A New Method of Vegetation Classification Based on Temporal Distribution of Vegetation Indices

Xiumin Zhang, Zhuotong Nan *, Yu Sheng, Lin Zhao, Jichun Wu
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China

Guoying Zhou
Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001, China

***Abstract*—Vegetation classification is the basis for various hydrological and ecological studies. Some vegetation classifications using NDVI data have been studied commonly, however, there are some limitations in classification of vegetation using the average NDVI value, especially, the study with spectrum characteristics of the NDVI is relatively rare. Based on this, this paper introduces a new vegetation classification method which combines the phase with the NDVI distribution on the basis of computing the probability. Firstly, the probability of each vegetation type in a period was calculated with two-dimensional random variable conditional probability. Then we synthesized the image at germination, maturity and senescence periods, and the probability was calculated according to the fact that various vegetation types may have different biometric information at different periods. At last, vegetation classification map was finished on the probability value. The Wenquan area over the Qinghai-Tibet plateau (QTP) was taken as an experimental area. The result showed that the overall accuracy and kappa coefficient was 70% and 0.60, respectively. So the vegetation classification method can classify and discriminate vegetation effectively.**

Keywords- *spectral analysis; vegetation; time domain analysis; probability computation*

Ⅰ. INTRODUCTION

Vegetation as an important ecological factor is sensitive to the change of ecological environment and climate [1]. Recently, with the development of the satellite remote sensing technology, there is a great progress on classification of regional and global vegetation. Especially, normalized difference vegetation index (NDVI) as an important vegetation physical parameter has been applied widely in many research fields, such as vegetation classification, relationship between vegetation and climate change and so on [1-5].

In recent years, based on different methods of vegetation classification, scholars have used NDVI data in different regions. Traditional methods are mainly supervised and unsupervised classification method. Supervised and unsupervised classification method only based on spectral data of remote sensing image to classify without considering the phenology of vegetation. Nowadays, some new methods are applied to vegetation classification, such as decision tree technology [6], object-based classification [7] and fuzzy classification [8-9]. However, average NDVI of each vegetation type was used by those typical vegetation classification methods. In fact, even though averaged NDVI profiles of different vegetation types differ from each other, the realistic NDVI distributions remain overlapped so that we cannot rely on averaged NDVI for classification [10]. Especially, the classification of vegetation types with similar NDVI spectrum characteristics was studied by few scholars. In other hand, different phenological characteristics of different vegetation types were showed in different growth periods. Thus phase information plays an important role in the vegetation classification. The high intra-class variability may lead to low separability at different composite periods and suggests a classification approach taking advantage of NDVI probability distribution should be developed [10]. So we try a new classification approach which combines the phase with the NDVI distribution on the basis of computing the probability to finish the vegetation map of Wenquan area in QTP through the analysis of NDVI spectrum characteristics and vegetation phenology.

Ⅱ. STUDY AREA

The Wenquan area is located in the southeastern part of the QTP. Administratively, it extends across four counties, Xinghai, Maduo, Maqin and Dulan of the Qinghai province, western China (Fig.1.b). Elevations in the study area range from 3430 to 5300m above sea level, with an average of 4327m. Two high and steep mountain ranges, the Ela Mountains and Jiangluling Mountains, are situated in the study area with the northwest-southeast direction. The Qing-Kang road traverses through the area in a northeast-southwest direction. There are two basins, the Wenquan basin and the Kuha basin, with lower elevations and flat terrains (Fig.1.a). Nearby meteorological Huashixia station reports average annual temperature of -3.2℃ and annual precipitation of up to500~600 mm in this area. According to the field investigation in 2009, the area is dominated by three vegetation types, i.e., alpine grassland, alpine meadow, and alpine swamp meadow, with a small portion of alpine shrub.

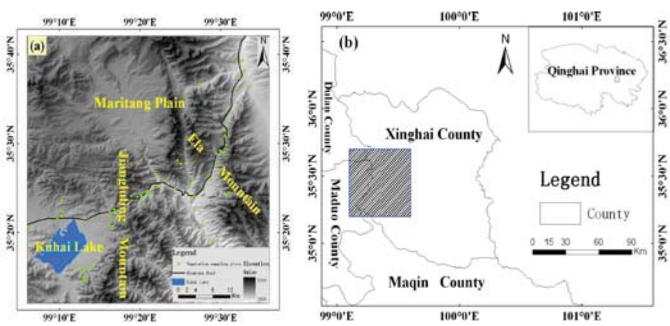

Figure 1.The terrain (a) and location (b) map of the study area

Foundation item: the Specific Fundamental Research Program of China (2008FY110200) and the High-tech Plan of MOST (2008AA12Z205).)

