Università degli Studi di Roma "Tor Vergata"

Facoltà di Ingegneria

Ph.D Geoinformation Programme

XXII Cycle

Neural network architectures for information extraction from hyper-spectral images

Geoinformation Ph.D Thesis

Tutor
Fabio Del Frate

Candidate
Giorgio Antonino Licciardi

March 2010

Contents

Abstract

Imaging spectroscopy, also known as hyper-spectral remote sensing, is an imaging technique capable of identifying materials and objects in the air, land and water on the basis of the unique reflectance patterns that result from the interaction of solar energy with the molecular structure of the material. Recent advances in aerospace sensor technology have led to the development of instruments capable of collecting hundreds of images, with each image corresponding to narrow contiguous wavelength intervals, for the same area on the surface of the Earth. As a result, each pixel (vector) in the scene has an associated spectral signature or "fingerprint" that uniquely characterizes the underlying objects.

Hyper-spectral sensors mainly cover wavelengths from the visible range (0.4μm- 0.7μm) to the middle infrared range (2.4μm). If we consider the consistency of this data, we can easily understand the importance of finding a method which can transform the data cube into one with reduced dimensionality and maintain, at the same time, as much information content as possible. These techniques are known under the general name of feature reduction. Besides enabling an easier storage and management of the data, features reduction procedures can be crucial for the implementation of op-

timum inversion algorithms.

This research work strives to give a contribution along the direction of extracting information from hyperspectral data. A major instrument is considered for this purpose, which is the use of neural networks algorithms, already recognized to represent a rather competitive family of algorithms for the analysis of hyperspectral data. Besides introducing a novel neural network approach for handling the dimensionality reduction of hyperspectral data, other specific issues will be considered, with a special focus on the unmixing problem, or sub-pixel classification.

While the first three chapters are dedicated to the presentation of the problems, to the current state of art and to the, theoretically sound, proposed solutions, the remaining sections are dedicated to the description and the assessment of the results obtained in different applicative scenarios. Some final considerations conclude the work.

Chapter 1

Hyper-spectral data

Both scientists and common people are becoming increasingly concerned with environmental phenomena such as the photosynthetic conditions of the vegetation, wide deforestation and fires, desertification, sea pollution, together with the general health of the Earth. The monitoring of these events and the understanding of the impact which they could have on the fragile biophysical mechanisms is becoming more and more important than in the past. For this reason, sensors like MERIS, MODIS, AVHRR and AATSR have been designed and placed in orbit. These measurements are performed using several spectral bands (up to 36 for MODIS) located into the visible and the infrared range in order to collect a noteworthy dataset for every kind of global investigation (land use, ocean color, snow cover, sea ice observation...). A following step has been the allocation of many contiguous and narrow bands (more than one hundred) available for the measurement. This technological evolution led to the hyper-spectral imagery, which has demonstrated very high performance in several cases of

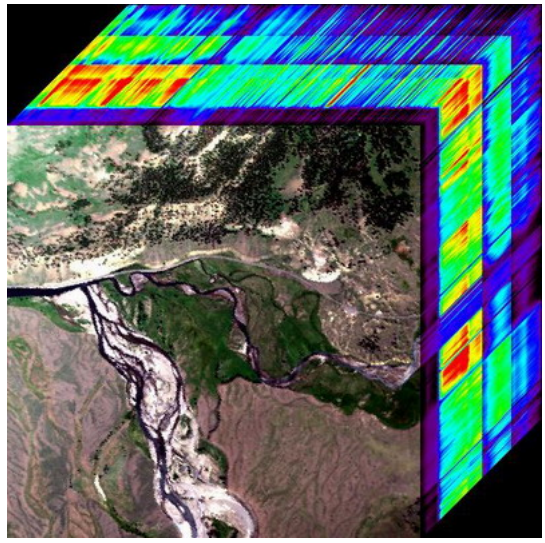


Figure 1.1: *Hyper-cube obtained from a AVIRIS dataset*

material identification and urban mapping, including sub pixel classification. The hyper-spectral sensors differ from each other in terms of number of bands, bandwidth, spatial resolution and spectral range, spatial acquisition and spectral selection modes. Managing such dissimilar type of data is not a simple task and requires the adoption of information extraction techniques that are appropriate for each specific sensor data.

1.1 Data acquisition principles

In general, hyper-spectral sensors can be divided into three different scanning systems for acquiring the image:

- Whiskbroom imagers (electromechanical scanners): on-axis optics or

telescopes with scan mirrors sweep from one edge of the swath to the other. The Field of View (FOV) of the scanner can be detected by a single detector or a single-line-detector. This means that the dwell time for each ground cell must be very short at a given Instantaneous Field of View (IFOV), because each scan line consists of multiple ground cells which will be detected.

- Pushbroom scanners: as electronic scanners they use a detector array to scan over a two dimensional scene. The number of across track pixels detector pixel is equal to the number of ground cells for a given swath. The motion of the aircraft or spacecraft provides the scan in along-track-direction. Pushbroom scanners are the standard for high resolution imaging spectrometers.

- Staring imagers: these imagers are also electronic scanners. They detect a two dimensional FOV at once. The IFOV along and cross track corresponds to the two dimensions of detector area array. Two sub-groups of staring imagers are Wedge Imaging Spectrometer (WIS) and Time Delay Integration Imager (TDI).

Hyper-spectral spectrometers can also have very different spectral selection modes:

Dispersion elements (grating, prism): This group collects spectral images by using a grating or a prism. The incoming electromagnetic radiation will be separated into different angles. The spectrum of a single ground pixel will be dispersed and focused at different locations of one dimension of the detector array. This technique is used for both whiskbroom

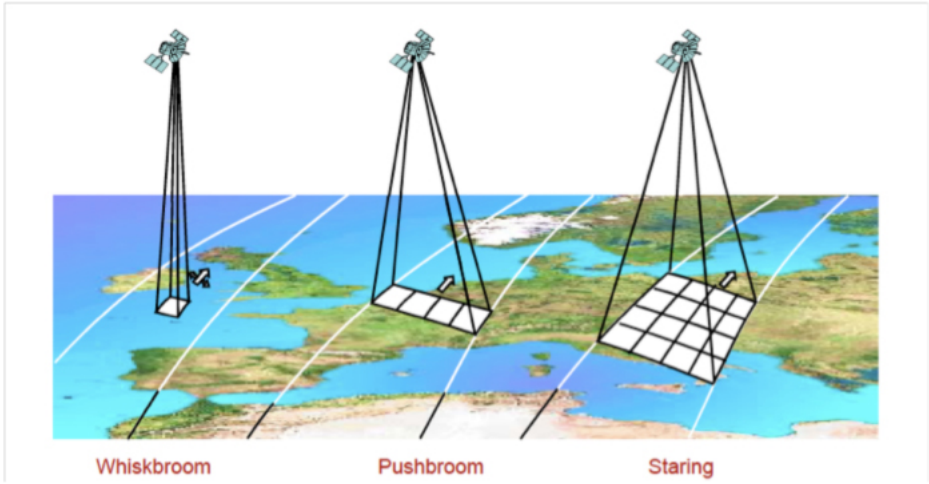


Figure 1.2: *Scanning approaches*

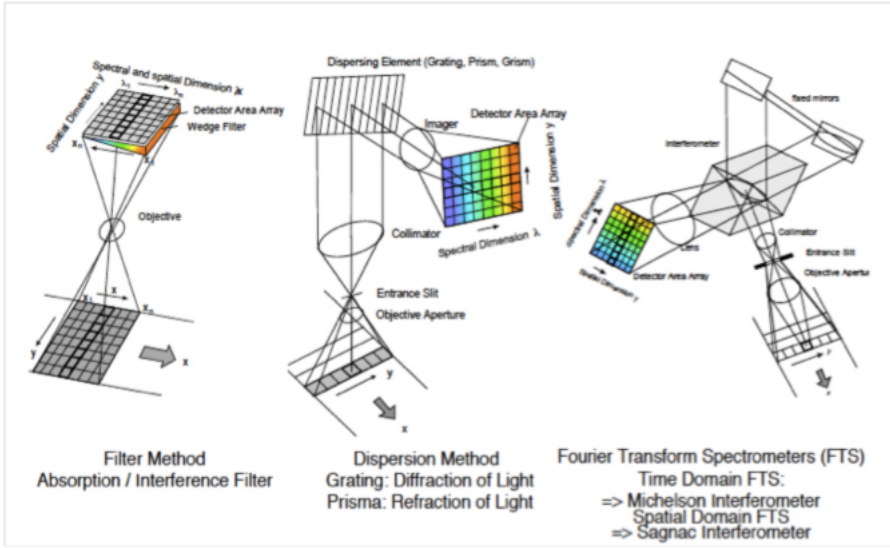


Figure 1.3: *Spectral selection modes*

and pushbroom image acquisition modes. Most hyper-spectral imagers are using grating as dispersive elements, whereas some use prisms.

Filter based systems: a narrow band of a spectrum can be selected by applying optical bandpass filters (tunable filters, discrete filters and linear wedge filters). A linear wedge transmits light at a centre wavelength that depends on the spatial position of the illumination in the spectral dimension. The detector behind the device receives light at different wavelengths of the scene.

Fourier-Transform Spectrometers (FTS): a Fourier-transform spectrometer is an adaption of the Michelson interferometer [1] where a collimated beam from a light source is divided into two by a beamsplitter and sent to two mirrors. These mirrors reflect the beams back along the same paths to the beamsplitter, where they interfere. The signal recorded at the output depends on the wavelength of the light and the optical path difference between the beamsplitter and each of the two mirrors. If the optical path difference between the two beams is zero or a multiple of the wavelength of the light then the output will be bright, otherwise if the optical path difference is an odd multiple of half the wavelength of the light then the output will be dark.

In the Fourier transform spectrometer, one of the mirrors is scanned in the direction parallel to the light beam. This changes the path difference between the two arms of the interferometer, hence the output alternates between bright and dark fringes. If the light source is monochromatic, then the signal recorded at the output will be modulated by a cosine wave; if it is not monochromatic then the output signal will be the Fourier transform

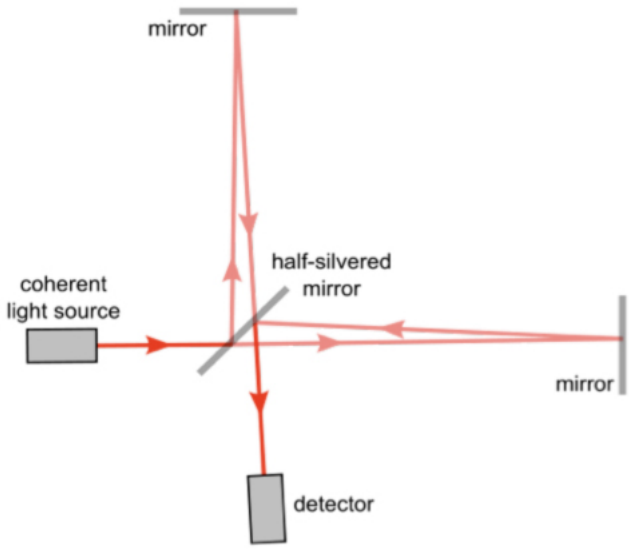


Figure 1.4: *Schematic of a Fourier-Transform Spectrometer*

of the spectrum of the input beam. The spectrum can then be recovered by performing an inverse Fourier-transform of the output signal.

1.2 Data uniformity

Efficient and accurate imaging spectroscopy data processing asks for perfectly uniform data in both spectral and spatial dimensions. The precision of a measurement is determined by the instrument response $r(z)$, with z the position coordinate. The transformation from the input physical quantity to the measurement $O(z)$ is described mathematically by a convolution:

$$O(z_0) = \int_W i(z) r(z - z_0) dz \tag{1.1}$$

Or in shorthand notation:

$$O(z) = i(z) \otimes r(z) \tag{1.2}$$

Where:

$i(z)$: input signal

$r(z - z_0)$: sensor response at the position z_0

$O(z_0)$: output signal, assigned to the position $z = z_0$

W: significant spatial range covered by the response of the system. The image of a scene viewed by the sensor is not completely its faithful reproduction. Small details are blurred relative to larger features; this blurring is characterized by the total sensor Point Spread Function (PSF). The response of a detector element depends principally from the PSF, which can be viewed as the spectral/spatial responsivity of the sensor.

PSF consists of several components:

- Optical PSF (PSF_{opt}), defined by the spatial energy distribution in the image of a point source. Being an optical system not perfect, the energy from a point source is spread over a small area in the focal plane. The extent of spreading depends on many factors, including optical diffraction, aberrations, and mechanical assembly quality.

- The image motion PSF (PSF_{IM}), caused by the motion of the carrier during the integration time, which lead normally to rectangular spatial pixel response shapes.

- The detector PSF (PSF_{det}) produces a spatial blurring caused by the non-zero spatial area of each detector in the sensor, and also normally

is not a quadratic shape.

- The electronics PSF (PSF_{el}), appearing by electronic filtering of the acquired data during acquisition, e.g. for correction of dark current or smear effects.

From these four influences, the total PFS can be expressed by a combination of these effects:

$$PSF = PSF_{opt} + PSF_{IM} + PSF_{det} + PSF_{el} \tag{1.3}$$

For pushbroom imaging spectroscopy, one image frame registers the spectral and spatial dimension simultaneously. Any non-uniformity in the system generates degrading artifacts, more in particular:

- *Spectral PSF non-uniformity*: is the non-uniformity of the spectral response within a sensor's spectral band and can be imaged on a detector row as shown in fig.1.5. This non-uniformity is typically represented by the position and shape of the spectral response function. The related artifacts of spectral misregistration are denoted as "smile" or "frown".

- *Spatial PSF non-uniformity*: is the non-uniformity of the spatial response within an acquired spectrum and is usually imaged on a detector column as shown in fig.1.5. This non-uniformity is represented by the position and shape of the spatial response function in both the along-track and across-track dimensions of a spatial pixel. The related artifacts in the across-track dimension are denoted as "keystone".

As for the spatial non-uniformity, the influence of the "keystone" effect results in a black pixel in the image that can be easily replaced. On the other hand the removal of the "smile" effect is not an easy task.

1.2.1 Influence of smile effect on pushbroom sensor

In many cases a pushbroom sensor can be affected by the "smile effect" [2]. The consequence of this effect is that the central wavelength of a band varies with spatial position across the width of the image in a smoothly curving pattern fig.1.6. Very often the peak of the smooth curve tends to be in the middle of the image and give it a shape of "smile" or "frown". That is why this spectral misalignment is termed as smile effect. The effect of the smile is not obvious in the individual bands. Therefore an indicator is needed to make evident whether or not a given image suffers from smile effect. A way to check for the smile effect is to look at the band around atmospheric absorption (760 nm)[reference smile]different images. In fact, the region of red-near infrared transition has high information content of vegetation spectra. This region is generally called "red-edge" (670-780 nm) and identifies the red-edge position (REP) [3] [4]. REP is a good indicator of chlorophyll concentration. Increase in amount of vegetation causes shift in red-edge slope and REP towards longer wavelengths. In contrast, low chlorophyll concentration causes shift in red-edge slope towards shorter wavelengths. The smile effect is acute due to sharp absorption at 760 nm, which is within the red-edge region, and for this reason atmospheric correction of smiled data will be incorrect. Although some research has been done on many hyper-spectral datasets to solve the smile problem, the researchers

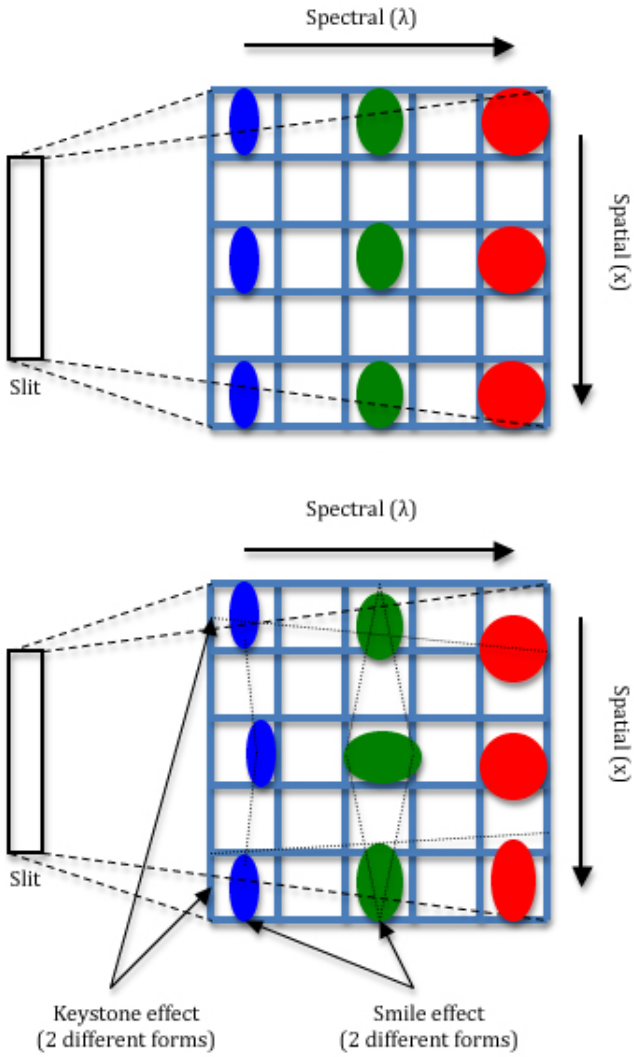


Figure 1.5: *PSF in ideal (top) and real (bottom) position: the smile and keystone effects on a detector*

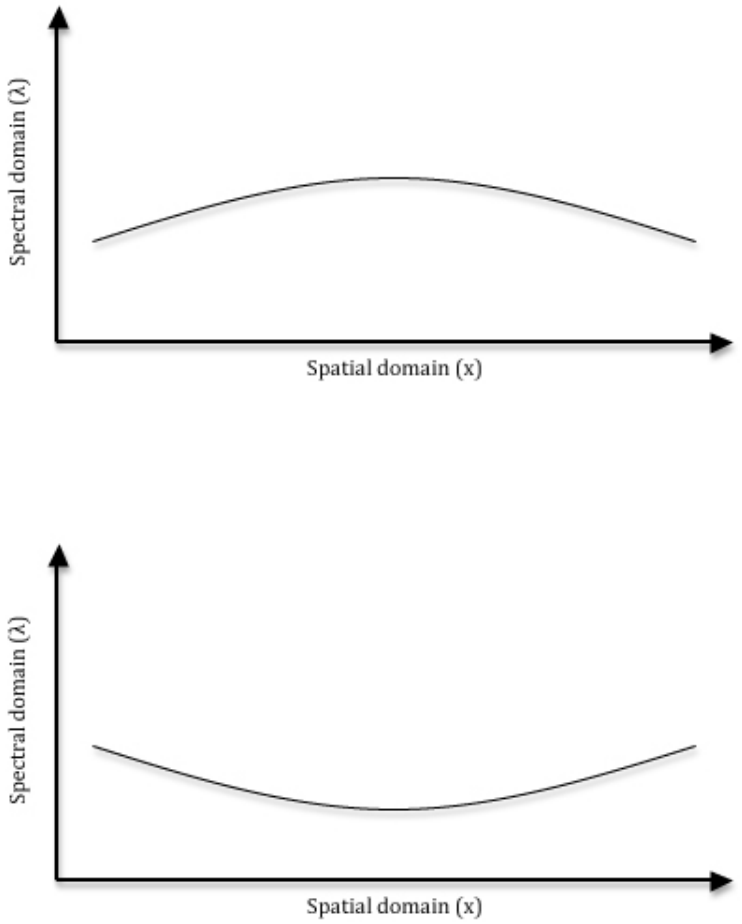


Figure 1.6: *Smooth curves representing the spectral variations along the spatial domain. Frown (top) and smile effect (bottom)*

have yet not come up with a complete solution. The methodologies developed so far can only reduce the intensity of smile effect but cannot remove it entirely, because during its life a detector element can change its response, therefore the knowledge and the correction of this phenomena became fundamental in the analysis of hyper-spectral images and more in general in multispectral images.

In the following paragraph I will list the main hyper-spectral sensors, and describe in details their main features, peculiarities and usage (employment, applications, etc).

1.3 Airborne Hyper-spectral sensors

The latest airborne payloads include sensors with measurements carried out at thousands of wavelengths and at the finest spatial resolution.

1.3.1 CASI

The Compact Airborne Spectrographic Imager (CASI) [5], produced by Itres Research of Canada, is a two-dimensional Charge-Coupled-Device (CCD) array based pushbroom imaging spectrograph.

One dimension of the 578x288 element array is used to obtain a 512 spatial pixels frame of the surface that builds up a flightline of data as the aircraft moves forward. The front side of the CASI camera head is equipped with a custom fore-optic lens with 54.4°FOV which has been designed to provide optimum focusing across the CASI wavelength range (achromatic focus).

After passing through a 15mm wide spectrographic slit, a reflection

Parameter	Description
IFOV	40°
Spectral Range	650nm between 380 and 1050nm
Spatial Samples	512 pixels
Bands	288
Bandwidth	< 3.5 nm
Dynamic Range	14 bits

Table 1.1: *CASI spectral parameters*

Figure 1.7: *CASI 1500 Hyper-spectral Imager*

grating disperses the light from each pixel over the 405nm to 950nm spectral range and is recorded by the 288 detectors on the orthogonal dimension of the CCD. The row spacing on the CCD equates with a spectral sampling of 1.8nm. The effective bandwidth of a single row has an approximate value of 2.2 nm FWHM (Full Width at only Half its Maximum value) at 650nm, resulting from the optical system and convolution of the slit width and detector size.

1.3.2 AHS

The Airborne Hyper-spectral Scanner (AHS) [6] is an 80-bands airborne imaging radiometer, developed by ArgonST (USA) and operating by INTA. It has 63 bands in the reflective part of the electromagnetic spectrum, 7 bands in the 3 to 5 microns range and 10 bands in the 8 to 13 microns region. The first element of the system is a rotating mirror, which directs the surface radiation to a cassegrain-type telescope. The telescope design includes a so-called *pfund-assembly*, that defines a 2.5 mrad IFOV and acts as a field stop. This field is therefore unique for all bands, and redirects the radiation to a spectrometer placed above the telescope. In the spectrometer, four dichroic filters are used to split the incoming radiation in five optical ports: Port 1 (corresponding to VNIR wavelengths), Port 2a (for a single band at 1.6 micrometers), Port 2 (SWIR), Port 3 (MIR) and Port 4 (TIR). For each of the ports, a grating disperses the radiation and a secondary optical assembly focuses it onto an array of detectors, which defines the final set of (contiguous) spectral bands. Table 1.2 displays the resulting spectral configuration.

