*Remote Sens.* **2014**, *6*, 10750-10772; doi:10.3390/rs61110750

OPEN ACCESS

*remote sensing*

**ISSN 2072-4292**

www.mdpi.com/journal/remotesensing

*Article*

# Estimation of Airborne Lidar-Derived Tropical Forest Canopy Height Using Landsat Time Series in Cambodia

**Tetsuji Ota [1,†,*], Oumer S. Ahmed [2,†], Steven E. Franklin [3], Michael A. Wulder [4], Tsuyoshi Kajisa [5], Nobuya Mizoue [1], Shigejiro Yoshida [1], Gen Takao [6], Yasumasa Hirata [6], Naoyuki Furuya [7], Takio, Sano [8], Sokh Heng [9] and Ma Vuthy [9]**

[1] Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, Japan; E-Mails: mizoue@agr.kyushu-u.ac.jp (N.M.); syoshida@agr.kyushu-u.ac.jp (S.Y.)

[2] Geomatics, Remote Sensing and Land Resources Laboratory, Department of Geography, Trent University, 1600 West Bank Drive Peterborough, Ontario K9J 7B8, Canada; E-Mail: osahmed@trentu.ca

[3] Department of Environmental and Resource Studies/Science, Department of Geography, and Office of the President, Trent University, 1600 West Bank Drive Peterborough, Ontario K9J 7B8, Canada; E-Mail: sfranklin@trentu.ca

[4] Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 West Burnside Road, Victoria, British Columbia V8Z 1M5, Canada; E-Mail: Mike.Wulder@NRCan-RNCan.gc.ca

[5] Faculty of Agriculture, Kagoshima University, Korimoto 1-21-24, Kagoshima 890-8580, Japan; E-Mail: k2083442@agri.kagoshima-u.ac.jp

[6] Department of Forest Management, Forestry and Forest Products Research Institute, Matsunosato 1, Tsukuba 305-8687, Japan; E-Mails: takaogen@affrc.go.jp (G.T.); hirat09@affrc.go.jp (Y.H.)

[7] Hokkaido Research Center, Forestry and Forest Products Research Institute, Hitsujigaoka 7, Toyohiraku, Sapporo 062-8516, Japan; E-Mail: nfuruya@affrc.go.jp

[8] Asia Air Survey Co., LTD, Shinyuri 21 Building, 1-2-2 Manpukuji, Asao-ku, Kawasaki 215-0004, Japan; E-Mail: tk.sano@ajiko.co.jp

[9] Forest-Wildlife Research and Development Institute, Forestry Administration, Khan Sen Sok, Phnom Penh 12157, Cambodia; E-Mails: sokhhengpiny@yahoo.com (S.H.); vuthydalin@yahoo.com (M.V.)

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: chochoji1983@gmail.com; Tel.: +81-926-422-868; Fax: +81-926-422-867.

External Editors: Nicolas Baghdadi; Prasad S. Thenkabail

*Received: 9 July 2014; in revised form: 3 October 2014 / Accepted: 8 October 2014 / Published: 6 November 2014*

**Abstract:** In this study, we test and demonstrate the utility of disturbance and recovery information derived from annual Landsat time series to predict current forest vertical structure (as compared to the more common approaches, that consider a sample of airborne Lidar and single-date Landsat derived variables). Mean Canopy Height (MCH) was estimated separately using single date, time series, and the combination of single date and time series variables in multiple regression and random forest (RF) models. The combination of single date and time series variables, which integrate disturbance history over the entire time series, overall provided better MCH prediction than using either of the two sets of variables separately. In general, the RF models resulted in improved performance in all estimates over those using multiple regression. The lowest validation error was obtained using Landsat time series variables in a RF model ($R^2$ = 0.75 and RMSE = 2.81 m). Combining single date and time series data was more effective when the RF model was used (opposed to multiple regression). The RMSE for RF mean canopy height prediction was reduced by 13.5% when combining the two sets of variables as compared to the 3.6% RMSE decline presented by multiple regression. This study demonstrates the value of airborne Lidar and long term Landsat observations to generate estimates of forest canopy height using the random forest algorithm.

**Keywords:** Landsat time series; airborne Lidar; canopy height; random forest

## 1. Introduction

Forests sequester and store more carbon than any other terrestrial ecosystem and are considered to be an important natural "brake" on climate change [1]. Tropical forests are especially important for carbon sequestration in the biosphere. For example, it is estimated that tropical forests store 229–247 Pg C [2,3]. However, tropical forests are subject to significant change, through agricultural expansion, forest management, and other natural and anthropogenic processes [4,5]. The policy and management decisions governing these forests require consistent and periodic information on forest structure; hence, there is an increasing need to generate accurate information regarding forest structural dynamics [6]. Field measurement, while typically of high accuracy, is costly, labor intensive, and limited in spatial scales, providing quality information for a limited number of stands in a given area [7]. Remote sensing may play an increasingly important role to provide estimates of required information on forest structure in a consistent and systematic manner across a variety of scales.

Airborne Lidar (Light Detection and Ranging) is an active remote sensing system well suited to measure specific forest information, including tree height [8], basal area [9], leaf area index [10], canopy

cover [11,12], vertical vegetation strata [13], successional stages [14], and canopy architecture [15]. In the case of tropical forests, Asner *et al.* [16,17] assessed the capability of mean canopy height (MCH) derived from airborne Lidar to estimate the above-ground carbon in Madagascar and Colombia, demonstrating that Lidar-derived MCH could be used to accurately estimate above ground carbon. Airborne Lidar is currently the most accurate remote sensing system to obtain specific site-level data on forest structure. However, wall-to-wall acquisitions of Lidar data remain cost prohibitive for large forest areas or in areas without a competitive commercial Lidar data acquisition community. In an effort to overcome cost limitation of Lidar data, recent studies report on the integration of Lidar data with optical satellite remotely sensed data to estimate canopy structure from a relatively small area covered by Lidar sample over a larger area covered by the optical scene [18–21]. These studies mainly focused on managed temperate forests and to date there is a paucity of such studies in unmanaged tropical forests.

Satellite remote sensing datasets, such as those from the Landsat sensors, Multispectral Scanner (MSS), Thematic Mapper (TM), and Enhanced Thematic Mapper Plus (ETM+), offer the capacity to map land cover, dynamics, and structural characteristics in a systematic, repeatable and cost effective fashion over long time periods and over a variety of spatial extents [22–24]. The free and open access of Landsat data since 2008 [25] has removed cost limitations to accessing large numbers of images while also reducing processing overhead through provision of a robust series of standard products [26]. Predicting forest structure variables with Landsat data has been a research topic of great interest (e.g., [27–30]. However, an asymptotic relationship is typically found when using Landsat data alone to make empirical predictions of forest structure [31], with the asymptote linked to canopy density, crown closure or canopy cover. The integration of such satellite multispectral remote sensing data with information from airborne Lidar provides opportunities to capitalize upon the distinctive characteristics of both data sources.

Since the forest structure at a given time is often related to the forest's disturbance history, one possible means of improving the prediction of forest structure using Landsat is by including information on forest disturbance trends prior to the date for which estimates are needed. For example, Gómez *et al.* [32] described the linkages between forest structure and various processes influencing stand condition, development, and disturbance legacy. Their results showed that forest structural conditions can be related to a complex suite of stand development trajectories and processes, such as regeneration status and rates that are influenced by forest management practices and natural disturbance regimes. The Landsat temporal depth and the increasingly well-understood radiometric characteristics of such data provide an opportunity to develop more accurate and reliable change detection and analysis methods. Recently, a trajectory-based automated method for characterizing forest disturbance was developed for reconstructing forest disturbance history and age structure using Landsat time series stacks [33]. This method provides estimates of discontinuous phenomena (disturbance date and intensity) as well as continuous phenomena (post-disturbance regeneration). This information enables the generation of a spectral trajectory that relates to the growth history of a given forest stand which may provide additional improvements to predict current forest structure [34]. For example, Pflugmacher *et al.* [35] demonstrated the inclusion of disturbance and recovery metrics derived from spectral profiles of annual Landsat time series provided an improved prediction of current forest structural attributes. Ahmed *et al.* [18] suggested that pre-stratification of the forest types based on the disturbance date and the addition of time since disturbance information improved forest canopy cover and height estimation in a managed temperate

forest. Li *et al.* [36] showed that inclusion of disturbance information characterized by vegetation change tracker algorithm yielded the lowest validation error for the estimation of young forest height calculated from space borne LiDAR. All these studies are mainly focused on managed temperate forests and to date there is a paucity of such studies in tropical forests where frequent small clear cutting and regrowth are expected to dominate. Helmer *et al.* [37] illustrated the relationship between field measured forest height and disturbance information derived from Landsat and ALI image mosaics using a regression tree model in tropical forest. However, in their study, a time series trajectory based approach was not implemented.

