CRANFIELD UNIVERSITY

Ruben Moya Torres

Non-Parametric Spatial Spectral Band Selection methods

Centre for Electronic Warfare
PhD in Remote Sensing

PhD Thesis
Academic Year: 2017 - 2021

Supervisor: Dr. Peter W. Yuen
Associate Supervisor: Dr. David James
May 2021

CRANFIELD UNIVERSITY

Centre for Electronic Warfare
PhD in Remote Sensing

PhD Thesis

Academic Year 2017 - 2021

Ruben Moya Torres

Non-parametric Spatial Spectral Band Selection methods

Supervisor: Dr. Peter W. Yuen
Associate Supervisor: Dr. David James
May 2021

ABSTRACT

This project is about the development of band selection (BS) techniques for better target detection and classification in remote sensing and hyperspectral imaging (HSI). Conventionally, this is achieved just by using the spectral features for guiding the band compression. However, this project develops a BS method which uses both spatial and spectral features to allow a handful of crucial spectral bands to be selected for enhancing the target detection and classification performances.

This thesis firstly outlines the fundamental concepts and background of remote sensing and HSI, followed by the theories of different atmospheric correction algorithms — in order to assess the reflectance conversion for band selection — and BS techniques, with a detailed explanation of the Hughes principle, which postulates the fundamental drawback for having high-dimensional data in HSI. Subsequently, the thesis highlights the performances of some advanced BS techniques and to point out their deficiencies.

Most of the existing BS work in the field have exhibited maximal classification accuracy when more spectral bands have been utilized for classification; this apparently disagrees with the theoretical model of the Hughes phenomenon. The thesis then presents a spatial spectral mutual information (SSMI) BS scheme which utilizes a spatial feature extraction technique as a pre-processing step, followed by the clustering of the mutual information (MI) of spectral bands for enhancing the BS efficiency. Through this BS scheme, a sharp ’bell’-shaped accuracy-dimensionality characteristic has been observed, peaking at about 20 bands.

The performance of the proposed SSMI BS scheme has been validated through 6 HSI datasets, and its classification accuracy is shown to be ~10% better than 7 state-of-the-art BS algorithms. These results confirm that the high efficiency of the BS scheme is essentially important to observe, and to validate, the Hughes phenomenon at band selection through experiments for the first time.

Keywords: hyperspectral imaging, remote sensing, classification, band selection, band clustering, image processing, mutual information, atmospheric correction

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Peter Yuen for his guidance and patience during the development of the present project. He offered me the possibility to expand my knowledge of hyperspectral imaging, especially on the field of band selection. This achievement was only possible thanks to his constant supervision and support.

I am grateful to the DSTL for funding me to attend Cranfield University and extend my technical knowledge as one of Cranfield's students.

Finally, I would like to thank my family for their support and faith in me. Especially my parents Antonio and Isabel, who proved our affection is strong enough to withstand impossible distances.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

AC	Atmospheric Correction
AS	Angular Similarity
ATCOR	Atmospheric and Topographic Correction
ATREM	Atmosphere Removal
AVIRIS	Airborne Visible / Infrared Imaging Spectrometer
BS	Band Selection
CCA	Convex Cone Analysis
CCP	Convex Cone Projection
DN	Digital Numbers
E-FDPC	Enhanced Fast Density-Peak-based Clustering
ELM	Empirical Line Method
EMS	Electromagnetic Spectrum
ENVI	Environment for Visualizing Images
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
FF	Flat Field
FOV	Field Of View
FWHM	Full Width at Half Maximum
GSD	Ground Sampling Distance
GT	Ground Truth
HSI	Hyperspectral Imaging
IARR	Internal Average Relative Reflectance
IR	Infrared
ISSC	Improved Sparse Subspace Clustering
KNN	K-Nearest Neighbours
KSC	Kennedy Space Center
LMM	Linear Mixing Model
LSR	Least-Squares Regression
MI	Mutual Information
MODTRAN	Moderate resolution atmospheric Transmission
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NP	Non-deterministic Polinomial time

OA	Overall Accuracy
OCF	Optimal Clustering Framework
PCA	Principal Component Analysis
QUAC	Quick Atmosphere Correction
RCS	Ranking on Clusters Strategy
RGB	Red / Green / Blue
ROI	Region Of Interest
ROSIS	Reflective Optics System Imaging Spectrometer
SBSS	Saliency Bands and Scale Selection
SLN	Single Layer Neural networks
SMACC	Sequential Maximum Angle Convex Cone
SNR	Signal-to-Noise Ratio
SSMI	Spatial Spectral Mutual Information
Std. dev.	Standard deviation
SVD	Single Value Decomposition
SVM	Support Vector Machines
SWIR	Short-Wave Infrared
TIR	Thermal Infrared
USGS	United States Geological Survey
VALERI	Validation of European Remote Sensing Instruments
VNIR	Visible and Near Infrared

1 Introduction

1.1 Motivation of research

Hyperspectral Imagery (HSI) contains detailed spectral information of the scene, which enables it to detect minute, unresolved sub-pixel targets, in great contrast to the conventional texture-based image processing techniques which are not capable to achieve despite decades of development. The extreme high detectability of HSI stems from the hundreds of bands of spectral information rather than primarily relying on spatial data for target detections and classifications. Henceforth HSI has been traditionally deployed for geographical earth observations such as land use/covers with unprecedented geographical information survey (GIS) performances. The high dimension of the HSI data requires huge data storage requirements as well as quite high computational cost for data processing, making it not practical for real time applications.

That was pointed out by Hughes [1] who postulated that the accuracy of a model was supposed to be scaled non-linearly with the dimensionality of the dataset. The classification accuracy is expected to be improved when more dimensions are utilized for the classification, but further increasing the dimensionality of the data would reduce the classification accuracy when the training data size is kept constant. This occurs because when the dimensionality is too large it increases the degrees of freedom, thereby the fitting of data into the model becomes more difficult, hence the error is increased with the data dimension. This theory has predicted a ‘bell’ shape of the classification accuracy versus the dimensionality (n), however, there is no experimental band selection data to support it, although it seems to be plausible.

The motivation of this work is to examine if Hughes phenomenon is valid and to see whether many spectral bands are really needed for achieving optimal accuracy.

In addition, we need to take into account that all spectral classification work that has been carried out to-date is performed by using the reflectance of the HSI

imagery, which, have to be converted from the raw digital number (DN) or radiance of the scene by atmospheric correction through radiance transfer methods. Hence, it is important to assess the integrity of the reflectance conversion. For this reason, we have compared the accuracy of three different atmospheric correction algorithms, to show their efficiency for different datasets, calibration targets and different types of surface.

1.2 Objectives of this work

There are two main objectives at this thesis:

1. Hyperspectral band selection (BS)

   The main focus of this work is to examine whether Hughes phenomenon is valid and to see whether many spectral bands are needed for achieving optimal classification accuracy. This research develops a combination of spatial and spectral features for band selection and to achieve optimal classification accuracies, presenting a spatial spectral mutual information (SSMI) BS scheme which utilizes a spatial feature extraction technique as a pre-processing step, followed by the clustering of the mutual information (MI) of spectral bands for enhancing the BS efficiency. This method will be compared to 7 recent state-of-the-art different band selection algorithms, in order to show the highest efficiency of SSMI with respect to other BS methods, for 6 different datasets; Hughes phenomenon validity will be also analysed for the band selection results obtained by each BS algorithm.

2. Accuracy of atmospheric correction (AC)

   As mentioned before, the main part of this work is to research the methodology for achieving optimal classification accuracy in HSI data processing. However, all spectral classification work that has been carried out to-date is performed by using the reflectance of the HSI

imagery, which, have to be converted from the raw digital number (DN) or radiance of the scene by atmospheric correction through radiance transfer methods. Hence, it is fundamentally important to assess the integrity of the reflectance conversion, thereby the assessment of the accuracy of atmospheric corrections has been formulated as part of the current research; three different atmospheric correction methods have been compared for two different datasets, in order to show their efficiency respect to different calibration targets and surfaces.

1.3 List of publications

- Spatial Spectral Band Selection for Enhanced Hyperspectral Remote Sensing Classification Applications

  Ruben Moya Torres, Peter W.T. Yuen, Changfeng Yuan, Jonathan Piper, Chris McCullough and Peter Godfree

  Journal of Imaging, 2020, 6(9), 87; doi:10.3390/jimaging6090087

1.4 Structure of the thesis

The thesis study is divided into six chapters. Chapter 1 primarily gives the aim and the objectives of the present research work as well as the organization of the thesis. Chapter 2 presents in detail the background of Hyperspectral imaging techniques. Chapter 3 is allocated for the study area and data used for the present study as well as the classification methods used in the thesis work. In Chapter 4 the theory and results of the analysed atmospheric correction methods are presented. Chapter 5 concludes the research work, deals with the data analysis consisting of the creation and implementation of the SSMI band selection algorithm, and its comparison with different band selection methods, as well as discussion of the findings of the study. Finally, summary, conclusions and future work ideas are presented in Chapter 6.

2 Overview of Remote Sensing and Hyperspectral imaging (HSI)

2.1 Overview of Remote Sensing

Remote sensing is one of a suite of geospatial technologies that are having a growing impact in a wide variety of areas from commerce to science in public policy. While remote sensing's origins and technological advance are an outgrowth of military and intelligence applications, commercial and scientific applications of the technology have been growing. [2]

The field of remote sensing evolved from the interpretation of aerial photographs to the analysis of satellite imagery, and from local area studies to global analyses, with advances in sensor system technologies and digital computing. Today remote sensor systems can provide data from energy emitted, reflected, and / or transmitted from all parts of the electromagnetic spectrum. These data can then be transformed into information products either by manual or machine-assisted interpretation and employed in an analytical or management decision-making context. [2]

Examples of applications of these data include population and demography studies, study of archaeological sites, energy studies using hydrological models, urban planning, environmental treaty enforcement, or agricultural yields. [2]

2.1.1 History of Remote Sensing

Previously to the advent of modern Remote Sensing techniques, in the 1840s photographs were taken from cameras secured to tethered balloons for purposes of topographic mapping. For the next 100 years or so camera technology improved but the major advances were in the platforms used to hold the camera systems. [3]

At first people experimented with platforms such as kites, rockets, and even pigeons; a major step forward was made with the invention of the airplane and the next leap occurred when cameras could be mounted on satellites, which provided a very stable and, of course, high-altitude platform. [3] Satellites also provide an ideal platform for acquiring systematic data from around the globe which has proved invaluable for large area ecological modelling.

By the 1940s instrument research was also becoming increasingly sophisticated, pushing remote-sensing technology beyond visible-spectrum photography into infrared detection and radar systems.

Continuing this research, in 1972 the National Aeronautics and Space Administration (NASA) began the Landsat program with the launch of the Earth Resources Technology Satellite 1 (ERTS 1), which was later renamed Landsat 1. The Landsat program is now the longest running program of satellite remote sensing focused on EO. [3]

Following the launch of Landsat 1 other satellites were launched carrying different types of instruments such as radar, lidar, and more precise optical sensors.

Satellite remote sensing has evolved to the point where most environmental systems (hydrologic, atmospheric, ecosystems) now have dedicated satellite instruments recording information to help us better monitor and manage Earth's environments, and providing valuable data for use in many different remote sensing applications. [3]

Some of the main sensors available now are Landsat ETM, ENVISAT and MODIS on the optical spectral region (wavelengths ranging from 0.4 to 2.5 micrometres), and ENVISAT on the radar spectral region (wavelengths ranging from 15 to 50 cm); some of them are shown on Table 2-1: [3]

Sensor name	*Type*	*Wavelength range*	*Resolution (m)*
IKONOS	Optical	450–900 nm	1–4
SPOT5a	Optical	500–1750 nm	2.5–10
IRS-P6 – LISS-4	Optical	520–860 nm	5.8
ALOS AVNIR-2	Optical	420–500 nm	10
ASTER	Optical	520–11 650 nm	15–60
Landsat ETM+	Optical	450–2350 nm	15–30
MODIS	Optical	459–14 385 nm	250–1000
AVHRR	Optical	580–12 500 nm	1000
ENVISAT	Radar	5.7 cm (C band)	25
RADARSAT-1	Radar	5.7 cm (C band)	10–100
RADARSAT-2	Radar	5.6 cm (C band)	3–100
ALOS – PALSAR	Radar	23.5 cm (L band)	10–100

Table 2-1: Active and passive satellite-based remote sensing instruments. [3]

2.1.2 Passive and active sensors

Remote sensing instruments are often categorized as having either active or passive sensors. An active sensor generates its own signal which is subsequently measured when reflected back by the Earth's surface. A passive sensor measures solar energy that is either reflected or emitted from features on the Earth's surface. [3]

Although most passive sensors operate in the visible and infrared portions of the Electromagnetic Spectrum (EMS), there are also some passive microwave sensors in use that measure a number of parameters such as wind speed, atmospheric and sea surface temperature, soil moisture, rainfall, and atmospheric water vapour. An advantage of passive sensors is that most rely on the Sun's energy to illuminate the target and therefore do not need their own energy source so in general they are simpler instruments.

A limitation for most passive optical sensors is that they require daylight to operate, although there are some sensors that record night-time lights and clouds at night and others that record energy emitted from the Earth's surface. Since most of these sensors operate in the visible and infrared wavelengths, they are adversely affected by weather and cloud cover. Lastly, since sunlight is primarily reflected from the top of a feature, such as a forest, it is not possible to 'see' under a canopy to measure vegetation structure. [3]

To obtain this kind of information it is necessary to use active sensors. Active sensors, such as radar and lidar emit their own energy to illuminate a target and are comprised of a signal generator and receiver. They measure the strength of the returned signal and the time delay between when the instrument emits the energy and when it receives the returned pulse. These two types of information are used to describe vegetation structure.

The advantage of Radar (acronym of “radio detection and ranging”) is that these systems operate in the long-wavelength microwave portion of the EMS and thus are largely unaffected by clouds and rain. They can be considered all-weather systems. [3][4]

Lidar is an acronym for “light detection and ranging” and these systems use lasers that emit light in the visible and near-infrared portions of the EMS. In a lidar system a single light pulse can reflect off of several features in vertical space such as different layers in a forest. A single emitted pulse will result in a wave or series of returned pulses that are recorded by the detector. These return pulses can be recorded as a wave or in discrete pieces that correspond to the peaks in the returned signal.

A number of different types of lidar systems have been developed but most provide the capability to record the first and last return of the light pulse. [3][4] The returns correspond to the top of an object (i.e., top of a tree canopy) and the base substrate that the object is resting on (i.e., the ground). This is ideal for measuring the height of features such as trees or buildings.

Radar systems behave differently from optical systems with respect to how they interact with materials. The signal from most radar systems can penetrate well into a forest canopy. It is useful for this type of application, although lidar is more used for this type of application. [3][4]

2.1.3 Types of image resolution in Remote Sensing

There are different characteristics that affect the detail that can be resolved (seen) in a digital image. These are traditionally referred to as the four types of image resolution. Most people think of 'resolution' as being synonymous with spatial resolution but other 'resolution' terms can also be used as follows: [3]

- **Spatial resolution**: This is often simply referred to as 'resolution' and is the size of a pixel (smallest discrete scene element and image display unit) in ground dimensions. In most cases an image's resolution is labelled with a single number, such as 30 m, which represents the length of a side of a square pixel if it were projected onto the Earth's surface.

- **Spectral characteristics**: This includes bandwidth, band placement, and the number of bands.

  Spectral bandwidth, or spectral resolution as it is often named, refers to the range of wavelengths that are detected in a particular image band. This is effectively a measure of how precisely an image band measures a portion of the EMS.

  Band placement defines the portion of the EMS that is used for a particular image band (wavelength where each band operates). The type of application of interest determined the bands of interest.

  The last spectral variable is the number of bands. The more bands that are available the more precisely spectral properties of a feature can be measured.

- **Acquisition dynamics**: This has two components.

  The first is the temporal resolution, which is the minimum time a particular feature can be recorded twice.

  Some sensors with a wide field of view can acquire multiple images of the same area in the same day whereas some sensors have a repeat frequency of several weeks. It should also be reiterated that most remote sensing satellites have a near-polar orbit and are not able to acquire imagery at the poles since their orbit does not go over these areas.

  The other component is the timing of the acquisitions. Dynamic features such as deciduous forests and events such as flooding or fires often have an optimum time for which they should be imaged, in order to perform the change detection processes.

- **Sensitivity of the sensor**: This is defined by the dynamic range of the sensor as well as the range of digital numbers (DN) that can be used to represent the pixel values.

  Sensors have lower limits below which a signal is not registered and upper limits above which the sensor saturates and is unable to measure increases in radiance. The detail that can be measured between these extremes is determined by the range between the minimum and maximum digital numbers permitted for a particular data type.

  For example, Landsat TM data values can range from 0 to 255; these values are the minimum and maximum values that the raw data (in digital numbers) can have at this sensor.

2.2 Overview of Hyperspectral imaging (HSI)

Hyperspectral imaging is a spectral sensing technique that takes hundreds of pictures of a scene or an object at contiguous narrow wavebands in the different regions of the electromagnetic spectrum [5]. Each pixel in the resulting image reflects the spectral characteristics of the material that are present in the scene.

Hyperspectral imaging was originally developed for identifying and differentiating different materials in mining and geology; today, it is common to see this technology being exploited commercially in agriculture, food industries and several military applications such as rescues and counter terrorism operations [6].

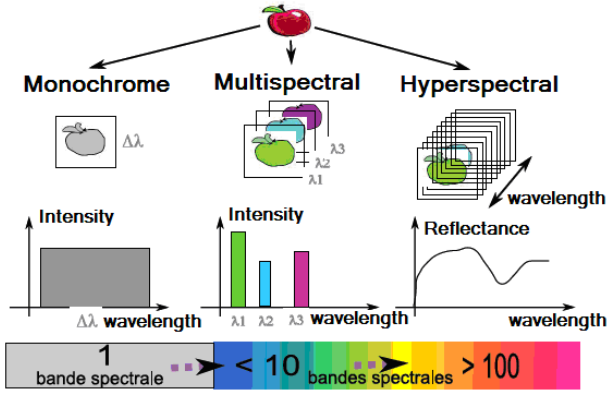


Figure 2-1: Precision and discrimination added by hyperspectral images. [5]

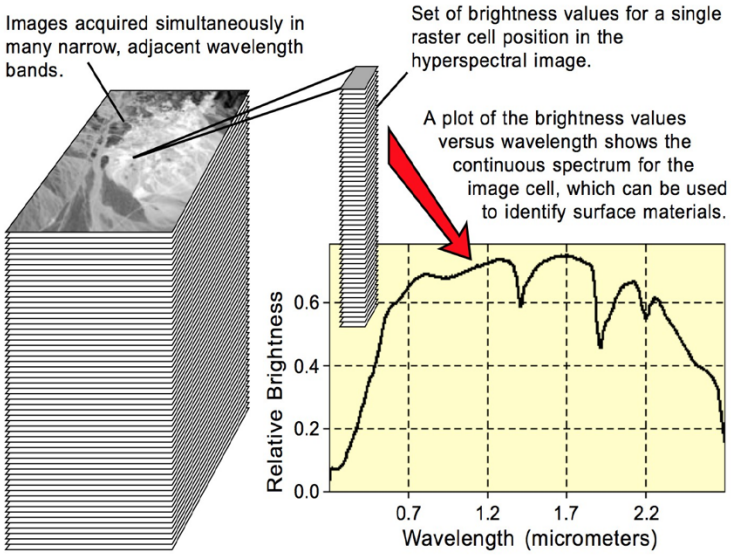


Figure 2-2: Spectral Characteristics of Pixels. [6]

Hyperspectral imaging has many fields of application; some of the main applications of HSI imaging [7] are hydrology, disaster preparedness and management, urban mapping, atmospheric study, geology, forestry, snow and ice parameterisation, soil mapping, environment, ecology, agriculture, oceanic survey, national security... Although a quite large variety of applications exists, they (and their related algorithms) can all be classified according to the application-specific tasks [7]:

- Anomaly detection: mainly used to detect small, man-made targets.
- Target recognition: finding the significant (to the user) changes between two hyperspectral scenes of the same geographic region.
- Background characterization: used normally to obtain image parameters, e.g. land surface temperatures, vegetation indices or atmospheric concentration of gases and aerosols.

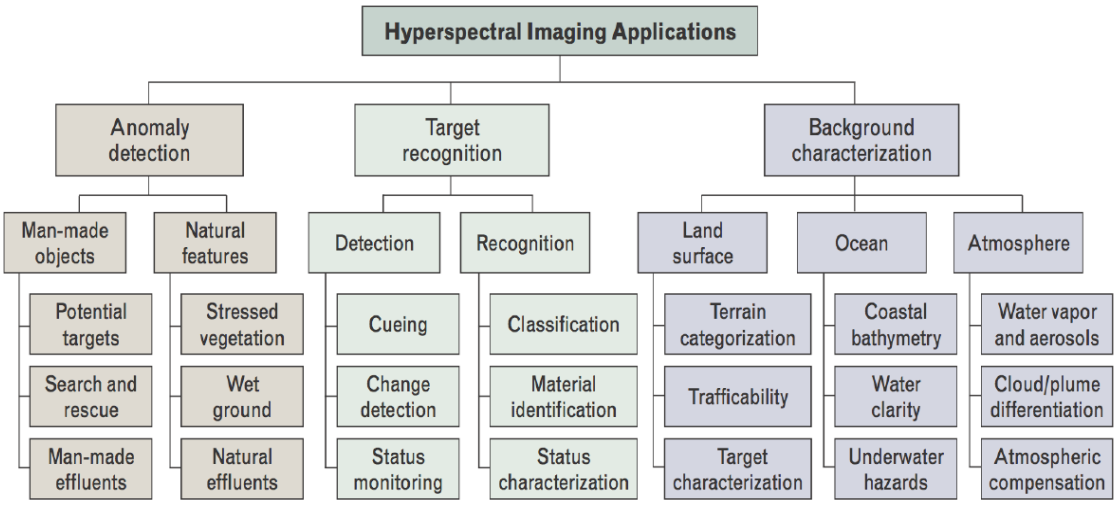


Figure 2-3: Simplified Taxonomy of HSI applications [7].

In a hyperspectral image there are normally different materials, each one with their unique molecular structure and physical shapes and sizes, thus they reflect, absorb, scatter and emit electromagnetic radiation in their own ways. Therefore, when measured at different wavelengths, the resulting spectral signature can be used to identify the different materials that appear in the scene. [7]

The concept of hyperspectral imaging is illustrated in Figure 2-4 where the data is collected using a sensor on an airborne platform [7]. The use of scan lines by the sensor allows us to obtain the spectral and spatial information simultaneously and to obtain a three-dimensional data cube. The x and y dimensions in the figure represent the spatial information (all the pixels of the image) while the spectral amplitudes of each pixel for all the bands are projected in the z dimension.

As discussed previously, hyperspectral imaging takes hundreds of images at different wavelengths; for this reason, a hypercube is expected to consist of a large number of stacked images (spectral bands). Each pixel or a group of pixels can represent a type of material and therefore, they can be selected to display its spectral signature. These provide the means for target discrimination and classification of materials in a scene. [7]

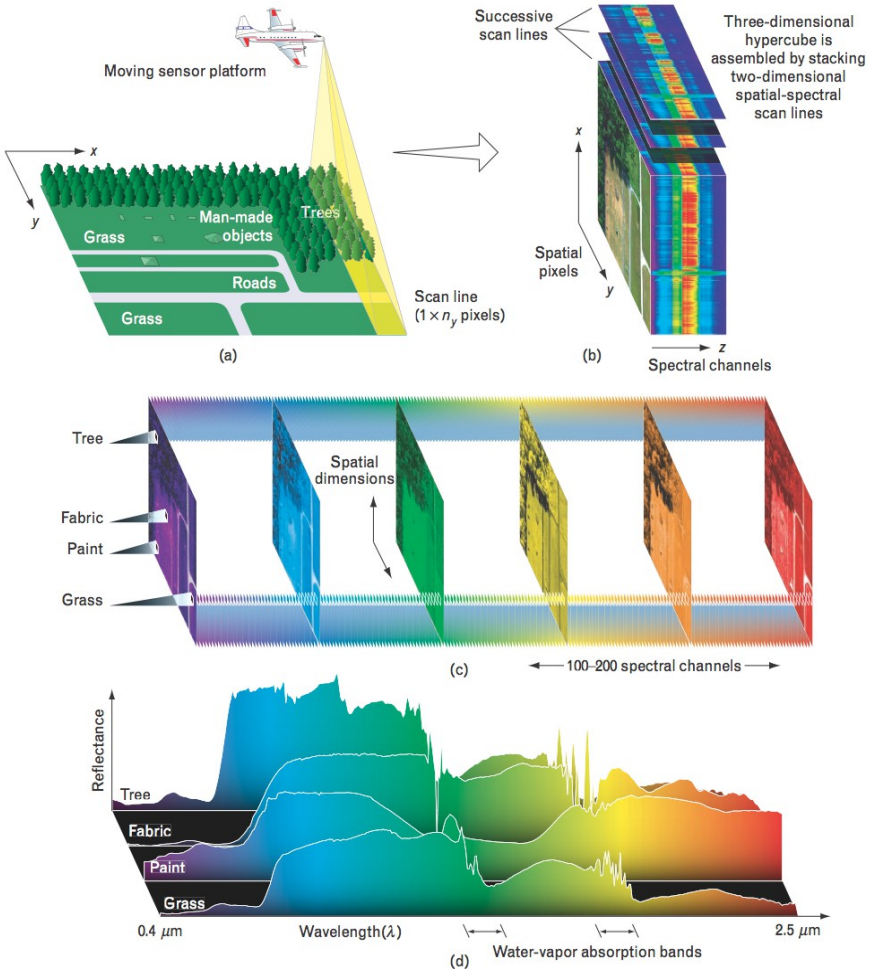


Figure 2-4: Details of a hyperspectral cube. [7]

2.3 Spectral imaging considerations

Before going straight into the details of data processing and detection, it is important to also introduce the considerations when performing spectral imaging. It is useful to go over some electro-optical quantities and their measurable definitions as indicated below [7]:

- Irradiance, which is specified in Wm^{-2}, refers to the light energy (per unit time – power) impinging on a surface, normalised by the surface area.

- Reflectance (unitless number ranging between 0 and 1) characterizes the fraction of incident light reflected by a surface. It can also be qualified by parameters such as the wavelength of reflected light, the angle of incidence, and the angle of reflection.

- Radiance, measured in $Wm^{-2}\ sr^{-1}$, is the irradiance normalised by the solid angle of the observation or the direction of propagation of the light from the source.

- Spectral radiance, specified in $Wm^{-2}\ \mu m^{-1}\ sr^{-1}$, is the radiance normalised by the wavelength of the light (specified in μm).

The composition of the material used in an object can define the colour and reflectivity because different materials react to electromagnetic radiation differently in terms of absorption, reflection and scattering. Thus, the following is a simplified expression that describes the spectral radiance (L_s) is the product of the impinging scene radiance (L_i) and the reflectance (ρ) of the material, where every variable depends on the wavelength:

$$\boldsymbol{L_s}(\boldsymbol{\lambda}) = \boldsymbol{\rho}(\boldsymbol{\lambda}) * \boldsymbol{L_i}(\boldsymbol{\lambda}) \qquad \textbf{(2-1)}$$

There are other factors not represented in the equation that can arise from environmental and sensing errors.

Firstly, although the solar radiation reaching to the earth atmosphere can be characterised, the problem lies at the part where the solar radiation propagates through the Earth's atmosphere. [7]

Secondly, errors in the sensor or in the process of sensing can distort the image, presenting an untruthful set of data. Thus, these would make it difficult to recover the real spectral reflectance of any materials. [7]

2.3.1 Factors that could distort the results of the remote sensing measures

Figure 2-5 illustrates some of the issues found when performing outdoor spectral imaging, such as spatial and spectral resolution of the sensor, atmospheric effects (e.g. absorption and scattering), spectral variability of surface materials in the scene, and other environmental effects such as viewing angle, secondary illumination, and shadowing [7].

We have represented the atmospheric transmissivity in Figure 2-6 for the spectral range of 0.3-14 μm [8]. We observe that the spectral regions with the highest transmissivities and therefore the best regions to obtain data from satellite and airborne sensors are located between 0.3-4 μm (Visible, NIR and SWIR) and between 8-14 μm (Thermal Infrared-TIR). At the time of measuring, we need to take into account the absorption bands of the water vapour and the carbon dioxide in the Visible-NIR (VNIR) and Short-Wave Infrared (SWIR) regions. The VNIR / SWIR regions are more used to applications related to anomaly detection and target detection, while the TIR region is normally used more to applications mainly related in order to obtain the land surface temperature of a hyperspectral image (e.g. detection of fires) [9]

2.3.2 Other factors

Other environmental factors that have an effect on spectral imaging include the angle of the sun, the viewing angle of the imager, and the surface orientation of the target; they can affect the amount of light to be reflected back into the spectral imager. In addition, passing clouds can cast shadows on the target, which will subsequently change the illumination on the target. Nearby objects such as windows on buildings and cars, may also reflect or scatter sunlight onto the target. [7]

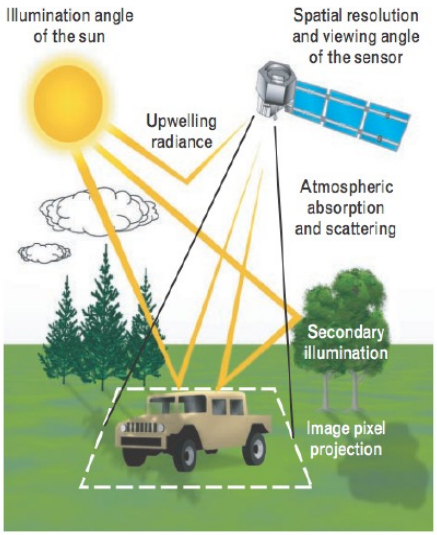


Figure 2-5: Practical issues during spectral imaging. [7]

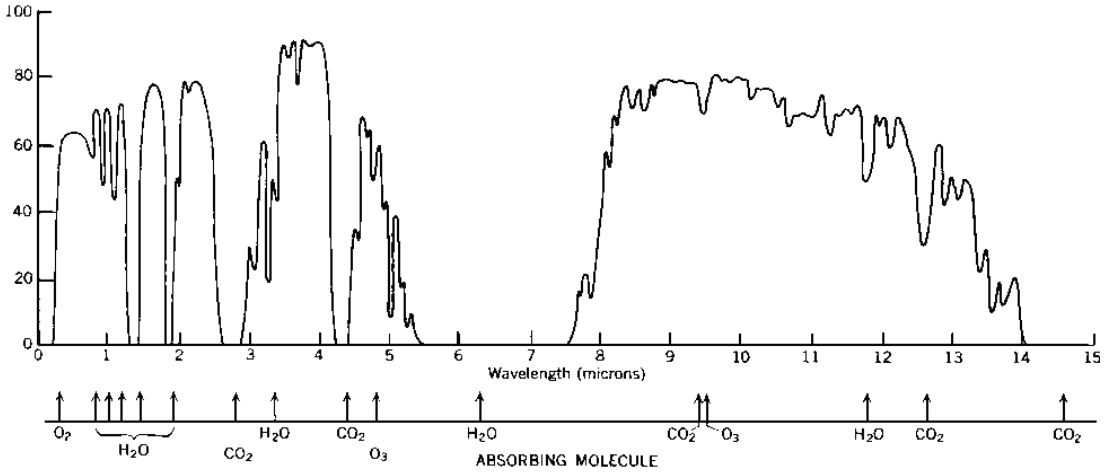


Figure 2-6: Atmospheric transmissivity of the electro-magnetic radiation in the range of wavelengths between 0.2 and 14 µm. [8]

2.4 Spectral unmixing for subpixel classifications

Spectral imaging sensors often record scenes in which numerous material substances contribute to the spectrum measured from a single pixel. Given such mixed pixels, we want to identify the individual constituent materials present in the mixture, as well as the proportions in which they appear. Spectral unmixing is the procedure by which the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, or endmembers, and a set of corresponding fractions, or abundances, that indicate the proportion of each endmember present in the pixel. [10] [11]

Endmembers normally correspond to familiar macroscopic objects at the scene, such as water, soil, metal, or any natural or man-made material. Unmixing provides a capability that is important in numerous tactical scenarios where subpixel detail is valuable. [11]

The most significant products of multispectral data processing have been classification maps that assign a class label to each pixel in an image. The relatively low numbers of spectral bands in multispectral sensors (usually no more than a dozen) have proved sufficient to provide classification maps for large scenes with numerous applications [12] to agriculture, forestry, oceanography, and environmental management and protection. [13] [14] [15]

Pixel-wise classification identifies the material class that a pixel spectrum most closely resembles, but it does not yield any further insight into the other substances— if any—that might also reside within the boundaries of the pixel. [16] Mixed pixels contain a mixture of more than one distinct material substance; [17] if the spatial resolution of an imaging sensor is low enough such that adjacent endmembers can jointly occupy a single pixel, then the resulting spectral measurement will be a composite of the individual endmembers.

The unmixing problem can be decomposed as a sequence of three consecutive stages: dimension reduction, endmember determination and inversion [10].

In the dimension reduction, we decrease the number of spectral bands with respect to the original image in order to reduce the computational cost of the algorithms. The endmember determination estimates the set of endmembers. These endmembers are the purest possible pixels for each class, although in many occasions, due to the sensor resolution or the class distribution it is not possible to find totally pure pixels in the scene. Finally, the inversion stage produces abundance planes that provide estimates of the fractional abundances for the endmembers in each pixel. [10]

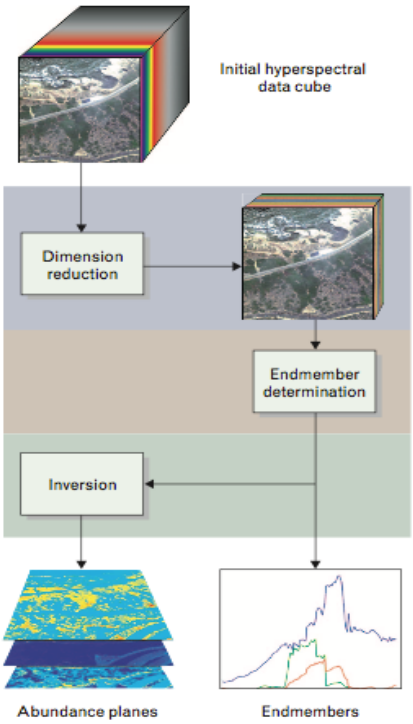


Figure 2-7: Main stages of the spectral unmixing process in a hyperspectral image [10].

2.5 Mixing models

Any approach for effectively unmixing hyperspectral data must begin with a model describing how constituent material substances in a pixel combine to yield the composite spectrum measured at the sensor. Mixing models attempt to represent the underlying physics that are the foundation of hyperspectral phenomenology, and unmixing algorithms use these models to perform the inverse operation, attempting to recover the endmembers and their associated fractional abundances from the mixed-pixel spectrum.

Figure 2-8 shows the two categories of mixing models— the linear mixing model and the non-linear mixing models [10].

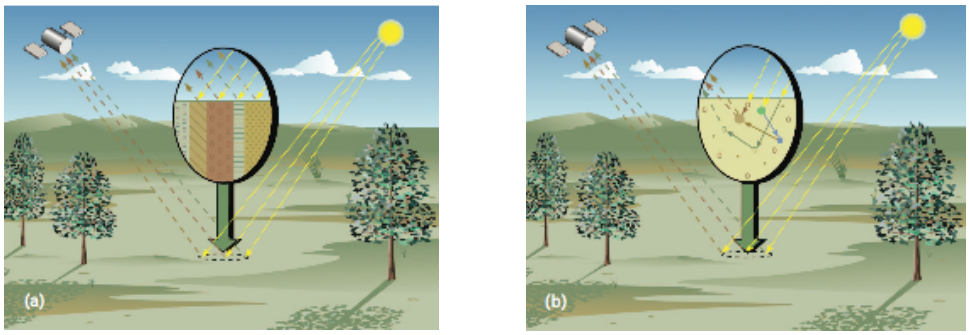


Figure 2-8: Representation of a linear mixing model (a) and a non-linear mixing model (b) related to their behaviour with respect to the solar and sensor irradiances [10].

The linear mixing model assumes a well-defined proportional checkerboard mixture of materials, with a single reflection of the illuminating solar radiation. On the other hand, non-linear mixing models assume a randomly distributed, homogeneous mixture of materials, with multiple reflections of the illuminating radiation. [10]

These models represent the underlying physics at the foundation of hyperspectral phenomenology. Unmixing algorithms use these models to recover endmembers and associated abundances from the mixed-pixel spectrum.

