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

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.

© 2010 Elsevier Ltd. All rights reserved.

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

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