Expert Systems with Applications 37 (2010) 8432–8444

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A logistic radial basis function regression method for discrimination of cover crops in olive orchards

C. Hervás-Martínez [a], P.A. Gutiérrez [a,*], J.M. Peña-Barragán [b], M. Jurado-Expósito [b], F. López-Granados [b]

[a] Department of Computer Science and Numerical Analysis, University of Cordoba, Campus de Rabanales, 14071-Cordoba, Spain
[b] Institute for Sustainable Agriculture, CSIC, Apdo. 4084, 14080-Córdoba, Spain

ARTICLE INFO

Keywords:
Spectral signature
Cover crops
Evolutionary neural networks
Radial basis functions
MultiLogistic regression
Multiclassification

ABSTRACT

Olive (*Olea europaea* L.) is the main perennial Spanish crop. Soil management in olive orchards is mainly based on intensive and tillage operations, which have a great relevancy in terms of negative environmental impacts. Due to this reason, the European Union (EU) only subsidizes cropping systems which require the implementation of conservation agro-environmental techniques such as cover crops between the rows. Remotely sensed data could offer the possibility of a precise follow-up of presence of cover crops to control these agrarian policy actions, but firstly, it is crucial to explore the potential for classifying variations in spectral signatures of olive trees, bare soil and cover crops using field spectroscopy. In this paper, we used hyperspectral signatures of bare soil, olive trees, and sown and dead cover crops taken in spring and summer in two locations to evaluate the potential of two methods (MultiLogistic regression with Initial and Radial Basis Function covariates, MLIRBF; and SimpleLogistic regression with Initial and Radial Basis Function covariates, SLIRBF) for classifying them in the 400–900 nm spectrum. These methods are based on a MultiLogistic regression model formed by a combination of linear and radial basis function neural network models. The estimation of the coefficients of the model is carried out basically in two phases. First, the number of radial basis functions and the radii and centres' vector are determined by means of an evolutionary neural network algorithm. A maximum likelihood optimization method determines the rest of the coefficients of a MultiLogistic regression with a set of covariates that include the initial variables and the radial basis functions previously estimated. Finally, we apply forward stepwise techniques of structural simplification.

We compare the performance of these methods with robust classification methods: Logistic Regression without covariate selection, MLogistic; Logistic Regression with covariate selection, SLogistic; Logistic Model Trees algorithm (LMT); the C4.5 induction tree; Naïve Bayesian tree algorithm (NBTree); and boosted C4.5 trees using AdaBoost.M1 with 10 and 100 boosting iterations. MLIRBF and SLIRBF models were the best discriminant functions in classifying sown or dead cover crops from olive trees and bare soil in both locations and seasons by using a seven-dimensional vector with green (575 nm), red (600, 625, 650 and 675 nm), and near-infrared (700 and 725 nm) wavelengths as input variables. These models showed a correct classification rate between 95.56% and 100% in both locations and seasons. These results suggest that mapping covers crops in olive trees could be feasible by the analysis of high resolution airborne imagery acquired in spring or summer for monitoring the presence or absence of cover crops by the EU or local administrations in order to make the decision on conceding or not the subsidy.

1. Introduction

Olive (*Olea europaea* L.) is the main perennial Spanish crop with a total area of about 2.5 M ha, of which 1.5 are in Andalusia (southern Spain; MAPYA, 2007). Soil management in olive orchards is mainly based on intensive tillage operations, which have a great relevancy in terms of the increase of atmospheric CO_2, desertification, erosion and land degradation (Hill, Megier, & Mehl, 1995; Schlesinger, 2000). Due to these negative environmental impacts, the European Union (EU) only subsidizes cropping systems which require the implementation of conservation agro-environmental techniques such as cover crops in olive orchards (Andalusian Administration Regulation, 2007). Traditionally, olive trees are separated 10–12 m each other and cover crops are 4–6 m wide. These include the cultivation with cover crops between the rows, usually grass species (sown cover crops), or recycled crop residues (dead cover crops). Sown cover crops are planted in autumn each year (mid November in Mediterranean conditions) and must be managed when the plants have completed their vegetative cycle by

* Corresponding author. Tel.: +34 957 218 349; fax: +34 957 218 630.
E-mail address: pagutierrez@uco.es (P.A. Gutiérrez).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.05.035

using either herbicides applied at the end of spring, i.e. end of March in our conditions, or through various passes of the chain mower just before cover crop starting to compete for water and nutrients with olive trees.

To control these agrarian policy actions a precise follow-up or monitoring of presence or absence of cover crops is required by the EU and Andalusian administrations. Current methods to estimate the cover crop soil coverage consist of sampling and ground visits to only 1% of the total olive orchards at any time from mid-March to late-June. However, this procedure is time-consuming and very expensive, delivering inconsistent result due to the fact that it covers relatively small areas or only target fields, and it does not sample inaccessible areas. Remotely sensed data may offer the ability to efficiently identify and map crops and cropping methods over large areas (South, Qi, & Lusch, 2004). These techniques may imply lower costs, faster work and better reliability than ground visits. At the same time, the accuracy of the thematic map is extremely important because this map could be used as a tool to help the administrative follow-up to make the decision on conceding or not the subsidy.

To detect and map olive trees and cover crops, it is necessary that suitable differences exist in spectral reflectance among them and bare soil. As part of an overall research programme to investigate the opportunities and limitations of remote sensed imagery in mapping accurately olive trees, bare soil, and cover crops, it is crucial to explore the potential for identifying variations in their spectral signatures using field spectroradiometry by analysing the ability of the discrimination at distinct cover crop phenological stages. Such an approach should indicate the wavelengths suitable for land use discrimination and classification. Previous works have demonstrated that the spectral signature of any plant species varies with time and therefore it is intrinsically related to the specific phenological stage when it was taken (López-Granados, Jurado-Expósito, Peña-Barragán, & García-Torres, 2006; Peña-Barragán, López-Granados, Jurado-Expósito, & García-Torres, 2006; Schmidt & Skidmore, 2003). To predetermine a subset of narrow wavelengths without losing any essential information of spectral signatures, several statistical methods have been applied. For instance, artificial neural networks to discriminate nitrogen status in corn (Goel et al., 2003) and to classify grass weeds in wheat in field conditions (López-Granados et al., 2008). Moreover, computational methods have been presented as very useful tools for improving decision making by olive oil growers (González-Andujar, 2009) or by pepper growers (González-Diaz, Martínez-Jimenez, Bastida, & González-Andujar, 2009).

Multispectral and medium spatial resolution satellite imagery such as Landsat Thematic Mapper and Spot has often proven to have an insufficient or inadequate accuracy for detailed vegetation studies (Harvey & Hill, 2001). Hyperspectral sensors offer an improvement over multispectral: hyperspectral sensors have many narrow and contiguous wavebands, usually around 25 nm width, whereas multispectral sensors collect data for several (3–7) broad bands. Hyperspectral scanner systems can detect small or local variations in absorption features that might otherwise be masked within the broader multispectral scanner systems (Koger, Shaw, Reddy, & Bruce, 2004; Schmidt & Skidmore, 2003). New satellites are being developed to provide high resolution hyperspectral data and high spatial resolution with the minimum spatial resolution (at least 1 m spatial resolution) to classify olive orchards at the tree scale and cover crops between trees. Airborne hyperspectral sensors such as Compact Airborne Spectral Imager (CASI) and artificial neural networks have already been considered to be a useful data source, which accurately determines agronomic variables such as prediction of corn yield (Uno et al., 2005) or detection of weeds (Karimi et al., 2005). The potential advantages are that hyperspectral satellite imagery usually cover higher surface and hyperspectral airborne sensors have superior flight versatility. CASI is capable of acquiring data up to 288 wavelengths at the spectral range of 400–1000 nm (visible and near-infrared) at 1.9 nm intervals. Moreover, CASI spectral collection is user programmable, and, if proper altitudes are maintained, it can achieve resolutions of 0.5–1 m, which are particularly useful for classifying vegetation classes. Thus, for effective olive tree-cover crop-bare soil discrimination, the identification of subtle differences in the spectral signatures at different seasons is required and it is also necessary the classification of the different spectra into the specific group to which they belong.

The problem of assigning a specific group to the different spectra analysed is treated in this paper using a pattern recognition technique. Multi-class pattern recognition is a problem of building a system that accurately maps an input feature space to an output space of more than two pattern classes. Whereas a two-class classification problem is well understood, multi-class classification is relatively less-investigated. In general, the extension from two-class to the multi-class pattern classification problem is not trivial, and often leads to unexpected complexity or weaker performances. This paper presents a MultiLogistic generalized regression where the linear predictor is replaced or extended using a non-parametric neural network model. The ideas introduced follow those presented in a recently proposed combination of neural networks and logistic regression (Gutiérrez, López-Granados, Peña-Barragán, Jurado-Expósito, & Hervás-Martínez, 2008a; Gutiérrez et al., 2009; Hervás-Martínez & Martínez-Estudillo, 2007; Hervás-Martínez, Martínez-Estudillo, & Carbonero-Ruz, 2008; Torres, Hervás-Martínez, & García, 2009) based on the hybridization of a linear MultiLogistic regression model and a non-linear Product-Unit Neural Network model for binary and multi-class classification problems. The presented methodology named Logistic Regression with Initial and Radial Basis Function covariates, LIRBF, combines different elements such as MultiLogistic regression, MLR, radial basis neural networks, RBFNNs, and evolutionary algorithms, EAs.

Logistic regression was used for the classification of spectral signatures because the LR may be preferred when the data distribution is not normal, or the group sizes are unequal (Neupane, Sharma, & Thapa, 2002). In Pu and Gong (2004) and Van Deventer, Ward, Gowda, and Lyon (1997), LR is applied for covariate selection from multispectral data used for binary classification. The results from these papers advocate the utility of the LR as a potential approach for the soft classification similar to the other recent ones such as the neural networks (Foody & Arora, 1996), possibilistic c-means clustering (Ibrahim, Arora, & Ghosh, 2005), and decision tree regression (Xu, Watanachaturaporn, Varshney, & Arora, 2005). A hard classification can be produced by assigning the spectrum with the class having a maximum probability.

Although LR is a simple and useful procedure, we cannot frequently formulate the stringent assumption of additive and purely linear effect of the covariates of the predictor function, so it is interesting to hybridize this classification model with other soft computing techniques (Kin, 2009). In this way, our technique overcomes these difficulties by augmenting the input covariates with new RBF covariates. From the opposite point of view, adding linear terms to a RBFNN in the predictor functions of a logistic regression yields models that are simpler and easier to interpret than models with only RBF covariates. In particular, if a covariate appears only linearly in the logistic final model, then the model is a traditional parametric model with regard to that covariate. A second reason is to reduce the variance associated with the overall modelling procedure, and a third is to reduce the likelihood of ending up with unnecessary terms in the final model. RBFNNs are an alternative to traditional Multilayer Perceptrons (Fukunaga, 1999; Lee, Chiang, Shih, & Tsai, 2009) and are based on localized hidden nodes (which have high non-zero outputs over only a localized region of the input space), instead of projection ones (which have high non-zero outputs over a large region of the input space).

