# Comparison of three supervised classification methods for deriving land cover maps from ASTER satellite images

J. S. Borges, A. R. S. Marçal, J. F. Pinto da Costa

***Abstract*— This work compares three classification methods (K-Nearest Neighbours, Logistic Discrimination and Support Vector Machines) for the production of land cover maps from multi-spectral images from the ASTER sensor. Data dimensionality reduction methods were also tested. The classification results were evaluated by a cross validation method and by an independent "ground truth" test set.**

***Index Terms*— Land cover classification, K-Nearest Neighbours, Logistic Discrimination, Support Vector Machines, Dimensionality Reduction, Ground truth validation.**

## I. Introduction

HIGH spatial resolution satellite sensors such as Landsat TM and ASTER are a privileged data source for the production of land cover maps. Data from these sensors has recently become available free of charge, or with very small costs, allowing for frequent updates in land cover mapping to be carried out. However, the production of accurate and meaningful land cover datasets is not a straight forward task. First of all, there is no single optimum classification method. Another problem is the evaluation of the results obtained by classification. Considerable field data is required in order to train and validate the classification process, but the human resources are usually limited and often there is not enough data to properly built the classifier and perform subsequent validation. The main objective of this work was to compare three classification methods for the production of land cover map: K-Nearest Neighbours (K-NN), Logistic Discrimination (LD) and Support Vector Machines (SVM). An evaluation of the cross validation method to estimate the classification error is carried out, comparing the results obtained by this method with an independent "ground truth" test set.

Manuscript received June 29, 2004. The authors wish to thank C.I.C.G.E. for the financial support through the POCTI FEDER program.

J. S. Borges and A. R. S. Marçal are with the Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, PORTUGAL (phone: 351-220100840; fax: 351-220100809; e-mail: jsborges@fc.up.pt and andre.marcal@fc.up.pt).

J. F. Pinto da Costa is with LIACC and the Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, PORTUGAL (e-mail: jpcosta@liacc.up.pt).

## II. Classification Procedures

Classification is the process of assigning a label (class) to an observation. Supervised classification methods require prior knowledge of the classes present in the data and the identification of training areas. Three supervised classification methods were tested, using a training set to built the classifiers, which are briefly presented in the following sections. Assuming the training set of size n as pairs of the form $(\mathbf{x}_i,\mathbf{y}_i)$, i=1,…,n, where $\mathbf{x}_i \in R^p$ is a p-dimensional vector and $\mathbf{y}_i \in C$, where C is the set of previously defined classes. The objective is then, given a new observation $\mathbf{x}_i$, to label it using the knowledge from the training set.

### A. K-Nearest Neighbours classifier

The K-Nearest Neighbours (K-NN) classifier is a well known non-parametric method that, although simple, is quite powerful. Given a new observation $\mathbf{x}_0$ to be classified, one must identify the nearest neighbours. The notion of proximity implies the existence of a metric. Usually the Euclidean distance is used to measure the proximity of $\mathbf{x}_0$ to the training objects. The class of $\mathbf{x}_0$ is determined as the most frequently on the set of the K nearest neighbours of $\mathbf{x}_0$ [1].

### B. Logistic Discrimination

The classification using Logistic Discrimination (LD) is based on Bayes rule, i.e., the class **c** assigned to the new observation $\mathbf{x}_0$ is the one that maximizes the probability $P(C=\mathbf{c}|X=\mathbf{x})$. These posterior probabilities are estimated using linear functions in **x**. To achieve the maximizing probability, the coefficients are determined by maximum likelihood method [2].

### C. Support Vector Machines

The Support Vector Machines (SVM) classifier is based on statistical learning theory proposed by Vapnik and Chervonenkis [3]. This classification procedure is an extension to the optimum margin algorithm.

Consider a two class classification problem, $\mathbf{y}\in\{-1,1\}$. The optimum margin classifier seeks the linear boundary defined by the hyperplane $w\cdot\mathbf{x}+b=0$ that maximizes the margin between the two classes. The solution is given by

$$f(\mathbf{x}) = sign\left(\sum_{i=1,\ldots,n} a_i \mathbf{y}_i (\mathbf{x}_i \cdot \mathbf{x}) + b\right). \quad (1)$$

