Application of Spatio-temporal Data Mining and Knowledge Discovery for Detection of Vegetation Degradation

Analysis of time-series remote sensing images using spatial statistics method

Xi-yong HOU[a], Lei Han[a,b], Meng GAO[a],
[a] Yantai Institute of Coastal Zone Research
Chinese Academy of Sciences
Yantai, 264003, China

Xiao-li BI[a], Ming-ming ZHU[a,b]
[b] Graduate School of the
Chinese Academy of Sciences
Beijing 100049, China

***Abstract*—Increasing time-series remote sensing images provide the information about the evolution processes of ecosystems on multi-spatial scales. Vegetation plays an important role in sustaining the natural environment and supporting human being with goods and ecosystem services. Detection of vegetation degradation has become a hot spot of multi-disciplinary researches recently. In this paper, a case study of spatio-temporal data mining and knowledge discovery for detection of vegetation degradation has been conducted. The special issues focused on the quantitative determination of historical evolutionary trend and furthermore, the sustainability of different trends in the future. Taking the Circum-Bohai-Sea region as the case study area, the Unary Linear Regression Model (ULRM) has been established based on the time-series SPOT-VGT images from 1998 to 2008, and then the Hurst index has been calculated by R/S method on the spatial scales of cell (1km^2) and the whole study area. It turned out that, the combined analysis between *Slope* of ULRM and Hurst index could effectively reveal the characteristics of vegetation changes, which included the degraded areas in the past as well as the risk level of degradation in the future. Overall, the areas of vegetation degradation in the future amount to 38.87 thousand square kilometers, which accounts for 7.55% of the whole study area. In addition, these degraded areas mainly distributed around the metropolitan regions, coastal zone, and so on. The findings will help us with more intelligent strategies of degradation prevention.**

Keywords-spatio-temporal data mining; knowledge discovery; vegetation degradation; unary linear regression model; Hurst index

I. INTRODUCTION

The remote sensing techniques has been developed rapidly over the past few decades, and today, more and more remote sensing images are available for multi-disciplinary studies. However, vast amounts of data have posed severe challenges to methodologies and techniques of spatial-temporal data mining and knowledge discovery. In recent years, scientists have developed different methods for data mining and knowledge discovery from time-series remote sensing images. For example, LI et al. (2002) had systemically discussed the theories and techniques of data mining and knowledge discovery of spatial data [1]. B. K. Sy et al. (2002) had proposed an information-statistical approach for the treatment of temporal-spatial data [2]. H.A. Barbosa et al. (2006) had carried out a study of NDVI (Normalized Difference Vegetation Index) variability for 20 years over the northeast region of Brazil [3]. LIU et al. (2006) had studied the forest disease spread by a spatial-temporal approach using multi-temporal high spatial resolution imagery [4]. R. Lasaponara (2006) had evaluated inter-annual vegetation anomalies by Principal Component Analysis method from temporal series NDVI data [5]. SONG et al. (2007) had studied the vegetation cover change in Northwest China based on time-series SPOT-VGT data [6]. Brian D. Wardlow et al. (2008) had developed the method for large-area crop mapping using time-series MODIS NDVI data [7]. GU et al. (2009) had presented a simplified data assimilation method for reconstructing time-series MODIS NDVI data [8]. HAN et al. (2009) had studied the spatial-temporal change of vegetation in the Yangtze River Delta based on time series remote sensing images [9]. Overall, it's still in the initial stages of this research field, and, in the future, more in-depth studies should be carried out in order to develop more effective methods and algorithms for spatio-temporal data mining and knowledge discovery.

In this paper, taking the Circum-Bohai-Sea region as the study area, the appropriate mathematical models and spatial statistics method for the monitoring of vegetation degradation have been studied based on the time-series SPOT-VGT data. The specific objectives included: 1) to reveal the change trends in the past; 2) to evaluate the sustainability of these trends in the future; 3) to detect regions showing trend of vegetation degradation in the future.

