Spatial entropy for the measurement of the spatial accuracy of classified remote sensing imagery

Didier G. Leibovici, Giles M. Foody and Doreen Boyd
School of Geography
University of Nottingham
Nottingham, United Kingdom
didier.leibovici@nottingham.ac.uk

Abstract— **Accuracy assessment is now widely accepted as an integral part of any mapping programme from remote sensing. Accuracy assessments are usually performed on either a global basis or at class level without consideration of spatial properties. Analysing the spatial distribution of the accuracy may be valuable, especially in providing information that aids evaluation of the quality of the derived map and of the methods used in its production. The paper focuses on spatial accuracy measurements and their spatial pattern using statistics, taking into account proximities of accurate and non accurate pixel values (*e.g.* classified image *versus* a reference). A methodological framework using spatial entropy measures based on co-occurrences of the predicates (accurate and non-accurate values) is described and illustrated with a one-class classification of land cover from Landsat ETM+ data.**

Keywords: Remote Sensing; land cover; Accuracy; Spatial Pattern; Entropy; Scale

I. INTRODUCTION

In natural resource and environmental science applications it is now widely accepted that the accuracy of maps derived from remote sensing be assessed and communicated to users. This is usually done for the algorithm used and the region of interest by comparing the classified image to some "ground truth data". Often the statistics used are global statistics, for example, the root mean squares error (RMSE) if the outcome is continuous or with overall measures of thematic accuracy such as a kappa coefficient or percentage of omission and commission error if the outcome is categorical (Congalton 1994, Liu et al. 2007). Such global assessments have important limitations, notably conveying no information on the spatial distribution of classification errors. Information on the spatial nature of errors is important and tools are, therefore, required to measure and visualize this aspect of accuracy.

The Shannon entropy has been used to reveal the spatial distribution of classification uncertainty (Foody 1995, Friedl et al. 1999), but this cannot reflect spatial structures (patterns). The entropy index, as the average of uncertainties of a series of outcomes, identifies: structured distribution - rare outcomes or some very frequent ones, or unstructured distribution – all outcomes have nearly the same probability. The uniform distribution has maximum uncertainty and its entropy, $\log(n)$, is maximum over all distributions with n outcomes. In order to use this index to describe spatial distributions, Leibovici (2009) and Leibovici et al. (2009) used the distribution of multiple co-occurrences of the events of interests. A co-occurrence is observed if 2 events are observed in a given vicinity *e.g.* 2 pixels wrongly predicting class A, or, 1 pixel correctly predicting class A and 1 pixel wrongly predicting class. A multiple co-occurrence is observed if, 3 or more (the order of co-occurrence) pixels are at proximity to each other (for a given distance), each pixel with specified values of the predicates *e.g.* as above but with 3 pixels. One set of valued predicates constitutes a co-occurrence class. The chosen multiple co-occurrence at the chosen order encompasses a multivariate property of the attributes to investigate, and a spatial constraint due to the order and the distance of co-occurrence. As the co-occurrences are defined according to a neighbourhood distance, the paradigm allows a multi-scale analysis of the spatial structure of errors: micro-scale and macro-scale spatial structures where scale here has to be understood as the extent of influence.

Overall accuracy assessment, one class-accuracy as well as multi-class accuracies can be achieved using the same paradigm. Localised co-occurrences counts can be used to provide some visual insights of the spatial assessments either by a spatial plot using the same index or by a multiway analysis based on the chi-square independence measure of the multiple co-occurrences. Illustrative examples of various situations using classified remote sensing images will demonstrate the added value of the approach compared to conventional accuracy assessment methods. The paper will close with a discussion of challenges requiring further attention.

II. ACCURACY ASSESSMENT

A. General Accuracy Assessment

The standard approach to accuracy assessment involves the comparison of the class labels predicted by a classifier with those in the ground reference data for a sample of cases extracted from the image. Crosstabulating the labels yields a confusion matrix from which a range of global accuracy statistics may be derived. Commonly attention focuses on measures of overall accuracy such as the kappa coefficient or per-class measures such as user's and producer's accuracy. (Strahler et al., 2006; Congalton and Green, 2009).

Although these measures of accuracy are useful, they are aspatial in nature (see Foody, 2002 for a discussion on this issue and other accuracy assessment issues in practice).

B. Example Dataset

A Landsat 7 Enhanced Thematic Mapper Plus (ETM+) image of the test site, located in North Norfolk, UK, acquired on June 19, 2000 was provided by the U.K.'s NERC Earth Observation Data Centre. These data were geometrically corrected to the British national grid with an estimated RMS error of 0.12 of a pixel.

Based on the six non-thermal waveband, a fuzzy image classification between saltmarsh and water classes was performed using the Fuzclass procedure in the Idrisi 32 package: with training sites located near the western edge of the image extract, and choice of zero membership occurring at a distance of 2 standard deviations from the class centroid. A crisp map of the marshland class was obtained by labeling all pixels with a membership of >0.1 to marshland as members of that class. The choice of this classification procedure was made mainly because it enabled a map of saltmarsh to be generated without a need to define and include all thematic classes that occur in the test site.

