Remote Sens. **2014**, *6*, 5325-5343; doi:10.3390/rs6065325

OPEN ACCESS

remote sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

A Circa 2010 Thirty Meter Resolution Forest Map for China

Congcong Li [1], Jie Wang [2], Luanyun Hu [3], Le Yu [3], Nicholas Clinton [3], Huabing Huang [2], Jun Yang [3] and Peng Gong [2,3,4,5,*]

[1] State Key Laboratory of Remote Sensing Science, and College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China; E-Mail: licc129@163.com

[2] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; E-Mails: wangjie@radi.ac.cn (J.W.); huanghb@radi.ac.cn (H.H.)

[3] Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing 100084, China; E-Mails: huluanyun@gmail.com (L.H.); leyu@tsinghua.edu.cn (L.Y.); nicholas.clinton@gmail.com (N.C.); larix001@gmail.com (J.Y.)

[4] Joint Center for Global Change Studies, Beijing 100875, China

[5] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720-3114, USA

***** Author to whom correspondence should be addressed; E-Mail: penggong@tsinghua.edu.cn; Tel.: +86-10-6279-5385; Fax: +86-10-6279-7284.

Received: 14 February 2014; in revised form: 22 May 2014 / Accepted: 23 May 2014 / Published: 10 June 2014

Abstract: This study examines the suitability of 30 m Landsat Thematic Mapper (TM), 250 m time-series Moderate Resolution Imaging Spectrometer (MODIS) Enhanced Vegetation Index (EVI) and other auxiliary datasets for mapping forest extent in China at 30 m resolution circa 2010. We calculated numerous spectral features, EVI time series, and topographical features that are helpful for forest/non-forest distinction. In this research, extensive efforts have been made in developing training samples over difficult to map or complex regions. Scene by scene quality checking was done on the initial forest extent results and low quality results were refined until satisfactory. Based on the forest extent mask, we classified the forested area into 6 types (evergreen/deciduous broadleaf, evergreen/deciduous needleleaf, mixed forests, and bamboos). Accuracy assessment of our forest/non-forest classification using 2195 test sample units independent of the training sample indicates that the producer's accuracy (PA) and user's accuracy (UA) are 92.0% and 95.7%, respectively. According to this map, the total forested area in China was

164.90 million ha (Mha) circa 2010. It is close to the forest area of 7th National Forest Resource Inventory with the same definition of forest. The overall accuracy for the more detailed forest type classification is 72.7%.

Keywords: classification; MODIS; TM; forest extent; forest type

1. Introduction

Forests cover about 31 percent of the world's land surface [1]. They provide habitats for 90% of terrestrial plants and most animals, and play an indispensable role in biogeochemical cycling, hydrological cycling, biodiversity conservation and climate change mitigation [2]. China is one of the five most forest-rich countries [1,3]. The 7th National Forest Resource Inventory (NFRI) in China conducted between 2004 and 2008 found that the forest extent was 195.45 million hectares (20.36% of China's terrestrial area) which was increased by 2.15% compared to the 6th NFRI conducted between 1999 and 2003. In the 7th NFRI, more than 20 thousand people spent 5 years to measure 415,000 permanent field sample plots and interpreted 2.8 million remote sensing sample plots [3]. NFRI is a sample based inventory, not a spatially explicit database. The Chinese government resolved to increase China's forest cover area by 40 million hectares from 2005 to 2020 [4]. Timely and accurate spatially explicit forest extent maps would be extremely valuable to forest managers at various administrative levels responsible for realizing the national afforestation goal.

Remote sensing is an important tool for monitoring environmental changes [5]. Liu *et al.* [6] developed land use maps for China by manual interpretation of Landsat TM and Enhanced TM plus (ETM+) imagery. In their map, 3 forest types were included, *i.e.*, "Forest Land", "Open Forest Land", "Other Forest Land". Zhang *et al.* [7] generated a Chinese land cover map for 2005 using a human-computer interaction method based on Landsat TM and China Brazil Earth Resources Satellite (CBERS) data and classified "Forest Land" from a land-cover perspective. Hu *et al.* [8] produced a 30 m land cover map of China by applying an efficient image clustering technique to Landsat TM data. Human-computer interaction is time- and labor-consuming and difficult to reproduce. In addition, the accuracy depends on the knowledge of interpreters. In recent years, more openly available remotely sensed data and new classification algorithms [9,10] have made it possible to map forests at large spatial scales more consistently and efficiently.

There are three existing global land cover maps that give some information on forest cover in China. European Space Agency (ESA) produced the GlobCover 2009 (Global land cover map) based on bi-monthly Medium Resolution Imaging Spectrometer (MERIS) data [11] with a resolution of 300 m. Forests are classified using the Land Cover Classification System (LCCS) [12] based on the cover, leaf type (*i.e.*, broadleaf, needleleaf) and leaf phenology (*i.e.*, evergreen, deciduous). The second forest extent map is the MODIS land-cover map (MCD12Q1) based on MODIS 500 m monthly data [13,14]. Forests are classified according to the leaf type and leaf phenology. However, many forests in China are distributed in patches smaller than 300 m in dimension. The third forest map is FROM-GLC (short for Finer Resolution Observation and Monitoring—Global Land Cover) [15], which is the first global land cover map at 30 m resolution created using Landsat TM/ETM+ imagery. Four forest types

(Level 2 type), *i.e.*, "Broadleaf Forest", "Needleleaf Forest", "Mixed Forest", "Orchard" were mapped (grouped as "Forest" in an aggregated class). In FROM-GLC, the producer's accuracy (PA) and user's accuracy (UA) for forest in China are 64.74% and 83.01%, respectively. Follow-up works of FROM-GLC used MODIS EVI time series and auxiliary datasets to improve the overall accuracies (OA), and the PA and UA for forest became 71.76% and 79.40%, respectively [16,17]. However, all these products aimed to map general land cover types not specifically forest cover.

There are also two global forest cover datasets around the year 2010. The MODIS Vegetation Continuous Fields (VCF) Tree Cover dataset was produced at 250 m resolution globally [18]. Hansen *et al.* [19] mapped 30 m resolution global forest cover extent based on Landsat time series. However, these two products concentrated on the percentage of tree cover not the forest types.

In this study, we aimed to produce a more detailed/accurate 30 m resolution forest map for China using multi-source remotely sensed data (*i.e.*, Landsat TM/ETM+, MODIS, SRTM) and additional features from ancillary sources (*i.e.*, spectral, topographical features). We followed the definition of forest provided by the FAO (Food and Agriculture Organization of the United Nations), which is "*land with tree crown cover of more than 10 percent. The height of trees should reach at least 5 m in situ*" [20]. Tree plantations such as orchards are excluded from forest in this classification system. We mapped forests at a per-pixel scale, insuring that all pixels meeting criteria of canopy coverage and height would be mapped as forest. Taking into account the spatial resolution, spectral-information limitations of the data and the characteristics of China's forest types, we further classified the forest cover into 6 types: evergreen broadleaf, deciduous broadleaf, evergreen needleleaf, deciduous needleleaf, mixed forests and bamboos. The definitions of the first five types are drawn from the International Global Biosphere Programme (IGBP) classification system [21]. Forests with more than 60% bamboo canopy cover are considered as bamboos. Bamboos mixed with other types (each type ≤ 60%) are considered as mixed forests.