Corresponding author: Zhuotong Nan

E-mail：nztong@lzb.ac.cn

978-1-4244-9170-4/11/$26.00 ©2011 IEEE

III. DATA AND METHOD

A. Field Investigation and Remote Sensing Data

A comprehensive vegetation investigation in the Wenquan area was carried out from September 12 to October 7, 2009. In order to validate MODIS NDVI data which represent an average condition of a 250m×250m cell, single plots were specifically selected in those sites where a simple vegetation type dominates. We set up a total of 99 vegetation transects in the area. Each vegetation transect was 50m long and extended along topographic gradients. 5 to 10 quadrats, each with a size of 1m by 1m, were randomly selected with an interval of 5 to 10m.Vegetationtypes, community structures, species, coverage, and average plant height were recorded for each Quadra whose geographic coordinates were logged with a GPS receiver. Among the total 99 transects, 49 are with alpine meadow, 30 with alpine grassland, 16 with alpine meadow marsh, and 4 with alpine shrub (see green dots on the Fig .1.a).

In this study, MODIS NDVI datasets (MOD13Q1) were acquired from the NASA WIST website. There were totally 223 16-day composite MODIS NDVI data for a 10-year period from March 5, 2000 to October 31, 2009. For the ease of use, NDVI data were projected into the WGS84 lat/lon reference system from their original sinusoidal projection and also converted to TIFF format form original HDF format.

B. Multi-temporal NDVI profilesy and Range Analysi sof the Vegetation

The multi-temporal averaged NDVI profiles of major vegetation types in the study area were presented in Fig.2. Similar phenological characteristics during the entire growth season of all vegetation types were observed. The vegetation phenology was determined by the curves of Multi-temporal NDVI profiles [11]. According to Fig.2, three points generally defined as germination, maturity and senescence. For all vegetation types, photosynthetic activity obviously starts in late May and early June with a rapid NDVI increases on around May 24 and June 9. NDVI values reach their peaks in the mid-summer (i.e. August 12). Senescence phases of all vegetations occur in late October as we saw rapid NDVI decreases in around October 15 and October 31.

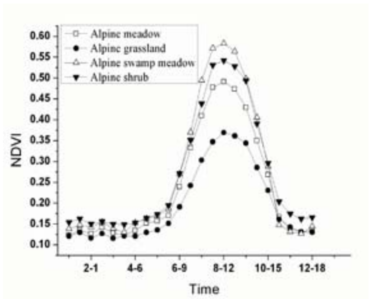

Figure 2. Multi-temporal averaged NDVI profiles of the major vegetation types

In order to much more clearly explain intra-class separability, a range analysis was conducted. As shown in Fig.3, different curves represent NDVI profiles from different sample plots and the vertical lines indicateing the variance of NDVI at given time points. For each vegetation type, greater variance of NDVI profiles was observed in the growth season, and smaller variance in dormancy periods. A maximum variance of NDVI was observed at the maturity time. Even though averaged NDVI profiles of different vegetation types differ from each other (Fig.2), the realistic NDVI distributions remain overlapped so that we cannot rely on averaged NDVI for classification. For example, in the maturity period, NDVIs of alpine meadow range from 0.15 to 0.70; NDVIs of alpine grassland range from 0.18 to 0.57. Given a NDVI of 0.50 at the maturity period at a loca tion with unknown vegetation type, it is hard to determine its vegetation type as this NDVI actually falls within all the four types, whereas according to averaged values (Fig.2) it might be of alpine meadow.

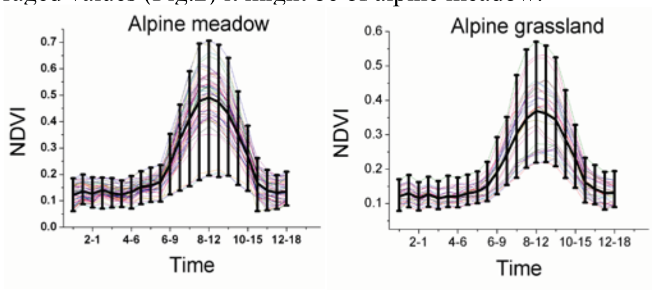

Figure 3. Range analyses with NDVI data of alpine meadow and alpine grassland

C. A New Method Based on Vegetation Indexs Data's Distribution

It was suggested in Fig.3 that a classification approach taking advantage of probability distribution instead of using averaged value should be developed. In other words, the probability values of vegetation types may be different in one NDVI interval, and then, the vegetation map would be finished on the different vegetation's probability value.

