Vegetation classification in eastern china using time series NDVI images

Guifeng Han[*a], Jianhua Xu[b]
[a]College of Architecture and Urban Planning, Chongqing University, 174 Shazheng Rd., Shapingba district, Chongqing, China, 400045; [b]Department of Geography, East China Normal University, Key Laboratory of Geographic Information Science of Education Ministry, 3663 Zhongshan Rd (N)., Shanghai, China, 200062

ABSTRACT

The SPOT/VGT NDVI (S10) time series data of eastern China (1998-2005) are smoothed with two methods, the moving average and the Savitzky–Golay filter, after they are downloaded from the official website of VITO. Then the monthly maximal NDVI images (total 93 images) are extracted from 279 NDVI (S10) images and the Principal Component Analysis (PCA) is applied on the 93 images. There are 3 components that each explains more than 1% of the variance, in which the principal components 1, 2 and 3 explain respectively 93.25%, 2.77% and 1.21% of the variance in the original 93 maximum NDVI images. The principal component 1 is interpreted as the "climate" component, and principal components 2 and 3 are interpreted as the "growth season" and "non-growth season" components respectively. Principal components 1, 2 and 3 are composed to a 3-band color image which is classified into 7 classes (including 18 subclasses) by ISODATA. The overall accuracy of classification in five samples is 83.6%, and the kappa index is 0.82. Finally, the unique intra-annual NDVI curve of each vegetation class is displayed.

Keywords: Time series NDVI dataset, vegetation classification, principal component analysis, unsupervised classification, eastern China

1. INTRODUCTION

Vegetation, both natural and cultivated, covers much of the earth and strongly influences the environment and is an important variable in many earth system processes[1]. Data from satellites provide our first opportunity to monitor the vegetation on earth surface in a systematic repetitive manner. Campbell (2002) indicated that remote sensing provides the only practical means of mapping and monitoring changes in major ecological regions that, although not directly used for production of food, have great long-term significance for mankind[2].

Normalized Difference Vegetation Index (NDVI) is well established in the literature as being a good representation of vegetation growth on land surface. The NDVI is derived from the red: near-infrared reflectance ratio, shown as NDVI=(NIR-RED)/(NIR+RED), where NIR and RED are the amounts of near-infrared and red light, respectively, reflected by the vegetation and captured by the sensor of the satellite. The formula is based on the fact that chlorophyll absorbs RED whereas the mesophyll leaf structure scatters NIR. NDVI is a nonlinear function that varies between-1 and 1, where negative values correspond to an absence of vegetation. Values of NDVI for vegetated land are generally greater than 0.1, and values exceeding 0.5 indicate dense vegetation[3]. Many studies show that the NDVI has strong correlation with vegetation productivity and this relationship is well established theoretically[4] and empirically[5]. Because the NDVI correlates directly with vegetation productivity[6], the NDVI can provide information about the spatial and temporal distribution of vegetation community, productivity and phenology in various ecosystems[7]. In general, the NDVI of vegetation is lower during the Start of Growing Season (SOS) and at the End of Growing Season (EOS), and the value is higher during the growing season. An exception to this is evergreen vegetation, which has a higher NDVI value through all four seasons. Different vegetation communities, however, have different inter-annual variabilities. So time series NDVI image represents the first useful tool with which to display directly the regulars of vegetation intra-annual variability and inter-annual variability. The latter reflects the vegetation responses to environmental change such as climate and soil[8-10].

[*] newseaboy@163.com; phone +86-21-62233861; fax +86-21-62232416

MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications, edited by Yongji Wang, Jun Li, Bangjun Lei, Jingyu Yang, Proc. of SPIE Vol. 6790, 67901N, (2007)
0277-786X/07/$18 · doi: 10.1117/12.749124

Different NDVI datasets are available, with different spatial and temporal resolutions, and different temporal coverage. Since the 1980s, pioneering researches have been conducted to map and monitor vegetation on continental scales using NDVI data acquired by the U.S. National Oceanographic and Atmospheric Administration's (NOAA) meteorological satellite, the Advanced Very High Resolution Radiometer (AVHRR)[11-14]. But the common disadvantage of their studies is that they only use short multi-temporal NDVI images, for example, one year or longer, i.e. the information of the vegetation inter-annual variability is not considered. The aim of this research is to analyze spatio-temporal variations of NDVI from 1998 to 2005 and classify vegetation of rapid urbanization area in eastern China using SPOT/VGT NDVI time series images which have better quality than NOAA/AVHRR NDVI dataset.

