A new way of calculating the sub-pixel confusion matrix: a comparative evaluation using an artificial dataset.

Stien Heremans and Jos Van Orshoven

KU Leuven, Department of Earth and Environmental Sciences
Celestijnenlaan 200E, 3000 Leuven (Belgium)
stien.heremans@ees.kuleuven.be

Abstract

This paper introduces a new, alternative method for calculating the sub-pixel confusion matrix. It calculates the off-diagonal matrix elements from the slope coefficients of the regression relations between the overestimations and the underestimations of the area fractions. The new method is set against existing approaches for calculating the sub-pixel confusion matrix through the use of an artificial dataset. The results show that it is able to compete with the existing approaches and in some cases even outperforms the latter.

Keywords: sub-pixel classification, confusion matrix, off-diagonal elements

1. Introduction

In remotely sensed image classifications for land cover and land use (LCU), the pixel size associated with the sensor often exceeds the average size of the land patches to be classified, giving rise to a considerable fraction of mixed pixels (Fisher, 1997). Sub-pixel classification approaches are especially designed to deal with these mixed pixels which explains their increasing popularity in the remote sensing community over the last decades (Silvan-Cardenas and Wang, 2010; Verbeiren *et al*., 2008; Walton, 2008; Xu *et al*., 2005, Zurita-Milla *et al*., 2011). Despite considerable advances in sub-pixel classifiers, a widely accepted method for the evaluation of their accuracy has until now not been established (Binaghi *et al*., 1999; Chen *et al*., 2010). This lack of an adequate and well-accepted validation method not only hinders the evaluation of a single classification for a particular application but also prevents a well-founded comparison of multiple competing classification techniques (Foody, 2004).

The most common method for accuracy assessment in hard classification is the cross-tabulation matrix between the obtained classification and a reference classification, also known as the error matrix or confusion matrix (Story and Congalton, 1986; Pontius and Cheuk, 2006). In the remainder of this paper, it will consistently be referred to as the confusion matrix. The confusion matrix not only provides a detailed assessment of the agreement between the reference data and the classification result, but also describes the confusions between each pair of classes (Stehman, 1997). When pixels are soft-classified, the best way of calculating a confusion matrix is not immediately obvious (Pontius, 2002). As the location of the within-pixel boundaries between the LCU classes is unknown, there is no way to determine the actual overlap among the classes based solely on the LCU fractions (Silván-Cárdenas and Wang, 2008).

Several authors have attempted to expand the crisp confusion matrix to sub-pixel classifications by extending the notion of 'crisp matching' to that of 'soft matching'. Binaghi et al. (1999) implemented the "min" operator introduced in the original formulation of fuzzy set theory (Zadeh, 1977). Lewis and Brown (2001) on the other hand proposed the use of a multiplication ("prod") operator to define agreement and disagreement between classes within a pixel. This operator is based on conventional probability theory, where the joint probability of two events is described as the product of the individual probabilities of each of the separate events. The "min" operator calculates the maximum possible overlap between two partitions while the "prod" operator calculates the expected overlap between two partitions . A third operator, the "least" operator, calculates the minimum possible overlap between two partitions.

As Pontius and Cheuk (2006) demonstrated, the operators described above exhibit some counterintuitive characteristics. Specifically, they do not result in a diagonal matrix when a classification is compared to itself. Moreover, for the "min" and the "least' operator the sum of all matrix entries does not match the total number of pixels. Silvan-Cardenas and Wang (2008) stated that a composite operator is necessary to warrant the diagonalization characteristic and that the "min" operator is the most appropriate candidate for the diagonal cells (agreement). At the same time, they admitted that there is no unique way to allocate the remaining sub-pixel proportions into the off-diagonal cells (disagreement or confusion). Their proposed solution of this dilemma is the use of confusion intervals, bounded by the "min-least" and "min-min" operators. The resulting intervals represent the range of possible confusions between the classes and the matrix containing these intervals was called the "sub-pixel confusion-uncertainty matrix". An alternative to the interval approach is to use the "min-prod" operator, a composite operator that calculates the statistical center of the confusions.

