



**International Doctorate School in Information and
Communication Technologies**

DISI - University of Trento

**ANALYSIS OF FOREST AREAS BY ADVANCED REMOTE SENSING
SYSTEMS BASED ON HYPERSPECTRAL AND LIDAR DATA**

Michele Dalponte

Advisor:

Prof. Lorenzo Bruzzone

Università degli Studi di Trento

Co-Advisor:

Dott. Damiano Gianelle

Fondazione Edmund Mach

March 2010

Abstract

Forest management is an important and complex process, which has significant implications on the environment (e.g. protection of biological diversity, climate mitigation) and the economy (e.g. estimation of timber volume for commercial usage). An efficient management requires a very detailed knowledge of forest attributes such as species composition, trees stem volume, height, etc. Hyperspectral and LIDAR remote sensing data can provide useful information to the identification of these attributes: hyperspectral data with their dense sampling of the spectral signatures are important for the classification of tree species, while LIDAR data are important for the study and estimation of quantitative parameters of forests (e.g. stem height, volume).

This thesis presents novel systems for the exploitation of hyperspectral and LIDAR data in forest application domain. In particular, the novel contributions to the existing literature are on both the development of new systems for data processing and the analysis of the potentialities of these data in forestry. In greater detail the main contribution of this thesis are: i) an empirical analysis on the relationship between spectral resolution, classifier complexity and classification accuracy in the study of complex forest areas. This analysis is very important for the design of future sensors and the better exploitation of the existing ones; ii) a novel system for the fusion of hyperspectral and LIDAR remote sensing data in the classification of forest areas. The system proposed exploits the complementary information of these data in order to obtain accurate and precise classification maps; iii) an analysis on the usefulness of different LIDAR returns and channels (elevation and intensity) in the classification of forest areas; iv) an empirical analysis on the use of multireturn LIDAR data for the estimation of tree stem volume. This study investigates in detail the potentialities of variables extracted from LIDAR returns (up to four) for the estimation of tree stem volume; v) a novel system for the estimation of single tree stem diameter and volume with multireturn LIDAR data. A comparative analysis on the use of three different variable selection methods and three different estimation algorithms is also presented; vi) a system for the fusion of hyperspectral and LIDAR remote sensing data in the estimation of tree stem diameters. This system is able to exploit hyperspectral and LIDAR data combined and separated: this is very important as the experimental analysis carried out with this system shows that hyperspectral data can be used for rough estimations of stem diameters when LIDAR data are not available.

The effectiveness of all the proposed systems is confirmed by quantitative and qualitative experimental results.

Keywords:

hyperspectral images, LIDAR, classification, estimation, forestry, stem volume, stem diameter, remote sensing.

Contents

1	INTRODUCTION AND THESIS OVERVIEW.....	3
1.1	INTRODUCTION TO REMOTE SENSING.....	3
1.2	FOREST INVENTORIES.....	4
1.3	REMOTE SENSING AND FORESTRY.....	5
1.4	OBJECTIVES OF THE THESIS.....	8
1.5	STRUCTURE OF THE THESIS.....	11
1.6	REFERENCES.....	11
2	THE ROLE OF SPECTRAL RESOLUTION AND CLASSIFIER COMPLEXITY IN THE STUDY OF COMPLEX FOREST AREAS.....	13
2.1	INTRODUCTION.....	13
2.2	DATA SETS DESCRIPTION.....	16
2.2.1	<i>Data Set 1: Bosco della Fontana.....</i>	<i>16</i>
2.2.2	<i>Data Set 2: Val di Sella.....</i>	<i>18</i>
2.3	METHODS.....	19
2.3.1	<i>Support Vector Machine classifier.....</i>	<i>19</i>
2.3.2	<i>Gaussian Maximum Likelihood classifier with Leave-One-Out-Covariance Estimator (GML-LOOC) 21</i>	
2.3.3	<i>Linear Discriminant Analysis classifier.....</i>	<i>22</i>
2.3.4	<i>Design of experiments.....</i>	<i>23</i>
2.4	EXPERIMENTAL RESULTS.....	24
2.4.1	<i>Experiment 1: analysis of the role of spectral resolution on classification accuracy.....</i>	<i>24</i>
2.4.2	<i>Experiment 2: effect of the number of spectral channels on the classification accuracies obtained by different classifiers using the highest spectral resolution.....</i>	<i>26</i>
2.4.3	<i>Analysis of results and discussion.....</i>	<i>29</i>
2.5	CONCLUSIONS.....	31
2.6	REFERENCES.....	32
3	FUSION OF HYPERSPECTRAL AND LIDAR REMOTE SENSING DATA FOR THE CLASSIFICATION OF COMPLEX FOREST AREAS.....	35
3.1	INTRODUCTION.....	35
3.2	DATA SET DESCRIPTION.....	39
3.3	PROBLEM DEFINITION AND SYSTEM ARCHITECTURE.....	41
3.3.1	<i>Problem definition.....</i>	<i>41</i>
3.3.2	<i>System architecture.....</i>	<i>41</i>
3.4	CLASSIFICATION TECHNIQUES.....	43
3.4.1	<i>Gaussian Maximum Likelihood with Leave-One-Out-Covariance algorithm (GML-LOOC).....</i>	<i>44</i>
3.4.2	<i>Support Vector Machine.....</i>	<i>45</i>
3.5	EXPERIMENTAL ANALYSIS AND DISCUSSION.....	46
3.5.1	<i>Experimental design.....</i>	<i>46</i>
3.5.2	<i>Experiment 1: analysis of the effectiveness of the proposed multisensor classification system.....</i>	<i>47</i>
3.5.3	<i>Experiment 2: detailed analysis of the complementary information contained in LIDAR returns.....</i>	<i>50</i>
3.5.4	<i>Experiment 3: generalization capability of the system.....</i>	<i>52</i>
3.6	CONCLUSIONS.....	53
3.7	ACKNOWLEDGMENTS.....	54
3.8	REFERENCES.....	54
4	ANALYSIS ON THE USE OF MULTIRETURN LIDAR DATA FOR THE ESTIMATION OF STEM VOLUME AT INDIVIDUAL TREE LEVEL.....	59
4.1	INTRODUCTION.....	59
4.2	DATA SET DESCRIPTION.....	61
4.3	METHODS.....	62
4.4	RESULTS.....	64

4.4.1	<i>Correlation between first return LIDAR data and tree heights.</i>	64
4.4.2	<i>Regression analysis of each variable extracted in the estimation of stem volume.</i>	64
4.4.3	<i>Regression analysis considering groups of variables for the estimation of tree stem volume.</i>	67
4.4.4	<i>Regression analysis using all the variables extracted for the estimation of tree stem volume.</i>	69
4.4.5	<i>Analysis on the relationship between the number of hits per return and the crown depth.</i>	70
4.4.6	<i>Discussion.</i>	71
4.5	CONCLUSIONS	72
4.6	ACKNOWLEDGMENTS	72
4.7	REFERENCES	73
5	A SYSTEM FOR THE ESTIMATION OF SINGLE TREE STEM DIAMETERS AND VOLUME USING MULTIRETURN LIDAR DATA	75
5.1	INTRODUCTION	75
5.2	MATERIAL AND METHODS	78
5.2.1	<i>Data set description</i>	78
5.2.2	<i>Data preprocessing and architecture of the system</i>	79
5.2.3	<i>Segmentation</i>	80
5.2.4	<i>Variable definition</i>	81
5.2.5	<i>Variable selection</i>	83
5.2.6	<i>Support Vector Regression</i>	84
5.3	EXPERIMENTAL RESULTS	86
5.3.1	<i>Design of experiments</i>	86
5.3.2	<i>Results</i>	87
5.3.3	<i>Discussion</i>	92
5.4	CONCLUSION	93
5.5	ACKNOWLEDGMENTS	93
5.6	REFERENCES	94
6	FUSION OF HYPERSPECTRAL AND LIDAR REMOTE SENSING DATA FOR THE ESTIMATION OF TREE STEM DIAMETERS	95
6.1	INTRODUCTION	95
6.2	DATA SET DESCRIPTION	96
6.3	METHODS	97
6.3.1	<i>Architecture of the proposed system</i>	97
6.3.2	<i>Preprocessing</i>	97
6.3.3	<i>Segmentation</i>	97
6.3.4	<i>Variable extraction and selection</i>	98
6.4	ESTIMATION TECHNIQUES	99
6.5	EXPERIMENTAL ANALYSIS AND DISCUSSION	99
6.6	CONCLUSION	101
6.7	REFERENCES	101
7	CONCLUSIONS	103

Chapter 1

1 Introduction and thesis overview

In this chapter an introduction to this dissertation will be given. In greater detail, we provide an overview on the remote sensing technology, on forest inventories and on the main issues related to analysis of forest areas with remote sensing data. The main objectives and the novel contributions of this thesis are also briefly presented. Finally, we describe the structure and the organization of this document.

1.1 Introduction to remote sensing

With the words “Remote Sensing” we represent a set of methods and techniques able to collect and interpret information regarding an object without being directly in contact with the item under investigation (from a remote point). Remote Sensing was born with photography but it was with the invention of airplanes and then of satellites that it assumed the meaning that has nowadays. Since the second World War remotely sensed images were widely used in the military field. It was in this domain that the main energies were used to develop the most efficient and reliable remote sensing systems. Since the 50s, when the first artificial satellites were launched, remote sensing started to be used also for civil operations, and in the recent years it became a key technology in many human activities. Thanks to the research in many fields of electronic, informatics and signal processing, there are now many kinds of sensors that are able to acquire different types of information for a great number of applications.

The remote sensing sensors can be divided into two categories: passive and active sensors. Passive sensors exploits the natural solar radiation, and in particular they collect: i) the energy coming from the sun and reflected by the Earth surface (that depends on the kind of land cover, the moisture content of vegetation and soil, the mineral content of the soil, etc), and ii) the energy spontaneously emitted by the Earth. The energy measured by the sensor is usually collected in several spectral bands (the spectral range of each single band defines the spectral resolution) in the range 0.4 – 15 μm , and over a certain elementary area (that defines the geometrical resolution). The number of spectral bands acquired ranges from very few to hundreds according to the kind of sensor. Sensors that acquire less than fifteen bands are usually called multispectral, while the others are called hyperspectral. Considering the spatial resolution we can divide the passive sensors into three categories: i) very high resolution (spatial resolution of less than one meter); ii) high resolution (spatial resolution of some meters); iii) medium resolution (spatial resolution of dozen meters); and iv) low spatial resolution (spatial resolution of hundreds meters).

Contrarily to passive sensors active sensors measure the radiation reflected by the Earth emitted by the sensor itself. There are two main kinds of active sensors which work in two different regions of the elec-

CHAPTER 1

tromagnetic spectrum: i) Radio Detection and Ranging (RADAR) sensors; and ii) Light Detection and Ranging (LIDAR) sensors. These sensors measure intrinsically the same information: i) the distance between the sensor and the target; and ii) the power of the returned waveform (called backscatter in radar systems and intensity in LIDAR ones). Due to the different working wavelengths and their specific peculiarities these two information are provided in different ways and they can be used in different application domains.

Almost all these sensors can be mounted on both satellite and aerial platforms. The main issues in using satellite platforms are: i) possibility to have acquisition over large area with a reduced cost; ii) possibility to have multiple acquisitions over the same area; and iii) possibility to have high spatial resolution but at the cost of reducing the spectral one (and thus to have less spectral bands). Concerning the use of aerial platforms: i) they allow to have both high spatial and spectral resolution; ii) they give the possibility to have immediate acquisitions in case of emergencies; and iii) usually they have a higher cost per square meter compared to satellite ones.

Due to the large amount of sensors available and their different peculiarities, remote sensing can be useful in many different applications: urban environment, agriculture, damage assessment, forestry, snow and ice monitoring, etc. In each of these applications each sensor has different potentialities allowing the user to retrieve different information from the data.

Among the possible applications of remote sensing, in this thesis we focus our attention on forestry and in particular on forest inventories. We focus on two specific topics of forest inventories: tree species classification, and forest attributes estimation.

1.2 Forest Inventories

Historically forestry has been concerned mainly with the assessment of timber resources and the management and utilization of closed forests for the production of wood. Attention was occasionally given to the other aspects of forests such as the wildlife and environmental protection. Only in the 20th century forests acquired a double relevance. On the one hand, they are important for timber exploitation; on the other hand, they are important from an environmental viewpoint. Many areas in the world base their economy on the exploitation of timber, which can be used for the local market as well as for exportation all over the world. The timber market has to relate and to find a trade off with environmental protection. Forest areas are the ecosystem for the living hood of wildlife species, and they preserve many plants that are important for the world biodiversity. From a different perspective, forests play a central role in the issues of greenhouse gasses and climate change, in particular in the context of the Kyoto protocol. Carbon cycle is a key point of the protocol and forests are strictly related to it, in particular concerning carbon stocks. Accurate estimates of carbon stocks are required to determine its role in the global carbon cycle, to estimate the level of anthropogenic disturbance (i.e. land use/land cover change) in changes that occurs in that cycle, and for monitoring mitigation efforts due to reforestation. In order to estimate the carbon stocks of forests, it is important to have detailed knowledge of them, in particular concerning the species

composition, and the biomass volume. Forest inventories are the instruments usually adopted to make such estimations. They consist in systematic collection of data and information over forests for assessment or analysis. Usually the following are important parameters to measure and note: species, diameter at breast height (DBH), height, site quality, age, and defects. From the data collected one can calculate the number of trees per hectare, the basal area, the stem volume and the value of the timber. Inventories can be done for other reasons than just calculating these parameters. The timber can be cruised to determine potential fire hazards and the risk of fire. The results of this type of inventory can be used in preventive actions and also awareness [1].

In the 18th century when forest inventories started to be carried out, they were done with visual inspection of the forest areas under investigation. Large forests were divided into smaller sections that were individually estimated by visual inspection. The estimates were added together to figure out the entire forest's resources. In the 19th century new relationships between diameter, height, and volume were discovered. These relationships allowed one for a quicker assessment of much larger forests. Thus, at the end of the 19th century, forest inventories started to be conducted through sample-based methods involving statistics. In the subsequent years these methods were better established and more accurate methods arose in the 20th century [1]. Nowadays the most common way to carry out a forest inventory is based on random sampling, statistics and only sometimes remote sensing. The area under analysis is divided into groups characterized by forests of the same age, stand structure, species, and location. Once these groups of forests have been created, random sampling points of circular shape are distributed over them. All the attributes measured in these sample areas are used to estimate through statistical models the attributes of the whole area. In this framework remote sensing data, especially high spatial resolution orthophoto are used for the definition of the homogeneous areas in which forests are divided and to draw detailed borders of the forest areas. Usually in the standard forest inventory procedure they are not involved in the estimation process. Only in the recent years remote sensing started to be a key technology in this field allowing foresters to increase the accuracy of the inventories. In fact the use of remote sensing data in the estimation process allow one to have more accurate models for the spatial estimation of forest attributes.

1.3 Remote sensing and forestry

As stated before, remote sensing can be a very useful technology for studying forest areas, as it provides objective data over all the scene analyzed; moreover, classification and regression techniques based on advanced pattern recognition approaches can provide accurate estimation of species composition and distribution, and the retrieval of the parameters useful for forest inventories (e.g. stem volume, forest structure, tree heights, etc). At the present, many different sensors are available, each with its own peculiarities and potentialities. Concerning passive sensors, they range from high spectral and spatial resolution sensors, to low spatial and spectral resolution ones. In the field of active sensors we can find Synthetic Aperture Radar (SAR) data, and LIDAR (Light Detection and Ranging) data.

CHAPTER 1

In the past years medium resolution multispectral sensors (as the Thematic Mapper of Landsat satellites) were widely used for the analysis and classification of forest areas. In particular, in the literature we can find studies on both classification and forest parameter estimation (e.g. [2]-[7]). Due to the different spectral and spatial resolutions of these sensors, it is possible to find studies that analyzed the problem with different levels of detail. Considering low spatial resolution satellite sensors (e.g., NOAA AVHRR, SPOT VGT, etc), in the classification field, the discrimination ability is mainly limited to the distinction between forested and non forested areas. As an example, in [2] Sedano *et al.* makes a land cover map of an area in Africa, distinguishing between forested and non forested areas. In the context of forest attributes estimation Xiao *et al.* in [3] underlines how SPOT-4 VEGETATION sensor can be useful for both the classification of forest types (they distinguish seven categories of forest types), and the identification of distinct growing pattern of these forests.

Regarding medium resolution sensors, they increase the level of spatial and spectral detail of the analysis respect to low resolution sensors. In the literature it is possible to find many studies with these kinds of data. Concerning classification, interesting results in the last years have been obtained with Landsat ETM+, like in [4] where eight different vegetation classes are analyzed with good results. These sensors can be useful also in the estimation of biophysical parameters, as detailed in many studies in the literature. As an example, in [5] Goodenough *et al.* compare different methods for the estimation of biophysical parameters necessary for Kyoto Protocol regulations, using multitemporal Landsat data. The results obtained are very accurate and comparable to those obtained by ground inventories.

High geometrical resolution multispectral sensors (e.g. Quickbird, Ikonos and SPOT5) provide more accurate geometrical information thanks to their high spatial resolution, but due to their low spectral resolution they do not give a real added value in the analysis of forest areas. The limited number of spectral bands of these data does not permit a detailed distinction of tree species in a forest, even though some studies are present in this field. As an example, in [6] Kosaka *et al.* analyze six forest types using Quickbird images. Some studies exist on the possibility to delineate tree crowns from high resolution satellite data, like in [7] where IKONOS data were used to define the number and the shape of tree crowns of a forest.

However, although interesting results have been obtained in forest analysis with multispectral data, when the number of species to distinguish increases, these sensors do not represent the best solution in order to achieve accurate results, as they acquire information in a relatively small number of large spectral intervals. This problem can be faced with passive hyperspectral sensors that, thanks to their ability to make dense sampling of the spectral signature, collect valuable information for an accurate and detailed forest analysis. Concerning the estimation of biophysical parameters, these data can be used in many different applications, from the estimation of chlorophyll concentration to the estimation of biomass volume. As an example, in [8] Zarco-Tejada *et al.* estimate the leaf chlorophyll content in coniferous pine forests using an hyperspectral image acquired by a CASI sensor. In [9] Kalacska *et al.* estimate forest biomass from EO-1 Hyperion hyperspectral satellite imagery obtaining accurate results. Regarding the classification

task, hyperspectral data have been used to solve several different problems, from the distinction of forest from other land covers, to the distinction of very similar tree species. As an example, Goodenough *et al.* in [4] present an interesting analysis comparing classification results on a forest area obtained with three different sensors, two multispectral (i.e. the Landsat-7 ETM+ and the EO-1 ALI) and one hyperspectral (i.e. the EO-1 Hyperion). The results of this study confirmed that with hyperspectral data (even if at medium spatial resolution) it is possible to reach much higher classification accuracies with respect to multispectral images. Moreover, as previously stated, the great potentialities of hyperspectral data in the classification step emerge when we have to distinguish very similar tree species. In [10], Clark *et al.* reached accuracies higher than 90% in distinguishing seven deciduous tree species with HYDICE sensor; in [11], Leckie *et al.* used the CASI sensor to separate five different coniferous species, obtaining promising results that demonstrate the importance of this kind of data in the classification of similar tree species.

Active remote sensing sensors are also widely used in forest analysis, with reference to both SAR (Synthetic Aperture Radar) and LIDAR systems. SAR systems are an important source of information for studies on forest environments, in particular concerning the estimation of biophysical parameters. SAR data allow one to estimate a wide range of forest parameters, from the forest structure to stem volume. In [12], Manninen *et al.* estimate Leaf Area Index (LAI) using ENVISAT ASAR data, reaching very low estimation errors. Concerning the classification domain, SAR data are mainly used for the distinction of forested from non-forested areas or in problems where classification is connected with tree parameters. In [13], Ranson *et al.* classify four vegetation classes in Siberia, according to the age of the trees and to some macro-species: young deciduous, old deciduous, young conifer and old conifer. Only few studies are used on the distinction of different tree species: an example of these kinds of studies is in [14], where the authors classify seven different vegetation classes (on a total of eight) using JPL-AIRSAR data.

An active remote sensing sensor that has been recently widely used in the study of forest areas is LIDAR. This sensor is an effective instrument for the study of tree heights, the forest structure, the forest stem volume and all the parameters connected with the vertical dimension of the scene under analysis. As an example, Andersen *et al.* in [15] analyze the potentialities of LIDAR in the estimation of some forest canopy fuel parameters, finding high correlation between LIDAR data and the biophysical parameters studied. In [17], Hyypä *et al.* estimate the stem volume using segmented first return LIDAR data, obtaining promising results. In the classification field, Holmgren *et al.* in [16] distinguish Scots pine from Norway spruce using features extracted from LIDAR data which characterize their structure and shape.

In the last years, the possibility to have acquisitions on the same area with different sensors resulted in studies focused on the combined use of multisensor data. In this context several papers have been presented on the joint use of multispectral (or hyperspectral) images and SAR data, for both classification and forest attributes estimation. In [18], Hyde *et al.* present a detailed study on the combined use of four different sensors (i.e. Landsat ETM+, Quickbird, SAR and LIDAR) for the analysis of the forest structure. In this work they underline that the combination of multispectral data (like Landsat ETM+) with LIDAR data provides good results. Regarding the joint use of LIDAR and hyperspectral data in the analy-

sis of forest areas, at the present only few works investigated their combined use. In the context of classification, it is possible to recall the study of Siemental *et al.* in [19], where the joint use of these kinds of data is considered for the separation of vegetation classes. From this paper it emerges the importance of LIDAR technology for the distinction of shrubs from trees.

On the basis of this overview on the use of remote sensing data for forestry applications we can conclude that many remote sensing data can be used in this field. In this thesis we focus our attention only on two kinds of data: hyperspectral and LIDAR. These data represent the most advanced remote sensing sensors that can be used in forestry and they allow one to reach the best results in the classification of tree species and in the estimation of forest attributes. We refer the reader to the introduction of each chapter of this thesis for a more detailed state-of-the-art on the analyzed problems.

1.4 Objectives of the thesis

In this thesis we present novel systems for the exploitation of remotely sensed data for the analysis of forest areas. In particular we focus our attention on hyperspectral and LIDAR data that, as we underline in the following chapters with a detailed analysis of the literature, are of primary importance in the study of forest areas. Moreover, our attention is also devoted to the use of advanced pattern recognition and machine learning techniques for the exploitation of the information contained in such data acquired over forest areas. Forest analysis is a wide context that can cover various aspects and themes. In this thesis we focus our attention on two topics: i) the classification of tree species; and ii) the estimation of tree stem attributes. In greater detail, the main novel contribution of this thesis can be summarized as follows:

- A. an empirical analysis on the relationship between spectral resolution, classifier complexity and classification accuracy in the classification of complex forest areas;
- B. a novel system for the fusion of hyperspectral and LIDAR remote sensing data in the classification of forest areas;
- C. an empirical analysis on the use of multireturn LIDAR data for the estimation of tree stem volume;
- D. a novel system for the estimation of single tree stem diameter and volume with multireturn LIDAR data
- E. a system for the fusion of hyperspectral and LIDAR remote sensing data in the estimation of tree stem diameters.

In the next sub-sections the main objectives and novelties of this research work will be briefly described.

A. Empirical analysis on the relationship between spectral resolution, classifier complexity and classification accuracy in the classification of complex forest areas

The processing of hyperspectral data is particularly complex both from a theoretical viewpoint (e.g. problems related to the Hughes phenomenon) and from a computational perspective. Despite many previous investigations have been presented in the literature on feature reduction and feature extraction in hyperspectral data, only few studies have analyzed the role of spectral resolution on the classification accuracy in different application domains. In this thesis, 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. In particular we analyze the behavior of the classification accuracy of different classifiers (based on different theoretical principles and characterized by different levels of complexity) versus: i) the spectral resolution of the sensor; and ii) the number of features acquired at the highest spectral resolution available with a given sensor. 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.

B. A novel system for the fusion of hyperspectral and LIDAR remote sensing data in the classification of forest areas

Hyperspectral and LIDAR remote sensing sensors are the most used in forest application. Usually hyperspectral data are used for the classification process, while LIDAR data are used in estimation problems. In this thesis we propose an analysis on the joint effect of hyperspectral and LIDAR data for the classification of complex forest areas. At the time of this work no studies existed in the literature that analyze the possibility of the joint exploitation of these data in the forest domain. In greater detail, we present: i) a novel system for the joint use of hyperspectral and LIDAR data in complex classification problems; ii) an investigation on the effectiveness of the very promising Support Vector Machines (SVM) and Gaussian Maximum Likelihood with Leave-One-Out-Covariance algorithm (GML-LOOC) classifiers for the analysis of complex forest scenarios characterized from a high number of species in a multisource framework; iii) an analysis on the effectiveness of different LIDAR returns and channels (elevation and intensity) for increasing the classification accuracy obtained with hyperspectral images, especially in relation to the discrimination of very similar classes. This novel system can be very useful for the exploitation of hyperspectral and LIDAR data in the classification domain. Moreover, it allows us to derive interesting conclusions on the effectiveness and potentialities of the joint use of these data in the forest analysis.

C. Empirical analysis on the use of multireturn LIDAR data for the estimation of single tree stems volume

Small footprint LIDAR data has been shown to be a very accurate technology to predict stem volume. In particular, most recent sensors can acquire multiple returns (more than 2) data at very high hit density, allowing one to have a detailed characterization of the canopy. These data contain information about the vertical structure of forests and trees. However, in the literature no detailed analysis on the contribution of each LIDAR return to the estimation of single tree stem volume has been carried out. In this thesis we propose an empirical analysis on this topic, with the goal to investigate the information content of each return and to point out which are the variables that maximize the information of each return. In particular our approach is as follows: individual trees are first extracted from the LIDAR data and a series of variables from both the 1st and non-first (multiple) hits associated with each crown are extracted. These variables are then correlated with ground truth individual estimates of stem volume.

The empirical analysis proposed provides useful information on the use of multireturn LIDAR data in the estimation task. In particular, the analysis presented on the exploitation of 2nd and 3rd return can drive future studies on stem volume estimation with LIDAR data.

D. A novel system for the estimation of single tree stem diameter and volume with multireturn LIDAR data

The estimation of the tree stem attributes (like stem diameter and volume) is a key point of forest inventories. In this thesis we present a system for the estimation of forest stem diameters and volume at individual tree level, which is based on multireturn LIDAR data and on Support Vector Regression (SVR). The system proposed is made up of a preprocessing module, a LIDAR segmentation algorithm (aimed at retrieving tree crowns), a variable extraction and selection procedure and an estimation phase. The variables derived from LIDAR data are computed from both the intensity and elevation channels of all available returns. Three different methods of variable selection are analyzed, and the sets of variables obtained are used in the estimation phase based on a multivariate linear regressor and a Support Vector Regression (SVR) technique. The stem volume is estimated with two approaches: i) estimation from the LIDAR variables; and ii) estimation obtained by combining the diameters and heights estimated from LIDAR variables, and the species information derived from a classification map, in a standard height/diameters relationship. Experimental results show that the system proposed is effective providing good accuracies in both the stem volumes and diameters estimation. Moreover, it provides useful information on the use of SVR in these kinds of problems.

E. A system for the fusion of hyperspectral and LIDAR remote sensing data in the estimation of tree stems diameters

As pointed out in previous subsection single tree estimations of stem parameters (such as height, diameter and volume) are usually carried out with systems based on LIDAR data. In recent years many forest areas

have been covered by hyperspectral acquisition for classification purposes. Nevertheless, no studies have been carried out on the possibility to use hyperspectral data (alone or combined with LIDAR ones) in the estimation of tree parameters at single tree level. In this thesis a system for the estimation of stem diameters with LIDAR and hyperspectral data (both separately and combined in a data fusion framework) is presented. A preliminary analysis on the effectiveness of these data in the estimation process and on the accuracy and robustness of different estimation algorithms is presented.

The system proposed and the preliminary analyses connected are important for the application of remote sensing in forestry. The possibility to exploit hyperspectral data in the context of tree parameters estimation is relevant in practical applications. Hyperspectral data are much less expensive than LIDAR ones and they are widely used for classification purposes.

1.5 Structure of the Thesis

The thesis is organized into seven chapters.

The present chapter pointed out the background and the motivations for this thesis, and highlighted the objectives as well as the novel contributions. Chapters 2 and 3 are focused on the classification of forest areas. Chapter 2 presents an empirical analysis on the role of the spectral resolution and classifier complexity in the classification of complex forest areas. The analysis is carried out on data sets with images at different spectral resolution and considering three classifiers with different levels of complexity. Chapter 3 describes a novel system for the fusion of hyperspectral and LIDAR data in the classification of forest areas. Different classifiers have been considered and experiments have been carried out with various features extracted from both hyperspectral and LIDAR data.

Chapters 4, 5 and 6 are focused on the estimation of physical parameters of trees, in particular tree height, diameter and stem volume. In chapter 4 an analysis on the use of multireturn LIDAR data for the estimation of tree stem volume is presented. Many experiments have been carried out with different predicting variables and different combinations of them. In Chapter 5 a novel system for the estimation of tree stem attributes with multireturn LIDAR data is presented. Different variable selection methods and estimation algorithms have been analyzed. Chapter 6 presents a system for the fusion of hyperspectral and LIDAR data for the estimation of tree stem diameters and volume.

In the final chapter of the thesis conclusions on the proposed systems and analysis are given. Furthermore, future developments of the dissertation are discussed.

1.6 References

- [1] J. W. Le Blanc, "What do we own: understanding forest inventories," *University of California Cooperative Extension*.
- [2] F. Sedano, P. Gong, and M. Ferrao, "Land cover assessment with MODIS imagery in southern African Miombo ecosystems," *Remote Sensing of Environment* 98, pp. 429 – 441, 2005.
- [3] X. Xiao, S. Boles, J. Liu, D. Zhuang, and M. Liu, "Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data," *Remote Sensing of Environment* 82 (2002) 335–348.

- [4] D. G. Goodenough, A. Dyk, K. O. Niemann, J. S. Pearlman, H. Chen, T. Han, M. Murdoch, and C. West, "Processing Hyperion and ALI for Forest Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, June 2003.
- [5] D. G. Goodenough, A.S. Bhogal, H. Chen, and A. Dyk, "Comparison of Methods for Estimation of Kyoto Protocol Products of Forest From Multitemporal Landsat," *IEEE International Geoscience and Remote Sensing Symposium, 2001*, vol. 2, pp. 764-767.
- [6] N. Kosaka, T. Akiyama, Bien Tsai and T. Kojima, "Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite imageries," *Proceedings IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05*. Volume 4, Issue , 25-29 July 2005, pp. 2980 – 2983.
- [7] R. Komura, M. Kubo, K. Muramoto, "Delineation of tree crown in high resolution satellite image using circle expression and watershed algorithm," *Proceedings IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS'04*, vol. 3, pp. 1577 – 1580.
- [8] P. J. Zarco-Tejada, J. R. Miller, J. Harrona, B. Hub, T. L. Nolandd, N. Goele, G. H. Mohammedd, and P. Sampson, "Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies," *Remote Sensing of Environment* 108 (2007) 82–96.
- [9] Kalacska, M.; Sanchez-Azofeifa, G.A.; Rivard, B.; Caelli, T.; White, H. Peter; Calvo-Alvarado, J.C., "Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy," *Remote Sensing of Environment*, Vol. 108, Issue: 1, May 15, 2007, pp. 82-96.
- [10] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sensing of Environment* 96, pp. 375 – 398, 2005.
- [11] D. G. Leckie, S. Tinis, T. Nelson, C. Burnett, F. A. Gougeon, E. Cloney, and D. Paradine, "Issues in species classification of trees in old growth conifer stands," *Canadian Journal of Remote Sensing*, Vol. 31, No. 2, pp. 175–190, 2005.
- [12] T. Manninen, P. Stenberg, M. Rautiainen, P. Voipio, and H. Smolander, "Leaf area index estimation of boreal forest using ENVISAT ASAR," *IEEE Transactions on Geoscience and Remote Sensing*, Volume 43, Issue 11, Nov. 2005, pp. 2627-2635.
- [13] K. J. Ranson, G. Sun, V. I. Kharuk, and K. Kovacs, "Characterization of Forests in Western Sayani Mountains, Siberia from SIR-C SAR Data," *Remote Sensing of Environment* 75, pp. 188–200, 2001.
- [14] S. S. Saatchi and E. Rignot, "Classification of Boreal Forest Cover Types Using SAR Images," *Remote Sensing of Environment* 60, pp. 270-281, 1997.
- [15] H. E. Andersen, R. J. McGaughey, and S. E. Reutebuch, "Estimating forest canopy fuel parameters using LIDAR data," *Remote Sensing of Environment* 94, pp. 441–449, 2005.
- [16] J. Holmgren and Å. Persson, "Identifying species of individual trees using airborne laser scanner," *Remote Sensing of Environment*, Vol. 90, No. 4, pp. 415-423, 2004.
- [17] J. Hyypä, O. Kelle, M. Lehtikoinen, and M. Inkinen, "A Segmentation-Based Method to retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 5, May 2001.
- [18] P. Hyde, R. Dubayah, W. Walker, J. B. Blair, M. Hofton, and C. Hunsaker, "Mapping forest structure for wildlife habitat analysis using multi-sensor (LIDAR, SAR/InSAR, ETM+, Quickbird) synergy," *Remote Sensing of Environment* 102, pp. 63–73, 2006.
- [19] E. Simental, D. J. Ragsdale, E. Bosch, R. Dodge Jr., and R. Pazak, "Hyperspectral Dimension Reduction and Elevation Data For Supervised Image Classification," *Proceedings of the 14th Annual American Society for Photogrammetry & Remote Sensing (ASPRS) Conference*, Anchorage, AK, May 3-9, 2003.

