Evaluating the reliability of time series land cover maps by exploiting the hidden Markov model

Guang Yang¹ · Shenghui Fang² · Wenbing Gong² · Yaolong Zhao¹ · Mengyu Ge²

Accepted: 17 October 2020 / Published online: 24 October 2020
© The Author(s) 2020

Abstract

Time series land cover maps are important materials for the work related to land use and land cover change. Satellite remote sensing images prove advantageous in fast mapping with low cost. In most time series land cover products yielded by the satellite remote sensing images, a number of illogical transitions exist between different time phases. The time series land cover products cannot exactly reflect the real land cover types and land cover changes for each pixel. The accuracy evaluation based on the limited ground truth cannot well guide the users because the reliability of different pixels of the land cover products is unknown. A generic model for the reliability evaluation of time series land cover products should be developed based on a strong theoretical frame. In order to better guide the use of the land cover products, this paper proposed an approach to evaluate the reliability of time series land cover products by exploiting the joint probability of hidden Markov model (HMM), in which the classification performance and the spatio-temporal relationships were taken into account. We applied the proposed evaluation method on the time series land cover maps of Poyang Lake Eco-economic Region in China. The reliability of the land cover products was presented by the grading of the joint probability of HMM. The results effectively reflected the classification performance, the spatio-temporal relationships and even the quality of the data source.

Keywords Land cover maps · Illogical transitions · Reliability · Joint probability · Hidden Markov model · Spatio-temporal relationships

1 Introduction

Time series land cover maps are the basis of change studies at the large scale (Peng et al. 2019; Shimabukuro et al. 2019; Xia et al. 2019). The usability of the land cover products closely relates to their reliability (Corves and Place 1994). Satellite remote sensing images are common for large scale LAND COVER mapping (Homer et al. 2004; Bartholome and Belward 2005). The reliability of the land cover maps produced by time series satellite remote

sensing images is quite low in certain areas owing to the classification strategies and the discrepancies of image qualities (Cai et al. 2014). Correct evaluation of land cover products will guide the application of the products, thus reducing the contrasts between the effect of practical applications and the expectation of the users. In other words, the application risk will be reduced. In this sense, the quality evaluation of land cover maps is really important.

There exist various global land cover products, like the International Geosphere-Biosphere Program Data and Information System's Land Cover Product (IGBP-DIS-Cover) (Loveland et al. 2000), the Global Land Cover database for year 2000 (GLC2000) (Bartholome and Belward 2005), University of Maryland Global Land Cover Classification (UMD GLCC) (Hansen et al. 2000), Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 4 Land Cover Product (Friedl et al. 2002), MODIS Collection 5 Land Cover Product (Friedl et al.

✉ Shenghui Fang
shfang@whu.edu.cn
Guang Yang
gyang@m.scnu.edu.cn

¹ School of Geography, South China Normal University, Guangzhou 510631, Guangdong, China

² School of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079, Hubei, China

2010) and GlobCover (Arino et al. 2008). Most products have low consistencies and continuities (Giri et al. 2005; Oort 2005). Furthermore, the accuracies estimated by the limited ground truths (GT) are generally higher than those verified from the third-party researchers (Chen et al. 2015). For example, a research based on over 2000 GT observations in West Siberia reported extreme low accuracies for IGBP-DISCover and MODIS Collection 4 Land Cover Product (Frey and Smith 2007). Another research found the significant overestimation of cropland cover by the global land cover products in Africa (Fritz et al. 2010). So, the existing time series land cover maps cannot reflect the dynamic land cover changes (LCC) for certain areas or classes.

Most studies on evaluating the reliability of land cover maps focused on calculating the uncertainty of remote sensing images by taking into account the spatial correlations (Foody 2005; Comber et al. 2012; Griffith and Chun 2016; Zhang et al. 2019) and modeling the classification uncertainty based on the classifiers (Loew et al. 2015). Besides, a study gave detailed recommendations for sampling and accuracy assessment of land cover maps (Olofsson et al. 2014). However, the accuracy assessment is usually based on the known samples. In fact, people may not have enough known samples for the areas like the edge between different land cover types. Limited samples do not have sufficient representativeness for large areas (Zhen et al. 2013), and the accuracy assessment based on limited samples should be added more criterions (Foody 2009). Moreover, it is rather difficult to acquire enough ground truth in multiple periods. For the applications based on change studies, most land cover products were produced in time series (Friedl et al. 2010; Yang et al. 2017). A generic evaluation model with strong theoretical frame has not been exploited for the time series circumstances, in which the classification accuracy, the spatio-temporal relationships and the image quality are simultaneously considered.

