Information extraction and dimensionality reduction of hyperspectral datasets through spectral region analyses

Hosseini Aria, Enayat

Citation (APA)
Hosseini Aria, E. (2018). *Information extraction and dimensionality reduction of hyperspectral datasets through spectral region analyses*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:72b4a11a-d394-433a-b1d1-1f40eb8bd8c6

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Information extraction and dimensionality reduction of hyperspectral datasets through spectral region analyses

Information extraction and dimensionality reduction of hyperspectral datasets through spectral region analyses

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
maandag 14 mei 2018 om 12:30 uur

door

Seyed Enayat HOSSEINI ARIA

Master of Science in Surveying & Geomatics Engineering & Remote Sensing
University of Tehran, Iran
geboren te Mashhad, Iran

Dit proefschrift is goedgekeurd door de promotor en de copromotor.

Samenstelling promotiecommissie bestaat uit:

Rector magnificus,	voorzitter
Prof. dr. M. Menenti	Technische Universiteit Delft, promotor
Dr. Ir. B.G.H Gorte	Universiteit New South Wales, copromotor

onafhankelijke leden:

Prof.dr. J. Zallat	Universiteit Strasbourg, France
Prof.dr.ir. W.G.M. Bastiaansen	Technische Universiteit Delft / IHE Delft
Prof.dr.ing. W. Verhoef	Universiteit Twente
Dr. S.L.M. Lhermitte	Technische Universiteit Delft
Prof.dr.ir. B.J.H. van de Wiel	Technische Universiteit Delft, reservelid

Front & Back: Designed by S.E. Hosseini Aria

An electronic version of this dissertation is avalable at
http://repository. tudelft.nl/.

Acknowledgment

Reflecting on the whole career path during my Ph.D. journey, I regard it as an incredibly unique experience and adventure. It has been a procedure consisting of both ups and downs with valuable struggles that shaped me and my future path. Many people contributed to the accomplishment of this dissertation in various ways, for which I am sincerely grateful.

First and foremost, I would like to thank my promoter prof. dr. Massimo Menenti, for providing me the opportunity to begin my PhD career at TUDelft, for his support, supervision, freedom and flexibility he gave me during the research, and for the careful reading of the dissertation and critical feedback he provided me. I appreciate all your contributions of time, great ideas, and funding to make my Ph.D. experience productive and stimulating. Second, I would like to thank my co-promotor dr. ir. Ben Gorte for help and guidance throughout these years. I appreciate your suggestions and challenging questions, that led me to think twice.

I am thankful to my (former) colleagues and friends at GRS department for sharing their knowledge and providing a friendly work environment. My sincere thanks also go to prof. Ramon Hanssen for his official support to accomplish my thesis, and to dr. Roderik Lindenbergh who always affably gave me helpful tips. I would, furthermore, like to appreciate the help of Marcel Kleinherenbrink for providing me the Dutch translation of the summary in the tight time I had, and Sami Samiei who devoted his time to discuss scientific challenges that I have encountered. I am grateful to Hamid Ghafarian, Prabu Dheenathayalan, Lorenzo Iannini, Siyavash Shakeri, Silvia Alfieri, Ramses Molijn, and Jinhu Wang for sharing their experiences with me.

I would like to express my appreciation to my defence committee members: prof. J. Zallat, prof. W.G.M. Bastiaansen, prof. W. Verhoef, dr. S.L.M. Lhermitte, and prof. B.J.H. van de Wiel for the time they spent at the thorough reading of this study and their constructive comments.

I gratefully thank Lidwein de Jong, Danko Roozemond, Dian Verbunt, Rebbeca Domino, Marjolein de Niet-de Jager, Debbie Rietdijk, Irma Zomerdijk and Suzanne de Hoog-Dolleefor for their kind administrative support.

My special thank goes to my family friends, Ali and Maryam, Sasan and Linda, and Petra Steuteknuel for their sincere help to my family and me.

Lastly, I would like to thank my family for all their love and encouragement. For my parents who believed in me and supported me in all my pursuits. For my beloved daughter, Asal, who brought a lot of joy and happiness to my life. And most of all for my loving, supportive, encouraging, and patient wife, Fariba, whose constant support and sacrifice during my Ph.D. is so appreciated. Thank you, my dear love.

S.Enayat Hosseini Aria
Delft, May 2018

Summary

Hyperspectral images present detailed spectral information of every pixel in the images where the spectral signal is sampled in hundreds of narrow and contiguous spectral channels, usually covering the 400-2500 nm spectral region where sunlight reflected by the Earth can be measured. Earth observation systems acquire spectral information by imaging spectrometers mounted in a platform flying over the Earth. Recent advances in technology make it possible to have miniaturised hyperspectral satellites in orbit. Much of the work presented in this thesis was inspired by the study of a CubeSat equipped with an imaging spectrometer and capable of onboard data processing.

Utilising hyperspectral datasets, however, is a big challenge. At first glance, the volume of the data compared to multispectral images is huge, thus making data transmission, analyses and storage a challenge. Secondly, hyperspectral datasets suffer from redundant spectral information, since adjacent narrow spectral channels are usually highly correlated. The third issue with hyperspectral images is that a huge number of samples are required to characterise observed targets, as in object recognition and classification, because of the large number of spectral variables (channels) and of the high radiometric resolution. So, the number of available samples in the hyper-dimension of the data makes the feature space inherently sparse. This issue is also called the "curse of dimensionality".

The primary solution to the mentioned problems is the dimensionality reduction (DR) of hyperspectral datasets. To investigate DR, we identified three leading objectives to extract optimal information and reduce the dimensionality. These are 1- accurate spectral representation 2-independent spectral information 3- optimal class separability. In this thesis, we propose a methodology to identify reduced spectral configurations to meet the three objectives. Our approach divides the spectrum into continuous spectral regions with different widths depending on the criterion used and the land cover in the scene, by applying either a top-down (the spectral region splitting (SRS) algorithm) or a bottom-up (the clustering-merging (CM) algorithm) searching strategy. It is similar to feature selection (FS) technique preserving the relationship between the physical properties of targets and the spectral features.

In this study, various criteria per objective have been investigated. For spectral reflectance representation (first objective), five criteria are used. They are RMSE (root-mean-square-error), ECS (Euclidean distance of cumulative spectrum), SA (spectral angle), SID (spectral information divergence), and a hybrid metric. The results have also been compared with the best signal approximation method using piecewise constant function approximation (PCFA). For identification of independent spectral regions (second objective), a metric named total dependence (TD) has been proposed and applied. Besides, a detailed investigation of TD and evaluating its capability for the objective, we have compared the final feature sets with the sets obtained by well-known FS algorithms focusing on the selection of independent channels using the mutual information metric and the similarity-based endmember extraction techniques. Six class separability metrics from the simple one that just considers the inter-class distance to the most advanced ones, e.g. Jeffreys-Matusita (JM) considering the intra-class distance as well, have been utilised to identify band-sets with the optimal separabilities between the available classes in a scene (third objective). The final feature sets have been compared with the sets obtained by the branch and bound (BB) algorithm, providing the best channel set regarding a given criterion, sequential forward selection (SFS), and sequential forward floating selection (SFFS).

The proposed methodology has been applied to different spectra of a complete spectral library, and various hyperspectral datasets and the band configurations obtained have been evaluated and compared with well-known and advanced FS algorithms concerning material detection, image classification and the complexity of the algorithms in the supervised and unsupervised situations.

The proposed algorithm, SRS, can make configuration that represent the spectral reflectance in hyperspectral images with a negligible error of representation, and, on average, is 40 times faster than the other method evaluated for this purpose. The independent band set obtained by the CM algorithm provides the set in a fraction of a second, while the result of image classification is better or comparable with the feature sets obtained by other algorithms with the same objective. Applying SRS to identify the spectral regions to optimal class separability give spectral bands providing a better class discrimination than the best channels selected by the BB algorithm using the same separability metric.

This thesis, in the framework of the three DR objectives on hyperspectral datasets, presents a detailed study through spectral region analyses which leads to the identification of different spectral band configurations. Comparing the configurations, we conclude that in the case of a supervised dimensionality reduction, it is worth to identify separable spectral bands leading to more accurate image classification. However, in the case of the lack of information about the classes in a scene (unsupervised) or the lack of interest in a specific region in the scene, we suggest identifying the independent and representative band sets, while the former requires short computation time, and the latter provides a better

classification accuracy. This argument could address the issue regarding the onboard preprocessing of hyperspectral images in a Cubesat as well.

Samenvatting

Hyperspectrale beelden geven gedetailleerde spectrale informatie van elke pixel weer, waarbij het spectrale signaal bemonsterd is in honderden smalle, aaneengesloten spectrale banden. Doorgaans omvatten de banden het bereik van 400 tot 2500 nm, waarin gereflecteerd zonlicht kan worden gemeten. Aardobservatiesystemen verzamelen spectrale informatie met behulp van spectrometers aan boord van een vliegend platform. Recente ontwikkelingen in technologie maken het mogelijk om kleine hyperspectrale satellieten in een baan om de aarde te brengen. Veel van het werk in deze thesis is geïnspireerd op het bestuderen van data van een Cubesat, die uitgerust is met een spectrometer en die aan boord data verwerkt.

Effectief gebruik van hyperspectrale data is een uitdaging. Ten eerste is het volume van de data enorm groot ten opzichte van multispectrale beelden, wat de datatransmissie, -analyse en -opslag bemoeilijkt. Ten tweede bevatten hyperspectrale data veel overtollige informatie, omdat aangrenzende spectrale kanalen vaak sterk gecorreleerd zijn. Het derde probleem is het grote aantal metingen dat nodig is om waargenomen objecten te beschrijven (in bijvoorbeeld objectherkenning en klassificatie), door het hoge aantal spectrale variabelen (kanalen) en door de hoge radiometrische resolutie. Dit zorgt ervoor dat de beschikbare gegevens de kenmerken van een object in de hyperspectrale ruimte slechts uiterst summier omschrijven. Dit wordt ook wel de 'vloek van de dimensionaliteit' genoemd.

De eerste oplossing voor de genoemde problemen is dimensionaliteitsreductie (DR) van de hyperspectrale data. Om DR te onderzoeken, hebben we drie hoofddoelen opgezet met als doel optimale informatie te extraheren, terwijl we de dimensionaliteit verkleinen. Dit zijn 1- nauwkeurige spectrale representatie, 2- onafhankelijke spectrale informatie, 3- optimale klassenscheiding. In dit proefschrift stellen we een methodologie voor om gereduceerde spectrale configuraties op te stellen, die de doelen bereiken. Onze methode verdeelt het spectrum in continue spectrale deelgebieden met verschillende breedtes, die afhankelijk zijn van de gebruikte criteria en de landbedekking in het geobserveerde gebied, door middel van een top-down (het spectrale-regio-splitsingalgoritme

(SRS)) of een bottom-up (het clusteren-samenvoegingsalgoritme (CM)) zoekstrategie. Het is vergelijkbaar met een kenmerkselectietechniek (FS), waarbij de verhouding tussen de fysieke eigenschappen en de spectrale kenmerken wordt behouden.

In deze studie zijn verschillende criteria voor de doelstellingen onderzocht. Voor de eerste doelstelling (spectrale reflectierepresentatie) zijn vijf criteria gebruikt, waaronder: RMSE (kwadratisch gemiddelde afwijking), ECS (Euclidische afstand van het cumulatieve spectrum), SA (spectrale hoek), SID (spectrale informatiedivergentie) en een hybride test. De resultaten zijn ook vergeleken met de beste signaal benaderingsmethode gebaseerd op een stuksgewijs continue functiebenadering (PCFA). Voor de identificatie van onafhankelijke spectrale deelgebieden (tweede doelstelling), is de totale afhankelijkheidsmetriek (TD) voorgesteld en toegepast. Naast een gedetailleerd onderzoek van TD en een evaluatie van haar geschiktheid voor de doelstelling, hebben we een vergelijking gemaakt tussen de uiteindelijke kenmerkencombinaties en degene die bepaald zijn met behulp van bekende FS algoritmes. Hierbij is gefocust op de selectie van onafhankelijke kanalen, gebruikmakende van de wederzijde informatiemetriek en de overeenkomstgebaseerde endmember extractietechnieken. Zes klassenscheidingsmetrieken zijn gebruikt, van de eenvoudige interklasseafstand tot aan de geavanceerde Jeffreys-Matusita (JM), om bandcombinaties te identificeren om de klasses optimaal van elkaar te scheiden (derde doelstelling). De uiteindelijke kenmerkencombinaties zijn vergeleken met combinaties die gegenereerd zijn met het vertak en bindalgoritme (BB) , dat de beste combinatie oplevert bij een gegeven criterium, met sequentiële voorwaartse selectie (SFS, SFSS).

De voorgestelde methodologie is toegepast op verschillende spectra van een complete spectrale bibliotheek en verschillende hyperspectrale datasets. Vervolgens zijn de verkregen bandcombinaties geëvalueerd en vergeleken met resultaten van de bekende en geavanceerde FS algoritmes voor materiaaldetectie, beeldklassificatie en op grond van de complexiteit van de algoritmes in gecontroleerde en ongecontroleerde situaties.

Het voorgestelde algoritme, SRS, kan configuraties maken, die de spectrale reflecties representeren in hyperspectrale beelden met verwaarloosbare fouten, en daarbij gemiddeld veertig keer sneller is dan andere methode, die voor dit doel worden gebruikt. De onafhankelijke bandcombinaties, verkregen met behulp van het CM algoritme leveren de kenmerkencombinaties in een fractie van een seconde, terwijl de klassificatie op zijn minst gelijkwaardig en meestal beter is dan bij bestaande algoritmes. Door SRS toe te passen om spectrale deelgebieden te herkennen voor optimale klassenscheiding, worden klasses beter gescheiden dan met de deelgebieden geselecteerd door het BB algoritme, gebaseerd op dezelfde metriek.

Deze thesis presenteert, in het kader van de drie DR doelstellingen voor hyperspectrale datasets, een gedetailleerde studie van spectrale deelgebiedsanalyse,

die geleid heeft tot de identificatie van verschillende spectrale bandconfiguraties. Een vergelijking tussen de configuraties laat zien dat het nuttig is om spectrale banden te scheiden ten behoeve van een betere beeldklassificatie. Echter, als men weinig informatie heeft over de klasses in een gebied, of alleen belangstelling heeft voor deelgebied, stellen wij voor om òf onafhankelijke òf representatieve bandcombinaties te identificeren, omdat het ene een korte berekeningstijd nodig heeft en het andere een hogere klassificatienauwkeurigheid geeft. Deze redenering kan ook gebruikt worden met betrekking tot het aan boord verwerken van hyperspectrale beelden in een Cubesat.

Contents

1.
Introduction

Hyperspectral remote sensing imaging exploits solar illumination to acquire information from the surface by measuring reflectance spectra. The goal of hyperspectral imaging is the identification of materials and objects existing on the surface or in the atmosphere by using spectral information acquired as images. Along the advantages of hyperspectral images, the hyper-dimensionality of the data is a big challenge. It may cause 1- a large volume of data for storage and transmission; 2- redundant information in the spectral domain; and 3- complexity in data analysis. In this chapter, we explained these issues and explored the dimensionality reduction techniques of hyperspectral datasets which can be the principal remedy for these issues.

1.1. Background

By definition, remote sensing is the collection of information about an object without physical contact with it (Shaw and Burke 2003; Elachi and Van Zyl 2006). The acquisition of information is by detecting and measuring energy received from an object, such as electromagnetic or acoustic. The electromagnetic techniques of information acquisition are most commonly used in connection with the "remote sensing" term (Elachi and Van Zyl 2006), and are based on the interactions between the electromagnetic radiation and matter. The discipline that deals with the measurement and interpretation of spectra arising from the interactions between the electromagnetic radiation and matter is 'spectroscopy' (Herrmann and Onkelinx 1986; Skoog, Holler et al. 2007; Penner 2010; Hapke 2012).

When an electromagnetic wave reaches a material, a fraction of the incident energy is scattered in all directions[1], and the scattered photons are reflected away from the surface so they may be detected and measured. The study of light as a function of wavelength that has been scattered from a solid, liquid, or gas is reflectance spectroscopy (Clark and National Research Council (U.S.). 1966; Clark and Roush 1984). Reflectance spectroscopy, which is the basis of hyperspectral remote sensing imaging, allows us to derive information about the characteristics of materials from its reflected light, and also determines the absorption spectral regions of the materials investigated (Piecuch, Maruani et al. 2009; Hapke 2012). For example, metals have high reflectivity, but the reflectivity reduces with

---

[1]When the electromagnetic radiation interacts with an object; some amount is scattered, some transmitted and/or some absorbed, which can be summarized as $\rho+\alpha+\tau=1$, where ρ, α, and τ are the ratio of reflected, absorbed, and transmitted radiation to the incident radiation within a specific wavelength and temperature respectively (Swain, P. H. and S. M. Davis (1978). Remote Sensing: The Quantitative Approach. New York, NY, McGraw-Hill.; Elachi, C. and J. Van Zyl (2006). Introduction to the physics and techniques of remote sensing. Hoboken, N.J., Wiley-Interscience.).

temperature and drastically decreases with the formation of an oxide layer (Elachi and Van Zyl 2006).

1.2. Hyperspectral remote sensing

Hyperspectral remote sensing imaging is an image-based measurement of the spectra of the reflected energy arriving at the sensor from objects remotely observed, where a vector representing the energy across a specified range of wavelengths is measured for each image element (pixel). This technique is also known as imaging spectrometry. These spectra are used to extract information based on the signature of the interactions between the electromagnetic radiation and the target being observed as mentioned.

The majority of hyperspectral sensors, focusing on Earth's surface features, operate from the visible region (400 to 700 nm) through the shortwave infrared (about 2500 nm) spectral regions, exploiting solar illumination to acquire information from the surface by measuring their reflectance spectra. This spectral range provides rich information to identify and distinguish spectrally unique materials and many targets on the Earth surface (Shaw and Burke 2003; Bioucas-Dias, Plaza et al. 2013). Hyperspectral sensors record the spectral information in hundreds of narrow contiguous bands, e.g. ten nanometers wide[1] (Green, Eastwood et al. 1998).

The spectral information in hyperspectral images can be used, often in combination with high-spatial-resolution, to identify terrestrial unknown objects remotely (Shaw and Burke 2003). For example, Thenkabail et al. (Thenkabail, GangadharaRao et al. 2007) used spectral matching techniques to recognize several objects using image data such as crop types, plant species and minerals. The spectral information is also utilized to identify the components of mixed pixels (Parra, Spence et al. 1999; Manolakis, Siracusa et al. 2001; Manolakis, Marden et al. 2003; Vishnu, Nidamanuri et al. 2013; Shanmugam and SrinivasaPerumal 2014).

The concept of hyperspectral imaging started in the mid-80's (Wang and Zhao 2015), and to this point, it has been utilised most widely by geologists for the

---

[1] In order to achieve a continious coverage of a hyperspectral sensor and the spectral resolution of the sensor, the full width at half maximum (FWHM) of center wavelength of the spectral response function of indivitual channels, which are overlapping each other, should be calculated (Green, R. O., M. L. Eastwood, et al. (1998). "Imaging spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS)." Remote Sensing of Environment 65(3): 227-248, Gao, Z. Y., R. D. Jia, et al. (2016). "Simulation and Analysis of Spectral Response Function and Bandwidth of Spectrometer." International Journal of Aerospace Engineering.).

mapping and identification of minerals (Adams and Filice 1967; Adams 1974). For example, Goetz et al. (Goetz, Vane et al. 1985) tested the capability of imaging spectroscopy for mineral identification over the Cuprite mining district in Nevada, USA. The first Earth-looking hyperspectral imager was AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) mounted on an aircraft platform in 1987 (Green, Eastwood et al. 1998). 13 years later after the commissioning of AVIRIS, the first satellite carrying a hyperspectral sensor, named Hyperion, was launched on the EO-1 satellite (Folkman, Pearlman et al. 2001). An impressive spaceborne hyperspectral sensor is CHRIS (Compact High-Resolution Imaging Spectrometer), the prime instrument of PROBA-1 (Project for On-Board Autonomy) mission (Barnsley, Settle et al. 2004). The technology objective was to explore the capabilities of imaging spectrometers on small satellite platforms. The sensor is fully programmable and allows collecting hyperspectral images in different spectral and spatial modes. Table 1.1 compares the salient features of the hyperspectral sensors- AVIRIS, Hyperion, and CHRIS-PROBA.

Table 1.1. Comparison of hyperspectral imaging systems.

Parameters	AVIRIS	Hyperion	CHRIS-PROBA
Nominal altitude (km)	20	705	600
Swath (km)	11	7.6	14
Spatial resolution (m)	20	30	18 and 36 (at nadir)
Spectral coverage (nm)	400-2500	400-2500	410-1050
Nominal spectral resolution (nm)	10	10	1.3 nm at 410nm to 12 nm at 1050nm (i.e. it varies across the spectrum)
Number of wavebands	224	220	19 bands at a spatial resolution of 18m 63 bands at a spatial resolution of 36m
Data-cube size	512*614*224	660*256*220	748*748*19 at best spatial resolution
Radiometric resolution	12 bits in 1994 and 16 bits in 2005	12 bits	0.5% at 20% albedo (dynamic range)

Figure 1.1 demonstrates how the three dominant earth surface materials, i.e. vegetation, soil, and water reflect the Sun irradiance in the VNIR /SWIR range. The reflectance spectrum has the potential of identifying the materials in a scene by matching the scene reflectance spectra to a spectral library of known materials. For example, the spectral reflectance of healthy green vegetation has a noticeable minimum of reflectance in the visible region of the spectrum resulting from the pigments in the green leaves. Alternatively, stressed vegetation can be detected because it has a lower reflectance in the infrared region (Grant, Daughtry et al.

1993; Gitelson, Stark et al. 2002; Sönmez, Emekli et al. 2010; Pleniou and Koutsias 2013).

Another example is the water spectral reflectance that is characterised by a high absorption in near infrared spectral region and beyond. Because of this property, water bodies can be easily detected with remote sensing data. Clearwater has a lower reflectance in the visible region than turbid water and waters containing high chlorophyll concentrations. These reflectance properties are utilised to detect algae colonies as well as pollutants such as industrial wastewater (Hirtle and Rencz 2003; Zhang, He et al. 2017).
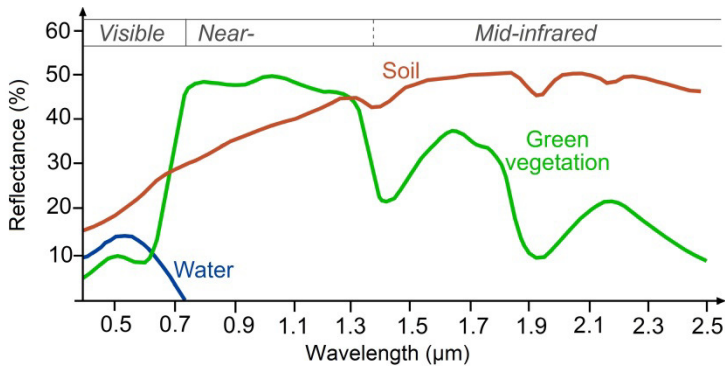
Figure 1.1. Reflectance of vegetation, soil, and water in different wavelengths.

1.3. Applications of hyperspectral imaging

As noted, the objective in hyperspectral imaging is the identification of materials and objects existing on the surface or in the atmosphere of the Earth by using spectral information acquired as images. We can apply hyperspectral imaging sensors to a variety of distinctly different problems, adapting the processing used to extract relevant information such as for anomaly detection and target recognition (Shaw and Burke 2003). These applications comprise environmental monitoring including atmosphere, ecology (Cochrane 2000; Wang, Menenti et al. 2007; Pontius, Martin et al. 2008; Ghiyamat and Shafri 2010), geology (Cloutis 1996; Crouvi, Ben-Dor et al. 2006), and hydrology (Schmid, Koch et al. 2005; Weng, Gong et al. 2008); sensors calibration (Guanter, Richter et al. 2006), modeling and algorithms such as autonomous atmospheric correction(Barducci, Guzzi et al. 2012).

1.4. N-dimensional space

Hyperspectral imagery is typically acquired (and represented) as a cube of spatial and spectral data where the spatial information is represented in the X-Y plane, and spectral information in the Z-direction (Figure 1.2, left).

The information for a given pixel corresponds to the reflectance spectrum of that pixel. The *n*-spectral bands sampled by an imaging spectrometer can be considered as the *n*-dimensions of a Cartesian space, called feature space. Each observed target (pixel) can be represented in such a space by a point with coordinates being the values of the observed radiometric magnitude, e.g. reflectance (Figure 1.2, right).
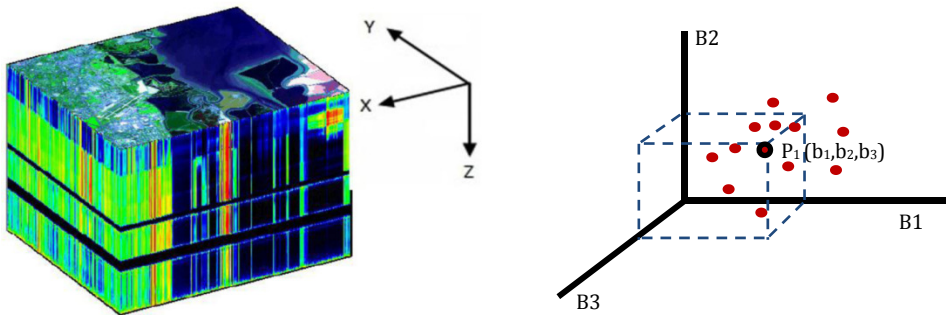
Figure 1.2. A hyperspectral data cube (left), and an example of 3D feature space with three nominal bands (right).

1.4.1. Issues with an n-dimensional feature space

Along the advantages of hyperspectral images, the hyper-dimensionality of the data is a big challenge (Manolakis, Marden et al. 2003; Shaw and Burke 2003; Bioucas-Dias, Plaza et al. 2013). The hyper-dimensionality of the data may cause 1- a large volume of data for storage and transmission; 2- redundant information in the spectral domain; and 3- complexity in data analysis. In this section, we review and explain these three main issues of hyperspectral data.

A large volume of data for storage and transmission

At first glance, the amount of data produced by hyperspectral imaging sensors is very large, and compression is of significance to reduce storage costs and transmission times. The computational time of analysis of such large dataset is also very high (Motta, Rizzo et al. 2006). One pixel of Landsat TM with seven bands, for example, requires seven bytes memory (8 bits per channels) (Barsi, Markham et al. 2007) for storage. A pixel of AVIRIS with 224 bands occupies 64 times more memory, i.e. 448 bytes memory per pixel[1]. So, a complete scene of Landsat TM with

---

[1]All AVIRIS data is stored in 16-bit integers per pixel per channel) (https://directory.eoportal.org/web/eoportal-/airborne-sensors/aviris#foot9%29, https://-aviris.jpl.nasa.gov/html/aviris.datafacility.html#detailed)

seven bands having 30 meters spatial resolution[1] occupies about 242 Mb, while the volume of a dataset collected by AVIRIS covering the same area as a Landsat TM scene exceeds 34843 Mb with 20 m spatial resolution[2].

For the reduction of the volume of hyperspectral data for the storage and transmission, data compression algorithms might be used. These methods are lossless or lossy (Pickering and Ryan 2006). By definition, lossless compression is the reduction of the amount of data without loosing any information (Sayood 1992). The Morse code is probably the earliest known lossless compression technique (Carron 1991). Lossless compression techniques are used in reducing various types of data. In hyperspectral domain for example, (Mielikainen, Toivanen et al. 2003) and (Mielikainen 2006) applied two lossless compression methods to different hyperspectral scenes acquired by AVIRIS and reduced the volume of data by a compression ratio about 3.4 and 3.3 respectively.

Lossy compression can be tolerated when the loss of information is not perceptually significant (Sayood 1992). Garcia-Vilchez, Munoz-Mari et al. (Garcia-Vilchez, Munoz-Mari et al. 2011) quantitatively substantiated the impact of different lossy compression techniques such as JPEG2000 (Taubman and Marcellin 2012) on spectral unmixing and hyperspectral image classification and revealed that different stages of analysing the data exhibit different sensitivities to lossy data compression.

Lossless compression is usually required at the archive level to reduce storage capacity, while in many cases the lossy compression is used to decrease redundant information (Motta, Rizzo et al. 2006). Sayood (Sayood 1992) introduced image classification and clustering as the most efficient form of data compression in remote sensing applications. He mentioned that if an image is assumed to be constructed of a small number of objects, by assigning each pixel in the image to one of the objects the data is highly compressed (Sayood 1992), although the information carried by the class spectral heterogeneity is lost. The dimensionality of a classified image is substantially less than the dimensionality of the original hyperspectral datacube, which can lead to transmitting the data in a short time from a satellite to a ground station if the data is classified or clustered onboard.

For example, Hilbert (Hilbert 1977) proposed the cluster compression algorithm (CCA) to reduce the costs in relation to transmitting, storing, and interpreting of Landsat multispectral images. In (Conticello, Esposito et al. 2016;

---

[1] The approximate scene size of The Landsat Thematic Mapper (TM) is 170 km north-south by 183 km east-west, with the spatial resolution of 30 meters (https://landsat.usgs.gov/what-are-band-designations-landsat-satellites).

[2] The pixel size of the AVIRIS data depend on the altitude from which the data is collected. When collected by the ER-2 (20km above the ground) the spatial resolution is approximately 20 meters on the ground, and when collected by the Twin Otter (4km above the ground), each ground pixel is 4m square (https://aviris.jpl.nasa.gov/html/aviris.concept.html).

Soukup, Gailis et al. 2016), a data processing chain was developed to convert hyperspectral datacubes into a segmented map onboard which can also reduce the transmission time of the data from a miniaturised hyperspectral satellite to ground stations.

Redundant information in spectral domain

The second point is more about the use of the data for different applications. Most of the time, we do not need all the data collected by imaging spectrometers to be analysed for specific applications (Shaw and Burke 2003; Bioucas-Dias, Plaza et al. 2013). For example, if the task is to distinguish a particular target material from the background, a few well-chosen wavelengths are usually sufficient to discriminate between the target and background materials (Manolakis, Marden et al. 2003). These wavelengths can be determined by knowing the interaction between the electromagnetic waves and the given target. For example, the absorption features of soil reflectance spectra are characterised by a few wavelengths at around 1.4, 1.9, and 2.2 µm due to OH^- (Ben-Dor, Goldlshleger et al. 2003), at 2.33 µm to CO_3^{2-} (Gaffey 1987), or at around 0.5-0.9 µm to Fe^{3+} and Fe^{2+} (Hunt, Salisbury et al. 1971).

In reflectance spectra of plant pigments, for instance, an index employing three wavelengths at 0.51, 0.55, and 0.7 µm appears to give a robust solution to estimate total Carotenoids at the leaf scale (Blackburn 2007). Gamon and Surfus (Gamon and Surfus 1999) proposed a ratio of red to green spectral regions for quantifying Anthocyanins concentration in vegetation canopies. Another example is oil dispersed in water using AISA (Airborne Imaging Spectro-radiometer for Applications) hyperspectral images (Salem and Kafatos 2001). The study showed that high absorption at 0.67 µm indicates high oil concentration.

As shown by some examples, different materials present different spectral properties. Therefore for recognising different materials, narrow and contiguous wavebands are essential to finding proper wavebands that help differentiate spectra since critical spectral features may be in various locations of the spectrum for each material (Clark and Roush 1984; Roy 1989; Clark 1999). Capturing all the spectral information in narrow contiguous bands, however, produces redundant and highly correlated data (Shaw and Burke 2003; Chang 2007). The high correlation between adjacent channels of reflectance spectra can be investigated by evaluating the dependency matrices of the data; such as a correlation matrix (Gu and Zhang 2003) or a dependency matrix based on the mutual information between the channels (Martinez-Uso, Pla et al. 2007).

Therefore, efficient data analysis needs specialized algorithms in the extraction of the required information, the reduction of redundant information for a given application, and identification of targets of interest (Schmid, Koch et al. 2005; Lanthier, Bannari et al. 2008; Pontius, Martin et al. 2008). This concept is

summarised in the classification of a hyperspectral scene when the focus is on the identification of objects in a scene (Fauvel, Tarabalka et al. 2013), not just of an individual pixel (Manolakis, Marden et al. 2003). The classification result can be considered as the general outcome of assigning individual pixels of a scene to a small number of classes.

In other applications, such as the retrieval of object properties, the data redundancy of hyperspectral images also creates some difficulties (Lee and Carder 2002; Bajcsy and Groves 2004; Lee, Carder et al. 2007; Zhou, Marani et al. 2017). For example, (Bajcsy and Groves 2004) showed that the redundancy of data could cause convergence problems, and noise in spectral band propagates to the retrieval results. (Zhou, Marani et al. 2017) also revealed that hyperspectral data have limited advantages over multispectral data for suspended sediment concentration retrieval, mainly due to information redundancy and cross-band correlation.

Complexity in data analysis

Frequently used image processing algorithms are usually applied to multispectral images (Landgrebe 2005), i.e. up to ten bands, and when the number of bands or features increases, they encounter problems (Bickel and Levina 2004; Fan and Li 2006; Fan and Fan 2008). For example, a classifier like k-NN, which looks for k nearest neighbouring pixels of a target pixel in a feature space, becomes a prohibitive method when increasing the dimensionality of the searching space (Marimont and Shapiro 1979; Parsons, Haque et al. 2004; Houle, Kriegel et al. 2010; Chen, Leon et al. 2016). In addition, usually in supervised classification methods, the classification accuracy increases gradually with the increase of the number of bands but decreases when the number of spectral bands becomes very large. This is called Hughes phenomenon, and it is due to the limited number of samples (relative to the number of bands) available in the training sets used in the supervised classification procedures and radiometric resolution (Hughes 1968; Kamandar and Ghassemian 2011). This phenomenon in a broader sense was named "curse of dimensionality" by Bellman (Bellman 1954; Bellman 2013). The term was later used in different domains such as numerical analysis (Zimek, Schubert et al. 2012), machine learning (Keogh and Mueen 2010), and data mining (Bernecker, Houle et al. 2011).

Curse of dimensionality describes the problem caused by the exponential increase in volume associated with adding extra dimensions to Euclidean space (Bellman 2013). When the dimensionality rises, the data in feature space become sparse; and to obtain a reliable and statistically sound result, the number of samples needed to support the result increases exponentially with the dimensionality (Donoho 2000; Keshava and Mustard 2002). Jimenez and Landgrebe (Jimenez and Landgrebe 1998) illustrated the specific multivariate data behaviour with several examples that as the dimensionality increases the data will

concentrate in the outer boundary of the volume of the data in feature space, and neighbourhoods of the samples are empty. In one of their experiments, they randomly generated normally distributed data with N(0,1) for different numbers of dimensions and calculated the distance between the random vectors. They intelligibly revealed that when the dimensionality increases the distance from the zero coordinate increases as well.

In three other experiments, they computed the fraction of the volume of 1-a hypersphere inscribed in a hypercube, 2- a hypersphere of radius r-ε inscribed inside a hypersphere of radius r, and 3- two hyperellipsoids in the same situation as no. 2. They proved that by increasing the dimension, the volume explored by a hypercube, hypersphere, and hyper-ellipsoid in three cases concentrated in the outer boundary of the explored space. They concluded that high dimensional space is mostly empty, and if the high dimensional data is projected to a lower dimensional subspace, the data will not lose significant information regarding separability of the different classes. Moreover, in high-dimensional spaces, generally, the data distribution tends to concentrate in the tails, which seems to be contradictory with the assumption of having data with bell-shaped density functions of widely used algorithms (Jimenez and Landgrebe 1998; Landgrebe 2002; Parsons, Haque et al. 2004; Zimek, Schubert et al. 2012).

The Hughes phenomenon is more problematic for parametric methods which assume the sample data follow a probability distribution based on a fixed set of parameters (Alonso, Malpica et al. 2011). The relationship between the dimensionality of the data and the required size of the training set is extensively studied in the pattern recognition literature (Kanal and Chandrasekaran 1971; Foley 1972; Chandrasekaran and Jain 1974; Kulkarni 1978; Jain, Duin et al. 2000; McLachlan 2004). In general, the larger the sample size, the more confident an investigator can be that the answers truly express the reality, i.e. they are statistically significant[1].

It is apparently essential to have as many training pixels per class in a scene as possible, especially when the dimensionality of the feature space increases, since in higher dimensionality spaces there is an increased chance of having some individual dimensions poorly sampled (Richards and Jia 2006). Swain and Davis (Swain and Davis 1978) recommended that at least $10n$ samples per class are needed for training of supervised image processing algorithms, where n is the dimensionality of the feature space, and $100n$, is highly desirable if it can be attained.

---

[1] In statistical hypothesis testing, when an observation, that involves drawing sample from a population, is statistical significant it means that an investigator may conclude that the samples reflect the characteristics of the whole population (Sirkin, R. M. (2005). Statistics for the Social Sciences, SAGE Publications.).

1.4.2. Dimensionality reduction

The primary solution for the issues mentioned is to reduce the dimensionality while retaining the information required for specific applications. An alternative is a lossless compression when the objective is just to reduce the storage of a vast volume of hyperspectral images (Motta, Rizzo et al. 2006). In the case of lossless compression, there is still a need for DR again for analysing the images. So, dimensionality reduction (DR) can be the principal remedy for the large volume of the data, the redundant information and the "curse of dimensionality". In other words, by applying DR to hyperspectral datasets, the volume of the data would be smaller, the information would be less dependent, and the data in feature space would not be sparse.

In general, DR is the process of reducing the number of variables under consideration and can be categorized in several ways. Below is a short review of the categories most applied to hyperspectral images.

Feature extraction vs feature selection

a) *Feature extraction (FE)* transforms the observed feature space into a new data space with a particular criterion (Kumar, Ghosh et al. 2001; Jimenez-Rodriguez, Arzuaga-Cruz et al. 2007), such as eliminating the linear dependency between the components (new features) in the new feature space, as obtained by Principal Component Analysis (PCA) (Kaewpijit, Le Moigne et al. 2002). The main problem with this technique is the loss of some critical and crucial information that might have been compromised or distorted by the transformation since the original spectral measurements are no longer mapped in the new data space.

b) *Feature selection (FS)* approaches try to find a subset of the original variables. This method is also called band/channel selection in the hyperspectral community. Compared with FE, this technique has the advantage of preserving the original information that can be used to understand how spectral features relate to the properties of observed materials (Chein and Su 2006; Martinez-Uso, Pla et al. 2007; Carmona, Mart et al. 2011; Jia, Ji et al. 2012; Jia, Zhu et al. 2014).

Supervised vs unsupervised DR

a) *Supervised DR:* Having labelled information, i.e. prior knowledge, about land-cover in several areas of a scene paves the way for supervised DR, i.e. by selecting image samples for each class, a supervised DR algorithm provides a class-specific feature set which improves the accuracy of classification. An example is the selection of the bands to achieve the

maximum discrimination between the given classes in the feature space (Huang and He 2005; Yang, Du et al. 2011).

b) *Unsupervised D*R: contrary to supervised algorithms, unsupervised DR techniques do not require any prior information and can be applied when information on the observed land cover is scarce (Du and Yang 2008; Cariou, Chehdi et al. 2011; Jia, Ji et al. 2012).

1.5. Scope and objectives

This study was inspired by the design of a nanosatellite to be equipped with a hyperspectral imager (Esposito and Marchi 2015). Recent advances in remote hyperspectral sensor and electronics miniaturisation technology allow utilising such sensors in a platform as small as 3-unit CubeSat and even capable of real-time processing (Conticello, Esposito et al. 2016; Soukup, Gailis et al. 2016). The design of a nanosatellite equipped with a hyperspectral imager is constrained by the insufficient downlink and storage capacity due to the platform power and size. A reduction of the hyperspectral data volume by an onboard data processing chain can be a solution for the limitation of the nano-satellite (Soukup, Gailis et al. 2016).

This study is focused on dimensionality reduction of the data. The critical point in the application of DR techniques to hyperspectral datasets is to identify a criterion to select/extract the proper features. Different criteria can be used for the definition of the optimal spectral configuration for a given application and a given image. Investigating various criteria, we distinguish three overall criteria relevant to the analysis of hyperspectral images, which are explained in the next section. Here, we highlight two issues as the rationale of the study:

a) As mentioned, the FS techniques preserve the connections between the spectral features and the physical properties of the observed materials, making a correct interpretation of the data more accessible. Hence, the feature selection methods are preferable in the hyperspectral domain (Chein and Su 2006; Martinez-Uso, Pla et al. 2007; Sotoca, Pla et al. 2007). In our study, we proposed a new DR algorithm for hyperspectral images that can be an alternative to FS.

b) In this approach, we construct parsimonious spectral configurations by applying the DR-criteria investigated to hyperspectral images. The proposed strategy determines a new spectral configuration instead of selecting individual spectral channels from the original dataset. It is subdividing the spectrum in a number of intervals of different widths. The intervals are fewer and usually wider than the original channels. The value of each interval per pixel is the mean value of its constitutive channels at

that pixel. Through the following chapters, we present more detail of the methodology and demonstrate the advantages of determining such spectral configurations over channel selection. In this strategy, the final spectral configuration consists of contiguous spectral regions covering the entire spectrum. Furthermore, the final band set comprises narrow and broad bands with respect to a predefined criterion.