All approaches described above incorporate some concept of spatial overlap within the pixel. As sub-pixel classifications do not provide any information on the within-pixel spatial allocation of the classes, validation methods that do not rely on this spatial concept are in our opinion to be preferred.

This study has two objectives. The first is to present an alternative method for calculating the sub-pixel confusion matrix based solely on the statistical (regression) relations between over- and underestimations. This approach does not include any presumptions about the spatial allocation of the classes within a pixel. The second objective is to assess and compare the appropriateness of different methods for calculating the sub-pixel confusion matrix, including the newly developed "statistical confusion matrix".

2. The statistical confusion matrix

As stated before, the off-diagonal elements of the statistical confusion matrix are calculated without any reference to the spatial allocation of the LCU classes within the pixels. To ensure this, the use of an intersection operator was deliberately avoided. Instead, the calculation is based on the statistical regression relations between the over- and underestimation of the LCU fractions of the individual classes. Under- and overestimation matrices were already introduced by Silvan-Cardenas and Wang (2008), but these authors implemented another method for calculating the off-diagonal elements from the over- and underestimation values. In our method, regression relations are calculated between the overestimation of one class and the underestimation of another class:

$$over_i = a_{ij} + b_{ij} \times under_j \;\; (i \neq j) \tag{1}$$

$$under_i = a'_{ij} + b'_{ij} \times over_j \;\; (i \neq j) \tag{2}$$

with i and j two different classes. The b_{ij} and b'_{ij} values are the slope coefficients associated with these relations. After the negative slopes are set equal to zero, these coefficients can be used to distribute the over- and underestimation values of the classes over the individual cells of the statistical confusion matrix. In general, confusions are perceived as

$$P_{ij} = c_{ij} \times under_j \tag{3}$$

$$P_{ij} = c'_{ij} \times over_i \tag{4}$$

with P_{ij} the entry at location ij in the confusion matrix. In our method, c_{ij} and c'_{ij} are calculated through normalization of the original slope coefficients (b_{ij} and b'_{ij}). In general, large slope values representing strong relations will result in large values for the matrix element located at the intersection of these classes. For the calculation of the diagonal elements the minimum operator was used, as recommended by Silvan-Cardenas and Wang (2008).

3. Methods

3.1. Confusion matrices

A number of existing methods for calculating sub-pixel confusion matrices were selected to reflect the advances in this area over the last decade. Details on their mathematical formulation and calculation method can be found in the references in table 1. The statistical confusion matrix that we recently developed, was also added to the list.

Table 1: Sub-pixel confusion matrices compared in this paper.

Confusion matrix	***Reference***
“min” matrix	Binaghi *et al*. (1999)
“prod” matrix	Lewis and Brown (2001)
“least” matrix	Pontius and Connors (2006)
“min-min”, “min-prod” and “min-least” matrix	Silvan-Cardenas and Wang (2008)
sub-pixel confusion-uncertainty matrix	Silvan-Cardenas and Wang (2008)
statistical confusion matrix	see section 2

3.2. Artificial confusion scenarios

All of the confusion matrices listed in table 1 were evaluated by means of an artificial dataset of 250 by 250 pixels containing area fractions for four LCU classes. Working with artificial datasets has the advantage that confusions can be induced artificially and are hence exactly known.

For each class, an artificial raster containing the reference fractions was first created. As they represent actual LCU classes, these fractions were set to be positive and to sum to one for each pixel. The reference fractions were then subjected to each of the three confusion scenarios described in table 2. The confusion values (over- and underestimations) for each pixel were randomly extracted from a Gaussian distribution and subsequently rescaled to make sure the output fractions would also be positive and sum to one. These output fractions were calculated through summation of the reference fractions and the confusion values (positive for overestimation and negative for underestimation). Also, the 'real confusion matrix' was calculated. The entries of this matrix contain the sum of the actual (induced) confusions over all pixels. All sub-pixel confusion matrices (table 1) were evaluated with respect to their ability to correctly reproduce the 'real confusion matrix'. This evaluation consists of two components: the ability to correctly identify the position of the confused classes in the matrix (i) and the ability to accurately predict the magnitude of the 'real' confusions (ii).

