Continuous tree distribution in China: A comparison of two estimates from Moderate-Resolution Imaging Spectroradiometer and Landsat data

Ronggao Liu,[1] Shunlin Liang,[2] Jiyuan Liu,[1] and Dafang Zhuang[1]

Received 2 April 2005; revised 27 October 2005; accepted 5 December 2005; published 18 April 2006.

[1] Forest change is a major contributor to changes in carbon stocks and trace gas fluxes between terrestrial and atmospheric layers. This study compares two satellite estimates of percent tree distribution data sets over China. One estimate is from the Chinese National Land Cover Data Set (NLCD) generated by a multiyear national land cover project in China through visual interpretation of Landsat thematic mapper (TM) and the Enhanced Thematic Mapper Plus (ETM+) images primarily acquired in the year 2000. The other estimate is the Moderate-Resolution Imaging Spectroradiometer (MODIS) standard product (MOD44B) from the same year. The two products reveal some common features, but significant discrepancies exist. Detailed analyses are carried out with different land cover types and over different regions. Comparison results show that the difference of the total tree canopy area for the whole country is 159,000 km^2. The pixel counts in the NLCD data set for dense forest are ~4 times those in the MODIS data set with the reverse holding for sparse forest. Generally, the percent tree canopy area of the NLCD data set is larger in eastern China and lower in the Tibetan plateau margin region. For different land cover types the percentage of tree canopy areas shows a good agreement for evergreen forests but a large discrepancy for deciduous forests. The largest variations are associated with grassland and nonvegetation classes. Regarding the spatial distributions of their differences, Inner Mongolia is the place where both data sets show a diverse result, but Guizhou and Fujian present the least divergence among those provinces with the tree canopy area being more than 20,000 km^2.

Citation: Liu, R., S. Liang, J. Liu, and D. Zhuang (2006), Continuous tree distribution in China: A comparison of two estimates from Moderate-Resolution Imaging Spectroradiometer and Landsat data, *J. Geophys. Res.*, *111*, D08101, doi:10.1029/2005JD006039.

1. Introduction

[2] Forests, the core of terrestrial ecosystems, not only provide habitats and food for animals and fiber and fuel for human beings, but they also control global climate and biogeochemical cycles [*Sellers et al.*, 1997]. Forest variations are the major contributor to changes in carbon stocks and trace gas fluxes between terrestrial and atmospheric layers [*Houghton*, 1999]. Since the continuous increase of atmospheric carbon dioxide derived from fossil fuel burning is considered as the main cause of global warming, and forests can remove carbon dioxide from the atmosphere and store it in wood through photosynthesis, the Kyoto protocol appeals to the participant members to partake in the ''reforestation and afforestation'' campaign to improve the carbon sequestration of terrestrial ecosystem so as to mitigate the global warming trend.

[3] China, one of the most populous and the longest-cultivated civilization in history, is the most rapidly developing economic regions in the last twenty years. Owing to its government's reforestation and afforestation policies since the late 1950s, China is believed to be a major location of forest carbon sinks today [*Fang et al.*, 2001; *Streets et al.*, 2001]. However, the estimates of the magnitude of China's terrestrial carbon sink vary considerably, from net sink of 0.1 PgC yr^{-1} [*Streets et al.*, 2001] to 0.02 PgC yr^{-1} [*Fang et al.*, 2001]. Understanding the carbon sequestration potential and its variation responses to climate variability in China is also a challenge [*Cao et al.*, 2003]. The forest distribution information is a precondition for reducing these estimates' uncertainties.

[4] Satellite remote sensing technology provides a powerful and independent approach to estimate area extent and spatial distribution of forests at the regional and global scale. Coarse-resolution images have been used to automatically map land cover from the regional to continental scales over the last few decades. Examples include the data from Advanced Very High Resolution Radiometer (AVHRR) [*Townshend*, 1994], Moderate-Resolution Imaging Spectroradiometer (MODIS) [*Friedl et al.*, 2002] and SPOT VEGETATION (VGT) [*Bartalev et al.*, 2003]. High-

[1]Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.
[2]Department of Geography, University of Maryland, College Park, Maryland, USA.

0148-0227/06/2005JD006039$09.00

resolution images, such as those from Landsat Enhanced Thematic Mapper (TM) [*Liu*, 1996] and IKONOS [*Goetz et al.*, 2003], have also been tasked to automatically or manually map regional land cover and land use. However, land cover mapping according to predefined classification schemes has several disadvantages [*DeFries et al.*, 1995, 1999; *Fernandes et al.*, 2004]: (1) The approach does not fully utilize the information content of remotely sensed data to describe gradients and mosaics in the landscape, (2) the heterogeneous characteristics in the same land cover are removed, (3) comparing vegetation maps from different classification schemes is difficult, and (4) the approach does not offer a means to estimate the land cover change over large areas with high-repetitive-frequency observations. Since most of China's forests are distributed over mountainous terrain, with very fragmental and heterogeneous nature as a result of the long historically anthropogenic land cover changes, these disadvantages are particularly detrimental for understanding Chinese forest characteristics [*Xiao et al.*, 2003]. Continuous percent forests, which describe the details of forest distribution characteristics, may be more beneficial to biogeochemical cycle modeling [*Liang*, 2003; *DeFries et al.*, 2002].

[5] There are two continuous percent tree data sets available over China. One is the 1 km $\times$ 1 km National Land Cover Data Set's (NLCD) continuous land cover data (hereafter referred to as NLCD data) derived from an interpretation of Landsat TM/ETM. The other is global percent tree data from NASA MODIS MOD44B standard product, developed by the University of Maryland's Department of Geography using a pixel unmixing algorithm [*Hansen et al.*, 2003, 2002a] (hereafter referred to as UMD data). Since the percent tree data over large regions are very rare (though it is important to evaluate them), there are only a few related works available to validate them. The objective of this study is to compare the two percent tree distribution data sets available over China derived from different methods. It attempts to (1) improve the understanding of tree distribution of China with respect to the spatial distribution variation from the different approaches and (2) identify any common features and differences between the UMD data and NLCD data through comprehensive analyses of the products over some special regions and with some land cover types. The comparison at the pixel and national scales can help quantify the uncertainties of tree distribution estimates over China from both high-resolution Landsat images and coarse-resolution MODIS data.

[6] The data sets employed are introduced in section 2 of the paper. Regional characteristics and variations of tree distribution are analyzed in section 3. The conclusion and discussion are given in section 4.

2. Data and Method

2.1. Forests of China

[7] China has a total area of 9.6×10^6 km^2 and a west-east gradient in land terrain with numerous mountain ranges, high plateaus, uplands, basins, and plains. The temperature decreases from south to north with the increase of latitude. The precipitation primarily concentrates on the eastern portions of China. The western region is dominated by arid or semiarid land. These characteristics make the ecosystem very variable. The forests are mainly distributed in the eastern section of the country. From north to south, the natural ecosystems from the boreal forests in the northeastern provinces, through the deciduous forests, to the mixed evergreen and deciduous subtropical forests in the mountainous region of central China and the evergreen tropical forests in the south [*Hou*, 1983]. With less precipitation and higher elevations in the western portions, the forests are replaced by steppe land and desert ecosystems [*Houghton and Hackler*, 2003].

