

Article

A New Tree Cover Percentage Map in Eurasia at 500 m Resolution Using MODIS Data

Toshiyuki Kobayashi ^{1,*}, Javzandulam Tsend-Ayush ² and Ryutaro Tateishi ³

¹ Geosystem and Biological Sciences Division, Graduate School of Science, Chiba University, 1-33 Yayoi-cho, Inage, Chiba 263-8522, Japan

² Department of Natural Resources and Environmental Management, University of Hawaii at Manoa, 1910 East-West Road, Sherman 101, Honolulu, HI 96822, USA; E-Mail: javzan@hawaii.edu

³ Center for Environmental Remote Sensing, Chiba University, 1-33 Yayoi-cho, Inage, Chiba 263-8522, Japan; E-Mail: tateishi@faculty.chiba-u.jp

* Author to whom correspondence should be addressed; E-Mail: kobayashit@graduate.chiba-u.jp; Tel.: +81-43-290-2964.

Received: 13 September 2013; in revised form: 5 November 2013 / Accepted: 12 December 2013 / Published: 20 December 2013

Abstract: Global tree cover percentage is an important parameter used to understand the global environment. However, the available global percent tree cover products are few, and efforts to validate these maps have been limited. Therefore, producing a new broad-scale percent tree cover dataset is valuable. Our study was undertaken to map tree cover percentage, on a global scale, with better accuracy than previous studies. Using a modified supervised regression tree algorithm from Moderate Resolution Imaging Spectroradiometer (MODIS) data of 2008, the tree cover percentage was estimated at 500 m resolution in Eurasia. Training data were created by simulation using reference data interpreted from Google Earth. We collected approximately 716 high-resolution images from Google Earth. The regression tree model was modified to fit those images for improved accuracy. Our estimation result was validated using 307 points. The root mean square error (RMSE) between estimated and observed tree cover was 11.2%, and the weighted RMSE between them, in which five tree cover strata (0%–20%, 21%–40%, 41%–60%, 61%–80%, and 81%–100%) were weighted equally, was 14.2%. The result was compared to existing global percent-scale tree cover datasets. We found that existing datasets had some pixels with estimation error of more than 50% and each map had different characteristics. Our map could be an alternative dataset and other existing datasets could be modified using our resultant map.

Keywords: land cover; classification; estimation; high resolution; modeling

1. Introduction

Forests, by playing an important role in regulating the climate and water resources, and by providing habitats for many species, are of great importance for life on Earth. Nevertheless, they have recently been converted to unsustainable forms of land use due to urbanization and deforestation by expanding human populations. About 13 million hectares of forest were lost annually worldwide in the last decade, though the rate of loss decreased compared with about 16.1 million hectares per year during the 1990s. At a continent level, the forest area significantly decreased in South America and Africa between 2000 and 2010. The rate of deforestation was decreased in Asia, Africa, and America compared with the 1990s, especially in Brazil and Indonesia, while the loss of forest increased in Australia since 2000 [1,2]. Trees are important structural members of forests that remove carbon dioxide from the atmosphere as they grow, and emit it when they decay or burn. Some attempts to estimate the global tree cover percentage have been made to date [3–8]. Some reports describe studies of tree cover percentage on a regional scale [9–11]. Continuous field maps present some advantages to measuring the change in spatially complex land cover compared with traditional discrete classifications. Estimates of proportional tree coverage can distinguish dense forest from sparse forest, and forests with small patches of cleared areas [4,12,13]. Maps showing the tree cover percentage are useful in many fields [14]. They were used as one independent variable to model the global forest canopy height for mapping ecosystem vertical structures [15]. They are also useful for making environmental policies and elucidating the present environmental situation for education. However, the available global percent tree cover products are few, and efforts to validate these maps have been limited to some regions or countries due to the difficulty of obtaining reference data or field data [12,14,16,17]. There are studies showing that existing maps' accuracy of estimating global tree cover is not high, particularly in sparsely forested areas of the circumpolar taiga–tundra transition zone and special areas and ecosystems [14,16,17]. For global land cover maps, some datasets have been released in the last decade [18,19], and users can choose the most suitable map for their application among them. There are also some efforts to evaluate and compare the quality of these maps [20,21]. Producing a new percent tree cover dataset on a global or continental scale is valuable by providing users with more choices for their application, as each of datasets generally has different characteristics. It can also enable the comparison analysis among maps. Users of these maps can identify areas of potential low or high map uncertainty by it. The users also can know the regions where the estimate of tree cover percentage is difficult and the land cover is complex. Each map can be modified by knowing the regions.

The objective of our study is to map tree cover percentage on a global scale with better accuracy compared to existing datasets. In this paper, we present a new continuous field map of tree cover in Eurasia as the first step of mapping on a global scale. We tried to improve the estimation of tree cover by using a great deal of training data, which cover many kinds of croplands, trees and other vegetation in different ecoregions. We used high-resolution imagery in Google Earth as reference data as reference data consisting of various land cover types can be collected by it, and the estimation model

was fitted with the interpreted result of Google Earth images. Training data were created by simulation. Data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments onboard Terra and Aqua satellites were used for estimation. We also compared our tree cover estimation results with two available global-scale datasets: (1) Vegetation Continuous Fields MOD44B, 2008 Percent Tree Cover Collection 5 product (MOD44B), published by University of Maryland [22]; and (2) Global Map—Percent Tree Cover of Global Mapping project [5].

Definitions of “tree” and “tree cover percentage (or tree crown cover)” differ among reports of the literature [6,10,23]. For MOD44B, tree cover percentage is the amount of light obstructed by tree canopies equal to or greater than 5 m in height per 500-m MODIS pixel (percent canopy cover) [6]. In botany, a tree is defined according to its characteristics: (a) whether it is perennial or not; (b) whether it has a self-supporting stem or not; (c) whether the thickness of secondary tissues is increasing or not; (d) whether it has a lignified stem or not; (e) whether the girth of its stem increases or not; (f) its height, *etc.* [24–27]. However, it is difficult to distinguish trees from other vegetation using satellite remote sensing techniques. In this report, tree cover percentage refers to the percentage of the ground surface area covered by a vertical projection of the foliage and branches of trees when the leaves are at full growth. Small openings inside each crown are included (percent crown cover). A reasonable relation was found between canopy cover and crown cover by which 80% canopy cover corresponded to 100% crown cover, although this relation differed by tree type [6]. A “tree” is a woody perennial with a single self-supporting main stem, with minimum height of *ca.* 3–6 m. Trees for agricultural production or in gardens, and trees on plantations are included. Bamboos and palms are also included as trees. This definition is conceptual as it is difficult to ascertain these characteristics from satellite images.

2. Study Area and Data

2.1. Study Area

Percent tree cover was estimated for Eurasia, extending from 11°W to 180°E and 10°S to 80°N. Eurasia includes various climate zones: tropical, temperate, cold, and arid climates [28].

2.2. Satellite Data

Global MODIS 2008 data processed by CEReS Chiba University [29] were used for estimating the tree cover percentage in 2008. These data include three bands in the visible, one band in near-infrared, and three bands in shortwave infrared. These data were produced from the TERRA/AQUA Nadir BRDF-Adjusted Reflectance 16-day L3 Global 500 m product (MCD43A4 NBAR) [30–32]. MCD43A4 NBAR product was provided in surface reflectance data corrected to the common nadir-view geometry at the solar angle at local solar noon time of the start of the observation period using a bi-directional reflectance distribution function (BRDF) model. Data were aggregated temporally to 16-day data, with a pixel size of about 500 m. Original data (MCD43A4 NBAR) were mosaicked and reprojected to geographic coordinates using nearest-neighbor resampling at CEReS, Chiba University. Furthermore, missing data in some pixels were linearly interpolated using data obtained in 2007 and 2009 to make them cloud-free.

2.3. Reference Data

We used high resolution imagery available in Google Earth as reference data for producing training data, modifying models, and validating results. Only images in Google Earth were used where individual tree crowns were visually interpretable. In recent years, high-resolution imagery in Google Earth has been valuable and used for reference data in the field of remote sensing [17,33]. It can be used for validation data of continuous field products [33]. The salient advantage of using Google Earth is that high-resolution images of inaccessible places are obtainable at no cost. We can also obtain several images in different years at specific places. Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Scan Line Corrector Off (SLC-Off) data, which are available from the US Geological Survey, were also used for reference data.

2.4. Percent Tree Cover Datasets Used for the Comparison

We used two available global-scale datasets for comparison to our tree cover estimation result: Vegetation Continuous Fields MOD44B, 2008 Percent Tree Cover Collection 5 product and Global Map—Percent Tree Cover of Global Mapping project. Vegetation Continuous Fields MOD44B was derived through an automated procedure using supervised regression tree algorithm, and produced in annual time steps since 2000. The pixel size of the dataset was 7.5 s. The input data used for the product were MODIS Bands 1–7 and Land Surface Temperature data. Training data were created by aggregating discretely classified Landsat images to the MODIS cells. They were refined using fine resolution data [6,13,22]. In this research, tree-cover percentage in MOD44B was divided by 0.8, as the definition of tree cover percentage in this product differed from our map. Global Map—Percent Tree Cover of Global Mapping project (hereafter PTC of GM project) was produced using regression tree method from MODIS 32-day composite data of 2003. The pixel size was 30 s. Training data were collected using 68 high-resolution QuickBird images and 153 images in Google Earth [5,8]. The definition of tree cover percentage was the same as our map.