Firstly, in a certain phase, the value of NDVI was divided into some intervals. According to the condition that two types were in one interval at least, the number of intervals was decided by the user and expert's knowledge. Then the sample number of each vegetation type was statisticed in each NDVI interval. Finally, the probability of each vegetation type in each interval at a certain period was calculated according to the method of the two-dimensional random variable conditional probability. The probability of each vegetation type in an interval was given by [12]:

$$\mathrm{P}\{\mathrm{X}=\mathrm{i} \mid Y=j\}=\frac{P_{ij}}{P_{.\mathrm{j}}} \tag{1}$$

$$P_{\mathrm{ij}}=\frac{n_{ij}}{N_i} \tag{2}$$

$$P_{.j}=\sum_{i=1}^{4} P_{\mathrm{ij}} \tag{3}$$

Where X refers to the variables about vegetation types, Y denotes the NDVI intervals, i is one vegetation class, j represents a NDVI interval, n_{ij} representatives the number of sample point about i vegetation class in j interval, N_i refers to the number of all samples point about i vegetation class.

The growth of vegetation experienced different periods with the changes in physiology, appearance, structure and so on; their changes also brought out the corresponding changes of NDVI. Therefore the accuracy of classification could be improved with using multi-temporal remote sensing images According to Fig.2, three points generally defined as germination, maturity and senescence. The images of three periods were synthesized when considering the fact that various vegetation types might have different biometric information at different periods. And then the probability of vegetation classification about the synthesized image was computed basing on the formula of the conditional probability [12]. The prerequisite was that the probability of each vegetation type as an independent event before computing the probability. The probability of each vegetation type about the synthesized image was given by [12]:

$$\mathrm{P}(x_i \mid y_1 y_2 y_3) = \frac{\mathrm{P}(x_i \mid y_1)\mathrm{P}(x_i \mid y_2)\mathrm{P}(x_i \mid y_3)}{\mathrm{P}(\mathrm{x_i})\mathrm{P}(\mathrm{x_i})} \quad (4)$$

Where, x_i denotes some vegetation types; y_1, y_2 and y_3 response to germination, maturity and senescence respectively.

The probability of each vegetation type was calculated through the formulas above. The probability of each pixel may be different due to the different phenology of three periods. According to the probability values of several vegetation types in a pixel, the category of vegetation was finished. In other words, the classification of one pixel was to set the vegetation classification with the maximum probability, when the probability value was zero, the classification of the pixel was defined non-vegetation. This method can reflect actual distribution of vegetation with considering phase information and practical distribution characteristics of NDVI. It also makes full use of different phenology of the different vegetation in three periods.

Ⅳ. RESULTS AND ANALYSIS

A. The Results of New Classification Method

According to"(1) ~ (3)", probability value of each vegetation type was calculated in the divided NDVI interval at the period of germination, maturity and senescence (Fig.4). In Fig.4, horizontal coordinate denoted the NDVI intervals and perpendicular coordinate was the probability values. In each period, each vegetation type had different distribution curves in the same NDVI interval. Although the probabilities of two vegetation classes were the same in some NDVI interval, the probability of two vegetation type about the synthesized image may be different. About the probability of each vegetation type distributed in space, the probability value of alpine meadow and alpine grassland in period of maturity were taken as examples (Fig.5). The color from red to green denoted that the probability value was higher.

