MERIS BASED LAND COVER CHARACTERIZATION: A COMPARATIVE STUDY

Hugo Carrão, PhD. Student
Mário Caetano, Head of Remote Sensing Unit
Portuguese Geographic Institute
Rua Artilharia Um, 107
1099-052 Lisboa, Portugal
hugo.carrao@igeo.pt
mario.caetano@igeo.pt

Paulo Gonçalves, Research associate
INRIA Rhône-Alpes
ZIRST, 655 avenue de l'Europe,
38 334 St. Ismier, Cedex. France
paulo.goncalves@inria.fr

ABSTRACT

Knowledge of land cover spatial distribution is important for many activities, including environmental monitoring, land planning, and resource management. Land cover information at the appropriate time and over large geographical regions can only be derived from satellite images. With the recent launch of MERIS, a wide range of new possibilities for the periodic land cover characterization at regional scale is available. This sensor offers a combination of innovative features, such as high spectral and temporal resolutions, wide geographical coverage and improved atmospheric correction. We believe that the exploitation of data obtained by this new sensor fills previous technological gaps, improving automatic land cover classes' discrimination. At the same time, the extra spectral information provided by MERIS can introduce some difficulties on land cover characterization, since the dimensionality of feature space can be overloaded. In this paper we report the comparison of Support Vector Machine (SVM), a new supervised classification system that is independent of feature space dimensionality, with most frequently used supervised classifiers for the land cover classification of high dimensional MERIS imagery. To reduce feature space dimensionality and filter out redundant information a Fisher's Discriminant Analysis (FDA) was performed for spectral data rearrangement. The original and rearranged/reduced data sets were then used as inputs for the several classifiers. The results of the classification allow identifying the efficiency of SVM and FDA to improve land cover classification of high dimensional imagery data. The study was carried out with MERIS Full Resolution data from 2004 for the continental Portuguese territory.

INTRODUCTION

Land cover cartography production is an essential task for scientific research and earth science applications, such as landscape management and biodiversity assessment, as well as, in general, to support environmental, social, and economic policies. Therefore, mapping landscape regimes that are persistently changing under the influence of several factors (e.g., forest fires, clear cuts, and urban fabric growth) must be a regular process. Land cover maps become rapidly out of date, while the need for updated intermediate-scale cartography is constantly increasing to help near real-time spatial decision-making. Remote sensing data are attractive for deriving land cover cartography by means of digital image classification, because updated information for continuous time periods and over extensive geographical regions can only be derived from satellite systems.

Visual interpretation and manual classification of remote sensing data is reliable and accurate, but cannot assure any longer the production in time of land cover knowledge that is required within many applications, given that is time consuming and economically expensive. To cope with this situation, several supervised and unsupervised algorithms have been developed and successively implemented for land cover classification of multi-spectral satellite imagery (Shah et *al.*, 2003). Even though the automatic exploitation of satellite data for land cover cartography production has proved to be cost-effective, some difficulties associated with imagery characteristics and classification specificities still hamper their effective employment (Gonçalves *et* al., 2005): (1) different land covers share similar spectral signatures and cannot always be automatically separated using classical remote sensing data; (2) conventional image classification techniques fail at separating a large number of land cover classes with acceptable misclassification rates.

Recently launched Earth Observation (EO) sensors, such as the MEdium Resolution Imaging Spectrometer (MERIS) and the Moderate Resolution Imaging Spectroradiometer (MODIS), exhibit enhanced spectral and temporal resolutions and offer new potentials and challenges to data analysis. Jackson and Landgrebe (2001) affirm that the availability of a large number of spectral bands makes it possible to identify more detailed land cover classes with higher accuracy than would be possible with the data from earlier sensors. Quite as well, several studies have proved the advantages of performing a land cover classification based on multi-temporal satellite imagery data (e.g., Gonçalves *et* al., 2005; Price *et* al., 2002; Maxwell *et* al., 2002; Vieira *et* al., 2001). However, the exploitation of spectral and temporal characteristics of MERIS imagery data, for the improvement of land cover spatial and thematic characterization, is still of confidential use in remote sensing community. Presently, few research studies and operational programs are focused on the analysis of MERIS data for land cover cartography production (e.g., Clevers *et* al., 2004; Gessner *et* al., 2004; Santos *et* al., 2005).

The introduction of these new generation sensors resulted in a large increase of remote sensing data volumes available for land cover classification purposes. Still, high dimensional remote sensing imagery provides a challenge to the current classification techniques. Existing algorithms often fail to deliver high classification accuracies and tend to suffer from the problem of the “curse of dimensionality”, many times referred to as the Hughes phenomenon (Hughes, 1968). Although it is intuitive to think that increasing the dimension of the features should never reduce the classification performance, since we are providing a larger amount of information, the performance can in fact decrease while we feed more data to the system (Guyon and Elissee, 2003). A large number of land cover classes of interest and a large number of spectral bands available require numerous training samples for classification task that not always are available, because are expensive and tedious to acquire (Jackson and Landgrebe, 2001). Thus, the development of new methods to analyze high dimensional satellite imagery data became extremely necessary (Haertel and Landgrebe, 1999; Landgrebe, 2002; Shah et *al.*, 2003). As a result, the dimension reduction of satellite imagery data, via a pre-processing transformation technique that maps data from a high order dimension to a low order dimension, is required. The goal of dimension reduction, by feature extraction or selection, is to obtain a few features that discriminate the necessary classes with a high degree of accuracy (Guyon and Elissee, 2003). Principal Component Analysis (PCA) and Fisher’s Discriminant Analysis (FDA) are dimension reduction techniques that could substantially increase classifiers performance by filtering out redundant information. This pre-classification procedure must improve classification accuracy of high dimensional imagery by means of frequently used classifiers.

