Sequential pattern mining of land cover dynamics based on time-series remote sensing images

Huichan Liu [1,2,3] · Guojin He [1,3] · Weili Jiao [1,3] ·
Guizhou Wang [1,3] · Yan Peng [1,3] · Bo Cheng [1,3]

Abstract Remote sensing images constitute a new type of multimedia data well suited to land cover change detection tasks, as they can repetitively provide information about the land surface and its changes over large and inaccessible areas. With plans for more missions and higher resolution earth observation systems, the challenge is increasingly going to be the efficient usability of the millions of collected images, especially the decades of remote sensing image time series, to describe land cover and/or scene evolution and dynamics. 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. 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. Experimental results showed that this approach not only quantifies land cover changes in terms of the percentage area affected and maps the spatial distribution of these land cover changes but also reveals possibly interesting or useful information regarding the trajectories of change. This method is a valuable complement to existing bi-temporal change detection methods.

Keywords Remote sensing satellite · Image time series · Data mining · Sequential pattern mining · Land cover change

---

✉ Guojin He
hegj@radi.ac.cn

1   Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, No. 9 Dengzhuang South Road, Haidian District, Beijing 100094, People's Republic of China

2   University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

3   Hainan Key Laboratory for Earth Observation, Sanya 572029, China

1 Introduction

With the development of remote sensing technologies, satellite images with the characteristics of multi-scale, multi-band, and multi-date make it tend to be big data [7, 16, 22, 25]. There is a major computational challenges to extract the useful information from remote sensing big data in an efficient manner, such as to analyze, aggregate, and store, especially for time-series remote sensing images. Remote sensing has long been used as a means of detecting and classifying temporal changes in the condition of the land surface [5, 15]. 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 [13]. 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 [6]. In order to efficiently use the huge amounts of data that will be produced by plans for more missions and higher resolution earth observation systems, a review of the currently used approaches is needed. On the one hand, most previous change detection studies have relied primarily on examining the differences between two or more satellite images acquired on different dates [3]. These procedures can be categorized into three types. Procedures in the first category consist of combining the values of the image at time t and those at time $(t - 1)$ in order reveal the intrinsic temporal structure of the data. The combination operator can be a subtraction [18], a division [11], or more sophisticated ones [20]. And then, the resulting combined image is classified to map the change areas. Further, procedures in the second category involve building a vector from two multi-band values (one before and one after the change) in a multi-dimensional space. The norm and angle of this vector is used as information on the type and intensity of the change [4]. Finally, procedures in the third category considers that the pixel value at time t is linearly correlated to the pixel value at time $(t - 1)$. The parameters of the regression (e.g., the residual) are studied to map and characterize the change [12]. These methods mentioned above are only suitable for two images at one time. Thus, in order to handle time-series images, they all have to be applied for several times (e.g., twelve images combined in one in [23]), which leads to hardly understandable results.

On the other hand, although some research progress has been reported with regard to the analysis of temporal trajectories, the method can be improved further to suit time-series remote sensing images processing. Some methods use a Fourier or a wavelet decomposition to analyze the time series images. Although these robust methods can handle time series as a whole, they require a regular sampling of the time series, which means data must be regularly sensed. For remote sensing time series images, this constraint is difficult to keep, since the acquisitions depend on several factors (operational, meteorological, etc.). Sequential pattern mining method is another way to deal with the time-series remote sensing images. This type of methods extract frequent sub-sequences of radiometric evolutions or land cover change trajectories by just using the ordering of the sequence. It has the advantage of being robust to noise and extracting meaningful patterns. As remote sensing data are recognized as "big data" in a certain sense [17], it is necessary to consider the performance efficiency of finding meaningful or interesting patterns from time-series remote sensing images. Therefore, there is a critical need for a method that can enable efficient and reliable characterization of spatio-temporal patterns contained in an image time series [9]. The objective of this study is to

generalize the application of the sequential pattern mining method to time-series remote sensing images by using the Continuous Association Rule Mining Algorithm (CARMA) [10] to improve the mining efficiency.

The remainder of this article is organized as follows. Section 2 describes the sequential pattern mining method. Section 3 presents a case study and an analysis of the results. Finally, Section 4 discusses the principal findings and states our conclusions.

2 Methodology

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 [9]. One basic problem when analyzing sequences of items is to find frequent episodes [14], or in other words, to extract regular patterns from temporal data. To perform such extraction tasks, we can rely on the various algorithms that have been designed to extract frequent sequential patterns. Mining sequential and spatial patterns is an active area of research in data mining, which is used for string mining or itemset mining in transactions analysis.

In data mining field, several algorithms can be used to extract sequential patterns, and most of them can be divided into two kinds: the Apriori-like, breadth-first search methods and the pattern-growth, depth-first search methods. The bottlenecks of the breadth-first search methods (such as AprioriAll [1], AprioriSome [1] and GSP [21]) include potentially huge sets of candidate sequences, multiple scans of databases and difficulties at mining long sequential patterns. And for the depth-first search methods (such as cSPADE [24], FreeSpan [8] and PrefixSpan [19]), the major cost is the construction of projected databases. All traditional algorithms operate offline: given minimal values for the support and the confidence the algorithms scan and rescan the sequence set, often several times, and eventually produce all association rules. However in general, the user does not know the appropriate thresholds in advance. An inappropriate choice yields, after a long wait, either too many or too few association rules. Among the various algorithms, CARMA is an efficient two-pass method for finding sequences. CARMA is an alternative to Apriori that reduces Input / Output costs, time, and space requirements. It uses only two data passes to deliver results which is much lower support levels than Apriori. In addition, it allows changes in the support level during execution. The time consumed is a crucial factor in remote sensing image processing, so in this paper we aim to apply CARMA to exploit sequential pattern mining within the field of data mining in order to analyze time-series remote sensing images.

This section introduces the sequential pattern mining method based on the CARMA algorithm. 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 $I = \{I_1, I_2, \ldots, I_p\}$ 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 $s = <e_1, e_2, \ldots, e_j>$, itemset e_1 appears before e_2, and e_2 appears before e_3, and so on. Itemset e_j is also an element of the sequence s denoted as $(x_1, x_2, \ldots, x_q)$ in which $x_q \in I$. A sequence that contains k itemsets is a k-sequence. If there exists $1 \leq i_1 < i_2 < \ldots < i_n \leq m$ such that $a_1 \subseteq b_{i_1}$,

$a_2 \subseteq b_{i_2}, \ldots, a_n \subseteq b_{i_n}$ appears in sequence $A = \langle a_1, a_2, \ldots, a_n \rangle$ and sequence $B = \langle b_1, b_2, \ldots, b_n \rangle$, 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 $sup(s)$). 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 $s = <e_1,e_2,\ldots,e_j>$, sequence $s' = <e_1,e_2,\ldots,e_{j-1}>$ is called the antecedent and sequence $s'' = <e_j>$ is called the consequent. The confidence for a sequential pattern can be defined as the ratio of $sup(s)$ to $sup(s')$. The confidence for each generated sequential pattern should be calculated, and those sequential patterns whose confidence is larger than the minimum confidence threshold are recognized as more interesting (or more useful) than the other sequential patterns.

2.2 CARMA algorithm

The CARMA algorithm uses an effective two-step method to discover sequential patterns. The first step is used mainly to find frequent sequences, and the second step is used to generate sequential patterns based on these frequent sequences.