3. Methods

3.1. Estimate of Tree Cover Percentage

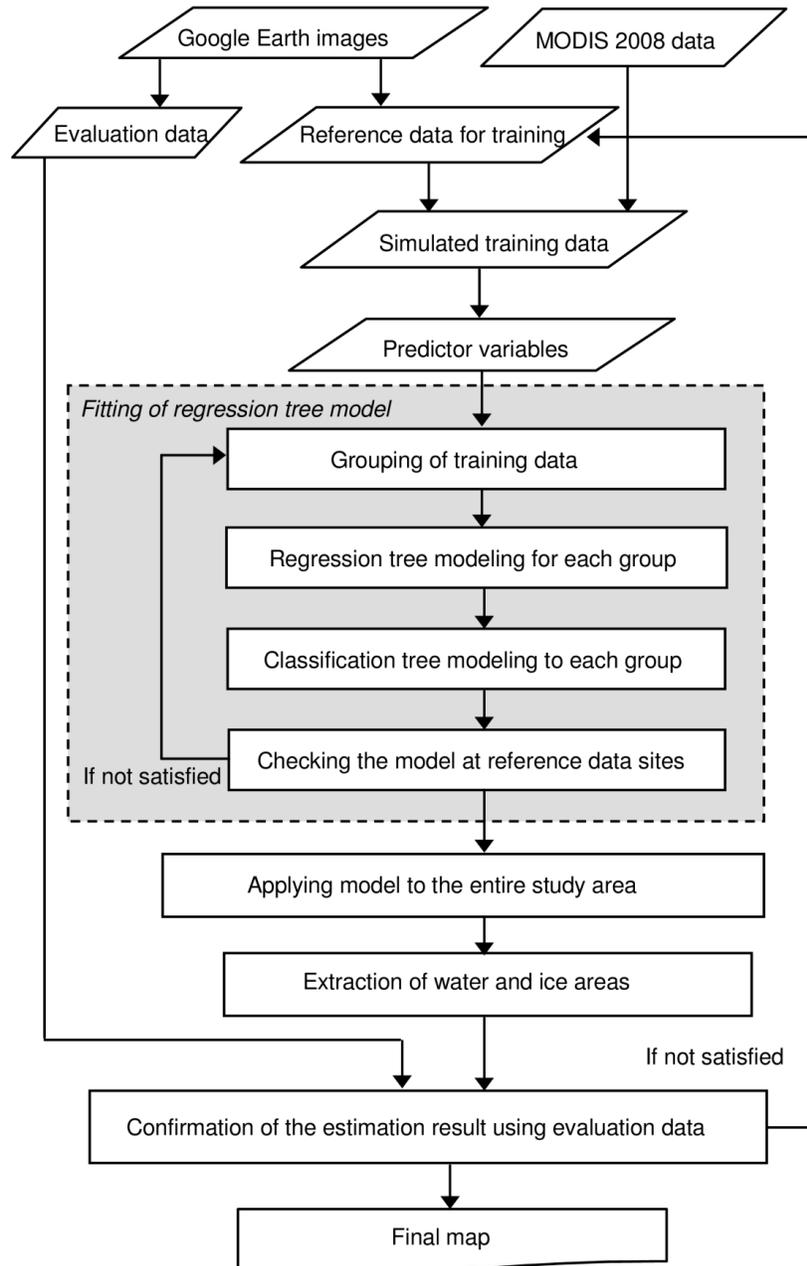
In this research, a regression tree algorithm was used for estimating tree cover percentage. Some methods are useful for estimating tree cover percentage: spectral mixture, neural networks, and multiple linear regression analysis. At a local scale, Berberoglu *et al.* [9] compared multiple linear regression, a linear mixture model, an artificial neural network and a regression tree model using Envisat Medium Resolution Imaging Spectrometer imagery. They showed that the regression tree method was the most accurate though the distribution of the error varied with altitude and tree cover percentage. Joshi *et al.* [34] also compared an artificial neural network, a multiple linear regression, a forest canopy density mapper and a maximum likelihood classification method using a Landsat ETM+ image, and showed that the artificial neural network outperformed the other methods and the multiple linear regression performed worse. We followed the existing global percent tree cover mapping for selecting the method. The regression tree model is built by partitioning each node into child nodes

starting with a root node. A target variable, tree cover percentage in this case, was calculated at terminal nodes. This algorithm can account for nonlinear relations between the predictor and target variables [13]. For our research, all training data for creating the estimation model were divided into groups, and a classification tree model for classifying all data into groups was created. Tree-cover percentage of each group was estimated using the regression tree model fit to training data of each group. To create an unbiased model, we collected an unbiased sample of the reference data across the study area in advance using Google Earth, and the tree model was modified to fit the reference data. Hereafter, we refer to these reference data collected for modifying the tree model as evaluation data. For mapping percent tree cover, a great deal of continuous tree-cover training data are essential. However, it is difficult to collect training data. In this research, training data for creating the estimation model were created in a simulation by combining reference data sets. Simulated training data created from the reference data of water or ice and other land covers were not used for this study. The percentage of tree cover for water and ice areas was set as 0% after they were independently extracted using the classification tree method. A flowchart depicting these steps is provided in Figure 1.

3.1.1. Creation of Evaluation Data for Modifying the Model

The evaluation data were collected in advance to create an unbiased model, and used for modifying the estimation model of tree cover. They were collected across the study area using high-resolution images in Google Earth for reference. To collect an unbiased sample of evaluation data from throughout the study area, we used PTC of GM project and Global Map—Global Land Cover (GLCNMO) in 2003 [5]. GLCNMO had 20 classes, which were five types of forest, tree open, shrub, three types of herbaceous and sparse vegetation, three types of agricultural area, two types of bare area, mangrove, wetland, urban, snow/ice, and water bodies. We divided each land cover class (except the water class) of GLCNMO into five strata according to PTC of GM project (0%–20%, 21%–40%, 41%–60%, 61%–80%, and 81%–100% tree cover), and 6 sites were randomly sampled from each stratum of each land cover class. The size of each site was 2.5×2.5 min. The sampled sites were overlaid on Google Earth, and they were used as evaluation data. In this step, for some cases, the sampled sites were unsuitable as evaluation sites. In such cases, random sampling was repeated for each stratum until 6 sites were sampled. The following situations are examples of unsuitable cases.

- It was difficult to estimate the tree cover percentage as the available images in Google Earth over the sampled site were not of high resolution.
- It was difficult to distinguish a tree from a shrub.
- Geo-location error of MODIS pixel caused poor estimate of tree cover percentage.
- Estimating the tree cover percentage at the peak of the growing season was difficult as the displayed image in Google Earth had been acquired in winter or during the dry season.
- Land cover change during the observation time of images in Google Earth and the year 2008 may result in wrong estimate of tree-cover percentage.

Figure 1. Flowchart showing percent tree cover estimation in this study.

The judgment of tree characteristics was based mainly on color and texture. The presence of shadows was used to distinguish a tree from a shrub. As most of the sites with greater than 60% tree cover were classified as forest in GLCNMO, another 50–60 sites for each stratum were also randomly sampled, leaving the land cover class out of consideration to obtain more than 100 evaluation sites for each stratum of the observed tree cover percentage. For the evaluation data, totally, 716 sites were collected (Figure 2). Confirmation that the tree-cover percentage of the acquired high-resolution images in Google Earth is almost identical to the percentage at the peak of the growing season in 2008 was made by comparison with the temporal profile of Normalized Difference Vegetation Index (NDVI) data calculated from MODIS data in 2003 and 2008. All sampled sites were also examined by

comparison with Landsat images to ascertain the possibility of the change of the tree cover percentage. A flowchart of these steps is presented in Figure 3.

Figure 2. Distribution map of evaluation data. Evaluation data were used to create a tree-cover estimation model.

