

Spatiotemporal Mining of Time-Series Remote Sensing Images Based on Sequential Pattern Mining

H. C. Liu^{a,b}, G. J. He^{a,*}, X. M. Zhang^{a,b}, W. Jiang^{a,b}, S. G. Ling^{a,b}

^a Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China –
(hcliu, gjhe, xmzhang, weijiang, sgling)@ceode.ac.cn

^b University of Chinese Academy of Sciences, Beijing 100049, China

KEY WORDS: Remote Sensing, Satellite Image, Time Series, Data Mining, Sequential Pattern Mining, Land Cover Change

ABSTRACT:

With the continuous development of satellite techniques, it is now possible to acquire a regular series of images concerning a given geographical zone with both high accuracy and low cost. Research on how best to effectively process huge volumes of observational data obtained on different dates for a specific geographical zone, and to exploit the valuable information regarding land cover contained in these images has received increasing interest from the remote sensing community. In contrast to traditional land cover change measures using pair-wise comparisons that emphasize the compositional or configurational changes between dates, this research focuses on the analysis of the temporal sequence of land cover dynamics, which refers to the succession of land cover types for a given area over more than two observational periods. Using a time series of classified Landsat images, ranging from 2006 to 2011, a sequential pattern mining method was extended to this spatiotemporal context to extract sets of connected pixels sharing similar temporal evolutions. The resultant sequential patterns could be selected (or not) based on the range of support values. These selected patterns were used to explore the spatial compositions and temporal evolutions of land cover change within the study region. Experimental results showed that continuous patterns that represent consistent land cover over time appeared as quite homogeneous zones, which agreed with our domain knowledge. Discontinuous patterns that represent land cover change trajectories were dominated by the transition from vegetation to bare land, especially during 2009–2010. This approach quantified land cover changes in terms of the percentage area affected and mapped the spatial distribution of these changes. Sequential pattern mining has been used for string mining or itemset mining in transactions analysis. The expected novel significance of this study is the generalization of the application of the sequential pattern mining method for capturing the spatial variability of landscape patterns, and their trajectories of change, to reveal information regarding process regularities with satellite imagery.

1. INTRODUCTION

Land use and land cover change are becoming increasingly recognized as important drivers of global environmental change (Turner et al., 2007). The characteristics of land cover can have important effects on the local climate, radiation balance, biogeochemistry, hydrology, and the diversity and abundance of terrestrial species (Randerson et al., 2006). Therefore, the study of land cover change is an important problem within the geoscience domain. Advances in earth observation technologies have led to the acquisition of vast volumes of accurate, timely, and reliable environmental data, which encompass wide-ranging information about the earth's land, ocean, and atmosphere (Karpatne et al., 2013). Remote sensing imagery consisting of satellite-based observations of the land surface, biosphere, solid earth, atmosphere, and oceans, combined with historical climate records and predictions from ecosystem models, offer new opportunities for understanding how the earth is changing, determining the factors causing these changes, and predicting future changes (Boriah et al., 2009). Therefore, remote sensing satellite imagery has emerged as the most useful data source for characterizing and quantitatively measuring landscape-scale land cover changes (Hudak and Wessman, 1998).

Detecting and characterizing change over time is the natural first step towards identifying the drivers of the change and

understanding the change mechanism (Verbesselt et al., 2010). Satellite remote sensing has long been used as a means of detecting and classifying temporal changes in the condition of the land surface (Coppin et al., 2004; Lu et al., 2004). Satellite sensors are well suited to this task because they provide consistent and repeatable measurements on a spatial scale appropriate for capturing the processes of change (Jin and Sader, 2006). A given scene may be observed repeatedly from space, resulting in a times series of satellite images. The high spatial resolution of current sensors provides detailed information on spatial structures, which after a series of revisits, can be extended to spatiotemporal data structures. It follows that a time series of satellite images represents a highly complex data set that potentially contains valuable spatiotemporal information (Gueguen and Datcu, 2007).

Although the value of remotely sensed long-term data sets to change detection has been firmly established (De Beurs and Henebry, 2005), only a limited number of time-series change-detection methods have been developed. Most previous change-detection studies have relied primarily on examining the differences between two or more satellite images acquired on different dates (Boriah, 2010). These procedures can be categorized into three types. Procedures of the first category identify simple proportional differences of certain classes within a certain area between two points in time without being spatially explicit (Godoy and Contreras, 2001; Sierra, 2000).