Parameter	**Description**
IFOV	2.5 mrad
Spectral Ranges	
VIS/NIR (Port 1)	441-1018nm
NIR (Port 2A)	1.491-1.650 μm
NIR (Port 2)	2.019-2.448 μm
MIR (Port 3)	3.03-5.41 μm
LWIR (Port 4)	7.950-13.17 μm
Spatial Samples	750 pixels
Bands	80
Bandwidths	
VIS/NIR (Port 1)	30 nm
NIR (Port 2A)	0.2 μm
NIR (Port 2)	0.013 μm
MIR (Port 3)	0.3 μm
LWIR (Port 4)	0.4-0.5 μm
Dynamic Range	12 bits

Table 1.2: *AHS spectral parameters*

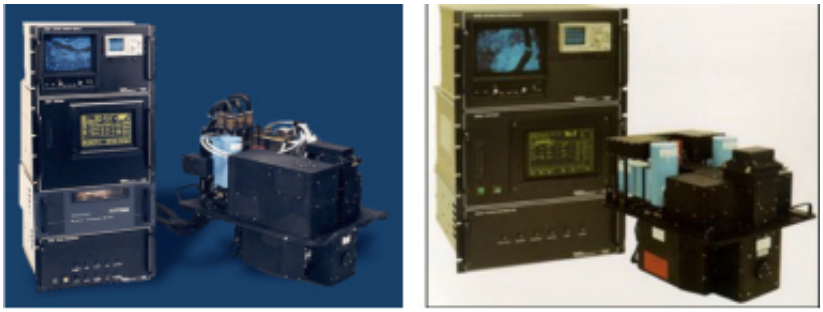


Figure 1.8: *MIVIS (left) and AHS (right) Hyper-spectral Imagers*

1.3.3 MIVIS

MIVIS [7] (Multispectral Infrared and Visible Imaging Spectrometer) is a modular hyper-spectral scanner composed of 4 spectrometers, which simultaneously measure the electromagnetic radiation of the Earth's surface recorded by 102 spectral bands. The instrument can be considered as one of the imaging spectrometers of second generation, that best meets the research needs because it enables advanced applications in environmental remote sensing, like Agronomy, Archaeology, Botanic, Geology, Hydrology, Oceanography, Pedology, Urban Planning, Atmospheric Sciences, and so on.

The simultaneous scanning in a great number of channels with a high spectral and spatial resolution require the highly technological optics and sensors, electronic pre-processing and registration of a large data quantity. The combination of a high resolution in the Mid Infrared region with a good sensitivity in the Thermal Infrared region has caused many problems during the design phase. The resulting system is a mechanical scanning

Parameter	Description
IFOV	2.0 mrad
Spectral Ranges	
VIS	0.43-0.83 μm
NIR	1.15-1.55 μm
MIR	2.0-2.5 μm
TIR	8.2-12.7 μm
Spatial Samples	755 pixels
Bands	102
Bandwidths	
VIS	0.02 μm
NIR	0.05 μm
MIR	0.009 μm
TIR	0.34-0.54 μm
Dynamic Range	12 bits

Table 1.3: *MIVIS spectral parameters*

optical instrument provided with a sensor for each spectral region that collects energy from a common Field Stop for all channels.

1.3.4 AVIRIS

AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) is a premier instrument in the realm of Earth Remote Sensing developed by NASA [8]. AVIRIS contains 224 different detectors, each with a bandwidth of approximately 10 nanometers, allowing it to cover the entire range between 380 nm and 2500 nm. AVIRIS uses a whiskbroom scanning mirror producing 677 pixels for the 224 detectors at each scan. The pixel size and swath width of the AVIRIS data depend on the altitude from which the data is collected.

Parameter	Description
IFOV	1.0 mrad
Spectral Ranges	380-2500 nm
Spatial Samples	677 pixels
Bands	224
Bandwidths	10 nm
Dynamic Range	12 bits

Table 1.4: *AVIRIS spectral parameters*

The ground data is recorded on board the instrument along with navigation and engineering data and the readings from the AVIRIS on-board calibrator. When all of this data is processed and stored on the ground, it yields approximately 140 Megabytes (MB) for every 512 scans (or lines) of data. Each 512 line set of data is called a "scene", and corresponds to an area about 10km long on the ground. Every time AVIRIS flies, the instrument takes several *runs* of data (also known as *flight lines*). A full AVIRIS disk can yield about 76 Gigabytes (GB) of data per day.

1.3.5 ROSIS

ROSIS (Reflective Optics System Imaging Spectrometer) [9], a compact airborne imaging spectrometer, developed jointly by MBB Ottobrunn (now EADS-ASTRIUM), GKSS Geesthacht (Institute of Hydrophysics) and DLR Oberpfaffenhofen (former Institute of Optoelectronics) based on an original design for a flight on ESA's EURECA platform. The design driver for ROSIS was its application for the detection of spectral fine structures especially in coastal waters. This task determined the selection of the spectral range, bandwidth, number of channels, radiometric resolution and its tilt

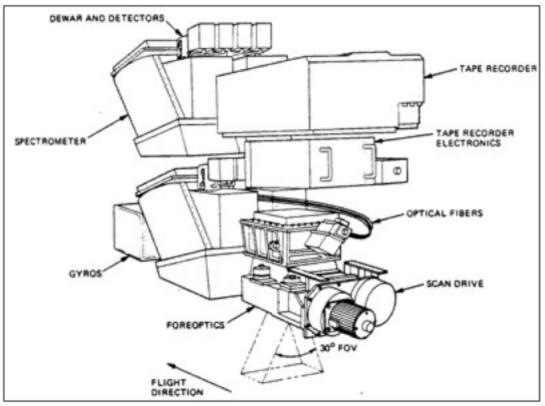


Figure 1.9: *AVIRIS hyper-spectral Imager*

Parameter	**Description**
IFOV	0.56 mrad
Spectral Ranges	430-860 nm
Spatial Samples	512 pixels
Bands	115
Bandwidths	4.0 nm
Dynamic Range	14 bits

Table 1.5: *ROSIS spectral parameters*

capability for sun glint avoidance. However, ROSIS can be exploited for monitoring spectral features above land or within the atmosphere.

1.4 Satellite Hyper-spectral sensors

The development of hyper-spectral technology for the space satellites remains difficult and very expensive in terms of payload design, maintenance and calibration. However, these difficulties have not deterred the space

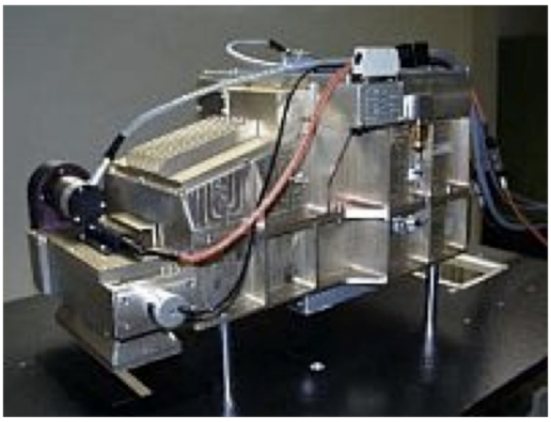


Figure 1.10: *ROSIS instrument*

agencies from finding interesting missions carrying on board hyper-spectral payloads. This is the case of Hyperion developed by NASA, CHRIS Proba-1 developed by a European consortium founded by ESA, and the upcoming PRISMA developed by ASI (Agenzia Spaziale Italiana), and EnMAP developed by DLR.

1.4.1 HYPERION

Hyperion instrument [10], mounted onboard of the National Aeronautics and Space administration (NASA) EO-1 satellite, provides a high resolution hyper-spectral imager capable of resolving 220 spectral bands (from 0.4 to 2.5 μm) with a 30 meter spatial resolution. The instrument covers a 7.5 km by 100 km land area per image and provides detailed spectral mapping across all 220 channels with high radiometric accuracy. The major components of the instrument are the System fore-optics design based

Parameter	Description
Spectral Ranges	410-2500 nm
Bands	220
Bandwidths	10 nm
Spatial resolution	30 m
Swath width	7.5 km

Table 1.6: *HYPERION spectral parameters*

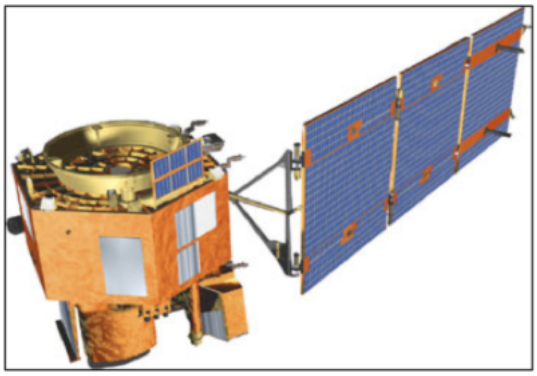


Figure 1.11: *EO1 satellite carrying the HYPERION instrument*

on the Korea Multi-Purpose Satellite (KOMPSAT) Electro Optical Camera (EOC) mission and the telescope that is provided with two different grating image spectrometers, with the purpose of improving signal-to-noise ratio (SNR).

1.4.2 CHRIS

CHRIS (Compact High Resolution Imaging Spectrometer) is a high resolution hyper-spectral sensor installed onboard the PROBA (Project for On-Board Autonomy) satellite, managed by the European Space Agency

Parameter	**Description**
Spectral Ranges	400-1050 nm
Bands	From 19 to 62
Bandwidths	From 5 nm to 11 nm
Spatial resolution	From 34 m up to 1m
Swath width	13 km

Table 1.7: *CHRIS spectral parameters*

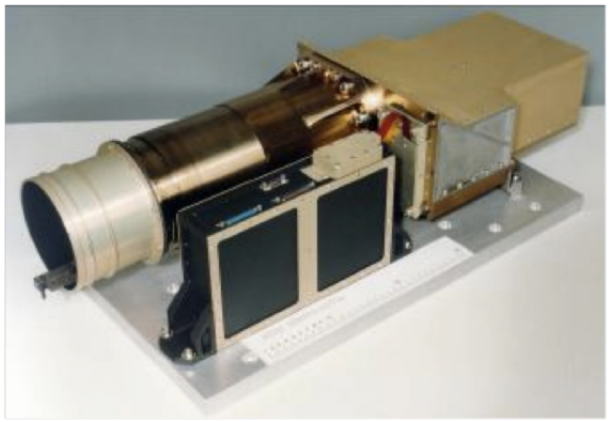


Figure 1.12: *CHRIS instrument*

(ESA) [11]. Distinctive feature of CHRIS is its ability to observe the same area under five different angle of view (nadir, $\pm 55\,^{\circ}$, $\pm 36\,^{\circ}$), in the VIS/NIR bands. CHRIS provides acquisitions up to 62 narrow and quasi-contiguous spectral bands with the spatial resolution of 34-40 meters and a radiometric resolution of 5-10 nm. This device can allow high-resolution observations at 18 meters, using only a subset of 18 spectral bands [9]. CHRIS is also designed to acquire up to 150 spectral bands with a spectral resolution of 1.25nm.

Parameter	Description
Spectral Ranges	400-2500 nm
Bands	210
Bandwidths	10 nm
Spatial resolution	20-30 m
Swath width	30-60 km

Table 1.8: *PRISMA spectral parameters*

1.4.3 PRISMA

PRISMA (PRecursore IperSpettrale della Missione Applicativa) [12] is a new earth observation project led by Agenzia Spaziale Italiana (ASI), integrating a hyper-spectral sensor with a pan-chromatic camera. The advantage of using both sensors is to integrate the classical geometric feature recognition, to the capability offered by the hyper-spectral sensor to identify the chemical/physical feature present in the scene. The primary applications are the environmental monitoring, geological and agricultural mapping, atmosphere monitoring and homeland security. The satellite launch is scheduled in 2011.

1.4.4 EnMAP

EnMAP (Environmental Mapping and Analysis Program) is a German hyper-spectral satellite mission designed to provide high quality hyper-spectral image data on a timely and frequent basis [13]. The main goal of this project is to investigate a wide range of ecosystem parameters encompassing agriculture, forestry, soil and geological environments, coastal zones and inland waters. The EnMAP HYPERSPECTRAL IMAGER (HSI) is a

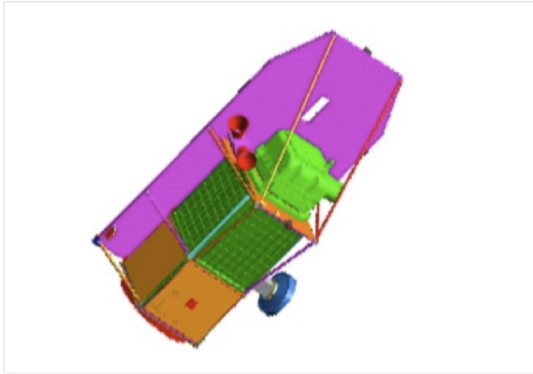


Figure 1.13: *A schematic design of PRISMA*

hyper-spectral imager of pushbroom type working with two separate spectral channels: one for VNIR range from 420 to 1000 nm and one for the SWIR range from 900 to 2450 nm. The channels share a common telescope (TMA) equipped with a field splitter placed on its focal plane. The field splitter features two entrance slits - one for each spectral channel. By placing a micro mirror directly behind the entrance slit of the SWIR channel both channels can be separated and fed into distinct spectrometer branches. Furthermore both spectrometers are designed as prism spectrometers thus providing the highest optical transmission with low polarization sensitivity. The sensor covers a swath width of 30 km, with a 30x30 m Ground Sampling Distance (GSD). Thanks to the chosen sun-synchronous orbit and a $\pm 30^{\circ}$ off-nadir pointing feature, each point on earth can be investigated and revisited within 4 days. In addiction to that, sun-synchronous orbit enables the satellite to pass over any given point of the Earth's surface at the same local solar time, which results in a consistent illumination.

Parameter	Description
Spectral Ranges	420 to 1000 nm VNIR 900 to 2450 nm SWIR
Bands	228
Bandwidths	6.5 nm VNIR 10 nm SWIR
Spatial resolution	30 m
Swath width	30 km

Table 1.9: *EnMAP spectral parameters*

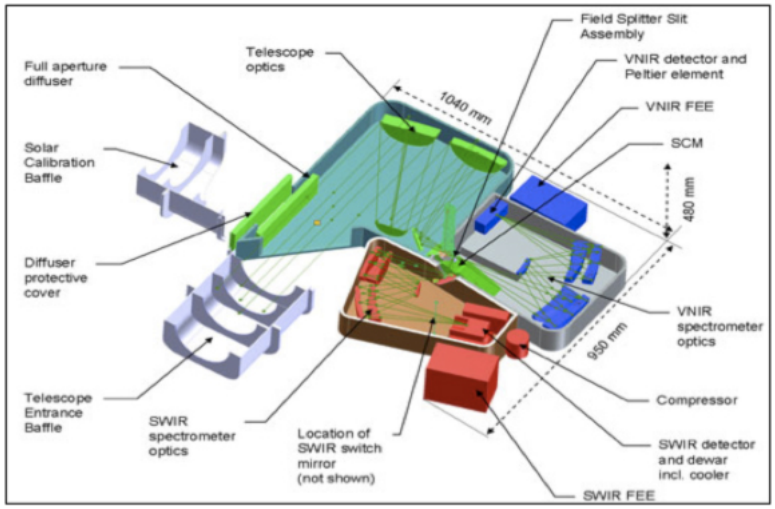


Figure 1.14: *EnMAP sensor scheme*

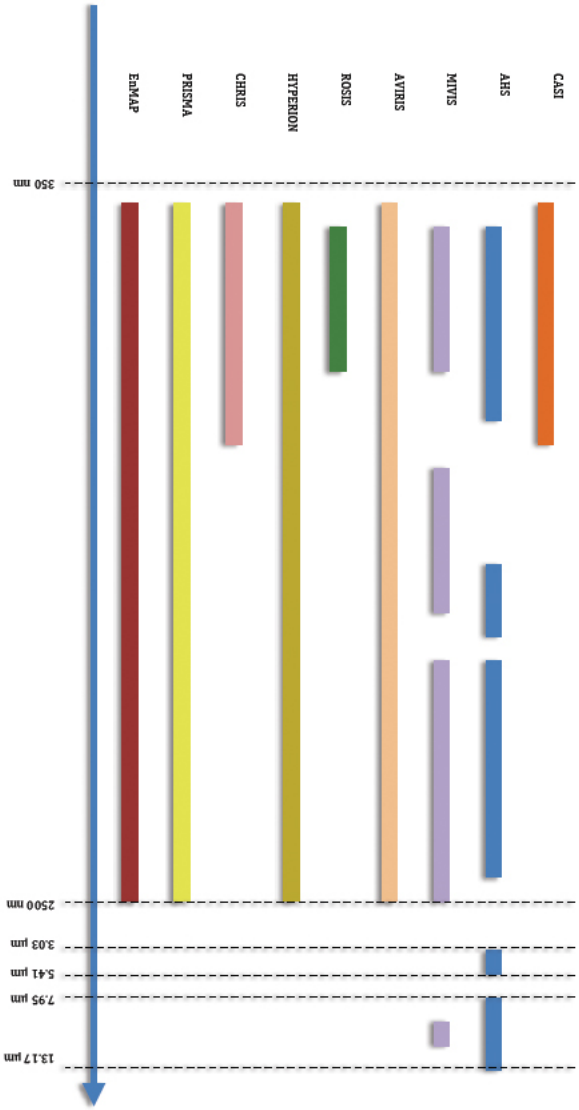


Figure 1.15: *A comparison of sensors bandwidth*

Chapter 2

Information extraction from hyper-spectral data

The great quality and quantity of spectral information provided by last-generation sensors has given ground-breaking perspectives in many applications, such as environmental modeling and assessment, target detection for military and defense/security deployment, urban planning and management studies, risk/hazard prevention and response including wild-land fire tracking, biological threat detection, monitoring of oil spills and other types of chemical contamination. Many of these applications require information extraction techniques, which are algorithms whose goal is conceived to automatically extract structured information, from unstructured machine-readable data.

For hyper-spectral data the design of such algorithms can be very challenging; in particular, the price paid for the accuracy of spatial and spectral information offered by the sensors is the very expensive amounts of data that they generate. For instance, the incorporation of hyper-spectral

sensors on NASA airborne/satellite platforms (AVIRIS, Hyperion) is currently producing a nearly continual stream of high-dimensional data, and it is estimated that NASA collects and sends to Earth more than 950GB of hyper-spectral data every day. Therefore, to develop fast, unsupervised techniques for near real-time information extraction has become a highly desired goal yet to be fully accomplished.

2.1 Data vs information

Information is data within a given context. Without the context, the data are usually meaningless. Once you put the information into meaningful context, you can use it to make decisions. The transfer of information to people who needs it, such as a data analyst or policy maker, can increase the ability of that person to make better decisions. Probably the most important characteristic of good information is its relevance to the problem. Information is usually considered relevant if it helps to improve the decision-making process. If the information is not specific to the problem set, it is irrelevant. Timeliness and accuracy are also strong considerations for the value of the information. Timeliness of data or information is directly related to the gap between the occurrences of the event to the transfer of information to the user. A system is considered *real time* when the gap between data collection and product development (such as target detection) is very short. Accuracy is the comparison of the data to actual events. Many times, a data authentication process is used to determine the validity of the data collected.

In hyper-spectral remote sensing, the ability to derive information from

spectral data is the key to a successful collection. The vast amount of spectral data must be culled to define the spectral signature of interest for the material under consideration. In spectral terms, the pure spectral signature of a feature is called an endmember. One method of collecting pure endmembers is from a laboratory spectro-radiometer that is focused on a single surface or material. These signatures are then used in the spectral sensor, and detection algorithms are used to define and refine the spectral scene collected so a material or materials with similar characteristics can be defined. However, when the material of interest is not available for laboratory measurements, it must be defined within the spectral scene collected. Several information extraction techniques, and some specific developments for hyper-spectral imagery, will be presented and discussed in the following subsections.

2.2 Clustering

Image Clustering can be defined as finding out similar image primitives such as pixels, regions, line elements etc. and grouping them together [14]. Image quantization, segmentation, and coarsening are different classes of image clustering. Image clustering approaches can be broadly categorized to two classes: supervised and unsupervised. Supervised clustering is known as classifications. In unsupervised approach there is no need of specifying the class value by the user. It clusters similar objects according to similarity measures.

An hyper-spectral pixel is generally a mixture of different materials present in the pixel with various abundance fractions. These materials ab-

sorb or reflect within each spectral band. K-MEAN and ISODATA are the most widely used clustering algorithms for hyper-spectral image analysis [15][16][17]. Both algorithms use a spectral-based distance as a similarity measure to cluster data elements into different classes. A drawback of the K-MEAN is that the number of clusters is fixed, so once k is chosen it always return k clusters. On the other hand ISODATA algorithm avoids this problem by removing redundant clusters. However, in high-dimensional spaces as hyper-spectral data are, the data space becomes sparsely populated and the distances between individual data points become very similar.

2.3 Classification

Classification and visualization software requires complex algorithms that are usually not cost effective for the evaluation of ordinary tasks and data sets. An image analyst determines the classification approach and decides between using spectral classes or information classes. A cluster of pixels with nearly identical spectral characteristics is considered part of a spectral class. An analyst uses an information class, such as pine trees, orange trees, or gravel, when trying to identify specific items or groups within an image. The primary goal of an image analyst is to try to match the spectral class to an information class. Once the analyst has decided to use spectral or information classes, the classification process can be either supervised or unsupervised. A supervised classification is based on detection algorithms using pixels from known reference samples, usually located within a scene, as a basis for comparison to other pixels from objects in the same scene. For example, if the analyst knows one specific area is a

gravel road, then all other areas with the same detection algorithm will be a gravel road. Therefore, in supervised classification, the analyst usually starts with known information classes that are then used to define representative spectral classes that closely match the reference samples. In contrast to the supervised classification, unsupervised classification do not require the user to specify any information about the features contained in the images. An unsupervised classification algorithm select natural grouping of pixels based on their spectral properties. However, an unsupervised classification algorithm still requires user interaction, in fact, decisions need to be made concerning which types each category falls within. To make these decisions, other materials and knowledge of the area are useful. Once performed, a classification can be refined considering more specific "themes" or "thematic maps".

In the last years a number of classification algorithms for multispectral image data have been developed [18][19][20][21]. However, with the first appearance of hyper-spectral sensors, the use of the same algorithms became troublesome for two main reasons. First, the training sample of hyper-spectral images at disposal is limited. Secondly, the hyper-spectral data contain a lot of information about the spectral properties of the land cover in the scene. In fact, with the increasing of the dimensionality of the measurements vector, the reliability of a classification algorithm decreases. This effect is better known as the *Hughes Effect* or *curse of dimensionality* [22]. Classification of hyper-spectral data has been discussed in some recent papers dealing with advanced pixel classifiers and feature extraction techniques based on decision boundaries [23][24], features simi-

larity [25][26], morphological transformations [27] and statistical approaches [28][29]. Among these approaches, advanced statistical classifiers such as neural networks and support vector machines (SVM), seem to be rather competitive for hyper-spectral data classification[30]. Moreover, Foody et al. [31] stated that an artificial neural network is less susceptible to the Hughes effect than other approaches. However, a great improvement in the classification accuracy can be extected by reducing the number of inputs, regardless from the adopted classifier. Therefore the dimensionality reduction can be an indispensable operation in the classification task.

