Geoinformation for European-wide Integration, Benes (ed.)

A rigorous protocol for post-classification land cover and land use change detection

Xavier Pons
Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Universitat Autònoma de Barcelona, Spain

Pere Serra & David Saurí
Departament de Geografia, Universitat Autònoma de Barcelona, Spain

Keywords: land-cover and land-use changes, overlay, post-classification, RMS, image erosion

ABSTRACT: This paper presents a protocol for rigorous accuracy assessment of land-cover and land-use changes between two dates (1977-1993) through the overlay of two independent classifications (post-classification method). Although postclasification overlay is a usual method, there are only a few works considering those factors that can distort results; these factors are thematic accuracy, spatial misregistration, fragmentation of the landscape, pixel size and grid origin. The methodology is applied over an area located at the NE of the Iberian Peninsula. Results clearly show that without correcting these factors the thematic accuracy of the change map would be only 43.9%, although the thematic accuracy of the maps to be overlaid is quite high (about 90%).

1 INTRODUCTION

Satellite remote sensing, because of its temporal resolution, provides an excellent historical framework for estimating the spatial extent of land cover and land use (LCLU) changes. Using satellite images, different types of LCLU changes have been monitored, for instance in urban development (Dai and Khorram 1998), in agricultural crop rotation (Congalton et al. 1998), in forest fire mapping (Salvador et al. 2000) or in deforestation assessment (Mertens and Lambin 1997).

There are two basic approaches for LCLU change detection (Singh 1989):

1. Comparative analysis of independently produced classifications from different dates (post-classification comparison: map-to-map comparison).
2. Simultaneous analysis of multitemporal data (multidate classification and others: image-to-image comparison).

Both approaches have advantages and disadvantages, but the most common method is the post-classification comparison (Congalton and Macleod, 1994). Moreover, this approach allows using legends more detailed than with the second approach; therefore, it was the chosen approach.

The main sources of uncertainty of this method are:

1. Misregistration of the polygon boundaries (locational inaccuracy) in the different classifications and, therefore, the presence of border pixels with false positive or negative changes.
2. Problems derived from classification errors: a false positive change may be recorded when no change has taken place because a polygon in one or both of the two maps is misclassified, or false negative changes, when no change is identified but a change has taken place.
3. Furthermore, this approach requires very good accuracy in both classifications because the accuracy of the change map is the product of the accuracies of the individual classifications (Singh 1989, Lambin and Strahler 1994).
4. Moreover, if the images used for each classification are from different seasons of the year, the comparison can be more difficult, especially for some legend items due to vegetation phenology.

Although it can be thought that the result of multiplying the accuracies of each individual classification could suffice, this would be only true if no planimetric error existed in either of the two layers. Indeed, all misregistration problems will decrease the accuracy because they introduce false positive or negative changes (Townshend *et al.* 1998).

When remotely sensed data come from different sensors, for instance MSS with SPOT in Jensen *et al.* (1995), or MSS with TM in Lodhi *et al.* (1998), some extra problems appear for both approaches:

1. Different pixel size affects the classification because some elements are not detected in the

coarser resolution images which do appear in the finer resolution images.
2. Overlaying for LCLU change analysis is complicated by different pixel size and/or grid origin, producing an extra problem to the misregistration caused by the different geometric corrections of the different images.
3. The number of bands and their wavelengths (spectral information) is different (Fung 1992), as well as the sensitivity of the sensors, this factor being more critical in the second approach.

Although working with different sensors is not ideal, it is often unavoidable. For example, in time series analysis, one of the sensors may not have existed at the earlier dates, or it stopped collecting data for technical or political reasons; in other cases, we simply do not have any other data available.

The general objective of our research was to detect and compare LCLU changes between 1977 and 1993 from Landsat image data of a Mediterranean agricultural plain. Unfortunately, comparison is usually made by a simple crossing of the results regardless of the implications of the factors discussed above, crucial to understanding and quantifying their effects and, especially, for properly interpreting the results of the LCLU comparison.

Thus, the specific objective of this paper is to discuss the implications of the usage of different sensors in the comparison of classified images and to propose a protocol for this type of situation.

