CHINA FOREST COVER EXTRACTION BASED ON GOOGLE EARTH ENGINE

GUO Yan-tao^{1,2}, ZHANG Xiao-mei¹, LONG Teng-fei¹, JIAO Wei-li¹, HE Guo-jin¹, YIN Ran-yu^{1,2}, DONG Yun-yun^{1,2}

Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
 University of Chinese Academy of Sciences, Beijing 100049, China

KEY WORDS: Google Earth Engine; Landsat 8; random forest; China forest cover; China forest area

ABSTRACT:

Forest cover rate is the principal indice to reflect the forest acount of a nation and region. In view of the difficulty of accurately calculating large-scale forest area by traditional statistical survey methods, it is proposed to extract China forest area based on Google Earth Engine platform. Trained by the enough samples selected through the Google Earth software, there are nine different random forest classifiers applicable to their corresponding zones. Using Landsat 8 surface reflectance data of 2018 year and the modified forest partition map, China forest cover is generated on the Google Earth Engine platform. The accuracy of China's forest coverage achieves 89.08%, while the accuracy of Global Forest Change datasets of Maryland university and Japan's ALOS Forest/Non-Forest forest product reach 87.78% and 84.57%. Besides, the precision of tropical/subtropical forest , temperate coniferous forest as well as nonforest region are 83.25%, 87.94% and 97.83%, higher than those of other's accuracy. Our results show that by means of the random forest partition can be extracted more accurately. Through the computation of forest cover, our result shows that China has a area of 220.42 million hectare in 2018.

1.INTRODUCTION

Among many earth systematical processings, vagetation land cover is the indispensible element. Vegetation land cover is required by a number of general to be the boundary layer of execution model(Sellers et al. 1997). As a significant component of land cover research topics, forest cover detection is now more than ever becoming the focus of scientific research and resource management projects, such as investigating climate change, food security, habitat loss(Foley et al. 2005). The purpose of mapping large area forest is producing globally consistent characters possessing local relevance and practicability, in other words, cross-scale reliable information(Hansen et al. 2013). Due to the significance of forest cover data, countries in the world and international research institutes conduct a series of investigations on the topic of different scale land cover mapping.

Forest detection already raises wide concern of international society and achives s series of results. Recently, remote sensing satellite data reveal a greening pattern that is strikingly prominent in China and India and overlaps with croplands world-wide and China alone accounts for 25% of the global net increase in leaf area with only 6.6% of global vegetated area(Chen et al. 2019). Meanwhile, mangrove forests along the coastal zones in China were mapped by integration of the GEE platform, time series Landsat and Sentine-1A SAR images(Chen et al. 2017). Besides, PALSAR-based forest map in China

demonstrate the potential of integrating PALSAR and MODIS images to map forests in large areas(Oin et al. 2015). On the other hand, some novel approachs were proposed to produce more accurate 25m forest maps by integrating PALSAR/PALSAR-2 and MODIS NDVI data during 2007-2010 and reconstruct annual 25m forest maps from time-series MODIS NDVI images during 2011–2014(Zhang et al. 2019). To minimize the influence of the changing ground footprint of MODIS, there were two new algorithms and a new assessment framework for near real-time monitoring of tropical forest disturbance(Tang et al. 2019). Referring to the global forest cover datasets, there are four representitive products alternative: (1) The Global Forest Change map (GFC) product proviede by the Maryland university(Hansen et al. 2013). (2) The PALSAR/PALSAR-2 mosaic and forest/non-forest (FNF) map produced by Japan Aerospace Exploration Agency (JAXA) (Shimada et al. 2014). (3) The first 30m and 10m resolution global land-cover maps created by Landsat Themtic Mapper (TM) and Enhanced Thematic Mapper Plus (EMT+) data(Gong et al. 2019; Gong et al. 2013). (4) The Global Land Cover (GLC) mapping at 30m resolution based on a POK-based operational approach supplied by National Geomatics Center of China(Chen et al. 2015). Comparing above products, there are some problem existing on the data processing and reprocessing, or the precision to be improved. Therefore, how to produce big scale forest maps efficiently and precisely is a puzzle to be solved.