[8] The current forest area is about 1.4×10^6 km^2 [*DeFries et al.*, 2002; *Fang et al.*, 2001], and the Chinese Agency of Forest declared the forest area as being 1.75×10^6 km^2 in January 2005. By 1996, 2.06×10^6 km^2 of the afforestation had been finished. However, the predisturbed forest area was estimated as 4.19×10^6 km^2, which is more than twice of that which remains today. Much of the loss occurred before 1700 [*Houghton and Hackler*, 2003], which could be ascribed to the highly developed agriculture of that time. It is reported that the forest area of China reached its minimum in 1949, and that, since the beginning of the People's Republic of China, forest area has increased as a result of massive afforestation campaigns. The planting of trees became a national objective in 1949, with an announced goal of returning 30% of the country to forestland. The program has been referred to as the greatest land use change project of all times, which mainly focused on afforestation for the protection of forests, fuelwood plantations, and economic plantations [*Li et al.*, 1999]. However, since the survival rate was very low in the early days of afforestation [*Richardson*, 1990] and the rapid development of New China requires more agricultural land and woodland, the national survey data show that the minimum forest area was not in 1949 but during 1977–1981 [*Fang et al.*, 2001].

2.2. UMD Continuous Tree Data Set

[9] The UMD continuous percent tree data in this study were produced from MODIS data between 31 October 2000 and 9 December 2001 [*Hansen et al.*, 2003] (the data are archived at http://redhook.gsfc.nasa.gov/~imswww/pub/imswelcome). The MOD44B Version 3 product, available from EROS data center, is delivered as a set of HDF-EOS files divided into the standard MODIS tiles in the Integerized Sinusoidal projection, which is an equal area projection. Each tile, which covers a spatial extent of 10 degrees of latitude by 10 degrees of longitude at the equator, contains 2400 samples by 2400 rows for 500-m resolution data. The product includes proportions of woody vegetation, herbaceous vegetation, and bare ground. The product was initially produced by a regression tree algorithm, and the outputs from the regression tree were then further modified by a stepwise regression and bias adjustment. The seven MODIS land bands were used as inputs: blue (459–479 nm), green (545–565 nm), red (620–670 nm), near infrared (841–876 nm) and three middle infrared (1230–1250 nm, 1628–1652 nm, 2105–2155 nm). The MODIS composite data were transformed into annual metrics that capture the salient points in the phenological cycle. A total of 68 metrics were derived from the composite data for the seven bands and the NDVI, which are used

Table 1. Hierarchical Land Cover Classification System of Chinese National Land Cover Data Set

Level I	Level II Class	Description
Croplands	paddy	crop-producing fields with enough water for irrigation
Croplands	nonflooded cropland	crop-producing fields without or that do not need water for irrigation
Forests	dense forest	natural and artificial forest with area >0.005 km^2 and tree canopy cover >30%
Forests	sparse forest	sparse forest with tree canopy cover >10% and <30%
Forests	shrub	tree canopy cover >40% but canopy height <4 m
Forests	other forest	plantation forest with age >3 years and <5 years
Grasses	dense grass	grass coverage >50%
Grasses	moderately dense grass	grass coverage >20% and <50%
Grasses	sparse grass	grass coverage >5% and <20%
Water bodies	river	natural or man-made river with water throughout the year
Water bodies	lake	natural reservoir with water throughout the year
Water bodies	reservoir	man-made reservoir with water throughout the year
Water bodies	snow and ice	land under snow/ice cover throughout the year
Water bodies	ocean beach	land under water cover while tiding
Water bodies	lake and river beach	land under water cover while flooding
Built-up land	urban	city and town built-up land
Built-up land	construction land	industrial built-up land outside of urban regions
Built-up land	other built-up land	country residential site, road, and airport
Unused land	sandy desert	land under sand cover and vegetation <5%
Unused land	harsh desert	land under gravel cover and vegetation <5%
Unused land	salinification	salinificated soil land with very sparse vegetation
Unused land	wetland	land with a permanent mixture of water and herbaceous or woody vegetation
Unused land	bare soil land	land under soil cover and very sparse vegetation
Unused land	bare rock land	rocks cover >50%
Unused land	other unused land	e.g., tundra

as the inputs for estimating tree cover percentages. The approach to deriving the metrics and training data is fully described by *Hansen et al.* [2002a]. The products have been tested at a woodland site, Western Province, Zambia [*Hansen et al.*, 2002b].

2.3. NLCD Continuous Percent Tree Data Set

[10] In the early 1990s, the Chinese Academy of Sciences organized its eight research institutions, which are located in different Chinese provinces, and about 100 scientists from the disciplines of agriculture, forestry, and geography to conduct Chinese national land cover and land use classification projects using satellite data. The goals of the projects were mainly focused on understanding the status of Chinese land use and land cover to support governmental decision-making. Therefore the land cover classification system was designed to consider these factors: (1) It should be as simple as possible to be used in land surveying and monitoring throughout the country, (2) it should represent the physical characteristics of land as a resource, (3) the land classes can be distinguished easily by Landsat TM/ETM image, and (4) the land classes represent the land change characteristics. To achieve these goals, a hierarchical land cover classification system was used, consisting of six classes at Level I and 25 classes at Level II (Table 1), including four forest types at Level II classes: dense forest, sparse forest, shrubs and other forests. With the diligent work spanning more than a decade, the Chinese National Land Cover/Use Data Sets (NLCD), which cover the whole country over three periods in the 1980s, medium 1990 and 2000 were created. About 520 TM images were used during each period. The images were georeferenced using field-collected ground control points and high-resolution digital elevation models, with an average horizontal error of 50 m. The identification of spectral characters of TM/ETM images for different land cover types was based on extensive field surveys. For example, more than 7,900 field photos were taken, using cameras equipped with global positioning system receivers, totaling about 75,000 km of transects across China. The 1:100,000-scale Chinese relief maps, surveyed by Chinese Mapping Agency in the early 1980s, which contain the land use information, were used as supplemental materials for interpretation. A combination of these two different sources of information, the land cover thematic data at a scale of 1:100,000 were generated by the local scientists. The validation results showed that the overall accuracy of the land cover classification is about 98.7% [*Liu*, 1996]. The 1:100,000-scale vector land cover data was then converted into a 1-km cell database by calculating the fractional percentages of each cover within the 1-km grid cell, which still contains all of the high-resolution land cover distribution information in each cell [*Liu et al.*, 2001]. The NLCD fractional data sets include 25 layers, each layer corresponding to one land class of Level II. The data sets were archived in the database system of the Data Center for Resources and Environment Sciences (http://www.resdc.cn), Chinese Academy of Sciences. They can be provided to users as ESRI Arc/Info or Geotiff format with geographical coordination or Albers Equal Area projection by CD-ROM.

