Sampling scheme optimization from hyperspectral data

Pravesh Debba

Promotors:

Prof. Dr. Ir. A. Stein
Professor of Mathematical and Statistical methods for geodata
International Institute for Geo-Information Science and Earth Observation (ITC), and
Wageningen University, The Netherlands

Prof. Dr. F. D. van der Meer
Professor of earth subsurface systems analysis International Institute for Geo-Information Science and Earth Observation (ITC), and
Professor of Geological remote sensing
University of Utrecht, The Netherlands

Co-Promotor:

Dr. E. J. M. Carranza
Assistant Professor at the International Institute for Geo-Information Science and Earth Observation (ITC)

Examining Committee:

Prof. Dr. N. Veraverbeke (University of Haselt (UHasselt), Belgium)
Prof. Dr. A. I. Dale (University of KwaZulu-Natal (UKZN), South Africa)
Prof. Dr. J. Grasman (Wageningen University (WU), The Netherlands)
Dr. Ir. G. B. M. Heuvelink (Wageningen University (WU), The Netherlands)

Sampling scheme optimization from hyperspectral data

Pravesh Debba

THESIS

To fulfil the requirements for the degree of Doctor
on the authority of the Rector Magnificus of Wageningen University
Prof. Dr. M. J. Kropff
to be publicly defended on Tuesday 20 June 2006 at 15:00 hrs
in the auditorium at ITC, Enschede, The Netherlands

Sampling scheme optimization from hyperspectral data

ISBN: 90-8504-462-6
International Institute for Geo-information Science and Earth Observation,
Enschede, The Netherlands (ITC)
ITC Dissertation number 136
ITC, P.O. Box 6, 7500 AA Enschede, The Netherlands

To my mother
&
To Nalinee, Harshil and Ayojana

Abstract

This thesis presents statistical sampling scheme optimization for geo-environmental purposes on the basis of hyperspectral data. It integrates derived products of the hyperspectral remote sensing data into individual sampling schemes. Five different issues are being dealt with.

First, the optimized sampling scheme is presented to select samples that represent different ontological categories. The iterated conditional modes algorithm (ICM) is used as an unsupervised segmentation technique. Within each category, simulated annealing is applied for minimizing the mean shortest distance (MMSD) between sampling points. The number of sampling points in each category is proportional to the size and variability of the category. The combination of the ICM algorithm for image segmentation with simulated annealing for optimized sampling, results in an elegant and powerful tool in designing optimal sampling schemes using remote sensing images. A validation study conducted shows that the optimized sampling scheme gives best estimates for commonly used vegetation indices compared to simple random sampling and rectangular grid sampling.

Next, optimal sampling schemes, which focus on ground verification of minerals derived from hyperspectral data, are presented. Spectral angle mapper (SAM) and spectral feature fitting (SFF) classification techniques are applied to obtain rule mineral images. The rule images provide weights that are utilized in objective functions of the sampling schemes which are optimized by means of simulated annealing. Three weight functions intensively sample areas where a high probability and abundance of alunite occurs. Weight function I uses binary weights derived from the SAM classification image, leading to an even distribution of sampling points over the region of interest. Weight function II uses scaled weights derived from the SAM rule image. Sample points are arranged more intensely in areas where there is an abundance of alunite. Weight function III combines information from several different rule image classifications. Sampling points are distributed more intensely in regions of high probable alunite as classified by both SAM and SFF, thus representing the purest of pixels. This method leads to an efficient distribution of sample points, on the basis of a user-defined objective.

This is followed by a quantitative method for optimally locating exploration targets based on a probabilistic mineral prospectivity map, which was created by means of weights-of-evidence (WofE) modeling. Locations of discovered mineral occurrences were used as a training set and a map of distances to faults/fractures

and three channel ratio images of HyMap hyperspectral data were used as evidences. The WofE posterior probability map was applied to an objective function that optimized location of exploration targets. Optimized exploration target zones spatially coincide with undiscovered mineral occurrences, namely, those not used to train the WofE model input, and show other zones without mineral occurrences within delineated prospective ground. The results indicate usefulness of the described optimization method to allocate exploration targets for undiscovered mineral occurrence, based on probabilistic mineral prospectivity maps.

A method for estimating the partial abundance of spectrally similar minerals in complex mixtures follows. Linear mixtures are generated with varying proportions of individual spectrum, from a spectral library, of a set of iron-bearing oxide/hydroxide/sulfate minerals. The first and second derivatives of each of the different sets of mixed spectra and the individual spectrum are evaluated. This method for spectral unmixing requires formulating a linear function of individual spectra of the minerals. The error between these derivative functions and the respective derivative function of the mixed spectrum is minimized by means of simulated annealing. Experiments are made on several different mixtures of selected end-members, which could plausibly occur in real situations. The variance of the differences between the second derivatives of the observed spectrum and the second derivatives of the end-member spectra give most precise estimates for the abundance of each end-member.

Lastly, a method by which an optimal ground sampling scheme can be obtained for a variable of interest is described. The variable of interest is the spatial distribution of a suite of heavy metals in mine tailings. Derivation of an optimal sampling scheme makes use of covariates of the spatial variable of interest, which are readily but less accurately obtainable by using airborne hyperspectral data. The covariates are abundances of secondary iron-bearing minerals estimated through spectral unmixing. Via simulated annealing, an optimal retrospective sampling scheme for a previously sampled area is derived having fewer samples but having almost equal mean kriging prediction error as the original ground samples. Via simulated annealing, an optimal prospective sampling scheme for a new unvisited area is derived based on the variogram model of a previously sampled area. The results of this study demonstrate potential application of hyperspectral remote sensing and simulated annealing to surface characterization of large mine tailings having similar climatic and terrain characteristics to the mine tailings in the case study area.

Samenvatting

Dit proefschrift richt zich op het optimaliseren van ruimtelijke bemonsteringsschema's met een accent op het gebruik van hyperspectraal beelden binnen geo-milieu toepassingen. Het integreert deze beelden, die met satellieten of vliegtuigen zijn verkregen, binnen bemonsteringsschema's. Vijf verschillende aspecten worden hierbij belicht.

In de eerste plaats wordt het optimale bemonsteringsschema gevonden om waarnemingen te plaatsen die ontologische klassen representeren. Voor dat doel wordt het ICM (Iteratieve Conditionele Modus) algoritme gebruikt als een ongecontroleerd classificatie algoritme. Optimalisatie met simulated annealing wordt toegepast binnen iedere categorie voor het vinden van de gemiddeld kortste afstand tussen bemonsteringspunten. Het aantal punten per categorie is recht evenredig met de grootte van een klasse en met de variatie daarbinnen. De combinatie van het ICM algoritme voor beeldsegmentatie en simulated annealing voor optimalisatie resulteert in een elegante en krachtige methode voor het vinden van het optimale bemonsteringsschema bij gebruikmaking van hyperspectraalbeelden. Een validatiestudie toont aan dat we hiermee betere schattingen krijgen voor veelgebruikte vegetatie indices dan met toevalsbemonstering of roosterbemonstering.

Vervolgens worden optimale bemonsteringsschema's gepresenteerd die zich richten op het verifiëren van de aanwezigheid van mineralen, die met hyperspectraal beelden worden gedentificeerd. Spectrale hoek karteringsclassificatie (SAM) en spectrale object schattingsclassificatie (SFF) worden gebruikt om mineraal referentiebeelden te krijgen. Deze mineraal referentiebeelden leveren de gewichten die gebruikt worden bij het optimaliseren van bemonsteringsschema's met behulp van simulated annealing. Drie gewichtsfuncties worden gebruikt om gebieden intensief te bemonsteren die een grote kans hebben op het aantreffen van aluniet. De eerste gewichtsfunctie gebruikt de gewichten die afkomstig zijn uit de SAM classificatie, hetgeen leidt tot een regelmatige verdeling van de bemonsteringspunten van het studiegebied. De tweede gewichtsfunctie gebruikt geschaalde gewichten afkomstig uit de mineraal referentiebeelden die met SAM zijn gemaakt. Dit levert een bemonsteringsschema dat intensiever wordt naarmate aluniet meer voorkomt. De derde gewichtsfunctie combineert de informatie van mineraal referentiebeelden die met verschillende classificatie methoden zijn vervaardigd. Dit levert een grotere waarnemingsdichtheid op in gebieden waar de kans groot is dat aluniet voorkomt, volgens zowel de SAM als de SFF classificatie. Als zodanig representeert het de meest zuivere pixels. De methode leidt tot een efficiënte verdeling van de waarnemingspunten op basis van een doel dat

door de gebruiker gedefinieerd is.

Het volgende aspect is een kwantitatieve methode voor het optimaal definiëren van een exploratiedoel. Deze methode is gebaseerd op een probabilistische verwachtingskaart voor mineralen. De kaart is vervaardigd op basis van modelleren via de weights-of-evidence (WofE) methode. Plaatsen waar de aanwezigheid van een mineraal bekend is zijn gebruikt als trainingsgegevens, terwijl het bewijs bestond uit een kaart van breuken en verschuivingen en drie beelden van verhoudingen tussen hyperspectraal banden. De *a posteriori* kaart van deze WofE modellering is gebruikt om een doelfunctie te optimaliseren voor het identificeren van exploratie doelen. De optimale exploratie zones vallen ruimtelijk samen met onontdekte mineraalverschijningen en met mineraalverschijningen die niet gebruikt zijn om het WofE model te trainen; ze laten ook zones zien zonder mineralen binnen gebieden met een vermoedelijke opbrengst. De resultaten tonen het nut van de optimalisatiemethode aan om exploratiedoelen te stellen voor onontdekte mineralen, gebaseerd op een probabilistisch mineraal prospectkaarten.

Vervolgens volgt een methode voor het schatten van het fractionele gehalte van mineralen met een identieke spectrale signatuur in complexe mengingen. Lineaire mengingen zijn gesimuleerd van verschillende verhoudingen van individuele spectra uit een spectrale bibliotheek van oxide/hydroxide/sulfaat mineralen. De eerste en de tweede afgeleide van ieder van de verschillende verzamelingen van de gemengde spectra en het individuele spectrum zijn daarbij geanalyseerd. Deze methode voor spectrale ontmenging vereist een formulering van een lineaire functie van de individuele spectra van de mineralen. De afwijking tussen de afgeleiden en de respectievelijke afgeleide van het gemengde spectrum is geminimaliseerd met behulp van simulated annealing. Simulatiestudies zijn gedaan met verschillende mengingen van geselecteerde eindleden die in de praktijk zouden kunnen voorkomen. De variantie van de verschillen tussen de 2^e afgeleide van het waargenomen spectrum en de 2^e afgeleide van de spectra van de eindleden leverde de meest precieze schatting voor het gehalte van ieder eindlid.

Als laatste wordt een methode beschreven waarmee het optimale bemonsteringsschema voor een doelvariabele op de grond kan worden gevonden. In dit onderdeel van de studie is de doelvariabele een verzameling zware metalen in een mijnstort. Het bepalen van het optimale schema maakt gebruik van co-variabelen van de doelvariabele. Deze kunnen eenvoudig maar minder precies worden verkregen uit hyperspectraal gegevens die met vliegtuigen zijn opgenomen. De co-variabelen zijn de gehaltes van secundaire ijzerhoudende mineralen, verkregen via een ontmengingprocedure. Door middel van simulated annealing is het optimale retrospectieve bemonsteringsschema bepaald voor een gebied dat al eerder was bemonsterd. Dit schema bevat minder waarnemingen, maar levert vrijwel een gelijke variantie van de voorspelfout als het oorspronkelijke schema. Door middel van simulated annealing is vervolgens het optimale bemonsteringsschema voor een nieuw, onbezocht gebied bepaald, onder aanname van geldigheid van het variogram uit het eerder bemonsterde gebied. De resultaten van deze studie tonen het potentiële gebruik aan van een combinatie van hyperspectrale remote sensing beelden en simulated annealing voor het karakteriseren van de bovengrond van een mijnstortgebied. Dit onder voorwaarde van vergelijkbare klimatologische en terrein karakteristieken van de verschillende mijnstortgebieden.

Acknowledgements

The many people who have helped in some way or the other to complete this thesis cannot all be mentioned. I take this opportunity to thank some of them.

I express my sincere gratitude to my two promoters and supervisors Prof. dr. Alfred Stein and Prof. dr Freek van der Meer for their joint contribution, support, invaluable guidance, great insight and encouragement throughout this work. As a statistician, Alfred provided great help with his invaluable knowledge in driving the thesis to an end. As a geologist, Freek provided me with information that I seriously lacked and felt as though I am embarking on an unknown terrain. Freek was also very understanding and I constantly bugged him whenever I needed a moral boost. When I first arrived at ITC, Alfred made arrangements so that my stay in the Netherlands was comfortable and provided me with the infrastructure to start research immediately. Freek arranged a quick tour of the study area in Spain, provided me with geological details of the area, an overview of hyperspectral imaging, data collection (both field and airborne) and funding for the field trip prior to my own research funding being accepted. To Alfred and Freek, thank you very much! Alfred, thanks for translating the summary into Dutch. I extend my sincere thanks to Prof. Noel Veraverbeke for referring me to Alfred for this project and the role he has played during my MSc and PhD period, which included major advancements in my career. Also, I would like to thank ITC for granting me the fully funded scholarship that made this thesis possible and to the NRF for the additional funding they provided over the years.

I thank the support and enthusiasm from my assistant promoter John Carranza. John became involved at a later stage of the project and with his help, the work progressed with incredible speed. John always found time in his busy schedule to have lengthy but constructive discussions and helped steer many of the papers to international peer reviewed journals. I also had the good fortune to work with Frank van Ruitenbeek as we had numerous interesting discussions on geology and statistics and together we successfully made research publications. Thanks to Arko Lucieer who introduced me to programming in Java and also contributed to an article. There were many others that helped me with programming in Java and RGui including Roger Bivand, Jan van der Kassteele, Peter Diggle, Søren Lophaven, Edzer Pebesma, Boen Liem and Willem Kruijer. Also thanks to Rolf de By for the LaTeXcourse he presented and to Javier Morales for the LaTeXclass used to type this thesis. It made typing so much easier.

There have been several people and teams involved in collecting field data that were used for this thesis. First, I would like to thank the EREG GRMEG and MRE students at ITC from 2002–2003 and Frank van Ruitenbeek (ESA Department, ITC) for field spectral data

collection in the Los Tollos area. Second, I acknowledge the efforts taken by Gift Rukezo (ITC MRE MSc alumni of 2003) and E. J. M. (John) Carranza (ESA Department, ITC) for collection of the mine tailings samples in the Lahóca-Recsk area, with field technical information and logistics assistance from Tibor Zelenka (Hungarian Geological Survey), Aniko Somody (Recsk Ore Company), and Zoltan Vekerdy (WRS Department, ITC). Gift Rukenzo and Boudewijn de Smeth (ESA Department, ITC) were also involved for laboratory determination of heavy metal contents of the mine tailings samples from the Lahóca-Recsk area. Third, I would again like to thank Zoltan Verkerdy (WRS Department, ITC) and János Tamás (University of Debrecen) for providing field information and the use of the hyperspectral data in the Tedej area, Hungary.

There are many PhD and MSc friends that I would like to thank during my stay in Enschede. Among them, my office-mates: Uday Bhaskar Nidumolu, Grace Nangendo, Javier Martinez, Cheng Jianquan, Chudamani Joshi, Istiak Sobhan, Richard Onchanga and Tiejun Wang. Aside from my office mates several other fellow PhD students to whom I am most grateful for their support, encouragement and discussions in and outside of ITC include: Prasun Gangopadhyay, Ulanbek Turdukulov, Gabrielle Iglesias, Sophia Woodcock, Marleen Noomen, Bobba Bharath, Peter Minang, Peter Beck, Jane Bemigisha, Ahmad Mobin-ud Din and Arta Dilo. Other friends that have made my stay here in the Netherlands more comfortable and deserve special mention include: Supawit Yawsangratt (Pae-pe), Marleen, Fabian, Celeste, Rodrigo Sicat, Si Yali, Ashwini Panday, Swarn Pant, Parth Shah, Umamaheshwaran Rajasekar, Veerendra Verma, Lemmy Namayanga, Barbara Casentini, ... I have also socialized with many Filophinos through two close friends Anna Dagohoy and Jonald Pimentel who are PhD students at University of Twente and also former MSc classmates at Limbergs University Centrum in Belgium.

I cannot forget the warmth extended by staff at ITC especially Martin Hale, Loes Colenbrander, Mireille Meester, Christie Agema, Ard Kosters, Carla Gerritsen, Petra Weber, Marga Koelen, Marie Chantal Metz, Theresa van den Boogaard, Bettine Geerdink, Marion Pierik. They were also responsible for making the frequent trips to and from Enschede and South Africa a relatively painless effort. I thank Marjolein Woerlee and Saskia Groenendijk for always providing me with very nice accommodation after returning from South Africa and their friendly warm faces at the reception on my way to the pool tables.

This project was in agreement with the University of KwaZulu-Natal in granting me leave certain periods of the year. At times it became really difficult to return to the university to resume my post while in the midst of my research but I somehow survived. I thank Prof. Sunil Maharaj as head of school at the time and Prof. Ahmed Bawa for granting me permission to pursue my PhD. My leave application and the decision to pursue the PhD in the Netherlands was also supported by Prof. Andrew Dale to whom I am very grateful. I also thank the entire staff in the School of Statistics and Actuarial Sciences for having to share my workload during my absence. Dr. Delia North has also inspired me and has tried effortlessly to grant me permission to take leave at crucial moments.

To my mum, whom I also dedicate this thesis. Without your support and the support I received from my family over the years towards my education, all of this would not have been possible. I have gained great strength from all of you.

I would finally like to thank Nalinee for your love, patience and sacrifice over the last few years. Thanks for taking care of young Harshil and Ayojana in my absence. Without your help none of this was possible. I have seen great strength in you, committing fully to the

children, while progressing with your own academic career. Even after I returned to South Africa, you still willing accepted the long hours I spent at the university in wrapping up this thesis. The smiling faces on my two children always gave me hope and strength to continue. Thanks. I also extend my thanks to my in-laws for the assistance they provided to my family while I was away.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Faith is taking the first step even when you don't see the whole staircase.

Martin Luther King, Jr

1.1 Land characterization: problems in deriving optimal sampling schemes

Land has many components. The various components, such as vegetation, and in the absence of vegetation the rocks and sands with all their minerals make up land cover. To adequately characterize the vegetation components or the mineral components of land, detailed maps describing the spatial distributions of, for example, certain crops or certain minerals are required. The spatial distributions of crops or minerals, however, vary from one place to another according to factors at local settings. Therefore, thorough sampling of land is required to generate detailed maps accurately depicting spatial variability of either crops or minerals and associated metals. Such an undertaking would require money, time, and manpower in order to achieve spatial information of interest at the desired level of accuracy. Therefore, planning *where* and *how many* samples should be collected, in order to map accurately the spatial distributions of either crops or minerals and associated metals, is a non-trivial task.

A sampling plan or scheme refers to positions of samples on the ground. There are two types of sampling schemes, (a) a retrospective scheme, whereby sample locations are either removed from or added to an existing sampling scheme, and (b) a prospective scheme, whereby sample locations are pre-determined before actual sampling in the field. A sampling scheme design is considered optimal if there is (i) a reduction in the number of samples but resulting in estimates of population parameters of interest with the same or similar uncertainty, (ii) a reduction in the variability or mean squared error in estimates of population parameters of interest, (iii) a more correct distribution of samples representing the distribution of the population of interest, or a combination of these criteria. Development of optimal sampling requires *a priori*

spatial information about a study area.

Around the mid-20th century and a few decades thereafter, those who studied crops (Johnson, 1969; Driscoll & Coleman, 1974; Everitt et al., 1980) and those who searched for minerals (Eardley, 1942; Laylender, 1956; Allum, 1966; Longshaw & Gilbertson, 1976; Gilbertson et al., 1976) developed their sampling schemes by using geographical information from topographic maps and/or stereoscopic aerial photographs and from visual observations during field reconnaissance surveys. From the 1970s, technological developments in remote sensing resulted in the collection of spaceborne multispectral data, which were to a larger extent useful to derive *a priori* spatial information required in sampling campaigns to study agricultural crops (Everitt et al., 1979; McGraw & Tueller, 1983) but were to a lesser extent useful to derive *a priori* spatial information required in searching for minerals (Houston, 1973; Siegal & Abrams, 1976; Lowman, 1976; Iranpanah, 1977; Siegal & Gillespie, 1980). The reasons for the relative contrast of usefulness spaceborne multispectral data to crop vegetation studies and to search for minerals are that multispectral sensors collect broad wavelength data (a) mostly in the visible to near infrared range of the electromagnetic spectrum, where vegetation has diagnostic spectral features, but (b) partly in the shortwave infrared range of the electromagnetic spectrum, where most minerals have diagnostic spectral features. Multispectral data allow mapping of individual crop species quite accurately (Richardson et al., 1985; Brisco et al., 1989; Bouman & Uenk, 1992), but allow mapping of groups and not individual minerals such as in hydrothermally altered rocks (Rowan et al., 1977; Kowalik et al., 1983; Abrams, 1984; Carranza & Hale, 2002).

From the 1990s, however, advanced technological developments in remote sensing resulted in acquiring airborne hyperspectral data, which are better sources of *a priori* information for those who optimize their respective sampling schemes to study crop vegetation or search for minerals and associated metals. The advantage of hyperspectral data over multispectral data can be attributed to their high spatial resolution and much higher spectral resolutions in the visible to the shortwave infrared regions (Clark, 1999; Polder & van der Heijden, 2001), which allow distinction between plant species (Thenkabail, 2002; Thenkabail P. S. & De-Pauw, 2002; Okina et al., 2001; Chang, 2006) or minerals and associated metals (Cudahy et al., 2000; Papp & Cudahy, 2002; Martini et al., 2003; Martini, 2003; Martini et al., 2003). Nevertheless, the ability to process and analyze multi-dimensional hyperspectral data promptly requires improved or novel techniques in order to extract and then further process vital information to derive optimal sampling schemes.

The availability of airborne hyperspectral data, therefore, raises two problems in deriving optimal sampling schemes to study crops and to search for minerals and associated metals: (1) how to extract accurate *a priori* information of interest; and (2) how to further process *a priori* information of interest to derive an optimal sampling scheme. The first problem is related to the fact that processing and analysis of hyperspectral data results in only estimates of certain parameters such as (a) vegetation indices, which could reflect crop health (Knipling, 1970; Ausmus & Hilty, 1972; Carter, 1994), and (b) mineral indices, which are estimates of relative abundance of

minerals (Smith et al., 1985; Resmini et al., 1997; Crósta et al., 1998; Chabrillat et al., 1999). Accurate estimation of these parameters is undermined by several factors that, for example, distort spectral signal from materials of interest on the ground to the hyperspectral sensor in the air (Richards, 1993; Lillesand et al., 1994; Sabins, 1996; Gupta, 2003). The second problem is related to the statistical correlation or spatial association between parameters estimated from hyperspectral data and the primary variables of interest, which in this thesis are crops or minerals and associated metals. To investigate potential solutions to these two problems in deriving optimal sampling schemes given hyperspectral data, it is important to first understand hyperspectral remote sensing and optimization of schemes separately and to then merge the disparate knowledge gained. The following two sections provide brief literature reviews on hyperspectral remote sensing and optimization of sampling schemes, respectively.

1.2 Hyperspectral remote sensing

In the study of electro-magnetic physics, when energy in the form of light interacts with a material, part of the energy at certain wavelength is absorbed, transmitted, emitted, scattered, or reflected due to the property or characteristics of the material (Sabins, 1996). The three most common ways of measuring the reflectance of a material are by (a) using a hand-held spectrometer over the material in the field or laboratory, (b) using a sensor mounted on an aircraft over a land terrain, or (c) using a sensor mounted on a spacecraft over the earth's surface.

Available hyperspectral data are mostly obtained by aircrafts. Hyperspectral data are reflectance measurements at very narrow wavelengths, approximately 10 nm or less, and are acquired simultaneously over a large spectral range, usually between 0.4 μm and 2.5 μm (Chang, 2006). This spectral range includes the visible, near infrared and short wave infrared regions of the electro-magnetic spectrum, resulting in a large number (often > 100) of contiguous spectral bands or channels. Reflectance data in each spectral channel can be pictorially represented as an image, which is composed of discrete picture elements or pixels. The brightness of a pixel represents the reflective value of materials at specific wavelengths of the electro-magnetic spectrum. Every material has unique spectral features (Hapke, 1993), which are distinct arrays of spectral values at certain regions of the electro-magnetic spectrum. Because hyperspectral sensors acquire spectral data from narrow and contiguous bands of the electro-magnetic spectrum, they provide much better capability to identify materials than broad-band sensors (Sabins, 1999). For example, analysis of changes in narrow absorption features (Van der Meer, 2004), which are usually not recorded by broad-band sensors, is a powerful tool in remote identification and estimation of individual materials instead of groups of materials.

A vast amount of scientific knowledge has been and is currently being developed in the field of hyperspectral remote sensing of the environment (Sabins, 1996; Gupta, 2003; Chang, 2006). There are several international peer reviewed journals specif-

ically publishing innovative procedures and advancements on hyperspectral remote sensing of the environment. Integration of hyperspectral data or information derived from hyperspectral data into optimization of sampling schemes has been relatively neglected (Stein et al., 1999).

1.3 Optimization of sampling schemes

Spatial sampling has been addressed by statisticians for many years. In comparing traditional sampling schemes Burgess et al. (1981) found that a regular grid results in only slightly less precise estimates than a triangular grid, for the same sampling density. They concluded that a small loss of precision or small increase in sampling density to achieve a given precision corresponds with a small increase in price to pay for the practical convenience of regular grids. Christakos & Olea (1992) present a case-specific methodology for choosing between different grid designs.

In optimization of model-based sampling schemes, Spruill & Candela (1990) considered the prediction accuracy of chloride concentration in groundwater by removing or adding locations to an existing sampling network. In a similar way, Royle & Nychka (1998) used a geometrical criterion in order to optimize spatial prediction. Brus & de Gruijter (1997) compared design-based and model-based sampling schemes.

With applications of geostatistical methods, it has been previously shown that for spatially correlated data a triangular configuration of sampling points is most efficient and for isotropic variations the grid should be equilateral (Burgess et al., 1981). McBratney et al. (1981) and McBratney & Webster (1981) presented procedures for optimizing the spacing grid of a regular rectangular or triangular lattice design by maximizing the prediction variance, given an *a priori* variogram. If a variogram, however, shows a relatively high nugget and sampling density is relatively scarce, then a hexagonal grid can be most efficient (Yfantis et al., 1987). By removing or adding locations to an existing sampling network, Ben-Jemaa et al. (1995) used ordinary cokriging between sediment concentration of mercury and a sediment grain size index to maximize the prediction accuracy. Lloyd & Atkinson (1999) used ordinary kriging and ordinary indicator kriging to optimize a sampling scheme. Diggle & Lophaven (2006) use a Bayesian criterion to optimize geo-spatial prediction by (a) deleting locations from an existing sampling design and (b) choosing positions for a new set of sampling locations. Other studies of variogram application to optimize sampling schemes include Russo (1984), Warrick & Myers (1987), Zimmerman & Homer (1991) and Müller & Zimmerman (1999).

With applications of simulated annealing, Sacks & Schiller (1988) presented several algorithms for optimizing a sampling scheme out of a small grid of possible locations. McGwire et al. (1993) investigated the impact of sampling strategies on the stability of linear calibrations by enforcing various sample distance constraints in a Monte Carlo approach. Van Groenigen & Stein (1998) extended this design by presenting the optimal sampling scheme using spatial simulated annealing that could handle earlier data

points and complex barriers. Van Groenigen & Stein (1998) also developed further the Warrick & Myers (1987) criterion to optimize sampling schemes. Van Groenigen et al. (1999) used spatial simulated annealing to construct sampling schemes with minimal kriging variance. They found that anisotropy of the variogram had considerable influence on the optimized sampling scheme, with the highest sampling density in the direction of the highest variability. Van Groenigen et al. (1999) used spatial simulated annealing and the criterion for minimizing the maximum kriging variance in obtaining the optimal sampling scheme. Van Groenigen et al. (2000b) showed how conditional probabilities of exceeding environmental threshold values of several contaminants could be pooled into one variable, indicating health risk and thereby used simulated annealing to optimize the sampling scheme. Van Groenigen et al. (2000a) used yield maps to optimize, via spatial simulated annealing, soil sampling for precision agriculture in a low-tech environment. Lark (2002) maximized the likelihood estimation for the Gaussian linear model, which results in designs consisting of fairly regular array supplemented by groups of closely spaced locations.

In sampling for field spectral measurements to support remote sensing, Curran & Atkinson (1998) used co-kriging to define the optimal 'multiple' sampling design, which could be used to simultaneously sample ground and remote sensing data. Tapia et al. (2005) applied a multivariate k-means classifier to delineate vegetation patterns from remote sensing data together with the Van Groenigen & Stein (1998) criterion in order to prioritize the survey to areas with high uncertainty. In this present thesis, sampling schemes are optimized based on remote sensing data or remotely sensed information and the application of simulated annealing.

1.4 Simulated Annealing in context of sampling scheme optimization

Simulated annealing is a general optimization method that has been widely applied to find the global optimum of an objective function when several local optima exist. Details on simulated annealing can be found in Kirkpatrick et al. (1983), Bohachevsky et al. (1986) and Aarts & Korst (1989).

In application of simulated annealing to sampling scheme optimization, a fitness function $\phi(\mathbf{S})$ has to be minimized, depending on the sampling configuration $\mathbf{S}$. Starting with a random sampling scheme $\mathbf{S}_0$, let $\mathbf{S}_i$ and $\mathbf{S}_{i+1}$ represent two solutions with fitness $\phi(\mathbf{S}_i)$ and $\phi(\mathbf{S}_{i+1})$, respectively. Sampling scheme $\mathbf{S}_{i+1}$ is derived from $\mathbf{S}_i$ by randomly replacing one of the points of $\mathbf{S}_i$ by a new point not in $\mathbf{S}_i$. A probabilistic acceptance criterion decides whether $\mathbf{S}_{i+1}$ is accepted or not. This probability $P_c(\mathbf{S}_i \rightarrow \mathbf{S}_{i+1})$ of $\mathbf{S}_{i+1}$ being accepted can be described as:

$$P_{\mathbf{c}}(\mathbf{S}_i \rightarrow \mathbf{S}_{i+1}) = \begin{cases} 1\,, & \text{if } \phi(\mathbf{S}_{i+1}) \leq \phi(\mathbf{S}_i) \\ \exp\left(\dfrac{\phi(\mathbf{S}_i) - \phi(\mathbf{S}_{i+1})}{\mathbf{c}}\right)\,, & \text{if } \phi(\mathbf{S}_{i+1}) > \phi(\mathbf{S}_i) \end{cases} \tag{1.1}$$

where **c** denotes a positive control parameter (usually called the temperature in simulated annealing problems). Several cooling schedules are possible to reduce the temperature. At each value of **c**, several transitions have to be made before the annealing can proceed, and **c** can take its next value. A transition takes place if $\mathbf{S}_{i+1}$ is accepted. Next, a solution $\mathbf{S}_{i+2}$ is derived from $\mathbf{S}_{i+1}$, and the probability $P_c(\mathbf{S}_{i+1} \rightarrow \mathbf{S}_{i+2})$ is calculated according to an acceptance criterion (Equation 1.1).

1.5 Purposes of this study

In this thesis, estimates of parameters of interest derived from hyperspectral data or statistical correlation between parameters estimated from hyperspectral data and the primary variables of interest are here referred to as a model. It is hypothesized that model-based optimal sampling schemes can be derived by (a) improving the precision of a model, (b) improving the accuracy of a model, (c) improving the estimates of a model, (d) reducing the variability of a model, (e) reducing the error of model; or by a combination of any of these aspects. Accordingly, to investigate the hypothesis, the main purpose of this thesis is to use airborne hyperspectral data to obtain models for input into simulated annealing in order to derive optimal sampling schemes. The main purpose of this thesis is supported by the following specific aims.

(i) To formulate optimization criteria that meet the specific demands for ground sampling to study spatial variability of crops or minerals and associated metals.

(ii) To formulate new hyperspectral remote sensing techniques useful for deriving optimal sampling schemes.

(iii) To incorporate ancillary spatial information, derived hyperspectral data and maps, in optimization of sampling schemes.

(iv) To develop and apply optimal sampling schemes to studies of crops or minerals and associated metals.

(v) To validate results by comparing derived optimal sampling schemes with classical sampling schemes or with existing sampling schemes.

1.6 Outline of thesis

This thesis is essentially a collection of papers, chapters 2–6, which either have been or will be published in international peer-reviewed journals. For this reason, gaps and overlaps may occur between individual chapters. The thesis, however, represents a coherent line of research. The thesis structure has the advantage that some chapters have had the necessary feedback from other scientists during the peer-review process and thus allow to spread the contents of the thesis to a wider audience.

Chapter 2 presents a method of deriving the optimal prospective sampling scheme based on a segmented hyperspectral image. The idea is to stratify an image into homogeneous segments or regions and then to determine the optimum number and locations of samples in each segment. The method is applied to a study of an agricultural field in Hungary, which contains several different crops. A validation study conducted shows that the optimized sampling scheme gives best estimates for commonly used vegetation indices compared to simple random sampling and rectangular grid sampling.

Chapter 3 presents a method for deriving optimal prospective sampling schemes to support ground verification of remotely-sensed distribution of minerals. Optimization of the sampling schemes are based on the results of spectral feature fitting and spectral angle mapper to detect particular minerals using airborne hyperspectral data. The optimization of sampling schemes are illustrated in a case study in the Rodalquilar mineral district (SE Spain). A validation study conducted shows that there is high probability of occurrence of the mineral of interest at the samples selected. The derived optimal sampling schemes are also compared to conventional sampling schemes.

The method presented in chapter 4 can be used to derive the optimal prospective sampling scheme or the optimal retrospective sampling scheme. In this chapter, a mineral prospectivity map, created by weights-of-evidence, was used to optimally locate exploration targets. The method was applied to the Rodalquilar mineral district (SE Spain). Additional maps, such as distance to faults and fractures, were used in conjunction with airborne hyperspectral data to optimize location of exploration targets. On the one hand, the derived optimal prospective sampling design shows that optimized exploration target zones spatially coincide with undiscovered mineral occurrences. On the other hand, the derived optimal retrospective sampling design shows other zones without discovered mineral occurrences within delineated prospective ground.

In chapter 5, a new method for unmixing of hyperspectral data is presented rather than a method for optimization of sampling scheme. It was deemed important to present the new unmixing technique in a separate chapter mainly due to the following three reasons. Firstly, development of such a new technique was deemed essential for the development of another method of optimization of sampling scheme presented in the following chapter. Secondly, the chapter is congruent to one of the specific aims of this thesis; that is, to formulate new techniques useful for designing optimized sampling schemes. Lastly, the chapter is represented by a separate paper in a peer-reviewed journal. The method thus presented in chapter 5 estimates relative abundance of spectrally similar materials in complex mixtures. Using spectra from a spectral library, mixtures of spectrally similar materials were generated. The new method of spectral unmixing uses simulated annealing. Minimization of variance of the differences between the second derivatives of the observed spectrum and the second derivatives of the end-member spectra results in the most precise estimates for the relative abundance of each end-member.

Chapter 6 presents methods to derive the optimal retrospective sampling scheme and

the optimal prospective sampling scheme for surface characterization of mine tailings dumps associated with sulphide-rich mines. The study area is the Recsk-Lahóca copper mining area in Hungary. Relative abundance of minerals of interest, estimated by the method presented in chapter 5, are used as covariates of quantified signature of heavy metal distribution in order to develop optimal retrospective and prospective sampling schemes. For the optimal retrospective sampling scheme, simulated annealing is used to minimize the kriging with external drift prediction error. For the optimal prospective sampling scheme, simulated annealing is used to minimize the kriging with external drift variance.

Finally, chapter 7 summarizes the main findings and conclusions of the thesis, and gives some recommendations for further research.

Chapter 2

The optimal field sampling scheme on a remote sensing segmented image

The tendency of the casual mind is to pick out or stumble upon a sample which supports or defies its prejudices, then to make it the representative of a whole class.

Walter J. Lippmann

This chapter is based on P. Debba, A. Stein, F. D van der Meer, E. J. M. Carranza and A. Lucieer (**In Prep**) The optimal field sampling scheme on a remote sensing segmented image.[1]

[1]This work was sponsored by ITC International Institute for Geo-Information Science and Earth Observation, project number 3083022 and NRF National Research Foundation, project number 10317, gun 2053944.

Abstract

This chapter presents a statistical method for deriving the optimal prospective field sampling scheme on a remote sensing image to represent different categories in the field. The iterated conditional modes algorithm (ICM) is used for segmentation followed by simulated annealing within each category. No pre-sampling of field data is required. The number of sampling points in each category is proportional to the size and variability of the category. Derived field sampling points are more intense in heterogenous segments. This method is applied to airborne hyperspectral data from an agricultural field. The combination of the ICM algorithm with simulated annealing for optimized sampling, results in an elegant and powerful tool for designing the optimal prospective sampling scheme. Such a scheme is useful as it is constructed prior to fieldwork, for categories obtained from remote sensing images. The optimized sampling scheme shows superiority to simple random sampling and rectangular grid sampling in estimating common vegetation indices and is thus more representative of the whole study area.

Keywords

Optimized sampling, iterated conditional modes, segmentation, simulated annealing, stratified.

2.1 Introduction

Sampling entails the selection of a part of a population to draw inference about the whole population. Random sampling is attractive as it implies unbiasedness, resulting in independent observations, which are a basic requirement for statistical inference. In geological and vegetational studies, though, it may conflict with the desire for representativeness. For example, points could be confined to only one part of the area of interest and specific local features may be missed. Geostatistical data often show spatial autocorrelation, hence random sampling may no longer be optimal and for some purposes equally spaced samples or clustered samples are more useful. In addition, estimation of the population mean may benefit from partitioning into homogeneous strata (Thompson, 1992).

Research over the past three decades has advanced in segmenting images composed of multiple bands. Image segmentation may be viewed as a form of data mining, creating objects from pixels, as such resulting into a more informative image than any of its individual spectral bands or individual pixels (Theiler & Gisler, 1997). Image segmentation applies to tomography, remote sensing and pattern recognition.

Segmented hyperspectral images have various applications. In this study the design of the optimal prospective sampling scheme is considered. Design of prospective sampling schemes for classification using ground truth samples has been problematic in geological and vegetational studies. In the past samples have been randomly selected aided with some form of disciplinary judgement (Lu et al., 2003; Brus et al., 1999). High spatial and spectral resolution hyperspectral imagery now makes it possible to select sampling locations in advance of fieldwork. Such images provide a synoptic overview of a large area and often provide topographic information that are more accurate and detailed than ground truth maps (Curran & Williamson, 1985).

In context of vegetational studies, biophysical parameters, such as leaf area index (LAI), biomass, chlorophyll concentration, and photosynthetic activity, are important for estimating foliage cover and forecasting vegetation growth and yield (e.g., Haboudanea et al. (2004)). By selecting appropriate bands, a segmented image can reflect spatial information of variability of certain biophysical parameters, and one could potentially optimize field visits to better estimate these parameters of interest.

Prospective sampling scheme optimization, in support of agricultural studies, based on remotely-sensed segmented images has not been demonstrated in the literature. This chapter considers the design of the optimal prospective sampling scheme for field visits in an agricultural study, using a segmented hyperspectral image. The optimal prospective sampling scheme will be representative of the whole study area for various parameters embedded by the segmentation and bands selected for the segmentation. For image segmentation the iterated conditional modes (ICM) algorithm (Besag, 1986) is used. ICM was chosen since it is reliable and computationally less demanding. Subsequentially, simulated annealing, is applied to optimize sampling within each segmented category. The chapter is illustrated with airborne hyperspectral data,

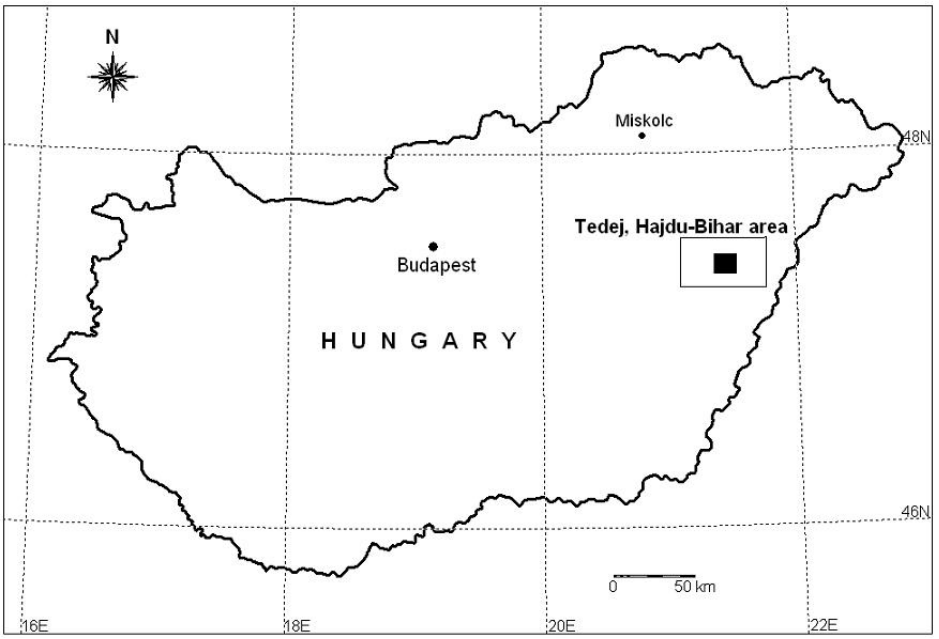


Figure 2.1: Study area in Tedej, Hajdu-Bihar area, Hungary.

DAIS-7915 acquired over the Tedej area in Hungary.

2.2 Study site

2.2.1 Study area

Tedej, Hajdu-Bihar area, Hungary (see Figure 2.1), approximately 1500 ha, is an intensively cultivated agricultural land neighboring a natural protection park area (Kardeván et al., 2003). Soil categories characteristic to this area are Chernozems, Phaeozems, Solonchaks and Solonetz (Kardeván et al., 2003) and major crops are barley, maize, sugar beet, sunflower and alfalfa.

This study area also includes non-vegetation areas (cultivated areas and pathways between fields), and as such was ideal for constraining the sampling. An appropriate sampling scheme, representative of the different crop categories, consist of samples distributed evenly over the respective categories of interest, at the same time avoiding the boundaries of crop categories because of the higher levels of uncertainty at the boundaries.

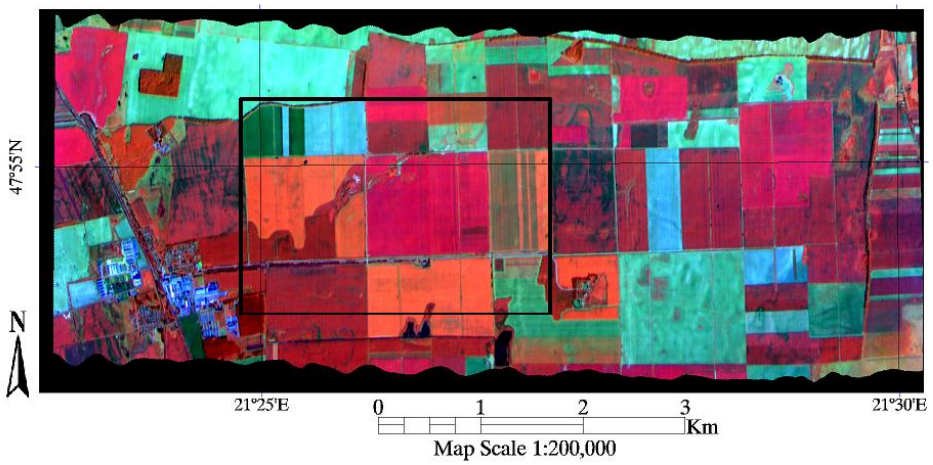


Figure 2.2: Hyperspectral image of study area in Tedej, Hajdu-Bihar area, Hungary.

2.2.2 Remote sensing

In this study, a subset of the Digital Imaging Spectrometer (DAIS-7915), is used. The resulting data is a 79 channel hyperspectral image that was acquired over the Tedej area (see Figure 2.2). DAIS-7915 is a whisk broom sensor, covering a spectral range from visible (0.4 μm) to thermal infrared (12.3 μm) at variable spatial resolution from 3–20 m depending on the carrier aircraft altitude (Lehmann et al., 1995). The 79 channels consist of varying bandwidths which are useful for a large variety of applications.

Flights took place on the 17th and 18th August 2002. Absence of clouds and wind on the second day were appropriate and hence this image was used for the study. Atmospheric effects were minimized by applying an empirical line calibration method (Roberts et al., 1985) to match field spectra measurements. The study area is shown in Figure 2.1 and the hyperspectral image is shown in Figure 2.2 at 5 m nominal resolution on the ground.

Assuming that ground truth data are unavailable at the time of designing a sampling scheme, the bands selected were in close proximity to the wavelengths used in Thenkabail (2002) to identify discriminating crops. The 18 bands selected were: 1 (0.496 μm), 4 (0.551 μm), 10 (0.657 μm), 11 (0.675 μm), 12 (0.693 μm), 13 (0.710 μm), 14 (0.727 μm), 15 (0.744 μm), 23 (0.886 μm), 25 (0.921 μm), 29 (0.988 μm), 32 (1.035 μm), 33 (1.539 μm), 37 (1.668 μm), 39 (1.727 μm), 49 (2.084 μm), 54 (2.158 μm) and 59 (2.232 μm). These include a series of visible, near-infrared and short-wave-infrared bands. Bands 41–48 (1.958–2.068 μm) and 62–72 (2.275–2.412 μm) were noisy and were not considered for selection since they could affect the results of the segmentation. Furthermore, the wavelengths of these bands are usually not considered as being very effective for discriminating between crops as opposed to other bands

(Thenkabail, 2002).

The methodology is illustrated on a 600×400 pixel hyperspectral image, displayed as a black box in Figure 2.1, using the 18 spectral bands. The number of categories in the image is determined by executing an iterative procedure using the Bayesian Information Criteria (BIC). To avoid slow convergence and insufficient computer memory (Fraley & Raftery, 2003), a random sample of pixels was selected. The number of categories was determined repeatedly on subsets of the image for more stable results. The results indicate either seven or eight categories in the image. Since ICM allows merging of categories by resulting in empty categories if they are similar, the algorithm was applied to eight categories. Other, more complex methods of determining the number of categories can be found in Stanford & Raftery (2002) and Forbes & Peyrard (2003) using Pseudolikelihood Information Criteria (PLIC) instead of BIC. It was not intended here, however, to compare performance of these methods or the actual ICM algorithm.

2.3 Methods

2.3.1 The ICM Algorithm

Adequate image segmentation takes into account both spectral features and spatial information. Applications of the Markov Random Fields (MRF) have been useful in this respect (Besag, 1986). The choice of ICM was rather subjective and various other segmentation methods are possible.

The ICM algorithm was applied to include K categories and m spectral bands on an image $\mathbf{I}$ of N pixels, where $(i, j) \in \mathbf{I}$ index the pixel in the image at row i column j and f_{ij} is the m-dimensional feature vector for pixel (i, j).

For each category $k = 1, 2, \ldots, K$, let

- $\mathbf{C}_k^{(\alpha)}$ denote the set of pixels which belongs to the kth category and $\mathbf{C}^{(\alpha)} = \bigcup_{k=1}^{K} \mathbf{C}_k^{(\alpha)}$ the segmented image at the αth iteration, $\alpha = 0, 1, 2, \ldots,$
- $N_k^{(\alpha)}$ denote the number of elements in $\mathbf{C}_k^{(\alpha)}$, i.e. the number of pixels in the kth category at the αth iteration,
- $\mu_k^{(\alpha)} = \sum_{(i,j) \in \mathbf{C}_k^{(\alpha)}} f_{ij} / N_k^{(\alpha)}$ be the m-dimensional mean vector of the kth category at the αth iteration.

For initialization, assume that the feature vectors f_{ij} are conditionally independent and follow a multivariate Gaussian distribution. Only spectral information was utilized to obtain the mean vectors, $\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_K^{(0)}$, for the K categories.

1	2	2
1	(i,j)	2
3	3	5

$N_{ij}^{(\alpha)}(1) = 2$
$N_{ij}^{(\alpha)}(2) = 3$
$N_{ij}^{(\alpha)}(3) = 2$
$N_{ij}^{(\alpha)}(4) = 0$
$N_{ij}^{(\alpha)}(5) = 1$

Figure 2.3: Calculation of $N_{ij}^{(\alpha)}(k)$ for an arbitrary interior pixel (i,j) belonging to category k.

In this study, the K-means multivariate clustering algorithm (Hartigan & Wong, 1979) was used to determine initial mean vectors. For each pixel (i,j) with feature vector f_{ij}, the category k closest to the mean vector is determined, namely, $\min_k \left(f_{ij} - \mu_k^{(0)}\right)^T \left(f_{ij} - \mu_k^{(0)}\right)$ and (i,j) is assigned to category k, i.e., $(i,j) \in \mathbf{C}_k^{(0)}$. The initial segmented image, based only on the spectral reflectance, is denoted as $\mathbf{C}^{(0)} = \bigcup_{k=1}^{K} \mathbf{C}_k^{(0)}$.

