Remote Sensing of Environment 199 (2017) 401–414

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Predicting individual pixel error in remote sensing soft classification

Reza Khatami[a], Giorgos Mountrakis[a,*], Stephen V. Stehman[b]

[a] Department of Environmental Resources Engineering, State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, United States

[b] Department of Forest and Natural Resources Management, State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, United States

ARTICLE INFO

Keywords:
Sub-pixel land-cover mapping
Classification accuracy assessment
Spectral unmixing
Error map
Local accuracy
Image classification

ABSTRACT

Accuracy assessment of remote sensing soft (sub-pixel) classifications is a challenging topic. Previous efforts have focused on constructing a soft classification error matrix and producing summary measures to describe overall and per-class map accuracy. However, these summary assessments do not provide information on the spatial distribution of the soft classification error as distributed at the individual pixel level. This is important because the map error of a given class may vary considerably over different regions. Spatial interpolation has been previously used for predicting soft classification error at the pixel level. Here, we propose two alternative domains for soft classification error interpolation, the spectral and mapped class proportion domains. In the spectral domain we interpolate errors in the classification feature space, whereas in the mapped class proportion domain interpolation takes place in a space with dimensions defined by the mapped class proportions (i.e., the output of the soft classification). The two newly proposed prediction methods (spectral domain and mapped class proportion domain), spatial interpolation, and a summary measure method were evaluated using 23 test regions, each 10 km × 10 km, distributed throughout the United States. These 10 km × 10 km blocks had complete coverage reference data (where the reference classification was determined by manual interpretation) and the predicted error maps were then evaluated by comparing them to these complete coverage reference error maps. Mean absolute error was used to quantify the agreement of the predicted error maps to the reference error maps. The spectral and mapped class proportion methods generally outperformed the spatial interpolation and the summary measure methods both in terms of smaller mean absolute error and visual similarity of predicted error maps to the reference error maps. The superiority of the new methods over spatial interpolation is an important result because spatial interpolation is a familiar method analysts would commonly consider for modeling spatial variation of classification error. The predicted soft classification error maps provide a straightforward visual assessment of the spatial patterns of error that can accompany the original classification products to enhance their value in subsequent analysis and modeling tasks. Furthermore, from the standpoint of implementation, our methods do not require additional datasets; the same test dataset currently used for confusion/error matrix construction can be used for our error interpolation methods.

1. Introduction

Classified land-cover maps have become one of the most important products of remote sensing science and industry enabling environmental and natural resources monitoring, modeling and management from local to global spatial extents. Land-cover maps are essential inputs for a broad range of applications such as forest and carbon monitoring (Carreiras et al., 2012; Dong et al., 2003; Eva et al., 2012; Réjou-Méchain et al., 2014); environmental change detection (Roy et al., 2014; Wulder et al., 2008); climate studies (Grimm et al., 2008; Seneviratne et al., 2010); and hydrological modeling (Khan et al., 2011; Nie et al., 2011; Sorooshian et al., 2014). Significant work has been done by the remote sensing community to improve the associated classification processes and to increase the accuracy of classified land-cover maps (Cihlar, 2000; Franklin & Wulder, 2002; Gómez et al., 2016; Khatami et al., 2016; Lu & Weng, 2007). Land-cover classification can be generally divided into two major categories, hard and soft classifications. In hard classifications, each pixel is assigned to a single class, whereas in soft classification, a pixel may belong to multiple classes and different levels of class membership or proportion are assigned. Soft classifications can potentially be very useful when a large number of mixed pixels exists in an image (Foody & Doan, 2007; Paneque-Gálvez

* Corresponding author.
E-mail addresses: sgkhatam@syr.edu (R. Khatami), gmountrakis@esf.edu (G. Mountrakis), svstehma@syr.edu (S.V. Stehman).

http://dx.doi.org/10.1016/j.rse.2017.07.028
Received 23 January 2017; Received in revised form 18 July 2017; Accepted 25 July 2017
Available online 04 August 2017

et al., 2013; Tsutsumida et al., 2016), as for example when the scene is heterogeneous and the pixel size is larger than the size of the objects of interest.

Whether a hard or soft classification is implemented, it is important to quantify classification error. The accuracy assessment of a hard classification is typically reported through the error or confusion matrix and summary measures derived from it that describe the accuracy of the entire map or a class (Foody, 2002; Olofsson et al., 2014; Stehman & Czaplewski, 1998; Story & Congalton, 1986). In addition, per-pixel classification accuracy prediction methods for hard classification have been investigated to produce maps depicting the spatial distribution of classification accuracy (Comber et al., 2012; Comber, 2013; Foody, 2005; Khatami et al., 2017; Kyriakidis & Dungan, 2001; Steele et al., 1998; Tsutsumida & Comber, 2015) or classification confidence (Mountrakis & Xi, 2013). Khatami et al. (2017) also investigated factors affecting per-pixel accuracy interpolation of hard classifications.

