Misclassification error propagation in land cover change categorization

ZHANG Jingxiong* and TANG Yunwei

School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

(Received 1 June 2011; final version received 7 August 2012)

It is important to describe misclassification errors in land cover maps and to quantify their propagation through geo-processing to resultant information products, such as land cover change maps. Geostatistical simulation is widely used in error modeling, as it can generate equal-probable realizations of the fields being considered, which can be summarized to facilitate error propagation analysis. To fix noninvariance in indicator simulation, discriminant space-based methods were proposed to enhance consistency in area-class mapping and replicability in uncertainty modeling, as the former is achieved by imposing means while the latter is ensured by projecting spatio-temporal correlated residuals in discriminant space to geographic space through a mapping process. This paper explores discriminant models for error propagation in land cover change detection, followed by experiments based on bi-temporal remote sensing images. It was found that misclassification error propagation is effectively characterized with discriminant covariate-based stochastic simulation, where spatio-temporal interdependence is taken into account.

Keywords: error propagation; area-class maps; land cover change; discriminant space; data class; information class; stochastic simulation

1. Introduction

There is a large literature on errors in spatial information and applications (1, 2). It is important to evaluate the errors in land cover change maps given estimates about errors in single-date land cover maps. Land cover change maps can be seen as products of change detection techniques based on bi-temporal or multi-temporal land cover maps. Thus, all factors affecting single-date land cover maps will act jointly in land cover change analysis, as errors in single-date land cover maps will be propagated into the resultant land cover change information. For instance, the accuracy with which single-date land cover maps are co-registered will affect the accuracy of change detection (3). Although there are different errors affecting accuracy in change detection, this paper focuses only on effects of misclassification upon change detection.

Misclassification is usually analyzed by confusion matrices where classification maps and their references are compared and the hits and misses between them are cross-tabulated on samples of pixels or parcels. However, such measures neither consider the spatio-temporal dependence of bi-temporal or multi-temporal land cover maps, nor offer hints about spatial variation in misclassification.

To assess errors in land cover change, it is not only necessary to be able to figure out single-date misclassification errors, but also to quantify and account for spatio-temporal interdependence between errors in single-date classification maps, as misclassification in land cover change is not a simple union of disjoint events of single-date misclassifications. In addition to estimation of misclassification probability, variance in misclassified land cover change must also be estimated, relying on knowledge of spatial and temporal interdependence among misclassification over specific types or spatial extents (4). This paper explores discriminant models under the geostatistical framework for misclassification error propagation in land cover information, where spatio-temporal interdependence is taken into account.

This research is pursued by devising methods for mapping area classes in the space of discriminant covariates and projecting errors from discriminant space to area classes in the geographic space. Spatially explicit quantification of spatial uncertainty in land cover classes and their change is facilitated by transforming and summarizing stochastically simulated class-defining discriminant covariates. Misclassification due to data and information class semantics will be discussed separately from misclassification due to measurement errors. Experiments with real data-sets will be carried out to demonstrate the effectiveness of discriminant models in misclassification analysis and how to deal with the semantic discrepancy between data and information classes.

2. Methods

In order to facilitate misclassification error quantification and propagation, it is important to establish class models in the discriminant space, also known as Z space, which

*Corresponding author. Email: jxzhang@whu.edu.cn

ISSN 1009-5020 print/ISSN 1993-5153 online
© 2012 Wuhan University
http://dx.doi.org/10.1080/10095020.2012.715874
http://www.tandfonline.com
is spanned by discriminant covariates. Suppose a training sample $T$ of size $n$ is collected from a population comprised of elements $T(x) = (Z(x); C(x))$, where $Z(x)$ and $C(x)$ refer to the measurement and area-class defined for location $x$. Let $\eta$ denotes a classification rule trained on $T$, which obtains a prediction of $C$ based on measurement $Z$. A classifier can be seen as the mapping from measurement to class labels and expressed as:

$$ \hat{C}(x) = \eta(Z(x)) = \arg \max_{k=1\ldots \nu} f_k(Z(x)) $$  

(1)

where $f_k$ indicates categorical similarity to class $k$, with the predicted class $\hat{C}(x)$ at location $x$ taking the maximum utility. A classifier can also be formulated as an estimator of posterior probabilities for location $x$, with $x$ labeled as the class with the maximum posterior probability.

