THE UNIVERSITY OF
SYDNEY

A probabilistic framework for classification and fusion of remotely sensed hyperspectral data

Sven Schneider

A thesis submitted in fulfillment
of the requirements of the degree of
Doctor of Philosophy

Australian Centre for Field Robotics
School of Aerospace, Mechanical and Mechatronic Engineering
The University of Sydney

31 March 2013

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the University or other institute of higher learning, except where due acknowledgement has been made in the text.

Sven Schneider

31 March 2013

Abstract

Sven Schneider
The University of Sydney

Doctor of Philosophy
31 March 2013

A probabilistic framework for classification and fusion of remotely sensed hyperspectral data

Reliable and accurate material identification is a crucial component underlying higher-level autonomous tasks within the context of autonomous mining. Such tasks can include exploration, reconnaissance and guidance of machines (e.g. autonomous diggers and haul trucks) to mine sites. This thesis focuses on the problem of classification of materials (rocks and minerals) using high spatial and high spectral resolution (hyperspectral) imagery, collected remotely from mine faces in operational open pit mines. A new method is developed for the classification of hyperspectral data including field spectra and imagery using a probabilistic framework and Gaussian Process regression. The developed method uses, for the first time, the Observation Angle Dependent (OAD) covariance function to classify high-dimensional sets of data. The performance of the proposed method of classification is assessed and compared to standard methods used for the classification of hyperspectral data. This is done using a staged experimental framework. First, the proposed method is tested using

high-resolution field spectrometer data acquired in the laboratory and in the field. Second, the method is extended to work on hyperspectral imagery acquired in the laboratory and its performance evaluated. Finally, the method is evaluated for imagery acquired from a mine face under natural illumination and the use of independent spectral libraries to classify imagery is explored. A probabilistic framework was selected because it best enables the integration of internal and external information from a variety of sensors. To demonstrate advantages of the proposed GP-OAD method over existing, deterministic methods, a new framework is proposed to fuse hyperspectral images using the classified probabilistic outputs from several different images acquired of the same mine face. This method maximises the amount of information but reduces the amount of data by condensing all available information into a single map. Thus, the proposed fusion framework removes the need to manually select a single classification among many individual classifications of a mine face as the 'best' one and increases the classification performance by combining more information. The methods proposed in this thesis are steps forward towards an automated mine face inspection system that can be used within the existing autonomous mining framework to improve productivity and efficiency. Last but not least the proposed methods will also contribute to increased mine safety.

Acknowledgements

This thesis would not have been possible with the help of my supervisor Richard Murphy. I would like to thank you for supporting me throughout the years, being patient, and for the invaluable advice and wisdom that you have shared with me. Thank you also for having an open-ear and for your support with all my troubles. It is ALL greatly appreciated!

I would also like to thank Arman Melkumyan, my associate supervisor, for helping me with technical things and being patient in the beginning when I asked many questions.

Thank you to both of you for encouraging me throughout the years!

Thank you also to Sildomar Monteiro who pushed me very early to start writing papers and gave me guidance in the very early stages.

Thank you to Ruth and Lisa for shielding me from all the bureaucracy and paperwork.

Thank you to the guys at the workshop, especially Javier and Bang for your help with preparations for our field trials.

Thank you to Rio Tinto for their support during field trials and making them possible. I would also like to thank the RTCMA for my financial support and my scholarship.

Thanks Nasir, Calvin, Sisir for helpful and interesting research discussions.

Thanks to Andre, Layth, Calvin, Franzi, Nick, Victor, Tariq, Wesley and Ekin for fun times such as trips, dinner parties, going out or just hanging out at the ACFR.

Special thanks to Layth for being such a good friend, even in difficult times. I won't forget it!

Many thanks to Peter, Corinna and Helena for reminding me to eat, being supportive and distracting me from my thesis and other stuff. Thank you for hanging out and having fun times together.

Thank you to Lia for a wonderful year which I will never forget!

Last but not least, thank you to my family! Thank you to Mum and Dad and my sister Senta for always believing in me and supporting me. Thank you also to my grandparents, Omas und Opa, who supported me and always encouraged me. Thank you also for reminding me of home.

List of Publications

Murphy, R.J., **Schneider, S**., & Monteiro, S.T. (accepted, in press). Consistency of Measurements of Wavelength Position from Hyperspectral Imagery: Use of the Ferric Iron Crystal Field Absorption at ∼900 nm for Estimating Ore Type and Grade. *Geoscience and Remote Sensing, IEEE Transactions on*, 2013

Chlingaryan, A., Melkumyan, A., Murphy, R.J., & **Schneider, S**. (accepted, in press). Automated Multi-Class Classification of Hyperspectral Data using Gaussian Processes. In: *Application of Computers and Operations Research in the Mineral Industry (APCOM)*. 2013, Porto, Alegre, Brazil: The Australasian Institute of Mining and Metallurgy (The AusIMM).

Murphy, R.J., Monteiro, S.T. & **Schneider, S**. (2012). Evaluating Classification Techniques for Mapping Vertical Geology Using Field-Based Hyperspectral Sensors. *Geoscience and Remote Sensing, IEEE Transactions on.* 50(8). p.pp. 3066-3080.

Schneider, S., Murphy, R.J., Melkumyan, A. & Nettleton, E. (2012). A geological perception system for autonomous mining. In: *Proceedings of IEEE International Conference on Robotics and Automation* (ICRA). 2012, St. Paul, Minnesota, USA.

Schneider, S., Melkumyan, A., Murphy, R.J. & Nettleton, E. (2011). Classification of hyperspectral imagery using GPs and the OAD covariance function with automated endmember extraction. In: *Proceedings of IEEE International Conference on Tools with Artificial Intelligence*. 2011, Boca Raton, Florida, USA.

Schneider, S., Murphy, R.J., Melkumyan, A. & Nettleton, E. (2011). Classification of Hyperspectral Data of Mine Face Geology Using Machine Learning for Autonomous Open Pit Mining. In: *Application of Computers and Operations Research in the Mineral Industry (APCOM)*. 2011, Wollongong, Australia: The Australasian Institute of Mining and Metallurgy (The AusIMM).

Schneider, S., Melkumyan, A., Murphy, R.J. & Nettleton, E.W. (2010). Gaussian Processes with OAD covariance function for hyperspectral data classification. In: Proceedings of IEEE International Conference on Tools with Artificial Intelligence. 2010, Arras, France.

Schneider, S., Murphy, R.J., Monteiro, S. & Nettleton, E.W. (2009). On the development of a hyperspectral library for autonomous mining. In: Proceedings of Australasian Conference on Robotics and Automation (ACRA). 2009, Sydney, Australia.

To my family

Contents

List of Figures

List of Tables

List of Acronyms

AIS	airborne imaging spectrometer
ASD	analytical spectral devices
ASTER	advanced spaceborne thermal emission and reflection radiometer
AVIRIS	airborne visible-infrared imaging spectrometer
BIF	banded iron formation
CASI	compact airborne spectrographic imager
CCD	charged coupled device
CCSM	cross correlogram spectral matching
CHT	cherty BIF
CLY	clay
CMOS	complementary metal-oxide-semiconductor
DVI	difference vegetation index
EMR	electromagnetic radiation
ERE	effective resolution element
FOV	field of view
FWHM	full width at half maximum
GOL	goethite limonite
GP	Gaussian Process
HG	higher-grade or high-grade
IFOV	instantaneous field of view
LabImg283	hyperspectral image with 283 bands acquired in the laboratory
LabImg359	see LabImg 283, except with 359 bands
LG	lower-grade or low-grade
LOI	loss on ignition
LWIR	long-wave infrared
MAR	martite-goethite
MED	minimum Euclidean distance
MIR	mid-wave infrared
MSE	mean squared error
MSS	multispectral scanner
NDVI	normalised difference vegetation index
NNET	neural network (refers to the NNET kernel)
NS3+4	shale (volcanic shale)

NVCL	national virtual core library
OAD	observation angle dependent
OvA	one-versus-all
OvR	one-versus-rest
PCA	principle component analysis
PSD	point spread function
PSVM	probabilistic SVM
QXRD	quantitative x-ray diffraction analysis
RBD	relative band-depth
RBF	radial basis function
RGB	red, green, blue (e.g. camera or filters)
RMSE	root mean square error
RVI	ratio vegetation index
SAM	spectral angle mapper
SAM 0.1	angular threshold of 0.1 used for SAM
SAM m.a.	minimum angle, i.e. SAM using a minimum angle criterion
SAR	synthetic aperture radar
SAVI	soil adjusted vegetation index
SE	squared exponential (kernel)
SFF	spectral feature finding
SHL	West Angelas shale
SHN	manganiferous shale
SNR	signal-to-noise ratio
SVM	support vector machine
SWIR	short-wave infrared
TM	Thematic Mapper
TP	Tom Price
USGS	US geological survey
VIS	visible
VI	vegetation indices
VNIR	visible-near infrared
WA	West Angelas
WA-MF1	Mine face image acquired in West Angelas, mine face 1
WA-MF2	Mine face image acquired in West Angelas, mine face 2
WA-MF2b	same as WA-MF2 but with more shade present on mine face 2
XRD	x-ray diffraction
XRF	x-ray fluorescence
XRS	x-ray sample

Chapter 1

Introduction

In Australia, gold, iron, coal and copper among other raw materials are extracted from the Earth by open-pit mining operations. For example, iron-ore deposits in Western Australia extend over areas of 80,000 km^2 (Morris, 1980). Mines in this region cover extensive areas that are often inhospitable or dangerous to humans. Because of this, many tasks in the mining process carry considerable risks to human workers. Due to increasing demands of the market for raw materials (e.g. iron ore and copper), mining companies are forced to increase the productivity and efficiency of the mining process, leading to uninterrupted mining operations (Yellishetty et al., 2010). In order to improve the productivity and efficiency, and to increase mine safety, mining companies have recognised the potential of robotics applications and autonomous systems. In recent years, the automation of individual mining tasks has taken place. For example, the most common automated vehicles and machineries in the context of mine automation are excavators (Sarata et al., 2006), autonomous drills (Caterpillar, 2008) and driverless haul trucks (Agamennoni et al., 2011; Maekawa et al., 2010).

Automation is expected to be increasingly used in the mining industry. An example of this is Rio Tinto's 'mine of the future' project which aims to develop methods and technology to create an end-to-end automation system (e.g. RioTinto, 2007; 2013). Such an end-to-end automation system would enable detection of ore in-ground, extraction and processing as well as shipping of ore using autonomous systems. A large

commitment towards this goal was made by Rio Tinto in creating and funding the Rio Tinto Centre for Mine Automation, thus emphasizing the need for such autonomous mining systems in order to create a more productive and safer mine environment (University of Sydney, 2009). One of the most fundamental aspects of mine automation is the provision of information about the mine environment prior to the extraction and processing of the ore. Specifically, this involves the ability to map rocks and minerals as they are presented on vertical mine faces and to determine the boundaries between discrete geological units.

This thesis provides research towards an automated system for the identification of rocks and minerals by providing a framework for the analysis of data acquired by hyperspectral sensors.

1.1 Motivation

Mining companies require up-to-date models of the exposed geology and of specific minerals and rock types on mine faces. These models are required at several stages in the mining process. For example, relationships between boundaries of ore and waste can change after blasting and a determination of this change is critical in order to make decisions about what is, respectively, ore and waste. Specific information that is needed include: (i) the type of ore; (ii) identification of waste products, i.e. rocks which are not mineralised or which are economically unimportant; (iii) information about materials that hinder the mining process because they are very hard or swell when wet, making them difficult to mine or process; and (iv) the position of boundaries between these different materials or geological units on a mine face.

Presently, these sorts of information are obtained by a geologist who makes a geological survey of a mine face prior to excavation. Mine faces are notoriously unstable and pose a considerable threat to geologists doing field surveys. Due to the sometimes complex geometry of mine faces, there is a constant risk from sliding rocks and falling debris, even where mining operations have been suspended for the duration of the survey. For safety reasons, only the lower part of a mine face may be approached to

collect samples for detailed analysis, thus limiting the area and the resolution over which a survey can be made. In addition, an accumulation of loose material (rill) is often found at the base of a mine face. The composition of rill is not necessarily representative of the ore-body towards the upper parts of a mine face.

The present methods for generating maps of the geology of a mine face by human inspection is also problematic because: (i) the determination of a rock type is subjective; (ii) truly quantitative measures of mineral identity and abundance cannot be provided unless further extensive sampling and laboratory analysis using X-ray diffraction (XRD) and / or X-ray fluorescence (XRF)[1] are made; (iii) the temporal and the spatial resolution, of maps made by human sampling, is irregular and extremely coarse; and (iv) the location of geological boundaries are approximate and are often crudely determined by boundary markers or tape.

In order to resolve these issues and other limitations associated with current methods of mapping geology on vertical mine faces, new methods need to be introduced which aim to improve the quantity and quality of information available. Ultimately, this has the overarching objective of increasing efficiency and productivity through the use of high-resolution information in planning and scheduling. Last, but by no means least, if information can be obtained autonomously from a safe distance without direct physical interaction with the mine face, this has the potential to radically improve safety in operational mining.

1.2 Remote Sensing

Because approaching a mine face and collecting samples from the bottom of it is dangerous, and cannot be representative of the whole mine face, new ways to obtain information about the geology from a distance at a high spatial resolution and over large areas are required.

[1] portable XRF devices are available, however, they only provide information about the chemistry of rocks and lack the spatial resolution and collection speed of imaging spectrometers.

Remote sensing is a way of acquiring information of a scene or a target object without being in physical contact with it, providing an efficient way of collecting information over large, inaccessible or dangerous areas. There are various platforms and sensors that can be used for remote sensing (Schowengerdt, 1997). Sensors have conventionally been used from airborne (e.g. balloons, aircraft) or orbital satellite platforms (Goetz, 2009). Typical sensors used in remote sensing include, amongst others, radar (e.g. synthetic aperture radar; SAR), lidar, and passive spectral imaging sensors, the latter being subject of this thesis. While aircraft or satellite platforms can cover larger areas, there are limitations in the observability of vertical objects due to the nature of their angle of observation of a target - normally from an overhead perspective (i.e. from nadir). As a consequence, it is difficult or impossible to obtain information from vertical surfaces (e.g. a mine face) using these platforms.

In order to map vertical surfaces such as mine faces, field-based remote sensing platforms need to be deployed. Field-based platforms have the advantage of obtaining information reliably from vertical surfaces and have a greatly increased spatial resolution (i.e. ground-pixel size) compared to satellite or airborne-platforms. Field-based platforms have been deployed in recent studies and demonstrated the potential of field-based remote sensing (e.g. Ramanaidou et al., 2002; Kurz et al., 2008; 2011; Ramanaidou and Wells, 2012; Murphy and Monteiro, 2013).

1.3 Hyperspectral imaging for remote sensing

Spectral imaging sensors for remote sensing can be divided into two broad groups, (i) multispectral imaging sensors and (ii) hyperspectral imaging sensors. Because of the limited discriminative power of multispectral sensors (e.g. Rowan et al., 1977), i.e. because they cannot resolve diagnostic absorption features, they are used mainly for separating materials. Hyperspectral sensors, however, have the capability to remotely identify materials based on diagnostic information within the spectral curve (Goetz et al., 1982; Vane and Goetz, 1988; Goetz, 2009).

Both types of sensors measure reflected light from a target object. The main differ-

ence between these types of sensors is the number of bands they discretely detect. Multispectral sensors have only a few (generally less than 10) broad bands, e.g. a three-band RGB camera (with a Red, Green and Blue channel) or the Landsat Thematic Mapper; Rowan, 1975). Hyperspectral imaging sensors have many (several hundreds), narrow, contiguously-spaced bands in which reflected light from a target can be detected (Goetz et al., 1985; Vane and Goetz, 1993; Green et al., 1998). Examples of commonly used hyperspectral sensors include the Airborne Visible Infrared Imaging Spectrometer (AVIRIS), the Compact Airborne Spectrographic Imager (CASI) and Hyperion. These sensors are mounted on airborne or, in the case of Hyperion, spaceborne platforms. The observation distance of such platforms, directly affects the pixel size. For example, for measurements taken 4-20 km above the ground using AVIRIS, the pixel size would be 4-20 m (Jet Propulsion Laboratory, 2007).

Recently, smaller and more compact hyperspectral sensors are commercially available enabling this technology to be deployed on field-based platforms. As a consequence, hyperspectral sensors can be used to obtain spectral information from vertical surfaces such as a mine face in an open pit mine, at a high spatial resolution (i.e. smaller pixel size) and of high spectral resolution. Recent work has demonstrated the use of a field-based single short-wave infrared (SWIR) hyperspectral imaging sensor to model outcrops of dolomite and limestone (Kurz et al., 2008). Only few studies have used separate hyperspectral imaging sensors (i.e. visible-near infrared, VNIR and SWIR sensors) to acquire images of mine faces in open pit mines (e.g. Kruse et al., 2011). The work presented in this thesis is different in that is uses an independent reference library acquired using a non-imaging spectrometer with artificial light[2] to construct geological thematic maps of mine faces acquired using an imaging spectrometer under natural light (similar to Murphy et al., 2012).

[2]similar to the HyLogger system (Huntington et al., 2004)

1.4 Problem statement

Hyperspectral imaging sensors allow identification of minerals based on their diagnostic absorption features. This is because different materials absorb and scatter different wavelengths (i.e. colours) of incoming light in different amounts (Elachi and Van Zyl, 2006). Thus, materials reflect only a fraction of the incoming light, with different intensities, back to an observer or to a sensor. One of the first applications of hyperspectral sensors was the identification of materials (e.g. geology and vegetation) by matching their diagnostic absorption features with known signatures (e.g. Goetz et al., 1985; Vane and Goetz, 1988; Kruse, 1988; Hutsinpiller, 1988). Such absorption feature based methods are effective as diagnostic absorption features can be directly related to the presence of certain materials, e.g. iron minerals or, in the case of vegetation, photosynthetically-active pigments (e.g. Clark et al., 2003; Rodger and Cudahy, 2009; Rodger et al., 2012). There are, however, instances were a spectrum is featureless - that is, the spectrum has no diagnostic absorption features (e.g. Cloutis et al., 1990; 2011; Gaffey et al., 1992). Featureless spectra appear when the analysed material has no diagnostic absorption features in the measured wavelength range or when acquisition or instrument or processing defects prevent their detection (e.g low signal-to-noise ratio). Some methods, which identify spectra based on their diagnostic absorption features, are sometimes not suitable for the analysis of hyperspectral data because they assume the presence of diagnostic absorption features, when in reality there may be none. A solution to this problem, to develop methods which work in the whole spectral curve shape, irrespective of whether diagnostic absorption features are present or not, e.g. the Spectral Angle Mapper (SAM; Kruse et al., 1993). Thus, any unknown spectrum can be objectively matched with any number of spectra in a spectral library where identities (i.e. class labels) are known (e.g. van der Meer, 2004; Gianinetto and Lechi, 2004; Cho et al., 2010).

Absorption feature based methods and methods which work on the whole spectral curve (e.g. SAM) are considered to be deterministic methods within the context of this thesis. “Deterministic” in the context of this thesis means that a method yields a definitive answer to a query, for example, an unknown pixel will be assigned the

class label of the library spectrum with which it has the closest match according to some quantitative measure (e.g. the smallest spectral angle or smallest Root Mean Square Error, RMSE). For many applications this is sufficient or even required (e.g. Murphy et al., 2012). However, deterministic methods do not provide measures of confidence associated with an assigned class label. In the context of autonomous mining, deterministic methods are insufficient. The main reason for this is that outputs from deterministic methods are difficult to fuse (i.e. to incorporate) with other information acquired by other sensors or at different times. For example, if a specific mine face is scanned (i.e. observed) several times during the day by a hyperspectral imaging sensor, for each acquired image, a classified map can be created using SAM. It is, however, unlikely that all pixels in each of the images will be assigned to the same class across all the classified images because of the variability introduced in extraneous effects such as changes in illumination, dust and shadow, causing changes in the spectral distribution of incoming light between measurements. A fundamental question to answer is: "Which of these classifications is the correct one - if any?"

For autonomous mining, methods of classification which are not deterministically-based are required. A method for the classification of remotely sensed hyperspectral data of mine faces needs to fulfill certain key requirements: (i) accurate classification of rocks irrespective of the presence, absence or the quality of diagnostic absorption features, (ii) it must provide a measure of certainty when classifying an unknown spectrum, (iii) have the capability to integrate (i.e. fuse) information from other sensors and (iv) have the capability to be integrated into an existing framework currently used for other tasks in autonomous mining.

This thesis presents research towards an automated mine face inspection system by developing a method which fulfils the above requirements and which is therefore capable of classification and integration of hyperspectral data for use in autonomous mining.

1.5 Contributions of this thesis

This thesis is concerned with the identification and classification of natural rocks, minerals and geological zones on a vertical mine face in an open pit mine using hyperspectral remote sensing. A new spectral matching algorithm was developed to do this. The algorithm, which was proposed and developed in this thesis, conforms to the fundamental requirements outlined in Section 1.4. The proposed method uses a Gaussian Process (GP) framework with a specific covariance function (Observation Angle Dependent; OAD). Classified maps of rocks and minerals are probabilistic, in the sense that they have been derived using a fully probabilistic method and, at the same time, each pixel in the map has a probability and uncertainty associated with it.

Probabilistic thematic maps of rocks and minerals are important to enable their incorporation into a comprehensive in-ground geological model. This might not be possible using either existing manual mapping methods currently employed in the mining industry or conventional (i.e. deterministic) methods for the analysis of hyperspectral data. There is no underlying structure to these data which can form any basis for the integration of this information (i.e. hyperspectral data). This thesis focuses on the use of spectral-angle based methods, either operating in a probabilistic or deterministic framework. Comparison to feature-based methods are beyond the scope of this thesis.

The principle contributions of this thesis are:

- Construction of a spectral library under various, environmental conditions, e.g. spectra acquired from cores of rock which were acquired using different types of illumination and from different viewing angles. See Schneider et al. (2009).
- Development and validation of the Gaussian Process - Observation Angle Dependent (GP-OAD) algorithm for classification of hyperspectral data (Schneider et al., 2010).

- Extension of the GP-OAD method to hyperspectral imagery and a detailed assessment of the performance of the classifier across several experiments.
- Rigorous comparison of different measures of similarity and their effects on performance of classification using hyperspectral data. This was done to determine which type or group of classifiers are most effective in generating the most accurate maps of mine faces. An assessment of the effects of shade on the classification performance of the different methods was also done by applying them to two images of the same mine face. One of the images was fully-illuminated, the other image was shaded in some parts of the mine face (Schneider et al., 2011a).
- Development of a framework for fusing classified probabilistic outputs of the GP-OAD. A demonstration of the framework was done using classified maps of geology derived from hyperspectral images of the same mine face. Images were acquired of the same mine face at different times of the day.

1.6 Thesis outline

Chapter 2 describes the fundamental principles of reflectance spectroscopy and spectrometry. Details about specific types of sensors used in this thesis and their basic working principles are introduced. The chapter provides an overview to the physical processes which govern the causes of absorption, scattering and reflection of electromagnetic radiation.

Chapter 3 provides an overview of selected methods used for the analysis of hyperspectral data. The first part is concerned with conventional methods which are, generally, deterministic in nature. Advantages and disadvantages of the mentioned methods are discussed. The second part of this chapter provides a general introduction into the field of machine learning. It introduces Gaussian Processes as they are an integral part of the contribution of this thesis. Advantages of machine learning over conventional methods are discussed.

Chapter 4 gives an overview of the data used in this work and explains the methodology for the acquisition of the spectral library (training sets) and the imagery of mine faces for the two study sites investigated in this thesis.

Chapter 5 describes a major contribution of this thesis which is the development and implementation of the GP-OAD for the classification of hyperspectral data. The method is validated using several approaches using high-quality, high-resolution field spectrometer data.

Chapter 6 extends the probabilistic method to hyperspectral imagery. An assessment of the performance of the proposed method is done and compared to standard methods for the classification of hyperspectral data. This chapter also discusses the reasoning for selecting the Observation Angle Dependent covariance function over other covariance functions, specifically in relation to data where the training and test sets were acquired under different types of illumination.

Chapter 7 describes the development of a new framework for the integration (fusion) of multiple classified maps derived from hyperspectral imagery. Using this new framework, the full potential of the proposed GP-OAD method is demonstrated by classifying several images of the same target and fusing the individual, probabilistic classification outputs to increase the overall accuracy of classification.

Chapter 8 concludes this thesis with a summary of the major contributions and a look at the future work required to take these techniques forward towards a practical implementation in a mining environment.

Chapter 2

Principles of Spectroscopy

2.1 Physical processes

2.1.1 Basic properties of electromagnetic radiation

The signal that is measured in hyperspectral remote sensing is electromagnetic radiation (EMR) in the range of the visible near-infrared (VNIR) and short-wave-infrared (SWIR). The radiation can be considered either as a stream of particles (i.e. photons) or waves (Hecht, 2002). These two contrasting theories, i.e. Newton's theory regarding particles and the theories by Snell, Huygens and Fresnel regarding waves are reconciled under the dual representation in Quantum physics, introduced by Planck at the beginning of the last century (Townsend, 2012; Born, 1948). This theory views electromagnetic radiation in its dual formalism as both wave and particle and describes it as a superposition of sinusoidal waves of the electric and magnetic field. Each wave is characterised by a frequency f and a wavelength λ, according to

$$\lambda = c/f \ , \tag{2.1}$$

where c is the speed of light in vacuum ($299792458\, m \cdot s^{-1}$). The wavelength λ is linked to the angular frequency ν and wavenumber $\bar{\nu}$ by $\lambda = c/\nu = 2\pi/\bar{\nu}$. Generally,

λ is expressed in μm or nm and $\bar{\nu}$ is expressed in cm^{-1}.

In a vacuum, electromagnetic radiation travels at the speed of light, c. In a medium, EMR travels with the velocity c_n, where n is the refractive index of the medium.

At an atomic scale, energy is quantised; meaning that each atom has discrete energy states. The lowest energy state of an atom is called the *fundamental state*. When exposing an atom to EMR, the atom can change from a lower to a higher energy state, i.e. the atom becomes excited. This is called the *photoelectric effect* (Hertz, 1887) and is the basic principle of absorption and emission; in the case of emission, the atom changes from a higher to a lower energy state. For a photon to be absorbed by an atom the energy of the photon (E_{photon}) has to match the exact energy difference (ΔE) between two energy states (e.g. E_1 and E_2). Only discrete energy states are allowed for absorption and emission, thus a material (or atom) only absorbs specific wavelengths which are exactly proportional to the energy difference of the electron states (*Bohr-Rutherford's atom model*, Figure 2.1; Bohr, 1913), i.e.

$$\Delta E = h \cdot f = \frac{h \cdot c}{\lambda} \ , \tag{2.2}$$

where h is the Planck's constant (Millikan, 1916). This means a photon can be characterised by its wavelength (λ), its angular frequency (ν) or by its energy (Eq. 2.2).

Electromagnetic spectrum

Radiation is classified into spectral regions according to their frequency or wavelength in a vacuum, (Figure 2.2). The foci of this thesis are the visible to short-wave infrared spectral regions (0.35 to 2.5 μm)[1].

The visible spectral region (0.39 to 0.7 μm) is the region of the electromagnetic spectrum which can be perceived by the human eye (Figure 2.2). In the visible region,

[1] This is the VNIR region often specified in the context of hyperspectral remote sensing contexts. The visible light which can be perceived by humans ranges from 0.39 to 0.7 μm

mechanisms like electron energy are important; if the incident radiation is highly energetic (visible and near ultraviolet), the resulting spectrum is generally complex due to the superposition of different physical effects (i.e. electronic charge transfers and crystal field effects, Section 2.3.1).

The infrared region covers wavelengths between $0.7\,\mu m$ and $1\,mm$. In spectroscopy, however, this spectral region is conventionally divided into three sub-categories known as (according to Rencz, 1999):

1. near-infrared (NIR), 0.78 to 3 μm;
2. mid-infrared (MIR), 3 to 30 μm;
3. thermal infrared or long-wave infrared (LWIR) or far infrared (FIR), 30 μm up to 1 mm.

The boundaries between regions of the infrared are not consistently defined. The main reason for this division is related to the wavelength of the transmitters, receivers, and transmission bands of the atmosphere. In remote sensing, however, wavelengths between $0.35 - 1\,\mu m$ are commonly referred to as VNIR and between $1 - 2.5\,\mu m$ as SWIR.

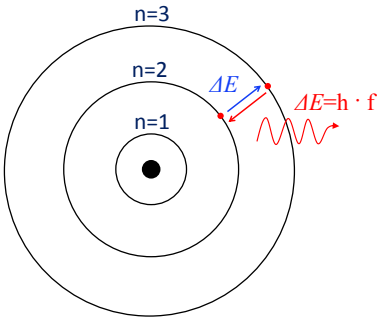


Figure 2.1 – Illustration of the Bohr-Rutherford model of a Hydrogen atom (Hecht, 2002) . The negatively charged electron is confined to an atomic shell. An electron transition between orbits is accompanied by an emitted or absorbed amount of electromagnetic energy ($h \cdot f$). The orbits in which the electron may travel are shown as black circles.

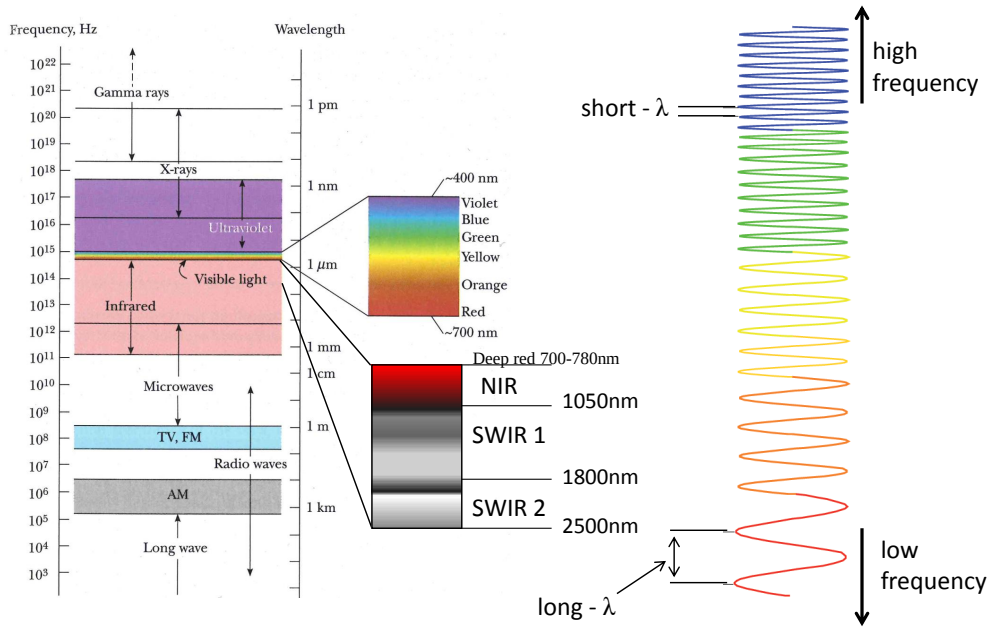


Figure 2.2 – Schematic representation of the electromagnetic spectrum. On the left side of the image names of spectral regions, their frequency and wavelength ranges are shown. The middle part of the diagram magnifies the VIS, NIR and SWIR region and shows their respective wavelength boundaries in nm. The inverse relationship between frequency and wavelength is depicted on the right side (i.e. $\uparrow f \propto \downarrow \lambda$). Image modified from Serway and Jewett (2004).

Electromagnetic radiation interacts with a surface through a combination of different processes such as reflection, transmission, absorption, diffraction, and multiple scattering (e.g., Wendlandt and Hecht, 1966; Kortum, 1969). The chemical structure of the mineral as well as the crystal structure in which the absorption occurs determines the position, shape, depth and width of the individual absorption feature. Thus, mineralogy of a sample can be related to spectral response and vice versa (Hunt, 1977; Clark et al., 1990).

2.2 Reflection processes

When photons from one medium (e.g. air) enter into another medium with a different refractive index, some are reflected away from the medium and others are refracted into the medium (Hecht, 2002; Kuehlke, 2007). The refractive index is comprised of a real part, n, and a complex part jK and is written as :

$$m = n - iK \ , \tag{2.3}$$

with $i = \sqrt{-1}$. The imaginary part is also called the extinction coefficient. When photons enter an absorptive medium they are absorbed according to *Beer's Law* (Elachi and Van Zyl, 2006; Beer, 1852):

$$I = I_0 \cdot e^{-k\ z} \ , \tag{2.4}$$

where I is the observed intensity, k is an absorption coefficient which depends on the properties of the medium and z is the distance traveled in the medium. The absorption coefficient z is dependent on the imaginary part of the refractive index of the material through

$$k = \frac{4\pi K}{\lambda} \ , \tag{2.5}$$

where λ is the wavelength of radiation (e.g. light). The absorption coefficient is traditionally expressed in cm^{-1} and z in cm. The above equations are wavelength dependent, i.e. their values change with varying wavelength.

2.3 Absorption Processes

2.3.1 Electronic Processes

Isolated atoms and ions have discrete energy states at which absorption of photons of a specific wavelength (which is proportional to a specific energy) can occur. When a photon is absorbed, it is usually emitted at a longer (i.e. less energetic) wavelength.

For example, absorption of a photon can cause heating of the material, resulting in grey-body emission at longer wavelengths (Serway and Jewett, 2004). In solid materials, electrons may be shared between individual atoms. The energy level of shared electrons may become smeared over a range of values called *energy bands* (e.g. Burns, 1993; 1970; Hunt, 1977, 1982 and Gaffey et al., 1993).

Crystal Field Effects

Electronic processes in the spectra of minerals are due to unfilled electron shells of transition elements (e.g. Ni, Cr, Fe, etc.), iron being the most common one. For all transition elements, the energy levels split in the d-orbital when the atom is located in a crystal field (e.g. Burns, 1993; 1970, Van de Meer and De Jong, 2001). This splitting of energy states enables an electron to be moved from a lower level into a higher level by absorption of a photon having an energy matching the energy difference between the states. The energy levels are determined by a number of properties (e.g. the valence state of the atom, coordination number, and the symmetry of the site it occupies). The crystal field varies with crystal structure from mineral to mineral and, thus, the amount of splitting varies. The same ion (e.g. Fe^{3+}; Burns, 1993) can produce, however, different absorptions, making specific mineral identification possible from spectroscopy. For example, absorption caused by Fe^{3+} for goethite (FeOOH) and hematite (Fe_2O_3) show different split energies of the d-orbital. An example of this is shown in spectra of goethite and hematite (Figure 2.3).

Charge Transfer Absorptions

Absorptions can also be caused by charge transfers, inter-element transitions where absorptions of photons cause electrons to move between ions or between ions and ligands (Elachi and Van Zyl, 2006). Transition can also occur between the same metal in different valence states, such as between Fe^{2+} and Fe^{3+}.

Generally, absorption bands caused by charge transfers are diagnostic of mineralogy and the depth and width can typically be hundreds to thousands of times stronger

than those of crystal field transitions (Morris et al., 1985; Burns, 1993). Band centres of charge transfer absorptions occur in the ultraviolet, with the wings of the absorption extending into the visible (e.g. Goethite spectrum Figure 2.3; Clark and Roush, 1984). For example, the red colour of iron oxides is mainly caused by charge transfer absorptions (Morris et al., 1985).

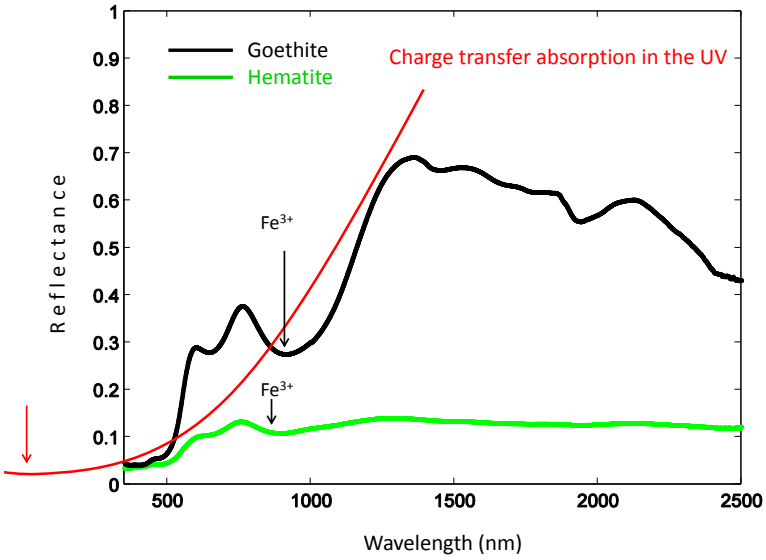


Figure 2.3 – Goethite (black) and Hematite spectrum (green) with arrows indicating their diagnostic absorption features and different absorptions caused by the different split energies of the d-orbital (see text). The red curve indicates a charge transfer absorption occurring in the ultraviolet and with wings extending into the visible and near infrared part of the spectrum.

Colour Centres

Absorption by colour centres is the reason for colours in minerals. Colour centers are caused by solar ultraviolet radiation of imperfect crystals (i.e. disruptions in the periodicity of a crystal). These defects, caused by impurities, can produce discrete energy states and electrons can become bound to them (Hunt, 1977). The transition of an electron into a defect requires photon energy and is considered as an electronic process of absorption (Gaffey et al., 1993).

2.3.2 Vibrational Processes

Vibrational processes occur in molecules rather than isolated atoms or ions. The bonds in a molecule between the different atoms (e.g. H_2O molecule) act like springs and the atoms like attached weights; thus the entire system can vibrate (Hunt, 1977). The frequency of vibration depends on the strength of each bond and their masses. For a molecule with N atoms, there are $3N - 6$ normal modes of vibrations called *fundamentals* (Rencz, 1999). Each vibration can produce *overtones* which occur at multiples of a single fundamental frequency. *Combinations* are due to combinations of different fundamental vibrations (Figure 2.4). Vibrations from two or more modes can occur at the same frequency which makes it impossible to distinguish their sources[2]. A free molecule can rotate and move laterally. In solids, partial rotation and slight translational movements (or lateral movements) can occur. Such lattice modes typically occur at very low energies (longer to mid-infrared wavelengths), beyond $20\,\mu m$. The higher the overtones, the weaker their influence on a particular absorption feature. From one overtone to the next higher one the influence is reduced by a factor of 30 to 100. In reflectance spectroscopy, these weak absorptions can be measured and diagnostic information is gained from second and third overtones and combinations (Farmer, 1974). An example of such overtone is the diagnostic kaolinite (Al-OH) feature which is caused by vibrational processes (Hunt, 1977, 1982). This absorption feature is intense and relatively narrow (Clark et al., 1990; Kruse, 1988) compared to absorptions caused by electronic processes (Figure 2.5).

Hematite, Fe_2O_3, has three strong stretching modes between 16 and 30 μm. Because iron oxides typically tend to be fine grained, often less than the wavelength of a mid-infrared photon, and due to the strong absorption in the mid-infrared, iron oxides tend to be dark in reflectance.

[2]They are impossible to distinguish, when they are not compared with their inherently linked combinations and/or overtones

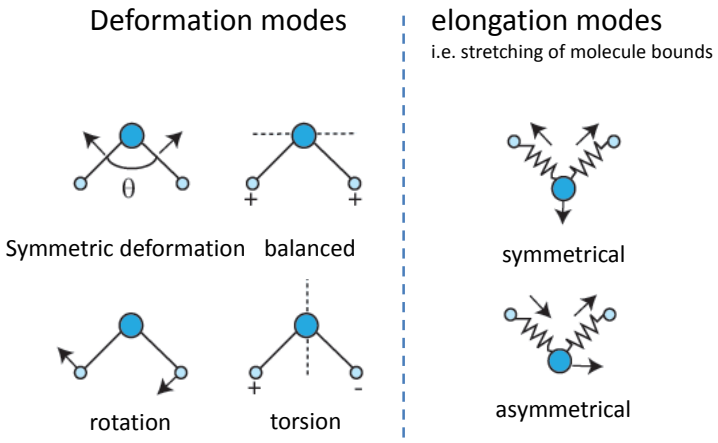


Figure 2.4 – Vibrational processes of a water molecule. The angle θ is the angle between the two H atoms in the H_2O configuration. From Verpoorter (2009).

2.4 Measurement of reflectance

Spectrometers and imaging spectrometers measure radiance ($Wm^2sr^{-1}nm^{-1}$), i.e. reflected radiation from an object at a solid angle (denoted in Steradian, sr) with a certain distribution of wavelengths. Spectrometers provide means to utilise remotely sensed reflectance spectra as a basis for estimating the material composition in, for example, geological (e.g. Clark et al., 1990, Scheinost, 1998, Kokaly et al., 2003) and oceanographic (e.g. Murphy et al., 2001) applications. This is because at visible and near-infrared wavelengths, reflectance spectra exhibit absorption features that are characteristic of the composition and crystal structure of the absorbing material (e.g. Burns, 1970; Singer, 1981; Hunt and Salisbury, 1970, 1971).

2.4.1 Units of measurement

The spectrum of a remotely sensed material is largely dominated by the solar irradiance curve, characterised by a peak at 504 nm and decreasing with longer wave-

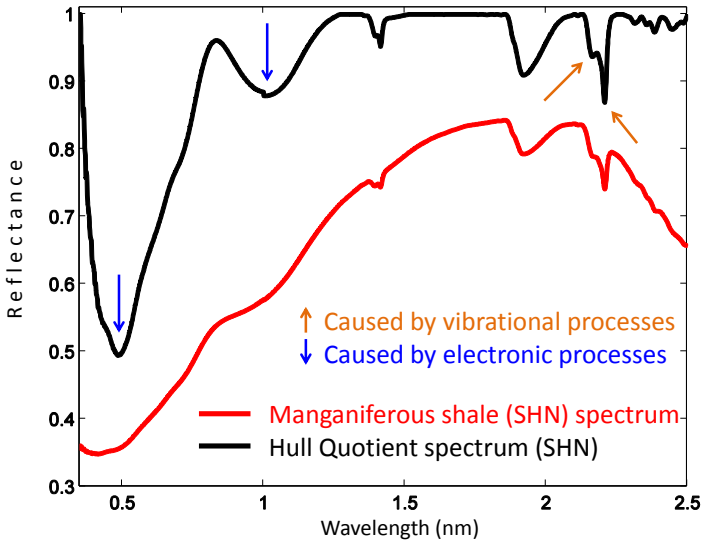


Figure 2.5 – Visual comparison between absorptions due to electronic and vibrational processes in reflectance and hull quotient spectra. Absorptions due to electronic processes are generally very broad (minimum indicated by downward blue arrows). Vibrational processes cause narrow features, e.g. kaolinite doublet (minimum indicated by upward orange arrows).

lengths[3]. The spectrum is further modified by absorptions caused by the atmospheric gasses (e.g. ozone, oxygen, carbon-dioxide) and water vapour (Strutt, 1899). Absorption by the atmosphere is wavelength-dependent and reduces the amount of light incident upon objects on the Earth's surface. Atmospheric scattering can increase the amount of radiation detected by the sensor (e.g. increased Rayleigh scattering towards shorter wavelengths; Sneep and Ubachs, 2005). Spectra measured using artificial light do not contain these effects caused by the atmosphere, however, they are still dominated by the wavelength distribution of the light source (Figures 2.6 and 2.7).

To compensate for these effects, target spectra have to be normalised to a reference with known reflectance properties, i.e. the reflectance coefficients for all wavelengths are known. Spectralon is a commonly used material and is manufactured so that it is diffuse, spectrally flat and highly reflective (99%) across the VNIR and SWIR

[3]Note that this is not the case for proximal sensing or lab technologies.

part of the spectrum. By dividing a target spectrum by the reference spectrum a unitless, normalised measure (reflectance) is obtained in the range of 0-1 (Figure 2.7 b). This is called relative reflectance because it describes the target brightness as a function of wavelength relative to the standard it was measured against. If the reflectance factor of the calibration panel is known, this can be multiplied to the reflectance spectrum to give an *absolute reflectance*. Conversion to units of absolute reflectance compensates for specific reflectance characteristics of the calibration panel and thus makes spectra more comparable between studies. This method, called *flat field calibration*, is commonly used for the calibration of field spectra acquired using a spectrometer or imaging spectrometer (e.g. Murphy et al., 2008, 2012), however, it does not compensate for additive effects. Spectra in absolute reflectance units are used in spectral libraries, where they serve as endmember spectra for training of classification algorithms. Other approaches to remove the effects of the atmosphere include atmospheric modelling (e.g. MODTRAN, Berk et al., 1998). These models are now commonly used for calibration of data acquired by airborne sensors. Such

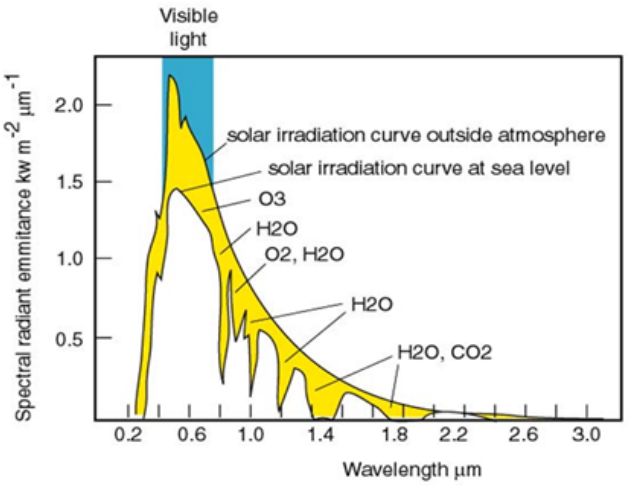


Figure 2.6 – Solar irradiance curve outside the atmosphere and at sea-level. Outside the atmosphere the irradiance curve is relatively smooth and has higher values (i.e. is offset against the curve at sea-level). At sea-level the solar irradiance curve shows gaps caused by absorption of specific wavelengths due to particles and gases in the atmosphere (e.g. water, water-vapour, carbon-dioxide). From Serway and Jewett (2004)

methods are, however, beyond the scope of this thesis.

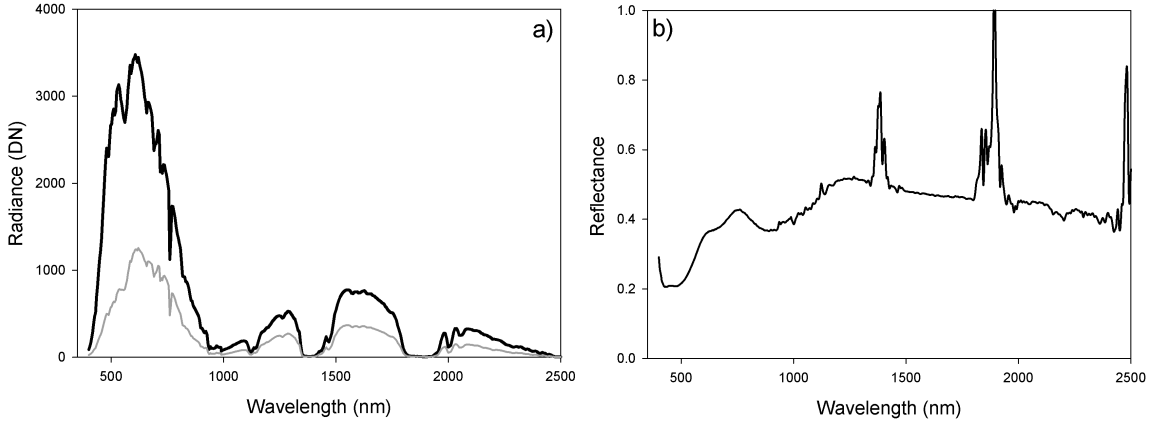


Figure 2.7 – Comparison of radiance and reflectance spectra. a) Radiance spectrum of shale (grey line) and radiance spectrum (black line) of a calibration target (Spectralon); b) Reflectance spectrum after dividing the target spectrum by the reference spectrum. Note that in a) it is difficult to detect diagnostic absorption features in the shale spectrum or even detect differences from a reflectance target (except differences in albedo) because the spectrum is dominated by the illumination characteristics. b) shows noise around 1450 and 1900 nm caused by atmospheric absorptions.

2.4.2 Types of sensors used for reflectance spectroscopy

Spectrometers operate in one or more spectral regions (e.g. VNIR or SWIR) and detectors are designed to detect radiation over these specific wavelength regions. Generally, two or more detector arrays are required to obtain a continuous spectrum between several spectral regions. For example, wavelengths larger than 1 μm are largely reflected by Silicon-Dioxide (SiO_2) CCD arrays which means that photons at these wavelengths are 'invisible' for these detector arrays. Thus, CCD or CMOS detectors of different materials have to be used, such as Indium-Gallium-Arsenide (InGaAr) and Mercury-Cadmium-Telluride (HgCdTe) among others, which have a much higher quantum efficiency than SiO_2 at longer wavelengths.

Two different categories of spectrometers are presently used (i) non-imaging spectrometers and (ii) imaging spectrometers. The first category measures the light dis-

tribution over a small area. This type of spectrometer averages the spectral response over a discrete target area and represents it by a single spectrum.

Imaging spectrometers, as their name implies, record an image where each pixel in the image is a complete spectral curve. Imaging spectrometers are mainly scanning sensors (i.e. line-scanners). They record a single spectral line of data for all bands using a 2D detector with one dimension sensing the spatial and the other the spectral domain. The forward motion of an aircraft or movement of a scanning table enables an entire image or image cube to be constructed. The resulting data-set is not a single spectrum but a data cube, where x and y dimensions represent the 2-d image and the z-dimension represents the spectral domain (Figure 2.8). Pushbroom scanners are commonly used to obtain such a hyperspectral data-set. For example, Townshend (1980) gives a short overview of sensing systems (see also De Jong and van der Meer, 2004; Schowengerdt, 1997).

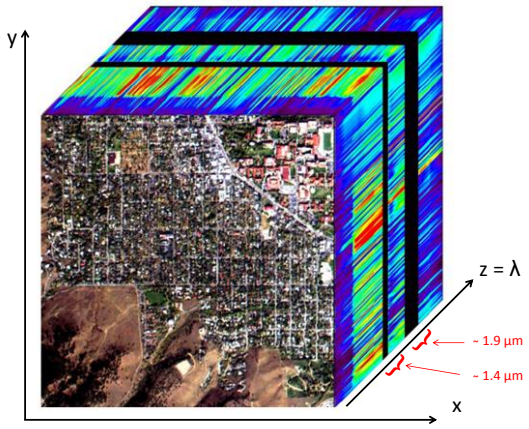


Figure 2.8 – Imaging spectrometer data hypercube. Colours in the z-dimension represent reflectance values. Blue colours indicate low values, red high values. Bands absorbed in the atmosphere due to water vapour are removed (black). Image modified from (Goetz, 2009).

2.5 Constraints of measurements

There are five parameters or variables which characterise the capability of a spectrometer: (i) spectral range, (ii) spectral bandwidth, (iii) spectral sampling, and (iv) radiometric resolution and (v) the signal-to-noise ratio. For imaging spectrometers, a sixth parameter is used to characterise the spatial information, i.e. (vi) the spatial resolution.

2.5.1 Spectral range

The spectral range over which a hyperspectral sensor detects light is important and sensors need to cover enough diagnostic spectral absorptions to solve a certain problem. In spectral remote sensing, the range that most sensors use, and which is the subject of this work, is from the VNIR to the SWIR (i.e. 0.35 - 2.5 μm). Diagnostic absorptions for minerals occur in both the VNIR and the SWIR.

2.5.2 Spectral resolution

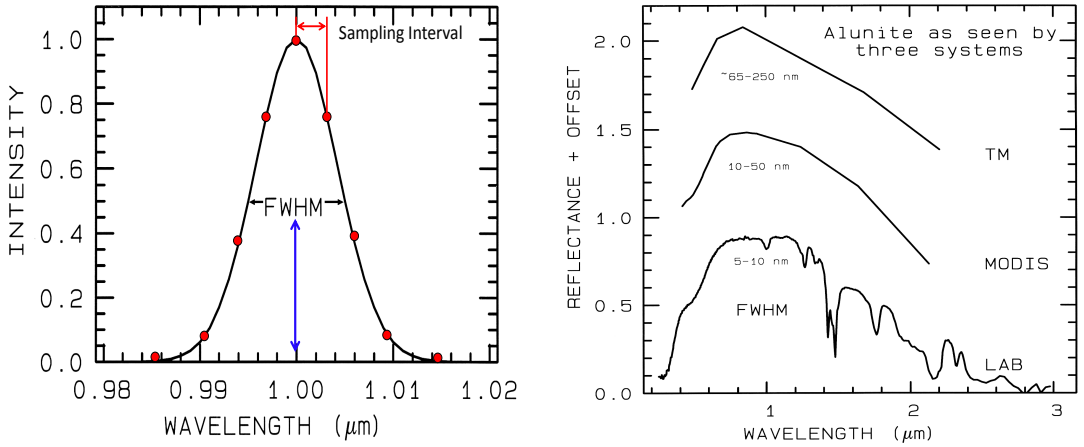
The location, width and the number of spectral bands determines the spectral resolution of the sensor (Mather and Magaly, 2010). The term spectral resolution is often confused with the spectral bandwidth and / or spectral sampling. In fact, the spectral resolution considers both spectral sampling and spectral bandwidth (Figure 2.9).

Spectral sampling and spectral bandwidth

Spectral sampling is the distance in wavelength between the spectral bandpass profiles for each channel in the spectrometer as a function of wavelength. Some sensors collect multispectral or multi-band images, which are individual images separately recorded at discrete spectral bands (Mather and Magaly, 2010). Spectral bandwidth refers to the width of these spectral bands measured in nanometres (nm). Sampling interval

and bandwidth describe the spectral resolution of a system. In some hyperspectral sensors the sampling can occur several times per bandpass (Figure 2.9a).

Spectral bandwidth is the width of an individual spectral channel in the spectrometer. The narrower the spectral bandwidth the narrower the absorption feature the spectrometer can resolve.



(a) Sampling Interval **(b)** Spectral bandwidth

Figure 2.9 – Spectral resolution is defined by the sampling interval and the spectral bandwidth. a) Spectral sampling interval and bandpass of a spectrometer. The bandpass is characterised by the width of the spectral band which is assumed to has a Gaussian shape. The width is measured at half the maximum intensity (i.e. FWHM). Spectral sampling interval is indicated by small red dots on the Gaussian curve. b) Alunite spectrum acquired with three different sensors. Top spectrum shows the spectrum as seen by a multispectral sensor (Landsat-TM) with broad spectral resolution. The middle spectrum shows the same spectrum with a higher spectral resolution using the moderate resolution imaging spectroradiometer (MODIS). The bottom spectrum is a high-resolution spectrum acquired using a field, or non-imaging, spectrometer showing diagnostic absorptions. Values in the figure indicate the FWHM of the sensors. Both figures from Rencz (1999), a) was slightly modified.

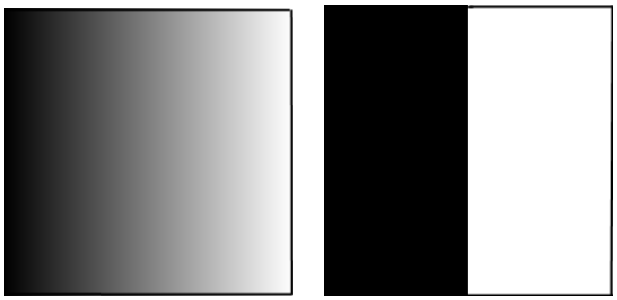
For example, the multispectral scanner (MSS) of the Landsat-5 Thematic Mapper (TM) resolves a spectrum with only six spectral bands in the spectral range of 0.4-2.5 μm. These spectral bandpasses are sufficient to separate, for example, clay material (e.g. kaolinite or alunite) from iron-oxides. The TM SWIR band does, however, not allow identification of any mineral, including clay which have important diagnos-

tic absorption features in this spectral region. It is also not possible to unambiguously identify any minerals using these broad bandpasses without *a priori* knowledge about the scene. In contrast, hyperspectral sensors (e.g. AVIRIS, Figure 2.9b) generally have sufficient spectral bandwidth (i.e. many narrow bandwidths, contiguously spaced) to resolve even narrow spectral features, and thus enable identification of minerals based on their diagnostic absorption features. A review of effects of spectrometer bandpasses, sampling interval, and signal-to-noise ratios on classification accuracy is given in Schowengerdt (1997).

2.5.3 Radiometric resolution

Radiometric resolution refers to the quantisation from an analogue or continuous signal to a digital, discrete signal such as intensity values in a standard camera image. There are three properties that characterise radiometric resolution. These are (i) the inherent sensor sensitivity, (ii) quantisation level of the sensor (bit-depth) and the electronics and (iii) the signal-to-noise ratio (SNR or S/R). The inherent sensor sensitivity is determined by the material used for the light detecting sensor and is usually referred to as quantum efficiency. The level of quantisation generally determines the level of detail that can be collected by a sensor (Schowengerdt, 1997). This can be illustrated by imaging a scene acquired with a standard camera with 256 quantisation levels compared to a binary image with only two quantisation levels (Figure 2.10). Older multispectral scanners (e.g. Landsat-1) had 6-bit sensors, which limits the differences in reflectance that can be recorded due to the inherent sensor noise.