Most studies that predict forest structure variables with remote sensing data commonly apply parametric approaches [12,20,21,38]. The most typical parametric approach used is multiple regression, which defines relationships between image data and forest structure variables. Although it is frequently used and simply interpreted, multiple regression often lacks the capacity to characterize forest complexity [19]. This type of parametric approach assumes a linear relationship between independent and dependent variables, and the model errors depend on a pre-defined normal distribution. This will often have a negative impact on the estimation accuracy if the statistical assumptions are not fulfilled. Alternatively, the use of non-parametric machine learning models has been reported for implicitly inferring unknown relationships underlying a given dataset [39]. As opposed to linear regression, most of these machine learning models are versatile enough to uncover complicated non-linear relationships. Machine learning algorithms, such as the random forest (RF), do not have any assumptions about the relationship between dependent and independent variables and are well suited for analyzing complex non-linear and possibly hierarchical interactions in large data sets [40]. Additionally, some machine learning algorithms are also useful in alleviating or even circumventing the difficulties caused by data dimensionality considerations [39,41], which is a common problem in fitting models with a large number of predictors. Powell *et al.* [30] compared three empirical modeling approaches: reduced major axis regression, gradient nearest neighbor, and RF, to predict live above ground biomass from Landsat data and demonstrated that the RF model yielded the best results in terms of RMSE. In addition, Hudak *et al.* [42] compared several nonparametric regression methods including RF for estimating basal area and tree density in managed mixed-conifer forests. Their study also concluded that the RF model was the most robust and flexible model among the considered models.

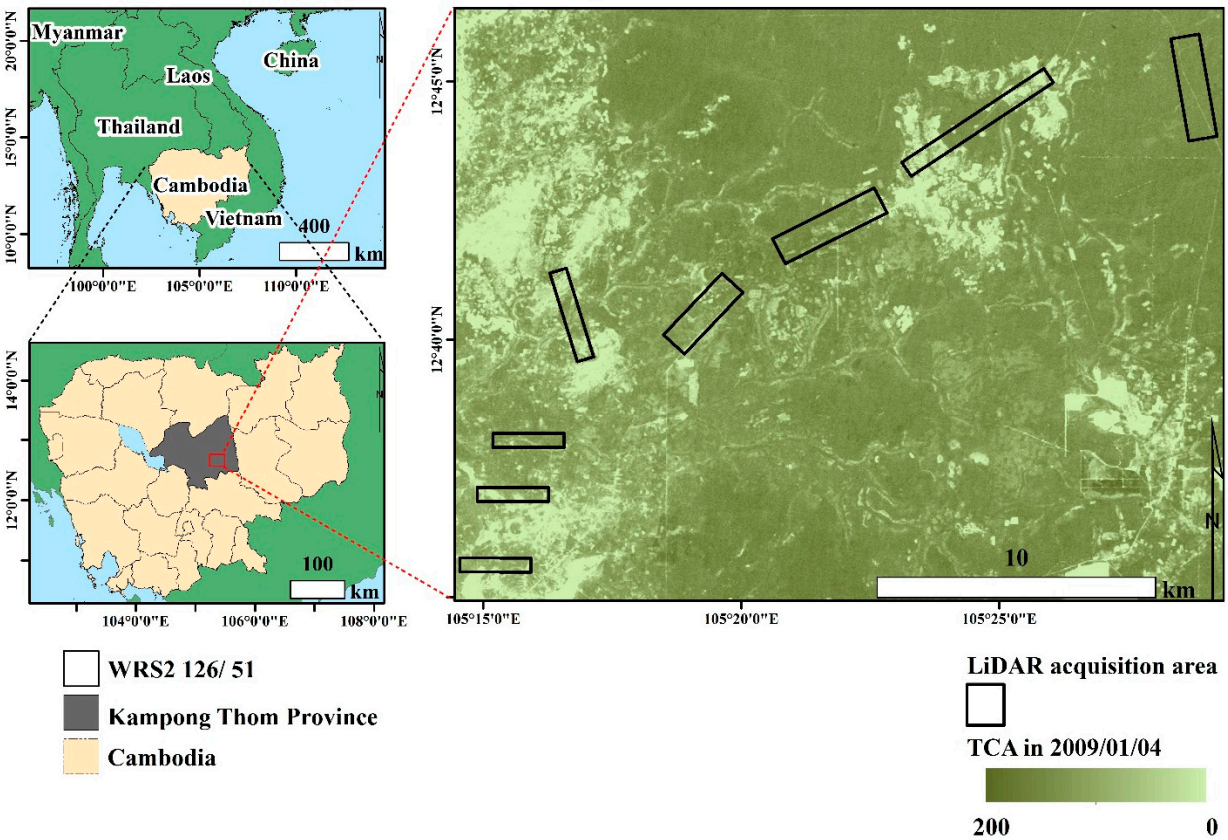
In this study, our main objective is to test and demonstrate the utility of disturbance and recovery data from Landsat time series to predict Lidar-derived canopy height in a tropical forest using a machine learning model. The temporally inclusive results, based upon metrics generated from the image time series, are compared to the more common approach of utilizing single-date Landsat derived variables for estimation in a multiple regression model. To address the main objective, we first apply a trajectory-based disturbance characterization approach to extract disturbance information from the Landsat time series to be used in the canopy height modeling. Next, we develop multiple regression and RF models to estimate forest mean canopy height using single date Landsat variables and time series variables derived from the trajectory characterization method. Finally, we compare the model performance between single date, time series, and combined single date and time series derived models. We hypothesized that the use of disturbance history information generated from a time series of Landsat imagery will improve our predictive capacity in the machine learning approach. The improved predictive capacity is desired in support of large area mapping of forest structural attributes using Landsat imagery and samples of Lidar data.

## 2. Materials and Methods

### *2.1. Study Area*

The study area is located in Kampong Thom Province in central Cambodia. This province, which is between 12°11′ and 13°26′N and 104°12′ and 105°44′E, has a total land area of 12,447 km$^2$, about 7% of the country (Figure 1) [43]. Ecological and geographical conditions are relatively uniform throughout the province. Atmospheric humidity is high throughout the year, ranging between 72% and 87% with an annual mean of 80%. The climate is tropical with a bi-annual change in monsoonal wind systems. The rainy season extends from May to October and the dry season from November to April. Mean annual rainfall and temperature have been measured at 1700 mm and 28.0 °C, respectively.

**Figure 1.** The study area located in Kampong Thom Province, Cambodia and the location of LiDAR data strips.




Within this province, we selected a 1632 ha unmanaged forest area with available remote sensing imagery to support the analysis. The study area is mainly covered by evergreen, deciduous, and other forest, based on the definitions of the Forestry Administration of Cambodia. Typically, the evergreen forest is a dense forest with a dominant tree height of 30–40 m, whereas the sparse deciduous forest has a dominant tree height between 10 and 15 m. Other forest contains a variety of forests such as re-growth area after logging, inundated forests, and plantation forest. Even though the study area is located in tropical forest where shifting cultivation (repeated small clear cutting and regrowth) is expected to be common, this practice is rare in the study area. In Cambodia, shifting cultivation is mainly found in

upland (mountainous) forests (e.g., [44,45]), and is rare in lowland forests such as the forests in central Cambodia. Even in areas where such practice used to be dominant (e.g., Ratanakiri province) studies indicate that shifting cultivation is decreasing and replaced by rubber and palm oil plantation [46].

### *2.2. Landsat Data*

The Landsat imagery for this study were obtained from the United States Geological Survey (USGS) Landsat archive and consist of 17 dry-season (November–February) images acquired between 1983 and 2011 (Table 1). We used a combination of Landsat Multispectral Scanner (MSS), Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) imagery. All of the MSS, TM and ETM+ data were terrain corrected and were converted to "top-of-atmosphere-radiance" (L1T data product). The study area falls within WRS2 path/row 126/51. Based on the quality and cloud-free status, the 2009 TM image was selected as the reference image for all image to image registrations and subsequent atmospheric correction and normalization procedures. Landsat ETM+ images following the Scan Line Corrector failure in May 2003 [47] were clipped to exclude the scene edges that have missing data.

**Table 1.** Landsat imagery used in this study.

| Year/Month/Day | Satellite |
|---|---|
| 1983/12/1 | Landsat 4 MSS |
| 1984/12/11 | Landsat 5 MSS |
| 1986/12/1 | Landsat 5 MSS |
| 1990/12/4 | Landsat 4 TM |
| 1994/11/21 | Landsat 5 TM |
| 1996/1/27 | Landsat 5 TM |
| 1997/1/13 | Landsat 5 TM |
| 2000/12/31 | Landsat 7 ETM+ |
| 2003/2/7 | Landsat 7 ETM+ |
| 2003/12/24 | Landsat 7 ETM+ |
| 2004/11/8 | Landsat 7 ETM+ |
| 2006/11/30 | Landsat 7 ETM+ |
| 2008/1/20 | Landsat 7 ETM+ |
| 2009/1/14 | Landsat 5 TM |
| 2009/12/24 | Landsat 7 ETM+ |
| 2010/12/27 | Landsat 7 ETM+ |
| 2011/12/30 | Landsat 7 ETM+ |

### *2.3. Airborne Lidar Data*

Airborne Lidar data were acquired from a helicopter platform using a discrete-return small-footprint Airborne Laser Terrain Mapper (Optech, Inc., ALTM 3100) in 18–20 January 2012 (Table 2). The Lidar data were acquired in eight separate strips within three consecutive days (Figure 1) using similar instrument parameters. Based on the pulse frequency, the Lidar data had a hit density of more than 20 pulses/m$^2$. The last return pulses were used to model the ground surface by filtering earlier returns.