2.4. Land Cover Classification Data

[11] The land cover classification data was used as background data for comparison of these two continuous percent tree data. There are several land cover data sets available covering all of China, such as the GLC 2000, MODIS MOD12Q1 product. In this study, a regional land cover classification data set, NLCD land cover classification data set (NLCD LCC), from the year 2000 was used

because it has been validated extensively in China [*Liu et al.*, 2003]. NLCD LCC was produced from the classification of one year of AVHRR composite data and geophysical data sets. Before classification, China was divided into nine climatic regions, on the basis of the mean climate data for ten years. For each region, the training data were selected from Landsat TM and survey maps, independently. The 1-year 10-day composite AVHRR band 1, band 2 and the derived NDVI, plus annual mean temperature, annual precipitation, and elevation were classified by a supervised classification algorithm to generate land cover maps for individual regions. Then, the nine land cover maps for each region were assembled together [*Liu et al.*, 2003]. Although this may produce inconsistent boundaries of land cover types among different regions, the local training data sample may be more suitable for the Chinese complex conditions. The land cover data set, derived from Landsat TM, was used to assess the classification accuracy. The data set consists of eighteen land cover classes: evergreen needleleaf forest, deciduous needleleaf forest, evergreen broadleaf forest, deciduous broadleaf forest, mixed forest, alpine forest, shrub, dense grassland, moderate dense grassland, sparse grassland, cropland, wetland, city, water body, ice and snow, harsh desert, sandy desert, and bare rock. Detailed definitions of these classes were given by *Liu et al.* [2003].

2.5. Data Preprocessing

[12] All the data were collected and then projected to an Albers conic equal-area map projection using bilinear interpolation. The part of the UMD global percent tree data, outside of the Chinese administrative boundary, was masked. The NLCD fractional land cover data sets include four layers of forest data: dense forest, sparse forest, shrub, and other forests (Table 1); whereas, the UMD data set is composed of only one tree layer. To make the two data sets comparable, the four forest layers in the NLCD data sets must be merged into one tree layer. Because the forest definitions in the two land cover classification systems are different, it is impossible to make the two data sets completely consistent with one another. From the forest definition of the NLCD data sets for different classes during Landsat TM/ETM interpretation (Table 1), it is assumed that the tree cover is 100% with dense forest, 30% with sparse forest, 60% with shrub and 60% with other forest. Therefore the aggregation formula from the four NLCD forest layers to one tree layer was set as

$$\begin{aligned}\text{Tree Layer} = {} & \text{Dense forest} * 1.0 + \text{Sparse forest} * 0.3 \\ & + \text{Shrub} * 0.6 + \text{Other forest} * 0.6\end{aligned}$$

After aggregation, it is assumed that the tree layer of the NLCD data sets is equivalent to the UMD data in terms of forest definition.

[13] Apart from the difference in definition, the NLCD tree data represent the percent crown cover, but the UMD tree data represent percent tree canopy, which is the amount of skylight obstructed by tree canopies equal to or greater than 5 m in height (crown cover = canopy cover + within crown skylight) [*Hansen et al.*, 2003]. Conversion of the two data sets for the purpose of comparison is needed. To establish the relationship between the percent crown cover and the percent canopy cover, fieldwork was performed in which both crown and canopy cover values were measured by UMD scientists. Initial work by *Hansen et al.* [2003] suggests that the mean forest label used in deriving canopy cover (80 percent) training data corresponds to a 100 percent forested area in terms of crown cover. Their results show that different tree types have different relationships between canopy and crown cover. For example, subalpine fir from four sites in Colorado revealed a 0.9 ratio of canopy to crown cover. However, broadleaf Kalahari woodland trees in Western Zambia have a 0.76 ratio [*Hansen et al.*, 2003]. It is suggested that it is suitable to multiply by 0.8 to derive canopy cover from crown cover. In this study, the NLCD aggregated percent tree was converted to percent tree canopy cover by multiplication of 0.8 as suggested by *Hansen et al.* [2003], which is hereafter referred to as NLCD percent tree canopy. After these procedures, we assume the forest data in two data sets are consistent.

[14] Some classes in the NLCD land cover data are regrouped into new types: dense grassland, moderate dense grassland, sparse grassland, cropland, and wetland are merged into a grassland class; city, water body, ice and snow, harsh desert, sandy desert, bare rock are merged to a nonvegetation class. The rest of the classes remain unchanged. As a result, the new classification types consist of seven forest types plus grassland and nonvegetation land.

3. Results

3.1. Forest Distribution of the Whole Country

[15] For the NLCD fractional forest data set, the forest area is calculated from four forest layers. Their spatial distribution can be found in Figures 1a–1d. The pixel counts, area and distribution region for different forest classes are listed in Table 2. From the four forest layers, the total tree crown area can be calculated using the aggregation formula. There are 1.741×10^6 km^2 tree crown in 3.803×10^6 pixels in the new aggregation layer. Because we assume the new tree layer in NLCD data sets represents tree crown but that in UMD data set is tree canopy, the canopy cover area in NLCD data sets, which was calculated from the new NLCD tree layer multiplying 0.8, is 1.392×10^6 km^2. The fractional tree canopy distribution map for the aggregated layer is shown in Figure 1e. The figure shows that the forests are mainly distributed in south China and the mountainous regions of northeastern China, with a few spots in the arid, semiarid or cold regions of western China and in the cropland of the northeastern plain. In the UMD percent tree data set, there are 1.233×10^6 km^2 of tree canopy cover in 5.158×10^6 pixels (Figure 1f). The tree canopy area in the NLCD data sets is 0.159×10^6 km^2 larger than that in the UMD data sets, with the relative difference about 12.0%. From Figures 1e and 1f, we can find that, on the whole, the spatial distribution patterns of forest in the two data sets are visually similar, which also corresponds well with the forest pattern of the NLCD LCC data set (Figure 2).

[16] The pixel frequency counts for different percentages of tree canopy are shown in Figure 3. It is found that the pixel frequencies of tree canopy percentage in the two data sets are similar for fractions between 10% and 70%; the

