http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (21): 6143-6148, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6143.6148

Poyang Lake Wetland Information Extraction and Change Monitoring Based on Spatial Data Mining

[1]Liu Chuanli, [1]Liu Xiaosheng and [2]Liao Qiumin
[1]School of Architectural and surveying Engineering,
Jiangxi University of Science and Technology, Jiangxi, Ganzhou, 341000, China
[2]Faculty of Foreign Studies, Jiangxi University of Science and Technology,
Jiangxi, Ganzhou, 341000, China

Abstract: In order to realize high-precision, automatic abstraction and live, dynamic monitoring of wetland information in Poyang Lake Area, this study researches on RS image classification and change monitoring based on knowledge rules. It first reviews the development of wetland RS monitoring and probes RS classification methods based on Spatial Data Mining; secondly it gives analysis on spectrum features of RS images of Poyang Lake wetland and sets up the Decision Trees Model; and then it classifies RS image by Decision Trees Model in Expert Classifier module of Erdas Imagine 2010 software and achieves a relatively high precision. Finally it gives dynamic change monitoring on Poyang Lake Wetland based on multi-temporal RS image and master the dynamic change to provide scientific proof for government decision making.

Key words: Spatial data mining, wetland, change detection, decision tree, DEM

INTRODUCTION

Wetlands are one of the most valuable ecosystems in the world because of the numerous ecosystem services they provide, including flood attenuation, carbon sequestration, groundwater recharge, water purification and habitats for biodiversity (Kayastha *et al*., 2012). As one of the most international important Wetlands, Poyang Lake Wetland is regarded as the largest “kidney of the mainland” and is listed in globally important ecological areas by WWF (World Wide Fund for Nature) for its important role in maintaining regional and national ecological safety. But the number and quality of wetland is keeping changing due to various kinds of natural and human factors, which has caught great attentions from many sources. Particular attention has been drawn therefore Poyang Lake Eco-economic Region was formally ratified by State Council, because it indicates that the construction of Poyang Lake Eco-economic Region has been exalted as a national strategy.

Now the prevailing method to monitor Poyang Lake wetland is to do on-the-spot survey and visual interpretation of RS image with low efficiency and long term, which contains a lot of human resources and means. What’s more, since the precision is greatly influenced by the experiences and patience of operators, it could not meet the requirements of lively information statistics and dynamic information monitoring in Poyang Lake wetland.

This study gives analysis on the spectrum characteristics of different vegetation in Poyang Lake wetland and abstracts Poyang Lake wetland information based on RS data and expert classification; meanwhile, it monitors the information change lively and dynamically based on Multi-temporal RS data and spatial data mining so as to provide scientific proof for monitoring, administrating and protection of Poyang Lake wetland. Also, monitoring by RS based on Spatial Data Mining is applicable in many fields such as eco-environment and land use change monitoring, which also provide scientific basis for decision-making in environment protection and economic development.

APPLICATION OF REMOTE SENSING IN WETLAND STUDY

Remote sensing is a cost efficient tool to identify and monitor wetlands over a large area. As remote sensing has synoptic coverage and better spatial, spectral and temporal resolutions, it is an effective tool to collect information on natural resources (Prabhadevi and Reddy, 2012).

Foreign application of RS in wetland study began from in 1980s. Developed countries such as the United

Corresponding Author: Liu Chuanli, School of Architectural andsurveying Engineering, Jiangxi University of Science and Technology, Jiangxi, Ganzhou, 341000, China

States, UK and Netherlands has made extensive and profound studies on wetland and a lot of work has been done in wetland dynamic change monitoring, functional analysis, landscape analysis and wetland protection by RS (Liu, 2007). For example, Sader *et al*. (1995) and Augusteijn and Warrender (1998), Munyati (2000) and Hirano *et al*. (2003) investigated wetland resources and monitored wetland resources by SAR, LandsatTM, Mss, RadarsatsAR, LIDAR RS image.

Domestic application of RS in wetland study began in 1990s. Remote Sensing Classification Method applied by scholars in wetland identification promotes from visual interpretation to supervised classification, unsupervised classification and mixed classification by computers, to data mining classification combined with earth sciences. Scholars such as Wang *et al*. (2012) and Zeng *et al*. (2013) has in turn made classification on various wetland resources by TM, Landsat, Radarsat, SPOT RS image and put forward some effective classification methods.

