all four study sites of each vegetation type (i.e., tree height prediction results from Model_{ENFm}, Model_{EBFm}, Model_{DBFm} and Model_{MFm}). The accuracy difference is denoted as the difference in R^2 and RMSE (mixed models minus site specific models).

	Site 1		Site 2		Site 3		Site 4	
	ΔR^2	$\Delta RMSE$ (m)						
ENF	0.12	−0.72	0.11	−0.10	0.07	−0.60	0.07	−1.26
EBF	0.08	−0.29	0.02	−0.03	0.06	−0.22	0.03	−0.24
DBF	0.07	−0.33	0.01	−0.07	0.06	−0.29	0.10	−0.25
MF	−0.03	0.30	0.05	−0.25	0.08	−0.15	0.01	−0.14

4.3. The Transferability of RF across Different Vegetation Types

Similar to the influence of study site locations on the RF-based canopy height prediction results, a RF model trained for a certain vegetation type can hardly be used to predict canopy height for other vegetation types. Although the four mixed models for each vegetation type can be used to well estimate canopy height within their study sites ($R^2 > 0.6$), the prediction accuracy dropped significantly when they were used for cross-vegetation type prediction (Figure 5). The R^2 for cross-vegetation type prediction results was over 10 times smaller than self-prediction results of a certain vegetation type on average, and the RMSE was over 50% larger. Nevertheless, a universal RF-based canopy height prediction model can be achieved by including training samples from all study sites of different vegetation types. A single model trained from all study sites without using vegetation type as a dummy variable Model_{TNV} can generate equivalent canopy height prediction accuracy to the four self-predicting mixing models. Their R^2 and RMSE were almost identical (Table 6). Moreover, including vegetation type to train a RF model (Model_{TV}) could not improve the canopy height prediction accuracy.

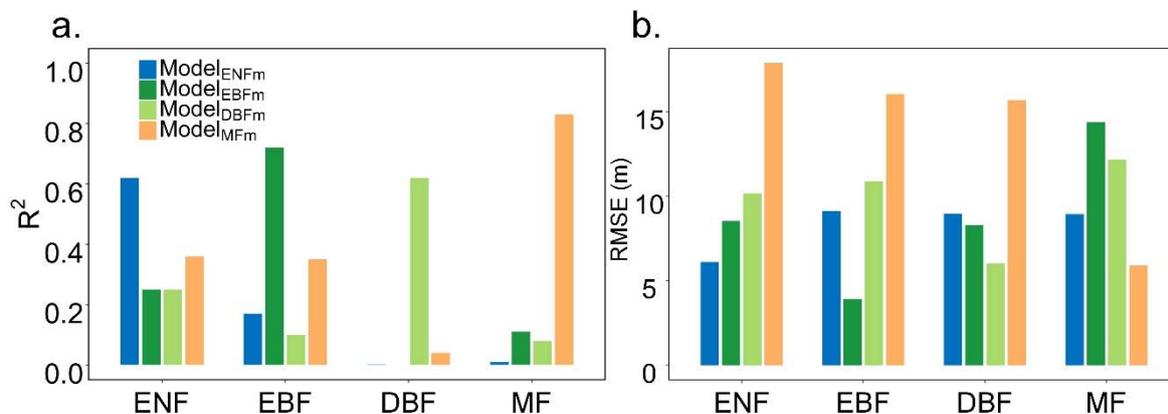


Figure 5. The transferability of RF models built from training samples mixed from all four study sites of each vegetation type (i.e., Model_{ENFm}, Model_{EBFm}, Model_{DBFm} and Model_{MFm}). The x-axis represents different vegetation types, and the y-axis is either the R^2 (a) or the RMSE (b) calculated by directly comparing the prediction results with the independent LiDAR-derived validation pixels.

Table 6. The accuracy differences between tree height prediction results using Model_{T_{NV}} or Model_{T_V} and results using Model_{ENF_m}, Model_{EBF_m}, Model_{DBF_m} or Model_{MF_m}. Model_{T_{NV}} and Model_{T_V} represent the RF tree height prediction models trained from pixels from all 16 study sites without/with considering vegetation type as a dummy variable. The accuracy difference is denoted as the difference in R^2 and RMSE (Model_{T_{NV}} or Model_{T_V} minus Model_{ENF_m}, Model_{EBF_m}, Model_{DBF_m} or Model_{MF_m}).