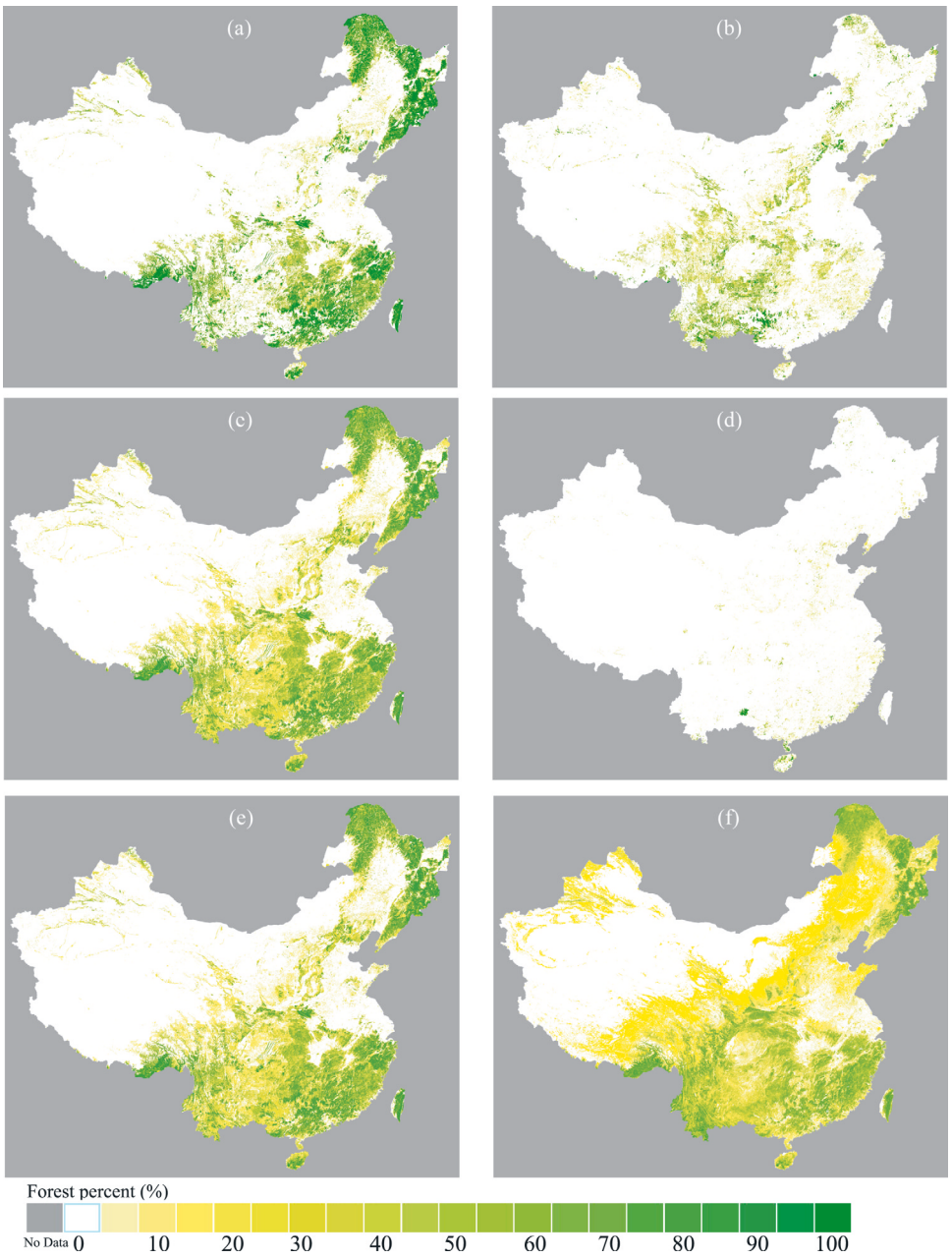

Figure 1. A comparison of tree percentages between the NLCD and UMD data sets at 1-km resolution in 2000. (a) Dense forest layer of NLCD data set, (b) sparse forest layer of NLCD data set, (c) shrub layer of NLCD data set, (d) other forest layer of NLCD data set, (e) aggregation forest layer of NLCD data set, and (f) UMD forest data set.

relative differences ((UMD counts minus NLCD counts)*2/(UMD counts + NLCD counts)*100%, hereafter referred to as the same definition) are 2.1%, −10.8%, 24.7%, 25.1%, 42.8% and 33.5%, respectively. For very dense forest (canopy cover more than 70%), the pixel counts in the NLCD data are more than 3 times that in the UMD data, but for sparse forest with tree canopy less than 10%, this is reversed. It can be explained that, for the NLCD data, the Landsat TM pixels with canopy cover more than 30% are classified as dense forest, yet the nonforest area in these pixels could lead to an overestimation of tree distribution. However, for the very sparse forest region, the Landsat TM pixels with canopy cover less than 10% are not classified as forest in the NLCD data, although there are still some trees in these pixels, which will lead to an underestimate of tree percentage. The pattern can be found in Figures 1e and 1f. In the grassland region, there is a great deal of sparse forest distribution in the UMD data set but little in the NLCD data set.

[17] To locate and enhance the tree spatial distribution agreement/disagreement and to determine the most uncertain regions between the two data sets, the NLCD data are subtracted by the UMD data grid cell by grid cell, and the absolute value is then calculated. If the pixel value of the

Table 2. Forest Area of NLCD Data Set for Different Land Cover Classes

Land Class	Pixel Count	Area	Main Distribution Region
Dense forest	2.407×10^6	1.343×10^6 km^2	northeastern China, eastern Tibetan plateau, and southern China
Sparse forest	1.378×10^6	0.492×10^6 km^2	southwestern China
Shrub	1.244×10^6	0.370×10^6 km^2	central China
Other forest	0.209×10^6	0.047×10^6 km^2	southern and northeastern China

UMD data is larger, it is rendered as red, otherwise as blue. If there is no difference in the pair of grid cells, the result is zero and is rendered as white. A new difference map was produced (Figure 4), which shows that, in general, the tree canopy percentages of the UMD data are larger than those of the NLCD data in the southwestern region of China, such as Yunnan, Sichuan, Chongqing, and Shannxi provinces; but in the eastern region, the trend is reversed. There is only one exception: Fujian province is located in southeastern China in which tree percentages in the UMD data are larger than those in the NLCD data and are enclosed by the administrative boundary. This may reflect the errors of the visual interpretation of Landsat TM data because the NLCD data are produced by province. Indeed, the internal quality report of the NLCD data accounts for some misclassifications forest to cropland. The difference between the two data sets shows a clear administrative boundary effect, which is not limited to the Fujian Province. In other provinces, such as Guizhou, Hubei, and Henan, this effect also exists. This effect is derived from the NLCD data set, and is caused by different operators who may interpret the same type of forest as a different forest type, such as misclassifying dense forest as sparse forest. This interpretation difference is magnified while aggregating the four NLCD forest layers to one tree layer. It implies that some subjective operator errors could be introduced when interpreting the NLCD data sets. The detailed analyses of the regional agreement/disagreement for each province are in section 3.3.

3.2. Forest Distribution With Different Land Cover Types

[18] To determine the agreement/disagreement of the two data sets, the forest area of the two data sets in different land cover types is calculated, and the pixels with area more than zero are counted. Meanwhile, mean forest percentages per pixel, mean percentages difference and the total area difference in different land cover types are calculated. The results