At the first and subsequent iterations, $\alpha = 1, 2, \ldots$, the mean $\mu_k^{(\alpha)}$ is determined for each category k, as well as and the within-category variance $\nu^{(\alpha)}$

$$\nu^{(\alpha)} = \frac{1}{N} \sum_{k=1}^{K} \sum_{(i,j) \in \mathbf{C}_k^{(\alpha)}} \left(f_{ij} - \mu_k^{(\alpha)}\right)^T \left(f_{ij} - \mu_k^{(\alpha)}\right). \tag{2.1}$$

A second order MRF was applied in which the neighbors of each pixel consists of its eight adjacencies, with border pixels adjusted appropriately. Let $N_{ij}^{(\alpha)}(k)$ represent the number of neighbors of pixel (i,j), currently classified as belonging to category k at iteration α. An example to determine the values of $N_{ij}^{(\alpha)}(k)$ is illustrated in Figure 2.3. Determine, for each pixel (i,j),

$$\min_k \left\{ \left(f_{ij} - \mu_k^{(\alpha)}\right)^T \left(f_{ij} - \mu_k^{(\alpha)}\right) - \beta \nu^{(\alpha)} N_{ij}^{(\alpha)}(k) \right\}, \tag{2.2}$$

where β is the spatial penalization parameter. After testing various values of β on several images, it was decided to set β to 1.5. Let $k^\star$ be the category that minimizes Equation 2.2 for pixel (i,j). Then (i,j) moves to category $k^\star$ if it presently belongs to another category. After considering all pixels $(i,j) \in \mathbf{I}$, a categorized image $\mathbf{C}^{(\alpha)} = \bigcup_{k=1}^{K} \mathbf{C}_k^{(\alpha)}$ is formed. Next, segmentation $\mathbf{C}^{(\alpha+1)}$ is derived from $\mathbf{C}^{(\alpha)}$ by updating the means $\mu_k^{(\alpha)}$, within category variance $\nu^{(\alpha)}$ and neighbors $N_{(ij)}^{(\alpha)}(k)$. This procedure is applied repeatedly until convergence, that is, no pixels change category, or, in practice for a predefined number of iterations to arrive at a final segmented image. Suppose completion of ICM occurs at iteration r. Then the final segmented image is denoted by

$\mathbf{C}^{(r)} = \bigcup_{k=1}^{K} \mathbf{C}_k^{(r)}$. Hence, the sampling scheme is to be optimized over the ensemble of the K categories, $\mathbf{C}_k^{(r)}$.

2.3.2 Sampling per category

The section now describes the method for obtaining the optimal prospective sampling scheme within each category, after segmentation.

Sample size per category

For a pre-specified number of n samples, suppose a proportional allocation is assigned to each category (Thompson, 1992). An elementary adjustment to the formula presented in Thompson (1992) was made to allow the minimum number of samples per category to be $n_{(0)}$ (see Equation 2.3) so that all categories are represented. The sample size per category is then obtained by distributing the remainder of the samples $\left(n - K \cdot n_{(0)}\right)$ proportionally according to the variability $\left(\nu_k^{(r)}\right)$ and size $\left(N_k^{(r)}\right)$ of the category. The sample size for category k equals

$$n_k = n_{(0)} + \left(n - K \cdot n_{(0)}\right) \cdot \frac{N_k^{(r)} \sqrt{\nu_k^{(r)}}}{\sum_{k=1}^{K} N_k^{(r)} \sqrt{\nu_k^{(r)}}} \,, \tag{2.3}$$

where $\nu_k^{(r)} = \dfrac{1}{N_k^{(r)}} \sum_{(i,j) \in \mathbf{C}_k^{(r)}} \left(f_{ij} - \mu_k^{(r)}\right)^T \left(f_{ij} - \mu_k^{(r)}\right)$.

Sampling is optimized over each category since segmentation is governed by categories, and sampling is carried out to characterize these categories. Choice of an appropriate optimization technique and a relevant criterion is of importance as the samples are to be spread over the category, consisting of several disjoint segments.

Simulated annealing per category

Simulated annealing is a general applicable optimization method to find the global optimum of an objective function in the presence of local optima. In simulated annealing, a fitness function $\phi(\mathbf{S})$, depending on the sampling configuration $\mathbf{S}$, has to be minimized. Starting with a random sampling scheme $\mathbf{S}_0$, let $\mathbf{S}_i$ and $\mathbf{S}_{i+1}$ represent two solutions with fitness $\phi(\mathbf{S}_i)$ and $\phi(\mathbf{S}_{i+1})$, respectively. Sampling scheme $\mathbf{S}_{i+1}$ is derived from $\mathbf{S}_i$ by randomly replacing one of the points of $\mathbf{S}_i$ towards a new point not in $\mathbf{S}_i$. A probabilistic acceptance criterion decides whether $\mathbf{S}_{i+1}$ is accepted or

not:

$$P_{\mathbf{c}}(\mathbf{S}_i \rightarrow \mathbf{S}_{i+1}) = \begin{cases} 1\ , & \text{if } \phi(\mathbf{S}_{i+1}) \leq \phi(\mathbf{S}_i) \\ \exp\left(\dfrac{\phi(\mathbf{S}_i) - \phi(\mathbf{S}_{i+1})}{\mathbf{c}}\right)\ , & \text{if } \phi(\mathbf{S}_{i+1}) > \phi(\mathbf{S}_i) \end{cases} \tag{2.4}$$

where $\mathbf{c}$ denotes a positive control parameter, usually called the temperature in simulated annealing problems. The parameter $\mathbf{c}$ is lowered according to a cooling schedule as the process evolves, to find the global minimum. A transition takes place if $\mathbf{S}_{i+1}$ is accepted. Next, a solution $\mathbf{S}_{i+2}$ is derived from $\mathbf{S}_{i+1}$, and the probability $P_c\left(\mathbf{S}_{i+1} \rightarrow \mathbf{S}_{i+2}\right)$ is calculated with a similar acceptance criterion as Equation 2.4. At each value of $\mathbf{c}$, several transitions have to be made before the annealing can proceed, and $\mathbf{c}$ can take its next value.

A linear cooling schedule has been chosen, which is slow but prevents solutions at local minima to increase the chance of arriving at the global minimum. The cooling schedule starts with an initial value $\mathbf{c}_0$ which has an acceptance ratio (γ) of 0.95 or higher for alternative solutions. For $l = 0, 1, 2, \ldots$, the decrements of $\mathbf{c}$ is given by $\mathbf{c}_{l+1} = \gamma \cdot \mathbf{c}_l$, with $0 < \gamma < 1$.

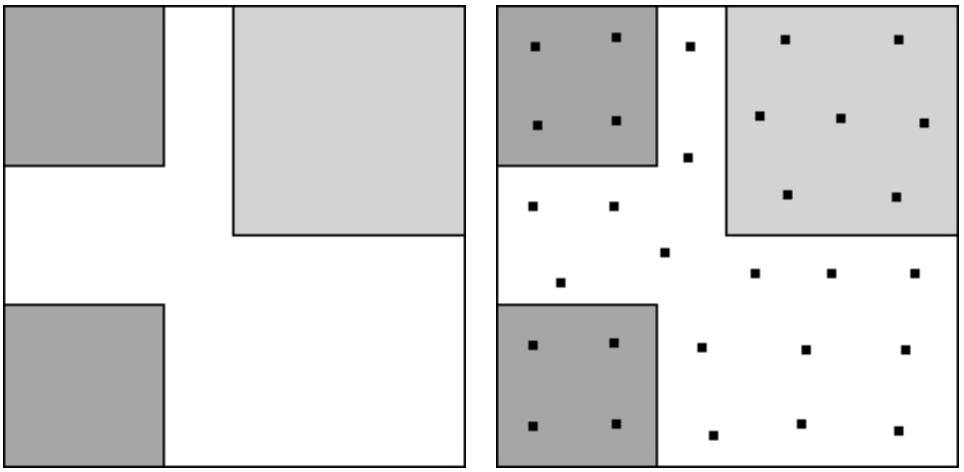
Fitness function per category

Various choices of fitness functions $\phi(\mathbf{S})$ can be made. For example, minimization of the Mean Shortest Distances (MMSD)-criterion aims at even spreading of all sampling points over the sampling region by minimizing the expectation of the distance between an arbitrary chosen point and its nearest observation (Van Groenigen & Stein, 1998). Each category is considered separately as a sampling region. The MMSD-criterion was chosen since it is able to spread the sampling points over each category, each category consisting of possibly several segments. In effect most segments in the image will also be sampled since the objective of this criterion will force sampling points to move to different segments, belonging to a common category, depending on the distance between nearby segments.

The final segmented image $\mathbf{C}^{(r)}$ with $N_k^{(r)}$ observation points is considered on a finely meshed grid in each of the $\mathbf{C}_k^{(r)}$ categories. Since a remote sensing image is raster based, the image is already on a finely meshed grid consisting of individual pixels. Hence all chosen points correspond to pixels. Regular spreading can be formulated as minimizing the expectation of the distance between an arbitrarily chosen point within the category, and its nearest sampling point.

The initial sampling scheme for the kth category $\mathbf{S}_k^{(0)}$ is a random selection of n_k [see Equation 2.3] points from category k. For $\mathbf{S}_k$, the fitness function equals

$$\phi_{\text{MMSD}}(\mathbf{S}_k) = \frac{1}{N_k^{(r)}} \sum_{(i,j) \in \mathbf{C}_k^{(r)}} \left|\left| c_{k(ij)} - W_{\mathbf{S}_k}(c_{k(ij)}) \right|\right|\ , \tag{2.5}$$



(a) (b)

Figure 2.4: (a) Segmented image and (b) Optimized sampling locations constrained to each category.

where $c_{k(ij)} \in \mathbf{C}_k^{(r)}$ is a location vector denoting the (i,j)th pixel belonging to category k and $W_{\mathbf{S}_k}(c_{k(ij)})$ denotes the location vector of the nearest sampling point in $\mathbf{S}_k$.

Minimizing Equation 2.5 results in an even spreading of sampling points within category k, i.e. points will arrange at an equilateral triangular configuration. The final sampling scheme $\mathbf{S}$ consists of all sampling points in the k categories, $\mathbf{S} = \bigcup_{k=1}^{K} \mathbf{S}_k$.

2.4 Results

The MMSD-criterion is first illustrated on a segmented image with distinct boundaries and regular segments. Next, it is applied to the DAIS-7915 image acquired over an agricultural field in Hungary.

2.4.1 Experimental study

The method of deriving the optimal sampling scheme is applied to a segmented image with three catergories. The objective was to illustrate the sampling technique, as the ICM algorithm has already been proven in its field of image segmentation.

The image (Figure 2.4(a)) consists of three categories, and assume the standard deviations of 10, 8 and 12 for each of the three categories. One category has two disjoint segments on the left top and bottom part of the image to demonstrate sampling being distributed over these segments. This allowed illustrating that sampling was

constrained by the categories and segments.

A total sample size of 30 was distributed as 7, 8 and 15 for each of the three categories using Equation 2.3. Simulated annealing was applied to the segmented image using the MMSD-criterion as the objective function. The initial temperature was set at 10, the cooling rate ($\alpha = 0.95$), the number of iterations at each temperature 20000 and the number of iterations without change 2000. The resulting sampling scheme is shown in Figure 2.4(b).

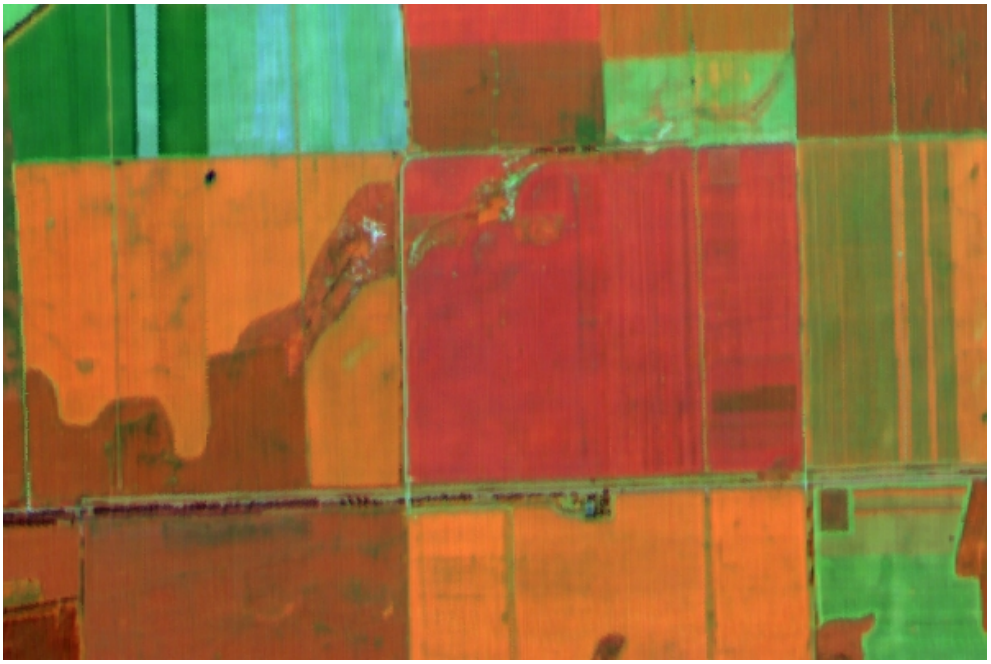
2.4.2 Application

ICM was then applied to the subset hyperspectral data. The resulting segmented image with eight categories is shown in Figure 2.5(b), which are the mean reflectance values for bands 29, 39 and 1. The sampling scheme was optimized by using the segmented image with eight categories. The image was first analyzed and adjustments were made to prevent sampling in areas of no interest. First, pixels belonging to pathways, between fields and non-vegetation categories (cultivated areas) were removed after segmentation but prior to sampling. Second, disjoint segments smaller than 10 pixels were removed from further analysis in the sampling stage. This decreased the chance of sampling in segments that were a result of pure noise in the image as it seemed impractical to sample in such small areas. However, if these are meaningful segments, the above procedure can be performed without removal of these pixels. Four of the eight categories were cultivated crops and hence combined with pixels belonging to pathways between the fields to form a category that constrained sampling. The resulting image is shown in Figure 2.6(a) where sampling was optimized over categories 1–4.

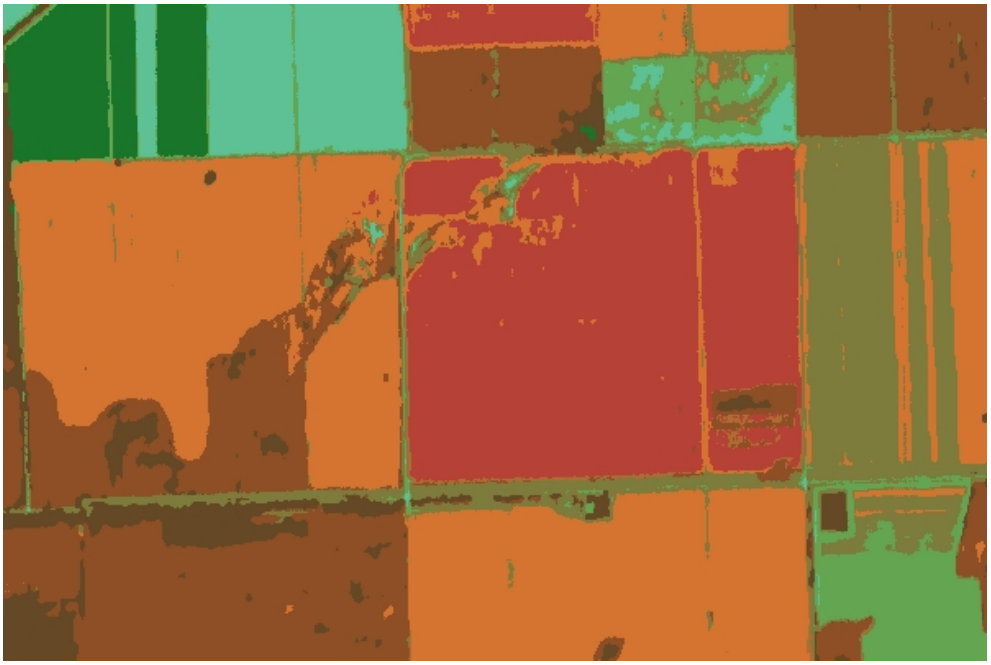
As an illustration, a total sample size of 50 points was used. The number of samples for each category was then determined using Equation 2.3 by stipulating the minimum number of samples per category as 3. This resulted in 7, 11, 22 and 10 samples for each of the four categories. Samples for each category were then optimized by using simulated annealing and the MMSD-criterion as the fitness function. The distribution of the 50 sampling points is shown in Figure 2.6(b). In this image, the optimized prospective sample points tend to arrange in a triangular formation, while being governed by the actual complexity of the segments. Most samples are also arranged away from the borders of the segments.

2.5 Validation

This section demonstrates that the optimal prospective sampling scheme is suitable to estimate various vegetation parameters. It is known that the Leaf Area Index (LAI) is often used as a key variable in estimating foliage cover and forecasting vegetation growth and yield (Haboudanea et al., 2004). Several researchers have shown that there is a relationship between ground-measured LAI and vegetation indices (Spanner et al.,

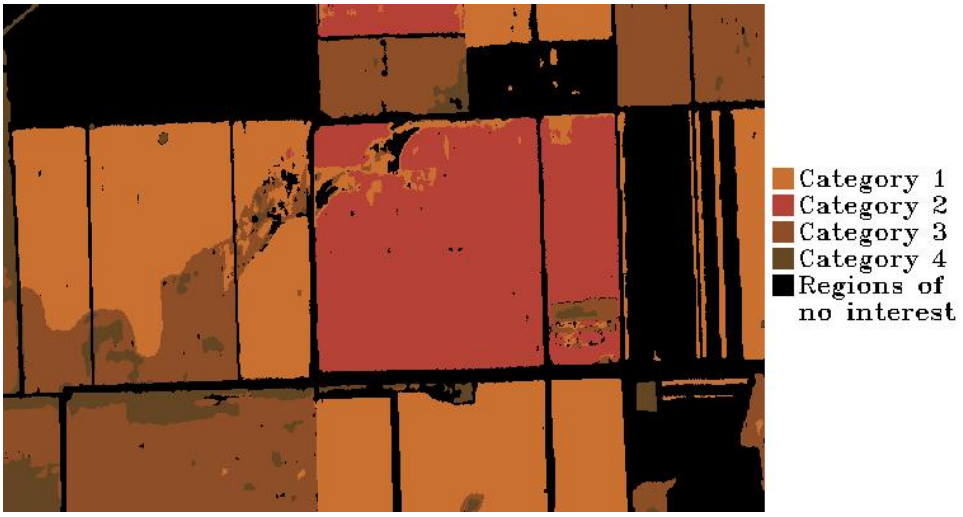


(a) Original hyperspectral image



(b) ICM Segmented image with eight categories

Figure 2.5: ICM image segmentation on the original DAIS-7915 hyperspectral image.



(a) Segmented image confining sampling regions to the four categories

(b) Optimized sampling locations of 50 points

Figure 2.6: Optimized sampling locations of 50 points distributed over the 4 categories.

1990; Baret & Guyot, 1991; Chen, 1996; Chen & Cihlar, 1996; Fassnacht et al., 1997; Brown et al., 2000).

The reason for the varying vegetation indices is that different indices are used to capture different vegetation parameters of interest. Some indices are used to capture the photochemical processes associated with photosynthesis (Chappelle et al., 1992; Gamon et al., 1992; Daughtry et al., 2000; Broge & Leblanc, 2000; Haboudanea et al., 2004), while others for retrieval of LAI estimates (Chen & Cihlar, 1996; Brown et al., 2000; Qi et al., 2000), or to quantify vegetation fraction (Gitelson et al., 2001). In all, vegetation indices are well correlated with various vegetation parameters such as LAI, biomass, chlorophyll concentration, and photosynthetic activity.

The Normalized Difference Vegetation Index (NDVI) developed by (Rouse et al., 1974) is the most widely used vegetation index. In terms of the hyperspectral bands, NDVI is defined as

$$\mathrm{NDVI} = \frac{R_{886} - R_{675}}{R_{886} + R_{675}} \tag{2.6}$$

where R_x is the reflectance of the wavelength at x nm. The NDVI is based on the contrast between the maximum absorption in the red due to chlorophyll pigments to the maximum reflection in the infrared caused by leaf cellular structure. Figure 2.7(a) shows the NDVI image, where brighter pixels are indicative of healthy vegetation.

The Renormalized Difference Vegetation Index (RDVI) was developed by Rougean & Breon (1995) to improve estimates of LAI. RDVI for the hyperspectral image is defined as

$$\mathrm{RDVI} = \frac{R_{886} - R_{675}}{\sqrt{R_{886} + R_{675}}}\,. \tag{2.7}$$

Figure 2.7(b) shows the RDVI image, where brighter pixels are indicative of healthy vegetation.

The Modified Simple Ratio (MSR) developed by Chen (1996) was also designed to improve estimates of LAI. MSR for the hyperspectral image is defined as

$$\mathrm{MSR} = \left(\frac{R_{886}}{R_{675}} - 1\right) \Big/ \sqrt{\frac{R_{886}}{R_{675}} + 1}\,. \tag{2.8}$$

Figure 2.7(c) shows the MSR image, where brighter pixels are indicative of healthy vegetation.

In attempting to improve the vegetation indices with regard to background soil, Qi et al. (1994) modified the Soil-Adjusted Vegetation Index (MSAVI), which is defined as

$$\mathrm{MSAVI} = \frac{1}{2}\left[(2R_{886} + 1)^2 - 8(R_{886} - R_{675})\right]\,. \tag{2.9}$$

Figure 2.7(d) shows the MSAVI image, where brighter pixels are indicative of healthy vegetation. Also noticeable is the region with high salinity, running diagonally across the center of the image, has high values for MSAVI since the effect of the background soil is diminished.

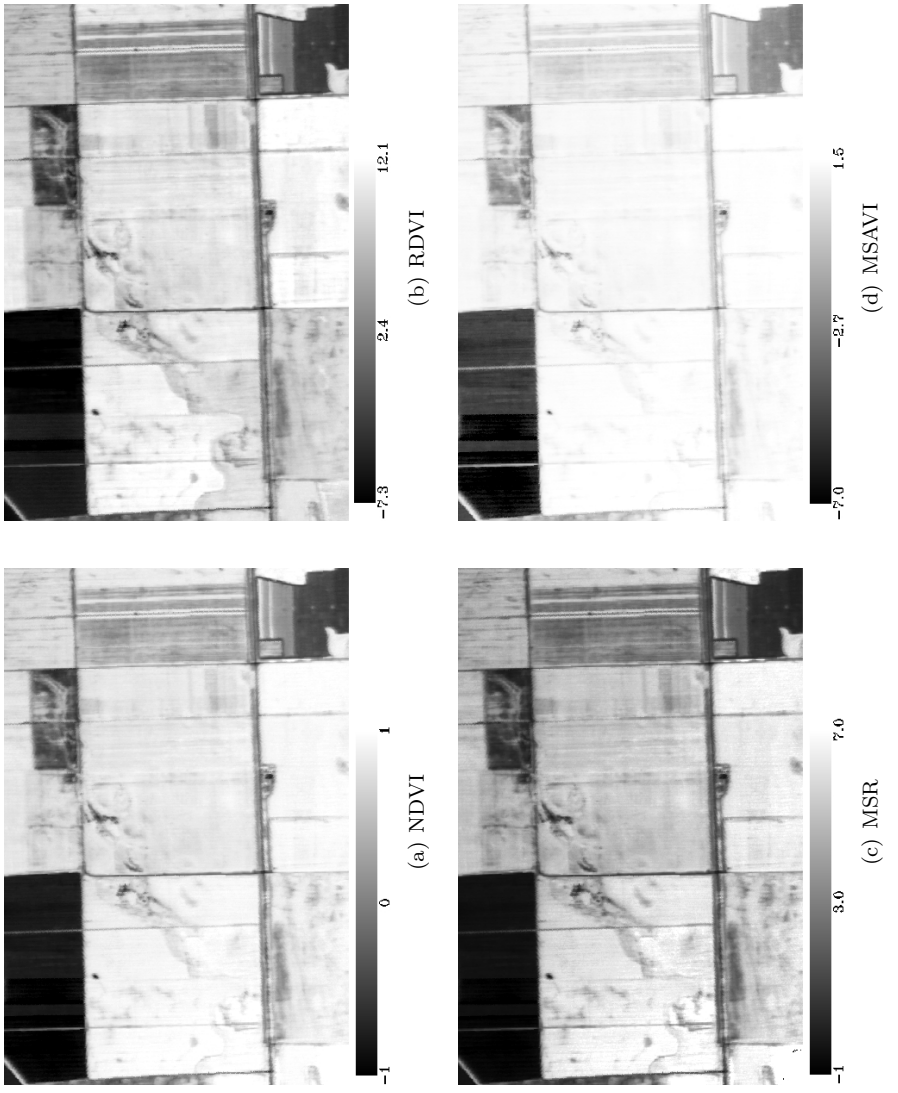


Figure 2.7. Vegetation indices maps.

Table 2.1: Comparison of sampling schemes.

		Mean			
		NDVI	RDVI	MSR	MSAVI
Image		0.59	8.8	1.34	1.24
Optimized sampling scheme		0.58	8.6	1.32	1.22
Random sampling scheme	1	0.49	7.9	1.18	1.09
	2	0.38	6.1	0.94	0.89
	3	0.45	7.0	1.11	1.06
Grid sampling scheme	1	0.49	7.8	1.14	1.13
	2	0.53	8.2	1.25	1.13
	3	0.53	8.3	1.26	1.15

Values in the table represent the average vegetation indices for each sampling scheme.

The above four indices are used to show that the optimal prospective sampling scheme gives most precise estimates for each vegetation index compared to simple random sampling and grid sampling. For simple random sampling, three sampling schemes are designed, by randomly selecting 50 samples from the region of interest, to show inconsistency in the estimates. For the rectangular grid scheme, three sampling schemes were designed to again show inconsistency in the estimates. Initially, for each grid, 54 samples were selected, with a grid spacing of 300 m. The first sample was randomly selected from a 300×300 m grid. Nine samples were arranged in a row and six in a column, thus totalling to the 54 samples. For each grid sampling scheme, the samples falling in the non-interest area were removed. This resulted in either three or four samples removed from each grid sampling scheme. The average vegetation index for each sampling scheme is then calculated and compared to the respective average vegetation index for all pixels in the area of interest. The results are shown in Table 2.1. The optimized sampling scheme gives most accurate estimates for each vegetation index compared to the simple random and rectangular grid sampling schemes. Inconsistencies are apparent for each of the latter two sampling designs.

2.6 Discussion

The sampling methodology presented in this chapter was tested on a synthetic segmented image with three categories, one of which had two disjoint segments. Each segment had a distinct boundary with a regular shape. The 30 samples were well distributed over the three categories. The sampling points for the category consisting of two disjoint segments were distributed evenly among these segments. This was useful as the MMSD-criterion was appropriately spacing the sampling points. The sampling

scheme in this simple illustration implies that with irregular segments created in realistic applications, the MMSD-criterion will distribute the samples over each category appropriately. The resulting samples will thus be representative of these categories.

The sampling methodology was further tested on a subset of a DAIS-7915 hyperspectral image by initially applying the ICM algorithm. The samples per category were obtained in proportion to the size and variability of the category as fewer samples are required to estimate smaller, more homogenous categories. The assumption that categories, comprising of possibly several segments, are homogeneous, logically makes it possible to spread sampling points over each category by using the MMSD-criterion as the fitness function in simulated annealing. Although simulated annealing is a slow process, it arrives at the global optimum. Disjoint segments for each category results in many local minima for the fitness function and hence necessitates such an algorithm. Most of the samples are away from the boundaries of these categories where uncertainty is high. This highlights the appropriateness of the implemented fitness function. This is further supported by samples spread over the categories, forming an equilateral triangular structure. It depended though on the spatial complexity of the category. Since these sampling points are geo-referenced, the location can be determined for field visits at these sampling points on the ground.

Field spectral measurements of agricultural crops could be used, for example, to study health of crops and thus are important for estimating foliage cover and forecasting vegetation growth and yield. This necessitates regular field visits. Because fields of agricultural crops may be homogeneous (planted with only one particular type of crop) or heterogeneous (planted with a group of different types of crops), the lack of this prior information can make field sampling non-optimal. Classification of agricultural fields is therefore useful in the optimization of prospective sampling schemes to support provisions that will sustain economic crop productivity. This hypothesis converges with Thompson (1992) suggestion that estimation of the population mean may benefit from partitioning a study area into homogeneous strata.

Using appropriate band ratios, commonly used vegetation index maps were used to compare the optimal prospective sampling scheme to simple random sampling and rectangular grid sampling schemes. The optimal prospective sampling scheme results in estimates for each of the four vegetation indices considered that are closest to the actual averages for all pixels in the image. The rectangular grid sampling schemes also have reasonably accurate estimates but depends highly on the initial random sample. Inconsistencies between each rectangular grid sampling scheme can be observed. The estimates are poor using simple random sampling and inconsistencies for these sampling schemes are high. The optimized sample points therefore represent the whole study area more accurately than either of the other two sampling schemes for each of the different parameters of interest.

The ICM algorithm has regularly been applied for various purposes as it is quick and produces reasonably accurate categories. The drawback is that it may arrive at a local optimum, hence emphasizing the choice of initial means. For this reason, the K-means multivariate clustering algorithm was used as opposed to randomly selecting

points in the image to represent the means for each category. Choice of spectral bands to use and restriction of categories to sample depends on the application and actual data. This is inevitable as relying solely on segmentation can result in sampling categories of no interest. Although bands selected in this study are not necessarily the optimal discriminating bands, they are commonly used for calculating the red-edge, vegetational indices or the amount of stress in vegetation. The segmented image thus inherit these qualities and resulted in an optimal prospective sampling scheme that best estimates the various vegetation indices.

Sampling of categories is not restrictive to the use of the ICM algorithm and in this sense any other segmentation method may be used. This methodology uses remote sensing in designing optimal prospective sampling schemes on the ground for field visits as opposed to the traditional way of selecting samples randomly or on the basis of disciplinary judgement. The selected samples will in effect have image characteristics, such as, gray tone, texture or pattern, depending on the type of segmentation performed.

2.7 Conclusions

This study resulted into three main conclusions.

- The combination of the ICM algorithm for image segmentation and simulated annealing for optimized sampling provides an elegant and powerful tool in designing prospective sampling schemes using hyperspectral data.
- The optimized prospective sampling scheme shows superiority to simple random sampling and rectangular grid sampling in estimating common vegetation indices and is thus more representative of the whole study area.
- The synthetic segmented image illustrates the use of the MMSD-criterion by ensuring allocation of sampling points in disjoint segments belonging to a common category. In regular segments, samples appeared in an almost equilateral triangular design, thereby implying the appropriateness of the MMSD-criterion for complicated practical applications.

Chapter 3

Optimization of field sampling for target minerals identified from hyperspectral data

Dig where gold is . . . unless you just need some exercise.

John M. Capozzi

This chapter is based on P. Debba, F. J. A. van Ruitenbeek, F. D. van der Meer, E. J. M. Carranza and A. Stein (2005). Optimal field sampling for targeting minerals using hyperspectral data, *Remote Sensing of Environment*, Vol 99, (pp. 373–386).[1]

[1]This work was sponsored by ITC International Institute for Geo-Information Science and Earth Observation, project number 3083022 and NRF National Research Foundation, project number 10317, gun 2053944.

Abstract

This chapter presents a statistical method for deriving optimal spatial sampling schemes. It focuses on ground verification of minerals identified and estimated from hyperspectral data. Spectral angle mapper (SAM) and spectral feature fitting (SFF) classification techniques were applied to obtain rule mineral images. Each pixel in these rule images represents the similarity between the corresponding pixel in the hyperspectral image to a reference spectrum. The rule images provide weights that are utilized in objective functions of the sampling schemes, which are optimized through a process of simulated annealing. A HyMAP 126-channel airborne hyperspectral data acquired in 2003 over the Rodalquilar area in Spain serves as an application to target those pixels with the highest likelihood of occurrence of a specific mineral, and as a collection, the location of these sampling points selected represent the distribution of that particular mineral. In this area, alunite being a predominant mineral in the alteration zones was chosen as the target mineral. Three weight functions are defined to intensively sample areas where a high probability and abundance of alunite occurs. Weight function I uses binary weights derived from the SAM classification image, leading to an even distribution of sampling points over the region of interest. Weight function II uses scaled weights derived from the SAM rule image. Sample points are arranged more intensely in areas of abundance of alunite. Weight function III combines information from several different rule image classifications. Sampling points are distributed more intensely in regions of high probable alunite as classified by both SAM and SFF, thus representing the purest of pixels. This method leads to an efficient distribution of sample points, on the basis of a user-defined objective.

Keywords

Optimized sampling, simulated annealing, spectral angle mapper, spectral feature fitting, weighted mean shortest distance, rule image, minerals, alunite, hyperspectral

3.1 Introduction

Remote sensing imagery has contributed significantly to mineral exploration. For example, mapping of faults and fractures that localize ore deposits and recognize hydrothermally altered rocks based on their spectral signatures (Sabins, 1999). A major problem with remote sensing approaches to mineral exploration using broad-band multispectral sensors is the insufficient spectral resolution to map hydrothermal alteration minerals, which exhibit subtle differences in spectral signatures (Clark, 1999). The advent of new hyperspectral sensor technology, in terms of both sensor and technique development, has provided opportunity to revisit previous remote sensing approaches to mineral exploration as well as development of improved methods.

Small bandwidths distinguish hyperspectral sensors from multispectral sensors, acquiring spectral information of materials usually over several hundreds of narrow contiguous spectral bands, with high spectral resolution on the order of 20 nm or narrower (Polder & van der Heijden, 2001). As such, they allow identification of specific materials, whereas broad-band multispectral data only allow discrimination between classes of materials (Kruse et al., 2003). In effect, hyperspectral imaging systems are useful in identifying individual iron and clay minerals, which can provide details of hydrothermal alteration zoning (Sabins, 1999) based on specific absorption features of these minerals. Thorough discussions on absorption features of hydrothermal alteration minerals can be found in Hapke (1993), Salisbury et al. (1991), Van der Meer (2004), and Clark (1999). Various mapping of minerals using hyperspectral data can be found in Kruse & Boardman (1997), Sabins (1999), Vaughan et al. (2003), Rowan et al. (2000), and Crósta et al. (1998).

Surface sampling in the field is often advantageous for starting surveys. Identification of hydrothermal alteration minerals like alunite, kaolinite and pyrophyllite, from hyperspectral images leads to a better understanding of the geology and alteration patterns in a region. As such, the analysis of airborne hyperspectral imagery can aid in selecting follow-up targets on the ground before fieldwork is performed. In this study the focus is on the mineral alunite as it is characteristic of hydrothermal alteration zones in the Rodalquilar area in Spain (Arribas et al., 1995). Alunite has a distinct spectral signature and is often, although not always, related to high sulphidation epithermal gold (Hedenquist et al., 2000). The purpose was to guide field sampling collection to those pixels with the highest likelihood for occurrence of alunite, while representing the overall distribution of alunite. The method offers an objective approach to selecting sampling points in order to, for example, create a mineral alteration map. However, this method can be easily extended to other hydrothermal alteration minerals that have diagnostic absorption features. Combination of several mineral images can then be used in classification of the image to create an alteration map.

Purposive sampling, where observations are made in linear traverses utilizing the skills and knowledge of the field worker is a common geological sampling method (Gupta, 2003). Consequences of such sampling include extrapolating observations for

the study area as a whole, and leads to subjectivity (Gupta, 2003). A different way of sampling is given by de Gruijter & ter Braak (1990). This requires definition of a mathematical objective function, related spatially to the desired sampling points. The objective function can be optimized through simulated annealing (Sacks & Schiller, 1988; McGwire et al., 1993; Van Groenigen & Stein, 1998). Simulated annealing is a computer intensive search technique to find the optimum value of the objective function by continually updating this function at successive steps. Optimal sampling schemes have also been derived to guide sampling to target specific areas or "hot spots" (Van Groenigen et al., 2000a,b). So far, these sampling schemes have not been combined with remote sensing. Model-based sampling differs from designed-based sampling that requires a random component, so that all parts of the area should have a non-zero inclusion probability (de Gruijter & ter Braak, 1990).

The present study aims to use the spectral angle mapper (SAM) and spectral feature fitting (SFF) to classify alunite and obtain rule images. Each pixel in a rule image represents the similarity between the corresponding pixel in the hyperspectral image to a reference spectrum. These rule images are then used to govern sampling to areas with a high probability of alunite occurring and to intensively sample in areas with an abundance of alunite. This effectively delineates favorable areas from unfavorable ones and provides an objective sampling scheme as an initial guideline. The optimal sampling scheme to target areas of a particular hydrothermal alteration mineral is the objective of this study. Such an optimal prospective sampling scheme could potentially be used for mineral exploration, which can be time-consuming, cost-prohibitive and involve a high degree of risk in terms of accurate target selection (Srivastav et al., 2000). The study is illustrated with hyperspectral data acquired over the Rodalquilar area.

3.2 Study area

3.2.1 Altered rocks in Los Tollos area

The Los Tollos area is part of the Rodalquilar mineral district in the Sierra del Cabo de Gata volcanic field, in south-eastern Spain (Figure 3.1). Volcanic rocks in the district range in composition from pyroxene-bearing andesites to rhyolites. These rocks have been intensely altered primarily due to volcanic geothermal activity (hypogene alteration) and secondarily due to chemical weathering (supergene alteration). Associated with some parts of the intensely altered rocks are high sulphidation gold deposits and low sulphidation base metal deposits. Arribas et al. (1995) distinguish five hydrothermal alteration zones: silicic, advanced argillic, intermediate argillic, sericitic, and propylitic (Table 3.1). In addition to hypogene advanced argillic alteration is supergene advance argillic alteration, which Arribas et al. (1995) call Stage 2 alunite. The Los Tollos area was selected because it remains relatively undisturbed by previous mining activities in the district. The volcanic rocks in the Los Tollos area have been affected by intense silicic, advanced argillic, and stage 2 alunite alterations.

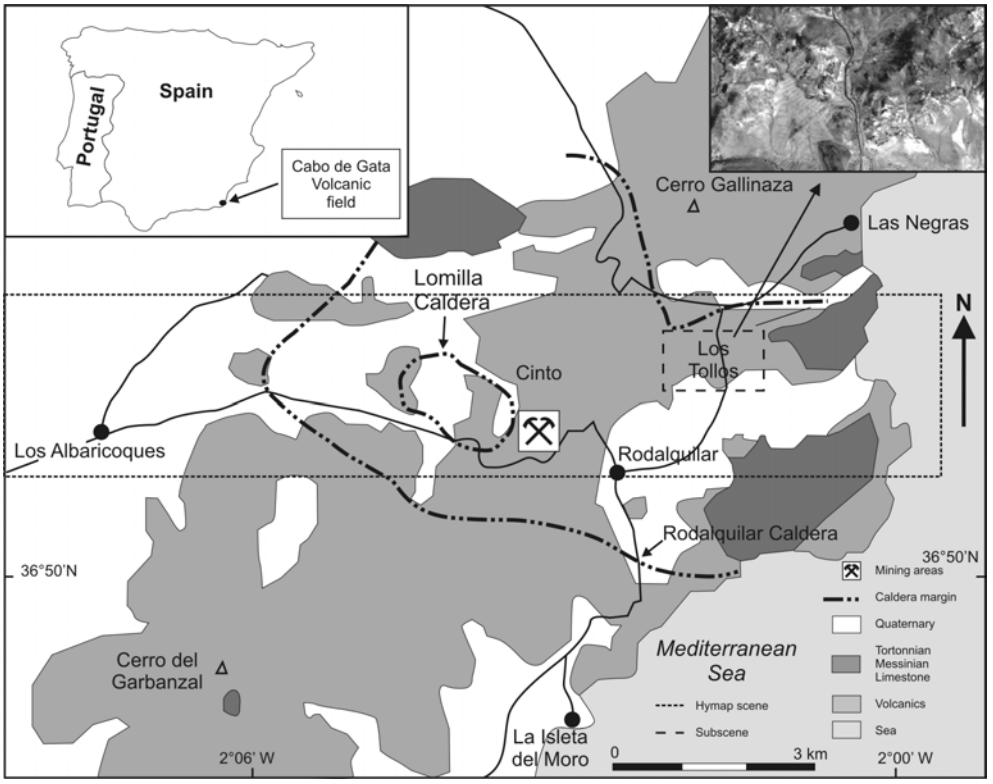


Figure 3.1: A generalized geological map (modified after Cunningham et al. (1990)) of the Rodalquilar study area showing the flight line (dotted box) and the hyperspectral data (top right corner and dashed box) used in the present chapter.

Table 3.1: Summary of alteration zones and dominant minerals in the Rodalquilar area (Arribas et al., 1995).

Alteration Zone	Alteration Minerals
Silicic	quartz; chalcedony; opal
Advanced Argillic	quartz; alunite; kaolinite; pyrophyllite; illite; illite-smectite
Intermediate Argillic	quartz; kaolinite; illite; illite-smectite
Sericitic	quartz; illite
Propylitic	quartz; illite; montmorillonite
Stage 2 Alunite	alunite; kaolinite; jarosite

3.2.2 Data

A sub-scene, consisting of 350 × 225 pixels of airborne imaging spectrometer data was used. The data were acquired by the Hyperspectral Mapper (HyMAP) in July 2003 during the HyEUROPE 2003 campaign (Figure 3.1). HyMAP is a 126-channel instrument that collects data in a cross-track direction by mechanical scanning and in an along-track direction by movement of the airborne platform. The instrument acts as an imaging spectrometer in the reflected solar region of the electromagnetic spectrum (0.4–2.5 μm). Spectral coverage is nearly continuous in the SWIR and VNIR regions with small gaps in the middle of the 1.4 and 1.9 μm atmospheric water bands. The spatial configuration of the instrument accounts for an IFOV of 2.5 mrad along track and 2.0 mrad across track resulting in a pixel size on the order of 3–5 m for the data presented in this chapter. Due to instrument failure the SWIR 1 detector did not function during acquisition, thus no data were acquired in the 1.50–1.76 μm window. The HyMAP data were atmospherically and geometrically corrected using the Atmospheric and Topographic Correction (ATCOR 4) model (Richter, 1996).

In support of the imaging spectrometer data, field spectra were collected from some parts of the study area during the over-flight using the Analytical Spectral Device (ASD) fieldspec-pro spectrometer. This spectrometer covers the 0.35–2.50 μm wavelength range with a spectral resolution of 3 nm at 0.7 μm and 10 nm at 1.4 and 2.1 μm. The spectral sampling interval is 1.4 nm in the 0.35–1.05 μm wavelength range and 2 nm in the 1.0–2.5 μm wavelength range.

The SWIR 2 detector of HyMap, with a spectral range 1.95–2.48 μm (bandwidth 16 nm), is potentially useful for mapping alteration assemblages as well as regolith characterization (Abrams et al., 1977; Goetz & Srivastava, 1985; Cudahy et al., 2000; Papp & Cudahy, 2002; Kruse, 2002). HyMAP has been used successfully to map minerals (Martini et al., 2003; Martini, 2003; Papp & Cudahy, 2002; Cudahy et al., 2000) and detect faults and fractures (Martini et al., 2003). Dimensionality of the data was reduced by considering only the channels in the spectral range 1.970–2.468 μm. This spectral range covers the most prominent spectral absorption features of hydroxyl-bearing minerals, sulfates and carbonates, which are common to many geologic units and hydrothermal alteration assemblages (Kruse, 2002). These minerals also exhibit distinctive absorption features at wavelengths in the partly missing range of 1.4–1.7 μm, a range also affected by the water absorption features in the atmosphere.

Figure 3.2 shows spectral plots of seven of the most prominent alteration minerals in the study area (Arribas et al., 1995), at a spectral resolution coinciding with HyMAP after continuum removal was applied. Continuum removal normalizes the respective spectra to enable comparison of absorption features from a common baseline. The continuum is a function of the wavelength that is fitted over the top of the spectrum between two local spectra maxima. A straight line segment joins the first and last spectral data values taken as the local maxima (Clark et al., 1991; Clark & Roush, 1984). This figure shows differences in absorption features of the different minerals, in terms of shape, size, symmetry, depth and wavelength position. These distinct characteristics enable researchers to identify individual minerals from hyperspectral

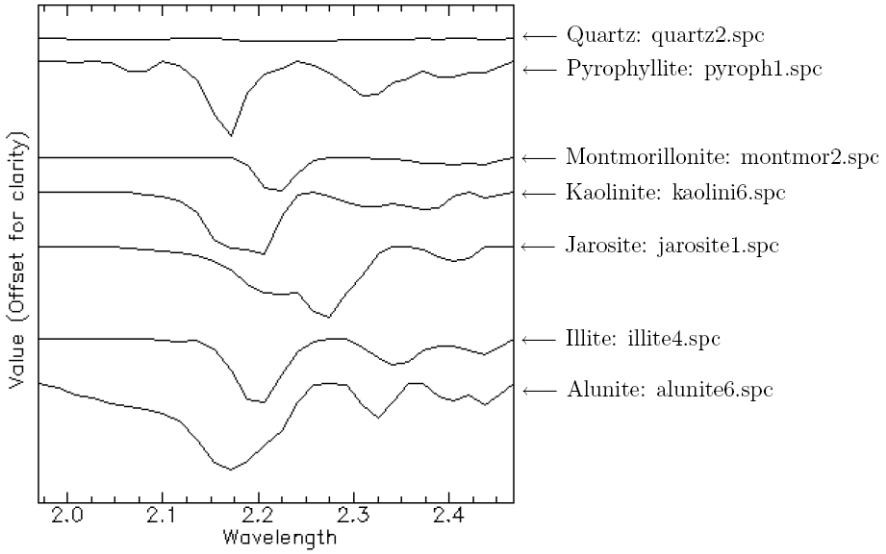


Figure 3.2: Plot of 7 end-members from USGS spectral library (Clark et al., 1993) for the 30 selected channels, enhanced by continuum removal.

data. The spectrum of quartz has no distinctive absorption feature (in this spectral range), but the remaining spectra have distinctive absorption features at wavelengths near 2.2 μm, each differing slightly in position and geometry.

Alunite was chosen among the seven most prominent alteration minerals in the area (Hedenquist et al., 2000) because it has distinct absorption characteristics (Hapke, 1993; Salisbury et al., 1991; Van der Meer, 2004; Clark, 1999), which are recognizable from hyperspectral images (Kruse & Boardman, 1997; Sabins, 1999; Vaughan et al., 2003; Rowan et al., 2000; Crósta et al., 1998). Although this study concentrates on one hydrothermal mineral, namely alunite, the method demonstrated can easily be extended to other minerals of interest. The test image selected was in an area that was relatively undisturbed through excavation, hence between 2–3 km from the nearest gold mining area as indicated in Figure 3.1.

3.3 Methods

The method for obtaining the optimal sampling scheme commences with application of two classification techniques used, namely, spectral angle mapper (SAM) (Kruse et al., 1993) and spectral feature fitting (SFF) (Clark et al., 1991) to obtain rule images. The digital number (DN) values in a rule image represent similarity between each corresponding pixel's spectrum to a reference mineral spectrum, resulting in one

rule image for each mineral considered. Binary and scaled weights are then derived from the rule images. These weights are used in a mathematical objective function (defined in Equation 3.7, see also Van Groenigen et al. (2000b)), which is optimized in relation to the spatial distribution of the georeferenced image pixels representing a collection of alunite samples in the field. The aim of optimizing the objective function is to spread the location of the alunite sampling points over the region while targeting pixels that have a high probability of being alunite. In effect, the location of these samples in the field will be dense if distributed in areas with an abundance of alunite and where pixels have a high probability of being alunite. Optimization of the objective function is an exhaustive combinatorial problem. The complexity of the objective function and the iterative process of randomly selecting a pixel in the image, as a new sampling point replacing an old one from the collection give rise to many local optima, which is solved through simulated annealing.

3.3.1 Spectral Angle Mapper (SAM) Classifier

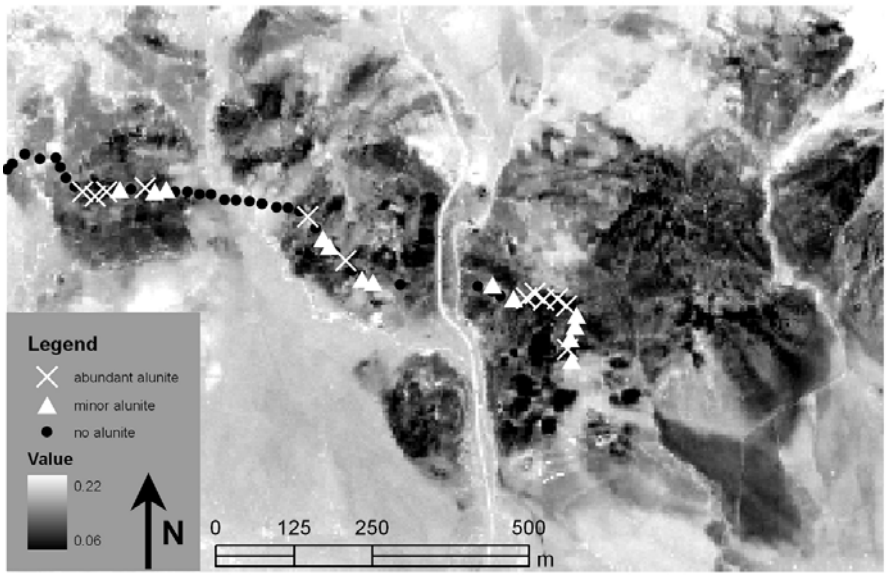
SAM is a pixel based supervised classification technique that measures the similarity of an image pixel reflectance spectrum to a reference spectrum from either a spectral library or field spectrum (Kruse et al., 1993). This measure of similarity is the spectral angle (in radians) between the two spectra, where each is an m-dimensional feature vector, with m being the number of spectral channels. Small angles indicate a high similarity between pixel and reference spectra. For an image $\mathbf{I}$, the spectral angle $\theta(\overrightarrow{\mathbf{x}})$, for $\overrightarrow{\mathbf{x}} \in \mathbf{I}$, is given by

$$\theta(\overrightarrow{\mathbf{x}}) = \cos^{-1}\left(\frac{f(\lambda) \cdot e(\lambda)}{||f(\lambda)|| \cdot ||e(\lambda)||}\right) , \tag{3.1}$$

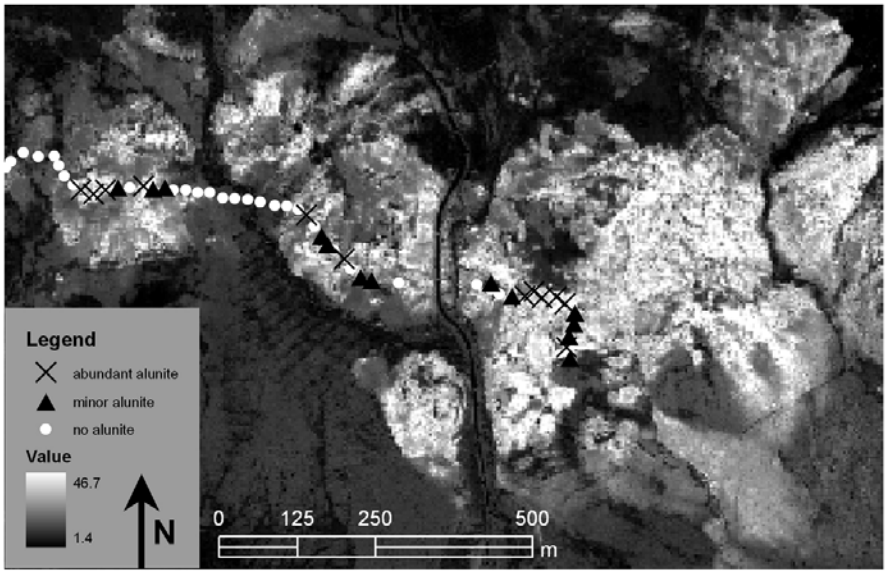
where λ is the wavelength range of the m spectral channels, $f(\lambda)$ is an unclassified m-dimensional image reflectance spectrum under observation and $e(\lambda)$ is an m-dimensional reference spectrum. SAM is directionally dependent, but independent of the length of the spectral vector, thus insensitive to illumination or albedo effects (Crósta et al., 1998). It is also dependent on the user-specified threshold and wavelength range. The result of using Equation 3.1 are grayscale images (SAM's Rule Images), one for each reference mineral, with DN value representing the angular distance in radians between each pixel spectrum and the reference mineral spectrum (see Figure 3.3(a)). Darker pixels in the rule image indicate greater similarity to the reference mineral spectrum. Further, if this angular distance is smaller than a user specified threshold, the pixel is assigned to the category of the respective reference mineral, leading to image classification. This algorithm has been implemented in ENVI$^{\text{TM}}$ image analysis commercial software.