2. DATA AND METHODOLOGY

2.1 Study area

Eastern China is selected as study, located at a longitude from 113°E to 123°E and latitude from 21.5°N to 35.5°N. The study area, approximately 903,717 km^2, contains Shanghai city, Jiangsu province, Zhejiang province, Anhui province, Fujian province and Jiangxi province (Figure 1). The climate is defined as a sub-tropical humid monsoon climate with four distinct seasons. Besides crop, the dominant vegetation of this area is evergreen broadleaf vegetation and a mixture of evergreen broadleaf vegetation and deciduous broadleaf vegetation. There are large areas of sub-tropical evergreen broadleaf forest clustering in the south of this study area such as Zhejiang province, Fujian province and Jiangxi province[15]. The study area has 38.33% of China's total population and 50.15% of China's total gross domestic product (GDP) in 2000. At the same time, population density is very high and the urbanization speed is very fast in this area. The study area includes part of south-north transition in eastern China which was selected as the fifteenth international standard transition in 2000 in order to study the change of terrestrial ecosystems and global climate and the relationship between socioeconomic development and land use/land cover[16].

2.2 Data

SPOT/VGT-NDVI (1 kilometer resolution) datasets are used in this study, because it has four channels designed specially to monitor vegetation, and has higher precision on spatial resolution than NOAA/AVHRR NDVI dataset[17]. The first VGT instrument is onboard SPOT4, was launched in March 1998. It is equipped with four spectral bands: b0 (blue, 0.43–0.47 Am); b2 (red, 0.61–0.68 Am); b3 (near-infrared, 0.78–0.89 Am); and SWIR (short wave infrared, 1.58–1.75 Am). Synthetic products on a daily (S1) and 10-day basis (S10) have been prepared by the Centre de Traitement des Images VEGETATION (CTIV or VITO), the production center located in Belgium, and further provided to us by JRC. These products are composed using the maximum value of the NDVI within the temporal window of composition. This is based on the traditional Maximum Value Composite (MVC) technique aimed at minimizing simultaneously the impact of atmosphere and surface anisotropy effects to reveal only clear and more standardized scenes[18-19]. We downloaded SPOT/VGT-NDVI (S10) and Satus Maps (SM, S10) time series data from the VGT internet site (http://free.vgt.vito.be) for the period from April 1998 to December 2005 (total 279 NDVI images and total 279 SM images). A subset was created using the CROPVGT program for the study area.

Some noises still present in these images, although NDVI datasets are generally well-documented and quality-controlled data sources and are synthesized in 10-day intervals based on the MVC in order to remove some effect resulted from atmosphere condition and topography. Therefore, time-series NDVI images need to be smoothed in order to display real vegetation growth before being used. Firstly, moving average is used to replace the false value. For the value on a specific point in a specific period, if it is labeled false value according to the information derived from Satus Map, it will be replaced with average values of 2 dekads before and 2 dekads after this point, that is 4 adjacent points' values on the point. If it is not labeled false value, the original value is retained. In addition, values of points with a random NDVI increase more than 0.6 during 2 dekads are also replaced by average value using the 4 adjacent points' values, as such increase cannot be caused by natural vegetation change[20]. Values of time series NDVI in same point should follow the gradual process of the annual vegetation cycle, so sudden falls that are not compatible with the process can be regarded as noise. The Savitzky–Golay (SG) filter is used to smooth the NDVI noise values and obtain a new time series NDVI data recovering long-term trend of vegetation activity. Through experiments on random samples, two best parameters are chosen for SG filter: the degree is 4 and span is 7. The processing is accomplished in ArcGIS and Matlab software with a set of codes. Flowchart of data processing is displayed in Figure 2.