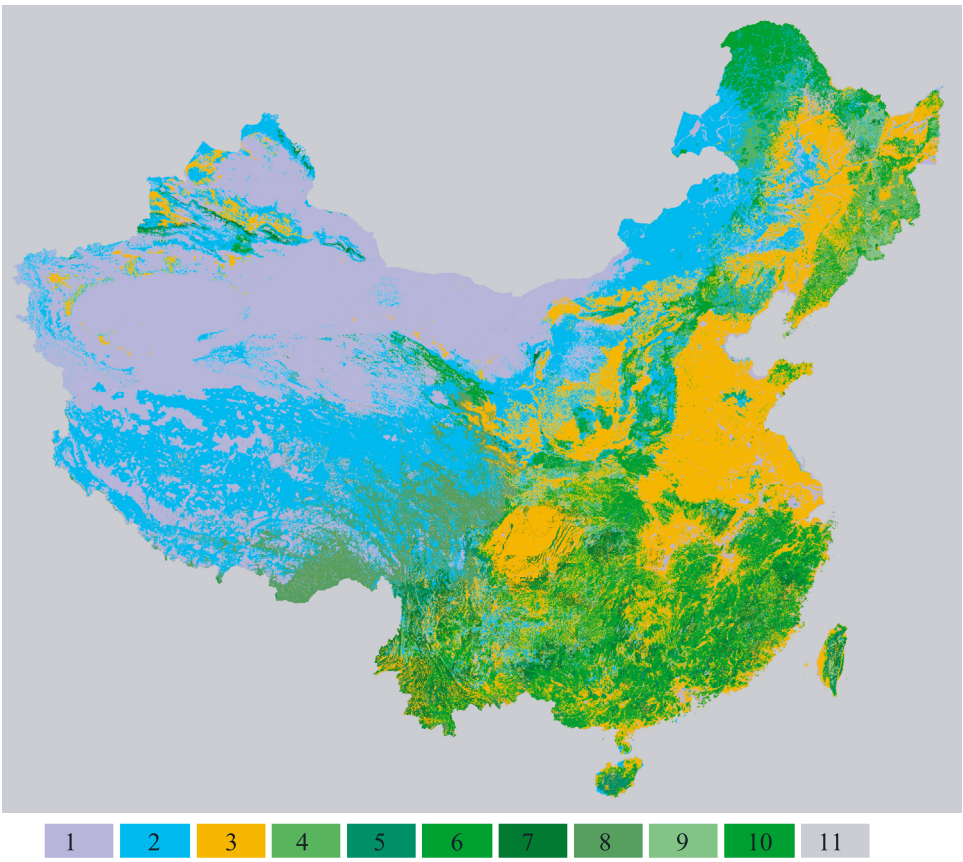

Figure 2. Land cover map of China. The legend codes are as follows: 1, nonvegetation; 2, grass; 3, farmland; 4, deciduous broadleaf forest; 5, evergreen broadleaf forest; 6, deciduous needleleaf forest; 7, evergreen needleleaf forest; 8, alpine forest; 9, mixed forest; 10, shrub; 11, no data.

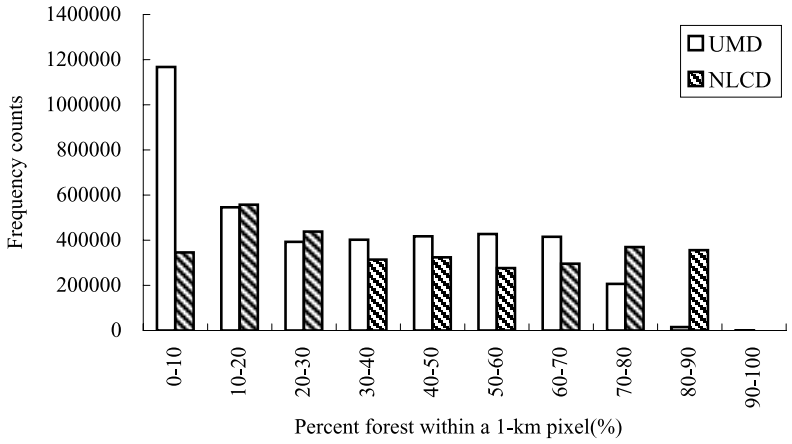

Figure 3. Pixel frequency of different percents of forest in 1-km pixel from NLCD and UMD data sets. The zero percent tree pixels are excluded.

are shown in Table 3. From the table, we can see that there is 25.3% forest in the UMD data set and 21.4% in the NLCD data set, distributed in the grass class with 1.8% in the UMD data set and 1.7% in the NLCD data set in the nonvegetation class. The estimations of the total forest area between the two data sets are in good agreement for an evergreen needleleaf forest, nonvegetation, evergreen broadleaf forest and grass, with the relative differences of

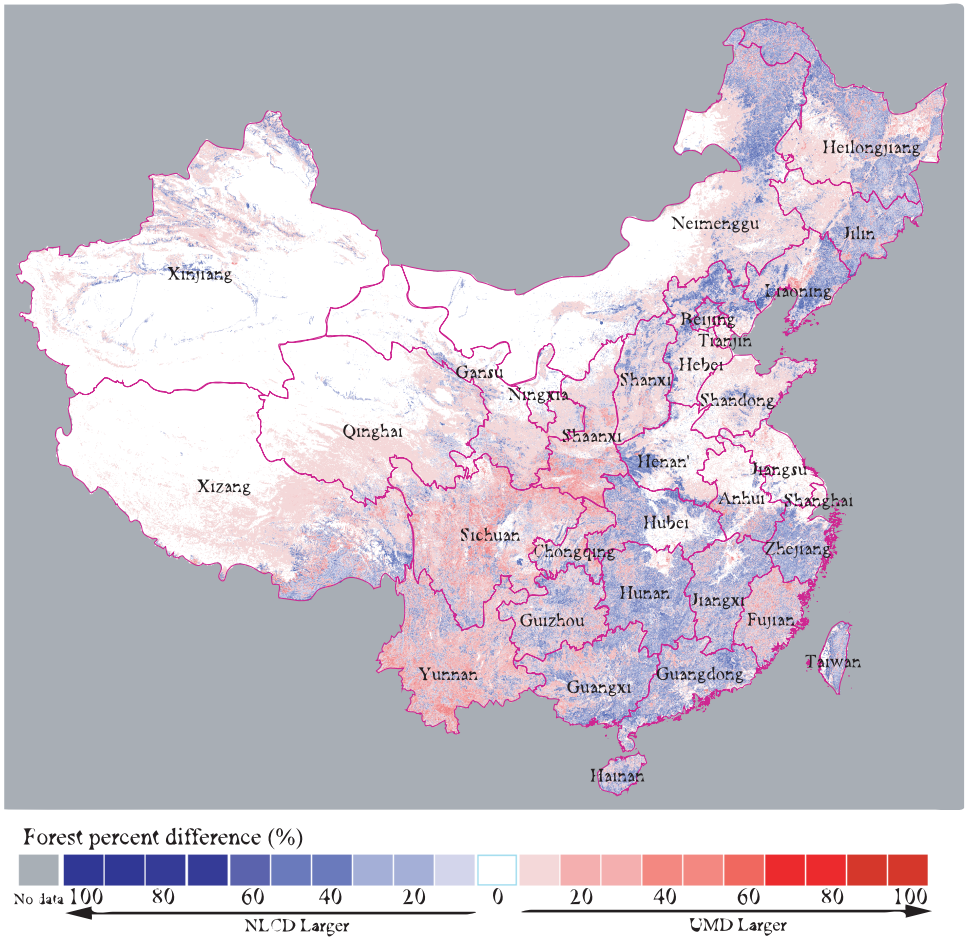

Figure 4. Spatial difference of the percent tree between NLCD and UMD data sets. Blue color means the NLCD data values are larger, red color indicates that UMD data values are larger, and zero means no difference.