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: $count(v)$, $firstTrans(v)$, and $maxMissed(v)$. The meanings of these three parameters are introduced below and shown in Fig. 1:

$count(v)$: number of occurrences of itemset v since v was inserted in the lattice.
$firstTrans(v)$: index of the sequence data at which v was inserted in the lattice.
$maxMissed(v)$: 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:
In Fig. 1, $t_1,t_2,\ldots,t_n$ are the sequences in the database. When t_j was under the scan process, itemset v was inserted into V. Suppose that we are reading sequence t_i; therefore, $maxMissed(v)$
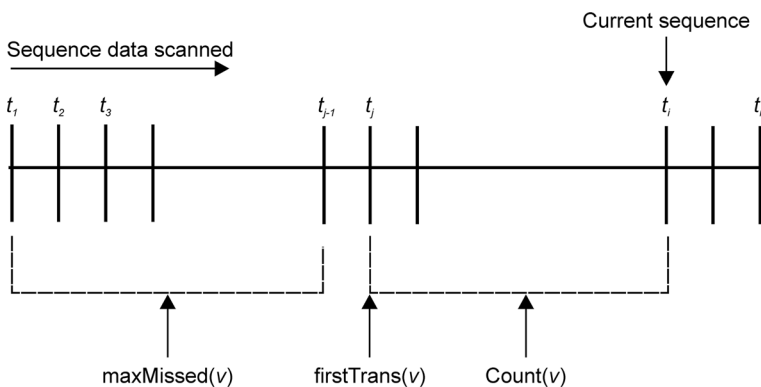

Fig. 1 Illustration of the meanings of the three parameters of CARMA

Table 1 Example of the lattice

Itemset	Count	firstTrans	maxMissed
i_1	$count(i_1)$	$firstTrans(i_1)$	$maxMissed(i_1)$
i_2	$count(i_2)$	$firstTrans(i_2)$	$maxMissed(i_2)$
…	…	…	…
i_k	$count(i_k)$	$firstTrans(i_k)$	$maxMissed(i_k)$

denotes the number of occurrences of v from t_1 to t_{j-1}, the value of *firstTrans(v)* is j, and *count(v)* denotes 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 $count(v)/i$ and upper bound $[maxMissed(v) + count(v)]/i$. We denote these bounds by *minSupport(v)* and *maxSupport(v)*, respectively. The detailed flowchart of CARMA is shown in Fig. 2:

1. The framework of CARMA algorithm has two parts: First, we initialize V to $\{\emptyset\}$ and set $count(\emptyset) = 0$, $firstTrans(\emptyset) = 0$, and $maxmissed(\emptyset) = 0$. Thus, V is a support lattice for an empty sequence. Suppose that V is a support lattice up to sequence i-1; if we are reading the *i-th* sequence t_i and want to transform V into a support lattice up to i, we have to go through three steps.

    1) For each itemset v in V, if v is contained in t_i, let $count(v) = count(v) + 1$.
    2) 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 $maxSupport(w) \geq \sigma_i$ (where σ_i is the current user-defined support
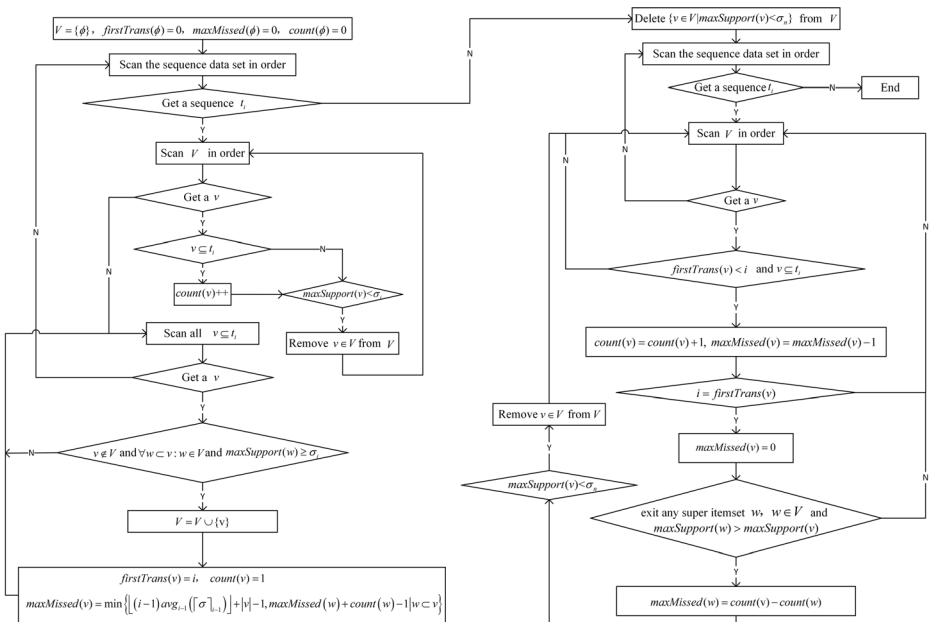

Fig. 2 Flowchart of CARMA

threshold). As v is contained in the current sequence t_i, let $count(v) = 1$, $firstTrans(v) = i$, and we compute the value of $maxMissed(v)$.

- As w is a subset of v, we obtain $maxSupport(w) \geq maxSupport(v)$. Furthermore, according to Formula (2.1), Formula (2.2) and Formula (2.3), we obtain Formula (2.4).

$$maxSupport(w) = [maxMissed(w) + count(w)]/i \qquad (2.1)$$

$$maxSupport(v) = [maxMissed(v) = count(v)]/i \qquad (2.2)$$

$$count(v) = 1 \qquad (2.3)$$

$$maxMissed(v) \leq maxMissed(w) + count(w) - 1 \qquad (2.4)$$

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 Formula (2.5), where $|v|$ denotes the number of items in v.

$$support_{i-1}(v) \leq avg_{i-1}(\lceil \sigma \rceil_{i-1}) + \frac{|v-1|}{i-1} \qquad (2.5)$$

- In addition, considering Formula (2.6), we obtain Formula (2.7)

$$maxMissed(v) = support_{i-1}(v) \times (i-1) \qquad (2.6)$$

$$maxMissed(v) \leq \lfloor (i-1)avg_{i-1}(\lceil \sigma \rceil_{i-1}) \rfloor + |v| - 1 \qquad (2.7)$$

- Based on all of the above, we can define Formula (2.8).

$$maxMissed(v) = min\{\lfloor (i-1)avg_{i-1}(\lceil \sigma \rceil_{i-1}) \rfloor + |v| - 1, maxMissed(w) + count(w) - 1 | w \subset v\} \qquad (2.8)$$

3) We compute Formula (2.9) 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 $maxSupport < \sigma_i$, we delete v from V.

$$maxSupport = (maxMissed + count)/I \qquad (2.9)$$

2. 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 $maxSupport < \sigma_n$ where σ_n is the last threshold of $minSupport$.
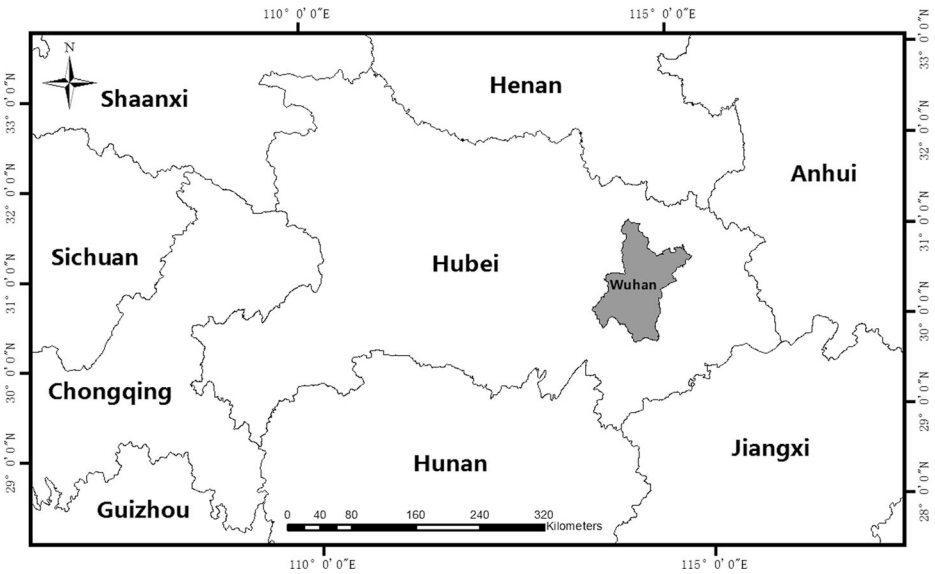
Fig. 3 Location of Wuhan city in Hubei province, China. The spatial extent of the experiment is shaded in light grey

While performing the scanning, all itemsets v of V are checked and the parameters associated with v are updated. Two situations may arise:

1)   If $firstTrans(v) < i$, then v is considered as a large itemset. If the current sequence index is greater than firstTrans for all itemsets in the lattice, the second part of the CARMA algorithm stops.