2.5.1 Linear mixing model

The dynamics behind the mixing of two or more substances depends largely upon the kind of mixture within a pixel that scatters the incident solar radiation. Figure 2-8(a) illustrates the reflecting surface as a mixture of endmembers, and the incident radiation scatters only once upon its surface. [10]

By this model, if the total surface area is divided proportionally according to the fractional abundances of the constituent substances, then the reflected radiation conveys with the same proportions the characteristics of the associated materials. In this sense, a linear relationship exists between the fractional abundance of the substances comprising the area being imaged and the spectrum of the reflected radiation. [10]

If we have *K* spectral bands, and we name the i_{th} endmember spectrum as $\mathbf{s}_i$ and the abundance of the i_{th} endmember as a_i, the observed spectrum **x** for any pixel in the scene can be expressed as: [10]

$$\mathbf{x} = \mathbf{a_1 s_1} + \mathbf{a_2 s_2} + \cdots + \mathbf{a_M s_M} + \mathbf{w} = \sum_{\mathbf{i=1}}^{\mathbf{M}} \mathbf{a_i s_i} + \mathbf{w} = \mathbf{Sa} + \mathbf{w} \qquad \textbf{(2-2)}$$

where **M** is the number of endmembers, **S** is the matrix of endmembers, and **w** is an error term to take the additive noise into account.

2.5.2 Non-linear mixing model

In the figure 2-8(b) seen previously, we observe that the arrangement of the constituent material substances is not as ordered as in the figure (a) because the substances comprising the soil are not organized proportionally on the surface. This mixture of materials happens when each component is randomly distributed in a non-homogeneous way. [10]

As a result, the incident radiation can experience reflections with multiple substances, and the aggregate spectrum of reflected radiation may no longer uphold the linear proportions (either in mass fraction or in volume) of the constituent substance spectrum. Because the LMM (Linear Mixing Model) is inappropriate to describe this interaction, this scenario, which has many variations, is referred to as non-linear mixing. [10]

2.6 Spectral unmixing algorithms

2.6.1 SMACC method

In this method, the endmember extraction is performed based on a convex cone model in order to represent vector data. In SMACC the endmembers are selected directly from the dataset [18].

The algorithm for finding the endmembers is sequential: the convex cone model starts with a single endmember and increases incrementally in dimension. At the same time, the abundance maps are updated at each new step. To identify new endmembers, this method uses the angle made with the existing cone as a main parameter. The algorithm chooses the vector that makes the maximum angle with the existing cone as the next endmember to add to enlarge the endmember set.

To organize hyperspectral in matrix form, we are going to assign the spectral channels to matrix rows and scene pixel spectra to columns. In the image matrix, H, the element, $H_{i,j}$, is the radiance in the i_{th} channel of the j_{th} pixel. Each row of *H* is a channel image and each column contains the spectrum of a single pixel. [18] The data matrix will be represented by a column or row expansion; that expansion in a spectral background, *S*, leads to:

$$\boldsymbol{H_{i,j}} = \sum_{\boldsymbol{n}}^{\boldsymbol{N}} \boldsymbol{S_{i,n}} \boldsymbol{F_{n,j}^{N}} + \boldsymbol{R_{i,j}^{N}} \qquad \textbf{(2-3)}$$

where N is the expansion length, F^N is a matrix of expansion coefficients that contains the contribution of each spectrum to each pixel and R^N is the error matrix resulting from truncation of the expansion to a set of *N* basis functions.

2.6.2 CCP method

On this section, a new method to find the endmember spectra in order to unmix or use as a target vectors for classification is presented [19]. This method, called convex cone analysis (CCA), is based on the fact that some physical quantities such as radiance are always non-negative. The vectors formed by discrete radiance spectra are linear combinations of positive components, and they lie inside a non-negative, convex region. The object of CCA is to find the boundary points of this region, which can be used as endmember spectra for unmixing or as target vectors for classification.

To implement this concept, the eigenvectors of the sample spectral correlation matrix of the image are found. Given the number of endmembers or classes, we

select as many eigenvectors corresponding to the largest eigenvalues. These eigenvectors are used as a basis to form linear combination that have only positive elements, and thus they lie inside a convex cone.

The vertices of the convex cone will be those points whose spectral vector contains as many zero elements as the number of eigenvectors minus one. Accordingly, a mixed pixel can be decomposed by identifying the vertices that were used to form its spectrum. [19]

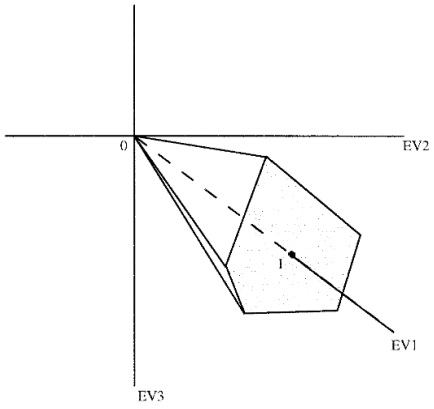


Figure 2-9: Representation of a three-dimensional convex cone, where the eigenvectors (EV) 1–3 are the axes defined by the eigenvectors of the spectral correlation matrix. [19]

2.6.3 N-FINDR method

One of the main methods for endmember determination was developed by Winter (1999) [20]; this algorithm is based upon the geometry of convex sets in order to find a unique set of the purest pixels inside each dataset. The usefulness of this technique is higher for the endmember determination and spectral unmixing for several datasets when we do not have *a priori* information of the constituent spectra (unsupervised case).

This algorithm is based on the fact that in N spectral dimensions, the N-volume contained by a simplex formed of the purest pixels is larger than any other volume formed from any other combination of pixels. The algorithm works by “inflating” a simplex inside the data, beginning with a random set of pixels. [20]

For each pixel and each endmember, the endmember is replaced with the spectrum of the pixel and the algorithm performs a volume recalculation. If it increases, the spectrum of the new pixel replaces that endmember. This procedure is repeated until no more replacements are done. The key part of the computation efficiency of this method is the application of some linear algebra procedures in order to reduce the number of operations performed by the algorithm, allowing it to reduce the processing time that takes to process all the image pixels and endmembers. [20]

Generally, the spectra for a given pixel in an image is assumed to be a linear combination of the endmember spectra:

$$p_{ij} = \sum_{k} e_{ik} c_{kj} + \varepsilon \qquad \textbf{(2-4)}$$

where p_{ij} is the i_{th} band of the j_{th} pixel, e_{ik} is the i_{th} band of the k_{th} endmember, c_{kj} are the mixing proportions for the j_{th} pixel from the k_{th} endmember, and ε is the Gaussian random error (assumed to be small).

Since the pixel compositions c_{kj} are assumed to be percentages, the mixing proportions are assumed to sum to one:

$$\sum_{k} c_{kj} = 1 \qquad \textbf{(2-5)}$$

If the number of image bands is exactly one less than the number of endmembers, the pixels in an image occupy a space formed by a simplex. A simplex is the simplest geometric shape that can enclose a space of a given dimension. [20] The simplex can be a line (1-D), a triangle (2-D) and a tetrahedron (3-D).

These simplex's vertices are the endmember spectra (one of which is the specification about how the sensor respond to a totally dark pixel, a "shade" point). Given the great number of pixels present in a large image, the probability

that some of the pixels will be pure (unmixed) is quite high; those pure pixels will be composed of only one of the endmembers present on the scene. [20]

Having that properties, we can simplify the process by finding the purest pixels in our image, or by analogy, finding the data points that represent the simplex' vertices containing the data. The premise that a hyperspectral image can be represented as a simplex is the fundament of the datasets geometrical interpretations. [20]

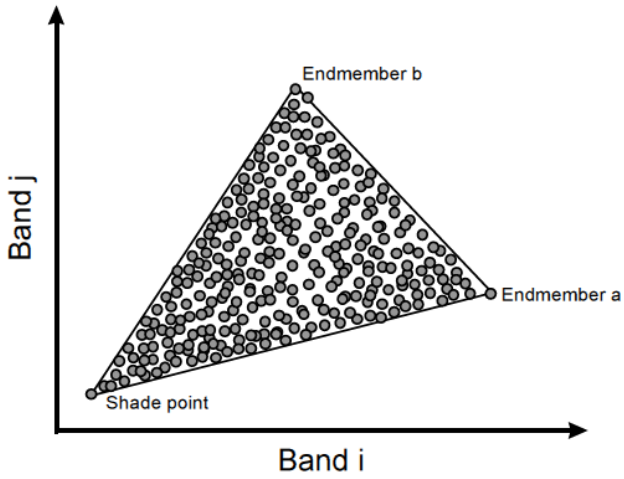


Figure 2-10: Scatter plots between two bands [20].

The scatter plots between two bands normally show a triangular shape, with the data radiating away from the shade point. The point set that specifies the simplex' vertices (the set of endmembers) possess the subsequent properties [20]:

1. They represent the point set present in the data with the largest possible volume.

2. The same pixels are endmembers if the entire data cloud suffers a translation movement.

3. The same pixels are endmembers if any or all of the axes are scaled.

4. The identity of the endmembers is constant under any rotation.

The basis of this method is the fact that the endmembers define the simplex of greatest volume.

The core of this N-FINDR algorithm is the autonomous determination of spectral endmembers. The input to this process is the full spectral image cube, which after performing the subspace projection process has been reduced in dimension. This procedure needs to examine the full dataset in order to find those pure pixels that can be used to describe the various mixed pixels in the scene. For this reason, that method is extremely slow compared to other unmixing methods, because it is necessary to search in all the pixels of the hyperspectral image in order to find the endmembers.

In addition, artificial surfaces, such as buildings and roads, can also produce characteristic unique spectra, being a possible source of problems for this algorithm, although these issues can be avoided by means of using a sensor with enough spectral resolution. Once the endmembers are found, the unmixing process of the image can be done by means of standard methods. [20]

2.7 Spectral reflectance

The radiance values captured by an HSI imager need to be converted to reflectance before comparing with the reference reflectance spectra for target identification and classification [10] [21]. For the SWIR (Short-wave Infrared) spectra between 2.0 to 2.5 μm, it is common to use conversion methodologies such as flat field conversion or average relative reflectance conversion [22] [23]; these two methods are the most used due to their minimal additive effect of atmospheric path radiance; on the other hand, the empirical line method (ELM) is typically used to account for path radiance effect between the visible and near IR ranges [24].

3 Datasets, classifiers and performance metrics used on this thesis

3.1 HSI datasets from public domains

Six widely studied publicly available datasets — Pavia University, Indian Pines, Barrax, Salinas, Kennedy Space Center, and Botswana [25], have been employed in this work for the validation of the different band selection methods used on this thesis. In addition, Barrax dataset is also used at the atmospheric correction section, to compare the results between the atmospheric correction algorithms analysed there.

Barrax dataset was acquired during the 1999–2006 VALERI campaigns [26]. This dataset was acquired by the DAIS sensor over the 5 km × 5 km Barrax site in Albacete, Spain. It consists of 400 × 400 pixels and 128 bands with 18 classes of vegetation and crops.

Pavia University hyperspectral dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. The ground sampling distance (GSD) is 1.3 m, and the dataset dimension is 340 × 610 pixels with 103 bands.

Indian Pines dataset has been one of the most widely studied imagery in the remote sensing research. It was acquired by the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor, and the imagery contains 145 × 145 pixels with 224 bands ranging from 400 to 2500 nm. There are 4 bad bands and low signal-to-noise ratio (SNR) bands due to water absorption effects such as those between (104–108), (150–163), and also the band 220: they have all been removed before the data analysis.

Salinas scene was collected by the 224-band AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor over Salinas Valley, California, and it is characterized by high spatial resolution of GSD 3.7 m with 86 lines and 83 samples. For this dataset, 20 water absorption bands between the (108–112), bands, (154–167), bands, and the 224 band have been removed, leaving 204 bands for data analysis.

Kennedy Space Center (KSC) dataset was acquired by the AVIRIS sensor in Florida on the 23 March 1996. The imagery was acquired at high altitude of 20 km with a GSD of 18 m. The dataset consists of 521 × 614 pixels in 224 bands covering a spectral region of 400–2500 nm with a narrow 10 nm FWHM per band. After removing water absorption and low SNR bands, there are 176 bands remaining for data analysis [27].

Botswana dataset was acquired by the Hyperion sensor (EO-1) on 31 May 2001, and it has dimensions of 1476 × 256 pixels in 220 bands. The GSD of Botswana is 30 m, which is the lowest spatial resolution with respect to the other 5 datasets that have been employed in this study. After removing the noisy and water absorption bands, there are 145 good quality bands remaining for data analysis.

The pseudo-RGB picture and the ground truth (GT) – pixel information obtained by direct observation and measurement – classification maps of these 6 datasets are presented in Figure 3-1, and the class information, such as the size and nature of each class, is tabulated in Table 3-1.

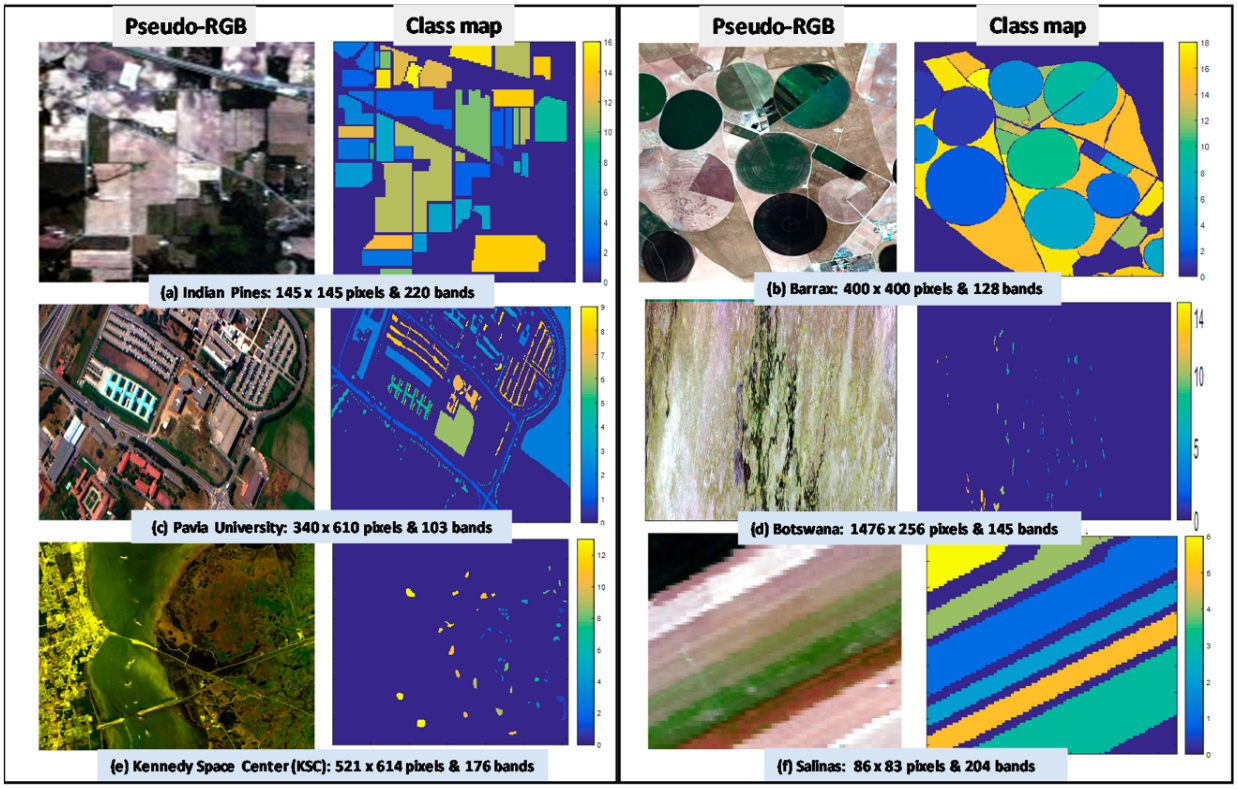


Figure 3-1: pseudo-RGB images and their corresponding ground truth (GT) maps, for all the datasets used on this thesis: Indian Pines (a), Barrax (b), Pavia University (c), Botswana (d), Kennedy Space Center (KSC) (e), and Salinas (f).

Class	Pavia University		Indian Pines		Barrax		Salinas		Kennedy Space Center		Botswana	
	Class label	Number of samples	Class label	Number of samples	Class label	Number of samples	Class label	Number of samples	Class label	Number of samples	Class label	Number of samples
1	Asphalt	6631	Alfalfa	46	Alfalfa	20606	Brocoli Green weeds_1	391	Scrub	875	Water	270
2	Meadows	18649	Corn-not ill	1428	Corn (two leaves)	13839	Corn_senesced green_weeds	1343	Willow swamp	279	Hippo grass	101
3	Gravel	2099	Corn mint ill	830	Corn (five leaves)	4921	Lettuce_romaine 4wk	616	Cabbage palm hammock	294	Floodplain grasses 1	251
4	Trees	3064	Corn	237	Corn (six leaves)	2063	Lettuce_romaine 5wk	1525	Cabbage palm / oak hammock	290	Floodplain grasses 2	215
5	Painted metal sheets	1345	Grass-pasture	483	Beet	5496	Lettuce_romaine 6wk	674	Slash pine	185	Reeds 1	269
6	Bare soil	5029	Grass-trees	730	Legumes	298	Lettuce_romaine 7wk	799	Oak / broad leaf hammock	263	Riparian	269
7	Bitumen	1330	Grass-pasture-mowed	28	Wheat	11554			Hardwood swamp	121	Fire scar 2	259
8	Self-blocking bricks	3682	Hay-windrowed	478	Experimental plots (legumes)	4965			Graminoid marsh	496	Island interior	203
9	Shadows	947	Oats	20	Experimental plots (papaver)	5118			Spartina marsh	598	Acacia woodlands	314
10			Soybean-not ill	972	Lignose	1972			Cattail marsh	465	Acacia shrublands	248
11			Soybean mint ill	2455	Vineyard	949			Salt marsh	482	Acacia grasslands	305
12			Soybean-clean	593	Test plots	3245			Mud flats	578	Short mopane	181
13			Wheat	205	Lysimeter station	534			Water	1066	Mixed mopane	268
14			Woods	1265	Water body site	62					Exposed soils	95
15			Buildings-Grass-Trees-Drives	386	Non-irrigated barley	26132						
16			Stone-Steel-Towers	93	Irrigated barley	976						
17					Bare soil	11357						
18					Ploughed soil	1196						

Table 3-1: class sizes and the nature of the class for all six datasets that have been utilized in this study.

3.2 DSTL datasets (Selene)

We will take different datasets from the Selene scene, taken on the Salisbury area in August 2014 by DSTL. The dataset used is the image called 140804_PTN_H03_03_DUAL_SCI.

This image is composed by 448 bands on the VNIR and SWIR spectral regions, between 0.4 and 2.5 μm; the dimensions of the full scene are 384 x 1876 x 448 pixels, at 944m of flight altitude (ground altitude: 50m), and it was taken on 04/08/14.

Figure 3-2: pseudo-RGB image extracted from the Selene HSI scene (bands 57, 38 and 10).

Later, we will use a smaller section of the scene: we will use the ground truth of several targets on the upper part of the HSI image of Selene:

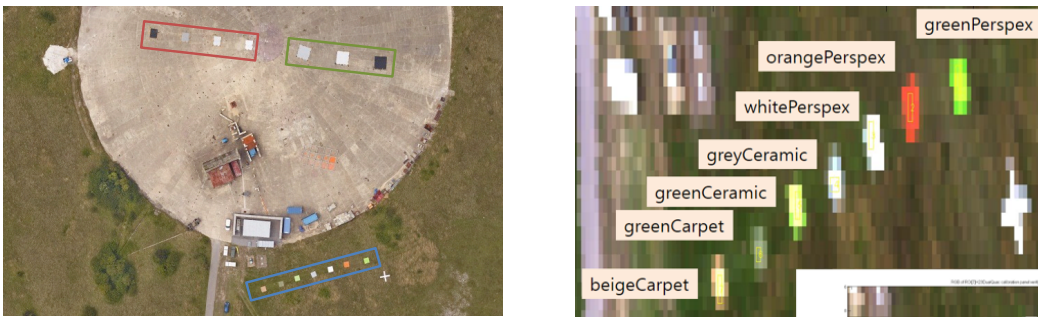


Figure 3-3: study area where several target surfaces are located, in order to perform a calibration of the Selene HSI full scene.

3.3 Classifiers

Two different classifiers (SVM and KNN) will be used on this thesis; in order to implement them, a percentage of training pixels will be taken from the scene in order to train each classifier. We will use 10% and 3% of training pixels for both cases, in order to have a fairer comparison of the band selection results, while the validation and test data will be all the available pixels for each band, for each one of the datasets mentioned at sections 3.1 and 3.2.

Respect to the number of features (number of selected spectral bands) for each case, their exact number will be selected by us; normally, we will start selecting 5 bands, continuing with 10 spectral bands, 20, 30, and so on, except at the areas where accuracy peaks are clearly visible, where we need more precision to determine the exact number of bands that maximise the accuracy results for each dataset. At these zones, we will increase the number of bands one by one, until the accuracy peak is clearly determined.

3.3.1 SVM classifier

Support vector machines (SVM) are an effective technique for solving classification and regression problems. SVM is originally an implementation of a Structural Risk Minimization (SRM) principle [28], which is known to have low

generalization error or equivalently does not suffer much from overfitting to the training dataset. [29]

A model is said to overfit or has a high generalization error if it performs poorly on instances not present in the training set. SVM is particularly effective on datasets that are linearly separable, i.e. where hyperplane H can be found that partitions the instances into two classes such that instances in one class (almost) entirely fall on one side of H.

Since there is an infinite number of candidate hyperplanes that can be selected, SVM selects the hyperplane H so that it maximizes its distance to the nearest data points in either class. This is referred to as margin maximization. [29]

However, for many real-life datasets, such a hyperplane may not exist. In these cases, SVM uses a function to map the data into a different feature space where such separability is then possible. This transformation often comes in the form of mapping to a high-dimensional space.

Thanks to these characteristics, SVM is one of the most reliable classifiers.

3.3.2 KNN classifier

The nearest neighbour approach is one of the simplest and oldest methods used for pattern classification. It often yields efficient performance and, in certain cases, its accuracy is greater than state-of the-art classifiers [30], although it normally obtains worse results than other classifiers, especially for complex images with a high number of different surfaces.

The KNN classifier categorises an unlabelled test example using the label of the majority of examples among its k-nearest (most similar) neighbours in the training set. The similarity depends on a specific distance metric; therefore, the performance of the classifier depends significantly on the distance metric used [31]. The KNN classifier is one of the most popular neighbourhood classifiers in pattern recognition [32] [33], because the technique is quite simple, and highly efficient in the field of pattern recognition, machine learning, text categorization, data mining, object recognition, etc. [34]. However, it has limitations, such as

memory requirement and time complexity, because it is fully dependent on every example in the training set.

There are two major problems inherited from the design of the KNN [33]:

1. There is no output trained model to be used; the algorithm has to use all the training examples on each test, therefore its time complexity is linear O(n). [35]
2. Its classification performance depends on choosing the optimal number of neighbours (k), which is different from one data sample to another. [35]

3.4 Performance metrics calculation

3.4.1 OA

OA is the most used tool to estimate the efficacy of a classifier. It is defined as simply the division between the correctly classified pixels, and the total number of pixels, using a Ground Truth (GT) reference image that has the correct classification result for each pixel [36]:

$$OA\ (\%) = \frac{number\ of\ correctly\ classified\ pixels}{total\ number\ of\ pixels} * 100 \quad \textbf{(3-1)}$$

3.4.2 Kappa coefficient

The Kappa Coefficient (also known as Cohen's kappa coefficient) is generated from a statistical test to evaluate the accuracy of a classification. Kappa essentially evaluates how well the classification performed as compared to just randomly assigning values for the different pixels of each dataset, i.e. if the classification did better than a random distribution [37]:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad \textbf{(3-2)}$$

where Pr (a) represents the actual observed distribution, and Pr (e) represents a random distribution, where pixels are randomly assigned to a class of the scene.

4 Atmospheric correction (AC) for HSI analysis

4.1 Why atmospheric correction is needed in this research

Atmospheric correction (AC) seems not much related to the main theme of this work, i.e., hyperspectral band selection for the reduction of dimensionality in order to improve the classification accuracy and efficiency in the processing of HSI data. However, all spectral classification work has been carried out by using the reflectance of the HSI scenes, which have to be converted from the raw digital number (DN) or radiance of the scene by atmospheric correction through radiative transfer methods. Hence, it is fundamentally important to assess the integrity of the reflectance conversion, and the accuracy of the atmospheric correction which formulates part of the current research.

4.2 Overview of AC

Atmospheric corrections try to compensate the atmospheric effects provoked by absorption and scattering, as well as variations on the illumination angle on hyperspectral images, by converting the at-sensor radiance to the target surface reflectance. [38]

The three main types of methods about atmospheric correction are the following: scene-based empirical approach, model-based and hybrid-based. The scene-based empirical approach was the first atmospheric correction type; it was developed during the early 1980s. [38] The most important scene-based algorithms are the Internal Average Relative Reflectance (IARR) [22], Flat Field (FF) [23] and Empirical Line Method (ELM) [24].

In the late 1980s, the first model-based atmospheric correction method was implemented, named as Atmosphere Removal Algorithm (ATREM) [39]. This was then followed by the DLR (German Aerospace Centre) algorithm called Atmospheric and Topographic Correction (ATCOR) [40], and later followed by the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [41]

For the hybrid approach, developers have implemented combinations of approaches based on radiative modelling together with empirical approaches, in

order to estimate the reflectance on the surface for hyperspectral scenes [42] [43], e.g. ATCOR and ELM combination [44]. The atmospheric correction method selection depends on the quality of our data, on the radiometric calibration and atmospheric parameters availability, on the previous knowledge of the scene and on our spectral measurements of the scene ground truth [45].

For the purpose of this work, we will explain the characteristics of several atmospheric correction algorithms with more detail, based on empirical approach and model-based correction types.

4.3 Variety of AC models

4.3.1 Empirical Line Method (ELM)

The Empirical Line Method (ELM) uses measurements of the scene field reflectance in order to obtain the apparent reflectance of the scene targets. Given an at-sensor raw scene as digital numbers (DN), the linear equation used to convert DN into radiance (L) is the following one: [24]

$$L = d^2(c_o + c_1 DN) \quad \textbf{(4-1)}$$

where c_o and c_1 are the offset and slope calibration coefficients respectively, d is the distance between the light source and the target (normally it is assumed to be as 1), and DN is the digital number of the pixel at-sensor measurement.

For two standard calibration panels, when their downwelling radiance and their reflectance appear in the scene, we will be able to obtain the two calibration coefficients c_o and c_1. In order to measure the offset and slope, this method needs to put at least two calibration panels (white and black spectralon) in the measurement area. [24]

With the purpose of obtaining these calibration coefficients, the ELM algorithm generates the linear fit to calculate the gain and offset, according to equation (2), from the mean spectra of both calibration panels, as shown schematically in Figure 4-1. [24] Once that we have the at-sensor radiance L, we will be able to obtain the reflectance ρ for each pixel in each band:

$$\rho = c_1\left(L - c_0\right) \tag{4-2}$$

where the term of c_1 is equal to the atmospheric effect. This method will achieve poor results if the standard calibration panels do not represent the 'true' radiance of the scene.

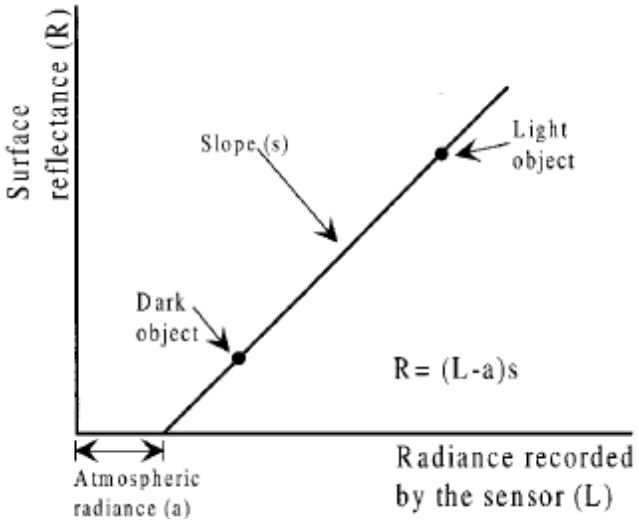


Figure 4-1: Equation formed by the white and black spectralon. [24]

4.3.2 Flat Field conversion (FF)

This method is a scene-based algorithm, as it requires the ability to identify flat field materials in the image. [23] The purpose of this method is to convert the radiance into reflectance; in order to perform this conversion, this algorithm uses the mean spectrum for our measurement area to convert the radiance image (in digital numbers, DN) into 'relative reflectance' (ρ) by equation (4-3):

$$\rho = \frac{DN - DN_D}{DN_w - DN_D} \tag{4-3}$$

where DN is the digital number pixels for each band, $\mathrm{DN_D}$ is the dark flat field target mean spectrum, and $\mathrm{DN_W}$ is the mean spectrum of a bright flat field target. [23]

It is known that a perfect dark pixel will have reflectance equal to zero in one or all bands; that property eases the process of converting the radiance into reflectance values. For this case [23], ρ now it can be simply achieved using the next expression:

$$\rho = \frac{DN}{DN_w} \tag{4-4}$$

4.3.3 Internal Average Relative Reflectance (IARR)

The purpose of this Internal Average Relative Reflectance (IARR) algorithm is to extract the spectral information contained on the raw data by means of measures of the entire image's mean spectra [22]. This global spectral mean will be used later to produce a 'relative reflectance' image for every pixel vector of the scene.

Given x as the image pixels, $i*j$ is the image coordinate location, n is the number of bands present on the image and μ_n is the global spectral mean of the entire scene for each band, this Internal Average Relative Reflectance (IARR) correction method obtains the reflectance by means of the following equation: [44]

$$\rho = \frac{x_{i*j,n}}{\mu_n} \tag{4-5}$$

This correction method is normally used when we do not have access to the atmospheric parameters and measurements of the scene ground spectra. On the other hand, in some cases it is possible to misinterpret some of the results as real spectral features, provoking large errors in this atmospheric correction method.

In order to solve this issue, Ben-Dor et. al. suggest to employ subsets of the scene for the mean estimation in order to avoid inducing additional errors [44].

4.3.4 Atmosphere Removal program (ATREM)

This algorithm called Atmosphere Removal program or ATREM uses the radiative transfer model in order to obtain the scaled surface reflectance from a hyperspectral scene [39]. The development of this atmospheric correction method started by the University of Colorado during the late 1980s. The first step is to calculate the solar zenith angle; this variable is obtained based on the acquisition time of our scene, its date and geographic location. Then, this algorithm uses the water absorption bands existent between 940 and 1140 nm, by means of extracting the water vapour column for all pixels. [39]

The next step of this method is to obtain the values of the transmissivity of the main atmospheric gases, such as water vapour, carbon dioxide, ozone, oxygen, nitrous oxide, carbon monoxide or methane; to perform this process, this method uses the radiative transfer model named as *MODerate resolution atmospheric TRANsmission* model (MODTRAN), in order to obtain 'look-up' tables generated from the absorption of water vapour, that will allow us to calculate the transmissivity of the main atmospheric gases present on our scene. [39]

Once having the solar zenith angle and the atmospheric transmissivity values, the apparent reflectance spectra are obtained based on the following expression:

$$\boldsymbol{L_o}\,(\boldsymbol{\lambda}) = \boldsymbol{L_{sun}}\,(\boldsymbol{\lambda})\,\boldsymbol{\tau}(\boldsymbol{\lambda})\,\boldsymbol{\rho}(\boldsymbol{\lambda})\,\cos\boldsymbol{\theta} \qquad \textbf{(4-6)}$$

where:

$L_o\,(\lambda)$: observed at-sensor radiance

$L_{sun}\,(\lambda)$: solar radiance above the atmosphere

$\tau(\lambda)$: total atmospheric transmittance

$\rho(\lambda)$: surface reflectance

θ : incidence angle

However, this code is rarely used for hyperspectral analysis, due to its inefficiency for atmospheric correction [39].

4.3.5 Atmospheric and Topographic Correction (ATCOR)

In a similar way to other model-based atmospheric correction methods, this algorithm known as Atmospheric and Topographic Correction (ATCOR) estimates the scene reflectance by means of the MODTRAN model of atmospheric propagation on a pixel-by-pixel basis, given the surface radiance (L) for a flat terrain background [40], as seen on Figure 4-2:

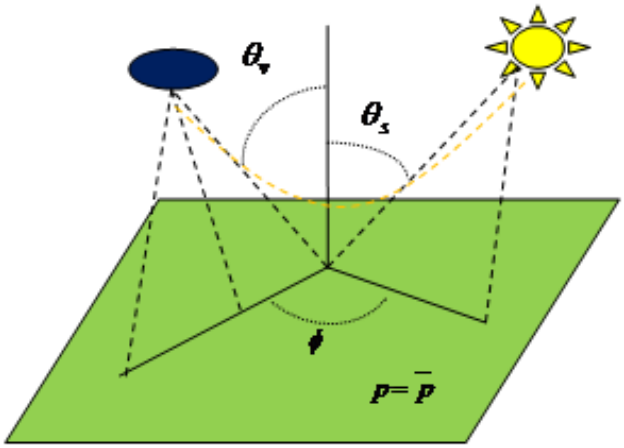


Figure 4-2: Illumination and viewing geometry for ATCOR atmospheric correction [40].

$$L = L_p(\theta_v, \theta_s, \phi) + \tau_v(\theta_v)\frac{\rho}{\pi(1-\rho_r)s}\frac{E_g(0)}{} \quad \textbf{(4-7)}$$

where:

L : at-sensor radiance for a surface reflectance, ρ

L_p : path radiance

τ_v : total ground-to-sensor atmospheric transmission

E_g : global flux of horizontal surface (sum of the direct component of the irradiance, E_{direct}, and the diffuse component, $E_{diffuse}$)

ρ : scene reflectance

ρ_r : background reflectance that determines the effective global flux

s : atmospheric spherical albedo

MODTRAN implements the atmospheric transmittance, τ_v, the direct and diffuse solar flux, E_g, and the path radiance, L_p, as functions of the solar irradiance and the viewing and illumination geometry of the Sun [40].

As shown in Figure 4-3, the at-sensor radiance consists of the next components:

L_1 : scattered radiance or path radiance,

L_2 : reflected energy coming from the target pixel, which consists of the global flux of irradiance (the sum of direct and diffuse irradiance),

L_3 : radiation reflected from the neighbouring pixel and scattered into the sensor field of view, e.g. the adjacency effect that includes the atmospheric backscattering and volume scattering.

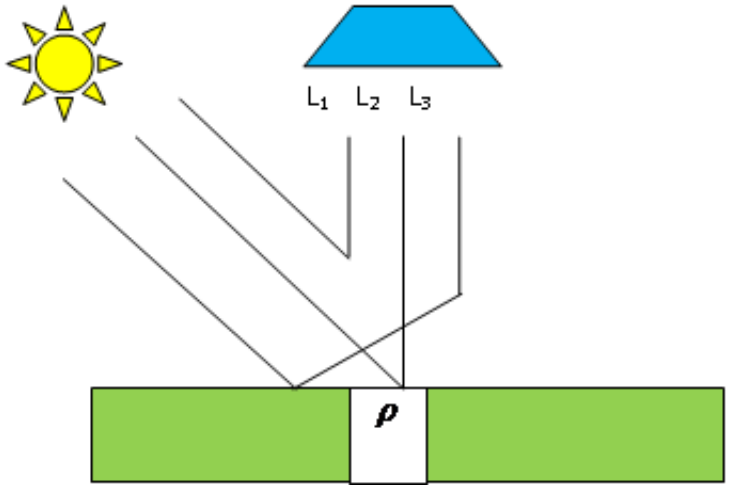


Figure 4-3: Schematic sketch of components of the solar radiation, as seen by the HSI system [40].

Once having obtained all these components, the total radiance signal L received by the sensor can be calculated as the result of the following equation: [40]

$$L = L_1 + L_2 + L_3 \qquad \textbf{(4-8)}$$

In this case, the surface properties of the target pixels are given by component L_2: L_1 and L_3 are terms that need to be removed during the process of atmospheric correction.

Respect to the adjacency radiation term, L_3 depends on the reflectance or brightness difference between the currently considered pixel and the large-scale (0.5-1 km) neighbourhood. For short-distance measurements (< 2m) between spectrometer and target, which are typically used for inflight calibration experiments, the influence of the adjacency effect is pretty small and it also decreases strongly for wavelengths higher than 1.5 µm. [45]. For this reason, we will consider L_3 as negligible on this case.