	Model _{T_{NV}}		Model _{T_V}	
	ΔR^2	$\Delta RMSE$ (m)	ΔR^2	$\Delta RMSE$ (m)
ENF	−0.01	0.08	−0.01	0.06
EBF	−0.02	0.15	−0.02	0.16
DBF	−0.01	0.11	−0.01	0.08
MF	−0.01	0.11	−0.01	0.10

4.4. The Transferability of RF across Different Spatial Scales

Four study sites with the best self-prediction accuracy of each corresponding vegetation type, i.e., site ENF3, EBF4, DBF3 and MF1 (Figure 4), were selected to evaluate the influence of spatial extent and spatial resolution on RF-based canopy height prediction results. Overall, the spatial extent had a positive logarithmic correlation with the canopy height prediction accuracy (Figure 6). The R^2 of the prediction results could increase by 100%, 15%, 37% and 25% with the spatial extent at site ENF3, EBF4, DBF3 and MF1, and the RMSE could decrease by 15%, 37%, 53% and 25%, respectively. Site ENF3 and MF1 were the two study sites that were more influenced by spatial extent. Moreover, the influence of spatial extent on the canopy height prediction accuracy was not linear. After the spatial extent reached to a certain threshold, the increase in canopy height prediction accuracy was hardly noticeable (Figure 6). By using the first-order derivation analysis on the fitted line between spatial extent and R^2 (or RMSE), we found that the thresholds for all these four study sites were around 50 km². Please note that the above-mentioned analysis was conducted at a spatial resolution of 30 m. The threshold of the optimal spatial extent might be different at a different targeted spatial resolution.

Spatial resolution also has a positive logarithmic correlation with the RF-based canopy height prediction accuracy. As can be seen in Figure 7, the canopy height prediction accuracy generally increased with the spatial resolution, but this pattern was not as strong as the spatial extent, except the MF1 study site. With the increase of spatial resolution (smaller pixel size), the R^2 between the predictions and independent LiDAR-derived canopy height could increase by around 80%, 5%, 40%, and 50% at the ENF3, EBF4, DBF3 and MF1 sites, and RMSE could decrease by around 5%, 5%, 10% and 180%, respectively. The influence of spatial resolution on the canopy height prediction accuracy was also nonlinear. After the spatial resolution being coarser than 500 m, the canopy height prediction accuracy was not noticeably decreased anymore (Figure 7).