Chapter 2

2 The role of spectral resolution and classifier complexity in the study of complex forest areas

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 [6]) and from a computational perspective. Despite many previous investigations 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 chapter, 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.

2.1 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

This chapter has been published on *Remote Sensing of Environment*, Vol. 113, pp. 2345 – 2355, November 2009, with the title: “The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas”. Authors: Michele Dalponte, Lorenzo Bruzzone, Loris Vescovo and Damiano Gianelle.

CHAPTER 2

carried out on the classification of forest areas with multispectral sensors (e.g. [1],[2]). 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 species), often it is not suitable in the study of dense natural forests with many mixed species. 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.* in [3] 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. In [4] Clark *et al.* studied seven deciduous tree species with the HYDICE sensor, using three different classifiers, reaching accuracies to the order of 90%. Martin *et al.* in [5] 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 2-1 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 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 (i.e. [6]). 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 chapter is defined as the capability of a classifier to model highly nonlinear 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.

Table 2-1. Main recent hyperspectral sensors and their related spectral properties.

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

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 (i.e.[7]) and the Steepest Ascent (i.e. [8]). As a separability measure, we find the Bhattacharyya distance (i.e. [9][10]), the Jeffries-Matusita distance (i.e. [11]), as well as the transformed divergence distance (i.e. [12]). Other studies have analyzed

CHAPTER 2

the location of the most informative channels in the spectrum by considering the physical meaning of each band (e.g. [13][14]). Among them, we recall the study of Becker *et al.* [13], 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 chapter 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 chapter is focused on forest application, the proposed analysis is quite general and can be easily extended to other domains.

The chapter is organized into four sections. The next section presents the data sets used in the study, while the preprocessing procedures applied to the data, and the classifiers used in the analysis are presented in section 2.3. Section 2.4 illustrates and discusses in detail the empirical results obtained. Finally, section 2.5 draws the conclusions of the work.

2.2 Data Sets Description

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.

2.2.1 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 2-2 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*, *Quercus cerris*, *Quercus robur* and *Quercus rubra*).

The hyperspectral image (see Figure 2-1) 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.

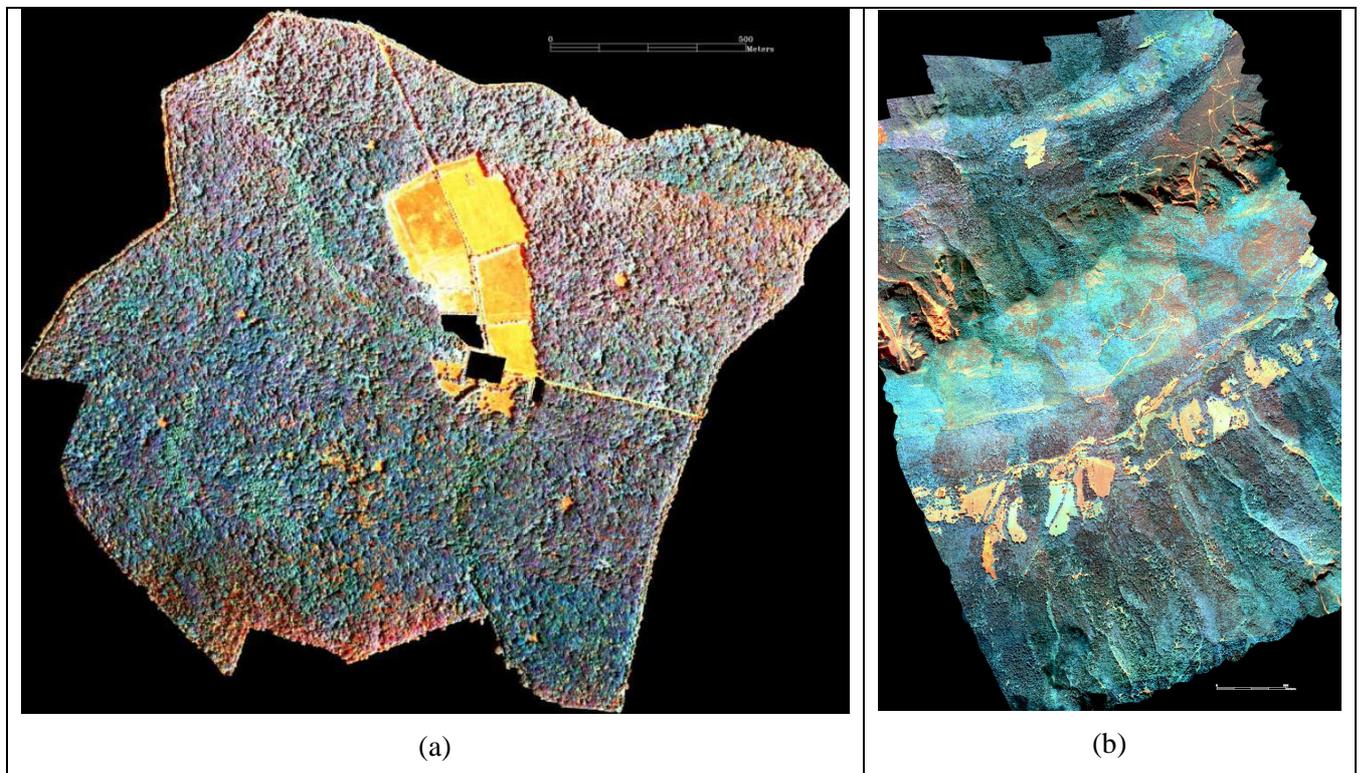


Figure 2-1. False color composition (channels 20, 70 and 110) of the hyperspectral image of *Bosco della Fontana* (a) and *Val di Sella* dataset (b).

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

form 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 at each tree corresponds more than one pixel. The total number of reference data samples (16,816 pixels) represented about 0.7% of the whole investigated area.

Table 2-2. 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
<i>Acer campestre</i>	170 (10)	<i>Juglans regia</i>	1573 (35)	<i>Quercus rubra</i>	1137 (21)
<i>Acer negundo</i>	48 (3)	<i>Morus sp.</i>	164 (5)	<i>Robinia pseudoacacia</i>	1008 (40)
<i>Alnus glutinosa</i>	507 (27)	<i>Platanus hybrida</i>	2048 (68)	<i>Rubus</i>	661
<i>Carpinus betulus</i>	910 (68)	<i>Populus canescens</i>	244 (5)	Shadows	290
<i>Corylus avellana</i>	58 (6)	<i>Populus hybrida</i>	211 (7)	Snags	205 (10)
<i>Fraxinus angustifolia</i>	787 (28)	<i>Prunus avium</i>	261 (19)	<i>Tilia cordata</i>	507 (10)
Grassland	496	<i>Quercus cerris</i>	1796 (47)	<i>Ulmus minor</i>	403 (17)
<i>Juglans nigra</i>	1283 (50)	<i>Quercus robur</i>	2049 (63)		

2.2.2 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 2-3 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 twelve 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.

Table 2-3. 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	Class Name	Reference Data Samples
<i>Abies alba</i>	179 (28)	Grassland	1010
<i>Acer pseudoplatanus</i>	146 (20)	<i>Picea abies</i>	314 (42)
<i>Alnus incana</i>	76 (7)	<i>Pinus sylvestris</i>	239 (37)
<i>Fagus sylvatica</i>	604 (57)	Shadows	192

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 at each tree corresponds more than one pixel. The total number of reference data samples (2,760 pixels) represents about 0.2% of the whole investigated area.

2.3 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. [15][16]). In our study, we adopted a simple linear normalization based on the mean-standard deviation normalization algorithm ([16]). 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., [3][17]). 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 ([17]).

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.

2.3.1 Support Vector Machine classifier

The Support Vector Machine (SVM) ([18]) is an effective distribution free classifier that has been widely used in the recent years for solving hyperspectral classification problems ([19][20]). 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 $\mathcal{X} = \{\mathbf{x}_n\}_{n=1}^N$ (where $\mathbf{x}_n \in \mathfrak{R}^q$). 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 $w \in \mathfrak{R}^{q'}$ (which is orthogonal to the hyperplane) and a bias $b \in \mathfrak{R}$ (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 $\text{sign}[f(x)]$, where $f(x)$ represents the discriminant function associated with the hyperplane and is written as:

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \quad (2.1)$$

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:

$$\Psi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{p=1}^Q \xi_p \quad (2.2)$$

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:

$$\begin{cases} \mathbf{y}_p \cdot (\mathbf{w} \cdot \mathbf{x}_p + b) \geq 1 - \xi_p, \quad \forall p = 1, \dots, Q \\ \xi_p \geq 0, \quad \forall p = 1, \dots, Q \end{cases} \quad (2.3)$$

where ξ_p are the so called slack variables and are introduced to take into account non-separable data [21]. The above optimization problem can be reformulated through a Lagrange functional as a dual optimization leading to a Quadratic Programming (QP) solution [18]. The final result is a discriminant function conveniently expressed as a function of the data in the original (lower) dimensional feature space:

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (2.4)$$

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 [22] 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 [20].

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:

1. *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.
2. *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 [20] for greater details on SVM classifiers in remote sensing and on the related multiclass strategies.

2.3.2 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) [23]. 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. [23]-[27]). In our study, we chose the algorithm proposed in [23], 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 $i=1, \dots, L$) the mean value and the covariance matrix of the i -th investigated class, respectively, and $\Omega = \{\omega_1, \omega_2, \dots, \omega_L\}$ the set of the L land-cover classes in the considered classification problem. The decision rule is as follows:

$$\mathbf{x}_n \in \omega_j \Leftrightarrow d_j(\mathbf{x}_n) > d_i(\mathbf{x}_n) \quad \forall i \neq j \quad (2.5)$$

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

$$d_i(\mathbf{x}_n) = (\mathbf{x}_n - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_i) + \ln |\boldsymbol{\Sigma}_i| \quad (2.6)$$

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: $\mathbf{S} = \frac{1}{L} \sum_{i=1}^L \boldsymbol{\Sigma}_i$ [28]. 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 $\boldsymbol{\Sigma}_i^{LOOC}$ of the i -th class is estimated as follows:

$$\boldsymbol{\Sigma}_i^{LOOC}(\alpha_i) = \begin{cases} (1 - v_i)diag(\boldsymbol{\Sigma}_i) + v_i\boldsymbol{\Sigma}_i & 0 \leq v_i \leq 1 \\ (2 - v_i)\boldsymbol{\Sigma}_i + (v_i - 1)\mathbf{S} & 1 < v_i \leq 2 \\ (3 - v_i)\mathbf{S} + (v_i - 2)diag(\mathbf{S}) & 2 < v_i \leq 3 \end{cases} \quad (2.7)$$

where v_i is a mixing parameter. The value of this parameter is selected according to the following procedure: i) removing one sample from 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 [28]. 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 [29].

2.3.3 Linear Discriminant Analysis classifier

The last technique that we consider is a very simple linear discriminant analysis (LDA) classifier ([30][31]). 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 [32]. In the following we recall the main concepts associated with LDA. We refer the reader to [32] for more details.

Let us consider a 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 according to the following equation:

$$\mathbf{a} = \mathbf{G}^T \mathbf{x} \in R^l \quad (2.8)$$

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

$$\mathbf{v}^* = \arg \max_{\mathbf{v}} \{J(\mathbf{v})\} = \arg \max_{\mathbf{v}} \left\{ \frac{\mathbf{v}^T \mathbf{S}_b \mathbf{v}}{\mathbf{v}^T \mathbf{S}_w \mathbf{v}} \right\} \quad (2.9)$$

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, $S_w = \frac{1}{n} \sum_{k=1}^L \sum_{i \in A_k} n_i (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$ 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:

$$\mathbf{v}^* = \arg \max_{\mathbf{v}} \{J(\mathbf{v})\} = \arg \max_{\mathbf{v}} \left\{ \frac{\mathbf{v}^T \mathbf{S} \mathbf{v}}{\mathbf{v}^T \mathbf{S}_w \mathbf{v}} - 1 \right\} \quad (2.10)$$

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:

$$\mathbf{G}^* = \arg \max_{\mathbf{G}} \left\{ \text{trace} \left[\left(\mathbf{G}^T \mathbf{S} \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{S}_b \mathbf{G} \right] \right\} \quad (2.11)$$

The solution can be obtained by applying an eigen-decomposition to the matrix $\mathbf{S}^{-1} \mathbf{S}_b$, if \mathbf{S} is nonsingular. 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 $k-1$ [33].

In this paragraph we have presented the standard LDA algorithm that we have used in this chapter. In presence of a reduced number of reference samples it is possible to use some regularization algorithms to avoid bad estimations of the within and between scattering matrices. For a detailed description of such algorithms we refer to [32].

2.3.4 Design of experiments

In order to achieve the goals of this chapter, 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 num-

ber 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 chapter 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 [28], while for the LDA we used the implementation contained in the MATLABArsenal software [34].

2.4 Experimental results

2.4.1 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. Figure 2-2 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.

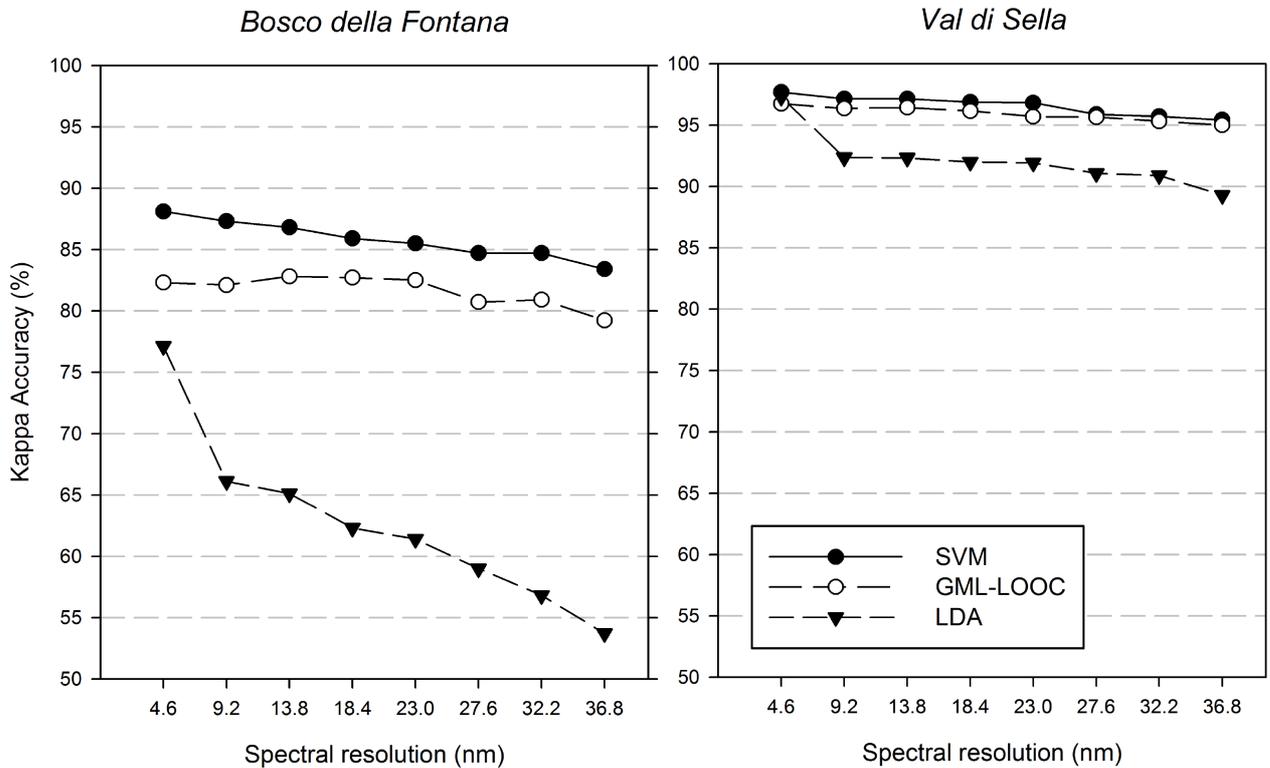


Figure 2-2. Behavior of the kappa accuracy of the analyzed classifiers versus the spectral resolution for the two datasets considered.

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

Figure 2-3 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.

2.4.2 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 [7] and on the Jeffreys-Matusita (JM) distance [11] 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.

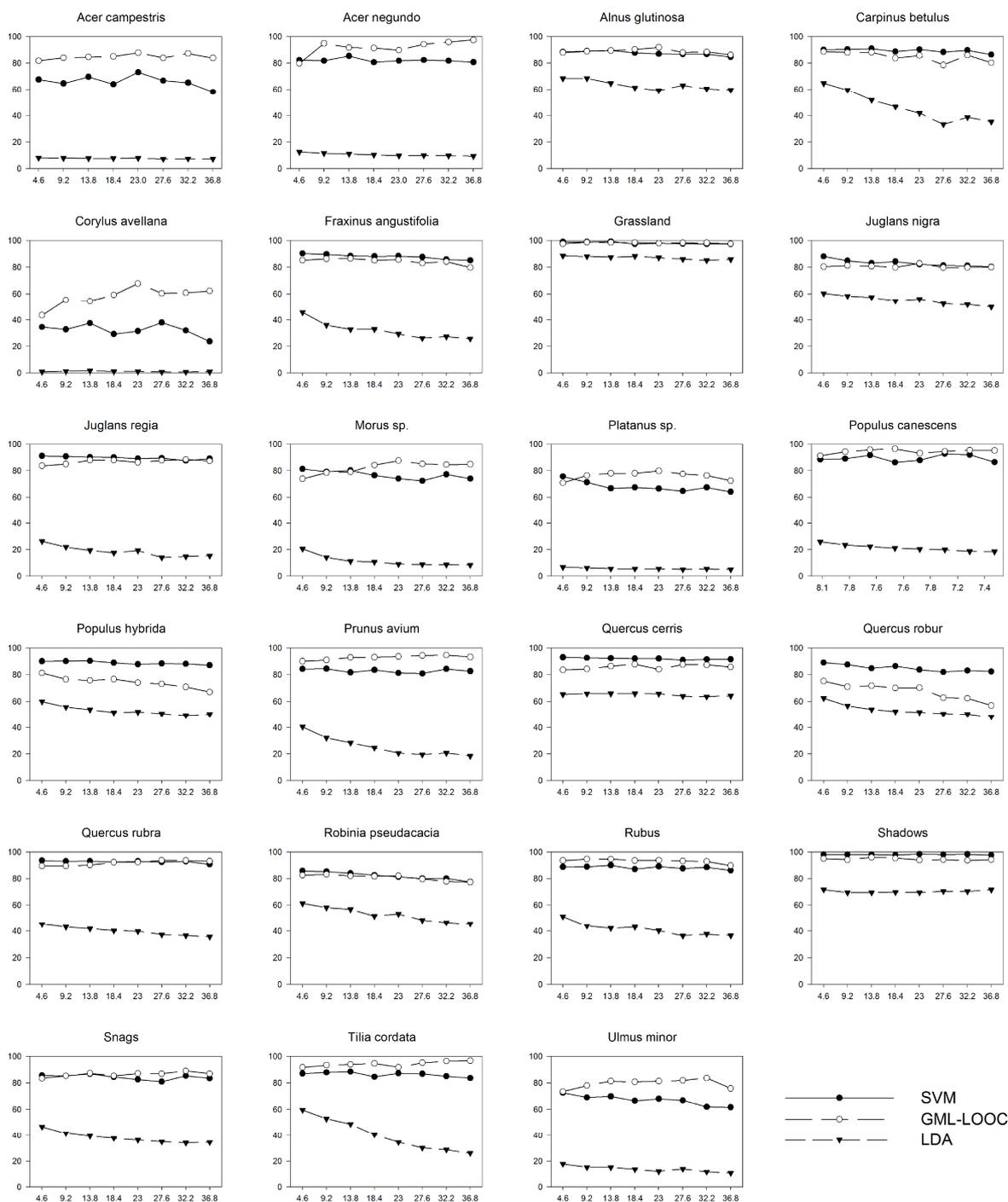


Figure 2-3. Behavior of the Producer Accuracy versus the spectral resolution for the classes analyzed in Bosco della Fontana dataset.

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.

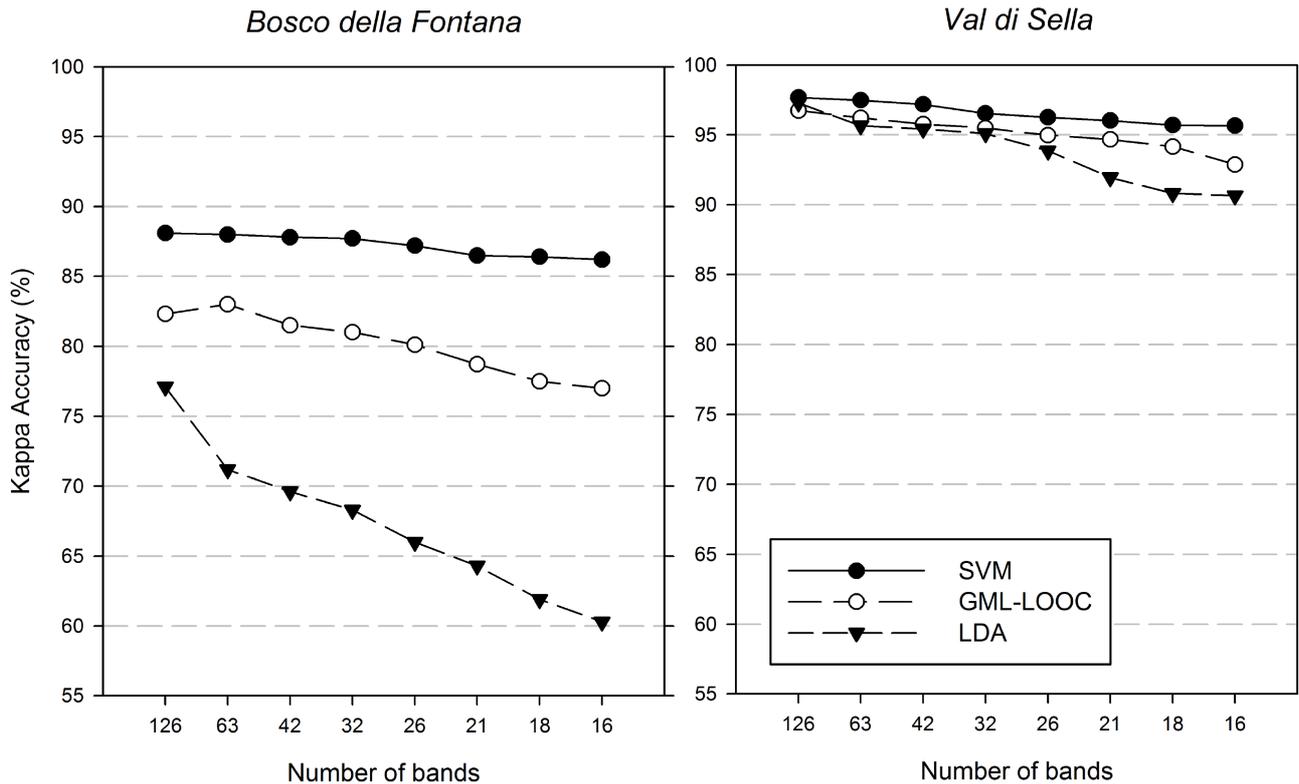


Figure 2-4. Behavior of the kappa accuracy provided by the analyzed classifiers versus the number of spectral bands at a spectral resolution of 4.6 nm for the two dataset considered.

Figure 2-4 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 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 experiment. In addition, this classifier increased its accuracy when the number of spectral channels was increased.

2.4.3 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 Figure 2-2, 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., [20][21]). 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 can be recalled to the use of a standard LDA algorithm. The use of a regularized LDA algorithm in some cases could improve the performances [32].

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 da-

CHAPTER 2

ta. 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 performances of the GML-LOOC classifier could be increased applying a decomposition of the information classes in cluster data classes. As explained previously this operation allows one to avoid the possible multimodality of the information classes distributions.

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. Figure 2-5 shows the distribution on the spectrum of 32 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 - ~500 nm), characterized mainly by carotenoides absorption peaks [35], 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 - ~600 nm) and red spectra (~600 - ~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 [36], especially around 550 nm [37]. Bands around 531 and 570 nm (two bands 535 and 573 nm were selected in our trials) were used in [38][39] 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. in [40] 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.* in [41] 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, [35][42]) and ~750 nm and ranges between the absorption region of the visible and the reflective region of the near infrared. Its position and behavior 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.

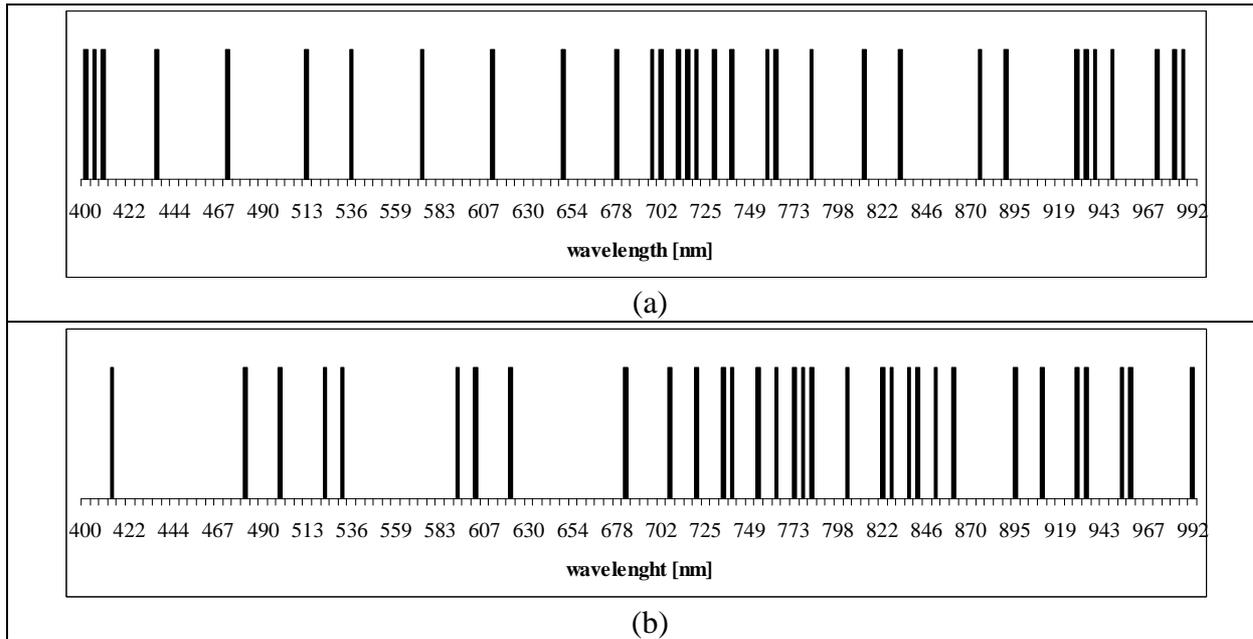


Figure 2-5. Spectral distribution of 32 hyperspectral bands selected by the feature selection algorithm for the *Bosco della Fontana* dataset (a) and the *Val di Sella* dataset (b).

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 [44]. These are linked with others parameters, such as leaf structure (that makes light scattering highly efficient), Leaf Area Index (LAI) [45], and the presence of water in the leaf (for example in the band around 970 nm [46]). 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 [47].

2.5 Conclusions

In this chapter 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.

2.6 References

- [1] F. Sedano, P. Gong, and M. Ferrao, "Land cover assessment with MODIS imagery in southern African Miombo ecosystems," *Remote Sensing of Environment*, vol. 98, pp. 429 – 441, 2005.
- [2] L. Wang, W. P. Sousa, P. Gong, and G. S. Biging, "Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama," *Remote Sensing of Environment*, vol. 91, pp. 432-440, 2004.
- [3] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416-1427, 2008.
- [4] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sensing of Environment*, vol. 96, pp. 375 – 398, 2005.
- [5] M. E. Martin, S. D. Newman, J. D. Aber, and R. G. Congalton, "Determining Forest Species Composition Using High Spectral Resolution Remote Sensing Data," *Remote Sensing of Environment*, vol. 65, pp. 249-254, 1998.
- [6] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. IT-14, pp. 55–63, 1968.
- [7] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, 1994.
- [8] S.B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1360-1367, 2001.

- [9] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by probability distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.
- [10] A. Djouadi, O. Snorrason and F. Garber, "The quality of Training-Sample estimates of the Bhattacharyya coefficient," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 12, pp. 92-97, 1990.
- [11] L. Bruzzone, F. Roli, and S. B. Serpico, "An extension to multiclass cases of the Jeffreys–Matusita distance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 1318–1321, 1995.
- [12] J. A. Richards, and X. Jia, "Remote Sensing Digital Image Analysis," Springer-Verlag, 1999.
- [13] B.L. Backer, D. P. Lusch, and J. Qi, "A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery," *Remote Sensing of Environment*, vol. 108, pp. 111-120, 2007.
- [14] K. S. Schmidt, and A. K. Skidmore, "Spectral discrimination of vegetation types in a coastal wetland," *Remote Sensing of environment*, vol. 85, no. 1, pp. 92-108, 2003.
- [15] Y. Du, J. Cihlar, J. Beaubien, and R. Latifovic, "Radiometric Normalization, Compositing, and Quality Control for Satellite High Resolution Image Mosaics over Large Areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 623-634, 2001.
- [16] D. Yuan, and C.D. Elvidge, "Comparison of relative radiometric normalization techniques," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 51, pp. 117-126, 1996.
- [17] P. Hsieh, and D. Landgrebe, "Low-pass filter for increasing class separability," *Proceedings of the IEEE Geoscience and Remote Sensing Symposium 1998*, vol. 5, pp. 2691-2693, 1998.
- [18] V.N. Vapnick, "Statistical Learning Theory," John Wiley and Sons Inc, 1998.
- [19] G. Camps-Valls, and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351-1362, 2005.
- [20] F. Melgani, and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778-1790, 2004.
- [21] L. Bruzzone, M. Chi, and M. Marconcini, M., "A Novel Transductive SVM for the Semisupervised Classification of Remote-Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3363-3373, 2006.
- [22] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London*, 1909.
- [23] J. P. Hoffbeck, and D. A. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763-767, 1996.
- [24] J. H. Friedman, "Regularized Discriminant Analysis", *Journal of the American Statistical Association*, vol. 84, pp. 165-175, March 1989.
- [25] S. P. Lin and M. D. Perlman, "A Monte Carlo comparison of four estimators of a covariance matrix," *Proceedings of the Sixth International Symposium on Multivariate Analysis*, 1985, pp. 411-429.
- [26] S. Marks and O.J. Dunn, "Discriminant Functions when the Covariance Matrices are unequal," *Journal of the American Statistical Association*, vol., 69, pp. 555-559, 1974
- [27] P. W. Wahl and R. A. Kronmall, "Discriminant Functions when Covariances are Equal and Sample Sizes are Moderate," *Biometrics*, vol. 33, pp. 479-484, 1977.
- [28] L. Biehl, "An Introduction to MultiSpec," *School of Electrical and Computer Engineering Purdue University*, 2001 (<http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/>).
- [29] M. Chi, and L. Bruzzone, "Semisupervised Classification of Hyperspectral Images by SVMs Optimized in the Primal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1870-1880, 2007.
- [30] R. O. Duda, P. E. Hart, and D. H. Stork, "Pattern Classification (2nd ed.)", Wiley Interscience, 2001.
- [31] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, vol. 7, pp. 179-188, 1936.
- [32] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 862-873, 2009.