* Corresponding author

The second category procedures involve the calculation of annual rates of change between end-member periods (Lambin and Ehrlich, 1997; Puyravaud, 2003). Methods in the third category quantify changes in the spatial configuration and composition of land cover on multiple dates using pair-wise comparisons (Mertens and Lambin, 2000; Schneider, 2012). One limitation of these methods is the inadequacy of detailed observation and discrimination of the process of land cover change. Actually, temporally, the change process of land cover types can be viewed as a trajectory, which highlights the dynamic character of change (Petit et al., 2001). By analyzing change trajectories, we can investigate the detailed dynamics of change processes (i.e., sequences of successive changes in land cover types). However, such methods used for analyzing change trajectories can usually be found for use with green cover; for example, vegetation (Cai and Wang, 2010), cropland (Liu et al., 2005), forest land (Chen et al., 2012; Gimmi et al., 2010; Kennedy et al., 2007; Lambert et al., 2011), and grassland (Dubinin et al., 2010; Dusseux et al., 2011) research. These methods are rarely applied in the change analysis of other land cover types. Therefore, there is an urgent need for a new universal technique for transforming remote sensing data into information on the patterns and processes of land cover change. The effective use of a satellite image time series to characterize and monitor land cover change trajectories requires the analysis of temporal variations in spatial patterns (Henebry and Goodin, 2002). When analyzing a sequence of items, one basic problem to be addressed is to find frequent episodes (Mannila and Toivonen, 1996), or in other words, to extract regular patterns from temporal data. Therefore, there is a critical need for a method that can enable efficient and reliable characterization of spatiotemporal patterns contained in an image time series (Henebry and Goodin, 2002). Mining sequential and spatial patterns is an active area of research in artificial intelligence (Le Ber et al., 2006), and it has been used for string mining or itemset mining in transactions analysis. The objective of this study is the generalization of the application of the sequential pattern mining method for capturing the spatial variability of landscape and their temporal trajectories of change, to reveal information regarding process regularities with satellite imagery. In this study, a sequence of repeated satellite images of the same scene was used to establish the land cover change sequence set, and a typical sequence mining algorithm—the Continuous Association Rule Mining Algorithm (CARMA)—used on this sequence set to analyze the land cover change in the form of change patterns.

This article is organized as follows. Section 1 discussed the significance of land cover change and introduced the objectives of this study. Section 2 covers the sequential pattern mining method and Section 3 presents a case study and an analysis of the results. Finally, Section 4 discusses the principal findings and offers our conclusions.

2. METHODOLOGY

The methodology used in this study includes remote sensing image processing, spatial analysis, and image data mining. In addition to image pre-processing and remote sensing image classification, a sequential pattern mining method is applied in the analysis of land cover change, which forms the main part of the research. Therefore, in this section, the sequential pattern mining method based on the CARMA algorithm is introduced. First, some basic concepts of pattern mining are provided in

Section 2.1, followed by an introduction to the CARMA algorithm in Section 2.2.

2.1 Basic concepts

Let $I = \{I_1, I_2, \dots, I_p\}$ be the set of all items; a set of items is referred to as an itemset and a sequence is an ordered set of one or more itemsets. For example, in a sequence $s = \langle e_1, e_2, \dots, e_j \rangle$, itemset e_1 shows before e_2 , and e_2 shows before e_3 , and so on. Itemset e_j is also an element of the sequence s denoted as (x_1, x_2, \dots, x_q) in which $x_q \in I$. A sequence that contains k itemsets is a k -sequence. If there exists $1 \leq i_1 < i_2 < \dots < i_n \leq m$ such that $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$ appears in sequence $A = \langle a_1, a_2, \dots, a_n \rangle$ and sequence $B = \langle b_1, b_2, \dots, b_m \rangle$, then it is said that A is contained in B (denoted as $A \subseteq B$). A set of sequences is referred to as a sequence set. In a sequence set, if a sequence s is not contained in any other sequence, then s can be called the largest sequence. Additionally, in the sequence set, the number of sequences that contain s is known as the support of s (written as $\text{sup}(s)$). In the mining process, if a sequence satisfies the pre-determined minimum support, then it is a frequent sequence. Sequential pattern mining is the mining of the largest sequences from the frequent sequences in a sequence set, and the sequences found by sequential pattern mining can be called the sequential pattern. In a sequential pattern $s = \langle e_1, e_2, \dots, e_j \rangle$, sequence $s' = \langle e_1, e_2, \dots, e_{j-1} \rangle$ is called the antecedent and sequence $s'' = \langle e_j \rangle$ is called the consequent. The confidence for a sequential pattern can be defined as the ratio of $\text{sup}(s)$ to $\text{sup}(s')$. The confidence for each generated sequential pattern should be calculated and those sequential patterns for which confidence is larger than the minimum confidence threshold are recognized as more interesting (or more useful for us) than the other sequential patterns.

2.2 CARMA algorithm

2.2.1 First part of CARMA

For the first step, the main target is to form a set V of all potentially large itemsets in a lattice. There are three parameters for each itemset in the lattice: $\text{Count}(v)$, $\text{firstTrans}(v)$, and $\text{maxMissed}(v)$. The meanings of these three parameters are introduced below and shown in Fig. 1:

$\text{Count}(v)$: the number of occurrences of itemset v since v was inserted in the lattice.

$\text{firstTrans}(v)$: the index of the sequence data at which v was inserted in the lattice.

$\text{maxMissed}(v)$: upper bound on the number of occurrences of v before v was inserted in the lattice.

