Accuracy Assessment of The Classified Landsat TM Satellite Imagery Data for Aried and Semiarid Areas

Dr. Tarek Elaswed
Associate professor
Department of Geography, Faculty of Arts, University of Zawia
T.elaswed@zu.edu.ly

Abstract

Depending on the availability of land cover and land use (LCLU) data and their importance in studying the impacting changes in environmental and climatic systems, and as these data provide opportunities to increase scientific research in the environmental field at the landscape level. Reports of accuracy for this data can be high and acceptable, but at the same time they are untrue and misleading. From this point of view, one of the first concerns of the remote sensing community has become to improve the quality of data and the methodology for extracting land cover information and land uses, in addition to the advantages provided by satellite methodologies, there are limitations that must be realistically measured to be made clear to users of this data so that they can make correct decisions about it and the possibility of its use. Accuracy assessment of these products is the procedure used to measure the quality of these products. Using remote sensing techniques to detect the changes during the period 1988 till 2000 using Landsat TM5 dates (1988, 1992, 1996, and 2000). Also using the different kind of maps which integrated with the remote sensing data to find the relationship between the changes in the land cover in the study area, west of Tripoli at Lon (12: 33:18 - 13:21:47) and Lat (32:55:10 - 32:35:44). Supervisor classification carried out using Maximum likelihood method chosen to classify the images. High resolution data such as Quick Bird (2002) and Spot 5 (2000) have been used as reference to choose the training sets and to apply the accuracy assessment for the classification results. The accuracy assessment has been applied was between 67% and 76%, obtained by using the high-resolution data as reference.

Keywords: Land use; Land cover; Maximum likelihood; Accuracy assessment; classification

تقييم دقة بيانات صور الأقمار الصناعية المصنفة من لاندسات TM للمناطق الجافة وشبه الجافة

د. طارق المختار الاسود
استاذ مشارك بقسم الجغرافيا/ كلية الآداب/ جامعة الزاوية
T.elaswed@zu.edu.ly

الملخص

تبعاً لتوفر بيانات الغطاء الأرضي واستعمالات الاراضي (LCLU) وأهميتها في دراسة التغيرات المؤثرة في النظم البيئية والمناخية، وحيث أن هذه البيانات توفر فرص لزيادة البحث العلمي في المجال البيئي على مستوى المناظر الطبيعية. تقارير الدقة بالنسبة لهذه البيانات يمكن ان تكون عالية ومقبولة ولكنها في نفس الوقت تكون غير حقيقية ومضللة. ومن هذا المنطلق اصبح من أولى اهتمامات مجتمع الاستشعار عن بعد هي تحسين جودة البيانات ومنهجية استخلاص معلومات الغطاء الارضي واستعمالات الاراضي، بالإضافة إلى المزايا التي توفرها منهجيات الاقمار الاصطناعية فهناك قيود يجب قياسها بشكل واقعي لتوضيحها للمستخدمين لهذه البيانات حتى يتمكنوا من اتخاذ قرارات صحيحة بشأنها وامكانية استخدامها. وتقييم الدقة لهذه المنتجات هو الإجراء المستخدم لقياس جودة هذه المنتجات. استخدام تقنيات الاستشعار عن بعد لكشف التغيرات خلال الفترة من 1988 حتى 2000 باستخدام مرئيات القمر الاصطناعي لاندسات TM5 للسنوات (1988 و 1992 و 1996 و 2000). أيضا باستخدام أنواع من الخرائط المختلفة بالتكامل مع بيانات الاستشعار عن بعد لإيجاد العلاقة بين التغيرات في الغطاء الأرضي في منطقة الدراسة الواقعة بين خطي طول ودائرتي عرض Lon (12: 33:18 -13:21:47) and Lat (32:35:44 - 32:55:10). تم اختيار تقنية التصنيف الموجه باستخدام طريقة احتمالية القصوى المختارة (Maximum likelihood) (ML) لتصنيف المرئيات. مرئيات عالية الدقة مثل Quick Bird (2002) و Spot 5 (2000) استخدمت كمرجع لاختيار مجموعات التدريب ولتطبيق تقييم واختبار الدقة لنتائج التصنيف. تم تطبيق تقييم الدقة والحصول على نتيجة ما بين 67 % و 76 % والتي تم الحصول عليها باستخدام البيانات عالية الدقة كمرجع.

الكلمات المفتاحية: استخدامات الأراضي، الغطاء الأرضي، احتمالية قصوى، تقييم الدقة، تصنيف.