- [33] J. Ye, T. Xiong, Q. Li, R. Janardan, J. Bi, V. Cherkassky, and C. Kambhamettu, "Efficient Model Selection for Regularized Linear Discriminant Analysis," *Proceedings of the 15th ACM international conference on Information and knowledge management*, Arlington, Virginia, USA., pp. 532–539, 2006.
- [34] R. Yan, "MATLABArsenal A MATLAB package for classification algorithms," 2006. (<http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/MATLABArsenal.htm>)
- [35] Y. Zur, A.A. Gitelson, O.B. Chivkunova, and M. N. Merzlyak, "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.
- [36] A.A. Gitelson, Y.J. Kaufman, and M.N. Merzlyak, "Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS," *Remote Sensing of Environment*, vol. 58, pp. 289-298, 1996.
- [37] A.A. Gitelson, and M.N Merzlyak, "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*, vol. 143, pp. 286–292, 1994.
- [38] J. A. Gamon, J. Pen˜uelas, and C. B. Field, "A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency," *Remote Sensing of Environment*, vol. 41, pp. 35– 44, 1992.
- [39] J. A. Gamon, L. Serrano, and J. S. Surfus, "The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types and nutrient levels," *Oecologia*, vol. 112, pp. 492– 501, 1997.
- [40] R. Darvishzadeh, A. K. Skidmore, M. Schlfer, C. Atzberger, F. Corsi, and M. A. Cho, "LAI and chlorophyll estimated for a heterogeneous grassland using hyperspectral measurements," *ISPRS Journal of Photogrammetry and Remote sensing*, vol. 63, no. 4, pp. 409–426, 2008.
- [41] P. Ceccato, N. Gobton, S. Flasse, B. Pinty, and S. Tarantola, "Designing a spectra index to estimate vegetation water content from remote sensing data: Part 1 Theoretical approach," *Remote Sensing of Environment*, vol. 82, pp. 188-197, 2002.
- [42] D.A. Sims, and J.A. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote Sensing of Environment*, vol. 81, pp. 337– 354, 2002.
- [43] D.M. Gates, H.J Keegan, J.C. Schleter, and V.R. Weidner, "Spectral properties of plants," *Applied Optics*, vol. 4, pp. 11-20, 1965.
- [44] E.B. Knipling, "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation," *Remote Sensing of Environment*, vol. 1, pp. 155-159, 1970.
- [45] P. Thenkabail, S. Enclona, E. A. Ashton, M.S. Legg, and M.J. De Dieu, "Hyperion, IKONOS, ALO, and ETM+ sensors in the study of African rainforests," *Remote Sensing of Environment*, vol. 90, no. 1, pp. 23-43, 2004.
- [46] J. Pe˜uelas, I. Filella, and L. Sweeney, "The reflectance at the 970/900 region as an indicator of plant water status," *International Journal of Remote sensing*, vol. 14, pp. 1887-1905, 1993
- [47] I.B. Strachan, E. Pattey, and J. B. Boisvert, "Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance," *Remote Sensing of Environment*, vol. 80, pp. 213-224, 2002.

Chapter 3

3 Fusion of hyperspectral and LIDAR remote sensing data for the classification of complex forest areas

In this chapter we propose an analysis on the joint effect of hyperspectral and LIDAR data for the classification of complex forest areas. In greater detail, we present: i) an advanced system for the joint use of hyperspectral and LIDAR data in complex classification problems; ii) an investigation on the effectiveness of the very promising Support Vector Machines (SVM) and Gaussian Maximum Likelihood with Leave-One-Out-Covariance algorithm (GML-LOOC) classifiers for the analysis of complex forest scenarios characterized from a high number of species in a multisource framework; iii) an analysis on the effectiveness of different LIDAR returns and channels (elevation and intensity) for increasing the classification accuracy obtained with hyperspectral images, especially in relation to the discrimination of very similar classes. Several experiments carried out on a complex forest area in Italy, provide interesting conclusions on the effectiveness and potentialities of the joint use of hyperspectral and LIDAR data and on the accuracy of the different classification techniques analyzed in the proposed system. In particular, the elevation channel of the first LIDAR return resulted very effective for the separation of species with similar spectral signatures but different mean heights, and the SVM classifier proved to be very robust and accurate in the exploitation of the considered multisource data.

3.1 Introduction

Forest preservation and management are important and complex processes, which have significant implications on the environment (e.g. protection of biological diversity, climate mitigation) and on the economy (e.g. estimation of timber volume for commercial usage). An efficient prevention and management policy requires a detailed knowledge of species composition, distribution and density. However, the assessment of the distribution of tree species in large forests by ground inventory is a difficult and time expensive task. Remote sensing is a very useful technology to perform such kind of study. This technology, if properly integrated with automatic processing techniques, allows one the analysis of large areas in a fast and accurate way. Several studies have been carried out in this field, analyzing the potentialities of different remote sensing sensors, including passive multispectral and hyperspectral sensors, as well as ac-

This chapter has been published on *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46, No. 5, pp. 1416-1427, May 2008, with the title: "Fusion of hyperspectral and LIDAR remote sensing data for the classification of complex forest areas". Authors: Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle.

CHAPTER 3

tive LIDAR (Light Detection And Ranging) and SAR (Synthetic Aperture Radar) systems (e.g. [1]-[36]). All these sensors, with their different peculiarities and characteristics, can provide different information about the analyzed forest, allowing to reach different targets, such as classification of tree species or estimation of biophysical parameters.

Standard passive multispectral sensors (like the Thematic Mapper of the Landsat satellites) have been widely used in the past years for forest classification and analysis. In the literature several studies are present on both classification and estimation of forest parameters (e.g. [1]- [4]). Regarding classification, due to the different spectral and geometrical characteristics of multispectral sensors available, it is possible to find works that analyze the problem with different levels of geometrical detail. Regarding low resolution multispectral data, the analysis is generally limited to the discrimination between forested and non-forested areas (see for example [5]). With medium resolution sensors the level of detail can be increased and thus the analysis can be focused on more specific classes, like in the study presented in [6] where, using Landsat ETM+ images, eight different vegetation classes are analyzed. High geometrical resolution multispectral sensors (e.g. Quickbird, Ikonos and SPOT5) allow a more detailed geometrical analysis considering the high spatial resolution, but due to the poor spectral information acquired by these sensors, they do not allow a detailed analysis of tree species. As an example, in [7] Kosaka *et al.* analyze six forest types using Quickbird images, and in [8] Wang *et al.* distinguish three kinds of mangrove using Ikonos and Quickbird data.

However, although significant results in forest analysis can be obtained with these kinds of data, in forest characterized by a high number of similar tree species, these sensors do not allow a detailed analysis of the different forest species, as they acquire information in a relatively small number of bands with large spectral intervals. The new generation of passive hyperspectral sensors, thanks to their ability to make a dense sampling of the spectral signature, can instead collect valuable information for a detailed classification and analysis of similar forest types. In particular, these data can be used in a wide range of different analyses of forest environments. Several studies have addressed the capability of hyperspectral data to estimate particular biophysical parameters, like chlorophyll concentration or biomass volume (e.g [9]-[11]). Concerning classification problems, hyperspectral images have been used in a wide number of forest applications, ranging from general cases focusing on the discrimination between forest and other land covers, to more detailed analysis dealing with the distinction of different tree species (e.g. [6], [12]-[14]). In [6], for example, Goodenough *et al.* present an interesting analysis comparing classification results on a forest area obtained with three different sensors, two multispectral (i.e. the Landsat-7 ETM+ and the EO-1 ALI) and one hyperspectral (i.e. the EO-1 Hyperion). The results of this study confirmed that with hyperspectral data it is possible to reach much higher classification accuracies than with multispectral images. In [13], Clark *et al.* studied seven deciduous tree species with HYDICE sensor, using three different classifiers, reaching accuracies in the order of 90%. In [14], Leckie *et al.* used CASI hyperspectral images to separate five different coniferous species, demonstrating the high importance of these kinds of data in

classification of similar tree species. In [12] Martin *et al.* separated eleven forest classes using AVIRIS data.

Active SAR and LIDAR remote sensing sensors are also widely used in forest analysis. SAR system is an important source of information for studies on forest environments. With SAR data it is possible to estimate a wide range of forest parameters, ranging from structure to biophysical indexes, like forest fuel load (e.g. [15]-[17]). In the classification domain, SAR data are mainly used for the separation of forested from non-forested areas [18], or in problems where classification is connected with tree parameters. In this context, in [19] Lee *et al.* classify different stages of the age of coniferous and deciduous trees using L-band polarimetric InSAR data. Ranson *et al.* [20] present a similar work, studying Siberian trees, dividing the vegetation in four classes: young deciduous, old deciduous, young conifer and old conifer. In [21], Saatchi *et al.* classify seven different vegetation classes (out of a total of eight) using JPL-AIRSAR data.

The use of LIDAR sensor is increasing in the context of forest applications. LIDAR is an effective information source for studies related to tree height, forest structure, biomass and all the parameters that are mainly related to the vertical dimension of the scene under analysis (e.g. [22]-[25]). LIDAR potentially allows a very precise and detailed analysis of different forest parameters. For example, in [25] Andersen *et al.* study the potentialities of LIDAR in the estimation of some forest canopy fuel parameters, finding high correlation between LIDAR data and biophysical parameters. Some studies have also been done in using LIDAR data in classification problems, in particular in cases where a reduced number of classes are investigated, such as the case of discrimination between deciduous and conifer trees (e.g. [26]-[28]). In [26], Brennan *et al.* present a study with nine classes, obtaining high classification accuracies for all classes, and emphasizing that LIDAR data can be very effective in the distinction between coniferous and deciduous trees. In [28], Holmgren *et al.* identify species of individual trees using high-density airborne laser scanner data characterizing the structure and the shape of different tree species.

The high number of remote sensing sensors available in these last years, as well as the possibility to have images acquired by different sensors on the same area, has resulted in several studies on the use of multisensor information for forest applications. In this context, many papers have been published on the joint use of multispectral (or hyperspectral) images and SAR data (e.g. [29],[30]). Recently, some works have also addressed the joint use of LIDAR and other active and passive sensors in forest parameter estimation problems (e.g. [31]-[34]). For example, in [34] Hyde *et al.* describe the results of an analysis on forest structure using four different sensors (i.e. LIDAR, SAR, Landsat ETM+ and Quickbird), underlining that for the estimation of forest parameters the combination of LIDAR and ETM+ data achieves good accuracy. Concerning classification problems, Simental *et al.* [35] explore the joint use of hyperspectral and LIDAR data for the separation of vegetation classes, underlining that LIDAR can be very useful in the separation of shrubs from trees. In [36], Lemp *et al.* exploit hyperspectral and LIDAR data for the classification of urban areas, using LIDAR for the segmentation of the scene, and then hyperspectral data for the classification of the resulting regions. In [37], Mundt *et al.* present a study on the joint use of hyper-

spectral and LIDAR data for the classification of sagebrush distribution, reporting accuracies of about 80%. In [38] Sugumaran *et al.* address the joint use of hyperspectral and LIDAR data for the identification of tree species in an urban environment, showing the effectiveness of LIDAR bands in the classification phase. Other studies exploit LIDAR data in the preprocessing phase. For example, in [39] Perry *et al.* use the DTM derived from LIDAR in the phase of geometric correction of hyperspectral images.

All the above-mentioned papers indicate a good complementary relationship between hyperspectral and LIDAR data, as they contain very different information: hyperspectral images provide a detailed description of the spectral signatures of classes, but no information on the height of ground covers; LIDAR data give detailed information about the height, but no information on the spectral signatures. However, most of the studies do not approach the integration of LIDAR and hyperspectral signals from a real data fusion perspective, but address the problem in terms of separate use of these information sources in different processing phases. In this scenario, at the present only very few investigations have been carried out on both the design of advanced classification systems capable of properly exploiting the complementary information present in these data, and the possibility to jointly use LIDAR and hyperspectral data for classification of complex forest areas in presence of many tree species.

In this chapter we address the above issues by proposing an advanced classification system for the joint exploitation of LIDAR and hyperspectral data, and by studying the importance of LIDAR data when fused with hyperspectral images in solving complex forest classification problems. The main motivation of this work is that at the present time it is becoming more common to acquire both LIDAR and hyperspectral data on forest areas. Generally, these data are used separately; in particular, hyperspectral data are exploited for forest classification and LIDAR data for forest parameter estimation. However, the availability of both data can be properly exploited in a data fusion framework both at the classification and the estimation level. In this work we focus our attention on the classification problem. The main contributions of this work to the literature are as follows:

- i. definition of an advanced system for the joint use of hyperspectral and LIDAR data in classification of complex forest areas. In particular the proposed system can properly manage: a) the hyperdimensionality of the features vector intrinsic in hyperspectral data; b) the different statistical properties of hyperspectral and LIDAR data; c) the complementary role that LIDAR data can play with respect to hyperspectral data for the discrimination of some important forest species;
- ii. investigation on the effectiveness of the very promising distribution free Support Vector Machines (SVM) and the parametric Gaussian Maximum Likelihood with Leave-One-Out-Covariance algorithm (GML-LOOC) classifiers in the analysis of complex forest scenarios characterized by a high number of species in a multisource framework;
- iii. analysis on the effectiveness of different LIDAR returns and channels (elevation and intensity) for increasing the classification accuracy obtained with hyperspectral images, especially in relation to the discrimination of very similar classes.

The proposed system was tested on a dense forest area characterized by a very high number of complex tree species (i.e. 19 species). In the experiments, we considered airborne hyperspectral images and LIDAR data with a very high geometrical resolution (1 m) and a density higher than 5 points per square meter, respectively. The results obtained confirm the effectiveness of the proposed system, and achieve interesting conclusions on the importance of the joint use of LIDAR and hyperspectral data in forest classification.

The chapter is organized into six sections. The next section describes the data set used in our analysis, while Section 3.3 presents the problem definition and the architecture of the proposed system, as well as the main preprocessing techniques adopted. The classification methods investigated in the proposed system are analyzed in Section 3.4. Section 3.5 describes and discusses the experimental results obtained. Finally, the last section draws the conclusions of this work.

3.2 Data Set Description

The study area selected is a complex forest scene that corresponds to the natural reserve of the “*Bosco della Fontana*” in the Po Plain near the city of Mantua (Italy). The central point of the area has the following coordinates: 45° 12’ 1.68” N, 10° 44’ 35.53” E. The topography of this area is almost perfectly flat and it extends across an area of approximately 230 ha. This area represents one of the best preserved forest relicts on the Po Plain. Due to the absence of a significant human impact in the last century, this area has the following interesting properties: i) it contains a high number of vegetation species (more than twenty); ii) it consists of several similar tree species, including *Quercus cerris*, *Quercus robur* and *Quercus rubra*; iii) it does not exhibit a preordered spatial tree distribution.

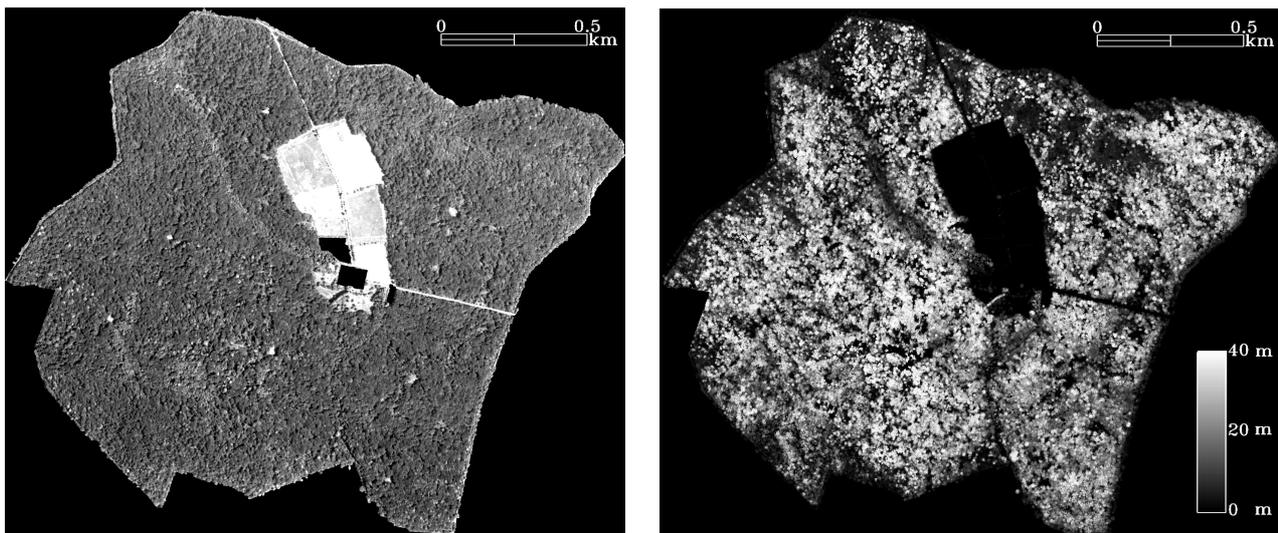
We investigated 19 different tree species, to which we added a further four classes in order to have an exhaustive representation of land covers of the whole area analyzed. In total 23 classes were represented (see Table 3-1 for a complete description of the classes investigated). It is worth noting that among the 19 tree species under analysis there are classes belonging to the same family, which have very similar spectral signatures. Another important consideration with respect to this data set is that from the analyzed area not all the vegetation classes have the same relative frequency, and that there are some dominant species (e.g. *Carpinus betulus*, *Quercus cerris*, *Quercus robur* and *Quercus rubra*).

The hyperspectral and LIDAR data (see Figure 3-1) were acquired simultaneously on June 28th, 2006 between 9:04 AM and 9:36 AM. The hyperspectral data consist 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 4.5 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 consistent, at approximately 750 m. The LIDAR data were acquired by a sensor Optech ALTM 3100, with a mean density of 5.6 points per square meter. The laser pulse wavelength and the laser repetition rate were 1064 nm and 100 kHz, respectively. The data used in our investigation refer to the first four LIDAR returns: in particular the elevation and the intensity channels of each return. The total number of LIDAR points per return is as follows: 20’271’067

points for the first return, 5'096'256 for the second, 1'110'799 for the third and 85'741 for the fourth. A Digital Terrain Model (DTM) of the investigated area with a spatial resolution of 1 m was extracted from the LIDAR data.

Table 3-1. Distribution of Reference Data Samples (Pixels) Among Investigated Classes.

Class Name	Reference Data samples	Class Name	Reference Data samples	Class Name	Reference Data samples
<i>Acer campestre</i>	170	<i>Juglans regia</i>	1573	<i>Quercus rubra</i>	1137
<i>Acer negundo</i>	48	<i>Morus sp.</i>	164	<i>Robinia pseudoacacia</i>	1008
<i>Alnus glutinosa</i>	507	<i>Platanus hybrida</i>	2048	<i>Rubus</i>	661
<i>Carpinus betulus</i>	910	<i>Populus canescens</i>	244	Shadows	290
<i>Corylus avellana</i>	58	<i>Populus hybrida</i>	211	Snags	205
<i>Fraxinus angustifolia</i>	787	<i>Prunus avium</i>	261	<i>Tilia cordata</i>	507
Grassland	496	<i>Quercus cerris</i>	1796	<i>Ulmus minor</i>	403
<i>Juglans nigra</i>	1283	<i>Quercus robur</i>	2049		



(a)

(b)

Figure 3-1. Example of images used in the experiments: a) channel 34 (550 nm) of the hyperspectral image; b) Digital Canopy Model (DCM) of the analyzed area.

The ground truth samples (approximately 550 points) were collected with a ground survey in autumn 2006. Samples were collated on a laptop within an orthophoto (with a geometrical resolution of 20 cm) of the area analyzed according to ground observations. We extracted these 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: i) of the species (the ground truth is exhaustive, i.e. it represents all the species present in the area; furthermore, it takes into account the relative frequency of each class); and ii) of the spatial distribution (samples have a uniform distribution all over the scene). All points were then converted to Region of Interests (ROIs) on the co-

registered hyperspectral and LIDAR data, and used for the generation of the training and test sets. The total number of ground truth samples (16816 pixels) represents about 0.7% of the whole investigated area.

3.3 Problem Definition and system architecture

3.3.1 Problem definition

Generally, the analysis of large forest areas with hyperspectral scanners (usually characterized from a relatively small FOV) requires the acquisition of different images which are then integrated according to a mosaic procedure. In this context, let us consider a series of M hyperspectral images with X_i ($i=1, \dots, M$) acquired in partially overlapping portions of the investigated area, and a LIDAR image L taken simultaneously with the hyperspectral ones. Let H be the radiometric normalized mosaic of these images and H_{nr} the corresponding noise reduced hyperspectral image. X_i ($i=1, \dots, M$), H and H_{nr} are n -dimensional images, where n is the total number of spectral bands. Let L denote the LIDAR interpolated image consisting of the elevation and the intensity channels of the first m LIDAR returns. The total number of bands of L is $2*m$, due to the fact that for each return we have both elevation E and intensity I image (i.e., $L = E \cup I$). Thus E and I are m -dimensional images, representing the elevation and the intensity of the first m LIDAR returns, respectively. Let x_p be the q -dimensional feature vector that represents the p -th pattern in input to the classifier. Finally let $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ be the set of the K land-cover classes in the considered classification problem, with ω_i the i -th class.

As stated in the introduction, we focus on a specific problem: the fusion of hyperspectral and LIDAR data for classification of trees species. To reach this objective, we propose a system based on an architecture that processes both hyperspectral and LIDAR data, exploiting the complementary role that these data can play. The architecture of the proposed system (with the above-defined notation) is shown in Figure 3-2. In the following we present in detail the different parts of the system.

3.3.2 System architecture

The analysis of two different kinds of data (hyperspectral and LIDAR) requires the use of two different preprocessing schemes. For hyperspectral data, according to what previously described, it is necessary to mosaic various images in order to achieve coverage of the whole site. Before this phase, a relative radiometric normalization should be applied to the single images in order to obtain a uniform mosaic image. Several normalization algorithms have been proposed in literature (e.g. [40],[41]). Since the investigated area is almost perfectly flat, and the data were acquired in a reduced interval of time (about 30 minutes), it is reasonable to assume that all six hyperspectral images were taken under the same illumination conditions. Therefore, and taking into account that in the classification phase we use a supervised classification system, we applied a relative radiometric normalization to the images without any specific atmospheric correction. In greater detail, we adopted a simple linear normalization based on the mean-standard deviation algorithm [40].

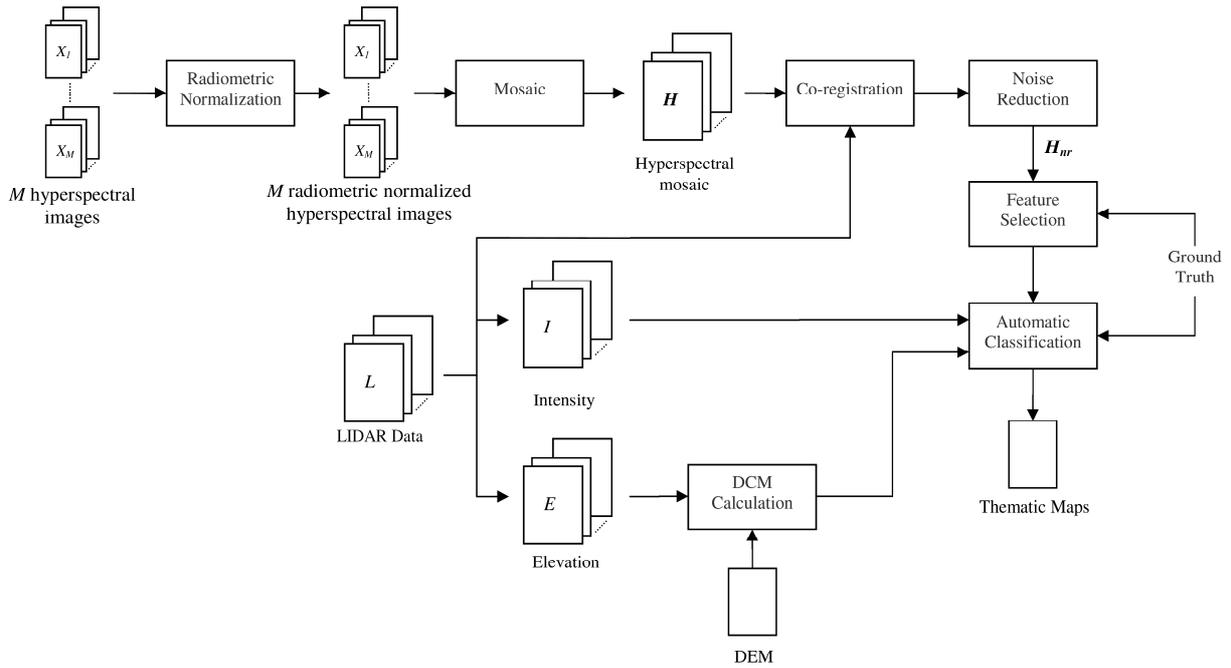


Figure 3-2. Architecture of the system developed.

After creating the mosaic, we co-registered the hyperspectral data to the LIDAR images, using approximately 75 ground control points (GCPs) distributed across the entire image. In particular, to warp the image we selected a polynomial transformation of third order and a nearest neighbor resampling of the pixels. The RMS error resulting after the co-registration phase was 0.76. The hyperspectral data were then de-noised with a simple low-pass filter with a window size of 3×3 pixels. In the previously published literature several studies pointed out the usefulness of this operation (e.g. [42][43]). In our case, given the high geometrical resolution of the images, the spatial degradation involved by the filter is acceptable with respect to both the reduction of the noise present in the images and the expected increase in class separability [42].

From a methodological viewpoint, the automatic analysis of hyperspectral data in the presence of a high number of forest classes is not a trivial task. In particular, the complexity can be attributed to: i) the high computational cost; ii) the need of advanced classification systems capable of adequately modeling the non-linear hyperdimensional discrimination functions associated with the presence of many tree species; and iii) the curse of dimensionality. In the context of supervised classification, one of the main difficulties is related to the usually small ratio between the number of available training samples and the number of features (Hughes phenomenon [44]), that makes it difficult (or impossible) to estimate the parameters of the classifier (e.g. with the Gaussian Maximum Likelihood algorithm).

A possible solution to this problem is to use a feature-selection technique. The rationale of this approach is to reduce the number of features, by selecting a representative subset of the original spectral channels. A feature-selection technique is made up of a search strategy and of a separability criterion. In the literature several algorithms have been proposed for both these tasks. Concerning the search strategy, we can

find optimal procedures (e.g. Branch and Bound [43]), which allow us to identify the subset of features that maximizes the separability criterion, or suboptimal ones (e.g. Sequential Forward Floating Selection (SFFS) [45], Steepest Ascent [46]), which find a suboptimal solution with a reduced computational cost. For our study, we adopted the SFFS algorithm, which is widely used in the literature as it provides solutions reasonably close to the optimal one. Regarding the separability criterion, several measures have been presented in the literature, including the Bhattacharyya distance, Jeffries–Matusita distance and transformed divergence [43], [47]. We selected the Jeffries-Matusita distance, which is associated to the Chernoff upper bound to the Bayesian error also in the multiclass case [48].

For the LIDAR analysis, we rasterized the raw data (corresponding to the LAS format) of all the returns. The elevation and the intensity channels were converted into a raster image with a spatial resolution of 1 m, assigning to each pixel the mean value of points within the corresponding area on the ground. The few pixels with missing data in the first return were replaced by a linear interpolation, whereas no interpolation was applied to the other returns. After this phase, in order to determine the height of vegetation with respect to the ground, we extracted the Digital Canopy Model (DCM) by subtracting the DTM to the elevation channel of the LIDAR return. This procedure was applied to the elevation band of all four LIDAR returns.

3.4 Classification Techniques

In the definition of the proposed system, we analyzed and compared two advanced classification techniques, specifically suitable to the analysis of hyperdimensional features spaces, in order to evaluate their effectiveness in classifying complex forest areas in a multisource framework. The first technique is a parametric regularized Gaussian Maximum Likelihood (GML) classifier that applies the Leave-One-Out-Covariance (LOOC) procedure [48] to the estimation of the statistics of the classifier. The second technique is a distribution-free machine learning classifier based on the Support Vector Machine (SVM) [49], [50]. The main motivations for this choice are: i) GML-LOOC and SVM have been widely used in previous studies on classification of hyperspectral data (e.g., [48]-[50]) proving their effectiveness in hyperdimensional feature spaces; ii) both techniques are intrinsically able to solve ill-posed classification problems, in which the ratio between the number of available training samples and the number of features is relatively small (this is a typical situation with hyperspectral data); iii) despite the above-mentioned common properties, GML-LOOC and SVM represent a good sampling of two different categories of classification algorithms. GML-LOOC is a parametric classifier based on the Gaussian model for the approximation of the class distributions. It represents an effective version of the widely used standard ML classifier for the analysis of hyperspectral data. The SVM classifier is a distribution-free complex classifier, which is based on machine learning and thus on a completely different theoretical background with respect to GML-LOOC. SVM proved to be very effective for classification of hyperspectral data (e.g., [49],[50]).

In the following we briefly recall the main properties of these classifiers.

3.4.1 Gaussian Maximum Likelihood with Leave-One-Out-Covariance algorithm (GML-LOOC)

This algorithm belongs to the family of parametric techniques, and is based on the Gaussian Maximum Likelihood (GML) classifier. The standard GML procedure is effective when the ratio between the number of training samples and the dimension of the feature space is relatively high, but its performance degrades when this ratio decreases (Hughes phenomenon [43]). In particular when the number of training samples is smaller than the number of features, the covariance matrix used in the decision rule become singular, and thus the GML can not be used. To avoid this problem, several algorithms have been developed for the estimation of a non-singular covariance matrix (e.g. [48]-[54]). In our study, we chose the algorithm proposed in [48], called *Leave-One-Out-Covariance* (LOOC) algorithm. In the following we give more details on this classifier.

Let \mathbf{x}_p be the p -th pattern to be classified, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ (with $i=1, \dots, K$) the mean value and the covariance matrix of the i -th investigated class, respectively. The decision rule is as follows:

$$\mathbf{x}_p \in \omega_j \Leftrightarrow d_j(\mathbf{x}_p) > d_i(\mathbf{x}_p) \quad \forall i \neq j \quad (3.1)$$

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

$$d_i(\mathbf{x}_p) = (\mathbf{x}_p - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_p - \boldsymbol{\mu}_i) + \ln |\boldsymbol{\Sigma}_i| \quad (3.2)$$

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: $\mathbf{S} = \frac{1}{K} \sum_{i=1}^K \boldsymbol{\Sigma}_i$ [48]. 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 $\boldsymbol{\Sigma}_i^{LOOC}$ of the i -th class is estimated as follows:

$$\boldsymbol{\Sigma}_i^{LOOC}(\alpha_i) = \begin{cases} (1 - \alpha_i) \text{diag}(\boldsymbol{\Sigma}_i) + \alpha_i \boldsymbol{\Sigma}_i & 0 \leq \alpha_i \leq 1 \\ (2 - \alpha_i) \boldsymbol{\Sigma}_i + (\alpha_i - 1) \mathbf{S} & 1 < \alpha_i \leq 2 \\ (3 - \alpha_i) \mathbf{S} + (\alpha_i - 2) \text{diag}(\mathbf{S}) & 2 < \alpha_i \leq 3 \end{cases} \quad (3.3)$$

where α_i is a mixing parameter, whose value is selected according to the following procedure: i) removing one sample, 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 [55]. This implementation has proved to be particularly effective in hyperspectral data classification¹.