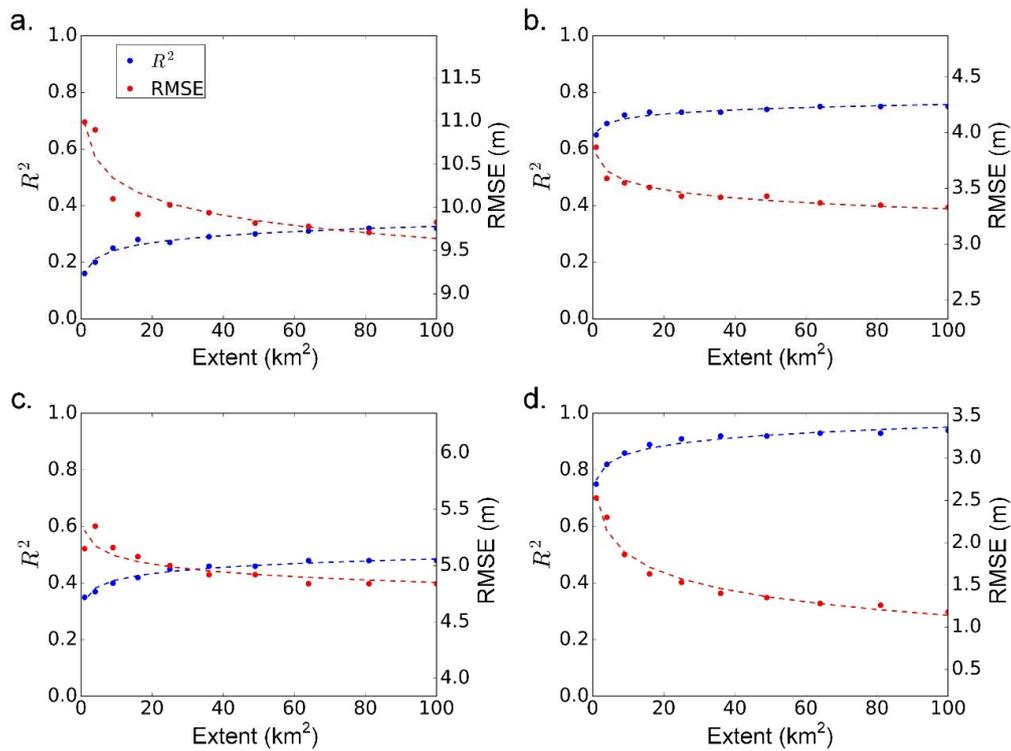


Figure 6. The influence of spatial extent on the RF-based tree height prediction accuracy. (a–d) represent the evaluation results at site ENF3, EBF4, DBF3, and MF1, respectively.

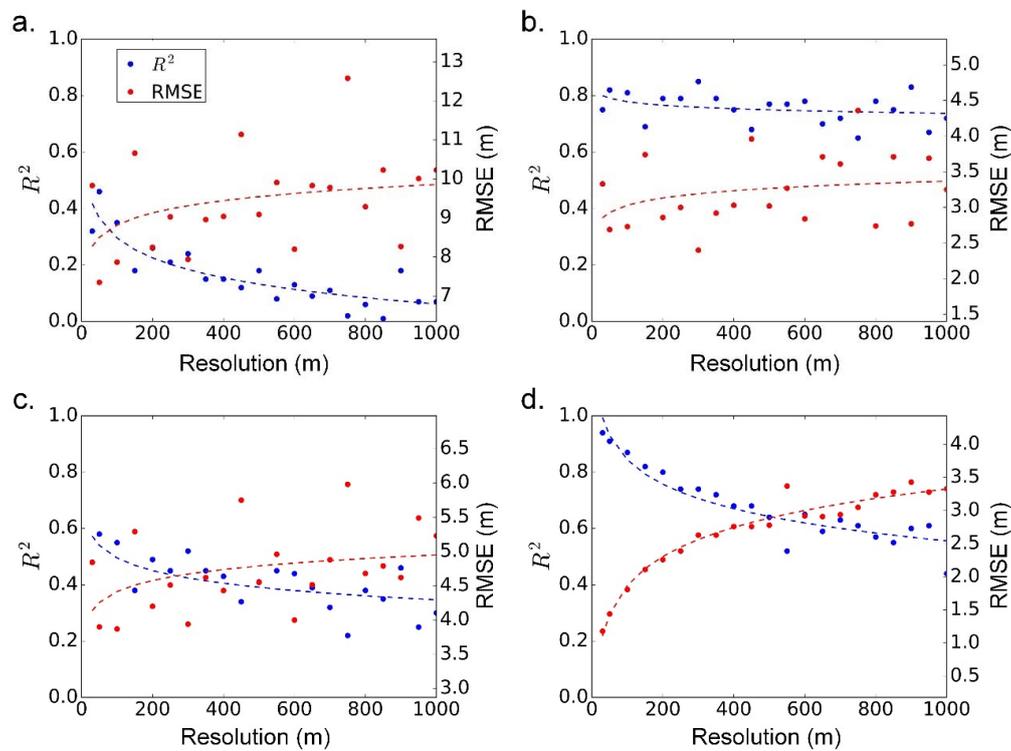


Figure 7. The influence of the targeted spatial resolution on the tree height prediction accuracy. (a–d) represent the evaluation results in site ENF3, EBF4, DBF3, and MF1, respectively.