Ordnance Survey MasterMap Topography layer was used to derive the areas of marshland (saltmarsh or grazing marshland) in the study area, used as the reference classification.

Figure 1 displays the image and derived products, notably a map of marshland presence and absence along with its difference from the reference (ground truth).
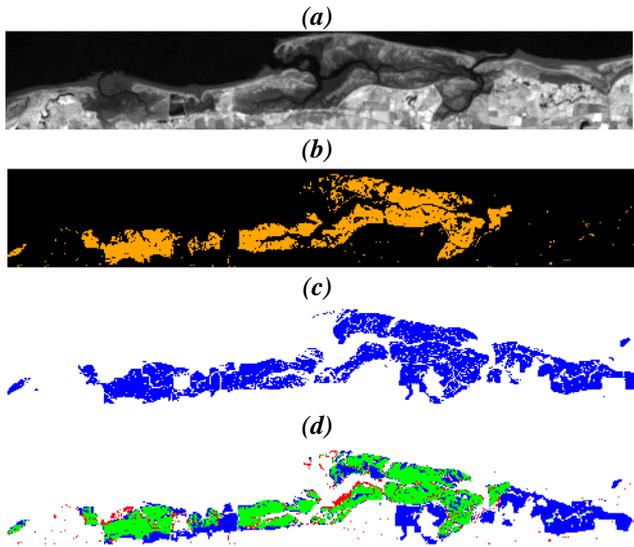
(a)

(b)

(c)

(d)

Figure 1. *(a) ETM+ imagery, (b) Map of marshland derived from the image classification, (c) Ground data on marshland reference, (d) difference between the image classification and ground reference data (classification – reference)- blue=error Marshland (eM) (or marshland omission), white=match Other (O), green=match Marshland (M), red=error Other (eO) (or marshalnd commission)*

A simple error assessment is summarized in Table 1; the good overall accuracy is the result of the high frequency of the "Land or Water" pixels. The omission of "Marshland" is large: large areas of marshland omitted, but also delineating errors of the marshlands.

TABLE I. CONFUSION MATRIX (KAPPA=57%, OVERALL ERROR=13%) ASSOCIATED WITH MARSHLAND PREDICTION OF FIGURE 1

. RS prediction	Reference		User's Accuracy
	Marshland	*Other*	
Marshland	6494	1179	**85%**
Other	5782	40417	**87%**
Omission	47%	3%	
Commission	10%	14%	
Producer's Accuracy	**53%**	**97%**	**overall 87%**

III. SPATIAL DISTRIBUTION OF ERRORS

The focus is on omission, commission and agreement for the marshland class (Fig.1), and on the spatial distribution of the corresponding labels: **eM**, **eO** and **M**. The right prediction of land or water (**O**) is not of interest here. To assess the spatial structures and proximities of these labels, Leibovici (2009) proposed to define a spatial entropy index based on the co-occurrences distribution of the labels: $p_{c_{oo}}$. The measure index depends on the distance d of co-occurrence, and on the choice of the order of co-occurrence, C_{oo}, that is the number of pixels involved in one co-occurrence of the labels involved in the multi-index c_{oo} (this defines as a multivariate co-occurrence, see also Leibovici et al.(2009)):

$$H_{Su}(C_{oo}, d) = -\frac{1}{\log(N_{C_{oo}})} \sum_{c_{oo}} p_{c_{oo},d} \log(p_{c_{oo},d}) \qquad (2)$$

Where $N_{C_{oo}}$ is the length of the table of counts of co-occurrences. Here the probabilities $p_{c_{oo},d} = p_{ijk,d}$, where $i, j,$ and k are one the labels above, reflecting the joint distribution of observing three "close" pixels with labels $i, j,$ and k. When $i, j,$ and k are restricted to be all equal (here, either to **eM** or **eO** or **M**), the spatial entropy is termed self-spatial entropy, H_{sSu}, as only co-occurrences of the same mark or label (at the given order) are looked for.

A. Exploring and visualising scale effect

Plotting the indices against distance gives some insight into the scale effect of the spatial structure of accuracies, Fig. 2. As each measure index is standardized relative to uniformity, the distribution of co-locations becomes uniform if the index value is close to 1, therefore reflecting no structure.

Theoretical curves, for a chosen hypothesis of distribution of the labels, such as Poisson processes or of their co-occurrences such as independence, provide evidence of effect or not for the observed values. Notice the lower self-spatial entropy (dotted line on Fig.2), under independence of co-occurrence, is due to more skewed distributions (at all d) with more co-location of matches (**M**) and less co-locations for the errors (**eO, eM**) under this