Today, hyperspectral sensors can have radiometric resolutions of up to 16-bits (e.g. non-imaging spectrometers). Imaging spectrometers generally have around 12 bits for the VNIR part and 14 bits for the SWIR sensors. A high bit-depth is necessary, especially for SWIR sensors because of a higher inherent sensor noise (or dark-current).



(a) 256 levels of quantisation - 8 bits

(b) 2 levels of quantisation - 1 bit

Figure 2.10 – Effects of quantisation. a) A greyscale image width 256 shades of grey quantised with 8 bits. b) Same image quantised with 1 bit.

2.5.4 Signal-to-Noise Ratio

The quality of the spectral signal is an important determinant for the identification of diagnostic spectral signatures, i.e. a spectrometer must measure a spectrum with enough precision to record subtle changes in the curve shape of the spectrum (Landgrebe, 1978). The quality of the signal is determined by the amount of random sensor noise compared to the amount of signal, i.e. signal-to-noise ratio (S/N or SNR). Large amounts of noise can degrade diagnostic absorption features making the identification of mineral types and their abundance more difficult (Murphy et al., 2013).

The signal-to-noise required to solve a particular problem will depend on the strength of the spectral features under study (Goetz and Calvin, 1987). The signal-to-noise ratio is dependent on the detector sensitivity, spectral bandwidth, and intensity of the light reflected or emitted from the surface being measured. Swayze et al. (2003) discussed the inherent sensor characteristics (e.g. spectral resolution and SNR) with respect to spectral identification.

2.5.5 Spatial resolution

The spatial resolution of a system can be characterised in several different ways (Mather and Magaly, 2010). Townshend (1980) describes four of them; each of these use different criteria to define the spatial resolution. These criteria are the

ability to (i) distinguish between two point targets, i.e. uses the Rayleigh criterion (Perrin, 1966; Slater, 1977); (ii) resolve periodic patterns by quantifying the optical transfer function (OTF) and the modulation transfer function (MTF; Steiner and Salerno, 1975); (iii) measure the spectral properties of small finite targets by defining the effective resolution element (ERE, Colvocoresses, 1979). The final criterion (iv) is based on the geometrical properties of imaging systems which is probably the most commonly used criterion. The geometrical properties of imaging systems can be expressed by the instantaneous field of view (IFOV) which is defined as the area on the target (e.g. a mine face) viewed by the instrument from a given distance (Mather and Magaly, 2010). The quality of the geometric resolution is subject to the abberations and diffraction occurring in an optical system which can be described by the point spread function (Figure 2.11) of an optical system (Welch, 1977; Hecht, 2002).

2.6 Summary

This chapter discussed the basic principles of spectroscopy which forms the basis for hyperspectral remote sensing in the VNIR-SWIR spectral region. The physical mechanism that lead to the reflection, absorption and scattering of light were described and linked to spectral curve shape and the appearance of some rock types (e.g. the yellow colour of goethite). Charge transfer and crystal field effects were introduced which cause broad absorption in the VNIR part of the spectrum (e.g. ferric iron absorption around 900 nm). Furthermore, vibrational processes were explained, these processes lead to narrow absorption features in the SWIR (e.g. the kaolinite doublet).

The measurement of reflected light from target objects was explained and the problems associated with the acquisition of radiance spectra using natural illumination (e.g. atmospheric windows) were discussed. One method to compensate for the spectral effects of light sources, which are unrelated to the material composition of a target, is a method called flat-field calibration. This method converts the measured radiance to a unitless measure of reflectance by measuring the radiance spectrum of a standard material (e.g. Spectralon) with known reflectance characteristics.

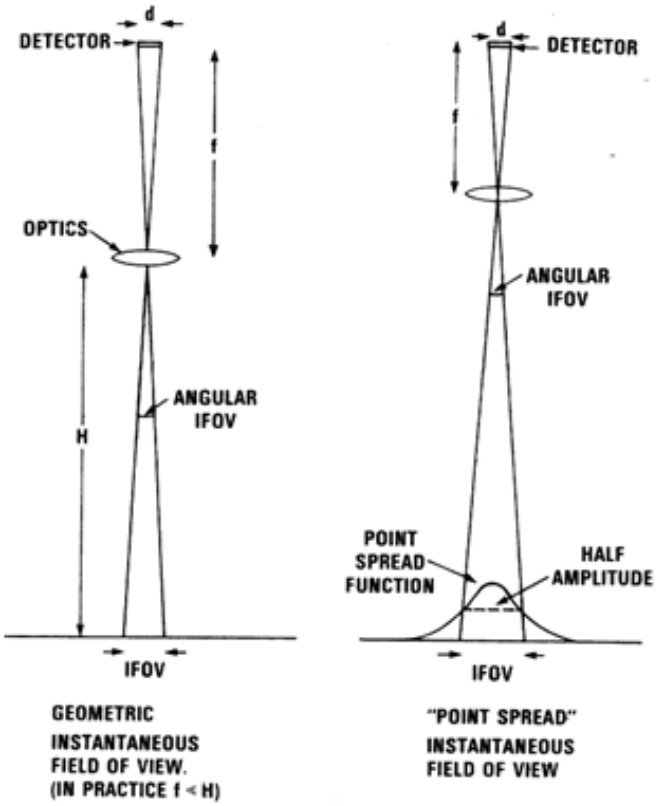


Figure 2.11 – Comparison of geometric IFOV and the actual IFOV. Left: Geometric field of view is illustrated using geometric optics. Right: The field of view in reality is blurred out due to physical limits of optics, i.e. a single point is imaged as a spread point in the image plane (detector), the size of the spread can be accurately calculated using Fourier optics and the Point Spread Function (PSF). Object distance (H), focal length (f'), detector size (d). From Townshend (1980).

Non-imaging spectrometers (also called discrete spectrometers) and imaging spectrometers were introduced as instruments to measure reflected light from target objects. Both instruments were used in this thesis and represent state-of-the-art instruments. Often, non-imaging spectrometers acquire reflected light with more spectral bands and narrower bandpasses than imaging spectrometers do, thus the higher spectral resolution leads to a better resolution of narrow absorption features (e.g. the kaolinite doublet). Another technical differences are the radiometric resolution (i.e. bit depth) and the signal-to-noise ratio of spectrometers which are often higher for discrete spectrometers.

The spatial resolution is a critical factor for imaging spectrometers. The spatial resolution has an influence on the minimum object size that can be resolved in an image. However, because objects in an image can sometimes occupy only fraction of a pixel (instead of whole pixels), the radiometric resolution (i.e. the bit depth) is coupled to the spatial resolution and the target and background reflectance in order to make a statement about the visibility of an object in an image (Schowengerdt, 1997, Chapter 3).

This chapter has provided a brief introduction to the principles of hyperspectral remote sensing. The discussed topics for the basis to understand the following chapters of this thesis. Detailed discussions of the topics mentioned in this chapter can be found in the literature cited throughout this chapter.

Chapter 3

Methods for hyperspectral data analysis and machine learning

The composition of materials can be determined by the analysis of specific absorption bands that comprise a reflectance spectrum (Van de Meer and De Jong, 2001; De Jong and van der Meer, 2004). The analysis of these absorptions according to their wavelength-position and their shape is used to identify and quantify the constituting minerals, pigments or chemicals. In this chapter, some selected conventional techniques used to analyse the shape of absorption features, resolved in hyperspectral data, are introduced and discussed. The second part of this chapter introduces the concept of machine learning for pattern recognition and matching. Further discussion will show that machine learning techniques have the potential to greatly-enhance hyperspectral scene analysis and classification.

3.1 Traditional methods for analysis and classification of hyperspectral data

Imaging spectrometers (hyperspectral scanners) provide sufficient spectral sampling to resolve diagnostic absorption features of materials. This enables direct identifica-

tion of materials based on their properties of absorption (Clark et al., 1990; Goetz et al., 1985). The spectral dimensionality, volume and complexity of these data impacts the users' ability to interact with the data and its statistical properties (Goetz et al., 1985). This has required the development of new methods to extract information from hyperspectral data in order to exploit these data to their full potential. Thus, to process hyperspectral data, several methods exploiting different aspects of the data have been proposed. Some of these methods were designed to work on individual absorption features in spectra, whereas others operate on the entire spectral curve. It is neither possible nor desirable to discuss all methods developed for hyperspectral data analysis, however, some concepts and methods need to be introduced as they are relevant to this work.

The objective behind the use of hyperspectral data is that materials can be identified by matching the spectrum (or pixel-spectrum in the case of imagery) to a library of spectra of known identity (reference data). In order to do this, unknown spectra must be in the same units as the reference spectra (i.e. reflectance). Reflectance spectra acquired in field settings can be influenced by a number of factors which change their brightness or their shape, when compared to reflectance spectra acquired from the same material assemblage using a spectrometer or a laboratory instrument.

In order to characterise absorption features in spectra, external factors, influencing the spectral curve, need to be removed from the spectral curve shape to improve the accuracy of matching unknown and library spectra based on their spectral absorption features. External factors can cause changes in spectra which are unrelated to mineral composition of the rock. An example for such an external factor are changes in brightness or the type of illumination in general (e.g. wavelength-intensity distribution) which is incident upon the surface of the rock. There are inherent characteristics of the rock that can change the reflectance spectrum of the rock, for example, the grain-size of the mineral of which it is constituted or organic content (Clark and Roush, 1984; H, 2012). Differences in the material brightness of rocks or different illumination can mask absorption features caused by material composition. Reflectance calibration compensates for variations in illumination but not for inherent brightness

differences of materials. Thus, removing these inherent brightness variations of a rock, i.e. removing the background brightness or continuum of a spectrum, effectively normalises the spectral curve shape to a common baseline. This baseline makes it easier to improve the accuracy of algorithms to characterise and match absorption features.

Deterministic techniques can be grouped into two categories, absorption feature based methods and spectral curve shape methods, i.e. methods which consider the spectral curve shape as a whole and not just individual features. Several deterministic methods identify absorption features in spectra and parameterise them to find correlations with a reference library. Matching parameters or spectral absorption features with those of library spectra is commonly done using correlation coefficients or least square fits (e.g. Clark et al., 2003). Correlation coefficients and least-squares can, however, also be used to match entire spectral curves by determining the cross-correlation along the spectral curve shape (e.g. van de Meer and Bakker, 1997; van der Meer, 2006). Absorption feature based methods are often used on a specific part of the spectral curve where diagnostic absorption features occur (e.g. 2.0-2.5 μm). Diagnostic absorption features must be present in order for these methods to work. Other methods, however, operate on the entire spectral curve shape and calculate some measure of similarity using a specific metric between unknown and reference spectra, which describes the degree of match. A selection of techniques from each of these categories is discussed. Some of these methods rely on preprocessing of spectra by removal of a spectral continuum, as discussed below.

3.1.1 Continuum removal

Continuum removal does not belong to either feature-based methods or to spectral curve shape methods for classification of hyperspectral data. Continuum removal is, however, an important preprocessing step for methods using diagnostic absorption features. It therefore is discussed in this section.

The spectral continuum describes the general brightness of the spectrum which is largely unrelated to specific absorption features of particular materials. To place all

diagnostic absorption features on the same reference plane the spectral continuum must be removed. There are several approaches in the literature to determine the continuum of spectra (e.g. Clark and Roush, 1984; Sunshine et al., 1990). In this work, only the traditional continuum removal is discussed (Clark and Roush, 1984; Clark et al., 1987).

The method by Clark and Roush (1984) estimates the upper convex hull (continuum) as a set of straight line segments tangent to the local maxima of absorption bands, spanning the entire spectrum[1] (Figure 3.1). The continuum is then removed from the reflectance data by dividing the reflectance in each channel by its respective continuum value. This places the continuum removed or 'Hull Quotient' spectrum on a reference plane (i.e. a reflectance value) of '1'. Absorption features are measured relative to their reference plane and have values smaller than '1'. A variation of this method is to subtract the reflectance values from the hull to produce a 'Hull Difference' spectrum, this places the continuum removed spectrum on a reference plane of '0'. Clark et al. (1984, 2003) argues that 'Hull Difference' spectra are appropriate for absorbance spectra because multiple components have additive effects on spectra. For reflectance, emittance and transmittance spectra, however, the continuum is removed by division due to exponential scattering processes (Clark and Roush, 1984).

3.1.2 Absorption feature based methods

Derivative Analysis

Although derivative analysis does not require the presence of absorption features in spectra, it is included here because it is commonly used to enhance their presence in spectra. Derivative analysis quantifies changes in slope from adjacent parts of the spectral curve and removes the background brightness (continuum). By applying derivatives to reflectance spectra, resulting spectra are set to a common baseline

[1] It is also possible to remove the continuum locally, i.e. over a couple of spectral bands. This methods is useful when data are noisy, however, it requires, *a priori* knowledge about the minerals in question because node points have to be placed at sensible wavelength positions for the method to work.

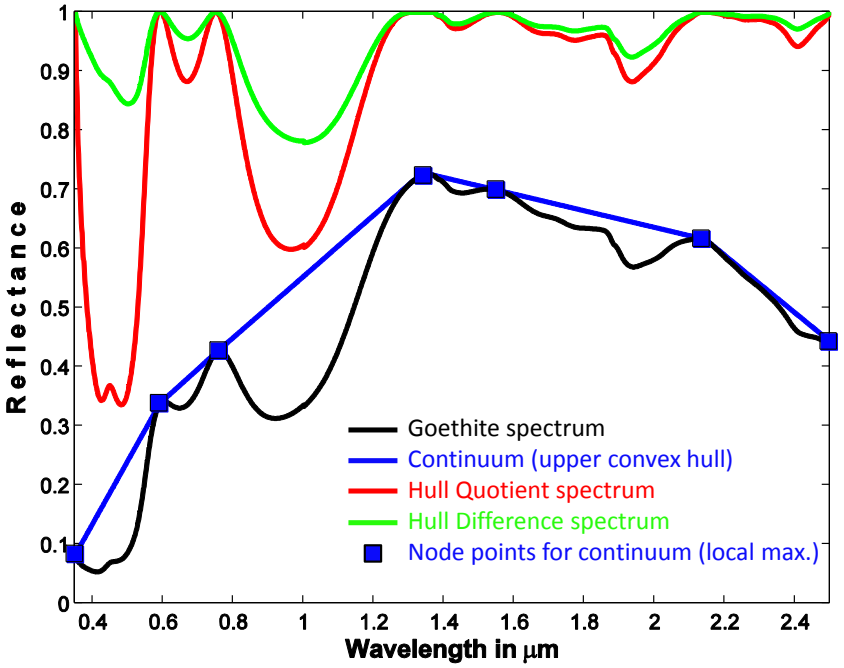


Figure 3.1 – A spectrum of goethite (black curve) for which the convex hull or continuum (blue curve) was determined. Six segments are connected by seven nodes (blue squares). The hull quotient (red curve) and hull difference was determined using the described methodology. Note that '1− hull difference' is shown (green curve).

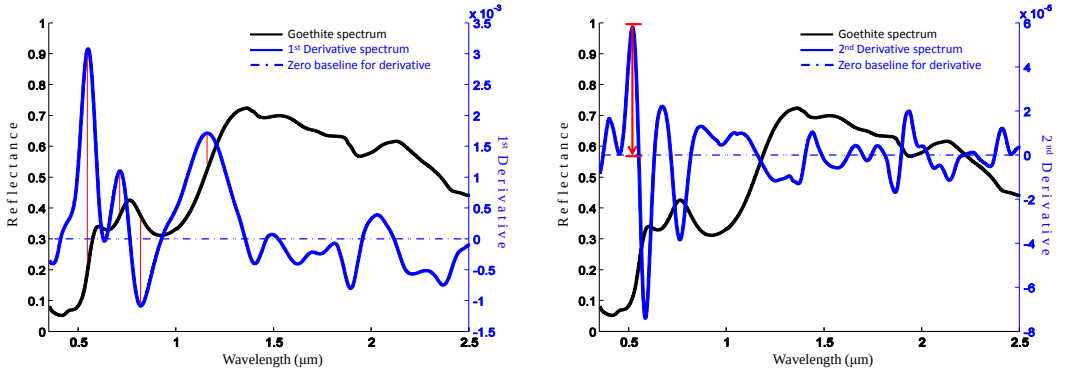
of *zero* with positive and negative peaks (Talsky et al., 1978). The continuum in spectra can have a profound impact on the spectral curve shape and consequently influence classification methods if they are not removed. For example, the background brightness in vegetation spectra is mostly due variations in the brightness of the soil background. These variations can have a large impact on classification and quantification. With an increasing order of derivative, however, soil background effects are progressively minimized. By the second derivative, these background effects are virtually eliminated (Demetriades-Shah et al., 1990). Derivatives can also be used to enhance very subtle absorption features and separate overlapping features which may appear as one feature in reflectance data (Huguenin and Jones, 1986; Murphy et al., 2008).

Removal of brightness variations from spectra by derivative analysis allows different orders of derivatives to provide information about different aspects of absorption. The first derivative has peaks and dips at the wavelengths of the positive and negative inflection points of spectra (Figure 3.2a). Such information can be used in analysis of vegetation spectra to find the red-edge – an important parameter which relates to the biophysical properties of vegetation (e.g. Zarco-Tejada et al., 2003; Smith et al., 2004; Cho and Skidmore, 2006). The second derivative of reflectance has peaks at the wavelengths of maximal absorption, i.e. the centre of absorption features (Figure 3.2b). The height of the peaks above the zero baseline provides information on the strength of the absorption and therefore the amount of the material causing the absorption. Derivation to the fourth order increasingly resolves smaller, more subtle absorption features in the spectrum. Derivative analysis has a large number of applications. In ecological applications for example, derivatives can be used for estimating chlorophyll-a in terrestrial vegetation (e.g. Horler et al., 1983; Blackburn, 1999) or quantifying amounts of chlorophyll-a and the pigments in algae (e.g. Murphy et al., 2005).

Derivative analysis are often used in geological and mineralogical applications where they have been applied successfully to identify and quantify minerals (e.g. Fe oxide or clays) in soils (e.g. Scheinost, 1998; Sellitto et al., 2009). The effectiveness of derivative analysis was demonstrated by achieving a detection limit of the hematite/-

goethite ratio which was found to be an order of magnitude lower ($5\,\mathrm{g\,kg^{-1}}$) than the detection limit for X-ray diffraction (XRD) analysis (Scheinost, 1998).

A major problem with the use of derivatives is that they introduce noise into the spectrum which can erroneously be identified as features. The amount of noise increases with the order of derivatives. If derivatives are calculated geometrically (i.e. dx/dy), the spectrum needs to be smoothed beforehand. Other methods which are based on numerical differentiation allow smoothing of the spectrum to be done at the same time as differentiation. A number of smoothing techniques have been proposed in the literature (reviewed by Tsai and Philpot, 1998). Another problem with derivatives is that *satellite peaks*[2] can be erroneously identified as peaks related to absorption (Talsky et al., 1978). A way to avoid this confusion is to find correspondent features of the true absorption in the first and second order derivatives.



(a) First order derivative **(b)** Second order derivative

Figure 3.2 – First and second order derivative spectrum. a) The minima and maxima of the first derivative (blue curve) coincide with the inflection points of the respective goethite spectrum (black curve) below and above the zero baseline (dashed blue line). A visual aid is provided using red lines which connect some of the major peaks in the derivative spectrum with the respective inflection points in the goethite spectrum. b) The height of the peaks in the second derivative (red arrow) can be used to determine abundances of specific minerals relating to the absorption features.

A method for smoothing and simultaneously differentiating hyperspectral data has been developed by Savitzky and Golay (1964). This method uses simplified least-

[2] e.g. double peaks in the second, forth and consequent derivatives

square fit convolution for smoothing and computing derivatives. A least mean-square smoothing method by Kawata and Minami (1984), similar to Savitzky and Golay's method, is effective when the signal mean and variance is known. Both methods can distinguish better between fine absorption features and noise than naive smoothing methods (e.g. moving average) and thus preserve absorption features. In contrast, simple smoothing techniques such as the 'mean filter smoothing' are quite effective in smoothing spectra but they do suppress fine spectral features. For all methods, an appropriate smoothing window size needs to be determined to obtain optimal results.

Extraction of parameters from absorption features

The relative wavelength-position, depth, area, asymmetry and width of individual absorption features, the *feature parameters*, in a spectrum provide valuable information about a material's composition such as mineral type, abundance and crystallinity (e.g. Kruse, 1988, Murphy et al., 2005). Extracting these feature parameters enables the characterisation of each absorption feature in the spectrum to be compared in a consistent way with a reference library. Feature parameters can be used to identify and classify materials and/or quantify the abundance of materials (Clark et al., 2003). The extracted parameters can themselves be used to provide maps of different minerals. For example, the wavelength position of the deepest absorption feature in the spectrum can itself be considered to be a map of different types of iron minerals (e.g. hematite or goethite). Parameterisation can also be used as a method to reduce the dimensionality of the data, thus discriminating the information contained in the numerous bands of the original spectrum. By representing the spectrum in much fewer parameters, the performance of mapping techniques and machine learning algorithms may be improved.

For example, the parameters of absorption features (Figure 3.3 and 3.4) are specific for goethite. Parameters describing the shape of the absorption directly relate to physical and chemical properties of goethite (Cudahy and Ramanaidou, 1997, Haest et al., 2012a,b). Parameterisation of absorption features generally operates on spectra after the reflectance continuum is removed using a hull-quotient (i.e. division)

procedure. Features are parameterised in terms of their:

1. Wavelength position
2. Total Area
3. Depth
4. Asymmetry
5. Full Width at Half Maximum depth (FWHM)
6. Full width at the shoulders of the absorption feature

'Wavelength position' is simply defined as the wavelength where the absorption is maximal (i.e. reflectance is minimal). The total area is the area of the absorption feature bounded by its shoulders and is calculated as the sum of the left and right halves of the absorption feature. The left and right halves are calculated using equations 3.1 and 3.2:

$$A_{left} = \sum_{i=min\text{-}1}^{i=LS} R_{max} - R_i \tag{3.1}$$

$$A_{right} = \sum_{i=min+1}^{i=RS} R_{max} - R_i \ , \tag{3.2}$$

where R is the hull quotient reflectance (denoted R_b in Figure 3.4), 'min' wavelength of maximal absorption, LS and RS are the left and right shoulders of the absorption feature and R_{max} is the maximum hull quotient reflectance in the spectrum (denoted R_c in Figure 3.4)

The depth is calculated as the hull quotient's reflectance at the absorption maximum of the absorption feature minus the maximum hull quotient's reflectance in the spectrum, i.e. $R_b - R_c$ (Figures 3.3 and 3.4).

Asymmetry is calculated by taking the natural log of the right area divided by the left area (Equation 3.3). This yields a value of zero for an absorption feature if it is

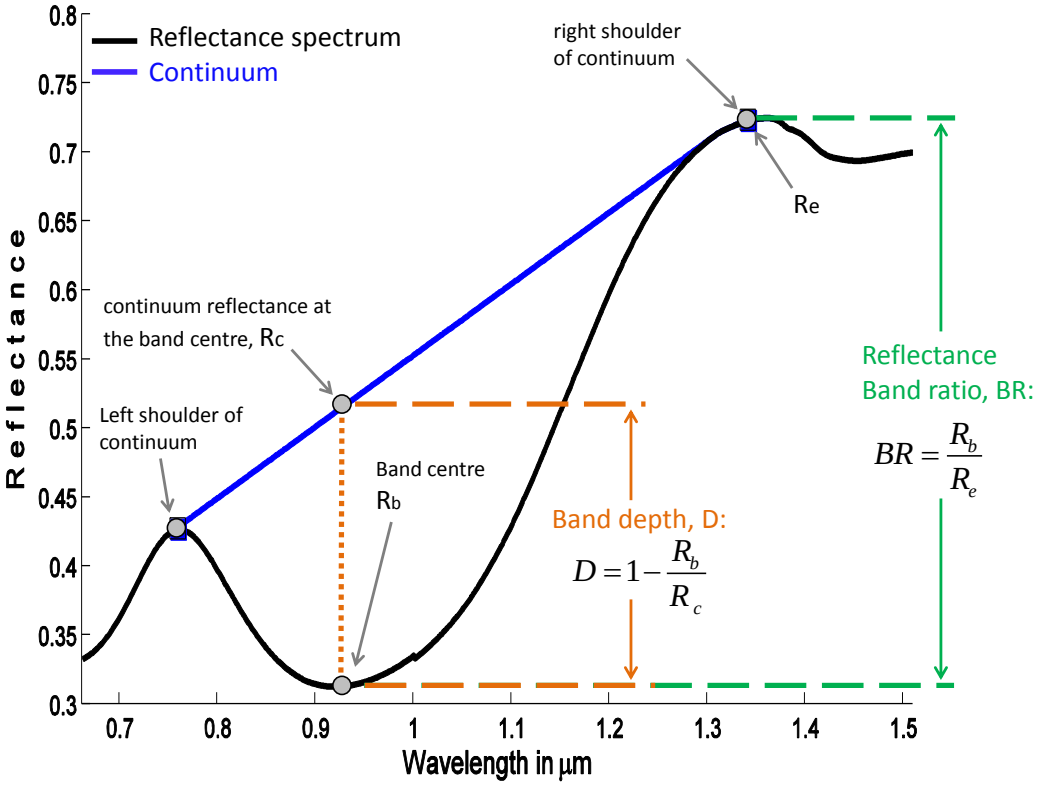


Figure 3.3 – Conventional criteria for estimating the shape of the absorption bands in the case of goethite ($Fe^{3+}O(OH)$) of the band centered at 0.93 microns. The gray dots represent the bands used to calculate parameters such as the band ratio (BR) and band depth (D).

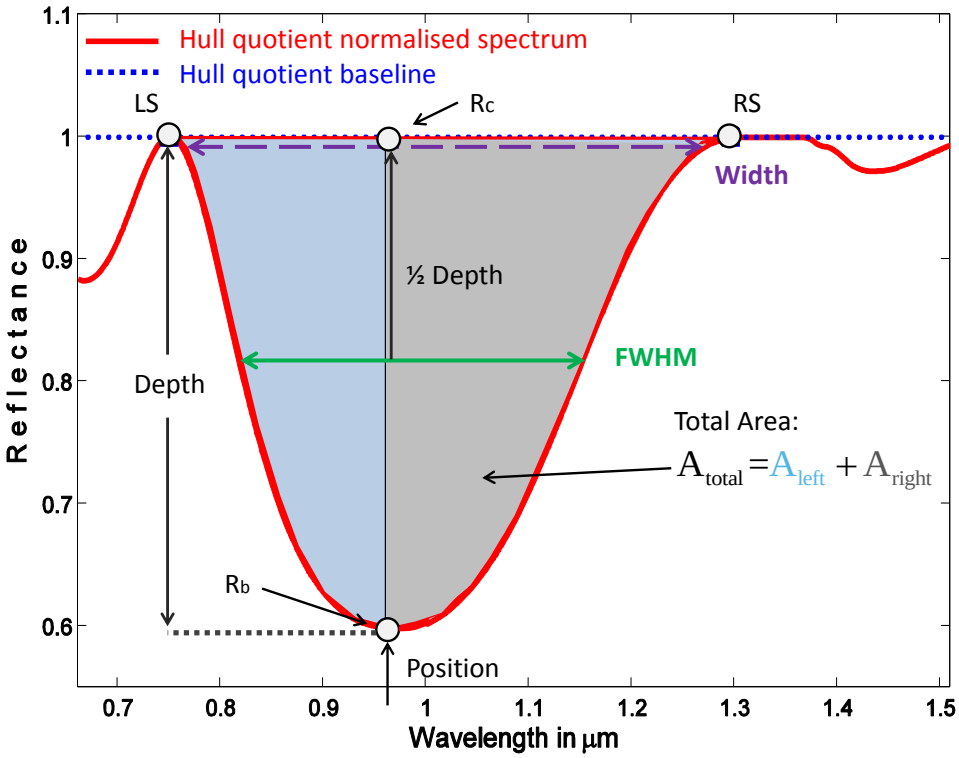


Figure 3.4 – Description of the parameters on the example of the Fe^{3+} absorption feature typical for goethite. From this figure, the following features can be obtained: position of the minimum (R_b), the depth (R_c-R_b), the positions of the left and right shoulders (LS and RS respectively), the Full Width at Half Maximum (FWHM), the shoulder width (ShWidth), the area left and right to the absorption minimum (A_{left} and A_{right}, respectively) as well as the total area of the absorption feature (A_{total}). Dividing A_{right} by A_{left} gives the asymmetry of the feature.

perfectly symmetrical about its reflectance minimum. A negative value is obtained for a left-skewed absorption features; a positive value is obtained for right-skewed absorption features, with values increasing positively with increasing right-skewness.

$$\text{Asymmetry} = \ln\left(\frac{A_{\text{right}}}{A_{\text{left}}}\right) \tag{3.3}$$

The FWHM is determined at exactly the half of absorption feature's depth. It is calculated as the wavelength on the left hand side of the absorption feature at half depth subtracted from the wavelength on the right hand side of the absorption feature at half depth. The feature's shoulder width is calculated in a similar way, except using the reflectance values at the shoulders.

A well known method which depends on the presence of diagnostic absorption features is the spectral features fitting (SFF) method described in Clark's 'Tricorder'-paper (Clark et al., 1990) which was later extended and renamed to 'Tetracorder' (Clark et al., 2003). The algorithm compares absorption features of an unknown spectrum to a known set of reference spectra and correlates the features using a least-square fit (Clark et al., 1990).

Ratios of reflectance - 'Band ratios'

Ratios of reflectance at different wavelengths have been routinely used for many years in remote sensing. Traditionally, band ratios were commonly used for analysis of multispectral images, e.g. Landsat TM or MSS images. Band ratios were developed to describe, numerically, changes in spectral curve shape between bands. At least two bands (or spectral channels) are required. Conventionally, the denominator band in the ratio is selected at the location (spectral channel) where a material is strongly absorptive and the numerator (or reference band) where the material is not absorptive. One of the advantages of ratio-analysis is that it suppresses variations in reflected light in a scene caused by topography whilst preserving changes in spectral curve shape caused by absorption (scattering of light of objects within a scene).

There are a number of categories in which the different ratios or spectral indices, which are proposed in the literature, can be classified. Vegetation indices (VI) are designed to enhance vegetation in a scene. Examples of VIs are the Ratio Vegetation Index (RVI; Jordan, 1969), Normalised Difference Vegetation Index (NDVI; Rouse et al., 1973), and the Difference Vegetation Index (DVI; Richardson and Everitt, 1992). Note that some indices can range from 0 to infinity while others (e.g. NDVI) are normalised and have a fixed range (e.g. -1 to 1). Other indices were developed to minimise soil background noise such as the Soil Adjusted Vegetation Index (SAVI; Huete and Jackson, 1988) for which some modifications exist (e.g TSAVI; Baret et al., 1989).

Band ratios were developed to map distributions of iron-oxides using MSS data from the Landsat-1 (Rowan et al. 1974, 1975, 1977). MSS bands 4/5 or their inverse (MSS 5/4) were also used by Vincent (1975) for detection of zones of iron-oxides. MSS 4/5, MSS 4/6, and MSS 4/7 ratios are related to intense ferric-absorption in the visible bands in the hematite spectrum. However, some rocks and/or minerals (e.g. clay minerals) such as montmorillonite and basalt, have small MSS 4/5 differences and, therefore, a composite of two or more ratio images was needed for discriminating among them (Rowan and Wetlaufer, 1974).

Some methods using band ratios were specifically developed for the analysis of hyperspectral data. Crowley et al. (1989) proposed the relative band-depth (RBD) analysis and applied it to Airborne Imaging Spectrometer (AIS) data. The RBD-technique uses spectral channels that closely define each absorption feature in the spectrum. In essence, this method provides a local continuum removal, removes small channel to channel radiometric offsets and the variable atmospheric absorption and solar irradiance components for each pixel spectrum. To produce an RBD image, reflectance values from several data channels from left and right absorption band shoulders are summed and then divided by the sum of several channels from the absorption band minimum. Crowley et al. (1989) combined as many as seven spectral channels to create RBD images for mineral identification. After each absorption feature in the spectrum is quantified using RBD analysis, a principle component analysis can then

be used to remove redundant information.

The main advantage of this method is that it does not require reflectance-calibrated data and thus can work on radiance spectra because it provides a local continuum removal. This method may, however, not be able to distinguish between different materials which have absorption maxima at similar wavelengths. Major disadvantages of this method are, and this is generally true for feature based methods, that i) *a priori* knowledge about the scene is required; ii) absorption features have to be present (e.g. not obliterated by noise); iii) loss of specificity may occur for minerals which have absorption features at similar wavelengths (Clark et al., 2003).

3.1.3 Methods operating on the entire spectral curve

Absorption feature based methods calculate ratios between some of the variables (spectral bands) or describe absorption features using geometric parameters (e.g. width, depth, asymmetry). In hyperspectral data, reflectance in bands adjacent to or in close proximity to each other can be highly correlated. This redundancy brings into question whether all bands are actually necessary for an effective classification algorithm (Hughes phenomena - 'curse of dimensionality', Bellman, 1957; Hughes, 1968). Some algorithms have been developed to operate on the entire spectral curve shape , i.e. using every single band in the spectrum. Thus, such methods need to use some mathematical metric which is used to determine the degree of similarity between an unknown and a reference spectrum.

Three basic metrics (i.e. similarity-measures) are introduced in this section which are most commonly used in remote sensing (Keshava, 2004) but also more widely for pattern matching and classification problems (van der Meer, 2006). These metrics are the:

1. Euclidean distance
2. Correlation metric
3. Angular metric

Euclidean distance metric

The Euclidean distance is a measure often used to determine the closeness between objects, for example, as used in the K-means classifier. If the closeness between two spectra is to be determined using the Minimum Euclidean Distance (MED) measure:

$$MED = \Delta = ||\mathbf{x} - \mathbf{x}'|| = \sqrt{\sum_{i=1}^{L} (\mathbf{x}_i - \mathbf{x}'_i)^2}\,, \tag{3.4}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two spectra, e.g. a target and a reference spectrum and Δ is the MED between points given by the two spectra of length L (Figure 3.5).

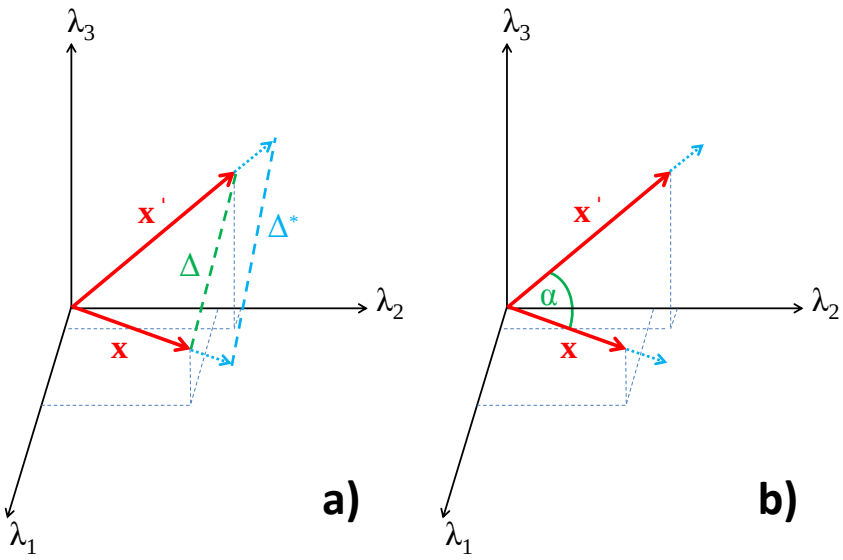


Figure 3.5 – Illustration of two points in a vector space given by the vectors $\mathbf{x}$ and $\mathbf{x}'$. $\lambda_1, \lambda_2, \lambda_3$ represent three bands of a spectrum, i.e. three elements of a vector. a) Δ is the minimal Euclidian distance (MED; dashed green line) that connects the endpoints of the two vectors. Δ^* is the MED after extending the vectors (blue arrows) which is larger than Δ. b) α is the angle between the vectors $\mathbf{x}$ and $\mathbf{x}'$. Note that the angle remains constant as $\mathbf{x}$ and $\mathbf{x}'$ are extended (blue arrows).

Correlation measure

Spectral correlation measure or cross correlogram spectral matching (CCSM; van de Meer and Bakker, 1997, 1998) is a measure of correlation describing the similarity between a target spectrum and a reference spectrum. To minimise effects of noise, the cross correlation is calculated at several positions or offsets, between the unknown and a reference spectrum. The greatest correlation across all offsets is then used as the measure of closeness between the unknown and training spectrum. The CCSM is normalised between '-1' and '1'. It is also centered around the mean of both spectra. The CCSM can be calculated using Equation 3.5

$$CCSM = \frac{\sum(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x}' - \bar{\mathbf{x}}')}{\sqrt{\sum(\mathbf{x} - \bar{\mathbf{x}})^2 \sum(\mathbf{x}' - \bar{\mathbf{x}}')^2}} \ , \tag{3.5}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are for example a target and reference spectrum and $\bar{\mathbf{x}}, \bar{\mathbf{x}}'$ their respective means. The CCSM yields a value of 1 for a perfect match (high correlation), a value of -1 for perfect inverse correlation (spectra are reverse-images of one another) and 0 for no correlation between two spectra. Van de Meer and Bakker (1998) use the root mean square error (RMSE) to create a map of correlations between an unknown spectrum's cross correlogram and the ideal cross correlogram for the reference spectrum.

Angular metric

The angle between two spectra (or vectors) is the determining measure of similarity between an unknown and a reference spectrum using an angular metric. Mathematically, a spectrum of L bands can be seen as a vector in $\mathbb{R}^L$, i.e. each spectral band can be seen as a dimension in this vector space (e.g. as shown for 2 bands in Figure 3.6). The elements of a vector also define the coordinates of a point. The distances from the origin to the endpoint of this vector vary depending on the reflectance values (brightness) of the bands. If the reflectance values of all bands vary by the same amount, the direction of a vector with respect to the vector coordinate frame remains

constant. The angle between two vectors (i.e. spectra) can thus be used as a measure of similarity between spectra even if they differ in brightness, i.e. spectral magnitude (Figure 3.5a).

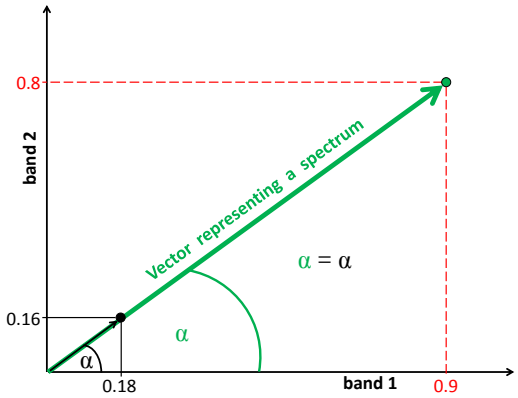


Figure 3.6 – Spectral angle concept of a 2 dimensional (2d) vector using two spectra (i.e. vectors) with two spectral bands. The black and green vector represent a dark and a bright (five times brighter) spectrum of the same material, respectively. The green dot represents a distance from the origin of the coordinate frame and describes the location of the green vector in this 2d space. The black vector has a different distance to the origin, however, the angle is the same as the green vector with respect to the coordinate frame.

Spectral Angle Mapper (SAM)

The spectral angle mapper (Kruse et al., 1993) is one of the most commonly used classical hyperspectral classifiers (e.g. Rowan and Mars, 2003; Dennison et al., 2004; Park et al., 2007; Baissa et al., 2011; Murphy et al., 2012). SAM calculates the similarity of two spectra in a high dimensional space using the spectral angle α (Equation 3.6). The main reason for SAM being very popular for classification of hyperspectral data is that it uses the entire reflectance curve and is not limited to spectra which exhibit specific absorption features. This means that SAM can be used to classify materials which have no diagnostic absorption features, including spectra where, due to effects of noise important features are suppressed, as long as the overall curve shape is more or less preserved. In addition, SAM is relatively insensitive to differ-

ences in albedo / brightness and topographic variability in a scene. This makes it ideally suited for classification of scenes with variable illumination conditions using an independent spectral library. Such libraries are acquired under controlled conditions of illumination. The brightness of a spectrum does not influence the spectral angle, i.e. the norm of a vector does not cause a change in angle between two vectors. The spectral angle α is calculated using Equation 3.6:

$$\alpha = cos^{-1}\left(\frac{\vec{\mathbf{x}} \cdot \vec{\mathbf{x}}'}{||\vec{\mathbf{x}}|| \cdot ||\vec{\mathbf{x}}'||}\right), \tag{3.6}$$

which can also be written as

$$\alpha = cos^{-1}\left(\frac{\sum_{i=1}^{L} \mathbf{x}_i \cdot \mathbf{x}'_i}{\sqrt{\sum_{i=1}^{L} x_i} \cdot \sqrt{\sum_{i=1}^{L} \mathbf{x}'_i}}\right), \tag{3.7}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two spectra (e.g. target and reference spectrum) and $||\mathbf{x}||$, $||\mathbf{x}'||$ are their respective norms (i.e. vector length). The dimensionality of the spectrum (e.g. number of spectral bands) is given by L. The spectral angle α is bound between values of -1 and 1 (or 0 and π), however, only values between 0 and $\frac{\pi}{2}$ are applicable, since reflectance cannot take negative values.

SAM is often implemented by applying a user-specified angular threshold (e.g. Dennison et al., 2004, Hecker et al., 2008). This means that if the angle between the target and reference spectrum is smaller than the defined threshold, the pixel is classified as belonging to the class of the reference spectrum (Figure 3.7). For example, a spectral angle of zero would indicate that target and reference spectrum are the same, although differences in albedo are not accounted for by SAM. Because for any particular material, the optimal threshold for SAM is rarely known *a priori*, and the threshold severely effects classification accuracy, other implementations can be used to improve scene classification (Murphy et al., 2012). For example, a 'minimum angle' criterion can be applied, whereby a target vector is compared to all reference spectra in a spectral library. A class label is then assigned by comparing the angles of all

target-reference combinations and selecting the class of the reference spectrum which has the smallest angle with the unknown (target) spectrum. This approach assumes that the mineral is present in the spectral library. Using the minimum angle always yields a classification, as there will always be a smallest angle between an unknown and a reference spectrum. Using a threshold, however, pixels may be unclassified in cases where the spectral angles between unknown and reference spectra are larger than the threshold. There is no general recommendation for the use of either approach as it is dependent on the application and the assumptions made to solve a specific problem.

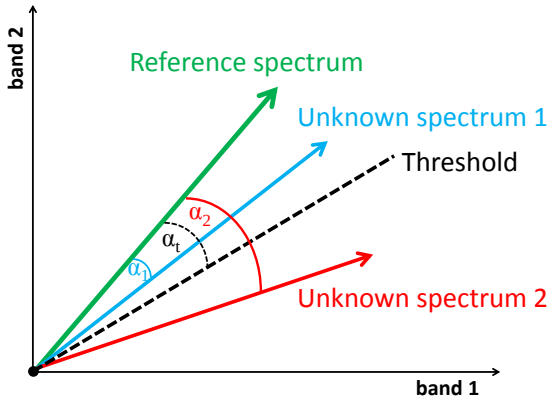


Figure 3.7 – Schematic representation of SAM in two dimensions (i.e. two spectral bands). A reference spectrum (green vector) is obtained from a spectral library and compared against two unknown spectra (blue and red arrow) using a fixed threshold (black dashed line). Spectral angles α_1 and α_2 are calculated between each reference-target spectrum combination. These spectral angles are compared against α_t which is the threshold angle.

3.2 Machine learning

Machine Learning is a concept that allows programming of machines or robots which cannot be programmed explicitly for a certain task (Bishop, 2006). For example, it may be possible to define input and output pairs but maybe not a concise relationship between inputs (e.g. reflectance values) and desired outputs (e.g. class labels, or function values). Machines need to be able to adjust their behaviour to produce correct outputs for a large number of sample inputs, thus suitably constrain their input/output function to approximate the relationship implicit in the training examples (Mjolsness and DeCoste, 2001). An example for machine learning applications is to extract important relationships and correlations in data which may be hidden among large amounts (or high dimensionality) of input data. For field-deployed robots for example, it is impossible to program or describe all possible environments which they might encounter prior to deployment. Further, it is very difficult to describe a single environment (e.g. a mine face) because it might change over time (e.g. on the scale of minutes to days).

By learning features of the environment, machines can adapt to changing environmental conditions, reducing the need for constant redesign or reprogramming (Nilsson, 1996). For example, input data (e.g. measurements) are always contaminated with random noise which introduces variability. Standard regression methods, may not achieve good fits to the data, however, with machine learning methods such noise can be modeled and in some cases, estimates of uncertainties (e.g. 2σ bounds) can be obtained. In addition, naturally occurring objects can have large amounts of variability (e.g. surface contaminants or bio-films) which can change reflectance properties of materials. Thus, changing properties may cause two objects of the same kind to appear differently (Rencz, 1999), so that traditional classification methods would categorise them as two different things (Schneider et al., 2010). Machine learning methods, are able to learn the *within-class* and *between-class* variability by learning 'hyper-parameters' which describe the trend of the data and, thus, achieve a better class separation than deterministic methods (Bishop, 2006).

There are two major types of machine learning, *supervised* and *unsupervised* learning. In supervised learning it is assumed that function values are known for the samples in the training set, at least approximately. The aim is to find a hypothesis (often denoted h) which closely agrees with the underlying function f of the training data. Curve fitting is a simple example of supervised learning. In *unsupervised* learning a training set of vectors without function values (labels) is given. Typically, the aim is to partition the training set into subsets in some appropriate way, so that the output represents the actual distribution of objects. The unsupervised learning problem is equivalent to learning a function; the value of the function, for a given sample, is the name or label of the subset of the data.

Some of the most common supervised machine learning methods used today are support vector machines (SVMs; Cortes and Vapnik, 1995; Vapnik, 1998) and Neural Network classifier (Bishop, 1995; Hastie et al., 2009). A new method which operate in a Bayesian framework, is *Gaussian Processes* (GP; Rasmussen, 2004; Seeger, 2004).

Traditional implementations of SVMs and Neural Network classifiers generally output discrete class labels, for example, class one and class zero. While this is an adequate implementation for most applications, it is regarded as a limitation for others, for example, when fusion of data (e.g. from different sensors) is required. A Gaussian Process (Rasmussen and Williams, 2006) is a stochastic process which is roughly a generalization of probability distributions over functions and allows the use of a fully Bayesian framework. It is thus often used in machine learning to solve a variety of problems (Kim and Kim, 2012; Nguyen-Tuong and Peters, 2008; Reid et al., 2011, Silversides et al., 2011; Melkumyan et al., 2012).

Each of the two learning settings, i.e. supervised and unsupervised learning, can be divided into two categories: *classification*, for discrete outputs and *regression*, for continuous outputs. These subsets are best explained using two examples.

A well known example for classification is to categorise images of handwritten digits into *discrete* categories. The training set consists of small digitised images, together with a classification from 0 to 9, normally provided by a human. A typical goal is

to learn a function from images that divides the handwritten numbers in discrete categories within the training set. This function can then be used on new, unseen images to obtain a label from an image of a handwritten digit.

A typical regression problem is a curve fitting problem (e.g. predicting housing prices; Oliveira, 2005). A more complex example in robotics is to learn the mechanical dynamics of a robot arm. The aim could be to map from the state[3] of the robot arm given by a set of variables (e.g. positions, velocities and accelerations of the joints) to the corresponding torques on the joints. Such a model could then be used to determine the torques needed to move the robot arm along a specified trajectory (Rasmussen and Williams, 2006).

3.2.1 Gaussian Processes (GPs)

Gaussian Processes (GPs) for machine learning are a supervised machine learning method (Rasmussen and Williams, 2006). Gaussian Processes can be used for either regression problems, producing continuous outputs, or classification, when discrete outputs are desired. In this section, only GPs for regression are introduced as they are subject of this thesis.

Rasmussen and Williams (2006) define a Gaussian Process as a collection of random variables which have a joint Gaussian distribution. Given a set of training data, it is possible to make predictions for new inputs (e.g. unknown test data) which are not present in the training data. This means one needs to move from a finite training set to a function which enables predictions for all possible input values. To do this, assumptions about the characteristics of the underlying function must be made, otherwise any function which describes the training data could be considered as a valid function. Two common ways of finding a predictive function are: i) restricting the class of functions which are considered; ii) assigning probabilities to every possible function, where higher probabilities are given to more likely functions representing

[3] the state in a machine learning problem can be, for example, the location of a robot or the pitch and yaw of an aircraft

the data (e.g. because of their smoothness). The limitation of the first approach is that the complexity of the function needs to be known (i.e. linear, quadratic, higher polynomial, etc.). If a certain class of functions is selected but the data is not well modeled by this function class, then predictions for new, 'unseen' test points (inputs) will be poor. If the flexibility of a certain class of functions is increased, other problems such as overfitting may occur. Overfitting means that the selected modeling function achieves a good fit to the training data but performs poorly when making predictions to new points. The second approach has the problem that there is an infinite set of possible functions that fit the test data, but an infinite number of functions cannot be calculated in finite time.

A Gaussian Process is a generalization of the Gaussian probability distribution that describes random variables which may be scalars or vectors, in the case of multivariate distributions. Gaussian Processes extend multivariate Gaussian distributions to infinite dimensionality, i.e. functions or infinite long vectors. A GP generates data located throughout some domain so that any finite subset follows a multivariate Gaussian distribution, thus inference in a GP on an infinite number of points can be done by considering only a few finite points. Such an approach gives the same answer as considering all infinite points and is consistent with any other finite query to this GP.

3.2.2 Support Vector Machines (SVMs)

Algorithms based on non-linear kernels (e.g. logistic regression, maximum likelihood classifier and GPs) need to evaluate the kernel function $k(\mathbf{x}, \mathbf{x}')$ at every possible pair of training points leading to large computational times during training and making predictions for new data points. In contrast, SVMs provide sparse solutions which means that predictions for new inputs depend only on the kernel function $k(\mathbf{x}, \mathbf{x}')$ evaluated at a subset of the training set, however, sparse solutions for GPs are also proposed in the literature (e.g. Melkumyan and Ramos, 2009). An important prop-

erty, especially in contrast to GPs, is that cost minimisation[4] (i.e. finding the 'best' model parameters) corresponds to a convex optimisation problem. This means that every local minimum is also the global minimum (Schölkopf and Smola, 2002). This section introduces the concept and a brief description of SVMs. For a deeper discussion of this topic the reader is referred to Vapnik (1998), Burges (1998), Schölkopf and Smola (2002), among others.

The SVM is traditionally a decision machine and does not provide posterior probabilities. Some benefits of having posterior probabilities are discussed in the summary of this chapter (Section 3.3). An alternative sparse kernel technique, which is based on a Bayesian formulation, is the *Relevance vector machine*. RVMs are, however, equivalent to GPs with covariance functions (Tipping, 2001; Bishop, 2006).

The contributions of this thesis are not based on SVMs, however, they are included as a method against which the algorithms developed in this thesis can be compared.

Large Margin Classifier

SVM is a large margin or maximum margin classifier. The concept of the *margin* is defined to be the smallest distance between the decision boundary and any of the samples; meaning that if multiple solutions exist which classify the data, then the one that gives the smallest generalisation error will be found (Burges, 1998). In other words, a SVM will give a decision boundary that has the smallest distance between the separating hyperplane and the training samples (Figure 3.8).

Given a two class problem and a linear model of the form

$$\hat{y}(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \tag{3.8}$$

where $\phi(\mathbf{x})$ denotes a fixed feature-space transformation, i.e. is a non linear function that maps the input space into a higher dimensional space, $\mathbf{w}$ is a weight vector which

[4] Optimising a cost function is common practice in machine learning or image processing, which is, for example, done in gradient descent methods.

is the SVM notation equivalent for θ[5] and b is a bias parameter of the hyperplane (e.g. intercept term in two dimensional space). The optimisation problem can be formulated as the minimisation of a cost function J which is defined as

$$J(w, b, r) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma\frac{1}{2}\sum_{i=1}^{n} r_i^2 \ , \tag{3.9}$$

with the regularisation term γ. The optimisation is subject to the constraints of $y_i = \mathbf{w}^T\phi(\mathbf{x}_i) + b + r_i$, with $r_i = y_i - \hat{y}_i$. Solving this optimization problem in dual representation using the kernel trick (Smola et al., 1998) leads to finding the optimal coefficients of the function

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i K(\mathbf{x}, \mathbf{x}_i') + b \ , \tag{3.10}$$

where the kernel function $K(\mathbf{x}, \mathbf{x}_i')$ is the dot product between the $\phi(\mathbf{x})^T$ and $\phi(\mathbf{x})$. The kernel trick is that the kernel function can be written as a dot product in feature space, which avoids the need to calculate the feature vectors explicitly, instead they can be replaced by kernel values. This procedure corresponds to a nonlinear classification performed in the input space. Implementations use *quadratic programming* to optimize the SVM. This means that instead of maximising $||\mathbf{w}||^{-1}$, one minimises $||\mathbf{w}||^2$ which is equivalent to each other. So the optimisation problem becomes

$$\underset{\mathbf{w},b}{\operatorname{argmin}} \frac{1}{2}||\mathbf{w}||^2 \tag{3.11}$$

3.2.3 Covariance functions

The covariance functions presented in this thesis can all be used within the GP and the SVM framework as they both comply with Mercer's condition (Schoelkopf et al., 1999). A covariance function K performs a critical task in a supervised learning

[5] In the machine learning literature θ denotes a vector of weights or hyper-parameters to be learned.

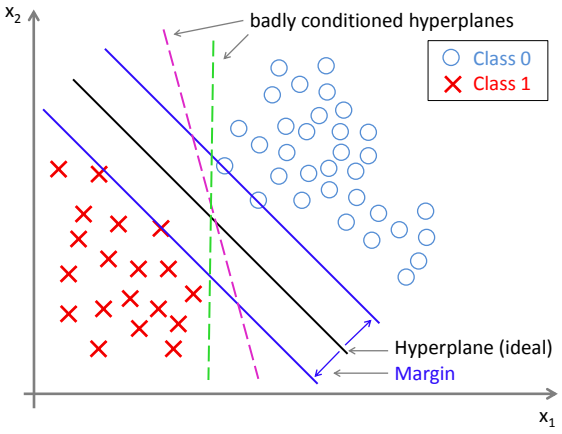


Figure 3.8 – Illustration of decision boundaries obtained using SVMs for a two dimensional data-set with $(\mathbf{x}_1, \mathbf{x}_2)$. All hyperplanes separate the two classes, however, only one of them (black solid line) gives the smallest margin between the two classes and the hyperplane and, thus, is the best decision boundary to separate the two classes.

method because it encodes the assumption about the function which is to be learned. It essentially controls how data points are related to each other or, more specifically, it encodes some aspect or measure of similarity between data points. A basic assumption is that points which are close to the inputs $\mathbf{x}$, are likely to have similar target values y, and thus training points that are near a test point should be informative about the prediction at that point.

An arbitrary function of input pairs $\mathbf{x}$ and $\mathbf{x}'$ will not generally be a valid covariance function because it may not be *positive semidefinite*. This section gives examples of some popular choices of covariance functions and briefly examines their properties.

Covariance functions can have different properties, two important ones are *stationarity* and *isotropy*. A covariance function is stationary if it is dependent on the difference between points, i.e. $\mathbf{x} - \mathbf{x}'$, thus it is invariant to translations in the input space[6]. For example, the squared-exponential (SE) covariance function (Eq. 5.11)

[6]In a statistical process there exist distinct definitions of stationarity. Generally, when dealing with groups of signals or sequences it is important to show whether these statistical properties hold true for the entire random process. A random process where all of its statistical properties do not vary when taking random subsets of the process is called *stationary*.

is a stationary kernel. A kernel which is invariant to all motions (depending upon directions) is a function of $|\mathbf{x} - \mathbf{x}'|$, and thus is isotropic (e.g. the SE kernel, Eq. 5.11). As the covariance function $k(\mathbf{x}, \mathbf{x}')$ is a function of $r = |\mathbf{x} - \mathbf{x}'|$, it is also known as *radial basis function* (RBF).

If a covariance function depends only on $\mathbf{x}$ and $\mathbf{x}'$ through $\mathbf{x} \cdot \mathbf{x}'$ it is called a *dot product covariance function* (Rasmussen and Williams, 2006). Examples of such a kernel are $k(\mathbf{x}, \mathbf{x}') = \sigma_0^2 + \mathbf{x} \cdot \mathbf{x}'$ and the polynomial kernel $k(\mathbf{x}, \mathbf{x}') = (\sigma_0^2 + \mathbf{x} \cdot \mathbf{x}')^p$ where p is a positive integer. Dot product kernels are invariant to rotation of the coordinates about the origin, but not to translations.

A few specific examples of commonly used covariance functions / kernels are presented in the following subsections.

