

# Methods of the National Vegetation Classification Based on Vegetation Partition

Y.Y. Hao

*College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

*Chinese Academy of Sciences Institute of Remote Sensing and Digital Earth State Key Laboratory of Remote Sensing, Beijing 100101, China*

X.B. Luo

*College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

B. Zhong & A.X. Yang

*Chinese Academy of Sciences Institute of Remote Sensing and Digital Earth State Key Laboratory of Remote Sensing, Beijing 100101, China*

**ABSTRACT:** The high precision land cover classification products has a very important significance for the study of quantitative remote sensing and remote sensing applications. Now there has a lot of free and global coverage land classification products, which is mostly developed by foreign research institutions and personnel. Due to the China regional complex terrain, differences in vegetation structures and in crop planting structures have not been fully studied, the classification accuracy of these products in the area of China is very low, especially the vegetation classification accuracy. Therefore, it is necessary to produce a vegetation type classification product of China region. So according to Chinese regional topography, soil and other information, and based on existing vegetation regionalization, we developed a method of Chinese vegetation classification based on vegetation partition. This method is based on long time series, which can capture the surface information changed with time with the high time resolution, and can improve the classification accuracy by the differences in the time dimension of the object. And we complete the 2012 national land cover classification by using the method. At last, the method of stratified random sampling was carried out to assess the accuracy of the classification results, found that the overall accuracy and Kappa coefficient of the classification products is greatly improved, which the overall accuracy was 90.78%, Kappa coefficient was 0.86. And the product was compared with the MODIS land cover data products, found that the accuracy of this product increased by 61.38% in vegetation than the MODIS land cover data product..

**KEYWORD:** varied topography; vegetation types; land cover classification; vegetation partition

## 1 INTRODUCTION

The vegetation is one of the biggest changes in the earth's surface, which directly affects the accuracy of quantitative remote sensing. Land cover is an important information based on the ecosystem, and is important to the research of carbon cycle, habitat and biodiversity and public health, and plays a key role in improving hydrological and atmospheric models. There are five kinds of global land cover datasets derived from the medium resolution satellite: global land cover characteristic (GLCC), the classification system of which is 17 kinds types of International Geosphere-Biosphere Programme Data and Information System Cover (IGBP-DISCOVER) developed by Geological survey of the United States (USGS). This data set is produced by unsupervised classification, the global overall accuracy (OA) of which is 66.9%. The University of Maryland (UMD) land-cover map, the system of which is simplified 14 types of IGBP, using supervised classification decision tree method. The OA is 69%. The

global land cover 2000 (GLC2000) map using the land cover classification system (LCCS) of food and agricultural organizations (FAO) generated by European Commission's Joint Research Center (EC-JRC) is completed with supervised classification, the OA is between 63.6% and 73.6%. MODIS land cover map (MODIS LC) with the 22 types of FAO LCCS developed by European Space Agency (ESA), uses unsupervised classification. The OA is 67.1%. This five global land cover maps are developed by automation program of supervised or unsupervised classification, the sample capacity and the subjective factors of the sample caused the poor precision. From some third party researchers of the world, the degree of accuracy assessment of these land cover products is different in different parts of the world. And the consistency of these kinds of global land cover products are less than 20% in addition to South America. Especially the understanding of China is not sufficient, the accuracy evaluation results of the products in China are very low, and the highest is GLC2000 with the 59.28% of OA. The

land cover data set in China has not been fully concerned in considering the accuracy evaluation. The elevation and latitude of China is larger and the terrain is broken, China has the characteristics of high diversity and segmentation. The single rule does not meet the needs of high precision. Therefore it is necessary to produce a land cover map which is suitable for China land characteristics modeling and is compatible with the acceptable legend system. In this paper, the whole country is divided into 8 zones according to the law of vegetation and soil, precipitation and terrain. Each partition was analyzed with the phenology and time series and set rules with the law of each type of planting.

## 2 DATA PROCESSING OF STUDY AREA

### 2.1 Data source

In this paper, the MODIS vegetation index product is used as the classification data source. The resolution of MODIS product includes four kinds: 250m, 500m, 1000m and 5600m. Most of the standard MODIS products use sine projection. The classification data used in the paper are MOD13A2 vegetation index products, equipped with a platform for Terra. The spatial resolution is 1km, the time resolution is 16 days and the sine projection is used. The auxiliary data is the ASTER GDEM data. The spatial resolution of the data is 30m, and then resampling for 1km.