It is worth noting that since this classifier models the class distributions according to a Gaussian function, its application to multisensor data imply a Gaussian approximation of the distribution of classes on the stacked features vector. This approximation is reasonable from an application viewpoint, but it is not rigorous from a theoretical perspective.

¹ In this chapter we used the implementation contained in the MultiSpec software [55].

3.4.2 Support Vector Machine

Support Vector Machines (SVMs) are distribution-free classifiers that overcome the aforementioned approximation of the GML-LOOC classifier. Developed by Vapnik [56], then SVM classifiers have undergone great development in last ten years and have been successfully applied to several remote sensing problems (e.g. [49],[50]). Their success is justified from four main properties: i) their relatively high classification accuracy and very good generalization capability with respect to other classifiers; ii) the limited effort required for architecture design and training phase if compared to other machine learning algorithms (such as multilayer perceptron neural networks); iii) the convexity of the cost function that finds always the optimum solution; and iv) their effectiveness in ill-posed classification problems (problems with a low ratio between number of training samples and number of features) [50]. In the following, we briefly relate the main concepts and the mathematical formulation of SVMs².

Let us consider a binary classification problem. Let us assume that the training set consists of Q vectors $\mathbf{x}_p \in R^q$, with the corresponding target $\mathbf{y}_p \in \{-1; +1\}$, where “+1” and “-1” denote the labels of the considered classes. 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 it looks for a separation between the two classes by means of an optimal hyperplane defined by a weight vector $\mathbf{w} \in R^{q'}$ and a bias $b \in R$. In particular, \mathbf{w} is a vector orthogonal to the separating hyperplane, b is a scalar value such that the ratio $b/\|\mathbf{w}\|$ represents the distance of the hyperplane from the origin, and the function Φ represents a non-linear transformation. The membership decision rule is defined according to $sign[f(\mathbf{x})]$, where $f(\mathbf{x})$ represents the discriminant function associated with the hyperplane and is written as:

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \quad (3.4)$$

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

$$\Psi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{p=1}^Q \xi_p \quad (3.5)$$

This cost function minimization is subject to the following constraints:

$$\mathbf{y}_p \cdot (\mathbf{w} \cdot \mathbf{x}_p + b) \geq 1 - \xi_p, \quad \forall p = 1, \dots, Q \quad (3.6)$$

and

$$\xi_p \geq 0, \quad \forall p = 1, \dots, Q \quad (3.7)$$

where ξ_p are the so called *slack variables* and are defined as follows:

$$\xi((\mathbf{x}_p, \mathbf{y}_p), (\mathbf{w}, b)) = \xi_p = \max(0, 1 - \mathbf{y}_p (\mathbf{w} \cdot \Phi(\mathbf{x}_p) + b)) \quad (3.8)$$

and they are introduced to take into account non-separable data. The constant C represents a regularization parameter that controls the shape of the discriminant function, and consequently the decision bound-

² We used our own implementation of SVM which is based on the SMO procedure.

ary when data are non-separable. The above optimization problem can be reformulated through a Lagrange functional for which the Lagrange multipliers can be found by means of a dual optimization leading to a Quadratic Programming (QP) solution [56]. The final result is a discriminant function conveniently expressed as a function of the data in the original (lower) dimensional feature space:

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i \Phi(\mathbf{x}_i, \mathbf{x}) + b \quad (3.9)$$

where $\Phi(\cdot)$ is a kernel function and S is the subset of training samples corresponding to the nonzero Lagrange multipliers. 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* [50]. 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*. The kernel $\Phi(\cdot)$ must satisfy the condition of Mercer's theorem so that it corresponds to some type of inner product in the transformed (higher) dimensional feature space [56].

The SVM classifier was developed to solve binary classification problems, but it can be easily extended to multiclass problems. The two main strategies used for K class problems are: i) One-Against-One (OAO) - the K -class problem is decomposed into $K(K-1)/2$ 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. ii) One-Against-All (OAA) - the K -class problem is decomposed into K 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 [50] for greater details on SVM classifiers and on the related multiclass strategies.

3.5 Experimental Analysis and Discussion

3.5.1 Experimental design

In order to assess the effectiveness of the proposed system and to achieve the goals of this chapter, we defined three different experiments: i) analysis of the importance of the joint use of hyperspectral images and first LIDAR return on the classification of complex forest areas; ii) analysis on the usefulness of multiple LIDAR returns and of the different information contained in elevation and intensity channels; iii) analysis on the generalization capability of the proposed system.

For the first two experiments, we carried out the learning of the classifier (with the model selection) and the accuracy assessment according to a k -fold cross-validation procedure. This allowed us to analyze, from a rigorous statistical perspective, the potential of the proposed system, and of the hyperspectral and LIDAR sensors, in the considered scenario. We randomly divided the available ground truth data into 5 subsets, and we then adopted a 5-fold cross-validation procedure, with training samples (pixels) distributed all over the scene. The samples of ground truth data available were used as follows: 20% in the training set (about 3300 samples) and 80% in the test set (about 13500 samples). It is worth noting that the use

of only 20% of the ground truth samples for learning tends to result in minority classes with very few training samples. However, this choice is reasonable as it represents a typical condition of real operational applications.

With regards to the last experiment, we defined the training and the test sets by considering samples from different spatially disjoint areas in order to analyze the generalization capability of the system with respect to the variability and the non-stationary behavior of the spectral signatures of the classes. In further detail, for this experiment the training and test samples were selected purposefully avoiding that they share pixels belonging to the same tree crown (i.e. all the pixels of a tree crown are completely included in only one of the two sets).

The performances of the system were assessed by using error matrices. We derived the overall kappa coefficients from these matrices, as described by Congalton *et al.* in [57], and analyzed the statistic significance of results according to the Zeta test [57].

In our experiments we used also the k -Nearest Neighbor (k -NN) classifier in order to compare the accuracy provided by the advanced classifiers included in the proposed system with a simple distribution-free classification technique. For the model selection of the SVM classifier, we chose a Gaussian kernel function, and applied a grid search strategy in a range between 50 and 240 for C , and in a range between 1 and 1000 for γ . For the k -NN classifier, the value of k varied from 1 to 29.

3.5.2 Experiment 1: analysis of the effectiveness of the proposed multisensor classification system.

Let us consider the noise reduced hyperspectral data, as well as the intensity and the corrected elevation of the first LIDAR return. In this experiment we analyze the effectiveness of first LIDAR return channels, at first considering only the global kappa accuracy, and then analyzing in greater detail the class-by-class accuracies. Experiments were conducted with three classifiers: SVM, GML-LOOC and k -Nearest Neighbor (k -NN). We carried out different trials using 126, 40, and 25 spectral channels derived according to the feature selection algorithm. Figure 3-3 shows the behavior of the average Jeffries-Matusita distance versus the number of hyperspectral channels selected with the SFFS search strategy. It is worth noting that the typical trend of this distance, which reaches saturation when the number of features used do not change the separability among information classes. We reached saturation with about 25 features but, in this experiment and in the following, we also analyzed what occurred with 40 hyperspectral features. This was done for consideration of some margin on the minimum number of input channels derived from the feature-selection phase. This is reasonable to better considering also the accuracy of minority classes that less affect the behavior of the average Jeffries-Matusita distance.

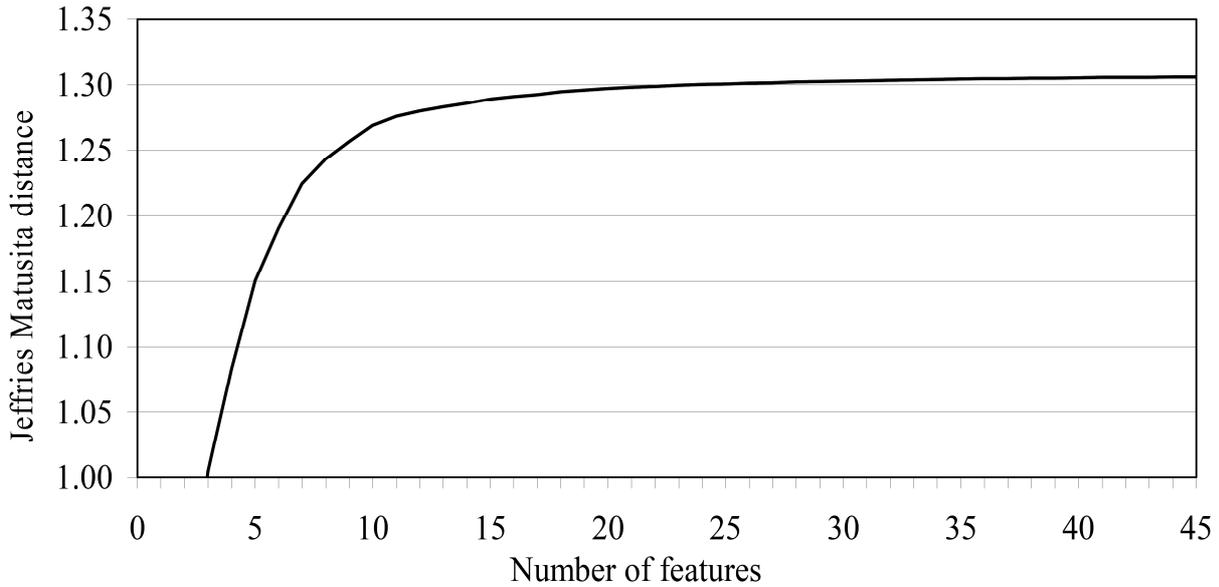


Figure 3-3. Behavior of the average Jeffries-Matusita distance versus the number of selected features.

Table 3-2. Kappa Accuracies Obtained on the Test Set with Different Spectral Features and Classifiers.

Features	Kappa Accuracy		
	SVM	GML-LOOC	k-NN
25 hyperspectral bands	0.872	0.778	0.649
40 hyperspectral bands	0.879	0.782	0.666
126 hyperspectral bands	0.881	0.823	0.676

Table 3-3. Kappa Accuracies Obtained on the Test Set with Different Spectral and LIDAR Features and Classifiers.

Features	Kappa Accuracy		
	SVM	GML-LOOC	k-NN
25 hyperspectral bands + elevation and intensity of the first LIDAR return	0.885	0.809	0.698
40 hyperspectral bands + elevation and intensity of the first LIDAR return	0.890	0.809	0.714
126 hyperspectral bands + elevation and intensity of the first LIDAR return	0.892	0.840	0.714

Table 3-2 illustrates the kappa accuracies obtained with different classifiers when varying the features used. From the analysis of these accuracies, we can infer some important points. For the SVM classifier, the accuracies obtained are particularly high considering the number of classes (23), and the number of training samples per class (as shown in Table 3-1, for some classes in the training phase we have only 10 samples). In particular, it is possible to observe that the SVM classifier always provided significantly higher accuracy than both the *k*-NN and the GML-LOOC techniques. In greater detail, thanks to its strongly non-linear properties, the SVM classifier obtained with 25 features a kappa accuracy which is higher than that obtained by the GML-LOOC technique with all the 126 channels. These results confirm the superior performances of the SVM technique, which also involves an intrinsically better generalization ability. The higher potentialities of the SVM classifier can be explained by the fact that it is a distri-

bution-free technique that does not approximate the distribution of classes with any predefined statistical model (the GML-LOOC assume Gaussian approximation), but models the decision boundary on the basis of the available training data. This results in the capability to model also strongly non-linear decision boundaries. Another important issue to note is that both SVM and GML-LOOC do not seem significantly affected by the Hughes phenomenon, since the classification accuracies increases with increasing the number of features. In the case of k-NN, the kappa accuracies for all the experiments were much smaller than those obtained by the other classifiers. This confirms that *k*-NN is not able to manage hyperdimensional feature spaces. This is especially true when classes with very few training samples are considered. The small kappa accuracies also illustrate the importance of using specific classifiers that exhibit a high generalization ability.

Let us now analyze the effect of first LIDAR return channels on the classification accuracy. Table 3-3 shows the accuracies obtained adding to different spectral features subsets the elevation and intensity channels of the first LIDAR return. Comparing the results of Table 3-2 and Table 3-3, reveals an increase in kappa accuracy from 1% to 4%, which is less relevant for the SVM and GML-LOOC classifiers, and more significant for the k-NN technique (which however does not obtain acceptable accuracies). From these results it seems that LIDAR channels provide relatively sparse information for discriminating between tree species. However, if we analyze the class-by-class accuracies, with and without LIDAR channels, the conclusions are quite different. Table 3-4 shows class-by-class accuracies obtained with the SVM classifier adding the two LIDAR channels to different spectral band subsets (25 and 40). From the analysis of the table we observe that, in general, very high accuracies were reached for very similar tree species, including *Quercus cerris*, *Quercus robur* and *Quercus rubra*. Concerning the role of LIDAR channels, we have different classification behaviors varying the number of spectral bands used. When 40 bands were used, an increase in classification accuracy occurs for classes characterized by a low height. In particular for four of the classes the increment was higher than 5%. The increase in classification accuracy becomes more relevant when reducing the number of hyperspectral bands used. With 25 hyperspectral channels, the accuracy increased by more than 10% for two classes, and more than 5% for seven classes. For example *Acer negundo* increased in accuracy by 13.56% adding LIDAR bands to 25 hyperspectral channels. Analyzing the confusion matrices this class increases its separability with respect to *Carpinus betulus*, *Platanus hybrida* and *Quercus robur* (which are characterized by a very different mean height with respect to *Acer negundo*). It is possible to draw similar conclusions also for others classes that have relevant increase in classification accuracy. It is worth noting that the classes that significantly increase their accuracy by introducing LIDAR features are the underrepresented classes. This is the motivation for the relatively small impact of this improvement on the overall classification accuracy.

The Zeta test [57] was computed between kappa accuracies obtained with 40 hyperspectral bands with and without LIDAR channels using SVM and GML-LOOC classifiers. All the differences in accuracy were statistically significant at 95% of the confidence interval.

Table 3-4. Class-by-Class Accuracy Obtained by the SVM Classifier with LIDAR and without LIDAR with a Different Number of Hyperspectral Channels.

Class Name	25 Hyperspectral Features			40 Hyperspectral Features		
	Accuracy (%)		Δ (%)	Accuracy (%)		Δ (%)
	LIDAR	no LIDAR		LIDAR	no LIDAR	
<i>Acer campestris</i>	76.76	70.29	6.47	75.59	71.03	4.56
<i>Acer negundo</i>	85.45	71.89	13.56	87.53	80.19	7.34
<i>Alnus glutinosa</i>	90.63	91.12	-0.49	91.02	90.73	0.30
<i>Carpinus betulus</i>	92.39	91.54	0.85	92.31	91.68	0.63
<i>Corylus avellana</i>	45.19	38.75	6.44	45.62	38.34	7.28
<i>Fraxinus angustifolia</i>	90.34	90.82	-0.48	90.25	89.83	0.41
Grassland	100.00	98.64	1.36	100.00	99.09	0.91
<i>Juglans nigra</i>	86.57	84.32	2.26	88.25	85.19	3.06
<i>Juglans regia</i>	89.67	90.40	-0.73	90.67	90.69	-0.02
<i>Morus sp.</i>	87.34	75.92	11.42	87.34	78.67	8.67
<i>Platanus hybrida</i>	89.99	90.23	-0.24	90.20	90.93	-0.73
<i>Populus canescens</i>	87.29	84.84	2.46	89.04	87.19	1.84
<i>Populus hybrida</i>	89.10	85.66	3.44	89.92	87.31	2.61
<i>Prunus avium</i>	79.79	72.89	6.90	80.65	76.53	4.12
<i>Quercus cerris</i>	92.87	93.07	-0.19	93.35	92.80	0.54
<i>Quercus robur</i>	86.49	86.91	-0.41	88.25	88.24	0.01
<i>Quercus rubra</i>	92.22	93.42	-1.21	91.93	93.36	-1.43
<i>Robinia pseudacacia</i>	88.54	85.05	3.50	88.22	84.90	3.32
<i>Rubus</i>	93.72	86.46	7.26	93.19	87.75	5.45
Shadows	98.02	97.93	0.09	98.28	97.84	0.43
Snags	86.34	85.98	0.37	85.61	85.85	-0.24
<i>Tilia cordata</i>	89.69	84.07	5.62	89.30	86.04	3.26
<i>Ulmus minor</i>	70.22	65.81	4.40	71.65	69.17	2.48

From these results, it is possible to conclude that first LIDAR return can be very useful in the discrimination of specific tree species. In addition, if LIDAR channels are available, it is convenient to use a reduced number of spectral channels, and to add to these channels the LIDAR information. In this perspective, on the one hand, the LIDAR channels compensate the minor lost of information due to the reduced number of spectral channels; on the other hand, the resulting smaller number of features allows both a reduction of computation time and an increase in the generalization capability of the system.

3.5.3 Experiment 2: detailed analysis of the complementary information contained in LIDAR returns.

In the first part of this experiment we considered 40 hyperspectral bands and the two channels of the first LIDAR return. From the results of the previous experiment, it is clear that these channels are useful in classification of complex forest areas, especially for discriminating between specific tree species. The next step is to understand the amount of information present in each channel (i.e. elevation and intensity).

Table 3-5. Kappa Accuracies Obtained with SVM Varying the First LIDAR return Channels Used.

Features Used	Kappa Accuracy
40 hyperspectral features	0.879
40 hyperspectral features + Elevation and Intensity of the First LIDAR return	0.890
40 hyperspectral features + Elevation of the First LIDAR return	0.888
40 hyperspectral features + Intensity of the First LIDAR return	0.876

Table 3-5 presents the kappa accuracies obtained with the SVM, with either: i) 40 hyperspectral bands, ii) 40 hyperspectral bands plus elevation and intensity of the first LIDAR return, iii) 40 hyperspectral bands plus elevation of the first LIDAR return, and iv) 40 hyperspectral bands plus intensity of the first LIDAR return. As it is clear from the table, the increase in classification accuracy obtained with LIDAR data is mainly due to the elevation channel, whereas the intensity channel does not give any relevant information for the classification of the considered forest area (it slightly decreases the overall kappa accuracy).

The second part of this experiment was focused on multiple LIDAR returns available in the data set used for this study. To analyze the information contained in these channels for the classification process, we carried out a series of trials incrementally adding the first, the second, the third and the fourth return channels (elevation and intensity) to the 40 hyperspectral bands.

Table 3-6. Kappa Accuracies Obtained with SVM Varying the Number of LIDAR Returns Jointly Used with the 40 Hyperspectral Features Selected.

LIDAR Features Used	Kappa Accuracy
1 st Return channels	0.890
1 st + 2 nd Return channels	0.878
1 st + 2 nd + 3 rd Return channels	0.872
1 st + 2 nd + 3 rd + 4 th Return channels	0.872

Table 3-6 shows the overall kappa accuracies obtained in these trials. From these results it seems that returns different from the first do not increase kappa accuracy. On the contrary, they result in a slight decrease of the accuracy with respect to that yielded using hyperspectral features plus first LIDAR return channels. These results depend on the properties of available multiple LIDAR returns. As described in section 3.2, the number of pulses is different for each return, and in particular it decreases by increasing the return number. This can be explained by the fact that the analyzed area is characterized by a very dense tree crown coverage that precludes the generation of secondary returns in many portions of the scene. For this reason, during the rasterization phase of the LIDAR data not all the pixels were associated with a value. In particular, for returns 3 and 4 we have many pixels with no data points. This introduces a noise in the classification process, thus balancing possible advantages in the characterization of the canopy of different species. In general, we expect that this issue should be better investigated using data with a higher number of representative samples from multiple returns.

3.5.4 Experiment 3: generalization capability of the system.

As described in section 3.5.1, with this last experiment we simulate a borderline case, in which training and test samples are as disjoint as possible. This allows us to verify the behavior of the proposed system when test samples belong to a significantly different area from the one considered for the training of the system. It is worth noting that in this specific case we have some classes with a reduced number of ground truth samples; this means that only few (three or four) trees in the whole scene were available for these classes. Thus, the exclusive assignment of a tree to the training or the test set makes ground truth for minority classes unrepresentative of the variability of the spectral signature over the scene, resulting in a very difficult classification problem.

Table 3-7. Kappa Accuracies Obtained on Test Set with Different Classifiers Using Disjoint Training and Test Sets.

Features Used	Kappa Accuracy		
	SVM	GML-LOOC	k-NN
40 Hyperspectral Bands	0.691	0.629	0.468
40 Hyperspectral Bands + Elevation and Intensity of the first LIDAR return	0.717	0.658	0.484

From Table 3-7, one can see that the kappa accuracies decrease with respect to the previous experiments for all the three classifiers considered. On the contrary, the differences in accuracies between the two subsets of features (with and without LIDAR) remain almost the same. The kappa accuracy of SVM was still significantly higher than those provided by others classifiers (i.e. 0.717 versus 0.658 and 0.484), but there was a large decrease with respect to those yielded in previous experiments. Also, the GML-LOOC significantly decreased the kappa accuracy. However, as expected, the more relevant degradation was associated with the k-NN classifier, that resulted in a kappa accuracy lower than 50%. Analyzing the SVM class-by-class accuracies, we observed that for some dominant classes, including *Carpinus betulus*, *Juglans regia*, *Platanus hybrida* and *Quercus rubra*, the accuracies are still in the order of 85-90% (88.41%, 87.43%, 84.35%, and 93.47%, respectively), whilst for the minority classes we have a dramatic decrease of accuracies. For example, for *Acer campestre*, *Populus hybrida*, *Prunus avium* and *Ulmus minor* the accuracies were lower than 50% (42.86%, 30.56%, 32.39%, and 24.31%, respectively). These results were expected in this very critical scenario (see [58]), that should be addressed by using semi-supervised classification techniques (like semi-supervised SVM [59]) especially developed for strongly ill-posed problems.

The differences in kappa accuracy between trials with and without LIDAR channels, with SVM and GML-LOOC classifiers were also tested with the Z test [57]. All the differences resulted to be statistically significant at 95% of the confidence interval.

3.6 Conclusions

In this chapter we investigated the joint use of hyperspectral and LIDAR remote sensing data for the classification of complex forest areas. We analyzed this issue by proposing a novel classification system, based on different possible classifiers, that were able to properly integrate multisensor information. From an analysis of the results of all the experiments carried out using the proposed system, we can conclude that in general it provided high accuracies, managing in an effective way the complementary information contained in hyperspectral and LIDAR data. In greater detail, we verified that:

- i. the presented system is very effective for classifying hyperspectral and LIDAR data, providing high accuracy on almost all the considered forest classes (it yielded accuracies of over 90% for certain classes);
- ii. the distribution-free SVM classifier provided much higher accuracies than the other classifiers investigated. The parametric GML-LOOC, even if less effective than SVM, yielded acceptable accuracies, whereas the k -NN technique (used for comparison) was unsuitable for the solution of hyperdimensional problems;
- iii. the elevation channel of the first LIDAR return data played the most important role for increasing the discriminability (and thus the accuracy) of the forest classes by having similar spectral signatures. This was due to the different average elevation of some of some forest classes;
- iv. LIDAR returns that are different from the first return do not seem capable to improve the kappa accuracy when used jointly with hyperspectral channels. However, this issue should be better analyzed on other data sets by considering a more complex feature extraction phase;
- v. in critical cases, with a large difference between training and test samples, the system based on the SVM classifier should provide an acceptable accuracy. However, in this extreme case the performances were degraded significantly and it is recommended to use specific classification techniques developed for ill-posed problems (e.g. semi-supervised [59]).

As a final remark, it is important to observe that the proposed system and study seem particularly relevant when considering that in several forest areas both hyperspectral and LIDAR data are acquired for species classification and parameter estimation, respectively. In these situations it is important to properly integrate LIDAR data in the classification process because: i) the use of hyperspectral and LIDAR data increases the separability of tree species having similar spectral signatures but different height; ii) the introduction of the first LIDAR return elevation channel produces, with a limited number of spectral features, accuracies similar to those yielded with a significantly higher number of features. This results in a lower computational time and in an increase of the generalization capability of the system.

In terms of future developments of this work we are planning to: i) introduce in the classification phase semi-supervised classifiers in order to increase the generalization ability of the system and improve the modeling of the non-stationarity of the spectral signatures of classes in the scene [58]; ii) jointly exploit hyperspectral images and LIDAR data for the estimation of biophysical forest parameters (e.g., biomass, structure, etc.).

3.7 Acknowledgments

The authors wish to thank Dr. F. Mason (*Centro Nazionale per lo Studio e la Conservazione della Biodiversità Forestale - Bosco della Fontana - Verona*) for providing the hyperspectral and LIDAR data used in this study.

3.8 References

- [1] A. A. Abuelgasim, R. A. Fernandes, and S. G. Leblanc, "Evaluation of National and Global LAI Products Derived From Optical Remote Sensing Instruments Over Canada," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 7, July 2006.
- [2] F. Maselli, and M. Chiesi, "Evaluation of Statistical Methods to Estimate Forest Volume in a Mediterranean Region," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, August 2006.
- [3] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic Spectral Rule-Based Preliminary Mapping of Calibrated Landsat TM and ETM+ Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, September 2006.
- [4] C. Song, T. A. Schroeder and W. B. Cohen, "Predicting temperate conifer forest successional stage distributions with multitemporal Landsat Thematic Mapper imagery," *Remote Sensing of Environment* 106, pp. 228–237, 2007.
- [5] F. Sedano, P. Gong, and M. Ferrao, "Land cover assessment with MODIS imagery in southern African Miombo ecosystems," *Remote Sensing of Environment* 98, pp. 429 – 441, 2005.
- [6] D. G. Goodenough, A. Dyk, K. O. Niemann, J. S. Pearlman, H. Chen, T. Han, M. Murdoch, and C. West, "Processing Hyperion and ALI for Forest Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, June 2003.
- [7] N. Kosaka, T. Akiyama, Bien Tsai and T. Kojima, "Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite imageries," *Proceedings IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05*. Volume 4, Issue , 25-29 July 2005, pp. 2980 – 2983.
- [8] L. Wang, W. P. Sousa, P. Gong, and G. S. Biging, "Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama," *Remote Sensing of Environment* 91, pp. 432–440, 2004.
- [9] D. M. Gates, H. J. Keegan, J. C. Schleter and V. R. Weidner, "Spectral properties of plants," *Applied Optics* 4, pp. 11-20, 1965.
- [10] E. B. Knipling, "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation," *Remote Sensing of Environment* 1, 155-159, 1970.
- [11] P. J. Zarco-Tejada, J. R. Millera, J. Harrona, B. Hub, T. L. Nolandd, N. Goele, G. H. Mohammedd and Paul Sampson, "Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies," *Remote Sensing of Environment*. 89, pp. 189–199, 2004.
- [12] M. E. Martin, S. D. Newman, J. D. Aber, and R. G. Congalton, "Determining Forest Species Composition Using High Spectral Resolution Remote Sensing Data," *Remote Sensing of Environment*. 65, pp. 249–254, 1998.
- [13] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sensing of Environment* 96, pp. 375 – 398, 2005.
- [14] D. G. Leckie, S. Tinis, T. Nelson, C. Burnett, F. A. Gougeon, E. Cloney, and D. Paradine, "Issues in species classification of trees in old growth conifer stands," *Canadian Journal of Remote Sensing*, Vol. 31, No. 2, pp. 175–190, 2005.
- [15] M. Watanabe, M. Shimada, A. Rosenqvist, T. Tadono, M. Matsuoka, S. A. Romshoo, K. Ohta, R. Furuta, K. Nakamura, and T. Moriyama, "Forest Structure Dependency of the Relation Between L-Band σ^0 and Biophysical Parameters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, November 2006.