L_2 is given by:

$$L_2 = L = d^2\left(c_o + c_1 DN\right) \tag{4-9}$$

In order to solve this equation, the scene radiance can be estimated using the Empirical Line Method (ELM) algorithm, where the offset and slope are the calibration coefficients (c_0 and c_1), and DN is the value in digital numbers for each pixel.

Then, the final surface reflectance can be calculated by the combination of the expression (4-9) into (4-7), considering negligible the adjacency term: [40]

$$\rho = \frac{\pi\left[d^2\left(c_o + c_1 DN\right) - L_p\right]}{\tau_v\left(E_g\right)} \tag{4-10}$$

where:

L_p : term equal to the path radiance, L_1

E_g : global flux of horizontal surface ($E_{\text{direct}} + E_{\text{diffuse}}$)

d^2 : distance Sun-Earth (in astronomical units)

We have shown how the scene-based empirical approach can be an easier form of estimating the reflectance of a surface in remote sensing. However, this type of atmospheric correction algorithm relies on accurate spectral calibration results for the estimation of the calibration coefficients of our scene. In order to achieve better precision on the estimation of the path radiance and the surface reflectance, atmospheric and topographic effects need to be taken into account. [40].

4.3.6 Quick Atmospheric Correction (QUAC)

The atmospheric correction algorithm called Quick Atmosphere Correction (QUAC) [46] tries to convert the radiance data into reflectance by trying to find a target with near-zero reflectance, in order to determine the baseline (offset). This method is based on the empirical fact that the average reflectance spectra values of many different types of material (except highly structured surfaces, such as mud, vegetation or shallow water) does not depend on the scene.

This baseline obtained will be subtracted from the data, in order to discard for performing the procedure to obtain the gain spectrum. This variable is calculated by finding the spectral mean for all the endmembers. [46] When QUAC will have these data, it will apply the offset and gain linearly, using the same calibration expression as the ELM method, in order to derive the reflectance data.

This atmospheric correction method only uses information provided on the scene to obtain the target reflectance, information directly from the scene to allow the retrieval of target reflectance even when we do not have a radiometric or wavelength calibration for our data, or we do not have solar illumination intensity data. [46] The underlying assumption of this algorithm is that there must be a sufficiently low non-zero reflectance in order to estimate the offset correctly.

The implementation of the algorithm in ENVI software contains the following improvements to increase the accuracy of atmospheric correction [46]:

- Application of a mud filtering (defined as the Normalized Difference Mud Index, NDMI),

$$NDMI = \frac{(\rho_{795} - \rho_{990})}{(\rho_{795} + \rho_{990})} \quad \textbf{(4-11)}$$

  where ρ_{795} and ρ_{990} are the reflectances at the wavelengths of 795 and 990 nm, respectively, to exclude highly structured materials.

- Selection of endmembers based on a small subset of available bands for most sensors. When a sensor spans both the visible and NIR-SWIR spectral regions, the algorithm excludes bands in the visible region.

- Constrain of the gain curve to be constant for wavelengths below 650 nm.

- Suppression of the effects caused by dense vegetation.

- Elimination of the cloud endmembers lying between 940 to 1020 nm (water absorption bands).

4.3.7 Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH)

FLAASH is an atmospheric correction code based on MODTRAN 4. It uses the standard equation for spectral radiance at a sensor pixel, L*, in the solar wavelength range (neglecting thermal emission) from a flat Lambertian surface or its equivalent [41]:

$$L^* = A\rho /(1-\rho_e S) + B\, \rho_e /(1-\rho_e S) + L^*_a \qquad \textbf{(4-12)}$$

where ρ is the pixel surface reflectance, ρ_e is a surface reflectance averaged over the pixel on a surrounding region, S is the spherical albedo of the atmosphere, L^*_a is the radiance backscattered by the atmosphere, and A and B are coefficients that depend on atmospheric and geometric conditions but not on the surface. Each of these variables depends on the spectral channel. [41]

The first term in Eq. (4-12) corresponds to radiance that is reflected from the surface and travels directly into the sensor. The second term corresponds to radiance from the surface that is scattered by the atmosphere into the sensor, resulting in a spatial blending, or adjacency, effect. For A, B, S and L*a in Eq. (4-12), the results are dependent on the viewing and solar angles and surface and sensor elevations, as well as on the atmospheric parameters of the water vapour column, the type of aerosol used, and the visibility [41].

In order to obtain these parameters, FLAASH uses the combination of a radiance ratio and an out-of-band radiance to integrate a MODTRAN 2-D Look-Up Table for the column water vapour obtained for each pixel.

Once the atmosphere is correctly characterized and the Eq. (4-12) constants are derived, in order to calculate the image reflectance, FLAASH uses a method that involves the calculations of a spatially averaged radiance image L^*_e, [41] from which the spatially averaged reflectance ρ_e is estimated using the relationship (4-13), with the same parameters used at Eq. (4-12):

$$L^*_e \approx (A+B)\,\rho_e / (1-\rho_e\,S) + L^*_a \qquad \textbf{(4-13)}$$

Once having this averaged radiance, FLAASH performs the spatial averaging by means of a point-spread function, which describes the relative contributions to the pixel radiance from points on the ground at different distances from the direct line of sight. FLAASH approximates this function as a nearly exponential function of radial distance.

Since clouds can be a severe contaminant in the spatial averaging process for the L^*_e calculation, FLAASH automatically identifies cloudy pixels and replaces them with an average radiance [41].

FLAASH precision can be increased up to one magnitude order by using an approximation in which the convolved reflectance and water vapour are averaged within pixel groups (“superpixels”), and Eq. (4-13) is reduced to a simple linear form. This method, implemented with 4x4 superpixels, is the default in FLAASH, and is suitable for sensors with a spatial resolution better than 100 m.

4.3.8 Modtran-based atmospheric correction method

Later, we will compare the results obtained using a Modtran-based algorithm; this atmospheric correction algorithm implemented in Matlab is based on the atmospheric correction by means of the subtraction of the path radiance term [47].

In order to obtain an expression dependent on the Modtran parameters, we need to take into account that the radiances measured by a hyperspectral camera are a combination of the atmospheric path and the surface-reflected solar radiances, as seen on Figure 4-4.

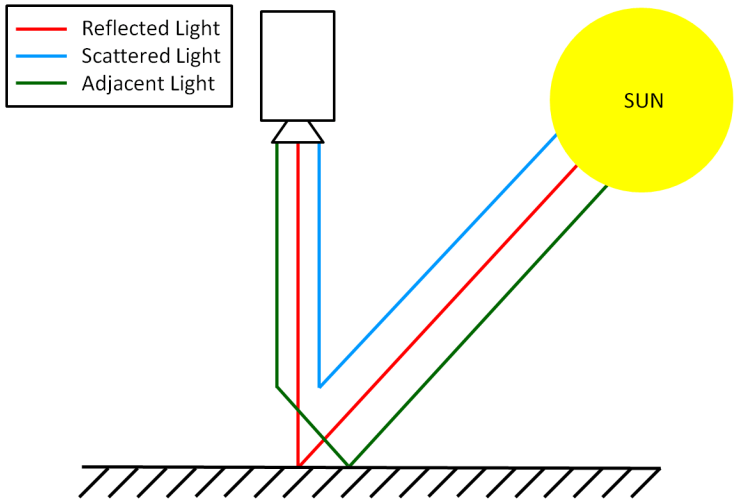


Figure 4-4: atmospheric scattering between the Sun and the sensor [38].

These radiance images can be converted to apparent reflectance (ρ_{obs}) using the next equation [38]:

$$\boldsymbol{\rho_{obs}}\left(\lambda, \boldsymbol{\theta}, \boldsymbol{\rho_{obs}}\left(\lambda, \boldsymbol{\theta}, \boldsymbol{\theta_0}, \boldsymbol{\phi_0}\right)\right) = \frac{\boldsymbol{\pi L_{obs}}(\lambda, \boldsymbol{\theta}, \boldsymbol{\theta_0}, \boldsymbol{\phi_0})}{\boldsymbol{\mu_0}\, \boldsymbol{F_0}(\lambda)} \qquad \textbf{(4-14)}$$

where:

- θ_0 is the solar zenith angle
- ϕ_0 is the solar azimuth angle
- θ is the sensor zenith angle
- ϕ is the sensor azimuth angle
- λ is the wavelength
- L_{obs} is the radiance measured at the satellite
- F_0 is the solar flux at the top of the atmosphere when the solar zenith angle is equal to zero
- μ_0 is the cosine of the solar zenith angle

According to Tanre et al. (1986) [47], when the surface is assumed to be Lambertian and the adjacency effect is neglected, $\rho_{\text{obs}}(\lambda, \theta, \theta_0, \phi_0)$ can be expressed approximately as:

$$\boldsymbol{\rho_{\text{obs}}}(\lambda, \boldsymbol{\theta}, \boldsymbol{\theta_0}, \boldsymbol{\phi_0}) = \left[\boldsymbol{\rho_{\text{atm}}}(\lambda, \boldsymbol{\theta}, \boldsymbol{\theta_0}, \boldsymbol{\phi_0}) + \frac{\boldsymbol{t_d}(\boldsymbol{\lambda}, \boldsymbol{\theta_0})\, \boldsymbol{t_u}(\boldsymbol{\lambda}, \boldsymbol{\theta})\boldsymbol{\rho}(\boldsymbol{\lambda})}{(\mathbf{1} - \boldsymbol{s}(\boldsymbol{\lambda})\boldsymbol{\rho}(\boldsymbol{\lambda}))}\right] \boldsymbol{T_g}(\lambda, \boldsymbol{\theta}, \boldsymbol{\theta_0}) \quad \textbf{(4-15)}$$

where:

ρ_{atm} is the path reflectance,

t_d is the downward scattering transmittance,

t_u is the upward scattering transmittance,

s is the spherical albedo of the atmosphere ρ is the surface reflectance, and

T_g is the total gaseous transmittance in the Sun-surface-sensor path.

Operating with this equation, we are going to obtain an expression for the reflectance on the terms of the Modtran variables [48], where L_{obs} is the observed radiance:

$$\rho_{\text{obs}} = T_g \left(\rho_{\text{atm}} + \frac{t_d\, t_u\, \rho}{(1 - s\rho)}\right) = T_g\, \rho_{\text{atm}} + \frac{T_g t_d\, t_u\, \rho}{(1 - s\rho)} = SOL\_SCAT + \frac{T_g t_d\, t_u\, \rho}{(1 - s\rho)}$$

$$\rho_{\text{obs}} - SOL\_SCAT = \frac{T_g t_d\, t_u\, \rho}{(1 - s\rho)} = TOTAL\_RAD * \rho$$

$$\boldsymbol{\rho} = \frac{\boldsymbol{L_{obs} - SOL\_SCAT}}{\boldsymbol{TOTAL\_RAD}} \quad \textbf{(4-16)}$$

However, when we tried to implement this expression on our Selene scene, we realized that we needed an extra term crucial for eliminate the albedo effect on the atmospheric correction process. We found this term on Ientilucci [49]: in order

to perform this additional correction, we will fix the ground albedo to zero, and the TOTAL_RAD value obtained on this case will be the path radiance.

Once we have those results, we will obtain the reflectance using the following expression:

$$\boldsymbol{\rho} = \frac{\boldsymbol{L_{obs}} - \boldsymbol{TOTAL\_RAD}\ (albedo=0) - \boldsymbol{SOL\_SCAT}}{\boldsymbol{TOTAL\_RAD}(albedo=1)} \quad \textbf{(4-17)}$$

Comparing this method with ATCOR, both are based on the use of Modtran to implement the atmospheric correction, but ATCOR has additional atmospheric correction tools in order to reduce the error provoked by the water vapour absorption bands between 0.8 and 1.1 μm; in addition, you can select a group of pixels in ATCOR whose ground truth reflectance results are known, in order to increase the precision of the atmospheric correction to obtain better reflectance results.

For this reason, this Modtran code is a raw version of the ATCOR algorithm, but without these additional atmospheric correction tools that ATCOR (and QUAC) possess, therefore Modtran is a worse atmospheric correction method than ATCOR and QUAC, due to not having additional tools for correcting the water absorption band effects.

4.4 Plant spectra

For some remote sensing applications, including target detection, it is important to understand the characteristic spectra of green plants [6]. At the visible portion of the spectrum (0.3-0.7 μm), the absorption effects from chlorophyll and other pigments in the leaf dominate. As known, chlorophyll in the leaves is the substance that absorbs light for photosynthesis.

The blue and red wavelengths are being absorbed by the chlorophyll more than green, which results in the peak at the green portion (0.5 μm) of the spectrum, because the reflectance (ρ) is defined as the difference between the unity and the absorptance (A), as seen on the following equation:

$$\boldsymbol{\rho}\,(\boldsymbol{\lambda}) = \mathbf{1} - \boldsymbol{A}\,(\boldsymbol{\lambda}) \tag{4-18}$$

therefore, when the absorption is minimum, the reflectance is maximum.

The 'red edge' is defined as the boundary between the red and NIR portion (0.7-0.8 μm), where reflectance values around 0.4-0.5 are observed in trees. This occurs due to the interactions with the internal cellular structure of leaves.

From 1.3μm (Middle Infrared), reflectance decreases because of the effects created by the water vapour absorption bands in 1.5 and 1.9 μm. The dry grass is affected less for this issue, because its water content is lower than green grass.

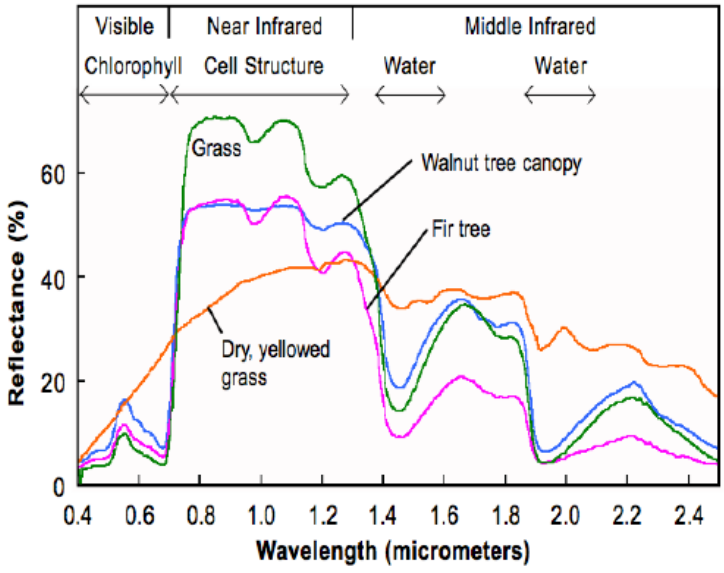


Figure 4-5: Reflectance spectra of different types of vegetation. [6]

4.5 Accuracy of atmospheric correction methods

4.5.1 Accuracy of AC for the Barrax scene

4.5.1.1 Ground truth (GT) spectra of the Barrax scene

The ground truth (GT) reflectance of the Barrax scene was acquired by the University of Valencia as part of the VALERI campaign. The GT spectra was measured for alfalfa at the locations of zones A2-1 and A2-2, and for corn at zone SF-1, as depicted in Figure 4-6; in addition, soil spectra was obtained from areas BS2, BS7, BS3 and BS9. These GT spectra will be used for assessing the accuracy of the AC methods as described in the next subsections.

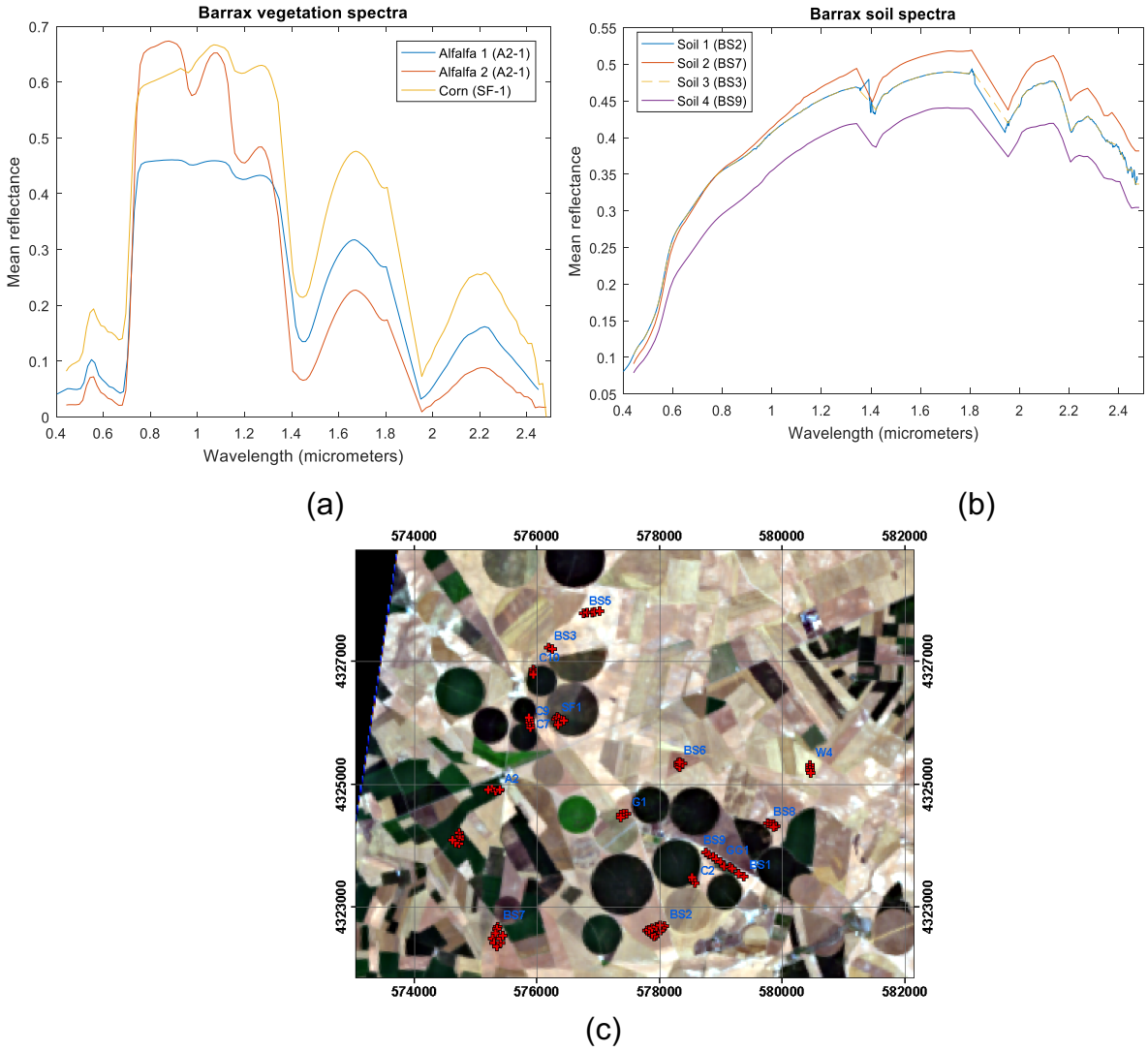


Figure 4-6: Average reflectance spectra for areas of alfalfa, corn (a) and soil (b), for the Barrax hyperspectral image, together with the location of the areas in which ground truth data was obtained for this scene (c).

4.5.1.2 ATCOR parameters

There are different user parameters in ATCOR, including some atmosphere variables such as the type of aerosol, altitude, water column density and visibility; in addition, flight properties such as the field of view (FOV) and flight directions, are needed to set up for ATCOR to work properly. The aerosol package of "h01000_wv20" with 1km altitude, water column of 2.0 g/cm^2, visibility of 35km, flight heading of 180˚ and the sensor FOV of 61 degrees, has been used for the creation of the scan profile (i.e. the *.sca file) for the AC of Barrax scene. The band at 820nm has been self-selected for the water vapour estimation, and the ATCOR utilities of "Variable water vapour" and "Haze or Sun Glint Removal" have also been selected.

Due to the relatively unknown conditions of the atmosphere at the time of data acquisition, ATCOR provides a facility for enhancing the AC accuracy through the in-situ spectral calibration. This works in a similar way to the principle where ELM is based, but it also includes the multiple scattering factor into account. Furthermore, multiple ground truth spectra can be used in order to average out the material variability issue.

In this work, the ground truth spectra of various classes (alfalfa2, corn, soil1, soil2, soil3, soil4) have been used for the calibration of ATCOR to work for the AC of the Barrax scene. The radiance of the scene data is selected manually from the locations of these classes, taking a ROI with the average of 20 x 20 pixels around the calibration pixel.

4.5.1.3 Assessment metric

The Euclidean distance d has been used for measuring the AC error:

$$d\ (\%) = 100 * (\sqrt{\sum(GT_i - x_i)})/||\text{GT}|| \qquad \textbf{(4-20)}$$

where *i* is the band number, x_i is the i_{th} band of the AC data, and GT is the field correct spectrum of each pixel used for calibration.

4.5.1.4 Accuracies of ATCOR, QUAC & ELM for the Barrax scene

This section gives a summary of the accuracy of the most common AC methods for the Barrax scene. It is appreciated that the scene corrected by ATCOR gives slightly higher reflectance in the visible (especially in the green colour region) and at the SWIR bands, with respect to QUAC reflectance, as seen in Figure 4-7. The highest prominence of reflectance in these bands has been commonly observed from scenes with higher proportions of vegetation [50] [51].

Figure 4-7 plots a typical spectrum of alfalfa extracted from the average of a 20x20-pixel ROI centred at the location of [127, 78], after the atmospheric correction by the ELM, QUAC and ATCOR, when soil 2 is used as calibration pixel in ATCOR. It is seen that the ELM corrected spectrum exhibits erratic peaks, while both ATCOR and QUAC give approximately the same extent of errors, with a slight advantage for ATCOR. Table 4-1 tabulate the accuracy of the AC for the 7 classes of vegetation in the scene: note that the errors are evaluated without water absorption bands.

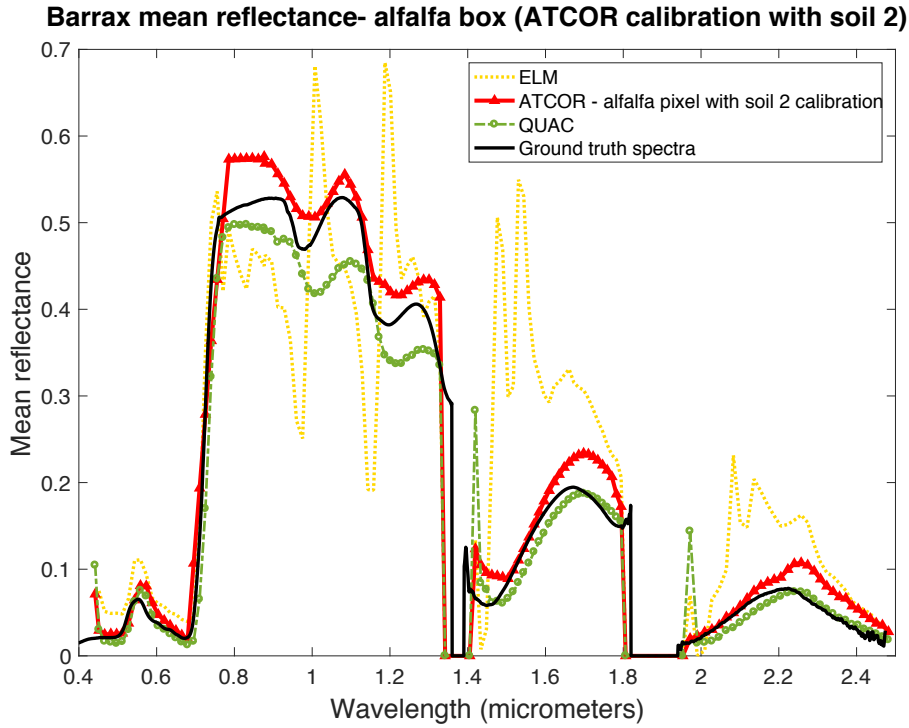


Figure 4-7: average reflectance spectra of alfalfa after the scene is atmospheric corrected by QUAC, ELM and ATCOR and compared with the GT measurement, when soil 2 is used as calibration pixel in ATCOR.

		Euclidean distance (d) of vegetation classes wrt GT (%)- The Barrax scene						
AC methods	ATCOR	Soil 1	Soil 2	Soil 3	Soil 4	Alfalfa 1	Alfalfa 2	Corn
	Alfalfa 2	9.4 ± 0.3 %	7.6 ± 0.2 %	5.7 ± 0.1 %	15.3 ± 0.3 %			15.2 ± 0.2 %
	Corn	18.7 ± 0.4 %	19.0 ± 0.4 %	17.8 ± 0.3 %	22.8 ± 0.4 %	10.7 ± 0.3 %	11.5 ± 0.3 %	
ATCOR	Soil 1		17.2 ± 0.3 %	7.5 ± 0.2 %	8.7 ± 0.2 %	7.5 ± 0.2 %	7.9 ± 0.2 %	17.2 ± 0.3 %
	Soil 2	13.1 ± 0.2 %		5.0 ± 0.1 %	23.5 ± 0.5 %	7.9 ± 0.4 %	6.8 ± 0.2 %	19.9 ± 0.3 %
	Soil 3	6.9 ± 0.2 %	3.9 ± 0.2 %		16.1 ± 0.3 %	4.3 ± 0.1 %	4.2 ± 0.1 %	17.6 ± 0.2 %
	Soil 4	4.3 ± 0.1 %	1.7 ± 0.1 %	2.9 ± 0.2 %		5.4 ± 0.2 %	5.6 ± 0.1 %	17.8 ± 0.4 %
QUAC		9.6 ± 0.3 %	11.1 ± 0.2 %	13.6 ± 0.3 %	5.5 ± 0.1 %	13.4 ± 0.4 %	14.2 ± 0.3 %	19.5 ± 0.5 %
ELM		22.6 ± 0.8 %	21.8 ± 0.3 %	21.7 ± 0.4 %	22.6 ± 0.6 %	17.7 ± 0.5 %	17.9 ± 0.6 %	23.0 ± 0.8 %
Mean of errors	**ATCOR**	**10.9 ± 0.2 %**						
	QUAC	**12.4 ± 0.3 %**						
	ELM	**21.0 ± 0.7 %**						

Table 4-1: mean error of the atmospheric correction by ATCOR, QUAC and ELM for 7 classes of vegetation of the Barrax dataset.

It is seen from Table 4-1 that the ELM is the worst performing algorithm, with error values that almost double the errors obtained by QUAC and ATCOR, giving an average error of ~21% over the 7 class of vegetation. The large error in the ELM may be ascribed to the large area of the scene (~4 km^2) without sufficient number of calibration panels to cover the study site entirely. In addition, the lack of a water vapour atmospheric correction tool for ELM also increases the error for this method with respect to QUAC and ATCOR.

4.5.2 Accuracy of AC for the Selene scene

4.5.2.1 AC by using ATCOR

The atmosphere package known as “h01000_wv20”, which is suitable for scenes acquired at 1000 m of altitude, and an atmospheric water vapour concentration of 2.0 g/cm^2, has been adopted for the AC. The visibility is set for 50 km, Flight heading is 325˚, sensor FOV is 27˚, and the 820nm band has been selected for performing the water vapour correction in ATCOR for the creation of the scan file. The illumination and solar geometry parameters of the scene, such as the flight altitude (0.94 Km), the solar zenith angle (38.5˚), the ground elevation (0.050 Km) and the solar azimuth angle (162.8˚), have been employed according to the information supplied by DSTL.

4.5.2.2 Accuracies of AC- ATCOR, QUAC & ELM for the Selene scene

Similar to the section above, the noisy and water absorption bands (234-251 and 318-361 bands) have been removed; only the remaining 386 spectral bands have been processed for assessing the accuracy of the AC methods. As mentioned in the experiment & method section (Section 3.2.1) this scene contains a set of 7 colour panels and a set of different constant reflectance calibration panels for spectral and illumination calibrations. Figure 4-8 and 4-9 show the RGB pictures of these panels and their spectral characteristics respectively.

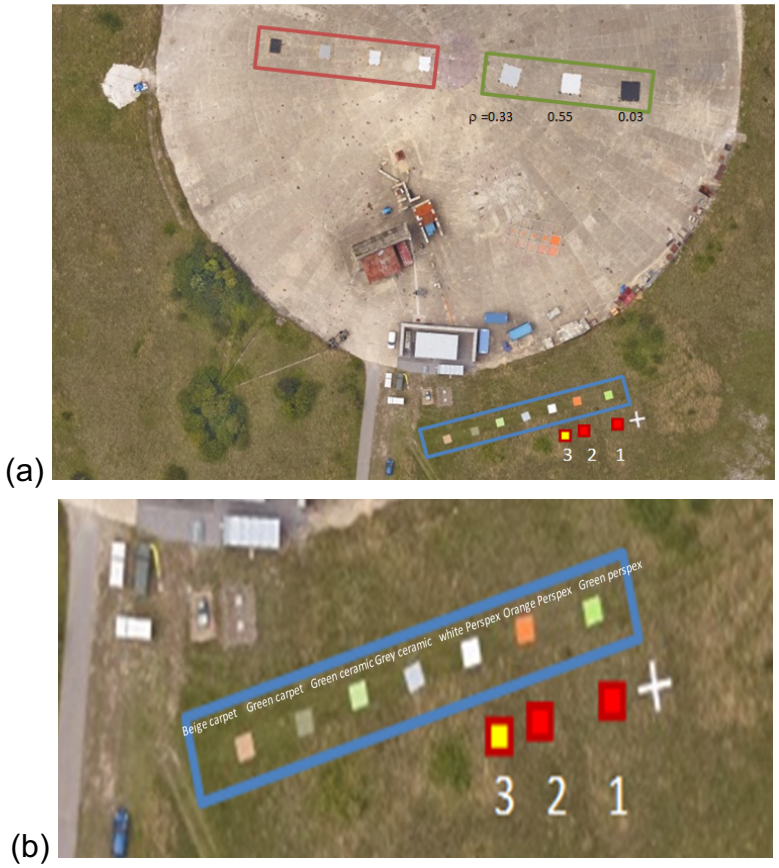
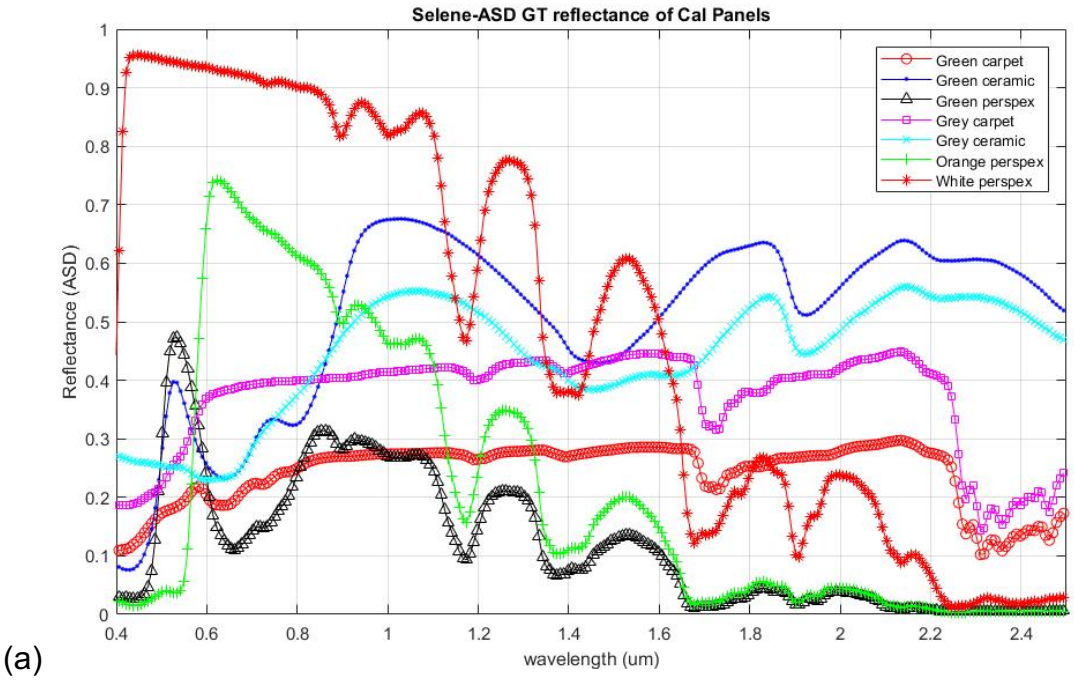
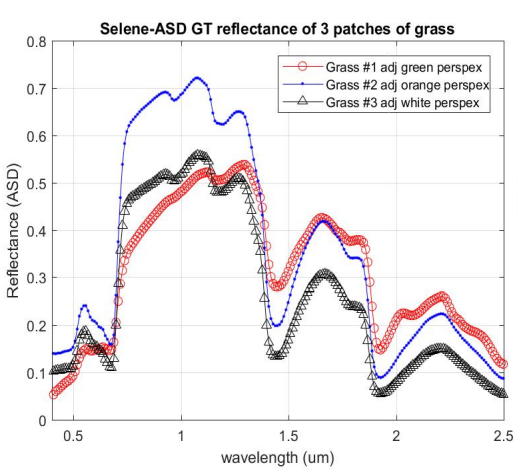


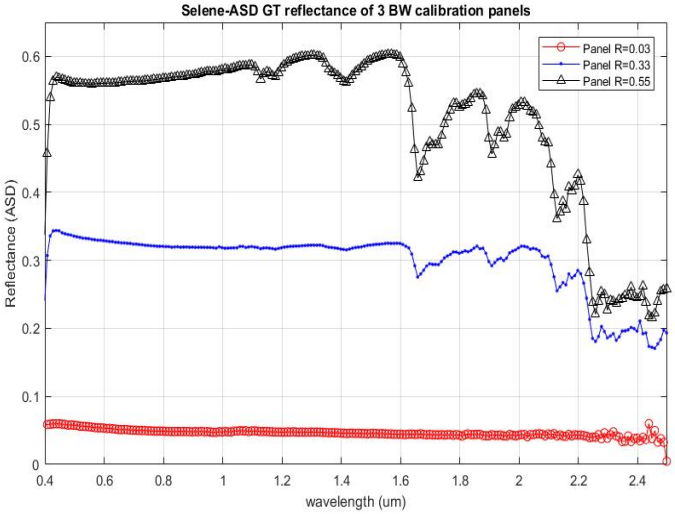
Figure 4-8: RGB picture of the calibration panels (a) the 3 different levels of grey panels for illumination calibration, (b) the 7 colour panels and 3 patches of grass for spectral calibration and validation.



(a)



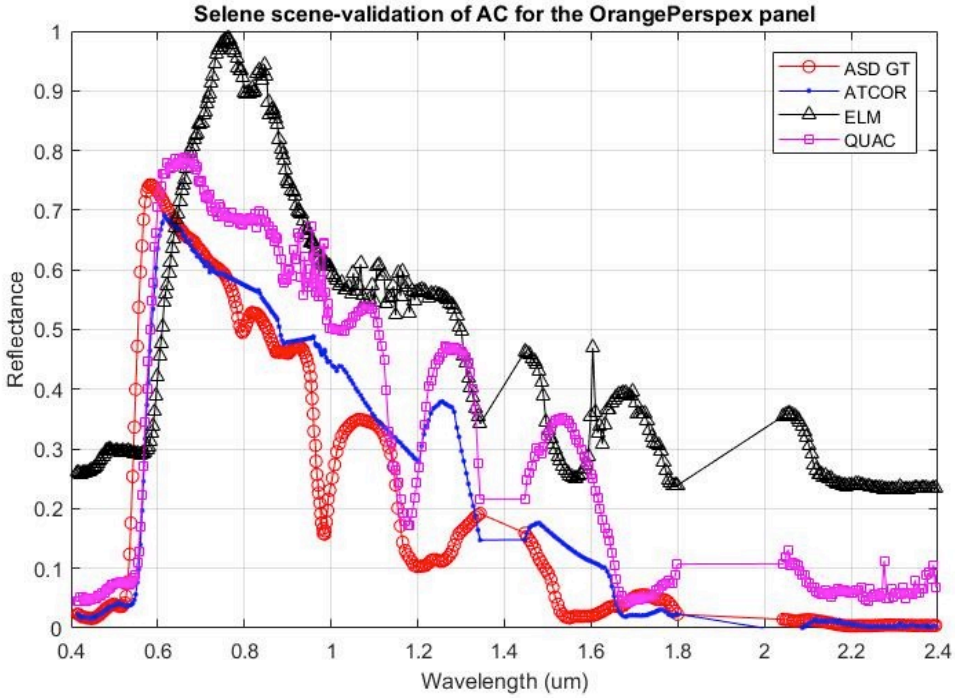
(b)



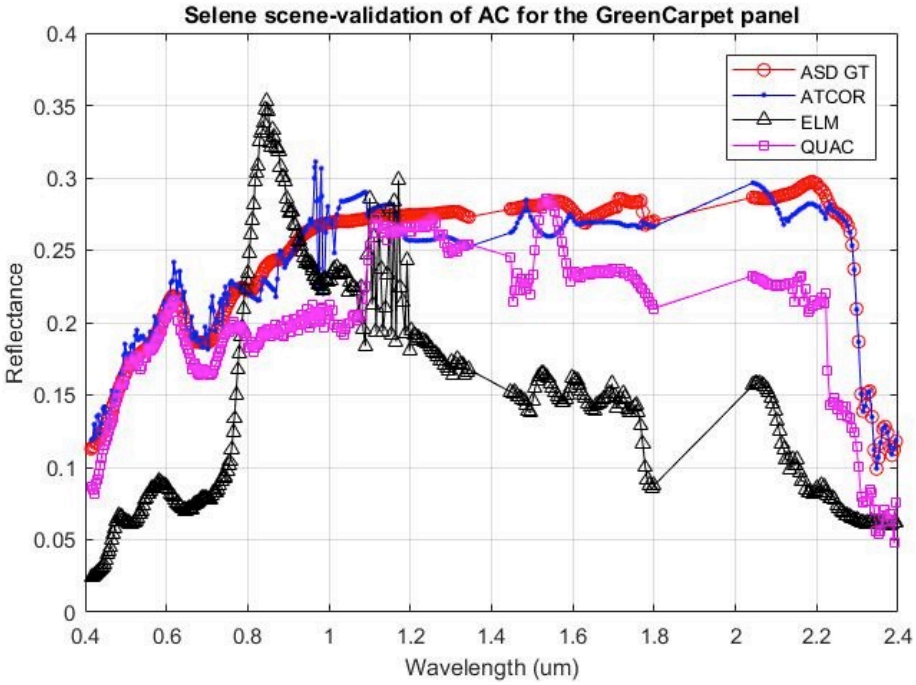
(c)

Figure 4-9: spectral characteristics of (a) the 7 colour panels, (b) the 3 patches of grass and (c) the 3 illumination calibration panels located at the top of the concrete slab.

