

Article

Testing Multivariate Adaptive Regression Splines (MARS) as a Method of Land Cover Classification of TERRA-ASTER Satellite Images

Elia Quirós, Ángel M. Felicísimo and Aurora Cuartero *

Polytechnic School, University of Extremadura, 10071 Cáceres, Spain;

E-Mails: equiros@unex.es (E.Q.); amfeli@unex.es (A.M.F.)

* Author to whom correspondence should be addressed; E-Mail: acuartero@unex.es;
Tel.: +34-927257195; Fax: +34-927257324.

Received: 24 September 2009; in revised form: 1 November 2009 / Accepted: 2 November 2009 /
Published: 13 November 2009

Abstract: This work proposes a new method to classify multi-spectral satellite images based on multivariate adaptive regression splines (MARS) and compares this classification system with the more common parallelepiped and maximum likelihood (ML) methods. We apply the classification methods to the land cover classification of a test zone located in southwestern Spain. The basis of the MARS method and its associated procedures are explained in detail, and the area under the ROC curve (AUC) is compared for the three methods. The results show that the MARS method provides better results than the parallelepiped method in all cases, and it provides better results than the maximum likelihood method in 13 cases out of 17. These results demonstrate that the MARS method can be used in isolation or in combination with other methods to improve the accuracy of soil cover classification. The improvement is statistically significant according to the Wilcoxon signed rank test.

Keywords: multi-spectral classification; multivariate adaptive regression splines (MARS); area under the ROC curve (AUC); TERRA-ASTER image

1. Introduction

Conventional classification methods used in remote sensing have some basic problems due to the fact that they are not adapted to the real characteristics of image data. In addition, they lack proper configurations, and there is generally minimal user interaction.

Traditional remote sensing classification methods are divided into two large families. The first family is parametric, and includes the ML, bayesian methods, *etc.*; in this family, initial conditions (such as Gaussian distributions of reflectances or homoscedasticity) are not usually met in the remote sensing images. This means that the power of these tests is seriously undermined, and thus, classifications can be unnecessarily weak. The second family includes non-parametric methods (e.g., neural networks and classification trees). This second group addresses the problems posed by the first group by making no assumptions regarding reflectance distributions. Some of these methods have problems related to their “black box” format that seriously undermine any understanding of the relationship between initial variables and classification results, thereby greatly limiting the generalizability of the results.

According to [1], “parametric methods are those in which the decision rules are obtained from probability density functions for each class, whose parameters are estimated from samples collected for each learning category, without taking into account the others.” In this kind of classification, it is commonly assumed that the probability distribution functions for each class follow a normal multivariate distribution.

In addition to the assumption of a probability distribution that may be inappropriate for the data under analysis, parametric methods have added another substantial problem. Namely, if the data have a high dimensionality, many samples are required for the learning stage of these methods. Overall, the assumption of a normal spectral distribution is often violated, especially in complex landscapes. In addition, insufficient, non-representative, or multimode distributed training samples can further introduce uncertainty to the image classification procedure [2].

Many studies have shown that non-parametric methods provide better classification results. In studies such as [3], it is demonstrated that, even with small training samples, non-parametric classification algorithms provide better results than parametric ones. Also, [4] concluded that, under the assumption of overlapping training samples, a non-parametric algorithm is preferable as a classification method. With non-parametric classifiers, it is not required to assume that the data follow a normal distribution. No statistical parameters are needed to separate image classes. The decision function coefficients corresponding to separation boundaries between classes in representation space are obtained from samples of all classes, and such functions do not follow mathematically defined structures [1].

MARS is a non-parametric regression method in which no assumption is made regarding the functional relationship between dependent and independent variables. Instead, MARS builds this relationship from a set of coefficients and basic functions, which in turn are heavily influenced by the regression of the data. The operating method involves partitioning the area of entry into regions, each with its own regression equation [5].

This method was proposed by [6], and it is essentially an algorithm based on recursive partitioning and multi-stage regression that uses spline functions to align data with an arbitrary regression

function [7]. Thus far, the MARS method has been widely used for predictive and simulation efforts, as in [8-11], and [12]. Other authors have compared this technique with other existing models. In this sense, [13] indicated that "MARS is better suited to model situations that include a high number of variables, non-linearity, multicollinearity and/or a high degree of interaction among predictors." In addition, works such as [14] and [15] have praised the rigorous statistical basis upon which the MARS technique MARS has been built, and they have highlighted the clear advantage of MARS results as compared to the "black box" results generated by neural networks.

As shown in [16], the MARS method can be extended to handle classification problems. This study analysed two types of classifications. The first type was comprised of those that included only two classes; in this case, the MARS method without any variant is solid and competitive. A second type of classification with more than two classes was also considered; in this second case, the use of a hybrid method called PolyMARS was recommended. We do not detail the latter technique because in our case, classification is carried out by means of a pairwise procedure that corresponds to the first of these two types [16].

It should be emphasised that [17] and [18] included the MARS method in the neural multilayer networks classification group, but it appears that the MARS method has not yet been applied in the context of multi-spectral imaging classification efforts. In fact, [18] refers to the MARS and autonomous classification engine (ACE) methods by expounding that "we believe that these methods will have a place in classification practice, once some relatively minor technical problems have been solved. As yet, however, we cannot recommend them on the basis of our empirical trials."

Only [19] has applied the MARS technique to remote sensing data, but since the author of that paper worked with radar images, the work reported in [20] is not comparable to that in our study. That is, radar data are active data, while the data used in this study are multi-spectral passive optical images, also known as advanced space-borne thermal emission and reflection radiometer (ASTER) images.