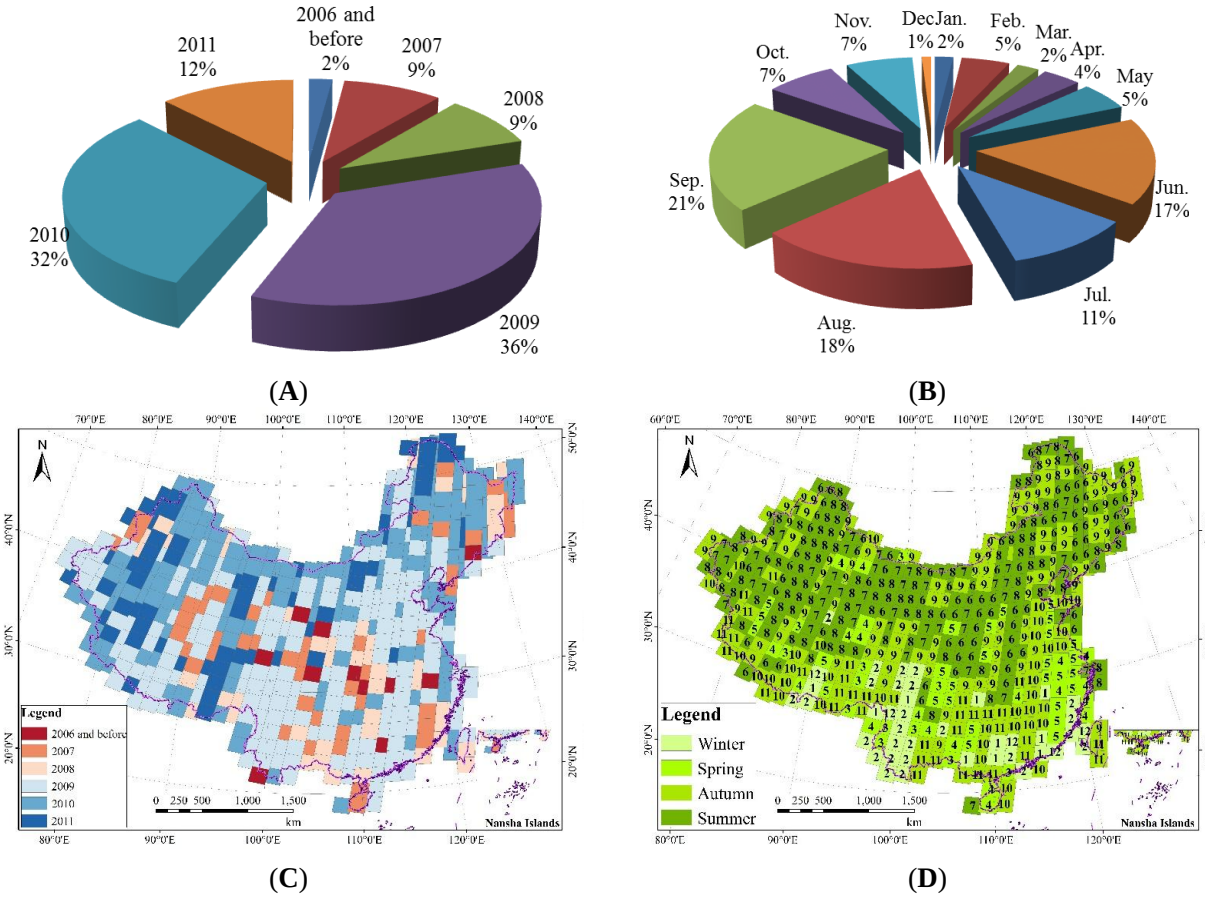
2. Data Sets

2.1. Landsat Data

We collected 520 Landsat TM scenes from United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) data center [22] and Center for Earth Observation and Digital Earth [23]. Scenes with fewer clouds obtained in the growing season were selected (Figure 1) covering a time span from 2005 and 2011 with 98% of them acquired after 2007 (Figure 1A). These images were the best that we could collect from those data sources after 2006 based on seasonality and low cloud cover criteria. As a result, only 8% of the images were obtained in winter, but they were mainly distributed in the subtropical regions (Figure 1B,D). The images acquired in spring were mainly distributed in the South and some non-forest areas in the west (Figure 1D). The deciduous forests in these regions get green in April. Although many images were acquired in autumn, they are mostly in September and October (Figure 1B). The deciduous forests in the Northeast begin losing their leaves the earliest in China. The images in these regions are mainly obtained in September, before the leaf-falling stage. The images acquired in November are mainly distributed in the regions where evergreen forests or no forests grow (Figure 1D). Thus, forest land cover was only minimally affected by phenological differences. All Landsat TM/ETM+ images were calibrated and atmospherically

corrected using the Global Mapper (GM) software, which implements an automatic Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm [15].

Figure 1. (**A**) Acquisition years of Landsat TM data. (**B**) Acquisition months of Landsat TM data. (**C**) Annual distribution of Landsat TM data. (**D**) Seasonal and monthly distribution of Landsat TM data (the labels are acquisition months).



2.2. MODIS

250 m resolution MODIS data provide additional time-series information which is important to capture vegetation dynamics useful in classification. We collected the MOD13Q1 product [24] for the whole year of 2010. It contains NDVI, EVI, quality assessment bands, and 4 reflectance bands (Bands 1, 2, 3, and 7). We selected EVI time-series only for further analysis, the reasons being: (1) EVI is a composite index from 3 reflectance bands. It works as a dimensionality reduction process by feature extraction; Using EVI instead of the original reflectance bands requires much less computational time; (2) EVI is more sensitive than NDVI to high LAI especially when LAI is above 2, and EVI performs better in areas contaminated by smoke plumes [25–27].

2.3. Shuttle Radar Topography Mission (SRTM) DEM

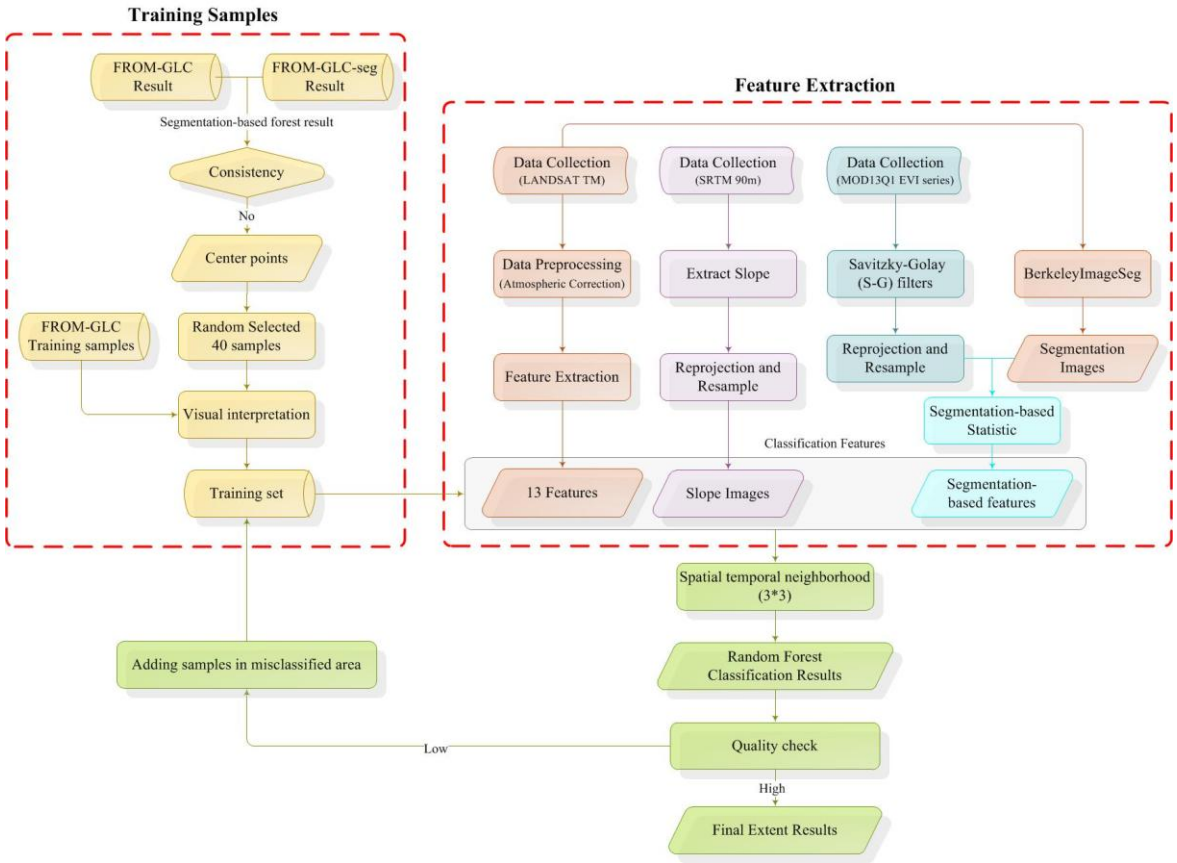
Considering the complex topography of China and soil moisture conditions on different slopes, topography is an important factor to forest distribution. We collected 90 m digital elevation data from the Shuttle Radar Topography Mission (SRTM) [28]. Slope information for each pixel was calculated

based on the SRTM. Slope data were then reprojected and resampled to the Landsat TM extent and resolution for subsequent use.