1.5.1. The criteria for DR

a) *Representation accuracy:* A spectral configuration may provide an accurate representation of the spectrum with fewer spectral bands than the number of channels at full spectral resolution. The optimal configuration is the one that yields the smallest error with a given number of bands or the one that needs fewest bands for a given maximum error (Price 1975; Angelopoulou 2000; Angelopoulou, Molana et al. 2001; Huynh and Robles-Kelly 2008). When expanding the values of the reduced spectral configuration for each pixel back to the original channel configuration (by duplicating the values of each band over the merged channels) and computing the difference between the expanded and the original spectra, summed up over all channels and all pixels, an error measure is obtained which drives the search for the required spectral configuration.

b) *Independent spectral regions*: since adjacent channels in a hyperspectral dataset are often highly dependent, we encounter redundant information when working with such a dataset. In the process of DR of hyperspectral datasets, decreasing redundant information is one of the main objectives (Peng, Long et al. 2005; Martinez-Uso, Pla et al. 2007; Sotoca, Pla et al. 2007; Kamandar and Ghassemian 2011). By choosing a band set with low dependence, the number of features is reduced, while the redundant information would be decreased. So, the optimal configuration would then be the one with the lowest dependence between spectral features.

c) *Class separability:* Having "ground truth" available, i.e. known class labels at a sufficient number of pixels, various methods to compute class separability can be applied. The Jeffreys-Matusita distance is well-known (Richards and Jia 2006) as a class separability metric, but alternatively, one can apply a classifier, determine the error matrix (after separating the ground truth into training and evaluation sets) and compute the overall accuracy or Kappa-statistic (Zhang, Zhang et al. 2013). Again, the maximum value of class discrimination identifies the optimal configuration with a given number of bands (Chang, Du et al. 1999;

Kumar, Ghosh et al. 2001; Ifarraguerri and Prairie 2004; Skurichina, Paclik et al. 2004; Gu, Li et al. 2012; Xiang, Nie et al. 2012).

This last type of band selection methods using the labelled class information is called supervised band selection (Jensen and Solberg 2007; Yang, Du et al. 2011; Feng, Jiao et al. 2014). The metrics of representation and information content can also be applied to an entire hyperspectral scene without any prior knowledge (unsupervised band selection) (Martinez-Uso, Pla et al. 2007; Sotoca, Pla et al. 2007; Bioucas-Dias and Nascimento 2008). The class separability criterion gives in principle results different from the two other criteria, as it depends on an additional data source, the ground truth. We expect different band configurations between the first and the second criterion too. The comparison of the configurations obtained with each criterion is dealt with in Chapter 5.

1.5.2. Research questions

The core of this thesis is to identify the spectral regions covered by most of the hyperspectral sensors with respect to the three aforementioned criteria of dimensionality reduction. The research question addressed by this thesis is to determine:

a) *What is the advantage of determining spectral configuration versus individual channel selection per objective?*
b) *How similar or dissimilar are the spectral configuration obtained based on different DR criteria? How large and significant are the differences of spectral configurations in the cases of using the entire scene (unsupervised) or using specific classes in the scenes (supervised)? Is it necessary to determine different spectral configurations?*
c) *What are the performances of different band configurations in distinguishing various classes in a scene and what is the accuracy of image classification?*
d) *Finally, which band configuration would be suitable for miniaturised hyperspectral satellites?*

To answer the first question and to determine the spectral configurations, the three objectives are investigated in detail in three separate chapters (Chapter 2, 3, and 4). We identify the spectral configurations through top-down or bottom-up search strategy, applying various metrics per objective, and compare them with several well-known feature selection algorithms.

According to the algorithms proposed and the band configurations obtained in the three chapters, we present a comparative evaluation of the spectral configurations determined to meet each objective in Chapter 5. This assessment is

on the basis of three different comparisons: 1- the positions of the splits dividing the full spectrum into a set of spectral bands, 2- the computational load to apply different algorithms, and 3- the accuracy of scene classification using the different spectral configurations. In all the comparisons, the performance of the evaluation is presented versus a different number of spectral features obtained to reveal the trend of variations.

1.5.3. Practical consideration

In order to either determine a new spectral configuration by the proposed methods or select channels by feature selection techniques, the searching algorithm has an impact. The searching algorithm is a method to find m items with specified properties among n collections of items ($m \leq n$), where in our approach m is the number of bands of the final spectral configuration and n is the number of channels of the original hyperspectral dataset. To this end, two strategies are used in different experiments presented in this study:

a) *Exhaustive search:* This method, also called brute-force search, is the simplest one but it is computationally heavy (Nievergelt 2000). It considers and evaluates all m-combinations of bands out of n possible situations based on a criterion. This method guarantees to find the best subset of variables, in a practical situation, however, the computational cost for large datasets like hyperspectral images is prohibitive. For example, if we consider selecting 20 bands out of a 200-channel hyperspectral dataset, the total number of combinations is $1.61*10^{27}$. We may assume that our computer can evaluate 10000 combination per second, which is reasonable for a very simple criterion. The time required to evaluate all the combinations is $5.12*10^{15}$ years.

b) *Heuristic search algorithms:* Heuristic algorithms are ones designed to solve a problem in a faster and more efficient fashion than exhustive search by sacrificing optimality or accuracy for speed. These algorithms solve a problem heuristically, i.e. they give nearly the right solution or provide a solution not for all instances of the problem (Kokash 2005). For example, a greedy search algorithm is a heuristic algorithm which does not produce an optimal solution, makes a locally optimal choice at each iteration with the hope of finding a global optimum in a reasonable time (Cormen, Leiserson et al. 2001; Bendall and Margot 2006).

The proposed algorithms, in this thesis, suggest a greedy search strategy which are evaluated by comparison either with the best solution provided by simulating a

dataset with the lower number of channels than the origenal dataset or with the existing and well-known algorithms applied with the same objectives.

1.6. Outline

This thesis is organised as follows:

Chapter 2 provides band configurations according to the spectral representation criterion. It addresses the DR issues when the representation of reflectance spectra is investigated and reviews the methods and metrics used. A newly developed algorithm is introduced, and the performance is evaluated with comparisons with an algorithm providing the best spectral representation. Different metrics mostly used for spectral representations are also studied in this chapter.

Chapter 3 focuses on the second objective, i.e. independent spectral regions. This chapter reviews the studies on reducing the dependence in the hyperspectral domain. A new metric to compute the overall dependence is introduced and its properties are investigated. More detailed, mathematical and numerical evidence is provided on the metric and the properties are presented in Appendix B. Then the metric is applied to a new bottom-up search algorithm to identify independent spectral regions. An assessment with comparisons with other existing algorithms is presented.

Chapter 4 investigates the importance of having broad spectral regions over narrow channels in class separability (third objectives). A review of frequently used search algorithms in the hyperspectral domain and the class separability metrics is presented. The advantage of broad spectral regions are presented mathematically, first, and later demonstrated experimentally by several implementations.

Chapter 5 combines and compares the results of three main chapters in terms of the location of the spectral regions, time consumption and the classification accuracy to evaluate different spectral configurations. Chapter 6 concludes the dissertation by presenting the most relevant products and findings of the research, followed by recommendations for further investigation.

2.
Representation of reflectance spectra

Hyperspectral imaging acquires the detailed spectral of reflectance and emittance of land surfaces, which can be used to describe, understand and model surface processes and ecosystem changes. However, efficiently analyzing huge hyperspectral datasets is very challenging. This chapter introduces an algorithm to reduce the dimensionality of hyperspectral images. In this study, we focused on the representation of the reflectance spectra using a reduced number of bands. The algorithm, spectral region splitting (SRS), merges adjacent narrow channels and constructs wider spectral bands whenever it is necessary. We compared the algorithm with other existing algorithms with respect to the error of representation and the running time over single spectra and an entire hyperspectral scene. We applied the algorithm using different spectral matching metrics to a complete spectral library for a material detection purpose and revealed how many spectral bands are needed to correctly detect the material signatures.

2.1. Introduction

[1]Hyperspectral imagers, also termed imaging spectrometers, capture reflected radiance from a scene in a regular, rectangular grid of pixels, where every pixel contains detailed spectral information represented by using a contiguous set of spectral channels, corresponding to narrow wavelength intervals. A well-known hyperspectral imager is the NASA AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) instrument (Green, Eastwood et al. 1998). When mounted on a high-altitude aircraft, it covers areas 11 km wide (at 20 km flying height) and up to 800 km long with 20m spatial resolution, with a spectral range of 400-2500 nanometers, subdivided into 224 narrow contiguous channels, each about 10 nm wide.

We propose in this study a method to reduce the dimensionality of the feature space, named the Spectral Region Splitting (SRS) method. Depending on the image, it yields a set of contiguous spectral bands that is much smaller than the original number of spectral channels. Each band covers a number of adjacent channels of the original image, and each channel is used in exactly one band. At every pixel, the signal value in a band is the average of the values of the adjacent channels included in each band. In this study, we distinguish between the spectral channels and the spectral bands, i.e. "channels" refer to the narrow width spectral features in the original hyperspectral dataset, and "bands" refer to the spectral features having different widths in the final dataset. Expected advantages of band formation include the reduction of random noise (due to averaging) compared to

---

[1] Parts of this chapter have been published in Hosseini Aria, S. E., M. Menenti, et al. (2017). "Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data." Journal of Applied Remote Sensing **11**(4): 046010.

narrow channel, as well as the fact that no data gets just thrown away (like the not-selected channels in a channel selection).

Different criteria can be used for the definition of the optimal spectral configuration for a given image, under constraints such as the (maximum) number of bands or the minimum (or maximum) bandwidth. We distinguish the three criteria mentioned in Chapter 1. In this chapter, we focus on the representation accuracy as the criterion to obtain reduced spectral configurations. Our objective is to represent a spectral signal using a spectral configuration with fewer spectral bands, while minimizing the loss in accuracy. The spectral signal sampled by an imaging spectrometer may be represented by using fewer and / or different spectral bands at the required accuracy (Price 1994; Jensen and Solberg 2007). When expanding the reduced spectral configuration at each pixel back into the original channel configuration, the values of each band are duplicated in the adjacent channels covered by the band, then the reconstructed and full spectra are compared. The difference between the expanded and the original spectra can be computed and averaged over all channels and all pixels to obtain an error measure.

Different error metrics may be applied instead of the absolute error or the square root of the mean squared error (RMSE). Spectral Angle (SA) (Zhang, Zhang et al. 2015), Spectral Information Divergence (SID) based on the mean divergence (Chang 2000), and Goodness-Fitting Coefficient (GFC) (Romero, Garcia-Beltran et al. 1997; Hernandez-Andres, Romero et al. 2001) have been used as metrics of the representation of spectral signals.

The accuracy of the representation of spectral data is relevant to applications based on the direct comparison of observed against reference spectra, such as recognition of objects and materials (Manolakis, Marden et al. 2003; Kolodner 2007) or the interpretation of absorption features (Cochrane 2000; Gitelson, Zur et al. 2002). Our method constructs a spectral configuration which can represent the data accurately for such applications but using fewer spectral bands.

To evaluate our proposed algorithm, we have compared it with two existing algorithms with respect to the accuracy of spectral representation and the running time. The algorithms used in the comparison represent a spectrum with a piecewise constant function as well, i.e. they also average adjacent channels to form wider spectral bands. Therefore, the band configurations, i.e. the spectral locations of breakpoints (or splits) over the spectrum identified by different algorithms, exert the major influence on the spectral representations.

The first algorithm is the Adaptive Piecewise Constant Approximation (APCA) method (Chakrabarti, Keogh et al. 2002), a well-known signal approximation method used in time series analysis, and the second one is the Piecewise Constant Function Approximations (PCFA) algorithm (Jensen and Solberg 2007) that finds a globally optimal representation. Jensen and Solberg (Jensen and Solberg 2007) proposed the PCFA algorithm and applied it to the hyperspectral signatures of a scene and found one spectral configuration suitable to represent the spectra of all

the classes in the scene. They partitioned the full spectra into a number of contiguous intervals, where the spectral locations of the breakpoints were the same for all pixels. We applied the method to identify the best spectral configuration separately for each spectrum as well. In this case, different spectra would be represented using different spectral configurations and the mean error of representation would be smaller.

In the second experiment, the spectral representations obtained by SRS using various spectral matching metrics were applied for material detection, and the accuracy of the detection was computed and compared with the PCFA results. The second experiment illustrates how the reduced spectral configurations can capture the signatures of vegetation and minerals at the accuracy required for identification, but using fewer spectral bands. In the third experiment, the methods were applied to all the spectra of an entire hyperspectral scene to obtain the same spectral configuration for all pixels. The error of representation and the running time of the algorithms were compared.

The chapter is organized as follows. Section 2.2 articulates the different aspects of the approximation of detailed reflectance spectra. Section 2.3 gives a full description of the SRS algorithm. The characteristics of the applied datasets are presented in Section 2.4 Section 2.5 presents the evaluation of the proposed algorithms with three different experiments and their results. Section 2.6 summarizes the study.

2.2. Review of methods

Two issues have to be addressed when the representation of reflectance spectra is investigated: 1- the algorithm applied to determine the spectral configuration and 2- the metric applied to compare the original with the new spectral configuration. In the following paragraphs, we review known algorithms and metrics, and then, we explain the need for the new proposed algorithm.

2.2.1. Algorithms

There are several methods for spectral approximation and representation (Price 1975; Price 1990; Li, Becker et al. 1999; Angelopoulou 2000; Angelopoulou, Molana et al. 2001; Wang, Menenti et al. 2007; Huynh and Robles-Kelly 2008), mostly used in colorimetric sciences. Their scope is to describe spectra accurately with a limited number of bands. These methods usually use a set of basis functions to represent the spectra. For example, Price (Price 1975) developed a method to represent high-resolution spectra with a limited number of independent variables or Mansouri et.al. (Mansouri, Sliwa et al. 2008) proposed an adaptive principal component algorithm to estimate reflectance spectra. Both methods require basis

functions to reconstruct the spectrum, and the reconstruction is an approximation of the original reflectance spectrum.

There are other methods to represent spectra directly by performing dimensionality reduction on the data, e.g. Discrete Fourier Transform (DFT) (Agrawal, Faloutsos et al. 1993), Singular Value Decomposition (SVD) (Keogh, Chakrabarti et al. 2001) or Discrete Wavelet Transform (DWT) (Kahveci and Singh 2001). A common way for direct approximation is the replacement of local variations in a spectrum with a constant value over a small range in wavelength. Chakrabarti et.al. (Chakrabarti, Keogh et al. 2002) presented such a technique, named Adaptive Piecewise Constant Approximation (APCA), and proved that this technique yields a better representation than other existing methods such as DFT and DWT for approximating signals in time series analyses. The APCA algorithm degrades a curve into a constant segment-based approximation, where the user specifies the number of segments. It includes two main steps. At first, it converts the signal approximation issue into a wavelet compression problem, for which there are well-known optimal solutions, and next, it converts the solution back to the APCA representation and makes minor modifications. The term 'segment' is equivalent to 'band' in our approach. More details on APCA can be found in (Chakrabarti, Keogh et al. 2002). Approximating spectra by piecewise constant functions has also been used in other fields, taking into account the physical characteristics of the spectra to determine the location of the spectral segments (Zehentbauer and Kiefer 2012; Thomson, Lue et al. 2014).

Konno and Kuno (Konno and Kuno 1988) proposed a method that provides the best piecewise constant approximation. They used the maximum norm and the Euclidean norm to find the approximation of a function of a single variable with less than a predefined number constant-value segments. Another study applied the Bayesian approach for piecewise smoothing of one-dimensional signals (Winkler and Liebscher 2002). Later, an extension of this method was used for multiple spectral curves to reduce the dimensionality of hyperspectral scenes (Jensen and Solberg 2007). The goal of the last approach was also to partition the spectra of a hyperspectral scene into a fixed number of continuous intervals using the piecewise constant function approximations (PCFA) algorithm. The intensity in a spectral band is the mean value of the signal in its constitutive channels per pixel. Considering the number of bands, the algorithm examines all the possible spectral locations for the breakpoints and finds the best approximation having the lowest error of representation. The authors applied it to a set of sampled spectra of a hyperspectral scene and identified a single spectral configuration for all the spectra to reduce the dimensionality of the data.

The point here is the possibility of finding the best spectral representation of a signal, which may seem the contrary of obtaining the best selection of m out n ($m \leq n$) spectral channels of hyperspectral data (Section 1.4.2). However, if the criterion used for the selection has the vector addition property of linear functions,

i.e. $c(f_1) + c(f_2) = c(f_1 + f_2)$, where $c(f_1)$ is the criterion function of feature f_1, finding the best solution is doable, since there is no need to evaluate all the situations. (Jensen and Solberg 2007) used the sum of squared error (SSE) that met the mentioned property and provided the best spectral representation for a given numebr of spectral bands.

Cariou et.al. (Cariou, Chehdi et al. 2011) proposed an unsupervised band clustering method for DR of hyperspectral images with the criterion of minimizing the mutual information between two adjacent band-clusters. The algorithm, named BandClust, follows a recursive binary search to find the spectral boundaries of band-clusters, and all the previously created sub-bands are split unless no lower value of the criterion can be found across any pair of consecutive band boundaries. Finally, the signal in each cluster is represented by the average over its constitutive channels.

2.2.2. Metrics

The algorithms reviewed above yield representations of reflectance spectra which deviate from the original spectra at full spectral resolution. Different metrics can be applied to measure the difference between the original and the approximated spectra. For example, some metrics give more weight to lower reflectance values, or vice versa. Giving weights to specific spectral ranges leads to a higher discrimination capability of particular materials. For instance, the visible spectral region of the vegetation reflectance is frequently used to retrieve plant properties such as chlorophyll or carotenoid content (Cochrane 2000; Gitelson, Zur et al. 2002; Bell and Baranoski 2004). (Imai, Rosen et al. 2002) carried out a comparative study of eight metrics to evaluate spectral matching quality and to document advantages and disadvantages of each metric. Some metrics are well-known and are widely used, e.g. the Root-Mean-Square-Error (RMSE) and Spectral Angle (SA) in hyperspectral studies (Price 1994; Chang 2000; Cochrane 2000; Du, Chang et al. 2004; Deborah, Richard et al. 2015). Other metrics performed differently when applied to identify materials using reflectance spectra. For example, by using the Spectral Information Divergence (SID) (Du, Chang et al. 2004), many concepts of information theory can be readily applied to evaluate the representation of reflectance spectra.

To illustrate this brief review five spectral matching metrics: RMSE, SA, SID, SID-SA, ECS (Euclidean distance of cumulative spectrum) are explained in detail.

Given n samples of an original (reference) signal $R = \{r_1, r_2, \dots, r_n\}$, and of its approximated one, $A = \{a_1, a_2, \dots, a_n\}$, the metrics are obtained as follows:

RMSE: the absolute error of estimate of the spectral signal (radiance or reflectance) is calculated as:

$$RMSE(R, A) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (r_i - a_i)^2} \qquad (2.1)$$

If the term $\frac{1}{n}$ is omitted, the metric is the Euclidean distance used by (Chakrabarti, Keogh et al. 2002).

SA: is normalized to the magnitude of the signal so that the metric is insensitive to the magnitude. It determines the similarity between the reference and the approximated spectra by calculating the angle between the two vectors in the feature space:

$$SA(R, A) = cos^{-1} \frac{\sum_{i=1}^{n} r_i \cdot a_i}{\sqrt{\sum_{i=1}^{n} r_i^2} \cdot \sqrt{\sum_{i=1}^{n} a_i^2}} \text{ (in radians)} \qquad (2.2)$$

In the remote sensing community, these two metrics have been widely used as spectral similarity metrics for material detection.

SID: this metric has been derived from the spectral information measure (SIM) which considers the spectral variability as a result of signal randomness (Chang 2000). The SIM can be assumed as a single-pixel-based information measure. Spectral information divergence or SID uses the same concept to measure the difference between the two spectra, R and A applying their probability mass functions $P = \{p_1, p_2, ..., p_n\}$ and $Q = \{q_1, q_2, ..., q_n\}$, where $p_i = \frac{r_i}{\sum_{r \in R} r}$ and $q_i = \frac{a_i}{\sum_{a \in A} a}$. SID is calculated as

$$SID(R, A) = \sum_{i=1}^{n} [p_i \log(p_i/q_i) + q_i \log(q_i/p_i)] \qquad (2.3)$$

SID-SA: Du et al. (Du, Chang et al. 2004) combined SID and SA to form a new mixed metric as $SID(R, A) * \tan(SA(R, A))$ and $SID(R, A) * \sin(SA(R, A))$ and showed that the new metric has the strengths of both SID and SA metrics.

ECS: The Euclidean distance of cumulative spectrum is computed as

$$ECS(R, A) = \left(\sum_{\lambda} \left| \int r_\lambda d\lambda - \int a_\lambda \, d\lambda \right|^2 \right)^{1/2} \qquad (2.4)$$

where λ is the wavelength and $\lambda \subset [1, n]$. Deborah et al. (Deborah, Richard et al. 2015) using simulated reflectance spectra demonstrated that ECS, in comparison with more than ten other measures, has a better performance in the cases of magnitude changes in hyperspectral data across image pixels.

2.2.3. Required improvements

One issue about the approximation methods mentioned above is that they have mostly been developed applying the l_2 norm of the representation error, such as RMSE, and they cannot easily utilize other metrics, like SID, used in the hyperspectral domain. Furthermore, they are mainly focused on finding the approximation of a single spectrum, not an entire scene. The method proposed by Jensen and Solberg (Jensen and Solberg 2007), although it provides the optimal representation of all spectra in a scene, is slow. The complexity of the algorithm is $O\left(kN_pN_c^3\right)$; where k, N_p, and N_c are the number of bands in the final bandset, the number of pixels, and the number of channels in the original dataset, respectively(Jensen and Solberg 2007). The BandClust (Cariou, Chehdi et al. 2011) algorithm is a suboptimal method, since each previously created band cluster (comprising more than one spectral channel) is divided by the algorithm again.

The review above suggests the need for an algorithm that easily accepts different metrics to approximate the spectrum with an adequate accuracy, taking into account different specific objectives which may apply to the further use of the approximated spectral signal. For this purpose, we introduce the spectral region splitting (SRS) method to determine a reduced spectral configuration to represent spectral signals.

2.3. Method

The proposed algorithm constructs a reduced spectral configuration, consisting in a set of contiguous bands with variable spectral width, to minimize the difference between the approximated and original (reference) spectrum. The reduced spectral configuration determined by the algorithm consists of wider, contiguous, spectral bands with the signal value being the average over all spectral channels merged into each band. The latter filters out random noise and increases the signal to noise ratio (SNR). The reduced spectral configuration, i.e. a set of contiguous bands, represents the full spectrum without using a spectral interpolation function. SRS can be applied using different performance metrics to

address various applications. In addition, this method preserves the physical meaning of the original data, unlike feature extraction techniques. In feature extraction (Richards and Jia 2006; Fauvel, Chanussot et al. 2009), the features in the low-dimensional space are combinations of individual original channels which are obtained by a transformation. The transformation compromises and distorts the data, and some critical information may be lost (Chein and Su 2006; Jensen and Solberg 2007; Feng, Jiao et al. 2014; Sun, Geng et al. 2014; Tan, Li et al. 2014).

2.3.1. Spectral Region Splitting

In the Spectral Region Splitting (SRS) method, the spectrum is split into spectral bands of different spectral width. This method is an iterative top-down algorithm, and therefore, a termination point must be chosen to stop the iterations.

Let R be an original hyperspectral image with N_c channels and N_p pixels: $R = \{R_{ij} | 1 \leq i \leq N_c, 1 \leq j \leq N_p\}$, and R_{ij} is the ith signal value in the jth pixel. Assume that a reduced spectral configuration (A) with k bands, $k \in \mathbb{Z}^+$ and $k \leq N_c$, is $A = \{A_{tj} | 1 \leq t \leq k, 1 \leq j \leq N_p\}$, and A_{tj} is the signal value in the tth band in the jth pixel. The number of splits is k-1, with $N_c - 1$ possible spectral locations. The set $S = \{s_1, s_2, \dots, s_{k-1}\}$ gives the split locations sorted with respect to their wavelength, so $s_1 < s_2 < \dots < s_{k-1}$. Determining the splits does not necessarily occur in the same order, e.g. the first split found can be any $s \in S$. By defining s_0 and s_k as the beginning and the end of the spectrum in S, then $S = \{s_0, s_1, \dots, s_{k-1}, s_k\}$, and A is defined by:

$$A_{tj} = \begin{cases} \langle R_{ij} \rangle_{i \in [s_{t-1}, s_t)} & if\ t < k \\ \langle R_{ij} \rangle_{i \in [s_{t-1}, s_t]} & if\ t = k \end{cases} \qquad (2.5)$$

where $\langle . \rangle$ is the average of all elements in the subset, and when $t=k$, all the spectral signals in the closed interval between s_{t-1} and s_t are averaged.

If one band is required, there will be no split and the signal per pixel in this band will be the mean value of all spectral channels. If the measurements in the individual channels are noisy due to the narrow spectral bandwidth of the spectrometer, a wider band will have a much better signal to noise ratio than the individual channels. Therefore, an advantage of the broadband over the narrow channels is that a radiance is obtained with less random noise, but with less spectral details. If k > 1, then the algorithm will search for the location of the first split. The location is determined based on the predefined criterion (see Introduction). The critical point is to translate the criterion into a score that gives a

value depending on the split location for the entire spectrum. By scanning all possible locations of the split, the maximum or the minimum score determines the best location of the split in that iteration. The new band configuration is evaluated using Eq. (2.5) for every location of the split, and calculates the score. Finally, the best location of the split at that iteration will be found, then the subsequent locations of splits are determined sequentially, and at each iteration, a new split is placed in the spectrum. After each iteration, the termination point must be checked. In conclusion, the algorithm yields a continuous bandset comprising several narrow and wide bands, identified by the spectral locations of the splits. SRS can be applied to a single spectrum and an entire hyperspectral scene. The SRS procedure is described schematically in Table 2.1.

Table 2.1. Spectral region splitting algorithm (pseudo-code).

Algorithm SRS(R, T)	// R is original dataset, and T is the termination point
begin	
$S \leftarrow \{s_0, s_k\}$	*// The set of split locations*
$P \leftarrow \{p_1, \dots, p_{N_c-1}\}$	*// The set of potential split locations*
Sc=Sc(A(S))	// Computing the score for one broad band covering the entire spectrum
while T is not valid	
for all p in P	
$A(S \cup p)$ is generated (Eq. (2.5)) and the score *Sc=Sc(A(S $\cup$ p))* is computed	
end_for	
Split at the position p_{max} with the highest score (Sc_{max})	
$S \leftarrow S \cup p_{max}$	*// Adding the location of the new split to S*
$P \leftarrow P \backslash p_{max}$	*// Subtracting the new split from P*
end_while	
return S	
end	

Using a spectral matching metric (Section 2.2.2) for spectral representation, the SRS algorithm yields a band configuration minimizing the error of estimate with the reduced spectral configuration. To calculate the error, such as Eq. (2.1), each band of the reduced spectral configuration is expanded into the full set of its constitutive spectral channels.

2.3.2. Computational complexity of the proposed algorithm

The SRS algorithm complexity alone for a single spectrum is of the order $O(kN_c)$, where k is the number of bands, and N_c is the number of channels in the original dataset. As mentioned SRS accepts various criteria. We assumed that the complexity of the metric for spectra representation is of the order $O(N_p N_c)$, where N_p is the number of pixels. Hence, the overall computation time of the SRS algorithm is $O(kN_p N_c^2)$.

2.4. Dataset

The SRS algorithm was evaluated by applying it to the spectra acquired by a hyperspectral imager and to a spectral library. The hyperspectral image was captured over a vegetation/urban/water area in Moffett Field, CA, by the imaging spectrometer AVIRIS in 1997 with 224 channels covering 400-2500nm wavelengths (Figure 2.1). It is made available free-of-charge on the Internet by the Jet Propulsion Laboratory, California Institute of Technology, USA (http://opticks.org/confluence/display-/opticks/Sample+Data). We have done all the necessary pre-processing steps before using the datasets, including atmospheric correction and removal of the noisy channels. The noisy channels are those that do not have any signal, located at water absorption spectral regions, and the ones having low signal to noise ratio (SNR). The channels with low SNR were identified by estimating the SNR using the geostatistical method described in (Curran and Dungan 1989), and visual inspection (Appendix A). The final hyperspectral image data has 188 channels, with the omitted channels located at 365.9-385.3, 1353.1-1432.9, 1811.5-1948.1, and 2437.5-2497.0 nm wavelength.

Figure 2.1. True color images of Moffett Field.

The second dataset used for the evaluation is a spectral library that comprises 1365 spectra from different materials and was developed by researchers at the Spectroscopy Lab, USGS, in 2007. The library is divided into six chapters: 1.Minerals, 2.Mixtures, 3.Coatings, 4.Volatiles, 5.Man-Made; and 6.Plants, Vegetation Communities, Mixtures with Vegetation, and Microorganisms. For completeness of the spectral library, different factors have been considered during

the collection of the spectra such as the type of the spectrometer, the spectral resolution, the purity of the materials, the grain size, the presence of other elements in the sample, etc. There might be more than one spectrum for a material. The chapters contain 881, 138, 12, 24, 110, 200 spectra respectively. The library is used as a reference for material identification in remote sensing images. More details of the spectral library can be found at https://speclab.cr.usgs.gov/spectral-lib.html. The database is over 6000 webpages. We used the convolved version of the library corresponding to the AVIRIS channels. After analyzing the spectral library, we found out that three pairs of the spectra are exactly the same and they cannot be distinguished from each other. These spectral pairs are from Mixtures, and Plants chapters and the duplicated spectra were removed from the chapters mentioned. The final number of spectra in Chapter 2. and 6. became 136, and 199 respectively.

2.5. Results and discussion

Three different experiments were performed to evaluate the SRS algorithm. In the first experiment, the SRS algorithm was compared with two other existing algorithms to assess the error of representation using various spectra. Both existing algorithms represent spectra with a set of constant signal value segments as SRS. The first algorithm, named piecewise constant function approximations (PCFA)(Jensen and Solberg 2007), provides the best spectral configuration minimizing the error of representation for a given spectrum. We applied the method to find the best spectral configuration per spectrum. The second method, adaptive piecewise constant approximation (APCA) (Keogh, Chakrabarti et al. 2001; Chakrabarti, Keogh et al. 2002), outperformed other methods in approximating time series signals. In this experiment, the three algorithms, PCFA, APCA, and SRS were compared using the RMSE metric and the running time. It should be noticed that PCFA was developed to minimise SSE as the approximation error, however, minimising SSE with respect to the approximated spectrum with m predefined number of bands is equivalent to minimising RMSE with the same conditions.

In the second experiment, the reduced spectral configurations obtained by the algorithms were evaluated for material detection, i.e. different materials were detected by comparing the approximated spectra with the full spectra available in the spectral library. In this experiment, we applied five different error metrics to SRS.

The third experiment compares the PCFA and SRS algorithm with respect to the error of representation and the running time to find one spectral configuration for all the spectra in an entire hyperspectral scene.

2.5.1. Evaluation of the algorithm using single spectra

This evaluation was performed as a benchmark to identify the error of representation of single spectra using the three methods. The reflectance spectra of soil, water, and vegetation were compared. First, the algorithms were applied to represent the spectra with 5, 10, and 15 bands. These reflectance spectra were obtained from the pixels with the same land cover in the Moffett Field scene. We applied the RMSE metric to evaluate the reduced spectral configurations. Figure 2.2 illustrates the results, and Table 2.2 gives the error of estimate by different algorithms for all the spectral configurations.

As expected the PCFA algorithm gave the smallest error of estimate in all cases, while APCA gave the largest error (Table 2.2). Interestingly, SRS represented the spectra almost twice as accurately as APCA and with an accuracy comparable with the PCFA algorithm. For example using ten bands to approximate the soil spectrum, the error of estimate was 0.0230 with APCA, 0.0115 with SRS, and 0.0096 with the PCFA algorithm. The difference in the error between SRS and PCFA is less than 0.002, and it became lower when 15 bands were used, with the difference in RMSE being 0.0005 only. It is also observed (Figure 2.2) that the locations of the splits (breakpoints) determined by the PCFA and SRS algorithms are almost identical.

Table 2.2. The approximation error in reflectance units of three dominant types of spectra using different number of spectral bands.

RMSE		APCA	SRS	PCFA
	Water	0.0108	0.0070	0.0068
Maximum 5 bands	Soil	0.0321	0.0214	0.0211
	Vegetation	0.0675	0.0409	0.0352
	Water	0.0062	0.0038	0.0038
Maximum 10 bands	Soil	0.0230	0.0115	0.0096
	Vegetation	0.0424	0.0200	0.0194
	Water	0.0049	0.0031	0.0029
Maximum 15 bands	Soil	0.0128	0.0078	0.0073
	Vegetation	0.0251	0.0139	0.0133

The SRS and PCFA algorithm always divide the spectrum into the predefined number of bands, while APCA does not, as shown in this experiment. This situation occurred, for example, when seeking to approximate the water spectrum, with ten bands by the APCA algorithm. In this case, the reduced spectral configuration had one spectral band less than the prescribed number of bands, while the error of estimate would have been lower with one additional band.
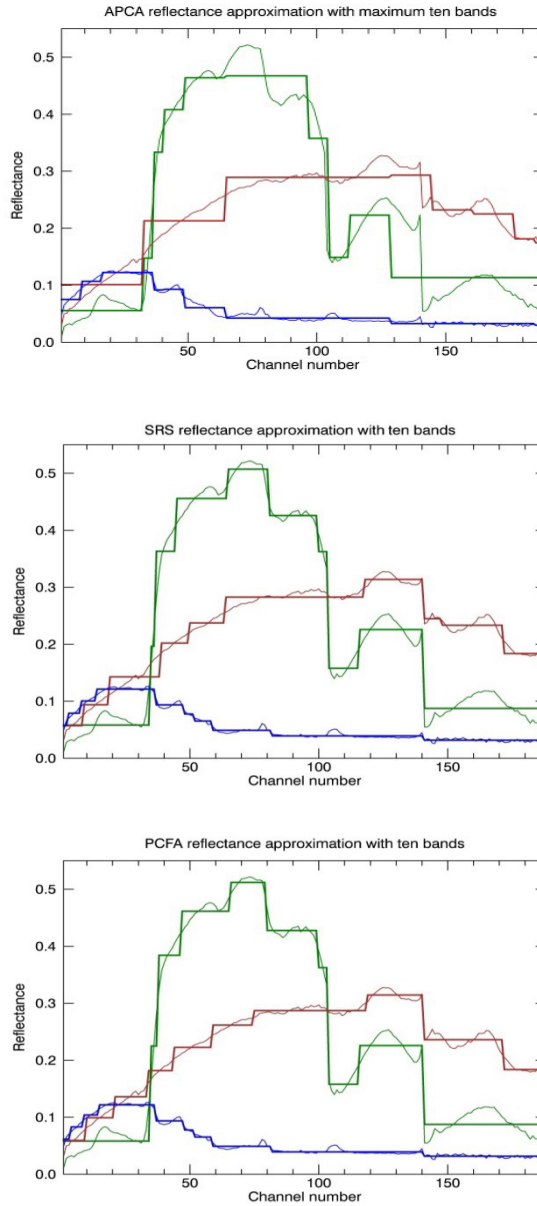
Figure 2.2. Spectral representation of three different land-cover spectra using APCA (upper row), SRS (middle) and PCFA (lower row). The original and the approximate representations of the spectra are the blue, brown and green lines for water, soil and vegetation respectively.

APCA is based on the Haar wavelet transform, so the number of samples in the original signal fed into the algorithm has to be a power of two. In the case that the signal does not have enough samples, it is padded with zeros, and later truncated. This process sometimes may yield fewer bands than expected.

We repeated the same experiment using more than 1000 pixels of the Moffett Field AVIRIS image with different reflectance spectra. The pixels were chosen to sample various land cover types including different types of water, soil, vegetation, man-made features such as buildings, roads, etc. The reflectance spectra were represented separately for each pixel with a different number of bands starting from 5 to 30, in steps of 5. The mean RMSEs decrease with increasing number of bands (Figure 2.3). Similar to the previous results, the APCA RMSE was the largest one, while the RMSE for SRS and PCFA algorithms were very similar and lower than when using APCA. The mean RMSE difference between PCFA and SRS over all spectra samples was about 0.0003 with the 5-band representation to 0.0001 with the 30-band one (the RMSE is in the unit of the spectral reflectance, i.e. in [0,1]). The difference between SRS and PCFA representation errors decreased with increasing number of bands.
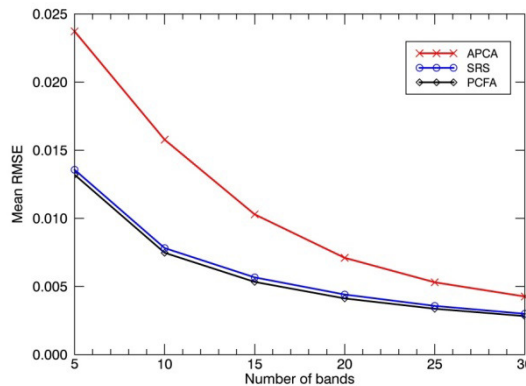
Figure 2.3. The mean approximation error of various spectra obtained by three methods with respect to the maximum number of bands.

We also considered the run time required to carry out the numerical experiment on the dataset with more than 1000 spectra (Table 2.3). The time estimate is based on the implementation of the algorithms on a desktop computer having following characteristics: Operating system: Windows 7, Processor: Intel Core 2, and 16 GB RAM. The algorithms were written in IDL programing language, version 8.2.

The APCA algorithm is fast. The running time was less than two seconds for all the spectral configurations, while it was increasing with the number of bands for SRS: started at less than 14 seconds for the 5-band representations and reached more than two and half minutes for the 30-band ones. PCFA consumed much more time than the two other algorithms. In the worst case, i.e. the 30-band representation, the run time was more than two days to find the representations for the 1089 spectra, which was about 1200 times slower than SRS. On average, SRS ran 660 times faster than PCFA. The main issue affecting the running time of the PCFA algorithm is that the algorithm recursively calls itself with respect to the number of bands, and checks all the possible situations. When the number of bands increases, the run time increases dramatically.

Table 2.3: Run time (sec.) of the three representation algorithms applied to 1089 spectra.

Time (sec.)	APCA	SRS	PCFA
5 bands	1.418	13.318	694.304
10 bands	1.420	34.476	3711.606
15 bands	1.445	59.049	17810.871
20 bands	1.449	85.471	37822.459
25 bands	1.466	124.688	67871.63478
30 bands	1.486	161.561	187344.1304

2.5.2. Material detection using approximated spectral signatures

The previous experiment showed that the SRS algorithm yields comparable spectral configurations as the best representation provided by PCFA with a much shorter run time. In the second experiment, we evaluated the spectral configurations obtained by SRS by applying them to material detection using five spectral matching metrics: RMSE, SA, SID, SID-SA, and ECS. The SID-SA mixed metric was calculated by the multiplication of SID and the tangent of the SA, since using the sine function would produce similar result (Du, Chang et al. 2004).

This experiment reveals how many bands are needed to correctly identify a target spectral signature using the reduced spectral configurations. We used the second dataset, the spectral library on specific materials, that contains different and well–defined spectra. We also applied the PCFA algorithm to material detection, to evaluate the accuracy of detection when using the reduced band configuration determined with SRS. The APCA algorithm was omitted, since the spectral representations provided by this algorithm are not as accurate as the representations obtained by the other algorithms.