### 2.2 Data processing

MOD13A2 vegetation index product data is HDF image, which need to convert to TIF file. The HDF data was converted to TIF data through the batch conversion process. Then the TIF data was band synthesized to a vegetation index time series with 23 bands.

### 2.3 The data partition and legend system

China has a vast territory and a large number of land cover. Single rule does not represent the growth of all parts in the country. And China's topography and biodiversity caused that the simple terrain division is not reasonable. Therefore, this paper proposes a national land classification method based on vegetation regionalization. In this paper, the vegetation zoning used is the most authoritative and most extensive vegetation zoning map. The zoning map divided the country into 8 vegetation regions according to the geography of the vegetation distribution principle, taking into account the relationship between vegetation and topography, soil, climate, hydrological factors, farming system and crop type combination. In this paper, the nation was vectored into 8 regions us-

ing the vegetation regionalization map. The following is the zoning map.

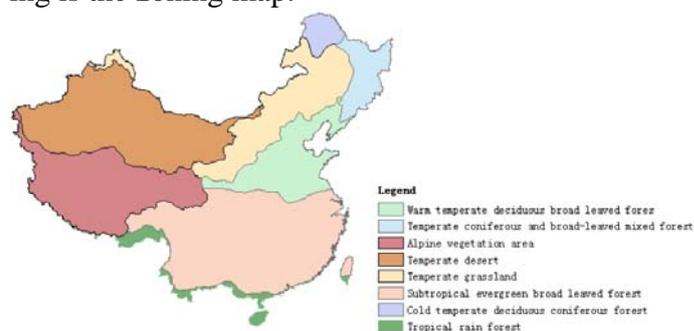


Figure 1 China vegetation regionalization map

In this paper, the map is 1km national land cover first class mapping. In order to make the product more widely applicable, the classification system uses Globland30 system (Jun Chen). The following is the system.

Table 1 Classification system

Code	Type	Content
10	Cultivated land	used for agriculture, horticulture and gardens including paddy fields, irrigated and dry farmland, vegetation and fruit garden, etc.
20	Forest	Covered with trees, with vegetation cover over 30%, including deciduous and coniferous forest and sparse woodland with cover 10-30%, etc.
30	Grassland	Covered by natural grass with cover over 10%, etc.
40	Shrubland	Covered with shrubs with cover over 30%, including deciduous and evergreen shrubs and desert steppe with cover over 10%, etc.
50	Wetland	Covered with wetland plants and water bodies including inland marsh, lake marsh, river floodplain wetland, forest/shrub wetland, peat bogs, mangrove and salt marsh, etc.
60	Water bodies	Water bodies in the land area, including river, lake, reservoir, fish pond, etc.
70	Tundra	Covered by lichen, moss, hardy perennial herb and shrubs in the polar regions, including shrub tundra, herbaceous tundra, wet tundra and barren tundra, etc.
80	Artificial surfaces	Modified by human activities, including all kinds of habitation, industrial and mining area, transportation facilities and interior urban green zones and water bodies.
90	Bareland	Vegetation cover lower than 10%, including desert, sandy fields, Gobi, bare rocks, saline and alkaline lands, etc.
100	Permanent snow and ice	Covered by permanent snow, glacier and icecap.

### 3 NATIONAL VEGETATION CLASSIFICATION PROCESS

#### 3.1 Classification procedure for national vegetation

National time series data is divided into 8 areas in the vegetation partition map. To determine the category and phenological period, information collection and vegetation analysis are carried out for each area according to the time series curve, rules are established according to the analysis results and the time series curve, and then the regional vegetation coverage diagram is figured out according to rules; finally, the results of 8 areas are integrated and the classification map of national vegetation form is obtained. The classification procedure is shown below.

