Remote Sensing of Environment 113 (2009) 2345–2355

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas

Michele Dalponte [a,b], Lorenzo Bruzzone [a,*], Loris Vescovo [b], Damiano Gianelle [b]

[a] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 14 I-38123, Povo, Trento, Italy
[b] Centro di Ecologia Alpina, Fondazione E. Mach, Viote del Monte Bondone, I-38040 Trento, Italy

ARTICLE INFO

Article history:
Received 19 November 2008
Received in revised form 18 June 2009
Accepted 20 June 2009

Keywords:
Hyperspectral images
Forestry
Spectral resolution
Channel selection
Feature selection
Classification techniques
Remote sensing

ABSTRACT

Remote sensing hyperspectral sensors are important and powerful instruments for addressing classification problems in complex forest scenarios, as they allow one a detailed characterization of the spectral behavior of the considered information classes. However, the processing of hyperspectral data is particularly complex both from a theoretical viewpoint [e.g. problems related to the Hughes phenomenon (Hughes, 1968) and from a computational perspective. Despite many previous investigations that have been presented in the literature on feature reduction and feature extraction in hyperspectral data, only a few studies have analyzed the role of spectral resolution on the classification accuracy in different application domains. In this paper, we present an empirical study aimed at understanding the relationship among spectral resolution, classifier complexity, and classification accuracy obtained with hyperspectral sensors for the classification of forest areas. We considered two different test sets characterized by images acquired by an AISA Eagle sensor over 126 bands with a spectral resolution of 4.6 nm, and we subsequently degraded its spectral resolution to 9.2, 13.8, 18.4, 23, 27.6, 32.2 and 36.8 nm. A series of classification experiments were carried out with bands at each of the degraded spectral resolutions, and bands selected with a feature selection algorithm at the highest spectral resolution (4.6 nm). The classification experiments were carried out with three different classifiers: Support Vector Machine, Gaussian Maximum Likelihood with Leave-One-Out-Covariance estimator, and Linear Discriminant Analysis. From the experimental results, important conclusions can be made about the choice of the spectral resolution of hyperspectral sensors as applied to forest areas, also in relation to the complexity of the adopted classification methodology. The outcome of these experiments are also applicable in terms of directing the user towards a more efficient use of the current instruments (e.g. programming of the spectral channels to be acquired) and classification techniques in forest applications, as well as in the design of future hyperspectral sensors.

1. Introduction

In the study of forest environments, and in particular of complex forest areas, the choice of the most suitable spectral and spatial resolution for classification is a very important problem. Many studies have been carried out on the classification of forest areas with multispectral sensors (e.g. Sedano et al., 2005; Wang et al., 2004). However, satellite multispectral data are usually characterized by a low spectral resolution that decreases when the spatial resolution increases. Despite, a relatively low spatial resolution can be useful in the study of plantation forests (or of forests characterized by the presence of only one tree specie), often it is not suitable in the study of dense natural forests with many mixed specie. Thus, the requirement to have accurate maps at a high spatial resolution increases the need to use airborne hyperspectral data, which can acquire images having both high spectral and spatial resolutions. These sensors acquire images in hundreds of spectral channels, providing a huge amount of useful data on the analyzed area. As an example, Dalponte et al. (2008) studied a forest area in Italy characterized by 23 different classes reaching accuracies of about 90% with hyperspectral data acquired at a spectral resolution of 4.6 nm in 126 bands. Clark et al. (2005) studied seven deciduous tree species with the HYDICE sensor, using three different classifiers, reaching accuracies to the order of 90%. Martin et al. (1998) discriminated 11 forest classes with AVIRIS data, obtaining an overall kappa accuracy of 68% using 9 spectral bands.

An important property of modern hyperspectral sensors (see Table 1 for a review of the most recent instruments) is that they have a programmable definition of the spectral resolution and of the distribution of the channels in the spectrum. This means that, within the boundaries of each sensor and depending also on the considered portion of the spectrum, it is possible to tune the channels acquired by the sensor

* Corresponding author.
E-mail addresses: michele.dalponte@disi.unitn.it (M. Dalponte), lorenzo.bruzzone@disi.unitn.it (L. Bruzzone), vescovo@cealp.it (L. Vescovo), gianelle@cealp.it (D. Gianelle).

0034-4257/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2009.06.013

Table 1
Major recent hyperspectral sensors and their related spectral properties.

Sensor name	Manufacturer	Platform	Maximum number of bands	Maximum spectral resolution	Spectral range
Hyperion on EO-1	NASA Goddard Space Flight Center	Satellite	220	10 nm	0.4–2.5 μm
MODIS	NASA	Satellite	36	40 nm	0.4–14.3 μm
CHRIS Proba	ESA	Satellite	up to 63	1.25 nm	0.415–1.05 μm
AVIRIS	NASA Jet Propulsion Lab	Aerial	224	10 nm	0.4–2.5 μm
HYDICE	Naval Research Lab	Aerial	210	7.6 nm	0.4–2.5 μm
PROBE-1	Earth Search Sciences Inc.	Aerial	128	12 nm	0.4–2.45 μm
CASI 550	ITRES Research Limited	Aerial	288	1.9 nm	0.4–1 μm
CASI 1500	ITRES Research Limited	Aerial	288	2.5 nm	0.4–1.05 μm
SASI 600	ITRES Research Limited	Aerial	100	15 nm	0.95–2.45 μm
TASI 600	ITRES Research Limited	Aerial	64	250 nm	8–11.5 μm
HyMap	Integrated Spectronics	Aerial	125	17 nm	0.4–2.5 μm
ROSIS	DLR	Aerial	84	7.6 nm	0.43–0.85 μm
EPS-H (Environmental Protection System)	GER Corporation	Aerial	133	0.67 nm	0.43–12.5 μm
EPS-A (Environmental Protection System)	GER Corporation	Aerial	31	23 nm	0.43–12.5 μm
DAIS 7915 (Digital Airborne Imaging Spectrometer)	GER Corporation	Aerial	79	15 nm	0.43–12.3 μm
AISA Eagle	Spectral imaging	Aerial	244	2.3 nm	0.4–0.97 μm
AISA Eaglet	Spectral imaging	Aerial	200	–	0.4–1.0 μm
AISA Hawk	Spectral imaging	Aerial	320	8.5 nm	0.97–2.45 μm
AISA Dual	Spectral imaging	Aerial	500	2.9 nm	0.4–2.45 μm
MIVIS (Multispectral Infrared and Visible Imaging Spectrometer)	Daedalus	Aerial	102	20 nm	0.43–12.7 μm
AVNIR	OKSI	Aerial	60	10 nm	0.43–1.03 μm

to the characteristics of the specific problem under analysis. As an example, it is possible to have a denser sampling of the spectral signature in a given region of the spectrum, and a sparser sampling in others regions.

On the one hand, if the use of hyperspectral data allows one to face complex classification problems, on the other hand the hyperdimensionality of the feature space produces some drawbacks connected with the classification algorithm to use. Indeed, only a few classification algorithms are able to fully exploit the huge amount of data provided by hyperspectral sensors. One of the main problems in classification of hyperspectral data is the Hughes phenomenon (Hughes, 1968). This phenomenon arises when the ratio between the number of input features (and thus of classifier parameters) to the number of training samples is small (ill-posed problems), and so results in a decrease of the accuracy in the estimation of the classifier parameters when increasing the number of features used and thus in poor generalization ability of the classifier. This is the case for the Gaussian Maximum Likelihood classifier where estimations of the covariance matrices and mean vectors are affected by a small ratio between the number of training samples and the number of features used. Thus it becomes very critical in the hyperspectral case in which the number of features is higher than the number of training samples for each class, thus resulting in singular covariance matrices that cannot be used in the classification task.

Another important variable to consider in the analysis of hyperspectral data is the "complexity" of the classification algorithm, which in this paper is defined as the capability of a classifier to model highly non-linear decision boundaries. Usually classifiers with higher complexity are potentially more effective than algorithms with smaller complexity, especially for difficult classification problems. However, effective distribution free classifiers require the estimation of a high number of parameters in the learning phase, thus resulting intrinsically more vulnerable to the Hughes phenomenon. In this framework, it is very important, given a specific application, to identify the limit of the spectral resolution over which the discrimination between classes does not change significantly. This limit is also determined by the capability of the classifier to exploit features with a very detailed characterization of the spectral signature, and thus it depends on the complexity of the classification algorithm. It is worth noting that by fixing the Instantaneous Field Of View (IFOV) and the radiometric resolution of the sensor, a decrease in the spectral resolution will produce a better signal-to-noise ratio (SNR) on the acquired signal. This means that relatively simple classifiers could take advantage of a decrease in spectral resolution, especially if they cannot address the complexity of hyperdimensional classification problems.