The Google Earth Engine based on cloud compute platform combines the high-performance abilities with large=scale geographic data processing missions. This solution settles a train of major information technology challenges, such as data acquisition and storage, file pattern analysis, database management and equipment distribution(Gorelick et al. 2017).

In this study, we produce China forest cover maps of different partitions in 2018 using Google Earth Engine for data acquisition and operation platform. This forest distribution product is made from Landsat image data and random forest classification method. To guarantee the accuracy of this map, this study compares the forest map with Global Forest Change data and Forest/Non-Forest data.

2.CHINA FOREST PARTITION

Global land covers are usually divided into fourteen biocoenosis and eight geographic zones and China has eight biocoenosis(Olson and Dinerstein 2002; Olson et al. 2001). In this study, we merge eight biocoenosis in China into five forest partitions to assist extraction of different forest. The five forest partitions are boreal forest, temperate coniferous forest, temperate cross forest, tropical/subtropical forest and nonforest.

Forest Partitions	Biocoenosis
Boreal Forests	Boreal Forests/Taiga
Temperate Conifer	Temperate Conifer Forests
Forests	
Temperate Mixed	Temperate Broadleaf and
Forests	Mixed Forests
Tropical and	Tropical and Subtropical
Subtropical	Broadleaf Forests
Broadleaf Forests	
Nonforest	Temperate Grasslands,
	Savannas and Shrublands
	Flooded Grasslands ,Savannas
	Montane Grasslands and
	Shrublands
	Deserts and Xeric Shrublands





Figure 1. China forest partitions

3.DATASETS AND RESEARCH METHODS

Google Earth Engine (GEE) contains a a range of Landsat image collections, among which is Landsat-8 Surface Reflectance Tier and it comes from Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS).

In this paper, the USGS Landsat-8 Surface Reflectance Tier datasets are used for import data, which is provied by GEE. These data have been atmospherically corrected using Landsat Surface Reflectance Code (LaSRC) (Vermote et al. 2016) and includes a cloud, shadow, water and snow mask produced using C Function of Mask (CFMASK) (Zhu and Woodcock 2014) , as well as a per-pixel saturation mask. Meanwhile, we select the median time images after cloud clear in different partitions to serve as the train data and classification data of the partition. The influence of cloud and shadow can greatly be avoided by mean of the operations.



Figure 2. China forest cover detection flowchart

3.1 Index computation

In order to avoid the cloud and shadow influence, quality assessment band is used for mask band to reject the cloud pixels. After that operation, six spectral

indexes are computed to act as character indexes for different land cover species. These indexes include Normalized Difference Vegetation Index (NDVI) (Tucker 1979), Normalized Difference Water Index (NDWI) (Gao 1996), Normalized Difference Built-up Index (NDBI) (Zha et al. 2003), Normalized Difference Moisture Index (NDMI) (Wilson and Sader 2002), Global Environment Monitoring Index (GEMI) (Pinty and Verstraete 1992) and Soil Adjusted Vegetation Index (SAVI) (Huete 1988). Normalized Difference Vegetation Index is sensitive to vegetation greenness and can be employed to monitor the photosynthetically active biomass of plant canopies. Normalized Difference Water Index is sensitive to changes in liquid water content of vegetation canopies and less sensitive to atmospheric effects than Normalized Difference Vegetation Index. Normalized Difference Built-up Index is much more effective and advantageous in mapping general built-up areas, which can serve as a worthwhile alternative for quickly mapping nonforest land. Some study shows that the less common Normalized Difference Moisture Index method (utilizing the middle infrared band instead of the visible red) produced significantly higher accuracies for detecting forest harvest in all classification trials. Global Environment Monitoring Index reduces the relative effects of these undesirable atmospheric perturbations, while maintaining the information about the vegetation cover. Soil-adjusted vegetation index is found to be an important step toward the establishment of simple "global" models that can describe dynamic soil-vegetation systems from remotely sensed data. The six spectral indexes are combined with six band reflectances.