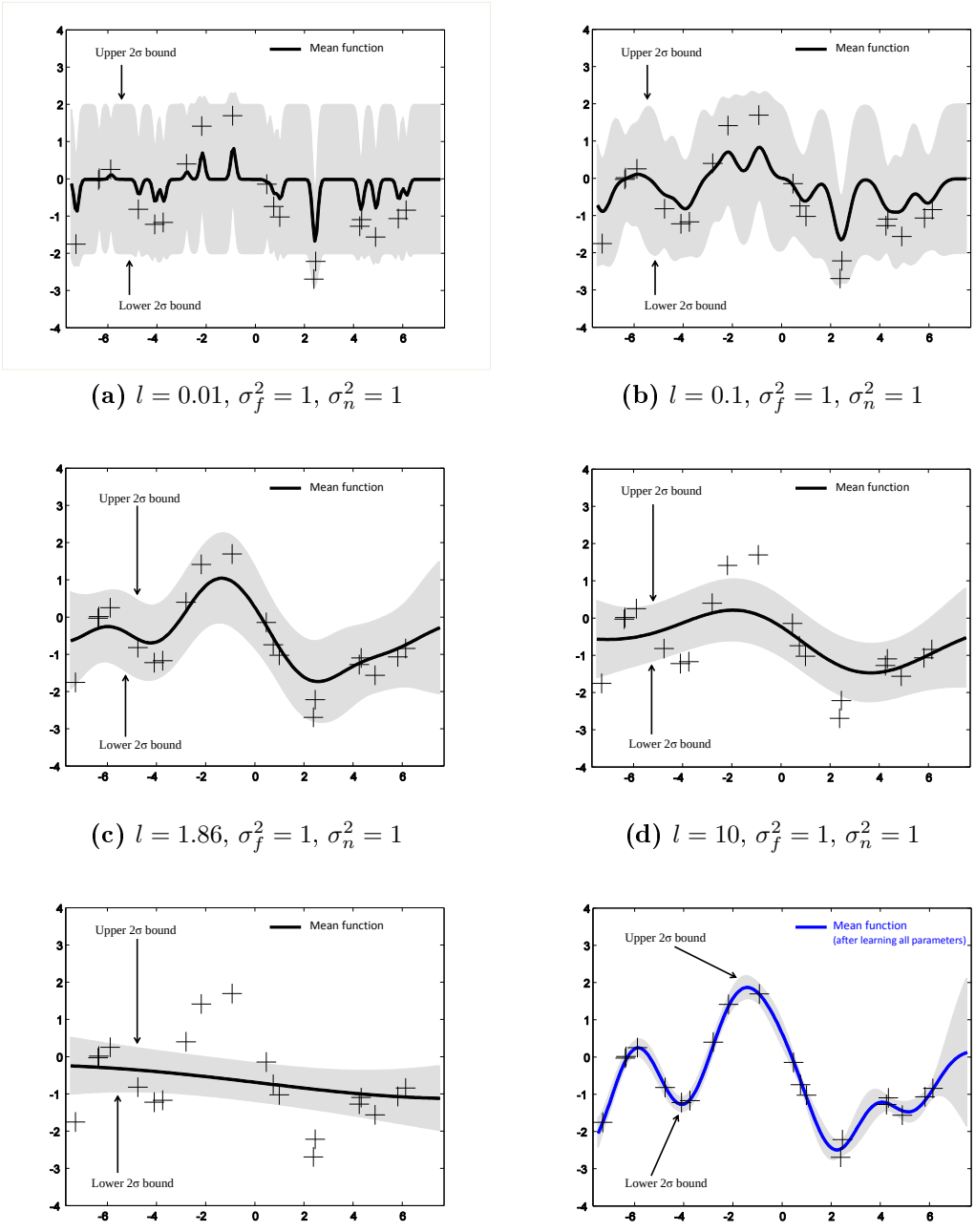
Squared Exponential - SE

The squared exponential (SE) covariance function is defined as:

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[\frac{-(\mathbf{x} - \mathbf{x}')^2}{2l^2}\right] \tag{3.12}$$

The SE covariance function is infinitely differentiable because of its exponential term. This is useful for functions which are relatively smooth, i.e. no large discontinuities. An output from a GP using this covariance function will also be quite smooth. Further, it is *stationary* which means that it is invariant against translation which can be seen from the $\mathbf{x} - \mathbf{x}'$ term. The σ_f^2 is one of the hyper-parameters of this covariance function which is the signal variance. The second hyper-parameter is the *characteristic lengthscale* l which determines how quickly the sample function can vary. If $\mathbf{x}$ is not a scalar but a vector, additional l parameters can be included to account for the variation in each direction / dimension. This can then be written as:

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \, \Omega(\mathbf{x} - \mathbf{x}')\right] \, , \tag{3.13}$$



(a) $l = 0.01$, $\sigma_f^2 = 1$, $\sigma_n^2 = 1$

(b) $l = 0.1$, $\sigma_f^2 = 1$, $\sigma_n^2 = 1$

(c) $l = 1.86$, $\sigma_f^2 = 1$, $\sigma_n^2 = 1$

(d) $l = 10$, $\sigma_f^2 = 1$, $\sigma_n^2 = 1$

(e) $l = 100$, $\sigma_f^2 = 1$, $\sigma_n^2 = 1$

(f) $l = 1.86$, $\sigma_f^2 = 1.54$, $\sigma_n^2 = 0.14$

Figure 3.9 – Example of the squared exponential covariance function used within the GP framework for a one dimensional data-set. In this sequence the signal variance (σ_f^2) and noise variance (σ_n^2) are constant while the characteristic lengthscale parameter (l) is varied. Panels a) and b) show over-fitting (too small ls) while c) shows an adequate fit. d) and e) show under-fitting (too large ls) while f) shows an optimal fit after learning all parameters. Note that c) and f) have the exact same lengthscale but different σ_f^2 and σ_n^2. Note that σ_n^2 describes the noise in the GP and this parameter is not available for use within the SVM framework. After Rasmussen and Williams (2006).

with the diagonal matrix $\Omega = \text{diag}(L)^{-2}$ containing the lengthscale parameter for each dimensions, i.e.

$$\Omega = \begin{bmatrix} l_1 & 0 & 0 & 0 & 0 \\ 0 & l_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & l_L \end{bmatrix}^{-2} \tag{3.14}$$

Neural Net covariance function

The non-stationary neural network (NNET or NN) kernel (Neal, 1996; Williams, 1998) takes the form

$$k_{NN}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \cdot \frac{2}{\pi} \sin^{-1}\left(\frac{2\tilde{\mathbf{x}}^{\mathrm{T}} \Sigma \tilde{\mathbf{x}}'}{\sqrt{(1 + \tilde{\mathbf{x}}^{\mathrm{T}} \Sigma \tilde{\mathbf{x}})(1 + \tilde{\mathbf{x}}'^{\mathrm{T}} \Sigma \tilde{\mathbf{x}}')}} \right) , \tag{3.15}$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ are augmented input vectors (each point is augmented with a 1) and Σ is $(d+1) \times (d+1)$ length-scale matrix with an additional bias term. The length-scale matrix determines how quickly the modeled function changes in each direction and d is the dimensionality of the input data. σ_f^2 is the signal variance and the set of variables $\{\sigma_f^2,\ l_1,\ \ldots,\ l_d\}$ constitute the kernel hyperparameters, where $l_1,\ \ldots,\ l_d$ are the length-scales for each of the d-dimensions of the data. The NN kernel represents the covariance function of a neural network with a single hidden layer between the input and output, infinitely many hidden nodes and using a Sigmoidal transfer function for the hidden nodes (Neal, 1996).

3.3 Summary

Some of the discussed algorithms are better suited than others for use in autonomous mining. Among the classical methods for classification of hyperspectral data, SAM, which operates on the entire curve shape, has advantages over feature based methods. This is because spectra acquired from a mine face are acquired using natural light, while spectral reference libraries are acquired using artificial light. Although data are reflectance calibrated, this change in the type of illumination causes differences in the spectral curve shape. In addition, effects of shade have profound impacts on feature based methods compared to SAM because parameters of diagnostic absorption features (e.g. symmetry, depth and width) change dramatically due to consequences of shade. SAM is more robust, although not entirely insensitive to these effects, however, it might be better suited than feature based methods for use in autonomous mining.

Machine learning methods are well suited to extract information from large amounts of data (Han and Kamber, 2006). For this reason, it can be expected that they perform well for the classification of hyperspectral data of mine face imagery. However, SVMs and GPs are merely frameworks rather than methods, as their effectiveness depends on the problem at hand and the kernel or covariance function used to solve the specific problem. For example, as it is well known in the machine learning literature, Matern kernels are unsuited to model discontinuities in data (Rasmussen and Williams, 2006), because of this, it would be impractical to 'force' this kernel onto a problem which is known to produce discontinuities in the data. Instead, one would probably chose a Neural Network (NNET) kernel which is known to perform well on data-sets with discontinuities (Neal, 1996). Similarly, for the classification of mine faces, it is reasonable to use a kernel that can handle the problems inherent with hyperspectral data. Specifically, because the limitations and capabilities of traditional classifiers for the classification of hyperspectral data are known, it is sensible to apply the same principles in the kernels used within either the SVM or GP framework.

The remainder of this thesis will focus on the two machine learning frameworks, SVMs

and GPs, combined with different kernels that use different rationales to determine similarities between data points (i.e. spectra). Comparisons between these machine learning frameworks and the spectral angle mapper are assessed in various experiments, in order to determine the effectiveness of the different methods for use in autonomous mining.

Chapter 4

Construction and development of spectral libraries for identification and mapping of geology

This chapter introduces spectral libraries which are used throughout this thesis as training and reference sets. The geographical study area from where samples and data were acquired is also introduced.

Comprehensive spectral libraries, containing several hundred minerals and materials are readily available, for example, the United States Geological Survey (Clark et al., 1993) and the ASTER spectral library (Grove et al., 1992, Baldridge et al., 2009). These libraries were, however, constructed to create a reference for a large number of different materials from relatively *pure mineral* samples, rather than from whole rocks or cores of rock. Libraries of pure, prepared minerals (e.g. cleaned, dried and ground to a standard grain size) containing spectra of specific *endmembers* of minerals, however, do not represent rocks on a mine face which are a composite of several minerals, biofilm and dust. There are, however, libraries which were also acquired from cores of rock, for example the National Virtual Core Library (NVCL; CSIRO and AuScope, 2008) which contains more than 61,000 m of scanned drill cores from across the Australian continent. This library aims to build a high-resolution miner-

alogy map of the upper 1-2 km of the Australian continent based on spectra acquired from drill cores. Spectra in the NVCL are acquired by an automated spectroscopic core logging machine ('HyLogger'™; Huntington et al., 2004, CSIRO, 2012).

Cores of rock were used to construct the spectral libraries used in this thesis. The library acquired for use in this thesis differs from existing libraries in several key aspects. First, the library is designed for a specific application - the automated recognition of ore deposits and waste materials in a specific mine. Second, the library incorporates variability with respect to illumination and viewing geometry. This is necessary because unlike airborne or spaceborne sensors, some field based remote sensing applications have to deal with complex surface geometry which is variable at relatively small scales. This is particularly evident on a mine face which can cast deep shadows at scales of centimetres to metres. From the perspective of autonomous mining, inclusion of variables such as different types of illumination and variable viewing geometry are essential to enable a seamless mapping of complex geological surfaces.

4.1 Geology of the region and the mines

In this thesis two study areas are used, the West Angelas and the Tom Price mines. Both mines are open pit mines and located in the Pilbara and lay within the Hamersley Basin Iron Province in Western Australia. The region extends roughly 80,000 km^2 and is characterised by late Archaean to Palaeoproterozoic formations. The main commodities in the Hamersley province are iron-oxides (Fe) hosted with volcanic and sedimentary sequences of the Mount Bruce Supergroup. This supergroup rests on the Pilbara Block in the far north-west of Western Australia (Thorne et al., 2008).

The Hamersley Group was formed by chemical sedimentation of minerals in a marine environment (Bekker et al., 2010). During formation, layers of rock with different chemical composition were deposited. Iron-rich layers alternate with silica-rich layers (e.g. shales) in the rock formation which are interspersed with layers of sedimentary rock (Clout and Simonson, 2005). This type of rock formation is called a Banded Iron

Formation (BIF) (Thorne et al., 2008). The Hamersley Group is approximately 2.5 km thick and contains several large units of BIF: rock with bands of iron minerals (e.g. magnetite and hematite) as well as gangue minerals, i.e. a commercially worthless material such as carbonates, silicates and chert. The most important BIF units for commercial purposes are the Brockman Iron Formation and the Marra Mamba Iron Formation (Lascelles, 2000). These layers host many of the large mineralised iron ore deposits in the Province. The Marra Mamba Iron Formation is older than the Brockman Iron Formation and is situated below the Brockman unit in the original rock sequence.

West Angelas is characterised by extensive areas of BIF which has, in some areas, become mineralised through the influence of weathering and ground-water leaching (Evans et al., 2013). In this process, silica, a major component of BIF, is leached from the rock matrix, thus concentrating deposits of iron in the form of goethite and martite (hematite). The other major rock type in this area is shale (including shale bands and extensive deposits of West Angeles Shale) and manganiferous shale, which contains large amounts of the clay mineral kaolinite and/or halloysite.

In its unmined state, mount Tom Price represented the second largest orebody with a high accumulation of high grade hematite (after Mount Newman) in the Hamersley province. The orebody consists mainly of hematite in a sequence of the Brockman Iron Formation (Morris and Ramanaidou, 2007). Most ore (around 90 %) occurs in the Dales Gorge Member of the Brockman Iron Formation. Remaining ore is associated with the Colonial Chert and Whaleback Shale Members.

4.2 West Angelas spectral libraries

A series of different libraries were created to test different hypotheses. The first set was acquired from cores of rock (10 cm wide) which were obtained using a diamond core drill. Cores of rock are produced in the mining process for exploration of the ore-body underground in order to plan the blasting and further mining processes. Drill cores should therefore be the relatively pure samples of a particular rock type

exposed on the mine face, but without any coating of dust or debris. Two sets of spectral libraries from cores of rock were created. The first set was created to be used for the development of the GP-OAD method (Chapter 5) and is used as a training set throughout this thesis. The second set was acquired from the same cores of rock as part of an extensive validation experiment. Several samples from rock cores were physically removed following acquisition of spectral data using a field spectrometer and a hyperspectral imaging spectrometer, these were then sent for XRF analysis to determine their chemistry. These analysis are used for validation of the rock types by comparing their results with the spectral response of the spectrometer and the hyperspectral imaging spectrometer.

The following libraries were created from cores of rock: The *WAcoreLib1* library was created from spectra acquired under artificial illumination and the *WAcoreLib2* library under natural illumination and at oblique angles (*WAcoreLib2a*). During the validation experiment, for each sample of rock, spectra were acquired from the core prior to sample collection and combined to create a spectral library (*WAcoreLibVal*). All spectral libraries from cores of rock are comprised of eight rock types (Table 4.1).

4.2.1 Acquisition of spectra

Artificial illumination

To provided the most accurate measurements of reflectance, a field spectrometer (Analytical Spectral Devices, Boulder, Co; ASD) was used to acquire measurements of reflectance (350-2500 nm). Spectra were acquired of homogeneous parts of the core surface, identified by visual inspection. The core surface was generally clean, smooth and free of debris, biofilm and dust. The cores were illuminated using a halogen lamp which was placed approximately $30\,cm$ away from the surface. Incident light illuminated the core surface at an angle of approximately 45° away from normal. The ASD spectrometer was fitted with an eight degree foreoptic placed in front of the optical fibre. The spectra were acquired from areas each of approximately $3.5\,cm^2$. Each spectrum represented the average reflectance of this area. The foreoptic was

Table 4.1 – List of the eight principle rock types (classes) in a WA mine. The last column indicates if the rock type is mineralised.

WA rock types	Description	M.
Banded Iron Formation (BIF)	Dark-red. Hard. High silica .	No
Chert or Cherty BIF (CHT)	Hard. Very high silica content. Light, creamy colour .	No
Martite (Hematite) (MAR)	Soft (very fragile) to hard texture. Dark brown, gray colour. High iron content.	Yes
Clay (CLY)	Soft. White to pink colour. Little to no iron, high kaolin content.	No
Goethite-Limonite (GOL)	Ochreous goethite. Chalky in appearance. Yellow colour. Medium to high iron content.	Yes
WA Shale (SHL)	Soft. Pink to yellow colour. Little iron content with clay content. Soft.	No
Shale (NS3+4)	Pink to yellow or cream colour. Kaolinitic. Volcanic shale. Used as marker horizon	No
Manganiferous Shale (SHN)	Shale with abundant manganese. Soft and light. Dark brown to black. Kaolinitic with veins of pyrolusite	No

placed perpendicular above the rock surface at an constant height of $15\,cm$ (Figure 4.1).

Care was taken not to acquire spectra of the shaded flanks of the cores. A single spectrum that was recorded from an area was itself an average of 30 individual spectra, acquired within a few seconds by the spectrometer. The spectral averaging was done to increase the signal-to-noise ratio of the spectrum. A white reference was taken immediately prior to acquiring each target spectrum. The white reference was a 100 % Spectralon panel (Labsphere, North Sutton, NH, UK) with known reflectance. Data were calibrate to absolute reflectance by dividing each spectrum by the spectrum from the reflectance panel and multiplying the result by the panel calibration factor, provided by its manufacturer.

For the *WAcoreLibVal*, acquisition of spectra was done as follows. The field spectrom-

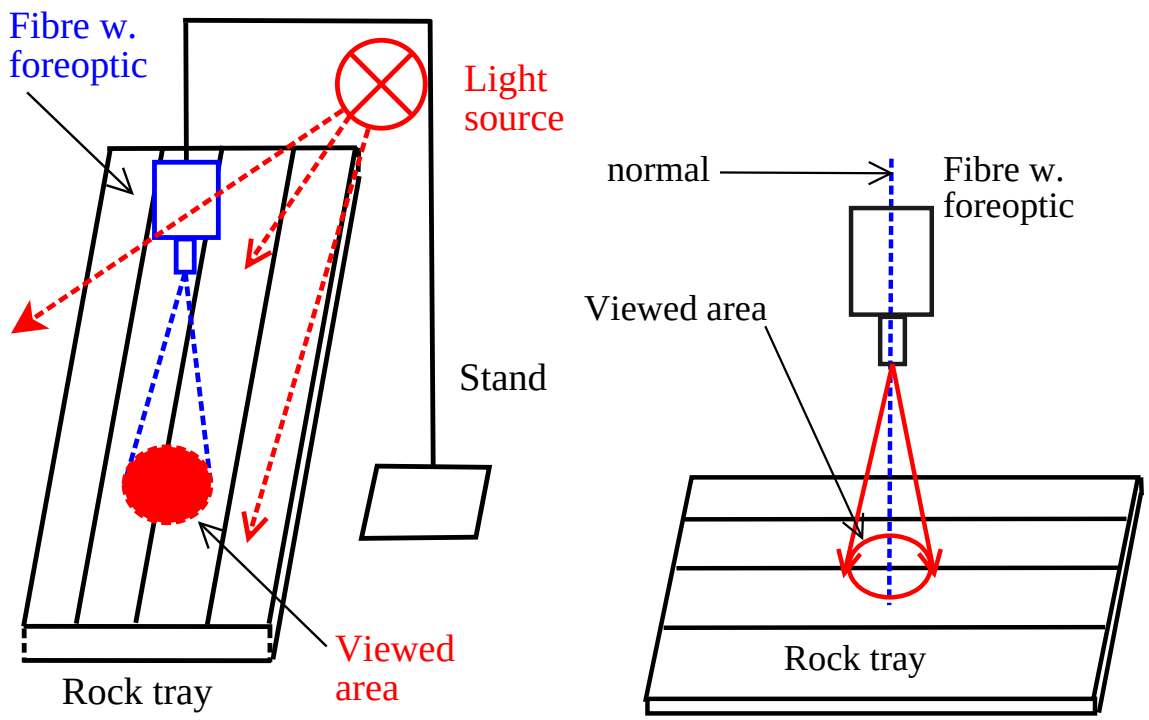


Figure 4.1 – Setup for acquisition of spectra from cores of rock in the laboratory using artificial illumination. Setup for the acquisition of spectra under laboratory conditions (a and b). (c) Example of a tray containing cores of rock. Red ellipsoids indicate different rock types; blue ellipsoids indicate different regions of a rock type. Green circles indicate positions from where replicate spectra were acquired. The area covered by the blue ellipsoid was removed for sample analysis using XRF and/or XRD.

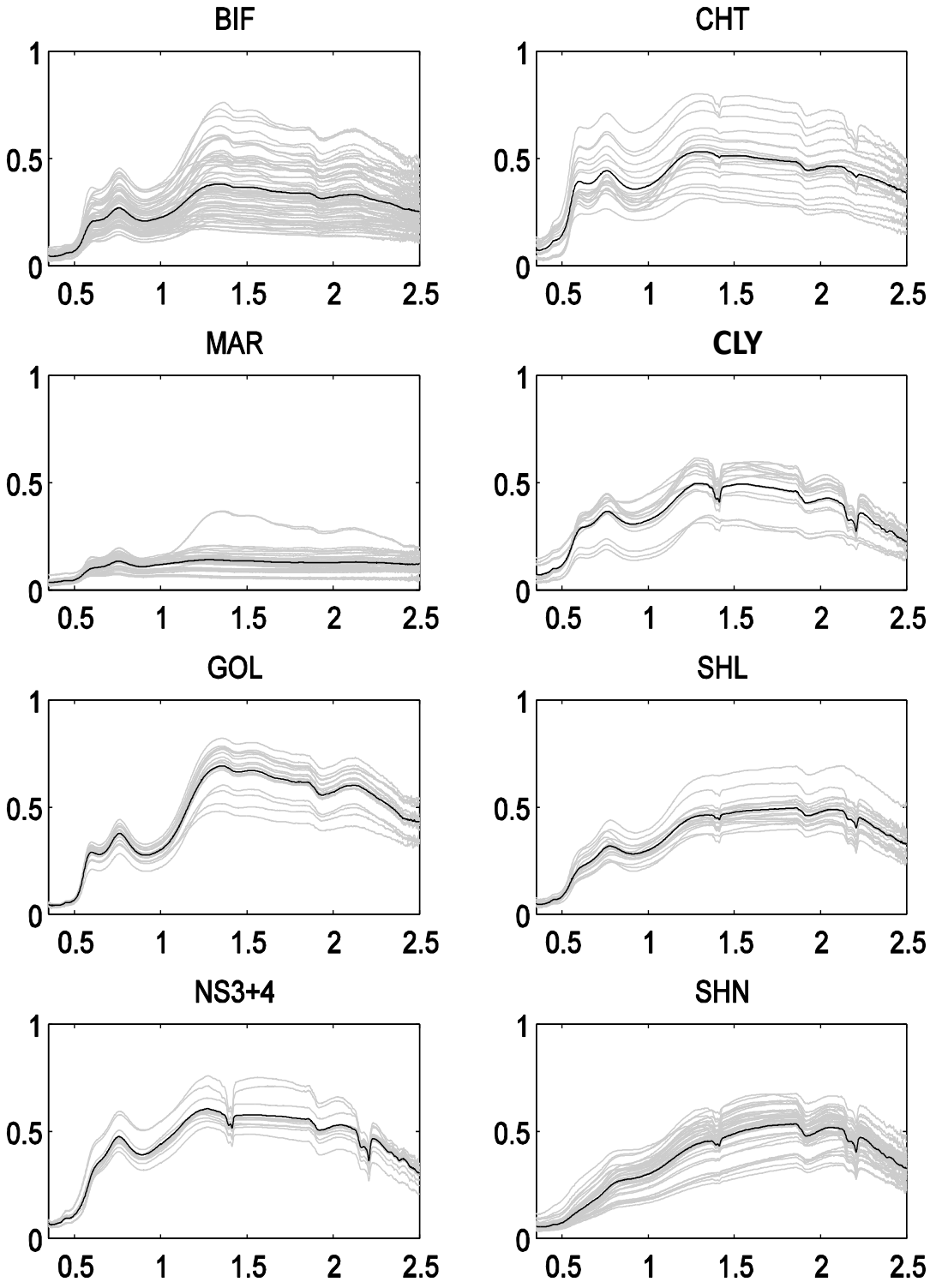


Figure 4.2 – Overview of the *WAcoreLib1* core library comprised of eight rock types. Axis labels are omitted for clarity. The abscissa is in units of wavelengths (μm) and the ordinate is in reflectance. All spectra of each rock type are shown in each panel including the mean (black).

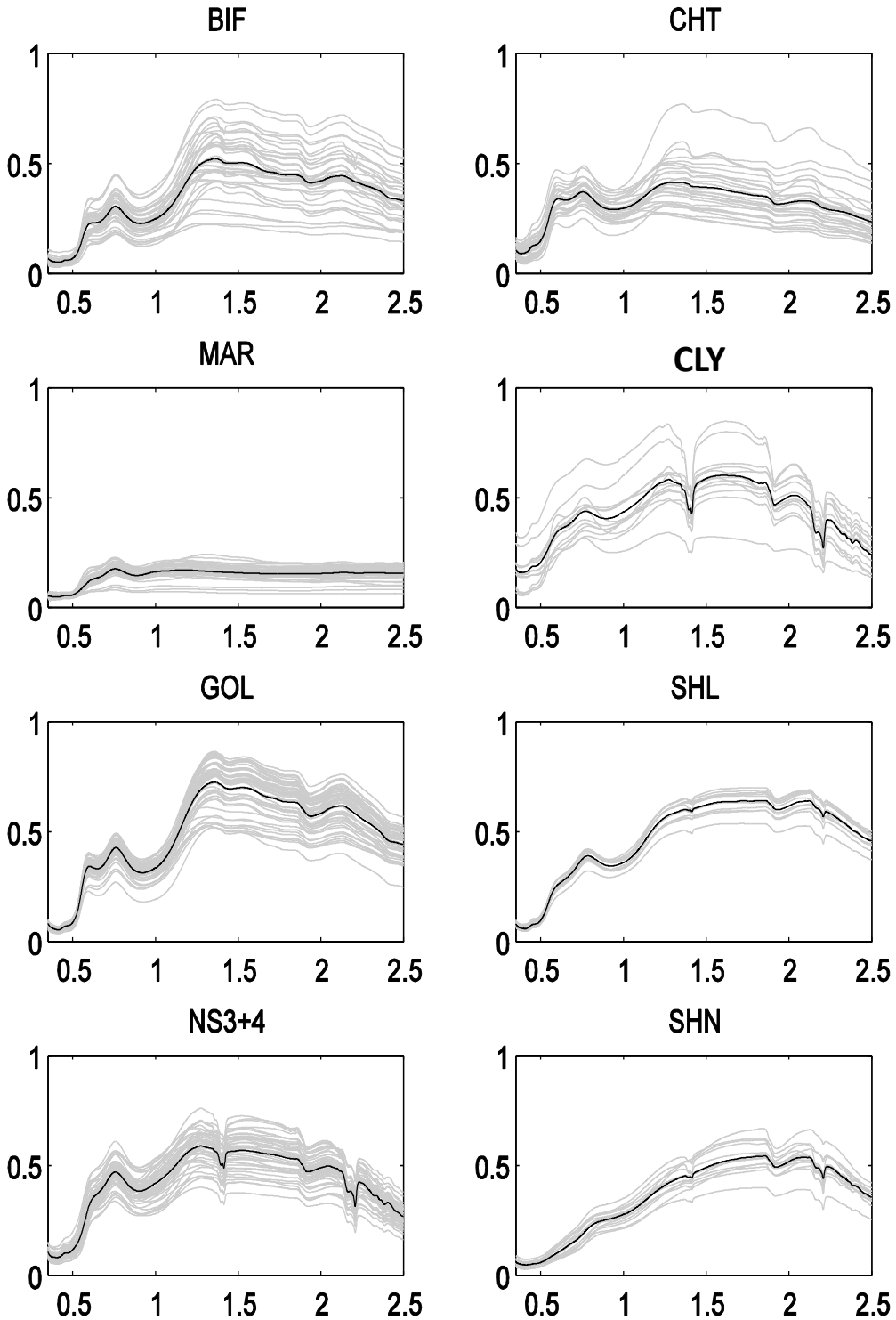


Figure 4.3 – Overview of the *WAcoreLibVal* core library, comprised of eight rock types. Axis labels are omitted for clarity. The abscissa is in units of wavelengths (μm) and the ordinate is in reflectance. All spectra of each rock type are shown in each panel including the mean (black).

eter was fitted with a reflectance probe with a high-intensity, integrated light-source. The probe measuring window was 2 cm in diameter. Acquisition was done similar to the acquisition of the *WAcoreLib1*, however, for measurement of the reference and target spectrum, the probe was placed into direct contact with the calibration panel and the target. The spectrometer was configured to average forty spectra to produce a single recorded spectrum. Between each measurement, the quartz window of the probe was cleaned using lens-wipes. Each spectrum was processed to absolute reflectance using the same method as above.

Natural illumination

There are several important differences between spectra acquired under artificial light and spectra acquired under natural light. The intensity of natural light reaching the Earth at some wavelengths around is reduced by gases and aerosols in the atmosphere, thus reducing quantities of light in these regions leading to small signal-to-noise ratios at affected wavelengths. In the spectral region between 0.35 and 2.5 μm, there are two major water absorption bands centred around 1.4 and 1.9 μm as well as two minor absorption bands occurring near 0.96 and 1.1 μm (Irons et al., 1989)[1]. Wavelengths around the two major water absorptions, i.e. around 1.4 and 1.9 μm), need to be removed. The water feature around 1.4 μm, however, contains diagnostic information about clay minerals which is lost as a consequence of natural light. Further, wavelength-dependent scattering causes shifts in the wavelength-intensity distribution of the indirect light which can cause apparent changes in colour of rocks.

The objective was to obtain spectral data acquired under natural illumination as well as to modify conditions of measurement by acquiring spectra from oblique viewing angles. This was done to introduce variability in the spectra due to naturally changing illumination conditions such as sun angle and elevation and to introduce variability due to different viewing angles. The latter was done to simulate conditions on a mine face where a hyperspectral sensor is placed in the middle of the mine face but acquires

[1] Another major water absorption band occurs just outside the the used wavelength region at around 2.7 μm.

data from a range of angles of the mine face.

To use the sun as a light source, a clear (mist free / particle free) and cloudless sky is preferable. Six spectra of each rock sample were acquired: two spectra normal to the rock surface; two spectra oblique under an angle of around 15° and two spectra oblique from the other side, i.e. −15° (Figure 4.4). Spectra acquired under natural light were measured within two hours of local noon (solar azimuth of 34.69° − 321.87° and solar elevation of 34.08° − 35.79°).

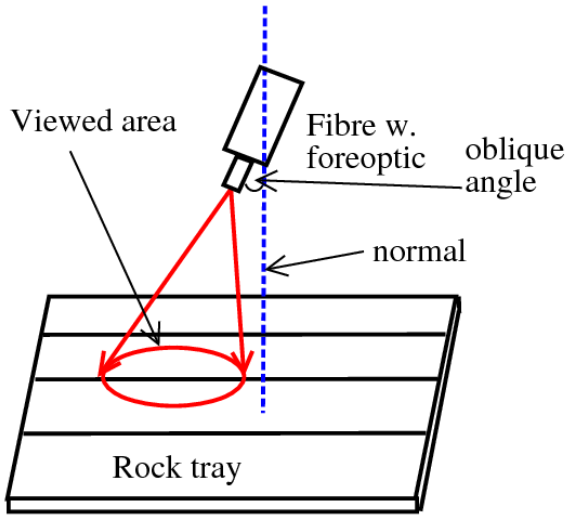


Figure 4.4 – Setup for acquisition of spectra from cores of rock using natural illumination.

4.2.2 Mineralogy and Chemistry of rocks from WA

X-ray diffraction (XRD) and X-ray fluorescence (XRF) provide information about the mineralogy and chemistry of rocks, respectively. These data were used to obtain a better understanding of the mineralogy and chemistry of samples of the *WAcoreLib1* library. To obtain this information, XRD and XRF samples were removed from the same position of the core where field spectra had been previously acquired (Figure 4.1 c), i.e. spectral measurements and XRD and XRF samples were collocated. Two to four samples per rock type were sampled from different parts of the core / cores. Ideally, XRD and XRF analysis should be done for all samples in a spectral library,

however, because these type of analysis are very expensive only a subset could be analysed.

Eight classes of rocks are used throughout this work for the analysis of WA data (Table 4.1). XRD samples were, however, divided into nine groups because shales were divided into different types. Table 4.2 shows the classes and rock types which were sampled for XRD and XRF analysis of the spectral library.

Table 4.2 – List of the nine classes used for XRD and XRF. NS3 and NS4 classes are combined to a NS3+4 supergroup in experiments throughout the thesis.

#	**Class division for XRD and XRF**	**Acronym**
1	Banded Iron Formation	BIF
2	Chert or Cherty BIF	CHT
3	Martite-Goethite	MAR
4	Clay	CLY
5	Goethite-Limonite	GOL
6	WA Shale	SHL
7	Shale	NS3
8	Shale	NS4
9	Manganiferous Shale	SHN

Results of XRF analysis show that the rock types MAR have very similar abundance of all the listed elements, although they showed variability in grain size and hardness (Table 4.3). MAR showed the highest amounts of Fe (63-68 w%; Table 4.4). The other high grade iron ore in WA is Goethite-Limonite (GOL) which has also a relatively high iron content (58% Fe). This was determined to be pure goethite without any significant amounts of hematite. Classes NS3, NS4, CLY and SHN had low amounts of Fe (5-30 w%), high amounts of clay minerals (10-35 w% Al_2O_3) and slightly higher amounts of silica (12-40 w%). QXRD determined that the clay mineral was kaolinite. Similar amounts of variability of Fe, SiO_2 and Al_2O_3 were found for classes NS3 and NS4. Because of this these two classes were grouped to the supergroup 'NS3+4'. A low-grade material in WA is the WA shale (SHL) which contains some amounts of Fe ($\sim 40 - 49$ w%), and is composed of goethite, varying amounts of kaolinite (~ 12 w%) and silica (between 9-14 w%). All clay minerals exhibit high amounts of bound water compared to the high grade iron ores which can be seen from the

loss on ignition (LOI) values (Table 4.3). Two major waste materials BIF and CHT can be distinguished from each other using QXRD and XRF data. Chert has a high silica content ($\sim$ 88 w%) and no significant iron content; BIF has a lower silica content ($\sim$ 65 w%) than CHT with increased amounts of goethite ($\sim$ 30 w%). To better visualise relationships between classes in terms of their chemistry, a principle component analysis (PCA) was applied to the data from the XRF analysis. A plot of the first two principal components which together show 99 % of the variability in the data is shown in Figure 4.5.

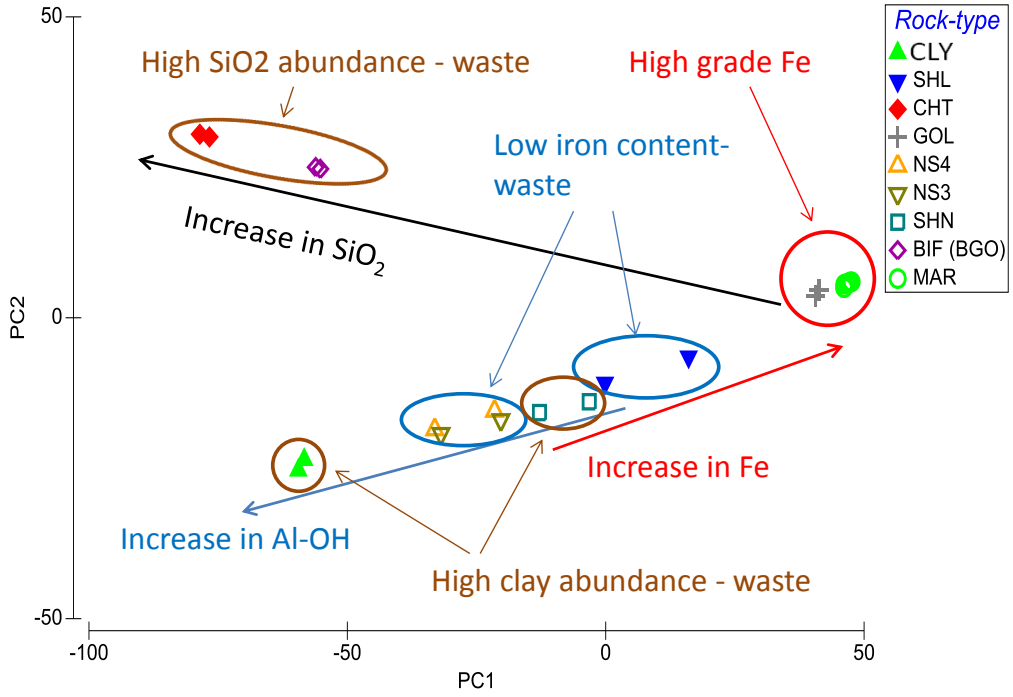


Figure 4.5 – First two principal components for the XRF analysis data. Data show the separability of the rock types using their chemistry. Arrows indicate an increase of abundances of different minerals (e.g. iron, silica and clay) for the different rock types. Rock types NS3 and NS4 are shown separately (i.e. not as supergroup NS3+4).

Table 4.3 – Overview of the chemistry (XRF) of the 24 rock samples used for validating the spectral library.

Sample	Fe	SiO2	Al2O3	MnO	LOI1000	LOI110-425	LOI425-650	LOI650-1000	Fe mean	SiO2 mean	Al2O3 mean	
N1 CLY	6.15	42.19	33.42	0.04	12.29	1.97	9.86	0.46	5.595	42.07	34.08	N1 CLY
N2 CLY	5.04	41.95	34.74	0.5	12.88	1.98	10.37	0.53				
N5 SHL	38.94	14.42	13.98	1.02	12.63	8.93	3.38	0.32	42.605	11.495	11.605	N5 SHL
N6 SHL	46.27	8.57	9.23	1.23	12.53	10.39	1.9	0.24				
N7 MAR	63.57	1.05	0.57	0.03	7.22	6.79	0.46	0.01	63.71	1.01	0.53	N7 MAR
N8 MAR	63.85	0.97	0.49	0.01	6.84	6.43	0.48	0.01				
N9 CHT	8.24	86.61	0.27	0.01	1.23	1.06	0.27	0.01	7.645	87.6	0.24	N9 CHT
N10 CHT	7.05	88.59	0.21	0.01	0.96	0.87	0.2	0.01				
N11 BIF	20.28	66.22	0.99	0.01	3.35	2.88	0.48	0.01	20.47	65.66	1.03	N11 BIF
N12 BIF	20.66	65.1	1.07	0.01	3.91	3.38	0.54	0.01				
N13 GOL	58.12	2.68	2.1	0.02	11.16	10.46	0.67	0.03	58.185	2.65	2.135	N13 GOL
N14 GOL	58.25	2.62	2.17	0.02	11.21	10.5	0.66	0.05				
N15 MAR	63.92	1.03	0.98	0.01	6.01	5.63	0.4	0.01	63.895	1.13	0.835	N15 MAR
N16 MAR	63.87	1.23	0.69	0.01	6.22	5.83	0.4	0.01				
N17 GOL	59.37	2.74	1.54	0.06	10.14	9.52	0.58	0.04	59.07	2.755	1.555	N17 GOL
N18 GOL	58.77	2.77	1.57	0.02	10.97	10.29	0.62	0.06				
N19 MAR	66.89	0.94	0.39	0.01	2.9	2.73	0.25	0.01	66.74	0.89	0.375	N19 MAR
N20 MAR	66.59	0.84	0.36	0.01	3.34	3.13	0.19	0.02				
N21 NS4	29.04	24.14	21.09	0.06	11.67	5.06	6.31	0.3	25.725	26.69	23.11	N21 NS4
N22 NS4	22.41	29.24	25.13	0.05	12.01	3.97	7.7	0.34				
N23 NS3	29.17	22.73	21.67	0.11	12.2	5.71	6.19	0.3	25.99	25.44	23.48	N23 NS3
N24 NS3	22.81	28.15	25.29	0.03	12.26	4.28	7.66	0.32				
N3 MSHL	34.25	12.05	12.17	10.8	13	8.6	3.43	0.97	32.15	14.33	13.87	N3 MSHL
N4 MSHL	30.05	16.61	15.57	8.76	13.05	7.76	4.44	0.85				

Table 4.4 – Overview of the mineralogy obtained using quantitative XRD (QXRD) analysis of the 24 rock samples used for validating the spectral library.

Sample Nx	RockTYpe	Kaolinte	Illite.Mica	Gt	Hm	Mag	Qtz	Dsp	Rt	Amorph
1	CLY	78	6	3	0	0	0	1	2	8
2	CLY	89	3	3	0	0	0	2	2	0
3	SHN	31	2	46	0	0	0	1	2	18
4	SHN	39	5	38	0	0	0	1	1	16
5	SHL	42	3	49	0	0	0	0	1	5
6	SHL	16	3	79	0	0	0	0	0	2
7	MAR	0	0	43	16	1	0	0	0	41
8	MAR	0	0	42	20	1	0	0	0	37
9	CHT	0	0	7	2	0	91	0	0	1
10	CHT	0	0	7	2	0	91	0	0	0
11	BIF	0	0	28	3	0	67	0	0	2
12	BIF	0	0	30	1	0	69	0	0	1
13	GOL	0	0	95	0	0	0	0	0	5
14	GOL	0	0	94	0	0	0	0	0	6
15	MAR	0	0	37	26	0	0	0	0	37
16	MAR	0	0	38	28	0	0	0	0	34
17	GOL	0	0	82	3	0	0	0	0	15
18	GOL	0	0	96	0	0	0	0	0	4
19	MAR	0	0	21	40	0	0	0	0	39
20	MAR	0	0	23	44	1	0	0	0	32
21	NS4	60	0	34	6	0	0	0	0	1
22	NS4	71	0	22	6	0	0	0	0	1
23	NS3	57	0	38	5	0	0	0	0	1
24	NS3	70	0	24	5	0	0	0	0	1

4.3 Summary

This chapter introduced the study area, i.e. the Hamersley Basin Iron Province and its geology. It provides the rational for constructing a new spectral library rather than using public ones (e.g. the USGS library). This is that modifications to the acquisition setup can be easily introduced and the materials represent real materials rather than pure samples of minerals which is the case in public ones. Furthermore, the constructed spectral libraries provide a principle and prior training set for the supervised classifiers used in this thesis. This training set is constructed from materials collected from several drill cores across a particular mine but is thought to be used to classify most of the mine faces within this mine.

The two spectral libraries or training sets (*WAcoreLib1* and *WAcoreLibVal*) are comprised of eight rock types (Table 4.1 and Figures 4.2 and 4.3). Two of the rock types (GOL and MAR) in this library are considered ore, and thus considered materials of interest; four classes are considered waste as they contain large amounts of clay and little or no iron ore (CLY, SHL, SHN and NS3+4); the remaining two classes (BIF and CHT) are also waste because they contain large amounts of silica and are thus difficult to mine and process.

There are spectral differences between the two ore classes, the clay-rich waste and the silica-rich waste. Both ore types have a deep and wide absorption feature around 900 nm caused by charge transfer absorptions in the ultra-violet. They are generally flat in the SWIR with no diagnostic absorption features. Rock types in the clay-rich group exhibit a diagnostic absorption feature around 2200 nm due to the presence Al-OH (e.g. kaolinite). Spectra of this group also contain diagnostic absorption features around 1450 nm and indicate large amounts of bound water. Some rock types in this group (e.g. SHL and NS3+4) also exhibit the main iron absorption feature which is also indicative of iron, however, the feature is generally less pronounced compared to the two ore rock types. BIF and CHT are two classes with high abundance of silica and are therefore relatively bright in the SWIR; both rock types have a large within-class variability. Some spectra of BIF are very similar to spectra of GOL. Some CHT

spectra show absorption features at 1450 nm and 2200 nm, however, no indication of Al-OH or kaolinite was found in neither XRF nor XRD analysis, respectively. A reason for the presence of these two absorption features in the *WACoreLib1* data set might be coatings of dust from materials which do contain water and / or Al-OH. Another reason for this might be that the field of view of the spectrometer foreoptic was larger than the rock sample for some samples of CHT, thus leading to the acquisition of spectra from surrounding materials. Both explanations seem reasonable because neither the water feature near 1450 nm nor the Al-OH features near 2200 nm are present in CHT spectra in the *WACoreLibVal* data set. Unlike *WACoreLib1* data set, this library was acquired by bringing the measuring probe in direct contact with the rock sample, thus avoiding accidently acquiring spectra from surrounding materials.

This chapter provides a spectral, chemical and mineralogical overview of the different rock types used in this thesis and highlights possible problems for classification due to spectral similarities and within-class variability of some classes. References to the rock types and their acronyms presented in this chapter (Table 4.1) will be made frequently and throughout this thesis.

Chapter 5

A probabilistic framework for identification and mapping of geology using hyperspectral data

5.1 Probabilistic versus deterministic spectral angle mapper

The main contribution of this thesis was the development of a method for improving the identification and mapping of geology on vertical mine faces. For this, a key requirement was that a method must operate in a fully probabilistic (i.e. Bayesian) framework to enable a probabilistic representation of the geology. A probabilistic representation of geology in this context means that the output of a classifier does not only provide a class prediction or an estimate for a class of rock but also gives a measure of uncertainty for any given estimate. Using the estimate and the uncertainty, a probability can be associated with any given prediction. Such a framework provides capabilities for fusion of data from different sensors and aids in the construction of a comprehensive and probabilistic picture of the mine. To address these requirements a new method for the classification of hyperspectral data is hereby proposed. The

proposed method, named *GP-OAD*, was developed for automated identification of types of geology using data from non-imaging and imaging spectrometers. The proposed GP-OAD is, however, not limited to geological applications. The GP-OAD is based on Gaussian Processes for machine learning (Rasmussen and Williams, 2006) and a specific Observation Angle Covariance function (Section 5.2.2). The GP-OAD applies the OAD covariance function to hyperspectral data for the first time.

For the analysis of hyperspectral data, a large number of deterministic methods are available and commonly used in remote sensing (reviewed in section 3.1). A deterministic method in this context is defined as a method that yields a class label for a given test spectrum but does not provide any measure of uncertainty. Many deterministic methods require diagnostic absorption features to be present in the spectrum. For example, the SFF method (Section 3.1.2) works well when classifying spectra of minerals that have very distinct and diagnostic absorption features (e.g. kaolinite or alunite). Some spectra from geological materials do not exhibit distinct or diagnostic absorption features (Gaffey and McCord, 1977). For spectra without diagnostic absorption features, however, SAM (Section 3.1.3) is often more effective for classification because it considers the whole spectral curve shape which may be typical (although not diagnostic) of lithologies or bulk mineralogy. This is a great advantage over feature based methods as there are no guaranties for the presence of diagnostic absorption features neither are there guaranties for high quality spectra with high signal-to-noise ratios which is an important factor for feature-based methods (Swayze et al., 2003). This forms part of the rationale behind the development of the GP-OAD method which uses SAM at its core.

But why should the GP-OAD perform better than the traditional implementation of SAM if it uses SAM at its core? Gaussian Processes allow the correlation of observations (i.e. spectra from a training set or a spectral library) with each other which is the fundamental principle of GPs and presents an advantage over SAM. What is equally important is the way correlations between observations are created. For example, in linear regression, a common way to determine the 'goodness of fit' of an estimate is to calculate the squared distance between observations and an estimate.

Generally, the larger the squared distance, the less correlated are the observations with the estimates. In the specific case of the GP-OAD, correlations are calculated using the spectral angle. The major difference between SAM and the GP-OAD is that correlations are created between all observations, i.e. between all spectra in the training set. This is very different to simply comparing spectral angles between pairs of spectra against a predefined threshold. Further, GPs allow to model the amount of noise in the data independently by learning the amount of noise from the data themselves. This is a large advantage compared to deterministic methods which do not allow the amount of noise to be learned or modeled from the spectral signal. As a consequence, confusion between similar signals (i.e. similar types of rock) occur much more readily in SAM than the GP-OAD as amounts of noise increase.

These characteristics make the GP-OAD very attractive for environmental remote sensing applications. In the specific context of autonomous mining, the GP-OAD has further advantages compared with deterministic methods: (i) The probabilistic framework of the GP-OAD allows fusion of information from various sensors in a systematic and consistent fashion, (ii) separation of ground cover materials (e.g. classes of rock) is possible even for featureless signals and (iii) it provides a measure of uncertainty for each estimate (e.g. a class of rock).

This chapter provides the formulation of the GP-OAD in the context of the GP framework. The GP-OAD is first tested on hyperspectral data from non-imaging spectrometers (this chapter) and imaging spectrometers (Chapter 6). The GP-OAD is compared to other methods to assess its performance against standard methods.

5.2 The GP-OAD method for classification of hyperspectral data

5.2.1 Gaussian Processes

In general, an input (e.g. a spectrum) can be defined as $\mathbf{x}$, and the output (or target) as y. The input observation $\mathbf{x}$ is usually a vector, i.e. $\mathbf{x} = [x_1, x_2, \ldots, x_L]$, comprised of L variables, thus $\mathbf{x} \in \mathbb{R}^L$ and $L \in \mathbb{N}$. The target y may either be continuous $y \in \mathbb{R}$ (regression) or discrete $y \in \mathbb{Z}$ (classification). A dataset of n observations is denoted by $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, \ldots, n\}$ which can be simplified to $\mathcal{D} = (X, \mathbf{y})$, where X is a (design-) matrix of n observations $X = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}]$ and $\mathbf{y}$ is a one dimensional target vector with n values, i.e. $\mathbf{y} = [y_1, y_2, \ldots, y_n]$. In the context of a supervised learning problem, $\mathcal{D}$ is the training set with n training samples (observations) and the associated class labels. In this case, X is an aggregate of the L dimensional input vector (e.g. a spectrum with L bands), thus, throughout the rest of this thesis, X will be arranged as a matrix of size $L \times n$.

A Gaussian Process can be defining as a distribution over the space of function variables $f(\mathbf{x})$ mapping input to output space. The objective is to compute the predictive distribution $f(\mathbf{x}_*)$ at a new test point $\mathbf{x}_*$. Inference can be done directly in the space of functions, the *function-space* (Seeger, 2004). Gaussian Processes are non-parametric, thus there are no specific basis functions with a set of parameters defining how inputs relate to outputs. Gaussian Processes can be fully specified by a mean function ($m(\mathbf{x})$) and a covariance function ($k(\mathbf{x}, \mathbf{x}')$).

The mean function can be defined as:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{5.1}$$

The covariance is defined as:

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \tag{5.2}$$

From these equations an equivalent form of the $\mathcal{GP}$ can be written as:

$$f(\mathbf{x}) = \mathcal{GP}\ (m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))\ \ , \tag{5.3}$$

where $\mathbf{x}$ may be scalar or multidimensional.

In some applications, the mean function does not serve any purpose, the best model can then be obtained by the training data and the covariance function. An advantage of GPs is that no prior basis function needs to be defined, thus the mean function can be set to zero, i.e. $m(\mathbf{x}) = 0$. The training data is incorporated by forming a prior or joint distribution (Equation 5.4) between the training data points X and $\mathbf{y}$ and a set of inference points (test set), denoted by X_*. In supervised learning, data is generally partitioned into two sets, a training and a test set. These sets are denoted by $(X, \mathbf{f}, \mathbf{y}) = (\{\mathbf{x}_i\}, \{f_i\}, \{y_i\})_{i=1}^n$ for the training set and $(X_*, \mathbf{f}_*, \mathbf{y}_*) = (\{\mathbf{x}_{*,i}\}, \{f_{*,i}\}, \{y_{*,i}\})_{i=1}^n$ for the test set. The joint distribution is defined as:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right). \tag{5.4}$$

$\mathcal{N}(\mu, \mathbf{\Sigma}))$ is a multivariate Gaussian distribution with mean μ for the output function $\mathbf{f}_*$ and covariance $\mathbf{\Sigma}$. K is the covariance matrix between all points in the data-set, i.e. $K(X,X)$ corresponds to the covariance between the training points, $K(X,X_*)$ and $K(X_*,X)$ corresponds to the covariance between the training and test set and $K(X_*,X_*)$ is the covariance between the test points.

Equation 5.4 is only true for measurements which are free of noise. In reality, however, every measurement has some noise or uncertainty associated with it. Thus, a noise parameter needs to be included so that the general linear regression model can be written as $\mathbf{y} = f(\mathbf{x}) + \epsilon$ which assumes a Gaussian noise distribution $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$, where σ_n^2 is the signal variance (e.g. random sensors noise in spectral measurements). Thus, the joint Gaussian distribution can be extended with the signal variance term

and becomes:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right). \tag{5.5}$$

Inference for new points

Inference is done to predict labels y for new test points (e.g. unknown spectra) which might be acquired using a spectrometer of some kind. By conditioning the prior distribution on the observed training points, the predictive or *posterior* distribution for new points can be obtained using the Bayes rule, so that

$$p(f_*|X_*, X, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu_*}, \boldsymbol{\Sigma}_*) \tag{5.6}$$

where

$$\boldsymbol{\mu_*} = K(X_*, X)[K(X,X)]^{-1}\mathbf{y}\,, \tag{5.7}$$

and

$$\boldsymbol{\Sigma}_* = K(X_*,X_*) - K(X_*,X)[K(X,X)]^{-1}K(X,X_*)\ . \tag{5.8}$$

Including the noise associated with the training points, the mean and covariance can be written as:

$$\boldsymbol{\mu_*} = K(X_*, X)[K(X,X)+\sigma_n^2 I]^{-1}\mathbf{y}\,, \tag{5.9}$$

and

$$\boldsymbol{\Sigma}_* = K(X_*,X_*) - K(X_*,X)[K(X,X)+\sigma_n^2 I]^{-1}K(X,X_*) + \sigma_n^2 I\ , \tag{5.10}$$

in order to use the additional parameter σ_n^2 to account for the error in measurements.

Equation 5.9 is a linear combination of N kernel functions each centred on a training point, $\boldsymbol{\mu}_* = \sum_{i=1}^{N} \alpha_i k(\mathbf{x}_i, \mathbf{x}_*)$, with $\boldsymbol{\alpha} = (K(X,X)+\sigma_n^2\ I)^{-1}\ \mathbf{y}$. Thus, a GP can be

viewed as an unbiased linear estimator (Cressie, 1993). During inference, most of the computational cost occurs during the inversion of the covariance matrix (Equation 5.10), which is in the order of $\mathcal{O}(N^3)$.

Learning in a Gaussian Process framework

The covariance function $k(\mathbf{x}, \mathbf{x}')$ is an important part of a Gaussian Process. In a similar way to how the variance controls the width of a Gaussian distribution, the covariance function affects the properties (e.g. smoothness) of a GP. The problem of learning in Gaussian Processes is exactly the problem of finding suitable properties for the covariance function which it is parameterised by its hyper-parameters $\boldsymbol{\theta}$, i.e. $k(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$. The covariance function and its properties control how data points are related to each other, i.e. the characteristics of the model (e.g. smoothness and degree of variation). To obtain an accurate model of the data, it is important to choose the 'correct' (or appropriate) covariance function. The 'correct' covariance function is the one that captures the uncertainty in the data, is unbiased and can capture the variations in the data (Rasmussen and Williams, 2006). As a consequence, the output function should be smooth and should follow the *trend* of the data, rather than fitting every point - which can lead to overfitting (e.g. as in Figure 3.9).

For the purpose of discussing the learning process in a GP, the squared exponential covariance function will be used (Section 3.2.3). It is re-defined from equation 3.12:

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[\frac{-(\mathbf{x} - \mathbf{x}')^2}{2l^2}\right] \tag{5.11}$$

The SE kernel[1] has L+1 hyper-parameters (i.e. two hyper-parameters for 1-dimensional problems, $L = 1$) which can be learned from the training data. These hyper-parameters, in the case of the SE kernel (for 1D), are σ_f^2 and l. In the machine learning literature, the notation for a set of hyper-parameters is denoted by $\boldsymbol{\theta}$. Using this notation, the set of hyper-parameters for the SE kernel is defined as $\boldsymbol{\theta}_{SE} = [l, \sigma_f^2, \sigma_n^2]$,

[1]Note that *kernel* will be used as a synonym for covariance function from this point onwards

where the kernel descriptor $_{SE}$ is often omitted. The third parameter σ_n^2 is the signal noise component (Equation 5.5) and is independent of the covariance function (Seeger, 2004; Rasmussen and Williams, 2006).

