Effects of landscape characteristics on accuracy of land cover change detection

Yingying Mei, Jingxiong Zhang

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

Abstract

The effects of landscape characteristics on land cover change detection accuracy were evaluated. Logistic regression models with change/no-change classes were extended to include bi-temporal explanatory variables: patch density, largest patch index and landscape shape index at both "from" and "to" times. It was found that, these bi-temporal variables had significant effects on change detection accuracy. To validate the models, a leave-one-out procedure was applied in which the absolute difference between the actual and the model-estimated was summarized over 20 sample area units in Wuhan. The sum of differences reduced from 1.68 to 1.47 after adding such bi-temporal variables as 'patch density', 'largest patch index' and 'landscape shape index' to the models concerning binary change/ no-change classes. Therefore, these landscape characteristics led to a substantial improvement in the estimation of change detection accuracy.

Keywords: classification, change detection accuracy, landscape characteristics, logistic, bi-temporal.

1. Introduction

Change detection, which seeks to identify differences in the states of spatially distributed phenomena, has been widely used in resources manage and environmental monitoring (Macleod, 1994). Quantifying the extent and rate of land-cover change, and developing models to relate the processes driving changes in land use to observed changes in land-cover are important challenges in the field of global change science. The limited knowledge concerning the accuracy of land-cover change products directly affects our ability to make accurate predictions of change (Foody, 2002). Error matrix, an extension of the single-date classification error matrix, has been widely used in accuracy assessment in change detection (Congalton and Macleod, 1994; Conglaton and Green, 1999), and is important in an error propagation analysis or estimation of the fitness for use of the data for a specific region. However, accuracy assessment is hampered by sorts of difficulties (Achard *et al.*, 2002; De Zeeuw and Hazeu, 2001) implying that much work is needed for quantitative estimation of change detection accuracy.

Up to now, many methods for the estimation of change detection accuracy have been proposed. Although simple methods may be adopted to derive accuracy measures of change detection based on individual classification maps (Congalton and Green,1999; Carmel et al.,2001; Sohl et al.,2004), more elaborated approaches should be explored for improved reliability in accuracy assessment, when dealing with spatial data which is characterized by data dependency and heterogeneity (Van Oort, 2005; Burnicki et al., 2007). More recent developments applying transition rules to land cover change trajectories to assess the accuracy of changes predicted across more than two classified images (Liu and Zhou, 2004).

Logistic regression models have been used to analyse the relationship between land cover classification accuracy at single time and landscape pattern indices (Smith *et al.*, 2002; Van Oort *et al.*, 2002). A model was established to illustrate the impact of locations on the uncertainty of land-cover maps (Carnel and Dean, 2004). Also, landscape pattern analysis (LPA) is an important tool to quantify changes in land cover maps (Turner et al.2001; Farina, 2006).

This paper seeks to extend landscape characteristics as potential explanatory variables for uncertainty evaluation from single-date data sets to change categorization. We established the logistic regression models with landscape characteristics as the explanatory variables with change detection accuracy. In order to validate the models, a leave-one-out procedure was applied. A substantial improvement in the estimation of change detection accuracy was achieved.

2. Methodology

2.1 Data

Two sub-scene ETM+ images of Wuhan flown on May 2012 (time1) and August 2013 (time2) with bands 1-5 and 7 at 30 m resolution were used to calculate landscape variables. Twenty area units, each of which contains 500*697 pixels, were sampled from the bi-temporal images at the same location. Post-classification was used to determine the changes in land cover information. Two high resolution land cover maps of the same region at corresponding time were used as reference data to estimate change detection accuracy. Figure1 shows the distribution of the 20 units. The black unit was fitted to validate the percentage correctly classified of the other 19 units (white).

