CHINESE GEOGRAPHICAL SCIENCE
Volume 9, Number 1, pp.49－56, 1999
Science Press, Beijing, China

INTEGRATED VEGETATION CLASSIFICATION AND MAPPING USING REMOTE SENSING AND GIS TECHNIQUES

Zhuang Dafang(庄大方)　　Ling Yangrong(凌扬荣)
Institute of Remote Sensing applications, the Chinese Academy of Sciences. Beijing 100101, *P.R.China*
Yoshio Awaya
Forest and Forestry Production Research Institute. Japan
(Received 4 May 1998)

ABSTRACT: NOAA-AVHRR data have been more and more used by scientists because of its short temporal resolution, large scope, inexpensive cost and broad wave bands. On macro and middle scale of vegetation remote sensing, NOAA-AVHRR possesses an advantage when compared with other satellites. However, because NOAA-AVHRR also possesses problem of low resolution, data distortion and geometrical distortion, in the area of application of NOAA-AVHRR in large scale vegetation－mapping, the accuracy of vegetation classification should be improved. This paper discuss the feasibility of integrating the geographic data in GIS(Geographical Information System) and remotely sensed data in GIS. Under the environment of GIS, temperature, precipitation and elevation, which serve as main factors affecting vegetation growth, were processed by a mathematical model and qualified into geographic image under a certain grid system. The geographic image were overlaid to the NOAA-AVHRR data which had been compressed and processed. In order to evaluate the usefulness of geographic data for vegetation classification, the area under study was digitally classified by two groups of interpreter: the proposed methodology using maximum likelihood classification assisted by the geographic database and a conventional maximum likelihood classification only. Both result were compared using Kappa statistics. The indices to both the proposed and the conventional digital classification methodology were 0.668(very good) and 0.563(good), respectively. The geographic database rendered an improvement over the conventional digital classification. Furthermore, in this study, some problems related to multi-sources data integration are also discussed.

KEY WORDS: NOAA-AVHRR, NDVI(Normal Division Vegetation Index), geographic image, integrated image, remote sensing, supervised classification, GIS

1 INTRODUCTION

Research on global environmental change, especially global vegetation change, has been more and more emphasized by scientists and governments. However, the investigation of vegetation covered in large area need a great deal of fund and labor. As the satellite remote sensing data have been available since early 1980s, these data are being employed towards to the improvement of vegetation classification. NOAA-AVHRR data have been more and more used by scientists because of its short temporal resolution, large scope, inexpensive cost and broad wave bands. On macro and middle scale vegetation mapping, NOAA-AVHRR possessed advantage when compared with other satellites. However, because of the scanning width of NOAA-AVHRR is so large(2800km), the earth's curvature, characteristics and angle of reflection from earth's object and atmosphere as well as the angle of scanner and deviation of sun's height cause a serious effect on the data, NOAA-AVHRR also possessed problem of low resolution, data distortion and geometrical distortion. As the result, the accuracy of vegetation classification should be improved when apply NOAA-AVHRR data for large scale vegetation mapping. In order to improve the single spectral information structure of remote sensing information, scientists have tried to use different methodologies of data integrated analysis. Employing some geographic data as ancillary data, for example, to improve the accuracy of remote sensing. With the development of GIS, scientists recognized that, although remote sensing and GIS are two relatively independent techniques, they are related to each other because of the same study objective. By means of integrating different data in GIS environment. Not only the accuracy of remote sensing information interpretation can be improved, but remote sensing can be served as a data resources for GIS system analysis. In this sense, this two techniques are mutual assisted.

Now, many scientists believed that, an inevitable way to improve the quantification of remote sensing application is to integrate remote sensing data with non－remote sensing data. In this direction, the objective of this program is, by means of integrating remote sensing data with geographic data such as temperature, precipitation and elevation, to improve the simple spectral structure of remote sensing information and the accuracy of macro scale vegetation mapping in test site using NOAA-AVHRR data.