Mathematically, the learning in a GP framework is done by optimising the set of hyper-parameters (i.e. finding optimal values for the elements of the vector $\boldsymbol{\theta}$) to find the distribution of functions which best fits the target $\mathbf{y}$ of the training data, given the observations of the training data (X). In a Bayesian framework, this can be done by maximising the log of the marginal likelihood with respect to $\boldsymbol{\theta}$:

$$\begin{aligned} \log p(\mathbf{y}|X,\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T[K(X,X)+\sigma_n^2 I]^{-1}\mathbf{y} \\ -\frac{1}{2}\log|K(X,X)+\sigma_n^2 I| \\ -\frac{N}{2}\log 2\pi \end{aligned} \tag{5.12}$$

To explicitly state that the optimisation is done for the training set (X and $\mathbf{y}$), the latter equation can be simplified by substitution $K_{\mathbf{y}} = K(X,X)+\sigma_n^2\, I$, which gives

$$\log p(\mathbf{y}|X,\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T K_{\mathbf{y}}^{-1}\ \mathbf{y} - \frac{1}{2}\log|K_{\mathbf{y}}| - \frac{N}{2}\log 2\pi\ . \tag{5.13}$$

The three terms on the right side of (5.13) are (from left to right) the data fit term, complexity penalty term and a normalization constant. The first term specifies the best fit of the data with the hyper-parameters; the second term penalises overly complex data fits (i.e. reflects Occam's Razor principle; Kolmogorov, 1965, McCall, 2004, Hutter, 2006), and thus reduces problems caused by overfitting; the third term is a normalisation constant. Equation 5.13 is a non-convex function and therefore only local maxima can be found. Optimising this function is non-trivial. Optimisation can, however, be done using numerous multivariate optimization algorithms (e.g. conjugate gradients, Nelder-Mead simplex) which are gradient descent based techniques. A review of non-convex optimisation techniques is provided in Nocedal and Wright (2006). Such methods require computation of partial derivatives, such as:

$$\frac{\partial}{\partial \theta_j} \log p(\mathbf{y}|X, \boldsymbol{\theta}) = \frac{1}{2}\mathbf{y}^T K_{\mathbf{y}}^{-1} \frac{\partial K_{\mathbf{y}}}{\partial \theta_j} K_{\mathbf{y}}^{-1}\ \mathbf{y} - \frac{1}{2}\mathrm{tr}\left(K_{\mathbf{y}}^{-1}\frac{\partial K_{\mathbf{y}}}{\partial \boldsymbol{\theta}}\right). \tag{5.14}$$

In addition, Equation 5.14 also requires the computation of the partial derivatives of the covariance function with respect to $\boldsymbol{\theta}$. To increase the chance of convergence to the global maximum, such optimisers can be applied several times from multiple starting points, i.e. different start values for $\boldsymbol{\theta}$ (Melkumyan and Ramos, 2009).

5.2.2 The Observation Angle Dependent (OAD) covariance function

The Observation Angle Dependent covariance function (OAD; Melkumyan and Nettleton, 2009) is less known in the machine learning literature, although, it was demonstrated that it outperforms more classical covariance functions (e.g. Neural Net, Squared Exponential and Matern) with respect to Mean Square Error (MSE). The OAD depends, not on the difference $\mathbf{x} - \mathbf{x}'$, but on the spatial location of the points $\mathbf{x}$ and $\mathbf{x}'$ with respect to the observation centre $\mathbf{x}_c$, thus the OAD kernel is *non-stationary*. For the application presented in this thesis, however, the OAD covariance function does not take into account[2] the observation center $\mathbf{x}_c$ and thus can be simplified from its original form. The OAD is then defined as:

$$k_{OAD}(\mathbf{x}, \mathbf{x}') = \sigma_0^2 \left(1 - \frac{1 - \sin\varphi}{\pi}\alpha(\mathbf{x}, \mathbf{x}')\right)\ , \tag{5.15}$$

where σ_0, φ are scalar hyper-parameters of the covariance function and $\alpha(\mathbf{x}, \mathbf{x}')$ represents the angle between two L dimensional points $\mathbf{x}$ and $\mathbf{x}'$. The observation angle $\alpha(\mathbf{x}, \mathbf{x}')$ is calculated using the dot product (Equation 3.6) and can thus be written as:

$$k_{OAD}(\mathbf{x}, \mathbf{x}') = \sigma_0^2 \left(1 - \frac{1 - \sin\varphi}{\pi} \cdot \cos^{-1}\left[\frac{\mathbf{x} \cdot \mathbf{x}'}{||\mathbf{x}|| \cdot ||\mathbf{x}'||}\right]\right)\ , \tag{5.16}$$

[2]The observation centre is the same for a reference and a target spectrum, i.e. the origin in a vector space at Cartesian coordinates (0,0), and thus $\mathbf{x}_c$ is set to zero.

where $||\mathbf{x}||$ is the norm of the vector $\mathbf{x}$. The angular similarity measure $\alpha(\mathbf{x}, \mathbf{x}')$ within the OAD kernel represents (in the context of the thesis) the spectral angle which is described in section 3.1.3 and defined in Equation 3.6. The GP-OAD therefore incorporates SAM at its core.

5.3 Implementation of the GP-OAD

5.3.1 Training the GP model

Learning a model in a GP is equivalent to learning the hyper-parameters which control the GP (Williams and Rasmussen, 1996). The hyper-parameters can be estimated from the data using the maximum likelihood approach (Equation 5.13). The hyper-parameters are initialized with random numbers and then used in a gradient descent based method to search for the optimal values for the hyper-parameters (Snyman, 2005). To avoid becoming stuck in a local minimum, the search step is repeated several times with different random starting points. After this step, the best parameter set is selected by comparing the magnitude of the log marginal likelihood for each starting point and selecting the one with the largest value of the log marginal likelihood.

5.3.2 Prediction of classes using the GP model

The predictive distribution for a test sample x_* is obtained from the joint Gaussian distribution of the training samples and the test samples by conditioning on the observed targets in the training set. Generally, the predictive distribution is Gaussian with mean and covariance given by Equations 5.9 and 5.10. The prediction step yields the most likely label for a sample x_* with some uncertainty around it.

5.3.3 Classification of predictions to class labels

So far the GP framework is described in the context of a regression problem, where the predictions are real numbers. However, both, classification and regression in the context of GPs can be viewed as *function approximations* (Rasmussen and Williams, 2006). In other words, the GP yields a guess for a class label and does not take the form of *class predictions*, i.e. no discrete class labels are predicted using the regression framework (Williams and Rasmussen, 1996). The set of numbers for the predictions and the targets are different, i.e $y \in \mathbb{Z}$ and $p(f_*|X_*, X, \mathbf{y}) \in \mathbb{R}$; therefore, a class or decision boundary needs to be defined. Real valued predictions can be used to do classification in discrete sets $\mathbb{Z}$. This is simply done by defining the boundary as the equal distance from both class labels $C_{1,2} = \{-1, +1\}$ which yields a decision boundary of zero:

$$C = \begin{cases} -1, & \text{if } \mathbb{E}(f_*|X_*, X, \mathbf{y}) < 0 \\ 1, & \text{if } \mathbb{E}(f_*|X_*, X, \mathbf{y}) \geq 0 \end{cases} \tag{5.17}$$

where $\mathbb{E}(f_*|X_*, X, \mathbf{y})$ stands for the expected value for f_*. The described classification method was used to perform the cross-validation experiment (Section 5.4.1). This was done so that spectra could be assigned to a discrete class to generate confusion matrices to assess the performance of the algorithm.

5.4 Assessment of the GP-OAD and comparison with other methods

To assess the performance of classification of the proposed method, it was applied to hyperspectral data from a field spectrometer and compared against standard methods. The first method used for comparison, the Spectral Angle Mapper, was selected because it is a standard method and at the same time the GP-OAD uses the same angular measure at its core. SAM has a long history of applications for classification of hyperspectral data (e.g. Dennison et al., 2004; Park et al., 2007), it was therefore important to assess how a traditional, deterministic method would measure against

the GP-OAD. One of the most commonly used methods for classification tasks in a variety of fields are SVMs (Xu et al., 2003; Zhao and Takagi, 2007; Braun et al., 2011; Wang, 2012). Because of this and the increasing use of SVMs for hyperspectral data classification, it was considered important to compare the GP-OAD also against SVMs. For the two learning frameworks, GPs and SVMs, the same covariance functions were used; these were the OAD (as part of the GP-OAD method), the squared exponential and the neural net covariance function (Rasmussen and Williams, 2006). The latter two kernels were used because both are widely used kernels (Vasudevan, 2012). These methods, the GP-OAD, GP-SE, GP-NNET, SAM, SVM-OAD, SVM-SE and SVM-NNET are used throughout most experiments in this thesis to maintain a set of standard methods for comparison. This framework also allows the strengths and weaknesses of a method under various experimental conditions to be assessed.

In order to validate the performance of the different classifiers, field spectra of consistent high quality, obtained using an ASD field spectrometer were used. This was done so that any differences in classification performance could be attributed to the classifiers themselves and not to the 'quality' of the data (i.e. high signal-to-noise ratio).

For the classification of each rock type, all spectra within the training and test set which were of the rock type being classified, were labeled as '-1', other spectra from all other rock types were labeled '1' (Section 4.1). The aim was to identify all samples of classes '-1' and '1' correctly. This is a binary approach of classification, opposed to a multi-class approach, and is known as 'One versus All' (OvA; illustrated in Figure 5.1), also called 'One versus Rest', and requires that any algorithm is applied n times for n different classes in a data-set (Rifkin and Klautau, 2004). For example, in the case of the *WAcoreLib1*, eight groups of rock were present, thus the binary algorithm had to be applied eight times to classify all different rock types. Because all the target samples are known in this experiment, it is possible to construct confusion matrices from the numbers of true-positive, false-positive, true-negative, and false-negative classifications for each rock type.

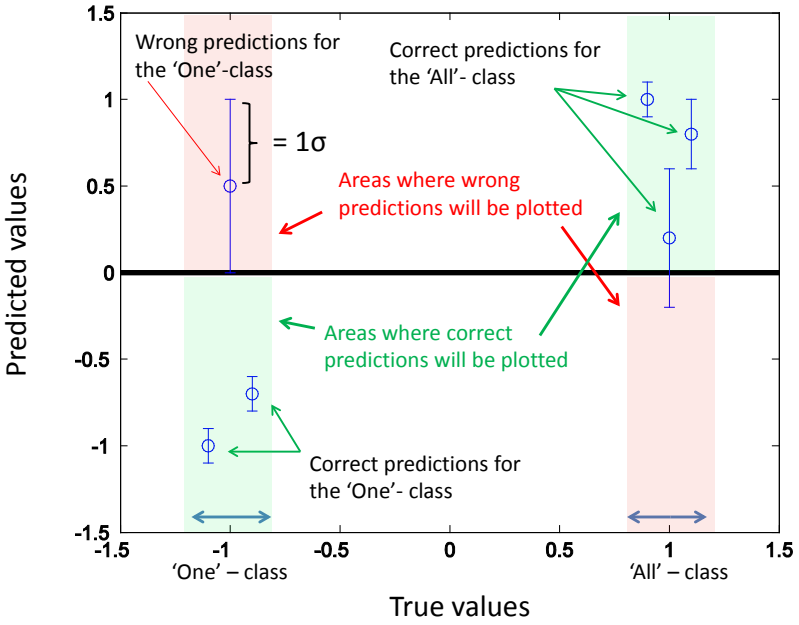


Figure 5.1 – Because the identity of samples is known for the following experiments, classification results can be visuallised in this way. The two blue double arrows at the bottom of the figure indicate the artificially introduced variability for the predictions, this was done for clarity of presentation. A predicted value would normally be plotted either at $x = -1$ or $x = 1$, where x is the 'true value'. The predicted value determines if the prediction is correct for a given value of x. In this example, there are no wrong classifications for the 'All' class (i.e. true value = 1), however, one for the 'One' class (i.e. true value = -1). Error bars are one standard deviation in either direction and indicate the confidence of a prediction.

These standard sets of statistics derived from the confusion matrices were used to determine the performance of each classifier (Congalton, 1991). These statistics include accuracy, precision, and recall. Recall measures the quantity of positive results predicted by the classifier. It is the number of positive results predicted divided by the total number of results that should have been returned. Precision is a measure of the quality of the results predicted and is the number of positive results predicted, divided by the total number of results returned. Neither precision nor recall alone are robust enough measures for the assessment of performance of a classifier (Olson and Delen, 2008). Because of this, the F-score (also known as F1-score) was used which is the harmonic mean between those two measures (van Rijsbergen, 1979). The expres-

sive power of the F-score is thus higher than the individual measures alone because high F-scores can only be obtained if both measures have similarly large values. The F-score is defined as

$$\text{F-score} = \frac{(2 \cdot \text{precision} \cdot \text{recall})}{(\text{precision} + \text{recall})} \ . \tag{5.18}$$

In addition to those statistics, the Kappa coefficient (κ; Congalton et al., 1983; Hudson and Ramm, 1987) was also determined. Although a linear relationship between Kappa and F-score is evident, the Kappa coefficient is often used in remote sensing applications to assess the classification performance over a scene and is thus provided here for ease of comparison. Kappa is defined as

$$\kappa = \frac{N \sum_{i}^{k} x - \sum_{i=1}^{k} x_i + x + i}{N - \sum_{i=1}^{k} n_i + x + i} \tag{5.19}$$

where k is the number of classes and N the number of test data samples. Because a OvA approach is applied throughout this thesis, k is always equal to two.

5.4.1 Experiment 1 - Assessment of classifier performance using cross validation

For many machine learning methods, the performance of the classifier is commonly assessed using a process of cross-validation (Schaffer, 1993; Horwood, 1994). Experiment 1 assesses the performance of different classifiers according to the machine learning literature using cross-validation (e.g. Huang et al., 2002; Arlot and Celisse, 2010; Amari et al., 1997).

Materials and Methods

Two spectral libraries were used in separate cross-validation tests to ensure that the classifier performance was independent of the data-sets. The first data-set was the *WAcoreLib1* which is the spectral library (training set) used throughout most experiments in this thesis. The second data-set was the *WAcoreLibVal* library which

was acquired slightly differently and contains the same classes as the *WAcoreLib1* library. Both data-sets were acquired from the same materials (drill cores) and using artificial illumination for the acquisition of spectra. The *WAcoreLib1* library was acquired from a distance of 15 *cm*; the *WAcoreLibVal* library, however, was acquired using a reflectance probe which was placed in direct contact with the sample (Section 4.2.1).

Table 5.1 – Principal rock types and number of samples for the *WAcoreLib1* and the *WAcoreLibVal* libraries. Drill cores were used for the construction of both libraries. The *WAcoreLib1* library was acquired from a distance of 15 *cm*; the *WAcoreLibVal* library, however, was acquired using a reflectance probe which was placed in direct contact with the sample (Section 4.2.1).

Rock types	**Code**	***WAcoreLib1***	***WAcoreLibVal***
Banded Iron Formation	BIF	63	37
Cherty Banded Iron Formation	CHT	19	28
Martite-Goethite	MAR	38	41
Clay	CLY	15	13
Goethite-Limonite	GOL	17	46
Shale	SHL	19	11
Shale (containing volcanic dust)	NS3	14	43
Manganiferous Shale	SHN	43	14
Total number of spectra:		228	233

Cross-validation uses a single data-set with known labels for every sample which is then divided into a training and a test set. To make such tests more general in the context of the general performance of the classifiers, the data (i.e. spectra) within the training and test set are permutated so that every sample is used at least once as part of the training and the test set. The number of permutations or times the training and test sets are generated using randomly selected samples from the entire data-set is generally defined as the 'fold'. The magnitude of the 'fold' in turn defines the partitioning of the data-set. Generally, a five- to ten-fold cross-validation is recommended (e.g. Breiman and Spector, 1992; Kohavi, 1995; Arlot and Celisse, 2010). A ten-fold cross-validation can, however, overestimate the true prediction error and thus introduce a bias into the results. Lower folds, i.e. three-fold cross-validation, can cause high variance (Hastie et al., 2009). In order to avoid high variance or bias, a five-fold cross-validation process was used in this experiment.

In this five-fold cross-validation process, the data-set was divided into $4/5^{th}$ of training

and $1/5^{th}$ of test data. All classifiers were then applied in a normal manner, where the classifiers learned from the training set and made predictions using the 'unknown' test set. This procedure was repeated five times. For each of the five folds a confusion matrix was generated, stored and accumulated during the remaining folds. A set of standard statistics (e.g. accuracies, F-scores and Kappa) can then be generated for a quantitative assessment of the classifier performance.

For SAM an angular threshold of 0.1 radians was applied to decide if an unknown spectrum matched a library spectrum or not. This angle is standard for many studies (e.g. Murphy et al., 2012) and image processing software (e.g. ENVI; Exelis Visual Information Solutions, Boulder, Colorado). If the angle between a training and a test spectrum was smaller or equal to 0.1 radians, the target spectrum was considered to be a match. Angles above this threshold were considered as too large, and thus spectra were considered to be misclassified.

Results

The different classifiers were assessed using accuracies, F-scores and the Kappa coefficient of agreement. Although accuracies are high, i.e. above 90 % for machine learning methods and above 80 % for SAM, the F-scores and Kappa values showed large differences in the performance of the classifiers for both data-sets (Figures 5.2 and 5.3).

The Gaussian Process framework outperformed SAM and SVMs in terms of the accuracy, F-scores and Kappa values, irrespective of the covariance function used. The GP-OAD achieved on average the highest classifier performance for both data-sets with respect to the three measures of performance. Using the SVM framework, the SVM-NNET outperformed both the SVM-SE and the SVM-OAD. A ranking of the performance of the different classifiers, based on the F-score measure, is summarised in Table 5.2.

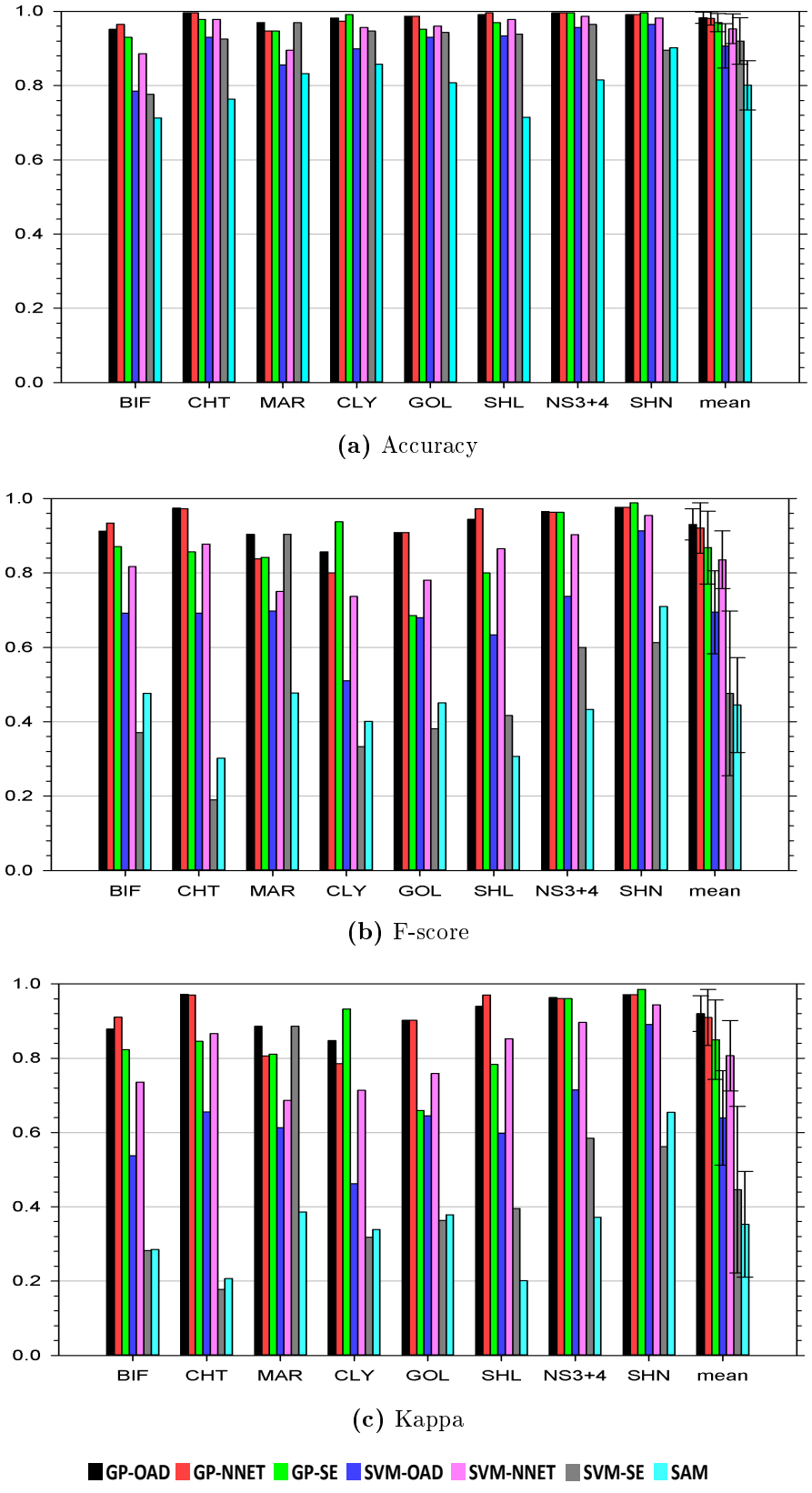
Table 5.2 – Ranking of F-score classifier performance in Experiment 1 for each of the data-sets. Results were obtained using five-fold cross-validation. Table summarises the average F-scores across all classes (see Figures 5.2 b and 5.3 b). Smaller numbers indicate better classification results.

Method	Rank *WAcoreLib1*	Rank *WACoreLibVal*
GP-OAD	1	1
GP-NNET	2	2
GP-SE	3	3
SAM (0.1)	7	6
SVM-OAD	5	5
SVM-NNET	4	4
SVM-SE	6	7

F-scores and Kappa (Figure 5.2 b and c) values showed a very similar relationship between the different methods for both data-sets. The relative differences in performance of all methods compared to the GP-OAD are provided in Table 5.3. F-scores and Kappa values showed that GPs outperformed SVMs using the respective covariance functions. SAM obtained the lowest average F-score value (45 %) but was closely followed by the SVM-SE (48 %) for the first data-set. SAM, however, outperformed the SVM-SE marginally using the *WAcoreLibVal* data-set, with F-scores of 55 % and 50 %, respectively. Within the GP framework, the GP-OAD and the GP-NNET achieved similar F-scores, with a slight advantage for the GP-OAD by 1 % (*WAcoreLib1*) and 3 % (*WAcoreLibVal*). The difference in F-scores between the GP-OAD and the GP-SE was 6 % and 9 % for *WAcoreLib1* and *WAcoreLibVal*, respectively. The SVM-NNET outperformed all other kernels within the SVM framework. The differences in performance of the classifier in terms of the Kappa coefficient of agreement were similar to those of the F-score values and, therefore, were not further discussed. See Table 5.3 and Figures 5.2 and 5.3 for a detailed tabulation of the classifier performance and the relative difference between them.

The performance of all method was variable across the different rock types, however, the smallest variability in terms of the standard deviation (indicated by error bars in Figures 5.2 and 5.3) were smallest for the GP-OAD and were generally small for the GP methods. SVMs and SAM showed large variability in performance. For example,

the SVM-SE showed a very low F-score values for rock types GOL, CHT and BIF but a very high F-score for MAR. A reason why the SVM-SE performed well for MAR might be that MAR is a rock type with little spectral variability within the MAR class (Figures 4.2 and 4.3). This should make it relatively easy for the SE kernel to classify these spectra correctly under ideal conditions of illumination, i.e. artificial light. On the other hand, GOL, CHT and BIF show large within-class variability and class-overlap. This is a problem for many algorithms and as shown by the results presented here, this is clearly a problem for the SVM-SE method.



(a) Accuracy

(b) F-score

(c) Kappa

Figure 5.2 – Performance of classification of the GP-OAD and other methods using a five-fold cross validation approach, applied to the *WAcoreLib1* library. Classification performance was determined using the (a) Accuracy, (b) F-score and (c) Kappa measures. Error bars indicate one standard deviation of the average classification performance.

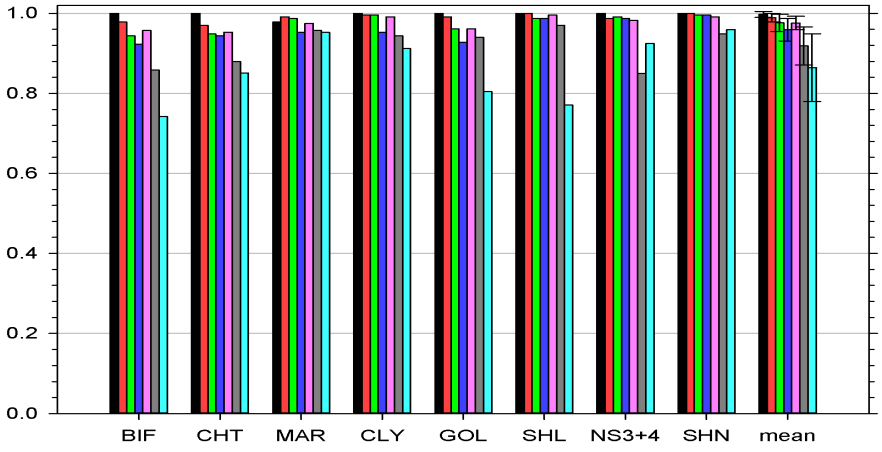
Table 5.3 – Performance of the different classifiers relative to the GP-OAD method using cross-validation. Values are in percent. Results were obtained using two data-sets, the *WAcoreLib1* (a) and the *WAcoreLibVal* (b). A larger value indicates a weaker classifier performance, i.e. a larger relative distance to the performance of the GP-OAD. A negative value would indicate that the GP-OAD was outperformed.

WAcoreLib1	**GP-NNET**	**GP-SE**	**SAM**	**SVM-OAD**	**SVM-SE**	**SVM-NNET**
Δ Accuracy	0.167	1.339	18.537	7.752	6.414	3.067
Δ F-score	1.026	6.701	52.208	25.356	48.830	10.190
Δ Kappa	1.123	7.607	61.650	30.463	51.507	12.303

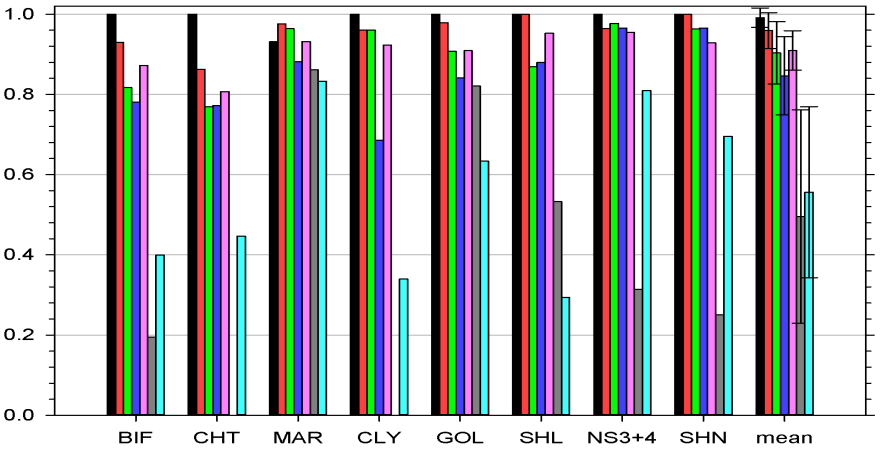
(a) - *WacoreLib1*

WACoreLibVal	**GP-NNET**	**GP-SE**	**SAM**	**SVM-OAD**	**SVM-SE**	**SVM-NNET**
Δ Accuracy	0.801	2.092	13.304	3.867	7.902	2.146
Δ F-score	3.273	8.889	43.907	14.619	50.009	8.236
Δ Kappa	3.746	10.132	50.603	16.926	64.350	9.533

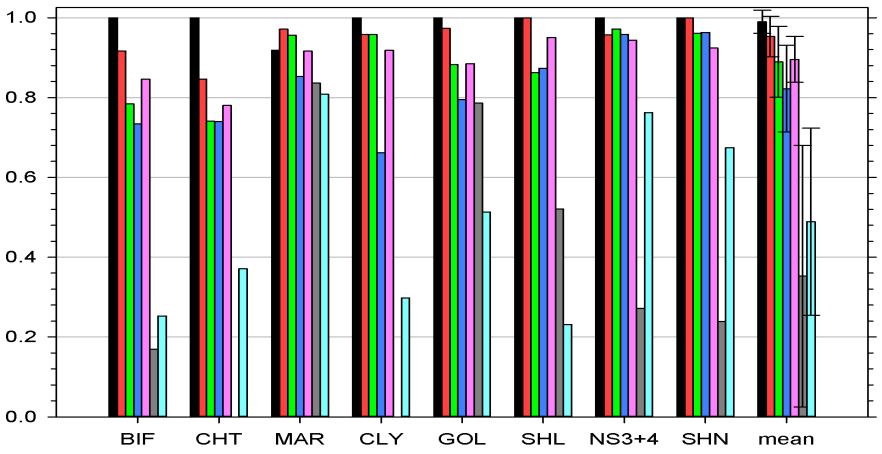
(b) - *WacoreLibVal*



(a) Accuracy



(b) F-score



(c) Kappa

Figure 5.3 – Classification performance of the GP-OAD and other methods using a five-fold cross validation approach, applied to the *WAcoreLibVal* library. Classification performance was determined using the (a) Accuracy, (b) F-score and (c) Kappa measures. Error bars indicate one standard deviation of the average classification performance.

5.4.2 Experiment 2 - Assessment of classifier performance using independent training and test sets

In the context of autonomous mining, it is not feasible to manually extract spectra from a hyperspectral data-set (e.g. a hyperspectral image) to generate a training set in order to classify the same image. Because of this, spectral libraries comprised of classes of rocks typically found in a particular mine need to be created *a priori* to serve as a standard training set. This training set can ideally be used to classify any standard mine face without specific *a priori* knowledge about that mine face[3]. Generally, such a library or training set is acquired under controlled laboratory conditions using artificial illumination to ensure high quality spectra (e.g. high signal-to-noise ratios) and generally only relatively homogenous samples without debris or biofilm are used. Experiment 2 was done to test the hypothesis that the GP-OAD can classify a set of 'unknown' spectra by learning the differences between classes from a different, independent training set, without using cross-validation. Independent in this context means that the training and test data-sets were acquired in separate experiments from different samples of the same rock type or mineral. The GP-OAD was compared to the same methods used in Experiment 1 because using independent training and test sets allows for a more rigorous test of the performance of classifiers than does cross-validation (Zhu and Rohwer, 1996).

Materials and Method

In this experiment no cross-validation was done, instead, the training and test sets were acquired independently. Hyper-parameters were learned from the training data (*WAcoreLib1*) and then used to classify an 'unknown' data-set (*WAcoreLibVal*). For both the training and test set, independent validation data from XRD / XRF analysis were available. These types of data were used to ensure that samples in the training and test set were comprised of very similar chemistry and mineralogy. XRD and XRF

[3] *A priori* knowledge is, however, required of the mine or parts of the mine in form of a training set. This training set is more or less specific to a certain mine, however, independent of a specific mine face.

data were thus used as ground-truth to confirm the 'true' classes of the samples. For the training and the test sets, however, XRD and XRF data were only available for a fraction of the samples from which spectra were acquired. This was due to high costs of such assays, however, it was ensured during the acquisition of spectra that all samples of a particular type of rock were visually similar to each other.

The same methods as in the previous experiment were used, however, a variation in the implementation of SAM was added. In the previous experiment a fixed angular threshold of 0.1 radians was used to decide if an unknown spectrum was a match with a library spectrum of a particular class. This implementation of SAM remains in the suite of methods for reasons of comparability across experiments. However, Murphy et al. (2012) showed that this threshold had a profound impact on the performance of the classification. They showed that the classification performance of SAM was a function of the magnitude of the threshold. In addition, a given threshold increased classification performance for one class but reduced it for another class. For this reason, SAM using the minimum angle criterion was implemented. In this implementation a class label is assigned according to the smallest angle between an unknown spectrum and numerous spectra with known class labels in a spectral library.

Results

The performance of each method is summarised in Figure 5.4. Accuracies showed high values for most classifiers. On average, the GP-OAD showed the highest accuracies value (95 %) followed by the GP-NNET (94 %) and the GP-SE (93 %). SAM using a threshold of '0.1' radians achieved the smallest accuracy (83 %). SAM using the minimum angle criterion outperformed SVMs irrespective of the kernel used. The GP-SE method achieved the same average accuracy as the GP-NNET (94 %). The SVM-NNET achieved the highest average accuracy (91%) amongst the SVM classifiers, followed by the SVM-OAD (89 %) and the SVM-SE (87 %). Most classifiers achieved accuracies of 100 % for SHL and SHN, these were two rock types which had the least amounts of samples in the test set and variability within these classes was relatively

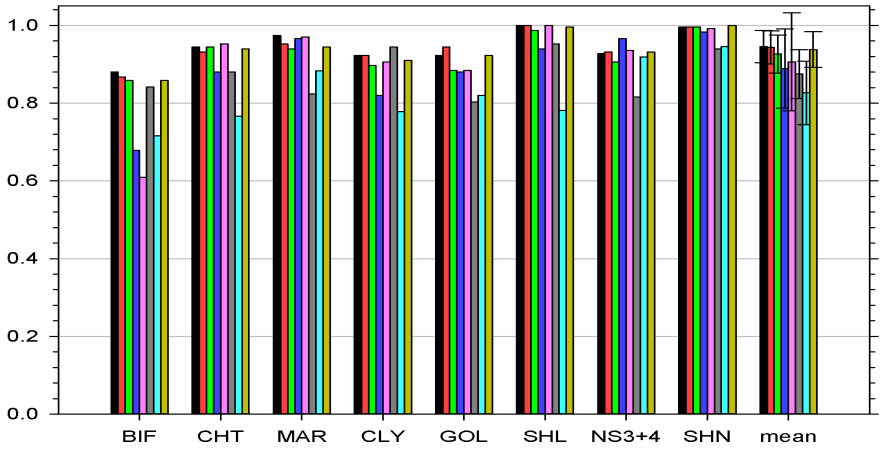
small (Figure 4.3). The lowest accuracies were detected for BIF with an accuracy of 61 % achieved by the SVM-NNET kernel.

F-scores provided results with a much larger expressive power than accuracies. Using the F-scores a distinct pattern in the performance of the classifiers could be detected. Methods or kernels which use an angular metric (e.g. SAM and the GP-OAD) or which use a metric other than the Euclidian distance (e.g. NNET) seemed to perform generally better compared to methods which use the Euclidian distance (e.g. SE). A ranking of the classifier performance and the relative difference to the performance of the GP-OAD is provided in Table 5.4 and 5.5, respectively.

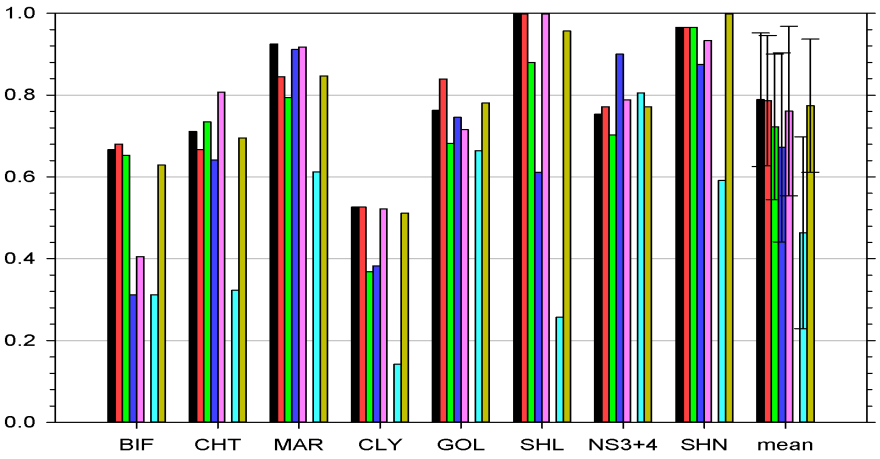
Table 5.4 – Ranking of classifier performance in Experiment 2 for each of the data-sets. Results were obtained using independent training (*WAcoreLib1*) and test data-sets (*WAcoreLibVal*). Table summarises the average F-scores across all classes (Figures 5.4 b). Smaller numbers indicate better classification performance.

Method	**Ranking**
GP-OAD	1
GP-NNET	2
GP-SE	5
SAM (0.1)	7
SAM minAngle	3
SVM-OAD	6
SVM-NNET	4
SVM-SE	8

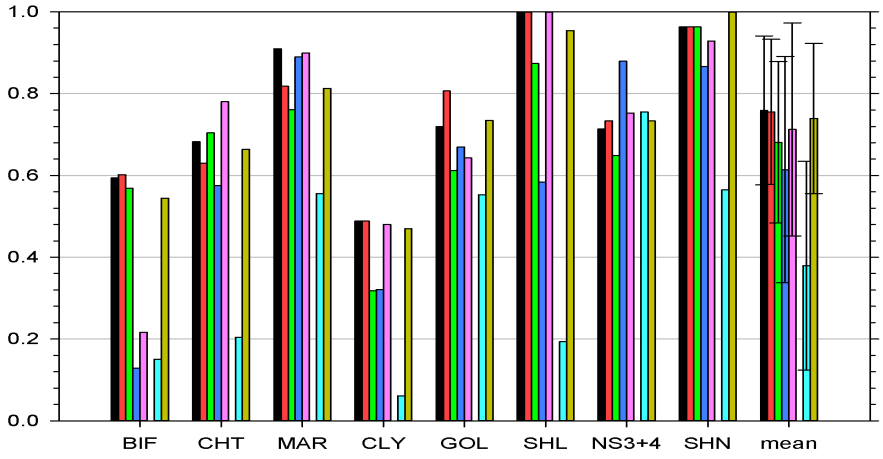
The GP-OAD method outperformed all other methods with the largest average of F-scores across all classes (F-score: 79 %, κ: 76 %), closely followed by the GP-NNET (F-score: 78 %, κ: 75 %), leaving the GP-SE as the weakest GP method (F-score: 72 %, κ: 68 %). GP-NNET was, however, the best method to classify GOL. SAM, using the fixed '0.1' threshold, was the weakest classifier on average (F-score: 46 %, κ: 38 %); SAM, using the minimum angle criterion, however, outperformed all SVMs and the GP-SE method (F-score: 77 %, κ: 74 %). The strongest SVM kernel was the SVM-NNET (F-score: 76 %, κ: 71 %), followed by the SVM-OAD (F-score: 67 %, κ: 61 %). The SVM-SE method failed to produce any results, i.e. F-score and Kappa coefficient were zero.



(a) Accuracy



(b) F-score



(c) Kappa

Figure 5.4 – Classification performance of the GP-OAD and other methods using independent training and test set. The training set was the *WAcoreLib1* library, test set was the *WAcoreLibVal* library. Classification performance was determined using the (a) Accuracy, (b) F-score and (c) Kappa measures. Note that the SVM-SE achieved a value of zero for F-scores and Kappa.

Table 5.5 – Performance of the different classifiers relative to the GP-OAD method using independent data-sets. Values are in per cent. Results were obtained using the *WAcoreLib1* as training and the *WacoreLibVal* as test set. A higher value indicates a weaker classifier performance, i.e. a larger relative distance to the performance of the GP-OAD. A negative value indicates that the GP-OAD was outperformed.

	GP-NNET	**GP-SE**	**SAM-0.1**	**SAM m.a.**	**SVM-OAD**	**SVM-SE**	**SVM-NNET**
Δ Accuracy	0.284	2.042	12.623	0.851	6.012	7.487	4.197
Δ F-score	0.261	8.413	41.257	1.892	14.789	100.000	3.516
Δ Kappa	0.490	10.250	49.951	2.590	19.054	100.000	6.108

5.4.3 Experiment 3 - Assessment of classifier performance using independent training and test sets acquired under artificial and natural light

This experiment is an extension of Experiment 2 and uses the same rationale, that is, that for autonomous mining, independent training data are required to classify geology / mineralogy of a mine face. Experiment 2 assessed the performance of classifiers using independent training and tests when they were acquired using the same type of illumination (artificial light). These data also had high signal-to-noise ratios and all the spectral bands recorded by the spectrometer were used. This experiment assesses the classification performance using independent training and test sets acquired using different types of illumination, i.e. under artificial and natural light. The wavelength-intensity distribution of artificial and natural light are different. For example, natural light has a greater proportion of blue light but it is also affected by absorptions by atmospheric gasses and aerosols. This makes parts of the spectrum affected by absorption unusable (see Sections 2.4.1 and 4.2.1).

This experiment was, therefore, designed to assess affects on the classification performance of the GP-OAD and other methods caused by differences in the type of illumination between a training (artificial light) and a test set (natural light). The same set of methods as in Experiment 2 were used and compared against the performance of the GP-OAD.

Materials and Methods

This experiment uses a library acquired under artificial light as the training set (*WAcoreLib1*) and a test set (*WAcoreLib2*) which was acquired under natural illumination. The test set was acquired from the same cores of rock as the training library. Spectra for the test set (*WAcoreLib2*) were acquired from nadir, at the same distance as the training set. Incorporated into the test data-set were spectra which were acquired from oblique viewing angles (*WAcoreLib2a*; Section 4.2). This was done to introduce variability into the test data-set. This is similar to conditions found in spectra in the field environment in later experiments, i.e. when a mine face image is observed from a fixed observation centre at different angles. Spectral bands around 1.4 and 1.9 μm were removed in the spectra acquired in natural light because these wavelengths were blocked by the Earth's atmosphere (Figure 2.6). An increase in noise towards the extreme end of the spectrum ($>$2.4 μm) was observed for spectra acquired in sunlight and were thus removed [4]. To have the same number of bands in the training and test set, the same bands which were removed in the test set (i.e. spectra acquired under natural light) were also removed from the training set.

To assess the influence of different types of illumination (i.e. natural and artificial light) on the performance of classifiers, the experiment was divided into two parts. In the first part of the experiment, the classifiers were applied to reflectance spectra as in the previous experiments. In the second part of the experiment, both data-sets were converted to hull quotient spectra. This process removes the background brightness of a spectrum (Section 3.1.1). If variability in illumination is a factor in the performance of a classifier, its performance would be better when applied to hull-quotient spectra compared with the original reflectance spectra. Conversely, if variability in illumination was not a factor in the performance of a classifier, then its performance would be similar for hull-quotient and the original reflectance spectra. It is predicted that kernels or methods which use a distance metric to measure spectral similarity will perform poorly when applied to the original reflectance spectra compared with

[4] This noise is probably mainly due to atmospheric water features in the mid infrared, centred around 2.7 μm.

the hull-quotient spectra. Kernels or methods which use an angular metric to measure spectral similarity are predicted to have similar performance when applied to the original reflectance spectra and hull-quotient spectra.

Results

Accuracies were high for all classifiers, however, they were not representative of the actual classification performance as shown by the previous experiments. Thus, they are not discussed in detail here, the reader is however referred to Figures 5.5 and 5.6 for a detailed summary of the performance of all measures, including accuracy, F-score, and Kappa coefficient of agreement.

Using reflectance spectra, the GP-OAD achieved the highest average F-score (F-score: 92%, κ: 91%), closely followed by the GP-NNET (F-score: 90%, κ: 89%). The GP-SE was the weakest kernel among the Gaussian Process methods (F-score: 86%, κ: 84%). SAM using the threshold of '0.1' achieved the lowest F-score overall (F-score: 47%, κ: 38%). SAM using the minimal angle criterion gave much better results (F-score: 88%, κ: 87%), and outperformed all SVM methods and the GP-SE method. The best SVM method was the SVM-NNET (F-score: 74%, κ: 71%), however, it was only marginally better than the SVM-OAD (F-score: 72%, κ: 70%) for both data-set, when looking at the average performances, but not at the performances of the classifiers for single rock types. The SVM-SE, did not achieve any useful classification results and thus F-score and Kappa values were zero. A ranking of the classifier performance and the relative difference to the performance of the GP-OAD is provided in Table 5.6 and 5.7, respectively.

Table 5.6 – Ranking of classifier performance in Experiment 3 for each of the data-sets. Results were obtained using independent training (*WAcoreLib1*) and test sets from combined libraries (i.e. *WAcoreLib2* and *WAcoreLib2a*). Table summarises the average F-scores across all classes (Figures 5.5 b and 5.6 b). Lower numbers indicate better classification results. An asterisk (*) indicates equal performance of methods.

Method	Reflectance spectra	Hull Quotient spectra
GP-OAD	1	3
GP-NNET	2	2
GP-SE	4	1
SAM (0.1)	7	5
SAM minAngle	3	4
SVM-OAD	6	6
SVM-NNET	5	7*
SVM-SE	8	7*

Using hull quotient spectra, an interesting change in the performance of classifiers was noted. The GP-SE achieved the highest F-score (F-score: 92%, κ: 91 %), closely followed by the GP-NNET (F-score: 91%, κ: 90 %), which, in turn, was slightly ahead of the GP-OAD method (F-score: 90%, κ: 89 %). SAM using the 0.1 threshold was the weakest classifier (F-score: 33 %, κ: 17 %) which achieved consistent results for each class. It outperformed, however, the SVM-SE and SVM-NNET which performed equally badly (F-scores: 9 %, κ: 8 %). The SVM-OAD (F-score: 14 %, κ: 12 %) was the best among the SVM methods. SAM using the minimum angle criterion performed well and only slightly worse than the GP-OAD (F-score: 88 %, κ: 87 %).

In comparison between the two data-sets, i.e. reflectance and hull quotient reflectance, SAM using the fixed threshold performed slightly worse using hull quotient spectra. The GP-OAD performed similarly for both data-set. SAM using the minimum angle criterion performed equally well for both data-set. The GP-NNET performed marginally better when normal reflectance data were used. The GP-SE performed better in the case of hull quotient spectra than in reflectance spectra. All SVM methods performed poorly after removing the spectral continuum, except the SVM-SE method which showed a slight improvement, similar to the GP-SE method. Both the SVM-NNET and SVM-OAD, performed slightly worse after removing the continuum

from the spectral data.

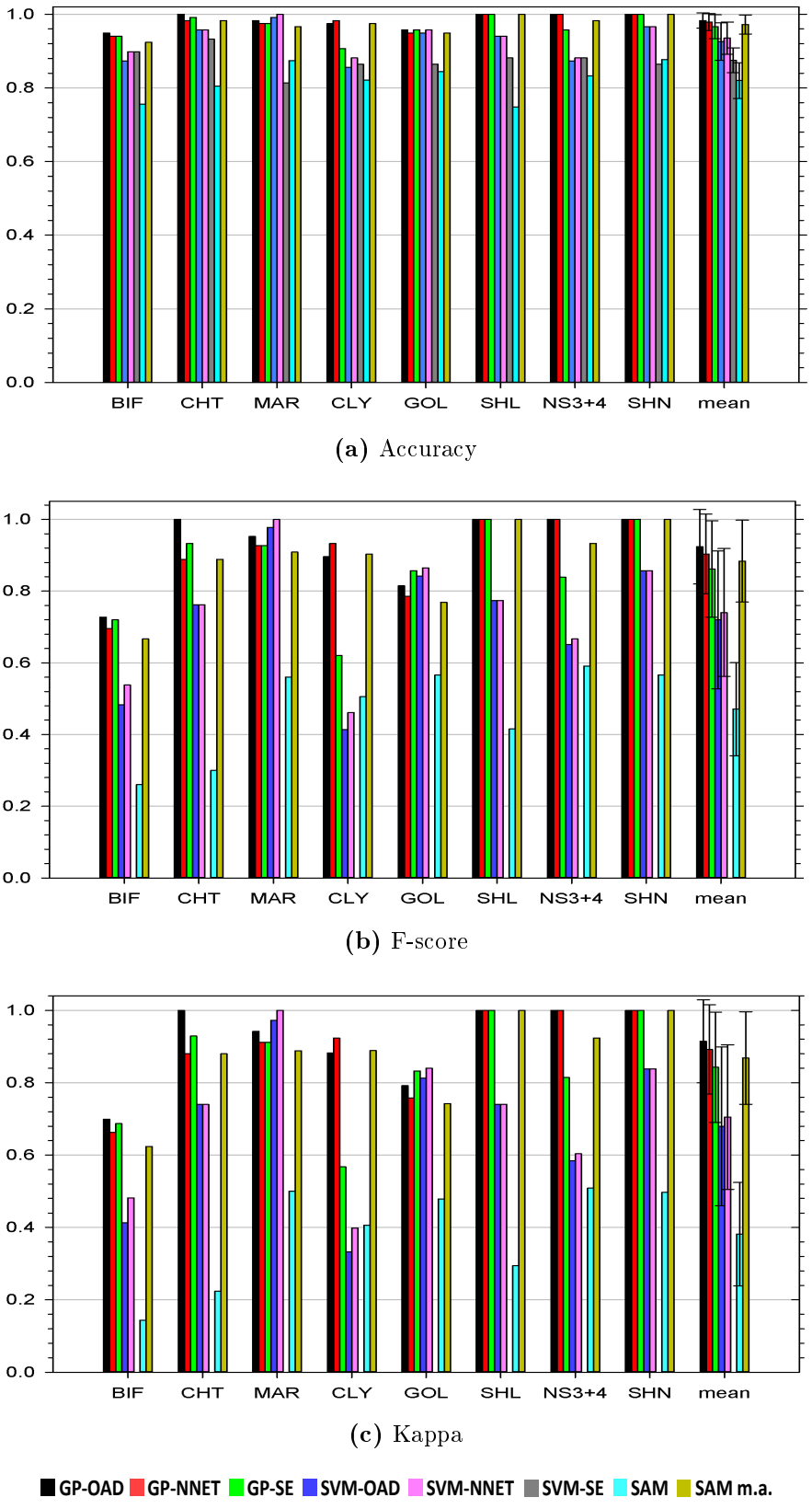
Table 5.7 – Performance of the different classifiers relative to the GP-OAD method using independent data-sets. Values are in percent. Results were obtained using the *WAcoreLib1* as training and the combined *WacoreLib2* and *WacoreLib2a* libraries as test set. A higher value indicates a weaker classifier performance, i.e. a larger relative distance to the performance of the GP-OAD. A negative value indicates that the GP-OAD was outperformed. Results are shown for reflectance spectra (a) and hull-quotients spectra (b).

Reflectance	**GP-NNET**	**GP-SE**	**SAM-0.1**	**SAM m.a.**	**SVM-OAD**	**SVM-NNET**	**SVM-SE**
Δ Accuracy	0.431	1.724	16.619	1.078	5.819	4.849	10.991
Δ F-score	2.173	6.688	49.046	4.337	22.056	19.838	100.000
Δ Kappa	2.463	7.834	58.256	5.040	25.718	22.854	100.000

(a) - reflectance spectra

Hull Quotient	**GP-NNET**	**GP-SE**	**SAM-0.1**	**SAM m.a.**	**SVM-OAD**	**SVM-NNET**	**SVM-SE**
Δ Accuracy	-0.108	0.000	44.784	0.757	10.270	10.162	10.162
Δ F-score	-1.385	-1.755	63.218	2.080	84.883	90.405	90.405
Δ Kappa	-1.449	-1.738	81.185	2.599	87.065	91.495	91.495

(b) - hull-quotient spectra

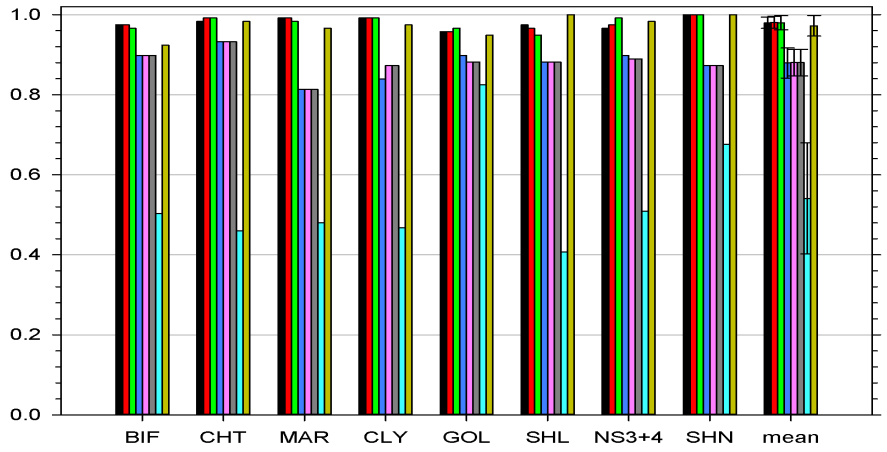


(a) Accuracy

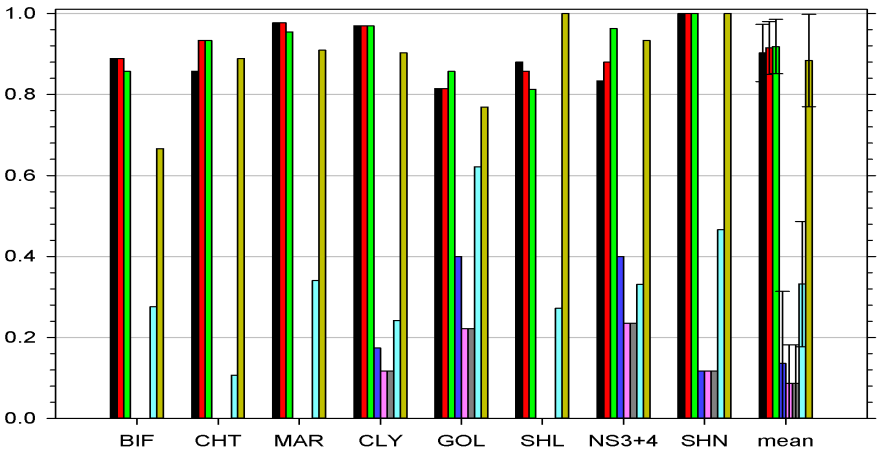
(b) F-score

(c) Kappa

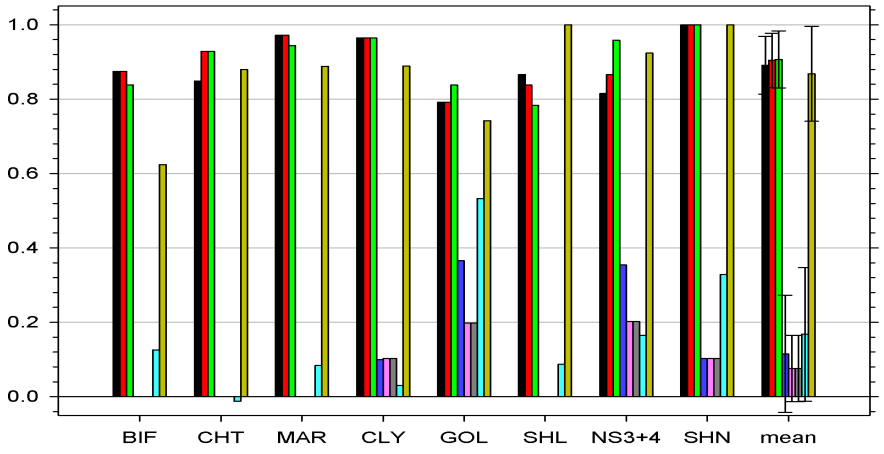
Figure 5.5 – Classification performance of the GP-OAD and other methods using independent training and test set. The training set was the *WAcoreLib1* library, test set was spectra from *WAcoreLib2* and *WAcoreLib2a* libraries. Classification performance was determined using the (a) Accuracy, (b) F-score and (c) Kappa measures. Note that some methods achieved a value of zero for F-scores and Kappa.



(a) Accuracy



(b) F-score



(c) Kappa

Figure 5.6 – Classification performance of the GP-OAD and other methods using independent training and test set. The training set was the *WAcoreLib1* library, test set were spectra from *WAcoreLib2* and *WAcoreLib2a* libraries. Both libraries were converted to hull quotient spectra for this test. Classification performance was determined using the (a) Accuracy, (b) F-score and (c) Kappa measures. Note that some methods achieved a value of zero for F-scores and Kappa.

5.5 Discussion

This chapter introduced the GP-OAD method and assessed and compared the classification performance using hyperspectral data from drill core samples. Three experiments were done to test the GP-OAD's ability to classify rocks based on their spectral curve shape. Experiment 1 was done using cross-validation, Experiment 2 was done using independent training and test sets and Experiment 3 was done using different sources of illumination (and spectra acquired from oblique angles) between training and test sets.

Experiment 1 evaluated and compared the methods under ideal (albeit unrealistic) conditions using high quality field spectrometer data from the same data-set (i.e. using cross-validation). In remote sensing, topographic changes and differences in albedo within a scene occur frequently and can cause classifiers to confuse classes. It is because of this that an assessment using cross-validation does not provide a robust and general test of the performance of a classifier in the context of the classification of remotely sensed hyperspectral data. It can, however, be assumed that a method which does not perform well under these 'ideal' conditions would not yield good classification results of hyperspectral data when illumination conditions are different between the training and the test set. For example, the SVM-SE kernel was among the two poorest performing classifiers in Experiment 1, the other method was SAM using a fixed threshold. This lack of performance of the SVM-SE was later observed in all other experiments. All methods performed less well for Experiment 2 compared with Experiment 1. This was because training and test set were acquired independently and using slightly different methods. This demonstrates the poor expressive power for the evaluation of classifier performance using cross-validation to make statements about the general efficiency of algorithms.

Using independent training and test sets (Experiment 2), both the GP-OAD and GP-NNET outperformed the other methods. SAM using the minimum angle criterion performed almost equally well as the GP-OAD and GP-NNET methods. A reason why these three methods performed similarly is that the data were relatively noise

free and / or the amounts of noise in the training and test set were very similar. Another reason was that these three methods did not use a distance measure such as the squared distance (e.g. as used by GP-SE and SVM-SE). Both the GP-OAD and SAM calculate similarities between spectra in the exact same way. An important aspect within the calculation of the spectral angle is the arccosine-function. Because there is no negative reflectance and the vectors (i.e. spectra) are normalised by their norm (i.e. length), the argument of the arccosine is always within the range of $[0, 1]$. This in turn sets the output range of the arccosine to the range of $[0, \frac{\pi}{2}]$. This means that values do not go beyond this range and thus the arccosine-function is similar to a process of normalisation. The NNET kernel uses a similar rationale by using the arcsin-function which has a different output range of values $[-\frac{\pi}{2}, \frac{\pi}{2}]$ but acts in a similar fashion to the arccosine-function. The SE kernel, on the other hand, uses the absolute distance to calculate similarities, and thus values can become very large and unbound. This is a particular problem when training and test sets were acquired using different types of illumination. Large squared distances occur between different types of rock (i.e. between different classes) using the SE kernel but also between a darker and a brighter spectrum of the same rock. For example, when spectra were from independent data-sets, but acquired using the same light source (i.e. using artificial illumination), the SVM-SE failed (Experiment 2). Differences between the training and test sets were that the measurement distance and the distance of the light source were different. In the first part of Experiment 3, the light sources between the training and test set were different, i.e. where the training set was acquired using artificial light and the test set using natural light. The SVM-SE did not produce any results (i.e. it had zero performance) in Experiment 2 and the first part of Experiment 3 using reflectance spectra.