A new supervised classification system based on the statistical learning theory (Vapnik, 1998), defined as Support Vector Machine (SVM), has recently been applied to the problem of high dimensional remote sensing data classification (Marçal *et* al., 2005; Mercier and Lennon, 2003; Huang et al., 2002; Zhu and Blumberg, 2002). This technique is said to be independent of the dimensionality of feature space as the main idea behind this classification technique is to separate the classes with a surface that maximizes the margin between them (Pal and Mather, 2005). SVM can obtain a better classification performance and have a more generalization capacity than classifiers that aim to minimize the training error rate alone, such as neural network classifiers (Shah et *al.*, 2003). Besides, a small number of training samples is enough to train the algorithm, which reveals interesting properties for high dimensional image processing (Mercier and Lennon, 2003).

In this paper we exploit several supervised pixel based techniques for the land cover classification of MERIS high dimensional imagery data. This is a preliminary study in the framework of an on going research work that aims at developing a systematic classification methodology for the regular production of land cover cartography from medium spatial resolution satellite imagery. Specifically, the main goal of this study is to assess the utility of SVM for the land cover classification of MERIS high dimensional imagery data in comparison with most frequently used kNN, K-means and LDA classifiers. Since these current classifiers are not well adapted to the classification of high dimensional data, we perform a dimension reduction of the original spectral information with Fisher’s Discriminant Analysis and in addition appraise the accuracy of the several classification methods with the compressed information. To completely evaluate the ability of the high spectral resolution MERIS data for land cover characterization, we seek for the separation between 19 land cover classes by means of the several classifiers. We experimentally evaluate the performance of each classifier on a MERIS image of Portugal Continental for the date of August 14th, 2004.

STUDY AREA AND DATA SET

The study area is the entire Portuguese mainland territory. Portugal is in a transition zone featuring diverse landscapes representing both Mediterranean and Atlantic climate environments. This landscape heterogeneity allows for the extrapolation of the developed methodologies and of the obtained results to many other regions of the world.

MERIS data

The satellite imagery used in this study was acquired by the MEdium Resolution Imaging Spectrometer (MERIS). This sensor is on board of ENVISAT, an advanced polar-orbiting Earth Observation (EO) satellite launched by the European Space Agency (ESA) in 2002. MERIS is a 68.5° field-of-view push-broom imaging spectrometer that covers a swath width of 1,150km and allows a global coverage of the Earth in 3 days. This sensor measures the solar radiation reflected by the Earth at a maximum ground spatial resolution of 300m, in 15 spectral bands or groups of wavelengths, ranging from the visible to the near infrared (390 nm to 1040 nm) radiance, and provides the most radiometrically accurate data on Earth surface that is currently acquired from space (Curran and Steele, 2005). One of the most outstanding features of MERIS is the programmability of its spectral bands in their width and position.

MERIS data can be provided at 3 different levels of processing - Level 0, Level 1b, and Level 2 - and at 3 different spatial resolutions – Full (FR), Reduced (RR) and Low (LR). FR products have a spatial resolution of 300m, RR of 1200m and LR of 4800m. The Level 0 product corresponds to the data in its most raw format and will not generally be available to the end users. Level 1b is annotated engineering data that has been calibrated radiometrically and geolocated – this is the Top-Of-the-Atmosphere (TOA) calibrated radiance measured in mWm-2sr-1nm-1. Level 2 products contain both geolocated geophysical parameters and surface radiance/reflectance depending on the surface type. In addition to the core geophysical data, MERIS Level 2 products contain geometric information, data describing the sun and viewing geometry, terrain height, some meteorological data and several flags that address the quality and the validity of the image.

In this study we exploit the Level 2 Full Resolution MERIS imagery. These data consists of calibrated surface reflectances in 13 groups of wavelengths (original bands 11 and 15 were removed from this product since they address O_2 content and water vapor) that are atmospherically corrected for Rayleigh scattering, ozone, water vapor absorption and aerosol content. We make use of a single image that was acquired by 14 August of 2004. This date was chosen due to several criteria. First, and according to Gonçalves *et al*. (2005), this is the period of the year that allows for a better separation among general land cover classes in this region of transitional climate; secondly, this image is the one that presents the minor cloud coverage among a set of 12 MERIS images acquired for Portugal for the year of 2004.

The MERIS original data was exported from raw format and converted to GeoTIFF using the BEAM VISAT 3.4 ® software application. Also, in order to combine these data with already existing auxiliary information it was necessary to project the image into a proper map projection (Hayford-Gauss, Datum Lisboa). This was easily accomplished since MERIS Level 2 imagery are already geolocated.