Standard practices were followed [48] to produce a triangulated irregular network (TIN) that allowed for development of a digital terrain model (DTM). The DTM allowed for generation of terrain surface height values for each point above ground. The relative height of each point was computed as the difference between the height of the first return and the interpolated terrain surface height. Finally, the mean canopy height (MCH) for the study area was derived with 30-m grid cell size to approximate the available Landsat spatial resolution.

**Table 2.** Airborne Lidar parameters.

| Parameter | Optech ALTM 3100 |
|---|---|
| Flight altitude (above ground) | 500 m |
| Flying Speed | 25 m/s |
| Pulse repetition frequency | 100 kHz |
| Scan frequency | 53 Hz |
| Laser beam divergence | 0.5 mrad |
| Wave length | 1064 nm |

### *2.4. Methods*

#### 2.4.1. Landsat Image Processing

In order to characterize the complex nature of long-term continuous (e.g., defoliation by insects, growth) and short-term abrupt (e.g., fire, harvest) forest disturbances that may exist in the study area, we used annual Landsat time series between 1983 and 2011. This required integrating Landsat data from MSS, TM and ETM+ sensors.

Based on image quality and cloud-free status, the 2009 TM image was selected as the reference image, and all images were co-registered to the 2009 TM image with an RMSE of less than 15 m using an automated registration and orthorectification package (AROP) [49]. Clouds and cloud shadows were masked out using manual approaches and were excluded from subsequent processing and analysis. All the MSS imagery (Table 1) were resampled using nearest neighbor methods to 30 m to match the TM and ETM+ spatial resolution.

To minimize annual variation in atmospheric conditions, we normalized each image in the Landsat time series to the 2009 TM reference image. First, the reference image was converted to surface reflectance using the cosine estimation of atmospheric transmittance (COST) absolute radiometric correction model developed by Chavez [50]. All other TM and ETM+ images in the time series were radiometrically normalized to the COST image by means of MADCAL (Multivariate Alteration Detection and Calibration algorithm [51] as recommended by Schroeder *et al.* [52]. To normalize MSS data with the TM and ETM+ time series we used Tasseled Cap Transformation (TCT) components with MADCAL rather than the individual spectral bands [53]. The coefficients used to create the TCT components were derived statistically from images and empirical observations and are specific to each imaging sensor. The first TCT component corresponds to the overall "brightness" of the image, and, by definition, is a positive value. The second TCT component corresponds to "greenness" and is typically used as an indicator of the amount of photosynthetically active vegetation. The values for greenness depend on the contrast between the visible and near-infrared bands, with exposed soil having negative values [54] and vegetated

areas having positive values. The third TCT component is often labeled "wetness" and is usually interpreted in vegetated areas as an index of canopy structure, soil or surface moisture, or possibly an estimate of the amount of dead or dried vegetation [23]. The brightness and greenness components derived from the 2009 Landsat TM image were used as the reference source for MADCAL. Brightness and greenness for the MSS imagery were generated using the modified coefficients developed by Pflugmacher *et al.* [35]. For the MSS data, the TCT wetness component cannot be generated due to the lack of the requisite short-wave infrared band. TCT brightness and greenness components for the 2009 TM reference image were generated using the coefficients for reflectance [55].

The image normalization process, above, creates a spectral space that is relatively consistent across sensors. The TCT was then used for all Landsat images in the time series. TCT is a well-known linear transformation widely employed [23] to characterize forest structure [56,57]. (For all TM/ETM+ images TCT was applied using the coefficients for reflectance data [55] and we obtained brightness (TCB), greenness (TCG), and wetness (TCW). Brightness and greenness define the vegetation plane [58] and provide a practical bridge between the earlier MSS imagery and more recent TM and ETM+ imagery [53].

MSS sensors lack short-wave infrared bands, and therefore wetness cannot be computed for Landsat MSS imagery. To compensate, Powell *et al.* [30] developed an index called the Tasseled Cap Angle (TCA) that is based on the angle formed by brightness and greenness in the vegetation plane (Equation (1)). TCA has been interpreted as an indicator of the proportion of vegetation to non-vegetation within a Landsat MSS, ETM+ or TM pixel [30,32]:

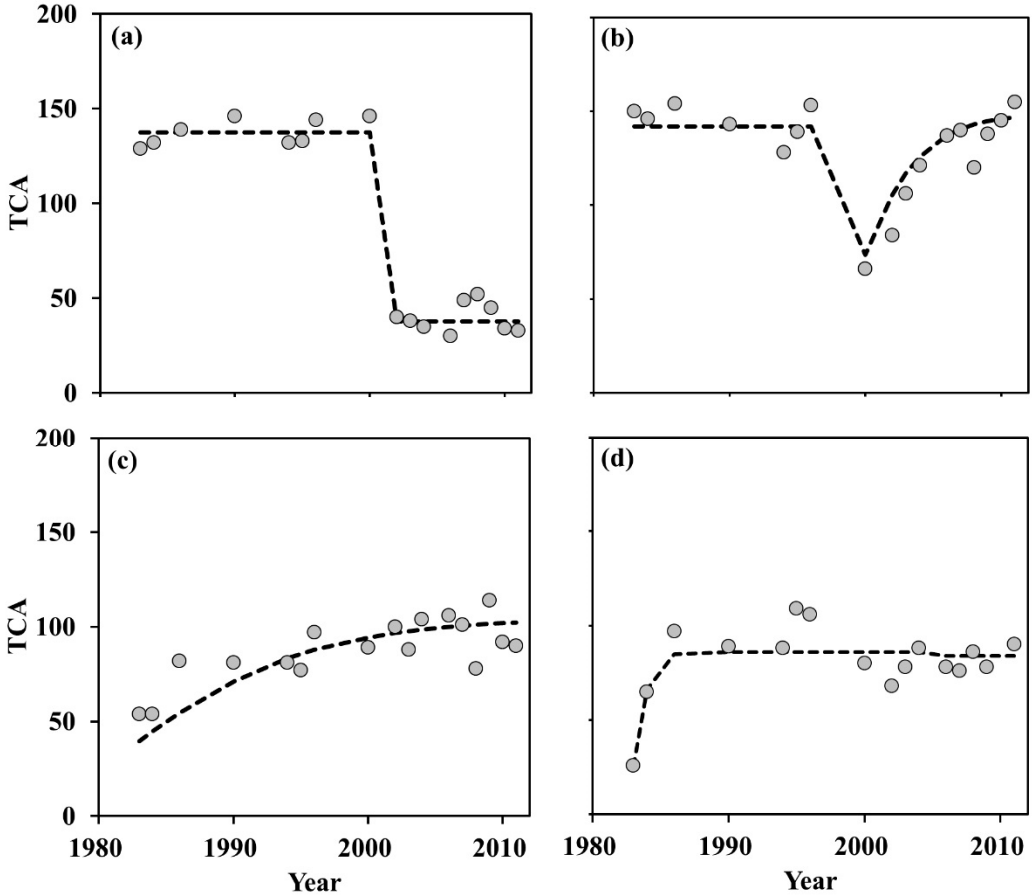
$$\mathrm{TCA} = \arctan(\mathrm{TCG}/\mathrm{TCB}) \tag{1}$$

Higher values of greenness and lower values of brightness are often found over forests, and therefore, higher values of TCA are more likely in dense cover classes when compared with open stands or clearcuts [59,60]. TCA was used to describe disturbance history in this study. We used the Tasseled Cap Transform components TCB, TCG, and TCW of the 2011 image in our statistical modeling to estimate mean canopy height.

#### 2.4.2. Disturbance History from Landsat Spectral Trajectories

An automated curve-fitting algorithm was used to characterize change in TCA over the available 28-year Landsat time series (1983–2011) building upon the logic of Kennedy *et al.* [33] and Ahmed *et al.* [34]. In addition to "intact and undisturbed forest", using the greatest disturbance in the time series, the trajectory-based change detection method identified four other disturbance classes of temporal trajectories of change in TCA: Class 1) disturbance, Class 2) disturbance followed by revegetation, Class 3) ongoing revegetation, and Class 4) revegetation to a stable state (Figure 2). The interpretation of spectral characteristics of these classes was similar to Gómez *et al.* [32], who found dense forest stands with higher TCA values than more open stands or bare soil, and TCA values in recent clearcuts significantly lower than in any other cover stage of the forest. Generally, TCA displays a clear increasing trend with time-since-disturbance.

**Figure 2.** Example of the four disturbance class spectral trajectories and parameters fit to each trajectory. (**a**) disturbance, (**b**) disturbance followed by revegetation, (**c**) ongoing revegetation, and (**d**) revegetation to a stable state (After Ahmed *et al.* [34]).