3. Methods

The forest mapping effort involved two main parts (Figure 2), forest/non-forest classification and more specific forest type classification. For the forest/non-forest classification, there were three main steps. First, numerous features in support of forest identification were calculated from TM, MODIS, and the DEM. Second, training samples for forest and other vegetation cover types (*i.e.*, Cropland, Grassland, Shrubland) were collected and refined. Finally, an advanced machine learning algorithm—Random Forest was used, followed by quality checking and refinement.

Figure 2. The workflow for forest/non-forest classification.



3.1. Forest/Non-Forest Classification

3.1.1. Feature Extraction

Three indices, *i.e.*, enhanced vegetation index (EVI), bare soil index (BI), and shadow index (SI) were calculated from TM images. Forests are expected to have higher EVI values, higher SI values and lower BI values than grassland in the growing season [29]. The wetland we mentioned here are marsh land. In these regions, the herbs have sufficient water supply and grow well. Their spectral characteristics are similar to that of broadleaf forests. In order to distinguish forests from wetlands,

we used the wetness from the Kauth—Thomas (KT) transformation [30,31]. Also greenness and brightness from the KT transform were used when there was a lack of EVI and BI indices.

We also used information in the EVI time-series from the MODIS data. We took the advantage of such data to distinguish forest from farmland. Although forests have similar phenological characteristics to other natural vegetation types such as grass and shrub, the phenological information of crops derives from human activities. In the northeast of China, spring wheat is harvested in the summer, while paddy fields are flooded in early June. In double cropping regions, harvest is usually conducted in early summer. Thus, EVI seasonal features of crops are usually shorter than forests because of harvest and irrigation. MODIS EVI was processed by a Savitzky-Golay filter (S-G) [32] to minimize the effect of the clouds and to obtain reliable EVI values. The EVI time-series obtained in the winter were averaged. Then a segmentation based multi-resolution integration approach [16] was used to link TM pixels with MODIS EVI pixels and pixels in other auxiliary data. We averaged the EVI values in the same segments, and then assigned these pixels with the same EVI values. For segment extraction, we used BerkeleyImageSeg [33] to perform image segmentations. The segmentation parameters are threshold, shape and compactness. The segmentation parameters are threshold, shape and compactness. The threshold determined the sizes of the segments. Higher thresholds will produce larger segments. The shape parameter controls the weighting between shape and spectral characteristics during the merging process. A greater shape parameter means a higher weighting of shape compared to spectral heterogeneity. The compactness parameter controls the compactness of the objects. A larger value produces more compact segments while a smaller value produces more smooth segments [34]. The threshold, shape parameter and compactness parameter were set to 15, 0.2 and 0.7, respectively. All calculated features used for forest/non-forest classification are listed in Table 1.

Table 1. List of input features for forest/non-forest classification.

Feature	**Formula or Algorithm**
Spectral features	Surface reflectance Band 1–Band 5, Band 7
Temperature	Calibrated from thermal infrared band
Enhanced Vegetation Index (EVI)	$EVI = G \times \frac{NIR - RED}{NIR + C1 \times RED - C2 \times BLUE + L}$
Shadow Index (SI)	$((10000 - BLUE) \times (10000 - GREEN) \times (10000 - RED))^{1/3}$
Bare soil Index (BI)	$\frac{(SWIR + RED) - (BLUE + NIR)}{(SWIR + RED) + (BLUE + NIR)}$
Greenness	−0.2848 (Band 1) − 0.2435 (Band 2) − 0.5436 (Band 3) + 0.7243 (Band 4) + 0.0840 (Band 5) − 0.1800 (Band 7)
Wetness	0.1509 (Band 1) + 0.1973 (Band 2) + 0.3279 (Band 3) + 0.3406 (Band 4) − 0.7112 (Band 5) − 0.4572 (Band 7)
Brightness	0.3037 (Band 1) + 0.2793 (Band 2) + 0.4743 (Band 3) + 0.5585 (Band 4) + 0.5082 (Band 5) + 0.1863 (Band 7)
Slope	Calculated from SRTM
Mean Winter EVI	Averaged EVI values of December, January and February
EVI time series	All the EVI values except those in winter

3.1.2. Training and Test Sample Collection

First, the training sites used in the FROM-GLC project were examined. Each training sample unit in China was checked according to TM images and Google Earth to ensure its validity. There were a total of 9037 sample units in China including 1559 forest sample units, 1175 crop sample units, 903 grass sample units, and 224 shrub (including orchards) sample units. Other sample units came from water bodies, wetland, bareland, and snow/ice. Since the number of orchards was too small to justify an additional class, they were merged into shrubs during the classification. This number of sample units was far from sufficient to get satisfactory classification results. However, adding representative training samples could increase classification accuracy more than changing classification algorithms [10].

Adding more training sample units became the most important part of the current research. We took advantage of the classification results from FROM-GLC and FROM-GLC-seg. They share the same training set but different features and classification methods. Specifically, FROM-GLC was generated using Landsat TM images as the only input feature; FROM-GLC-seg was produced using a segmentation based approach to integrate many features (*i.e.*, TM, MODIS, DEM, *etc.*) with different spatial resolution. They are all freely available online [35]. A majority statistic was calculated in each segment. In total, forty segments with discrepant classification of forest *vs.* non-forest between FROM-GLC and FROM-GLC-seg were randomly selected from each TM scene. The center points of these segments were interpreted and added into the training sample. These additional sample units often fall into difficult-to-map areas, the addition of which is conducive to improving the classification results.

Test samples were collected from the common global land cover validation database [36]. Two experienced interpreters checked further through the forest sample, especially when a sample unit was contaminated by atmospheric interference.

3.1.3. Forest/non-forest Classification

We compared 15 common classification algorithms including support vector machine (SVM), Random Forest (RF), Bagging, and AdaBoost in an urban region [10]. RF was chosen here for its relatively low computational cost but high accuracy performance while handling a large number of features with large training sample sizes. The random forests algorithm performed in this study is in R. There are two critical parameters in RF, *i.e.*, number of trees (numTrees) and the number of attributes to be used at each node (numFeatures). A value of numFeatures a bit smaller than $\sqrt{N}$ was shown to be effective for classification, where N is the number of features [10]. There are 32 features in total. Therefore, for each node, 5 features were randomly chosen to be used in classification tree induction. Using random features reduces the correlation between trees while maintaining generalization ability. Since increasing the number of trees will not hurt classifier performance, but may help [37], we set numTrees to be 200.

Considering the consistency in space and time, we used locally adaptive training samples from 8 neighboring scenes with an acquisition time in the ± 30 day range from an image to be classified. In other words, training samples in local spatial and temporal neighborhoods of an image were used to

train an RF model (classifier) with input features listed in Table 1. The RF model was only applied to classify the one image.

3.1.4. Post-Classification

We checked the quality of every initially classified image to ensure a reliable forest map could be produced from the individual scene result. Additional training samples were collected and used in poorly classified areas and images were re-classified until a satisfactory result was achieved. The criterion was that there was less than 5% misclassified area according to visual interpretation.