The ASTER sensor was developed in an attempt to use detailed geological data to understand phenomena such as volcanic activities that can significantly impact the global environment [20]. However, the current implementation of this technology is substantially different than expected, with the images produced by this sensor having been used for several purposes other than geologic ones. For example, various studies have used ASTER scenes to characterise urban areas [21], use information from five TIR bands for the discrimination of agricultural crops [22], estimate the biomass of boreal forests [23], validate ASTER scenes for the geometric reconstruction of cloud masses [25], investigate glacier geometries and movements [24,25], predict natural hazards [26,27], and detect soil temperatures [28,29]. Finally, it must be highlighted that [30] employed ASTER image to identify different types of land uses and also compared two types of classifiers.

On the other hand, on how to evaluate the accuracy of the classification, a receiver operating characteristic (ROC) curve is a 2D graph representing both the specificity and the sensitivity of a binary (i, j) classifier. The sensitivity of a classifier is the probability of classifying an item as belonging to class i when it actually belongs to class i . The specificity is the probability of classifying an item as not belonging to class j when it actually does not belong to class j .

AUC provides a measure of how well a classification algorithm performs. [13] pointed out that AUC summarises performance across all possible thresholds and is independent of balance among classes.

ROC curves are generated by varying a threshold across the output range of a scoring model and then observing the corresponding classification performances. This graph is necessary to obtain the AUC statistic. The AUC statistic has an important property; namely, the AUC of a classifier is equivalent to the probability that classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [31]. A random guess classifier will produce a diagonal line on the ROC diagram with an area of 0.5. Hence, the AUC of a realistic classifier lies between 1 and 0.5. The high value of 1 corresponds with an ideal classifier, so the closer the AUC statistic is to 1, the better the classification is expected to be.

[32,33] concluded that the AUC statistic is more of a discriminating measure than an accuracy statistic. The use of the AUC statistic has been widely extended in predictive modelling techniques. However, only a few remote sensing works, such as [34] or [35], have used AUC for evaluating classification performance. This study uses the AUC statistic to evaluate three classification methods.

In brief, the aim of this study is to evaluate the MARS algorithm as a remote sensing classifier. For this purpose, the same TERRA-ASTER scene was classified by MARS and by two classical remote sensing classifiers (the ML and parallelepiped methods) to compare class probabilities derived from the AUC statistic. Section 2 describes the study's framework and data processing methods. Section 3 introduces the classification methods, and section 4 analyses and discusses the results.

2. Materials

2.1. TERRA-ASTER scene

The test area was a roughly 60×60 km area in the Spanish province of Badajoz, which is located in Extremadura in southeastern Spain (Figure 1). Elevation in this area ranges from 77 to 855 m, with an average of 360 m. The area was captured by the ASTER sensor on 4 August 2000.

Figure 1. Location of the Extremadura test area in southeastern Spain.



The platform is composed of three different subsystems. First, the visible and near-infrared (VNIR) has three bands with a spatial resolution of 15 m and an additional backward telescope for stereoscopic

use. Second, the shortwave infrared (SWIR) has six bands with a spatial resolution of 30 m. Finally, the thermal infrared (TIR) has five bands with a spatial resolution of 90 m. Each subsystem operates in a different spectral region with its own telescope [36]; see Table 1.

Table 1. Main characteristics of the ASTER sensor systems.

Sub-system	Band n ^o	Spectral range (mm)	Spatial resolution (m)	Quantization levels
Visible & Near infrared (VNIR)	1	0.52–0.60	15	8 bits
	2	0.63–0.69		
	3N	0.78–0.86		
	3B	0.78–0.86		
Shortwave infrared (SWIR)	4	1.60–1.70	30	8 bits
	5	2.145–2.185		
	6	2.185–2.225		
	7	2.235–2.285		
	8	2.295–2.365		
	9	2.360–2.430		
Thermal infrared (TIR)	10	8.125–8.475	90	12 bits
	11	8.475–8.825		
	12	8.925–9.275		
	13	10.25–10.95		
	14	10.95–11.65		

The ASTER data can be downloaded free of cost from the website <http://asterweb.jpl.nasa.gov/>.

2.2. Vegetation map

The Regions of Interest (ROIs) used in the classification process were obtained from the Spain forest map (see Figure 2). This map provides a polygonal distribution of territory based on homogeneous vegetation units characterised by information on vegetation type. This study used a version synthesised by the Extremadura Regional Government in 2001 [37]. This version includes the main forest categories groups existing in the region. Taking only the area under study into account, a total of 18 categories must be analysed (see Table 2).