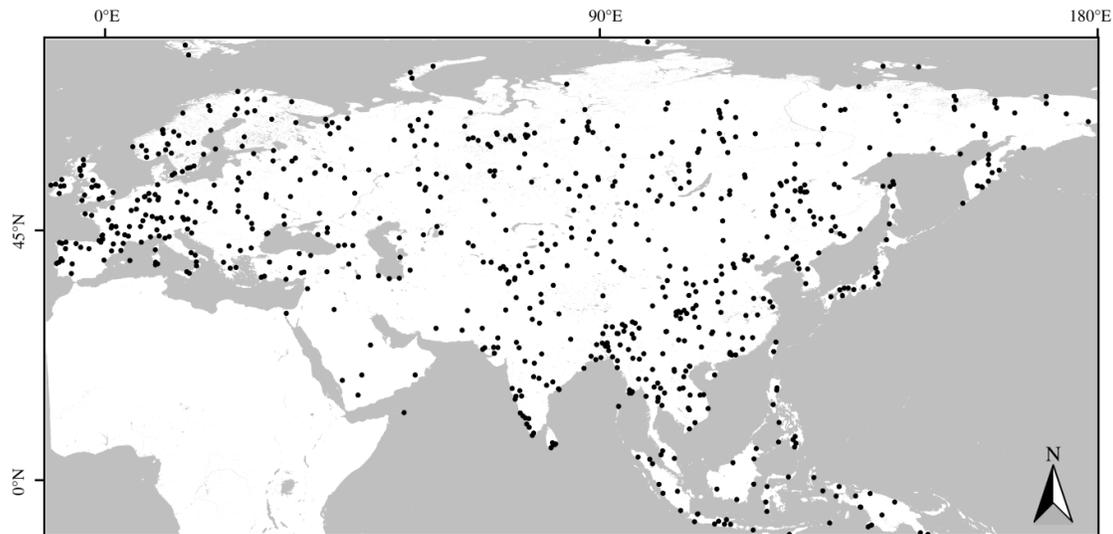
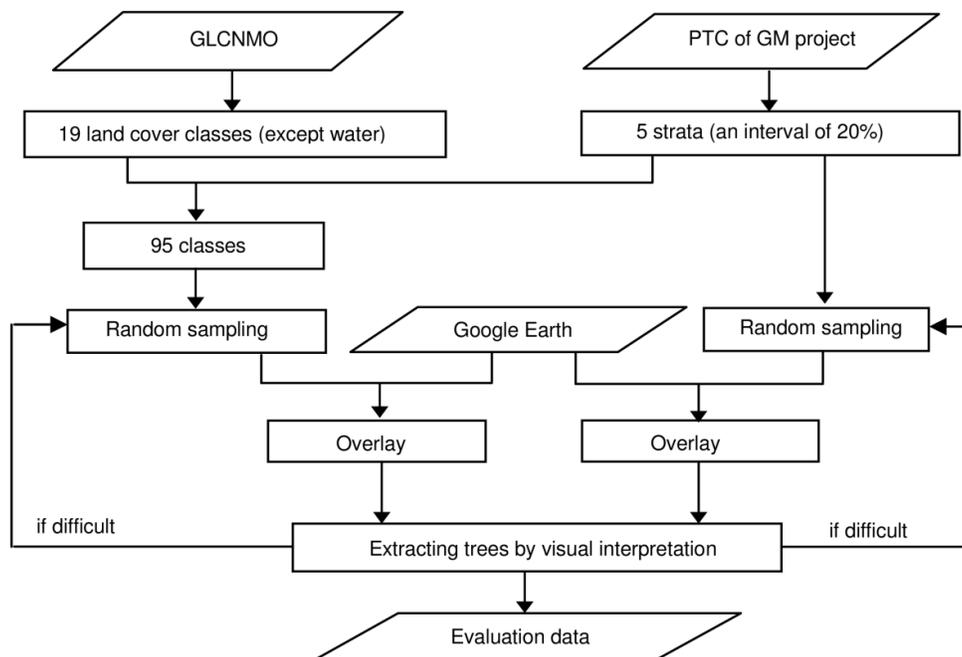


Figure 3. Flowchart showing production of evaluation data.



3.1.2. Creation of Training Data from Images in Google Earth and MODIS Band Values

In our research, the reference data for training were collected only from areas where the tree cover percentage was greater than 85% or less than 20%, as judged from images in Google Earth, and training data of continuous tree cover percentage from 0% to 100% were created by simulation

using reference data. This was because it was sometimes too difficult to distinguish trees in images in Google Earth. In most cases, the error of visually interpreted tree cover percentage from the images displayed in Google Earth was less than 10% for areas with greater than 85% or less than 20% tree cover. The reference data of less than 20% tree cover were obtained from various land cover types, and the reference data of greater than 85% tree cover were obtained from many kinds of forest. They were collected from areas where the influence of geo-location error was small because the Global MODIS 2008 data had geo-location error of less than half pixel [29]. The number of pixels within one reference site varied between 4 and 100 pixels. 1,040 sites in all were collected around the whole study area (Table 1, Figure 4). Each site had 7 bands' reflectance values at 23 periods of MODIS data in 2008. Each value of a reference site was obtained by averaging values of all pixels of the reference site. Table 2 presents the period of 16-day composite data. Training data of 0%–100% tree cover percentage were created using MODIS data by simulation. These data were used mainly for modeling classification trees. The following linear spectral mixture model was used for this simulation:

$$S_{ij} = aV_{F_k,ij} + bV_{V_l,ij} + cV_{N_m,ij}, (a + b + c = 1) \quad (1)$$

where S_{ij} = simulated reflectance of MODIS band i at period j

$V_{F_k,ij}$ = reflectance of reference site k (land cover type: forest) of MODIS band i at period j

$V_{V_l,ij}$ = reflectance of reference site l (land cover type: vegetation except forest) of MODIS band i at period j

$V_{N_m,ij}$ = reflectance of reference site m (land cover type: non-vegetation) of MODIS band i at period j

a, b, c = area ratio of reference site (=0, 0.2, 0.4, 0.6, 0.8 or 1)

$i = 1-7$

$j = 1-23$.

S_{ij} was calculated at 20% intervals of the area ratio of reference sites. The 1,040 reference data collected using Google Earth were grouped into three land cover types, which were forest, vegetation except forest and non-vegetation. Then the simulated reflectance was calculated for the combinations of these groups. In this study, each of three groups was divided into about 20 subgroups beforehand from the perspective of the temporal profile of NDVI data, and the temporal profile of NDVI data of subgroups was also considered for the combination of the reference data. The combination of two reference data that are mutually distant, for example the combination of a grassland in Kazakhstan and a forest in Japan, was not considered. In all, more than 5,000,000 training data were created.

Table 1. Number of collected reference sites for training and their dominant land cover types. Vegetation includes grassland, agricultural areas, shrub, and their mosaic without forests. Non-vegetation includes water, urban, and bare areas.

Dominant Land Cover Types	Number of Collected Reference Sites
Forest	320 sites
Vegetation except forest	575 sites
Non-vegetation	145 sites

Figure 4. Distribution map of reference data for training. Vegetated areas include grasslands, agricultural areas, shrub, and their mosaic without forests. Non-vegetated areas include water, urban, and bare areas.

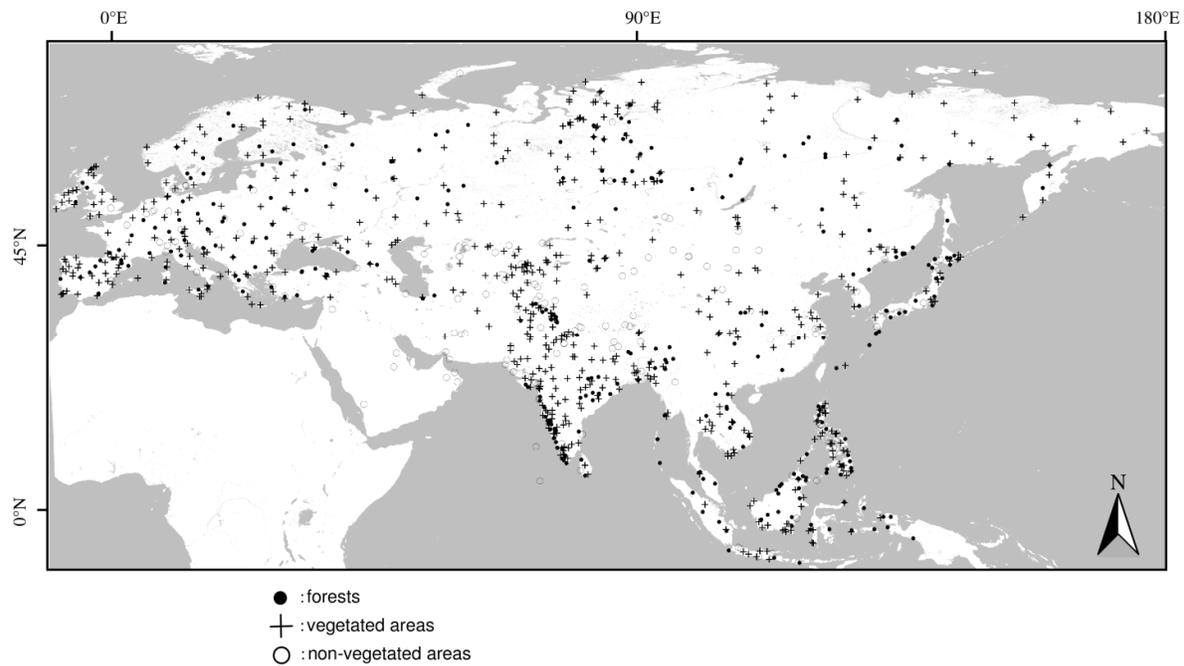


Table 2. Start and end dates of 23 periods of 16-day composite Moderate Resolution Imaging Spectroradiometer (MODIS) data in 2008.

Period No.	Start Date	End Date
1	2008/01/01	2008/01/16
2	2008/01/17	2008/02/01
3	2008/02/02	2008/02/17
4	2008/02/18	2008/03/04
5	2008/03/05	2008/03/20
6	2008/03/21	2008/04/05
7	2008/04/06	2008/04/21
8	2008/04/22	2008/05/07
9	2008/05/08	2008/05/23
10	2008/05/24	2008/06/08
11	2008/06/09	2008/06/24
12	2008/06/25	2008/07/10
13	2008/07/11	2008/07/26
14	2008/07/27	2008/08/11
15	2008/08/12	2008/08/27
16	2008/08/28	2008/09/12
17	2008/09/13	2008/09/28
18	2008/09/29	2008/10/14
19	2008/10/15	2008/10/30
20	2008/10/31	2008/11/15
21	2008/11/16	2008/12/01
22	2008/12/02	2008/12/17
23	2008/12/18	2009/01/02