5. Discussion

The RF-based method showed great potential to upscale canopy height estimations from LiDAR footprints to larger areas. However, the trained RF model performed differently in different areas, and the controlling factors also varied with study sites and vegetation types. In high-elevation ENF sites, topographic related parameters (i.e., elevation, slope and aspect) were the more important factors for the canopy height modelling, since high-elevation forests were more likely energy-limited and topographic parameters can determine the amount of solar radiation that can be received by trees [49,50]. Annual mean precipitation and vegetation indices from Landsat imagery played a much bigger role in canopy height prediction for the low to medium elevation EBF and DBF sites. This may be caused by the fact that broadleaf forests are more water-limited rather than energy-limited as broadleaf forests intercept less precipitation than coniferous forests, and the vegetation indices from optical imagery can well represent the forest structure and vegetation health [51–53]. The four MF forests had site-specific variable importance, since the composition of tree species and topographic climate conditions may change significantly with sites. Moreover, it was found that the RF-based canopy height prediction accuracy decreased with the rise of elevation (Tables 1 and 4), which is consistent with the suggestion that ecological modelling studies were preferable in low-elevation areas [54]. Integrating spaceborne LiDAR data, which can provide forest structure information, may further help to improve the model performance in high-elevation areas [14].

The performance of the site-specific canopy height prediction model shows that a single RF model trained within a LiDAR footprint cannot be used to upscale canopy height estimations in a larger area, which may be composed of different vegetation types, topographic characteristics, and climate conditions. The reason may be the RF canopy height prediction model trained within one LiDAR footprint was very sensitive to the study site location [55]. The canopy height prediction accuracy dropped significantly while one model was used to estimate canopy height at other study sites even with the same vegetation type (Figure 4), and the weakest transferability of site-specific RF canopy height prediction model appeared at the four MF sites. As above-mentioned, the complexity in vegetation compositions and topographic and climate conditions may lead to significantly different variables controlling the tree growth. In addition, MF sites naturally have more complex vegetation, topographic and climate conditions comparing to the other three vegetation types, which may be the reason for their weak transferability. Moreover, the model transferability was also largely diminished among different vegetation types (Figure 5), which may also be caused by that different vegetation types have different controlling factors for tree growth (Figure 3).

However, the weak transferability of RF models can be resolved by including representative training samples from all study sites and vegetation types. The barrier of RF canopy height prediction model in locations can be removed by using the training samples from all study sites of each vegetation type (Table 5). It has been documented that the heterogeneity in the training samples from different study sites is beneficial for ecological modelling, and training samples from a range of environmental sets should be included to improve the model performance [56,57]. The four mixed models of each vegetation type can even slightly improve the canopy height prediction accuracy, except at the MF1 site. The R^2 value derived from the Model_{MFm} was dropped by around 3% compared with the model trained from its own samples. The possible explanation for this may be that MF forests naturally contain mixed vegetation types with sufficient heterogeneity, and the RF model may not gain improvements by increasing training samples from different study sites. Moreover, errors introduced in the data pre-processing steps, such as different LiDAR data qualities at different study sites and spatial mis-registration among different data sources, may also influence the performance of mixed models [55].

Moreover, the transferability of a RF model can be further improved by selecting representative training samples covering all vegetation types, and a universal canopy height prediction model for continental and maybe even global scale can be obtained. The estimated canopy height from a universal RF model was almost equivalently accurate to that from the model built for each vegetation type

(Table 6). Vegetation type was not a necessary input for generating the universal RF canopy height prediction model. This may indicate that the RF algorithm can pick out the vegetation type information from the training samples if they can well represent all vegetation types [58].

Besides, the spatial extent of LiDAR footprints and the targeted spatial resolution of canopy height estimations are also two important factors that influence model performance. The canopy height prediction accuracy was positively influenced by both the spatial extent and spatial resolution (Figures 6 and 7). This might be mainly caused by two reasons: (1) the heterogeneity can have a significant increase at the beginning of the increase of spatial extent and the decrease of spatial resolution (bigger pixel size), and reaches a sill gradually, which corresponds to the rule that the number of patches change quickly at the beginning which brings much heterogeneity change, and reaches stability with the larger extent and coarser resolution [59,60]; (2) the number of training samples increases with the spatial extent and resolution, which can indirectly increase the spatial heterogeneity. This influence was non-linear because the mean value of different variables may not have the same changing rate in a simple linear way, which brings inconsistency among various variables and canopy height at each pixel. The larger variance (errors) may arise at coarser scales, which was also suggested by Moody and Woodcock [61]. Moreover, the influence of spatial resolution varied more significant than extent, because when the resolution decreases, not only the number of training pixels decreases, but also the heterogeneity of each pixel might be reduced also. In addition, the influence of spatial scales of broadleaf (i.e., EBF, DBF) was less, that may be caused by the more stable forest characteristics of the broad-leaved forest [62]. Therefore, we need to comprehensively consider the influence of spatial extent and resolution when we build the RF model in various vegetation types.