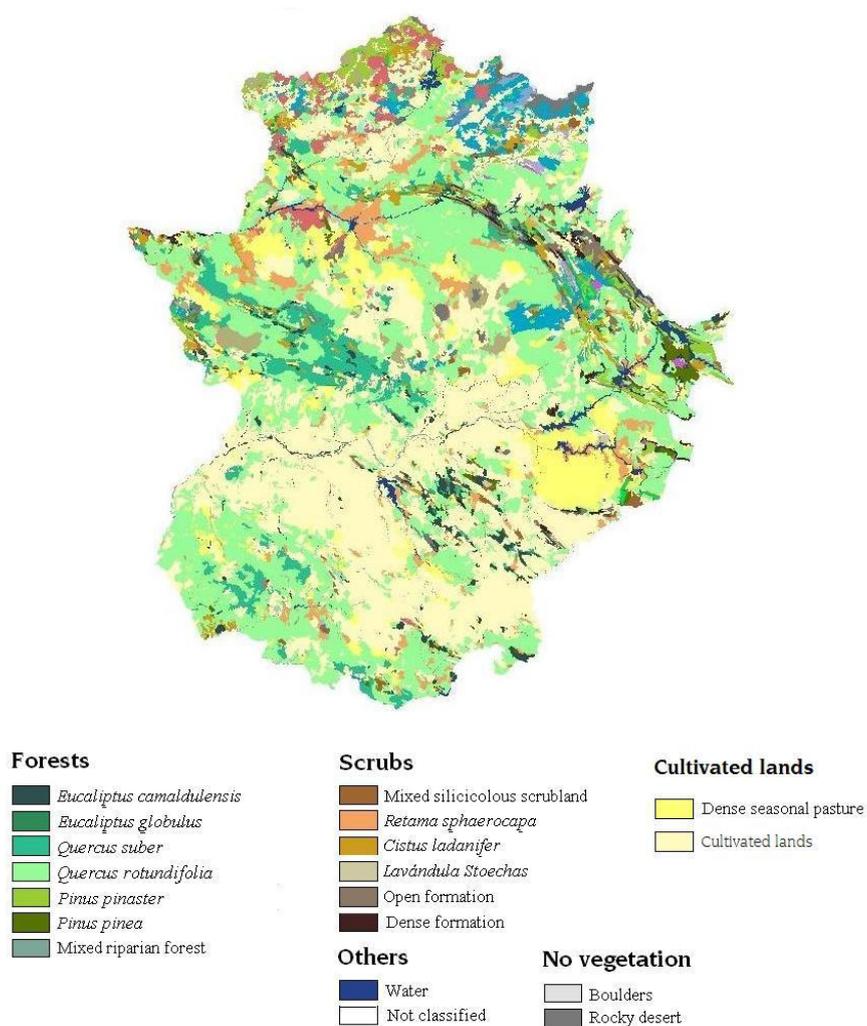
Table 2. Forest categories of area under study.

Forest map cod.	Legend	Total area (Km ²)	Percentage of the area under study
999	Water	49.6	1.3%
547	Mixed silicicolous scrubland	47.5	1.2%
534	Agricultural land	2,422.9	61.4%
507	Mixed riparian forest	23.5	0.6%
458	Dense seasonal pasture	135.6	3.4%

Table 2. Cont.

454	Open formation	8.7	0.2%
453	Dense formation	8.6	0.2%
337	Boulders	2.1	0.1%
329	Rocky desert	42.2	1.1%
309	<i>Retama sphaerocapa</i>	109.2	2.8%
303	<i>Cistus ladanifer</i>	22.0	0.6%
221	<i>Lavandula stoechas</i>	10.4	0.3%
62	<i>Eucalyptus camaldulensis</i>	194.7	4.9%
61	<i>Eucalyptus globulus</i>	10.1	0.3%
46	<i>Quercus suber</i>	53.5	1.4%
45	<i>Quercus rotundifolia</i>	786.0	19.9%
26	<i>Pinus pinaster</i>	1.4	0.0%
23	<i>Pinus pinea</i>	17.8	0.5%

Figure 2. Extremadura forest map (EFM).



2.3. Software

All operations were performed using a variety of softwares as ENVI (image processing), ArcView and ArcGIS (geographic information systems), SPSS (statistical analysis), and MARS (predictive modelling).

3. Methods

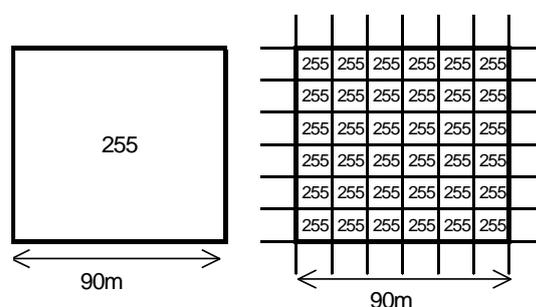
In this study, the data were pre-processed prior to any classification procedures (see Section 3.1). A brief introduction to the MARS algorithm is presented before developing different classifications (see Section 3.2.). Two different processes were performed in this study. On the one hand, we generated three traditional classification maps using the ML, parallelepiped and MARS algorithms (see Section 3.3.). On the other hand, class probability maps were also obtained using these three algorithms; their accuracies were calculated using the AUC statistic (see Section 3.4.).

3.1. Image preprocessing

Two different operations were implemented on the original image before starting classification processes. First, regardless of whether the data included L1-B images, that is, radiometrically calibrated and geometrically registered images, an additional geometric registration was performed using nine ground control points. These points were previously measured using GPS techniques.

Second, we re-sampled the SWIR and TIR bands to obtain the same geometric resolution for the entire image. With this action, no alteration was introduced to the original pixel information because the re-sampling was performed from a larger pixel resolution up to a smaller pixel resolution (see Figure 3). This re-sampling was necessary because the classifier requires homogeneous data to operate, and this method is simpler than introducing different band resolutions for different classifier calculations.

Figure 3. Re-sampling example for TIR bands.



After these two processes, the image data were ready for classification.

3.2. Introduction to the MARS algorithm

The MARS model is a spline regression model that uses a specific class of base functions as predictors in place of the original data [38].

A spline is a special function defined piecewise by polynomials, and it is used to refer to a wide class of functions that are used in applications requiring data interpolation. [16] defined the cubic spline as a function that has first and second continuous derivatives. Cubic splines are also called basic functions.

On the other hand, basic functions have a key attribute known as a “knot”. A knot marks the end of one region of data and the beginning of another [38]. The number of knots and their placement are fixed for regression splines, and in the MARS procedure, knots are determined by a search that occurs both forwards and backwards in a stepwise fashion. First, MARS generates a model with an excessive number of knots; then, knots that contribute least to the overall fit are eliminated.