3.1.3. Creation of Predictor Variables from MODIS Band Reflectance

Original MODIS band values of training data were converted into predictor variables for estimating the tree cover percentage. Selection of predictor variables is important to estimate the tree cover percentage using the regression tree method. Previous studies used annual variables derived from some temporal profiles. Some examples are the maximum value of NDVI, average Band 1–7 reflectance at the three or seven highest NDVI periods, the minimum Band 1 reflectance, the maximum Band 2 reflectance, average reflectance in the four darkest reflectance periods and amplitude for the minimum and maximum reflectance [6,7,13]. In PTC of GM project, annual variables derived from the 32-day composited data of MODIS Bands 1–7, NDVI, NDSI and SI were used as predictor variables [8]. In MOD44B, the 16-day composited data of MODIS Bands 1–7, and brightness temperature from MODIS Bands 20, 31, 32 were used [6,22]. Hansen *et al.* [35] compared MODIS inputs to mapping land cover, and the difference of standard errors of the tree cover percentage using annual variables and composited variables was less than 1% at the global and the continent scales. In this study, NDVI, Normalized Difference Soil Index (NDSI), Shadow Index (SI) [36], and reflectance for Bands 1–7 of individual 16-day periods were used. We created several models using different combinations of predictor variables, and selected the model with the best fit to reference sites. Table 3 presents the predictor variables used for modeling the tree cover percentage estimation in this study. These predictor variables were calculated for MODIS data in the entire study area and simulated training data. The NDVI, NDSI, and SI [36] were calculated from the following equations.

$$NDVI = \frac{b_2 - b_1}{b_2 + b_1} \quad (2)$$

$$NDSI = \frac{b_6 - b_2}{b_6 + b_2} \quad (3)$$

$$SI = [(1 - b_1) \times (1 - b_3) \times (1 - b_4)]^{1/3} \quad (4)$$

In those equations, b_i = reflectance of MODIS band i . Annual mean values of MODIS Bands 1–7 were created by averaging 23 periods of composite data of MODIS Bands 1–7. Mean values at the three highest periods of each band reflectance were created by averaging three brightest reflectance values from among 23 periods of composite data of MODIS Bands 1–7. Minimum values of each band reflectance meant the darkest value of 23 periods of composite MODIS data. Mean values at the three greenest periods, based on the temporal profile of NDVI, were created by averaging three periods of data of MODIS Bands 1–7, NDVI, NDSI, and SI. Three periods were derived based on ranked 23 periods of NDVI values. Individual values based on the temporal profile of NDVI meant 23 values of each of MODIS Bands 1–7, NDVI, NDSI, and SI, which were put in order of ranked NDVI values.

3.1.4. Grouping of Training Data and Regression Tree Modeling

Training data were divided into groups from the perspective of geographical location and the temporal profile of NDVI data. For example, training data created from the reference data in Europe and East Asia were divided into different groups. Then, small regression tree models, in which the number of multiple linear regression equations at terminal nodes was less than four, were fitted for

each group using Cubist, which is commercially available software for regression tree models by RuleQuest Research. This software generates models consisting of rules associated with linear expressions based on the sample's average error, standard deviation, and correlation coefficient on newly created nodes [37,38]. This software has been used for regression tree modeling in the field of remote sensing [8,9,39]. Nine annual predictor variables at the peak of the growing season and average reflectance in the three brightest reflectance periods for Bands 1–7 were mainly used for predictor variables. The growing season peak was inferred from the three periods of data with the highest NDVI values. These annual variables can use the data for the snow-free period and can reduce the number of input data. The first regression tree model was fitted using all training data for each group. The second regression tree model was fitted using training data whose original reference data had more than 20% error in the first regression tree model. This procedure was continued until the error of all reference data became less than 20% using any regression tree model.

Table 3. Predictor variables derived for modeling to estimate tree cover percentage.

Types	Variables
Individual values for all the 16-day periods	Nadir BRDF-Adjusted Reflectance (Bands 1–7)
	NDVI
	NDSI
	SI
Mean values at the three greenest periods based on the temporal profile of NDVI	Nadir BRDF-Adjusted Reflectance (Bands 1–7)
	NDVI
	NDSI
	SI
Individual values based on the temporal profile of NDVI	Nadir BRDF-Adjusted Reflectance (Bands 1–7)
	NDVI
	NDSI
	SI
Annual mean values	Nadir BRDF-Adjusted Reflectance (Bands 1–7)
Mean values at the three highest periods of each band reflectance	Nadir BRDF-Adjusted Reflectance (Bands 1–7)
Minimum values of each band reflectance	Nadir BRDF-Adjusted Reflectance (Bands 1, 3)

3.1.5. Creation of Classification-Tree Model and Modification

A classification tree model for classifying the data into groups was fitted from all training data. This modeling of the classification into groups was accomplished using a software tool for a classification tree (See5; RuleQuest Research). It automatically fits decision trees by recursive partitioning of a subset of training data to maximize the Gain ratio at each node [40–42]. The Gain ratio is based on the entropy or amount of information, which is expressed as:

$$Information(D) = - \sum_{j=1}^m p(D, j) \times \log(p(D, j)) \quad (5)$$

where $p(D, j)$ denotes the proportion of a set of cases in D that belong to the j th class and m signifies the number of classes. The Gain ratio of cases in D by a test T is expressed as:

$$\text{Gainratio}(D,T) = \frac{\text{Gain}(D,T)}{\text{Split}(D,T)} \quad (6)$$

$$\text{Gain}(D,T) = \text{Information}(D) - \sum_{i=1}^n \frac{|D_i|}{|D|} \times \text{Information}(D_i) \quad (7)$$

$$\text{Split}(D,T) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \times \log \left(\frac{|D_i|}{|D|} \right) \quad (8)$$

where D_i is the i th subset and n is the total number of subsets by a test T .

Tree models were fitted using following combination of predictor variables.

- (A) Mean NDVI, NDSI, and SI values and mean reflectance for Bands 1–7 at the peak of the growing season.
- (B) NDVI values at odd periods (periods 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23).
- (C) Mean NDVI, NDSI, and SI values and mean reflectance for Bands 1–7 at the peak of the growing season, and NDVI and SI values at odd periods.
- (D) Mean NDVI, NDSI, and SI values and mean reflectance for Bands 1–7 at the peak of the growing season, and NDVI, SI and Bands 2, 5, and 7 at odd periods.

The model among them with the best fit to reference sites was adopted as the final model. When the estimation result was not good for some reference sites, fitting of a regression tree model and modification of grouping training data was conducted again. Water and ice areas were classified independently using See5 software as the percent tree cover estimated by Cubist was not good when the simulated training data made of these two classes were used. These two areas were set to 0% tree cover.

3.1.6. Application of the Model to the Whole Study Area and Modification of the Model According to Evaluation Data

The created tree model was applied to all pixels in the entire study area. The estimation result was compared with evaluation data by visual interpretation of tree cover. This comparison was conducted, taking account of the range of geo-location and image interpretation errors. We modified the model further to agree with the evaluation data until the difference between the estimated and observed tree cover percentage became less than 20%. We collected more reference data for training when it was necessary for modification to adjust to the evaluation data. In case of no suitable pixels were available, reference data of tree cover percentages of 20%–85% were also collected. The predictor variables used to calculate multiple linear regression equations were also changed for some groups.

3.2. Validation

Validation data were created to assess the accuracy of the final estimation result. They were collected across the study area using two kinds of random sampling approaches. Sea areas were excluded. In the first approach, 477 pixels were randomly sampled in our resultant map. In the second approach, 50 other pixels were collected using stratified random sampling of the pixel in our resultant map. In this approach, all pixels in our resultant map were divided into five strata by the estimated tree cover percentage (0%–20%, 21%–40%, 41%–60%, 61%–80%, and 81%–100%), and ten pixels for each stratum were randomly sampled. This sampling approach was conducted because it was difficult

to collect enough validation data of greater than 20% tree cover by random sampling of the first approach. Then, 3×3 pixel areas within our resultant map centered on each sampled pixel were also sampled because a pixel-level validation is not accurate when the geo-location error of less than half pixel of our resultant map caused the error of tree cover percentage. The sampled areas were overlaid on Google Earth and examined using the same method as that used for the creation of evaluation data. For the collection of validation data by the second sampling approach, random sampling was repeated until ten areas enough for using as validation data were collected for each stratum as it was sometimes difficult to interpret the images displayed in Google Earth. In total, 307 areas were collected for validation from 477 and 123 areas sampled by the first and second sampling approaches, respectively. The tree cover percentage of sampled areas was estimated using on-screen digitization, which was based on visual interpretation of tree cover compared with images displayed in Google Earth for reference. Observers interpreted high-resolution images displayed in Google Earth, corresponding to each sampled area, and calculated the percentage of trees of the area. We also estimated the error range of each area with visual interpretation. Using collected validation data, RMSE was calculated in this assessment. RMSE was calculated as:

$$RMSE = \left[\left(\sum_{i=1}^n (\hat{y}_i - y_i)^2 \right) / n \right]^{1/2} \quad (9)$$

where y_i is the observed value of the tree cover percentage at pixel i , \hat{y}_i denotes the estimated value and n signifies the number of observations. We also calculated a weighted RMSE, in which interpreted tree cover strata of 0%–20%, 21%–40%, 41%–60%, 61%–80%, and 81%–100% were weighted equally, because tree cover percentage of most of the collected validation data was less than 10%. The weighted RMSE was calculated as:

$$WRMSE = \left[\left(\sum_{j=1}^M \sum_{i=1}^{N_j} (\hat{y}_{j,i} - y_{j,i})^2 / N_j \right) / M \right]^{1/2} \quad (10)$$

where $y_{j,i}$ is the observed value of the tree cover percentage at pixel i for a tree cover stratum j , $\hat{y}_{j,i}$ denotes the estimated value, N_j signifies the number of observations for a tree cover stratum j and M signifies the number of tree cover strata. M was 5 in this study, because we divided the observed tree cover percentage into five strata.

This assessment was made at the pixel-level and for 3×3 windows of pixels within our resultant map to minimize misregistration errors.