The construction of the lattice is shown in Table 1:

Itemset	Count	firstTrans	maxMissed
i_1	$\text{Count}(i_1)$	$\text{firstTrans}(i_1)$	$\text{maxMissed}(i_1)$
i_2	$\text{Count}(i_2)$	$\text{firstTrans}(i_2)$	$\text{maxMissed}(i_2)$
...
i_k	$\text{Count}(i_k)$	$\text{firstTrans}(i_k)$	$\text{maxMissed}(i_k)$

Table 1. Example of the lattice

In Figure 1, t_1, t_2, \dots, t_n are the sequences in the database. When t_j was under the scan process, itemset v was inserted into V . Suppose we are reading sequence t_i ; therefore, $\text{maxMissed}(v)$ means the number of occurrences of v from t_1 to t_{j-1} , the value of $\text{firstTrans}(v)$ is j , and $\text{Count}(v)$ means the number of occurrences of itemset v from t_j to t_i . For any itemset v in the

lattice, we get a deterministic lower bound $\text{Count}(v)/i$ and upper bound $[\text{maxMissed}(v) + \text{Count}(v)]/i$. We denote these bounds by $\text{minSupport}(v)$ and $\text{maxSupport}(v)$, respectively.

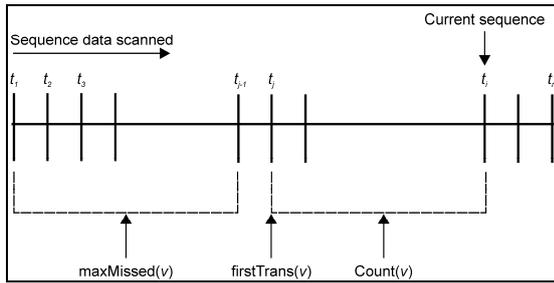


Figure 1. Scan of the sequence database

First, we initialize V to $\{\emptyset\}$ and set $\text{Count}(\emptyset) = 0$, $\text{firstTrans}(\emptyset) = 0$, and $\text{maxMissed}(\emptyset) = 0$. Thus, V is a support lattice for an empty sequence. Suppose V is a support lattice up to sequence $i-1$, we are reading the i -th sequence t_i , and we want to transform V into a support lattice up to i ; there are three steps to go through.

First, for each itemset v in V , if v is contained in t_i , let $\text{Count}(v) = \text{Count}(v) + 1$. We insert a subset v of t_i into V , if and only if all subsets w of v are already contained in V and satisfy $\text{maxSupport}(w) \geq \sigma_i$ (where σ_i is the current user defined support threshold). As v is contained in the current sequence t_i , let $\text{Count}(v) = 1$, $\text{firstTrans}(v) = i$, and we compute the value of $\text{maxMissed}(v)$. As w is a subset of v , we obtain $\text{maxSupport}(w) \geq \text{maxSupport}(v)$.

Second, for $\text{maxSupport}(w) = [\text{maxMissed}(w) + \text{Count}(w)]/i$, $\text{maxSupport}(v) = [\text{maxMissed}(v) + \text{Count}(v)]/i$ and $\text{Count}(v) = 1$; therefore, we obtain $\text{maxMissed}(v) \leq \text{maxMissed}(w) + \text{Count}(w) - 1$. When we are inserting a subset v into V , the set v is not yet contained in V . Hence, the support of v for the first $(i-1)$ sequences satisfies

$$\text{support}_{i-1}(v) \leq \text{avg}_{i-1}(\lceil \sigma \rceil_{i-1}) + \frac{|v|-1}{i-1}, \text{ where } |v| \text{ means the}$$

number of items in v . In addition, considering $\text{maxMissed}(v) = \text{support}_{i-1}(v) \times (i-1)$, we obtain

$$\text{maxMissed}(v) \leq \lfloor (i-1) \text{avg}_{i-1}(\lceil \sigma \rceil_{i-1}) \rfloor + |v| - 1. \text{ Based on all of}$$

the above, we can define that $\text{maxMissed}(v) = \min\{\lfloor (i-1) \text{avg}_{i-1}(\lceil \sigma \rceil_{i-1}) \rfloor + |v| - 1, \text{maxMissed}(w) + \text{count}(w) - 1 | w \subset v\}$.

And third, We compute $\text{maxSupport} = (\text{maxMissed} + \text{Count})/i$ for each itemset v of V when every k sequences (the value of k is defined by the user) are scanned. For any itemset v whose $\text{maxSupport} < \sigma_i$, we delete v from V .

2.2.2 Second part of CARMA

For the second step, the main aim is to scan the sequences a second time and generate sequential patterns based on the frequent sequences found in the first part of CARMA. In the second step, we compute the precise support of all itemsets v in V and continually remove itemsets with $\text{maxSupport} < \sigma_n$ where σ_n is the last threshold of minSupport . While performing the scanning, all itemsets v of V are checked and the parameters associated with v updated. Two situations may arise: first, if $\text{firstTrans}(v) < i$, then v is considered as a large itemset. If the current sequence index is past firstTrans for all itemsets in the lattice, the second part of the CARMA algorithm stops. And second, if the current sequence contains itemset v of V , we set

$\text{Count}(v) = \text{Count}(v) + 1$ and $\text{maxMissed}(v) = \text{maxMissed}(v) - 1$, and if $\text{firstTrans}(v) = i$, we set $\text{maxMissed}(v) = 0$. However, setting $\text{maxMissed}(v) = 0$ for an itemset v , might yield $\text{maxSupport}(w) > \text{maxSupport}(v)$ for some superset w of v . Thus, we set $\text{maxMissed}(w) = \text{Count}(w) - \text{Count}(v)$ for all supersets w of v with $\text{maxSupport}(w) > \text{maxSupport}(v)$. We also remove the itemsets v from V with $\text{maxSupport} < \sigma_n$.