2.4 Spectral mixture analysis

The underlying assumption governing the clustering and classification techniques aforementioned is that each pixel vector measures the response of a single underlying material. However, if the spatial resolution of the sensor is not high enough to separate different materials, these can jointly occupy a single pixel: the resulting spectral measurement will thus be a "mixed pixel" i.e., a composite of the individual pure spectra [32]. In order to deal with this problem, spectral mixture analysis techniques first identify a collection of spectrally pure constituent spectra, called *endmembers*, and then define the measured spectrum of each mixed pixel as a combination of *endmembers* weighted by fractions or *abundances* that indicate the proportion of each *endmember* present in the pixel [33][34]. More precisely, in hyper-spectral imagery, mixed pixels are a mixture of distinct substances, and they exist for one of two reasons. First, if the spatial resolution of a sensor is low enough that disparate material can jointly occupy a single pixel, the

resulting spectral measurement will be some composite of individual spectra. Second, mixed pixels can result when distinct materials are combined into homogeneous mixture. This circumstance can occur independently of the spatial resolution of the sensor. The basic premise of mixture modeling is that, within a given scene, the surface is dominated by a small number of distinct materials, all having relatively constant spectral properties, the so-called *endmembers*. If we assume that most of the spectral variability within a scene results from the varying proportions of the *endmembers*, it consequently follows that some combinations of their spectral properties can model the spectral variability observed. If the *endmembers* in a pixel appear in spatially segregated patterns similar to a square checkerboard, these systematics are basically linear. In this case the spectrum of a mixed pixel is a linear combination of the *endmember* spectra weighted by the fractional area coverage of each *endmember* in a pixel. This model can be expressed by:

$$x = \sum_{i=1}^{M} a_i s_i + w = Sa + w \tag{2.1}$$

where x is the received pixel spectrum vector, S is the matrix whose columns are the $M = 1, .., i$ endmembers, a is the fractional abundance vector and w is the additive observation noise vector. Otherwise, if the components of interest in a pixel are in an intimate association, like sand grains of different composition in a beach deposit, light typically interacts with more than one component as it is multiply scattered, and the mixing systematics between these different components are nonlinear. Which process (linear or nonlinear) dominates the spectral signature of mixed pixel is still an unresolved issue. Several applications have demonstrated that the

linear approach is a useful technique for interpreting the variability in remote sensing data [26]. Despite the obvious advantages of using a nonlinear approach for intimate mixtures, it has not been widely applied to remotely acquired data, because the particle size, together with composition, and alteration state of the endmembers are essential controlling parameters of the solutions. For this reason, the Linear Mixing Model is considered to be the most frequently used model for representing the synthesis of mixed pixels from distinct endmembers [32]. The complete linear unmixing problem can be decomposed as a sequence of three consecutive procedures:

- **Dimensionality reduction**: Reduce the dimensionality of the input data vector;
- **Endmember determination**: Estimate the set of distinct (reduced) spectra in the scene;
- **Inversion**: Estimate the fractional abundances of each mixed pixel from its spectrum and the endmember spectra.

As seen for the classification task, the dimensionality reduction seems to be an essential operation also to solve the unmixing problem.

2.5 Geophysical parameter retrieval

In remote sensing data analysis, estimating biophysical parameters is a special relevance task to better understand the environmental dynamics at local and global scales. Geophysical parameter estimation from remotely sensed data has been an outstanding field of research in recent years, and it

is still a challenge for remote sensing scientists all over the world. In the next years, services to users will include production of biophysical parameters at global scales to support the implementation and monitoring of international conventions. In this context, there is an urgent need for more robust and accurate inversion models.

The use of analytical models can represent a first solution but it is characterized by a higher level of complexity and induces an important computational burden. In addition, with such an approach, ill-posed problems are usually encountered and sensitivity to noise becomes an important issue [35]. Consequently, the use of empirical models adjusted to learn the relationship between the acquired data and actual ground measurements has become very attractive. The original attempts introduced general linear models, but they produced poor results since biophysical parameters are commonly characterized by more complex (nonlinear) relationships with the measured reflectances [36]. More sophisticated models were also developed, including exponential or polynomial terms, but these models are often too simple to capture the relationships between remote sensing reflectance and the investigated biophysical parameters. Parametric models have some important drawbacks, which could lead to poor prediction results on unseen (out-of-sample) data. For instance, they assume explicit relationships among variables, and an explicit noise model is adopted. As a consequence, nonparametric and potentially nonlinear regression techniques have been effectively introduced for the estimation of biophysical parameters from remotely sensed images [2]. Nonparametric models do not assume a rigid functional form; they rely on the available data, and no a

priori assumptions on variable relations are made.

In hyper-spectral remote sensing most of the studies dealing with the retrieval of parameters are dedicated to the characterization of vegetation and water [37][38][39][40]. A very specific problem is often addressed and a deep analysis of the retrieval performance provided by possible different techniques is seldom provided. However, among the already attempted approaches, the neural network inversion seems to be among the most promising as shown by [41][42]. Indeed neural networks could result particularly suitable in discovering the subtle pieces of information hidden in the complex spectra measured by the hyper-spectral sensors. On the other hand, in biophysical parameter estimation, few ground measurements are typically available (in contrast to the wealth of unlabeled samples in the image), and also very high levels of noise and uncertainty are present in the data. Hence the use of optimum features extraction techniques is even more necessary when a statistical technique as neural networks is considered to handle the inversion problems with hyper-spectral data.

Chapter 3

Feature reduction of hyper-spectral data

In hyper-spectral data, pixel vectors (or spectra) are commonly defined as the vectors formed of pixel intensities from the same location, across the bands [38]. If we assume that each pixel corresponds to a certain region of the scene surveyed, it will represent the spectral information for that region. Due to the narrow bandwidth and the abundance of observations, the pixel vector for each pixel location resembles a continuous function of wavelengths. This function describes the reflectance of the material for wavelengths within the frequency interval covered by the sensor. However, a dataset composed of hundreds of narrowband channels may cause problems in the:

- acquisition phase (noise),
- storage and transmission phases (data size),

- processing phase (complexity),
- inversion phase (Hughes phenomenon).

Therefore, dimensionality reduction may become a key parameter to obtain a good performance. Many methods have been developed to tackle the issue of high dimensionality and some of them already have been tried on hyper-spectral data [36]. Summarizing, we may say that feature-reduction methods can be divided into two classes: "feature-selection" algorithms (which suitably select a sub-optimal subset of the original set of features while discarding the remaining ones) and "feature extraction" by data transformation (which projects the original feature space onto a lower dimensional subspace that preserves most of the information) [27][43][44]. First analysis suggests that feature selection is a more simple and direct approach compared to feature extraction, and that the resulting reduced set of features is easier to interpret. Nevertheless, extraction methods can be expected to be more effective in representing the information content in lower dimensionality domain.

Feature-selection techniques can be generally considered as a combination of both a search algorithm and a criterion function [45][46]. The solution to feature-selection problem is offered by the search algorithm, which generates a subset of features and compares them on the basis of the criterion function. From a computational viewpoint, an exhaustive search for the optimal solution becomes intractable even for moderate values of features [47]. In addition, computational saving is not enough to make it feasible for problems with hundreds of features. Despite these apparent difficulties, many feature selection approaches have been developed

[48][49][50]. The *sequential forward selection* (SFS) and the *sequential backward selection* (SBS) techniques [47][50], are the simplest suboptimal search strategies: they can identify the best feature subset achievable by adding (to an empty set in SFS) or removing (from the complete set from SBS) one feature at a time, until the desired number of features is achieved. Both methods does not allow backtracking, in fact, once a feature is selected in a given iteration, it cannot be removed in any successive iteration. The *sequential forward floating selection* (SFFS) and the *sequential backward floating selection* (SBFS) methods improve the standard SFS and SBS techniques by dynamically changing the number of features included (SFFS) or removed (SBFS) at each step and by allowing the revision of the features included or removed at the previous steps [50][51]. Several other methods based on interesting concepts were also explored in the literature: feature similarity measures [35], graph searching algorithms [52], neural networks [53], support vector machines [54], genetic methods [55][56][57][58], simulated annealing [59], finite mixture models [60][61], "tabu search" metaheuristics [62], spectral distance metrics [50], parametric feature weighting [63], and spatial autocorrelation and band ratioing [64][43].

A feature-extraction technique aims at reducing the data dimensionality by mapping the feature space onto a new lower-dimensional space. Both supervised and unsupervised methods have been developed. Unsupervised feature-extraction methods do not require any prior knowledge or training data, even though are not directly aimed at optimizing the accuracy in a given classification task. The class comprises the "principal component analysis" (PCA) [65][66], where a set of uncorrelated transformed features

is generated, the "independent component analysis" [67], a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signal, and the "maximum noise fraction" [68], where an operator calculates a set of transformed features according to a signal-to-noise ratio optimization criterion. Further unsupervised transforms are reviewed in [69]. On the other hand a supervised feature extraction technique directly takes into account the training information available for the solution of a given supervised classification problem. Three main approaches based on discriminant analysis, decision boundary analysis, and correlated feature grouping, have been proposed. The first one is based on the maximization of a functional (i.e., the Rayleigh coefficient) expressed as the ratio of a between-class scatter matrix to an average within-class scatter matrix [65] [69]. This technique has some drawbacks, such as the possibility of extracting at most $(C-1)$ features, where C is the number of classes. The second approach employs information about the decision hyper-surfaces associated with a given parametric Bayesian classifier to define an intrinsic dimensionality useful for the classification problem and the corresponding optimal linear mapping. A third strategy consists of grouping the original features into subsets of highly correlated features to transform the features separately in each subset [30][70][71]. Further techniques, based on image processing approaches, have been proposed in [33][44][71][72][73].

One of the main topics of this thesis work is the development of a novel unsupervised feature extraction procedure based on neural networks (NN). NN are already recognized to represent a rather competitive family of al-

gorithms for the analysis of hyper-spectral data [30]. In fact, they have already been successfully applied for the design of one of the first end-to-end processing scheme dedicated to hyper-spectral imagery provided by the Compact High-Resolution Imaging Spectrometer (CHRIS) on board of the Project for On-Board Autonomy (PROBA) satellite [74]. Although their promising potential, the application of NN to feature extraction in the processing of hyper-spectral data has not been investigated yet. For this purpose, in this work we consider the autoassociative neural networks AANN, which can be seen as a method to generate nonlinear features from the data under analysis, hence to contribute to minimize overfitting problems associated to high dimensionality [75]. The AANN are of a conventional type, featuring feedforward connections and sigmoidal nodal transfer functions, trained by backpropagation or similar algorithms [44]. The particular network architecture used employs three hidden layers, including an internal "bottleneck" layer of smaller dimension than either input or output. The network is trained to perform the identity mapping, where the input is approximated at the output layer. Since there are fewer units in the bottleneck layer than the output, the bottleneck nodes must represent or encode the information in the inputs for the subsequent layers to reconstruct the input. Hence a feature extraction from the input vector is performed. In the following, before introducing AANN, I'll briefly describe two among the most common unsupervised techniques: the "principal component analysis" (PCA) and the "minimum noise fraction" (MNF). In fact, the performance given by these techniques will be considered for an assessment of the results obtained by the application of AANN to hyper-spectral data.

3.1 Principal component analysis

PCA, also known as the Karhunen-Loeve (K-L) transformation, uses a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components [65][66]. The conventional PCA techniques rely on eigenvector expansion stemming from the variance-covariance matrix describing the variability of the observed quantity. Mathematically if $X^T = [X_1, X_2, ..., X_N]$, where T denotes transpose of matrix, is a N-dimensional random variable with mean vector M, the covariance matrix $[B]$ associated to the unknown vector $[X]$ can be evaluated. The generic element of such a matrix is:

$$B_{ij} = \langle X_i X_j \rangle \tag{3.1}$$

Then a new set of variables, Y_1, Y_2,..., Y_N, known as principal components, can be calculated by:

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + ... + a_{Nj}X_N \tag{3.2}$$

where

$$a_j^T = [a_{1j} + a_{2j} + ... + a_{Nj}] \tag{3.3}$$

are the normalized eigenvectors of the covariance matrix $[B]$, solution of the eigenvalue problem:

$$[B][u] = \lambda[u] \tag{3.4}$$

Fig.3.1 shows an example of a PCA reduction from e to 2 dimensions. Given a set of points (*a, b,...,z*) in a 3-dimension Euclidean space ($S1, S2, S3$), the first principal component $PCA1$ (the eigenvector with the largest eigenvalue) corresponds to a line that passes through the mean and minimizes the sum of squared error with those points. The second principal component $PCA2$ corresponds to the same concept after all the correlation with the first principal component has been subtracted out from the points.

The principal component transformation has several interesting characteistics:

- The total variance is preserved in the trasformation i.e.

$$\sum_{i=1}^{N} \sigma_i^2 = \sum_{i=1}^{N} \lambda_i \tag{3.5}$$

  where σ_i^2 are variances of the original variables and λ_i the eigenvalues of B with $\lambda_1, \lambda_2, ..., \lambda_N$,.

- It minimizes the mean square approximation error.

- In a geometrical sense, the transformation may rotate highly correlated features in N-dimensions to a more favourable orientation in the feature space, where components are till orthogonal to each other, such that maximum amount of variance is accounted for in decreasing magnitutde along the ordered components.

The applicability of PCA is limited by the assumptions made in its derivation. These assumptions are:

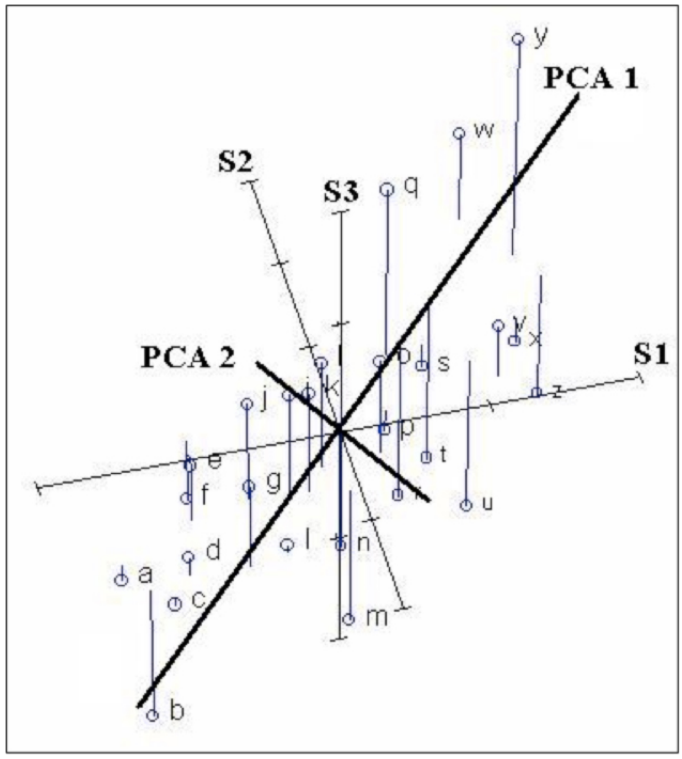


Figure 3.1: *A PC projection from 3-dimensional space to 2-dimensional space*

Assumption on Linearity : It is assumed that the observed data set is linear combinations of certain basis.

Assumption on the statistical importance of mean and covariance: PCA uses the eigenvectors of the covariance matrix and it only finds the independent axes of the data under the Gaussian assumption. For non-Gaussian or multi-modal Gaussian data, PCA simply de-correlates the axes. There is no guarantee that the directions of maximum variance will contain good features for discrimination.

Assumption that large variances have important dynamics: PCA simply performs a coordinate rotation and scaling that aligns the transformed axes with the directions of maximum variance. It is only when we believe that the observed data has a high signal-to-noise ratio that the principal components with larger variance correspond to interesting dynamics and lower ones correspond to noise.

3.2 Minimum noise fraction

Minimum Noise Fraction (MNF), also called Maximum Noise Fraction [68], has been used to determine the inherent dimensionality of image data removing noise from the image, and to reduce the computational requirements for subsequent processing. The signal-to-noise ratio is one of the most common measures of image quality, thus, instead of choosing new components to maximize variance, as the principal components transform does, it is preferred to choose them in order to maximize the signal-to-noise ratio. This technique can be viewed as a two cascaded principal components transformation. The first transformation, based on the estimated

noise covariance matrix, decorrelates and rescales the noise in the data so that the noise has unit variance and no band-to-band correlations. At this stage, the information about between-band noise is not considered. The second step is a standard principal components transformation of the noise-whitened data that takes into accounts the original correlations and creates a set of components that contains weighted information about the variance across all bands in the raw data set. The algorithm retains specific channel information because all original bands contribute to the weighting of each component. Often, most of the surface reflectance variation in a data set can be explained in the first few components, with the rest of the components containing variance as contributed primarily by noise [76]. Weighting values for each component can also be examined, pointing to the raw bands that are contributing most to the information contained in the dominant components. The transformation can be defined in the following way. Let us consider a multivariate dataset of p-bands with grey levels $Z_i(x)$, $i = 1, ...p$, where x gives the coordinate of the sample. We can assume that:

$$Z(x) = S(x) + N(x) \tag{3.6}$$

where

$$Z^T(x) = \{Z_1(x), Z_2(x), ..., Z_p(x)\} \tag{3.7}$$

And $S(x)$ and $N(x)$ are the uncorrelated signal and noise components of $Z(x)$. Thus the covariance of $Z(x)$ is defined by:

$$cov\{Z(x)\} = \Sigma_Z = \Sigma_S + \Sigma_N \tag{3.8}$$

Where Σ_S and Σ_N are the covariance matrices of $S(x)$ and $N(x)$ respectively. The noise fraction NF of the *i-th* band can be defined as the ratio between the noise variance $var\{N_i(x)\}$ and the total variance for that band $var\{Z_i(x)\}$:

$$NF_i = var\{N_i(x)\}/var\{Z_i(x)\} \tag{3.9}$$

The maximum noise fraction (MNF) transform chooses linear transformations:

$$Y_i(x) = a_i^T Z(x) \quad i = 1, ...p \tag{3.10}$$

In such a way that the noise fraction for $Y_i(x)$ is maximum among all linear transformations orthogonal to $Y_j(x)$, $j = 1, ..., i$. As for the derivation of principal components, it can be shown that the vectors a_i are the left-hand eigenvectors of $\Sigma_N\Sigma^{-1}$ and that μ_i, the eigenvalue corresponding to a_i equals the noise fraction in Yi(x). Hence, from the definition of the MNF transform, we see that

$$\mu_1 \geq \mu_2 \geq ... \geq \mu_p \tag{3.11}$$

and so the MNF components will show steadily increasing image quality (unlike the usual ordering of principal components). The first step in MNF transformation is to calculate the noise covariance matrix, which can be estimated from either the dark reference measurements (Dark Current)

or the near neighbor differences. The former is the signal observed while the foreoptics shutter of the detector is closed. It represents the detector's background data and the instrument's noise [77]. In the radiometric calibration processing of hyper-spectral data, the total dark current could be derived by subtracting the dark current values of each channel from the DN values [78]. Most instruments do not produce the dark image, therefore a valid alternative could be to use the near-neighbor differences, which can be calculated from a procedure known as minimum/maximum autocorrelation factors (MAF). This procedure assumes that the signal at any point in the image is strongly correlated with the signal at neighbor pixels while the noise shows only weak spatial correlations [79]. It can be assumed that the eigenvectors a are normed so that:

$$a_i^T = \sum a_i = 1 \quad i = 1, ..., p \tag{3.12}$$

It is also convenient at certain points to express the MNF transform in the matrix form:

$$Y(x) = A^T Z(x) \tag{3.13}$$

where

$$Y^T\{Y_1(x), Y_2(x), ...Y_p(x)\} \tag{3.14}$$

and

$$A = \{a_1, a_2, ..., a_p\} \tag{3.15}$$

Two of the most relevant properties of the MNF transform (not shared by principal components) are: the scale invariance, because it depends on signal-to-noise ratios, and the ability to orthogonalizes $S(x)$ and $N(x)$, as well as $Z(x)$. If we want to obtain the MNF transform, we need to know both Σ_Z and Σ_N. In many practical situations, these covariance matrices are unknown and need to be estimated. Usually, Σ_Z is estimated through the sample covariance matrix of $Z(x)$.

Once data have been transformed with decreasing noise fraction (increasing S/N ratio), it is logical to spatially filter the noisiest components and subsequently to transform back to the original coordinate system. It has been demonstrated that MNF successfully orders components with reference to image quality, unlike the PCA, which could not reliably separate signal and noise components [68][35].

3.3 Nonlinear principal component analysis

Nonlinear principal component analysis (NLPCA) is commonly seen as a nonlinear generalization of standard principal component analysis. Multilayer neural networks can themselves be used to perform nonlinear dimensionality reduction of the input space, overcoming some of the limitations of linear principal component analysis. Consider a multi-layer perceptron of the form shown in fig.3.2 having d inputs, d output units and a *bottleneck* layer of M hidden units, with $M < d$ [60]. The targets used to train the network are simply the input vectors themselves, so that the network is attempting to map each input vector onto itself.

Due to the reduced number of units in the hidden layer, a perfect re-

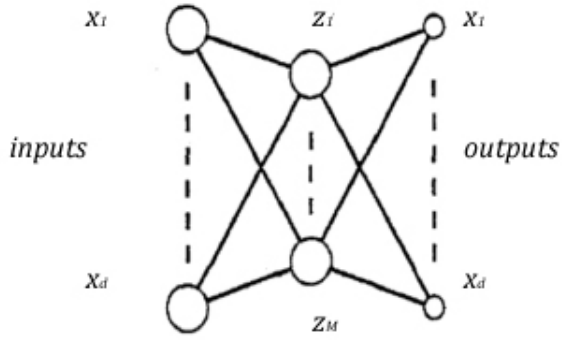


Figure 3.2: *An autoassociative multi-layer neural network*

construction of all input is not in general possible. However, if the network training finds an acceptable solution, a good reduction of the input data must exist in the bottleneck layer. The network can be trained by minimizing the sum-of-squares error of the form:

$$E = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{d}\{y_k(x^n) - x_k^n\}^2 \tag{3.16}$$

where y_k $(k = 1, 2, ..., d)$ is the output vector. Such a network is said to form an autoassociative mapping. In this case, error minimization represents a form of unsupervised training, due to the fact that no independent data is provided. The limitations of a linear dimensionality reduction, such as PCA, can be overcome by using nonlinear (sigmoidal) activation functions for the hidden units in the network. Let's consider the topology in fig.3.2, only with the bottleneck as hidden layer between inputs and outputs. If the nodes of this layer were linear, the projection into the M-dimensional

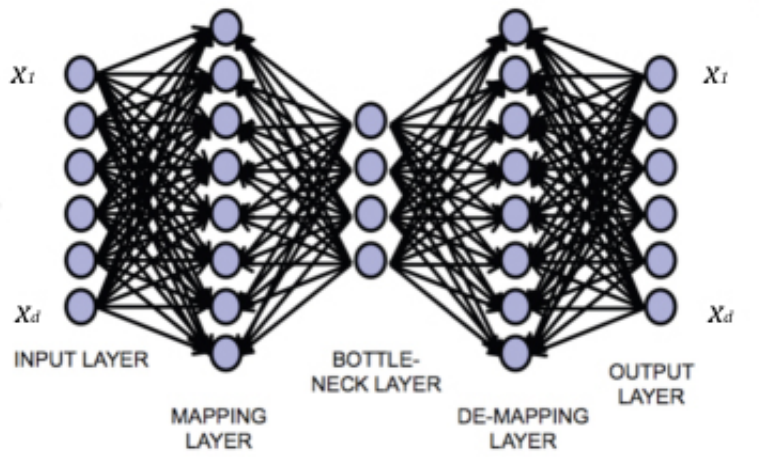


Figure 3.3: *A three hidden layer autoassociative neural network*

subspace would correspond exactly to linear PCA [80]. Also if the activation functions in the *bottleneck* nodes were sigmoidal, the projection into the sub-space would still be severely constrained; only linear combinations of the inputs compressed by the sigmoid into the range [-1,1] could be represented. Therefore the performance of an autoassociative neural network with only one internal layer of sigmoidal nodes is often no better than linear PCA [81]. Starting from these premises, Kramer demonstrated that to perform an effective NLPCA, exactly one layer of sigmoidal nodes and two layers of weighted connections are required [75], as depicted in fig.3.3.