- [16] S. Saatchi, K. Halligan, D. G. Despain, and R. L. Crabtree, "Estimation of Forest Fuel Load From Radar Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, June 2007.
- [17] H. Balzter, L. Skinner, A. Luckman and R. Brooke, "Estimation of tree growth in a conifer plantation over 19 years from multi-satellite L-band SAR," *Remote Sensing of Environment* 84, pp. 184–191, 2003.
- [18] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, June 2004.
- [19] J. S. Lee, K. P. Papathanassiou, I. Hajnsek, T. Mette, M. R. Grunes, T. Ainsworth, and L. Ferro-Famil, "Applying polarimetric SAR interferometric data for forest classification," *Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium. IGARSS '05*. Volume 7, 25-29 July 2005, pp. 4848–4851.
- [20] K. J. Ranson, G. Sun, V. I. Kharuk, and K. Kovacs, "Characterization of Forests in Western Sayani Mountains, Siberia from SIR-C SAR Data," *Remote Sensing of Environment* 75, pp. 188–200, 2001.
- [21] S. S. Saatchi and E. Rignot, "Classification of Boreal Forest Cover Types Using SAR Images," *Remote Sensing of Environment* 60, pp. 270-281, 1997.
- [22] M. Maltamo, K. Eerikainen, J. Pitkanen, J. Hyypäe and M. Vehmas. "Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions," *Remote Sensing of Environment*, no. 90, pp. 319–330, 2004.
- [23] B. Koetz, F. Morsdorf, G. Sun, K. J. Ranson, K. Itten, and B. Allgöwer, "Inversion of a LIDAR Waveform Model for Forest Biophysical Parameter Estimation," *IEEE Geoscience and Remote Sensing Letters*. vol. 3, no. 1, January 2006.
- [24] N. Skowronski, K. Clark, R. Nelson, J. Hom and M. Patterson, "Remotely sensed measurements of forest structure and fuel loads in the Pinelands of New Jersey," *Remote Sensing of Environment* 108, pp. 123–129, 2007.
- [25] H. E. Andersen, R. J. McGaughey, and S. E. Reutebuch, "Estimating forest canopy fuel parameters using LIDAR data," *Remote Sensing of Environment* 94, pp. 441–449, 2005.
- [26] R. Brennan and T.L. Webster, "Object-oriented land cover classification of LIDAR-derived surfaces", *Canadian Journal of Remote Sensing*, Vol. 32, No. 2, pp. 162–172, 2006.
- [27] A. P. Charaniya, R. Manduchi, and S. K. Lodha, "Supervised Parametric Classification of Aerial LIDAR Data," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)* 1063-69.
- [28] J. Holmgren and Å. Persson, "Identifying species of individual trees using airborne laser scanner," *Remote Sensing of Environment*, Vol. 90, No. 4, pp. 415-423, 2004.
- [29] M. Moghaddam, J. L. Dungan, and S. Acker, "Forest Variable Estimation From Fusion of SAR and Multispectral Optical Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, October 2002.
- [30] R. M. Lucas, N. Cronin, M. Moghaddam, A. Lee, J. Armston, P. Bunting and C. Witte, "Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia," *Remote Sensing of Environment* 100, pp. 388 – 406, 2006.
- [31] A. T. Hudak, M. A. Lefsky, W. B. Cohen and M. Berterretche, "Integration of LIDAR and Landsat ETM+ data for estimating and mapping forest canopy height," *Remote Sensing of Environment* 82, pp. 397–416, 2002.
- [32] G. A. Blackburn, "Remote sensing of forest pigments using airborne imaging spectrometer and LIDAR imagery," *Remote Sensing of Environment* 82, pp. 311– 321, 2002.
- [33] P. Hyde, R. Nelson, D. Kimes, and E. Levine, "Exploring LIDAR–RaDAR synergy—predicting aboveground biomass in a southwestern ponderosa pine forest using LIDAR, SAR and InSAR," *Remote Sensing of Environment* 106, pp. 28–38, 2007.
- [34] P. Hyde, R. Dubayah, W. Walker, J. B. Blair, M. Hofton, and C. Hunsaker, "Mapping forest structure for wildlife habitat analysis using multi-sensor (LIDAR, SAR/InSAR, ETM+, Quickbird) synergy," *Remote Sensing of Environment* 102, pp. 63–73, 2006.
- [35] E. Simental, D. J. Ragsdale, E. Bosch, R. Dodge Jr., and R. Pazak, "Hyperspectral Dimension Reduction and Elevation Data For Supervised Image Classification," *Proceedings of the 14th Annual*

- American Society for Photogrammetry & Remote Sensing (ASPRS) Conference*, Anchorage, AK, May 3-9, 2003.
- [36] D. Lemp, and U. Weidner, "Improvements of roof surface classification using hyperspectral and laser scanning data," Proceedings of the ISPRS joint conference: 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas (URBAN 2005) and 5th International Symposium Remote Sensing of Urban Areas (URS 2005), Tempe, AZ, USA, March 14-16 2005.
- [37] J. T. Mundt, D. R. Streutker and N. F. Glenn, "Mapping Sagebrush Distribution Using Fusion of Hyperspectral and LIDAR Classifications," *Photogrammetric Engineering & Remote Sensing* Vol. 72, No. 1, pp. 47–54, January 2006.
- [38] R. Sugumaran and M. Voss, "Object-Oriented Classification of LIDAR-Fused Hyperspectral Imagery for Tree Species Identification in an Urban Environment," *Urban Remote Sensing Joint Event 2007*, Paris 11-13 April 2007.
- [39] E.M. Perry, H.P. Foote, G.M. Petrie, K.L. Steinmaus, D.E. Irwin, and A.J. Stephan, "Exploitation of hyperspectral imagery and lidar for landuse classification," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Seattle (USA) 1998.
- [40] D. Yuan and C. D. Elvidge, "Comparison of relative radiometric normalization techniques," *ISPRS Journal of Photogrammetry and Remote Sensing*, 51, pp. 117-126, 1996.
- [41] Y. Du, J. Cihlar, J. Beaubien, and R. Latifovic, "Radiometric Normalization, Compositing, and Quality Control for Satellite High Resolution Image Mosaics over Large Areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, March 2001.
- [42] P. Hsieh and D. Landgrebe, "Low-pass filter for increasing class separability," *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98*.
- [43] J. A. Richards, and X. Jia, "Remote Sensing Digital Image Analysis," *Springer-Verlag*, 1999.
- [44] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 55–63, 1968.
- [45] P. Pudil, J. Novovicova and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, 15, pp. 1119-1125, 1994.
- [46] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, July 2001, pp. 1360-1367.
- [47] L. Bruzzone, F. Roli and S. B. Serpico, "An extension to multiclass cases of the Jeffries–Matusita distance", *IEEE Transactions on. Geoscience and Remote Sensing*, vol. 33, pp. 1318–1321, Nov. 1995.
- [48] J. P. Hoffbeck and D. A. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol.18, no. 7, pp. 763-767, July 1996.
- [49] G. Camps-Valls and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, No. 6, pp. 1351-1362, June 2005.
- [50] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778-1790, August 2004.
- [51] J. H. Friedman, "Regularized Discriminant Analysis", *Journal of the American Statistical Association*, Vol. 84, pp. 165-175, March 1989.
- [52] S. Marks and O.J. Dunn, "Discriminant Functions when the Covariance Matrices are unequal," *Journal of the American Statistical Association*, vol., 69, pp. 555-559, 1974
- [53] S. P. Lin and M. D. Perlman, "A Monte Carlo comparison of four estimators of a covariance matrix," *Proceedings of the Sixth International Symposium on Multivariate Analysis*, 1985, pp. 411-429.
- [54] P. W. Wahl and R. A. Kronmall, "Discriminant Functions when Covariances are Equal and Sample Sizes are Moderate," *Biometrics*, vol. 33, pp. 479-484, 1977.
- [55] L. Biehl, "An Introduction to MultiSpec", *School of Electrical and Computer Engineering Purdue University* (<http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/>).
- [56] V. N. Vapnick, "Statistical Learning Theory," *John Wiley and Sons Inc*, 1998.
- [57] R. G. Congalton and K. Green, "Assessing the Accuracy of Remotely Sensed Data: Principles and Practices," *CRC Press*, 1999.

- [58] M. Chi, and L. Bruzzone, "Semisupervised Classification of Hyperspectral Images by SVMs Optimized in the Primal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, part 2, June 2007, pp. 1870-1880.
- [59] L. Bruzzone, M. Chi, and M. Marconcini, "Semisupervised Support Vector Machines for Classification of Hyperspectral Remote Sensing Images. Hyperspectral Data Exploitation Theory and Applications," *Wiley-Interscience*, 2007, p:275-311.

CHAPTER 3

Chapter 4

4 Analysis on the use of multireturn LIDAR data for the estimation of stem volume at individual tree level

Small footprint Light Detection and Ranging (LIDAR) data have been shown to be a very accurate technology to predict stem volume. In particular most recent sensors are able to acquire multiple return (more than 2) data at very high hit density, allowing one to have detailed characterization of the canopy. In this chapter we utilize very high density (> 8 hits per m^2) LIDAR data acquired over a forest stand in Italy.

Our approach was as follows: individual trees were first extracted from the LIDAR data and a series of attributes from both the 1st, and non-first (multiple) hits associated with each crown were then extracted. These variables were then correlated with ground truth individual estimates of stem volume.

Our results indicate that: i) non-first returns are informative for the estimation of stem volume (in particular the 2nd return); ii) some attributes (e.g., maximum at the power of n) better emphasize the information content of returns different from the 1st respect to other metrics (e.g., minimum, mean); and iii) the combined use of variables belonging to different returns slightly increases the overall model accuracy. Moreover we found that the best model for stem volume estimation ($adj-R^2 = 0.77$, $P < 0.0001$, $SE = 0.06$) comprised four variables belonging to three returns (1st, 2nd, and 3rd).

The results of this analysis are important as they underline the effectiveness of the use of multiple return LIDAR data, underling the connection between LIDAR hits different from the 1st and tree structure and characteristics.

4.1 Introduction

Prediction of stem volume is an important goal of sustainable forestry, with estimates critical for both forest inventories as well as for assessing terrestrial carbon stocks as a key component of carbon accounting (i.e. [1],[2]). Although tree stem volume is generally estimated using ground based measurements, a large number of studies have demonstrated the capacity of using remotely sensed data for this purpose (e.g., [2]-[13]). There are a number of advantages of using remote sensing for the estimation of forest stem volume including the possibility to have measurements from every location in the forest, or the ability to collect data in areas difficultly accessible on the ground.

This chapter is in press on *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, with the title: "Analysis on the use of multiple return LiDAR data for the estimation of tree stems volume". Authors: Michele Dalponte, Nicholas C. Coops, Lorenzo Bruzzone, and Damiano Gianelle.

CHAPTER 4

One remote sensing technology which has been widely investigated over the past decade to estimate forestry attributes is Light Detection And Ranging (LIDAR) (e.g., [2]-[13]). These investigations can be divided into studies at stand level (e.g., [3]-[6]) and studies at single tree level (e.g., [2],[7]-[12]), with stand approaches consisting of estimating the stem volume of groups of trees usually starting from circular plot of a given radius, while single tree approaches estimate individual stem volume of each tree.

Among these two scales of application, the majority of the studies have focused on the stand level, principally due to the ready availability of plot level data from forest inventory. Moreover, in the past the majority of the LIDAR sensors did not acquire data with a sufficiently high posting density to allow multiple hits per tree crown thus making single tree level prediction of volume difficult. Naesset [3] analyzed the effects of different sensors (Optech ALTM1233 and ALTM3100), flying altitudes (1100, 1200 and 2000 m), and pulse repetition frequencies (PRF) (33, 50 and 100 kHz) on the estimation of stem volume and mean height at stand level using 1st and last return LIDAR data. The study concluded that: i) different sensors produce point clouds with different properties; ii) low PRFs tend to produce upward shifted canopy height distributions compared to higher PRFs; iii) all the datasets acquired in different conditions appear to be suitable for the estimation of volume (the “best” model developed has a R^2 of 0.92) and mean height, with a mean error of up to 10.7% for stem volume and 2.5% for mean height [3]. In [4] Coops *et al.* estimated the canopy structure of a Douglas-fir forest with 1st return LIDAR data and found high correlations between field data and LIDAR derived data ($R^2 = 0.85$ ($P < 0.001$, $SE = 1.8$ m) for the mean height, and $R^2 = 0.65$ ($P < 0.05$, $SE = 14.1$ m²ha⁻¹) for basal area). Patenaude *et al.* in [5] estimated the aboveground carbon content in a number of plots using first and last return LIDAR data and also found strong correlations ($R = 0.74$, $P < 0.01$, $SE = 4.06$ t ha⁻¹).

At the single tree scale Popescu *et al.* in [7] estimated forest volume and biomass at the individual tree level using LIDAR 1st return and a crown extraction algorithm with encouraging results (83% of the variance explained for the estimation of volume). Similarly, Hyypä *et al.* in [9] proposed a method for the estimation of stem volume using 1st return at single tree level, based on the segmentation of the individual tree crowns. Bortolot *et al.* [2] used an individual tree-based approach to estimate forest biomass using 1st return LIDAR data, obtaining good results with R ranging between 0.59 and 0.82. In [10] Wang *et al.* proposed a procedure for the analysis of the vertical canopy structure and the 3D modeling of forest. From their analysis they derived parameters from 1st return LIDAR data characterizing crown volume tree diameter and height. Likewise Falkowski *et al.* in [11] proposed an automated technique for the estimation of tree crowns based on spatial wavelet analysis and accurately predicted crown diameters ($R = 0.86$).

In the majority of these single stem volume analyses 1st return LIDAR data have been used with little investigation into the information content and applicability of returns different from the 1st or the last. This lack of investigation is principally due to the fact that, until recently, most sensors only recorded dual returns (1st and last hit); however, more recently multiple return, discrete small footprint LIDAR systems have become available allowing multiple returns (more than 2) to be recorded and subsequently analyzed.

However, whilst multiple return system may have the capacity to record more than 2 returns per LIDAR pulse, numerous factors influence the number of returns [4] including the amount of energy needed to trigger a return, the minimum time differences between two echoes, and the specific method used to detect an echo. All these factors affect the minimum distance between returns. For example, in the Optech ALTM3100 (the sensor used in this study) the minimum distance detected between the 1st and the 2nd return is 2.1 m, which increases to 3.8 m for any subsequent returns [3]. Despite these potential limitations, multiple LIDAR returns potentially provide an increase in the information provided by these sensors, in particular in applications such as predicting crown and stem attributes where multiple returns are expected. The goal of this chapter therefore is to examine the differences in the capacity of LIDAR pulse returns to predict individual stem volume based on their relative return. Our analysis is focalized on: i) single variables; ii) group of variables according to their characteristics (e.g., standard metrics, percentiles, etc) and returns (1st, 2nd, 3rd, and 4th); and iii) all the variables. Moreover we analyze the generalization ability of the best model developed with a cross-validation analysis.

This chapter is organized as follows: in section 4.2 we describe the study area and data used; in section 4.3 we present our approach with a particular focus on the phase of variables extraction. Section 4.4 illustrates the experimental results, with important discussions on the outcomes of the experiments, and finally in section 4.5 we draw some conclusions.

4.2 Data set description

The focus area for this study is a 500 ha forest stand located in the Trento Province in the north of Italy in the Italian Alps. It has a variable topography with Norway spruce (*Picea abies*) and Silver Fir (*Abies alba*), the dominant species and subdominant species including *Fagus sylvatica*, *Larix decidua* and *Pinus sylvestris*.

The field data for this study were collected in 2007 with the relascope technique. Fifty plots were randomly distributed over the study area. Within each sampling point, a standard cluster of five angle count sampling (ACS) was used to estimate mean basal area around the point, while the diameter at breast height (DBH) (1.30 m) was measured for all trees with DBH > 17.5 cm. For each sample plot, some tree heights (about 4-6 of tallest trees for species that were present in the central ACS) were measured with a Vertex hypsometer, in order to select an acceptable height-diameter function for the estimation of tree volume. For trees for which only the diameter was measured, the height was estimated using a local height-diameter function selected using the information provided by the heights measured. The height-diameter relationships were provided by the Forest Service of the Province of Trento (Italy).

The LIDAR data were acquired on September 4th, 2007, using an Optech ALTM 3100 laser scanner, with a mean density of 8.6 points per square meter. The laser pulse wavelength and the PRFs were 1064 nm and 100 kHz, respectively, with the system recording up to four returns per pulse.

Table 4-1. Summary of the field measurements (N=number of trees; DBH=diameter at breast height (1.30 m); CBH=crown base height).

Characteristic	Species						
	All	<i>Abies alba</i>	<i>Picea abies</i>	<i>Fagus sylvatica</i>	<i>Larix decidua</i>	<i>Pinus sylvestris</i>	
N	243	111	106	14	10	2	
%	100	45.68	43.62	5.76	4.12	0.82	
Tree Height (m)	Range	11.1 - 37.1	13.9 - 36.6	15.4 - 37.1	11.1 - 28.7	15.5 - 29.2	14.1 - 16.2
	Mean	26.27	25.81	27.73	21.87	24.15	15.15
	S.D	4.88	4.45	4.76	4.2	4.21	1.48
DBH (cm)	Range	16 - 74	16 - 74	22 - 72	18 - 47	35 - 63	26 - 34
	Mean	44.68	43.71	47.56	29.71	48.8	30
	S.D	11.09	9.53	11.23	8.72	9.33	5.66
CBH (m)	Range	1.4 - 23.3	2.1 - 20.8	1.4 - 23.3	2 - 15.4	1.5 - 16.5	6.8 - 10.1
	Mean	11.33	12.22	10.86	8.46	11.13	8.45
	S.D	4.57	3.80	5.23	3.70	4.33	2.33
Volume (m ³)	Range	0.16 - 6.50	0.19 - 6.50	0.29 - 5.69	0.16 - 2.24	0.63 - 2.90	0.33 - 0.66
	Mean	2.01	1.95	2.3	0.8	1.7	0.5
	S.D	1.12	0.99	1.2	0.61	0.76	0.23

In order to eliminate the effect of the topography on the elevation of the LIDAR hits and to retrieve the exact height of each tree it was necessary to subtract from each LIDAR return the height of the underlying terrain. To this end, a Digital Terrain Model (DTM) with a spatial resolution of 1 m was generated starting from the data acquired. The DTM was provided by the company that acquired the LIDAR data. This surface was then subtracted from all returns points.

4.3 Methods

The approach followed in this chapter is shown in Figure 4-1.

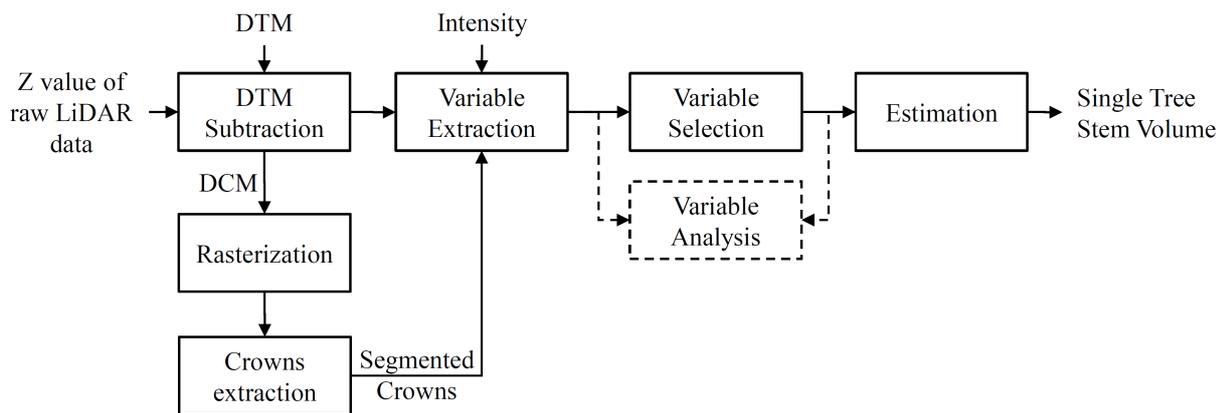


Figure 4-1. Architecture of the system adopted.

In order to derive individual crowns we first derived a Digital Canopy Model (DCM), which was calculated as the mean height of all first return hits within a 1 × 1 m grid.

To retrieve the individual tree crowns from the DCM we applied the algorithm described in [7], implemented in the software TreeVaW¹. This algorithm assumes a circular shape for the tree crowns and it is based on two main steps: i) the individual trees are located using a moving window; ii) starting from the individual tree positions the diameter of each crown is estimated.

As described in Popescu *et al.* [7][8], in the first step the local maximum (LM) technique is used to locate the tree tops. In particular this algorithm operates with a square window of $n \times n$ pixels and a circular window of variable sizes. After this step the crown diameter is identified. In this phase at first the algorithm applies a median 3x3 filter in order to reduce the outliers, preserving the edges. The crown diameter is computed as the average between two values measured along two perpendicular directions from the tree top location. In order to describe the crown profile along these two perpendicular directions the algorithm fits them with a 4th degree polynomial using the singular value decomposition (SVD). The lengths of these profiles are determined by the window size, and they are usually double of the window size. The use of a 4th degree polynomial allows one to exploit a concave shape with three extreme values. These values could be both local maxima and minima, and the values of the independent variable at extreme functional are called critical points. The algorithm finds these points, and analyzes them with a derivative analysis (first and second derivative). In particular the sign of the second derivative allows one to know if the concavity has changed. If it happens we have a point of inflection that usually occurs on the edges of a crown profile. The distance between these points is used to compute the tree crowns. The final value of the crown diameter is the average between the diameters measured on the two directions.

All tree locations were overlaid onto both a 20 cm orthophoto and the derived TreeVaW crown polygons. The size of the tree crown and tree species from the field data were used to ensure the individual tree data matched the extracted crown information to avoid errors connected with tree positions in the final model (see Table 4-1 for a detailed description of the final ground truth available). Only tree crowns which were positively matched to the LIDAR data were used in the analysis. Once the tree position and the diameter of the crown were extracted, a cylinder is defined representing the individual tree within the dataset, and all LIDAR hits were extracted.

From each identified crown we extracted a series of variables from both the elevation and the intensity information of each pulse. We divided the variables extracted into five different groups: i) “standard metrics”: *minimum*, *maximum*, *mean* and *range* value of the elevation of each return (e.g., [4][7][13]); ii) “distributional metrics”: *standard deviation*, *kurtosis*, *skewness*, *coefficient of variation* of the elevation of each return (i.e. [13]), crown *radius*, crown *area* and crown *volume* (calculated as a cylinder having as area the crown area and as height the difference between the DCM and the average height of the 2nd, 3rd or 4th return according to which is the last return available after the 1st); iii) “intensity metrics”: the *mean* value of the intensity for each return; iv) “percentiles”: the percentiles of the elevation from the 5th to the 95th for each return (e.g., [13]); and v) “maximumⁿ”: the maximum of each return elevation at the power of n (with $n=0.1, \dots, 5$) (e.g., [6]).

¹ http://www-ssl.tamu.edu/personnel/s_popescu/TreeVaW/

In order to assess the relationships between the LIDAR extracted variables and the volume we utilized a stepwise selection procedure. This approach has widely been used in previous research (e.g., [1][13]), and it is an enhancement of the forward stepwise selection. In this technique variables are added and deleted from the model according to their significance (see [14] for a more detailed description).

No predictor variable was left in the model with a significance value of the F statistic greater than 0.01. This value was applied instead of the most common 0.05 as a model with a reduced number of variables allow us to obtain a more stable model with a higher generalization ability.

In the estimation phase we utilized multivariate linear regression. In the analysis we used all the ground truth points for the creation of the model. Subsequently with the best model we applied a 10-fold cross-validation analysis using 90% of the data (about 219 trees) for the training and 10% for the test (about 24) in order to analyze the generalization ability of the model.

4.4 Results

Four sets of analysis were undertaken. First we analyzed the relationship between the LIDAR data and the tree heights (section 4.4.1). Secondly we focused on the stem volume estimation by analyzing its relationship with the extracted variables, considering each variable separately (section 4.4.2), groups of variables (section 4.4.3), and all the variables together (section 4.4.4).

4.4.1 Correlation between first return LIDAR data and tree heights.

The relationship between individual tree height and the *maximum of the 1st return* is shown in Table 4-2. The overall relationship across all species is highly significant ($\text{adj-R}^2 = 0.91$, $P < 0.0001$, $\text{SE} = 0.3$). When stratified by species the relationship remains highly significant ($\text{adj-R}^2 = 0.90$ to 0.92).

Table 4-2. Correlation between the maximum of the first return inside the crown and the tree height.

Characteristic	N	RMSE	adj-R ²
All trees	243	1.44	0.91
<i>Abies alba</i>	111	1.38	0.90
<i>Picea abies</i>	106	1.54	0.90
<i>Fagus sylvatica</i>	14	1.26	0.91
<i>Larix decidua</i>	10	1.21	0.92
<i>Pinus sylvestris</i>	2	-	-

4.4.2 Regression analysis of each variable extracted in the estimation of stem volume.

The relationship between individual stem volume and the extracted LIDAR variables presented in section 3.3 is shown in Table 4-3. Results are shown for all the reference points and for the two main species present in the investigated area.

Among the “Standard metrics” the variable which emerges to be the most highly correlated with the stem volume is the *maximum of the 1st return* ($\text{adj-R}^2 = 0.7$, $P < 0.0001$, $\text{SE} = 0.06$). This result was anticipated

as the ground truth tree stem volume was computed as a function of both height and the DBH of the stem. The second highest correlation occurs with the *maximum of the 2nd return* (adj-R² of 0.69, P < 0.0001, SE = 0.06).

Table 4-3. Variables extracted from each crown and their adj-R² relative to the volume estimation considering all the reference points and the points divided by species.

	Return	Variable	adj-R ²		
			All	<i>Abies alba</i>	<i>Picea abies</i>
Standard metrics	1 st	maximum	0.70	0.62	0.75
		minimum	0.10	0.06	0.13
		Mean	0.47	0.37	0.55
		range	0.46	0.36	0.51
	2 nd	maximum	0.69	0.62	0.74
		minimum	0.02	0.01	0.03
		mean	0.41	0.34	0.50
		range	0.50	0.38	0.56
	3 rd	maximum	0.49	0.44	0.51
		minimum	0.00	0.00	0.01
		mean	0.32	0.29	0.35
		range	0.45	0.40	0.45
	4 th	maximum	0.31	0.28	0.31
		minimum	0.04	0.06	0.03
		mean	0.22	0.22	0.20
		range	0.28	0.26	0.27
Maximum ⁿ	1 st	n=0.1, ..., 5	0.65 - 0.74	0.58 - 0.67	0.72 - 0.77
	2 nd	n=0.1, ..., 5	0.61 - 0.74	0.55 - 0.68	0.69 - 0.77
	3 rd	n=0.1, ..., 5	0.12 - 0.63	0.12 - 0.58	0.10 - 0.67
	4 th	n=0.1, ..., 5	0.19 - 0.32	0.18 - 0.33	0.18 - 0.32
Percentiles	1 st	5 th to 95 th	0.00 - 0.70	0.00 - 0.61	0.01 - 0.75
	2 nd	5 th to 95 th	0.00 - 0.66	0.00 - 0.60	0.00 - 0.72
	3 rd	5 th to 95 th	0.00 - 0.46	0.01 - 0.41	0.00 - 0.49
	4 th	5 th to 95 th	0.04 - 0.31	0.05 - 0.28	0.03 - 0.31
Distributional Metrics	1 st	standard deviation	0.47	0.38	0.44
		kurtosis	0.00	0.00	0.01
		skewness	0.01	0.00	0.06
		coefficient of variation	0.22	0.20	0.14
	2 nd	standard deviation	0.25	0.15	0.29
		kurtosis	0.00	0.00	0.00
		skewness	0.04	0.02	0.06
		coefficient of variation	0.03	0.02	0.03
	3 rd	standard deviation	0.30	0.28	0.27
		kurtosis	0.01	0.03	0.01
		skewness	0.01	0.02	0.02
		coefficient of variation	0.17	0.21	0.11
	4 th	standard deviation	0.25	0.22	0.24
		kurtosis	0.12	0.13	0.10
		skewness	0.02	0.03	0.01
		coefficient of variation	0.20	0.21	0.16
		area	0.50	0.38	0.53
		radius	0.52	0.41	0.56
		cylinder volume	0.38	0.25	0.44
Intensity metrics	1 st	mean	0.10	0.09	0.03
	2 nd	mean	0.00	0.03	0.00
	3 rd	mean	0.04	0.06	0.01
	4 th	mean	0.08	0.14	0.07

CHAPTER 4

Among the “Distributional metrics” the variables most highly correlated with volume are the *radius* and the *area*, however in both cases the correlation is quite low (adj-R² of 0.52 and 0.5, respectively).

Figure 4-2 shows a corrollogram of the relationship between the stem volume and the “percentiles”, based on the four returns. Results indicate the most significant percentile is the 95th for all the returns, with the 1st return the most informative (adj-R² = 0.70, P < 0.0001, SE = 0.06), followed by the 2nd return (adj-R² = 0.66, P < 0.0001, SE = 0.06).

The behavior of the “maximum” metric is shown in Figure 4-3. It is worth nothing that these are the variables that provide the highest levels of correlation, with a maximum of adj-R² of 0.74 (P < 0.0001, SE = 0.06). In particular, for these variables there is no difference between the 1st and the 2nd return. Moreover, in this case also the 3rd return has quite high correlations, exhibiting a maximum adjusted R² of 0.63 (P < 0.0001, SE = 0.06). This underlines the potential of returns different from the first.

Regarding the variables extracted from the intensity information, they resulted in a very low level of information (adj-R² = 0.1, P < 0.0001, SE = 0.02).

From Table 4-3 it is also possible to see the behavior of adj-R² for the two main species present in the area. As these species belong to the same family, the values of adj-R² are quite similar for all of them, with slightly higher values for the *Picea abies* with respect to the *Abies alba*. Moreover, the values obtained for these species are quite similar to the ones obtained considering all the reference points.

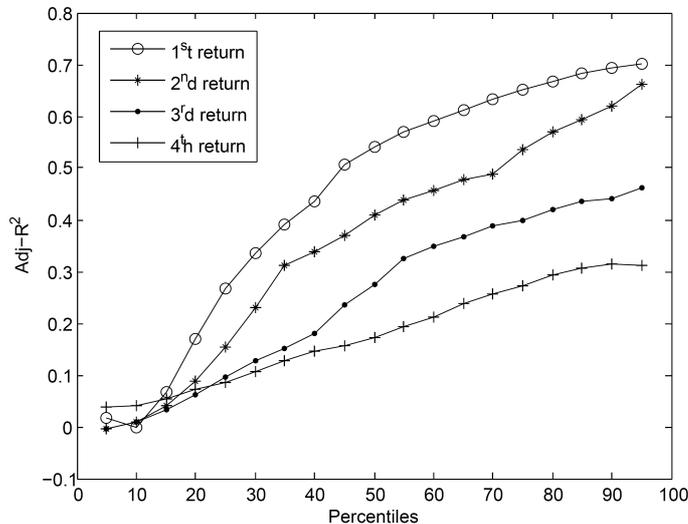


Figure 4-2. Adj-R² of the percentiles of the elevation of the different returns.

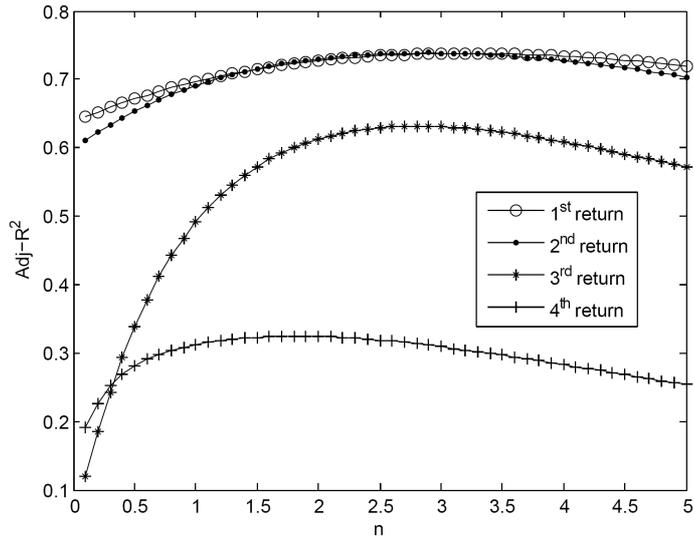


Figure 4-3. Adj-R² of the maximum of the different returns at the power of n.

4.4.3 Regression analysis considering groups of variables for the estimation of tree stem volume.

Once individual correlations were assessed, we performed regression analysis based on the groups of variables. Table 4-4 shows the results of the stepwise selection applied to different groups of variables. Interestingly the best model incorporates all classes of variables. This is important, as it underlines that the combined use of these variables increases the predictive capacity of the model. Indeed the model created with all the returns provides always higher values of adj-R² with respect to the ones generated with only variables belonging to one return.

Concerning the “standard metrics”, the variable that was always selected is the *maximum*. In two cases also other variables were selected, like the *range of the 4th return* and the *mean of the 1st one*.

The model created using the “distribution metrics” has the largest number of variables (6) (adj-R² = 0.75, P < 0.0001, SE = 0.06). These variables belong to different sources (1st and 4th return) and they are connected also to the geometry of the tree (*area* and *cylinder volume*).

Among the “percentiles” the variables derived from the 1st return provides the regression model with the highest accuracy (adj-R² = 0.75, P < 0.0001, SE = 0.06); however in most cases the 2nd return does equally well (adj-R² = 0.66, P < 0.0001, SE = 0.06). The model extracted with all the variables (adj-R² = 0.75, P < 0.0001, SE = 0.06) is made up by three variables belonging to the 1st return and one variable from the 4th, even if this variable is the last one selected.

Concerning the “intensity metrics”, also in this case they do not provide good results, with an adjusted R² of only 0.13 (P < 0.0001, SE = 0.02).

The set “maximumⁿ” included the variables that provide the highest correlations (adj-R² = 0.75, P < 0.0001, SE = 0.06). In this case, it is worth noting that the model developed with the variables belonging to the 1st (adj-R² = 0.74, P < 0.0001, SE = 0.06) and the 2nd (adj-R² = 0.74, P < 0.0001, SE = 0.06) return provide the same results, underling the effectiveness of these variables, as well as also the amount of in-

CHAPTER 4

formation contained in the 2nd return. Moreover, also the variables extracted from the 3rd return provide quite good correlations ($\text{adj-R}^2 = 0.63$, $P < 0.0001$, $SE = 0.06$).

Table 4-4. Selected models for the different sets of variables for the estimation of tree stem volume.