In the literature several studies have focused on the selection of the optimal sets of hyperspectral channels for use in the classification phase. Of these many focused on the development of algorithms for the selection of the optimal features, given a certain classification problem. In this context, we can recall the feature selection algorithms based on a search strategy and a separability measure. Common search strategies on hyperspectral data are the Sequential Forward Floating Selection (Pudil et al., 1994) and the Steepest Ascent (Serpico & Bruzzone, 2001). As a separability measure, we find the Bhattacharyya distance (Bhattacharyya, 1943; Djouadi et al., 1990), the Jeffries–Matusita distance (Bruzzone et al., 1995), as well as the transformed divergence distance (Richards & Jia, 1999). Others studies have analyzed the location of the most informative channels in the spectrum by considering the physical meaning of each band (e.g. Backer et al., 2007; Schmidt & Skidmore, 2003). Among them, we recall the study of Backer et al. (2007), where the authors analyzed different band selection methodologies and different spectral resolutions on a CASI 2 image acquired in 46 bands.

Despite the aforementioned papers addressing the analysis and the selection of the spectral channels, little attention has been devoted to the study of the relationship among spectral resolution and classifier complexity in forest applications. Nevertheless, given a certain classifier it is interesting to know the optimal spectral resolution to use in the classification of complex forest areas. Thus, the objective of this paper is to present an empirical analysis on the relationship among spectral resolution, classifier complexity and classification accuracy on a complex forest area with hyperspectral data. In particular we analyzed the behavior of the classification accuracy of different classifiers (based on different theoretical principles) versus: i) the spectral resolution of the sensor; and ii) the number of features acquired at a high spectral resolution (4.6 nm). This analysis has practical applications in terms of directing more efficient application of the current instruments (e.g. programming of the spectral channels to be acquired) and in terms of selection of classification techniques in forest applications, as well as being useful for the design of future hyperspectral sensors. Although this paper is focused on forest application, the proposed analysis is quite general and can be easily extended to other domains.

Table 2
Distribution of reference data samples (pixels) among investigated classes of the *Bosco della Fontana* dataset (in brackets the number of trees).

Class name	Reference data samples	Class name	Reference data samples	Class name	Reference data samples
Acer campestris	170 (10)	*Juglans regia*	1573 (35)	*Q. rubra*	1137 (21)
Acer negundo	48 (3)	*Morus sp.*	164 (5)	*Robinia pseudoacacia*	1008 (40)
Alnus glutinosa	507 (27)	*Platanus hybrida*	2048 (68)	*Rubus*	661
C. betulus	910 (68)	*Populus canescens*	244 (5)	Shadows	290
Corylus avellana	58 (6)	*P. hybrida*	211 (7)	Snags	205 (10)
Fraxinus angustifolia	787 (28)	*Prunus avium*	261 (19)	*Tilia cordata*	507 (10)
Grassland	496	*Q. cerris*	1796 (47)	*Ulmus minor*	403 (17)
Juglans nigra	1283 (50)	*Q. robur*	2049 (63)		

The paper is organized into four sections. The next section presents the data sets used in the study, while the pre-processing procedures applied to the data, and the classifiers used in the analysis are presented in Section 3. Section 4 illustrates and discusses in detail the empirical results obtained. Finally, Section 5 draws the conclusions of the work.

2. Data set descriptions

In this study we considered two data sets related to forest areas with different properties. These data sets are described in the following two subsections.

2.1. Data set 1: Bosco della Fontana

The first data set considered is the natural reserve of *Bosco della Fontana*, which is a Floodplain forest near the city of Mantua (Italy), and is one of the best preserved forest relicts on the Po Plain. The central point of the area has the following coordinates: 45°12′1.68″ N, 10°44′35.53″ E. This area extends across approximately 230 ha and its topography is almost perfectly flat. It can be considered a complex forest area as, thanks to the absence of a significant human impact in the last century, it exhibits the following interesting properties: i) it is a very dense forest area; ii) it contains a high number of different species; iii) it consists of several similar tree species, including *Quercus cerris*, *Quercus robur* and *Quercus rubra*; iv) it does not exhibit a preordered spatial tree distribution.

In this area 19 tree species were identified, and four land-cover types were considered in the classification procedure in order to have an exhaustive coverage of all the classes present in the image (see Table 2 for a detailed description of the investigated classes). It is worth noting that among the 19 tree species under analysis there are classes belonging to the same genus, which have very similar spectral signatures. Another important consideration with respect to this data set is that in the analyzed area the vegetation classes do not have the same relative frequency, and that there are some dominant species (e.g. *Carpinus betulus*, *Q. cerris*, *Q. robur* and *Q. rubra*).

The hyperspectral image (see Fig. 1) was acquired on June 28th, 2006 between 9:04 AM and 9:36 AM. It consists of six partially overlapping images acquired by an AISA Eagle sensor in 126 spectral bands, ranging from 400 nm to 990 nm, with a spectral resolution of about 4.6 nm and a spatial resolution of 1 m. The flight direction of the plane was the same for all the six images (from East to West) and the flight height was approximately 750 m.

The reference data samples were collected during a ground survey in autumn 2006 (approximately 540 trees). Samples were collated on

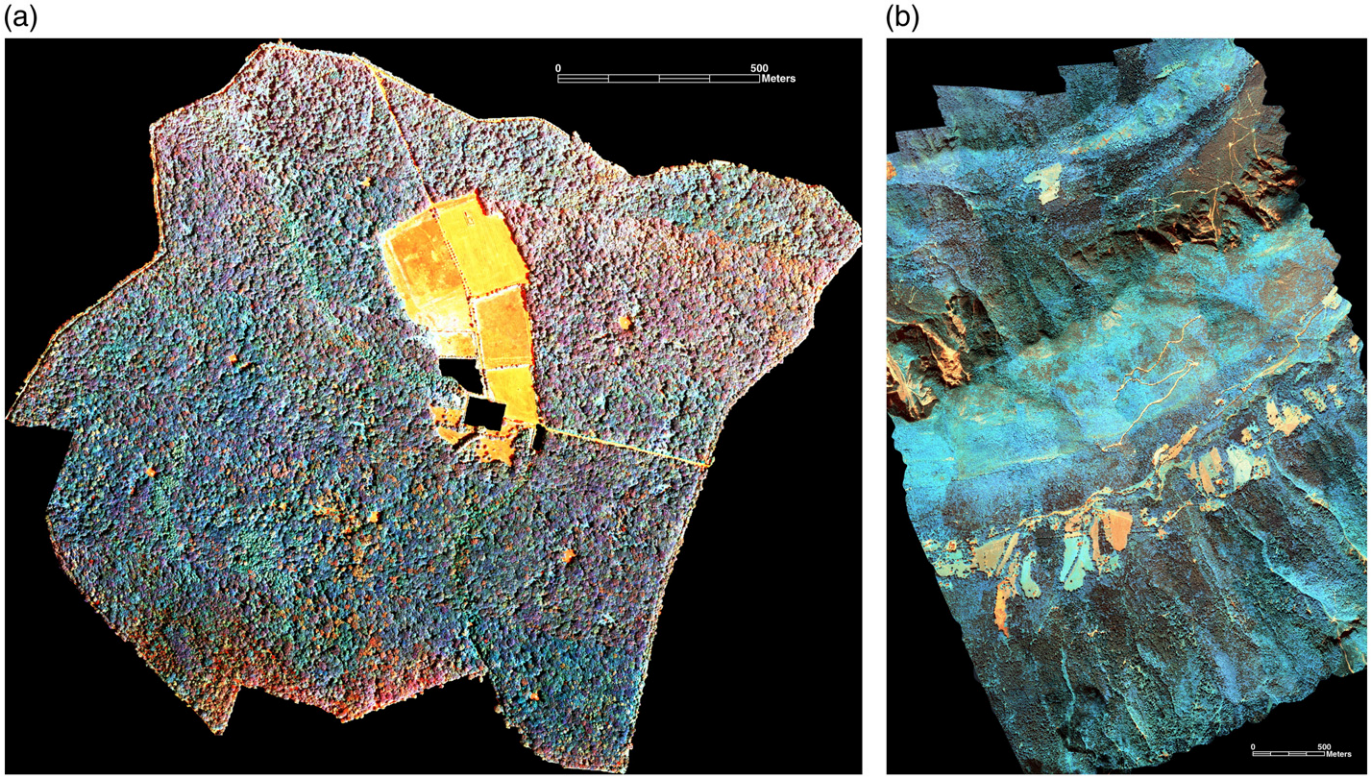

Fig. 1. False color composition (channels 20, 70 and 110) of the hyperspectral image of *Bosco della Fontana* (a) and *Val di Sella* dataset (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

field within an orthophoto (with a geometrical resolution of 0.20 m) of the area analyzed according to ground observations. We extracted these sample points from the entire study area, thus ensuring a precise matching between the ground observations and the aerial ones (e.g. we considered trees near roads, grassland, etc.). The samples were collected on the basis of: i) the species (the reference data was exhaustive, i.e. it represented all the species present in the area; furthermore, it took into account the relative frequency of each class); and ii) the spatial distribution (samples had a uniform distribution across the scene). Starting from all the points collected we draw the Region of Interests (ROIs) of the tree crowns on the mosaicked hyperspectral data, and used them for the generation of the training and test sets. This means that many pixels are extracted from each tree. The total number of reference data samples (16,816 pixels) represented about 0.7% of the whole investigated area.