Accuracy assessment of soft classifications is more challenging because the concept of the error matrices typically used in hard classifications cannot be directly applied for soft classifications. Efforts to construct error matrices for soft classification analogous to those applicable to a hard classification include fuzzy error matrix (Binaghi et al., 1999; Stehman et al., 2007) and soft classification error matrix (Latifovic & Olthof, 2004; Pontius Jr. & Cheuk, 2006). Generally, the objective is to build an error matrix for each test pixel based on the reference class proportions and mapped (from soft classification) class proportions. Error matrices for all test pixels can then be aggregated to produce a single estimated error matrix for the entire map. Summary measures such as overall, user's, and producer's accuracies can be estimated from the aggregated error matrix. However, because the spatial distribution of the reference and mapped class proportions within each test pixel is unknown, it is not possible to exactly determine the true overlap among reference and mapped classes and obtain the true error matrix for each test pixel. This issue is known as "sub-pixel area allocation problem" (Silván-Cárdenas & Wang, 2008). Many approaches or operators have been devised to allocate the overlap among reference and mapped class proportions to construct the error matrix of a given test pixel. Some of these methods include fuzzy minimum operator (Binaghi et al., 1999); composite operator (Pontius Jr. & Cheuk, 2006); product operator (Lewis & Brown, 2001); similarity index (Townsend, 2000); and confusion intervals (Silván-Cárdenas & Wang, 2008). Silván-Cárdenas & Wang (2008) introduced a series of characteristics for an ideal per-pixel confusion matrix and discussed whether different operators could result in error matrices that hold those characteristics.

Another group of methods employed for accuracy assessment of soft classifications is based on directly measuring the proximity, similarity, or correlation among the reference and mapped class proportions. Commonly, an accuracy summary measure is calculated using test data to describe how close the reference and mapped class proportion values are for the entire classification or for a given class. Some of these measures include Euclidian and city block distance (Foody, 1996; Foody & Arora, 1996); root mean squared error (RMSE) (Carpenter et al., 1999; Chen et al., 2010; Lu & Weng, 2006; Olthof & Fraser, 2007); correlation coefficient (Foody & Cox, 1994; Maselli et al., 1996); entropy (Finn, 1993; Maselli et al., 1994); cross-entropy (Foody, 1995); information closeness (Foody, 1996); weighted disagreement (Gómez et al., 2008); kappa coefficient (Homer et al., 2012); and Morisita's index (Ricotta, 2004). Similarity can also be assessed in the context of fuzzy logic (Foody, 1999; Gopal & Woodcock, 1994; Laba et al., 2002; Woodcock & Gopal, 2000).

The summary measures derived from the two aforementioned general categories of soft classification accuracy assessment are useful to describe the classification accuracy at the general map scale. However, the summary measures do not provide specific information about the spatial distribution of the classification error. This is an important issue because the classification accuracy would likely vary over different regions of the map (Campbell, 1981; Chen & Wei, 2009; Congalton, 1988) and the summary measures may not be useful when the local accuracy for an area of interest differs from the global accuracy (Mcgwire & Fisher, 2001).

Classification errors affect the reliability of subsequent map use for environmental analyses and modeling (Castilla & Hay, 2007; Ge et al., 2007; Jin et al., 2014; McMahon, 2007; Straatsma et al., 2013). Because classification accuracy varies over different map regions, subsequent models would also inherit this spatial accuracy variation. Thus, environmental modeling can be greatly enhanced if local estimates of classification accuracy or error are available (DeFries & Los, 1999; Gahegan & Ehlers, 2000; Miller et al., 2007). Consequently, in this research we focus on pixel-level error map construction for land-cover maps created by soft classification of remotely sensed imagery. Spatial interpolations have been previously suggested to create error maps for soft classifications (Comber, 2013; Foody, 2005). In this manuscript, we introduce spectral and mapped class proportion domains as the explanatory domain for error prediction, domains that to the best of our knowledge have not been previously explored for error interpolation of soft classifications. The performances of the spectral and mapped class proportion interpolation methods are compared to two benchmark methods, a map-level summary measure and a spatial interpolation method.

The rest of the manuscript is organized as follows. In Section 2, the datasets including the reference data and satellite images used to evaluate the soft classification error mapping methods along with input data preprocessing are presented. The details of the four error prediction methods are explained in Section 3. In Section 4, the research experimental design and the overall workflow is elaborated, including details of the four main steps: (i) input data preprocessing, (ii) land-cover soft classification, (iii) classification error map predictions (using the methods discussed in Section 3), and (iv) evaluation of the predicted error maps. In Section 5, results of evaluation of the error prediction methods are discussed. Evaluations are based on (i) quantitative analysis using mean absolute error (MAE) as a measure to quantify prediction error and (ii) visual investigation and comparison of the reference and predicted error maps. Discussion and conclusions are presented in Section 6.

2. Datasets used to evaluate methods for predicting per-pixel error

The performances of error prediction methods were investigated using reference data from the United States Geological Survey (USGS) Land-Cover Trends project (Loveland et al., 2002). Twenty-three Trends blocks (Fig. 1) were used to provide a diverse set of examples to evaluate the error prediction methods. Each block represents a special case study. The 2011 Trends reference data for each block were obtained using manual interpretation of all pixels in the block providing a census of reference data at a 30 m pixel size. Each pixel was assigned a single class (hard classification) based on a modified Anderson (Anderson et al., 1976) Level I classification scheme including the following 11 land-cover classes: water, developed/urban, mechanically disturbed, barren, mining, forests/woodlands, grassland/shrubland, agriculture, wetland, non-mechanically disturbed, and ice/snow (see http://landcovertrends.usgs.gov/main/classification.html for the specific class definitions, last accessed April 2017). Each of the 23 blocks covered a 10 km × 10 km (333 pixels × 333 pixels) area (blocks are enlarged to enhance visualization in Fig. 1). Because the Trends reference data represent a hard classification, a recoding and resampling process was applied to convert these data to a soft classification for use as reference data as discussed below.

To evaluate the error prediction methods it was necessary to produce land-cover soft classification maps. The land-cover classification for each of the 23 Trends blocks was implemented by applying a spectral unmixing method using six reflective bands (excluding thermal) from Landsat TM images for 2011. This method required that the number of target classes not be larger than the number of spectral