3.3.2 Spectral Feature Fitting (SFF)

SFF matches the image pixel reflectance spectrum to reference spectrum from either a spectral library or a field spectrum by examining specific absorption features in the



(a) SAM classification rule image for alunite. Dark areas indicate smaller angles, hence, greater similarity to alunite. This figure also shows the location of the field data.



(b) SFF fit image for alunite. Lighter areas indicate better fit values between pixel reflectance spectra and the alunite reference spectrum. This figure also shows the location of the field data.

Figure 3.3: SAM and SFF (fit) Rule Images.

spectrum after continuum removal has been applied to both the image and reference spectra (Clark et al., 1991). Denote the continuum for the image reflectance spectrum as $c_f(\lambda)$ and for the reference spectrum as $c_e(\lambda)$. The continuum is removed (Clark & Roush, 1984) using

$$\begin{array}{rcl} e_c(\lambda) & = & e(\lambda)/c_e(\lambda) \\ f_c(\lambda) & = & f(\lambda)/c_f(\lambda) \end{array} \tag{3.2}$$

where $e_c(\lambda)$ is the continuum removed reference spectrum and $f_c(\lambda)$ is the continuum removed image reflectance spectrum. The resulting normalized spectrum reflect levels equal to 1.0 if continuum and the spectrum match and less than 1.0 in the case of absorption.

Similarly, the absorption feature depth is defined as

$$\begin{array}{rcl} D[e_c(\lambda)] & = & 1 - e_c(\lambda) = 1 - e(\lambda)/c_e(\lambda) \\ D[f_c(\lambda)] & = & 1 - f_c(\lambda) = 1 - f(\lambda)/c_f(\lambda) \end{array} \tag{3.3}$$

for each spectrum. The absorption feature depth has a unique magnitude and location, both depending on the mineral and its chemical composition.

Scaling is usually necessary for reference spectra because absorption features in library data typically have greater depth than image reflectance spectra. A simple scaling function of the form $e_c^s(\lambda) = a_0 + a_1 e_c(\lambda)$ is useful, where $e_c^s(\lambda)$ is the modified continuum removed reference spectrum that best matches the image spectrum. For an image $\mathbf{I}$, the scale $\tau_S(\overrightarrow{\mathbf{x}})$, for $\overrightarrow{\mathbf{x}} \in \mathbf{I}$, is determined using least squares that gives the best fit to the image spectrum $f_c(\lambda)$

$$D[f_c(\lambda)] = a + \tau_S(\overrightarrow{\mathbf{x}}) D[e_c(\lambda)] \ . \tag{3.4}$$

Hence the scale image, produced for each reference mineral, is the image of scaling factors used to fit the unknown image spectra to the reference spectrum. The result is a grayscale scale image, whose DN value corresponds to $\tau_S(\overrightarrow{\mathbf{x}})$.

The total root-mean-squares (RMS) errors, $\tau_E(\overrightarrow{\mathbf{x}})$, was defined as

$$\tau_E(\overrightarrow{\mathbf{x}}) = \sqrt{\frac{1}{m} \sum_b \left(D[f_c(\lambda_b)] - D[e_c^s(\lambda_b)] \right)^2} \tag{3.5}$$

where λ_b denotes the wavelength of channel b, $b = 1, \ldots, m$. The result is a grayscale RMS error image, with DN value corresponding to $\tau_E(\overrightarrow{\mathbf{x}})$.

The fit image equals

$$\tau_F(\overrightarrow{\mathbf{x}}) = \tau_S(\overrightarrow{\mathbf{x}}) / \tau_E(\overrightarrow{\mathbf{x}}) \tag{3.6}$$

providing a measure of how well an image pixel reflectance spectrum match a reference spectrum. A large value of $\tau_F(\overrightarrow{\mathbf{x}})$ corresponds to a good match between the image spectrum and the reference spectrum. The fit values were used as a rule image to

weigh each pixel to a reference mineral, namely alunite (see Figure 3.3(b)). This algorithm has been implemented in ENVI^{TM} image analysis commercial software. Further details on SFF can be found in Clark et al. (1991), Clark et al. (1992), Clark & Swayze (1995) and Clark et al. (2003).

3.3.3 Sampling

Sampling by simulation annealing requires definition of a mathematical objective function, called the fitness function.

Simulated Annealing

Simulated annealing is a random search technique that is analogous to the way in which a metal cools and freezes into a minimum energy crystalline structure, called the annealing process (Aarts & Korst, 1989). It forms the basis of an optimization technique for combinatorial problems by finding the optimal value of a fitness function numerically. In terms of sampling, it starts by randomly selecting a collection of points in an image. A new point in the image is then randomly selected and replaces a randomly selected old point from the current collection. This replacement occurs, based on a probabilistic criterion, if the fitness function decreases, and if the fitness function increases. Hence, the process allows inferior moves. Initially, the probabilistic criterion is high, allowing a large probability of inferior moves. A parameter in the annealing process is then reduced by a factor of 0.95 at each successive step, thereby decreasing the probability of accepting inferior moves until the process stabilizes. The final solution is independent of the initial random selection of points as the process reaches the global optimum.

Consider a two-dimensional image $\mathbf{I}$. Let the collection of all possible sampling schemes with n observations on $\mathbf{I}$ be denoted by $\mathbf{S}^n$. A fitness function $\phi(\mathbf{S}^n) : \mathbf{S}^n \rightarrow \mathbb{R}^+$ that has to be minimized to optimize the sampling scheme must be defined.

Fitness function

The Weighted Means Shortest Distance (WMSD)-criterion is a weighted version of the Minimization of the Mean Shortest Distances (MMSD)-criterion (Van Groenigen et al., 2000b). The fitness function is extended with a location dependent weight function that is scaled to $[0, 1]$, namely, $w(\overrightarrow{\mathbf{x}}) : \mathbf{I} \rightarrow [0, 1]$ by

$$\phi_{\text{WMSD}}(\mathbf{S}^n) = \frac{1}{N} \sum_{\overrightarrow{\mathbf{x}} \in \mathbf{I}} w(\overrightarrow{\mathbf{x}}) \left|\left| \overrightarrow{\mathbf{x}} - W_{\mathbf{S}^n}(\overrightarrow{\mathbf{x}}) \right|\right| \ , \tag{3.7}$$

where $W_{\mathbf{S}^n}(\overrightarrow{\mathbf{x}})$ is the location vector of the sampling point in $\mathbf{S}^n$ nearest to $\overrightarrow{\mathbf{x}}$, N is the number of pixels in the image and $w(\overrightarrow{\mathbf{x}})$ is a weight for the pixel with location

vector $\overrightarrow{\mathbf{x}}$. The weights express knowledge or assumptions about the occurrence of alunite in some parts of the region by controlling the sampling density in these areas. Larger weights result in a higher likelihood of a pixel being selected in the final sampling scheme.

This fitness function also spreads the location of the sampling points over the region classified as alunite. Since these points on the image are georeferenced, they will appropriately serve as target points to be sampled in the field. There will be a high probability that the field sample points suggested are alunite and these points will be spread according to the distribution of alunite as in the classified image. The fitness function chosen should, in addition, be able to achieve the study purpose of obtaining a collection of sampling points in the field that appropriately represent the distribution of the mineral of interest. Three different weight functions were then considered to meet the objectives.

Weight function I: Binary weights are used to indicate whether a pixel has a high probability of alunite being present or not. Using SAM rule image derived by application of Equation 3.1, a threshold θ^t is selected. The weight $w(\overrightarrow{\mathbf{x}})$, for an arbitrary pixel $\overrightarrow{\mathbf{x}} \in \mathbf{I}$, is defined as

$$w(\theta(\overrightarrow{\mathbf{x}})) = \begin{cases} 0\,, & \text{if} \quad \theta(\overrightarrow{\mathbf{x}}) > \theta^t \\ 1\,, & \text{if} \quad \theta(\overrightarrow{\mathbf{x}}) \leq \theta^t \end{cases} \tag{3.8}$$

Weight function I, if used in the fitness function, will be restricted to those pixels with a spectral angle smaller than the chosen threshold. This results in a collection of sample points in the field that corresponds to the alunite classified image. Each pixel being georeferenced, when selected by the algorithm, will be a point to be sampled on the ground.

Weight function II: For Weight function II, scaled weights indicate the degree of a pixel's probability of being classified as alunite and to sample intensively where an abundance of alunite occurs. Using SAM rule image derived by application of Equation 3.1, a threshold θ^t is selected. Pixels with $\theta(\overrightarrow{\mathbf{x}}) > \theta^t$ have zero weight, otherwise the weight is a function of $\theta(\overrightarrow{\mathbf{x}})$. Small spectral angles between image and alunite reference spectrum correspond to a large weight. Here, the weight $w(\overrightarrow{\mathbf{x}})$, for each pixel $\overrightarrow{\mathbf{x}}$, scaled to $[0, 1]$ was used.

$$w(\theta(\overrightarrow{\mathbf{x}})) = \begin{cases} 0\,, & \text{if} \quad \theta(\overrightarrow{\mathbf{x}}) > \theta^t \\ \dfrac{\theta^t - \theta(\overrightarrow{\mathbf{x}})}{\theta^t - \theta_{\min}}\,, & \text{if} \quad \theta(\overrightarrow{\mathbf{x}}) \leq \theta^t \end{cases} \tag{3.9}$$

where $\theta_{\min}$ is the minimum spectral angle occurring. Weight function II if used in the fitness function will be restricted to those pixels with a spectral angle smaller than the chosen threshold. The probability is largest to select a pixel that is most similar to the alunite reference spectrum. The georeferenced location of each pixel chosen by the algorithm in the final sampling scheme will be a point to be sampled on the ground.

Weight function III: For Weight function III, scaled weights are used based on several rule images to guide sampling to areas with a high probability of being alunite and to sample more intensely where an abundance of alunite occurs. Using SAM rule image and SFF rule image, derived, respectively, by application of Equations 3.1 and 3.6, thresholds θ^t and τ_F^t are selected for SAM and SFF, respectively. Pixels exceeding either of these threshold angles receive zero weight, otherwise the weight is a function of the spectral angle and the fit value. Higher weights will emerge from smaller spectral angle between the image pixel reflectance spectrum and reference alunite spectrum, and a larger fit value between these two spectra. Here, the weight $w(\overrightarrow{\mathbf{x}})$, for each pixel $\overrightarrow{\mathbf{x}}$, scaled to $[0, 1]$ was used.

$$w(\theta(\overrightarrow{\mathbf{x}}), \tau_F(\overrightarrow{\mathbf{x}})) = \begin{cases} \kappa_1 w_1(\theta(\overrightarrow{\mathbf{x}})) + \kappa_2 w_2(\tau_F(\overrightarrow{\mathbf{x}}))\,, \\ \qquad \text{if } \theta(\overrightarrow{\mathbf{x}}) \leq \theta^t \text{ and } \tau_F(\overrightarrow{\mathbf{x}}) \geq \tau_F^t \\ 0\,, \quad \text{if otherwise} \end{cases} \tag{3.10}$$

where $0 \leq \kappa_1, \kappa_2 \leq 1$ and $\kappa_1 + \kappa_2 = 1$. The weight for SAM: $w_1(\overrightarrow{\mathbf{x}})$ is defined in Equation 3.9 and the weight for SFF: $w_2(\overrightarrow{\mathbf{x}})$, for each pixel $\overrightarrow{\mathbf{x}}$, scaled to $[0, 1]$ is defined as

$$w_2(\tau_F(\overrightarrow{\mathbf{x}})) = \begin{cases} 0\,, & \text{if} \quad \tau_F(\overrightarrow{\mathbf{x}}) < \tau_F^t \\ \dfrac{\tau_F(\overrightarrow{\mathbf{x}}) - \tau_F^t}{\tau_{F,\max} - \tau_F^t}\,, & \text{if} \quad \tau_F(\overrightarrow{\mathbf{x}}) \geq \tau_F^t \end{cases} \tag{3.11}$$

where τ_F^t is the minimum fit threshold value chosen and $\tau_{F,\max}$ the maximum value.

Weight function III, if used in the fitness function, will be restricted to those pixels with a spectral angle smaller than the threshold chosen and with a fit larger than the chosen threshold. The probability is largest to select a pixel that is most similar to the alunite reference spectrum, in terms of both the angle between these spectra and absorption feature fit. The georeferenced location of each pixel chosen by the algorithm in the final sampling scheme will be a point to be sampled on the ground.

Weight function III (Equation 3.10), is based on two rule images. This can easily be extended to more than two rule images, by using different proportions κ_i for each rule image i conditional on $\sum \kappa_i = 1$. Also, in terms of the method of SFF, several absorption features could be considered for a particular mineral, producing a fit image for each feature. These images could be combined in the same way, thereby increasing the weights of image pixels having a spectrum similar to the mineral. This in effect increases the probability of the mineral being selected in the sampling scheme.

3.4 Results

For each of the three weight functions, 40 samples were arbitrarily specified to illustrate the distribution of these points for the proposed sampling scheme. Prior

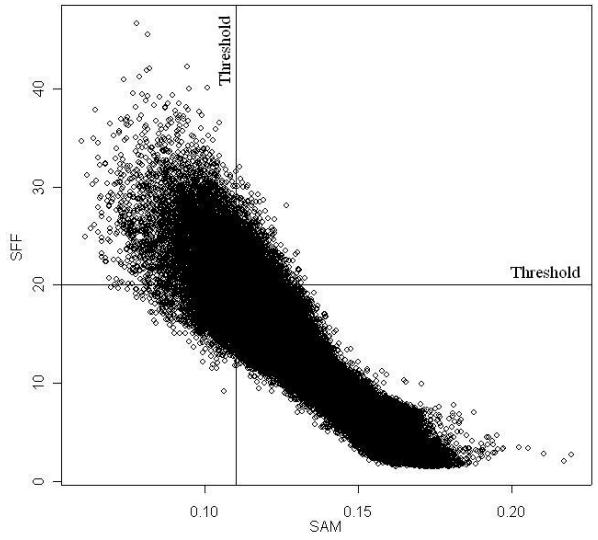


Figure 3.4: Scatter plot of values in rule images obtained through SAM and SFF and the respective thresholds chosen to represent similarity or fit to alunite.

to sampling, isolated segments (< 10 pixels) were removed. This was performed as there was a high chance that they were a result of noise in the image and it seemed impractical to sample in such small areas. However, if these are meaningful targets, with very high probability of alunite, the above procedure can be performed without removal of these pixels.

Weight function I: SAM was applied to each pixel in the image for the alunite reference spectrum, resulting in a rule image (Figure 3.3(a)). By observing the individual spectral reflectances of several pixels and choosing different thresholds for θ^t it was decided to use a threshold of 0.11 radians. The values of SAM rule image for alunite can be seen with respect to the horizontal axis of Figure 3.4. Pixels less than 0.11 radians have weight one, otherwise zero. The binary weights are given in Figure 3.5. The prospective sampling points resulting from using Weight function I are also displayed in Figure 3.5. Sampling intensity is incorrectly represented because the prospective sampling points are spread evenly over the distribution of alunite.

Weight function II: The DN values, $\theta(\overrightarrow{\mathbf{x}})$, from SAM rule image in Figure 3.3(a) were used in Equation 3.9 to obtain scaled weights. A threshold, $\theta^t = 0.11$ radians, was used. Pixels lying left of the 0.11 threshold (Figure 3.4) correspond to positive weights. The resulting scaled weights are shown in Figure 3.6, corresponding to a greater similarity to alunite reference spectrum. The prospective sampling points result from using Weight function II are also displayed in Figure 3.6. The sample points are distributed over the alunite region and most

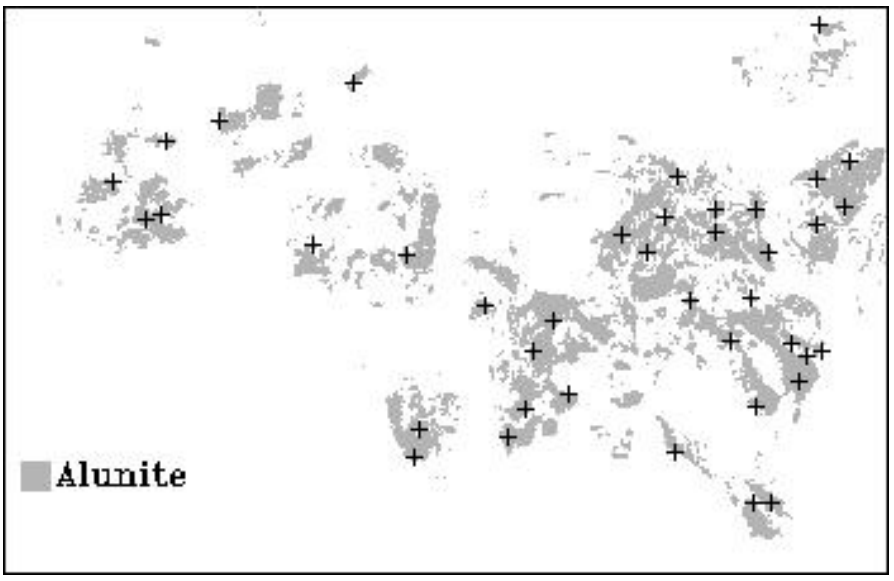


Figure 3.5: Weight function I: SAM classified image based on a threshold of 0.11 radians for alunite. Distribution of 40 sampling points.

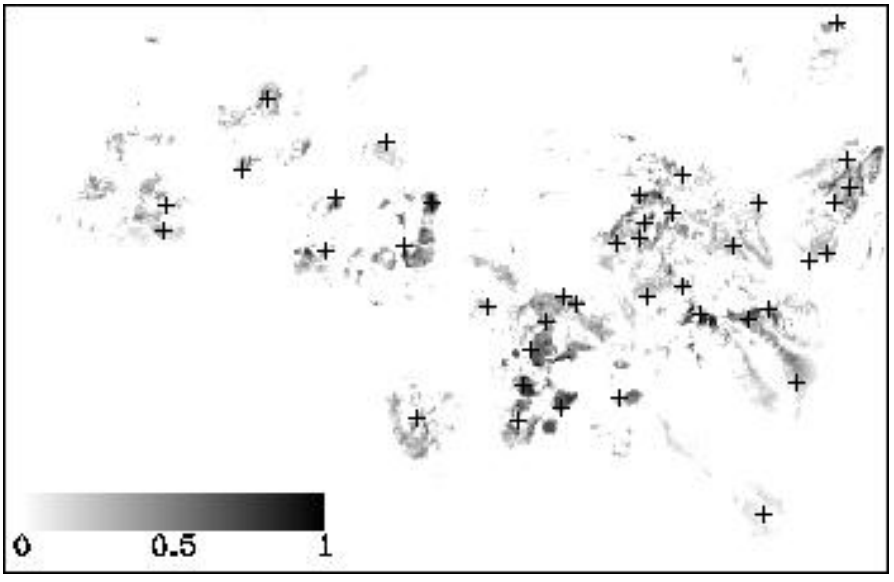


Figure 3.6: Weight function II: SAM scaled map for alunite using a threshold of 0.11 radians. Distribution of 40 sampling points. Darker areas indicate greater similarity to alunite.

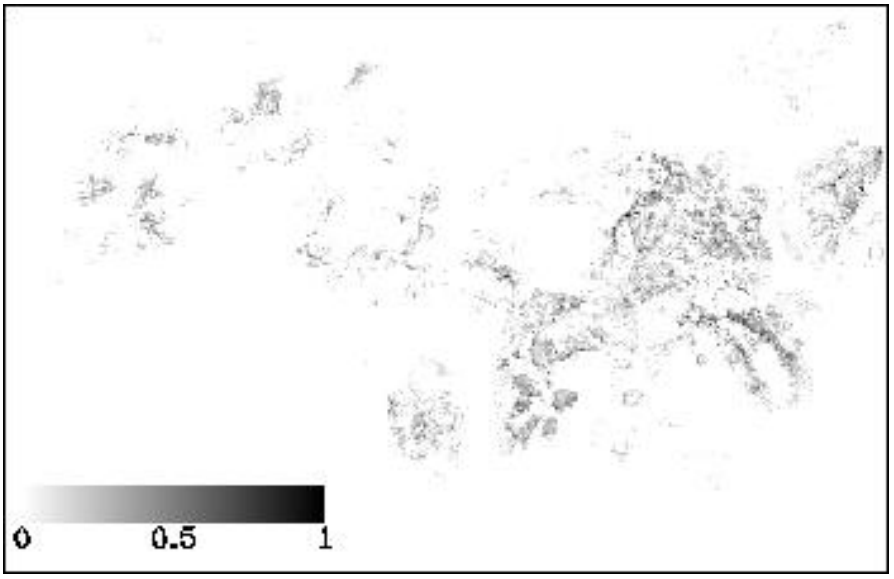


Figure 3.7: SFF scaled map for alunite using a threshold of 20. Darker areas indicate greater similarity to alunite.

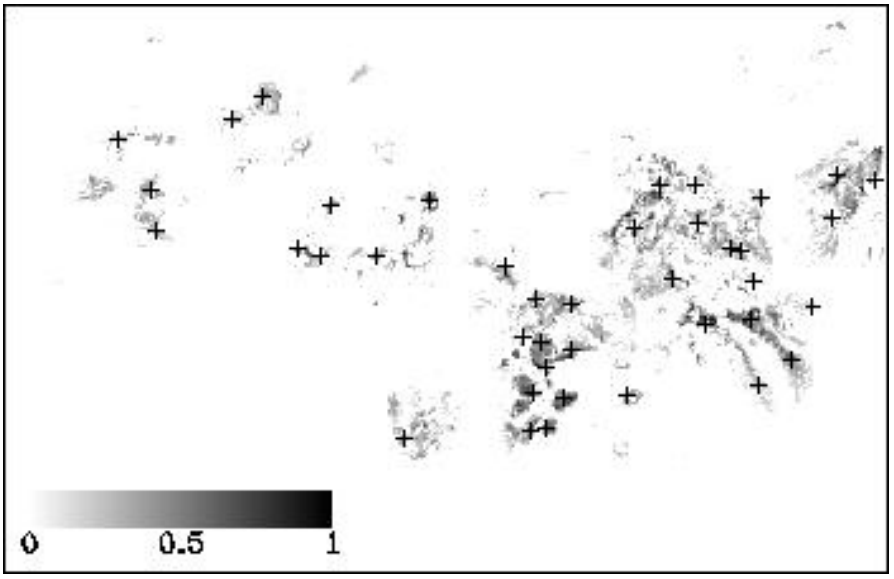


Figure 3.8: Weight function III: Scaled weights derived using SAM and SFF rule images for alunite using their respective thresholds. Distribution of 40 sampling points. Darker areas indicate greater similarity to alunite.

Table 3.2: Weights derived from SAM (column) and SFF (row).

SFF	SAM					
	0.0	(0.0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	(0.6, 0.8]	(0.8, 1.0]
0.0	70603	615	238	84	40	8
(0.0, 0.2]	2546	739	304	126	78	8
(0.2, 0.4]	1183	710	332	111	37	7
(0.4, 0.6]	304	333	156	49	10	1
(0.6, 0.8]	33	42	31	6	1	0
(0.8, 1.0]	4	3	4	4	0	0

Values in the table represent the number (frequency) of pixels that match in a certain range.

points are arranged in the darker areas of the image.

Weight function III: SFF was applied to the alunite reference spectrum, resulting in a scale image and an RMS error image. The ratio of these images, produces a fit image (Figure 3.3(b)). The bright pixels represent the best fit to the alunite reference spectrum. The DN values from the fit image, $\tau_F(\overrightarrow{\mathbf{x}})$, was used in Equation 3.11 to obtain the weights for SFF using a threshold value of 20 for τ_F^t. This threshold was chosen after individual spectral analysis of some pixels and selecting several thresholds. The values of the rule images of SAM and SFF can be seen in Figure 3.4. Pixels in the upper left quadrant correspond to positive weights. In Equation 3.10, $\kappa_1 = \kappa_2 = \frac{1}{2}$ was used. The resulting weights scaled to $[0, 1]$ are shown in Figure 3.7. Darker areas have higher weights and hence greater similarity to the alunite reference spectrum. SAM rule image for alunite was used to obtain weights for SAM by the procedure described by the previous weight function. Table 3.2 summaries the weights derived by SAM and the weights derived by SFF. From the first row and first column, 6.5% of the pixels receive zero weight from one classification but weights larger than zero from the other classification. This can also be seen in Figure 3.4 corresponding to the pixels in the upper right and lower left quadrants. These weights were then combined using Equation 3.10 and are displayed in Figure 3.8. Darker areas have higher weights and hence greater similarity to alunite reference spectrum in terms of both SAM and SFF. The prospective sampling points resulting from using Weight function III are also displayed in Figure 3.8. The sample points are again distributed over the alunite areas and most of the points are found in the darker areas of the image.

Sensitivity analysis

A sensitivity analysis of the proposed sampling schemes was performed using different thresholds, namely, $\theta^t = 0.10$ & 0.12 radians for SAM and $\tau_F^t = 18$ & 22 for SFF, and using $m = 10$ and 15 spectral channels, as a subset of the original 30 channels.

Changing the threshold, produce slightly different rule images. SAM and SFF also produce different rule images when different channels are selected. For these different possibilities and applying each weight function separately, it was desired that the prospective sampling points result in accurate representation of the overall distribution of alunite and the location of these points to be consistent. For $m = 15$, every alternative channel was chosen starting from the first of the original 30 channels, and for $m = 10$, every third channel was selected starting from the second of the original 30 channels. Each prospective sampling scheme derived from a change of threshold and input channels was compared to the corresponding initial prospective sampling scheme as illustrated in Figures 3.5, 3.6 and 3.8 to determine the robustness of each of the weight functions used. For each prospective sampling scheme derived, the average and the maximum distance (in m) to the nearest sampling point from the initial prospective sampling scheme were calculated. Table 3.3 contains the results for these comparisons. The effect of choosing different thresholds and channels varied most, in terms of location of these sampling points, when the sampling scheme was derived from binary weights. This was reduced when scaled weights were used. Combining SAM and SFF weights, resulted in the most robust collection of prospective sampling points, in terms of location, over the alunite region.

For each prospective sampling scheme, the method applied involves quadratic kernel smoothing of the 40 samples as point pattern with a kernel width parameter of 25 (Berman & Diggle, 1989; Rowlingson & Diggle, 1993). The distribution of the prospective sample points for each scheme is illustrated in Figure 3.9. Nearby prospective sampling points effectively produce darker patches in the images, thereby indicating greater abundance of alunite. These images are comparable to the rule images (Figure 3.3) or the classified image using SAM (Figure 3.5). This means that, for most prospective sampling schemes, the distribution of the sample points corresponds closely to the distribution of the alunite. For the binary weights, the intensity of alunite occurring is incorrectly represented. In this respect, scaled weights produce slight improvements.

From the classified SAM image, 40 sampling points were selected randomly (see Figure 3.9, the plot in the second row, last column). The distribution of alunite is incorrectly represented. Furthermore, a different set of random samples will produce different results. Samples with the highest weights derived from Weight functions II (see Figure 3.9, the plot in the fifth row, last column) and III (see Figure 3.9, the plot in the eighth row, last column) were also selected. The distribution of samples in these two sets of prospective sampling schemes can be seen in the same figure. Almost all prospective sample points are restricted to one area in the image. Hence by using the highest weights, prospective samples in the field will often be limited to a small area.

Validation

Ground data collected using an ASD fieldspec-pro spectrometer were used to support the derived prospective sampling schemes by validating the SAM classified image and

Table 3.3: Sensitivity analysis of sampling schemes.

Weight function			Channels 10	15	30
I	0.10 radians	mean	40.8	34.0	41.9
		max	124.3	104.3	119.9
	0.11 radians	mean	52.1	37.9	•
		max	141.4	102.4	•
	0.12 radians	mean	53.2	41.1	47.0
		max	144.2	120.9	124.2
II	0.10 radians	mean	28.4	25.1	28.4
		max	89.4	54.4	115.4
	0.11 radians	mean	39.8	32.7	•
		max	138.7	119.3	•
	0.12 radians	mean	46.3	36.1	44.2
		max	155.6	134.4	144.2
III	0.10 radians	mean	27.3	22.8	20.8
	& 22 fit	max	88.1	62.6	70.9
	0.11 radians	mean	35.2	23.2	•
	& 20 fit	max	103.8	80.0	•
	0.12 radians	mean	33.2	27.7	33.8
	& 18 fit	max	102.0	83.2	109.8

Values represent the mean and maximum distance (in meters) between sampling points in each new sampling scheme to the nearest sampling point in the corresponding initial sampling scheme. The proposed sampling scheme for each of the three weight functions is denoted by a •.

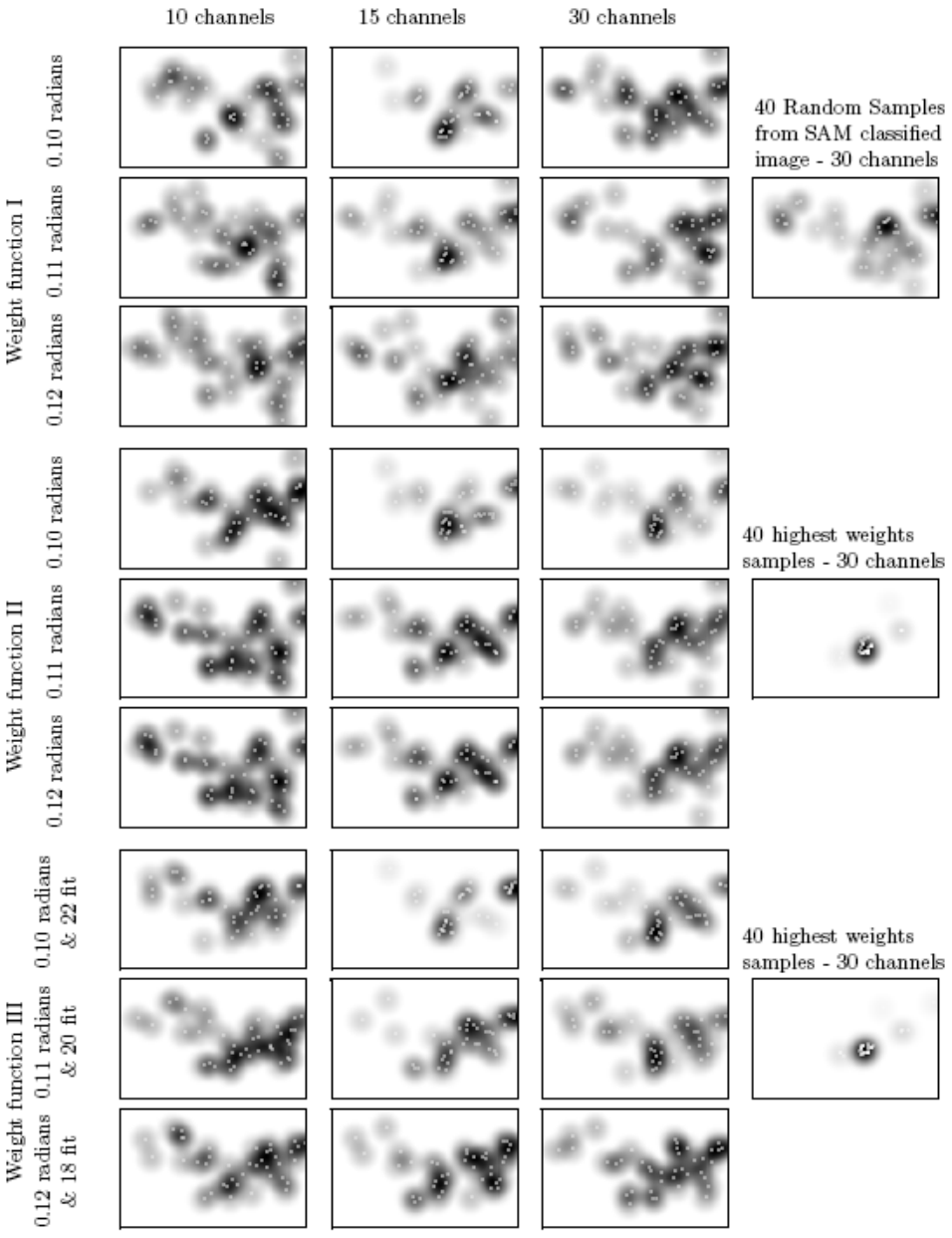


Figure 3.9: Sensitivity Analysis Plots. Darker patches in the images indicate sampling points are near to each other. This effectively implies greater abundance of alunite.

Table 3.4: Validation using ground data.

		Weight	Ground data		
			No alunite	Minor alunite	Abundant alunite
Image data	Weight function I	0.0	26	7	4
		1.0	0	6	8
	Weight function II	0.0	26	7	4
		(0.0, 0.2]	0	3	2
		(0.2, 0.5]	0	3	5
		(0.5, 1.0]	0	0	1
	Weight function III	0.0	26	8	7
		(0.0, 0.2]	0	2	1
		(0.2, 0.5]	0	2	3
		(0.5, 1.0]	0	1	1

Values in the table represent the number (frequency) of ground data that match image data.

the images of the weights used. Reflectance spectra of 51 ground measurements (see Figure 3.3) were analyzed individually for their alunite content and classified into one of three classes, namely, "no alunite", "minor alunite" and "abundant alunite". Table 3.4 summarizes the validation results for each of the three images used in deriving the prospective sampling scheme. Using the ground data of those pixels classified as alunite or not, the accuracy of SAM is $100 \times (26 + 6 + 8)/51 = 78\%$. Table 3.4 also shows that the selected thresholds were conservative since all pixels above the threshold contain alunite.

Forty samples were also selected randomly from the entire image and then from SAM classification image and their the average values were compared to the average values of the prospective samples obtained by using Weight functions I, II and III (Table 3.5). The randomly selected samples over the entire image were poor, with average values for each weight function low, average SAM value high and average SFF value low as compared to the samples randomly selected from the classified image and to the samples in the optimum prospective sampling schemes. The samples in the optimum prospective sampling schemes performed best, with highest average values for each weight function, lowest average SAM value and highest average fit value. Comparing the three optimal sampling schemes, the sampling scheme derived from Weight function II is an improvement to that derived from Weight function I since it has a lower average SAM value. The sampling scheme derived from Weight function III performed best since it has a higher average weight compared to that of Weight function II, lowest average SAM and highest average SFF value.

Table 3.5: Validation of sampling pixels.

	Image used	Average pixel value of 40 samples		
		Random: from image	Random: from SAM classified image	Optimum sampling scheme
Weight function I	Weights	0.07	1	1
	SAM (rad)	0.139	0.101	0.099
Weight function II	Weights	0.01	0.18	0.29
	SAM (rad)	—	—	0.095
Weight function III	Weights	0.01	0.16	0.33
	SAM (rad)	—	—	0.093
	SFF (fit)	11.0	22.4	25.5

3.5 Discussion

Deriving optimum prospective sampling schemes that target areas with high probability and having greater abundance of alunite occurring was demonstrated by using three different weight functions for the WMSD-criterion as an objective function in simulated annealing. Predefined weights allow distinction between areas with different priorities. Hence, sampling can be focused in areas with a high potential for the occurrence of a mineral of interest and reduces sampling in areas with low potential. This could effectively reduce time and costs in the field. Randomly selecting points in the image, as potential sites to sample on the ground, could result in the location of these samples clustered and/or having a low probability of being alunite (Table 3.5). Figure 3.9 shows that the abundance of alunite is incorrectly represented when 40 random sample points within the classified region of alunite were selected. These samples are not optimal as indicated by the average values of SAM, SFF and the weights derived from these images (Table 3.5). Selecting a collection of sampling locations that have the highest probability of being alunite could result in the location of most sampling points clustered in the image (Figure 3.9). This implies sampling in a limited area on the ground, and effectively these samples will not represent the overall distribution of alunite over the entire study area. In the proposed sampling schemes there is a balance between selecting samples that have a high probable alunite and the location of samples not to be clustered in the field. A good sampling scheme will target areas with high probability of alunite and the distribution of sample points will correspond closely to the distribution of alunite. This means intensive sampling in the area with an abundance of alunite.

For Weight function I, binary weights were obtained from a classified image derived using SAM. These weights guided sampling to points in the image that were classified as alunite. This sampling scheme depends upon the classified image using SAM and,

in turn, upon the chosen threshold (Table 3.3). This, in effect, is a limiting factor of the sampling design. In practice, the sampling scheme is designed before fieldwork and hence no training data is available for selecting appropriate thresholds. The threshold was selected prior to analyzing the field data and was purposely chosen to be conservative, so that it is unlikely to classify a pixel as alunite when in fact the pixel represents some other mineral on the ground. This can be seen in Table 3.4, where no pixels other than alunite has the possibility of being selected as a point to be sampled in the field. This was important for this study as it was preferred not to incorrectly sample minerals other than alunite.

The prospective sampling scheme (Figure 3.5) resulting from Weight function I shows a more or less even distribution of 40 sampling points for alunite in the area. A shortfall of using binary weights is that, once the image is classified, each pixel in the region has an equal chance of being selected and the sampling points will be arranged in a way that any three points form roughly equilateral triangles depending on the complexity of the classified region. If the mineral is spatially concentrated in homogeneous areas, the location of the sample points obtained from Weight function I will be most appropriate. In most cases, however, the distribution of a mineral on the ground could be dispersed, hence the result using binary weights were improved by defining scaled weights to target high probability areas of being classified as alunite. This effect is observable in Table 3.5 as the average pixel value for SAM is lower for the sampling scheme derived when using Weight function II as opposed to Weight function I.

Weight functions II and III each use the image derived from the respective scaled weight functions, based on hyperspectral data, to guide the location of sample points to those pixels with a high likelihood of alunite. Weight function II was derived from the SAM rule image for alunite. The same threshold of 0.11 radians was used as in Weight function I. The threshold chosen in this case can be set higher to include some pixels with a reflectance spectrium similar to that of other minerals, example kaolinite and pyrophyllite. This is not considered to be a major problem, as the scaled weights used by the optimal prospective sampling scheme will be low, thereby reducing the probability of selecting that pixel's location as a point to be sampled on the ground. This effect can be seen in the validation results in Table 3.4, where each pixel is given a different weight depending on how similar the image reflectance spectrum is to alunite spectrum. Use of a high threshold in Weight function I, however, can result in selection of pixels representing presence of minerals other than alunite. This is another disadvantage of using binary weights in the sampling scheme. The results using scaled weights are promising, as more sample points are in areas with high probability of alunite (dark areas in the image) and some samples are close to each other, but not clustered, implying more sample points in the region with an abundance of alunite. These results are quite valid, as one would sample more intensively in these areas. The effect of using different thresholds and selecting different channels has been reduced as compared to using binary weights (Table 3.3). The distribution of alunite is also more accurately portrayed in the image (Figure 3.9). The major disadvantage of Weight function II is that these scaled weights are derived solely from

the SAM results and other useful information contributing to classification of a pixel are not taken into account. Introducing another classification provides additional information, and the procedure to combine the information uses scaled weights, with higher weights corresponding to areas of higher probability of alunite.

Weight function III uses two rule images, one derived from SAM and another from SFF. A comparison of the scaled weights derived from SAM and SFF (Table 3.2), indicates that the methods for SAM and SFF do not always agree. This can also be seen in Table 3.4. Only the purest pixels classified as alunite have positive weights. The advantage of combining SAM and SFF classification methods in the weights function results in a classified image that is more robust for the thresholds and selected channels. The weights derived from SAM and from SFF were then combined into a single weight image, which was used for the design of the optimal sampling scheme. In terms of robustness of the thresholds and channels selected, the prospective sampling scheme shows improvements to both the binary weights used and the scaled weights derived solely from the rule image of SAM (Table 3.3). A suitable range for the thresholds has to be known. This can be obtained by observing individual spectra and the purest of these can be selected to train the thresholds. Using the combined weights from SAM and SFF, sample points can be concentrated in the region with a high probability of alunite, which are more robust against the thresholds selected. The distribution of sample points corresponds closely to the distribution of alunite (Figure 3.9). Weight function III also produces the best sampling result, with highest average weight, lowest average SAM values and highest average SFF values (Table 3.5).

The prospective sampling schemes derived are of interest to (a) exploration geologists for specified target locations of hydrothermally altered minerals (e.g. alunite) with distinct absorption features, (b) researchers trying to understand the geothermal system and hydrothermal zones in a specific region and (c) engineers to better collect field data in relation to flights by improving on ground truthing and calibration measurements. With the aid of new spaceborne launched hyperspectral sensors, e.g. Hyperion and ARIES-1, data are available for most regions and hence will be helpful to geologist's planning phase of selecting important mineral targets in the field. The methods presented here could result in reduction of time and effort in the field, but by no means replace the field geologist. It is merely an aid for target selection of minerals as an initial survey, followed by denser surface sampling of interesting anomalies.

3.6 Conclusions

This study resulted into three main conclusions.

- Combination of SAM and SFF rule images results in robust weights to focus sampling in areas of high probability of alunite. The resulting sample scheme also produces the best sampling result, with highest average weight, lowest

average SAM and highest average SFF values. Prospective sample points are arranged more intensely in areas with an abundance of alunite.

- SAM and SFF both lead to a relevant classification of the study area with respect to alunite, as observed from the rule images and validation of the rule images using ground measurements.
- A sensitivity analysis showed that prospective sample points derived from the combination of SAM and SFF classification are more stable against changes in thresholds and channels selected. Moreover, distribution of the location of these prospective samples in space corresponds closely to the intensity of alunite.

Chapter 4

Optimization of exploration targets on mineral prospectivity maps

Prospecting is not what it once was. Unless you want to walk the same ground and not find anything either, you have to be prepared to prospect differently than the Old-Timers did.

Clyde H. Spencer

This chapter is based on P. Debba, E. J. M. Carranza, A. Stein and F. D. van der Meer (**In review**) Optimum allocation of exploration targets on mineral prospectivity maps, *Mathematical Geology*.[1]

[1]This work was sponsored by ITC International Institute for Geo-Information Science and Earth Observation, project number 3083022 and NRF National Research Foundation, project number 10317, gun 2053944.

Abstract

This chapter describes a quantitative method for optimally locating exploration targets based on a probabilistic mineral prospectivity map, which was created by means of weights-of-evidence (WofE) modeling. Locations of discovered mineral occurrences were used as a training set and a map of distances to faults/fractures and three channel ratio images of HyMap hyperspectral data were used as evidences. The WofE posterior probability map was input to an objective function that optimized location of exploration targets. The method was applied to the Rodalquilar mineral district (SE Spain). Optimized exploration target zones spatially coincide with undiscovered mineral occurrences, namely, those not used to train the WofE model input, and show other zones without mineral occurrences within delineated prospective ground. The results indicate usefulness of the described optimization method to allocate exploration targets for undiscovered mineral occurrences based on probabilistic mineral prospectivity maps.

Keywords

Simulated annealing, epithermal mineralization, weights-of-evidence, hyperspectral remote sensing, hydrothermal alteration

4.1 Introduction

In mineral districts, locations of mineral occurrences are available in the form of mines and prospects. These locations are used for training in data-driven predictive mapping of prospective ground. There are several ways of doing so. Particular possibilities include the weights-of-evidence (WofE) method (Agterberg et al., 1990; Bonham-Carter et al., 1988), linear regression (Chung & Agterberg, 1980), logistic regression (Agterberg & Bonham-Carter, 1999), canonical favorability analysis (Pan, 1993), neural networks (Porwal et al., 2003; Rigol-Sanchez et al., 2003) and evidential belief functions (Carranza & Hale, 2003). Mineral prospectivity maps are then usually used to guide further mineral exploration. A logical question regarding efficacy of mineral prospectivity maps is: "Where should targets of exploration for undiscovered mineral occurrences be focussed?" This question is likely to be answered on the basis of any of the predictive mapping methods.

To provide a plausible answer to this question in a case study, a mineral prospectivity map was first created by employing the WofE method. The results were then used to define foci of exploration targets. WofE modeling is based on a Bayesian probability framework to update the prior probability of mineral occurrence per unit cell or pixel in a study area, given a set of geological evidences spatially associated with the mineral occurrences. This leads to the posterior probability of mineral occurrence per unit cell or pixel in that area. The WofE method is easily implementable in common Geographic Information System (GIS) software. It is postulated, however, that the proposed method of locating foci of exploration targets is adaptable to using other methods for mineral prospectivity mapping.

The proposed methodology aims to derive the optimal allocation scheme to set exploration targets on mineral prospectivity maps, whereby each target represents a focal zone that coincides spatially with or is proximal to undiscovered mineral occurrences. Derivation of the optimal allocation scheme is done with simulated annealing. Previous studies of simulated annealing applications to obtain optimal sampling schemes to guide sampling to target specific areas, involved stratification (Van Groenigen et al., 2000a), using thresholds (Van Groenigen et al., 2000a), defining a weight function (Van Groenigen et al., 2000b) and using ordinary kriging (Shieh et al., 2005). Studies on the design of optimal sampling schemes in combination with remote sensing used fuzzy classification (Tapia et al., 2005) and targeting a particular intense hydrothermal alteration mineral (Debba et al., 2005b). These are typically model based sampling approaches (de Gruijter & ter Braak, 1990), which depend upon a model derived from explanatory variables. In this study, the optimal allocation scheme was derived based on a mineral prospectivity model to indicate targets for further exploration. Ideally, each target should (a) be in zones of high posterior probability (or other indices) of mineral occurrence and (b) provide opportunity for discovery of mineral occurrences. The proposed method was tested and demonstrated within the Rodaquilar mineral district, where several epithermal gold occurrences are present.

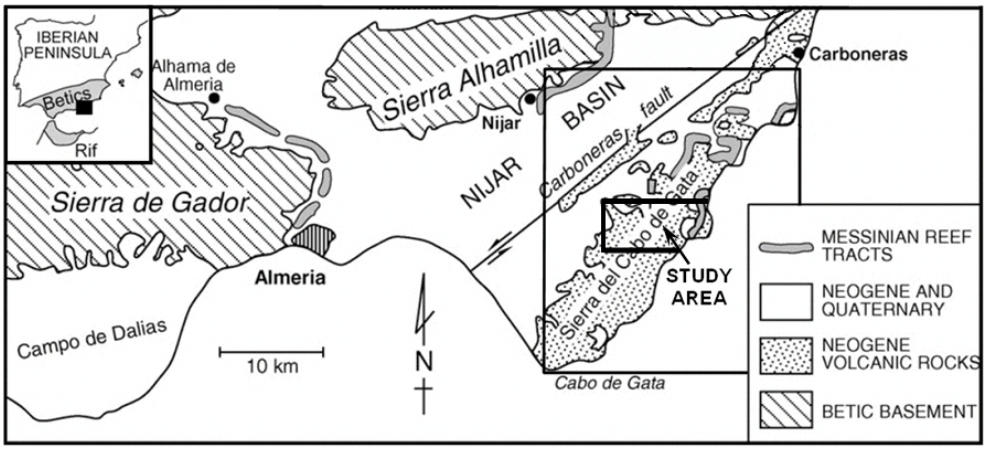


Figure 4.1: A generalized geological map of the Rodalquilar area mineral district.

4.2 Study area

4.2.1 Geology and mineralization of the Rodalquilar mineral district

The Rodalquilar mineral district is located in the Sierra del Cabo de Gata volcanic field, in the south-eastern part of Spain (Figure 4.1), consisting of pyroxene andesites to rhyolites of the late Tertiary age. Extensive hydrothermal alteration of the volcanic rocks resulted in formation of high to low temperature minerals as: silica → alunite → kaolinite → illite → chlorite. Occurrences of epithermal precious- and base-metals are in veins or in hydrothermal breccias (i.e. fracture controlled) associated with hydrothermally altered rocks (Arribas et al., 1995). High sulphidation precious-metal occurrences are associated with advanced argillic (alunite±kaolinite) and intermediate argillic (kaolinite±illite) zones, whereas low sulphidation precious- and base-metal occurrences are associated with intermediate argillic to pyropylitic (illite±chlorite) zones (Arribas et al., 1995). The epithermal mineral occurrences have been localized along faults and fractures that cut through the volcanic host rocks. Based on these generalized geological characteristics of discovered epithermal mineral occurrences in the district, two recognition criteria were applied for mapping zones with high potential of epithermal mineral occurrence, (1) hydrothermal alteration evidence and (2) structural evidence.

