

Extreme learning machines for soybean classification in remote sensing hyperspectral images

Ramón Moreno^a, Francesco Corona^b, Amaury Lendasse^b,
Manuel Graña^{a,*}, Lênio S. Galvão^c

^a Computational Intelligence Group, Universidad del País vasco, San Sebastián, Spain

^b Aalto University, School of Science and Technology, Aalto, Finland

^c Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil

ARTICLE INFO

Article history:

Received 7 September 2012

Received in revised form

14 February 2013

Accepted 4 March 2013

Available online 7 November 2013

Keywords:

Extreme learning machine

Hyperspectral images

Agricultural remote sensing

ABSTRACT

This paper focuses on the application of Extreme Learning Machines (ELM) to the classification of remote sensing hyperspectral data. The specific aim of the work is to obtain accurate thematic maps of soybean crops, which have proven to be difficult to identify by automated procedures. The classification process carried out is as follows: First, spectral data is transformed into a hyper-spherical representation. Second, a robust image gradient is computed over the hyper-spherical representation allowing an image segmentation that identifies major crop plots. Third, feature selection is achieved by a greedy wrapper approach. Finally, a classifier is trained and tested on the selected image pixel features. The classifiers used for feature selection and final classification are Single Layer Feedforward Networks (SLFN) trained with either the ELM or the incremental OP-ELM. Original image pixel features are computed following a Functional Data Analysis (FDA) characterization of the spectral data. Conventional ELM training of the SLFN improves over the classification performance of state of the art algorithms reported in the literature dealing with the data treated in this paper. Moreover, SLFN-ELM uses less features than the referred algorithms. OP-ELM is able to find competitive results using the FDA features from a single spectral band.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

This paper is concerned with specific problem of soybean variety classification [11] from hyperspectral data. Precise thematic map building by pixel classification is relevant for the development of precision agriculture where crops are monitored by remote sensing. We report classification experiments on real life hyperspectral data obtained from the Tanguru farm in Brazil [11]. This hyperspectral image is a strong benchmark because it covers many crop plots (about 70) of five different soybean classes, which have very similar spectral signatures so that their accurate classification is difficult. Hyperspectral imagery is an emerging technology with many applications for remote sensing on earth observation [1–4]. The rich spectral information provided by hyperspectral images is useful for ecological applications such as detection of early-stages of vegetation stress, identifying small differences in vegetation cover abundances, discriminating land cover types, measuring leaf pigment concentrations, modeling quantitative biophysical and yield characteristics of agricultural crops, such as the estimation of citrus fruit yield [5]. Space borne

remote sensing field allows us to observe forest growing and shrinking, the evolution of burned regions, and to monitor agricultural crops, i.e. soybean [4,6–8], canopy chlorophyll and nitrogen content studies [9]. Machine learning has been used in connection to these applications, for instance, canopy species identification by pixel-wise classification in humid tropical forest [10].

In this paper, classification is performed by Single hidden Layer Feedforward Networks (SLFN) trained with Extreme Learning Machines (ELM) [14] and Optimally Pruned ELM (OP-ELM) [15,16] because their quick learning processes allow us to achieve the wrapper feature selection processes in an affordable time frame. The ELM approach to training SLFN consists in the random generation of the hidden layer weights, followed by solving a linear system of equations by least squares for the estimation of the output layer weights. This learning strategy is very fast, gives good classification accuracy, and it has been proven to provide universal approximation [17]. The OP-ELM algorithm performs the selection of the hidden layer units based on their ranking by a greedy sparse regression exploration of their effect on the SLFN output followed by a Leave One Out (LOO) validation, which is computed efficiently by a PREdiction Sum of Squares (PRESS) statistical approximation.

The process of the hyperspectral images involves the following steps: First, the image is spatially segmented into crop regions by

* Corresponding author. Tel.: +34 94 3018000.

E-mail address: manuel.grana@ehu.es (M. Graña).

applying a robust gradient operator, which allows us to find the boundaries of the regions. Second, we train a classifier using a selected sample of labeled pixels. This classifier is applied to the whole image. Finally, the identified crop regions serve for the spatial regularization of the classification results. It is assumed that the contents of the crop regions are single class soybean pixels, therefore all pixels are assigned to the winning class in the region. Validation is performed using a hold-out approach, where we use the same sample sizes as in [11] to be able to perform direct comparison of the achieved results. Additionally, we have performed one experiment of 10-fold stratified cross-validation to assess the impact of using stratified sampling. Two additional tasks are the definition of the pixel features used for classification, and the search for the minimal representation, that is the feature selection for each pixel.

A Functional Data Analysis (FDA) [12,13] approach for feature extraction is based on the idea of characterizing high dimensional data as functions, and using functional information as features for the classification. That is, the data samples may be interpolated by parametric functions and the parametric representation used as features. Following this approach, in this paper we compute the first and second derivatives of the spectral signature, so that we have at each band the second order Taylor expansion of the spectral signature. They are used as features for classification together with the original spectral signature of the pixel.

Dimensionality reduction aims to find the minimal size data representation with better classification accuracy results. In this paper, we achieve dimensionality reduction by a greedy wrapper feature selection process: for each hyperspectral image band we perform a separate hold-out validation experiment to assess its value, measured by the test classification accuracy achieved by a classifier specifically trained and tested on its information. Once we have the hyperspectral bands ranked, we perform an incremental greedy search for the optimal subset of bands. The bands are added in order of their separate classification accuracies, and a classifier is trained and tested on them. The process stops when the accuracy of the aggregate sets of bands decreases. Finally, we test the accuracy of the OP-ELM on the best single band data compared with other state of the art classifiers.

The outline of this paper is as follows: Section 2 gives a brief introduction to FDA. Section 3 describes the hyperspectral data used for the computational experiments. Section 4 gives a short review of ELM and OP-ELM. Section 5 describes the methodology and image preprocessing followed in the computational experiments. Section 6 gives the experimental results. Finally Section 7 gives the conclusions of this work.

2. Functional data analysis

Our interest in Functional Data Analysis (FDA) lies in its ability to provide compact parametric representations with increased discriminative power. Spectral signatures, such as those shown in Fig. 3, can be modeled by a parametric function. Sometimes the value of the function is less relevant for a precise application than other functional informations, such as the function derivatives. FDA [12,13] is an extension of traditional data analysis to functional data; that is, observations that can be thought as real-valued curves over some domain rather than vectors in a high-dimensional space. In practice, a single functional data sample is a curve corresponding to an observation constituted of J pairs $\{(x_j, y(x_j))\}_{j=1}^J$ of a function $f : \Omega \rightarrow \mathbb{R}$ defined over an open interval $\Omega \subset \mathbb{R}$ such that $\Omega = \{x \in \mathbb{R} : a < x < b\}$. Each sample curve follows the model $y(x_j) = f(x)|_{x_j} + r(x_j)$, with $j = 1, \dots, J$, where $y(x_j)$ is the observed value of the function $f(x)$ at argument x_j and $r(x_j)$ is the additive measurement error. The function $f(x)$ is also assumed to be smooth (or, regular), so

that values $f(x)|_{x_j}$ at contiguous arguments are necessarily linked together and unlikely to be too different from each other. In that sense, the function is implicitly assumed to be continuous and to possess a finite number of derivatives.