Table 2 Earth observation data available and processed for the study

Sensor	Year	Acquisition time	Path/Row	Level
TM	2010	20091017	122/039	L4
TM		20090906	123/038	L4
TM		20090906	123/039	L4
TM	2005	20040917	122/039	L4
TM		20050911	123/038	L4
TM		20050911	123/039	L4
ETM+	2000	20000914	122/039	L4
ETM+		20000820	123/038	L4
ETM+		20001007	123/039	L4
TM	1990	19921018	122/039	L4
TM		19930926	123/038	L4
TM		19910719	123/039	L4
MSS	1980	19791020	132/038	L4
MSS		19781016	132/039	L4

2) If the current sequence contains itemset v of V, we use Formula (2.10) and Formula (2.11), and if $firstTrans(v)=i$, we use Formula (2.12). However, using Formula (2.12)for an itemset v might yield $maxSupport(w)>maxSupport(v)$ for some superset w of v. Thus, we use Formula (2.13) for all supersets w of v with $maxSupport(w)>maxSupport(v)$. We also remove the itemsets v from V with $maxSupport<\sigma_n$.
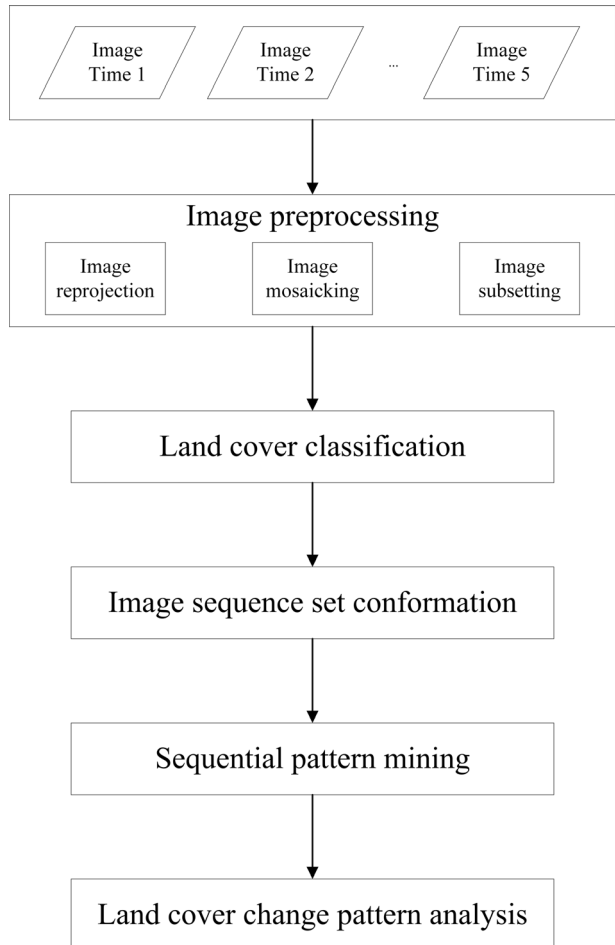
$$count(v) \;=\; count(v) \,+\, 1 \tag{2.10}$$

$$maxMissed(v) = maxMissed(v) - 1 \tag{2.11}$$

$$maxMissed(v) \;=\; 0 \tag{2.12}$$

$$maxMissed(w) = count(v) - count(w) \tag{2.13}$$

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

Fig. 4 Flowchart of the experimental procedure

3 Case study and analysis

The main objective of this paper is to generalize the typical sequential pattern mining algorithm in order to obtain patterns for remote sensing time-series images that could be useful for reaching meaningful conclusions regarding land cover change. Here, we analyze some selected patterns obtained by the CARMA algorithm, which represent the process of land cover change. This approach quantifies land cover changes in terms of the percentage area affected and maps the spatial distribution of these changes.

3.1 Study area and data

The present paper is based on a study of Wuhan city in Hubei province, China. As shown in Fig. 3, Wuhan is located in the east of Hubei province, between 113°41′–115°05′E and 29°58′–31°22′N, and covers an area of around 8494.41 km². The terrain of Wuhan city is dominated by plains and supplemented by hills with a surface elevation ranging from 11.3 m to 873.7 m. The landform falls under the hilly regions in the southeast Hubei province. It is a transitional area between the eastern margin of Jianghan plain and the southern low mountains and hills of the Dabie Mountains. Wuhan city has witnessed rapid urbanization and industrialization owing to a booming economy; the population growth and economic development have contributed to drastic changes in land use/land cover in Wuhan.

Time-series Landsat MSS/TM/ETM+ images for five periods (i.e., 1980, 1990, 2000, 2005, and 2010) were selected as the data source (Table 2) for the study. The images used in the experiment were all downloaded free of charge from two websites: http://ids.ceode.ac.cn/ and http://glovis.usgs.gov/.
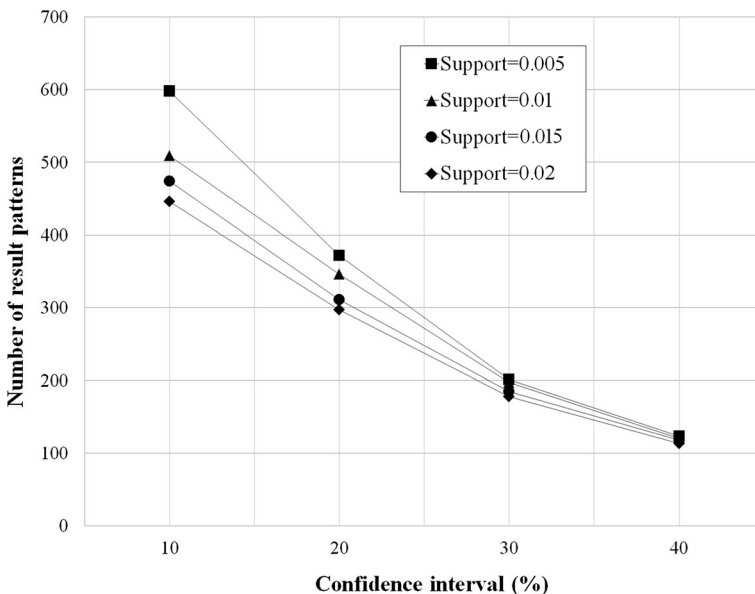

Fig. 5 Relationship of support, confidence, and number of patterns: different combinations to find the appropriate support and confidence values

3.2 Experimental procedure

For rational and effective analysis of land cover changes, after the image pre-processing, we first classified the time-series images of the study area into land cover maps. Second, based on the land cover maps, 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. Third, we applied the sequential pattern mining algorithm to the image sequence set to search for sequential patterns. Finally, we analyzed some interesting
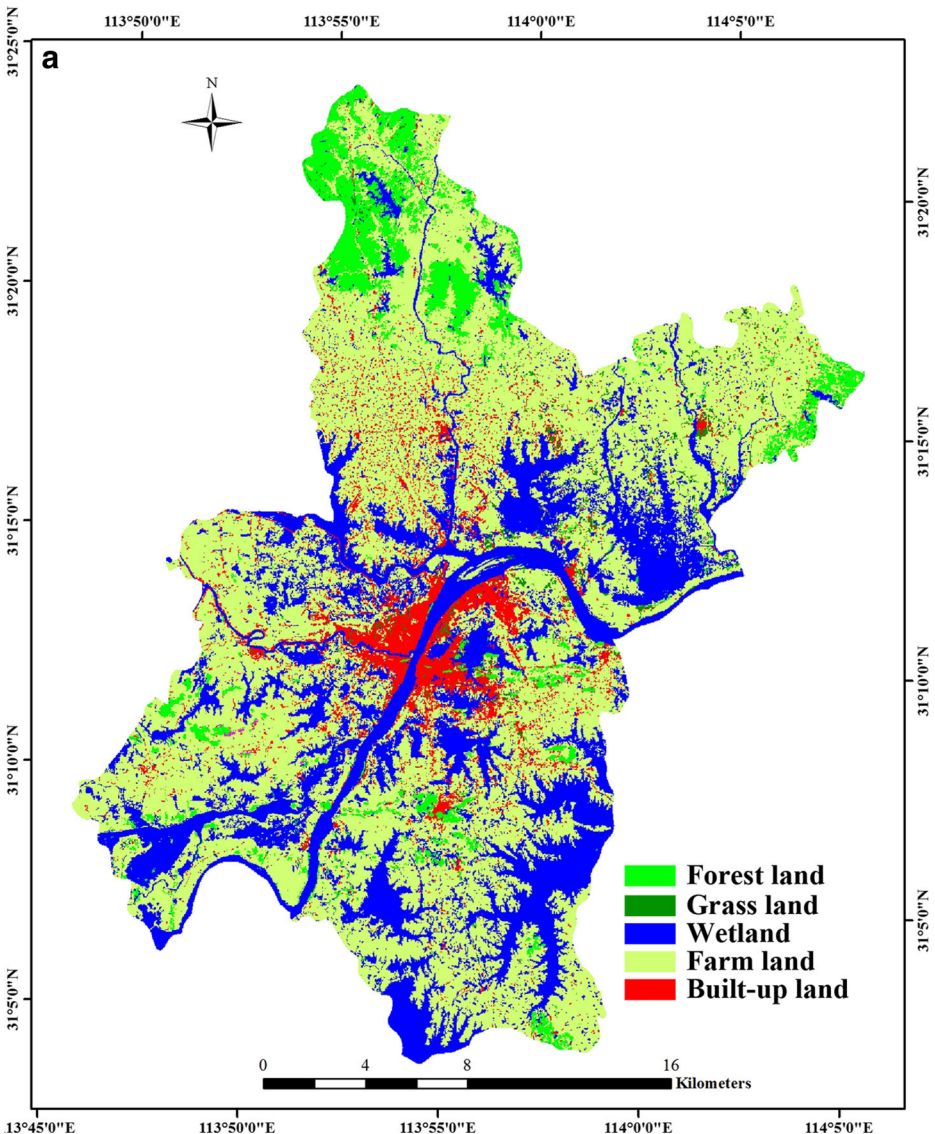