Such a network effectively performs a nonlinear principal component analysis, having the advantage of not being limited to linear transformation, although it contains standard principal component analysis as a special case. Moreover, the dimensionality of the sub-space could be specified

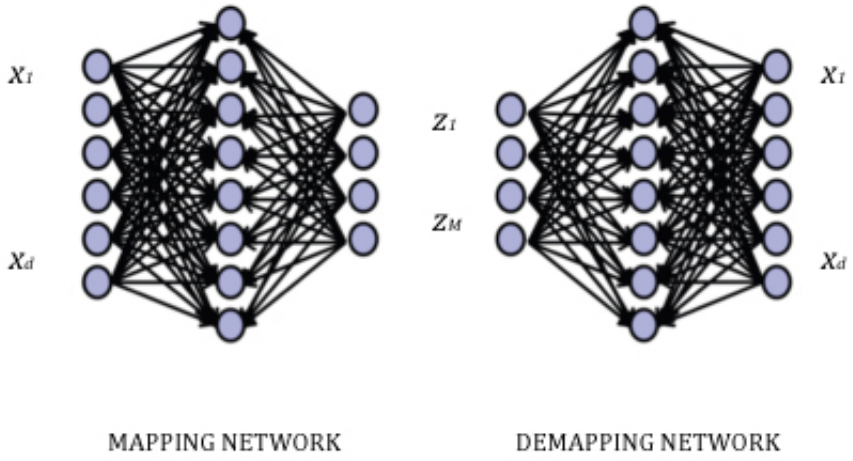


Figure 3.4: *The mapping and demapping layers of an autoassociative neural network*

in advance of training the network. As the NLPCA method finds and eliminates nonlinear correlations in the data, analogous to principal component analysis, this method can be used to reduce the dimensionality of a given data by removing its redundant information. Our intent is to apply this methodology to perform the dimensionality reduction of the large measurements vector typical of the hyper-spectral data. As illustrated in fig.3.4, the network can be viewed as two successive functional mapping networks.

The first mapping network projects the original d dimensional data onto a lower dimensional sub-space defined by the activations of the units in the bottleneck layer. Because of the presence of nonlinear units, this mapping is essentially arbitrary, and in particular not restricted to being linear. Similarly the second demapping network defines an arbitrary mapping from

the lower dimensional space back into the original d-dimensional space. Once the sum-of-squares error has reached its minimum, the projection in the low dimensional feature space was obtained from the outputs of the mapping layer. As the number of nodes in the input and output layers, as well as in the bottleneck layer, can be considered as a fixed parameter, the only varying value in the design of a autoassociative neural network are the number of nodes in the mapping layers. However, there is no definitive method for deciding a priori the dimensions of these layers. The number of mapping nodes is related to the complexity of the nonlinear functions that can be generated by the network. If too few mapping nodes are provided, accuracy might be low because the representational capacity of the network is limited. However, if there are too many mapping nodes, the network will be prone to "over-fitting". In the following chapters, the use of the NLPCA, intended as a dimensionality reduction technique of very different types of hyper-spectral data, will be described for different cases.

Chapter 4

Land cover maps from AHS dataset

4.1 Introduction

In this section we present the assessment of a methodology based on AANN with respect to more standard features extraction approaches such as Principal Component Analysis (PCA) and MNF (Minimum Noise Fraction). The study has been carried out for a set of hyper-spectral data collected by the Airborne Hyper-spectral line-Scanner radiometer (AHS) over a test site in Northeast Germany. This is a test area for which an extensive ground-truth was also available. The results have been quantitatively evaluated and critically analyzed either in terms of their capability of representing the hyper-spectral data with a reduced number of components or in terms of the accuracy obtained on the final derived product. This latter consists of a land cover classification map performed using another NN scheme, this time with the standard topology of a Multi-Layer Perceptron (MLP),

having as input the reduced vector provided by the AANN. It has to be observed that, dealing with a NN classification scheme, features extraction assumes an even more crucial role. Minimizing the number of inputs of a NN, avoiding significant loss of information, generally affects positively its mapping ability and computational efficiency. A network with fewer inputs has fewer adaptive parameters to be determined, which need a smaller training set to be properly constrained. This leads to a network with improved generalization properties providing smoother mappings. In addition, a network with fewer weights may be faster to train. All these benefits make the reduction in the dimension of the input data a normal procedure when designing NN, even for a relatively low dimensional input space.

4.2 Data and methodology

The potential of AANN has been investigated for a set of data acquired by the INTA-AHS instrument, in the framework of the ESA AGRISAR measurement campaign [91]. The test site is the area of DEMMIN (Durable Environmental Multidisciplinary Monitoring Information Network). This is a consolidated test site located in Mecklenburg-Western Pomerania, in Northeast Germany, which is based on a group of farms within a farming association covering approximately 25,000 ha. The field sizes are very large in this area, in average 200-250 ha. The main crops grown are wheat, barley, rape, maize and sugar beet. The altitudinal range within the test site is around 50 m. The AHS has 80 spectral channels available in the visible, shortwave infrared and thermal infrared. The data processing level is the L1b (at-sensor radiance): the VIS/NIR/SWIR bands were converted to at

sensor radiance applying the absolute calibration coefficients obtained in the laboratory. The MIR/TIR bands were converted to at-sensor radiance using the information from the onboard blackbodies and the spectral responsivity curves obtained by the AHS spectral calibration. The resulting files were converted to BSQ format + ENVI header and scaled to fit an unsigned integer data type. In this paper the acquisition taken on the 06/06/2006 has been considered. At that time 5 bands in the SWIR region became blind due to loose bonds in the detector array so they were not used in this study [82].

A double-stage processing has been applied to the data. In the first stage a features reduction has been performed by means of AANN, in the second stage the reduced measurement vector has been used as input to a new NN scheme for a pixel-based classification procedure. It has to be noted that the vector reduction should positively affects the training of the classification neural network algorithm under two points of view: it reduces both over-fitting and learning time. As far as the feature extraction is concerned, the comparison with PCA and MNF techniques has been carried out using processing libraries available within the ENVI package. Another important aspect regards the design of the network topologies, in particular how many units should be considered in the hidden layers. This is a crucial issue because using a too little number of units may weaken the capability of the network to perform the desired mapping. On the other hand, over-dimensioned hidden layers may lead to poor generalization properties. In this study, different strategies have been adopted. However, most of the efforts have been dedicated to the design of the AANN, being this aspect

the focus of the work. The number of neurons in the bottleneck layer was guided by the necessity of comparing the features extraction performance of different approaches. A preliminary analysis based on PCA was carried out to determine the number of PCA components containing most of the statistical information (more than 99%). For the sake of consistency, the same number was also considered for the NLPCA, hence for the units in the bottleneck layer, and for the MNF. The decision on the size of the intermediate hidden layers was based on a more extended analysis where the size was systematically varied and the corresponding autoassociative network mapping MSE error evaluated.

The size minimizing such an error was selected for determining the networks topology. Finally, for the network dedicated to the classification scheme, a more soft strategy among those already existent in literature has been chosen. In particular we considered the rule used by Palmason et al. [83] suggesting that the number of neurons in the hidden layer should be set as geometrical mean of the number of inputs and outputs, i.e., the square root of the product of the number of input features and the number of information classes.

4.3 Results

4.3.1 Feature extraction

According to Kramer [75], the AANN has been trained considering all pixels in the image (2061000). After about 2000 epochs no significant decrease in the error could be noted so the training phase was stopped at that stage. From the PCA analysis it resulted that the first 5 PCA components con-

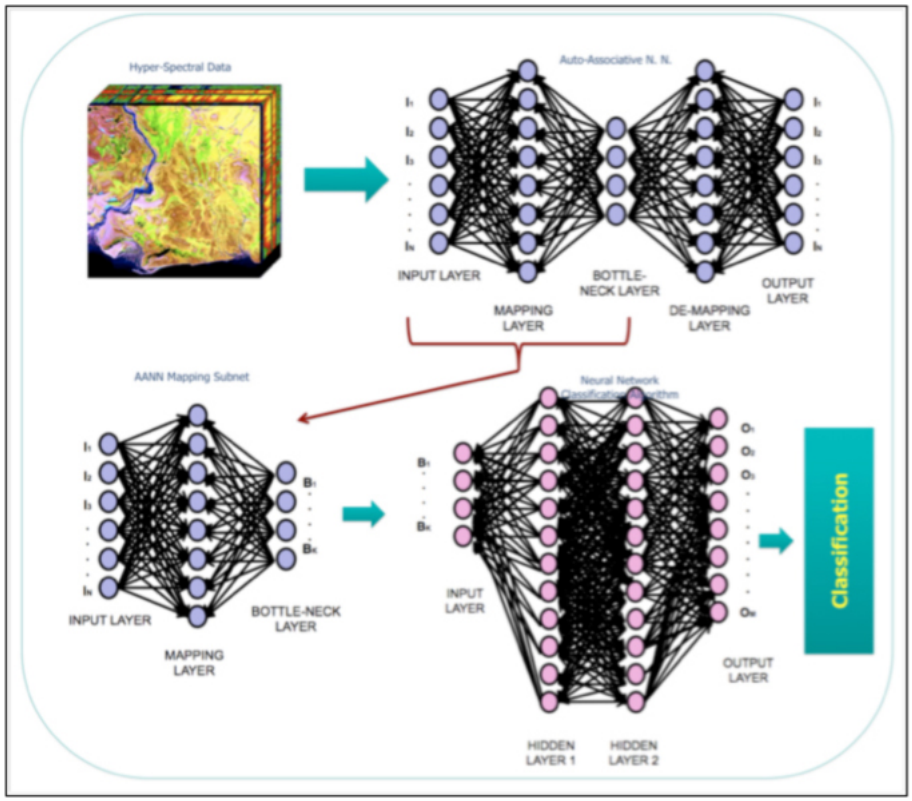


Figure 4.1: *Complete feature reduction and classification scheme*

tained almost the 99.9% of the whole statistical information and for this reason 5 components has been considered as a benchmark for the comparison. The dimension of the outer hidden layers of the AANN, as explained previously, was established after a numerical analysis aiming at minimizing the mean square error (MSE), where, if E is given by expression (3.16) introduced in section 3, MSE is :

$$MSE = \sqrt{\frac{E}{N}} \tag{4.1}$$

With N number of training patterns. The plot of fig.4.2 shows the result of this investigation and motivates the choice of 25 units for the two intermediate layers. This mean that the best topology for the AANN is composed by 75 nodes for both input and output layers, 25 nodes for the mapping and de-mapping hidden layers and 5 nodes for the bottleneck layer. The Stuttgart neural network simulator (SNNS) [84], made available by the University of Stuttgart, Germany, has been used to design the network topology and perform the learning phase.

In fig.4.3, fig.4.4 and fig.4.5, we show the five components for each feature extraction method. It can be seen that the MNF components are clearly disturbed, especially the first one, by the "smile" or "frown" effect introduced in Section 1. In fact, due to the intrinsic light dispersion properties of grating spectrometers and to minor misalignments of optical components, the wavelengths for pixels near the center of an array and those near the edge of the same array can be slightly different [2]. Conversely, NLPCA technique appears to be rather robust to this type of noise while a slightly disturbing pattern due to the smile effect might be the cause

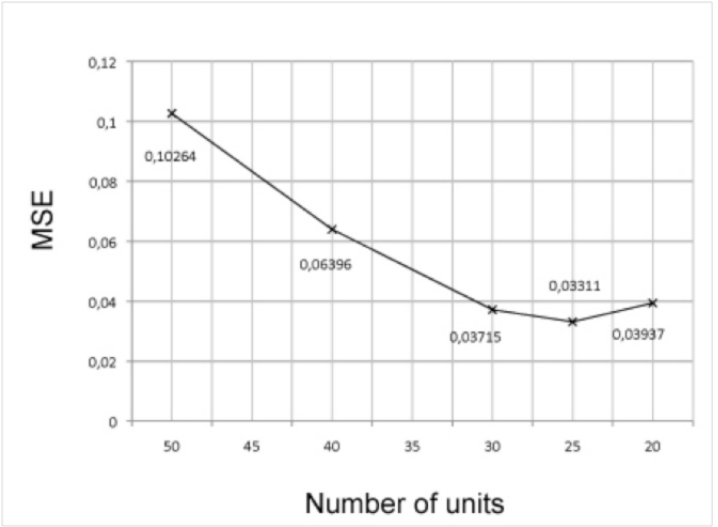


Figure 4.2: *Plot of the number of units depending on the MSE*

of a brighter area in the right side if the first PCA component. Another comment regards a different behavior between PCA and NLPCA. NLPCA, with respect of PCA, seems to be more consistent in representing similarities in spectral signatures among pixels, hence in extracting objects. An example is reported in fig. 4.6. Here we considered the big bright object in the center: a) as it looks like considering an RGB image obtained using the original band 11, band 9 and band 8, b) as it appears in the 5^{th} PCA component and c), as it appears in the 5^{th} NLPCA component. We see that, at the 5^{th} component level, the considered object can be still clearly detected by the NLPCA while, according to PCA, its response is more similar to adjacent fields. We put the following explanation for this result: at the level of the 5th component the statistical content as derived from PCA is

0.11% and such an analysis starts being conditioned by low order processes that may be not characteristic of a specific crop. NLPCA, thanks to its capabilities of investigating about nonlinear dependencies among the data, may be better capable in all its components of coding physical behavior that are peculiar of a specific agricultural field.

In fig. 4.7 and fig 4.8 we report on different type of analysis consisting in investigating on the capabilities of reconstructing the original spectra starting from the extracted features. Two signature samples reconstructed considering only 5 components are shown. The figures have been obtained by averaging over pixels of the same area of interest. The two considered land cover types are water and trees. Also in this case the MNF performance is lower than NLPCA and PCA. In the case of water (about 650 pixels) it can be noted that the NLPCA is significantly more effective than PCA in encoding the spectral information. In particular the behavior in the visible and near infrared with strong curvatures is better resembled. Differently, the forest case (again about 650 pixels) is an example where the two techniques are rather comparable, even though the PCA shows slight discrepancies with the true spectra in the long infrared bands. Similar trends have been observed for the other land cover types.

4.3.2 Classification

The 5 components extracted from each pixel spectral signature have been used for the implementation of a pixel-based classification algorithm. Also the classification task is based on a NN procedure. In this case the network topology is a MLP of 5-25-25-10. About 450.000 pixels were considered for

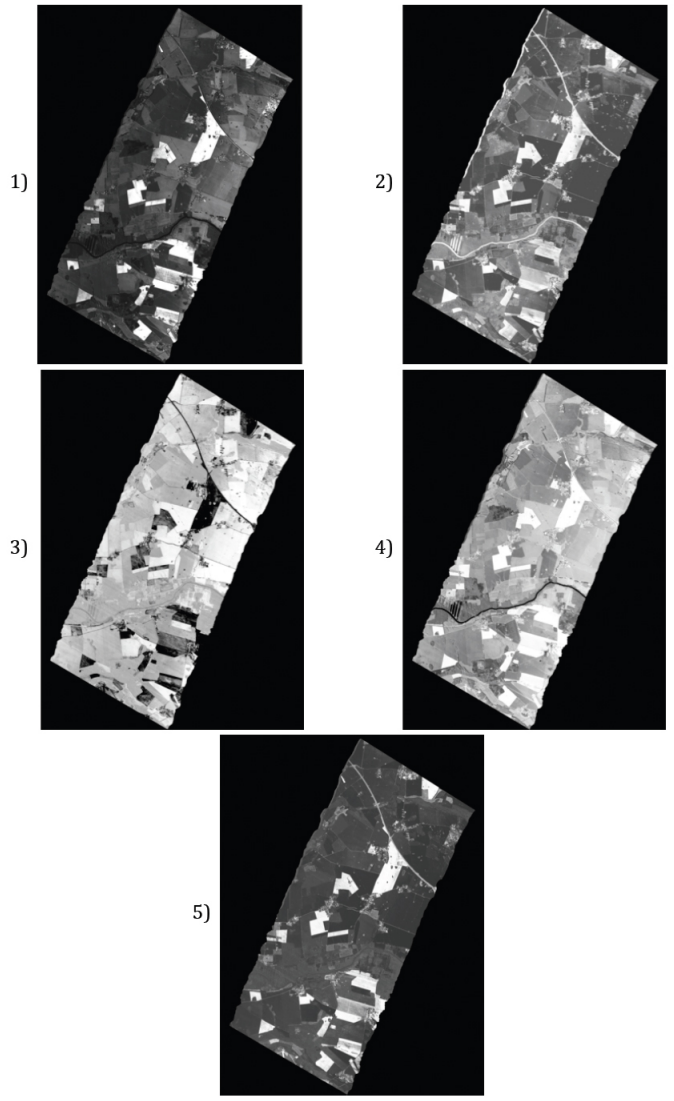


Figure 4.3: *5 nonlinear principal components derived from NLPCA*

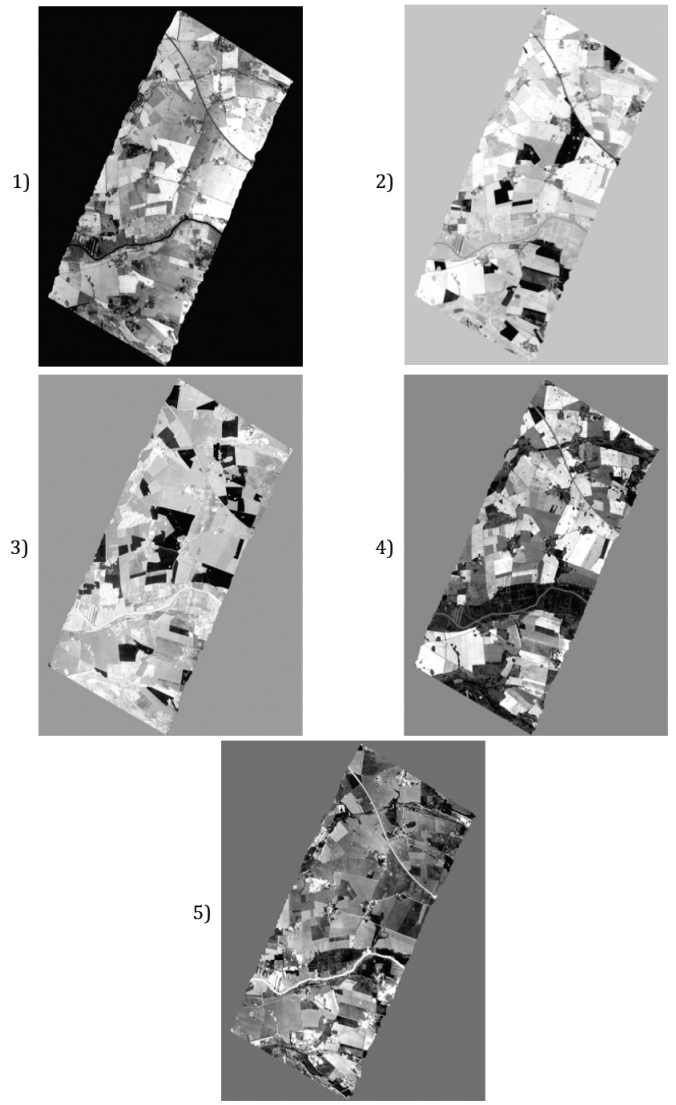


Figure 4.4: *The first 5 principal components derived from PCA*

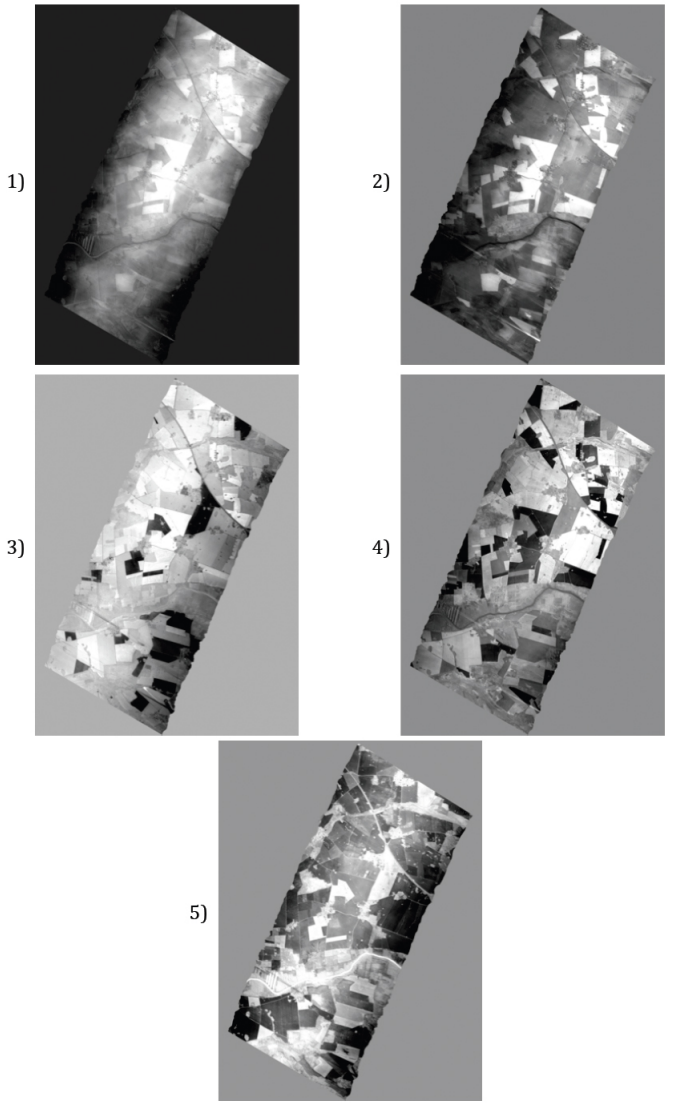


Figure 4.5: *The first 5 components derived from MNF transformation*

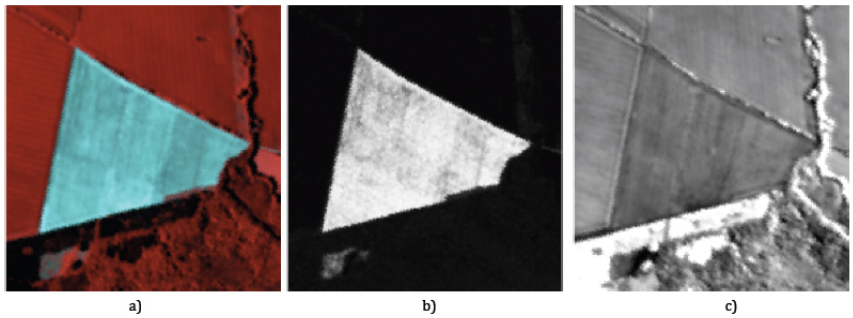


Figure 4.6: *a) RGB (bands 11, 9, 7) detail of the original AHS image; b) 5^{th} NLPCA component; c) 5^{th} PCA component*