In the second part of Experiment 3, the continuum was removed from the reflectance spectra. The process of reflectance calibration compensates for different light sources and makes spectra acquired from different scenes but of the same material comparable. There are, however, other effects which cannot be entirely removed by using reflectance calibration (Section 2.1). Although for many minerals, absorptions occur

in the VNIR and SWIR there are also absorptions and scattering processes that occur beyond these spectral regions. For example, Figure 2.3 shows an absorption occurring in the Ultra Violet (UV) region (i.e. absorptions due to charge transfers) whose wing extends roughly to the end of the VNIR. In order to analyse spectral features in the VNIR region, effects caused by absorptions in the UV need to be removed, particularly for feature based methods to work. This is done by removing the continuum (Section 3.1.1) to produce hull-quotient spectra. Results obtained from the classification of hull-quotient spectra showed that the SVM-SE was able to classify some of the classes. While the performance of the GP-SE was very poor compared to other methods, it was nevertheless an improvement to when reflectance spectra were used. The GP-SE performed reasonably well using reflectance spectra (4^{th} rank, based on F-scores) but outperformed all other methods when continuum removed spectra were used. These results indicate that the SE kernel might be considered as a feature based method and not as a method that operates on the entire spectral curve shape (i.e. SAM or the GP-OAD) because it is implicitly dependent on relative proportions (e.g. widths, depths) of diagnostic absorption features. Also, the SE kernel can only operate effectively on normalised spectra which have absorptions in them. It is, however, not always feasible to remove the continuum of a spectrum. If data are degraded due to noise, it may be very difficult to determine the hull of a spectrum without smoothing the spectrum. Smoothing can, however, introduce artifacts and may remove narrow absorption features, thus potentially reducing classification performance.

Contrasting results were obtained from the SVM-SE and GP-SE. The SVM-SE performed very poorly throughout all experiments but the GP-SE performed relatively well, even outperforming competing methods in one of the experiments. These results can be related to the different underlying principles of SVMs and GPs. SVMs can only model amounts of noise in the data appropriately if the kernel allows for this, for example, by learning a parameter relating to the amounts of noise. GPs can do the same, however, they are also capable of learning another parameter which is external to the data and the kernel, which describes the noise in the 'system' or

'process' within the GP[5]. Thus, SVMs cannot model noise independently and, therefore, perform, in circumstance outlined in this thesis, more poorly than do GPs. This limitation and the fact that SVMs can only obtain deterministic answers without a measure of uncertainty, makes them deterministic and thus unsuited for use in autonomous mining.

This chapter introduced the GP-OAD method and demonstrated its capabilities to classify high quality hyperspectral data acquired from an ASD field spectrometer. It was shown that, on average, the method was relatively insensitive to changes in illumination. The GP-OAD provides a solid basis for the use in the field as well as using hyperspectral image data. The proposed method was assessed in various experiments and the performance was generally better than for other methods. This is an advantage due to the stochastic and fully Baysian framework compared to deterministic methods. Deterministic methods have no common basis for fusing information of the same scene acquired from different angles of observation, or fusing data of the same scene acquired by different hyperspectral sensors. Data fusion requires some measure of probability that a particular spectrum has been classified correctly. Thus, it is these probabilities which can be used to fuse information acquired of the same location but from different positions or at different times. This capability was not demonstrated in its entirety but this will be done in the following chapters of this thesis.

[5]A visual illustration of this was provided in Section 3.2.3 using Figure 3.9, panels c) and f), where σ_n^2 determines the uncertainty around the mean function.

Chapter 6

Effects of sensors and measuring conditions on the performance of classifiers

In the previous chapter, the GP-OAD method was introduced and applied to field spectrometer data and compared to other methods (GP-NNET, GP-SE, SVM-OAD, SVM-SE, SVM-NNET and SAM). These data had high spectral resolution and a high signal-to-noise ratio. The acquisition of such high quality data was achieved using a high-resolution field spectrometer under artificial illumination and natural light using a constant target-sensor-illumination geometry.

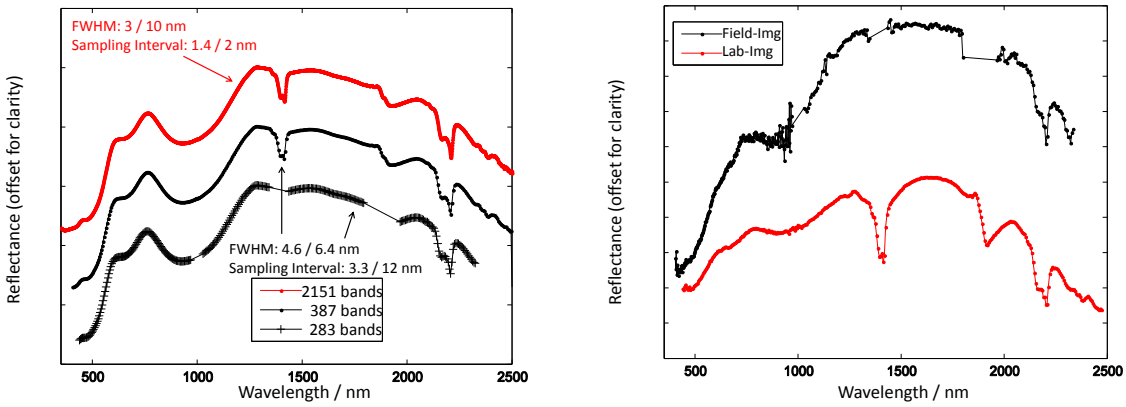
Hyperspectral data, acquired using hyperspectral imaging sensors, differ in four key aspects from data acquired from high-resolution field spectrometers. (i) Hyperspectral image data often have a lower spectral resolution (i.e. fewer and wider bandpasses, Figure 2.9b); (ii) the VNIR and SWIR parts of the spectrum are acquired by two, physically separate sensors[1]. (iii) Data acquired by separate sensors need to be spatially registered to generate continuous spectra across the VNIR and SWIR. Spectral data can therefore be impacted by residual errors caused by misregistration of spatial

[1]The used ASD field spectrometer has three detectors (VNIR, SWIR1, SWIR2) that use the same fibreoptic.

data. (iv) Data acquired from imaging spectrometers often have a smaller signal-to-noise ratio due to the size of the sensor elements and the smaller dynamic range (levels of quantisation). For hyperspectral imagery acquired under laboratory conditions, with a fixed target-sensor-illumination geometry, and a sufficiently bright and uniform illumination, problems arising from poor signal-to-noise ratios are minimised (e.g. Figure 6.1). A specific problem with the particular VNIR imaging sensor used in this thesis caused an artifact in the VNIR spectrum towards shorter wavelengths ($< 500\,\text{nm}$). This was due to a build-up of electrical charge in the VNIR detector array, causing an increase in reflectance at wavelengths below 500 nm (Figure 6.2). Often, the only effective way to remove this artifact from spectra is to truncate the affected bands.

Hyperspectral imagery acquired in the field, using natural illumination, is often much more affected by the effects of poor signal-to-noise ratios than imagery acquired in the laboratory (e.g. Figure 6.1 b). In addition, the process to acquire the data themselves in a dynamic environment is much more difficult. For example, if objects within the scene are very dark and very bright, then compromises have to be made in setting appropriate integration times for both sensors. This generally means that the darker objects in the scene (i.e. those with inherently small reflectance) have a lower signal-to-noise ratio compared to the brighter objects. This is also the case for laboratory data, but in this case it is practical to select samples for imaging which are of similar brightness. Thus, it removes the need to select a smaller integration time to prevent saturation of very bright pixels within the image. Ultimately, the time of the day at which data are acquired is also important, as different sun angles and elevations cause different interaction of the incoming light with the target object (e.g. a mine face with complex geometry), thus causing shadows in certain areas within the target scene.

This chapter provides an assessment of the performance of classifiers on data acquired from imaging and non-imaging spectrometers under different measurement conditions. The GP-OAD method is extended to hyperspectral imagery and is compared to the same classifiers as described in the previous chapter. This extension



(a) data from a non-imaging field spectrometer **(b)** data from an imaging spectrometer

Figure 6.1 – A kaolinite spectrum with different numbers of spectral bands, acquired under different conditions. a) Top spectrum acquired using a field spectrometer at full spectral resolution (2151 bands); middle spectrum a convolved spectrum (387 bands) emulating data from imaging sensors and bottom spectrum a spectrum where atmospheric water absorption bands were removed as well as bands at the junction of the VNIR and SWIR sensors (283 bands). FWHM and the sampling interval for the VNIR and SWIR sensors are given for comparison. b) Differences in signal-to-noise ratios between an image spectrum acquired under laboratory conditions (red, bottom spectrum) and at an operational mine site (black, top spectrum). The amount of noise in the field image spectrum is relatively large, particularly at the junction between sensors and at the extremes of the spectrum. The laboratory spectrum has smaller amounts of noise. Note: Noisy wavelengths due to absorption in the atmosphere were removed from the field spectrum.

enables a direct comparison of classifier performance between hyperspectral sensors, in particular between an imaging and a non-imaging sensor (ASD spectrometer). In addition, the performance of classifiers are evaluated for data acquired in the laboratory and in the field, as these different environments can have a great impact on the resulting signal-to-noise ratio of the data and, potentially, the spectral curve shape. It is important to assess the performance of methods prior to their application to large image scenes (e.g. a mine face) for which ground truth may be sparse or not available and measuring conditions are not ideal.

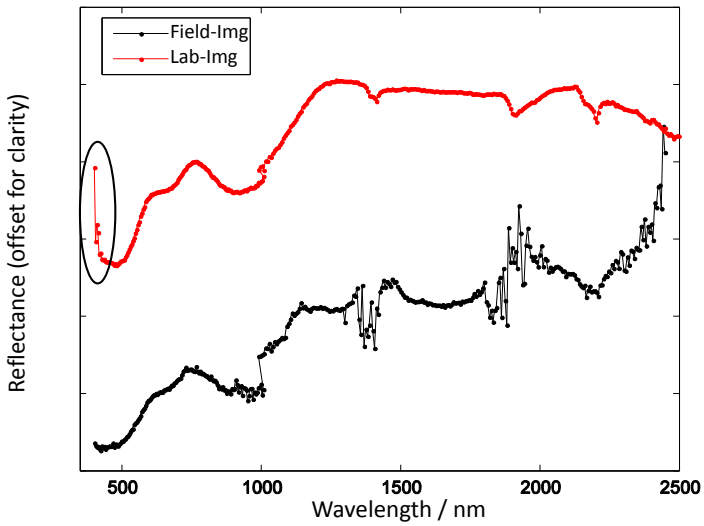
The first experiment in this chapter (Experiment 4, Section 6.2) was done to assess the performance of classifiers using data from different sensors (i.e. an imaging and non-imaging spectrometer). The primary reason for doing this was to determine if

the GP-OAD and other classifiers performed as well on noise free ASD spectra. The amount of noise in ASD spectra is consistently less than in image spectra (cf. Figures 6.1a and b). A direct comparison could not fairly be made between the performance of the GP-OAD on ASD and image spectra in terms of their noise-content. This is because any observed difference may simply be due to the fact that the former has a greater spectral resolution than the latter. In other words, it would be impossible to separate the effects of sensor resolution (number and width of bands) from other properties of image data such as the amount of noise, data quantisation and the use of physically separate sensors to measure the VNIR and SWIR. If the difference in performance between the GP-OAD was small between full-resolution and convolved spectra, then future experiments (e.g. Experiment 5) could make a direct comparison between ASD and image spectra.

Results from full-resolution spectra (presented in Experiment 2, Section 5.4.2) were compared with results from reduced-resolution spectra obtained in Experiment 4. Results from Experiment 4 could, therefore, be used as a baseline against which classification results obtained from the imaging data can be compared (Experiment 5).

Experiment 5 (Section 6.3) was done to determine if the performance of the classifiers changed when training data and test data were acquired from different types of sensors, but under the same conditions of measurement - i.e. using artificial light in the laboratory. The two types of sensors used in this experiment were a field spectrometer and an imaging spectrometer.

In the final experiment of this chapter (Experiment 6, Section 6.4), hyperspectral imagery from several mine faces were classified. The previous experiments used data which were of relatively high quality (e.g. high SNR and constant illumination). Experiment 6 assesses the performance of the classifiers under more challenging conditions found in a dynamic mine environment (high temperatures, wind-blown dust) and differences in sources of illumination (i.e. artificial and natural) between test and training sets.



(a) imaging spectrometer data

Figure 6.2 – Comparison of a field and laboratory spectrum from an imaging spectrometer. The signal-to-noise ratio of the field spectrum is much lower and degrades diagnostic absorption features. The laboratory spectrum presented here is strongly affected by the 'smear effect' (black ellipsoid) compared to the spectrum acquired in the field.

6.1 Implementation of the GP-OAD on Hyperspectral Imagery

The implementation of the GP-OAD classifier to work on hyperspectral imagery is done in a *One versus All* (OvA, also called *One versus Rest*) approach (Rifkin and Klautau, 2004) because it performed slightly better than the multi-class approach in initial tests of the method. A subsequent study using an improved multi-class implementation (Chlingaryan et al., 2013) showed a slight improvement compared to the OvA approach, however, at the cost of prolonged processing time. Therefore, the OvA approach was used as it gave a good compromise between classification performance and classification accuracy. In such a OvA paradigm, a test set of data (e.g. a hyperspectral image) is classified N times by an algorithm, where N is the number of classes in the training set. Each time the test set is classified, a mean ($p(f_*|X_*, X, \mathbf{y}) \simeq \mathbb{E}(f_*|X_*, X, \mathbf{y})$) and variance (σ) are obtained for each image

pixel yielding a 2D array of mean values mv and another 2D array with the associated variances v. The dimensionality of this array is given by the spatial dimension of the input image (i.e. $x \times y$). These arrays are stored and concatenated to generate separate 3D arrays of mean (MV) and variances (V). The dimensionality of either 3D array (Figure 6.3) is of size $x \times y \times z$, where z has the dimensionality N, i.e. the number of classes in the training set and the number of OvA classifiers applied to the test set.

The main difference of the GP-OAD for hyperspectral imagery compared to its implementation for field spectra is in the way class labels are assigned. Thematic maps of geology are created by assigning a class label C to each pixel within the (x, y) dimensional image, with labels ranging between one and N, i.e. $C_{x,y} = \{1, 2, \ldots, N\}$. This is achieved by determining the z position (index) at (x, y) in the 3D array (MV) where the mean is the closest to '-1' (the 'true' label for the 'One' class). This can formally be expressed as:

$$C_{(x,y)} = \underset{z}{\operatorname{argmin}}\, MV(x, y, z) \ . \tag{6.1}$$

6.2 Pre-assessment of classifier performance on convolved field spectra - Experiment 4

Hyperspectral image data are acquired at a coarser spectral resolution than field spectrometer data. In order to compare performances of different classifiers when applied to field non-imaging and imaging data, it is first necessary to determine if there is a decrease in the performance caused by a reduction in the number of spectral bands.

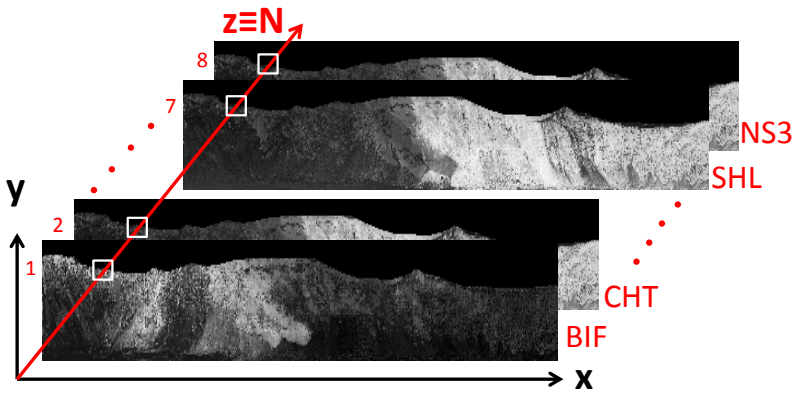


Figure 6.3 – Concept for the classification of hyperspectral imagery using a OvA approach. Output images, containing mean values in each pixel, are shown as grayscale images. Different layers correspond to different OvA classifier outputs for a different rock type (red numbers / labels). The (x, y) coordinates represents the spatial dimensionality of the output image (which is equal to the input image). z has dimensionality N and corresponds to the number of classes in the training set. White boxes symbolise the same pixel position across all layers. The smallest mean value is found across this dimension and a class label is then assigned according to the index of z in which the lowest mean value was found for a particular pixel in (x, y).

6.2.1 Materials and Methods

Two spectral libraries, a training and a test library, were created from high-resolution spectra by converting them to the bandpasses of imaging spectrometers (used in later experiments). This was done to isolate the number of spectral bands from other properties or extraneous factors such as variations in illumination among pixels and greater amounts of noise inherent with hyperspectral imaging data. The use of these two libraries enables the assessment of the influence of the number of spectral bands on the performance of different classifiers. Results obtained in this experiment were compared to a baseline (i.e. high-resolution spectra) which is provided by the results obtained in Experiment 2 (Section 5.4.2).

Table 6.1 – Data used in Experiment 4. 'Original' indicates the ASD field spectrometer data, the following two rows have the same bandpasses and wavelengths except that some bands were removed (indicated by 'BR') to form a data-set with 283 bands. The 'WAcoreLib1' and 'WAcoreLibVal' were both converted to emulate the bandpasses of hyperspectral imaging data (with respect to the number of spectral bands and their width) as given below.

Type of Data	Bands	Sampling Interval	FWHM	Spectral range
original	2151	1.4 nm (VNIR), 2 nm(SWIR)	3 nm (VNIR), 10 nm (SWIR)	350 - 2500 nm
convolved	359	3.3 nm (VNIR), 12 nm (SWIR)	4.6 nm (VNIR), 6.4 nm (SWIR)	401 - 2447 nm
convolved + 'BR'	283	3.3 nm (VNIR), 12 nm (SWIR)	4.6 nm (VNIR), 6.4 nm (SWIR)	439 - 2325 nm

Both the training (*WAcoreLib1*) and test (*WAcoreLibVal*) libraries were convolved to the bandpasses of the imaging spectrometers. Conversion of the high-resolution spectra to imaging spectrometer bandpasses was done using a standard method (i.e. Gaussian convolution) using the image processing software ENVI (Exelis Visual Information Solutions, Boulder, Colorado). Noisy bands at the extreme ends of the spectrum were removed so that the convolved spectra had the identical number of bands as the image data used in Experiment 5, i.e. 359 bands compared to 2151 bands (Table 6.1). Bands at the short end of the spectrum were removed due to the artefact caused by a build up of charge in the VNIR detector.

A second data-set was created to include the same pre-assessment for the ultimate experiment in which data were acquired under natural illumination, where bands affected by atmospheric absorption were removed (Experiment 6). For the second data-set (*convolved+'BR'*, Table 6.1) the number of bands were reduced further to 283 bands. Noisy bands at the junction of the VNIR and SWIR sensor were removed as well as bands affected by atmospheric water absorption (1345-1432 nm and 1800-1962 nm). In addition, bands at the end of the SWIR were also removed, yielding a spectral range of 439nm to 2325nm.

6.2.2 Results and discussion

The performance of classifiers was evaluated using the same quantitative statistics as in Chapter 5, these were Accuracy, F-score and the Kappa coefficient of agreement. The performance of the different classifiers using the convolved libraries (i.e. with 359 and 283 bands) was compared to the full spectral resolution libraries (2151 bands). Results are presented relative to those obtained in Experiment 2 (Section 5.4.2) by calculating the root mean squared error (RMSE) between the average values across all classes for accuracies, F-scores and Kappa from the full-resolution library and the convolved libraries (Table 6.2).

Table 6.2 – Comparison between classifier performance using data-sets with different numbers of spectral bands. The classifier performance was evaluated using the *WAcoreLib1* library (training set) and the *WAcoreLibVal* library (test set). Values in tables represent the RMSE between the performance of classifiers using the full-resolution data-sets of Experiment 2 and performance achieved using the convolved data-set. Negative values indicate that the classifier achieved a higher performance when a) 359 band b) 283 bands were used, respectively, compared to the 2151 band data-set. Positive values indicate the opposite. c) compares the classifier performance using 359 against 283 spectral bands. Negative values indicate a higher classifier performance using the 283 data-set and positive values indicate higher classifier performance using 359 band data-set. 'SAM m.a.' indicates the 'minimum angle criterion'.

2151vs359 bands	GP-OAD	GP-NNET	GP-SE	SAM-0.1	SAM m.a.	SVM-OAD	SVM-SE	SVM-NNET
Δ Accuracy	0.51	0.34	-0.70	-0.37	1.03	0.60	0.12	0.00
Δ F-score	2.48	1.08	-2.83	-1.27	4.18	-4.52	n/a	2.35
Δ Kappa	2.96	1.41	-3.50	-1.62	5.15	-4.85	n/a	2.66

(a)

2151vs283 bands	GP-OAD	GP-NNET	GP-SE	SAM-0.1	SAM m.a.	SVM-OAD	SVM-SE	SVM-NNET
Δ Accuracy	0.62	0.97	-0.75	0.30	1.14	0.66	44.51	-0.24
Δ F-score	2.52	3.97	-3.16	-1.29	3.43	-5.25	n/a	2.07
Δ Kappa	3.09	4.87	-3.91	-1.20	4.47	-5.73	n/a	2.27

(b)

359vs283 bands	GP-OAD	GP-NNET	GP-SE	SAM-0.1	SAM m.a.	SVM-OAD	SVM-SE	SVM-NNET
Δ Accuracy	0.62	0.97	-0.75	0.30	1.14	0.66	-0.06	-0.24
Δ F-score	0.04	2.92	-0.32	-0.03	-0.78	-0.70	-59.45	-0.29
Δ Kappa	0.13	3.52	-0.40	0.42	-0.72	-0.84	-62.10	-0.39

(c)

In the following, descriptions and discussions of results are made in relation to the performance of the classifiers achieved using the full-resolution ASD spectra obtained in Experiment 2 (Section 5.4.2).

For the data-sets with 359 bands (Table 6.2 a) and 283 bands (Table 6.2 b) a slight

decrease in classifier performance relative to the full-resolution data-set was found for most classifiers with respect to the average accuracy. Average F-scores and Kappa values obtained from classification of data-sets with 359 and 283 bands were slightly greater than for the full spectral resolution data. The GP-OAD performed around 2.5 % better with full spectral resolution compared to either reduced data-set. The GP-NNET performed 1 % better with the full-resolution library compared to the library with 359 bands, however, a further decrease in performance was noted when bands were removed to 283 (approximately 5 % difference to the full-resolution library). The GP-SE showed around 3 % improvement for the 283 and 359 band data-sets compared to the full-resolution library. SAM, using a fixed threshold, showed a 1.3 % improvement. SAM, using the minimum angle criterion, had a decreased performance of approximately 4 % for either data-sets. The SVM-OAD improved by approximately 5 % using either reduced library. The SVM-SE could not be evaluate quantitatively since it did not yield any meaningful results in Experiment 2, however, average absolute F-scores values were 0.143 (359 bands) and 0.228 (283 bands) which were extremely small values compared to performances achieved by the other methods (which ranged between 0.47 and 0.78). The SVM-NNET method performed slightly better (approximately 2 %) using the full-resolution library than using either one of the reduced data-sets. The main differences between the 359 band and the 283 band data set are that noise at the beginning and the end of the spectrum is removed and water absorption bands near 1400 and 1900 nm are removed. Water features, particular at 1400nm, contains diagnostic absorption features for clay minerals. However, because this feature is present in many rock types (e.g. SHL, CLY, SHN, NS3+4) it can cause ambiguity between these classes for some methods. Thus, by removing this feature, ambiguity can be reduced for some classifiers because they can rely only on the more diagnostic Al-OH doublet (2208nm).

By comparing the average classifier performance across all classes between the two reduced band libraries (i.e with 359 and 283 bands), only slight differences in the GP-OAD's performance were noted - less than 1 % improvement - for the 359 band library. The GP-NNET also performed better (by 3 %) when using 359 bands instead

of 283. The GP-SE, however, performed better with respect to all measures when 359 bands were used ($< 1\,\%$). SAM, using a fixed threshold, performed similarly in both libraries across all measures of performance. The minimum angle criterion showed a slight improvement ($< 1\,\%$) in the F-score and Kappa values when using the 283 band library, however, a slight decrease in the average accuracy was observed ($< 1\,\%$). Similar results were observed for the SVM-OAD and SVM-NNET (i.e. an improvement of $< 1\,\%$) for the 283 band library. The SVM-SE showed significant improvement when the 283 band library was used compared to 359 band data-set. Although only a marginal improvement in the average accuracy was observed, average F-scores and Kappa values showed an improvement of around 60 %.

From these results the following conclusions can be made. Using convolved field spectrometer data with and without atmospheric water bands showed a small reduction in the performance for some classifiers compared to the full-resolution data. A decrease in performance in terms of the F-score using 359 bands compared to 2151 bands was found. Differences were between 1 and 4 % for all methods. Exceptions to this were the GP-SE, SVM-OAD and SAM using a fixed threshold which showed improvements between 1.3 and 4.5 % using the reduced, 359 band data-set. Using the 283 band data-set, a similar performance compared to the full-resolution data-set was found. This suggests two things: (i) GPs yield more reliable classifications than SVMs when applied to hyperspectral data where correlations between neighboring bands are present. (ii) The SE kernel is more affected by the number of spectral bands than angle based methods (i.e. the OAD) causing a very poor classification performance for high dimensional data-sets. Using the 283 band data set provides a sufficiently high resolution to detect most narrow absorption features (e.g. Al-OH feature). The 359 band data set contains the diagnostic water absorptions at 1400 and 1900 nm and more noise bands at the end and beginning of the spectrum, however, the reduced resolution is insufficient to resolve subtle differences of these diagnostic absorption features between similar minerals. By reducing the number of bands from 2151 bands to either 359 or 283, correlations between neighbouring bands are significantly reduced, thus reducing redundant information. This helps some methods (e.g.

SVMs) and some kernels (e.g. SE) to perform slightly better with fewer numbers of bands while GPs can generalise better over the higher dimensional data sets.

The overarching conclusion from these results is that the number of bands does not affect the performance of classification algorithms significantly. Therefore, if the classification performance of a method decreases in the following experiments, it is a consequence of other factors (e.g. amounts of noise and/or variability in illumination).

6.3 Application and validation of the GP-OAD using hyperspectral imagery acquired in the laboratory - Experiment 5

Data from field spectrometers are generally of high quality in terms of the signal-to-noise ratio and the spectral resolution. Imaging spectrometer data can often have lower signal-to-noise ratios and fewer and broader spectral bands than non-imaging spectrometers. In the previous experiment, the effects of reducing the number of spectral bands of data from a field spectrometer were assessed using two independent spectral libraries, one for training and one for testing. The spectral signal of imaging spectrometers can be affected by (i) residual errors during the spatial image registration of the VNIR and SWIR sensors which can cause a jump in spectra between the VNIR and SWIR; (ii) a potentially smaller signal-to-noise ratio due to the size of the sensor elements and (iii) a smaller dynamic range (i.e. levels of quantisation). In this experiment, the GP-OAD and other methods were applied to imaging spectrometer data which contained these effects to investigate their influence on the classification performance. Data were acquired using artificial illumination so that sources of illumination between the training and the test set were similar, however, target-light-sensor geometries and the types of sensors between these two data-sets were different.

6.3.1 Materials and Methods

The aim of this experiment was to assess the performance of the classifiers when training and test data were acquired using different sensors but under similar illumination conditions. Although artificial light was used to acquire the training set (*WAcoreLib1* spectral library) and the test set (i.e. hyperspectral imagery), different light sources and different target-light-sensor geometries were used for both data-sets. Training data were acquired using a field spectrometer (*WAcoreLib1*) which were convolved to the bandpasses of the imaging spectrometer (see Experiment 4, section 6.2). Separate images were acquired from several trays containing exploration drill cores from a West Angelas mine. Imagery was acquired using two Specim imaging spectrometers (VNIR; 400-970 nm) and shortwave infrared (SWIR; 971-2516 nm) sensors (Specim, Finland). The VNIR and SWIR imagers had full-width at half-maximum spectral resolutions of 4.6 and 6.4 nm, respectively. Up to 40 frames of data were acquired from the white reference (Spectralon®) covering the entire width of the image swath. After image acquisition, dark-current was subtracted from the image on a line by line basis. Calibration to reflectance was done on a line by line basis, i.e. each image line was divided by the reflectance of the calibration panel for the same image line. To obtain an optimal SNR, the white reference and the target were acquired in separate images, each with their own integration time. The maximal integration time was set without saturating any pixels in the image. For example, to obtain a VNIR reflectance image, several frames of the white reference were acquired and averaged at a certain exposure time t_{WR}; hereafter, the tray was acquired using a different exposure time t_{tray}. Both exposure times were set so that values occupied the full dynamic range of the sensors. This method yielded two images, one 'calibration' image and one 'target' image; both images were then used to obtain a normalised reflectance image using Equation 6.2.

$$\rho(\lambda)_{norm.tray} = \frac{DN(\lambda)_{tray} \cdot \rho\,'(\lambda)_{WR}}{DN(\lambda)_{WR}} \cdot \frac{t_{tray}}{t_{WR}} \tag{6.2}$$

where $\rho(\lambda)_{norm.tray}$ is the reflectance of a tray image at band λ, $DN(\lambda)_{tray}$ is the digital

Number (DN) of an individual pixel in the tray image, $DN(\lambda)_{WR}$ is the line average of forty image frames of the white reference and $\rho\,'(\lambda)_{WR}$ is the reflectance factor of the calibration panel at wavelength λ. The difference in illumination (exposure time) was taken into account during normalisation using the last term in Equation 6.2. The same process was repeated for the SWIR sensor. The separately normalised VNIR and SWIR images were then spatially registered, yielding a continuous spectrum at each image pixel. Sixteen images from exploration drill cores were acquired and processed in this way.

The 16 images acquired covered several tens of metres (between 50 and 60 m) of rock cores. Only specific rock types of interest were sampled from the cores and these were separated by large unknown sections of core which were not used for this analysis. So, for ease of processing and to enable concise presentation of data, the areas of core which were of interest, were combined into a single image (8 classes in total). Interactive image analysis was used to define regions of interest for the areas of core. Each region of interest contained pixels of a single rock type which was assigned to one of the eight classes. This combined image is the test set used in this experiment. Only XRF data were used to assign appropriate class labels as ground truth.

For consistency across experiments, classifiers were applied to data with the same number of bands as in the previous experiment (Section 6.2), i.e. data-sets with 359 and 283 bands (Table 6.3).

Table 6.3 – Data-sets *LabImg359* and *LabImg283* used in Experiment 5. Data-set *LabImg359* retained atmospheric water absorption bands which were removed for data-set *LabImg283*. Some additional bands at the end of the spectra in data-set *LabImg283* were also removed due to noise. Data-set *LabImg283* has the same number of bands as data acquired under natural sunlight (Experiment 6).

Data-sets	LabImg359	LabImg283	Comments
Training set	*WACoreLib1*	*WACoreLib1*	Training set contains 228 spectra in 8 classes; the test set contains 49 rock samples.
Test set	*LabImg359*	*LabImg283*	The test set is a hyperspectral image opposed
Bands	359	283	to previous experiments (ASD field spectra).

6.3.2 Results and Discussion

Classified images derived using the different classifiers show different classification accuracies for different classes (Figure 6.4). Classification using SVMs with hard decision boundaries and several different kernels (as used in previous experiments) yielded poor results (Figure 6.5).

The other SVM results in Figure 6.4 were obtained by applying Platt's probabilities (Platt, 2000) instead of traditional SVMs with hard decision boundaries. Probabilistic SVMs yielded much better classification results than did the traditional non-probabilistic SVMs (cf. Figures 6.4 and 6.5). A few samples of GOL and NS3 and most CHT samples were misclassified by the GP-OAD (Figure 6.4). The remaining rock types appeared visually to be classified correctly. Very similar results were obtained for the GP-NNET. SAM using the minimum angle (m.a.) criterion showed similar results to the GP-OAD and GP-NNET. SAM (m.a.) appeared to confuse GOL and BIF classes. The SVM methods with Platt's probabilities performed relatively well, in particular, SVM-OAD and SVM-NNET showed a slightly better agreement with the ground truth of the GOL and CHT class compared to their GP competitors. However, the SVM-OAD and SVM-NNET misclassified more MAR and BIF classes compared to the respective GP methods. The SVM-SE showed the poorest agreement with the ground truth.

Quantitative measures of performance, Accuracy, F-score and Kappa values are shown in Figure 6.6. The GP-OAD method outperformed all other GP methods by a small amount. In comparison to probabilistic SVMs the GP-OAD performed better and by larger margin. SVMs with hard decision boundaries performed very poorly as indicated by the qualitative results (Figure 6.5) and are thus not suited for the classification of hyperspectral imagery in this context. Quantitative results were, therefore, not produced for these images.

The performance of classifiers when applied to data with 359 bands instead of 283 bands did not change results significantly, however, all methods performed slightly better using 283 bands, except the SVM-SE which performed better using 359 bands

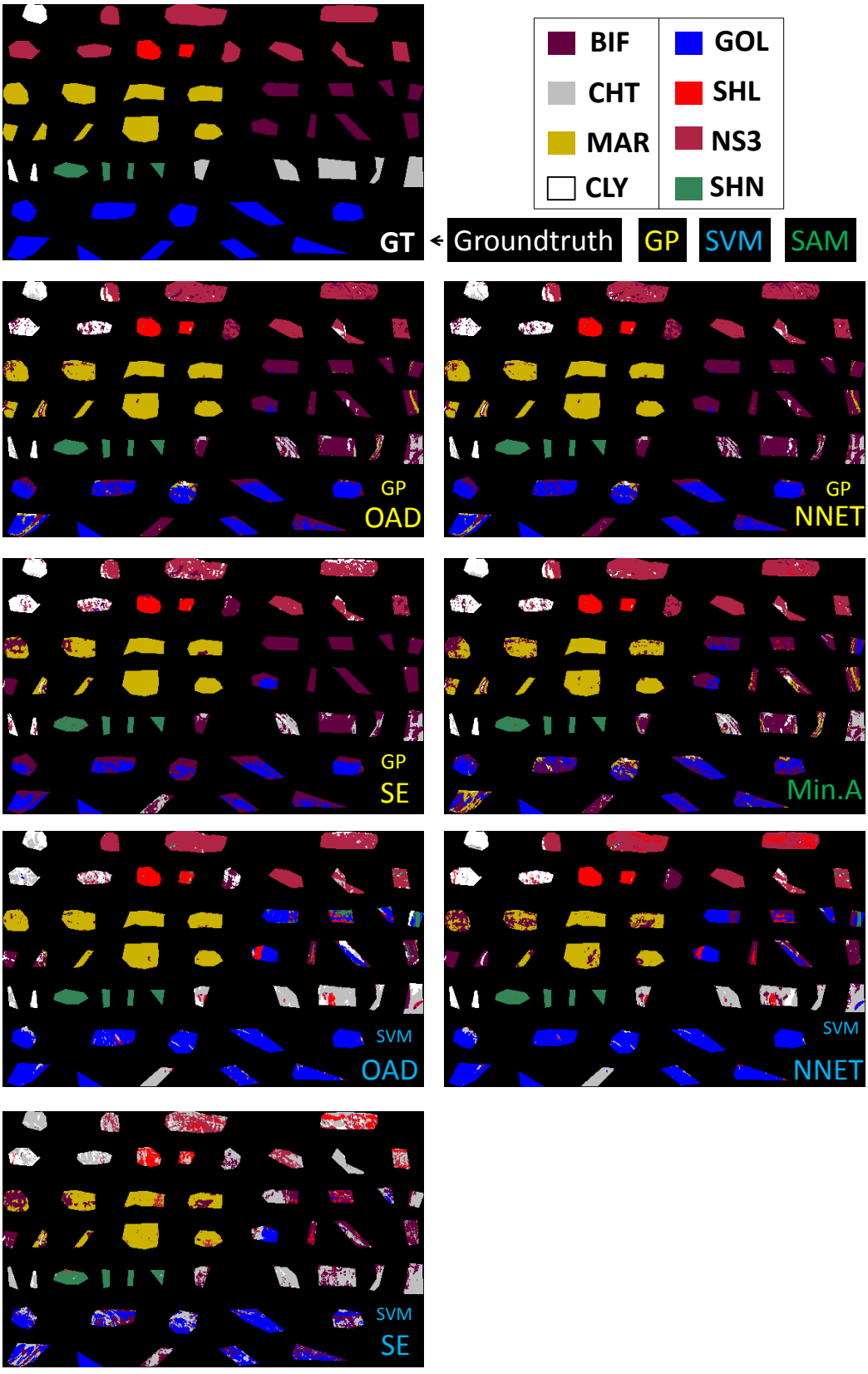


Figure 6.4 – Qualitative classification results using 283-band data-sets - *LabImg283*.

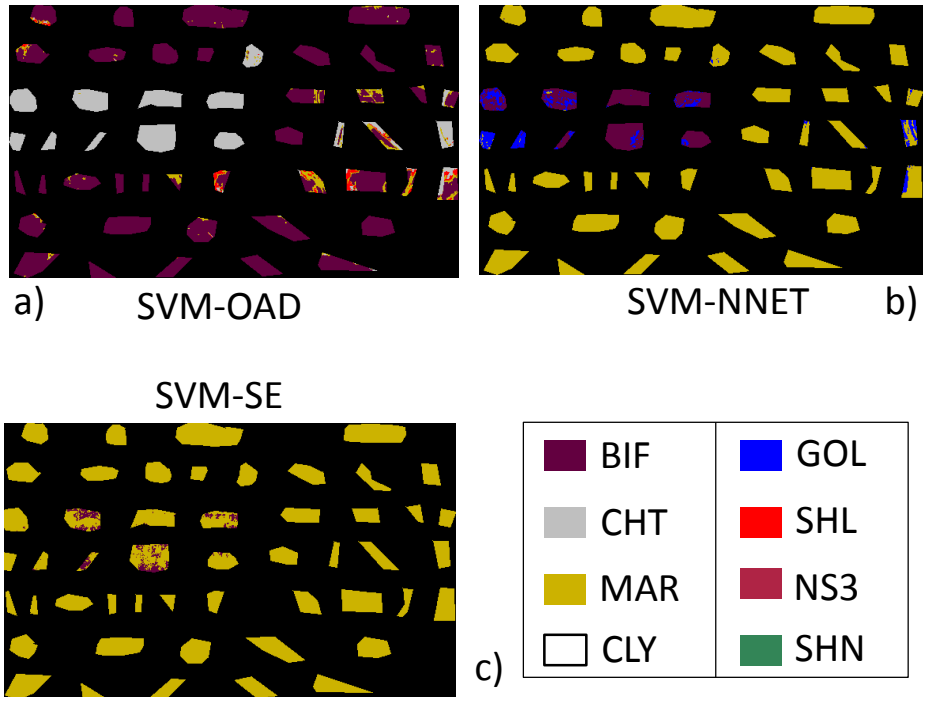


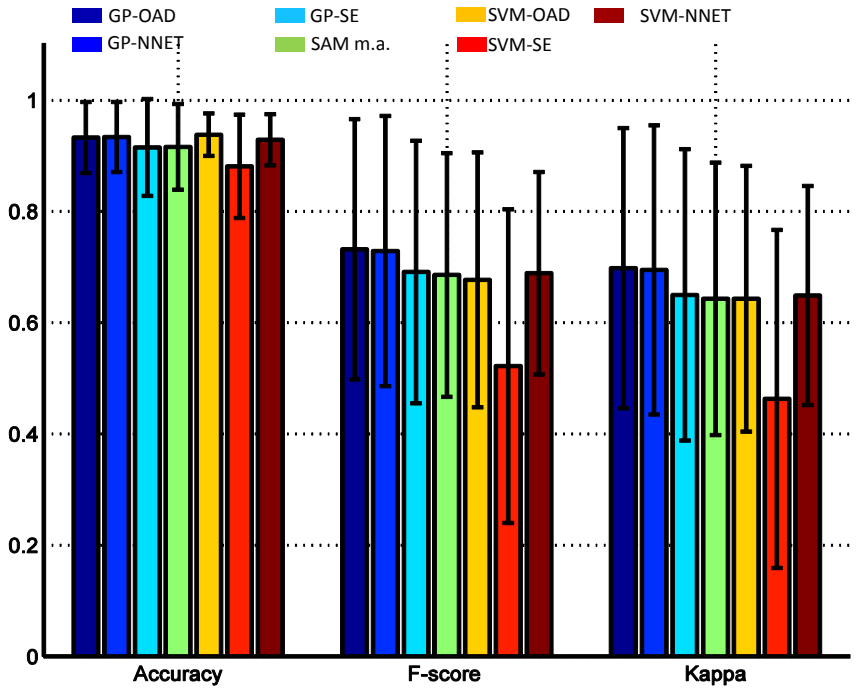
Figure 6.5 – Qualitative results using SVMs with hard decision boundaries. *LabImg283* was used for these results.

(cf. 6.6 a and b). Quantitative results for the *LabImg359* and *LabImg283* data-sets are compared against each other and also against the performance of classifiers obtained using independent training and test data. These results were obtained from data acquired using a field spectrometer (i.e. *WAcoreLib1* and *WAcoreLibVal*) in Experiment 2 (Section 5.4.2). This was done to provide a direct comparison between the performance of classifiers using independent training sets from a field spectrometer and for independent training and test sets using field spectra and imagery, respectively.

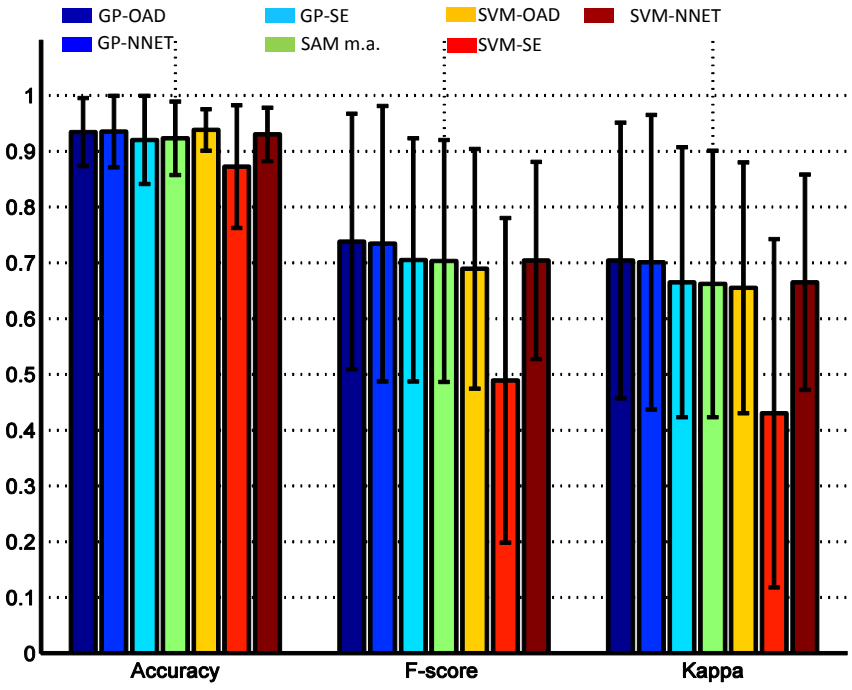
Generally, the GP framework appears to benefit from an increase in spectral bands as well as SAM (m.a.). Using the SVM framework, the number of spectral bands seem to affect some kernels more than others. For example, the SVM-OAD performed on average slightly worse using fewer bands compared to when the full spectral resolution data-set was used (2151 bands). The SVM-SE failed to produce any useful classification results (using 2151 bands). The SVM-NNET showed an improvement in

classification performance with respect to Accuracy, F-score and Kappa when using more spectral bands.

GP based methods and SAM generally distinguish better between CLY, CHT, SHL, and NS3+4 although CHT is often misclassified with many false-negative classifications this is probably due to the fact that some CHT spectra in the training set contained diagnostic water features at 1400 nm and also the diagnostic Al-OH feature at 2208 nm. This indicates that the overall curve shape is more important than individual absorption features. On the other hand, the SVM based methods are relaying more on absorption features to distinguish different rock types.

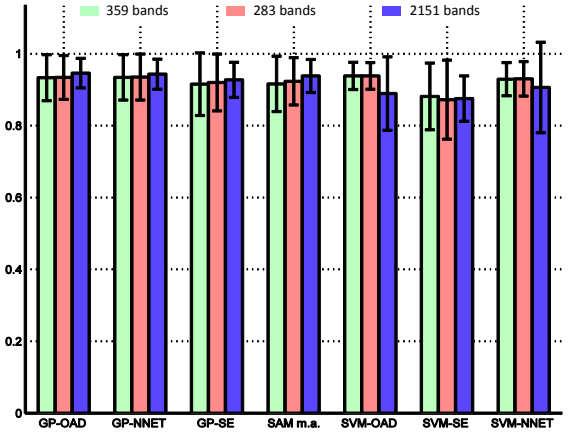


(a) 359 bands

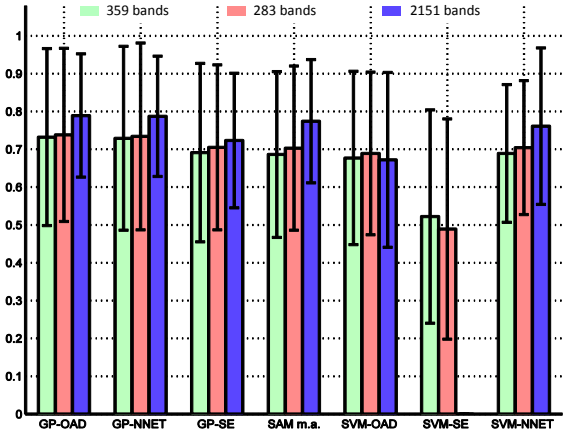


(b) 283 bands

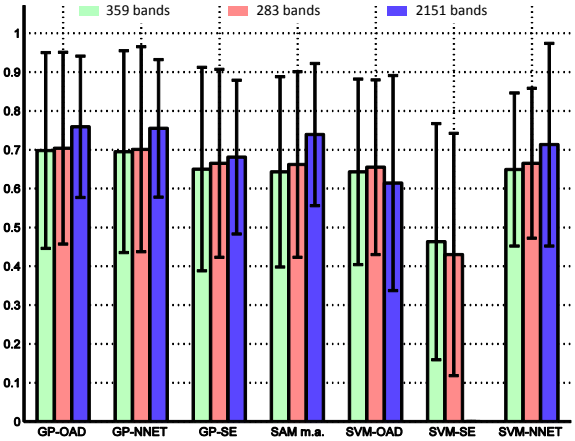
Figure 6.6 – Overview of quantitative classification results using 359 bands (a - *LabImg359*) and 283 bands (b - *LabImg283*).



(a) Accuracy



(b) F-score



(c) Kappa

Figure 6.7 – Comparison of performance of classifiers applied to imagery with different numbers of spectral bands (283 and 359 bands; Table 6.3). Results using 2151 bands were obtained from Experiment 2 (Section 5.4.2) using ASD field spectra of the same targets (rock samples). The values in each graph are the averages across all rock types (classes) and their standard deviation.

6.3.3 Causes of misclassification of samples *S-24* and *S-25*

The classification performance of most classifiers was relatively similar with respect to accuracy, F-score and the Kappa coefficient although the GP-OAD method performed slightly better on average. Some samples of rock were consistently misclassified by all methods, suggesting that these samples may not actually belong to the classes to which they were assigned during the ground-truthing process. For example, pixels of rock sample *S-25* (see Figure 6.4 and 6.8) were misclassified by all classifiers. This particular rock sample was assigned manually to the class of GOL on the basis of its visual appearance and chemical composition determined by XRF (Table 6.4). It was, however, classified as CHT using SVMs and as BIF using GPs and SAM. Because such inconsistent results made it difficult to judge the relative performance of classifiers, it was necessary to further investigate the reasons for this inconsistency with respect to the samples identity.

The chemistry of the *S-25* sample suggested that it was assigned to the GOL class (Table 6.4) because its chemistry was within the standard deviation of the average of the major components. These were content of Iron, Silicon-dioxide and Aluminium-(III)-Oxide and the amounts of bound water as determined by loss on ignition analysis (LOI). Sample *S-24* was assigned to the same class (i.e. GOL) on the same basis. The chemistry of *S-24* was very similar to *S-25* in terms of the amounts and proportions of its elements. The LOI of *S-24* was slightly greater than for *S-25*, i.e. 11.11 and 10.65 w%, respectively.

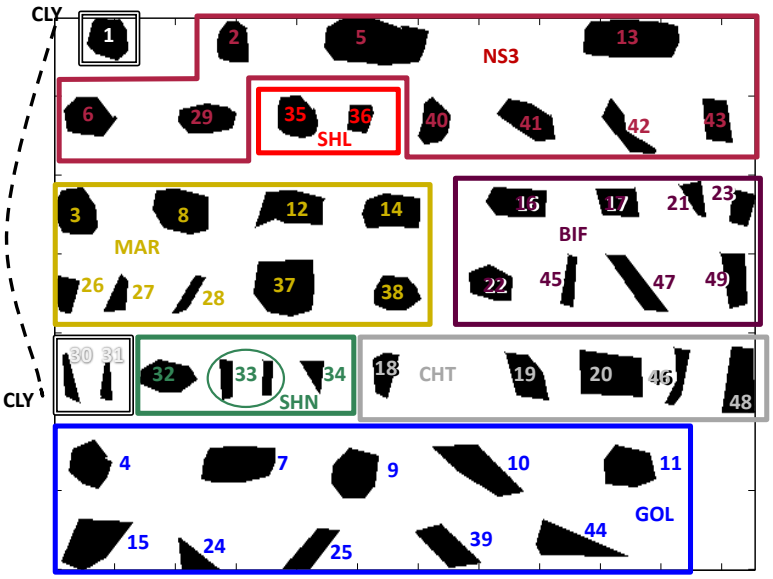


Figure 6.8 – Overview of rock samples used in Experiment 5. Coloured boxes indicated which sample belongs to which class. The colour of the outline of each box corresponds to the class label printed in the same colour (e.g. SHL or MAR).

Materials and Methods

To investigate the causes if the misclassification of *S-25*, four types of analysis were done: (i) inspection of the spectral variability among different pixels. (ii) Comparison of the spectral curve shape using visual inspection and numerical analysis using the spectral angle, (iii) examination of the regression values obtained using the GP-OAD for both samples and (iv) determination of the minimum of the major iron absorption feature, this feature is important as it can shift position according to the amount of goethite in the sample (referred to as F_{900} according to Murphy et al., 2013). To extract image spectra from the image (246 spectra from *S-24* and 396 from *S-25*), a region of interest (ROI) was placed over each of the two samples covering most of their surface area. Spectra were extracted from these ROIs for further analysis.

Results and Discussion

The variability of extracted spectra was relatively small for the *S-24* and *S-25* samples compared to the variability within the GOL class of the training library (Figure 6.9). A large within-class variability in the spectral library was, however, intentionally incorporated during its construction so that the natural variability of each rock type was captured. Visually, there were distinct differences in the spectral curve shapes of samples *S-24* and *S-25* (Figure 6.10). Sample *S-25* had a spectral curve shape that was distinctly different from sample *S-24* and the GOL library spectrum. The latter two spectra were almost identical, especially after differences in the background brightness (continuum) were removed (Figure 6.10 b). Numerical evaluation of the similarity using the spectral angle of these samples confirmed observations made by visual inspection (Table 6.5). A very small spectral angle (0.016 rad) was calculated between the *S-24* sample spectrum and the GOL library spectrum. In contrast, relatively large spectral angles were calculated between *S-24* and *S-25* (0.143 rad) and between *S-25* and the GOL library spectrum (0.141 rad). These results showed, visually and numerically, that the sample *S-25* was most likely not a true representative of the GOL class or it contained more vitreous than ochreous goethite as indicated by the higher SWIR reflectance values of *S-24* (Figure 6.10), possibly due to higher amounts of amorphous silica (Tables 6.4 and 6.6).

Table 6.4 – XRF results for *S-24* and *S-25* including mean and standard deviation for the GOL class. Numbers are given in relative abundance. Values are in weight percent (w%).

Sample	Fe	SiO2	Al2O3	MnO	MgO	K2O	Cu	LOI140-425	LOI425-650	LOI1000
S24	57.89	3.44	1.78	0.02	0.09	0.004	-0.001	10.42	0.54	11.11
S25	58.57	2.98	1.74	0.02	0.08	-0.001	-0.001	9.97	0.56	10.65
GOL mean	58.54	2.90	1.92	0.02	0.11	0.00	0.00	10.07	0.55	10.76
std	0.57	0.46	0.26	0.01	0.03	0.00	0.00	0.33	0.02	0.36

The GP-OAD provides a class estimate and uncertainty for each test point in a test set (e.g. a pixel in an image), it was thus used to assess the confidence of the classified results for rock samples *S-24* and *S-25*. For a correct classification of GOL, the predicted regression values of the GP-OAD should be around '-1' or at least yield the lowest values among the eight 'One versus All' - classifications. This is because

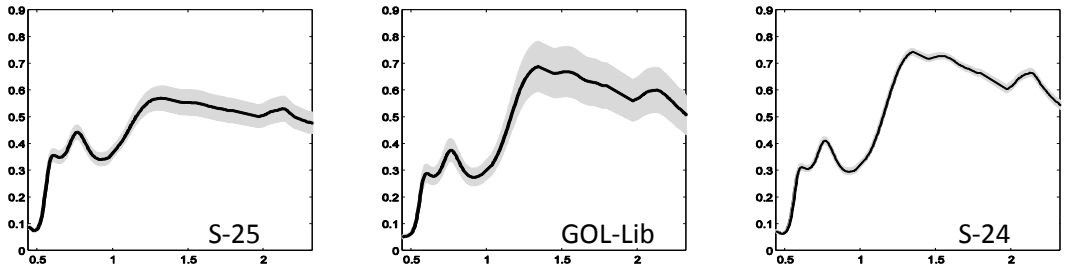
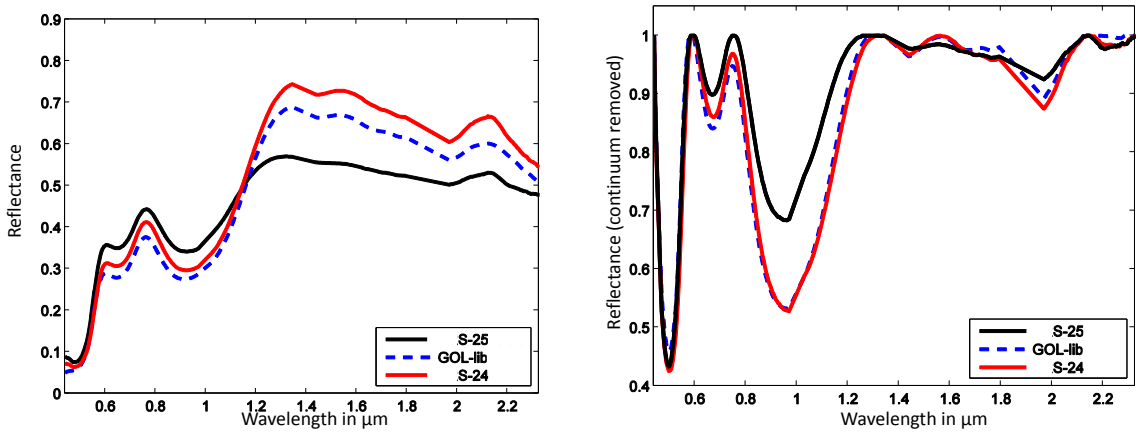


Figure 6.9 – Spectral variability and comparison of spectra for the misclassified sample *S-25* (left), GOL library spectra (from training set; middle) and sample *S-24* (right). The black line indicates the spectral average and the shaded area is the standard deviation at a particular wavelength. Abscissa is in units of wavelengths in μm and the ordinate is in units of reflectance.

'-1' is the expected label or value for the 'One' class and the value '1' is the expected label for the 'All' class.



(a) Reflectance **(b)** HullQuotient

Figure 6.10 – *S-24* and *S-25* spectral comparison with the average of GOL spectrum of the class GOL from the training library in reflectance (a) and with the continuum removed (b).