II. DATA AND METHODOLOGY

A. SPOT-VGT 1km NDVI composites

The SPOT-VGT NDVI data are available free of charge at the website of the image processing and archiving centre, VITO, Belgium (http://free.vgt.vito.be/). The imagery data are 10-day composites with 1-km spatial resolution, and they are derived from the vegetation sensor on board the SPOT satellites (SPOT-4 and SPOT-5). The Maximum Value

Supported by the Knowledge Innovation Program of the Chinese Academy of Sciences (No.kzcx2-yw-224) and the National Natural Science Foundation of China (No.40801016).

Composite method(MVC) has been used to expand the time-scale from one day to 10-day and simultaneously to reduce the contamination effects caused by sensor, residual clouds, atmospheric perturbations, observation angle, variable illumination and so on [5,10]. In this paper, based on the 10-day composites data from 1998 to 2008, an 11-year time-series SPOT-VGT NDVI data set (Figure1) are obtained using the MVC method.

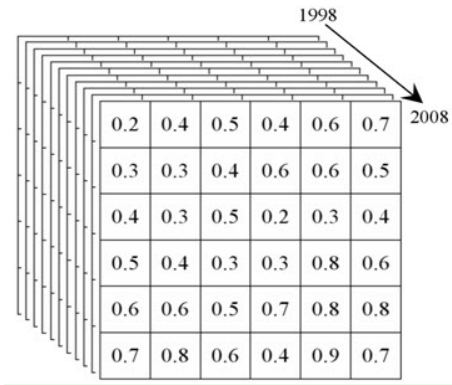

Figure 1. Schematic diagram of the time-series SPOT-VGT NDVI data

B. *Unary linear regression model*

The Unary Linear Regression Model (ULRM) is used to calculate the slope of time-series NDVI data based on the cell values in every year. Therefore, it can overcome the bias of range analysis method caused by occasional factors of both the starting and the ending year. The formula is as follows,

$$Slope = \frac{n \times \sum_{i=1}^{n} i \times NDVI_i - \sum_{i=1}^{n} i \sum_{i=1}^{n} NDVI_i}{n \times \sum_{i=1}^{n} i^2 - \left(\sum_{i=1}^{n} i \right)^2} \quad (1)$$

where n is the temporal range, $NDVI_i$ is the NDVI value in the year i, and the *Slope* obtained by ULRM can quantitatively depict the historical trend of NDVI changes. When *Slope*> 0, it reveals that the vegetation has been increased, and otherwise, decreased.

C. *Hurst index*

Vegetation cover change is a kind of natural process similar to hydrology, climate, geo-chemistry and geology, which show strong self-similarity and long-range dependence. Hurst index (H) is one of the most effective methods to depict the self-similarity and long-range dependence for the time-series data. And the R/S method is one of the most common methods to calculate the Hurst index; the basic principles are as follows,

for a time-series, $\{\zeta(t)\}$, $t = 1, 2, 3, \ldots, N$;

define another sequence, $\tau = 1, 2, 3, \ldots, N$,

for a certain τ, define the mean sequence,

$$< \xi >_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t) \quad (2)$$

and for the cumulative deviation of time t,

$$X(t,\tau) = \sum_{u=1}^{t} \left(\xi(u) - < \xi >_\tau \right), \quad 1 \le t \le \tau \quad (3)$$

next, to define the range sequence,

$$R(\tau) = \max_{1 \le T \le \tau} X(t,\tau) - \min_{1 \le T \le \tau} X(t,\tau) \quad (4)$$

then, define the standard deviation sequence,

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} (\xi(t) - < \xi >_\tau)^2} \quad (5)$$

There is an exponential law as follows,

$$\frac{R(\tau)}{S(\tau)} = (c\tau)^H \quad (6)$$

$$\ln \frac{R(\tau)}{S(\tau)} = H \ln c + H \ln \tau \quad (7)$$

Based on the equation (7), H can be calculated through observed data. If $H = 0.5$, the time-series will be a independent random process; if $H > 0.5$, the time-series is a persistent series, that is, if there is a positive increment in the average sense in the past, there will be a increase in the future; if $H < 0.5$, the time-series is a anti-persistent series, that is, the decreasing trend in the past will change to increase in the future.