METHODS OF SPATIAL SPATIAL DATA MINING

Spatial data mining (SDM), together with Spatial Knowledge Discovery, was initiated by Academician Li Deren in GIS international conference in Ottawa, Canada in 1994, which brought a spatial concept to data mining and knowledge discovery. SDM, or Spatial data mining, is the process of extracting space and non-space general knowledge rules that is connotative, unknown, potentially useful and ultimately understandable from spatial data (Li *et al*., 2006).

After decades of development, abundant theories and methods of spatial data mining and spatial knowledge have emerged, in which Decision Trees, Artificial Neural Networks, Support Vector Machine, Rough sets, Bayesian Networks and Association Rules are all often applied in RS field. Decision tree is a machine learning algorithm and a non-parametric classifier involving a recursive partitioning of the feature space, based on a set of rules learned by an analysis of the training set (Kumar *et al*., 2010). Decision Trees is to deduce classification rules in a tree form from a group of irregular and disorder data sets. The tree form classification is a fundamental form in natural species classification and is particularly suitable for RS image classification for its superior classification precision. Algorithms of Decision tree that are often used includes CLS presented by Hunt *et al*. (1966), ID3 learning Algorithms by Quinlan (1979) and ID3 See 5.0/C 5.0 Algorithms which is founded on the basis of ID3 learning Algorithms.

DATA SOURCES AND PREPROCESSING

RS image: The RS image data adopted here are Landsat-5 TM images of the studies area on 15th July, 1990 and on 29th September, 2005. Both orbit number are 121/40 and resolution 30 m. As the original images always involve some significant distortions, such as scanning distortion, instability of RS platform height, speed and RS platform posture and changes caused by earth curvature and air refraction, geometrical correction of original RS images is necessary so as to relieve image distortion caused by systematic and unsystematic factors First, selecting on the image a series of ground control points and inspection points that are evenly distributed with 1:50000 relief map as the reference; Ground control points should be set on significant surface features such as crossroads. Secondly, calculating the index of polynomial correction models by least square method regression after ground control points are fixed; and then conducting coordinate conversion to all pixels in the whole image by bilinear interpolation, re-sampling and correcting into the integrated Gauss Projection system. Image Residual in the process of correction should be controlled within half pixel.

Topography and landform map: Maps in 1:50 000 scale that covered the studied area are collected, scanned by computer and corrected by its control points. Thus contour lines and elevation points in the studied area are vectorized. Digital Elevation Model and SLOPE are generated by crag interpolation in ArcGIS software. As Fig. 1 is shown is overlaid RS image.

Field survey: By field survey, 1500 test samples were selected in the studied area, with their vegetation types recorded by GPS for later accuracy tests. And according

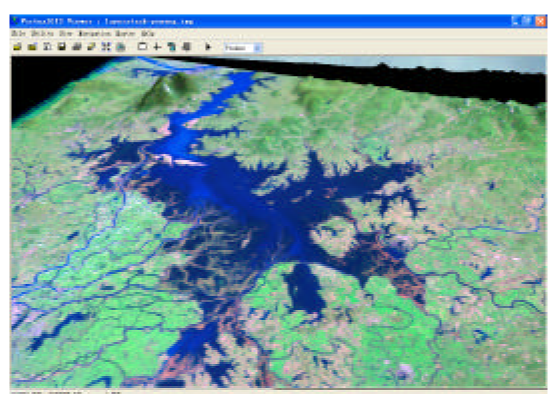


Fig. 1: RS images the overlaid DEM in poyang lake area

to existing historic monitoring record, spatial distribution characteristics of wetland vegetation in the studied area are basically grasped and spectrum interpretation keys of different land use type in the studied area are established. Referring to relevant classification method of land use or cover and combined with knowledge from field survey, 6 types of land cover were defined as paddy fields, forest land, swamp wetland, waters, settlement places and beaches respectively, in view of the spatial resolution of Landsat-5 TM image.

WETLANDINFORMATIONIDENTIFICATIONAND EXTRACTION

Image feature analysis: This study first takes samples in the studied area by AOI in Erdas Imagine 2010 and its seed pixel growth and then censuses Spectrum Characteristics information and makes out the spectrum Statistical Graphics. What's shown in Fig. 2 is spectrum statistical data of water samples.