Statistical learning methods generally establish probabilistic models by data and then forecast data by these models in turn. Labelling/forecast and probability calculation problems are the main applications of these methods (Ionescu and Limnios 1999). There have been studies exploiting the labelling problems to map the time series land cover by remote sensing images (Kasetkasem and Varshney 2002; Wolfe et al. 2015; Abercrombie and Friedl 2016; Gong et al. 2017). The methods in these studies were based on the forecast problem of probabilistic graphical model (PGM) (Jordan et al. 1999). In this problem, the most possible label sequence of the time series land cover was determined by the highest conditional probability or joint probability. However, these methods have not been exploited to evaluate the time series land cover classification results. In fact, by exploiting the probability

calculation problem of PGM, we may use the model to assess the reliability of land cover maps. Although a few studies quantified the uncertainty in land cover maps by the probabilistic methods (Li and Zhang 2011; Cripps et al. 2013), a known sample set were still needed to support the frameworks.

Inspired by the previous studies, this paper introduced an approach to evaluate the reliability of land cover maps by taking into account the spatio-temporal context. In this approach, the rules of land cover transitions and the classification probabilities were put in the hidden Markov model (HMM) (Miller et al. 1999). A spatial reliability indicator was designed based on the idea of local binary pattern (LBP) (Ojala et al. 2002). The spatio-temporal relationships and the classification performance were simultaneously employed to calculate the joint probability of HMM under the premise that the traditional 'hidden' state layer (land cover class sequence) was already known. We further exploited the approach to evaluate the land cover maps for circumstances of both time series and single moment. We applied the model on the evaluation of the time series land cover maps of Poyang Lake Eco-economic Region and further discussed the results.

2 Methods

2.1 Area and data processing

The study area, Poyang Lake Eco-economic Region, is located in the middle and lower reaches of Changjiang River, Middle China. Most of the study area belongs to Jiangxi Province (Fig. 1). The longitude range is between 114° and 117°, and the latitude range is between 27° and 40°. The total area is 51,100 km². The climate consists of temperate ecosystem and subtropical ecosystem. A large part of the territory is covered by croplands, forests and water body, among which the core lake area is quite stable and other land cover areas relatively change frequently.

The study took advantage of multi-source images to produce the time series land cover maps. The initial satellite images comprised Landsat 5 Thematic Mapper (TM) images (Chander et al. 2009) from year 2007 to 2011, Huanjing Charge Coupled Device (CCD) images (Hu and Tang 2012) for year 2012 and Landsat 8 Operational Land Imager (OLI) images (Roy et al. 2014) from year 2013 to 2015. The Landsat path and row of the study area were respectively 120, 121, 122, 123 and 39, 40, 41. The images, all covering the study area, were georeferenced and the resolution was 30 m. The samples were collected by the field survey and human interpretation from Google Earth. We selected 7308 pixels and recorded their ground

Fig. 1 Study area

information in terms of class label, longitude and latitude for each year. The sample numbers of water, cropland, forests, artificial cover and bare land were 2972, 951, 1939, 1085 and 361, respectively. We used 5-fold cross validation for the accuracy calculation (Kohavi 1995). That means the samples were split into 5 mutually exclusive subsets. Each subset was employed as the test set in turn and the others were employed as the training sets. The accuracy was then calculated by the average values.

To make full advantage of the remote sensing images with 30 m resolution, the pixel-wised compositing approach is commonly used in land cover mapping of large areas (Roy et al. 2010). To overcome the influence of the clouds, the study employed the minimum-value compositing to yield the composited images covering the whole study area. For the pixels covered by different images, the minimum values of the spectral bands were selected. After the images of all the years were yielded, the support vector machine (SVM) classifier (Vapnik and Cortes 1995; Ben-Hur et al. 2000) was employed to produce the land cover maps year by year. The study used the Environment for Visualizing Images (ENVI) software to deal with the classification procedure. Simultaneously, the posterior probability of each pixel was acquired and saved by ENVI.

2.2 Probability calculation of hidden Markov model

2.2.1 Generation of hidden Markov model

We will describe the generation of hidden Markov model (HMM) corresponding to time series remote sensing land cover labelling. Suppose $C = \{c_1, c_2, \dots, c_N\}$ is the state set that containing all the possible land cover classes. $O = \{o_1, o_2, \dots, o_M\}$ represents the observation set that containing all the possible spectral values, in which each

element belonging to O is a spectral vector. N is the number of the classes and M is the total number of all the possible spectral vectors. $L = \{l_1, l_2, \dots, l_T\}$ is the label sequence of any pixel for all the T moments, where each element in L belongs to C . $S = \{s_1, s_2, \dots, s_T\}$ is the spectral vector sequence of any pixel for all the T moments, where each element in S belongs to O . A is the transition probability matrix and B is the observation probability matrix

$$A = [a_{ij}]_{N \times N}, \quad (1)$$

$$B = [b_{jk}]_{N \times M}, \quad (2)$$

where

$$a_{ij} = P(l_{t+1} = c_j | l_t = c_i) \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, N \quad (3)$$