D. *Spatial statistics method*

It's the powerful spatial analysis techniques that make Geographic Information System (GIS) very different from and superior to many Computer-aided design (CAD) systems and Graphics Processing Systems. The ArcGIS software, which is the most popular GIS platform in the world, provides users abundant spatial analysis and Geo-computation algorithms. For example, there are lots of commands and functions in its Cell-based Modeling tools by which users can carry out local, focal and zonal analysis.

In this paper, the Unary Linear Regression Model and the Hurst index were calculated on the 1 km cell scale mainly based on the commands and functions of local operations in ArcGIS software platform. Figure 2 shows the flow chart of data processing and the outputs by spatial statistics techniques in this paper.

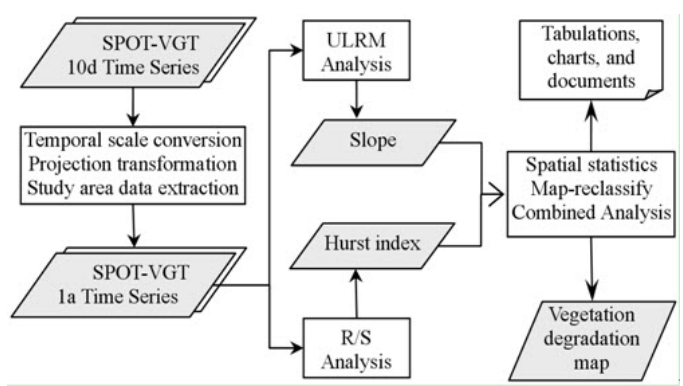

Figure 2. Flow chart of data processing by spatial statistics techniques

III. RESULTS

Figure 3 and Table I were the analysis results of ULRM.

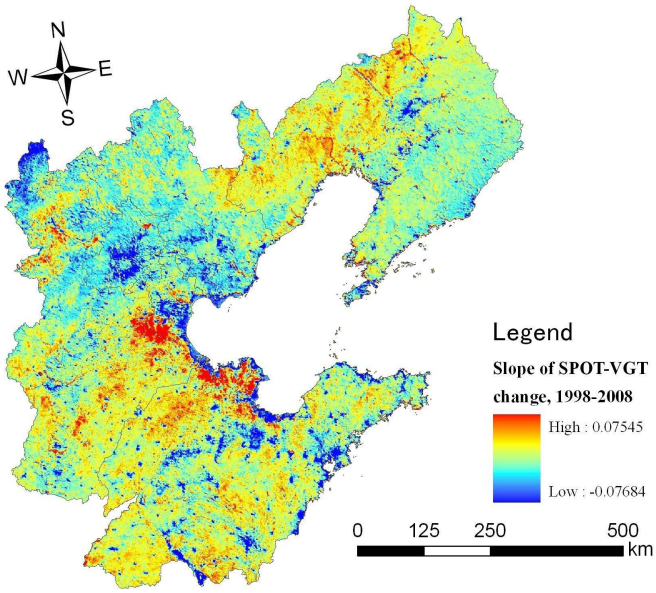

Figure 3. The slopes of SPOT-VGT NDVI in 1998-2008 on 1 km cell scale

TABLE I. SLOPES OF SPOT-VGT NDVI IN 1998-2008

Class	Slope	NDVI change	Area/km^2	Area ratio /%
1	< -0.01	Severely degraded	10313	2.00
2	-0.01~-0.005	Moderately degraded	9619	1.87
3	-0.005~-0.001	Slightly degraded	16447	3.19
4	-0.001~0.001	Stable	15458	3.00
5	0.001~0.005	Slight improvement	66122	12.84
6	0.005~0.01	Moderate improvement	181685	35.28
7	> 0.01	Significantly improved	215375	41.82