Often, $i = 1, \dots, I$ functional observations $\{(x_j, y_i(x_j))\}_{j=1}^J$ are provided over the same domain, so that unsupervised exploration of underlying curve variations becomes feasible. For this task, there exist functional extensions of descriptive statistics [18], principal component analysis [19], canonical correlation [20,21] and discriminant analysis [22] and methods for cluster analysis [23], all discussed in detail in [24,12]. In general, the observed functions could also be associated to a target variable z_i (again with $i = 1, \dots, I$) and we want to learn a supervised model that allows us to predict the output from the input curves. Such variable can be real-valued or a class label, thus leading to functional extensions of regression or classification, respectively. There exist functional extensions of analysis of variance and various linear regression and classification models [12], multi-layer perceptrons and radial basis function networks [25,26], support vector machines [27] and other non-parametric techniques [13]. In order to perform mathematical manipulations of functions, most of the functional data analysis methods instead of analyzing the actual (discrete) observations replace them with estimated functional representations (continuous and smooth) obtained through basis fitting approximations. For instance, Saeys et al. [28] suggest that for spectroscopic data, especially in the NIR domain, B-Splines are the most appropriate choice because of their resemblance with absorption peaks and their compact support. According to Rossi [25], multi-layer perceptrons can be directly used on sampled functional data, thus leading to functional classification and regression machines that do not require the simplified representation from basis expansions.

Traditionally, most applications of functional data analysis have been on longitudinal data in fields like econometrics, medicine, meteorology, agricultural and behavioral sciences [24]. However, the use of functional data analysis techniques is growing with applications to parameter estimation in differential [29] land usage prediction based on satellite images [30], analysis of mass spectra for proteomics [31,32] and fluorescence curves [33], to mention a few.

3. Materials: hyperspectral image

In this section we will describe some basic ideas about hyperspectral images, then we present the specific dataset that is treated in the computational experiments, and finally we discuss the data structures used in the experiments. The main application is identification of soybean crops for precision agriculture monitored by remote sensing.

3.1. Hyperspectral images

An image sensor captures the irradiance reflected from an observed scene. RGB sensors capture light energy in the red, green and blue bands roughly corresponding to the human vision sensitivity in the visible spectrum (from 400 nm to 780 nm). Hyperspectral cameras overcome the human visual system perceptual wavelength resolution. Current hyperspectral cameras have a sensitivity ranging from 100 nm to 2400 nm, covering also part of the infrared spectrum. This wavelength detection range is divided into n disjoint and contiguous intervals. The energy detected in the n th wavelength interval is stored into the n th image band. Hyperspectral cameras provide image spectral resolution over hundreds of bands. Fig. 1 illustrates the structure of a hyperspectral image. Different hyperspectral sensors have different

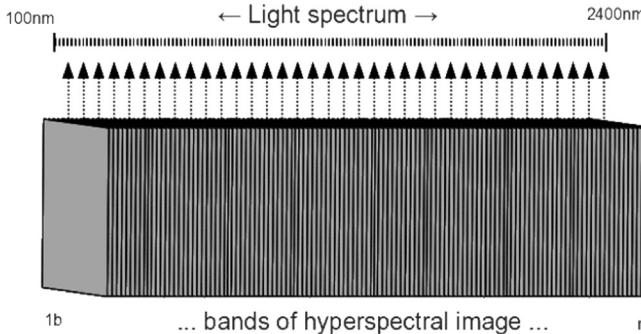


Fig. 1. Representation of an hyperspectral image as a cube of data.

resolutions: the SOC 710, *Surface Optics*, covers the wavelength range from 333 to 900 nm into 128 bands, the AVIRIS sensor covers wavelength range from 400 to 2500 nm into 224 bands, the Hyperion satellite sensor covers the same range in 220 bands with a 30-meter spatial resolution imaging a 7.5 km by 100 km land area per image.

3.2. Soybean remote sensing data

This paper focuses on a Brazilian soybean crop area. Brazil is the second world producer of soybean. The state of Mato Grosso alone produces nearly 30% of the worlds' oilseed crop. The study area is the Tanguru farm located in Querência municipality, Mato Grosso state,¹ central Brazil (Fig. 2(a)). This area is characterized by a flat topography (350 m altitude) with Latossolo Vermelho-Amarelo distrófico soil predominance (Typic Acrustox, in the USA Soil Taxonomy). The climate type is designated as tropical, with mean temperature of 26 °C and annual rainfall of 1850 mm, with a well-defined dry season from May to September and a rainy season from October to April. The native vegetation comprises the transition between savannas and Amazonian tropical rainforest.

Five soybean varieties were planted at this farm in the 2004/2005 growing season, covering approximately 8500 ha: Perdiz, Monsoy 8411, Monsoy 9010, Kaiabi, and Tabarana. Perdiz was predominant in the study area and Monsoy 8411 presented the shortest life cycle (116 days). The soybean varieties were planted in different dates from October to November and from March to April. Fig. 2(c) shows the distribution of different classes: Perdiz in white, Monsoy 8411 in green, Tabarana in yellow, Kaiabi in red and Monsoy 9010 in blue. The hyperspectral data was obtained by the Hyperion sensor.

Hyperion acquired the image used in the computational experiments on February 8, 2005, with 21° view angle in the forward scattering direction, in which shade effects predominated. Image pre-processing involved the application of an algorithm for striping removal [34] and another one for wavelength recalibration by spectral matching of both the MODTRAN-modeled and observed radiance spectra [35]. This spectral matching procedure is also useful to reduce the smile effect, which may produce variations between 2.6 and 3.6 nm in the visible and near-infrared (VNIR) and less than 1 nm in the SWIR [36]. Such variations have a small impact on crop classification considering the 10-nm bandwidth of Hyperion. Keystone effects were not corrected. After radiometric calibration 196 spectral bands were retained for further analysis.

Fig. 3(a) shows the spectral signatures of a sample pixel from each crop class in the normalized intensity range [0–1] after calibration, the abscissa corresponds to the actual spectral band number. The color of each soybean varieties spectral representative is the same as in Fig. 2(c) except for the Perdiz variety that is black. Differences between signature representatives of soybean classes are very small along the 196 bands. Bands with low signal to noise ratio have been removed at the calibration step. Fig. 3(b) shows the compacted spectral plots, obtained removing the bands that have no signal information. This second and final representation retains 145 bands from five continuous sections of the original bands: 1–44, 45–61, 62–80, 81–112, 113–145, hence spectra have four discontinuities that must be taken into account for feature extraction processes: 44–45, 61–62, 80–81 and 112–113.