Ancillary Data

To help the selection of samples for algorithm training and testing we made use of existent land cover databases and high spatial resolution earth observation data, namely:

- CORINE Land Cover map for 2000 (CLC2000) (Painho and Caetano, 2005) – land cover map with 44 classes derived from visual interpretation of Landsat-7 ETM+ images acquired in the summer of 2000;
- National cover of SPOT High Resolution Geometric (HRG) images acquired in 2003, owned by the Portuguese Geographic Institute (IGP);
- National cover of orthorectified colour infrared aerial photography for 1995;
- National Forest Inventory for 1995.

METHODOLOGY

Land Cover Nomenclature

In this study we used the land cover nomenclature developed by the Remote Sensing Unit of the Geographic Portuguese Institute (IGP) (Table 1). The nomenclature classes were defined through the Land Cover Classification System (LCCS) from Food and Agriculture Organization (FAO) (Di Gregorio, 2000). The description of land cover classes is based on LCCS classes definition. The rationale behind the development of the nomenclature was two-fold: (1) a nomenclature that is well adapted to the type of landscapes existent in regions with characteristics similar to the Portuguese mainland, and (2) a nomenclature that is compatible with established ones (e.g., CORINE Land cover, Global Land cover and the International Geosphere-Biosphere Programme nomenclatures) in order to make possible the comparison of our maps with the ones that used other nomenclatures.

Table 1. Land Cover nomenclature and classes description

Land Cover Class	Acronym	Description
Continuous Artificial Areas	11	The land cover consists of non-linear built up areas which can be further specified into industrial area(s) or urban area(s). The density of the impermeable surface(s) is high.
Discontinuous Artificial Areas	12	The land cover consists of non-linear built up areas which can be further specified into industrial area(s) or urban area(s). The density of the impermeable surface(s) is medium or low.
Rain fed Herbaceous Crops	21	Field(s) are covered by rain fed herbaceous crops. The crop covers the land during the cultivation period of a shifting system
Irrigated Herbaceous Crops	22	Field(s) are covered by irrigated herbaceous crops. The irrigation systems commonly used are surface, sprinkler and drip irrigation. The crop covers the land during the cropping period of a fallow system.
Rice Crops	23	Field(s) are covered by herbaceous crops. The crop uses infiltrated water as reserve for establishment. The crop covers the land during the cropping of a fallow system.
Permanent Evergreen Crops (Trees or Shrubs)	241	Evergreen tree crops cover a defined area. // Evergreen shrub crops cover a defined area.
Permanent Deciduous Crops (Trees or Shrubs)	242	Deciduous tree crops cover a defined area. // Deciduous shrub crops cover a defined area.
Broadleaved Closed Trees	311	The main layer consists of broadleaved closed trees. The crown cover is more than (70-60)%. The height is in the range 30-3 m.
Broadleaved Open Trees	312	The main layer consists of broadleaved woodland. The crown cover is between (70-60) and (20-10)%. The height is in the range 30-3 m.
Needleleaved Closed Trees	321	The main layer consists of needleleaved closed trees. The crown cover is more than (70-60)%. The height is in the range 30-3 m.
Needleleaved Open Trees	322	The main layer consists of needleleaved woodland. The crown cover is between (70-60) and (20-10)%. The height is in the range 30-3 m.
Mixed Closed Trees	331	The main layer consists of broadleaved closed trees. The crown cover is more than (70-60)%. The height is in the range 30-3 m. / The main layer consists of needleleaved closed trees. The crown cover is more than (70-60)%. The height is in the range of 30-3 m.
Shrubland	35	The main layer consists of closed to open shrubland. The crown cover is between 100 and 15% The height is in the range of 5-0.3 m.
Natural grassland	37	The main layer consists of closed to open herbaceous vegetation. The crown cover is between 100 and 15%. The height is in the range of 5-0.3 m.
Sparse Vegetation	38	The main layer consists of sparse herbaceous vegetation. The crown cover is between (20-10) and 1%. // The main layer consists of sparse shrub vegetation. The crown cover is between (20-10) and 4%. // The main layer consists of sparse tree vegetation. The crown cover is between (20-10) and 4%.
Recently Burnt (Trees or Shrubs)	310	The main layer consists of closed to open trees affected by forest fires. The crown cover is between 100 and 15%. The height is in the range of 30-3m. // The main layer consists of closed to open shrubland affected by forest fires. The crown cover is between 100 and 15%. The height is in the range of 5 - 0.3m.
Permanent Wetlands	5	Primarily vegetated areas containing more than 4 percent vegetation during at least two months a year. The environment is influenced by the presence of water over extensive periods of time, i.e. water is present for more than three months a year and when water is present less than three months a year, it is present 75 percent of the flooding time. The vegetative cover is characterized by the presence of (semi)natural vegetation which species composition, its environmental and ecological processes are indistinguishable from, or in a process of achieving, its undisturbed state. The vegetative cover is not artificial and does not need to be managed nor maintained. This class

		includes floating vegetation but excludes areas with only occasional flooding.
Barren	6	Primarily non-vegetated areas containing less than four percent vegetation during at least 10 months a year. The environment is influenced by the edaphic substratum. The cover is natural. Included are areas like bare rock and sands.
Water Bodies	7	The land cover consists of natural waterbodies. A further specification can be made in flowing or standing water. // The land cover consists of artificial water bodies. A further specification can be made in flowing or standing water.