3.2. Forest Type Classification

Forest type classification was done based on the forest extent mapping result described above. Because differences in the acquisition time of the Landsat images cause inconsistency in forest type classification, we used the MODIS EVI time series instead of the LANDSAT TM spectral data in the forest type classification. The phenology of different forest types varies, so they could be classified using these EVI features. For example, evergreen forests have higher winter EVI values than the deciduous forests. And the mixed evergreen and deciduous would have medium winter EVI values. To distinguish needleaf and broadleaf forests, the summer EVI values are helpful. Bamboos are mostly cultivated. So their phenology is different from natural forests in the same regions. Although the MODIS EVI values are combined with TM based on segments, the resolution is still lower than 30 m. Only large sample units of forest, homogeneous in 250 m × 250 m around a Landsat image pixel, were used.

Vegetation growth has regional characteristics as a result of climatic features such as temperature and precipitation within an ecological zone. However, vegetation growth is continuous, there is no clear-cut distinction of forest types in neighboring ecological zones. Instead of using ecological zones, we used geographical coordinate (*i.e.*, Latitude and Longitude) to represent spatial neighborhood in this research. With these features, training sample units in the same region will be assigned to the same branch during tree induction if, at that node, geographic location is the most effective predictor of forest class.

All large sample units in the training sample were used to train an RF model (classifier) using the input features (MODIS EVI time series, slope, and geographical coordinates). Then this RF model was applied to classify all TM scenes with the mask of forest extent to get the forest types.

4. Results and Analysis

4.1. Results

We mosaicked all 520 scenes into a forest map for China. In this process, we clipped each scene using boundary metadata downloaded from the USGS website [38] to eliminate the sawtooth edges. For overlapping areas, a pixel was classified as forest when at least one scene indicated so. This step also helped to eliminate cloud contamination. Finally, we reprojected the mosaicked map to the Albers projection. From the mosaicked forest map, we estimated the total area of forest in China to be 164.90 million ha (Mha) (17.38% of land area), which is 41.96 Mha lower than the area reported by

FAO FRA in year 2010 (2010). According to 7th NFRI, there are 1,555,899 km^2 of "arbor forests" (with tree canopy cover more than 20%), 53,810 km^2 of bamboo forests [39] and 48,222 km^2 of "open forest" (with tree canopy cover 10%–19%) [40]. As defined by FAO, the forest area of the 7th NFRI in China is 1,657,931 km^2 (about 165.79 Mha). This is in close agreement to the result obtained here, with a difference of only 0.89 Mha, which is less than 0.1% of the territory area.

Figure 3 illustrates the forest map of China. From the map, we can conclude that deciduous broadleaf forests are mainly distributed in Shaanxi, Northeast, and the North China Plain regions. Evergreen broadleaf forests are mainly distributed in South China up to the latitude of 25 °N. Deciduous needleleaf forests are mostly found in Daxingan region. Evergreen needleleaf forests are mainly in the northwest and southwest area. Mixed forests are widely distributed. Most pure Bamboos are found in Anhui, Zhejiang and Fujian Provinces.

Figure 3. The forest map of China.

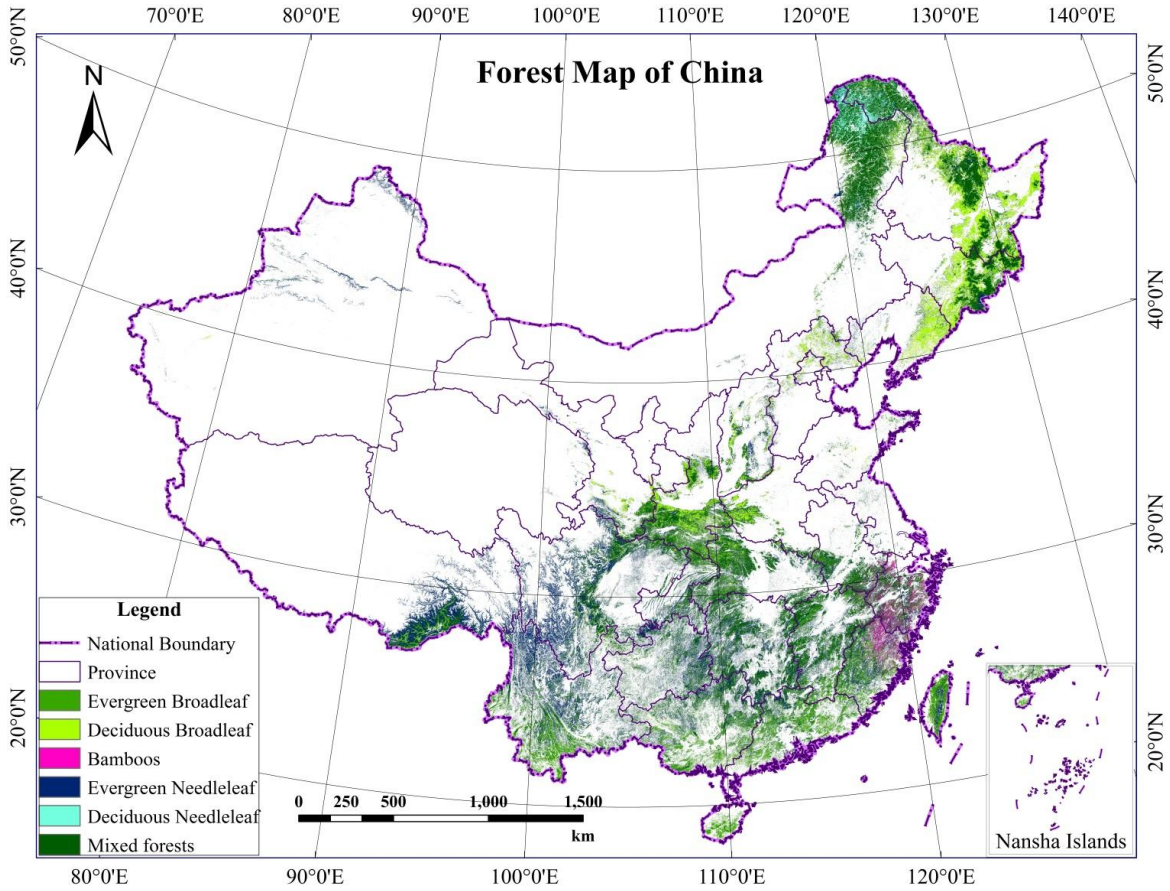


Table 2 is forest extent estimated at the provincial level. The top 5 provinces with the greatest forest extent in China are Heilongjiang, Yunnan, Sichuan, Inner Mongolia and Hunan. The top 5 provinces with the greatest percentage of forest cover are Taiwan, Fujian, Zhejiang, Jiangxi and Hong Kong, which are located in the south of China. The 5 provinces with lowest forest area are Shanghai, Qinghai, Xinjiang, Ningxia, and Tianjin. These regions are either economically developed or ecologically fragile areas.

Table 2. Forest extent in different provinces.

Province	Forest Area (10,000 ha)	Forest Coverage Rate	Province	Forest Area (10,000 ha)	Forest Coverage Rate
Anhui	322.691	23.00%	**Liaoning**	313.901	21.57%
Macao	0.028	12.47%	**Inner Mongolia**	1338.422	11.67%
Beijing	25.334	15.46%	**Ningxia**	5.194	1.00%
Fujian	674.005	55.40%	**Qinghai**	23.512	0.33%
Gansu	276.577	6.82%	**Shandong**	52.075	3.39%
Guangdong	646.289	36.56%	**Shanxi**	248.106	15.86%
Guangxi	909.395	38.49%	**Shaanxi**	766.779	37.24%
Guizhou	593.855	33.75%	**Shanghai**	0.055	0.09%
Hainan	80.465	23.80%	**Sichuan**	1347.007	27.84%
Hebei	176.739	9.44%	**Taiwan**	210.065	58.25%
Henan	219.540	13.26%	**Tianjin**	1.380	1.19%
Heilongjiang	1822.689	40.27%	**Tibet**	790.407	6.57%
Hubei	714.795	38.43%	**Hong Kong**	4.929	46.45%
Hunan	938.114	44.24%	**Xinjiang**	157.376	0.96%
Jilin	735.188	38.52%	**Yunnan**	1450.589	37.85%
Jiangsu	37.491	3.71%	**Zhejiang**	514.054	50.44%
Jiangxi	797.623	47.75%	**Chongqing**	295.540	35.84%
Total	16,490.208	17.38%			

4.2. Classification Accuracy

4.2.1. Forest/Non-Forest Classification Accuracy

The accuracy of forest extent was assessed by using 2195 samples (including 388 forest sample units, 1790 non-forest sample units, and 17 uncertain sample units). The producer's accuracy (PA) and user's accuracy (UA) for forest based on sample units with high certainty are 92.0% and 95.7%, respectively. If the 17 uncertain sample units were included as forest, then the producer's accuracy (PA) and user's accuracy (UA) for forest would become 88.4% and 95.7%, respectively. If they were taken as non-forest, then UA and PA accuracies would still be 92.0% and 95.7%, respectively. That is to say, the omission error of the forest extent is between 11.6% and 8%, while the commission error of the forest extent is 4.5%.