3.3. Comparison with Other Datasets

We compared our results with those obtained in two previous studies: (1) MOD44B in 2008; and (2) PTC of GM project in 2003, although the years and the resolution of their maps are slightly different. First, we assessed the results of PTC of GM project in 2003 and MOD44B in 2008 using the new validation data, which were collected from the sites including the center of the pixels used for the validation of our result. Second, we compared the difference of three maps. As the resolution and the position of pixel differed for three maps, each pixel of MOD44B and our result was resampled into the same position and size of PTC of GM project in this assessment. Third, we conducted the statistical comparison of tree cover percentage among maps. The total number of pixels, according to the estimated tree cover percentage, were examined for this comparison.

4. Results

4.1. The Result of Our New Map

The map showing the tree cover percentage of Eurasia in 2008 is shown in Figure 5. Randomly sampled pixels used for the validation are also shown. The number of pixels used for validation was smaller for the observed tree cover strata of 41%–60% and 61%–80% than for the observed tree cover strata of a 0%–20%. The result of the accuracy assessment is portrayed in Figure 6. For validation data, we estimated the maximum and the minimum tree cover percentage because interpreting the image on Google Earth was difficult and the geo-location error of less than half pixel caused the error of tree cover percentage. Error bars represent the range of these errors. For some pixels, tree cover was overestimated. These pixels were in agricultural land or in grassland (Figure 7). One pixel showed a difference of more than 50% between our result and observed data. This pixel was in grassland and it was surrounded by pixels in forest. Table 4 reveals that the RMSE and the WRMSE for all validation data was 11.2% (mean absolute error = 6.3%) and 14.2%, respectively. The validation for each 3×3 window of pixels centered on each randomly sampled pixel showed the RMSE and the WRMSE of the estimated tree cover percentage was 8.6% (mean absolute error = 5.1%) and 12.3%, respectively. For this window size, the overestimation which showed a difference of more than 50% between our result and observed data was alleviated.

Figure 5. Percent tree cover estimation for Eurasia in 2008. Distribution of pixels, which were randomly sampled for the validation, are also shown.

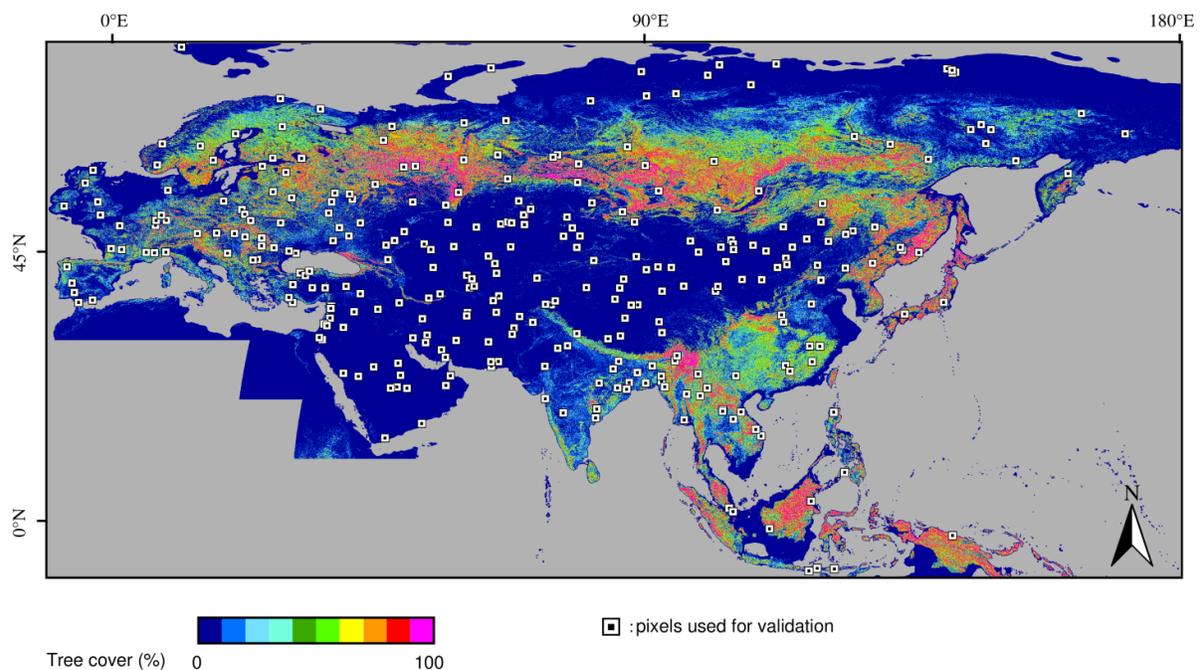


Figure 6. Scatterplot of the observed *versus* estimated tree cover percentage for validation data. (a) is the single-pixel level validation; and (b) is validation using a 3 × 3 pixel window. Validation data were randomly sampled. The horizontal error bars show the range of minimum and maximum visually interpreted tree cover percentage of images in Google Earth. The diagonal shows a linear 1:1 relation; dotted lines show the range of 20% error.

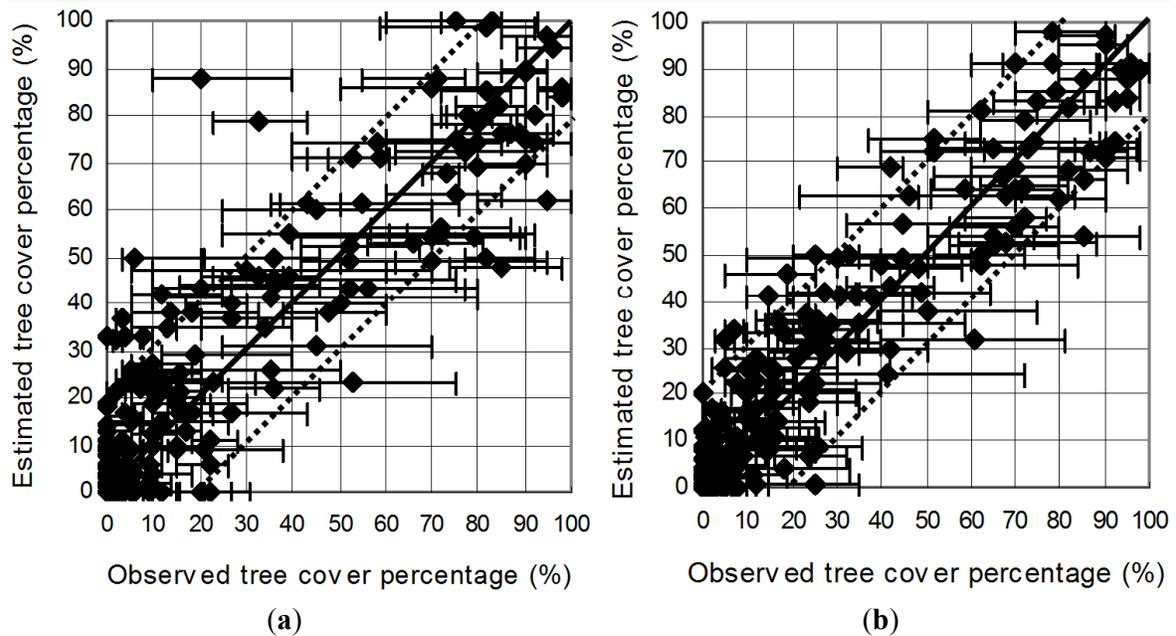


Figure 7. Percentage tree cover at agricultural fields in a part of China (a) and grassland in a part of Russia (b). The tree cover percentage was overestimated in these areas.

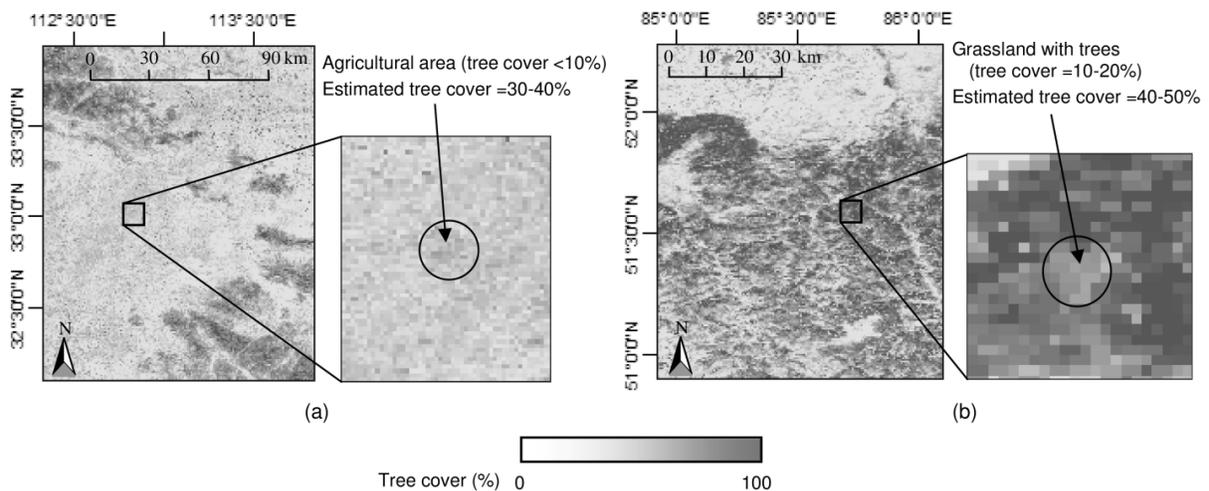


Table 4. Root Mean Square Error (RMSE) and Weighted Root Mean Square Error (WRMSE) of estimated tree cover percentage. Validation data sampled using two kinds of random sampling approaches were used for this assessment.