2.2. Data set 2: Val di Sella

The second data set considered is *Val di Sella*, a forest area in the Italian Alps near the city of Trento. The central point of the area has the following coordinates: 46°0′55.06″ N, 11°25′39.67″ E. This area extends across approximately 1500 ha and its morphology includes both valleys and mountains.

Differently from the first data set, in this case we have only 6 tree species, plus two other additional classes, i.e. shadows and grassland (see Table 3 for a description of the investigated classes). Also in this case the distribution of the species is random and the relative frequency differs among all the species.

The hyperspectral data were acquired on July 2008. They consist of 12 partially overlapping images acquired by an AISA Eagle sensor in 126 spectral bands, ranging from 400 to 990 nm, with a spectral resolution of about 4.6 nm and a spatial resolution of 1 m.

The reference data samples (approximately 190 trees) were collected according to the same strategy used for the previous data set. Starting from all the points collected we draw the Region of Interests (ROIs) on the mosaicked hyperspectral data, and we used them for the generation of the training and test sets. This means that also on this case many pixels are extracted from each tree. The total number of reference data samples (2760 pixels) represents about 0.2% of the whole investigated area.

3. Methods

Before carrying out the analysis of the hyperspectral bands, we applied some pre-processing to the images. First of all, we mosaicked the available images, in order to obtain a single image for each study site. A relative radiometric normalization was applied to the single images in order to obtain a uniform mosaic image. Several algorithms have been proposed in literature to apply these corrections (e.g. Du et al., 2001; Yuan & Elvidge, 1996). In our study, we adopted a simple linear normalization based on the mean-standard deviation normalization algorithm (Yuan & Elvidge, 1996). After that, data were de-noised with a simple low-pass filter. In the literature several studies have pointed out the usefulness of this method (e.g., Dalponte et al., 2008; Hsieh & Landgrebe, 1998). In our case, given the high geometrical resolution of the images, the spatial degradation caused by the filter was acceptable given both the reduction of the noise present in the images and the expected increase in the separability of analyzed classes (Hsieh & Landgrebe, 1998).

Table 3
Distribution of reference data samples (pixels) among investigated classes of the *Val di Sella* dataset (in brackets the number of trees).

Class name	Reference data samples
Abies alba	179 (28)
Acer pseudoplatanus	146 (20)
Alnus incana	76 (7)
Fagus sylvatica	604 (57)
Grassland	1010
Picea abies	314 (42)
Pinus sylvestris	239 (37)
Shadows	192

In our investigation we considered three supervised classification techniques characterized by different levels of complexity. There are different ways to define the level of complexity of a classifier. In this study we consider empirically the level of complexity of a classifier as its ability to define non-linear decision boundaries between the investigated classes. The supervised classifiers considered are: i) a non-linear Support Vector Machines (high complexity); ii) a Gaussian Maximum Likelihood with Leave-One-Out-Covariance estimation (medium complexity); and iii) a Linear Discriminant Analysis (low complexity). In the following we provide greater details on these classifiers and motivate the reasons for these choices.

3.1. Support vector machine classifier

The Support Vector Machine (SVM) (Vapnick, 1998) is an effective distribution free classifier that has been widely used in the recent years for solving hyperspectral classification problems (Camps-Valls & Bruzzone, 2005; Melgani & Bruzzone, 2004). The main reason for the choice of this classifier is associated with its properties that are: i) high generalization ability and high classification accuracies (with respect to others classifiers); ii) convexity of the cost function (which allows one to identify always the optimal solution); iii) effectiveness in addressing ill-posed problems (which are quite common with hyperspectral data); iv) limited effort required for architecture design and training phase if compared to other machine learning algorithms (such as multilayer perceptron neural networks). The main concepts associated with non-linear SVM are briefly described in the following.

Let us consider for simplicity a binary classification problem, characterized by a set of N training samples $\chi=\{\mathbf{x}_n\}_{n=1}^{N}$ (where $\mathbf{x}_n\in\Re^q$). Each pattern is a vector of M features that represents the value that the considered pixel assumes on the considered hyperspectral bands. Let $\psi=\{\mathbf{y}_n\}_{n=1}^{N}$, $\mathbf{y}_n\in\{-1;+1\}$ be the set of related reference labels, where "+1" and "−1" are associated with one of the two classes investigated. The non-linear SVM approach consists of mapping the data into a higher dimensional feature space, i.e., $\Phi(\mathbf{x}_p)\in R^{q'}$ ($q'\gg q$), where the two classes are separated by an hyperplane defined by a weight vector $\boldsymbol{w}\in\Re^{q'}$ (which is orthogonal to the hyperplane) and a bias $b\in\Re$ (which is a scalar value such that the ratio $b/||w||$ represents the distance of the hyperplane from the origin). The function Φ represents a non-linear transformation. The membership decision rule is defined according to $sign[f(x)]$, where $f(x)$ represents the discriminant function associated with the hyperplane and is written as:

$$f(\boldsymbol{x}) = \mathbf{w}\cdot\Phi(\boldsymbol{x}) + b \tag{1}$$

The optimal hyperplane is the one that minimizes a cost function which expresses a combination of two criteria, i.e., margin maximization and error minimization. It is defined as:

$$\Psi(\mathbf{w},\xi) = \frac{1}{2}|\ \mathbf{w}|^{\ 2} + C\sum_{p=1}^{Q}\ \xi_p \tag{2}$$

where the constant C represents a regularization parameter that controls the shape of the discriminant function, and consequently the decision boundary when data are non-separable. This cost function minimization is subject to the following constraints:

$$\begin{cases} \mathbf{y}_p\cdot\left(\mathbf{w}\cdot\mathbf{x}_p + b\right)\geq 1-\xi_p, \forall p=1,\ldots,Q \\ \xi_p\geq 0, \forall p=1,\ldots,Q \end{cases} \tag{3}$$

where ξ_p are the so called slack variables and are introduced to take into account non-separable data (Bruzzone et al., 2006). The above optimization problem can be reformulated through a Lagrange functional as a dual optimization leading to a Quadratic Programming (QP) solution (Vapnick, 1998). The final result is a discriminant function conveniently expressed as a function of the data in the original (lower) dimensional feature space:

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{4}$$

where $K(\cdot,\cdot)$ is a kernel function and S is the subset of training samples corresponding to the nonzero Lagrange multipliers. A kernel function is a function that satisfies the Mercer's theorem (Mercer, 1909) and that makes it possible to avoid a direct explicit representation of the transformation of the feature vectors, i.e. $K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})$.

It is worth noting that the Lagrange multipliers α_i effectively weight each training sample according to its importance in determining the discriminant function. The training samples associated with nonzero weights are termed support vectors. In particular the support vectors where $\alpha_i = C$ are referred to as bound support vector, and support vectors with $0 < \alpha_i < C$ are called non bound support vectors (Bruzzone et al., 2006).

The SVM classifier was developed to solve binary classification problems, but it can be easily generalized to multiclass problems. The two main strategies used for L-class problems are:

- One-Against-One (OAO) — the L-class problem is decomposed into L $(L-1)/2$ binary problems, each focused on the recognition of a pair of classes. A generic pattern is associated with the class that receives the majority of the votes from the ensemble of binary classifiers.
- One-Against-All (OAA) — the L-class problem is decomposed into L binary problems, each focused on the recognition of one class against all the others. The "winner-takes-all" rule is used for the final decision, i.e. the winning class is the one corresponding to the SVM with the highest output (discriminant function value). We refer the reader to (Melgani & Bruzzone, 2004) for greater details on SVM classifiers in remote sensing and on the related multiclass strategies.