Basis functions are used to search for knots; these functions serve as a set of functions representing the relationship between the predictor variables (X) and the target variable (y):

$$y = f(x) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X) \quad (1)$$

This function consists of an interceptor parameter β_0 and the weighted sum of other basic functions $h_m(X)$.

MARS uses what [16] denotes as “reflected pairs”. These are linear basic functions of the form $(x - t)_+$ and $(t - x)_+$, with t being the knot.

The MARS procedure is divided into three steps. First, a forward algorithm selects all possible basic functions and their corresponding knots. Second, a backward algorithm eliminates all basic functions in order to generate the best combinations of existing knots, and finally, a smoothing operation is performed to obtain continuous partition borders.

1. The selection of basic functions from the initial set is achieved by determining a constant function $h_0(X) = 1$ so that all functions from set C are candidates. New pairs of functions are considered at each stage until the model has the maximum number of terms set by the user at the beginning of the process.
2. The backward removal is performed by suppressing those model terms that contribute to a minimal residual error. This stage consists of reducing the complexity of the model complexity by increasing its generalisability [7]. This process can be conducted by means of generalised cross-validation (GCV).

$$GCV(\lambda) = \frac{\sum_{i=1}^N (y_i - \hat{f}_\lambda(x_i))^2}{(1 - M(\lambda)/N)^2} \quad (2)$$

With this GCV function, the optimum number of model terms (λ) can be estimated with:

$$M(\lambda) = r + cK \quad (3)$$

The value $M(\lambda)$ is the effective number or parameters in the model, and it is expressed in terms of r (*i.e.*, the number of linearly independent basic functions) and K (*i.e.*, the number of knots selected in the forward process).

The process stops when the number of model terms reaches $GCV(\lambda)$.

3. Finally, smoothing is necessary for removing discontinuities within regional borders and ensuring the continuity of first and second derivatives.

[16] included two possibilities for classification with MARS. The first one relates to pairwise classification in which the output can be coded as 0 or 1, thereby treating classification as a regression. The second possibility involves more than two classes, with the classification serving as a hybrid of MARS called PolyMARS as expounded in [39]. This study adopts the first technique.

3.3. Classification maps

In this process, 17 out of the total 18 categories were considered for classification. The “agricultural cultivations” class (cod 534) was excluded due to its heterogeneity. All the crops of dry regions and irrigated regions are incorporated into this EFM category so that very different spectral responses are included. Considering this class would cause serious errors in classification processes.

The original format of the EFM categories was the ArcInfo exchange coverage extension “*.e00”. Thus, an import operation became necessary to convert all classes into “shp” format. This process was performed using both ArcInfo and ArcView software.

The second preliminary operation, as referred to above, was to reduce all “shp” polygons through a buffer of –100 m in order to purify the ROIs. Not all polygons were chosen for the classification task. As shown in Table 3, the area of the ROIs fluctuate between 78.18% of the total area for class 221 and 31.89% of the total area for class 454.

Table 3. Regions of interest (ROI) used in the classifications.

Forest map cod.	Legend	Category areas at image (km ²)	ROI areas (km ²)	ROI area percentages
999	Water	49.6	33.3	67.19%
547	Mixed silicicolous scrubland	47.5	28.4	59.73%
507	Mixed riparian forest	23.5	11.6	49.18%
458	Dense seasonal pasture	135.6	52.2	38.47%
454	Open formation	8.7	2.8	31.89%
453	Dense formation	8.6	5.6	64.68%
337	Boulders	2.1	0.8	36.22%
329	Rocky desert	42.2	23.5	55.82%
309	<i>Retama sphaerocapa</i>	109.2	69.6	63.69%
303	<i>Cistus ladanifer</i>	22.0	10.6	48.33%
221	<i>Lavandula stoechas</i>	10.4	8.1	78.18%
62	<i>Eucaliptus camaldulensis</i>	194.7	140.6	72.25%
61	<i>Eucaliptus globulus</i>	10.1	5.7	55.93%
46	<i>Quercus suber</i>	53.5	26.0	48.48%
45	<i>Quercus rotundifolia</i>	786.0	544.2	69.23%
26	<i>Pinus pinaster</i>	1.4	0.7	52.24%
23	<i>Pinus pinea</i>	17.8	12.1	67.92%

Thus, we relied on a variety of ROI sizes to confirm whether there was relationship between training size and final classification results. As shown in Table 3, the smallest training size was for class 26, and the greatest size was for class 45.

This study uses the ML, parallelepiped, and MARS methods. A brief summary of the properties of each of these classifiers is given in the next section.

3.3.1. ML classification

ML is the most popular parametric classification method used in remote sensing. This method assigns observation X to class ω_l if the function $g_l(X)$ is larger than any other $g_j(X)$ for all $i \neq j$ so that:

$$g_i(X) = -\frac{1}{2} \ln(|\sum_i|) - \frac{1}{2} (X - \mu_i)^T \sum_i^{-1} (X - \mu_i) \quad (4)$$