Tree Cover Strata (Observed)	Single-Pixel Level		3 × 3 Pixel Window	
	RMSE	Number of Sites	RMSE	Number of Sites
0–20	9.5%	231	6.1%	223
21–40	15.7%	21	11.8%	29
41–60	14.2%	15	14.6%	13
61–80	14.5%	17	13.1%	24
81–100	16.2%	23	13.9%	18
0–100	11.2%	307	8.6%	307
0–100 (WRMSE)	14.2%	307	12.3%	307

4.2. Comparison with Other Datasets

The result of the accuracy assessment for other two datasets using newly collected validation data is presented in Figure 8 and Table 5. RMSE was calculated for five strata by the observed tree cover percentage (0%–20%, 21%–40%, 41%–60%, 61%–80% and 81%–100%) of each map. The RMSEs of PTC of GM project and MOD44B were, respectively, 18.6% and 14.0%. The WRMSEs of PTC of GM project and MOD44B were, respectively, 27.5% and 20.4%. The estimation result in our study and in MOD44B was much better than that of PTC of GM project. The RMSE and the WRMSE of our result were, respectively, about 2.8% and 6.2% better than that of MOD44B. Table 6 portrays the results of pixel-based comparison among three maps. Our result was more coincident with MOD44B than with PTC of GM project. In fact, 90.6% of pixels showed a difference of less than 20% between our result and MOD44B. Moreover, 0.3% of pixels showed a difference of more than 50% between our result and MOD44B, whereas about 75% of pixels had a difference of less than 20% and about 5% of pixels had a difference of more than 50% between PTC of GM project and our result or MOD44B. For tree-cover strata, the number of pixels of greater than 20% tree cover showed much more difference between maps than the pixels of less than 20%. Figure 9 illustrates the difference of tree-cover estimation between our result and other two maps: 25 pixels were randomly sampled from pixels, where the difference between two maps was greater than 50%. When compared with PTC of GM project, the difference of tree cover estimation was large at Northern Eurasia, Southeast Asia, India, and Southern China. At the sampled pixels, our estimation result was better in 20 pixels, and the judgment was impossible or both were not good in 5 pixels. PTC of GM project overestimated tree cover in herbaceous areas with trees or water bodies and underestimated it in some forests. When compared with MOD44B, the difference of tree cover estimation was large at Northern Eurasia, Ireland, UK, Eastern India, and Indonesia. At the sampled pixels, our estimation result was better in 13 pixels and worse in 7 pixels. The judgment was difficult in 5 pixels. MOD44B overestimated tree cover in grasslands and underestimated deciduous forest and tropical rainforest. Our result overestimated tree cover in herbaceous areas with water bodies, and underestimated it at deciduous forest and there were isolated pixels with no trees in forest. Figure 10 presents the frequency of pixels for the estimated tree cover percentage in the entire study area. Some differences were apparent among three maps. Our result had large numbers of pixels of a 100% tree cover percentage, although the other

two maps had fewer pixels in the maximum tree cover percentage. MOD44B had only 20% of pixels with a 0% tree cover and much more pixels with a 1%–10% tree cover than other maps. In PTC of GM project, about 50% of pixels were a 0% tree cover, and the numbers of pixels of some specific values of tree-cover percentage were extremely large. For example, the number of pixels was too large at 6% and 46% tree cover. MOD44B and our result displayed a tendency for the number of pixels to decrease as tree cover increases, whereas PTC of GM project had fewer pixels for tree cover at an intermediate percentage (for 31%–70%).

Figure 8. Scatterplot of observed *versus* estimated tree cover percentage for PTC of GM project in 2003 (a) and Vegetation Continuous Fields MOD44B in 2008 (b). This assessment was conducted using the new validation data, which were collected from the pixels including the center of each pixel used for the validation of our result. The tree-cover percentage was divided by 0.8 in MOD44B. The diagonal shows a linear 1:1 relation; dotted lines show the range of 20% error.

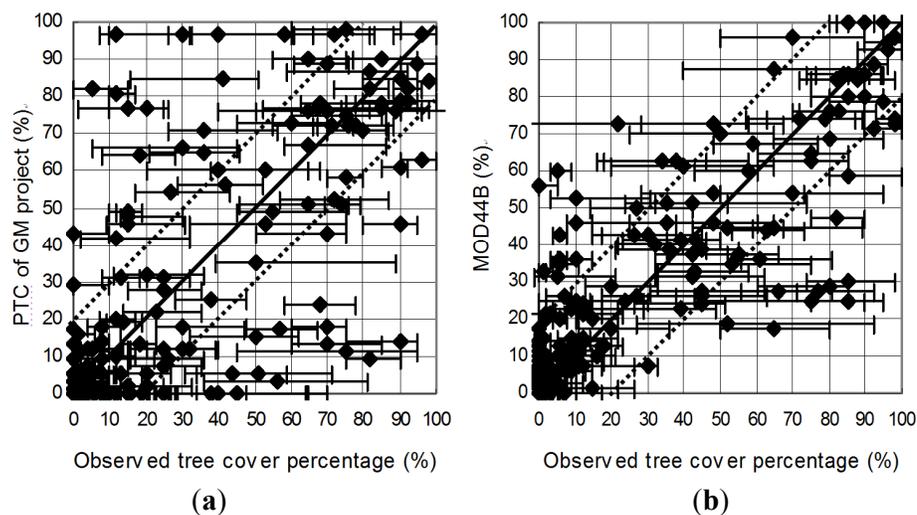


Table 5. RMSE and WRMSE of tree cover percentage in MOD44B and PTC of GM project. RMSE was calculated for five strata by the observed tree cover percentage (0%–20%, 21%–40%, 41%–60%, 61%–80% and 81%–100%). The tree-cover percentage was divided by 0.8 in MOD44B.

Tree Cover Strata (Observed)	PTC of GM Project		MOD44B	
	RMSE (%)	Number of Sites	RMSE (%)	Number of Sites
0–20	13.2	231	9.7	216
21–40	29.8	20	20.3	16
41–60	33.2	14	15.2	19
61–80	27.0	23	29.3	17
81–100	29.8	18	21.9	23
0–100	18.6	306	14.0	291
0–100 (WRMSE)	27.5	306	20.4	291

Table 6. Frequency of the difference of tree-cover percentage among three maps. Water bodies were excluded.

Tree Cover Strata	Difference of Estimated Tree Cover Percentage	Frequency of the Number of Pixels					
		Our Result vs. MOD44B		Our Result vs. PTC of GM Project		MOD44B vs. PTC of GM Project	
		Our Result	MOD44B	Our Result	PTC of GM Project	MOD44B	PTC of GM Project
0–20	0–10	88.4%	89.0%	82.5%	75.1%	83.1%	75.2%
	11–20	8.9%	8.6%	10.6%	11.0%	11.0%	11.5%
	21–30	1.9%	1.5%	1.9%	4.7%	1.8%	4.8%
	31–40	0.5%	0.5%	1.4%	2.9%	1.3%	2.9%
	41–50	0.2%	0.2%	1.1%	2.0%	0.8%	2.0%
	51–60	0.0%	0.1%	0.9%	1.6%	0.7%	1.5%
	61–70	0.0%	0.1%	0.7%	1.2%	0.6%	1.0%
	71–80	0.0%	0.0%	0.5%	0.9%	0.4%	0.5%
	81–90	0.0%	0.0%	0.3%	0.5%	0.2%	0.3%
	91–100	0.0%	0.0%	0.2%	0.1%	0.1%	0.2%
21–40	0–10	52.6%	48.8%	17.4%	28.5%	19.0%	35.2%
	11–20	31.2%	32.1%	23.2%	26.8%	23.1%	28.0%
	21–30	11.7%	14.1%	28.3%	23.4%	27.4%	19.8%
	31–40	3.0%	3.6%	16.2%	11.4%	15.5%	8.4%
	41–50	1.0%	1.0%	8.1%	4.7%	8.0%	3.9%
	51–60	0.4%	0.4%	4.5%	3.1%	4.5%	2.7%
	61–70	0.1%	0.1%	1.9%	1.6%	1.8%	1.4%
	71–80	0.0%	0.0%	0.6%	0.4%	0.6%	0.5%
41–60	0–10	44.5%	45.8%	12.9%	23.3%	14.7%	25.1%
	11–20	30.7%	30.5%	16.7%	22.6%	18.7%	24.6%
	21–30	16.2%	16.1%	25.1%	21.4%	24.8%	22.9%
	31–40	6.6%	6.1%	22.3%	17.8%	21.0%	17.7%
	41–50	1.8%	1.4%	16.4%	12.0%	15.4%	8.2%
	51–60	0.2%	0.1%	6.5%	2.9%	5.4%	1.5%
61–80	0–10	47.7%	49.9%	34.2%	30.1%	39.4%	33.7%
	11–20	30.1%	31.0%	24.0%	24.3%	23.7%	23.5%
	21–30	14.8%	13.5%	13.5%	17.5%	11.3%	15.5%
	31–40	4.8%	4.0%	5.5%	11.5%	4.3%	11.4%
	41–50	1.5%	1.2%	5.1%	8.0%	5.1%	8.8%
	51–60	0.8%	0.3%	7.2%	5.2%	7.8%	5.2%
	61–70	0.3%	0.1%	3.0%	2.7%	6.2%	1.7%
	71–80	0.0%	0.0%	0.0%	0.8%	2.1%	0.2%
81–100	0–10	50.7%	49.1%	37.1%	26.6%	49.1%	35.5%
	11–20	29.5%	28.2%	24.6%	20.0%	24.0%	20.1%
	21–30	12.4%	13.5%	11.8%	15.8%	7.6%	12.2%
	31–40	4.5%	5.9%	6.0%	11.9%	3.5%	8.5%
	41–50	1.5%	2.3%	2.7%	8.4%	2.2%	6.8%
	51–60	0.6%	0.8%	1.6%	6.0%	1.8%	5.9%
	61–70	0.4%	0.2%	3.8%	4.3%	3.0%	5.0%
	71–80	0.3%	0.0%	5.8%	3.2%	3.2%	3.8%
	81–90	0.1%	0.0%	5.5%	2.3%	3.4%	1.7%
	91–100	0.0%	0.0%	1.1%	1.5%	2.2%	0.4%

Table 6. Cont.