Negative values of *Slope* meant that the vegetation cover had degenerated in the past, while positive values of *Slope* meant that the vegetation cover had improved (Table I). It turned out that, from 1998 to 2008, vegetation cover has improved markedly in most regions. The severely degraded area with the *Slope* less than -0.01 amounted to 10.31 thousand square kilometers and accounted for 2% of the whole study area, which was mainly distributed at the north-west of Hebei province as well as areas around the mega-cities such as Beijing, Tianjin, Jinan, and Qingdao, and so on. The moderately degraded area with the *Slope* ranging from -0.01 to -0.005 amounted to 9.6 thousand square kilometers and accounted for 1.87% of the whole study area, which was mainly distributed at the middle part of Hebei province, the east part of Liaoning province and areas around the mega-cities(Figure 3). The slightly degraded area with the *Slope* ranged from -0.005 to -0.001 amounted to 16.45 thousand square kilometers and accounted for 3.19% of the whole study area, which was mainly distributed at the north part of Hebei province, the Liao-dong peninsula and the Jiao-dong peninsula.

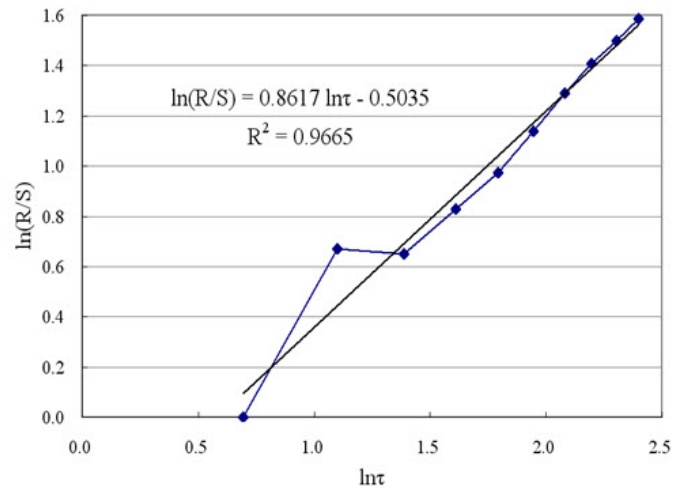

Figure 4. The R/S analysis for the SPOT-VGT NDVI changes in the Circum-Bohai-Sea region

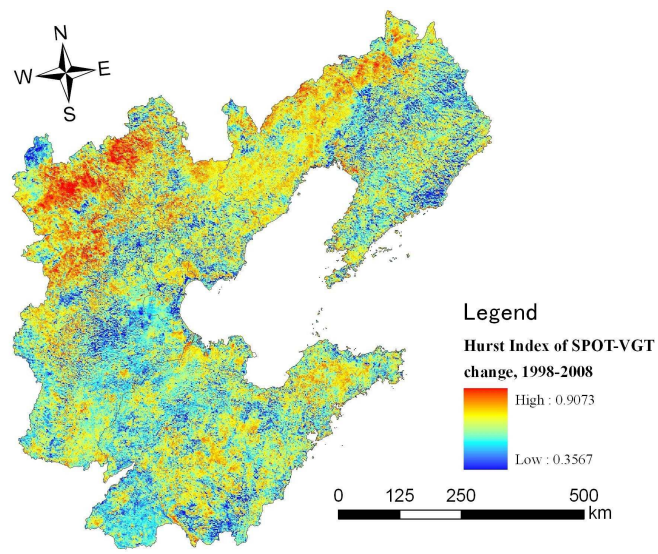

Figure 5. Hurst index: Sustainability of SPOT-VGT NDVI changes

The Hurst index both on 1km cell scale and on the whole study area scale turned out that, for the whole study area, it arrived at 0.8617and the significance level amounted to 0.9665 (Figure 4), which showed that the increasing trends of vegetation cover between 1998 and 2008 would be significantly sustainable in the future. Overall, spatially, the north-west part of Hebei province, the west part of Liaoning province, the mid-part of Shandong province and the mid-part of Jiao-dong peninsula were areas that with very high values of Hurst index. On the contrary, the Liao-dong peninsula, the north-west corner and the south part of Hebei province, the south part of Shandong province were areas that with low values of Hurst index (Figure 5).