2 DATA COLLECTION AND PROCESSING

2.1 NOAA-AVHRR Data and Correction

The study area is localized in Northeast China, corresponding to the north temperate zone, humid area of temperate zone and semi-humid area. The dominant vegetation covered are forest, grassy marshland and grassland. The NOAA-AVHRR data used are from 14 dates of June 1995, mainly because these were the best available cloud-free scenes for the season. The NOAA-AVHRR scenes were preprocessed for radiometric standard calibration and geometric correction(Achard and D'souza, 1994). After this correction, the NOAA-AVHRR data need to be corrected again by means of geometrical transformation. In geometrical transformation, GCPs(Ground Control Points) must be selected from maps. Fig. 1 is the rectified NOAA－AVHRR image of northeast China.

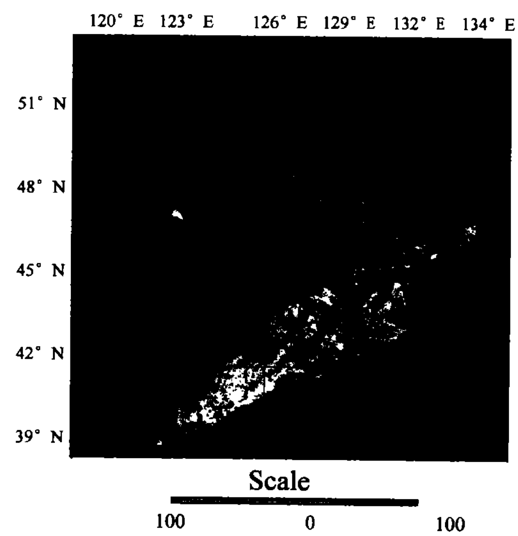

Fig. 1 Rectified NOAA－AVHRR image of northeast China

2.2 NDVI MaximumValue Composites and Optimization of NOAA-AVHRR data

2.2.1 NDVI maximum value composites

The Normalized Difference Vegetation Index (NDVI) is calculated from the measured radiance on the red and near-infrared part of the electromagnetic spectrum. The index is based on the fact that highly vegetated area will have high reflectance on the near-infrared and low reflectance on the red, while water and bare soil will generally have higher reflectance on the red than on the near－infrared part of the spectrum(Gowerd *et al.*, 1991). There are several factors that can cause a considerable change to the NDVI measured by satellite sensors. The most important of these factors are: changing illumination and viewing conditions within a single image and also between images of different days, the presence of cloud cover, variation in atmospheric constituents, in particular the variation in water vapor and aerosol concentration. In this study, we use NDVI Maximum Value Composites(MVCs) to eliminate the undesirable influence of these factors. The use of NDVI MVCs is generally accepted as an effective way of eliminating the undesirable influences (Holben, 1986). The assumption behind the MVC technique is that the maximum value of NDVI a set of images will correspond to the ideal conditions: low solar zenith angle and viewing angle, low concentration of water vapor and aerosols and cloud - free conditions. This assumption is true for a NDVI MVC produced from a large number of images. However, a large number of images per MVC will correspond to long composing periods and the loss of meaningful vegetation biomass change. It is therefore important to achieve a compromise between consistency and frequency of multi-temporal NDVI MVC data. In this study, we select NOAA-AVHRR data from 14 dates of June 1995 for NDVI MVC

2.2.2 Optimization of NOAA-AVHRR data

NOAA-AVHRR have five bands from red to thermal bands. To improve visual effect, we often use colored or false colored image in data processing: choose three bands or channels from multi-band data as red, green and blue channel after data processing. In this study, we would integrate temperature, precipitation and elevation data as one band, NDVI MVC image as the second band. Therefore, we can only employed one band from NOAA-AVHRR for false-colored image processing. Because principal components analysis (PCA) can be used to simplify data processing and compress satellite multi-spectral imagery (Richard, 1986), and one of the most important results of PCA is its ability to change pixel definition from M channel sets of number (counts) to K-principal-component (PC) sets (K < M) without remarkable loss of (relative) information, we use principal components analysis (PCA) to compress multi-spectral NOAA-AVHRR data.

At the image analysis system, principal component analysis of NOAA-AVHRR data acquired during June 1995 in test site was performed after digitally co-registering and merging. Eigenvalues and eigenvectors were subsequently computed. For convenience, comparison of the information content in each principal component (PC) have been titled in accending order (Table 1).