Tree Cover Strata	Difference of Estimated Tree Cover Percentage	Frequency of the Number of Pixels					
		Our Result vs. MOD44B		Our Result vs. PTC of GM Project		MOD44B vs. PTC of GM Project	
		Our Result	MOD44B	Our Result	PTC of GM Project	MOD44B	PTC of GM Project
0–100	0–10	73.5%	73.5%	60.2%	60.2%	62.0%	62.0%
	11–20	17.1%	17.1%	14.9%	14.9%	15.3%	15.3%
	21–30	6.4%	6.4%	9.2%	9.2%	8.6%	8.6%
	31–40	2.1%	2.1%	5.9%	5.9%	5.5%	5.5%
	41–50	0.6%	0.6%	3.9%	3.9%	3.6%	3.6%
	51–60	0.2%	0.2%	2.5%	2.5%	2.4%	2.4%
	61–70	0.1%	0.1%	1.6%	1.6%	1.4%	1.4%
	71–80	0.0%	0.0%	1.0%	1.0%	0.7%	0.7%
	81–90	0.0%	0.0%	0.5%	0.5%	0.3%	0.3%
	91–100	0.0%	0.0%	0.2%	0.2%	0.2%	0.2%

Figure 9. Difference map between our resultant tree cover estimation map and MOD44B (a) or PTC of GM project (b). Pixels of MOD44B and our map were resampled into the same position and size of PTC of GM project.

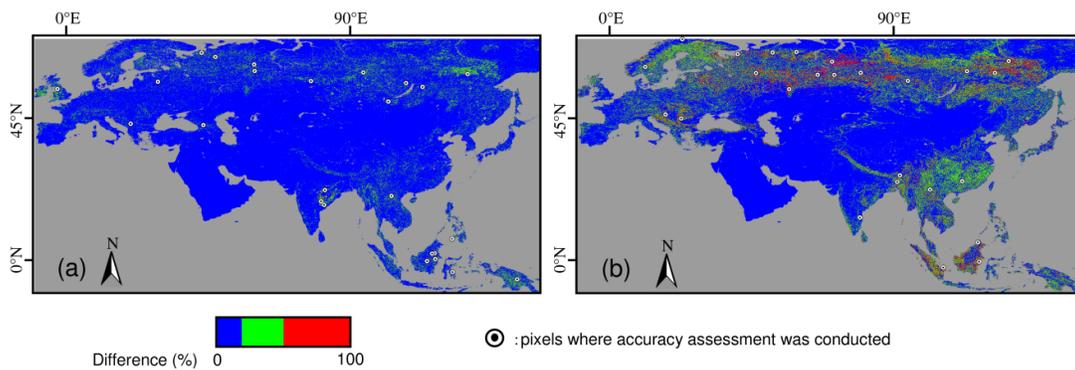
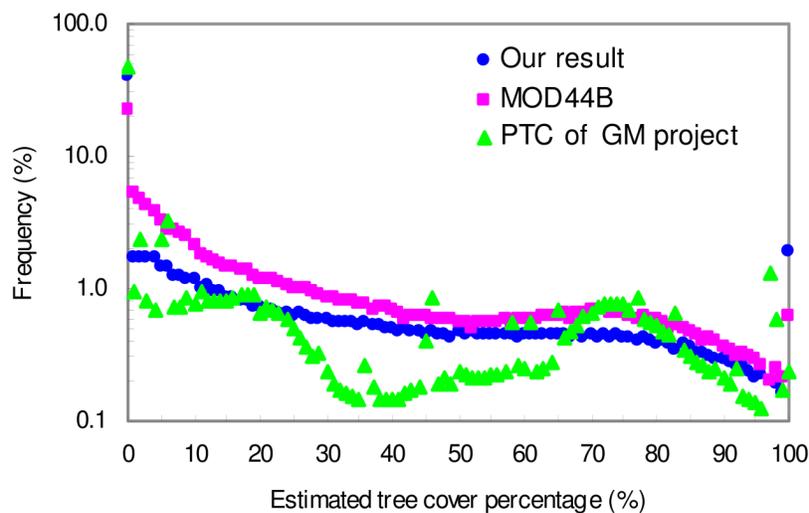


Figure 10. Histograms of pixels according to the tree cover percentage in the whole study area. Water bodies were excluded in advance.



5. Discussion

5.1. The New Tree Cover Percentage Map

We used reference data collected and interpreted from Google Earth to create training data. Actual land cover is extremely complicated. Therefore, various land cover types of training data are necessary to produce more precise estimates. For instance, cropland, urban areas, and trees and soil of many kinds are portrayed in one pixel (500 m × 500 m). However, it is difficult to collect many training data. To accommodate this difficulty, training data were created in a simulation by combining reference data. Our tree cover estimation model was created to fit the evaluation data. Our method presents several advantages. First, when the Google Earth archive of high-resolution images is used, reference data consisting of various land cover types can be collected. Then the decision tree model can be fitted with numerous training data. More than 5,000,000 training data, created from 1,040 sites of reference data, and 456 evaluation data, were used in this study. About 270,000 training data from over 250 Landsat scenes for global were used for initial MOD44B [6]. Training data from only 62 scenes of QuickBird data were used for PTC of GM project, and 26 scenes among them were in Eurasia [8]. The size of one scene of QuickBird image was about 5 km × 5 km. Consequently, MOD44B tend to over-estimate areas of low percent tree cover in the taiga-tundra transition zone, and tree-cover percentage varied greatly depending on factors, such as ecoregion and latitude [17]. Second, this method can partially solve the problem of stability, which is common among decision tree algorithms. Generally, small variations in the training dataset can result in different trees and different predictions for the same validation example. However, deleting some training data or adding new data affects only a small number of pixels of estimation in this method because small regression tree models are applied for each group after each group was classified.

However, the use of images in Google Earth presents some problems in our research. First, it is somewhat difficult to interpret the images as they are displayed only in true color. It is particularly difficult to distinguish between trees and shrubs. Second, many places have no high-resolution images. Third, the images in Google Earth were not always acquired at the peak of the growing season. The tree cover percentage in the growing season could not be estimated in this case. Moreover, the year of image acquisition is not always 2008. This means that training and evaluation data were collected randomly from areas where high-resolution images were available and visual interpretation was not so difficult. This might cause bias to the result. Therefore, we compared our result with other datasets.

5.2. Accuracy Assessment

It is better to assess the accuracy using accurate ground-truth data by random sampling method, especially for shrubs and deciduous forest. But obtaining ground-truth data with the estimation error of less than 10% is difficult. Field survey is necessary for collecting accurate ground-truth data, but it is difficult for random sampling method. Therefore, we used images in Google Earth as reference data for the accuracy assessment. The use of Google Earth are effective for collecting the reference data, especially for validation by random sampling methods, despite the drawbacks of using Google Earth in image availability, quality, and date. The problem of our validation method was that validation sites might be biased as high-resolution images of Google Earth were limited to specific areas and specific

land cover types. Actually, 293 pixels were excluded from the 600 randomly sampled pixels in the accuracy assessment of our new map. In some cases, it was difficult to interpret shadowed areas in forests: they might be on the tree canopy, but might be on gaps between trees. Image-interpretations of percent tree cover from QuickBird color images had an RMSE of 14.8% by different interpreters [17]. Our result should be regarded as better than those of other studies only in the areas where validation data were collected.

5.3. Comparison among Three Datasets

As our validation method had some problems as described above, we assessed our result and other two maps in randomly sampled pixels, where the difference of the estimation between them was greater than 50%. This was because it was difficult to get the accurate reference data in randomly sampled pixels with the estimation error of less than 50%. Our estimation result was more coincident with the reference data interpreted from images in Google Earth than other two datasets. However, several pixels were difficult to judge because some pixels were on the border of different land cover types and some pixels were difficult to distinguish trees from other vegetation. The statistical analysis showed that no artificial characteristic was included in the total number of pixels in our estimation as occurs with PTC of GM project, which had many pixels at the specific values of tree cover percentage. The original MOD44B had almost no pixels whose tree cover was greater than 80% because of the limitation of training data. In MOD44B, tree canopy cover was used for the definition of tree cover instead of tree crown cover. Canopy cover is the crown cover excluding gaps inside the crown. For this reason, tree-cover percentage in MOD44B was divided by 0.8 in this assessment. It is noteworthy that the resolution, the year, and the data used for the generation of products differed among the three maps.

