Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data

Allan Aasbjerg Nielsen

Institute of Mathematical Modelling
Ph.D. Thesis No. 6
Lyngby 1994

IMM

ISSN 0909–3192

Printed by IMM/Technical University of Denmark
2nd Edition

Some of the work reported in this thesis has previously been reported in

Conradsen, K., Ersbøll, B. K., Nielsen, A. A., Pedersen, J. L., Stern, M. & Windfeld, K. (1991). *Development and Testing of New Techniques for Mineral-Exploration Based on Remote Sensing, Image Processing Methods and Multivariate Analysis. Final Report*. The Commission of the European Communities, Contract No. MA1M-0015-DK(B). 196 pp.

Conradsen, K., Nielsen, A. A., Windfeld, K., Ersbøll, B. K., Larsen, R., Hartelius, K. & Olsson, C. K. (1993). *Application and Development of New Techniques Based on Remote Sensing, Data Integration and Multivariate Analysis for Mineral Exploration. Final Report. Technical Annex*. The Commission of the European Communities, Contract No. MA2M-CT90-0010. 96 pp.

Conradsen, K. & Nielsen, A. A. (1991). *Remote Sensing in Forecasting Agricultural Statistics in Kenya*. Danida, the Danish International Development Agency, Contract No. 104.Dan.8/410. 191 pp.

Conradsen, K. & Nielsen, A. A. (1994). Multivariate alteration detection (MAD) in multispectral, bi-temporal image data: a new approach to change detection studies. Submitted to *Remote Sensing of Environment*.

Conradsen, K., Nielsen, B. K., & Nielsen, A. A. (1991). Noise removal in multichannel image data by a parametric maximum noise fractions estimator. In Environmental Research Institute of Michigan (Ed.), *Proceedings of the 24th International Symposium on Remote Sensing of Environment*, pp. 403–416. Rio de Janeiro, Brazil.

Conradsen, K., Nielsen, A. A., & Windfeld, K. (1992). Analysis of geochemical data sampled on a regional scale. Invited contribution in Walden, A. & Guttorp, P. (Eds.), *Statistics in the Environmental and Earth Sciences*, pp. 283–300. Griffin.

Elsirafe, A. M., Nielsen, A. A., Conradsen, K. (1994). Application of image processing techniques and geostatistical methods to the aerial gamma-ray spectrometric and magnetometric survey data of a sample area from the Central Eastern Desert of Egypt as an aid to geological mapping and mineral exploration. In prep. Institute of Mathematical Modelling, Technical University of Denmark. 94 pp.

GAF, MAYASA, IMSOR, & DLR (1993). *Application and Development of New Techniques Based on Remote Sensing, Data Integration and Multivariate Analysis for Mineral Exploration. Final Report*. The Commission of the European Communities, Contract No. MA2M-CT90-0010. 117 pp.

Nielsen, A. A. & Larsen, R. (1994). Restoration of GERIS data using the maximum noise fractions transform. In Environmental Research Institute of Michigan (Ed.), *Proceedings from the First International Airborne Remote Sensing Conference and Exhibition, Volume II*, pp. 557–568. Strasbourg, France.

Nielsen, A. A. (1993). 2D semivariograms. In Cilliers, P. (Ed.), *Proceedings of the Fourth South African Workshop on Pattern Recognition*, pp. 25–35. Simon's Town, South Africa.

Nielsen, A. A. (1994a). Geochemistry in Eastern Erzgebirge: data report. Institute of Mathematical Modelling, Technical University of Denmark. 38 pp.

Nielsen, A. A. (1994b). Geophysics and integration with geochemistry in Eastern Erzgebirge: data report. Institute of Mathematical Modelling, Technical University of Denmark. 25 pp.

Pendock, N. & Nielsen, A. A. (1993). Multispectral image enhancement neural networks and the maximum noise fraction transform. In Cilliers, P. (Ed.), *Proceedings of the Fourth South African Workshop on Pattern Recognition*, pp. 2–13. Simon's Town, South Africa.

Schneider, T., Petersen, O. H., Nielsen, A. A., & Windfeld, K. (1990). A geostatistical approach to indoor surface sampling strategies. *Journal of Aerosol Science*, **21**(4), 555–567.

Preface

This thesis was prepared at the IMSOR Image group of the Institute of Mathematical Modelling (formerly the Institute of Mathematical Statistics and Operations Research), Technical University of Denmark in partial fulfillment of the requirements for acquiring the Ph.D. degree in engineering.

The thesis describes different methods that are useful in the analysis of multivariate data. Some of the methods focus on spatial data (sampled regularly or irregularly), others focus on multitemporal data or data from multiple sources. The latter methods can be used for change detection studies in multivariate, multitemporal data also. The thesis does not intend to cover all aspects of relevant data analysis techniques in this context. The methods presented have proven useful in several research programs, primarily in the fields of geologic mapping and mineral exploration. I see no reason why application of these methods should not be equally successful in any other field of application where studies are based on the analysis of collected data. Potential application areas besides geologic mapping and mineral exploration include monitoring and surveillance in environmental studies, oceanography, agriculture, forestry, geobotany etc.

Behind many data analysis concepts there is a simple idea. Sometimes this idea is dressed up or it disappears in long, intricate descriptions. This has *not* been my intended approach. I hope I have succeeded in giving straight forward and reasonably concise descriptions of traditional as well as new concepts. I also hope that the new techniques presented here will stand the test of time, the only real judge of true essence.

The thesis is fairly application oriented. This tendency might have been even stronger had the work been carried out in day-to-day contact with experts in the relevant fields of application be it geology, mineral exploration or other (earth) sciences. I am convinced that a narrow cooperation between experienced application experts and data analysts is *the* key to success in studies where the analysis of collected data is important.

Reading this thesis requires a basic knowledge of linear algebra and multivariate statistics.

Lyngby, October 1994

Allan Aasbjerg Nielsen

Driven by the wish to make this book available on the Internet an unusal 2nd edition of my thesis has been prepared. In the 2nd edition all photographic illustrations and a few sketches have been replaced by PostScript files, and other linear stretches of imagery in Chapter 3 have been applied. Also, a few minor misprintings have been corrected.

Lyngby, November 1995

AA

Acknowledgements

In carrying out the work reported here I have received important assistance from many people.

My supervisor Professor Knut Conradsen has been instrumental as a teacher, as my boss, and as a colleague and friend. I thank him for many tutorials, for good advise and good discussions not only on statistics and data analysis. Over a little more than a decade Knut with never ending enthusiasm has built up a professionally and socially well-working group with good facilities in hard- and software, and in natural intelligence.

Another instrumental person is my long-time colleague Dr. Bjarne Kjær Ersbøll. Bjarne and Knut first introduced me to and taught me statistics and image processing. Also, Dr. Kristian Windfeld (now Novo Nordisk A/S) has been important for me in the process of becoming a data analyst. Bjarne and Kristian taught me a lot in terms of data analysis concepts and working with them and getting hands-on experience has been of paramount importance for me.

In the early phase of this work (now Dr.) Rasmus Larsen as a student on my initiative and under my supervision wrote some of the software described here. Since then Rasmus has become a colleague with whom I enjoy working. Especially, I appreciate Rasmus's permanent support and involvement, and his good overview of most aspects of the activities of the IMSOR Image Group.

I also wish to thank Karsten Hartelius, Henrik Juul Hansen and Anders Rosholm for writing some of the software described here on my initiative and under my supervision.

The software development in the IMSOR Image Group takes place on a network of UNIX platforms. The programs are written mostly in C under the HIPS format. An important person in this context (and in many others that do not link directly to this work) is Dr. Jens Michael Carstensen. Without his important work in this area, which is not beneficial in a traditional academic sense, the toolbox of computer programs that we have built up would not be as powerful as it is.

The many other persons in the IMSOR Image Group not mentioned above have been important in that they constitute a good professional and social environment. Often little problems that could grow big for an individual are handled rapidly by a member of the group for whom it turns out to be no problem at all.

I wish to thank Dr. Arne Drud for the immediate interest he took in the optimization problems involved in some of the techniques and for writing the generic code to solve these problems. I also wish to thank Professor Kaj Madsen for introducing me to the world of non-linear optimization.

On the application side I wish to thank Chief Geologist Dr. Enrique Ortega, Minas de Almadén y Arrayanes, S. A., and Chief Geologist John L. Pedersen, Nunaoil A/S, for their interest and encouragement during the development of several of the data analysis concepts described and illustrated. I also wish to thank the entire team behind the projects funded by the Commission of the European Communities (now the European Union). These are our scientific officer with the commission Dr. Klaus Kögler and employees of

- Gesellschaft für Angewandte Fernerkundung (GAF), Munich, Germany,
- Minas de Almadén y Arrayanes, S.A. (MAYASA), Almadén, Spain, and
- Institut für Optoelektronik (OE), Deutsche Forschungsanstalt für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany.

Allan Aasbjerg Nielsen

The project coordinator is Dr. Peter Volk, GAF.

During a three month contract period some of the methods described in Chapters 2 and 3 were implemented at the DLR/OE Spektrometrie und Modelle (SM) group headed by Dr. Frank Lehmann.

I thank Professor James J. Simpson, Scripps Institution of Oceanography, University of California, San Diego, for the immediate interest he showed in the MAD and MUSECC concepts. Professor Simpson's group has—with my assistance—implemented the MAD transformation, and Professor Simpson has an excellent oceanographic application awaiting publication, hopefully jointly with that of Professor Conradsen and myself. Thanks also to Professor Simpson for taking on the task as my external examiner.

Also, I would like to thank Senior Geologist Agnete Steenfelt, the Geological Survey of Greenland (GGU), for the interest she took in especially the concept of MAF/MNFs of irregularly sampled data.

I thank Minas de Almadén y Arrayanes, S. A. (MAYASA) and the Geological Survey of Greenland (GGU) for permission to use some of their geochemical data. The access to the six years of geometrically corrected Landsat Thematic Mapper data from northern Sweden given by the Swedish Space Corporation is also acknowledged.

The funding from the Commission of the European Communities under contracts MA1M-0015-DK(B), MA2M-CT90-0010 and BRE2-CT92-0201 is highly appreciated.

Finally, I would like to thank my friends and family, my parents and my wife, Karin, and our children, Ida and Pi, for being there and for bearing with me in times of high work load.

Without all the fine people mentioned above and without the funding received I would not have been able to carry out this work.

Summary

This thesis describes different methods that are useful in the analysis of multivariate data. Some methods focus on spatial data (sampled regularly or irregularly), others focus on multitemporal data or data from multiple sources. The thesis covers selected and not all aspects of relevant data analysis techniques in this context.

Geostatistics is described in Chapter 1. Tools as the semivariogram, the cross-semivariogram and different types of kriging are described. As an independent re-invention 2-D sample semivariograms, cross-semivariograms and cova functions, and modelling of 2-D sample semi-variograms are described. As a new way of setting up a well-balanced kriging support the Delaunay triangulation is suggested. Two case studies show the usefulness of 2-D semivariograms of geochemical data from areas in central Spain (with a geologist's comment) and South Greenland, and kriging/cokriging of an undersampled variable in South Greenland, respectively.

Chapters 2 and 3 deal with various orthogonal transformations. Chapter 2 describes principal components (PC) analysis and two related spatial extensions, namely minimum/maximum autocorrelation factors (MAF) and maximum noise fractions (MNF) analysis. Whereas PCs maximize the variance represented by each component, MAFs maximize the spatial autocorrelation represented by each component, and MNFs maximize a measure of signal-to-noise ratio represented by each component. In the literature MAF/MNF analysis is described for regularly gridded data only. Here, the concepts are extended to irregularly sampled

data via the Delaunay triangulation. As a link to the methods described in Chapter 1 a new type of kriging based on MAF/MNFs for irregularly spaced data is suggested. Also, a new way of removing periodic, salt-and-pepper and other types of noise based on Fourier filtering of MAF/MNFs is suggested. One case study successfully shows the effect of the MNF Fourier restoration. Another case shows the superiority of the MAF/MNF analysis over ordinary non-spatial factor analysis of geochemical data in South Greenland (with a geologist's comment). Also, two examples of MAF kriging are given.

In Chapter 3 the two-set case is extended to multiset canonical correlations analysis (MUSECC). Two new applications to change detection studies are described: one is a new orthogonal transformation, multivariate alteration detection (MAD), based on two-set canonical correlations analysis; the other deals with transformations of minimum similarity canonical variates from a multiset analysis. The analysis of correlations between variables where observations are considered as repetitions is termed R-mode analysis. In Q-mode analysis of correlations between observations, variables are considered as repetitions. Three case studies show the strength of the methods; one uses SPOT High Resolution Visible (HRV) multispectral (XS) data covering economically important pineapple and coffee plantations near Thika, Kiambu District, Kenya, the other two use Landsat Thematic Mapper (TM) data covering forested areas north of Umeå in northern Sweden. Here Q-mode performs better than R-mode analysis. The last case shows that because of the smart extension to univariate differences obtained by MAD analysis, all MAD components—also the high order MADs that contain information on maximum similarity as opposed to minimum similarity (i.e. change) contained in the low order MADs—are important in interpreting multivariate changes. This conclusion is supported by a (not shown) case study with simulated changes. Also the use of MAFs of MADs is successful. The absolute values of MADs and MAFs of MADs localize areas where big changes occur. Use of MAFs of high order multiset Q-mode canonical variates seems successful. Due to lack of ground truth data it is very hard to determine empirically which of the five multiset methods described is best (if any). Because of their strong ability to isolate noise both the MAD and the MUSECC techniques can be used iteratively to remove this noise.

Resumé

Denne afhandling beskriver forskellige nyttige multivariate dataanalysemetoder. Nogle metoder fokuserer på spatielle data (regulært eller irregulært indsamlede), andre fokuserer på multitemporale data eller data fra flere kilder. Afhandlingen omhandler udvalgte og ikke alle aspekter af relevante dataanalyseteknikker i denne sammenhæng.

Kapitel 1 omhandler geostatistik. Værktøjer som semivariogrammet, kryds-semivariogrammet og forskellige former for kriging er beskrevet. Som en uafhængig genopfindelse er 2-D eksperimentelle semivariogrammer, kryds-semivariogrammer og cova funktioner samt modellering af 2-D eksperimentelle semivariogrammer beskrevet. Som en ny måde, hvorpå man kan udvælge velbalancerede kriging naboskaber foreslås anvendelse af Delaunay triangulering. To eksempler viser nytten af 2-D semivariogrammer af geokemiske variable fra det centrale Spanien (med geologkommentar) og Sydgrønland hhv. kriging/cokriging af en undersamplet variabel i Sydgrønland.

Kapitel 2 og 3 omhandler forskellige ortogonale transformationer. Kapitel 2 beskriver principal komponent (PC) analyse og to beslægtede spatielle udvidelser, nemlig minimum/maksimum autokorrelationsfaktor- (MAF) og maksimum støjfraktionsanalyse (MNF). Hvor PCer maksimerer variansen i hver komponent, maksimerer MAFer den spatielle autokorrelation i hver komponent, og MNFer maksimerer et mål for signal-støj forholdet i hver komponent. I litteraturen er MAF/MNF analyse kun beskrevet for data indsamlet på et regulært net. Her er begreberne via Delaunay trianguleringen udvidet til irregulært indsamlede

data også. I sammenhæng med de i kapitel 1 beskrevne metoder foreslås en ny type kriging baseret på MAF/MNFer for irregulært indsamlede data. Yderligere foreslås en ny metode baseret på Fourier filtrering af MAF/MNFer til fjernelse af periodisk støj, salt-og-peber støj og andre former for støj. Ét eksempel viser med sukces effekten af MNF Fourier restaureringen. Et andet eksempel viser, at MAF/MNF analyse er almindelig ikke-spatiel faktoranalyse af geokemiske data i Sydgrønland overlegen (med geologkommentar). Desuden gives to eksempler på MAF kriging.

I kapitel 3 udvides to-sæt tilfældet til multisæt kanonisk korrelationsanalyse (MUSECC). To ny anvendelser til forandringsdetektion beskrives: en er en ny ortogonal transformation, multivariat forandringsdetektion (MAD), baseret på to-sæt kanonisk korrelationsanalyse; en anden omhandler transformationer af minimum similaritets kanoniske variable fra multisætanalyse. Analyse af korrelationer mellem variable, hvor observationer betragtes som gentagelser, kaldes R-modus analyse. I Q-modus analyse af korrelationer mellem observationer betragtes variable som gentagelser. Tre eksempler viser metodernes styrke; ét anvender SPOT High Resolution Visible (HRV) multispektrale (XS) data, som dækker økonomisk vigtige ananas- og kaffeplantager nær Thika, Kiambu District, Kenya, to andre anvender Landsat Thematic Mapper (TM) data fra skovdækkede områder nord for Umeå i det nordlige Sverige. Q-modus analyse giver her bedre resultater end R-modus. Det sidste eksempel viser, at på grund af den smarte udvidelse af univariate differenser, der opnås med MAD analyse, er alle MAD komponenter – også højere ordens MADer, som i modsætning til minimum similaritetsinformationen (altså forandring) i lavordens MADer indeholder maksimum similaritetsinformation – vigtige for en tolkning af multivariate forandringer. Denne konklusion støttes af et (ikke vist) eksempel med simulerede forandringer. MAFer af MADer kan også bruges med sukces. Den numeriske værdi af MADer og MAFer af MADer lokaliserer områder, hvor store forandringer forekommer. Brug af MAFer af højordens multisæt kanoniske variable ser lovende ud. Grundet manglende *ground truth* information er det meget vanskeligt at bestemme empirisk, hvilken (om nogen) af de fem beskrevne multisætanalyse metoder er bedst. På grund af både MAD og MUSECC teknikkernes evne til at isolere støj, kan begge anvendes iterativt til fjernelse af denne støj.