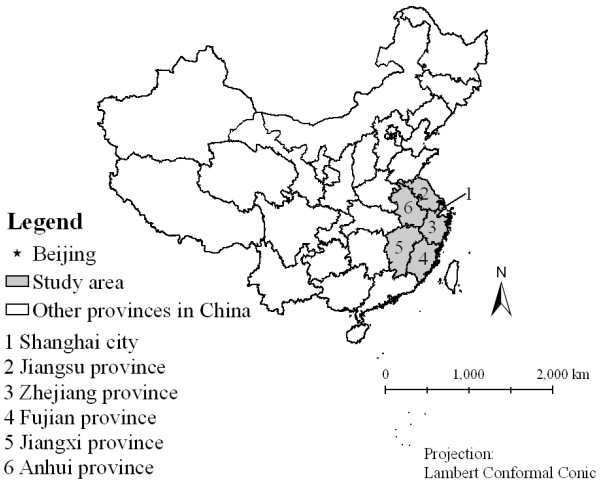

Fig. 1. Study area

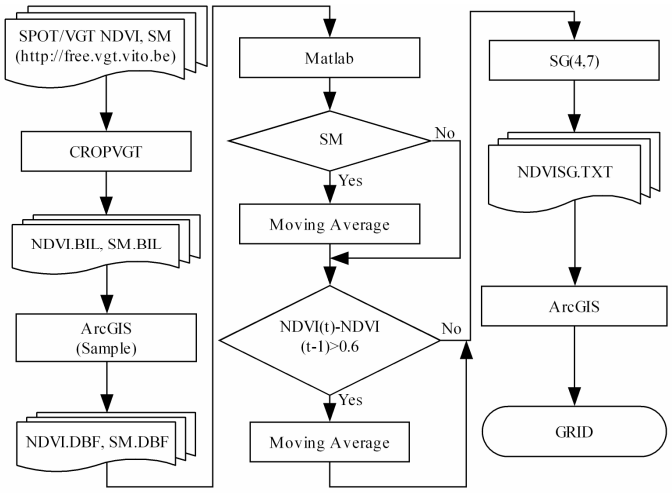

Fig. 2. Flowchart of data processing

2.3 Method

2.3.1 Principal component analysis of time series NDVI images

After smoothed, monthly maximum NDVI are extracted over 3 dekads in every month using MVC, and 93 images are produced. Then the Principal Component Analysis (PCA) is applied to compress data and to produce 93 principal components which include both temporal and spatial patterns. At present, up to 30 images can be analyzed simultaneously in ArcGIS9.0, so PCA for 93 images have to be implemented with modeling in ERDAS software. Several principal components in front carry most of information, and each principal component expresses underlying themes (trends, shifts, periodicities and so on) in the original series[21]. Principal components are uncorrelated with one another and are ordered in terms of the amount of variance. The principal component images indicate the spatial patterns of major elements of variability over the series. Moreover, the correlation coefficients between principal component images and original images are calculated.

2.3.2 Unsupervised classification with ISODATA

The aim of unsupervised classification is to uncover the major vegetation classes that exist in the image without prior knowledge of what they might be. Unsupervised classification techniques search for clusters of pixels with similar reflectance characteristics in a multi-band image. They are concerned with uncovering the major vegetation classes, and thus tend to ignore those that have very low frequencies of occurrence. The Iterative Self-Organizing Data Analysis Technique (ISODATA) as the most common unsupervised classification scheme is applied to classify the 3-band color image which is composed from principal components 1, 2 and 3.

3. RESULTS

3.1 Principal components

After PCA is applied, 93 principal components are produced from monthly maximum NDVI images from April 1998 to December 2005. Table 1 lists the eigenvalues and percentages of variance explained by components in the monthly maximum NDVI images of eastern China. There are 3 components that each explains more than 1% of the variance in the original 93 images. The components 1, 2 and 3 explain respectively 93.25%, 2.77% and 1.21% of the variance in the original 93 maximum NDVI data. Table 1 indicates that the first 3 components explain 97.23% of the total variance. Comparing with similar studies with only NDVI time series images of one year or 12 successive months[22-23,15], results of PCA in this study show that almost all information of original NDVI images concentrate in less principal components. It is also said that longer NDVI time series through appropriate smooth processing can filter effectively out incidental fluctuation and that corresponding principal components can well extract the information of vegetation change both spatial and temporal. Based on the principle of PCA that components with cumulative percentage of variance more than 85% may be selected for sequent analysis[24], only the first 3 principal components are selected to classify vegetation

cover in this study and the information of other 90 principal components can be ignored. Figure 3 displays the images of the first 3 components.