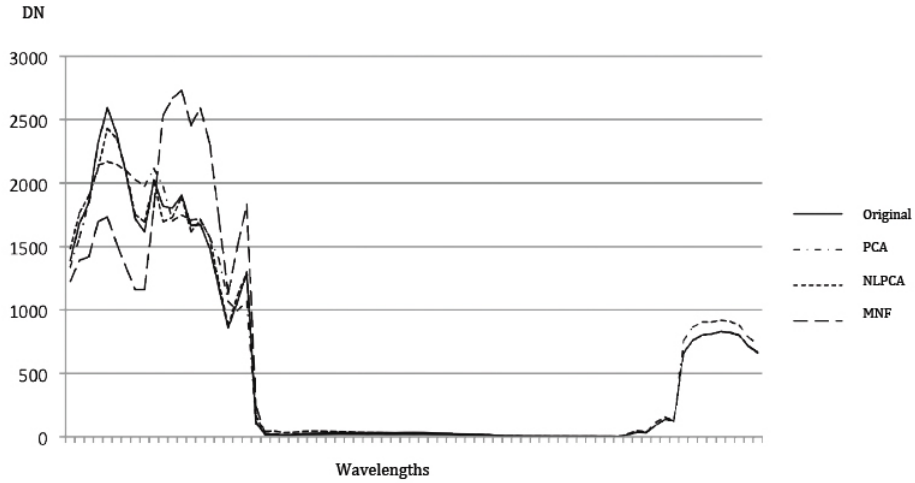


Figure 4.7: *Original and reconstructed spectral signatures of a water surface*

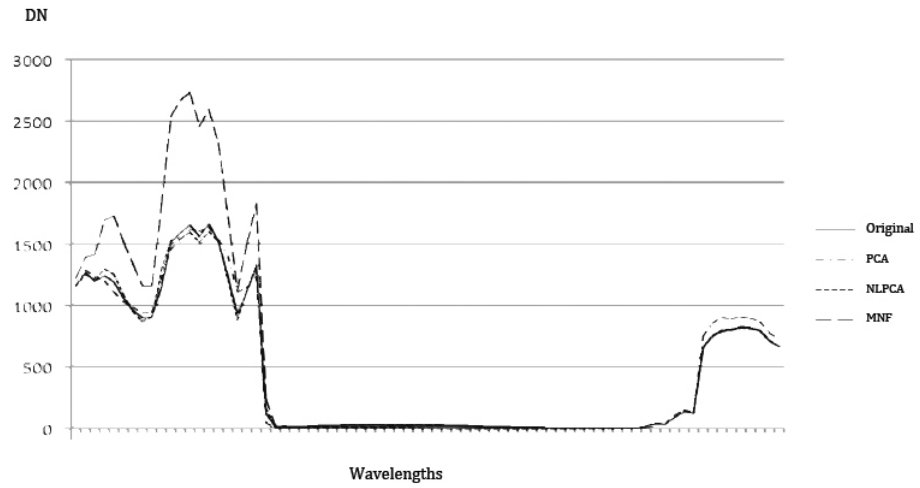


Figure 4.8: *Original and reconstructed spectral signatures of a deciduous trees forest*

the training phase which consisted of less than 150 epochs for the NLPCA and more than 1000 epochs for both PCA and MNF. It has to be noted that, once the training phase of all involved networks is completed, the overall "end-to-end" scheme providing the processing chain from the hyper-spectral measurement to the classification response is rather compact, consisting of one single neural architecture where the first stage perform the feature extraction and the second one the classification (fig.4.9). In particular, for our case, the whole architecture consists of the following layers: one input layer of 75 units, 4 hidden layer of 25, 5, 25, 25 neurons, respectively, and an output vector of 10 components. The NN-MLP classification has been applied, using the same training dataset, using the reduced components given by each features extraction technique. In fig 4.13, we report the confusion matrix computed for NLPCA on a set of about 50.000 pixels (not used in the training phase). The overall classification accuracy is of about 97% with a value for k-coefficient of 0.96. The only confusion element regards the class "Non deciduous trees". In fact, in some of the area assigned to this class the trees are rather sparse and mixed with wheat fields, causing some adjacency noise in the signal. In fig. 4.10 and 4.11 the original image (bands 11, 9 and 8) and the classification result based on the NLPCA are reported, respectively. The classification performances obtained with the other two methods are significantly lower, even if the data considered for the training and the test phase are the same. In both cases some classes are not recognized at all and the computed final accuracies values are 43,75% and 80,68% for PCA and MNF, respectively, as shown in the table reported in fig. 4.13 and 4.14. This latter result confirms a better

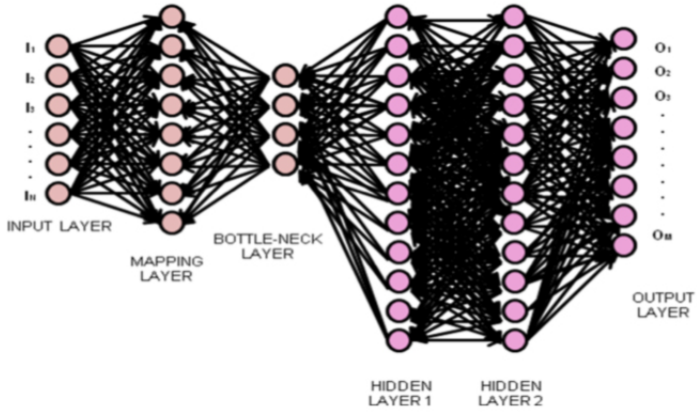


Figure 4.9: *"end-to-end" scheme providing the processing chain from the hyper-spectral dimensionality reduction to the classification phase*

behavior of MNF technique, with respect to PCA, when the high-dimension data are used for pixel-based classification.

4.4 Spectral unmixing

In the previous paragraph we assumed that each pixel vector measures the spectral response of a single material. In an hyper-spectral image the wide existence of mixed pixels is a nearly unavoidable problem. Within the reflective regime, the remotely sensed spectral signal of a mixed pixel is the combination of the spectral signatures of the constituent materials, usually known as endmembers, present in the pixel. To increase the accuracy of characterizing land surface, a measured mixed spectrum must be decomposed into a set of endmembers and their corresponding fractional abundances within the pixel. Theoretically, to achieve a good unmixing,

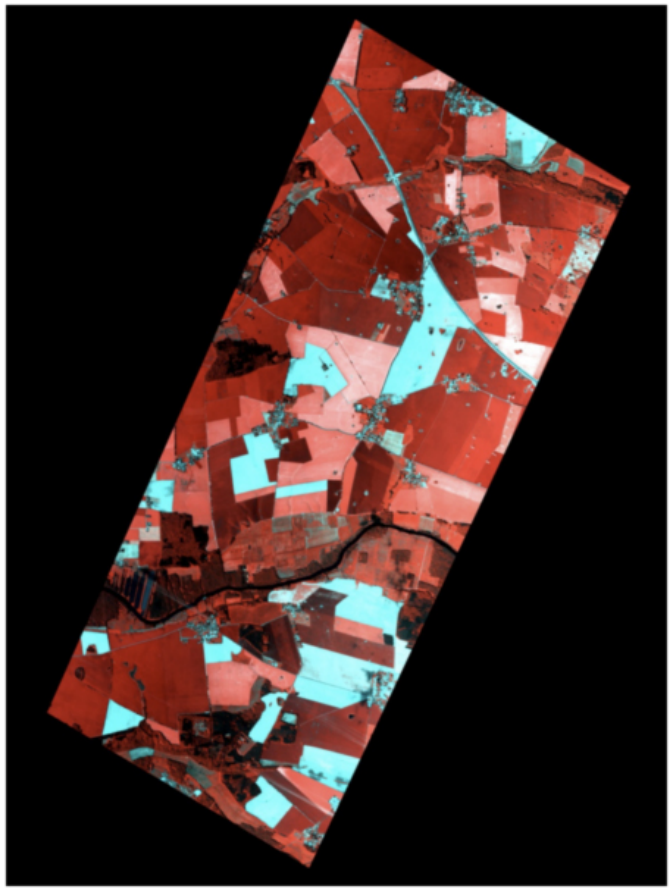


Figure 4.10: *AHS false color composition (bands 11, 9 and 7)*

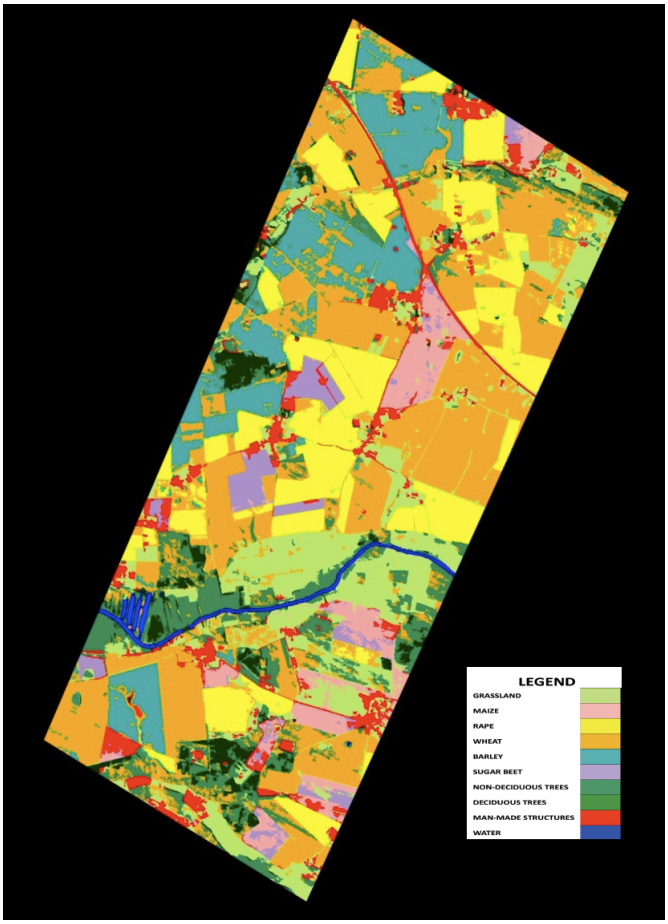


Figure 4.11: *Classification map derived from NLPCA method*

Classification Pixels \ Ground Truth Pixels	GRASSLAND	MAIZE	RAPE	WHEAT	BARLEY	SUGAR BEET	NON-DECIDU.	DECIDUOUS	MAN-MADE	WATER	TOTAL
GRASSLAND	**98,47%**	0,00%	**0,15%**	0,00%	0,00%	0,00%	**0,01%**	0,00%	**2,91%**	0,00%	4,27%
MAIZE	0,00%	**95,25%**	0,00%	0,00%	0,00%	**4,28%**	0,00%	0,00%	0,00%	0,00%	5,37%
RAPE	0,00%	0,00%	**99,78%**	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	33,28%
WHEAT	**1,53%**	0,00%	0,07%	**99,36%**	**2,77%**	0,00%	**35,09%**	0,00%	**0,10%**	0,00%	33,15%
BARLEY	0,00%	0,00%	0,00%	**0,57%**	**97,19%**	0,00%	0,00%	0,00%	0,00%	0,00%	10,00%
SUGAR BEET	0,00%	**4,75%**	0,00%	0,00%	0,00%	**95,72%**	0,00%	0,00%	0,00%	0,00%	3,37%
NON-DECIDUOUS TREES	0,00%	0,00%	0,00%	0,07%	**0,04%**	**0,00%**	**64,80%**	**1,59%**	**0,06%**	0,00%	4,04%
DECIDUOUS TREES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,10%**	**98,41%**	0,00%	**0,29%**	2,37%
MAN-MADE STRUCTURES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**96,93%**	0,00%	1,14%
WATER	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**99,71%**	3,00%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Figure 4.12: *Confusion matrix for the NLPCA method with overall Accuracy: 97.74% and K coefficient: 0.9574*

Classification Pixels \ Ground Truth Pixels	GRASSLAND	MAIZE	RAPE	WHEAT	BARLEY	SUGAR BEET	NON-DECIDU.	DECIDUOUS	MAN-MADE	WATER	TOTAL
GRASSLAND	**0,00%**	0,00%	**0,00%**	0,00%	0,00%	0,00%	**0,00%**	0,00%	**0,00%**	0,00%	0,00%
MAIZE	0,59%	**0,02%**	0,00%	0,00%	0,00%	**0,00%**	0,00%	29,26%	5,53%	10,32%	1,10%
RAPE	37,36%	0,14%	**29,03%**	0,00%	0,00%	0,00%	0,70%	12,44%	3,54%	0,00%	11,66%
WHEAT	**60,84%**	0,00%	70,97%	**100,00%**	**100,00%**	0,00%	**99,30%**	2,43%	**0,00%**	0,00%	73,34%
BARLEY	0,00%	0,00%	0,00%	**0,00%**	**0,00%**	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
SUGAR BEET	1,22%	**99,84%**	0,00%	0,00%	0,00%	**100,00%**	0,00%	55,87%	90,93%	89,68%	13,90%
NON-DECIDUOUS TREES	0,00%	0,00%	0,00%	0,00%	**0,00%**	**0,00%**	**0,00%**	**0,00%**	**0,00%**	0,00%	0,00%
DECIDUOUS TREES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,00%**	**0,00%**	0,00%	**0,00%**	0,00%
MAN-MADE STRUCTURES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,00%**	0,00%	0,00%
WATER	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,00%**	0,00%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Figure 4.13: *Confusion matrix for the PCA method with overall Accuracy: 42.75% and K coefficient: 0.4121*

Classification Pixels \ Ground Truth Pixels	GRASSLAND	MAIZE	RAPE	WHEAT	BARLEY	SUGAR BEET	NON-DECIDL	DECIDUOUS	MAN-MADE	WATER	TOTAL
GRASSLAND	**0,00%**	0,00%	**0,00%**	0,00%	0,00%	0,00%	**0,00%**	0,00%	**0,00%**	0,00%	0,00%
MAIZE	0,00%	**100,00%**	0,00%	0,00%	0,00%	**0,00%**	0,00%	0,00%	0,00%	0,00%	7,97%
RAPE	32,25%	0,00%	**99,98%**	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	35,10%
WHEAT	**6,13%**	0,00%	0,01%	**94,78%**	**0,53%**	0,00%	**0,00%**	0,00%	**0,00%**	0,00%	31,91%
BARLEY	1,30%	0,00%	0,00%	**0,61%**	**99,47%**	0,00%	0,00%	0,00%	0,00%	0,00%	10,39%
SUGAR BEET	8,22%	**0,00%**	0,01%	0,00%	0,00%	**0,00%**	0,00%	25,60%	0,00%	0,00%	1,77%
NON-DECIDUOUS TREES	0,00%	0,00%	0,00%	0,00%	**0,00%**	**0,00%**	**0,00%**	**0,00%**	**0,00%**	0,00%	0,00%
DECIDUOUS TREES	52,09%	0,00%	0,00%	4,61%	0,01%	0,00%	**0,00%**	**74,40%**	0,00%	**0,00%**	12,86%
MAN-MADE STRUCTURES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,00%**	0,00%	0,00%
WATER	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	**0,00%**	0,00%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	0,00%	0,00%	100,00%	0,00%	0,00%	100,00%

Figure 4.14: *Confusion matrix for the MNF method with overall Accuracy: 80.68% and K coefficient: 0.5002*

the endmembers should be uncorrelated each other as much as possible. Usually the endmembers are extracted from the image and correspond to macroscopic object such as water, bare soil or vegetation. In our case the endmembers correspond to the classes previously used in the classification.

In this work, to explore on the potential of neural networks in managing the unmixing problem, the abundances estimation was carried out through the analysis of the output of the classification provided by the neural network algorithm. In other words, the value of each output has been used as an estimator of the fractional abundances of each endmember.

More analytically the abundance a_i corresponding to i-esm class is given by the following expression:

$$a_i = \frac{o_i}{\sum_{k=1}^{M} o_k} \tag{4.2}$$

where o_k indicates the neural network output associated to the k-esm endmember and M is the total number of endmembers, in this case $M = 10$.

	LSU	NN
MA	0,00%	0,00%
WW	3,87%	0,57%
WB	12,03%	99,08%
RA	7,58%	0,35%
SB	8,34%	0,00%
PS	15,35%	0,00%
WA	16,07%	0,00%
BU	16,25%	0,00%
CT	12,34%	0,00%
DT	8,17%	0,00%

Figure 4.15: *High resolution image and abundances matrix of a pixel covering entirely a winter barley area (MA: Maize; WW: winter wheath; WB: winter barley; RA: rape; SB: sugar beet; PS: pasture; WA: water; BU: built-up area; CT: conuferous trees; DT: deciduous trees)*

To test the feasibility of such an approach, a comparison exercise with the most affordable method known as Linear Spectral Unmixing (LSU) was made. More in particular, some pixels from the original image were selected to evaluate the efficiency of the two methodologies. The comparison of the results shows that the LSU came to a wrong result, providing abundances values always above 0, also in those pixels where some elements are supposed not to be present. The proposed method, on the other hand, does not show this kind of problems, providing a good accuracy as shown in fig.4.15-4.20.

	LSU	NN
MA	0,00%	0,00%
WW	9,55%	69,88%
WB	9,25%	1,02%
RA	7,34%	0,00%
SB	14,34%	0,00%
PS	14,54%	0,00%
WA	14,24%	0,00%
BU	16,03%	0,00%
CT	9,16%	0,00%
DT	5,54%	29,10%

Figure 4.16: *High resolution image and abundances matrix of a pixel covering a winter wheath area mixed with a tree (MA: Maize; WW: winter wheath; WB: winter barley; RA: rape; SB: sugar beet; PS: pasture; WA: water; BU: built-up area; CT: conuferous trees; DT: deciduous trees)*

4.5 Conclusions

In this work a novel approach based on AANN for the extraction of nonlinear principal components from AHS hyper-spectral data has been developed. Such an approach should be more suitable to eliminate nonlinear correlations in the data hence to optimize the design of successive inversion schemes. A NN algorithm with MLP topology has been also exploited to handle a successive classification task leading to a final single architecture performing the two processing stages: the feature extraction and the classification. The results show that the feature extraction based on AANN outperforms that obtained considering more traditional approaches such as PCA and MNF, suggesting the potential of the technique to reduce dimensionality of hyper-spectral data even only for the storage or trans-

	LSU	NN
MA	1,22%	0,00%
WW	6,81%	0,02%
WB	11,19%	0,02%
RA	3,33%	0,00%
SB	12,23%	0,01%
PS	18,71%	0,00%
WA	17,62%	0,02%
BU	20,42%	0,00%
CT	0,00%	37,15%
DT	8,45%	62,78%

	LSU	NN
MA	23,79%	0,00%
WW	5,12%	0,00%
WB	14,50%	0,00%
RA	7,51%	0,00%
SB	0,00%	0,00%
PS	9,94%	0,00%
WA	9,66%	0,16%
BU	8,03%	45,17%
CT	1,26%	54,64%
DT	20,18%	0,02%

	LSU	NN
MA	26,51%	0,05%
WW	6,48%	0,39%
WB	14,85%	0,00%
RA	5,66%	0,00%
SB	0,00%	0,00%
PS	7,97%	80,83%
WA	7,71%	0,00%
BU	4,97%	3,61%
CT	4,49%	0,00%
DT	21,36%	15,11%

	LSU	NN
MA	21,79%	0,02%
WW	3,79%	4,57%
WB	16,61%	0,00%
RA	0,09%	0,00%
SB	0,00%	0,00%
PS	12,16%	53,47%
WA	11,70%	0,00%
BU	9,28%	0,03%
CT	3,17%	0,00%
DT	21,42%	41,90%

Figure 4.17: *High resolution image and abundances matrices of pixels covering a mixed urban area (MA: Maize; WW: winter wheath; WB: winter barley; RA: rape; SB: sugar beet; PS: pasture; WA: water; BU: built-up area; CT: conuferous trees; DT: deciduous trees)*

	LSU	NN
MA	25,31%	0,03%
WW	5,71%	0,64%
WB	15,12%	0,00%
RA	4,88%	0,00%
SB	0,00%	0,00%
PS	9,31%	66,79%
WA	8,93%	0,00%
BU	6,59%	1,88%
CT	2,87%	0,00%
DT	21,27%	30,67%

Figure 4.18: *High resolution image and abundances matrix of a pixel relative to a garden (MA: Maize; WW: winter wheath; WB: winter barley; RA: rape; SB: sugar beet; PS: pasture; WA: water; BU: built-up area; CT: conuferous trees; DT: deciduous trees)*

	LSU	NN
MA	11,47%	0,00%
WW	0,00%	96,49%
WB	16,24%	1,80%
RA	2,91%	0,00%
SB	0,52%	0,00%
PS	16,43%	0,00%
WA	16,60%	0,00%
BU	15,63%	0,00%
CT	3,03%	0,00%
DT	17,19%	1,70%

Figure 4.19: *High resolution image and abundances matrix of a pixel covering entirely a winter wheath area (MA: Maize; WW: winter wheath; WB: winter barley; RA: rape; SB: sugar beet; PS: pasture; WA: water; BU: built-up area; CT: conuferous trees; DT: deciduous trees)*

mission purposes. Moreover, the reduced vector, allows to yields land cover maps with rather satisfactory accuracy. In fact, from the comparison with the ground truth, an overall successful classification rate of about 97% is observed. This means that the NLPCA are able to retain most of the information content of the raw data while the other two techniques seem less effective under this point of view. Finally NLPCA, compared with MNF and PCA, allowed performing the training of the neural algorithm for the classification map in a limited number of epochs, which, besides involving a faster training time, may be important to avoiding overfitting.

Another important point is that NLPCA, by comparison with other feature reduction techniques applied on hyper-spectral data, seems to be more robust to the smile effect. Hence specific pre-processing routine may

	LSU	NN
MA	0,00%	0,00%
WW	6,83%	0,00%
WB	8,27%	0,11%
RA	16,10%	0,00%
SB	11,39%	0,00%
PS	14,88%	0,00%
WA	15,04%	17,38%
BU	18,05%	0,00%
CT	4,58%	82,50%
DT	4,86%	0,00%

	LSU	NN
MA	20,22%	0,11%
WW	8,31%	0,00%
WB	6,90%	0,00%
RA	34,12%	0,14%
SB	1,58%	0,00%
PS	0,00%	0,00%
WA	2,20%	99,75%
BU	0,90%	0,00%
CT	15,19%	0,00%
DT	10,58%	0,00%

	LSU	NN
MA	10,69%	0,00%
WW	4,46%	0,00%
WB	8,57%	0,00%
RA	25,48%	0,00%
SB	4,00%	0,00%
PS	11,19%	0,00%
WA	11,87%	78,67%
BU	14,26%	0,00%
CT	0,00%	21,33%
DT	9,48%	0,00%

Figure 4.20: *High resolution image and abundances matrices of pixels related to water surfaces mixed with wetlands (MA: Maize; WW: winter wheath; WB: winter barley; RA: rape; SB: sugar beet; PS: pasture; WA: water; BU: built-up area; CT: conuferous trees; DT: deciduous trees)*

be avoided using NLPCA.

