ISPRS Journal of Photogrammetry and Remote Sensing 144 (2018) 379–389

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A practical sampling method for assessing accuracy of detected land cover/land use change: Theoretical analysis and simulation experiments

Yang Li[a], Jin Chen[a], Yuhan Rao[a,b,*]

[a] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[b] Department of Geographical Sciences, University of Maryland, College Park, MD, 20742, United States

ARTICLE INFO

Keywords:
Land use land cover change
Accuracy assessment
Transition accuracy
Sample design

ABSTRACT

Accuracy assessment plays a crucial role in the implementation of change detection, which is commonly used to track land surface changes and ecosystem dynamics. There are currently two major indicators for accuracy assessment of change detection: the binary change accuracy (*ca*) and the overall transition accuracy (*ta*). The overall transition accuracy has been recommended over change accuracy, because the binary change accuracy does not consider the accuracy of the types of changes of the underlying land cover classes. However, the application of overall transition accuracy has been limited by the challenge of collecting enough representative samples with a practical sampling strategy to meet the users' requirement of precision. This study provides an iterative sampling framework to ensure that the precision of the estimated overall transition accuracy meets the users' predefined requirement. We use a set of simulated change maps to comprehensively examine the effectiveness and robustness of the proposed sampling strategy. The simulation-based results demonstrate that the proposed framework can achieve satisfactory performance for transition accuracy assessment and it is robust against different properties of classification results and target landscapes, including the degree of fragmentation, proportions of land cover types, and temporal correlation of the classification error between individual dates. The effectiveness, robustness and practicality of the proposed sampling strategy will enable producers and users of land cover/land use change maps to get reliable and meaningful accuracy assessment for further applications.

1. Introduction

Change detection is a commonly used technique in the remote sensing community to track land surface changes and ecosystem dynamics, such as urbanization, forest dynamics, desertification and wildfire (Boyd et al., 2018; Giglio et al., 2003; Hansen et al., 2013; Jeon et al., 2014; Mas, 1999; Zhu, 2017). When remotely sensed data acquired from two different years are available, there are mainly two types of methods that have been applied for change detection. The first type of method compares spectral properties of two images acquired in different years, such as change vector analysis (Chen et al., 2011, 2003; Lambin and Strahlers, 1994; Lu et al., 2016; Ye et al., 2016) and cross-correlogram spectral matching (Wang et al., 2009). The second type compares two independently classified land cover/land use maps of different years, such as post-classification comparison methods (Lu et al., 2005; Mas, 1999). In addition to these two methods, because of the increasing availability of long-term remotely sensed data (e.g., AVHRR, MODIS, SPOT), especially the open archive of Landsat after 2008 (Woodcock et al., 2008), time-series based methods have become increasingly popular among the scientific community (Broich et al., 2011; DeVries et al., 2015; Hansen et al., 2013; Zhu, 2017; Zhu and Woodcock, 2014). Time-series based methods usually identify changes by detecting abnormal behaviors in the time series of spectral reflectance and/or derived indices (e.g., NDVI, spatial context). In addition to the occurrence of change, time-series based methods can also identify when the detected changes occurred, which is a critical aspect for near-real time change is monitoring (e.g., forest disturbance and natural disaster monitoring). Even though time-series based methods have their advantages, they usually focus on specific change types (e.g., deforestation, urbanization). Due to the easiness of implementation and interpretation, post-classification comparison is still one of the most popular methods for multi-class change analysis (Lu et al., 2005).

To use detected changes in other scientific studies, management, or policy supporting activities, it is necessary to provide the estimated accuracy and changed area with the associated bias and precision based on a set of samples and probability inference, otherwise the results would be just "pretty pictures" (McRoberts, 2011). Ideally, the estimation should be both independent from change detection techniques

* Corresponding author at: Department of Geographical Sciences, University of Maryland, College Park, MD, 20742, United States.
E-mail address: yrao@terpmail.umd.edu (Y. Rao).

https://doi.org/10.1016/j.isprsjprs.2018.08.006
Received 14 June 2016; Received in revised form 29 June 2018; Accepted 3 August 2018

and robust against data noise as well as other factors (e.g., analyst biases, land cover transition types, and imperfect reference data) (Olofsson et al., 2014). Accordingly, guidance on how to assess accuracy and to estimate areas of change of produced maps in a consistent and statistically sound manner is necessary for strengthening the confidence of users and to help them understand uncertainties before further applications (Foody, 2002; Olofsson et al., 2013; Yang et al., 2017). Recently, accurately estimating areas of change has become increasingly important in this field because the areas of change are required for many further analyses, such as estimating carbon emissions/sequestration and economic loses/gains due to land cover changes. Various area estimators of change analysis that use information from the error matrix have been proposed (Czaplewski and Catts, 1992; McRoberts, 2011, 2006; Stehman, 2013). Although the most straight forward method is counting changed pixels from the map, this estimation may suffer from biases caused by classification and change detection errors. Therefore, area estimators with adjustments for classification and change detection errors have been recommended in practice (McRoberts, 2011; Olofsson et al., 2014, 2013; Stehman, 2013).