Table. 1. Results for the extraction of Principal Components

Principal Components	Eigenvalue	Percentage of Variance (%)	Cumulative Percentage of Variance (%)	Principal Components	Eigenvalue	Percentage of Variance (%)	Cumulative Percentage of Variance (%)
1	8.65	93.25	93.25	6	0.02	0.17	97.96
2	0.26	2.77	96.01	7	0.01	0.14	98.10
3	0.11	1.21	97.23	8	0.01	0.11	98.21
4	0.03	0.33	97.55	9	0.01	0.10	98.32
5	0.02	0.24	97.80	10-93	-	-	100

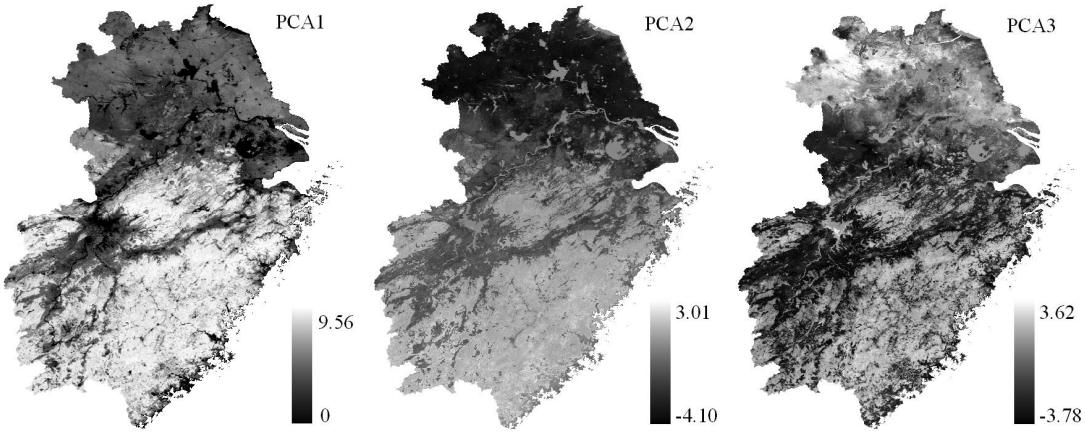

Fig. 3. Images of principal component 1, 2 and 3

Because principal component analysis eliminates the season characteristic of NDVI time series data, it is necessary to interpret the principal components. The first principal component image (PCA1) contains positive values for all cells whose variety is consistent with transition of climate (temperature and precipitation) from south to north. Furthermore, the PCA1 has strong correlation with annual integral NDVI[25], so it can be used to indicate directly vegetation primary product. In this study, the correlation coefficients between PCA1 image and original 93 monthly maximum NDVI images are all positive ranging from 0.85 to 0.96, of which maximal correlation coefficient occurs at February 2003, and minimal correlation coefficient occurs at July 1999 (Figure 4). Therefore PCA1 is interpreted as "climate principal component" indicating climate fluctuation. The correlation coefficients between the second principal component image (PCA2) and original 93 NDVI images range from -0.38 to 0.22. And the correlation coefficient's fluctuation appears obvious regular cycle where correlation coefficient is negative from February to September with minimal troughs in April, and positive correlation coefficients appear from October to January next year. So PCA2 can be interpreted as "non-growth season principal component". The correlation coefficients between the third principal component image (PCA3) and original 93 NDVI images range from -0.51 to -0.17. And the minimal correlation coefficient appears in June, and the maximal one appears in January or February. Based on this fluctuation rule, PCA3 can be interpreted as "non-growth season principal component".