Finally, we analysed the potential of a new neural network technique for the spectral unmixing of hyper-spectral imagery. In fact, this approach provided the mixed distribution of the considered endmembers. The results stemming from a comparison exercise with the well-known LSU method was made show that the proposed approach can be effective.

Chapter 5

Production of land cover maps from multi-temporal and multi-angular hyper-spectral data

5.1 Introduction

In the previous chapter, Auto Associative Neural Network feature reduction technique, applied on an airborne hyper-spectral dataset (AHS), has been investigated. In this chapter we applied the NLPCA methodology to CHRIS-PROBA datasets, to evaluate the ability of the AANN to detect correlation among hyper-spectral data when combined with multi-temporal and multi-angular information. The chosen test site is the area surrounding Tor Vergata University and Frascati. This is a mainly flat area located in the southeast of Rome (41°51'26"N, 12°40'26"E), which represent an interesting heterogeneous landscape. Permanent crops, such as vineyards

and olive trees and other fruit trees are mixed with agricultural areas characterized by a growth cycle, mainly corn fields, and uncultivated areas or pasture. Artificial land cover consists of residential urban areas, industrial and commercial units, and different kinds of road networks. The NLPCA methodology for dimensionality reduction was applied on two different datasets:

- A multi-temporal dataset composed by three CHRIS-PROBA mode-3 acquisitions, each consisting of 18 measurements acquired on different dates, for a total number of 54 inputs.
- A combination of hyper-spectral multi-angular and multi-temporal dataset, consisting of 72 measurements obtained adding a 36° acquisition to the previous dataset.

Also in this study a reduced vector has been given as input to a MLP to produce a land cover classification map. For each dataset, a confusion matrix was produced and evaluated.

5.2 Multi-temporal dataset

The multi-temporal dataset was composed by three acquisitions, taken on February 28, 2006, August 19, 2006 and October 9, 2006. Such dates are, in principle, particularly suitable to sample the crops' growth cycle, hence to catch the differences among the multi temporal signatures associated to each land cover type. In fact, the acquisition at the end of winter is particularly useful to observe the plough phase of some agricultural fields

such as corn fields. During summer most of the crops present in this area are in the maximum of the photosynthetic phase, while uncultivated areas, covered by dry grass or sparse vegetation, are characterized by a lower reflectance in the infrared range. Finally the early-autumn acquisition permits to monitor the harvesting of cereals, crops and vineyards, as well as the plough of agricultural fields that in this period have to be ready for the winter sowing. Fig.5.1 shows the multitemporal iamges in fasle colors. It should be added that the images underwent atmospheric calibration and other pre-processing stages such as destriping according to the procedures indicated in [85].

5.2.1 Feature extraction

To reduce the number of input measurements avoiding loss of information, a NLPCA, obtained through an AANN, was performed. Also in this case the number of nodes in the bottleneck was chosen through a comparison with the PCA. The selected topology consists of 54-25-4-25-54 nodes, and the 4 nonlinear components have been computed feeding the AANN with all pixels in the image. In fact, it resulted that the first 4 PCA components contained almost the 99% of the whole statistical information.

The dimension of the outer hidden layers of the AANN was again established after a numerical analysis aiming at minimizing the mean-square-error MSE. As in the previous chapter, the choice of 25 units for the two intermediate layers resulted from a comparison of various network topologies.

The NPCA components are illustrated in Fig.5.2. It can be noted that

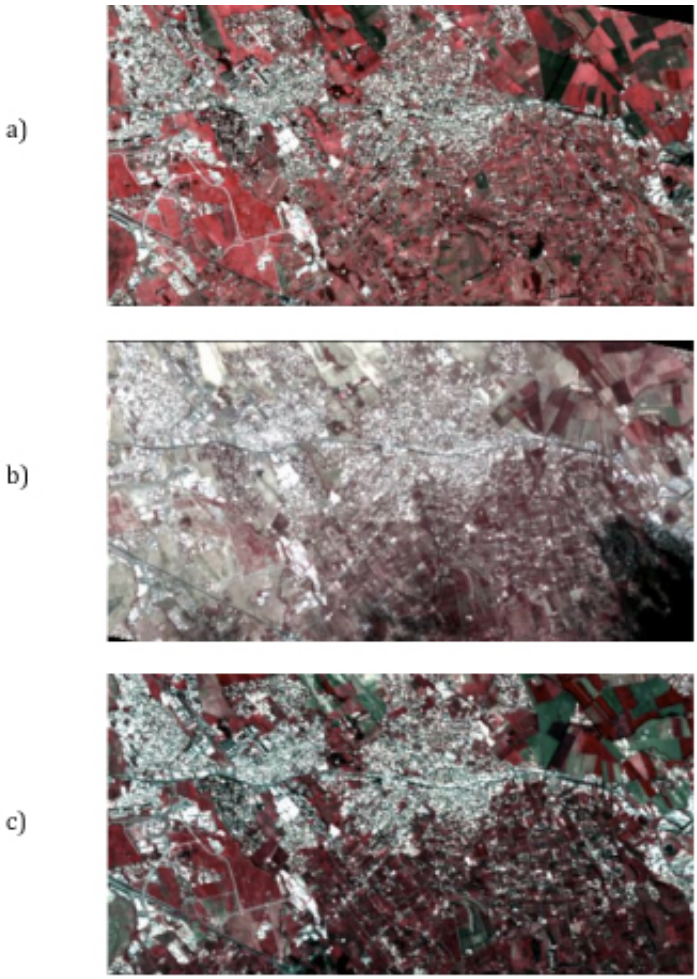


Figure 5.1: *False color CHRIS-Proba images (bands 11, 9 and 7) of three different 2006 acquisitions. a) 28 February, b) 19 August, c) 09 October*

the NLPCA components are polarized on different types of information contained on the image. For example, component 4 is more focused on buildings while components 3 and 1 are more sensitive to natural areas; component 2 shows the presence of a cloud in one of the acquisitions while this is not the case for the other components. Moreover, analyzing the composite image, obtained from components 4, 3 and 2, as showed in fig.5.3 and fig.5.4, it can be noted that the NLPCA tend to emphasize the small watercourses present in the area, that are not so evident in the original images.

5.2.2 Classification

In a successive step the 4 nonlinear components have been used to produce the land cover map of the test area. A neural network algorithm has been considered for the classification. This time the MLP topology is 4-16-16-11. The neural algorithm has been trained using a training set and a test set of 3300 and 1975 patterns, respectively. A third set of 2766 ground truth pixels has been considered for the computation of the classification accuracy. The number of training epochs necessary to get the network trained is about 130, which is significantly lower respect to the case where the 54 measurements are given straightforward to the net. Also in this case fig.4.9 depicts the entire processing scheme where the specific topology of 54-25-4-16-16-11is considered. The final land cover map is shown in Fig.5.5, and the corresponding confusion matrix in Fig.5.6. The overall accuracy is 92.32%. For the sake of comparison it has to be noted that a direct, that is without dimensionality reduction, neural network classification of the same

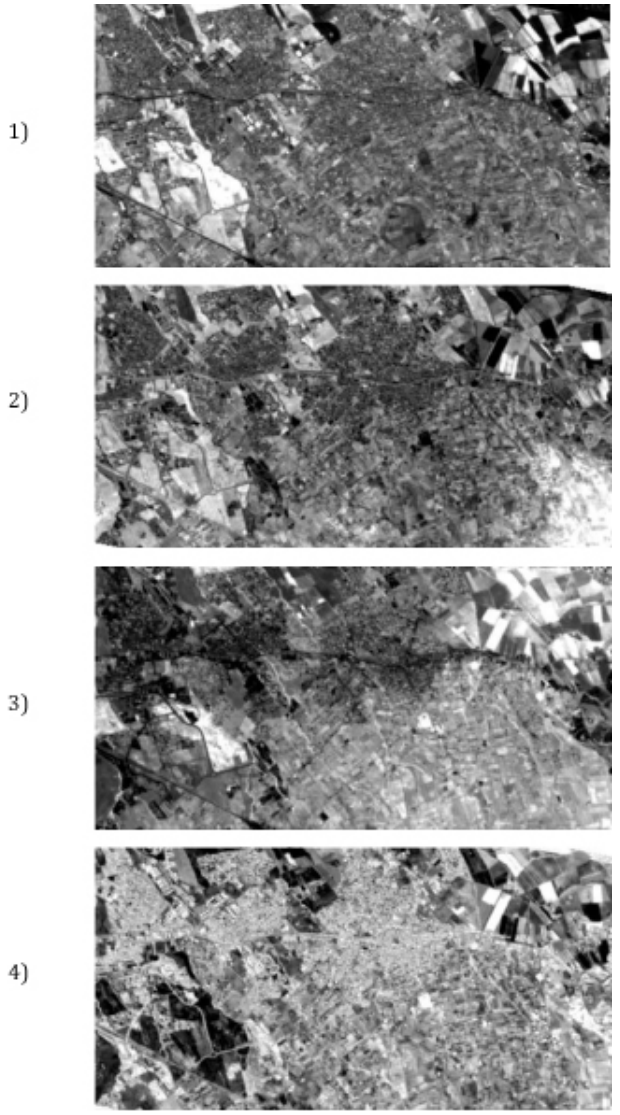


Figure 5.2: *Four nonlinear principal components obtained from a multi-temporal CHRIS-Proba dataset*

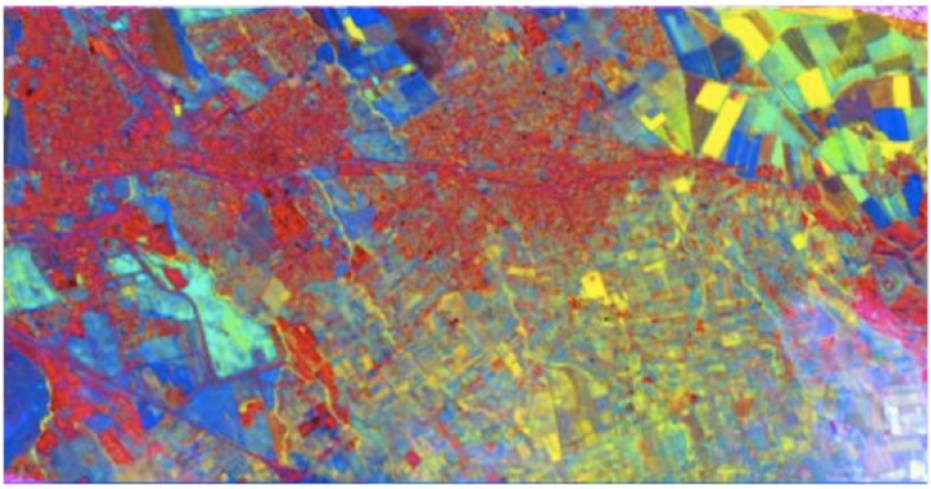


Figure 5.3: *Composition of three nonlinear principal components (4, 3, 2) obtained from the CHRIS-Proba dataset*

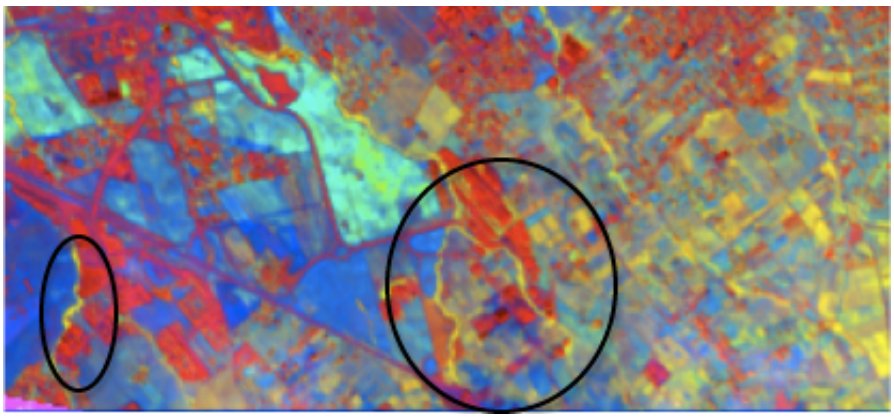


Figure 5.4: *Watercourses highlighted in the composition of three nonlinear principal components (4, 3, 2)*

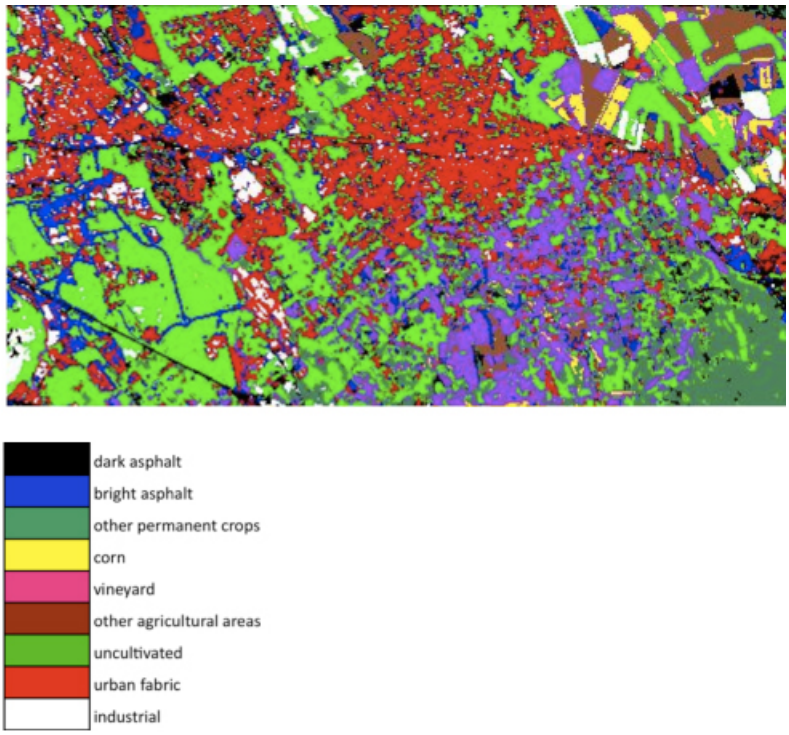


Figure 5.5: *Classification map of the multi-temporal dataset*

data set has been performed by [85]. As in this latter case the final overall accuracy value was of 85,7%, we may conclude that about 7% in accuracy is gained by using the NLPCA.

To assess the accuracy of the classification, the obtained result has been evaluated through a comparison with that obtained by

	vineyard	uncultivated	other permanent crops	industrial	dark asphalt	corn	urban fabric	bright asphalt	other agricultural areas	Total
vineyard	**92,09%**	0,00%	0,00%	0,00%	0,00%	4,27%	0,00%	0,00%	0,00%	9,30%
uncultivated	0,96%	**99,35%**	9,34%	0,26%	2,67%	2,40%	0,00%	0,00%	2,74%	19,54%
other permanent crops	0,00%	0,00%	**79,77%**	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	4,77%
industrial	0,00%	0,00%	0,00%	**84,69%**	0,00%	0,00%	11,53%	6,06%	0,00%	9,33%
dark asphalt	0,00%	0,00%	0,00%	0,00%	**94,00%**	0,00%	0,38%	0,00%	0,00%	3,33%
corn	5,28%	0,00%	5,84%	0,00%	0,00%	**93,33%**	0,00%	0,00%	0,16%	9,05%
urban fabric	0,00%	0,00%	0,00%	13,27%	0,67%	0,00%	**85,82%**	12,12%	0,70%	12,37%
bright asphalt	0,00%	0,65%	0,00%	1,79%	2,67%	0,00%	1,32%	**81,82%**	0,08%	3,07%
other agricultural areas	1,68%	0,00%	5,06%	0,00%	0,00%	0,00%	0,95%	0,00%	**96,33%**	29,24%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Figure 5.6: *Confusion matrix of the multi-temporal dataset, with an overall accuracy of 92.32%*

5.3 Multi-angular/temporal dataset

The PROBA mission acquisition plan allows taking images over the same test area in different dates and with different angles. This characteristic allows the concurrently exploitation of hyper-spectral, multi-angular and multi-temporal information. In this experiment the classification accuracy of a hyper-spectral dataset composed by measurements taken in different dates and angles was evaluated. The new dataset was derived from the multi-temporal one with the addition of one new acquisition at FZA 36° of the August date. The choice of this date stems from the fact that the multi-angular information should carry the best contribution when the agricultural fields are at the stage of full development. A higher number of angular acquisitions were not taken into account, due to different concurrent

reasons such as the sensible shifts affecting the images acquired at different angles, sensitivity to clouds and the co-registration errors.

5.3.1 Feature extraction

As for the multi-temporal experiment, an AANN algorithm to perform NLPCA was applied to perform the dimensionality reduction of the data. In this case, the chosen topology consists of 72-25-5-25-72 nodes, and the 5 nonlinear components have been computed feeding the AANN with all pixels in the image. In the same way as the previous experiments, the number of the bottleneck nodes was set by a comparison with the PCA components, from which it resulted that the first 5 components contained almost the 99% of the whole statistical information. The dimension of the outer hidden layers of the AANN was then established after a numerical analysis aiming at minimizing the MSE error. Also in this experiment it can be noted a polarization of the NLPCA components on different types of information contained on the image. More in particular, in fig.5.7 it can be noted that component 2 is more focused on buildings while component 3 is more sensitive to natural areas; component 5, besides retrieving the information on the presence of a cloud in one of the acquisitions, also highlights the road network.

5.3.2 Classification

The successive step was to produce a land cover classification map of the test area using the 5 nonlinear components obtained by the NLPCA as input to a new neural network algorithm. This time the topology is that

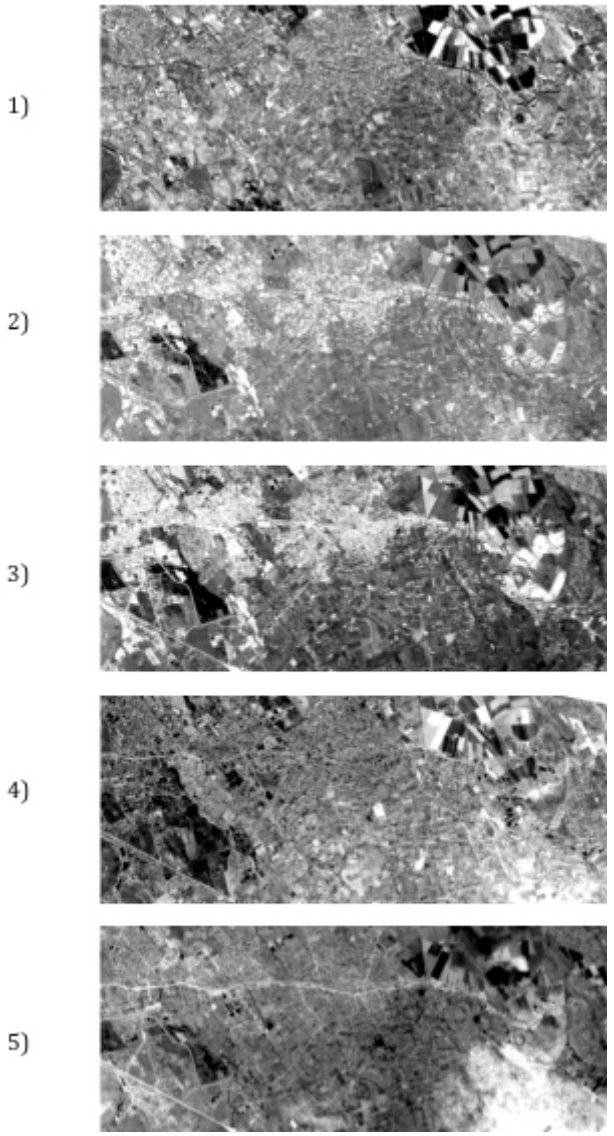


Figure 5.7: *Five nonlinear principal components obtained from a fusion of multi-temporal and multi-angular CHRIS-Proba dataset*

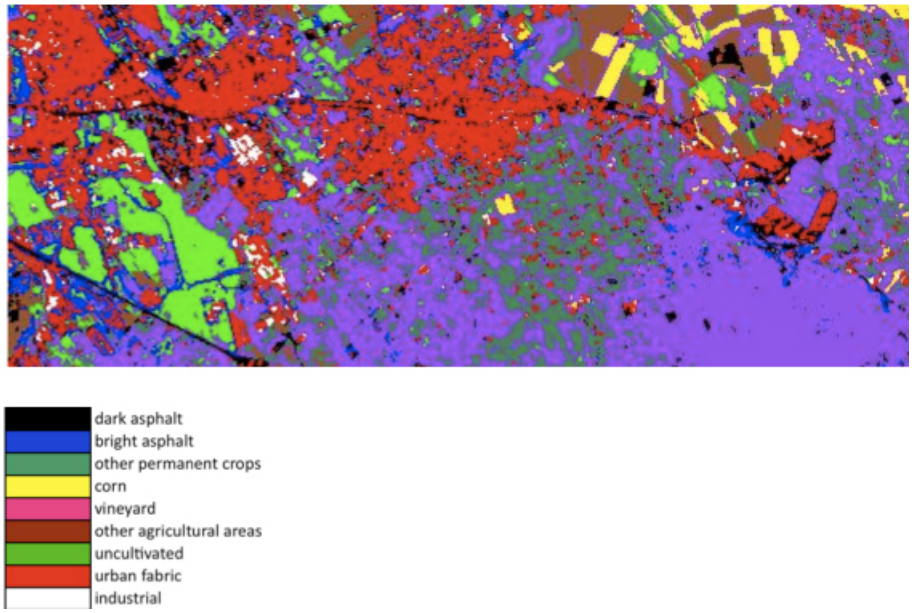


Figure 5.8: *Classification map of the multi-temporal and multi-temporal dataset*

of a standard MLP with a 5-36-36-11 topology. The neural algorithm has been trained using a training set and a test set of 4404 and 1216 patterns, respectively. A third set of 2776 ground truth pixels has been considered for the computation of the classification accuracy. Also in this case the number of training epochs necessary to get trained the network was very low (150). The final land cover map is shown in Fig.5.8, and the corresponding confusion matrix in Fig.5.9. The final overall accuracy obtained a value over 94%, improving the accuracy obtained only using the multi-temporal dataset.

	vineyard	uncultivated	other permanent crops	industrial	dark asphalt	corn	urban fabric	bright asphalt	other agricultural areas	Total
vineyard	**99,56%**	0,00%	9,16%	0,00%	0,00%	0,00%	0,08%	0,00%	0,00%	16,65%
uncultivated	0,00%	**100,00%**	0,00%	0,00%	0,00%	0,00%	0,33%	3,57%	0,00%	6,41%
other permanent crops	0,44%	0,00%	**89,69%**	0,00%	0,00%	0,00%	0,49%	0,00%	0,00%	8,55%
industrial	0,00%	0,00%	0,00%	**86,51%**	0,00%	0,00%	0,57%	0,00%	0,00%	9,05%
dark asphalt	0,00%	0,00%	0,76%	0,00%	**93,33%**	0,00%	0,33%	0,00%	0,00%	1,20%
corn	0,00%	0,00%	0,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%	0,00%	3,45%
urban fabric	0,00%	0,00%	0,00%	12,80%	0,00%	0,00%	**96,00%**	10,71%	0,00%	42,84%
bright asphalt	0,00%	0,00%	0,38%	0,69%	3,33%	0,00%	2,20%	**85,71%**	0,00%	1,94%
other agricultural areas	0,00%	0,00%	0,00%	0,00%	3,33%	0,00%	0,00%	0,00%	**100,00%**	9,93%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Figure 5.9: *Confusion matrix of the multi-temporal and multi-angular dataset, with an overall accuracy of 95.67%*

5.4 Spectral unmixing

From an accurate analysis of the confusion matrices, it became evident the not all the classes obtained a good classification accuracy. More in particular, the permanent crop class derived from areas covered by fruit and olive trees, and the class related to asphalts, were subject to misclassification. As far as the permanent crops are concerned, these kinds of cultivations are characterized by a peculiar pattern. In fact, as shown in fig.5.10, the trees are spatially distributed and do not completely cover the cultivation surface. Depending on the distance between each tree, the prevailing spectral signature in a pixel could not even be that related to the permanent crop species. Other signatures present in the area covered by the pixel, such as pasture or bare soil, may prevail. In some cases, mixtures of all the considered signatures, lead the algorithm to a wrong classification.