3.2. Gaussian Maximum Likelihood classifier with Leave-One-Out-Covariance Estimator (GML-LOOC)

The second classifier that we consider in this study is a Gaussian Maximum Likelihood with Leave-One-Out-Covariance estimator (GML-LOOC) (Hoffbeck & Landgrebe, 1996). This technique is based on the Gaussian Maximum Likelihood (GML) classifier and is suitable for managing hyperdimensional feature spaces. The GML is a parametric classifier based on the Bayesian decision theory. Differently from the SVM, this classifier assumes Gaussian distributions for the class densities. The GML-LOOC approach differs from the standard GML in the phase of estimation of the covariance matrices of the analyzed classes. In fact, when the ratio between the number of training samples for each class and the dimension of the feature space is near one, the standard GML degrades its performances (Hughes phenomenon). In the limit case when the number of training samples is smaller than the number of features, the covariance matrices used in the decision rule become singular, and thus the GML cannot be used. In the literature several algorithms have been developed for the estimation of a non-singular covariance matrix (e.g. Friedman, 1989; Hoffbeck & Landgrebe, 1996; Lin & Perlman, 1985; Marks & Dunn, 1974; Wahl & Kronmall, 1977). In our study, we chose the algorithm proposed in (Hoffbeck & Landgrebe, 1996), which is called Leave-One-Out-Covariance (LOOC) algorithm. In the following we give some more details on this classifier.

Let $\mathbf{x}_n$ be the n-th pattern to be classified, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ (with $i = 1,\ldots, L$) the mean value and the covariance matrix of the i-th investigated class, respectively, and $\Omega = \{\omega_1, \omega_2, \ldots \omega_L\}$ the set of the L land-cover classes in the considered classification problem. The decision rule is as follows:

$$\mathbf{x}_n \in \omega_j \Leftrightarrow d_j(\mathbf{x}_n) > d_i(\mathbf{x}_n) \quad \forall\, i \neq j \tag{5}$$

where $d_i(\mathbf{x}_n)$ is computed as:

$$d_i(\mathbf{x}_n) = (\mathbf{x}_n - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_i) + \ln |\boldsymbol{\Sigma}_i| \tag{6}$$

Usually the true values of the mean vectors and of the covariance matrices are not known and they should be estimated from the training samples. When a reduced number of samples is available, the covariance matrices can be replaced with the common covariance matrix, defined as: $\boldsymbol{S} = \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{\Sigma}_i$ (Hoffbeck & Landgrebe, 1996). The LOOC algorithm proposes a more refined way to estimate the covariance matrices for classes characterized by a reduced number of training samples. In particular the covariance matrix Σ_i^{LOOC} of the i-th class is estimated as follows:

$$\Sigma_i^{\mathrm{LOOC}}(\alpha_i) = \begin{cases} (1-\upsilon_i)\mathrm{diag}(\boldsymbol{\Sigma}_i) + \upsilon_i \boldsymbol{\Sigma}_i & 0 \le \upsilon_i \le 1 \\ (2-\upsilon_i)\boldsymbol{\Sigma}_i + (\upsilon_i - 1)\boldsymbol{S} & 1 < \upsilon_i \le 2 \\ (3-\upsilon_i)\boldsymbol{S} + (\upsilon_i - 2)\mathrm{diag}(\boldsymbol{S}) & 2 < \upsilon_i \le 3 \end{cases} \tag{7}$$

where υ_i is a mixing parameter. The value of this parameter is selected according to the following procedure: i) removing one sample form the training set, ii) computing the mean and covariance from the remaining samples, iii) computing the likelihood of the sample which was left out, given the mean and covariance estimates. Each sample is removed in turn, and the average log likelihood is computed. The value that maximizes the average log likelihood is selected (Biehl, 2001). This implementation proved to be particularly effective in hyperspectral data classification.

In our experiments we used this classifier under the unimodal Gaussian assumption for the distribution of information classes. This assumption is widely used in the literature, even if a more complex and accurate approach based on the decomposition of each information class in a set of unimodal Gaussian data classes could be used. This could be done by applying clustering to the training samples of each class. However, when a high number of information classes is present in the classification problem, this process results time consuming (also because an adequate number of clusters for each class should be identified). In addition, when few training samples for each class are available, this may involve a high risk to overfit the training set in the modeling of the multimodal class distributions. This can be particularly critical when hyperspectral images are considered, where a significant spatial variability of the spectral signature of each class in the image is usually present (Chi & Bruzzone, 2007).

3.3. Linear discriminant analysis classifier

The last technique that we consider is a very simple linear discriminant analysis (LDA) classifier (Duda et al., 2000; Fisher, 1936). The rationale of this classifier can be considered as the opposite of that at the basis of the SVM classifier. LDA projects high dimensional feature spaces into a low-dimensional space, with the target to keep information classes as more separated as possible. This transformation is obtained by minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum discrimination. Given its simplicity, this classifier is less suitable to the analysis of hyperspectral data with respect to the previous ones, even if some studies exist on the application of LDA techniques to hyperspectral data (Bandos et al., 2009). In the following we recall the main concepts associated with LDA. We refer the reader to (Bandos et al., 2009) for more details.

Let us consider L classes classification problem. The idea of the classical LDA classifier is to find a linear transformation $\mathbf{G}$ that project

the sample $\mathbf{x}_n$ from the original m-dimensional feature space to a lower dimensional space a according to the following equation:

$$\mathbf{a} = \mathbf{G}^T\mathbf{x} \in R^l \tag{8}$$

where $l<m$. The goal of this transformation is to choose the direction of $\mathbf{v}$ in the feature space along which the distances of the class means are at maximum and the variances around these means are at minimum. This corresponds to maximize the following criterion:

$$\mathbf{v}^* = \arg\max_{\mathbf{v}}\{J(\mathbf{v})\} = \arg\max_{\mathbf{v}}\left\{\frac{\mathbf{v}^T\mathbf{S}_b\mathbf{v}}{\mathbf{v}^T\mathbf{S}_w\mathbf{v}}\right\} \tag{9}$$

where $S_b = \frac{1}{n}\sum_{i=1}^{L} n_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$ is the between-class variance, $S_w = \frac{1}{n}\sum_{k=1}^{L}\sum_{i\in A_k} n_i(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^T$ is the within-class variance, $\boldsymbol{\mu}_i$ is the sample mean, and A_i denotes the index set for class i. As the total scatter matrix (which is the estimate of the common covariance matrix) can be written as $\mathbf{S} = \mathbf{S}_w + \mathbf{S}_b$, the maximization criterion becomes:

$$\mathbf{v}^* = \arg\max_{\mathbf{v}}\{J(\mathbf{v})\} = \arg\max_{\mathbf{v}}\left\{\frac{\mathbf{v}^T\mathbf{S}_b\mathbf{v}}{\mathbf{v}^T\mathbf{S}_w\mathbf{v}} - 1\right\} \tag{10}$$

In this case the optimization problem maximizes the total scatter of the data while minimizing the within scatter of the classes. The criterion can be rewritten as follows:

$$\mathbf{G}^* = \arg\max_{\mathbf{G}}\left\{\mathrm{trace}\left[\left(\mathbf{G}^T\mathbf{S}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{S}_b\mathbf{G}\right]\right\} \tag{11}$$

The solution can be obtained by applying an eigen-decomposition to the matrix $\mathbf{S}^{-1}\mathbf{S}_b$, if $\mathbf{S}$ is non-singular. Note that there exist no more than $k-1$ eigenvectors corresponding to nonzero eigenvalues, since the rank of the matrix $\mathbf{S}_b$ is bounded by $k-1$. Therefore, the reduced dimension of classical LDA is at most $k-1$ (Ye et al., 2006).

In this paper we have considered the standard LDA algorithm. However, in presence of very small numbers of training samples, it is possible to use some regularization algorithms to avoid singularity or poor estimations of the within and between scattering matrices. For a detailed description of such algorithms we refer the reader to (Bandos et al., 2009).