Table 3. Forest Distribution and Pixel Counts in Different Land Cover Types

Land Cover Type	Pixel Counts, $\times 10^6$		Forest Area, $\times 10^6$ km^2		Mean Forest, km^2/pixel		Mean Forest Difference, km^2/pixel	Total Area Difference, $\times 10^6$ km^2
	UMD	NLCD	UMD	NLCD	UMD	NLCD		
Deciduous broadleaf forest	0.225	0.225	0.094	0.129	0.418	0.573	−0.156	−0.035
Deciduous needleleaf forest	0.128	0.128	0.066	0.083	0.516	0.648	−0.133	−0.017
Nonvegetation	0.282	0.113	0.022	0.023	0.078	0.204	−0.126	−0.001
Mixed forest	0.058	0.057	0.033	0.039	0.569	0.684	−0.115	−0.006
Shrub	1.012	0.999	0.359	0.461	0.355	0.461	−0.107	−0.102
Grass	2.577	1.332	0.313	0.294	0.121	0.221	−0.099	0.019
Alpine forest	0.37	0.272	0.115	0.102	0.311	0.375	−0.064	0.013
Evergreen needleleaf forest	0.363	0.356	0.164	0.172	0.452	0.483	−0.031	−0.008
Evergreen broadleaf forest	0.144	0.142	0.071	0.071	0.493	0.500	−0.007	0

4.8%, 4.4%, 0%, and 6.3%, respectively. However, the disagreement is relatively large for deciduous broadleaf forest, shrub, deciduous needleleaf forest, and mixed forest, with relative differences of 31.4%, 24.9%, 22.8%, and 16.7%. These differences also indicate that the estimations of forest area in the NLCD data sets are larger than those in the UMD data sets for these land cover types. The pixel counts are very close for deciduous broadleaf forest, deciduous needleleaf forest, shrub, evergreen broadleaf forest, mixed forest, and evergreen needleleaf forest, with relative differences of 0%, 0%, 1.3%, 1.4%, 1.8%, and 1.9%. However, for Alpine forest, grass and nonvegetation, the forest pixel counts are discrepant, with relative differences of 30.5%, 63.7%, and 85.6%. The mean percentage of tree cover is very close between the two data sets for evergreen broadleaf forest and evergreen needleleaf forest, with relative differences of −1.4% and −6.6%. However, for deciduous broadleaf forest, nonvegetation, grass, shrub, deciduous needleleaf forest, mixed forest and Alpine forest there are large disagreements, with relative differences of −31.3%, −89.4%, −58.5%, −26.0%, −22.7%, −18.7%, and −18.4%. In particular, the nonvegetation and grass classes have the most discrepancy, which may be explained partly by the difference in defining the forest between the two data sets, and demonstrate also that the UMD data set is inclined to contain more sparse forest than the NLCD data set.

[19] Figure 5 shows a comparison of the pixel frequency counts of different tree canopy percentages in the nine land cover types. These histograms show that there are more pixels in the NLCD data set than in the UMD data set with forest percentage more than 70% for all seven forest land cover types. However, with tree canopy percentage less than 10%, the pixel counts of the UMD data set are much larger than those of the NLCD data set for all land cover types, especially for grass and nonvegetation types, where the count differences are ∼4 times.

[20] According to these comparisons, we can find that the tree canopy percentages are very consistent for evergreen needleleaf forest and evergreen broadleaf forest in terms of pixel counts, total forest area, and mean percent forest. For deciduous broadleaf forest and deciduous needleleaf forest, there are almost identical counts. However, the total forest area is very discrepant between the two data sets; the forest area of the NLCD data set is more than that of the UMD data set, with relative differences of −31.3% and −22.7%. This may indicate that the UMD data set underestimates the forest area in these two land cover types. For nonvegetation and grass, which contain both sparsities of forest, the total forest area is close in the two data sets, but the pixels count vary greatly, which suggests the UMD data set contains a larger percentage of the sparse forest. The area agreement is favorable for shrub and mixed forest types, but there are relatively high pixel variations between the two data sets. The statistical analysis of the pixel frequency suggests that the NLCD data set has higher pixel counts for high forest percentage, while the UMD data set has higher pixel counts for low forest percentage.

[21] From the two data sets, it can be found that more than 30% of Chinese forests are distributed in shrublands and more than 20% in grasslands. This statistics demonstrates that many forests are the shrubs from secondary or man-made forest. Many forests are dispersed in grassland or agricultural regions, though they are sparse, owing to the large area of grassland, their total area is large. If we use the hard classification data, these forests would not be accounted for.

3.3. Forest Distribution in Different Provinces

[22] From Figure 4, it is clear that the difference between the NLCD and the UMD data sets has some visually spatial patterns. In this study, statistics are analyzed in more detail by provinces. The results are shown in Table 4. The table suggests that, in general, the NLCD data set shows larger total forest area. The UMD data set is larger in these provinces: Jiangsu, Gansu, Qinghai, Chongqing, Fujian, Shannxi, Sichuan, Yunnan. Gansu, Qinghai, Chongqing, Shannxi, Sichuan and Yunnan are located in the margin of the Tibetan plateau. They have similar climates with steep mountains and fragmental landscape in the forest regions. We have referred the case of Fujian in section 3.1. Jiangshu, Tianjin, and Shanghai are regions with little forest distribution. This fact may indicate that the UMD data set is larger in steep mountain regions and sparse forest regions. For the other provinces, the NLCD data set is larger. The most discrepant provinces are Sichuan and Yunnan, where the UMD data set is larger, and Neimenggu, Hunan, Heilongjiang, where the NLCD data set is the larger. For those provinces of Qinghai, Hainan, Chongqing, Anhui, Gansu, Guizhou, Fujian and Xizang with the total forest area of more than 0.01×10^6 km^2, there is very good agreement.

[23] For the mean percent forest area, the largest discrepancies occur in Neimenggu, Henan, Xizang, Hebei, Jilin, Liaoning, and Heilongjiang provinces, where the NLCD data set is larger. In these provinces, there are more pixel counts but less total area. This fact suggests that, for the