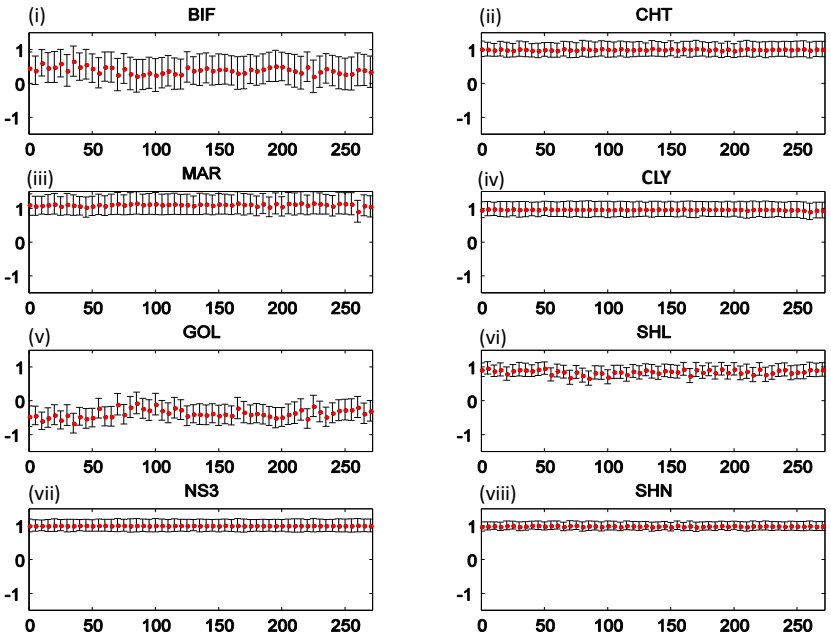
Table 6.5 – Spectral angles between samples *S-24* and *S-25* and the average of all GOL spectra from the training library.

	S-25	*GOL-Lib-ref*	*S-24*
S-25	0	0.141	0.143
GOL-Lib-ref	-	0	0.016
S-24	-	-	0

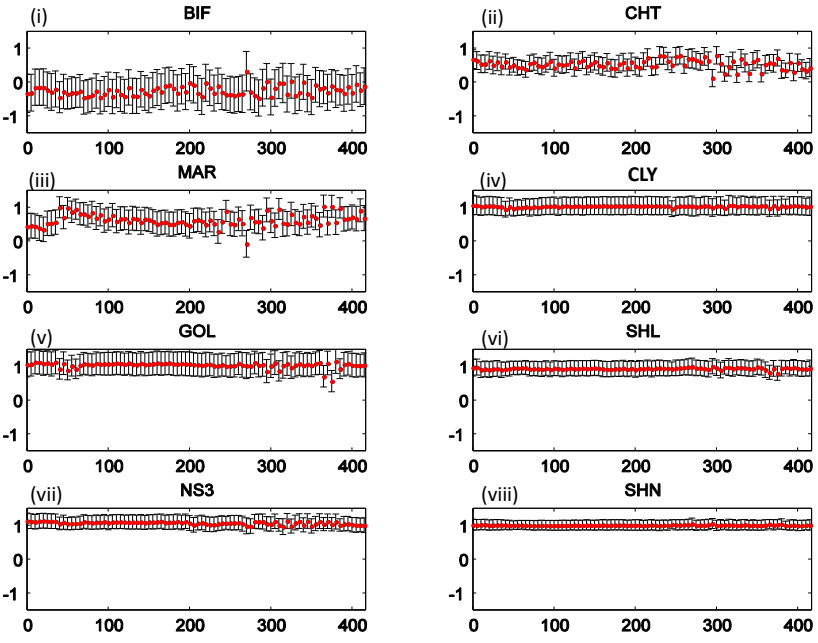
Values obtained from the GP-OAD for the GOL class when applied to *S-24* were around '-0.5' (6.11 a, panel v). These were by far the lowest predicted values associated with relatively small standard deviations (uncertainties) compared with other classes. Prediction values for all other classes (i.e. not GOL) were generally around '1' also with very small uncertainties (cf. panels in Figure 6.11 a). For BIF, however, values of '+0.5' with large uncertainties were obtained; these were the largest uncertainties across the eight classes. This indicates that the GP-OAD classification obtained for the *S-24* sample (Figure 6.11 a) can be regarded as a confident classification result. Because the predicted values for the BIF class were further away from '1' than other non-GOL classes, it can be assumed that there is some similarity between *S-24* and BIF or that the variability within the GOL class and within the BIF class of the spectral library are relatively large.

Values obtained from the GP-OAD for the GOL class, when applied to *S-25* were, however, around '+1' (6.11 b, panel v). Predicted values for all classes were close to '1', except for BIF (6.11 b, panel i). For the BIF 'One versus All' classification of the test image, values were slightly below zero, however, with the largest standard deviation. This indicates that the sample does not actually belong to the GOL class.

Analysis of the wavelength position of the ferric iron feature near 900 nm showed that its wavelength position was different for *S-24* and *S-25*. Cudahy and Ramanaidou (1997) showed that with decreasing amounts of goethite and increasing amounts of hematite the F_{900} feature shifts to shorter wavelengths (see also Townsend, 1987). This enabled accurate estimation of hematite/goethite ratios from reflectance spectra.



(a) *S-24*



(b) *S-25*

Figure 6.11 – *S-24* and *S-25* predicted regression values (red dots) and uncertainties (black error bars) obtained by the GP-OAD method. Each panel shows the predicted value, i.e. the value which determines if a pixel is of class '-1' or '1'. The uncertainties are shown as one standard deviation either side (i.e. $\pm\ 1 \cdot \sigma$). Note only a fifth of the pixels contained in each sample ROI is shown for clarity. The Abscissa denotes the number of pixels in the ROI, the ordinate the predicted value obtained by the GP-OAD.

To use the position of the minimum of the F_{900} feature in hyperspectral imagery to determine if the spectral difference of two samples (*S-24* and *S-25*) was statistically different, an interpolation method was applied over the F_{900} feature between 770-1150 nm, i.e. the spectral range describing this feature. This was necessary because of relative large amounts of noise in the data in this spectral region. A fifth order polynomial (following to Murphy et al, in review) was fitted over the F_{900} feature which gave very consistent results compared to the raw data (example shown in Figure 6.12).

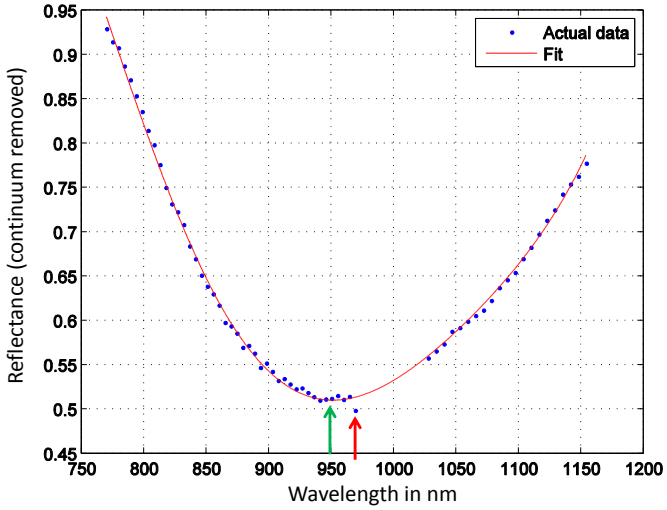


Figure 6.12 – Polynomial fit of the major iron absorption (F_{900}) for the *S-25* spectrum. The red arrow indicates the wavelength position of the iron absorption feature minimum using the raw data while the green arrow indicates the wavelength position of the same feature using the fitted data (5^{th} degree polynomial).

Figure 6.13 (a) shows hull quotient spectra extracted from both samples (*S-24* and *S-25*). Spectra of each rock type were clearly separated from each other with respect to albedo and the locations of the minimum of the F_{900} feature (coloured dots). The difference becomes clearer in Figure 6.13 (b) where the black lines indicate the mode of each distribution, i.e. the wavelength position of the minimum of the F_{900} feature with the highest count. The histograms of the wavelength position of F_{900} (Figure 6.13 c), derived from a 5^{th} degree polynomial fit to the original data, showed a bimodal distribution which enabled a separation of rock types and suggests that both samples

were potentially different from a mineralogical point of view. The distributions of the determined F_{900} feature for samples *S-24* and *S-25* were distinctly different as indicated by the modes with 949 nm (*S-24*) and 939 nm (*S-25*). From these data, it was shown that *S-25* had more hematite in the sample than did *S-24*.

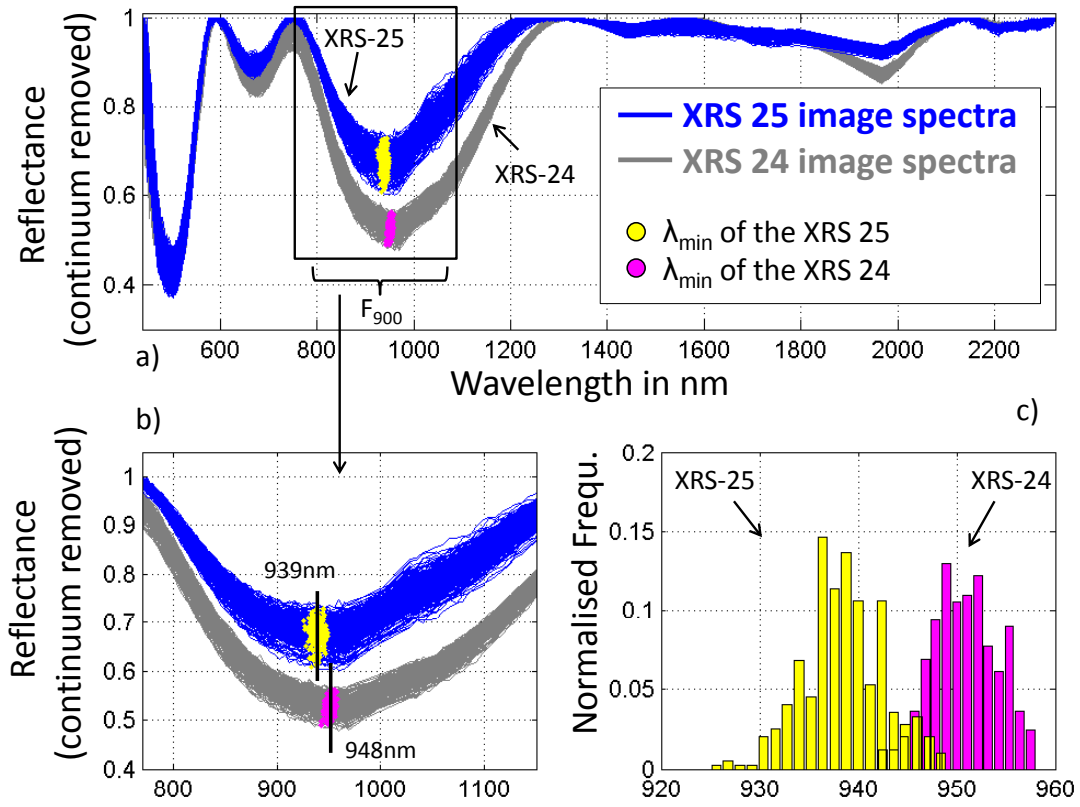


Figure 6.13 – Comparison of wavelength position between sample *S-25* and *S-24*. (a) shows all spectra of each of the samples plotted in hull quotient reflectance as well as the wavelength position of the major iron absorption feature minimum (dots) for each spectrum. (b) shows a zoomed version of the wavelength region of interest (770-1150 nm). The black line in each of distributions of dots indicates the mode. (c) shows a histogram for both samples resulting in a bimodal distribution. Both histograms were normalised to the number of pixels in each region of interest representing each rock sample.

To confirm the hypothesis that both samples have different mineralogical compositions, *S-24* and *S-25* were sent for XRD analysis. XRD assays confirmed that sample *S-25* contained hematite (Table 6.6). This explains the shift in the wavelength po-

sition of spectra from sample *S-25* relative to sample *S-24*. This shift is caused by inclusion of the iron mineral hematite in sample *S-25*. The presence of hematite would be enough to cause changes in spectral curve shape through changes in the wavelength position of F_{900} and changes in brightness in some parts of the spectrum as a consequence of transopaque behaviour (Ramanaidou et al., 2008).

Table 6.6 – Summary of XRD results for samples *S-24* and *S-25*. Numbers are given in weight percent (wt%). Amorph. is the amorphous content of each sample.

sample	Goethite	Hematite	Amorph.
S-24	92	0	8
S-25	90	4.3	5.7

XRD data show that although the Goethite content was nearly the same for both samples, there were some amounts of hematite (4.3 wt%) in sample *S-25*. Murphy et al. 2013 showed that the wavelength position of the F_{900} can change with the addition of only small amounts of goethite and hematite. While they investigated mixtures by adding goethite to pure hematite, it can be assumed that the reverse relationship is also statistically significant. An ANOVA test confirmed that the two samples had statistically different F_{900} wavelength positions ($P < 0.0001$).

This analysis shows the advantage of the staged experimental framework used in this chapter. Applying classifiers directly to field imagery without first testing them on laboratory-acquired data would not have permitted such deliberate analysis of their performance.

6.4 Application of the GP-OAD to mine face imagery acquired in the field - Experiment 6

Mapping geology and mineralogy on vertical mine faces at the required spatial resolution and orientation requires that hyperspectral imagery is acquired from field-based platforms. Open pit mines, however, pose significant challenges to the acquisition

and analysis of hyperspectral data. The complex geometry and angular protrusions on the mine face can cast deep shadows and the changing position of the sun with respect to the mine face can cause large variability in incident illumination during the course of the day. Acquisition of the data themselves is more difficult in a dynamic environment such as an open pit mine. Imagery acquired from a mine face differs from data acquired in the laboratory in several key aspects.

Firstly, setting the appropriate integration time used to acquire the data is critical in order to acquire data of the highest possible quality. An integration time must be set for the VNIR and SWIR sensors, so that the brightest objects in the scene are not saturated. This is problematic in field settings due to the rapidly changing conditions of illumination. Furthermore, in the context of iron ore mines, minerals display large variability in spectral brightness - for example, goethite is very bright and hematite is very dark. Setting an appropriate integration time to ensure that pixels dominated by goethite do not saturate, by default, is inappropriate for dark pixels dominated by hematite. The consequence of this is that pixels dominated by hematite are underexposed compared to pixels dominated by goethite. Thus, the former have a smaller signal-to-noise ratio (i.e. they are noisier) than the latter. Secondly, the ground pixel size of images acquired in the field is much larger then laboratory-acquired imagery. That is, each pixel in the image can cover an area of several tens of centimetres in size. This reduced spatial resolution relative to laboratory-acquired imagery increases the likelihood that several minerals and/or materials may be contained within each ground pixel.

The spectral signals of all component materials in a pixel are averaged or convolved into a single component pixel spectrum. This composite pixel spectrum may not resemble any of the training spectra. Thirdly, data are usually acquired from a mine face using natural (solar) illumination. Absorptions by atmospheric gasses and aerosols increase noise in certain parts of the spectrum by reducing the amount of incident light. These spectral regions may contain important spectral information about minerals.

Finally, field imagery are acquired by mounting the sensors on a rotating stage in

order to build up the spatial dimension of the image. The rotational movement of the sensors significantly complicates the spatial registration of the VNIR and SWIR (Nieto et al., 2010). All these factors make the classification of imagery acquired in the field more challenging than for laboratory imagery.

For hyperspectral data to be used for autonomous mining, classification algorithms need to conform to several key requirements: (i) they must provide a probabilistic output so that maps can be fused with data from other sensors; (ii) algorithms must yield consistent results under variable conditions of illumination (including shade and shadow) and (iii) enable minerals and rocks to be classified without *a priori* knowledge of any particular mine face, i.e. classification must be possible using an independent set of training data.

6.4.1 Materials and Methods

Hyperspectral imagery was acquired from two vertical mine faces (Figures 6.14 and 6.15 and Table 6.7) in the WA mine from a distance of 30 m, using separate visible infrared (VNIR; 400-970 nm) and shortwave infrared (SWIR; 971-2516 nm) sensors (Specim, Finland) mounted adjacently on a rotating stage. The VNIR and SWIR imagers had full-width at half-maximum spectral resolutions of 4.6 and 6.4 nm, respectively. The spatial resolution was 6 cm (VNIR), and 12 cm (SWIR). A calibration panel (100% reflective Spectralon; 30 cm by 30 cm) was placed in front of the mine face during image acquisition. Integration time of each sensor was adjusted so that the brightest object within the scene did not saturate. The first of the two mine faces ('WA-MF1'), displayed several exposed geological zones characteristic of the West Angelas ore-body. This image was used to test how well the classification algorithms would perform to distinguish different geological zones and classify geology on the mine face. The second mine face ('WA-MF2'), had a variable surface topography and was oriented so that it experienced large variations in the angle and amount of incident illumination during the course of the day. Thus, imagery was acquired at two different times during the day, at 11 A.M. (WA-MF2a) and 1:30 P.M. (WA-MF2b)

local time. The WA-MF2b image had areas of shade which had been fully illuminated in the WA-MF2a image. These images were used to test the effects of shade on the classification performance of algorithms.

All image data were calibrated to reflectance using a standard method by dividing each image pixel by the average pixel spectrum (Equation 6.3) over the calibration panel (in accordance with Murphy et al., 2008). Wavelengths affected by atmospheric absorptions were removed from the image spectra, as were noisy bands at the extreme ends of spectra and at the junction of the two sensors. Calibration was done using:

$$\rho(\lambda)_{norm} = \frac{DN(\lambda) \cdot \rho\,'(\lambda)_{WR}}{DN(\lambda)_{WR}} \ , \tag{6.3}$$

where $\rho(\lambda)_{norm}$ is the reflectance of a mine face image at band λ, $DN(\lambda)$ is the digital number (DN) of an individual pixel in the image, $DN(\lambda)_{WR}$ is the average of all pixels over the calibration panel in the image and $\rho\,'(\lambda)_{WR}$ is the reflectance factor of the particular calibration panel used at wavelength λ.

For the classification of the imagery, an independent training set, i.e. the *WAcoreLib1* library which was acquired under artificial illumination from cores of rock (Section 4.2) was used.

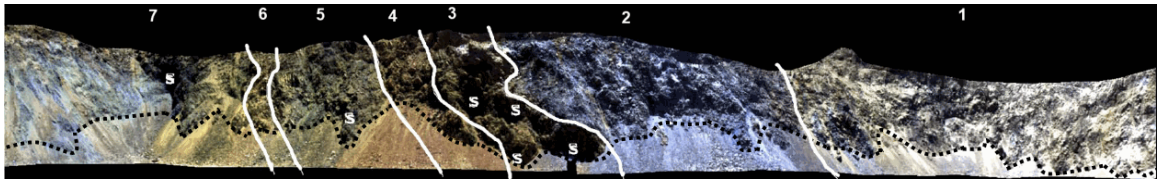
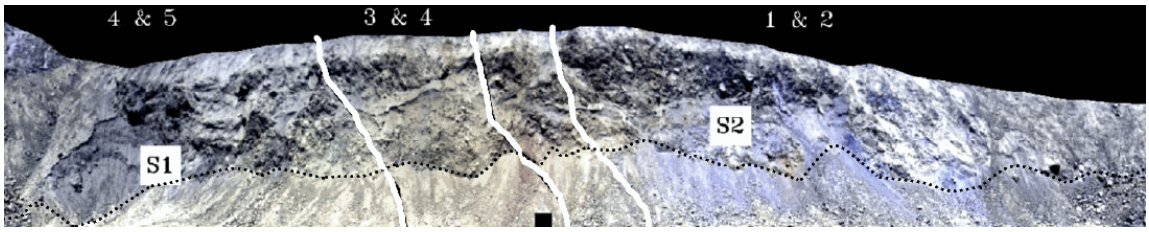
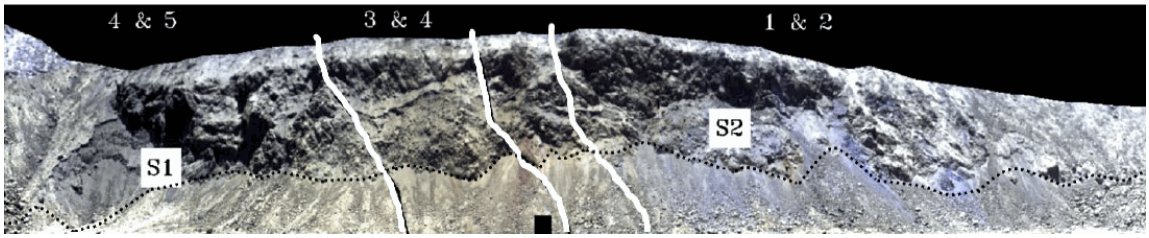


Figure 6.14 – Contrast-enhanced true-colour composite with overlaid geological boundaries mapped in the field. Geological zones are indicated by numbers: (1) Shale; (2) Manganiferous shale; (3) Mixed shale and ore zone (Goethite, Shale, and Martite); (4) Ore zone (dominant Goethite with Martite background); (5) Ore zone (dominant Martite with Goethite background); (6) Ore zone (dominant Goethite with Martite); (7) Ore zone (dominant Martite with Goethite background, Chert and BIF). Shaded areas indicated by "S". Black dotted line marks the boundary between rill (loose rocks) and the ore-body.



(a) WA-MF2-fully illuminated, 11 am local time



(b) WA-MF2-shade, 1:30 pm local time

Figure 6.15 – Contrast-enhanced true-colour composites of WA-MF2. Images were acquired at different times of the day. Areas which are affected by shade are indicated by "S1" and "S2". Zones 1 & 2 contain mainly SHL with some patches of SHN; Zones 3 & 4 contain MAR and GOL and Zones 4 & 5 contain mainly MAR. Black dotted line marks the boundary between rill (loose rocks) and the ore-body.

Table 6.7 – Data-sets used in Experiment 6. Atmospheric water absorption bands were removed for both data-sets. Additional bands at the extremes of the spectra were also removed.

Data-sets			Comments
Training set	*WACoreLib1*	*WACoreLib1*	*WACoreLib1* is comprised of 228 spectra in 8 classes. See Section 4.2
Test set	*WA-MF1*	*WA-MF2(a)*	Hyperspectral images from two mine faces. Two images are available for MF2.
Bands	283	283	Wavelengths range was between 439-2325 nm

In this section, data from a mine face, acquired under natural illumination, were classified using only the GP-OAD, GP-SE and SAM (using the minimum angle criterion). This was done to compare the performance of methods when applied to hyperspectral data acquired in the field using natural illumination. The GP-OAD was compared to the GP-SE and SAM because the GP-OAD provided the best classification performance in previous experiments and it is subject of this thesis. The GP-SE performed relatively well using laboratory data, but was expected to have a decreased perfor-

mance when there were large variabilities in brightness. The GP-SE was also selected because it uses a distance measure rather than an angular measure to calculate similarities between spectra. The spectral angle mapper was selected because, like the GP-OAD, it should be relatively insensitive to differences in brightness. SAM (m.a.) is commonly used for the classification of hyperspectral imagery and was thus used as a reference for comparison. Classification using either algorithm was done using the same methodology as applied in Experiment 5 (Section 6.3) using laboratory imagery. Class labels were assigned using the same criteria. SVMs were excluded from this analysis as they performed relatively poorly in previous experiments and thus similar poor results were expected for this experiment. The Neural Network (NNET) kernel was excluded because of similar performance to the OAD kernel in previous experiments which was based on the same rationale, which was discussed in Experiment 5.

The performance of these classifiers was assessed by applying them to data which contained the combined effects of: (i) Using different types of sensors for the acquisition of the training and test sets. The training set (*WAcoreLib1* library) was acquired using an ASD field spectrometer. The test set, i.e. imagery of mine faces, was acquired using two separate imaging spectrometers. (ii) Using different sources of illumination. The training set was acquired using artificial illumination and the test set was acquired using natural (solar) illumination. (iii) Using different types of materials for the acquisition of training and test sets such as drill cores (*WAcoreLib1* training set) and whole or bulk rocks on a mine face (test set). (iv) Using a test set that contains pixel spectra from shaded rocks. This is a particular challenge for the different classifiers because the training set does not include training examples of spectra from shaded rocks (Figure 6.16).

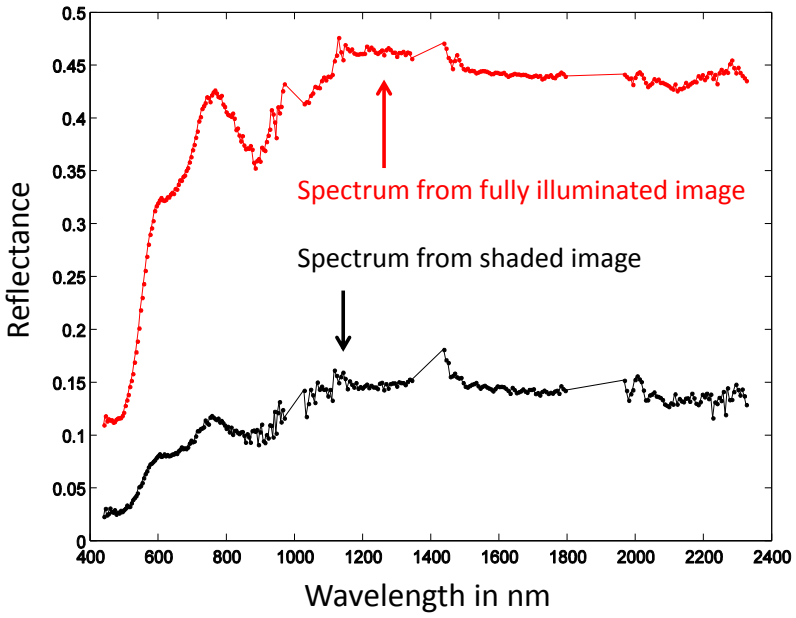


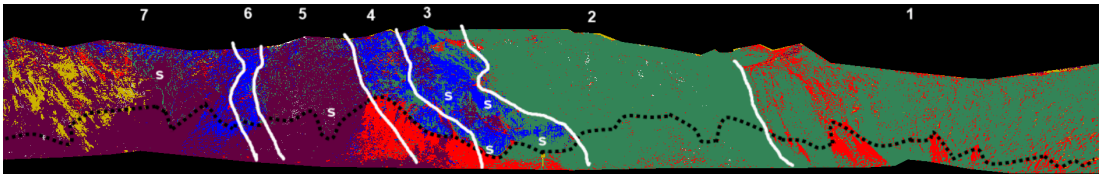
Figure 6.16 – Spectral difference between a shade and no-shade pixel. Changes in brightness and curve shape due to effects of shade. The two spectra are from the same pixel from an area of the mine face which was fully-illuminated in the no-shade image but became fully-shaded in the shade image.

6.4.2 Results and Discussion

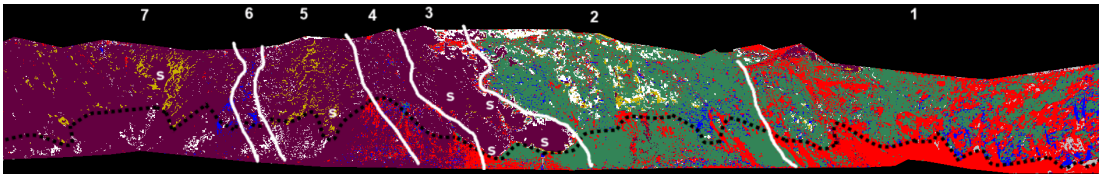
Mine face 1

Classification results obtained from the three methods (GP-OAD, GP-SE and SAM) were relatively similar for this mine face (Figure 6.17). Classifications using the GP-OAD and SAM yielded relatively similar results, both methods classified pixels in Zone 7 as MAR and BIF which was mapped during field surveys. The GP-SE classified Zone 7 nearly entirely as BIF with some isolated patches of MAR. Zone 6, mapped in the field to be mainly GOL with some MAR, was classified correctly in most areas by SAM and the GP-OAD as GOL, however, no MAR was detected by either method and some BIF was classified as false positive at the bottom of the mine face (rill). The GP-SE classified this area as BIF, however, with very small and isolated patches of GOL. Zone 5 was misclassified by all methods as BIF, however, MAR and GOL

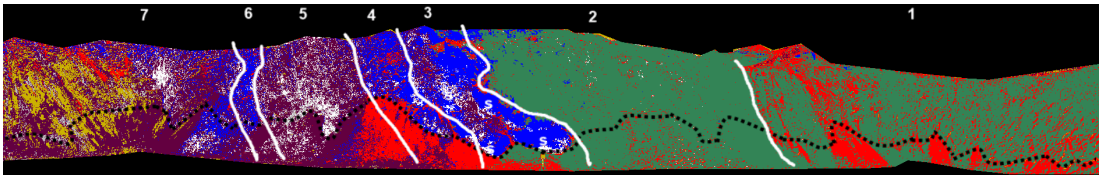
were expected from field geological maps made in the field. Classification of Zone 4 appeared to be visually correctly classified by the GP-OAD and SAM, however, both methods showed small and isolated patches of SHN (GP-OAD) and CLY (SAM) which was not mapped in the field. The GP-SE did misclassify Zones 4 and 3 as BIF. The GP-OAD and SAM, however, obtained correct classification results for most of Zones 4 and 3. Areas of misclassification coincided with areas affected by shade (marked with 'S' in 6.14 and 6.17). Zones 2 and 1 were mostly classified correctly by all methods, however, a large variability in the classification of Zone 2 was observed for the GP-SE. The GP-OAD and SAM gave very similar results, but showed less variability in classes of pixels in these zones than the GP-SE.



(a) GP-OAD classification



(b) GP-SE classification



(c) SAM classification

BIF CHT MAR CLY GOL SHL NS3+4 SHN

(d) Class labels

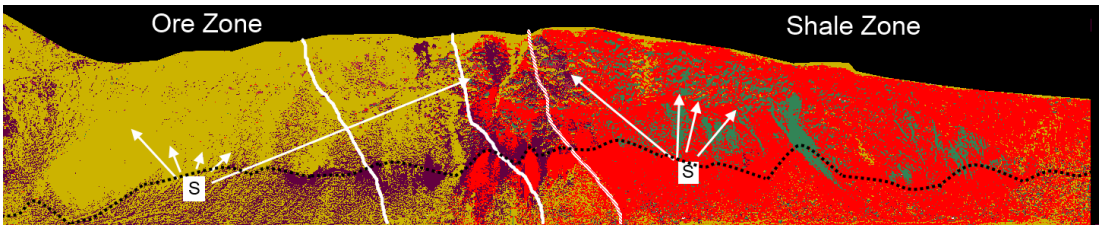
Figure 6.17 – Classification results of WA-MF1. The *WAcoreLib1* library acquired from drill cores was used as the training set / spectral library to produce these results. Zones (1-7) are in accordance with geological maps made in the field (e.g. Figure 6.14. 'S' indicates shaded areas.

Mine face 2

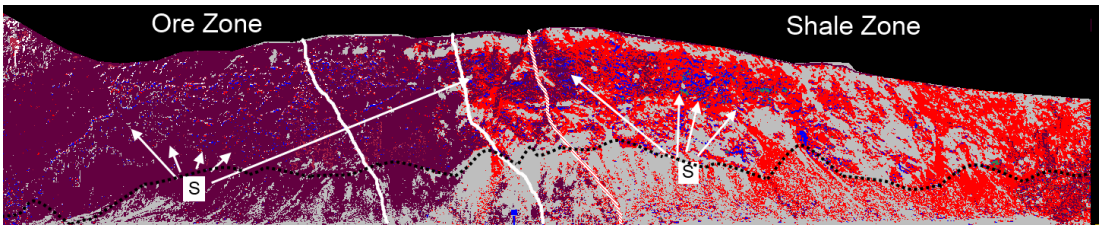
In the second investigated mine face, the GP-OAD performed relatively well on the unshaded image, i.e. ore and shale were identified correctly. The SAM classification showed a similar result to the GP-OAD in the shale zone, however, high-frequency spatial variability was observed in the ore zone. Some pixels in the ore zone were classified as BIF, however, this could not be confirmed from the geological map made in the field. The classification of the classes SHL and SHN in the shale zone were very similar between the GP-OAD and SAM results (cf. Figures 6.18 a and c) and were in closer agreement with the geology mapped in the field (Figure 6.15). However, no GOL was mapped using neither GP-OAD nor SAM, thus misclassifying on group of ore.

The GP-SE (Figure 6.18 b) distinguished the ore-zone from the shale zone, however, using mostly the wrong classes for either zone, and thus the classification of the mine face was inconsistent with the geology mapped in the field. Specifically, the entire shale zone was misclassified as CHT and some patches of SHL. Moreover, similar to SAM, the entire ore zone was misclassified as BIF with high-frequency spatial variability of GOL and SHL. The results from the GP-SE were entirely inconsistent with the known distribution of ore and shale mapped in the field.

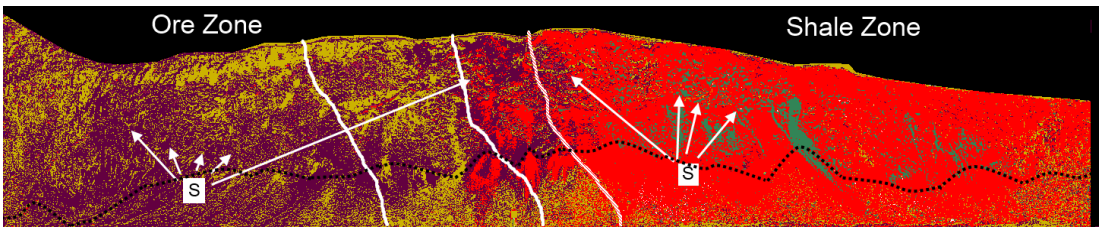
Because the images were of the same mine face, no changes in the distribution of pixel classes between the images acquired at different times of the day were expected. SAM gave broadly consistent classification results for both images (cf. Figures 6.18 and 6.19); there was, however, high-frequency spatial variability, particularly in the shaded image. A large number of pixels, however, changed class in areas which were not shaded in the no-shade image but which subsequently became shaded in the shade image (Table 6.8). On a pixel-by-pixel basis more than one third (36 %) of pixels changed class using SAM (Table 6.8).



(a) GP-OAD classification



(b) GP-SE classification



(c) SAM classification

BIF CHT MAR CLY GOL SHL NS3+4 SHN

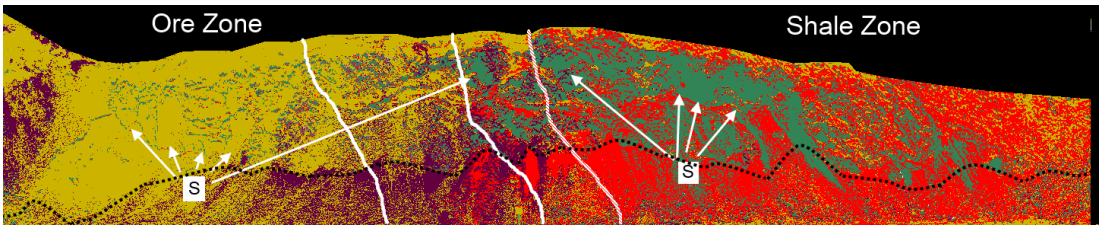
(d) Class labels

Figure 6.18 – Classification results of WA-MF2 using the no-shade image. The *WA-coreLib1* library acquired from drill cores was used as the training set / spectral library to produce these results.

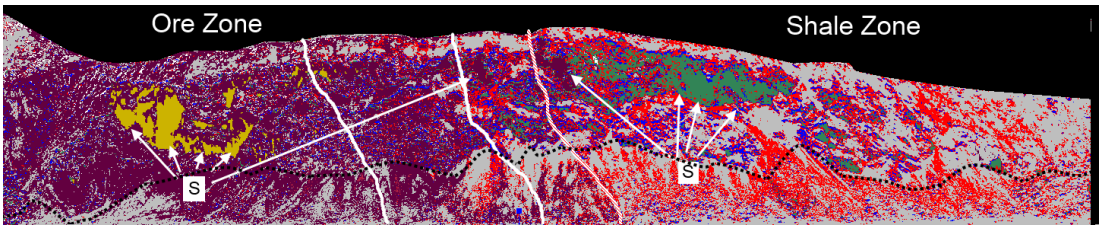
Table 6.8 – Number of pixels that changed class between the no-shade and shade classification are shown. Absolute and relative values are shown. The total number of pixels in each image was n=417,027. Absolute and relative change of pixels is also shown for Median filter processed images with a 5×5 filter mask.

Method	GP-OAD	GP-SE	SAM
n pixels original	130,301	130,300	149,855
pixel change (%)	31.2	31.2	35.9
n pixels median	62,254	102,501	102,503
pixel change (%)	14.9	24.6	24.6

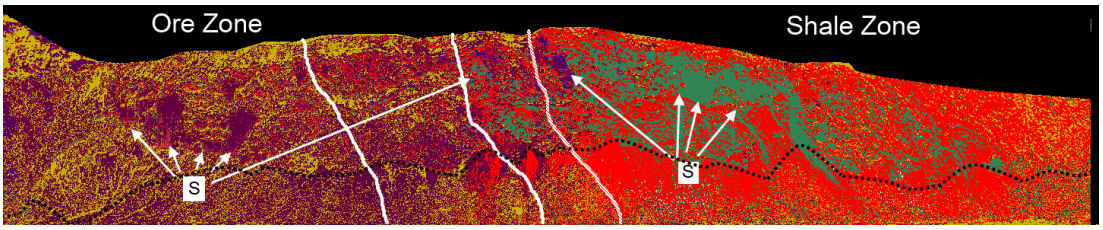
For the GP-SE and the GP-OAD the same number of pixels changed classes between the shade and no-shade images (31 %). In order to reduce the influence of high-frequency spatial variability and the influence of changes of individual pixels on the expressive power of the quantitative results, the same assessment was repeated using a smoothed version of the classified images. A 2D median filter with a 5×5 smoothing window was applied to the classified shade and no-shade image before calculating the changes in class on a pixel-by-pixel basis. This analysis suggested that the changes in class occurred predominantly for individual, not spatially grouped pixels using the GP-OAD. Only 15 % of the pixels changed class after application of the median filter to the classified image. For the GP-SE and SAM, however, 25 % of the pixels changed class between the median filtered shade and no-shade images. This indicates that the pixels classified by the GP-SE and SAM were grouped together. These groupings were in the discrete areas of the mine face that had been impacted by shade and which changed class between the shade and no shade images. Analysis of images classified by the GP-SE and SAM showed that pixels which changed class between the no-shade and shade images were grouped together in areas of the image which were affected by shade. For the GP-OAD classified images, changes of pixel from one class to another were not localised at these areas but changes occurred mostly due to factors other than shade. Although only 6.6 % and 11.3 % of pixels changed classes due to factors other than shade for the GP-SE and SAM, respectively, both methods did misclassify a large proportion of pixels as a consequence of shade. The GP-OAD was less affected by shade as evidenced by the fact that shaded areas were only partly misclassified (e.g. only within the shale zone).



(a) GP-OAD classification



(b) GP-SE classification



(c) SAM classification

(d) Class labels

Figure 6.19 – Classification results of WA-MF2b using the shade image. The *WAcore-Lib1* library acquired from drill cores was used as the training set / spectral library to produce these results. 'S' marks areas affected by shade.

The GP-OAD produced results which were qualitatively and quantitatively similar, with 15 % of pixels changing class between the no-shade and shade image. Crucially, however, classification changed only slightly in areas of shade within the ore zone from MAR to SHN and these were isolated pixels only. Slightly more change between classes from SHL to SHN were observed in the shale zone (cf. Figures 6.18 a and 6.19 a). The GP-SE consistently misclassified the mine face for the shade image, however, classes changed distinctly in areas affected by shade. Coincidentally, pixels

changed classes from the wrong class in the no-shade image to the correct class in the shade image. The classification of the shade image obtained using SAM was largely consistent with the classification obtained using the no-shade image.

Classified maps from *WA-MF1* were relatively consistent with geological maps made in the field for all classifiers. The GP-SE showed, visually, a classification which was in poorer agreement with maps made in the field than the GP-OAD and SAM. The GP-SE showed also more high-frequency spatial variability within pixels in most zones (e.g. Zones 1 and 2) on the mine face than the other two methods. This indicates that the method is more effected by small variations in illumination and / or debris and biofilm than the GP-OAD or SAM. Using the *WA-MF2* data-set, the GP-SE produced classifications which were entirely inconsistent with the geology mapped in the field. Qualitatively it performed very poorly compared to SAM and the GP-OAD. This highlights an inherent problem using the GP-SE with independent libraries acquired using different sources of illumination because it cannot compensate for offsets in brightness between spectra of the same class (rock) in the training set and the test set.

There were large variations in illumination within and between images which caused the GP-SE to, almost consistently, misclassify geology using the *WA-MF2* data-set. Examination of the proportion of pixels which changed class between the shade and no-shade image showed that a larger proportion of pixels changed class for the GP-SE (26 %) than for the GP-OAD (16 %) after spatially smoothing the classified images. In comparison to SAM, the GP-OAD produced a classification that was less sensitive, qualitatively and quantitatively, to changes in illumination.

These results highlight two important issues: The SE covariance function is based on the distance of two vectors, therefore, if test data have the same type of illumination as the training data, classification will be relatively accurate (as seen in Experiments 1-4). If, however, the test set shows different properties, such as an offset due to brightness differences, classification results degrade. Because SAM and the OAD covariance function use an angular metric rather than a distance metric (such as the SE), they have a higher chance of successfully classifying data with different brightness.

This means an independent training library acquired with a non-imaging field spectrometer using artificial illumination can be used to classify hyperspectral imagery acquired in the field using natural sunlight using spectral angle based methods.

6.5 Discussion and Conclusions

This chapter extended the use of the GP-OAD to hyperspectral imagery and applied it to such data for the first time. The GP-OAD was applied to field spectra which were convolved to the bandpasses of an imaging scanner to determine the effects of reduced spectral resolution on the classification performance (Experiment 4). The GP-OAD was applied to hyperspectral imagery acquired under artificial illumination with a relatively high SNR (Experiment 5). The GP-OAD and other classifiers were then applied to hyperspectral imagery of vertical mine faces acquired using natural illumination (Experiment 6).

Results of Experiment 4 showed that the GP-OAD outperformed the other methods on most occasions with respect to Accuracy, F-score and Kappa measures. Results also showed that the GP-OAD was least affected by the number of spectral bands used. It performed better using more bands (i.e. 2151 bands compared to either 359 or 283 bands), however, the difference in performance to when fewer bands were used was smaller compared to other methods. The performance of the GP-OAD seemed unaffected when water bands were removed as the shown by a slight increase in performance after removal of these bands. This is also true for the other GP methods using the SE and NNET kernel. The GP-SE failed to deliver any useful classification results for data-sets with 2151 and 359 spectral bands but performed slightly better using data-sets with 283 bands. The performance of the SVM-SE was still very poor compared to the other classifiers. This suggests that the dimensionality of the data has larger impacts on its performance than on other methods (e.g. GP-OAD and SAM). SVMs using hard decision boundaries performed poorly when applied to hyperspectral data, thus probabilistic SVMs, using Platt's probabilities, were used. A reasons why some classifier performed better with less spectral bands could be that

neighboring bands are correlated and not entirely independent from each other which provides redundant information. This can cause classifier to perform worse even though 'more features' are used. This is often referred to as 'Hughes phenomena' or the 'curse of dimensionality' in the literature (Hughes, 1968, Bellman, 1957).

In Experiment 5, the training and test sets were acquired with different sensors under different illumination geometries. Both ASD spectra and hyperspectral imagery were acquired using artificial illumination. The GP-OAD classified several types of rock correctly but many samples of CHT were misclassified as BIF. One explanation for this is that BIF is mineralogically similar to CHT, except that it contains less silica[2]. SVMs using hard decision boundaries performed poorly when applied to hyperspectral imagery, instead, probabilistic SVMs using Platt's probabilities were used. In this implementation the SVM using the NNET and OAD kernels performed nearly as well as the GP-OAD and outperformed the SE kernel. This showed that SVMs using hard decision boundaries fail when applied to hyperspectral data, even when the training and test set were acquired under ideal laboratory conditions. The spectral angle mapper using a minimum angle criterion outperformed the SVMs by a small margin. The classification performance of the GP-SE was poorer than other GP methods and SAM. The SVM-SE showed the poorest performance. This suggests that the SE kernel can perform relatively well in cases where training and test sets are acquired using different types of sensors but using the same light source, but only if the framework models the noise between the different data-sets appropriately (e.g. using the GP-SE). Angular based methods (e.g. the GP-OAD and SAM) can compensate for slight differences in brightness (e.g. albedo offsets) due to different illumination geometries much better than distance based kernels and are therefore less affected by effects of variability in brightness and illumination.

Some rock samples from laboratory imagery (Experiment 5) were misclassified by all methods (e.g. *S-25*, which was determined to be of class GOL according to XRF data and visual inspection). Detailed analysis of one of these rock samples was done to determine the cause of their misclassification. It was shown that, although the

[2]Silica is a key parameter for the processing of ore, however, both CHT and BIF are considered waste which is difficult to process.

chemistry was within the standard deviation of the sample mean of the chemistry of all the GOL samples, the mineralogy was different from other GOL samples. Small amounts of hematite were detected in the sample (using XRD) in addition to the major component goethite. This presence of the hematite caused a significant shift in wavelength position of the major iron absorption feature as well as changes to the curve shape, leading to consistent misclassification of the sample. These observations also raise questions as to whether the training set was influenced by outliers due to the variability in the quantities of minerals within some classes. The large variability in the training set was, however, designed to incorporate the natural within-class and between-class variability of the rocks. No outliers were removed on the basis that all spectra in the training set should represent the natural variability of a specific rock type on a mine face.

In Experiment 6, the performance of classifiers, including the GP-OAD, was compared by applying them directly to hyperspectral imagery acquired in the field. This was the most difficult test for all the classifiers due to the smaller number of bands, increased noise and other artifacts present in the field imagery. Results for the WA-MF1 data-set showed that the GP-OAD and SAM performed similarly, with both methods classifying most geological zones mostly correctly. There were, however, zones (e.g. Zone 5 on mine face 1) which were mostly misclassified as BIF by these methods. This was probably due to the large within- and between-class variability and the spectral similarity of the BIF and GOL classes. The GP-SE, however, performed well in some areas (e.g. Zones 1 and 2) but poorly in the remaining zones. Thus, results were visually inconsistent with the zones mapped in the field. Therefore the SE-methods seem at least good in mapping shales. This is possibly because of their similar spectral curve shape with rather shallow positive slope in the VNIR, when compared to other, iron oxide rich samples, which have usually a steep slope in the VNIR and then flatten (Figure 4.2). Using the *WA-MF2* data-set, the GP-SE also performed extremely poorly across the entire mine face in both the shade and no-shade images, with virtually all pixels incorrectly classified. Changes in pixel classes were relatively small between the shade and no-shade image, however, they

were localised in areas affected by shade. This indicates that the performance of the GP-SE was consistently poor across imagery acquired at different times of the day. SAM performed much better than the GP-SE by correctly classifying most of the shale zone in both shade and no-shade images. Similar to the GP-SE classification, changes in class of pixels for the SAM classification were localised in shade affected areas and a smaller portion of pixels changed class in areas which were not affected by shade. Regarding the different types of illumination, the GP-OAD performed markedly better than both the GP-SE or SAM. Changes in class between the no-shade and shade images occurred at the scale of individual pixels rather than clusters of pixels located in areas of the image affected by shade. For example, individual pixels changed to SHN in the ore zone on mine face 2 using the shaded image this may be due to coatings of dust from. This could be because the diagnostic clay features still appear in 'darker' and potentially shaded spectra (Figure 4.2). This indicates that the performance of the GP-OAD was less affected by shade than were GP-SE and SAM. Most of the ore zone and the shale zone were correctly identified using the GP-OAD. The performance was assessed qualitatively because the available ground-truth for both data-sets was spatially coarse.

Contrasting results were found in the performance of the GP-SE across Experiments 4, 5 and 6. The GP-SE performed relatively well in Experiments 4 and 5 using artificial illumination but the performance was, on average, very poor for imagery acquired from a mine face (Experiment 6). A reason for this might be the stable illumination conditions during the acquisition of data, in terms of a constant sensor-target-illumination geometry used in Experiments 4 and 5. For example, a white reference was taken prior to each spectral acquisition with an ASD spectrometer for the data used in Experiment 4.

In laboratory images (Experiment 5), calibration was done on a line by line basis. Separate calibration and target images were acquired in order to optimise the SNR of the spectral data. In addition, the target surface was relatively homogenous and mostly flat and was free of surface coatings. Mine faces in operational iron ore mines are different in that their surfaces are coated with dust and they have highly-variable

and complex geometries causing variability in illumination which are unrelated to the rock type or mineralogy. Furthermore, the calibration panel was placed in a fully-illuminated part in the middle of the mine face. Therefore, reflectance spectra of shaded pixels which were calibrated using the fully-illuminated calibration panel can cause offsets and changes in curve shape which are unrelated to the mineralogy of the rock. These are factors that lead to a poor classification performance. However, the main reason for the performance exhibited by the GP-SE using imagery from mine faces is due to the metric encoded within the kernel. This is a particular problem when training and test data-sets are acquired using different light sources (i.e. artificial and natural light). The Euclidian distance measure used in the SE kernel is unbound, thus the range of values for the distance between the training and test set can take values which are not representative of the distances between classes learned from the training set. This can lead to misclassification due to offsets in brightness and changes in curve shape in spectra from the test set relative to the training set.

Because of the sensitivity of the SE kernel to variability in illumination, the within image variability for a particular class is relatively large, as seen in Zones 1 and 2 in the *WA-MF1* image. The difference of classifications between images (i.e. between the *WA-MF2* shade and no-shade images) is, therefore, even larger. This was, however, difficult to evaluate as classifications of the shade and no-shade images were both very poor.

This chapter has the following conclusions:

1. The GP-OAD outperformed standard methods of classification when applied to laboratory-imagery and using imagery of mine faces acquired in the field.

2. The GP-OAD is less sensitive to the number of bands used than were competing methods, specifically, than an Euclidian distance based method (e.g. GP-SE or SVM-SE). Also, methods using an angular metric perform generally better than distance based methods using data presented in this chapter.

3. SVMs perform similarly as GPs when amounts of noise between training and

test set are similar and the light sources used to acquire either data-set are the same (i.e. using artificial illumination).

4. GPs outperform SVMs irrespective of the kernel and the conditions of the data. This is especially evident when the light sources used to acquire the training and test set are different (i.e. artificial and natural light).

5. Small amounts of change in mineralogy can cause some classes to be misclassified even though samples may be nearly identical in terms of their chemistry. Hyperspectral data can, however, be used to detect such differences in the case of the major iron absorption (i.e. the F_{900} feature).

6. Reasons for misclassification are generally related to illumination geometry, noise, and type of illumination (e.g. artificial or natural light) and, in some cases, small changes in mineral composition. The performance of classifiers is generally less affected by the number of spectral bands than by other factors. This means that there is a relatively small reduction in the specificity of data from imaging spectrometers compared to high-resolution non-imaging spectrometers.

7. These results show that angular-based methods, such as SAM or the OAD kernel, are better suited for the classification of mine faces using hyperspectral imagery compared to distance-based methods (e.g. the SE kernel). Furthermore, a method is required which can model the noise in the data and in the entire system (e.g. the classifier). GPs learn two parameters which are related to the amounts of noise in the data and the system. The parameter that describes the noise in the data is generally provided in the kernel. Thus, SVMs can learn this parameter and model the noise in the data as well. The other parameter, however, describes the noise in the Gaussian Process, i.e. the noise in the system which is directly related to the uncertainty of an estimate. The capability of GPs to model the noise in a system is a major advantage over SVMs, especially in cases where the amount of noise between the training and test data-sets are very different. Of the classifiers tested in this research, GPs

using an angular based kernel (i.e. the OAD kernel) are therefore most suited for the classification of hyperspectral imagery of mine faces.

Chapter 7

Fusion of hyperspectral images

7.1 Introduction and problem statement

The ability to integrate multiple sources of information is important because at any particular time in a dynamically changing environment (e.g. an operational mine) some important parts of the mine face may be obscured by moving objects. Objects may include pockets of airborne dust and moving equipment such as digger and haul trucks. Similarly, due to changes in the sun angle and elevation interacting with the complex geometry of a mine face, some areas of the mine face may be shaded at times during the day but fully illuminated at others and vice versa. Classification of shaded spectra is more challenging because they have been acquired with indirect light which has a different wavelength-intensity distribution than the library spectra which were acquired using direct light. Classification becomes more challenging as spectra acquired from areas of the image which are affected by both multiplicative and additive effects (i.e. due to changing proportions of skylight in shaded areas) cannot be reliably matched with library spectra, acquired under direct light. Classification based on the spectral angle does in fact change in response to changes in the geometry of illumination and, particularly, in areas of the mine face affected by overt shade (Murphy et al., 2012). These authors created a library of spectra which had been measured under conditions of shade, with the direct solar beam occluded for the target

but not for the calibration measurement. The library was incorporated into a meta-library which included spectra acquired in direct light. When this library was used to classify a hyperspectral image, containing both shaded and unshaded areas, the library spectra acquired in shade were not preferentially selected over spectra acquired in direct light to classify the shaded areas of the image. The probable reasons for this were that the shaded library spectra were measured in a different location with a different proportion of skylight and light reflected from surrounding structures. Thus, the use of such libraries, acquired in shade to classify shaded areas of imagery is problematic as the wavelength-intensity distribution of incident light is specific to a particular location. One obvious and rather simple solution to this problem is then to acquire the same target at different times during the day with different illumination conditions, i.e. so that all areas of the mine face are illuminated in at least in one of the repeated image acquisitions.

Deterministic methods for classification of hyperspectral data (e.g. SAM), do not provide mechanisms to integrate hyperspectral information from other sources (e.g. several scans of a mine face). Deterministic methods as well as probabilistic methods can provide classifications (e.g. thematic maps of geology) at each point in time for all images acquired during a day, however, classification results will be different to one another, particularly in areas affected by shade. An important question to answer is, therefore, which, if any, of these classifications is the correct one? To answer this, it may be necessary to manually classify individual pixel spectra from the affected areas or even over the entire scene. Because this is not feasible in autonomous mining or in any other automated processes, it is better to avoid making this decision altogether. Instead, it may be beneficial to reconcile all the available information from multitemporal images into a single classification.

Information in this context is the classified imagery, each derived independently from the imagery acquired of the same target at a different time of the day. Given the nature of this information, a fundamental question arises - what is an appropriate framework under which this information can be integrated or 'fused'? Fusing several classified images derived from deterministic methods of classification would be prob-

lematic because there is no easy or an intuitive way to fuse deterministic classifier outputs. Probabilistic methods, however, which operate in a fully Bayesian framework, provide the necessary underlying structure which serves as a basis to fuse more information into an existing classification output. The objective for the probabilistic classifier changes slightly from obtaining the 'correct' classification for a single image to obtaining "the most probable estimate or class" for a pixel given multiple classifications of the same target, acquired at different times of the day.

Advantages of fusing multiple probabilistic outputs from Gaussian Processes over individual classifications are: (i) large amounts of data from several classifications from hyperspectral images acquired at different times of the day are reduced to a single output, (ii) a single output removes the need to decide - potentially without any actual knowledge about the scene - which classification is *probably* the correct one, (iii) no manual interaction is required to make such a decision and (iv) as new information becomes available, the estimate of the class becomes more certain due to the incorporation of new information into the existing estimate.

This chapter introduces a framework for fusing multiple Gaussian Process outputs as a way of improving classification of a mine face. The presented framework is probabilistic, i.e. operating in a Bayesian framework, and, as such, it is able to fuse classifications of images acquired at multiple times during the day and/or from different positions. The proposed new framework fuses Gaussian Process outputs, obtained from the GP-OAD method, of several hyperspectral images. While it is possible to fuse also GP-SE and/or GP-NNET outputs, for reasons of clarity the fusion framework is demonstrated by fusing only several GP-OAD outputs.

In the following sections, the mathematical formulism is provided for fusing two Gaussian Process outputs. The framework is (i) generalised to fuse any number (n) of GP outputs sequentially and the functionality of the framework is extended to (ii) fuse any n GP outputs at the same time. The ability to fuse data sequentially is important if an existing estimate is to be updated as new data becomes available. Fusing several observations at once is necessary if many observations are available at the same time, thus reducing computational requirements. A practical implementation

of the new framework is demonstrated using several data-sets. Comparison of qualitative and quantitative results between the proposed fusion framework and individual classifications are provided and discussed.

7.1.1 Related work

Gaussian Processes provide means for fusing multi-modal data because they operate within a Bayesian framework. There are three established means of data fusion using GPs (e.g. Rasmussen, 2004; Vasudevan, 2012).

The simplest method for fusing data using GPs is to use the additional data (e.g. a second hyperspectral image) within the existing GP model. The assumption is that the existing GP model comprising the kernel and hyper-parameters, provides an adequate representation of the new data. In this case, all previous hyper-parameters would be identical and only the training data (i.e. X using the notation from section 5.2.1) and their corresponding function values ($\mathbf{y}$) for each data-set need to be considered for the data fusion process (using Equations 5.5, 5.9, 5.10, in Section 5.2.1).