Similarly to the last section, ATCOR has been calibrated by using various colour panels for enhancing the AC accuracy; the spectral errors obtained after the corrections by QUAC, ATCOR and ELM are obtained in Figure 4-10 and their accuracies for the Selene scene are tabulated in Table 4-2.



(a)



(b)

Figure 4-10: Average reflectance spectra of the atmospheric corrected spectra of the orange Perspex (a) and green carpet (b) by QUAC, ATCOR and ELM in comparison with that of the ground truth data measured by ASD.

		Euclidean distance (d) of vegetation classes wrt GT (%)- The Selene scene									
AC methods	ATCOR calibration	Green perspex	Orange perspex	White perspex	Green ceramic	Grey ceramic	Green carpet	Grey carpet	Vegetation	Vegetation 2	Vegetation 3
	Green perspex		7.9 ± 0.2 %	8.9 ± 0.2 %	10.0 ± 0.3 %	10.5± 0.2 %	10.0 ± 0.3 %	11.4 ± 0.4 %	11.1 ± 0.3 %	11.7 ± 0.5 %	11.4 ± 0.1 %
	Orange perspex	9.1 ± 0.1 %		9.2 ± 0.2 %	10.4 ± 0.2 %	10.7 ± 0.3 %	10.7 ± 0.4 %	10.5 ± 0.3 %	11.8 ± 0.3 %	12.2 ± 0.3 %	12.1 ± 0.2 %
ATCOR	White perspex	9.7 ± 0.2 %	10.1 ± 0.3 %		16.7 ± 0.3 %	10.3 ± 0.3 %	9.8 ± 0.1 %	11.0 ± 0.4 %	12.0 ± 0.4 %	12.4 ± 0.5 %	12.2 ± 0.4 %
	Green ceramic	10.1 ± 0.2 %	10.5 ± 0.2 %	9.9 ± 0.3 %		10.2 ± 0.4 %	9.5 ± 0.2 %	10.3 ± 0.3 %	12.0 ± 0.3 %	11.5 ± 0.4 %	11.3 ± 0.5 %
	Grey ceramic	10.8 ± 0.3 %	11.2 ± 0.4 %	11.5 ± 0.4 %	10.7 ± 0.1 %		9.3 ± 0.3 %	9.6 ± 0.2 %	12.3 ± 0.2 %	12.6 ± 0.4 %	12.5 ± 0.3 %
	Green carpet	9.5 ± 0.1 %	10.4 ± 0.1 %	10.8 ± 0.2 %	10.5 ± 0.3 %	11.9 ± 0.3 %		9.0 ± 0.2 %	11.3 ± 0.4 %	11.7 ± 0.3 %	11.2 ± 0.2 %
	Grey carpet	11.2 ± 0.2 %	11.0 ± 0.3 %	11.5 ± 0.3 %	11.0 ± 0.4 %	11.7 ± 0.2 %	9.0 ± 0.1 %		11.6 ± 0.4 %	11.7 ± 0.6 %	11.5 ± 0.3 %
	Vegetation	11.3 ± 0.2 %	11.8 ± 0.2 %	11.9 ± 0.4 %	11.4 ± 0.1 %	12.2 ± 0.4 %	11.6 ± 0.2 %	12.9 ± 0.5 %		11.4 ± 0.5 %	11.3 ± 0.2 %
	Vegetation 2	11.1 ± 0.1 %	12.0 ± 0.4 %	11.7 ± 0.2 %	11.3 ± 0.3 %	12.4 ± 0.3 %	11.8 ± 0.2 %	13.0 ± 0.3 %	11.5 ± 0.1 %		11.8 ± 0.3 %
	Vegetation 3	11.7 ± 0.1 %	12.3 ± 0.2 %	12.3 ± 0.4 %	11.6 ± 0.1 %	12.7 ± 0.5 %	12.4 ± 0.4 %	13.1 ± 0.5 %	11.3 ± 0.2 %	12.0 ± 0.3 %	
QUAC		12.1 ± 0.4 %	11.6 ± 0.4 %	12.5 ± 0.3 %	13.1 ± 0.4 %	12.2 ± 0.3 %	13.4 ± 0.4 %	13.0 ± 0.2 %	13.1 ± 0.4 %	14.3 ± 0.5 %	13.4 ± 0.3 %
ELM		20.5 ± 0.7 %	20.3 ± 0.6 %	19.6 ± 0.6 %	20.2 ± 0.4 %	19.1 ± 0.3 %	19.8 ± 0.5 %	20.5 ± 0.3 %	20.9 ± 0.7 %	21.9 ± 0.8 %	21.1 ± 0.7 %
Mean of errors	**ATCOR**	**11.3 ± 0.3 %**									
	QUAC	**12.9 ± 0.4 %**									
	ELM	**20.4 ± 0.6 %**									

Table 4-2: mean error of the atmospheric correction by ATCOR, ELM and QUAC for 7 calibration panels and 3 vegetation in the Selene scene.

The results of the atmospheric correction assessments for ATCOR, ELM and QUAC, as indicated by the Table 4-2, shows a clear trend which is consistent with Barrax data results (see section 4.5.1.4); the absolute accuracies of these AC assessments are similar in both cases. The improvement of the absolute accuracies in the Selene scene for ELM happens due to the use of the man-made targets for the validation (white and black panels), and these panels exhibit negligible material variability in these man-made materials.

4.6 Summary of AC chapter

By summing up the above two sections the following conclusions can be drawn:

1. The errors of the ELM AC method are ~100% higher than the error values obtained for ATCOR/QUAC methods.

2. The errors of ATCOR/QUAC are in the range of ~11%-13% depending on the flight parameters and atmospheric properties of the scene.

3. By using the in-flight scene calibration method, ATCOR has achieved the best accuracy, with error values ranging around 8-12%.

4. This means that all classification assessments carry an uncertainty of ~10% with respect to the true spectral characteristics of the scene.

5. These error values are obtained with respect to the true spectral characteristics of the scene and therefore it is a systematic error in our case, as we compare with the results of other work which had been done using the same corrected datasets.

5 Hyperspectral band selection

5.1 Curse of dimensionality

Hughes research [1] shows the theoretical accuracy of a model which scales non-linearly with the dimensionality of the dataset: the accuracy is expected to be improved when more spectral bands are utilized for the classification, but further increasing of the data dimensionality would reduce the classification accuracy when the training data size is kept constant. This occurs because when the dimensionality is too large it increases more degrees of freedom for fitting the data, hence the error is increased with the data dimension. This behaviour is plotted in Figure 5-1 which exhibits the well-known 'bell' shape of the classification accuracy versus the dimensionality (n).

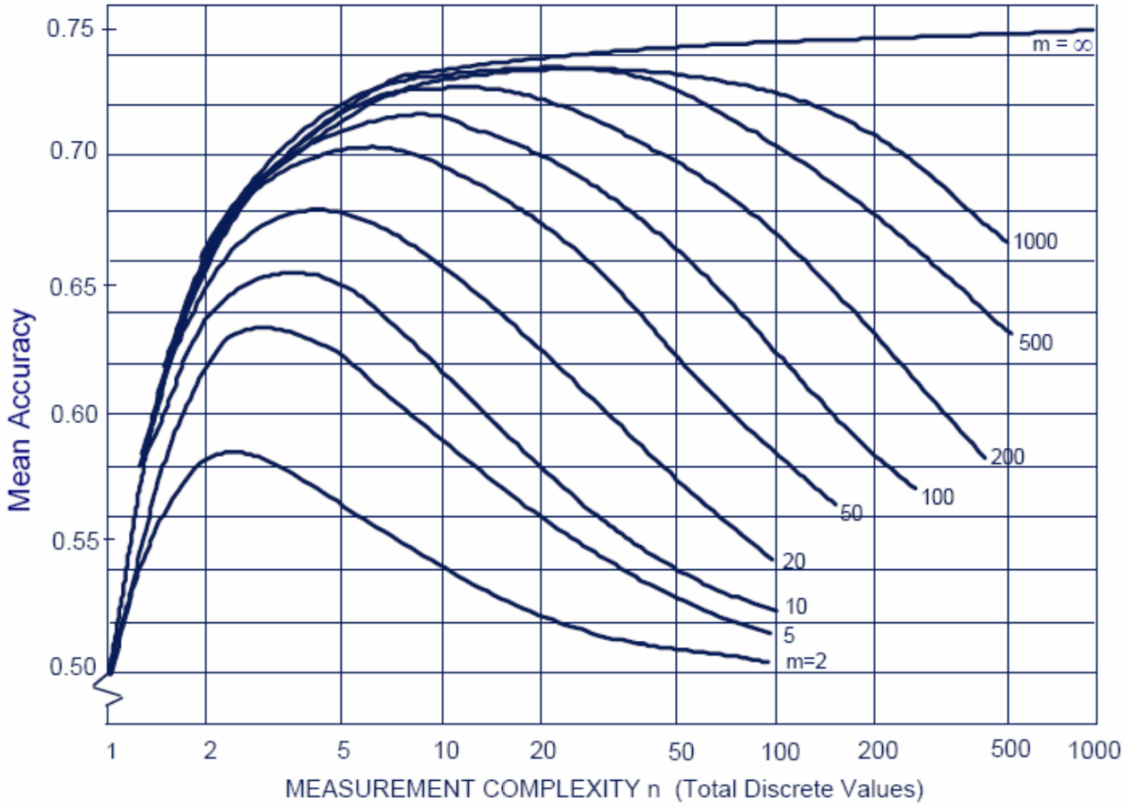


Figure 5-1: The main result of Hughes analysis which explains how the accuracy of supervised classification models varies with the dimensionality of dataset (n) with a running parameter of the number of the training data (m) which is needed to define the class [52].

5.1.1 Why band selection is important at remote sensing and at this thesis

Hyperspectral imaging sensor collects detailed spectral responses from ground objects using hundreds of narrow bands; this technology is used in many real-world applications. Band selection aims to select a small subset of hyperspectral bands to remove spectral redundancy and reduce computational costs while preserving the significant spectral information of ground objects. [53]

Due to the advancement of imaging spectrometer technology, a typical hyperspectral sensor now collects the spectral reflectance of ground objects on Earth's surface using hundreds of narrow bands [54][55]. Because of high spectral resolution, the acquired hyperspectral imagery (HSI) can better differentiate ground objects using subtle spectral discrepancies. It greatly benefits many practical applications, e.g., biophysical parameter retrieval [56][57], fine classification of complex environments [58][59], small target detection [60][61], ...

However, HSI produces a very large volume of data. For example, the Hyperspectral Infrared Imager launched by NASA's Jet Propulsion Laboratory (JPL) has the continuous averaged data rate of 65 Mb/s, which produces a data volume of 372 Gb/orbit and 5.2 Tb/day [62]. The resulting high computational cost makes fast on-board processing difficult.

Its vast data volume also renders the common image processing algorithms designed for multispectral imagery ineffective. In particular, the high dimensionality of HSI data brings about the "curse of dimensionality" problem - as mentioned before, under a fixed, small number of training samples, the classification accuracy of HSI data decreases when the dimensionality of HSI data increases [1].

In remote sensing applications, collecting a larger number of training samples is costly, time-consuming, and even impossible. Therefore, it is necessary to

develop new dimensionality reduction methods and protocols for hyperspectral image processing [63] – [65].

Adjacent bands in HSI are highly correlated, and some spectral bands may not carry discriminant information in a specific application [66][67]. One of the most important pre-processing steps for HSI analysis is dimensionality reduction, which aims to remove spectral redundancy while preserving critical information for subsequent applications [68][69].

Hyperspectral dimensionality reduction can be achieved through either band selection or feature extraction [70].

Feature extraction transforms the original data into another feature space with certain criteria [71][72]. Most feature extraction methods linearly combine all of the original bands and may complicate the interpretation of results if the adopted criterion is inappropriate [73]. On the other hand, band selection selects a desired band subset from the original bands and well preserves the spectral meaning of spectral channels [74].

For all of these reasons, it is important to implement band selection methods at remote sensing. Due to its importance, we have studied and implemented several band selection methods at this thesis, for different datasets.

To conclude this part, atmospheric correction affects band selection due to the errors provoked by those processes when converting the digital numbers / radiance results into reflectance.

At this case, the datasets used on this section have been obtained from public sources, as explained on chapter 3. The atmospheric correction has already been performed by the publishers of these public datasets; for this reason, the atmospheric correction is not taken into account at this chapter 5. It is a good future work idea to try to study the effects of these atmospheric correction errors from raw datasets into the final classification results.

5.2 Methods of band selection

To mitigate this problem observed by Hughes [1] for achieving the optimal classification accuracy, data compression in the dimensional domain has been generally used [75]. Most work takes the advantage of spectral correlation of multispectral/hyperspectral images for achieving data compression. Many approaches have been proposed for finding the most crucial and significant information of the HSI data [76] - [78]. For instance, distance measures (Bhattacharyya distance, Jeffreys - Matushita distance), information-theoretic approaches (divergence, transformed divergence, mutual information) and eigen analysis have been the main concepts for capturing the information of the HSI data by using a handful of spectral bands. In particular, the use of the divergence measure for band selection has received considerable interest in the HSI research community and it has been utilized for hyperspectral pixel classification with good results [78]. However, the technique involves formidable computational complexity for calculating the divergences of all possible combinations of band subsets which consists of > 200 bands in a typical HSI dataset.

Different approaches have also been used, such as the feature extraction based methods [79] [80], which transforms the original data into the destination feature space through projections like projection pursuit (PP) [81][82], principal component analysis (PCA) [83][84], independent component analysis (ICA) [85] etc. Other methods use band (or feature) selection [86] of the richest characteristics from the dataset. The former generally achieves higher classification accuracy, while the latter could preserve the relevant original information of the dataset [87].

To alleviate this dimensionality problem, this research uses a combination of spatial and spectral features for achieving the optimal band selection and classification accuracy. Our work has demonstrated the validity of the Hughes phenomenon [1] through experiments at band selection for the first time.

5.3 BS algorithms

We are also going to investigate new ways of improving the most recent state-of-the-art band selection algorithms.

5.3.1 HyperBS algorithm

We are going to test first the band selection method named as HyperBS [88]. This algorithm was developed by Rashwan et al. (2017) [88]; this paper proposes an unsupervised band selection approach based on a split-and-merge concept.

This new approach provides relevant spectral bands by splitting low-correlated bands and then merging high-correlated bands, in order to reduce the dimensionality of the hyperspectral image. The algorithm presented on this article performs a split step once two adjacent spectral bands were detected as weakly correlated.

The splitting step is conducted by creating two virtual sub-bands between the original two bands; this process is repeated for the sub-bands as long as they are weakly correlated. Two consecutive bands are split if the sample-based Pearson correlation coefficient $\check{r}_{l+1}^{(t)}$ between two consecutive bands $\check{x}_l^{(t)}$ and $\check{x}_{l+1}^{(t)}$ is lower than a given threshold value ρ, i.e.,

$$\check{\boldsymbol{r}}_{\boldsymbol{l,l+1}}^{\boldsymbol{(t)}} < \boldsymbol{\rho} \qquad \textbf{(5-1)}$$

For the merging stage, the algorithm applies a merge step by fusing all bands that are determined as highly correlated.

Once a synthetic HSI image $\check{X}$ with an extended number $\check{L}$ of bands has been obtained from the splitting process, these $\check{L}$ spectral bands follow a merging

process, that aims at gathering most of the information into a significantly lower number of bands $\hat{L}$, with $\widehat{L} \leq L \leq \check{L}$.

This merging process consists in successively analysing the spectral bands of $\check{X}$ to identify the most correlated ones, following the opposite decision rule specified in the expression (5-1).

5.3.2 Saliency algorithm

The next method that we are going to test on our datasets is a saliency-based algorithm, obtained by Su et al. (2018) [89]. The SBSS (Saliency Bands and Scale Selection) method is used to obtain the hidden information of hyperspectral images effectively, while its underlying assumptions are:

1) It is reasonable to combine spectral and spatial information to excavate the intrinsic property of a hyperspectral image

2) There are some saliency bands that can represent a hyperspectral image without significant information loss in data exploitation; and

3) Saliency, scale, and image description have an intrinsic connection.

For a hyperspectral image, as the target may mix with the background in one pixel, and no edges of a target can be detected, it is reasonable to use extrema points to select proper saliency bands. However, it is feasible to use other features to implement the SBSS method when the features are appropriate. In addition, the trace and determinant of a Hessian matrix have a strong ability to describe local structures of an image, so the Hessian matrix is used to measure the saliency of a band.

For a hyperspectral image H, the trace and determinant of each pixel in H are calculated, and then the results are denoted by *tr H* and *det H*, respectively, which have the same size as the original hyperspectral image:

$$\text{tr } H = H_{xx} + H_{yy} \tag{5-2}$$

$$\det H = H_{xx} H_{yy} - H_{xy}^{2}. \tag{5-3}$$

where H_{xx}, H_{yy} and H_{xy} are the two-order difference results of H; H_{xx}, H_{yy} and H_{xy} are calculated by the following expression, where ∇ is the gradient operator:

$$H_{xx} = \nabla_x (\nabla_x H), \qquad H_{yy} = \nabla_y (\nabla_y H), \qquad H_{xy} = \nabla_y (\nabla_x H) \tag{5-4}$$

Extrema points here refer to points that are simultaneously local extrema with respect to both spatial and spectral dimensions; this means that spectral coordinates should be treated the same as the spatial coordinates.

5.3.3 SLN algorithm

Now we are going to look deeper on the comparison of our band selection methods with respect to other algorithms. We are going to compare them with the neural network algorithm [90], proposed by Habermann et al. (2019), where they propose a new supervised filter-based approach for band selection based on neural networks, named as SLN (Single Layer Neural networks).

This algorithm performs a classification by means of a binary single layer neural network classifier between each pixel of our datasets and the rest of the data, for each spectral band. After that, the bands with the highest classification score are taken.

This process iterates until the previously defined number of bands is achieved.

5.3.4 OCF algorithm

The next step for us is starting to use an optimal clustering algorithm for band selection, obtained by Wang et al. (2018) [91], using the following contributions, as seen on Figure 5-2:

1) An optimal clustering framework (OCF) is proposed to search for the optimal clustering structure in HIS. We need to take into account that the proposed OCF is a general framework (different kinds of objective function can be optimized via the same procedure once they comply with the specific form).

2) A ranking on clusters strategy (RCS) is proposed as an effective criterion to select the representative bands under the achieved clustering structure. By applying a ranking algorithm on the clustering result, RCS can better exploit the advantages of clustering-based and ranking-based methods, and generate a band subset with lower correlation and more discriminative information.

3) An automatic method to determine the required number of bands is proposed. Through reducing the correlation among bands, we aim to uncover how much distinctive information can be produced by certain number of bands.

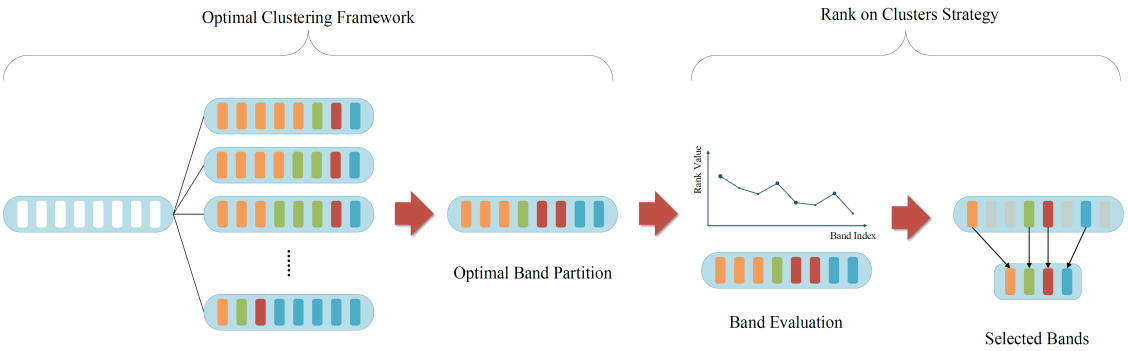


Figure 5-2: flowchart of the overall procedure to conduct band selection performed by the optimal clustering method. [91]

As seen on Figure 5-2, first, given an objective function, the algorithm searches for the optimal band partition via OCF (in the figure, each colour refers to a cluster of bands); then all the bands are evaluated through a ranking method, and finally, the top-rank bands are selected in each cluster.

5.3.5 E-FDPC algorithm

Once performed this, we are going to add the results obtained for an algorithm named as E-FDPC (Enhanced Fast Density-Peak-based Clustering), obtained by Jia et al. (2016) [92].

This band selection method is based on the computation of the local density and the intra-cluster distance of each point; the product of the two factors is sorted in decreasing order, and cluster centres are recognized as points with anomalously large values.

This method works as shown:

At the first step the ranking score of each band is computed by weighting the normalized local density and the intra-cluster distance. Second, an exponential-based learning rule is used to adjust the cut-off threshold for a different number of selected bands, where it is fixed in the FDPC.

The proposed approach is thus named the enhanced FDPC (E-FDPC). Finally, a process called isolated-point-stopping criterion, is developed to automatically determine the appropriate number of bands to be selected.

Weighted Ranking Score Scheme for Each Band: The original FDPC algorithm is a rapid and effective method to find the cluster centres according to local density ρ and intra-cluster distance δ. Although the idea of combining local

density and distance from points of higher density is far more reliable than using either of the two factors alone, how to effectively integrate the two factors for hyperspectral band selection is still a difficult issue. The weight value is important to promote the performance of the algorithm, and the ranking score of each band should be computed through weighting the two factors.

Specifically, after the points with anomalously large scores (considered the cluster centres) have been found by the FDPC, the following chosen bands should be those with large δ and relatively small ρ i.e., the points that are away from the chosen points and are the centres in smaller regions, which means that parameter δ should be given more weight than ρ.

Based on this observation, instead of taking δ and ρ into account equally, the proposed E-FDPC algorithm increases the weight of δ through a normalization and square product process.

In order to estimate the local density ρ_i for each image, a Gaussian kernel function is adopted to decrease the negative impact of the statistical errors caused by the small number of bands, i.e.,

$$\boldsymbol{\rho_i} = \sum_{\boldsymbol{j=1, j\neq i}}^{\boldsymbol{L}} \mathbf{exp}\left(-\left(\frac{\boldsymbol{D_{ij}}}{\boldsymbol{d_c}}\right)^{\mathbf{2}}\right) \quad \textbf{(5-5)}$$

where $D_{ij} = \sqrt{S_{ij}}/L$ is a variable depending on the similarity matrix S, for each band image i, 1 ≤ i ≤ L, which can be considered a data point, and d_c is a cut-off distance used to keep a region for each data point. Likewise, δ_i is measured by computing the minimum distance between point i and any other point (j) with a higher density, i.e.,

$$\boldsymbol{\delta_i} = \min_{\boldsymbol{j}:\, \boldsymbol{\rho_j} > \boldsymbol{\rho_i}} (\boldsymbol{D_{ij}}) \quad \textbf{(5-6)}$$

Meanwhile, for the data point whose density is the highest among all the points, i.e., the global maximum in the density, δ_i should be much larger than the typical nearest neighbour distance, which is:

$$\boldsymbol{\delta_j = \underset{j}{max}(D_{ij}),\ if\ \rho_i = \max(\rho)} \quad \textbf{(5-7)}$$

Thus, the cluster centres can be recognized as points with anomalously large values of δ, which is the core principle of the FDPC algorithm.

E-FDPC improves the values of δ by means of a normalization of these values on the scale of [0, 1], which is done by:

$$\boldsymbol{\delta = \frac{(\delta - \delta_{min}).}{(\delta_{max} - \delta_{min})}} \quad \textbf{(5-8)}$$

where ./ represents the element-wise division operator, and ρ is normalized in the same way.

5.3.6 ISSC algorithm

Now we are going to implement a new algorithm called ISSC (Improved Sparse Subspace Clustering), implemented first by Sun et al. (2015) [93]. This band selection method assumes that the band vectors are sampled from a union of low-dimensional orthogonal subspaces and each band can be sparsely represented as a linear combination of other bands within its subspace.

This algorithm has four main steps:

1. ISSC represents band vectors by solving the L_2-norm optimization problem using the LSR (least-squares regression) algorithm. The sparse and block diagonal structure of the LSR coefficient matrix leads to the correct segmentation of our band vectors.

2. The angular similarity (AS) measurement is presented and used to construct the similarity matrix; we assume that the space representation of two similar bands from the same subspace should have a small angle between both of them, since both are sparsely represented in a similar combination using other bands.

For each pairwise sparse coefficient vectors $\widehat{z_i}$ and $\widehat{z_j}$, which represent the coefficient vectors of y_i and y_j (two spectral bands without noise) respectively, the AS measurement is defined as:

$$W_{AS_{ij}} = \left(\frac{\widehat{z_i} * \widehat{z_j}}{||\widehat{z_i}||^2 \, x \, ||\widehat{z_j}||^2}\right)^2 \quad \textbf{(5-9)}$$

3. The distribution compactness plot algorithm is used to estimate an appropriate size of the band subset.

4. Finally, spectral clustering is implemented to segment the similarity matrix and the desired ISSC band subset is formed.

5.3.7 CNN algorithm

The last method that we are going to implement is the CNN (Convolutional Neural Network): this band selection method, obtained by Sharma et al. (2016) [94], is based on the union of 5 x 5 convolutional layers, to classify the best hyperspectral bands on the neural network. The network is firstly trained through a substantial amount of band images as training data. Subsequently, it extracts features from each test band image, and the weight of the neurons is updated depending on whether the predicted label is correctly or incorrectly classified. Then, the network combines all models into an ensemble, and the mostly weighted bands are selected by voting.

Having the CNN work on a single band as a generic CNN helps to improve the classification performance: treating the data in this flat way handles the requirement of many images in the CNN training process.

5.4 Proposed spatial spectral mutual information (SSMI) BS algorithm

5.4.1 Spatial feature through Edge algorithm

We are also going to implement the combination of saliency with respect to a different band selection method obtained by Cao et al. (2016) [95], which uses the spatial structure of each spectral band, based on the hypothesis that, though the amplitudes of different classes or materials may have a different variation tendency in each band, the spatial structure of every high-discriminating band image should be similar; bands with high discriminating power often have well-defined edges between different materials.

The objective of this spatial pre-processing is to remove spectral bands that are low in morphological features, i.e. non-discriminative bands. Given a hyperspectral image which consists of N bands B = [B_1, B_2, …, B_N], the morphology of the bands can be approximated by using edge detection:

$$E_i = \text{Edge}(B_i) \quad \textbf{(5-10)}$$

where E_i is the edge feature map of the band B_i, and Edge(.) is the edge detector operator which can be Canny/Sobel/Roberts detectors. There are various options to classify the edge feature maps into highly structured and to distinguish them from the featureless low morphological maps.

One straightforward means is to rank the E_i by comparing it with respect to the mean of all E_i:

$$C_i = \text{Corr}(E_i, E). \quad \textbf{(5-11)}$$

where $$E=\text{mean}(E_i) \quad \textbf{(5-12)}$$

Corr(.) is the correlation function, and an abrupt change of C_i will give an indication of the threshold boundary between the highly structured and the featureless morphological maps. In this work, all bands below this threshold are removed and the highly structured bands are then passed onto the next stage of spectral processing. The validity of this formulation is more applicable to high quality datasets (i.e. high signal-to-noise ratio) of scenes with very small portion of low reflectance objects like water or hard shadows.

There are three different types of edge filters to use into our algorithm:

Canny filter, which finds edges by looking for local maxima of the gradient of our image [96]; on this case, the edge function calculates the gradient using the derivative of a Gaussian filter.

Sobel filter, where the operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives – one for horizontal changes, and one for vertical [97]. If we define **I** as the source image, and $\mathbf{G}_x$ and $\mathbf{G}_y$ are two images which contain the vertical and horizontal derivative approximations at each point respectively; the computations are performed as follows:

$$\boldsymbol{G_x} = \begin{bmatrix} \mathbf{-1} & \mathbf{0} & \mathbf{1} \\ \mathbf{-2} & \mathbf{0} & \mathbf{2} \\ \mathbf{-1} & \mathbf{0} & \mathbf{1} \end{bmatrix} * \boldsymbol{I} \quad \text{and} \quad \boldsymbol{G_y} = \begin{bmatrix} \mathbf{-1} & \mathbf{-2} & \mathbf{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{2} & \mathbf{1} \end{bmatrix} * \boldsymbol{I} \qquad \textbf{(5-13)}$$

where * denotes the convolution operation.

Roberts filter, that provides a simple approximation to the gradient magnitude [98]:

$$\boldsymbol{G}\,[\boldsymbol{f}(\boldsymbol{i},\boldsymbol{j})] \;=\; |\boldsymbol{G_x}| \;+\; |\boldsymbol{G_y}|, \qquad \textbf{(5-14)}$$

where G_x and G_y are calculated using the following masks:

$$\mathbf{G_x} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{-1} \end{bmatrix} \quad \text{and} \quad \mathbf{G_y} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{-1} & \mathbf{0} \end{bmatrix} \qquad \textbf{(5-15)}$$

5.4.2 Spectral feature extraction through Joint Entropy algorithms

Entropy has been a quantity that has been widely used in communication, computing, cryptography and many other data-related applications. Entropy is a measure of the unpredictability of the state, therefore it is not only the content of the state; how the state is chosen determines its entropy. For a variable A, where the probability of the event A_i is $p(A_i)$, the amount of information $I(A_i)$ acquired due to the observation of the event A_i [99], is defined as:

$$\boldsymbol{I(A_i) = \log\left(\frac{1}{p(A_i)}\right) = -\log\left(p(A_i)\right)} \quad \textbf{(5-16)}$$

The entropy (H) of the variable A (represented by H(A)) is the expectation value of the information function I(A), i.e., the expectation of the probability p(A) for choosing elements of A:

$$\boldsymbol{H(A) = -\sum_A p(A) \log p(A)} \quad \textbf{(5-17)}$$

A large entropy H(A) means that A is quite unpredictable, and the averaged amount of the information conveyed by the identification of A is large. The band image can be considered as A and the pixels of the band as A_i; thus, the entropy can be used for encoding the information of every bands in the dataset by using Eq. 5-18 [99-105]. Alternative methodology has been using joint entropy between two variables A and B [106-108].

From Eq. 5-16 the joint information of variables A and B is given by the mutual information (MI) I(A,B) which is in the form of:

$$\mathrm{I(A,B)} = \sum_{\mathrm{A,B}} \mathrm{p(A,B)} \log \frac{\mathrm{p(A,B)}}{\mathrm{p(A)} * \mathrm{p(B)}}$$

$$\mathbf{I(A,B)} = \mathbf{H(A)} + \mathbf{H(B)} - \mathbf{H(A,B)} \quad \textbf{(5-18)}$$

where p(A) and p(B) are the marginal probability distributions of variables A and B respectively, and p(A, B) is the joint probability distribution of variables A & B.

Eq. 5-18 implies that I(A,B) can be used to measure the similarity of two variables A & B. For two spectral bands A and B present at the dataset, I(A,B) measures the independency (or similarity) of the information conveyed in these two bands.

In the case when one variable is chosen to be the reference data, and the other variable is the set of band images of the scene, then the joint information of I(A,B) measures how close (in the context of information) are the band images with respect to the reference data such as the labelled data that has been used previously [103].

In this work, the marginal probabilities of the bands are estimated from the normalised histograms of each band image, and their corresponding entropies are then calculated by using Eq. 5-18. We select the bands over a determined threshold, obtained empirically for each different dataset. The joint entropy is subsequently evaluated for each pairs of adjacent image bands and the mutual information (MI) of the image pair is evaluated according to Eq. 5-18, once having taken out the low discrimination bands.

The computational cost of the algorithm is reduced due to the band selection process; the edge filter increases its computational cost with a quadratic relationship with respect to the number of bands [$O(n^2)$], due to the processing of the correlation matrix for the spatial part, while the MI increases its time linearly [O(n)] due to the processing of each pair of bands to take out the redundant bands of each dataset. For this reason, the selection of a number of bands at the start helps us to reduce this time cost associated to this method.

On this section, a basic BS scheme which utilizes the mutual information (MI) for the selection of bands from hyperspectral datasets has been adopted in this study. The algorithm has been a basic one which ranks the MI of each spectral band of the dataset, and they are then selected from the top of the list for classification. Since this BS method utilizes spectral information only, a pre-processing method which exploits spatial information for the elimination of low discrimination bands, has been added as a pre-processing technique prior to the MI band selection.

This method is termed as “Spatial Spectral Mutual Information (SSMI)”, and the purpose of the spatial technique is to eliminate bands which do not convey much information towards the morphological property of the dataset for classification.

5.5 Experimental results

5.5.1 Competing algorithm parameters

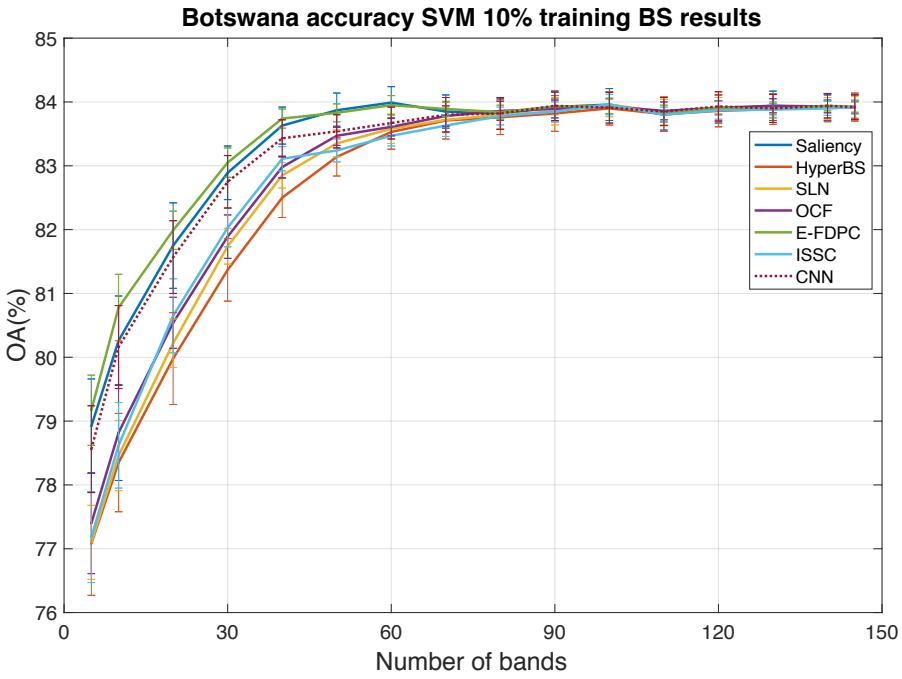
The user parameters of the competing algorithms have been set as according to these authors’ original settings as follows:

- Saliency: use the trace + determinant combination;
- HyperBS: the threshold value has been set to 1;
- E-FDPC: the standard Euclidean distance has been used for the clustering process;
- ISSC: the scalar value of $\beta = 0.1$ has been set for the more complex datasets such as the Indian Pines, Barrax and Botswana, while the rest (Pavia University, Salinas and KSC) the $\beta = 0.001$ have been used [93].
- SBSS: the use of the Roberts filter has obtained the best results.
- Other parameters: Edge algorithm: threshold = 0.2; MI: band information threshold = 0.5 for the selection of the bands which have more information.