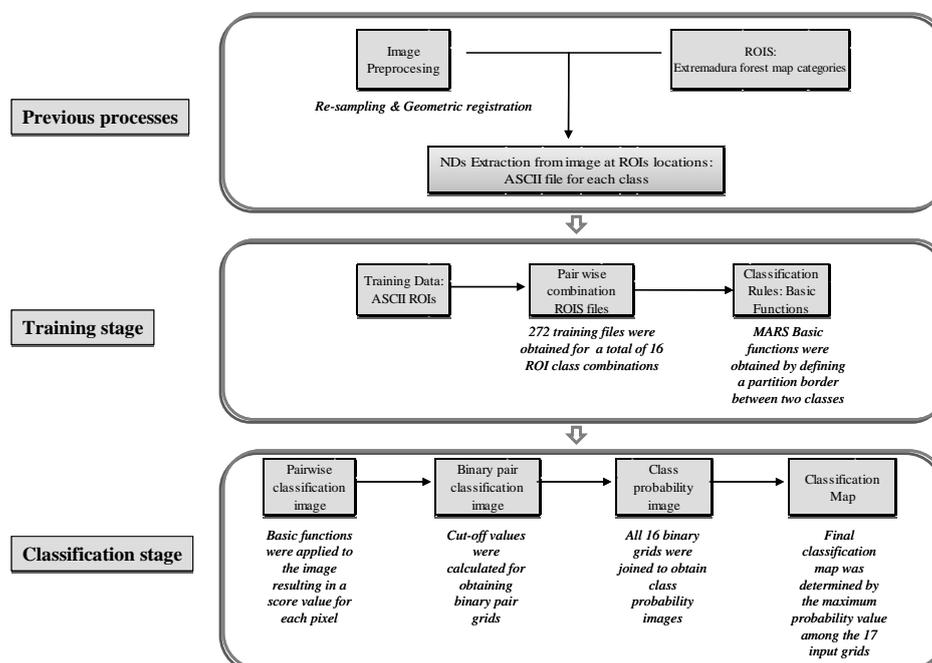
3.3.2. Parallelepiped classification

The parallelepiped classification method is a non-parametric method. It is one of the simplest classification methods based on a radiometric model rather than on the measurement of distances or probabilities [40]. The parallelepiped-like subspace is defined using training samples, and its boundaries can be determined in several ways, such as maximum or minimum pixel values, a corresponding class, a multiple of the standard deviation, and so on. The decision rule simply checks whether the point representing a pixel in the feature space lies inside any of the parallelepipeds [41]. The Boolean operator of the decision rule is based on the standard deviation.

3.3.3. MARS classification

In order to monitor the whole battery process more effectively, a chart diagram is displayed, summarizing the methodology for MARS classifications (Figure 4).

Figure 4. Work flow for the MARS process classification.



For MARS classification, it is necessary to know the digital numbers (DNs) of pixel values at all 14 bands. Thus, once all shp ROI polygons were superimposed onto the ASTER scene, we performed an ASCII exportation using ENVI software. The 17 resulting files contained image and map coordinates of all pixels inside the corresponding ROI and their corresponding DN values for all bands. See Figure 5.

Figure 5. Fragment of the ROI ASCII export file.

ID	X	Y	Map X	Map Y	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14
1	1380	1927	210840.27	4302737.64	87	79	81	92	74	77	69	66	68	1555	1687	1779	2078	2152
2	1379	1927	210825.27	4302737.64	86	80	77	92	74	77	69	66	68	1541	1668	1762	2056	2135
3	1380	1928	210840.27	4302722.64	84	77	82	92	74	77	69	66	68	1555	1687	1779	2078	2152
4	1381	1928	210855.27	4302722.64	86	78	75	92	75	80	72	67	69	1555	1687	1779	2078	2152
5	1379	1928	210825.27	4302722.64	84	75	77	92	74	77	69	66	68	1541	1668	1762	2056	2135
6	1378	1929	210810.27	4302707.64	86	73	78	87	69	72	65	64	64	1541	1668	1762	2056	2135
7	1379	1929	210825.27	4302707.64	85	73	78	87	70	72	65	64	66	1541	1668	1762	2056	2135
8	1380	1929	210840.27	4302707.64	84	76	78	87	70	72	65	64	66	1555	1687	1779	2078	2152
9	1381	1929	210855.27	4302707.64	86	78	74	89	72	74	69	65	66	1555	1687	1779	2078	2152
10	1382	1929	210870.27	4302707.64	88	83	80	92	75	80	72	67	69	1555	1687	1779	2078	2152
11	1378	1930	210810.27	4302692.64	86	74	76	87	69	72	65	64	64	1541	1668	1762	2056	2135

First, the image must be classified by discriminating each class from the rest. The class with which all other classes were compared was called the “fixed class”, and the fixed class was compared with the “comparing class”. Next, these classifications were merged into a unique probability image per class, and finally, the probabilities per class were joined into a final classification map. Based on this premise, the classification process was designed as follows.

a) Training stage

As discussed in [42], training is the classification stage used to estimate the parameters of the classifier algorithm; these parameters were the equations that defined partitions in multi-spectral space. The present study used the basic functions obtained by MARS for the multi-spectral space partition in each pair of classes.

a.1. Pairwise combination of ROI ASCII files

The pairwise combinations of ROI files were developed using SPSS. A new variable was introduced to distinguish between the fixed and comparing classes. This variable was assigned the value 1 for the fixed class and 0 for the comparing class. Overall, 272 training files were obtained, for a total of 16 ROI class combinations.

a.2. Obtaining basic functions

All 272 training files were introduced into the MARS software, and basic functions were obtained by defining a partition border between the two classes. Before validating these basic functions, they were applied to the input data again to verify that they could discriminate between class training data and that they could consistently extrapolate image holes for discriminating classes. This validation process was developed using the AUC statistic, and it was assumed to show the degree of adjustment of the MARS model relative to the input data.

b) Classification stage

Once the basic functions were obtained from the training data, the second stage consisted of applying them to the entire image in order to perform the actual classification.

b.1. Application of pairwise basic functions to the entire scene

Basic functions were introduced into an ArcInfo macro language (AML) file. This AML file was programmed to apply MARS basic functions to the image, resulting in a probability value for each pixel. This value indicated the probability of belonging to the fixed class (i.e., higher probability values) or to the comparing class (i.e., lower values).

b.2. Generating binary pair classification images

Until now, we have used 272 probability files for the combinations of all working classes. Now, we intend to generate a binary map in which the value 1 represents “pixel belongs to the fixed class” and value 0 represents “pixel belongs to the comparing class”, or equivalently, “pixel does not belong to the fixed class”.