represents the conditional probability that a pixel with Class c_i will transfer to Class c_j from Time t to $t + 1$. And

$$b_{jk} = P(s_t = o_k | l_t = c_j) \quad j = 1, 2, \dots, N; \quad k = 1, 2, \dots, M \quad (4)$$

represents the probability that a spectral vector o_k will be generated by a class c_j or a class c_j will be observed by a spectral vector o_k at any moment. In addition, an initial state probability vector should be decided as

$$\pi = (\pi_i) \quad i = 1, 2, \dots, N, \quad (5)$$

where

$$\pi_i = P(l_1 = c_i) \quad i = 1, 2, \dots, N. \quad (6)$$

Consequently, HMM consists of three necessary elements: transition probability matrix, observation probability matrix and initial state probability vector. They are usually expressed by $\lambda = (A, B, \pi)$.

2.2.2 Solutions for the reliability evaluation of land cover products

Traditionally, the probability calculation problem of HMM acquires the probability of the occurrence of the observation sequence S when the model parameters $\lambda = (A, B, \pi)$ and the observation sequence S are fixed:

$$P(S|\lambda) = \sum_L P(S|L)P(L). \quad (7)$$

In this case, the hidden label sequence is unknown, and the solution of the problem is generally the Viterbi algorithm (Li et al. 1999; Gong et al. 2017), which is derived from the dynamic programming (DP) algorithm in the operational Research (Howard 1966). However, unlike the traditional probability calculation problem of HMM, the class label sequence (state sequence) is available in the land cover maps and the traditional joint probability of HMM does not meet the purpose of the evaluation.

Another case is to calculate the probability of the occurrence of the state sequence (class label sequence) L (Eq. 8) when we consider the land cover product evaluation problem:

$$P(L|\lambda) = \pi_1 \prod_{t=1}^T a_{l_t l_{t+1}}, \quad (8)$$

where π_1 is the first element of the initial state probability vector. In this case, we cannot accurately evaluate the reliability just by the state sequence and the predetermined transition probability unless the initial state vector and transition probability matrix of the model are so accurate that we do not have to consider the probability of classification and the spatial rationality of the land cover product. In fact, we cannot estimate the objective transition rules of land cover types by a limited number of samples.

This study transforms the traditional situation into the calculation of the joint probability of the concurrent appearance of the observation sequence S and the state sequence L . That means the ‘hidden’ state sequence is known in this problem, and more importantly, the classification performance and the transition matrix will be simultaneously taken into account. As a result, the solution is the multiplication of the transition probability and observation probability belonging to each moment:

$$P(L, S) = P(S|L)P(L|\lambda) = \pi_1 \prod_{t=1}^T a_{l_t l_{t+1}} b_{l_t s_t} \quad i = 1, 2, \dots, N. \quad (9)$$

We have to mention that HMM cannot work without two primary premises. In simple words, the state in any moment only depends on the state of the previous moment, and is independent of the other moments; the observation at

any moment only depends on the state at the moment, and is independent of the other states (Rabiner 1989).

2.3 Determination of the essential elements

First, in this study, the transition probability matrix represents the transition probability between different land cover types within 1 year. Different land cover types comply with different transformation rules: some areas are relatively stable, like forest cover areas; while others change quite fast, like urban expansion areas (Gómez et al. 2016). We believe it should be close to the physical truths of the study area as much as possible. The physical truths include the location, development planning and the ecosystem etc. The study area has a relatively higher level of land cover change rate in china (Li et al. 2017). According to the logical transitions and the statistical studies about the transition rate of the land cover in the study area (Hui et al. 2008; Feng et al. 2012; Michishita et al. 2012; Zhang et al. 2014; Li et al. 2017), the rates and types of LCC present differences between different classes, and the conclusions of these studies are not consistent. To give a simplified expression and avoid disputes, we refer to the study in Abercrombie and Friedl (2016) and give a 10% probability for annual land cover change. That means, in this model, the transition probability will be set to 90% when a pixel's class labels of the adjacent two years are the same.

Second, each element in the initial state probability vector is set to be the same in this study because it is multiplied only once, instead of many times according to Eq. 9. Consequently, the magnitude of the final calculation result will not change significantly due to an initial value. So, the initial state probability vector is ignored when we evaluate the relative reliability.

Third, the observation probability values are different for different pixel locations. In order to acquire the values, we first calculate the posterior probabilities which represent the probabilities of the correct classification for each pixel. The probabilities can be acquired from various classifiers (Friedl et al. 2002). In this study, we get the posterior probabilities from ENVI classification output. Then the observation probability values are calculated by Bayes formula (Zadeh 1984).