Contents

Preface v

Acknowledgements vii

Summary xi

Resumé xiii

Contents xv

List of Tables xvii

List of Figures xix

1 Geostatistics 1

1.1 The Semivariogram . . . 3

1.1.1 1-D Semivariogram models . . . 5

1.2 The Crossvariogram . . . 6

1.3 The Sample 2-D Crossvariogram . . . 9

1.3.1 2-D Semivariogram Models . . . 9

1.4 Regularization . . . 15

1.5 Ordinary Kriging . . . 15

1.6 Universal Kriging . . . 21

1.7 Cokriging . . . 25

1.8 Choice of Kriging Support . . . 27

1.8.1 Delaunay Triangulation . . . 27

1.9 Case Studies . . . 29

1.9.1 2-D Semivariograms . . . 29

1.9.2 Kriging versus Cokriging . . . 34

2 Dimensionality Reduction **45**

2.1 Principal Components . . . 46

2.2 Min/Max Autocorrelation Factors . . . 48

2.2.1 Linear Transformations of MAFs . . . 51

2.3 Maximum Noise Fractions . . . 53

2.3.1 Estimation of the Noise Covariance Matrix . . . 57

2.3.2 Periodic Noise . . . 58

2.4 MAF/MNFs of Irregularly Spaced Data . . . 58

2.5 Case Studies . . . 60

2.5.1 Noise in Hyperspectral GERIS Data . . . 60

2.5.2 MAFs and Irregular Sampling . . . 64

3 Multiset Data Analysis **75**

3.1 Two-set Canonical Correlations 77

3.1.1 MAD Transformation 82

3.2 Multiset Canonical Correlations 87

3.2.1 Maximize Sum of Covariances 90

3.2.2 Maximize Sum of Squared Covariances 96

3.2.3 Maximize Largest Eigenvalue 96

3.2.4 Minimize Smallest Eigenvalue 98

3.2.5 Minimize Determinant 99

3.3 Case Studies 99

3.3.1 SPOT HRV Data in Agriculture (MAD) 99

3.3.2 Landsat TM Data in Forestry (MUSECC) 120

3.3.3 Landsat TM Data in Forestry (MAD revisited) 143

A Redundancy Analysis **155**

B Multiset Canonical Correlations **159**

B.1 Maximize Sum of Squared Covariances 159

C Computer Implementations **165**

C.1 Geostatistics 166

C.2 Dimensionality Reduction 167

C.3 Multiset Data Analysis 167

References **171**

Ph.D. theses from IMSOR **181**

Ph.D. theses from IMM **187**

List of Tables

3.1 Simple statistics for 1987 and 1989 SPOT HRV XS data 111

3.2 Correlations among original variables 111

3.3 Canonical correlations 112

3.4 Raw canonical coefficients 113

3.5 Correlations between original variables and canonical variables 114

3.6 Correlations between original variables and MADs 114

3.7 Levels of MADs in three pineapple areas and in the town . . . 115

3.8 Variance of 1987 XS explained by the individual canonical variates for 1987 and 1989 117

3.9 Variance of 1989 XS explained by the individual canonical variates for 1989 and 1987 117

3.10 Squared multiple correlations (R^2) between 1987 (1989) XS and the first M canonical variates of 1989 (1987) XS 117

3.11 Correlations between R-mode canonical variates 1 for all five methods 141

3.12 Correlations between Q-mode canonical variates 1 for all five methods 142

3.13 Optimization criteria for all five methods, R-mode 143

3.14 Optimization criteria for all five methods, Q-mode 144

3.15 Correlations between MADs and original variables 147

3.16 Correlations between MAFs of MADs and MADs 147

3.17 Correlations between MAFs of MADs and original variables . . 148

List of Figures

1.1 Sketch of 2-D semivariogram concept 10

1.2 Sketch of elliptic cone 2-D semivariogram model 12

1.3 Examples of 2-D elliptic cone/spherical semivariogram models 16

1.4 Voronoi tessellation (top), Delaunay triangulation (bottom) . . . 28

1.5 2-D semivariograms for 16 geochemical elements in central Spain, 21×21 1 km pixels 30

1.6 2-D semivariograms for 16 geochemical elements in central Spain, 81×81 250 m pixels 30

1.7 2-D semivariograms for 16 geochemical elements in central Spain, 21×21 1 km pixels as contour plots 32

1.8 2-D semivariograms for 16 geochemical elements in central Spain, 21×21 1 km pixels as perspective plots 33

1.9 2-D semivariograms for 41 geochemical elements in South Greenland, 21×21 5 km pixels 35

1.10 2-D semivariograms for 41 geochemical elements in South Greenland, 31×31 2 km pixels 36

1.11 2-D semivariograms for 41 geochemical elements in South Greenland, 21×21 5 km pixels as contour plots 37

1.12 2-D semivariograms for 41 geochemical elements in South Greenland, 21×21 5 km pixels as perspective plots 38

1.13 Cross-semivariograms for Nb, Ta and Eu 39

1.14 Estimation variances as functions of undersampling, left: cokriging, center: separate kriging, right: ratio of empirical variances of separate kriging and cokriging; square: kriging variance, cross: empirical variance 40

1.15 Separately kriged Nb, no undersampling 41

1.16 Separately kriged Nb, 90% undersampling 42

1.17 Cokriged Nb, 90% undersampling 43

2.1 Principal components of 62 GERIS bands 62

2.2 Minimum/maximum autocorrelation factors of 62 GERIS bands 63

2.3 MNF number 4 before (bottom) and after (top) MNF Fourier destriping 65

2.4 Fourier spectra of MNF number 4 before (bottom) and after (top) peak removal 66

2.5 Channel number 1 before (bottom) and after (top) MNF Fourier destriping 67

2.6 The first three restored MNFs as RGB (bottom) and IHS (top) . 68

2.7 Geological map of South Greenland 69

2.8 South Greenland: Varimax rotated factors 1, 2 and 3 as RGB . 70

2.9 South Greenland: MNFs 1, 2 and 3 as RGB 71

2.10 South Greenland: MAF kriged factors 1, 2 and 3 as RGB . . . 72

2.11 Southern Spain: MAF kriged factors 1, 2 and 3 as RGB 73

3.1 Sketch of areas of interest 100

3.2 False color composite of SPOT HRV XS, 5 Feb 1987 102

3.3 False color composite of SPOT HRV XS, 12 Feb 1989 103
3.4 False color composite of simple difference image 104
3.5 1989 NDVI as red and 1987 NDVI as cyan 105
3.6 Canonical variates of SPOT HRV XS, 5 Feb 1987 107
3.7 Canonical variates of SPOT HRV XS, 12 Feb 1989 108
3.8 MAD1, 2 and 3 in red, green and blue 109
3.9 Absolute value of MAD1, high values in red 110
3.10 Canonical variates geometrically 119
3.11 Sketch of R-mode multiset canonical correlations analysis . . . 121
3.12 Sketch of Q-mode multiset canonical correlations analysis . . . 122
3.13 Order of variables in following images, left: False color and R-mode, right: Q-mode 123
3.14 Landsat TM channels 4, 5 and 3 as red, green and blue 124
3.15 R-mode canonical variates 1, 2 and 3 as red, green and blue . . 126
3.16 R-mode canonical variates 6, 5 and 4 as red, green and blue . . 127
3.17 R-mode canonical variates 6 128
3.18 Absolute values of R-mode canonical variates 6 129
3.19 MAFs of R-mode canonical variates 6 130
3.20 Absolute values of MAFs of R-mode canonical variates 6 . . . 131
3.21 Q-mode canonical variates 6 132
3.22 Absolute values of Q-mode canonical variates 6 133
3.23 MAFs of Q-mode canonical variates 6 134
3.24 Absolute values of MAFs of Q-mode canonical variates 6 . . . 135
3.25 Correlations between R-mode CVs 1 and original data 136

3.26 Correlations between R-mode CVs 6 and original data 137
3.27 Correlations between Q-mode CVs 1 and original data 138
3.28 Correlations between Q-mode CVs 6 and original data 139
3.29 Sum of absolute value of MAFs 1 and 2 of Q-mode CVs 6 . . 140
3.30 MADs of TM bands from 1986 and 1988 145
3.31 Absolute values of MADs of TM bands from 1986 and 1988 . 146
3.32 MAFs of MADs of TM bands from 1986 and 1988 149
3.33 Absolute values of MAFs of MADs of TM bands from 1986 and 1988 150
3.34 Sum of absolute values of MAFs 1 and 2 of MADs of TM bands from 1986 and 1988 151
3.35 MAFs 1, 2 and 3 of MADs of TM bands from 1986 and 1988 as RGB 152
3.36 Absolute values of MAFs 1, 2 and 3 of MADs of TM bands from 1986 and 1988 as RGB 153

Chapter 1

Geostatistics

The basis of geostatistics is the idea of considering the observed values of a geochemical, a geophysical or another natural variable at a given set of positions as a realization of a stochastic process in space. For each position $\boldsymbol{x}$ in a domain $\mathcal{D}$ there exists a measurable quantity $z(\boldsymbol{x})$, a so-called **regionalized variable**. $\mathcal{D}$ is typically a subset of $\mathcal{R}^2$ or of $\mathcal{R}^3$. $z(\boldsymbol{x})$ is considered a particular outcome or realization of a **random variable** $Z(\boldsymbol{x})$. The set of random variables $\{Z(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{D}\}$ constitutes a **random function**. $Z(\boldsymbol{x})$ has mean value $\mathrm{E}\{Z(\boldsymbol{x})\} = \mu(\boldsymbol{x})$ and covariance $\mathrm{Cov}\{Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h})\} = C(\boldsymbol{x}, \boldsymbol{h})$. If $\mu(\boldsymbol{x})$ is constant over $\mathcal{D}$, i.e. $\mu(\boldsymbol{x}) = \mu$, Z is said to be first order stationary. If $C(\boldsymbol{x}, \boldsymbol{h})$ is constant over $\mathcal{D}$ also, i.e. $C(\boldsymbol{x}, \boldsymbol{h}) = C(\boldsymbol{h})$, Z is said to be second order stationary. Often $Z(\boldsymbol{x})$ is assumed to follow a normal or a lognormal distribution. If more variables are studied simultaneously, the cross-covariance functions $C_{ij}(\boldsymbol{x}, \boldsymbol{h}) = \mathrm{Cov}\{Z_i(\boldsymbol{x}), Z_j(\boldsymbol{x}+\boldsymbol{h})\}$ apply also. This statistical view on natural phenomena was inspired by work of Georges Matheron in 1962–1963 and is described in great detail in David (1977) and in Journel & Huijbregts (1978). An introductory textbook is Clark (1979). David (1988) looks back on ten years of application of geostatistics. Journel (1989) is a good concise survey of many important topics in geostatistics. Isaaks & Srivastava (1989) give an excellent practically and

data analytically oriented introduction to geostatistics. Cressie (1991) gathers a decade of development in statistics for spatial data. The official journal of the International Association for Mathematical Geology “Mathematical Geology” (Ehrlich, ed.) is *the* vehicle for publishing of research and applications in the field of geostatistics.

The application of a stochastic approach to spatial phenomena in e.g. geology and mining is sometimes questioned. The phenomenum under study is considered unique and a statistical approach where one considers that unique phenomenum as a realization or an outcome of an underlying random function seems awkward to some. However, thinking in terms of a data-analytical line of attack, the samples and the variables available represent one realization. The data material could be discarded and another set of samples could be collected and maybe analyzed (e.g. chemically) in a different fashion. This would then constitute another realization. Even if we just repeat the sampling process and have samples analyzed by the same laboratory using the same chemical techniques there would be a natural variation. Also, in terms of the concept of random functions, one can easily conceive of other areas that are statistically similar to the study region (e.g. sub-areas within the study region or geographically distant areas with similar geology). This can also be thought of as another realization of the same random function.

Point measurements of geochemical, geophysical or other natural variables or measurements taken over areas or volumes, also known as **supports**, are in principle continuous phenomena in space. If “dense” sampling is performed the continuous nature of the variable in question will be reflected in the covariation of neighboring samples. If taken further apart from each other there will be little or no covariation between samples. Whether samples are “dense” depends on the variable in question and sample sizes. Also, the autocorrelation revealed will depend on the scale at which one is operating. Different autocorrelation structures can be present simultaneously at different scales (mineralizations at the size of a few meters vs. regional variations at the size of several kilometers); this is referred to as **nested structures**.

The classical application of geostatistics is the calculation of ore reserves. Another application is the description of the spatial distribution and the interpolation of natural variables, e.g. geochemical elements, over large areas (in the order of several kilometers by several kilometers). In general, geostatistical methods are useful whenever spatial phenomena can be considered as being of probabilistic nature.

The term "geostatistics" refers to many different techniques where (spatial) statistical methods are applied in (earth) science(s). It also refers to topics not mentioned in the following sections (e.g. simulation techniques, relative semi-variograms, intrinsic random functions, and non-linear estimation techniques such as disjunctive kriging).

In the following sections I describe the semivariogram, the crossvariogram including 1- and 2-D sample versions, regularization and several forms of kriging. Also, the choice of kriging support is described. The use of the Delaunay triangulation in this context is believed to be new.

1.1 The Semivariogram

Consider two scalar values $z(\boldsymbol{x})$ and $z(\boldsymbol{x}+\boldsymbol{h})$ measured at two points in space $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$ separated by $\boldsymbol{h}$. z is considered a particular realization of a random variable Z. The variability is described by the **autocovariance** function (assuming or imposing first order stationarity)

$$C(\boldsymbol{x},\boldsymbol{h}) = \mathrm{E}\{[Z(\boldsymbol{x})-\mu][Z(\boldsymbol{x}+\boldsymbol{h})-\mu]\}. \tag{1.1}$$

The **variogram** is defined as

$$2\gamma(\boldsymbol{x},\boldsymbol{h}) = \mathrm{E}\{[Z(\boldsymbol{x})-Z(\boldsymbol{x}+\boldsymbol{h})]^2\}. \tag{1.2}$$

In general the variogram will depend on the location in space $\boldsymbol{x}$ and on the displacement vector $\boldsymbol{h}$. Note, that the variogram represents a more general concept than that of the covariance function since the increment process $Z(\boldsymbol{x})-Z(\boldsymbol{x}+\boldsymbol{h})$ may have desired properties which the basic process $Z(\boldsymbol{x})$ does not possess. The **intrinsic hypothesis** in geostatistics states that the variogram is independent of the location in space and that it depends on the displacement vector only, i.e.

$$2\gamma(\boldsymbol{x},\boldsymbol{h}) = 2\gamma(\boldsymbol{h}). \tag{1.3}$$

Second order stationarity of $Z(\boldsymbol{x})$ implies the intrinsic hypothesis (but not the other way around).

Assuming or imposing second order stationarity the autocovariance function and the **semivariogram**, γ, are related by

$$\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h}). \tag{1.4}$$

Note that $C(\boldsymbol{0}) = \sigma^2$, the variance of the random function.

An estimator for the semivariogram is the mean of the squared differences between any two measurements $z(\boldsymbol{x}_k)$ and $z(\boldsymbol{x}_k+\boldsymbol{h})$

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{k=1}^{N(\boldsymbol{h})}[z(\boldsymbol{x}_k)-z(\boldsymbol{x}_k+\boldsymbol{h})]^2, \tag{1.5}$$

where $N(\boldsymbol{h})$ is the number of point-pairs separated by $\boldsymbol{h}$. $\hat{\gamma}$ is called the **experimental or sample semivariogram**. Similarly we get for the **experimental or sample autocovariance**

$$\hat{C}(\boldsymbol{h}) = \frac{1}{N(\boldsymbol{h})} \sum_{k=1}^{N(h)} [z(\boldsymbol{x}) - \bar{z}][z(\boldsymbol{x}+\boldsymbol{h}) - \bar{z}], \tag{1.6}$$

where $\bar{z}$ is the estimated mean value of Z. Averaging over intervals of both magnitude and argument of $\boldsymbol{h}$ of $\hat{C}$ or $\hat{\gamma}$ is often performed. Averaging over intervals of the magnitude of $\boldsymbol{h}$ – i.e. creating distance or lag classes – is done to obtain a sufficiently high $N(\boldsymbol{h})$ to ensure a small estimation variance (the estimation variance is proportional to $1/N(\boldsymbol{h})$). Averaging over intervals of the argument of $\boldsymbol{h}$ – i.e. creating angular classes – is done to check for anisotropy.

1.1.1 1-D Semivariogram models

In order to be able to define characteristic quantities for the semivariogram (and in order to apply the semivariogram in kriging, see below) a model is often assumed. An often used semivariogram model is the spherical model with nugget effect. A reason for this is the easy interpretability of the parameters. Assuming isotropy and setting $|\boldsymbol{h}| = h$ the form of this model is

$$\gamma^*(h) = \begin{cases} 0 & \text{if } h = 0 \\ C_0 + C_1 \left[\frac{3}{2}\frac{h}{R} - \frac{1}{2}\left(\frac{h}{R}\right)^3\right] & \text{if } 0 < h < R \\ C_0 + C_1 & \text{if } h \geq R\,, \end{cases} \tag{1.7}$$

where C_0 is the **nugget effect** and R is the **range of influence**. $C_0/(C_0+C_1)$ is the relative nugget effect and $C_0 + C_1$ is the **sill** (= σ^2). The nugget effect is a discontinuity in the autocorrelation function at $h = 0$ due to both measurement errors and to micro-variability the structure of which is not available at the scale of study. This variability thus turns up as noise. The range of influence is the distance at which covariation between measurements stops; measurements taken further apart are uncorrelated. The spherical semivariogram model with nugget

effect can easily be extended to e.g. a double spherical model with nugget effect to allow for nested structures

$$\gamma^*(h) = \begin{cases} 0 & \text{if } h = 0 \\ C_0 + C_1 \left[\frac{3}{2}\frac{h}{R_1} - \frac{1}{2}\left(\frac{h}{R_1}\right)^3\right] & \\ \quad + C_2 \left[\frac{3}{2}\frac{h}{R_2} - \frac{1}{2}\left(\frac{h}{R_2}\right)^3\right] & \text{if } 0 < h < R_1 \\ C_0 + C_1 + C_2 \left[\frac{3}{2}\frac{h}{R_2} - \frac{1}{2}\left(\frac{h}{R_2}\right)^3\right] & \text{if } R_1 \leq h < R_2 \\ C_0 + C_1 + C_2 & \text{if } h \geq R_2\,, \end{cases} \tag{1.8}$$

where C_0 is the nugget effect and R_2 is the range of influence. $C_0/(C_0+C_1+C_2)$ is the relative nugget effect and $C_0 + C_1 + C_2$ is the sill. Other models for the semivariogram such as linear, bi-linear and exponential models are often used also.

The parameters in the above semivariogram models γ^* can be estimated from the experimental semivariograms $\hat{\gamma}$ by means of iterative, non-linear least squares methods. Different weights of the estimated values in the experimental semivariogram $\hat{\gamma}$ may be considered. A weighting with the number of point pairs included in the estimation for each lag distance seems natural. Also, if one is interested in a good model for small lags a weighting with inverse lag distance (or similar) seems appropriate.

It might be possible to estimate the above models directly from the data also.

There is an extensive literature on the problems one encounters when estimating experimental semivariograms on real world data, cf. e.g. Journel & Froidevaux (1982), Cressie (1985, 1991).

1.2 The Crossvariogram

What is said above about autocovariance functions and variograms is readily extended to cross-covariance functions and cross-variograms if more variables are studied simultaneously.