From all the above, it is clear that CARMA only requires two scans of the sequences to obtain the sequential pattern.

2.3 Data

A time series of Landsat TM images, path 123/row 32, for 2006–2011 were selected as the data source (Table 2). The chosen acquisition time was June–July; this corresponds to the spring and summer for most areas in northern China, which are the best seasons for analyses of land cover change. Satellite data used in the experiment were all downloaded free of charge from two websites, <http://ids.ceode.ac.cn/> and <http://glovis.usgs.gov/>.

Image data	Acquisition date
I	7.12.2006
II	5.28.2007
III	5.30.2008
IV	7.20.2009
V	6.5.2010
VI	6.8.2011

Table 2. Acquisition time of the six Landsat TM images used in the experiment

As cloud and fog can seriously affect the results of the mining, we chose the northern part of Beijing as the study area, which was clear in all six images (Figure 2).

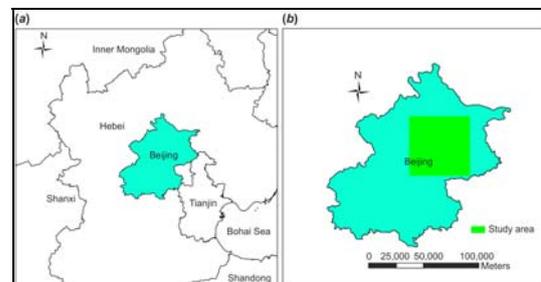


Figure 2. Overview of study area, (a) location of Beijing within the People's Republic of China, (b) location of the study area within Beijing

2.4 Experimental Procedure

For the rational and effective analysis of land cover changes, after the image pre-processing, we firstly classified the six time-series images of the study area into land cover maps. Secondly, based on the land cover map, we constructed the image sequence set within which each sequence is a land cover class trajectory at pixel level that is described through the classified images assembled in the time series. Thirdly, we applied the sequential pattern mining algorithm to the image sequence set to search for sequential patterns. Finally, we analyzed some interesting sequential patterns to reveal the trajectory of land cover change and evaluated the degree of change. The flowchart of the experimental procedure is shown in Figure 3.

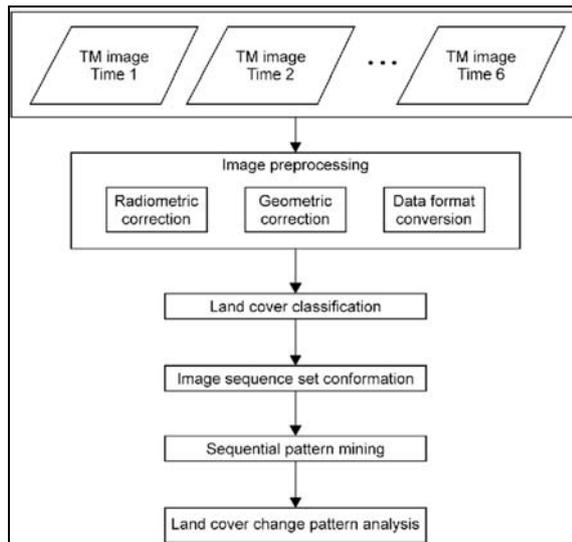


Figure 3. The flowchart of the experimental procedure

2.5 Pre-processing and classification

The pre-processing steps included atmospheric correction, geometric correction, and data format conversion. Landsat TM image from 2011, obtained from the United States Geological Survey, is orthorectified and therefore, it was selected as the master or reference data for geometric correction. Ground control points were collected from the reference image for rectifying the remaining images from 2006 to 2010 (i.e., relative registration) (Kennedy and Cohen, 2003). In this paper, the Gauss-Kruger projection was used in the geometric correction, and a minimum of 20 evenly distributed ground control points were selected to ensure geometric precision of 0.5 pixel (<15 m) for all images. Thus, all the data were arranged in the same coordinate system to form a data set with consistency and integrity, suitable for spatial and sequential comparative analyses.

For image classification, we used eCognition software, and for visualization, the band composite included bands 5, 4, and 3 in the TM data, shown as red, green, and blue, respectively. Because differences and disagreements may appear in the classification process when interpreting land cover types, the classification for all six remote sensing images was undertaken by a single expert in a manner combining software and manual techniques. The land cover types were classified into four categories: built-up area, vegetation, bare land, and water bodies, using a modified Anderson land cover classification scheme (Anderson, 1976). After all the pre-processing steps, the original digital number values for every pixel in the six images were transformed into the value of land cover type. Then, the land cover types or classes were converted into symbols. The experiment used four characters "1", "2", "3", and "4" as a representation for built-up area, bare land, vegetation, and water bodies, respectively. The classification results are shown in Figure 4 and the statistical results for the classification are shown in Table 3.