A secondary objective is to discuss which classifier is more suitable, given that the accuracy of each classification is very important in LCLU change detection.

2 MATERIAL AND METHODS

2.1 *Study area and materials*

The study area covers 30,170 ha and includes 22 municipalities located in the Alt (Upper) Empordà Plain (in the northeast of Spain), with the following UTM 31-North zone coordinates: 489 990, 515 010, 4 660 410 and 4 689 990. The area does not exceed 100 metres above sea level and is replenished with neogen and quaternary sediments. The landscape is drained by two primary river systems: the Muga and the Fluvià.

Traditionally, this plain has specialised in herbaceous crops, mainly cereals and fodder, located in fragmented parcels and often with the same crop sown at different dates through the year. The area of irrigated fields has increased since the mid 1960s due to the Boadella reservoir project, the canalisation of the Muga River and the intensive exploitation of the aquifer around the Fluvià river mouth.

The spectral response of many cover types varies throughout the year: categories that appear very similar in spring may become distinguishable at earlier or later stages of the annual cycle. For this reason three multispectral images were used for each of the periods considered: three Landsat MSS images for the 1970s (17 July 1977, 2 June 1978 and 18 September 1978; the 4 spectral bands were considered), and three Landsat TM images for the 1990s (16 May 1992, 28 June 1993 and 31 August 1993; thermal bands were not used).

2.2 *Methodology*

2.2.1 *Geometric and radiometric corrections*

The first step was the geometric correction using the procedure developed by Palà and Pons (1995). During the geometric correction, MSS images were resampled to 60 m x 60 m while TM images were resampled to 30 m x 30 m (nominal resolution for MSS images is 79 m x 57 m (Campbell 1996), while for TM images it is 30 m x 30 m). In both cases the resampling method was the nearest neighbour to preserve the original image radiometry. Georeferencing was done using a mean of 26 Ground Control Points (GCPs) per image. The accuracy of the georeferencing was assessed through the root mean square (RMS) of the location of independent test GCPs (a mean of 14 GCPs per image). In our case, the MSS images had a RMS error of about 0.9 pixels while in the TM images the error was about 0.7 pixels.

The second step was the radiometric correction, through which digital numbers were converted into reflectance values using the sensor calibration parameters and other factors such as atmospheric effects, solar incident angle accounting for the relief, etc. (Pons and Solé-Sugrañes, 1994). The resultant corrected images presented a coherent range of reflectance values.

2.2.2 *Legend*

As mentioned before, herbaceous crops are predominant in the study area. According to our field experience and to agricultural studies (Ministerio de Agricultura, Pesca y Alimentación 1982; Pujol 1985), we defined the following categories: in the case of 1977, eleven LCLU categories were established: dry and irrigated maize, other dry herbaceous with fallow land, other irrigated herbaceous, fruit trees, olive trees, vineyards, meadows and pastures, woodlands and shrublands, uncultivated pastured lands, unproductive lands (quarries, etc.), urban surfaces (villages, etc.) and rivers and lagoons. For 1993 we had the same classes as in 1977 plus two new categories: rice, which was reintroduced in 1985, and dry and irrigated sunflower, which was favoured by the Common Agriculture Policy subsidies since 1986.

2.2.3 *Classification*

Traditionally, classification strategies have been divided into two broad categories: supervised and unsupervised. The supervised approach involves the selection of areas on the image which statistically characterise the informational categories of interest, while the unsupervised approach attempts to identify spectrally homogenous groups within the image that are later assigned to informational categories (Richards 1993, Chuvieco 1996). A third category would be the hybrid classification approaches (Estes *et al.* 1983, Townshend 1992).

The most commonly applied supervised classification method is the maximum likelihood procedure because of its robustness; nevertheless, it has the underlying assumption of a normal (Gaussian) distribution of the data within each class. If a class is multimodal, the modelling is not very effective (Richards, 1993). In our research this method was not considered adequate because crops did not follow normal distributions due to the different stages of growth in the different fields covered by the training areas (different crop development) and water availability (dry and irrigated).