Consider two scalar values $z_i(\boldsymbol{x})$ and $z_j(\boldsymbol{x}+\boldsymbol{h})$ measured at two points in space $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$ separated by $\boldsymbol{h}$. z_i and z_j are considered particular realizations of random variables. The covariability is described by the **cross-covariance** function (again assuming or imposing first order stationarity)

$$C_{ij}(\boldsymbol{x},\boldsymbol{h}) = \mathrm{E}\{[Z_i(\boldsymbol{x}) - \mu_i][Z_j(\boldsymbol{x}+\boldsymbol{h}) - \mu_j]\}. \tag{1.9}$$

The **cross-variogram** is defined as

$$2\gamma_{ij}(\boldsymbol{x},\boldsymbol{h}) = \mathrm{E}\{[Z_i(\boldsymbol{x}) - Z_i(\boldsymbol{x}+\boldsymbol{h})][Z_j(\boldsymbol{x}) - Z_j(\boldsymbol{x}+\boldsymbol{h})]\}. \tag{1.10}$$

It is readily seen that $\gamma_{ij}(\boldsymbol{x},\boldsymbol{h}) = \gamma_{ji}(\boldsymbol{x},\boldsymbol{h})$. Furthermore we have

$$2\gamma_{ij}(\boldsymbol{x},\boldsymbol{h}) = C_{ij}(\boldsymbol{x},\boldsymbol{0}) + C_{ij}(\boldsymbol{x}+\boldsymbol{h},\boldsymbol{0}) - [C_{ij}(\boldsymbol{x},\boldsymbol{h}) + C_{ji}(\boldsymbol{x},\boldsymbol{h})]. \tag{1.11}$$

Similarly to the case of the variogram, we assume (or impose) the intrinsic hypothesis

$$2\gamma_{ij}(\boldsymbol{x},\boldsymbol{h}) = 2\gamma_{ij}(\boldsymbol{h}). \tag{1.12}$$

Assuming or imposing second order stationarity we get

$$C_{ij}(\boldsymbol{h}) = C_{ji}(-\boldsymbol{h}), \tag{1.13}$$

$$2\gamma_{ij}(\boldsymbol{h}) = 2\gamma_{ij}(-\boldsymbol{h}), \tag{1.14}$$

$$2\gamma_{ij}(\boldsymbol{h}) = 2C_{ij}(\boldsymbol{0}) - [C_{ij}(\boldsymbol{h}) + C_{ji}(\boldsymbol{h})]. \tag{1.15}$$

In general $C_{ij}(\boldsymbol{h}) \neq C_{ij}(-\boldsymbol{h})$, i.e. the cross-covariance is not symmetric in $\boldsymbol{h}$, whereas the cross-variogram is symmetric in $\boldsymbol{h}$.

If $\gamma^+_{ij}(\boldsymbol{h})$ denotes the semivariogram for $Z_i(\boldsymbol{x}) + Z_j(\boldsymbol{x})$, $\gamma_i(\boldsymbol{h})$ and $\gamma_j(\boldsymbol{h})$ denote the semivariograms for $Z_i(\boldsymbol{x})$ and $Z_j(\boldsymbol{x})$ respectively, then

$$\gamma^+_{ij}(\boldsymbol{h}) = \gamma_i(\boldsymbol{h}) + \gamma_j(\boldsymbol{h}) + 2\gamma_{ij}(\boldsymbol{h}). \tag{1.16}$$

Hence we see that the crossvariogram $\gamma_{ij}(\boldsymbol{h})$ can be modelled by sums of the same models that were mentioned for the semivariogram $\gamma(\boldsymbol{h})$.

An estimator for the crossvariogram is

$$\hat{\gamma}_{ij}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})} \sum_{k=1}^{N(h)} [z_i(\boldsymbol{x}_k) - z_i(\boldsymbol{x}_k+\boldsymbol{h})][z_j(\boldsymbol{x}_k) - z_j(\boldsymbol{x}_k+\boldsymbol{h})], \tag{1.17}$$

where $N(\boldsymbol{h})$ is the number of point-pairs separated by $\boldsymbol{h}$. $\hat{\gamma}_{ij}$ is called the **experimental or sample crossvariogram**. Similarly we get for the **experimental or sample crosscovariance**

$$\hat{C}_{ij}(\boldsymbol{h}) = \frac{1}{N(\boldsymbol{h})} \sum_{k=1}^{N(h)} [z_i(\boldsymbol{x}_k) - \bar{z}_i][z_j(\boldsymbol{x}_k+\boldsymbol{h}) - \bar{z}_j], \tag{1.18}$$

where $\bar{z}_i$ is the estimated mean value of Z_i.

In practical applications, averaging over intervals of both distance and direction of $\boldsymbol{h}$ is often applied in 1-D representations of $\hat{C}_{ij}$ or $\hat{\gamma}_{ij}$. This 1-D representation is often shown for each direction class at a time plotting $\hat{C}_{ij}$ or $\hat{\gamma}_{ij}$ as a function of distance classes. To reflect this averaging, $\boldsymbol{\Sigma}_k = \{\hat{C}_{ij}\}, k = 1, \ldots, n$, where n is the number of 1-D distance and direction classes, can be calculated.

When averaging is performed over sufficiently large areas, the variables z_i and z_j need not be sampled in exactly the same locations to estimate $\hat{C}_{ij}$. In case of such non-corresponding spacing, $\hat{C}_{ij}$ is referred to as a **cova function**, cf. Herzfeld (1990).

Another estimator for the crossvariogram is

$$\hat{\gamma}_{ij}(\boldsymbol{h}) = \frac{1}{2}[\hat{\gamma}^{+}_{ij}(\boldsymbol{h}) - \hat{\gamma}_i(\boldsymbol{h}) - \hat{\gamma}_j(\boldsymbol{h})]. \tag{1.19}$$

After separate modelling of $\hat{\gamma}_i$, $\hat{\gamma}_j$ and $\hat{\gamma}^{+}_{ij}$ it is necessary to verify that this equation holds for the models also.

1.3 The Sample 2-D Crossvariogram

If averaging over the Cartesian coordinates of $\boldsymbol{h}$, h_x and h_y, rather than averaging over the polar coordinates of $\boldsymbol{h}$ is considered, 2-D representations of $\hat{C}_{ij}$ or $\hat{\gamma}_{ij}$ can be estimated as ordinary image data with pixel size $h_x \times h_y$. In Figure 1.1 this 2-D crossvariogram concept is sketched. For each point in the data set we consider all points with relative position inside each square, e.g. the solid one. For all such point pairs we calculate $\hat{C}_{ij}$ or $\hat{\gamma}_{ij}$ and place the estimate as a pixel value in the pixel situated as indicated by the solid square. If $i = j$ the resulting image is the experimental 2-D semivariogram. This 2-D notion of the crossvariogram, which is also described in GAF, MAYASA, IMSOR, & DLR (1993), Conradsen, Nielsen, Windfeld, Ersbøll, Larsen, Hartelius,

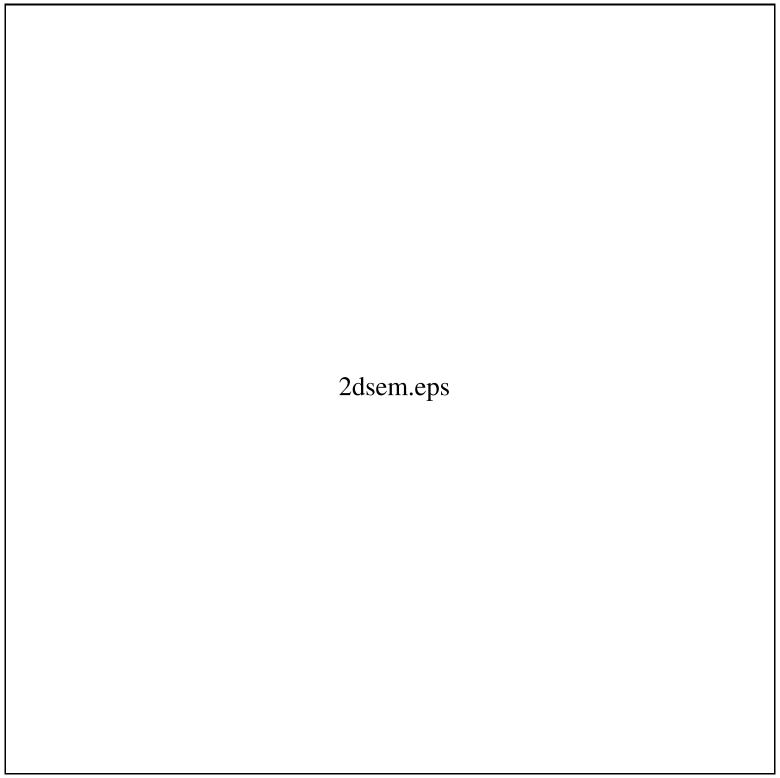

Figure 1.1: Sketch of 2-D semivariogram concept

& Olsson (1993), Nielsen (1993), is very powerful in revealing the degree and directions of anisotropy of the variables under study and also in its depiction of range of influence and nugget effect. The circles in Figure 1.1 are the ordinary 1-D lag limits for averaging in the magnitude of $\boldsymbol{h}$.

1.3.1 2-D Semivariogram Models

In this section I introduce several anisotropy models, an elliptic cone model, single and multiple elliptic spherical models, all with nugget effect. The multiple spherical models allow for range and sill anisotropy and for nested or un-nested spheres. All these models are intended for use with small lags and are not meant to describe the long range spatial behavior of the phenomena under study. Similar models are hinted in Isaaks & Srivastava (1989). The models presented allow for neither nugget effect anisotropy, periodicity nor non-linear behavior for $|\boldsymbol{h}| \to 0$. The parameters in these models can be estimated by means of iterative, non-linear least squares methods from the experimental semivariograms. It might be possible to estimate them directly from the data also.

An Elliptic Cone Model

The linear model with nugget effect is one of the simplest 1-D semivariogram models traditionally in use. A natural extension of this model into 2-D is a cone. If we want the ability to detect range of influence anisotropy (also known as geometric as opposed to zonal anisotropy) we must apply an elliptic cone. A sketch of this model is shown in Figure 1.2.

The equation for the elliptic cone is

$$(\frac{x_1}{a_1})^2 + (\frac{y_1}{b_1})^2 - (\frac{\gamma^* - c_0}{c})^2 = 0 \tag{1.20}$$

where x_1 and y_1 are the Cartesian coordinates of the displacement vector $\boldsymbol{h}$ in a coordinate system with x- and y-axes parallel to the major and minor axes

/home1/aa/phd/2Dsemivar/2dsemmod.eps

Figure 1.2: Sketch of elliptic cone 2-D semivariogram model

of the elliptic cone. To allow for anisotropy in any direction we introduce new coordinates corresponding to a rotation through α_1 relative to the coordinate system in which the data are recorded,

$$\begin{aligned} x &= x_1 \cos\alpha_1 - y_1 \sin\alpha_1 \\ y &= x_1 \sin\alpha_1 + y_1 \cos\alpha_1 \end{aligned} \tag{1.21}$$

$$\begin{aligned} x_1 &= x \cos\alpha_1 + y \sin\alpha_1 \\ y_1 &= -x \sin\alpha_1 + y \cos\alpha_1 . \end{aligned} \tag{1.22}$$

For the semivariogram model we get

$$\gamma^* = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + \sqrt{(\frac{x_1}{a})^2 + (\frac{y_1}{b})^2} & \text{if } h > 0 \end{cases} \tag{1.23}$$

with $a = a_1/c$ and $b = b_1/c$. $h = \sqrt{x^2 + y^2}$ is of course the magnitude of the displacement vector. c_0 is the nugget effect. The range anisotropy ratio is a/b and the range anisotropy direction is α_1. If inspection of the experimental 2-D semivariogram reveals different range anisotropies at different displacement distances, this simple model can also be used for instance by omitting observations with high respectively low h from the estimation. Alternatively a more complicated model can be used.

If isotropy can be assumed we simply omit α_1 from the model and set $a = b$. Thus we get the well known 1-D linear model with nugget effect

$$\gamma^*_{iso} = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + \frac{h}{a} & \text{if } h > 0. \end{cases} \tag{1.24}$$

Elliptic Spherical Models

The above elliptic cone model can be extended into a single elliptic spherical model that allows for ranges of influence and a sill

$$\gamma^* = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_1 \left[\frac{3}{2}\sqrt{(\frac{x_1}{a_1})^2 + (\frac{y_1}{b_1})^2} - \frac{1}{2}\left(\sqrt{(\frac{x_1}{a_1})^2 + (\frac{y_1}{b_1})^2}\right)^3 \right] & \\ & \text{if } 0 < (\frac{x_1}{a_1})^2 + (\frac{y_1}{b_1})^2 < 1 \\ c_0 + c_1 & \text{if } (\frac{x_1}{a_1})^2 + (\frac{y_1}{b_1})^2 \geq 1 . \end{cases} \tag{1.25}$$

If isotropy can be assumed this model reduces to the well known 1-D single spherical model with nugget effect.

The above single elliptic spherical model is readily extended into multiple elliptic spherical models that take different anisotropy directions for the elliptic spheres into account. This is done by introducing new coordinates x_i and y_i rotated through α_i relative to x and y and establishing relations similar to the ones noted for x_1 and y_1. These spheres may be nested (as in the 1-D double spherical model) so that only one sphere is effective in a certain range or they may all be effective simultaneously (or even combinations hereof). It is probably most sensible to have nested spheres in the same direction only.

One way of allowing for sill anisotropy, if so wanted, is by letting ($i > 0$)

$$\begin{aligned} c_i &= c_{i0} + c_{i1}\cos^2(\alpha - \alpha_i) \\ &= c_{i0} + c_{i1}(\frac{x_i}{h})^2 \end{aligned} \tag{1.26}$$

where α is the current angle for each pixel in the experimental 2-D semivariogram and the sill anisotropy direction is α_i. This forces the directions for range and sill anisotropy to be equal. The sill anisotropy ratio is $(c_{i0} + c_{i1})/c_{i0}$.

$c_{i0} + c_{i1}$ and c_{i0} are the sills reached in the direction of the major and minor axes respectively. c_{i1} can of course be negative. If we have nested spheres in some direction, only the sphere with the longest range should be modeled in this fashion.

Modeling sill anisotropy in the above fashion causes the semivariogram models to have "peanut-shaped" contours. Another way of performing this modeling is to apply additive models where range and sill anisotropies are modeled in separate additive terms. This type of modeling allows the sill anisotropy direction to be independent of the range anisotropy direction.

Figure 1.3 shows examples of the models suggested in the equations in this section, elliptic cone (top-left with paper in landscape mode), single elliptic spherical (top-right), single elliptic spherical with sill anisotropy (bottom-left), and un-nested double elliptic spherical with sill anisotropy for both spheres (bottom-right). For practical estimation purposes some of these models may contain too many parameters. The parameter estimation software for these 2-D models presently needs some further attention.

1.4 Regularization

Another important concept in geostatistics is **regularization**, i.e. the averaging of a random function over a domain $\mathcal{D}$. Let $\boldsymbol{x} \in \mathcal{D}$. The regularized value $\bar{Z}_{\mathcal{D}}$ of Z is

$$\bar{Z}_{\mathcal{D}} = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\boldsymbol{x})\, d\boldsymbol{x}, \tag{1.27}$$

where $|\mathcal{D}|$ is the area (or volume) of $\mathcal{D}$. Similarly for the moment functions, e.g.

$$\bar{\gamma}(\mathcal{A}, \mathcal{B}) = \frac{1}{|\mathcal{A}||\mathcal{B}|} \int_{\mathcal{A}} \int_{\mathcal{B}} \gamma(\boldsymbol{x} - \boldsymbol{y})\, d\boldsymbol{y}\, d\boldsymbol{x}. \tag{1.28}$$

/home1/aa/phd/2Dsemivar/3dsem4.eps

Figure 1.3: Examples of 2-D elliptic cone/spherical semivariogram models

Thus $\bar{\gamma}(\mathcal{A}, \mathcal{B})$ is the regularized semivariogram when one end of the displacement vector $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{y}$ varies in $\mathcal{A}$ and the other end of $\boldsymbol{h}$ varies in $\mathcal{B}$. This integral can be solved either analytically for certain semivariogram models and supports of simple geometry or it can be solved numerically.

1.5 Ordinary Kriging

Suppose that the random variable $Z(\boldsymbol{x})$ is sampled on a number of supports (could be points) $\mathcal{D}_1, \ldots, \mathcal{D}_N$ giving scalar measurements $z(\boldsymbol{x}_1), \ldots, z(\boldsymbol{x}_N)$ considered as particular realizations of $Z_{\mathcal{D}_1}, \ldots, Z_{\mathcal{D}_N}$. We now want to estimate $Z_{\mathcal{D}_0}$ on a support $\mathcal{D}_0$ where Z is not sampled (or Z is sampled on a part of $\mathcal{D}_0$ only). We are looking for a linear, unbiased estimator

$$\hat{Z}_{\mathcal{D}_0} = \sum_{i=1}^{N} w_i Z_{\mathcal{D}_i} \tag{1.29}$$

$$\mathrm{E}\{Z_{\mathcal{D}_0} - \hat{Z}_{\mathcal{D}_0}\} = 0. \tag{1.30}$$

If the mean is constant $\mu(\boldsymbol{x}) = \mu$ over the study area, the unbiasedness gives

$$\sum_{i=1}^{N} w_i = 1. \tag{1.31}$$

Without this non-bias constraint the estimation method is referred to as **simple kriging**. The estimation variance (or the mean squared error) is ($w_0 = -1$)

$$\begin{aligned} \sigma_E^2 &= \mathrm{E}\{(Z_{\mathcal{D}_0} - \hat{Z}_{\mathcal{D}_0})^2\} \\ &= \mathrm{Var}\{Z_{\mathcal{D}_0} - \hat{Z}_{\mathcal{D}_0}\} \\ &= \sum_{i=0}^{N} \sum_{j=0}^{N} w_i w_j C(\mathcal{D}_i, \mathcal{D}_j). \end{aligned} \tag{1.32}$$

The **ordinary kriging (OK) estimate** is defined by the values of the weights w_i that minimize the estimation variance σ_E^2 subject to the above constraint that the sum of the (ordinary kriging) weights is unity. These weights can be found by introducing a Lagrangian multiplier and setting each of the N partial derivatives $\partial[\sigma_E^2 - 2\lambda(\sum_{i=1}^N w_i - 1)]/\partial w_i = 0$ leading to the $(N+1) \times (N+1)$ set of equations

$$\sum_{i=1}^{N} w_i C(\mathcal{D}_j, \mathcal{D}_i) - \lambda = C(\mathcal{D}_j, \mathcal{D}_0)\,,\ j = 1, \ldots, N \tag{1.33}$$

$$\sum_{i=1}^{N} w_i = 1. \tag{1.34}$$

with the **ordinary kriging variance** (or the minimum mean squared error)

$$\sigma_{OK}^2 = -\sum_{i=1}^{N} w_i C(\mathcal{D}_i, \mathcal{D}_0) + \lambda + C(\mathcal{D}_0, \mathcal{D}_0). \tag{1.35}$$

Of course this can be expressed in terms of the regularized semivariogram also

$$\sum_{i=1}^{N} w_i \gamma(\mathcal{D}_j, \mathcal{D}_i) + \lambda = \gamma(\mathcal{D}_j, \mathcal{D}_0)\,,\ j = 1, \ldots, N \tag{1.36}$$

$$\sum_{i=1}^{N} w_i = 1. \tag{1.37}$$

with the (ordinary) kriging variance

$$\sigma^2_{OK} = \sum_{i=1}^{N} w_i \gamma(\mathcal{D}_i, \mathcal{D}_0) + \lambda - \gamma(\mathcal{D}_0, \mathcal{D}_0). \tag{1.38}$$

Both kriging systems can be written in matrix form here expressed by means of the autocovariance functions

$$\boldsymbol{C}\boldsymbol{w} = \boldsymbol{D} \tag{1.39}$$

with

$$\boldsymbol{C} = \begin{bmatrix} C(\mathcal{D}_1, \mathcal{D}_1) & \cdots & C(\mathcal{D}_1, \mathcal{D}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(\mathcal{D}_N, \mathcal{D}_1) & \cdots & C(\mathcal{D}_N, \mathcal{D}_N) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \tag{1.40}$$

and

$$\boldsymbol{w}^T = [w_1, \ldots, w_N, -\lambda] \tag{1.41}$$

and

$$\boldsymbol{D}^T = [C(\mathcal{D}_1, \mathcal{D}_0), \ldots, C(\mathcal{D}_N, \mathcal{D}_0), 1]. \tag{1.42}$$

The autocovariance formulation is sometimes preferred because of programming efficiency (the largest elements in the symmetric covariance matrix $\{C_{ij}\}$ are on the diagonal; hence pivoting is not needed for the first N equations when solving the system).