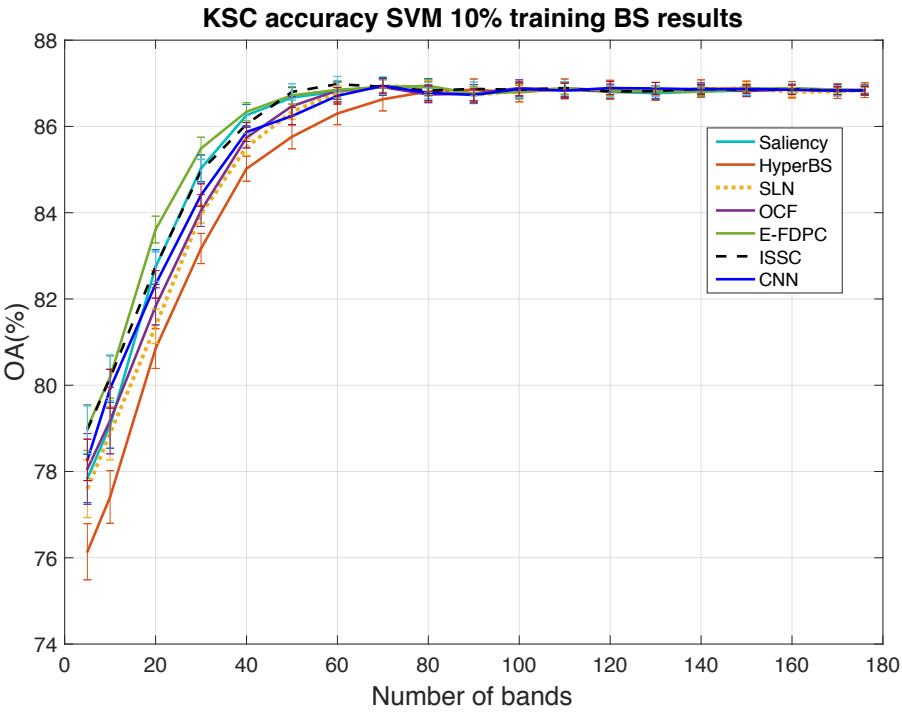
5.5.2 Competing algorithm results

Two classifiers, SVM and KNN (see section 3.3 above) have been used with 10% and 3% training data, randomly selected from each class for training. The experiment is repeated for five times to obtain the mean and standard deviations to represent the goodness of the classification performance for each number of bands selected.

Seven competing algorithms (saliency, HyperBS, SLN, OCF, E-FDPC, ISSC and CNN), have been used for this experiment, obtaining the results shown now:

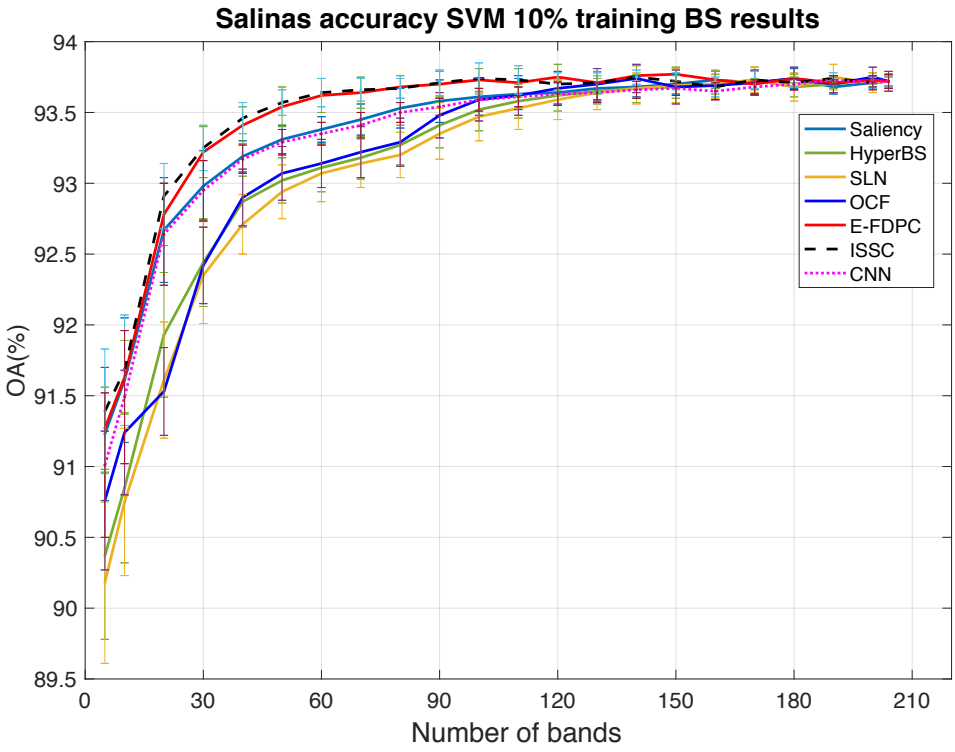


(a)

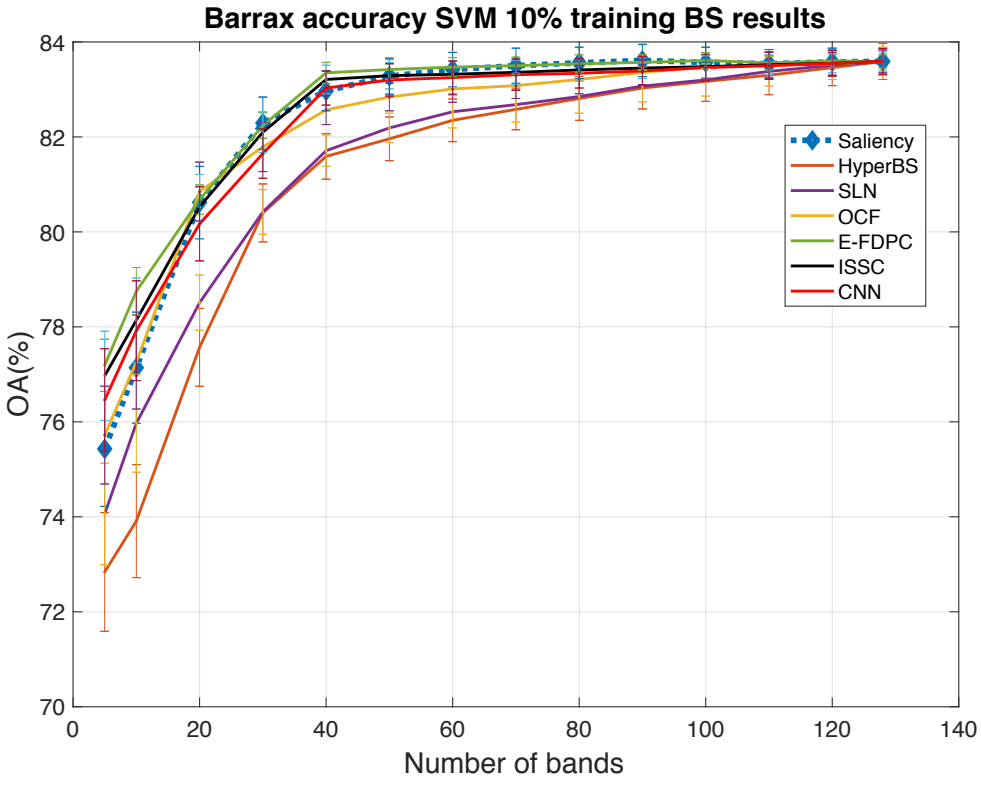


(b)

Figure 5-3: SVM classification accuracy (OA) of all the BS competing algorithms used in this study for the datasets of a) Botswana, b) KSC, c) Salinas, d) Barrax, e) Indian Pines and f) Pavia. The training data is 10% and experiments have been repeated for 5 times to get the average and the standard deviation.

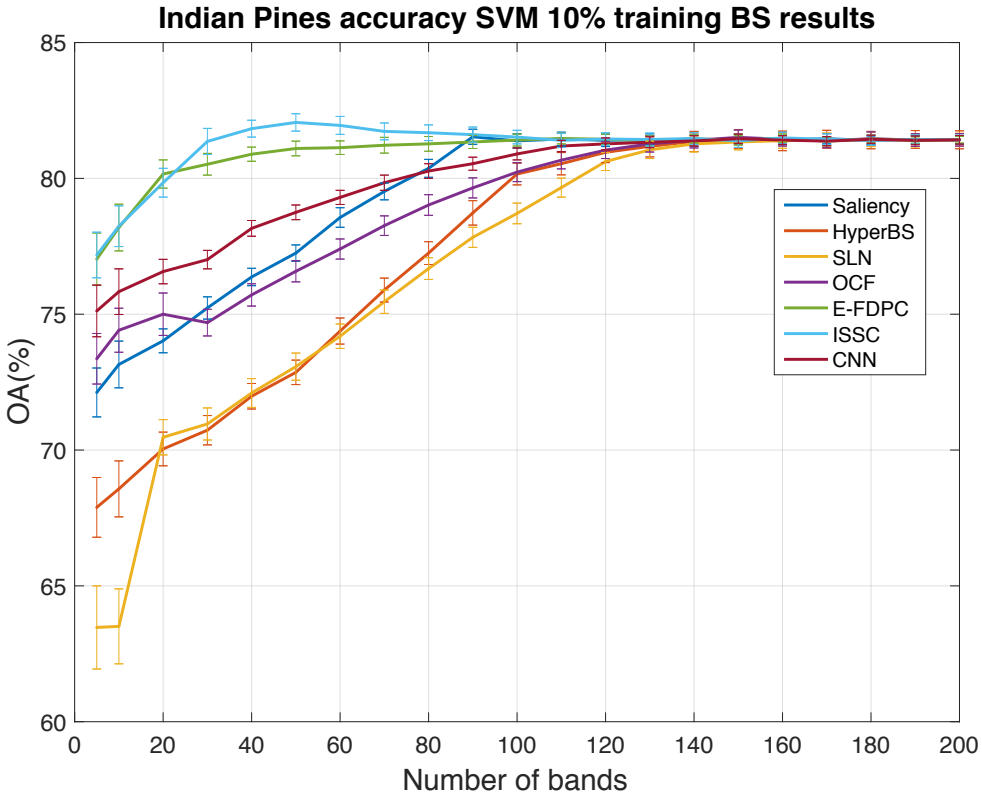


(c)

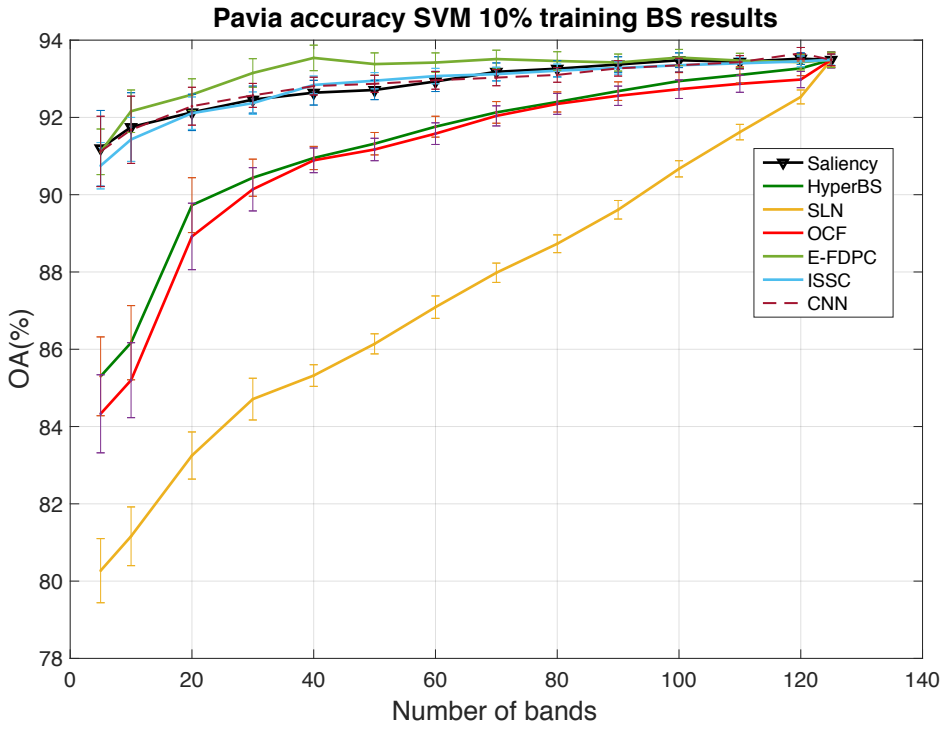


(d)

Figure 5-3: continuation.

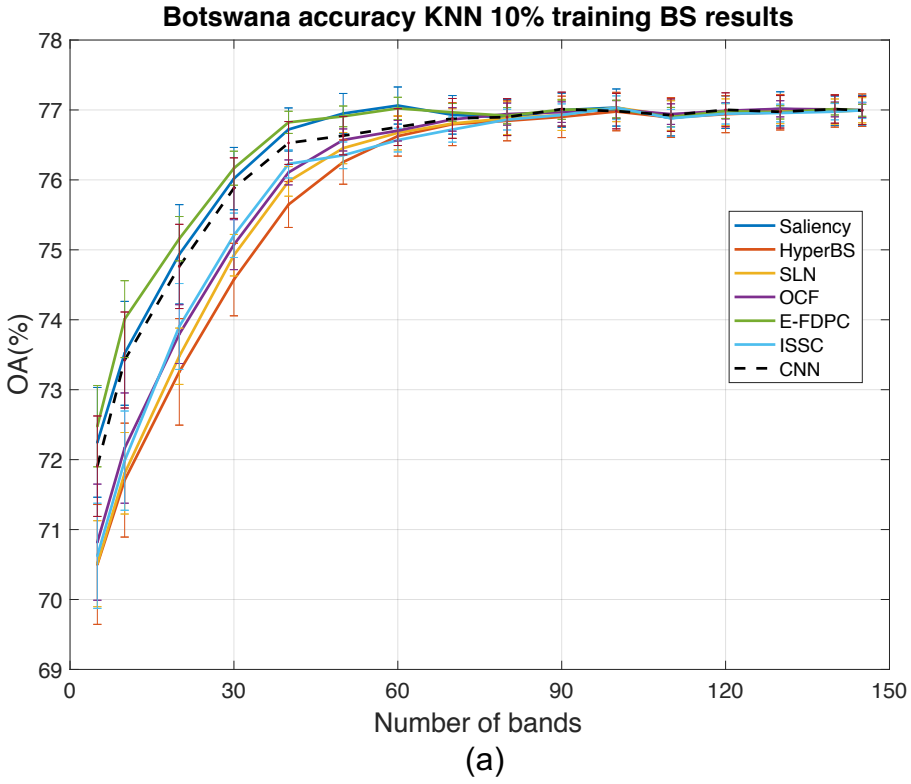


(e)



(f)

Figure 5-3: continuation.



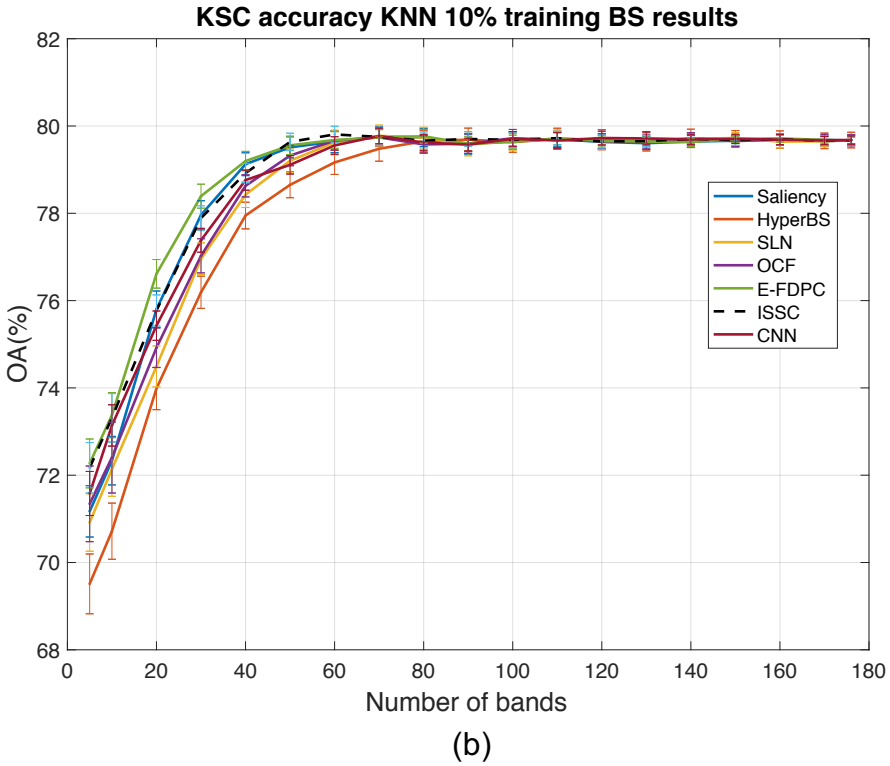
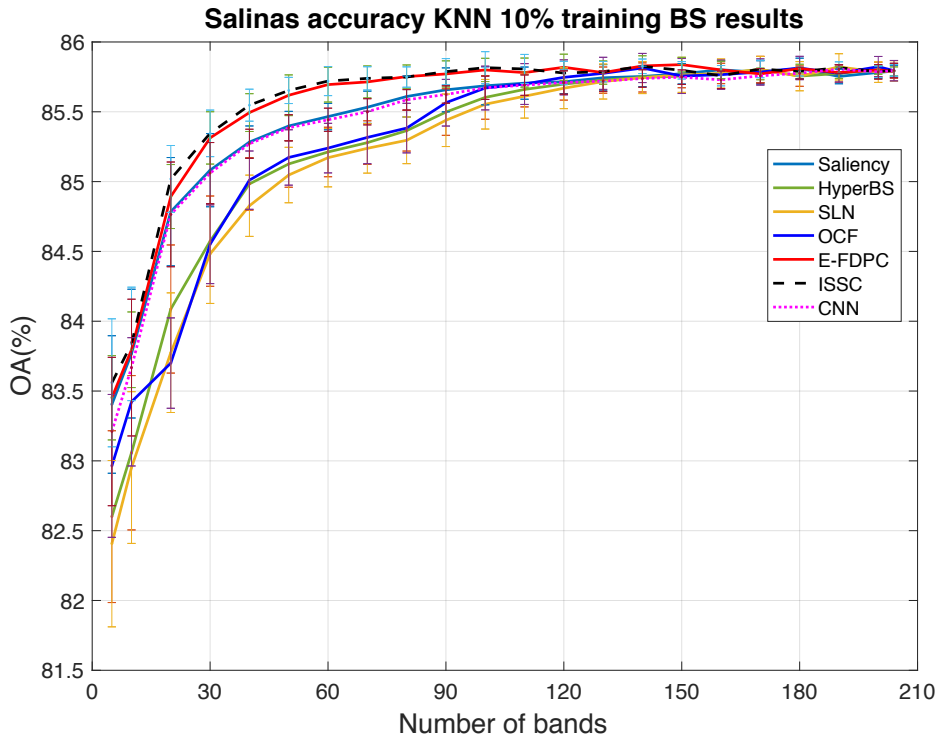
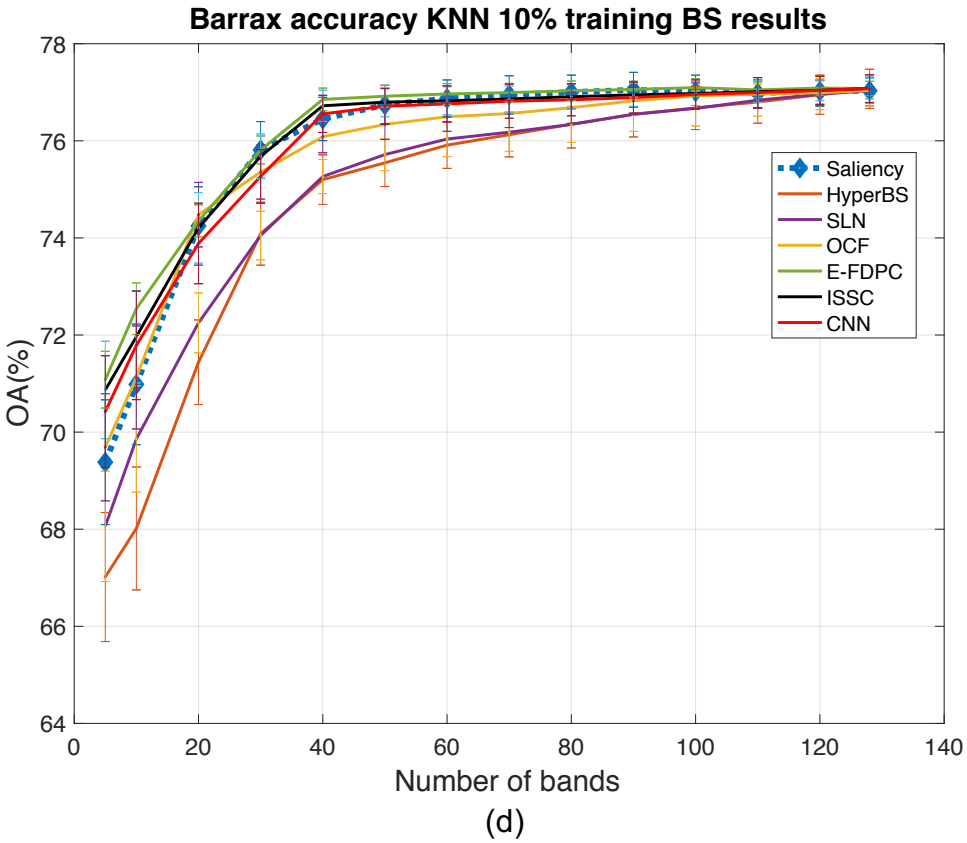


Figure 5-4: KNN classification accuracy (OA) of all the BS competing algorithms used in this study for the datasets of a) Botswana, b) KSC, c) Salinas, d) Barrax, e) Indian Pines and f) Pavia. The training data is 10% and experiments have been repeated for 5 times to get the average and the standard deviation.

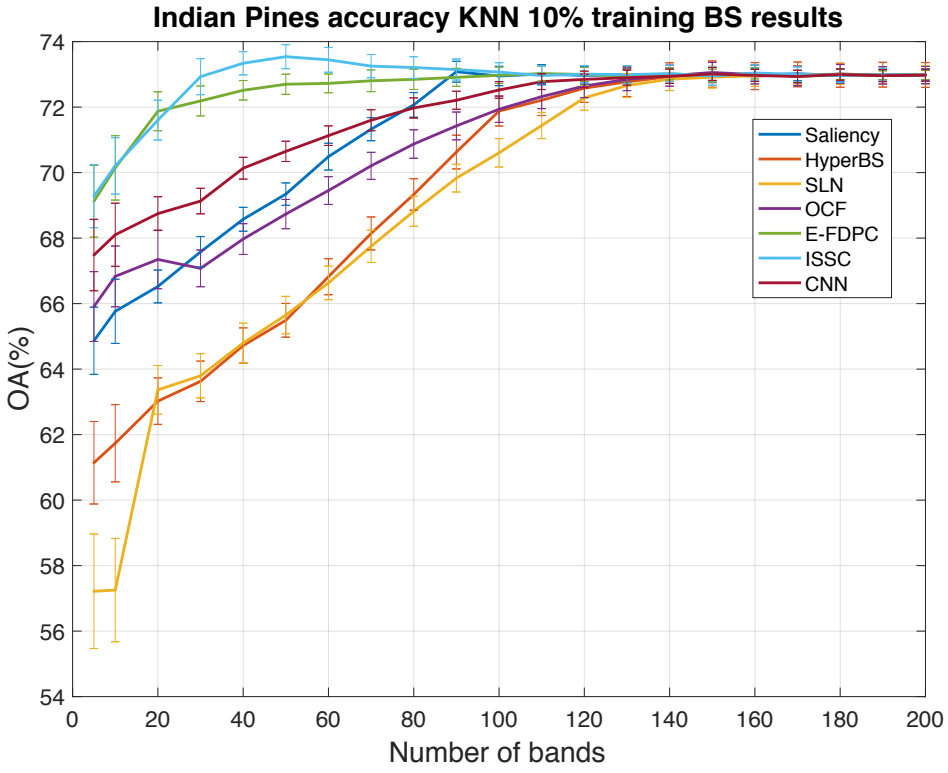


(c)

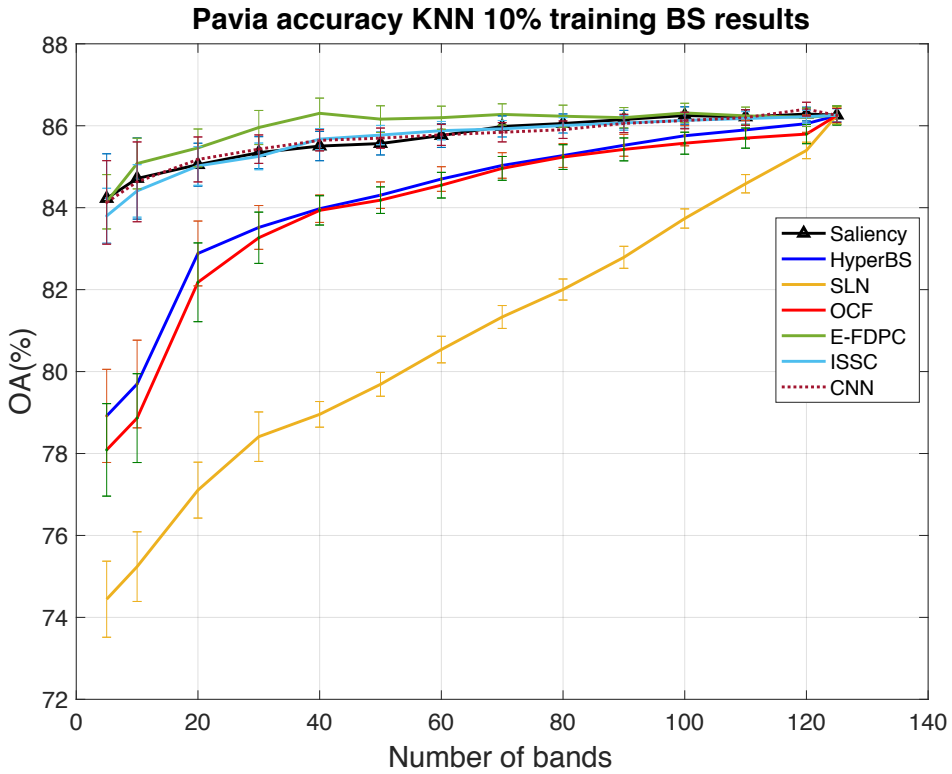


(d)

Figure 5-4: continuation.

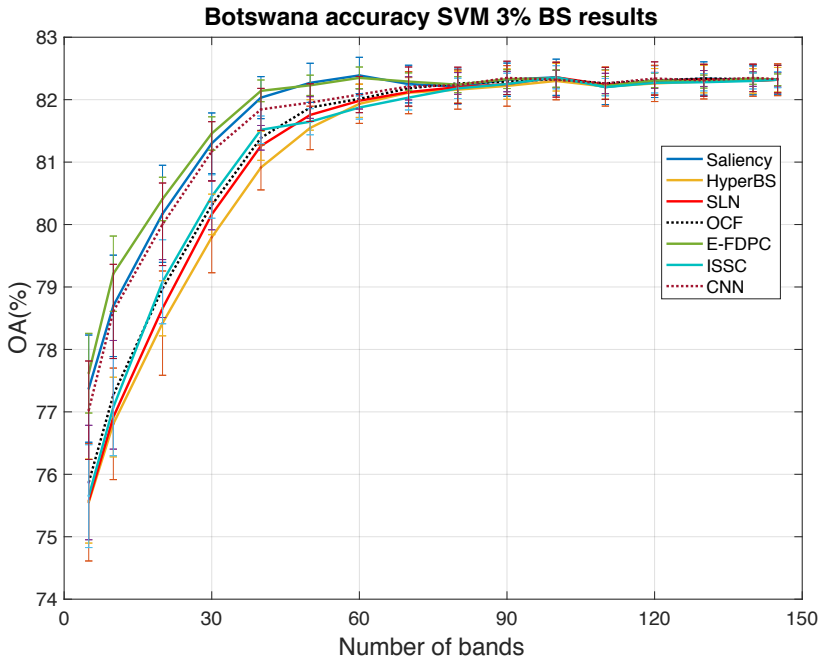


(e)

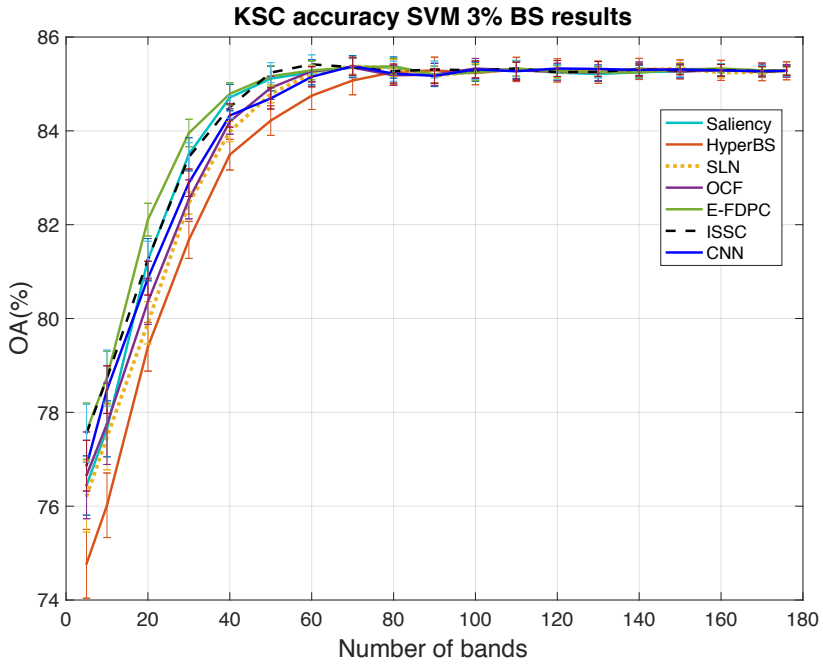


(f)

Figure 5-4: continuation.

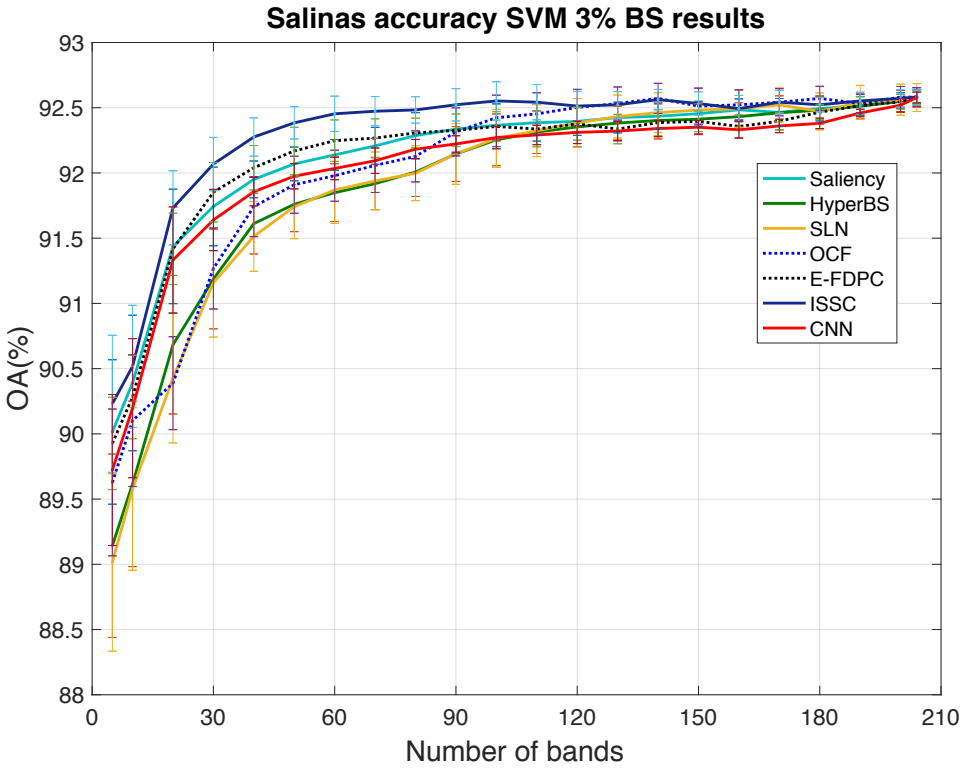


(a)

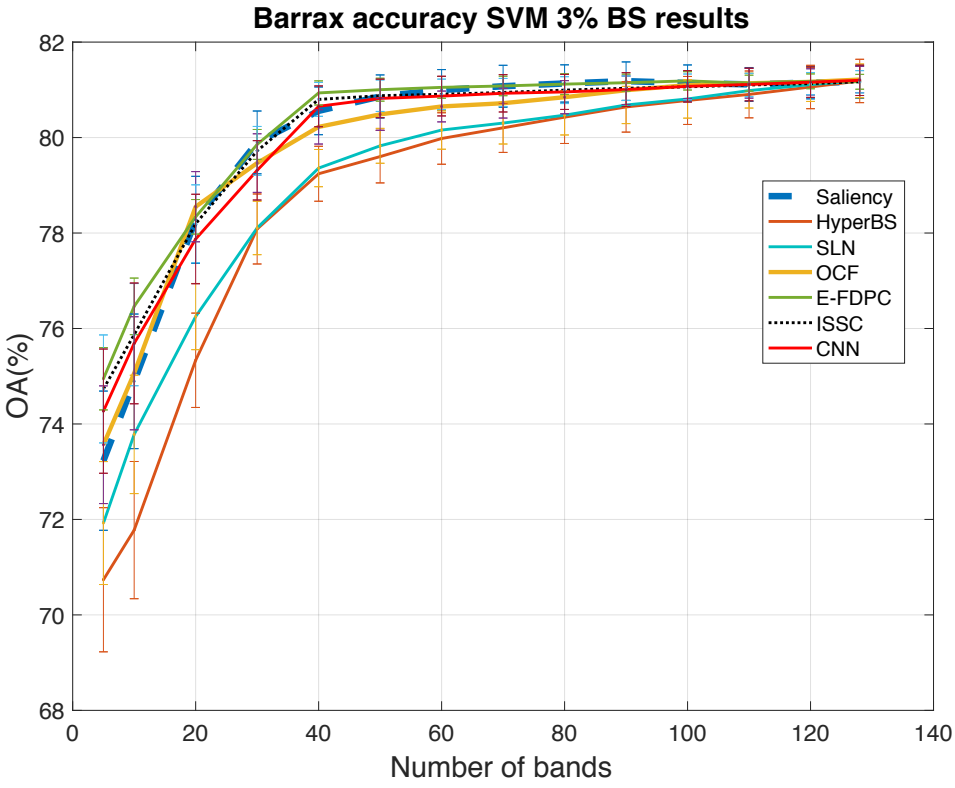


(b)

Figure 5-5: SVM classification accuracy (OA) of all the BS competing algorithms used in this study for the datasets of a) Botswana, b) KSC, c) Salinas, d) Barrax, e) Indian Pines and f) Pavia. The training data is 3% and experiments have been repeated for 5 times to get the average and the standard deviation.