Table 2: Confusion scenarios imposed on the artificial dataset

Confusion scenario	*Description*
Scenario 1: Random confusion between class 1 and class 2	Class 1 is underestimated Class 2 is overestimated } equal Classes 3 and 4 are not confused
Scenario 2: Random confusion between class 1 and classes 2 and 3	Class 1 is overestimated Classes 2 and 3 are underestimated } equal Class 4 is not confused
Scenario 3: Random confusion between class 1 and class 2 and between class 3 and class 4	Class 1 is underestimated Class 2 is overestimated } equal Class 3 is underestimated Class 4 is overestimated } equal

4. Results and discussion

The sub-pixel confusion matrices from table 1 were compared with respect to their ability to correctly reproduce the position and the magnitude of the artificially induced confusion patterns. Table 3 summarizes the results for the three confusion scenarios.

Table 3: Evaluation of the sub-pixel confusion matrices. A value (+) was assigned to correctly estimated locations/magnitudes, a (-) to incorrectly estimated ones.

	scenario 1		*scenario 2*		*scenario 3*	
Confusion matrix	*location*	*magnitude*	*location*	*magnitude*	*location*	*magnitude*
"min" matrix	-	-	-	-	-	-
"prod" matrix	-	-	-	-	-	-
"least" matrix	+	-	+	-	+	-
"min-min" matrix	+	+	+	+	-	-
"min-prod" matrix	+	+	+	+	-	-
"min-least" matrix	+	+	+	+	+	-
sub-pixel confusion-uncertainty matrix	+	+	+	+	-	-
statistical confusion matrix	+	+	+	-	+	-

Scenario 1 is the most simple scenario since only two classes are confused. For this scenario, the composite matrices – "min-min", "min-prod" and "min-least" – as well as the statistical confusion matrix are able to correctly predict both the location and the magnitude of the introduced confusions. The "least" matrix correctly identifies the confused classes, but largely underestimates the magnitude of the confusion.

For scenario 2, comparable results were obtained. The composite matrices are again able to correctly identify the confusion patterns. The statistical confusion matrix correctly predicts the location, but its estimates of the confusion values are not completely accurate. The same holds for the "least" matrix, although the errors found here are much larger than for the statistical confusion matrix.

For scenario 3, the composite matrices are no longer able to correctly estimate the magnitude of the induced confusions and the "min-min" and "min-prod" matrices even fail to correctly identify the location of the confusions. They display large confusions between classes that were in reality not confused with each other at all. This can be attributed to the fact that the cross-tabulation values they predict are based solely on the concurrence of an underestimation of one class with an overestimation of another, regardless of whether this concurrence represents an actual confusion pattern between the classes concerned. Consequently, the sub-pixel confusion-uncertainty matrix displays large confusion intervals at the wrong locations. Although these intervals do include zero, they can easily lead to a misleading or even wrong interpretation of the actual confusion patterns present in a given classification.

Three matrices are able to correctly predict the location of the confusions in scenario 3: the "least" matrix, the "min-least" matrix and the statistical confusion matrix. The former two however largely underestimate the magnitude of the confusions thus encouraging an overoptimistic estimation of the classification accuracy. The statistical confusion matrix is the only one able to approximate the real confusion values to a somewhat acceptable degree of accuracy.

The results seem to imply that as soon as more than one class is underestimated and at the same time more than one class is overestimated, the newly developed statistical confusion matrix is able to outperform the existing sub-pixel confusion matrices including the sub-pixel confusion uncertainty matrix. We would also like to emphasize that we believe further refinement of the statistical confusion matrix is possible through the incorporation of uncertainty, replacing the current estimated values by accuracy intervals. We are however aware of the fact that since our confusion values are calculated from the entire population of pixels instead of per-pixel, their accuracy/reliability will largely depend on the number of pixels (samples) available in the dataset. This issue is partially resolved through the use of accuracy intervals, as the uncertainty will be reflected by the size of the intervals, but it does remain an important constraint to the use of our approach.