In addition, the spatial continuity also reflects the reliability of the land cover products. Inspired by the ideas of local binary pattern (LBP) (Guo et al. 2010), we had introduced a spatial reliability indicator taking into account the spatial continuity (Gong et al. 2017). Here we give a more detailed description along with necessary legends. Suppose A is the label of a certain pixel P at any moment. A1, A2, A3, A4, A5, A6, A7 and A8 are the class labels of the 8 neighboring pixels of P. Let n be the number of the neighboring pixels of the same labels with P. We initially

define the patterns ‘1’ and ‘0’ for the neighboring pixels. ‘1’ stands for the class label consistent with P and ‘0’ stands for the class label inconsistent with P. Let v be the variation times of the patterns from any directions (clock-wise or anticlockwise). The results with the highest n and the lowest v are the most reliable. Then the indicator I is preliminarily defined by

$$I \propto \frac{n}{v}. \quad (10)$$

The possible patterns and all the values of the indicator are listed by Fig. 2. We give a sample pattern for each possible value of n and v . We set I to 0 when $n = 0$.

Fig. 2 Patterns and values of the spatial reliability indicator. The higher I indicates the higher reliability in space. And I is determined by the local binary pattern

Besides, when $n = 8$ and $v = 0$, I is set to 4 which is higher than all the other values. The value of I should be normalized to fit the model and all the values should be divided by 4. Consequently, the range of I is between 0 and 1. This range is suitable for the probabilistic expression. As a result, the indicator is calculated by

$$I = \begin{cases} \frac{n}{4v}, & v > 0 \\ 0, & v = 0 \text{ and } n = 0 \\ 1, & v = 0 \text{ and } n \neq 0 \end{cases}. \quad (11)$$

The higher value of the indicator means the more reliable spatial relationship. In contrast, the low values indicate the land cover patterns are unreliable. Figure 3 illustrates several common patterns corresponding to the possible land cover distributions. In the central lake area, like the first example, the labels of both the central pixel and the 8 neighboring pixels are water. According to Eq. 11, the indicator value is the maximum of 1. For pixels at the boundary between water and farmland (the second example), the spatial reliability is also quite high. In this circumstance, the value of n and v are 5 and 2, respectively. And the indicator value is $5/8$. In case of more classes surrounding (like the third example), a relatively lower indicator value is more likely to be yielded for the central pixel. According to the equation, the value of n and v are 3 and 2, respectively. And the indicator value is $3/8$. However, since the third example also belongs to a reasonable and objective land cover distribution pattern, the indicator value of the spatial reliability is still higher than many other cases, especially the unreasonable land cover distribution which may exist in any land cover map.

To consider the spatial reliability in the model, the observation probability b_{ls} should be multiplied by I at each moment without destroying the value range of the probability in Eq. 9. The joint probability is then calculated by

$$P(L, S) = P(S|L)P(L) = \pi_1 \prod_{t=1}^T a_{l_t l_{t+1}} b_{l_t s_t} I_t \quad i = 1, 2, \dots, N, \quad (12)$$

where I_t means the spatial indicator of the t th moment.

Fig. 3 Samples of the common patterns corresponding to the possible land cover distributions. The values of the spatial reliability indicators are 1, $5/8$ and $3/8$

2.4 Reliability evaluation for different circumstances

2.4.1 Time series evaluation

The evaluation of the time series land cover products is calculated by the joint probability introduced above (Eq. 12). It reflects the overall reliability of the time series land cover products expressed by the probability values. The elements in the model are all normalized values. That means the probability values of each moment are between 0 and 1. Consequently, the classification performance, temporal relationships and spatial continuities are simultaneously considered in HMM.

However, the values of joint probability are too exhaustive. Small gaps in the values cannot tell the differences of the reliability. This study distinguishes the levels of the reliability of the land cover products by the orders of magnitudes. Values within one order of magnitude are deemed to own the same reliability level. The grading results will be landed in Sect. 3.

2.4.2 Single moment evaluation

The evaluation of the land cover map of a certain moment is calculated by the same model in which only one moment is involved. As a result, Eq. 12 is transformed into

$$P(L_t, S_t) = P(S_t|L_t)P(L_t) = \pi_1 a_{l_t l_{t+1}} b_{l_t s_t} I_t \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T - 1. \quad (13)$$

However, this form only takes into account the temporal relationship between the target moment and next moment. Obviously, the former moment also has direct relationship with the target moment. In this study, the transition probability from the former moment to the target moment is taken as the weight. Then, the evaluation of the single moment is calculated by

$$P(L_t, S_t) = P(S_t|L_t)P(L_t) = \pi_1 a_{l_{t-1} l_t} a_{l_t l_{t+1}} b_{l_t s_t} I_t \quad i = 1, 2, \dots, N; \quad t = 2, \dots, T - 1. \quad (14)$$

Specially, for the first moment, $a_{l_{t-1} l_t}$ is replaced by $a_{l_t l_{t+1}}$; while for the last moment, $a_{l_t l_{t+1}}$ is replaced by $a_{l_{t-1} l_t}$. In this way, the evaluation of the first and the last moment is treated equally with other moments.