3.3. Data structures

In this section we will specify some notation and information about the data structures used for the computational experiments. Let us denote a hyperspectral image as I , a pixel spectra as $I_{x,y}$ where x,y refer to the domain coordinates, and a single spectral value as $I_{x,y,b}$ where b is the index band. After pre-processing and segmentation of the image, detecting image regions corresponding to the spatial crop plots, we copy all pixel spectra corresponding to soybean into a bidimensional matrix : $TS_{S \times B}$ where S is the number of soybean pixels and B is the number of spectral bands. A vector lts of size S contains the ground truth providing the class of each sample. We extract randomly n samples of each class without repetition from TS , building training data matrix $TR_{S' \times B}$ and the corresponding ground truth vector ltr . Then we have a training set composed of TR and ltr , and a test set composed of TS and lts . In addition, the domain coordinates of each sample in the test set is saved in an structure ctr , used to visualize the class prediction as color image representations of the thematic map. In our case study: $S=53,694$, $S'=1500$, $B=145$, and $n=300$. We have selected $n=300$ pixels from each class for the training set, in order to compare with previous results reported on the same data [10]. On the other hand, the samples in TS are distributed into the classes as shown in Table 1. Specifically in this work, the FDA point of view leads to the consideration of the first and second order derivatives of the original spectral signature. Therefore, the data structures TS and TR are of size $(S \times 3*B)$ and $(S' \times 3*B)$, respectively, after the addition of the spectral derivatives. Summarizing, Table 2 enumerates the data structures needed for the implementation of the presented approach.

4. Extreme learning machines

In this section we review the fundamental definitions of the two learning algorithms applied in the experiments reported below to classify the soybean pixels in the hyperspectral image described in the previous section: the ELM and the OP-ELM.

4.1. ELM

Extreme Learning Machine (ELM) [14,37] is a simple supervised learning algorithm for Single-hidden Layer Feedforward Neural Network (SLFN). Let us have N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where input vector patterns are $\mathbf{x}_i = [x_1^i, x_2^i, \dots, x_n^i]^T \in \mathbb{R}^n$ and target output values are $\mathbf{t}_i = [t_1^i, t_2^i, \dots, t_L^i]^T \in \mathbb{R}^m$. The output of a SLFN with L hidden nodes is computed as follows:

$$f_L(\mathbf{x}) = \sum_{i=1}^L \beta_i \cdot g(\mathbf{w}_i \mathbf{x}_j^T + b_i) \quad (1)$$

¹ http://upload.wikimedia.org/wikipedia/commons/4/4f/Brazil_State_Mato_Grosso.svg.

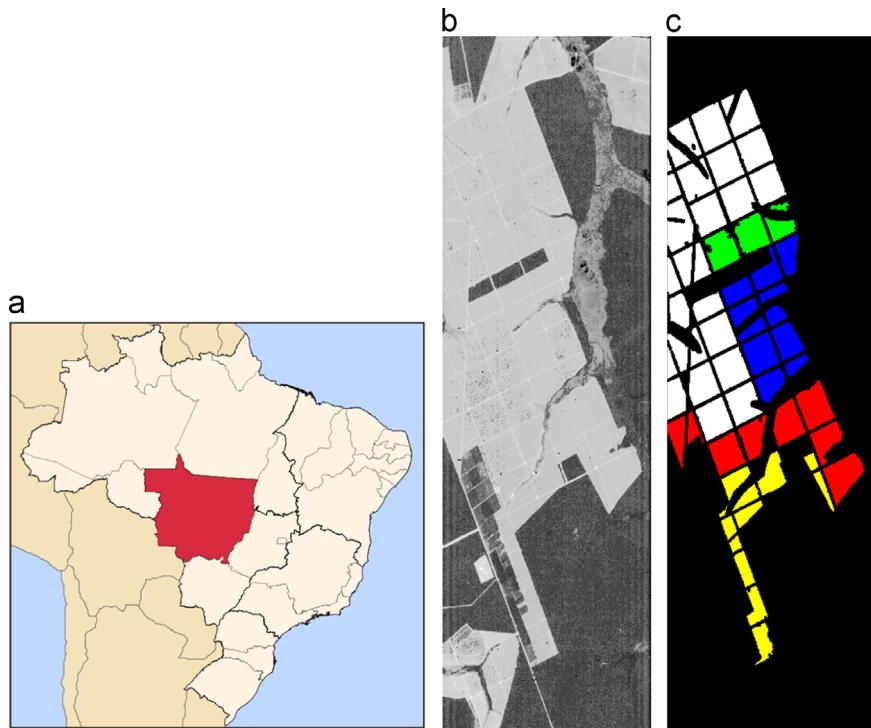


Fig. 2. Location of the Tanguru farm in Brazil. (a) The Mato Grosso State localization, (b) intensity image corresponding to one band, and (c) the ground truth, where soybean classes are identified by colors. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

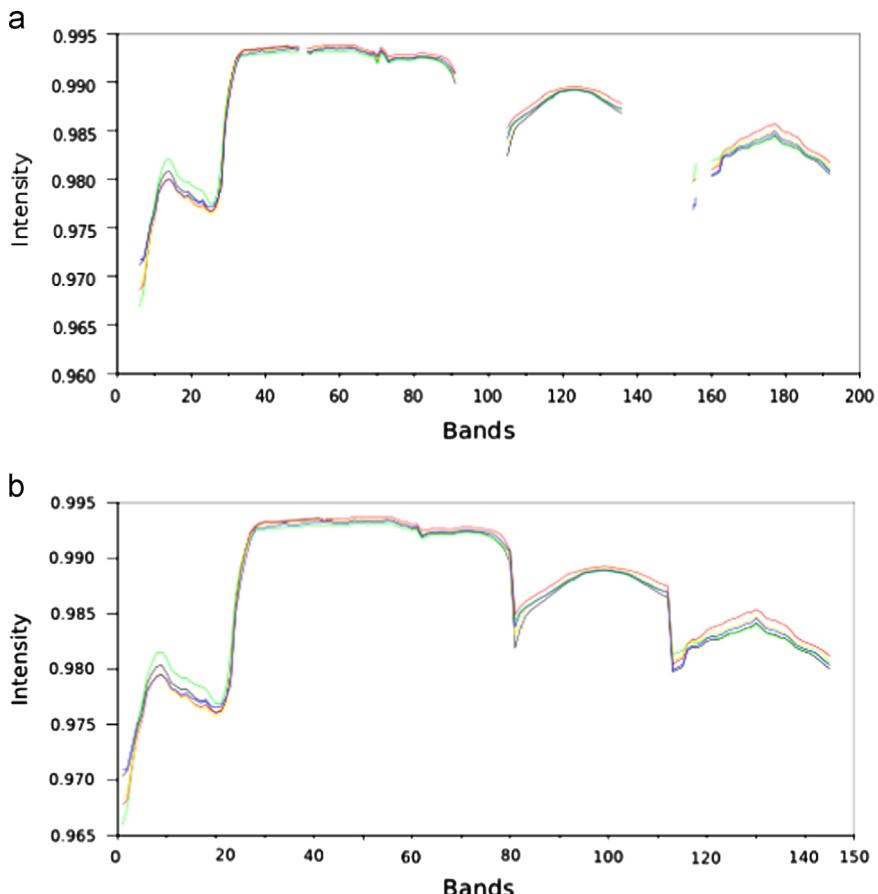


Fig. 3. Spectral signatures of soybean varieties. The color of each soybean variety is the same as in Fig. 2(c) except for the Perdiz variety which is black in the signature plots. (a) Original plot including low signal-to-noise bands as gaps in the plot, (b) compacted plot after removal of low signal-to-noise bands. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

where $\mathbf{x} \in \mathbb{R}^n$, $\beta_i[\beta_1^i, \beta_2^i, \dots, \beta_L^i]^T \in \mathbb{R}^m$ are the output weights from the hidden unit to the output units, $\mathbf{w}_i = [w_1^i, w_2^i, \dots, w_n^i]^T$ are the hidden unit weights, b_i is the threshold of the i th hidden neuron,

Table 1

Classes in the dataset, number of samples and color in the visualization of the ground truth.