The algorithms were applied to every spectrum of the spectral library, i.e. they start with a single wide band and calculate the split points in such a way that the full spectrum is represented by the reduced spectral configuration with increasing accuracy. At each iteration, the reduced spectral signature of the target spectrum was compared with all the spectra in the spectral library to check whether the approximated spectrum could correctly identify a material, i.e. the spectral approximation of a target spectrum has the maximum similarity (or minimum error) when compare with its original spectrum. The iterative procedure was ended when the reduced spectral configuration of the target signature had been correctly identified or the number of bands was more than 30. We used a distance-based identifier and a spectral angle based identifier to measure the difference between the known and unknown spectra (Kruse, Lefkoff et al. 1993; Price 1994; Cochrane 2000). The identifiers use Eq. (2.1) and Eq. (2.2), respectively, to identify the unknown materials with the reduced band configurations. Finally, we calculated the percentage of the spectra correctly identified vs. the number of bands, and called it as the accuracy of detection (Figure 2.4).
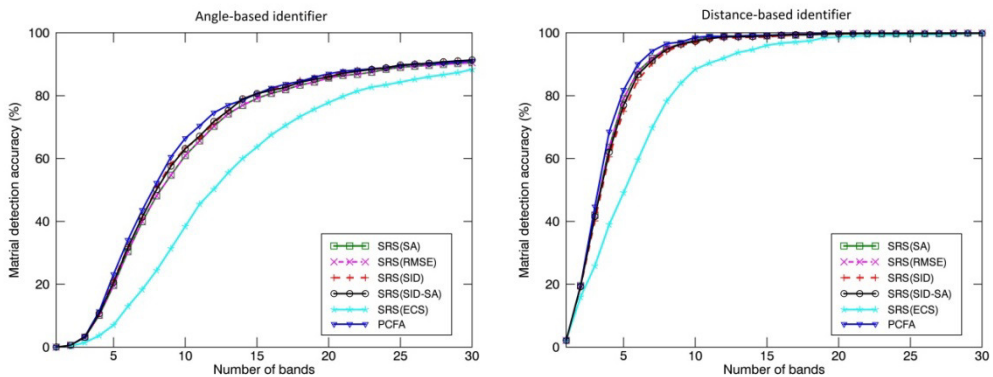
Figure 2.4. The percentage of materials correctly identified using the reduced band configurations obtained by different methods/metrics. In the legend, the metric utilized in SRS per experiment is identified.

In general, the experiment demonstrates that materials can be identified using the reduced spectral configuration with much fewer spectral bands than the detailed full spectra. When the distance-based identifier is used, the accuracy is much higher. For instance out of 224 spectral channels, the 10-, 15-, and 20-band spectral configurations achieved correct identification of materials in 97% , 99%, and 100% of cases with the distance-based identifier and 61%,79 %, 85% with the angle-based identifier. The latter normalizes the spectra, and removes the signal intensity (reflectance in this experiment) dependence.

Using either identifier, the reduced spectral configurations obtained with PCFA and SRS using RMSE, SA, SID, and SID-SA gave a comparable accuracy in material detection. When ECS was used in SRS, the accuracy of detection was lowest. SRS gave similar accuracy of detection when applied with either RMSE or SA. This is due to the fact that when the spectral angle is small, SA and RMSE are essentially equivalent as regards spectral discrimination (Chang 2000; Chang 2003). The spectral configurations obtained with PCFA gave slightly more identifications than the ones obtained with either SRS(RMSE) or SRS(SA). The difference in performance between PCFA and SRS(RMSE) was higher when the number of bands is small, and it decreased with increasing number of bands. However, if the spectral configuration obtained with SRS(RMSE) has just one band more than the PCFA configuration, the detection accuracy for SRS(RMSE) is higher. For example using the angle–based identifier and the 16-band configuration obtained by SRS(RMSE), the identification of materials was correct in 80.7% of the cases, while the 15-band configuration obtained with PCFA was correct in 80.6 % of cases. It should be noted that the computational cost of the 15-band PCFA configuration is much higher than the one of the 16-band configuration obtained with SRS. In the example mentioned, SRS was more than 2000 times faster than PCFA.

When applying SRS with the metrics SID and SID-SA, the accuracy of identification was often higher than when using RMSE or SA, and, interestingly in some cases, it is even higher than the accuracy obtained with PCFA. This occurred mostly when the angle-based identifier was used. As an example, the accuracy of detection with the 14-band configuration was 78.8% with SRS(SID) and 79.0% with SRS(SID-SA), while it was 78.6% with PCFA. The accuracy of identification when applying SRS(SID) and SRS(SID-SA) was higher than with PCFA when the number of bands increased. It is because the SID and SID-SA are more comprehensive spectral discrimination metrics than RMSE and SA and they model a spectrum as a probability distribution (see Eq. (2.3)), so the spectral variabilities across spectral channels can be captured more efficiently in a stochastic manner. Therefore, when SID and SID-SA are used in SRS, the spectral configuration provided have better discriminability. Du et al. (Du, Chang et al. 2004) also illustrated that the discriminatory power of SID and SID-SA is much better than a metric like SA.

As mentioned, the distance-based identifier had a much better performance for material identification than the angle-based one. With a few exceptions, most materials were identified correctly using the reduced band configuration with 11 spectral bands and a 99% accuracy. Table 2.4 shows the number of bands of the reduced spectral configurations required for 99% accurate identification of different categories of materials using SRS(RMSE).

Table 2.4: The number of bands of the reduced spectral configurations obtained with SRS(RMSE) needed for correct material identification with 99% accuracy.

Chapter	No. of bands
Minerals	14
Mixtures	9
Coatings	8
Volatiles	11
Man-Made	7
Plants	11

These results confirm that materials can be identified correctly with much fewer bands than the original narrow spectral channels (e.g. 224 channels). For instance, we selected six vegetation materials and three minerals in the spectral library to illustrate the spectral representations obtained by SRS(RMSE) (Figure 2.5). The figure demonstrates the location and the number of bands required to identify correctly the selected materials, when using the distance-based identifier.
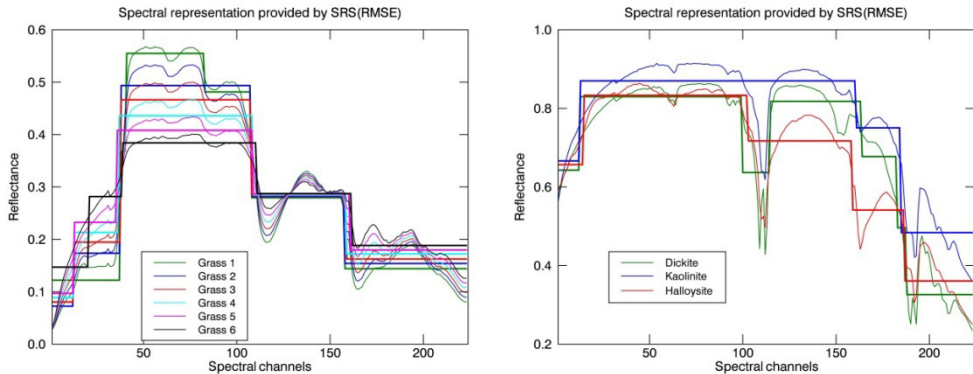
Figure 2.5. Reflectance spectra and the representations of two sets of material signatures: the vegetation set (left) containing six types of grass and the mineral set (right) containing three clay minerals.

The first set contains six reflectance spectra of various types of grass starting from Grass 1 with the highest greenness and the lowest dryness and ending to Grass 6 with the opposite characteristics. So, the greenness and the dryness gradually change in reverse order. The SRS(RMSE) gave a spectral configuration with six bands for Grass 1 and five bands for the remaining grass types. As illustrated, three split locations are almost identical for all the configurations, i.e. at

the start and the end of the near infrared plateau (~710nm and 1320nm) and at 1870nm. There is another split (at 510 nm) in the visible region of the spectrum for the grasses with lower greenness. So, this region is divided into two spectral regions where the split location of Grass 6 with the most depressed greenness is closer to the red spectral region (590nm). For the greenest grass (Grass 1), there is a need for two other splits, besides the first three, to be detected correctly. So for this type of grass, the near infrared plateau was divided into two bands where there is a sudden drop at 1150nm, and the red edge (720-750nm) was also selected as a separate band.

The second data set is minerals spectra. We took three clay minerals having similar and not easily distinguishable spectral signatures (Crowley and Vergo 1988; Frost 1995). These minerals are dickite, kaolinite, and halloysite. As already mentioned, some materials have more than one spectrum. In the cases of selected minerals, there exist 4, 17, 7 spectral reflectances for dickite, kaolinite, and halloysite respectively, which have even closer spectral signatures and would make the detection more difficult. The spectral configuration obtained by SRS(RMSE) demonstrated that the configuration with seven bands for dickite, with four bands for Kaolinite, and with five bands for halloysite was sufficient for correct identification. The splits were at 510, 1292, 1442, 1920, 2110, and 2160nm for dickite; 510, 1890, and 2130nm for kaolinite; and 510, 1322, and 2150nm for halloysite.

2.5.3. Evaluation of the algorithms using entire scene

When the number of pixels is large, like an entire scene of a hyperspectral image, applying the algorithms to the individual pixel is a time-consuming process. Furthermore, the spectral locations of the splits must be the same for all pixels in a hyperspectral scene if the dimensionality reduction of the image is the goal. In this experiment, we extended the PCFA and SRS(RMSE) algorithm to apply to an entire hyperspectral scene. The final band configuration for all the pixels per algorithm is the same, despite the previous experiment having different band configurations for each single pixel. In this case, we computed an error representing the mean spectral approximation error over all pixels. This error is the average of the spectral approximation error for individual pixels. In this experiment, we computed the RMSE of the spectral configurations provided by the PCFA and SRS(RMSE) algorithms for the Moffiet field scene with respect to the number of bands in the reduced spectral configuration. The APCA algorithm is not applicable in this experiment since its error of approximation was much higher than the two other algorithms (Section 2.5.1). Figure 2.6 shows the results and Table 2.5 presents more details of six different number of bands.
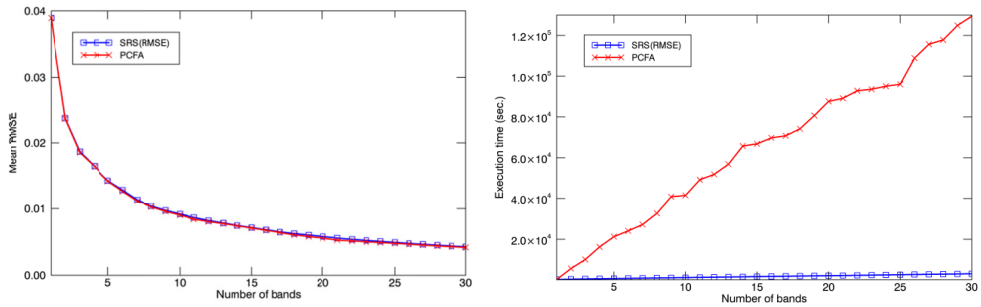
Figure 2.6. The mean approximation error (left) and the execution time (right) obtained by two methods for the entire Moffiet Field scene with respect to the number of bands.

Table 2.5: Run – time(s) and RMSE of the representation algorithms applied to the entire Moffiet Field scene spectra.

No. of bands	SRS		PCFA	
	Error	Run time	Error	Running time
5 bands	0.0165	653.16	0.0164	16258.2
10 bands	0.0097	1120.56	0.0096	40804.68
15 bands	0.0075	1587.72	0.0074	65752.92
20 bands	0.0060	2052	0.0058	80703.6
25 bands	0.0051	2515.2	0.0049	95084.16
30 bands	0.0044	2995.8	0.0043	124891.6

The error of spectral representation using the same configuration for an entire scene reveals an almost complete overlap between the two graphs showing the mean representation error obtained by the PCFA and SRS algorithm. Table 2.5 indicates that the difference between the mean error of two methods is about 0.0001. On the other hand, the PCFA was about 37 times slower than SRS, on average. The higher the number of bands, the slower PCFA than SRS, as clearly illustrated by Figure 2.6 (right). SRS provides a spectral configuration for the entire scene as accurate as the best spectral configuration provided by the PCFA algorithm but in a much faster way. For target identification using a spectral configuration explained in this section, we refer to Chapter 5.

2.5.4. Discussion

The experiments show that, although the PCFA algorithm gives the best spectral configuration, on the basis of RMSE, it is rather slow, especially when the

number of bands is high. As an alternative, we proposed an algorithm, SRS, that is much faster and yields spectral representation almost as accurate as the ones provided by PCFA and the location of splits are almost identical. This result was observed in three experiments. Furthermore, we can obtain different spectral configurations by applying other metrics with SRS. This option may lead to a better outcome in applications. For example, in the second experiment the reduced spectral configurations obtained by for example SRS(SID) gave better identification accuracy than the ones obtained with PCFA. In addition, it is worth to have spectral configurations provided by SRS(RMSE) with one more band than the representations obtained with PCFA, since the accuracy of material identification is similar or even higher, while the SRS algorithm is faster. When the same spectral configuration is required for dimensionality reduction of an entire scene, SRS represents the spectra with the same error as PCFA, while the time required to obtain the configuration is much shorter.

2.6. Conclusions

In this chapter, we proposed an algorithm, the spectral region splitting (SRS) method, to represent the spectral signals captured by imaging spectrometers and reduce the dimensionality of the data. The final band configurations obtained by the algorithm are sets of continuous spectral bands covering the whole spectrum. Compared with the PCFA method, which provides the best spectral configuration minimizing the representation error, SRS yields band configurations almost as accurate as PCFA, but in a much faster way. Applying the algorithms to more than 1000 diverse spectra to provide spectral configurations with 5 to 30 bands, SRS was 50 to 1200 times faster than PCFA, while the mean difference in RMSE was 0.0002 on reflectance scale, i.e. in [0,1] . The difference was even less when both algorithms provided the same spectral configuration for an entire hyperspectral scene, i.e. 0.0001. SRS readily accepts various spectral matching metrics as well. We have applied the SRS method using five different metrics to the spectra of a complete spectral library for material identification. The configurations obtained with SRS(RMSE) led to material identification nearly as accurate as with PCFA, while sometimes with SRS(SID) and SRS(SID-SA) the accuray of identification was higher. Overall, a spectral configuration per material signature with 11 spectral bands out of 224 channels was sufficient to identify the materials in a complete spectral library with 99% accuracy.

3.

A rapid algorithm to select independent spectral regions in hyperspectral images

Hyperspectral channels are highly dependent (Motta, Rizzo et al. 2006; Bioucas-Dias, Plaza et al. 2013), especially within adjacent spectral regions, i.e. many channels provide redundant information to some extent. Therefore, a method that can reduce the redundancy of hyperspectral datasets without losing much information is a solution to deal with constraints on data transmission and storage. As explained in the Introduction, our objective is to achieve a method that can be implemented onboard on a nano-satellite, i.e. it must be fast and uncomplicated. A fast approach can be developed by applying an unique score for measuring dependence to identify highly dependent spectral features in a hyper-dimensional space. Finding the score and a dimensionality reduction (DR) method to address these objectives is the main research question of the study described in this Chapter.

3.1. Introduction

There are different objectives (see Chpater 1) to apply DR to hyperspectral datasets. One of them is to select the channels that have a lower dependence with other channels (Chein and Su 2006; Martinez-Uso, Pla et al. 2007; Sotoca, Pla et al. 2007; Li and Liu 2009; Feng, Jiao et al. 2014). By choosing a channel set with low dependence, the number of channels and redundant information are reduced. Furthermore, by selecting the channels carrying independent information, the physical meaning of the spectral properties of observed objects is retained. This type of DR techniques is named feature selection (Bruzzone, Roli et al. 1995; Serpico and Bruzzone 2001). The alternative is feature extraction using a transformation from the original data space to a new feature space having the minimum dependence between the new components (Kumar, Ghosh et al. 2001; Serpico and Moser 2007; Serpico, Dellepiane et al. 2012). The transformation leads to a more complex relation between spectral attributes and object properties. Hence, feature selection techniques are preferable in the hyperspectral domain (Chein and Su 2006; Martinez-Uso, Pla et al. 2007).

The other categorization of DR in the hyperspectral domain is supervised vs. unsupervised techniques, where the former usually need a priori knowledge of the land-cover in a scene, and the latter can be implemented without extra information. Using the dependence as the criterion for DR of hyperspectral images paves the way for applying unsupervised methods, i.e. based on the inherent complexity of the spectral signal rather than on ancillary information on relevant classes. Available unsupervised algorithms to select independent channels from the datasets are usually complicated and time-consuming (Martinez-Uso, Pla et al. 2007; Sotoca, Pla et al. 2007; Du and Yang 2008; Jia, Ji et al. 2012). A fast approach can be developed by applying an unique score for measuring dependence to identify highly dependent channels in a hyper-dimensional space.

There are several metrics of dependence that can be applied to hyperspectral datasets. Usually, these metrics just calculate the dependence between two variables, e.g. the correlation coefficient. For more than two variables, the dependence is evaluated for paired variables, and shown in a pairwise dependence matrix. There exist some dependence metrics based on information theory (Timme, Alford et al. 2014), e.g. mutual information (Studholme, Hill et al. 1996; Masulli and Valentini 2001; Bettencourt, Gintautas et al. 2008), Kullback-Leibler divergence (Cover and Thomas 2006) and total correlation (Watanabe 1960) that can provide a unique score by measuring shared information for more than two dimensions. These metrics are mostly computationally expensive, however, especially when the number of bands is large. Furthermore, difficulties may arise when they are applied to more than two variables. For, example, Matsuda (Matsuda 2000) analyzed higher-order mutual information (HMI) to show the correlation properties of general complex systems. The author concluded that HMI can be either positive or negative whereas the usual mutual information is always non-negative. Sotoca, et al. (Sotoca, Pla et al. 2007) also showed that the total correlation metric is not appropriate for multivariate dependence since part of the dependence information between variables is recounted several times. They proposed another metric to measure dependence, but this metric is time-consuming and needs large computer memory, since the metric is estimated from the joint entropy of different bands and the computational cost grows exponentially with the number of bands (Sotoca, Pla et al. 2007).

In this study, we propose and use a metric of dependence in multi-dimensional spaces with low computational cost. This metric, called Total Dependence (TD), is based on the eigendecomposition concept and makes it technically possible to compare the overall dependence of multi-dimensional datasets having different dimensionality.

 Considering the properties of TD (see Section 3.2), a DR technique is applied to hyperspectral images with the aim of identifying the spectral regions with minimum dependence. In this study, we apply the metric to a clustering approach to define a new, reduced spectral configuration with minimum dependence. At first, the adjacent channels are clustered based on their dependencies in a bottom-up searching strategy. Then, instead of selecting an individual channel from each cluster to represent the cluster, we merge the spectral channels by averaging the signal, obtaining a continuous spectral configuration covering the entire spectrum. This implies that the least dependent bandset is chosen among different spectral configurations via a single dependence metric. This chapter is structured as follows:

Detailed explanations on the TD metric and its properties are provided in Section 3.2. Section 3.3 presents the details of the proposed bottom-up search strategy to define the reduced spectral configuration. The characteristics of the datasets used in this study are given in Section 0. In section 3.5, TD properties are

exploited in three experiments on multispectral and hyperspectral data, both real and synthetic ones. Section 3.6 gives an evaluation of the spectral configurations obtained by the algorithm. Section 3.7 summarises the approach and the results.

3.2. Total Dependence (TD)

The general approach to represent the dependence in multi-dimensional spaces is to calculate the dependence between each pair of variables to obtain a dependence matrix. This matrix is a square matrix where each element shows the dependence between the named variables in the specific row and the specific column. This matrix is symmetric. For example, given n random variables such as spectral bands $X = \{X_1, X_2, \ldots, X_n\}$, the dependence matrix can be presented as follows:

$$\Sigma = \begin{bmatrix} f(X_1, X_1) & f(X_1, X_2) & \ldots & f(X_1, X_n) \\ f(X_2, X_1) & f(X_2, X_2) & \ldots & f(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(X_n, X_1) & f(X_n, X_2) & \ldots & f(X_n, X_n) \end{bmatrix} \quad (3.1)$$

where Σ is the dependence matrix of X, and $f(X_i, X_j)$ is the value of the dependence of the named variables. The most common dependence matrix is the variance-covariance matrix (covariance matrix). This matrix includes the variance of each variable on the main diagonal and the covariance in the non-diagonal positions. The covariance matrix can easily be replaced by the correlation matrix showing the correlation between each pair of variables. In this matrix, if all the non-diagonal elements have the absolute minimum (zero) or maximum (one) value then the dependence among all the variables is minimum or maximum respectively. However, if not, the overall dependence of the entire set of variables cannot be directly estimated.

The simplest single measure of dependence of multiple variables is the mean value of the pairwise correlation coefficients. The same mean value, however, can be obtained with different matrices. As an example, the non-diagonal elements of two following correlation matrices have the same mean value (0.3); while in the left matrix two variables are highly correlated, i.e. leaving out one of the two may preserve the required information content. In the case of the right matrix, there might be a need to have all the three variables, because of the low correlation between the variables.

$$\Sigma_1 = \begin{bmatrix} 1 & 0.9 & 0.0 \\ 0.9 & 1 & 0.0 \\ 0.0 & 0.0 & 1 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix}$$

One way to measure the linear dependence in a multivariate data set with a single metric is to use eigendecomposition of the dependence matrix. The eigendecomposition of a square dependence matrix, e.g. the correlation matrix, gives a transformed matrix, where the eigenvalues lie on the diagonal and the non-diagonal elements are zero. The first (largest) eigenvalue contains the maximum information in the transformed space (Shannon 1948; Garner and Mcgill 1956; Nering 1970). The first eigenvalue also carries information about the dependencies between the variables in the original data space (Friedman and Weisberg 1981; Salinelli and Sgarra 2011).

Kaiser (Kaiser 1968) showed that a correlation matrix can be expressed by a single metric based on the eigendecomposition theory. The metric, that we named Total Dependence (TD), is a function of the first eigenvalue of a correlation matrix. Friedman and Weisberg (Friedman and Weisberg 1981) proved that the first eigenvalue is a function of the central tendency of the correlations when the correlations are all positive. We also investigated the first eigenvalue numerically in this study, and showed that by increasing the dependence between any paired variables, the first eigenvalue becomes larger (see Appendix A). The first eigenvalue is an indicator of the dependence of the variables sampled by the data. To be scalable, the first eigenvalue is normalized to its range of variation, to vary between zero and one. The final metric, *TD*, can measure the dependence in a *n*-dimensional space and makes it possible to compare the overall dependence of multi-dimensional datasets having different dimensionality. In the context of hyperspectral imaging, the metric can be calculated quickly, irrespective of the number of bands. Experimental results on multispectral and hyperspectral data, both real and synthetic ones, shed light on the capability of TD to measure the overall dependence (see Section 0). In this research, we have extracted some properties of the metric found useful for hyperspectral DR purposes.

As mentioned, the metric is a function of the first eigenvalue of the absolute correlation matrix. The eigenvalues are obtained by a method like Principal Component Analysis (PCA), which transforms a dataset with dependent variables into independent variables. To be more accurate, PCA is defined (Jolliffe 2002) as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. By considering the first eigenvalue of a correlation matrix that all its elements are replaced by their absolute values, it has been shown that the higher the dependence, the greater the first eigenvalue (Mayer 1976; Friedman and Weisberg 1981). The proposed metric is expressed as follows:

$$\text{TD} = \frac{\lambda_1 - 1}{n - 1} \qquad\qquad (3.2)$$

where λ_1 is the first eigenvalue derived of the correlation matrix of a data set on spectral data in n − bands, in which the elements have the absolute values. In the case of two example matrices mentioned above, the left and right matrices have the total dependence of 0.45 and 0.3 respectively, showing the overall dependence among the variables in the left matrix is higher than the right one.

For a given bandset as $X = \{X_1, \ X_2, \dots, X_n\}$, TD satisfies

a) $0 \leq TD(X_1, X_2, \dots, X_n) \leq 1$.
b) $TD(X_1, X_2) = |r(X_1, X_2)|$ where r is the correlation coefficient.
c) $TD(X_1, X_2, \dots, X_n) = 0$ if and only if $X_1, X_2, \dots, X_n$ are linearly fully independent.
d) $TD(X_1, X_2, \dots, X_n) = 1$ if and only if $X_1, X_2, \dots, X_n$ are linearly fully dependent.
e) $TD(X_1, X_2, \dots, X_n) = TD\big(X_{1(\sigma)}, X_{2(\sigma)}, \dots, X_{n(\sigma)}\big)$ for every σ permutation of $I_n \coloneqq \{1,2,\dots,n\}$.
f)   TD   is   a   non-monotonic   function   i.e.
$TD(X_1, X_2, \dots, X_{n-1}) \leq TD(X_1, X_2, \dots, X_n)$ or
$TD(X_1, X_2, \dots, X_n) < TD(X_1, X_2, \dots, X_{n-1})$

As TD is a function of λ_1, the properties mentioned above are mostly obtained from the characteristic of λ_1. The proof of the properties with some experimental results is given in Appendix B.

3.3. The Clustering-Merging (CM) method

For identification the spectral regions with low dependence, the SRS method is not applicable. The reason is that the SRS algorithm is a top-down algorithm. Using the minimum correlation as the criterion, SRS would determine the noisiest channels, i.e. in each iteration, it divides the spectrum into the spectral regions having the lowest correlation leading to identification of several narrow channels usually located in the ends of the spectrum and one broad band in the middle. This issue can be solved by considering a bottom-up search algorithm and instead of looking for the minimum correlated bands, searching for the maximum correlated ones.

Therefore, we use, first, the TD metric in a bottom-up clustering search algorithm, named Clustering-Merging (CM) algorithm, to identify highly dependent adjacent channels and cluster them. The intra-cluster correlation is high, while the inter-clusters correlation is low. The next procedure is the creation of the new spectral bands by averaging the signal in the channels per cluster and creating spectral band. Therefore, no channel spectral information gets just thrown away. Moreover, wider spectral bands increase the signal to noise ratio and

improve the quality of the image for further investigation. Finally, the algorithm produces a spectral configuration comprising both narrow and wide bands, which covers the whole spectrum. The narrow or wide bands are scene-dependent, i.e. the final bandset of a soil scene would be different from a vegetation scene. Furthermore, we expect that this method selects in a noise-free hyperspectral dataset the most uncorrelated spectral bands, causing a reduction in redundant information.

The method starts with the correlation matrix of the channels in a hyperspectral dataset, where the elements are positive, i.e. the absolute value of the correlation coefficients. In this step, the number of clusters and channels is equal, since every spectral channel is in one cluster. Given n spectral channels as $X = \{X_1, X_2, \ldots, X_n\}$ there are m clusters, $C = \{C_1, C_2, \ldots, C_m\}$, where $m=n$ initially. There are $m-1$ correlations between the adjacent clusters, $R = \{r_1, r_2, \ldots, r_{m-1}\}$ and $TD = \{TD_1, TD_2, \ldots, TD_{m-1}\}$ where $TD_i = |r_i(C_i, C_{i+1})|$ (see Property (b) in Section 3.2). Next, the algorithm searches for the most correlated adjacent clusters and insert them in one cluster, by replacing two highly dependent clusters with one cluster. The algorithm computes the dependence with the new cluster comprising the two highly dependent adjacent clusters using TD and update the TD sets to find the most highly correlated adjacent clusters in the next iteration. Based on Property (f) of TD, it is possible to compare the dependence of different clusters with a different number of channels. So, the algorithm finds highly dependent neighboring clusters and merges them in an iterative manner, until it reaches a predefined termination point based on the number of bands.

Table 3.1. Clustering-Merging algorithm

Algorithm CM(X, t)   // X is n-channel dataset, and t is the termination point
begin
Σ_{n*n} = abs(corr(X))   // Σ_{n*n} is the correlation matrix of X comprising the absolute value of the correlation coefficients
$C_{set} = \{C_1, C_2, \ldots, C_m\}$
$TD_{set} = \{TD_1, TD_2, \ldots, TD_{m-1}\}$
m=n
while ($m > t$)
i =index(max(TD_{set}))
$C_i \leftarrow \{C_i, C_{i+1}\}$ // *merging two clusters and update* C_{set}
$TD_{set} \leftarrow TD_{set}/TD_i$ // *Subtracting the ith TD form* TD_{set}
$TD_{i-1} = TD(C_{i-1}, C_i)$ // TD function calculates the total dependence of two adjacent clusters using Σ_{n*n}
$TD_i = TD(C_i, C_{i+1})$
m=m-1
endwhile
Final_bandset=the average of the channel signals per cluster in C_{set}
return Final_bandset
end

The final step is to average the channel signals per cluster and substitute a spectral region (band) instead of the cluster. The pseudocode of the algorithm is presented in Table 3.1.

3.4. Dataset

First, we examined the potential of TD to compute the overall dependence in multi- and hyper-dimensional spaces: by using a multispectral, a simulated and a hyperspectral dataset in three different experiments. Then the algorithm was evaluated by applying it to two hyperspectral datasets acquired by AVIRIS (airborne visible/infrared imaging spectrometer). The AVIRIS sensor covers the spectrum from 400nm to 2500nm in 224 spectral channels (Vane, Green et al. 1993; Green, Eastwood et al. 1998). We used the following datasets for the experiments:

a) A Landsat 8 image that covers agriculture, open water, and bare lands (Figure 3.1). We used six bands of the dataset (Table **3.2**) in two experiments.

Table 3.2: six bands of Landsat 8	
Band #	*Wavelength (nm)*
1	0.43 - 0.45
2	0.45 - 0.51
3	0.53 - 0.59
4	0.64 - 0.67
5	0.85 - 0.88
6	1.57 - 1.65

Figure 3.1: True-colour image of Salton Sea in Southern California and nearby irrigated fields, taken Mar. 24, 2013, by Landsat 8.

b) Moffett field: The data has been acquired by AVIRIS in California with 224 bands. The bandset covers the spectrum from 365nm to 2497nm continuously with approximately 10nm-wide bands. The bands located at 366-385,1353-1433,1811-1948,2337-2497nm wavelength were removed

due to noise and water absorption. So the final dataset has 177 bands (Figure 3.2).

Figure 3.2: True color of the Moffied Field hypesrpectral scene taken by AVIRIS in 1997.

c) Indian Pines: the scene consists of 145*145 pixels with a spatial resolution of about 20m. Two-thirds of the Indian Pines scene is covered by agriculture, and one-third by forest and other natural perennial vegetation (Figure 3.3). The ground truth available documents sixteen classes, not mutually exclusive. Since three classes in the scene contain less than 50 samples, we do not use them for the experiments. After the atmospheric correction and the removal of bad channels, the number of channels was reduced to 178. We removed water absorption channels (104-108, 150-163, and 220), noisy bands (1-4, 103, 109-111, 148-149, 164-166, and 217-219), and seven channels that are spectrally overlapping channels (32, 33, 95, 96, 158, 191, and 192). The Indian Pines dataset is available free of charge via Purdue University website: https://engineering.purdue.edu/~biehl/-MultiSpec/hyperspectral.html.

d) Salinas: This scene (Figure 3.4) is characterized by high spatial resolution (3.7m). The area covered comprises 512 lines by 217 pixels. The dataset is available at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral-_Remote_Sensing_Scenes only as at-sensor radiance.

Figure 3.3: True colour image of the Indian Pines scenes and the reference data of the classes used taken in 1992 (see Table 4.1 for the legend of the land-cover classes).

So, it has been atmospherically corrected and the noisy and duplicated channels have been removed. The final dataset has 190 channels. The ground-truth is also available and documents 16 classes including vegetables, bare soils, and vineyard fields, which we used in the experiments.
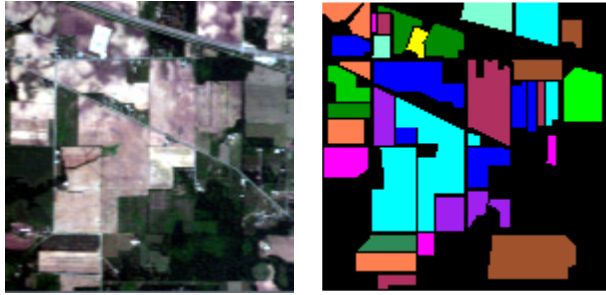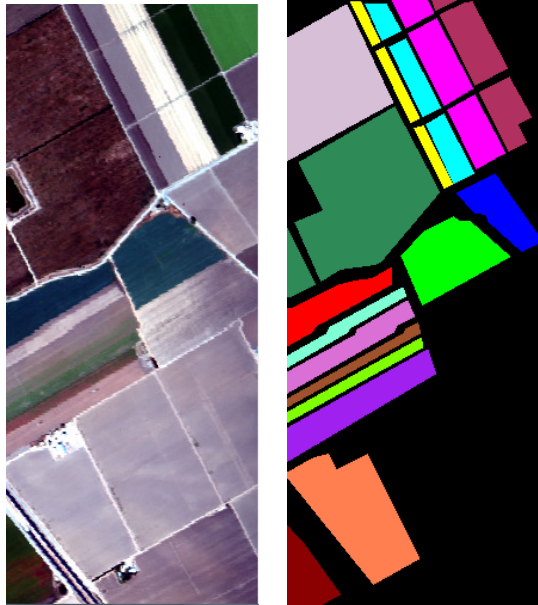

Figure 3.4: True colour image of the Salinas scenes and the reference data of the classes (see Table 4.2 for the legend of the land-cover classes).

3.5. **Evaluation of the TD metric**

In this section, The characteristics of TD, which are useful for identifying independent spectral features are explained and evaluated in practical cases. In each case, the total dependence (*TD*) metric is computed, and some of its properties are shown. Three different data sets: multispectral, simulated multispectral, and hyperspectral data are applied to evaluate *TD*. It should be noticed that this section is devoted to the TD metric, and in each experiment individual spectral channels are evaluated and selected, and there is no aggregation of channels.

3.5.1. **TD for comparing n-D spaces**
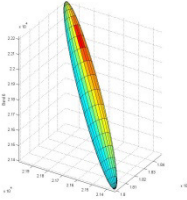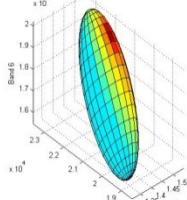
In the first experiment, three bands of Landsat 8 image are used, and the overall dependence is calculated for two different situations: 1-computing TD for the entire scene, and 2- computing TD for a soil-dominated sub-scene. This experiment shows the behaviour of the TD metric, when different land covers are compared on dependence. The results are also compared with a metric proposed by (Sotoca, Pla et al. 2007), (Θ_{DI}), that measures dependence on the basis of information theory. This measure is already used for band selection in multispectral cases. In the second experiment, *TD* is used to identify the three lowest dependence bands from the spectral bands mentioned in Table 3.2.

Experiment 1: Bands 4, 5, and 6 are selected from the dataset. The dependence metrics are calculated for the two given areas of the scene using the three bands (Table 3.3). For both areas the ellipsoid of the data distributions and the correlation matrix are also given in Table 3.3.

As expected, the correlation between bands in one type of land cover (soil) is higher than the correlation for a generic scene including several land cover types. This fact can be observed from the shape of the ellipsoid in the feature space and the correlation matrix. The three bands are highly correlated in the bare soil area, so the ellipsoid is narrow. Using the entire scene, the correlation between just two bands (5 & 6) is high (0.90). The data distribution is wider than the first situation as well.

Comparing the dependence metrics, Θ_{DI} gave higher dependence for the entire scene than the soil subscene, while TD has the opposite behaviour. The main reason is that Θ_{DI} is a cumulative measure for dependence (Sotoca, Pla et al. 2007), not a normalized measure, since this metric is a summation of all dependences between bands. If an image covers a larger area with different land types, it has more information. When the image comprises several spectral bands, it is more likely that the amount of shared information among the various spectral channels is greater as well. So Θ_{DI} that is the summation of shared information gives a higher value.

Table 3.3: The Ellipsoid shape of the data distribution, correlation matrix, the dependent information measure, and the total dependence metric for bands 4, 5, and 6 of the given areas in the Landsat 8 scene.

	Bare soil subset				*Complete Scene*		
Data distribution							
	Band#	4	5	6	Band #	4	5	6
Correlation matrix	4	1	0.94	0.83	4	1	0.15	0.90
	5	0.94	1	0.92	5	0.15	1	0.44
	6	0.83	0.92	1	6	0.90	0.44	1
Θ_{DI}	2.40				4.25			
TD	0.90				0.53			

In this experiment, the complete scene gives a higher Θ_{DI} value than the bare soil subset (4.25 vs. 2.40), although it can be seen from the correlation matrix that the overall dependence of the soil scene must be higher than the complete scene. This fact illustrates that Θ_{DI} may not be a suitable measure to compare the overall dependence of different datasets. On the other hand, the TD metric reveals that the complete scene is more independent than the soil sub-scene (0.53 vs. 0.90), which emphasizes that TD provides acceptable results in terms of comparing dependence of different data sets: when TD was used, the 3-bandset of the complete scene was identified as the dataset with lower dependence.

Experiment 2: in the second experiment we investigated how to identify the least dependent three out of six bands for the entire scene. TD was computed for 20 band combinations ($C_3^6 = 20$), which are shown in Table 3.4. The correlations between the band pairs are also given in Table3.5.

Table 3.4: Total dependence (TD) and the first eigenvalue (λ_1) for all combinations of three out of six bands of the Landsat-8 dataset.

Combination #	*Bands*	*TD*	Combination #	*Bands*	*TD*	Combination #	*Bands*	*TD*
1	1,2,3	0.99	8	1,4,5	0.51	15	2,4,6	0.91
2	1,2,4	0.98	9	1,4,6	0.91	16	2,5,6	0.52
3	1,2,5	0.52	10	1,5,6	0.52	17	3,4,5	0.52
4	1,2,6	0.91	11	2,3,4	0.98	18	3,4,6	0.92
5	1,3,4	0.98	12	2,3,5	0.53	19	3,5,6	0.54
6	1,3,5	0.53	13	2,3,6	0.92	20	4,5,6	0.53
7	1,3,6	0.92	14	2,4,5	0.51			

From Table 3.4, it is observable that the most independent band combinations are number 8 and 14. The bands are 1, 4, 5 and 2, 4, 5 with TD = 0.51. Bands 1, and 2 are completely dependent (r_{12} = 1.00), so substituting one instead of the other does not change TD. Looking at the different combinations and the correlations of the band pairs in that combination, it can be seen that TD is equivalent to the overall dependence. It is also obvious that the most dependent combination is the first one with three consecutive bands :1, 2, and 3.

Table3.5: Correlation matrix of the six bands of the entire Landsat-8 scene.

Band #	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
Band 1	1	1.00	0.99	0.97	0.16	0.87
Band 2	1.00	1	0.99	0.97	0.16	0.87
Band 3	0.99	0.99	1	0.98	0.20	0.89
Band 4	0.97	0.97	0.98	1	0.15	0.90
Band 5	0.16	0.16	0.20	0.15	1	0.44
Band 6	0.87	0.87	0.89	0.90	0.44	1

3.5.2. Sensitivity of TD

This experiment indicates the sensitivity of the TD metric to an additional band in a dataset. The new band may have either high dependence or low dependence with other bands in the dataset. We generated a synthetic data set for this experiment. The first seven bands (X_1, X_2, ..., X_7) were generated with random integer values in [0,10]. Each band has 100 pixels, and they have very low correlation with each other. Then, different combinations of the simulated bands are generated, and the metrics are calculated. These combinations start with two bands (X_1 and X_2), and then X_2 is duplicated twice and added to the bandset. Next, the other bands (X_3 to X_7) are added one by one. So, the final band combination is given by $X_1, X_2, X_2, X_2, X_3, ..., X_7$. Table **3.6** shows the correlation matrix for the complete combination with nine bands. *TD* is computed whenever a new band is added, and the results are illustrated in Figure 3.5. The values of the metric are given in Table **3.7**.

Table 3.6: Correlation matrix for the simulated multispectral data

Band #	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 9
Band 1	1	-0.05	-0.05	-0.05	-0.15	0.11	-0.02	0.05	0.00
Band 2	-0.05	1	1.00	1.00	-0.04	0.06	0.02	0.02	0.02
Band 3	-0.05	1.00	1	1.00	-0.04	0.06	0.02	0.02	0.02
Band 4	-0.05	1.00	1.00	1	-0.04	0.06	0.02	0.02	0.02
Band 5	-0.15	-0.04	-0.04	-0.04	1	0.08	0.16	0.04	-0.17
Band 6	0.11	0.06	0.06	0.06	0.08	1	0.10	0.04	-0.06
Band 7	-0.02	0.02	0.02	0.02	0.16	0.10	1	0.06	-0.06
Band 8	0.05	0.02	0.02	0.02	0.04	0.04	0.06	1	0
Band 9	0.00	0.02	0.02	0.02	-0.17	-0.06	-0.06	0	1
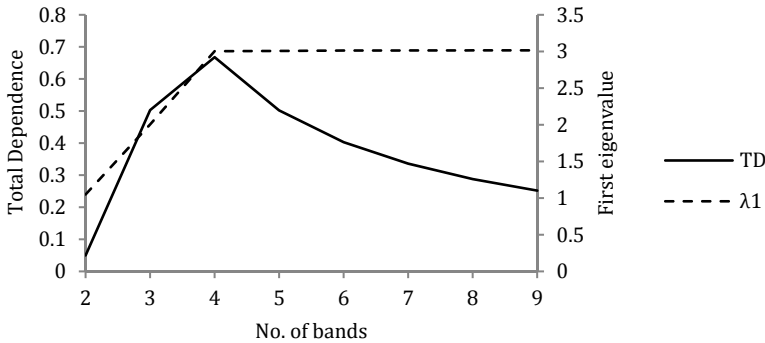
Figure 3.5: Variation of TD and λ_1 with respect to different band combinations for simulated multispectral data.

Table 3.7: Total dependence and the first eigenvalue for the simulated dataset.