The most important factors affecting the spectral reflectance among wetland vegetation are the biochemical and biophysical parameters of the plants' leaves and canopy such as chlorophyll a and b, carotene and xanthophylls (Adam *et al.,* 2010). Vegetation has the capacity of selectively absorbing, transmitting and reflecting electromagnetic radiation of all length from the sun. When the sun shines on the vegetation, most of the near-infrared (TM 4) is reflected back to form a reflecting peak and most visible light, particularly TM3, is absorbed to form an absorbing peak (Lei *et al.*, 2009). Reflection value of swamp wetland is relatively small on TM3 band and peaks at TM4 band. The same feature cannot be observed on non-vegetation. Therefore Normalized Difference Vegetation Index (NDVI) can be applied to enhance vegetation information and reduce non-vegetation in formation, so as to reduce the influence from soil and non-vegetation spectrum. Normalized Difference Vegetation Index (NDVI) can be defined as the ration of difference between the near-infrared (NIR) and visible light (RED) to the sum of the two, that is (Li and Yu, 2007):

$$\text{NDVI} = \frac{\text{IR-R}}{\text{IR+R}} \tag{1}$$

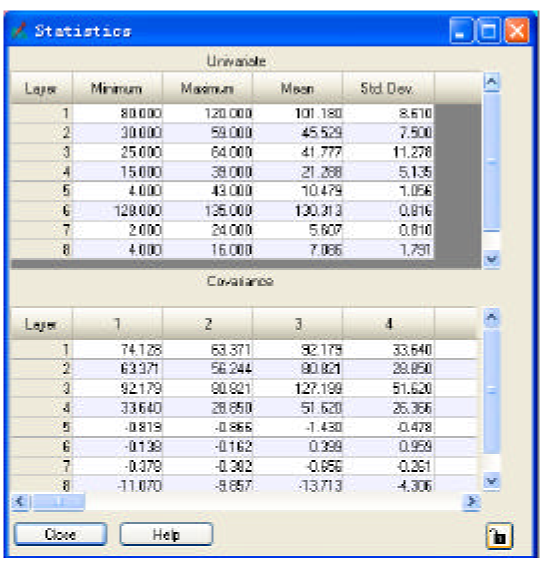
For TM image, the calculating formula of Normalized Difference Vegetation Index (NDVI) is:

$$\text{NDVI} = \frac{\text{TM4-TM3}}{\text{TM4+TM3}} \tag{2}$$

As to non-vegetation area, the main surface features are beach and settlement places consisting of building land, traffic land and mining area under human influences. According to spectrum mean value curves of the two types of surface features, great similarity can be found in the first three bands, while significant contrary features can be observed in band TM4 and TM5. That is, there is a downward tendency from TM4 to TM5 due to the absorption of the plenty water in beach; while as to the settlement places, it is just on the opposite side.

According to spectrum Statistical Graphics of different types of samples in the studied area, it can be found that TM3 is larger than TM4 for waters, while smaller for swamp wetlands, beaches, forest lands and paddy fields. So, by this way water can be preliminarily distinguished from swamp wetlands, beaches, forest lands and paddy fields.

Comparison of growing environment of swamp wetlands with forest lands shows that forest lands are mainly distributed on the hilly country in the southwest of the studied area with generally high terrain; while swamp wetlands are mainly distributed near the lake or on the bottomland of delta plain with generally low terrain. Therefore, a properly set elevation threshold value K3 could distinguish swamp wetlands from forest lands.



Statistics

Univariate

Layer	Minimum	Maximum	Mean	Std. Dev.
1	80.000	120.000	101.180	8.610
2	30.000	59.000	45.529	7.500
3	25.000	64.000	41.777	11.278
4	15.000	39.000	21.268	5.135
5	4.000	43.000	10.479	1.056
6	128.000	135.000	130.313	0.816
7	2.000	24.000	5.607	0.810
8	4.000	16.000	7.066	1.791

Covariance

Layer	1	2	3	4
1	74.128	63.371	92.179	33.640
2	63.371	56.244	90.821	28.850
3	92.179	90.921	127.199	51.620
4	33.640	28.650	51.620	26.366
5	-0.819	-0.866	-1.430	-0.478
6	-0.138	-0.162	0.399	0.959
7	-0.378	-0.392	-0.656	-0.261
8	-11.070	-9.657	-13.713	-4.306

Close Help

Fig. 2: Spectrum statistical data of water samples

Setting up decision tree model: According to RS image analysis of the studied area, decision tree model is set as Fig. 3 in this study.