Consider change categorization based on bi-temporal measurement. Let $\mathbf{x}(t)$ and $\mathbf{z}(t)$ be random vectors and stand for the observations at times $t_1$ and $t_2$, respectively. Suppose $\{v_i, i = 1, \ldots, K_1\}$ and $\{w_j, j = 1, \ldots, K_2\}$ are the sets of possible classes at times $t_1$ and $t_2$, respectively. A comprehensive classification takes place when each pair of pixels with observations is analyzed with the aim of finding the best pair of class labels. According to Bayes theorem, this process is written as:

$$ \hat{C}_1(x) = v_i \text{ and } \hat{C}_2(x) = w_j \text{ if and only if } p(v_i, w_j | \mathbf{z}(t_1), \mathbf{z}(t_2)) = \max_{i,j} p(v_i, w_j | \mathbf{z}(t_1), \mathbf{z}(t_2)) $$

(2)

where $\hat{C}_1(x)$ and $\hat{C}_2(x)$ are the predicted class labels for location $x$ at time $t_1$ and time $t_2$, respectively.

With class-conditional dependence in the time domain assumed, it is possible to replace the quantity in the right side of Equation (2) with:

$$ p(v_i, w_j | \mathbf{z}(t_1), \mathbf{z}(t_2)) \propto p(\mathbf{z}(t_1) | v_i) p(\mathbf{z}(t_2) | w_j) p(v_i, w_j) $$

(3)

where $p(v_i, w_j)$ is the prior joint probabilities. To map single-date class probabilities $p(v_i | \mathbf{z}(t_1))$ and $p(w_j | \mathbf{z}(t_2))$ in $Z$ space, kernel density estimation can be used as a nonparametric way of estimating the density function of a random variable. Equations (1) and (2) imply that class probabilities vary in the $Z$ space, so does misclassification probability, as shown below.

Let $I_c(x)$ and $I_e(x)$ denote indicator variables for correct and incorrect classification at $x$ given measurement $Z(x)$, respectively. Their means refer to probability of correct and incorrect classification conditional to $Z(x)$, respectively, which are denoted by $P[I_c(Z(x)) = C(x) | Z(x)]$ and $P[I_e(Z(x)) = C(x) | Z(x)]$, respectively. It is proved mathematically that the probability that $x(x) = (Z(x); C(x))$ is correctly classified, given $Z(x)$ is the maximum posterior probability (6). Thus, misclassification probability conditional to $Z(x)$ is:

$$ \begin{align*}
\bar{I}_c(x) &= P[\eta(Z(x)) = C(x) | Z(x)] \\
&= 1 - P[\eta(Z(x)) = C(x) | Z(x)] \\
&= 1 - \arg \max_{k=1\ldots \nu} p(C(x) = k | Z(x)) 
\end{align*} $$

(4)

The classifier $\eta$ allows construction of estimator $P(C = k | Z(x))$, for each class $k$, thus misclassification probability.

Such estimators can be written for both single-date classification maps and bi-temporal change categorization. In change categorization, the class labels $k$ refer to categorized changes so that $P[\eta(Z(x)) = C(x) | Z(x)]$ is still cognate but needs to be interpreted properly, i.e. class label $k$ denoting a “from-to” class itself. In this case, it is tempting to reduce the computing for misclassification probability to certain combination of single-date ones. Unfortunately, provision of single-date misclassification probability does not make it straightforward to arrive at estimation of misclassification in the resultant land cover change.

Misclassification probability maps corresponding to the land cover maps at times $t_1$ and $t_2$ are derived from Equation (1), which are denoted as $P[I_c^{t_1}(x)]$ and $P[I_c^{t_2}(x)]$, respectively. The probability of incorrectly classified land cover change, $P[I^{t_1\rightarrow t_2}(x)]$, can be computed as the probability of the union of two events, i.e. $I_c^{t_1}(x)$ and $I_c^{t_2}(x)$:

$$ P(I_c^{t_1\rightarrow t_2}(x)) = P(I_c^{t_1}(x)) + P(I_c^{t_2}(x)) - P(I_c^{t_1\cap t_2}(x)) $$

(5)

where $P(I_c^{t_1\cap t_2}(x))$ stands for the probability of misclassification at both times $t_1$ and $t_2$, which discounts the amount of otherwise double-counted misclassification probability. $P(I_c^{t_1\cap t_2}(x))$ may be computed as $P(I_c^{t_1}(x))P(I_c^{t_2}(x)$ based on the assumption that $I_c^{t_1\cap t_2}(x)$ is the intersection of two independent events of incorrect classification $I_c^{t_1}(x)$ and $I_c^{t_2}(x)$, i.e. $x$ being misclassified at times $t_1$ and $t_2$, respectively.