3.4. Design of experiments

In order to achieve the goals of this paper, we defined two different kinds of experiments: i) analysis of the effects of the spectral resolution on the classification accuracy; ii) analysis of the effects of the number of spectral channels selected with a feature selection algorithm (applied to the original bands at full resolution) on the classification accuracy. For both the experiments we carried out the training of all the considered classifiers (including the model selection) and the accuracy assessment according to a fivefold cross-validation procedure. This allowed us to conduct the analysis from a rigorous statistical perspective. We randomly divided the available ground-truth data into five subsets, and then we adopted a fivefold cross-validation procedure, with training samples distributed across the scene. The samples (pixels) of the reference data available were used as follows: 20% in the training set and 80% in the test set. It is worth noting that the goal of this paper was not to analyze the generalization ability of the classifiers, but to assess their role in managing hyperdimensional feature spaces. Thus the choice to use a cross-validation procedure appears to be the most suitable one for a correct statistical analysis of the problem in hand. We used the same cross-validation subsets for all the classifiers analyzed.

The SVM classifier used was based on our own implementation. We selected Gaussian RBF kernel functions and applied a grid search strategy in a range between 5 and 240 for C, and in a range between 1 and 1000 for γ. The multiclass architecture adopted was based on the One-Against-One multiclass strategy. With regards to the GML-LOOC classifier we used the MultiSpec software (Biehl, 2001), while for the LDA we used the implementation contained in the MATLABArsenal software (Yan, 2006).

4. Experimental results

4.1. Experiment 1: analysis of the role of spectral resolution on classification accuracy

The first experiment focused on the analysis of the role of the spectral resolution on the classification accuracy by varying the classifier adopted. To develop this analysis, we simulated data with different spectral resolutions averaging contiguous spectral bands of the acquired image. Specifically, we degraded the resolution from 4.6 nm to 36.8 nm, using a step of 4.6 nm. It is worth noting that to obtain a precise simulation of the reduction of the spectral resolution, it would be necessary to consider the frequency response of the spectral filter associated with each channel. However, for the purpose of our analysis, it was reasonable to approximate the frequency response as constant for all the channels and to use an average operator for approximating the reduction of the spectral resolution. Fig. 2 shows the behavior of the kappa accuracies obtained with different spectral resolutions for each of the classifiers used on the two datasets considered. From an analysis of the figure, it is possible to derive some inferences of the effect of changing spectral resolution upon the different classifiers. First of all, the SVM classifier obtained higher accuracies than all the other classifiers for all the spectral resolutions considered and in both the datasets. The difference in accuracy between SVM and the other classifiers was higher in the *Bosco della Fontana* dataset where we have a very high number of classes. This result underlines the effectiveness of the SVM classifier in managing complex hyperspectral classification problems. LDA was not able to model the complexity of the problem assessed with the *Bosco della Fontana* dataset. This is mainly due to the oversimplification obtained by projecting the high dimensional feature space in a low-dimensional space. Concerning the *Val di Sella* dataset there is a small difference between the accuracy provided by the SVM and the GML-LOOC, and also LDA resulted in reasonable accuracies (at the maximum spectral resolution the accuracies are very similar for all the classifiers). This depends on the simplicity of the second problem which is characterized by a small number of classes.

Secondly, it is interesting to analyze the behavior of different classifiers to the degrading of the spectral resolution. Concerning the *Bosco della Fontana* dataset, both the SVM and LDA classifiers reduced their accuracy as the spectral resolution of the sensor was reduced. In particular the LDA classifier was strongly influenced by the spectral resolution. It reduced noticeably its accuracy as the spectral resolution decreased. Also the SVM classifier decreased its accuracy as the spectral resolution was reduced (approximately 1% from 4.6 nm to 9.2 nm, and 5% from 9.2 nm to 36.8 nm). Despite this, the lowest accuracy of the SVM classifier was still higher than the highest accuracy of the other classifiers considered. The GML-LOOC presented the most stable accuracy and in particular it did not result in significant differences between the kappa accuracies obtained with a resolution in the range between 4.6 and 23 nm. On the contrary, it exhibited a slight increase in accuracy between 9.2 and 13.8 nm. Regarding the *Val di Sella* dataset, given the simplicity of the problem, the behavior of the three classifiers was very similar. Also in this case the LDA degraded its accuracy reducing the spectral resolution, even if the degradation is limited with respect to the previous dataset. On this dataset the performance remains acceptable also with a spectral resolution of 36.8 nm. SVM and GML-LOOC provided very similar results, exhibiting a kappa accuracy always higher than 95% for all the spectral resolution considered. It is worth

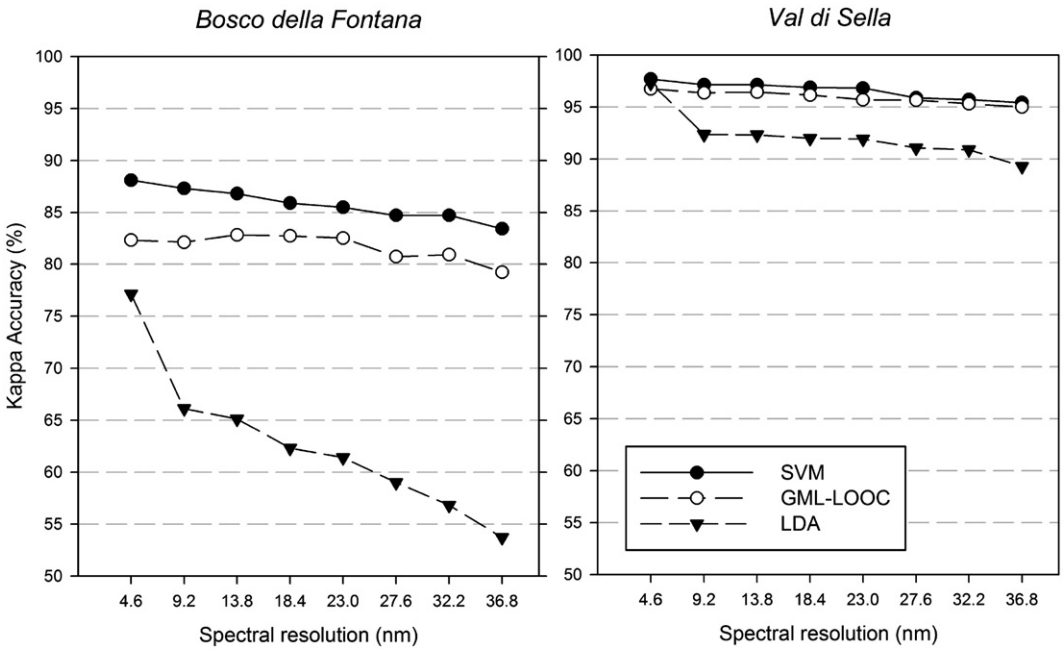

Fig. 2. Behavior of the kappa accuracy of the analyzed classifiers versus the spectral resolution for the two datasets considered.

noting that in this case it seems that also a low spectral resolution is enough to separate the considered forest classes.

Fig. 3 shows the behavior of the class producer accuracies versus the spectral resolution obtained by the three classifiers analyzed on the *Bosco della Fontana* dataset. Firstly, it is worth noting that the LDA classifier always provided the lower accuracies and it reduced its performances by reducing the spectral resolution, confirming the behavior of the overall kappa accuracy. Concerning SVM and GML-LOOC the behavior is quite different on the different classes analyzed. In general, SVM provided the highest accuracy on the majority and most relevant classes, thus confirming the results obtained in terms of kappa accuracy. Nevertheless, as expected, some classes exhibited higher accuracy on the maps produced by the GML-LOOC classifier. This is intrinsic in the solution of a multiclass problem, where different classifiers obtain different accuracies on many different classes. Thus, the overall accuracy remains the most important performance for a general estimation of the results in our study.

4.2. Experiment 2: effect of the number of spectral channels on the classification accuracies obtained by different classifiers using the highest spectral resolution

In this second experiment, we analyzed the effect of the number of spectral channels on the classification accuracies obtained by different classifiers, keeping the original spectral resolution of the sensor (in this case 4.6 nm). In particular with this experiment we wanted to determine: i) if all the bands at the highest spectral resolution were significant, and to examine the behavior of the different classifiers with respect to their selection; ii) if, given a fixed number of bands, the selection of channels at the highest resolution is more effective than the acquisition of bands at a lower resolution; and iii) the physical meaning of the bands selected by the feature-selection algorithm on our test areas. To achieve these goals, we applied a feature selection algorithm based on the Sequential Forward Floating Selection search strategy (Pudil et al., 1994) and on the Jeffreys–Matusita (JM) distance (Bruzzone et al., 1995) to the original image. The JM distance was adopted as it is correlated with the Chernoff upper bound to the error probability of the Bayesian classifier. This means that the feature-selection process adopted is nearly optimum for the GML-LOOC classifier. Concerning the SVM classifier, in the literature it is possible to find few methods for feature selection which are especially developed for such a classifier; however, in this study we preferred to use for all the three classifiers the same feature-selection algorithm (and thus the same set of features). This is reasonable at an operational level as confirmed from many studies published in the literature that combine such an algorithm with different kinds of classifiers (including the SVM). It is worth noting that we did not consider other feature selection algorithms as we aim at analyzing the behavior of the classification techniques considered versus the number of spectral channels at the maximum resolution, and of comparing such results with those obtained in the first experiment. Thus, it is reasonable to consider just one reference feature-selection algorithm rather than exploring results obtained by different methods.