“All models are wrong, but some are useful” [63]. This is especially true for models transferability built from remote sensing data [64,65]. This study can provide a comprehensive knowledge in using RF model not only to predict regional to global-scale canopy height but also to estimate biomass and other forest parameters. It can provide guidance on how to choose enough LiDAR data footprints to ensure the modelling accuracy. However, the current study did not consider the temporal transferability of RF-based method in estimating forest parameters, and we did not compare the RF algorithm with other machine learning algorithms. These questions need to be further explored in the future.

6. Conclusions

This study aims to evaluate the RF model transferability among various vegetation types, locations, and spatial scales through the combination of multi-source remote sensed datasets. In total, 16 study sites with an area of 100 km² covering four vegetation types (four study sites for each vegetation type) were chosen to address this mission. The results demonstrate that RF method can be used to upscale canopy height measurements from airborne LiDAR footprints to areas without LiDAR coverage. However, the RF models built at different study sites have different controlling factors on the canopy height estimations, which constrain their transferability among different locations and vegetation types. A RF model built in a specific location cannot transfer to other locations; as well as a RF model built in a vegetation type cannot transfer to other vegetation types. By including the training samples from all study sites of each vegetation type, a universal RF canopy height prediction model can be achieved for each vegetation type that can generate equivalent accuracy as site-specific models. A universal RF canopy height prediction model can even be generated by including training samples covering all vegetation types and study sites, but vegetation type alone, as a dummy variable, does not play a significant role in the RF model. In addition, spatial extent of the LiDAR footprint and spatial resolution of the targeted canopy height products both have a positive correlation with the canopy height prediction accuracy. They need to be carefully evaluated in the RF-based canopy height upscaling process based on the study extent, available airborne LiDAR data, and the available ancillary datasets.

Author Contributions: Conceptualization, S.J., Y.S. and Q.G.; Methodology, S.J., Y.S. and S.G.; Software, S.J. and T.H.; Validation, S.J., Y.S., T.H., J.L. and Q.G.; Formal Analysis, S.J., Y.S., and Q.G.; Investigation, S.J. and S.G.;

Resources, Y.S., Q.G. and S.J.; Data Curation, S.J., Y.S., S.G. and Q.G.; Writing-Original Draft Preparation, S.J.; Writing-Review & Editing, Y.S., Q.G. and S.J.; Visualization, S.J.; Supervision, Q.G. and Y.S.; Project Administration, S.J., Y.S. and Q.G.; Funding Acquisition, Q.G.

Funding: This research was funded by the National Key R&D Program of China (Grant Number 2016YFC0500202, 2017YFC0503905), the Frontier Science Key Programs of the Chinese Academy of Sciences (Grant Number QYZDY-SSW-SMC011), the National Science Foundation of China (Grant Number 41471363, 31741016), CAS Pioneer Hundred Talents Program, and the US National Science Foundation (Grant Number DBI 1356077).

Acknowledgments: These authors contributed equally to this work.