The solution to the kriging system expressed in matrix form by means of the covariance function is

$$\boldsymbol{w} = \boldsymbol{C}^{-1}\boldsymbol{D}. \tag{1.43}$$

Of course, inversion of $\boldsymbol{C}$ is not needed to solve the system. A preferred way of solving the kriging system is to factorize $\boldsymbol{C}$ and solve the system by backward substitution.

If the support $\mathcal{D}_0$ can be considered as points the kriging performed is referred to as **point kriging**, otherwise it is referred to as **block kriging** or **panel kriging**. If block kriging is performed $\boldsymbol{D}$ is replaced by regularized values described in Section 1.4.

A few remarks on some very important properties of kriging:

- Kriging is an interpolation form that provides us with not only an estimate based on the covariance structure of the variable in question but it also provides us with an estimation variance.
- The kriging system has a unique solution if and only if the covariance matrix $\{C_{ij}\}$, $i, j = 1, \ldots, N$ is positive definite; this also ensures a non-negative kriging variance.
- The kriging estimator is a best linear unbiased estimator (BLUE) and it is also exact, i.e. if the support to be estimated coincides with any of the supports of the data included in the estimation, kriging provides an estimator equal to the known measurement and a zero kriging variance.
- The kriging system and the kriging variance depend only on the covariance function (semivariogram) and on the spatial lay-out of the sampled supports and not on the actual data values. If a covariance function is known (or assumed) this has important potential for minimizing the estimation variance in experimental design (i.e. in the planning phase of the spatial lay-out of the sampling scheme).

- Through the covariance function (semivariogram) $\boldsymbol{D}$ performs a statistical distance weighting of the data points in the support. $\boldsymbol{C}^{-1}$ rescales the weights in $\boldsymbol{D}$ to add to one, and also – again through the covariance function (semivariogram) – $\boldsymbol{C}^{-1}$ allows for possible redundancies in the support, i.e. it attempts to compensate for any possible clustering of the support points. The power of kriging is due to this combination.

1.6 Universal Kriging

In applications of ordinary kriging the problem of assuming stationarity arises. **Universal kriging (UK)** is a technique that allows for some forms of non-stationarity. The non-stationarity is modelled as a trend in the mean value described as a linear combination of known functions or as local Taylor expansions.

As in the case of ordinary kriging, suppose that the random variable $Z(\boldsymbol{x})$ is sampled on a number of supports (could be points) $\mathcal{D}_1, \ldots, \mathcal{D}_N$ giving scalar measurements $z(\boldsymbol{x}_1), \ldots, z(\boldsymbol{x}_N)$ considered as particular realizations of $Z_{\mathcal{D}_1}, \ldots, Z_{\mathcal{D}_N}$. We now want to estimate $Z_{\mathcal{D}_0}$ on a support $\mathcal{D}_0$ where Z is not sampled (or Z is sampled on a part of $\mathcal{D}_0$ only). Again, we are looking for a linear, unbiased estimator

$$\hat{Z}_{\mathcal{D}_0} = \sum_{i=1}^{N} w_i Z_{\mathcal{D}_i} \tag{1.44}$$

$$\mathrm{E}\{Z_{\mathcal{D}_0} - \hat{Z}_{\mathcal{D}_0}\} = 0. \tag{1.45}$$

As stated the mean is not assumed constant over the study area. Consider a trend in the mean of the form

$$\mu(\boldsymbol{x}) = \sum_{\ell=0}^{L} a_\ell f_\ell(\boldsymbol{x}), \tag{1.46}$$

where f_ℓ are known functions of $\boldsymbol{x}$ and a_ℓ are unknown parameters to be estimated. By convention $f_0 = 1$. In $\mathcal{R}^2$ (such as horizontal space) low order polynomials are often used for $f_\ell(\boldsymbol{x}) = f_\ell(x, y)$. E.g. in case of a quadratic trend ($L = 5$) we get

$$\mu(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy. \tag{1.47}$$

In the case of the stated trend in the mean, the unbiasedness gives

$$\sum_{i=1}^{N} w_i f_\ell(\mathcal{D}_i) = f_\ell(\mathcal{D}_0)\,,\ \ell = 0, \ldots, L, \tag{1.48}$$

where f_ℓ is the regularized value over the domain in question. The estimation variance (or the mean squared error) is ($w_0 = -1$)

$$\begin{aligned} \sigma_E^2 &= \mathrm{E}\{(Z_{\mathcal{D}_0} - \hat{Z}_{\mathcal{D}_0})^2\} \\ &= \mathrm{Var}\{Z_{\mathcal{D}_0} - \hat{Z}_{\mathcal{D}_0}\} \\ &= \sum_{i=0}^{N}\sum_{j=0}^{N} w_i w_j C(\mathcal{D}_i, \mathcal{D}_j). \end{aligned} \tag{1.49}$$

The **universal kriging estimate** is defined by the values of the weights w_i that minimize the estimation variance σ_E^2 subject to the constraint above on the weighted sum of the (universal kriging) weights. These weights can be found by introducing L Lagrangian multipliers and setting each of the N partial derivatives $\partial[\sigma_E^2 - 2\lambda_\ell(\sum_{i=1}^N w_i f_\ell(\mathcal{D}_i) - f_\ell(\mathcal{D}_0))]/\partial w_i = 0$ leading to the $(N + L + 1) \times (N + L + 1)$ set of equations

$$\sum_{i=1}^{N} w_i C(\mathcal{D}_j, \mathcal{D}_i) - \sum_{\ell=0}^{L} \lambda_\ell f_\ell(\mathcal{D}_j) = C(\mathcal{D}_j, \mathcal{D}_0)\,,\ j = 1, \ldots, N \tag{1.50}$$

$$\sum_{i=1}^{N} w_i f_\ell(\mathcal{D}_i) = f_\ell(\mathcal{D}_0)\,,\ \ell = 0, \ldots, L \tag{1.51}$$

with the **universal kriging variance** (or the minimum mean squared error)

$$\sigma^2_{UK} = -\sum_{i=1}^{N} w_i C(\mathcal{D}_i, \mathcal{D}_0) + \sum_{\ell=0}^{L} \lambda_\ell f_\ell(\mathcal{D}_j) + C(\mathcal{D}_0, \mathcal{D}_0). \tag{1.52}$$

Of course this can be expressed in terms of the regularized semivariogram also

$$\sum_{i=1}^{N} w_i \gamma(\mathcal{D}_j, \mathcal{D}_i) + \sum_{\ell=0}^{L} \lambda_\ell f_\ell(\mathcal{D}_j) = \gamma(\mathcal{D}_j, \mathcal{D}_0)\,,\ j = 1, \ldots, N \tag{1.53}$$

$$\sum_{i=1}^{N} w_i f_\ell(\mathcal{D}_i) = f_\ell(\mathcal{D}_0)\,,\ \ell = 0, \ldots, L \tag{1.54}$$

with the (universal) kriging variance

$$\sigma^2_{UK} = \sum_{i=1}^{N} w_i \gamma(\mathcal{D}_i, \mathcal{D}_0) + \sum_{\ell=0}^{L} \lambda_\ell f_\ell(\mathcal{D}_j) - \gamma(\mathcal{D}_0, \mathcal{D}_0). \tag{1.55}$$

Both kriging systems can be written in matrix form here expressed by means of the autocovariance functions

$$\boldsymbol{C}\boldsymbol{w} = \boldsymbol{D} \tag{1.56}$$

with

$$\boldsymbol{C} = \begin{bmatrix} C(\mathcal{D}_1,\mathcal{D}_1) & \cdots & C(\mathcal{D}_1,\mathcal{D}_N) & f_0(\mathcal{D}_1) & \cdots & f_L(\mathcal{D}_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C(\mathcal{D}_N,\mathcal{D}_1) & \cdots & C(\mathcal{D}_N,\mathcal{D}_N) & f_0(\mathcal{D}_N) & \cdots & f_L(\mathcal{D}_N) \\ f_0(\mathcal{D}_1) & \cdots & f_0(\mathcal{D}_N) & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_L(\mathcal{D}_1) & \cdots & f_L(\mathcal{D}_N) & 0 & \cdots & 0 \end{bmatrix} \tag{1.57}$$

and

$$\boldsymbol{w}^T = [w_1, \ldots, w_N, -\lambda_0, \ldots, -\lambda_L] \tag{1.58}$$

and

$$\boldsymbol{D}^T = [C(\mathcal{D}_1, \mathcal{D}_0), \ldots, C(\mathcal{D}_N, \mathcal{D}_0), f_0(\mathcal{D}_0), \ldots, f_L(\mathcal{D}_0)]. \tag{1.59}$$

Although some of the symbols from the deduction of the OK system are reused deducing the UK system, obviously the values they represent need not (and will indeed not) be the same.

The comments given with the OK system also apply here. Also, it is clear from the above that there is nothing universal about the universal kriging system of equations. The equations allow for a pre-defined trend in the mean, nothing more. Furthermore, we have what Armstrong (1984) calls a "chicken-and-egg" problem: We need the autocovariance function (or the semivariogram) for the UK system; if we try to estimate that we need the drift estimated from the UK system! Setting up iterative schemes to solve this problem has been reported to fail (Armstrong, 1984). The semivariogram of residuals gives a very biased estimate (an underestimate) of the true semivariogram, and it is extremely difficult to determine either the order of the drift or the type of the true semivariogram from the semivariogram of residuals. This problem limits the practical applicability of universal kriging.

Journel & Rossi (1989) report that if kriging is used for interpolation and data to be included in the kriging system is found in local windows (this is common practice) ordinary and universal kriging yield the same estimates, using data with a trend, for both the variable in question and its trend component. This apparent paradox is understandable when remembering that any type of kriging with data selected in local windows implies reestimating the mean at each new location. Modelling the trend matters only when extrapolating. Journel & Rossi (1989) thus recommend the simpler ordinary kriging scheme for this type of analysis. This author has good experience with estimating a possible regional trend before kriging and then adding this trend back again after.

1.7 Cokriging

A possible approach when interpolating multivariate observations is **cokriging**. Here one takes the spatial covariation between different variables into account. Algebraically, cokriging is not different from kriging. Suppose that the *multivariate* random variable $\boldsymbol{Z}(\boldsymbol{x})$ is sampled on a number of supports (could be points) $\mathcal{D}_1, \ldots, \mathcal{D}_N$ giving scalar measurements $z_i(\boldsymbol{x}_1), \ldots, z_i(\boldsymbol{x}_N), i = 1, \ldots, m$ where m is the number of variables sampled. The z_is are considered as particular realizations of $\boldsymbol{Z}_{\mathcal{D}_1}, \ldots, \boldsymbol{Z}_{\mathcal{D}_N}$. We now want to estimate $\boldsymbol{Z}_{\mathcal{D}_0}$ on a support $\mathcal{D}_0$ where $\boldsymbol{Z}$ is not sampled (or $\boldsymbol{Z}$ is sampled on a part of $\mathcal{D}_0$ only). Again, we are looking for a linear, unbiased estimator:

$$\hat{\boldsymbol{Z}}_{\mathcal{D}_0} = \sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{Z}_{\mathcal{D}_i} \tag{1.60}$$

where the cokriging weights $\boldsymbol{w}_i$ are found be solving the following system of linear equations

$$\begin{bmatrix} \boldsymbol{C}_{11} & \ldots & \boldsymbol{C}_{1N} & \boldsymbol{I} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{C}_{N1} & \ldots & \boldsymbol{C}_{NN} & \boldsymbol{I} \\ \boldsymbol{I} & \ldots & \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_N \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_{10} \\ \vdots \\ \boldsymbol{C}_{N0} \\ \boldsymbol{I} \end{bmatrix} \tag{1.61}$$

where $\boldsymbol{C}_{ij}$ is the cross-covariance between support points i and j, $\boldsymbol{C}_{i0}$ is the cross-covariance between the interpolation point and support point i, and $\boldsymbol{\mu}$ is a Lagrange multiplier. $\boldsymbol{C}_{ij}$, $\boldsymbol{w}_i$ and $\boldsymbol{\mu}$ are $m \times m$ matrices, where m is the dimensionality of $\boldsymbol{Z}$. $\boldsymbol{I}$ is the unit matrix and $\boldsymbol{0}$ is the null matrix, both of order m. If $m = 1$ the system is similar to the ordinary separate kriging system.

The estimation dispersion is

$$\boldsymbol{\sigma}^2_{CK} = \boldsymbol{C}_{00} - \boldsymbol{\mu} - \sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{C}_{i0}, \tag{1.62}$$

where $\boldsymbol{C}_{00}$ is the ordinary dispersion matrix of $\boldsymbol{Z}$. The diagonal elements of $\boldsymbol{\sigma}^2_{CK}$ are the **cokriging variances** of the individual variables.

Isaaks & Srivastava (1989) report that using a single non-bias condition (all weights, primary and secondary, sum to one) rather than the traditionally used m conditions, where primary weights sum to one and secondary weights sum to zero (as in Equation 1.61), gives better results because of the lack of inherent negative weights on secondary variables. This cokriging estimator requires additional information in the form of (sensible) estimates of the mean value of $\boldsymbol{Z}$.

In practice, the use of cokriging rather than individual kriging thus using more computer resources seems valuable only if the primary variable is undersampled compared to other variables with which the primary variable is spatially correlated.

1.8 Choice of Kriging Support

The support set used to krige to a given point can be chosen in several fashions. The traditional way is to choose all points within the range of influence no matter the possible under-representation of certain directions. If this is not feasible because of the number points involved, a search radius smaller than the range of influence is often chosen. As a first step to avoid under-representation of certain directions Conradsen, Ersbøll, Nielsen, Pedersen, Stern, & Windfeld (1991) suggest to use the same number of support points from each quadrant in a coordinate system with the kriging point as origin. An elaboration of this technique is based on the Delaunay triangulation.

1.8.1 Delaunay Triangulation

The Delaunay triangulation partitions the plane into triangles, where the edges constitute connections between neighboring points. No edges intersect. There are many possible triangulations in a point-set, but the Lawson criterion (Ripley, 1981) states that the most open triangles, that is, the triangles with the smallest angle as large as possible, should be chosen. The formal definition of the triangulation comes from the definition of the dual state, the so-called Voronoi tessellation.

Given a point C in the point-set $\mathcal{P}$. Points within the set having a Euclidean distance to C smaller than to any other point in $\mathcal{P}$ constitute the Voronoi polygon of C. Points equidistant from a pair of points, constitute the polygon boundaries. Points equidistant from three or more points constitute vertices. The Voronoi polygons of $\mathcal{P}$ constitute a Voronoi tessellation. The polygons are convex.

The Delaunay triangulation may now be defined in the following way: Two points are connected with an triangulation edge, if the Voronoi polygons of the two points have an edge in common.

Figure 1.4 shows a Delaunay triangulation and the dual Voronoi tessellation of the same point-set. Note that the triangulation edges form a convex hull spanned by $\mathcal{P}$, and that the points on the edge of the convex hull have open polygons.

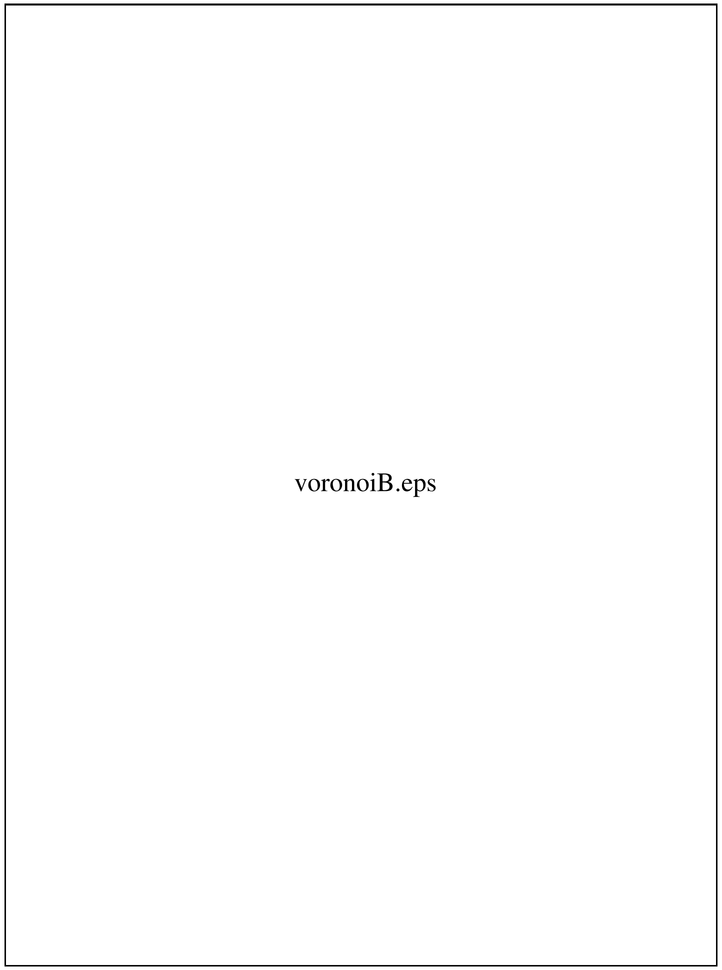

Figure 1.4: Voronoi tessellation (top), Delaunay triangulation (bottom)

The Delaunay triangulation has the property that the lines drawn from a point are fairly well distributed in all directions, regardless of the distance to the neighbors. This property may be used to define a neighborhood hierarchy in irregularly distributed point-sets. In regular grids, the first order neighborhood of a point (i, j) consists of the points $(i-1, j), (i, j-1), (i+1, j), (i, j+1)$, and the second order neighborhood consists of the first order neighborhood and the corners $(i-1, j-1), (i+1, j-1), (i+1, j+1), (i-1, j+1)$. Increasing the order is similar to including a set of points in the neighborhood which have equal Euclidean distances to (i, j) and are well balanced around (i, j). In the case of irregularly distributed points a similar neighborhood hierarchy may be defined in the following way: all points connected directly to the center-point are denoted first order neighbors, points which are connected through first order neighbors are denoted second order neighbors etc.

When performing interpolation by e.g. kriging the application of the Delaunay triangulation to set up the interpolation support introduces a problem of establishing the correct neighborhood for an new point not included in the original triangulation. This problem is solved in Conradsen, Nielsen, Windfeld, Ersbøll, Larsen, Hartelius, & Olsson (1993). This application of the Delaunay triangulation in choosing the kriging support set is believed to be new.

1.9 Case Studies

An interesting case is a regional study with nearly 34,000 stream sediments samples analyzed by ICP for the contents of 26 geochemical elements covering nearly 10,000 square kilometers given in Conradsen, Nielsen, & Windfeld (1992). An example on application of geostatistical methods in indoor surface sampling is Schneider, Petersen, Nielsen, & Windfeld (1990).