4.2.2 Data for hydrothermal alteration evidence

A sub-scene, consisting of 2640×1300 pixels, of airborne imaging spectrometer data was used. The data were acquired by the Hyperspectral Mapper (HyMAP) in July 2003 during the HyEUROPE 2003 campaign over the study area and its vicinity. HyMAP is a 126-channel instrument that collects data in a cross-track direction by

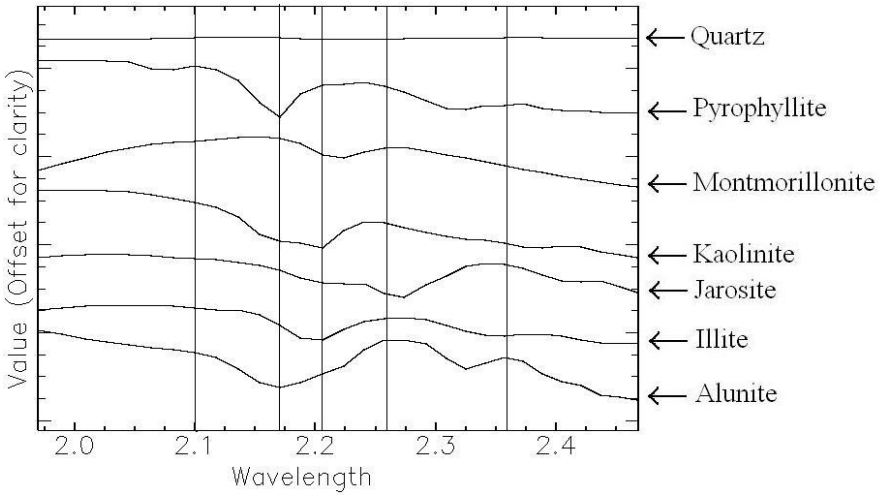


Figure 4.2: Plot of 7 end-members from USGS spectral library (Clark et al., 1993) in the spectral range 1.95–2.48 μm. Vertical lines indicate the channel centers used to obtain channel ratio images (see text for further information).

mechanical scanning and in an along-track direction by movement of the airborne platform. The instrument acts as an imaging spectrometer of the reflected solar radiation within the 0.4–2.5 μm wavelength region of the electromagnetic spectrum. Spectral coverage is nearly continuous in the visible-to-near-infrared (VNIR) and shortwave-infrared (SWIR) regions with small gaps in the middle of the 1.4 and 1.9 μm atmospheric water absorption bands. The spatial configuration of the instrument accounts for an instantaneous-field-of-view (IFOV) of 2.5 mrad along track and 2.0 mrad across track resulting in a pixel size in the order of 3–5 m for the data used in this study. Due to instrument failure, the SWIR 1 detector did not function during acquisition, thus no data were acquired in the 1.50–1.76 μm window. The HyMAP data were atmospherically and geometrically corrected using the Atmospheric and Topographic Correction (ATCOR 4) model (Richter, 1996).

Data acquired by the SWIR 2 detector (bandwidth 16 nm), within the 1.95–2.48 μm spectral range are potentially useful for mapping hydrothermal alteration assemblages as well as for regolith characterization (Abrams et al., 1977; Goetz & Srivastava, 1985; Cudahy et al., 2000; Papp & Cudahy, 2002; Kruse, 2002). This spectral region covers the most prominent spectral absorption features of hydroxyl-bearing minerals, sulfates and carbonates, which are common to many geologic units and hydrothermal alteration assemblages (Kruse, 2002).

Figure 4.2 shows plots of spectra of the seven most prominent alteration minerals in the study area (Arribas et al., 1995), at spectral intervals coinciding with the HyMAP SWIR 2 data. This figure shows differences in absorption features of the different minerals, in terms of shape, size, symmetry, depth and position. With the

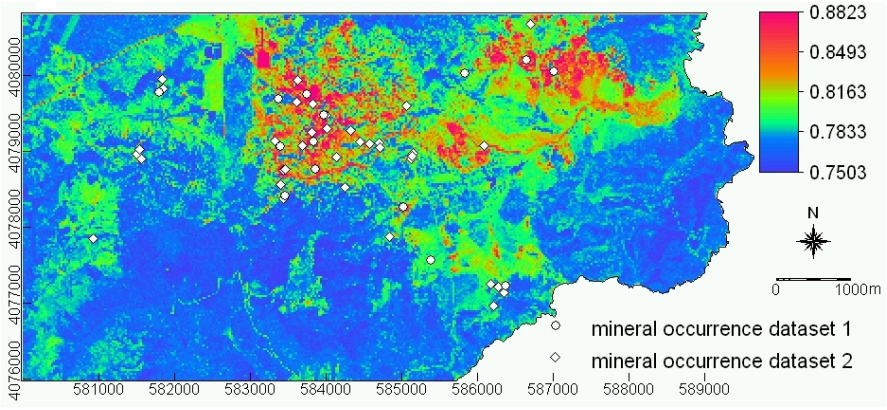
exception of the quartz spectrum, all the other spectra have distinctive absorption features at wavelengths of approximately 2.2 μm, although each absorption feature differ slightly in position and depth. To delineate predominant minerals in the alteration zones associated with the epithermal occurrences, channel ratio images (Lillesand et al., 1994) were created using the channels as indicated in Figure 4.2, namely, channels 103/107 (2.100/2.171 μm), channels 107/109 (2.171/2.205 μm) and channels 118/112 (2.357/2.258 μm). Channel ratioing is a form of enhancing material of interest from spectral images by dividing spectral data in a channel with spectral data in another channel. Images derived from channel ratioing can convey information, due to spectral properties of surface minerals, independent on variations in scene illumination. An arctan transformation was applied to the channel ratios (Lillesand et al., 1994), which considers the gradient of spectral data between two channels. Figure 4.3 displays the respective images of channel ratios to be used as input evidence layers in WofE modeling. Pixels in image of channel ratio 1 (2.100/2.171μm) are red/orange (i.e. higher ratios) for alunite, kaolinite and pyrophyllite but green (i.e. lower ratios) for illite (Figure 4.3(a)). The first three minerals are predominant in the advanced argillic zones. Pixels in image of channel ratio 2 (2.171/2.205μm) are red/orange for illite and kaolinite but are green for alunite and pyrophyllite (Figure 4.3(b)). The red/orange pixels in the image for channel ratio 2 thus enhance predominant minerals associated with intermediate argillic zones. Pixels in image of channel ratio 3 (2.357/2.258μm) are blue/green for minerals predominant in advanced argillic zones but red/orange for minerals predominant in intermediate argillic to pyropylitic zones (Figure 4.3(c)).

4.2.3 Data for structural evidence

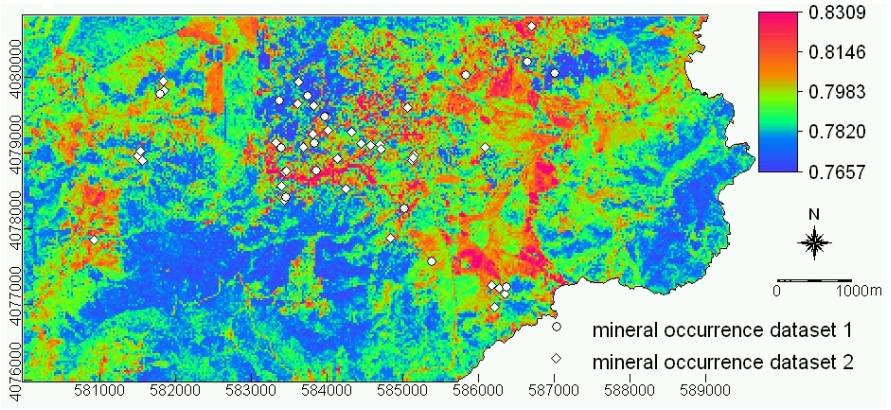
Mapped faults and fractures were digitized from georeferenced published (IGME, 1981) and unpublished reports. In addition, faults and fractures were interpreted and digitized on shaded-relief images of digital elevation model (DEM) derived from Advanced Spaceborne and Thermal Emission Radiometer (ASTER) data acquired on 26 May 2002. A map of distances to mapped and interpreted faults and fractures was then created (Figure 4.4) and used in WofE modeling.

4.2.4 Data for training of WofE model and allocation of exploration targets

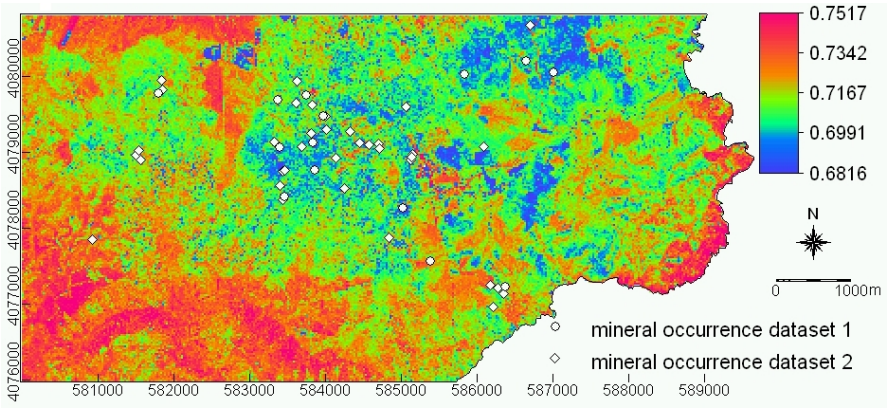
Two sets of mineral occurrence data were used in WofE modeling. One set, of 14 epithermal occurrences, was digitized from a 1:50,000 scale geological map of Spain (IGME, 1981). The other set, of 47 epithermal occurrences, was digitized from the mineral prospectivity map of Rigol-Sanchez et al. (2003), which actually shows 49 epithermal occurrences although two of these fall outside the study area. In this latter set, 11 epithermal occurrences were discarded because each of them lie within 100 m of an epithermal occurrence in the first set, which indicates a high likelihood that those 11 in the second set are the same as 11 of the 14 in the first set. Thus, the



(a) Channel Ratio 1: arctan transformation on channels 103/107 (2.100/2.171).



(b) Channel Ratio 2: arctan transformation on channels 107/109 (2.171/2.205).



(c) Channel Ratio 3: arctan transformation on channels 118/112 (2.357/2.258).

Figure 4.3: Input layers for WofE modeling. Map coordinates are in meters (UTM projection, zone 30N).

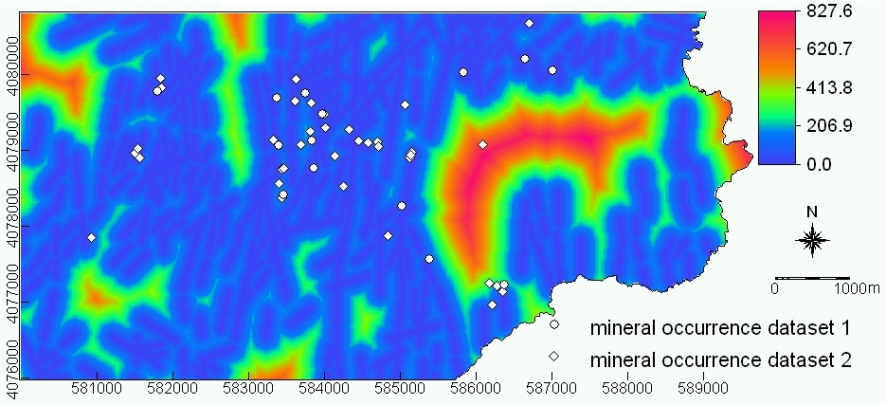


Figure 4.4: Distance to fault and fracture. Pixels ranging from blue to red in this image indicates increasing distance from a fault or fracture. Input layer for WofE modeling. Map coordinates are in meters (UTM projection, zone 30N).

second set has 36 epithermal occurrences, each of which is believed to be different from the 14 epithermal occurrences in the first set in terms location, but not in terms of mineralization type. Each of the two sets of epithermal occurrence data were then used for training and for cross-validation of a WofE model. A training set is assumed to represent discovered mineral occurrences, whereas a cross-validation set is assumed to represent undiscovered mineral occurrences. Prediction rate (discussed later) was the criterion applied to select the better of the two WofE models created as input for allocating exploration targets.

4.3 Methods

To obtain the optimal schemes of exploration targets, the initial step was to create an image with distances to faults and fractures and three channel ratio images. These four images were resampled to 25 m pixel size (sub-scene 389 × 193 pixels), corresponding to the pixel or unit cell representation of point locations of epithermal precious- and base-metal occurrences in WofE modeling using a raster-based GIS software. Next, the images are used as evidences in WofE modeling to produce a posterior probability map. Further, this posterior probability map was thresholded and used as a weight in an objective function. Finally, the objective function is optimized to derive optimal locations of exploration targets. The objective function also considers the uncertainty of the posterior probability. The aim of optimizing the objective function is to spread foci of exploration targets over the district such that they are represented in the posterior probability map by pixels with high probability of being proximal to a discovered or undiscovered mineral occurrence. Ideally, such exploration targets should be spatially coincident or proximal to undiscovered mineral

occurrences from the cross-validation set, which was not used to train a WofE model.

4.3.1 WofE Method

WofE modeling (Agterberg et al., 1990; Bonham-Carter et al., 1988) is a Bayesian method that combines information from multiple evidences to predict the occurrence of a binary pattern. Each mineral occurrence is treated as a binary object, being either present or absent, within a small unit cell or pixel. In mineral prospectivity mapping, each evidence has either a positive or negative spatial association with a set of discovered mineral occurrences. It thereby either increases or decreases the posterior probability for mineral occurrences at unvisited locations. The posterior probability is an index of degree for mineral occurrence.

Let D represent discovered mineral occurrence, with one and only one occurrence in a unit cell and let $P(D)$ be an estimate of prior probability of mineral occurrence. Further, let B^t denote a binary pattern for an evidence, thresholded at t, which is a spatial data attribute (e.g. ratio or distance) initially arbitrarily chosen. Definition of conditional probability results in:

$$P(D|B^t) = P(D) \cdot \frac{P(B^t|D)}{P(B^t)} \,, \tag{4.1}$$

where $P(D|B^t)$ is the posterior probability of mineral occurrence, given the presence of the binary pattern. Similarly, the posterior probability of mineral occurrence, given the absence of the binary pattern, $\overline{B^t}$ can be defined.

In WofE, the posterior probability is converted to posterior odds ratio ($O(A) = P(A)/P(\overline{A})$ for any event A), by dividing both sides of Equation 4.1 by $P(\overline{D}|B^t)$ and simplifying by replacing $P(\overline{D}|B^t) \cdot P(B^t) = P(B^t|\overline{D}) \cdot P(\overline{D})$, yielding

$$O(D|B^t) = O(D) \cdot \frac{P(B^t|D)}{P(B^t|\overline{D})} \,, \tag{4.2}$$

where $O(D|B^t)$ is the posterior odds of D given B^t and $O(D)$ is the prior odds of D. Taking the natural logarithm on both sides of Equation 4.2, results in

$$\ln O(D|B^t) = \ln O(D) + W^+ \,, \tag{4.3}$$

where $W^+ = \ln \dfrac{P(B^t|D)}{P(B^t|\overline{D})}$ is the weight of evidence for the presence of B^t. Similarly, using the posterior probability of mineral occurrence given the absence of B^t, the natural logarithm of posterior odds of mineral occurrence is defined as

$$\ln O(D|\overline{B^t}) = \ln O(D) + W^- \,, \tag{4.4}$$

where $W^- = \ln \dfrac{P(\overline{B^t}|D)}{P(\overline{B^t}|\overline{D})}$ is the weight of evidence for absence of B^t.

Statistical significance of the weights can be determined based on their variances, which are approximated from Bishop et al. (1975) as

$$\begin{aligned} s^2(W^+) &= \frac{1}{N(B^t \cap D)} + \frac{1}{N(B^t \cap \overline{D})} \quad \text{and} \\ s^2(W^-) &= \frac{1}{N(\overline{B^t} \cap D)} + \frac{1}{N(\overline{B^t} \cap \overline{D})} \,, \end{aligned} \tag{4.5}$$

where $N(\cdot)$ denotes the number of counts, for example, $N(B^t \cap D)$ is the number of mineral occurrences in the presence of binary pattern B^t. Once the weights W^+ and W^- are determined from Equations 4.3 and 4.4 for evidences B^t at several different thresholds, the maximum contrast, $C = W^+ - W^-$ usually gives the optimum threshold value. If the number of mineral occurrence is small, the maximum studentized contrast, $C/s(C)$ can be used, where $s(C)$ is the standard deviation of the contrast values. A studentized contrast > 1.96 (Bonham-Carter, 1994) is a useful criteria for optimum thresholds to create binary predictor maps, B_i. The binary predictor maps are then used to determine the posterior probability of mineral occurrence. For k sets of evidences, resulting in $B_1, B_2, \ldots, B_k$ binary predictor maps,

$$\begin{aligned} P\left(D|B_1, \ldots, B_k\right) &= \frac{P\left(B_1, \ldots, B_k \,|D\right) \cdot P(D)}{P(B_1, \ldots, B_k)} \\ &= \frac{P\left(B_1, \ldots, B_k \,|D\right) \cdot P(D)}{P\left(B_1, \ldots, B_k|D\right) \cdot P(D) + P\left(B_1, \ldots, B_k|\overline{D}\right) \cdot P(\overline{D})} \,. \end{aligned} \tag{4.6}$$

Equation 4.6 allows estimation of the posterior probability of mineral occurrence given presence of binary evidences. Because of their interaction, the terms $P\left(B_1, \ldots, B_k \,|D\right)$ and $P\left(B_1, \ldots, B_k|\overline{D}\right)$ are difficult to estimate, unless conditional independence (CI) is assumed. Assuming CI, then $P\left(B_1, \ldots, B_k \,|D\right) = \prod_{i=1}^{k} P(B_i|D)$. A similar expression applies for the second term in the denominator of Equation 4.6. With k binary predictor maps, 2^k possible combinations exist, depending on whether binary predictor map pattern B_i is present or not. This also means that there are 2^k unique conditions in the posterior probability map, being equivalent to 2^k polygons or grid cells in which the same combination of evidence occurs. After assuming CI in Equation 4.6 and some simplifications, formulation of posterior odds of mineral occurrence, given k binary predictor maps, is obtained as

$$\ln O_j\left(D \mid B_1, \ldots, B_k\right) = \ln O(D) + \sum_{i=1}^{k} W_i^j \,, \tag{4.7}$$

where W_i^j denotes weights (W_i^+ or W_i^-) contributed by spatial evidence from binary predictor map B_i ($i = 1, 2, \ldots, k$) to the jth unique condition ($j = 1, 2, \ldots, 2^k$). The posterior probabilities are then obtained from the posterior odds using

$$P_j = P_j(D|B_1, \ldots, B_k) = \frac{O_j(D|B_1, \ldots, B_k)}{1 + O_j(D|B_1, \ldots, B_k)} \,. \tag{4.8}$$

The variance of the posterior odds is

$$s^2(O) = \sum_{i=1}^{k} s^2(W_i) \,, \tag{4.9}$$

where $s^2(W_i)$ is either $s^2(W_i^+)$ or $s^2(W_i^-)$ (defined by Equation 4.5 for each binary predictor map B_i) depending on whether the binary predictor map B_i is present or not.

The images of channel ratios and the map of distances to faults and fractures were optimized to binary predictor maps B_i by finding the optimum threshold, which maximizes positive spatial associations of these evidential data with the target variable D as an indicator of mineral occurrence. In the thresholded map, the corresponding values of W^+ and W^- were then assigned to the pattern indicating presence or absence, respectively. The binary predictor maps were then combined using Equation 4.7 and the posterior probability was then estimated using Equation 4.8.

WofE modeling assumes CI among the evidence maps with respect to a set of mineral occurrences. Violation of this assumption causes the posterior probabilities to be either over- or under-estimated. Assumption of CI was tested using the new omnibus test (NOT) (Agterberg & Cheng, 2002; Thiart et al., 2004). The NOT compares the number of training mineral occurrences $N(D)$ to the number of predicted training mineral occurrences $N(D)_{\text{pred}}$, where

$$N(D)_{\text{pred}} = \sum_{j=1}^{2^k} P_j \{N(A)\}_j \tag{4.10}$$

and $\{N(A)\}_j$ is the area in unit cells for the jth unique condition. The test statistic (Agterberg & Cheng, 2002; Thiart et al., 2004) under the null hypothesis H_0 : $N(D)_{\text{pred}} = N(D)$ is

$$\text{NOT} = \frac{N(D)_{\text{pred}} - N(D)}{s[N(D)_{\text{pred}}]} \,, \tag{4.11}$$

where the variance of the predicted number of training mineral occurrences $s^2[N(D)_{\text{pred}}]$ is estimated by

$$s^2[N(D)_{\text{pred}}] = \sum_{j=1}^{2^k} [\{N(A)\}_j]^2 \times s^2(P_j) \,, \tag{4.12}$$

and variance of P_j is estimated based on variance of the weights (Bonham-Carter et al., 1989) by

$$s^2(P_j) = \frac{1}{N(D)} + \sum_{i=1}^{k} s^2(W_i^j) \times P_k^2 \,. \tag{4.13}$$

Values of NOT are assumed to approximate the standard Gaussian distribution and the hypothesis will be rejected in favor of $H_1 : N(D)_{\text{pred}} > N(D)$ for a statistically larger difference.

An integrated model showing non-violation of CI assumption was then used to create a posterior probability map. The prediction rate of a posterior probability map was estimated as the proportion of predicted undiscovered occurrence in a cross-validation set that corresponds with posterior probability greater than the prior probability in a WofE model created with a training set of discovered mineral occurrences. By interchanging the roles of the two sets of mineral occurrences as training and cross-validation, the WofE model with the higher prediction rate is chosen as input into the proposed method for allocating exploration targets.

4.3.2 Scheme for allocating exploration targets

The number of exploration targets has to be chosen in advance of allocating these targets.

Number of exploration targets

Each exploration target is a composite of adjoining unit cells where mineral occurrence can be investigated further by an appropriate field sampling technique. Since an exploration target may or may not contain at least one undiscovered mineral occurrence, the number of exploration targets is invariably greater but can be equal to the number of undiscovered mineral occurrence. Mostly expert-driven methods for estimation of the latter is discussed in Singer (1993). Here, a data-driven procedure is proposed to estimate the number of exploration targets.

Each exploration target is represented by a focal point or unit cell. To estimate the number of exploration targets, the binomial distribution is employed because mineral occurrence is considered to be a binary variable, being either present or absent. Thus, estimation of n exploration targets to yield at least r mineral occurrences, with a probability of success p, at a 95% confidence, requires solving the following equation for n:

$$\sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 0.95\,. \tag{4.14}$$

The obtained values of n are used in the allocation scheme of exploration targets, which is derived by simulation annealing. This requires definition of an objective function, called the fitness function.

Simulated Annealing of exploration targets

For a two-dimensional region A divided into $N(A)$ unit cells, let the spatial configurations of n exploration targets be denoted by $\mathbf{S}^n$. Denote the posterior probability of a mineral occurrence in a unit cell in A derived from WofE modeling, and thresholded by values less than the prior probability, by $P(\overrightarrow{\mathbf{x}}) = \{P_j(\overrightarrow{\mathbf{x}}) | \overrightarrow{\mathbf{x}} \in A\}$, where $\overrightarrow{\mathbf{x}}$ is the location vector of the unit cell in A, with a corresponding pixel in an image I, for unique condition j. A fitness function $\phi(\mathbf{S}^n) : \mathbf{S}^n \rightarrow \mathbb{R}^+$ that has to be minimized to optimize the allocation scheme is an extension to the Weighted Means Shortest Distance (WMSD)-criterion (Debba et al., 2005b; Van Groenigen et al., 2000b).

$$\phi_{\mathrm{WMSD+V}}(\mathbf{S}^n) = \frac{\lambda}{N(A)} \sum_{\overrightarrow{\mathbf{x}} \in A} p(\overrightarrow{\mathbf{x}}) \left|\left| \overrightarrow{\mathbf{x}} - Q_{\mathbf{S}^n}(\overrightarrow{\mathbf{x}}) \right|\right| + (1-\lambda) s^2(O_{\mathbf{S}^n}) \ , \qquad (4.15)$$

where $Q_{\mathbf{S}^n}(\overrightarrow{\mathbf{x}})$ is the location vector of an exploration target in $\mathbf{S}^n$ nearest to $\overrightarrow{\mathbf{x}}$, $p(\overrightarrow{\mathbf{x}})$ is the posterior probability for a unit cell with location vector $\overrightarrow{\mathbf{x}}$, $s^2(O_{\mathbf{S}^n})$ is the variance of the posterior odds (Equation 4.9) under the current allocation scheme $\mathbf{S}^n$ and $\lambda \in [0, 1]$ is a constant controlling the effect of the two terms. The posterior probabilities express the knowledge or assumptions about mineral occurrence in some parts of the region A by controlling density of mineral occurrence in these areas. Larger posterior probabilities result in a higher chance of a unit cell being selected in the final allocation scheme. The variance of the posterior odds controls allocation of exploration targets to areas with a high positive spatial association between evidential patterns and discovered mineral occurrences.

4.4 Results

WofE modeling of mineral prospectivity

The study area consists of 65253 unit cells of 25×25 m. Each location of mineral occurrence in training set 1 (with 14 epithermal occurrences) was buffered to a minimum of 25 m, which increased the number of pixels to 70, because of their representation in IGME (1981). Estimate of $P(D)$ based on training set 1 is 0.00107, whereas estimate of $P(D)$ based on training set 2 (with 36 epithermal occurrences) is 0.00055. Table 4.1 shows the results of WofE modeling to create binary predictor patterns using the sets of hydrothermal alteration evidence and structural evidence with respect to either set of epithermal occurrences.

Zones with high values of channel ratio 1 (CR1) and of channel ratio 2 (CR2) have positive spatial associations with epithermal occurrences in either set of training data. Positive spatial association between zones with high values of CR1 and epithermal occurrences is stronger than positive spatial association between zones with high values of CR2 and epithermal occurrences as indicated by magnitude of W^+ and C. Zones with high values of channel ratio 3 (CR3) have negative spatial association

Table 4.1: Results of WofE calculations for binary predictor patterns based on range of spatial data attributes (in brackets under column 1) having optimum spatial associations with the training epithermal occurrences.

Binary predictor patterns[a]	N_B^b	N_{DB}^c	W^+	$s(W^+)$	W^-	$s(W^-)$	C	st.C
Using training set 1 ($N(D) = 14$ epithermal occurrences) for WofE modeling[d]								
CR1 (≥ 0.80)	13038	46	1.19	0.15	-0.85	0.20	2.04	8.10
CR2 (≥ 0.79)	32509	44	0.23	0.15	-0.30	0.20	0.53	2.16
CR3 (≥ 0.71)	52290	38	-0.56	0.18	1.01	0.16	-1.57	-5.97
DFF (≤ 70 m)	24800	48	0.59	0.14	-0.68	0.21	1.27	4.55
Using training set 2 ($N(D) = 36$ epithermal occurrences) for WofE modeling								
CR1 (≥ 0.81)	9819	19	1.26	0.23	-0.59	0.24	1.84	5.52
CR2 (≥ 0.79)	22791	19	0.41	0.23	-0.32	0.24	0.73	2.20
CR3 (≥ 0.70)	55419	21	-0.38	0.22	1.02	0.26	-1.39	-4.12
DFF (≤ 170 m)	45396	32	0.25	0.18	-1.01	0.50	1.25	2.36

[a] Values in brackets indicate attributes of spatial data within pattern representing presence of binary evidence.

[b] Values in this column refer to number of pixels within pattern representing presence of binary evidence, i.e. $N(B)$

[c] Values in this column refer to number of pixels of training data within pattern representing presence of binary evidence i.e. $N(D \cap B)$.

[d] Each location of epithermal occurrence was buffered to 25 m, which increased number of training pixels from 14 to 70.

with epithermal occurrences in either set of training data. These results are consistent with field observations, as most epithermal occurrences in the area are associated with intermediate argillic to advanced argillic alteration zones while some are associated with argillic to propylitic zones.

Spatial association between faults and fractures (DFF) and the 14 epithermal occurrences in the first training set is positive and optimal at 70 m. Spatial association between faults and fractures and the 36 epithermal occurrences in the second training set is also positive but optimal at 170 m. These results suggest that epithermal occurrences in training set 1 are mostly vein-type rather than disseminated-type, whereas in training set 2 there is a higher proportion of disseminated-type epithermal occurrences than in the first set. The pattern of type and strength of spatial association (as indicated by C or Studentized C) between the spatial evidences and the epithermal occurrences, however, is the same. This indicates that the epithermal occurrences in either set of training data have similar geological characterics, which means that a mineral prospectivity map derived through WofE modeling using each training set can be used to predict a large proportion of epithermal occurrence in the other set. The similar results for training set 1 and training set 2 also indicate that minimal buffering

Table 4.2: Results of tests of CI assumption based on NOT.

Predictor map combination	NOT value	p(NOT)	CI test
Integrated models based on training set 1 ($N(D) = 14$ epithermal occurrences) for WofE modeling			
CR1-CR2-CR3	1.77	0.038	Fail
CR1-CR2-DFF	0.49	0.312	Pass
CR1-CR3-DFF	2.09	0.018	Fail
CR2-CR3-DFF	1.00	0.159	Pass
CR1-CR2-CR3-DFF	2.24	0.012	Fail
Integrated models based on training set 2 ($N(D) = 36$ epithermal occurrences) for WofE modeling			
CR1-CR2-CR3	1.34	0.090	Pass
CR1-CR2-DFF	0.35	0.363	Pass
CR1-CR3-DFF	0.85	0.198	Pass
CR2-CR3-DFF	0.17	0.432	Pass
CR1-CR2-CR3-DFF	1.37	0.085	Pass

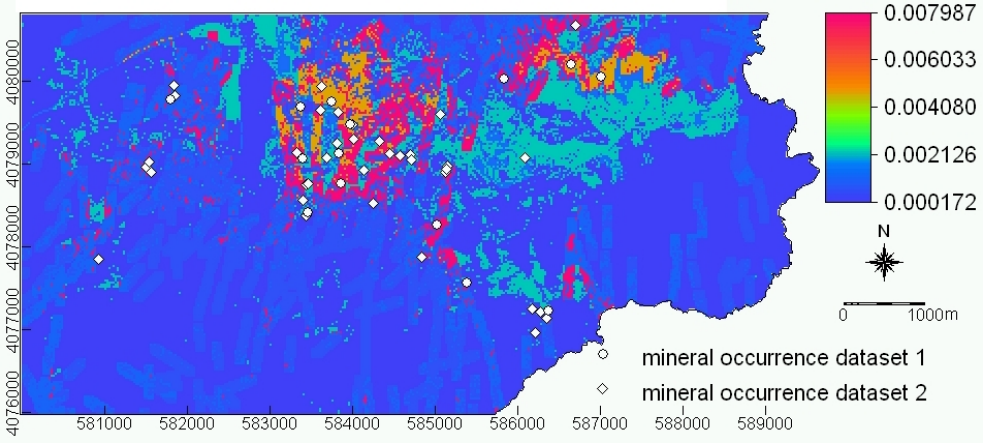
of mineral occurrences in training set 1 was sufficient to adequately quantify spatial associations with the evidences. This is reasonable, considering that all epithermal occurrences in the district were formed by similar mineral-forming processes.

Table 4.2 shows results of tests of CI assumption on 3-layer and full 4-layer models of posterior probabilities of epithermal occurrence based on each set of training data.

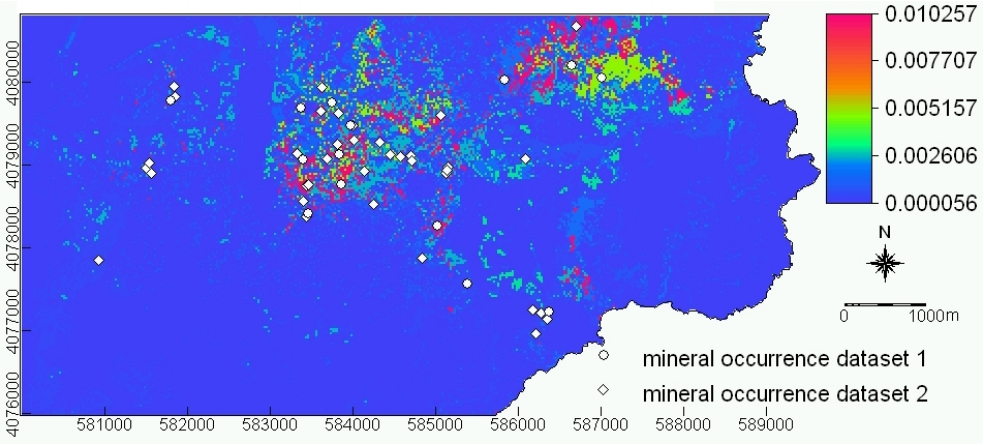
Only two 3-layer models based on training set 1 pass the NOT for CI assumption. The models in which CI assumption is not violated do not involve CR1 and CR3 together, whereas the models in which CI assumption is violated involve CR1 and CR3 together. Violation of CI assumption is mainly due to overlap between positive values in W^+ patterns of CR1 and positive values in W^- patterns of CR3 (Table 4.1), which results in over-estimation of posterior probability. Each of the two 3-layer models, which pass the CI test, can be used mainly to map prospective zones for epithermal occurrences associated with intermediate argillic to advance argillic zones. The CR1-CR2-DFF model has a prediction rate of 0.58, namely, 21 predicted occurrences out of 36 cross-validation occurrences. The CR2-CR3-DFF model has a prediction rate of 0.47, namely, 17 predicted occurrences out of 36 cross-validation occurrences.

All the 3-layer models and the full 4-layer model based on the training set 2 pass the NOT of CI assumption. The models involving CR1 and CR3 together, however, barely pass the NOT. This indicates some degree of conditional dependence between binary predictor patterns of CR1 and CR3, which is mainly due to overlap between positive values in W^+ patterns of CR1 and positive values in W^- patterns of CR3 (Table 4.1). Because all integrated models based on the training set 2 pass the CI test, each can be used to map zones with potential for epithermal occurrence. Only the full 4-layer model was considered further in the analysis because it involves all the

spatial evidences, which suggests its usefulness not only for mapping zones prospective for epithermal occurrences associated with intermediate to advanced argillic zones but also for epithermal occurrences associated with intermediate argillic to propylitic zones. The posterior probability map, representing the CR1-CR2-CR3-DFF model based on the training set 2 of 36 epithermal occurrences, has a prediction rate of 0.64 (i.e. nine predicted occurrences out of 14 cross-validation occurrences), which is better than the prediction rate (see above) of the CR1-CR2-DFF model based on the training set 1 of 14 epithermal occurrences. The posterior probability maps based on the CR1-CR2-DFF model is shown in Figure 4.5(a) and CR1-CR2-CR3-DFF model in Figure 4.5(b).



(a) Using training set 1 ($N(D) = 14$ epithermal occurrences) for WofE modeling.



(b) Using training set 2 ($N(D) = 36$ epithermal occurrences) for WofE modeling.

Figure 4.5: Maps of posterior probability, representing prospectivity of epithermal occurrence.

4.4.1 Optimum allocation scheme of exploration targets

The posterior probability map (Figure 4.5(b)) based on training set 2 was used as input into the allocation scheme. Training set 1 was used as reference for the number of undiscovered epithermal occurrences and for validation of locating the exploration targets result.

Estimated number of exploration targets

To apply Equation 4.14, it was assumed that $r = 9$ based on the nine predicted out of 14 undiscovered epithermal occurrences in training set 1 and $p = 0.0025$ based on the average posterior probabilities greater than the prior probability in the input WofE model. Based on these assumptions, $n = 6280$. Such number of exploration targets is intractable. It is interpreted and shown later that 6280 is approximately the total number of unit cells within plausible exploration target zones. Instead of $p = 0.0025$, $p = 0.6$ was used based on the approximate prediction rate of the input WofE model. Accordingly, $n = 22$, which is a plausible number of exploration targets to search for the nine (assumed) undiscovered epithermal occurrences.

Focal points of exploration targets

Using the posterior probability map shown in Figure 4.5(b) as input, and specification of $n = 22$ and $\lambda = 0.5$ in Equation 4.15, derived focal points of exploration targets are shown in Figure 4.6. Each of the 22 allocated focal points of exploration targets occupies a unit cell with high estimated posterior probability based on training set 2. This indicates that the algorithm was effective in allocating the focal points of exploration targets into prospective ground. The allocated focal points of exploration targets do not fall exactly on but are proximal to a unit cell representing epithermal mineral occurrence belonging to either set of training data. The focal points of exploration targets thus suggest sites where to focus further prospecting.

Exploration target zones

Proximity to an undiscovered occurrence was quantified by utilizing the number of allocations of 6280 unit cells required to delineate the nine predicted occurrences out of the 14 cross-validation occurrences. The total area represented by the 6280 unit cells is approximately $6280 \times 25^2 = 3925000$ m^2. If each of the nine undiscovered occurrences, predicted by the WofE model out of the 14 cross-validation undiscovered occurrences, is within a delineated sub-area of $3925000/22 = 178409$ m^2 containing the allocated focal points of exploration targets, then this indicate that a focal point of exploration target is proximal to at least one undiscovered occurrence. This also means that, if each of the nine predicted undiscovered occurrences, delineated by

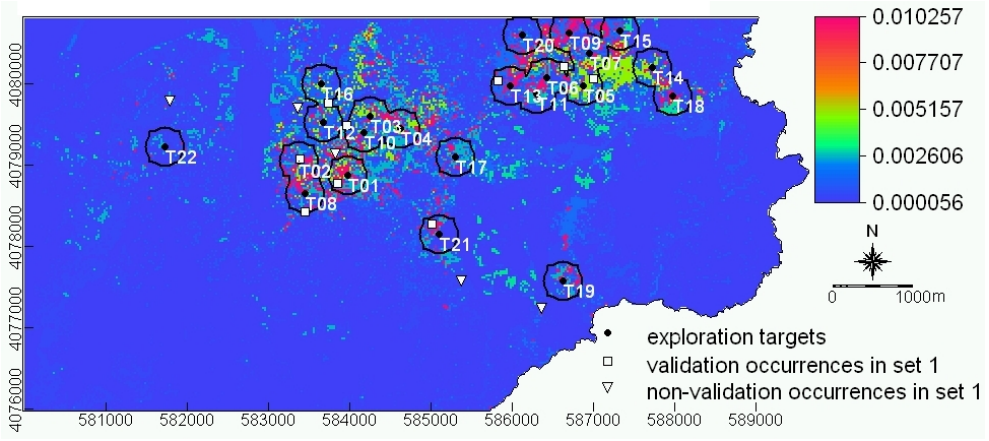


Figure 4.6: Optimized allocation scheme of exploration targets, each target buffered to 238 m (black polylines) for epithermal occurrence.

the WofE model out of the 14 cross-validation undiscovered occurrences, is within a radius of $\sqrt{178409/\pi} = 238$ m (area of circle $= \pi \times \text{radius}^2$) around a focal point of exploration target, then a focal point of exploration target is in close proximity to at least one undiscovered occurrence.

Each of the 22 allocated focal points of exploration targets was thus buffered to a radius of 238 m to delineate exploration zones. Seven of the nine (assumed) undiscovered occurrences, delineated by the WofE model out of the 14 cross-validation undiscovered occurrences, are within the 238 m buffered exploration target zones. The result of this analysis indicates that allocated focal points of exploration targets are proximal to undiscovered epithermal occurrences. The average of posterior probabilities of unit cells occupied by each exploration target is 0.010, which is higher than the average posterior probability (0.0024) of unit cells occupied by discovered epithermal occurrences (training set 2) and the average posterior probability (0.0029) of unit cells occupied by (assumed) undiscovered epithermal occurrences (training set 1). These indicate that the algorithm of the allocation scheme is efficient in targeting unit cells in prospective ground. The results also suggest that buffered exploration target zones are favorable sites for further mineral prospecting. This suggestion is validated below.

4.4.2 Prioritization and validation of exploration targets

In practice, exploration targets are prioritized or ranked according to some criteria. The criteria applied to prioritize each of the 22 buffered focal points of exploration targets are (a) number of unit cells with posterior probability greater than prior probability and (b) average posterior probability of unit cells. The first criterion represents a measure of whether or nor a target zone is wholly made up of prospective cells, whereas the second criterion is an index of mineral occurrence. To each exploration

target, descending ranks of 1 through 22 were assigned according to decreasing values per criterion. The ranks per criterion were then added, which represent a measure of relative prospectivity with low sums indicating higher prospectivity. Hence, descending sums were assigned ascending ranks. Table 4.3 summarizes the priority/rank derived for each exploration target.

To validate the ranked exploration targets, presence of at least one undiscovered occurrence in each 238 m buffer zone of each exploration target was determined. In addition, distance from each exploration target focal point to nearest undiscovered occurrence was determined. Seven buffered exploration targets (T01, T02, T05, T08, T10, T13 and T21) contain at least one (assumed) undiscovered occurrence (Table 4.3). Five of the top 10 priority exploration targets (T01, T02, T05, T08 and T10) contain at least one (assumed) undiscovered occurrence. The 238 m buffer zones of three exploration targets (T06, T12 and T16) are only about 10–30 m away from an undiscovered occurrence. For the other 12 exploration targets, whose 238 m buffer zones do not contain an (assumed) undiscovered occurrence and whose buffer limits are at least 50 m away from an (assumed) undiscovered occurrence, the average distance to a nearest (assumed) undiscovered occurrence is about 600 m. Figure 4.6 also shows that all the nine (assumed) undiscovered occurrences are within or very close ($\sim$15 m on average) to an exploration target zone. The other five undiscovered occurrences not predicted by the WofE model (and thus not assumed undiscovered occurrences in the allocation scheme) are, on average, about 230 m away from the limits of buffered exploration targets. These results indicate that the focal points of exploration targets are adequately positioned such that further mineral prospecting within (and up to a few tens of meters beyond) their 238 m buffer limits could potentially lead to mineral occurrence discovery.

4.5 Discussion

4.5.1 On predictive modeling of mineral prospectivity

Optimum allocation of exploration targets depends on accuracy of an input mineral prospectivity map, which, in turn, depends, on number and accuracy of evidential data sets and, in case of data-driven methods, on number and accuracy of training data used in modeling. In the present work, only four sets of evidential data and two sets of training data were used. Three sets of evidential data used are remotely-sensed information in the form of channel ratio images, the accuracy of each which is reliable based on a number of spectral measurements of ground samples. Several ways exist to use channel ratio images as hydrothermal evidences. For example, principal component scores from several channel ratio bands could be used as evidence maps in WofE modeling. With principal component analysis, however, it is not always easy to judge the hydrothermal alteration assemblages associated with epithermal mineralization. Selecting hyperspectral channels that enhance these features proved to be more intuitive and practical. The accuracy of interpreted faults/fractures, which

Table 4.3: Results of prioritization and validation of allocated exploration targets. Rows in bold and italics, respectively, indicate buffered target zones containing and proximal to (< 30 m) an undiscovered occurrence.

	Prioritization of targets						Validation of targets	
ID[a]	Crit1[b]	R1[c]	Crit2[d]	R2[e]	SR12[f]	Final Rank[g]	Within zone[h]	Dist[i]
T01	**237**	**1**	**0.00512**	**1**	**2**	**1**	**Yes**	**159.7**
T02	**222**	**4**	**0.00395**	**4**	**8**	**2**	**Yes**	**33.9**
T03	231	2	0.00370	7	9	3	No	314.6
T04	216	5	0.00376	6	11	4	No	638.9
T05	**226**	**3**	**0.00353**	**9**	**12**	**5.5**	**Yes**	**150.0**
T06	*199*	*9*	*0.00401*	*3*	*12*	*5.5*	*No*	*266.2*
T07	212	7	0.00358	8	15	7	No	334.0
T08	**191**	**11**	**0.00378**	**5**	**16**	**8.5**	**Yes**	**217.8**
T09	186	14	0.00430	2	16	8.5	No	430.8
T10	**215**	**6**	**0.00311**	**13**	**19**	**10**	**Yes**	**222.6**
T11	187	12.5	0.00323	11	23.5	11	No	464.6
T12	*210*	*8*	*0.00255*	*16*	*24*	*12*	*No*	*261.4*
T13	**185**	**15**	**0.00345**	**10**	**25**	**13**	**Yes**	**164.6**
T14	187	12.5	0.00292	15	27.5	14	No	759.9
T15	172	17	0.00318	12	29	15	No	706.6
T16	*196*	*10*	*0.00189*	*21*	*31*	*16*	*No*	*246.8*
T17	142	19	0.00302	14	33	17	No	905.1
T18	180	16	0.00237	18	34	18	No	1021.2
T19	116	20	0.00245	17	37	19.5	No	421.1
T20	143	18	0.00223	19	37	19.5	No	663.1
T21	**106**	**21**	**0.00195**	**20**	**41**	**21**	**Yes**	**150.0**
T22	47	22	0.00069	22	44	22	No	551.8

[a] Target ID, with numeric characteristics indicating priority.
[b] Values for criterion 1 (= number of unit cells with posterior probability > prior probability within each of the 238 m buffered exploration target).
[c] Descending ranks assigned to decreasing values for criterion 1.
[d] Values for criterion 2 (= average posterior probability of all unit cells within each of the 238 m buffered exploration target).
[e] Descending ranks assigned to decreasing values for criterion 2.
[f] Sum of ranks for criteria 1 & 2.
[g] Ascending ranks assigned to increasing sum of R1 & R2.
[h] Presence of (assumed) undiscovered occurrence within the 238 m buffer zone.
[i] Distance to nearest (assumed) undiscovered occurrence (m).

were combined with published mapped faults/fractures, was also considered reliable based on ground-checking.

The two sets of mineral occurrences data, used for training and cross-validation were also derived from published literature and were considered accurate mainly in terms of their locations but not in terms of mineralization (sub-)type homogeneity. The rather low prediction rate (of 64%) of input probabilistic prospectivity map could be attributable partly to (a) the small number of evidential data sets used and partly to (b) the presence of two (precious- and base-metal) sub-types of epithermal occurrences used in modeling prospectivity. In regard to the latter, relatively homogeneous set(s) of training data (say, for precious-metal epithermal occurrences) could have been prepared by classification of all the epithermal occurrences in the district through application of artificial neural network (Singer & Kouda, 1997, 2003). Required data sets (e.g., mineralogy, grade and tonnages) for typing of mineral occurrence or deposits through neural network classification are, however, incomplete to unavailable. Hence, it was decided to model prospectivity for occurrence of epithermal mineralizations, in general, and considered the 64% prediction rate of one of the WofE models to be adequate in view of small number of evidential data sets. Inclusion of other evidences (e.g. geophysical anomalies) may help to further improve the prediction rate of the WofE modeling.

4.5.2 On exploration target allocation scheme

Locations of exploration targets based on mineral prospectivity maps are, in previous works, determined by expert-driven analysis. That is, portions of delineated prospective ground distal and not containing discovered mineral occurrences are considered new exploration targets. In this study, locations of new exploration targets are determined based on estimated posterior probabilities of mineral occurrence and using an objective function in simulated annealing. The results show that, based on cross-validation mineral occurrence data, allocated focal points of exploration targets derived from the proposed method could provide opportunity for mineral occurrence discovery. Thus, exploration targets based on mineral prospectivity maps can be allocated objectively, rather than subjectively, through application of the proposed method.

Although the suggested exploration target allocation scheme is based upon a posterior probability map, it has potential to accommodate other maps of indices of mineral prospectivity in the range $[0, 1]$, such as fuzzy prospectivity membership values or degrees of evidential belief. It would, however, require other indices of mineral prospectivity not in the range $[0, 1]$ to be transformed into this range. In addition, a threshold index of prospective ground is required (in the present case, this is the prior probability) so that exploration targets are allocated properly to locations with index of prospectivity above this threshold. The graphical technique showed by Porwal et al. (2003) to define objectively a threshold index of prospective ground can be useful for methods of mineral prospectivity modeling (e.g., knowledge-driven fuzzy approaches) that do not involve updating of prior probability of mineral occurrence.

Finally, it requires a measure of uncertainty of the index of mineral prospectivity.

A critical consideration in application of the objective function in simulated annealing to determine optimum locations of exploration targets is the plausible number of exploration targets, which should practically be at least the number of undiscovered occurrences of mineral deposits. Estimation of undiscovered deposits in a geologically-permissive terrane (a) is based on grade-and-tonnage model of deposit-type of interest and (b) is practically performed through consensus by a group of geoscience experts in mineral deposits (Singer & Kouda, 1997). Partly due lack of the first criterion (as epithermal mineralizations according to our database are mostly occurrences and few are deposits or measured reserves) and partly because the second criterion is not met, number of undiscovered deposits in the district was not estimated. Instead, the theorem of binomial distribution was applied to estimate a plausible number of exploration targets based on assumed undiscovered mineral occurrences. Using an estimated number of exploration targets based on actually discovered mineral occurrences, however, does not negate usefulness of the proposed exploration target allocation scheme. Rather, it illustrates that the proposed exploration target allocation scheme is a potential link between predictive modeling of mineral prospectivity and assessment of undiscovered resources. For example, if estimates of number of undiscovered deposits and their corresponding confidence levels are available, then these variables could be used, respectively, in lieu of r and 0.95 in Equation 4.14. For p in Equation 4.14, it is shown here that using prediction rate of mineral prospectivity model results in a plausible number of exploration targets. If, in case, all discovered mineral occurrences are used for modeling of mineral prospectivity, then p in Equation 4.14 could be represented by success rate of mineral prospective model (i.e., proportion of training points delineated in prospective ground). However, in such a case, estimates of number of undiscovered deposits or occurrences and their corresponding confidence levels should be obtained to estimate number of exploration targets.

Understandably, the only convincing way to validate whether or not the proposed method for allocating exploration targets is useful to guide mineral prospecting to undiscovered occurrences is to visit and perform sampling in the delineated and prioritized exploration target zones. A way forward from this work, however, is to test further the proposed method in control areas where works on both mineral prospectivity modeling and assessment of undiscovered mineral resources have been carried out.

4.6 Conclusions

This study resulted in four main conclusions.