No. of bands	TD	λ_1
2 bands	0.050	1.050
3 bands	0.503	2.005
4 bands	0.668	3.004
5 bands	0.502	3.007
6 bands	0.403	3.013
7 bands	0.336	3.014
8 bands	0.288	3.015
9 bands	0.252	3.016

In Figure 3.5, the TD graph starts with very low dependence for two independent bands (X_1 and X_2). After adding bands, one by one, the *TD* metric varies. When a band is duplicated and added (e.g. band 4), the overall dependence should be larger which is clearly illustrated by *TD*. On the contrary, *TD* shows that the total dependence decreases when an independent bands is added, as expected. For example, 5-band dataset has lower overall dependence than the 4-band one in this experiment, i.e. as expected due to the design of the synthetic data set.

We also computed the first eigenvalue of the correlation matrix to compare it with TD. Looking at the λ_1 graph, it is observable that λ_1 is always ascending (Figure 3.5). It is already proven that by increasing the dependence the first eigenvalue also rises. λ_1 increases sharply when X_2 is duplicated. Later the independent bands are added, and there is slight increase in the first eigenvalue (Table 3.7). It is because, the new bands have very low correlation with the other existing bands (see Table 3.6).

In general, λ_1 does not appear to be an appropriate metric of dependence as it monotonically increases, and we can not compare the dependence of different band sets. On the contrary, TD increases by adding a dependent band to the other bands and decreases when an independent band is added, as it should.

3.5.3. TD applied to real hyperspectral data

Here we carried out the same experiments as in Sections 3.5.1 and 3.5.2, but using a real hyperspectral image, Moffett field. First, the sensitivity of TD is examined with specific land cover types and, second, TD is applied to band selection from a hyperspectral data set.

Experiment 1: A region of interest from a vegetation area was extracted from the scene with 900 pixels. Similar to the experiment with synthetic data (Section 3.5.2), TD is calculated for the first two bands at the beginning of the spectrum, and then the other bands are added one by one sequentially at longer wavelengths, and the metric is calculated at each step.

Figure 3.6 (left) shows the TD values. It reveals that the metric can be computed for *n* bands and shows the overall dependency in an hyper-dimensional space. When the graph is ascending it means that the added bands are highly dependent with the other existing bands, and vice versa. For example, the bands located in the red edge spectral region for a vegetation scene, 34-37 (700-735nm wavelength), have very low correlations with the visible bands (the absolute mean correlation is less than 0.1). So by adding these bands to the previous bands, TD decreases (Figure 3.6 (left)).

Figure 3.6 (right) demonstrates the changes of the first eigenvalue with respect to the band numbers (corresponding to the wavelengths). Likewise the second case, the eigenvalue increases when a new band is added due to the dependency of the new band on the other bands. The first eigenvalue is always ascending, while the *TD* metric increases or decreases, depending on how dependent a new added band is on the bands already included in the data set.

Experiment 2: Suppose, five channels with the lowest dependence should be selected from a original hyperspectral bandset. In general, all the combinations should be considered. Because the number of combinations is huge, we have constructed 45 contiguous bands with 20 nm spectral width covering visible and Near Infra-Red (NIR) spectral regions. This bandset is derived from the Moffett Field dataset by averaging every two adjacent bands from 400 to 1240nm. Then, two sub-scenes were selected from vegetation and bare soil areas. TD is computed for all band combinations for the two scenes to find the spectral configuration with the lowest overall dependence.
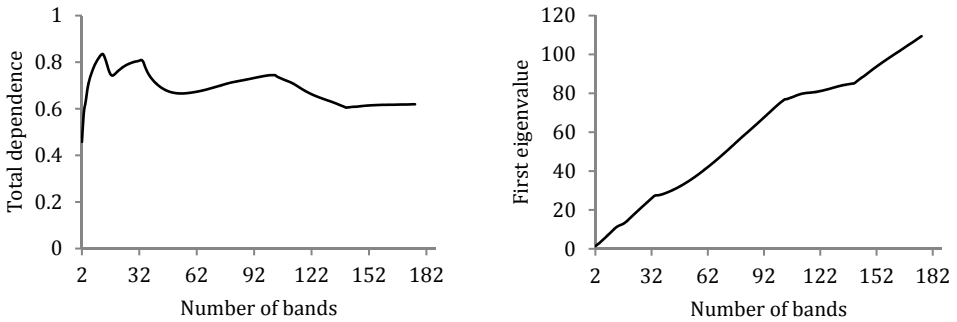
Figure 3.6: Variation of *TD* (left panel) and the first eigenvalue (right panel) as a function of number of bands from real hyperspectral data.

The number of combinations is $C_5^{45} = 1221759$. Figure 3.7 shows the TD values sorted from the minimum to maximum value for different band combinations and the two scenes. The minimum TD of the vegetation scene is less than the TD in the scene with bare soil areas 0.353 vs 0.451 respectively.
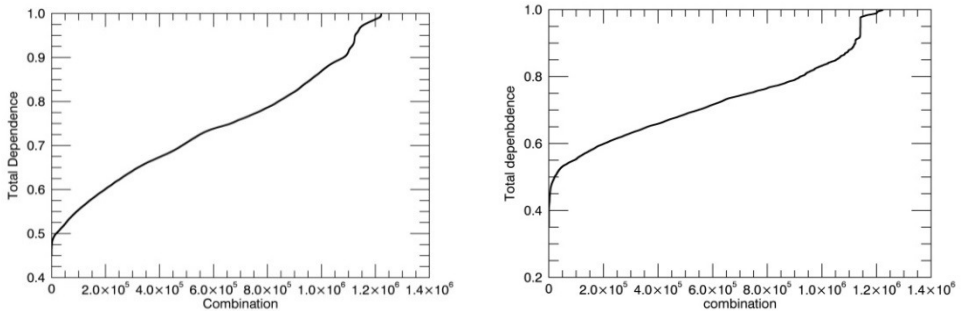
Figure 3.7: Sorted *TD* for all 5-band combinations of a hyperspectral scene over soil (right), and vegetation (left) areas.

Table 3.8 gives the five bands with the lowest, the mean, and the highest overall dependence. The 5 bands with minimum TD were selected from different parts of the spectral region. For the soil scene, three channels selected from the blue spectral region, and two separated bands from NIR, while for the vegetation scene, there are one blue, two green, and two continuous bands from NIR. On the other hand, the highest dependency 5 bands are contiguous for both scenes, with TD about one.

Table 3.8: Five selected bands from a hyperspectral image with the lowest, themean, and the highest overall dependence

	Combination with	Bands	Wavelength (nm)	TD
Soil scene	Min (TD)	1, 4, 5, 35, 40	404, 463, 483, 1043, 1139	0.451
	Avg (TD)	3, 24, 39, 40, 42	444, 832, 1120, 1139, 1178	0.739
	Max (TD)	34, 35, 36, 37, 38	1024,1043, 1062, 1082, 1101	0.999
Vegetation scene	Min (TD)	1, 8, 9, 36, 37	404, 542, 562, 1062, 1082	0.353
	Avg (TD)	3, 5, 36, 37, 42	444, 483, 1062, 1082, 1178	0.722
	Max (TD)	23, 24, 25, 26, 27	813, 832, 851, 870, 889	1.000

3.5.4. Discussion

TD was evaluated for three different cases in Section 0. From the first case, it appears that TD gives a normalized score of dependence to compare two or more different configurations of a multispectral dataset. The metric can be used to select a bandset with the minimum overall dependency from different band combinations, which is a key requisite in feature selection techniques. The second case reveals the behavior of TD when a new band is added to an existing bandset. The metric is not always ascending, unlike the first eigenvalue. Based on how dependent the new band is on the existing bandset TD might ascend or descend. Also, the second case illustrates that TD can be applied for a higher (any) number of spectral bands. This fact is also confirmed by the third case with 177 bands. In the third example, a real hyperspectral image from a vegetation scene is analysed. Here, it is shown that TD is applicable in practice. The sensitivity of TD to new additional bands is also clearly observable in the third experiment. The second experiment in case 1 and 3 demonstrated that TD can be applied to select the most independent band subset from multi and hyperspectral bandset.

Selecting bands with minimum dependence is a common approach to choose the optimal bands from hyperspectral data (Kumar, Ghosh et al. 2001; Martinez-Uso, Pla et al. 2007; Sotoca, Pla et al. 2007; Jia, Kuo et al. 2013). The TD metric can be used in feature selection techniques to provide a feature-set having the minimum overall dependency which can lead to a significant reduction in the dimensionality of hyperspectral datasets. In the next section, we applied TD to CM algorithm to identify independent spectral regions, instead of the selection of individual channels.

3.6. Evaluation of the spectral regions identified by CM

3.6.1. Comparison based on independence

For the evaluation the CM algorithm, we applied it to two entire hyperspectral scenes (Indian Pines and Salinas) and obtained reduced spectral configurations.

Table 3.9: TD computed for different band configurations over the entire scenes: Indian Pines (up) and Salinas scene (down). The band configurations were obtained in unsupervised and supervised cases.

| | No. of bands | Unsupervised | | Supervised | |
		Split location (nm)	TD	Split location (nm)	TD
Indian Pines scene	4	725, 1138, 1302	0.547	725, 1138, 1322	0.552
	5	715, 725, 1138, 1302	0.548	715, 725, 1138, 1322	0.551
	10	706, 715, 725, 937, 1129, 1138, 1148, 1302, 1322	0.578	706, 715, 725, 735, 937, 1138, 1302, 1322, 1342	0.583
	15	518, 706, 715, 725, 735, 937, 1129, 1138, 1148, 1283, 1302, 1322, 1332, 1789	0.590	567, 706, 715, 725, 735, 937, 1129, 1138, 1148, 1302, 1322, 1332, 1343, 1789	0.590
	20	439, 518, 706, 715, 725, 735, 937, 1120,1129, 1138, 1148, 1158, 1283, 1302, 1313, 1322, 1332, 1343, 1789	0.591	567, 706, 715, 725, 735, 851, 937, 1120, 1129, 1138, 1148, 1283, 1302, 1313, 1322, 1332, 1343, 1789, 2450	0.606
	25	439, 518, 706, 715, 725, 735, 812, 937, 1120,1129, 1138, 1148, 1158, 1177, 1196, 1206, 1273, 1283, 1302, 1313, 1322, 1332, 1343, 1789	0.589	518, 567, 706, 715, 725, 735, 745, 851, 937, 1120, 1129, 1138, 1148, 1158, 1283, 1302, 1313, 1322, 1332, 1343, 1789,2430, 2440, 2450	0.638
	178	All channels	0.724	----	----
Salinas scene	4	711, 1143, 1324	0.465	702, 1143, 1324	0.466
	5	702, 711, 1143, 1324	0.514	702, 711, 1143, 1324	0.514
	10	517, 566, 702, 711, 721, 1143, 1324, 1344, 1802	0.596	517, 702, 711, 721, 1143, 1314, 1324, 1344, 1813	0.606
	15	428, 517, 566, 692, 702, 711, 721, 942, 1143, 1314, 1324, 1344, 1434, 1802	0.617	428, 517, 566, 702, 711, 721, 942, 1143, 1314, 1324, 1334, 1344, 1574, 1813	0.620
	20	428, 517, 566, 692, 702, 711, 721, 731, 942, 1143, 1305, 1314, 1324, 1335, 1344, 1434, 1544, 1802, 1813	0.614	409, 428, 478, 517, 566, 693, 702, 711, 721, 731, 942, 1143, 1305, 1314, 1324, 1334, 1344, 1574, 1813	0.621
	25	409, 428, 517, 566, 606, 692, 702, 711, 721, 731, 942, 1134, 1143, 1305, 1314, 1324, 1335, 1344, 1434, 1544, 1793, 1802, 1813, 1948	0.615	409, 418, 428, 478, 517, 566, 636, 693, 702, 711, 721, 731, 942, 1134, 1143, 1295, 1305, 1314, 1324, 1334, 1344, 1574, 1803, 1813	0.621
	190	All channels	0.686	----	----

The spectral regions identified by the configurations have overall lower dependence. The number of bands started with 2 and ended with 30. TD was computed for of all bandsets. Table 3.9 gives the details of five band-sets (5-, 10-, 15-, 20- and 25-band) per scene. This information is under the heading "*unsupervised*" Additionally, Table 3.9 presents the band configuration that has the minimum TD, and the total dependence of all the spectral channels of the scenes.

In most cases, the dependence increases with the number of bands, although the TD metric is normalised to the number of bands, e.g., TD is 0.578 with 10-band and 0.591 with 20-band of Indian Pines scene. Using all the spectral channels gives the maximum TD = 0.724 and 0.686 for Indian Pines and Salinas scenes respectively. The minimum TD was obtained with four band configurations with 0.547 and 0.465 for the first and second datasets.

In the last experiment, we used the entire scenes to identify the independent spectral regions, i.e. without any a priori knowledge of specific areas in the scenes in an unsupervised manner. On the other hand, if some regions of interest in a scene are analysed with the CM method (supervised), and a new band configurations is obtained, how would the band configurations in supervised and unsupervised cases compare, as regards the dependence and the location of the splits? In the next experiment, we use particular pixels of the scenes, not all of them. Therefore, an equal number of samples per class of each scene was collected and utilised with the CM method. This is an experiment on supervised dimensionality reduction.

To identify how the overall dependence changes when different band configurations are used, we calculated TD for the entire scene using the bandsets obtained in the second experiment. This experiment reveals how different the independent spectral bands are when the method is applied in supervised and unsupervised ways.

In the supervised experiment, we selected 200 samples per class of the Salinas scene and 90 samples per class of Indiana Pines. So the number of samples used for identifying independent spectral regions are 3200 (200*16) and 1170 (90*13) for the Salinas and Indian Pines scenes respectively. Table 3.9 shows the details.

It is observed that the overall dependence between the bands is lower in the unsupervised than supervised cases, showing that changes in the spectral configurations increase the dependence between the spectral bands. For example, the total dependence is 0.596 in a 10-band dataset of the Salinas scene, while TD is more, 0.606, when the band configuration in supervised case was used. In the other words, using the entire scene provides a bandset with the lowest overall dependence among the bands.

Looking at the split locations obtained in unsupervised and supervised DR, several split locations are identical, since a large fraction of the scene is covered by the classes selected by the experiment. For instance, 7 out of 9, 12 out of 14, and 16

out of 19 split locations are the same in 10-, 15-, and 20-band sets of the Indian Pines scene.

3.6.2. Running time of the CM algorithm

We kept track of the running time of the CM algorithm and presented it in Table 3.10 for four band sets with 5, 10, 15, and 20 bands of the Indian Pines and Salinas datasets.

Table 3.10: The running time (sec.) of the CM algorithm with respect to the number of bands for two hypespectral scenes.

No. of bands	Indian Pines	Salinas
5	0.2510	0.2150
10	0.2210	0.1670
15	0.1950	0.1460
20	0.1750	0.1170

As observed, the CM method selects independent spectral regions for all the cases in less than half a second which is suitable for on-board data processing. It is because, in the CM algorithm, the dependence between the channels of the original dataset is, first, collected in a correlation matrix and then the CM algorithm is applied to the matrix in a clustering-merging procedure; therefore the method is so fast. Using TD in the clustering process is not time-consuming as calculating the metric is simple. Table 3.10 also reveals that when the number of bands is higher, the elapsed time is less since the algorithm is a bottom-up algorithm, i.e. it starts with the maximum number of bands and reduces it in each iteration. Therefore, the higher the number of bands, the less the elapsed time.

3.6.3. Comparison with existing algorithms

A common methodology to compare different feature sets obtained by various algorithms is to compare the accuracy in the detection of materials and objects in an image from reduced band configurations (Shaw and Burke 2003). The result of material detection per pixel in an image can be summarised in a similar way as classification accuracy since image classification is formally the process of assigning a label to the observations collected in an image dataset. Although selecting independent bands aims to a reduction of redundant information and does not have a direct influence on classification, it is expected that an independent bandset does not lose the required information for correct recognition of materials in a scene.

To validate the performance of the proposed method, we presented a comparison with three other unsupervised feature selection methods, focusing on

minimizing the shared information between the channels of hyperspectral data. The evaluation has been done by examining the number of features selected by the proposed and reference methods vs. the classification accuracy using different classifiers to check the relevance of the features selected. Since the running time of the algorithm is of importance for onboard processing, we compared the running time of all algorithms evaluated. These algorithms are as follows:

Clustering-based band selection: The method was proposed by Martínez-Usó et al. (Martinez-Uso, Pla et al. 2007) through a clustering-based strategy. This method, first, clusters the channels based on a dissimilarity measure and then selects the best representative per cluster. The advantage of this method is that it is not a ranking or incremental method that selects bands taking into account the previously selected bands, i.e. *m* selected bands in the clustering-based method is not equal to the *m-1* selected bands plus another relevant band.

Two dissimilarity criteria based on information theory were used: 1) mutual-information-based criterion and 2) divergence-based criterion. The authors utilized the measures to compute the amount of dissimilar information between two channels, and make a symmetric dissimilarity matrix for the entire hyperspectral dataset. Then, a hierarchical clustering process (Jain and Dubes 1988) was applied, to form clusters of channels as similar as possible within each cluster. After obtaining *K* desired clusters, a channel was selected by using a weighted method to provide the best representative channel predicting the information content of the other channels in each cluster. This algorithm used the Ward's linkage method (Ward 1963) in hierarchical clustering, so it is named WaLuMI (Ward's Linkage strategy using Mutual Information), and WaLuDi (Ward's Linkage strategy using Divergence). Since WaLuMI and WaLuDi usually obtain the same results (Martinez-Uso, Pla et al. 2007), we use just WaLuMI in this experiment.

Linear prediction and orthogonal subspace projection: Du et al. (Du and Yang 2008) applied two similarity-based endmember extraction algorithms (Plaza, Martinez et al. 2004) to select bands in an unsupervised manner. They searched the most distinctive bands based on linear unmixing methods to model a hyperspectral image pixel as a linear mixture of a set of finite image endmembers (Du, Ren et al. 2003). The two methods are 1) linear prediction (LP) and 2) orthogonal subspace projection (OSP), which were used as the criteria to select the most distinctive bands in a sequential forward selection searching strategy. They can jointly evaluate the similarity between a channel and multiple channels. The former criterion gives a linear estimate of the selected channels and searches for the most dissimilar one. The latter maps a channel onto an orthogonal subspace defined by the previously selected channels, and then, the channel that yields the maximum orthogonal component is considered as the most dissimilar channel to the selected ones.

In this approach, the authors proposed to use noise whitening after bad channel removal to obtain informative bands as well. They applied it to reduce the effect of the noise in the channels with large noise components which might be identified as informative while they are not. The noise whitening can be achieved by the eigendecomposition of the data covariance matrix. After noise whitening the linear-based methods are applied. The proposed methods are slow for large images, and the authors suggested to resample the data spatially since each band image is spatially highly correlated.

Comparison based on classification accuracy

We applied the three mentioned methods to compare the results with the feature sets obtained by the CM method. These comparisons were performed with regards to the classification accuracy of the two datasets.
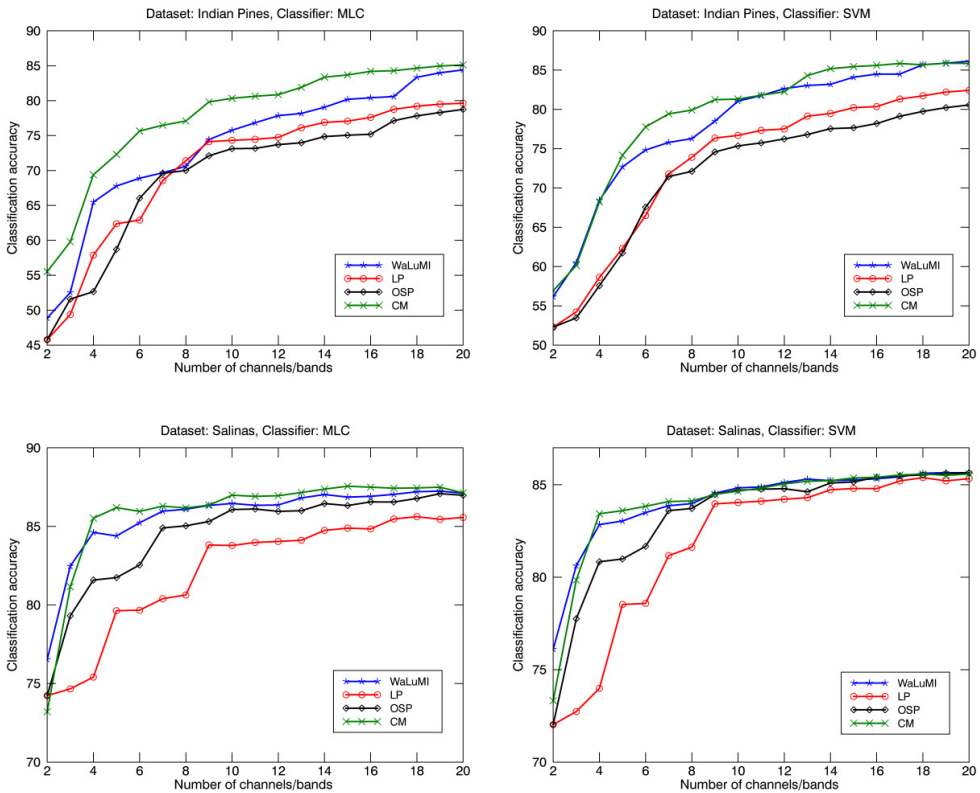

Figure 3.8: Classification accuracy of different feature sets for the Indian Pines (up) and Salinas (down) scenes.

We used two types of classifier: maximum likelihood classifier (MLC), and support vector machine (SVM), the former is a parametric classifier and the latter is non-parametric. Figure 3.8 shows the results.

In general, the proposed method gives better classification accuracy than its competitors. When MLC is applied, the classification results are clearly better than the other methods. When, using SVM, WaLuMI sometimes provides channel sets with comparable classification accuracy. The channel selection algorithms based on linear prediction always give less accurate results than CM.

1.1.1.1. *Comparison based on the running time*

We compared the CM method to the other ones on the basis of running time. Table 3.11 gives the elapsed time for the predefined number of channels or bands using different methods. As observed, the CM method is much faster than other methods. Therefore, it selects independent spectral regions rapidly which is suitable for an on-board band selection process. The main reason is that CM is independent of the number of pixels in a scene since all the calculation is based on the correlation coefficient matrix derived from the original dataset (assuming the effect of the number of pixels on running time to compute the correlation matrix is negligible).

Table 3.11: The running time (sec.) of different independent feature identification algorithms vs. number of channels/bands.

Number of features	Indian Pines				Salinas			
	LP	OSP	WaLuMI	CM	LP	OSP	WaLuMI	CM
5	6.580	11942	301.216	0.251	49.519	168442	555.06	0.215
10	18.000	20848	301.214	0.221	170.028	422118	555.03	0.167
15	45.870	30092	301.189	0.195	398.528	694587	554.96	0.146
20	83.398	39811	301.203	0.175	707.345	991460	555.76	0.117

The other three methods depend on the number of pixels in the scene, so the higher the number of pixels in a scene, the longer the time needed for implementation e.g. Salinas scene with 111104 pixels is more costly than Indian Pines with 21025 pixels. The LP and OSP use a sequential forward selection searching algorithm, so a higher number of bands requires more time. Both methods are very time consuming for large datasets, especially OSP.

Although WaLuMI uses a hierarchical algorithm that is fast for clustering the channels of hyperspectral datasets, computing the dependencies matrixes fed to the algorithm is time-consuming. As shown, the difference between the elapsed time using different number of channels is negligible, while the total time is very long,which is caused by the computation of the mutual information matrix derived from the hyperspectral dataset.

3.7. Conclusion

In this study, we proposed a fast band selection algorithm for onboard implementation on miniaturised hyperspectral platforms. Using TD, as a measure of overall dependence in a multivariate system, in a bottom-up searching algorithm (CM), we identified independent spectral regions from a hyperspectral dataset. We clustered adjacent highly dependent channels by considering a property of TD that allows comparing the overall dependence in various band sets. The final band set yields the relevant spectral regions in a contiguous manner. Band configurations selected by CM in an unsupervised manner were compared with the band configurations obtained by supervised DR, i.e. a set of sampled pixels from the available classes in the scenes were fed to the algorithm. The results showed the band configurations identified in unsupervised mode had a lower dependence. Besides, CM is a very fast algorithm that can be implemented onboard Cubesats. The elapsed time was independent of the size of the scene and was less than 0.3 seconds making the algorithm perfect for missions such CubeSats. The final band configuration is scene-dependent and includes uncorrelated spectral regions, which leads to significant reduction of the data dimensionality.

The band configurations selected by CM were compared with three other unsupervised band selection techniques on the classification accuracy of the given scenes and the running time. The results showed the CM method provided band sets giving a higher or comparable classification accuracy, and it was quicker than other algorithms.

4.
Spectral region identification vs. individual channel selection in supervised dimensionality reduction

Hyperspectral images may be applied to classify objects in a scene. The redundancy in hyperspectral data implies that fewer channels or spectral features might be sufficient to discriminate the objects captured by a scene. The availability of labeled classes of several areas in a scene paves the way for a supervised dimensionality reduction (DR), i.e. by using a discrimination measure between the classes in a scene to select spectral channels or spectral features. In this study, we show that averaging adjacent spectral channels and using wider spectral regions yields a better class separability than the selection of the individual channels from the original hyperspectral dataset. We used the spectral region splitting method which creates a new feature space by averaging neighboring channels. Besides the common benefits of channel selection methods, the algorithm constructs wider spectral regions when it is useful. Using different class separability measures over various datasets resulted in a better discrimination between the classes than the best-selected channels. The reason is that the wider spectral regions leads to a reduction in intra-class distances and an improvement of the class discrimination. The overall classification accuracy of two hyperspectral scenes gave about a two-percent increase when using the spectral regions determined by applying SRS.

4.1. Introduction

[1]Hyperspectral imagery can provide a complete spectral coverage in the 400–2500 nm spectral range with narrow spectral channels, causing an increase in the data dimensionality (Vane, Green et al. 1993; Green, Eastwood et al. 1998; Goetz 2009). On the one hand, the high dimensionality of hyperspectral data has the potential to identify observed targets and patterns reliably. On the other hand, it increases the complexity of the data that must be analysed (Kumar, Ghosh et al. 2001; Le Moan, Mansouri et al. 2011; Chen, Li et al. 2014). The main challenge of using hyperspectral images is to reduce the cost of data inspection without degrading the potential of image classification and pattern recognition (Shaw and Burke 2003; Jia, Ji et al. 2012; Fauvel, Tarabalka et al. 2013).

Dimensionality reduction (DR) techniques are usually applied to hyperspectral images before classification procedures, to mitigate the so-called curse of the dimensionality of hyperspectral datasets (Plaza, Martinez et al. 2005; Oommen, Misra et al. 2008; Peng, Zhou et al. 2015). In classification procedures, for example, the problem of hyper-dimensionality can usually be observed, i.e. the classification accuracy increases gradually with the number of features or dimensions but decreases when the number of features (spectral bands) becomes

---

[1] Chapter 4 is based on: Hosseini Aria, S. E., M. Menenti, et al. (2017). "Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data." Journal of Applied Remote Sensing 11(4): 046010.

large (Hughes 1968; Kamandar and Ghassemian 2011). Sometimes, using algorithms less dependent on the number of dimensions, such as the Support Vector Machine (SVM) (Oommen, Misra et al. 2008; Gheyas and Smith 2010; Alonso, Malpica et al. 2011), is a solution. Another way is to obtain a limited number of features and applying frequently used classification algorithms to achieve a high classification accuracy (Huang and He 2005; Martinez-Uso, Pla et al. 2007; Yang, Du et al. 2011; Jia, Ji et al. 2012; Han, Lee et al. 2013).

The main objective of the DR method used in our approach is to evaluate the potential of wider spectral regions versus the narrow channels of original hyperspectral data in providing a better class separability and a good classification accuracy with a limited number of features. Having sufficient labelled classes based on ground truth in an image captured by a hyperspectral sensor, it is feasible to measure the separability of the classes in the feature space. In general, the objective is to achieve maximum separability of the labelled classes, since by maximising class separability, the minimum-error in the classification is expected (Duda and Hart 1973; Swain and Davis 1978; Schowengerdt 1997). It can be achieved by selecting the spectral bands increasing the class discrimination. This type of band selection using a set of pre-defined classes is a supervised method (Huang and He 2005; Richards and Jia 2006; Yang, Du et al. 2011).

Ideally, the best bands for an image classification and class separability are those which provide larger inter-class distances and smaller intra-class diameters in the selected feature space (Duda and Hart 1973; Richards and Jia 2006; Han, Lee et al. 2013). For example, the inter-class distance can be obtained simply by measuring the Euclidian distance between the class means; and the class variances can provide the intra-class diameter. Separability measures such as Mahalanobis distance and Jeffreys-Matusita (Schowengerdt 1997) account for both the inter- and the intra-class distance implicitly, i.e. a larger Mahalanobis distance gives a higher ratio of inter- to intra-class distances. These separability measures can be plugged in a DR method to obtain a channel set that maximizes class discrimination (Huang and He 2005).

There are several supervised band selection methods (Chang, Du et al. 1999; Kumar, Ghosh et al. 2001; Huang and He 2005; Yang, Du et al. 2011; Han, Lee et al. 2013), mostly aiming at maximising class separability. In this Chapter, we apply a method called the Spectral Region Splitting (SRS) to identify spectral regions which maximize class separability. This method allows the use of different metrics to obtain the required spectral configuration, given a hyperspectral dataset. SRS, similar to other dimensionality reduction techniques, mitigates the curse of dimensionality (Hughes 1968) although it may not be categorised into the standard DR methods.

Usually dimensionality reduction is obtained by feature extraction and feature selection techniques with the aim of exploiting the most useful information carried by hyperspectral data (Martinez-Uso, Pla et al. 2007; Le Moan, Mansouri et al.

2011; Jia, Ji et al. 2012). Feature extraction methods transform the data into a new coordinate system to provide the required features, but the transformation may compromise critical information associated with the original spectral bands (Wang and Chang 2006). On the other hand, feature (channel) selection techniques identify a subset of the original channels for a given purpose, having the advantage of preserving the original information in the dataset (Guo, Gunn et al. 2006; Yuan, Zhu et al. 2015).

SRS is an alternative method to channel selection techniques, which merges adjacent channels into spectral regions (bands), instead of selecting a subset of the original spectral channels. Each band covers a number of adjacent channels of the original image, and each channel is used in exactly one band. The final band configuration is also contiguous, so it covers the whole spectrum captured by an imaging spectrometer. At every pixel, the signal value in a band is the average of the values of the adjacent channels included in this band. The primary advantage over channel selection is that no information gets just thrown away.

There are also other methods that cluster adjacent spectral channels for different purposes and substitute them with a value per pixel, e.g. the mean reflectance spectrum (Kumar, Ghosh et al. 2001; Jensen and Solberg 2007; Prasad and Bruce 2008; Cariou, Chehdi et al. 2011). For example, Jensen, et al. (Jensen and Solberg 2007) averaged adjacent spectral channels when needed to minimize the square representation error of spectral curve in a scene (see also Chapter 2); or Cariou, et al. (Cariou, Chehdi et al. 2011) used an unsupervised method to find the least dependent spectral regions by averaging adjacent channels (see also Chapter 3). The final bandset obtained by these methods including SRS is a set of contiguous spectral regions with different widths covering the whole spectrum. Besides the higher signal to noise ratio of the wider bands than of the narrow channels of the original hyperspectral data, in this approach, we demonstrate the advantage of averaging adjacent channels in increasing class discrimination.

We applied different separability metrics with SRS, and we evaluated all the metrics with two hyperspectral datasets. The separability metrics used were Euclidean, Mahalanobis, divergence, Bhattacharyya, transformed divergence and Jeffreys-Matusita distance (Schowengerdt 1997). We also compared SRS with three well-known search algorithms used in channel selection: Branch and Bound (BB) (Narendra and Fukunaga 1977), Sequential Forward Selection (SFS) (Whitney 1971), and Sequential Forward Floating Selection (SFFS) (Pudil, Novovicova et al. 1994).

The selected feature sets were also evaluated regarding classification accuracy, although we obtained the features independently of the classification method. This type of FS process is called filter approach (Kohavi and John 1997; Sebban and Nock 2002; Lodha and Kamlapur 2014), where there is no classifier involved in the feature selection procedure, and the features selected are entirely independent of the classifier applied in the subsequent classification procedure. As mentioned, in

our study, we identify features by focusing on class separability, which is just one of the factors to determine the final results of the classification. The other approaches are so–called wrapper models (Kohavi and John 1997; Sebban and Nock 2002; Lodha and Kamlapur 2014) that select features on the basis of classification accuracy. In general, the wrapper approaches search for a subset of features using the induction algorithm, and utilise its performance, usually, accuracy, as the criterion to select features. We also compared our classification results with the ones obtained with an algorithm based on the wrapper approach.

This study is organized as follows. Next section reviews frequently used search algorithms in channel selection and the separability metrics and their properties utilized in this approach. Section 4.3 explains the concept of the SRS method, and Section 4.4 describes the datasets utilized. Section 4.5 presents four different experiments and the results including evaluation of SRS by comparing it with the BB, SFS, and SFFS techniques and the classification accuracy achieved. A discussion is presented at the end of the section. Finally, the findings of this study are summarized in Section 4.6.

4.2.   State of the art

In this section, we review search algorithms widely applied in supervised channel selection methods. These algorithms select the channels where spectral information is more relevant to assign pixels to pre–defined classes in the given scene and "relevance" is measured by a specific separability metric. The selected channels can later be used in a classification procedure. The review also covers the separability metrics used in such algorithms. The review provides a framework to evaluate the application of SRS in combination with a separability metric for supervised selection of spectral features.

4.2.1.   Search algorithms

The aim of the feature (channel) selection methods is to select the best combination of n out of N variables on the basis of a predefined metric of performance. In the supervised feature selection for DR of hyperspectral images, finding a truly optimal combination is very challenging due to the dimensionality of data. The selection of a limited number of channels, e.g. 10, from an original hyperspectral image, with e.g. 224 channels, requires the evaluation of a huge number of combinations, i.e. $7.15*10^{16}$, which makes the evaluation of all of them impossible. This problem is known to be NP-hard (Amaldi and Kann 1998; Guyon, Andr et al. 2003). Therefore instead of an exhaustive search, a greedy algorithm is the solution, which solves the problem heuristically to find a near optimal solution (Cormen 2009).

There are several search strategies to select n out of N spectral channels of an original hyperspectral dataset. Here, we briefly explain three frequently used search algorithms, which later will be compared with the SRS method.

Branch and Bound (BB): This technique was developed to select the best subset of n features from an N-feature set (Narendra and Fukunaga 1977). The algorithm avoids the exhaustive search by rejecting suboptimal subsets without direct evaluation and guarantees that the selected subset yields the globally best value of any monotonic metric. However, the BB algorithm is applicable for small datasets, and if the number of features in the original dataset is high, the utilization of BB is expensive (Pudil, Novovicova et al. 1994). This is due to the rapid growth with the number of features of the enumeration scheme (solution tree) in the BB algorithm, leading to a dramatic increase in computational cost (Somol, Pudil et al. 2004; Bader, Hart et al. 2005).

Sequential Forward Selection (SFS): This method has been used in many feature selection approaches (Sotoca, Pla et al. 2007; Le Moan, Mansouri et al. 2011; Yang, Du et al. 2011; Han, Lee et al. 2013) and can be applied to large datasets. This method is much faster than BB (Pudil, Novovicova et al. 1994). SFS is an iterative process that selects a single element (such as a channel) that appears to be the best when combined with a previously selected subset of elements (Whitney 1971). The principle of SFS and SRS is similar (see Section 4.3). The method reduces the complexity of selection significantly by progressively ranking the evaluated selections. It permits an analyst to make trade-offs between system complexity and performance. However, SFS suffers a problem, the so-called "nesting effect", i.e. a feature once selected cannot be discarded later.

Sequential Forward Floating Selection (SFFS): This method is also a sequential search method that is characterized by dynamically changing number of features included or eliminated at each step (Pudil, Novovicova et al. 1994). Features selected can be later discarded in a search strategy which avoids the nesting effect problem. There is no guarantee, however, that SFFS always provides a better subset of features than SFS. SFFS is also faster than BB (Pudil, Novovicova et al. 1994).

4.2.2. Separability metrics

The separability distance shows how well the given classes are separated in the feature space, which provides guidance for the actual classification of images. Swain and Davis define the separability analysis as a performance assessment based on the training data to estimate the expected error in the classification for various feature combinations (Swain and Davis 1978). The training samples per class in a scene should be selected from the homogenous areas which provide a

unimodal histogram[1]. Having multimodal training data suggests that there are at least two different types of land cover existing in the training area. This condition is not suitable for discriminating individual classes and should be discarded and be retrained by focusing on parts of the polygon of interest until a unimodal histogram is optained per class.

There are several separability metrics, with the simplest ones just taking into account the inter-class distances, while most consider both inter- and intra-class distances. Figure 4.1 schematically shows two normally distributed classes ($N(\mu_a, \sigma_a^2), N(\mu_b, \sigma_b^2)$) in one dimensional feature space. The distance obtained by subtracting the mean values ($\mu_a - \mu_b$) gives the inter-class distance and the variances of the classes (σ^2) indicate the intra-class distances.

This section reviews the separability measures for two given classes a and b. In the following formulas, $\mu = \{\mu_1, \mu_2, .., \mu_n\}$ and Σ are the mean vector and covariance matrix of a class in an n-dimensional feature space. σ_i is the variance of a class in the ith dimension.
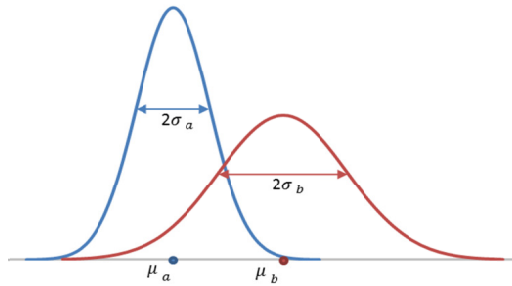
Figure 4.1. Two hypothetical classes with normal distribution in one dimensional feature space.

The simplest separability metrics just use the class means, e.g. the *Euclidean distance* (*ED* in Eq. (4.1)), which does not take into account the class variance, i.e. the intra-class distance.

$$ED = \|\mu_a - \mu_b\| = [(\mu_a - \mu_b)^T(\mu_a - \mu_b)]^{1/2} \qquad (4.1)$$

Other separability metrics take into account the mean and the variance of the classes. The multivariate generalization of the Euclidean distance is the

---

[1] A unimodal histogram or distribution is a histogram with only one peak. The mode in "unimodal" refers to the local maximum in the diagram.

Mahalanobis distance (*Mh* in Eq.(4.2)). It comprises both the inter- and intra-class distances.

$$Mh = \left[(\mu_a - \mu_b)^T \left(\frac{\Sigma_a + \Sigma_b}{2} \right)^{-1} (\mu_a - \mu_b) \right]^{1/2} \tag{4.2}$$

One of the problems of this metric is that the distance is zero when the class means are the same, no matter what the variances are. For example in Figure 4.1, if $\mu_a = \mu_b$ while $\sigma_a^2 \neq \sigma_b^2$, the *Mh* metric equals zero, showing that the metric is not comprehensive. The divergence (*D* in Eq.(4.3)) and Bhattacharyya (*B* in Eq.(4.4)) metrics overcome this problem, where $tr[A]$ and $|A|$ denote the trace and determinant of matrix A.

$$D = \frac{1}{2} tr \left[(\Sigma_a - \Sigma_b)(\Sigma_a^{-1} - \Sigma_b^{-1}) \right]$$
$$+ \frac{1}{2} tr \left[(\Sigma_a^{-1} + \Sigma_b^{-1})(\mu_a - \mu_b)(\mu_a - \mu_b)^T \right] \tag{4.3}$$

$$B = \frac{1}{8} Mh + \frac{1}{2} ln \left[\frac{|(\Sigma_a + \Sigma_b)/2|}{(|\Sigma_a||\Sigma_b|)^{1/2}} \right] \tag{4.4}$$

These two metrics are never zero unless both the means and the covariance matrices for the two classes are identical. Both metrics are the summation of two terms, where one of the terms is independent of class means keeping the metrics non-zero when the class means are equal. However, these two metrics are not yet entirely satisfactory as they do not have a maximum value and can get any positive value. The *Mh* and *ED* metrics have the same issue as well, which becomes a problem when evaluating the separability of multiple classes. In these cases, the separability is estimated by averaging the separability values over all class pairs. When the separability metric has no maximum value, the average may be biased due to one or more classes being far away from other overlapping classes in feature space (Swain and Davis 1978; Han, Lee et al. 2013). A separability metric with a known maximum value always provides a value showing whether the given classes have overlap or not. Furthermore, we can estimate how far the classes are. The closer the separability metric to the maximum value, the better separated the classes are. For example assume that there are four classes and the maximum value of the separability metric is *M*. To compute the separability in this multi-class case, we have to calculate the separability for the six class pairs and average them. If the average value is equal to *M*, it means that there is no overlap between the classes, otherwise there is.