In this analysis we applied the feature selection so as to identify eight sets of bands made up of the same number of features that we obtained in the previous experiment by reducing the spectral resolution. This allowed us to make some further considerations comparing the results of the two experiments.

Fig. 4 shows the kappa accuracies versus the number of selected features in the two datasets considered obtained with the three different classifiers. From these results we firstly note that none of the classifiers under investigation was significantly affected by the Hughes phenomenon. This was due to both the intrinsic robustness of these classifiers to ill-posed problems and the relatively high number of reference data samples available. Secondly, we observed that the SVM always obtained the highest accuracy with respect to the other classifiers. In particular, the difference in accuracy between using 16 or 126 spectral channels was less than 2% in both the datasets considered. This is a point that we would like to stress as it underlines the high discrimination ability of high spectral resolution hyperspectral data. Moreover, these results underline that with a high complexity classifier, like the SVM, it is possible to work with a subset of hyperspectral bands, thereby reducing the computational costs but not the classification accuracy. Additionally, the results confirm the robustness of the SVM classifier to

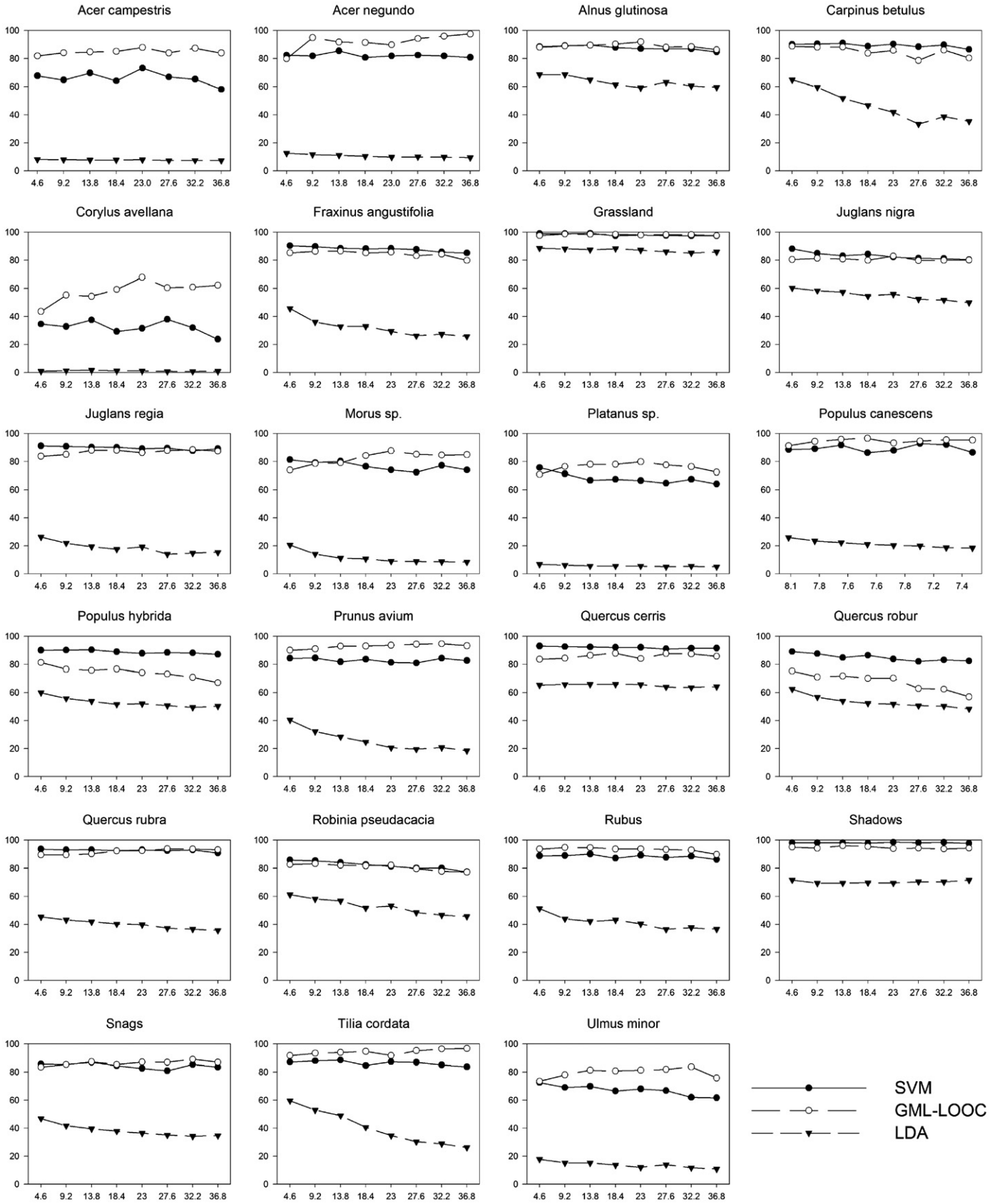

Fig. 3. Behavior of the Producer Accuracy versus the spectral resolution for the classes analyzed in *Bosco della Fontana* dataset.

hyperdimensional feature spaces. Also the LDA classifier seems to take advantage of using features at the highest spectral resolution. Nevertheless, this classifier produced the lowest accuracies, but its kappa coefficient increased in comparison to the previous experiment. For the GML-LOOC classifier the behavior was quite different as it in general provided lower kappa accuracies with respect to the previous

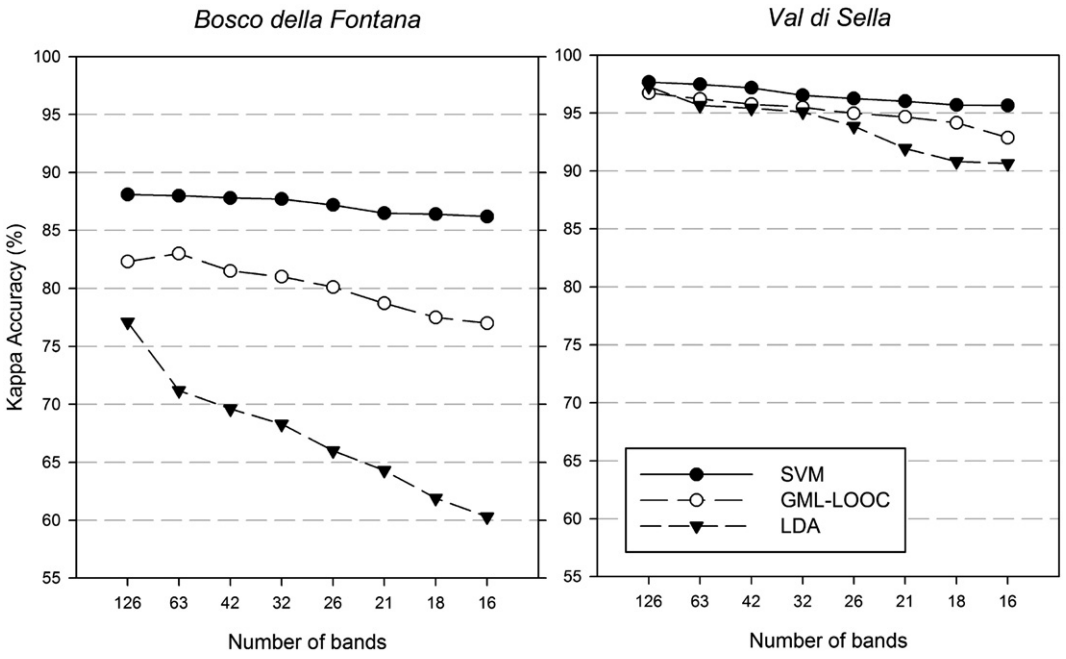

Fig. 4. Behavior of the kappa accuracy provided by the analyzed classifiers versus the number of spectral bands at a spectral resolution of 4.6 nm for the two dataset considered.

experiment. In addition, this classifier increased its accuracy when the number of spectral channels was increased.