There are other approaches such as heteroscedastic Gaussian Process (HGP; Rasmussen, 2004; Le et al., 2005) and multi-output Gaussian Processes (termed multi-task GPs or Dependent GPs) which extend Gaussian Processes to handle multiple correlated outputs simultaneously (Boyle and Frean, 2005). All the aforementioned approaches for fusing GPs require the construction of a new covariance matrix from several data-sets. This may not be possible as data are not necessarily available all at once, thus, sequential fusion may be necessary as new data becomes available.

There are also other methods proposed in the literature for fusing hyperspectral images which are not based on GPs. Alajlan et al. (2012) proposed an approach which uses SVMs and a fuzzy C-mean clustering algorithm (FCM; Dunn, 1973). Both methods (i.e. SVMs and FCM) were applied to the same image and their outputs were then fused using voting-rules and Fisher-Markov Selector (Cheng, 2011). Other work has used a combination of statistical methods and neural networks to fuse multisource data as well as hyperspectral data (e.g. Benediktsson et al., 1990; Benediktsson and

Kanellopoulos, 1999). In cases where the different classifiers yielded different class labels for the same image pixel, a decision process based on consensus theory was used (Benediktsson and Sveinsson, 2000). Waske and Benediktsson (2007) proposed a method of decision fusion for SVMs using multisource data from SAR and optical imagery. Individual SVM classifications for each data source were then fused using SVM-based decision functions as well as majority voting and the absolute maximum strategy. Other approaches for decision fusion of hyperspectral and multisource data are discussed in Licciardi et al. (2009) and the references therein as well as Fauvel et al. (2006) and Plaza et al. (2009) and references therein.

The proposed method for fusion hyperspectral data described in this chapter differs to those methods described in the literature in several ways: (i) Data can be fused simultaneously or sequentially which does not change the outcome of the resulting fused classified map, and (ii) no majority voting or other decision strategies are required because decisions are made automatically based on the most probable estimate. This is important because probabilities allow the class of a pixel to be changed on the basis of the two inputs (i.e. images), specifically, the 'old' thematic map, i.e. the one which already exists, and the 'new' thematic map, i.e. the map just obtained. In this case, the 'old' map may already contain information from several previous classifications. If a majority voting is employed, all previous classifications need to be stored in order to make a new 'vote' every time a new observation becomes available. (iii) The proposed method allows to fuse two classifications of the same image which were obtained using different methods or kernels, for example, it would be possible to fuse three classifications of a single image using the GP-SE, GP-NNET and the GP-OAD. Whether such an approach is useful would require further investigation, however, theoretically, fusion of several kernel outputs may be beneficial for some applications. (iv) Neither the classification nor the fusion process requires manual interaction (e.g. extraction or selection of training pixels) with the target images themselves because an independent training set is provided. Many methods in the literature require some interaction with the target image in order to select training data. This is not suitable nor desirable for applications in autonomous mining.

The presented approach is simple to use and fits well within the probabilistic context of GPs. It is believed that a simple fusion approach, such as the one proposed here, can lead to an improved classification estimate of a mine face. By doing data fusion external to the GP framework, it is possible to fuse any number of new arriving images to previous estimates, thus providing greater flexibility in the way data can be fused.

7.2 Mathematical formulation of the fusion framework

The framework to fuse several GP outputs is based on the 'recursive Bayes Update' algorithm which is similar to the Kalman filter update (Grewal and Andrews, 1993). The main reference used to describe the equations used in this section is Bertsekas and Tsitsiklis (2002). Given a joint probability density function $P(\mathbf{x}, \mathbf{z})$ on two random variables $\mathbf{x}$ and $\mathbf{z}$, using the chain rule, one can obtain the conditional densities

$$\begin{aligned} P(\mathbf{x}, \mathbf{z}) &= P(\mathbf{x}|\mathbf{z})P(\mathbf{z}) \\ &= P(\mathbf{z}|\mathbf{x})P(\mathbf{x}) \ . \end{aligned} \tag{7.1}$$

Rearranging of 7.1 gives Bayes Theorem, which is the principle of integrating information

$$P(\mathbf{x}|\mathbf{z}) = \frac{P(\mathbf{z}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{z})} \ . \tag{7.2}$$

Assuming that $\mathbf{x}$ is a state (i.e. class of rock) from $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z}$ are observations (i.e. predicted values for a class from the GP outputs) of that state then $\mathbf{z}^n$ denotes a set of observations

$$\mathbf{z}^n = \{\mathbf{z}_1, \ldots, \mathbf{z}_n \in \ \mathcal{Z}_n\} \tag{7.3}$$

It is desired to use this information to construct a posterior distribution $P(\mathbf{x}|\mathbf{z}^n)$ describing the relative likelihoods of the various values of the state of interest $\mathbf{x}$ given

the information obtained. Bayes theorem can be directly employed to compute this distribution function using

$$P(\mathbf{x}|\mathbf{z}^n) = \frac{P(\mathbf{z}^n|\mathbf{x})P(\mathbf{x})}{P(\mathbf{z}^n)}$$

$$= \frac{P(\mathbf{z}_1, \ldots, \mathbf{z}_n|\mathbf{x})P(\mathbf{x})}{P(\mathbf{z}_1, \ldots, \mathbf{z}_n)} \tag{7.4}$$

In practice this may not be possible, since it requires that the joint distribution $P(\mathbf{z}_1, \ldots, \mathbf{z}_n|\mathbf{x})$ is completely known, i.e. the joint distribution of all possible combinations of observations conditioned on the underlying state. However, if it is assumed that, given the true state $\mathbf{x}$, the information obtained from the i^{th} information source is independent of the information obtained from other sources, then conditional independence can be expressed as

$$P(\mathbf{z}_i|\mathbf{x}, \mathbf{z}_1, \ldots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_n) = P(\mathbf{z}_i|\mathbf{x}) \ , \tag{7.5}$$

which can be written as a product

$$P(\mathbf{z}_1, \ldots, \mathbf{z}_n|\mathbf{x}) = P(\mathbf{z}_1|\mathbf{x}) \ldots P(\mathbf{z}_n|\mathbf{x}) = \prod_{i=1}^{n} P(\mathbf{z}_i|\mathbf{x}). \tag{7.6}$$

Substituting this back into 7.4 gives a simple and direct mechanism for computing the relative likelihood in different values of a state from any number of observations or other pieces of information and is known as the *independent likelihood pool*:

$$P(\mathbf{x}|\mathbf{z}^n) = [P(\mathbf{z}^n)]^{-1} P(\mathbf{x}) \prod_{i=1}^{n} P(\mathbf{z}_i|\mathbf{x}) \ . \tag{7.7}$$

The updated likelihoods in the state, i.e. the posterior distribution on $\mathbf{x}$, is proportional to the product of prior likelihood and individual likelihoods from each information source. The marginal distribution $P(\mathbf{z}^n)$ acts as a normalising constant.

The integration of information using Equation 7.7 would, in principle, require that upon the arrival of new information, $P(\mathbf{z}_k|\mathbf{x})$, the total likelihood would need to be recomputed based on all information gathered.

Equation 7.7 enables the incremental or recursive addition of new information in determining a revised posterior distribution on the state $\mathbf{x}$ with $\mathbf{z}^k = \{\mathbf{z}_k, \mathbf{z}^{k-1}\}$

$$P(\mathbf{x}, \mathbf{z}^k) = P(\mathbf{x}|\mathbf{z}^k)P(\mathbf{z}^k) \tag{7.8}$$

$$= P(\mathbf{z}_k, \mathbf{z}^{k-1}|\mathbf{x})P(\mathbf{x})$$

$$= P(\mathbf{z}_k, |\mathbf{x})P(\mathbf{z}^{k-1}|\mathbf{x})P(\mathbf{x}) \ , \tag{7.9}$$

where it is assumed conditional independence of the observation sequence. Equating both sides of this expansion gives

$$P(\mathbf{x}|\mathbf{z}^k)P(\mathbf{z}^k) = P(\mathbf{z}_k|\mathbf{x})P(\mathbf{z}^{k-1}|\mathbf{x})P(\mathbf{x}) \tag{7.10}$$

$$= P(\mathbf{z}_k|\mathbf{x})P(\mathbf{x}|\mathbf{z}^{k-1})P(\mathbf{z}^{k-1}). \tag{7.11}$$

Rearranging with $P(\mathbf{z}^k)/P(\mathbf{z}^{k-1}) = P(\mathbf{z}_k|\mathbf{z}^{k-1})$ gives

$$P(\mathbf{x}|\mathbf{z}^k) = \frac{P(\mathbf{z}_k|\mathbf{x})P(\mathbf{x}|\mathbf{z}^{k-1})}{P(\mathbf{z}_k|\mathbf{z}^{k-1})} \ . \tag{7.12}$$

Using Equation 7.12, only the posterior likelihood $P(\mathbf{x}|\mathbf{z}^{k-1})$ which contains a complete summary of all past information needs to be computed. When the next piece of information $P(\mathbf{z}_k|\mathbf{x})$ arrives, the previous posterior takes on the role of the current prior and the product of the two becomes the new posterior after normalisation. Equation 7.12 provides a significant improvement in computational efficiency and

memory requirements over Equation 7.7.

7.2.1 Fusing two Gaussian Process outputs

A GP output, in the case of GP regression, is a mean and a variance (Section 5.2.1). The GP-OAD, for example, outputs mean values and uncertainties (e.g. variances), thus they provide the basis for representing a Normal distribution such as $\mathbf{y} \sim \mathcal{N}(\mu, \sigma)$. Given a classified hyperspectral image, the probability for a pixel of belonging to a certain class can be calculated by evaluating the integral of the cumulative distribution function between the interval of minus infinity to zero, i.e.

$$P = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left(\frac{-(\mathbf{z}-\mu)^2}{2\sigma^2}\right) d\mathbf{z} \tag{7.13}$$

The result, P, is the probability that a single observation from a normal distribution with parameters μ and σ (obtained from the GP) will fall within the interval $(-\infty, 0]$.

Assuming two GP outputs obtained from application of the GP-OAD to two hyperspectral images of the same mine face, the aim is to increase the predictive accuracy by fusing those two outputs. Specifically, the posterior distribution of a scalar, $\mathbf{x}$, is sought under the assumption that the likelihood function for the observations given the true state is Gaussian with known variance σ^2,

$$P(\mathbf{z}_k|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mathbf{z_k}-\mathbf{x})^2}{2\sigma^2}\right) \tag{7.14}$$

Assuming that the posterior distribution of $\mathbf{x}$ after the first $k-1$ observations is also Gaussian with mean $\mathbf{x}_{k-1}$ and variance σ^2_{k-1} then

$$P(\mathbf{x}|\mathbf{z}^{k-1}) = \frac{1}{\sqrt{2\pi\sigma^2_{k-1}}} \exp\left(-\frac{(\mathbf{x}_{k-1}-\mathbf{x})^2}{2\sigma^2_{k-1}}\right) \tag{7.15}$$

then the posterior distribution in $\mathbf{x}$ after the first k observations (i.e. the first two in this example) is given by

$$P(\mathbf{x}|\mathbf{z}^k) = K\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\mathbf{z_k}-\mathbf{x})^2}{2\sigma^2}\right)\cdot\frac{1}{\sqrt{2\pi\sigma_{k-1}^2}}\exp\left(-\frac{(\mathbf{x}_{k-1}-\mathbf{x})^2}{2\sigma_{k-1}^2}\right) \tag{7.16}$$

$$= \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left(-\frac{(\mathbf{x}_k-\mathbf{x})^2}{2\sigma_k^2}\right)\ , \tag{7.17}$$

where K is a constant independent of $\mathbf{x}$ which normalises the posterior, and $\mathbf{x}_k$ and σ_k^2 are the new weighted means and variances given by

$$\mathbf{x}_k = \frac{\sigma_{k-1}^2}{\sigma_{k-1}^2+\sigma^2}\mathbf{z}_k + \frac{\sigma^2}{\sigma_{k-1}^2+\sigma^2}\mathbf{x}_{k-1} \tag{7.18}$$

and

$$\sigma_k^2 = \frac{\sigma^2\sigma_{k-1}^2}{\sigma^2+\sigma_{k-1}^2}\ . \tag{7.19}$$

$\mathbf{x}_{k-1}$ is the weighted mean of all previous observations $\mathbf{z}_1,\ldots,\mathbf{z}_{k-1}$ and σ_k is the harmonic mean of the previous $k-1$ variances, for the specific case of fusing two Gaussians Equations 7.18 and 7.19 can be simplified to

$$\mathbf{x}_\star = \frac{\sigma_{\mathbf{z}_2}^2}{\sigma_{\mathbf{z}_1}^2+\sigma_{\mathbf{z}_2}^2}\mathbf{z}_1 + \frac{\sigma_{\mathbf{z}_1}^2}{\sigma_{\mathbf{z}_1}^2+\sigma_{\mathbf{z}_2}^2}\mathbf{z}_2 \tag{7.20}$$

and

$$\sigma_\star^2 = \frac{\sigma_{\mathbf{z}_1}^2\sigma_{\mathbf{z}_2}^2}{\sigma_{\mathbf{z}_1}^2+\sigma_{\mathbf{z}_2}^2}\ , \tag{7.21}$$

where subscripts $\mathbf{z}_1$ and $\mathbf{z}_2$ indicate to which observation $\mathbf{z}$ the variance belongs. The star subscript ($\star$) indicates the fused mean and variance, respectively.

7.2.2 Fusion of n Gaussian Process outputs

The formalism introduced in the previous section can now be used to fuse n observations sequentially by treating the previous posterior as the new prior distribution. This is useful if data are acquired sequentially. If, however, more observations are available at the same time, fusing all n observations at the same time can be an advantage in terms of simplicity and also in terms of computational requirements. Equations in this section are based on the Maximum Likelihood classifier and are derived in Bishop (2006).

To fuse n observations at once, Equations 7.20 and 7.21 can be rewritten as

$$\begin{aligned} \sigma_\star^2 &= \frac{\sigma_{\mathbf{z}_1}^2 \sigma_{\mathbf{z}_2}^2}{\sigma_{\mathbf{z}_1}^2 + \sigma_{\mathbf{z}_2}^2} \\ &= \left(\frac{1}{\sigma_{\mathbf{z}_1}^2} + \frac{1}{\sigma_{\mathbf{z}_2}^2} \right)^{-1} , \end{aligned} \tag{7.22}$$

thus, the new variance $\sigma_\star^2$ is equal to the parallel combination of the original variances. Equation 7.22 leads to a more general form for n observations

$$\sigma_{\mathbf{z}_1,\ldots,\mathbf{z}_n}^2 = \left(\frac{1}{\sigma_{\mathbf{z}_1}^2} + \cdots + \frac{1}{\sigma_{\mathbf{z}_n}^2} \right)^{-1} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_{\mathbf{z}_i}^2}} . \tag{7.23}$$

For the new mean $\mu_\star$ Equation 7.20 can be rearranged to

$$\begin{aligned} \mu_{\mathbf{z}_1,\mathbf{z}_2}(\sigma_{\mathbf{z}_1}^2 + \sigma_{\mathbf{z}_2}^2) &= \sigma_{\mathbf{z}_1}^2 \mathbf{z}_2 + \sigma_{\mathbf{z}_2}^2 \mathbf{z}_1 \\ \mu_{\mathbf{z}_1,\mathbf{z}_2} \left(\frac{1}{\sigma_{\mathbf{z}_1}^2} + \frac{1}{\sigma_{\mathbf{z}_2}^2} \right) &= \frac{\mathbf{z}_1}{\sigma_{\mathbf{z}_1}^2} + \frac{\mathbf{z}_2}{\sigma_{\mathbf{z}_2}^2} \end{aligned} \tag{7.24}$$

Substituting Equation 7.23 into Equation 7.24, the final rearrangement of Equation 7.20 can be obtained from

$$\frac{\mu_{\mathbf{z}_1,\mathbf{z}_2}}{\sigma^2_{\mathbf{z}_1,\mathbf{z}_2}} = \frac{\mathbf{z}_1}{\sigma^2_{\mathbf{z}_1}} + \frac{\mathbf{z}_2}{\sigma^2_{\mathbf{z}_2}} \ . \tag{7.25}$$

If a further observation is fused with the two previous observations, Equation 7.25 can be expanded to

$$\frac{\mu_{\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3}}{\sigma^2_{\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3}} = \frac{\mathbf{z}_1}{\sigma^2_{\mathbf{z}_1}} + \frac{\mathbf{z}_2}{\sigma^2_{\mathbf{z}_2}} + \frac{\mathbf{z}_3}{\sigma^2_{\mathbf{z}_3}} \ . \tag{7.26}$$

This can be expressed as a general statement for n observations in the form

$$\frac{\mu_{\mathbf{z}_1,\ldots,\mathbf{z}_n}}{\sigma^2_{\mathbf{z}_1,\ldots,\mathbf{z}_n}} = \sum_{i=1}^{n} \frac{\mathbf{z}_i}{\sigma^2_{\mathbf{z}_i}} \tag{7.27}$$

again substituting the general form for the variance of n observations the final expression for the mean for fusing n observations is obtained

$$\mu_{\mathbf{z}_1,\ldots,\mathbf{z}_n} = \frac{\sum_{i=1}^{n} \frac{\mathbf{z}_i}{\sigma^2_{\mathbf{z}_i}}}{\sum_{i=1}^{n} \frac{1}{\sigma^2_{\mathbf{z}_i}}} \tag{7.28}$$

7.2.3 Data Fusion - Practical Implementation

Each time the GP-OAD (as well as GP-NNET and GP-SE) are used to classify a hyperspectral image, they yield two three-dimensional arrays (as described in Section 6.1). Each array has the dimensionality of the spatial size of hyperspectral image for the x and y dimensions and the dimensionality of z is determined by the number of classes in the training set (Section 6.1). One of these arrays contains mean values and the other variances (uncertainties). From these, probabilities can be calculated using Equation 7.13. However, instead of fusing probabilities, one can calculate a new mean and variance from all inputs (Equations 7.28 and 7.25) and simply calculate

probabilities from the fused (or updated) mean and variances. By determining the maximum probability across the z dimension for each pixel. Class labels can then be assigned to obtain a final thematic map of the geology (as described in Section 6.1). The 'final classification' can, however, be updated if the fused mean and variances are stored and fused with newly arriving data (Figure 7.1).

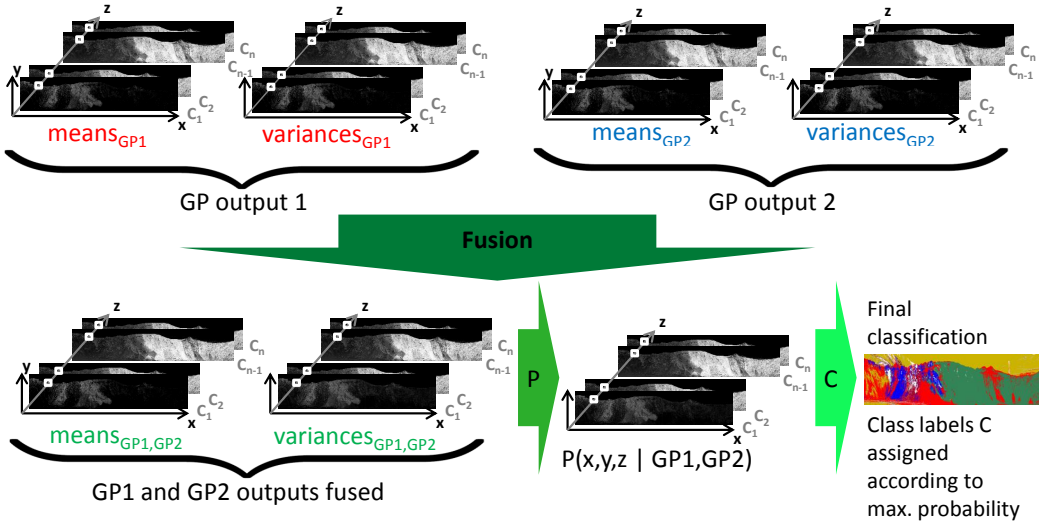


Figure 7.1 – Concept for fusing two GP outputs. Two GP outputs, each comprised of two arrays, one representing means and the other the variances. The two mean arrays and variance arrays - from classifications of GP1 and GP2 - are fused to single mean and variance array, according to the Equations given in the previous section. The probability is then calculated from the fused arrays. Using the maximum probability across the z dimension, class labels can be assigned to obtain a thematic map of the geology, i.e. the final classification.

7.3 Materials and Methods

7.3.1 Tom Price

Geology of Tom Price

The field site was located in an open pit iron ore mine near Tom Price, Western Australia. The geology of the mine face was divided into zones during field trials by

an experienced geologist (Figure 7.2). Zone 1 was mapped to be mainly comprised of Banded Iron Formation (BIF); Zones 2 and 5 were mainly shale (SHL). Zones 3 and 4 were categorised as iron ore, mainly hematite but with some goethite. The distinction between Zones 3 and 4 is that Zone 3 was observed to be low grade iron ore (LG) and Zone 4 high grade iron ore (HG). There was, however, no fixed or distinct boundary between the LG and HG zones, instead a gradation between the two rock types was recognised, i.e. due to decreasing amounts of goethite. The direction of the gradation was from left to right, i.e. from LG to HG. Both rock types LG and HG are valuable iron ores, however, with different hematite to goethite ratios.

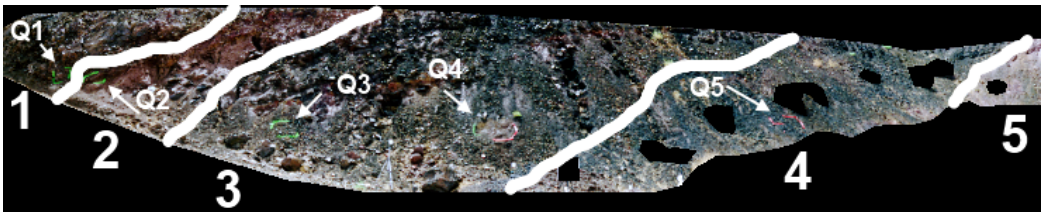


Figure 7.2 – Zones of different geology of a Tom Price mine face. See text for reference to the numbers. The five quadrats (Q1-Q5) mark areas on the mine face from which samples were collected for determination of 'ground truth'.

Acquisition of ground-truth for the Tom Price mine face

Sampling of the mine face was strictly controlled due to safety considerations. Quadrats (70 cm × 70 cm) were placed over areas identified as BIF, SHL (2 quadrats), LG and HG which were the four major types of rock found on this particular mine face (Figure 7.2). The rocks within each quadrat were of relatively homogenous composition. Five representative rock samples were taken from each quadrat for XRD analysis to serve as a surrogate for ground-truth of each zone. Due to the irregular spatial sampling across the mine face, it may not be appropriate to assume that the entire geological unit or zone (numbers in Figure 7.2 indicate the zones) is of the same type of rock as the one in the quadrat, however, this is the best that could be done considering the constraints due to safety consideration.

Training set

Training data were created using bulk rocks collected from the mine face from each of the five quadrats. Rocks from each of the quadrats were collected and used to create a spectral library (Table 7.1). Spectra were acquired using a high reflectance probe with high intensity illumination and an ASD field spectrometer. The reflectance probe was placed in direct contact with the sample to obtain spectra with very high signal-to-noise ratios. Spectra were acquired from areas of the rocks which showed relatively small amounts of micro-topography. The areas on the rock from which spectra were acquired were also those previously exposed on the mine face, i.e. the bottom side was marked during sample collection.

Table 7.1 – Tom Price training data and principle rock types. 'Zone' indicates where on the mine face each rock type was expected and 'Quadrat' indicates from where samples were collected for each of the rock types (Figure 7.2). The number of spectra was reduced to 15 per class so that the same number of spectra was available in the training set.

Type of Rock	Code	Zone	Quadrat
Banded Iron Formation	BIF	1	Q1
Shale	SHL	2+5	Q2 + Q3
Low-Grade (Hematite + Goethite)	LG	3	Q4
High-Grade (Hematite)	HG	4	Q5

Hyperspectral imagery

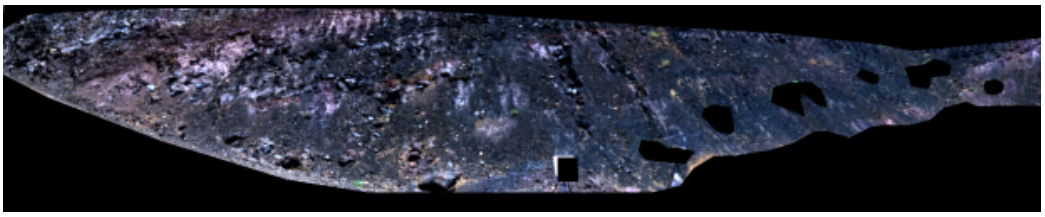
Hyperspectral imagery was acquired from a vertical mine face in the Tom Price mine from a distance of approximately 20 m, using the same VNIR and SWIR imaging sensors and acquisition methodology as used to acquired data from the West Angelas mine (Section 6.4.1). Calibration to reflectance was also done using the same method as used to calibrate the West Angelas data-sets. Hyperspectral images were acquired with 'temporal' differences from a mine face in Tom Price. Temporal difference between images means that a scene or an object such as a mine face is acquired at several times during a short period of time (e.g. a day). Partial or full cloud cover of the sky, the orientation of the sun and the proportion of light reflected from the

surroundings of the mine face relative to the skylight causes variations in illumination. As a consequence the appearance of the mine face changes significantly between images (Table 7.2, Figures 7.3 and 7.4). Also, the orientation of the sun may cause some areas on the mine face to be shaded at one point in time and be fully illuminated at a later point or vice versa (Figure 7.3). Thus, being able to scan a mine face at several times during the day provides, theoretically, a greater chance of observing a certain area under optimal conditions (e.g. fully illuminated). Figure 7.5 shows example spectra of one of the hyperspectral images to illustrate the quality of the data (e.g. high noise) and the spectral difference between the classes. All spectra are very noise because the mine face was partly shaded and rock types such as HG and LG are inherently dark materials.

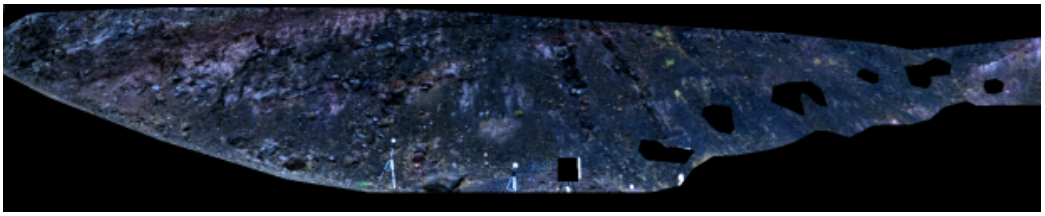
Table 7.2 – Summary of information about the Tom Price imagery. Information contains local time of acquisition, solar azimuth and elevation angles in degree at the specific mine face location. All images had 283 bands after preprocessing. The penultimate column references the corresponding RGB image in Figure 7.4.

image	time	sol azimuth	elevation	Fig. 7.4	condition
3	11.35am	58.06	84.55	a)	clear sky, little shade on mine face
25	1.02pm	277.47	74.06	b)	thick cloud cover, diffuse illumination, little shade
33	1.35pm	272.43	66.48	c)	clear sky, shade in different areas (left)
40	2.27pm	267.4	54.51	d)	clear sky, variable intensity towards end of scan, possible high altitude clouds, increase in shade (left)

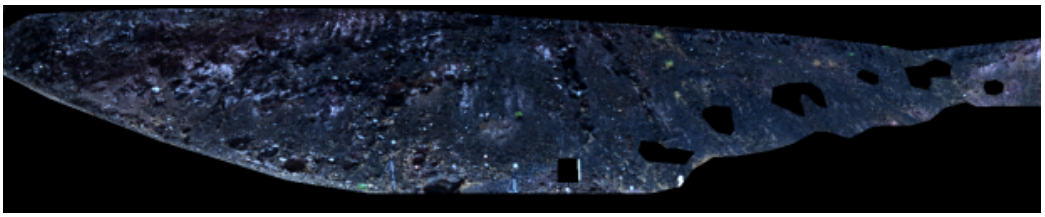
Figure 7.3 – Illustration of sun-angle relative to the position. Labels 'sunrise' and 'sunset' indicate the apparent sunrise or sunset at that location. Yellow lines indicate the direction from which the sun is shining (Table 7.2). Image source: Google Earth.



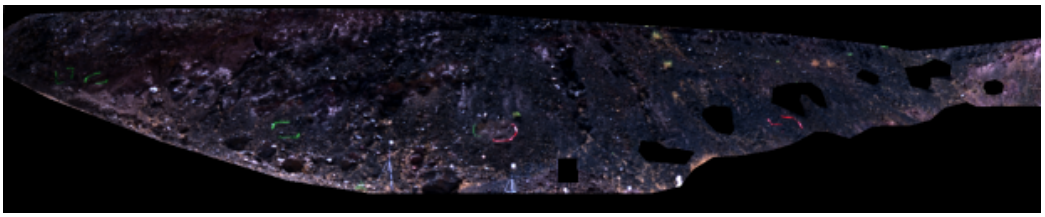
(a) image 3 at 11.35am local time



(b) image 25 at 1.02pm local time



(c) image 33 at 1.35pm local time



(d) image 40 at 2.27pm local time

Figure 7.4 – Tom Price RGB composite mine face images. Images are not colour enhanced and show a colour composite of bands 54, 35 and 14 from the reflectance calibrated image. These bands represent the RGB wavelengths at 639.7 nm, 551.9 nm and 457.8 nm. More blue light is present in the first three images (particular, b and c) which were acquired close to the local solar noon compared to the last image acquired which has a reddish hue in the afternoon. These colour variations are likely due to wavelength-dependent Rayleigh scattering ($\propto 1/\lambda^4$) and absorptions of the blue wavelengths when passing through the atmosphere in the afternoon and evening. Example image spectra are shown in Figure 7.5.

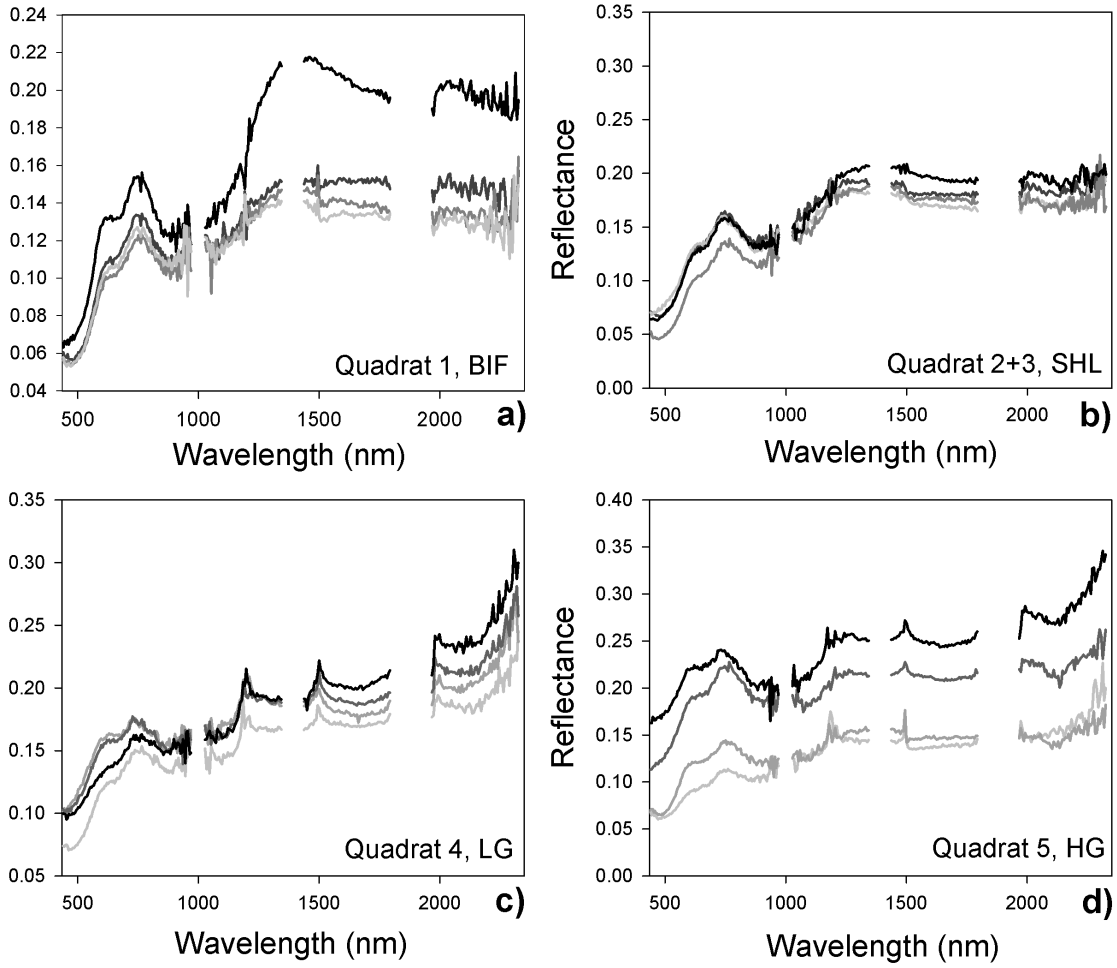


Figure 7.5 – Four randomly selected Tom Price example spectra from image 40 for each of the quadrants. Spectra from Quadrant 2 and 3 are combined in b). Note the different scaling of the Ordinates for each of the panels a)-d) which was done for clarity.

7.3.2 West Angelas

The West Angelas, Mine face 2 (WA-MF2), data-set is used to test the proposed fusion framework, in this case only two images with temporal difference were available. This imagery (WA-MF2, Figure 6.15) was acquired in the morning and in the afternoon during the same day (see Section 6.4.1 for more information about the imagery). In one of the images, some areas of the mine face were shaded. The aim was to investigate if the classification of the shaded image can be improved by fusing it with

the fully illuminated image.

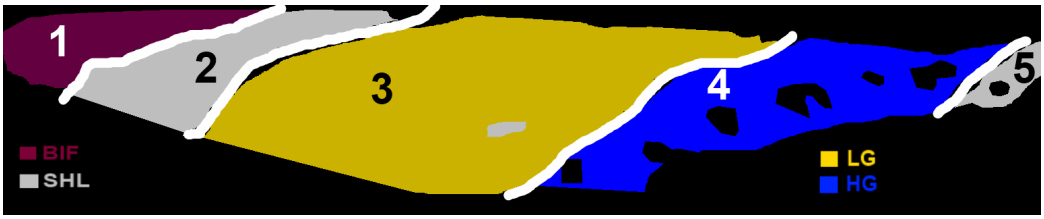
7.4 Results and discussion

7.4.1 Tom Price

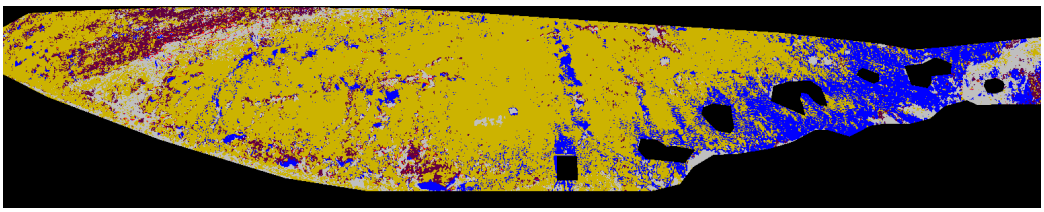
Classification of the four individual images of the mine face at Tom Price using the GP-OAD are shown in Figure 7.6. Within each zone there was variability in the classification of pixels which were inconsistent with the broad classification made in the field (Figure 7.6). This was confirmed by the quantitative results (Figure 7.7) which show variable and inconsistent F-scores across the zones for each of the four classifications. For example, the classification of image 25 is relatively good for the LG and HG classes but yields a lesser F-score for SHL and a poor performance for BIF (see Figure 7.7). The classification of image 40 (img40) yields the highest F-score among all individual classifications and the second highest score for LG, however, the classification of SHL is the poorest and only the third highest F-score for HG. Similar inconsistent patterns are present for the other two individual image classifications. This inconsistent performance of classification makes it difficult to choose an overall best classification. Fusing the individual GP outputs yields a map of geology which is qualitatively in closer agreement with the ground-truth. Also, quantitative results show a significant improvement in the performance of classification with respect to F-scores. The F-scores for all rock types, with the exception of BIF, and the average F-score (across all types of rock) is higher than any of the individual classifications.

Fusing GP outputs improves the classification performance as defined by increased F-scores for most classes compared to individual image classifications. For the BIF class, however, there is a slight decrease in classification performance although the average F-score value across all classes has improved. In theory, this should not be possible because every observation will either improve the classification or will remain the same if the new input is very uncertain. So why does the classification performance decline after fusing four GP outputs together for the class BIF?

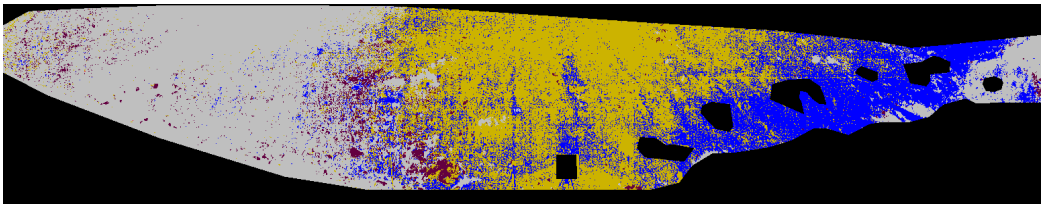
A decline in performance as mentioned above is likely to occur due to 'over-confident' classifications of pixels from one or more of the individual classifications. An over-confident classification in this context means that a small uncertainty (variance) is obtained for an incorrect estimate, i.e. a misclassified pixel. If the GP-OAD estimates the correct class for a given pixel, the variance should be small. If the estimate is wrong, the variance should be large compared to the variance of a correctly classified image pixel. In some cases, the variance is, however, small although the image pixel is misclassified. This is then referred to as 'over-confident'. Because a small uncertainty is equivalent to a large weight when calculating the new, weighted mean during the fusion procedure (Equations 7.23 and 7.28), the new mean and variance will be biased in the wrong direction (i.e. towards the wrong class) and become more certain about the wrong estimate. This diminishes the probability for a given pixel and hence the quality of the classification decreases. In this context,'over-confident' classifications can be obtained if there are large amounts of noise in the data and which cannot be appropriately modeled by the GP. Learning the noise hyper-parameter incorrectly is mostly a consequence of large differences in the amount of noise between the training and test set, however, in some cases the learning process converges to a local minimum. For the specific example of the data used here, the reason for over-confident classifications is most likely due to very different amounts of noise in the training and test data. The data presented used in this chapter was a training set which was acquired using a high-resolution field spectrometer with an artificial light source, thus yielding relatively noise-free spectra. The test set, i.e. the mine face imagery, was under exposed and does not utilise the full dynamic range of the sensor, resulting in data with large amounts of noise. Together with the inherent sensor noise (i.e. dark current), the resulting signal was very flat, nearly featureless and dominated by noise.



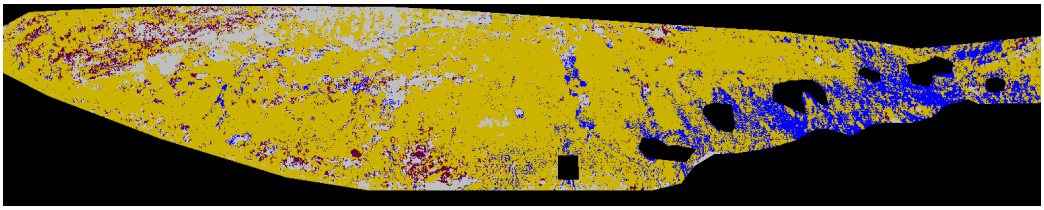
(a) Approximate ground truth according to XRD and XRF analysis of the 4 quadrats



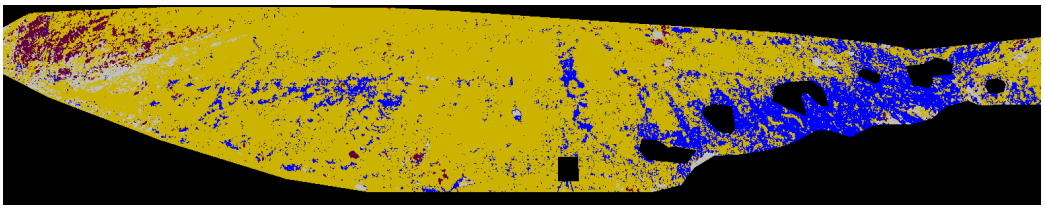
(b) classification of image 3 at 11.35am local time



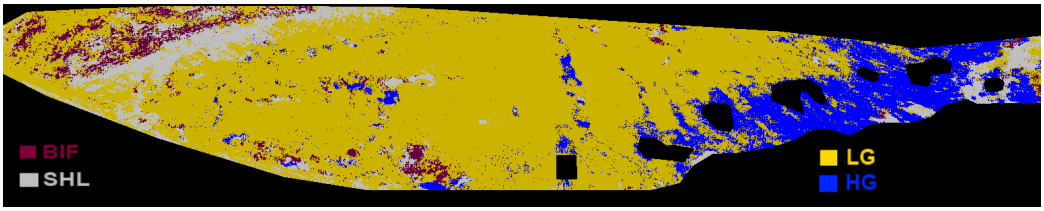
(c) classification of image 25 at 1.02pm local time



(d) classification of image 33 at 1.35pm local time



(e) classification of image 40 at 2.27pm local time



(f) classification after fusion of the four individual classifications

Figure 7.6 – Tom Price approximate ground truth (a) and individual mine face classifications (b-e) and the classification after fusion (f) of the four individual GP outputs. Numbers in (a) mark distinct zones referenced in the text.

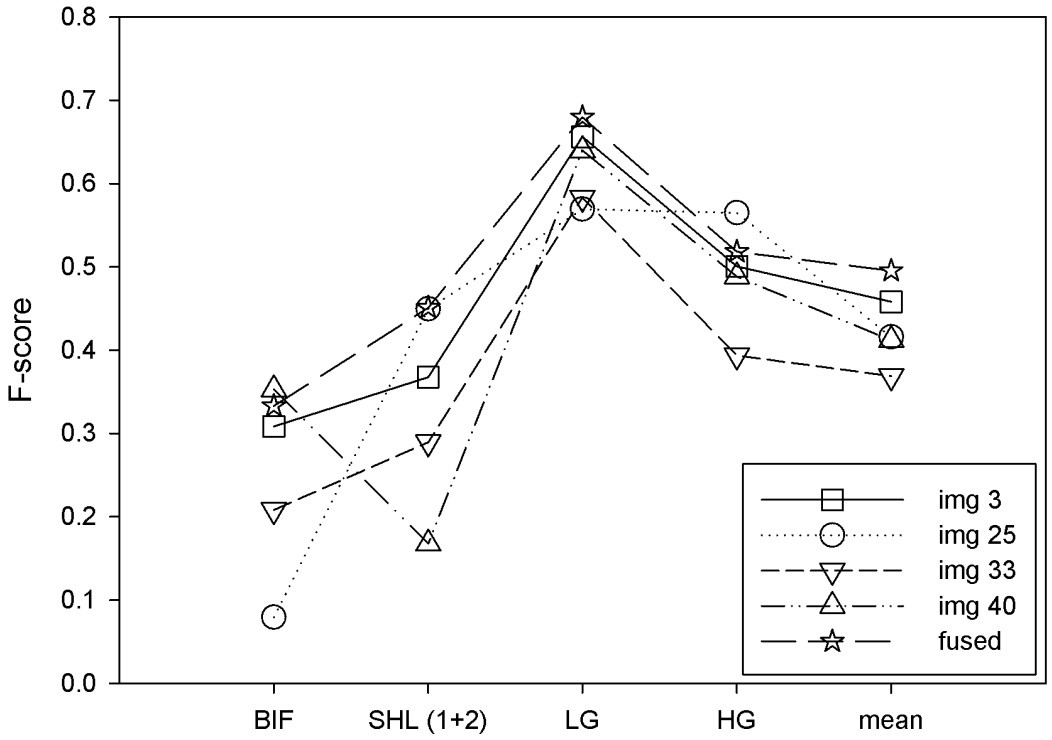


Figure 7.7 – Comparison of individual GP-OAD classifications vs. the fused GP-OAD classification based on F-scores. Lines between symbols are visual aids only. 'mean' denotes the mean or average across the four rock types (BIF, SHL, LG and HG). Note that these quantitative results are only approximate, since the ground truth could not be mapped with such high-resolution and accuracy as is illustrated in the ground-truth map (Figure 7.6 a).

7.4.2 West Angelas - Mine face 2

Contrast-enhanced true-colour composites show changes in illumination as an increase in shade or as variations in intensity or color due to shading. Qualitatively, the GP-OAD effectively delineates the shale zones (right, 1 and 2) and the ore zones (left, 3-5) in both the fully illuminated and the shaded images (cf. Figures 7.8, 7.9 a and b). As expected, due to the way the GP-OAD works internally, i.e. using an angular and not a distance metric, only minor changes between the fully illuminated and the shade images were detected. Nevertheless, overt shade or shadow has profound and complex effects on the spectral curve shape which cannot be remedied using an angular metric to calculate similarities between a spectral library and the imagery. Therefore, the most obvious changes in classification are in areas of shade, e.g. marked by S_1 and S_2 and indicated by arrows in Figure 7.8. The dominant class near location S_1 in the fully illuminated image was MAR, in the shade image the dominant rock type is also MAR, however, a small number of pixels changed class to SHN, CLY but mainly to SHL. The reason for the relatively small numbers of pixels which changed classes between the shade and no-shade images is due to the nature of the metric used in the GP-OAD. Results obtained in Experiment 6 (Section 6.4.2) showed qualitatively that using a distance measure, for example, used in the squared exponential kernel, the mine face was misclassified almost entirely as BIF and SHL. Such a classification can, however, not be used as a reliable classification. Similarly to results obtained in Experiment 6, no GOL was detected by the GP-OAD and most likely confused with BIF due to their very similar spectra curves and the very large within-class variability of BIF.

Fusing the two GP outputs used to generate the previously discussed images (Figure 7.9 a and b) yields a classification result that is very similar to the classification of the fully illuminated image. On the left side of the image, the dominant rock type is MAR as the classification of the fully illuminated image shows, however, this region also seems to show less spatial variability at the scale of pixels. On the right side of the fused and classified image, SHL is the dominant rock type, however, there is a larger portion of the SHN class than in the original classification of the

fully illuminated image. The right side of the fused and classified image, however, shows less high-frequency spatial variability, i.e. less individual pixels of a different class occurring in this area. An increase in BIF on the mine face was, however, observed in the fused image which could not be confirmed with mappings made in the field. Possible reasons for this may be that BIF has a very large within-class variability in the training set and spectra can be very similar to spectra of GOL due to similar mineralogy. BIF, however, has a higher silica content than GOL which mostly influences the brightness in the SWIR, it is thus more difficult to distinguish from GOL or CHT. Another reason for the increase in BIF may be due to the over-confident classifications obtained from the GP-OAD algorithm. These may occur if the amount of noise is very different between the training and the test set which was the case for the training and test data used. Furthermore, an unbalanced training set may introduce a class-bias towards one class. Notwithstanding, the presented results suggest that both classifications were, overall, quite accurate except in areas of shade. Furthermore, the overall qualitative performance of the classification can be improved using the proposed framework which can be seen as taking the best parts from either classification on the basis of uncertainties. In this case, the GP-OAD was not over-confident compared to some of the results obtained using the Tom Price data-set. These results confirm, experimentally, that if no over-confident classifications occur and one of two results is inaccurate, the fused product will be at least as accurate as one of the two classifications.

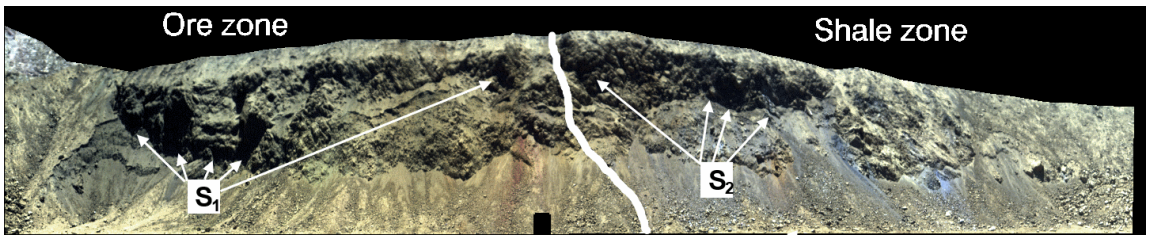
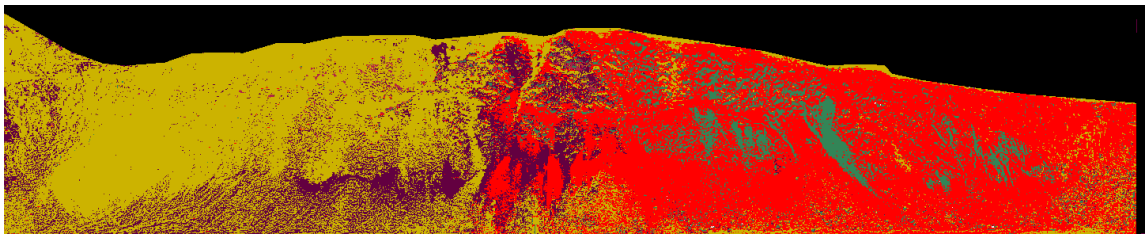
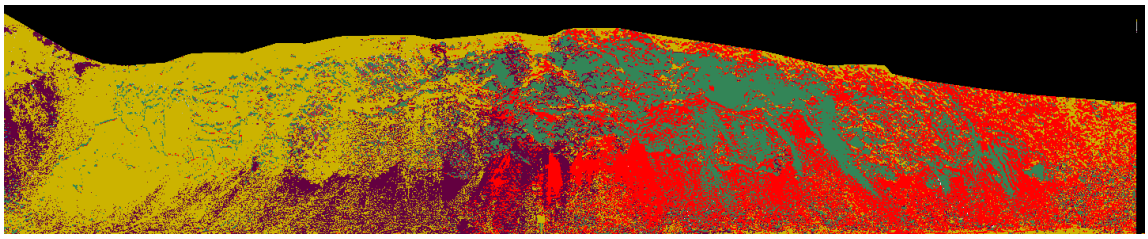


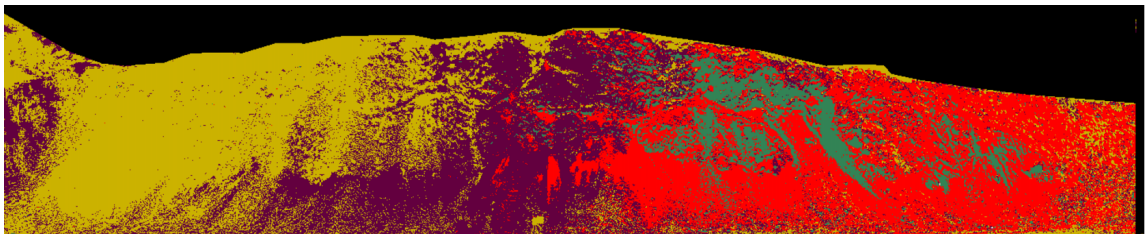
Figure 7.8 – WA-MF2 colour composite with ore and shale zones delineated. Areas affected by shade are indicated by arrows and labels S_1 and S_2 in the image. This image serves as an approximate ground-truth for this mine face.



(a) Classification of the fully illuminated mine face - 11:30am



(b) Classification of the partly shaded mine face - 1:30pm



(c) Classification after fusing the two GP outputs of the previous classifications

(d) Legend for WA-MF2 for 8 classes

Figure 7.9 – West Angelas mine face 2 classifications. GP-OAD classifications (a) and (b) are obtained from Figure 6.19. Classification (c) is the result obtained after fusing the GP outputs of the image classifications from (a) and (b). (d) Legend.

7.5 Summary and Conclusions

A framework for fusing GP outputs from hyperspectral data was introduced. The framework allows the fusion of more than two (up to n) classifications of hyperspectral images, in order to combine the information contained in all images. Fusion can be done simultaneously or sequentially for any number of observations. The necessary

equations were derived in Section 7.2. The concept of this method requires probabilities for the fusion process. They are obtained by modeling a normal probability distribution using a mean and variance from which probabilities can be calculated. Because of this, probabilistic SVMs (using Platt's probabilities) cannot be used within this fusion framework because they calculate probabilities directly (internally) and do not provide a mean and a variance to model a normal probability distribution. Probabilistic SVMs thus face similar problems as deterministic methods such as SAM when it comes to fusing several images using the proposed framework. However, all stochastic or Bayesian classifiers that yield a class estimate and a standard deviation should, in theory, be able to utilise this framework for fusing multiple probabilistic outputs.

The reflectance, of any given mineral surface, irrespective of the illumination source used, should theoretically remain constant, when calibrated using a reflectance standard. Absolute reflectance takes into account the reflectance coefficient of the calibration panel used, and thus makes a spectrum comparable to spectra of the same material acquired under different illumination conditions or using a different calibration panel. This rationale is the basis of many open source spectral libraries such as Jet Propulsion Laboratory (JPL) libraries (Grove et al., 1992), USGS (Clark et al., 1993), ASTER (Baldridge et al., 2009), which allow classification of unknown scenes using a well-defined set of spectra. The same rationale was applied in this work, where a spectral library was constructed from rocks collected from a mine face. Spectra were acquired from these rocks under the most optimal conditions possible, i.e. using a high-resolution field spectrometer and artificial illumination which optimised the signal-to-noise ratio. Similarly, a spectral library was acquired from drill cores of rock using laboratory conditions and applied to bulk rock material on a mine face, illuminated by natural sun light (i.e. using the WA-MF2 data-set).

This work demonstrates, that hyperspectral images of the same target acquired from the same position at different times during the day do, in fact, differ. Differences are (i) overt shade occurring in different areas on the mine face with the coarse of the day and (ii) changes in the colour temperature (i.e. the hue) between images. These

effects can be observed when comparing the four colour composites of the Tom Price data-set (cf. Figures 7.4a - d) which show different colours, yet are calibrated to the same reflectance standard. As a consequence, classification of these individual images does differ when using an independent training set, not constructed from each of the target images themselves. This suggests that there are contributing factors which cannot be compensated for using a reflectance standard or other means of reflectance calibration.

Variability in illumination on the mine face and between different images of the same mine face is caused by differing proportions of direct and indirect light. In the case of the latter, the light comes from a contribution of skylight and light reflected from surrounding fully-illuminated structures, including the mine face itself. The amount of light reflected from surrounding structures would depend on their composition and orientation. For example, this effect causes a red-shift in the scene across images because of the predominant red and yellow colours of the surroundings of the mine face (e.g. opposite and or adjacent mine faces). This means that a spectral library acquired with a halogen light source cannot take into account light reflected from surrounding structures and is inadequate for classifying shaded pixels. An indication of this was observed by comparing the quantitative results of the Tom Price images. The image acquired closest to the solar noon (img3) showed on average and for most rock types the highest F-scores because the wavelength distribution in the image and the spectral library were relatively similar. This is, however, a conjecture based on the presented data and cannot be generalised without further studies.

7.5.1 Implications for autonomous mining

As demonstrated, classifications using an independent library vary due to the mentioned effects, however, it is also impractical to generate spectral libraries from each image itself. If, however, multiple observations are fused, effects due to the wavelength distribution but also due to shading of certain areas at a certain point in time can be compensated. The proposed method for fusing several hyperspectral images

has large advantages for autonomous mining: (i) Classifications of a mine face using a probabilistic framework provides a measure of uncertainty around an estimate (in this case a classified pixel). (ii) Due to the proposed fusion framework, an existing estimate can become more certain once more information becomes available and thus can improve the overall classification accuracy. (iii) The need for human interaction to select representative spectra manually from an image for a particular class can be removed as well as the need to select 'the best' classification out of several available image classifications. (iv) The ability to fuse hyperspectral imagery acquired at different times - and possibly from different locations - can improve the classification of a mine face significantly, if, for example, the quality of one or more images is not high enough (e.g. due to dust). These key features are critical in the context of operational autonomous mining because they have the underlying structure to blend into existing frameworks currently used in autonomous mining.

This research has shown that an independent spectral library or libraries can be used to classify field based imagery of mine faces. This has significant implications for autonomous mining as a single large meta-library can be used to classify any number of images at a particular mine site without the need for human input. Although the use of spectral libraries may not compensate for all effects caused by variations in natural illumination, this approach does not require any *a priori* knowledge of a mine face. Once a spectral library has been built it can be applied to any image without input from experts - in particular the manual selection of training data from the image itself.

Chapter 8

Conclusions

8.1 Summary of the contributions of this thesis

Research contained within this thesis has investigated machine learning methods for the classification and fusion of hyperspectral data. Hyperspectral data were acquired in the laboratory and from field-based platforms using different sensors. Data therefore exhibited very different properties with respect to the signal-to-noise ratio of the data and the number of spectral bands.

The performance of classifiers when applied to these data were determined quantitatively and qualitatively. The GP-OAD method proposed in this work was compared to other commonly used methods for the classification of hyperspectral data.

A new framework that allows the integration of several hyperspectral data-sets from a particular target or scene was presented. This framework, probabilistic in nature, allows the probabilistic outputs from the GP-OAD method to be maintained. Improvements in the classification of the fused data-sets over individual classifications were presented. This chapter discusses and summarises the contributions of this thesis and proposes future research.