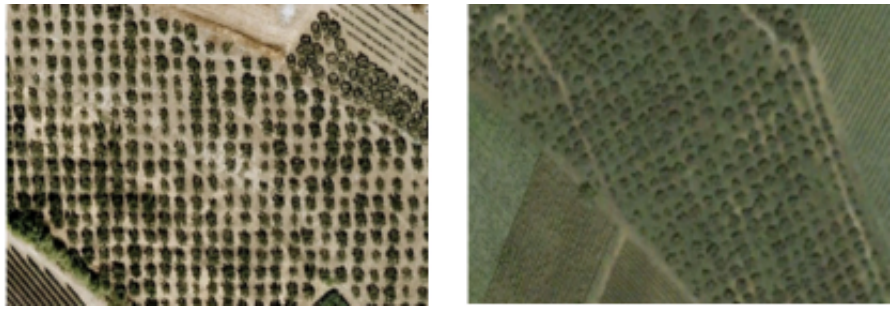


Figure 5.10: *Different olive trees cultivations*

Figure 5.11: *Different road types in the same area*

	LSU	NN	GT
VY	14,23%	0,11%	0,00%
PS	16,54%	0,03%	2,00%
PC	13,83%	0,14%	0,00%
IN	14,91%	0,75%	0,00%
DA	7,74%	0,00%	0,00%
MA	2,14%	0,46%	0,00%
BU	1,59%	11,22%	18,00%
BA	10,01%	86,98%	80,00%
AA	19,01%	0,31%	0,00%

Figure 5.12: *The abundances obtained by neural network unmixing technique and LSU, on a bright asphalt pixel. In this picture, it can be noted that the pixel surface contain not only a road but also a part of a building and some vegetation (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	14,88%	4,85%	5,00%
PS	21,61%	0,09%	0,00%
PC	11,04%	83,27%	80,00%
IN	12,73%	0,00%	0,00%
DA	0,57%	0,29%	0,00%
MA	0,00%	0,00%	0,00%
BU	0,18%	0,00%	0,00%
BA	19,31%	0,00%	0,00%
AA	19,68%	11,50%	15,00%

Figure 5.13: *A permanent crop pixel. The total pixel signature is a mixture of olive trees, vineyards, grass and a very small road (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	8,10%	49,29%	50,00%
PS	10,82%	0,00%	0,00%
PC	0,00%	24,17%	20,00%
IN	15,38%	0,00%	0,00%
DA	11,17%	23,41%	0,00%
MA	11,50%	0,27%	0,00%
BU	7,93%	1,58%	25,00%
BA	3,60%	0,35%	0,00%
AA	31,51%	0,94%	5,00%

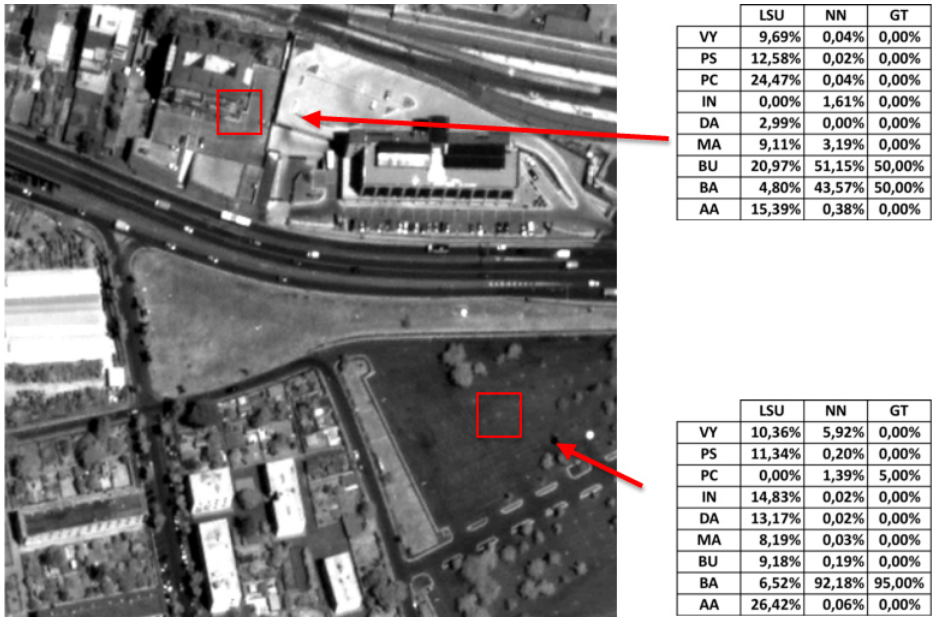
Figure 5.14: *In this pixel the total signature is a mixture of trees, vineyards, roads and buildings (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	40,39%	40,87%	40,00%
PS	0,55%	2,08%	10,00%
PC	0,00%	40,00%	40,00%
IN	3,35%	0,02%	0,00%
DA	6,52%	0,00%	0,00%
MA	11,96%	0,65%	0,00%
BU	2,30%	2,50%	0,00%
BA	20,28%	13,15%	10,00%
AA	14,65%	0,72%	0,00%

Figure 5.15: *A pixel covering an area equally distributed among trees and vineyards (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	1,04%	0,10%	0,00%
PS	9,83%	42,38%	40,00%
PC	21,05%	0,09%	0,00%
IN	0,00%	5,88%	0,00%
DA	7,59%	0,00%	0,00%
MA	2,02%	0,00%	0,00%
BU	18,79%	0,01%	0,00%
BA	12,62%	51,38%	60,00%
AA	27,06%	0,16%	0,00%

Figure 5.16: *In this image the pixel cover an uncoltivated area crossed by a road (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*



	LSU	NN	GT
VY	9,69%	0,04%	0,00%
PS	12,58%	0,02%	0,00%
PC	24,47%	0,04%	0,00%
IN	0,00%	1,61%	0,00%
DA	2,99%	0,00%	0,00%
MA	9,11%	3,19%	0,00%
BU	20,97%	51,15%	50,00%
BA	4,80%	43,57%	50,00%
AA	15,39%	0,38%	0,00%

	LSU	NN	GT
VY	10,36%	5,92%	0,00%
PS	11,34%	0,20%	0,00%
PC	0,00%	1,39%	5,00%
IN	14,83%	0,02%	0,00%
DA	13,17%	0,02%	0,00%
MA	8,19%	0,03%	0,00%
BU	9,18%	0,19%	0,00%
BA	6,52%	92,18%	95,00%
AA	26,42%	0,06%	0,00%

Figure 5.17: *The top the pixel cover a building with an asphalted courtyard, the bottom the pixel relies entirely on a parking lot (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	13,50%	0,54%	0,00%
PS	10,59%	0,02%	0,00%
PC	0,00%	56,57%	60,00%
IN	12,24%	0,08%	0,00%
DA	8,42%	0,10%	0,00%
MA	8,38%	0,08%	0,00%
BU	10,91%	1,55%	0,00%
BA	9,30%	0,05%	0,00%
AA	26,66%	41,01%	40,00%

Figure 5.18: *In this image the pixel relies on a olive trees cultivation (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	18,06%	50,94%	50,00%
PS	8,74%	0,07%	0,00%
PC	0,00%	48,67%	50,00%
IN	3,96%	0,00%	0,00%
DA	3,53%	0,16%	0,00%
MA	4,84%	0,00%	0,00%
BU	23,44%	0,00%	0,00%
BA	17,73%	0,04%	0,00%
AA	19,70%	0,11%	0,00%

Figure 5.19: *A pixel covering an area between a vineyard and some trees (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	NN
VY	22,07%	5,42%	0,00%
PS	5,65%	79,34%	90,00%
PC	0,00%	0,72%	0,00%
IN	11,61%	0,27%	0,00%
DA	12,73%	0,00%	0,00%
MA	6,27%	0,00%	0,00%
BU	2,35%	0,00%	0,00%
BA	6,04%	14,19%	10,00%
AA	33,28%	0,07%	0,00%

Figure 5.20: *A pixel at the border of an uncoltivated area (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	LSU	NN	GT
VY	7,78%	0,22%	0,00%
PS	6,63%	99,45%	100,00%
PC	9,46%	0,14%	0,00%
IN	6,75%	0,16%	0,00%
DA	16,28%	0,00%	0,00%
MA	11,07%	0,00%	0,00%
BU	9,74%	0,00%	0,00%
BA	0,00%	0,00%	0,00%
AA	32,28%	0,03%	0,00%

Figure 5.21: *An uncoltivate area pixel (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

As the same way, in urban context, pixels related to the road network do not show pure asphalt signatures as can be noted in fig.5.11, but in many cases it can be possible to have a mixture with the buildings signatures. Also in rural context, where there are no buildings surrounding the road, it is possible to have a mixed pixel, especially in those pixels where the road width does not cover the entire pixel dimension.

To overcome these problems, it can be important to "unmix" the signatures, and then evaluate the abundance of each class present in a pixel. The abundances estimation was obtained through the analysis of the output of the classification neural network algorithm adopting the procedure considered for AHS imagery and explained in the previous chapter. The value of each output was used as an estimator of the fractional abundances of each class. Again the technique was compared with the Linear Spectral Unmixing (LSU) method [86] but in this case a more quantitative assessment was carried out. In particular a ground-truth in terms of percentages of abundances of the considered classes was determined by visual inspection on Google Earth for a certain number of pixels. As the main purpose was a preliminary quantitative comparison between the LSU and the NN technique, it has been assumed that a very detailed percentages measurement was not necessary. Table 5.1 reports the mean and the standard deviation values of each endmember considering the whole set of measured pixels.

In figs.5.12-5.21 we report some examples of results obtained on selected pixels extracted from the data set, the LSU was unable to correctly estimate the abundances of the elements, providing values always above 0 also in those pixels where the corresponding elements are not present. This be-

	MEAN	ST. DEV.
VY	0.17	0.26
PS	0.16	0.30
PC	0.17	0.26
IN	0.06	0.21
DA	0.10	0.26
MA	0.0	0.0
BU	0.12	0.24
BA	0.16	0.29
AA	0.07	0.15

Table 5.1: *Mean and standard deviation of each endmember (VY: vineyards; PS: pasture; PC: permanent crops; IN: industrial; DA: dark asphalt; MA: maize; BU: built-up area; BA: bright asphalt; AA: agricultural area)*

	RMSE
NN	0.0801
LSU	0.3492

Table 5.2: *RMSE computed over the entire dataset and considering all classes*

havior came from the assumption that standard unmixing methodologies such as LSU, to produce good result, require the endmembers to be the most uncorrelated as possible. Usually the endmembers extracted from the image, correspond to macroscopic objects in the scene such as water, soil or vegetation. In our case, the endmembers correspond to classes that may be very correlated each other, leading a standard unmixing technique to a wrong result. On the other hand, as the abundancies matrices of test pixels demonstrate, the unmixing technique using neural network provides a good accuracy even if the chosen endmembers are closely correlated. Finally, in table 5.2 we report the quantitative assessment in terms of RMSE computed over the entire data set considering all classes. From the reported values we see that NN seems to be definetely more effective than LSU.

5.5 Conclusions

In this work we applied the NLPCA technique to reduce the dimensionality of multi-temporal and multi-angular hyper-spectral datasets. A NN algorithm with MLP topology has been exploited to handle the successive classification of the reduced datasets. The results show a satisfactory overall accuracy, increasing that obtained not reducing the dataset. This means that NLPCA is able to retain most of the information content of the original data, reducing the worsening influence of the dimensionality curse. Some other issues emerged from the analysis of the results. Not good classification accuracies were obtained for permanent crops and bright asphalt classes. These problems are due to the not very high spatial resolution of the image, allowing some pixels to have a mixed spectral signature. To

solve those problems, a novel spectral unmixing technique based on neural network was proposed. The quantitative results in terms of comparison with LSU show a good ability in the estimation of the percentage of each component in a pixel.

Chapter 6

Urban area classification using high resolution hyperspectral data

6.1 Introduction

This chapter is dedicated to the outcome of the 2008 GRS-S Data Fusion Contest. The Data Fusion Contest has been organized by the Data Fusion Technical Committee (DFTC) of the IEEE Geoscience and Remote Sensing Society and has been annually proposed since 2006. It is a contest open not only to DFTC members but also to everyone. The aim of the Data Fusion Contest is to evaluate existing methodologies at the research or operational level to solve remote sensing problems using data from different sensors. The main aim of this contest is to provide a benchmark to the researchers interested in a class of data fusion problems, starting with a contest and then allowing the data and results to be used as reference for

the widest community, inside and outside the DFTC. In 2008, the contest was dedicated to the classification of very high-resolution hyper-spectral data. A hyper-spectral data set was distributed to every participant, and the task was to obtain a classified map as accurate as possible with respect to the ground truth data, depicting land-cover and land-use classes. The ground truth was kept secret, but training pixels could be selected by the participants by photo-interpretation in order to apply supervised methods. The data set consisted of airborne data from the Reflective Optics System Imaging Spectrometer (ROSIS-03) optical sensor. The flight over the city of Pavia, Italy, was operated by the Deutschen Zentrum fur Luft-und Raumfahrt (the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. The number of bands of the ROSIS-03 sensor is 115 with a spectral coverage ranging from 0.43 to 0.86 μm. Thirteen noisy bands have been removed. The dimension of the distributed data set is hence 102. The spatial resolution is 1.3 m per pixel. For the contest, five classes of interest were considered, namely, buildings, roads, shadows, vegetation, and water. Everyone could enter the contest and download the data set. After classification, the participant could upload the resulting map for an automatic evaluation of the classification performances (confusion matrix and average accuracy). The participating teams were allowed to upload as many different results as they wished.

The contest was open for three months. At the end of the contest, the participant teams had uploaded over 2100 classification maps. The five best individual classification maps have been fused together. The fi-

nal corresponding teams have been awarded with an IEEE Certificate of Recognition during the Chapters and Technical Committees' Dinner at the IEEE International Geoscience and Remote sensing Symposium in Boston in July 2008. The five best algorithms were:

1. An algorithm where different standard classifiers [three neural networks (NNs) and two maximum likelihood (ML) classifiers] were used, and then a majority voting (MV) between different outputs was developed by Giorgio Licciardi and Fabio Pacifici.

2. Devis Tuia and Frederic Ratle use both spectral and spatial features. The spectral features are a six-principal-component (PC) analysis (PCA) extraction of the initial pixel's vector value. The spatial information is extracted using morphological operators. These features are classified by combining several support vector machines (SVM) using MV.

3. Saurabh Prasad and Terrance West use wavelet-based preprocessing of the initial spectra followed by a linear discriminant analysis (LDA) and an ML classifier.

4. Ferdinando Giacco and Christian Thiel use a PCA to reduce the dimension of the data. Spatial information is taken into account with some textural features. The classification is achieved using SVM one-versus-one classifiers, and a spatial regularization is performed on the classification map to eliminate isolated pixels.

5. Jordi Inglada and Emmanuel Christophe perform a Bayesian fusion of different classifiers (such as SVM classifiers). The weight assigned

to each classifier is determined by the quantitative results it obtained. All these algorithms are available with the ORFEO Toolbox, which is an open source library of image processing algorithms for remote sensing applications (http://www.orfeo-toolbox.org).

Among these five algorithms, our technique was ranked as first.

6.2 Majority voting between NN and ML classifiers

The aim of the contest implicates the reduction of data set dimensionality to both decrease the complexity of the classifier and the computational time required, preserving most of the relevant information of the original data [87][88]. The proposed technique was based on a dimensionality reduction of the input measurements vector, and then a majority voting between the results obtained from neural networks and maximum likelihood classifiers applied on the reduced dataset.

6.2.1 Dimensionality reduction

The pre-processing procedure exploited, divided the hyper-spectral signatures into adjacent regions of the spectrum and approximates their values by piecewise constant functions PCF as in fig.6.1. This simple representation has shown to outperform most of the linear feature reduction methods proposed in literature, such as principal components analysis, minimum noise fraction, sequential forward selection or decision boundary feature extraction [89]. Assume S_{ij} to be the value of the ith pixel in the jth band, with a total of N pixels. The spectral signatures of each class extracted from

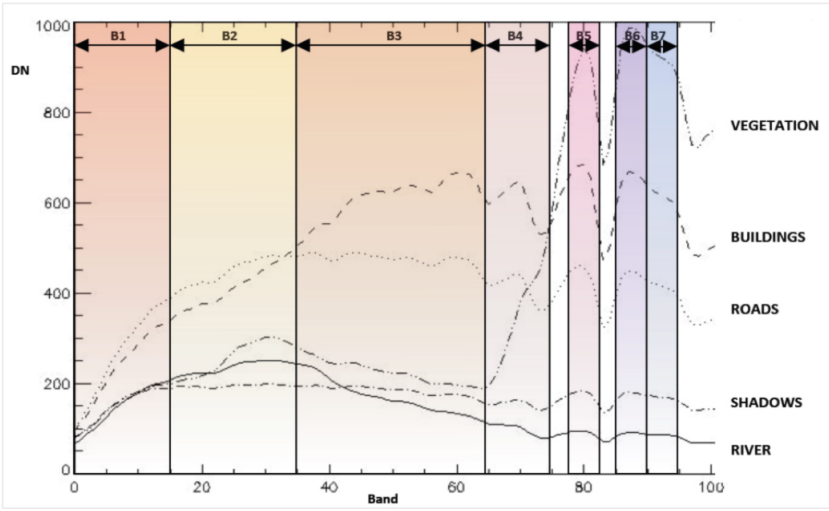


Figure 6.1: *Spectral signatures divided into adjacent regions*

ground truth pixels have been partitioned into a fixed number of contiguous intervals with constant intensities minimizing the mean square error:

$$H = \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j \in I_k} (s_{ij} - \mu_{ik})^2 \tag{6.1}$$

Where a set of K breakpoints define continuous intervals Ik, while μ_{ik} represents the mean value of each pixels interval between breakpoints. A number of K = 7 breakpoints was found to be a reasonable compromise between model complexity and computational time as show in table 6.1. Once the data has been reduced we proceeded with the classification phase.

	Sensor bands	Wavelength (μm)
	from-to	from-to
B1	1-15	430-486
B2	16-35	490-566
B3	36-65	570-686
B4	66-75	690-726
B5	78-82	730-766
B6	86-90	770-786
B7	91-95	790-834

Table 6.1: *Resulting subbands*

6.2.2 Classification

The classifier scheme exploited here is based on a combination of single decision maps. In [90], it has been demonstrated that combining the decisions of independent classifiers can lead to better classification accuracies. The combination can be implemented using a variety of strategies, among which majority voting (MV) is the simplest, and it has been found to be as effective as more complicated schemes[90][91]. Majority voting was used on five independent maps resulting from two different methods, i.e. three neural networks and two maximum likelihood classifiers, derived using three different training sets as shown in table 6.2. The five classifiers with the best individual classification accuracy drove this choice.

For each method, the seven features obtained by the reduction of input dimensionality composed the input to the classifier, while the outputs were the five class of interest. The algorithm of majority voting was implemented following two simple rules:

Set	**Buildings**	**Roads**	**Shadows**	**Vegetation**	**Water**
Set 1	132.369	18.914	20.356	53.065	43.104
Set 2	33.168	6.525	3.260	14.323	26.816
Set 3	45.268	5.210	1.524	17.485	20.367

Table 6.2: *training samples used for the supervised classifiers*

- A class is the winner if it recognized from the majority of the classifiers
- In case of a balance voting, the winner class is the one with the highest K-coefficient.

The final accuracy obtained using this method was 98.84% and the final land cover map is shown in fig.6.5.

6.3 NLPCA approach

Starting from the results obtained from the Contest, it became interesting to test the ability of the NLPCA with the same dataset.

6.3.1 Dimensionality reduction

Starting from the same 102 bands, a reduction of the input vector was performed through the NLPCA. The number of nonlinear components obtained training an AANN was set to 3, leading to a 102-36-3-36-102 topology. This choice was driven by the fact that the statistical information contained in the first three components of the PCA was over 99%. For this reason the number of nodes of the bottleneck layer of the AANN was set to 3. A first analysis of the nonlinear components, shown in fig.6.2, highlights

Figure 6.2: *3 nonlinear principal components obtained from the original image*

an optimal separation of the types existent in the scene. More in particular, component 1 exalts the water component. As the same way, in the component 2 and 3 the red-clay roofs of the buildings and the vegetation are evidenced, respectively. Another important note regards the component 1, which is not influenced by the buildings shadows. More in particular the roads covered by shadows and the roads without shadows in the original image, present in the component 1 the same values, as show in fig.6.2.

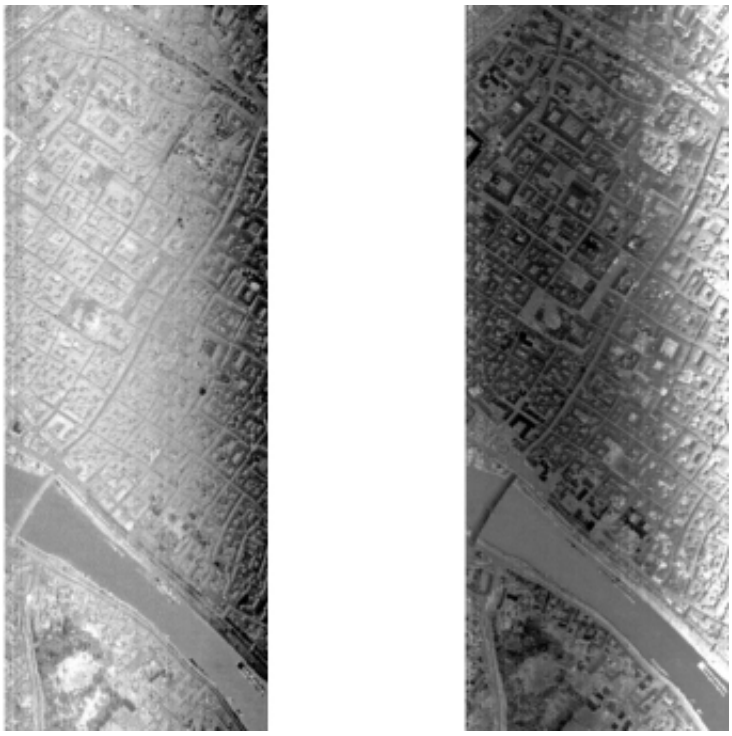


Figure 6.3: *Smile effect on one component of PCA (left) and MNF (right)*

From an accurate analysis of the original image emerged that it is influenced by the presence of the "smile" effect [2]. This effect cannot be detected by a simple band inspection, but became evident if we apply the MNF or PCA transformations, as shown in fig.6.3. PCA and MNF, as dimensionality reduction techniques, seem to be strongly influenced by the presence of "smile" effect artifacts. In such cases a further preprocessing stage to mitigate this problem is necessary. On the other hand, the use of NLPCA does not require any preprocessing as can be seen in fig.6.2.

