CORRECTING ESTIMATES OF LAND COVER CHANGE AND CHANGE DETECTION ACCURACY FOR ERROR IN GROUND REFERENCE DATA

Giles M. Foody
School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK
(e-mail - giles.foody@nottingham.ac.uk)

Land cover change is a major component of global environmental change, notably as a cause and consequence of climate change as well as representing the greatest threat to biodiversity [1]. There is, consequently, considerable desire for accurate and up-to-date information on land cover and its dynamics. Although remote sensing has considerable potential as a source of information on land cover and land cover change there are many challenges to face in its use for accurate land cover monitoring. Indeed, the monitoring of land cover from remotely sensed data is complicated by factors ranging from semantic issues connected with the definitions of land cover classes and of change [2] through to technical issues connected with the data and information extraction methods used [3]. One common problem noted in studies of land cover change by remote sensing is a tendency to overestimate the amount of change [4]. Given the many sources of error and uncertainty in the study of land cover change by remote sensing, there is a need to know their likely impacts on analyses. One fundamental problem, however, is that error in the ground reference data used in validation activities that seek to determine change detection accuracy can be a major source of bias in estimates of change and of the perceived accuracy of change detection by remote sensing.

Although the presence of error in the ground reference data set is well-known, the magnitude of its impact may not be apparent. However, the confusion matrix used in accuracy assessment has a role that can extend beyond the description of accuracy to the provision of information to refine estimates of land cover and land cover change. Furthermore, if the ground reference data set used in forming the confusion matrix is of known quality it may be possible to remove the impact of its imperfections on the estimates derived from the matrix.

The **aim** of this presentation is to illustrate the possible impacts arising from the use of imperfect ground reference data on change detection by remote sensing and illustrate the potential to correct for ground data error in common analyses.

The presentation will use standard **methods** based on a binary confusion matrix, showing allocations to change and no-change classes, which is the common basis for the estimation of measures of change and of change detection accuracy. Attention focuses on such matrix based analyses using simulated data. This approach allows the evaluation of specific scenarios in which the effect of other error sources (e.g. due to spatial mis-registration of data sets) is avoided. Attention is focused on two properties commonly estimated from the binary confusion matrix. These are the producer's accuracy of change detection and the amount of change that has occurred; many other properties may, of course, be estimated (e.g. user's accuracy).

The presentation will introduce equations that may be used to remove the effect of known ground data error. These equations provide estimates of the producer's accuracy and change extent that have been corrected for the presence of known error in the ground reference data set.

As a guide to the **results**, one scenario considered had a remote sensing change detection technique that was 80% accurate and a ground data set that was 90% accurate. For the confusion matrix derived with this scenario, the producer's accuracy of change was estimated to be 61%, a value that differs greatly from the real situation (which is known for the simulated data). Given that the data set was generated using a classifier that actually had a producer's accuracy of 80%, an apparent 19% reduction in accuracy has arisen because the ground data set was imperfect, with a 90% accuracy.

The amount of change inferred from the matrix was also biased. In this case the bias was upwards with the perceived amount of change being 26% while the actual amount was 20%.

The results show that a large bias may be introduced into estimates of land cover change related variables, even if the data sets used were of high accuracy. The presentation will contain other scenarios for illustrative purposes and consider additional features such as the impact of correlated errors.

The **conclusions** will stress that ground reference data error can result in systematic bias in estimates of change occurrence and change detection accuracy. The examples presented will highlight substantial bias, with the producer's accuracy of change systematically underestimated and the amount of change systematically overestimated. Moreover, the magnitude of the bias introduced may be very large even if the ground data set is of a very high accuracy. Fortunately, it is sometimes possible to correct for the systematic biases caused error in the ground reference data and provide refined estimates of land cover change and of the accuracy of change detection by remote sensing.

References

1. Sala, O. E. *et al*. (2000) Global biodiversity scenarios for the year 2100, *Science*, **287**, 1770-74.
2. Fritz, S. and See, L. (2008) Identifying and quantifying uncertainty and spatial disagreement in the comparison of global land cover for different applications, *Global Change Biology*, **14**, 1057-75.
3. Lu, D. and Weng, Q. (2007) A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing*, **28**, 823-70.
4. van Oort, P. A. J. (2005) Improving land cover change estimates by accounting for classification errors, *International Journal of Remote Sensing*, **26**, 3009-24.