The problem concerning the maximum value of the separability metrics is solved with the transformed divergence (*D^t* in Eq. (4.5)) and Jeffreys-Matusita (Bruzzone, Roli et al. 1995) (*JM* in Eq. (4.6)) metrics, which have a maximum value

of 2 and 1.4142, respectively. The maximum value is achieved when classes do not overlap with each other.

$$D^t = 2[1 - e^{-D/8}] \tag{4.5}$$

$$JM = [2(1 - e^{-B})]^{1/2} \tag{4.6}$$

The square root in the *JM* metric is sometimes omitted, which makes the maximum value of *JM* equal to 2 (Richards and Jia 2006). All of the mentioned separability metrics compute the distance between two classes. In multiclass cases, a common solution is to calculate the mean distance over all class pairs (Jensen 2015):

$$L_{avg} = \frac{\sum_{a=1}^{m-1} \sum_{b=a+1}^{m} L_{ab}}{C} \tag{4.7}$$

where L_{ij} is a separability measure between the classes a and b, m is the number of classes and C is the number of class pairs.

4.3. Proposed Method

4.3.1. Spectral Region Splitting

The spectral region splitting (SRS) method is a top-down algorithm i.e. it starts with a single broadband and breaks it down into narrower spectral bands. The signal in the initial broad band per pixel is the mean signal (radiance or reflectance) of all channels of the original dataset captured by an imaging spectrometer. The algorithm searches for the best spectral location to split the broad band into two bands, then applies iteratively the same procedure to the new spectral bands..

The search is based on the given objective and corresponding metric, which in this Chapter is the best class discrimination. The first two spectral regions give the maximum separability on average of the classes defined in the scene. Next, the algorithm searches for the second spectral location giving the highest separability, taking into account the location of the first split. This process continues iteratively till either a predefined mean separability (Eq. 4.7) is achieved with the selected bands, or the number of spectral regions reaches the predefined termination point. Finally, we have a set of continuous bands with different widths covering the entire spectrum. The SRS procedure is described schematically in Table 2.1.

4.3.2. SRS with the class separability metric

The channel selection methods operate keeping the original channel width (Huang and He 2005; Yang, Du et al. 2011; Han, Lee et al. 2013). Such methods search in the original hyperspectral space consisting of hundreds of narrow channels and select the channels providing the optimal value of the given metric. If the target application is classification, this metric is usually a separability metric., which is calculated based on the known classes in the scene. In each channel, the class attributes, such as mean or variance, are known and constant, and the search algorithm finds the best channel combination maximizing separability.

In the SRS algorithm, besides the possibility of selecting narrow channels, there is also the freedom to average the narrow channels and construct wider spectral regions with new class means and variances. The algorithm can average the channels to yield a better class discrimination than the selection of just individual channels. By averaging the channels, two situations may cause a better discrimination among the classes: 1) a larger distance between the class means (i.e. an increase in the inter-class distance), and 2) a smaller class variance (i.e. a reduction in the intra-class distance).

Here, we prove that the intra-class distance becomes smaller in the new feature space. Let $X = \{X_1, X_2, \dots, X_n\}$ be a class in a hyperspectral dataset having n independent spectral channels with m sample pixels per channel. The class variance is a vector given as $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$. The variance of the linear combination $Y = \sum_{i=1}^{n} a_i X_i$, where $a_1, a_2, \dots. a_3$ are real constants ($a_i \in R$), is:

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 \tag{4.8}$$

In SRS, when the neighbouring channels are averaged to yield a wider band, the variance of each class will be the summation of all the class variances in the channels divided by the factor n^2, where n is the number of constitutive channels of the band. In fact, if the class variances in all channels are equal to σ^2, the class variance for a band obtained by averaging those channels is σ^2/n^2, leading to a smaller intra-class distance. This argument can be the main point to achieve a better class separability when the SRS method is applied.

4.3.3. SRS iterations

SRS, likewise most search algorithms, detects a suboptimal feature subset based on the selected criterion, to avoid an exhaustive search for all possible combinations. The total number of subsets of size n out of N possible elements, $n<N$, is $C_n^N = \frac{N!}{(N-n)!n!}$. In the case of hyperspectral images, the number of

combinations is indeed very large, e.g. the number of possible selections of 10 out of 200 channels is $2.25*10^{16}$. Whitney (Whitney 1971) proposed the sequential forward selection method, to reduce the number of subsets evaluated and obtain a suboptimal subset. This iterative process selects a single element (like a channel) that appears to be the best when combined with a previously selected subset of elements. SRS follows the same rule as well. Since SRS searches for the location of the split to divide the spectrum, the number of possible locations is *N-1* at each iteration, where N is the number of channels. Therefore, the number of subsets searched to find a subset of n bands from an original dataset ($2 \leq n \leq N$) is given by

$$\sum_{i=1}^{n-1} (N - i) = (n - 1)\left(N - \frac{n}{2}\right) \tag{4.9}$$

Thus, the previous example of selecting 10 bands out of 200 channels requires evaluating 1746 band sets, i.e. 13 orders of magnitude (10^{13}) less than the number of band-sets searched by the exhaustive search strategy.

4.4. Dataset

The algorithm was evaluated by applying it to two hyperspectral datasets acquired by AVIRIS over a mostly vegetated area in North-Western Indiana, and the Salinas Valley, California, USA. The AVIRIS sensor covers the spectrum from 400nm to 2500nm in 224 spectral channels (Vane, Green et al. 1993; Green, Eastwood et al. 1998). The details of the images are as follows:

Indian Pines: The scene consists of 145*145 pixels with a spatial resolution of about 20m. Two-thirds of the Indian Pines scene is covered by agriculture, and one-third by forest and other natural perennial vegetation. The available ground truth on land cover is based on sixteen classes, but it does not completely cover the entire scene. Since three classes in the scene contain less than 50 samples, we did not use them for the experiments. After the atmospheric correction and the removal of noisy channels (Appendix A), the number of channels was reduced to 178. We removed water absorption channels (104-108, 150-163, and 220), noisy channels (1-4, 103, 109-111, 148-149, 164-166, and 217-219), and seven channels that are spectrally overlapping (channels 32, 33, 95, 96, 158, 191, and 192). The Indian Pines dataset is available free of charge via Purdue's University website: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html.

Figure 4.2 shows the scene and its reference ground truth data with 13 classes. Table 4.1 is the legend of the reference data giving information about all the available classes in the scene in addition to the unclassified pixels (unknown class).

The classes not included in our evaluations are shown as black, the same as the
unclassified pixels.

Figure 4.2. True color image of Indian Pines scene and the reference data.

Table 4.1. The legend of the Indian Pines reference data and their respective samples
number.

#	Class	Samples	Color		
1	Alfalfa	46			
2	Corn-notill	1428			
3	Corn-mintill	830			
4	Corn	237			
5	Grass-pasture	483			
6	Grass-trees	730			
7	Grass-pasture-mowed	28			
8	Hay-windrowed	478			
9	Oats	20			
10	Soybean-notill	972			
11	Soybean-mintill	2455			
12	Soybean-clean	593			
13	Wheat	205			
14	Woods	1265			
15	Buildings-Grass-Trees-Drives	386			
16	Stone-Steel-Towers	93			
17	Uknown samples	10776			

Salinas: This scene is characterized by high spatial resolution (3.7m). The area
covered comprises 512 lines by 217 samples. The dataset is available at
http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes only
as at-sensor radiance with 20 water absorption channels (108-112, 154-167, and
224) discarded. We corrected the data atmospherically and removed the noisy (see
Appendix A) (1-4, 107, 113, and 220-223) and duplicated channels (33, 34, 97, 98).
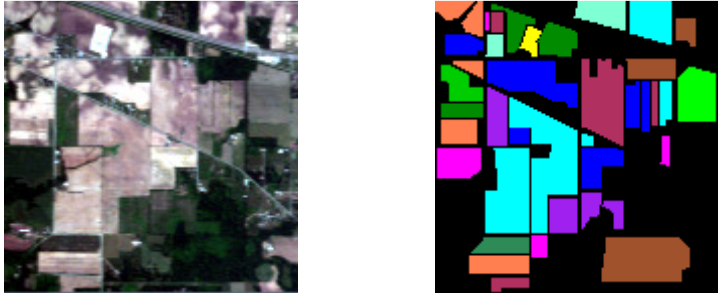The final dataset has 190 channels.
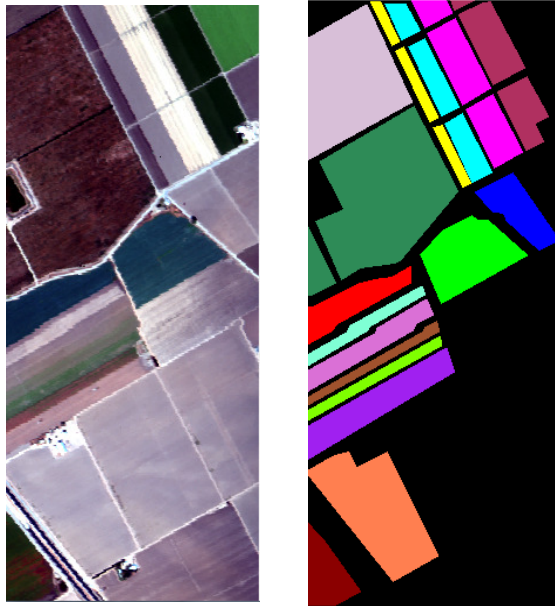
Figure 4.3. True color image of the Salinas scene and the reference data.

Table 4.2. The legend of the Salinas reference data and their respective samples number.

#	Class	Samples	Color
1	Brocoli_green_weeds_1	2009	
2	Brocoli_green_weeds_2	3726	
3	Fallow	1976	
4	Fallow_rough_plow	1394	
5	Fallow_smooth	2678	
6	Stubble	3959	
7	Celery	3579	
8	Grapes_untrained	11271	
9	Soil_vinyard_develop	6203	
10	Corn_senesced_green_weeds	3278	
11	Lettuce_romaine_4wk	1068	
12	Lettuce_romaine_5wk	1927	
13	Lettuce_romaine_6wk	916	
14	Lettuce_romaine_7wk	1070	
15	Vinyard_untrained	7268	
16	Vinyard_vertical_trellis	1807	
17	Uknown samples	56975	

The ground-truth on the land cover is also available, which contains 16 classes including vegetables, bare soils, and vineyard fields, which all the classes used in

the our experiments. Figure 4.3 shows the scene and the ground – truth reference data, and Table 4.2 shows the legend.

4.5. Experiments and results

SRS was applied to the two hyperspectral datasets with the criterion of optimal separability between the classes in the scenes. To assess SRS, we compared it with the search algorithms used for channel selection and described in Section 4.2.1. This evaluation was performed by comparing the mean separability (Eq. (4.7)) and classification accuracy vs. the number of features in each step. We also compared the classification accuracy obtained by SRS with a wrapper algorithm.

For the experiments, we separated the available samples (pixels) for each land-cover class into training and testing data using a random subsampling method, so that the training set represents the distribution of class attributes well. The training datasets were generated by taking 35%, 50% and 70% of the total number of samples (pixels) per class and the remaining samples were considered as validation data. Finally, the results obtained with the three datasets were averaged. We repeated random sub-sampling validation using different partitions of the samples per class and averaged the results (Bubitzky, Granzow et al. 2007) to derive a more accurate estimate of the model performance and generalise the statistical analyses.

4.5.1. SRS vs. best-selected channels

In the first experiment, we compared the spectral bands constructed by SRS with the best channels having the maximum separability selected by the BB search algorithm. As mentioned, BB selects the best n features with the highest value of the given metric out of an N-feature dataset. All other channel selection methods would choose n channels giving a lower or at best equal value of the metric than the channel set selected by BB. However, BB is costly when applied to large datasets (Pudil, Novovicova et al. 1994). Therefore, we used smaller datasets with 20 channels in this experiment.

The 20-channel datasets were chosen from two spectral regions of the Indian Pines scene: 1- visible (VIS) spectral region 400-637nm, and 2- near-infrared (NIR) spectral region 647-822nm. The two common separability measures; Mh and JM, were applied as separability metrics with the SRS and BB algorithms. We also applied the two other search algorithms to select channels: SFS and SFFS. These two algorithms do not select the channels having a better result than BB since they select a suboptimal channel set (see Section 4.2.1). We used them in this experiment to benchmark their performance against BB, prior applying them to the entire scene in the second experiment.
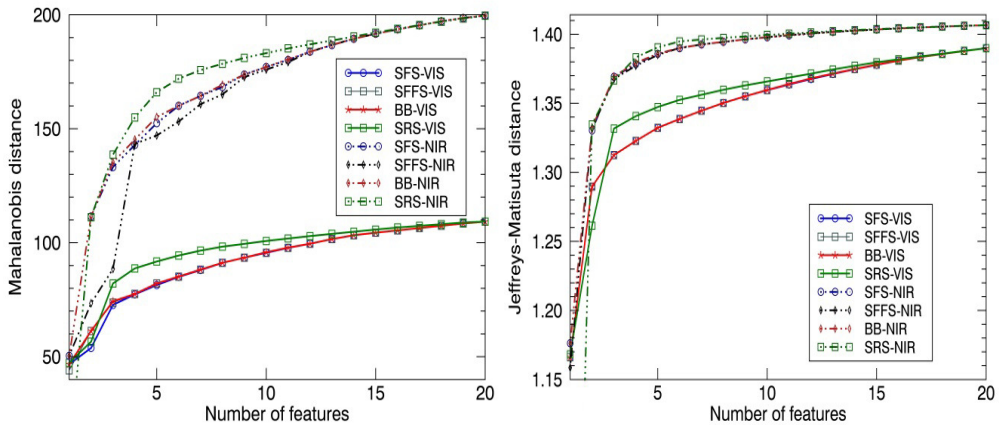
Figure 4.4. The mean class separability vs. the number of features obtained by applying the Mahalanobis (left panel) and Jeffreys-Matusita (right panel) distances and the four algorithms described in the text; the datasets used are two spectral subsets with 20 channels of the Indian Pines scene in the visible (VIS) and near infrared (NIR) spectral region.

The first experiment shows the SRS method can provide a better class discrimination than the best selection of spectral channels (Figure 4.4). When the number of bands is small (in the worst case was three); SRS may not give a better class separability than channel selection. After a few iterations, however, the SRS method creates spectral regions providing better separability. The spectral regions identified by SRS would discriminate the classes better than the same number of selected spectral channels. This result was obtained in all cases, regardless of whether Mh (Eq. (4.2)) or JM (Eq. (4.6)) was used. Table 4.3 also gives the value of JM on the NIR subset for more clarification.

The performance with both SFS and SFFS is almost as good as the BB selection in most cases. SFFS sometimes gave a lower separability than SFS, e.g. when Mh is used over NIR spectral regions. The result of the final selection, i.e. 20 features, was the same for all the algorithms, since all available features were applied to determine separability with all the algorithms.

Overall, a spectral configuration determined by SRS identifies features that give better class separability than the other search algorithms used for channel selection in this study (Figure 4.4 and Table 4.3). The widely used search methods, SFS and SFFS, result in channel sets providing lower class separability than the channels selected by BB per dataset to maximize separability, although the differences are rather small.

Table 4.3. Mean Jeffreys-Matusita distance
between the class pairs using the bands
and channels obtained by different
algorithms over the NIR dataset.

No. of features	SFS	SFFS	BB	SRS
1	1.1762	1.1585	1.1762	0.7635
2	1.3304	1.3314	1.3325	1.3347
3	1.3692	1.3676	1.3694	1.3666
4	1.3784	1.3772	1.3793	1.3833
5	1.3858	1.3850	1.3859	1.3905
6	1.3899	1.3901	1.3901	1.3948
7	1.3926	1.3926	1.3926	1.3962
8	1.3945	1.3944	1.3945	1.3974
9	1.3963	1.3963	1.3963	1.3984
10	1.3978	1.3979	1.3979	1.3995
11	1.3992	1.3994	1.3994	1.4004
12	1.4004	1.4006	1.4006	1.4013
13	1.4017	1.4017	1.4017	1.4022
14	1.4026	1.4025	1.4026	1.4030
15	1.4034	1.4034	1.4034	1.4037
16	1.4042	1.4042	1.4042	1.4044
17	1.4049	1.4049	1.4049	1.4050
18	1.4055	1.4055	1.4055	1.4056
19	1.4061	1.4061	1.4061	1.4061
20	1.4065	1.4065	1.4065	1.4065

4.5.2. SRS vs. conventional channel selection methods

The second experiment compares the SRS with SFS and SFFS by applying them to the complete datasets. The BB algorithm is not applicable for a large dataset (see Section 4.2.1). We used the six separability measures mentioned in Section 4.2.2, and both datasets were used, Indian Pines and Salinas. Figure 4.5 and Figure 4.6 illustrate the trends in mean separability vs. the number of features for the two different hyperspectral scenes, where the number of features increases from two to maximum 30.

Overall, SRS yields a better separability with a higher number of bands in all cases except when Euclidean distance was applied. The main difference between the Euclidean distance and the other metrics is the class variance that is not taken into account by the Euclidean distance, which only accounts for inter – class distance.
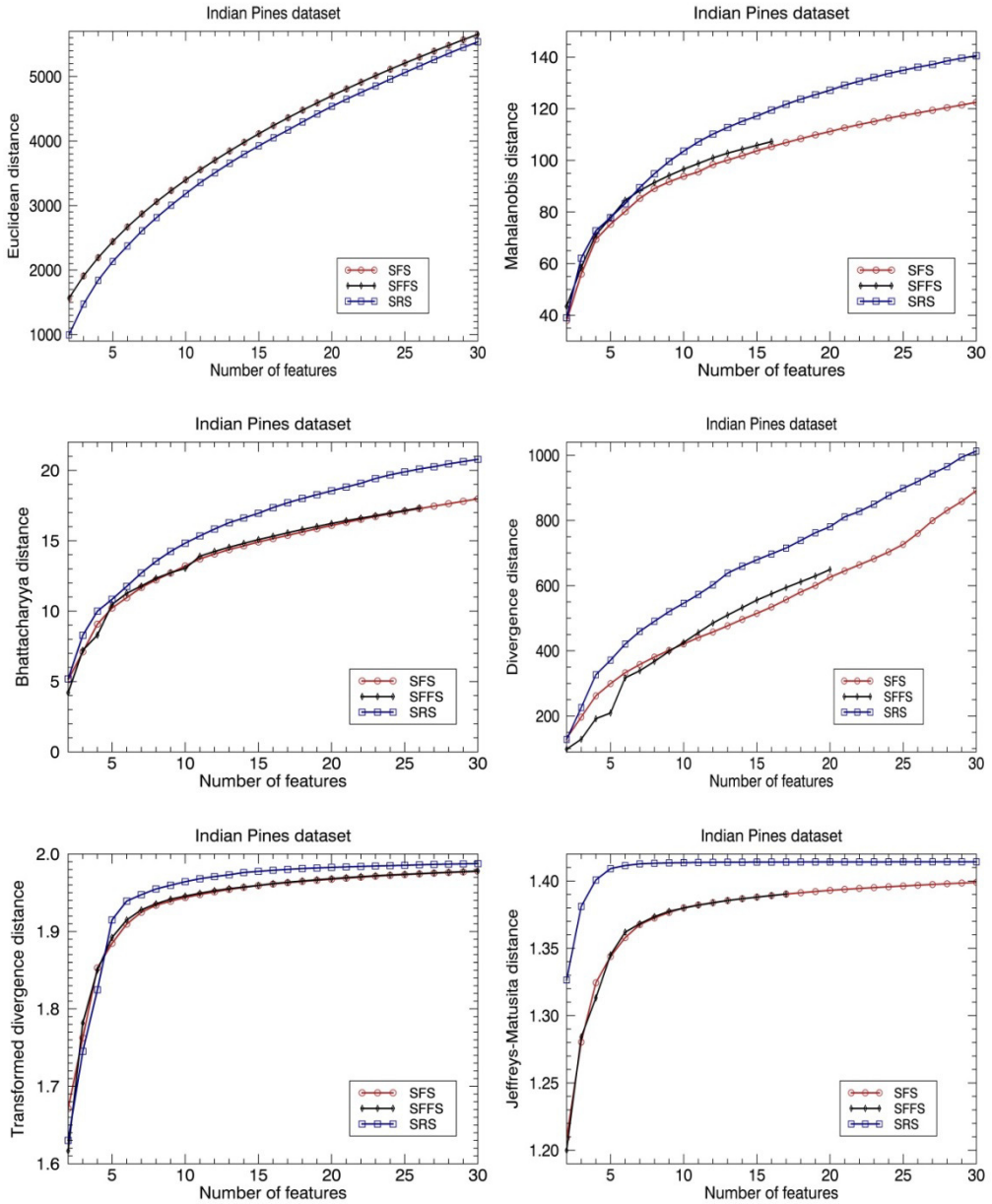
Figure 4.5. The mean separability of the features obtained by different algorithms for the Indian Pines dataset.
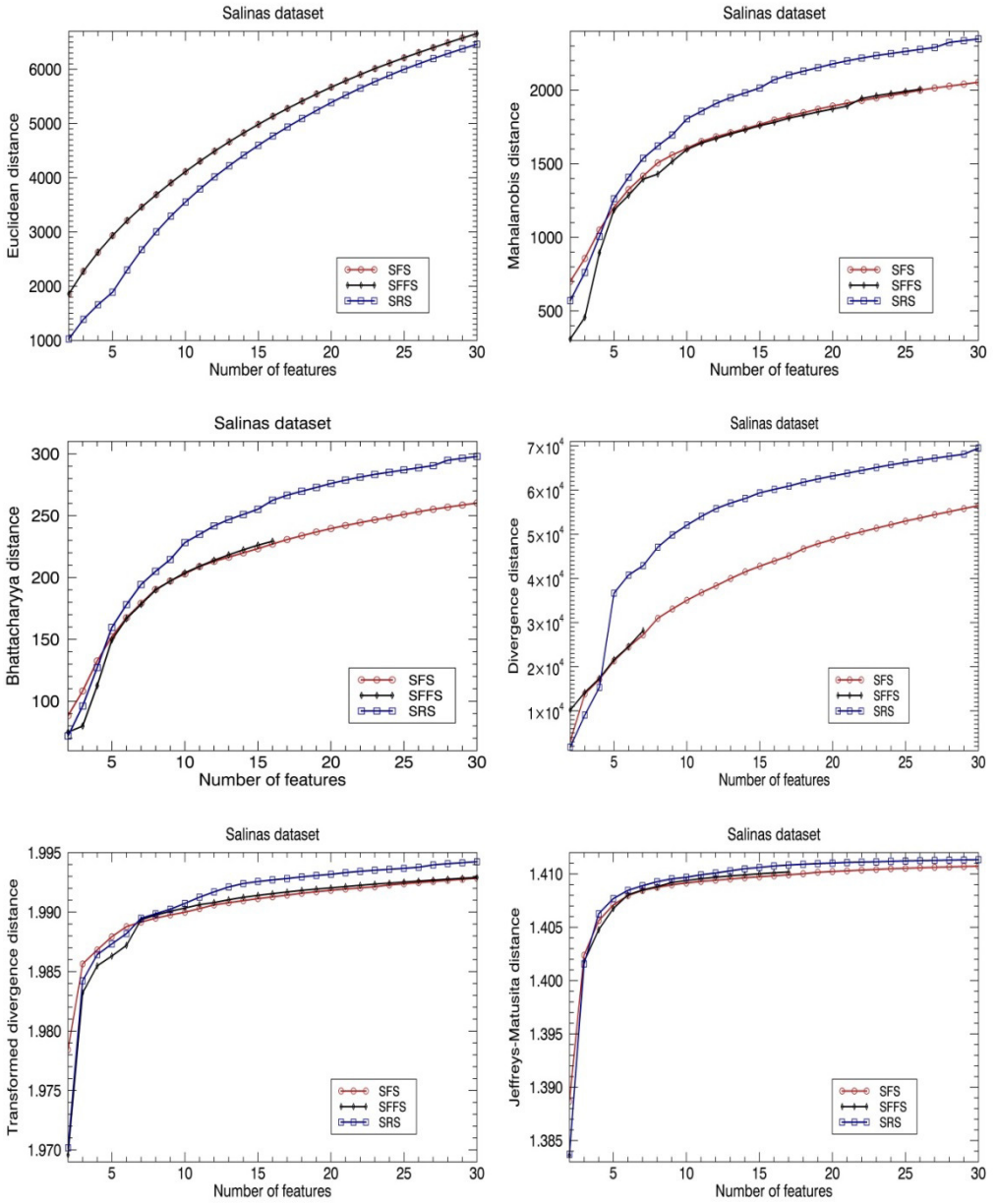
Figure 4.6. The mean separability of the features obtained by different algorithms for
the Salinas dataset.

It means that a new feature space is generated by SRS, but if based on the class mean only, it gives a worse discrimination. On the other hand, when other separability metrics are used, SRS achieves a better class discrimination than SFS and SFFS. These metrics consider both the class mean and the class variance, which contributes to improve the separability. Initial selections usually give better performance with individual channels than with wider spectral bands. The number of wider bands needed for higher separability than with spectral channels, is different from case to case, but at some points SRS always provides a better separability. For example, even with two bands SRS gave a better result than channel selection methods when Bhattacharyya or Divergence metrics were applied with the Indian Pines dataset. In the worst case, at least seven bands are needed to achieve a better SRS performance for the Salinas scene, when the D^t metric is applied.

Sometimes SFFS does not select the number of channels predefined by the termination point since SFFS remains in a local loop during the search and cannot determine the required number of features. It occurs during the "conditional exclusion" of the features already selected in the backward process of the algorithm (see reference (Pudil, Novovicova et al. 1994) for the details).

Table 4.4: A hypothetical example of steps in a run of the SFFS algorithm remaining in a local loop.

Step	Action	k	S_k	$J(S_k)$	$J(S_{k-1})$
1	Add feature 1	1	[1]	1	--
2	Add feature 2	2	[1,2]	3	1
3	Add feature 3	3	[1,2,3]	5	3
4	Add feature 4	4	[1,2,3,4]	7	5
5	Add feature 5	5	[1,2,3,4,5]	9	7
6	Remove feature 2	4	[1,3,4,5]	8	4
7	Add feature 6	5	[1,3,4,5,6]	10	8
8	Add feature 2	6	[1,3,4,5,6,2]	12	10
9	Remove feature 5	5	[1,3,4,6,2]	11	6
10	Remove feature 6	4	[1,3,4,2]	7	4
11	Remove feature 4	3	[1,3,2]	5	2
12	Remove feature 3	2	[1,2]	3	1

We developed a hypothetical example (Table 4.4) of the steps in a run of the SFFS algorithm, since in the practical cases of this experiment, the number of steps

is too large to reveal this issue. Let k be the number of features in the set S_k and $J(S_k)$ be the criterion function for the given set, then the example is as in Table 4.4.

At Step 2, the features are [1,2], which are again obtained ten steps later, at Step 12. When the iteration continues, the same loop will be repeated, and the other features of the original data set will not be selected. This situation is more likely to occur when the number of features is large. In these cases, we considered the maximum number of channels provided by SFFS.

Using the Euclidean distance, the result of SFS and SFFS were the same. Furthermore, SFFS does not always provide a higher separability than SFS (Spence and Sajda 1998), and the result with the two methods are almost the same, while the SRS results were better. Table 4.5 also gives more details about the spectral location of the channels and the splits selected by the two algorithms using the JM metric.

Table 4.5. The spectral locations (in nm) of the channels and splits selected by SFS-JM and SRS-JM respectively for the Indian Pines dataset in the order of selection.

SFS-JM	677, 976, 1641, 2113, 754, 528, 697, 1983, 716, 1091, 627, 783, 1293, 1512, 1770, 587, 2262, 1139, 706, 1760, 1993, 548, 667, 2093, 2202, 725, 745, 617, 947, 1283.
SRS-JM	697, 1148, 2003, 735, 687, 1571, 841, 657, 577, 1789, 1323, 1120, 1621, 1283, 1730, 1983, 1442, 985, 899, 617, 1462, 488, 1091, 2153, 2083, 2242, 2192, 956, 677.

4.5.3. Separability with wider spectral bands

The second experiment suggests that the class variance (intra-class distance) is the major determinant of the better separability achieved by SRS, since we just obtained the worst result by SRS when the Euclidean distance was used. We evaluated in detail the role of the class variance (intra-class distance) in determining class separability by analysing the variations of both intra- and inter-class distances. This evaluation was performed by averaging of all the intra- and inter-class distances for Indian Pines dataset using the features identified by SRS and SFS at each iteration. There are 13 intra-class (class variance) per feature and 78 inter-class distances per feature set. The number of features in a set is increasing in each iteration. Therefore, for a given feature set with k bands or channels, we calculated the mean value of k*13 intra-class and 78 inter-class distances. We consider the results of the two most comprehensive separability measures: JM and D^t (Figure 4.7).
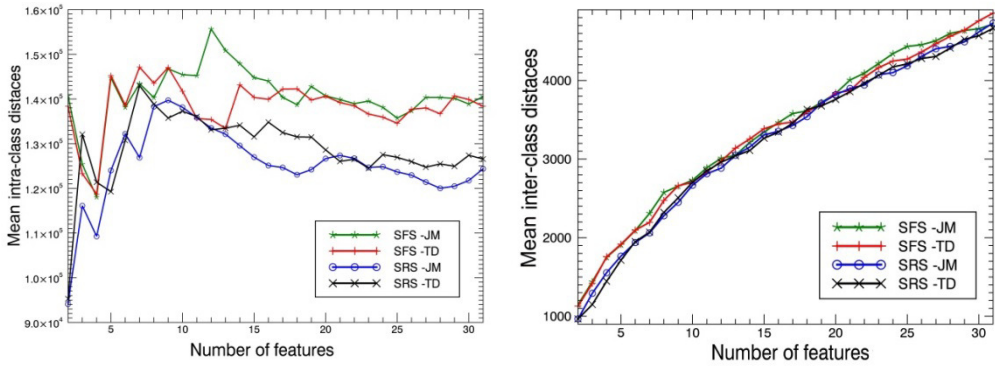
Figure 4.7. The mean intra-class (left panel) and inter-class (right panel) distance of 13 classes for the Indian Pines dataset with different number of features when Jeffreys-Matusita (JM) and transformed divergence (D^t) measures are applied with the SRS and SFS search algorithms.

As expected (see Section 4.3.2), the mean intra-class distance (class variance) obtained with the spectral bands is smaller than with individual channels in most cases, i.e. the mean intra-class distance becomes smaller when the channel signals are averaged. The difference increases with the number of bands. Analysing the inter-class distances (Figure 4.7 right) reveals that none of the methods performs better than the other ones, since each method sometimes gives a higher mean inter-class distance (distance between the class centroids).

From the three experiments, it can be concluded that a bandset created by SRS with the proper number of bands gives better separability than any other channel selection method for a given class set, and this is mainly due to the reduction of the intra-class distances, because of averaging the channel signals. The SRS separability is lower in the initial iterations, because the initial SRS bands are usually very broad. In this case the class variances of the constitutive channels are not usually equal, and an individual channel having a small class variance may yield a better class separability. For example, a vegetation class has a larger NIR than VIS variance, so a narrow VIS channel is likely to give a smaller variance than the average of all the variances across the spectrum. After the first iterations, however, the newly formed spectral bands give comparable variances and yield a better separability of the classes in a scene. As an example, the sequence of the first 20 splits selected by SRS using JM on the Indian Pines scene is: 702, 1115, 2048, 1596, 721, 856, 642, 1517, 1794, 582, 990, 692, 523, 2158, 740, 894, 1685, 1998, 2257 and 493 nm.

4.5.4. Classification accuracy

Finally, we applied two classifiers to the selected bands and channels, and computed the accuracy of classification to evaluate and compare performance. Although the classification accuracy depends on several factors, the separability of classes usually has a significant impact on the final result of the classification. In this experiment, Maximum Likelihood Classifier (MLC) and Support Vector Machine (SVM) were applied to the final channels and bands identified by SFS, SFFS, and SRS using the JM and D^t metrics. The classification accuracies of the two datasets were averaged and illustrated in Figure 4.8 for the two scenes. able 4.6 also gives the accuracy of six instances of feature sets determined by SFS-JM, SRS-JM.
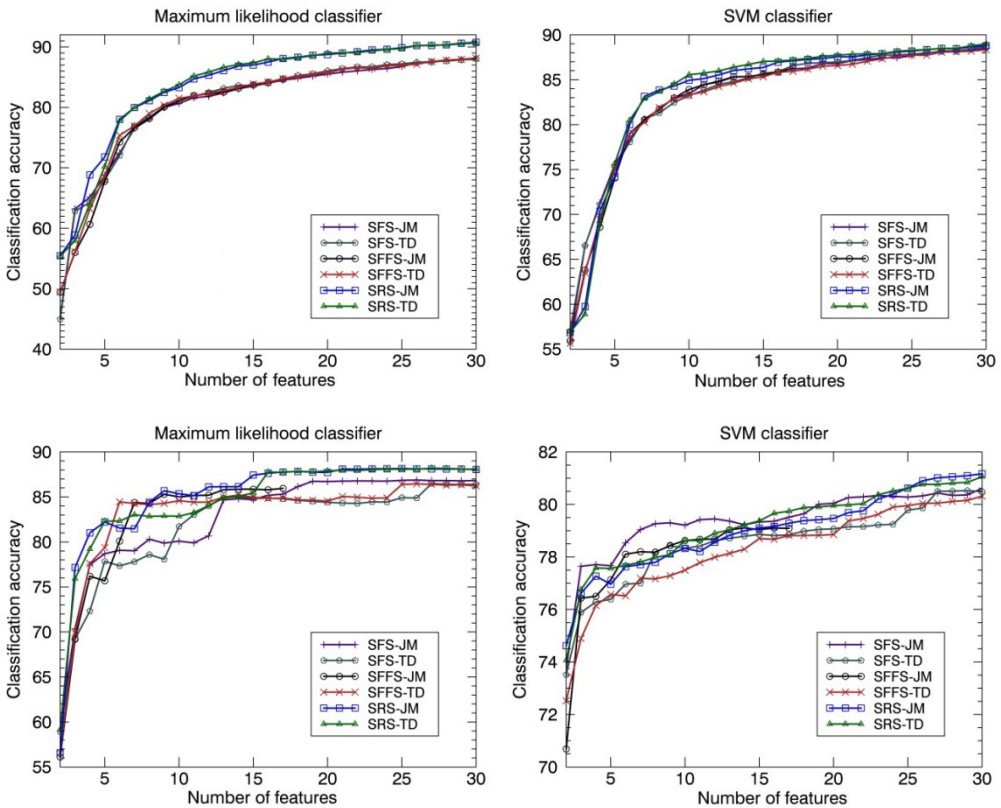
Figure 4.8. The classification accuracy obtained with the classifiers MLC and SVM applied to the bands and channels identified by SRS, SFS, and SFFS and applying the Jeffreys-Matusita and Transformed Divergence separability metrics ; Indian Pines (upper panel) and Salinas (lower panel) scenes.

Table 4.6 The classification accuracy of two DR methods (SFS-JM and SRS-JM) based on the filter approach and one method (SVM-RFE) based on the wrapper approach.

Dataset		Indian Pines					Salinas				
Classifier		MLC		SVM			MLC		SVM		
Method		SFS-JM	SRS-JM	SFS-JM	SRS-JM	SVM-RFE	SFS-JM	SRS-JM	SFS-JM	SRS-JM	SVM-RFE
Number of features	5	68.73	71.78	75.53	74.18	43.01	78.69	82.22	77.67	76.96	73.38
	10	80.65	83.27	83.43	84.94	72.22	80.09	85.35	79.21	78.33	81.28
	15	83.53	86.99	83.43	84.94	76.70	84.71	87.42	79.33	79.10	83.16
	20	85.50	88.72	86.80	87.55	81.21	86.71	87.72	80.04	79.47	87.03
	25	86.76	89.86	87.68	88.19	84.21	86.85	88.15	80.28	80.62	88.05
	30	87.95	90.76	88.41	88.78	87.76	86.82	88.04	80.61	81.17	88.17

The classification accuracy obtained with the spectral bands identified by SRS is higher than the accuracy obtained with the channel selection methods. When the number of features is small, SFS and SFFS may give better results. On the other hand, with maximum four spectral features, SRS gives higher accuracy than either SFS or SFFS for the Indiana Pines scene when MLC was used as the classifier. For the Salinas scene, there are larger fluctuations in the classification accuracy, i.e. sometimes the channel selection methods gave better results when the number of bands is low, while with higher number of bands SRS had better accuracy. MLC gave a larger improvement in classification accuracy with the bandset created by SRS than with SVM. On average, the improvements are about 3.24% and 0.96% when MLC and SVM were used respectively. This is due to the reduction in class variances having a larger impact on a parametric classifier like MLC that considers the distributions of class attributes explicitly.

4.5.5. Comparison with a wrapper approach

The last results revealed that the band configurations identified by SRS gave a noticeable improvement in classification accuracy compared with the channel selected by SFS and SFFS, especially when the MLC was used. Although our focus was on feature selection based on the filter approach, we compared the results of classification with an algorithm based on the wrapper models (Kohavi and John 1997; Sebban and Nock 2002; Lodha and Kamlapur 2014) as well. In the wrapper models, features (channels) are usually selected given the classifier based on a the accuracy of classification. It is usually assumed that the wrapper approaches achieve better classification accuracy than the filter approaches since they identify features that better suit the classification algorithm regarding the performance

(Guyon, Weston et al. 2002; Bazi and Melgani 2006; Kuo, Ho et al. 2014). The wrapper approaches, on the other hand, are computationally more expensive than the filter approaches (Santos, de S Celes et al. 2012; Sarrhrouni, Hammouch et al. 2012). We examined a well-known wrapper algorithm for channel selection on the basis of classification accuracy in comparison with a method giving a better classification results in our experiments: SRS-JM combined with MLC. Infact, SRS-JM identifies features based on maximising class separability, to be applied with the classification algorithm, while a wrapper algorithm selects features to maximise classification accuracy.

We used SVM-RFE (recursive feature elimination) as a wrapper algorithm (Guyon, Weston et al. 2002; Pal 2006; Zhang and Ma 2009; Huang, Hung et al. 2014; Kuo, Ho et al. 2014) in this study. In RFE, the decision function of the SVM, i.e. finding an optimal hyperplane that maximizes the marginal distance between two classes, is used as the criterion to select features in a backward elimination approach. It computes ranking weights based on the training samples for all the features and sorts the features according to weight vectors (Guyon, Weston et al. 2002; Huang, Hung et al. 2014).

For the SVM-RFE algorithm implemented in this experiment, we considered the commonly used "one-against-all' strategy (Bazi and Melgani 2006; Huang, Hung et al. 2014). The "One-against-all" strategy converts the problem of k classes ($k>2$) into k dual-class problems. The radial basis function (RBF) (Pal 2006; Zhang and Ma 2009; Kuo, Ho et al. 2014) is utilised as the kernel function to map the original feature space into the higher dimensional space. For the kernel width factor (γ) and regularization parameter (C), we applied different values suggested in literature and chosen the ones giving the best classification performance, i.e. γ =0.1 and C=2000 proposed by

In this experiment, we investigated which combination of features and classifiers would give a higher classification accuracy. So, the accuracy achieved with the spectral configuration constructed by SRS-JM, combined with MLC, was compared with the accuracy obtained by SVM-RFE for each dataset (Figure 4.9). Table 4.6 also gives the accuracy obtained with six feature sets determined by SVM-RFE in comparison with SRS-JM.
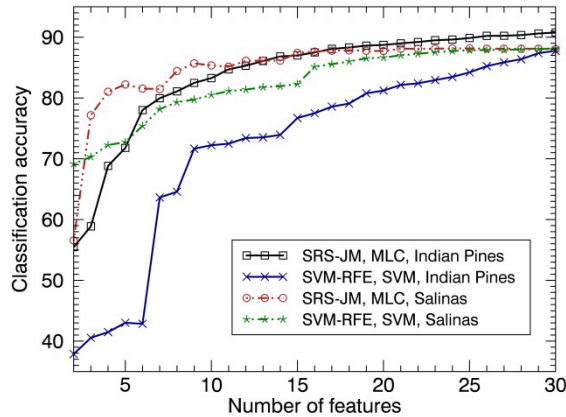
Figure 4.9. The classification accuracy obtained by a filter approach i.e. features identified by SRS-JM and the classification with MLC, and one wrapper approach (SVM-RFE) for two datasets. The legend gives in each row the feature identification algorithm, the classification method, and the dataset used respectively.