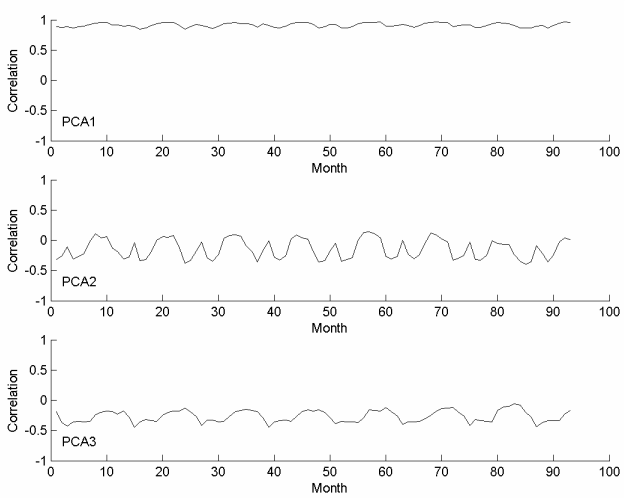

Fig. 4. The correlation between principal components and original NDVI images. (1= April 1998, 50=May 2002, 90= September 2005, 93= December 2005)

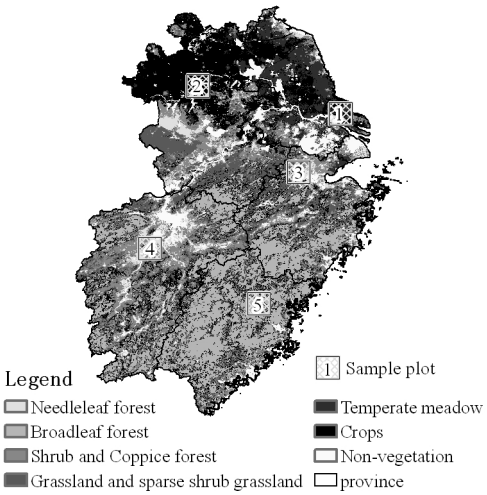

Fig.5. Vegetation classification in eastern China

3.2 Vegetation classification

PCA1, PCA2 and PCA3 are endowed Red, Green and Blue, and are composed into a 3-band color image which contains almost all information of 93 monthly maximal NDVI images. Firstly, this image is classified into 36 clusters using unsupervised classification (ISODATA). This classification statistically separates the entire collection of pixels into homogenous clusters in terms of spectral and temporal characteristics. Obvious false classes are manually corrected based on available field knowledge, ancillary information such as the State Forest Map (1:1,000,000), State Land-use Map (1:1,000,000) and a visual analysis of spatial distribution patterns. In addition to this, small patches are merged to big patches by applying the tool: majority filter with 3×3 moving windows in ArcGIS9.0. Finally, 6 vegetation classes and 1 non-vegetation class[*], with a total of 18 subclasses, are identified according to Chinese vegetation classification system (Figure 5 and Table 2).

Table 2. Each vegetation class' area and classification accuracy

Vegetation Class (subclass Number)	Area (km^2)	Area Percentage (%)	User Accuracy (%)	Producer Accuracy (%)
Needleleaf forest (2)	112308	12.43	84.36	75.45
Broadleaf forest (5)	197379	21.84	84.67	80.93
Shrub and coppice forest (3)	152192	16.84	72.18	78.76
Grassland and sparse shrub grassland (1)	51861	5.74	63.64	79.5
Temperate meadow (1)	42216	4.67	79.17	90.48
Crops (5)	291174	32.22	76.41	82.11
Non-vegetation (1)	56587	6.26	90.23	92.31

It is difficult to validate the correctness of vegetation classification at region scale using coarse resolution remote sensing image. In general, only some sample plots are selected to evaluate accuracy combining limited field survey and records. In this study, five random sample plots (100km×100km) are designed, and each plot contains 100 sample points distributing evenly (Figure 5). Referring to the national land cover map (1:4,000,000), kappa indices in five plots are all more than the standard of kappa index of land cover classification[26], which are 0.87, 0.74, 0.88, 0.79 and 0.85 respectively. The overall accuracy of vegetation classification is 83.6%, and kappa index is 0.82 which indicates that this classification is more satisfying than the classification with NDVI images of one year or successive 12 months.

[*] Non-vegetation contains water and urban which cannot be differentiated because of coarse resolution (1 kilometer) and similar low NDVI value during all four seasons.