In the case of having a non-linear separable binary problem, a slack variable $\xi$ is introduced in order to penalise observations falling on the "wrong" side of the hyperplane. The extension of this method to SVM is made by mapping the input space, into another of higher dimension (eventually infinite) – the feature space. This mapping is done using non-linear functions $f$. The objective of mapping the initial space into a higher dimension space is to spread out the data in a way that facilitates the finding of a linear hyperplane. So, after replacing $\mathbf{x}$ by its mapping $f(\mathbf{x})$ in the feature space, the separating hyperplane is given by

$$\sum_{i=1,\ldots,n} a_i \mathbf{y}_i (f(\mathbf{x}_i) \cdot f(\mathbf{x})) + b = 0 \ . \quad (2)$$

Using a kernel function defined as $K(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) \cdot f(\mathbf{y})$, the linear boundary on the feature space corresponds to a non-linear boundary on the initial space. In this way, the solution is given by

$$f(\mathbf{x}) = sign\left(\sum_{i=1,\ldots,n} a_i \mathbf{y}_i K(\mathbf{x}_i \cdot \mathbf{x}) + b\right). \quad (3)$$

For each classification problem, the user must choose the kernel as well as the associated parameters. The most commonly used kernels on SVM for classification are the Gaussian, and the Polynomial kernels.

The SVM classifier was described as a binary classification problem, thus when dealing with a multi-class problem, one needs to apply appropriate methods, such as one-against-one or one-against-rest method. As suggested by Platt [4] the one-against-one approach is the best method, and was therefore used in this work.

### D. Data reduction

One common stage in data processing is dimensionality reduction. The objective of this process is to reduce significantly the computational complexity while preserving most of the information. The most widely used algorithm to reduce the data dimensionality is Principal Component Analysis (PCA). Mathematically, the problem consists in finding a subspace of the original space that maximizes the dispersion of the points projected on that subspace [1]. Another method to reduce the dimensionality of the input space is the Linear Discriminant Analysis (LDA). The process searches for the linear combination that maximises the dispersion between the classes and at the same time minimises the dispersion within the classes of the projected points [1]. The variables resulting from this procedure are called canonical variates. While PCA examines the difference between all variables, LDA examines the structure between the different classes. So it is not always possible to perform LDA since the information about the classes is not always available.

### E. Validation

After performing the classification, it is important to evaluate the quality of the results. The ideal process is to have an independent set of test data. Unfortunately this kind of information is rarely available. An alternative form is to split the training data into two sets: one to be used for training and the other for validation. This introduces the problem of reducing the amount of information available for the training stage. There are alternative ways to avoid this problem, such as using a cross-validation strategy. The method consists on randomly splitting the training data into a number of subsets (e.g. 10). Each of the subsets is left out to be used as validation data, using the remaining subsets to train the classifier. The process is repeated for each of the subsets as validation. The final error is estimated by the average of the individual errors (e.g. average of 10 error estimates). Cross-validation should in principle be a valid indicator of the accuracy provided by the classifier, as long as the training data set is big enough.

## III. THE DATA

The test area for this work was the Vale do Sousa (river Sousa valley) region, in the North-West of Portugal, with an area of 764 km$^2$ (about 300 square miles). About 70% of the region is forested or uncultivated. The uncultivated areas are mainly due to numerous forest fires. The land use is generally small private fields (4ha or less) with the forest areas in much larger portions of the land.

### A. Data reduction

Two image granules were acquired consecutively, at 11:42, 24 October 2001 by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER).

The ASTER multi-spectral images have 14 bands: 3 in the visible and near infrared (VNIR instrument, 0.52-0.86 μm), 6 in the short-wave infrared (SWIR instrument, 1.60-2.43 μm) and 5 in the thermal infrared (TIR instrument, 8.12-11.65 μm). The pixel size is 15x15 meters (VNIR), 30x30 meters (SWIR) and 90x90 meters (TIR) [5]. Only the VNIR and SWIR image bands were used, as the TIR data has much lower spatial resolution and should not provide useful information about the vegetation cover [6].

Due the morphology of the interest area, the two ASTER images were ortho-rectified using ground control points and a Digital Elevation Model (DEM). The two ortho-rectified images were combined into a single multi-channel image file of 2060 by 3340 pixels, of 15x15 meters (30.9 by 50.1 km), covering the whole of the Vale do Sousa region.

### B. Data pre-processing

The segmentation of an image allows for groups of pixels to be considered as a single unit, or object, within the image. The image segmentation was performed using the eCognition software [7]. This multi-resolution segmentation is a bottom-up region-merging technique starting with one-pixel objects. In numerous subsequent steps, smaller image objects are merged into bigger ones. Throughout this clustering process, pairs of adjacent image objects are merged according with several

adjustable criteria of homogeneity or heterogeneity in colour and shape. This procedure generates a segmented image where the average size of the objects depends on the scale parameters previously defined.