Initial variables set	Returns	RMSE	adj-R ²	N° var.	Variables selected
Standard metrics	All	0.60	0.72	2	maximum of the 1 st return range of the 4 th return
	1 st	0.61	0.71	2	maximum mean
	2 nd	0.62	0.69	1	maximum
	3 rd	0.80	0.49	1	maximum
	4 th	0.93	0.32	1	maximum
Other metrics	All	0.57	0.75	6	skewness of the 1 st return area cylinder volume standard deviation of the 1 st return standard deviation of the 4 th return coefficient of variation of the 1 st return
	1 st	0.60	0.72	4	standard deviation coefficient of variation kurtosis skewness
	2 nd	0.78	0.52	2	standard deviation coefficient of variation
	3 rd	0.83	0.45	2	standard deviation coefficient of variation
	4 th	0.93	0.32	2	standard deviation coefficient of variation
Intensity metrics	All	1.04	0.13	2	mean of the 1 st return mean of the 4 th return
Percentiles	All	0.57	0.75	4	10 th percentile of the 1 st return 55 th percentile of the 1 st return 85 th percentile of the 1 st return 90 th percentile of the 4 th return
	1 st	0.58	0.74	3	10 th percentile 55 th percentile 85 th percentile
	2 nd	0.65	0.66	1	95 th percentile
	3 rd	0.82	0.46	1	95 th percentile
	4 th	0.90	0.37	3	5 th percentile 15 th percentile 90 th percentile
Maximum ⁿ	All	0.56	0.75	3	maximum of the 1 st at the power of 2 maximum of the 2 nd at the power of 3.5 maximum of the 4 th at the power of 4.2
	1 st	0.57	0.74	1	maximum at the power of 3.2
	2 nd	0.57	0.74	1	maximum at the power of 2.9
	3 rd	0.69	0.63	2	maximum at the power of 1.1 maximum at the power of 1.4
	4 th	0.91	0.35	2	maximum at the power of 0.3 maximum at the power of 3.8

4.4.4 Regression analysis using all the variables extracted for the estimation of tree stem volume.

In this final analysis we considered all the variables extracted from all the four returns. Table 4-5 shows that the model developed using all the variables has the highest correlation ($\text{adj-R}^2 = 0.77$, $P < 0.0001$, $SE = 0.06$). In this case the model is made up of four variables belonging to the 1st, the 2nd and the 3rd return. This is important as the selected variables represent different sources of information. However, the *maximum* variable is always selected in all the five selections, and also the variables of the group “maximumⁿ” are always present. It is worth noting that another important source of information for the estimation of volume is that associated with the “percentiles”. Figure 4-4 shows the relationship between the predicted vs. observed stem volume.

As the model derived from the variables of all returns is the one that provides the highest accuracy, we decided to use it in the cross-validation analysis. The results are shown in Table 4-6.

Concerning the results on the training set, they are quite similar to the ones presented in Table 4-5, whereas for the test set there is a slight decrease of the adj-R^2 (while the RMSE remains unchanged).

Table 4-5. Selected models for the estimation of tree stem volume considering all the variables extracted.

Initial variables set	Returns	RMSE	adj-R ²	N° var.	Variables in the final model
All the variables extracted	All	0.55	0.77	4	maximum of the 1 st return maximum of the 2 nd return at the power of 4.8 10 th percentile of the 1 st return 10 th percentile of the 3 rd return
	1 st	0.56	0.75	3	maximum standard deviation maximum at the power of 5
	2 nd	0.58	0.74	1	maximum at the power of 3.4
	3 rd	0.68	0.63	3	maximum maximum at the power of 1.1 maximum at the power of 5
	4 th	0.91	0.34	2	maximum maximum at the power of 5

Table 4-6. Results obtained with a 10-fold cross validation.

Variables in the final model	Training		Test	
	RMSE	adj-R ²	RMSE	adj-R ²
maximum of the 1 st return maximum of the 2 nd return at the power of 4.8 10 th percentile of the 1 st return 10 th percentile of the 3 rd return	0.55	0.76	0.55	0.71

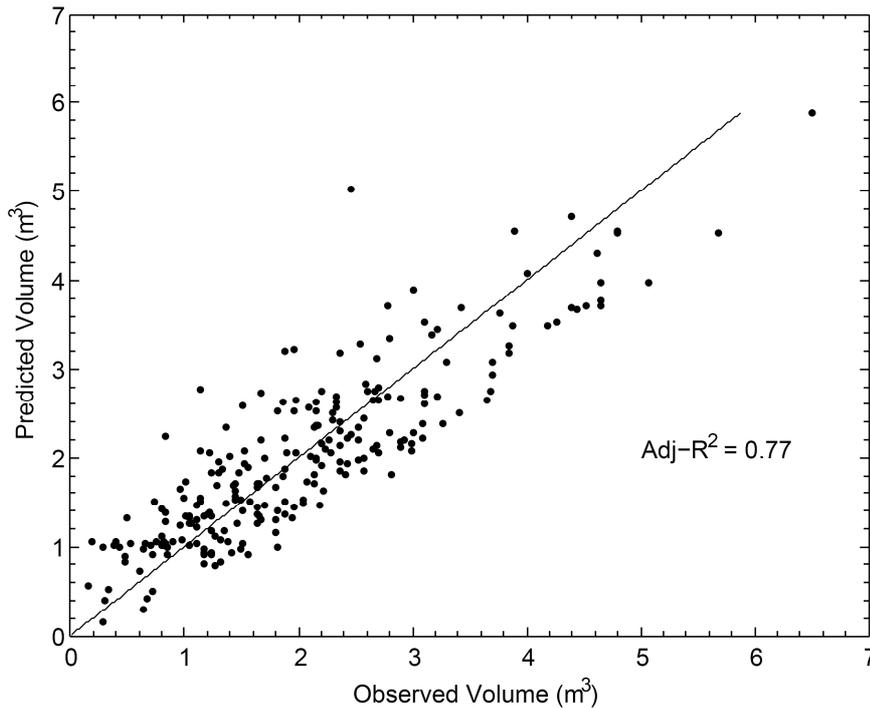


Figure 4-4. Observed volume vs. predicted volume for all the 243 trees of the ground truth.

4.4.5 Analysis on the relationship between the number of hits per return and the crown depth.

In this final analysis we examined if a relationship exists between the number of hits per return and the depth of the tree crowns. Figure 4-5 shows crown depth vs. the percentage of hits on the total for all the four returns considered. Twelve groups of crown depth were defined from 6 m to 28 m. Only trees with a height between 20 and 40 m were considered. From a theoretical viewpoint we expect that as much the crown is depth as high the possibility to have hits over the 1st return is. In greater detail, analyzing the specification of the sensor considered in this study we know that the minimum distance between the first and the second pulse is 2.1 m, and 3.8 m for any subsequent return [3]. This is confirmed from our analysis. In Figure 4-5 it is possible to see that there is a slight trend for which we have a reduction of the percentage of 1st return hits, in favor to the hits belonging to the other returns. In particular we move from an 81.5% of the 1st returns for the range between 6 and 8 m to 51.1% for the range from 26 to 28 m. Meanwhile we have an increase of the 2nd return (from 15.8% to 34%), of the 3rd (from 2.6% to 13%) and of the 4th one (from 0.1% to 1.9%).

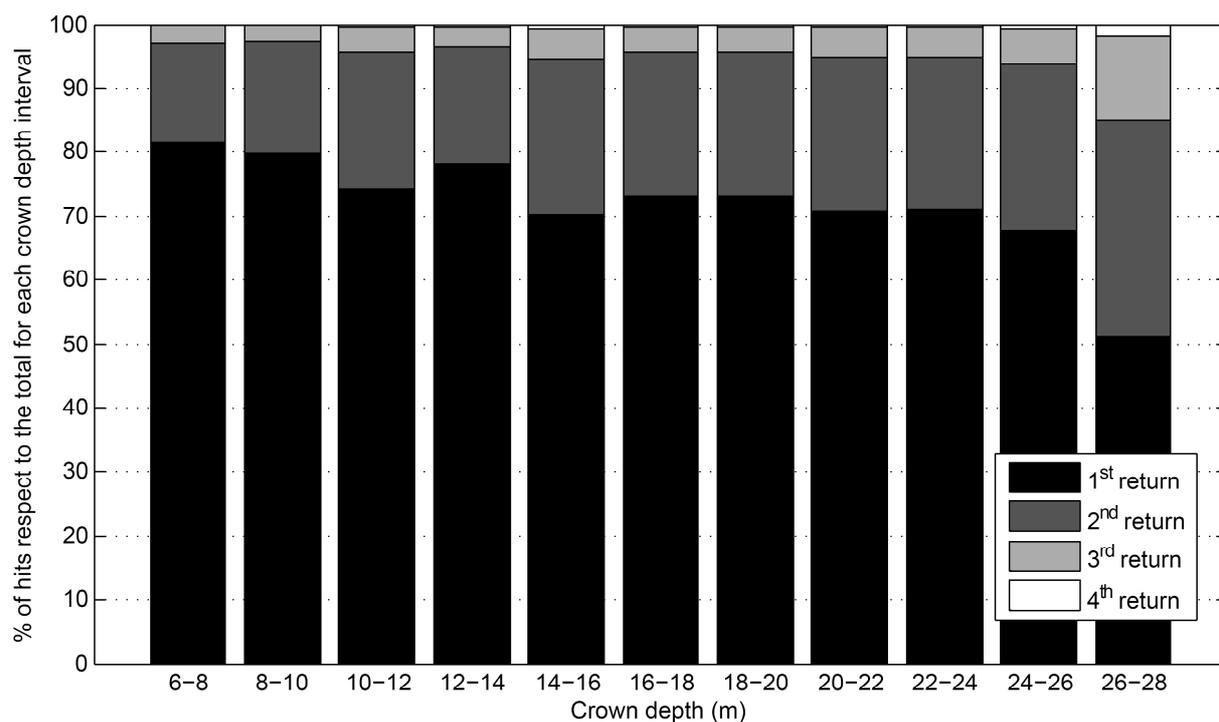


Figure 4-5. Crown depth vs. percentage of hits per return.

4.4.6 Discussion.

From these results it is possible to draw a number of conclusions on the use of LIDAR variables to predict individual stem volume and on the exploitation of information contained in the non-first returns.

In this study ground truth individual stem volume was estimated using an equation of the form $V = \beta D^\gamma H^\delta$ where V is the stem volume, D is the diameter at breast height (DBH), H is the height of the tree, and β , γ , δ are parameters dependent on the species, the geographical area, and the terrain characteristics.

This equation, explain the reason for which throughout the analysis the *maximum of the 1st return* is considered to be informative in the stem volume estimation. This variable is highly correlated with tree height, likewise other variables such as the percentiles over the 80th. The variables “maximum” in particular emerge to be highly correlated with the stem volume. This comes from the fact that in the computation of the volume the height of the tree at a certain power is used. This could be also a reason for the efficiency on how this kind of variable emphasizes the information content of the 2nd and 3rd return. In particular the maximum of the 2nd return at the power of 2.9 provides an adjusted R^2 of 0.74, while the maximum of the 3rd return moved from a correlation of 0.49 to 0.63 at the power of 2.7.

Concerning the percentiles, many studies in the literature used this kind of variables in the estimation phase (e.g, [13]). This is mainly due to the fact that high percentage percentiles usually represent better the tree height with respect to the “maximum” (the *maximum* could be an outlier), and that the percentiles around the 50th could be used as a measure of crowns density. We can expect a connection between the

density of the crown and the tree stem volume, and in particular trees with a higher crown density have a higher stem volume.

Moreover, from our analysis it is clear that the returns different from the 1st are informative in the estimation of the tree stem volume. In greater detail the 2nd return provides good results comparable to those obtained with the 1st return. Also in this case the information contained in the 2nd return can be related to the crown density, and thus to the volume. The same consideration holds for the variables of the 3rd return.

Concerning the variables descriptive of the tree crown (e.g., the *radius* and the *area* of the crown), they are correlated with the stem volume, as confirmed from some literature studies (e.g., [7]).

It is worth noting that the combined use of variables belonging to different returns allows one to increase the estimation accuracy. In all the models developed starting from ensembles of variables belonging to different returns, the stepwise selection included variables extracted from almost all the returns. In particular in the final model used we have variables belonging to the 1st, the 2nd, and the 3rd return.

4.5 Conclusions

In this chapter we have presented an analysis on the effectiveness of the use of multireturn LIDAR data in the estimation of tree stem volume at individual tree level. We have studied a multireturn LIDAR data set characterized by four returns. We have also analyzed different kinds of variables extracted from the different returns, deriving some interesting conclusions:

1. the use of variables belonging to all the returns allows one to obtain an increase of the estimation accuracy. In our particular case, the final best model is based on variables extracted from the 1st, the 2nd, and the 3rd returns;
2. the variables “maximumⁿ” allow one to emphasize the information contained in all the returns, and in particular to obtain good correlations only using the 2nd or the 3rd returns;
3. there exists a correlation between the crown depth and the number of hits per return; in greater detail increasing the crown depth the probability to have returns different from the 1st increases.

As future developments of this work we plan to: i) analyze the effectiveness of different kinds of variable-selection techniques; ii) study other kinds of non-linear estimators (e.g. Support Vector Regression); iii) investigate the interaction of LIDAR data with other sources of information (e.g. multispectral and hyperspectral remote sensing images); iv) analyze the effects of the undetected crowns (e.g. in multilayer forests) on the estimation of the of stem volume in forest inventories; v) study the possibility to identify information on the dominated layers from the analysis of different LIDAR returns in multilayer forests.

4.6 Acknowledgments

We would like to thank Sergio Tonolli for the field work and for the preparation of the ground truth data. This work was partially supported by the CARBOITALY project funded by the FISIR program of the Italian Ministry of University and Research.

4.7 References

- [1] Y. Malhi, P. Meir, S. Brown, “Forests, carbon and global climate,” *Philosophical Transactions of the Royal Society of London*. vol. 360, no. 1797, pp. 1567– 1591.
- [2] Z. J. Bortolot, and R. H. Wynne, “Estimating forest biomass using small footprint LIDAR data: An individual tree-based approach that incorporates training data”, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, pp. 342-360, 2005.
- [3] E. Naesset, “Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data”, *Remote Sensing of Environment*, vol. 113, pp. 148-159, 2009.
- [4] N. C. Coops, T. Hilker, M. A. Wulder, B. St-Onge, G. Newnham, A. Siggins, and J.A. Trofymow, “Estimating canopy structure of Douglas-fir forest stands from discrete-return LIDAR,” *Trees*, vol. 21, pp. 295-310, 2007.
- [5] G. Patenaude, R. A. Hill, R. Milne, D. L. A. Gaveau, B. B. J. Briggs, and T. P. Dawson, “Quantifying forest above ground carbon content using LIDAR remote sensing,” *Remote Sensing of Environment*, vol. 93, pp. 368-380, 2004.
- [6] P. Corona, G. Chierici, A. Lamonaca, D. Travaglini, F. Mason, E. Minari, M. Marchetti, and A. Montagni, “LIDAR-supported estimation of growing stock of broadleaved forests”, *Forestsat 2007 (Forest and Remote Sensing: Methods and Operational Tools)*, Montpellier 2007.
- [7] S. C. Popescu, R. H. Wynne, and R. F. Nelson, “Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass”, *Canadian Journal of remote sensing*, vol. 29, no. 5, pp. 564-577, 2003.
- [8] S. C. Popescu, R. H. Wynne, and R. E. Nelson, “A simplified method of predicting percent volume in log portions,” *Computers and Electronics in Agriculture*, vol. 37, no. 1-3, pp. 71-95.
- [9] J. Hyypä, O. Kelle, M. Lehikoinen, and M. Inkinen, “A Segmentation-Based Method to Retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 5, 2001.
- [10] Y. Wang , H. Weinacker and B. Koch, “A Lidar Point Cloud Based Procedure for Vertical Canopy Structure Analysis And 3D Single Tree Modelling in Forest,” *Sensors*, vol. 8, pp. 3938-3951, 2008.
- [11] M. J. Falkowski, A. M.S. Smith, A. T. Hudak, P. E. Gessler, L. A. Vierling, and N. L. Crookston, “Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data,” *Canadian Journal of Remote Sensing*, vol. 32, no. 2, pp. 153-161, 2006.
- [12] M. Heurich, “Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park,” *Forest Ecology and Management*, vol. 255, pp. 2416-2433, 2008.
- [13] E. Naesset, “Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data,” *Remote Sensing of Environment*, vol. 80, pp. 88-89, 2002.
- [14] A. C. Rencher, *Methods of multivariate analysis*, Wiley-IEEE, 2003.

CHAPTER 4

Chapter 5

5 A system for the estimation of single tree stem diameters and volume using multireturn LIDAR data

Forest inventories are important tools for the management of forests. In this context the estimation of the tree stem volume is a key issue. In this chapter we present a system for the estimation of forest stem diameters and volume at individual tree level, which is based on multireturn LIDAR data and on Support Vector Regression (SVR). The system proposed is made up of a preprocessing module, a LIDAR segmentation algorithm (aimed at retrieving tree crowns), a variable extraction and selection procedure and an estimation procedure. The variables derived from LIDAR data are computed from both the intensity and elevation channels of all available returns. Three different methods of variable selection are analyzed, and the sets of variables obtained are used in the estimation phase based on a multivariate linear regressor and a Support Vector Regression (SVR) technique. The stem volume is estimated with two approaches: i) estimation from the LIDAR variables; and ii) estimation obtained by combining the diameters and heights estimated from LIDAR variables with the species information derived from a classification map according to standard height/diameter relationships. Experimental results show that the system proposed is effective and provides good results in both the stem volume and diameter estimation. Moreover it provides useful information on the use of SVR in these kinds of problems.

5.1 Introduction

In the last years forest management and protection have become a very important task, having many implications in many different fields, from the economical one (e.g. estimation of timber volume for commercial usage) to the environmental one (e.g., protection of biological diversity). Important tools for forest management are forest inventories. These procedures are used to measure and estimate the most important attributes of a forest, like the species composition, the tree stem heights, diameters at breast height and volume, the age, the health, etc. Among these attributes one of the most important is the tree stem volume. In fact, as an example, the knowledge of this parameter drives the economical exploitation of a forest: a low value of stem volume per hectare means that probably only little trees are present in that area and thus it is not a relevant site for timber exploitation. From the environmental viewpoint the knowledge of the stem volume is important as it is related to the carbon stored by the forest analyzed. Carbon

This chapter has been submitted to *IEEE Transactions on Geoscience and Remote Sensing* with the title: "A system for the estimation of single tree stem diameters and volume using multireturn LIDAR data". Authors: Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle.

stocks are important in the context of the Kyoto protocol. This protocol states that each nation has to reduce the CO₂ emissions under a certain threshold and this threshold is computed taking into account also the carbon stocks of each country. Thus, it is of primary importance to have a detailed knowledge of forest stem volume at national and regional level in order to have a precise computation of this amount. In this context, remote sensing is a very useful technology for a precise and objective analysis of forest areas.

Nowadays, different kinds of remote sensing sensors exist, with different characteristics and peculiarities. In the literature it is possible to find studies that analyze the estimation of stem volume with active and passive sensors. Concerning passive sensors, we can find data with different spectral and spatial resolutions. Some studies exploited low spectral and spatial resolution data for the estimation of stem volume, like the study of Marsden *et al.* in [1]. This study analyzes the relationship between NDVI time-series extracted from MODIS data and stand structural characteristics (volume, dominant height, mean annual increment) in Eucalyptus plantations finding good agreements. In [2] Muukkonen *et al.* use MODIS data to estimate stem volumes in Finnish forests. The results that they obtain are significant, as the difference between the volumes of the national forest inventories and their estimation differs only of 3.6%. These kinds of data allow one to make raw estimations of stem volumes and are effective when the area analyzed is wide and uniformly characterized by the same tree species. If this condition is not satisfied, due to the low spatial (and sometimes spectral) resolution, they do not result in precise analysis of complex forest areas. In this case, it is better to use high spatial resolution sensors that result in a more detailed analysis. In the literature several studies exploit such kind of data. In [3], Hall *et al.* use Landsat ETM+ data to estimate forest volume in Canada, obtaining good accuracies. They find that Landsat derived forest volumes are statistically moderately correlated to the inventory-derived volumes with values of adjusted R² of 0.63, 0.68, and 0.70 for conifer, deciduous, and mixed species, respectively. Also Luther *et al.* [4] estimate Canada forest volumes using ETM+ data, while in [5] Muukkonen *et al.* use ASTER images for the same purpose in Finland, obtaining predictions significantly close to the municipality-level mean values provided by the National Forest Inventory of Finland.

Concerning active remote sensing sensors, in the literature several studies have been presented that exploit Synthetic Aperture RADAR (SAR) data. In [6], Wang *et al.* exploit high resolution polarimetric SAR data to estimate volume of Tomakomai forests in Hokkaido (Japan). Experimental results on ground-truth data collected in 2005 show an accuracy of approximately 86% with a correlation coefficient of 0.91. In [8], Quiñones *et al.* analyze the limits of the use of SAR data in the estimation of forest volumes, while in [9] Mette *et al.* use polarimetric SAR interferometry to estimate volumes, analyzing the effects of parameters tuning in the final result.

A relatively recent technology that has demonstrated to be effective in precise estimations of forest volumes is laser scanning. Light Detection and Ranging (LIDAR) sensors allow one to acquire precise measures of both tree height and structure. These sensors can be classified according to the dimension of the footprint, and according to the system used to record the data (full waveform or discrete return). Con-

cerning the estimation of forest volumes, the most widely used systems are discrete return sensors. In this context, in the literature it is possible to find studies that exploit both large and small footprint LIDAR data. In [9] Drake *et al.* use large footprint LIDAR data on a neotropical rainforest, finding high correlations (R^2 up to 0.94) between LIDAR metrics and aboveground biomass. Small-footprint LIDAR data results in a more detailed analysis, in particular at tree level. This is a very challenging approach as it allows one to estimate the stem volume value for each tree present in the area under analysis. In this context, in [11] Hyypä *et al.* estimate single trees volume using small footprint LIDAR data with a segmentation based method, achieving good results (a standard error of 10%). Bortolot *et al.* [10] use an individual tree-based approach to estimate forest volumes using small footprint LIDAR data, obtaining a correlation coefficient (R) ranging between 0.59 and 0.82.

At present the sensor that potentially seems to be most promising for the estimation of forest volumes is small footprint LIDAR. This sensor allows one to have detailed and precise analysis at local level, in particular at individual tree level. Moreover, high density LIDAR data results in very detailed analysis of the structure of the trees. However, a drawback of high density LIDAR data is that they should be taken by airplane (or helicopter); thus acquiring high density measures is quite expensive. In this context, from an analysis of the literature, it emerges that: i) only few papers exploited multireturn LIDAR data in the estimation of stem volume at single tree level; and ii) no studies exist that exploited advanced machine learning techniques, like Support Vector Regression (SVR), for the estimation of tree stem volume starting from LIDAR data.

The goals of this chapter are: i) to propose a system that, starting from small footprint multireturn LIDAR data, exploits the SVR technique to derive both tree diameters and volume; ii) to compare different variable selection techniques; iii) to compare the SVR technique (with two different kernel functions) with the standard multivariate linear estimator; and iv) to compare different approaches in stem volume estimation. To reach these goals, we adopt a segmentation-based method that identifies single tree crowns from LIDAR data, extracts and selects the most effective variables, and at the end estimates tree diameters and volume. This is accomplished according to an architecture made up of the following modules: i) preprocessing, ii) variable extraction, iii) variable selection, and iv) estimation.

The rest of the chapter is organized as follows. In Section 5.2 materials and methods are presented. At first the data set used is described (Section 5.2.1), and then Section 5.2.2 presents the architecture of the system adopted and the data preprocessing operations. The segmentation algorithm is illustrated in section 5.2.3, while section 5.2.4 and 5.2.5 present the techniques of variables extraction and selection. The SVR estimator technique is described in section 5.2.6. Section 5.3 reports experimental results. Finally, conclusions are drawn in section 5.4.

5.2 Material and methods

5.2.1 Data set description

The study area analyzed in this chapter is a forest site in the Italian Alps located at Lavarone (near the city of Trento) in the Trentino province (Italy). The central point of the area has the following coordinates: 45° 57' 30.09" N, 11° 16' 25.17" E. The topography of this area is complex: it includes hill sides of different inclinations with an altitude that ranges from 1200 to 1600 meters on the sea level. The area has a size of approximately 495 ha. This site represents a typical example of Alpine forest with the presence of three main species (Norway Spruce, Silver Fir and European Beech) and some other species like European Larch and Scots Pine.



Figure 5-1. Digital Canopy Model of the investigated area.

Table 5-1. Distribution of ground truth points in the training, test and validation sets. The species composition of each set and the values of height, Diameter at Breast Height (DBH) and stem volume are also presented (N= total number of samples).

	N	%	Height (m)		DBH (cm)		Stem Volume (m ³)	
			Mean	Range	Mean	Range	Mean	Range
Training	174	100	26.2	7.5 – 38.1	45.2	9 – 90	2.34	0.04 – 10.93
<i>Silver Fir</i>	74	42.5	26.3	7.5 – 38.0	45.8	13 – 90	2.37	0.05 – 10.93
<i>Norway Spruce</i>	79	45.4	28.1	15.4 – 37.7	49.0	25 – 74	2.68	0.37 – 7.12
<i>Other specie</i>	21	12.1	19.6	11.6 – 28.8	29.7	9 – 63	0.95	0.04 – 3.74
Test	147	100	25.8	11.1 – 37.0	45.0	13 – 78	2.27	0.08 – 7.67
<i>Silver Fir</i>	71	48.3	25.8	12.9 – 36.8	44.2	19 – 73	2.20	0.20 – 6.11
<i>Norway Spruce</i>	59	40.1	27.4	15.9 – 37.0	48.6	21 – 78	2.66	0.31 – 7.67
<i>Other specie</i>	17	11.6	20.6	11.1 – 29.2	36.3	13 – 60	1.21	0.08 – 3.51
Validation	160	100	26.44	9.4 – 38.1	45.2	12 – 74	2.30	0.05 – 7.21
<i>Silver Fir</i>	67	41.9	26.1	12.4 – 35.0	43.7	16 – 71	2.17	0.19 – 6.32
<i>Norway Spruce</i>	79	49.4	28.3	9.4 – 38.1	49.6	9.4 – 38.1	2.72	0.05 – 7.21
<i>Other specie</i>	14	8.7	17.6	12.4 – 24.5	27.7	14 – 44	0.53	0.10 – 1.47

The LIDAR data were acquired on September 4th, 2007, between 11:29 AM and 12:07 AM. These data were taken by a sensor Optech ALTM 3100EA, with a mean density of 8.6 points per square meter for the first return. The laser pulse wavelength and the laser repetition rate were 1064 nm and 100 kHz, respectively. The number of recorded returns for each laser pulse is up to four.

Ground truth data were collected in summer 2007. We collected 481 points (trees) distributed in 50 sample sites randomly selected across the investigated area. These points were then divided into three sets: training, test and validation sets. Table 5-1 shows the distribution of the points in these sets, the species composition, the values of height, the Diameter at Breast Height (DBH) and the stem volume.

5.2.2 Data preprocessing and architecture of the system

Figure 5-2 presents the architecture of the system proposed in this chapter. As described in the introduction, the goal of the system is to obtain a map of tree stem volume by integrating the information provided by both multireturn LIDAR data and a classification map.

Regarding the preprocessing phase we have two steps: i) rasterization of the raw LIDAR data, and ii) subtraction of the Digital Terrain Model (DTM) to the elevation information of the LIDAR data. The rasterization was performed with a ground resolution of 1 m. The average values of the different returns of the LIDAR pulses included in a square meter were assigned to each pixel. Concerning the first return, pixels with no value were interpolated with the nearest neighbor technique, while for returns different from the first we left value 0. The rasterization was performed for each recorded return and for both the elevation and intensity values. After that, the Digital Terrain Model (DTM) of the area considered was subtracted from the elevation image of each return. This allowed us to correct the raw LIDAR elevations from the topography of the scene. In particular subtracting the DTM from the elevation of the first LIDAR return we obtained the digital Canopy Height Model (CHM) (see Figure 5-1).

The CHM was used in the segmentation phase (see section 5.2.3), and then the LIDAR bands plus the segmented image were used in the variable extraction phase (see section 5.2.4). After that, the most significant variables were selected (see section 5.2.5) and used as input to the SVR algorithm (see section 5.2.6). In the following sections the main blocks of the system are detailed.

In the proposed architecture hyperspectral data are used for the identification of the tree species of each tree. The classification map obtained from the classification of these data is aggregated at crown level in order to have an information on the species for each crown. In the rest of this study we focus our attention only on the estimation part. We refer the reader to [19] for more details on the classification architecture adopted for the hyperspectral images.

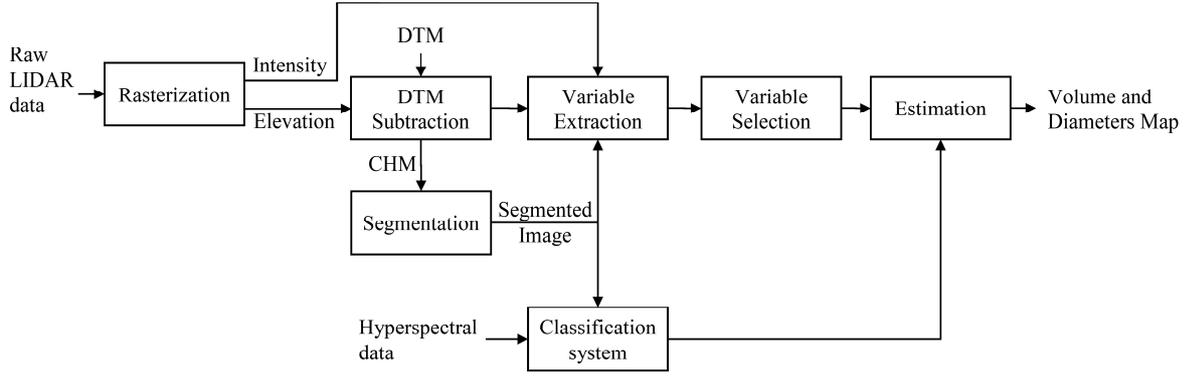


Figure 5-2. Architecture of the proposed system.

5.2.3 Segmentation

An important phase in the proposed system is the segmentation. This phase drives the next steps of variable extraction and diameter/volume estimation. The rationale of this step is to identify and delineate individual tree crowns. The segmentation algorithm used in this chapter is derived from the algorithm presented by Hyypä *et al.* in [10]. This algorithm, is divided into three main steps: i) prefiltering; ii) seed point extraction; iii) seed region growing.

According to [10], in the prefiltering phase the CHM was filtered with a convolution filter for emphasizing local maxima and tree crowns. The coefficient of the filter using a 3x3 window are defined as follows:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} / 16 \quad (5.1)$$

After this phase it is necessary to identify the seed points, corresponding to the tree tops, from which the region growing procedure start. In order to consider only trees higher than a given value, seeds are defined as the local maxima higher than a certain threshold $thSeed$. A mobile window of a given size (defined by the user) is used to detect them. As an example, if we consider a window of size 3x3, the pixel with coordinates (i, j) of the image I is a seed if:

$$\begin{cases} I(i, j) = \arg \max_{\substack{x \in [i-1; i+1] \\ y \in [j-1; j+1]}} [I(x, y)] \\ I(i, j) > thSeed \end{cases} \quad (5.2)$$

where $I(i, j)$ is the elevation value of the pixel of coordinates (i, j) . At the end of this process we obtain the set of the seed points $S = \{s_1, \dots, s_N\}$, where s_n identifies the n -th seed point. Figure 5-3 shows an example of seed points extracted from a prefiltered CHM image.

The last phase consists in the seed region growing and it is aimed at the identification of the crowns of the trees. Seed region growing starts from each seed and grows iteratively the region from the first order neighborhood system to the n -th. A pixel $I(i, j)$ is added to the considered region if it satisfies two condi-

tions that take into account both the dimension and the shape of the crown. If we define the set of the regions $R = \{r_1, \dots, r_N\}$, where r_n identifies the region around the seed point s_n , we can write as follows:

$$I(i, j) \in r_n \text{ if } \begin{cases} I(i, j) > P * I_{s_n} \\ D[r_n + I(i, j)] < thDiameter \end{cases} \quad (5.3)$$

where I_{s_n} is the height of the considered seed point, $P \in (0, 1]$, $D[r_n + I(i, j)]$ is the diameter of the considered region including the new pixel $I(i, j)$, and $thDiameter$ is the maximum acceptable diameter of a region.

The algorithm stops when no pixels are added to any region.