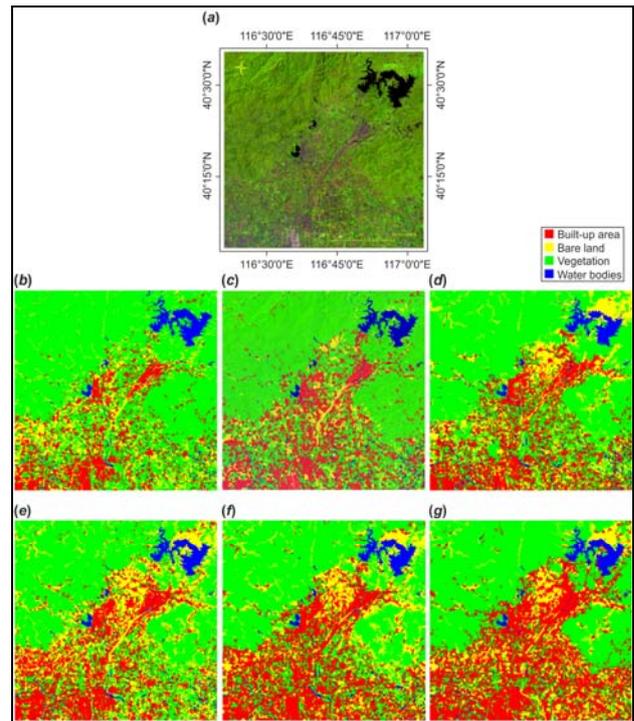


Figure 4. (a) Original satellite image of the study area, (b)-(g) the classification results of the study area from 2006 to 2011

Land cover types	1	2	3	4
2006				
Number of pixels	590487	525164	2735895	148454
Percentage of the research area (%)	14.76	13.13	68.40	3.71
2007				
Number of pixels	848970	638272	2362560	150198
Percentage of the research area (%)	21.22	15.96	59.06	3.76
2008				
Number of pixels	879715	275484	2692544	152257
Percentage of the research area (%)	21.99	6.89	67.31	3.81
2009				
Number of pixels	900253	889216	2071981	138550
Percentage of the research area (%)	22.51	22.23	51.80	3.46
2010				
Number of pixels	1094414	679474	2094474	131638
Percentage of the research area (%)	27.36	16.99	52.36	3.29
2011				
Number of pixels	1330093	510013	2039183	120711
Percentage of the research area (%)	33.25	12.75	50.98	3.02

Table 3. Statistical results for the classification of the six remote sensing images

Considering the image from 2006 as an example: 590487 pixels were interpreted as built-up area, ac Counting for up to 14.76% of the research area; 525164 pixels were interpreted as bare land, ac Counting for 13.13% of the research area; and 2735895 and 148454 pixels, ac Counting for 68.40% and 3.71% of the

research area were interpreted as vegetation and water bodies, respectively.

Because of the lack of quality ground data, it is difficult to quantify the error in classifications at the pixel level (Foody, 2002), especially for multi-temporal analysis (Lunetta, 1999). Although the classification accuracy of year 2011 and year 2010 nearly reached 87% and 85%, the degree of uncertainty is not possible to quantify.

2.6 Construction of land cover change sequence set

An image time series portraying the same scene can be transformed into a landscape trajectory by decomposing the sequence image-by-image and projecting it as a time-ordered series of coordinates in a pattern metric space (Henebry and Goodin, 2002). In our research, as land classes were converted to symbols, a categorical land cover change sequence set that contains the pixel history, or the land cover trajectory, at the pixel-level, was created by obtaining each sequence for every pixel transition. Suppose the classification results for the pixel located in position (1, 1) in the image is vegetation in year 2006, vegetation in 2007, vegetation in 2008, bare land in 2009, bare land in 2010, and built-up area in 2011; then, the land cover change sequence for this pixel can be denoted as “333221”. Therefore, the land cover change sequence set can be generated by copying the land cover change sequence sequentially for each pixel in the study area from the beginning of the image to the end.

2.7 Sequential pattern mining

To identify the typical land cover changes within the six-year period, sequential pattern mining was performed on the constructed land cover change sequence set. In the mining process, the two most important parameters are the support and confidence of the sequence mode. We selected a number of different combinations to establish the most appropriate support and confidence values and tested the resultant sequential patterns, as shown below.

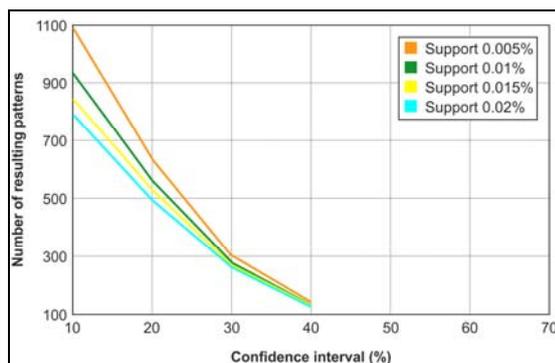


Figure 5. Performance comparison of different values for the CARMA parameters (“support” and “confidence”)

In Figure 5, the numbers of generated patterns of different support tend to coincide as confidence increases. Therefore, in subsequent experiments, we selected a confidence level of 40% and a support rate of 0.02% as the parameters for the sequential pattern mining. To select the most representative land cover change patterns in the different periods, the resultant patterns were subdivided into two types. If the support rate of a pattern was more than 0.1%, it was considered as a selected pattern; if

not, it was not selected. Furthermore, to characterize the direction of change, a distinction was made between continuous and discontinuous patterns within the selected patterns. Continuous patterns are characterized by pixels that belong to the same class (e.g.: 111111 and 222222), whereas discontinuous patterns are characterized by pixels that change class (e.g.: 111112, 121211) through time.