TABLE II. SUSTAINABILITY OF SPOT-VGT NDVI CHANGES

Class	Hurst	Sustainability	Area/km^2	Area ratio /%
1	< 0.45	Strong anti- sustainability	474	0.09
2	0.45~0.5	Weak anti- sustainability	3108	0.60
3	0.5~0.65	Weak sustainability	88093	17.10
4	0.65~0.8	Moderate sustainability	364677	70.81
5	> 0.8	Strong sustainability	58667	11.39

Table II was the statistical results of Figure 5. It turned out that most regions of the study area had the Hurst index greater than 0.5, which meant that the trend of the vegetation cover change between 1998 and 2008 would be sustainable in the future. The area of strong anti-sustainability amounted to 474 square kilometers and accounted for 0.09% of the whole study area. The area of weak anti-sustainability amounted to 3.11 thousand square kilometers and accounted for 0.60% of the whole study area. The area of weak sustainability amounted to 88.09 thousand square kilometers and accounted for 17.10% of the whole study area. The rest of the study area were regions of moderate to strong sustainability, with the area totally amounted to 423.34 thousand square kilometers and accounted for 82.20% of the whole study area.

It's not enough to study the *Slope* and Hurst index separately because *Slope* alone can reveal the change characters in the past only, and Hurst index alone can reveal the long-range dependence only. Therefore, combined analysis between *Slope* and Hurst index had been carried out (Figure 6, Figure 7 and Table III).

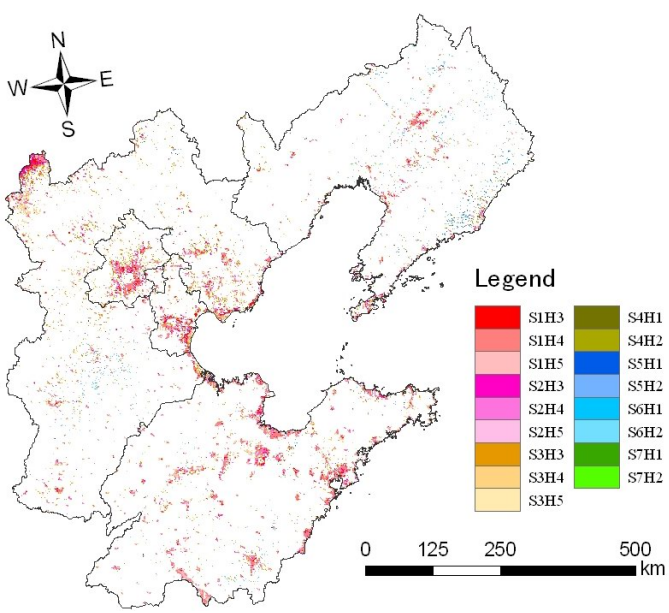

Figure 6. Degradation trends of vegetation cover

In figure 6, naming rules for the legend are as follows, 'S' means the '*Slope*', and 'H' means the Hurst index, the numeral followed with 'S' or 'H' means the hierarchical classes of them, and the specific classification criteria of *Slope* and Hurst index are absolutely the same as that in Table I and Table II respectively. For example, 'S1H4' means that the class of '*Slope*' is 1 and the class of Hurst index is 4, and all the cells marked with 'S1H4' are the regions that vegetation cover had 'severely degraded' in the past and the degrading trend in the future is 'moderate sustainability'.