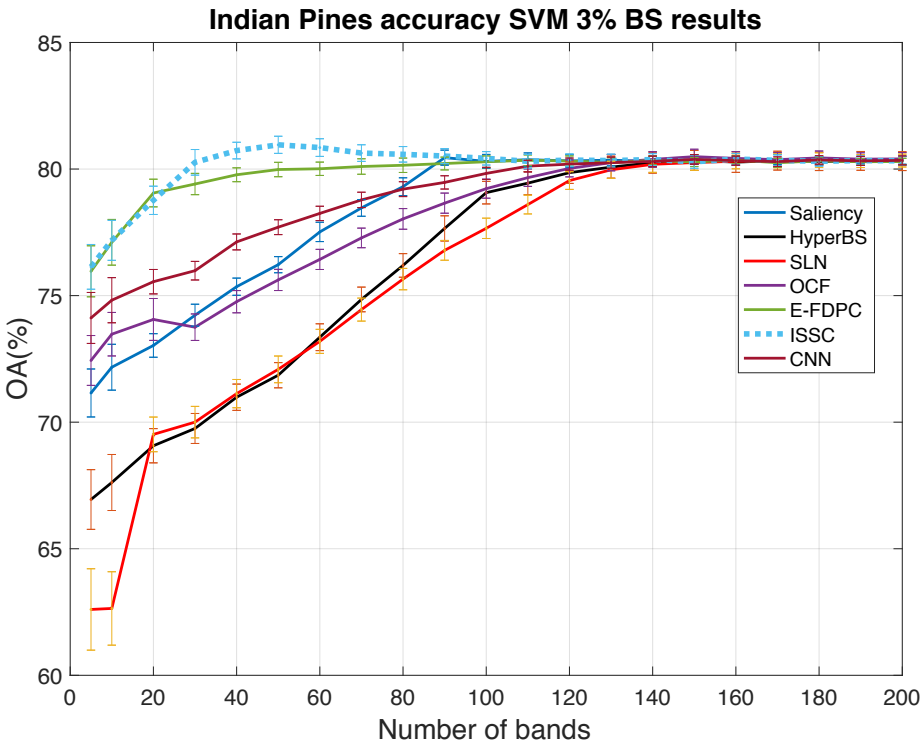


(c)

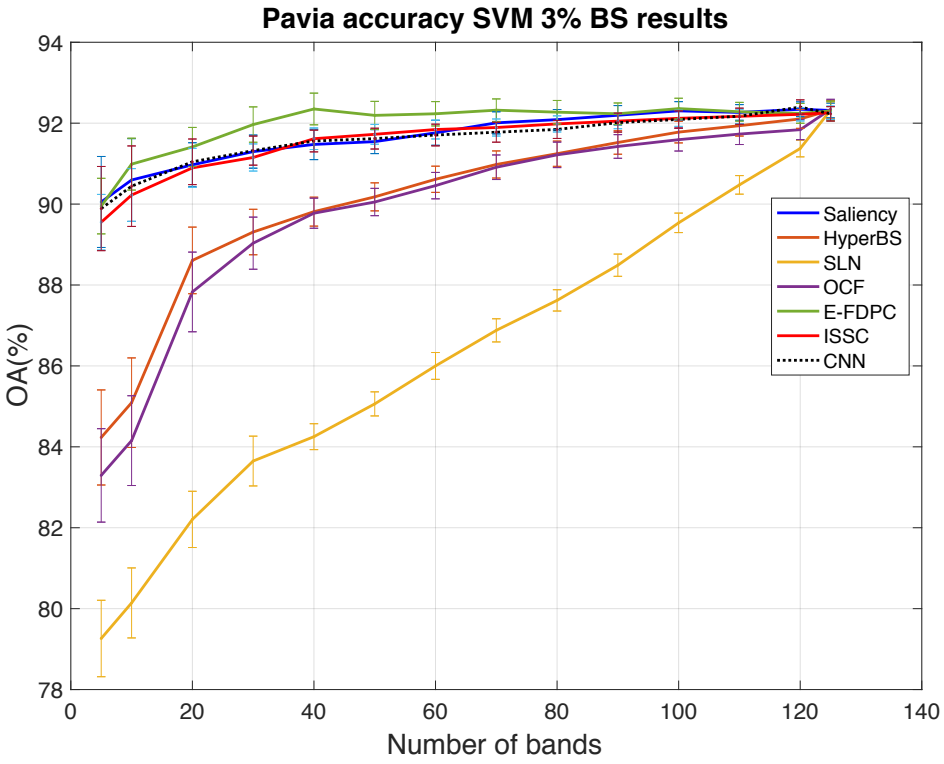


(d)

Figure 5-5: continuation.

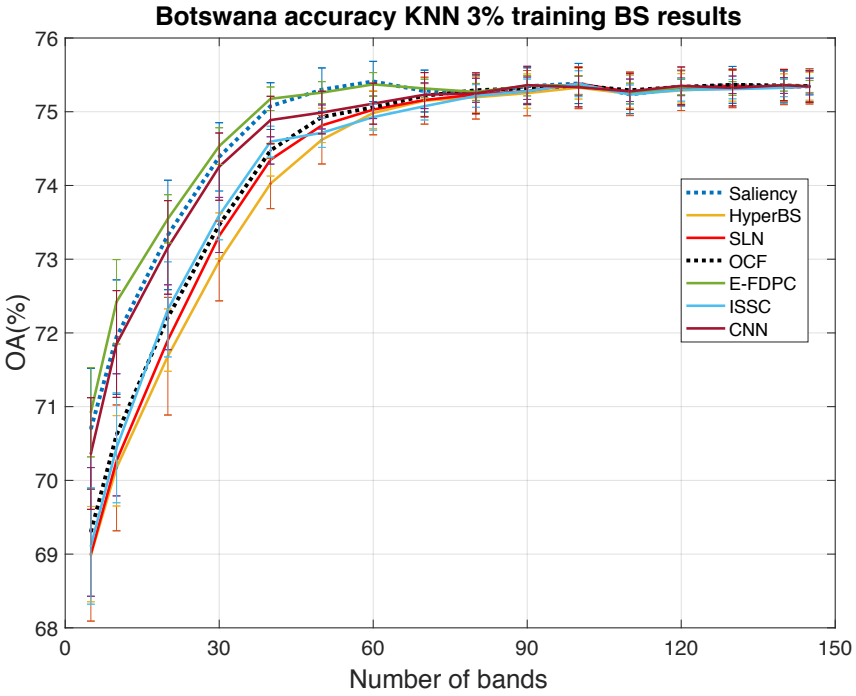


(e)

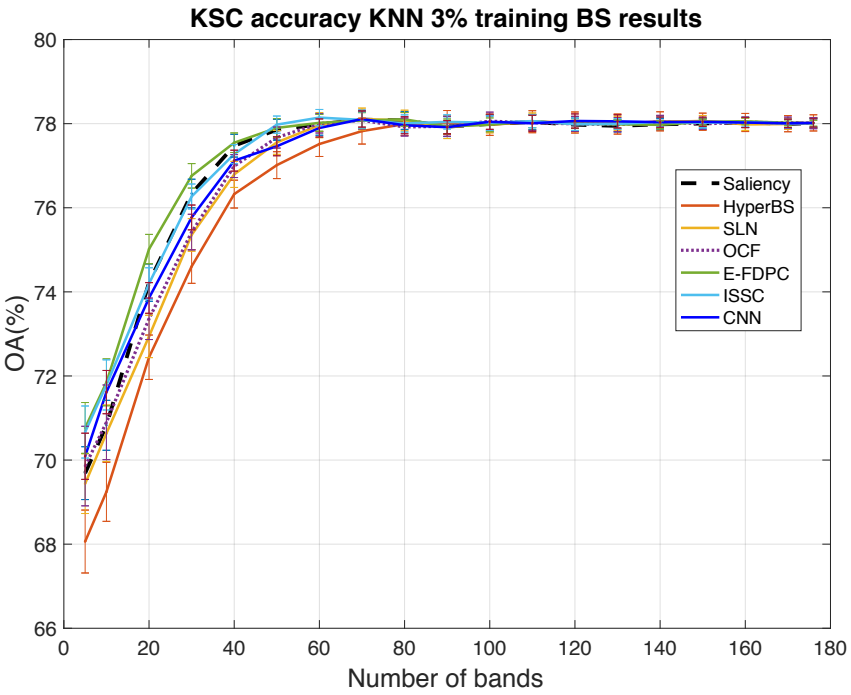


(f)

Figure 5-5: continuation.

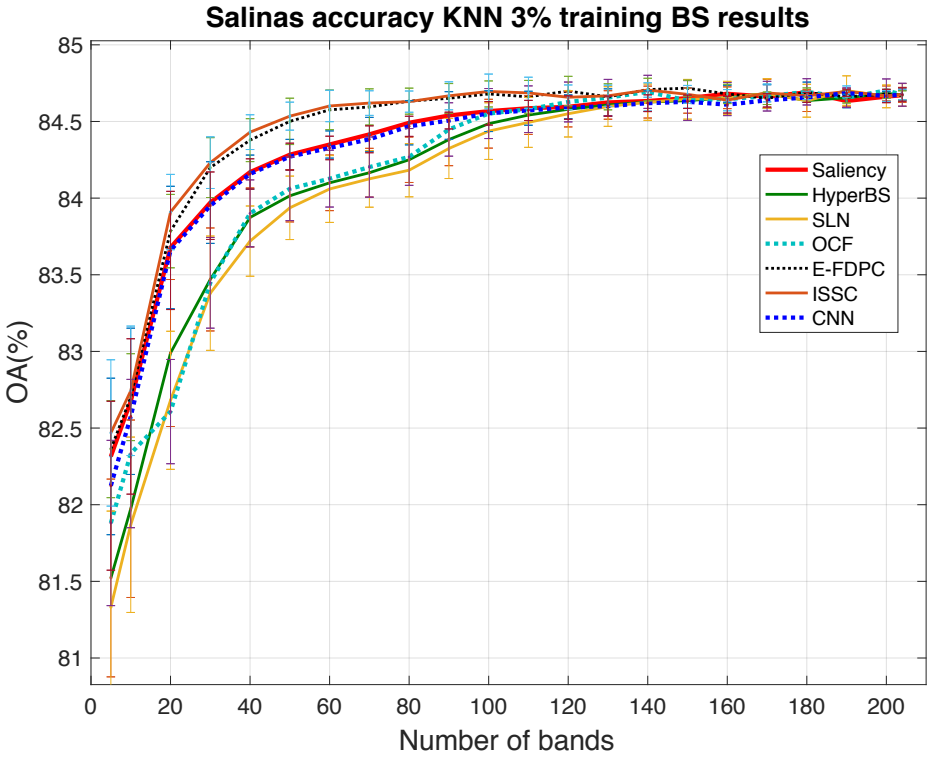


(a)

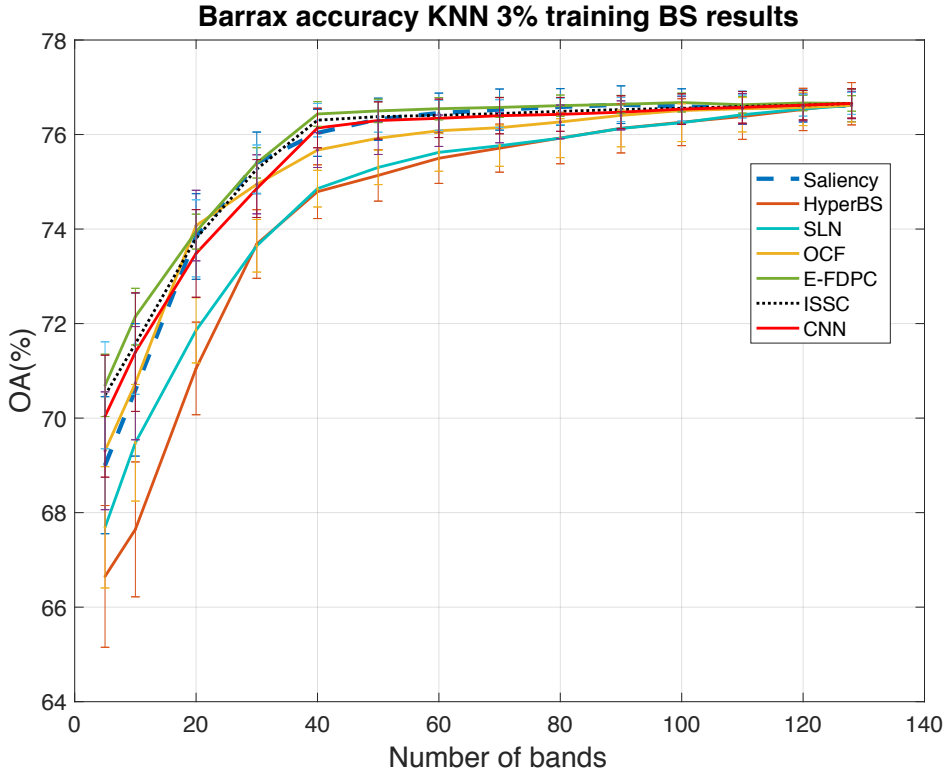


(b)

Figure 5-6: KNN classification accuracy (OA) of all the BS competing algorithms used in this study for the datasets of a) Botswana, b) KSC, c) Salinas, d) Barrax, e) Indian Pines and f) Pavia. The training data is 3% and experiments have been repeated for 5 times to get the average and the standard deviation.

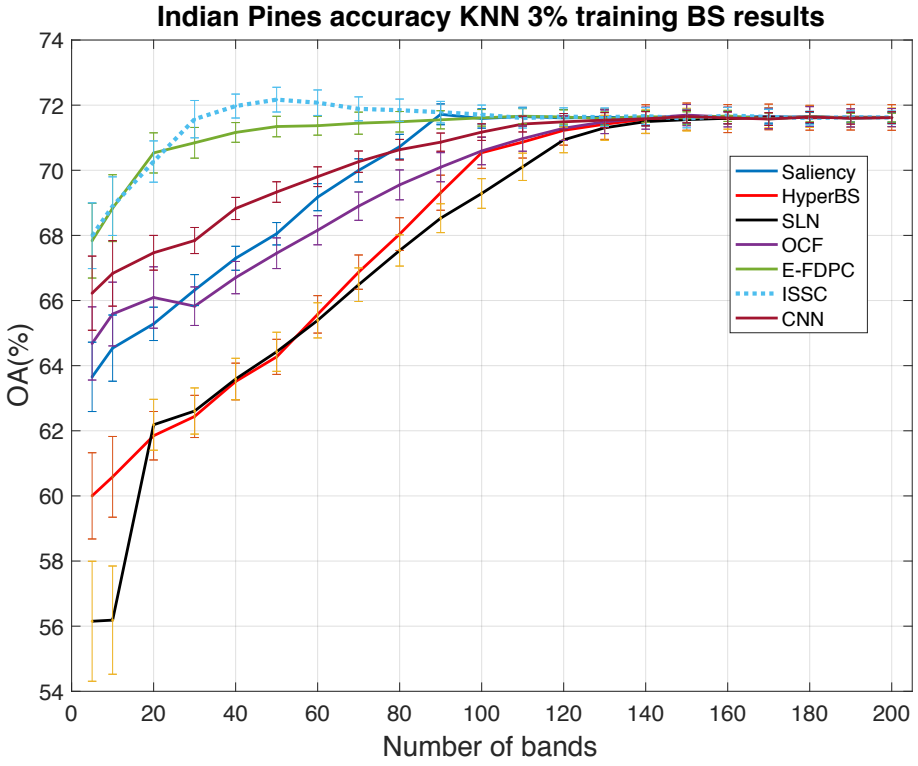


(c)

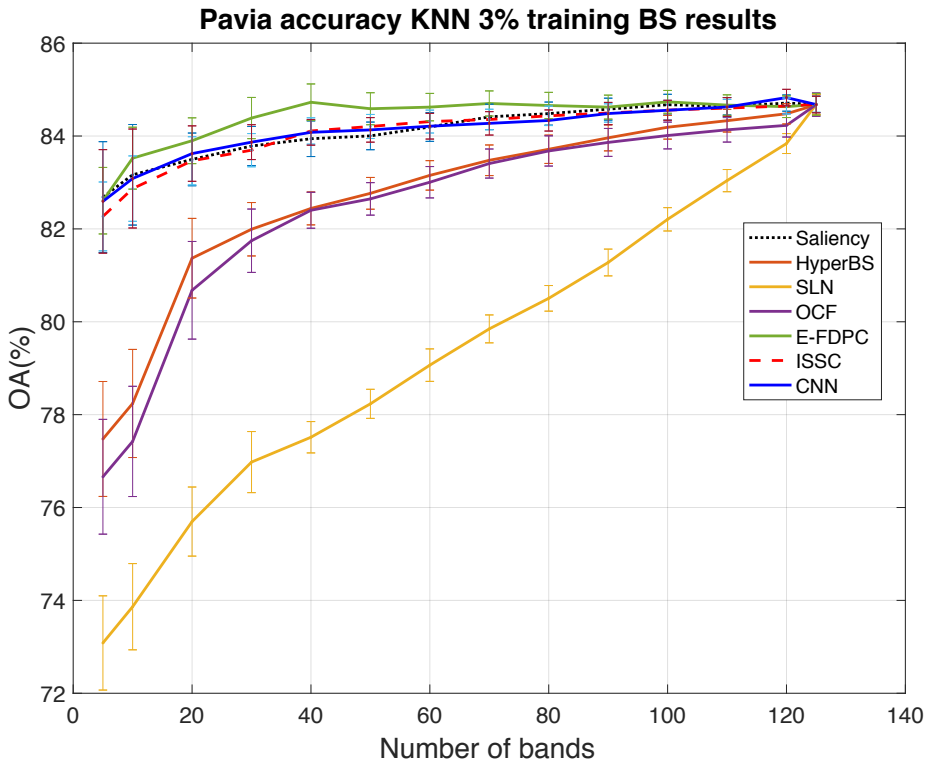


(d)

Figure 5-6: continuation.



(e)



(f)

Figure 5-6: continuation.

It can be seen from Figure 5-3 to 5-6 that the classification performance of SVM and KNN for these datasets is quite similar and the two figures gives almost identical trends of the classification performances, but the absolute accuracy values of SVM are ~5% higher than KNN values, for 10% and 3% training pixels. Amongst the competing algorithms, the E-FDPC and the ISSC perform the best while the SLN is the BS algorithm with the worst results. It is noted that all classification accuracy performance exhibits a 'knee'-shaped behaviour showing better accuracy when more bands have been used for the classification. This observation may contradict Hughes prediction [1], which can be because the BS algorithms that have been utilised here may not be efficient enough to show the 'intrinsic' accuracy curve as modelled by Hughes.

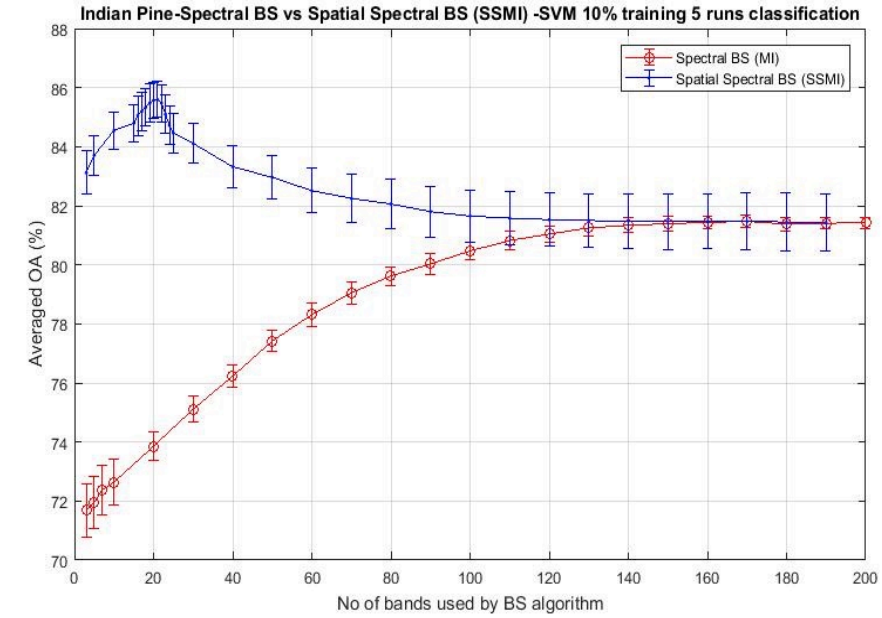
5.5.3 BS results of the proposed algorithm (SSMI)

Now we will compare the results obtained for the combination of edge [95] and MI (named as SSMI) with the results obtained for each one of our individual band selection algorithms. In this work, the marginal probabilities of the bands are estimated from the normalised histograms of each band image, and their corresponding entropies are then calculated. The joint entropy is subsequently evaluated for each pair of adjacent image bands and the mutual information (MI) of the image pair is evaluated. The I(A,B) (i.e. Mutual Information) of each dataset is then ranked and the topmost ones are then chosen in the order of the MI ranking.

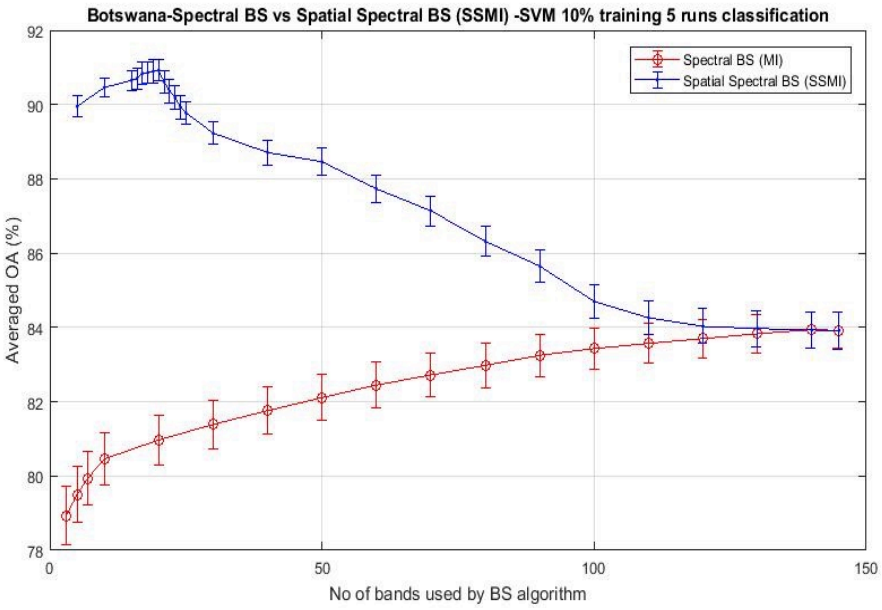
The spatial information has been traditionally regarded as an added advantage for the remote sensing data analysis, especially in applications related to ground-use classification; in order to see this, we are going to look first at the correlation coefficient values for the Indian Pines dataset:

First, the correlation coefficients C_i are ranked and the threshold is selected for band elimination. The remaining bands are then subsequently processed by the

basic MI scheme; the effectiveness of the MI and the proposed SSMI for the classification of a couple of well-studied datasets (Indian Pines and Botswana) is shown in Figure 5-7.



(a)



(b)

Figure 5-7: mean classification accuracy of two BS algorithms: the MI and the proposed SSMI, as function of the number of bands used for the classification: a) Indian Pines and b) Botswana datasets.

As seen on Figure 5-7, the classification accuracy reaches a peak at about 20 bands when the data is treated by the proposed SSMI, and the accuracy is then reduced steadily after the peak, when more bands are added to it; the training data that has been utilized in this experiment is also 10% per class for the classification of both datasets (Indian Pines and Botswana). We can also appreciate that the use of the edge algorithm first allows us to obtain an important improvement on the accuracy results.

It is seen that the peak accuracy is saturated at the point when all bands have been used for the classification; similar trends have been observed over many band selection papers reported in the literature [53, 91, 109-120], based on Hughes analysis [1]. Hughes research shows that the theoretical accuracy of a model scales non-linearly with the dimensionality of the dataset: the accuracy should be improved when more spectral bands are utilized for the classification, and further increasing of data dimensionality used to classify would reduce the classification accuracy when the training data size is kept constant.

This trend is observed for SSMI, while for MI the lack of a method that eliminated the low discrimination bands provokes that this decrease is not appreciated.

Figure 5-8 plots the I(A,B) value of the Indian Pines scene, which reveals that the I(A,B) of bands that are processed through SSMI (i.e. after the spatial treatment), have ~2 times higher contrasts (i.e. the peak and valley) than the I(A,B) processed by the MI.

This is probably one of the reasons that explains why SSMI performs better than MI alone; the red and black solid square markers shown in Fig. 5-8 depict the bands that have been chosen by the MI and the SSMI respectively, when the dimensionality is set to 5 (bands). In addition, we observe that MI alone seems to choose slightly suboptimal bands than SSMI.

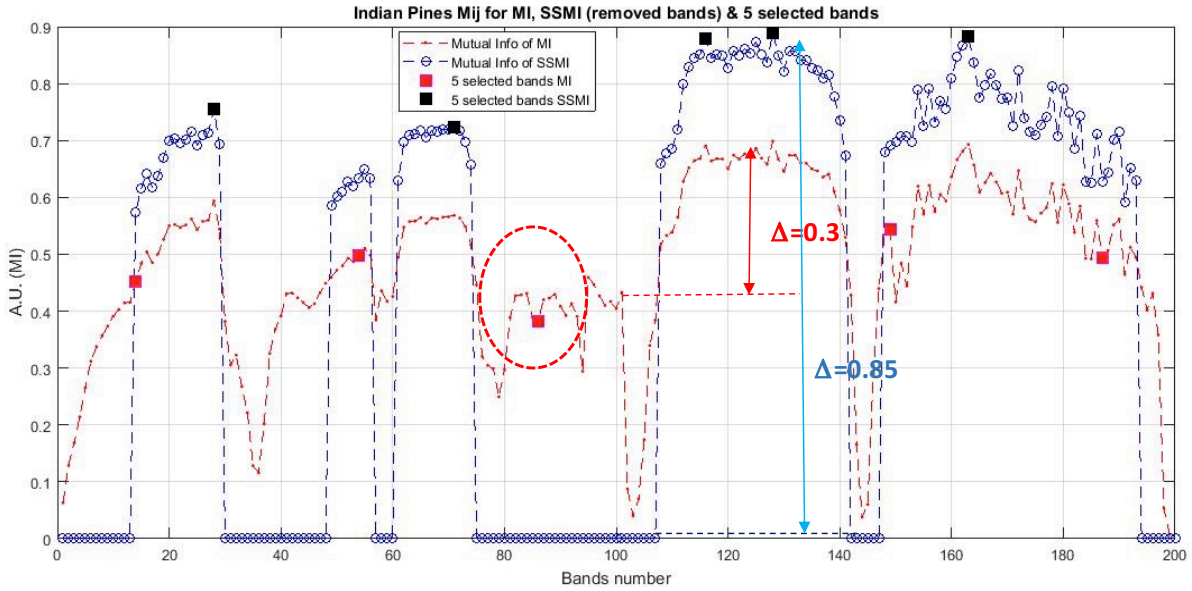


Figure 5-8: I(A,B) for the bands of the Indian Pines dataset given by the MI (red plot) and SSMI (blue plot) band selection methods. We note how that SSMI, which utilizes both spatial and spectral information, enhances the contrast of the I(A,B) in comparison to the basic MI band selection results without spatial pre-processing. The bands chosen by MI are: {14, 54, 86, 149, 187} and the bands chosen by SSMI are: {28, 71, 116, 128, 163}.

The robustness of the proposed SSMI is further examined by looking at its band selection capability with respect to the other competing algorithms: a) Saliency, b) HyperBS, c) SLN, d) OCF, e) FDPC, f) ISSC and g) CNN.

All experiments were conducted under the same parameters:

i) 10% per class of training data,

ii) The experiments were repeated 5 times,

iii) The SVM classifier was used for all cases.

Now the classification results of six HSI datasets: Botswana, Indian Pines, Barrax, KSC, Salinas and Pavia University, are presented at Figures 5-9, 5-10

and 5-11, together with Table 5-1. We can appreciate from these figures that the proposed SSMI achieves the best performance over our algorithms, reaching the classification accuracy peak at about 20 selected bands for our datasets.

At 20 bands, the SSMI band selection scheme obtains an improvement of ~10.5% with respect to the competing algorithms used on this thesis.

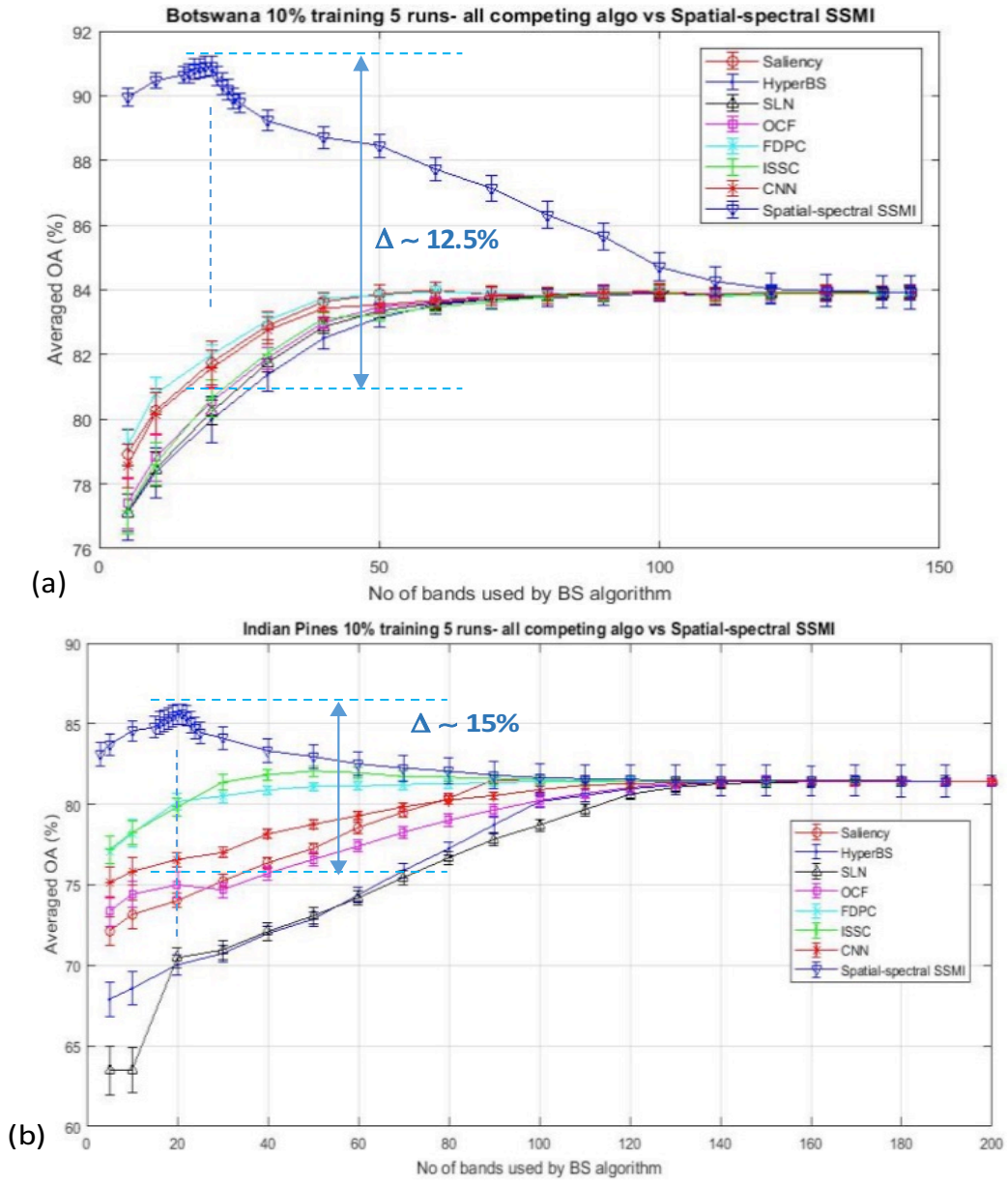
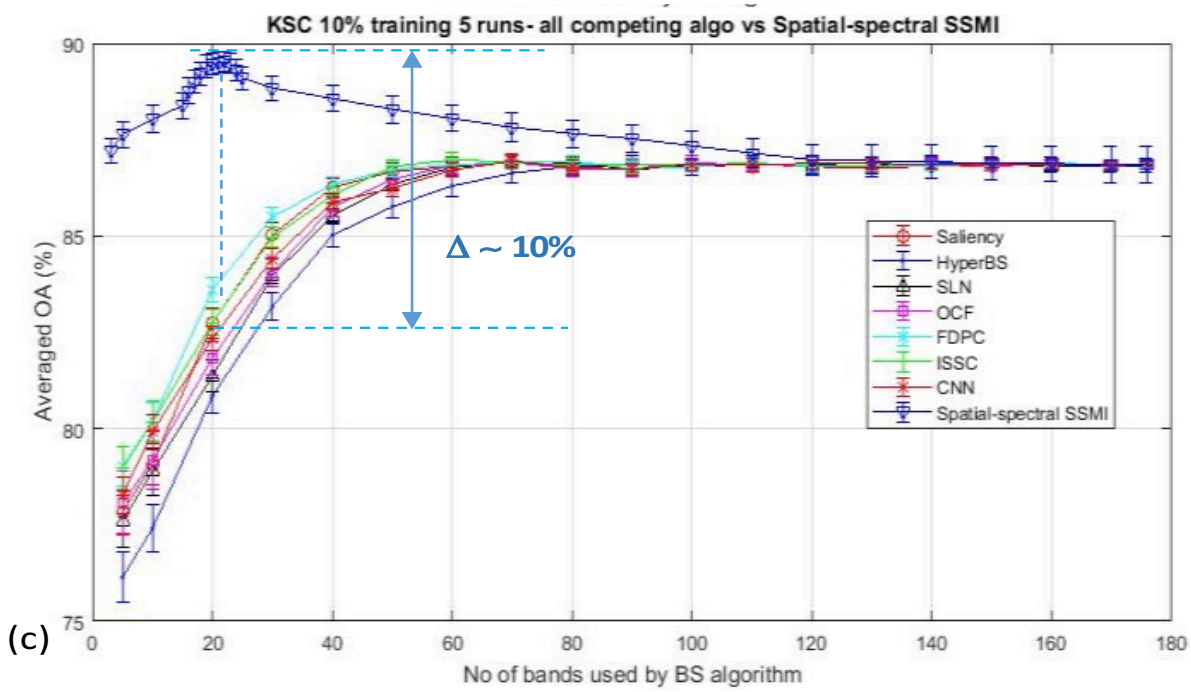
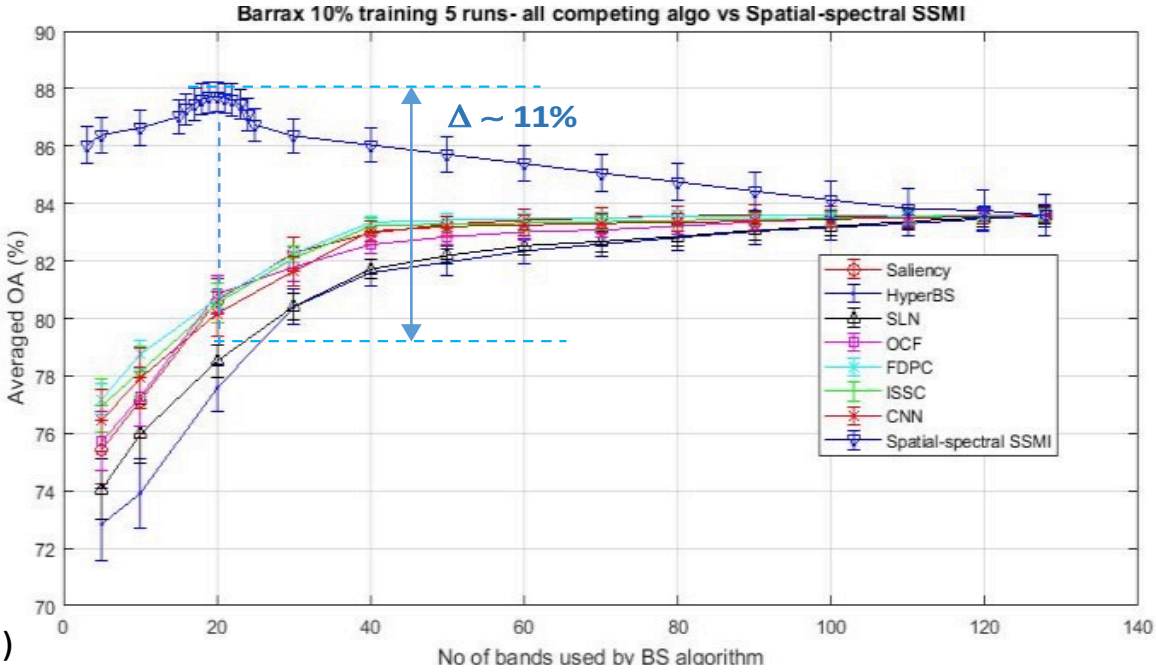


Figure 5-9: classification accuracy results of the proposed SSMI and the other band selection algorithms used on this section, for SVM (10% training pixels), for the datasets of a) Botswana, b) Indian Pines, c) KSC; d) Barrax, e) Salinas and f) Pavia University.