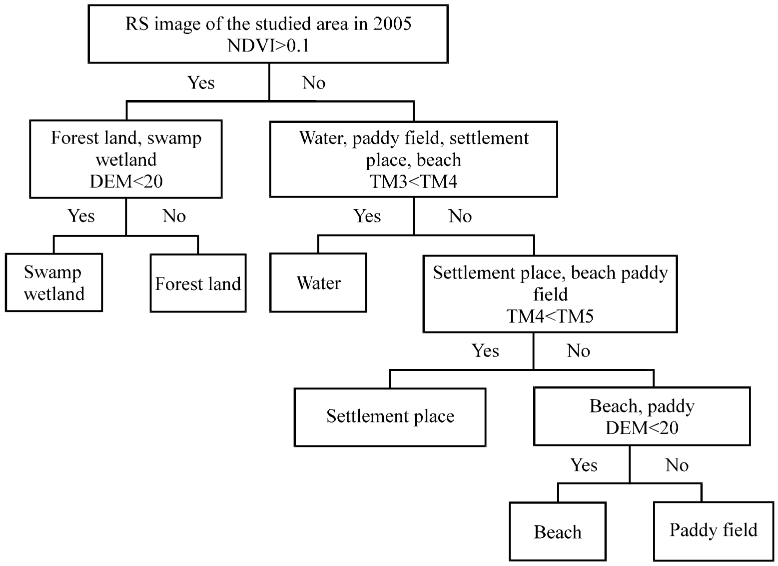


Fig. 3: Decision tree classification model for poyang lake area

First, set up inter-spectrum relation according to spectrum knowledge and interpret vegetation and non-vegetation. As the rice has been reaped when imaging, paddy fields are classified as non-vegetation features.

Second, distinguishing swamp wetlands from forest lands with the aid of Digital Elevation Model (DEM). As forest lands are mainly distributed on the hilly country in the southwest of the studied area with generally high terrain, while swamp wetlands are mainly distributed near the lake or on the bottomland of delta plain with generally low terrain, a properly set elevation threshold value could distinguish swamp wetlands from forest lands.

In all types of non-vegetation, TM3 of waters is larger than TM4, while it is just the opposite for paddy fields, settlement places and beaches. Thus land type as water can be extracted from others. Meanwhile, since TM4 of settlement places is larger than TM5, it can also be singled out. What's left are paddy fields and swamp wetlands. According to their corresponding distributing characteristics and law, non RS data are introduced to classify and extract. As swamp wetlands lies in the areas between high water level and lower one and periodically inundated, their elevation is usually lower than paddy fields. So, paddy fields can be distinguished with the help of DEM.

Image classification and precision analysis: According to above decision tree classification model, decision tree can be created by Knowledge Engineer module of Erdas Imagine 2010 software and knowledge database can be set up as Fig. 4. Then decision tree classification can be executed by Knowledge Classifier module and the result is generated (shown in Fig. 5). Finally, classification precision was assessed by test samples from field survey and is estimated by rough statistics to be above 85%.

DYNAMIC WETLAND CHANGE MONITOR

In order to analyze Poyang Lake Wetland change in many years and predict its tendency, RS dynamic monitoring function is applied. RS dynamic monitoring is the process of identifying surface feature or phenomenon changes on the basis that the images are different in spectrum features for of the same area in different years. Usually RS images data of different phrases (at least 2 phrases) are compared and pasting dynamic changes are analyzed quantitatively for space position and number change and future tendency is predicted (Li and Jia, 2008).

In this study two RS image of the same area in 2005 and in 1990 respectively are adopted to monitor 15 years' dynamic changes in Poyang Lake Wetland. Image preprocessing, including mainly Image Registration and radiation normalization, was necessary before change detection.

There are many algorithms for RS image change detection, in which Magnitude Difference, Tasseled Cap Difference, Primary Color Difference, Single-Band Difference and Band-Slope Difference are most often applied (Dang *et al*., 2010). Since this study aims at obtaining changes of many surface features, Magnitude Difference is adopted here with the threshold as 30%. Result of changes detection is shown in Fig. 6.