Table 1 Eigenvalues and eigenvectors of Principal Component Analysis

		PC#1	PC#2	PC#3	PC#4	PC#5
Bands	1	0.448	-0.472	0.498	-0.186	0.543
	2	0.369	-0.223	0.246	0.728	0.473
	3	0.320	0.260	-0.508	0.473	0.590
	4	0.727	0.091	0.345	-0.460	-0.363
	5	0.170	0.808	0.561	-0.090	0.037
Eigenvalues		2880.2	297.2	151.3	62.4	32.1
Variance (%)		84.8	8.8	3.7	1.8	0.9
Cumulative variance (%)		84.8	93.6	97.3	99.1	100

A close look at table 1 reveals That the first principal component (PC#1) account for 84.8 percent of the total scene variance. In an other word, PC#1 contained 84.8 percent information of NOAA-AVHRR data. From chart 1 which indicated the distribution of each band of NOAA-AVHRR on PC#1, we can also learn that PC#1 is the positive sum of each band of NOAA-AVHRR. Among them, the contribution of band 4 (10.5 - 11.5μm) is the largest. Therefore, PC#1 have represented the main information of the five bands of NOAA-AVHRR and the humidity and thermal distribution of the ground. It can be used as a representative band for vegetation interpretation and we can use it to co-register with digital image of geographic data and NDVI MVC data.

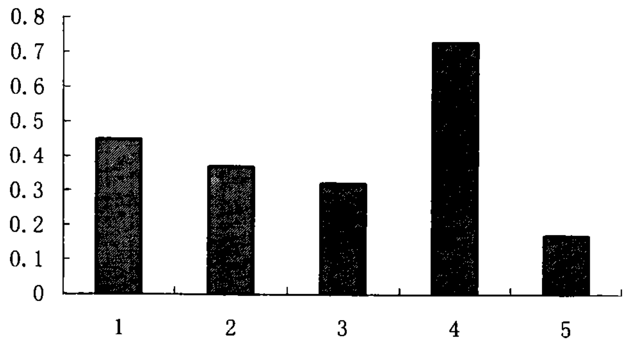

Chart 1 The information distribution of each band of NOAA-AVHRR on PC#1

2.3 The Generation and Processing of Digital Image from Geographic Data

2.3.1 Collection and quantification of geographic data

Temperature, precipitation and elevation are three dominant factors which affect the ground vegetation distribution. Corresponding to NOAA-AVHRR, historical temperature and precipitation data in June from 1991 – 1995 have been collected from local meteorological stations. The mean values of this two kinds of data have also been calculated. Considering the co-registering with NOAA-AVHRR data after quantification, we used 1km * 1km ground grid system and acquired isogram from separated point value of elevation and mean value of temperature and precipitation, then, process the isogram by interpolation and digital quantifying. After transferring the vector data to raster format, we can get three geographic digital images.

2.3.2 The weighting of digital geographic data

In the study area, because there is intern-connection among local temperature, precipitation and elevation, and this connection is related to the vapor path in vegetation growth season, in information integrating, we only employed this three data as one band which we called it integrated geographic image.

In order to get the integrated geographic image, we set value system and perform weighting analysis for this three factors. There are many mathematical processes such as Analytical Hierarchy Process(AHP) and Sequence Synthetical Process for factor-weighting in regional resources evaluation. In this study, we employed AHP for factor – weighting of temperature, precipitation and elevation. AHP is an effective process for quantitatively analyzing un-quantitative factors. In AHP, we firstly compute the maximum eigenvalue and its corresponding eigenvector from assessing matrix of evaluating factors and then get the weights for each factor by means of normalizing the maximum eigenvector. The following list the preliminary procedure of AHP:

1) To construct assessing matrix

Given the assessing objective is A and assessing factors are $F = \{f_1, f_2, \ldots, f_n\}$, then the assessing matrix P can be set as:

$$P = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}$$

where f_{ij} is the value which accounts for the relative importance of f_i compared with f_j, the meaning of the value listed in Table 2.