2.4.3 Reliability evaluation without probabilities

Sometimes, the probabilities are not easy to be acquired for many classifiers. Besides, the probabilities of all the pixels are not always retained in practice. In this case, the reliability can also be estimated by HMM, in which the

observation probabilities are set to 1. That means the classifier is recognized to give the most credible output. The joint probability is then calculated by

$$P(L, S) = P(S|L)P(L) = \pi_1 \prod_{t=1}^T a_{l_t l_{t+1}} I_t, \quad (15)$$

where the spatial indicator I_t takes place of the observation probability. We believe the classification performance can be indirectly reflected by the spatio-temporal relationships if the probabilities and the accuracies are unavailable. The evaluation results without probabilities will be shown in the next section.

3 Results

3.1 Time series land cover maps

Figure 4 shows the time series land cover maps of the study area. Even the pixel-wised compositing was employed, the strips in the results of certain years cannot be fully removed owing to the quality of the images. However, since this study aims to assess the quality of the products, the classification results are not important. The average accuracies calculated by the ground truth is 90.5%.

3.2 Evaluation results of the reliability

The evaluation results of the time series products are presented in Fig. 5. Figure 5a presents the pixel-level distribution of the reliability level for two cases. In one case, the (posteriori) probability is available and the results are calculated by Eq. 12; in another case, the probability data is not acquired or retained in practice and the results are calculated by Eq. 15. Figure 5b counts the distribution of both cases.

According to the figure, the areas of higher temporal stability, higher spatial continuity and higher classification accuracies are more likely to acquire higher levels. For example, the water body in the center of the lake is not possible to change within several years, and these pixels have higher spatial continuities. Moreover, the separability of water is obviously the highest among all the classes. As a result, these pixels are given the higher levels. In contrast, other areas are not likely to acquire such a high level: the riverbed for the frequently temporal variations, the

Fig. 5 Evaluation results of time series land cover products. **a** Spatial distribution of the reliability level, **b** statistics of the level distributions

farmland for the variations and mixed classification with the forests and water, the mountain forests for the uncertain spatial relationship and the classification performance, the urban marginal areas for the speed of the expansion.

The results of two cases (with probability input and without probability input) are close in both visual perspective and statistics. That means the classifier has reported a quite high posteriori probability for most pixels in the land cover maps. And the temporal relationships play an important role in the proposed evaluation scheme of time series land cover maps. Even without regard to the classification performance, the spatio-temporal relationships will reflect the quality of time series land cover maps.

Figure 6 presents the evaluation results of year 2008 and 2014 corresponding to the land cover maps. Although the reliability evaluation of the single year only considers the spatial relationships of the target year and the temporal transitions of the adjacent years, the differences of the reliability for different areas are clear in the evaluation

Fig. 4 Time series land cover maps for reliability evaluation. The period is from year 2007 to 2015

Fig. 6 Evaluation maps of year 2008 and 2014

maps. For example, the reliability of the central lake area is high, while areas with complex spatial relationships have lower reliability.

Tables 1 and 2 present the evaluation levels (expressed by the brightness of the gray level) corresponding to the orders of the magnitude of the joint probability (for example, if a pixel gets the reliability level like 3, that means the joint probability calculated by Eq. 12 is between 10^{-7} and 10^{-6}). Note that the elements in the transition probability matrix are multiplied by 100 in the calculation procedure in this study to avoid extremely low values. Additionally, we did not list all the levels, especially for those with the extremely low orders of the magnitude. For

time series evaluation, all pixels with the joint probability of lower than 10^{-9} will be given the same level of 0. For single year evaluation, all pixels with the joint probability of lower than 10^{-7} will be given the same level of 0.

4 Discussion

The experimental results lead to the following discussions. Specially, the advantages, theoretical comparisons and limitations of the study will be mentioned.

4.1 Advantages from users' perspective

The proposed method will reflect the influence of the stripes and the shadows owing to the spatial indicator. The undesirable results of the stripes and the shadows can hardly be expressed by the accuracy. In addition, the marginal areas and the areas of frequent variations will be labelled as a lower level of reliability, for the joint probability is more likely to be low for frequently changing pixels. These advantages will benefit the users in practice. Figure 7 shows the typical regions from the classification results and the evaluation output. The frequently-changing areas like the riverbed and the paddy fields are given the lower levels (Fig. 7a). Also, the misclassification owing to the stripes can be discovered in the evaluation maps (Fig. 7b).