Fig. 6 **a–e** Results of the classification for remote sensing images from 1980 to 2010, respectively
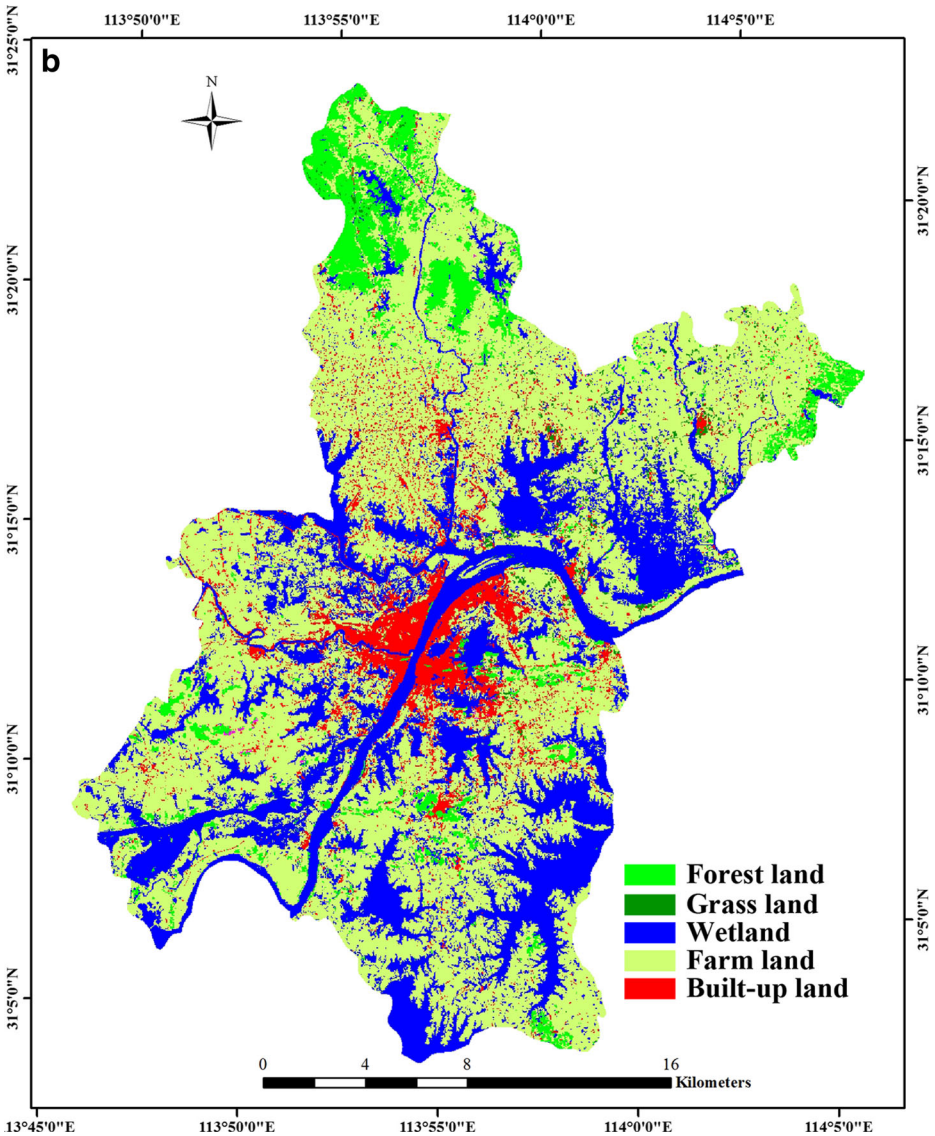
Fig. 6 (continued)

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 Fig. 4.

3.3 Pre-processing and classification

The pre-processing steps included reprojection, image mosaicking, and subsetting. The images were projected to the Albers Conical Equal Area projection coordinate system with detailed parameters as follows: central longitude 105° E and WGS84 spheroid.
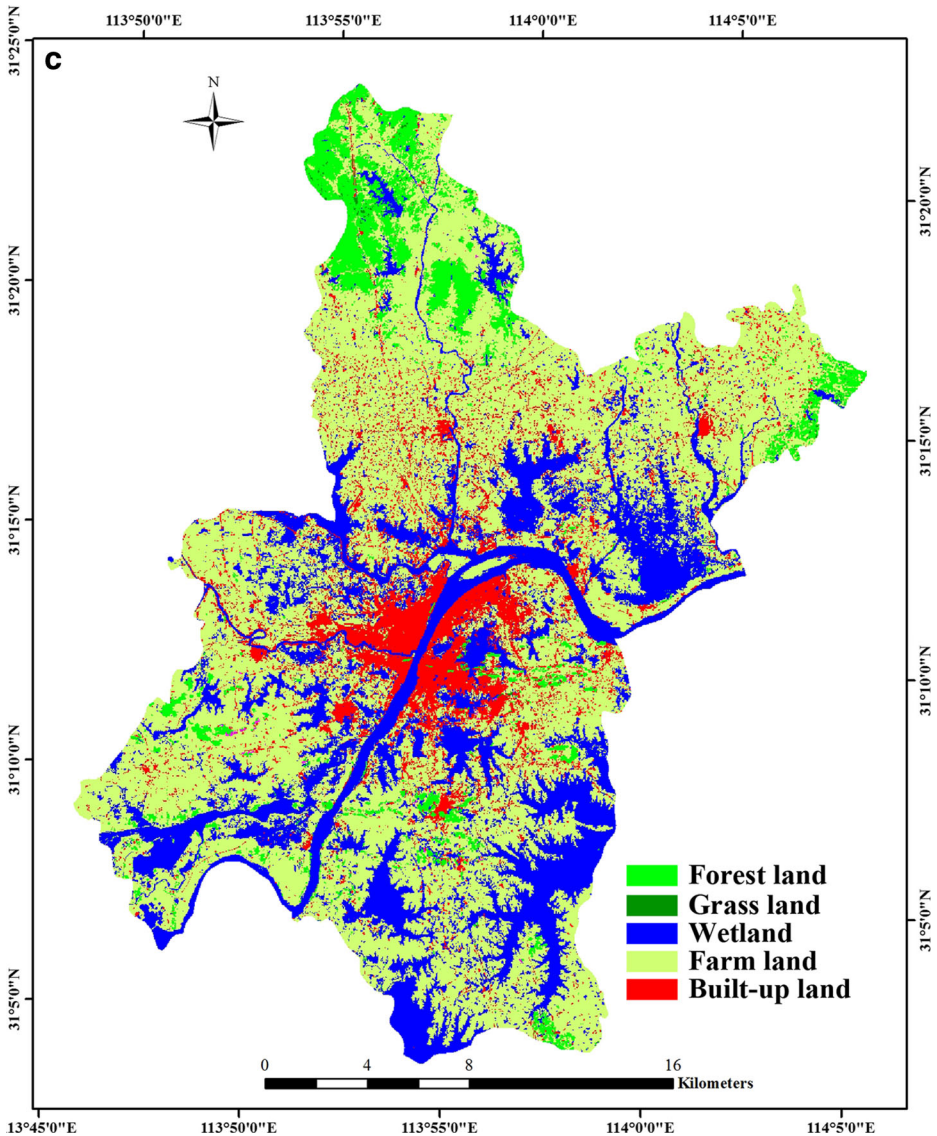
Fig. 6 (continued)

Then, the images for each year were mosaicked in the same coordinate system, and the study area was clipped from the mosaicked images against the municipal boundary layer. 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.