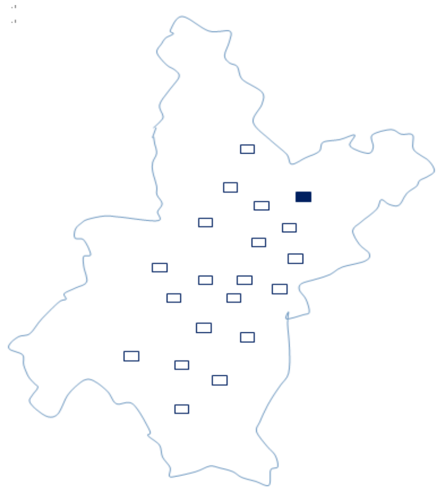

Figure1. Wuhan with the sample area units (rectangles) used.

2.2 Explanatory variables

Table1 shows the explanatory variables used.

Table 1: Explanatory variables.

Variable	Ab.	Key aspects of variable
No-change/Change Class	Class	two binary variables are used to indicate change or no-change
Patch Density	PD	the extent to which the landscape is fragmented
Largest Patch Index	LPI	a standardized measure of the plaque concentration and landscape dominance
Landscape Shape Index	LSI	a standardized measure of total edge or edge density that adjusts for the size of the landscape

2.3 Statistical analysis

2.3.1 Logistic Regression

The logistic regression (Smith *et al.*, 2002; Lin *et al.*, 2008) was used to calculate the distribution of cells correct classification p as a function of the landscape variables introduced above. The logistic regression model with intercept β_0 and with $k = 1 \cdots K$ explanatory variables x_k equals:

$$\mathrm{p} = \frac{\exp(\beta_0 + \sum_{k=1}^{K} \beta_k \cdot x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^{K} \beta_k \cdot x_k)} \tag{1}$$

Formula (1) can be expressed as follows:

$$\mathrm{logit(p)} = \ln \frac{p}{1-p} = \beta_0 + \sum_{k=1}^{K} \beta_k \cdot x_k \tag{2}$$

Regression coefficients are calculated by minimizing the -2 log likelihood of the model.

In order to evaluate the impact of these variables on land-cover change accuracy, a set of models including different variables were built (Table2), where PD1、LPI1 and LSI1 express the landscape variables on time1, PD2、LPI2 and LSI2 on time2.

Table 2: Description of models.

Model number(m)	Model	Variables
0	β_0	Interception
1	$\beta_0+\beta_{1-2}\cdot Class$	Interception, Class
2a	$\beta_0+\beta_{1-2}\cdot Class+\beta_3X_3+\beta_4X_4$	Interception, Class,PD1,PD2
2b	$\beta_0+\beta_{1-2}\cdot Class+\beta_5X_5+\beta_6X_6$	Interception, Class,LPI1,LPI2
2c	$\beta_0+\beta_{1-2}\cdot Class+\beta_7X_7+\beta_8X_8$	Interception,Class,LSI1,LSI2
3a	$\beta_0+\beta_{1-2}\cdot Class+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6$	Interception, Class,PD1,PD2, LPI1,LPI2
3b	$\beta_0+\beta_{1-2}\cdot Class+\beta_3X_3+\beta_4X_4+\beta_7X_7+\beta_8X_8$	Interception, Class,PD1,PD2, LSI1,LSI2
3c	$\beta_0+\beta_{1-2}\cdot Class+\beta_5X_5+\beta_6X_6+\beta_7X_7+\beta_8X_8$	Interception, Class, LPI1,LPI2, LSI1,LSI2
4	$\beta_0+\beta_{1-2}\cdot Class+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+\beta_7X_7+\beta_8X_8$	Interception, Class, PD1,PD2, LPI1,LPI2, LSI1,LSI2

2.3.2 Validation

A leave-one-out procedure (Van Oort *et al.*, 2004) was used to find the model containing the highest number of significant explanatory variables. Each one of the 20 area units was used to test the model based on the other 19 units. The resulting 20 parameter sets formed versions of a model. The measure SM_m (Equation (3)) gives the residuals of each model, which is better with a small value. $R0_m$ (Equation (4)) and $R1_m$ (Equation (5)) describe the relative improvement of model m to the model 0 assuming the same probability of correct classification for all cells and model 1 containing no-change/change class, respectively.

$$\mathrm{SM_m}=\sum_{r=1}^{20}\sum_{c=1}^{n(r)}\left|p_m(\mathrm{c})-\mathrm{y(c)}\right| \tag{3}$$

$$R0_m=100\%\cdot(\mathrm{SM}_0-\mathrm{SM}_m)/\mathrm{SM}_0 \tag{4}$$

$$R1_m=100\%\cdot(\mathrm{SM}_1-\mathrm{SM}_m)/\mathrm{SM}_1 \tag{5}$$

where $p_m(c)$ is the correct classification probability of model m, $n(r)$ is the number of cells in unit r, the binary variable $y(c)$ indicts if the cells c is correctly classified or not.