hypothesis. For each measure-index, an envelope of variability under the null hypothesis of, for example, random labelling can also be used to test the spatial structure. To refine the evolution of these indices, plots of probabilities of co-occurrences may be analyzed. For example for the self spatial entropy, p_{iii}, with i as **eM**, is steadily decreasing, and even more with i as **eO**, while a steady increase occurs with i as **M**.
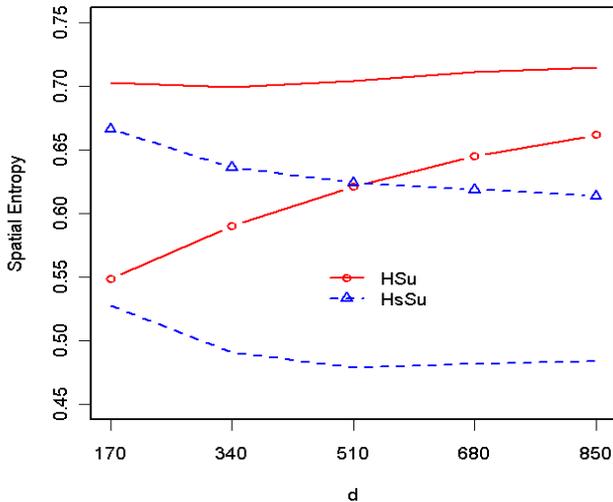

Figure 2. *Spatial entropy (H_{Su}) and self-spatial entropy (H_{sSu}) of order 3 for co-occurrences using (M), (eM), (eO): line without observation points are the theoretical curves under independence of co-occurence.*

B. Spatial pattern of errors

Localising the third order co-occurrences counts, that is those for a given pixel, leads to co-occurrence of order 2 which the entropy can be mapped to assess visually the pattern of errors and matches. Fig.3 expresses at d=340m the spatial structure of errors. The whole areas of marshland not predicted (mainly on the east of the map) appear distinctly (lowest values), as only co-occurrences of (**eM**) can be recorded,, as well the isolated pixels resulting from the "noise" in the classification image. Three areas pointed by a arrow, appear also unbalanced about the distribution of co-occurrences of errors.
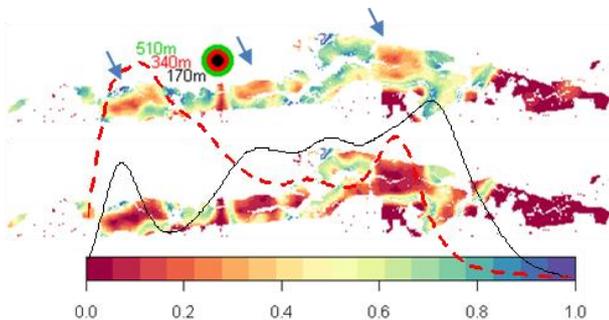

Figure 3. *Map and overall density of localised Spatial entropy from order 3 collocations at distance d=340 m (red disk): HSu(top- black) and HsSu(bottom-red)*

IV. FURTHER ISSUES AND DISCUSSION

The approach taken here is exploratory and focuses on co-occurrences of order greater than 2, but 2nd order analysis using geostatistics can be used for spatial assessment of accuracy (Kyriakidis and Zhang, 2003).While focusing on the errors (match and mismatches of pixels), the spatial entropy indices are standardized relative to uniformity, but another approach would be to compare directly the distribution of co-occurrences of the classes for each image, then a ratio of spatial entropy of the two images. Computationally this is more demanding at all the pixels of image have to be considered, whilst for the approach taken here the over represented **O** is discarded. Using a mask could be a solution. The number of classes involved using this type of analysis can be greater than 2.

The distinction of land and water was not made, and the focus was on one class: marshland. Land Use and Land Cover Changes, or LUCC (2 dates or more) can utilize the methodology described here, leading to spatial structure of change, or profile changes: each profile constituting one "class" over time.

REFERENCES

Congalton, R.G.(1994) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing thematic accuracy assessment: a compendium*, 37, 3.

Congalton, R.G. and Green, K. (2009) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd edition, CRC Press.

Foody, G.M. (1995) Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(5), 2-12.

Foody, G.M. (2002) Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185-201.

Foody, G.M. (2009) The impact of imperfect ground reference data on the accuracy of land cover change estimation. *International Journal of Remote Sensing*, 30, 3275-3281

Friedl, M.A., Brodley, C.E., and Strahler, A.H. (1999) Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 969–977

Kyriakidis, P.C. and Zhang, J. (2003) A Geostatistical Framework for Accuracy Assessment of Remotely Sensed Land-cover Information. *Geocomputation*, 2003.

Leibovici, D.G., Bastin, L., and Jackson, M. (2009) Higher Order Cooccurrences in Point Pattern Analysis and Decision Tree Clustering. StatGIS 2009, Milos Greece, June 17-19

Leibovici, D.G. (2009) Defining Spatial Entropy from Multivariate Distributions of Co-Occurrences. In: *Lecture Notes in Computer Science*: Volume 5756/2009, 392-404.

Liu, C. Frazier, P. and Kumar, L. (2007) Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, 107, pp. 606–616.

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V..and. Woodcock, C.E. (2006) *Global Land Cover Validation: Recommendations For Evaluation And Accuracy Assessment Of Global Land Cover Maps*, EUR 22156 EN – DG, Office for Official Publications of the European Communities, Luxembourg, 48 pp