Table 2 The value of f_{ij} and its meaning

Value of f_{ij}	Meaning
1	f_i and f_j possessed the same importance for objective A
3	f_i is a little more important than f_j for objective A
5	f_i is obviously more important than f_j for objective A
7	f_i is much more important than f_j for objective A
9	f_i is extremely more important than f_j for objective A
2,4,6,8	the mean value between 1 and 3, 3 and 5, 5and 7, 7and 9 respectively
$F_{ji} = 1/f_{ij}$	

In study area, when we analyze the relationship between vegetation and temperature, precipitation and elevation, we can learned that: temperature is a dominant factor affecting the vegetation growth; precipitation affects the vegetation growth coordinated with temperature; the influence on vegetation growth from elevation is somewhat less important than that of temperature and precipitation. Therefore, according to AHP, we can define the assessing matrix as follows:

$$\begin{array}{lccc} & F_1 & F_2 & F_3 \\ F_1(\text{temperature}) & 1 & 2 & 3 \\ F_2(\text{precipitation}) & \frac{1}{2} & 1 & 3 \\ F_3(\text{elevation}) & \frac{1}{3} & \frac{1}{3} & 1 \end{array}$$

2) To calculate eigenvalue (d) and eigenvector (X) of assessing matrixare:

$$d = \begin{bmatrix} 3.0536 & 0 & 0 \\ 0 & -0.0268+0.4038i & 0 \\ 0 & 0 & -0.0268-0.403i \end{bmatrix}$$

$$X = \begin{bmatrix} 0.8257 & 0.5674+0.5998i & 0.5674-0.5998i \\ 0.5201 & -0.5060+0.1206i & -0.5060-0.1206i \\ 0.2184 & 0.0624-0.2094i & 0.0624+0.2094i \end{bmatrix}$$

3) To calculate the weight of each factor

by means of normalizing the eigenvector, we can get the weight of each factor:

$$\alpha_1 = \frac{0.8257}{0.8257+0.5201+0.2184} = 0.527$$

$$\alpha_2 = \frac{0.5201}{0.8257+0.5201+0.2184} = 0.333$$

$$\alpha_3 = \frac{0.2184}{0.8257+0.5201+0.2184} = 0.140$$

Then, we can define the integrated geographic image composed with Fig. 2 − Fig. 4 as:

$$G(x,y) = 0.527 \times T(x,y) + 0.333 \times P(x,y) + 0.140 \times E(x,y)$$

$G(x, y)$ can be used as an independent band for multi-resource data integrating.

3 INTEGRATED IMAGE AND VEGETATION CLASSIFICATION

when integrating remote sensing data and geographic data, the accurate co − registration is very important. In this study, we transferred integrated geographic data, NDVI MVC data and PC # 1 at UTM projection, then we integrated them and get the integrated data.

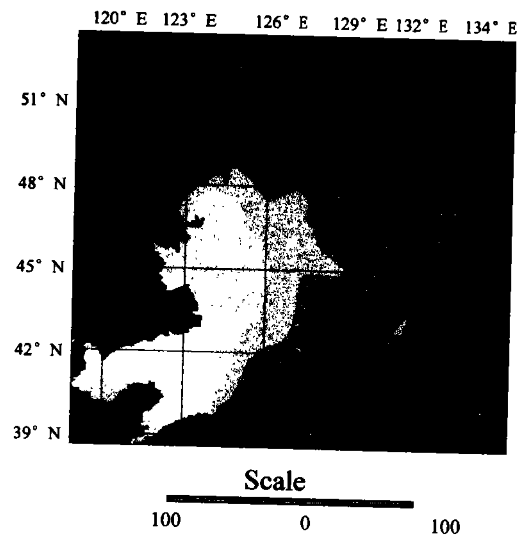

Fig. 2 Multi-year average temperature image of June in northeast China

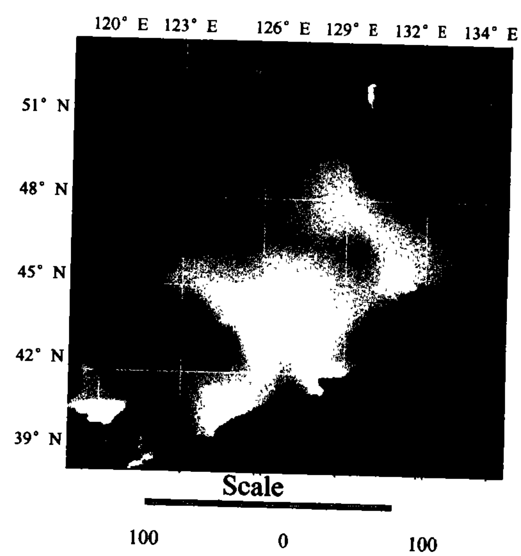

Fig. 3 Multi-year average precipitation image of June in northeast China

In vegetation classification on integrated image, we can regard the geographic data as one channel of multi-spectral remote sensor and all of the layers over laid. In theoretical, we can proved that integrated image is still in conformity with the principle of