RBFNNs have been found to be very helpful to many engineering problems because: (1) they are universal approximators (Park & Sandberg, 1991); (2) they have more compact topology than other neural networks (Lee & Kil, 1991); and (3) their learning speed is fast because of their locally tuned neurons (Moody & Darken, 1989). The learning procedure of a RBFNN mainly includes two parts: one is the adjustment of the connection weights, and the other is the modification of the parameters of the RBF units, namely, the hidden centres and the RBF widths or radii.

MLR models are in general fit by maximum likelihood, where the Newton–Raphson algorithm is the traditional way to estimate the maximum "a posteriori" parameters. Usually, the algorithm converges, since the log-likelihood is concave. However, in our approach, the non-linearity of the RBFs with regard to the centres and radii implies that the corresponding Hessian matrix is generally indefinite and the likelihood could have local maxima. These reasons justify, in our opinion, the use of an EA (Goldberg, 1989) as an alternative heuristic procedure to estimate the parameters of the model.

The estimation of the coefficients is carried out basically in two steps. In a first step, an EA, which we have called Evolutionary RBF (ERBF) algorithm, determines the number of RBFs in the model and their corresponding centres and radii. This step can be seen as a global search in the coefficients' model space. In a second step, once the basis functions have been determined by the ERBF algorithm, we consider a transformation of the input space by adding the non-linear transformations of the input variables given by the obtained basis functions. The final model is linear in the set of covariates formed by these new covariates and the initial covariates. Now, the Hessian matrix is definite and fitting proceeds with the standard maximum likelihood optimization method. Finally, we use a forward stepwise procedure, adding variables sequentially to form the model and including a cross-validation for assessing the test performance.

This methodology is tested for discriminating cover crops in olive orchards as affected by their phenological stage using a high-resolution field spectroradiometer. The objectives of this study were: (1) to determine the hyperspectral reflectance curves of sown (live and desiccated) and dead cover crops, bare soil, and olive trees, (2) to select the best hyperspectral wavelengths and phenological stages to assess the different classification models based in the LIRBF methodology for reaching the best discrimination approach, (3) to compare the accuracy performance for a spectrum classification into the group to which it belongs, and (4) to establish the misclassification percentage and validate the classification accuracy of this analysis by using a 10-fold approach for cross-validation procedure. Five were the models tested: (a) evolutionary radial basis functions neural networks, ERBF; (b) MultiLogistic regression with RBF covariates, MLRBF; (c) SimpleLogistic regression with RBF covariates, SLRBF; (d) MultiLogistic regression with Initial and RBF covariates, MLIRBF; (e) SimpleLogistic regression with Initial and RBF covariates, SLIRBF. These objectives would provide information to programme the suitable wavelengths of airborne hyperspectral sensors such as CASI for administrative follow-up and monitoring of agro-environmental measures in olive orchards under conservation agriculture.

The remainder of the paper is structured as follows: Section 2 is based in Materials and methods (the study sites, the spectral readings and the description of the LIRBF methods and the comparison methods). In Section 3, the results and an associated discussion are provided and, finally, the work is summarized and conclusions drawn in Section 4.

2. Materials and methods

2.1. Study sites and spectral readings

The study was conducted in Andalusia, southern Spain, in two locations named "Cortijo del Rey" and "Matallana" in early spring and early summer. Sown cover crop species composition was made of *Lolium rigidum* Gaudin. (ryegrass) in "Cortijo del Rey" and of *Hordeum murinum* L. (barley) in "Matallana". Before the herbicide treatment, sown cover crops in spring had the typical green colour of the vegetative growing phase (life cover crops, Fig. 1a), while cover crops were yellow-like colour in early summer as they had been previously desiccated with herbicide (desiccated cover crops, Fig. 1b). Dead cover crops (Fig. 1c) consisted of remaining of the corresponding olive spring pruning. Covers crops had been established for over 6 years in both locations. Twenty spectral signatures of live, desiccated and dead cover crops, and ten of olive trees and bare soil were taken in 2007, on 22 and 23 March, and on 23 and 24 June, in "Cortijo del Rey" and "Matallana", respectively (this number of spectral signatures and the main characteristics of each dataset can be seen in Table 1).

Measurements were collected using an ASD Handheld FieldSpec Spectroradiometer (Analytical Spectral Devices, Inc., Boulder) under sunny conditions between 12.00 and 14.00 h (Salisbury, 1999) and measuring an area of about 0.15–0.20 m^2. A telescopic pole (Fig. 2) was used for positioning the spectroradiometer at 80–100 cm above the olive tree canopy. In addition, each measurement was georeferenced using the sub-meter differential DGPS TRIMBLE PRO-XRS (Trimble, Sunnyvale, CA) provided with TDC-1 unit for further remote sensing analysis. The spectral data were converted into reflectance, which is the ratio of energy reflected off the target to an energy incident on the target, with reference to a barium sulphate standard (Spectralon, Labsphere, North Sutton, NH) before and immediately after each measurement. The hyperspectral range was between 325 and 1075 nm (1.5 nm bandwidth). However, the reflectance spectra data were noisy at the extremes on the range and only the measurements between 400 and 900 nm were analysed. Previous studies have shown that neighbouring wavelengths can frequently provide similar information (Peña-Barragán et al., 2006; Thenkabail, Enclona, Ashton, & Van Der Meer, 2004; Thenkabail, Smith, & De-Pauw, 2000). Thus, the high spectral resolution of hyperspectral measurements collected was demeaned and averaged to represent 20-, 25-nm-wide measurements between 400 and 900 nm. These 20 wavelengths of the spectral signatures were considered in a previous cross-validation hold out analysis which basically showed that only the seven wavelengths between 575 and 725 nm were always selected in the discriminant functions. Therefore, we used a seven-dimensional vector with green (575 nm), red (600, 625, 650, and 675 nm), and near-infrared (700 and 725 nm) wavelengths as the input variables. Results from 20 and 7 wavelengths did not show relevant differences in terms of accuracy, but in number of links of the discrimination functions (data not shown), leading to a worse model interpretation.

2.2. Methods

The methods presented in this paper are based on a MultiLogistic regression model (LIRBF), whose coefficients are obtained using a hybrid learning procedure.

2.2.1. Logistic regression using radial basis function and initial covariates model (LIRBF)

In classification problems, measurements $\mathbf{x}_i$, $i = 1, 2, \ldots, k$, are taken on a single individual (or object), and the individuals are to

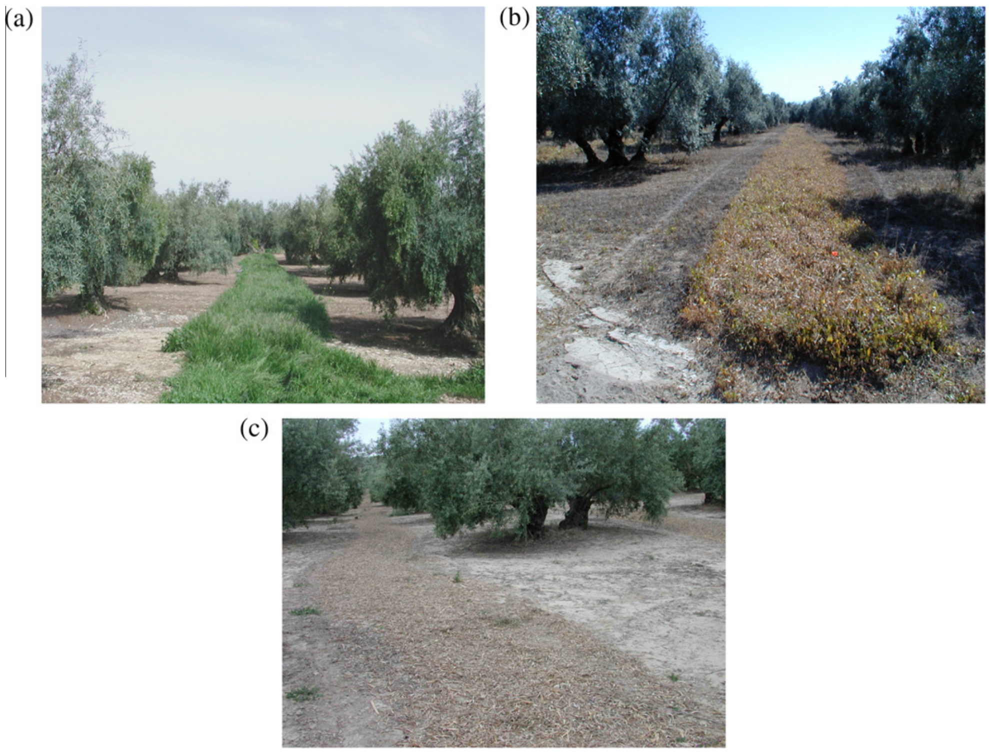

Fig. 1. Images of the different elements to discriminate: (a) live cover crops in March; (b) desiccated cover crops in June; and (c) dead cover crops.

Table 1
Main characteristics of each location and season tested: total number of instances or spectral signatures (# Instances), number of wavelengths representing each spectral signature (# Inputs), number of classes to discriminate (# Classes) and number of signatures per each class (Distribution).

Location and seasons	# Instances	# Inputs	# Classes	Distribution
"Cortijo del Rey" spring	80	7	4	(40, 20, 10, 10)
"Cortijo del Rey" summer	50	7	4	(10, 20, 10, 10)
"Matallana" spring	60	7	4	(20, 20, 10, 10)
"Matallana" summer	60	7	4	(20, 20, 10, 10)

be classified into one of J classes on the basis of these measurements. It is assumed that J is finite, and the measurements $\mathbf{x}_i$ are random observations from these classes. Based on the training sample $D = \{(\mathbf{x}_n, \mathbf{y}_n); n = 1, \ldots, N\}$, where $\mathbf{x}_n = (x_{1n}, \ldots, x_{kn})$ is the vector of measurements taking values in $\Omega \subset R^k$, and $\mathbf{y}_n$ is the class level of the n-th individual, we wish to find a decision function $C:\Omega \to \{1, 2, \ldots, J\}$ for classifying the individuals. A misclassification occurs when the decision rule C assigns an individual of the training sample to a class j when it is actually coming from a class $l \neq j$. To evaluate the performance of the classifiers, we define the Correct Classification Rate by $\mathrm{CCR} = \frac{1}{N}\sum_{n=1}^{N} I(C(\mathbf{x}_n) = \mathbf{y}_n)$, where $I(.)$ is the zero-one loss function and the common technique of representing the class levels using a "1-of-J" encoding vector $\mathbf{y} = (y^{(1)}, y^{(2)}, \ldots, y^{(J)})$ is adopted, such as $y^{(l)} = 1$ if $\mathbf{x}$ corresponds to an example belonging to class l and $y^{(l)} = 0$ otherwise. A good classifier tries to achieve the highest possible CCR in a given problem for the generalization set. It is usually assumed that the training data are independent and identically distributed sample from an unknown probability distribution.