- The optimal allocation scheme indicates exploration targets of undiscovered epithermal occurrences. Nine of 14 (assumed) undiscovered epithermal occur-

rences, predicted by WofE modeling, are either within a buffered zone or at most 30 m from a buffered zone of an exploration target. This indicates that the algorithm was able to delineate the nine undiscovered epithermal deposits with reasonable accuracy.

- An objective data-driven way to buffer and then to prioritize predicted foci of exploration targets was demonstrated to produce meaningful results. The results of this method indicate its usefulness because an undiscovered epithermal occurrence is delineated by (a) 60% of the top 5 buffered exploration targets and (b) 50% of the top 10 buffered exploration targets.
- Each of the 22 exploration targets are on prospective ground, with high posterior probability of mineral occurrence. This suggests that new exploration targets located distal to any (actually) discovered epithermal occurrences warrant further investigations in the Rodalquilar district.
- Hyperspectral images provide evidential information for predicting epithermal occurrences in this district, although they need support by other evidences, such as proximity to faults and fractures.

Chapter 5

Abundance estimation of spectrally similar minerals by using derivative spectra in simulated annealing

In solving a problem of this sort, the grand thing is to be able to reason backward. This is a very useful accomplishment, and a very easy one, but people do not practise it much ... Most people, if you describe a train of events to them, will tell you what the result would be. They can put those events together in their minds, and argue from them that something will come to pass. There are a few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backward ...

Sherlock Holmes

This chapter is based on P. Debba, E. J. M. Carranza, F. D. van der Meer and A. Stein (**In review**) Abundance estimation of spectrally similar materials in mine wastes using hyperspectral data and simulated annealing, *IEEE Transactions of Geosciences and Remote Sensing*.[1]

[1]This work was sponsored by ITC International Institute for Geo-Information Science and Earth Observation, project number 3083022 and NRF National Research Foundation, project number 10317, gun 2053944.

Abstract

This chapter presents a method for estimating the partial abundance of spectrally similar minerals in complex mixtures. The method requires formulation of a linear function of individual spectra of individual minerals. The first and second derivatives of each of the different sets of mixed spectra and the individual spectra are determined. The error is minimized by means of simulated annealing. Experiments were made on several different mixtures of selected end-members, which could plausibly occur in real situations. The variance of the differences between the second derivatives of the observed spectrum and the second derivatives of the end-member spectra give most precise estimates for the partial abundance of each end-member. We conclude that the use of second order derivatives provides a valuable contribution to unmixing procedures.

Keywords

Unmixing, simulated annealing, hyperspectral remote sensing, abundance, fraction

5.1 Introduction

Hyperspectral remote sensing technology has proven to be useful in mineral exploration and in monitoring environmental of mining-induced impact. Usefulness of this technology stems from the fact that abundance of various indicator materials on the ground, such as hydrothermal and/or secondary acid-generating minerals, can be characterized by their respective spectral signature. An important step in characterization of surface materials by means of hyperspectral image analysis is spectral unmixing, aiming to determine abundances of some set of materials that contribute to the observed spectrum. At the pixel scale of hyperspectral imagery, it is postulated that the observed spectrum can be modeled as a linear combination of the products of the individual material reflectance functions and their surface areas (Singer & McCord, 1979). With this assumption, most spectral unmixing techniques are variants of algorithms involving matrix inversion (Boardman, 1989; Shimabukuro & Smith, 1991; Boardman et al., 1995; García-Haro et al., 1996; Metternicht & Fermont, 1998; Ferrier, 1999; Drake et al., 1999; Van der Meer & De Jong, 2000; Robinson et al., 2000). Two major problems, however, exists in spectral unmixing: (1) selection of end-members; (2) non-orthogonality of end-members. Theories behind each of these are becoming well-established, as well as intrinsic difficulties of the mathematical solutions to the problems of end-member selection (Smith et al., 1985; Green et al., 1988; Boardman, 1993; Craig, 1994; Jimenez & Landgrebe, 1999) and of non-orthogonality of end-members (Boardman et al., 1995; Van der Meer & De Jong, 2000).

The two major problems become even more serious when it is desired to determine and map abundance of iron-bearing oxide/hydroxide/sulfate minerals associated with sulfide-bearing mine waste impoundment, instead of aiming to map distribution of certain mine tailings within mining districts (Farrand & Harsanyi, 1997; Ferrier, 1999; Swayze et al., 2000). The reason is that certain heavy metals are either absorbed on surfaces or incorporated in molecular structures of iron-bearing oxide/hydroxide/sulfate minerals. The latter is important to determine which, and an indication of how much, metals could potentially be released from such types of mine waste impoundments to nearby ecosystems (Audry et al., 2005; Moncur et al., 2005; Sidenko & Sherriff, 2005). Metals released could be a result of further weathering. Each secondary iron-bearing oxide/hydroxide/sulfate mineral within a weathering sulfide-bearing mine waste shows distinctive spectral features in the visible to the shortwave infrared (0.4–2.5μm) regions of the electromagnetic spectrum (Crowley et al., 2003). Such distinctive spectral features would enable identification of individual iron-bearing oxide/hydroxide/sulfate minerals in non-complex mixtures, by means of relatively simple spectral analytical techniques. The ability to estimate their abundance in complex mixtures through spectral unmixing techniques, however, would be complicated by their similar spectral signatures (Crowley et al., 2003). The complexity would be related to (a) selection of end-members of iron-bearing oxide/hydroxide/sulfate minerals based on only image data (for a "true" remote sensing case) and (b) estimation of partial abundances of end-members.

This chapter addresses the question: how could estimates of abundances of spec-

trally similar iron-bearing oxide/hydroxide/sulfate minerals in complex mixtures be obtained using hyperspectral data. To do so, spectral mixtures were generated with varying linear proportions of individual spectra of a set of iron-bearing oxide/hydroxide/sulfate minerals. The set of end-members is commonly associated with sulphide-bearing mine wastes. The first and the second derivatives were then calculated for each of the different sets of mixed spectra and the individual spectra of the minerals. It is shown here that most pairs of the derivatives for individual spectra have lower correlation coefficients than the pairs of original individual spectra. Finally, a method is presented for spectral unmixing, which requires formulation of a linear function of the individual spectra.

5.2 Method of spectral unmixing

Spectral unmixing is a deconvolution process for estimating the contribution of individual e ($e = 1, \ldots, ||E||$) component spectra to an observed spectrum containing a set M of unknown $||M||$ spectral end-members, where $E \subset M$, $||E||$ and $||M||$ are the number of end-member spectra in E and in M, respectively. Each component spectrum e, which can be derived from a spectral library, consists of L discrete wavelengths λ_l ($l = 1, \ldots, L$). It is denoted by $R^e = (R^e(\lambda_1), \ldots, R^e(\lambda_L))$, where $R^e(\lambda_l)$ is the reflectance value at wavelength λ_l. An observed spectrum $U = (U(\lambda_1), \ldots, U(\lambda_L))$ is assumed to be a linear combination of the $||M||$ end-members. It is difficult, if not impossible, to model U for all possible components or end-members in a complex spectral mixture. Instead an exhaustive set of end-members of interest (subset E of M) is considered. The proportion contribution of each of these end-members can then be estimated. Accordingly, a spectrum at λ_l can be modeled as

$$\widehat{U}(\lambda_l) = \sum_{e=1}^{||E||} p_e R^e(\lambda_l) + p_0 R^{M \setminus E}(\lambda_l)\,, \tag{5.1}$$

where $0 \leq p_e \leq 1$ and $p_0 + \sum_{e=1}^{||E||} p_e = 1$ is the contribution of each end-member of interest in the spectral library and $0 \leq p_0 \leq 1$ is the contribution of end-members in the exhaustive set of end-members in $M \setminus E$. $R^{M \setminus E}(\lambda_l)$ is an unknown linear combination of end-members from the exhaustive set of end-members in $M \setminus E$. Previous studies have arrived at either a constrained unmixing solution ($0 \leq p_e \leq 1$ and $\sum_{e=1}^{||E||} p_e = 1$) (Boardman, 1989; Shimabukuro & Smith, 1991; García-Haro et al., 1996; Metternicht & Fermont, 1998; Ferrier, 1999; Drake et al., 1999; Robinson et al., 2000) or an unconstrained unmixing solution (Van der Meer & De Jong, 2000), where the fraction of abundance p_e can be negative and/or the total abundance can exceed 1. Here, a partial constraint is applied because the abundance of a material is between 0 and 1, and the total abundance due to end-members in E is at most equal to 1. This is more realistic, as a mixture can consist of end-members that are either not of interest or are unknown. The resulting values of p_e represent estimates, as proportions, of partial abundance of the material of each considered component or end-member in an observed spectrum.

The difference between the estimated and actual spectra at λ_l equals

$$\epsilon_l = U(\lambda_l) - \sum_{e=1}^{||E||} p_e R^e(\lambda_l)\,. \tag{5.2}$$

To achieve the optimal set of coefficients, it was endeavored to find values p_e that minimize some function of ϵ_l, e.g. SumSpec $= \sum_{l=1}^{L} |\epsilon_l|$ or VarSpec $= \mathrm{var}(\epsilon_l)$. Because remote sensing images are subject to albedo effects and the reflectance seldom match end-member reflectances, it is appealing to use either the differences in the first derivative or the second derivative instead of the actual differences. The difference in the second derivative between an estimated and an actual spectrum at λ_l is

$$\epsilon'_l = \frac{\Delta^2 U(\lambda_l)}{\Delta^2 \lambda_l} - \sum_{e=1}^{||E||} p_e \left(\frac{\Delta^2 R^e(\lambda_l)}{\Delta^2 \lambda_l} \right), \tag{5.3}$$

where $\Delta^2 x_l = \Delta x_{l+1} - \Delta x_l = x_{l+1} - 2x_l + x_{l-1}$ and $\Delta x_l = x_{l+1} - x_l$. With these equations $\Delta^2 U(\lambda_l)$, $\Delta^2 \lambda_l$ and $\Delta^2 R^e(\lambda_l)$ can be calculated. Minimization of a loss function of Equation 5.3, e.g. SumDeriv $= \sum_{l=1}^{L-2} |\epsilon'_l|$ or VarDeriv $= \mathrm{var}(\epsilon'_l)$, results in estimates of p_e, which are the fractions or partial abundance of each end-member e with reflectance R^e. The minimization is achieved through simulated annealing, using either SumSpec, SumDeriv, VarSpec or VarDeriv as the fitness function to optimize.

Simulated annealing

Simulated annealing is a general optimization method that has been widely applied to find the global optimum of an objective function called the fitness function $\phi(\epsilon)$. The fitness function depends on the configuration of the estimates p_e, corresponding to ϵ that is to be minimized. As such, simulated annealing (Aarts & Korst, 1989) is a computer intensive search technique to find the optimum value of a function of the absolute difference between an image (mixed) spectra and a linearly combined reference spectrum, by continually updating this function at successive steps. The problem of non-orthogonality in matrix inversion is thus avoided and reduced to solving a finite state space combinatorial problem. Unmixing of image spectra by means of optimization was previously addressed by applying simulated annealing (Penn, 2002) and by using a genetic optimization algorithm (Linforda & Platzman, 2004). The method of spectral unmixing presented here is demonstrated by means of a comparative study using synthetic spectra with different mixing properties.

Starting with a random configuration of p_e, $\phi(\epsilon^0)$ is calculated. Let ϵ^i and ϵ^{i+1} represent two solutions with fitness $\phi(\epsilon^i)$ and $\phi(\epsilon^{i+1})$, respectively. Configuration ϵ^{i+1} is derived from ϵ^i by randomly replacing one point p_j of ϵ^i by a new point p_k in $\left[0, 1 + p_j - \sum p_e\right]$, so that $\sum p_e \leq 1$. A probabilistic acceptance criterion decides whether ϵ^{i+1} is accepted or not. This probability $P_c(\epsilon^i \rightarrow \epsilon^{i+1})$ of ϵ^{i+1} being accepted

equals

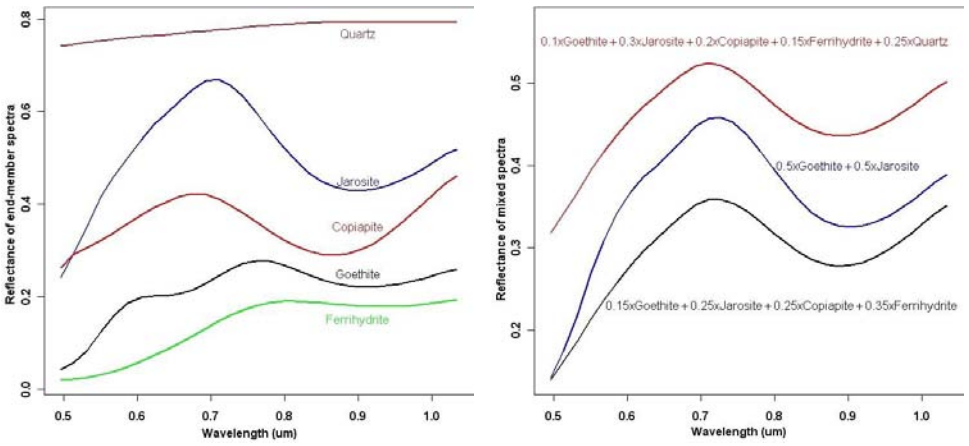
$$P_c(\epsilon^i \rightarrow \epsilon^{i+1}) = \begin{cases} 1\ , & \text{if } \phi(\epsilon^{i+1}) \leq \phi(\epsilon^i) \\ \exp\left(\dfrac{\phi(\epsilon^i) - \phi(\epsilon^{i+1})}{\mathbf{c}}\right)\ , & \text{if } \phi(\epsilon^{i+1}) > \phi(\epsilon^i) \end{cases} \tag{5.4}$$

where $\mathbf{c}$ denotes a parameter. This parameter is reduced by a factor of 0.95 after several transitions are made, thereby decreasing the probability of accepting inferior moves. Reduction stops when the process stabilizes. A transition takes place if ϵ^{i+1} is accepted. Next, a solution ϵ^{i+2} is derived from ϵ^{i+1}, and the probability $P_c(\epsilon^{i+1} \rightarrow \epsilon^{i+2})$ is calculated with a similar acceptance criterion as Equation 5.4. The fitness function will be one of SumSpec, VarSpec, SumDeriv or VarDeriv.

5.3 End-member spectra and synthetic mixtures

Synthetic spectral mixtures were created to test the proposed spectral unmixing methodology. Based on expert knowledge and geoscience literature, five minerals were selected to compose a set of end-members, namely: ferrihydrite; copiapite; jarosite; goethite and quartz. The first four minerals represent the spectra of secondary iron-bearing sulfate hydroxide oxide minerals that could form from pyrite-rich mine wastes (Swayze et al., 2000). The fifth mineral usually represents gangue (non-economic) materials in ores and thus forms part of mine wastes. Although each secondary iron-bearing oxide/hydroxide/sulfate mineral within a weathering sulfide-bearing mine waste shows distinctive spectral features in the 0.4–2.5 μm regions of the electromagnetic spectrum (Crowley et al., 2003), this study was limited to the spectral range 0.5–1.1 μm, because this is where most of the iron-bearing oxide/hydroxide/sulfate minerals of interest have many and strong spectral features. The individual spectrum of each of the five end-members (Figure 5.1(a)) was selected from the USGS spectral library (Clark et al., 1993) and then linearly mixed with each other according to some proportions of each end-member (Figure 5.1(b)). Note that quartz is spectrally featureless (i.e., has no diagnostic spectral features) in the spectral range of interest (Clark, 1995), which makes it difficult, but not impossible, to identify and estimate from spectral data in the range of interest. The mixed spectrum was then degraded to an approximate 15 nm spectral resolution. The resampling was performed (a) to simulate data with lower spectral resolution hyperspectral sensors (e.g., HyMap, DAIS, etc.) as compared to the spectral resolution of the original end-members in the library, (b) to reduce dimensionality of the data, and (c) because it is a practical technique found effective for prediction of different soil properties (Ben-Dor & Banin, 1994). It is acknowledged, however, that spectral channel degradation potentially increases correlation between end-members, which would undermine spectral unmixing (Van der Meer & De Jong, 2000). For this reason, it is proposed and demonstrated to "decorrelate" end-members (see further below) using either their first or their second derivative (Figure 5.2).

Experiments were made on several different mixtures of selected end-members (with



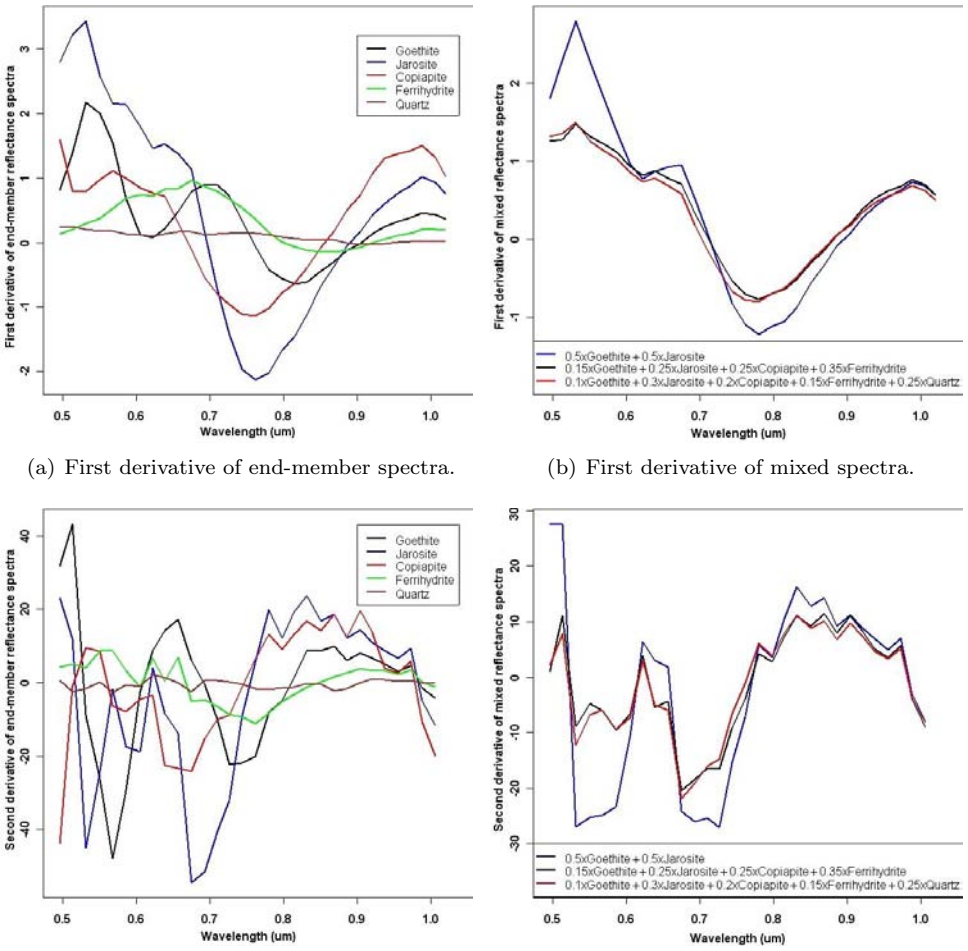
(a) Spectra. (b) Mixed spectra.

Figure 5.1: Five end-member spectra from USGS library, resampled to DAIS VIR region and mixed spectra.

or without quartz), which could plausibly occur in real situations: (a) the observed spectrum is a pure end-member spectrum and is included in the set of end-members considered for unmixing ($R^e \in E$); (b) the observed spectrum is a pure end-member spectrum but is excluded (because, e.g. not identified or not known) in the set of end-members considered for unmixing ($R^e \in M \backslash E$); (c) the observed spectrum is a mixture of end-member spectra and all end-members are included in the set of end-members considered for unmixing (all $R^e \in E$) and (d) the observed spectrum is a mixture of end-member spectra and some end-members are excluded (for similar reasons as in (b)) in the set of end-members considered for unmixing (some $R^e \in M \backslash E$). For (c) and (d), two simple mixtures of the end-member spectra were considered, namely, mixed spectrum 1 with 50% goethite and 50% jarosite, and mixed spectrum 2 with 15% goethite, 25% jarosite, 25% copiapite and 35% ferrihydrite (Figure 5.1(b)). Note that (c) also considers the case where the mixed spectrum results from a smaller set of end-members in E. This is useful if prior information suggests including end-members in the set E, when in fact these end-members does not contribute to the mixed spectrum. The resulting abundance for these end-members, in such a case, should then be equal to zero.

5.4 Results

Only results of using the second derivatives of observed and end-member spectra are reported, because these are more accurate in the estimation than the results using the first derivatives.



(a) First derivative of end-member spectra.

(b) First derivative of mixed spectra.

(c) Second derivative of end-member spectra.

(d) Second derivative of mixed spectra.

Figure 5.2: First and second derivatives of end-member and mixed spectra.

Table 5.1: Estimated mineral abundance based on observed spectra of mixtures containing known proportions of quartz and four iron-bearing minerals. The dash (—) indicate that the end-member spectrum was excluded from the set E of end-members considered for unmixing.

Known abundance					Estimated abundance					
Goe	Jar	Cop	Fer	Qua	Goe	Jar	Cop	Fer	Qua	$M \backslash E$[1]
					Using observed spectra & SumSpec[2]					
End-member spectrum included in E										
0.00	0.00	0.00	0.00	1.00	0.01	0.00	0.00	0.00	0.94	0.05
End-member spectrum excluded from E										
0.00	0.00	0.00	0.00	1.00	0.04	0.93	0.00	0.01	—	0.02
Mixtures										
0.10	0.30	0.20	0.15	0.25	0.01	0.70	0.27	0.01	—	0.01
0.10	0.30	0.20	0.15	0.25	0.05	0.33	0.14	0.16	0.27	0.05
					Using observed spectra & VarSpec[3]					
End-member spectrum included in E										
0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.02	0.10	0.70	0.18
End-member spectrum excluded from E										
0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.27	—	0.72
Mixtures										
0.10	0.30	0.20	0.15	0.25	0.11	0.31	0.18	0.15	—	0.25
0.10	0.30	0.20	0.15	0.25	0.08	0.30	0.20	0.17	0.22	0.03

[1] represents the set of end-member spectra that are in the exhaustive set containing all possible end-member spectra but excluding the end-member spectra used for unmixing.
[2] represents the sum of the absolute difference between the observed spectrum to the end-member spectra.
[3] represents the variance of the differences between the observed spectrum to the end-member spectra.

Table 5.2: Estimated mineral abundance based on second derivative of observed spectra of mixtures containing known proportions of quartz and four iron-bearing minerals. The dash (—) indicate that the end-member spectrum was excluded from the set E of end-members considered for unmixing.

Known abundance					Estimated abundance					
Goe	Jar	Cop	Fer	Qua	Goe	Jar	Cop	Fer	Qua	$M \backslash E^1$
					Using second derivatives of spectra & SumDeriv[2]					
End-member spectrum included in E										
0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.03	0.87	0.09
End-member spectrum excluded from E										
0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.01	0.10	—	0.86
Mixtures										
0.10	0.30	0.20	0.15	0.25	0.00	0.00	0.00	0.98	—	0.02
0.10	0.30	0.20	0.15	0.25	0.11	0.30	0.20	0.14	0.23	0.02
					Using second derivatives of spectra & VarDeriv[3]					
End-member spectrum included in E										
0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.04	0.87	0.09
End-member spectrum excluded from E										
0.00	0.00	0.00	0.00	1.00	0.02	0.00	0.00	0.01	—	0.97
Mixtures										
0.10	0.30	0.20	0.15	0.25	0.11	0.30	0.20	0.16	—	0.23
0.10	0.30	0.20	0.15	0.25	0.10	0.30	0.19	0.17	0.18	0.06

[1] represents the set of end-member spectra that are in the exhaustive set containing all possible end-member spectra but excluding the end-member spectra used for unmixing.
[2] represents the sum of the absolute differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.
[3] the variance of the differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.

Table 5.3: Estimated mineral abundance based on observed spectra of mixtures containing known proportions of four iron-bearing minerals. The dash (—) indicate that the end-member spectrum was excluded from the set E of end-members considered for unmixing.

Known abundance				Estimated abundance				
Goe	Jar	Cop	Fer	Goe	Jar	Cop	Fer	$M \backslash E^1$
				Using observed spectra & SumSpec[2]				
End-member spectrum included in E								
1.00	0.00	0.00	0.00	0.88	0.02	0.01	0.08	0.01
0.00	1.00	0.00	0.00	0.01	0.93	0.02	0.01	0.03
0.00	0.00	1.00	0.00	0.00	0.03	0.96	0.00	0.01
0.00	0.00	0.00	1.00	0.03	0.01	0.01	0.88	0.07
End-member spectrum excluded from E								
1.00	0.00	0.00	0.00	—	0.24	0.00	0.69	0.07
0.00	1.00	0.00	0.00	0.00	—	0.98	0.00	0.02
0.00	0.00	1.00	0.00	0.00	0.68	—	0.00	0.32
0.00	0.00	0.00	1.00	0.70	0.00	0.00	—	0.30
Mixtures								
0.50	0.50	0.00	0.00	0.49	0.51	—	—	0.00
0.50	0.50	0.00	0.00	0.39	0.53	0.01	0.05	0.02
0.15	0.25	0.25	0.35	0.63	0.31	—	—	0.06
0.15	0.25	0.25	0.35	0.15	0.26	0.24	0.35	0.00
				Using observed spectra & VarSpec[3]				
End-member spectrum included in E								
1.00	0.00	0.00	0.00	0.91	0.02	0.01	0.05	0.01
0.00	1.00	0.00	0.00	0.02	0.92	0.01	0.00	0.05
0.00	0.00	1.00	0.00	0.00	0.03	0.96	0.00	0.01
0.00	0.00	0.00	1.00	0.01	0.02	0.00	0.92	0.05
End-member spectrum excluded from E								
1.00	0.00	0.00	0.00	—	0.26	0.01	0.69	0.04
0.00	1.00	0.00	0.00	0.43	—	0.56	0.00	0.01
0.00	0.00	1.00	0.00	0.00	0.37	—	0.00	0.63
0.00	0.00	0.00	1.00	0.93	0.00	0.00	—	0.07
Mixtures								
0.50	0.50	0.00	0.00	0.50	0.50	—	—	0.00
0.50	0.50	0.00	0.00	0.47	0.49	0.02	0.02	0.00
0.15	0.25	0.25	0.35	0.58	0.24	—	—	0.18
0.15	0.25	0.25	0.35	0.13	0.25	0.25	0.37	0.00

[1] represents the set of end-member spectra that are in the exhaustive set containing all possible end-member spectra but excluding the end-member spectra used for unmixing.
[2] represents the sum of the absolute difference between the observed spectrum to the end-member spectra.
[3] represents the variance of the differences between the observed spectrum to the end-member spectra.

Table 5.4: Estimated mineral abundance based on second derivative of observed spectra of mixtures containing known proportions of four iron-bearing minerals. The dash (—) indicate that the end-member spectrum was excluded from the set E of end-members considered for unmixing.

Known abundance				Estimated abundance				
Goe	Jar	Cop	Fer	Goe	Jar	Cop	Fer	$M \backslash E^1$
				Using second derivatives of spectra & SumDeriv[2]				
End-member spectrum included in E								
1.00	0.00	0.00	0.00	0.98	0.00	0.01	0.00	0.01
0.00	1.00	0.00	0.00	0.00	0.94	0.05	0.01	0.01
0.00	0.00	1.00	0.00	0.01	0.02	0.96	0.00	0.01
0.00	0.00	0.00	1.00	0.02	0.00	0.01	0.96	0.01
End-member spectrum excluded from E								
1.00	0.00	0.00	0.00	—	0.21	0.00	0.78	0.01
0.00	1.00	0.00	0.00	0.36	—	0.63	0.00	0.01
0.00	0.00	1.00	0.00	0.00	0.43	—	0.01	0.56
0.00	0.00	0.00	1.00	0.10	0.06	0.01	—	0.83
Mixtures								
0.50	0.50	0.00	0.00	0.50	0.50	—	—	0.00
0.50	0.50	0.00	0.00	0.45	0.50	0.00	0.04	0.01
0.15	0.25	0.25	0.35	0.00	0.00	—	—	1.00
0.15	0.25	0.25	0.35	0.15	0.26	0.24	0.34	0.01
				Using second derivatives of spectra & VarDeriv[3]				
End-member spectrum included in E								
1.00	0.00	0.00	0.00	0.97	0.00	0.01	0.01	0.01
0.00	1.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00
0.00	0.00	1.00	0.00	0.00	0.01	0.98	0.01	0.00
0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.99	0.00
End-member spectrum excluded from E								
1.00	0.00	0.00	0.00	—	0.25	0.00	0.52	0.23
0.00	1.00	0.00	0.00	0.43	—	0.56	0.00	0.01
0.00	0.00	1.00	0.00	0.00	0.29	—	0.00	0.71
0.00	0.00	0.00	1.00	0.05	0.03	0.00	—	0.92
Mixtures								
0.50	0.50	0.00	0.00	0.50	0.50	—	—	0.00
0.50	0.50	0.00	0.00	0.48	0.50	0.00	0.01	0.01
0.15	0.25	0.25	0.35	0.10	0.35	—	—	0.55
0.15	0.25	0.25	0.35	0.15	0.25	0.25	0.35	0.00

[1] represents the set of end-member spectra that are in the exhaustive set containing all possible end-member spectra but excluding the end-member spectra used for unmixing.
[2] represents the sum of the absolute differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.
[3] represents the variance of the differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.

5.4.1 Abundance estimation from spectra containing quartz

Tables 5.1 and 5.2 show the results of the above experiments using observed spectra and second derivatives of the spectra, respectively, containing quartz and performed using SumSpec, VarSpec, SumDeriv and VarDeriv.

If quartz is the observed spectrum and is included in the set of end-members, the accuracy of estimated abundance for quartz (between 0.70 and 0.94) is generally lower than the results obtained for other minerals. In this case, abundance estimates are most accurate by using SumSpec.

If quartz is the observed spectrum but is excluded in the set of end-members, the accuracy of estimated abundance for quartz (between 0.02 and 0.97; see column $M \backslash E$) is also generally lower than the results obtained for the other minerals. In this case, abundance estimates are most accurate by using VarDeriv.

For the same mixed spectrum of 10% goethite, 30% jarosite, 20% copiapite, 15% ferrihydrite and 25% quartz (Figure 5.1(b)), estimated abundances have high accuracy (≤ 0.06 inaccuracy, see second row under mixtures in Tables 5.1 and 5.2) when all end-members are included in the set of end-members considered for unmixing (all $R^e \in E$). In this case, abundance estimates by using the second derivative of the spectra are more accurate than by using the original spectra.

For the same mixed spectrum (Figure 5.1(b)), but excluding quartz from the set of end-member spectra considered for unmixing, accuracy of abundance estimates for each end-member decreases (see first row under mixtures in Tables 5.1 and 5.2). Estimated abundance for quartz, which can be seen in column $M \backslash E$, ranges from 0.01 to 0.02 out of the actual abundance of 0.25 for SumSpec and SumDeriv and from 0.23 to 0.25 for VarSpec and VarDeriv. These results show that estimated abundances for the other minerals are accurate by using VarSpec and VarDeriv if quartz is not considered as end-member for unmixing even if it is present. In all, abundance estimates are most accurate using VarDeriv.

5.4.2 Abundance estimation from spectra containing no quartz

Tables 5.3 and 5.4 show the results of the experiments using observed spectra and second derivatives of spectra, respectively, containing no quartz and performed using SumSpec, VarSpec, SumDeriv and VarDeriv.

If the observed spectrum is a pure end-member, and is included in the set of end-members considered for unmixing, application of the proposed method is able to estimate abundance of all materials in set E (Tables 5.3 and 5.4) with a high degree of accuracy, with estimated abundance ≥ 0.88 for the correct end-member and ≤ 0.08 for an incorrect end-member. Abundance estimates by using the second derivative of the spectra are more accurate than abundance estimates by using the original spectra. Moreover, abundance estimates by using VarDeriv are more accurate than abundance

estimates by using SumDeriv.

If the observed spectrum is a pure end-member, but is excluded from the set of end-members considered for unmixing, application of the proposed method results in estimated abundances of the materials with poor accuracies. Estimated abundances vary between 0.01 and 0.92 for the correct end-member (see the columns $M\backslash E$) and between 0.00 and 0.98 for an incorrect end-member. Nevertheless, abundance estimates by using the second derivative of the spectra (Table 5.4) are slightly more accurate than abundance estimates by using the original spectra (Table 5.3), particularly for copiapite and ferrihydrite.

If the end-member spectra contributing to the observed mixed spectrum are included in the set of end-members considered for unmixing (all $R^e \in E$), the estimated abundances have a very high accuracy for each end-member (≤ 0.11 inaccuracy, see first, second and fourth rows under mixtures in Tables 5.3 and 5.4). Using the second derivative of the spectra produced slightly better abundance estimates (Table 5.4) than when using the original spectra (Table 5.3). In total, abundance estimates obtained by using VarDeriv are most accurate.

If some end-member spectra contributing to the observed spectrum are excluded from the set of end-members considered for unmixing (some $R^e \in M\backslash E$), estimated abundances have a low accuracy for each end-member (between 0.00 to 0.48 inaccuracy, see third row under mixtures in Tables 5.3 and 5.4). Also in this case, use of VarDeriv results in the most accurate abundance estimates.

5.4.3 Abundance estimates by exclusion of contributing end-members

The results show that abundance estimates are generally inaccurate if end-members contributing to the observed pure spectrum are excluded from the set of end-members considered for unmixing (some $R^e \in M\backslash E$). If the observed spectrum is a pure end-member but is excluded from the set of end-members for unmixing, then the abundance of other end-members is often overestimated, particularly when using the original spectra (Tables 5.1 to 5.4).

If the observed spectrum is pure quartz but is excluded from the set of end-member spectra considered for unmixing, then some end-members, particularly jarosite or ferrihydrite not contributing to the observed spectrum are over-estimated (Tables 5.1 and 5.2). The results are similar if quartz is excluded from the set of end-member spectra considered to unmix an observed mixed spectrum containing quartz. Overestimation of either jarosite or ferrihydrite is particularly remarkable when the original spectrum is unmixed rather than the second derivative of the original spectrum. A plausible reason is that the unmixing procedure yields an end-member in E with spectral signature similar to the observed spectrum even though the actual end-member is excluded from E. From Table 5.5, it can be seen that that there is positive correlation between quartz spectrum and jarosite spectrum and high positive correlation between quartz spectrum and ferrihydrite spectrum. The relatively slight over-estimation of

ferrihydrite when the second derivative of the original spectrum is unmixed could also be due to positive albeit low correlation between second derivative of quartz spectrum and second derivative of ferrihydrite spectrum.

If the observed spectrum is pure mineral but is excluded from the set of end-member spectra considered for unmixing, then other minerals not contributing to the observed spectrum are over-estimated (Tables 5.3 and 5.4). For example, if the observed spectrum is jarosite but is excluded from the set of end-member spectra considered for unmixing, then copiapite is over-estimated; and vice versa. Similarly, if the observed spectrum goethite but is excluded from the set of end-member spectra considered for unmixing, then ferrihydrite is over-estimated; and, vise versa. Over-estimation of another mineral is particularly remarkable when the original spectrum is unmixed rather than the second derivative of the original spectrum. A plausible reason for this is the high correlation between copiapite spectrum and jarosite spectrum and between goethite spectrum and ferrihydrite spectrum (Table 5.5). The relatively slight over-estimation of another mineral when the second derivative of the original spectrum is unmixed is also plausibly due to positive albeit low correlation between the second derivative spectra of copiapite and jarosite and between the second derivative spectra of goethite and ferrihydrite.

5.4.4 Performance of fitness functions

Results of abundance estimates in Tables 5.1 to 5.4 already indicate relative good performance by VarDeriv among the four fitness functions. To properly compare performance of the four fitness functions, error in estimated abundance is quantified as an absolute difference between the actual (known) and estimated abundance of each end-member (Err). These errors were ranked accordingly, and the average error and the average rank error in abundance estimates by using each of four fitness functions were determined and shown in Tables 5.6, 5.7, and 5.8.

If the observed spectrum is a pure end-member but is either included or excluded in the set of end-members considered for unmixing, VarDeriv performs best in terms of the average error and average rank error. For example, when observed spectrum is pure ferrihydrite but is excluded in the set of end-members considered for unmixing, abundance estimates by using VarDeriv are mostly accurate (see Tables 5.6 and 5.7). If all or some end-member spectra contributing to the observed spectrum are either included or excluded in the set of end-members considered for unmixing, then VarDeriv outperforms the other fitness functions in terms of average error and average rank error (Table 5.8). If the observed spectrum consists of mixtures of quartz spectrum either with or without the other four spectra, then VarDeriv outperforms the other fitness functions in terms of average error and average rank error (Table 5.8). The superior performance of VarDeriv in unmixing of the different experimental spectra can be attributed to the lower correlations among the second derivatives of the end-member spectra as compared to the correlations of the original spectra (Table 5.5).

Table 5.5: Correlation coefficient between pairs of original spectra, pairs of first derivative of spectra and pairs of second derivative of spectra.

Original spectra:					
	goethite	jarosite	copiapite	ferrihydrite	quartz
goethite	1.00				
jarosite	0.67	1.00			
copiapite	0.43	0.72	1.00		
ferrihydrite	0.86	0.29	0.16	1.00	
quartz	0.85	0.23	0.15	0.98	1.00
First derivative of spectra:					
	goethite	jarosite	copiapite	ferrihydrite	quartz
goethite	1.00				
jarosite	0.71	1.00			
copiapite	0.35	0.79	1.00		
ferrihydrite	0.44	0.24	-0.15	1.00	
quartz	0.60	0.36	-0.20	0.57	1.00
Second derivative of spectra:					
	goethite	jarosite	copiapite	ferrihydrite	quartz
goethite	1.00				
jarosite	0.35	1.00			
copiapite	-0.14	0.43	1.00		
ferrihydrite	0.22	0.18	-0.02	1.00	
quartz	0.17	-0.04	-0.23	0.15	1.00

Table 5.6: Errors in mineral abundance estimates derived from observed spectra of mixtures containing known proportions of four iron-bearing minerals.

Known abundance				Estimated abundance			
				Spectra			
				Using SumSpec[1]		Using VarSpec[2]	
Goe	Jar	Cop	Fer	|Err|[3]	Rank[4]	|Err|	Rank
End-member spectrum included in E							
1.00	0.00	0.00	0.00	0.23	4.0	0.16	3.0
0.00	1.00	0.00	0.00	0.11	2.5	0.11	2.5
0.00	0.00	1.00	0.00	0.07	3.0	0.07	3.0
0.00	0.00	0.00	1.00	0.17	4.0	0.11	3.0
Averages				0.145	3.375	0.113	2.875
End-member spectrum excluded in E							
1.00	0.00	0.00	0.00	0.93	2.0	0.96	3.0
0.00	1.00	0.00	0.00	0.98	1.0	0.99	3.0
0.00	0.00	1.00	0.00	0.68	4.0	0.37	2.0
0.00	0.00	0.00	1.00	0.70	3.0	0.93	4.0
Averages				0.823	2.500	0.813	3.000
Mixtures							
0.50	0.50	0.00	0.00	0.02	4.0	0.00	2.0
0.50	0.50	0.00	0.00	0.20	4.0	0.08	2.0
0.15	0.25	0.25	0.35	0.54	4.0	0.48	3.0
0.15	0.25	0.25	0.35	0.02	2.0	0.04	4.0
Averages				0.195	3.500	0.150	2.750

[1] represents the sum of the absolute differences between the observed spectrum to the end-member spectra.
[2] represents the variance of the differences between the observed spectrum to the end-member spectra.
[3] represents the absolute differences between the actual and estimated abundance of each end-member.
[4] represents the ranked values for the error in the estimated abundance to the known abundance of each end-member.

Table 5.7: Errors in mineral abundance estimates derived from second derivative of observed spectra of mixtures containing known proportions of four iron-bearing minerals.

Known abundance				Estimated abundance			
				Second derivatives of spectra			
				Using SumDeriv[1]		Using VarDeriv[2]	
Goe	Jar	Cop	Fer	|Err|[3]	Rank[4]	|Err|	Rank
End-member spectrum included in E							
1.00	0.00	0.00	0.00	0.03	1.0	0.05	2.0
0.00	1.00	0.00	0.00	0.12	4.0	0.02	1.0
0.00	0.00	1.00	0.00	0.07	3.0	0.04	1.0
0.00	0.00	0.00	1.00	0.07	2.0	0.02	1.0
Averages				0.100	2.500	0.080	1.250
End-member spectrum excluded in E							
1.00	0.00	0.00	0.00	0.99	4.0	0.77	1.0
0.00	1.00	0.00	0.00	0.99	3.0	0.99	3.0
0.00	0.00	1.00	0.00	0.46	3.0	0.29	1.0
0.00	0.00	0.00	1.00	0.17	2.0	0.08	1.0
Averages				0.653	3.000	0.533	1.500
Mixtures							
0.50	0.50	0.00	0.00	0.00	2.0	0.00	2.0
0.50	0.50	0.00	0.00	0.09	3.0	0.03	1.0
0.15	0.25	0.25	0.35	0.40	2.0	0.15	1.0
0.15	0.25	0.25	0.35	0.03	3.0	0.00	1.0
Averages				0.130	2.500	0.045	1.250

[1] represents the sum of the absolute differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.
[2] represents the variance of the differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.
[3] represents the absolute differences between the actual and estimated abundance of each end-member.
[4] represents the ranked values for the error in the estimated abundance to the known abundance of each end-member.

Table 5.8: Errors in mineral abundance estimates derived from observed spectra and from second derivative of observed spectra of mixtures containing known proportions of quartz and four iron-bearing minerals.

Known abundance					Estimated abundance			
Goe	Jar	Cop	Fer	Qua	|Err|[1]	Rank[2]	|Err|	Rank
					Spectra			
					Using SumSpec[3]		Using VarSpec[4]	
End-member spectrum included in E								
0.00	0.00	0.00	0.00	1.00	0.07	1.0	0.42	4.0
End-member spectrum excluded in E								
0.00	0.00	0.00	0.00	1.00	0.98	4.0	0.28	3.0
Mixtures								
0.10	0.30	0.20	0.15	0.30	0.70	3.0	0.04	2.0
0.10	0.30	0.20	0.15	0.30	0.17	4.0	0.07	2.0
Averages					0.480	3.000	0.203	2.750
					Second derivatives of spectra			
					Using SumDeriv[5]		Using VarDeriv[6]	
End-member spectrum included in E								
0.00	0.00	0.00	0.00	1.00	0.17	2.5	0.17	2.5
End-member spectrum excluded in E								
0.00	0.00	0.00	0.00	1.00	0.14	2.0	0.03	1.0
Mixtures								
0.10	0.30	0.20	0.15	0.30	1.33	4.0	0.02	1.0
0.10	0.30	0.20	0.15	0.30	0.04	1.0	0.10	3.0
Averages					0.420	2.375	0.080	1.875

[1] represents the absolute differences between the actual and estimated abundance of each end-member.
[2] represents the ranked values for the error in the estimated abundance to the known abundance of each end-member.
[3] represents the sum of the absolute differences between the observed spectrum to the end-member spectra.
[4] represents the variance of the differences between the observed spectrum to the end-member spectra.
[5] represents the sum of the absolute differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.
[6] represents the variance of the differences between the second derivative of the observed spectrum to the second derivative of end-member spectra.

5.5 Discussion

Technology for mapping, within mining districts, of sources of acid-generating minerals (e.g., pyrite) or their weathering products using imaging spectroscopy and hyperspectral data is a potentially useful tool for environmental remediation studies. However, techniques for mapping, within only mine waste impoundments, of abundance and distribution of secondary iron-bearing sulfates/hydroxides/oxides (as an indirect way to find acid-generating minerals) using hyperspectral data are still desirable. Development of such techniques is not straightforward because of the high spectral similarity among the target materials (i.e., only, if not mostly, weathering products of, say, pyrite), which undermines current algorithms for spectral unmixing of hyperspectral data.

Previous techniques of spectral unmixing by means of inversion of a matrix of the spectral data can become problematic because of (a) highly corrected data in hyperspectral bands and (b) high correlation between the target materials within mine wastes areas. The idea of using singular-valued decomposition of the end-member matrix in lieu of matrix inversion was suggested by Boardman (1989). In contrast, Van der Meer & De Jong (2000) proposed to enhance orthogonality of end-member matrix by application of the MNF (minimum noise fraction) transformation, which was developed by Green et al. (1988) for noise removal in multispectral data. Spectral unmixing of MNF-transformed hyperspectral data, however, requires that the noise covariance matrix of the data be known or estimated thereby separating the noise from the data. Although this proves to reduce the correlation among end-member spectra, this study considered explicitly de-correlating noise-free end-member spectra. It was decided to follow Boardman (1989) suggestion to use derivatives of spectral data in order to reduce correlations among end-members. Accordingly, estimated abundances of each of the spectrally similar materials of interest based on either the first or the second derivatives of the observed spectra are more accurate than those based on only the observed spectra. It was found further and thus only reported here that using the second derivatives of the observed spectra results in most accurate estimates of abundance of spectrally similar minerals of interest.

A successful application, via simulated annealing, of second derivatives of observed spectra to estimate abundance of spectrally similar iron-bearing minerals in mine tailings has not been reported yet, until now. In the same field of study, Penn (2002) also used simulated annealing in spectral unmixing to estimate mineral abundance within and around a porphyry copper deposit, although he did not use second derivates of observed spectra but simply normalized the hyperspectral data and the end-member spectra to alleviate albedo effects in the data. In the present work, normalization of either the observed spectra or the end-member spectra is not required. In a different field of study, Loethen et al. (2004) also found that using a second-derivative variance minimization procedure outperforms other methods to estimate a dilute component (solute) from a liquid mixture whose spectrum is dominated by a major component (solvent). To estimate the dilute component, however, Loethen et al. (2004) minimize the variance by assuming the solute and solvent spectra are non-correlated. For spec-

trally similar end-members, as in mine tailings, assumption of non-correlation must be demonstrated.

Results of the study showed that using the fitness function VarDeriv results in the most accurate estimates of abundance of spectrally similar minerals in a complex mixture such as mine tailings. The VarDeriv functions to minimize the variance of the difference between the second derivatives of the observed spectra and the reference end-member spectra. The results therefore demonstrate that the simulated annealing procedure is efficient in minimizing the variance of the difference between the second derivatives of the target spectra and the reference end-member spectra. The findings reported are, nonetheless, consistent with previous findings that error in abundance estimates, is a function of the difference between the target spectra and the reference end-member spectra, and is proportional to the amount of variability in the individual end-member spectra relative to the mixture (Van der Meer & De Jong, 2000).

Results of the study show that abundance estimates inclusive of end-members other than those that contribute to the observed spectra are more accurate than abundances estimates exclusive of end-members that do not contribute to the observed spectra. The results therefore imply that a set of end-members should include not only major components but also minor to trace components. It was not investigated here, however, whether or not accurate estimates would be obtained if an end-member set contains all 15 secondary iron-bearing sulfates/hydroxides/oxides listed in Crowley et al. (2003), because it is known that distinction between end-member spectra becomes difficult as number of reference spectra increases. Penn (2002) used in simulated annealing, however, 10 end-members (inclusive of spectral similar mineral such as jarosite, goethite, and hematite, as well as other spectrally different materials). This suggests that accurate estimation of abundances of up to 10 spectrally similar minerals in mine wastes is achievable via simulated annealing. This suggestion, however, needs further verification. Selection of which spectrally similar minerals to be included in an end-member set should, nonetheless, be based on expert judgment or based on spectral analysis of calibration spectroscopic data usually collected immediately following an airborne hyperspectral campaign.

Results of the study also show that abundance estimates based on an end-member set inclusive of quartz are more accurate than abundance estimates based on an end-member set exclusive of quartz. The results indicates that quartz should be included in an end-member set for unmixing if mine wastes or geological materials of interest indeed contain quartz. The results also imply that quartz, which is spectrally featureless but nevertheless invariably present in mine wastes, (a) can be estimated with reasonable accuracy and (b) should be included in the end-member set. The results also suggest that, for estimation of spectrally similar minerals in complex mixtures such as mine tailings, end-member selection through image data analysis may not be necessary at all in a practical case exercise of the method described. However, this suggestion also needs further verification.

The study presented here shows potential application to hyperspectral remote sensing for surface characterization of mine tailings, especially those associated with ore

deposits in which pyrite and quartz are the major gangue minerals. A shortcoming of this study, however, is that actual hyperspectral data were not used to demonstrate the methods described. Even so, the potential usefulness of the methods described is supported by using synthetic spectral mixtures end-members commonly found in weathered pyrite-rich mine wastes (Swayze et al., 2000). It is acknowledged, nevertheless, that further demonstration of the methods described by using actual hyperspectral and ground measurements are important for validation purposes (e.g., Bajcsy & Groves (2004)).

5.6 Conclusions

This study resulted in three main conclusions.

- Abundances of spectrally similar minerals in mine wastes can be estimated with relatively high accuracy by unmixing of second derivatives of target spectra, in which contributing components are decorrelated.
- Simulated annealing proved efficient in minimizing variance of the difference spectrum to estimate abundance of spectrally similar minerals.
- Minimization of variance of the difference spectrum shows that using end-member spectra inclusive components, not contributing to target spectra does diminish according to abundance estimates. In contrast, using an end-member spectra exclusive of component contributing to target spectra reduces accuracy of abundance estimation.

Chapter 6

Optimum sampling scheme for surface geochemical characterization of mine tailings

Errors using inadequate data are much less than those using no data at all.

Charles Babbage

This chapter is based on P. Debba, E. J. M. Carranza, A. Stein and F. D. van der Meer (**In Prep.**) Optimum sampling scheme for surface geochemical characterization of mine tailings.[1]

[1]This work was sponsored by ITC International Institute for Geo-Information Science and Earth Observation, project number 3083022 and NRF National Research Foundation, project number 10317, gun 2053944.