We used eCognition software for image classification, and image enhancement was performed to increase the visual discrimination between features from the data. Because differences and disagreements may appear in the classification process when
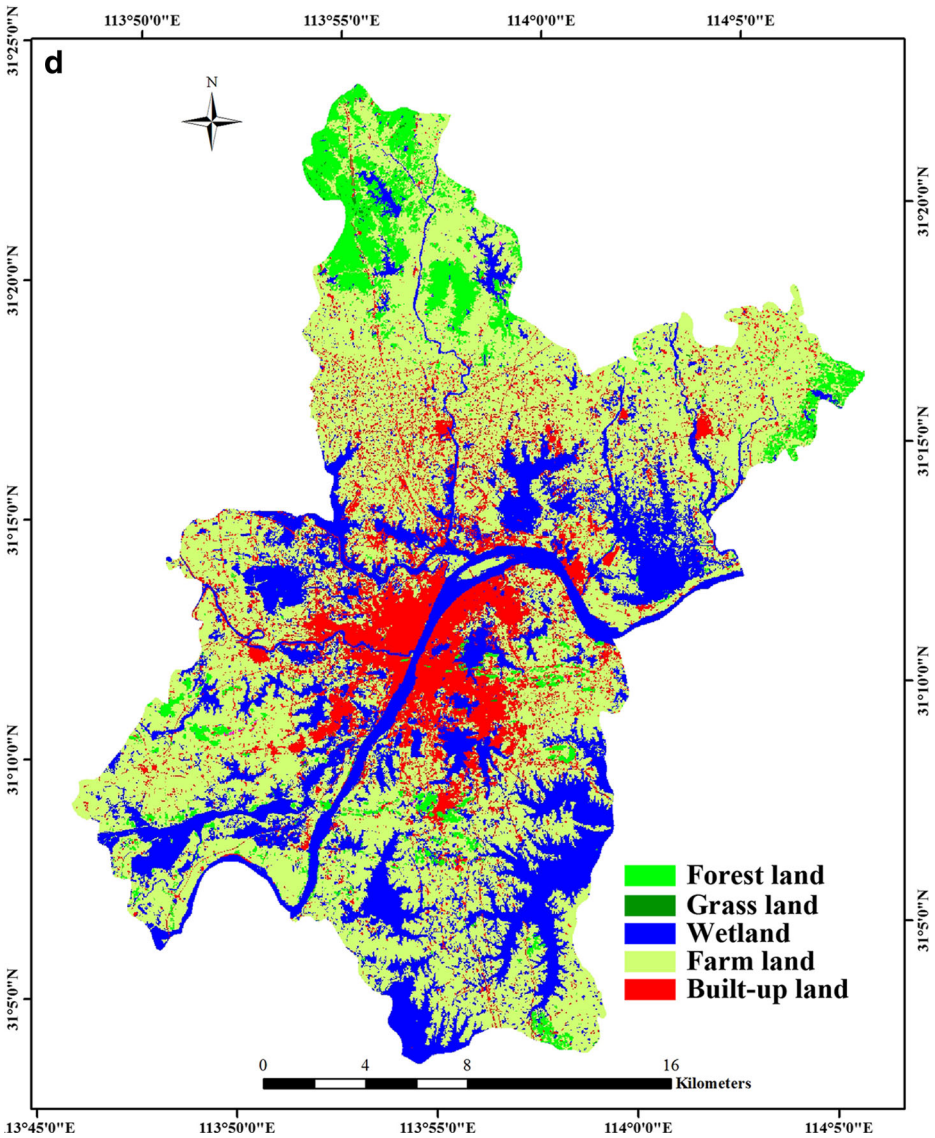
Fig. 6 (continued)

interpreting land cover types, the classification for all the images was undertaken by a single expert by combining software and manual techniques. The land cover types were classified into five categories, namely, forest land, grassland, wetland (including rivers, ponds, and reservoirs), farm land, and built-up land, using a modified Anderson land cover classification scheme [2]. After all the pre-processing steps, the original digital number values for every pixel in the five images were transformed into land cover type values. Then, the land cover types or classes were converted into
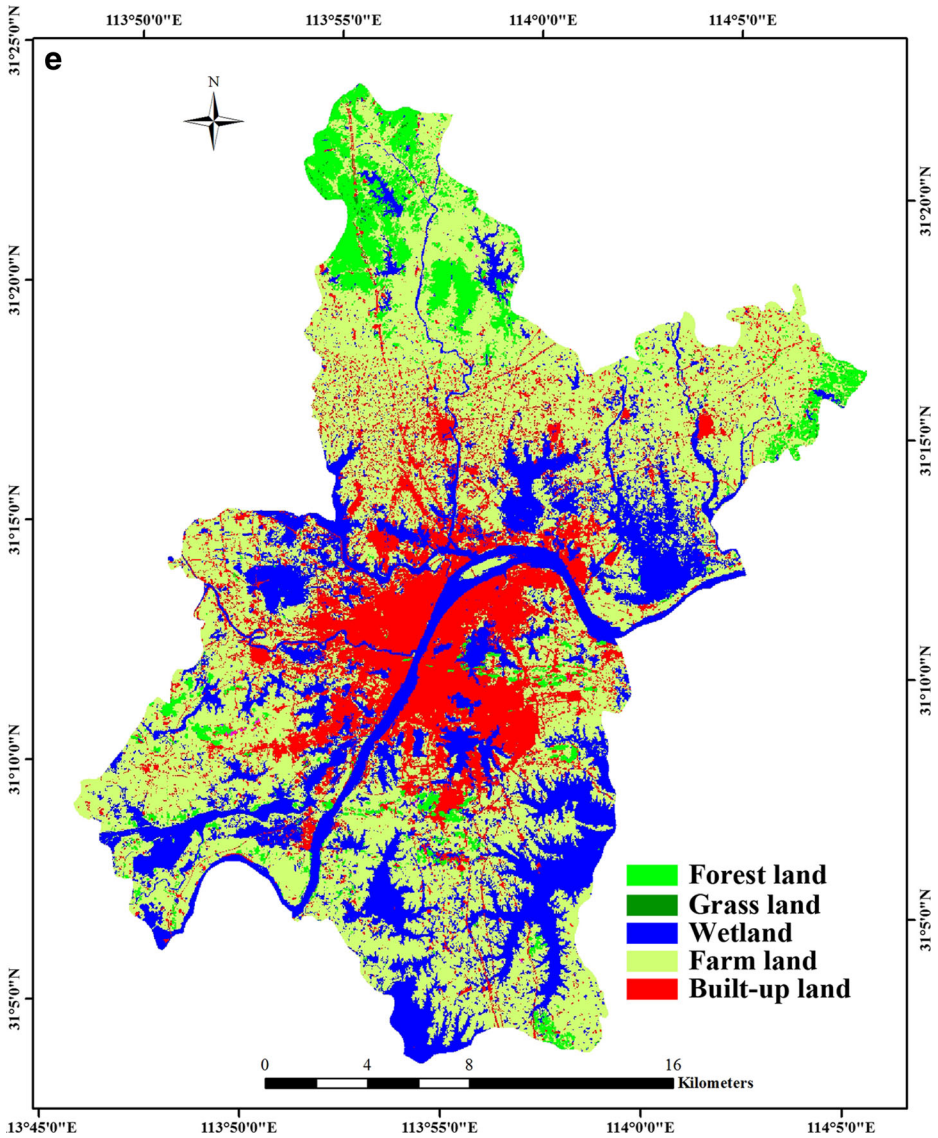
Fig. 6 (continued)

symbols. The experiment used five characters, namely, "1", "2", "3", "4", and "5", to represent forest land, grassland, wetland, farm land, and built-up land, respectively.

3.4 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. 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 that the classification result for the pixel located in position (1, 1) in the image is grassland in 1980, 1990, and 2000, and built-up land in 2005 and 2010; then, the land cover change sequence for this pixel can be denoted as "22255". 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. All the land cover change sequence for each pixel constitute the land cover change sequence set, and this sequence set is the data source for sequential pattern mining.