The most common and informative way to assess the accuracy of a map is to report the error matrix by simply cross tabulating the reference labels and the map classifications (Foody, 2002). For change detection, there are currently two different practices: the binary change/no-change error matrix and the transition error matrix (van Oort, 2007). The commonly used change/no-change error matrix compares maps and references for two simplified situations, change and no-change. On the other hand, the transition error matrix reports the agreement between map- and reference-labels for all possible change scenarios (from/to each land cover types). The change/no-change method has been widely adopted because it is easy to report and to compare across different studies. However, it has also been widely criticized because it does not consider mislabeled change/no-change resulting from the classification error or change detection methods. This can be more serious for studies using post-classification comparison methods. The mislabeled change/no-change units are defined as units that have been labeled as change/no-change in both the map and the reference due to classification errors. An example of this error could be assuming a unit has been converted from forest to cropland, when instead this unit has been classified as shrubland and grassland at two dates respectively. When constructing the change/no-change error matrix, this unit will be labeled as correctly detected change. However, this is only a chance agreement because of the classification errors occurred at both dates. By reporting land cover types for both dates in the transition error matrix, it barely suffers the mislabeled change/no-change. Therefore, the transition error matrix has been recommended to assess the accuracy of change detection in practice (van Oort, 2007). It should be noted that the dimension of the full transition error matrix will be the square of the number of the land cover type involved, which is much larger than the binary change/no-change error matrix (Foody, 2002).

Accompanying the error matrix, studies usually summarize it with one or several indicators (Foody, 2002; Stehman, 2009). One of the most commonly used indicators is the overall accuracy defined as the proportion of units that are correctly matched between maps and references. For change detection studies, the change detection accuracy derived from the binary change/no-change error matrix has been widely reported since it is easy to implement. However, it suffers from the defect inherited from the change/no-change error matrix (Burnicki, 2011; van Oort, 2007). Therefore, scholars suggest that the transition accuracy derived from the transition error matrix should be reported instead since it can notably eliminate influences of classification errors to represent the real accuracy of detected changes (Olofsson et al., 2014; van Oort, 2007). Unfortunately, the challenge of collecting sufficient representative samples to calculate the overall transition accuracy limits its applications. Consequently, there are few attempts to report overall transition accuracy in recent land cover/land use change detection studies (van Oort, 2007).

In the process of accuracy assessment, there are two essential steps before final calculation of the accuracy indicator: sampling design and response design (Olofsson et al., 2014). Sampling design determines the basic principles for selecting subsets of spatial units (e.g. pixels or objects) from the detected change map instead of unfeasibly conducting the accuracy assessment for the whole change map. After sampling subsets have been selected, response design provides essential rules for determining agreements between map labels and corresponding reference labels for a sample of units (e.g., pixels or objects). As a basic step for accuracy assessment, sampling design is usually required to be practical with reasonable cost and statistical accuracy. Unfortunately, the sampling design of change detection accuracy assessment has not attracted adequate attention in the remote sensing community, and only a few publications have noticed the importance of practical sampling design (Almutairi and Warner, 2010; Foody, 2010; McRoberts et al., 2018). Olofsson et al. (2014) provide a thorough overview of area estimation and accuracy assessment for land cover change analysis. In the review, they provide general suggestions regarding the sample design for accuracy assessment of land cover change analysis, including sample size estimation and sample allocation. Stehman (2012) compares four different sample allocation strategies when stratified sampling is applied to land cover change studies, which suggests there is no simple best sample allocation for all accuracy indicators and area estimations (Stehman, 2012).

The objective of this study is to provide a general practical sampling framework for accuracy assessment of the detected change results. The practical sample design presented in this article uses probability sampling theory to ensure that estimated accuracy is unbiased and meets users' requirements of precision. We use the indicator of overall transition accuracy as an example to present our framework. However, the direct product of this framework is the error matrix resulted from the sampling strategy. This error matrix can be further used for probability-based inference of change area and other indicators of users' interests. In order to test the effectiveness of the proposed sampling strategy, a series of classification and reference images reflecting different landscapes are simulated by SIMMAP 2.0 (Saura and Mart, 2000) rather than collecting actual classification and reference (also called as ground truth) maps. We use simulation data here because (1) in this way reliable reference data can be acquired without labor-intensive fieldwork or visual interpretation, and (2) the classification error can be controlled to the desired range for theoretical analysis. The reset of this article is presented as follow: Section 2 describes the theoretical basis and implementation processes of the general sampling strategy guided by users' requirement of precision; Section 3 presents the experiments using simulated data to test the effectiveness and robustness of the sampling framework; Section 4 discusses the implications of this sampling framework for user-oriented area estimation, and provides suggestions for accuracy assessment for change detection studies. In Section 2, we only provide the minimum amount of mathematical derivation based on sampling theory and recommend readers who are interested in detail refer to Appendix A and corresponding references.

2. Method

2.1. Definition of change accuracy and overall transition accuracy

As described before, change accuracy and overall transition accuracy are two major accuracy indicators for change analysis. The change accuracy (i.e. *ca*), derived from binary change/no-change error matrix (Table 1), is defined as follow,

$$ca = x_{11} + x_{22} \tag{1}$$

where x_{11} and x_{22} are the percentage of pixels whose change status are consistent with the reference change labels (i.e. percentage of correct change and no change) respectively.