Land Cover Sampling

In order to test the several classification approaches using the defined nomenclature, we carry out a collection of samples for the 19 nomenclature classes for the year of 2004. The sampling process was performed by visual interpretation of high spatial resolution Earth Observation data, namely SPOT5-XS images acquired during the year of 2003 and orthorectified colour infrared aerial photography of 1995. To help sampling collection we also made use of the land cover databases described in the Study Area and Dataset section.

Each sample represents a specific land cover class covering at least 90% of the area covered by a MERIS 300-by-300 m pixel (same as the nominal resolution of used satellite images). The goal was to have, for each land cover class, the samples distributed all over the mainland territory in order to obtain a high-quality representation of possible regional within class differences. For each class we collected 40 samples and its spatial distribution is shown in Figure 1.

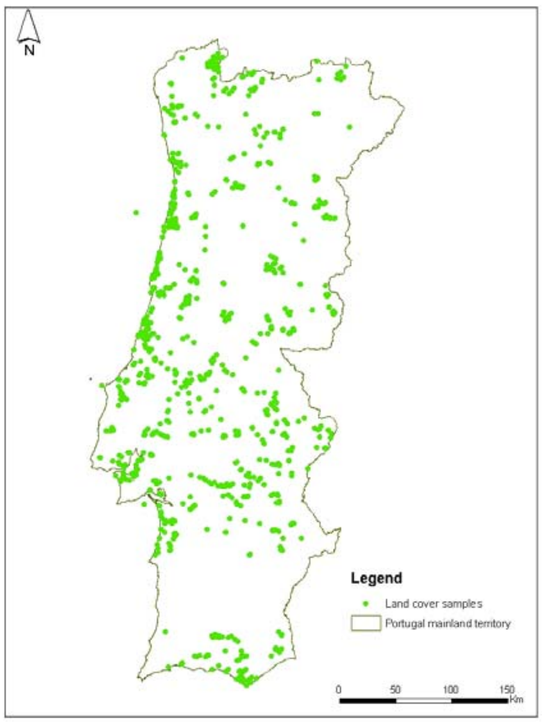

Figure 1. Spatial distribution of collected samples.

Dimension Reduction

When data to be classified lie in multi-dimensional vector spaces, it is often the case that not all dimensions show the same discrepant power. Ever because components are highly correlated, or simply because all classes merge in a unique cluster when looking at marginalized data in some particular direction, redundant or information less coordinates can be discarded from a classification viewpoint. As computational cost of classifiers increases rapidly with the feature space dimension, it is in general profitable to keep the dimensionality of the problem as small as possible, as long as it does not degrade the classification accuracy. The dimensionality issue can turn even worse in the supervised case: if the feature space dimension is excessive compared to the size of the available training set, the classifier can fail at finding a simple decision rule. All this makes data dimension reduction an important step in classification task.

Feature selection and feature extraction are the two common ways to perform dimensionality reduction (Guyon and Elissee, 2003). Feature selection refers to techniques that identify the best subspace within the input features space, whereas extraction methods aim at finding the best transformation (by means of coordinate's combinations) of the input feature space. The choice between those two approaches depends on the physical meaning of the coordinates dimension.

Principal Component Analysis (PCA) (Thomaz *et al.*, 1999; Jain *et al.*, 2000) is a widely used feature extraction that relies on the eigenvectors decomposition of the data covariance matrix. The corresponding linear transformation is an axis rotation that yields a system of systematically uncorrelated vectors (and statistically independent if data are normally distributed), aligned with the directions of maximum overall data variance. Mapped in this new coordinate system, data are compactly represented and it is likely that retaining only the few first components associated to the largest eigenvalues, suffices to explain most of the overall data variability. Clearly though, this is an unsupervised method that does not take into account any prior information about classes, and therefore nothing guarantees that the directions of maximum variance are also the directions of maximum classes discrepancy.

Fisher's Discriminant Analysis (FDA) is a supervised extension of PCA, which identifies a reduced dimension system that enhance inter-classes separability (Mousavi *et* al., 2003; Jain *et al.*, 2000). Inter-classes separation is here measured by the Fisher criterion, calculated as the eigenvalues of the between-class scatter matrix (B) divided by the within-class scatter matrix (W). The within-class scatter matrix is defined by:

$$W = \sum_{j=1}^{c} \sum_{\mathbf{x_i} \in E_j} (\mathbf{x_i} - \overline{\mathbf{x_j}})^t (\mathbf{x_i} - \overline{\mathbf{x_j}}) \tag{1}$$

where $\boldsymbol{x}$ denotes the vector data, c is the number of classes, and $\overline{\mathbf{x_j}}$ is the mean vector of class j. The between-class scatter is defined by:

$$B = \sum_{j=1}^{c} q_j (\overline{\mathbf{x_j}} - \overline{\mathbf{x}})^t (\overline{\mathbf{x_j}} - \overline{\mathbf{x}}) \tag{2}$$

where q_j is the number of samples in class j, and $\overline{\mathbf{x}}$ is the mean of the overall data distribution.

The representation maximizing classes discrepancy was proposed by Fisher as the projection onto the eigenvectors associated to the largest eigenvalues of the matrix $W^{-1}B$.

Classification

k-Nearest Neighbor. kNN is an instance-based learning algorithm (Cover and Hart, 1967) based on a distance function (e.g. Euclidian distance) between pairs of observations. The algorithm is quite simple: given a feature test vector, the system finds the k-nearest neighbors among the training vectors, and uses the labels of these k neighbors to determine the class of the unknown sample.