1. Introduction

When using satellite images and classifying them to produce maps of Landcover and Landuse (LCLU) by conducting supervised classification, collecting a set of different classes and training samples is used to establish classification rules and multi-class boundaries in the feature space. The training sample data used in the Maximum Likelihood (ML) also provide extra information such as the shape of the distribution of the members of each class as well as the location of the center of each cluster; therefore, the resulting classification might be expected to yield a more accurate result than those produced by the other statistical supervised classifiers (Mather, 2004).

The ML algorithm considers the relative likelihood of overlapping pixels using the training data as a means of estimating class variances and also using the variability of brightness of each class to maximize the probability of correct classification (Campbell, 2006). The algorithm was used to identify LCLU changes, especially in vegetation cover (agriculture activities), classifying Landsat TM5 data of the study area during the period from 1988 to 2000, using four images acquired in 1988, 1992, 1996 and 2000. Nine different classes were collected from each image represented in class1: Other trees (OT) (Olive, Palm, Almonds), class2: Citrus fruits (CF) (Orange and Lemon), class3: Annual Crops (AC) (Cereal, alfalfa, market-gardening, etc), class4: Urban areas (UA), class5: Pasture land with natural vegetation (PLNV), class6: Sand dunes, sand covered areas and drifted sand (SD), class7: Forest, reforestation (F), class8: Sea (S) and class9: Bare rocks (BR). High spatial resolution data (Quick Bird 2002, Spot5 2000) and Spot XS 1987) and existing land use map were used to choose the samples to classify the images, and were also used to assess the accuracy of the classification.

Although accuracy assessment is a vital component in any study involving LCLU classification, which is being used increasingly to produce thematic land cover maps (Foody, 2002 and Boschetti *et al*., 2004). Since spectral similarity of some classes and the complexity of bounders between them in the classification process might be one of the sources of the error and that the basic challenge of the accuracy assessment (Powell *et al*., 2004). Maps provided from remote sensing are often judged with reference data and found to be of insufficient quality for operational applications (Foody, 2002 and Latifovic *et al*., 2004). Accuracy is usually based on an evaluation of the classified images with a reference data set and the dissimilarities between the two data sets are typically interpreted as errors in the derived land cover map (Stehman, 1997a and Foody, 2002).

For assessment of classification accuracies different classification accuracy of the reference samples are then summarized in a confusion matrix and performance of the LCLU classification. Analysing the critical assumption of the classification accuracy that the confusion matrix essentially representative of the classification results of the entire study area (Cheng *et al*., 2019). All accuracies or errors are characteristically associated with uncertainties due to variability or uncertainty in selection of training and reference samples (Weber and Langille, 2007).

2. **Classification Accuracy Assessmen**t

A Stratified Random Sampling technique was applied in order to produce the accuracy of the classified images. Many remote sensing analysts prefer this method (Jensen, 2005), in which a minimum number of samples are selected from each class after the thematic map has been prepared. Stratified random sampling involves two steps. First, the study area is classified into land cover classes on what is found in the remote sensing classification. Sample locations are then randomly distributed throughout an existing land use map (CEDEX, Land use map 1981), and a high-resolution images Quick Bird (2002) and Spot 5 (2000). Points were randomly created using the Accuracy Assessment Package in ERDAS to all classified images, only selected confidence points were used in the statistic accuracy analysis (Kappa Analysis). These points have been indicated by applying two rules:

- Selected classes – only points for selected classes were chosen,

- Confidence point – the point should belong clearly to one class.

In the assessment of the accuracy, to produce the report for the classified images, it needs to be compared with an existing land use map, and the high-resolution images that was considered a clear to discover the features, on the other hand the doublecheck during the field work was followed. When the images had been classified, ground survey was done to ensure that the classes, which were mapped effectively, correspond to the thematic classes they were supposed to be.

To produce accuracy statistic of classified images in this study, Error Matrices Analysis and Kappa Analysis (K_{hat}) were applied to define overall accuracy and a K_{hat} value. Short explanations of these methods are shown below:

The most common and typical method used by researchers to assess classification accuracy is with the use of an error matrix (Congalton, 1991). An error matrix is a square assortment of numbers defined in rows and columns that represent the number of sample units (i.e., pixels, clusters of pixels, or polygons) assigned to a particular category relative to the actual category as confirmed on the ground. The rows in the matrix represent the remote sensing derived land use map (i.e., Landsat data), while the columns represent the reference data (i.e., aerial photo) (Jensen, 1996). The error matrix was applied to produce overall accuracy for the classified images this study. The overall accuracy of the classification map is determined by dividing the total number of correct pixels (sum of the major diagonal) by the number of pixels in the error matrix (N).