Based on the strategy proposed by this paper, people can use the land cover products selectively as needed in practice. For example, users can only choose the areas with no less than a certain level, or give a weight according to the estimated reliability when they use the products. This is important in the management and policy decision (Fritz

Table 1 Orders of magnitude of the joint probability corresponding to the level for time series evaluation

Level	0	1	2	3	4	5	6	7	8	9	10
Orders of magnitude	$(-\infty, -9)$	$[-9, -8)$	$[-8, -7)$	$[-7, -6)$	$[-6, -5)$	$[-5, -4)$	$[-4, -3)$	$[-3, -2)$	$[-2, -1)$	$[-1, 0)$	$[0, 1)$
Level		11	12	13	14	15	16	17	18		
Orders of magnitude		$[1, 2)$	$[2, 3)$	$[3, 4)$	$[4, 5)$	$[5, 6)$	$[6, 7)$	$[7, 8)$	$[8, \infty)$		

Table 2 Orders of magnitude of the joint probability corresponding to the level for single year evaluation

Level	0	1	2	3	4	5	6	7	8	9
Orders of magnitude	$(-\infty, -7)$	$[-7, -6)$	$[-6, -5)$	$[-5, -4)$	$[-4, -3)$	$[-3, -2)$	$[-2, -1)$	$[-1, 0)$	$[0, 1)$	$[1, \infty)$

Fig. 7 Details of the evaluation maps of the time series products. **a** The evaluation of the frequently changed region like the riverbed, **b** the evaluation of the striped area. Left: the land cover maps; right: the evaluation maps

Fig. 8 Distribution maps of with different evaluation levels. Users can select pixels as needed in practice

and See 2008; Tsendbazar et al. 2017). Figure 8 lists the distribution maps under different evaluation levels. With the increase of the standards, the pixel number tends to decrease (white areas). For example, there is almost only the central lake area left in the map when the target level is higher than 14.

4.2 Comparisons in theory

HMM and Maximum A Posteriori-Markov Random Field (MAP-MRF) have both been employed to time series land cover mapping in previous studies (Cai et al. 2014; Abercrombie and Friedl 2016). However, the models have to modify a large number of pixels in the post-classification process. To some extent, if the parameters are not accurate, the models will make the classification results more reasonable, rather than objective and accurate. Moreover, the parameters of the spatio-temporal weights are difficult to be decided. Thereby, we believe the models are more

suitable in evaluation than in mapping. Specially, compared to undirected graph models like MRF (Li 2001), the directed graph models [Bayesian Network (Pearl 1988)] like HMM are more suitable, for it can give the joint probabilities of all the moments with much less time cost. According to Gibbs distributions, the joint probability of MRF can be expressed by the energy function (Hazan et al. 2013). However, it can just act on the maximum clique, instead of the whole chains.

The previous studies generally considered the spatial relationships in reliability evaluation (Comber et al. 2012; Zhang et al. 2019), but it is rare to consider both temporal and spatial contexts. This study integrated the temporal and spatial relationships into a unified model for the reliability evaluation of time series land cover maps. On one hand, by considering the characteristics of remotely-sensed land cover maps, the study exploited the joint probability calculation problems of HMM and applied a generic model to the problem. Because the probability calculation problem

of the model was solved by continuous multiplication, instead of the Viterbi algorithm (Li et al. 1999), it was easy to reproduce. On the other hand, the proposed strategy took into account the spatial and temporal relationships simultaneously without violating the premise of the basic elements and requirements of HMM.

4.3 Limitations

In this study, we refer to the previous study to decide the transition probability matrix. A more accurate setting may be different according to the ecological law and the development characteristics of the study area. However, we believe the transition probability matrix reflects the trends of the land cover transitions. Similarly, the spatial indicator also reflects the rationality of the spatial relationships of the land cover types.

The single moment evaluation only takes into account the neighboring years, and the influence of the other years is ignored. In this sense, the undirected graph model may also be the applicable model because the time cost of the calculation of the joint probability will be relatively low (Cai et al. 2014).

The evaluation scheme only gives the relative reliability levels calculated by the joint probabilities of the regional land cover maps, instead of the absolute reliability. In other study areas, the order of magnitudes may vary with multiple factors, like the number of moments, the overall classification performance and the regional ecological laws. A criterion should be unified in the future. A possible solution is to select the water body in the center of the lake/river/ocean as the highest level, because these pixels have a higher identification accuracy in most land cover products. Although this paper provided the evaluation results under different model parameters and application modes, the verification of the evaluation effect is still limited. However, different from the verification of classification accuracy, it is not sufficient to test the whole time series land cover products by limited samples.