4.3. Analysis of results and discussion

Comparing the results of all the experiments carried out on the two considered datasets it is possible to draw some interesting conclusions on the relationship among accuracy, classifier complexity and spectral resolution. From an analysis of Fig. 2, it is clear that the different classifiers have different behaviors with respect to the spectral resolution. This underlines the complexity and the importance of our study.

First of all, let us consider the behavior of the SVM classifier in the two experiments. The ability of this classifier in managing hyperspectral feature spaces and its robustness to noisy pixels is well known in the literature (e.g., Bruzzone et al., 2006; Melgani & Bruzzone, 2004). The analysis of the overall kappa accuracies confirms these characteristics: SVM classifier provided the highest overall kappa accuracy for all the spectral resolutions and it was not significantly affected by the Hughes phenomenon. Moreover, comparing the results of the two experiments, it seems that for the SVM it was better to apply a feature selection to the original spectral bands, rather than reducing the spectral resolution (and thus increasing the SNR of each spectral channel). This was dependent upon on the effectiveness of the SVM to define effective non-linear discrimination function in the original feature space starting from high information content data like the original channels rather than from those with reduced spectral resolution. Such a capability is due to two main reasons: i) the potentially high complexity of the decision boundary associated with the SVM classifier; ii) high robustness of the SVM classifier to the outliers, and thus to the lowest SNR present in the original spectral channels.

Concerning the LDA classifier, it decreased its accuracy reducing both the number of original spectral channels considered and the spectral resolution of the sensor. This behavior can be explained by the intrinsic properties of LDA; this algorithm applies a transformation of the original feature space into a space with a lower dimensionality, by maximizing classes' separability. It is reasonable to expect that LDA performs better this transformation when more discriminant information (higher number of informative spectral channels) is available. Moreover the reduced performances of this classifier in all the experiments considered could also be due to the quality of estimations of the within and between scattering matrices in presence of high dimensional feature spaces and relatively few training samples. Possible improvements might be obtained by using regularized method in the matrix estimation process (Bandos et al., 2009).

The GML-LOOC classifier has a different behavior. From our results it was possible to note that this classifier exhibited a higher accuracy if the feature reduction was carried out by decreasing the spectral resolution of the sensor rather than selecting original channels according to a feature-selection algorithm. As observed in Experiment 1, it provided almost the same accuracies in a range of spectral resolutions from 4.6 to 23 nm. This behavior can be explained as follows: i) by decreasing the spectral resolution we increased the SNR of the signal acquired in each channel by introducing a low-pass spectral filtering that reduces the noise in the spectral domain; ii) the Gaussian assumption of the GML-LOOC and the regularization method adopted resulted in relatively simple quadratic decision boundaries that cannot seize the complexity of the problem modeled with the original spectral channels. In other words, as shown with the SVM classifier, the original spectral channels at the highest resolution contain the maximum amount of information for discriminating classes, but the GML-LOOC classifier cannot effectively exploit these data. To illustrate this point, by comparing the results of the SVM and the GML-LOOC classifier obtained by using 32 bands at 2 different spectral resolutions (4.6 and 18.4 nm) for the *Bosco della Fontana* dataset, we observe completely different behaviors: the SVM provided the highest accuracy at the highest resolution considered (4.6 nm), while GML-LOOC yielded the highest accuracy at 18.4 nm. It is worth noting that the accuracy of the GML-LOOC classifier could be increased by applying a more detailed modeling of the distributions of the information classes through a decomposition in unimodal Gaussian data classes. However this process would increase significantly the complexity of the design of the classifier (especially when many information classes are considered) and would result in the risk of overfitting the training data when few training samples are available.

In order to better understand the effectiveness of the SVM classifier at the highest resolution in relation to the specific considered forest problem, it is also important to analyze the physical meaning of the selected features. Fig. 5 shows the distribution on the spectrum of 32

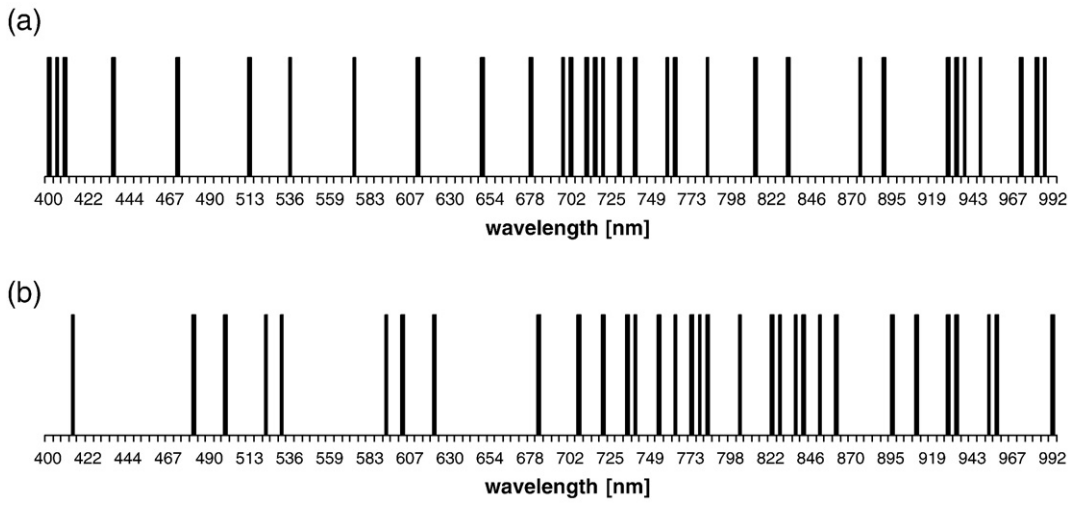

Fig. 5. Spectral distribution of 32 hyperspectral bands selected by the feature selection algorithm for the *Bosco della fontana* dataset (a) and the *Val di Sella* dataset (b).

spectral bands selected at the spectral resolution of 4.6 for the two datasets. All the main regions of the spectrum analyzed by the sensor have an important role in species classification. In the visible range 11 bands were selected for the *Bosco della Fontana* dataset and 7 for the *Val di Sella* dataset; specifically, five and three bands were chosen in the blue range (~400 to ~500 nm), characterized mainly by carotenoides absorption peaks (Zur et al., 2000), but also by chlorophyll *a* with a maximum absorption peak around 430 nm (a band at 435 nm was selected for the *Bosco della Fontana* dataset). In the green (~500 to ~600 nm) and red spectra (~600 to ~650 nm) 5 bands were selected. Chlorophyll has a reflectivity peak in the green area that gives the green color to the vegetation, and the reflectance is strongly linked to chlorophyll content (Gitelson et al., 1996), especially around 550 nm (Gitelson & Merzlyak, 1994). Bands around 531 and 570 nm (two bands 535 and 573 nm were selected in our trials) were used by Gamon et al. (1992, 1997) for PRI index calculation to estimate rapid changes in the relative levels of xanthophyll cycle pigments and thus serves as an estimate of photosynthetic light use efficiency. Neighbouring bands in the green region (529 and 564 nm) were proposed by Darvishzadeh et al. (2008) for leaf chlorophyll content measurements. The red spectra region is well known for chlorophyll peaks absorption (chlorophyll *b*, with a maximum absorption of ~642 nm and a band at 649 nm was selected).

As described by Ceccato et al. (2002) these first regions of the spectra are primarily influenced by the pigment content and secondly by the internal structure parameters. This aspect is more important in the red edge region, where 8 and 6 bands were localized for the *Bosco della Fontana* dataset and the *Val di Sella* dataset, respectively. This region is between ~680 nm (the main red absorption peak of chlorophylls, Zur et al., 2000; Sims & Gamon, 2002) and ~750 nm and ranges between the absorption region of the visible and the reflective region of the near infrared. Its position and behaviour is affected by many factors including changes of chlorophyll content, leaf area index, biomass and hydric status, vegetation age, plant health levels, and seasonal patterns. The interesting issue, from a classification viewpoint, is that the exact wavelength and strength of the red edge depends upon the species considered, and thus bands in this region are important for classification.