Suppose that the conditional probability that $\mathbf{x}$ belongs to class l verifies: $p_l = p(y^{(l)} = 1|\mathbf{x}) > 0,\ l = 1,\ 2,\ \ldots,\ J,\ \mathbf{x} \in \Omega$. Let $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k, \mathbf{Y})$ be a set of variables, under a multinomial logistic regression, the probability that $\mathbf{x}$ belongs to class l, $l = 1, \ldots, J$ is based in the equation:

$$p_l(\mathbf{x},\ \boldsymbol{\beta}_l) = \frac{\exp(f_l(\mathbf{x},\ \boldsymbol{\beta}_l))}{\sum_{l=1}^{J}\exp(f_l(\mathbf{x},\ \boldsymbol{\beta}_l))},$$

Fig. 2. Measuring olive tree canopy reflectance using a hand-held spectroradiometer with a telescopic pole under sunny conditions. Every olive tree was georeferenced for future remote sensing investigations.

where the prediction function $f_l(\mathbf{x}, \boldsymbol{\beta}_l) = \boldsymbol{\beta}_l^{\mathrm{T}}\mathbf{x}$ is linear in the $\mathbf{x}$ covariates, $\boldsymbol{\beta}_l = (\beta_{l,0}, \beta_{l,1}, \ldots, \beta_{l,m})$, and the $\mathbf{x}_0 = 1$ value has been added to the input covariates vector $\mathbf{x}$. The vector components $\boldsymbol{\beta}_l$ are estimated from the training data set D. If we use the probability axiom of the normalization, we have $f_J(\mathbf{x}, \boldsymbol{\theta}_J) = 0$, and, in this way, we reduce the number of parameters to be estimated.

The LIRBF model is based on the combination of the standard initial covariates and non-linear radial basis functions transformed covariates. The model in matrix form is given by the following expression:

$$f_l(\mathbf{x}, \boldsymbol{\theta}_l) = \boldsymbol{\alpha}_l^{\mathrm{T}}\mathbf{x} + \boldsymbol{\beta}_l^{\mathrm{T}}\mathbf{B}(\mathbf{x}, \mathbf{W}), \quad l = 1, \ldots, J-1, \tag{1}$$

where $\mathbf{x} = (1, x_1, \ldots, x_k)$ and $\mathbf{B}(\mathbf{x}, \mathbf{W}) = (B_1(\mathbf{x}, \mathbf{w}_1), \ldots, B_m(\mathbf{x}, \mathbf{w}_m))$ for

$$B_j = B_j(\mathbf{x}, \mathbf{w}_j) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{r_j^2}\right), \quad j = 1, \ldots, m. \tag{2}$$

Then, the prediction function of the LIRBF model, is

$$f_l(\mathbf{x}, \boldsymbol{\theta}_l) = \alpha_0^l + \sum_{i=1}^{k} \alpha_i^l x_i + \sum_{j=1}^{m} \beta_j^l \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{r_j^2}\right),$$
$$l = 1, 2, \ldots, J = 1.$$

Let $\boldsymbol{\theta}_l = (\boldsymbol{\alpha}^l, \boldsymbol{\beta}^l, \mathbf{W})$ be the parameters of the model, where $\boldsymbol{\alpha}^l = (\alpha_0^l, \alpha_1^l, \ldots, \alpha_k^l)$ and $\boldsymbol{\beta}^l = (\beta_1^l, \ldots, \beta_m^l)$ are the coefficients of the MultiLogistic regression model and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ are the parameters of the RBF transformations, $\mathbf{w}_j = (w_{j0}, w_{j1}, \ldots, w_{jk})$, $\mathbf{c}_j = (w_{j1}, \ldots, w_{jk})$ is the centre of the j-th Gaussian RBF and $r_j = w_{j0}$ is the corresponding radius.

To perform the maximum likelihood estimation of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{J-1})$, one can minimize the negative log-likelihood function:

$$\begin{aligned} L(\boldsymbol{\theta}) &= -\frac{1}{N}\sum_{n=1}^{N} \log p(\mathbf{y}_n|\mathbf{x}_n, \boldsymbol{\theta}) \\ &= \frac{1}{N}\sum_{n=1}^{N}\left[-\sum_{l=1}^{J-1} y_n^{(l)} f_l(\mathbf{x}_n, \boldsymbol{\theta}_l) + \log\sum_{l=1}^{J-1}\exp f_l(\mathbf{x}_n, \boldsymbol{\theta}_l)\right]. \end{aligned}$$

The classification rule coincides with the optimal Bayes' rule. In other words, an individual should be assigned to the class $C(\mathbf{x}) = \hat{l}$ which has the maximum probability, given the vector measurement $\mathbf{x}$, i.e. $\hat{l} = \arg\max_l p_l(\mathbf{x}, \hat{\boldsymbol{\theta}}_l)$ for $l = 1, \ldots, J-1$. The non-linearity of the model with regard to the $\boldsymbol{\theta}_l$ parameters and the indefinite character of the associated Hessian matrix do not recommend the use of gradient-based methods to maximize the log-likelihood function. Moreover, the optimal number of basis functions of the model (i.e. the number of hidden nodes in radial basis neural network) is unknown. Thus, the estimation of the vector parameter $\hat{\theta}$ is carried out by means of a hybrid procedure, combination of an EA and a standard maximum likelihood optimization method.

2.2.2. Estimation of the LIRBF coefficients

The process is structured basically in two steps. The first step obtains a RBFNN best model, and the second obtains the LIRBF model. Once the basis functions have been determined by the EA, we consider a transformation of the input space by adding the non-linear transformations of the input variables given by theses basis functions. The model is now linear in these new variables and the initial covariates. The remaining coefficient vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are calculated by the maximum likelihood optimization method.

2.2.2.1. Step 1: Evolutionary radial basis functions neural networks (ERBF). We apply an evolutionary neural network algorithm (which we have called Evolutionary RBF algorithm, ERBF) to find the basis functions:

$$\mathbf{B}(\mathbf{x}, \mathbf{W}) = (B_1(\mathbf{x}, \mathbf{w}_1), \ldots, B_m(\mathbf{x}, \mathbf{w}_m)),$$

corresponding to the non-linear part of $f(\mathbf{x}, \boldsymbol{\theta})$ in Eq. (1). We have to determine the number of basis functions m and the weight vector $\mathbf{W}$. To apply evolutionary neural network techniques, we consider RBFNNs with the standard structure: an input layer with a node for every input variable; a hidden layer with several nodes; and an output layer with nodes, one for each category. There are no connections between the nodes of a layer and none between the input and output layers either. The activation function of the j-th node in the hidden layer is given by Eq. (2). The activation function of the output node l is given by:

$$g_l(\mathbf{x}, \boldsymbol{\beta}^l, \mathbf{W}) = \beta_0^l + \sum_{j=1}^{m} \beta_j^l B_j(\mathbf{x}, \mathbf{w}_j),$$

where β_j^l is the weight of the connection between the hidden node j and the output node l. The transfer function of all output nodes is the identity function.

The weight vector $\mathbf{W}$ is estimated by means of the ERBF algorithm (detailed next) that optimizes the error function given by the negative log-likelihood for N observations associated with the RBFNN model:

$$L(\boldsymbol{\beta}, \mathbf{W}) = \frac{1}{N}\sum_{n=1}^{N}\left[-\sum_{l=1}^{J-1} y_n^{(l)} g_l(\mathbf{x}_n, \boldsymbol{\beta}^l, \mathbf{W}) + \log\sum_{l=1}^{J-1}\exp g_l(\mathbf{x}_n, \boldsymbol{\beta}^l, \mathbf{W})\right]. \tag{3}$$

Although in this step the evolutionary process obtains a concrete value for the $\boldsymbol{\beta}$ vector, we only consider the estimated weight vector $\hat{\mathbf{W}} = (\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \ldots, \hat{\mathbf{c}}_m, \hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m)$, which builds the basis functions.

Among the different paradigms of Evolutionary Computation, we have chosen Evolutionary Programming due to the fact that we are evolving artificial neural networks. The population-based algorithm for architectural design and the estimation of real-coefficients have points in common with other EAs in the bibliography (Angeline, Saunders, & Pollack, 1994; Martínez-Estudillo, Hervás-Martínez, Martínez-Estudillo, & García-Pedrajas, 2006a; Martínez-Estudillo, Martínez-Estudillo, Hervás-Martínez, & García-Pedrajas, 2006b; Yao & Liu, 1997). The search begins with an initial population and in all iterations the population is updated using a population-update algorithm. The population is subject to the operations of replication and mutation. Crossover is not used due to its potential disadvantages in evolving artificial networks (Angeline et al., 1994). The basic steps the ERBF algorithm are represented in Fig. 3.

A brief description of the algorithm is given next. The algorithm receives a training dataset D and returns the optimized RBFNN that minimizes the cross-entropy error $L(\boldsymbol{\beta}, \mathbf{W})$ given by Eq. (3). In this way the fitness function of an individual $g(\mathbf{x}, \boldsymbol{\beta}, \mathbf{W})$ is a strictly decreasing transformation of the error function $L(\boldsymbol{\beta}, \mathbf{W})$ given by $A(\boldsymbol{\beta}, \mathbf{W}) = \frac{1}{1+L(\boldsymbol{\beta}, \mathbf{W})}$, where $0 < A(\boldsymbol{\beta}, \mathbf{W}) \leqslant 1$ (Fig. 3, steps 2, 6a and 8). The algorithm evolves architectures and connection weights simultaneously, each individual being a fully specified RBFNN. The nets are represented using an object-oriented approach and the algorithm deals directly with the neural network phenotype. The population is initialized (Fig. 3, steps 1–5) in the following way: first we generate 5000 RBFNNs and evaluate them; then, we select the best 10% of these RBFNNs and improve their centres by using standard k-means clustering algorithm (Fukunaga, 1999). Each of the iteration of the algorithm is performed as follows: we evaluate the individuals, sort them by increasing fitness (Fig. 3, step 6b) and apply two different mutations (parametric mutation to the best 10% of the individuals and structural mutation to the best 90% of individuals minus one for considering the elitism). It is important to note that the algorithm always maintains the best individual (Fig. 3, step 6c), resulting in an elitist algorithm. The

ERBF Algorithm:

Input: Training Dataset D

Output: Best optimized RBFNN (p^{B})