5 Conclusion

The paper contributed to the previous studies on time series land cover mapping, post-classification and statistical learning methods. We introduced a strategy to evaluate the time series land cover products. The spatio-temporal relationships and the rules of local environment were simultaneously taken into account. The evaluation framework did not rely on the ground truth information. Based on the joint probability of HMM, we exploited the models for the time series evaluation, single moment evaluation and the reliability evaluation without probabilities. A nine-year

time series land cover maps were employed to apply the proposed method. The evaluation results of the graded joint probability reflected the reliability pixel by pixel. The results demonstrated the importance of the spatio-temporal relationships in the evaluation of time series land cover classification. Meanwhile, the results present the influences of the stripes, shadows and marginal areas in the land cover products. The users of the land cover products will be better guided by the proposed method. Though the paper described a framework to evaluate the time series land cover products by the probability values, some details should be hashed out. For example, the standard of the level may be stipulated according to the length of time series.

Nowadays, different versions of global time series land cover products are in service, and the reliability of these products presents differences in different areas. By using the proposed method, we can equally and objectively evaluate the reliability of these products for any area. It will be contributory to design a unified scheme to evaluate the reliability of the existing global land cover products.

Authors contribution GY is the main author who proposed the basic idea and completed the experiments. SF and WG provided the useful suggestions on designing the approaches involved in our proposed strategy. YZ and MG helped to complete the manuscript.

Funding The study was funded by Certificate of China Postdoctoral Science Foundation Grant (No. 2019M662949); the National Natural Science Foundation of China (No. 41871292); the Science and Technology Program of Guangdong Province, China (No. 2018B020207002); the Science and Technology Program of Guangzhou, China (Nos. 201803030034, 201802030008).

Availability of data and materials The experiments of the study were based on the Landsat images. The data processing was done on <https://earthengine.google.com>. The images are available on <https://earthexplorer.usgs.gov> and <https://earthengine.google.com>.

Code availability The custom code is not available online now.

Compliance with ethical standards

Conflict of interest The authors declared no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abercrombie SP, Friedl MA (2016) Improving the consistency of multitemporal land cover maps using a hidden Markov model. *IEEE Trans Geosci Remote Sens* 54(2):703–713
- Arino O et al (2008) GLOBCOVER: the most detailed portrait of Earth. *Esa Bull Bull Ase Eur Space Agency* 2008(136):24–31
- Bartholome E, Belward AS (2005) GLC2000: a new approach to global land cover mapping from earth observation data. *Int J Remote Sens* 26(9):1959–1977
- Ben-Hur A et al (2000) A support vector clustering method. In: *Proceedings of the international conference on pattern recognition 2000*
- Cai S et al (2014) Enhancing MODIS land cover product with a spatial-temporal modeling algorithm. *Remote Sens Environ* 147:243–255
- Chander G et al (2009) Summary of current radiometric calibration coefficients for landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens Environ* 113(5):893–903
- Chen J et al (2015) Global land cover mapping at 30 m resolution: a POK-based operational approach. *Isprs J Photogramm Remote Sens* 103:7–27
- Comber A et al (2012) Spatial analysis of remote sensing image classification accuracy. *Remote Sens Environ* 127:237–246
- Corves C, Place CJ (1994) Mapping the reliability of satellite-derived landcover maps—an example from The Central Brazilian Amazon Basin. *Int J Remote Sens* 15(6):1283–1294
- Cripps E et al (2013) Quantifying uncertainty in remotely sensed land cover maps. *Stoch Env Res Risk Assess* 27(5):1239–1251
- Feng L et al (2012) Assessment of inundation changes of Poyang lake using MODIS observations between 2000 and 2010. *Remote Sens Environ* 121(2):80–92
- Foody GM (2005) Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *Int J Remote Sens* 26(6):1217–1228
- Foody GM (2009) Sample size determination for image classification accuracy assessment and comparison. *Int J Remote Sens* 30(20):5273–5291
- Frey KE, Smith LC (2007) How well do we know northern land cover? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia. *Glob Biogeochem Cycles* 21(1):1435–1440
- Friedl MA et al (2002) Global land cover mapping from MODIS: algorithms and early results. *Remote Sens Environ* 83(1–2):287–302
- Friedl MA et al (2010) MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sens Environ* 114(1):168–182
- Fritz S, See L (2008) Identifying and quantifying uncertainty and spatial disagreement in the comparison of global land cover for different applications. *Glob Change Biol* 14(5):1057–1075
- Fritz S et al (2010) Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *Int J Remote Sens* 31(9):2237–2256
- Giri C et al (2005) A comparative analysis of the global land cover 2000 and MODIS land cover data sets. *Remote Sens Environ* 94(1):123–132
- Gómez C et al (2016) Optical remotely sensed time series data for land cover classification: A review. *Isprs J Photogramm Remote Sens* 116:55–72
- Gong W et al (2017) Using a hidden Markov model for improving the spatial-temporal consistency of time series land cover classification. *Isprs Int J Geo-Inf* 6(10):292
- Griffith DA, Chun Y (2016) Spatial autocorrelation and uncertainty associated with remotely-sensed data. *Remote Sens* 8(7):535
- Guo Z et al (2010) A completed modeling of local binary pattern operator for texture classification. *IEEE Trans Image Process Publ IEEE Signal Process Soc* 19(6):1657
- Hansen MC et al (2000) Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int J Remote Sens* 21(6–7):1331–1364
- Hazan T et al (2013) On sampling from the gibbs distribution with random maximum a-posteriori perturbations. *Adv Neural Inf Process Syst* 26:1268–1276
- Homer C et al (2004) Development of a 2001 national land-cover database for the United States. *Photogramm Eng Remote Sens* 70(7):829–840
- Howard RA (1966) Dynamic programming. *Manag Sci* 12(5):317–348
- Hu C, Tang P (2012) Automatic algorithm for relative radiometric normalization of data obtained from Landsat TM and HJ-1A/B charge-coupled device sensors. *J Appl Remote Sens* 6:063509
- Hui F et al (2008) Modelling spatial-temporal change of Poyang Lake using multitemporal Landsat imagery. *Int J Remote Sens* 29(20):5767–5784
- Ionescu DC, Limnios N (1999) *Statistical and probabilistic models in reliability*. Birkhäuser, Boston
- Jordan MI et al (1999) An introduction to variational methods for graphical models. *Mach Learn* 37(2):183–233
- Kasetkasem T, Varshney PK (2002) An image change detection algorithm based on Markov random field models. *Geosci Remote Sens IEEE Trans* 40(8):1815–1823
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International joint conference on artificial intelligence*
- Li SZ (2001) *Markov random field modeling in image analysis*. Springer, Berlin, pp 344–357
- Li W, Zhang C (2011) A Markov Chain geostatistical framework for land-cover classification with uncertainty assessment based on expert-interpreted pixels from remotely sensed imagery. *IEEE Trans Geosci Remote Sens* 49(8):2983–2992
- Li J et al (1999) Image classification by a two dimensional hidden Markov model. In: *International conference on 1999 IEEE. Acoustics, speech, and signal processing, 1999*
- Li H et al (2017) Using land long-term data records to map land cover changes in China over 1981–2010. *IEEE J Sel Topics Appl Earth Observ Remote Sens* 10(4):1372–1389
- Loew F et al (2015) Analysis of uncertainty in multi-temporal object-based classification. *Isprs J Photogramm Remote Sens* 105:91–106
- Loveland TR et al (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int J Remote Sens* 21(6–7):1303–1330
- Michishita R et al (2012) Monitoring two decades of urbanization in the Poyang Lake area, China through spectral unmixing. *Remote Sens Environ* 117(1):3–18
- Miller DRH et al (1999) A hidden Markov model information retrieval system. In: *International Acm sigir conference on research and development in information retrieval*
- Ojala T et al (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
- Olofsson P et al (2014) Good practices for estimating area and assessing accuracy of land change. *Remote Sens Environ* 148:42–57