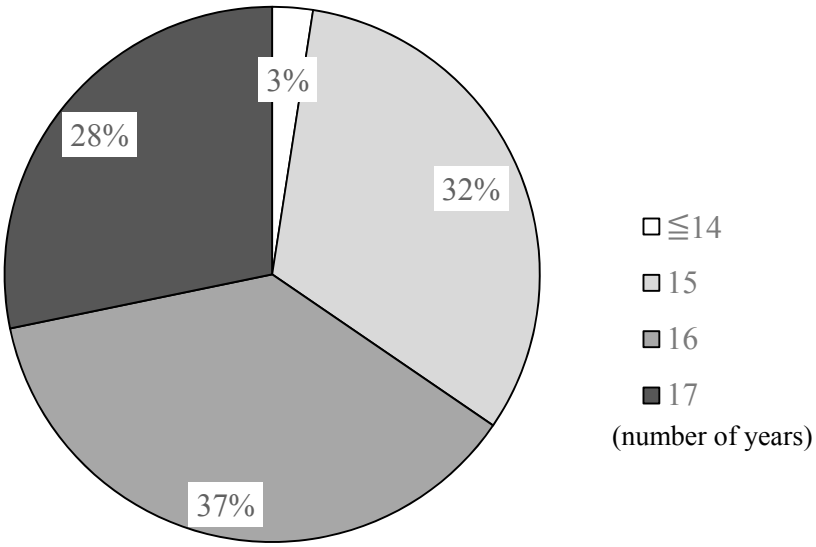



First, we extracted TCA for each pixel in the Landsat time series (1983–2011). Most pixels have more than 14 years of data (Figure 3). The shortest data for a pixel was composed of 12 year time series. For each of the four classes (Figure 2), we created initial estimates of spectral trajectory shape parameters, and then used a curve fitting function to adjust these initial parameters to find the best fit of the potential trajectory to the observed trajectory. A detailed description of the curve fitting process and the procedures used for selecting the best model can be found in Ahmed *et al.* [34]. For the curve fitting processing, we used minpack.lm, which is add-on packages to implement Levenberg-Marquardt algorithm in R, a multiplatform, open-source language and software for statistical computing [61].

The temporal curve-fitting was applied to the time-series of Landsat-based TCA values listed in Table 1. This trajectory characterization method produces a suite of products, including disturbance maps and measures for characterizing the detected disturbances and for tracking post-disturbance processes [33]. To limit the scope of our research, this study, will not attempt to fully characterize the forest disturbance condition in the study area. However, we extract relevant forest disturbance data from the Landsat time series and test the utility of the extracted disturbance and recovery data in predicting Lidar-derived canopy height. From the fitted trajectories, we calculated disturbance variables that characterize the disturbance and recovery history of each pixel. First, we identified the greatest disturbance as the point with the greatest negative change in the trajectory. For each of the detected greatest disturbances, time

since disturbance (TSD) is calculated as the difference between model target year (2011) and the disturbance year.

**Figure 3.** The proportion of time series data length (in years) for pixels.




Disturbance intensity was also calculated using TCA value at the time of the greatest disturbance. In this study, in addition to the Landsat Tasseled Cap Transform derived variables mentioned in the previous section, TSD, disturbance intensity and class were also used as predictor variables to model canopy height in the study area. This helped to examine the added canopy height estimation advantage of time series derived predictor variables over single date derived spectral variables.

2.4.3. Datasets for Mean Canopy Height Estimation

Canopy height was modeled separately based on two different reference datasets using multiple regression and RF algorithm, and the results were compared. For the dependent and independent variables random data subsets were created for model fitting (reference dataset) and a second subset for validation (target dataset), which will be used to assess the performance of the prediction model. First, the LiDAR coverage was used to create datasets containing canopy height estimates from the LiDAR data and coincident spectral values from the Landsat image. These datasets were then randomly split in to calibration (2/3 of the data) and validation datasets (1/3 of the data). The estimation was performed within the extent of Lidar strips (Figure 1). Finally, the number of calibration points and that of validation points were 12,084 and 6042, respectively. To evaluate the contribution of the available disturbance history from Landsat trajectory to the canopy height prediction which will otherwise be based on single date spectral variables, in this study, the estimation of canopy height was mainly performed based on three different reference datasets *i.e.*, for each model, three sets of independent variables were used: (1) Single date variables (TCB, TCG, TCW, and TCA of the 2011 image); and (2) Time series variables (Time since disturbance (TSD), disturbance intensity, disturbance class); (3) Combination single date and time series variables (1 and 2 above). When we used the combination of single date and time series variables, we constructed seven data sets using the single date variables and all possible combinations of

time series variables. Single date variables, TSD, and disturbance intensity were represented by a numeric data while disturbance class was represented by nominal data. The models were assessed using RMSE and the predictions were validated against the validation datasets.

### 2.4.4. Mean Canopy Height Estimation Using Multiple Regression and Random Forest

Before performing the multiple regression prediction, a check for the presence of multicollinearity between predictor variables in the regression analysis was carried out among the independent variables. Chen *et al.* [62] indicated that models formulated using a combination of exponential and quadratic form performed better than a simple linear model for estimating Lidar-measured canopy height using multispectral imagery. Therefore, the same type of nonlinear multiple regression model was employed in this study to relate Landsat-derived independent variables with the Lidar-derived mean canopy height:

$$\mathrm{MCH} = \exp\left(\sum_{i=1}^{n} a_i X_i + c\right) \tag{2}$$

where MCH is mean canopy height from Lidar data; $X_i$ is the $i$th independent variable; $a_i$, is coefficients for the $i$th variable; $c$ is constant value; and n is the number of independent variables.

The other estimation method used in this study is the RF algorithm which is an ensemble method comprised of many decision trees. This method fits several decision trees to a data set and then combines the predictions from all the trees, where each decision tree is constructed from a different bootstrap sample of the original dataset. The random selection of independent variables is performed at each level of the tree. A comprehensive explanation of the algorithm can be found in Breiman [63,64] and its application for forest parameter estimation can be found in Hudak *et al.* [42], McInerney and Nieuwenhuis [65], Powell *et al.* [30], Gleason and Im [66] and Ahmed *et al.* [18]. The RF modeling was performed using the "random forest" package [67] in R statistical language [61]. The "caret" package [68] was also used for tuning RF associated parameters. The k-fold cross validation was selected to create and optimize the regression model. In this study, a 10-fold cross validation technique was applied following Duro *et al.* [69].

It should be indicated that RF used the same time series and single date spectra derived input variables as those used in multiple regression. This allows the two types of models (multiple regression and RF) to be compared in a straightforward way.

### 2.4.5. Comparison of Model Performance for Estimating Lidar-Measured Mean Canopy Height

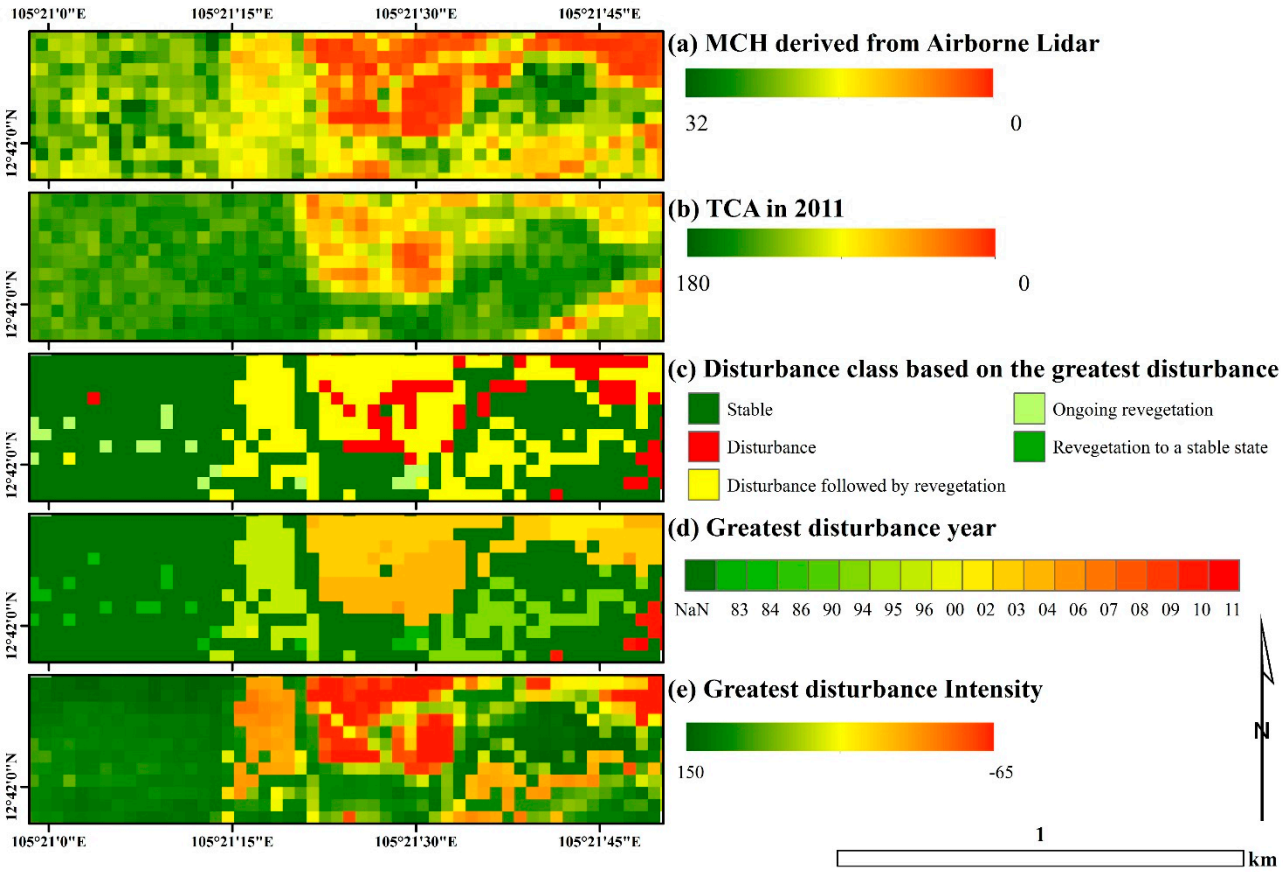
Multiple regression and RF models were developed using separate training samples from single date variables and Landsat trajectory derived time series variables. Additional models were also developed by combining samples from both single date and time series variables. The variables extracted from Landsat were imported into both models to estimate Lidar-measured canopy height. Finally, the best models were assessed using RMSE and the predictions were validated against the entire validation (target) dataset which was not used in the model training process, $R^2$ and the RMSE are reported. Specifically, the comparison of model performance was made for three types of reference dataset (*i.e.*, single date, time series and the two combined).