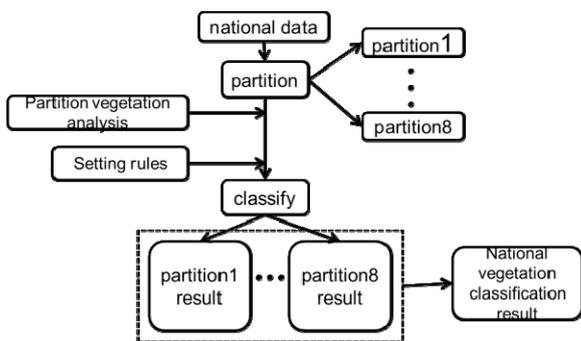


Fig.2 Classification procedure for national vegetation

#### 3.2 Regional classification system

The temperate grassland, which mainly consists of forest land, cultivated land, grassland and bare land, is taken as an example. Through observation and analysis, layering stripping classification method is adopted. The maximum value of bare land and vegetation is easier to distinguish; first, extract the bare land and remove it from the time series curve through masking, and the rest is the time series of vegetation. Due to an earlier growth cycle, the forest land can be extracted and stripped, and distinguished in the rest cultivated land and grassland according to growth situation. The flow chart of classification is shown below.

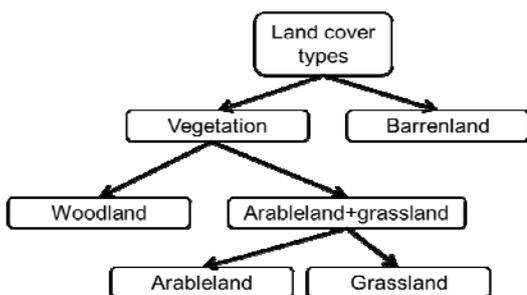


Fig. 3 Flow chart of the temperate grassland

#### 3.3 Classification rule

It is observed that the land type in this area is unimodal with different growth cycles and situations. The peak value of the bare land is the lowest, lower than 2500, while the peak values of other land types are higher than 2500. The forest land has a higher peak value and an earlier growth cycle, and forests grow tall on the 145th day (the last ten-day of May). Thus, the forest land can be extracted when the peak value is higher than 8000 and when it is higher than 5000 in the last ten-day of May. With similar growth cycle, the growth situation of the cultivated land and the grassland is different, especially in the last ten-day of July, so they can be distinguished when the peak value is higher than 5000 in the last ten-day in July. After the above extraction, the time series of other areas can be obtained through masking and land types will not be interfered. The time series curve of main land types in this area is shown below:

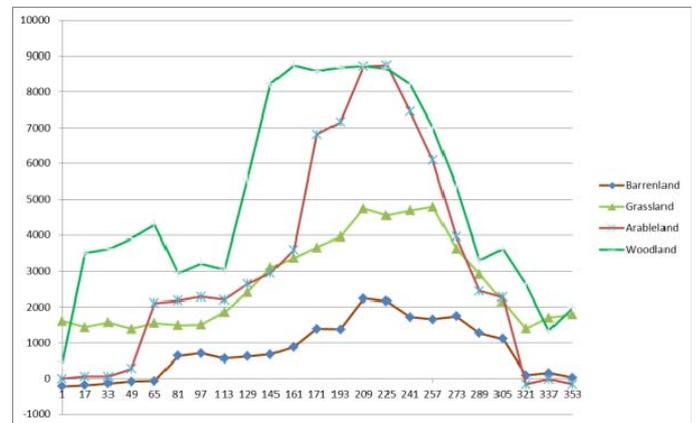


Fig. 4 Time series for temperate grassland land type

### 4 CLASSIFICATION RESULT AND PRECISION EVALUATION

Through the above analysis and regulations, the paper has carried out classification and gained divisional classification results.

Table 2 Confusion Matrix of this Product

Correct category \ Category herein	Arable land	Wood land	Grass land	Total
Arable land	268	22	10	300
Woodland	20	276	4	300
Grassland	22	5	273	300
Total	310	303	287	900

In addition, the final classification is generally evaluated. To make evaluation more object and accurate, the paper automatically and randomly collects sample points of each category. Furthermore, it generates 300 random sample points respectively for

three categories including woodland, arable land, and grassland, and distributes them to different people to compare Google earth for examination. Then the following confusion matrix is gained.

As a result, the overall accuracy is 90.78%, and Kappa coefficient is 0.86. From the previous part, it can be seen that five global products, which have been widely applied at present, present a relatively low precision evaluation result in China. The highest result is GLC2000, while the overall accuracy is 59.28%. Thus, precision result herein is relatively high with better applicability.