#	Class	#Samples	Color
1	Perdiz	27,614	White
2	Monsoy 8411	3495	Green
3	Monsoy 9010	8451	Blue
4	Kaiabi	9651	Red
5	Tabarana	5634	Yellow

Table 2

Data structures used in the implementation of the proposed approach.

$TS_{S \times 3B}$	Matrix containing all soybean spectra samples (Test set)
$lts_{S \times 1}$	Vector containing the ground truth class of each sample in TS
$cts_{S \times 2}$	Vector containing the domain coordinates (x, y) of each sample in TS
$TR_{S' \times 3B}$	Matrix containing S' samples (Train set)
$ltr_{S' \times 1}$	Vector with the ground truth class of each sample in TR

and g denotes the non-linear activation function of the hidden node. It can be the identity, sigmoid or Gaussian function, among a large collection of polynomial functions [38]. Training the SLFN to learn the map between the input and output patterns corresponds to solve the following set of non-linear equations:

$$\sum_{i=1}^L \beta_i \cdot g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = t_j, \quad j = 1, \dots, N. \quad (2)$$

The non-linearity of Eq. (2) comes from the activation function of the hidden units. The contribution of ELM is the realization that once the output of the hidden units is fixed, the remaining problem becomes a set of linear equations, which can be solved by minimum square error estimation. Eq. (1) can be written in matrix form as

$$\mathbf{H}\beta = \mathbf{T}. \quad (3)$$

where \mathbf{H} , of size $N \times L$, is the matrix composed of the SLFN hidden layer outputs activated by the input samples, β is the matrix of hidden-to-output weights of size $L \times m$, and \mathbf{T} is the target matrix with size $N \times m$. Training of SLFN is accomplished by first setting randomly the hidden units' weights, and, second, computing the least-squares estimation $\hat{\beta}$ solving the linear system $\mathbf{H}\beta = \mathbf{T}$, given

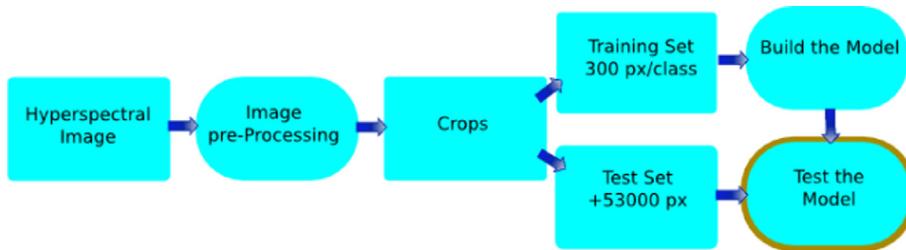


Fig. 4. Experimental flow diagram of the computational experiments. Squares correspond to data objects whereas circles correspond to computational processes.



Fig. 5. Image pre-processing sequence of results, (a) image intensity, (b) chromatic gradient, (c) intensity gradient, (d) crop plot region identification, (e) class assignment to crop plot regions.

in closed form by $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$, where \mathbf{H}^\dagger is the Moore–Penrose inverse of \mathbf{H} . The ELM training algorithm is summarized in [Algorithm 1](#).

Algorithm 1. ELM training algorithm.

Given a training set $\mathcal{N} = (x_i, t_i) | x_i \in \mathbb{R}^d, t_i \in \mathbb{R}^m, i = 1, \dots, L$, activation function $g(x)$, and L hidden neurons,

- (1) Assign arbitrary input weight w_i and bias $b_i, i = 1, \dots, L$.
- (2) Generate the random hidden layer weight matrix \mathbf{H} .
- (3) Calculate the output weights $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$

4.2. Optimally pruned ELM

The proposition of a pruned ELM to increase robustness of the classifier in [\[16\]](#) led to the definition of the Optimally pruned ELM (OP-ELM) [\[15\]](#), which is a variation of the ELM introducing an optimal selection of the number of hidden units and variables modeling the problem, alternative to incremental approaches to ELM model selection [\[39,40\]](#). The OP-ELM works in three steps:

- (1) Building an over-parameterized ELM.
- (2) Ranking of the hidden layer neurons by their contribution to the linear explanation of the ELM output by the Multi-Response Sparse Regression (MRSR). The MRSR profits from the linearity of the output layer. It proceeds by adding columns of the regressor matrix \mathbf{H} in $\mathbf{H}\hat{\beta} = \mathbf{T}$, and corresponding non-zero rows in $\hat{\beta}$, obtaining a series of approximations $\mathbf{H}^k\hat{\beta}^k = \mathbf{T}^k$. Hidden nodes are ordered by the corresponding decrease in the prediction error $\|\mathbf{T}^k - \mathbf{T}\|$ obtained including them in the model.
- (3) Leave one out (LOO) validation: compute the LOO for a given number of hidden units, which can be done efficiently by the PRESS statistic in the linear case:

$$\epsilon^{PRESS} = \frac{t_i - \mathbf{h}_i \mathbf{b}_i}{1 - \mathbf{h}_i \mathbf{P} \mathbf{h}_i^T}, \quad (4)$$

where \mathbf{h}_i and \mathbf{b}_i are the i th column and i th row of \mathbf{H} and $\hat{\beta}$, respectively. The process is greedy incremental, units are added in order until the LOO decreases below a preset threshold. Efficiency is increased if the order of unit inclusion is guided by the results of the previous MRSR.

5. Experimental design

[Fig. 4](#) shows the flow diagram of the computational experiments carried out. Rectangle boxes correspond to data and rounded boxes correspond to computational processes. From left to right, the *hyperspectral image* enters the image pre-processing, which segments the image into *crops*, corresponding to image regions containing separate soybean crop plots. Separate *training set* (*TR*) and *test set* (*TS*) are extracted from the identified crop regions as described in [Section 3.3](#). The size per class of the training and test data is 300 and 53,000 samples, respectively, as in [\[11\]](#). The training set is used to build the model, which is tested against the test data in the *Test the model* process represented by the lower rightmost box.