3.5 Sequential pattern mining

The greater the number of land cover classes, the greater is the number of change trajectories. For three successive land cover classifications with 10 land cover classes each, the potential number of land cover change trajectories is 1000. Analyzing and interpreting all the possible change trajectory results is a time-consuming task; moreover, it is difficult to draw conclusions regarding the land cover change for the entire area. In order to identify the typical land cover changes and determine the bases of the sequences, sequential pattern mining was performed on the constructed land cover change sequence set. Here, we used the CARMA algorithm to explore the sequential patterns that represent the typical land cover changes. In the mining process, the two most important parameters are the support and the confidence of the sequence mode. We selected a number of different combinations to establish the most appropriate support and confidence values and tested the obtained sequential patterns, as shown in Fig. 5.

In Fig. 5, the numbers of generated patterns for different supports tend to coincide as the confidence increases. Therefore, in subsequent experiments, we selected a

Table 3 Area statistics of the land cover types over the 30-year study period

| Year	Land cover types	1 Forest land	2 Grassland	3 Wetland	4 Farm land	5 Built-up land
1980	Area (km^2)	587	86	2321	4944	555
	Percentage of the research area (%)	6.91	1.02	27.33	58.21	6.54
1990	Area (km^2)	587	75	2321	4922	588
	Percentage of the research area (%)	6.91	0.89	27.33	57.95	6.93
2000	Area (km^2)	577	12	2154	4893	858
	Percentage of the research area (%)	6.79	0.14	25.35	57.61	10.11
2005	Area (km^2)	568	11	2148	4660	1107
	Percentage of the research area (%)	6.69	0.13	25.29	54.86	13.03
2010	Area (km^2)	567	11	2112	4307	1497
	Percentage of the research area (%)	6.67	0.13	24.87	50.71	17.63
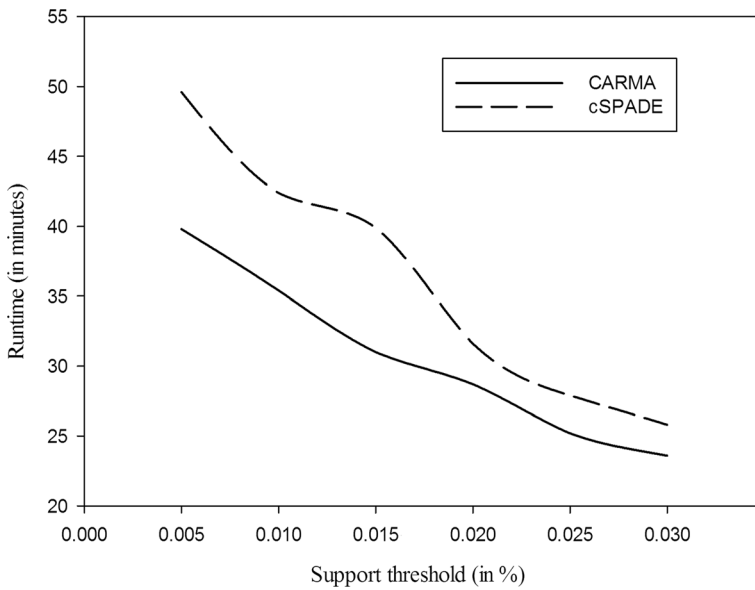
Fig. 7 The performance of CARMA and cSPADE algorithm

confidence level of 40 % and a support rate of 0.02 % as the parameters for the sequential pattern mining.

Table 4 Top 18 land cover change patterns

Pattern	Support (%)	Confidence (%)	Pattern category
44444	49.36	92.80	stable pattern
33333	23.95	97.39	stable pattern
11111	6.59	98.50	stable pattern
55555	6.47	93.82	stable pattern
44455	2.34	48.07	human-induced pattern
44555	1.64	40.93	human-induced pattern
33444	1.16	92.64	human-induced pattern
22555	0.74	97.95	human-induced pattern
44433	0.64	71.70	natural pattern
33555	0.64	50.80	human-induced pattern
33355	0.56	48.09	human-induced pattern
33343	0.25	51.95	human-induced pattern
45555	0.25	43.37	human-induced pattern
33345	0.20	41.79	human-induced pattern
44434	0.16	47.58	human-induced pattern
25555	0.13	44.7	human-induced pattern
22222	0.12	97.67	stable pattern
11555	0.11	62.96	human-induced pattern
Total	95.31 %		

3.6 Result and analysis

3.6.1 Classification result and area statistics

Five classes were considered to represent the different types of land cover for the entire study area. The classification results are shown in Fig. 6 and the statistical results for the classification are listed in Table 3.

According to the records of the Wuhan Planning and Design Institute, since the 1990s, the city has expanded considerably owing to rapid advances of the development zone and district economy. As a result of urban development, the city has
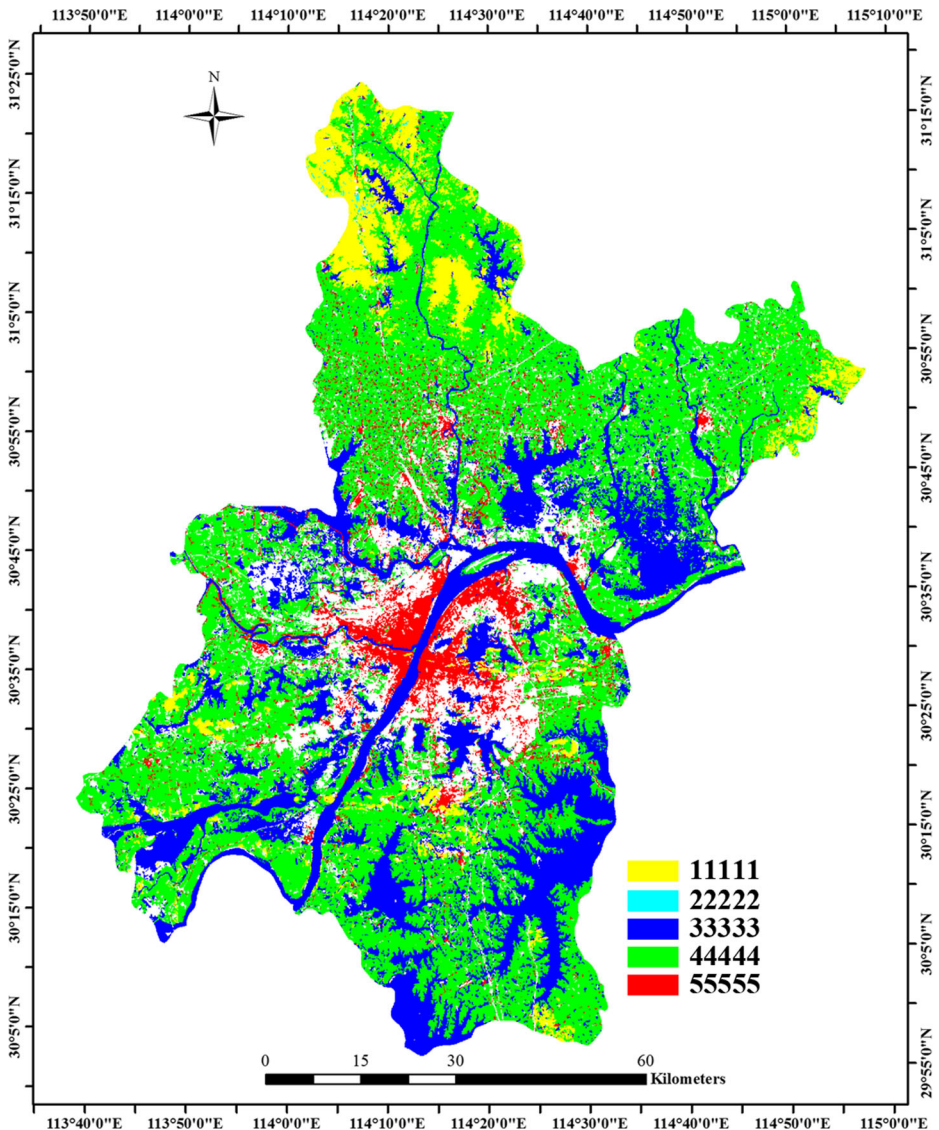

Fig. 8 Spatial distribution of the stable patterns

expanded along the main lines of communication. This change is confirmed by the findings of the present study. According to the area statistics of the land cover types and their changes over the 30-year study period, the built-up area has grown remarkably from 555 km^2 in 1980 to 1497 km^2 in 2010, and the grassland area has decreased by nearly 90 % from 86 km^2 in 1980 to 11 km^2 in 2010. At the same time, other types of land cover have reduced by varying degrees.