2.8 Analysis

For the study area, the mining process over the six images led to the identification of 118 sequential patterns, each with their own proportion. Table 4 presents the top ten land cover pixel trajectories, the support rates for which were more than 0.1%, and which were selected as land cover patterns.

Pattern	Support (%)	Confidence (%)
333333	44.30	83.43
111111	12.83	58.56
322322	3.97	67.71
222322	2.92	53.39
444444	2.58	80.19
433333	0.80	52.55
322423	0.60	40.79
442444	0.17	71.83
443444	0.17	68.07
322422	0.10	70.98
Total	68.44%	

Table 4. The resultant top 10 patterns

Among all the resultant sequential patterns, the selected top 10 patterns account for 68.44% of the land cover change of the entire study area (Table 4). The remainder is spread among the other 108 patterns. Table 5 shows the composition of land cover change in the selected top 10 patterns for each year, which reveals that the land cover change trajectories are dominated by the transition from vegetation to bare land, especially during 2009 to 2010. It also shows that during 2008 to 2009, the highest percentage (7.83%) of the area was affected by land cover change.

Land cover change	2006–2007	2007–2008	2008–2009	2009–2010	2010–2011	Total
2–3	0	0	6.89%	0	0.6%	7.49%
2–4	0	0	0.77%	0	0	0.77%
3–2	4.67%	0	0	6.89%	0	11.56%
3–4	0	0	0.17%	0	0	0.17%
4–2	0	0.17	0	0.7	0	0.87%
4–3	0.8%	0.17	0	0	0	0.97%
Total	5.47%	0.34%	7.83%	7.59%	0.6%	21.83%

Table 5. Composition of land cover change in the selected top 10 patterns

The percentage area of a land cover type involved in land cover change is illustrated in Table 6. From this table, we can conclude that the trajectories are dominated by transitions that contain vegetation, especially during 2009 to 2010, and that the most stable land cover type is built-up area. These conclusions are also interesting because they reflect our domain knowledge that vegetation is liable to change and built-up area are not.

Land cover type in land cover change	2006–2007	2007–2008	2008–2009	2009–2010	2010–2011	Total
1	0	0	0	0	0	0
2	0	0	7.66%	0	0.6%	8.26%
3	4.67%	0	0.17%	6.89%	0	11.73%
4	0.8%	0.34%	0	0.7	0	1.84%
Total	5.47%	0.34%	7.83%	7.59%	0.6%	21.83%

Table 6. Different land cover types contained in the land cover change

Although the support and confidence in the mining process can help identify the sequential pattern, alone they are insufficient. Not all the resultant patterns can be considered as meaningful (i.e., to reflect the needs and interests of a particular user). Therefore, an analysis of pattern interestingness is essential for the resultant patterns. Typically, only the user can make a judgment of the degree of pattern interestingness, and this judgment is subjective. As the image sequence mining is a data-driven model, it requires expert knowledge to analyze and interpret the semantic meaning of the generated sequential patterns. Therefore, we describe just a few interesting patterns for demonstration purposes.

Among the selected patterns, three are continuous from 2006 to 2011 and they jointly represent 59.71% of the study area. Therefore, the majority of the study area has remained in the cluster to which it belonged in 2006 for the duration of the entire time series. Considering the continuous pattern “111111” as an example, the spatial distribution of this pattern is shown as the pixels within the red area in Figure 6. The support of the pattern is 12.83%, and it shows that 86.9% of the built-up area of cluster 1 in 2006 still belonged to that same cluster in 2011, and that 38.59% of the built-up area in 2011 originated from that same cluster in 2006. Thus, the current land area share of the same cluster can have different histories. Thus, this method can provide new ideas and approaches for the study of urban development.

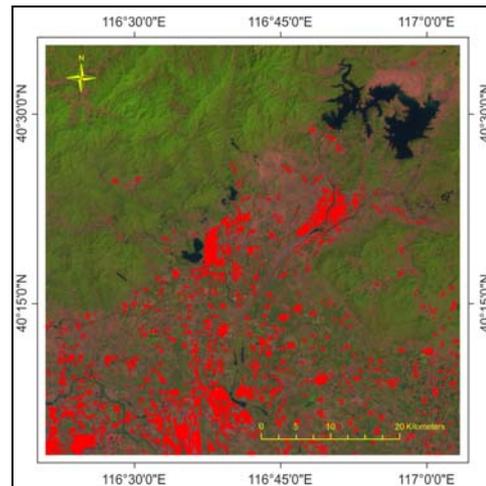


Figure 6. Spatial distribution of the continuous pattern “111111” (in red)

Considering the discontinuous pattern “442444” as an example, the support of the pattern is 0.17% and its spatial distribution is shown in Figure 7.