3.2 Training samples

On the basis of the forest density in various partitions, a diifferent number of samples are selected as training data. Referring to near-real time high-resolution image collection on the Google Earth software, these datasets are labelled on forest,nonforest or water, in order to correspond to the classification labels of GFC and FNF products. The training pixels are extracted randomly. To ensure only clearly forest pixels were selected, the

Study	Sample Number		
Area	Forest	Non-	Water
		Forest	
Nonforest	66	1474	36
Boreal Forest	39	10	4
Tropical and	626	444	10
Subtropical			
Broadleaf Forest			
Temperate	824	1505	68
Mixed Forest			
Temperate	1017	889	37
Conifer Forest			
China	2572	4322	155

forest samples were collected carefully to avoid pixels near the boundary of forest(Bastarrika et al. 2011).

Table 2. The training sample numbers for the forest



Figure 3. The samples map of distribution

3.3Model Training

The random forest algorithm provided by GEE was applied to train the forest classifier. Compared with decision tree algorithm creating GFC maps, the random forest classifier contains more numbers of decision trees. The Random Forest classifier uses bootstrap aggregating for form an ensemble of classification and induction tree like tree classifiers. This structure means that random forest model has more robustness and higher anti-interference. Random forest has become one of the high accuracy and widely applicable algorithms(Pal 2005). Random forest is a ensamble composed of decision trees acted as basic learners, while all learners in the forest have same distribution and every attribute of decision tree depends on the spectral characters choosed independently and

randomly. The generalization error of random forest is dicided by the individual structural strength and relevance of decision trees in the forest. The formula of random forest is expressed as

$$H(x) = \arg\max_{Y} \sum_{i=1}^{k} I(h_i(x) = Y)$$
(1)

Where H(x) indicates the random forest integrated classification model, h_i indicates single decision tree, Y indicates the output variable , $I(\cdot)$ indicates the indicator function which ranges from 0 and 1.

The advantage of random forest algorithm lies in usage of out-of-bag cross-validaton (OOBCV) to realise evaluation of classification accuracy. overall Compared with Adaboost, random forest utilizes characters selected randomly to split error ratio of every node domain, thus this model has better stability in noise reduction and anti-interference aspects. Overall evaluation manages the interior error, structure and relevance of forest model by means of estimating the posterior probability of every node in the decision tree of random forest. These indices indicate the response of character number increase which are applied by segmentation. Not only that, overall estimation also evaluates the significance of variable (Breiman 2001).

Because of multiplicate input characters including six bands of surface reflectance and six specific spectral indexes, the number of decision trees in the random forest model was limited to 500 to balance accuracy and timeliness.

Classification results are hard voting of three labels. This is the overall consequences of all decision trees in the classifier. In the other hand, random forest classifier can also export the probability of each category. This results act as the confidence level output of every category, which contain the confidence index ranging from 0 to 1.

4. CLASSIFICATION RESULT AND ACCURACY ANALYSIS

4.1 Classification results

In view of area forest partion, different number of validation points were produced for every study region. After that, extract the real categories of points based on the high resolution images in the Google Earth software and the three forest maps, thus we obtained the error matrixs of classification results. At Last, the overall accuracy (OA) and kappa coefficients (KCs) were computed by the error matrixs.

$$OA = \frac{\sum_{i=1}^{n} x_{ii}}{N}$$
(2)

$$KC_{s} = \frac{\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} (x_{i+} \times x_{+i})}{N^{2} - \sum_{i=1}^{k} (x_{i+} \times x_{+i})}$$
(3)

Where N indicates the pixels used in the accuracy assessment, x_ii indicates the overall number of the class i pixels classified correctly, $x_{(i+)}$ indicates the number of class i pixels in classification results, $x_{(+i)}$ indicates the number of class i pixels in validation results.