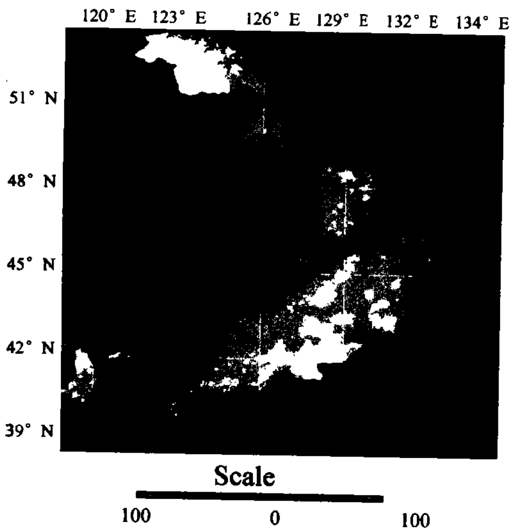

Fig. 4 DEM image of northeast China

Maximum Likelihood Classification, ground samples and published maps in test area have been refereed to select training spot in integrated image. From the classification result, the integrated image can be classified as broadleaf forest, mixed conifer and broadleaf forest, conifer forest, shrub, miscellaneous forest, poplar and birch, farmland, meadow, marsh, grassland, reed marshes and lake twelve kinds of land – cover types.

4 CLASSIFICATION ACCURACY ASSESSMENT

Accuracy assessment is very necessary for classification. In mathematics, the computation of probable error of classification is very complicated. But in practice, we often assess the classification accuracy by means of examining samples and computing the error-matrix from the statistical comparison between each interpretation and the ground data. The error matrix and Kappa statistics of integrated classification is presented in Table 3. In order to evaluate the usefulness of geographic data for vegetation classification, we also used conventional maximum likelihood classification to NDVI MVCs data only, the error-matrix and Kappa statistics are presented in Table 4.

Table 3 Error-matrix for integrated classification assisted by digital geographic data

Training area class classifi. Output	Broadleaf forest	Mixed forest	Conifer forest	Shrub	Poplar & birch	Miscell. f.	Farmland	Meadow	Marsh	Grassland	Reed marsh	Lake	N. of samples
Broadleaf forest	88	8	16	4									116
Mixed forest		16	2										18
Conifer forest	2	8	36										46
Shrub		1		3									4
Poplar & birch	1				2								3
Miscell. forest	1	1				3							5
Farmland							96	20	4	2			122
Meadow							2	56			2		60
Marsh							8		5				13
Grassland							12	2		14			28
Reed swamp									j		5		6
Lake												3	3

The overall kappa statistics = 0.668

From Table 3 and Table 4, we can calculate the classification accuracy of integrated classification assisted by geographic data(R_1) and conventional classification respectively (R_2):

We can also learned from the Kappa statistics that integrated classification obtained the quality index "very good" and the conventional classification obtained the index "good"(Table 5).

5 CONCLUSION

1) The methodology of using multi-temporal remote sensing images, coupled with digital geographic image integrated in a GIS for vegetation classification

Table 4 Error-matrix for the conventional classification

Training area class classifi. Output	Broad forest	Mixed forest	Coni. forest	Shrub	Poplar & birch	Miscell. forest	Farmland	Meadow	Marsh	Grass-land	Reed marsh	Lake	N. of samples
Broadleaf forest	69	14	11	3									97
Mixed forest	7	11	5										23
Conifer forest	7	10	19	3									39
Shrub		4	1	3									8
Poplar & birch	2	1			1								4
Miscell. forest		1	1	1		2							5
Farmland							93	28	13	7			141
Meadow							8	25	3	10			46
Marsh							7	3	9	2			21
Grassland							8	11	2	14			35
Reed swamp								1	1		2		4
Lake												2	2

The overall Kappa statistics = 0.563

Table 5 classification quality associated to a Kappa statistics value(Landis and Koch 1977)

Kappa	Quality
<0.00	worst
0. 00 - 0.2	poor
0.2 - 0.40	reasonable
0.4 - 0.60	good
0.6 - 0.80	very good
0.8 - 1.00	excellent

proved to be better than the conventional digital classification method alone. Not only the global performance index was better(77.1 percent against 58.8 percent) but also the Kappa statistics was assigned as "very good" against "good" respectively.