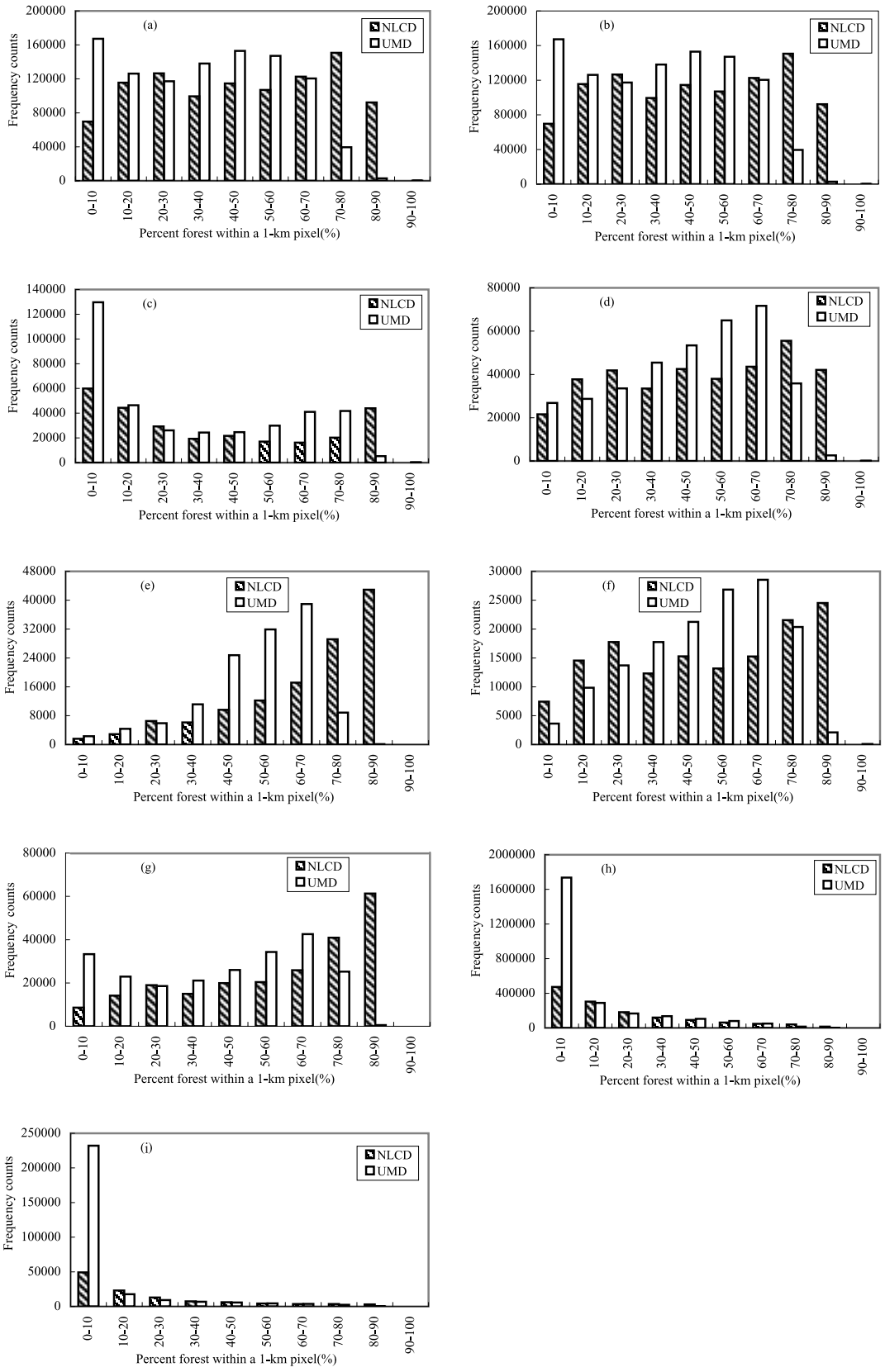

Figure 5. Pixel frequency of different percent forest ranges in 1-km pixel from NLCD and UMD data sets. The zero percent tree pixels are excluded: (a) shrub, (b) mixed forest, (c) alpine forest, (d) evergreen needleleaf forest, (e) deciduous needleleaf forest, (f) evergreen broadleaf forest, (g) deciduous broadleaf forest, (h) grass, and (i) nonvegetation.

Table 4. Forest Distribution and Pixel Counts in Different Provinces

Province	Forest Area, $\times 10^6$ km^2		Pixel Counts, $\times 10^6$		Mean Value Difference, km^2/pixel	Total Area Difference, $\times 10^6$ km^2
	UMD	NLCD	UMD	NLCD		
Neimenggu	0.075	0.111	0.497	0.277	−0.250	−0.036
Hunan	0.062	0.089	0.184	0.193	−0.124	−0.027
Heilongjiang	0.127	0.15	0.39	0.274	−0.222	−0.023
Guangdong	0.057	0.077	0.154	0.153	−0.133	−0.02
Guangxi	0.074	0.092	0.216	0.216	−0.083	−0.018
Liaoning	0.022	0.038	0.124	0.095	−0.223	−0.016
Hubei	0.036	0.052	0.123	0.133	−0.098	−0.016
Hebei	0.011	0.026	0.132	0.083	−0.230	−0.015
Jiangxi	0.056	0.07	0.143	0.148	−0.081	−0.014
Zhejiang	0.037	0.049	0.085	0.085	−0.141	−0.012
Jilin	0.051	0.062	0.169	0.117	−0.228	−0.011
Henan	0.01	0.019	0.08	0.052	−0.240	−0.009
Xinjiang	0.013	0.022	0.248	0.103	−0.161	−0.009
Taiwan	0.015	0.019	0.029	0.029	−0.138	−0.004
Guizhou	0.038	0.041	0.156	0.157	−0.018	−0.003
Shanxi	0.013	0.016	0.141	0.08	−0.108	−0.003
Hainan	0.011	0.014	0.031	0.032	−0.083	−0.003
Anhui	0.019	0.021	0.071	0.053	−0.129	−0.002
Shandong	0.004	0.006	0.096	0.032	−0.146	−0.002
Xizang	0.08	0.081	0.424	0.191	−0.235	−0.001
Beijing	0.002	0.003	0.013	0.011	−0.119	−0.001
Ningxia	0.001	0.001	0.017	0.009	−0.052	0
Tianjin	0	0	0.006	0.001	0	0
Shanghai	0	0	0.003	0.001	0	0
Jiangsu	0.003	0.002	0.028	0.009	−0.115	0.001
Gansu	0.02	0.019	0.173	0.091	−0.093	0.001
Qinghai	0.01	0.009	0.265	0.075	−0.082	0.001
Chongqing	0.018	0.015	0.069	0.059	0.007	0.003
Fujian	0.056	0.052	0.117	0.114	0.022	0.004
Shannxi	0.04	0.024	0.173	0.105	0.003	0.016
Sichuan	0.124	0.086	0.419	0.32	0.027	0.038
Yunnan	0.15	0.109	0.376	0.327	0.066	0.041

dense forest, the NLCD data set has larger forest area and less forest pixels; for sparse forest, the UMD data set has less forest area but more forest pixels. For forest area more than 0.02×10^6 km^2 in Guizhou, Shannxi, Jiangxi, Fujian, Guangxi, Sichuan and Yunnan provinces, these two data sets are in good agreement.

4. Summary and Discussion

[24] Understanding the Chinese forest area and its distribution pattern is important for locating carbon dioxide sequestration. However, these parameters still demonstrate many uncertainties. *Fang et al.* [2001] use data from the latest forest inventory of 1994–1998 conducted by the Chinese Ministry of Forestry to report a total forest area, including both planted and natural forests, of 1.058×10^6 km^2. Elsewhere in their record they report a forest area of 1.337×10^6 km^2, citing the same official inventory. The reasons for the difference of 0.28×10^6 km^2 are not given, but the higher value seems to include additional nontimber plantations [*Houghton and Hackler*, 2003] or the forest definition is different in two cites. In this paper, the forest areas are 1.233×10^6 km^2 for the UMD data set and 1.392×10^6 km^2 for the NLCD data set respectively. We can't determine which data set is more accurate directly because the area of forest depends largely on the definition of forest.