Figure 4. The forest extraction result in China

Study Area	Overall Accuracy (%)		
	RF	FNF	GFC
Tropical and	83.25%	69.25%	73.25%
Subtropical			
Broadleaf			
Forests			
Temperate	87.94%	81.81%	86.87%
Conifer			
Forests			
Nonforest	97.83%	95.50%	96.83%
Temperate	87.90%	87.40%	88.70%
Mixed Forests			
Boreal Forests	90.00%	93.00%	97.00%
China	89.08%	84.57%	87.78%

4.2 Classification accuracy

Table 3. The overall accuracy of the three forest maps in five partitions

Study Area	Kappa Coefficient		
	RF	FNF	GFC
Tropical and	0.6735	0. 3985	0. 4794
Subtropical			
Broadleaf			
Forests			
Temperate	0.7513	0.6197	0.7209
Conifer	0.1010		0
Forests			
Nonforest	0. 4957	0.2100	0. 2879
Temperate	0. 6335	0. 5663	0.6057
Mixed Forests			
Boreal Forests	0. 00	0. 00	0.5640
China	0. 7503	0.6367	0. 7088

Table 4. The kappa coefficients of the three forest maps in five partitions

4.3 Local comparision of forest detection

In this part, four forest extractions of four forest partitions are compared to analysis the reason of misclassification and missorting. Fig 4 shows the original images and forest extraction results of four forest partitions including Tropical and Subtropical Broadleaf Forests (a1,a2,a3,a4), Temperate Conifer Forests of Notheast (b1,b2,b3,b4), Temperate Conifer Forests of Southwest (c1,c2,c3,c4) and Temperate Mixed Forests of East (d1,d2,d3,d4), while the classification differences are highlighted by the yellow circles.

In the tropical and subtropical broadleaf forests, forests weren't extracted in the FNF product compared with the RF product. When it comes to GFC product, some non-forest regions weren't classified , such as road and lake. In the temperate conifer forests of northeast, FNF map and GFC map appear large area of forest misclassification while RF map seperates forest and non-forest well. The reason of this situation may be local forest density is sparse and unlikely to be discrinible considering the origin image. In the temperate conifer forest of southwest, there are different levels of forest misclassification cases in the mountain regions of FNF product and GFC product. In the temperate mixed forest of east, FNF product has serious missorting situations when roads and grassland in the city region are divided into forest. Some trees on the boundary of forest in the GFC product are classified as non-forest. On the contrary, RF product extracts the boundary of forest better and reduces the missortting of non-forest vegetation in the urban area. In summary, the classification accuracy of forest product created by random forest algoriothm in the GEE platform is higher than those of GFC and FNF.



Figure 5. Local comparision of forest detection in different partitions

4.4 Area computation

Forest partition	Area(Ten
	thousand hectare)
Nonforest	663.1828
Boreal Forests	0.6885
Tropical and Subtropical	10195.1802
Broadleaf Forests	
Temperate Mixed Forests	7491.9180

Temperate Conifer Forests	3691.2586
China	22042.23

Table 5. Forest area in different forest partitions.

5.CONCLUSION

This work demonstrates that the classifier trained by random forest algorithm extracts China forest cover better and achives the overall accuracy of 89.08%. It's not to be ignored that this classifier has better performance in the extraction of forest in tropical and subtropical broadleaf forests, temperate conifer forests and temperate mixed forests. In addition, the process speed in the Google Earth Engine platform is high, thus the time cost of data download and data selection can be saved.