The model building process includes the search for the minimal set of features providing the maximal pixel-wise classification accuracy. Following the FDA approach we add the first and second order derivatives of the spectra as features of each band. We have found in the experiments that the second derivative is highly

informative for classification. Summarizing, the process followed to select the best bands for ELM and OP-ELM is:

- First step: Build separate classifiers on the features corresponding to each band, performing separate validations on the test set. The test accuracy is assumed as the saliency of the band, denoted *BB*.
- Second step: Sort bands by their saliency.
- Third step: Afterward, proceed by an incremental training of the classifiers on the orderly adding band features to the

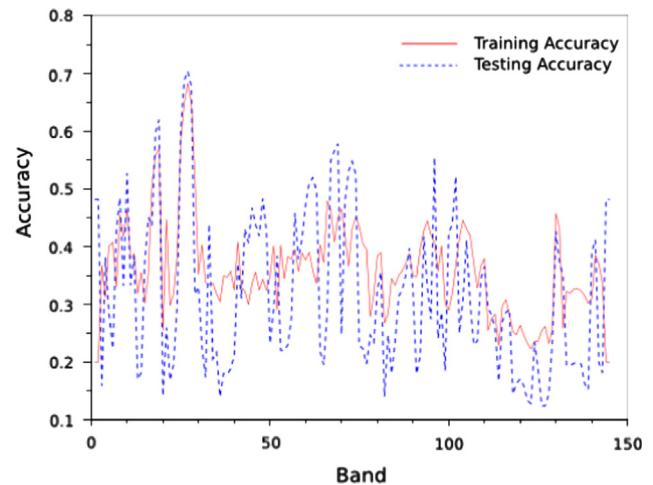


Fig. 6. Accuracy obtained training independent ELMs on each band's features. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

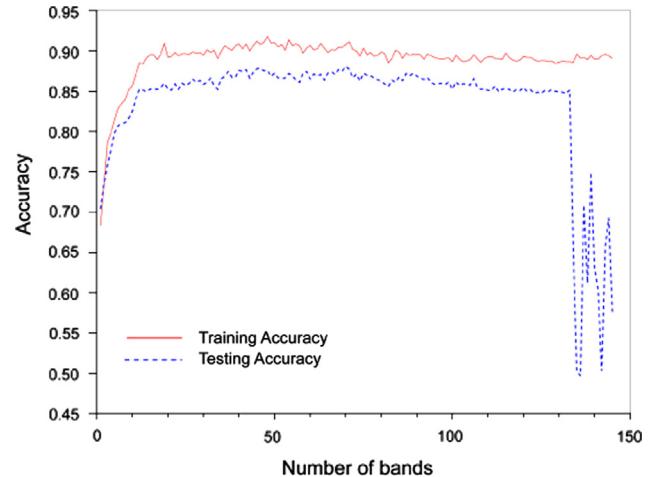


Fig. 7. Accuracy obtained during the greedy incremental search adding features of bands according to their saliency.

Table 3

Summary of ELM classification performance results with diverse number of spectral bands selected as explained in the text. Training and testing datasets selected according to [\[11\]](#).

#Bands	Accuracy on testing	MAE	Kappa
12	0.855557	0.144443	0.791923
37	0.865093	0.134907	0.806339
46	0.878549	0.121451	0.824533
70	0.882432	0.117568	0.830018

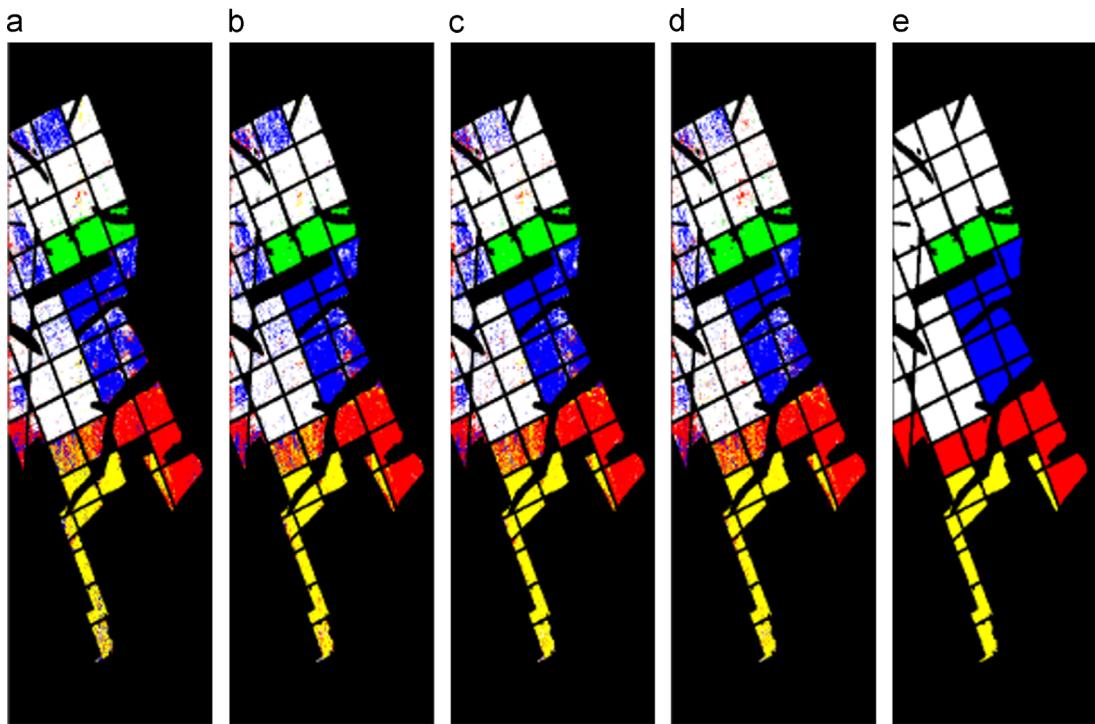


Fig. 8. Soybean variety thematic map obtained from different numbers of bands included in the feature vector during the greedy process. (a) 12, (b) 37, (c) 46, (d) 70, (e) ground truth. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

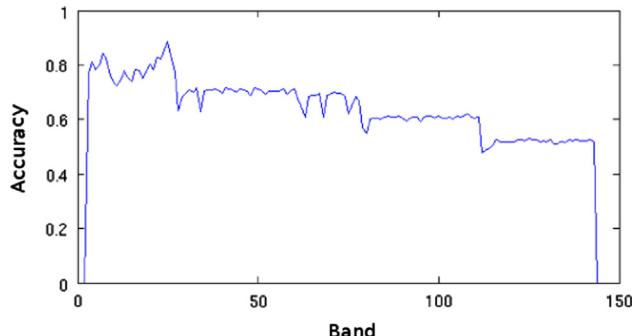


Fig. 9. Testing accuracy obtained by independent OP-ELM trained on each band's features.

training and test sets according to their saliency. At most, the number of classifiers trained and tested is the number of bands.

- Fourth step: Select the classifier with the best testing accuracy.