In the conventional procedure of the unsupervised classification, spectral classes of pixels are first identified by cluster analysis. ISODATA (Interactive Self Organizing Data Analysis) is a non-hierarchical clustering algorithm commonly used in remote sensing (Richards, 1993). Once the clusters are obtained, 'rules of correspondence' between the spectral and the LCLU categories are established; these rules are normally known through fieldwork or ancillary information (ground data). The standard procedure of unsupervised classification is based on the assumption that each spectral class corresponds to one and only one LCLU category and vice-versa, but this does not always work because there are different possible patterns of correspondence (Lark 1995):

1. *One spectral class corresponds to one LCLU category*. This is the ideal situation but it is often not the case due to the reasons explained above.
2. *Several spectral classes correspond to one LCLU category* or, equivalently, an LCLU category is formed by several spectral classes. Usually this situation does not cause problems if using appropriate classification techniques taking advantage of the fact that, as a result of an unsupervised classification, we can have more spectral classes than LCLU categories.
3. *One spectral class corresponds to more than one LCLU category*. For example, the response of a bare soil can sometimes be allocated to a recently harvested field or to unproductive land. This is the most problematic situation and it is caused because an LCLU category is indistinguishable from another using the available spectral data (from all dates), or because the spectral classes are too broad. In this case, the LCLU categories may, perhaps, be distinguished if a larger number of narrower spectral classes are sought.

In our case, the classification was performed by means of a hybrid classifier, which may be able to deal with most of the situations mentioned above. This classifier combines two modules of the MiraMon software (Pons 2000): ISODATA and CLSMIX. The procedure both involves unsupervised classification and training areas (collected as in the first stage of a conventional supervised classification). The unsupervised classification uses the ISODATA algorithm, and then the CLSMIX module assigns the spectral classes (obtained by the ISODATA algorithm) to LCLU categories by using the training areas. As the ISODATA module allows a large number of spectral classes to be found, the third situation described above is avoided in most cases (we worked with 84 final classes for the 1970s and with 98 final classes for the 1990s). We did not run ISODATA with the original images, but rather with the first principal components resulting from a principal component analysis performed on each period: 12 MSS bands were input for the 1970s (we used the first four PCs, explaining 93.9% of the variance) and 18 TM bands were input for the 1990s (we used the first six PCs, explaining 96.1% of the variance).

In order to achieve an accurate classification, perhaps the most important part of the process lies in the CLSMIX module. As input parameters the module needs:

1. The image resulting from the unsupervised classifier.
2. The training areas.
3. The threshold proportion at which to accept a spectral class as being a part of an LCLU category in terms of the proportion of the spectral class that is inside the LCLU category. For example, 0.6 will mean that if 60% or more of the spectral class is inside the training areas of a given LCLU category, then this spectral class will be assigned to this LCLU category.
4. The threshold proportion at which to accept a spectral class as being a part of an LCLU category in terms of the proportion of the LCLU category that is formed by a given spectral class. For example, 0.01 will mean that if 1% or more of the LCLU category is formed by a given spectral class, this spectral class will be assigned to the LCLU category.

The required thresholds (points 3 and 4 above) must be obtained empirically, but since the execution of CLSMIX is extremely fast (as it only performs comparisons), they can be easily adjusted after some iteration, especially if independent ground data is available, to converge to an optimal solution.

Note that when classifying a given pixel, the module chooses the category that has the more 'reasonable' assignation:

1. Spatial correspondence between the spectral class and the training areas of that LCLU category (the spectral class is inside the training area).
2. The spectral class is mainly inside this LCLU category (an important proportion of the spectral class belongs to the category).
3. The spectral class is a not insignificant part of the LCLU category.

Conversely, a pixel will remain unclassified if no training area covers pixels in the same spectral class or if, given the input thresholds, no spectral class is adequate for it: that is either the pixel belongs to a class that is split too much between two or more LCLU categories (no clear LCLU tendency of the spectral class) or the pixel belongs to a class that is poorly representative of the total area of any LCLU category (perhaps the spectral class is noisy).

In our case, the two required thresholds were 42% and 1% for MSS and 36% and 1% for TM.