5. Conclusions

This paper introduces a new method for calculating the sub-pixel confusion matrix. Based on a comparison using artificial datasets, it was found to be a promising competitor to a number of existing methods including the sub-pixel confusion uncertainty matrix which is considered as the current reference for sub-pixel classification accuracy assessment. Applications on real datasets are however needed to consolidate and possibly extend our understanding of this new approach.

References

Binaghi, E., Brivio, P.A., Ghezzi, P., Rampini, A. (1999), "A fuzzy set-based accuracy assessment of soft classification". *Pattern Recognition Letters*, Vol. 20: 935-948.

Chen, J., Zhu, X., Imura, H., Chen, X. (2010), "Consistency of accuracy assessment indices for soft classification: simulation analysis". *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol.65(2): 165-164.

Fisher, P. (1997), "The pixel: a snare and a delusion". *International Journal of Remote Sensing*, Vol. 18(3): 679-685.

Foody, G.M. (2004), "Thematic map comparison: evaluating the statistical significance of differences in classification accuracy". *Photogrammetric Engineering & Remote Sensing*, Vol. 70(5): 627-633.

Lewis, H.G., Brown, M. (2001), "A generalized confusion matrix for assessing area estimates from remotely sensed data". International Journal of Remote Sensing, Vol. 22: 3223-3235.

Pontius, R.G. (2002), "Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions". Photogrammetric Engineering & Remote Sensing, Vol. 68(10): 1041-1049.

Pontius, R.G., Cheuk, M.L. (2006), "A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions". *International Journal of Geographic Information Science*, Vol. 20(1): 1-30.

Pontius, R.G. and Connors J. (2006), "Expanding the conceptual, mathematical and practical methods for map comparison". *In*: Caetano, M., Painho, M. (eds.). *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisboa, Portugal, pp. 64-79.

Silvan-Cardenas, J.L., Wang,, L. (2008), "Sub-pixel confusion-uncertainty matrix for assessing soft classifications". *Remote Sensing of Environment*, Vol. 112: 1081-1095.

Silvan-Cardenas, J.L., Wang, L. (2010), "Retrieval of subpixel Tamarix canopy cover from Landsat data along the Forgotten River using linear and nonlinear spectral mixture models". *Remote Sensing of Environment*, Vol. 114(8): 1777-1790.

Stehman, S.V. (1997), "Selecting and interpreting measures of thematic classification accuracy". *Remote Sensing of Environment*, Vol. 62(11): 77-89.

Story, M., Congalton, R.G. (1986), "Accuracy assessment: a user's *perspective". Photogrammetric Engineering & Remote Sensing*, Vol. 53(3): 397-399.

Verbeiren, S., Eerens, H., Piccard, I., Bauwens, I., Van Orshoven, J. (2008), "Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium". *International Journal of Applied Earth Observation and Geoinformation*, Vol. 10(4): 486-497.

Walton, J.T. (2008), "Subpixel urban land cover estimation: comparing cubist, random forests and support vector regression". *Photogrammetric Engineering & Remote Sensing*, Vol. 74(10): 1213-1222.

Xu, M., Watanachaturaporn, P., Varshney, P.K., Arora, M.K. (2005), "Decision tree regression for soft classification of remote sensing data". *Remote Sensing of Environment*, Vol. 97(3): 322-336.

Zadeh, L.A. (1965), 'Fuzzy sets". *Information and Control,* Vol. 8: 338-353.

Zurita-Milla, R., Gómez-Chova, L., Guanter, L., Clevers, J.G.P.W., Camps-Valls, G. (2011), "Multitemporal unmixing of medium-spatial-resolution satellite images: as case study using MERIS images for land-cover mapping". *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49(11): 4308-4317.