3.6.2 Performance of CARMA

To compare the performance of CARMA to traditional algorithms, we implemented CARMA along with the cSPADE algorithm developed by Zaki in 2001, and the recorded time consume for both algorithms are given in Fig. 7. The comparison results (Fig. 7) shows the runtime of CARMA and cSPADE in mining sequential patterns with different support thresholds. According to Fig. 7, we can see that CARMA outperformed cSPADE especially on low support thresholds. That due to cSPADE requires more time for removing all useless sub-sequences that are computed by the algorithm. It also requires, for a given evolution, browsing the whole dataset for identifying pixels that are concerned by this evolution. In addition, CARMA is typically by an order of magnitude more memory efficient and readily computes association rules in cases which are intractable for cSPADE.

3.6.3 Sequential pattern mining result

For the study area, the mining process of CARMA over the land cover change sequence set led to the identification of 113 sequential patterns, each with their own proportion. In order to highlight the most significant land cover change trajectories, the top 18 land cover trajectories whose support rates were more than 0.1 % were selected as the land cover change patterns.

The analysis of land cover change trajectories has focused on the selected top 18 patterns with a total cover equal to 95.31 % of the entire study area (Table 4). The remainder is spread among the other 95 patterns. To analyze the temporal human impact on the environment, we divided the selected top 18 patterns into three categories, namely, stable patterns, human-induced patterns, and natural patterns. Stable patterns are characterized by pixels that are constant over time (e.g., 11111 and 22222). In particular, the trajectories are dominated by five stable patterns accounting for 86.49 % of the study area, where the land cover type never changed between 1980 and 2010. This

Table 5 Composition of land cover change in the human-induced patterns (each land cover change area taken from the entire area)

Land cover change	1980–1990	1990–2000	2000–2005	2005–2010	Total
1–5	0	0.11 %	0	0	0.11 %
2–5	0.13 %	0.74 %	0	0	0.87 %
3–4	0	1.16 %	0.45 %	0.16 %	1.77 %
3–5	0	0.64 %	0.56 %	0	1.2 %
4–3	0	0	0.16 %	0.25 %	0.41 %
4–5	0.25 %	1.64 %	2.34 %	0.2 %	4.43 %
Total	0.38 %	4.29 %	3.51 %	0.61 %	8.79 %

implies that the majority of the land cover types have remained the same since 1980 for the entire time series. The spatial distribution of the stable pattern "11111" is shown by the pixels in the yellow region in Fig. 8. The support of the pattern is 6.59 %, and in 1980, forest land accounted for 6.91 % of the entire area. This implies that more than 95 % of the forest land in 1980 belonged to that the same land cover type in 2010. Similarly, for pattern "55555" (the red region in Fig. 8), we concluded that 98.93 % of the built-up land in 1980 belonged to the same land cover type in 2010. This also indicates that the areas that are not covered by the stable patterns have undergone some kind of land cover change. Thus, the current land area share of the same land cover type can have different histories.
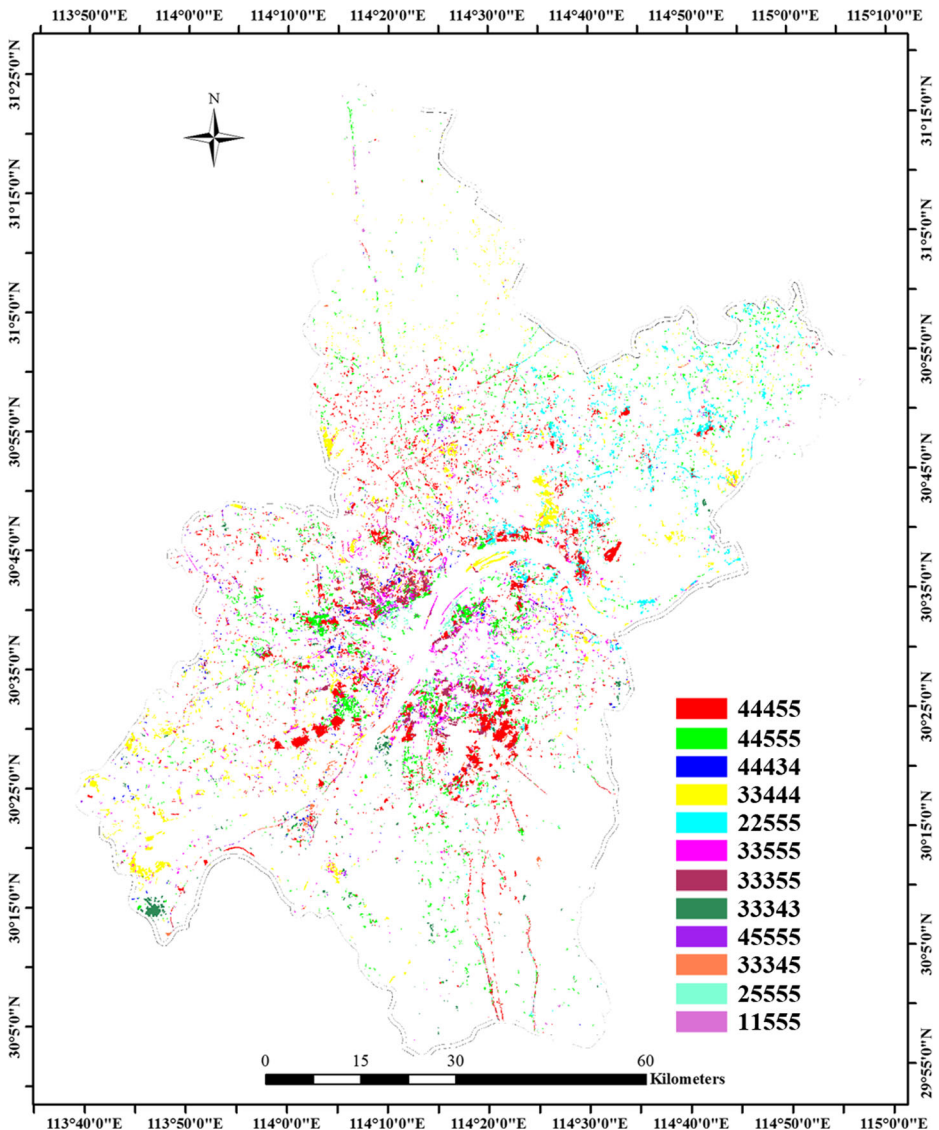

Fig. 9 Spatial distribution of the human-induced patterns

Human-induced patterns are represented by pixel histories with decisive changes due to human activities, such as the change from farm land to built-up land. Table 5 shows the composition of the 12 human-induced patterns accounting for 8.79 % of the study area. Human-induced land cover change trajectories are dominated by the transition from farm land to built-up land, especially during 2000 to 2005. Further, during 1990 to 2010, the highest percentage (4.29 %) of the area was affected by human-induced land cover change.

Figure 9 shows the spatial distribution of human-induced land cover change patterns over the past 30 years. In the study area, the human-induced land cover change is more concentrated in urban areas and becomes gradually sparse toward the suburbs. It also can be noted that some land cover change patterns are continuous in time and space, such as "44555" and "44455", and "33555" and "33355". For example, consider "44555" and "44455"; in 2005, areas of human-induced change from farm land to built-up land most likely appeared in the farm land adjacent to the built-up land that had been converted from farm land in 2000. This implies that urban expansion has spatial connectivity and temporal continuity.

Natural change patterns include indecisive changes due to natural processes. Only one sequential pattern falls into this category, i.e., "44433"; it indicates that the land cover type was farm land before 2005, which changed to wetland after 2005 probably because of floods.

4 Conclusion

The method developed in this study allows the identification, classification, and spatial localization of land cover types and their trajectories of change for a temporal series. It quantifies the land cover changes in terms of the percentage of area affected and maps the spatial distribution of these changes. Further, it provides a different measure for the description of land cover changes 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. Discovering interesting patterns is an important requirement in this field. In future research, we intend to develop interestingness and mining methods that are more sophisticated in order to improve the utility and efficiency of applying sequential pattern mining to remote sensing data.

and Assessment project of the National Ecological Environment Decade of Change (STSN-10-03). The majority of the 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/). Additional data were downloaded from the USGS Landsat archive via the GloVis interface (http://glovis.usgs.gov/).