These tables produce many statistical measures of thematic accuracy including overall classification accuracy (the sum of the diagonal elements divided by the total number), KAPPA analysis yields a K_{hat} statistic (an estimate of KAPPA) that is a measure of agreement or accuracy between the remote sensing-derived classification map. The K_{hat} statistic is computed as below:

$$K_{hat} = \frac{N\sum^{rows} x_{ii} - \sum^{rows} x_{class}\, x_{ref}}{N^2 - \sum^{rows} x_{class}\, x_{ref}}$$

Where, N = is the total number of observations
X_{ii} = are the observations along the diagonal
X_{class}=are the observations for classified data
X_{ref} = are the observations for reference data

As the results of the classification which locate the change on the land cover. Samples (pixels) for each image were selected randomly for comparison with the same samples in the reference data. Firstly, to assess the classified image of the 2000 Landsat TM5, the samples were evaluated with the same points in a Spot 5 image from 2000 and Quick Bird image from 2002. The confusion matrix Table (1), illustrate the overlap between the classes.

Table (1). Confusion matrix of the classification accuracy of 2000

	OT	**CF**	**AC**	**U**	**PLNV**	**SD**	**F**	**S**	**BR**	**Users Accuracy (%)**
OT	**126**	1	59	2	31	7	0	0	0	56
CF	13	**121**	39	0	4	0	3	0	0	67
AC	37	8	**149**	1	4	5	1	0	0	73
U	0	0	0	**19**	3	1	0	0	0	83
PLNV	13	0	27	1	**163**	1	0	0	0	80
SD	0	0	0	0	1	**20**	0	0	0	95
F	0	1	1	0	0	0	**12**	0	0	86
S	0	0	0	0	0	0	0	**22**	0	100
BR	0	0	0	0	3	2	0	0	**16**	76
Producers Accuracy %	67	92	54	83	78	56	75	100	100	

Secondly, 1996 Landsat TM5 image classification was compared with the Spot 5 image from 2000 Table (2), shows the confusion matrix of the accuracy assessment.

Table (2). Confusion matrix of the classification accuracy of 1996

	OT	CF	AC	U	PLNV	SD	F	S	BR	Users Accuracy (%)
OT	**184**	8	76	1	11	0	4	0	0	65
CF	24	**176**	44	0	0	0	3	0	0	71
AC	1	1	**56**	0	1	0	0	0	0	95
U	0	0	0	**13**	0	0	0	0	0	100
PLNV	13	1	9	0	**174**	1	2	0	0	87
SD	0	0	0	0	0	**11**	0	0	0	100
F	0	0	0	0	0	0	**8**	0	0	100
S	0	0	0	0	0	0	0	**14**	1	93
BR	0	0	0	0	0	0	0	0	**9**	100
Producers Accuracy %	83	95	30	93	94	92	47	100	90	

Thirdly, because there is no high spatial resolution data as valid to assess 1992 Landsat TM5 image classification. Therefore 1996 classified image was compared with the classified 1996 Landsat TM5 image and the 2000 Spot 5 image to test the accuracy the confusion matrix as shown in Table (3).

Table (3). Confusion matrix of the classification accuracy of 1992

	OT	CF	AC	U	PLNV	SD	F	S	BR	Users Accuracy (%)
OT	**155**	7	80	0	9	0	5	0	0	61
CF	38	**130**	61	0	4	0	5	0	0	55
AC	0	1	**65**	0	0	0	0	0	0	98
U	0	0	0	**17**	1	3	0	0	0	81
PLNV	15	0	4	1	**155**	0	2	0	0	87
SD	1	0	0	1	0	**47**	0	0	0	96
F	1	0	0	0	1	0	**6**	0	0	75
S	0	0	0	0	0	0	0	**26**	0	100
BR	0	0	0	1	0	3	0	0	**6**	60
Producers Accuracy%	74	94	31	85	91	89	33	100	100	

Finally, for the same reason of the absence of data as reference to test the accuracy of the result of the classification image of 1988, hence 1987 Spot XS image, with a spatial resolution of 20 m and assisted by Spot 5 image of 2000 to assess the accuracy of the classified 1988 Landsat TM5 and the confusion matrix was as shown in Table (4). Clearly it would have been preferable to have had independent land cover data for each date with which to assess the accuracy of the classified images but such information was not available and the above comparisons deemed the closest evaluations (in terms of timeliness of data acquisition) on this accession.