K-means. The decision rule adopted in K-means classifiers is to attribute a sample to-be-classified to the class which centre of gravity is closest.

Linear Discriminant Analysis. This classifier combines a Fisher Discriminant Analysis with a linear (or quadratic) separator function. Moreover, LDA implicitly assumes that data are normally distributed within each class (Johnson and Wichern, 1998).

Support Vector Machines. Support Vector Machines (SVM) are a new generation of supervised learning systems based on recent advances in statistical learning theory (Cristianini and Shawe-Taylor, 2000). Pioneered by the work on learning strategy by Vapnik and collaborators (Boser *et al.*, 1992; Vapnik, 1998), they have rapidly and successfully been applied to numerous real-world classification problems.

Conceptually, SVM rationale is that a classification problem that does not have a satisfying solution in its own observation space may have one simple and efficient in a more complicated representative system. As so, SVM use a hypothesis space of linear indicator functions to draw classification in a high dimensional (possibly infinite) feature space, image of the observation space by a non-linear mapping .

Consider a binary classification-learning task(*)and the following data training set:

$$S\{(\mathbf{x}_1, y_1),...,(\mathbf{x}_N, y_N)\}, \mathbf{x}_i \in \mathbf{X} \subseteq \Re^{n,x}, y_i \in \{-1,+1\}, \tag{3}$$

with samples $(\mathbf{x}_i, y_i)$ drawn i.i.d. according to some unknown but fixed probability distribution $P(\mathbf{x}, y)$. The non-linear mapping   transforms the n_x-dimensional input space $\mathbf{X}$ into the feature space $\mathbf{F}$, and let the set of hypotheses be linear hypothesis functions of the type:

$$h(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle + b. \tag{4}$$

One remarkable fact about this hypothesis function (written here in its dual form) is that it implies the data only through their inner products in the feature space. Therefore, if   is properly chosen so that it obeys the so-called *kernel* condition,

$$\langle \Phi(\mathbf{x}), \Phi(\mathbf{z}) \rangle = K(\mathbf{x}, \mathbf{z}), \tag{5}$$

we do not even need to know the underlying feature map   to be able to learn in the feature space.

Assuming so, the problem is then to determine the coefficients $\{\alpha_i, i = 1,..., N\}$ that minimize the classification error on unseen samples. Ideally, the best classifier should minimize the expected value of the *loss* or *risk*:

$$R(h) = \mathrm{E}_{P(\mathbf{x},y)}\{L(y, h(\mathbf{x}))\} = \int L(y, h(\mathbf{x}))dP(\mathbf{x}, y), \tag{6}$$

where the loss function $L(y, h(\mathbf{x})) \geq 0$, penalizes the deviations. In practice however, the joint probability function $P(\mathbf{x}, y)$ is unknown, and the expected loss (also called *generalization loss*) is approximated by the *empirical classification error* based on the available information (i.e. the training set). Then, the empirical risk functional reads:

(*)For the sake of simplicity we restrict this introductory study to the case of a two-classes classification. Generalization to multiple classes problems is straightforward using a one-versus-the-rest strategy.

$$R_{emp}(h) = \sum_{i=1}^{N} L(y_i, h(\mathbf{x}_i)) \, . \tag{7}$$

Often though, finding the hypothesis h^* that minimizes this empirical risk leads to an ill-posed problem that results in the well-known phenomenon described as *over-fitting* in the literature (i.e. the selected hypothesis is too complex). One way to avoid over-fitting tolerating noise and outliers, is to restrict the complexity of the hypothesis function, by introducing a regularization term (Vapnik, 1982). This regularization term is closely related to the notion of margin, another important concept for SVM that reflects the sensitivity and tolerance of the classifier to the samples $\{\mathbf{x}_i, i = 1,\ldots,N\}$ that stand "close" to the (non-linear) separator. Those points are called support vectors, and the solution that minimizes the regularized empirical risk function is referred to as a soft margin SVM (in contrast to a hard margin SVM that systematically zeroes the empirical loss). Solution to this quadratic optimization problem is classically obtained by Lagrangian theory and comes out to find the Lagrangian multipliers * that maximize the following quantity:

$$W(\alpha) = \sum_{i=0}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \left(K(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{C} \delta_{ij} \right)$$

$$\text{subject to } \sum_{i=1}^{N} y_i \alpha_i = 0 \text{ and } \alpha_i \geq 0 \text{, } i=1,\ldots\text{N}. \tag{8}$$

C is the regularization parameter that bounds the Lagrangian multipliers (i.e. the weights associated to the support vectors) controlling this way the capacity of the hypothesis function class(*).

Finally, a Radially Basis Function (RBF) kernel, $K(\mathbf{x}, \mathbf{z}) = e^{-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{\sigma^2}}$, with user pre-defined sample variance ^{2}, is chosen because it maps the input space into an infinite dimensional feature space and often yields good results for nonlinear regression (Suykens *et al.*, 2002; Seeger, 2004).

RESULTS AND DISCUSSION

All classification results we report in this section were obtained using a cross validation technique that permits evaluating classifiers' efficiency (in terms of generalized bias) using only the collected sample set. We fixed to five the number of cross validation folds. Thus, the initial set of 40 samples per class was divided randomly into five subsets: each time four subsets are used for training purpose, while the remaining fifth is used for classification test.