Table 2 shows that the 3 largest areas of classes are crops, broadleaf forest, and shrub and coppice forest. From the viewpoint of the user accuracy and producer accuracy, the classification result of non-vegetation is the most distinct as a result of a low NDVI during the entire cycle of the year. Temperate meadow with a producer accuracy of more than 90% has the smallest area percentage of study area, and clusters in the northeast of Jiangsu province. Crops are located widely in the north of the Yangtze River and low lands and flat area in the south of the Yangtze River. Broadleaf forest clusters in mountains area in south of the Yangtze River such as Zhejiang province, Fujian province and Jiangxi province. Needleleaf forest distributes sporadically in the mountain area with high elevation. Other vegetation classes with relative low accuracies distribute widely in the transition zones, for example, from crops to broadleaf forest and from crops to non-vegetation where blending of multi-class vegetation results in many misjudgments.

3.3 Intra-annual NDVI curve of each vegetation class

The mean NDVI value of each vegetation class in each monthly maximal NDVI image can be calculated based on spatial pattern of vegetation class patches. Then the mean NDVI values of each vegetation class in every month from 1999 to 2005 can be calculated. These mean values are averaged over 7 year in order to remove inter-annual variability resulting from climate change and data error. Finally, 7 vegetation classes have their own unique intra-annual NDVI curves (Figure 6). Broadleaf forest (including evergreen and deciduous broadleaf) has the most high NDVI through all the year round, and crops and grassland have two NDVI peaks as a result of two crops cycles in a year. The non-vegetation class including water and city area has very low NDVI values throughout the year. The other 3 vegetation classes have higher NDVI values during the growth season with one peak respectively.

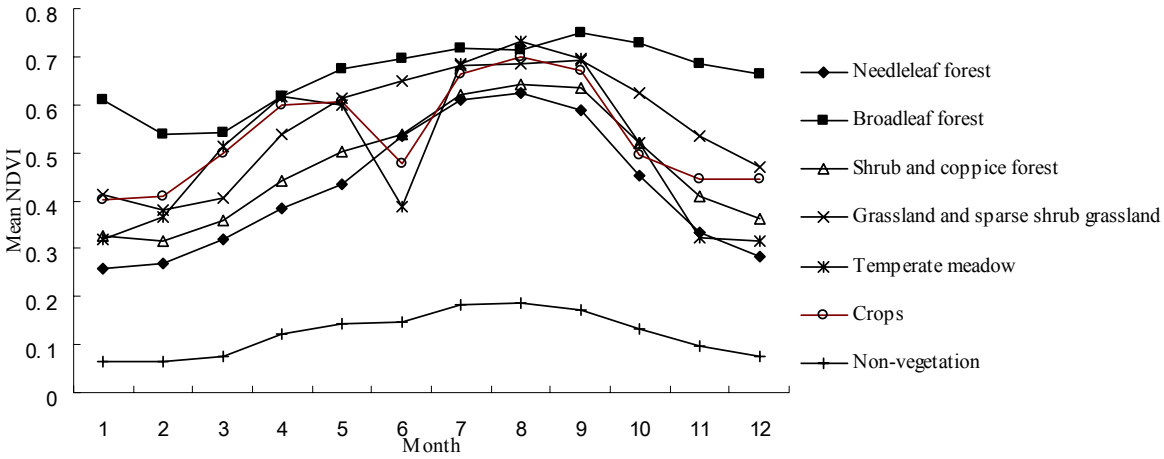

Fig. 6. Inter-annual NDVI curve of each vegetation class throughout all the year

4. CONCLUSION

Time series NDVI images can be well smoothed combining the moving average and SG filter. Based on principal components of longer time series NDVI images, vegetation classification has a satisfied accuracy for coarser resolution remote sensing images at regional scale. Comparing with traditional methods which identify spectral difference on images at only a moment, this classification method is a good innovation in vegetation classification, and can be well applied in other regions.

REFERENCES

1. F.F. Sabins, *Remote Sensing: Principles and Interpretation*, W.H. Freeman and Company, New York, 1996.
2. J.B. Campbell, *Introduction to Remote Sensing*, The Guilford Press, New York, 2002.
3. R.B. Myneni, *et al.*, "The interpretation of spectral vegetation indexes," *IEEE Trans. Geosci. Rem. Sens.* 33, 481-486 (1995).

