Estimation of probabilistic accuracy measures in remotely sensed land cover change information

Na Yao[1]*, Jingxiong Zhang*[2]*, Wangle Zhang*[2]

[1]Beijing Institute of Space Mechanics and Electricity, Beijing, China

[2]School of Remote Sensing Information Engineering, Wuhan University, China

Abstract

It is well known that change detection and categorization are of limited accuracy due to various man-machine factors. However, it is a complicated matter to estimate accuracy levels in land cover change information derived from remote sensing data due to spatio-temporal interdependence (and inhomogeneity) in error occurrences, non-trivial extension of accuracy measures from single-date classifications to change detection, the often inadequacy of ground data, and other limitations. This paper seeks to provide a synthesis of some of the potentially useful methods for estimating accuracy in land cover change information. Geostatistical approaches will be valuable tools, though with computational and modeling costs, for providing accuracy-related information as they can incorporate spatio-temporal interdependence in estimating pixel-specific probability of correct classification.

Keywords: land cover change, accuracy, geostatistics, spatial and temporal dependence.

1. Introduction

Remotely sensed land cover change detection is becoming increasingly important for environmental monitoring and modeling. It is crucial to perform accuracy assessment in land cover change information so that derivatives and analysis results based on such information can be accountable scientifically and practically (Wickham *et al*., 2013). Statistical measures of uncertainty in land cover and other categorical information are mostly probabilistic (i.e., percent correctly classified (*PCC*) pixels, and per-class user's accuracy). Pixel-level accuracy estimates can also be derived from combined use of validation samples, classifications, and the source image data employed (Steele *et al*., 1998).

Although these measures can be used for land cover change information (which can be treated as a broad kind of categorical information), obtaining reliable estimates about accuracies in change information is hardly a simple matter. This is because product of accuracy measures for single-date classifications can not be used as an accurate measure for accuracy in change detection, unless temporal independence among events of single-date misclassification errors is assumed, supposing post-classification comparison is employed for change detection. Moreover, it tends to be more difficult to estimate pixel-wise (rather than global and class-level) accuracies in change detection, as local accuracy estimates for single-date classifications are usually less easier to obtain than global or class-level ones due to limited validation samples, as is usually the case in practice.

We provide a synthesis of some of the methods that may be used to derive location-specific estimates of classification accuracy (or the opposite, misclassification probabilities) in land cover change information derived from remotely sensed data. As will be reviewed below, there are approached based on probabilistic reasoning, maximum posteriori probabilities, regression analysis, resampling and interpolation, kriging, and geostatistical simulation, although they are not claimed to be an exhaustive list.

2. Methods

2.1. Probabilistic intervals

This method provides interval-valued estimates for classification accuracies based on single-date probabilistic accuracy measures when applying post-classification comparison for change detection (Khorram *et al.*, 1999). The probability of correct classification of multi-temporal images lies between a minimum, the product of individual single-date classifications when assuming independence of correct classifications of single-date images, and a maximum, which is the minimum of single-date probabilities of correct classification. The logic is applicable for PCC, user's accuracy, and pixel-wise probability of correct classification. The formulation is:

$$p^{LCC}(x) = [\prod_i p^{(i)}(x), \min(p^{(i)}(x))] \tag{1}$$

where $p^{LCC}(x)$ is the probability of correctly categorized land cover change at pixel x, and $p^{(i)}(x)$ the probability of correct classification at pixel x at time i.These single-date probabilities of correct classification need to be made available to feed in Equation (1) . For this, there are various options to obtain pixel-specific estimates of accuracy measures, as shown below.

2.2. Maximum posteriori probabilities as indices of accuracies

It is constructive to explore alternatives for quantifying misclassification. A good example is to use maximum posterior class probability as indices or surrogates of classification accuracy indicators, as described by Steele (2005). We can extend this method to change detection, as below.

In the context of change categorization, a compound classification according to Bayes theorem takes place when each pair of pixels is analyzed with the aim of finding the best pair of class labels. This Bayesian method is useful as it allows for tabulation of to-from transitions and accommodates temporal dependence. Let $\mathbf{Z}^1(x)$ and $\mathbf{Z}^2(x)$ be random vectors representing the observations at times t_1 and t_2, respectively. Suppose $\{\nu_i, i = 1, \ldots, K_1\}$ and $\{\omega_j, j = 1, \ldots, K_2\}$ are the sets of possible classes at times t_1 and t_2, respectively. The compound classification is written as:

$$\begin{aligned}
&\hat{\hat{C}}_1(x) = \nu_k \quad \text{and} \quad C_2(x) = \omega_l \quad \text{if and only if} \\
&p(\nu_k, \omega_l \mid \mathbf{z}^1(x), \mathbf{z}^2(x)) = \max_{i,j} p(\nu_i, \omega_j \mid \mathbf{z}^1(x), \mathbf{z}^2(x)) \\
&\qquad \approx \max_{i,j} p(\mathbf{z}^1(x) \mid \nu_i) p(\mathbf{z}^2(x) \mid \omega_j) p(\nu_i, \omega_j) \\
&\qquad = \max_{i,j} \frac{p(\nu_i \mid \mathbf{z}^1(x)) p(\omega_j \mid \mathbf{z}^2(x)) p(\nu_i, \omega_j)}{p(\nu_i) p(\omega_j)}
\end{aligned} \tag{2}$$

where $\hat{C}_1(x)$ and $\hat{C}_2(x)$ are the predicted class labels for location x at times t_1 and t_2, respectively; $p(v_i \mid \mathbf{z}^1(x))$ and $p(\omega_j \mid \mathbf{z}^2(x))$ are single-date *a posteriori* probabilities, and $p(v_i)$, $p(\omega_j)$, and $p(v_i, \omega_j)$ are *a priori* probabilities and joint probability, respectively; approximation occurs due to assumption of conditional independence (Zhang *et al*. 2014). $p(v_k, \omega_l \mid \mathbf{z}^1(x), \mathbf{z}^2(x))$ in Equation (2) can be used as an index of change detection (and categorization) accuracy, and need to be calibrated by validation samples, as in single-date classifications (Steele, 2005).

2.3. Regression analysis

The calibration in the previous sub-section was actually done by linear regression (Steele, 2005). Regression analysis has been proposed for accuracy modeling in which pixel values, their textural/contextual derivatives (such as local spatial pattern indices or landscape shape indices (Smith *et al*., 2002)), and other covariates may be incorporated as explanatory variables to predict local classification (in)accuracy. Logistic regression for indicator-coded (in)correctness in map classes based on landscape pattern indices and other covariates has been proposed for estimating misclassification probabilities (Smith *et al*., 2003; Van Oort *et al*., 2004). This may be implemented for global (taking a categorical map as a whole for global estimates of *PCC* or the like), class-specific, or location-specific accuracy measures, providing a comprehensive framework for estimating classification accuracy.

There are also challenges as to how localized estimates of misclassification probabilities in a change detection context may be derived from regression analysis by exploring relationships between occurrences of erroneous change detection and categorization and landscape shape indices (Van Oort *et al*., 2004). Regression analysis mentioned above may be applied to incorporate maximum posteriori probabilities.

2.4. Resampling and interpolation

To lessen the requirements for validation samples and to incorporate certainty-related information about image classification, we can employ resampling to estimate classification accuracy. Steele *et al*. (1998) presented a method using nearest neighbor classifier with bootstrap technology to estimate the misclassification probability of training samples prior to using kriging to generate surfaces of pixel-level misclassification probability. While it is a simple matter to change the indicators of misclassification to those denoting correct classification to facilitate estimation of probability of correct classification, extension of resampling-based methods to change detection scenarios relies, again, on the assumption of collocated validation samples over individual single-date classifications.

The bootstrap procedure mentioned above simulates the process of sampling and classification many times by a Monte Carlo method, and estimates the (in)correct classification probability at training points by the proportion of times that the training observation is (in)correctly classified across the simulations. This leads to a population concept of classification (in)accuracy. Because of sampling variation and measurement error (say in remote sensing images denoted as ovariate vectors **Z**), multiple repetitions of this procedure will yield different training samples, different classification rules, and different predictions of the class based on **Z**. The probabilistic (in)accuracy measure is the probability that the sampling and classification rule construction process will yield an (in)correct prediction of class at a location (or polygon) x when the rule is applied to **Z** at x (Zhang *et al*., 2006).

Note that (in)accuracy measures derived at training sample locations need to be densified to generate surfaces of (in)correct classification probabilities by kriging. This implies that training samples should be well distributed, as is usually required in accuracy assessment (Stehman, 2012).

2.5. Kriging

As shown previously, we can explore interpolation-based approaches to mapping probabilities of (in)correct classification over space, given sample data at a set of validation sites. Clearly, indicator krigimg can be used to estimate probabilities of (in)correct classification, based on indicator-transformed validation samples. While indicator kriging can be applied given adequately sampled validation data, extending methods for estimating single-date classification accuracy to those working for change detection accuracy is not so easy, as we would need to have validation samples collocated over the individual classification maps.

The aforementioned kriging-based approach to mapping classification accuacry is actually a kind of integration of probabilistic measures of accuracy derived from resampling (of training data) and locations of validated sample pixels. The resulting maps of classification accuracy indicate spatially varied confidence levels in pixel class labeling.

Indeed, we can combine the strengths of regression and kriging for localized estimates of classification accuracy. This may be done by predicting local means of probabilistic accuracy measures based on logistic regression mentioned previously. With adequately sampled validation data, we can apply kriging to generate estimates of local classification accuracy. Note that the kriged accuracy measures are somehow calibrated by the validation samples.