However, it is often the case that (in)correctly classified pixels or parcels on bi-temporal land cover maps tend to be positively correlated, because the relationships between misclassification and landscape characteristics, such as patch size and land cover heterogeneity, were confirmed (7). Thus, evaluation of uncertainty on the land cover change map should be conducted through a more elaborate procedure that takes account of co-occurrence of (in)correct classification on the bi-temporal land cover maps, i.e. $P(I_c^{t_1\cap t_2}(x))$. This is where stochastic simulation can contribute and $Z$-based simulation should be pursued for replicability in simulated area-class maps (8, 9). What is needed is an error model or mechanism that can simulate realized maps conforming to observed distributions of categories and their variogram models (10).

With simulated $Z$ maps, simulated area-class maps can be derived for times $t_1$ and $t_2$ by using Equation (1)
and simulated land cover change using Equation (2). These simulated area-class maps can be compared with mean maps, which are generated by evaluating class memberships (using Equations (1) and (2)) with \( m_Z \) maps, to obtain realized binary maps depicting (in)correct classification at individual locations with respect to mean classes.

With validation samples available, we can further assess if the simulated area-class or change at individual location are misclassified with respect to reference data. The probability of misclassification at location \( x \) at both times \( t_1 \) and \( t_2 \) may be quantified through calculating the relative frequency of incorrect classification concurring at both time points:

\[
p_{mis}^{1/2}(x) = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} p_i(\xi_{t_1}(x)) \times p_i(\xi_{t_2}(x))
\]

where \( n_{sim} \) is the number of realizations generated for the purpose of quantifying (in)correct classification in land cover change, \( p_i(\xi_{t_1}(x)) \) and \( p_i(\xi_{t_2}(x)) \) stand for indicators checking if the classes at \( x \) on realized maps at times \( t_1 \) and \( t_2 \) are misclassified by testing against validation samples. The location set \( \{x\} \) can be made to exhaust any patches of pixels or parcels of irregular polygons to facilitate the quantification of uncertainty in any complex queries about land cover information over space and time.

The notion of mean area-class maps mentioned previously hints on that of data classes. As discussed in the introductory section, applications of discriminant models usually assume that class zones in the \( Z \) space can be accurately delineated on the basis of \( m_Z \). However, discriminant models constructed on empirical data do not always support 1–1 relationships between \( m_Z(x) \) and \( C(x) \). Thus, it may not be sensible to directly describe misclassification with respect to information classes \( C \), as misclassification due to \( Z \) deficiency in class discrimination and measurement errors would become interwoven and difficult to disentangle. To separately evaluate misclassification due to model imperfection and measurement errors, data classes denoted \( D \in \{1, \ldots, P\} \) should be discerned from \( Z \) space to act as reference for the latter, while the former is quantified on the basis of transition probabilities from data classes \( D \) to information classes \( C \).

3. Experiments

An area of central Montana, USA, located at 46°25’ ~ 48°30’ N and 108°04’ ~ 111°10’ W, was chosen as the study area. The NLCD 1992–2001 Land Cover Change Retrofit product developed by scientists at MRLC was used to derive more accurate and useful land cover change data than would be possible by direct comparison of NLCD 1992 and NLCD 2001. Along with the MRLC land cover change product also downloaded were Landsat 5 Thematic Mapper (TM) images (P38/R27) obtained on 17 July 1992 (time 1) and 11 August 2001 (time 2) with bands 1–5 and 7 at 30 m resolution. A subset covering 500 by 500 Landsat TM image pixels was selected as the dataset for the studies. This product contains unchanged pixels (6 class labels) and changed pixels labeled with a “from-to” class code (12 change classes). The 6 land cover labels are 1-open water, 2-barren (sand), 3-forest, 4-grassland/ shrub, 5-agriculture, and 6-wetlands. The bi-temporal Landsat TM images were tasseled cap transformed (11) and bands of brightness and greenness were selected and further transformed via Choleski factorization, so that Euclidean distance can be computed in lieu of Mahalanobis distance. This resulted in the discriminant covariates at times 1 and 2, respectively.