It was observed in this experiment that the combination of the features identified by SRS-JM and MLC gave a better or comparable classification accuracy than the wrapper method. The SVM-RFE combination starts with classification accuracy of about 70% for Salinas scene and increases gradually, and reaches the accuracy of SRS-JM with MLC when the number of features is about 25. For the Indian Pines dataset, SVM-RFE had a lower accuracy than SRS-JM in all cases, while the difference became less by increasing the number of features. The SRS-JM-MLC gave about 3% higher classification accuracy with 30 features.

In comparison with the SVM-RFE algorithm, the spectral configuration determined by SRS gave a better or comparable classification accuracy when MLC was used as the classifier and the number of features was less than 30.

4.5.6. Discussion

We compared SRS with three search strategies used in channel selection: branch and bound, sequential forward selection, and sequential forward floating selection, by applying different separability metrics: Euclidean, Mahalanobis, Bhattacharyya, Divergence, Transformed Divergence, and Jeffreys-Matusita.

In the first experiment, the comparison between SRS and BB indicated that SRS gives better results than the best-selected channels over small data sets. Then, SRS was compared with widely used algorithms in channel selection, i.e. SFS and

SFFS, using different separability metrics over complete hyperspectral datasets. The second experiment had a very similar outcome, i.e. better class discrimination by SRS with a higher number of bands. On average, at least four spectral bands are needed to have better separability than by selecting narrow channels.

The third experiment analysed the effect of the broader spectral regions formed by SRS on two main factors of class discrimination: inter-class and intra-class distances. The results of this experiment and the second one revealed that the main reason why SRS provides a better class separability is the reduction of intra-class distances due to the broader spectral bands identified by SRS. Having a better separability between the classes in a scene leads to a higher classification accuracy in a filter-based approach, and finally, a better identification of observed targets, as shown by the fourth experiment. Comparison with a wrapper approach, in the fifth experiment, revealed that the combination of SRS and MLC gave a better or comparable accuracy of classification.

4.6. Conclusion:

The approach we propose demonstrates the importance of averaging narrow channels in improving the class separability by utilizing a new dimensionality reduction method of hyperspectral data, which identifies spectral regions with the aim of optimal class discrimination. We have shown that the algorithm, i.e. the Spectral Region Splitting (SRS), applied with a class separability metric can provide a bandset with a better class separability than the best channels selected from the original dataset with the same criterion. The reason is that SRS does not preserve the width of the original spectral channels, unlike the channel selection methods, and defines a new feature space by merging the adjacent spectral channels if it is necessary. Averaging the narrow adjacent channels results in smaller intra-class distance by reducing the class variances leading to an increase in class separability. We concluded that whenever the separability measures include the intra-class parameters, SRS provides a better class discrimination when there is an adequate number of bands. The experiments were implemented on two different hyperspectral datasets including various types of classes. Eventually, the scenes were classified applying the selected bands showing that SRS increased the accuracy of classification by about 2%.

5.
Comparative evaluation of spectral configurations

This chapter presents a comparative assessment of the bandsets obtained with the methods described in the previous chapters and designed to meet the three objectives articulated in the Introduction. There are three different comparisons were performed for the assessment using two different hyperspectral datasets. These comparisons are based on the spectral configurations determined, the running performance of the algorithms, and the image classification accuracy. After each comparison, a discussion about the results is presented.

5.1. Introduction

The main scope of this thesis was to develop and evaluate methods to reduce the dimensionality of hyperspectral images by extracting the necessary spectral information. For this purpose, we identified three objectives: spectral representation, independent spectral bands, and spectral bands for optimal class separability (Section 1.5.1). The implementation of the methods to achieve these three objectives has been presented in detail in the Chapters 2, 3 and 4, along with numerical experiments to evaluate alternate solutions.

In this chapter, we present a comparative evaluation of the spectral configurations determined to meet each objective.

This assessment is based on three different comparisons. First, the spectral configurations determined for each separate objective were compared with each other on the basis of the positions of the splits dividing the full spectrum into a set of spectral bands. This evaluation shows how different the spectral configurations are to meet each one of the three identified objectives. In addition, we compared the computational load to apply different algorithms. Finally, we compared the accuracy of scene classification using the different spectral configurations, i.e. the ones obtained to meet each objective. The last comparison yields an assessment of the relevance of determining objective–specific spectral configurations. Image classification is a frequently used application to evaluate reduced spectral configurations. The three comparisons are presented in Section 5.2, 5.3, and 5.4 respectively.

We conducted all the experiments on the two hyperspectral scenes having ground truth data: Indian Pines and Salinas AVIRIS scenes (Section 4.4). As already mentioned, the Indian Pines dataset has 178 channels, while the landcover was mapped into 13 classes and comprises 21025 pixels (Figure 4.1 and Table 4.1). The Salinas scene has 190 channels, while the landcover was mapped into 16 classes and comprises 111104 pixels (Figure 4.2 and Table) 4.2).

The band configurations obtained based on the two first objectives can be determined either in a generic (unsupervised) way or taking into account the spectral features of the available classes in the scene (supervised). In this assessment, both supervised and unsupervised band sets were produced and evaluated. As mentioned in Introduction, supervised DR techniques are used when

there are some regions of interest, such as classes, in a scene so that a band configuration may be obtained based on these regions or for a specific class. In this comparison, we will assess how different the band configurations are if they are obtained based on a portion of the scene or based on the entire scene.

5.2. Comparison of spectral configurations.

In the first comparison, we determined different spectral configurations for each objective and compared them on the basis of the position of the splits. We applied supervised and unsupervised dimensionality reduction whenever possible. It should be noted that in each experiment, one spectral band configuration is finally achieved for all the pixels in the datasets. For the supervised situation, we used almost half of the pixels having known land cover of each scene. To obtain the representative and independent band configurations, we applied the algorithms to the selected pixels together. As mentioned, in this situation we would achieve a band configuration derived from a portion of a given scene where is of interest to an analyser. We selected the samples from the known classes in order to compare different experiments and to better interpret the final results. The "known pixels" are specified by the ground truth maps. The details about the available classes per scene and their locations were presented in Section 4.4.

In each case, 20 spectral bands are identified, the details of the split locations are tabulated, the 20-band configurations are illustrated, and the results are discussed in each section.

5.2.1. Spectral representation

The spectral region splitting (SRS) method was used to obtain the spectral configurations to represent the spectra of the datasets accurately (see Chapter 2). In Chapter 2, we compared the SRS results with the best band configurations with the minimum error of the representation obtained by PCFA and showed that SRS represents the spectra with comparable representation accuracy with the PCFA representation, but it requires a much shorter computation time. In this experiment, we used RMSE as the metric of the quality of the representation. Table 5.1 and 5.2 give the spectral locations of the splits obtained by SRS and the sequence of selection for the Indian Pines and Salinas scene respectively. From the tables, the spectral configurations with a predefined number of bands (less than 21) can be derived.

The spectral configurations with 20 bands are illustrated as well in Figure 5.1 for both the supervised and unsupervised case. In the figures, we also illustrated μ, and μ±σ per channel of the original datasets and per band of 20-band dataset shown by dashed and continuous lines respectively, where μ is the mean and σ is the standard deviation in each spectral feature.

Table 5.1: The spectral configurations of representative bands for the Indian Pines scene.

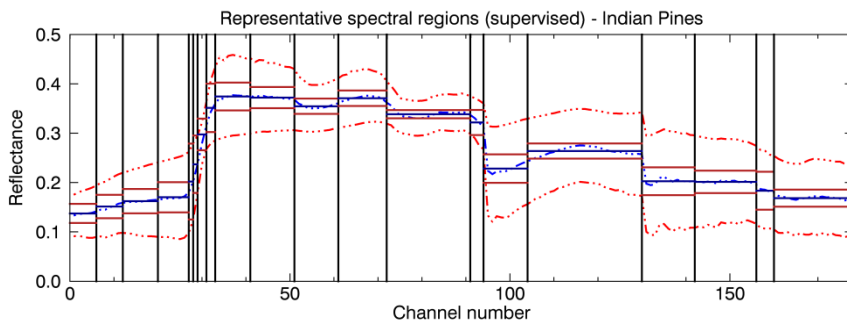| No. of splits	Supervised			Unsupervised		
	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)
1	28	706	20	94	1343	39
2	94	1343	36	29	716	62
3	130	1789	48	130	1789	82
4	72	1129	61	72	1129	99
5	156	2242	72	156	2242	114
6	31	735	79	31	735	129
7	12	548	87	105	1542	143
8	104	1532	95	13	557	157
9	51	928	102	51	928	171
10	29	716	109	27	697	184
11	20	627	116	91	1313	198
12	6	488	123	61	1024	211
13	142	2083	130	8	508	224
14	61	1024	137	142	2083	237
15	91	1313	144	33	754	251
16	160	2282	152	30	725	264
17	27	697	159	100	1492	277
18	41	831	166	160	2282	290
19	33	754	172	28	706	304

To calculate μ, and σ of a spectral feature, in the unsupervised case, all pixels in the image are used, and in the supervised one, the pixels located in the portions of the image are used, where ground truth is available. As discussed in Section 4.3.2, the standard deviation of broader spectral regions created by averaging the adjacent narrow channels was reduced noticeably, which is demonstrated in the figures. These illustrations give a general view of the spectral reflectance in original datasets which is split into different spectral regions by applying the objective.

On the one hand, the spectral representations of two scenes are rather different. Although the majority portions of both datasets are covered by plants, the vegetation types are different (see Table 4.1 and Table 4.2). It reveals that the spectral band representation obtained by SRS is highly dependent on the land cover types of a scene. On the other hand, it can be observed for both datasets that the spectral regions identified in the supervised and unsupervised case had some similarity. For example, there are several narrow bands in the red spectral region for both datasets, giving the idea that the red and red edge spectral regions are essential in investigating the entire scene or some parts of the scene covering vegetation areas.

Table 5.2: The spectral configurations of representative bands for the Salinas scene.

No. of splits	Supervised			Unsupervised		
	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)
1	98	1345	78	31	702	221
2	32	712	126	98	1345	399
3	138	1813	168	138	1813	517
4	76	1134	205	75	1125	623
5	171	2269	237	169	2249	733
6	16	557	268	16	557	822
7	35	741	300	34	731	909
8	110	1534	333	54	923	997
9	54	923	364	110	1534	1085
10	30	693	394	9	488	1167
11	42	808	424	158	2139	1248
12	10	497	454	24	636	1328
13	95	1315	484	42	808	1408
14	155	2109	517	93	1295	1489
15	33	721	550	174	2299	1569
16	64	1019	580	32	712	1650
17	23	626	610	65	1029	1727
18	114	1574	639	144	1999	1805
19	128	1713	668	77	1144	1883

The reason of the similarity in the spectral configurations for the supervised and unsupervised cases can be, firstly, that the pixels selected in supervised cases cover about half of the scenes, not a small portion (Figure 4.1 and 4.2). Secondly, a visual inspection of the spectral reflectance of the pixels that are not used in the supervised experiments suggests that their reflectance characteristics were almost the same as the pixels used. As noticed in Figure 5.1, μ and σ for the two datasets in both situations (supervised and unsupervised) are very similar, which may not be even distinguishable; showing that the general spectral behaviour of pixels used in the supervised and unsupervised cases are almost the same in these experiments.
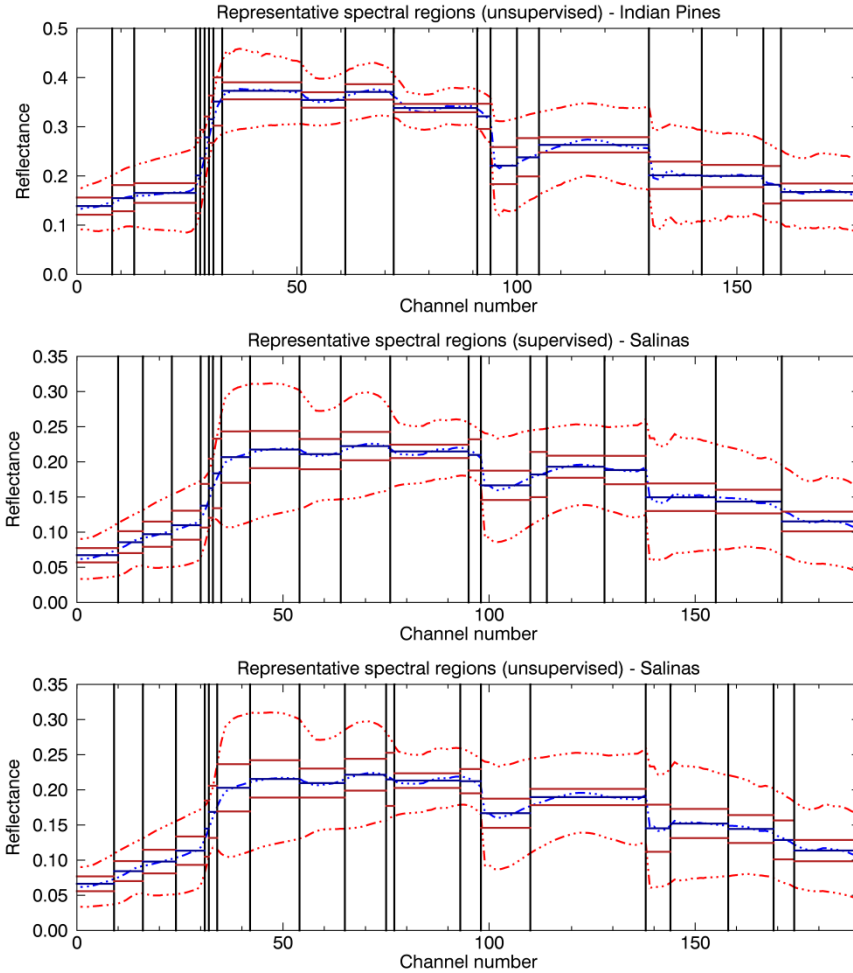
Figure 5.1: 20-band configurations obtained by SRS for spectral representation per scene in the supervised and unsupervised cases. The vertical black lines indicate the location of the splits. The blue and red lines illustrate μ and $\mu \pm \sigma$ of the spectral reflectance per channel in the original datasets (dashed lines), and per band in the final dataset (solid lines), where μ is the mean and σ is the standard deviation of all the pixels used in the experiments.

When looking at the figure and the tables in the supervised and unsupervised cases more accurately, dissimilarities appear. For example, the locations of several splits were in the same spectral regions while the order of the selection might be different. E.g., the first and the second splits were at the start of the NIR region and around 1340nm in the supervised and unsupervised case in both datasets, but the order of the selection was not the same (Table 5.1 and Table 5.2). The dissimilarity of the split locations gradually increased with the iterations. The number of exactly

similar bands for the supervised and unsupervised cases for the two finally datasets was 11 and 2 out of 20 identified bands of the Indian Pines and Salinas dataset respectively.

5.2.2. Independent spectral bands

To obtain spectral configurations with independent bands we applied the clustering-merging (CM) algorithm using the total dependence metric (see Chapter 3). These experiments were also performed for the supervised and unsupervised cases. Table 5.3 and Table 5.4 give the split locations in each iteration for the two datasets, and Figure 5.2 illustrates the 20-band spectral configurations for each case.
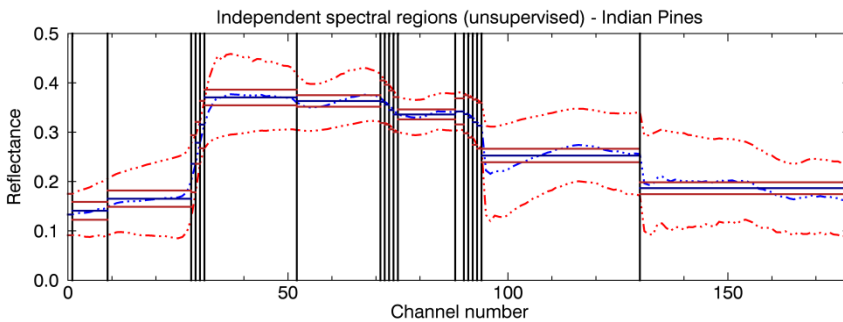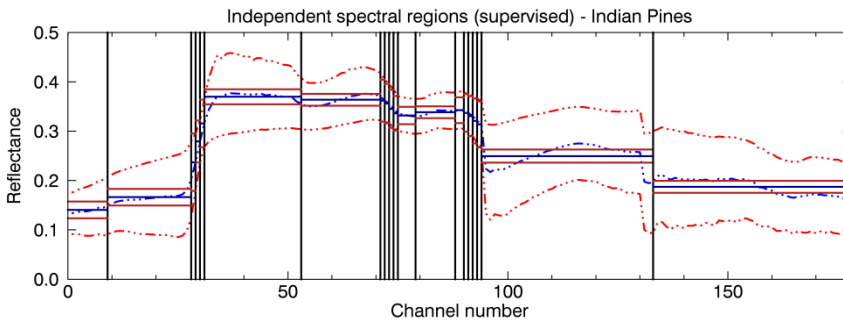
Table 5.3: The spectral configurations of independent spectral bands for the Indian Pines scene.

No. of split	Supervised			Unsupervised		
	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)
1	30	725	0.3	30	725	0.4
2	73	1139	0.6	73	1139	0.6
3	90	1303	0.9	90	1303	0.9
4	29	716	1.1	29	716	1.1
5	53	947	1.3	52	937	1.4
6	92	1323	1.5	92	1323	1.6
7	74	1148	1.7	74	1148	1.8
8	72	1129	1.9	28	706	2.0
9	93	1333	2.1	72	1129	2.2
10	31	735	2.3	93	1333	2.5
11	28	706	2.5	31	735	2.7
12	133	1993	2.7	130	1789	2.9
13	88	1283	2.9	9	518	3.0
14	91	1313	3.1	88	1283	3.2
15	9	518	3.2	71	1120	3.4
16	71	1120	3.4	91	1313	3.6
17	75	1158	3.6	75	1158	3.8
18	94	1343	3.8	94	1343	4.0
19	79	1196	3.9	1	439	4.2

In this experiment, there were splits located in the same place in the supervised and unsupervised cases for both datasets as well. There were more splits in the same spectral locations than in the previous experiment to identify the representative bands. For the Indian Pines dataset, 13 out of 20 bands and for the Salinas dataset 12 out of 20 bands were exactly identical in both cases.

Table 5.4: The spectral configurations of independent spectral bands for the Salinas scene.

No. of split	Supervised			Unsupervised		
	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)
1	32	712	0.4	32	712	0.3
2	96	1325	0.6	96	1325	0.6
3	77	1144	0.8	77	1144	0.8
4	17	567	1.0	31	702	1.0
5	98	1345	1.2	98	1345	1.2
6	31	702	1.4	17	567	1.3
7	33	721	1.6	33	721	1.5
8	95	1315	1.7	137	1803	1.7
9	138	1813	1.9	12	517	1.8
10	55	933	2.0	56	942	2.0
11	6	458	2.2	95	1315	2.1
12	97	1335	2.3	3	428	2.2
13	94	1305	2.5	30	693	2.4
14	30	693	2.6	111	1544	2.5
15	115	1584	2.7	97	1335	2.6
16	34	731	2.9	94	1305	2.8
17	76	1134	3.0	100	1435	2.9
18	1	409	3.1	34	731	3.0
19	137	1803	3.2	138	1813	3.1

Independent spectral regions (supervised) - Indian Pines

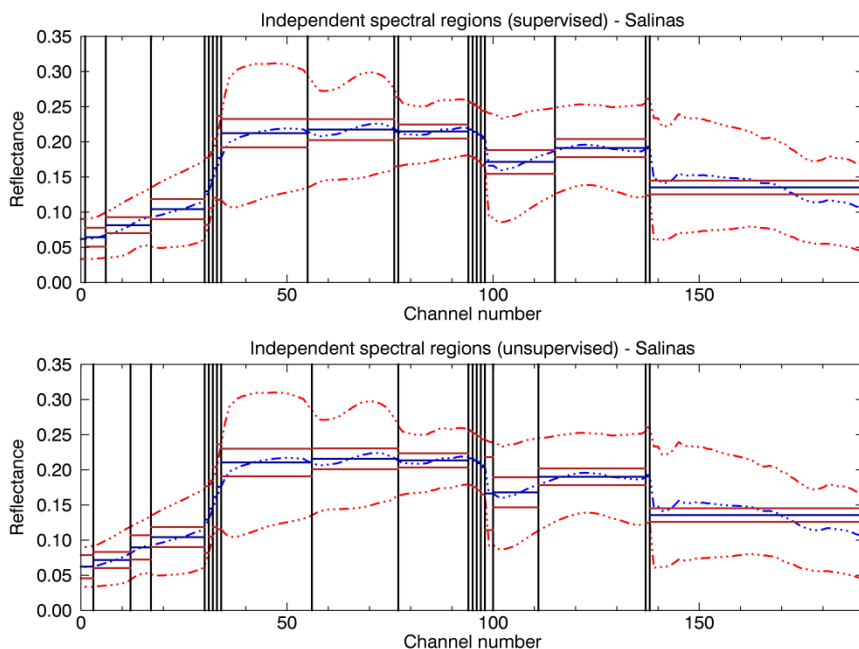Independent spectral regions (unsupervised) - Indian Pines

Figure 5.2: 20-band configurations obtained by CM for independent spectral regions per scenes in the supervised and unsupervised cases. The vertical black lines indicate the location of the splits. The blue and red lines illustrate μ and μ±σ of the spectral reflectance per channel in the original datasets (dashed lines), and per band in the final dataset (solid lines), where μ is the mean and σ is the standard deviation of all the pixels used in the experiments.

There were several narrow adjacent bands identified in all cases. These narrow spectral regions can be categorized into two parts. First, they were adjacent to the water absorption spectral regions that had been already removed from the original datasets, e.g. around 1330nm. It seems that those selected bands were still contaminated with noise due to the presence of carbon dioxide, water and oxygen in the atmosphere, which led to the identification of them as narrow independent bands by the algorithm. Second, the red edge spectral region also contained narrow bands, revealing that the narrow channels in this region have more uncorrelated information and should be kept as independent spectral bands.

5.2.3. Spectral regions for class separability

We applied the SRS algorithm to the datasets to identify the spectral regions being most useful for class separability (see Chapter 4). We used the JM distance as the criterion in SRS. The results of the split selection of both datasets were tabulated in Table 5.5, and Figure 5.3 showes 20-band configurations of so-called

separable bands. There is no very narrow band in these band configurations, except the red edge although the visible spectral region has been divided into narrower spectral regions. The overall comparison of different band configuration is given in next section.

Table 5.5: The spectral band configurations that maximize the class separability using JM for the two datasets.

| No. of split | Indian pines | | | Salinas | | |
	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)	Location of split (channel #)	Location of split (wavelength (nm))	Time (sec.)
1	27	697	6	100	1435	21
2	70	1110	11	54	923	35
3	138	2043	14	48	865	46
4	110	1591	16	58	961	58
5	29	716	19	136	1793	71
6	43	851	22	66	1038	82
7	21	637	24	118	1614	94
8	102	1512	27	24	636	106
9	130	1789	30	78	1154	118
10	15	577	33	72	1096	131
11	57	985	37	33	721	145
12	26	687	40	29	683	159
13	9	518	44	36	750	174
14	149	2153	49	11	507	191
15	31	735	53	17	567	208
16	47	889	58	92	1285	226
17	119	1680	63	4	438	245
18	133	1993	69	124	1674	265
19	157	2252	75	32	712	286

5.2.4. Overall comparison

In this section, we compared the spectral regions identified to meet the three objectives discussed in the previous sections. The band configurations obtained in three different experiments were compared in relation to the spectral representation, the independence between the bands, and the separability between classes. For the first and second experiments (Sec. 5.2.1 and 0), we used all the pixels of the scenes (unsupervised) as well as a part of the pixels of the scenes (supervised) and showed the differences regarding the spectral locations of the splits. The third experiment (Sec. 5.2.3) was performed in the supervised case only.
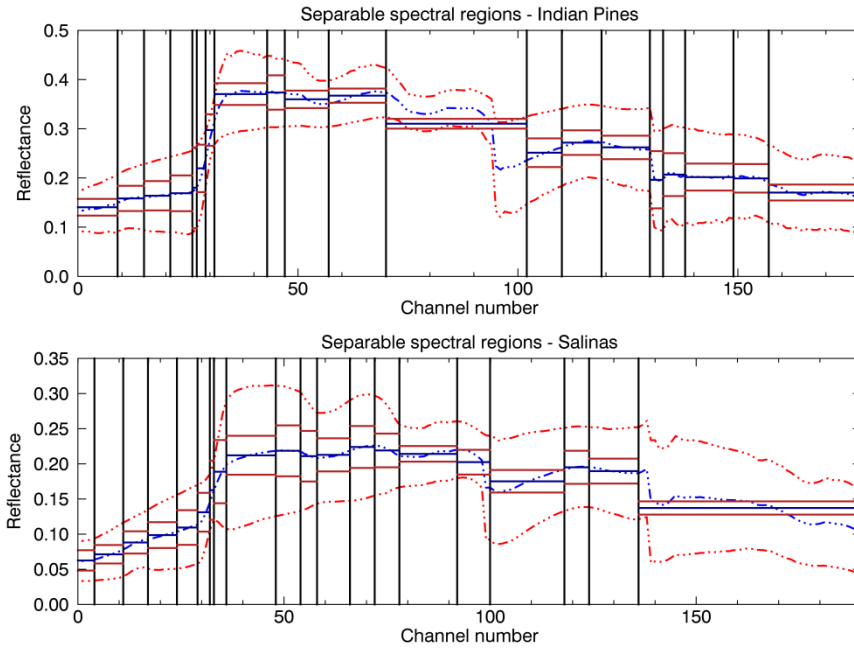
Figure 5.3: 20-band configurations obtained by SRS for the spectral regions maximizing class separability per scene. The vertical black lines indicate the location of the splits. The blue and red lines illustrate μ and μ±σ of the spectral reflectance per channel in the original datasets (dashed lines), and per band in the final dataset (solid lines), where μ is the mean and σ is the standard deviation of all the pixels used in the experiments.

The band configurations in these three experiments were obtained by employing the metrics specified and the algorithm developed per objective (Chapter 2, 3, and 4) as follows:

1-   The spectral band representations were determined by applying the SRS algorithm to the spectra and minimising the RMSE between the approximated spectra and the original ones.

2-   The independent bands were obtained using the Total dependence (TD) metric applied to CM algorithm to identify the spectral regions with the lowest correlation.

3-   The separable bandset was identified by applying the Jeffreys–Matusita separability distance plugged into the SRS algorithm to find the spectral regions maximising the discrimination between the classes in the scenes.

Now, in this section, we evaluated the spectral region configurations obtained to meet the three objectives in relation with each other. For example, how

separable the independent bands or representative bands are; or how much the accuracy of representation is if we use the separable bands. These comparisons would help us to answer the research question mentioned in the Introduction about the relevance of having objective-specific band configurations.

Table 5.6 presents the details of the comparison. In this table, the RMSE and TD metrics are computed using entire hyperspectral scenes, and the JM distance is calculated based on the mean of the class separabilities.

Table 5.6: Overall comparison of different band configurations obtained to meet the three objectives.
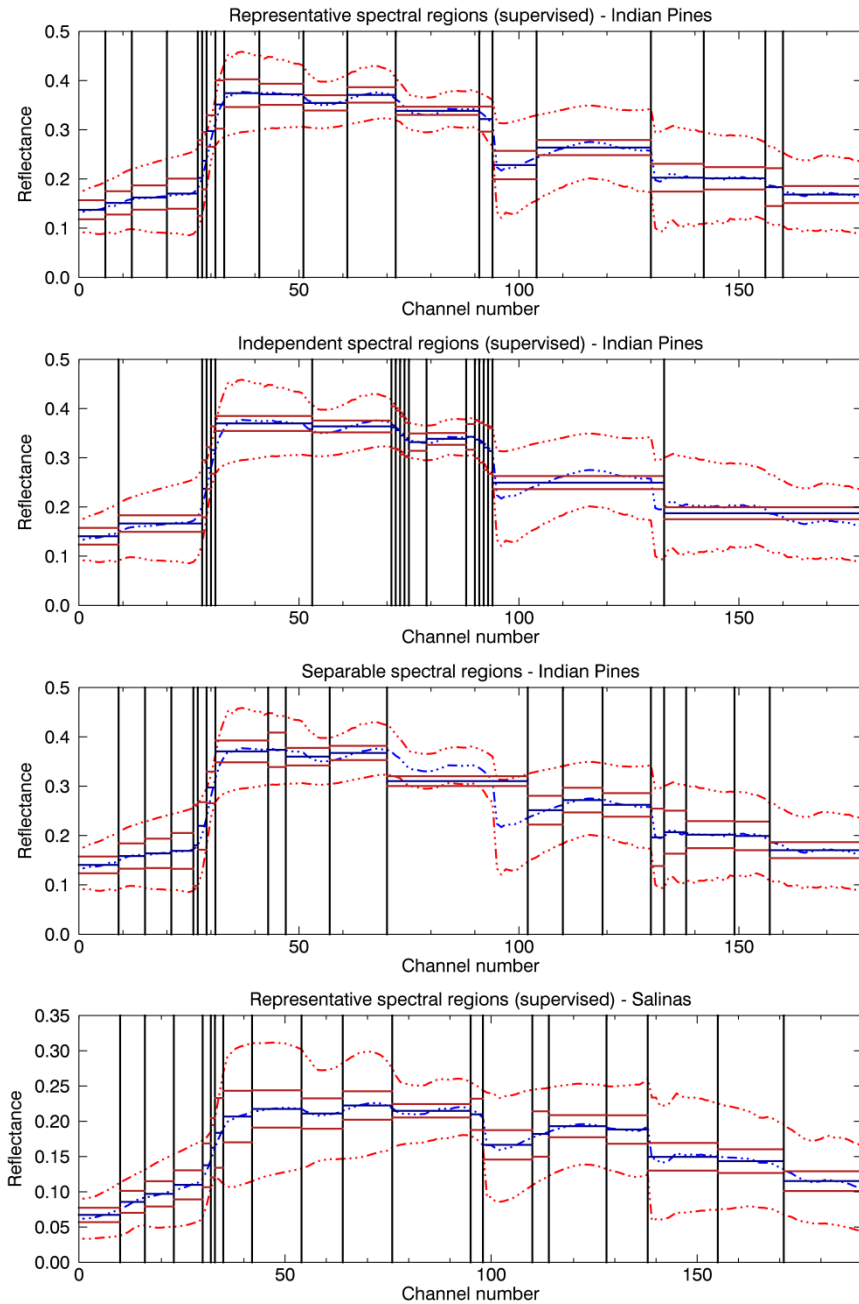
20-band sets	Indian Pines			Salinas scene		
	RMSE	TD	JM	RMSE	TD	JM
Representative bands-supervised	0.00826	0.6786	1.3959	0.00632	0.6258	1.4104
Representative bands-unsupervised	0.00814	0.6749	1.3950	0.00627	0.6262	1.4105
Independent bands-supervised	0.01740	0.6064	1.3866	0.01208	0.6211	1.4098
Independent bands-unsupervised	0.01570	0.5911	1.3854	0.01190	0.6143	1.4100
Separable bands	0.02415	0.7568	1.4010	0.01488	0.6268	1.4110

As observed in both cases, the minimum error of representation (RMSE), the minimum dependence (TD) and the maximum class separability (JM) were obtained by the representative bands-unsupervised, independent bands-unsupervised, and by separable bands for both datasets respectively. For instance, the minimum RMSE and TD are 0.0081 and 0.59, and the maximum JM is 1.40 for the Indian Pines dataset, which apply to the objective - specific cases. This result is in agreement with the design and motivation of this research, i.e., in each case, we achieved the best result with the best configuration obtained to meet each objective.

The difference between RMSE of various band configurations revealed that the representative bands, either supervised or unsupervised, had almost half representation error than the other band sets, e.g. RMSE was 0.006 and 0.012 for the representative and independent bands of the Salinas scene respectively. TD showed that the overall dependence was the least for the independent bands, then representative bands and finally the separable bands, while the TD values were closer to each other for the Salinas scene than Indian Pines. It can also be observed from the experiments on both datasets that the mean class separability of representative bands was higher than for the independent band configuration.

Figure 5.4 is a collection of the 20-band configurations determined to meet the three objectives for both datasets. We used the band configurations determined in the supervised mode so that the pixels used in each evaluation are the same. The

comparison between the supervised and unsupervised cases per objective are given in Figure 5.1, Figure 5.2, and Figure 5.3.
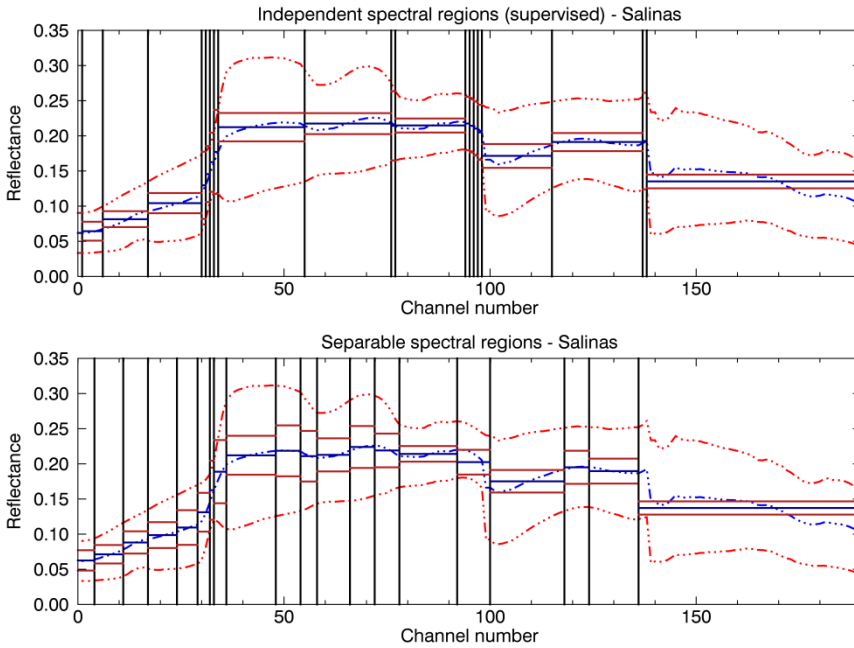
Figure 5.4: 20-band spectral configurations with respect to the three objectives for two datasets: Indian Pines and Salinas. The vertical black lines indicate the location of the splits. The blue and red lines illustrate μ and μ±σ of the spectral reflectance per channel in the original datasets (dashed lines), and per band in the final dataset (solid lines), where μ is the mean and σ is the standard deviation of all the pixels used in the experiments.

In general, there was a slight similarity in the spectral configurations determined for the three objectives. Although there were splits located at the same spectral locations, adjacent splits of them were mostly placed at different locations creating different spectral regions. In the Salinas scene for example, there were splits exactly at the same spectral location, 1813nm, identified in both the representative and the independent bandset, but the adjacent bands created by this split were different, i.e. 1713-1813nm and 1813-2109nm identified as the representative bands, and 1803-1813nm and 1813-2460nm identified as independent ones.

The first and the third objectives led to more similar spectral configurations, while the ones determined for the second objective were rather different. It appears that the spectral configuration determined to represent full spectra with fewer bands is more sensitive to spectral gradients to minimise the error of the spectral representation. Independent bands consider the correlation of samples in a band with other bands, and random noise increases the independence of spectral bands, thus leading to selection. The spectral configuration to maximise separability

depends on land cover and pre–defined classes, so is likely to differ from the previous two (Figure 5.4).

A fine spectral sampling of the red – edge region, i.e. from 680 nm to 730 nm, was determined when pursuing all three objectives. At least two narrow bands in this spectral region were identified in all spectral configurations and for both datasets. This region was necessary to represent the full spectrum with fewer bands (Objective 1) due to the large spectral gradient in this region. There is also less correlation between the adjacent spectral regions within the red – edge region (Objective 2), so the region was divided into narrow independent bands as well. It is also known that the red edge is important for recognising different types of vegetation classes. Both datasets have several vegetation classes. Therefore, narrow spectral bands were determined when pursuing class separability (Objective 3).

5.3. Comparison of execution time

The details of the execution time of the methodology for identifying spectral configuration were presented in Section 2.3.2, 3.6.2, and 4.3.3. In this section, we evaluated and compared the execution time required to obtain the spectral configurations to meet each objective. As already mentioned, given the application of DR algorithms for onboard processing, the computational performance is a critical issue. Table 5.1, Table 5.2, Table 5.3, Table 5.4, and Table 5.5 give the details of the running times of the algorithms per iteration. Table 5.7 summarised the times for 20-band datasets. The execution time obtained is based on the implementation of the algorithms on a desktop computer having following characteristics: Operating system: Windows 7, Processor: Intel Core 2, and 16 GB RAM. The algorithms were written in IDL[1] programing language, version 8.2.

As already mentioned in Chapter 2, the execution time of the SRS algorithm depends on the number of samples (pixels) analysed. This algorithm was utilised in this comparison experiments to obtain the representative and separable bands. To determine the representative bands in the unsupervised case an entire scene is

---

[1] IDL, stands for Interactive Data Language, is a programming language used for data analysis. It is famous in particular areas of science, such as atmospheric physics, astronomy, and medical imaging.It was developed and used by the solar physics group at NASA in the early 1980s for analysing satellite data (Landsman, W. (1993). The IDL astronomy user's library. Astronomical Data Analysis Software and Systems II, Hassanien, A., M. Tolba, et al. (2014). Advanced machine learning technologies and applications. Second international conference, AMLTA.) It offers all the power, adaptability, and programmability of high-level languages like FORTRAN, C, and C++. Today, IDL has become the universal treatment system for analysing big data. More detailed information about IDL can be found at http://www.idlcoyote.com/documents/idllinks.php.

used and this requires a longer execution time than the supervised case, where only the samples (pixels) in the training set are used, e.g. 304 sec. vs. 172 sec. needed to provide the representative bands in unsupervised and supervised cases respectively. The comparison between the two datasets showed that the execution time of SRS was longer for the Salinas dataset: this was due to the number of samples since the Salinas dataset is more than five times larger than the Indian Pines one.

Table 5.7: The execution time(s) required to determine spectral configurations to meet each objective; the maximum number of bands was constrained to 20; the experiments were performed for both the supervised and unsupervised cases and the two datasets.

Objectives	Indian Pines		Salinas	
	Supervised	Unsupervised	Supervised	Unsupervised
Representative bands	172	304	668	1883
Independent bands	4	4	3	3
Separable bands	75	---	286	---

When the execution time to obtain the representative bands and the separable bands was compared, it was observed that the execution time was more than twice in the former case than in the latter, i.e. 172 sec. vs. 75 sec., with the same number of samples. It is because, the full spectral reflectance per pixel have to be evaluated for representative bands, while for the separability, the mean and the variance-covariance matrix of a class of pixels are evaluated.

On the other hand, the CM algorithm to obtain the independent bands was much faster than SRS since at each iteration it computes the correlation matrix of the entire data set, rather than a metric per pixel. The number of samples had a negligible effect on the running time and, because of that, there was no difference in time for the supervised and unsupervised cases. So, obtaining a spectral configuration based on the second objectives was much faster than for the two other objectives.

5.4. Comparison of image classification

The third comparison is based on the image classification and the accuracy of correct recognition of the pixels in the scenes in relation to the ground-truth reference maps. In this experiment, we applied two frequently used classifiers to the band configurations obtained for the three objectives in the supervised and unsupervised cases. The classifiers were maximum likelihood classifier (MLC), and support vector machine (SVM).