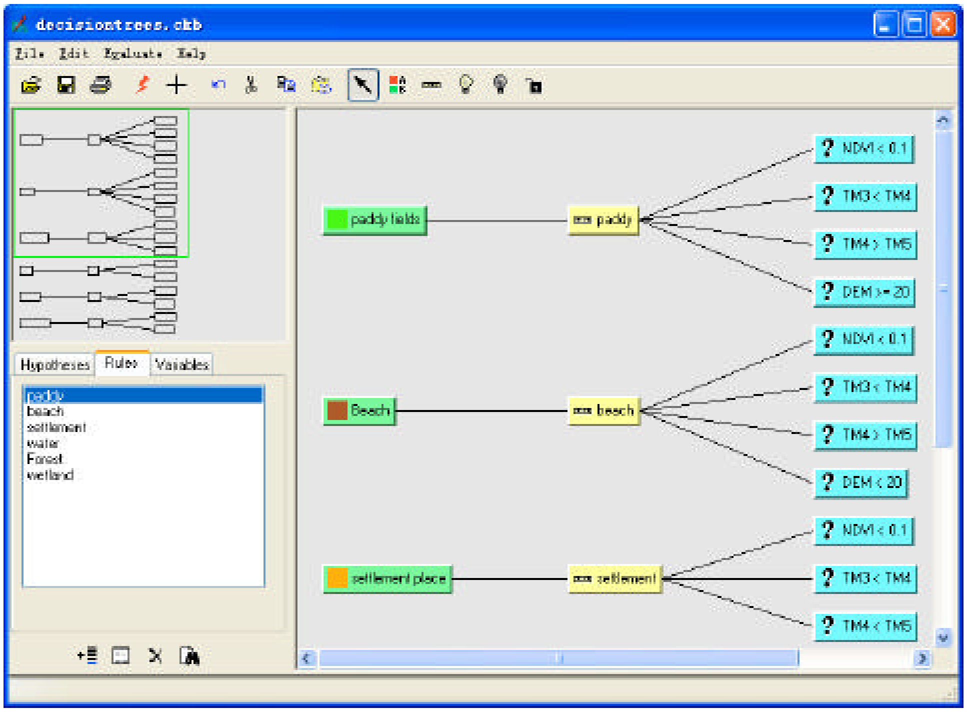


Fig. 4: Decision tree model set up in Erdas

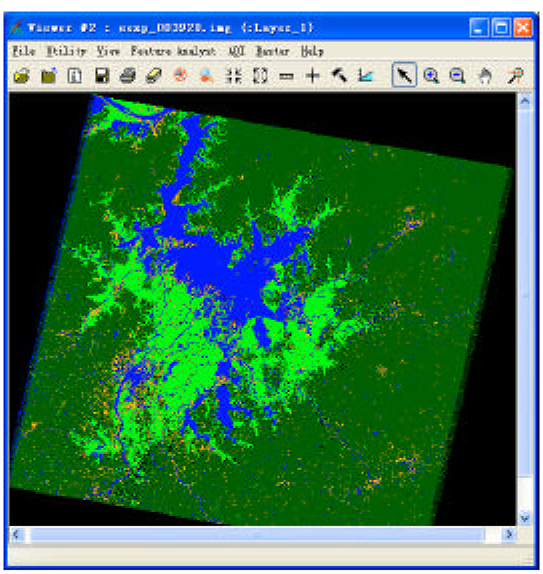


Fig. 5: Result of RS classification

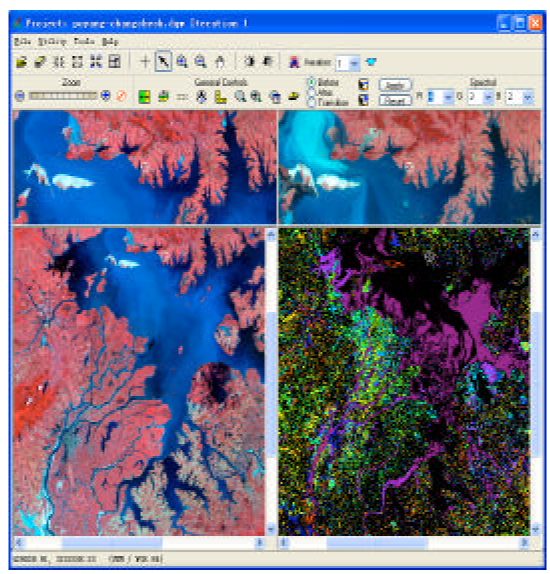


Fig. 6: Result of change detection of poyang lake wetland from 1990 to 2005

Result of change detection shows that forest lands, waters and settlement places all has increased to some different degree in 2005 compared with those of 1990, while beaches, swamp wetlands and paddy fields has decreased. Every change in land use type is motivated both by nature and by human. For example, the increase of water area is both due to inflating forest coverage, which benefit from the 20 years' "mountain, river, lake project" implemented in Jiangxi Province and due to decreased soil erosion and increased rain. The reason for shrinking swamp wetlands and beaches is both the water level rise and artificial development and utilization. While the main reasons for shrinking paddy fields are the policy of Grain for Green Project and Return the Field in Lakeside Areas to Lake and urban sprawl. As to forest lands and settlement places, the main factors that influence the changes are social ones such as policies and economic factors.