Data classes were calculated by plotting all pixels in the \( Z \) space discretized into a grid of 256 by 256 cells and summarizing land cover class labels of pixels falling into individual grid cells. The resultant class probability vectors estimated were considered as the mean class probabilities. The class labels with maximum probabilities in these grid cells were taken as the labels of data classes, so that all pixels were separable in the \( Z \) space. This gave rise to maps showing data classes of land cover, which correspond but do not equal to NLCD land cover classes (considered to be information classes).

It was further possible to identify pure pixels based on tallied proportions of information class labels in grid cells over the discretized \( Z \) space. To make sure the training samples can be separated completely in geographic space, random sampling was applied to select 2841 training pixels from a total of 5682 high purity (97%) pixels. As data classes revised all pixels to pure pixels, training samples for information classes and data classes should have similar proportion. In fact, two sets of data shared the same labels in all 97% purity pixels, thus only one set of training data was applied to 1992 and 2001 land cover mapping.

The training samples were used for kernel-based density estimation in the \( Z \) space. The results were probability vectors for all 250,000 pixels in the study area, one set for land cover in 1992 and the other for land cover in 2001. It was possible to derive land cover change from these single-date land cover probability maps based on the Bayesian classification rule specified in Equation (3).

The trend surfaces were discerned for \( Z_1 \) and \( Z_2 \) fields at time 1 and time 2. The corresponding residual surfaces for the bi-temporal vectors were obtained, denoted as \( \{R_1^{(1)}, R_2^{(1)}\} \) and \( \{R_1^{(2)}, R_2^{(2)}\} \). By the discriminant method of error modeling, stochastic simulation in the \( Z \) space was performed using \( \text{sgsim} \) in GSLIB on these de-trended data-sets. Normal score transforms necessary for \( \text{sgsim} \) were performed and their variogram models derived, which are 0.6°exp(-h/100) +0.4°nug(0), 0.55°exp(-h/95)+0.45°nug(0), 0.55°exp(-h/90)+0.45°nug(0), and 0.55°exp(-h/90)+0.45°nug(0) for \( R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, \) and \( R_2^{(2)} \), respectively. The realized residual surfaces were added to the corresponding trend.
surfaces to simulate equal-probable and error-contaminated $Z$ data. The discriminant variables $Z_1(1)$, $Z_2(1)$ and $Z_1(2)$, $Z_2(2)$ were independently simulated, each with 100 realizations.

By checking realized land cover change maps against reference classes, misclassification probabilities for individual pixels were obtained. These probabilities were then averaged over all land cover change types. Results of misclassification of estimated land cover change referring to information classes and data classes are reported in Table 1. The proportion of the areas of different change types is also shown in Table 1, along with misclassification probabilities that were computed based on Equation (4) and averaged over the land cover change types. The differences between calculated and simulated misclassification probabilities of land cover change were tested for significance using $t$-test. The proportions of pixels within individual change types, which were confirmed significant in terms of the difference between calculated and simulated probabilities of misclassification, were calculated in order to compare the methods with or without temporal correlation incorporated.

As can be seen from Table 1, the area proportion of a certain class type has a great effect on its misclassification probability. For example, the changed type “grassland/shrub” has the most serious misclassification. However, its result is not shown in Table 1, because its training samples are too rare to be put into classification. In contrast, the changed types of “barren (sand)” and “forest” are less misclassified, mainly due to their large area proportion and the amount of training samples with high purity, accounting for over 70 and 98.9%, respectively.

Comparing misclassification under two references in Table 1, the percentage of misclassification is generally lower in reference to data classes than in reference to information classes, suggesting that data classes can bridge the discrepancy between measurements and information classes.