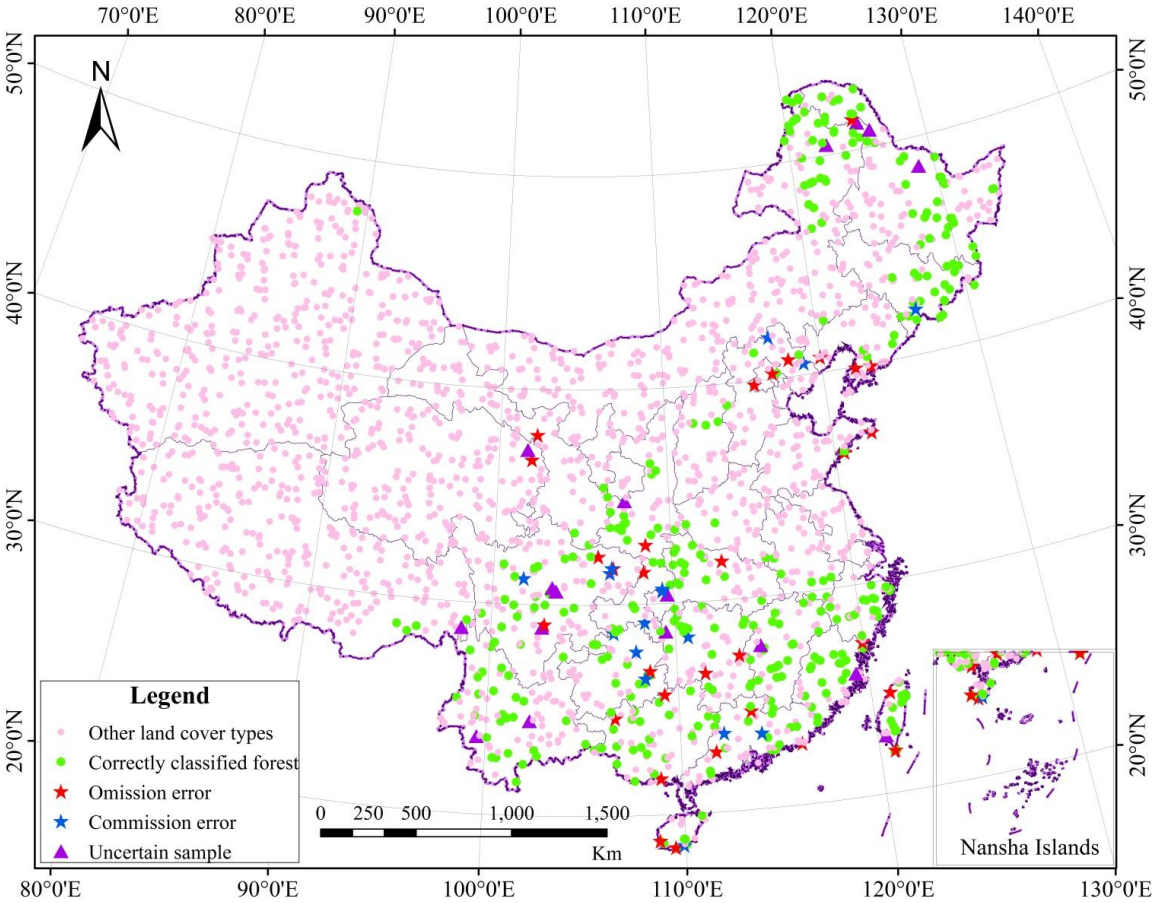
We further explored the main causes of those misclassified sample units (Table 3). Most errors occur at the edges of different land cover types. The labels of those sample units and mixed pixels are ambiguous as indicated by visual interpretation.

Figure 4 shows that most of the errors (omission and commission) are happening in the central region of Southern China, Hebei and Liaoning Provinces. For the central region of Southern China, many images are contaminated by clouds or haze. As a result, the uncertainty of classification increases in these areas. For Hebei and Liaoning Provinces, forests are mixed with shrubs. Their spectral information is similar and difficult to separate. Also, large areas of orchard in these regions increase the classification difficulty.

Table 3. The main causes of the errors.

Errors	Main Causes	Number of Samples
Omission Errors	Mixed pixels	4
	The edge of land cover type	17
	Shadow of mountains	2
	Shadow of clouds or contaminated by clouds	3
	Pepper and Salt	1
	Others	4
Commission Errors	Mixed pixels	1
	The edge of land cover type	8
	Shadow of clouds or contaminated by clouds	1
	Others	6

Figure 4. The distribution of the test samples and errors.



4.2.2. Forest Type Classification Accuracy

For the validation of Level-2 classification results, we used the out-of-bag (OOB) estimate in the Level-2 classification instead of the test sample validation. The reasons are: (1) there are only 388 ± 17 test sample units for forest (averaged about 65 ± 3 samples for each type) for the entire China, insufficient for evaluating the accuracies of each type; (2) As Table 2 shows, many test sample units

are located at the edges of land cover types, because they are randomly selected. The resolution of Level-2 classes is coarser than 30 m. So it is not suitable to validate the Level-2 classification results using these test samples; (3) The OOB error estimate is unbiased [37,41].

Table 4 shows that evergreen broadleaf and evergreen needleleaf forests are classified with relatively high accuracies. Mixed forests are confused with all forest types because the EVI time-series of different mixed species and mixing ratios show different change curves in a year. There are few test samples for deciduous needleleaf forests and bamboos because they are not widely distributed.