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

References

1. Costanza, R.; d'Arge, R.; de Groot, R.; Farber, S.; Grasso, M.; Hannon, B.; Limburg, K.; Naeem, S.; O'Neill, R.V.; Paruelo, J.; et al. The value of the world's ecosystem services and natural capital. *Nature* **1997**, *387*, 253–260. [[CrossRef](#)]
2. Houghton, R. Aboveground forest biomass and the global carbon balance. *Glob. Chang. Biol.* **2005**, *11*, 945–958. [[CrossRef](#)]
3. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G. A large and persistent carbon sink in the world's forests. *Science* **2011**, *333*, 988–993. [[CrossRef](#)] [[PubMed](#)]
4. Balzter, H.; Rowland, C.S.; Saich, P. Forest canopy height and carbon estimation at monks wood national nature reserve, UK, using dual-wavelength SAR interferometry. *Remote Sens. Environ.* **2007**, *108*, 224–239. [[CrossRef](#)]
5. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.; Chambers, J.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99. [[CrossRef](#)] [[PubMed](#)]
6. Chen, G.; Ozelkan, E.; Singh, K.K.; Zhou, J.; Brown, M.R.; Meentemeyer, R.K. Uncertainties in mapping forest carbon in urban ecosystems. *J. Environ. Manag.* **2017**, *187*, 229–238. [[CrossRef](#)] [[PubMed](#)]
7. Hu, T.; Su, Y.; Xue, B.; Liu, J.; Zhao, X.; Fang, J.; Guo, Q. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sens.* **2016**, *8*, 565. [[CrossRef](#)]
8. Xue, B.L.; Guo, Q.; Hu, T.; Xiao, J.; Yang, Y.; Wang, G.; Tao, S.; Su, Y.; Liu, J.; Zhao, X. Global patterns of woody residence time and its influence on model simulation of aboveground biomass. *Glob. Biogeochem. Cycles* **2017**, *31*, 821–835. [[CrossRef](#)]
9. Zhang, G.; Ganguly, S.; Nemani, R.R.; White, M.A.; Milesi, C.; Hashimoto, H.; Wang, W.; Saatchi, S.; Yu, Y.; Myneni, R.B. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sens. Environ.* **2014**, *151*, 44–56. [[CrossRef](#)]
10. Prush, V.; Lohman, R. Forest canopy heights in the pacific northwest based on InSAR phase discontinuities across short spatial scales. *Remote Sens.* **2014**, *6*, 3210–3226. [[CrossRef](#)]
11. Donoghue, D.; Watt, P. Using LiDAR to compare forest height estimates from IKONOS and Landsat ETM+ data in Sitka spruce plantation forests. *Int. J. Remote Sens.* **2006**, *27*, 2161–2175. [[CrossRef](#)]
12. McCombs, J.W.; Roberts, S.D.; Evans, D.L. Influence of fusing LiDAR and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *For. Sci.* **2003**, *49*, 457–466.
13. Lefsky, M.A.; Keller, M.; Pang, Y.; De Camargo, P.B.; Hunter, M.O. Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms. *J. Appl. Remote Sens.* **2007**, *1*, 013537. [[CrossRef](#)]
14. Su, Y.; Ma, Q.; Guo, Q. Fine-resolution forest tree height estimation across the sierra nevada through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery. *Int. J. Digit. Earth* **2016**, *10*, 307–323. [[CrossRef](#)]
15. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. LiDAR remote sensing for ecosystem studies: LiDAR, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience* **2002**, *52*, 19–30.