1.9.1 2-D Semivariograms

Figures 1.5 and 1.6 show 2-D semivariograms for 2,625 samples analyzed for the contents of 16 geochemical elements in an area in central Spain (data from

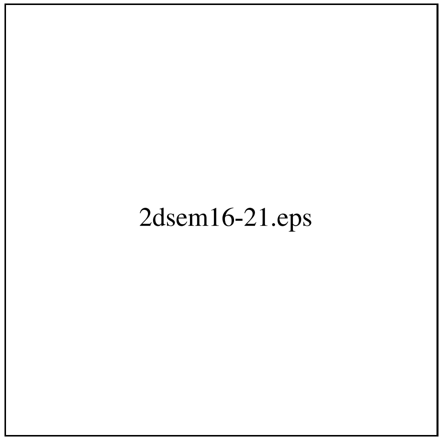

Figure 1.5: 2-D semivariograms for 16 geochemical elements in central Spain, 21×21 1 km pixels

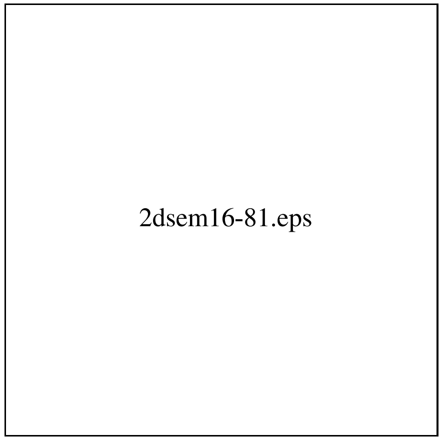

Figure 1.6: 2-D semivariograms for 16 geochemical elements in central Spain, 81×81 250 m pixels

Minas de Almadén y Arrayanes, S. A.). This case is given in GAF, MAYASA, IMSOR, & DLR (1993) also. In Figure 1.5 there are 21×21 1×1 km^2 pixels and in Figure 1.6 there are 81×81 250×250 m^2 pixels. Therefore the degree of detail revealed is different in the two images. The order is (row-wise from top-left) Pb, Zn, Cd, Hg, Cu, Ba, Mn, Ni, Co, Cr, Sn, W, Mo, V, Sb and Ag. Figures 1.7 and 1.8 show the same data as Figure 1.5 as contour plots and as perspective plots, respectively. These 2-D semivariograms clearly indicate anisotropies and differences herein. For example, for Mn we see a clear short range anisotropy in the NNE-SSW direction and a long range anisotropy in the NW-SE direction; for Ag we see a long range anisotropy in the E-W direction. In the words of Chief Geologist Dr. Enrique Ortega, MAYASA: "This result is very interesting because it indicates the spatial behaviour of each element as characterized by its migration capability. The elements with the highest mobility or with a uniform distribution over the entire test area, are logically represented as isotropic. Contrary to this, the fixed and low mobility elements are clearly anisotropic. The directions of anisotropy are closely related to the directions of the geological features (mainly faults) revealing their presence, position and orientation. This is valuable in future explorations campaigns because it provides information on the orientation of the mineralized structures which could facilitate location of drill holes. For these reasons MAYASA recommends continued application of this technique, e.g. on soils geochemistry data."

Figures 1.9 and 1.10 show 2-D semivariograms for 2,097 stream sediments samples analyzed by INAA or EDX for the contents of 41 geochemical elements from South Greenland (the Syduran Project, data from the Geological Survey of Greenland, GGU). In Figure 1.9 there are 21×21 5×5 km^2 pixels, and in Figure 1.10 there are 31×31 2×2 km^2 pixels. Again, the degree of detail revealed is different in the two images. The order is (row-wise from top-left) Au, Ag, As, Ba, Br, Ca, Co, Cr, Cs, Cu, Fe, Ga, Hf, K, Mn, Mo, Na, Ni, Pb, Rb, Sb, Sc, Se, Sr, Nb, Ta, Th, Ti, U, W, Y, Zn, Zr, La, Ce, Nd, Sm, Eu, Tb, Yb and Lu. Figures 1.11 and 1.12 show the same data as Figure 1.9 as contour plots and as perspective plots, respectively. In this case there are no formal comments from geologists but differences in anisotropy structure similar to those of the above case from central Spain are seen. For example, according to Chief

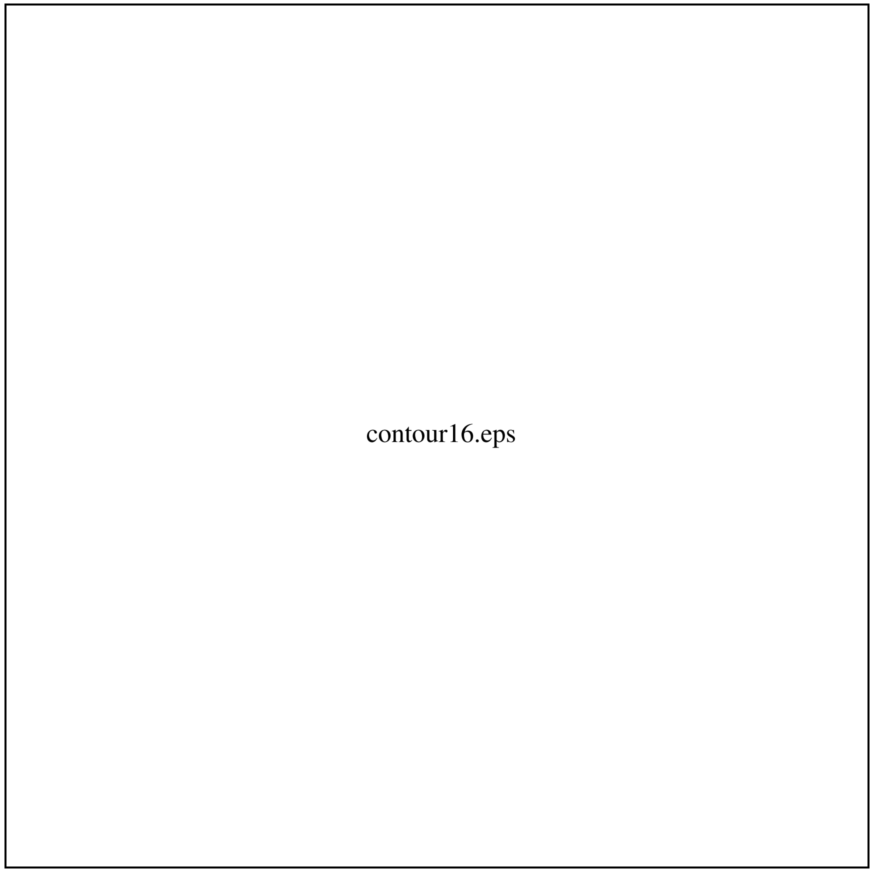

Figure 1.7: 2-D semivariograms for 16 geochemical elements in central Spain, 21×21 1 km pixels as contour plots

persp16.eps

Figure 1.8: 2-D semivariograms for 16 geochemical elements in central Spain, 21×21 1 km pixels as perspective plots

Geologist John Pedersen, Nunaoil, the content of Br in the samples is a question of distances to the sea only. We see that the 2-D semivariogram for Br depicts this fact: high autocorrelation between samples is seen along the coast (cf. Figures 1.15, 1.16, 1.17 or 2.7) where we have low values of the semivariogram.

It is expected that the value of this analysis is similar to that of the above case from central Spain.

1.9.2 Kriging versus Cokriging

Another case is kriging and cokriging (with Ta and Eu) of Nb from the above South Greenland study. All three variables are standardized to unit variance here. This case is given in GAF, MAYASA, IMSOR, & DLR (1993) also.

The performances of the two interpolation methods are examined by "leave-one-out" crossvalidation: all data points are estimated by leaving out the point itself and the estimate is then compared with the true value. An undersampled feature is simulated by replacing the data value with a missing value. After the estimation, the mean of the kriging variances (estimated from the kriging system) and the mean of the empirical variances (the mean of the squared differences between the true and the estimated value) are calculated.

The semivariograms and cross-semivariograms are shown in Figure 1.13.

A number of datasets with varying levels of undersampling in Nb are simulated. These datasets are interpolated by kriging and cokriging. The results expressed as the estimation variances and the ratio of the empirical variances of separate kriging and cokriging as functions of the degree of undersampling are shown in Figure 1.14. The neighborhood used in the interpolations are the 12 nearest neighbors.

When the level of undersampling is increased the estimation variance increases, but the increments are greater in the case of separate kriging than cokriging which shows that the cross-correlation between the features is indeed utilized. In the case of 95% undersampling the increment of the empirical cokriging

2dsem41-21.eps

Figure 1.9: 2-D semivariograms for 41 geochemical elements in South Greenland, 21×21 5 km pixels

2dsem41-31.eps

Figure 1.10: 2-D semivariograms for 41 geochemical elements in South Greenland, 31×31 2 km pixels

contour41.eps

Figure 1.11: 2-D semivariograms for 41 geochemical elements in South Greenland, 21×21 5 km pixels as contour plots

persp41.eps

Figure 1.12: 2-D semivariograms for 41 geochemical elements in South Greenland, 21×21 5 km pixels as perspective plots

NbTaEu.new.sem.ps

Figure 1.13: Cross-semivariograms for Nb, Ta and Eu

krig.new.ps

Figure 1.14: Estimation variances as functions of undersampling, left: cokriging, center: separate kriging, right: ratio of empirical variances of separate kriging and cokriging; square: kriging variance, cross: empirical variance

variance constitutes only 56% of the increment of the empirical separate kriging variance.

Figure 1.14 also shows that the kriging variance is a reliable estimate of the estimation variance.

Figures 1.15, 1.16 and 1.17 (north is to the left) show separately kriged Nb (no undersampling), separately kriged Nb (90% undersampling), and cokriged Nb (90% undersampling). The area shown is 321×211 km^2. The krigings in the undersampled cases were performed using all first, second, third and fourth order Delaunay neighbors. It is seen that the structures in the no undersampling case are best preserved by cokriging.

The results obtained have important practical implications for future mapping and exploration projects, namely the possibility of saving an important percentage of sampling and analysis. This translates directly into mapping and exploration cost savings.

Figure 1.15: Separately kriged Nb, no undersampling

Figure 1.16: Separately kriged Nb, 90% undersampling

Figure 1.17: Cokriged Nb, 90% undersampling

Chapter 2

Dimensionality Reduction

When one collects multivariate data in some field of application a redundancy effect often arises because of covariation between variables. An interesting issue in reduction of dimensionality of the data is the desire to obtain simplicity for better understanding, visualizing and interpreting the data on the one hand, and the desire to retain sufficient detail for adequate representation on the other hand. E.g. a remote sensing device typically measures the emitted intensity at a number of discrete wavelengths or wavelength intervals for each element in a regular grid. This "repetition" of the measurement at different wavelengths induces a high degree of redundancy in the dataset. This can be used for noise reduction and data compression.

A traditional method used in this context is the celebrated principal components transformation. This is a pixel-wise operation that does not take the spatial nature of image data into account. Also, principal components will not always produce components that show decreasing image quality with increasing component number. It is perfectly imaginable that certain types of noise have higher variance than certain types of signal components.

First I shall briefly consider the theory of principal components (see also Anderson (1984)), following this, two procedures for transformation of multivariate

data given a spatial grid (images) with the purpose of isolating signal from noise and data compression. These are the minimum/maximum autocorrelation factor transformation, which was first described by Switzer & Green (1984) and the maximum noise fractions transformation which was described by Green, Berman, Switzer, & Craig (1988). The methods are also described in Conradsen, Nielsen, & Thyrsted (1985), Ersbøll (1989), Conradsen & Ersbøll (1991), Conradsen, Nielsen, & Nielsen (1991a), Conradsen, Ersbøll, Nielsen, Pedersen, Stern, & Windfeld (1991), Larsen (1991). Also, an application of MAF/MNFs to remove periodic noise in hyper-channel airborne scanner data is described. Finally, a new concept of MAF/MNFs of irregularly spaced data is suggested. This also gives rise to a new form of kriging, namely maximum autocorrelation factorial or maximum signal-to-noise factorial kriging.

Non-orthogonal techniques such as projection pursuit (Friedman & Tukey, 1974; Windfeld, 1992) and non-linear techniques such as generalized principal components (Gnanadesikan, 1977) and MAFs by the alternating conditional expectations (ACE) algorithm (Breiman & Friedman, 1985; Windfeld, 1992) are not considered here.

2.1 Principal Components

Based on a technique described by Pearson in 1901 Hotelling (1933) developed principal components (PC) analysis. The principal components of a stochastic multivariate variable is a linear transformation which produces uncorrelated variables of decreasing variance.

The application of this transformation requires knowledge of or an estimate of the dispersion matrix. **The PCs maximize the variance represented by each component.** PC one is the linear combination of the original bands that explains maximum variance in the original data. A higher order PC is the linear combination of the original bands that explains maximum variance subject to the constraint that it is uncorrelated with lower order PCs. PC analysis performs a pixel-wise operation that does not take the spatial nature of an image into account.

Let us consider a multivariate data set of m bands with grey levels $Z_i(\boldsymbol{x}), i = 1, \ldots, m$, where $\boldsymbol{x}$ denotes the coordinates of the sample, and the dispersion is $\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{\Sigma}$, where $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$. We assume that $\mathrm{E}\{\boldsymbol{Z}\} = \mathbf{0}$.

We are looking for linear combinations of $\boldsymbol{Z}$

$$X(\boldsymbol{x}) = \boldsymbol{p}^T \boldsymbol{Z}(\boldsymbol{x}) \tag{2.1}$$

with maximum variance

$$\mathrm{Var}\{X(\boldsymbol{x})\} = \boldsymbol{p}^T \boldsymbol{\Sigma} \boldsymbol{p}. \tag{2.2}$$

If $\boldsymbol{p}$ is a solution so is $c\boldsymbol{p}$ where c is any scalar. We choose $\boldsymbol{p}$ so that $\boldsymbol{p}^T\boldsymbol{p} = 1$.

Let $\lambda_1 \geq \lambda_2, \geq \cdots \geq \lambda_m$ be the eigenvalues and $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_m$ corresponding orthogonal eigenvectors of $\boldsymbol{\Sigma}$. Then

$$X_i(\boldsymbol{x}) = \boldsymbol{p}_i^T \boldsymbol{Z}(\boldsymbol{x}), i = 1, \ldots, m \tag{2.3}$$

with variance λ_i is the i'th principal component.

We see that

$$\boldsymbol{X} = \boldsymbol{P}^T \boldsymbol{Z} \text{ with } \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}. \tag{2.4}$$

For the dispersion of $\boldsymbol{X}$ we get

$$\mathrm{D}\{\boldsymbol{X}\} = \boldsymbol{P}^T \boldsymbol{\Sigma} \boldsymbol{P} = \boldsymbol{\Lambda} \text{ with } \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix}. \tag{2.5}$$

This means that

$$\boldsymbol{\Sigma} = \boldsymbol{P}^{-T} \boldsymbol{\Lambda} \boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T = (\boldsymbol{P}\boldsymbol{\Lambda}^{\frac{1}{2}})(\boldsymbol{P}\boldsymbol{\Lambda}^{\frac{1}{2}})^T. \tag{2.6}$$

If we seek the i variables describing as much as possible of the total variance of the original variables the solution is the first i principal components. The fraction of the total variance described by these is given by

$$\frac{\lambda_1 + \cdots + \lambda_i}{\lambda_1 + \cdots + \lambda_i + \cdots + \lambda_m}. \tag{2.7}$$

A drawback of principal components analysis is that its results depend on the unit of measurement of the original variables. This problem can be circumvented by considering the standardized variables instead, i.e. by performing the PC transformation on the correlation matrix instead of on the dispersion matrix.

2.2 Min/Max Autocorrelation Factors

As opposed to the principal components transformation the minimum/maximum autocorrelation factors (MAF) transformation allows for the spatial nature of the image data. The application of this transformation requires knowledge of or an estimate of the dispersion matrix and the dispersion matrix of the spatially shifted image. The MAF transform minimizes the autocorrelation rather than maximizing the data variance (PC). **In reverse order the MAFs maximize the autocorrelation represented by each component.** MAF one is the linear combination of the original bands that contains minimum autocorrelation between neighboring pixels. A higher order MAF is the linear combination of the original bands that contains minimum autocorrelation subject to the constraint that it is orthogonal to lower order MAFs. The MAF procedure thus constitutes a (conceptually) more satisfactory way of orthogonalizing image data than PC analysis. The MAF transform is equivalent to a transformation of the data to a coordinate system in which the covariance matrix of the spatially shifted image data is the identity matrix followed by a principal components transformation. An important property of the MAF procedure is its invariance to linear transforms, a property not shared by ordinary PC analysis. This means that it doesn't matter whether the data have been scaled e.g. to unit variance before the analysis is performed.

The minimum/maximum autocorrelation factors procedure was suggested by Switzer & Green (1984). PCs, MAFs and other orthogonal transforms are described in Ersbøll (1989), Conradsen & Ersbøll (1991).

Again we consider the random variable $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$ and we assume that

$$\begin{aligned} \mathrm{E}\{\boldsymbol{Z}(\boldsymbol{x})\} &= \boldsymbol{0} && (2.8)\\ \mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} &= \boldsymbol{\Sigma}. && (2.9) \end{aligned}$$

We denote a spatial shift by $\boldsymbol{\Delta}^T = (\Delta_1, \Delta_2)$. The spatial covariance function is defined by

$$\mathrm{Cov}\{\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{\Gamma}(\boldsymbol{\Delta}). \tag{2.10}$$

$\boldsymbol{\Gamma}$ has the following properties

$$\begin{aligned} \boldsymbol{\Gamma}(\boldsymbol{0}) &= \boldsymbol{\Sigma} && (2.11)\\ \boldsymbol{\Gamma}(\boldsymbol{\Delta})^T &= \boldsymbol{\Gamma}(-\boldsymbol{\Delta}). && (2.12) \end{aligned}$$

We are interested in the correlations between projections of the variables and the shifted variables. Therefore we find

$$\begin{aligned} \mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a} && (2.13)\\ &= (\boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a})^T\\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T\boldsymbol{a}\\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(-\boldsymbol{\Delta})\boldsymbol{a}\\ &= \frac{1}{2}\boldsymbol{a}^T\,(\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}))\,\boldsymbol{a}. \end{aligned}$$

Introducing

$$\begin{aligned} \boldsymbol{\Sigma}_\Delta &= \mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} && (2.14)\\ &= \mathrm{E}\{[\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})][\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})]^T\}, \end{aligned}$$

which considered as a function of $\boldsymbol{\Delta}$ is a multivariate variogram, see Equation 1.2, we have

$$\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}) = 2\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta \tag{2.15}$$

and thus

$$\mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{a}^T(\boldsymbol{\Sigma} - \frac{1}{2}\boldsymbol{\Sigma}_\Delta)\boldsymbol{a} \tag{2.16}$$

wherefore

$$\mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}. \tag{2.17}$$

If we want to minimize that correlation we must maximize the Rayleigh coefficient

$$R(\boldsymbol{a}) = \frac{\boldsymbol{a}^T \boldsymbol{\Sigma}_\Delta \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}}. \tag{2.18}$$

Let $\kappa_1 \geq \cdots \geq \kappa_m$ be the eigenvalues and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ corresponding conjugate eigenvectors of $\boldsymbol{\Sigma}_\Delta$ with respect to $\boldsymbol{\Sigma}$. Then

$$\boldsymbol{Y}_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{Z}_i(\boldsymbol{x}) \tag{2.19}$$

is the i'th minimum/maximum autocorrelation factor or shortly the i'th MAF.