1. $P^{\mathrm{I}} \leftarrow \{p_1^{\mathrm{I}}, \ldots, p_{5000}^{\mathrm{I}}\}$ // Each p_i^{I} is a randomly generated individual
2. $\forall p_i^{\mathrm{I}} \in P^{\mathrm{I}}, f_i^{\mathrm{I}} \leftarrow A(p_i^{\mathrm{I}})$ // Evaluate the fitness of the individuals
3. $P \leftarrow \{p_{(1)}, \ldots, p_{(5000)}\}, \quad (p_{(i)} \prec p_{(j)}) \Leftrightarrow (f_i^{I} > f_j^{I})$ // Sort individuals in P by increasing f_i^{I}
4. $P \leftarrow \{p_{(1)}, \ldots, p_{(500)}\}$ // Retain the best 500 individuals
5. $\forall p_{(i)} \in P, p_{(i)} \leftarrow k\text{-means}(p_{(i)})$ // Improve individuals' centres
6. **while not** Stop Condition
   a. $\forall p_i \in P, f_i \leftarrow A(p_i)$ // Evaluate the fitness of the individuals
   b. $P \leftarrow \{p_{(1)}, \ldots, p_{(500)}\}, \quad (p_{(i)} \prec p_{(j)}) \Leftrightarrow (f_i > f_j)$ // Sort individuals in P by increasing f_i
   c. $p^{\mathrm{B}} \leftarrow p_{(1)}$ // Store the best individual
   d. $P^{\mathrm{P}} \leftarrow \{p_{(1)}, \ldots, p_{(50)}\}$ // Parametric mutation parents (best 10% of individuals)
   e. $P^{\mathrm{S}} \leftarrow \{p_{(1)}, \ldots, p_{(449)}\}$ // Structural mutation parents (best 90% of individuals)
   f. $\forall p_{(i)}^{\mathrm{P}} \in P^{\mathrm{P}}, p_{(i)}^{\mathrm{P}} \leftarrow \text{parametricMutation}(p_{(i)}^{\mathrm{P}})$ // Apply parametric mutation
   g. $\forall p_{(i)}^{\mathrm{S}} \in P^{\mathrm{S}}, p_{(i)}^{\mathrm{S}} \leftarrow \text{structuralMutation}(p_{(i)}^{\mathrm{S}})$ // Apply structural mutation
   h. $P \leftarrow P^{\mathrm{P}} \cup P^{\mathrm{S}} \cup \{p^{\mathrm{B}}\}$ // Offspring including the elite
7. **end while**
8. $\forall p_i \in P, f_i \leftarrow A(p_i)$ // Evaluate the fitness of the individuals
9. $P \leftarrow \{p_{(1)}, \ldots, p_{(500)}\}, \quad (p_{(i)} \prec p_{(j)}) \Leftrightarrow (f_i > f_j)$ // Sort individuals in P by increasing f_i
10. $p^{\mathrm{B}} \leftarrow p_{(1)}$ // Store the best individual
11. **return** p^{B}

Fig. 3. Evolutionary radial basis functions algorithm (ERBF).

parametric mutation (Fig. 3, step 6f) alters the value of the weights of the neural net and it is accomplished for each coefficient c_{ji}, r_j of the $\mathbf{W}$ and β_j^l of the model with the same Gaussian noise. The structural mutation (Fig. 3, step 6g) modifies the number of radial basis functions of the RBFNNs and the number of connections of the model. For more details about this algorithm and the different mutations aforementioned, we recommend consulting some previous works where it is more deeply explained (Gutiérrez et al., 2008b; Martínez-Estudillo et al., 2006b).

As it can be observed, the ERBF algorithm includes several parameters that have to be defined in order to apply it. All these parameters are common for the four data sets analysed below. The first thing that we have to take into account is that we have done a simple linear rescaling of the input variables in the interval $[-2, 2]$, λ_i^* being the transformed variables. In this way, the centres c_{ji} are initialized in this interval and the coefficients β_j^l are initialized in the $[-5, 5]$ interval. The initial value of the radii r_j is obtained as a random value in the interval $(0, d_{\max}]$, where $d_{\max}$ is the maximum distance between two training input examples. The size of the population is $N = 500$. The minimum and maximum values for the number of hidden nodes in the hidden layer are 2 and 4, respectively. The stop criterion is reached whenever one of the following two conditions is fulfilled: a number of generations is reached (50 generations in our experiments, a low value that reduces the computational cost) or the variance of the fitness of the best ten percent of the population is less than10^{-4}.

2.2.2.2. Step 2: Optimization of the rest of coefficients. We transform the input space by including the non-linear basis functions obtained by the ERBF algorithm, $z_1 = B_1(\mathbf{x}, \hat{\mathbf{w}}_1), \ldots, z_m = B_m(\mathbf{x}, \hat{\mathbf{w}}_m)$, and then minimize the negative log-likelihood function for N observations defined in (2). Now, the Hessian matrix of the negative log-likelihood in the new variables $x_1, x_2, \ldots, x_k, z_1, \ldots, z_m$ is semi-definite positive. In this paper, two different algorithms have been considered for obtaining the maximum likelihood solution for the MultiLogistic regression model, both available in the WEKA machine learning workbench (Witten & Frank, 2005):

(1) *MultiLogistic*: It is an algorithm for building a multinomial logistic regression model with a ridge estimator to guard against over-fitting by penalizing large coefficients, based on work by Le Cessie and Van Houwelingen (1992). In order to find the coefficient matrix $\boldsymbol{\theta}$ for which $L(\boldsymbol{\theta})$ is minimized, a Quasi-Newton method is used. Specifically, the method used is the active-sets' method with Broyden–Fletcher–Goldfarb–Shanno (BFGS) update (Gill, Murray, & Wright, 1982).

(2) *SimpleLogistic*: This algorithm builds multinomial logistic regression models fitting them by using the LogitBoost algorithm, which was proposed by Friedman, Hastie, and Tibshirani (2000) for fitting additive logistic regression models. These models are a generalization of the (linear) logistic regression models described in Section 2.2.1.

This results in two different models, one with all $x_1, x_2, \ldots, x_k, z_1, \ldots, z_m$ covariates present in the model (MultiLogistic algorithm) and the other with only those variables selected by the SimpleLogistic algorithm. These two approaches will be called MLIRBF and SLIRBF, respectively. For comparison purposes, we have also considered the MultiLogistic regression models that are obtained with these two algorithms but constructed only from the non-linear transformations given by the RBFNN of the ERBF algorithm, i.e. $z_1, \ldots, z_m$. This results in two other approaches which we will call MLRBF and SLRBF.

2.3. Comparison to other classification methods

In this paper, the experimental design for the ERBF, MLIRBF,and SLIRBF methods was conducted using a 10-fold cross-validation procedure (Kohavi, 1995), with 10 repetitions per each fold using the seven wavelengths as the input variables. We compare our LIRBF approaches (MLIRBF and SLIRBF) to the results obtained using eight recent and competitive methodologies (Landwehr, Hall, & Frank, 2005): Logistic Regression without covariate selection, MLogistic; Logistic Regression with covariate selection, SLogistic; Logistic Model Tree algorithm (LMT) (Landwehr et al., 2005); the C4.5 induction tree (Quinlan, 1993); Naïve Bayesian tree algorithm: NBTree (Kohavi, 1996); boosted C4.5 trees using AdaBoost.M1 with 10 and 100 boosting iterations (Freund & Schapire, 1996), and our evolutionary radial basis functions neural networks, ERBF.

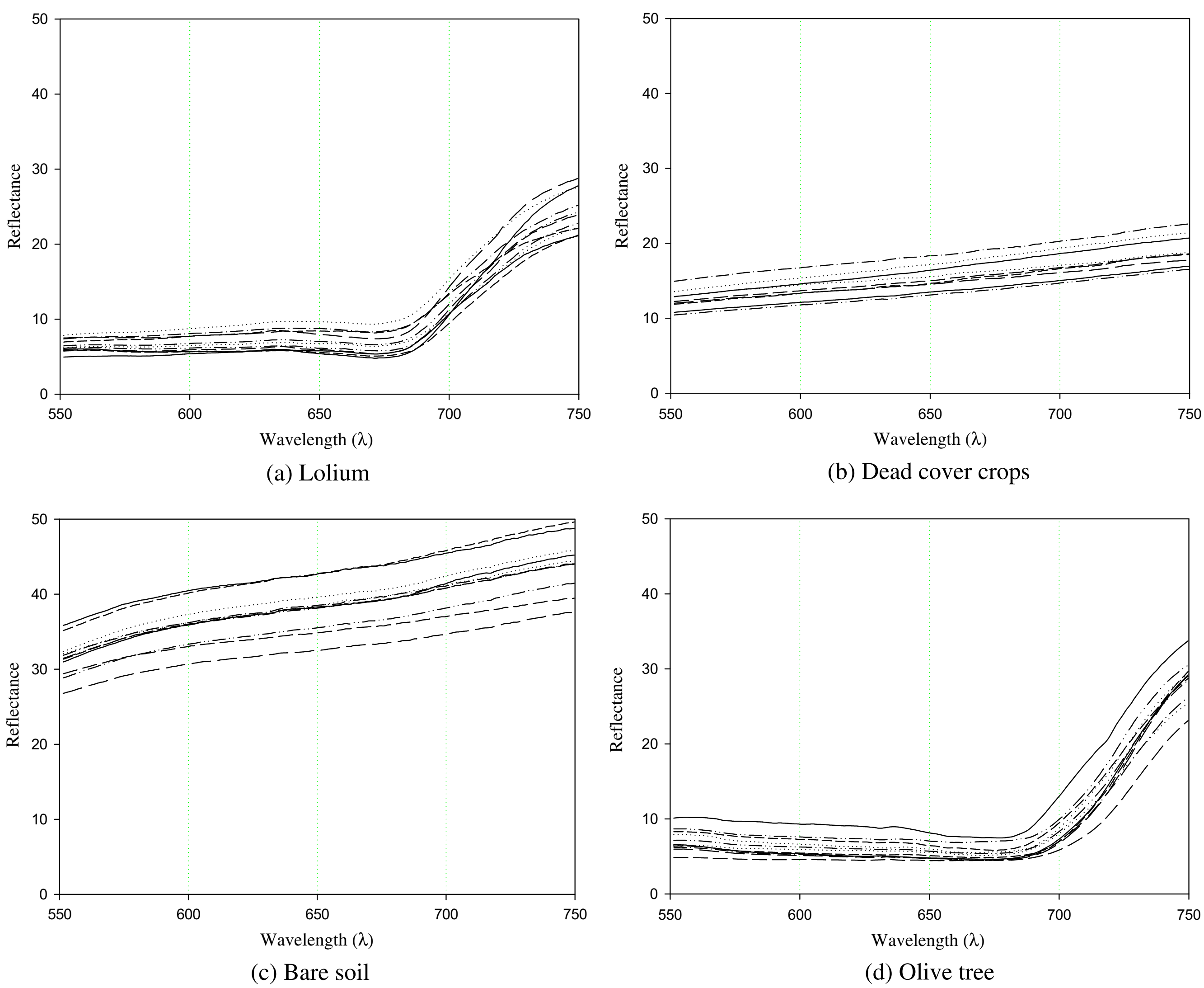

Fig. 4. Spectral reflectance (%) curves in "Cortijo del Rey" spring: (a) live cover crop (*Lolium rigidum*), (b) dead cover crop, (c) bare soil, and (d) olive tree.

3. Results and discussion

First of all, the different reflectance curves used for the experiments are going to be briefly analysed. Figs. 4–7 show these reflectance curves, which are consistently affected by the kind of land use and the cover crop phenological stage. There were apparent reflectance differences in several wavelengths for the different land uses, which showed that there was a potential for discrimination. The overall shape of the reflectance curves for live cover crops and olive trees in "Cortijo del Rey" and "Matallana" in spring were similar and exhibited the characteristic highest reflectance from 700 to 750 nm (near-infrared) of the green vegetation (Figs. 4a and d, and 6a and d). This demonstrates that they were in an active living stage. Desiccated and dead cover crops reflectance curves for "Cortijo del Rey" and "Matallana" summer steadily increased as wavelengths increased, indicating that both cover crops were no photosynthetically active vegetation and, consequently, their spectral signatures were very similar to bare soil (Figs. 4b and c, 5a–c, 6b and c, and 7a–c).