3. Results and Discussion

3.1 Change/no change error matrix

Table 3 shows accuracy measures at two dates.

Table 3: Accuracy measures.

	Users' accuracy(%)				PCC(%)
	water	urban	forest	farmland	
time1	98.79	87.91	84.27	92.34	96.85
time2	97.84	88.12	72.17	93.32	94.04

The change/no change error matrix is shown in Table 4.

Table 4: Change/no change error matrix.

	No Change	Change	Total
No Change	**30262**	2264	32526
Change	796	**851**	1647
Total	31058	3115	34173

Change detection accuracy= (30262+851)/34173=91.05%.

3.2 Model selection

Table 5 gives the estimated regression coefficients for a selection of models. Model1 shows the effect of Class on change detection accuracy, model 2a-2c show the effect of the landscape characteristics variables PD、LPI and LSI when Class is accounted for, model 3a-3c show the interaction of two different landscape indices accompanied with Class. Model 4 is the model that shows the comprehensive effects of the four explanatory variables.

Table 5: Estimated regression coefficients.

Model number(m)	Regression coefficients								
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0	2.3198								
1	-2.0919	3.4799	0.1216						
2a	0.4933	0.0087	-3.1415	0.0840	0.0180				
2b	-0.5648	2.4245	0.0123			-0.0260	-0.0018		
2c	0.5741	0.0002	-3.2919					0.0094	-3.0652
3a	-0.6150	2.4176	0.0156	-0.0200	0.0306	-0.0296	0.0019		
3b	0.6097	0.006	-3.1544	0.0902	0.0230			-0.0020	-5.1736
3c	-0.1910	2.4215	0.0089			-0.0311	0.8258	-0.0091	0.0058
4	-0.2190	2.4321	0.1134	-0.0057	0.0398	-0.0320	0.0030	-0.0077	0.9989

3.3 Model validation

Table 6 gives the validation results of models established. It indicates that all model estimates were better than model 0($R0_m$ >0 for all m). Addition of a landscape variable to a model containing Class (models 2a-2c) improved model 1 estimates by 2.69 to 10.67%. The interaction of two of the three landscape indices to the model containing variable Class (models 3a-3c) was significant: $R1_{3a}$ =12.06%, $R1_{3b}$ =5.12%. Relatively to models 0 and 1, the result of model 4 is closest to the actual value: $R0_4$ =38.39% and $R1_4$ =12.53%.

Table 6: Validation.

Model number(m)	Model description	Relative to models 0 and 1		
		SM_m	$R0_m$ (%)	$R1_m$ (%)
0	same PCC in all units	2.39		
1	Class(=error matrix)	1.68	29.57	
2a	Class&PD1&PD2	1.59	33.35	5.36
2b	Class&LPI1&LPI2	1.50	37.09	10.67
2c	Class&LSI1&LSI2	1.64	31.47	2.69
3a	Class&PD1&PD2& LPI1&LPI2	1.48	38.07	12.06
3b	Class& PD1&PD2& LSI1&LSI2	1.59	33.18	5.12
3c	Class& LPI1&LPI2& LSI1&LSI2	1.48	37.95	11.89
4	Class&PD1&PD2& LPI1&LPI2& LSI1&LSI2	1.47	38.39	12.53

4. Conclusion

It is difficult for remote sensing classification methods only based on spectral response to distinguish objects with similar spectral characteristics but of totally different classes. This means that accuracy in classification and change detection will be affected by class inseparability inherent to remotely sensed data and other factors in labelling class and detecting change. To estimate accuracy in classification and change detection, it is important to investigate the methods by which accuracy can be inferred from the spatial patterns in the maps being examined. Existing literature indicates that landscape variables are significant predictors of classification accuracy, and can be used to improve the prediction reliability of change detection accuracy. This paper evaluated landscape indices as potential explanatory variables of variability in change detection accuracy considering the class is change or not. Logistic regression models were developed to include three landscape indices: patch density, largest patch index and landscape shape index as the explanatory variables for uncertainty evaluation in change categorization. In order to quantify the impact of landscape indices on accuracy of change detection, a leave-one-out procedure was applied to evaluate the estimated residuals of each model.