8.2 A probabilistic method for the classification of remotely sensed hyperspectral data

In chapter 5, Gaussian Processes for regression were used to learn a model of the spectral differences between classes of rock. For this, a *One versus All* (OvA) framework was chosen to separate (classify) one class of rock from all others. The OvA framework was applied to a data-set by classifying unknown spectra on a class-by-class basis, i.e. classification was run multiple times, independently, for each class. The number of OvA classifications performed for the classification of a single data-set is determined, therefore, by the number of classes in the training set.

The main contribution of Chapter 5 was the development of the GP-OAD algorithm which uses Gaussian Process Regression and a new covariance function (kernel), the Observation Angle Dependent (OAD) covariance function, first proposed by Melkumyan and Nettleton (2009). The proposed GP-OAD is a general framework for the classification of hyperspectral or high-dimensional data and is thus suitable to classify scenes of materials of unknown identity, without any *a priori* knowledge of the particular scene in question. This is the first time that the OAD covariance function has been applied to remotely sensed data of any kind, including hyperspectral data. As it is a supervised learning method, the GP-OAD is not restricted to classifying mineral materials but can be applied to a variety of tasks, e.g. to map vegetation or different soil types, as long as a suitable spectral library (training set) is provided.

Application of the GP-OAD method in this context was inspired by the spectral angle mapper (SAM) which was designed to be insensitive to brightness variations between spectral vectors. Spectra can exhibit differences in two basic ways; (i) in their curve shape and/or (ii) their brightness (albedo). Differences in spectral curve shape are caused, almost exclusively, by the intrinsic absorption and scattering characteristics of the minerals or rocks in question. Absorption of light at specific wavelengths is caused by specific physical-chemical processes associated with the composition of minerals or rocks. Scattering of light is dependent upon the physical characteristics of the rock such as its hardness and the grain-size of minerals of which it is comprised.

Brightness also varies due to broad absorptions across the entire spectral curve shape (400-2500 nm). These broad absorptions can be caused by, for example, the organic content of the rock and its thermal maturity (e.g. Rowan et al., 1991). In the context of remote sensing, variations in brightness of spectra are largely caused by differences in illumination geometry or intensity of incident light. These effects are unrelated to the intrinsic mineral composition of the rock and can have large negative impacts on the performance of classification. This was the motivation behind the development of SAM which was designed to be insensitive to brightness variations by using the spectral angle between the spectral vectors, rather than the distance between vectors or their length - as used for example in the Squared Exponential kernel. The relative insensitivity of SAM to variations in brightness has both positive and negative implications for classification of imagery of mine faces. A major positive consequence is that it enables SAM to classify rocks under conditions - such as those found in operational mines - where geological surfaces are complex and where illumination characteristics can change, often drastically, during the course of the day. If different types of rock have spectra which exhibit differences in brightness, but not in their spectral curve shape, then classification of these rocks, using SAM, would potentially become less reliable.

Because the GP-OAD also uses the spectral angle as an internal measure to construct correlations between training and test set, in theory, it would be affected in the same way as SAM. However, because the GP-OAD operates in a probabilistic framework, it can learn parameters from the data themselves (i.e. from an entire ensemble of spectra) which encapsulates the natural spectral variability of a class of rocks. In this way, variability among numerous library spectra is taken into account during classification, an approach which is fundamentally different to the traditional implementation of SAM where matches can only be made between individual pairs of spectra. Incorporation of variability into the spectral library makes the GP-OAD more effective in discriminating different classes of rocks. The OAD kernel is thus at least as insensitive to variations in brightness as is SAM. The Squared Exponential kernel, however, is very sensitive to variations in brightness.

Evaluations of the GP-OAD method were made using cross validation and using independent libraries and test data-sets. Results were compared to traditional classifiers, such as SAM and SVMs and to different kernels. The proposed GP-OAD outperformed the conventional methods by distinguishing more effectively between classes of rocks. This suggests that the GP-OAD is capable of modeling differences between inherent brightness of the materials and illumination conditions better than SAM and SVMs as well as other kernels (i.e. the SE kernel).

8.3 Application of the GP-OAD method to hyperspectral imagery and comparison with other methods

Often in the development of algorithms for classification, the development of the algorithm is done on spectra acquired by a field spectrometer using artificial light. The effectiveness of the algorithm is then directly tested using imagery acquired in the field. Such an approach is not able to test and isolate the causes of any changes in the performance of the developed algorithm when applied to data obtained under a different set of conditions using a different sensor. For example, it would not be possible to test if variations in the classification performance was due to differences in types of illumination (artificial versus natural) or due to differences in sensors (non-imaging versus imaging).

The contribution of Chapter 6 is to extend the use of the GP-OAD to hyperspectral imagery. A further contribution is a detailed evaluation of the performance of different classifiers under different sources of illumination and using different sensors. In addition, evaluation of the performance of the algorithm is often done using cross-validation and not using independent data-sets. Generally, no indications are made towards the capabilities of a method for use in conditions other than the one used for the development and testing of the algorithm. For example, when data have lower signal-to-noise ratios than field spectra, or when rocks and minerals are exposed to

a qualitative and quantitative variability in the incident illumination. A thorough assessment and comparison to other methods of the GP-OAD was made in a series of experiments to answer the following questions: (i) How does the number of spectral bands affect the performance of various classifiers when applied to field spectrometer data which were converted to imaging bandpasses and wavelengths? This is important to understand before moving to actual imagery because it enables isolating the different effects and properties of imaging and non-imaging hyperspectral data. (ii) How is the performance of classifiers affected when a GP model is learned from a non-imaging (field spectrometer) data-set but applied to hyperspectral imagery, with both data-sets acquired under similar illumination conditions (i.e. both using artificial illumination)? This separates the effects of number of bands and the use of different sensors from effects caused by different sources of illumination. (iii) How well do classifiers perform when a model is learned from convolved field spectrometer data acquired under controlled conditions (e.g. in the laboratory), but used to classify imagery acquired under natural illumination using a hyperspectral imaging sensor? This a fundamental question to answer in the context of autonomous mining because a standard library may be used to classify minerals and rocks on many different mine faces within a mine. The final contributions were an experimental assessment which included (iv) determining the best measure of similarity, i.e. an angular or a distance based metric for classification of hyperspectral data and (v) assessment of classifier performance when training classifiers on independent training sets acquired using a different sources of illumination than the test set.

8.4 A probabilistic framework for fusing thematic maps of hyperspectral images

A framework was developed that enables fusing several classified GP outputs derived from hyperspectral imagery (Chapter 7). This framework demonstrates the capabilities of a fully probabilistic framework by fusing several hyperspectral image classifications of the same scene acquired at different times during a day.

Using the proposed framework for fusing GP outputs, the benefits of a probabilistic method (e.g. GP-OAD) over deterministic methods (e.g. SAM or SVMs), are clearly demonstrated. The advantages of fusing outputs from probabilistic methods over retaining several individual classifications are (i) there is no requirement to objectively decide from several different, individual classifications which (if any) is the best one. (ii) The amount of data is reduced, however, not the amount of information; the information increases when new information becomes available because the uncertainty of the estimate is decreased. In the worst case, when new information is not useful, the uncertainty remains the same. (iii) Using class labels together with probabilities is superior to using just categorical class labels because probabilities can be fused easily and intuitively, a process which is not easily done using non-probabilistic methods. For example, from several SAM or SVM classifications of a given scene, one can chose the most frequently occurring class for a given pixel (spectrum) across all classifications and take this as the 'correct' label. Technically, this is also a way to fuse information, however, in this case, all the individual estimates would have equal weight - a situation which may or may not be appropriate. To assign weights to fuse deterministically-derived maps, one would need to devise another method to weight the different outputs appropriately, i.e. according to some discriminative function. Using the proposed probabilistic framework, weights are assigned automatically using the uncertainties which are provided by the GP itself, making it much easier to use.

8.5 Future Research

The research presented in this thesis has demonstrated the ability of hyperspectral remote sensing to provide maps of rocks and minerals on mine faces. In doing so, the research has provided a framework for the integration of hyperspectral data into automated processes within an operational mine. Further work will build upon this research along the following lines.

8.5.1 Comparison with other methods

This thesis investigated the GP-OAD and compared it to other kernels and even SVM and SAM. Further research could investigate the performance of the GP-OAD against feature-based methods (e.g. the SFF) in similar staged analysis. Such feature-based methods may be either used directly as deterministic methods or they may be used as a way to reduce the dimensionality of the data by extracting important spectral features (e.g. wavelength position, feature width and depth). These features may be used in a probabilistic framework (e.g. GP-OAD or GP-SE) for the classification of hyperspectral data. Furthermore, new machine learning methods maybe explored and compared to the GP-OAD (e.g. Boosting and / or Particle Swarm Optimisation; Monteiro et al., 2009, Tjiong and Monteiro, 2011, Wang et al., 2007).

8.5.2 Multi-class approach for the GP-OAD

The implementation of the GP-OAD has been limited to using the 'One versus all' framework in which classification is done on a class-by-class basis. Thus, the classification is run multiple times, independently, as many times as there are classes. In practice, the 'One versus all' approach is not optimal, however, depending on the application, it may nevertheless yield good classification results, especially where the spectral differences between classes are large. In the context of the classification of different minerals, spectral differences between classes can be subtle. The 'One versus All' approach, by operating on each class independently, only considers information for that class and not information amongst the constellation of all classes.

Further research will aim to extend the 'One versus All' classification approach into a multi-class framework using Gaussian Processes. Such an approach would allow the correlations between different classes to be learned and would integrate the inference stage for several classes into a unified process. It is necessary to allow information to propagate between classes so that information present in one class may improve the classification performance for other classes, particularly where classes are spectrally

similar. Such an approach could also, theoretically, reduce 'over-confident' classifications because an estimate of one class can be adjusted by considering the outcomes in the other classes.

8.5.3 Integration of other sensors for fusing hyperspectral data

The integration of hyperspectral data with data from other sensors is likely to become an increasing focus of research. Integration of data in a probabilistic framework has the potential to improve significantly maps of minerals and rocks on mine faces. For example, a laser scanner can be used to generate a three dimensional point cloud of a mine face during acquisition of a hyperspectral image (e.g. Nieto et al., 2010). From this point cloud, a continuous surface model can be estimated and registered with the hyperspectral image. Given the time of day of the acquisition, the orientation of the mine face and its exact location (GPS coordinates), a model can be generated that describes the diffuse light field over the mine face by taking into account its complex geometry. Using such a geometric illumination model, shaded areas on the mine face can be potentially identified. Further, such an illumination model could serve as a prior estimate for the mine face where areas that appear to be shaded could be set to a lower probability. When fused with other observations of the mine face, where different areas of the the mine have different priors (according to their own models of illumination), the estimate of the geology of the mine face may become more certain and more accurate overall. Similar approaches may be used to incorporate information from other sensor sources (e.g. RGB cameras) to aid in estimating the geology or geometry of a mine face.

8.6 Implications of this research for autonomous mining

The experiments done in this thesis showed that there was a slight reduction in the performance of classifiers as a consequence of reducing the spectral resolution of the

data from 2151 bands sensed by a non-imaging spectrometer to 283 bands sensed by an imaging spectrometer. A further reduction in the performance for all classifiers was observed when training and test sets were acquired using different sensors, i.e. a non-imaging and a imaging spectrometer. Reductions in performance was relatively small because of the constant illumination (artificial illumination) used for both the training and test data-sets. A further reduction in performance was observed when, in addition to using different types of sensors, the sources of illumination between the training and test set were different (i.e. artificial versus natural light). Yet given these constraints, the GP-OAD yielded classification results which were largely in close agreement with the distribution of minerals and rocks of the different mine faces mapped in the field. The GP-OAD was generally the method which performed best among the tested classifiers in terms of qualitative and quantitative results. A direct comparison of the GP-OAD algorithm developed as part of this research with other more frequently used classifiers, has provided a valuable insight into which types of classifiers are most appropriate for the classification of imagery of mine faces.

These results highlight important issues in the context of remote sensing of vertical mine faces. On a vertical mine face, variations in brightness are predominantly caused by intensity variations in the illumination and the geometry of the mine face. These effects are entirely unrelated to the mineral composition of the rock but can have large, negative impacts on the classification of a scene. The GP-OAD was designed to be insensitive to variations in brightness by using the spectral angle between the spectral vectors. Results presented here show that this was the case when compared to other standard methods for classification of hyperspectral data (Chapter 6).

The GP-OAD and the GP-SE performed similarly well in conditions where the sources of illumination between training and test set were the same (i.e. when artificial illumination was used; Experiments 4 and 5). For example, similar performance of the GP-OAD and GP-SE was obtained when applied to laboratory imagery (Experiment 5). In conditions where different sources of illumination were used for the training and test data-sets (Experiment 6), the GP-OAD and spectral angle based methods in general outperformed the SE kernel, irrespective of the framework used (e.g. GP

or SVMs). This means that the similar performance of the GP-OAD and GP-SE in Experiments 4 and 5 are largely caused by the consistent measurement geometry and consistent illumination used for both the training and test data. Lack of variability in the geometry and the amount of incident illumination enable the GP-SE to perform well compared with the GP-OAD and SAM. However, because large variations in illumination occur within individual images and between images acquired at different times during the day (Experiment 6) caused the GP-SE to consistently and inaccurately classify the types of rock on mine faces. Training for all classifiers was done using a spectral library acquired under a particular set of conditions (i.e. using artificial light). Variations in illumination found on mine faces could not be adequately modeled by any single spectral library acquired under artificial light.

These results highlight two important issues between the OAD and the SE kernels: The Squared Exponential covariance function calculates the distance of two vectors, therefore if the training and test data were acquired under similar and constant illumination, classification will be relatively accurate. When the test set was acquired using a different source of illumination than the training set, an offset in the curve shape due to brightness differences occurred. Thus, the Squared Exponential kernel was no longer able to account for offsets in the spectral curve shape, i.e. differences in brightness between the training and test sets. Because of this, the classifier is likely to confuse an inherently dark rock with, say, a bright one that is simply differently illuminated than the training examples. In other words, variability in brightness arising from variability in incident illumination is greater than the variability occurring as a result of geophysical-geochemical absorption and scattering processes. As a consequence, classification performance severely degraded for the SE kernel compared to the OAD kernel or conventional methods such as SAM.

The overarching conclusions from the above findings are that the GP-OAD can be successfully applied to any particular mine face (using natural illumination) to map minerals and rocks, even when the training data were acquired using a artificial illumination and a different type of sensor to that used to acquired imagery of a mine face. Although SAM uses an angular metric, and thus performs relatively well in

these situations, the GP-OAD has major advantages compared to the classical implementation of SAM as well as to other commonly used machine learning methods and kernels, including SVMs and the SE kernel. The GP-OAD (i) takes into account all the spectra in the training set, and thus models the variability within a certain class so that spectra that are unknown can be identified; and (ii) provides a probabilistic output (i.e. a mean or prediction and an uncertainty) for each sample, which is a requirement for data fusion and integration into autonomous systems. The results presented here show conclusively that the GP-OAD method produces consistently superior results in classification of rocks than does SAM and the GP-SE. The SE covariance function failed almost entirely when training and test sets were acquired under different illumination conditions. As the acquisition of hyperspectral data from field based platforms becomes more commonplace, algorithms need to be developed and tested to deal with spectral variability in illumination, including changing curve shape occurring as a result of variability in incident illumination. Furthermore, algorithms must be able to provide classifications based on independent spectral libraries. The GP-OAD method represents a significant step towards this goal.

For autonomous mining, there is a requirement that maps of geology are acquired frequently and within short periods of time. Information must be frequently and rapidly updated during the course of the day. Conventional and manual methods for mapping geology on mine faces require that mining operations are suspended for the duration of the survey. In addition, if XRD data are required in order to provide quantitative measures of minerals, it can take up to a day or several days until a geological map is accurately generated. The method presented in this thesis, can provide a thematic map of the geology of a mine within minutes. The time it takes to generate such a map depends on the size of the training data and the length of the swath of the image. For example, for the imagery presented in this thesis, processing times were between three to ten minutes. This is a significant improvement in terms of processing time compared to current methods employed in the mining process. In addition, there is no need to suspend the mining operations during these assessments. This is especially true if classified maps obtained from

several hyperspectral images can be fused. This is because fusing several hyperspectral images allows observing a mine face from several angles and at different times during the day, thus previously obscured or shaded[1] areas on the mine face may be observed clearly in a later image (Chapter 7). Alternatively, in cases where new data are poorly classified (i.e. high uncertainties), an autonomous algorithm could trigger a complete or partial recalibration of the classifier.

It is envisioned that the research presented in this thesis will provide tangible, practical benefits to the mining industry. With the increasing focus on mine automation and the timely acquisition of data, it is anticipated that robotic platforms will become a cost effective and reliable enhancement for the mining industry and other industries. Because for the operation of a mine many different tasks have to work 'hand-in-hand', different processes need to have the same underlying structure for the representation of current states of the individual processes. A probabilistic framework provides the the underlying structure that allows the integration and propagation of information across systems. Robotic platforms allow for a continuous acquisition of data during mining operations, for example, drones which acquire high spatial - and in due course - also high spectral resolution data. This research shows how these data can be used to provide information about the state of a mine face at any given period of time. Previous deterministic methods are unsuited for this task. In addition, due to data fusion capabilities of probabilistic methods, there is a smaller requirement for ground-truth measurements to validate maps. This has implications for improving safety as it removes the need to approach unstable mine faces. Using a probabilistic method such as the one presented in this thesis, timely information acquired with little human input can be fused autonomously to create a more accurate 'mine picture'. This ultimately leads to a safer, more efficient and productive mining process.

[1] In regard to shadow effects, multispectral laser scanners for illumination are under development and it can be expected that hyperspectral laser scanners may follow in due course.

Bibliography

[1] G Agamennoni, J I Nieto, and E M Nebot. Robust Inference of Principal Road Paths for Intelligent Transportation Systems. *Intelligent Transportation Systems, IEEE Transactions on*, 12(1):298–308, March 2011. ISSN 1524-9050.

[2] N Alajlan, Y Bazi, H Alhichri, and E Othman. Robust classification of hyperspectral images based on the combination of supervised and unsupervised learning paradigms. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 1417–1420, 2012. ISBN 2153-6996 VO -.

[3] S.-I. Amari, N Murata, K.-R. Muller, M Finke, and H H Yang. Asymptotic statistical theory of overtraining and cross-validation. *Neural Networks, IEEE Transactions on*, 8(5):985–996, 1997.

[4] S Arlot and A Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.

[5] R Baissa, K Labbassi, P Launeau, A Gaudin, and B Ouajhain. Using HySpex SWIR-320m hyperspectral data for the identification and mapping of minerals in hand specimens of carbonate rocks from the Ankloute Formation (Agadir Basin, Western Morocco). *Journal of African Earth Sciences*, 61(1):1–9, August 2011. ISSN 1464-343X.

[6] A M Baldridge, S J Hook, C I Grove, and G Rivera. The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113(4):711–715, April 2009. ISSN 0034-4257.

[7] F Baret, G Guyot, and D J Major. TSAVI: A vegetation vndex which minimizes soil brightness effects on LAI and APAR estimation. In *Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International*, volume 3, pages 1355–1358, 1989. ISBN VO - 3.

[8] A Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik und Chemie*, 86:78–88, 1852.

[9] A Bekker, J F Slack, N Planavsky, B Krapež, A Hofmann, K O Konhauser, and O J Rouxel. Iron Formation: The Sedimentary Product of a Complex Interplay among Mantle, Tectonic, Oceanic, and Biospheric Processes. *Economic Geology*, 105(3):467–508, 2010.

[10] R Bellman. *Dynamic Programming*. Princeton, Princeton University Press, 1st prince edition, 1957.

[11] J A Benediktsson and I Kanellopoulos. Classification of multisource and hyperspectral data based on decision fusion. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(3):1367–1377, 1999.

[12] J A Benediktsson and J R Sveinsson. Consensus Based Classification of Multisource Remote Sensing Data. In *Multiple Classifier Systems SE - 27*, volume 1857 of *Lecture Notes in Computer Science*, pages 280–289. Springer Berlin Heidelberg, 2000. ISBN 978-3-540-67704-8.

[13] J A Benediktsson, P H Swain, and O K Ersoy. Neural Network Approaches Versus Statistical Methods In Classification Of Multisource Remote Sensing Data. *Geoscience and Remote Sensing, IEEE Transactions on*, 28(4):540–552, 1990.

[14] A Berk, L S Bernstein, G P Anderson, P K Acharya, D C Robertson, J H Chetwynd, and S M Adler-Golden. MODTRAN Cloud and Multiple Scattering Upgrades with Application to AVIRIS. *Remote Sensing of Environment*, 65(3):367–375, September 1998. ISSN 0034-4257.

[15] D P Bertsekas and J N Tsitsiklis. *Introduction to probability*, volume 1. Athena Scientific, Belmont, MA, 1 edition, 2002. ISBN 188652940X.

[16] C M Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, New York, 1995.

[17] C M Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

[18] G A Blackburn. Relationships between Spectral Reflectance and Pigment Concentrations in Stacks of Deciduous Broadleaves. *Remote Sensing of Environment*, 70(2):224–237, November 1999. ISSN 0034-4257.

[19] N Bohr. I. On the constitution of atoms and molecules. *Philosophical Magazine Series 6*, 26(151):1–25, July 1913. ISSN 1941-5982.

[20] M Born. Max Karl Ernst Ludwig Planck. 1858-1947. *Obituary Notices of Fellows of the Royal Society*, 6(17):161–188, November 1948.

[21] P Boyle and M Frean. Dependent gaussian processes. In L K. Saul, Y Weiss, and L'eon Bottou, editors, *Advances in Neural Information Processing Systems 17*, volume 17, pages 217–224. MIT Press, Cambridge, MA, USA, 2005.

[22] A C Braun, U Weidner, and S Hinz. Support vector machines, import vector machines and relevance vector machines for hyperspectral classification; A comparison. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pages 1–4, 2011.

[23] L Breiman and P Spector. Submodel selection and evaluation in regression: X-random case. *International Statistical Review*, 60:291–319, 1992.

[24] C J C Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Knowledge Discovery and Data Mining*, 2(2), 1998.

[25] R Burns. *Mineralogical Applications of Christal Field Theory.* Cambrdige University Press, Cambridge, 1970.

[26] R G Burns. Origin of Electronic Spectra of Minerals in the Visible and Near-Infrared Region. In C.M; Pieters P. Englert, editor, *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pages 3–30. Cambridge University Press, New York, 1993.

[27] Caterpillar.
Caterpillar To Develop Autonomous System For Blasthole Drills. Press Release. http://australia.cat.com/cda/components/fullArticle?m=60164&x=7&id=1064002, September 2008.

[28] Q Cheng. The Fisher-Markov Selector: Fast Selecting Maximally Separable Feature Subset for Multiclass Classification with Applications to High-Dimensional Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(6):1217–1233, October 2011. ISSN 0162-8828.

[29] A. Chlingaryan, A Melkumyan, R J Murphy, and S Schneider. Automated mutli-class classification of hyperspectral data using Gaussian Processes. In *Application of Computers and Operations Research in the Mineral Industry (APCOM)*, Porto Alegre, Brazil, 2013.

[30] M A Cho and A. K. Skidmore. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment*, 101(2):181–193, March 2006. ISSN 0034-4257.

[31] M A Cho, P Debba, R Mathieu, L Naidoo, J van Aardt, and G P Asner. Improving Discrimination of Savanna Tree Species Through a Multiple-Endmember Spectral Angle Mapper Approach: Canopy-Level Analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(11): 4133–4142, 2010.

[32] R N Clark and T L Roush. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *Journal of Geographical Research*, 89, 1984.

[33] R N Clark, T V V King, and N Gorelick. Automatic continuum analysis of reflectance spectra. In *Proceedings, 3rd AIS workshop*, volume 87-30, pages 138–142, Jet Propulsion Laboratory, Pasadena, California, 1987. JPL Publication.

[34] R N Clark, A J Gallagher, and G A Swayze. Material Absorption Band Depth Mapping of Imaging Spectrometer Data Using a Complete Band Shape Least-Squares Fit with Library Reference Spectra . In *Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, volume 90-54, pages 176–186. JPL Publication, 1990.

[35] R N Clark, G A Swayze, A J Gallagher, T V V King, and W.M. Calvin. The U. S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 microns. *U.S. Geological Survey Open File Report*, 93-592:1340, 1993.

[36] R N Clark, G A Swayze, K E Livo, R F Kokaly, S J Sutley, J Brad Dalton, R R McDougal, and C A Gent. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geographical Research*, 108, 2003.

[37] Roger N Clark, T V V King, M Klejwa, G A Swayze, and N Vergo. High spectral resolution reflectance spectroscopy of minerals. *Journal of Geographical Research*, 95(B8):12653–12680, 1990.

[38] J.M.F. Clout and B.M Simonson. Precambrian iron formation and iron formation-hosted iron ore deposits. *in Hedenquist, J.W., Thompson, J.F.H., Goldfarb, R.J. and Richards, J.P. (eds.), Economic Geology*, (100th Anniversary Volume Society of Economic Geologists):643–679, 2005.

[39] E A Cloutis, M J Gaffey, D G W Smith, and R St. J Lambert. Reflectance spectra of "featureless" materials and the surface mineralogies of M- and E-class asteroids. *Journal of Geophysical Research: Solid Earth*, 95(B1): 281–293, January 1990. ISSN 2156-2202.

[40] E A Cloutis, T Hiroi, M J Gaffey, C.M.O'D. Alexander, and P Mann. Spectral reflectance properties of carbonaceous chondrites: 1. CI chondrites. *Icarus*, 212(1):180–209, March 2011. ISSN 0019-1035.

[41] A.P Colvocoresses. Multispectral linear arrays as an alternative to Landsat D. *Photogrammetric Engineering and Remote Sensing*, 45(1):67–69, 1979.

[42] R G Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing Of Environment*, 37(1):35–46, 1991.

[43] R. G. Congalton, R. G. Oderwald, and R. A. Mead. Assessing Landsat classification accuracy using discrete multivariate-analysis statistical techniques. *Photogrammetric Engineering & Remote Sensing*, 49(12): 1671–1678, 1983.

[44] C Cortes and V Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995. ISSN 0885-6125.

[45] N Cressie. *Statistics for spatial data*. Wiley, New York, 1993.

[46] J K Crowley, D W Brickey, and L C Rowan. Airborne imaging spectrometer data of the Ruby Mountains, Montana: Mineral discrimination using relative absorption band-depth images. *Remote Sensing Of Environment*, 29(2):121, 1989.

[47] CSIRO. HyLogging systems. 2012.

[48] CSIRO and AuScope. National Virtual Core Library. 2008.

[49] T J Cudahy and E R Ramanaidou. Measurement of the hematite:goethite ratio using field visible and nearinfrared reflectance spectrometry in channel iron deposits, Western Australia. *Australian Journal of Earth Sciences*, 44(4): 411–420, 1997.

[50] S M De Jong and F van der Meer. *Remote Sensing and Digital Image Analysis : Including the Spatial Domain*. Remote sensing and digital image processing ; v. 5. Kluwer Academic, New York, Boston, Dordrecht, London, Moscow, 2004. ISBN 9781402025594.

[51] T H Demetriades-Shah, M D Steven, and J A Clark. High resolution derivative spectra in remote sensing. *Remote Sensing Of Environment*, 33(1):55, 1990.

[52] P E Dennison, K Q Halligan, and D A Roberts. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment*, 93(3):359–367, November 2004. ISSN 0034-4257.

[53] J C Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3:32–57, 1973. ISSN 0022-0280.

[54] C Elachi and J Van Zyl. *Introduction to the Physics and Techniques of Remote Sensing*. John Wiley & Sons, 2nd edition, 2006. ISBN 978-0-471-47569-9.

[55] K A Evans, T C McCuaig, D Leach, T Angerer, and S G Hagemann. Banded iron formation to iron ore: A record of the evolution of Earth environments? *Geology*, 41(2):99–102, February 2013.

[56] V. C. Farmer. *The Infra-red Spectra of Minerals*. Mineralofical Society, London, 1974.

[57] M Fauvel, J Chanussot, and J A Benediktsson. Decision Fusion for the Classification of Urban Remote Sensing Images. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(10):2828–2838, 2006.

[58] M. J. Gaffey and T. B McCord. Asteroid surface materials - Mineralogical characterizations and cosmological implications. In *8th Lunar Science Conference*, pages 113–143, Houston, TX, 1977. New York, Pergamon Press, Inc.

[59] M J Gaffey, K L Reed, and M S Kelley. Relationship of E-type Apollo asteroid 3103 (1982 BB) to the enstatite achondrite meteorites and the Hungaria asteroids. *Icarus*, 100(1):95–109, November 1992. ISSN 0019-1035.

[60] S. J. Gaffey, L. A. McFadden, D. Nash, and C.M. Pieters. Ultraviolet, visible, and near-infrared reflectance spectroscopy: laboratory spectra of geologic materials. In C. M. Pieters and P. A. J. Englert, editors, *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pages 43–78. Cambrdige University Press, Cambridge, 1993.

[61] M Gianinetto and G Lechi. The development of Superspectral approaches for the improvement of land cover classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(11):2670–2679, 2004.

[62] A F H Goetz and Wendy M Calvin. Imaging Spectrometry: Spectral Resolution And Analytical Identification Of Spectral Features. In *Imaging spectroscopy II*, pages 158–165, San Diego, CA, January 1987. Society of Photo-Optical Instrumentation Engineers 1988.

[63] A F H Goetz, L C Rowan, and M J Kingston. Mineral Identification From Orbit - Initial Results From The Shuttle Multispectral Infrared Radiometer. *Science*, 218(4576):1020–1024, 1982.

[64] A F H Goetz, G Vane, J E Solomon, and B N Rock. Imaging spectrometry for earth remote sensing. *Science*, v228:p1147(7), 1985.

[65] Alexander F H Goetz. Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing Of Environment*, 113(Supplement 1): S5, 2009.

[66] R O Green, M L Eastwood, C M Sarture, T G Chrien, M Aronsson, B J Chippendale, J A Faust, B E Pavri, C J Chovit, M Solis, M R Olah, and O Williams. Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing Of Environment*, 65(3):227–248, 1998.

[67] M Grewal and A Andrews. *Kalman Filtering : Theory And Practice*. John Wiley, New York, 1993.

[68] C. I. Grove, S. J. Hook, and E. D Paylor. Laboratory reflectance spectra for 160 minerals 0.4 - 2.5 micrometers. *Jet Propulsion Laboratory, Padadena, CA.*, (JPL Publication 92-2), 1992.

[69] Bruce H. *Theory of reflectance and emittance spectroscopy*, volume 3. Cambrdige University Press, Camebridge, New York, Melbourne, 2 edition, 2012. ISBN 0-521-30789-9.

[70] Maarten Haest, Thomas Cudahy, Carsten Laukamp, and Sean Gregory. Quantitative Mineralogy from Infrared Spectroscopic Data. I. Validation of Mineral Abundance and Composition Scripts at the Rocklea Channel Iron Deposit in Western Australia. *Economic Geology*, 107(2):209–228, 2012.

[71] Jiawei Han and Micheline Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[72] T Hastie, R Tibshirani, and J Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009.

[73] E Hecht. *Optics*. Addison-Wesley, 4th edition, 2002.

[74] C Hecker, M van der Meijde, H van der Werff, and F D van der Meer. Assessing the Influence of Reference Spectra on Synthetic SAM Classification Results. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(12): 4162–4172, 2008.

[75] H Hertz. Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung. *Annalen der Physik*, 267(8):983–1000, January 1887. ISSN 1521-3889.

[76] D N H Horler, M Dockray, and J Barber. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2):273–288, January 1983. ISSN 0143-1161.

[77] Ellis Horwood. Machine Learning, Neural and Statistical Classification. 1994.

[78] C Huang, L S Davis, and J R G Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, 2002.

[79] W.D. Hudson and C.W. Ramm. Correct formulation of the Kappa coefficient of agreement. *Photogrammetric Engineering & Remote Sensing*, 53:421–422, 1987.

[80] A R Huete and R D Jackson. Soil and atmosphere influences on the spectra of partial canopies. *Remote Sensing of Environment*, 25(1):89–105, June 1988. ISSN 0034-4257.

[81] G Hughes. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14(1):55–63, 1968.

[82] R L Huguenin and J L Jones. Intelligent Information Extraction From Reflectance Spectra: Absorption Band Positions. *Journal of Geographical Research*, 91, 1986.

[83] G R Hunt. Spectral Signatures of particulate minerals, in the visible and near-infrared. *Geophysics*, 42:501–513, 1977.

[84] G. R. Hunt. Spectroscopic properties of rocks and minerals. In R.S. Carmichael, editor, *Handbook of Physicial Properties of Rocks*, pages 295–385. CRC Press, Boca Raton, Fla, vol 1 edition, 1982.

[85] G R Hunt and J W Salisbury. Visible and near-infrared spectra of minerals and rocks; I, Silicate minerals. *Modern Geology*, 1(7):283–300, 1970.

[86] G R Hunt J. W. Salisbury and C.J. Lenhoff. Visible and near-infrared spectra of minerals and rocks; III, Oxides and Hydroxides. *Modern Geology*, 1: 195–205, 1971.

[87] J Huntington, A Mauger, R Skirrow, and E Bastrakov. Automated mineralogical logging of core from the Emmie Bluff, iron oxide copper-gold prospect, south Australia. In *Pacrim 2004 Congress*, pages 223–230, Parkville, Victoria, 2004.

[88] A Hutsinpiller. Discrimination of hydrothermal alteration mineral assemblages at Virginia City, Nevada, using the Airborne Imaging Spectrometer. *Remote Sensing of Environment*, 24(1):53–66, February 1988. ISSN 0034-4257.

[89] Marcus Hutter. On generalized computable universal priors and their convergence. *Theoretical Computer Science*, 364(1):27–41, November 2006. ISSN 0304-3975.

[90] J.R. Irons, R.A. Weismiller, and G.W. Petersen. Soil Reflectance. In G. Asrar, editor, *Theory and application of optical remote sensing*, pages 66 – 106. John Wiley & Sons, Inc., New York, NY, USA, 1989.

[91] Jet Propulsion Laboratory. AVIRIS Concept. 2007.

[92] C F Jordan. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, 50(4):663–666, July 1969. ISSN 00129658.

[93] S Kawata and S Minami. Adaptive Smoothing of Spectroscopic Data by a Linear Mean-Square Estimation. *Appl. Spectrosc.*, 38(1):49–58, January 1984.

[94] N Keshava. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(7):1552–1565, 2004.

[95] S Kim and J Kim. Building occupancy maps with a mixture of Gaussian processes. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4756–4761, 2012.

[96] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Morgan Kaufmann, editor, *International Joint Conference on Artificial Intelligence*, pages 1139–1143, 1995.

[97] R F Kokaly, D G Despain, R N Clark, and K E Livo. Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing Of Environment*, 84(3):437–456, 2003.

[98] A N Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1):4–7, 1965.

[99] G. Kortum. *Reflectance Spectroscopy: Principles, Methods, Applications*. Springer, New York, NY, 1969.

[100] F A Kruse. 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, 1988.

[101] F A Kruse, A B Lefkoff, J W Boardman, K B Heidebrecht, A T Shapiro, P J Barloon, and A F H Goetz. The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data. *Remote Sensing Of Environment*, 44(2-3):145, 1993.

[102] F A Kruse, R L Bedell, J V Taranik, W A Peppin, O Weatherbee, and W M Calvin. Mapping alteration minerals at prospect, outcrop and drill core scales using imaging spectrometry. *International Journal of Remote Sensing*, 33(6): 1780–1798, 2011.

[103] D. Kuehlke. *Optik Grundlagen und Anwendungen*. Verlag Harri Deutsch, Frankfurt, Germany, 2nd edition, 2007.

[104] T Kurz, S Buckley, J Howell, and D Schneider. Geological outcrop modelling and interpretation using ground based hyperspectral and laser scanning data fusion. *ISPRS, International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, XXXVII(8):1229–1234, 2008.

[105] T.H Kurz, S.J Buckley, J.A Howell, and D. Schneider. Integration of panoramic hyperspectral imaging with terrestrial lidar data. *The Photogrammetric Record*, 26:212–228, 2011.

[106] D.A. Landgrebe. Useful information from multispectral image data: Another look. In P. H. Swain and S. M. Davis, editors, *Remote Sensing: The Quantitative Approach*, page 366. McGraw-Hill, New York, 1978.

[107] D F Lascelles. Marra Mamba Iron Formation stratigraphy in the eastern Chichester Range, Western Australia. *Australian Journal of Earth Sciences*, 47(4):799–806, August 2000. ISSN 1440-0952.

[108] Q V Le, A J Smola, and S Canu. Heteroscedastic Gaussian process regression. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, pages 489–496, New York, NY, USA, 2005. ACM. ISBN 1-59593-180-5.

[109] G Licciardi, F Pacifici, D Tuia, S Prasad, T West, F Giacco, C Thiel, J Inglada, E Christophe, J Chanussot, and P Gamba. Decision Fusion for the Classification of Hyperspectral Data: Outcome of the 2008 GRS-S Data Fusion Contest. 2009.

[110] T Maekawa, T Noda, S Tamura, T Ozaki, and K Machida. Curvature continuous path generation for autonomous vehicle using B-spline curves. *Computer-Aided Design*, 42(4):350–359, 2010.

[111] P Mather and K Magaly. *Computer Processing of Remotely-Sensed Images: An Introduction*. John Wiley & Sons, Chichester, West Sussex, UK, 4th edition, 2010.

[112] John J McCall. Induction: From Kolmogorov and Solomonoff to De Finetti and Back to Kolmogorov. *Metroeconomica*, 55(2-3):195–218, May 2004. ISSN 1467-999X.

[113] A Melkumyan and E Nettleton. An Observation Angle Dependent Nonstationary Covariance Function for Gaussian Process Regression. In *Lecture Notes in Computer Science*, volume Neural Inf, pages 331–339. Springer, 2009.

[114] A. Melkumyan and F.T Ramos. A Sparse covariance function for exact gaussian process inference in large datasets. In *In International Joint Conferences on Artificial Intelligence*, 2009.

[115] A. Melkumyan, P.J. Hatherly, and H. Zhou. Fusion of drill monitoring data with geological borehole assays. In *Proceedings of the 12th ISRM International Congress on Rock Mechanics*, pages 2125–2128, Beijing, China, 2012.

[116] R A Millikan. A Direct Photoelectric Determination of Planck's "h". *Physical Review*, 7(3):355–388, March 1916.

[117] E Mjolsness and D DeCoste. Machine Learning for Science: State of the Art and Future Prospects. 2001.

[118] S T Monteiro, J R Richard, R Fabio, and Juan Nieto. Applying Boosting For Hyperspectral Classification Of Ore-Bearing Rocks. In *IEEE International Workshop on Machine Learning for Signal Processing*, volume XIX, pages 1–6, 2009.

[119] R Morris and E. Ramanaidou. Genesis of the channel iron deposits (CID) of the Pilbara region, Western Australia. *Australian Journal of Earth Sciences*, 54:733–756, 2007.

[120] R C Morris. A textural and mineralogical study of the relationship of iron ore to banded iron-formation in the Hamersley iron province of Western Australia. *Economic Geology*, 75(2):184–209, 1980.

[121] R V Morris, H V Lauer, C A Lawson, E K Gibson, G A Nace, and C Stewart. Spectral and other physicochemical properties of submicron powders of hematite (α-Fe2O3), maghemite (γ-Fe2O3), magnetite (Fe3O4), goethite (α-FeOOH), and lepidocrocite (γ-FeOOH). *Journal of Geophysical Research: Solid Earth*, 90(B4):3126–3144, March 1985. ISSN 2156-2202.

[122] R J Murphy and Sildomar T Monteiro. Mapping the distribution of ferric iron minerals on a vertical mine face using derivative analysis of hyperspectral imagery (430 - 970 nm). *ISPRS Journal of Photogrammetry & Remote Sensing*, 75:29–39, 2013.

[123] R J Murphy, M H Pinkerton, K M Richardson, J M Bradford-Grieve, and P W Boyd. Phytoplankton distributions around New Zealand derived from SeaWiFS remotely sensed ocean colour data. *New Zealand Journal of Marine and Freshwater Research*, 35(2):343–362, June 2001. ISSN 0028-8330.

[124] R J Murphy, A J Underwood, M H Pinkerton, and P Range. Field spectrometry: New methods to investigate epilithic micro-algae on rocky shores. *Journal of Experimental Marine Biology and Ecology*, 325(1):111–124, 2005.

[125] R J Murphy, A J Underwood, T J Tolhurst, and M G Chapman. Field-based remote-sensing for experimental intertidal ecology: Case studies using hyperspatial and hyperspectral data for New South Wales (Australia). *Remote Sensing Of Environment*, 112(8):3353, 2008.

[126] R J Murphy, S T Monteiro, and S Schneider. Evaluating Classification Techniques for Mapping Vertical Geology Using Field-Based Hyperspectral Sensors. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(8): 3066–3080, 2012.

[127] R J Murphy, S Schneider, and S T Monteiro. Consistency of Measurements of Wavelength Position From Hyperspectral Imagery: Use of the Ferric Iron Crystal Field Absorption at 900 nm as an Indicator of Mineralogy. *Geoscience and Remote Sensing, IEEE Transactions on*, PP(99):1–15, 2013.

[128] M. Neal. Bayesian Learning for Neural Networks. In *Lecture Notes in Statistics 118*. Springer, 1996.

[129] D Nguyen-Tuong and J Peters. Learning Robot Dynamics for Computed Torque Control Using Local Gaussian Processes Regression. In *Learning and Adaptive Behaviors for Robotic Systems, 2008. LAB-RS '08. ECSIS Symposium on*, pages 59–64, 2008.

[130] J I Nieto, S T Monteiro, and D Viejo. 3D geological modelling using laser and hyperspectral data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 4568–4571, 2010. ISBN 2153-6996.

[131] N J Nilsson. Introduction to Machine Learning: An Early Draft of a Proposed Textbook. http://robotics.stanford.edu/people/nilsson/mlbook.html. 1996.

[132] J Nocedal and S J Wright. *Numerical Optimization*. Springer-Verlag New York, Inc., 2006.

[133] Stanley Robson de Medeiros Oliveira. Data Transformation For Privacy-Preserving Data Mining. University of Alberta Library, 2005.

[134] D. L. Olson and D. Delen. *Advanced Data Mining Techniques*. Springer-Verlag, Berlin, Germany, 2008.

[135] B Park, W R Windham, K C Lawrence, and D P Smith. Contaminant Classification of Poultry Hyperspectral Imagery using a Spectral Angle Mapper Algorithm. *Biosystems Engineering*, 96(3):323–333, March 2007. ISSN 1537-5110.

[136] F.H. Perrin. The structure of the developed image. In T. James, editor, *The theory of photographic process*, pages 499–551. MacMillan, New York, 1966.

[137] J.C. Platt. Probabilities for SV Machines. In A. Smola, P. Bartlett, B. Schoelkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 2000.

[138] A Plaza, J A Benediktsson, J W Boardman, J Brazile, L Bruzzone, G Camps-Valls, J Chanussot, M Fauvel, P Gamba, A Gualtieri, M Marconcini, J C Tilton, and G Trianni. Recent advances in techniques for hyperspectral image processing. *Remote Sensing Of Environment*, 113 (Supplement 1):S110, 2009.

[139] E Ramanaidou and M Wells. Hyperspectral Imaging of Iron Ores. In Maarten A T M Broekmans, editor, *Proceedings of the 10th International Congress for Applied Mineralogy (ICAM) SE - 68*, pages 575–580. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-27681-1.

[140] E. Ramanaidou, P Connor, A Cornelius, and S Fraser. Imaging spectroscopy for iron ore mine faces. In *Iron Ore Conference*, pages 155–157, Perth, Australia, 2002.

[141] E Ramanaidou, M Wells, D Belton, M Verrall, and C Ryan. Mineralogical and Microchemical Methods for the Characterization of High-Grad Banded Iron Formation - Derived Iron Ore. *Society of Economic Geologists Reviews*, 15: 129–156, 2008.

[142] C Rasmussen. Gaussian processes in machine learning. In O Bousquet, U Luxburg, and G Rätsch, editors, *Lecture Notes in Artificial Intelligence 3176 - Advanced Lectures on Machine Learning*, chapter Gaussian P, pages 63–71. Springer-Verlag, Heidelberg, 2004. ISBN 978-3-540-23122-6.

[143] C E Rasmussen and C K I Williams. *Gaussian Process for Machine Learning*. The MIT Press, Cambridge, Massachusetts, 2006.

[144] A Reid, F Ramos, and S Sukkarieh. Multi-class classification of vegetation in natural environments using an Unmanned Aerial system. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2953–2959, 2011.

[145] A N Rencz. *Remote Sensing for the Earth Sciences - Manual of Remote Sensing*. John Wiley & Sons, New York, third edition, 1999. ISBN 978-0471-29405-4.

[146] A J Richardson and J H Everitt. Using spectral vegetation indices to estimate rangeland productivity. *Geocarto International*, 7(1):63–69, March 1992. ISSN 1010-6049.

[147] R Rifkin and A Klautau. In Defense of One-Vs-All Classification. *J. Mach. Learn. Res.*, 5:101–141, December 2004. ISSN 1532-4435.

[148] RioTinto. Australia wins Rio Tinto's mining automation effort. http://www.riotinto.com/media/18435_media_releases_6473.asp. Rio Tinto, July 2007.

[149] RioTinto. Mine of the Future. http://www.riotinto.com/ourapproach/17203_mine_of_the_future.asp. Rio Tinto, January 2013.

[150] A Rodger and T Cudahy. Vegetation corrected continuum depths at 2.20µm: An approach for hyperspectral sensors. *Remote Sensing of Environment*, 113 (10):2243–2257, October 2009. ISSN 0034-4257.

[151] A Rodger, C Laukamp, M Haest, and T Cudahy. A simple quadratic method of absorption feature wavelength estimation in continuum removed spectra. *Remote Sensing of Environment*, 118(0):273–283, March 2012. ISSN 0034-4257.

[152] J.W. Rouse, R.H. Hass, J.A. Schell, and D.W Deering. Monitoring vegetation systems in the great plains with ERTS. In *Third NASA ERTS Symposium*, pages 309–317, 1973.

[153] L C Rowan. Application of Satellites to Geologic Exploration: Recent experiments in two spectral regions, the visible and near-infrared and the thermal-infrared, confirm the value of satellite observations for geologic exploration of the earth. *American Scientist*, 63(4):393–403, July 1975. ISSN 00030996.

[154] L C Rowan and J C Mars. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Remote Sensing of Environment*, 84(3):350–366, March 2003. ISSN 0034-4257.

[155] L C. Rowan and P H. Wetlaufer. Iron-absorptionband analysis for the discrimination of iron-rich zones. In *U.S. National Aeronautics and Space Administration Report*, 1974.

[156] L C Rowan, A F H Goetz, and R P Ashley. Discrimination of hydrothermally altered and unaltered rocks in visible and near infrared multispectral images. *Geophysics*, 42(3):522–535, 1977.

[157] L C Rowan, J W Salisbury, M J Kingston, N Vergo, and N H Bostick. Evaluation of visible and near-infrared and thermal-infrared reflectance spectra for studying thermal alteration of Pierre Shale, Wolcott, Colorado. *Journal of Geophysical Research: Solid Earth*, 96(B11):18047–18057, October 1991. ISSN 2156-2202.

[158] S Sarata, N Koyachi, T Tubouchi, H Osumi, M Kurisu, and K Sugawara. Development of Autonomous System for Loading Operation by Wheel Loader. In *The 23rd International Symposium on Automation and Robotics in Construction*, pages 466–471, 2006.

[159] A Savitzky and M J E Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

[160] C Schaffer. Technical Note: Selecting a Classification Method by Cross-Validation. *Machine Learning*, 13(1):135–143, 1993. ISSN 0885-6125.

[161] A. C. Scheinost. Use and Limitations of Second-Derivative Diffuse Reflectance Spectroscopy in the Visible to Near-Infrared Range to Identify and Quantify Fe Oxide Minerals in Soils. *Clays and Clay Minerals*, 46(5):528–536, 1998. ISSN 0009-8604.

[162] S Schneider, R J Murphy, S Monteiro, and E W Nettleton. On the development of a hyperspectral library for autonomous mining. In *Proceedings of Australasian Conference on Robotics and Automation (ACRA)*, Sydney, Australia, 2009.

[163] S Schneider, A Melkumyan, R J Murphy, and E W Nettleton. Gaussian Processes with OAD covariance function for hyperspectral data classification. In *Proceedings of IEEE International Conference on Tools with Artificial Intelligence*, Arras, France, 2010.

[164] S Schneider, R J Murphy, A Melkumyan, and E Nettleton. Classification of Hyperspectral Data of Mine Face Geology Using Machine Learning for Autonomous Open Pit Mining. In *Application of Computers and Operations Research in the Mineral Industry (APCOM)*, Wollongong, Australia, 2011. The Australasian Institute of Mining and Metallurgy (The AusIMM).

[165] B. Schoelkopf, C. Burges, and A. Smola. Introduction to SupportVector Learning. Chap.1. In *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1999.

[166] B Schölkopf and A J Smola. *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, Massachusetts, 2002.

[167] R A Schowengerdt. *Remote Sensing - Models and Methods for Image Processing*. Academic Press, New York, second edition, 1997.

[168] M. Seeger. Gaussian Processes for Machine Learning. *International Journal of Neural Systems*, 14(02):69–106, April 2004. ISSN 0129-0657.

[169] V M Sellitto, R B A Fernandes, V Barrón, and C Colombo. Comparing two different spectroscopic techniques for the characterization of soil iron oxides: Diffuse versus bi-directional reflectance. *Geoderma*, 149(1–2):2–9, February 2009. ISSN 0016-7061.

[170] R. A. Serway and J.W. Jewett. *Physics for Scientists and Engineers with modern Physics*. Thomson Brooks / Cole, California, USA, 6th edition, 2004.

[171] K Silversides, Melkumyan, A., D. Wyman, and P.J. Hatherly. Identification of shale and ore boundaries using gaussian processes. In *Iron Ore 2011 Proceedings*, pages 179–183, Perth, Australia, 2011.

[172] R B Singer. Near-infrared spectral reflectance of mineral mixtures: Systematic combinations of pyroxenes, olivine, and iron oxides. *Journal of Geophysical Research: Solid Earth*, 86(B9):7967–7982, September 1981. ISSN 2156-2202.

[173] P N Slater. Comparison of Photographic and Digital Imagery from Film and Solid-State-Array Remote Sensing Cameras. *Photographic Science and Engineering*, 21:198, 1977.

[174] K L Smith, M D Steven, and J J Colls. Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks. *Remote Sensing of Environment*, 92(2):207–217, August 2004. ISSN 0034-4257.

[175] A J Smola, B Schölkopf, and K-R Müller. The connection between regularization operators and support vector kernels. *Neural Networks*, 11(4): 637–649, June 1998. ISSN 0893-6080.

[176] M Sneep and W Ubachs. Direct measurement of the Rayleigh scattering cross section in various gases. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 92(3):293–310, May 2005. ISSN 0022-4073.

[177] J A. Snyman. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Springer Publishing, 2005.

[178] D. Steiner and A.E Salerno. Remote sensor data systems, processing and management. In R. G. In Reeves, editor, *Manual of remote sensing*, pages 611–803. American Society of Photogrammetry, Falls Church, Virginia, 1975.

[179] J Strutt. On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. *Philosophical Magazine*, 47(5):375–394, 1899.

[180] J M Sunshine, C M Pieters, and S F Pratt. Deconvolution of Mineral Absorption Bands: An Improved Approach. *Journal of Geographical Research*, 95, 1990.

[181] G A Swayze, R N Clark, A F H Goetz, T G Chrien, and N S Gorelick. Effects of spectrometer band pass, sampling, and signal-to-noise ratio on spectral identification using the Tetracorder algorithm. *Journal of Geographical Research*, 108(E9):5105, 2003.

[182] G Talsky, L Mayring, and H Kreuzer. High-Resolution, Higher-Order UV/VIS Derivative Spectrophotometry. *Angewandte Chemie International Edition in English*, 17(11):785–799, 1978.

[183] W. Thorne, S.G. Hagemann, A. Webb, and J. Clout. BIF-related deposits of the Hamersley province, Western Australia. *in Hagemann, S.G., Rosiere, C. , Gutzmer, J., and Beukes, N., (eds). Reviews In Economic Geology. High-grade BIF-related iron mineralization*, 15:197–222, 2008.

[184] M.E. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

[185] A S J Tjiong and S T Monteiro. Feature selection with PSO and kernel methods for hyperspectral classification. In *IEEE Congress on Evolutionary Computation*, pages 1762–1769, New Orleans, LA, USA, 2011.

[186] J S. Townsend. *A modern approach to quantum mechanics*. University Science Books, Mill Valley, Calif., 2nd edition, 2012.

[187] T E Townsend. Discrimination of iron alteration minerals in visible and near-infrared reflectance data. *Journal of Geophysical Research: Solid Earth*, 92(B2):1441–1454, February 1987. ISSN 2156-2202.

[188] J R G Townshend. The spatial resolving power of earth resources satellites. *Progress in Physical Geography*, 5(1):32–55, 1980.

[189] F Tsai and W Philpot. Derivative Analysis of Hyperspectral Data. *Remote Sensing Of Environment*, 66(1):41, 1998.

[190] University of Sydney. Rio Tinto Centre For Mine Automation Website. www.usyd.edu.au/cma. University of Sydney, www.usyd.edu.au/cma, July 2009.

[191] F van de Meer and W Bakker. Cross correlogram spectral matching: Application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada. *Remote Sensing of Environment*, 61(3):371–382, September 1997. ISSN 0034-4257.

[192] F Van de Meer and W Bakker. Validated surface mineralogy from high-spectral resolution remote sensing: a review and a novel approach applied to gold exploration using AVIRIS data. *Terra Nova*, 10(2):112–119, March 1998. ISSN 0954-4879.

[193] F D Van de Meer and S M De Jong. *Imaging Spectroscopy: Basic Principles and Prospective Applications*, volume 4. KLUWER ACADEMIC PUBLISHERS, DORDRECHT, BOSTON, LONDON,, 2001. ISBN 1-4020-0194-0.

[194] F van der Meer. Analysis of spectral absorption features in hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 5(1):55, 2004.

[195] F van der Meer. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 8(1):3–17, 2006.

[196] C J van Rijsbergen. *Information retrieval*. Butterworths, London, 1979.

[197] G Vane and A F H Goetz. Terrestrial imaging spectroscopy. *Remote Sensing Of Environment*, 24(1):1, 1988.

[198] G Vane and A F H Goetz. Terrestrial imaging spectrometry: Current status, future trends. *Remote Sensing Of Environment*, 44(2-3):117, 1993.

[199] V N Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[200] S Vasudevan. Data fusion with Gaussian processes. *Robotics and Autonomous Systems*, 60(12):1528–1544, December 2012. ISSN 0921-8890.

[201] C Verpoorter. *Télédétection hyperspectrale et cartographie des faciès sédimentaires en zone intertidale : application à la Baie de Bourgneuf*. PhD thesis, Universite de Nantes, 2009.

[202] R K Vincent. Oil, gas exploration tool- Composite mapping of earth from satellite information. *The Oil and Gas Journal*, 17:141–142, 1975.

[203] S Wang. Support Vector Machines Classification for High-Dimentional Dataset. In *Multimedia Information Networking and Security (MINES), 2012 Fourth International Conference on*, pages 315–318, 2012.

[204] Xiangyang Wang, Jie Yang, Xiaolong Teng, Weijun Xia, and Richard Jensen. Feature selection based on rough sets and particle swarm optimization. *Pattern Recognition Letters*, 28(4):459–471, 2007.

[205] B Waske and J A Benediktsson. Fusion of Support Vector Machines for Classification of Multisensor Data. 2007.

[206] R. Welch. Progress in the specification and analysis of image quality. *Photogrammetric Engineering and Remote Sensing*, 43:709–719, 1977.

[207] W W Wendlandt and H G. Hecht. *Reflectance spectroscopy*. Interscience Publishers, New York, 21 edition, 1966.

[208] C K I Williams and C E Rasmussen. Gaussian processes for regression. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 514–520. The MIT Press, Cambridge, MA, USA, 1996.

[209] C.K.I. Williams. Computation with infinite neural networks. *Neural Computation*, 10(5):1203–1216, 1998.

[210] C Xu, M C Maddage, X Shao, F Cao, and Q Tian. Musical genre classification using support vector machines. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, volume 5, pages V–429–32 vol.5, 2003.

[211] M Yellishetty, P G Ranjith, and A Tharumarajah. Iron ore and steel production trends and material flows in the world: Is this really sustainable? *Resources, Conservation and Recycling*, 54(12):1084–1094, 2010. ISSN 0921-3449.

[212] P J Zarco-Tejada, J C Pushnik, S Dobrowski, and S L Ustin. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment*, 84(2):283–294, February 2003. ISSN 0034-4257.

[213] L Zhao and N Takagi. An application of support vector machines to chinese character classification problem. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 3604–3608, 2007.

[214] H Zhu and R Rohwer. No Free Lunch for Cross-Validation. *Neural Computation*, 8(7):1421–1426, October 1996. ISSN 0899-7667.