It is shown that there is a general underestimation of the misclassification probabilities by the analytical
method. This is because the relationship between bi-temporal misclassification \( P(t^{1/2}(x)) \) can be computed as the product of single-temporal data-sets \( P(I^1(x))P(I^2(x)) \) only if the \( I^1(x) \) and \( I^2(x) \) are assumed to be two independent events of incorrect classification, but they are usually spatio-temporal correlated in reality. For example, the unchanged type of “open water,” of which only 23.92% was expected to be misclassified according to the analytical method, but was assessed to be 96.12 and 92.71% misclassified referring to information and data classes, respectively, due to the interdependence between single-date maps. By testing the difference between simulated and calculated misclassification probabilities, the results show significant difference between these two methods, which suggest that the spatio-temporal interdependence existing in the single-date maps cannot be ignored. Also, \( P(t^{1/2}(x)) \) may have an opposite effect on misclassification of estimation sometimes. The unchanged type of “forest” shows a lower misclassification probability when evaluated by simulation than by analytical estimation.

4. Conclusion

This paper explored discriminant models for quantifying misclassification in single-date land cover maps and their propagation in land cover change maps. For this, stochastic simulation can generate equal-probable area-class maps based on realized \( Z \), which can be used to compute summary statistics, such as means of individual class proportions over specific areas or different land cover transitions for land cover change. It was confirmed that spatio-temporal interdependence plays an important role on misclassification error propagation in land cover change products, and simple applications of the law of variance and covariance propagation without accounting for spatial dependence would lead to biased quantification of standard errors in derived results. For assessing errors in change categorization, interpixel interactions and temporal correlation should be studied before arriving at unbiased estimation of error variance.

As seen in the experiments, there are different concepts and definitions about the “classes” so that clarification is necessary for objective evaluation of misclassification. Referring to data classes, misclassification probabilities are attributable solely to measurement errors, which are modulated by data-information class transition probabilities when referring to information classes. It was shown that discriminant models are effective to characterize uncertainty due to measurement error in \( Z \), and data-information class definition discrepancy should be dealt with separately when calculating misclassification error. Further research should be focused on detecting misclassification when the predicted class does not agree with the true class through the notion of bias and variance decomposition to isolate the effects due to semantics vs. measurement errors.

Acknowledgments

Supported by the National Natural Science Foundation of China (Nos. 41071286 & 41171346), Hubei Provincial Science and Technology Department (2007ABA276).

Notes on contributor

Zhang Jingxiong obtained his PhD in geography from the University of Edinburgh and he is currently a professor in the School of Remote Sensing and Information Engineering, Wuhan University. He has research interests in spatial analysis, remote sensing, and GIScience.

References

(1) Goodchild, M.F.; Guoqing, Sun; Shiren, Yang Development and test of an error model for categorical data. Int. J. Geogr. Inform. Syst. 1992, 6, 87–104.
(2) Jingxiong, Zhang; Goodchild, M.F. Uncertainty in Geographical Information; Taylor & Francis: London, 2002.
(3) Verbyla, D.L.; Boles, S.H. Bias in land cover change estimates due to misregistration. Int. J. Remote Sens. 2000, 21 (18), 3553–3560.
(4) Murtaugh, P.A.; Phillips, D.L. Temporal correlation of classifications in remote sensing. J. Agric. Biol. Environ. Stat. 1998, 3 (1), 99–110.
(5) Parzen, E. On estimation of a probability density function and model. Ann. Math. Stat. 1962, 33 (3), 1065–1076.
(6) Steele, B.M.; Patterson, D.A.; Redmond, R.L. Towards estimation of map accuracy without a probability test sample. Ecol. Environ. Stat. 2003, 10, 333–356.
(7) Smith, J.H.; Wickham, J.D.; Stehman, S.V.; Yang, L. Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy. Photogr. Eng. Remote Sens. 2002, 68 (1), 65–70.
(8) Arbia, G.; Griffith, D.A.; Haining, R.P. Spatial error propagation when computing linear combination of spectral bands: The case of vegetation indices. Environ. Ecol. Stat. 2003, 10 (3), 375–396.
(9) Goodchild, M.F.; Jingxiong, Zhang; Kyriakidis, P. Discriminant models of uncertainty in nominal fields. Transactions in GIS 2009, 13 (1), 7–23.
(10) Journel, A.G.; Huijbregts, C.H.J. Mining Geostatistics; Academic Press: London, 1978.
(11) Crist, E.P.; Cicone, R.C. A physically-based transformation of Thematic Mapper data- the TM tasseled cap. IEEE Trans. Geosci. Remote Sens. 1984, 22 (3), 256–263.