- Oort PAJv (2005) Improving land cover change estimates by accounting for classification errors. *Int J Remote Sens* 26(14):3009–3024
- Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. *Comput Sci Artif Intell* 70(2):1022–1027
- Peng F et al (2019) Content-based search of earth observation data archives using open-access multitemporal land cover and terrain products. *Int J Appl Earth Obs Geoinf* 81:13–26
- Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77(2):257–286
- Roy DP et al (2010) Web-enabled landsat data (WELD): landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens Environ* 114(1):35–49
- Roy DP et al (2014) Landsat-8: science and product vision for terrestrial global change research. *Remote Sens Environ* 145:154–172
- Shimabukuro YE et al (2019) Monitoring deforestation and forest degradation using multi-temporal fraction images derived from Landsat sensor data in the Brazilian Amazon. *Int J Remote Sens* 40(14):5475–5496
- Tsendbazar N-E et al (2017) Integrating global land cover datasets for deriving user-specific maps. *Int J Digital Earth* 10(3):219–237
- Vapnik V, Cortes C (1995) Support vector networks. *Mach Learn* 20(3):273–297
- Wolfe JM et al (2015) Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution. *Isprs J Photogramm Remote Sens* 103(4–8):38–47
- Xia CY et al (2019) Analyzing spatial patterns of urban carbon metabolism and its response to change of urban size: a case of the Yangtze River Delta, China. *Ecol Ind* 104:615–625
- Yang Y et al (2017) Accuracy assessment of seven global land cover datasets over China. *Isprs J Photogramm Remote Sens* 125:156–173
- Zadeh LA (1984) Review of a mathematical theory of evidence. *AI Mag* 5(3):235–247
- Zhen Z et al (2013) Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *Int J Remote Sens* 34(19):6914–6930
- Zhang X et al (2019) Uncertainty assessment in multitemporal land use/cover mapping with classification system semantic heterogeneity. *Remote Sens* 11(21):2509
- Zhang ZB et al (2014) Studying changes in land use within the Poyang Lake region. *J Indian Soc Remote Sens* 42(3):633–643

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.