The minimum/maximum autocorrelation factors satisfy

i) $\mathrm{Corr}\{\boldsymbol{Y}_i(\boldsymbol{x}), \boldsymbol{Y}_j(\boldsymbol{x})\} = 0,\ i \neq j,$

ii) $\mathrm{Corr}\{\boldsymbol{Y}_i(\boldsymbol{x}), \boldsymbol{Y}_i(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\kappa_i,$

iii) $\mathrm{Corr}\{\boldsymbol{Y}_1(\boldsymbol{x}), \boldsymbol{Y}_1(\boldsymbol{x}+\boldsymbol{\Delta})\} = \inf_a \ \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\},$
$\mathrm{Corr}\{\boldsymbol{Y}_m(\boldsymbol{x}), \boldsymbol{Y}_m(\boldsymbol{x}+\boldsymbol{\Delta})\} = \sup_a \ \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\},$
$\mathrm{Corr}\{\boldsymbol{Y}_i(\boldsymbol{x}), \boldsymbol{Y}_i(\boldsymbol{x}+\boldsymbol{\Delta})\} = \inf_{a \in \mathcal{M}_i} \ \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\},$
$\mathcal{M}_i = \{\boldsymbol{a} \mid \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{Y}_j(\boldsymbol{x})\} = 0,\ j = 1, \ldots, i-1\}.$

The reverse numbering of MAFs so that the signal MAF is referred to as MAF1 is often used.

2.2.1 Linear Transformations of MAFs

We now consider the problem of transforming the original variables. If we set

$$\boldsymbol{U}(\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) \tag{2.20}$$

where $\boldsymbol{T}$ is a transformation matrix, we have that the MAF solution for $\boldsymbol{U}$ is obtained by investigating

$$R_1(\boldsymbol{b}) = \frac{\boldsymbol{b}^T \boldsymbol{T} \boldsymbol{\Sigma}_\Delta \boldsymbol{T}^T \boldsymbol{b}}{\boldsymbol{b}^T \boldsymbol{T} \boldsymbol{\Sigma} \boldsymbol{T}^T \boldsymbol{b}}. \tag{2.21}$$

The equation for solving the eigenproblem is

$$\begin{aligned} \boldsymbol{T}\boldsymbol{\Sigma}_\Delta \boldsymbol{T}^T \boldsymbol{v}_i &= \lambda_i \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T \boldsymbol{v}_i \Leftrightarrow \\ \boldsymbol{\Sigma}_\Delta(\boldsymbol{T}^T\boldsymbol{v}_i) &= \lambda_i \boldsymbol{\Sigma}(\boldsymbol{T}^T\boldsymbol{v}_i) \end{aligned} \tag{2.22}$$

i.e. the eigenvalues are unchanged and $\boldsymbol{T}^T\boldsymbol{v}_i = \boldsymbol{u}_i$ is an eigenvector for $\boldsymbol{\Sigma}_\Delta$ with respect to $\boldsymbol{\Sigma}$. We find that the MAFs in the transformed problem are

$$\begin{aligned} \boldsymbol{v}_i^T\boldsymbol{U}(\boldsymbol{x}) &= \boldsymbol{v}_i^T\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) \\ &= (\boldsymbol{T}^T\boldsymbol{v}_i)^T\boldsymbol{Z}(\boldsymbol{x}) \\ &= \boldsymbol{u}_i^T\boldsymbol{Z}(\boldsymbol{x}) \\ &= \boldsymbol{Y}_i(\boldsymbol{x}). \end{aligned} \tag{2.23}$$

Therefore the MAF solution is invariant to linear transformations, which can be useful in computations. Let $\lambda_1 \geq \cdots \geq \lambda_m$ be the ordinary eigenvalues and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ corresponding orthonormal eigenvectors of $\boldsymbol{\Sigma}$. If we—inspired by Equation 2.6—set

$$\boldsymbol{T}^T = \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{2.24}$$

we have for the dispersion of the transformed variables

$$D\{\boldsymbol{U}(\boldsymbol{x})\} = D\{\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \boldsymbol{I}. \tag{2.25}$$

With this transformation the original generalized eigenproblem is reduced to an ordinary eigenproblem for

$$\begin{aligned} \boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T &= D\{\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} \\ &= D\{\boldsymbol{U}(\boldsymbol{x}) - \boldsymbol{U}(\boldsymbol{x}+\boldsymbol{\Delta})\} \end{aligned} \tag{2.26}$$

and the MAF solution can be obtained by solving two ordinary eigenproblems as follows

- calculate principal components from the usual dispersion matrix $\boldsymbol{\Sigma}$,
- calculate dispersion matrix for shifted principal components $\boldsymbol{P}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{P}$,
- calculate principal components for transformed data corresponding to $\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}$.

The original generalized eigenproblem can be solved by means of Cholesky factorization of $\boldsymbol{\Sigma}$ also.

As far as the practical computation of $\hat{\boldsymbol{\Sigma}}_\Delta$ is concerned Switzer & Green (1984) recommend the formation of two sets of difference images. The two sets are $\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta}_h)$ and $\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta}_v)$ where $\boldsymbol{\Delta}_h$ is a unit horizontal shift and $\boldsymbol{\Delta}_v$ is a unit vertical shift. Calculate $\hat{\boldsymbol{\Sigma}}_{\Delta_h}$ and $\hat{\boldsymbol{\Sigma}}_{\Delta_v}$ and pool them to obtain $\hat{\boldsymbol{\Sigma}}_\Delta$.

2.3 Maximum Noise Fractions

Principal components do not always produce components of decreasing image quality. When working with spatial data the maximization of variance across

bands is not an optimal approach if the issue is this ordering. In this section we will maximize a measure of image quality, namely a signal-to-noise ratio. This should ensure achievement of the desired ordering in terms of image quality. In the previous section another measure of image quality namely spatial autocorrelation was dealt with.

If we estimate the noise at a pixel site as the difference of the pixel value at that site and the value of a neighboring pixel, we obtain the same eigenvectors as in the MAF analysis.

The maximum noise fractions (MNF) transformation can be defined in several ways. It can be shown that the same set of eigenvectors is obtained by procedures that maximize the signal-to-noise ratio and the noise fraction. The procedure was first introduced by Green et al. (1988) where the authors in continuation of the MAF work by Switzer & Green (1984) choose the latter. Hence the name maximum noise fractions.

The MNF transformation maximizes the noise content rather than maximizing the data variance (PC) or minimizing the autocorrelation (MAF). The application of this transformation requires knowledge of or an estimate of the signal and noise dispersion matrices. **In reverse order the MNFs maximize the signal-to-noise ratio represented by each component.** MNF one is the linear combination of the original bands that contains minimum signal-to-noise ratio. A higher order MNF is the linear combination of the original bands that contains minimum signal-to-noise ratio subject to the constraint that it is orthogonal to lower order MNFs. The MNF transform is equivalent to a transformation of the data to a coordinate system in which the noise covariance matrix is the identity matrix followed by a principal components transformation. The MNFs therefore also bear the name noise adjusted principal components (NAPC), cf. Lee, Woodyatt, & Berman (1990). The MNFs share the MAFs' property of invariance to linear transforms.

First we will deduce the maximum noise fraction transformation. We will then briefly describe methods for estimating the dispersion of the signal and the noise.

Let us as before consider a multivariate data set of m bands with grey levels $Z_i(\boldsymbol{x}), i = 1, \ldots, m$, where $\boldsymbol{x}$ denotes the coordinates of the sample. We will assume an additive noise structure

$$\boldsymbol{Z}(\boldsymbol{x}) = \boldsymbol{S}(\boldsymbol{x}) + \boldsymbol{N}(\boldsymbol{x}), \tag{2.27}$$

where $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$, and $\boldsymbol{S}(\boldsymbol{x})$ and $\boldsymbol{N}(\boldsymbol{x})$ are the uncorrelated signal and noise components. We assume that $\mathrm{E}\{\boldsymbol{Z}\} = \boldsymbol{0}$. Therefore

$$\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_S + \boldsymbol{\Sigma}_N, \tag{2.28}$$

where $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_N$ are the dispersion matrices for $\boldsymbol{S}(\boldsymbol{x})$ and $\boldsymbol{N}(\boldsymbol{x})$ respectively. Note that the techniques described in this section can in principle be applied to multiplicative noise also by first taking logarithms of the observations.

We define the signal-to-noise ratio of the i'th band as

$$\frac{\mathrm{Var}\{S_i(\boldsymbol{x})\}}{\mathrm{Var}\{N_i(\boldsymbol{x})\}}, \tag{2.29}$$

i.e. the ratio of the signal variance and the noise variance. We define the noise fraction of the i'th band as

$$\frac{\mathrm{Var}\{N_i(\boldsymbol{x})\}}{\mathrm{Var}\{Z_i(\boldsymbol{x})\}}, \tag{2.30}$$

i.e. the ratio of the noise variance and the total variance. We define the maximum noise fraction transformation as the linear transformations

$$Y_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{Z}(\boldsymbol{x}), i = 1, \ldots, m \tag{2.31}$$

such that the signal-to-noise ratio for $Y_i(\boldsymbol{x})$ is maximum among all linear transforms orthogonal to $Y_j(\boldsymbol{x}), j = 1, \ldots, i-1$. Furthermore we shall assume that the vectors $\boldsymbol{a}_i$ are normed so that

$$\boldsymbol{a}_i^T \boldsymbol{\Sigma} \boldsymbol{a}_i = 1, i = 1, \ldots, m. \tag{2.32}$$

Maximization of the noise fraction leads to the opposite numbering, namely a numbering that gives increasing image quality with increasing component number.

The SNR for $Y_i(\boldsymbol{x})$ is

$$\begin{aligned} \frac{\mathrm{Var}\{\boldsymbol{a}_i^T \boldsymbol{S}(\boldsymbol{x})\}}{\mathrm{Var}\{\boldsymbol{a}_i^T \boldsymbol{N}(\boldsymbol{x})\}} &= \frac{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_S \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i} \\ &= \frac{\boldsymbol{a}_i^T (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_N) \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i} \\ &= \frac{\boldsymbol{a}_i^T \boldsymbol{\Sigma} \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i} - 1. \end{aligned} \tag{2.33}$$

If instead we work on the noise fraction, we get

$$\frac{\mathrm{Var}\{\boldsymbol{a}_i^T \boldsymbol{N}(\boldsymbol{x})\}}{\mathrm{Var}\{\boldsymbol{a}_i^T \boldsymbol{Z}(\boldsymbol{x})\}} = \frac{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma} \boldsymbol{a}_i}. \tag{2.34}$$

In both cases we will find the vectors $\boldsymbol{a}_i$ as eigenvectors to the real, symmetric, generalized eigenproblem

$$\det(\boldsymbol{\Sigma}_N - \lambda \boldsymbol{\Sigma}) = 0. \tag{2.35}$$

Thus the SNR for $Y_i(\boldsymbol{x})$ is given by

$$\frac{1}{\lambda_i} - 1, \tag{2.36}$$

where λ_i is the eigenvalue of $\boldsymbol{\Sigma}_N$ with respect to $\boldsymbol{\Sigma}$.

An important characteristic of the MNF transformation which is not shared by the PC transformation is the invariability to linear scaling (the signal-to-noise *ratio* is maximized).

As for the MAFs the reverse numbering of MNFs so that the signal MNF is referred to as MNF1 is often used.

2.3.1 Estimation of the Noise Covariance Matrix

The central problem in the calculation of the MNF transformation is the estimation of the noise with the purpose of generating a dispersion matrix that approximates $\boldsymbol{\Sigma}_N$. In this process we will make use of the spatial characteristics of the image. Five methods are suggested, see also Olsen (1993)

- Simple differencing. The noise is estimated as for MAFs as the difference between the current and a neighboring pixel. In this case we refer to $\boldsymbol{\Sigma}_N$ as $\boldsymbol{\Sigma}_\Delta$.
- Causal SAR. The noise is estimated as the residual in a simultaneous autoregressive (SAR) model involving the neighboring pixel to the W, NW, N and NE of the current pixel.
- Differencing with the local mean. More pixels could be entered in to the estimation by differencing between the current pixel and the local mean.
- Differencing with local median. Mean filters blur edges and other details. This could be avoided by using the local median instead of the local mean.
- Quadratic surface. The noise is estimated as the residual from a fitted quadratic surface in a neighborhood.

2.3.2 Periodic Noise

As satellite images and images obtained from airborne scanners often are corrupted by striping we will consider methods for eliminating this form of noise. As periodic noise such as various forms of striping often has a high degree of spatial correlation, it will often be considered signal by the MAF and MNF transformations. To some extent the striping will be isolated in some of the factors. Periodic noise can be removed by Fourier methods. It should be noted that periodic noise can be very disturbing as the regular pattern catches the viewer's eyes.

A "naïve" bandwise Fourier filtering may corrupt significant parts of the relevant signal. Therefore we shall minimize the amount of filtering by introducing a new concept, namely to eliminate this noise by filtering out the relevant structures in the MNF Fourier domain. In order not to create an inverse pattern by setting the Fourier values to zero we keep the phase and fill the magnitude values by an iterative algorithm that takes means of the neighboring values. If we want to remove other types of noise also (e.g. salt-and-pepper noise) we can filter or skip the MNFs that contain the noise pattern in question before transforming back from MNF space to the original image space.

2.4 MAF/MNFs of Irregularly Spaced Data

Consider the above formulation of the MAF/MNF problem. For irregularly spaced data an alternative to the estimate of the noise dispersion $\boldsymbol{\Sigma}_N$ is $\boldsymbol{\Sigma}_k, k = 1, \dots, n$ from the cross-covariance function described in Section 1.2. These $\boldsymbol{\Sigma}_k$'s are *not* dispersion but covariance matrices. Therefore this definition of a noise dispersion matrix with different data points used in the estimation of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_N$ does not ensure that solutions to the real, symmetric, generalized eigensystem exist. Another estimate of $\boldsymbol{\Sigma}_N$ (or rather $\boldsymbol{\Sigma}_\Delta$) is simply the dispersion matrix of a new variable consisting of the difference between a data value and its nearest neighbor for all variables. This defines a new data analytical concept namely minimum/maximum autocorrelation factors (MAFs) for irregularly spaced multivariate data.

Analogous to the extension of MAFs into MNFs for gridded data, a more elaborate model for $\mathbf{\Sigma}_N$ based on each observation's neighborhood can be defined. With gridded data the neighborhood is easily defined. With non-gridded or irregularly spaced data, the Voronoi tessellation of the plane and its dual concept, the Delaunay triangulation are useful. To each point in the plane we associate a Voronoi polygon which is the part of the plane that is nearer to that point than to any other point. From the Voronoi tessellation we can construct the Delaunay triangulation by joining points with common Voronoi polygon edges. The Delaunay triangulation is described in Section 1.8.1.

Two ways of estimating $\mathbf{\Sigma}_N$ come directly to mind: the use of the dispersion matrix of a new variable consisting of the difference between a data value and the mean or the median of all its, say, first order Delaunay neighbors for all variables. This defines a new data analytical concept namely maximum noise fractions (MNFs) for irregularly spaced multivariate data.

Of course, both the MAFs and MNFs defined in this fashion can be extended to allow for other neighborhoods, e.g. confined by distance and/or direction constraints.

Grunsky & Agterberg (1988, 1991) circumvent the problem of irregularity of the sampling pattern by fitting parametric surfaces to observed correlations. Because of the lack of positive definiteness of the joint correlation structures this approach seems less satisfactory than the method proposed here.

The MAF/MNFs for irregularly spaced data possess the characteristics that they are orthogonal and they are ordered by decreasing autocorrelation/signal-to-noise ratio. Typically, low order factors will contain a lot of signal, high order factors will contain a lot of noise. The MAF/MNFs thus relate beautifully to the above interpolation by kriging, Section 1.5. Because the low order MAF/MNFs contain signal they are expected to have low nugget effects and long ranges of influence. This tendency is expected to develop towards higher nugget effects and shorter

ranges of influence as the higher order MAF/MNFs contain increasingly more noise. As the factors are orthogonal there is no cross-correlation for small lags ($|\boldsymbol{h}|$ small). I therefore suggest the use of separate (as opposed to co-) kriging for interpolation purposes and introduce a new data analytic concept: maximum autocorrelation factorial or maximum signal-to-noise factorial kriging based on the above new MAF/MNF analysis of non-gridded, multivariate data. In order to obtain kriged versions of the original data the inverse MAF/MNF transformation can be applied.

2.5 Case Studies

2.5.1 Noise in Hyperspectral GERIS Data

In this section the above orthogonal transformations are applied to a hyperspectral scene, namely a dataset recorded over central Spain using the Geophysical Environmental Research Corporation (GER) airborne scanner with the purpose of isolating signal from noise.

The GER imaging spectrometer (GERIS) actually consists of three spectrometers, that view the ground through the same aperture via an optoelectronic scanning device. The three spectrometers record a total of 63 bands through the visible, near infrared and short wave-infrared wavelength range between 0.47 and 2.45 μm. The spectral resolution in the visible region between 0.47 and 0.84 μm is 12.3 nm. In the near infrared region from 1.40 nm to 1.90 nm it is much broader, around 120 nm. In the short wave-infrared region between 2.00 and 2.45 μm the sampling frequency is 16.2 nm.

The scanner uses a rotating mirror perpendicular to the flight direction to scan a line of 512 pixels with a scan angle of 45° to either side of the flight track. A flight altitude of 3,000 meters and an aperture setting of 2.5 mrad leads to a nominal pixel size of 7.5 meters. The recorded data are stored in 16 bits with a dynamic range of 12 bits. After recording, the dataset is corrected for aircraft roll by the use of roll data recorded by a gyroscope hard mounted to the scanner optics.

Apart from noise introduced by the atmosphere, the instrumentation, and from quantization and sampling, the GERIS data are corrupted by a heavy two line and four line banding. This is due to slight differences of the surfaces of the rotating mirrors in the scanning device. These differences in the optical properties probably stem from dirt and oil on the surfaces.

PC versus MAF

In Figures 2.1 and 2.2 the 62 principal components and the 62 minimum/maximum autocorrelation factors are shown. The images are ordered row wise with component/factor 1 in the top-left corner (paper in landscape mode). Each subimage consists of 340×500 7.5×7.5 m^2 pixels. Because of the extreme noise content channel 28 is omitted from the analyses. It is evident that the principal components transformation is not capable of producing a natural ordering of image quality. The minimum/maximum autocorrelation factors do a much better job in terms of ordering as well as separating signal from noise. One might describe the MAF transformation as a decomposition of spatial frequency.

MNF/Fourier Noise Filtering

In Figure 2.3 we see the effect of filtering out the peaks in the MNF Fourier domain that result from the striping. The effect is dramatic in terms of improved image quality. The line banding causes very distinct peaks in the Fourier domain as seen in Figure 2.4. Three peaks in each half plane are easily detected. These peaks are replaced with an iterated local mean.