16. Clark, M.L.; Clark, D.B.; Roberts, D.A. Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sens. Environ.* **2004**, *91*, 68–89. [[CrossRef](#)]
17. Alexander, C.; Bøcher, P.K.; Arge, L.; Svenning, J.-C. Regional-scale mapping of tree cover, height and main phenological tree types using airborne laser scanning data. *Remote Sens. Environ.* **2014**, *147*, 156–172. [[CrossRef](#)]
18. 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*, 78–82. [[CrossRef](#)]
19. 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](#)]
20. 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](#)]
21. Wang, Y.; Li, G.; Ding, J.; Guo, Z.; Tang, S.; Wang, C.; Huang, Q.; Liu, R.; Chen, J.M. A combined GLAS and MODIS estimation of the global distribution of mean forest canopy height. *Remote Sens. Environ.* **2016**, *174*, 24–43. [[CrossRef](#)]
22. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
23. Chen, G.; Hay, G.J. A support vector regression approach to estimate forest biophysical parameters at the object level using airborne LiDAR transects and quickbird data. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 733–741. [[CrossRef](#)]
24. Cutler, M.E.J.; Boyd, D.S.; Foody, G.M.; Vetrivel, A. Estimating tropical forest biomass with a combination of sar image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 66–77. [[CrossRef](#)]
25. Tetko, I.V.; Livingstone, D.J.; Luik, A.I. Neural network studies. 1. Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 826–833. [[CrossRef](#)]
26. Zhu, X.; Liu, D. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 222–231. [[CrossRef](#)]
27. Cao, L.; Pan, J.; Li, R.; Li, J.; Li, Z. Integrating airborne LiDAR and optical data to estimate forest aboveground biomass in arid and semi-arid regions of China. *Remote Sens.* **2018**, *10*, 532. [[CrossRef](#)]
28. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]
29. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
30. 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](#)]
31. Su, Y.; Guo, Q.; Xue, B.; Hu, T.; Alvarez, O.; Tao, S.; Fang, J. Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sens. Environ.* **2016**, *173*, 187–199. [[CrossRef](#)]
32. Keane, R.E.; Parsons, R.A.; Hessburg, P.F. Estimating historical range and variation of landscape patch dynamics: Limitations of the simulation approach. *Ecol. Model.* **2002**, *151*, 29–49. [[CrossRef](#)]
33. Hall, R.; Skakun, R.; Arsenaault, E.; Case, B. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *For. Ecol. Manag.* **2006**, *225*, 378–390. [[CrossRef](#)]
34. Turner, M.G.; Dale, V.H.; Gardner, R.H. Predicting across scales: Theory development and testing. *Landsc. Ecol.* **1989**, *3*, 245–252. [[CrossRef](#)]
35. Turner, M.G.; O'Neill, R.V.; Gardner, R.H.; Milne, B.T. Effects of changing spatial scale on the analysis of landscape pattern. *Landsc. Ecol.* **1989**, *3*, 153–162. [[CrossRef](#)]
36. Masek, J.; Vermote, E.; Saleous, N.; Wolfe, R.; Hall, F.; Huemmrich, F.; Gao, F.; Kutler, J.; Lim, T. *LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing Code, Version 2*; ORNL DAAC: Oak Ridge, TN, USA, 2013.
37. Broxton, P.D.; Zeng, X.; Sulla-Menashe, D.; Troch, P.A. A global land cover climatology using MODIS data. *J. Appl. Meteorol. Clim.* **2014**, *53*, 1593–1605. [[CrossRef](#)]