Two different kinds of vegetation degradation trends could be found (Figure 6), one was that the vegetation cover has degraded from 1998 to 2008 and the degrading trend would be sustainable in the future; and the other was that the vegetation cover has improved from 1998 to 2008, while the improving trend would be unsustainable in the future. Clearly, as it turned out by Figure 6, the whole study area was dominated by the former kind of vegetation degradation. Furthermore, spatially, the north-west corner of Hebei province, the Beijing-Tianjin-Tangshan metropolitan region, the west and south coastal area of Bohai Sea, and large number of urban areas both in Shandong province and in Liaoning province are the areas that vegetation cover would continue to degrade in the future.

TABLE III. AREA SHOWING TREND OF DEGRADATION

City or province	Area showing trend of degradation		
	Area /km^2	Ratio-A[a] /%	Ratio-B[b] /%
Beijing	3364	20.53	8.66
Tianjin	2187	18.82	5.63
Hebei	12781	6.82	32.89
Liaoning	6354	4.37	16.35
Shandong	14179	9.19	36.48

a. The proportion of degradation area to the total area of this city or province.

b. The proportion of degradation area to the total degradation area in the Circum-Bohai-Sea region.

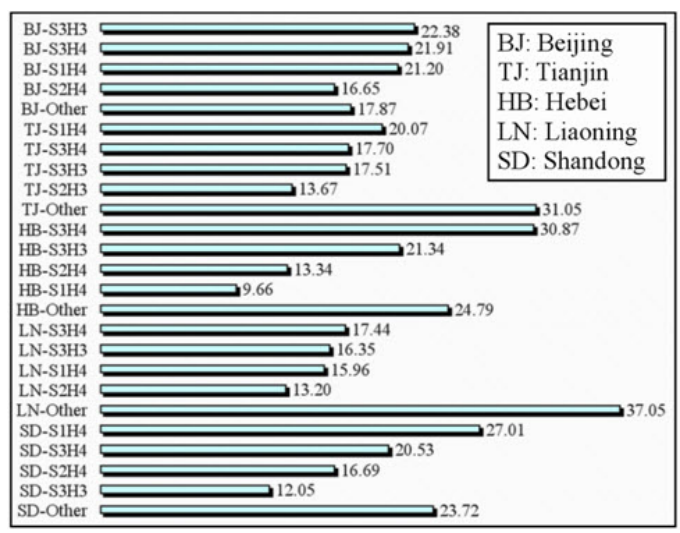

Figure 7. The structural characteristics of vegetation degradation in every province or city

Table III showed the result of zonal statistics. It turned out that, overall, the total area with the trend of vegetation degradation amounted to 38.87 thousand square kilometers and accounted for 7.55% of the whole study area. And it mainly distributed in Shandong province and Hebei province, the next

was Liaoning province, the fourth is Beijing City, and the last is Tianjin City. The former two provinces totally accounted for 69.37% of the vegetation degradation area in the whole study area. However, in terms of area ratio-A, i.e. the proportion of vegetation degradation area to the total area of this city or province, the highest is Beijing City, with 20.53%, the second is Tianjin City, with 18.82%, the third is Shandong province, with 9.19%, the fourth is Hebei province, with 6.82%, and the last is Liaoning province, with 4.37% only.

Figure 7 showed the structural characteristics of vegetation degradation in every province or city in the future. The former four kind of combination of '*Slope*' and Hurst index as well as its area ratio were shown as figures for every province or city. It turned out that the most wide spread four kind of vegetation degradation in every province or city belong to the former type of vegetation degradation trends, i.e. degraded in the past and will go on degrading in the future. And furthermore, the accumulative area ratio of these four former kind of vegetation degradation amount to 82.13%, 68.95%, 75.21%, 62.95% and 76.28% in Beijing, Tianjin, Hebei, Liaoning and Shandong respectively.

IV. CONCLUSIONS AND DISCUSSIONS

We have presented a case study of the application of spatio-temporal data mining and knowledge discovery techniques based on mathematical method and spatial analysis in GIS platform.

The results of time series SPOT-VGT analyses by Unary Linear Regression Model demonstrated that the vegetation degradation area was 36.38 thousand square kilometers and accounted for 7.06% of the whole study area. And the degradation area mainly distributed at the north-west of Hebei province, the metropolitan areas and the coastal areas etc.