After filtering of the twenty first maximum noise fraction components (signal) and replacing the remaining MNFs with their mean value, we transform them back to the original space. In this fashion we remove all types of noise isolated in MNF21 through MNF62 including salt-and-pepper noise and a herringbone-like noise isolated in MNF21. The effect of this on the original channel number 1 can be seen in Figure 2.5. It is evident that a considerable improvement is obtained.

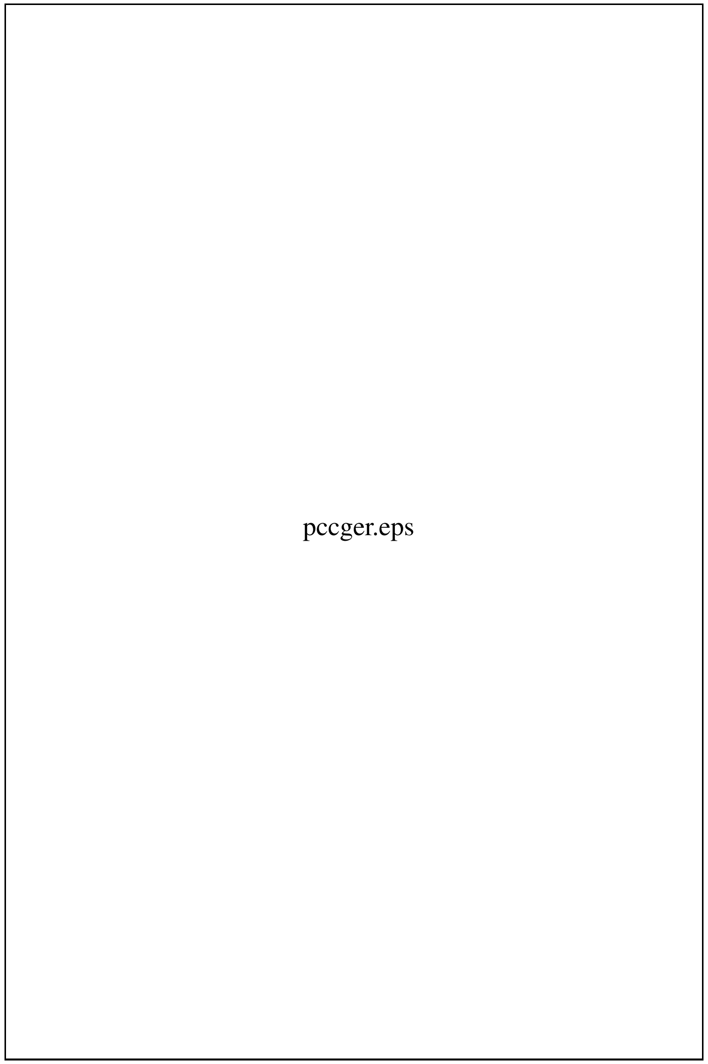

Figure 2.1: Principal components of 62 GERIS bands

mafger.eps

Figure 2.2: Minimum/maximum autocorrelation factors of 62 GERIS bands

In order to show the ability of the MNF transformation to concentrate the discriminatory power of the GERIS data in a few components we show in Figure 2.6 an RGB (red-green-blue) and an IHS (intensity-hue-saturation) combination of the restored MNFs 1, 2 and 3 respectively. As the first of the MNFs often contains most of the topographic features the IHS representation is often a good way of visualizing the MNF space. The intensity channel is stretched to match a beta distribution with parameters $\alpha = 4.0$ and $\beta = 4.0$, similar to a Gaussian. The hue channel is stretched to match a beta distribution with parameters $\alpha = 1.0$ and $\beta = 1.0$, this is histogram equalization. The saturation channel is stretched to match a beta distribution with parameters $\alpha = 2.0$ and $\beta = 1.0$, which is a linearly growing distribution.

This case is given in GAF, MAYASA, IMSOR, & DLR (1993), Nielsen & Larsen (1994) also. Similar noise reduction schemes are used in Conradsen, Nielsen, & Nielsen (1991a), Berman (1994). An artificial neural network approach to MNF noise filtering in a multichannel airborne magnetic survey is given in Pendock & Nielsen (1993).

2.5.2 MAFs and Irregular Sampling

This section deals with MAF/MNFs of irregularly spaced image data and MAF kriging. The data used are the 2,097 samples of 41 elements from South Greenland and the 2,625 samples of 16 elements from southern Spain used in the case studies of Chapter 1 also. Both are stream sediments geochemistry data. These cases are shown in GAF, MAYASA, IMSOR, & DLR (1993) also.

Figure 2.7 shows a geologic map of South Greenland (from Olesen (1984) who worked on geochemical data from South Greenland also). Figure 2.8 shows the sampling pattern in South Greenland with varimax rotated principal factors 1, 2 and 3 as red, green and blue. All images here are shown in landscape mode. Figure 2.9 shows MNFs 1, 2 and 3 as red, green and blue. Figure 2.10 shows the result of the MAF kriging procedure. MAFs 1, 2 and 3 are shown as red, green and blue respectively. A comparison of the results in the words of Senior Geologist Agnete Steenfelt (Geological Survey of Greenland, GGU):

Figure 2.3: MNF number 4 before (bottom) and after (top) MNF Fourier de-striping

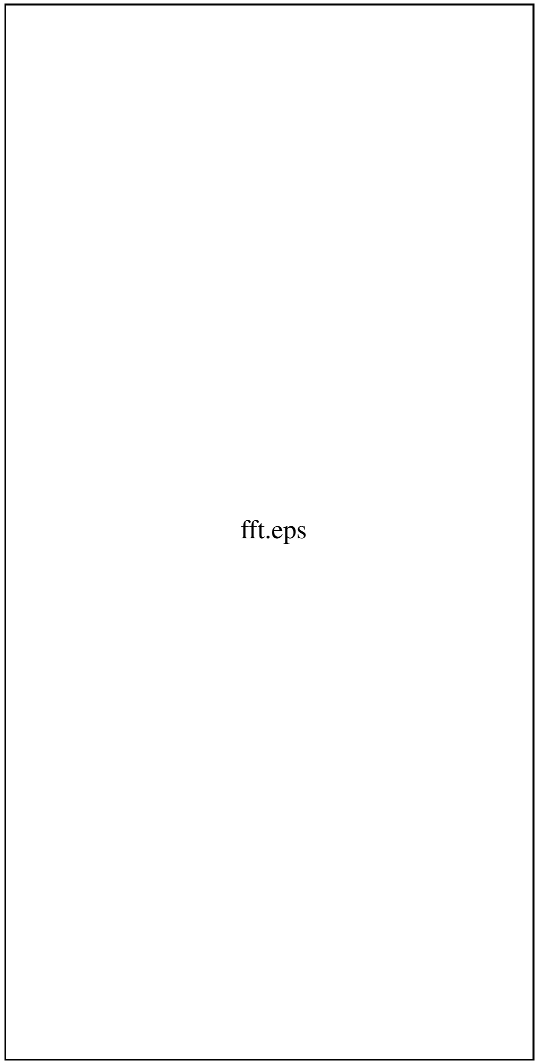

Figure 2.4: Fourier spectra of MNF number 4 before (bottom) and after (top) peak removal

ger1ba.eps

Figure 2.5: Channel number 1 before (bottom) and after (top) MNF Fourier destriping

mnf123.eps

Figure 2.6: The first three restored MNFs as RGB (bottom) and IHS (top)

geolmap.eps

Figure 2.7: Geological map of South Greenland

irrfac123.eps

Figure 2.8: South Greenland: Varimax rotated factors 1, 2 and 3 as RGB

irrmnf123.eps

Figure 2.9: South Greenland: MNFs 1, 2 and 3 as RGB

maf123OK.eps

Figure 2.10: South Greenland: MAF kriged factors 1, 2 and 3 as RGB

maf123ok.eps

Figure 2.11: Southern Spain: MAF kriged factors 1, 2 and 3 as RGB

"Contrary to results from ordinary non-spatial analysis of the data, this result clearly distinguishes the major lithotectonic units of South Greenland: northern Archean craton (dark blue), central Proterozoic unit (red and yellow) dominated by granite batholiths, Proterozoic alkaline intrusive complexes (white and bright magenta), and southern Proterozoic unit (cyan and green) dominated by supracrustal rocks and rapakivi granites." In the geological map in Figure 2.7 the northern Archean craton is denoted "Border Zone" and the southern Proterozoic unit is denoted "Migmatite Complex." The Proterozoic alkaline intrusive complexes is also referred to as the Gardar intrusions.

Figure 2.11 shows the result of the MAF kriging for the southern Spain area. Again MAFs 1, 2 and 3 are shown as red, green and blue respectively. In this case there are no written comments from geologists. However, an oral statement from Chief Geologist Dr. Enrique Ortega, Minas de Almadén y Arrayanes, S. A.(MAYASA), stresses the richness of the geological information about the area in that image.

In general, it is expected that this new spatial analysis technique will be very useful in the analysis of irregularly spaced multivariate data irrespective of the field of application. The technique has been applied successfully to environmental data, namely studies of leachate migration under a landfill and pollution with organic chemicals under a storage yard in Andersen (1994).

Chapter 3

Multiset Data Analysis

This chapter deals with a type of methods for multi-source data analysis namely linear canonical correlations analysis. Two-set canonical correlations analysis investigates the relationship between two groups of variables. It finds corresponding sets of linear combinations of the original two groups of variables. The first set of linear combinations are the ones with the largest correlation. This correlation is called the first canonical correlation and the two linear combinations are called the first canonical variates. The second set of linear combinations are the ones with the largest correlation subject to the condition that they are uncorrelated with the first canonical variates. This correlation is called the second canonical correlation and the two linear combinations are called the second canonical variates. Higher order canonical correlations and canonical variates are defined similarly.

First ordinary two-set canonical analysis is described in two slightly different ways. One way is the ordinary maximization of the correlation coefficient of the two transformed sets. The other way is the maximization of the variance of the sum of the transformed sets under unit variance constraint. An application of two-set canonical analysis is given with the introduction of a new orthogonal transformation, the multivariate alteration detection (MAD) transformation

(Conradsen & Nielsen, 1994) that finds linear combinations that give maximal multivariate differences. The method that maximizes the variance of the sum of the transformed sets under unit variance constraint is generalized to deal with more than two sets of variables. It is shown that this approach corresponds to maximizing the sum of the elements of the dispersion matrix of the transformed sets. Also, other methods of optimizing characteristics of this dispersion matrix are studied, namely (including the above method)

- maximization of the sum of the elements,
- maximization of the sum of the squared elements,
- maximization of the largest eigenvalue,
- minimization of the smallest eigenvalue, and
- minimization of the determinant.

These optimizations take place under different constraints, namely

- the projection vectors are unit vectors within each set,
- the sum of the projection vectors is a unit vector,
- the weighted projection vectors are unit vectors within each set, and
- the sum of the weighted projection vectors is a unit vector.

The above methods are used in change detection studies in remote sensing. Results from such analysis are linear combinations that transform the original sensor bands into new variables that show decreasing similarity over several (more than two) points in time. The higher order canonical variates have minimum similarity and they are therefore measures of change in all bands simultaneously. This technique is new in truly multi-temporal, multivariate change detection studies.

Other multi-source data analysis techniques such as the application of for example Mahalanobis' distance in joint distributions of multiset data to point out potentially interesting areas based on a training set, see Conradsen, Nielsen, & Nielsen (1991b), or spatial orthogonalization of joint distributions of multiset data, see Nielsen (1994a, 1994b), are not dealt with.

3.1 Two-set Canonical Correlations

Canonical correlations analysis was first introduced by Hotelling (1936) to analyze linear relations between two sets of variables. The technique is described in most standard textbooks on multivariate statistics, e.g. Cooley & Lohnes (1971), Anderson (1984). Work on non-linear canonical correlations analysis comprise van der Burg & de Leeuw (1983), Breiman & Friedman (1985), Buja (1990), Shi & Taam (1992), Windfeld (1992). This type of analysis will not be pursued here.

We consider a $p+q$ dimensional random variable ($p \leq q$) ideally following a Gaussian distribution split into two groups of dimensions p and q respectively

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \tag{3.1}$$

and we assume that $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ (and $\boldsymbol{\Sigma}$) are non-singular. Also, we assume that $\mathrm{E}\{\boldsymbol{X}\} = \mathrm{E}\{\boldsymbol{Y}\} = \boldsymbol{0}$.

We are searching for linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$\boldsymbol{U} = \boldsymbol{a}^T \boldsymbol{X}, \ \mathrm{V}\{\boldsymbol{U}\} = \boldsymbol{a}^T \boldsymbol{\Sigma}_{11} \boldsymbol{a} \tag{3.2}$$

$$\boldsymbol{V} = \boldsymbol{b}^T \boldsymbol{Y}, \ \mathrm{V}\{\boldsymbol{V}\} = \boldsymbol{b}^T \boldsymbol{\Sigma}_{22} \boldsymbol{b} \tag{3.3}$$

with maximum correlation

$$\rho = \mathrm{Corr}\{\boldsymbol{U}, \boldsymbol{V}\} = \frac{\mathrm{Cov}\{\boldsymbol{U}, \boldsymbol{V}\}}{\sqrt{\mathrm{V}\{\boldsymbol{U}\}\,\mathrm{V}\{\boldsymbol{V}\}}} = \frac{\boldsymbol{a}^T \boldsymbol{\Sigma}_{12} \boldsymbol{b}}{\sqrt{\boldsymbol{a}^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}\ \boldsymbol{b}^T \boldsymbol{\Sigma}_{22} \boldsymbol{b}}}. \tag{3.4}$$

If $(\boldsymbol{a}, \boldsymbol{b})$ is a solution so is $(c_1\boldsymbol{a}, c_2\boldsymbol{b})$ where c_i is any scalar. We choose $(\boldsymbol{a}, \boldsymbol{b})$ so that $\mathrm{V}\{\boldsymbol{U}\} = \mathrm{V}\{\boldsymbol{V}\} = 1$, introduce Lagrange multipliers $\nu/2$ and $\mu/2$ and maximize

$$F = \boldsymbol{a}^T \boldsymbol{\Sigma}_{12} \boldsymbol{b} - \frac{\nu}{2}(\boldsymbol{a}^T \boldsymbol{\Sigma}_{11} \boldsymbol{a} - 1) - \frac{\mu}{2}(\boldsymbol{b}^T \boldsymbol{\Sigma}_{22} \boldsymbol{b} - 1). \tag{3.5}$$

By setting $\partial F/\partial \boldsymbol{b} = 0$ and $\partial F/\partial \boldsymbol{a} = 0$, and inserting the results into the expression for ρ we get

$$\rho^2 = \frac{\boldsymbol{a}^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}} = \frac{\boldsymbol{b}^T \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{b}}{\boldsymbol{b}^T \boldsymbol{\Sigma}_{22} \boldsymbol{b}} \tag{3.6}$$

or

$$\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{a} = \rho^2 \boldsymbol{\Sigma}_{11} \boldsymbol{a} \tag{3.7}$$

$$\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{b} = \rho^2 \boldsymbol{\Sigma}_{22} \boldsymbol{b} \tag{3.8}$$

i.e. we find the desired projections for $\boldsymbol{X}$ by considering the conjugate eigenvectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ corresponding to the eigenvalues $\rho_1^2 \geq \cdots \geq \rho_p^2$ of $\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$ with respect to $\boldsymbol{\Sigma}_{11}$. Similarly, we find the desired projections for $\boldsymbol{Y}$ by considering the conjugate eigenvectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ of $\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$ with respect to $\boldsymbol{\Sigma}_{22}$ corresponding to the *same* eigenvalues ρ_i^2. If $p = q$ this will be all the eigenvalues and -vectors of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. If $q > p$ the last eigenvalue will be 0 with multiplicity $q - p$.

Another way of maximizing the above ρ with $\mathrm{V}\{\boldsymbol{U}\} = \mathrm{V}\{\boldsymbol{V}\} = 1$, consists of maximizing the variance of the sum of the linear combinations of the two sets of variables, i.e. maximize

$$V = \mathrm{V}\{\boldsymbol{a}^T\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{Y}\} = \boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a} + \boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b} + 2\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{b} = 2(1+\rho). \quad (3.9)$$

Again, we introduce Lagrange multipliers λ_1 and λ_2 and maximize

$$\begin{aligned} F &= \boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a} + \boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b} + 2\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{b} - \quad (3.10)\\ &\quad \lambda_1(\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a} - 1) - \lambda_2(\boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b} - 1) \end{aligned}$$

without constraints. Again, by setting $\partial F/\partial \boldsymbol{b} = 0$ and $\partial F/\partial \boldsymbol{a} = 0$, and inserting the results into the expression for V we get

$$\boldsymbol{\Sigma}_{11}\boldsymbol{a} + \boldsymbol{\Sigma}_{12}\boldsymbol{b} = \lambda_1\boldsymbol{\Sigma}_{11}\boldsymbol{a} \quad (3.11)$$

$$\boldsymbol{\Sigma}_{22}\boldsymbol{b} + \boldsymbol{\Sigma}_{21}\boldsymbol{a} = \lambda_2\boldsymbol{\Sigma}_{22}\boldsymbol{b}. \quad (3.12)$$

As $\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{b} = (\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{b})^T = \boldsymbol{b}^T\boldsymbol{\Sigma}_{21}\boldsymbol{a}$ we see that $\lambda_1 = \lambda_2 = \lambda$. Also, it is obvious that

$$\boldsymbol{a} = \frac{1}{\lambda - 1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{b} \quad (3.13)$$

$$\boldsymbol{b} = \frac{1}{\lambda - 1}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}. \quad (3.14)$$

The eigensystem written in this fashion is the same as the above eigensystem (Equations 3.7 and 3.8) with $\rho^2 = (\lambda - 1)^2$.

Letting δ_{ij} be the Kronecker delta ($\delta_{ij} = 1$ for $i = j$, $\delta_{ij} = 0$ otherwise) we have

$$\mathrm{Corr}\{\boldsymbol{a}_i^T\boldsymbol{X}, \boldsymbol{a}_j^T\boldsymbol{X}\} = \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}_j = \delta_{ij} \quad (3.15)$$

$$\mathrm{Corr}\{\boldsymbol{b}_i^T\boldsymbol{Y}, \boldsymbol{b}_j^T\boldsymbol{Y}\} = \boldsymbol{b}_i^T\boldsymbol{\Sigma}_{22}\boldsymbol{b}_j = \delta_{ij} \quad (3.16)$$

$$\mathrm{Corr}\{\boldsymbol{a}_i^T\boldsymbol{X}, \boldsymbol{b}_j^T\boldsymbol{Y}\} = \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{12}\boldsymbol{b}_j = \frac{1}{\rho_j}\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}_j = \rho_j\delta_{ij}. \quad (3.17)$$

We are now able to introduce the **canonical variates**

$$\boldsymbol{U}_i = \boldsymbol{a}_i^T\boldsymbol{X}, \; i = 1, \ldots, p \quad (3.18)$$

$$\boldsymbol{V}_i = \boldsymbol{b}_i^T\boldsymbol{Y}, \; i = 1, \ldots, p \quad (3.19)$$

and with an obvious choice of notation

$$\boldsymbol{U} = \boldsymbol{A}^T\boldsymbol{X} \quad \text{and} \quad \boldsymbol{V} = \boldsymbol{B}^T\boldsymbol{Y}, \quad (3.20)$$

where

$$\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p] \text{ is } p \times p \quad (3.21)$$

$$\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p] \text{ is } q \times p. \quad (3.22)$$

To facilitate interpretation we calculate the covariance between the original variables and the transformed variables

$$\mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{A}^T\boldsymbol{X}\} = \boldsymbol{\Sigma}_{11}\boldsymbol{A} \quad \text{and} \quad \mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{22}\boldsymbol{B} \tag{3.23}$$

$$\mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{12}\boldsymbol{B} \quad \text{and} \quad \mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{A}^T\boldsymbol{X}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{A}. \tag{3.24}$$

Because we are constructing new variables with maximum correlation there will be some redundancy in the transformed data. Canonical redundancy analysis is described in Appendix A.