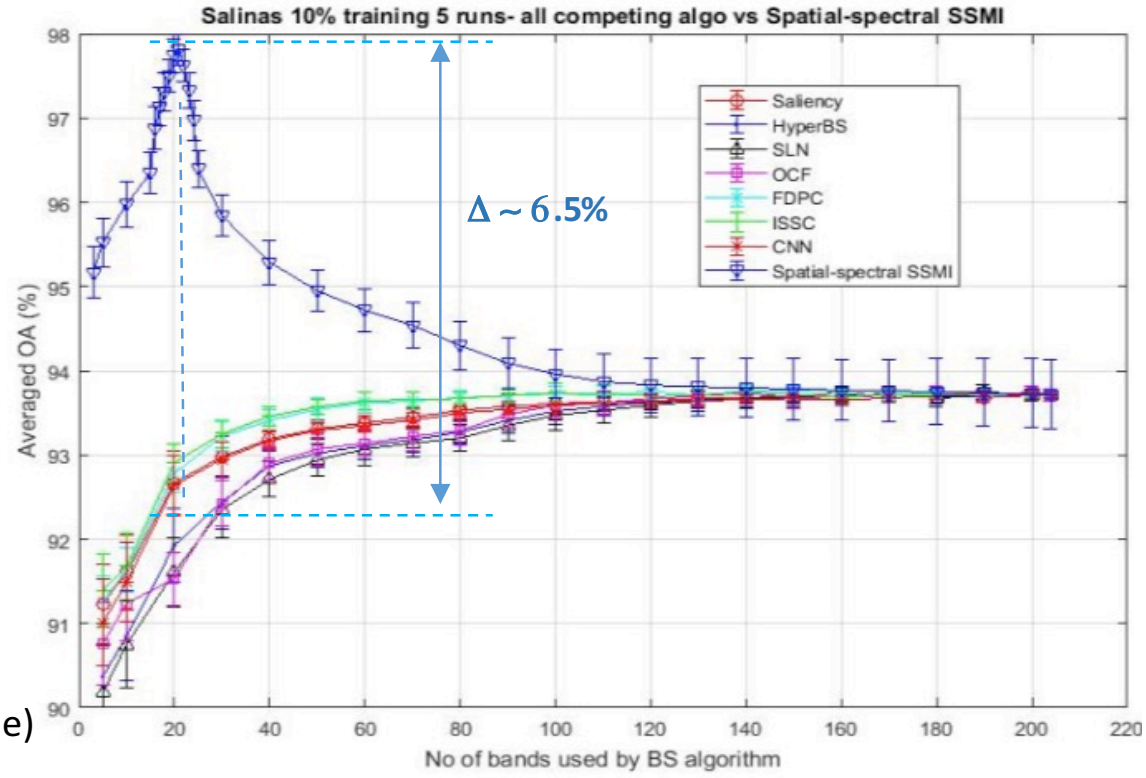
(c)



(d)

Figure 5-9: continuation.

(e)



(f)

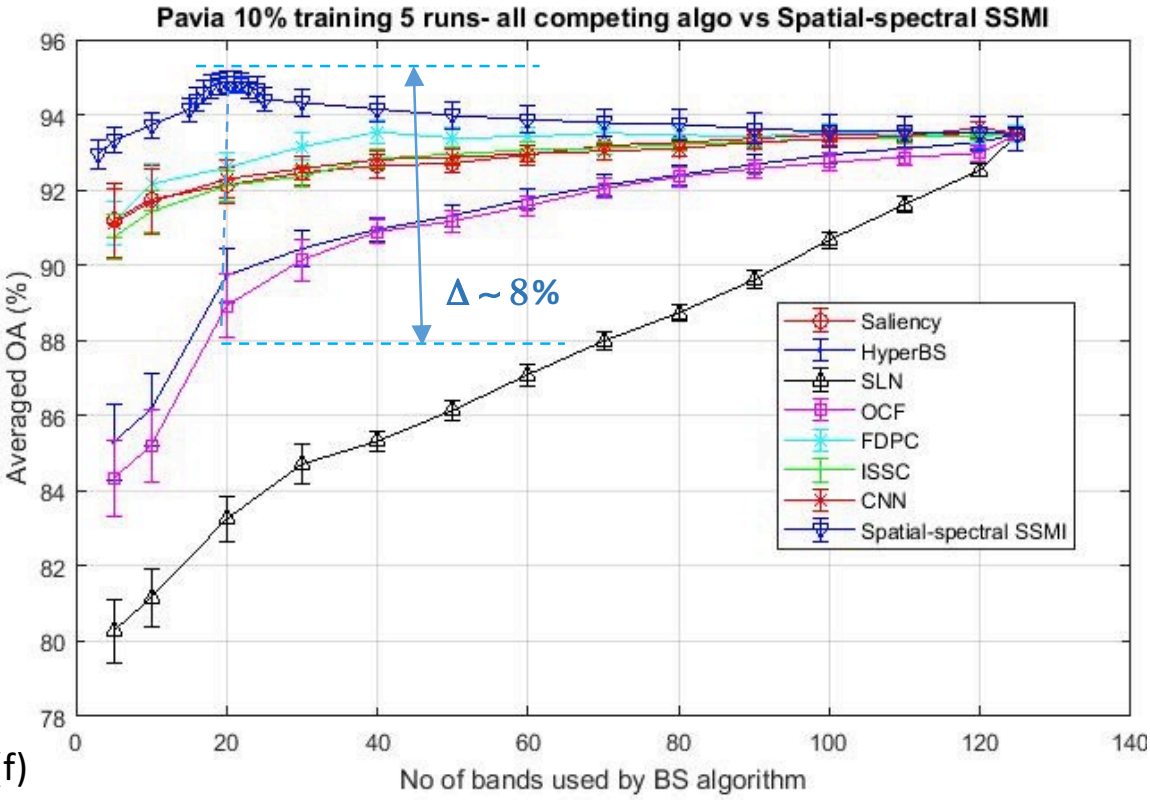


Figure 5-9: continuation.

	Botswana - SVM 10% training							
	CNN		E-FDPC		HyperBS		ISSC	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	78.56	0.68	79.17	0.55	77.08	0.81	77.18	0.71
10	80.16	0.65	80.78	0.52	78.35	0.77	78.62	0.67
20	81.57	0.57	81.99	0.3	79.98	0.72	80.65	0.58
30	82.75	0.41	83.05	0.23	81.37	0.49	82.03	0.3
40	83.43	0.29	83.74	0.15	82.5	0.31	83.11	0.19
50	83.54	0.26	83.83	0.14	83.14	0.3	83.24	0.18
60	83.67	0.25	83.95	0.15	83.53	0.27	83.47	0.16
70	83.8	0.27	83.89	0.12	83.71	0.29	83.63	0.17
80	83.82	0.25	83.84	0.11	83.76	0.27	83.78	0.14
90	83.94	0.23	83.93	0.13	83.82	0.28	83.85	0.15
100	83.91	0.24	83.94	0.12	83.9	0.26	83.96	0.16
110	83.85	0.22	83.82	0.13	83.81	0.27	83.8	0.12
120	83.93	0.23	83.91	0.1	83.86	0.25	83.87	0.14
130	83.9	0.22	83.89	0.09	83.89	0.24	83.88	0.12
140	83.94	0.19	83.94	0.08	83.91	0.22	83.9	0.11
145	83.92	0.19	83.92	0.08	83.92	0.22	83.92	0.11

	Botswana - SVM 10% training							
	OCF		Saliency		SLN		SSMI	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	77.4	0.79	78.92	0.74	77.1	0.58	89.95	0.28
10	78.82	0.75	80.26	0.7	78.46	0.55	90.46	0.24
20	80.54	0.4	81.75	0.67	80.22	0.38	90.92	0.29
30	81.89	0.34	82.89	0.42	81.74	0.28	89.23	0.31
40	82.98	0.17	83.63	0.29	82.85	0.2	88.71	0.34
50	83.47	0.15	83.87	0.27	83.35	0.21	88.46	0.36
60	83.61	0.14	83.99	0.25	83.58	0.23	87.73	0.37
70	83.78	0.16	83.85	0.26	83.72	0.2	87.14	0.4
80	83.86	0.15	83.81	0.24	83.8	0.22	86.32	0.41
90	83.89	0.14	83.92	0.23	83.85	0.21	85.65	0.43
100	83.93	0.12	83.96	0.25	83.96	0.19	84.7	0.44
110	83.86	0.14	83.8	0.24	83.84	0.2	84.26	0.46
120	83.91	0.11	83.89	0.22	83.89	0.18	84.03	0.47
130	83.94	0.11	83.94	0.23	83.93	0.19	83.97	0.49
140	83.93	0.1	83.91	0.2	83.9	0.18	83.93	0.5
145	83.92	0.11	83.92	0.21	83.92	0.17	83.92	0.49

(a)

Table 5-1: classification accuracy results of the proposed SSMI and the other band selection algorithms used on this section, for SVM (10% training pixels), for the datasets of a) Botswana, b) Indian Pines, c) KSC; d) Barrax, e) Salinas and f) Pavia University.

	KSC - SVM 10% training							
	CNN		E-FDPC		HyperBS		ISSC	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	78.27	0.48	79.02	0.53	76.14	0.65	78.97	0.55
10	79.92	0.45	80.19	0.49	77.41	0.61	80.17	0.53
20	82.34	0.32	83.61	0.31	80.85	0.46	82.75	0.34
30	84.41	0.26	85.49	0.26	83.17	0.35	84.97	0.27
40	85.87	0.22	86.34	0.21	85.02	0.29	86.05	0.21
50	86.24	0.2	86.72	0.2	85.76	0.28	86.8	0.19
60	86.71	0.19	86.85	0.18	86.3	0.26	86.98	0.18
70	86.94	0.17	86.91	0.17	86.63	0.27	86.92	0.18
80	86.78	0.18	86.94	0.15	86.81	0.24	86.83	0.19
90	86.73	0.14	86.76	0.16	86.85	0.25	86.87	0.16
100	86.87	0.16	86.8	0.15	86.79	0.22	86.85	0.17
110	86.83	0.17	86.89	0.14	86.87	0.23	86.89	0.15
120	86.89	0.15	86.83	0.11	86.85	0.22	86.81	0.14
130	86.88	0.14	86.81	0.1	86.81	0.21	86.82	0.12
140	86.86	0.11	86.8	0.09	86.88	0.2	86.87	0.11
150	86.87	0.1	86.87	0.07	86.87	0.18	86.85	0.12
160	86.85	0.11	86.89	0.08	86.85	0.19	86.86	0.11
170	86.83	0.09	86.85	0.07	86.82	0.17	86.83	0.1
176	86.84	0.08	86.84	0.08	86.84	0.16	86.84	0.1

	KSC - SVM 10% training							
	OCF		Saliency		SLN		SSMI	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	78.06	0.82	77.84	0.56	77.6	0.67	87.64	0.34
10	79.18	0.77	79.07	0.53	78.89	0.62	88.03	0.35
20	81.83	0.43	82.74	0.4	81.37	0.4	89.46	0.27
30	84.05	0.37	85.03	0.31	83.97	0.21	88.84	0.32
40	85.74	0.24	86.26	0.25	85.52	0.19	88.57	0.33
50	86.47	0.22	86.67	0.23	86.34	0.17	88.29	0.35
60	86.83	0.2	86.81	0.24	86.78	0.16	88.05	0.34
70	86.91	0.18	86.93	0.21	86.94	0.15	87.82	0.37
80	86.74	0.19	86.91	0.2	86.9	0.16	87.66	0.36
90	86.76	0.21	86.75	0.22	86.73	0.18	87.53	0.35
100	86.89	0.19	86.82	0.18	86.82	0.15	87.34	0.38
110	86.85	0.17	86.87	0.17	86.87	0.14	87.15	0.39
120	86.82	0.16	86.8	0.15	86.81	0.12	86.98	0.4
130	86.8	0.15	86.77	0.14	86.83	0.11	86.95	0.41
140	86.87	0.14	86.81	0.13	86.86	0.13	86.92	0.43
150	86.82	0.12	86.83	0.11	86.89	0.11	86.9	0.44
160	86.86	0.11	86.86	0.12	86.8	0.1	86.88	0.46
170	86.85	0.12	86.82	0.11	86.81	0.09	86.85	0.47
176	86.84	0.11	86.84	0.11	86.84	0.1	86.84	0.46

(b)

Table 5-1: continuation.

	Salinas - SVM 10% training							
	CNN		E-FDPC		HyperBS		ISSC	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	91.01	0.51	91.27	0.29	90.37	0.59	91.39	0.44
10	91.49	0.47	91.63	0.26	90.85	0.53	91.68	0.39
20	92.64	0.36	92.78	0.22	91.93	0.44	92.91	0.23
30	92.95	0.21	93.22	0.18	92.44	0.31	93.25	0.16
40	93.17	0.1	93.41	0.13	92.87	0.18	93.46	0.11
50	93.29	0.09	93.54	0.14	93.02	0.16	93.57	0.09
60	93.35	0.08	93.62	0.12	93.11	0.17	93.64	0.1
70	93.41	0.09	93.64	0.11	93.18	0.15	93.66	0.08
80	93.5	0.07	93.68	0.08	93.27	0.14	93.67	0.07
90	93.54	0.06	93.7	0.09	93.41	0.16	93.71	0.09
100	93.59	0.08	93.73	0.08	93.52	0.15	93.74	0.11
110	93.61	0.07	93.71	0.1	93.58	0.12	93.73	0.1
120	93.63	0.08	93.75	0.09	93.62	0.11	93.7	0.08
130	93.64	0.07	93.71	0.08	93.65	0.09	93.71	0.07
140	93.66	0.06	93.76	0.07	93.67	0.1	93.75	0.05
150	93.67	0.05	93.77	0.05	93.68	0.08	93.72	0.06
160	93.65	0.06	93.73	0.06	93.7	0.07	93.68	0.07
170	93.68	0.05	93.7	0.04	93.73	0.09	93.73	0.06
180	93.7	0.04	93.74	0.03	93.68	0.07	93.71	0.03
190	93.73	0.03	93.71	0.04	93.7	0.06	93.74	0.04
200	93.71	0.05	93.73	0.03	93.73	0.05	93.71	0.04
204	93.72	0.04	93.72	0.02	93.72	0.06	93.72	0.03

	Salinas - SVM 10% training							
	OCF		Saliency		SLN		SSMI	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	90.76	0.49	91.23	0.47	90.18	0.57	95.52	0.28
10	91.24	0.44	91.61	0.44	90.75	0.52	95.98	0.27
20	91.53	0.31	92.67	0.37	91.61	0.41	97.74	0.2
30	92.42	0.27	92.98	0.25	92.35	0.34	95.84	0.25
40	92.9	0.2	93.19	0.11	92.71	0.21	95.29	0.26
50	93.07	0.19	93.31	0.1	92.94	0.19	94.95	0.24
60	93.14	0.17	93.38	0.09	93.07	0.2	94.72	0.25
70	93.22	0.18	93.45	0.11	93.14	0.17	94.54	0.27
80	93.29	0.17	93.53	0.14	93.2	0.16	94.3	0.29
90	93.48	0.16	93.58	0.15	93.35	0.18	94.09	0.3
100	93.59	0.15	93.61	0.13	93.47	0.17	93.96	0.29
110	93.62	0.14	93.63	0.11	93.53	0.15	93.87	0.33
120	93.67	0.12	93.64	0.08	93.59	0.14	93.83	0.32
130	93.7	0.11	93.67	0.09	93.64	0.12	93.81	0.34
140	93.74	0.1	93.68	0.07	93.67	0.11	93.8	0.35
150	93.68	0.12	93.7	0.07	93.69	0.12	93.78	0.37
160	93.69	0.1	93.73	0.06	93.7	0.1	93.77	0.36
170	93.71	0.09	93.71	0.08	93.73	0.09	93.76	0.37
180	93.74	0.08	93.74	0.07	93.68	0.1	93.75	0.39
190	93.7	0.06	93.68	0.05	93.75	0.09	93.74	0.4
200	93.75	0.07	93.71	0.04	93.71	0.07	93.73	0.41
204	93.72	0.06	93.72	0.03	93.72	0.07	93.72	0.4

(c)

Table 5-1: continuation.

	Barrax - SVM 10% training							
	CNN		E-FDPC		HyperBS		ISSC	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	76.46	1.08	77.19	0.55	72.84	1.25	76.97	0.94
10	77.92	1.05	78.75	0.5	73.91	1.19	78.14	0.89
20	80.17	0.78	80.68	0.31	77.57	0.82	80.53	0.68
30	81.65	0.52	82.24	0.27	80.4	0.61	82.1	0.43
40	83.03	0.36	83.35	0.22	81.59	0.48	83.21	0.3
50	83.2	0.34	83.42	0.21	81.96	0.46	83.29	0.28
60	83.25	0.35	83.47	0.2	82.35	0.45	83.32	0.27
70	83.31	0.33	83.5	0.18	82.58	0.43	83.36	0.25
80	83.34	0.31	83.54	0.19	82.81	0.46	83.41	0.24
90	83.39	0.3	83.57	0.16	83.03	0.44	83.45	0.22
100	83.46	0.28	83.61	0.17	83.17	0.42	83.48	0.23
110	83.5	0.29	83.56	0.15	83.3	0.41	83.52	0.21
120	83.55	0.27	83.6	0.14	83.46	0.38	83.55	0.2
128	83.59	0.25	83.59	0.14	83.59	0.37	83.59	0.2

	Barrax - SVM 10% training							
	OCF		Saliency		SLN		SSMI	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	75.72	1.03	75.43	1.21	74.06	1.07	86.39	0.62
10	77.26	0.99	77.14	1.17	75.97	1.03	86.63	0.61
20	80.85	0.62	80.62	0.76	78.51	0.58	87.72	0.51
30	81.79	0.52	82.29	0.55	80.42	0.47	86.35	0.59
40	82.57	0.31	82.97	0.42	81.71	0.33	86.04	0.61
50	82.84	0.29	83.28	0.38	82.19	0.31	85.71	0.62
60	83.01	0.28	83.42	0.36	82.53	0.34	85.39	0.6
70	83.08	0.27	83.5	0.37	82.68	0.37	85.06	0.64
80	83.21	0.3	83.56	0.33	82.85	0.35	84.75	0.65
90	83.36	0.26	83.61	0.34	83.07	0.33	84.43	0.68
100	83.47	0.25	83.57	0.32	83.2	0.34	84.12	0.67
110	83.51	0.27	83.54	0.3	83.38	0.31	83.83	0.7
120	83.54	0.24	83.58	0.29	83.51	0.3	83.75	0.72
128	83.59	0.23	83.59	0.28	83.59	0.29	83.59	0.72

(d)

Table 5-1: continuation.

	Indian Pines - SVM 10% training							
	CNN		E-FDPC		HyperBS		ISSC	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	75.12	0.95	77.03	0.96	67.89	1.1	77.18	0.84
10	75.83	0.84	78.19	0.86	68.57	1.03	78.24	0.75
20	76.57	0.45	80.16	0.52	70.04	0.62	79.84	0.53
30	77.01	0.34	80.52	0.4	70.73	0.54	81.36	0.48
40	78.16	0.29	80.89	0.26	71.98	0.47	81.83	0.31
50	78.75	0.27	81.1	0.27	72.86	0.45	82.06	0.32
60	79.3	0.26	81.13	0.25	74.38	0.48	81.95	0.33
70	79.84	0.28	81.22	0.29	75.89	0.44	81.73	0.31
80	80.27	0.27	81.27	0.27	77.25	0.42	81.68	0.29
90	80.54	0.24	81.34	0.24	78.73	0.45	81.61	0.28
100	80.9	0.22	81.41	0.22	80.16	0.4	81.52	0.25
110	81.19	0.21	81.47	0.2	80.54	0.41	81.4	0.26
120	81.27	0.22	81.44	0.19	80.96	0.38	81.45	0.23
130	81.33	0.2	81.38	0.18	81.18	0.39	81.43	0.24
140	81.38	0.19	81.46	0.16	81.36	0.37	81.47	0.22
150	81.46	0.17	81.4	0.17	81.45	0.35	81.38	0.24
160	81.41	0.16	81.48	0.15	81.38	0.36	81.49	0.21
170	81.37	0.17	81.37	0.16	81.44	0.33	81.44	0.2
180	81.44	0.15	81.45	0.14	81.41	0.32	81.4	0.18
190	81.4	0.16	81.41	0.12	81.43	0.33	81.43	0.17
200	81.42	0.15	81.42	0.11	81.42	0.32	81.42	0.16

	Indian Pines - SVM 10% training							
	OCF		Saliency		SLN		SSMI	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	73.36	0.93	72.12	0.9	63.47	1.53	83.53	0.68
10	74.41	0.81	73.15	0.86	63.51	1.38	84.37	0.63
20	75	0.78	74.02	0.44	70.47	0.65	85.39	0.62
30	74.69	0.49	75.23	0.41	70.96	0.59	83.93	0.68
40	75.71	0.41	76.37	0.32	72.1	0.53	83.14	0.71
50	76.58	0.39	77.25	0.3	73.07	0.5	82.78	0.73
60	77.4	0.37	78.56	0.36	74.19	0.45	82.33	0.76
70	78.26	0.36	79.52	0.31	75.46	0.43	82.07	0.81
80	79.02	0.38	80.37	0.33	76.68	0.4	81.88	0.83
90	79.65	0.37	81.53	0.28	77.83	0.37	81.62	0.86
100	80.23	0.35	81.38	0.26	78.71	0.38	81.47	0.87
110	80.67	0.32	81.45	0.25	79.66	0.35	81.4	0.89
120	81.04	0.31	81.4	0.23	80.62	0.33	81.45	0.9
130	81.27	0.3	81.43	0.21	81.05	0.32	81.51	0.91
140	81.4	0.28	81.38	0.2	81.27	0.3	81.5	0.92
150	81.51	0.27	81.45	0.18	81.34	0.29	81.48	0.91
160	81.44	0.25	81.41	0.17	81.39	0.28	81.47	0.93
170	81.39	0.26	81.46	0.19	81.41	0.26	81.46	0.94
180	81.46	0.25	81.4	0.16	81.44	0.27	81.45	0.96
190	81.41	0.23	81.43	0.14	81.4	0.24	81.44	0.95
200	81.42	0.22	81.42	0.15	81.42	0.24	81.43	0.94

(e)

Table 5-1: continuation.

	Pavia University - SVM 10% training							
	CNN		E-FDPC		HyperBS		ISSC	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	91.12	0.91	91.11	0.59	85.3	1.02	90.75	0.6
10	91.68	0.87	92.16	0.55	86.17	0.96	91.43	0.57
20	92.29	0.49	92.59	0.41	89.73	0.71	92.11	0.42
30	92.57	0.31	93.15	0.37	90.44	0.48	92.37	0.29
40	92.81	0.23	93.54	0.33	90.95	0.3	92.84	0.23
50	92.87	0.22	93.38	0.29	91.32	0.29	92.95	0.21
60	92.96	0.23	93.42	0.25	91.76	0.27	93.07	0.2
70	93.03	0.21	93.51	0.23	92.13	0.28	93.12	0.18
80	93.1	0.19	93.46	0.24	92.4	0.26	93.21	0.17
90	93.27	0.2	93.42	0.22	92.68	0.24	93.28	0.16
100	93.35	0.18	93.55	0.21	92.94	0.22	93.35	0.17
110	93.43	0.17	93.47	0.19	93.1	0.21	93.4	0.15
120	93.66	0.15	93.43	0.2	93.27	0.19	93.45	0.14
125	93.49	0.14	93.49	0.19	93.49	0.18	93.49	0.15

	Pavia University - SVM 10% training							
	OCF		Saliency		SLN		SSMI	
Bands	OA(%)	Std	OA(%)	Std	OA(%)	Std	OA(%)	Std
5	84.33	1.01	91.2	0.98	80.27	0.83	93.31	0.34
10	85.2	0.97	91.75	0.89	81.16	0.76	93.72	0.35
20	88.92	0.86	92.13	0.47	83.25	0.61	94.87	0.28
30	90.14	0.56	92.46	0.35	84.71	0.54	94.32	0.36
40	90.89	0.32	92.64	0.32	85.32	0.28	94.15	0.34
50	91.17	0.29	92.71	0.25	86.14	0.26	93.98	0.35
60	91.58	0.28	92.93	0.26	87.09	0.29	93.87	0.37
70	92.04	0.26	93.18	0.23	87.98	0.25	93.79	0.38
80	92.35	0.27	93.26	0.21	88.73	0.23	93.75	0.4
90	92.56	0.25	93.37	0.2	89.61	0.24	93.63	0.41
100	92.73	0.24	93.48	0.19	90.67	0.21	93.56	0.43
110	92.87	0.22	93.44	0.16	91.62	0.2	93.55	0.42
120	92.98	0.21	93.52	0.15	92.53	0.18	93.51	0.45
125	93.49	0.2	93.49	0.14	93.49	0.19	93.49	0.45

(f)

Table 5-1: continuation.

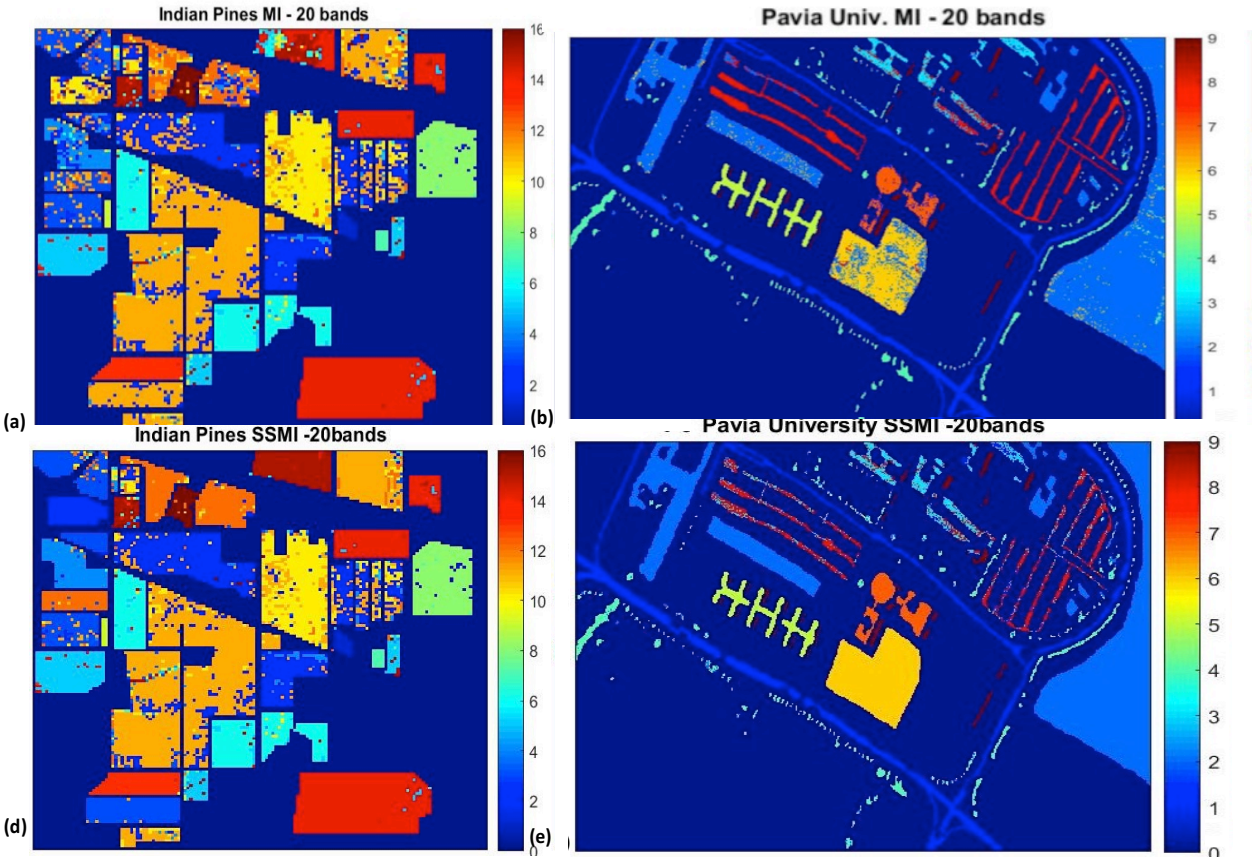
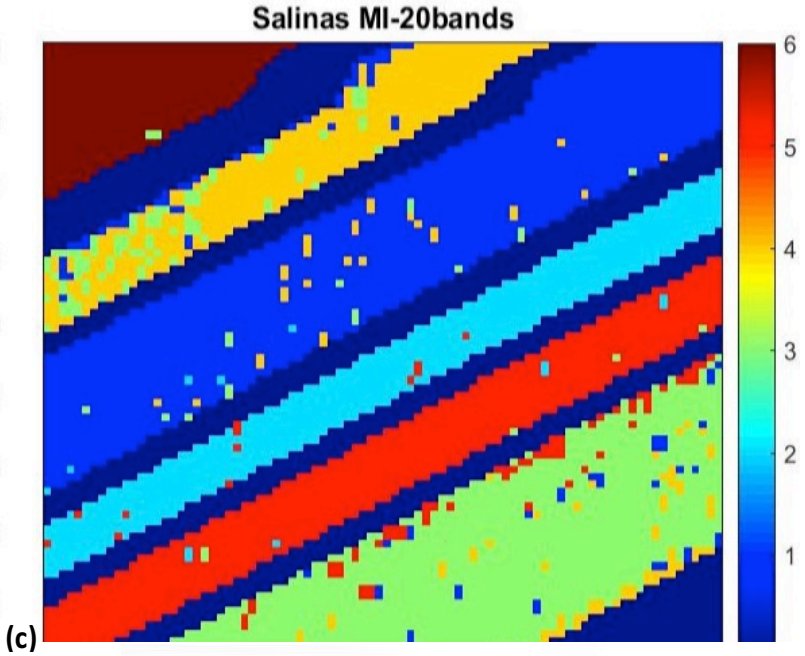
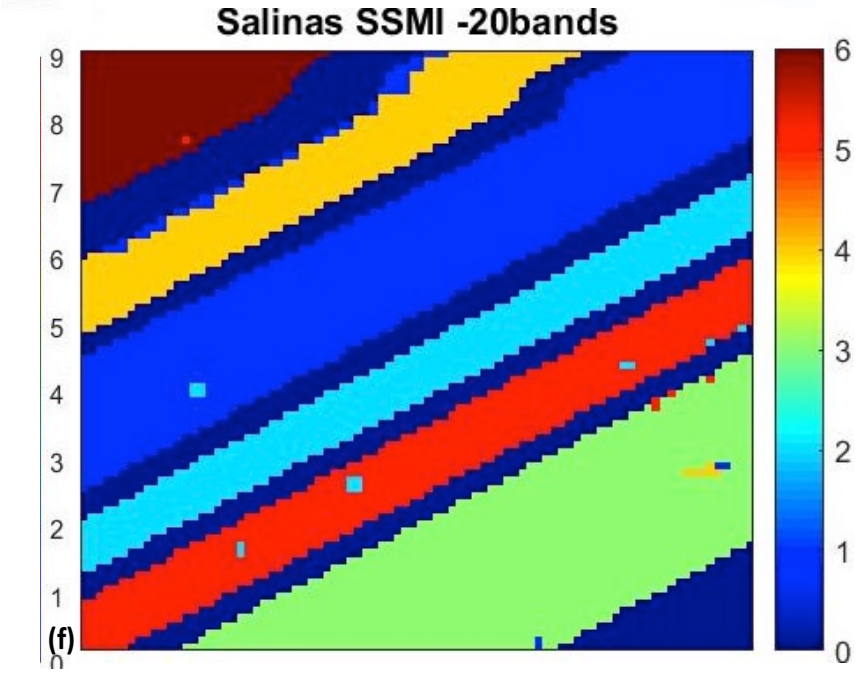


Figure 5-10: false colour classification maps of 3 datasets: (a) and (d): Indian Pines, (b) and (e): Pavia University, (c) and (f): Salinas; obtained by SVM classification (10% training) through the MI BS scheme (Upper panel) and the proposed SSMI method (Lower panel) which exhibits substantial reduction of false alarms.



(c)



(f)

Figure 5-10: continuation.

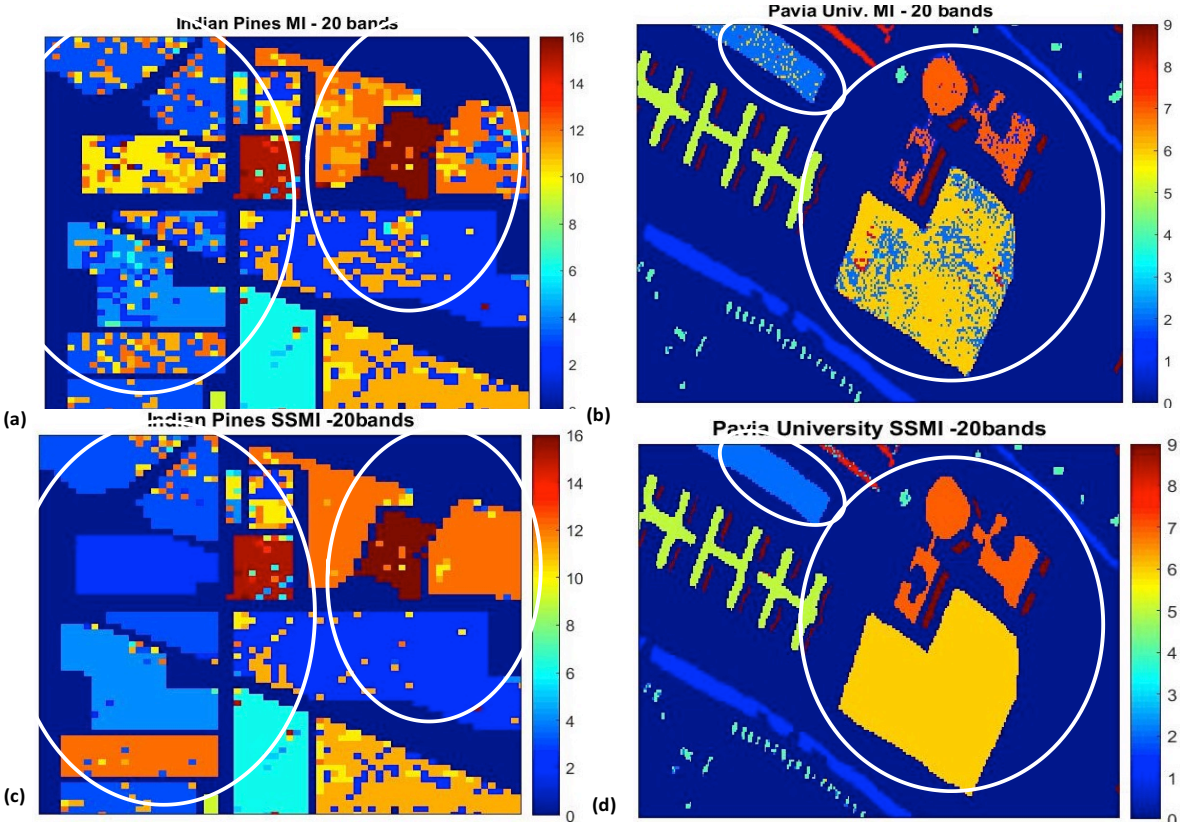


Figure 5-11: zoom-ins of the false colour classification maps of Figure 5-10 to highlight the substantial reduction of classification false alarms in the classes (circled) when the bands are selected through the SSMI BS method. Left column: Portion of Indian Pines, Right column: portion of Pavia University dataset.