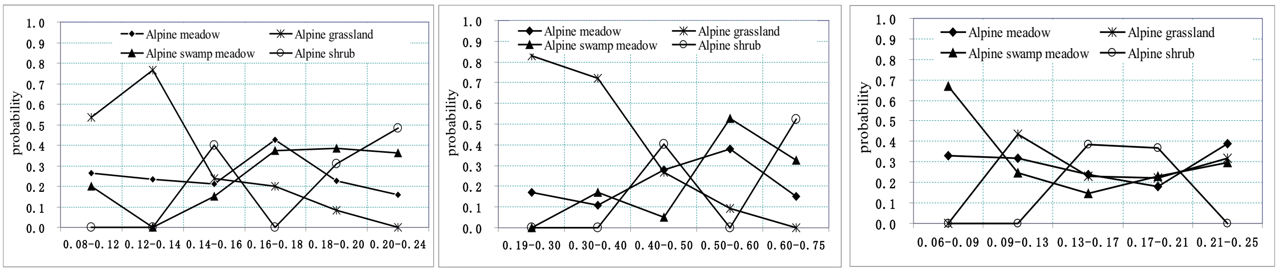

Figure 4. Probability distribution of four vegetation types in different NDVI intervals in three different periods (germination, maturity and senescence)

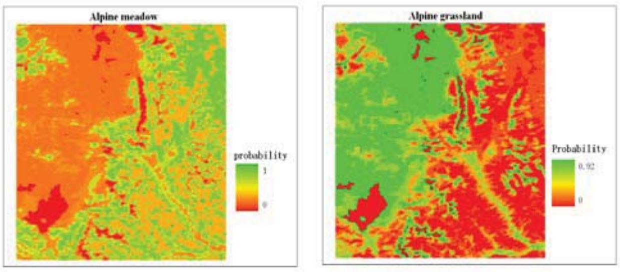

Figure 5. Probability map of alpine meadow and alpine grassland in the maturation period

According to "(4)", the probability of four types was calculated about the synthesized image. The probability value of alpine meadow and alpine grassland at the period of maturity were taken as examples (Fig.6). Compared with Fig.5, the probability distribution of alpine meadow and alpine grassland had changed a lot due to different NDVI physiological characteristics considered when synthesizing the different stages of multi-temporal image. NDVI as an important physical parameter can reflect the group characteristics of vegetation. A multi-temporal image contained the characteristics of phenology of the same vegetation in different phases, so it could reflect the distribution of vegetation better.

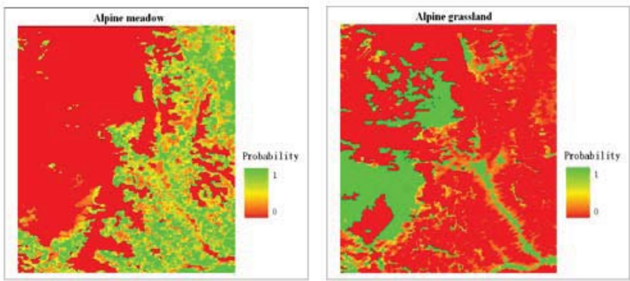

Figure 6. The composite probability distribution of alpine meadow and alpine grassland based on the germination, maturity and senescence

Through above work and analysis, the maximum synthesis algorithm was used for the vegetation of Wenquan area. According to probability values of the four types on the same pixel, vegetation type of the pixel was set to the vegetation type with the maximum probability value, if the probability value of a pixel equal to 0 and then the pixel was set to non-vegetation. Under the condition of this discriminated function, the vegetation type's map was finished (Fig.7).

Under the influences of permafrost and topography condition on the vegetation, alpine meadow and alpine swamp meadow are distributed mainly in area of low mountains and depressions area where there are water; alpine grassland is mainly distributed in high plain; Because evaporation and solar radiation intensity are weak in the north slope, alpine shrub is distributed mainly in the north slopes of the Ela Mountains and the Jiangluling Mountains (Fig.7).

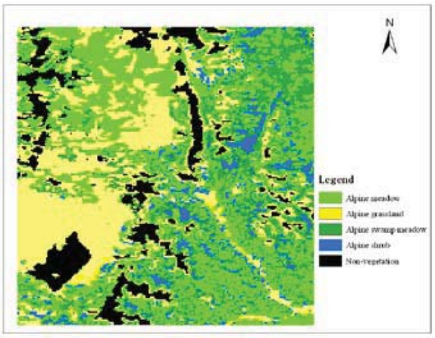

Figure 7. Vegetation classification map in the Wenquan area

B. Precision Evaluation

Accurate evaluation during classification using Remote sensing data is an indispensable work. Combining with the data of visual interpreting on TM images and the investigation, 318 points as a validation sample was selected randomly. Error matrix was calculated and shown in table1. In the error matrix, the total classification accuracy was 70% and Kappa coefficient was 0.61. From the aspect about the production precision of each vegetation type, the classification accuracy of alpine grassland was the highest and the alpine shrub's accuracy was the minimum. When it came to user precision of each vegetation type, the classification accuracy of alpine swamp meadow and alpine shrub were lower than other vegetations. Two reasons might contribute to it, Firstly, the sample points of alpine swamp meadow and alpine shrub was very little and the NDVI interval between them was very close. Secondly, the distribution of sample points of various vegetation types wasn't uniform in space. A number of samples were not selected by the method of simple random sampling, so the classification results were not idea.