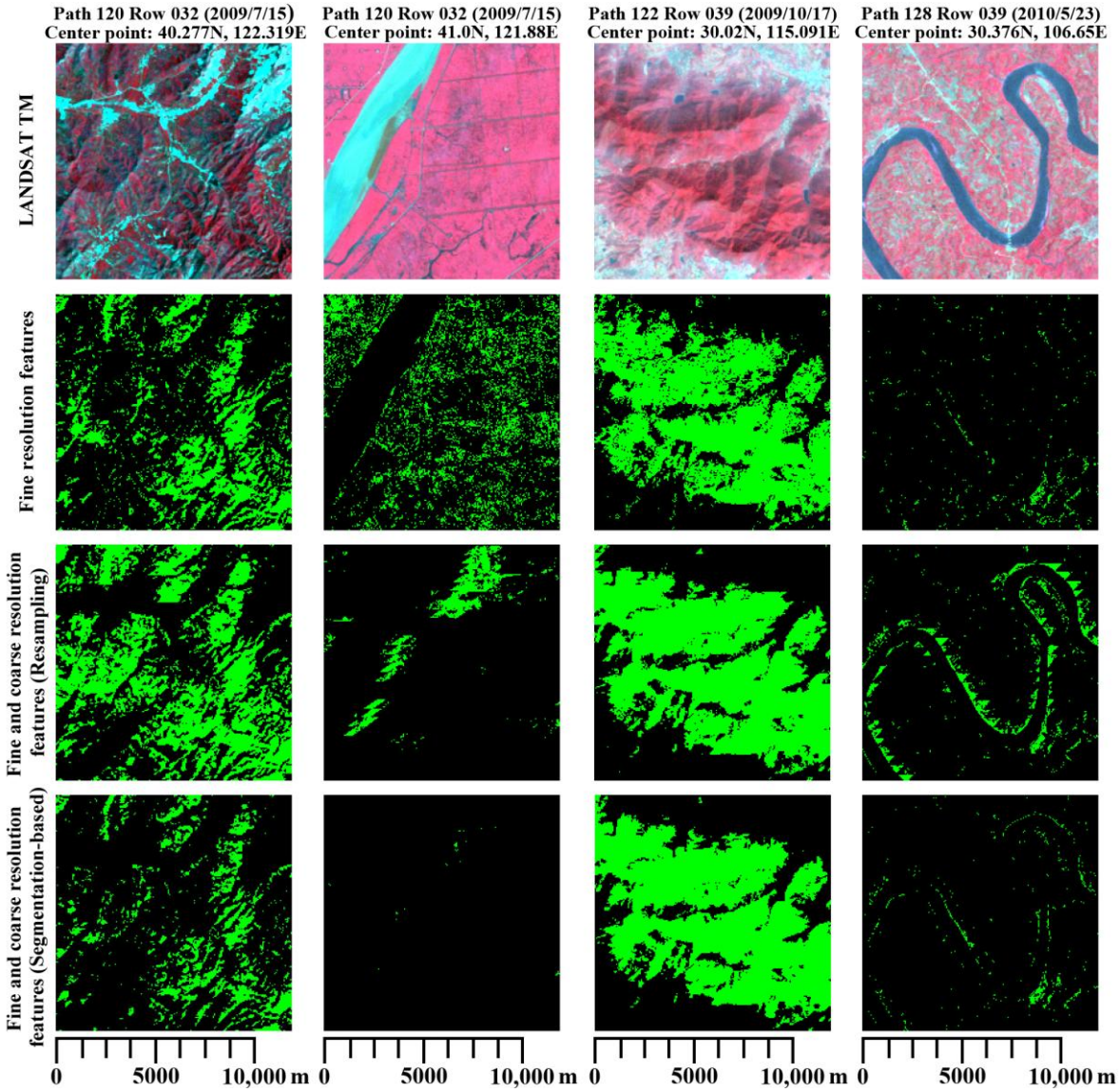
Table 4. The confusion matrix of Level-2 classification based on OOB estimate.

	Ground Truth						
	Evergreen Broadleaf	**Deciduous Broadleaf**	**Evergreen Needleleaf**	**Deciduous Needleleaf**	**Mixed Forests**	**Bamboos**	**Total**
Evergreen Broadleaf	**166**	0	7	0	31	7	211
Deciduous Broadleaf	0	**273**	7	2	90	0	372
Evergreen Needleleaf	7	5	**478**	14	66	8	578
Deciduous Needleleaf	0	1	9	**50**	10	0	70
Mixed forests	30	111	78	29	**437**	16	701
Bamboos	3	0	1	0	5	**28**	37
Total	206	390	580	95	639	59	1969
Overall Accuracy: 72.73%							

4.3. Influences of Resolution

We used thirteen 30 m resolution features, one 90 m resolution feature and eighteen 250 m resolution features as inputs to this forest mapping. A segmentation-based method was applied to combine those multi-resolution features. We averaged EVI values in the same segment. The advantage of this method is that it combines the spectral, spatial information of the TM images and ensures homogeneity of land cover types in the same segment. By so doing, the influences of the coarse resolution images at the edges of the land cover patches have been reduced. Figure 5 shows the influences of different resolution on the classification results where fine resolution features stand for classification results obtained by using the thirteen 30 m resolution features and the 90 m resolution feature. We compared two TM/MODIS combination methods, resampling of MODIS down to TM scale and segmentation-based methods. It can be seen that the classification results based on the segmentation method are constrained by the TM data. They show more details at the edges of land cover types than the resampling method (Path 128 Row 039). Combining TM and MODIS using a resampling method may produce poorer results than using TM features alone (Path 120 Row 032, Path 128 Row 039). Including MODIS features using segmentation method could improve the results without loss of resolution when the land cover types are not separable with TM features only (Path 120 Row 032), or the TM images are contaminated by fog (Path 122 Row 039).

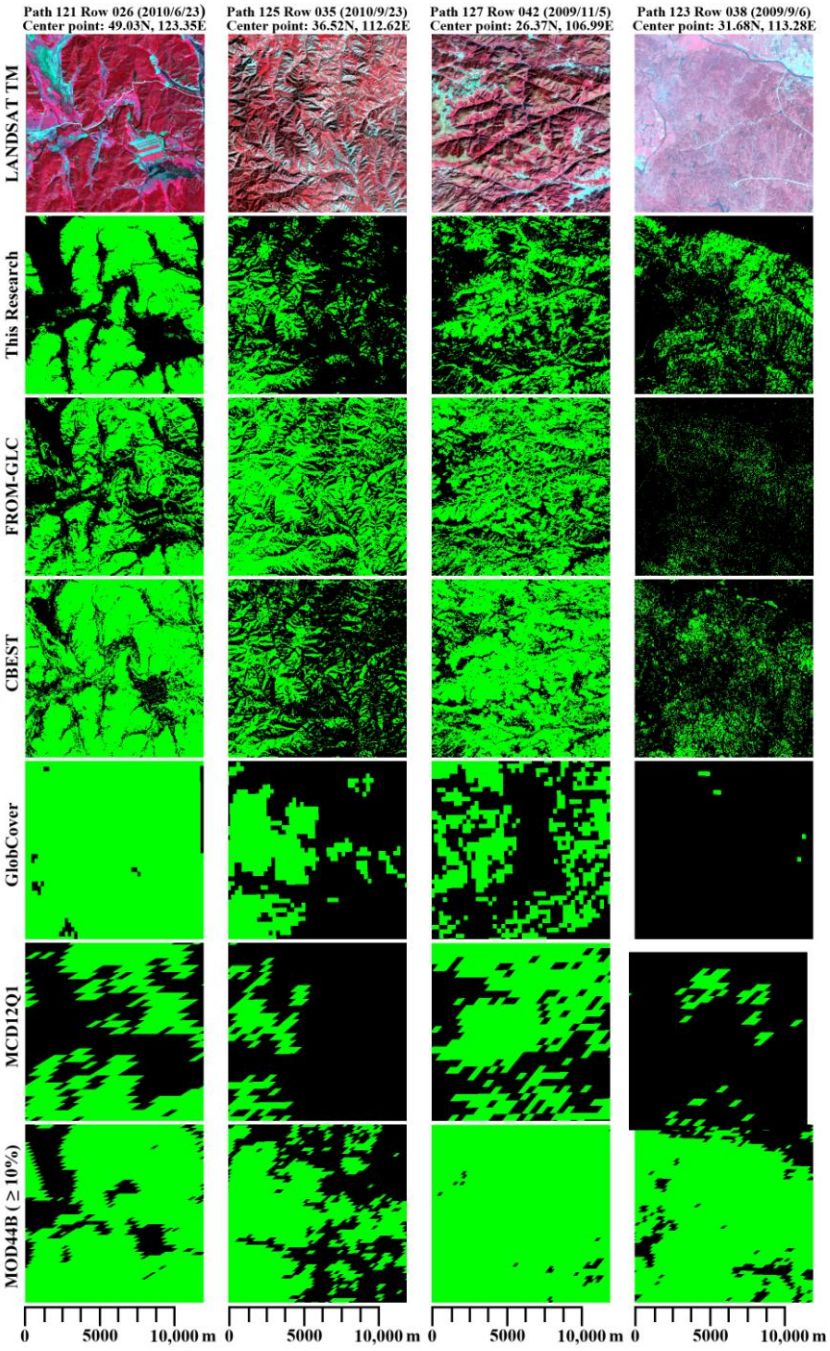
Figure 5. Selected case comparison of the resolution influences (Green color represents forest and black represents non-forest.) From left to right, the image segments were selected from Liaoning Province, Liaoning Province, Hubei Province, and Sichuan Province.