References

1. Agrawal R, Srikant R (1995) Mining sequential patterns. Proceedings of the Eleventh International Conference on Data Engineering, 6–10 Mar 1995. doi:10.1109/ICDE.1995.380415
2. Antrop M (1998) Landscape change: plan or chaos? Landsc Urban Plan 41(3–4):155–161. doi:10.1016/S0169-2046(98)00068-1
3. Boriah S (2010) Time series change detection: algorithms for land cover change. Dissertations, The University of Minnesota
4. Bovolo F, Bruzzone L (2007) A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. IEEE Trans Geosci Remote Sens 45(1):218–236
5. Coppin P, Jonckheere I, Nackaerts K, Muys B, Lambin E (2004) Digital change detection methods in ecosystem monitoring: a review. Int J Remote Sens 25(9):1565–1596
6. Gueguen L, Datcu M (2007) Image time-series data mining based on the information-bottleneck principle. IEEE Trans Geosci Remote Sens 45(4):827–838
7. Guo H, Wang L, Chen F, Liang D (2014) Scientific big data and digital earth. Chin Sci Bull 59(35):5066–5073. doi:10.1007/s11434-014-0645-3
8. Han J, Pei J, Mortazavi-Asl B, Chen Q, Dayal U, Hsu M-C (2000) FreeSpan: frequent pattern-projected sequential pattern mining. Paper presented at the Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, Boston, Massachusetts, USA
9. Henebry GM, Fountain T, Chomicki J, Ranson KJ (2002) Toward spatio-temporal models of biogeophysical fields for ecological forecasting. Paper presented at the Proceedings of the 2002 annual national conference on Digital government research, Los Angeles, California, USA, 2002
10. Hidber C (1999) Online association rule mining. Paper presented at the Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, New York, NY, USA
11. Jensen JR (1981) Urban change detection mapping using Landsat digital data. Am Cartographer 8(2):127–147
12. Jha CS, Unni N (1994) Digital change detection of forest conversion of a dry tropical Indian forest region. Int J Remote Sens 15(13):2543–2552
13. Jin S, Sader SA (2006) Effects of forest ownership and change on forest harvest rates, types and trends in northern Maine. For Ecol Manag 228(1):177–186
14. Le Ber F, Benoît M, Schott C, Mari JF, Mignolet C (2006) Studying crop sequences with CarrotAge, a HMM-based data mining software. Ecol Model 191(1):170–185
15. Lu D, Mausel P, Brondízio E, Moran E (2004) Change detection techniques. Int J Remote Sens 25(12):2365–2401
16. Ma Y, Wang L, Zomaya AY, Chen D, Rajiv R (2014) Task-tree based large-scale mosaicking for massive remote sensed imageries with dynamic DAG scheduling. IEEE Trans Parallel Distrib Syst 8:2126–2137
17. Ma Y, Wu H, Wang L, Huang B, Ranjan R, Zomaya A, Jie W (2015) Remote sensing big data computing: challenges and opportunities. Futur Gener Comput Syst 51:47–60. doi:10.1016/j.future.2014.10.029
18. Melgani F, Moser G, Serpico SB (2002) Unsupervised change-detection methods for remote-sensing images. Opt Eng 41(12):3288–3297. doi:10.1117/1.1518995
19. Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C (2001) PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. Paper presented at the Proceedings of the 17th International Conference on Data Engineering, Washington, DC, USA
20. Piles M, Entekhabi D, Camps A (2009) A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations. IEEE Trans Geosci Remote Sens 47(12):4125–4131
21. Srikant R, Agrawal R (1996) Mining quantitative association rules in large relational tables. Paper presented at the Proceedings of the 1996 ACM SIGMOD international conference on Management of data, Montreal, Quebec, Canada

22. Wang L, Lu K, Liu P, Ranjan R, Chen L (2014) IK-SVD: dictionary learning for spatial big data via incremental atom update. Comput Sci Eng 16(4):41–52. doi:10.1109/MCSE.2014.52
23. Young S, Wang C (2001) Land-cover change analysis of China using global-scale Pathfinder AVHRR Landcover (PAL) data, 1982–92. Int J Remote Sens 22(8):1457–1477
24. Zaki MJ (2000) Scalable algorithms for association mining. IEEE Trans Knowl Data Eng 12(3): 372–390. doi:10.1109/69.846291
25. Zhang W, Wang L, Ma Y, Liu D (2014) Design and implementation of task scheduling strategies for massive remote sensing data processing across multiple data centers. Software: Pract Experience 44(7):873–886. doi:10.1002/spe.2229

Huichan Liu received her B.Sc. degree from China Agricultural University, Beijing, China, and her M.Sc. degree from China Remote Sensing Satellite Ground Station (RSGS), Chinese Academy of Sciences (CAS), Beijing, China. In 2007, she joined the Center for Earth Observation and Digital Earth, CAS, Beijing, China, as a Research Assistant. At present, she is an Engineer with the Institute of Remote Sensing and Digital Earth (RADI), CAS. Her current research interests include land cover change detection, time-series image processing, and data mining.


Prof. Guojin He was born in Fujian, China, in 1968. He received his B.Sc. degree in geology from Fuzhou University, Fuzhou, China, in 1989, his M.Sc. degree in remote sensing of geology from China University of Geosciences, Wuhan, China, in 1992, and his Ph.D. degree in geology from the Institute of Geology, Chinese Academy of Sciences (CAS), Beijing, China, in 1998. From 1992 to 2007, he was with the Information Processing Department, China Remote Sensing Satellite Ground Station (RSGS), CAS. In 2001, he became the Deputy Director of the Information Processing Department, RSGS, CAS. Since 2004, he has been a Professor and the Director of the Information Processing Department, RSGS, as well as the head of the research group for

Remote Sensing Information Mining and Intelligent Processing. From 2008 to 2012, he was a Professor and the Director of the Value-added Product Department and the Deputy Director of the Spatial Data Center, Center for Earth Observation and Digital Earth, CAS. Since 2013, he has been a Professor and the Director of the Satellite Data Based Value-added Product Department and the Deputy Director of RSGS, Institute of Remote Sensing and Digital Earth, CAS. A major part of his previous research dealt with information processing and applications of satellite remote sensing data. His current research interests include processing optical high-resolution remote sensing images and using information retrieved from satellite remote sensing images in combination with other sources of data to support better understanding of the Earth.

Prof. Weili Jiao received her B.Sc. degree in cartography from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1988, and her M.Sc. degree in geoinformatics from the International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede, The Netherlands, in 1995. At present, she is a Professor with the Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences (CAS), Beijing, China. She has 26 years of experience in the field of remote sensing image processing. Her current research interests include digital photogrammetry, image registration, spatial data uncertainty, and land cover mapping and change detection.
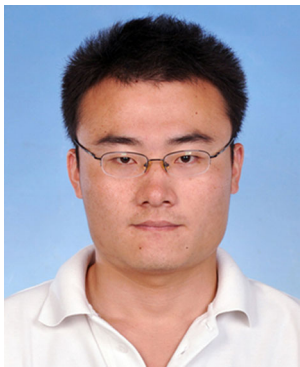
Dr. Guizhou Wang received his B.Sc. degree in surveying and mapping engineering from Jilin University, Changchun, China, in 2008. He received his M.Sc. degree in cartography and geographic information systems and his Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2011 and 2014, respectively. At present, he is a Research Assistant at the Institute of Remote

Sensing and Digital Earth (RADI), Chinese Academy of Sciences (CAS). His research interests include intelligent processing of high-spatial-resolution remote sensing images, such as image segmentation and classification, and spatial information mining from long-time-series Landsat remote sensing images.

Yan Peng was born in Hunan, China, in 1988. She received her B.Sc. degree from the University of South China in 2010, and her M.Sc. degree from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (CAS) in 2013. She joined the Institute of Remote Sensing and Digital Earth, CAS, as a Research Assistant in 2013. Her current research interests include atmospheric correction and remote sensing applications.

Prof. Bo Cheng received his Ph.D. degree in cartography and geographic information engineering from the China University of Geoscience, Beijing, China, in 2003. Since then, he has been working for over 10 years in the fields of photogrammetry and remote sensing at the Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences (CAS). He has been actively conducting research using remote sensing and GIS in support of natural resource mapping and management, urban land use monitoring, and environmental applications.