Abstract

This chapter describes a method by which an optimal ground sampling scheme can be obtained for a variable of interest, namely, spatial distribution of a suite of heavy metals in mine tailings. Derivation of an optimal sampling scheme makes use of covariates of the spatial variable of interest, which are readily but less accurately obtainable by using airborne hyperspectral data. The covariates are abundances of secondary iron-bearing minerals estimated through spectral unmixing. Via simulated annealing, an optimal retrospective sampling scheme for a previously sampled area is derived having fewer samples but having almost equal mean kriging prediction error as the original ground samples. Via simulated annealing, an optimal prospective sampling scheme for a new unvisited area is derived based on the variogram model of a previously sampled area. The results of this study demonstrate potential application of hyperspectral remote sensing and simulated annealing to surface characterization of large mine tailings having similar climatic and terrain characteristics to the mine tailings in the case study area.

Keywords

Optimized sampling, simulated annealing, hyperspectral data, spectral unmixing, mineral abundance, heavy metal concentration, kriging with external drift

6.1 Introduction

Mine wastes invariably contain very high albeit mostly non-economic concentrations of metals. Metals could potentially be leached from mine wastes and then be released to, thus contaminating, the nearby ecosystems. In mine wastes of sulphide-rich ores, leaching of metals is caused by acid drainage, which is generated by oxidation and hydrolysis of sulphide-rich, for e.g., pyrite, pyrrhotite, gangue minerals. Geochemical characterization of mine waste impoundments is thus important for rehabilitation, or for remediation, to protect the surrounding environment. Such geochemical characterization would entail surface (to subsurface) sampling, which could be labor or cost intensive, especially if not properly planned.

Metals in mine waste impoundments are usually hosted by acid-generating sulphide-rich minerals, for e.g., pyrite, pyrrhotite, or adsorb onto surfaces of weathering products of such sulphide-rich minerals. Unfortunately, such minerals are difficult to detect or identify by using current remote sensing techniques using multispectral or even hyperspectral data. It has been shown, however, that certain sulphide-rich minerals, particularly pyrite, weathers to a series of iron-bearing sulfates, hydroxides and oxides (Swayze et al., 2000). Such secondary iron-bearing sulfates/hydroxides/oxides have diagnostic spectral features (Crowley et al., 2003), which enable their detection or identification with analytical techniques using hyperspectral data. In a previous work (see chapter 5), the potential of using hyperspectral data to estimate abundances of spectrally similar iron-bearing sulfates/hydroxides/oxides was demonstrated (Debba et al., 2005a). It has also been shown that heavy metal contamination in soils can be quantified using reflectance spectroscopy (Kemper & Sommer, 2002). Thus, remote sensing technology potentially provides an indirect tool for surface characterization of mine waste impoundments with oxidizing sulphide-rich materials; namely, for mapping spatial distributions of secondary iron-bearing sulfates/hydroxides/oxides and heavy metals.

Certain groups of metals could be spatially associated with secondary iron-bearing oxides/hydroxides depending on geochemical conditions in a given environment (Levinson, 1974; Rose et al., 1979). Given a model of spatial distribution of secondary iron-bearing oxides/hydroxides, the problem is how to design a sampling scheme that would adequately capture spatial distribution of certain groups of metals. The case study presented in this chapter attempts to model spatial relationships between a multi-element signature and abundance estimates of secondary iron-bearing minerals in mine tailings dumps. The multi-element signature, on one hand, was modeled through factor analysis of element contents of mine tailings samples, which were measured in a laboratory. Abundances of secondary iron-bearing minerals, on the other hand, were estimated by the method demonstrated by Debba et al. (2005a). Spatial relationships between a multi-element signature and abundance estimates of secondary iron-bearing minerals were modeled through conventional kriging with external drift. Derived spatial relationship models are then used for sampling scheme optimization, by means of simulated annealing, for surface characterization of the mine tailings dumps.

Previous studies of simulated annealing applications to obtain optimal sampling schemes to guide sampling to target specific areas, involved stratification (Van Groenigen et al., 2000a), using thresholds (Van Groenigen et al., 2000a), defining a weight function (Van Groenigen et al., 2000b), using ordinary kriging variance (Shieh et al., 2005) and Sacks & Schiller (1988) presented several annealing based algorithms for optimizing a sampling scheme out of a small grid of possible locations. Studies on the design of optimal sampling schemes in combination with remote sensing used fuzzy classification (Tapia et al., 2005) and targeted a particular intense hydrothermal alteration mineral (Debba et al., 2005b).

In a recent paper, Diggle & Lophaven (2006) discuss a retrospective sampling design, which sequentially removes, from a sampling design, samples that contribute least to a Bayesian prediction of a response. They do, however, state that this is not the optimal design. In this chapter, an adaptation of the retrospective sampling methodology by Diggle & Lophaven (2006) is demonstrated, not through the same Bayesian approach but by incorporation of covariates in a conventional kriging with external drift model. In addition, a prospective sampling scheme is derived for nearby unsampled areas based on variogram model of the adjacent sampled area. The present case study area is in the Recsk-Lahóca copper mining area in Hungary.

6.2 The Study Area

The Recsk-Lahóca mining area is situated in the Mátra Mountains, about 110 km northeast of Budapest, Hungary (Figure 6.1). Base- and precious-metal mining of volcanogenic sulphide ore deposits in the Recsk-Lahóca area started about 200 years ago, with major mining activities taking place between the 1950s and 1998. The Recsk deposits consist of a large, buried porphyry copper-polymetallic vein and replacement-copper/gold skarn system. The Lahóca epithermal deposits contain enargite, luzonite, tetrahedrite, and pyrite in a gangue of quartz, kaolinite, and alunite. The Lahóca hill was mined for copper between 1852 and 1979. Development of Recsk Deeps mines resulted in underground shafts extending 1200 m deep. The mine was meant to produce 3 to 5 million tons of copper ore per year but it never went into production since its construction in 1970. Mining of ore deposits in the Recsk-Lahóca area resulted in the exposure of sulphide bearing-rocks to surface water and atmospheric oxygen, which accelerate oxidation, leaching and release of metals and acidity. Mine tailings and waste rock dumps resulting from the two-century mining of copper and gold are present in the area, such as those to the south and southwest of the Recsk mine (Figure 6.2).

This study pertains to the tailings dumps northwest of Lahóca mine, which consist actually of two dumps referred to as “East Tails” and “West Tails”. These tailings dumps were selected because, unlike the other tailings and waste rock dumps in the area, they are not covered with non-mine-waste soil materials derived from other places in the area. The “East Tails” and “West Tails” are situated immediately south

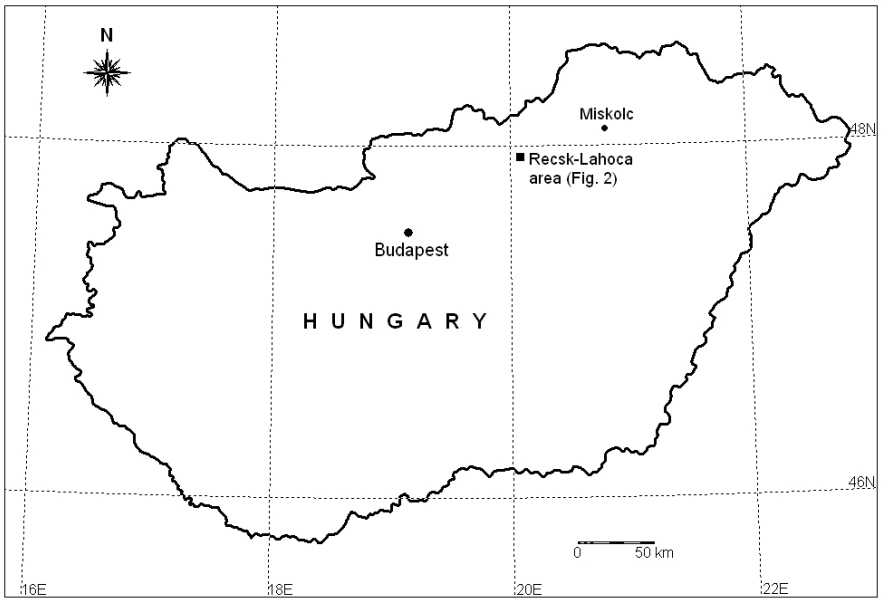


Figure 6.1: Study area: Recsk: Hungary.

of an unpaved road (Figure 6.3). They are separated from each other by a small active southwesterly-flowing stream, whose northwesterly-flowing tributaries were dry at the time of sampling.

6.3 Data

6.3.1 Hyperspectral

HySens 2002 Flight Campaign was initiated in August 2002 to undertake a survey of areas suspected of pollution due to mining in the Matra Mountains. In this study, a subset of the Digital Airborne Imaging Spectrometer (DAIS-7915) is used. The resulting data is a 79 channel hyperspectral image, acquired over the Recsk. DAIS-7915 is a whisk broom sensor, covering a spectral range from visible (0.4 μm) to thermal infrared (12.3 μm) at variable spatial resolution from 3–20 m depending on the carrier aircraft altitude (Lehmann et al., 1995).

Flights took place on the 17th and 18th August 2002. Absence of clouds and wind on the second day were appropriate and hence this image was used for the study. Atmospheric effects were minimized by applying an empirical line calibration method (Roberts et al., 1985) to match field spectra measurements. The study area and hyperspectral image are shown in Figures 6.2 and 6.3 at 5 m nominal resolution on the ground.

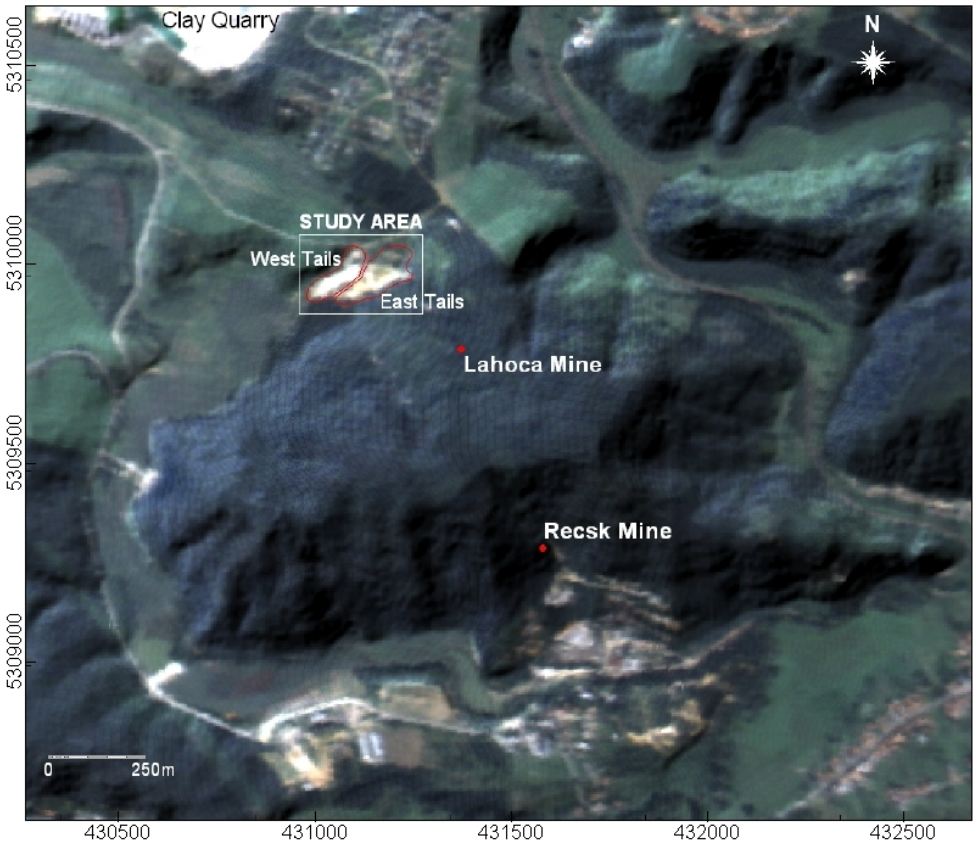


Figure 6.2: The Recsk-Lahóca area shown in pseudo-natural color composite image using DAIS data (red = ch10, green = ch5, blue = ch1) fused with a digital elevation model. Map coordinates are in meters (UTM projection, zone 34N).

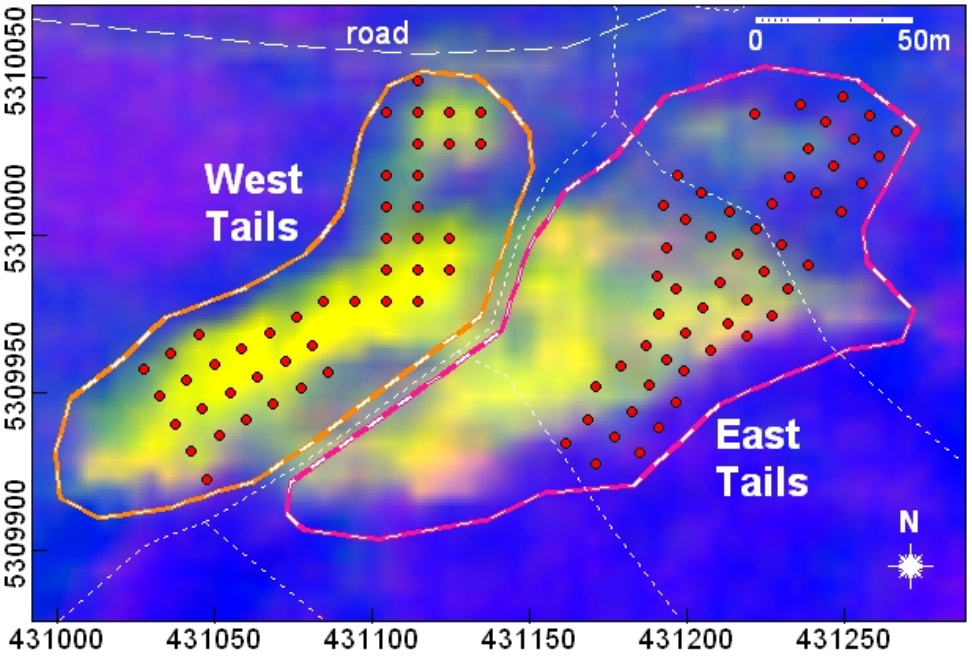


Figure 6.3: The "East Tails" and the "West Tails" shown in a color composite image of the DAIS data. Image of ratios of ch17 to ch28 (representing ferrihydrite reflectance and absorption peaks, respectively) was used as red band. Image of ratios of ch13 to ch25 (representing jarosite reflectance and absorption peaks, respectively) was used as green band. Image of ratios ch32 to ch1 (representing non-iron-bearing minerals) was used as blue band. Red dots are locations of mine tailings samples. Short dashed lines in the image represent drainage lines occupied by either active or non-active streams during the field sampling campaign.

Not all 79 channels were useful as many channels were too noisy and could not be corrected efficiently. Fortunately, the first 32 channels, spectral range 406-1035 nm, where iron-bearing oxides/hydroxides/sulphates have diagnostic features were found useful for this study.

6.3.2 Mine Tailings Data

Samples from the tailings (Figure 6.3) were collected a few minutes shortly after collection of the DAIS hyperspectral data on 18th August 2002. Fifty-three samples were collected in the East Tails and 44 in the West tails. Samples of tailings were collected at 10m×10m grid points in portions of the tailings dumps with almost no vegetation cover within 3 m radius. Portions of the tailings dumps, close to the active stream have steep slopes and were not sampled. Samples were collected from the top surface of the tailings and were air dried at the field camp. In the laboratory, they were further dried in an oven at 40^oC. The minus-80 mesh (< 180 μm) fraction of the tailings material, weighing 500 mg for each sample, was used for chemical analysis. Metal extraction used 63% hot HNO_3 followed by shaking on a vortex shaker and shaking in a water bath at 90^oC for 2 hours. Concentrations of As, Cd, Cu, Fe, Mn, Ni, Pb, Sb and Zn in the decomposed samples were determined using the ICP-AES analyzer. Analytical precision of the measured element concentrations either approaches or is better than 10% according to Thompson & Howarth (1978) control chart.

Table 6.1 shows the elementary statistics of the tailings geochemical data. As indicated by their skewness coefficients, all element distributions are positively skewed except for Fe in the East Tails. Some maximum values are large, exceeding three standard deviations from the mean. Such large values would have a dominant effect on the results of the statistical analysis. Natural logarithmic transformation of the data, shows that asymmetry of the element distributions is effectively reduced.

Of the nine elements studied, concentrations for Fe, Ni, Mn, and Cd are higher in the East Tails than in the West Tails (Figure 6.4), although Cd is enriched in the northern part of the West Tails. Concentrations for As, Cu, Pb, Sb, and Zn are higher in West Tails than in the East Tails. Note also in Table 6.1 that, for As, Sb, Cu, Pb and Zn, (a) concentrations are lower in the East Tails than in the West Tails and (b) standard deviations are lower in the East Tails than in the West Tails. On the one hand, these observations suggest that scavenging of Cd, Mn, and Ni by secondary Fe-bearing oxides/hydroxides is more prevalent in the East Tails than in the West Tails. On the other hand, the data suggest that As, Cu, Pb, Sb, and Zn in the East Tails have become more mobile and released from the tailings to the environment, whereas in the West Tails these metals are still relatively immobile and relatively intact in the tailings. The East Tails and the West Tails, therefore, have quite different geochemical characteristics.

Correlations between the different elements (Table 6.2) show that either the East Tails or the West Tails are characterized by antipathetic associations of As-Sb-Cu-Pb-Zn

Table 6.1: Elementary statistics of original geochemical data and skewness of $\log_e$-transformed data. All concentrations are in ppm except where stated.

East Tails samples ($n = 53$):						
Element	Min	Max	Mean	Std. Dev.	Skewness	Skewness ($\log_e$)
As	48.6	1568.0	266.3	273.9	2.74	0.67
Cd (ppb)	190.0	540.0	323.2	78.2	0.27	−0.28
Cu	85.5	1483.7	354.8	303.3	2.22	0.91
Fe (%)	1.5	3.7	2.8	0.4	-0.49	−1.11
Mn	17.7	766.4	128.4	140.2	3.17	0.41
Ni (ppb)	100.0	4340.0	1129.2	903.5	2.16	−0.36
Pb	14.0	251.8	50.9	52.1	2.27	0.92
Sb (ppb)	5.0	160.0	36.9	31.2	1.59	−0.17
Zn	42.6	762.8	124.4	111.8	3.96	1.14
West Tails samples ($n = 44$):						
Element	Min	Max	Mean	Std. Dev.	Skewness	Skewness ($\log_e$)
As	196.3	2789.3	625.9	452.0	3.05	0.73
Cd (ppb)	140.0	720.0	275.7	121.8	2.42	1.19
Cu	303.0	2064.7	889.9	476.2	1.12	0.09
Fe (%)	1.4	3.2	2.3	0.4	0.40	−0.06
Mn	15.1	207.6	52.6	34.8	2.40	0.33
Ni (ppb)	60.0	1370.0	371.1	285.6	2.07	−0.06
Pb	40.4	806.9	192.0	169.8	2.24	0.47
Sb (ppb)	5.0	420.0	84.4	76.9	2.75	−0.31
Zn	68.7	776.7	275.2	179.6	1.18	0.06

Table 6.2: Correlation coefficients for the tailings geochemical data. ^{a} denotes values are significant at $\alpha = 0.05$.

East Tails samples ($n = 53$):

	As	Cd	Cu	Fe	Mn	Ni	Pb	Sb	Zn
As	1.000								
Cd	0.106	1.000							
Cu	0.452^a	-0.358^a	1.000						
Fe	0.118	0.904^a	-0.457^a	1.000					
Mn	-0.412^a	0.373^a	-0.096	0.228	1.000				
Ni	-0.485^a	0.591^a	-0.228	0.464^a	0.835^a	1.000			
Pb	0.475^a	-0.513^a	0.576^a	-0.509^a	-0.458^a	-0.596^a	1.000		
Sb	0.429^a	-0.128	0.440^a	-0.191	-0.186	-0.333^a	0.490^a	1.000	
Zn	0.152	-0.373^a	0.596^a	-0.513^a	-0.027	-0.130	0.727^a	0.359^a	1.000

West Tails samples ($n = 44$):

	As	Cd	Cu	Fe	Mn	Ni	Pb	Sb	Zn
As	1.000								
Cd	0.446^a	1.000							
Cu	0.812^a	0.339^a	1.000						
Fe	0.253	0.423^a	-0.013	1.000					
Mn	-0.101	0.412^a	-0.096	0.403^a	1.000				
Ni	-0.300^a	0.279	-0.289	0.418^a	0.902^a	1.000			
Pb	0.626^a	0.315^a	0.481^a	-0.002	-0.075	-0.193	1.000		
Sb	0.623^a	0.445^a	0.655^a	0.287	0.155	0.030	0.506^a	1.000	
Zn	0.550^a	0.432^a	0.610^a	-0.250	-0.113	-0.229	0.645^a	0.464^a	1.000

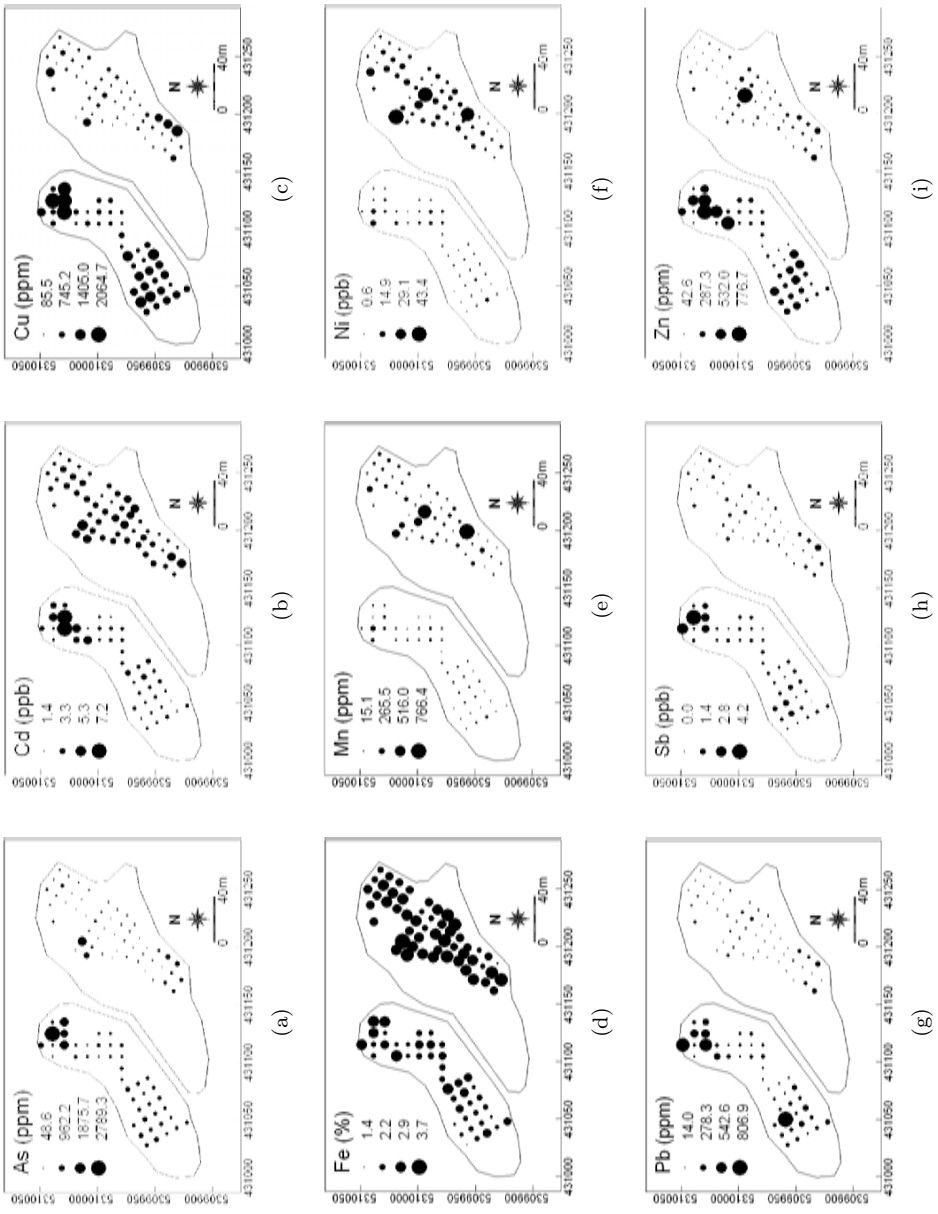


Figure 6.4: Element concentrations in samples from the East Tails and the West Tails.

(abbreviated as A_1) and of Cd-Fe-Ni-Mn (abbreviated as A_2). The A_1 association reflects the polymetallic epithermal deposits in the Lahóca mine. The A_2 association reflects surficial process related to (a) weathering of the pyrite (FeS2) in the mine tailings and (b) scavenging of Cd, Ni, and Mn by weathering products such as Fe-bearing oxides/hydroxides. The presence of the A_1 association in either the East Tails or the West Tails indicates that the tailings were probably derived mostly from a single source, i.e., the Lahóca mine. Based on magnitude of the correlation coefficients, the A_1 association is strongest in the West Tails, whereas the A_2 association is strongest in the East Tails. This suggests that (a) the two Tails represent different geochemical domains, (b) weathering in the East Tails is in a more advanced state than in the West Tails and (c) the East Tails are older than the West Tails. Elements in the A_2 association are enriched in the East Tails (Table 6.1) probably due to metal scavenging or adsorption by secondary Fe-oxide, Fe-hydroxide, and Fe-sulphate-hydrate minerals associated with weathering of the sulphide-bearing mine wastes. The apparent differences between the geochemical characteristics of the East Tails and the West Tails provide motivation to design optimal sampling scheme for new mine waste sites.

6.4 Methods

Since the East Tails and the West Tails have different geochemical characteristics, it was decided to split the data into two sets. The small stream between the East Tails and the West Tails provides a natural boundary to do so. Data from either sub-area are used to model a relationship between heavy metal associations and relative abundances of secondary iron-bearing minerals. The latter data are derived from spectral unmixing of hyperspectral data. A model relationship between heavy metal associations and mineral abundances in one sub-area is then used as basis for optimal sampling design in the same sub-area and in the other sub-area. Division of the area and the data thus provides calibration analysis and prediction/validation analysis for optimal sampling design.

6.4.1 Estimation of mineral abundance

Spectral unmixing of hyperspectral data was performed to estimate relative abundance or proportion, per 5 m pixel, of secondary iron-bearing minerals, with which metals in the mine tailings could be associated. Spectral unmixing is a deconvolution process for estimating proportional contributions of each end-member to spectra. Debba et al. (2005a) suggested that better abundance estimates are obtained if materials, not necessarily of interest but are probably present, and contribute to a pixel spectrum is also included in an end-member set for spectral unmixing. Accordingly, copiapite, jarosite, goethite, ferrihydrite, hematite, kaolinite, anhydrite, gypsum, quartz, and tumbleweed (grass) was considered to consist the end-member set (Figure 6.5). The first five minerals were considered because they are iron-bearing

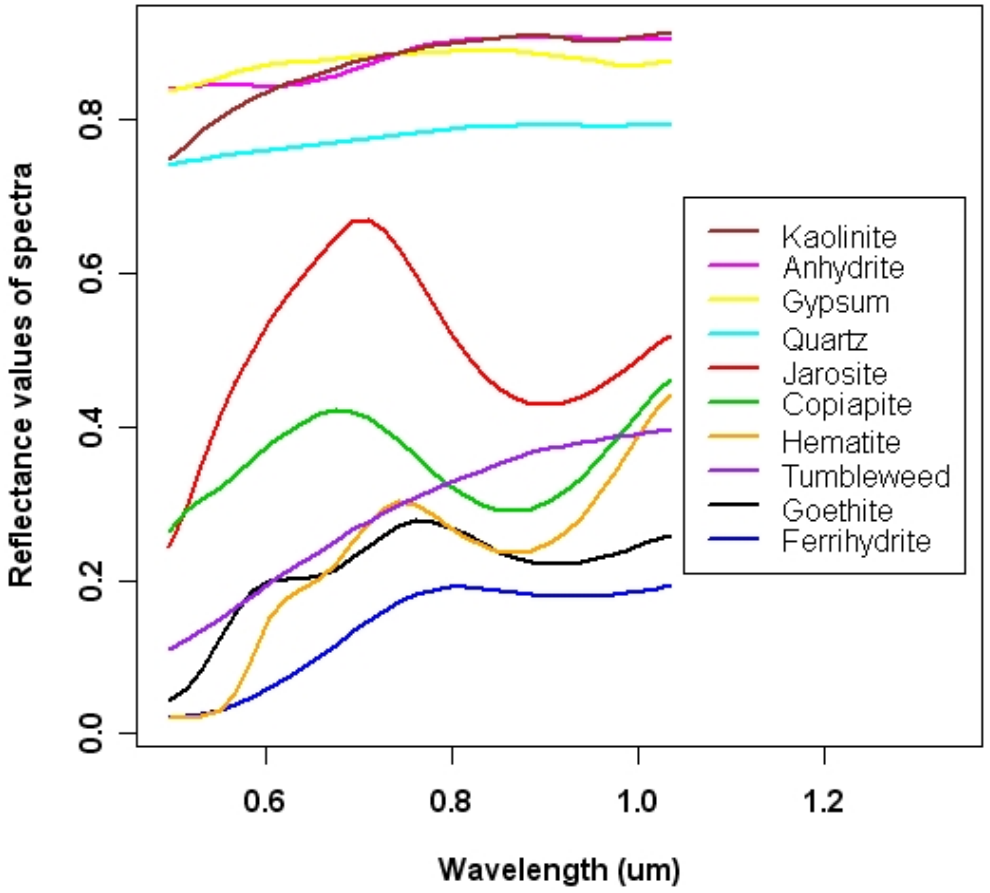


Figure 6.5: Reflectances of minerals which are common in contaminated areas.

secondary minerals that could be derived from weathering of pyrite (FeS_2) and each of them could, although not always, form spectral detectable mineral zones (e.g. pyrite → copiapite → jarosite goethite → ferrihydrite → hematite) around pyrite-rich zones in tailings dumps associated with sulphide-rich zones (Swayze et al., 2000). Kaolinite was considered, because it is a common clay mineral found in mine tailings (Staenz et al., 1999) and it is also known to scavenge metals in the weathering environment (Shahwan et al., 2005). The other three minerals were considered, because they are also often found in mine tailings dumps. Grass was considered because of their presence, albeit minimal, in small patches on the mine tailings dumps. Abundance estimates were determined according to the method of Debba et al. (2005a), which involves minimization of variance of the differences between the second derivative of an estimated spectrum and the second derivative of an actual spectrum. The elementary statistics of abundance estimates of selected end-members are given in Table 6.3.

Table 6.3: Elementary statistics of relative abundance estimates for end-members.

	East Tails Samples ($n = 53$)				West Tails Samples ($n = 44$)			
Endmember	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.
goethite	0.01	0.75	0.38	0.20	0.00	0.73	0.34	0.24
jarosite	0.00	0.15	0.03	0.03	0.00	0.39	0.10	0.10
hematite	0.00	0.70	0.05	0.13	0.00	0.16	0.02	0.03
ferrihydrite	0.00	0.11	0.03	0.03	0.00	0.13	0.03	0.03
kaolinite	0.00	0.18	0.04	0.04	0.00	0.25	0.05	0.06
quartz	0.00	0.08	0.02	0.02	0.00	0.10	0.02	0.02
copiapite	0.00	0.74	0.02	0.10	0.00	0.65	0.05	0.16
gypsum	0.01	0.71	0.35	0.19	0.01	0.73	0.33	0.22
anhydrite	0.00	0.11	0.03	0.03	0.00	0.06	0.02	0.02
tumbelweed	0.00	0.11	0.02	0.02	0.00	0.11	0.02	0.02
	East Tails Hyperspectral ($n = 575$)				West Tails Hyperspectral ($n = 383$)			
Endmember	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.
goethite	0.00	0.75	0.13	0.13	0.00	0.77	0.11	0.14
jarosite	0.00	0.69	0.08	0.11	0.00	0.78	0.13	0.17
hematite	0.00	0.84	0.51	0.19	0.00	0.82	0.50	0.21
ferrihydrite	0.00	0.16	0.02	0.03	0.00	0.18	0.03	0.03
kaolinite	0.00	0.16	0.03	0.03	0.00	0.14	0.03	0.03
quartz	0.00	0.16	0.02	0.02	0.00	0.13	0.02	0.02
copiapite	0.00	0.74	0.01	0.02	0.00	0.65	0.02	0.04
gypsum	0.00	0.81	0.13	0.14	0.00	0.76	0.11	0.15
anhydrite	0.00	0.12	0.02	0.02	0.00	0.14	0.02	0.02
tumbelweed	0.00	0.11	0.02	0.02	0.00	0.14	0.02	0.02

Table 6.4: Factor component analysis with varimax rotation of the heavy metal concentrations.

East Tails ($n = 53$)										
Factor	As	Cd	Cu	Fe	Mn	Ni	Pb	Sb	Zn	var[a]
FA1	0.59	-0.13	0.82	-0.26	-0.02	-0.18	0.72	0.65	0.76	2.65
FA2	0.43	0.91	-0.23	0.91	0.15	0.32	-0.36	0.04	-0.43	2.34
FA3	-0.57	0.31	-0.3	0.13	0.92	0.90	-0.41	-0.17	0.14	2.33
West Tails ($n = 44$)										
Factor	As	Cd	Cu	Fe	Mn	Ni	Pb	Sb	Zn	var
FA1	0.90	0.56	0.86	0.10	-0.08	-0.27	0.76	0.71	0.79	3.65
FA2	0.01	0.58	-0.09	0.71	0.90	0.88	-0.07	0.35	-0.15	2.58

[a] Variance explained by each factor.

The spatial distributions of abundance estimates of the minerals of interest in the optimal sampling design are shown in Figure 6.6. Copiapite is least abundant and hematite is highly abundant in the mine tailings dumps. There are no distinct trends in spatial distributions of ferrihydrite, hematite, and kaolinite. There is, however, a subtle elliptical zone characterized by high jarosite abundance and low goethite abundance extending from the East Tails to the West Tails or vice versa. Outside this elliptical zone, the abundance of jarosite is mostly low but abundance of goethite is mostly intermediate to high. The presence of this elliptical zone suggests that materials in the West Tails and the East Tails are partly, if not wholly, derived from the same source.

6.4.2 Modeling of heavy metal associations

Concentrations of several metals in soils can be estimated using reflectance spectroscopy (Kemper & Sommer, 2002). In addition, geochemical sampling addresses a suite of metals, which reflect intrinsic processes in a system, like a mine tailings dump. It was thus decided to model a heavy metals association reflecting scavenging of metals by secondary iron-bearing minerals in the mine tailings dumps. A factor component analysis with varimax rotation was performed on the logarithmic-transformed heavy metal concentrations to obtain the heavy metal association of interest. The results of the factor component analysis with varimax rotation are shown in Table 6.4.

The first factor (FA1) of heavy metal contents in either the East Tails or the West Tails show high positive loadings mostly on As, Cu, Pb, Sb, and Zn. This heavy metal association reflects the type of mineral deposits that were mined the Recsk-Lahóca area. The As-Cu-Pb-Sb-Zn association therefore reflects the source materials of the mine tailings and/or the relatively unweathered parts of the mine tailings dumps. The second factor (FA2) of heavy metal contents in either the East Tails or the West Tails shows high positive loadings on Fe. The FA2 also show that in either the East Tails or the West Tails there is a common heavy metal association of Fe-Cd-Ni-Mn. This heavy metal association reflects metal scavenging by not only secondary iron-bearing minerals but also secondary manganese bearing minerals. The third factor (FA3) in

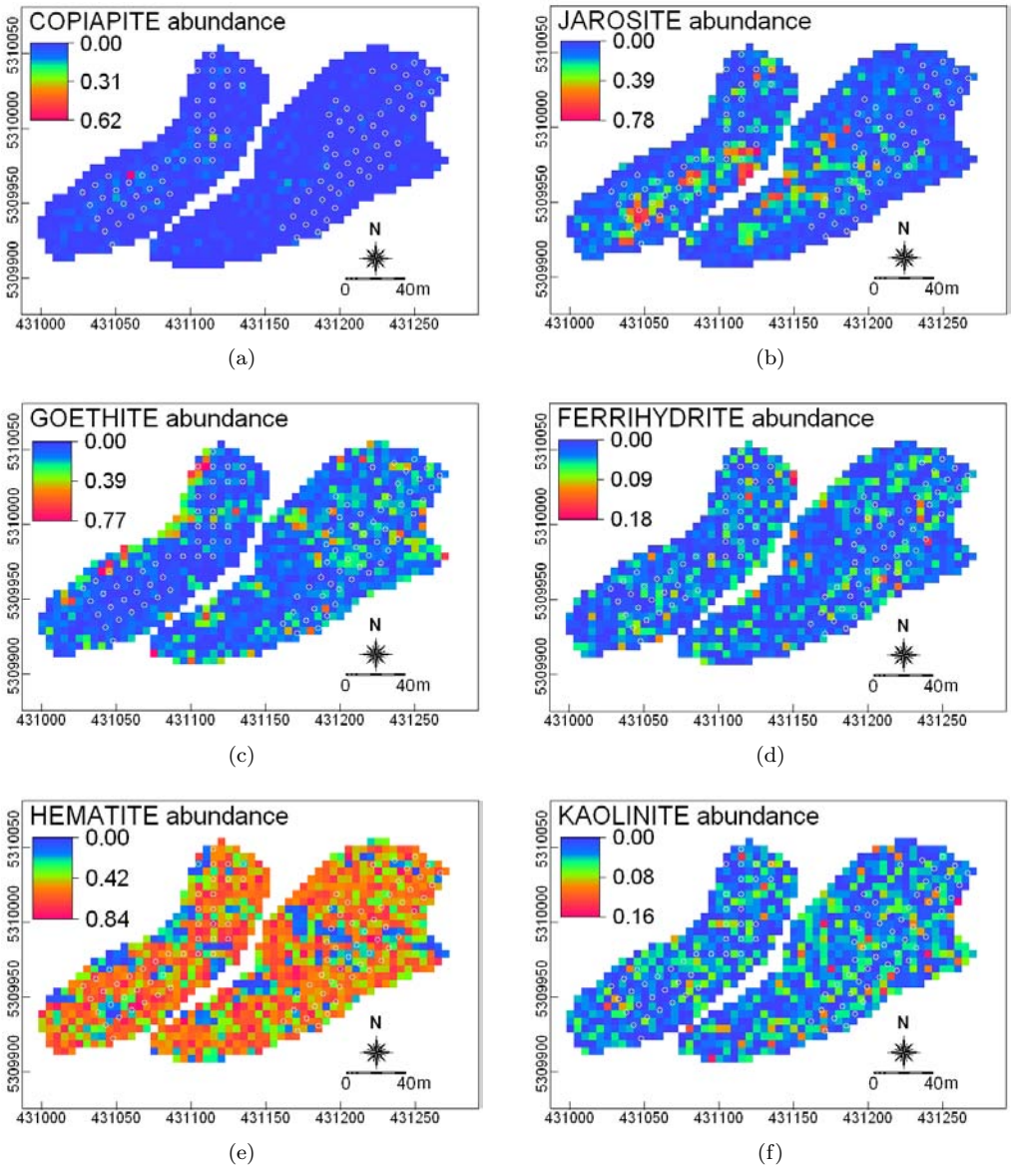


Figure 6.6: Spatial distributions of abundance estimates for minerals of interest to be used in optimal sampling design. Mineral abundance estimates are based on 5 m resolution hyperspectral data. White open circles are locations of mine tailings samples.

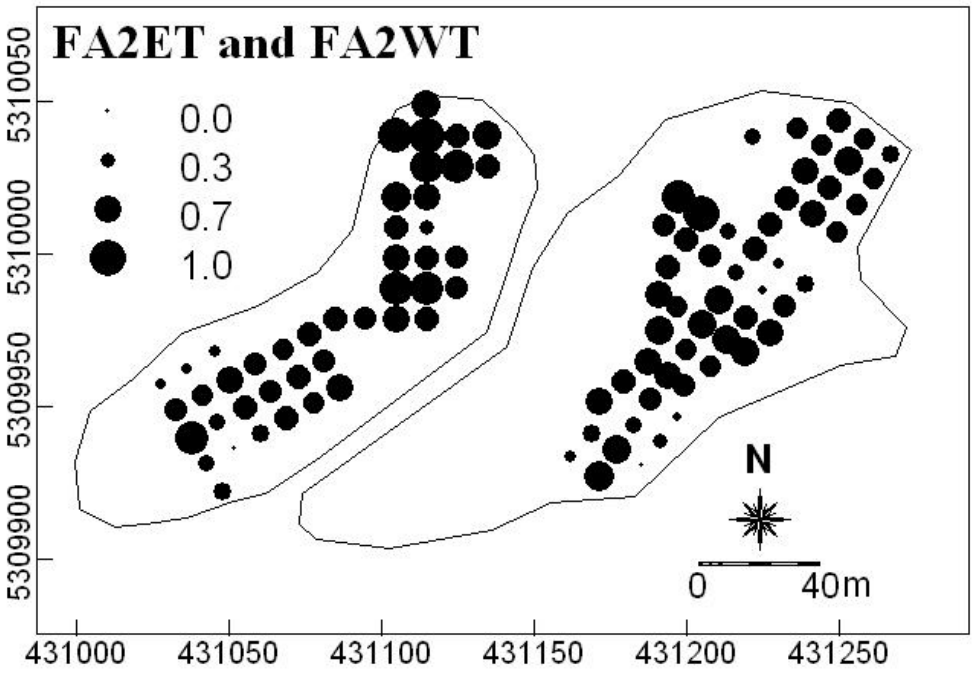


Figure 6.7: Spatial distributions of FA2ET and FA2WT scores. High scores indicate areas of Fe-Cd-Ni-Mn enrichment, whereas low scores indicate areas of As-Cu-Pb-Sb-Zn enrichment.

the East Tails shows high positive loadings on Mn and Ni. This Mn-Ni association mainly reflects scavenging of Ni by secondary manganese oxides/hydroxides.

The FA2 in both the East Tails and the West Tails are considered to represent the heavy metal association of interest. The second factor in the East Tails is labeled as FA2E and the second factor in the West Tails is labeled FA2W. Scores of FA2E AND FA2W were calculated using logarithmic transformed element concentrations and the corresponding factor component loadings from Table 6.4.

The scores of FA2E and FA2W are then linearly transformed to $[0, 1]$, for numerical compatibility with the mineral abundance estimates, and labeled FA2ET and FA2WT, respectively. On the one hand, high scores of FA2ET and FA2WT would indicate parts of the tailings dumps likely to be enriched in Fe, Cd, Ni, and Mn but depleted in As, Cu, Pb, Sb, and Zn. On the other hand, low scores of FA2ET and FA2WT would indicate parts of the tailings dumps likely to be enriched in As, Cu, Pb, Sb, and Zn but depleted in Fe, Cd, Ni, and Mn (Figure 6.7).

6.4.3 Kriging with external drift

Kriging with external drift is applicable to estimate primary variables of interest, which are practically measurable at only few sample sites, based on linearly related ancillary variables, which are measurable at much higher sampling density than the primary variables. Wackernagel (1998) suggests that such ancillary variables can be incorporated into a kriging system as external drift functions. Kriging with external drift is ideal if a primary variable could be measured more precisely and practically at a few locations, whereas possibly less accurate measurements of linearly related ancillary variables are available everywhere in the spatial domain. The present case study applies kriging with external drift to model a relationship between heavy metal association of interest and metal-savenging minerals in the mine tailings dumps. The distribution of heavy metal associations, represented as factor scores, were used as is the primary variable of interest and is based on field sampling. The relative abundances of metal-scavenging iron-bearing minerals, which were obtained from hyperspectral data, are the ancillary variables. Assumption of linear relationship between the primary variable and the ancillary variables is deduced from results of the factor component analysis.

For the modeling, consider $\mathbf{x} \in \mathbf{A} \subset \mathbb{R}^2$ to be a generic data location (x_u, x_v) in 2-dimensional Euclidean space and suppose the domain $Z(\mathbf{x})$ at spatial location $\mathbf{x}$ is a random quantity. The multivariate random field

$$\{Z(\mathbf{x}) : \mathbf{x} \in \mathbf{A}\} \ , \tag{6.1}$$

is generated by letting $\mathbf{x}$ vary over index set $\mathbf{A} \subset \mathbb{R}^2$. A realization of (6.1) is denoted $\{z(\mathbf{x}) : \mathbf{x} \in \mathbf{A}\}$, (Cressie, 1991).

The variogram is defined as the average squared difference between values separated by a given lag h, where h is a vector in both distance and direction, that is,

$$2\gamma(h) = E[Z(\mathbf{x}) - Z(\mathbf{x}+h)]^2 \ . \tag{6.2}$$

Hence the semi-variogram $\gamma(h)$ is defined as:

$$\gamma(h) = \frac{1}{2} E[Z(\mathbf{x}) - Z(\mathbf{x}+h)]^2 \ . \tag{6.3}$$

The experimental semi-variogram $\gamma^\star(h)$, where h is a fixed lag vector in both distance and direction, may be obtained from $\kappa = 1, 2, \ldots, P(h)$ pairs of observations $\{z(\mathbf{x}_\kappa), z(\mathbf{x}_\kappa + h)\}$ at locations $\{\mathbf{x}_\kappa, \mathbf{x}_\kappa + h\}$, as:

$$\gamma^\star(h) = \frac{1}{2 \cdot P(h)} \sum_{\kappa=1}^{P(h)} [z(\mathbf{x}_\kappa) - z(\mathbf{x}_\kappa + h)]^2 \ . \tag{6.4}$$

Suppose that precise measurements are available for a primary variable $Z(\mathbf{x})$ with n_B observations, which is assumed to be a second order random function with known

covariance function $C(h)$, hence the variogram $\gamma(h) = C(h)/C(0)$ is assumed to be known. The k ancillary variables represented as regionalized variables $y_i(\mathbf{x})$, $i = 1, \ldots, k$ with n_A observations, are less accurate measurements covering the whole domain $\mathbf{A}$ at small scale and are considered as deterministic. The values $\{y_i(\mathbf{x})\}$ needs to be known at all locations $\mathbf{x}_\alpha$ of the samples as well as at the nodes of the estimation grid.

Since $Z(\mathbf{x})$ and the set of $\{y_i(\mathbf{x})\}$ are two ways of expressing the same phenomenon, assume that $Z(\mathbf{x})$ is an average equal to a linear function of the set of $\{y_i(\mathbf{x})\}$ up to a constant b_0 and coefficients b_i, $i = 1, \ldots, k$,

$$E[Z(\mathbf{x})] = b_0 + \sum_{i=1}^{k} b_i \cdot y_i(\mathbf{x}) = \sum_{i=0}^{k} b_i \cdot y_i(\mathbf{x}) \ , \tag{6.5}$$

where $y_0(\mathbf{x}) = 1$. The method of merging both sources of information uses $\{y_i(\mathbf{x})\}$ as an external drift function for the estimation of $Z(\mathbf{x})$. The drift of $Z(\mathbf{x})$ is defined externally through the ancillary variables $\{y_i(\mathbf{x})\}$ rather than some smooth version of $Z(\mathbf{x})$ itself, (Chilès & Delfiner, 1999).

Assuming $Z(\mathbf{x})$ is a second order stationary random function, then

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n_A} \lambda_\alpha Z(\mathbf{x}_\alpha) \tag{6.6}$$

where λ_α denotes the weight of the αth observation and is constraint to unit sum.

In estimating the external drift coefficients, the following conditions,

$$\sum_{\alpha=1}^{n_A} \lambda_\alpha y_i(\mathbf{x}_\alpha) = y_i(\mathbf{x}_0) \ , \qquad i = 1, \ldots, k \ , \tag{6.7}$$

are added to the kriging system independently of the inference of the covariance function, hence the term "external". The kriging variance can then be written as

$$\sigma^2_{\mathrm{KED}}(\mathbf{x}_0) = C(0) \left(1 - \sum_{\alpha=1}^{n_A} \lambda_\alpha \gamma(\mathbf{x}_\alpha - \mathbf{x}_0) \right) - \sum_{i=0}^{k} b_i y_i(\mathbf{x}_0) \ . \tag{6.8}$$

The only factor influencing the kriging variance are the variogram $\gamma(h)$, the number of observations n_A, the sampling locations $\mathbf{x}_\alpha$ and the location $\mathbf{x}_0$. This means that the kriging variance does not depend on the observations themselves, but rather only on their relative spacing. The advantage is that in can be used to optimize sampling schemes in advance of data collection.

In this chapter, the location and the covariates as external drift were used to estimate the heavy metal concentration,

$$\begin{aligned} E[Z(\mathbf{x})] \quad = \quad & b_0 + b_1 \cdot x_u + b_2 \cdot x_v + b_3 \cdot \mathrm{GOE}(\mathbf{x}) \\ & + b_4 \cdot \mathrm{JAR}(\mathbf{x}) + b_5 \cdot \mathrm{FER}(\mathbf{x}) + b_6 \cdot \mathrm{HEM}(\mathbf{x}) \\ & + b_7 \cdot \mathrm{KAO}(\mathbf{x}) + b_8 \cdot \mathrm{COP}(\mathbf{x}) \, , \end{aligned} \tag{6.9}$$

namely, a first order polynomial on the coordinates and the abundance estimates of the metal-scavenging minerals. In Equation 6.9, GOE is abundance estimates for goethite, JAR is for jarosite, FER is for ferrihydrite, HEM is for hematite, KAO is for kaolinite, and COP is for copiapite.

6.4.4 Sampling scheme optimization

Sampling schemes derived are based on Equations 6.6 and 6.8 by minimization of a fitness function for a certain criterion. Minimization of a fitness function can be performed exclusive or inclusive of simulated annealing.