5.1. Image pre-processing

We apply the hyper-spherical coordinate transformation [41] on the hyperspectral image. This transformation allows us to separate image intensity from image chromaticity. Fig. 5(a) shows the image intensity. By using the chromatic gradient [41] we have found noisy pixels to be removed from the computations. Fig. 5(b) shows the chromatic gradient computed on the hyper-spherical coordinate representation of the image pixel spectra. By inspection it is easy to see that the edges between the crop plots are easy to detect on the intensity gradient shown in Fig. 5(c). Once the plot boundaries are found, crop plot image regions are identified as shown in Fig. 5(d). Finally, we apply the ground truth information, labeling each pixel with the correspondent class. Fig. 5(e) shows the resulting pixel assignment to classes according to the ground truth thematic map.

Table 4

Comparison of OP-ELM classification performance results on the testing data of single band #25 with state of the art classifiers. Training and testing datasets selected according to [11].

Classifier	A	K	MAE
BayesNet	0.636	0.494	0.192
SVM	0.675	0.547	0.259
1-NN	0.679	0.555	0.128
C 4.5	0.506	0.026	0.314
OP-ELM	0.832	0.762	0.167

6. Soybean classification results

First, we report the results of computational experiments carried out with ELM analyzing the accuracy versus the dimensionality reduction and their consequences on each class prediction. Second, we show the results of computational experiments carried out with OP-ELM, looking for the band with best accuracy. Finally, we report a comparison of the results of the OP-ELM against state of the art algorithms reported in [11] under the same conditions, i.e. the same dataset for training and testing.

6.1. ELM Experiment

6.1.1. First step: Independent ELM accuracy on each band

Fig. 6 plots the accuracy achieved training over the data of each band independently, the red line corresponds to the training data whereas blue line corresponds to the testing data. It can be appreciated that some bands achieve better accuracy than others. Note too, that there are many bands reaching higher accuracy on the testing data than on the training data, and others show just the opposite results. In the following, testing accuracy is assumed as the band saliency.

6.1.2. Second step: Incremental ELM training adding bands

After sorting the bands according to their saliency, the experiment follows a greedy feature space growing process. Band features are added in order of saliency, one at a time. The ELM is trained on the current feature space. Fig. 7 plots the accuracy achieved versus the number of bands considered in the training/testing data. During the early stages of this greedy process, accuracy improves quickly. After the addition of 12-th band accuracy growth stagnates until it begins to worsen after the addition of the 135th band.

6.1.3. Third step: Selecting the best ELM

Table 3 summarizes the results of Fig. 7. The first column shows the number of bands included in the feature vectors of the training/testing data, the second column shows the accuracy achieved on testing data, the third column shows the mean absolute error (MAE) and last one shows the Kappa coefficient. The image that uses the 70 best bands has the best accuracy, however the image that uses the 12 best bands is not far from these results. The balance between size and accuracy is a choice of the user, depending on the application requirements. The reduction in the number of bands can be interpreted as a new sensor with reduced spectral resolution. A similar experiment was reported in [11], where several sensors were simulated from the high spectral resolution data by compacting the bands. Comparing these results with those reported in [11], our approach with 12

bands provides better results than the classifications from the foremost multispectral sensors reported there (Table 4 in [11]) reports accuracies: 68.85, 85.94, 86.72, 80.47, 76.46, 82.87, 82.97). Using the 46 or 70 best bands the improvement that we obtain is even greater.

Fig. 8 shows the thematic maps produced by the trained ELMs. From left to right (a)–(d) show the maps obtained using the 12, 37, 46 and 70 most salient bands, respectively. Fig. 8(e) is the ground truth. Monsoy 8411 class (green) is well detected in all cases, as well as Tabarana class (yellow). Some pixels of Kaiabi class (red) are classified as Tabarana class, this misclassification is gradually reduced with the use of bigger feature vectors, the feature vector built with 70 bands gives the best result for this class. Perdiz class (white) is the more abundant having as many misclassification as Monsoy 9010 (blue). As the number of bands grows, results improve. These improvements can be better appreciated in the second crop at the top of the image. Finally Monsoy 9010 has some misclassifications as Perdiz.

6.2. OP-ELM experiments

In order to find the best spectral band, we train an OP-ELM on each band feature data looking for the best testing accuracy. Fig. 9 shows testing accuracy results. The best testing accuracy is found on the band #25. Further, in order to compare OP-ELM results with well-known classifiers, we train them on the data of band 25. We have used the WEKA platform [42], where we have used: (1) a Bayesian network (called BayesNet in Weka), (2) Support Vector Machine (SVM), (3) 1-NN, and (4) Quinlan's C 4.5 algorithm to build decision trees [43] (called J48 in Weka). These four classifiers are well-known. The Weka's default parameter settings have been used in the experiments for reproducibility of results.

Table 4 shows the results on testing data, when training and testing datasets selected according to [11]. The table shows the Kappa statistic (K), Accuracy (A), Mean absolute error (MAE). Best results are achieved by OP-ELM with an accuracy of 0.832, Kappa coefficient 0.762, and mean absolute error (MAE) 0.167. Therefore

Table 5
Comparison of classification performance results of OP-ELM against state of the art classifiers on a 10-fold stratified cross-validation of single band #25 data.

Classifier	A	K	MAE
BayesNet	0.711	0.659	0.132
SVM	0.664	0.530	0.279
1-NN	0.790	0.653	0.112
C 4.5	0.604	0.443	0.208
OP-ELM	0.886	0.815	0.134