Once the potential of discrimination by using these wavelengths have been assessed, a further analysis of the results obtained with the different methodologies was performed. Table 2 shows the mean value and the standard deviation (Mean ± SD) obtained from 100 values of the Correct Classification Rate for the generalization set (CCR_G) of the best models in all the experiments performed. Our models, MLIRBF and SLIRBF, produced very satisfactory results in mean and standard deviation with regard to the CCR_G. The MLIRBF methodology is the most accurate, in mean and variance, for "Matallana" summer, and the second best, in mean, for "Matallana" spring. SLIRBF yields the best results, in mean, for "Cortijo del Rey" spring and summer, with very low standard deviations values.

In order to assess the significance of the differences observed between the methods, a set of statistical tests have been performed. In general, the nature of the CCR_G results does not imply the normality and the equality of variances hypothesis. This reason justifies the use of a non-parametric Friedman test (Friedman, 1940) with the ranking of the mean CCR_G values of the different methods as the test variable with the aim of determining the statistical significance of the differences observed for each method in the different datasets, "Cortijo del Rey" and "Matallana" in spring and summer. This test is based on obtaining a ranking $R_{(i)}$ for each method for all datasets, where $1 \leqslant i \leqslant 12$ (with the same id for each method than that used in Table 2), $R_{(i)} = 1$ for the best method and $R_{(i)} = 12$ for the worst method. Then, the mean ranking $\overline{R}_{(i)}$ is obtained as the mean value of these rankings for the i-th

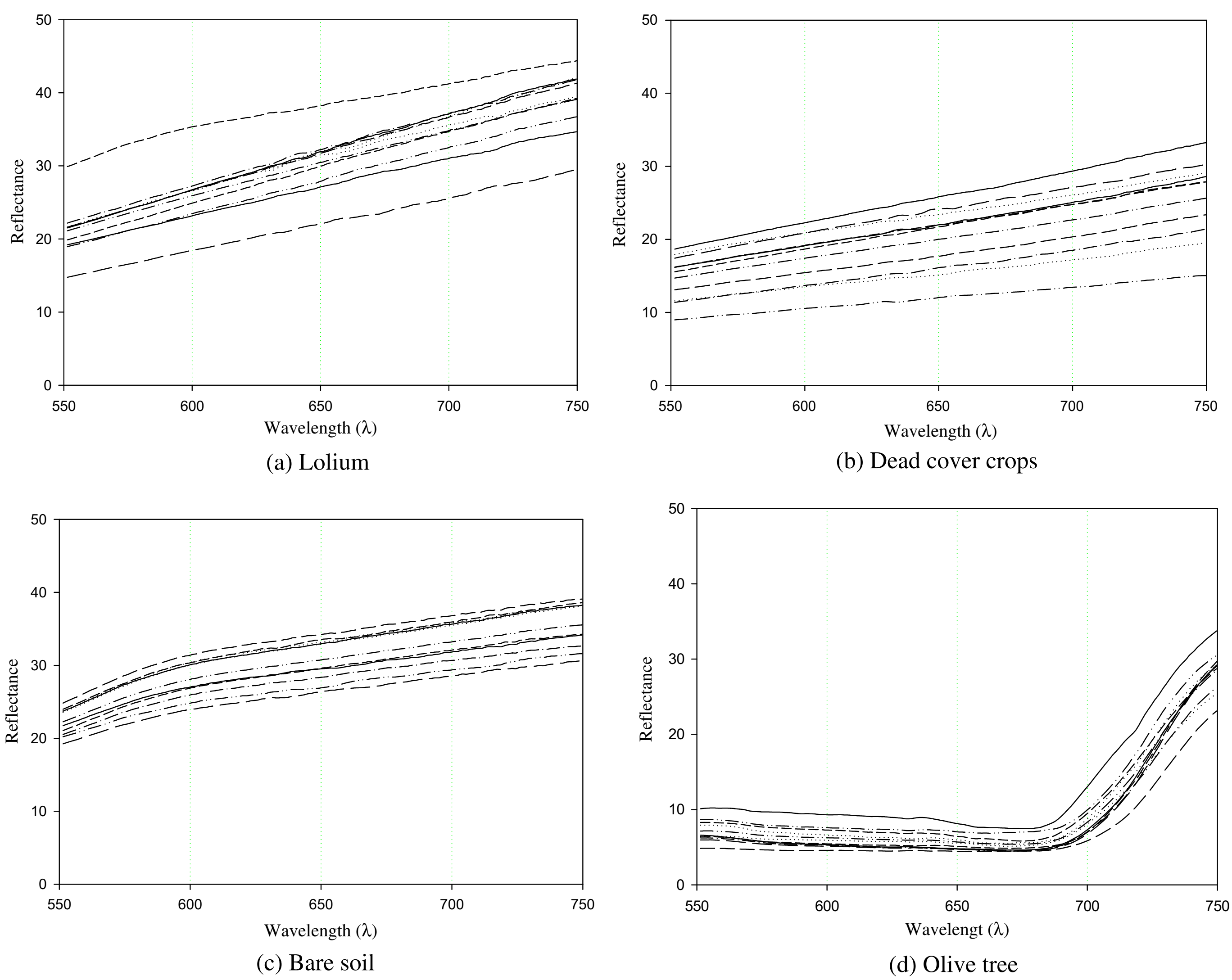

Fig. 5. Spectral reflectance (%) curves in "Cortijo del Rey" summer: (a) desiccated cover crop (*Lolium rigidum*), (b) dead cover crops, (c) bare soil, and (d) olive tree.

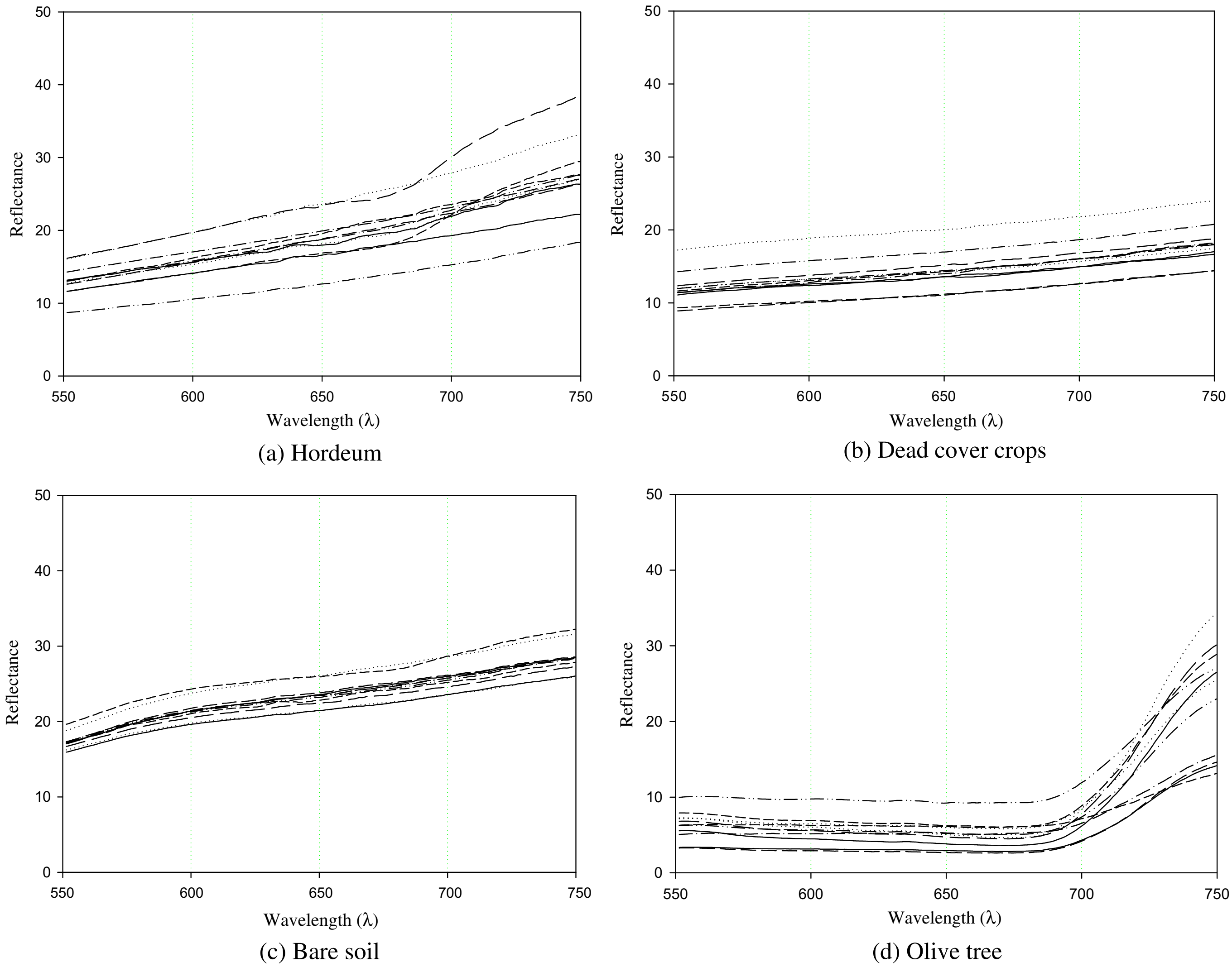

Fig. 6. Spectral reflectance (%) curves in "Matallana" spring: (a) live cover crop (*Hordeum murinum*), (b) dead cover crops, (c) bare soil, and (d) olive tree.

method (these mean ranking can be observed in the first column of Table 3). The null-hypothesis is $H_0 \equiv \overline{R}_{(1)} = \ldots = \overline{R}_{(12)}$ and, in our experiments, the test shows that the effect of the method used for classification is statistically significant at a significance level of 5%, as the confidence interval is $C_0 = (0, F_{0.05} = 2.09)$ and the F-distribution statistical value is $F^* = 15.09 \notin C_0$. Consequently, we reject the null-hypothesis stating that all algorithms perform equally in mean ranking.

On the basis of this rejection, two post-hoc non-parametric Bonferroni–Dunn tests (Hochberg & Tamhane, 1987) were applied with the best performing algorithms (MLIRBF and SLIRBF) as the control methods. The results of the Bonferroni–Dunn test for $\alpha = 0.1$ and $\alpha = 0.05$ can be seen in Table 3 using the corresponding critical values for the two-tailed test. The results show that SLIRBF obtains better significant mean ranking of CCR_G than C4.5 and NBTree for $\alpha = 0.05$. MLIRBF yields significant better results in mean CCR_G ranking than the same before two methods for $\alpha = 0.05$ and than ABoost(100) and ERBF for $\alpha = 0.1$. Therefore, our models MLIRBF and SLIRBF performed better or equally than the other classification methods herein presented (widely used in statistics and machine learning), whose percentage of correct classifications were already higher than 90% for most of the methods.