Sampling scheme exclusive of simulated annealing

This procedure is performed directly through kriging with external drift by application of a criterion called the Mean Kriging Prediction Error (MKPE), the fitness function of which is defined as,

$$\phi_{\text{MKPE}}(\mathbf{S}) = \frac{1}{n_p} \sum_{\mathbf{x} \in \mathbf{A}} \left\{ Z(\mathbf{x}) - Z^*(\mathbf{x}|\mathbf{S}) \right\}^2 \ , \tag{6.10}$$

where n_p is the number of observations in the sampling scheme, $Z(\mathbf{x})$ is the primary variable at location $\mathbf{x}$ and $Z^*(\mathbf{x}|\mathbf{S})$ is the predicted value at $\mathbf{x}$ (Equation 6.6) for sampling scheme $\mathbf{S}$ with n_p samples. The MKPE-criterion is useful only when data for the primary variable of interest available at all points to be sampled. Hence, the MKPE-criterion is proposed for removal of samples from an existing design to achieve an optimal retrospective sampling scheme. It should be noted that for an existing sampling scheme, the initial MKPE is zero. Removal of an existing sample is expected to result in an increase in MKPE value. For sequential removal of existing samples, the variogram and $Z(\mathbf{x})$ (Equation 6.6) are estimated in a sequence consisting of several steps whereby an existing sample is sought to be removed and finally discarded if its removal contributes to a highest increase in MKPE value. After one sequence with one less observation than an existing sampling scheme, the resulting MKPE value should be closest to zero. Sequential removal is existing samples is carried to attain a desired number of samples in a retrospective design.

Sampling scheme optimization inclusive of simulated annealing

For a two-dimensional area $\mathbf{A}$, let the collection of all possible sampling schemes with n observations on $\mathbf{A}$ be denoted by $\mathbf{S}^n$. A fitness function $\phi(\mathbf{S}^n) : \mathbf{S}^n \rightarrow \mathbb{R}^+$ is minimized through simulated annealing. In terms of sampling, the initial step in simulated annealing is to randomly select a set of points in $\mathbf{A}$. A new point in $\mathbf{A}$ is

then randomly selected and replaces a randomly selected old point from the current collection. This replacement occurs, based on a probabilistic criterion, if the fitness function decreases and if the fitness function increases. Hence, the process allows inferior moves. Initially, the probabilistic criterion is high, allowing a large probability of inferior moves. A parameter in the annealing process is then reduced by a factor of 0.95 at each successive step, thereby decreasing the probability of accepting inferior moves until the process stabilizes. The final solution is independent of the initial random selection of points as the process reaches the global optimum.

In a previously sampled area, estimates of $Z(\mathbf{x})$ (Equation 6.6) by kriging with external drift can be used in simulated annealing to optimize a retrospective sampling scheme using the MKPE-criterion. Cooling schedules are applied to derive the optimal retrospective sampling scheme with fewer samples than an existing sampling scheme. The predictions are at the existing sample sites and the optimal sampling scheme will consist of samples that least contribute to an increase in the MKPE value. At each step, an existing sample is randomly selected and consider for replacement by a random sample not in the scheme based on the probabilistic criterion. When there is no improvement after a specified number of steps the resulting samples form the retrospective scheme.

A variogram model for a previously sampled area can be used to estimate $\sigma^2(\mathbf{x})$ (Equation 6.8) in an unvisited area, if the latter is considered to have similar, if not the same, spatial characteristics as the former. In this case, estimates of $\sigma^2(\mathbf{x})$ in the unvisited area can be used to derive the optimal prospective sampling scheme. The optimization procedure by simulated annealing is then performed by application of a criterion called the Mean Kriging Variance with External Drift (MKVED), the fitness function of which is defined as

$$\phi_{\mathrm{MKVED}}(\mathbf{S}) = \frac{1}{n_{\mathbf{A}}} \sum_{j=1}^{n_{\mathbf{A}}} \sigma^2_{\mathrm{KED}}(\mathbf{x}_{\mathbf{A},j}|\mathbf{S}) \ , \tag{6.11}$$

where $n_{\mathbf{A}}$ is the number of raster nodes for which data for each of the covariates are available. The MKVED-criterion is ideal for deriving optimal prospective sampling schemes, since it does not depend on the actual data for the primary variable. It depends, however, on a covariance function, the spatial configuration of sampling locations and data locations of the covariates (Equation 6.8). Hence, the MKVED-criterion is proposed to derive the optimal prospective sampling scheme in an unvisited area based on a relevant model from a previously sampled area.

6.5 Results

6.5.1 Mineral abundance estimates

Spectral unmixing was performed for all image pixels of the hyperspectral data to estimate relative abundance of secondary metal-scavenging minerals. Table 6.3 con-

tains the summary of the spectral unmixing for end-members that are thought to be present in the mine tailings dumps. Six of the 10 end-members are weathering by-products that influence heavy metal distributions. Abundance estimates for copiapite, jarosite, goethite, ferrihydrite, hematite, and kaolinite were used as covariates in modeling by spatial kriging with external drift.

6.5.2 Optimal retroprective sampling schemes

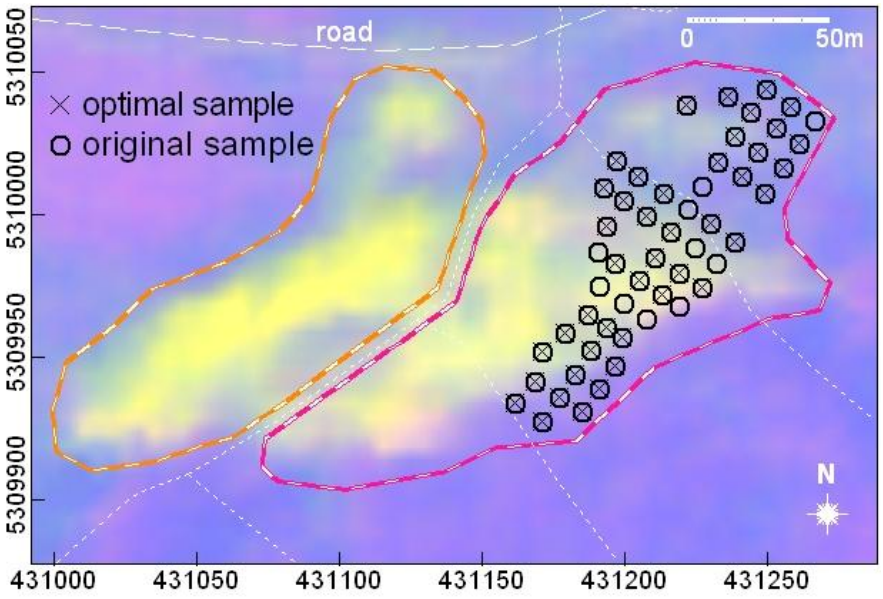
Decision on final number of samples removed from the existing design is dependent on level of prediction error considered and, perhaps, on economic considerations. In this study, it was decided to create retrospective sampling schemes with 10 less samples from an existing sampling scheme. The decision here to remove 10 existing samples in the East Tails and the West Tails is made not on the basis of economic consideration but simply for the purpose of comparing results from non-application of simulated annealing with results from application of simulated annealing.

Via sequential removal of existing samples

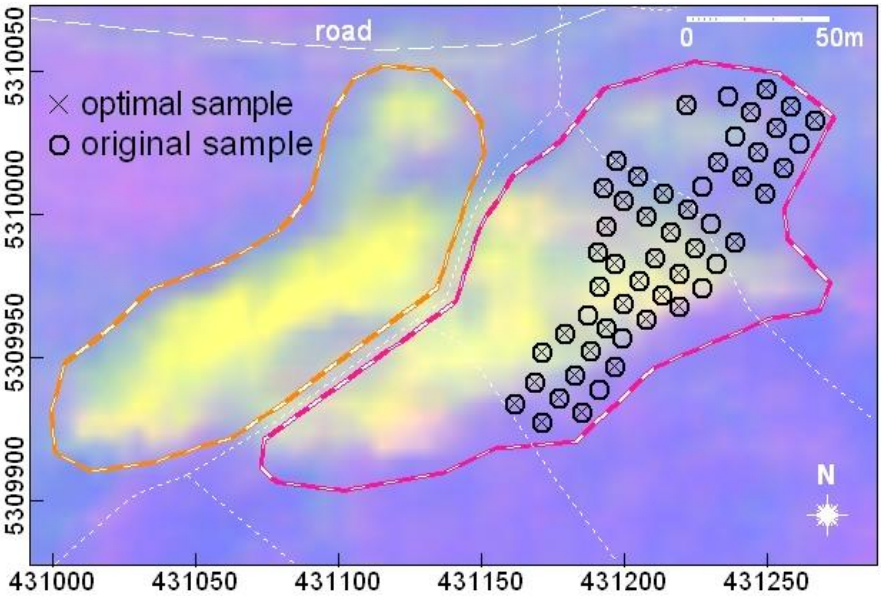
The MKPE-criterion was applied to derive the optimal retrospective sampling scheme for the East Tails and West Tails. In either of these test areas, data of the primary variable and the ancillary variables for existing samples were used for sequential estimation of the variogram and $Z^*(\mathbf{x})$ (Equation 6.6). Ten sequential removals were made to derive an optimal retrospective sampling scheme with 43 samples for the East Tails and an optimal retrospective sampling scheme with 34 samples for the West Tails. For the East Tails (Figure 6.8(a)), remaining 43 samples gave a mean prediction error (Equation 6.10) of 6.34×10^{-7}. For the West Tails (Figure 6.9(a)), remaining 34 samples gave a mean prediction error of 1.36×10^{-6}.

Via simulated annealing

Exactly the same data sets used in sequential removal of sampling were then used in simulated annealing to minimize Equation 6.10 through a cooling schedule to obtain a sampling scheme with 10 samples less than an existing sampling scheme. For the East Tails (Figure 6.8(b)), remaining 43 samples gave a mean prediction error (Equation 6.10) of 2.75×10^{-16}. For the West Tails (Figure 6.9(b)), remaining 34 samples gave a mean prediction error of 1.17×10^{-14}. Clearly, optimized retrospective sampling schemes derived via simulated annealing have considerably lower prediction errors than optimized retrospective samples schemes derived by sequential removal of samples.

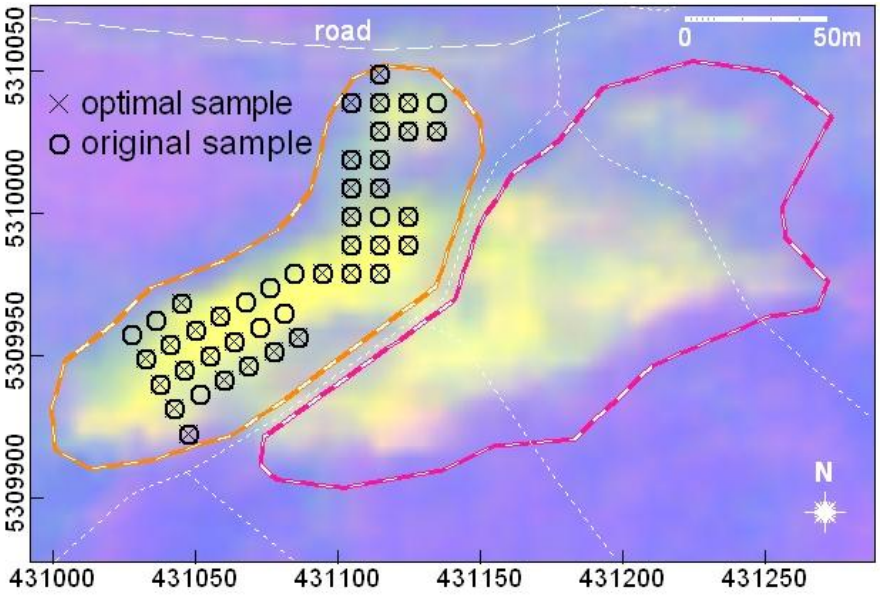


(a) Sequential removal

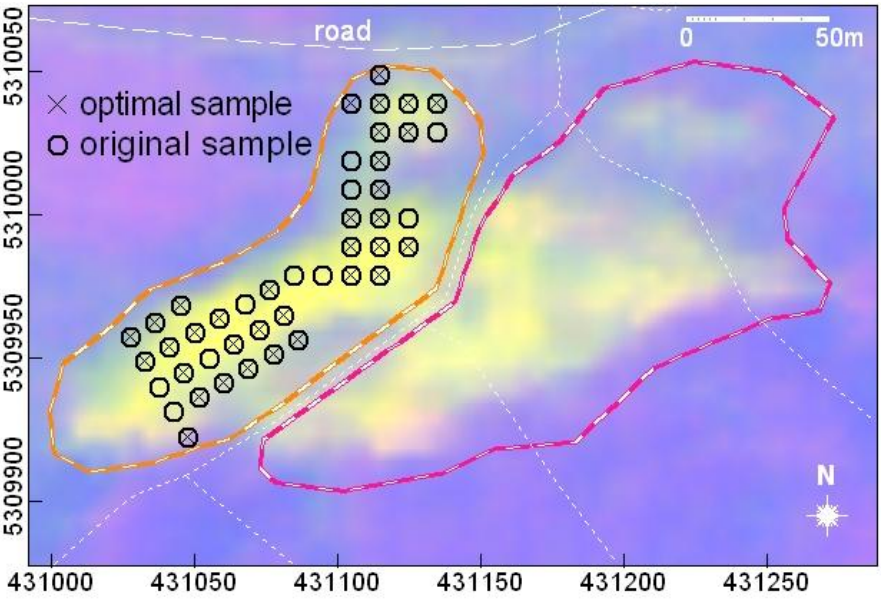


(b) Simulated annealing

Figure 6.8: A retrospective sampling design in the East Tails to compare the method of sequentially removing 10 samples from the current design to the optimal sampling scheme using simulated annealing.

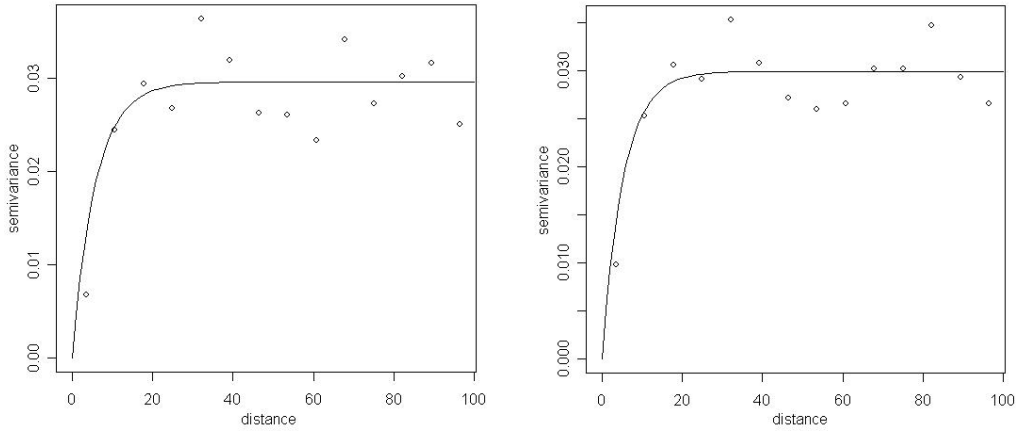


(a) Sequential removal



(b) Simulated annealing

Figure 6.9: A retrospective sampling design in the West Tails to compare the method of sequentially removing 10 samples from the current design to the optimal sampling scheme using simulated annealing.



(a) East Tails (b) East and West Tails combined

Figure 6.10: The exponential variogram for the East Tails data and the East and West Tails data combined.

6.5.3 Optimal prospective sampling schemes

By way of simulated annealing, a prospective sampling scheme for the West Tails is derived based on a model for the East Tails. As an illustration, it was decided to derive a prospective sampling scheme having 30 samples in the West Tails using the 53 samples from the East Tails. The exponential variogram was estimated with the data from the East Tails and is shown in Figure 6.10(a). To verify that this variogram is also appropriate for the West Tails, the East and West Tails data were combined and the variogram is shown in Figure 6.10(b). The similarity of the two variograms indicate that the variogram for the East Tails could be appropriate for modeling the West Tails.

Variogram model for the East Tails was applied to the West Tails data to derive a prospective sampling scheme via simulated annealing to minimize Equation 6.11. The resulting prospective sampling scheme, with 30 samples for the West Tails, is shown in Figure 6.11. The optimal sampling scheme constructed using the kriging external drift variance approach are spread over the West Tails region while retaining some close pairs of samples. These close pair samples are to improve the estimation of the variogram model. The mean kriging with external drift variance for the West Tails, using the combined East and West Tails sampling data, as illustrated in Figure 6.3, is 6.8×10^{-4} for the West Tails. This mean kriging variance was approximately the same when either of the two variograms was used. The optimal sampling scheme resulted in a mean kriging with external drift variance for the West Tails of 3.3×10^{-4} using the variogram derived from the East Tails data. This indicates that the optimal sampling scheme contains samples that reduces the mean kriging with external drift variance for the previously designed grid sampling scheme in the West Tails.

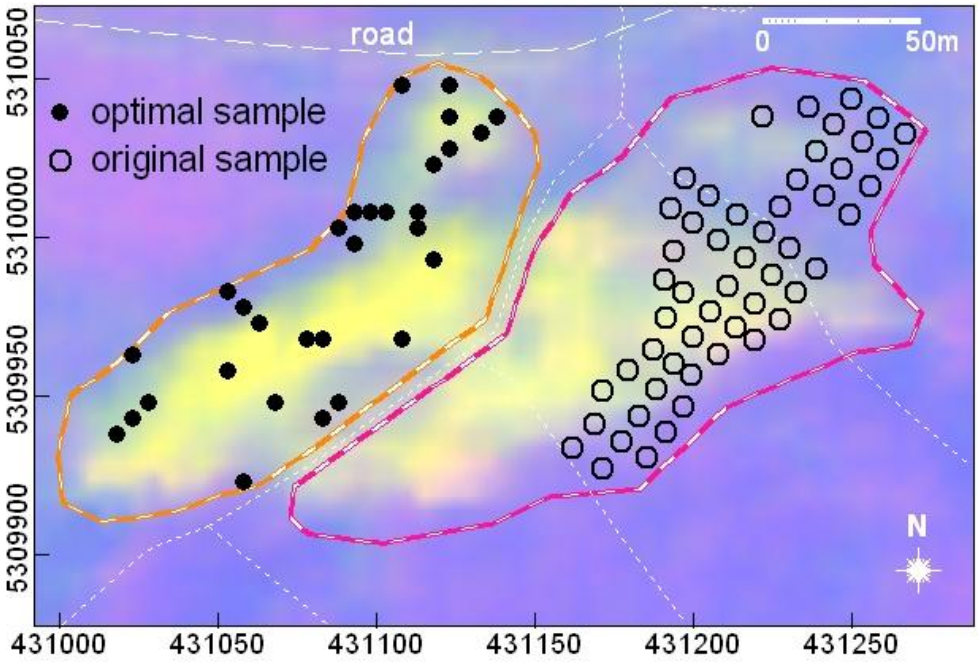


Figure 6.11: Prospective optimal sampling scheme in the West Tails using the samples in the East Tails.

6.6 Discussion

Surface characterization of mine tailings could provide essential information for protection of surrounding ecosystems. Planning where and how many mine tailings samples should be collected is therefore a crucial task. This is so because spatial distributions of undesirable heavy metals in mine tailings must be determined accurately. This study has demonstrated usefulness of airborne hyperspectral data to support optimization of sampling schemes for surface characterization of mine tailings.

Analysis of hyperspectral data can yield information about spatial distributions of secondary iron-bearing and clay minerals associated with weathering of pyrite-rich mine wastes. Heavy metals usually reside in such secondary minerals. Nevertheless, field samples are necessary to model spatial relationships between heavy metals and secondary minerals in mine tailings. For large mine tailings, say in Rio Tinto (Spain), this means necessity to undertake an orientation, preferably grid, sampling program prior to the main sampling program. In an orientation grid sampling program, not only samples of mines tailings should be collected but also measurements by field spectrometers. The latter type of data could also be useful in calibration of airborne hyperspectral data. The data collected in an orientation grid sampling program must allow determination of (a) spatial distributions of heavy metals, (b) spatial distributions of secondary metal-scavenging minerals, (c) spatial relationships between heavy metals and secondary metal-scavenging minerals, and (d) a variogram model of heavy metal associations due to metal-scavenging minerals. These four types of spatial information are essential to optimize a prospective sampling scheme to be carried out during a main sampling program, especially for large mine tailings. These four types of spatial information are also essential to optimize a retrospective sampling scheme in a previously sampled mine tailings area.

This study has shown that spatial relationships between heavy metals and metal-scavenging minerals can be modeled adequately by kriging with external drift. The kriging variance, being dependent only on the variogram, the spatial configuration of the sampling locations and the data locations, could then be used to derive optimal sampling schemes. Certainly, there are other ways by which spatial relationships between co-related variables can be modeled. As examples, spatial regression analysis (Anselin, 1988; Anselin & Bera, 1998) and Bayesian Kriging with external drift (Goovaerts, 1997; Deutsch & Journel, 1998) could be used to model the spatial relationships between co-related variables. It was not intended here, however, to compare performance of these methods. In the case of multivariate primary variables a possible solution might be to use the co-kriging variance, however, this requires modeling of variograms and cross-variograms and can only be applied in cases of few variables (Van Groenigen et al., 2000b). For these reasons, a combined primary variable and several covariates were used.

This study has also shown the efficiency of simulated annealing to optimize a retrospective or a prospective sampling scheme. Although this study is not concerned with economic considerations in sampling scheme optimization, the optimum sam-

pling schemes derived in this study are not equally spaced grids. Experience shows that grid sampling is labor intensive and time-consuming but not necessarily statistically optimal (Burgess et al., 1981; Yfantis et al., 1987; Van Groenigen & Stein, 1998). This suggests that a statistically optimal non-grid sampling scheme derived through the methods described in this study is also economical. This suggestion, however, requires further investigation.

The procedures described in this study are applicable to mine tailings areas in regions with similar climatic and terrain characteristics as the case study area. This is so because surficial geochemistry of heavy metals or surficial chemical stability of minerals are influenced by climatic and terrain factors. Applications of the procedures described here to mine tailings areas in other regions with different climatic and terrain characteristics would require changes in strategies. For example, in arid regions, where chemical weathering is not prevalent, analysis of spatial relationships between heavy metals and secondary iron-bearing minerals becomes trivial. In such a case, however, orientation sampling is still important to calibrate hyperspectral data for prediction of heavy metals (e.g., Kemper & Sommer, 2002) for each pixel. Because sampling addresses a suite of metals, a surface model of heavy metal association is still the desirable input to simulated annealing for optimization of sampling schemes.

6.7 Conclusions

This chapter results in three main conclusions.

- This study demonstrates that designing sampling schemes using simulated annealing results in much better selection of samples from an existing scheme in terms of prediction accuracy.
- The use of secondary information in designing optimal sampling schemes was also illustrated. Often these secondary information can be achieved at a relatively low cost and available over a greater region. These are the primary reasons for incorporating this information into the sampling design.
- Optimized sampling schemes using the mean kriging with external drift variance will result in sampling schemes that explicitly take into account the nature of spatial dependency of the data and together with hyperspectral data can be used to design sampling schemes in nearby unexplored areas.

Chapter 7

Research findings, conclusions and further research recommendations

If I have seen further than others, it is by standing upon the shoulders of giants.

Isaac Newton

7.1 Overview

In the study of spatial distribution of crops or minerals and associated metals, schemes for ground sampling must be optimized so that field data collection is cost-effective but the derived spatial information about primary variables of interest is accurate. On the one hand, for a previously sampled area, samples may be added or removed from an existing ground sampling scheme in order to develop an optimal retrospective sampling scheme. On the other hand, for an unvisited area, an optimal prospective sampling scheme must be developed to provide opportunity to capture accurate information about the primary spatial variables of interest. Although cost-effectiveness of a sampling scheme is not directly addressed in this research, optimal retrospective sampling schemes were developed mostly with fewer samples than in an existing sampling design and optimal prospective sampling schemes developed have few but realistic number of samples, which achieve statistically accurate spatial information.

Optimization of sampling schemes requires relevant a priori information about the population of interest. In this thesis, a priori information was derived from airborne hyperspectral data. Notwithstanding the variety of case studies presented in chapters 2–6, appropriate and optimal sampling schemes were developed by simulated annealing. The methods explained and demonstrated in chapters 2–6 for applications of hyperspectral data and simulated annealing make this research unique as compared to other researches in the same field.

7.2 Research purposes and findings

Major research findings relevant to specific research purposes stated in section 1.5 are described below.

(i) *To formulate optimization criteria that meet the specific demands for ground sampling to study spatial variability of crops or minerals and associated metals.*

In areas with homogeneous and heterogeneous agricultural fields (or segments), the MMSD (Minimization of the Means Shortest Distance) criterion was adapted and found useful in optimization of sampling schemes (chapter 2). Using a segmented hyperspectral image, adaptation of the MMSD-criterion resulted in a balanced distribution of sampling points over the whole segmented image such that more samples are positioned in heterogeneous segments, whereas fewer samples are positioned in homogeneous segments.

For ground verification of remotely sensed information for the presence of certain minerals of interest, the WMSD (Weighted Means Shortest Distance) criterion was adapted and found useful in optimization of sampling schemes (chapter 3). The WMSD-criterion was tested to target areas for ground truthing of a map for a mineral (in this case, alunite) derived from the processing and analysis of hyperspectral data. Rule images, derived from hyperspectral data, were used as weights for the application of the WMSD-criterion. The optimal sampling scheme was achieved by targeting ground locations that (a) have high probability of alunite occurrence and (b) maintain overall distribution of alunite. Application of the WMSD-criterion is not specific to mapping of alunite but also to mapping of any mineral having distinct absorption features.

Application of the WMSD-criterion was also found useful in positioning further mineral exploration targets (chapter 4). The WMSD-criterion was applied to set optimal exploration targets based on the variance of the odds ratio depicted in the prospectivity map. The prospectivity map was created by using evidential maps derived mostly from hyperspectral in weight-of-evidence modeling.

In chapter 5, new criteria for unmixing of hyperspectral data to estimate relative abundance of materials of interest are proposed, namely, SumSpec, VarSpec, SumDeriv, and VarDeriv. Using mixed spectra as inputs, each criterion is optimized to achieve accurate estimates for the relative abundance of materials of interest. It was found that minimization of VarDeriv, which is the differences between the second derivatives of an observed spectrum and the second derivatives of the end-member spectra, results in most accurate relative abundance estimates.

Finally, in chapter 6, two new optimization criteria are proposed. The first criterion, called the MKPE (Mean Kriging Prediction Error), is satisfactory for deriving a retrospective sampling scheme whereby the number of existing samples is reduced. The optimal retrospective sampling scheme derived does not result in much loss of prediction accuracy even when the number of samples

is reduced. The second criterion, called the MKVED (Mean Kriging Variance with External Drift), is satisfactory for deriving a prospective sampling scheme in an area based on a model from an existing sampling scheme in another area. Application of the MKVED-criterion indicates that, given a spatial model of interest in a previously sampled area, an optimal prospective sampling scheme can be developed for a new unvisited area having similar, if not the same, spatial characteristics as the previously sampled area.

(ii) *To formulate new hyperspectral remote sensing techniques useful for deriving optimal sampling schemes.*

A segmented hyperspectral image is considered useful to derive a prospective sampling scheme to study spatial variability of crops because (a) samples in each segment represent the same category of crops, and hence are similar and (b) variability of each segment influence the optimum number and locations of samples for homogeneous segments and for heterogeneous segments. Segmentation using hyperspectral images is hampered by high dimensionality of the data. In chapter 2, therefore, the iterated conditional modes (ICM) algorithm was slightly modified for hyperspectral image segmentation.

Robust estimates of materials of interest from given hyperspectral data is vital to develop a prospective sampling scheme. Rule images, which represent information about type and estimate of materials of interest, can be derived from hyperspectral data by using any established technique. It is demonstrated in chapter 3, however, that a combination of two established techniques, namely, spectral feature fitting (SFF) and spectral angle mapper (SAM), results in rule images that are more robust than the rule images derived by individual techniques. Using weights from the combined rule images produced the best prospective sampling scheme, as compared to prospective sampling schemes derived by using weights from SFF rule images and from SAM rule images.

A new technique of unmixing hyperspectral data to estimate relative abundance of spectrally similar materials in complex solid mixtures, such as mine wastes, is presented in chapter 5. This new technique involves the second derivatives of observed spectrum and the second derivatives of end-member spectra in order to achieve accurate relative abundance estimates of materials of interest. Application of this new technique to hyperspectral data results in abundance estimates of secondary iron-bearing minerals in mine tailings that were useful in developing optimal retrospective and prospective sampling schemes for surface characterization of heavy metals in mine tailings.

(iii) *To incorporate ancillary spatial information, derived hyperspectral data and maps, in optimization of sampling schemes.*

Hyperspectral data were used to create a segmented image (chapter 2) and rule images (chapter 3), which were used to develop the respective optimal sampling schemes. The segmented image allows calculation of the variability of each defined segment and hence the calculation of the sample size for a segment. The rule images allow calculation of the weight function so that sample points

are arranged more intensely in areas with high probability of the mineral of interest.

In chapter 4, maps of spatial distributions of hydrothermal alteration minerals, which are depicted by channel ratio images derived from hyperspectral data, and a map of distance to faults/fractures were used to derive optimal exploration targets. Hyperspectral data provide valuable spatial information for predicting epithermal occurrences in this district, although they need to be supported by other evidences, such as proximity to faults and fractures.

Finally, in chapter 6, mineral abundance estimates derived from hyperspectral data (see chapter 5) were used as covariates to obtain an empirical relationship between the minerals and the heavy metal concentration in order to design optimal retrospective and prospective sampling schemes.

(iv) *To develop and apply optimal sampling schemes to studies of crops or minerals and associated metals.*

Field spectral measurements of agricultural crops could be used, for example, to study health of crops and thus to predict crop yields. Because fields of agricultural crops may be homogeneous (planted with only one particular type of crop) or heterogeneous (planted with a group of different types of crops), the lack of this prior information can make field sampling non-optimal. In chapter 2, the ability to develop optimal prospective sampling schemes based on a given segmented hyperspectral image of an area and with application of simulated annealing is demonstrated to be potentially useful in agricultural studies.

Mineral mapping by hyperspectral remote sensing is becoming an indispensable tool in mineral exploration as well as in geo-environmental studies. Ground-truthing of spatial information derived from hyperspectral data, however, is always necessary. In chapter 3, the ability to develop optimal prospective sampling schemes to support ground verification for the spatial distribution of minerals derived from hyperspectral data is demonstrated by application of simulated annealing.

Surface characterization of mine tailings dumps is important to determine which parts of the environment could be at risk in terms of heavy metal contamination. In chapter 5, given hyperspectral data, the ability to identify and estimate metal-scavenging minerals in mine tailings is demonstrated to be potentially useful in surface characterization of mine tailings dumps. Furthermore, in chapter 6, given a model of spatial relationship between heavy metal associations and metal-scavenging minerals, application of hyperspectral data and simulated annealing to develop optimal retrospective sampling schemes and optimal prospective sampling schemes are demonstrated to be potentially useful for surface characterization of previously sampled mine waste dumps and unvisited mine waste dumps, respectively.

(v) *To validate results by comparing derived optimal sampling schemes with classical sampling schemes or with existing sampling schemes.*

In chapter 2, the optimized sampling scheme was compared to simple random sampling and rectangular grid base sampling. The optimal prospective sampling scheme, derived from a segmented image performed better in estimating the different vegetation indices than either of the other two sampling schemes. The optimal prospective sampling points are more intense in heterogeneous areas, thereby is able to better estimate parameters of interest.

In chapter 3, field measurements of ground spectra were used to validate rule images of a mineral of interest. It is shown that combined application of spectral angle mapper and spectral feature fitting result in accurate classification of high probability of alunite occurrence in certain parts of a study area. A sensitivity analysis was then performed to compare performance of the optimal prospective sampling scheme developed with performance of a stratified random sampling scheme. The optimal prospective sampling scheme, developed from rule images that were created through combined application of spectral angle mapper and spectral feature fitting, performed better. The optimal prospective sampling points are more stable against changes in thresholds and channels selected. In addition, the locations of prospective sampling points correspond closely to the spatial distribution of alunite intensity.

The probabilistic mineral prospectivity mapping presented in chapter 4 was validated by a sub-set of mineral deposit occurrences not used in weights-of-evidence modeling. The probabilistic mineral prospectivity map showed a prediction rate of 64%, which is considered adequate because of the limited number of evidential maps used. Buffer zones around focal points of exploration targets, which were optimally located by simulated annealing using the probabilistic model, contain nine of 14 (assumed) undiscovered mineral deposit occurrences. Analysis of prioritized exploration targets further show that at least one validation deposit occurrence is present in (a) 60% of the top 5 priority exploration targets and (b) 50% of the top 10 priority exploration targets. These validation results suggest that optimal positioning of focal points of exploration targets, based on a probabilistic mineral prospectivity model and application of simulated annealing, could potentially guide further mineral prospecting towards mineral deposit discovery.

In chapter 5, the known quantities of each mineral used to create a mixed spectrum were used to validate results of new spectral unmixing technique via simulated annealing. The validation results indicate, on one hand, the accuracy of relative abundance estimates is not diminished even if the new spectral unmixing technique uses a set of end-members inclusive of end-members not contributing to a mixed spectrum. On the other hand, the validation results indicate the accuracy of relative abundance estimates is diminished if the new spectral unmixing technique uses a set of end-members exclusive of end-members contributing to a mixed spectrum.

The optimal retrospective sampling scheme, presented in chapter 6, used a previously designed grid sampling scheme to measure its prediction accuracy. It was shown that deleting appropriate sampling points from an existing sampling scheme did not significantly decrease the prediction accuracy. The optimal ret-

rospective sampling scheme, by means of simulated annealing, was also demonstrated to be superior to the method of sequentially removing sampling points from the existing grid sampling scheme. The method using simulated annealing resulted in a lower mean prediction error. The optimal prospective sampling scheme, presented in chapter 6, for the West Tails similarly used the previously designed grid sampling scheme. In this case, the mean kriging variance with external drift for the optimal prospective sampling scheme was approximately half the mean kriging with external drift for the present grid sampling scheme in the West Tails.

7.3 Research conclusions

This research resulted in the following overall conclusions:

- With regard to specific research purpose (i), the various optimization criteria formulated are satisfactory in optimization of sampling schemes to study spatial variability of crops or minerals and associated metals.

  Experiences outside this research have shown that regular grid sampling is cost- and labor-intensive. In chapters 2–6, it is evident that the optimized sampling scheme is neither close to a regular grid design nor randomly selecting samples over the area. In chapter 2, this issue is observed as the sampling is governed by the segments and with irregular segments a grid design will not be optimal. In chapter 3 it was also demonstrated that the optimal sampling scheme was more accurate than a stratified random design. It is further demonstrated in chapter 6 that, based on optimal retrospective sampling designs, not all the existing samples are required in order to achieve the same prediction accuracy. This has great economic benefits as sampling costs and field research are extremely high.

- With regard to specific research purpose (ii), it is imperative to either adopt or adapt established techniques and to develop novel techniques for analysis of hyperspectral data in order to achieve accurate spatial information vital to the optimization of sampling schemes.

  Since the advent of hyperspectral remote sensing about two decades ago, several techniques have already been established to process and analyse hyperspectral data for specific studies. Therefore, in chapter 2 the ICM algorithm was adopted for hyperspectral image segmentation, whereas the band ratio technique was adopted in chapter 4 to create images depicting estimates of groups of minerals. In chapter 3, however, it shown that combining results of spectral feature fitting and of spectral angle mapping provides better weight functions for optimization of sampling schemes. In chapter 5, development of a novel technique for unmixing hyperspectral data was prompted by the difficulty in estimating relative abundances of spectrally similar minerals in complex mixtures such as mine wastes, which is not satisfactorily addressed by existing unmixing techniques.

- With regard to specific research purpose (iii), hyperspectral data are useful in deriving robust spatial ancillary information of interest vital to the optimization of sampling schemes.

  Other spatial variables that affect spatial variability of primary spatial variable of interest must be considered and integrated in sampling scheme optimization. Estimates of such spatial variables can be obtained from hyperspectral data. This was demonstrated in chapter 4, whereby multiple evidences derived from hyperspectral data were integrated in weights-of-evidence modeling. The derived posterior probability was combined with the variance of the odds ratio to satisfy the sampling objective. In chapter 6, various spatial factors were integrated into a single value to be optimized. Factor component analysis was used to integrate the geochemical factors, which was then modeled with covariates derived from hyperspectral data. The external drift kriging predictions and variance were then optimized.

- With respect to specific research purpose (iv), the methods for optimization of sampling schemes developed in this research, by application of hyperspectral data and simulated annealing, are potentially useful to precision agriculture, mineral exploration, and remediation/rehabilitation of mine wastes.

  Designing sampling schemes by using hyperspectral data has the advantage of availing spatial information at virtually all points on the ground in remote and unexplored areas. Ground sampling of control area is, however, still necessary in order to calibrate statistical spatial correlation between information derived from ground data and information derived from hyperspectral data. The objective of sampling is crucial to its application. For an optimized sampling scheme, the sampling objective has to be transformed to a numerical value to be optimized. Often several variables affecting the sampling design have to be incorporated into an objective function to form a single value to be optimized. This dependents on the application. Sometimes this is impossible and it is necessary to compromise.

- With respect to specific research purpose (v), propositions of potential usefulness of methods for optimization of sampling schemes developed in this research are supported by validation of results provided in each case study.

  It is necessary, however, to further test these methods for optimization of sampling schemes in the fields of study in this research or in other fields of study in order to fully reveal their usefulness.

7.4 Recommendation for further research

Several issues, in terms of designing optimal sampling schemes, still need addressing. The following are directly related to this thesis.

- An important optimal sampling design for improving classification accuracy of remote sensing images still remains unsolved. In this case, samples should fall in areas of high uncertainty.
- This thesis assumed that every point on the ground is accessible, however, there is a need to develop algorithms to incorporate quantitative expert knowledge, sampling costs and to handle complex sampling barriers.
- Chapter 5 contains the results for unmixing a mixed spectrum that was generated without any error. In remote sensing the mixed pixel contains various sources of error. For random error, the spectrum can be smoothed before the unmixing procedure is applied. However, for systematic error which might not even be additive and could be multiplicative, the proposed technique will fail.
- Scaling is another issue that has been ignored in this thesis, yet it is an important factor since a pixel is represented by an area on the ground whereas a sample is a point on the ground.
- Chapter 6 considers multivariate primary variables that was combined to form a single primary variable. A possible alternative might be to use the co-kriging variance. This requires modeling of variograms and cross-variograms which could complicate the model.
- Further testing of these methods for optimization of sampling schemes in these fields of study or in other fields of study, are necessary in order to fully appreciate their usefulness.

Of the making of books, there is no end.

Ecclesiastes

Bibliography

Aarts, E. & Korst, J. (1989). *Simulated Annealing and Boltzmann Machines*. New York: John Wiley.

Abrams, M., Ashley, R., Rowan, L., Goetz, A. F. H., & Kahle, A. (1977). Use of imaging in the .46–2.36 μm spectral region for alteration mapping in the Cuprite mining district, Nevada: USGS OFR-77-585.

Abrams, M. J. (1984). Landsat-4 thematic mapper and thematic mapper simulator data for a porphyry copper. *Photogrammetric Engineering and Remote Sensing*, 50, 1171–1173.

Agterberg, F. P. & Bonham-Carter, G. F. (1999). Logistic regression and weights of evidence modeling in mineral exploration. In *Proceedings of 28th International Symposium on Computer Applications in the Mineral Industries* (pp. 483–590).

Agterberg, F. P., Bonham-Carter, G. F., & Wright, D. F. (1990). Statistical pattern integration for mineral exploration. In G. Gaal & D. F. Merriam (Eds.), *Computer Applications in Resource Exploration and Assessment for Minerals and Petroleum* (pp. 1–21). Pergamon, Oxford.

Agterberg, F. P. & Cheng, W. (2002). Conditional independence test of weights-of-evidence modeling. *Natural Resources Research*, 11(4), 249–255.

Allum, J. A. E. (1966). *Photogeology and regional mapping*. Oxford : Pergamon Press.

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht: Kluwer.

Anselin, L. & Bera, A. K. (1998). *Handbook of applied economic statistics*, chapter Spatial dependence in linear regression models with an introduction to spatial econometrics. New York: Marcel Dekker.

Arribas, Jr., A., Cunningham, C. G., Rytuba, J. J., Rye, R. O., Kelley, W. C., Podwysocki, M. H., McKee, E. H., & Tosdal, R. M. (1995). Geology, geochronology, fluid inclusions, and isotope geochemistry of the Rodalquilar gold alunite deposit, Spain. *Economic Geology*, 90, 795–822.

Audry, S., Blanc, G., & Schäfer, J. (2005). The impact of sulphide oxidation on dissolved metal (Cd, Zn, Cu, Cr, Co, Ni, U) inputs into the Lot-Garonne fluvial system (France). *Applied Geochemistry*, 20(5), 919–921.

Ausmus, B. S. & Hilty, J. W. (1972). Reflectance studies of healthy, maize dwarf mosaic virusinfected, and Helminthosporium maydis-infected corn leaves. *Remote Sensing of Environment*, 2, 77–81.

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

Baret, F. & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, 35, 161–173.

Ben-Dor, E. & Banin, A. (1994). Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils. *Remote Sensing of Environment*, 48(3), 261–274.

Ben-Jemaa, F., Mariño, M., & Loaiciga, H. (1995). Sampling design for contaminant distribution in lake sediments. *Journal of Water Resource Planning Management*, 121, 71–79.

Berman, M. & Diggle, P. J. (1989). Estimating weighted integrals of the second-order intensity of spatial point patterns. *Journal of the Royal Statistical Society B*, 51, 81–92.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Royal Statistical Society*, B-48(3), 259–302.

Bishop, M. M., Fienberg, S. E., & Holland, P. W. (1975). *Disctete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge Massachusetts.

Boardman, J. W. (1989). Inversion of imaging spectrometry data using singular value decomposition. In *Proceedings 12th Canadian Symposium on Remote Sensing IGARSS89, SPIE Vol. 4* (pp. 2069–2072).

Boardman, J. W. (1993). Automated spectral unmixing of AVIRIS data using convex geometry concepts. In R. Green (Ed.), *Summaries 4th Annual JPL Airborne Geoscience Workshop, JPL Publication 93-26*, volume 1 (pp. 11–14).

Boardman, J. W., Kruse, F. A., & Green, R. O. (1995). Mapping target signatures via partial unmixing of AVIRIS data. In *Proceedings of the Fifth JPL Airborne Earth Science Workshop, JPL Publication 95-1 (Pasadena, USA: NASA JPL)*, volume 1 (pp. 22–26).

Bohachevsky, I. O., Johnson, M. E., & Stein, M. L. (1986). Generalized simulated annealing for function optimization. *Technometrics*, 28(3), 209–217.

Bonham-Carter, G. F. (1994). *Geographic Information Systems for Geoscientists: Modelling with GIS*, volume 13 of *Computer Methods in the Geosciences*. New York, Pergamon.

Bonham-Carter, G. F., Agterberg, F. P., & Wright, D. F. (1988). Integration of geological datasets for gold exploration in Nova Scotia. *Photogrammetry and Remote Sensing*, 54, 1585–1592.

Bonham-Carter, G. F., Agterberg, F. P., & Wright, D. F. (1989). Weights of evidence modelling: a new approach to mapping mineral potential. In F. P. Agterberg & Bonham-Carter (Eds.), *Statistical Applications in the Earth Sciences: Geological Survey Canada Paper 89-9* (pp. 171–183).

Bouman, B. A. M. & Uenk, D. (1992). Crop classification possibilities with radar in ERS-1 and JERS-1 configuration. *Remote Sensing of Environment*, 40, 1–13.

Brisco, B., Brown, R. J., & Manore, M. J. (1989). Early season crop discrimination with combined SAR and TM data. *Canadian Journal of Remote Sensing*, 15(1), 44–54.

Broge, N. H. & Leblanc, E. (2000). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76, 156–172.

Brown, L., Jin, M. C., Lablanc, S. G., & Cihlar, J. (2000). A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote Sensing Environment*, 71, 16–25.

Brus, D. J. & de Gruijter, J. J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80, 1–44.

Brus, D. J., Spatjens, L. E. E. M., & de Gruitjer, J. J. (1999). A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. *Geoderma*, 89, 129–148.

Burgess, T. M., Webster, R., & McBratney, A. B. (1981). Optimal interpolation and isarithmic mapping of soil properties. IV Sampling strategy. *Journal of Soil Science*, 32(4), 643–660.

Carranza, E. J. M. & Hale, M. (2002). Mineral imaging with landsat thematic mapper data for hydrothermal alteration mapping in heavily vegetated terrane. *International Journal of Remote Sensing*, 23(22), 4827–4852.

Carranza, E. J. M. & Hale, M. (2003). Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. *Ore Geology Reviews*, 22(1), 117–132.

Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, 15, 697–703.

Chabrillat, S., Goetz, A. F. H., Olsen, H. W., Krosley, L., & Noe, D. C. (1–3 March 1999). Use of AVIRIS hyperspectral data to identify and map expansive clay soils in the front range urban corridor in Colorado. In *Proceedings of the 13th International Conference on Applied Geologic Remote Sensing*, I (pp. 390–397).: Vancouver, British Columbia, Canada.

Chang, C.-I. (2006). *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer.

Chappelle, E. W., Kim, M. S., & McMurtrey III, J. E. (1992). Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and the carotenoids in soybean leaves. *Remote Sensing of Environment*, 39, 239–247.

Chen, J. (1996). Evaluation of vegetation indices and modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22, 229–242.

Chen, J. & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat Thematic Mapper. *Remote Sensing of Environment*, 55, 153–162.

Chilès, J.-P. & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty*. John Wiley & sons, INC.

Christakos, G. & Olea, R. A. (1992). Sampling design for spatially distributed hydrogeologic and environmental processes. *Advances in Water Resources*, 15(4), 219–237.

Chung, C. F. & Agterberg, F. P. (1980). Regression models for estimating mineral resources from geological map data. *Mathematical Geology*, 12(5), 473–488.

Clark, R. N. (1995). *Rock Physics and Phase Relations: Handbook of Physical Constants*, chapter Reflectance spectra, (pp. 178–188). AGU Reference Shelf 3, American Geophysical Union. Washington D. C.

Clark, R. N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy. In A. Rencz (Ed.), *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, volume 3 chapter 1, (pp. 3–58). John Wiley and Sons, New York.

Clark, R. N. & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89, 6329–6340.

Clark, R. N. & Swayze, G. A. (1995). Mapping minerals, amorphous materials, environmental materials, vegetation, water, ice, and snow, and other materials: The USGS Ticorder Algorithm. In *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, volume 1 (pp. 39–40).: JPL Publication 95-1.

Clark, R. N., Swayze, G. A., & Gallagher, A. (1992). Mapping the mineralogy and lithology of Canyonlands. Utah with imaging spectrometer data and the multiple spectral feature mapping algorithm. In *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, volume 1 (pp. 11–13).: JPL Publication 92-14.

Clark, R. N., Swayze, G. A., Gallagher, A. J., King, T. V. V., & Calvin, W. M. (1993). 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.

Clark, R. N., Swayze, G. A., Gorelick, N., & Kruse, F. A. (1991). Mapping with imaging spectrometer data using the complete band shape least-squares algorthm simultaneously fit to multiple spectral features from multiple materials. In *Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) workshop* (pp. 2–3).: JPL Publication 91-28.

Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., Sutley, S. J., Dalton, J. B., McDougal, R. R., & Gent, C. A. (2003). Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysical Research*, 108(E12), 5–1–5–44.

Craig, M. D. (1994). Minimum volume transforms for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 663–670.

Cressie, N. (1991). *Statistics for Spatial Data*. John Wiley and Sons, Inc., New York.

Crósta, A. P., Sabine, C., & Taranik, J. V. (1998). Hydrothermal alteration mapping at Bodie, California, using AVIRIS hyperspectral data. *Remote Sensing of Environment*, 65(3), 309–319.

Crowley, J. K., Williams, D. E., Hammarstrom, J. M., Piatak, N., Chou, I. M., & Mars, J. C. (2003). Spectral reflectance properties (0.4–2.5 μm) of secondary Fe-oxide, Fe-hydroxide, and Fe-sulphate-hydrate minerals associated with sulphide-bearing mine wastes. *Geochemistry: Exploration, Environment, Analysis*, 3(2), 219–228.

Cudahy, T., Okada, K., & Brauhart, C. (2000). Targeting VMS-style Zn mineralisation at Panorama, Australia, using airborne hyperspectral VNIR-SWIR HyMap data. In *ERIM Proceedings of the 14th International Conference on Applied Geologic Remote Sensing* (pp. 395–402).: Las Vegas.

Cunningham, C. G., Arribas, Jr., A., Rytuba, J. J., & Arribas, A. (1990). Mineralized and unmineralized calderas in Spain; Part I, evolution of the Los Frailes Caldera. *Mineralium Deposita*, 25 [Suppl.], S21–S28.

Curran, P. J. & Atkinson, P. M. (1998). Geostatistics and remote sensing. *Progress in Physical Geography*, 22(1), 61–78.

Curran, P. J. & Williamson, H. D. (1985). The accuracy of ground data used in remote-sensing investigations. *International Journal of Remote Sensing*, 6(10), 1637–1651.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., Brown de Colstoun, E., & McMurtrey III, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74, 229–239.

de Gruijter, J. J. & ter Braak, C. J. F. (1990). Model-free estimation from spatial samples: A reappraised of classical sampling theory. *Mathematical Geology*, 22(4), 407–415.

Debba, P., Carranza, E. J. M., van der Meer, F. D., & Stein, A. (2005a). Abundance estimation of spectrally similar materials in mine wastes using hyperspectral data and simulated annealing. *IEEE Geoscience and Remote Sensing (In Review)*.

Debba, P., van Ruitenbeek, F. J. A., van der Meer, F. D., Carranza, E. J. M., & Stein, A. (2005b). Optimal field sampling for targeting minerals using hyperspectral data. *Remote Sensing of Environment*, (pp. 373–386).

Deutsch, C. V. & Journel, A. G. (1998). *GSLIB Geostatistical Software LIBrary and User's Guide*. Oxford University Press, New York, second edition.

Diggle, P. & Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33, 53–64.

Drake, N. A., Mackin, S., & Settle, J. J. (1999). Mapping vegetation, soils, and geology in semiarid shrublands using spectral matching and mixture modeling of SWIR AVIRIS imagery. *Remote Sensing of Environment*, 68, 12–25.

Driscoll, R. S. & Coleman, M. D. (1974). Color for shrubs. *Photogrammetric Engineering and Remote Sensing*, 40, 451–459.