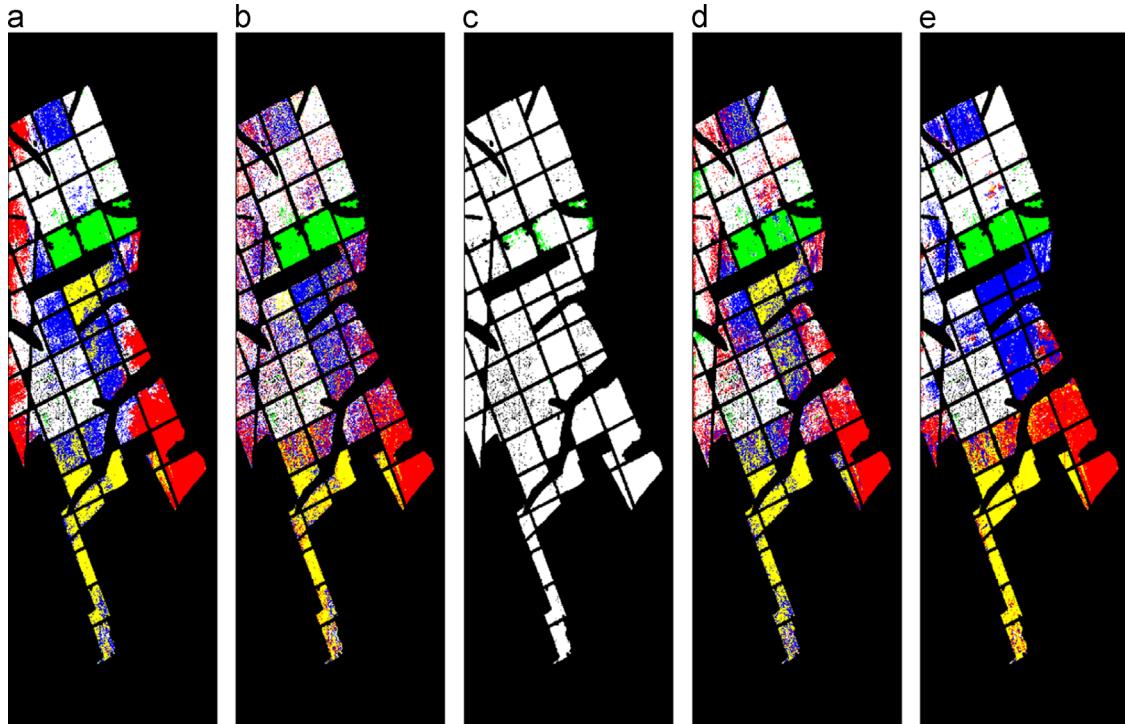


Fig. 10. Pixel-wise class estimation by using only the data in band 25 and its first and second derivatives: (a) SVM, (b) 1-NN, (c) C4.5, (d) BayesNet, (e) OP-ELM.

we can conclude that the OP-ELM gives the best and more stable results, improving also on previous results reported on this data [11] both in accuracy and in dimensionality reduction, which is maximum because a single band is used to obtain the reported results. The selection of data samples in [11] is not stratified, we have also performed a stratified 10-fold cross-validation experiment to assess the classifiers in the case of unbalanced data. Table 5 shows the results on this experiment. We appreciate an improvement of the results of all classifiers, maybe due to the improved sampling of the data distribution, with the exception of the SVM which reflects its well known sensitivity to sample balance. Still, the OP-ELM provides the best results with a strong improvement over the non-stratified data sampling.

Fig. 10 shows the thematic map produced by each classifier corresponding to results reported in Table 4. From left to right, (a) SVM, (b) 1-NN, (c) C4.5, (d) BayesNet and (e) ELM. The C4.5 decision tree provides the worst results classifying most pixels as Perdiz, with some exceptions (Monsoy 8011). Though BayesNet, SVM and 1-NN classifiers achieve similar accuracy, kappa and MAE values, we can find some salient differences in the visualization of their thematic maps. SVM classifies many Perdiz samples as Kaiabi, and many Kaiabi samples as Monsoy 9010. 1-NN distributes its misclassified pixels across the image, similar to BayesNet. The best result is offered by OP-ELM, which has the best accuracy, lower error values. Image on the right shows the results. Most of the misclassified pixels are Perdiz who are classified as Monsoy 9010. In most cases, Monsoy 8411 is correctly classified, with the exception of C 4.5.

7. Conclusions

In this paper we report results of ELM and OP-ELM on a very specific application: Soybean variety classification in hyperspectral images. We follow a FDA approach for feature extraction, consisting in the consideration of the first and second order spectral derivatives as additional features. Soybean classification has proven tough to achieve. The results reported in this paper improve our previous results on the same image data. The ELM has been tested in an incremental approach adding salient bands to the data, training a new ELM at each incremental step. Best results with 70 bands improve significantly previous results reported in the literature. We have also tested the OP-ELM, finding that it improves state-of-the art results using only the information from one band. Future works maybe addressed to a deeper and extensive study of FDA spectra representations, like modeling the spectra by B-splines or polynomial approximation, and their impact on classification accuracy.

Acknowledgments

Some authors received support from UFI11/07 of the UPV/EHU, SandS project EU grant agreement 317947, MECCO projects TIN2011-28753-C02-02, TIN2011-23823.