T ABEL Ⅰ. RECISION ESTIMATION OF VEGETATION CLASSIFICATION

classification	alpine meadow	non-vegetation	alpine grassland	alpine swamp meadow	alpine shrub	overall	user precision
alpine meadow	62	1	12	5	6	86	72. 1%
non-vegetation	0	45	0	0	0	45	100%
alpine grassland	15	1	60	4	0	80	75%
alpine swamp meadow	20	0	5	31	10	66	47%
alpine shrub	6	0	2	9	24	41	58.5%
overall	103	47	79	49	40	318	
production precision	60. 2%	95.7%	76%	63. 3%	60%		

Ⅴ. CONCLUSIONS

A new vegetation classification method which combined the phase and the NDVI distribution on the basis of computing the probability was used to finish the vegetation classification of the Wenquan area in QTP. The result showed that this method was feasible in the Wenquan area. This method has some advantages over the method using the mean NDVI to identify effectively various vegetation types with similar seasonal rhythms and phonological features. This method uses biological information of vegetation types comprehensively at different periods. Although this method in this study area had achieved good results, the algorithm of selecting parameters also brought out some problems, such as the number of NDVI interval and samples. How to build the relationship between the number of NDVI interval and the vegetation types is our direction in the future. This method also has some limitations which only use the NDVI data, comprehensively use terrain factors (slope, slope, elevation) and other texture images to get better accuracy is our next work.

REFERENCES

[1] J. Y. Chen, and Q. J. Tian, "Vegation classification based on high-resolution satellite image," Journal of Remote Sensing, vol.11(002), pp. 221–227, 2007.

[2] B. D. Wardlow, and S. L. Egbert, "Large-area crop mapping using time-series MODIS 250m NDVI data: an assessment for the US Central Great Plains," Remote Sensing of Environment, vol. 112, pp. 1096–1116, 2008.

[3] R. S. DeFries, and A. S. Belward, "Global and regional land cover characterization from satellite data: An introduction to the special issue," International Journal of Remote Sensing, vol. 21(6–7), pp. 1083–1092, 2000.

[4] R. S. DeFries, and J. R. G. Townshend, "NDVI-derived land cover classifications at a global scale," International Journal of Remote Sensing, vol. 15, pp. 3567–3586, 1994.

[5] D. W. Brian, L. E. Stephen, and H. K. Jude, "Analysis of time-series MODIS 250m vegetation index data for crop classification in the U.S Central Great Plains," Remote Sensing ofEnvironment, vol. 108, pp. 290–310, 2007.

[6] C. C. Yang, S. O. Praster, P. Enright, et al. "Application of decision tree technology for image classification using remote sensing data," Agricultural Systems, vol. 76(3), pp. 1101–1117, 2003.

[7] G. Mallinis, N. Koutsias, M. Tsakiri-Strati et al. "Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site, ISPRS," Journal of Photogrammetry & Remote Sensing, vol. 63, pp. 237–250, 2008.

[8] A. Filippi, and J. Jensen, "Fuzzy learning vector quantization for hyperspectral coastal vegetation classification," Remote Sensing of Environment, vol. 100(4), pp. 512–530, 2006.

[9] R. Tapia, A. Stein, and W. Biiker, "Optimization of samping schemes for vegetation mapping using fuzzy classification," Remote Sensing of Environment, vol. 99, pp. 425–433, 2005.

[10] X. Zhang, Z. Nan, Y. Sheng, et al, Analysis of time-series MODIS 250m vegetation index data for vegetation classification in the Wenquan area over the Qinghai-Tibet plateau[A].Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium. Honolulu, Hawaii, USA, pp. 2059–2062, 2010.

[11] B. Reed, and J. Brown, "Issues in characterizing phenology from satellite observations.Proceedings of the international workshop: Use of earth observation data for phonological monitoring," Joint Research Center, Italy, 2002.

[12] Z. Sheng, S. Q. Xie, and C. Y. Pan, "Probability Theory and Mathematical Statistics," Beijing: Advanced Education Press, 2005.