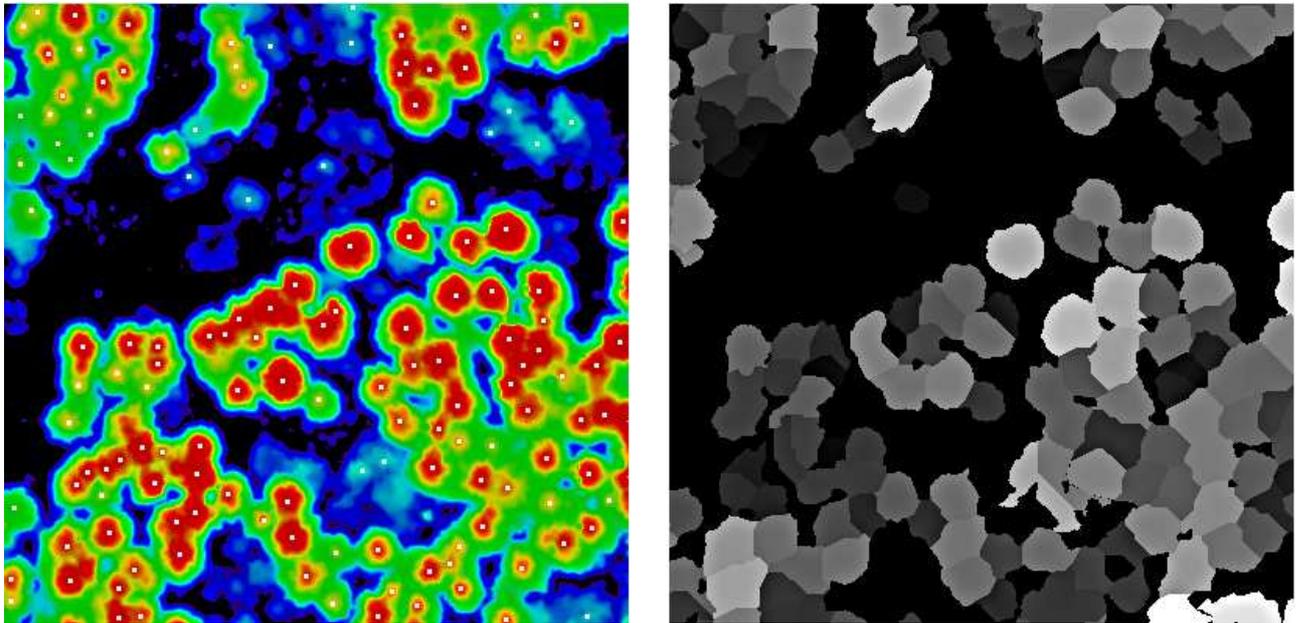


Figure 5-3. Example of i) seed points extracted from the elevation of the first LIDAR return, and ii) regions associated with the crowns.

5.2.4 Variable definition

In order to make a detailed characterization of each tree, a series of variables were defined and extracted from each segmented region from the elevation and intensity channels of the four returns available in our dataset. These variables describe the structure and the characteristics of the trees. Table 5-2 reports a summary of the variable extracted. As it is possible to see, we extracted some variables connected with LIDAR points distribution among the crown (e.g., mean, maximum, minimum, etc.) and other variables connected with the crown geometry (e.g., area, and radius). The area variable for the first LIDAR return represents the surface of the region in pixels, while for the other returns represents the number of pixels for which a return is present. Concerning the two radius variables, *radius 1* is the radius of the circle with area equal to the area of the region, while *radius 2* is computed as $(a_x + a_y)/4$, where a_x is the length of the region along the x axis and a_y is the length of the region along the y axis.

CHAPTER 5

Each variable extracted is related to a physical characteristic of the tree. We define five different groups of variables: i) tree height; ii) crown horizontal shape; iii) crown vertical shape; iv) crown internal structure; and v) species. It is worth noting that we do not know a priori which the most effective variable for each characteristic is. In general, each variable can be related to more than one characteristic. In Table 5-3 variables are divided according to their expected physical meaning.

Table 5-2. Variables extracted from each segmented region.

Return	Variable	Return	Variable
1 st	maximum	3 rd	maximum
	minimum		minimum
	mean		mean
	range		range
	variance		variance
	area		area
	mean intensity		mean intensity
2 nd	maximum	4 th	maximum
	minimum		minimum
	mean		mean
	range		range
	variance		variance
	area		area
	mean intensity		mean intensity
	mean 1 st – mean 2 nd		max 1 st – min 3 rd
	mean 1 st – mean 3 rd		radius 1
	mean 1 st – mean 4 th		radius 2
	mean 2 nd – mean 3 rd		
	mean 2 nd – mean 4 th		
	mean 3 rd – mean 4 th		

Table 5-3. Physical meaning attributed to each variable.

Physical characteristic of the tree	Variables	
Tree Height	1 st maximum	1 st minimum
	1 st mean	
Crown Horizontal Shape	1 st area	radius 2
	radius 1	
Crown Vertical Shape	1 st variance	mean 1 st – mean 3 rd
	1 st range	mean 1 st – mean 4 th
	mean 1 st – mean 2 nd	max 1 st – min 3 rd
Crown Internal Structure	2 nd maximum	4 th mean
	2 nd mean	4 th minimum
	2 nd minimum	4 th variance
	2 nd variance	4 th range
	2 nd range	4 th area
	2 nd area	mean 1 st – mean 2 nd
	3 rd maximum	mean 1 st – mean 3 rd
	3 rd mean	mean 1 st – mean 4 th
	3 rd minimum	mean 2 nd – mean 3 rd
	3 rd variance	mean 2 nd – mean 4 th
	3 rd range	mean 3 rd – mean 4 th
	3 rd area	max 1 st – min 3 rd
	4 th maximum	
Species	1 st intensity	3 rd intensity
	2 nd intensity	4 th intensity

5.2.5 Variable selection

In problems characterized by the presence of many potentially useful predicting variables an important phase is that of the variable selection. The importance of variable selection depends on many reasons, the most important of which are: i) the degradation of the generalization ability of the regression model when increasing the number of parameters to estimate; ii) the noise introduced by some variables; and iii) the high computational cost caused by a large number of input variables. Thus, variable selection becomes mandatory to improve the regression results both in terms of computational complexity and generalization ability. The goal of this selection is to find the smallest set of variables that provide estimates similar (or better) to those obtained with the whole set of available variables. It is worth noting that a small set of predicting variables results in a simple predicting model characterized by good generalization ability.

In the literature it is possible to find three main approaches to variable selection: i) the filter methods; ii) the wrapper methods; and iii) the embedded methods.

Filter methods perform the variable selection as a preprocessing step independently on the algorithm used for model construction. These methods are usually very simple and are based on a different principle with respect to that used in the final estimation process. This allows one to have a quick and general selection phase at the cost to select variables that are not explicitly optimized for the final model used. As an example, it is possible to rank the variables according to their correlation with the estimated parameter, and then to take arbitrary the first n variables. Another approach is to use a simple regression technique (e.g., a multivariate linear regression) to select the variables and then to exploit the variables chosen in the non-linear regression model [12].

Wrapper methods are different from the previous ones, as they are related to the algorithm used in the estimation process. The rationale of these methods is to select an optimized set of variables for a given estimation technique. Wrapper methods are sometimes considered “brute force” methods that require a very high computational load, but at the same time exhibit the advantage to select the set of variables that optimizes the final predictive model [13]. In their most general formulation, these methods consist in using the prediction performances of a given estimation technique to assess the usefulness of variables or subsets of them.

Similarly to the feature selection techniques used in the classification, filter and wrapper methods require a criterion to compare the performances of the different variable subsets (e.g., minimization of the mean square error on the training set), and a search strategy. Several search strategies exist in the literature (e.g., Genetic Algorithms, Sequential Forward Selection, Sequential Backward Elimination, etc) [12]. The two sub-optimal methods most commonly used for their simplicity are Sequential Forward Selection (SFS) and Sequential Backward Elimination (SBE).

Embedded methods incorporate the variable selection step in the training of the algorithm. This means that the variable selection is performed during the definition of the estimation model. Usually these methods require changes in the objective function of the algorithm considered, and thus they are specifically

developed for each predictor. In general they allow one to reach better accuracies with respect to the previous methods, even if they are more computationally demanding. Examples of these methods are decision trees (e.g., [14]) or the zero norm optimization (e.g., [15])

In this chapter we consider filter and wrapper methods. As performance evaluation of the considered variable set we used the minimization of the Mean Square Error (MSE) on the test set. Concerning the search strategy, we adopted a simple and fast Sequential Forward Selection (SFS) strategy for limiting the computational complexity.

The Sequential Forward Selection (SFS) is a suboptimal search strategy that has been used in many previous studies in the literature. It can be summarized in the following steps:

1. *Initialization*: the variable v^* that provides the model with the lowest MSE on the test set is selected and added to the empty set $V^{(i=0)}$. i is incremented by 1.
2. *Insertion*: the variable v^* that added to $V^{(i)}$ results in the model with the lowest MSE on the test set is selected and added to $V^{(i)}$, i.e. $V^{(i+1)} = V^{(i)} + v^*$.
3. *Convergence*: if $i+1 = M$ (where M is the total number of desired variables), then stop; otherwise set $i = i+1$ and return to step 2.

The set of variables that provides the best trade-off between the number of variables M and the MSE is selected. In particular, we give the priority to variable sets characterized by a low dimensionality, in order to obtain a model with a high generalization capability. The set of variables selected is then used as input to the considered estimator.

The training set is used to define the regression model; then, the test set is used in the variable selection and model selection phases. The validation set is used to evaluate the final performances of the system.

This method can be considered either a filter or a wrapper depending on the estimator used for the selection and for the subsequent estimation.

5.2.6 Support Vector Regression

In this section we briefly summarize the main principles of the non linear and multivariate ε -Insensitive Support Vector Regression (ε -SVR) algorithm used in our estimation system.

Let $\{(\mathbf{x}_i, y_i), i = 1, \dots, T\}$ be a training set, where $\mathbf{x}_i \in \mathfrak{R}^d$ is the d -dimensional vector of selected input variables, $y_i \in \mathfrak{R}$ is the target tree attribute to be estimated, and T is the number of training samples. The rationale of the SVR is to map the original variable space into a higher dimensional space $\Phi(\mathbf{x})$ using a non linear transformation function Φ , and to find a linear regression function $f(\mathbf{x})$ in this new space, as:

$$f(\mathbf{x}) = \langle w \cdot \Phi(\mathbf{x}) \rangle + b \quad (5.4)$$

where $w \in \mathfrak{R}^d$ is the weight vector, $b \in \mathfrak{R}$ is the bias, and $\langle w \cdot \Phi(\mathbf{x}) \rangle$ represents the dot product between w and $\Phi(\mathbf{x})$. This function should have at most deviation ε from the real targets y_i for all the training samples and, at the same time, should be as flat as possible. In other words, we neglect errors smaller than ε whereas we penalize errors larger than ε (ε -insensitive tube).

The optimal function $f(\mathbf{x})$ can be obtained solving the following constrained minimization problem:

$$\begin{cases} \min_{w, b, \xi} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^T (\xi_i + \xi_i^*) \right\} \\ y_i - (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i, \quad \forall i = 1, \dots, T \\ (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i^*, \quad \forall i = 1, \dots, T \\ \xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, T \end{cases} \quad (5.5)$$

The variables ξ_i, ξ_i^* are called slack variables and are used to consider the patterns outside of the ε -insensitive tube. Their values depend on the kind of penalization function adopted: linear or quadratic (see Figure 5-4 for an example of linear penalization function). C represents a regularization constant that should be tuned in the model selection phase in order to reach the best trade-off between the smoothness of the function $f(\mathbf{x})$ and the tolerance to the errors (due to the patterns outside the ε -insensitive tube).

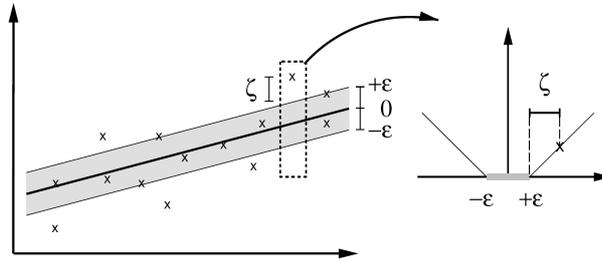


Figure 5-4. ε -insensitive tube of a linear SVR with a linear penalization function (figure source: [16]).

This minimization problem can be rewritten in a dual formulation and solved according to standard methods of quadratic programming based on the Lagrange multipliers. In the case of a linear penalty function, we obtain the following Lagrangian function:

$$L(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \sum_{i=1}^T y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^T (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^T \sum_{j=1}^T (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) \quad (5.6)$$

where $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \rangle$ is called kernel function. This function satisfies the Mercer theorem [17] and allows us to replace the dot product among the transformation functions $\Phi(\cdot)$. This is very important as the kernel functions avoid the need to know explicitly the transformation functions $\Phi(\cdot)$. Common examples of kernel functions are:

1. *Linear kernel function*: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$.
2. *Gaussian kernel function*: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-|\mathbf{x}_i - \mathbf{x}_j|^2 / 2\sigma^2\right)$, where σ^2 is the spread of the Gaussian.

Thus, the minimization problem in the dual formulation becomes:

$$\begin{cases} \max_{\alpha} \{L(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*)\} \\ \sum_{i=1}^T (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C, \quad \forall i = 1, \dots, T \\ 0 \leq \alpha_i^* \leq C, \quad \forall i = 1, \dots, T \end{cases} \quad (5.7)$$

The final estimation function in the original variable space is given by:

$$f(\mathbf{x}) = \sum_{i \in U} (\alpha_i - \alpha_i^*) y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (5.8)$$

The set U represents the set of training patterns corresponding to the Lagrangian multipliers different from zero. Only these patterns, defined as *support vectors* (SVs), affect the definition of the estimation function. The Lagrangian multipliers α_i and α_i^* (with $i=1, \dots, T$) allow us to define the contribution that each SV gives to the estimation function. From a geometrical viewpoint the SVs are the training patterns outside of the ε -insensitive tube.

5.3 Experimental results

5.3.1 Design of experiments

In order to verify the effectiveness of the proposed architecture and to achieve the goals of this chapter we define two main experiments. In the first one we estimate the stem diameters with the proposed system and then combine them with both the stem height (directly measured by the LIDAR) and the species information (derived from a classification map) in order to retrieve the stem volume. In the second experiment, we estimate directly the stem volume with the proposed system starting from the LIDAR variables. For both the experiments we investigated three different variable selection methods (SFS with multivariate linear estimator, SFS with SVR estimator with linear kernel function, and SFS with SVR estimator with RBF kernel function) and different estimators (multivariate linear, SVR with linear and RBF kernel functions).

In the learning of the SVR algorithm we performed a grid search for the value of the parameters ε , C and γ of the kernel function. The values for the grid search of ε and C were selected on the basis of the following empirical equations [18]:

$$C = \max\left(\left|\bar{y} + 3\rho_y\right|, \left|\bar{y} - 3\rho_y\right|\right) \quad (5.9)$$

$$\varepsilon = 3\rho_y \sqrt{\frac{\log(T)}{T}} \quad (5.10)$$

where \bar{y} is the mean value of the targets of the training set, ρ_y is their standard deviation, and T the size of the training set.

As mentioned in the data description section, we divided the available ground truth in three sets: training (174 points), test (147 points) and validation (160 points) sets. The training set was used for the variable selection and the learning of the estimation algorithm, the test set was used for the model selection, and the validation set for the final estimation of performance.

5.3.2 Results

Table 5-4 presents the results obtained for the estimation of stem diameters using different variable subsets and different estimation algorithms. From these results it is possible to derive important indications. Firstly, let us analyze the selection methods and the selected variables. Comparing the selection results obtained by the different methods, one can observe that the two selections based on linear models (multivariate linear and SVR-linear) have four variables in common. This points out that the use of a slow selection procedure like the one that exploits the SVR does not necessarily provide significantly different sets of variables compared to the faster technique based on a multivariate linear estimator. Moreover, we can observe that half of the variables selected are extracted from the first return, and the other half is computed on the other returns. This confirms that the first return is the most informative measure for the estimation of stem diameter, but that also the other returns significantly contribute to this task.

Analyzing in deeper detail the selected variables, one can note that the variables “1st maximum” and “1st mean” were selected by all the three algorithms. These two variables and the variable “1st minimum” are connected with the tree height which is strongly related to the stem diameter. There is then a group of variables that is connected with the crown internal structure (“3rd variance”, “2nd mean”, “2nd range”, “4th maximum”, and “4th mean”), a variable that can be linked to the crown vertical shape (“1st mean – 2nd mean”), and one to the species (“1st intensity”). All these variables model some characteristics of the tree that are connected with the diameter: the height, the species and the crown properties. Many of the selected variables can be linked to the crown internal structure. This is due to the fact that there can be a link between the crown internal structure and both the health of the tree and the species, two factors that affect the stem diameter.

Regarding the estimation algorithms, SVR-RBF provided the highest accuracy even if the difference with the other estimators is small. In particular, all the three estimators provided similar results with all the three variable sets considered. This is probably due to the fact that the variables considered are linearly related to the diameter and thus a simple multivariate linear estimator is enough to obtain good results. From the results one can also see that each estimator provided the highest accuracy when the selection was performed on the basis of the same estimator.

Concerning the estimation errors, it is worth noting that the MAE is very low, about 6 cm on an average value of the diameters around 45 cm. Moreover the estimation is unbiased as the mean error is almost zero in all the configurations considered.

A first conclusion that is possible to draw from these results is that with the considered variables it is not necessary to use complex and non linear techniques for the estimation of stem diameters. In particular, by using a simple linear estimator in both the phases of selection and estimation it is possible to obtain results comparable to that provided by SVR. However, the use of a linear estimator allows us to have a lower computational cost compared to SVR.

Figure 5-5 shows the distribution of the measured diameters vs. the estimated ones obtained by using the SVR-RBF technique for both variable selection and estimation for the test and validation sets. The correlation between estimated and measured diameters is good even if not excellent. The R^2 on the test and validation samples considered together is of about 0.63.

Table 5-4. Mean absolute error (MAE), mean square error (MSE) and coefficient of variation (R^2) of the estimates obtained on the test and validation sets using different variable sets and estimators.

Variables selected	Selection method	Estimator	Test			Validation		
			MAE	MSE	R^2	MAE	MSE	R^2
1 st maximum	SFS with linear estimator	linear	6.34	68.99	0.621	7.16	78.85	0.589
1 st mean		SVR-linear	6.29	69.03	0.625	7.00	75.44	0.608
1 st intensity		SVR-RBF	6.30	69.24	0.621	6.98	77.90	0.595
1 st mean – 2 nd mean	SFS with SVR-linear estimator	linear	6.42	69.77	0.617	7.28	79.71	0.585
3 rd variance		SVR-linear	6.20	67.78	0.630	7.15	78.24	0.594
2 nd range		SVR-RBF	6.35	67.52	0.631	7.13	79.09	0.588
1 st maximum	SFS with SVR-RBF estimator	linear	6.41	70.17	0.615	7.20	79.45	0.586
1 st mean		SVR-linear	6.28	68.86	0.625	7.12	78.97	0.591
1 st minimum		SVR-RBF	6.20	67.70	0.630	7.17	78.19	0.593
4 th maximum								
4 th mean								

In order to estimate the tree stem volume we used the diameters derived by the selection and estimation based on SVR-RBF, and the tree heights provided by LIDAR measurements. The estimation was carried out by using standard height/diameter relationships adopted for the estimation of the tree stem volume in forest inventories. These equations estimate the volume combining the tree diameter, the tree height and the species information. We considered as height of a tree the variable "maximum of the 1st return", as diameter the one estimated with the SVR-RBF, and as species that extracted from a classification map of the considered area. Table 5-5 and Table 5-6 show the results obtained on the stem volume estimation. As it is possible to see, the results in terms of MAE, MSE and R^2 are good, and in particular they are much better compared to the ones of the diameters. In terms of total volume it is worth noting that there is a underestimation of the volume for both the test and validation sets.

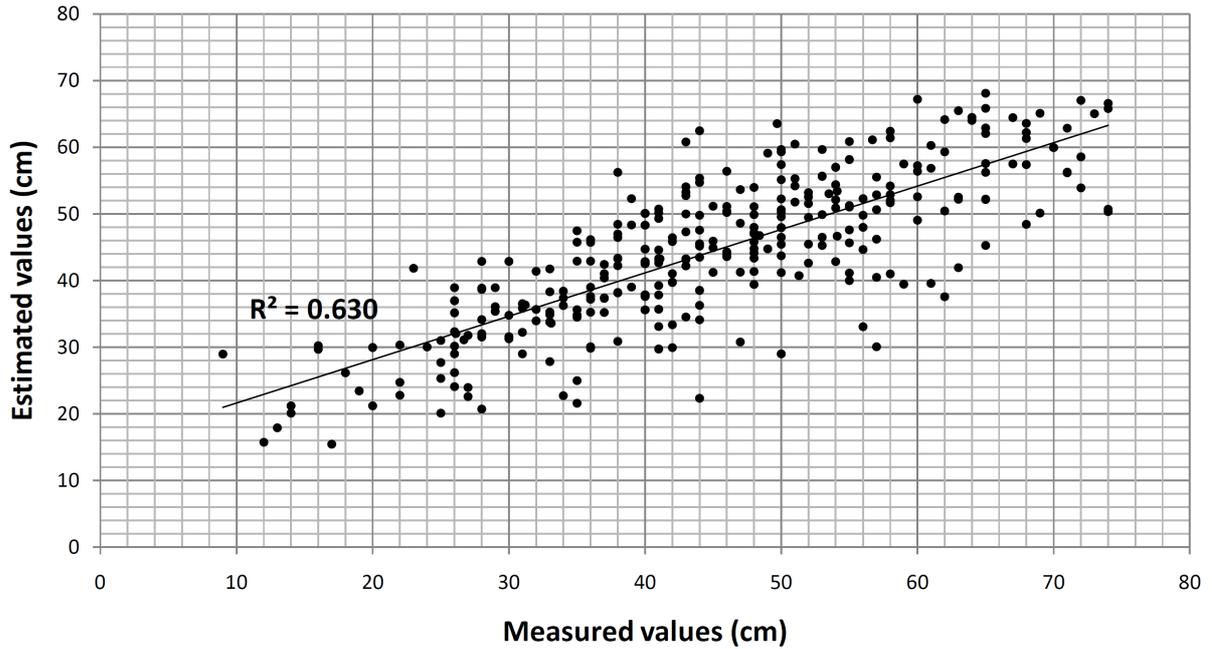


Figure 5-5. Measured diameters vs. predicted diameters for the 307 trees of the test and validation sets.

Figure 5-6 shows the distribution of the reference volume vs. the predicted one for the test and validation sets. It is possible to see, that the R^2 is of about 0.7, with a significant increase compared to that obtained in the diameter estimation (0.63).

Table 5-5. Mean absolute error (MAE), mean square error (MSE) and coefficient of variation (R^2) of the estimates obtained on the test and validation sets for the stem volume. The estimation was carried out with standard height/diameter relationships using the diameters estimated with LIDAR variables and as height the variable “1st maximum”.

	MAE	MSE	R^2
Test Set	0.59	0.66	0.726
Validation Set	0.65	0.82	0.674

Table 5-6. Tree stem volume estimations obtained on the test and validation sets with the proposed system (estimated volume) and with ground collected measures (reference volume).

	Reference Volume (m ³)		Estimated Volume (m ³)	
	Total	Mean	Total	Mean
Test Set	330.676	2.250	311.647	2.120
Validation Set	368.613	2.304	350.937	2.193

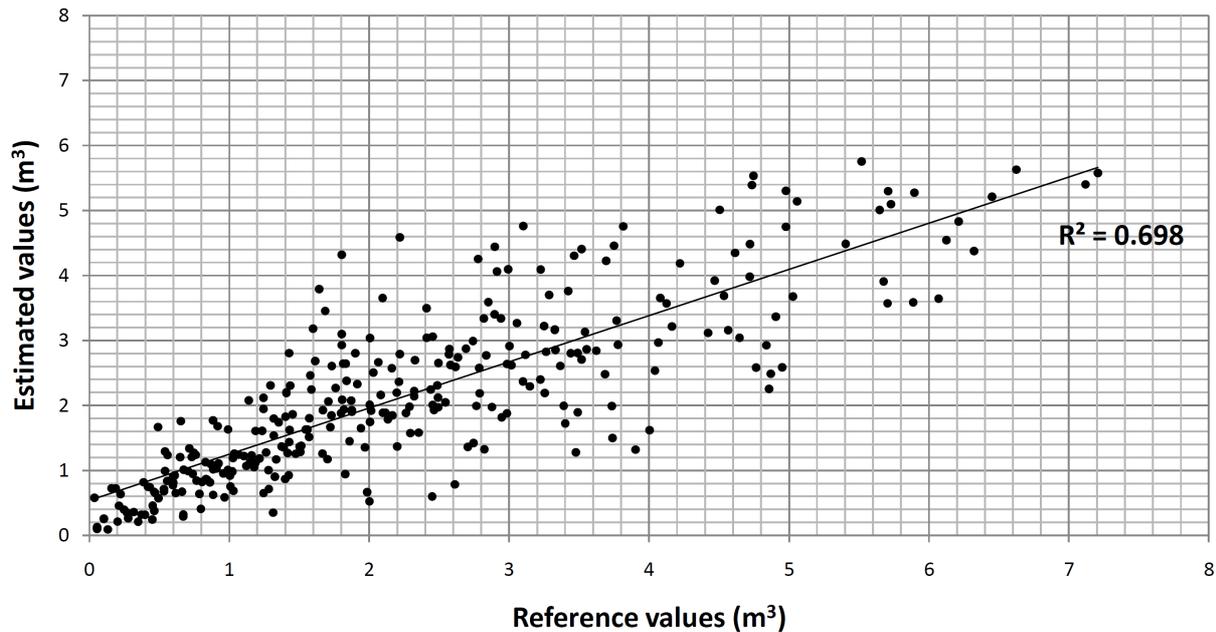


Figure 5-6. Reference volume vs. predicted volume for the 307 trees of the test and validation sets.

Starting from the architecture of Figure 5-2, it is also possible to estimate directly the stem volume. In this case we did not use the information on species present in the classification map but we exploited only the LIDAR variables presented in Table 5-2.

Table 5-7 shows the results obtained for the estimation of the stem volume with the proposed architecture by using different variable subsets and different estimators. Firstly, it is interesting to analyze the selected variables. It is worth noting that all the selection algorithms identified the variables “*1st maximum*” and “*1st mean – 3rd mean*”. The first one is the tree height, while the second one is connected with the vertical shape of the crown. The remaining variables differ for each selection method. In greater detail, the selection based on multivariate linear estimator chooses only variables based on the 1st return, and in particular variables connected with the crown vertical shape (“*1st mean – 2nd mean*”, “*1st mean – 3rd mean*”, “*1st mean – 4th mean*” and “*1st variance*”). The selection based on SVR with linear kernel selects variables connected with the crown vertical shape (“*1st mean – 3rd mean*” and “*1st mean – 4th mean*”) and the crown internal structure (“*3rd minimum*”, and “*4th mean*”). The selection method that chooses the widest range of variables is the one based on SVR-RBF, as it selects variables connected with the crown vertical shape (“*1st mean – 3rd mean*”), the tree height (“*1st minimum*”), the crown internal structure (“*4th range*”), the species (“*1st intensity*”) and the crown horizontal shape (“*radius 1*”).

These selections shows us that almost all the kinds of variables considered are connected with the stem volume. In particular, some of them have a strong linear relation (e.g., those connected with tree height and crown vertical shape), while others have a non linear relation (like the crown horizontal shape and the crown internal characteristics). The latter are selected only by a non linear model (SVR-RBF).

Regarding the estimation results, it is possible to observe that also in this case the selection based on SVR-RBF combined with the SVR-RBF estimator provided the best results. However, in this experiment the difference between the accuracies yielded by this technique and those achieved by the other one is

relevant. On the test set, it provided good results in terms of both MSE and R^2 compared to the other configurations (R^2 is of 0.71 on the test set compared to an average of 0.65 of the other methods). This behavior is confirmed also on the validation set, even if the difference is smaller. It is also worth noting that the estimations based on SVR provide always better accuracies than those based on the linear estimator even if in some configurations the differences are relatively small (e.g., in the selection based on linear estimator results obtained with linear estimator and SVR with linear kernel function are quite similar).

Table 5-7. Mean absolute error (MAE), mean square error (MSE) and coefficient of variation (R^2) of the estimates obtained on the test and validation sets using different variable sets and estimators.

Variables selected	Selection method	Estimator	Test			Validation		
			MAE	MSE	R^2	MAE	MSE	R^2
1 st maximum	SFS with linear estimator	linear	0.69	0.87	0.643	0.76	0.89	0.641
1 st mean-2 nd mean		SVR-linear	0.68	0.86	0.652	0.74	0.89	0.644
1 st mean-3 rd mean		SVR-RBF	0.67	0.83	0.661	0.68	0.87	0.653
1 st mean-4 th mean			0.69	0.88	0.636	0.77	0.91	0.629
1 st variance	SFS with SVR-lin estimator	linear	0.69	0.84	0.661	0.75	0.93	0.634
1 st maximum		SVR-linear	0.66	0.80	0.681	0.71	0.88	0.646
1 st mean-3 rd mean		SVR-RBF	0.70	0.88	0.639	0.78	0.91	0.628
4 th mean			SVR-linear	0.68	0.85	0.651	0.74	0.92
3 rd minimum	SFS with SVR-RBF estimator	linear	0.60	0.70	0.715	0.71	0.86	0.650
1 st mean-4 th mean		SVR-linear						
1 st maximum		SVR-RBF						
1 st mean-3 rd mean								
1 st minimum								
4 th range								
1 st intensity								
radius 1								

Table 5-8 reports the estimations of the volume in m^3 for the test and validation sets, compared to the reference ones. As one can see, the total and mean values are quite similar for both the sets.

Table 5-8. Tree stem volume estimation obtained on the test and validation sets with the proposed system (estimated volume) and with ground collected measures (reference volume).

	Reference Volume (m^3)		Estimated Volume (m^3)	
	Total	Mean	Total	Mean
Test Set	330.676	2.250	325.601	2.215
Validation Set	368.613	2.304	362.693	2.267

Figure 5-7 shows the distribution of the reference volume vs. the one estimated with the SVR-RBF algorithm for the test and validation sets.

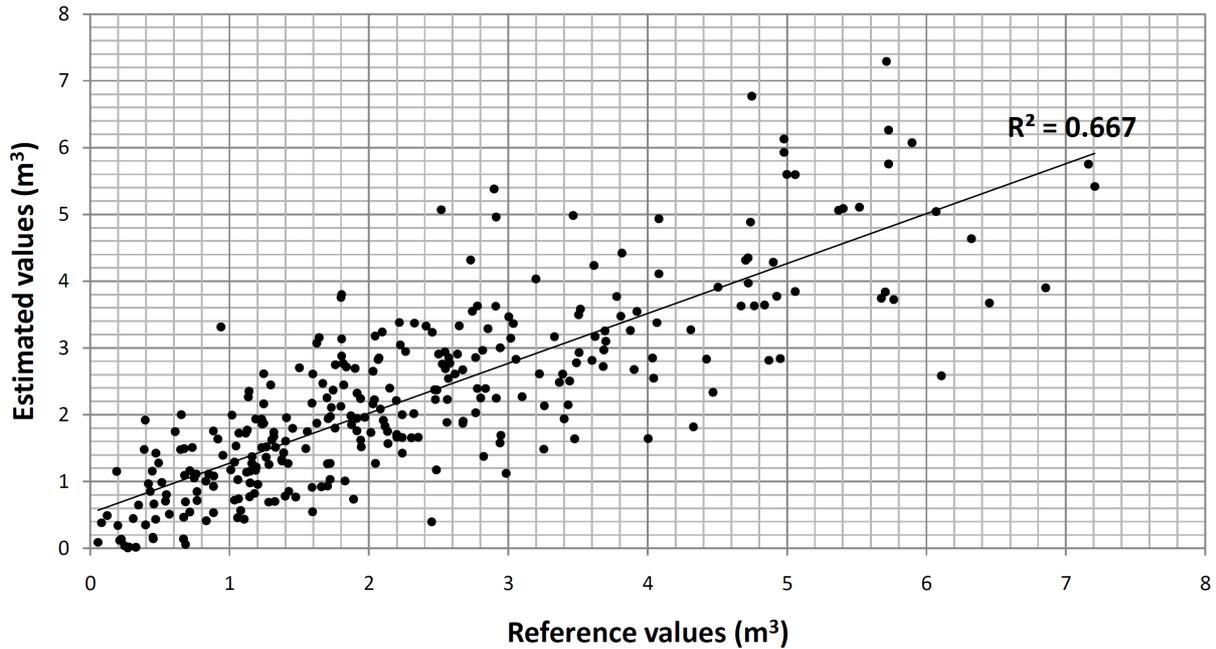


Figure 5-7. Reference volume vs. predicted volume for the 307 trees of the test and validation sets.