Finally, note that unsupervised classifications usually present two main problems:

1. the number of clusters to choose is critical and, if very large, it is hard to do the final assignation to LCLU categories (Richards 1993, Chuvieco 1996), and
2. it is necessary to manually define the rules of correspondence between clusters and LCLU categories (Lark 1995).

These two problems are not present in this hybrid classification because it is an automatic and objective process (simply choose a large number of clusters). Of course, in this case it is necessary to define training areas, but some ground data knowledge is always needed (even in conventional unsupervised classification) and we consider that the time devoted to digitising these training areas is compensated by the objectivity of the assignations.

3 PROTOCOL FOR REALISTIC ACCURACY ASSESSMENT OF LCLU CHANGES

As has been discussed above, several aspects should be considered in order to obtain a rigorous accuracy assessment of LCLU changes from a pair of LCLU maps of different dates obtained from different sensors. The proposed protocol should take into account the following three main issues:

1. To assess the overall accuracy resulting from the overlay, the overall accuracies of the two maps must be multiplied. In our case we decided to accept an accuracy of from 75% upwards for the final product of the two overlaid classifications (but higher for each individual classification).
2. Erode the boundaries of the polygons to avoid comparing areas with locational inaccuracy. Assuming that the calculated accuracy in point 1 is satisfactory, when the two maps to be compared are overlaid, the area around the polygon boundaries is not reliable because the maps have some planimetric error. Those areas where some error exists should not be used for LCLU change detection, because false positive or negative changes could result. Moreover, the number of mixed pixels (heterogeneous) increases substantially along polygon boundaries and thus severe spectral confusion may lead to classification errors, contributing also to this uncertainty (Campbell 1996). The solution is to eliminate these uncertain areas, which can be done, in raster mode, by applying an erode procedure.
   To apply the erosion it is necessary to define some parameter that indicates the area to erode. As all the images usually have the RMS error computed in the geometric correction stage and this parameter has the statistical meaning of the standard deviation of the errors (assuming they are normal), the RMS could be used in this sense.
   Nevertheless, the RMS is not the only parameter to take into account. The degree of fragmentation also heavily affects the number of points that present an erroneous location in polygon-based maps. Indeed, because this inaccuracy is related to the boundaries of polygons (patches), but not their inner parts, a very homogeneous landscape (large patches) is less affected than a very fragmented landscape. We carried out tests with some images corresponding to areas of different degrees of fragmentation and simulating several RMS errors (from less than 0.5 pixels to 2 pixels) (Figure 1). In each case, the images were mis-registered with a random function of mean 0 and standard deviation equal to the RMS, and then overlaid with the original to calculate the number of pixels erroneously located. When an image has a low RMS error (less than 0.5 pixels) and a low fragmentation, it is not necessary to erode because 99.2% of the points remain in the same location. Conversely, on an image in a very fragmented landscape with an RMS of 1 pixel, only 56.1% of the points remain in the same location, while 94.1% remain in the same location when an erosion of 1 pixel is applied. These results indicate that a conservative approach for medium to highly fragmented landscapes is to erode 1 pixel in order to avoid comparison artefacts. In our images, with an RMS of nearly 1 pixel and a quite fragmented landscape, eroding 1 pixel guaranteed, for the MSS images, that

97.9% of the points were correctly located while, if no erosion were applied, only 66.5% of the points would have been reliable (i.e., 33.5% of points would have been outside their corresponding polygon). It is important to note that these figures affect the overall accuracy resulting from multiplying the accuracy of each layer. For example, using an erosion of 1 pixel on the MSS classification only slightly decreases the final accuracy, but overlaying without erosion would reduce its accuracy by a factor of 0.665.
Finally, we note that the erosion process can also improve the comparison because it tends to eliminate some of those pixels that may be more difficult to classify (mixed pixels).

3. Resample the two layers accounting for the different pixel size and grid origin. This should take into account modal aspects of the resampling window.