Table 4 presents the best functions obtained from the 100 executions performed in our experiments. Best discriminant functions were obtained by SLIRBF for "Cortijo del Rey" and by MLIRBF for "Matallana". In "Cortijo del Rey" spring, the three discriminant functions of the best classification model have only four covariates λ^*_{575}, λ^*_{675}, λ^*_{700} and λ^*_{725} and three radial basis functions localized in λ^*_{625}, λ^*_{650} and λ^*_{725}, indicating that the six wavelengths were used for a successful classification of live cover corps, bare soil and olive trees. In "Cortijo del Rey" summer, the best model has only two covariates, λ^*_{675} and λ^*_{725}, and two basis radial functions localized in λ^*_{575} and λ^*_{600}. One of the discriminant functions has not any RBF transformation, which matches with Fig. 5a–c, where the form of the spectral signatures for desiccated and dead cover crops, and bare soil, is quasi-linear. In "Matallana" spring, the discriminant functions are composed of all covariates and three RBF transformations defined for λ^*_{575}, λ^*_{650} and λ^*_{725}. Finally, the seven wavelengths are also present in "Matallana" summer. Moreover, four additional RBF transformation are defined around λ^*_{575}, λ^*_{600}, λ^*_{650} and λ^*_{675} wavelengths. If we compare the wavelengths used in this case to those used in "Cortijo del Rey" summer, two extra wavelengths are necessary, λ^*_{650} and λ^*_{675}, because, as it can be seen in Fig. 7d, there is a non-linearity associated to the class olive tree over these wavelengths. The accuracy or CCR of these models for the training and generalization sets (CCR_T and CCR_G) are also included in Table 4. The obtained models perform a perfect classification in the generalization set ($CCR_G = 100\%$) in both seasons and locations. For the

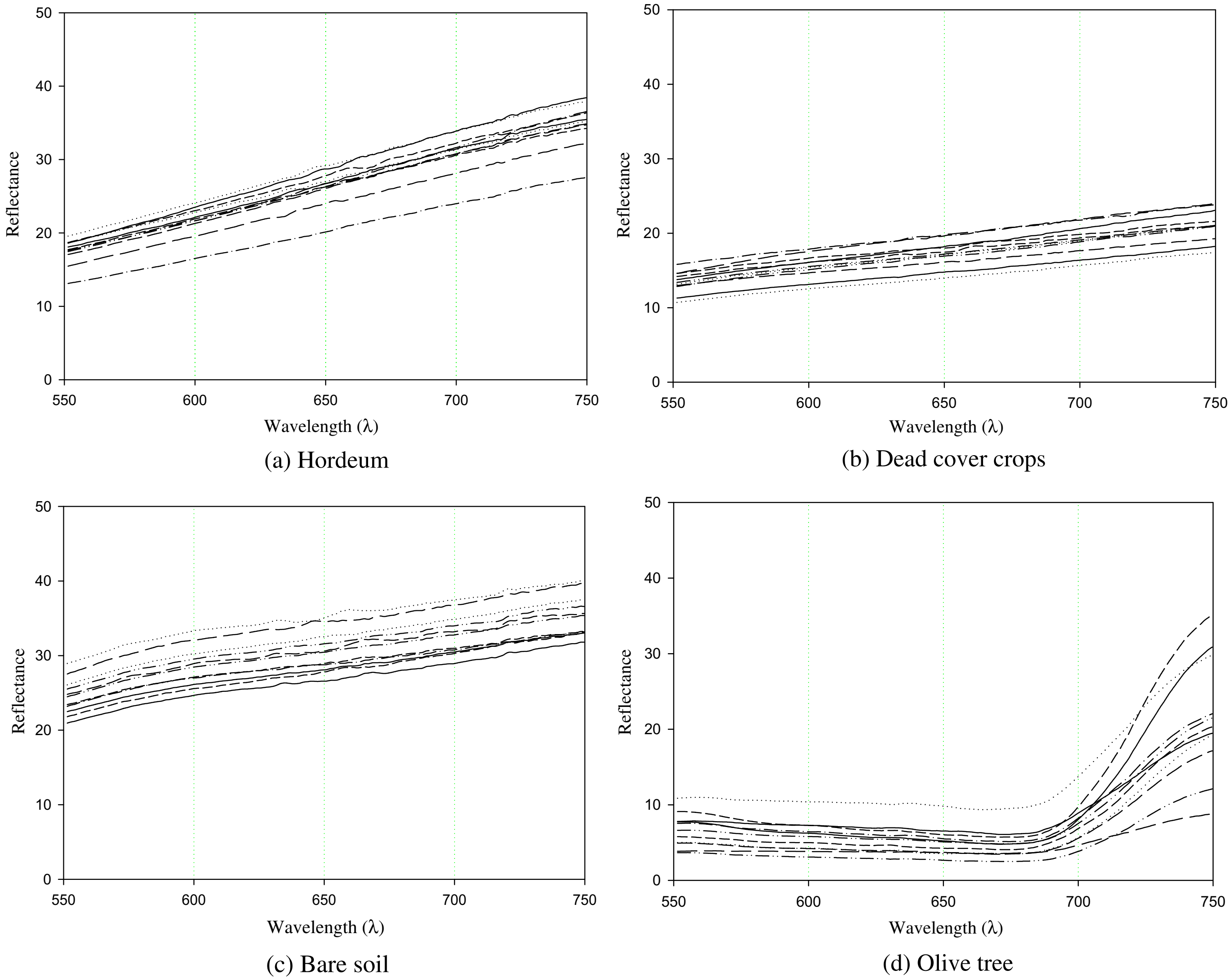

Fig. 7. Spectral reflectance (%) curves in "Matallana" summer: (a) desiccated cover crop (*Hordeum murinum*), (b) dead cover crops, (c) bare soil, and (d) olive tree.

Table 2
Statistical results (mean ± SD) of the CCR_G, for different competitive classifiers of bare soil, olive trees and sown, and dead cover crops in spring and summer in "Cortijo del Rey" and "Matallana".

Location and seasons	(1)	(2)	(3)	(4)	(5)	(6)
"Cortijo del Rey" spring	91.63 ± 9.74	*95.25 ± 7.70*	*95.25 ± 7.70*	87.50 ± 11.51	84.50 ± 11.53	91.00 ± 8.90
"Cortijo del Rey" summer	89.60 ± 14.06	91.00 ± 12.19	91.40 ± 11.81	77.40 ± 14.95	72.00 ± 12.39	80.00 ± 14.49
"Matallana" spring	**97.50 ± 5.98**	91.50 ± 11.24	91.33 ± 11.23	84.67 ± 14.92	86.50 ± 14.92	87.00 ± 13.52
"Matallana" summer	*98.50 ± 4.79*	95.83 ± 7.25	95.83 ± 7.25	91.00 ± 11.22	90.00 ± 11.36	91.00 ± 11.22
	(7)	(8)	(9)	(10)	(11)	(12)
"Cortijo del Rey" spring	91.25 ± 8.79	86.00 ± 5.11	88.63 ± 9.42	87.63 ± 8.79	92.13 ± 7.88	**96.13 ± 7.04**
"Cortijo del Rey" summer	79.60 ± 14.77	86.60 ± 15.06	86.20 ± 14.13	87.80 ± 13.30	*91.40 ± 10.73*	**92.80 ± 10.45**
"Matallana" spring	85.83 ± 13.69	86.17 ± 11.37	90.33 ± 10.64	88.33 ± 10.98	*97.00 ± 6.44*	92.00 ± 8.37
"Matallana" summer	91.00 ± 11.22	91.00 ± 9.60	94.33 ± 9.54	92.17 ± 9.01	**99.50 ± 2.86**	93.50 ± 9.14

(1) MLogistic; (2) SLogistic; (3) LMT; (4) C4.5; (5) NBTree; (6) ABoost(10); (7) ABoost(100); (8) ERBF; (9) MLRBF; (10) SLRBF; (11) MLIRBF; (12) SLIRBF. The best result for each location and season is in bold face, and the second best result in italics.

training set of "Matallana", the models result in a $CCR_T = 98.61\%$ in spring and a $CCR_T = 95.56\%$ in summer, while for "Cortijo del Rey" training set both seasons result in a perfect classification, $CCR_T = 100\%$.

Our study reveals that reducing the number of covariates to seven wavelengths allowed a better interpretability and a lower cost-efficiency ratio during the modelling process.

The results demonstrated that there were significant spectral differences between bare soil, the different cover crops and olive trees in spring and summer, and that our models successfully classified their spectral signatures at any of the seasons. Therefore, there would be different timeframes for future image acquisition. To have a wide timeframe is essential for a proper mapping of cover crops by using remote sensing; specially taking into account that

Table 3
Bonferroni–Dunn test for SLIRBF and MLIRBF control methodologies.

Mean ranking ($\overline{R}$)	Difference with $\overline{R}_{\text{SLIRBF}}$	Difference with $\overline{R}_{\text{MLIRBF}}$
$\overline{R}_{\text{MLogistic}} = 3.25$	$\lvert\overline{R}_{\text{MLogistic}} - \overline{R}_{\text{SLIRBF}}\rvert = 0.50$	$\lvert\overline{R}_{\text{MLogistic}} - \overline{R}_{\text{MLIRBF}}\rvert = 0.88$
$\overline{R}_{\text{SLogistic}} = 3.50$	$\lvert\overline{R}_{\text{SLogistic}} - \overline{R}_{\text{SLIRBF}}\rvert = 0.75$	$\lvert\overline{R}_{\text{SLogistic}} - \overline{R}_{\text{MLIRBF}}\rvert = 1.13$
$\overline{R}_{\text{LMT}} = 3.38$	$\lvert\overline{R}_{\text{LMT}} - \overline{R}_{\text{SLIRBF}}\rvert = 0.63$	$\lvert\overline{R}_{\text{LMT}} - \overline{R}_{\text{MLIRBF}}\rvert = 1.00$
$\overline{R}_{\text{C4.5}} = 10.63$	$\lvert\overline{R}_{\text{C4.5}} - \overline{R}_{\text{SLIRBF}}\rvert = 7.88(*)$	$\lvert\overline{R}_{\text{C4.5}} - \overline{R}_{\text{MLIRBF}}\rvert = 8.25(*)$
$\overline{R}_{\text{NBTree}} = 11.25$	$\lvert\overline{R}_{\text{NBTree}} - \overline{R}_{\text{SLIRBF}}\rvert = 8.50(*)$	$\lvert\overline{R}_{\text{NBTree}} - \overline{R}_{\text{MLIRBF}}\rvert = 8.88(*)$
$\overline{R}_{\text{ABoost10}} = 8.38$	$\lvert\overline{R}_{\text{ABoost10}} - \overline{R}_{\text{SLIRBF}}\rvert = 5.63$	$\lvert\overline{R}_{\text{ABoost10}} - \overline{R}_{\text{MLIRBF}}\rvert = 6.00$
$\overline{R}_{\text{ABoost100}} = 9.13$	$\lvert\overline{R}_{\text{ABoost100}} - \overline{R}_{\text{SLIRBF}}\rvert = 6.38$	$\lvert\overline{R}_{\text{ABoost100}} - \overline{R}_{\text{MLIRBF}}\rvert = 6.75(**)$
$\overline{R}_{\text{ERBF}} = 9.38$	$\lvert\overline{R}_{\text{ERBF}} - \overline{R}_{\text{SLIRBF}}\rvert = 6.63$	$\lvert\overline{R}_{\text{ERBF}} - \overline{R}_{\text{MLIRBF}}\rvert = 7.00(**)$
$\overline{R}_{\text{MLRBF}} = 6.75$	$\lvert\overline{R}_{\text{MLRBF}} - \overline{R}_{\text{SLIRBF}}\rvert = 4.00$	$\lvert\overline{R}_{\text{MLRBF}} - \overline{R}_{\text{MLIRBF}}\rvert = 4.38$
$\overline{R}_{\text{SLRBF}} = 7.25$	$\lvert\overline{R}_{\text{SLRBF}} - \overline{R}_{\text{SLIRBF}}\rvert = 4.50$	$\lvert\overline{R}_{\text{SLRBF}} - \overline{R}_{\text{MLIRBF}}\rvert = 4.88$
$\overline{R}_{\text{MLIRBF}} = 2.38$	$\lvert\overline{R}_{\text{SLIRBF}} - \overline{R}_{\text{MLIRBF}}\rvert = 0.38$	-
$\overline{R}_{\text{SLIRBF}} = 2.75$	-	$\lvert\overline{R}_{\text{SLIRBF}} - \overline{R}_{\text{MLIRBF}}\rvert = 0.38$