## 3. Results

### *3.1. Trajectory-Based Characterization of Forest Disturbance*

Figure 4 shows examples of MCH, TCA, and trajectory-based disturbance characterization. The trajectory curve fitting process resulted in a summary image with the attributes corresponding to the fitting parameters for each pixel whose trajectory matched one of the four potential trajectories in the disturbance classes at the $p$-value < 0.05. The disturbance class (Figure 4c) shows where disturbance has occurred. For each detected greatest disturbance, a disturbance year value, that is, the year when that disturbance occurred, is recorded (Figure 4d), and, disturbance intensity is calculated based on the TCA change at disturbance year (Figure 4e). The results from this trajectory characterization approach indicate over the period of 1983–2011, approximately 67.8% of the study area showed intact and undisturbed forest structure. The areas characterized by disturbance account for 23.6%, disturbance and revegetation 7.2%, ongoing revegetation 1.3% and revegetation to stable state account 0.1% of the study area, respectively. For forest stands regenerated immediately following stand clearing disturbances, this measure should be close to their actual age. The disturbance intensity was also estimated using the change in TCA value at disturbance.

**Figure 4.** Examples of MCH, TCA, and out puts from trajectory-based disturbance characterization: (**a**) MCH derived from Airborne LiDAR, (**b**) TCA in 2011, (**c**) Disturbance class based on the greatest disturbance, (**d**) Greatest disturbance year and (**e**) Greatest disturbance intensity.



### *3.2. Performance of Canopy Height Models*

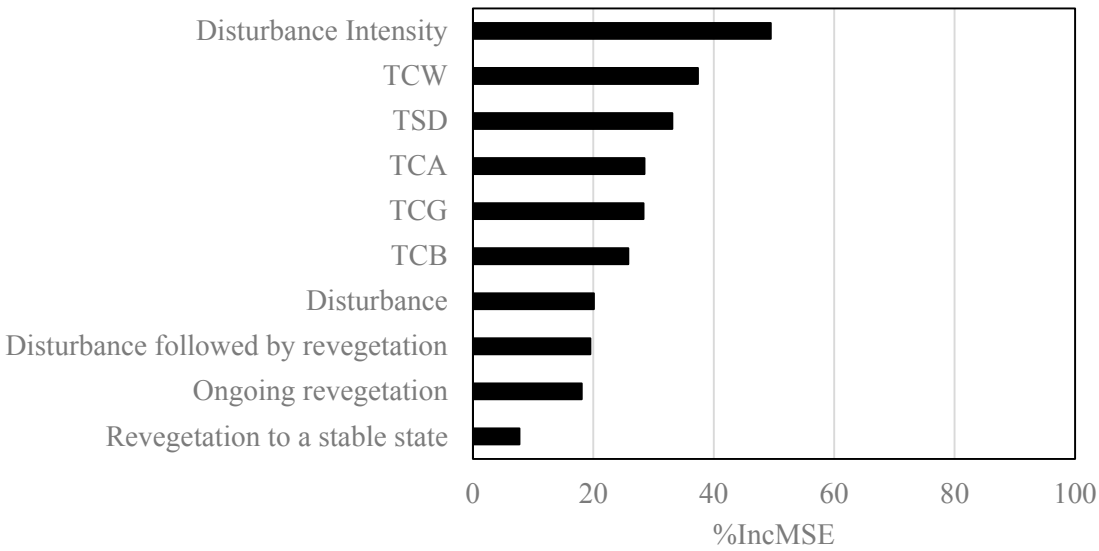
Table 3 shows the correlation between MCH and independent variables. The correlations were performed for the entire dataset. There was strong positive correlation between Lidar-derived MCH and TCT based variables, with lower though still significant correlations between MCH and trajectory derived variables. Figure 5 represents random forest variable importance, which is an indicator of a variables contribution to predictive accuracy. This indicates the average increase in mean square error (%IncMSE) when the variable in the model is replaced (permuted) with a random one. In this study, disturbance intensity was found to be the most important variable followed by Tasseled Cap Wetness and Time Since Disturbance.

**Table 3.** Correlations between independent variables and MCH.

| Variables | Correlation | |
|---|---|---|
| TCB | −0.70 | ** |
| TCG | 0.64 | ** |
| TCW | −0.73 | ** |
| TCA | 0.69 | ** |
| Disturbance Intensity | 0.64 | ** |
| TSD | 0.38 | ** |

** Significant at $\alpha < 0.01$.

**Figure 5.** Variable importance of random forest.




In this study, the RF models yielded higher $R^2$ and lower RMSE than multiple regression models both for single date and time series variables (Table 4). The combination of single date and time series variables, that integrated disturbance history over the entire time series, generally provided better MCH prediction than using either of the two sets of variables separately. The lowest error was obtained using Landsat time series variables in a RF model ($R^2$ = 0.75 and RMSE = 2.81 m) (Table 4, Figure 6). In the multiple regression models, single date variables provided better estimation compared to time series

variables, where as in the RF models both the variables provided nearly the same estimation. Combining single date and time series data was more effective when the RF model was used compared to multiple regression. The inclusion of time series variables showed slight improvement on the regression models. When the two variables are combined, the RMSE for RF was reduced by 13.5% as compared to the observed 3.6% decline in RMSE for multiple regression (Table 4). Among the time series variables disturbance intensity is found to be the most important variable with the highest contribution to the model estimation accuracy in both regression and RF models (Table 4).

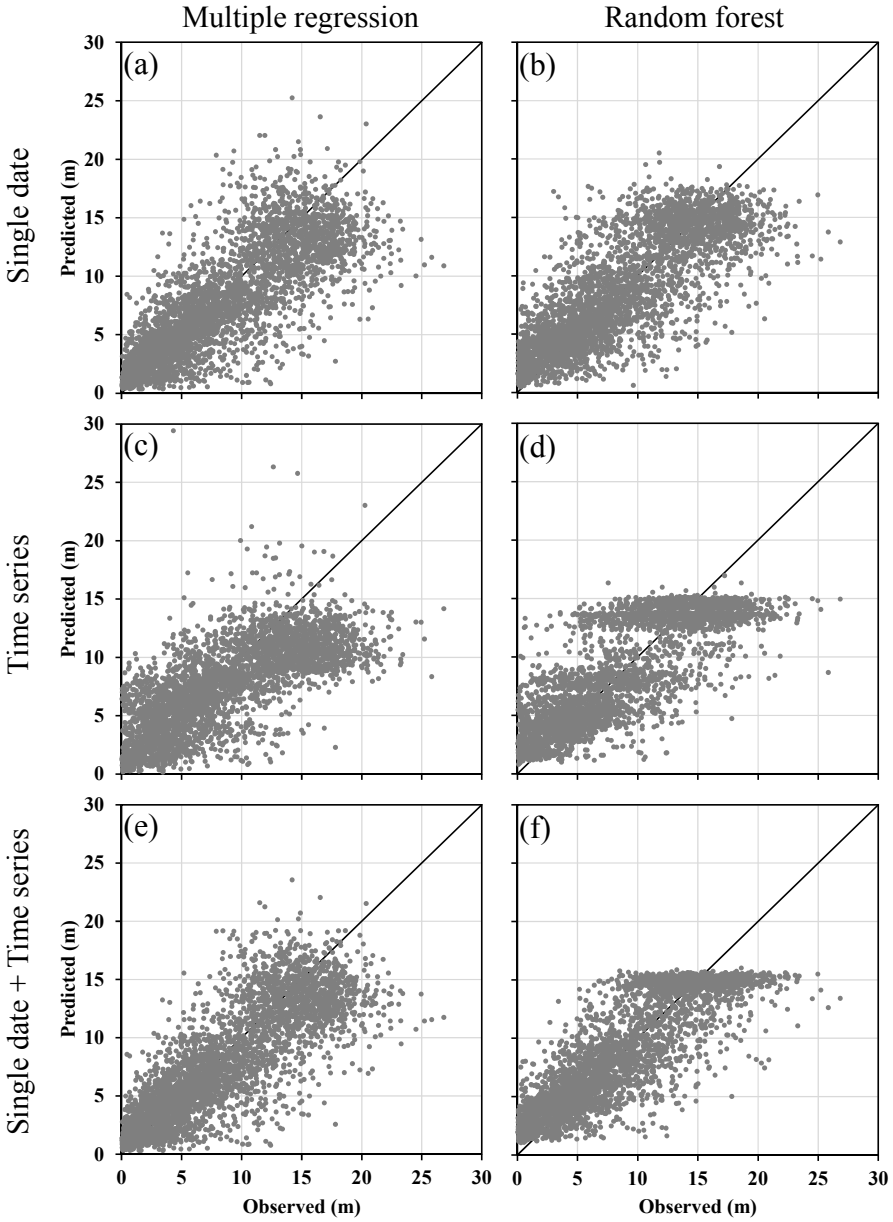
**Table 4.** Summary of MCH prediction using combination of variables with multiple regression and random forest.