Once having seen the good results of SSMI with respect to the competing BS algorithms, we are going to change the conditions to see if the “bell”-shaped curve still appears; we are going to implement SSMI, for SVM and KNN, using 10% and 3% of training pixels:

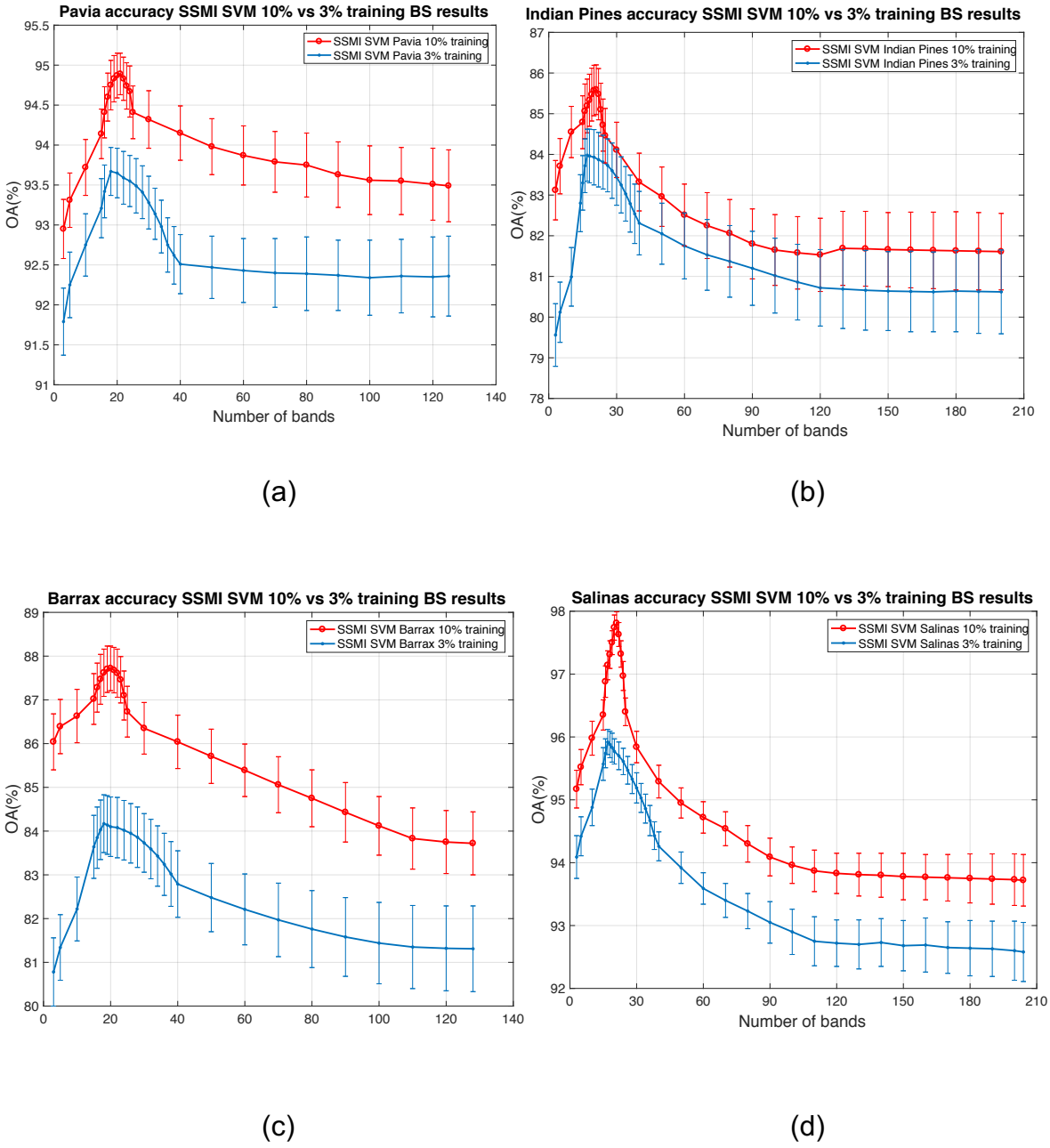
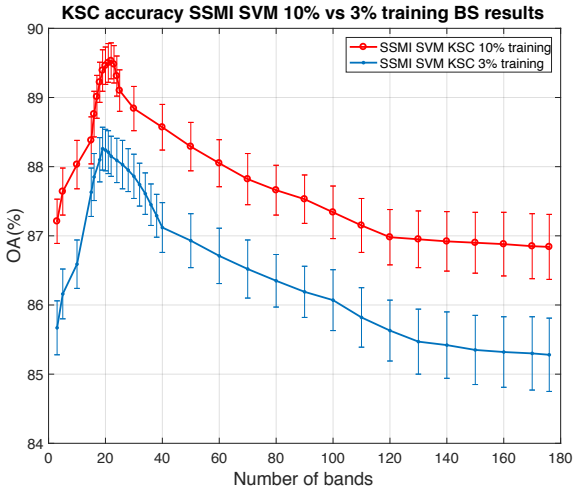
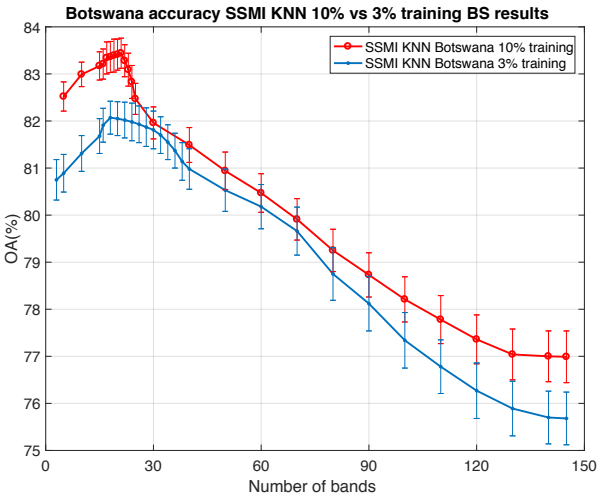


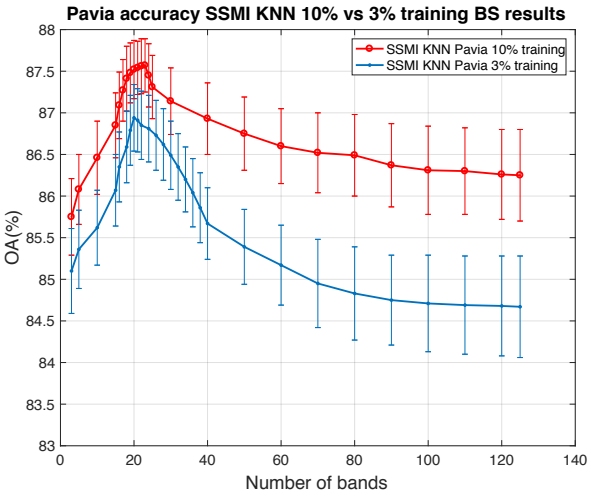
Figure 5-12: SVM classification accuracy results of the proposed SSMI and the other band selection algorithms used on this section, for 10% and 3% of training pixels, for the datasets of a) Pavia University, b) Indian Pines, c) Barrax; d) Salinas, e) KSC and f) Botswana.

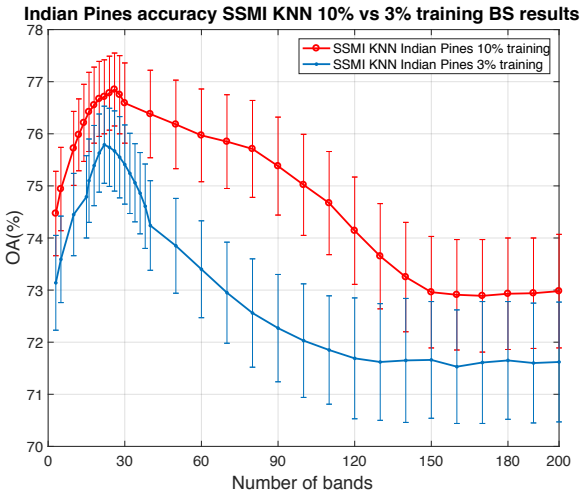




(e) (f)

Figure 5-12: continuation.





(a) (b)

Figure 5-13: KNN classification accuracy results of the proposed SSMI and the other band selection algorithms used on this section, for 10% and 3% of training pixels, for the datasets of a) Pavia University, b) Indian Pines, c) Barrax; d) Salinas, e) KSC and f) Botswana.

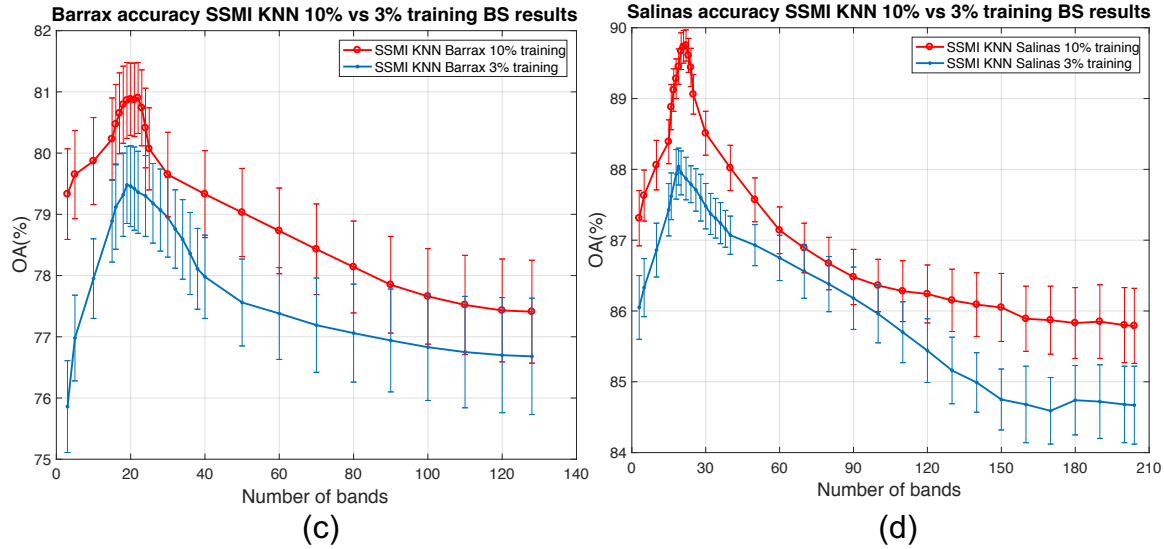


Figure 5-13: continuation.

Looking at the results obtained on Figures 5-12 and 5-13, we observe that the "bell" shape is maintained for SSMI, when SVM and KNN are used, and when 10% and 3% of training pixels are taken for each class. We also appreciate that the maximum accuracy values are obtained for a lower number of bands when a smaller number of training data is used; the 'bell' shape characteristics are sharper when the number of training data decreases, as predicted by Hughes, for all the 6 datasets used on this thesis.

6 Conclusions and Future Work

6.1 Conclusions

The main objective of this research is to establish a new band selection method that improves the accuracy of the previous band selection algorithms and complies with Hughes phenomenon. In order to implement this method, we have combined the spatial information given by the edge filter, taking into account the property that the most discriminative bands are the bands with the best defined edges between the different surfaces, with the spectral information obtained by the use of a Mutual Information algorithm in order to select the bands with the highest quantity of information, once having eliminated the low discrimination bands. This algorithm has been named as SSMI (Spatial Spectral Mutual Information); we have compared its results with respect to 7 different state-of-the-art recent band selection algorithms (Saliency, HyperBS, SLN, OCF, E-FDPC, ISSC and CNN), obtaining an improvement of 10% as average for SSMI with respect to the rest of the algorithms analysed at this thesis, for both SVM and KNN classifiers.

We have also compared three different types of atmospheric correction algorithms; our band selection and classification methods use the reflectance data, but the raw data for each dataset is in Digital Numbers (DN) or radiance. For this reason, a transformation process need to be done, and during this process the atmospheric correction methods are implemented.

We have observed that ATCOR and QUAC obtain much better results than ELM, due to the existence of an additional water absorption correction tool for both methods; ELM is a simpler method and it does not have this characteristic. In addition, ELM needs a bright pixel (ideally with reflectance values as close as possible to 1) and a dark pixel (ideally with reflectance values as close as possible to 0), and for most datasets it is quite difficult to find those types of pixels, increasing the error of ELM.

6.1.1 Conclusions summary for band selection

The main focus of this work has been to examine if Hughes phenomenon is valid and to see whether many spectral bands are needed for achieving optimal classification accuracy; to do this, we have developed a combination of spatial and spectral features for band selection.

Through experiments we have validated Hughes “curse of dimensionality” by experiments; we have observed a ‘bell’ shape of the classification accuracy vs dimensionality characteristics at band selection for the first time, for all the publicly available HSI datasets used on this thesis. We have achieved these results for two different well-known classification algorithms: SVM and KNN.

The ‘bell’ shape characteristics are seen to be sharper when the number of training data decreases, as predicted by Hughes; we observe these accuracy peak differences for our spatial spectral BS algorithm, for both SVM and KNN classifiers. In addition, our results have shown that the classification accuracy can be enhanced by >10% by using ~20 narrow (~10nm FWHM) bands with respect to the results obtained when using all spectral bands for classification.

6.1.2 Conclusions summary for atmospheric correction

In order to correctly assess the band selection methods, it is important to perform an adequate atmospheric correction, to increase the efficacy of the BS algorithms. Most band selection techniques are already optimised to use reflectance values instead of radiance or raw digital; for this reason, it is important to assess the integrity of the reflectance conversion.

To do this, we have looked at the effect of different atmospheric corrections techniques, concluding that atmospheric correction has typically an error of 10-20% error with respect to the true spectral characteristics of the scene. We appreciate that the errors obtained by ELM are much higher than ATCOR and QUAC error values.

6.2 Future Work

During the course of this work it is found that the future research related to the objectives of this thesis can be pursued in the following directions:

1. Although the SSMI technique has shown remarkably good results for shadow compensation, more future work is needed. One of the options that we have is to modify the edge algorithm; instead of using the correlation to choose the bands with the highest discrimination, we could use the variance instead. We can also add additional filters to smooth the edge information for each band, in order to reduce the local effects determined by the position of the surfaces for each dataset.

Respect to MI, we can improve its performance by performing a second round of the algorithm, but only taking into account the set of selected bands by MI on the first application of this algorithm.

2. The error observed for our atmospheric correction methods goes around 10-15%; it may be important to try to find better atmospheric correction methods to try to improve this error. This is important because our band selection methods use the reflectance of each dataset on their algorithm implementation.

One of the main options that we have is to study the combination of atmospheric correction algorithms; we have thought to take the combination of ATCOR and ELM, and QUAC and ELM, in order to look at the results of hybrid atmospheric correction algorithms, and to compare with the results obtained previously by each method alone.

Combining both atmospheric correction effects and band selection, atmospheric correction affects band selection due to the errors provoked by those processes when converting the digital numbers / radiance results into reflectance.

The datasets used on chapter 5 (band selection) have been obtained from public sources, as explained on chapter 3. The atmospheric correction has already been performed by the publishers of these public datasets; for this reason, the atmospheric correction is not taken into account at this chapter 5. It is a good future work idea to try to study the effects of these atmospheric correction errors from raw datasets into the final classification results.

REFERENCES

[1] Hughes, G.F., “On the mean accuracy of statistical pattern recognizers”, *IEEE Transactions on Information Theory*, 14, 55–63, 1968.

[2] Estes, J., Kline, K. and Collins, E., "Remote sensing", *International Encyclopedia of the Social & Behavioral Sciences,* Pergamon, 13144-13150, 2001.

[3] Horning, N., “Remote Sensing”, *Encyclopedia of Ecology*, Academic Press, 2986-2994, 2008.

[4] Henderson, F. M. and Lewis, A. J., “Principles and applications of imaging radar. Manual of remote sensing: Third edition, Volume 2”, John Wiley & Sons, 1998.

[5] Sarhrouni, E., Hammouch, A. and Driss A., "Band Selection and Classification of Hyperspectral Images using Mutual Information: An algorithm based on minimizing the error probability using the inequality of Fano", 2012 International Conference on Multimedia Computing and Systems (ICMCS), IEEE, 2012.

[6] Smith, R. B., “Introduction to Hyperspectral Imaging”, *MicroImages, Inc.*, Lincoln, Nebraska, 1(24), 2006.

[7] Shaw, G. A., and Hsiao-hua, K. B., "Spectral imaging for remote sensing", *Lincoln Laboratory Journal*, 14(1): 3-28, 2003.

[8] Willebrand, H. and Baksheesh, S. G., “Free space optics: enabling optical connectivity in today's networks”, *SAMS Publishing*, 2002.

[9] Weng, Q. and Fu, P., “Modeling annual parameters of clear-sky land surface temperature variations and evaluating the impact of cloud cover using time series of Landsat TIR data”, *Remote Sensing of Environment*, 140, 267-278, 2014.

[10] Keshava, N., "A survey of spectral unmixing algorithms", *Lincoln Laboratory Journal*, 14(1): 55-78, 2003.

[11] Keshava, N. and Mustard, J. F., “Spectral unmixing”, *IEEE Signal Processing magazine*, 19(1), 44-57, 2002.

[12] Ok, A. O., Akar, O. and Gungor, O., “Evaluation of random forest method for agricultural crop classification”, *European Journal of Remote Sensing*, 45(1), 421-432, 2012.

[13] Duro, D. C., Franklin, S. E. and Dubé, M. G., “A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery”, *Remote Sensing of Environment*, 118, 259-272, 2012.

[14] Huang, Y. and Fipps, G., “Landsat satellite multi-spectral image classification of land cover change for GIS-based urbanization analysis in irrigation districts: evaluation in the Lower Rio Grande Valley”, *Texas Water Resources Institute*, College Station, USA, 2006.

[15] Stehman, S. V. and Milliken, J. A., “Estimating the effect of crop classification error on evapotranspiration derived from remote sensing in the lower Colorado River basin, USA”, *Remote Sensing of Environment*, 106(2), 217-227, 2007.

[16] Sidike, P., Khan, J., Alam, M. and Bhuiyan, S., “Spectral unmixing of hyperspectral data for oil spill detection”, Optics and Photonics for Information Processing VI. International Society for Optics and Photonics, 8498, 84981B, 2012.

[17] Zortea, M. and Plaza, A., “Spatial preprocessing for endmember extraction”, *IEEE Transactions on Geoscience and Remote Sensing*, 47(8), 2679-2693, 2009

[18] Gruninger, J. H., Ratkowski, A. J. and Hoke, M. L., "The sequential maximum angle convex cone (SMACC) endmember model", *Defense and Security. International Society for Optics and Photonics*, 2004.

[19] Ifarraguerri, A. and Chang, C-I, "Multispectral and hyperspectral image analysis with convex cones", *IEEE Transactions on Geoscience and Remote Sensing*, 37(2): 756-770, 1999.

[20] Winter, M. E., "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data", *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics*, 1999.

[21] Kruse, F. A., “Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the northern grapevine mountains, Nevada, and California”, *Remote Sensing of Environment*, 24(1), 31-51, 1988.

[22] Racek, F., Balaz, T. and Melsa, P.,”Hyper-spectral data conversion in the case of military surveillance”, Advances in Military Technology, 10(1), 5-13, 2015.

[23] Zhang, X., Zhang, B., Geng, X., Tong, Q. and Zheng, L., “Automatic flat field algorithm for hyperspectral image calibration”, Third International Symposium on Multispectral Image Processing and Pattern Recognition, 5286, 636-640, International Society for Optics and Photonics, 2003.

[24] Smith, G. M., and Milton, E. J., "The use of the empirical line method to calibrate remotely sensed data to reflectance", *International Journal of Remote Sensing*, 20(13): 2653-2662, 1999.

[25] Validation of Land European Remote sensing Instruments. Available online: http://w3.avignon.inra.fr/valeri.

[26] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[27] Boukhechba, K., Huayi Wu, W. and Razika, B., “DCT-based preprocessing approach for ICA in hyperspectral data analysis”, *Sensors*, 18(4), 1138, 2018.

[28] Kuang, F., W. Xu, W. and Zhang, S., “A novel hybrid KPCA and SVM with GA model for intrusion detection”, *Applied Soft Computing*, 18, 178-184, 2014.

[29] Aburomman, A. A. and Reaz, M. B. I., “A novel SVM - kNN - PSO ensemble method for intrusion detection system”, *Applied Soft Computing*, 38, 360-372, 2016.

[30] Hamamoto, Y., Uchimura, S. and Tomita, S., "A Bootstrap Technique for Nearest Neighbor Classifier Design", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1), 73-79, 1997.

[31] Weinberger, K. and Saul, L. K., "Distance Metric Learning for Large Margin Nearest Neighbor Classification", *Journal of Machine Learning Research*, 10, 207-244, 2009.

[32] Bhatia, N. and Vandana, A., "Survey of Nearest Neighbor Techniques," (IJCSIS) *International Journal of Computer Science and Information Security*, 8(2), 302-305, 2010.

[33] Hassant, A. B. A., "Visual Speech Recognition”, *Speech and Language Technologies*, 1, 279-303, 2011.

[34] Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K., "KNN Model-Based Approach in Classification", *Lecture Notes in Computer Science*, 2888, 986-996, 2003.

[35] Hassanat, A. B., Abbadi, M. A., Altarawneh, G. A., & Alhasanat, A. A., “Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach”, *arXiv preprint arXiv:1409.0919*.

[36] Yan, G., Mas, J. F., Maathuis, B. H. P., Xiangmin, Z. and Van Dijk, P. M., “Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area”, *International Journal of Remote Sensing*, 27(18), 4039-4055, 2006.

[37] Hallgren, K. A., “Computing inter-rater reliability for observational data: an overview and tutorial”, *Tutorials in quantitative methods for psychology*, 8(1), 23, 2012.

[38] Gao, B.C., Montes, M-J., Davis, C.O. and Goetz, A.F., “Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean”, *Remote Sensing of Environment*, 113, S17–S24, 2009.

[39] Gao, B. C. and Goetz, A. F., ”Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data”, *Journal of Geophysical Research: Atmospheres*, 95(D4), 3549-3564, 1990.

[40] Richter, R. and Schläpfer, D., “Atmospheric/topographic correction for satellite imagery”, *DLR report DLR-IB*, 565-01, 2005.

[41] Matthew, M. W., Adler-Golden, S. M., Berk, A., Felde, G., Anderson, G. P., Gorodetzky, D. and Shippert, M., "Atmospheric correction of spectral imagery: evaluation of the FLAASH algorithm with AVIRIS data", Applied Imagery Pattern Recognition Workshop, 2002. Proceedings, 157-163, 2002.

[42] Gao, B.C., Davis, C. and Goetz, A., "A review of atmospheric correction techniques for hyperspectral remote sensing of land surfaces and ocean color", 2006 IGARSS Geoscience and Remote Sensing Symposium, 1979-1981, 2006.

[43] Clark, R. N., Swayze, G. A., Heidebrecht, K., Green, R. O., and Goetz, F. H., “Calibration to surface reflectance of terrestrial imaging spectrometry data: Comparison of methods”, 1995.

[44] Ben-Dor, E., and Kruse, F.A., "The relationship between the size of spatial subsets of GER 63 channel scanner data and the quality of the Internal Average Relative Reflectance (IARR) atmospheric correction technique", *Remote Sensing*, 15(3), 683-690, 1994.

[45] Richter, R., Bachmann, M., Dorigo, W. and Mueller, A., ”Influence of the adjacency effect on ground reflectance measurements”, *IEEE Geoscience Remote Sensing Letters*, 3, 565-569, 2006.

[46] Bernstein, L. S., Jin, X., Gregor, B. and Adler-Golden, S., "Quick Atmospheric Correction Code: Algorithm Description and Recent Upgrades", *Optical Engineering*, 51 (11), 1:11, 2012.

[47] Tanre, D. Deroo, C., Duhaut, P., Herman, M., Morcrette, J. J., Perbos, J. and Deschamps, P. Y., “Simulation of the Satellite Signal in the Solar Spectrum (5S), User's Guide”, Laboratoire d'Optique Atmospherique, Universite des Sciences et Techniques de Lille, 1986.

[48] Mousivand, A., Verhoef, W., Menenti, M. and Gorte, B., "Modeling top of atmosphere radiance over heterogeneous non-Lambertian rugged terrain", *Remote Sensing,* 7(6), 8019-8044, 2015.

[49] Ientilucci, E. J., "Using MODTRAN Predicting Sensor-Reaching Radiance", Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 2007.

[50] Prieto-Amparan, J. A., Villarreal-Guerrero, F., Martinez-Salvador, M., Manjarrez-Dominguez, C., Santellano-Estrada, E. and Pinedo-Alvarez, A., “Atmospheric and Radiometric Correction Algorithms for the Multitemporal Assessment of Grasslands Productivity”, *Remote Sensing*, 10(2): 219, 2018.

[51] Tuominen, J., and T. Lipping. "Atmospheric correction of hyperspectral data using combined empirical and model based method", Proceedings of the 7th European association of remote sensing laboratories SIG-imaging spectroscopy workshop, 1113, 2011.

[52] Bouzgou, H. "Development of multiple regression systema for hyperdimensional spectral spaces”, PhD dissertation, Batna, Université El Hadj Lakhder, Faculté des sciences de l'ingénieur, 2006.

[53] Sun, W. and Du, Q., “Hyperspectral band selection: A review”, *IEEE Geoscience and Remote Sensing Magazine*, 7(2), 118-139, 2019.

[54] Goetz, A. F., “Three decades of hyperspectral remote sensing of the Earth: A personal view”, *Remote Sensing of Environment*, 113, S5-S16, 2009.

[55] Tong, Q., Xue, Y. and Zhang, L., “Progress in hyperspectral remote sensing science and technology in China over the past three decades”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1), 70-91, 2013.

[56] He, K. S., Rocchini, D., Neteler, M. and Nagendra, H., “Benefits of hyperspectral remote sensing for tracking plant invasions”, *Diversity and Distributions*, 17(3), 381-392, 2011.

[57] Van der Meer, F. D., Van der Werff, H. M., Van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., Van der Meijde, M., E. Carranza, J. M., Boudewijn de Smeth, J. and Woldai, T., “Multi- and hyperspectral geologic remote sensing: A review”, *International Journal of Applied Earth Observation and Geoinformation*, 14(1), 112-128, 2012.

[58] Thenkabail, P. S. and Lyon, J. G., “Hyperspectral remote sensing of vegetation”, CRC press, 2016.

[59] Camps-Valls, G., Tuia, D., Bruzzone, L. and Benediktsson, J. A., “Advances in hyperspectral image classification: Earth monitoring with statistical learning methods”, *IEEE Signal Processing Magazine*, 31(1), 45-54, 2013.

[60] Nasrabadi, N. M., “Hyperspectral target detection: An overview of current and future challenges”, *IEEE Signal Processing Magazine*, 31(1), 34-44, 2013.

[61] Manolakis, D. and Shaw, G., “Detection algorithms for hyperspectral imaging applications”, *IEEE Signal Processing Magazine*, 19(1), 29-43, 2002.

[62] Hook, S. J., “NASA 2014 The hyperspectral infrared imager (HyspIRI)–science impact of deploying instruments on separate platforms”, Jet Propulsion Laboratory, Pasadena, California, National Aeronautics and Space Administration, *White Paper*, 2014.

[63] Varshney, P. K. and Arora, M. K., “Advanced image processing techniques for remotely sensed hyperspectral data”, *Springer Science & Business Media*, 2004.

[64] Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J. and Tilton, J. C., “Advances in spectral-spatial classification of hyperspectral images”, *Proceedings of the IEEE*, 101(3), 652-675, 2012.

[65] Landgrebe, D., “Hyperspectral image data analysis”, *IEEE Signal Processing Magazine*, 19(1), 17-28, 2002.

[66] Marshall, M. and Thenkabail, P., “Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 205-218, 2015.

[67] Bajcsy, P. and Groves, P., “Methodology for hyperspectral band selection”, *Photogrammetric Engineering & Remote Sensing*, 70(7), 793-802, 2004.

[68] Harsanyi, J. C. and Chang, C. I., “Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach”, *IEEE Transactions on Geoscience and Remote Sensing*, 32(4), 779-785, 1994.

[69] Lunga, D., Prasad, S., Crawford, M. M. and Ersoy, O., “Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning”, *IEEE Signal Processing Magazine*, 31(1), 55-66, 2013.

[70] Miao, X., Gong, P., Swope, S., Pu, R., Carruthers, R. and Anderson, G. L., “Detection of yellow starthistle through band selection and feature extraction from hyperspectral imagery”, *Photogrammetric Engineering and Remote Sensing*, 73(9), 1005, 2007.

[71] Sun, W., Halevy, A., Benedetto, J. J., Czaja, W., Liu, C., Wu, H., Shi, B. and Li, W., “UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 89, 25-36, 2014.

[72] Dópido, I., Villa, A., Plaza, A. and Gamba, P., “A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 421-435, 2012.

[73] Martinez-Uso, A., Pla, F., Martinez-Sotoca, J. and Garcia-Sevilla, P., “Clustering-based hyperspectral band selection using information measures”, *IEEE Transactions on Geoscience and Remote Sensing*, 45(12), 4158-4171, 2007.

[74] Yang, H., Du, Q. and Chen, G., “Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensin*g, 5(2), 544-554, 2012.

[75] Jia, S., Ji, Z., Qian, Y., and Shen, L., “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, 2012.

[76] Richards, J. A., “Remote Sensing Digital Image Analysis”, 3, Berlin: Springer, 1999.

[77] Conese, C. and Maselli, F., “Selection of optimum bands from TM scenes through mutual information analysis”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(3), 2–11, 1993.

[78] Stearns, S. D., B. E. Wilson, B . E. and Peterson, J. R., “Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery,” *Applications of Digital Image Processing XVI*, 2028, 118–127, 1993.

[79] Kumar, S., Ghosh, J. and Crawford, M. M., “Best-bases feature extraction algorithms for classification of hyperspectral data”, *IEEE Transactions on Geoscience and Remote Sensing*, 39(7), 1368–1379, 2001.

[80] Jimenez-Rodriguez, L. O., Arzuaga-Cruz, E. and Velez-Reyes, M., “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, 2007.

[81] Jones, M. and Sibson, R., “What is projection pursuit?”, *Journal of the Royal Statistical Society: Series A (General)*, 150(1), 1–36, 1987.

[82] Ifarraguerri, A. and Chang, C.-I., “Unsupervised hyperspectral image analysis with projection pursuit”, *IEEE Transactions on Geoscience and Remote Sensing*, 38(6), pp. 2529–2538, 2000.

[83] Agarwal, A., El-Ghazawi, T., El-Askary, H. and Le-Moigne, J., “Efficient hierarchical-PCA dimension reduction for hyperspectral imagery”, *2007 IEEE International Symposium on Signal Processing and Information Technology*, 353-356, 2007

[84] Fauvel, M., Chanussot, J., and Benediktsson, J. A., “Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas”, *EURASIP Journal on Advances in Signal Processing, 2009*, 1-14, 2009.

[85] Wang, J. and Chang, C.-I., “Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis”, *IEEE Transactions on Geoscience and Remote Sensing*, 6, 1586–1600, 2006.

[86] Guyon, I. and Elisseeff, A., “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, 3, 1157–1182, 2003.

[87] Chang, C.-I. and Wang, S., “Constrained band selection for hyperspectral imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, 44(6), 1575–1585, 2006.

[88] Rashwan, S. and Dobigeon, N., “A split-and-merge approach for hyperspectral band selection”, *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1378-1382, 2017.

[89] Su, P., Liu, D., Li, X. and Liu, Z., “A saliency-based band selection approach for hyperspectral imagery inspired by scale selection”, *IEEE Geoscience and Remote Sensing Letters*, 15(4), 572-576, 2018.

[90] Habermann, M., Fremont, V. and Elcio Shiguemori, E. "Supervised band selection in hyperspectral images using single-layer neural networks", *International Journal of Remote Sensing*, 40 (10), 3900-3926, 2019.

[91] Wang, Q., Zhang, F. and Li, X., “Optimal clustering framework for hyperspectral band selection”, *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 5910-5922, 2018.

[92] Jia, S., Tang, G., Zhu, J. and Li, Q., "A novel ranking-based clustering approach for hyperspectral band selection", *IEEE Transactions on Geoscience and Remote Sensing*, 54(1), 88-102, 2015.

[93] Sun, W., Zhang, L., Du, B., Li, W. and Lai, Y. M., “Band selection using improved sparse subspace clustering for hyperspectral imagery classification”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6), 2784-2797, 2015.

[94] Sharma, V., Diba, A., Tuytelaars, T. and Van Gool, L., “Hyperspectral CNN for image classification & band selection, with application to face recognition”, Technical report KUL/ESAT/PSI/1604, KU Leuven, ESAT, Leuven, Belgium, 2016.

[95] Cao, X., Wu, B., Tao, D. and Jiao, L., “Automatic band selection using spatial-structure information and classifier-based clustering”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9), 4352-4360, 2016.

[96] Ali, M. and Clausi, D., “Using the Canny edge detector for feature extraction and enhancement of remote sensing images”, IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), 5, 2298-2300, 2001.

[97] Vincent, O. R. and Folorunso, O., “A descriptive algorithm for Sobel image edge detection”, Proceedings of Informing Science & IT Education Conference (InSITE), California: Informing Science Institute, 40, 97-107, 2009.

[98] Shrivakshan, G. T. and Chandrasekar, C. “A comparison of various edge detection techniques used in image processing”, *International Journal of Computer Science Issues (IJCSI)*, 9(5), 269, 2012.

[99] Feng, J., Jiao, L., Zhang, X. and Sun, T., “Hyperspectral band selection based on trivariate mutual information and clonal selection”, *IEEE Transactions on Geoscience and Remote Sensing*, 52, 4092–4105, 2014.

[100] Guo, B., Gunn, S., Damper, B. and Nelson, J., “Adaptive band selection for hyperspectral image fusion using mutual information”, 2005 7th International Conference on Information Fusion, Philadelphia, PA, IEEE, 1, 25-28 July 2005.

[101] Gunduz, A. and Principe, J. C., “Correntropy as a novel measure for nonlinearity tests”, *Signal Processing*, 89(1), 14-23, 2009.

[102] Liu, Y., Chen, Y., Tan, K., Xie, H., Wang, L., Yan, X., Xie, W. and Xu, Z., “Maximum relevance, minimum redundancy band selection based on neighborhood rough set for hyperspectral data classification”, *Measurement Science and Technology*, 27(12), 125501, 2016.

[103] Tschannerl, J., Ren, J., Yuen, P., Sun, G., Zhao, H., Yang, Z., Wang, Z. and Marshall, S., “MIMR-DGSA: unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm”, *Information Fusion*, 51, 189-200, 2019.

[104] Feng, J., Jiao, L., Liu, F., Sun, T. and Zhang, X., “Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy”, *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2956-2969, 2014.

[105] Jin, J. and Wang, Q., “Selection of informative spectral bands for PLS models to estimate foliar chlorophyll content using hyperspectral reflectance”, *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(5), 3064-3072, 2018.

[106] Amankwah, A., “Spatial mutual information based hyperspectral band selection for classification”, *The Scientific World Journal*, 2015.

[107] Guo, B., Gunn, S.R., Damper, R. and Nelson, J., “Band selection for hyperspectral image classification using mutual information”, *IEEE Geoscience and Remote Sensing Letters*, 3, 522–526, 2006.

[108] Ren, J., Kelman, T. and Marshall, S., “Adaptive clustering of spectral components for band selection in hyperspectral imagery”, Hyperspectral Imaging Conference 2011, Glasgow, United Kingdom, 2011.

[109] Lin, L. and Zhang, S., “Superpixel Segmentation of Hyperspectral Images Based on Entropy and Mutual Information”, *Applied Sciences*, 10(4), 1261, 2020.

[110] Varade, D., Maurya, A. K. and Dikshit, O., “Unsupervised band selection of hyperspectral data based on mutual information derived from weighted cluster entropy for snow classification”, *Geocarto International*, 1-23, 2019.

[111] Gao, F., Wang, Q., Dong, J. and Xu, Q., “Spectral and spatial classification of hyperspectral images based on random multi-graphs”, *Remote Sensing*, 10(8), 1271, 2018.

[112] Adam, E., Mutanga, O. and Rugege, D., “Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review”, *Wetlands Ecology and Management*, 18(3), 281-296, 2010.

[113] Li, M., Zang, S., Zhang, B., Li, S. and Wu, C., “A review of remote sensing image classification techniques: The role of spatio-contextual information”, *European Journal of Remote Sensing*, 47(1), 389-411, 2014.

[114] Poojary, N., D'Souza, H., Puttaswamy, M. R. and Kumar, G. H., “Automatic target detection in hyperspectral image processing: A review of algorithms”, 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, IEEE, 1991-1996, 2015.

[115] Manolakis, D., Marden, D. and Shaw, G. A., “Hyperspectral image processing for automatic target detection applications”, *Lincoln Laboratory Journal*, 14(1), 79-116, 2003.

[116] Hennessy, A., Clarke, K and Lewis, M., “Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability”, *Remote Sensing*, 12(1), 113.2020, 2020.

[117] Cheng, J. H., Nicolai, B. and Sun, D. W., “Hyperspectral imaging with multivariate analysis for technological parameters prediction and classification of muscle foods: A review”, *Meat Science*, 123, 182-191, 2017.

[118] Lu, G. and Fei, B., “Medical hyperspectral imaging: a review”, *Journal of Biomedical Optics*, 19(1), 010901, 2014.

[119] Yuen, P. W. and Richardson, M., “An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition”, *The Imaging Science Journal*, 58(5), 241-253, 2010.

[120] Habermann, M., “Band selection in hyperspectral images using artificial neural networks”, Doctoral dissertation, Université de Technologie de Compiègne, Compiègne, France, 2018.