CONCLUSION

Conclusions can be drawn from the above study as follows:

- Compared with traditional supervised and unsupervised classification, RS image classification based on Decision Tree Model has a higher precision
- It is feasible to monitor Poyang Lake Wetland dynamically by multi-temporal RS image data

Constrained by RS image resolution and data quantity, not every land use type is considered in this study. And only two-temporal RS data are adopted in RS change monitoring. Integration of high resolution and high spectrum data can be applied in classification in order to classify land use type and wetland vegetation more precisely.

ACKNOWLEDGMENT

The study was subsidized by the National Natural Science Foundation of China (41061041).

REFERENCES

Adam, E., O. Mutanga and D. Rugege, 2010. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. Wetlands Ecol. Manage., 18: 281-296.

Augusteijn, C.E. and M.F. Warrender 1998. Wetland classification using optical and radar data and neural network classification. Int. J. Remote Sen., 19: 1545-1560.

Dang, A.R., H.F. Jia, X.F. Chen and J.B. Zhang, 2010. Eradas Imagine RS image process. Tsinghua University Press, Beijing, China.

Hirano, A., M. Madden and R. Welch, 2003. Hyperspectral image data for mapping wetland vegetation. Wetlands, 23: 436-448.

Hunt, E., J. Martin and P. Stone, 1966. Experiments in Induction. Academic Press, New York.

Kayastha, N., V. Thomas, J. Galbraith and A. Banskota, 2012. Monitoring wetland change using inter-annual landsat time-series Data. Wetlands, 32: 1149-1162.

Kumar, U., N. Kerle, M. Punia and T.V. Ramachandra, 2010. Mining land cover information using multilayer perceptron and decision tree from MODIS data. J. Indian Soc. Remote Sens., 38: 592-603.

Lei, T., G. Huang and Y. Lei, 2009. Remote sensing information recognition and sort abstraction of poyang lake wetland vegetation based on the DEM. Resour. Environ. Eng., 6: 844-847.

Li, D.R., S.L. Wang and D.Y. Li, 2006. Spatial Data Mining Theories and Applications. Science press, Beijing, China.

Li, F.F. and Y.H. Jia, 2008. Study on method of extracting wetland and its changed area based on Landsat TM images. Sci. Surv. Mapping, 33: 147-149.

Li, H. and M. Yu, 2007. Wetland information extraction based on decision tree model. Geo-Infor. Sci., 2: 60-65.

Liu, K., 2007. Study on knowledge discovery-based identification, detection and evalution law mining of pearl river estuary wetlands. Ph.D. Thesis, University of Chinese Academy of Sciences.

Munyati, C., 2000. Wetland change detection on the Kafue Flats, Zambia, by classification of a multitemporal remote sensing image dataset. Int. J. Remote Sens., 21: 1787-1806.

Prabhadevi, V. and C.S. Reddy, 2012. The use of remote sensing to quantify spatio-temporal land cover changes in point calimere a ramsar site. Natl. Acad. Sci. Lett., 35: 85-90.

Quinlan, J.R., 1979. Discovering Rules from Large Collections of Examples. In: Expert Systems in the Microelectronic Age, Michie, D. (Ed.). Edinburgh Universy Press, UK.

Sader, S.A., D. Ahl and W.S. Liou, 1995. Accuracy of landsat-TM and GIS rule-based methods for forest wetland classification in maine. Remote Sens. Environ., 53: 133-144.

Wang, C., H.Y. Liu, M.X. Hou and Q.M. Tan, 2012. Classification method of muddy tidal flat wetlands based on remote sensing. Geo-Inform. Sci., 15: 590-596.

Zeng, G., H.J. Gao and G. Zhu, 2013. Analysis of wetlands evolution process between Tarim Basin and Junggar Basin in the past 32 years. Remote Sens. Land Resour., 3: 118-123.