The suitability scores obtained from MARS were not in the standard [0,1] interval. Thus, it was necessary to improve the application to calculate what some authors have called a “cut for best classification”. In our study, this value had to fulfil two conditions:

- Maximise correct classification probabilities
- Minimise incorrect classification probabilities

The application was programmed using SPSS software, and it counted false positive and false negative frequencies and subsequently changed the cut-off point in probability space. The process was similar to the one used for the ROC curve.

Once the cut-off value was calculated for all 272 probability files, the binary grids were generated using ArcInfo software.

b.3. Generating the class probability image

At this stage of the work, we had 16 binary grids per class. The purpose of the class probability image was to join all these binary grids.

All grids were added in ArcView, so the result was another grid with values ranging from 16 (if the pixel was always classified as the fixed class) to 0 (if the pixel had never been classified as belonging to the fixed class).

These values were transformed to the standard [0,1] interval dividing the grid by 16. Thus, the maximum probability value of 1 was assigned to those pixels that the MARS classifier always denoted with the same classification.

b.4. Obtaining the final classification image

The final operation was to join the 17 grids using ArcInfo software. This operation determined the class with the maximum probability value among the 17 input grids for each pixel.

3.4. Probabilities maps

At the same time that the classifications were calculated, a probability map was generated for each class. This allowed us to conduct an exhaustive evaluation of the method's accuracy in predicting classifiers.

3.4.1. The ML probability maps

While performing classification processes, ENVI software allows for the calculation of rule images. Rule images for ML classifiers are grids containing the discriminant function expressed in [42] for multivariate normal class models:

$$g_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - m_i)^t \Sigma_i^{-1} (x - m_i) \quad (5)$$

The rule image obtained for this study contained values in the interval $[-300,000, 100]$ with a heterogeneous distribution. This property impeded the use of rule images for the development of probability maps.

An alternative probability map was calculated by following the same pairwise method used for the MARS classification. We generated 272 binary ML classifications and combined them as explained in the MARS procedure to obtain the final probability map.

This option was also useful to validate the pairwise classification method because we did not stop at the class probabilities map but rather completed the entire process and thus obtained a final ML classification map; from now on, this will be called the pairwise ML classification map. If the pairwise classification process is valid, then the final pairwise ML classification map should be the same as the original ML classification map from Section 3.3.1.

3.4.2. Parallelepiped probability maps

The ENVI rule images for parallelepiped classification provides for each pixel the number of bands that fulfils the parallelepiped condition. These pixel values were considered as probabilities with no transformations because the AUC statistic can be calculated with values not in the $[0, 1]$ interval.

3.4.3. MARS probability maps

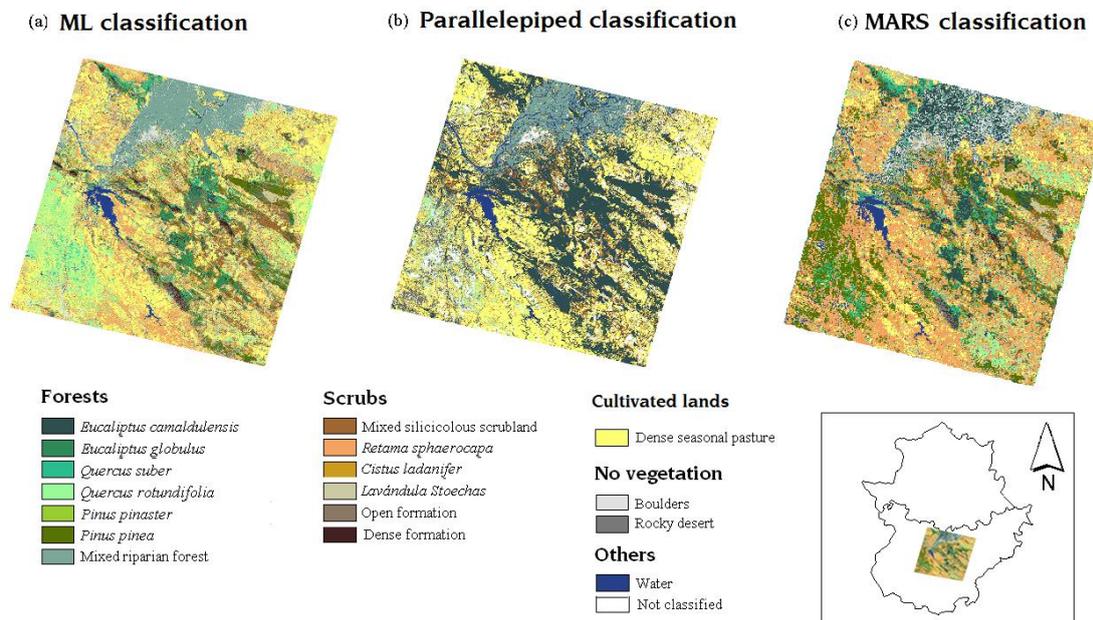
Probability maps were calculated during the classification process, so it was not necessary to recalculate them.

4. Results and Discussion

4.1. Classification maps

Figure 6 shows the three different classification maps obtained.