### C. Training sets

The classification requirement was to produce land cover maps following the Portuguese National Forest Inventory [8]. A total of 9 classes were considered, 6 main classes – Urban/Residential areas (SC), Water (HH), Burned areas (FG), Forest (Fl), Uncultivated/Non-productive(IC), Agricultural areas (AG); and 4 subclasses of Fl – Eucalyptus (FlEc), Mixed (FlMix), Hardwoods (FlFd), Pinus (FlPb).

The selection of training sites was done in two stages. Firstly a land cover vector dataset produced in 1995 by air-photo interpretation was used. The polygons from this dataset that overlaid the segmented image objects by at least 75% were selected as candidates. The second stage was to carry out field surveys (in 2002) to inspect the land cover type of some of those objects. A training set with a total of 582 objects was identified by this process. The distribution of these training objects amongst the 9 land cover classes was not uniform, as it can be seen in Table I. This was due to the uneven distribution of the land cover types over the study area.

### D. Validation Sites

A post procedure to inspect the accuracy of the classifiers used, consisted in the identification of the land cover class of a set of locations independently selected. A rectangular grid with intervals of 800 meters was overlaid in two areas of the Vale do Sousa region, providing a total 277 sites for validation. The land cover of each of these sites was obtained by field surveys carried out in 2003. The validation data was collected after the classification process and was therefore a completely independent process. The number of objects in this independent dataset is shown in Table II.

As it also happened on the training set, the validation set does not have an uniform distribution of all the classes. This will be reflected negatively on the evaluation of the classifiers.

## IV. Data Classification

The data used to train the classifiers was exported form eCognition to R software [9], resulting on a data-frame of 582 objects described by 19 variables: average and standard deviation of ASTER bands 1 to 9 plus the objects density. This was the dataset used for classification and for the accuracy rate estimation. The cross-validation method was used for the K-NN, LD and SVM classifiers, by partitioning the training dataset on 10 subgroups, each with 10% of the data. The classifiers were build with 9 of these subgroups, leaving the remaining subgroup for testing. This was carried out 10 times, changing the subgroup used for testing.

TABLE I
Number of objects for the training areas of each land cover class

| **Class** | SC | HH | FG | FlEc | Flmix | FlFd | FlPb | IC | AG |
|---|---|---|---|---|---|---|---|---|---|
| **No. Objects** | 222 | 13 | 65 | 35 | 70 | 12 | 16 | 58 | 91 |

TABLE II
Number of objects for the Test areas of each land cover class

| **Class** | SC | HH | FG | FlEc | Flmix | FlFd | FlPb | IC | AG |
|---|---|---|---|---|---|---|---|---|---|
| **No. Objects** | 36 | 1 | 3 | 79 | 88 | 4 | 15 | 20 | 31 |

TABLE III
Overall accuracy measured on the train dataset

| | **K-NN** | **LD** | **SVM** |
|---|---|---|---|
| **Original** | 79.7% | 82.8% | 81.8% |
| **PCA** | 76.3% | 79.4% | 79.0% |
| **LDA** | 84.0% | 84.0% | 85.4% |

### A. Learning process

The performance of three classifiers was evaluated on the original data and on two transformations of the data: (i) using the principal components (that explained at least 95% of the variance of the data); (ii) using the canonical variates. This gave rise to 9 different classification results.

The overall accuracy of each classification result is presented in Table III. The results achieved with the LD and SVM classifier, are better than with the K-NN classifier. A first inspection of these values seems to suggest that the classifiers perform poorly. However, it is worth pointing out that the classification is performed on image objects, most of them having a significant level of mixture between land cover types. Another point to be retained is that we are trying to distinguish classes with very similar spectral response, particularly the forest sub-classes. So, the results can in fact be considered quite good. The application of data compression methods, should in theory improve the results. However, it was observed that the use of LDA improved significantly the accuracy of the results but the PCA did not.