2.6. Geostatistical simulation

However, the maps of (in)correct classification probabilities or the like, which may be generated by kriging or other approaches, do not lend to straightforward quantification of spatial uncertainty unless spatial independence among neighboring pixels in a region is assumed. Stochastic simulation was proposed for generating multiple alternative realizations (maps) of the spatial distribution of class labels (validated with existing ground data but corroborated with image-based soft information about class occurrences) over the study area. The simulated alternative class label representations can be used for assessing joint spatial accuracy, i.e., classification accuracy regarding certain spatial features (Kyriakidis and Dungan, 2001).

Geostatistical simulation is very useful for accuracy estimation in the context of change detection, as it is possible to generate and summarize a large number of simulated land cover change maps conditional to available data (both validation data and image data) and conformal to spatio-temporal interdependence ascribed to the problem domain (Burnicki *et al*., 2007). The modeling and computing demands may be relieved by decomposing the bi-temporal change detection into two single-date classifications and then combining them through Equation (2).

For consistency in the simulation-based error modeling, phase space (or discriminant space)-based strategies were proposed, whereby discriminant covariates are simulated, with categorical maps simulated as equal-probable outputs of a classifier fed with the simulated equal-probable covariates (Goodchild *et al*., 2009). It is important to pursue extended use of discriminant models for error modeling from single-date categorical maps to those representing temporal changes (Zhang and Tang, 2012).

3. Conclusion

The methods reviewed in this paper, which can be seen as different generalizations of methods designed originally for single-date classifications, are shown to have their relative strengths and weakness. The complexity of change detection poses major challenges to their applications, and their performances remain to be evaluated with either simulated or real datasets.

These methods usually requires, implicitly, spatially well distributed sampled data, which are often not something that can be taken for granted due to cost of manpower and time. Furthermore, it should be noted that validation samples are always necessary for anchoring accuracy measures derived from the reviewed methods regardless of their sophistication.

Acknowledgments

The research reported in this paper is supported by the National Natural Science Foundation of China (grant no. 41171346).

References

Burnicki, A.C., Brown, D.G., Goovaerts, P. (2007), "Simulating error propagation in land-cover change analysis: the implications of temporal dependence". *Computers, Environment and Urban Systems*, 31(3): 282-302.

Goodchild, M.F., Zhang, J., Kyriakidis, P. (2009), "Discriminant models of uncertainty in nominal fields". *Transactions in GIS*, 13(1): 3–7.

Khorram, S., *et al.* (1999), *Accuracy assessment of remote sensing-derived change detection*. American Society for Photogrammetry and Remote Sensing, Bethesda, MD, U.S.A., 65p.

Kyriakidis, P.C., Dungan, J.L. (2001), "Stochastic simulation for assessing thematic classification accuracy and the impact of inaccurate spatial data on ecological model predictions". *Environmental and Ecological Statistics*, 17: 1083-1098.

Smith, J.H., Stehman, S.V., Wickham, J.D., et al. (2003), "Effects of landscape characteristics on land-cover class accuracy". *Remote Sensing of Environment*, 84(3): 342-349.

Smith, J.H., Wickham, J.D., Stehman, S.V., Yang, L. (2002), "Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy". *Photogrammetric Engineering and Remote Sensing*, *68*(1): 65-70.

Steele, B.M. (2005), "Maximum posterior probability estimators of map accuracy". *Remote sensing of environment*, 99(3): 254-270.

Steele, B.M., Winne, J.C., Redmond, R.L. (1998), "Estimation and mapping of misclassification probabilities for thematic land cover maps". *Remote Sensing of Environment*, 66(2): 192-202.

Stehman, S.V. (2012), "Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change". *Remote Sensing Letters,* 3: 111-120.

Van Oort, P.A., Bregt, A.K., de Bruin, S., de Wit, A.J., Stein, A. (2004), "Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database". *International Journal of Geographical Information Science*, *18*(6): 611-626.

Wickham, J.D., Stehman, S.V., Gass, L., et al. (2013), "Accuracy assessment of NLCD 2006 land cover and impervious surface". *Remote Sensing of Environment*, 130: 294-304.

Zhang, J., Atkinson, P.M., Goodchild, M.F. (2014), *Scale in Spatial Information and Analysis*, CRC Press, Boca Raton, U.S.A., 367p.

Zhang, J., Steele, B.M., Liu, J. (2006), Resampling-based accuracy assessment in remote sensed land cover information. In M. Raubal, H. J. Miller, A. U. Frank, & M. F. Goodchild (Eds.), Proceedings of Geographic Information Science: 4th International Conference, GIScience 2006, Münster, Germany, September 20-23, 2006, pp 213-216.

Zhang, J., Tang, Y. (2012), "Misclassification error propagation in land cover change categorization". *Geo-Spatial Information Science*, 15(3): 171-175.