Figure 6. (a) ML classification map, (b) Parallelepiped classification map, (c) MARS classification map.



These classification maps cannot be used in and of themselves to evaluate accuracy. It is necessary to perform a separate accuracy study, and thus as mentioned before, ROC curves and AUC statistics were calculated for this purpose.

4.2. Comparison of classification maps

Accuracy assessments of the probability maps were implemented; Table 4 summarises the AUC statistic for each method.

Table 4. Area under the ROC curve (AUC) statistics.

Forest map cod.	Legend	ROI areas (km ²)	MARS AUC	ML AUC	Parallelepiped AUC
999	Water	33.3	0.952	0.945	0.793
547	Mixed silicicolous scrubland	28.4	0.852	0.813	0.754
507	Mixed riparian forest	11.6	0.936	0.936	0.814
458	Dense seasonal pasture	52.2	0.844	0.714	0.687

Table 4. Cont.

454	Open formation	2.8	0.978	0.929	0.954
453	Dense formation	5.6	0.985	0.961	0.971
337	Boulders	0.8	0.963	0.969	0.791
329	Rocky desert	23.5	0.890	0.884	0.701
309	<i>Retama sphaerocapa</i>	69.6	0.724	0.699	0.670
303	<i>Cistus ladanifer</i>	10.6	0.856	0.826	0.728
221	<i>Lavandula stoechas</i>	8.1	0.906	0.898	0.657
	<i>Eucaliptus</i>				
62	<i>camaldulensis</i>	140.6	0.908	0.856	0.834
61	<i>Eucaliptus globulus</i>	5.7	0.949	0.939	0.870
46	<i>Quercus suber</i>	26.0	0.864	0.841	0.766
45	<i>Quercus rotundifolia</i>	544.2	0.688	0.577	0.600
26	<i>Pinus pinaster</i>	0.7	0.957	0.976	0.903
23	<i>Pinus pinea</i>	12.1	0.952	0.960	0.924

Based on AUC statistics (Table 4), the MARS method obtains better results in 14 out of 17 classes, which confirms the suitability of this method for multi-spectral classification. We can conclude from the results shown in Table 4 that MARS is better than MV and PP but these are subjective interpretations. We performed three Wilcoxon paired-sample tests to estimate the P values for each comparison. Results are: MARS vs MV ($P < 0.005$), MARS vs PP ($P < 0.001$), MV vs PP ($P < 0.001$). We conclude that MARS is statistically a better method with a significance level better than 0.005.

5. Conclusions

This paper has presented the novel application of the MARS classification method to a conventional multi-spectral image. The results show that this method can be useful to improve some classifications. The main advantages of this method are that MARS is a non-parametric method that can be used without prior assumptions regarding the statistical distribution of the data. In addition, MARS is not severely affected by data collinearity, a common issue in satellite multiband imagery, and its functions are clear and transparent, especially when compared to the “black box” functions of other methods.

We compared the accuracy of various methods using the AUC statistic, an independent and objective test that can be applied to very different classification methods and that is independent of the cut-off classification threshold. The AUC statistic is a tool for evaluating classification performance that has been widely used in other disciplines, but only infrequently employed in remote sensing.

As the main conclusion of this study, MARS is a robust classification method that can be used in remote sensing without any disadvantages or apparent problems: its non-parametric nature, its transparency in terms of the relevant variables, and its adaptability to the data give it great potential as a multi-spectral classifier.

Our future work will focus on the application of hyperspectral data to MARS in order to deepen our analysis of its performance by using highly correlated bands and very large numbers of data.

Acknowledgements

The authors are grateful for the wise suggestions by the two anonymous referees. This paper has been co-funded by the Junta de Extremadura (Consejería de Educación, Ciencia y Tecnología — II Plan Regional de Investigación, Desarrollo Tecnológico e Innovación de Extremadura) and FEDER (Fondo Europeo de Desarrollo Regional).

References

1. Ruiz, L.A. *Introducción al tratamiento digital de imágenes en teledetección*; Servicio de publicaciones de la Universidad Politécnica de Valencia: Valencia, Spain, 1998.
2. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870.
3. Sarunas, R. On dimensionality, sample size, and classification error of nonparametric linear classification algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 667–671.
4. Cortijo, F.J.; de la Blanca, N.P. The performance of regularized discriminant analysis versus non-parametric classifiers applied to high-dimensional image classification. *Int. J. Remote Sens.* **1999**, *20*, 3345–3365.
5. Hill, T.; Lewicki, P. *Statistics: Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*; StatSoft: Tulsa, OK, USA, 2006.
6. Friedman, J.H. *Multivariate adaptive regression splines*; Salford Systems: San Diego, CA, USA, 1990.
7. Bonilla, M.; Olmeda, I.; Puertas, R. Modelos paramétricos y no paramétricos en problemas de credit scoring. *Revista Española de Financiación y Contabilidad* **2003**, *31*, 833–869.
8. Heikkinen, R.K.; Luoto, M.; Kuussaari, M.; Toivonen, T. Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landsc. Urban Plan.* **2007**, *79*, 347–357.
9. Leathwick, J.R.; Elith, J.; Hastie, T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Modell.* **2006**, *199*, 188–196.
10. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199.
11. Felicísimo, A.M.; Gómez, A.; Muñoz, J. Potential distribution of forest species in dehesas of Extremadura. In *Sustainability of agrosilvopastoral systems: Dehesas, montados*; Schnabel, S., Ferreira, A., Eds.; Catena Verlag: Reiskirchen, Germany, 2004; pp. 231–246.
12. Moisen, G.G.; Frescino, T.S. Comparing five modelling techniques for predicting forest characteristics. *Ecol. Modell.* **2002**, *157*, 209–225.
13. Muñoz, J.; Felicísimo, A.M. Comparison of statistical methods commonly used in predictive modelling. *J. Veg. Sci.* **2004**, *15*, 285–292.
14. Francis, L.; FCAS; MAAA. *Martian Chronicles: Is MARS Better than Neural Networks?* Available online: <http://www.salfordsystems.com/publications.php> (accessed June 8, 2008).