| Model | Variables | | | | $R^2$ | RMSE |
|---|---|---|---|---|---|---|
| **Random forest** | Single date | | | | 0.66 | 3.25 |
| | Time series | | | | 0.66 | 3.24 |
| | Single date | Disturbance Intensity | | | 0.72 | 2.97 |
| | Single date | | TSD | | 0.68 | 3.14 |
| | Single date | | | Class | 0.70 | 3.05 |
| | Single date | Disturbance Intensity | TSD | | 0.73 | 2.88 |
| | Single date | Disturbance Intensity | | Class | 0.74 | 2.87 |
| | Single date | | TSD | Class | 0.71 | 3.00 |
| | Single date | Disturbance Intensity | TSD | Class | 0.75 | 2.81 |
| **Multiple regression** | Single date | | | | 0.65 | 3.38 |
| | Time series | | | | 0.55 | 3.91 |
| | Single date | Disturbance Intensity | | | 0.67 | 3.28 |
| | Single date | | TSD | | 0.66 | 3.35 |
| | Single date | | | Class | 0.66 | 3.35 |
| | Single date | Disturbance Intensity | TSD | | 0.67 | 3.28 |
| | Single date | Disturbance Intensity | | Class | 0.67 | 3.26 |
| | Single date | | TSD | Class | 0.66 | 3.25 |
| | Single date | Disturbance Intensity | TSD | Class | 0.67 | 3.26 |

## 4. Discussion

Several studies have demonstrated the value of using Lidar data to predict a range of forest characteristics with improved accuracies to conventional field inventory methods [70]. However, the cost associated with acquisition of airborne Lidar is high. Therefore, methods that integrate optical satellite imagery and airborne Lidar are required as a means to spatially extend Lidar measured attributes over larger areas. In this study, we evaluated the potential of combining samples of airborne Lidar data with Landsat time series to estimate forest canopy height in unmanaged tropical forest. We implemented Landsat-Lidar integration approach for modeling mean canopy height from historical Landsat observations in tropical forest located in Cambodia and compared the results with the more common approach of implementing single date Landsat image derived models.

**Figure 6.** Validation result observed *vs.* predicted for MCH estimation using single date (**a**,**b**), time series (**c**,**d**) and combining both single date, and time series (**e**,**f**) with multiple regression (a,c,e) and random forest (b,d,f). The solid line shows the 1:1 line.




Canopy height derived from airborne Lidar was correlated with both TCT variables and disturbance information obtained from a trajectory-based disturbance characterization method (Table 3), and a strong positive correlation is found between Lidar-derived MCH and the TCA. Our study has confirmed that TCA is particularly a well suited index to long-term change detection studies that include MSS data (e.g., [18,30]). Several other spectral variables or indices can also be used for disturbance characterization especially when MSS data are not included in the time series. Kennedy *et al.* [33] used short-wave infrared band reflectance for trajectory based disturbance characterization. Frazier *et al.* [71] used TCW to characterize disturbance with Landsat time series and they demonstrated the utility of the disturbance variables to estimate aboveground biomass. However, long-term disturbance characterization using the full temporal depth of the Landsat imagery e.g., [18] requires the inclusion of MSS imagery

and this limits the available variables to only few variables which can normalize the imagery across the Landsat sensors such as: Tasseled Cap transformation components (Brightness and Greenness) or the derivative spectral index (TCA), to be used in the long-term disturbance characterization. Here, it is important to note that, in our study, some explanatory power might be lost by restricting the data used to trends in TCA only and not including Tasseled Cap Wetness from the TM and ETM+ data.

The results from the trajectory characterization approach have a variety of potential uses. By detecting disturbance and recovery stages, the approach provides a better understanding of the evolving landscape for employing various land management techniques. Compared to other currently available disturbance characterization methods, trajectory based approach captured more parameters such as disturbance intensity. The examination of parameters captured by the trajectory based method and their relationship with forest vertical structure may indicate how disturbance history relates to current forest vertical structure conditions.

In this study, the correlation and variable importance results between the trajectory derived variables from the Landsat time series and Lidar MCH (Table 3, Figure 5) shows an interesting distinction between our study and previous studies that are conducted in a temperate forest setting. Unlike the studies in temperate managed forest that obtained higher correlation between MCH and TSD e.g., [18], in our study, TSD showed the lowest correlation with MCH. In managed temperate forest, mostly following stand clearing disturbances, there will be immediate regeneration of forest stands (e.g., through plantation), and the canopy closure can be achieved in a relatively shorter time as compared to the natural regeneration process that is expected in unmanaged tropical forest. Mostly, forest regeneration in tropical forest depends on the type and frequency of the disturbance and the ability of forest species to compete for nutrients and other resources. However, the trajectory derived variable "disturbance intensity" showed higher correlation with MCH (Table 3) and was the most important variable (Figure 5). These results indicate, in tropical forest the severity of disturbance is related more to the current canopy vertical structure than the time spent on regeneration (TSD) which depends on a number of factors in this region.

The RF models yielded higher $R^2$ and lower RMSE compared to multiple regression models in all estimations (Table 4). Previous studies showed the most robustness and flexibility of the RF model (e.g., [30,42]). The present study also demonstrated the utility of the RF model. Interestingly, the disturbance information from Landsat time series data was more effective when the RF model was used compared to multiple regression (Table 4). This result implies that the disturbance information from Landsat time series data might violate the assumptions required for applying multiple regression. Consequently, the RF models maximized the potential of disturbance information that is derived from Landsat time series data and provided the best estimation. Both the RF models with time series data showed a clear cut-off line around 15 m for the predicted canopy height and also the predicted canopy heights were overestimated when the heights are less than 5 m (Figure 6). One possible cause might be that since the RF model estimates values by averaging the estimation of many decision trees, it might tend to underestimate when the predicted value is close to the maximum value of training data. Similarly, when the estimated value is close to the minimum value of training data it might tend to overestimate. Further investigation is required on the functioning of the algorithm to reveal the actual causes.

We extracted disturbance information from the Landsat time series following the methodology developed by Kennedy *et al.* [33] and Ahmed *et al.* [18], and demonstrated that the resulting disturbance information improved the MCH prediction in a tropical forest. On the other hand, our MCH estimation

error is somewhat higher compared to Ahmed *et al.* [18], they estimated canopy height using sample of Lidar data and Landsat time series variables in a temperate managed forest with $R^2$ of 0.88 and RMSE of 2.39 m. In our study, the best model resulted in an $R^2$ of 0.75 with corresponding RMSE of 2.81 m using time series variables in a RF model (Table 4). The disturbance characterization method we used is mainly developed based on temperate managed forest and it considers only the single highest disturbance during the analysis period and it is not capable of detecting multiple disturbances. Forest disturbance from small scale illegal selective logging is anticipated in tropical forest which can result in multiple disturbances within a short observation period [72]. Complete characterization of forest disturbance in the study area being beyond the scope of our study, such continuous disturbances may not be readily detected by the disturbance characterization method, although focused investigations on subtle change have been successful using similar data to this study [73]. Following the methods presented in this research, we identify a multi-year temporal segment of change. The approach currently disregards other minor disturbances that occurred in the analysis period. Further study is required to develop algorithms which detect such multiple disturbances that are typical in tropical forest. Consequently, this will improve the time series information that can be extracted from Landsat which in turn might improve canopy height estimations.