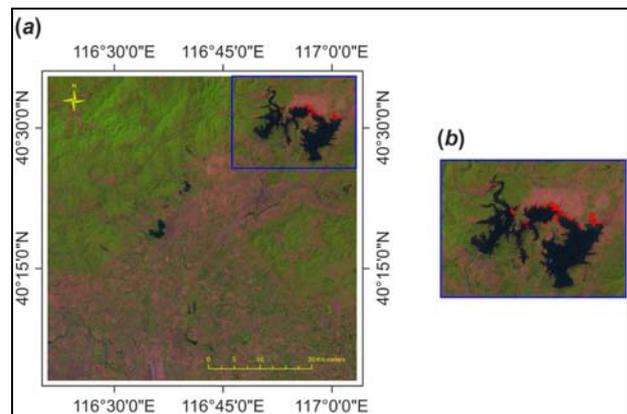


Figure 7. (a) Spatial distribution of the discontinuous pattern “442444” (in red), (b) Partial enlarged image detail of (a)

In Figure 7, the pixels within the red area represent the pattern “442444” and they are mainly located on the waterside of Miyun Reservoir. The “442444” pattern means that since 2006, except for 2008, the area represented by this pattern belonged to cluster 4 (water bodies). However, in 2008, this area belonged to cluster 2 (bare land). In other words, these areas were not covered by water at the end of May 2008 (the acquisition time of the image was May 30, 2008). Table 7 shows the end-of-year water storage of the Miyun Reservoir from 2006 to 2011, according to the “Beijing Water Resource Bulletin.”

Year	Water volume (billion cubic meters)
2006	1.093
2007	0.976
2008	1.13
2009	1.039
2010	1.066
2011	1.101

Table 7. End-of-year water storage of the Miyun Reservoir

Compared with other years, the water storage of the Miyun Reservoir at the end of 2008 was significantly higher, which is not consistent with the obtained pattern “442444”. However, according to the Beijing Water Authority, the water level of the Miyun Reservoir was reduced to a minimum of 853.1 million cubic meters by the end of June 2008. The main flood season began in late July and by September 6, the water volume of the Miyun Reservoir exceeded one billion cubic meters. Therefore, because of the lower water level at the time of image acquisition, some additional bare land was exposed, which is reflected in “442444” pattern in the sequence mining results.

3. CONCLUSION

The method developed in this study has allowed the identification, classification, and spatial localization of land cover types and their trajectories of change for a temporal series. It quantified the land cover changes in terms of the percentage of area affected, as well as mapping the spatial distribution of these changes. It has also provided a different measure for the description of land cover change according to their current characteristics and history. The expected novel significance of this study is the generalization of the application of the sequential pattern mining method for capturing the spatial variability of landscape patterns and their trajectories of change, to reveal information regarding process regularities with satellite imagery.

Although the presented case study clearly demonstrates that the sequential pattern mining method is a promising analytical tool for spatiotemporal data analysis, a number of issues warrant further investigation. As with other studies using historical data for studying landscape changes, the availability and quality of the data, their classification, and analysis all influence the typology of the landscape patterns and of the changes detected (Antrop, 1998). Discovering interesting patterns is also an important requirement in this field and in future research; we intend to develop interestingness and mining methods that are more sophisticated, to improve the utility and efficiency of applying sequential pattern mining to remote sensing data. Moreover, we will try to set a threshold of minimal covered area used to clump the isolated results, to avoid the outliers and other minority patterns which would introduce errors and further problems to understand the results. And most important of all, as the more snapshots included in the time series, the more complex the pattern code will be. So, providing more efficient method to dissolve and understand the pattern code would be the major challenge.

ACKNOWLEDGEMENTS

The research has been supported by grants from the National Natural Science Foundation of China (61271013), the National High Technology Research and Development Program of China (863 Program) (2012BAH27B05), the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (under the project “135 Strategy Planning”), and National Ecological Environment Change Assessment by Remote Sensing Survey Project 2000-2010 (STSN-10-03). The majority of Landsat image data were obtained free of charge from the Open Spatial Data Sharing Project of the Institute of Remote Sensing and Digital Earth Chinese Academy of Sciences (<http://ids.ceode.ac.cn/>). Additionally, data outside that

available in the archive were downloaded from the USGS Landsat archive via GloVis interface (<http://glovis.usgs.gov/>).

References

- Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer, 1976. A land use and land cover classification system for use with remote sensor data. *Usgs Prof Pap*, 964.
- Antrop, M., 1998. Antrop, M. (1997). Landscape change: plan or chaos?. *International Conference Perspectives of Rural Landscapes in Europe Prnu Estonia*, 41(98), pp. 155–161.
- Boriah, S., 2010. Boriah, S. (2010). Time series change detection: algorithms for land cover change. *Dissertations & Theses – Gradworks*, University of Minnesota, US, 161 pages; 3403358
- Boriah, S., V. Kumar, M. Steinbach, P. N. Tan, C. Potter, and S. Klooster, 2009. Detecting ecosystem disturbances and land cover change using data mining. *In Next Generation of Data Mining*, CRC, USA.
- Cai, Z., and X. Wang, 2010. Research on vegetation dynamic change simulation based on spatial data mining of ANN-CA model using time series of remote sensing images. *Advances in Information & Communication Technology*, pp. 551–557
- Chen, W., T. Sakai, C. Cao, K. Moriya, and L. Koyama, 2012. Detection of forest disturbance in the Greater Hinggan Mountain of China based on Landsat time-series data. *IGARSS*, pp.7232–7235.
- Coppin, P., I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, 2004. Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, 25(9), pp. 1565–1596.
- De Beurs, K. M., and G. M. Henebry, 2005. A statistical framework for the analysis of long image time series. *International Journal of Remote Sensing*, 26(8), pp. 1551–1573.
- Dubinina, M., P. Potapov, A. Lushchekina, and V. C. Radeloff, 2010. Reconstructing long time series of burned areas in arid grasslands of southern Russia by satellite remote sensing. *Remote Sensing of Environment*, 114(8), pp. 1638–1648.
- Dusseux, P., L. Hubert-Moy, R. Lecercq, X. Gong, and T. Corpetti, 2011. Identification of grazed and mown Grasslands using a time series of high spatial resolution remote sensing images. 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images, Trento, Italy.
- Foody, G.M., 2002. Status of land-cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), pp. 185–201.
- Gimmi, U., T. Wohlgemuth, A. Rigling, C. W. Hoffmann, and M. Bürgi, 2010. Land-use and climate change effects in forest compositional trajectories in a dry Central-Alpine valley. *Annals of Forest Science*, 67(7), pp. 701.
- Godoy, R., and M. Contreras, 2001. A comparative study of education and tropical deforestation among lowland Bolivian Amerindians: Forest values, environmental externality, and school subsidies. *Economic Development and Cultural Change*, 19(3), pp. 555–574.