The advantages of each classifier can be evaluated by comparing their performances in two different stages: (i) the learning process and (ii) the classification process. While the LD classifier performed best on the learning process, the SVM classifier is faster in the classification process. The advantages of the LD classifier were the accuracy, together with the facility to built the classifier. Although the SVM and LD classifiers performed similarly, the process of building the SVM classifier was a hard task. This was because the choice of correct kernel parameters conducted to an extensive search on the parameter space. Nevertheless, it is worth noticing that once the best classifier is achieved, the classification process is quite easy to perform and the results are quite good. If we want to evaluate the data compression together with the classifiers,

LDA and SVM is surely the best option. The K-NN method is easy to implement, but it has the drawback of requiring information from all of the object in memory each time it classifies a new object.

### B. *Independent Validation*

After having used the 582 objects to train and test the different classifiers and data reduction methods, the classifiers were also tested using an independent dataset. Once the classifiers were built, the one that provided best accuracy was used for the test with the independent dataset (277 objects). For the K-NN classifier a k=4 was selected and the SVM classifier was implemented with a gaussian kernel.

TABLE IV
OVERALL ACCURACY EVALUATED BY THE INDEPENDENT DATASET

| | K-NN | LD | SVM |
|---|---|---|---|
| **Cross-validation** | 84.0% | 84.0% | 85.4% |
| **Independent Validation** | 60.9% | 72.2% | 70.5% |

The results presented on Table IV give the overall accuracy of the classification methods estimated by cross-validation and using the independent data. The classification evaluation by the independent dataset was performed using LDA because this method produced the best results by cross-validation.

The classification accuracy evaluated from the independent dataset was worst than from the cross-validation. The accuracy differences varied from 11.8% (LD) to 23.1% (K-NN). This result was somehow expected, since the independent dataset was obtained from a completely new survey, carried out about one year after the identification of the training sites and almost two years after the satellite image acquisition. Another possible reason has to do with the uneven distribution of the training and validation sites between the classes on both datasets (see Tables I and II). The fact that the relative incidence of each land cover class in the training set is different from the test set reduces the overall accuracy. The cross-validation indicated that the main difficulty of all classifiers was to discriminate between the forest subclasses. This is also perceptible from the independent validation analysis. Nevertheless, the SVM and LD classifiers achieved the best results in the task of discriminating between the forest subclasses. The level of accuracy of the remaining classes was much better, particularly the land cover classes Water (HH) and Agricultural (AG). It is worth noticing that the relative accuracy between the different classifiers was maintained in the independent test set, even with an unbalanced distribution of training data. This shows that independently of the composition of training and the test sets used, the SVM and LD classifiers performed better than K-NN, in this multi-spectral image. The use of LDA resulted in better classification results as indicated in both evaluation methods.

## V. CONCLUSIONS

Two of the classification methods used are well known: the K-Nearest Neighbours (K-NN) and the Logistic Discrimination (LD). Competing with these two methods was the Support Vector Machine (SVM) classifier, a relatively new method that proved to be quite effective. Data dimensionality reduction methods were also used, and the classifiers performance on the reduced datasets was tested. The classifiers performed much better with data transformed by Linear Discriminant Analysis (LDA) than with Principal Component Analysis (PCA). Initially, due the lack of training data, a cross-validation estimation of the classification error was made. At a later stage, when an independent validation dataset was made available, the classifiers were further evaluated. This independent validation confirmed the results obtained by cross-validation, in terms of the relative performance of the classifiers, although the accuracy values were considerably lower.

## ACKNOWLEDGMENT

The authors wish to thank Alexandre Gomes and the ASTER ERSDAC for providing the satellite data through an Announcement of Research Opportunity (ARO-070).

## REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*,2nd ed.. John Wiley & Sons, 2001.
[2] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
[3] V. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
[4] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. "Large Margin DAGs for multiclass classification," *Advances in Neural Network Information Processing Systems*, vol. 12, pp. 547-553, 2000.
[5] Y. Yamaguchi, A. B. Kahle, H. Tsu, T. Kawakami, and M. Pniel, "Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 36, pp. 1062-1071, 1998.
[6] A. R. S. Marçal, J. S. Borges, J. A. Gomes, and J. P. Costa, "Land cover update by supervised classification of segmented ASTER images," *International Journal of Remote Sensing*, to be published.
[7] *eCognition Object Oriented Image Analysis User Guide*, Definiens Imaging, München, Germany, 2001.
[8] M. Tomé, B. Fidalgo, J. Gaspar, "Normas de fotointerpretação para as áreas de demonstração da Associação Florestal do Vale do Sousa," Relatórios Técnicos do GIMREF, No.1/2002, Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Lisboa, Portugal, 2002.
[9] The R Project for Statistical Computing, Available: http://www.r-project.org