38. Mohan, C.; Western, A.W.; Wei, Y.; Saft, M. Predicting groundwater recharge for varying land cover and climate conditions—A global meta-study. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2689–2703. [[CrossRef](#)]
39. Sharma, A.; Goyal, M.K. Assessment of ecosystem resilience to hydroclimatic disturbances in India. *Glob. Chang. Biol.* **2018**, *24*, 432–441. [[CrossRef](#)] [[PubMed](#)]
40. Zhao, X.; Guo, Q.; Su, Y.; Xue, B. Improved progressive tin densification filtering algorithm for airborne LiDAR data in forested areas. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 79–91. [[CrossRef](#)]
41. Guo, Q.H.; Li, W.K.; Yu, H.; Alvarez, O. Effects of topographic variability and LiDAR sampling density on several dem interpolation methods. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 701–712. [[CrossRef](#)]
42. Ma, Q.; Su, Y.; Guo, Q. Comparison of canopy cover estimations from airborne LiDAR, aerial imagery, and satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4225–4236. [[CrossRef](#)]
43. Jakubowski, M.K.; Guo, Q.; Kelly, M. Tradeoffs between LiDAR pulse density and forest measurement accuracy. *Remote Sens. Environ.* **2013**, *130*, 245–253. [[CrossRef](#)]
44. 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](#)]
45. Næsset, E.; Bjerknæs, K.-O. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sens. Environ.* **2001**, *78*, 328–340. [[CrossRef](#)]
46. Hwang, S.; Lee, I. Current status of tree height estimation from airborne LiDAR data. *Korean J. Remote Sens.* **2011**, *27*, 389–401. [[CrossRef](#)]
47. Van Leeuwen, W.J.; Huete, A.R.; Laing, T.W. MODIS vegetation index compositing approach: A prototype with AVHRR data. *Remote Sens. Environ.* **1999**, *69*, 264–280. [[CrossRef](#)]
48. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
49. Das, A.J.; Stephenson, N.L.; Flint, A.; Das, T.; Van Mantgem, P.J. Climatic correlates of tree mortality in water-and energy-limited forests. *PLoS ONE* **2013**, *8*, 69917–69927. [[CrossRef](#)] [[PubMed](#)]
50. Rich, P.M.; Hetrick, W.A.; Saving, S.C. *Modeling Topographic Influences on Solar Radiation: A Manual for the Solarflux Model*; Los Alamos National Lab.: Los Alamos, NM, USA, 1995.
51. Zon, R. *Forests and Water in the Light of Scientific Investigation*; U.S. Government Printing Office: Statesboro, GA, USA, 1927.
52. Kogan, F.; Stark, R.; Gitelson, A.; Jargalsaikhan, L.; Dugrajav, C.; Tsooj, S. Derivation of pasture biomass in mongolia from AVHRR-based vegetation health indices. *Int. J. Remote Sens.* **2004**, *25*, 2889–2896. [[CrossRef](#)]
53. Freitas, S.R.; Mello, M.C.S.; Cruz, C.B.M. Relationships between forest structure and vegetation indices in atlantic rainforest. *For. Ecol. Manag.* **2005**, *218*, 353–362. [[CrossRef](#)]
54. Linderman, M.A.; An, L.; Bearer, S.; He, G.; Ouyang, Z.; Liu, J. Modeling the spatio-temporal dynamics and interactions of households, landscapes, and giant panda habitat. *Ecol. Model.* **2005**, *183*, 47–65. [[CrossRef](#)]
55. Foody, G.M.; Boyd, D.S.; Cutler, M.E. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* **2003**, *85*, 463–474. [[CrossRef](#)]
56. Aarts, G.; Fieberg, J.; Brasseur, S.; Matthiopoulos, J. Quantifying the effect of habitat availability on species distributions. *J. Anim. Ecol.* **2013**, *82*, 1135–1145. [[CrossRef](#)] [[PubMed](#)]
57. Matthiopoulos, J.; Hebblewhite, M.; Aarts, G.; Fieberg, J. Generalized functional responses for species distributions. *Ecology* **2011**, *92*, 583–589. [[CrossRef](#)] [[PubMed](#)]
58. Chen, Q.; Laurin, G.V.; Battles, J.J.; Saah, D. Integration of airborne LiDAR and vegetation types derived from aerial photography for mapping aboveground live biomass. *Remote Sens. Environ.* **2012**, *121*, 108–117. [[CrossRef](#)]
59. Baldwin, D.J.; Weaver, K.; Schneckeburger, F.; Perera, A.H. Sensitivity of landscape pattern indices to input data characteristics on real landscapes: Implications for their use in natural disturbance emulation. *Landsc. Ecol.* **2004**, *19*, 255–271. [[CrossRef](#)]
60. Buyantuyev, A.; Wu, J. Effects of thematic resolution on landscape pattern analysis. *Landsc. Ecol.* **2007**, *22*, 7–13. [[CrossRef](#)]
61. Moody, A.; Woodcock, C. Scale-dependent errors in the estimation of land-cover proportions: Implications for global land-cover datasets. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 585–594.
62. Proença, V.; Pereira, H.M.; Vicente, L. Resistance to wildfire and early regeneration in natural broadleaved forest and pine plantation. *Acta Oecol.* **2010**, *36*, 626–633. [[CrossRef](#)]

63. Box, G.E.; Draper, N.R. *Empirical Model-Building and Response Surfaces*; John Wiley & Sons: New York, NY, USA, 1987.
64. Woodcock, C.E. *Uncertainty in Remote Sensing and GIS*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2002; pp. 19–24.
65. Vancoillie, F.; Verbeke, L.; De Wulf, R. *Artificial Neural Network Training for Savanna Vegetation Mapping: Transferring Previously Learned Experience to New Learning Tasks*; International Workshop on Geo-Spatial Knowledge Processing for Natural Resource Management: Varese, Italy, 2001.



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).