2) The spectral information structure of remote sensing data can be improved by means of integrating digital geographic data with remote sensing. The integrated image reflected not only the present condition of vegetation but also its internal causes(geographic factors).

3) Although remote sensing and GIS are two relatively independent techniques, they are related to each other because of the same study objective. By means of integrating different data in GIS environment, not only the accuracy of remote sensing information interpretation can be improved, but remote sensing can used as a data resource for GIS system analysis. In this sense, this two techniques are mutual assisted.

REFERENCES

Brain Berry, 1964. Approaches to Regional Analysis: A Synthesis. *Annuals of Association of American Geographers*. 54.

Defnies R. S., Townshend, J. R. G., 1994. NDVI - derived land cover classifications at a global scale. *Int. J. Remote Sensing*, 15, (17).

Donna J. Peuquet, 1988. Representations of Geographic Space: Toward a Conceptal Synthesis. *Annuals of the Association of American Geographers*, 78(3).

Eidenshink J. C., Faundeen J. L. 1994. The 1km AVHRR global land data set: first stages in implementation. *Int. J. Remote Sensing*, 15, (17).

Frederic Achard, Christine Estreguil, 1995. Forest Classification of Southeast Asia Using NOAA-AVHRR Data. *Remote Sens. Environ.*, 54.

Frederic Achard, Christine Estreguil, 1995. Forest Classification of Southeast Asia Using NOAA-AVHRR Date. *Remote Sens. Environ*. 54.

Gutman G. Garik, Aleksandr. M. Ignatov, Steve Orson, 1994. Towards Better Quality of AVHRR Composite Images over Land: Reduction of Cloud Contamination. *Remote Sens. Environ.*, 50.

Gong P., Zheng, X. Shi G. *et al.*, 1994. Large area mapping for global change studies: multitemporal image processing and pattern recognition. GIS'94, Vancouver, B. C.

James M. E., Kalluri S. N., 1994. The Pathfinder AVHRR land data set: An improved coarse resolution data set for terrestrial monitoring. *Int. J. Remote Sensing*,. 15, (17).

Loveland T. R., Mechent, J. W. Brown J. F. *et al.*, 1995. Seasonal land-cover region of the United States. *Annals of the Association of American Geographers*, 85.

Moody A., Strahler, 1994. Characteristics of composed AVHRR data and problems in their classification. *Int. J. Remote Sensing*, 15, (17).

Ortiz M. J., Formaggio, A. R. Epiphanio, J. C. N. 1997. Classification of croplands through integration of remote sensing, GIS and historical database. *Int. J. Remote Sensing*, 18 (1).

Ruth Defries, Mathew Hansen, John Townshend, 1995. Global Discrimination of Land Cover Types from Metrics Derived from AVHRR Patherfinder Data. *Remote Sens. Environ.*, 54.

Dwived, R. S. 1996. Monitoring of salt-affected soils of the Indo-Gangetic alluvial plains using principal component analysis. *Int. J. Remote Sensing*, 17 (10).

Townshend J. G. R., 1994. Global data sets for land applications from the Advanced Very High Resolution Radiometer: an introduction. *Int. J. Remote Sensing*, 15(17).

Tucker, C. J., Townshend, J. R. Goff, T. E. 1985. African land cover classification using satellite data, *Science*, 227.

Townshend J. R. G., Justice, C. O. Kalb, V. T. 1987. Characterization and classification of south American land cover using satellite data. *Int. J. Remote Sensing*, 8.

Tong Lee, John A. Richard, Philip H. Swain, 1987. Probabilistic and Evidential Approaches for Multisource Data Analysis. IEEE *Transactions on Geoscience and Remote Sensing*, GE-25 (3).

Zhu Z., D. L. Evans, 1994. US Forest types and predicted percent forest cover from AVHRR data. *PE&RS*, 15.