Although this work has attained superior result of forest cover extraction ,there are still some challenges remained to be solved. For example, due to the cloud cover and shadow in the data all the year round, there are data deficiency in the classification result. In the other hand, there are some salt pepper effect in some districts as calculated based on the pixel. So it's neccesary to introduce the method of object-based to avoid the misclassification. In addition, there are more data could be applied to classify the forest types, such as digital elevation model and hyperspectral images.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Programs of China [grant numbers 2016YFA0600302], the Strategic Priority Research Program of the Chinese Academy of Sciences [grant number XDA19090300], the program of the National Natural Science Foundation of China [grant number 61731022].

REFERENCES

Bastarrika, A., Chuvieco, E., & Martín, M.P. (2011). Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors. *Remote Sensing of Environment, 115*, 1003-1012 Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32

Chen, B., Xiao, X., Li, X., Pan, L., Doughty, R., Ma, J., Dong, J., Qin, Y., Zhao, B., Wu, Z., Sun, R., Lan, G., Xie, G., Clinton, N., & Giri, C. (2017). A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131, 104-120

Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R.K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tømmervik, H., Bala, G., Zhu, Z., Nemani, R.R., & Myneni, R.B. (2019). China and India lead in greening of the world through landuse management. *Nature Sustainability, 2*, 122-129

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., & Mills, J. (2015). Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 7-27 Cossu, R., Petitdidier, M., Linford, J., Badoux, V., Fusco, L., Gotab, B., Hluchy, L., Lecca, G., Murgia, F., & Plevier, C. (2010). A roadmap for a dedicated Earth Science Grid platform. *Earth Science Informatics*, *3*, 135-148

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., & Snyder, P.K. (2005). Global Consequences of Land Use. *Science, 309*, 570

Gao, B.-c. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58,* 257-266

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., & Song, L. (2019). Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin*

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., & Chen, J. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, *34*, 2607-2654

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27

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., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., & Townshend, J.R.G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*, 850

Huete, A.R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*, 295-309

Olson, D.M., & Dinerstein, E. (2002). The Global 200: Priority Ecoregions for Global Conservation. *Annals of the Missouri Botanical Garden, 89*, 199-224

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., & Kassem, K.R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*, 933-938

Pal, M. (2005). Random forest classifier for remote sensing classification *International Journal of Remote Sensing, 26,* 217-222

Pinty, B., & Verstraete, M.M. (1992). GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio, 101,* 15-20

Qin, Y., Xiao, X., Dong, J., Zhang, G., Shimada, M., Liu, J., Li, C., Kou, W., & Moore, B. (2015). Forest cover maps of China in 2010 from multiple approaches and data sources: PALSAR, Landsat, MODIS, FRA, and NFI. *ISPRS Journal of Photogrammetry and Remote Sensing, 109*, 1-16 Sellers, P.J., Dickinson, R.E., Randall, D.A., Betts, A.K., Hall, F.G., Berry, J.A., Collatz, G.J., Denning, A.S., Mooney, H.A., Nobre, C.A., Sato, N., Field, C.B., & Henderson-Sellers, A. (1997). Modeling the Exchanges of Energy, Water, and Carbon Between Continents and the Atmosphere. *Science, 275*, 502-509

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., & Lucas, R. (2014). New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment, 155*, 13-31

Tang, X., Bullock, E.L., Olofsson, P., Estel, S., & Woodcock, C.E. (2019). Near real-time monitoring of tropical forest disturbance: New algorithms and assessment framework. *Remote Sensing of Environment, 224*, 202-218

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8,* 127-150

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment, 185*, 46-56

Wilson, E.H., & Sader, S.A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment, 80*, 385-396 Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*, 583-594

Zhang, Y., Ling, F., Foody, G.M., Ge, Y., Boyd, D.S., Li, X., Du, Y., & Atkinson, P.M. (2019). Mapping annual forest cover by fusing PALSAR/PALSAR-2 and MODIS NDVI during 2007–2016. *Remote Sensing of Environment, 224*, 74-91

Zhu, Z., & Woodcock, C.E. (2014). Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment, 152*, 217-234