4. P.J. Sellers, *et al.*, "Canopy reflectance, photosynthesis, and transpiration III: A reanalysis using improved leaf models and a new canopy integration scheme," *Remote Sens. Environ*. 42, 187-216 (1992).
5. G. Asrar, *et al.*, "Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat," *Agricult. J.* 76, 300-306 (1984).
6. B.C. Reed, *et al.*, "Measuring phenological variability from satellite imagery," *J. Veg. Sci.* 5, 703-714 (1994).
7. N. Pettorelli, *et al*., "Using the satellite-derived NDVI to assess ecological responses to environmental change," *TRENDS in Ecology and Evolution*. 20, 503-510 (2005).
8. G.J. Roerick, *et al.*, "Assessment of climate impact on vegetation dynamics by using remote sensing. Phys," *Chem. Earth*. 28, 103-109 (2003).
9. F.F. Yu, *et al.*, "Response of seasonal vegetation development to climatic variations in eastern central Asia," *Remote Sens. Environ*. 87, 42-54 (2003).
10. J.Y. Zhang, *et al.*, "New evidence for effects of land cover in China on summer climate," *Chin. Sci. Bull*. 48, 401-405 (2003).
11. C.J. Tucker, *et al.*, "African land-cover classification using satellite data," *Science*. 227, 369-375 (1985).
12. F. Achard and F. Blasco, "Analysis of vegetation seasonal evolution and mapping of forest cover in West Africa with the use of NOAA AVHRR data," *Photogram. Engin. Remote Sens*. 56, 1359-1365 (1990).
13. F. Achard and C. Estreguil, "Forest classification of Southeast Asia using NOAA AVHRR data," *Remote Sens. Environ*. 54, 198-208 (1995).
14. J.W. Van Wagtendonk and R.R. Root, "The use of multitemporal Landsat Normalized Difference Vegetation Index (NDVI) data for mapping fuel models in Yosemite National Park, USA," *Int. J. Remote Sens*. 24, 1639-1651 (2003).
15. J. X. Li, *et al.*, "Vegetation classification of east china using multi-temporal NOAA-AVHRR data," *Acta Phytoecologica Sinica*. 29 (3), 436-443 (2005).
16. Y. Li, *et al.*, "The temporal-spatial distribution of primary production of forest and cropland in south-north transition, east China based on digital elevation model (DEM)," *Chin. Sci. Bull*. 49(7), 679-685 (2004).
17. P. Mayaux, *et al.*, "Mapping the Forest-Cover of Madagascar with SPOT4-VEGETATION data," http://vegetation.cnes.fr
18. B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data," *Int. J. Remote Sens*. 7, 1417-1434 (1986).
19. K. S. Han, *et al.*, "A land cover classification product over France at 1km resolution using SPOT/VEGETATION data," *Remote Sens. Environ*. 92, 52-66 (2004).
20. J. Chen, *et al.*, "A simple method for reconstructing a high quality NDVI time-series data set based on the Savitzky–Golay filter," *Remote Sens. Environ*. 91, 332-344 (2004).
21. S. W. Sheen, "A Spatial and Temporal Analysis of East Asian Normalized Difference Vegetation Index: 1982-2000," *Geog. Res*. 39, 95-104 (2003).
22. H. Yan, *et al.*, "Bio-implication of Principal Component Analysis to Land Cover Using Multitemporal AVHRR Data," *Remote Sensing Technology and Application*. 16(4), 209-213 (2001).
23. F. Yu, *et al.*, "Land cover classification in China based on the NDVI-Ts feature space," *Acta Phytoecologica Sinica*. 29(6), 934-944 (2005).
24. J.H. Xu, *Mathematic methods in contemporary geography*, High Education Press, Beijing, 2002.
25. S. D. Prince, "Satellite remote sensing of primary production: comparison of results for sahelian grasslands 1981-1988," *Int. J. Remote Sens*. 12, 1301-1311 (1991).
26. L. F. Janssen and F. J. Mvander Wel, "Accuracy assessment of satellite derived band-cover data: a review," *Photogram. Engin. Remote Sens*. 60(4), 419-426 (1994).