- Gueguen, L., and M. Datcu, 2007. Image Time-Series Data Mining based on the Information-Bottleneck Principle. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), pp. 827-838.
- Henebry, G. M., and D. G. Goodin, 2002. Landscape trajectory analysis: toward spatio-temporal models of biogeochemical fields for ecological forecasting. Workshop on Spatio-temporal Data Models for Biogeophysical Fields, La Jolla, CA, USA, April 8–10.
- Hudak, A. T., and C. A. Wessman, 1998. Textural analysis of historical aerial photography to characterize woody plant encroachment in South African savanna. *Remote Sensing of Environment*, 66(3), pp. 317–330.
- Jin, S. M., and S. A. Sader, 2006. Effects of forest ownership and change on forest harvest rates, types and trends in northern Maine. *Forest Ecology and Management*, 228, pp. 177–186.
- Karpatne, A., J. Faghmous, J. Kawale, L. Styles, M. Blank, V. Mithal, X. Chen, et al., 2013. Earth Science Applications of Sensor Data. *Managing and Mining Sensor Data*, USA: Springer, pp. 505–530.
- Kennedy, R. E., and W. B. Cohen, 2003. Automated designation of tie-points for image-to-image coregistration. *International Journal of Remote Sensing*, 24(17), pp. 3467–3490.
- Kennedy, R. E., W. B. Cohen, T. A. Schroeder, 2007. Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, 110(3), pp. 370.
- Lambert, J., A. Jacquin, J. P. Denux, and V. Cheret, 2011. Comparison of two remote sensing time series analysis methods for monitoring forest decline. In 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), Italy: The Institute of Electrical and Electronics Engineers, Inc.
- Lambin, E. F., and D. Ehrlich, 1997. Land-cover changes in sub-Saharan Africa (1982-1991). *Remote Sensing of Environment*, 61, pp. 81–200.
- Le Ber, F., M. Benoît, C. Schott, J.-F. Mari, C. Mignolet, 2006. Studying crop sequences with CarrotAge, a HMM-based data mining software. *Ecological Modelling*, 191, pp.170-185.
- Liu, J. Y., M. L. Liu, H. Q. Tian, W. Chen, T. Sakai, C. Cao, K. Moriya, and L. Koyama, 2005. Spatial and temporal patterns of China's cropland during 1990-2000: An analysis based on Landsat TM data. *Remote Sensing of Environment*, 98 (4), pp. 442–456.
- Lu, D., P. Mausel, E. Brondi'zio, and E. Moran, 2004. Change detection techniques. *International Journal of Remote Sensing*, 25(12), pp. 2365–2407.
- Lunetta, R., 1999. Applications, project formulation, and analytical approach. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, Taylor and Francis Ltd., London.
- Mannila, H., and H. Toivonen, 1996. Discovering generalized episodes using minimal occurrences. *Second International Conference on Knowledge Discovery and Data Mining (KDD '96)*, AAAI, Portland, OR, pp. 146–151.
- Mertens, B., and E. F. Lambin, 2000. Land-cover-change trajectories in Southern Cameroon. *Annals of the Association of American Geographers*, 90(3), pp. 467–494.
- Petit, C., T. Scudder, and E. Lambin, 2001. Quantifying processes of land-cover change by remote sensing: resettlement and rapid land-cover changes in south-eastern Zambia. *International Journal of Remote Sensing*, 22(17), pp. 3435–3456.
- Puyravaud, J. P., 2003. Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management*, 177(1–3), pp. 593–596.
- Randerson, J. T., H. Liu, M. G. Flanner, S. D. Chambers, Y. Jin, P. G. Hess, G. Pfister, et al., 2006. The impact of boreal forest fire on climate warming. *Science*, 314(5802), pp. 1130–1132.
- Schneider, A., 2012. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sensing of Environment*, 124, pp. 689–704.
- Sierra, R., 2000. Dynamics and patterns of deforestation in the western Amazon: The Napo deforestation front, 1986–1996. *Applied Geography*, 20(1), pp. 1–16.
- Turner, B. L., E. F. Lambin, and A. Reenberg, 2007. The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, 104(52), pp. 20666–20671.
- Verbesselt, J., R. Hyndman, G. Newnham, and D. Culvenor, 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1), pp.106–115.