The results of time series SPOT-VGT analyses by R/S method on different spatial scales revealed that the Hurst index was 0.8617 for the whole study area, showing that the trend of vegetation improvement between 1998 and 2008 will be sustainable in the future. However, results of Hurst index calculated on the local fine-scale of 1 square kilometer pixel distinctly reflected the spatial variations. In detail, areas with the value of Hurst index less than 0.5 amounted to 3.58 thousand square kilometers and accounted for 0.69% of the whole study area. The trend of vegetation cover change in these areas in the past would be unsustainable in the future.

The combined analysis of *Slope* and Hurst index effectively detected areas showing trend of degradation in the Circum-Bohai-Sea region, which totally amounted to 38.87 thousand square kilometers and accounted for 7.55% of the whole study area. And it mainly distributed around the metropolitan regions, a part of coastal zone, and so on. Detailed statistics show that the accumulative area ratio of the four former kind of vegetation degradation in Beijing, Tianjin, Hebei, Liaoning and Shandong amount to 82.13%, 68.95%, 75.21%, 62.95% and 76.28% respectively.

Obviously, the analyses of SPOT-VGT NDVI changes from 1998 to 2008 in the Circum-Bohai-Sea region highlight the spatio-temporal characters of vegetation cover change. In detail, the Slope of Unary Linear Regression Model has revealed both the directions and extents of vegetation changes in the past 11 years. And the Hurst index by the R/S method both on the regional macro-scale and the local fine-scale of 1 square kilometer pixel has reflected the sustainability of past changes. Furthermore, the combined analysis of *Slope* and Hurst index has clearly revealed the degradation trends of vegetation cover in the study area. The SPOT-VGT time series is very useful for vegetation degradation detection based on the techniques of spatio-temporal data mining and knowledge discovery.

ACKNOWLEDGMENT

Special thanks are due to Dr. De-juan JIANG for her insightful and constructive comments that helped to improve the manuscript, and to Dr. Liang-ju YU for his valuable advisements and assistance during the process of the research.

REFERENCES

[1] De-ren LI, Shu-liang WANG, De-yi LI, and Xin-zhou WANG, "Theories and technologies of spatial data mining and knowledge discovery," Geomatics and information Science of Wuhan University, vol. 27, pp. 221-233, June 2002.

[2] B.K. Sy, Arjun K. Gupta, "Information-statistical approach for temporal-spatial data with application," Engineering Applications of Artificial Intelligence, vol. 15, pp. 177-191, 2002.

[3] H.A. Barbosa, A.R. Huete, W.E. Baethgen. "A 20-year study of NDVI variability over the Northeast Region of Brazil," Journal of Arid Environments, vol. 67, pp. 288-307, 2006.

[4] Desheng Liu, Maggi Kelly, Peng Gong, "A spatial−temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery," Remote Sensing of Environment, vol.101, pp. 167−180, 2006.

[5] R. Lasaponara, "On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series," Ecological modelling, vol. 194, pp. 429-434, 2006.

[6] Yi SONG, Ming-guo MA, "Study on vegetation cover change in Northwest China based on SPOT VEGETATION data," Journal of Desert Research, vol. 27, pp. 89-93, January 2007.

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

[8] Juan Gua, Xin Li, Chun-lin Huang, and Gregory S. Okin, "A simplified data assimilation method for reconstructing time-series MODIS NDVI data," Advances in Space Research, vol. 44, pp. 501-509, 2009.

[9] Gui-feng HAN, Ke ZHAO, Jian-hua XU, "Spatial-temporal change of vegetation in the Yangtze River Delta based on time series remote sensing," Chinese Landscape Architecture, vol. 25, pp. 60-64, 2009.

[10] Maisongrande P, Duchemin B, and Dedieu G, "VEGETATION /SPOT: an operational mission for the Earth monitoring; presentation of new standard products," International Journal of Remote Sensing, vol. 25, pp. 9 -14, 2004.