[25] From the description above, it is clear that mapping forest accurately is very difficult. It is therefore reasonable to find large differences and similarities with different approaches. In this study we have compared the percent of trees from the NLCD data set and MODIS standard product over all of China at 1 km resolution. Though the two data sets are derived from completely different approaches, we have tried to compare the two data sets at the scale of the whole country with different land cover types. For the whole country, the difference in total forest area is 159,000 km^2. There are more pixels for dense forest and fewer pixels for sparse forest in the NLCD data set. Generally, forest percent areas of the NLCD data set in eastern China are larger and for the Tibetan plateau margin region, percent areas of the UMD data set are larger. For different land cover types, the percent areas of forest have excellent agreement for evergreen broadleaf forest and evergreen needleleaf forest, but counterproductive agreement for the nonvegetation and grass classes. At the province level, Inner Mongolia is the place where both data sets show a diverse result, but they have good agreement in Guizhou and Fujian among those provinces with forest area more than 20,000 km^2.

[26] Each method can cause errors. The NLCD method may ignore the sparse trees because it labels a pixel as nonforest if its tree cover is less than 10%. In the region with sparse forest, such as urban, grassland, cropland, the tree canopy areas in the NLCD data set is mostly zero. However, the NLCD method may overestimate the dense forest coverage because it labels a pixel as forest if the tree coverage is more than 30%. Although the crown cover in

the NLCD data was converted to canopy cover through multiplication by 0.8 before the data comparisons, the nonforest fraction was not excluded. The UMD method is based on multitemporal land surface spectral characteristics; the increased fragmentation of forest and the confusion in spectral space in mixed-type forest can result in the biased result [*Hansen et al.*, 2002b]. In comparing the two data sets, both correspond well for evergreen forest but the forest percentages for deciduous forest may be underestimated in the UMD data. According to the definition of a forest, the UMD data have the less omission error for very sparse forest distribution, such as grassland and cropland. Compared to the NLCD data made by hundreds of people over many years, the UMD data are automatically produced from MODIS data globally and are more objective and economical, which indicates the UMD method is a very promising technique.

[27] **Acknowledgments.** This research was supported in part by the National 973 program (2002CB4125) and the National Natural Science Foundation from China (40471098). The authors would like to thank the anonymous reviewers for their insightful comments.

References

Bartalev, S., A. S. Belward, D. Erchov, and A. S. Isaev (2003), A new SPOT4-VEGETATION derived land cover map of northern Eurasia, *Int. J. Remote Sens.*, *24*, 1977–1982.

Cao, M. K., S. D. Prince, K. Li, B. Tao, J. Small, and S. Shao (2003), Response of terrestrial carbon uptake to climate interannual variability in China, *Global Change Biol.*, *9*, 536–546.

DeFries, R. S., et al. (1995), Mapping the land surface for global atmosphere-biosphere models: Towards continuous distributions of vegetation's functional properties, *J. Geophys. Res.*, *100*, 20,867–20,882.

DeFries, R. S., J. R. G. Townshend, and M. C. Hansen (1999), Continuous fields of vegetation characteristics at the global scale at 1-km resolution, *J. Geophys. Res.*, *104*, 16,911–16,923.

DeFries, R. S., R. A. Houghton, M. C. Hansen, C. B. Field, D. Skole, and J. R. G. Townshend (2002), Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s, *Proc. Natl. Acad. Sci. U.S.A.*, *99*, 14,256–14,261.

Fang, J., A. Chen, C. Peng, S. Zhao, and L. Ci (2001), Changes in forest biomass carbon storage in China between 1949 and 1998, *Science*, *292*, 2320–2322.

Fernandes, R., R. Fraser, R. Latifovic, J. Cihlar, J. Beaubien, and Y. Du (2004), Approaches to fractional land cover and continuous field mapping: A comparative assessment over the BOREAS study region, *Remote Sens. Environ.*, *89*, 234–251.

Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, *83*, 287–302.

Goetz, S. J., R. K. Wright, A. J. Smith, E. Zinecker, and E. Schaub (2003), IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region, *Remote Sens. Environ.*, *88*, 195–208.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, R. Sohlberg, C. Dimiceli, and M. Carroll (2002a), Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data, *Remote Sens. Environ.*, *83*, 303–319.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, L. Marufu, and R. Sohlberg (2002b), Development of a MODIS tree cover validation data set for Western Province, Zambia, *Remote Sens. Environ.*, *83*, 235–320.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg (2003), Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm, *Earth Interact.*, *7*, pap. 10. (Available at http://EarthInteractions.org)

Hou, X. Y. (1983), Vegetation of China with reference to its geographical distribution, *Ann. Mo. Bot. Garden*, *70*, 508–548.

Houghton, R. A. (1999), The annual net flux of carbon to the atmosphere from changes in land use 1850–1990, *Tellus, Ser. B*, *51*, 298–313.

Houghton, R. A., and J. L. Hackler (2003), Sources and sinks of carbon from land-use change in China, *Global Biogeochem. Cycles*, *17*(2), 1034, doi:10.1029/2002GB001970.

Li, C., J. Koskela, and O. Luukkanen (1999), Protective forest systems in China: Current status, problems and perspectives, *Ambio*, *28*, 341–345.

Liang, S. (2003), *Quantitative Remote Sensing of Land Surfaces*, 550 pp., John Wiley, Hoboken, N. J.

Liu, J. Y. (1996), *Macro-scale Survey and Dynamic Study of Natural Resources and Environment of China by Remote Sensing* (in Chinese), 353 pp., Press of Sci. and Technol., Beijing.

Liu, J. Y., D. F. Zhuang, D. Luo, and X. Xiao (2003), Land-cover classification of China: Integrated analysis of AVHRR imagery and geophysical data, *Int. J. Remote Sens.*, *24*, 2485–2500.

Liu, M., X. Tang, J. Liu, and D. F. Zhuang (2001), Research on scaling effect based on 1 km grid cell data (in Chinese), *J. Remote Sens.*, *5*(3), 183–190.

Richardson, S. D. (1990), *Forests and Forestry in China*, Island, Washington, D. C.

Sellers, P. J., et al. (1997), Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, *Science*, *275*, 502–509.

Streets, D. G., K. Jiang, X. Hu, J. E. Sinton, X.-Q. Zhang, D. Xu, M. Z. Jacobson, and J. E. Hansen (2001), Recent reductions in China's greenhouse gas emissions, *Science*, *294*, 1835–1836.

Townshend, J. R. G. (1994), Global data sets for land applications from the Advanced Very High Resolution Radiometer: An introduction, *Int. J. Remote Sens.*, *15*, 3319–3332.

Xiao, X., J. Liu, D. Zhuang, S. Frolking, S. Boles, B. Xu, M. Liu, W. Salas, B. Moore III, and C. Li (2003), Uncertainties in estimates of cropland area in China: A comparison between an AVHRR-derived dataset and a Landsat TM-derived dataset, *Global Planet. Change*, *37*, 297–306.

S. Liang, Department of Geography, 2181 LeFrak Hall, University of Maryland, College Park, College Park, MD 20742, USA.

J. Liu, R. Liu, and D. Zhuang, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Datun Road, Chaoyang, Beijing, China 100101. (lronggao@yahoo.com)