6. Conclusions

We estimated the tree cover percentage in Eurasia at a pixel size of 500 m using the supervised regression tree method with modification using evaluation data collected in advance from the entire study area. Two new parts in the method of this study are: (1) training data of various kinds were produced through simulation from reference data based on images available in Google Earth; (2) tree-cover percentage was estimated for groups and the groups were further divided into subgroups so that the estimation agree with the randomly sampled evaluation data. Fundamentally a small regression tree model would not be applicable for all pixels on a global scale. It was better to produce many regression tree models and apply them area-by-area. The RMSE of the estimated tree cover percentage was better than existing datasets. For some pixels, the difference between our resultant map and the other two maps was greater than 50%. From the pixel-level assessment of those pixels, our result was better than MOD44B. Mainly, our result and MOD44B underestimated tree cover in those pixels. The statistical comparison of tree-cover percentage among maps showed that the three maps exhibited different characteristics. More accurate validation and comparison with previous studies are necessary for our result. Despite the limitations in the validation data, the map produced by our method is valuable. The accuracy of our result was not low compared with other studies. Users of these maps can modify each map for areas or pixels in which the difference is greater than 50%. They also can know which regions have the lower or higher map uncertainty by comparing them. Our map can be

substituted for existing percent tree-cover products. This study was done for mapping tree cover percentage on a global scale. We are planning to proceed to mapping the tree cover percentage in other continents using the method proposed in this report.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 22220011. This study was partially supported by Geospatial Information Authority of Japan (GSI). The authors especially thank students Xuehua Chen, Zhongkai Cheng, Xianmixinuer Kelimu, Ying Li, Xiaolin Wei, and Naijia Zhang for spending much valuable time interpreting Google Earth images and producing validation data. The authors also thank Rokhmatuloh for teaching us basic concepts and how to use some software, Nguyen Thanh Hoan and Nehal Soliman for contributing valuable comments and discussions, and Bayan Alsaaidh and Kalibinuer Yishamiding for helping to interpret high-resolution images.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. FAO (Food and Agriculture Organization of the United Nations). *Forest Resources Assessment Programme Working Paper 33: FRA 2000 on Definitions of Forest and Forest Change*; FAO: Rome, Italy, 2000.
2. FAO (Food and Agriculture Organization of the United Nations). *Global Forest Resources Assessment 2010 Main Report*; FAO: Rome, Italy, 2010.
3. DeFries, R.S.; Hansen, M.C.; Townshend, J.R.G. Global continuous fields of vegetation characteristics: A linear mixture model applied to multi-year 8 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1389–1414.
4. DeFries, R.S.; Hansen, M.C.; Townshend, J.R.G.; Janetos, A.C.; Loveland, T.R. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Glob. Chang. Biol.* **2000**, *6*, 247–254.
5. Geospatial Information Authority of Japan, Chiba University and Collaborating Organizations. Global Map—Global Land Cover (GLCNMO), Global Map—Percent Tree Cover, 2008. Available online: <http://www.iscgm.org/login.html> (accessed on 1 April 2013).
6. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Carroll, M.; Dimiceli, C.; Sohlberg, R.A. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interact.* **2003**, *7*, 1–15.
7. Rokhmatuloh; Nitto, D.; Al-Bilbisi, H.; Arihara, K.; Tateishi, R. Estimating percent of tree cover using regression tree method with very-high-resolution QuickBird images as training data. *J. Remote Sens. Soc. Jpn.* **2007**, *27*, 1–12.
8. Rokhmatuloh, R.; Tateishi, R.; Al-Bilbisi, H.; Arihara, K.; Kobayashi, T.; Nitto, D.; Lee, S.A.; Hirabayashi, K.; Lu, Y.Q.; Lu, C.; *et al.* Global percent tree cover mapping using regression tree method: An advantage of QuickBird images as training data. *Asian J. Geoinform.* **2010**, *10*, 21–28.

9. Berberoglu, S.; Satir, O.; Atkinson, P.M. Mapping percentage tree cover from Envisat MERIS data using linear and nonlinear techniques. *Int. J. Remote Sens.* **2009**, *30*, 4747–4766.
10. Heiskanen, J.; Kivinen, S. Assessment of multispectral, -temporal and -angular MODIS data for tree cover mapping in the tundra-taiga transition zone. *Remote Sens. Environ.* **2008**, *112*, 2367–2380.
11. Schwarz, M.; Zimmermann, N.E. A new GLM-based method for mapping tree cover continuous fields using regional MODIS reflectance data. *Remote Sens. Environ.* **2005**, *95*, 428–443.
12. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Marufu, L.; Sohlberg, R. Development of MODIS tree cover validation data set for Western Province, Zambia. *Remote Sens. Environ.* **2002**, *83*, 320–335.
13. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Sohlberg, R.; Dimiceli, C.; Carroll, M. Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sens. Environ.* **2002**, *83*, 303–319.
14. White, M.A.; Shaw, J.D.; Ramsey, R.D. Accuracy assessment of the vegetation continuous field tree cover product using 3954 ground plots in the south-western USA. *Int. J. Remote Sens.* **2005**, *26*, 2699–2704.
15. 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*, doi:10.1029/2010GL043622.
16. Heiskanen, J. Evaluation of global land cover data sets over the tundra-taiga transition zone in northernmost Finland. *Int. J. Remote Sens.* **2008**, *29*, 3727–3751.
17. Montesano, P.M.; Nelson, R.; Sun, G.; Margolis, H.; Kerber, A.; Ranson, K.J. MODIS tree cover validation for the circumpolar taiga-tundra transition zone. *Remote Sens. Environ.* **2009**, *113*, 2130–2141.
18. Arino, O.; Bicheron, P.; Achard, F.; Latham, J.; Witt, R.; Weber, J.L. GLOBCOVER: The most detailed portrait of Earth. *Eur. Space Agency Bull.* **2008**, *136*, 24–31.
19. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182.
20. Herold, M.; Mayaux, P.; Woodcock, C.E.; Baccini, A.; Shmullius, C. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens. Environ.* **2008**, *112*, 2538–2556.
21. Pflugmacher, D.; Krankina, O.N.; Cohen, W.B.; Friedl, M.A.; Sulla-Menashe, D.; Kennedy, R.E.; Nelson, P.; Loboda, T.V.; Kuemmerle, T.; Dyukarev, E.; *et al.* Comparison and assessment of coarse resolution land cover maps for Northern Eurasia. *Remote Sens. Environ.* **2011**, *115*, 3539–3553.
22. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Carroll, M.; Dimiceli, C.; Sohlberg, R.A. Vegetation Continuous Fields MOD44B, 2008 Percent Tree Cover, Collection 5. 2011. Available online: <http://www.glcfc.umd.edu/data/vcf/> (accessed on 1 April 2013).
23. FAO (Food and Agriculture Organization of the United Nations). *Global Forest Resources Assessment Update 2005: Terms and Definitions*; FAO: Rome, Italy, 2004.
24. Hallé, F.; Oldeman R.A.A.; Tomlinson, P.B. What is a Tree? In *Tropical Trees and Forests: An Architectural Analysis*; Springer-Verlag: Berlin, Germany, 1978; pp. 1–6.
25. Nicholson, B.E.; Clapham, A.R. *The Oxford Book of Trees*; Oxford University Press: London, UK, 1975.

26. Oldeman, R.A.A. Second Part: Forest Components. In *Forests: Elements of Silvology*; Springer-Verlag: Berlin, Germany, 1990; pp. 27–154.
27. Thomas, P. *Trees: Their Natural History*; Cambridge University Press: Cambridge, UK, 2000.
28. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Koppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644.
29. Hoan, N.T.; Tateishi, R.; Al-Bilbisi, H. *Global MODIS 2008 Data User's Manual*; Chiba University: Chiba, Japan, 2011.
30. Schaaf, C.B.; Gao, F.; Strahler, A.H.; Lucht, W.; Li, X.; Tsang, T.; Strugnell, N.C.; Zhang, X.; Jin, Y.; Muller, J.; *et al.* First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.* **2002**, *83*, 135–148.
31. Schaaf, C. *MODIS BRDF/Albedo Product (MCD43) User's Guide*; 2004. Available online: <http://www-modis.bu.edu/brdf/userguide/intro.html> (accessed on 1 April 2013).
32. Schaaf, C.; Liu, J.; Gao, F.; Jiao, Z.; Shuai, Y.; Strahler, A. Collection 005 Change Summary for MODIS BRDF/Albedo (MCD43) Algorithms. Available online: http://landweb.nascom.nasa.gov/QA_WWW/forPage/C005_Change_BRDF.pdf (accessed on 1 April 2013).
33. Clark, M.L.; Aide, T.M.; Grau, H.R.; Riner, G. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. *Remote Sens. Environ.* **2010**, *114*, 2816–2832.
34. Joshi, C.; Leeuw, J.D.; Skidmore, A.K.; van Duren, I.C.; van Oosten, H. Remotely sensed estimation of forest canopy density: A comparison of the performance of four methods. *Int. J. Appl. Earth Obs.* **2006**, *8*, 84–95.
35. Hansen, M.C.; Townshend, J.R.G.; DeFries, R.S.; Carroll, M. Estimation of tree cover using MODIS data at global, continental and regional/local scales. *Int. J. Remote Sens.* **2005**, *26*, 4359–4380.
36. Rikimaru, A.; Roy, P.S.; Miyatake, S. Tropical forest cover density mapping. *Trop. Ecol.* **2002**, *43*, 39–47.
37. Pal, M. M5 model tree for land cover classification. *Int. J. Remote Sens.* **2006**, *27*, 825–831.
38. Quinlan, J.R. Learning with Continuous Classes. In Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; World Scientific Press: Singapore, Singapore, 1992; pp. 343–348.
39. Li, A.; Huang, C.; Sun, G.; Shi, H.; Toney, C.; Zhu, Z.; Rollins, M.G.; Goward, S.N.; Zhu, Z. Modeling the height of young forests regenerating from recent disturbances in Mississippi using Landsat and ICESat data. *Remote Sens. Environ.* **2011**, *115*, 1837–1849.
40. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106.
41. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers: San Mateo, CA, USA, 1993.
42. Quinlan, J.R. Improved use of continuous attributes in C4.5. *J. Artif. Intell. Res.* **1996**, *4*, 77–90.