Then the paper compares the above result with modis product of the same year. Since modis product differs from product herein in category code, modis product is reclassified so that it have the same category and have same category code with this product. Then inconsistent areas are generated after subtraction between the two products. In that way, there might be multiple possibilities for inconsistent area code of each category. For instance, if the category code after minus is 10, there will be three possible situations: this product is grassland (code 30) while modis product is woodland (code 20); this product is woodland while modis product is arable land (code 10); this product is arable land while modis product is others (code 0). Therefore, each situation of inconsistent area has to be extracted and evaluated again. Then there will be an inconstant area result with specific explanations. According to area proportion and amount of actual pixels of reevaluated categories, the paper sets sample numbers of each category, and carries out even sampling according to sample points. Finally, precision evaluation table of this product is generated:

Table 3 Precision Evaluation Table of this Product

correct herein	Arable land	Wood land	Grass land	Town	Scrub	Water	Bare area
Arable land	79	10	7	2	1	1	0
Wood land	18	87	9	0	2	0	0
Grass land	6	3	77	0	0	0	8

Table 4 Precision Evaluation Table of Modis

correct modis	Arable land	Wood land	Grass land	Town	Scrub	Water	Bare area
Arable land	22	31	19	3	3	19	8
Wood land	14	9	20	0	2	0	0
Grass land	50	44	18	0	0	0	28

Statistics are acquired according to the above two evaluation tables. As for differences, arable land ac-

curacy of this result is 79% while the accuracy of modis is 20.95%; woodland accuracy of this result is 75% while the accuracy of modis is 20.93%; grassland accuracy of this result is 81.91% while the accuracy of modis is 12.86%. It can be seen that this product has more advantages than modis product. The following are charts of this classification result and modis product result:

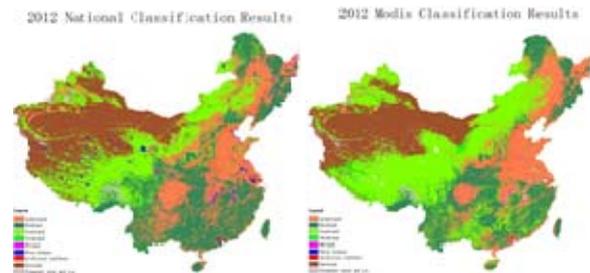


Fig.5 (a) 2012national and coverage result Fig.5 (b) 2012 modis classification result

## 5 CONCLUSION

By comparing verification between this product and other products, it can be concluded that classification results gained from targeted divisional plan and time series can lead to a higher precision. Through comparison with products in China, such as GLCC, UMD, GLC2000, MODIS LC, GlobCover, result in this paper has a higher accuracy and applicability. Compared with modis product of the same period, precision in the paper has been greatly increased. Therefore, this classification method can be taken as reference for follow-up global classification, while this result can be adopted in classification application.

## ACKNOWLEDGEMENT

This work was supported in part by the National High Technology Research and Development Program of China under Grant 2013AA12A301, in part by the Chinese Academy of Sciences Action Plan for West Development Project under Grant KZCX2-XB3-15.

## REFERENCES

- Hansen, M.C., Defries, R.S., Townshend, J.R.G. and Sohlberg, R. 2000. "Global land cover classification at 1 km spatial resolution using a classification tree approach". International Journal of Remote Sensing, 21, pp. 1331-1364.
- Loveland, T.R., Reed, B.C., Brown, J.F., etc. 2000. "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data". International Journal of Remote Sensing, 21, pp. 1303-1330.

- Scepan, J. 1999. "Thematic Validation of High-Resolution Global Land-Cover Data Sets." *Photogrammetric Engineering and Remote Sensing*, 65: 1051–60.
- Yan Bai, Min Feng, Hao Jiang, etc. 2014. "Assessing Consistency of Five Global Land Cover Data Sets in China". *Remote Sensing*, 6:8739-8759.
- Youhua Ran, Xin Li and Ling Lu.2010. "Evaluation of four remote sensing based land cover products over China". *International Journal of Remote Sensing*, Vol.31, No.2, 391-401, 20 January.