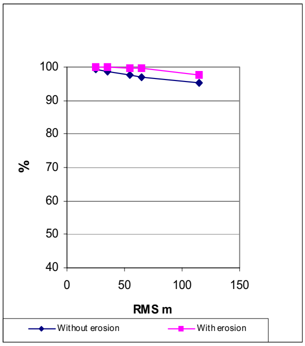
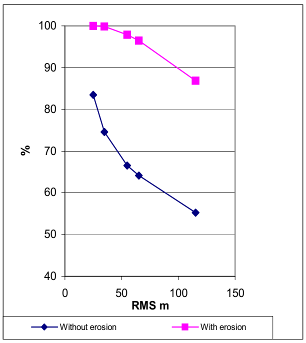
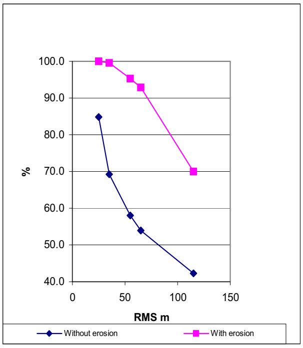


Figure 1. Percentage of pixels well located according to different geometric errors in the classifications (RMS in pixels) before and after eroding one pixel, in landscapes with low fragmentation (above), quite fragmented (below) and very fragmented (right), respectively.





4 RESULTS AND DISCUSSION

The test of the two final maps (1970s and 1990s) was performed by means of new, independent, training areas (not the same areas used to run clsmix) considered as ground data (also identified from field work, aerial photos and orthophotos). For the 1970s, the overall accuracy was 91.8% and, for the 1990s, it was 95.2%. According to our results, the product of the accuracies obtained from the independent training areas was 87.4%. It is important to note that 87.4% is not the final accuracy of the LCLU change analysis, this value must be decreased by the factor given by the locational inaccuracy of both layers, as we show below.

In our case we decided to erode 1 pixel all around each polygon. Given the RMS of our images and the degree of fragmentation of our landscape, eroding 1 pixel on the MSS images guaranteed that the 97.9% of the points were correctly located, while on the TM images this figure reached 99.5% (without erosion the figures are 65.8% and 76.3% respectively). These results indicate that eroding 1 pixel on each layer gives a final accuracy for the LCLU change analysis of 87.4%*0.979*0.995 = 85.1%. Note that, although the combined accuracy of 87.4% appears to be sufficient (and several authors suggest this indicator), if erosion is not applied the real resulting accuracy for the LCLU change analysis is 87.4%*0.658*0.763 = 43.9%.

After the erosion, 85.8% of the study area became nodata in the 1970s classified image and 73.6% in the 1990s classified image. Due to the spatial fragmentation present in our area, these are significant proportions but they permit a rigorous comparison (avoiding misregistration problems) between the two maps. Indeed, it is important to point out that not eroding of the polygon boundaries leads to very poor results when comparing the two classified images (43.9%, or less since the boundary pixels are often the most difficult to classify). It is also worth noting that our case is an extreme one (high landscape fragmentation, especially regarding the pixel size in the case of MSS) and that most users would erode a smaller fraction of their images. In other words, some users might be tempted not to use this methodology to avoid area losses, but they would risk reducing the reliability of the results of their comparisons.

5 CONCLUSIONS

This paper uses a post-classification comparison for LCLU change detection. This methodology requires each of the classifications to have a high accuracy, a goal not always reached when a legend with several agricultural categories is needed. In addition, it becomes more difficult in fragmented landscapes like our study area. The results of the hybrid classification method have been successful, solving the problem of choosing the number of clusters and the patterns of correspondence between spectral classes and LCLU categories, and giving a high degree of classification accuracy.

In this work we have emphasised the need, when carrying out LCLU change analyses, to take into account the different classification accuracies, fragmentation of the landscape, planimetric accuracies, pixel sizes and grid origins. The proposed protocol has been applied without a significant increase of effort and the results are more reliable than a direct overlay. The drawback of this method is that it reduces the useful area of comparison, in our case substantially. However, it should be noted that when directly overlaying two classified images with an RMS of about 1 pixel in quite fragmented landscapes, the amount of noisy results (false positives and negatives) can be critical for the interpretation of the outcomes (more than 30% of the information can be unreliable). For instance, studies may find a change of 10% deduced from an overlay, but probably this will be mainly due to problems in the boundaries of the polygons. With the protocol proposed in the present paper, the comparison of LCLU is based on a sample, but a sample taken from the more reliable part of the polygons (the inner part). From our point of view the choice is clear: renouncing part of the data produces conclusions that are far more reliable.