Figure 2 presents the mean spectral signatures of land cover classes derived from the MERIS Level 2 image of August 14th, 2004. All curves, except those for impervious land covers like urban or barren, display a minimum and very similar reflectance value along the visible wavelength range that is comprised between 400 and 700nm. This is mainly because these reflectance wavelengths are scattered by the atmosphere and at the same time largely absorbed by chlorophyll – land cover pixels with vegetation do not show up very brightly in this fraction of the spectrum. Although spectral similarities exist among vegetated land cover classes, these bands can be useful for soil/vegetation discrimination and mapping of man-made features. On the other hand, near infra-red wavelengths are good for mapping vegetation biomass content, providing an evident contrast between different types of vegetation. Vegetation classes show a steep slope between the red and the near infra-red reflectance that is

---

(*)There exists an alternative way for controlling the hypothesis function capacity, which amounts to fix the proportion of training samples that will lie between the boundary and the margin hyperplane. In SVM literature, this regularization strategy is often referred to as the -parameterization, and we will use this in our study.

characterized by a peak around 700nm. This portion of the spectrum appears as disjunction point amongst vegetated land cover classes that maximizes their separation. This suggests that the combination and exploitation of land cover spectral profiles along visible and near-infrared wavelengths is important for the separation of different land cover classes and should significantly improve classification scores obtained with a lower dimensional representation of the available spectral data.

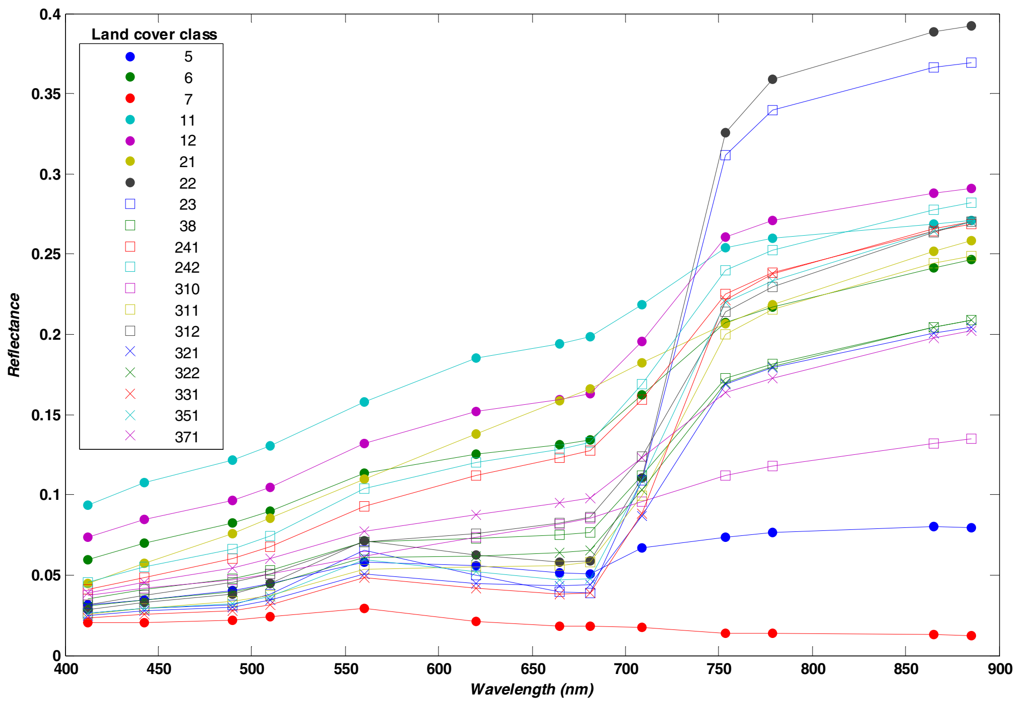

Figure 2. Mean spectral signatures for land cover classes derived from the MERIS image of August 14th, 2004.

We first performed an empirical estimation of the MERIS spectral bands correlations, which unquestionably evidenced the need of dimension reduction to remove redundant information. The averaged correlation coefficient between bands of the visible interval is about 95%, and it goes up to 99% for infra-red bands. In contrast, this same averaged coefficient falls down to 33% when measuring the correlation between visible and infra-red wavelengths. Then, it is not surprising that a Principal Component Analysis applied to the samples set reveals that more than 99% of the overall data variance summarizes in the first three principal components. As a result, on can expect that feature extraction, using for instance Fisher's Discriminant Analysis, will dramatically improve classification performances.

In Table 2, we present the overall classification scores obtained when applying the different classifiers described in the Methodology section, with and without prior dimension reduction.

Table 2. Overall classification accuracies achieved with different classifiers and feature data sets

Classifier	Overall accuracy (%)	
	Reflectance data	FDA output data
kNN	64	76
K-means	31	67
SVM	73	78
LDA	-	77

As expected, dimension reduction through FDA considerably improves the overall classification performances obtained with both kNN and K-means classifiers. This suggests that Land Cover classes are better separated (w.r.t Euclidean distance, at the core of theses two classifiers) in the reduced eigenvectors hyper plane than in the original highly dimensional feature space. Moreover, results obtained with LDA not only indicate that data are reasonably linearly separable in the transformed space, but also that multivariate canonical distribution law is in good agreement with the normal hypothesis. Yet, it is worth noticing that to ensure linear separability of data with a 99% confidence interval, 11 eigenvectors were still needed (recall that initially we are dealing with reflectance vectors of dimension 13).