4.4. Comparison with Other Sources

Figure 6 compares the forest extent results with other forest products mentioned in the introduction. As MOD44B provides percentage of tree cover, we assign the pixels with more than 10 percent as forest in this comparison. Because of the coarse resolution, forest extent from Globcover, MCD12Q1 and MOD44B show less detail. We obtained improved forest mapping results by reducing the confusion of forest cover with grass (path 121 row 026), with farmland (path 125 row 035) and with shrubs (path 127 row 042). In these regions, the commission errors are large in most of the other products. The results presented here are less affected by contamination of fog (path 123 row 38). Our results preserved more details of the forest distribution with less obvious misclassifications.

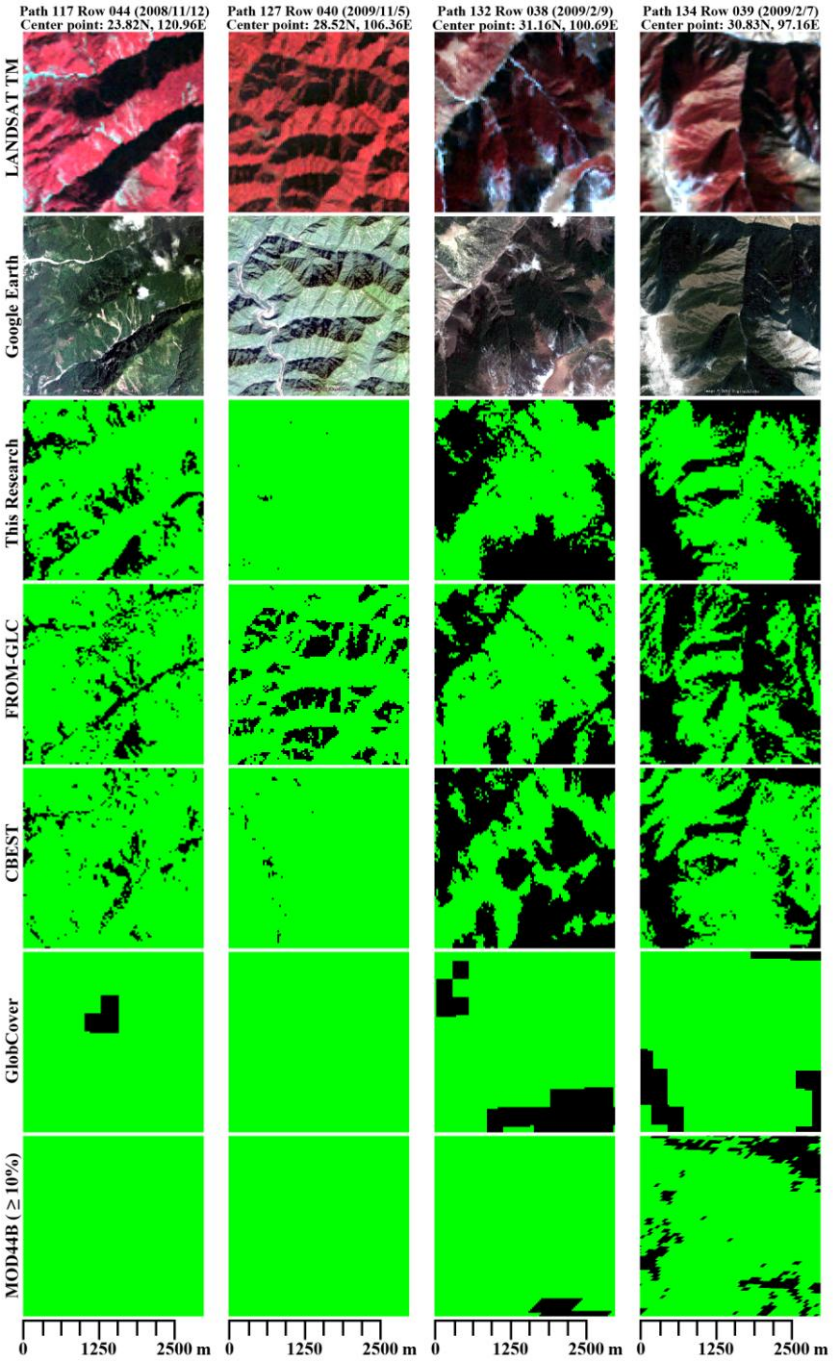
Figure 6. Selected case comparison of forest extent products with this product (Green color represents forest and black represents non-forest.) From left to right, the image segments were selected from Inner Mongolia, Shanxi Province, Guizhou Province, and Hubei Province.



In China, mountainous areas account for 2/3 of the national land area. The surface reflectance is different with different solar elevation and azimuth, slope and aspect. Thus, the quality of classification results will be influenced. Some algorithms were produced to reduce the topographic effect [42]. However, we did not do topographic correction. Topographic effects are not strong for Landsat images due to their ±7.5° viewing angle off nadir. There is still spectral information in shaded areas, particularly for the TM data. During the training sample collection process, we treated the forests distributed on sunny slopes and shaded slopes differently. Figure 7 shows the forest extent results in

selected mountainous areas. We only compared one MODIS product—MOD44B, because the resolution of MCD12Q1 is too coarse for this extent (3 km × 3 km). The shaded surface is darker than the sunlit surface. It can be seen that both forest in the shaded areas (path 117 row 044; path 127 row 040) and non-forest in the shaded area (path 132 row 038; path 134 row 039) are better identified by the 30 m forest map described here. The effect of shaded slopes seems to have little effect on the final image classification results.

Figure 7. Selected case comparison of forest extent products with this product in mountainous areas (Green color represents forest and black represents non-forest.) From left to right, the image segments were selected from Taiwan, Sichuan Province, Sichuan Province, and Tibet.



5. Conclusions

In this study, we presented a workflow to integrate Landsat TM images, MODIS images and auxiliary datasets and derived features to map forest in China at 30 m resolution. This map can be used as a base line for the researchers to understand the current state and for governmental agencies to understand changes of forest in a spatially explicit way; and help to develop forest management programs subsequently. Additionally, it could be helpful to understand the suitable afforestation conditions. It also could be used as ancillary data for biogeochemical cycling and climate change studies. For the forest/non-forest extent classification, the post-processing is very important. We modified the incorrectly classified images through adding samples and doing reclassification. This step would ensure the quality of classification results. Both the producers' accuracy and the user's accuracy of forest exceeded 90% based on test sample data. For forest type classification, the OOB estimate of overall accuracy is 72.7%. The forest extent area estimated from this map is close to the forest area of 7th NFRI using the forest definition defined here. Despite an exhaustive optimal search, 11% of the Landsat scenes were acquired before 2008 and some scenes were covered by heavy clouds. Future improvements can be made by making use of multi-temporal data composites, inclusion of Chinese satellite imagery (*i.e.*, HJ data, Gaofen satellite data) or fully utilizing the information in overlapping areas between scenes. SAR data or lidar data may be helpful to reduce confusion between classes [43]. It would be beneficial to compare our results with the national survey data in a more spatially-explicit manner but the survey data are not available in wall-to-wall mapping for China. For now, the forest extent map presented here is the most recent, high-resolution forest cover map of China with a relatively complete accuracy report. It would also be desirable to use the national survey data to validate this map but they are not currently accessible. Finally, more effort will be done on the forest change detection. Such information will be helpful for future afforestation planning and change monitoring.

Acknowledgments

This work was partially supported by the National High Technology Research and Development Program of China (2013AA122804), and a research grant from Tsinghua University (2012Z02287). Initial training samples were interpreted with assistance by Xiaoyi Wang, Caixia Liu and Wenyu Li at Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences.