Several remarks can be made as follows:

1) Logistic regression model can be used effectively for analyzing land-cover change and its environmental effects. Quantitative analysis of the relationship between the change detection accuracy and binary landscape index was used to simulate the land cover change tendency. The principle of the land-cover evolution can be used in change detection prediction.

2) Addition of landscape characteristics variables PD、LPI and LSI to spectral classification led to a substantial improvement in the estimation of change detection accuracy. The three landscape indices were all improved the reliability of estimation results in different degrees, where LSI had the weakest impact among the three landscape indices.

3) It was found that change detection accuracy was significantly higher for Logistic models with more explanatory variables. Synergistic effect of the three landscape indices is more than single variables above. The highest estimates of change detection

accuracy were obtained in models containing the four variables above, where the estimation results are closest to the actual values.

Future research will focus on how the impacts of landscape characteristics may vary across different land cover and change types and over space. Regression-based methods for analyzing accuracy of change detection can be usefully combined with other analytical approaches to accuracy estimation to make better use of temporal correlation and spatial patterns of misclassification errors.

Acknowledgments

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

References

Macleod, R. D., Congalton, R. G. (1998). A quantitative comparison of change-detection algorithms for monitoring eelgrass from remotely sensed data. *Photogrammetric engineering and remote sensing*, *64*(3), 207-216.

Foody G M(2002). Status of land cover classification accuracy assessment[J]. *Remote sensing of environment*, 80(1): 185-201.

Congalton R G, Macleod R D(1994). Change detection accuracy assessment on the NOAA Chesapeake Bay Pilot Study[C]. *American Society of Photogrammetry and Remote Sensing, Williamsburg*: 78-87.

Congalton R G, Green K(2008). Assessing the accuracy of remotely sensed data: principles and practices[M]. *CRC press*.

Achard, F., Eva, H. D., Stibig, H. J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J. P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*, *297*(5583), 999-1002.

Zeeuw, C. D., & Hazeu, G. W. (2001). Monitoring land use changes using geo-information: possibilities, methods and adapted techniques. *Alterra-rapport*.

Carmel Y, Dean D J, Flather C H(2001). Combining location and classification error sources for estimating multi-temporal database accuracy[J]. *Photogrammetric Engineering and Remote Sensing*, 67(7): 865-872.

Sohl T L, Gallant A L, Loveland T R(2004). The characteristics and interpretability of land surface change and implications for project design[J]. *Photogrammetric Engineering and Remote Sensing*, 70(4): 439-450.

Van Oort P A J(2005). Improving land cover change estimates by accounting for classification errors[J]. *International Journal of Remote Sensing*, 26(14): 3009-3024.

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

Liu H, Zhou Q(2004). Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison[J]. *International Journal of Remote Sensing*, 25(5): 1037-1050.

Carmel Y, Dean D J(2004). Performance of a spatio-temporal error model for raster datasets under complex error patterns[J]. *International Journal of Remote Sensing*, 25(23): 5283-5296.

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.

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.

Turner M G(2001). Landscape ecology in theory and practice: pattern and process[M]. *Springer*.

Farina A(2006). Principles and methods in landscape ecology: towards a science of the landscape[M]. *Springer*.

Lin, H., Wang, J., Bo, Y., & Bian, J. (2008). The Impacts of Landscape Patterns on the Accuracy of Remotely Sensed Data Classification. In*Proceedings of 8 th International Symposium on Special Accuracy Assessment in Natural Resources and Environmental Sciences (Shanghai, China* (pp. 219-225).