15. Dwinnell, W. *Exploring MARS: An Alternative to Neural Networks*. Available online: <http://www.salfordsystems.com/publications.php> (accessed June 8, 2008).
16. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*; Springer-Verlag: New York, NY, USA, 2001.
17. Duda, R.; Hart, P.; Stork, D. *Pattern Classification*; Wiley: New York, NY, USA, 2000.
18. Molina, R.; de la Blanca, N.P.; Taylor, C.C. Modern statistical techniques. In *Machine Learning, Neural and Statistical Classification*; Michie, D.J., Spiegelhalter D.J., Eds.; Ellis Horwood: Upper Saddle River, NJ, USA, 1994; pp. 29–49.
19. Solberg, A.H. Flexible nonlinear contextual classification. *Pattern Recognit. Lett.* **2004**, *25*, 1501–1508.
20. Kato, M.; Syoji, M.; Oyanagi, M. Aster status and data application use. In *Proceedings of the 22nd Asian Conference on Remote Sensing*, Singapore, November, 2001.
21. Stefanov, W.L.; Netzband, M. Assessment of ASTER land cover and MODIS NDVI data at multiple scales for ecological characterization of an arid urban center. *Remote Sens. Environ.* **2005**, *99*, 31–43.
22. Sobrino, J.A.; Jiménez-Muñoz, J.C.; Balick, L.; Gillespie, A.R.; Sabol, D.A.; Gustafson, W.T. Accuracy of ASTER level-2 thermal-infrared standard products of an agricultural area in Spain. *Remote Sens. Environ.* **2007**, *106*, 146–153.
23. Muukkonen, P.; Heiskanen, J. Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sens. Environ.* **2005**, *99*, 434–447.
24. Barry, R.G. The status of research on glaciers and global glacier recession: A review. *Prog. Phys. Geogr.* **2006**, *30*, 285–306.
25. Khromova, T.E.; Osipova, G.B.; Tsvetkov, D.G.; Dyurgerov, M.B.; Barry, R.G. Changes in glacier extent in the eastern Pamir, central Asia, determined from historical data and ASTER imagery. *Remote Sens. Environ.* **2006**, *102*, 24–32.
26. Tralli, D.M.; Blom, R.G.; Zlotnicki, V.; Donnellan, A.; Evans, D.L. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS-J. Photogramm. Remote Sens.* **2005**, *59*, 185–198.
27. Gillespie, T.W.; Chu, J.; Frankenberg, E.; Thomas, D. Assessment and prediction of natural hazards from satellite imagery. *Prog. Phys. Geogr.* **2007**, *31*, 459–470.
28. Liu, Y.; Hiyama, T.; Yamaguchi, Y. Scaling of land surface temperature using satellite data: A case examination on ASTER and MODIS products over a heterogeneous terrain area. *Remote Sens. Environ.* **2006**, *105*, 115–128.
29. Pu, R.; Gong, P.; Michishita, R.; Sasagawa, T. Assessment of multi-resolution and multi-sensor data for urban surface temperature retrieval. *Remote Sens. Environ.* **2006**, *104*, 211–225.
30. Jianwen, M.; Bagan, H. Land-use classification using ASTER data and self-organized neural networks. *Int. J. Appl. Earth Obs. Geoinf.* **2005**, *7*, 183–188.
31. Fawcett, T. *ROC Graphs: Notes and Practical Considerations for Researchers*; HP Laboratories: Palo Alto, CA, USA, 2004.
32. Huang, J.; Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 299–310.

33. Ling, C.X.; Huang, J.; Zhang, H. AUC: A statistically consistent and more discriminating measure than accuracy. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, August, 2003.
34. Marchant, J.A.; Andersen, H.J.; Onyango, C.M. Evaluation of an imaging sensor for detecting vegetation using different waveband combinations. *Comput. Electron. Agric.* **2001**, *32*, 101–117.
35. Tu, T.M.; Chen, C.H.; Chang, C.I. A posteriori least squares orthogonal subspace projection approach to desired signature extraction and detection. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 127–139.
36. Abrams, M.; Hook, S. *ASTER User Handbook*; JPL-NASA: Pasadena, CA, USA, 2001.
37. de la Torre, J.R. *Mapa forestal de España. Escala 1:200.000. Memoria general*; Ministerio de agricultura, pesca y alimentación: Madrid, Spain, 1990.
38. Nedjah, N.; de Macedo Mourelle, L. *Fuzzy Systems Engineering: Theory and Practice*; Springer: New York, NY, USA, 2005.
39. Stone, C.J.; Hansen, M.H.; Kooperberg, C.; Truong, Y.K. Polynomial splines and their tensor products in extended linear modelling: 1994 Wald memorial lecture. *Ann. Stat.* **1997**, *25*, 1371–1470.
40. Girard, M.C. *Processing of Remote Sensing Data*. Taylor & Francis: Paris, France, 2003.
41. Tso, B.; Mather, P.M. *Classification Method for Remotely Sensed Data*; CRC: New York, NY, USA, 2001.
42. Richards, J.A.; Jia, X. *Remote Sensing Digital Image Analysis: An Introduction*; Springer-Verlag: New York, NY, USA, 2006.

© 2009 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).