Author Contributions

Peng Gong and Congcong Li conceived and designed this research. Congcong Li did most of the analysis. All authors contributed extensively to data collection and writing of this paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Food and Agriculture Organization of the United Nations. *Global Forest Resources Assessment 2010*; Main Report; Food and Agriculture Organization of the United Nations: Rome, Italy, 2010.
2. Townshend, J.; Masek, J.G.; Huang, C.Q.; Vermote, E.F.; Gao, F.; Channan, S.; Sexton, J.O.; Feng, M.; Narasimhan, R.; Kim, D.; *et al.* Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *Int. J. Digit. Earth* **2012**, *5*, 373–397.
3. Department of Forest Resources Management. The 7th national forest inventory and status of forest resources. *Forest. Resour. Manag*. **2010**, *1*, 1–8. (In Chinese)
4. The Central People's Government of People's Republic of China. Available online: http://www.gov.cn/wszb/zhibo416/ (accessed on 27 May 2014).
5. Gong, P. Remote sensing of environmental changes over China: A review. *Chin. Sci. Bull*. **2012**, *57*, 2793–2801.
6. Liu, J.Y.; Kuang, W.H.; Zhang, X.Z.; Xu, X.L.; Qin, Y.W.; Jia, N.; Zhou, W.C.; Zhang, S.W.; Li, R.D.; Yan, C.Z.; *et al.* Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s. *Acta Geogr. Sin.* **2014**, *69*, 3–14. (In Chinese)
7. Zhang, Z.X. *Monitoring Land Cover of China Using Remote Sensing*; Planet Map Publishing House: Beijing, China, 2010. (In Chinese)
8. Hu, L.; Chen, Y.; Xu Y.; Zhao, Y.; Yu, L.; Wang, J.; Gong, P. 30 m land cover mapping of China with an efficient clustering algorithm CBEST. *Sci. China Earth Sci*. **2014**, in press.
9. Bishop, C. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2007.
10. Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sens.* **2014**, *6*, 964–983.
11. Bontemps, S.; Defourney, P.; van Bogaert, E.; Arino, O. GLOBCOVER2009 Products Description and Validation Report. Available online: http://due.esrin.esa.int/globcover/LandCover2009/GLOBCOVER2009_Validation_Report_2.2.pdf (Accessed on 20 January 2012).
12. Di Gregorio, A.; Jansen, L.J.M. *Land Cover Classification System (LCCS). Classification Concepts and User Manual* 2000. Available online: http://www.fao.org/docrep/003/x0596e/x0596e00.HTM (accessed on 27 May 2014).
13. Friedl, M.A.; Mciver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; *et al.* Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ*. **2002**, *83*, 287–302.
14. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X.M. MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ*. **2010**, *114*,168–182.
15. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.C.; Zhao, Y.; Liang, L.; Niu, Z.G.; Huang, X.M.; Fu, H.H.; Liu, S.; *et al.* Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654.
16. Yu, L.; Wang, J.; Gong, P. Improving 30 m global land cover map FROM-GLC with time series MODIS and auxiliary datasets: A segmentation based approach. *Int. J. Remote Sens*. **2013**, *34*, 5851–5867.

17. Yu, L.; Wang J.; Li, X.C.; Li, C.C.; Zhao, Y.Y.; Gong, P. From-Hierarchy, a multi-resolution global land cover data set. *Sci. China Earth Sci.* **2014**, in press.
18. Townshend, J.R.G.; Carroll, M.; Dimiceli, C.; Sohlberg, R.; Hansen, M.; DeFries, R. *Vegetation Continuous Fields MOD44B, 2000–2010 Percent Tree Cover, Collection 5*; University of Maryland: College Park, MD, USA, 2010.
19. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *342*, 850–853
20. Food and Agriculture Organization of the United Nations (FAO). *Global Forest Resources Assessment 2010: Terms and Definitions*; 2010. Available online: http://www. fao.org/docrep/014/am665e/am665e00.pdf (accessed on 27 May 2014).
21. Belward, A. *The IGBP-DIS Global 1 km Land Cover Data Set "DISCover": Proposal and Implementation Plans*; IGBP-DIS Working Paper 13, International Geosphere Biosphere Programme; European Commission Joint Research Center: Varese, Italy, 1996.
22. GLOVIS: The USGS Global Visualization Viewer. Available online: http://glovis.usgs.gov/ (accessed on 27 May 2014).
23. Open Spatial Data Sharing Project RADI. Available online: http://ids.ceode.ac.cn/query.html (accessed on 27 May 2014)
24. LP DAAC. Available online: https://lpdaac.usgs.gov/products/modis_products_table/mod13q1 (accessed on 14 April 2014)
25. Huete, A.; Justice, C.; van Leewen, W. MODIS Vegetation Index (MODIS13) Algorithm Theoretical Basis Document Version 3. Available online: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf (access on 30 April 1999).
26. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ*. **2002**, *83*, 195–213.
27. Miura, T.; Huete, A.R; van Leeuwen, W.J.D.; Didan, K. Vegetation detection through smoke-filled AVIRIS images: An assessment using MODIS band passes. *J. Geophys. Res.*, **1998**, *103*, 32001–32011.
28. CGIAR-CSI SRTM 90 m Digital Elevation Data Database. Available online: http://srtm.csi.cgiar.org/ (accessed on 19 August 2008).
29. Mon, M.S.; Mizoue, N.; Htun, N.Z.; Kajisa, T.; Yoshida, S. Estimating forest canopy density of tropical mixed deciduous vegetation using Landsat data: A comparison of three classification approaches. *Int. J. Remote Sens.* **2012**, *33*, 1042–1057.
30. Kauth, R.J.; Thomas, G.S. The Tasseled Cap-Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat. In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, IN, USA, 1976; pp. 4b41–4b51.
31. Crist, E.P.; Cicone, R.C. A physically-based transformation of Thematic Mapper data—The TM tasseled cap. *IEEE Trans. Geosci. Remote Sens*. **1984**, *22*, 256–263.

32. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.H.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series dataset based on the Savitzky-Golay filter. *Remote Sens. Environ*. **2004**, *91*, 332–344.
33. BIS Cloud. Available online: http://www.imageseg.com (accessed on 27 May 2014).
34. Clinton, N.; Holt, A.; Scarborough, J.; Yan, L.; Gong, P. Accuracy assessment measures for object based image segmentation goodness. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 289–299.
35. Finer Resolution Observation and Monitoring—Global Land Cover. Available online: http://data.ess.tsinghua.edu.cn (accessed on 14 March 2014).
36. Zhao, Y.; Yu, L.; Hu, L.Y.; Li, X.Y.; Li, C.; Zhang, H.Y.; Zheng, Y.M.; Zhao, Y.C.; Cheng, Q.; Liu, C.X.; *et al.* Towards a common validation sample set for global land cover mapping. *Int. J. Remote Sens*. **2014**, in press.
37. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
38. Landsat Bulk Metadata Service. Available online: http://Landsat.usgs.gov/consumer.php (accessed on 27 May 2014).
39. China's State Forestry Administration. *China Forest Resources Report—The 7th National Forest Resource Inventory*. Available online: http://wenku.baidu.com/link?url=yCgzhDhX0OFdntS4s-SMeyQwWYzaneh10tnU9br0pIVQCsakGH6Ogqey-uCdlrjMF6Z5TOO1FX3cO5oo300Gmf5QZd1pTdgK44BVaAcBMz3 (accessed on 27 May 2014).
40. The Central People's Government of People's Republic of China. Available online: http://www.gov.cn/zxft/ft205/content_1695019.htm (accessed on 27 May 2014).
41. Tibshirani, R. *Bias, Variance, and Prediction Error for Classification Rules*; Technical Report; University of Toronto: Toronto, ON, Canada, 1996.
42. Huang, H.; Gong, P.; Clinton, N.; Hui, F. Reduction of atmospheric and topographic effect on Landsat TM data for forest classification. *Int. J. Remote Sens*. **2008**, *29*, 5623–5642.
43. Waske, B.; van der Linden, S. Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Trans. Geosci. Remote Sens*. **2008**, *46*, 1457–1466.