Regarding SVM results, the influence of dimension reduction is not so drastic, in accordance with the well known fact that SVM is quite insensitive to the curse of dimensionality.

Now, in terms of performances and computational costs, SVM clearly leaders when no dimension reduction is possible. Whenever FDA is affordable, then k-NN and LDA attain results very comparable to the ones achieved with SVM. However, the latter classifier is computationally more expensive, and LDA might be in that case a better choice.

Let us now split the global classification accuracy into generalization losses per land cover class (Table 3), considering the SVM classification of reflectance data linearly mapped by FDA. A closer view at the producer's and user's accuracy shows that "Broadleaved Closed Trees" (311) is definitely the less distinctive land cover class, frequently confounded with "Broadleaved Open Trees" (312) and "Mixed Closed Trees" (331). On the contrary, pixels of "Irrigated Herbaceous Crops" (22) and "Rice Crops" (23), two land cover classes that were very close according to mean spectral profiles of Figure 1, surprisingly almost never mix up. These results suggest that main classification puzzlement occurs among pixels of classes that contain the same land cover features but in different proportion arrangements. Land cover classes that consist of absolutely different surface elements, even with vegetation within their composition, are well separated. This demonstrates that the visual spectral similarity amongst different land cover classes does not imply their mixing, as attested by the overall accuracy results obtained with the exploited classifiers.

Table 3. User's and producer's per class. This repartition corresponds to a SVM classification of FDA transformed data based on reflectances measured at a single date (August 14, 2004).

Land cover class	**5**	**6**	**7**	**11**	**12**	**21**	**22**	**23**	**38**	**241**	**242**	**310**	**311**	**312**	**321**	**322**	**331**	**351**	**371**	**User's Accuracy (%)**
5	39															1				98
6		25		1	3				7			1							3	63
7			40																	100
11		1	1	24	11		1				1		1							60
12		1		7	31				1											78
21						36				2									2	90
22							31	1		1	3			1			1	2		78
23							5	35												88
38		1			1				34			1				1		1	1	85
241						3		1		26	8			1					1	65
242										6	28		1	3				1	1	70
310												39	1							98
311									1	1		1	18	9		3	7			45
312													2	36		1	1			90
321															31	7	2			78
322									4			1	1		5	26		1	2	65
331							1						4	1	5		29			73
351									2						3			35		88
371						1			2		1	1			2		2		31	78
Producer's Accuracy (%)	100	89	98	75	67	90	82	95	67	72	68	89	64	70	67	66	6	88	76	

CONCLUSION

The main goal of the current study was to evaluate the usefulness of using SVM for the characterization of numerous land cover classes with high spectral dimensional MERIS imagery data of Portugal continental. The achieved results suggest that SVM classifier outperforms kNN, K-means and LDA classifiers in term of overall classification accuracy with both data sets – the full MERIS Level 2spectral imagery and the low-dimensional representation of such data derived through Fisher's discriminants. The level of classification accuracy achieved with SVM classifier is better than with the other classifiers when the whole spectral information is used. On the other hand, the kNN and LDA performed a land cover classification with similar accuracy to SVM when a low-dimensional and rearranged representation of spectral data is used as input for classification task. However, unlike kNN and LDA classifiers, SVM does not require a pre-processing technique to reduce data dimensionality in order to improve its classification accuracies, since they remained stable with full and low dimensional imagery data. The preliminary tests with the full spectral data and compressed information with FDA showed that common classifiers had to face the problem of insufficient training data, which logically degraded the classification scores. In contrast, SVM proved to be insensitive to the dimensionality of input data, since classification results with MERIS data are improved, more stable and reliable than with other classifiers. From a methodological viewpoint, SVM are powerful and easy to tune learning systems that should rapidly enter the standard classifiers' toolbox used in remote sensing applications.

Regarding the particular land cover classes discrimination we think that MERIS imagery have an enormous capability to distinguish between them. The results proved that large mixing occurs only between land cover classes that contain the same land cover features but in different proportion arrangements. Classes with completely different configuration characteristics present a classification accuracy obtained with SVM classifier that is over 80%.

Future work will rely on the multi-temporal analysis of MERIS data, land cover classes redefinition, and data fusion appliance, since we consider that is possible to increase the overall classification accuracy obtained with MERIS images. We will also compare the classification accuracy obtained with MODIS data, a sensor with similar spatial and temporal resolutions to MERIS, but that acquires spectral information within short wavelength infrared.

ACKNOWLEDGEMENTS

This study was carried out in the framework of the project "User driven land cover characterisation for multi-scale environmental monitoring using mulit-sensor earth observation data" funded by "Programa Dinamizador das Ciências e Tecnologias para o Espaço" from "Fundação para a Ciência e Tecnologia", and from the "Announcement of Oppurtunity for the Utilisation of ERS and ENVISAT Data" from European Space Agency (ESA).

REFERENCES

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In: Haussler, D., editor, Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, 22-27 July, USA, pp. 144-152.