The present study demonstrated the potential to predict airborne Lidar MCH from Landsat time series data. We showed Landsat time series data improved LiDAR derived canopy height estimation in unmanaged tropical forest (Table 4, Figure 6). The results are comparable to previous studies which have been conducted in temperate managed forests. The distinctive methodology implemented here can be used to spatially extend MCH beyond the Lidar extent which can in turn be used for large area, above ground carbon estimation.

## 5. Conclusions

In this study, we implemented Landsat-Lidar integration approach for modeling mean canopy height from single date and historical Landsat observations in a tropical forest in Cambodia. Mean Canopy Height (MCH) was estimated separately using single date, time series, and the combination of single date and time series variables in multiple regression and random forest (RF) models. The lowest estimation error was obtained using Landsat time series variables in a RF model ($R^2$ = 0.75 and RMSE = 2.81 m). The RMSE for RF mean canopy height prediction was reduced by 13.5% when including disturbance information derived from Landsat time series data as compared to the 3.6% RMSE decline presented by multiple regression. This study demonstrates the utility of disturbance and recovery data from Landsat time series to predict Lidar-derived canopy height in a tropical forest using a RF model, as compared to the more common approaches, that consider a sample of airborne Lidar and single-date Landsat derived variables.

Beneficial future research directions include implementing biophysical structure estimation over a large tropical forest area using Landsat time series data. The estimation of above ground biomass should be evaluated given the objectives of REDD+ (reducing emissions from deforestation and forest degradation) providing an additional characterization opportunity for enhancement of forest carbon stock estimation in developing countries. Further study is also required to detect multiple disturbances that are

typical in tropical forests to obtain a more accurate forest disturbance characterization, to possibly inform on degradation, and to consequently improve the canopy structure estimation in such environments.

## Acknowledgments

This research has been supported by Technology development for circulatory food production systems responsive to climate change from Ministry of Agriculture, Forestry and Fisheries, Japan and NSERC (Natural Sciences and Engineering Research Council) discovery grant awarded to Steven E. Franklin. The authors also acknowledge the valuable discussions with Joanne C. White at Canadian Forest Service, Natural Resources Canada.

## Author Contributions

All authors contributed extensively to the work. Tetsuji Ota and Oumer S. Ahmed performed the processing and lead writing of the manuscript. Steven E. Franklin and Michael A. Wulder reviewed the manuscript and gave comments and suggestions to the manuscript, especially with regards to the trajectory based disturbance characterization approach. Tsuyoshi Kajisa, Nobuya Mizoue, Shigejiro Yoshida supervised the research work, reviewed the manuscript and gave comments and suggestions to improve the manuscript. Gen Takao, Yasumasa Hirata and Naoyuki Furuya reviewed the manuscript and gave comments and suggestions to the manuscript, ensuring to include insights on tropical forests and related forest practices. Takio Sano contributed to acquiring airborne LiDAR data in Cambodia. Sokh Heng and Ma Vuthy liaised to obtain permission to research in Cambodian forest and provided significant knowledge about forests in Cambodia.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. 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*, 1–13.
2. Baccini, A.; Goetz, S.J.; Walker, W.S.; Laporte, N.T.; Sun, M.; Sulla-Menashe, J.; Hackler, P.; Beck, P.S.A.; Dubayah, R.; Friedl, M.A.; *et al.* Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nat. Clim.* **2012**, *2*, 182–185.
3. Saatchi, S.S.; Harris, N.L.; Brown, S.; Lefsky, M.; Mitchard, E.T.A.; Salas, W.; Zutta, B.R.; Buermann, W.; Lewis, S.L.; Hagen, S.; *et al.* Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9899–9904.
4. Geist, H.J.; Lambin, E.F. *What Drives Tropical Deforestation? A Meta-Analysis of Proximate and Underlying Causes of Deforestation Based on Subnational Case Study Evidence*; LUCC International Project Office: Louvain-la-Neuve, Belgium, 2001.
5. Geist, H.J.; Lambin, E.F. Proximate causes and underlying driving forces of tropical deforestation. *Bioscience* **2002**, *52*, 143–150.

6. Wulder, M.A.; Franklin, S.E. *Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches*; Taylor and Francis: Boca Raton, FL, USA, 2007.
7. Song, C.; Schroeder, T.A.; Cohen, W.B. Predicting temperate conifer forest successional stage distributions with multitemporal Landsat Thematic Mapper imagery. *Remote Sens. Environ.* **2007**, *106*, 228–237.
8. Hunter, M.O.; Keller, M.; Victoria, D.; Morton, D.C. Tree height and tropical forest biomass estimation. *Biogeosciences* **2013**, *10*, 8385–8399.
9. Stone, C.; Penman, T.; Turner, R. Determining an optimal model for processing Lidar data at the plot level: Results for a *Pinus radiata* plantation in New South Wales, Australia. *New Zeal. J. For. Sci.* **2011**, *41*, 191–205.
10. Moesera, D.; Roubineka, J.; Schleppib, P.; Morsdorfc, F.; Jonasa, T. Canopy closure, LAI and radiation transfer from airborne LiDAR synthetic images. *Agric. For. Meteorol.* **2014**, *197*, 158–168.
11. Donoghue, D.N.M.; Watt, P.J.; Cox, N.J.; Wilson, J. Remote sensing of species mixtures in conifer plantations using Lidar height and intensity data. *Remote Sens. Environ.* **2007**, *110*, 509–522.
12. Ahmed, O.S.; Franklin, S.E.; Wulder, M.A. Integration of LiDAR and Landsat to estimate forest canopy cover in coastal British Columbia. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 953–961.
13. Morsdorf, F.; Marell, A.; Koetz, B.; Cassagne, N.; Pimont, F.; Rigolot, E. Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. *Remote Sens. Environ.* **2010**, *114*, 1403–1415.
14. Falkowski, M.J.; Evans, J.S.; Martinuzzi, S.; Gessler, P.E.; Hudak, A.T. Characterizing forest succession with Lidar data: An evaluation for the Inland Northwest, USA. *Remote Sens. Environ.* **2009**, *113*, 946–956.
15. Palminteri, S.; Powell, G.V.N.; Asner, G.P.; Peres, C.A. LiDAR measurements of canopy structure predict spatial distribution of a tropical mature forest primate. *Remote Sens. Environ.* **2012**, *127*, 98–105.
16. Asner, G.P.; Clark, J.K.; Mascro, J.; Vaudry, R.; Chadwick, K.D.; Vieilledent, G.; Rasamoelina, M.; Balaji, A.; Kennedy-Bowdoin, T.; Maatoug, L.; *et al.* Human and environmental controls over aboveground carbon storage in Madagascar. *Carbon Balance Manag.* **2012**, *7*, doi:10.1186/1750-0680-7-2.
17. Asner, G.P.; Clark, J.K.; Mascaro, J.; Galindo García, G.A.; Chadwick, K.D.; Encinales, D.A.N.; Paez-Acosta, G.; Montenegro, E.C.; Kennedy-Bowdoin, T.; Duque, Á.; *et al.* High-resolution mapping of forest carbon stocks in the Colombian Amazon. *Biogeosciences* **2012**, *9*, 2683–2696.
18. Ahmed, O.S.; Franklin, S.E.; Wulder, M.A. 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.* **2014**, under review.
19. Chen, G.; Hay, G.J. A support vector regression approach to estimate forest biophysical parameters at the object level using airborne Lidar transects and quick bird data. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 733–741.
20. Hudak, A.T.; Lefsky, M.A.; Cohen, W.B.; Berterretche, M. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sens. Environ.* **2002**, *82*, 397–416.
21. Wulder, M.A.; Seemann, D. Forest inventory height update through the integration of Lidar data with segmented Landsat imagery. *Can. J. Remote Sens.* **2003**, *29*, 536–543.