We separated each known land-cover in the scenes into training and testing data using a random subsampling method so that the training set represents the distribution of class attributes well. Two training datasets were generated by taking 40%, and 60% of the total number of pixels per class and the rest were considered

as testing data to assess the classification accuracy. Finally, the classification accuracies of the two types of experiments were averaged and illustrated in Figure 5.5 for the two scenes.
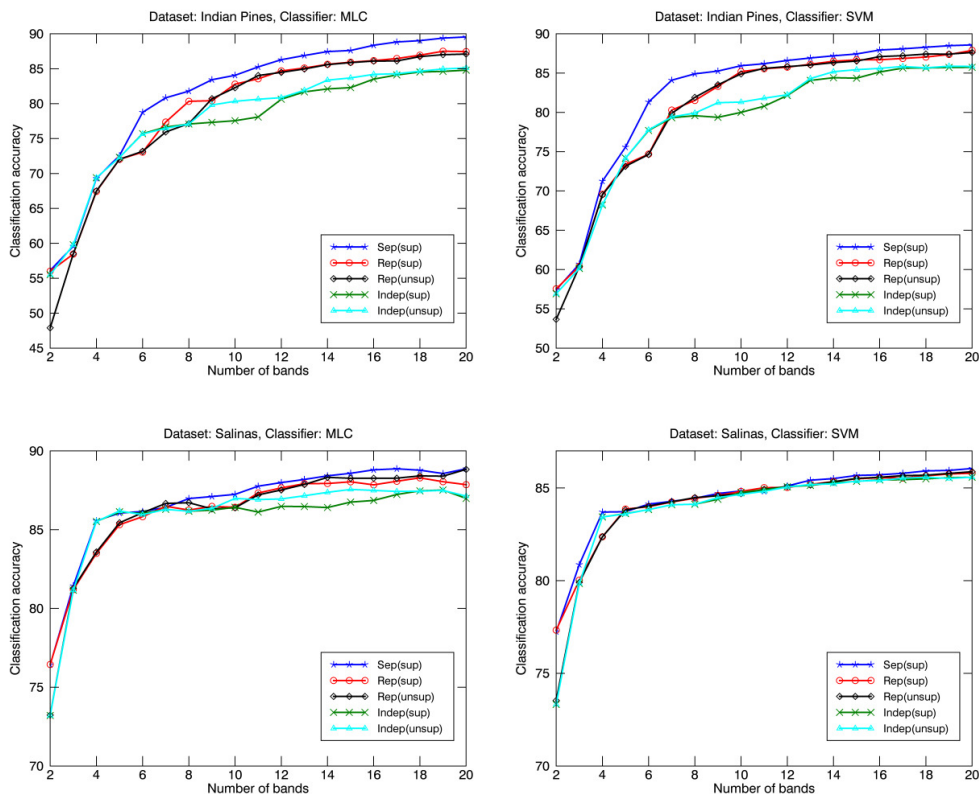

Figure 5.5: Classification accuracy of different band configurations obtained based on the three DR objectives of hyperspectral images for the Indian Pines (upper panel) and Salinas (lower panel) scenes. 'Sep', 'Rep', and 'Indep' stand for separable bands, representative bands, and independent bands respectively; 'sup' and 'unsup' for supervised and unsupervised. The results of MLC and SVM are depicted in left and right side pannels respectively.

This experiment demonstrated that overall the separable band configurations gave better classification accuracies than other configurations, then the representative bands and the last was independent bands. It was expected that the separable bands would provide more accurate classification, especially when applying MCL. This is because the SRS algorithm using the JM distance gave a spectral configuration maximising the class separability which is needed to improve the classification accuracy.

The representative bands capture the main spectral features, i.e. the key attributes related to land cover types. So, when the classifiers were applied to the representative bands, the classification accuracy was acceptable, although lower than with the separable bands. The supervised and unsupervised cases for the representative bands did not make much difference in the classification accuracy since as mentioned, there were similarities between the spectral features of pixels used in both cases. So, the band configurations obtained in one case were also suitable for the other one when applied to classification. The representative spectral configurations were obtained in a shorter time in the supervised case.

The independent bands gave a lower classification accuracy than two other band sets. It is because the independence of spectral bands does not relate to separability and accuracy of classification. By finding independent spectral regions, we reduce the dimensionality of hyperspectral data by eliminating redundant information. This spectral configuration may happen to increase the distance between the classes in the feature space, which is suitable for classification performance but is not an aspect "measured" by independence. Usually, in literature, the independence metric is applied for DR in the unsupervised case. In this experiment, we observed that unsupervised independent bands had almost better classification accuracy than supervised. It revealed that the unsupervised independent bands retained more information useful for classification performance.

5.5. Conclusion

Three different comparisons were conducted to evaluate different spectral configurations based on the mentioned objectives of hyperspectral dimensionality reduction. We conclude from the evaluation that in the case of a supervised dimensionality reduction, it is worth to identify separable spectral bands that grants a higher accuracy of image classification, compared with other spectral configurations. However, in the case of the lack of information about the classes in a scene (unsupervised) or the lack of interest in a specific region in the scene, we suggest identifying the independent and representative band sets. The independent bandset has an advantage of requiring very short computation time, so it is more suitable for the cases that execution time of the algorithm is of importance. The representative bands, though they are identified more slowly, provided more accurate classification result in comparison with the independent bands. One way to boost the running time of providing representative bands is to apply the SRS algorithm to some regions of interest in a scene, instead of using the entire scene.

6.
Conclusions and Recommendations

As mentioned through chapters, reducing the dimensionality (DR) of hyperspectral images is the most challenging step prior to application–specific analyses of data carrying a large amount of spectral information. Applying DR to hyperspectral data can be performed by applying various criteria providing reduced spectral configurations. Investigating different criteria in the literature, we categorised them into three main objectives covering broad criteria utilised in the hyperspectral community. A reduced spectral configuration can be determined to pursue three broad objectives: *a)* representing the spectral reflectance of pixels in a scene accurately, *b)* minimising redundant spectral information by selecting independent spectral features, and *c)* maximising class separability. Based on the objectives, four research questions were defined (Section 1.5.2).

In this final chapter, we summerize the answers to the research questions and highlight the most relevant products and findings of the research documented in this dissertation, followed by recommendations for further investigations.

6.1. Conclusions

a) What is the advantage of determining spectral configuration versus individual channel selection per objective?

As already mentioned, individual channel selection techniques preserve the connections between the spectral features and the physical properties of the observed materials, making a correct interpretation of the data easier. In our study, we developed a method for hyperspectral images to determine reduced spectral configurations instead of the selection of individual channels. In this method, the connections between the spectral feature and the physical properties of the materials in images are retained.

The primary advantage of identifying reduced spectral configurations and creating wider spectral bands by averaging narrow channels is to filter out random noise and increases the signal to noise ratio (SNR). It is known that the SNR of a sensor increases in proportion to the radiance signal from each pixel, which is decomposed into n spectral channels. Averaging the spectral channels causes a reduction in the variance of the signal and an increase in the SNR of the sensor.

The second advantage was pointed out in Chapter 2, where it was shown that a reduced spectral configuration, i.e. a set of contiguous bands, represents the full spectrum without using a spectral interpolation function. There are methods for spectral approximation and representation which use a set of basis functions to represent the spectra, while the proposed methodology represent spectra directly.

In Chapter 4, we specifically addressed this research question, and investigated the main advantage of having a reduced spectral configuration vs.

individual channel selection. This chapter revealed that the importance of creating wider spectral bands from narrow channels is in the discrimination between the classes in a scene and the improvement of image classification that can be the final result of a dimensionality reduction algorithm applied to hyperspectral datasets. Applying various class separability metrics to two datasets, we showed that the method provides band configurations with a better class separability than the channels selected from the original dataset and maximized the separability. The reason is that the method does not preserve the width of the original spectral channels, unlike the channel selection methods, and defines new spectral bands by merging the adjacent spectral channels if it is useful to improve separability. Averaging narrow adjacent channels results in smaller intra-class distances by reducing the class variances leading to an increase in class separability.

b)   How similar or dissimilar are the spectral configuration obtained based on different DR criteria? How large and significant are the differences of spectral configurations in the cases of using the entire scene (unsupervised) or using specific classes in the scenes (supervised)? Is it necessary to determine different spectral configurations?

According to the three main objectives to reduce the dimensionality of hyperspectral images, Chapters 2, 3, and 4 separately presented in detail the methodologies achieving these objectives, along with implementations and numerical experiments to compare with alternate solutions. These three chapters provided different band configurations based on the specified objectives. Chapter 5 presented a comparative evaluation of the spectral configurations determined to meet each objective, and answers these research questions.

In general, there was some similarity between the spectral configurations obtained based on the three objectives. The representative and separable spectral configurations were rather similar, while the ones identified for independent bands, were more different. It appears that the spectral configuration determined to represent full spectra with fewer bands is more sensitive to spectral gradients to minimise the error of the spectral representation. Independent bands consider the correlation of samples in a band with other bands, and random noise increases the independence of spectral bands, thus leading to selection. The spectral configuration to maximise separability depends on land cover and predefined classes, so is likely to differ from the previous two.

In the case of supervised and unsupervised DR for each objective, the comparison (Section 5.2) showed that the band configurations are highly dependent on the land cover types and the portion of the scene covering an image. Different land cover types gave different band configuration, and spectrally similar materials provide nearly identical bandset. If the region of interest covers a small

portion of an image and presents spectrally unique materials, the configuration is very different from the configuration obtained for entire scene and vice versa.

It was concluded that the necessity of determining different band configurations concerning the objectives is dependent on the application. In the case of a supervised dimensionality reduction, it is worth to identify separable spectral bands that give a higher accuracy of image classification, compared with other spectral configurations. However, in the case of the lack of information about the classes in a scene or the lack of interest in a specific region in the scene, we suggest identifying the independent and representative band sets. The independent bandset has an advantage of requiring very short computation time, so it is more suitable for the cases that execution time of the algorithm is of importance. The representative bands, though they are identified more slowly, provided more accurate classification result in comparison with the independent bands. One way to boost the running time of providing representative bands is to apply the SRS algorithm to some regions of interest in a scene, instead of using the entire scene.

c) What are the performances of different band configurations in distinguishing various classes in a scene and what is the accuracy of image classification?

Chapter 4 gave evidence of the advantage of broader spectral bands over narrow channels in image classification. Applying various class separability metrics to identify spectral bands compared with narrow spectral channels demonstrated that the land-cover classes in a scene are more distinguishable than the best channel selected for this purpose.

The assessment of different band configurations on image classification (Section 5.4) demonstrated that, overall, the separable band configurations gave better classification accuracies than other ones, then the representative bands and the last was independent bands. For example, Figure 5.5 shows that by applying the 20-band configuration of Indian Pines dataset to MLC, the classification accuracies were about 90, 87, and 85 for separable bands, representative bands, and independent bands respectively. The higher accuracy of separable bands is because the algorithm gave a spectral configuration maximising the class separability which is needed to improve the classification accuracy.

The representative bands captured the main spectral features, i.e., the critical spectral attributes related to land cover types. So, when the classifiers were applied to the representative bands, the classification accuracy was acceptable, although lower than with the separable bands. The independent bands gave a lower classification accuracy than two other band sets. It is because the independence of spectral bands does not relate to separability and accuracy of classification. By finding independent spectral regions, we reduce the dimensionality of hyperspectral data by eliminating redundant information. This spectral

configuration may happen to increase the distance between the classes in the feature space, which is suitable for classification performance but is not an aspect "measured" by independence. Usually, in literature, the independence metric is applied for unsupervised DR, where there is a lack of ancillary information about the classes.

d)   Finally, which band configuration would be suitable for miniaturised hyperspectral satellites?

Since the capability of a hyperspectral nanosatellite for onboard processing of is not explicitly specified, an absolute answer to this question is impossible. The first miniaturised hyperspectral satellite with the capability of onboard processing was launched on 2 February 2018, and to build a complete hyperspectral dataset onboard, a chain of processing stages have to be implemented. Then, the dataset is ready to be used in a DR algorithm. The faster the algorithm, the less power is needed for onboard processing. Based on this fact, three scenarios can be described as follows:

I-   Assume that the power budget is low, and there is no ancillary information about the classes in a scene. In this case, the best option is to identify the independent band configuration. It is shown in Chapter 3 and 5 that the CM algorithm provides independent bands in less a second.

II-   If the electric power is high, and the ancillary data of the classes in the scene is not available again, a better option is to apply the SRS algorithm to the dataset and provide the representative band configuration in an unsupervised way. In this situation, besides achieving more accurate classification results, the spectral features of materials in the image can be extracted.

III-   The third scenario is when extra information about the classes in a scene is available, and the power budget is high. In this case, the best option is to obtain the separable band configuration which leads to the most accurate classification results.

The methodology developed and presented in this thesis to provide reduced spectral band configurations is one solution among others for the major issues of hyperspectral images (HSI) introduced in Section 1.4.1. As mentioned in Section 1.4.2, the other main category of DR applied to HSI is feature extraction. If the only scope of analyzing hyperspectral datasets is to achieve a more accurate image classification, the feature extraction techniques may give higher accuracy; since the

data can be transformed into a new feature space providing a better separability between the classes in the scene. However, it should be noticed that the relation between the spectral features and the physical properties of materials in the classes is distorted or compromised; making the interpretation of final results obscure.

As mentioned in Section 1.5.3., in general achieving the best solution, i.e., finding the locations of m spectral splits from n possible locations obtaining the best score based on the given criterion, is a prohibitive method. Although in this thesis, the advantages of identifying reduced spectral band configurations compared to the selection of individual channels are revealed, there might be other heuristic solutions in the future that could determine band configurations more optimal than the algorithms developed in this study. What's more, since the SRS algorithm can easily accept various criteria, applying metrics based on the wrapper approaches might provide band configurations giving more accurate classification results than the filter approaches (see Section 4.5.5).

6.2. Contributions

Given the research objectives, focused on reducing the dimensionality of hyperspectral while minimizing the loss of information as defined in Chapter 1, the study led to the following innovative contributions:

1. A new DR algorithm applicable to hyperspectral images named Spectral Region Splitting (SRS) has been proposed. This method may not be categorised into the usual DR methods i.e. feature selection (FS) and feature extraction (FE). Similar to the FS algorithms, SRS provides a reduced feature set preserving the physical meaning of the original data. Unlike FE techniques, it does not throw any spectral information away, which the FS algorithms do. SRS is a greedy top-down searching strategy, that divides the spectrum into spectral regions. In each case, the spectral configuration is a set of contiguous bands with different widths. Furthermore, the wider bands identified by SRS have a better signal to noise ratio than narrow spectral channels, due to averaging the spectral signal over the spectral channels merged into the new spectral band.

2. SRS accepts different metrics and objectives to determine a spectral configuration. This capability allows to choose the best metric for a given objective to achieve a better outcome. For example, we applied five different metrics to SRS to represent the spectral reflectance of a complete spectral library. Then the approximated spectra were used in a material detection procedure. We found out that SRS using a metric such as SID

gives a better accuracy of material detection than a method providing the best spectral representation but just can use RMSE as the metric.

3. Comparing the spectral configurations obtained by SRS with the best representation of the spectra obtained by the piecewise constant function approximation (PCFA) algorithm, SRS provided almost identical spectral configurations with a negligible mean difference in the RMSE.

4. When SRS is applied to an entire hyperspectral scene, i.e. providing the same band configuration for all the pixels in the scene, the approximation error on spectral reflectance obtained by PCFA and SRS are identical, while SRS was about 40 times faster than PCFA. We can apply the method to hyperspectral images in supervised and unsupervised cases with the supervised one providing the representative spectral configuration much faster.

5. The Total dependence (TD) metric for measuring dependence in multivariate systems was introduced and developed to be used for the DR of hyperspectral images. We proved the mathematical properties of the metric and showed its utility for identifing independent spectral features. TD is a normalised metric, so the dependence of various spectral band combinations with different numbers of bands can be compared leading to a selection of a spectral configuration with minimum dependence. TD is sensitive to the dependence of a new band added to an existing band set, and it is not always increasing. It can easily be calculated regardless the number of bands.

6. We developed a new algorithm to determine a spectral configuration called Clustering-Merging (CM). This algorithm, likewise SRS, provides a continuous spectral configuration covering the whole of the spectrum by averaging the adjacent spectral channels, but in a bottom-up searching strategy. Instead of iteratively splitting the spectrum as in SRS, it merges the spectral channels iteratively, so the final spectral configurations have similar broad characteristics to the ones determined by SRS. With the utilisation of the TD metric in CM, we obtain independent spectral regions. We revealed that small changes in the spectral configuration obtained by CM increase the overall dependence. This is a rapid method i.e. less than a second to get a required spectral configuration for an entire hyperspectral scene.

7. By comparing the classification accuracies using the independent spectral regions obtained by CM and channels selected by unsupervised FS

algorithms with the aim of minimising information shared by different spectral channels, we revealed that the band sets obtained by CM gave better or comparable accuracy. Furthermore, CM determines the spectral configurations much faster than the other algorithms.

8. In the supervised DR strategy, we applied six different class separability metrics with the SRS algorithm and obtained spectral configurations maximising class separability. Comparing with the most separable channel set obtained with the branch and bounds algorithm, we showed that the SRS spectral configuration gave better separability than the original spectral channels. We demonstrated that averaging spectral channels to obtain wider spectral bands reduces the intra-class distance, causing an increase in class separability. Having a better separability of the classes in a scene leads to a higher classification accuracy and a better identification of materials.

9. The three proposed objectives lead to determine different spectral configurations. Large spectral gradients may lead to cases where independent spectral bands are identified which are also useful for class identification. The spectral configurations obtained to pursue the first and the third objectives where rather similar. The independent band configuration is more sensitive to the noise in the scene and it may select noisy narrow channels as separated spectral bands. The experiments on classification accuracy with the spectral configurations obtained for the three objectives showed that the highest accuracy was achieved with the separable configuration, followed by representative and independent bands respectively. The shortest execution time with the three algorithms was achieved when determining independent bands, followed by separable and then representative bands.

10. Dimensionality reduction of hyperspectral images through spectral region identification contains the following advantages: a) preserving the relationship between the characteristics of an object and its reflectance spectrum; b) having the freedom to identify spectral features with broader spectral resolution when it is necessary; c) increasing the signal to noise ratio by averaging narrow spectral channels; d) obtaining a higher class separability leading to more accurate image classification and material identificaion; and e) using all the spectral information of the original dataset.

6.3. **Recommendations**

This study covered the three primary objectives dimensionality reduction applied to hyperspectral images to tackle the principal preprocessing issue of these datasets. Following the research presented in this thesis, efforts should be made in the following domains for future studies.

1- In this thesis, we focused on analysing of hyperspectral datasets in the spectral domain, while these datasets carry spatial information as well. To reduce the volume of the datasets for more straightforward analyses of the data in different applications, focusing on methodologies for the reduction of information in both spectral and spatial domain while keeping the necessary information, is recommended. For instance, an integration of a method like segmentation applied to the spatial domain with an algorithm such as the SRS method, which has been used, in this thesis, in the spectral domain.

2- In this thesis, the dimensionality reduction in the context of image classification was based on the filter approaches, i.e. the identification of spectral regions was completely independent of image classification. It is worth to identifying spectral regions using the wrapper approaches as well, i.e. with the focus on the classifier that will be used for evaluation of the final bandset.

3- Applying the spectral region identification algorithms using different criteria to a specific target to determine predominant spectral regions for recognising the target from background is recommended. A collection of such experiments over different targets, would help to identify the best spectral regions per target and validate the spectral regions identified regarding the objectives. These experiments would also pave the way to determine the spectral regions needed for onboard processing of hyperspectral images which is devoted to a region of interest.

References

Adams, J. B. (1974). "Visible and near-infrared diffuse reflectance spectra of pyroxenes as applied to remote sensing of solid objects in the solar system." Journal of Geophysical Research 79(32): 4829-4836.

Adams, J. B. and A. L. Filice (1967). "Spectral reflectance 0.4 to 2.0 microns of silicate rock powders." Journal of Geophysical Research 72(22): 5705-5715.

Agrawal, R., C. Faloutsos, et al. (1993). Efficient Similarity Search In Sequence Databases. Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms, Springer-Verlag: 69-84.

Alonso, M. C., J. A. Malpica, et al. (2011). Consequences of the Hughes phenomenon on some classification techniques. ASPRS 2011 Annual Conference, Milwaukee, Wisconsin May.

Amaldi, E. and V. Kann (1998). "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems." Theoretical Computer Science 209(1-2): 237-260.

Angelopoulou, E. (2000). "Objective colour from multispectral imaging." Computer Vision - Eccv 2000, Pt I, Proceedings 1842: 359-374.

Angelopoulou, E., R. Molana, et al. (2001). "Multispectral skin color modeling." 2001 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, Proceedings: 635-642.

Asmat, A., P. M. Atkinson, et al. (2010). "Geostatistically estimated image noise is a function of variance in the underlying signal." International Journal of Remote Sensing 31(4): 1009-1025.

Atkinson, P. M., I. M. Sargent, et al. (2005). "Interpreting image-based methods for estimating the signal-to-noise ratio." International Journal of Remote Sensing 26(22): 5099-5115.

Bader, D. A., W. E. Hart, et al. (2005). Parallel Algorithm Design for Branch and Bound. Tutorials on Emerging Methodologies and Applications in Operations Research: Presented at Informs 2004, Denver, CO. H. J. G. New York, NY, Springer New York: 5-1-5-44.

Bajcsy, P. and P. Groves (2004). "Methodology for hyperspectral band selection." Photogrammetric Engineering & Remote Sensing 70(7): 793-802.

Barducci, A., D. Guzzi, et al. (2012). Blind atmospheric correction of hyperspectral images gathered by high spectral resolution sensors. 2012 IEEE International Geoscience and Remote Sensing Symposium.

Barnsley, M. J., J. J. Settle, et al. (2004). "The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere." Ieee Transactions on Geoscience and Remote Sensing 42(7): 1512-1520.

Barsi, J. A., B. L. Markham, et al. (2007). "Radiometric calibration status of Landsat-7 and Landsat-5." Sensors, Systems, and Next-Generation Satellites Xi 6744.

Bazi, Y. and F. Melgani (2006). "Toward an optimal SVM classification system for hyperspectral remote sensing images." Ieee Transactions on Geoscience and Remote Sensing 44(11): 3374-3385.

Bell, I. E. and G. V. G. Baranoski (2004). "Reducing the dimensionality of plant spectral databases." Ieee Transactions on Geoscience and Remote Sensing 42(3): 570-576.

Bellman, R. (1954). The theory of dynamic programming, RAND CORP SANTA MONICA CA.

Bellman, R. (2013). Dynamic programming, Courier Corporation.

Ben-Dor, E., N. Goldlshleger, et al. (2003). "The spectral reflectance properties of soil structural crusts in the 1.2- to 2.5-mu m spectral region." Soil Science Society of America Journal 67(1): 289-299.

Bendall, G. and F. Margot (2006). "Greedy-type resistance of combinatorial problems." Discrete Optimization 3(4): 288-298.

Bernecker, T., M. Houle, et al. (2011). "Quality of similarity rankings in time series." Advances in Spatial and Temporal Databases: 422-440.

Bettencourt, L. M. A., V. Gintautas, et al. (2008). "Identification of functional information subgraphs in complex networks." Physical Review Letters 100(23).

Bickel, P. J. and E. Levina (2004). "Some theory for Fisher's linear discriminant function,'naive Bayes', and some alternatives when there are many more variables than observations." Bernoulli: 989-1010.

Bioucas-Dias, J. M. and J. M. P. Nascimento (2008). "Hyperspectral subspace identification." Ieee Transactions on Geoscience and Remote Sensing 46(8): 2435-2445.

Bioucas-Dias, J. M., A. Plaza, et al. (2013). "Hyperspectral Remote Sensing Data Analysis and Future Challenges." Ieee Geoscience and Remote Sensing Magazine 1(2): 6-36.

Blackburn, G. A. (2007). "Hyperspectral remote sensing of plant pigments." Journal of Experimental Botany 58(4): 855-867.

Bruzzone, L., F. Roli, et al. (1995). "An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection." Ieee Transactions on Geoscience and Remote Sensing 33(6): 1318-1321.

Bubitzky, W., M. Granzow, et al. (2007). Fundamentals of data mining in genomics and proteomics. New York, Springer.

Cariou, C., K. Chehdi, et al. (2011). "BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing." Ieee Geoscience and Remote Sensing Letters 8(3): 565-569.

Carmona, P. L., A. Mart, et al. (2011). Band selection in spectral imaging for classification and regression tasks using information theoretic measures. Information Optics (WIO), 2011 10th Euro-American Workshop on.

Carron, L. P. (1991). Morse code: the essential language, Amer Radio Relay League.

Chakrabarti, K., E. Keogh, et al. (2002). "Locally adaptive dimensionality reduction for indexing large time series databases." Acm Transactions on Database Systems 27(2): 188-228.

Chandrasekaran, B. and A. K. Jain (1974). "Quantization complexity and independent measurements." Ieee Transactions on Computers 100(1): 102-106.

Chang, C.-I. (2003). Hyperspectral imaging : techniques for spectral detection and classification. New York, Kluwer Academic/Plenum Publishers.

Chang, C.-I. (2007). Hyperspectral data exploitation : theory and applications. Hoboken, N.J. :, Wiley-Interscience.

Chang, C. I. (2000). "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis." Ieee Transactions on Information Theory 46(5): 1927-1932.

Chang, C. I., Q. Du, et al. (1999). "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification." Ieee Transactions on Geoscience and Remote Sensing 37(6): 2631-2641.

Chein, I. C. and W. Su (2006). "Constrained band selection for hyperspectral imagery." Ieee Transactions on Geoscience and Remote Sensing 44(6): 1575-1585.

Chen, C., W. Li, et al. (2014). "Reconstruction of Hyperspectral Imagery From Random Projections Using Multihypothesis Prediction." Ieee Transactions on Geoscience and Remote Sensing 52(1): 365-374.

Chen, D., A. S. Leon, et al. (2016). "Dimension reduction of decision variables for multireservoir operation: A spectral optimization model." Water Resources Research 52(1): 36-51.

Clark, R. N. (1999). "Spectroscopy of rocks and minerals, and principles of spectroscopy." Manual of remote sensing 3: 3-58.

Clark, R. N. and T. L. Roush (1984). "Reflectance Spectroscopy - Quantitative-Analysis Techniques for Remote-Sensing Applications." Journal of Geophysical Research 89(Nb7): 6329-6340.

Cloutis, E. A. (1996). "Hyperspectral geological remote sensing: Evaluation of analytical techniques." International Journal of Remote Sensing 17(12): 2215-2242.

Cochrane, M. (2000). "Using vegetation reflectance variability for species level classification of hyperspectral data." International Journal of Remote Sensing 21(10): 2075-2087.

Conticello, S. S., M. Esposito, et al. (2016). Hyperspectral imaging for real time land and vegetation inspection from nanosatellite. Small satellite systems and services. MALTA, ESA.

Cormen, T. H. (2009). Introduction to algorithms. Cambridge, Mass., MIT Press.

Cormen, T. H., C. E. Leiserson, et al. (2001). Introduction To Algorithms, MIT Press.

Cover, T. M. and J. A. Thomas (2006). Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing), Wiley-Interscience.

Crouvi, O., E. Ben-Dor, et al. (2006). "Quantitative mapping of arid alluvial fan surfaces using field spectrometer and hyperspectral remote sensing." Remote Sensing of Environment 104(1): 103-117.

Crowley, J. K. and N. Vergo (1988). "Near-Infrared Reflectance Spectra of Mixtures of Kaolin-Group Minerals - Use in Clay Mineral Studies." Clays and Clay Minerals 36(4): 310-316.

Curran, P. J. and J. L. Dungan (1989). "Estimation of Signal-to-Noise - a New Procedure Applied to Aviris Data." IEEE Transactions on Geoscience and Remote Sensing 27(5): 620-628.

Deborah, H., N. Richard, et al. (2015). "A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing." Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(6): 3224-3234.

Donoho, D. L. (2000). "High-dimensional data analysis: The curses and blessings of dimensionality." AMS CONFERENCE ON MATH CHALLENGES OF THE 21ST CENTURY: 1-32.

Du, Q., H. Ren, et al. (2003). "A comparative study for orthogonal subspace projection and constrained energy minimization." Ieee Transactions on Geoscience and Remote Sensing 41(6): 1525-1529.

Du, Q. and H. Yang (2008). "Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis." Ieee Geoscience and Remote Sensing Letters 5(4): 564-568.

Du, Y. Z., C. I. Chang, et al. (2004). "New hyperspectral discrimination measure for spectral characterization." Optical Engineering 43(8): 1777-1786.

Duda, R. O. and P. E. Hart (1973). Pattern classification and scene analysis. New York,, Wiley.

Duggin, M. J., H. Sakhavat, et al. (1985). "Systematic and Random Variations in Thematic Mapper Digital Radiance Data." Photogrammetric Engineering and Remote Sensing 51(9): 1427-1434.

Elachi, C. and J. Van Zyl (2006). Introduction to the physics and techniques of remote sensing. Hoboken, N.J., Wiley-Interscience.

Fan, J. and Y. Fan (2008). "High dimensional classification using features annealed independence rules." Annals of statistics 36(6): 2605.

Fan, J. and R. Li (2006). "Statistical challenges with high dimensionality: Feature selection in knowledge discovery." arXiv preprint math/0602133.

Fauvel, M., J. Chanussot, et al. (2009). "Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data over Urban Areas." Eurasip Journal on Advances in Signal Processing.

Fauvel, M., Y. Tarabalka, et al. (2013). "Advances in Spectral-Spatial Classification of Hyperspectral Images." Proceedings of the Ieee 101(3): 652-675.

Feng, J., L. C. Jiao, et al. (2014). "Hyperspectral Band Selection Based on Trivariate Mutual Information and Clonal Selection." Ieee Transactions on Geoscience and Remote Sensing 52(7): 4092-4105.

Foley, D. (1972). "Considerations of sample and feature size." Ieee Transactions on Information Theory 18(5): 618-626.

Folkman, M., J. Pearlman, et al. (2001). "EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration." Hyperspectral Remote Sensing of the Land and Atmosphere 4151: 40-51.

Friedman, S. and H. F. Weisberg (1981). "Interpreting the 1st Eigenvalue of a Correlation Matrix." Educational and Psychological Measurement 41(1): 11-21.

Frost, R. L. (1995). "Fourier-Transform Raman-Spectroscopy of Kaolinite, Dickite and Halloysite." Clays and Clay Minerals 43(2): 191-195.

Fujimoto, N., Y. Takahashi, et al. (1989). Evaluation of SPOT HRV image data received in Japan. Quantitative Remote Sensing – as Economic Tool for the Nineties, Vancouver, BC, Canada, IGARSS '89: 12th Canadian Symposium on Remote Sensing.

Gaffey, S. J. (1987). "Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 um): Anhydrous carbonate minerals." Journal of Geophysical Research: Solid Earth 92(B2): 1429-1440.

Gamon, J. and J. Surfus (1999). "Assessing leaf pigment content and activity with a reflectometer." The New Phytologist 143(1): 105-117.

Gao, Z. Y., R. D. Jia, et al. (2016). "Simulation and Analysis of Spectral Response Function and Bandwidth of Spectrometer." International Journal of Aerospace Engineering.

Garcia-Vilchez, F., J. Munoz-Mari, et al. (2011). "On the Impact of Lossy Compression on Hyperspectral Image Classification and Unmixing." Ieee Geoscience and Remote Sensing Letters 8(2): 253-257.

Garner, W. R. and W. J. Mcgill (1956). "The Relation between Information and Variance Analyses." Psychometrika 21(3): 219-228.

Gheyas, I. A. and L. S. Smith (2010). "Feature subset selection in large dimensionality domains." Pattern recognition 43(1): 5-13.

Ghiyamat, A. and H. Z. M. Shafri (2010). "A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment." International Journal of Remote Sensing 31(7): 1837-1856.

Gitelson, Stark, et al. (2002). "Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction." International Journal of Remote Sensing 23(13): 2537-2562.

Gitelson, A. A., Y. Zur, et al. (2002). "Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy¶." Photochemistry and Photobiology 75(3): 272-281.

Goetz, A. F. H. (2009). "Three decades of hyperspectral remote sensing of the Earth: A personal view." Remote Sensing of Environment 113: S5-S16.

Goetz, A. F. H., G. Vane, et al. (1985). "Imaging Spectrometry for Earth Remote-Sensing." Science 228(4704): 1147-1153.

Gonzalez, R. C. and R. E. Woods (2008). Digital Image Processing, Pearson/Prentice Hall.

Grant, L., C. S. T. Daughtry, et al. (1993). "Polarized and specular reflectance variation with leaf surface features." Physiologia Plantarum 88(1): 1-9.

Green, R. O., M. L. Eastwood, et al. (1998). "Imaging spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS)." Remote Sensing of Environment 65(3): 227-248.

Gu, Q., Z. Li, et al. (2012). "Generalized fisher score for feature selection." arXiv preprint arXiv:1202.3725.

Gu, Y. F. and Y. Zhang (2003). "Unsupervised subspace linear spectral mixture analysis for hyperspectral images." 2003 International Conference on Image Processing, Vol 1, Proceedings: 801-804.

Guanter, L., R. Richter, et al. (2006). "Spectral calibration of hyperspectral imagery using atmospheric absorption features." Applied Optics 45(10): 2360-2370.

Guo, B. F., S. R. Gunn, et al. (2006). "Band selection for hyperspectral image classification using mutual information." Ieee Geoscience and Remote Sensing Letters 3(4): 522-526.

Guyon, I., Andr, et al. (2003). "An introduction to variable and feature selection." J. Mach. Learn. Res. 3: 1157-1182.

Guyon, I., J. Weston, et al. (2002). "Gene selection for cancer classification using support vector machines." Machine learning 46(1): 389-422.

Han, J. S., S. W. Lee, et al. (2013). "Feature subset selection using separability index matrix." Information Sciences 223: 102-118.

Hapke, B. (2012). Theory of reflectance and emittance spectroscopy. Cambridge, UK ; New York, Cambridge University Press.

Hassanien, A., M. Tolba, et al. (2014). Advanced machine learning technologies and applications. Second international conference, AMLTA.

Hernandez-Andres, J., J. Romero, et al. (2001). "Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain." Journal of the Optical Society of America a-Optics Image Science and Vision 18(2): 412-420.

Herrmann, R. and C. Onkelinx (1986). "Quantities and Units in Clinical-Chemistry - Nebulizer and Flame Properties in Flame Emission and Absorption Spectrometry - (Recommendations 1986)." Pure and Applied Chemistry 58(12): 1737-1742.

Hilbert, E. E. (1977). "Cluster compression algorithm: a joint clustering/data compression concept."

Hirtle, H. and A. Rencz (2003). "The relation between spectral reflectance and dissolved organic carbon in lake water: Kejimkujik National Park, Nova Scotia, Canada." International Journal of Remote Sensing 24(5): 953-967.

Hosseini Aria, S. E., M. Menenti, et al. (2017). "Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data." Journal of Applied Remote Sensing 11(4): 046010.

Houle, M. E., H.-P. Kriegel, et al. (2010). Can shared-neighbor distances defeat the curse of dimensionality? International Conference on Scientific and Statistical Database Management, Springer.

Huang, M.-L., Y.-H. Hung, et al. (2014). "SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier." The Scientific World Journal 2014.

Huang, R. and M. Y. He (2005). "Band selection based on feature weighting for classification of hyperspectral data." Ieee Geoscience and Remote Sensing Letters 2(2): 156-159.

Hughes, G. (1968). "On the mean accuracy of statistical pattern recognizers." Ieee Transactions on Information Theory 14(1): 55-63.

Hunt, G. R., J. W. Salisbury, et al. (1971). "Visible and near-infrared spectra of minerals and rocks; III, Oxides and Hydroxides." Modern Geology 2: 195-205.

Huynh, C. P. and A. Robles-Kelly (2008). "A NURBS-based spectral reflectance descriptor with applications in computer vision and pattern recognition." 2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12: 3345-3352.

Ifarraguerri, A. and M. W. Prairie (2004). "Visual Method for Spectral Band Selection." Ieee Geoscience and Remote Sensing Letters 1(2): 101-106.

Imai, F. H., M. R. Rosen, et al. (2002). "Comparative study of metrics for spectral match quality." Cgiv'2002: First European Conference on Colour in Graphics, Imaging, and Vision, Conference Proceedings: 492-496.

Jain, A. K. and R. C. Dubes (1988). Algorithms for clustering data, Prentice-Hall, Inc.

Jain, A. K., R. P. W. Duin, et al. (2000). "Statistical pattern recognition: A review." IEEE Transactions on pattern analysis and machine intelligence 22(1): 4-37.

Jensen, A. C. and A. S. Solberg (2007). "Fast hyperspectral feature reduction using piecewise constant function approximations." Ieee Geoscience and Remote Sensing Letters 4(4): 547-551.

Jensen, J. R. (2015). Introductory Digital Image Processing: A Remote Sensing Perspective, Prentice Hall Press.

Jia, S., Z. Ji, et al. (2012). "Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal." Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(2): 531-543.

Jia, S., Z. Zhu, et al. (2014). "A Two-Stage Feature Selection Framework for Hyperspectral Image Classification Using Few Labeled Samples." Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(4): 1023-1035.

Jia, X. P., B. C. Kuo, et al. (2013). "Feature Mining for Hyperspectral Image Classification." Proceedings of the Ieee 101(3): 676-697.

Jimenez-Rodriguez, L. O., E. Arzuaga-Cruz, et al. (2007). "Unsupervised Linear Feature-Extraction Methods and Their Effects in the Classification of High-Dimensional Data." Ieee Transactions on Geoscience and Remote Sensing 45(2): 469-483.

Jimenez, L. O. and D. A. Landgrebe (1998). "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data." Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews 28(1): 39-54.

Jolliffe, I. T. (2002). Principal component analysis. New York, Springer.

Kaewpijit, S., J. Le Moigne, et al. (2002). "A wavelet-based PCA reduction for hyperspectral imagery." Igarss 2002: Ieee International Geoscience and Remote Sensing Symposium and 24th Canadian Symposium on Remote Sensing, Vols I-Vi, Proceedings: 2581-2583.

Kahveci, T. and A. Singh (2001). Variable length queries for time series data. Data Engineering, 2001. Proceedings. 17th International Conference on.

Kaiser, H. F. (1968). "A Measure of Average Intercorrelation." Educational and Psychological Measurement 28(2): 245-&.

Kamandar, M. and H. Ghassemian (2011). Maximum relevance, minimum redundancy band selection for hyperspectral images. 2011 19th Iranian Conference on Electrical Engineering.

Kamandar, M. and H. Ghassemian (2011). Maximum relevance, minimum redundancy band selection for hyperspectral images. Electrical Engineering (ICEE), 2011 19th Iranian Conference on.

Kanal, L. and B. Chandrasekaran (1971). "On dimensionality and sample size in statistical pattern classification." Pattern recognition 3(3): 225-234.

Keogh, E., K. Chakrabarti, et al. (2001). "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases." Knowledge and Information Systems 3(3): 263-286.

Keogh, E. and A. Mueen (2010). Curse of Dimensionality. Encyclopedia of Machine Learning. C. Sammut and G. I. Webb. Boston, MA, Springer US: 257-258.

Keshava, N. and J. F. Mustard (2002). "Spectral unmixing." Ieee Signal Processing Magazine 19(1): 44-57.

Kohavi, R. and G. H. John (1997). "Wrappers for feature subset selection." Artificial Intelligence 97(1-2): 273-324.

Kokash, N. (2005). "An introduction to heuristic algorithms." Department of Informatics and Telecommunications: 1-8.

Kolodner, M. A. (2007). "An Automated Targert Detection System for Hyperspectral Imaging Sensors." Johns Hopkins APL technical digest 27(3): 208-217.

Konno, H. and T. Kuno (1988). "Best Piecewise Constant Approximation of a Function of Single Variable." Operations Research Letters 7(4): 205-210.

Kruse, F. A., A. B. Lefkoff, et al. (1993). "The Spectral Image-Processing System (Sips) - Interactive Visualization and Analysis of Imaging Spectrometer Data." Remote Sensing of Environment 44(2-3): 145-163.

Kulkarni, A. V. (1978). "On the mean accuracy of hierarchical classifiers." Ieee Transactions on Computers 27(8): 771-776.

Kumar, S., J. Ghosh, et al. (2001). "Best-bases feature extraction algorithms for classification of hyperspectral data." Ieee Transactions on Geoscience and Remote Sensing 39(7): 1368-1379.

Kuo, B.-C., H.-H. Ho, et al. (2014). "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification." Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(1): 317-326.

Landgrebe, D. (2002). "Hyperspectral image data analysis." Ieee Signal Processing Magazine 19(1): 17-28.

Landgrebe, D. A. (2005). "Multispectral land sensing: where from, where to?" Ieee Transactions on Geoscience and Remote Sensing 43(3): 414-421.

Landsman, W. (1993). The IDL astronomy user's library. Astronomical Data Analysis Software and Systems II.

Lanthier, Y., A. Bannari, et al. (2008). Hyperspectral Data Segmentation and Classification in Precision Agriculture: A Multi-Scale Analysis. IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium.

Le Moan, S., A. Mansouri, et al. (2011). "A Constrained Band Selection Method Based on Information Measures for Spectral Image Color Visualization." Ieee Transactions on Geoscience and Remote Sensing 49(12): 5104-5115.

Lee, Z., K. Carder, et al. (2007). "Determination of primary spectral bands for remote sensing of aquatic environments." Sensors 7(12): 3428-3441.

Lee, Z. and K. L. Carder (2002). "Effect of spectral band numbers on the retrieval of water column and bottom properties from ocean color data." Applied Optics 41(12): 2191-2201.

Li, X. and J. Liu (2009). An adaptive band selection algorithm for dimension reduction of hyperspectral images. Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on.

Li, Z.-L., F. Becker, et al. (1999). "Channel selection for soil spectrum reconstruction in 8–13 μm region." Journal of Geophysical Research: Atmospheres 104(D18): 22271-22285.

Lillesand, T. M., R. W. Kiefer, et al. (2008). Remote sensing and image interpretation. Hoboken, NJ, John Wiley & Sons.

Lodha, S. P. and S. Kamlapur (2014). "Dimensionality reduction techniques for hyperspectral images." International Journal of Application or Innovation in Engineering & Management (IJAIEM) 3(10).

Manolakis, D., D. Marden, et al. (2003). "Hyperspectral image processing for automatic target detection applications." Lincoln Laboratory Journal 14(1): 79-116.