References

- [1] M. Shimoni, G. Tolt, C. Pernel, J. Ahlberg, Detection of vehicles in shadow areas, in: 3rd Workshop on Hyperspectral Image and Signal Processing (WHISPERS), 2011, pp. 1–4.
- [2] B. Rivard, J. Feng, V. Bushan, M. Lipsett, Infrared reflectance hyperspectral features of athabasca oil sand ore and froth, in: 3rd Workshop on Hyperspectral Image and Signal Processing (WHISPERS), 2011, pp. 1–4.
- [3] I. Megameen, Y. Deville, S. Hosseini, H. Carfantan, M. Karoui, Extraction of stellar spectra from dense fields in hyperspectral muse data cubes using non-negative matrix factorization, in: 3rd Workshop on Hyperspectral Image and Signal Processing (WHISPERS), 2011, pp. 1–4.
- [4] D. Krezhova, E. Kirova, Hyperspectral remote sensing of the impact of environmental stresses on nitrogen fixing soybean plants (*Glycine max* L.), in: 5th International Conference on Recent Advances in Space Technologies (RAST), 2011, pp. 172–177.
- [5] X. Ye, K. Sakai, Application of airborne hyperspectral imagery to estimating fruit yield in citrus, in: 3rd Workshop on Hyperspectral Image and Signal Processing (WHISPERS), 2011, pp. 1–6.
- [6] L. Guozhu, S. Kaishan, N. Shuwen, Soybean LAI estimation with in-situ collected hyperspectral data based on BP-neural networks, in: 3rd International Conference on Recent Advances in Space Technologies (RAST '07), 2007, pp. 331–336.
- [7] S.T. Monteiro, Y. Minekawa, Y. Kosugi, T. Akazawa, K. Oda, Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery, *ISPRS J. Photogr. Remote Sens.* 62 (1) (2007) 2–12.
- [8] Z. Zhou, Y. Zang, B. Shen, X. Zhou, X. Luo, Detection of cowpea weevil (*Callosobruchus maculatus* (F.)) in soybean with hyperspectral spectrometry and a backpropagation neural network, in: Sixth International Conference on Natural Computation (ICNC), vol. 3, 2010, pp. 1223–1227.
- [9] J. Clevers, L. Koosstra, Using hyperspectral remote sensing data for retrieving total canopy chlorophyll and nitrogen content, in: 3rd Workshop on Hyperspectral Image and Signal Processing (WHISPERS), 2011, pp. 1–4.
- [10] J. Feret, G. Asner, S. Jacquemoud, Regularization of discriminant analysis for the study of biodiversity in humid tropical forests, in: 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011, pp. 1–4.
- [11] F.M. Breunig, L.S. Galvao, A.R. Formaggio, J.C.N. Epiphanio, Classification of soybean varieties using different techniques: case study with hyperion and sensor spectral resolution simulations, *J. Appl. Remote Sens.* 5 (1) (2011) 053533.
- [12] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, second edition, Springer-Verlag, New York, 2006.
- [13] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis*, Springer-Verlag, New York, 2006.
- [14] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, *Neurocomputing* 70 (1–3) (2006) 489–501.
- [15] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, A. Lendasse, OP-ELM: optimally pruned extreme learning machine, *IEEE Trans. Neural Netw.* 21 (January (1)) (2010) 158–162.
- [16] H.-J. Rong, Y.-S. Ong, A.-H. Tan, Z. Zhu, A fast pruned-extreme learning machine for classification problem, *Neurocomputing* 72 (1–3) (2008) 359–366.
- [17] L. Chen, G.-B. Huang, H.K. Pung, Systematical convergence rate analysis of convex incremental feedforward neural networks, *Neurocomputing* 72 (10–12) (2009) 2627–2635.
- [18] U. Grenander, *Abstract Inference*, Wiley, New York, 1981.
- [19] P. Besse, J.O. Ramsay, Principal component analysis of sampled curves, *Psychometrika* 51 (1986) 285–311.
- [20] J. Dauxois, G.M. Nkiet, Measure of association for Hilbertian subspaces and some applications, *J. Multivariate Anal.* 82 (2002) 136–154.
- [21] G.Z. He, H.-G. Müller, J.I. Wang, Functional canonical analysis for square integrable stochastic processes, *J. Multivariate Anal.* 85 (2003) 54–77.
- [22] G.M. James, T.J. Hastie, Functional linear discriminant analysis for irregularly sampled curves, *J. R. Stat. Soc. Ser. B* 63 (2001) 533–550.
- [23] C. Abraham, P.A. Cornillon, E. Matzner-Löber, N. Molinari, Unsupervised curve-clustering using B-Splines, *Scand. J. Stat.* 30 (2003) 581–595.
- [24] J.O. Ramsay, B.W. Silverman, *Applied Functional Data Analysis: Methods and Case Studies*, Springer-Verlag, New York, 2002.
- [25] F. Rossi, B. Conan-Guez, Functional multi-layer perceptron: a nonlinear tool for functional data analysis, *Neural Netw.* 18 (2005) 45–60.
- [26] F. Rossi, N. Delannay, B. Conan-Guez, M. Verleysen, Representation of functional data in neural networks, *Neurocomputing* 64 (2005) 183–210.
- [27] F. Rossi, N. Villa-Vialaneix, Support vector machine for functional data classification, *Neurocomputing* 69 (2006) 730–742.
- [28] W. Saeyns, B.D. Ketelaere, P. Darius, Potential applications of functional data analysis in chemometrics, *J. Chemom.* 22 (2008) 335–344.
- [29] J.O. Ramsay, G. Hooker, D. Campbell, J. Cao, Parameter estimation for differential equations: a generalized smoothing approach, *J. R. Stat. Soc. Ser. B* 69 (2007) 741–796.
- [30] H. Cardot, R. Faivre, M. Goulard, Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data, *J. Appl. Stat.* 30 (2003) 1185–1199.
- [31] J. Harezlak, M.C. Wu, M. Wang, A. Schwartzman, D.C. Cristianini, X. Lin, Biomarker discovery for arsenic exposure using functional data. Analysis and feature learning of mass spectrometry proteomic data, *J. Proteome Res.* 7 (2008) 217–224.
- [32] J.S. Morris, P.J. Brown, R.C. Herrick, K.A. Baggerly, K.R. Coombes, Bayesian analysis of mass spectrometry proteomics data using Wavelet-based functional mixed models, *Biometrics* 64 (2008) 479–489.
- [33] C. Ritz, J.C. Streibig, Functional regression analysis of fluorescence curves, *Biometrics* 65 (2009) 609–617.
- [34] D. Goodenough, A. Dyk, K. Niemann, J. Pearlman, H. Chen, T. Han, M. Murdoch, C. West, Processing hyperion and ALI for forest classification, *IEEE Trans. Geosci. Remote Sens.* 41 (June (6)) (2003) 1321–1331.
- [35] T. Wang, G. Yan, H. Ren, X. Mu, Improved methods for spectral calibration of on-orbit imaging spectrometers, *IEEE Trans. Geosci. Remote Sens.* 48 (November (11)) (2010) 3924–3931.

- [36] L.B. Liao, P.J. Jarecke, D.A. Gleichauf, T.R. Hedman, Performance characterization of the hyperion imaging spectrometer instrument, in: Proceedings of SPIE on Earth Observing Systems V (SPIE 2000), vol. 4135, 2000, pp. 264–275.
- [37] G.-B. Huang, D. Wang, Advances in extreme learning machines (ELM2010), Neurocomputing 74 (16) (2011) 2411–2412.
- [38] W. Xi-Zhao, S. Qing-Yan, M. Qing, Z. Jun-Hai, Architecture selection for networks trained with extreme learning machine using localized generalization error model, Neurocomputing 102 (1) (2013) 3–9.
- [39] G.-B. Huang, L. Chen, Convex incremental extreme learning machine, Neurocomputing 70 (16–18) (2007) 3056–3062.
- [40] G.-B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, Neurocomputing 71 (16–18) (2008) 3460–3468.
- [41] R. Moreno, M. Graña, About gradient operators on hyperspectral images, in: Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods, SciTePress, 2012, pp. 433–437.
- [42] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weak data mining software: an update, SIGKDD Explor. 11 (1) (2009) 10–18.
- [43] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.

Ramón Moreno has received his Ph.D. in Computer Science under the advice of Prof. Manuel Graña at the University of the Basque Country. His interests are image processing, specifically hyperspectral images, and computational intelligence, including Neuroinspired approaches. He has published over 5 papers in international journals. After holding a postdoc position he is currently working in the industry.

Francesco Corona is senior researcher at the Environmental and Industrial Machine Learning group of the Department of Information and Computer Science at the Aalto University/Helsinki University of Technology. He received the Laurea degree in chemical engineering from the Department of Chemical engineering and Materials at the University of Cagliari (Sardinia). He received the Ph.D. (Eng.) degree in industrial engineering from the Doctoral School in Industrial Engineering at the University of Cagliari. He officially joined the Aalto University/Helsinki University of Technology in January 2007. His interest concentrates on machine learning technologies with application to process modeling, visualization, control and optimization.

Amaury Lendasse born in Tournai, Belgium on April 16th, 1972. He got a Master Degree in Mechanical Engineering in the Universite de Louvain-la-Neuve in Belgium in 1997 and a second Master Degree in Control in the same university in 1997. He got a PhD degree in 2003 in Louvain-la-Neuve under the supervision of Prof. Vincent Wertz and Michel Verleysen. After being a Docent and senior researcher in ICS, HUT, he is currently Ikerbasque fellow at the University of the Basque Country.



Manuel Graña received his Computer Engineering degree and PhD at the Computer Science School of San Sebastian at the University of the Basque Country. He is Full Professor at the University of the Basque Country since 1998. He is head of the Computational Intelligence Group, evaluated as level A by the Basque Government Education, Universities and Research department. He has advised 23 PhD thesis, co-authored more than 100 research papers in international journals, organized 4 international conferences, edited more than 20 books on specific research topics or conference proceedings.

Lênio Soares Galvão is senior researcher at the Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil. His current interest included in the application of computational intelligence to remote sensing images and data.