5.3.3 Discussion

Comparing the results of the two experiments it is possible to draw some interesting considerations. Firstly, it is worth noting that the estimates of the stem volume with the two methodologies presented are accurate in both cases, with the first approach (based on stem diameters estimation) that seems to give better results in terms of estimation accuracy and coefficient of determination (we have a higher value of R^2 on both the test and validation sets: i.e., 0.72 and 0.67 vs. 0.71 and 0.65, respectively). These results show the effectiveness of the proposed approach. The main negative point of these results can be observed comparing Table 5-6 and Table 5-8. In fact it seems that in terms of total volume the first approach underestimates more than the second one, even if the results are quite similar.

Another consideration strictly connected with the previous one emerges by comparing the estimation results of stem diameters and volumes starting from LIDAR variables. It seems that the considered variables are more suitable for the estimation of stem volume than diameters. In the estimation of stem diameters the variables selected are quite similar for all the estimators considered; thus the results of the different estimators are very similar. This is not the case of volume estimation, where the variables selected differs from an estimator to another. In particular SVR-RBF (a complex non linear estimator) selects a pool of variables with very different physical meanings reaching higher accuracies with respect to other estimators.

By analyzing the variables selected, one can see that there are some variables strongly correlated with both the volume and the diameter, like the ones related to tree height. A variable belonging to this set is present in all selection results. This is quite obvious as the tree height is strongly related to both diameter and volume. Regarding the other variables, it seems that those related to the crown internal structure are those more correlated to the diameter. In fact almost one third of the selected variables belong to this set.

On the contrary, it seems that variables related to the horizontal and vertical shape of the crown do not provide much information in this context. Considering the estimation of stem volume, the situation is quite different. In this case we have very different variables set changing the estimator considered. It is worth noting that the selection performed by the SVR-RBF provides variables belonging to every set. Moreover, in this case the variables selected by the linear estimator belong all to the 1st return showing that probably these are the only variables that have a linear correlation with the volume.

5.4 Conclusion

In this chapter a system for the exploitation of discrete multireturn LIDAR data for the estimation of trees stem diameter and volume is presented. The system proposed is made up of four different blocks: pre-processing, segmentation, variable extraction and selection, and estimation. We presented and analyzed different kinds of variables extracted from LIDAR data, different variable selection algorithms and different estimation techniques. From the experimental results we can draw the following conclusions:

- i. the proposed system is effective for the estimation of tree stem diameters and volumes;
- ii. the approach to the estimation of stem volume based on the estimation of stem diameters seems to be the most effective. The results in terms of MAE, MSE and R^2 are better if compared to the direct estimation of stem volume;
- iii. the estimation accuracy is maxima when using the same regression technique in both the phases of variable selection and estimation;
- iv. when the relationship between the estimated variables and the targets can be approximated as linear, a simple estimator (like the multivariate linear regressor) provides results comparable to complex non-linear estimators (like SVR);
- v. a non-linear regression model (like SVR with RBF kernel function) provides always better results when compared to other estimators (like multivariate linear regression). The difference in accuracy is higher when there is a non linear relationship between the variables and the target;

As a future development of the proposed system, we plan: i) to improve the estimation of diameters considering other variables; ii) to compare the results of the SVR with other non-linear multivariate parametric regression techniques; iii) to analyze the impact of the posting density (number of LIDAR measures per square meter) on the estimation results; and iv) to estimate other attributes like the biomass volume.

5.5 Acknowledgments

We would like to thank Sergio Tonolli for the field work and for the preparation of the ground truth data. This work was partially supported by the CARBOITALY project funded by the FISIR program of the Italian Ministry of University and Research.

5.6 References

- [1] C. Marsden, G. le Maire, J. Stape, D. L. Seen, O. Roupsard, O. Cabral, D. Epron, A. M. Nascimento Lima and Y. Nouvellon, "Relating MODIS vegetation index time-series with structure, light absorption and stem production of fast-growing Eucalyptus plantations," *Forest Ecology and Management*, in press.
- [2] P. Muukkonen, and J. Heiskanen, "Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories", *Remote Sensing of Environment*, vol. 107, no. 4, pp. 617-624, 2007.
- [3] R.J. Hall, B.S. Case, E. Arsenault, D. T. Price, J. E. Luther, D. E. Piercey, L. Guindon, and R. A. Fournier, "Modeling and mapping forest biomass using forest inventory and Landsat TM data: results from the Foothills Model Forest, Alberta", *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2002*, vol. 3, pp. 1320-1323, 2002.
- [4] J. E. Luther, R. A. Fournier, R. J. Hall, C.-H. Ung, L. Guindon, D. E. Piercey, M.-C. Lambert, and A. Beaudoin, "A strategy for mapping Canada's Forest biomass with Landsat TM imagery", *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2002*, vol. 3, pp. 1312-1315, 2002.
- [5] P. Muukkonen, and J. Heiskanen, "Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data", *Remote Sensing of Environment*, vol. 99, no. 4, pp. 434-447, 2005.
- [6] H. Wang, and K. Ouchi, "Accuracy of the K-Distribution Regression Model for Forest Biomass Estimation by High-Resolution Polarimetric SAR: Comparison of Model Estimation and Field Data", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1058-1064, 2008.
- [7] M. J. Quiñones, and D. H. Hoekman, "Exploration of Factors Limiting Biomass Estimation by Polarimetric Radar in Tropical Forests", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 86-103, 2004.
- [8] T. Mette, K. P. Papathanassiou, I. Hajnsek, and R. Zimmermann, "Forest biomass estimation using polarimetric SAR interferometry", *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2002*, vol. 2, pp. 817-819, 2002.
- [9] J. B. Drake, R. O. Dubayaha, R. G. Knox, D. B. Clark, and J. B. Blair, "Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest", *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 378-392, 2002.
- [10] J. Hyypä, O. Kelle, M. Lehtikoinen, and M. Inkinen, "A Segmentation-Based Method to Retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 5, pp. 969-975, 2001.
- [11] Z. J. Bortolot, and R. H. Wynne, "Estimating forest biomass using small footprint LIDAR data: An individual tree-based approach that incorporates training data", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 6, pp. 342-360, 2005.
- [12] I. Guyon, A. Elisseeff, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [13] J. Guajardo, J. Miranda, and R. Weber, "A hybrid forecasting methodology using feature selection and support vector regression", *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems 2005*, vol. 6, 2005.
- [14] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees", Wadsworth & Brooks, Monterey, 1984.
- [15] J. Weston, A. Elisseeff, B. Schoelkopf, and M. Tipping, "Use of the zero norm with linear models and kernel methods", *Journal of Machine Learning Research*, vol. 3, pp. 1439-1461, 2003.
- [16] A. J. Smola, and B. Schoelkopf, "A Tutorial on Support Vector Regression", *NeuroCOLT2 Technical Report Series, NC2-TR-1998-030*, October, (1998).
- [17] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London*, 1909.
- [18] V. Cherkassky, and Y. Ma, "Practical Selection of SVM Parameters and Noise Estimation for SVM Regression", *Neural Networks*, vol. 17, no. 1, pp. 113-126, 2004.
- [19] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416-1427, 2008.

Chapter 6

6 Fusion of hyperspectral and LIDAR remote sensing data for the estimation of tree stem diameters

The estimation of stem diameters can be very useful in the study of forests, as together with height and tree species, it is one of the most important parameters used in forest inventories. In this chapter a system for the estimation of stem diameters with LIDAR and hyperspectral data (both separately and combined in a data fusion framework) is presented. An analysis on the effectiveness of these data in the estimation process and on the accuracy and robustness of different estimation algorithms is presented. Experimental results point out the effectiveness and the properties of the proposed system.

6.1 Introduction

The study of forests and their biophysical parameters is an important task that has many implications in many different fields (e.g., economy, environment). Forests are an important source of information for studies related to climate change and carbon balance. Quantitative studies on carbon exchange and stocks have become of great importance in the recent years due to the rules of the Kyoto protocol that require each nation to have an estimate of the CO₂ stored and exchanged by their forests.

In this context, it is important to study the structural parameters of the trees and forests, like the height, the stem volume, the basal area, the diameter at breast height (DBH) of the trees, etc. In particular the stem volume is directly related to the amount of CO₂ stored by trees. In order to compute this parameter it is necessary to know the DBH, the diameter and the species of a tree. In this study we focus our attention on the DBH as it is a variable that can be used for the estimation of various parameters of the trees.

Two types of remote sensing data have been widely used in the last years for the study of forests: hyperspectral and LIDAR data. Usually hyperspectral data are mainly exploited for the classification of forest species, while LIDAR data are mainly used for the estimation of biophysical parameters. Concerning the estimation of stem diameters, some studies exist in the literature, which are mainly based on LIDAR signal processing [1]. At the present at the best of our knowledge no studies exist on the estimation of stem diameters by hyperspectral data.

This chapter has been submitted at IEEE International Geoscience and Remote Sensing Symposium 2009 with the title: "Fusion of hyperspectral and LIDAR remote sensing data for the estimation of tree stem diameters". Authors: Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle.

In this scenario, the main goals of this chapter are: i) to analyze the use of LIDAR and hyperspectral data (both separately and combined) for the estimation of stem diameters; ii) to develop a system for the estimation of stem diameters that can exploit the aforementioned data both separately or in a data fusion framework; iii) to compare the accuracy and robustness of different estimation algorithms (i.e. Support Vector Regression and linear regression). Moreover, we study the possibility to use hyperspectral data alone to derive rough estimations of stem diameters. This issue is very interesting as hyperspectral data are widely used for the classification of forest areas (they provide a detailed description of spectral signatures of tree species), and thus it would be important to understand if they are a significant information source also to estimate physical parameters of the trees.

6.2 Data set Description

The data set considered in this study is a forest area on Mount Bondone in the Italian Alps near the city of Trento. The central point of the area has the following coordinates: 46°3'15.84"N, 10°59'59.35"E. This area has an extension of approximately 170 ha and an elevation that ranges from 200 m to 1400 m. The area is characterized by the presence of many broad leaved tree species (European Beech, Holm Oak, Downy Oak, Common Hazel, Silver Birch, etc) and some coniferous species (Scots Pine, European Larch, etc).

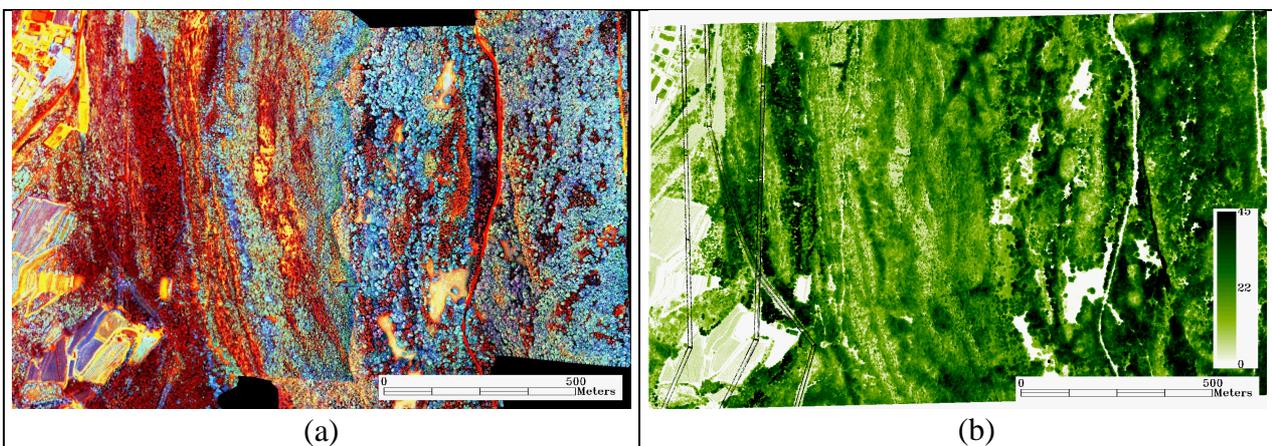


Figure 6-1. Example of images used in the experiments: a) false color composition of channels 10 (483 nm), 35 (718 nm) and 55 (911 nm) of the hyperspectral image; b) Digital Canopy Model (DCM) of the analyzed area.

The remotely sensed data were acquired on September 2007. The hyperspectral data consist of four partially overlapping images acquired by an airborne AISA Eagle sensor in 63 spectral bands, ranging from 400 to 990 nm, with a spectral resolution of about 9.2 nm and a spatial resolution of 1 m (see Figure 6-1). The LIDAR data were acquired by the sensor Optech ALTM 3100, with a mean density of 8.6 points per square meter for the first return. The laser pulse wavelength and the laser repetition rate were 1064 nm and 100 kHz, respectively. The data used in our investigation refer to the first four LIDAR returns. A Digital Terrain Model (DTM) of the investigated area with a spatial resolution of 1 m was extracted from the LIDAR data.

The reference data samples (178 trees) were collected during a ground survey in summer 2007. We extracted these sample points from the entire study area. The samples were collected on the basis of: i) the species and ii) the spatial distribution (samples have a uniform distribution across the scene). They have been divided into three sets: training (60 trees), test (59 trees) and validation (59 trees).

6.3 Methods

6.3.1 Architecture of the proposed system

Figure 6-2 shows the general architecture of the proposed multisensor system for the estimation of stem diameters. It can be divided into four main blocks: i) data preprocessing; ii) segmentation; iii) variables extraction and selection; and iv) estimation. These blocks are analyzed in greater detail in the following. It is worth noting that the system, with adequate simplifications, can be also used with only LIDAR or hyperspectral data.

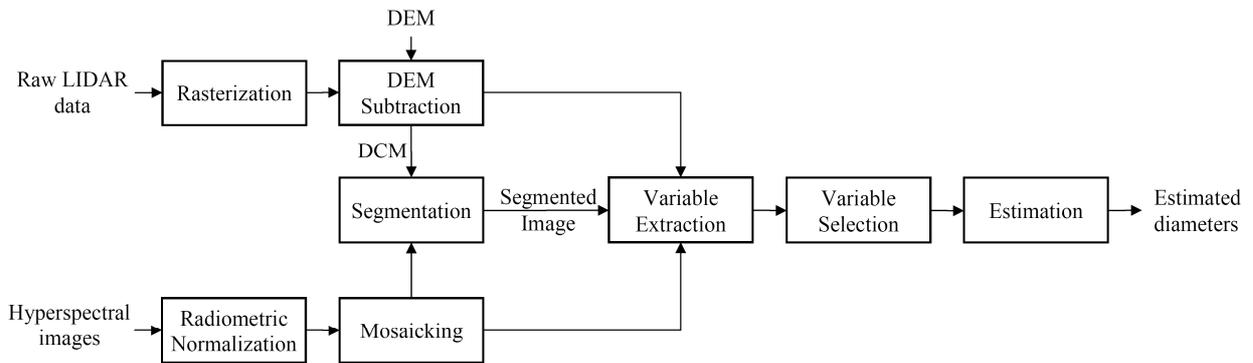


Figure 6-2. Architecture of the proposed system.

6.3.2 Preprocessing

The preprocessing applied to LIDAR data consisted in the rasterization of the raw point cloud, and in the computation of the Digital Canopy Model (DCM) (the Digital Elevation Model (DEM) of the area was subtracted to the elevation information). This last process allowed us to obtain the height of each pixel with respect to the ground.

Regarding the hyperspectral data, after the application of a radiometric normalization, the four overlapping images were mosaiked in order to cover the whole area analyzed.

6.3.3 Segmentation

An important phase in the system proposed is the segmentation. This phase drives the next steps of variable extraction and diameter estimation. The rationale of this step is to delineate individual tree crowns in order to have diameter estimation at the tree level.

Concerning LIDAR data, the segmentation algorithm used in this chapter is derived from the algorithm presented by Hyypä *et al.* in [2]. This algorithm, is divided into three main steps: i) prefiltering; ii) seed point extraction; iii) seed region growing.

According to [2], in the pre-filtering phase for emphasizing tree crowns the DCM was filtered with a convolution filter, whose coefficients are defined as follows:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} / 16 \quad (6.1)$$

The seed point extraction phase is aimed at the identification of the tree tops, from which starting the region growing procedure. Seeds are the local maxima higher than a certain threshold value. They are detected moving a window of a given size defined by the user through the image. At the end of this process we obtain the set of the seed points $S = \{s_1, \dots, s_N\}$, where s_n identifies the n -th point.

The last phase consists in the seed region growing for the identification of the crowns of the trees. Seed region growing starts from each seed and grows iteratively the region from the first order neighborhood system to the n -th. A pixel $I(i, j)$ is added to the considered region if it satisfies two conditions that take into account both the dimension and the shape of the crown. If we define the set of the regions $R = \{r_1, \dots, r_N\}$, where r_n identifies the region around the seed point s_n , we can write as follows:

$$I(i, j) \in r_n \quad \text{if} \quad \begin{cases} I(i, j) > P * I_{s_n} \\ D[r_n + I(i, j)] < thDiameter \end{cases} \quad (6.2)$$

where I_{s_n} is the height of the considered seed point, P is a value between 0 and 1, $D[r_n + I(i, j)]$ is the diameter of the considered region including the new pixel $I(i, j)$, and $thDiameter$ is the maximum acceptable diameter of a region. The algorithm stops when none pixel is added to any region.

Regarding the segmentation of hyperspectral data it was carried out with the Definiens eCognition software (i.e., [3]), which exploits a multiresolution segmentation algorithm. This is a region growing approach based on a heuristic optimization procedure which locally minimizes the average heterogeneity of image objects for a given resolution.

6.3.4 Variable extraction and selection

From each crown a series of variables were extracted from both the LIDAR and hyperspectral data. Regarding the LIDAR data, we extracted from the pulses of each return the variables reported in Table 6-1. With respect to the hyperspectral bands, for each spectral channel we computed the mean value among the pixel of each crown.

A stepwise variable selection [4] was then applied to select subsets of the above-mentioned measures relevant for the proposed estimation problem. The selection algorithm works as follows. In the initial step, each of the available variables is used for estimating the diameters according to a linear model, and then the one that results in the highest R^2 value is selected. In the next steps, each of the remaining variables is

added to those already selected. The one which results in the highest increase of R^2 is included in the set of selected variables if it meets the statistical criterion for entry. This criterion is based on the significance in the increase of the R^2 produced by the addition of the considered variable. Then variables that are already in the model are evaluated for removal. The criterion for removal is similar to the one for entry: the variable whose rejection would result in the lowest decrease of R^2 is removed. The process is iterated until no selected variable meets the removal criterion, and no unselected variable meets the entry criterion. The number of variables selected changes according to the characteristics of the problem considered and to the considered entry and removal criterion.

Table 6-1. Variables extracted from LIDAR data.

Variable	Variable
maximum	standard deviation
minimum	kurtosis
mean	skewness
range	coefficient of variation
maximum ⁿ with n=0.1, ..., 5	mean intensity
percentiles from the 5 th to the 95 th	

6.4 Estimation techniques

In the estimation phase we considered two different estimators: a linear multivariate estimator and a non-linear multivariate ϵ -Insensitive Support Vector Regression (ϵ -SVR) algorithm.

The linear estimator is very simple and it is associated with a low computational load. For this reason, it is widely used in the forest science domain for the estimation of stem volume and biomass. It provides good results when the correlation between the target variable and the predicting variables can be reasonably approximated by linear function.

Regarding the ϵ -SVR estimator, it is an advanced machine learning technique that allows one to seize complex regression problems characterized by: i) non linear correlation between the target variable and the predicting variables; and ii) a reduced number of training samples. It is based on the Support Vector Machine theory and its goal is to find a function $f(x)$ that has at most ϵ deviation from the real targets for the training data, and at the same time is as flat as possible. In other words, it does not consider errors if they are smaller than ϵ , whereas it does not accept any deviation larger than ϵ . For more details on this estimator we refer the reader to [5],[6].

6.5 Experimental analysis and discussion

In order to achieve the goals of this chapter we carried out three experiments for the estimation of the diameters with three different sets of variables: i) LIDAR variables; ii) hyperspectral variables; and iii) LIDAR and hyperspectral variables. In each of these experiments we first performed the variable selection phase, and then the estimation phase with a multivariate linear estimator and an SVR estimator.

CHAPTER 6

The variables selected in the three experiments are as follows: i) 1st return range, 2nd return skewness, maximum of the 1st return at the power of 0.2, and the 95th percentile of the 1st return; ii) 7 hyperspectral bands (see Figure 6-3 for the distribution in the spectrum of the selected bands); iii) maximum of the 1st LIDAR return at the power of 0.2, hyperspectral bands at about 429 and 979 nm.

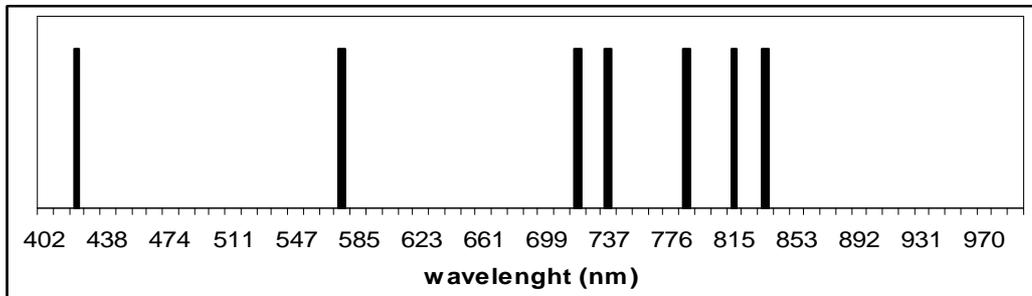


Figure 6-3. Hyperspectral bands selected.

As it is possible to observe from Figure 6-3, the spectral bands selected are strongly related to the characterization of the species of the single tree; thus they are directly correlated with the stem diameter. In fact the species is one of the key features that with the height characterize the diameter of a tree. All these bands can be connected to chlorophyll content (bands at 575, 718 and 735 nm), carotenoides content (band at 420 nm), leaves structure, and water content (bands at 785, 815 and 834 nm) ([7],[8]). These parameters change according to the species considered and thus the spectral signature on the related bands is useful in the species identification.

Table 6-2 presents the results obtained on test and validation sets in the three experiments. Concerning the SVR estimator, we considered two kernel functions: linear and RBF functions. The model selection phase was carried out through a grid search for the regularization parameters C , the tube insensitive to the errors ϵ , and the spread of the RBF kernel γ [6].

Firstly, we can observe that the use of both hyperspectral and LIDAR variables provides the best results with all the estimators considered. The MAE is only of about 4 cm on the test set, and of about 5 cm on the validation set.

Secondly, the SVR estimator with both the linear and RBF kernel functions provides higher accuracy than the linear estimator with almost all the variables sets considered. Concerning the linear kernel, it achieved the best results using the LIDAR variables and the fusion between the hyperspectral and the LIDAR ones. The SVR with RBF kernel provided the smallest error when using only the hyperspectral variables. This can be explained with the higher complexity (non linearity) of the regression function in the case in which only spectral channels are used for retrieving the diameters of the trees.

In general, it is worth nothing that in all the three experiments the performances in terms of MAE are similar for all the three estimators considered, while the R^2 provided by the SVR estimator is slightly better than those obtained with the linear estimator.

Concerning the different data sources the use of hyperspectral images results in the highest error in the estimation process. This is reasonable, given the complexity of the relation between the spectral signatures and the diameters of the trees. Nonetheless, these results can be considered interesting when a rough estimation of stem diameters should be done and only hyperspectral data are available.

Table 6-2. Results obtained on the test and validation sets in terms of mean absolute error (MAE), mean square error (MSE) and coefficient of variation (R^2) using different variable sets and estimators.

Variables selected	Estimator	Test set			Validation set		
		MAE	MSE	R^2	MAE	MSE	R^2
1 st return range	Linear	5.12	40.43	0.540	5.31	43.56	0.520
2 nd return skewness (1 st return maximum) ^{0.2}	SVR - linear	4.48	33.03	0.650	5.24	40.14	0.573
1 st return 95 th percentile	SVR - RBF	4.96	38.45	0.585	5.87	50.37	0.463
420 nm hyperspectral band	Linear	5.66	53.14	0.399	6.37	60.65	0.324
576 nm hyperspectral band							
718 nm hyperspectral band	SVR - linear	5.85	55.17	0.424	6.22	57.27	0.366
737 nm hyperspectral band							
786 nm hyperspectral band	SVR - RBF	5.78	51.33	0.439	6.15	54.90	0.400
815 nm hyperspectral band							
834 nm hyperspectral band	SVR - RBF	5.78	51.33	0.439	6.15	54.90	0.400
(1 st return maximum) ^{0.2}							
429 nm hyperspectral band	SVR - linear	4.15	28.41	0.681	5.20	40.14	0.570
979 nm hyperspectral band	SVR - RBF	4.52	32.75	0.679	4.96	41.10	0.564

6.6 Conclusion

In this chapter a system for the estimation of stem diameters with hyperspectral and LIDAR data has been proposed, and an empirical study on the effectiveness of these different information sources in the estimation task has been presented. Different kinds of variables (extracted from data acquired by different sensors) and different estimators have been compared. From the experimental results, it is possible to conclude that: i) the system proposed is effective for the estimation of stem diameters; ii) LIDAR data involve accurate estimates of tree diameters (MAE of about 5 cm on the validation set); iii) the combination of hyperspectral and LIDAR data allows one to slightly increase the performances (MAE reduced on both the test and validation sets); and iv) as expected hyperspectral data provide less accurate estimations than the LIDAR ones, but the retrieved tree diameters are still reasonable indications of the true values when only hyperspectral images are available.

6.7 References

- [1] M. Dalponte, L. Bruzzone, and D. Gianelle, "Estimation of tree biomass volume in alpine forest areas using multireturn LIDAR data and support vector regression," *Image and Signal Processing for Remote Sensing XIV. Proceedings of the SPIE*, vol. 7109, pp. 71090C-71090C-12, 2008.
- [2] J. Hyypä, O. Kelle, M. Lehtikoinen, and M. Inkinen, "A Segmentation-Based Method to Retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 5, 969-975 (2001).
- [3] <http://www.definiens.com/>
- [4] A. C. Rencher, "Methods of multivariate analysis," *Wiley-IEEE*, 2003.

CHAPTER 6

- [5] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Alvarez, and F. Melgani, "Robust Support Vector Regression for Biophysical Variable Estimation from Remotely Sensed Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 3, pp. 339-343, July 2006.
- [6] A. J. Smola, and B. Schoelkopf, "A Tutorial on Support Vector Regression," *NeuroCOLT Technical Report Series*, 1998.
- [7] Y. Zur, A. A. Gitelson, O. B. Chivkunova, and M. N. Merzlyak, "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.
- [8] P. Thenkabail, S. Enclona, E. A. Ashton, M. S. Legg, and M. J. De Dieu, "Hyperion, IKONOS, ALO, and ETM+ sensors in the study of African rainforests," *Remote Sensing of Environment*, vol. 90, no. 1, pp. 23-43, March 2004.

Chapter 7

7 Conclusions

This thesis investigated novel systems for the use of hyperspectral and LIDAR remote sensing data in the classification of forest areas and in the estimation of trees stem attributes. The main novel contributions of this thesis to the literature are: i) an empirical analysis on the relationship between spectral resolution, classifier complexity and classification accuracy in the classification of complex forest areas; ii) a novel system for the fusion of hyperspectral and LIDAR remote sensing data in the classification of forest areas; iii) an empirical analysis on the use of multireturn LIDAR data for the estimation of tree stem volume; iv) a novel system for the estimation of single tree stem diameter and volume with multireturn LIDAR data; and v) a system for the fusion of hyperspectral and LIDAR remote sensing data in the estimation of tree stem diameters.

These contributions are related to each other and they can be grouped into two areas: classification and estimation in the forestry application domain. Regarding the classification, at first an empirical analysis on the relationship among spectral resolution, classifier complexity and classification accuracy have been carried out. This analysis is very important as the indications derived from it can be used for both the design of new sensors, and for a more efficient use of the existing ones. The second novel contribution in the classification area is a system that allows one to jointly exploit hyperspectral and LIDAR data for the analysis of complex forest areas. This contribution proposes on the one hand a complete system that from raw hyperspectral and LIDAR data provides classification maps, and on the other hand important indications on the significance of LIDAR returns in the classification process, as well as on their complementary role with hyperspectral data.

Summarizing, the main conclusions that can be extracted from the classification part of the thesis are:

- i) advanced and complex classification systems (like SVM) are able to exploit high dimensional data providing very high classification accuracies, whereas more standard and simpler classifiers (like LDA) do not provide high classification accuracies;
- ii) other simple classifiers (like ML) are advantaged when the spectral resolution decreases and thus (considering the same kind of detector) the SNR of the images increases;
- iii) the proposed novel system for the fusion of hyperspectral and LIDAR data is effective in the classification of the considered complex forest areas providing very high classification accuracies and detailed maps;

CHAPTER 7

- iv) the proposed classification system properly exploits the complementary information contained in hyperspectral and LIDAR data; in particular, the elevation information provided by LIDAR data results useful in the separation of tree species with different mean heights but similar spectral signatures;
- v) LIDAR returns after the first one are not relevant for the classification task.

Regarding the estimation part of the work, we focused our attention on the estimation of single tree stem height, diameter and volume. Firstly, we propose an empirical analysis on the use of multireturn LIDAR data for the estimation of tree stem volume. This study is very important to understand the information content of LIDAR returns and the variables more suitable for the exploitation of this information. This analysis drives also the design of the novel systems proposed in this thesis. The first system exploits multireturn LIDAR data for the estimation of tree stem attributes, in particular tree stem diameter and volume. A detailed empirical analysis has been carried out comparing three variable selection methods and three estimators. The second system extends the previous one by combining and exploiting the spectral information of hyperspectral data with the spatial information of LIDAR data for the estimation of stem diameters. It fuses variables extracted from hyperspectral and LIDAR data for the estimation of tree stem diameters. This system which is based on advanced algorithms can be used to exploit these data separately and combined.

Summarizing, the main conclusions that can be extracted from this part of the thesis are:

- i) the systems proposed are effective for the estimation of tree stem diameters, volume and height;
- ii) 1st, 2nd and 3rd LIDAR returns are important for the estimation of tree stem volume;
- iii) the estimation accuracy is maxima when using the same regression technique in both the phases of variable selection and estimation;
- iv) the proposed system for the fusion of hyperspectral and LIDAR data is effective in the exploitation of the complementary information of these data. The combined use of variables belonging to both sensors allows one to increase the estimation accuracy;
- v) hyperspectral data provide low accuracy estimations of stem diameters. Nevertheless, they can be useful for a rough estimation of this variable as they are less expensive than the LIDAR ones and they can be used also for species classification.

There are various future developments for this thesis both in the classification and in the estimation fields. In the following we point out some main ideas:

- i) to consider in the classification and estimation process, other algorithms, in particular semi-supervised techniques. These techniques can be useful in real situation of forest studies where the number of training samples is small.
- ii) to consider in the analysis also satellite VHR multispectral images (e.g., Quickbird, GeoEye). It can be interesting to analyze how these data can interact with LIDAR ones in both classification and es-

timation. In fact these data, even if they have less spectral information, allow one to make studies at single tree level with a lower cost compared to aerial images;

- iii) in order to use the proposed systems for forest inventories of large areas, an issue that needs more investigation is that related to their generalization ability considering training and test areas derived from areas significantly different at regional scale;
- iv) an analysis on the most suitable point density of LIDAR data in forest analysis application, for both classification and estimation;
- v) the study of forest structure with LIDAR data. In particular it seems interesting to investigate the possibility to detect single trees in dominated forest layers.

CHAPTER 7