6.4 Classification

The 3 nonlinear components were used as input to a new neural network to obtain a pixel based classification of the original image. In this case the network topology is a MLP of 3-9-9-5. About 1.500.000 pixels were considered for the training phase, which consisted of less than 500 epochs. In particular, for this case, the whole architecture consists of the following layers: one input layer of 102 units, 4 hidden layer of 36, 3, 9, 9 neurons, respectively and an output vector of 5 components. In fig.6.4, we report the confusion matrices computed for NLPCA and the PCF approaches on a set of about 330.000 pixels (not used in the training phase) and in fig.6.5 the relative classification maps. The overall classification accuracy is of about 99% for the NLPCA approach and over 95% for the PCF approach.

6.5 Conclusions

The output of the contest provided some interesting conclusions and perspectives. First of all, it became evident that the best five results were obtained by the use of a supervised method. Other results, obtained by unsupervised methods, were outperformed by the supervised methods. Among the best five techniques, those implementing neural provide the best classification performances. Another point regards the reduction of the input dimensionality, in fact, most of the proposed methods, to mitigate the Hughes effect [22], used a dimension reduction as a preprocessing procedure. Most of them used the PCA or MNF, retaining various numbers of components. But, from an accurate analysis of the image, it became evident

	WA	VE	SH	BU	RD	TOTAL
WA	**99,90%**	0,00%	0,00%	0,00%	0,00%	5,59%
VE	0,00%	**100,00%**	0,00%	0,00%	0,00%	10,92%
SH	0,00%	0,00%	**94,61%**	0,00%	1,80%	2,95%
BU	0,00%	0,00%	2,63%	**100,00%**	5,00%	78,49%
RD	0,10%	0,00%	2,77%	0,00%	**93,20%**	2,05%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

	WA	VE	SH	BU	RD	TOTAL
WA	**99,73%**	0,00%	0,00%	0,00%	0,00%	5,58%
VE	0,00%	**99,96%**	0,00%	1,60%	0,00%	12,17%
SH	0,22%	0,00%	**95,78%**	2,32%	1,47%	4,81%
BU	0,00%	0,03%	4,22%	**94,30%**	0,40%	73,99%
RD	0,05%	0,01%	0,00%	1,77%	**98,13%**	3,45%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Figure 6.4: *Confusion matrices for NLPCA (top) and PCF (bottom) approaches with overall accuracies of 99.68% and 95,35% respectively*

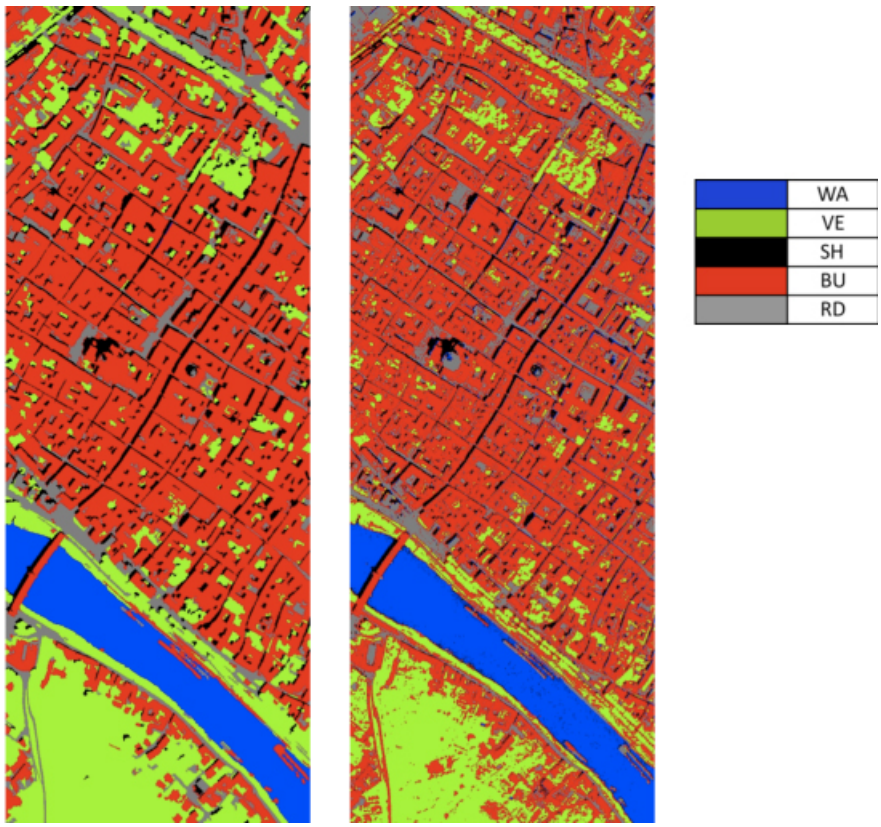


Figure 6.5: *Classification map for PCF (left) and NLPCA (right) approaches*

that both techniques could not be used until the smile effect persists. Compared to the other techniques, NLPCA does not emphasize the smile effect and does not require any preprocessing routine. Also the reduced dataset obtained through the use of PCF does not show any presence of smile artifacts but NLPCA, compared with PCF, allowed performing the training of the neural algorithm for the classification map in a limited number of epochs, which besides involving a faster training time, may be important to avoiding overfitting.

Conclusions

In this research work the potential of neural networks algorithms and architectures for extracting information from hyperspectral data has been investigated. Different aspects has been considered. Besides the one concerning with the actual inversion problem, in our case a pixel-based classification, a novel dimensionality reduction approach based on extraction of nonlinear principal components from hyper-spectral data has been developed. With the hyper-spectral imagery, the need of a technique to reduce the dimension of the huge input vector preserving as much information content as possible is an essential step for the analysis. Until now, the suitability of neural networks to handle nonlinear correlations in the data has not been investigated yet and one of the main puroses of our study was to give a contribution under this point of view. Moreover, neural networks have been also introduced as a tool for addressing the unmixing problem, which is also of particular importance in the context of hyperspectral imagery analysis.

The novel neural networks techniques have been applied to three hyperspectral datasets characterized by very different bandwidth, spatial resolution, acquisition mode and context:

- Acquisition by INTA-AHS instrument on a rural area. Image characterized by 75 bands with a spectral resolution from 30 to 500 nm and a spatial resolution of 5,5 meters.

- Acquisitions by CHRIS-Proba instrument on an area with the presence of urban settlements and large cultivations. The dataset was characterized by multi-angle and multi-date acquisitions with a spectral resolution varying from 5 to 11 nm and a spatial resolution of 18 meters.

- Acquisition by ROSIS instrument on dense urban area, characterized by 102 bands at 1 meter spatial resolution and 4 nm spectral resolution.

For each experiment a NN a final compact architecture performing both the feature extraction and classification phases has been designed. The experiments put in evidence different issues in favour of the presented NN methodology.

First of all the yielded land cover maps are characterized by accuracies which are better than those obtained using alternative techniques, for example considering PCA or MNF for features reduction. This means that the NLPCA are more effective in retaining the information content of the raw data. Indeed, NLPCA allowed performing the training of the neural algorithm for the classification map in a limited number of epochs, which, besides involving a faster training time, may be important to avoiding overfitting.

Another point is that the developed approach does not require any

pre-processing routine, resulting more robust to instruments distorsion artifacts. This is particularly evident in the two airborne experiments (AHS and ROSIS), where the strong influence of the smile effect prevents other dimensionality reduction techniques from being applied without a preliminary correction of the original data.

NLPCA technique seems also to have an interesting "object-oriented" property. It often happened that single NLPCA components were polarized on single objects, for example clouds or watercourses, which could be of great usefulness when there is a need of routines dedicated to the production of specific thematic maps.

Finally a new neural network technique for the spectral unmixing of hyper-spectral imagery was developed. Until now, the unmixing procedure relies on the assumption that the endmembers should be the most uncorrelated possible to achieve an acceptable result. In this work the ability of the neural networks to correctly estimate the abundances of strongly correlated endmembers in each pixel has been demonstrated.

Acronyms

AANN: AutoAssociative Neural Network
AATSR: Advanced Along-Track Scanning Radiometer
AHS: Advanced Hyper-spectral Scanner
ASI: Agenzia Spaziale Italiana (Italian Space Agency)
AVHRR: Advanced Very High Resolution Radiometer
AVIRIS: Airborne Visible Infra Red Imaging Spectrometer
BSQ: Band Sequentia
CASI: Compact Airborne Spectrographic Imager
CCD: Charge Coupled Device
CHRIS: Compact High Resolution Imaging Spectrometer
DEMMIN: Durable Environmental Multidisciplinary Monitoring Information Network
DFTC: Data Fusion Technical Committee
DLR: Deutschen zentrums fur Luft- und Raumfahrt (German Space Agency)
DN: Digital Number
EnMAP: Environmenta Mapping and Analysis Program
EO: Earth Observation
EOC: Electro Optical Camera

ESA: European Space Agency

FOV: Field Of View

FTS: Fourier Transform Spectrometer

FWHM: Full Width at only Half Maximum value

FZA: Fly-by Zenith Angle

GRS-S: Geoscience and Remote Sensing Society

GSD: Ground Sampling Distance

IEEE: Institute of Electrical and Electronics Engineers

IFOV: Istantaneous Field Of View

K-L: Karhunen-Loeve

KOMPSAT: Korea Multi-Purpose SATellite

LDA: Linear Discriminant Analysis

LSU: Linear Spectral Unmixing

MAF: Minimum/Maximum Autocorrelation Factor

MERIS: MEdium Resolution Imaging Spectrometer

MIR: Medium Infra Red

ML: Maximum Likelihood

MLP: Multi-Layer Perceptron

MNF: Minimum/Maximum Noise Fraction

MODIS: MODerate Resolution Spectroradiometer

MSE: Mean Square Error

MV: Majority Voting

NASA: National Aeronautics and Space Administration

NIR: Near Infra Red

NLPCA: NonLinear Principal Component Analysis

NN: Neural Network

PC: Principal Component

PCA: Principal Component Analysis

PCF: Piecewise Constant Function

PRISMA: PRecursore IperSpettrale della Missione Applicativa

PROBA: Project for On Board Autonomy

PSF: Point Spread Factor

REP: Red Edge Position

RMSE: Root Mean Square Error

ROSIS: Reflective Optics System Imaging Spectrometer

SBFS: Sequential Backward Floating Selection

SBS: Sequential Backward Selection

SFFS:Sequential Forward Floating Selection

SFS: Sequential Forward Selection

SNNS: Stuttgart Neural Network Simulator

SNR: Signal-to-Noise Ratio

SVM: Support Vector Machine

SWIR: Short Wave Infra Red

TDI: Time Delay Integration Imager

TIR: Termal Infra Red

TMA: Three Mirror Anastigmatic

VIS: VISible

VNIR: Visible Near Infra Red

WIS: Wedge Imaging Spectrometer

Bibliography

[1] F. Gires and P. Tournois, "Interférométre utilisable pour la compression d'impulsions lumineuses modulées en fréquence," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 258, pp. 6112–6115, 1964.

[2] H. P. White, K. S. Khurshidt, R. Hitchcock, R. Neville, S. Lixin, C. M. Champagne, and K. Staenz, "From at-sensor observtion to at-surface reflectance - calibration steps for earth observation hyperspectral sensors," *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04 Proceedings*, vol. 5, pp. 3241–3244, 2004.

[3] D. N. H. Horler, M. Dockay, and J. Barber, "the red edge of plant leaf reflectanceÓ," *International journal of remote sensing*, vol. 4, no. 2, pp. 273–288, 1983.

[4] J. Clevers, S. M. D. Jong, G. F. Epema, F. V. D. Meer, W. J. Bakker, A. K. Skidmore, and K. H. Scholte, "derivation of red edge index using meris standard band setting," *international journal of remote sensing*, vol. 23, no. 16, pp. 3169–3184, 2002.

[5] "http://www.itres.com/home."

[6] "http://www.inta.es."

[7] "http://www.cgrit.it/tecnologie/mivis.html."

[8] "http://aviris.jpl.nasa.gov/."

[9] "http://www.op.dlr.de/dais/hysens/."

[10] "http://eo1.usgs.gov/hyperion.php."

[11] "http://www.chris-proba.org.uk/."

[12] "http://www.asi.it/it/flash/osservare/prisma."

[13] "http://www.enmap.org/."

[14] M. Ichino and J. Sklansky, "Optimum feature selection by zero-one integer programming," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-14, no. 5, pp. 737–746, 1984.

[15] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–189, 1997.

[16] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2973–2987, 2009.

[17] J. Levesque and K. A. Staenz, "method for monitoring mine tailings revegetation using hyperspectral remote sensing," *Geoscience and*

Remote Sensing Symposium, 2004. IGARSS '04 Proceedings, vol. 1, pp. 20–24, 2004.

[18] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1552–1565, 2004.

[19] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognit.*, vol. 33, no. 1, pp. 25–41, 2000.

[20] M. Pesaresi and J. A. Benediktsson, "Classification of urban high-resolution satellite imagery using morphological and neural approaches," *Geoscience and Remote Sensing Symposium, 2000. IGARSS '00 Proceedings*, vol. 7, pp. 3066–3068, 2000.

[21] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 2, pp. 309–320, 2001.

[22] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[23] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Wiley-InterScience, 2003.

[24] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 15, no. 4, pp. 388–400, 1993.

[25] E. Merenyi, W. H. Farrand, L. E. Stevens, T. S. Melis, and K. Chhibber, "Mapping colorado river ecosystem resources in glen canyon: Analysis of hyperspectral low-amplitude aviris imagery," *Proc. ERIM, 14th Int. Conf. and Workshops Appl. Geologic Remote Sens.*, pp. 44–51, 200.

[26] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, 2002.

[27] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, 2005.

[28] P. R. Marpu, P. Gamba, and I. Niemeyer, "Hyperspectral data classification using an ensemble of class-dependent neural networks," *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop on*, pp. 1–4, 2009.

[29] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 8, pp. 1778–1790, 2004.

[30] P. Pudil, "Feature selection toolbox software package," *Pattern Recognit. Lett.*, vol. 23, no. 4, pp. 487–492, 2002.

[31] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis.* Springer, 1999.

[32] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, 1994.

[33] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *IPattern Recognit. Lett.*, vol. 10, no. 5, pp. 335–347, 1989.

[34] W. Siedlecki and J. Sklansky, "On automatic feature selection," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 2, no. 2, pp. 197–210, 1988.

[35] T. Li, C. Chen, and C. Su, "Comparison of neural and statistical algorithms for supervised classification of multi-dimensional data," *Int. J. Indus. Eng.-Theory Appl. Pract.*, vol. 10, no. 1, pp. 73–81, 2003.

[36] T. M. Lillesand and R. Kiefer, *Remote sensing and image interpretation.* John Wiley and Sons, 2000.

[37] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, K. Richter, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semisupervised support vector machine," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, no. 2, pp. 248–252, 2009.

[38] R. Pu, P. Gong, G. Biging, and M. R. Larrieu, "Retrieval of surface reflectance and lai mapping with data from ali, hyperion and aviris," *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02 Proceedings*, vol. 3, pp. 1411–1413, 2002.

[39] D. Haboudane, N. Tremblay, P. Vigneault, and J. R. Miller, "Indices-based approach for crop chlorophyll content retrieval from hyperspectral data," *Geoscience and Remote Sensing Symposium, 2007. IGARSS '07 Proceedings*, vol. 1, pp. 3297–3300, 2007.

[40] J. F. Moreno, F. Baret, M. Leroy, M. Menenti, M. Rast, and M. Shaepman, "Retrieval of vegetation properties from combined hyperspectral/multiangular optical measurements: results from the daisex campaigns," *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03 Proceedings*, vol. 4, pp. 2875 – 2877, 2003.

[41] F. D. Frate and D. Solimini, "On neural network algorithms for retrieving forest biomass from sar data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 1, pp. 24–34, 2004.

[42] F. D. Frate, M. Iapaolo, S. Casadio, S. Godin-Beekmann, and M. Petitdidier, "Neural networks for the dimensionality reduction of gome measurement vector in the estimation of ozone profiles," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 92, pp. 275–291, 2005.

[43] H. Bishof, W. Shneider, and P. A. J., "Multispectral classification of landsat images using neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 3, pp. 482–490, 1992.

[44] B. C., *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[45] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, 2003.

[46] F. D. Frate, J. Lichtenegger, and D. Solimini, "Monitoring urban areas using ers-sar data and neural networks," *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings*, vol. 5, pp. 2696–2698, 1999.

[47] G. Foody and M. Ajay, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, 2004.

[48] M. Friedl and C. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ*, vol. 61, pp. 399–409, 1997.

[49] P. Goel, S. Prasher, R. Patel, J. Landry, R. Bonnell, and A. A. Viau, "Classification of hyper-spectral data by decision trees and artificial

neural networks to identify weed stress and nitrogen status of corn," *Comput. Electron. Agricult.*, vol. 39, no. 2, pp. 67–93, 2003.

[50] W. Hsu, M. Lee, and J. Zhang, "Image mining: Trends and developments," *Remote Sensing Rev*, vol. 6, pp. 319–329, 2002.

[51] I. Kanellopoulos and G. Wilkinson, "Strategies and best practice for neural network image classification," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 711–725, 1997.

[52] D. McIver and M. Friedl, "Using prior probabilities in decision-tree classification of remotely sensed data," *Remote Sens. Environm.*, vol. 81, no. 2-3, pp. 253–261, 2002.

[53] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 grs-s data fusion contes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3857–3865, 2009.

[54] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03 Proceedings*, vol. 1, pp. 288–290, 2003.

[55] M. Molinier, J. Laaksonen, and T. Hame, "A self-organizing map framework for detection of man-made structures and changes in satellite imagery," *Geoscience and Remote Sensing Symposium, 2006. IGARSS '06 Proceedings*, vol. 1, pp. 1–4, 2006.

[56] H. K. Nabeel, A. A. Tawq, and H. A. Samaher, "Design and implementation of classification system for satellite images based on soft computing techniques," *Information and Communication Technologies, 2006. ICTTA '06*, vol. 1, pp. 430–436, 2006.

[57] J. D. Paola and R. A. Schowengerdt, "A review and analysis of backpropagation neural networks for classification of remotely-sensed multispectral imagery," *nt. J. Remote Sens.*, vol. 16, no. 16, pp. 3033–3058, 1995.

[58] D. R. Peddle and S. E. Franklin, "Multisource evidential classification of surface cover and frozen ground," *Int. J. Remote Sens.*, vol. 13, no. 17, pp. 3375–3380, 1992.

[59] D. Peddle, G. Foody, A. Zhang, J. Franklin, and E. L. Drew, "Multisource image classification ii: An empirical comparison of evidential reasoning and neural network approaches," *Can. J. Remote Sens.*, vol. 20, pp. 396–407, 1994.

[60] D. Rumelhart, G. Hinton, and R. Williams, *Learning internal representations by error propagation, in Parallel Distributed Processing.* MIT Press, 1986.

[61] G. Wilkinson and J. Megier, "Evidential reasoning in a pixel classification hierarchy-a potential method for integrating image classifiers and expert system rules based on geographic context," *Int. J. Remote Sens.*, vol. 11, no. 10, pp. 1963–1968, 1990.

[62] J. Rogan, J. Franklin, and D. Roberts, "A comparison of methods for monitoring multi-temporal vegetation change using thematic mapper imagery," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 143–156, 2002.

[63] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sens. Environ*, vol. 86, no. 4, pp. 554–565, 2003.

[64] C. Liu, L. Zhang, C. Davis, D. Solomon, T. Brann, and L. Caldwell, "Comparison of neural networks and statistical methods in classification of ecological habitats using fia data," *Forest Sci.*, vol. 48, pp. 619–631, 2003.

[65] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Academic, 2nd ed., 1990.

[66] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901.

[67] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4-5, pp. 411–430, 2000.

[68] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, 1988.

[69] M. Bressan and J. Vitri, "Nonparametric discriminant analysis and nearest neighbor classification," *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2743– 2749, 2003.

[70] X. Jia and J. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, 1999.

[71] T. A. Warner, K. Steinmaus, and H. Foote, "An evaluation of spatial autocorrelation feature selection," *Int. J. Remote Sens.*, vol. 20, no. 8, pp. 1601–1616, 1999.

[72] S. Mallat, *A Wavelet Tour Of Signal Processing.* Academic, 1999.

[73] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005.

[74] P. Pudil, J. Novovicova, N. Choakjarernwanit, and J. Kittler, "Feature selection based on the approximation of class densities by finite mixtures of special type," *Pattern Recognit.*, vol. 28, no. 9, pp. 1389–1398, 1994.

[75] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE J.*, vol. 37, pp. 233–243, 1991.

[76] J. Boardman, "Automating spectral unmixing of aviris data using convex geometry concepts," *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, vol. 1, 1993.

[77] R. E. Steinkraus and R. W. Hickok, *AVIRIS onboard data handling and control*, 1987. http://makalu.jlp.nasa. gov/docs/aviris87/TOC. HTM.

[78] G. Vane, T. G. Chrien, E. A. Miuer, and J. H. Reimer, *Spectral and radiometric calibration of the Airborne Visible/Infrared Imaging Spectrometer,*, 1987. http://makalu.jlp.nasa. gov/docs/aviris87/TOC. HTM.

[79] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295–304, 1990.

[80] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks*, vol. 2, p. 459, 1989.

[81] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biological Cybernetics*, vol. 59, pp. 291–294, 1988.

[82] J. A. Gomez, E. de Miguel, O. G. de la Camara, and A. Fernandez-Renau, "status of the inta ahs sensor," *5th EARSeL Workshop on Imaging Spectroscopy Proceedings*, 2007.

[83] A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using mor-

pholgical preprocessing and independent component analysis," *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05 Proceedings*, vol. 1, pp. 176–179, 2009.

[84] "http://www.ra.cs.uni-tuebingen.de/snns/."

[85] R. Duca and F. D. Frate, "Hyperspectral and multiangle chris-proba images for the generation of land cover maps," *IEEE Trans. Geosci. Remote Sens*, vol. 46, no. 10, pp. 2857 – 2866, 2008.

[86] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, 2002.

[87] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1368–1379, 2001.

[88] V. W. Samawi and O. Basheer, "The effect of features reduction on different texture classifiers," *Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on*, pp. 67–72, 2008.

[89] S. B. Serpico, M. D'Inca, F. Melgani, and G. Moser, "Comparison of feature reduction techniques for classification of yperspectral remote sensing data," *Proc. SPIE Image Signal Process. Remote Sens.*, vol. 4885, no. VIII, pp. 347–358, 2003.

[90] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, 1994.

[91] L. Lam and S. Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 27, no. 5, pp. 553–568, 1997.