Table (4). Confusion matrix of the classification accuracy of 1988

	OT	CF	AC	U	PLNV	SD	F	S	BR	Users Accuracy(%)
OT	**198**	8	134	4	15	2	7	0	0	54
CF	22	**153**	57	1	0	0	3	0	0	65
AC	0	0	**61**	0	0	0	1	0	0	98
U	0	0	1	**12**	0	0	0	1	0	86
PLNV	16	2	12	1	**147**	1	1	0	0	82
SD	0	0	0	0	0	**10**	0	0	0	100
F	1	0	0	0	0	0	**2**	0	0	67
S	0	0	0	0	0	0	0	**18**	0	100
BR	0	0	0	6	0	3	0	2	**13**	54
Producers Accuracy%	84	94	23	50	91	63	14	86	100	

3. Results and Analysis

As shown in previous section, the error matrix of the accuracy assessments of classified satellite images, the most confusion between the interesting vegetation classes (OT, CF, AC and PLNV). The spectral similarity of the classes is one of the most causes when the training samples were selected. On the other hand, the spaces between the lines of trees which sometimes more than 20 meter are using to grow the annual crops or natural vegetation might be grow naturally as other reasons to make the confusing of the classes. The number of the test points between the classified image and the reference data and the availability of the valid data as a reference to test the classified image, might be one of the factors which affect the percentage of the accuracy assessment results. Table (5) illustrate the overall accuracy and Kappa statistic of the classification, as shown the overall accuracy of image 2000 and 1996 were higher than image 1988 because the high spatial resolution (reference) was captured in 2000 and 2002 and that make it easy to test the classification.

Table (5). Summary of Landsat classification accuracy (%) for 1988, 1992, 1996 and 2000

Land cove classes	**1988**	**1992**	**1996**	**2000**
Overall accuracy	67.03%	71.24%	76.24%	70.67%
Kappa statistic	0.5848	0.6470	0.6952	0 .6335

Overall accuracy for each classified image is comparatively good with all of them indicating more than 67%. The highest accuracy map was 76.24%, for year 1996 and the lowest was 67.03%for 1988. K_{hat} statistic for all maps given values from 0.5848 to 0.6952, meaning moderate agreement between all classified images produced in this study compared with ground survey data.

4. Conclusion

The results demonstrate that ML supervised classification of Landsat TM-5 imagery can be used to produce accurate maps and statistics referring to land cover change. On the other hand, data with high spatial resolution such as Quick Bird and Spot 5 were useful to select samples to classify to classify Landsat TM-5 imagery and to assess the accuracy of the classification results. The confusion matrix is the simplest descriptive statistic used to compare a classification result with ground truth information. "...This accuracy measure indicates the probability of a reference pixel being correctly classified and is really a measure of omission error. It is difficult to have complete confidence in the accuracy measures for the earlier images as the reference data are not contemporary. Also, the accuracies are probably related to issues of training data selection, since it was more difficult to distinguish and select pure training areas in the earlier images because of a lack of independent reference data for training set selection. However, in general the resulting accuracy appears consistent with other studies that have attempted to classify land cover in semiarid areas and so deemed acceptable for further analysis. The results suggest that ML can be used to map land cover in this study, but errors persist and overall accuracies are not necessarily as high as they could be, e.g., Kappa accuracies described as 'good' rather than 'excellent'. Hence there is a need to investigate an alternative image classification method to either improve or at least validate the patterns in land cover observed. The accuracy of the classification depends on many issues; (i) Data availability; (ii) Quality of the data to be classified; (iii) The validity of the data used as reference and the gap in time between the classified images and the validation data; (iv) The similarity of some land cover classes making them difficult to separate. Whilst the accuracy of the classification was generally between 67% and 76%, this was based upon a pragmatic rather than an ideal approach to accuracy assessment, relying on only a limited set of available validation data. In addition, the ML algorithm is also prone to a number of influences that can affect the accuracy of the outputs, e.g., mixed pixel and atmospheric effects (Foody, 2002).

5. References

- Boschetti, L., Flasse, S. P., & Brivio, P. A. (2004). Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: *The Pareto Boundary*. *Remote Sensing of Environment*, 91, 280–292.
- Campbell, J. B., 2006. Introduction to remote sensing, 4rd Ed. London: Taylor and Francis, New York.
- Cheng, K. S., Ling, J.Y., Lin, T. W., Liu, Y. T., Shen, Y. C., Kono, Y. A new thinking of LULC classification accuracy assessment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-2/W13, 2019.
- Congalton, R. G. (1991); A Review of Accessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, 37:35-46.
- Foody, M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185-201.
- Latifovic, R. Olthof., I. (2004). Accuracy assessment using sub-pixel fractional error matrices derived from satellite data *Remote Sensing of Environment*, 90, 153-165.
- Mather, M. (2004), *Computer processing of remotely-sensed images*, Wiley, Chichester. 3rd Edition.
- Stehman, S. V. (1997a). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62, 77-89.
- Powell, L. R., Matzke, N., de Souza Jr, C., Clark, M., Numata, I., Hess, L. L., Roberts, A. D. (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote sensing of environment*, 90, 221-234.
- Weber, K.T., Langille, J., 2007. Improving classification accuracy assessments with statistical bootstrap resampling techniques. *GIScience & Remote Sensing*, 44, pp. 237–250.