Eardley, A. J. (1942). *Aerial photographs: their use and interpretation*. New York: Harper.

Everitt, J. H., Gerbermann, A. H., Alaniz, M. A., & Bowen, R. L. (1980). Using 70 mm aerial photography to identify rangeland sites. *Photogrammetric Engineering and Remote Sensing*, 46, 1339–1348.

Everitt, J. H., Richardson, A. J., Gerbermann, A. H., Wiegand, C. L., & Alaniz, M. A. (1979). Landsat-2 data for inventorying rangelands in south Texas. In *Proceedings of the 5th Symposium Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, Ind.* (pp. 132–141).

Farrand, W. H. & Harsanyi, J. C. (1997). Mapping the distribution of mine tailings in the Coeur dAlene River valley, Idaho, through the use of a constrained energy minimization technique. *Remote Sensing of Environment*, 59, 64–76.

Fassnacht, K. S., Gower, S. T., MacKenzie, M. D., Nordheim, E. V., & Lillesand, T. M. (1997). Estimating the leaf area index of north central Wisconsin forest using Landsat Thematic Mapper. *Remote Sensing Environment*, 61, 229–245.

Ferrier, G. (1999). Application of imaging spectrometer data in identifying environmental pollution caused by mining at Rodaquilar, Spain. *Remote Sensing of Environment*, 68(2), 125–137.

Forbes, F. & Peyrard, N. (2003). Hidden Markov Random Field Model selection criteria based on mean field-like approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9), 1089–1101.

Fraley, C. & Raftery, A. E. (2003). *Model-Based Clustering for Image Segmentation and Large Datasets Via Sampling*. Technical Report 424, University of Washington, Department of Statistics.

Gamon, J. A., Penuelas, J., & Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 41, 35–44.

García-Haro, F. J., Gilabert, M. A., & Meliá, J. (1996). Linear spectral mixture modelling to estimate vegetation amount from optical spectral data. *International Journal of Remote Sensing*, 17, 3373–3400.

Gilbertson, B., Longshaw, T. G., & Viljoen, R. P. (1976). Multispectral aerial photography as exploration tool. iv-v - an applications in the Khomas Trough region, South Africa; and cost effective analysis and conclusions (for mineral exploration). *Remote Sensing of Environment*, 5(2), 93–107.

Gitelson, A., Rundquist, D., Derry, D., Ramirez, J., Keydan, G., Stark, R., & Perk, R. (2001). Using remote sensing to quantify vegetation fraction in corn canopies. In *Proceedings of Third Conference on Geospatial Information in Agriculture and Forestry, Denver, Colorado, 3-7 November 2001.*

Goetz, A. F. H. & Srivastava, V. (1985). Mineralogical mapping Cuprite mining district, Nevada. In G. Vane & A. Goetz (Eds.), *Proc Airborne imaging spectrometer data analysis workshop*, Jet Propulsion Laboratory Publication 85-41 (pp. 22–31).

Goovaerts, P. (1997). *Geostatistics For Natural Resources Evaluation.* Oxford University Press.

Green, A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74.

Gupta, R. P. (2003). *Remote Sensing Geology.* Springer-Verlag New York, LLC, second edition.

Haboudanea, D., Millera, J. R., Patteyc, E., Zarco-Tejadad, P. J., & Strachane, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90, 337–352.

Hapke, B. (1993). Combined theory of reflectance and emittance spectroscopy. In C. Pieters & P. A. J. Englert (Eds.), *Remote Geochemical Analysis: Elemental and Mineralogical Composition* (pp. 31–42). Cambridge University Press, Cambridge, UK.

Hartigan, J. & Wong, M. (1979). A K-means clustering algorithm. *Applied Statistics*, 28, 100–108.

Hedenquist, J. S., Arribas, A. R., & Gonzalez-Urien, E. (2000). Exploration for epithermal gold deposits. *Reviews in Economic Geology*, 13, 245–277.

Houston, R. S. (1973). Geologic mapping using space images. *Contributions to geology*, 12(2), 77–79.

IGME (1981). Mapa Geologico de Espaa (Carboneras, 1.046/24-43; El Pozo de los Frailes, 1.060/24-44), e. 1:50,000. Instituto Geologico y Minero de Espaa (IGME), Servicio de Publicaciones, Ministerio de Industria y Energia, Madrid.

Iranpanah, A. (1977). Geologic applications of Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, 43, 1037–1040.

Jimenez, L. O. & Landgrebe, D. A. (1999). Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1), 2653–2667.

Johnson, P. L. (1969). *Remote sensing in ecology*. University of Georgia Press, Athens, GA.

Kardeván, P., Vekerdy, Z., Róth, L., Sommer, S., Kemper, T., Jordan, G., Tamás, J., Pechmann, I., Kovács, E., Hargitai, H., & László, F. (2003). Outline of scientific aims and data processing status of the first Hungarian hyperspectral data acquisition flight campaign, HYSENS 2002 Hungary. In M. Habermeyer, A. Mülle, & S. Holzwarth (Eds.), *Proceedings of the 3rd EARSeL workshop on imaging spectroscopy, Herrsching, Germany* (pp. 324–332).: EARSeL.

Kemper, T. & Sommer, S. (2002). Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental Science and Technology*, 36(12), 2742–2747.

Kirkpatrick, S., Gelatt, J. C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1, 155–159.

Kowalik, W. S., Lyon, R. J. P., & Switzwe, P. (1983). The effect of additive reaidance terms on ratios of Landsat data (for mineral exploration). *Photogrammetric Engineering and Remote Sensing*, 49, 659–669.

Kruse, F. A. (2002). Comparison of AVIRIS and Hyperion for hyperspectral mineral mapping. In *SPIE Aerospace Conference, 9-16 March 2002, Big Sky, Montana, published on CD-ROM, IEEE Catalog Number 02TH8593C, Paper 6.0102* (pp. 1–12).

Kruse, F. A. & Boardman, J. W. (1997). Characterization and mapping of Kimberlites and related diatremes in Utah, Colorodo, and Wyoming, USA, using the airborne visible/ infrared imaging spectrometer (AVIRIS). In *ERIM Proceedings of the 12th International Conference on Applied Geologic Remote Sensing* (pp. 21–28).: Colorado.

Kruse, F. A., Boardman, J. W., & Huntington, J. F. (2003). Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), 1388–1400.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., & Goetz, F. H. (1993). The spectral image processing system (SIPS) interactive visualization and analysis of imaging spectrometer data. *Remote Sensing Environment*, 44, 145–163.

Lark, R. M. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma*, 105, 49–80.

Laylender, P. A. (1956). A performance estimate comparing conventional geologic mapping with that accomplished with the aid of color photographs. *Photogrammetric engineering*, (pp. 953).

Lehmann, F., Oertel, D., Richter, R., Rothfuss, H., Strobl, P., Muller, A., Tischler, S., Mueller, R., Beran, D., Fries, J., Boehl, R., & Obermeier, P. (1995). Hyperspectral applications with a new sensor. Presentation at the ISSSR (International Symposium on Spectral Sensing Research) in Melbourne: DAIS-7915, The Digital Airborne Imaging Spectrometer DAIS-7915.

Levinson, A. A. (1974). *Introduction to Exploration Geochemistry*. Applied Publishing Ltd., Calgary.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (1994). *Remote Sensing and Image Interpretation*. New York, John Wiley & Sons.

Linforda, N. & Platzman, E. (2004). Estimating the approximate firing temperature of burnt archaeological sediments through an unmixing algorithm applied to hysteresis data. *Physics of the Earth and Planetary Interiors*, 147, 197–207.

Lloyd, C. D. & Atkinson, P. M. (1999). Designing optimal sampling configurations with ordinary and indicator kriging. *GeoComputation*.

Loethen, Y. L., Zhang, D., Favors, R. N., Basiaga, S. B. G., & Ben-Amotz, D. (2004). Second-derivative variance minimization method for automated spectral subtraction. *Applied Spectroscopy*, 58(3), 272–278.

Longshaw, T. G. & Gilbertson, B. (1976). Multispectral aerial photography as exploration tool. iii - two applications in the North-Western Cape Province, South Africa (for mineral exploration). *Remote Sensing of Environment*, 5(2), 79–92.

Lowman, P. D. (1976). Geologic structure in California: three studies with Landsat-1 imagery. *California Geology*, 29, 75–81.

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2003). Classification of successional forest stages in the Brazilian Amazon basin. *Forest Ecology and Management*, 181(3), 301–312.

Martini, B. A. (2003). Assessing hydrothermal system dynamics and character by coupling hyperspectral imaging with historical drilling data: Long Valley Caldera, CA, USA. In *Proceedings 25th New Zealand Geothermal Workshop*, volume 25 (pp. 101–106).

Martini, B. A., Silver, E. A., Pickles, W. L., & Cocks, P. A. (2003). Hyperspectral mineral mapping in support of geothermal exploration: Examples from Long Valley Caldera, CA and Dixie Valley, NV, USA. In *Geothermal Resources Council Transactions*, volume 27 (pp. 657–662).

McBratney, A. B. & Webster, R. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables - II: Program and examples. *Computers & Geosciences*, 7(4), 335–365.

McBratney, A. B., Webster, R., & Burgess, T. M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables - I: Theory and method. *Computers & Geosciences*, 7(4), 331–334.

McGraw, J. F. & Tueller, P. T. (1983). Landsat computer-aided analysis techniques for range vegetation mapping. *Journal of Range Management*, 36, 627–631.

McGwire, K., Friedl, M., & Estes, J. E. (1993). Spatial structure, sampling design and scale in remotely-sensed imagery of a California Savanna Woodlands. *International Journal of Remote Sensing*, 14(11), 2137–2164.

Metternicht, G. I. & Fermont, A. (1998). Estimating erosion surface features by linear mixture modeling. *Remote Sensing of Environment*, 64, 254–265.

Moncur, M. C., Ptacek, C. J., Blowes, D. W., & Jambor, J. L. (2005). Release, transport and attenuation of metals from an old tailings impoundment. *Applied Geochemistry*, 20(3), 639–659.

Müller, W. G. & Zimmerman, D. L. (1999). Optimal designs for variogram estimation. *Environmetrics*, 10(23–37).

Okina, G. S., Roberts, D. A., Murraya, B., & Okin, W. J. (2001). Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment*, 77, 212–225.

Pan, G. (1993). Canonical favorability model for data integration and mineral potential mapping. *Computers & Geosciences*, 19(8), 1077–1100.

Papp, É. & Cudahy, T. (2002). Geophysical and remote sensing methods for regolith exploration. In É. Papp (Ed.), *Hyperspectral remote sensing* (pp. 13–21). CRCLEME Open File Report 144.

Penn, B. S. (2002). Using simulated annealing to obtain optimal linear end-member mixtures of hyperspectral data. *Computers & Geoscience*, 28, 809–817.

Polder, G. & van der Heijden, G. W. A. M. (2001). Multispectral and hyperspectral image acquisition and processing. In Q. Tong, Y. Zhu, & Z. Zhu (Eds.), *Proceedings of SPIE*, volume 4548.

Porwal, A., Carranza, E. J. M., & Hale, M. (2003). Artificial neural networks for mineral-potential mapping : a case study from Aravallia province, Western India. *Natural Resources Research*, 12(3), 155–171.

Qi, J., Chehbouni, A., Huete, A. R., Keer, Y. H., & Sorooshian, S. (1994). A modified soil vegetation adjusted index. *Remote Sensing of Environment*, 48, 119–126.

Qi, J., Kerr, Y. H., Moran, M. S., Weltz, M., Huete, A. R., Sorooshian, S., & Bryant, R. (2000). Leaf area index estimates using remotely sensed data and BRDF models in a semiarid region. *Remote Sensing of Environment*, 73, 18–30.

Resmini, R. G., Kappus, M. E., Aldrich, W. S., Harsanyi, J. C., & Anderson, M. (1997). Mineral mapping with hyperspectral digital imagery collection experiment (HYDICE) senson at Cuprite, Nevada, U.S.A. *International Journal of Remote Sensing*, 18(7), 1553–1570.

Richards, J. A. (1993). *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, Berlin, second edition.

Richardson, A. J., Menges, R. M., & Nixon, P. R. (1985). Distinguishing weed from crop plants using video remote-sensing. *Photogrammetric Engineering & Remote Sensing*, 51(11), 1785–1790.

Richter, R. (1996). Atmospheric correction of DAIS hyperspectral image data. In *SPIE Proceedings*, volume 2756 (pp. 390–399). Orlando.

Rigol-Sanchez, J. P., Chica-Olmo, M., & Abarca-Hernandez, F. (2003). Artificial neural networks as a tool for mineral potential mapping with GIS. *International Journal of Remote Sensing*, 24(5), 1151–1156.

Roberts, D. A., Yamaguchi, Y., & Lyon, R. J. P. (1985). Calibration of airborne imaging spectrometer data to percentage reflectance using field spectral measurements. In *Proceedings of the Nineteenth International Symposium on Remote Sensing of the Environment* (pp. 21–25).: Ann Arbor, Michigan.

Robinson, G. D., Gross, H. N., & Schott, J. R. (2000). Evaluation of two applications of spectral mixing models to image fusion. *Remote Sensing of Environment*, 71, 272–281.

Rose, A. W., Hawkes, H. E., & Webb, J. S. (1979). *Geochemistry in Mineral Exploration*. Academic Press, London, 2nd edition.

Rougean, J. L. & Breon, F. M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51, 375–384.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., & Harlan, J. C. (1974). Monitoring the vernal advancements and retrogradation of natural vegetation. In *NASA/GSFC, Final Report, Greenbelt, MD, USA* (pp. 1–137).

Rowan, L. C., Crowley, J. K., Schmidt, R. G., Ager, C. M., & Mars, J. C. (2000). Mapping hydrothermally altered rocks by analyzing hyperspectral image (AVIRIS) data of forested areas in the Southeastern United States. *Journal of Geochemical Exploration*, 68(3), 145–166.

Rowan, L. C., Goetz, A. F. H., & Ashley, R. P. (1977). Discrimination of hydrothermally altered and unaltered rocks in visible and near infrared multispectral images. *Geophysics*, 42(3), 522–535.

Rowlingson, B. & Diggle, P. (1993). Splancs: spatial point pattern analysis code in S-Plus. *Computers and Geosciences*, 19, 627–655.

Royle, J. A. & Nychka, D. (1998). An algorithm for the construction of spatial coverage designs with implementation in S-PLUS. *Computational Geoscience*, 24, 479–488.

Russo, D. (1984). Design of an optimal sampling network for estimating the variogram. *Soil Science Society American Journal*, 52, 708–716.

Sabins, F. F. (1996). *Remote Sensing: Principles and Interpretation.* W.H. Freeman and Company, New York, third edition.

Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore Geology Reviews*, 14(Issues 3–4), 157–183.

Sacks, J. & Schiller, S. (1988). Spatial designs. In S. Gupta & J. Berger (Eds.), *Statistical Decision Theory and Related Topics*, volume 2 of *Papers from the forth Purdue symposium* (pp. 385–399).: Springer-Verlag, New York.

Salisbury, J. W., Walter, L. S., Vergo, N., & D'Aria, D. M. (1991). *Infrared (2.125 μm) spectra of minerals.* Johns Hopkins University Press, Baltimore, MD.

Shahwan, T., Zünbül, B., Erŏlu, A. E., & Yilmaz, S. (2005). Effect of magnesium carbonate on the uptake of aqueous zinc and lead Ions by natural kaolinite and clinoptilolite. *Applied Clay Science*, 30, 209–218.

Shieh, S., Chu, J., & Jang, S. (2005). An interactive sampling strategy based on information analysis and ordinary kriging for locating hot spot regions. *Mathematical Geology*, 37(1), 29–48.

Shimabukuro, Y. E. & Smith, A. J. (1991). The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 29, 16–20.

Sidenko, N. V. & Sherriff, B. L. (2005). The attenuation of Ni, Zn and Cu by secondary Fe phases of different crystallinity from surface and ground water of two sulfide mine tailings in Manitoba, Canada. *Applied Geochemistry*, 20(6), 1180–1194.

Siegal, B. S. & Abrams, M. J. (1976). Geologic mapping using Landsat data. *Photogrammetric Engineering and Remote Sensing*, 42, 325–337.

Siegal, B. S. & Gillespie, A. R. (1980). *Remote sensing in geology.* New York : Wiley.

Singer, D. A. (1993). Basic concepts in three-part quantitative assessments of undiscovered mineral resources. *Nonrenewable Resources*, 2(2), 69–81.

Singer, D. A. & Kouda, R. (1997). Classification of mineral deposits into types using mineralogy with a probabilistic neural network. *Nonrenewable Resources*, 6(1), 27–32.

Singer, D. A. & Kouda, R. (2003). Typing mineral deposits using their grades and tonnages in an artificial neural network. *Natural Resources Research*, 12(3), 201–208.

Singer, R. B. & McCord, T. B. (1979). Mars: large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance. In *Proceedings of the 10th Lunar and Planetary Science Conference, Houston, USA, 19-23 March 1979* (pp. 1835–1848).

Smith, M. O., Johnston, P. E., & Adams, J. B. (1985). Quantitative determination of mineral types and abundances from reflectance spectra using principal component analysis. *Journal of Geophysical Research*, 90, 797–804.

Spanner, M. A., Pierce, L. L., Peterson, D. L., & Running, S. W. (1990). Remote sensing of temperate coniferous forest leaf area index. the influence of canopy closure, understory vegetation and background reflectance. *Internation Journal of Remote Sensing*, 11, 95–111.

Spruill, T. B. & Candela, L. (1990). Two approaches to design of monitoring networks. *Ground Water*, 28, 430–442.

Srivastav, S. K., Bhattacharya, A., Kamaraju, M. V. V., Reddy, G. S., Shrimal, A. K., Mehta, D. S., List, F. K., & Burger, H. (2000). Remote sensing and GIS for locating favourable zones of lead-zinc-copper mineralization in Rajpura-Dariba area, Rajasthan, India. *International Journal Remote Sensing*, 21(17), 3253–3267.

Staenz, K., Neville, R. A., Lévesque, J., Szeredi, T., Singhroy, V., Borstad, G. A., & Hauff, P. (1999). Evaluation of CASI and SFSI hyperspectral data for environmental and geological applications - two case studies. *Canadian Journal of Remote Sensing*, 25(3), 311–322.

Stanford, D. C. & Raftery, A. E. (2002). Approximate bayes factors for image segmentation: The pseudolikelihood information criterion (PLIC). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11), 1517–1520.

Stein, A., van der Meer, F., & Gorte, B., Eds. (1999). *Spatial Statistics for Remote Sensing*, volume 1 of *Remote Sensing and Digital Image Processing*. Kluwer Academic Publishers.

Swayze, G. A., Smith, K. S., Clark, R. N., Sutley, S. J., Pearson, R. M., Vance, J. S., Hageman, P. L., Briggs, P. H., Meier, A. L., Singleton, M. J., & Roth, S. (2000). Using imaging spectroscopy to map acidic mine waste. *Environmental Science and Technology*, 34, 47–54.

Tapia, R., Stein, A., & Bijker, W. (2005). Optimization of sampling schemes for vegetation mapping using fuzzy classification. *Remote Sensing of Environment*, (pp. 425–433).

Theiler, J. & Gisler, G. (1997). A contiguity-enhanced K-Means clustering algorithm for unsupervised multispectral image segmentation. In *Proceedings of SPIE 3159* (pp. 108–118).

Thenkabail, P. S. (2002). Optimal hyperspectral narrowbands for discriminating agricultural crops. *Remote Sensing Reviews*, 20(4), 257–291.

Thenkabail P. S., Smith, R. B. & De-Pauw, E. (2002). Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering and Remote Sensing*, 68(6), 607–621.

Thiart, C., Bonham-Carter, G. F., Agterberg, F. P., Cheng, Q., & Panahi, A. (2004). An application of the new omnibus test for conditional independence in weights-of-evidence modeling. In J. Harris & D. Wright (Eds.), *Special Volume on GIS Applications in the Earth Sciences, Geological Association of Canada*.

Thompson, M. & Howarth, R. (1978). A new approach to the estimation of analytical precision. *Journal of Geochemical Exploration*, 9(1), 23–30.

Thompson, S. K. (1992). *Sampling*. John Wiley and Sons, Inc., NY.

Van der Meer, F. D. (2004). Analysis of spectral absorption features in hyperspactral imagery. *JAG: International Journal of Applied Earth Observation and Geoinformation*, 5(1), 55–68.

Van der Meer, F. D. & De Jong, S. M. (2000). Improving the results of spectral unmixing of Landsat Thematic Mapper imagery by enhancing the orthogonality of end-members. *International Journal of Remote Sensing*, 21(15), 2781–2797.

Van Groenigen, J. W., Gandah, M., & Bouma, J. (2000a). Soil sampling strategies for precision agriculture research under Sahelian conditions. *Soil Science Society American Journal*, 64, 1674–1680.

Van Groenigen, J. W., Pieters, G., & Stein, A. (2000b). Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics*, 11, 227–244.

Van Groenigen, J. W., Siderius, W., & Stein, A. (1999). Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 87, 239–259.

Van Groenigen, J. W. & Stein, A. (1998). Constrained optimization of spatial sampling using continuous simulated annealing. *Journal Environmental Quality*, 27, 1078–1086.

Vaughan, R. G., Calvin, W. M., & Taranik, J. V. (2003). SEBASS hyperspectral thermal infrared data: surface emissivity measurement and mineral mapping. *Remote Sensing of Environment*, 85(1), 48–63.

Wackernagel, H. (1998). *Multivariate Geostatistics: An Introduction with Applications*. Springer-Verlag Germany, second edition.

Warrick, A. W. & Myers, D. E. (1987). Optimization of sampling locations for variogram calculations. *Water Resources Research*, 23(3), 496–500.

Yfantis, E. A., Flatman, G. T., & Behar, J. V. (1987). Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology*, 19(3), 183–205.

Zimmerman, D. L. & Homer, K. E. (1991). A network design criterion for estimating selected attributes of the semivariogram. *Environmetrics*, 4, 425–441.

Curriculum Vitae

Pravesh Debba (1969) was born in Durban, KwaZulu-Natal, South Africa. He received a B.Sc. degree (Mathematics and Statistics) and B.Sc. (Hons)(Statistics) from the University of Durban-Westville, Durban, KwaZulu-Natal, South Africa in 1991 and 1992 respectively. Having received a scholarship from the Flemish Ministry of Education, he completed an M.Sc. degree in Biostatistics at the Limburgs Universitair Centrum, Diepenbeek, Limburg, Belgium in 1998. He started his career as a junior lecturer in the department of statistics at the University of Durban-Westville in 1993 and continued at the University of South Africa, Pretoria, Gauteng, South Africa, from 1994 until 1999. He then joined the School of Statistics and Actuarial Sciences at the University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa in 2000, where he is presently employed as a lecturer. He was awarded a scholarship in 2001 to pursue his Ph.D. degree at the ITC International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

CONFERENCE AND WORKSHOP PRESENTATIONS

Debba, P, Carranza, E. J. M., Stein, A and van der Meer, F. D. (2005). Optimal Allocation of Exploration Targets on Mineral Prospectivity Maps. Presented at the 55th annual South African Statistical Association (SASA), 02–04 November 2005, Grahamstown, Eastern Cape, South Africa.

Debba, P., van Ruitenbeek, F. J. A., van der Meer, F. D., Carranza, E. J. M. and Stein, A. (2005). Weights derived from hyperspectral data to facilitate an optimal field sampling scheme for potential minerals. Abstract: In: 4th workshop on imaging spectroscopy : new quality in environmental studies, 27–29 April 2005 Warsaw, Poland : abstract book. / ed. by B. Zagajewski, M. Sobczak and W. Próchnicki. EARSeL, 2005. pp. 47.

Debba, P, Stein, A, van der Meer, F. D. and Lucieer, A. (2004). Segmentation techniques applied to remote sensing images in deriving an optimal sampling scheme. Presented at the VOC Vereniging voor ordinatie en Classificatie conference in De Bergse Bossen, Driebergen, 11–12 November 2004, The Netherlands.

Debba, P (1999). Review of statistical methods to access the agreement between different measuring devices and the reproducibility of repeated measurement applied to haemodynamic variables from a hypertension study. Presented at the 49th annual South African Statistical Association (SASA), 03–05 November 1999, Durban, KwaZulu-Natal, South Africa.

PEER REVIEWED ARTICLES

Debba, P., van Ruitenbeek, F. J. A., van der Meer, F. D., Carranza, E. J. M. and Stein, A. (2005). Optimal field sampling for targeting minerals using hyperspectral data. *Remote Sensing of Environment*, Vol 99, pp 373–386.

van Ruitenbeek, F. J. A., **Debba, P.,** van der Meer, F. D., Cudahy, T., van der Meijde, M. and Hale, M. (2006). Mapping white micas and their absorption wavelengths using hyperspectral band ratios. *Remote Sensing of Environment*. In press.

Debba, P, Carranza, E. J. M., Stein, A and van der Meer, F. D. Optimal Allocation of Exploration Targets on Mineral Prospectivity Maps. *Mathematical Geology*. In review.

Debba, P, Carranza, E. J. M., van der Meer, F. D. and Stein, A. Abundance Estimation of Spectrally similar materials in mine wastes using hyperspectral data and simulated annealing. *IEEE Transactions on Geoscience and Remote Sensing*. In review.

Debba, P, Carranza, E. J. M., Stein, A and van der Meer, F. D. Optimum sampling scheme for surface geochemical characterization of mine tailings. In preparation.

Debba, P, Stein, A, van der Meer, F. D., Carranza, E. J. M., and Lucieer, A. The optimal field sampling scheme on a remote sensing segmented image. In preparation.

ITC Dissertations

[1] **Akinyede, Joseph O.**, 1990, *Highway Cost Modelling and Route Selection Using a Geotechnical Information System*, Delft University of Technology.

[2] **Pan, Ping He**, 1990, *A Spatial Structure Theory in Machine Vision and Applications to Structural and Textural Analysis of Remotely Sensed Images*, University of Twente, 90-9003757-8.

[3] **Bocco Verdinelli, Gerardo H. R.**, 1990, *Gully Erosion Analysis Using Remote Sensing and Geographic Information Systems: A Case Study in Central Mexico*, Universiteit van Amsterdam.

[4] **Sharif, Massoud**, 1991, *Composite Sampling Optimization for DTM in the Context of GIS*, Wageningen Agricultural University.

[5] **Drummond, Jane E.**, 1991, *Determining and Processing Quality Parameters in Geographic Information Systems*, University of Newcastle.

[6] **Groten, Susanne**, 1991, *Satellite Monitoring of Agro-ecosystems in the Sahel*, Westfälische Wilhelms-Universität.

[7] **Sharifi, Ali**, 1991, *Development of an Appropriate Resource Information System to Support Agricultural Management at Farm Enterprise Level*, Wageningen Agricultural University, 90-6164-074-1.

[8] **van der Zee, Dick**, 1991, *Recreation Studied from Above: Air Photo Interpretation as Input into Land Evaluation for Recreation*, Wageningen Agricultural University, 90-6164-075-X.

[9] **Mannaerts, Chris**, 1991, *Assessment of the Transferability of Laboratory Rainfall-runoff and Rainfall—Soil Loss Relationships to Field and Catchment Scales: A Study in the Cape Verde Islands*, University of Ghent, 90-6164-085-7.

[10] **Wang, Ze Shen**, 1991, *An Expert System for Cartographic Symbol Design*, Utrecht University, 90-3930-333-9.

[11] **Zhou, Yunxuan**, 1991, *Application of Radon Transforms to the Processing of Airborne Geophysical Data*, Delft University of Technology, 90-6164-081-4.

[12] **de Zuviría, Martín**, 1992, *Mapping Agro-topoclimates by Integrating Topographic, Meteorological and Land Ecological Data in a Geographic Information System: A Case Study of the Lom Sak Area, North Central Thailand*, Universiteit van Amsterdam, 90-6164-077-6.

[13] **van Westen, Cees J.**, 1993, *Application of Geographic Information Systems to Landslide Hazard Zonation*, Delft University of Technology, 90-6164-078-4.

[14] **Shi, Wenzhong**, 1994, *Modelling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems*, Universität Osnabrück, 90-6164-099-7.

[15] **Javelosa, R.**, 1994, *Active Quaternary Environments in the Philippine Mobile Belt*, Utrecht University, 90-6164-086-5.

[16] **Lo, King-Chang**, 1994, *High Quality Automatic DEM, Digital Elevation Model Generation from Multiple Imagery*, University of Twente, 90-9006-526-1.

[17] **Wokabi, S. M.**, 1994, *Quantified Land Evaluation for Maize Yield Gap Analysis at Three Sites on the Eastern Slope of Mt. Kenya*, University Ghent, 90-6164-102-0.

[18] **Rodríguez Parisca, O. S.**, 1995, *Land Use Conflicts and Planning Strategies in Urban Fringes: A Case Study of Western Caracas, Venezuela*, University of Ghent.

[19] **van der Meer, Freek D.**, 1995, *Imaging Spectrometry & the Ronda Peridotites*, Wageningen Agricultural University, 90-5485-385-9.

[20] **Kufoniyi, Olajide**, 1995, *Spatial Coincidence: Automated Database Updating and Data Consistency in Vector GIS*, Wageningen Agricultural University, 90-6164-105-5.

[21] **Zambezi, P.**, 1995, *Geochemistry of the Nkombwa Hill Carbonatite Complex of Isoka District, North-east Zambia, with Special Emphasis on Economic Minerals*, Vrije Universiteit Amsterdam.

[22] **Woldai, Tsehaie**, 1995, *The Application of Remote Sensing to the Study of the Geology and Structure of the Carboniferous in the Calañas Area, Pyrite Belt, South-west Spain*, Open University, United Kingdom.

[23] **Verweij, Pita A.**, 1995, *Spatial and Temporal Modelling of Vegetation Patterns: Burning and Grazing in the Páramo of Los Nevados National Park, Colombia*, Universiteit van Amsterdam, 90-6164-109-8.

[24] **Pohl, Christine**, 1996, *Geometric Aspects of Multisensor Image Fusion for Topographic Map Updating in the Humid Tropics*, Universität Hannover, 90-6164-121-7.

[25] **Bin, Jiang**, 1996, *Fuzzy Overlay Analysis and Visualization in Geographic Information Systemes*, Utrecht University, 90-6266-128-9.

[26] **Metternicht, Graciela I.**, 1996, *Detecting and Monitoring Land Degradation Features and Processes in the Cochabamba Valleys, Bolivia. A Synergistic Approach*, University of Ghent, 90-6164-118-7.

[27] **Chu Thai Hoanh**, 1996, *Development of a Computerized Aid to Integrated Land Use Planning (CAILUP) at Regional Level in Irrigated Areas: A Case Study for the Quan Lo Phung Hiep region in the Mekong Delta, Vietnam*, Wageningen Agricultural University, 90-6164-120-9.

[28] **Roshannejad, A.**, 1996, *The Management of Spatio-Temporal Data in a National Geographic Information System*, University of Twente, 90-9009-284-6.

[29] **Terlien, Mark T. J.**, 1996, *Modelling Spatial and Temporal Variations in Rainfall-triggered Landslides: The Integration of Hydrologic Models, Slope Stability Models and GIS for the Hazard Zonation of Rainfall-triggered Landslides with Examples from Manizales, Colombia*, Utrecht University, 90-6164-115-2.

[30] **Mahavir, J.**, 1996, *Modelling Settlement Patterns for Metropolitan Regions: Inputs from Remote Sensing*, Utrecht University, 90-6164-117-9.

[31] **Al-Amir, Sahar**, 1996, *Modern Spatial Planning Practice as Supported by the Multi-applicable Tools of Remote Sensing and GIS: The Syrian Case*, Utrecht University, 90-6164-116-0.

[32] **Pilouk, M.**, 1996, *Integrated Modelling for 3D GIS*, University of Twente, 90-6164-122-5.

[33] **Duan, Zengshan**, 1996, *Optimization Modelling of a River-Aquifer System with Technical Interventions: A Case Study for the Huangshui River and the Coastal Aquifer, Shandong, China*, Vrije Universiteit Amsterdam, 90-6164-123-3.

[34] **de Man, W. H. E.**, 1996, *Surveys: Informatie als Norm: Een Verkenning van de Institutionalisering van Dorp-surveys in Thailand en op de Filippijnen*, University of Twente, 90-9009-775-9.

[35] **Vekerdy, Zoltan**, 1996, *GIS-based Hydrological Modelling of Alluvial Regions: Using the Example of the Kisafld, Hungary*, Lorand Eotvos University of Sciences, 90-6164-119-5.

[36] **Gomes Pereira, Luisa M.**, 1996, *A Robust and Adaptive Matching Procedure for Automatic Modelling of Terrain Relief*, Delft University of Technology, 90-407-1385-5.

[37] **Fandiño Lozano, M. T.**, 1996, *A Framework of Ecological Evaluation oriented at the Establishment and Management of Protected Areas: A Case Study of the Santuario de Iguaque, Colombia*, Universiteit van Amsterdam, 90-6164-129-2.

[38] **Toxopeus, Bert**, 1996, *ISM: An Interactive Spatial and Temporal Modelling System as a Tool in Ecosystem Management: With Two Case Studies: Cibodas Biosphere Reserve, West Java Indonesia: Amboseli Biosphere Reserve, Kajiado District, Central Southern Kenya*, Universiteit van Amsterdam, 90-6164-126-8.

[39] **Wang, Yiman**, 1997, *Satellite SAR Imagery for Topographic Mapping of Tidal Flat Areas in the Dutch Wadden Sea*, Universiteit van Amsterdam, 90-6164-131-4.

[40] **Saldana Lopez, Asun**, 1997, *Complexity of Soils and Soilscape Patterns on the Southern Slopes of the Ayllon Range, Central Spain: a GIS Assisted Modelling Approach*, Universiteit van Amsterdam, 90-6164-133-0.

[41] **Ceccarelli, T.**, 1997, *Towards a Planning Support System for Communal Areas in the Zambezi Valley, Zimbabwe; A Multi-criteria Evaluation Linking Farm Household Analysis, Land Evaluation and Geographic Information Systems*, Utrecht University, 90-6164-135-7.

[42] **Peng, Wanning**, 1997, *Automated Generalization in GIS*, Wageningen Agricultural University, 90-6164-134-9.

[43] **Mendoza Lawas, M. C.**, 1997, *The Resource Users' Knowledge, the Neglected Input in Land Resource Management: The Case of the Kankanaey Farmers in Benguet, Philippines*, Utrecht University, 90-6164-137-3.

[44] **Bijker, Wietske**, 1997, *Radar for Rain Forest: A Monitoring System for Land Cover Change in the Colombian Amazon*, Wageningen Agricultural University, 90-6164-139-X.

[45] **Farshad, Abbas**, 1997, *Analysis of Integrated Soil and Water Management Practices within Different Agricultural Systems under Semi-arid Conditions of Iran and Evaluation of their Sustainability*, University of Ghent, 90-6164-142-X.

[46] **Orlic, B.**, 1997, *Predicting Subsurface Conditions for Geotechnical Modelling*, Delft University of Technology, 90-6164-140-3.

[47] **Bishr, Yaser**, 1997, *Semantic Aspects of Interoperable GIS*, Wageningen Agricultural University, 90-6164-141-1.

[48] **Zhang, Xiangmin**, 1998, *Coal fires in Northwest China: Detection, Monitoring and Prediction Using Remote Sensing Data*, Delft University of Technology, 90-6164-144-6.

[49] **Gens, Rudiger**, 1998, *Quality Assessment of SAR Interferometric Data*, University of Hannover, 90-6164-155-1.

[50] **Turkstra, Jan**, 1998, *Urban Development and Geographical Information: Spatial and Temporal Patterns of Urban Development and Land Values Using Integrated Geo-data, Villaviciencio, Colombia*, Utrecht University, 90-6164-147-0.

[51] **Cassells, Craig James Steven**, 1998, *Thermal Modelling of Underground Coal Fires in Northern China*, University of Dundee.

[52] **Naseri, M. Y.**, 1998, *Monitoring Soil Salinization, Iran*, Ghent University, 90-6164-195-0.

[53] **Gorte, Ben G. H.**, 1998, *Probabilistic Segmentation of Remotely Sensed Images*, Wageningen Agricultural University, 90-6164-157-8.

[54] **Ayenew, Tenalem**, 1998, *The Hydrological System of the Lake District Basin, Central Main Ethiopian Rift*, Universiteit van Amsterdam, 90-6164-158-6.

[55] **Wang, Donggen**, 1998, *Conjoint Approaches to Developing Activity-Based Models*, Technical University of Eindhoven, 90-6864-551-7.

[56] **Bastidas de Calderon, María**, 1998, *Environmental Fragility and Vulnerability of Amazonian Landscapes and Ecosystems in the Middle Orinoco River Basin, Venezuela*, University of Ghent.

[57] **Moameni, A.**, 1999, *Soil Quality Changes under Long-term Wheat Cultivation in the Marvdasht Plain, South-central Iran*, University of Ghent.

[58] **van Groenigen, J.W.**, 1999, *Constrained Optimisation of Spatial Sampling: A Geostatistical Approach*, Wageningen Agricultural University, 90-6164-156-X.

[59] **Cheng, Tao**, 1999, *A Process-oriented Data Model for Fuzzy Spatial Objects*, Wageningen Agricultural University, 90-6164-164-0.

[60] **Wolski, Piotr**, 1999, *Application of Reservoir Modelling to Hydrotopes Identified by Remote Sensing*, Vrije Universiteit Amsterdam, 90-6164-165-9.

[61] **Acharya, B.**, 1999, *Forest Biodiversity Assessment: A Spatial Analysis of Tree Species Diversity in Nepal*, Leiden University, 90-6164-168-3.

[62] **Abkar, Ali Akbar**, 1999, *Likelihood-based Segmentation and Classification of Remotely Sensed Images*, University of Twente, 90-6164-169-1.

[63] **Yanuariadi, Tetra**, 1999, *Sustainable Land Allocation: GIS-based Decision Support for Industrial Forest Plantation Development in Indonesia*, Wageningen University, 90-5808-082-X.

[64] **Abu Bakr, Mohamed**, 1999, *An Integrated Agro-Economic and Agro-Ecological Framework for Land Use Planning and Policy Analysis*, Wageningen University, 90-6164-170-5.

[65] **Eleveld, Marieke A.**, 1999, *Exploring Coastal Morphodynamics of Ameland (The Netherlands) with Remote Sensing Monitoring Techniques and Dynamic Modelling in GIS*, Universiteit van Amsterdam, 90-6461-166-7.

[66] **Hong, Yang**, 1999, *Imaging Spectrometry for Hydrocarbon Microseepage*, Delft University of Technology, 90-6164-172-1.

[67] **Mainam, Félix**, 1999, *Modelling Soil Erodibility in the Semiarid Zone of Cameroon*, University of Ghent, 90-6164-179-9.

[68] **Bakr, Mahmoud I.**, 2000, *A Stochastic Inverse-Management Approach to Groundwater Quality*, Delft University of Technology, 90-6164-176-4.

[69] **Zlatanova, Siyka**, 2000, *3D GIS for Urban Development*, Graz University of Technology, 90-6164-178-0.

[70] **Ottichilo, Wilber K.**, 2000, *Wildlife Dynamics: An Analysis of Change in the Masai Mara Ecosystem*, Wageningen University, 90-5808-197-4.

[71] **Kaymakci, Nuri**, 2000, *Tectono-stratigraphical Evolution of the Cankori Basin (Central Anatolia, Turkey)*, Utrecht University, 90-6164-181-0.

[72] **Gonzalez, Rhodora**, 2000, *Platforms and Terraces: Bridging Participation and GIS in Joint-learning for Watershed Management with the Ifugaos of the Philippines*, Wageningen University, 90-5808-246-6.

[73] **Schetselaar, Ernst**, 2000, *Integrated Analyses of Granite-gneiss Terrain from Field and Multisource Remotely Sensed Data. A Case Study from the Canadian Shield*, University of Delft, 90-6164-180-2.

[74] **Mesgari, M. Saadi**, 2000, *Topological Cell-Tuple Structure for Three-Dimensional Spatial Data*, University of Twente, 90-3651-511-4.

[75] **de Bie, Cees A. J. M.**, 2000, *Comparative Performance Analysis of Agro-Ecosystems*, Wageningen University, 90-5808-253-9.

[76] **Khaemba, Wilson M.**, 2000, *Spatial Statistics for Natural Resource Management*, Wageningen University, 90-5808-280-6.

[77] **Shrestha, Dhruba**, 2000, *Aspects of Erosion and Sedimentation in the Nepalese Himalaya: Highland-lowland Relations*, Ghent University, 90-6164-189-6.

[78] **Asadi Haroni, Hooshang**, 2000, *The Zarshuran Gold Deposit Model Applied in a Mineral Exploration GIS in Iran*, Delft University of Technology, 90-6164-185-3.

[79] **Raza, Ale**, 2001, *Object-Oriented Temporal GIS for Urban Applications*, University of Twente, 90-3651-540-8.

[80] **Farah, Hussein O.**, 2001, *Estimation of Regional Evaporation under Different Weather Conditions from Satellite and Meteorological Data. A Case Study in the Naivasha Basin, Kenya*, Wageningen University, 90-5808-331-4.

[81] **Zheng, Ding**, 2001, *A Neuro-Fuzzy Approach to Linguistic Knowledge Acquisition and Assessment in Spatial Decision Making*, University of Vechta, 90-6164-190-X.

[82] **Sahu, B. K.**, 2001, *Aeromagnetics of Continental Areas Flanking the Indian Ocean; with Implications for Geological Correlation and Gondwana Reassembly*, University of Capetown, South Africa.

[83] **Alfestawi, Yahia Ahmed M.**, 2001, *The Structural, Paleogeographical and Hydrocarbon Systems Analysis of the Ghadamis and Murzuq Basins, West Libya, with Emphasis on Their Relation to the Intervening Al Qarqaf Arch*, Delft Technical University, 90-6164-198-5.

[84] **Liu, Xuehua**, 2001, *Mapping and Modelling the Habitat of Giant Pandas in Foping Nature Reserve, China*, Wageningen University, 90-5808-496-5.

[85] **Oindo, Boniface Oluoch**, 2001, *Spatial Patterns of Species Diversity in Kenya*, Wageningen University, 90-5808-495-7.

[86] **Carranza, Emmanuel John M.**, 2002, *Geologically-constrained Mineral Potential Mapping: Examples from the Philippines*, Technical University of Delft, 90-6164-203-5.

[87] **Rugege, Denis**, 2002, *Regional Analysis of Maize-based Land Use Systems for Early Warning Applications*, Wageningen University, 90-5808-584-8.

[88] **Liu, Yaolin**, 2002, *Categorical Database Generalization in GIS*, Wageningen University, 90-5808-648-8.

[89] **Ogao, Patrick**, 2002, *Scientific Visualization*, Utrecht University, 90-6164-206-X.

[90] **Abadi, Abdulbaset Musbah**, 2002, *Tectonics of the Sirt Basin: Inferences from Tectonic Subsidence Analysis, Stress Inversion and Gravity Modeling*, Vrije Universiteit Amsterdam, 90-6164-205-1.

[91] **Geneletti, Davide**, 2002, *Ecological Evaluation for Environmental Impact Assessment*, Vrije Universiteit Amsterdam, 90-6809-337-1.

[92] **Sedogo, Laurent D.**, 2002, *Integration of Local Participatory and Regional Planning for Resources Management Using Remote Sensing and GIS*, Wageningen University, 90-5808-751-4.

[93] **Montoya, Ana Lorena**, 2002, *Urban Disaster Management: A Case Study of Earthquake Risk Assesment in Cartago, Costa Rica*, Utrecht University, 90-6164-2086.

[94] **Mobin-ud Din, Ahmad**, 2002, *Estimation of Net Groundwater Use in Irrigated River Basins Using Geo-information Techniques: A Case Study in Rechna Doab, Pakistan*, Wageningen University, 90-5808-761-1.

[95] **Said, Mohammed Yahya**, 2003, *Multiscale Perspectives of Species Richness in East Africa*, Wageningen University, 90-5808-794-8.

[96] **Schmidt, Karen S.**, 2003, *Hyperspectral Remote Sensing of Vegetation Species Distribution in a Saltmarsh*, Wageningen University, 90-5808-830-8.

[97] **López Binnqüist, Citlalli**, 2003, *The Endurance of Mexican Amate Paper: Exploring Additional dimensions to the Sustainable Development Concept*, University of Twente, 90-3651-900-4.

[98] **Huang, Zhengdong**, 2003, *Data Integration for Urban Transport Planning*, Utrecht University, 90-6164-211-6.

[99] **Cheng, Jianquan**, 2003, *Modelling Spatial and Temporal Urban Growth*, Utrecht University, 90-6164-212-4.

[100] **Campos dos Santos, José Laurindo**, 2003, *A Biodiversity Information System in an Open Data-Metadatabase Architecture*, University of Twente, 9061642140.

[101] **Hengl, Tomislav**, 2003, *Pedometric Mapping: Bridging the Gaps Between Conventional and Pedometric Approaches*, Wageningen University.

[102] **Barrera Bassols, Narciso**, 2003, *Symbolism, Knowledge and management of Soil and Land Resources in Indigenous Communities: Ethnopedology at Global, Regional and Local Scales*, University of Ghent.

[103] **Zhan, Qingming**, 2003, *A Hierarchical Object-based Approach for Urban Land-use Classification from Remote Sensing Data*, Wageningen University, 90-5808-917-7.

[104] **Daag, Arturo Santos**, 2003, *Modelling the Erosion of the Pyroclastic Flow Deposits and the Occurrences of Lahars at Mt. Pinatubo, Philipines*, Utrecht University, 90-6164-218-3.

[105] **Bacic, Ivan Luiz Zilli**, 2003, *Demand Driven Land Evaluation: With Case Studies in Santa Catarina, Brazil*, Wageningen University, 90-5808-902-9.

[106] **Murwira, Amon**, 2003, *Scale matters! A New Approach to Quantify Spatial Heterogeneity for Predicting the Distribution of Wildlife*, Wageningen University.

[107] **Mazvimavi, Dominic**, 2003, *Estimation of Flow Characteristics of Ungauged Catchments: A Case Study in Zimbabwe*, Wageningen University, 90-5808-950-9.

[108] **Tang, Xinming**, 2004, *Spatial Object Modeling in Fuzzy Topological Spaces: With Applications to Land Cover Change*, University of Twente, 90-6164-2205.

[109] **Kariuki, Patrick C.**, 2004, *Spectroscopy to measure the swelling potential of expansive soils*, University of Delft, 90-6164-221-3.

[110] **G., Javier M. Morales**, 2004, *Model-driven Design of Geo-Information Services*, University of Twente, 90-6164-222-1.

[111] **Mutanga, Onisimo**, 2004, *Hyperspectral remote sensing of tropical grass quality and quantity*, 90-5808-981-9.

[112] **Sliuzas, Richard V.**, 2004, *Managing informal settlements: a study using geo-information in Dar es Salaam, Tanzania*, 90-6164-223-X.

[113] **Lucieer, Arko**, 2004, *Uncertainties in segmentation and their visualisation*, 90-6164-222-1.

[114] **Corsi, Fabio**, 2004, *Applications of existing biodiversity information: capacity to support decision-making*, 90-8504-090-6.

[115] **Tuladhar, Arbind M.**, 2004, *Parcel-based geo-information system: concepts and guidelines*, 90-6164-224-8.

[116] **Corné P. Elzakker, van**, 2004, *The use of maps in the exploration of geographic data*, 90-6809-357-6.

[117] **Nidumolu, Uday B.**, 2004, *Integrating geo-information models with participatory approaches: applications in land use analysis*.

[118] **Koua, Etien**, 2005, *Computational and visual support for exploratory geovisualization and knowledge construction*, 90-6164-229-9.

[119] **Blok, Connie A.**, 2005, *Dynamic visualization variables in animation to support monitoring of spatial phenomena*, 90-6809-367-3.

[120] **Meratnia, Nirvana**, 2005, *Dynamic visualization variables in animation to support monitoring of spatial phenomena*, 90-6809-367-3.

[121] **Yemefack, Martin**, 2005, *Modelling and monitoring soil and land use dynamics within shifting agricultural landscape mosaic systems in Southern Cameroon*, 90-6164-233-7.

[122] **Zarkesh, Masoud Kheirkhah**, 2005, *Decision support system for floodwater spreading site selection in Iran*, 90-8504-256-9.

[123] **Nangendo, Grace**, 2005, *Changing forest-woodland-savanna mosaics in Uganda: with implications for conservation*, 90-8504-200-3.

[124] **Mohamed, Yasir A.**, 2005, *The Nile hydroclimatology: impact of the Sudd wetland*, 0-415-38483-4.

[125] **Duker, Alfred**, 2005, *Spatial analysis of factors implicated in Mycobacterium ulcerans infection in Ghana*, 90-8504-243-7.

[126] **Ferwerda, Jelle G.**, 2005, *Charting the quality of forage: measuring and mapping the variation of chemical components in foliage with hyperspectral remote sensing*, 90-8504-209-7.

[127] **Martinez, Javier A.**, 2005, *Monitoring intra-urban inequalities with GIS-based indicators: with a case study in Rosario, Argentina*, 90-6164-235-3.

[128] **Saavedra, Carlos P.**, 2005, *Estimating spatial patterns of soil erosion and deposition in the Andean region using geo-information techniques : a case study in Cochabamba, Bolivia*, 90-8504-289-5.

[129] **Vaiphasa, Chaichoke**, 2006, *Remote sensing techniques for mangrove mapping*, 90-8504-353-0.

[130] **Porwal, Alok**, 2006, *Mineral potential mapping with mathematical geological models*, 90-6164-240-X.

[131] **van der Werff, Harald M. A.**, 2006, *Knowledge-based remote sensing of complex objects: recognition of spectral and spatial patterns resulting from naturnal hydrocarbon seepages*, 90-6164-238-8.

[132] **van de Vlag, Daniël**, 2006, *Modelling and visualizing dynamic landscape objects and their qualities*, 90-8504-384-0.

[133] **Joshi, Chundamani**, 2006, *Mapping cryptic invader Chromolaena odorata in tropical forest ecosystems of Nepal*.

[134] **Dilo, Arta**, 2006, *Vague sets: definition, operators and implementation*.