22. Cohen, W.B.; Fiorella, M.; Gray, J.; Helmer, E.; Anderson, K. An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Photogramm. Eng. Remote Sens.* **1998**, *64*, 293–300.
23. Cohen, W.B.; Goward, S.N. Landsat's role in ecological applications of remote sensing. *Bioscience* **2004**, *54*, 535–545.
24. Wulder, M.A.; White, J.C.; Hay, G.J.; Castilla, G. Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery. *For. Chron.* **2008**, *84*, 221–230.
25. Woodcock, C.E.; Allen, R.; Anderson, M.; Belward, A.; Bindschadler, R.; Cohen, W.; Gao, F.; Goward, S.N.; Helder, D.; Helmer, E.; *et al.* Free access to Landsat imagery. *Science* **2008**, *320*, 1011.
26. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10.
27. Kajisa, T.; Murakami, T.; Mizoue, N.; Yoshida, S. Differences in spectral trajectory with stand volume development between Japanese larch and Japanese oak in Hokkaido, Japan. *J. For. Res.* **2007**, *12*, 435–441.
28. Kajisa, T.; Murakami, T.; Mizoue, N.; Top, N.; Yoshida, S. Object-based forest biomass estimation using Landsat ETM+ in Kampong Thom Province, Cambodia. *J. For. Res.* **2009**, *14*, 203–211.
29. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *For. Ecol. Manag.* **2004**, *198*, 149–167.
30. Powell, S.L.; Cohen, W.B.; Healey, S.P.; Kennedy, R.E.; Moisen, G.G.; Pierce, K.B.; Pierce, K.B.; Ohmann, J.L. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modelling approaches. *Remote Sens. Environ.* **2010**, *1145*, 1053–1068.
31. Duncanson, L.I.; Niemann, K.O.; Wulder, M.A. Integration of GLAS and Landsat TM data for aboveground biomass estimation. *Can. J. Remote Sens.* **2010**, *36*, 129–141.
32. Gómez, C.; White, J.C.; Wulder, M.A. Characterizing the state and processes of change in a dynamic forest environment using hierarchical spatio-temporal segmentation. *Remote Sens. Environ.* **2011**, *115*, 1665–1679.
33. Kennedy, R.E.; Cohen, W.B.; Schroeder, T.A. Trajectory-based changedetection for automated characterization of forest disturbance dynamics. *Remote Sens. Environ.* **2007**, *110*, 370–386.
34. Ahmed, O.S.; Franklin, S.E.; Wulder, M.A. The interpretation of forest disturbance using Landsat imagery time-series and airborne LiDAR-derived canopy cover data in British Columbia. *Can. J. Remote Sens*. **2014**, *39*, 521–542.
35. Pflugmacher, D.; Cohen, W.B.; Kennedy, R.E. Using Landsat-derived disturbance history 1972–2010 to predict current forest structure. *Remote Sens. Environ.* **2012**, *122*, 146–165.
36. Li, A.; Huang, C.; Sun, G.; Shi, H.; Toney, C.; Zhu, Z.; Rollins, M.G.; Goward, S.N.; Masek, J.G. Modeling the height of young forests regenerating from recent disturbances in Mississippi using Landsat and ICESat data. *Remote Sens. Environ.* **2011**, *115*, 1837–1849.

37. Helmer, E.H.; Ruzycki, T.S.; Wunderle, J.M.; Vogesser, S.; Ruefenacht, B.; Kwit, C.; Brandeis, T.J.; Ewert, D.N. Mapping tropical dry forest height, foliage height profiles and disturbance type and age with a time series of cloud-cleared Landsat and ALI image mosaics to characterize avian habitat. *Remote Sens. Environ.* **2010**, *114*, 2457–2473.
38. Hilker, T.; Wulder, M.A.; Coops, N.C. Update of forest inventory data with Lidar and high spatial resolution satellite imagery. *Can. J. Remote Sens.* **2008**, *34*, 5–12.
39. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
40. Olden, J.D.; Lawler, J.J.; Poff, N.L. Machine learning methods without tears: A primer for ecologists. *Q. Rev. Biol.* **2008**, *83*, 171–193.
41. Durbha, S.S.; King, R.L.; Younan, N.H. Support vector machines regression for retrieval of leaf area index from multi-angle imaging spectroradiometer. *Remote Sens. Environ.* **2007**, *107*, 348–361.
42. Hudak, A.T.; Cookston, N.; Evans, J.; Hall, D.; Falkowski. M. Nearest neighbor imputation of species-level, plot scale forest structure attributes from Lidar data. *Remote Sens. Environ.* **2008**, *112*, 2232–2245.
43. DFW. *Forest Cover Assessment, Cambodia (in Khmer)*; Department of Forest and Wildlife: Phnom Penh, Cambodia, 1999.
44. Bruun, T.B.; de Neergaard, A.; Lawrence, D.; Ziegler, A.D. Environmental consequences of the demise in Swidden cultivation in Southeast Asia: Carbon storage and soil quality. *Hum. Ecol.* **2009**, *37*, 375–388.
45. Padoch, C.; Coffey, K.; Mertz, O.; Leisz, S.J.; Fox, J.; Wadley, R.L. The demise of Swidden in Southeast Asia? Local realities and regional ambiguities. *Geogr. Tidsskr.* **2007**, *107*, 29–41.
46. Fox, J.; Vogler, J.B. Land-use and land-cover change in Montane Mainland Southeast Asia. *Environ. Manag.* **2005**, *36*, 394–403.
47. Markham, B.L.; Storey, J.C.; Williams, D.L.; Irons, J.R. Landsat sensor performance: History and current status. *IEEE Trans. Geosci. Remote Sens*. **2004**, *42*, 2691–2694.
48. Axelsson, P. Processing of laser scanner data—Algorithms and applications. *ISPRS J. Photogramm. Remote Sens*. **1999**, *54*, 138–147.
49. Gao, F.; Masek, J.G.; Wolfe, R.E. Automated registration and orthorectification package for Landsat and Landsat-like data processing. *J. Appl. Remote Sens.* **2009**, *3*, 033515. doi:10.1117/1.3104620.
50. Chavez, P.S., Jr. Image-based atmospheric corrections—Revisited and improved. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1025–1036.
51. Canty, M.J.; Nielsen, A.A.; Schmidt, M. Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sens. Environ.* **2004**, *91*, 441–451.
52. Schroeder, T.A.; Cohen, W.B.; Song, C.; Canty, M.J.; Yang, Z. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sens. Environ.* **2006**, *103*, 16–26.
53. Powell, S.L.; Cohen, W.B.; Yang, Z.; Pierce, J.D.; Alberti, M. Quantification of impervious surface in the Snohomish Water Resources Inventory Area of Western Washington from 1972–2006. *Remote Sens. Environ.* **2008**, *112*, 1895–1908.

54. Gillanders, S.N.; Coops, N.C.; Wulder, M.A.; Gergel, S.E.; Nelson, T. Multitemporal remote sensing of landscape dynamics and pattern change: Describing natural and anthropogenic trends. *Prog. Phys. Geog.* **2008**, *35*, 502–528.
55. Crist, E.P. A TM tasselled cap equivalent transformation for reflectance factor data. *Remote Sens. Environ.* **1985**, *17*, 301–306.
56. Cohen, W.B.; Spies, T.A.; Alig, R.A.; Oetter, D.R.; Maiersperger, T.K.; Fiorella, M. Characterizing 23 years (1972–1995) of stand replacement disturbance in Western Oregon forests with Landsat imagery. *Ecosystems* **2002**, *5*, 122–137.
57. Hansen, M.J.; Franklin, S.E.; Woudsma, C.G.; Peterson, M. Caribou habitat mapping and fragmentation analysis using Landsat MSS, TM, and GIS data in the North Columbia Mountains, British Columbia, Canada. *Remote Sens. Environ.* **2001**, *77*, 50–65.
58. Crist, E.P.; Cicone. R.C. A physically based transformation of Thematic Mapper data. The TM tasselled Cap. *IEEE Trans. Geosci. Remote Sens*. **1984**, *22*, 256–263.
59. Price, K.P.; Jakubauskas, M.E. Spectral retrogression and insect damage in lodgepole pine successional forests. *Int. J. Remote Sens.* **1998**, *19*, 1627–1632.
60. White, J.C.; Wulder, M.A.; Gomez, C.; Stenhouse, G. A history of habitat dynamics: Characterizing 35 years of stand replacing disturbance. *Can. J. Remote Sens.* **2011**, *37*, 234–251.
61. R: A Language and Environment for Statistical Computing. Available online: http://www.R-project.org/ (accessed on 12 January 2013).
62. Chen, G.; Hay, J.G.; Castilla, G.; St-Onge, B.; Powers, R. A multiscale geographic object-based image analysis to estimate Lidar measured forest canopy height using Quickbird imagery. *Int. J. Geogr. Inf. Sci.* **2010**, *25*, 877–893.
63. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
64. Breiman, L.; Cutler, A. Random forests—Classification description: Random forests. Available online: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed on 12 January 2013).
65. McInerney, D.; Nieuwenhuis, M. Comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory. *Int. J. Remote Sens.* **2009**, *30*, 4937–4955.
66. Gleason, C.J.; Im, J. Forest biomass estimation from airborne Lidar data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91.
67. Liaw, A.; Wiener, M. Classification and regression by random forest. *R News* **2002**, *23*, 18–22.
68. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *285*, 1–26.
69. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272.
70. Maltamo, M.; Eerikainen, K.; Packalen, P.; Hyyppa, J. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry* **2006**, *79*, 217–229.
71. Frazier, R.J.; Coops, N.C.; Wulder, M.A.; Kennedy, R. Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. *ISPRS J. Photogramm. Remote Sens*. **2014**, *92*, 137–146.

72. Mon, M.S.; Mizoue, N.; Htun, N.Z.; Kajisa, T.; Yoshida, S. Factors affecting deforestation and forest degradation in selectively logged production forest: A case study in Myanmar. *For. Ecol. Manag.* **2012**, *267*, 190–198.
73. Cohen, W.B.; Yang, Z.; Kennedy, R. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation. *Remote Sens. Environ.* **2010**, *114*, 2911–2924.

© 2014 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 license (http://creativecommons.org/licenses/by/4.0/).