Critical difference for $\alpha = 0.1$, $CD_{0.1} = 6.65$; and for $\alpha = 0.05$, $CD_{0.05} = 7.24$.
($*$ and $**$): Statistically significant differences for $\alpha = 0.05$ ($*$) and $\alpha = 0.1$ ($**$).

Table 4
Discrimination equations provided by the best SLIRBF and MLIRBF models in "Cortijo del Rey" and "Matallana" for classification of spectral signatures of bare soil, olive trees and sown (live in spring, desiccated in summer), and dead cover crops. The accuracy in the training (CCR_T) and generalization (CCR_G) sets are also included, together with the number of RBFs and the number of coefficients of the models.

"Cortijo del Rey" spring. SLIRBF model

$$F_1 = -18.83 - 17.64(\lambda^*_{575}) - 1.80(\lambda^*_{675}) + 9.96(\lambda^*_{700}) + 19.43(\text{RBF}_1)$$
$$F_2 = -16.53 - 10.58(\lambda^*_{575}) + 4.10(\lambda^*_{675}) + 9.02(\lambda^*_{700}) - 4.77(\lambda^*_{725}) + 4.71(\text{RBF}_1) + 21.68(\text{RBF}_3)$$
$$F_3 = 7.44 - 4.83(\lambda^*_{575}) + 9.02(\lambda^*_{700}) + 4.71(\text{RBF}_1) - 30.17(\text{RBF}_2)$$

$$\text{RBF}_1 = \exp\left(-0.5\left(\left((\lambda^*_{725} - 1.18)^2\right)0.5/(1.28)^2\right)\right)$$
$$\text{RBF}_2 = \exp\left(-0.5^*\left(\left((\lambda^*_{625} + 0.67)^2\right)^{0.5}/(0.93)^2\right)\right)$$
$$\text{RBF}_3 = \exp\left(-0.5 * \left((\lambda^*_{625} + 0.39)^2 + (\lambda^*_{650} + 0.21)^2\right)^{0.5}/(1.13)^2\right)$$

Number of RBFs = 3, number of coefficients=21, CCR_T = 98.61%, CCR_G = 100%

"Cortijo del Rey" summer. SLIRBF model

$$F_1 = 0.68 + 1.53(\lambda^*_{675}) + 1.28(\lambda^*_{725})$$
$$F_2 = -1.19 + 1.53(\lambda^*_{675}) + 5.57(\text{RBF}_1)$$
$$F_3 = -1.44 + 1.53(\lambda^*_{675}) + 4.30(\text{RBF}_2)$$

$$\text{RBF}_1 = \exp\left(-0.5\left((\lambda^*_{600} + 0.72)^2\right)0.5/(0.66)^2\right)$$
$$\text{RBF}_2 = \exp\left(-0.5\left((\lambda^*_{575} - 1.23)^2\right)^{0.5}/(0.87)^2\right)$$

Number of RBFs = 2, number of coefficients = 11, CCR_T = 95.56%, CCR_G = 100%

"Matallana" spring. MLIRBF model

$$F_1 = -66.76 - 49.91(\lambda^*_{575}) - 28.50(\lambda^*_{600}) - 5.45(\lambda^*_{625}) + 8.46(\lambda^*_{650}) + 21.41(\lambda^*_{675}) + 43.34(\lambda^*_{700}) + 13.89(\lambda^*_{725}) + 80.71(\text{RBF}_1) + 32.17(\text{RBF}_2) + 78.14(\text{RBF}_3)$$
$$F_2 = -46.36 + 24.86(\lambda^*_{575}) + 9.31(\lambda^*_{600}) + 2.19(\lambda^*_{625}) - 1.70(\lambda^*_{650}) - 9.35(\lambda^*_{675}) - 14.45(\lambda^*_{700}) - 27.96(\lambda^*_{725}) + 78.58(\text{RBF}_1) - 22.88(\text{RBF}_2) + 92.22(\text{RBF}_3)$$
$$F_3 = +37.93 + 15.64(\lambda^*_{575}) + 16.01(\lambda^*_{600}) + 8.67(\lambda^*_{625}) + 1.46(\lambda^*_{650}) - 2.92(\lambda^*_{675}) - 2.99(\lambda^*_{700}) - 1.94^*(\lambda^*_{725}) - 34.58(\text{RBF}_1) - 3.676(\text{RBF}_2) - 146.02(\text{RBF}_3)$$

$$\text{RBF}_1 = \exp\left(-0.5\left((\lambda^*_{650} - 0.36)^2\right)^{0.5}/(0.86)^2\right)$$
$$\text{RBF}_2 = \exp\left(-0.5\left((\lambda^*_{650} + 1.42)^2\right)^{0.5}/(1.04)^2\right)$$
$$\text{RBF}_3 = \exp\left(-0.5\left((\lambda^*_{575} + 0.17)^2 + (\lambda^*_{725} - 1.77)^2\right)^{0.5}/(1.00)^2\right)$$

Number of RBFs = 3 , number of coefficients = 37, CCR_T = 100%, CCR_G = 100%

"Matallana" summer. MLIRBF model

$$F_1 = 7.08 - 46.26(\lambda^*_{575}) - 21.87(\lambda^*_{600}) + 1.33(\lambda^*_{625}) + 18.80(\lambda^*_{650}) + 28.50(\lambda^*_{675}) + 24.34(\lambda^*_{700}) - 6.82(\lambda^*_{725}) - 0.34(\text{RBF}_1) - 51.50(\text{RBF}_2) - 64.85(\text{RBF}_3)$$
$$F_2 = -15.83 + 37.59(\lambda^*_{575}) + 25.85(\lambda^*_{600}) + 14.16(\lambda^*_{625}) + 6.80(\lambda^*_{650}) + 3.16(\lambda^*_{675}) - 15.27(\lambda^*_{700}) - 56.01(\lambda^*_{725}) + 58.83(\text{RBF}_1) + 8.31(\text{RBF}_2) + 14.59(\text{RBF}_3)$$
$$F_3 = +7.30 + 14.08(\lambda^*_{575}) + 9.39(\lambda^*_{600}) + 3.25(\lambda^*_{625}) - 1.32(\lambda^*_{650}) - 3.42(\lambda^*_{675}) - 6.81(\lambda^*_{700}) - 12.76(\lambda^*_{725}) + 19.51(\text{RBF}_1) + 36.91(\text{RBF}_2) - 24.58(\text{RBF}k_3)$$

$$\text{RBF}_1 = \exp(-0.5((\lambda^*_{575} + 0.10)^2 + (\lambda^*_{600} + 0.32)^2)^{0.5}/(1.07)^2)$$
$$\text{RBF}_2 = \exp(-0.5((\lambda^*_{575} - 1.54)^2 + (\lambda^*_{600} - 1.42)^2)^{0.5}/(0.74)^2)$$
$$\text{RBF}_3 = \exp(-0.5((\lambda^*_{650} + 1.10)^2 + (\lambda^*_{675} + 1.76)^2)^{0.5}/(1.32)^2)$$

Number of RBFs = 3, number of coefficients = 39, CCR_T = 100%, CCR_G = 100%

$F_{\text{i}} = \log\ \text{odd pi},\ \lambda^*_i \in (-2, 2)$

it is very frequent to have cloudy days in March and April (spring in our Mediterranean conditions) and no remote images can be taken in these circumstances. If image acquisition fails on spring, we could programme the remote images acquisition around June when there are plenty of sunny days. In addition, this double possibility could be used for re-monitoring a doubtful and specific field analysed in previous spring. This is essential for programming and implementing the control tools and for avoiding the annoying and frequent bottle neck of administrative follow-up to concede or not the subsidy. Once MLIRBF and SLIRBF have been shown as promising to successfully classify spectral signatures of bare soil, olive trees, and live, desiccated and dead cover crops, next investigations should explore their potential discrimination and mapping by image analysis of CASI imagery taken in spring and summer and programmed with seven wavelengths in the green (575 nm), red (600, 625, 650, and 675 nm), and near-infrared (700 and 725 nm) spectral range, rather than using the 288 available wavelengths. Granted that, in our case, computational requirements for training LIRBF models were nearly insignificant once the ERBF models are built, three considerations should be made: the improvement in correct classification obtained with our models, MLIRBF and SLIRBF, the computational or expertise requirements involved in the modelling process and the objective that we wish to achieve. From an agronomic point of view, if we would aim to create a map for crop inventory with detailed vegetation classifications of olive orchards, then a simpler, easier model, for example MLogistic or SLogistic, would be the best choice as its accuracy of classification was high enough and the higher computational requirements for MLIRBF and SLIRBF would not be justified; however, if we need to produce a very accurate thematic map ready to be used for decision-making procedures by administrations, the criteria for selecting MLIRBF and SLIRBF should not be based on decreasing computational requirements and complexity, but on the accuracy of the classifications and these more sophisticated and accurate models would be highly recommended.

Finally, a current trend is that farmers use dead cover crops rather than sown ones due to dead covers come from the recycling of olive tree residues. This has several economical and environmental advantages due to the non-use of fertilizers and herbicides. Dead cover crops only have to be cut through light farm equipment. However, seed for sown cover crops must be brought, sown, and plants fertilized and treated with herbicide in a specific time, which also implies the use of machinery and fuel. Fertilizers, herbicide and fuel are the most important crop inputs in any agrarian system. Thus, it is worth noting that our models clearly classified dead cover crops from olive trees and bare soil, this being very important for future remote sensing investigations.