Manolakis, D., C. Siracusa, et al. (2001). "Hyperspectral subpixel target detection using the linear mixing model." Ieee Transactions on Geoscience and Remote Sensing 39(7): 1392-1409.

Mansouri, A., T. Sliwa, et al. (2008). "An Adaptive-PCA Algorithm for Reflectance Estimation from Color Images." 19th International Conference on Pattern Recognition, Vols 1-6: 2941-2944.

Marimont, R. and M. Shapiro (1979). "Nearest neighbour searches and the curse of dimensionality." IMA Journal of Applied Mathematics 24(1): 59-70.

Martinez-Uso, A., F. Pla, et al. (2007). "Clustering-based hyperspectral band selection using information measures." Ieee Transactions on Geoscience and Remote Sensing 45(12): 4158-4171.

Masulli, F. and G. Valentini (2001). "Quantitative evaluation of dependence among outputs in ECOC classifiers using mutual information based measures." Ijcnn'01: International Joint Conference on Neural Networks, Vols 1-4, Proceedings: 784-789.

Matsuda, H. (2000). "Physical nature of higher-order mutual information: Intrinsic correlations and frustration." Physical Review E 62(3): 3096-3102.

Mayer, E. (1976). "A measure of the average intercorrelation." EDUCA-TIONAL AND PSYCHOLOGICAL MEASUREMENT 35: 67-72.

McLachlan, G. (2004). Discriminant analysis and statistical pattern recognition, John Wiley & Sons.

Mielikainen, J. (2006). "Lossless compression of hyperspectral images using lookup tables." Ieee Signal Processing Letters 13(3): 157-160.

Mielikainen, J., P. Toivanen, et al. (2003). "Linear prediction in lossless compression of hyperspectral images." Optical Engineering 42(4): 1013-1017.

Motta, G., F. Rizzo, et al. (2006). Hyperspectral Data Compression, Springer US.

Motta, G., F. Rizzo, et al. (2006). Hyperspectral data compression. New York, Springer Science+Business Media.

Narendra, P. and K. Fukunaga (1977). "Branch and Bound Algorithm for Feature Subset Selection." Ieee Transactions on Computers 26(9): 917-922.

Nering, E. D. (1970). Linear algebra and matrix theory. New York,, Wiley.

Nievergelt, J. (2000). Exhaustive search, combinatorial optimization and enumeration: Exploring the potential of raw computing power. International Conference on Current Trends in Theory and Practice of Computer Science, Springer.

Oommen, T., D. Misra, et al. (2008). "An objective analysis of Support Vector Machine based classification for remote sensing." Mathematical Geosciences 40(4): 409-424.

Pal, M. (2006). "Support vector machine-based feature selection for land cover classification: a case study with DAIS hyperspectral data." International Journal of Remote Sensing 27(14): 2877-2894.

Parra, L. C., C. Spence, et al. (1999). Unmixing Hyperspectral Data. NIPS.

Parsons, L., E. Haque, et al. (2004). "Subspace clustering for high dimensional data: a review." SIGKDD Explor. Newsl. 6(1): 90-105.

Peng, H., F. Long, et al. (2005). "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." Pattern Analysis and Machine Intelligence, IEEE Transactions on 27(8): 1226-1238.

Peng, J. T., Y. C. Zhou, et al. (2015). "Region-Kernel-Based Support Vector Machines for Hyperspectral Image Classification." Ieee Transactions on Geoscience and Remote Sensing 53(9): 4810-4824.

Penner, M. H. (2010). Basic Principles of Spectroscopy. Food Analysis. Boston, MA, Springer US: 375-385.

Pickering, M. R. and M. J. Ryan (2006). An Architecture for the Compression of Hyperspectral Imagery. Hyperspectral Data Compression. G. Motta, F. Rizzo and J. A. Storer. Boston, MA, Springer US: 1-34.

Piecuch, P., J. Maruani, et al. (2009). Advances in the Theory of Atomic and Molecular Systems: Conceptual and Computational Advances in Quantum Chemistry, Springer Netherlands.

Plaza, A., P. Martinez, et al. (2004). "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data." Ieee Transactions on Geoscience and Remote Sensing 42(3): 650-663.

Plaza, A., P. Martinez, et al. (2005). "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations." Ieee Transactions on Geoscience and Remote Sensing 43(3): 466-479.

Pleniou, M. and N. Koutsias (2013). "Sensitivity of spectral reflectance values to different burn and vegetation ratios: A multi-scale approach applied in a fire affected area." Isprs Journal of Photogrammetry and Remote Sensing 79(2): 199-210.

Pontius, J., M. Martin, et al. (2008). "Ash decline assessment in emerald ash borer-infested regions: A test of tree-level, hyperspectral technologies." Remote Sensing of Environment 112(5): 2665-2676.

Prasad, S. and L. M. Bruce (2008). "Decision fusion with confidence-based weight assignment for hyperspectral target recognition." Ieee Transactions on Geoscience and Remote Sensing 46(5): 1448-1456.

Price, J. C. (1975). "Information content of Iris spectra." <u>Journal of Geophysical Research</u> 80(15): 1930-1936.

Price, J. C. (1990). "On the Information-Content of Soil Reflectance Spectra." <u>Remote Sensing of Environment</u> 33(2): 113-121.

Price, J. C. (1994). "How Unique Are Spectral Signatures." <u>Remote Sensing of Environment</u> 49(3): 181-186.

Pudil, P., J. Novovicova, et al. (1994). "Floating Search Methods in Feature-Selection." <u>Pattern Recognition Letters</u> 15(11): 1119-1125.

Richards, J. A. and X. Jia (2006). <u>Remote sensing digital image analysis : an introduction</u>. Berlin, Springer.

Romero, J., A. Garcia-Beltran, et al. (1997). "Linear bases for representation of natural and artificial illuminants." <u>JOSA A</u> 14(5): 1007-1014.

Roy, P. S. (1989). "Spectral reflectance characteristics of vegetation and their use in estimating productive potential." <u>Proceedings: Plant Sciences</u> 99(1): 59-81.

Salem, F. and M. Kafatos (2001). <u>Hyperspectral image analysis for oil spill mitigation</u>. Paper presented at the 22nd Asian Conference on Remote Sensing.

Salinelli, E. and C. Sgarra (2011). "Some Results on Correlation Matrices for Interest Rates." <u>Acta Applicandae Mathematicae</u> 115(3): 291-318.

Santos, A., C. de S Celes, et al. (2012). <u>Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithm and cluster validity</u>. Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Sarhrouni, E., A. Hammouch, et al. (2012). "Band Selection and Classification of Hyperspectral Images using Mutual Information: An algorithm based on minimizing the error probability using the inequality of Fano." <u>2012 International Conference on Multimedia Computing and Systems (Icmcs)</u>: 156-160.

Sayood, K. (1992). "Data compression in remote sensing applications." <u>IEEE Geosci. Remote Sensing Newslett.</u> 84: 7-15.

Schmid, T., M. Koch, et al. (2005). "Multisensor approach to determine changes of wetland characteristics in semiarid environments (Central Spain)." <u>Ieee Transactions on Geoscience and Remote Sensing</u> 43(11): 2516-2525.

Schowengerdt, R. A. (1997). <u>Remote sensing, models, and methods for image processing</u>. San Diego, Academic Press.

Sebban, M. and R. Nock (2002). "A hybrid filter/wrapper approach of feature selection using information theory." <u>Pattern recognition</u> 35(4): 835-846.

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

Serpico, S. B., S. Dellepiane, et al. (2012). "Information Extraction From Remote Sensing Images for Flood Monitoring and Damage Evaluation." Proceedings of the Ieee 100(10): 2946-2970.

Serpico, S. B. and G. Moser (2007). "Extraction of spectral channels from hyperspectral images for classification purposes." Ieee Transactions on Geoscience and Remote Sensing 45(2): 484-495.

Shanmugam, S. and P. SrinivasaPerumal (2014). "Spectral matching approaches in hyperspectral image processing." International Journal of Remote Sensing 35(24): 8217-8251.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." Bell System Technical Journal 27(3): 379-423.

Shaw, G. A. and H.-h. K. Burke (2003). "Spectral Imaging for Remote Sensing." LINCOLN LABORATORY JOURNAL 14(1): 3-28.

Sirkin, R. M. (2005). Statistics for the Social Sciences, SAGE Publications.

Skoog, D. A., F. J. Holler, et al. (2007). Principles of instrumental analysis. Belmont, CA, Thomson Brooks/Cole.

Skurichina, M., P. Paclik, et al. (2004). "Selection/extraction of spectral regions for autofluorescence spectra measured in the oral cavity." Structural, Syntactic, and Statistical Pattern Recognition, Proceedings 3138: 1096-1104.

Somol, P., P. Pudil, et al. (2004). "Fast branch & bound algorithms for optimal feature selection." IEEE Transactions on pattern analysis and machine intelligence 26(7): 900-912.

Sönmez, N. K., Y. Emekli, et al. (2010). "Relationship between spectral reflectance and water stress conditions of Bermuda grass (Cynodon dactylonL.)." New Zealand Journal of Agricultural Research 51(3): 223-233.

Sotoca, J. M., F. Pla, et al. (2007). "Band selection in multispectral images by minimization of dependent information." Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews 37(2): 258-267.

Soukup, M., J. Gailis, et al. (2016). HyperScout: Onboard Processing of Hyperspectral Imaging Data on a Nanosatellite. Small satellite systems and services. MALTA.

Spence, C. and P. Sajda (1998). The role of feature selection in building pattern recognizers for computer-aided diagnosis. Proceedings of SPIE.

Studholme, C., D. L. G. Hill, et al. (1996). "Incorporating connected region labelling into automated image registration using mutual information." Proceedings

of the Ieee Workshop on Mathematical Methods in Biomedical Image Analysis: 23-31.

Sun, K., X. R. Geng, et al. (2014). "A New Band Selection Method for Hyperspectral Image Based on Data Quality." Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6): 2697-2703.

Swain, P. H. and S. M. Davis (1978). Remote Sensing: The Quantitative Approach. New York, NY, McGraw-Hill.

Tan, K., E. Z. Li, et al. (2014). "Hyperspectral Image Classification Using Band Selection and Morphological Profiles." Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(1): 40-48.

Taubman, D. and M. Marcellin (2012). JPEG2000 image compression fundamentals, standards and practice: image compression fundamentals, standards and practice, Springer Science & Business Media.

Thenkabail, P., P. GangadharaRao, et al. (2007). "Spectral matching techniques to determine historical land-use/land-cover (LULC) and irrigated areas using time-series 0.1-degree AVHRR Pathfinder datasets." Photogrammetric Engineering & Remote Sensing 73(10): 1029-1040.

Thomson, C., L. Lue, et al. (2014). "Mapping continuous potentials to discrete forms." Journal of Chemical Physics 140(3).

Timme, N., W. Alford, et al. (2014). "Synergy, redundancy, and multivariate information measures: an experimentalist's perspective." Journal of Computational Neuroscience 36(2): 119-140.

Vane, G., R. O. Green, et al. (1993). "The Airborne Visible Infrared Imaging Spectrometer (Aviris)." Remote Sensing of Environment 44(2-3): 127-143.

Vishnu, S., R. R. Nidamanuri, et al. (2013). "Spectral material mapping using hyperspectral imagery: a review of spectral matching and library search methods." Geocarto international 28(2): 171-190.

Wang, C., M. Menenti, et al. (2007). "Mapping mixed vegetation communities in salt marshes using airborne spectral data." Remote Sensing of Environment 107(4): 559-570.

Wang, J. and C. I. Chang (2006). "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis." Ieee Transactions on Geoscience and Remote Sensing 44(6): 1586-1600.

Wang, L. and C. Zhao (2015). Hyperspectral Image Processing, Springer Publishing Company, Incorporated.

Ward, J. H. (1963). "Hierarchical Grouping to Optimize an Objective Function." Journal of the American Statistical Association 58(301): 236-&.

Watanabe, S. (1960). "Information Theoretical Analysis of Multivariate Correlation." Ibm Journal of Research and Development 4(1): 66-82.

Webster, R. (1985). "Quantitative spatial analysis of soil in the field." Advances in Soil Science 3: 1-70.

Weng, Y. L., P. Gong, et al. (2008). "Soil salt content estimation in the Yellow River delta with satellite hyperspectral data." Canadian Journal of Remote Sensing 34(3): 259-270.

Whitney, A. W. (1971). "Direct Method of Nonparametric Measurement Selection." Ieee Transactions on Computers C 20(9): 1100-&.

Winkler, G. and V. Liebscher (2002). "Smoothers for discontinuous signals." Journal of Nonparametric Statistics 14(1-2): 203-222.

Xiang, S., F. Nie, et al. (2012). "Discriminative least squares regression for multiclass classification and feature selection." Neural Networks and Learning Systems, IEEE Transactions on 23(11): 1738-1754.

Yang, H., Q. A. Du, et al. (2011). "An Efficient Method for Supervised Hyperspectral Band Selection." Ieee Geoscience and Remote Sensing Letters 8(1): 138-142.

Yuan, Y., G. K. Zhu, et al. (2015). "Hyperspectral Band Selection by Multitask Sparsity Pursuit." Ieee Transactions on Geoscience and Remote Sensing 53(2): 631-644.

Zehentbauer, F. M. and J. Kiefer (2012). "Detection of unexpected species in soft modelling of vibrational spectra." Chimica Oggi-Chemistry Today 30(3): 54-56.

Zhang, L. F., L. P. Zhang, et al. (2013). "A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images." Isprs Journal of Photogrammetry and Remote Sensing 83: 30-39.

Zhang, L. F., L. P. Zhang, et al. (2015). "Compression of hyperspectral remote sensing images by tensor approach." Neurocomputing 147: 358-363.

Zhang, R. and J. Ma (2009). "Feature selection for hyperspectral data based on recursive support vector machines." International Journal of Remote Sensing 30(14): 3669-3677.

Zhang, X., S. He, et al. (2017). "Spectral sea surface reflectance of skylight." Optics express 25(4): A1-A13.

Zhou, X., M. Marani, et al. (2017). "Hyperspectral and Multispectral Retrieval of Suspended Sediment in Shallow Coastal Waters Using Semi-Analytical and Empirical Methods." Remote Sensing 9(4): 393.

Zimek, A., E. Schubert, et al. (2012). "A survey on unsupervised outlier detection in high-dimensional numerical data." Statistical Analysis and Data Mining: The ASA Data Science Journal 5(5): 363-387.

A.
Removal of noisy channels

A.1. Signal to noise ratio

Hyperspectral sensors acquire images in narrow and contiguous spectral bands throughout the visible, near infrared and shortwave-infrared regions, including several bands which do not provide surface information because of atmospheric particles, one of the most influential particle is water vapour. The effect of water vapour is significant at about 940, 1140, 1400 and 1880nm (Lillesand, Kiefer et al. 2008). Furthermore, the temporal and spatial distribution of atmospheric water vapour can vary significantly. Several atmospheric particles affect the radiance measured by hyperspectral sensors and cause weak signals at-sensor (irrespective of the surface observed), i.e. a low Signal to Noise Ratio (SNR). To avoid the effect of the noisy bands on the results obtained by different dimensionality reduction methods, they should be chosen and eliminated from a dataset. Sometimes noisy channels are selected through visual inspection. Jia et al. (Jia, Ji et al. 2012) were proposed a strategy to automatically choose channels without manual band removal by projecting the signal onto a set of wavelet bases with various scales. We calculated SNR for identifying noisy channels, as it is a good indicator that gives the amount of the contamination provided by random noise. However, there is no single definition of SNR, and it should be chosen to be meaningful for the problem at hand (Schowengerdt 1997). It is reasonable to assume that random noise has zero mean over a large area of an image, and therefore consists of positive and negative fluctuations in the noiseless signal. By definition, SNR is the ratio of the noise-free image contrast to the noise contrast. For computing the contrast of signal and noise, the standard deviation gives a reliable estimation (Schowengerdt 1997), so the formula of SNR can be as follows:

$$SNR = \frac{\sigma_{signal}}{\sigma_{noise}} \qquad (A.1)$$

Schowengerdt (Schowengerdt 1997) suggested that in the case of low noise level, it might suffice to use the entire image and a uniform area in an image for approximating the signal and the noise respectively.

An alternative, the noise level can be estimated from homogeneous areas of an image which are assumed to yield a constant signal (Duggin, Sakhavat et al. 1985; Fujimoto, Takahashi et al. 1989; Gonzalez and Woods 2008). A homogeneous area in a hyperspectral image has nearly constant reflectance. In this case, the standard deviation of the spectral radiance can be solely attributed to the noise, while the mean value (μ) reveals the signal. So, the signal to noise ratio can be estimated as:

$$SNR_i = \frac{\mu_{signal_i}}{\sigma_{noise_i}} \qquad (A.2)$$

where SNR_i is the signal to noise ratio in channel i. Deep water pixels in satellite images is considered as a suitable homogeneous area. The problem of this method is finding a perfectly homogenous area in a given scene, which in several cases is not available, due to the variety of objects on the ground having different reflectance and the coarse spatial resolutions of conventional hyperspectral sensors.

Another method is the GeoStatistical (GS) method, developed and applied to estimate the SNR for AVIRIS data (Curran and Dungan 1989). This method is preferable to the previous ones because it differentiates automatically between the underlying spatially correlated variance of interest (the true values) and the spatially uncorrelated random noise without requiring homogenous areas (Asmat, Atkinson et al. 2010). The GS method uses the semivariogram to describe the spatial dependence of the variable of interest and assumes that the nugget variance, a parameter of a model fitted to the semivariogram, is the variance of the spatially uncorrelated noise in the image. Intuitively, this is reasonable because the nugget variance is generally composed almost entirely of measurement error and micro-scale variation (Curran and Dungan 1989; Asmat, Atkinson et al. 2010). In the GS method, the SNR can be estimated by dividing the mean value by the square root of nugget variance:

$$SNR = \frac{\mu}{\sqrt{C_0}} \tag{A.3}$$

where C_0 is the nugget variance. The procedure and the rationale for using C_0 in this way are given in (Curran and Dungan 1989). After determining the SNR value for each channel, the channels which have SNR values lower than a predefined threshold are eliminated.

A.2. Implementation

The GeoStatistical method has been applied to estimate the SNR for all AVIRIS spectral channels in our study. On each image, three parallel transects were extracted across a uniform area to obtain a finer semivariogram and better estimation of the nugget value by an extrapolation. Since both the noise and the SNR are related to the land cover type (Atkinson, Sargent et al. 2005), the transects should sample the same land cover. These transects were long enough to enable the estimation of a semivariogram with at least 15 lags in the statistically significant first fifth of their length (Webster 1985). The semivariogram was calculated for each transect over the 224 wavebands and averaged. The mean signals were also calculated, and the nugget variance was determined by extrapolating the slope of the semivariogram to a zero lag. The signal and the square root of the nugget

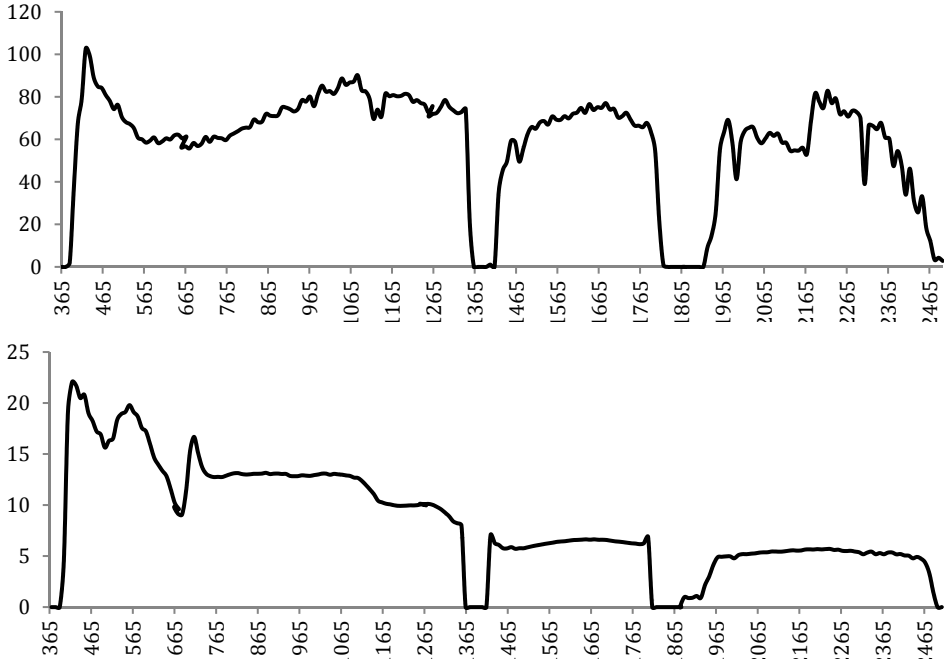variance were used to plot SNR versus wavelength for two hyperspectral scene in Figure A.1 .

Figure A.1.: Signal to noise ratio in different wavelength for Cuprite (up) and Moffett Field (down).

It is observable that the spectral regions which have low SNR value occur at almost two regions corresponding to H_2O and CO_2 absorption spectrum: around 1350-1400nm, and 1800-1900 nm. There are also some other sources of error like instrumental error affecting the signal value, and cause losing information, such as three channels at begin and end of the spectrum for two datasets.

After obtaining the SNR for every channel of hyperspectral data, the channels having low SNR were eliminated by choosing a threshold based on the visual inspection of the low-SNR channels. The thresholds were 22 and 4 for Cuprite and Moffett field respectively. So, 31 channels located at 365-385, 1353-1413, 1811-1938 and 2457-2497 nm and 29 channels placed at 370-390, 1363-1413, 1812-1949 and 2477-2507 nm were eliminated from Cuprite and Moffett scene respectively.

B.
Relation between the correlation coefficients (r) and the first eigenvalue (λ_1)

B.1. Introduction

Since $TD = f(\lambda_1, n)$, the proof of the properties mentioned in Secion 3.2. are based on the characteristics of the first eigenvalue of the correlation matrix. In (Kaiser 1968), there is a mathematical solution leading to Eq. 3.2. Below, we used a numerical experiment to prove the equation and later we provide proofs the properties of TD.
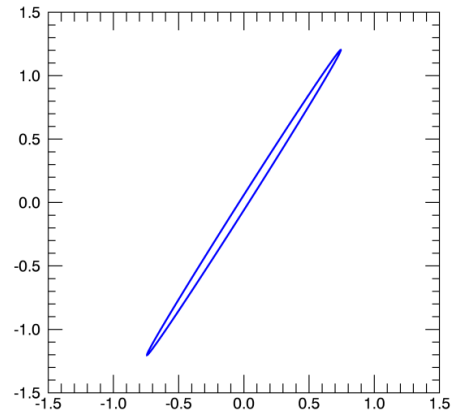

r=0.000 λ_1=1.000

r=0.333 λ_1=1.333

r=0.666 λ_1=1.666
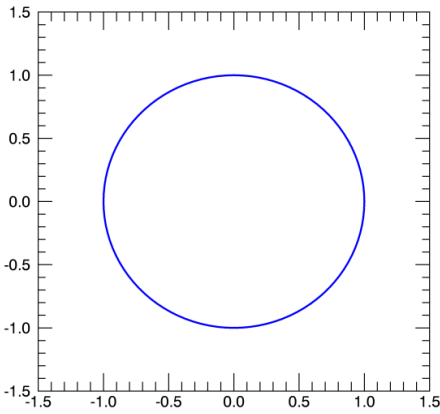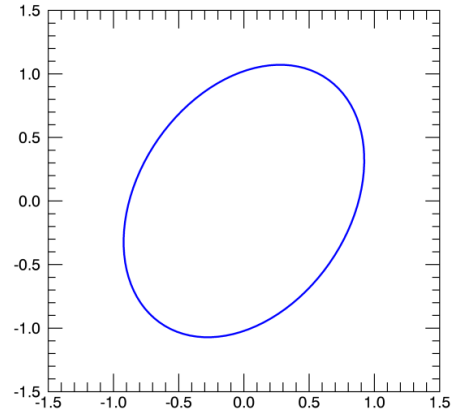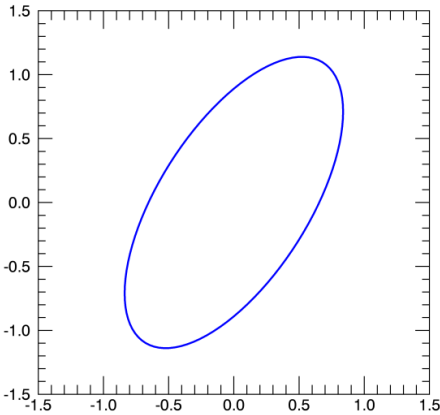
r=0.999 λ_1=1.999

Figure B.1: First eigenvalue ($\boldsymbol{\lambda_1}$) vs. the correlation coefficient (r) in a 2D space.

The eigenvalues are affected by two statistical parameter in multivariate systems: the variances and the covariances. The variance of a variable itself does not give any information about the dependence on other variables. To obtain a metric which is just influenced by the dependence of variables on each other, the correlation matrix may be applied. The 1st eigenvalue of this matrix is just a function of the dependence of variables on each other. The correlation between variables can be either positive or negative while their sign is not relevant to evaluate their interdependence. Consequently, the absolute values of matrix elements are used to compute the first eigenvalue. The first eigenvalue of an absolute correlation matrix is an approximation of the mean correlation of the variables (Friedman and Weisberg 1981). Figure B.1 demonstrates the relation between the correlation coefficient (r) and the first eigenvalue (λ_1) in a 2D feature space.

This concept is developed to explain the dependence in n-dimensional spaces. For ease of understanding in the following explanations, the assertion is proven analytically in the lowest possible dimension (2D), and for higher dimensions it is investigated numerically.

B.2. Two dimensinal space analysis

Let us consider two bands each consisting with pixel-wise intensities in m pixels, i.e. X = {x_1, x_2, ..., x_m} and Y = {y_1, y_2, ..., y_m}. The correlation matrix of these two variables is $\Sigma = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$. This matrix is symmetric. Solving the following equation gives the eigenvalues (λ's):

$$det(\Sigma - \lambda I) = 0 \tag{B.1}$$

where I is the identity matrix and *det(.)* is the determinant of Σ. The parametric solution of the equation gives the unknowns λ's :

$$\lambda = 1 \pm |r| \tag{B.2}$$

There is a linear relation between the correlation coefficient and the two roots of λ. We consider the first component (the greatest eigenvalue), max(λ) = λ_1. This value is obtained from Eq. A.2 with the plus sign between the parts of the equation. It shows that by an increase of the correlation, the first eigenvalue also increases. It means that if there is no dependence between the two variables (r = 0), λ_1 has its minimum value which is equal to one. On the other hand, if the dependence is higher, λ_1 is larger as well. The maximum λ_1 is when |r| = 1. This is the state where the maximum dependence is experienced.

B.3. Higher dimensional spaces

When the number of bands is larger than two, the analytical solution to obtain the variation of the first eigenvalue as a function of the pairwise dependence requires a long parametric expression. For this reason, we applied a numerical analysis in 3D space. A 3×3 correlation matrix is taken into account showing the pairwise correlations of three bands *X, Y,* and *Z*. This correlation matrix can be shown in a parametric way:

$$\Sigma = \begin{bmatrix} 1 & r_1 & r_3 \\ r_1 & 1 & r_2 \\ r_3 & r_2 & 1 \end{bmatrix} \tag{B.3}$$

Now, we change the correlations in equal increments within [-1,1]. Then as already mentioned, the absolute value of the correlations is taken to obtain the first eigenvalue. Given a new correlation matrix in each step, the eigenvalues are computed, and the first eigenvalue is inserted in a list with the corresponding correlations. Finally, we have a list showing different dependencies between each pair of the variables and the corresponding first eigenvalue. Now, it is possible to identify the variation of the first eigenvalue.

Figure B.2 illustrates these variations with respect to the changes of correlations between the pairs of the bands. The colors of the dots in the plot correspond to the values of the first eigenvalue; starting from one in blue to three in red.

The maximum of the first eigenvalue is obtained in the corners of the cube, which are shown with black circles. It reveals that the maximum λ_1 is achieved when the absolute values of correlation coefficients for all combinations are one. λ_1 is decreasing when the absolute values of correlations are lower (more bluish colors), and the minimum λ_1 is exactly at the center of the cube when r_1, r_2, and r_3 are equal to zero.

Figure B.3 demonstrates two cross-sections of Figure B.2, when $r_3 = 0$ and $r_3 = 1$. It shows the variation of the first eigenvalue on the changes of the correlation between two pairs of variables (r_1 and r_2), The third axis gives the changes in the first eigenvalue as a function of the correlations.

In this figure also, the lowest λ_1 is given when the correlations are zero, and the largest is at the maximum absolute correlation between the pairs. This numerical analysis can be extended for more than 3D as well, which the final result for each two pairs of variables is similar to the shapes demonstrated in Figure B.3. If the first two axes in Figure B.2 are used for correlation between the other variables the shape of the mesh always remains the same, i.e. it is always concave up. Just the concavity may be wider or narrower.
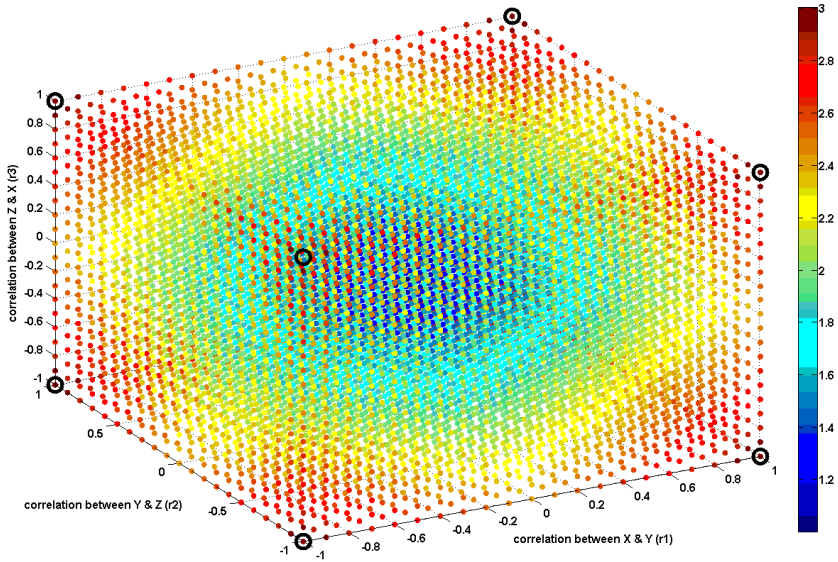
Figure B.2. Variation of the first eigenvalue with respect the correlation between pairs of variables in 3D space. The colours correspond to λ_1 which changes between one and three.
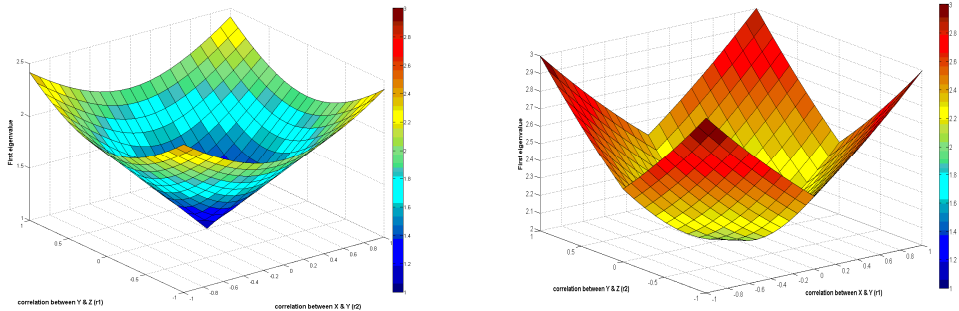

Figure B.3: Variations of the first eigenvalue with respect the correlation (r_1 and r_2) between two pairs of variables in 3D space in two cases: $r_3 = 0$ (left) , and $r_3 = 1$ (right).

B.4. The total dependence metric

Changes in the dependencies between different variables (spectral bands) affect the first eigenvalue of the data in such way that the first eigenvalue increases

when the dependence rises. However, using the first eigenvalue as a measure of dependence is not ideal, because it can get any positive real value depending on the number of bands and the correlation between the bands in the original dataset. Therefore, for having a normalized metric for dependence, the lower and upper bound of the first eigenvalue is taken into account.

In a variance-covariance matrix, the lower bound of the first eigenvalue is reached when there is no correlation between any pair of bands. In this case, the maximum eigenvalue equals the maximum variance of the original dataset. For the upper bound, the correlation between all pairs of spectral bands must be maximum. It means all the bands are completely correlated, and all observations lie on a line corresponding to the first eigenvector. In other words, there is just one eigenvalue higher than zero, and the rest are zero. Since $\sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{n} \lambda_i$, where σ^2_i is the variance of ith variable and n is the number of variables, by having just $\lambda_1 > 0$, λ_1 must be equal to the summation of all variances in the covariance matrix. Therefore the bounds on the first eigenvalue are:

$$\max_{i \in \{1,2,\dots,n\}} (\sigma^2_i) \leq \lambda_1 \leq \sum_{i=1}^{n} \sigma_i^2 \tag{B.4}$$

where the lower bound applies to a fully independent dataset, and the upper one to a completely dependent dataset. Using correlation matrix instead of covariance matrix changes Eq. A.4 to:

$$1 \leq \lambda_1 \leq n \tag{B.5}$$

where n is the number of bands. Hence, the total dependence metric (Eq. 3.2) is derived from Eq. A.5 by a simple normalization. Now, the properties of the metric given in Section 2 are proven as follows:

Proof.

(a) Eq. A.5 can be written as $0 \leq \lambda_1 - 1 \leq n - 1 \Rightarrow 0 \leq \frac{\lambda_1 - 1}{n-1} \leq 1$.

(b) Eq. A.2 gives $\lambda_1 = 1 + |r|$, by substituting in Eq. 3.2, $TD = \frac{1+|r|-1}{n-1} = |r|$.

(c) $\lambda_1 = 1$ in Eq. A.2 if and only if $X_1, X_2, \dots, X_n$ are linearly fully independent; meaning $TD = 0$, and if $r(X_1, X_2) \neq 0$ then $1 < \lambda_1 \Rightarrow TD \neq 0$.

(d) $\lambda_1 = n$ in Eq. A.2 if and only if $X_1, X_2, \dots, X_n$ are linearly fully dependent; meaning $TD = 1$, and if $r(X_1, X_2) \neq \pm 1$ then $\lambda_1 < n \Rightarrow TD \neq 1$.

(e) Since λ_1 is independent to the order of the variables so (e) follows.

(f) Since λ_1 is always a monotonic function, i.e. $\lambda_1(X_1, X_2, \ldots, X_{n-1}) \leq \lambda_1(X_1, X_2, \ldots, X_n)$. $\lambda_1(X_1, X_2, \ldots, X_{n-1}) = \lambda_1(X_1, X_2, \ldots, X_n)$, if and only if $|r(X_n, X_i)| = 0$, for every i, $i \in \{1, 2, \ldots, n-1\}$; then $(TD(X_1, X_2, \ldots, X_{n-1}) = \frac{\lambda_1 - 1}{n-2}) > (TD(X_1, X_2, \ldots, X_n) = \frac{\lambda_1 - 1}{n-1})$, so TD is not a monotonic function.

List of acronyms

APCA	Adaptive Piecewise Constant Approximation
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
B	Bhattacharyya
BB	Branch and Bound
CHRIS	Compact High-Resolution Imaging Spectrometer
CM	Clustering-Merging
D	Divergence
DR	Dimensionality Reduction
D^t	transformed Divergence
ECS	Euclidean distance of Cumulative Spectrum
ED	Euclidean Distance
EO	Earth Observing
FE	Feature Extraction
FS	Feature Selection
HSI	HyperSpectral Images
JM	Jeffreys-Matusita
LP	Linear Prediction
Mh	Mahalanobis
MLC	Maximum Likelihood Classifier
NIR	Near-InfraRed
OSP	Orthogonal Subspace Projection
PCA	Principal Component Analysis
PCFA	Piecewise Constant Function Approximations
PROBA	Project for On-Board Autonomy
RFE	Recursive Feature Elimination
RMSE	Root Mean Square Error
SA	Spectral Angle
SFFS	Sequential Forward Floating Selection
SFS	Sequential Forward Selection
SID	Spectral Information Divergence
SNR	Signal to Noise Ratio
SRS	Spectral Region Splitting
SVM	Support Vector Machine
SWIR	ShortWave InfraRed
TD	Total Dependence

VIS	VISible
VNIR	Visible and Near-InfraRed
WaLuDi	Ward's Linkage strategy using Divergence
WaLuMI	Ward's Linkage strategy using Mutual Information

List of publications

Journal publications

1. Hosseini Aria, S. E.; Menenti, M.; Gorte, B., A rapid algorithm for onborad dimensionality reduction of HyperScout images, planned.

2. Hosseini Aria, S. E.; Menenti, M.; Gorte, B., An unsupervised dimensionality reduction of hyperspectral images using the representation of reflectance spectra, under review**.**

3. Hosseini Aria, S. E., M. Menenti, et al. (2017). "Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data." Journal of Applied Remote Sensing 11(4): 046010.

4. Li, J.; Donselaar, M. E.; Hosseini Aria, S. E.; Koenders, R.; Oyen, A., Landsat imagery-based visualization of the geomorphological development at the terminus of a dryland river system. *Quaternary* International 2014, *352*, 100-110.

Conference proceedings

1. Soukup, M.; Gailis, J.; Fantin, D.; Jochemsen, A.; Aas, C.; Baeck, P.-J.; Benhadj, I.; Livens, S.; Delaure, B.; Menenti, M.; Gorte, B.; Hosseini Aria, E.; Esposito, M.; Van Dijk, C.; HyperScout: Onboard Processing of Hyperspectral Imaging Data on a Nanosatellite. In Small satellite systems and services, MALTA, 2016.

2. Conticello, S. S.; Esposito, M.; Foglia Manzillo, P.; Van Dijk, C.; Vercruyssen, N.; Baeck, P.-J.; Benhadj, I.; Livens, S.; Delaure, B.; Soukup, M.; Jochemsen, A.; Aas, C.; Gorte, B.; Hosseini Aria, E.; Menenti, M., Hyperspectral imaging for real-time land and vegetation inspection from nanosatellite. In Small satellite systems and services, ESA: MALTA, 2016.

3. Gorte, B.; Hosseini Aria, S. E.; Menenti, M. In Entropy-based noise reduction in hyperspectral images, XII International Scientific Congress and Exhibition, SIBERIA, 2016; SIBERIA, 2016.

4. Hosseini Aria, S. E.; Menenti, M.; Gorte, B. In Hyperspectral Band Selection Using the Total Dependence Metric, Hyperspectral Imaging and Sounding of the Environment, 2015; Optical Society of America: 2015; p HW3B. 3.

5. Hosseini Aria, S. E.; Menenti, M.; Gorte, B. G. H., Assessing reliability of classification in the most informative spectral regions of hyperspectral images. IOP Conference Series: Earth and Environmental Science 2014, 17, 012064.

6. Oyen, A. M.; Koenders, R.; Hosseini Aria, S. E.; Lindenbergh, R. C.; Li, J.; Donselaar, M. E. In Application of synthetic aperture radar methods for morphological analysis of the Salar De Uyuni distal fluvial system, Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 22-27 July 2012, 2012; pp 3875-3878.

7. Hosseini Aria, S. E.; Menenti, M.; Gorte, B. In Spectral discrimination based on the optimal informative parts of the spectrum, 2012; 2012; pp 853709-853709-853707.

8. Hosseini Aria, S. E.; Gorte, B.; Menenti, M. In Evaluation of Sentinel-2 Bands over the Spectrum, First Sentinel-2 Preparatory Symposium, 2012; 2012.

9. Hosseini Aria, S. E.; Donselaar, M. E.; Lindenbergh, R.; Koenders, R.; Li, J.; Oyen, A. In Monitoring a river channel network at Salar de Uyuni using Landsat ETM+ images, Proc SPIE 11/2012, 2012; 2012; pp 85310A-85310A-85319.

10. Hosseini aria, S. E.; R., A., Traffic asset database management using an enterprise GIS. In Geospatial World Forum, Hyderabad, India, 2011.

11. Hosseini Aria, S. E.; Saradjian, M. R.; Amini, J.; Lucas, C., Generalized cooccurrence matrix to classify IRS-1D images using neural network. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences 2004, 35, 117-122.

12. Hosseini Aria, S. E.; Saradjian, M. R.; Amini, J., Feature-based classification of IRS-1D satellite images using Artificial Neural Network. In International Seminar on Satellite Technology Applications in Communications and Remote Sensing, Tehran, Iran, 2004.

13. Hosseini Aria, E.; Amini, J.; Saradjian, M. In Backpropagation neural network for classification of IRS-1D satellite images, Joint Workshop

of High-Resolution Mapping from Space, Tehran University, Iran, 2003.