Although it may be advisable to avoid mixing sensors and spatial resolutions, currently, and even more in the future, the problem of overlaying remotely sensed data from different sources with a historical perspective will increase due to the availability of new sensors with higher spatial resolutions at 15, 10, 5, 1 metre and beyond; hence the importance of establishing protocols for LCLU change assessment.

REFERENCES

Campbell, J.B. 1996. Introduction to Remote Sensing, 2nd edn. (London: Taylor & Francis Ltd).

Congalton, R.G., Balogh, M., Bell, C., Green, K.., Milliken, J.A., Ottman, R. 1998. Mapping and monitoring agricultural crops and other land cover in the lower Colorado river basin. *Photogrammetric Engineering and Remote Sensing*, 64, 1107-1113.

Congalton, R.G., and Macleod, R.D., 1994. Change detection accuracy assessment of the NOAA Chesapeake bay pilot study. International Symposium on the spatial accuracy of natural resource data bases, 16 May 1994 (Virginia: American Society for Photogrammetry and Remote Sensing), pp. 78-87.

Chuvieco, E. 1996. *Fundamentos de Teledetección espacial*. Madrid, Rialp.

Dai, X., Khorram, S. 1998. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1566-1577.

Estes, J.E., Hajic, E.J., Tinney, l. 1983. Fundamentals of image analysis: analysis of visible and thermal infrared data. In: *Manual of Remote Sensing*, Colwell, R.N (Ed.). Virginia, American Society of Photogrammetry, 987-1124.

Fung, T. 1992. Land use land cover change detection with Landsat MSS and SPOT HRV data in Hong Kong. *Geocarto International*, 3, 33-40.

Jensen, J.R., Rutchey, K., Koch, M.S. y Narumalani, S. 1995. Inland wetland change detection in the everglades water conservation area 2A using a time series of normalized remotely sensed data. *Photogrammetric Engineering and Remote Sensing*. 61: 199-209.

Lambin, E.F., and Strahler, A.H. 1994. Change vector analysis in multitemporal space: a tool to detect and categorize land-cover change processes using high temporal resolution satellite data. *Remote Sensing of Environment*, 48, 231-244.

Lodhi, M., Echavarria, F.R. y Keithley, C. 1998. Using remote sensing data to monitor land cover changes near Afghan refugee camps in Northern Pakistan. *Geocarto International*. 13: 33-39.

Mertens, B. y Lambin, E.F. 1997. Spatial modelling of deforestation in Southern Cameroon. *Applied Geography*. 17: 143-162.

Ministerio de Agricultura, Pesca y Alimentación, 1982. Calendario de siembra, floración y recolección. Madrid.

Palà, V. y Pons, X. 1995. Incorporation of relief in polynomial-based geometric corrections. *Photogrammetric Engineering and Remote Sensing*. 61: 935-944.

Pons, X., Sole-Sugrañes, Ll. 1994. A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. *Remote Sensing of Environment*. 48: 191-204.

Pons, X. 2000. *MiraMon. Geographic Information System and Remote Sensing software.* Centre de Recerca Ecològica i Aplicacions Forestals, CREAF. Bellaterra. http://www.creaf.uab.es/miramon

Pujol, M., 1985. Els cereals: Generalitats, 2nd edition (Barcelona: Escola Universitària d'Enginyeria Tècnica Agrícola i d'Especialitats Agropequàries de Barcelona).

Richards, J.A. 1993. *Remote sensing digital image analysis.* Berlin, Springer-Verlag.

Salvador, R., Pons, X., Valeriano, J., and Diaz-Delgado, R. 2000. A semi-automatic methodology to detect fire scars in shrubs and evergreen forests with Landsat MSS time series. *International Journal of Remote Sensing*, 21, 655-671.

Singh, A. 1989. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*. 10: 989-1003.