4. Conclusions

This study demonstrated the capability of LR and RBFNN combination models, where the final coefficients were estimated using MultiLogistic regression. These models (MLIRBF and SLIRBF) were applied for the discrimination of cover crops in olive orchards as affected by their phenological stage using the spectral signatures obtained with a high-resolution field spectroradiometer. The objective was to differentiate bare soil, olive trees and cover crops (live or dead).

SLIRBF and MLIRBF models provided better accuracy models in the generalization sets than linear LR. Mean generalization accuracies of 96.13% and 92.80% were obtained using the SLIRBF methodology for the two seasonal stages of ‘‘Cortijo del Rey’’. Moreover, the best model in this location in spring resulted in an accuracy of 98.61% in the training set and 100% in generalization and 95.56% in the training set and 100% in generalization set in summer. Furthermore, a 97.00% and 99.50% of generalization mean accuracies was obtained using MLIRBF methodology for the two seasonal stages of ‘‘Matallana’’ and a 100% in training and generalization sets for the best models obtained in the two phenological stages. Last, seven advanced methodologies (MLogistic, SLogistic, LMT, C4.5, NBTree, AdaBoost 10 and 100) were compared to the methodologies herein presented, resulting in lower accuracy except for ‘‘Matallana’’ spring where MLogistic obtained a slightly higher result. From the statistical test results we can conclude that the best methodology is MLIRBF because it presents statistically significant differences for $\alpha = 0.05$ or $\alpha = 0.1$, for four out of the 11 compared methodologies.

To summarize, our models successfully discriminated bare soil, olive trees and all of the possible kind of cover crops used by farmers in our conditions in spring and summer. However, more research is needed to study if high spatial and spectral resolution airborne imagery would correctly classify and map any of the land uses proposed in this paper.

Acknowledgements

This work has been partially financed by the TIN 2008-06681-C06-03 project (Spanish Inter-Ministerial Commission of Science and Technology), FEDER funds, and the P08-TIC-3745 project (Junta de Andalucia, Spain).

References

Andalusian Administration Regulation (2007). *Andalusian official bulletin (BOJA), 119*, 7–37. Available from http://www.juntadeandalucia.es/boja/boletines/2007/119/d/2.html.

Angeline, P. J., Saunders, G. M., & Pollack, J. B. (1994). An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks, 5*, 54–65.

Foody, G. M., & Arora, M. K. (1996). Incorporating mixed pixel in the training, allocation and testing stages of supervised classification. *Pattern Recognition Letters, 17*, 1389–1398.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In Morgan Kaufmann (Ed.), *Proceedings of the thirteenth international conference on machine learning* (pp. 148–156).

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics, 11*, 86–92.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics, 38*, 337–374.

Fukunaga, K. (1999). *Introduction to statistical pattern recognition* (2nd ed.). Academic Press.

Gill, P. E., Murray, W., & Wright, M. H. (1982). *Practical optimization*. Academic Press.

Goel, P. K., Prasher, S. O., Patel, R. M., Landry, J. A., Bonnel, R. B., & Viau, A. A. (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture, 39*, 67–93.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, Massachusetts, USA: Addison-Wesley.

González-Andujar, J. L. (2009). Expert system for pests, diseases and weeds identification in olive crops. *Expert Systems with Applications, 36*, 3278–3283.

González-Diaz, L., Martínez-Jimenez, P., Bastida, F., & González-Andujar, J. L. (2009). Expert system for integrated plant protection in pepper (*Capsicum annuun* L). *Expert Systems with Applications, 36*, 8975–8979.

Gutiérrez, P. A., Fernández, J. C., Hervás-Martínez, C., López-Granados, F., Jurado-Expósito, M., & Peña-Barragán, J. M. (2009). Structural simplification for hybrid neuro-logistic regression in multispectral analyse of remote sensed data. *Neural Network World, 19*, 3–20.

Gutiérrez, P. A., López-Granados, F., Peña-Barragán, J. M., Jurado-Expósito, M., Gómez-Casero, M. T., & Hervás-Martínez, C. (2008b). Mapping sunflower yield as affected by *Ridolfia segetum* patches and elevation by applying Evolutionary Product Unit Neural Networks to remote sensed data. *Computers and Electronics in Agriculture, 60*, 122–132.

Gutiérrez, P. A., López-Granados, F., Peña-Barragán, J. M., Jurado-Expósito, M., & Hervás-Martínez, C. (2008a). Logistic Regression Product Unit Neural Networks for mapping *Ridolfia segetum* infestations in sunflower crop using multitemporal remote sensed data. *Computers and Electronics in Agriculture, 64*, 293–306.

Harvey, K. R., & Hill, G. J. E. (2001). Vegetation mapping of a tropical freshwater swamp in the Northern Territory, Australia: A comparison of aerial photography, Landsat TM and SPOT satellite imagery . *International Journal of Remote Sensing, 22*, 2911–2925.

Hervás-Martínez, C., & Martínez-Estudillo, F. J. (2007). Logistic regression using covariates obtained by product-unit neural network models. *Pattern Recognition, 40*, 52–64.
Hervás-Martínez, C., Martínez-Estudillo, F. J., & Carbonero-Ruz, M. (2008). Multilogistic regression by means of evolutionary product-unit neural networks. *Neural Networks, 21*, 951–961.
Hill, J., Megier, J., & Mehl, W. (1995). Land degradation, soil erosion, and desertification monitoring in Mediterranean ecosystems. *International Journal of Remote Sensing, 12*, 107–130.
Hochberg, Y., & Tamhane, M. (1987). *Multiple comparison procedures*. John Wiley & Sons.
Ibrahim, M. A., Arora, M. K., & Ghosh, S. K. (2005). Estimating and accommodating uncertainty through the soft classification of remote sensing data. *International Journal of Remote Sensing, 26*, 2995–3007.
Karimi, Y., Prasher, S. O., Mcnaim, H., Bonnell, R. B., Dutilleul, P., & Goel, P. K. (2005). Classification accuracy of discriminant analysis, neural networks and decision trees for weed and nitrogen stress detection in corn. *Transactions of the American Society of Agricultural Engineerings, 48*, 1261–1268.
Kin, M. (2009). Two-stage logistic regression model. *Expert Systems with Applications, 36*, 6727–6734.
Koger, C. H., Shaw, D. R., Reddy, K. N., & Bruce, L. M. (2004). Detection of pitted morningglory (*Ipomoea lacunosa*) with hyperspectral remote sensing. II. Effects of vegetation ground cover and reflectance properties. *Weed Science, 52*, 230–235.
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 1995 international joint conference on artificial intelligence (IJCAI'95)* (pp. 1137–1143). Montreal, Quebec, Canadá.
Kohavi, R. (1996). Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *Proceedings of the second international conference on knoledge discovery and data mining (KDD96)* (pp. 202–207). Menlo Park, CA: AAAI Press.
Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning, 59*, 161–205.
Le Cessie, S., & Van Houwelingen, J. (1992). Ridge estimators in logistic regression. *Applied Statistics, 41*, 191–201.
Lee, C.-C., Chiang, Y.-C., Shih, C.-Y., & Tsai, C.-L. (2009). Noisy time series prediction using M-estimator based robust radial basis function neural networks with growing and pruning techniques. *Expert Systems with Applications, 36*, 4717–4724.
Lee, S., & Kil, R. M. (1991). A Gaussian potential function network with hierarchically selforganizing learning. *Neural Networks, 4*, 207–224.
López-Granados, F., Jurado-Expósito, M., Peña-Barragán, J. M., & García-Torres, L. (2006). Using remote sensing for identification of late season grass weed patches in wheat. *Weed Science, 54*, 346–353.
López-Granados, F., Peña-Barragán, J. M., Jurado-Expósito, M., Francisco-Fernández, M., Cao, R., Alonso-Betanzos, A., et al. (2008). Multispectral classification of grass weeds and wheat (*Triticum durum*) crop using linear and nonparametric functional discriminant analysis and neural networks. *Weed Research, 48*, 28–37.
MAPYA (2007). Spanish Ministry of Agriculture, Fishery and Food. Available from http://www.mapa.es/es/estadistica/pags/anuario/introduccion.htm.
Martínez-Estudillo, A. C., Hervás-Martínez, C., Martínez-Estudillo, F. J., & García-Pedrajas, N. (2006a). Hybridization of evolutionary algorithms and local search by means of a clustering method. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, 36*, 534–545.
Martínez-Estudillo, A. C., Martínez-Estudillo, F. J., Hervás-Martínez, C., & García-Pedrajas, N. (2006b). Evolutionary product unit based neural networks for regression. *Neural Networks, 19*, 477–486.
Moody, J., & Darken, C. J. (1989). Fast learning in network of locally-tuned processing units. *Neural Computation, 1*, 281–294.
Neupane, R. P., Sharma, K. R., & Thapa, G. B. (2002). Adoption of agroforestry in the hills of Nepal: A logistic regression analysis. *Agricultural Systems, 72*, 177–196.
Park, J., & Sandberg, I. W. (1991). Universal approximation using radial basis functions network. *Neural Computation, 3*, 246–257.
Peña-Barragán, J. M., López-Granados, F., Jurado-Expósito, M., & García-Torres, L. (2006). Spectral discrimination of *Ridolfia segetum* and sunflower as affected by phenological stage. *Weed Research, 46*, 10–21.
Pu, R., & Gong, P. (2004). Determination of burnt scars using logistic regresión and neural network techniques from a single post-fire Landsat-7 ETM + image. *Photogrammetric Engineering and Remote Sensing, 70*, 841–850.
Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
Salisbury, J.W. (1999). Spectral measurements field guide. In *Report No. ADA362372*, 90 (pp. 2–9). Defense Tecnol. Info. Ctr., Fort Bervoi, USA.
Schlesinger, W. H. (2000). Carbon sequestration in soils: Some cautions amidst optimism. *Agriculture, Ecosystems & Environment, 82*, 121–127.
Schmidt, K. S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in coastal wetland. *Remote Sensing of Environment, 85*, 92–108.
South, S., Qi, J., & Lusch, D. P. (2004). Optimal classification methods for mapping agricultural tillage practises. *Remote Sensing of Environment, 91*, 90–97.
Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral wavebands performance of vegetation analysis applications. *Remote Sensing of Environment, 91*, 354–376.
Thenkabail, P. S., Smith, R. B., & De-Pauw, E. (2000). Hyperspectral vegetation indices for determining agricultural crop characteristics. *Remote Sensing of Environment, 71*, 158–182.
Torres, M., Hervás, C., & García, C. (2009). Multinomial logistic regression and product unit neural network models: Application of a new hybrid methodology for solving a classification problem in the livestock sector. *Expert Systems with Applications, 36*, 12225–12235.
Uno, Y., Prasher, S. O., Lacroix, R., Goel, P. K., Karimi, Y., Viau, A., et al. (2005). Artificial Neural Networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Computers and Electronics in Agriculture, 47*, 149–161.
Van Deventer, A. P., Ward, A. D., Gowda, P. H., & Lyon, J. G. (1997). Using Thematic Mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing, 63*, 87–93.
Witten, I.H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. In Morgan Kaufmann (Ed.), *Data management systems* (2nd ed., ser.). Elsevier.
Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment, 97*, 322–336.
Yao, X., & Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks, 8*, 694–713.