Clevers, J.G.P.W., Zurita Milla, R., Schaepman, M., and Bartholomeus, H. (2004). Using MERIS on ENVISAT for Land Cover Mapping. Proceedings of the 2004 Envisat & ERS Symposium, 6-10 September, 2004, Salzburg, Austria.

Cover, T. M., and Hart, P. E. (1967). Nearest Neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1): 21-27.

Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.

Curran, P.J. and Steele, C.M. (2005). MERIS: the re-branding of an ocean sensor. *International Journal of Remote Sensing*, 26(9): 1781-1798.

Di Gregorio, A., and Jansen, L.J.M. (2000). *Land Cover Classification System*. FAO, Rome, 179 pp.

Gessner, U., Gunther, K., P., and Maier, S. W. (2004). Landcover/land use map of Germany based on MERIS full resolution data. Proceedings of the 2004 Envisat & ERS Symposium, 6-10 September, 2004, Salzburg, Austria.

Gonçalves, P., Hugo Carrão, Andre Pinheiro and Mário Caetano (2005). Land cover classification with Support Vector Machine applied to MODIS imagery. Proceedings of the 25th EARSeL Symposium, June 6 - 11, 2005, Porto, Portugal.

Guyon I. and Elissee, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3: 1157-1182.

Haertel, Victor and David Landgrebe (1999). On the Classification of Classes with Nearly Equal Spectral Responses in Remote Sensing Hyperspectral Image Data. IEEE *Transactions on Geoscience and Remote Sensing*, l(37): 2374-2386.

Huang, C., Davis, L. S. and Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4): 725-749.

Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14: 55-63.

Jackson, Q. and David A. Landgrebe (2001). An Adaptive Classifier Design for High-Dimensional Data Analysis with a Limited Training Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, 39(12): 2664-2679.

Jain, Anil K., Robert P.W. Diun, and Jianchange Mao (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis Intelligence*, 22(1): 4 –37.

Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis* (4th ed.). Prentice Hall, Upper Saddle River, New Jersey.

Landgrebe, David (2002). Hyperspectral Image Data Analysis as a High Dimensional Signal Processing Problem, (Invited), *Special Issue of the IEEE Signal Processing Magazine*, 19(1): 17-28.

Marçal, A. R. S., Borges, J. S., Gomes, J. A., and Costa, J. F. Pinto Da (2005). Land cover update by supervised classification of segmented ASTER images. *International Journal of Remote Sensing*, 26(7): 1347-1362.

Maxwell, S.K., Hoffer R.M., and Chapman P. L. (2002). AVHRR channel selection for land cover classification. *International Journal of Remote Sensing*, 23(23): 5061-5073.

Mercier, G. and M. Lennon (2003). Support vector machines for hyperspectral image classification with spectral based kernels. Proceedings of IEEE Intl. Geoscience & Remote Sensing Symposium, IGARSS 2003 Conference, 21-25 July, Toulouse, France.

Mousavi, M., J., K. L. Butler-Purry, R. Gutierrez-Osuna and M. Najafi (2003). Classification of Load Change Transients and Incipient Abnormalities in Underground Cable Using Pattern Analysis Techniques. Proceedings of 2003 IEEE/PES Transmission and Distribution Conference, September 7-12, Dallas, TX.

Painho, M., and Caetano, M. (2005). *Cartografia de Ocupação do Solo, Portugal Continental, 1985-2000*. Instituto do Ambiente, Lisboa, Portugal.

Pal, M., and Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5): 1007-1011.

Price, K.P., X. Guo, and J.M. Stiles (2002). Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing*, 23(23): 5031-5042.

Santos, T., Tenedório, J., Rocha, J., and Encarnação, S. (2005). SATSTAT: Exploratory Analysis of Envisat-MERIS Data for Land Cover Mapping of Portugal in 2003. Proceedings of 14th European Colloquium on Theoretical and Quantitative Geography, September 9-13, 2005, Tomar, Portugal.

Seeger, M. (2004). Gaussian processes for machine learning. *Interrnational Journal of Neural Systems*, 14 (2):1-38.

Shah, C. A., Watanachaturaporn, P., Arora, M. K., and Varshney, P. K. (2003). Some Recent Results on Hyperspectral Image Classification. In IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data. October 27-28, 2003, NASA Goddard Spaceflight center, Greenbelt, MD.

Suykens, J., Gestel, T., de Brabanter, J., de Moor, B., and Vandewalle, J. (2002). *Least Squares Support Vector Machines*. World Scientific.

Thomaz, C.E., R.Q. Feitosa, and A. Veiga (1999). Separate-Group Covariance Estimation with Insufficient Data for Object Recognition. Proceedings from the Fifth All-Ukrainian International Conference, Ukraine, pp. 21-24.

Vapnik, V. (1982). *Estimation of dependencies based on empirical data*. Springer Verlag, New York.

Vapnik, V. (1998). *Statistical Learning* Theory. Wiley.

Vieira, C. A. O., P. M. Mather, and P. Aplin (2001). Multitemporal classification of agricultural crops using the spectral-temporal response surface. In Analysis of Multi-Temporal Remote Sensing Images: Proceedings of the First International Workshop 2001, pp. 290-297.

Zhu, G., and Blumberg, D. G. (2002). Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. *Remote Sensing of Environment*, 80: 233-240.