In the near infrared region (from ~750 to ~1000 nm) 12 and 18 bands were selected for *Bosco della Fontana* and *Val di Sella* dataset, respectively. For deciduous species (as found in our study site) there is a strong reflectance in this range (Gates et al., 1965). This is due to chlorophyll pigments that are very absorptive at visible wavelengths but are not at all absorptive at near infrared wavelengths (Knipling, 1970). These are linked with others parameters, such as leaf structure (that makes light scattering highly efficient), Leaf Area Index (LAI), (Thenkabail et al., 2004) and the presence of water in the leaf (for example in the band around 970 nm, Peñuelas et al., 1993). Greater transmittance occurs when water is more prevalent between the plant cells of leaves and more reflectance occurs when the spaces between cells are more filled with air. Additionally, the water content can be also linked to the last three bands selected, which are usually used to compute the floating-position Water Bands Index (Strachan et al., 2002).

5. Conclusions

In this paper an experimental analysis on the relationship among the spectral resolution, the classifier complexity and the classification accuracy has been presented. This analysis has focused on two complex forest areas characterized by different numbers of classes, and can be subdivided into three parts: i) analysis on the role of the spectral resolution on the classification accuracy versus the classifier complexity; ii) analysis of the effects on the classification accuracy of the number of spectral bands (given a fixed spectral resolution) versus the classifier complexity; iii) analysis of the effects on the classification accuracy of both reducing spectral resolution and selecting features at the highest resolution given a fixed number of channels as input to the classifier.

The experimental analysis resulted in interesting conclusions on the relationship among the aforementioned factors. In particular, our analysis points out that the option to acquire images at a certain spectral resolution should be driven not only from the considered problem, but also from the classifier used for the data processing. Not all the classifiers were able to exploit the wide amount of data provided by hyperspectral sensors, and not all the classifiers have the same behavior reducing the spectral resolution.

In greater detail, we verified on the considered data set the following issues:

i LDA (a simple classifier) even if obtained the highest accuracy with a high spectral resolution, does not achieve acceptable classification accuracies in complex forest classification problems with a high number of classes;

ii GML-LOOC (medium complexity classifier) provided high classification accuracies in all the considered experimental setups. Moreover, from our analysis, it seemed that it does not take advantage from increasing the spectral resolution over a given value (about 23 nm on the considered data sets);

iii SVM (complex classifier) fully exploited the discrimination ability of channels with very high spectral resolution. In our experiments SVM provided always the highest accuracies among the considered

classifiers. In addition it exhibits the best performances with the maximum spectral resolution (4.6 nm).

As a final remark it is important to observe that the proposed analysis provides important hints on the sensor and data analysis setup to use for classification of complex forest areas, as it supplies interesting indications on the trade-off between the spectral resolution and the classifier complexity in the study of such kinds of environments. It is worth noting that this research does not want to present an exhaustive analysis of the problem, but it should be considered as a starting point for future analysis on different areas (also in relation to applications different from forestry) and with different classifiers.

Acknowledgements

The authors would like to thank the two anonymous Reviewers and the Editor for their helpful comments. This work was supported by the CARBOITALY project funded by the FISR program of the Italian Ministry of University and Research.

References

Backer, B. L., Lusch, D. P., & Qi, J. (2007). A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery. *Remote Sensing of Environment*, *108*, 111−120.

Bandos, T. V., Bruzzone, L., & Camps-Valls, G. (2009). Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), 862−873.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by probability distributions. *Bulletin of the Calcutta Mathematical Society*, *35*, 99−109.

Biehl, L. (2001). *An introduction to multispec.*: School of Electrical and Computer Engineering Purdue University (http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/).

Bruzzone, L., Chi, M., & Marconcini, M. (2006). A novel transductive SVM for the semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(11), 3363−3373.

Bruzzone, L., Roli, F., & Serpico, S. B. (1995). An extension to multiclass cases of the Jeffreys–Matusita distance. *IEEE Transactions on Geoscience and Remote Sensing*, *33*, 1318−1321.

Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(6), 1351−1362.

Ceccato, P., Gobton, N., Flasse, S., Pinty, B., & Tarantola, S. (2002). Designing a spectra index to estimate vegetation water content from remote sensing data: Part 1 Theoretical approach. *Remote Sensing of Environment*, *82*, 188−197.

Chi, M., & Bruzzone, L. (2007). Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(6), 1870−1880.

Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, *96*, 375−398.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(5), 1416−1427.

Darvishzadeh, R., Skidmore, A. K., Schlfer, M., Atzberger, C., Corsi, F., & Cho, M. A. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *International Journal of Photogrammetry and Remote Sensing*, *63*(4), 409−426.

Djouadi, A., Snorrason, O., & Garber, F. (1990). "The quality of Training-Sample estimates of the Bhattacharyya coefficient". *IEEE Transactions on Pattern analysis and machine intelligence*, *12*, 92−97.

Duda, R. O., Hart, P. E., & Stork, D. H. (2000). *Pattern classification*, (2nd ed.) : Wiley Interscience.

Du, Y., Cihlar, J., Beaubien, J., & Latifovic, R. (2001). Radiometric normalization, compositing, and quality control for satellite high resolution image mosaics over large areas. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(3), 623−634.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179−188.

Friedman, J. H. (1989). Regularized Discriminant Analysis. *Journal of the American Statistical Association*, *84*, 165−175.

Gamon, J. A., Peñuelas, J., & Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, *41*, 35−44.

Gamon, J. A., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: An optical indicator of photosynthetic radiation use efficiency across species, functional types and nutrient levels. *Oecologia*, *112*, 492−501.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). *Spectral properties of plantsApplied Optics*, *4*, 11−20.

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*, 289−298.

Gitelson, A. A., & Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* and *Acer platanoides* leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, *143*, 286−292.

Hoffbeck, J. P., & Landgrebe, D. A. (1996). Covariance matrix estimation and classification with limited training data. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, *18*(7), 763−767.

Hsieh, P., & Landgrebe, D. (1998). Low-pass filter for increasing class separability. *Proceedings of the IEEE Geoscience and Remote Sensing Symposium 1998*, *5*, 2691−2693.

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

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, *1*, 155−159.

Lin, S. P., & Perlman, M. D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. *Proceedings of the Sixth International Symposium on Multivariate Analysis*, 411−429.

Marks, S., & Dunn, O. J. (1974). Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association*, *69*(346), 555−559.

Martin, M. E., Newman, S. D., Aber, J. D., & Congalton, R. G. (1998). Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment*, *65*, 249−254.

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(8), 1778−1790.

Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London*, *209*, 415−446.

Peñuelas, J., Filella, I., & Sweeano, L. (1993). The reflectance at the 970/900 region as an indicator of plant water status. *International Journal of Remote Sensing*, *14*, 1887−1905.

Pudil, P., Novovicova, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, *15*, 1119−1125.

Richards, J. A., & Jia, X. (1999). *Remote sensing digital image analysis.* : Springer-Verlag.

Schmidt, K. S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote sensing of environment*, *85*(1), 92−108.

Sedano, F., Gong, P., & Ferrao, M. (2005). Land cover assessment with MODIS imagery in southern African Miombo ecosystems. *Remote Sensing of Environment*, *98*, 429−441.

Serpico, S. B., & Bruzzone, L. (2001). A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(7), 1360−1367.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*, 337−354.

Strachan, I. B., Pattey, E., & Boisvert, J. B. (2002). Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sensing of Environment*, *80*, 213−224.

Thenkabail, P., Enclona, S., Ashton, E. A., Legg, M. S., & De Dieu, M. J. (2004). Hyperion, IKONOS, ALO, and ETM+ sensors in the study of African rainforests. *Remote Sensing of Environment*, *90*(1), 23−43.

Vapnick, V. N. (1998). *Statistical learning theory.* : John Wiley and Sons Inc.

Wahl, P. W., & Kronmall, R. A. (1977). Discriminant Functions when Covariances are Equal and Sample Sizes are Moderate. *Biometrics*, *33*, 479−484.

Wang, L., Sousa, W. P., Gong, P., & Biging, G. S. (2004). Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. *Remote Sensing of Environment*, *91*, 432−440.

Yan, R. (2006). *MATLABArsenal A MATLAB package for classification algorithms.* http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/MATLABArsenal.htm

Ye, J., Xiong, T., Li, Q., Janardan, R., Bi, J., Cherkassky, V., & Kambhamettu, C. (2006). Efficient Model Selection for Regularized Linear Discriminant Analysis. *Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA.*, 532−539.

Yuan, D., & Elvidge, C. D. (1996). Comparison of relative radiometric normalization techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, *51*, 117−126.

Zur, Y., Gitelson, A. A., Chivkunova, O. B., & Merzlyak, M. N. (2000). The spectral contribution of carotenoids to light absorption and reflectance in green leaves. *Second international conference on geospatial in agriculture and forestry, Lake Buena, Vista, 10–12 February, 2000*.