An Interpretation of Canonical Variates

Consider a regression $\hat{\boldsymbol{X}}$ of $\boldsymbol{X}$ based on $\boldsymbol{Y}$ respectively $\hat{\boldsymbol{Y}}$ of $\boldsymbol{Y}$ based on $\boldsymbol{X}$

$$\hat{\boldsymbol{X}} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{Y} \tag{3.25}$$

$$\hat{\boldsymbol{Y}} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X} \tag{3.26}$$

For the dispersions we get

$$\mathrm{D}\{\hat{\boldsymbol{X}}\} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{3.27}$$

$$\mathrm{D}\{\hat{\boldsymbol{Y}}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \tag{3.28}$$

Linear combinations $\boldsymbol{a}^T\hat{\boldsymbol{X}}$ respectively $\boldsymbol{b}^T\hat{\boldsymbol{Y}}$ that maximize $\frac{\mathrm{Var}\{\boldsymbol{a}^T\hat{\boldsymbol{X}}\}}{\mathrm{Var}\{\boldsymbol{a}^T\boldsymbol{X}\}}$ respectively $\frac{\mathrm{Var}\{\boldsymbol{b}^T\hat{\boldsymbol{Y}}\}}{\mathrm{Var}\{\boldsymbol{b}^T\boldsymbol{Y}\}}$ fulfill

$$\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{a}} = \rho^2 \tag{3.29}$$

$$\frac{\boldsymbol{b}^T\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{b}} = \rho^2. \tag{3.30}$$

As these Rayleigh coefficients are identical to the defining Equations 3.7 and 3.8 above, the canonical variates can be interpreted as being new variables that maximize the ratio of the variances between linear combinations of predicted values of one set of variables from the other set of variables and the same linear combinations of the actual values of the one set of variables.

We also see that canonical correlation analysis can be considered as a type of regression analysis with several independent as well as dependent variables.

3.1.1 MAD Transformation

The above technique is used by Conradsen & Nielsen (1994) to find linear combinations that give maximal multivariate differences. The name chosen for the transformation, multivariate alteration detection (MAD), is due to the application in change detection in remote sensing (and the flashy acronym). Although it is presented here as an change detection technique in remote sensing, the technique applies to non-spatial multivariate differences also. The MAD transformation was first sketched in Conradsen & Nielsen (1991). Shettigara & McGilchrist (1989) use a hybrid canonical correlation/principal components technique to enhance uncorrelated parts of Landsat TM equivalents of ATM data in a gold exploration study. Change detection techniques based on canonical variates are also described in Hanaizumi & Fujimura (1992), Hanaizumi, Chino, & Fujimura (1994).

As opposed to traditional univariate change detection schemes the MAD scheme transforms two sets of multivariate observations (e.g. two multispectral satellite images covering the same geographical area acquired at different points in time) into a difference between two linear combinations of the original variables explaining maximal change (i.e. the difference explaining maximal variance) in all variables simultaneously.

When analyzing changes in panchromatic images taken at different points in time it is customary to analyze the difference between two images, possibly after some normalization. The idea is of course that areas with no or little changes come out with zero or low absolute values and areas with large changes come out with large absolute values in the difference image.

If we have two multivariate images with demeaned outcomes at a given pixel written as vectors

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \quad \text{resp.} \quad \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} \tag{3.31}$$

where k is the number of spectral bands, then a simple change detection transformation is

$$\boldsymbol{X} - \boldsymbol{Y} = \begin{bmatrix} X_1 - Y_1 \\ \vdots \\ X_k - Y_k \end{bmatrix}. \tag{3.32}$$

If our image data has more than three channels it is difficult to visualize changes in all channels simultaneously. To overcome this problem and to concentrate information on change, linear transformations of the image data that optimize some design criterion can be considered. A linear transformation that will maximize a measure of change in the simple multispectral difference image is one that maximizes deviations from no change, e.g. the variance

$$\mathrm{V}\{v_1(X_1 - Y_1) + \cdots + v_k(X_k - Y_k)\} = \mathrm{V}\{\boldsymbol{v}^T(\boldsymbol{X} - \boldsymbol{Y})\}. \tag{3.33}$$

A multiplication of vector $\boldsymbol{v}$ with a constant c will multiply the variance with c^2. Therefore we must put a restriction on $\boldsymbol{v}$. A natural restriction is requesting that $\boldsymbol{v}$ is a unit vector, $\boldsymbol{v}^T\boldsymbol{v} = 1$. This then amounts to finding principal components of the simple difference images. A disadvantage of this technique is that principal components are sensitive to the scale at which the individual variables are measured. Therefore they depend on for instance gain settings of a measuring device. A more versatile measure of change that allows different coefficients for $\boldsymbol{X}$ and $\boldsymbol{Y}$ and different number of spectral bands in the two sets, p and q respectively, are linear combinations

$$\boldsymbol{a}^T\boldsymbol{X} = a_1X_1 + \cdots + a_pX_p \tag{3.34}$$

$$\boldsymbol{b}^T\boldsymbol{Y} = b_1Y_1 + \cdots + b_qY_q \tag{3.35}$$

and the difference between them

$$\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}. \tag{3.36}$$

This measure also accounts for situations where the spectral bands are not the same but cover different spectral regions, for instance if one set of data comes from Landsat Thematic Mapper and the other set comes from SPOT High Resolution Visible. In this case one must be more cautious when interpreting the multivariate difference as multivariate change. In principle, any choice of $\boldsymbol{a}$ and $\boldsymbol{b}$ will give a measure of change. One could use principal components analysis on $\boldsymbol{X}$ to find an optimal $\boldsymbol{a}$ and on $\boldsymbol{Y}$ to find an optimal $\boldsymbol{b}$ (independent of $\boldsymbol{a}$). An improvement of this technique is to use principal components analysis on $\boldsymbol{X}$ and $\boldsymbol{Y}$ considered as *one* variable, cf. Fung & LeDrew (1987).

The Fung & LeDrew (1987) approach does not guarantee an optimal separation of $\boldsymbol{X}$ and $\boldsymbol{Y}$. It defines $\boldsymbol{a}$ and $\boldsymbol{b}$ simultaneously but the design criterion is not necessarily the one we want (for example, bands are treated similarly whether or not they come from different points in time). A potentially better approach is to define an optimal set of $\boldsymbol{a}$ and $\boldsymbol{b}$ simultaneously in the fashion described below. Again, let us maximize the variance, this time $\mathrm{V}\{\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}\}$. A multiplication of $\boldsymbol{a}$ and $\boldsymbol{b}$ with a constant c will multiply the variance with c^2. Therefore we must put some restrictions on $\boldsymbol{a}$ and $\boldsymbol{b}$, and natural restrictions in this case are requesting unit variance of $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{Y}$.

The criterion then is

$$\text{maximize } \mathrm{V}\{\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}\} \tag{3.37}$$

subject to the constraints

$$\mathrm{V}\{\boldsymbol{a}^T\boldsymbol{X}\} = \mathrm{V}\{\boldsymbol{b}^T\boldsymbol{Y}\} = 1. \tag{3.38}$$

Under these constraints we have

$$\begin{aligned} \mathrm{V}\{\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}\} &= \mathrm{V}\{\boldsymbol{a}^T X\} + \mathrm{V}\{\boldsymbol{b}^T\boldsymbol{Y}\} - 2\mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\} \\ &= 2(1 - \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\}). \end{aligned} \tag{3.39}$$

As we are talking difference (or change) detection here, we shall request that $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{Y}$ are positively correlated, i.e. $\mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\} \geq 0$. Therefore, determining the difference between linear combinations with maximum variance corresponds to determining linear combinations with minimum correlation (≥ 0). We assume that $\boldsymbol{a}$ and $\boldsymbol{b}$ are chosen so that the correlation between $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{Y}$ is positive. Positive correlation may simply be obtained by a change of sign if necessary.

The main idea behind the MAD transformation is now that as corresponding pairs of canonical variates are linear combinations of the original variables ordered by correlation or similarity between pairs, it seems natural to base a change detection scheme on differences between pairs of variates that show minimum similarity, i.e. the higher order canonical variates.

The main mathematical idea is to modify the theory used in defining canonical variates. This could be viewed as a time analog to the introduction of minimum/maximum autocorrelation factors in the spatial domain, cf. Section 2.2.

We define the **multivariate alteration detection (MAD)** transformation as

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{a}_p^T\boldsymbol{X} - \boldsymbol{b}_p^T\boldsymbol{Y} \\ \vdots \\ \boldsymbol{a}_1^T\boldsymbol{X} - \boldsymbol{b}_1^T\boldsymbol{Y} \end{bmatrix}, \tag{3.40}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are defined as in Section 3.1, i.e. $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the defining coefficients from a standard canonical correlations analysis. The MAD transformation has the very important property that if we consider linear combinations of two sets of p resp. q ($p \leq q$) variables that are positively correlated then the p'th difference shows maximum variance among such variables. The $(p-j)$'th difference shows maximum variance subject to the constraint that this difference is uncorrelated with the previous j ones. In this way we may sequentially extract uncorrelated difference images where each new image shows maximum difference (change) under the constraint of being uncorrelated with the previous ones.

If $p < q$ then the projection of $\boldsymbol{Y}$ on the eigenvectors corresponding to the eigenvalues 0 will be independent of $\boldsymbol{X}$. That part may of course be considered the extreme case of multivariate change detection.

The MAD transformation is objective in the sense that given the same geographical areas and variables there is full reproducibility and repeatability of the numerical results. If an initial MAD analysis to find areas of maximum change in all spectral bands simultaneously is performed, the areas located will always be the same. The analyst can control the method if used in a more directed fashion in that she or he can choose spatial and/or spectral subsets on which the analysis should be performed.

To facilitate interpretation we calculate the covariance between the original variables and the transformed variables

$$\mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{A}^T\boldsymbol{X} - \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{11}\boldsymbol{A} - \boldsymbol{\Sigma}_{12}\boldsymbol{B} \tag{3.41}$$

$$\mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{A}^T\boldsymbol{X} - \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{A} - \boldsymbol{\Sigma}_{22}\boldsymbol{B}. \tag{3.42}$$

The multivariate alteration detection (MAD) transformation gives an optimal (in the sense of maximal variance) detection of alterations (differences, changes) from one scene to the other in all spectral channels simultaneously. The transformation is invariant to linear scaling. It also provides a statistical analysis and it offers an interpretation of the nature of the alterations. The MAD transformation can be used iteratively. First, it can be used to detect outliers (such as drop-outs) and in a second iteration, it can be used to perform the actual change detection after appropriate action on the outliers.

Irrespective of the application in question and the individual analyst's favorite change detection scheme, the absolute value of the first MAD component will always outline the areas with the largest overall changes in all channels simultaneously. The correlations between the original image channels and the MADs form a basis for interpretation of the MADs. Based on the absolute values of the MADs and the interpretation, a more physically oriented change study can be performed if so desired.

We conclude that the MAD transformation is a useful supplement to univariate and existing multivariate change detection schemes. The MAD technique is believed to be useful with multichannel data in monitoring and surveillance in environmental studies, oceanography, agriculture, geobotany etc. For an oceanographic application of the MAD transformation suggested here, cf. Simpson (1994).

3.2 Multiset Canonical Correlations

Generalized or multiset canonical correlations analysis is a technique for analyzing linear relations between more (than two) sets of variables. Earlier work in this field comprise Vinograde (1950), Steel (1951), Horst (1961), Kettenring (1971). Royer & Mallet (1982) give an interesting example using satellite data and two types of geochemical data.

We consider an $m = m_1 + m_2 + \cdots + m_n$ dimensional random variable $\boldsymbol{X}$ ideally following a Gaussian distribution split into n groups of dimensions m_1, m_2 to m_n ($m_1 \leq m_2 \leq \cdots \leq m_n$) respectively

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix} \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \tag{3.43}$$

$$N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_n \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \right)$$

and we assume that the appropriate dispersion matrices are non-singular (depending on the constraints given below). Also, we assume that $\boldsymbol{\mu}_i = \boldsymbol{0}$. Of course $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}_{ji}^T$.

An obvious extension from the two-set case is to search for linear combinations $\boldsymbol{U}^T = [\boldsymbol{U}_1^T, \boldsymbol{U}_2^T, \ldots, \boldsymbol{U}_n^T]$ of $\boldsymbol{X}^T = [\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \ldots, \boldsymbol{X}_n^T]$

$$\boldsymbol{U}_1 = \boldsymbol{a}_1^T\boldsymbol{X}_1, \ \mathrm{V}\{\boldsymbol{U}_1\} = \boldsymbol{a}_1^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}_1 \tag{3.44}$$

$$\boldsymbol{U}_2 = \boldsymbol{a}_2^T\boldsymbol{X}_2, \ \mathrm{V}\{\boldsymbol{U}_2\} = \boldsymbol{a}_2^T\boldsymbol{\Sigma}_{22}\boldsymbol{a}_2 \tag{3.45}$$

$$\vdots$$

$$\boldsymbol{U}_n = \boldsymbol{a}_n^T \boldsymbol{X}_n, \ \mathrm{V}\{\boldsymbol{U}_n\} = \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{nn} \boldsymbol{a}_n \tag{3.46}$$

with maximum variance

$$V = \mathrm{V}\{\boldsymbol{a}^T \boldsymbol{X}\} = \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a} \tag{3.47}$$

where $\boldsymbol{a}^T = [\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \ldots, \boldsymbol{a}_n^T]$. Another way of expressing this variance is of course

$$\begin{aligned} V &= \mathrm{V}\{\sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{X}_i\} \\ &= \mathrm{V}\{\boldsymbol{a}_1^T \boldsymbol{X}_1 + \boldsymbol{a}_2^T \boldsymbol{X}_2 + \cdots + \boldsymbol{a}_n^T \boldsymbol{X}_n\} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j. \end{aligned} \tag{3.48}$$

We see that maximizing the above variance corresponds to maximizing the sum of the elements in the covariance matrix of the transformed variables

$$\boldsymbol{\Sigma}_U = \begin{bmatrix} \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}_1 & \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{1n} \boldsymbol{a}_n \\ \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{21} \boldsymbol{a}_1 & \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{22} \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{2n} \boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{n1} \boldsymbol{a}_1 & \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{n2} \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{nn} \boldsymbol{a}_n \end{bmatrix}. \tag{3.49}$$

The sum of covariances is only one of several natural and sensible measures to optimize. A list of measures including the above is

1. maximize sum of elements ($V = \sum_{i=1}^n \sum_{j=1}^n \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j$),

2. maximize sum of squared elements ($W = \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j)^2$),
3. maximize largest eigenvalue (λ_1),
4. minimize smallest eigenvalue (λ_n) and
5. minimize determinant ($\det \boldsymbol{\Sigma}_U = \prod_{i=1}^n \lambda_i$).

Kettenring (1971) lists all these possibilities and solves the problems involved; he names them 1. Sumcor, 2. Ssqcor, 4. Maxvar, 3. Minvar and 5. Genvar.

Several natural constraints under which to carry out the optimizations come to mind

1. $\boldsymbol{a}_i^T \boldsymbol{a}_i = 1$,
2. $\boldsymbol{a}^T \boldsymbol{a} = \sum_{i=1}^n \boldsymbol{a}_i^T \boldsymbol{a}_i = 1$,
3. $\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = 1$, or
4. $\boldsymbol{a}^T \boldsymbol{\Sigma}_D \boldsymbol{a} = \sum_{i=1}^n \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = \mathrm{tr}\ \boldsymbol{\Sigma}_U = 1$

with

$$\boldsymbol{\Sigma}_D = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix}. \tag{3.50}$$

In the two-set case all the above methods with constraints 3 and 4 reduce to the standard Hotelling case described in Section 3.1.

Horst (1961) examines the Sumcor method and Kettenring (1971) examines all the above methods using constraint 3.

3.2.1 Maximize Sum of Covariances

To maximize the sum of covariances under constraints we use a Lagrange multiplier technique. The four above constraints are examined.

Constraint 1: $\boldsymbol{a}_i^T \boldsymbol{a}_i - 1 = 0$

Introduce

$$F = V - \sum_{i=1}^{n} \lambda_i (\boldsymbol{a}_i^T \boldsymbol{a}_i - 1). \tag{3.51}$$

By setting $\partial F / \partial \boldsymbol{a}_i = \mathbf{0}$ we get

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j = \lambda_i \boldsymbol{a}_i,\ i = 1, \ldots, n \tag{3.52}$$

or

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \boldsymbol{a}_1 \\ \lambda_2 \boldsymbol{a}_2 \\ \vdots \\ \lambda_n \boldsymbol{a}_n \end{bmatrix}. \tag{3.53}$$

This is not a normal eigensystem.

Constraint 2: $\boldsymbol{a}^T \boldsymbol{a} - 1 = \sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{a}_i - 1 = 0$

Introduce

$$F = V - \lambda (\sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{a}_i - 1). \tag{3.54}$$

By setting $\partial F / \partial \boldsymbol{a}_i = \mathbf{0}$ we get

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j = \lambda \boldsymbol{a}_i,\ i = 1, \ldots, n \tag{3.55}$$

or

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}. \tag{3.56}$$

This is a normal real, symmetric eigensystem. It corresponds to principal components analysis of $\boldsymbol{X}$.

Constraint 3: $\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 = 0$

Introduce

$$F = V - \sum_{i=1}^{n} \lambda_i (\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1). \tag{3.57}$$

By setting $\partial F/\partial \boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j = \lambda_i \boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i, \; i = 1, \dots, n \tag{3.58}$$

or

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{\Sigma}_{11} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2\boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \lambda_n\boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}. \tag{3.59}$$

This is not a normal generalized eigensystem.

Constraint 4: $\boldsymbol{a}^T\boldsymbol{\Sigma}_D\boldsymbol{a} - 1 = \sum_{i=1}^{n} \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1 = 0$

Introduce

$$F = V - \lambda(\sum_{i=1}^{n} \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1). \tag{3.60}$$

By setting $\partial F/\partial \boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j = \lambda \boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i, \; i = 1, \dots, n \tag{3.61}$$

or

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}. \tag{3.62}$$

This is a normal real, symmetric, generalized eigensystem, i.e. we find the desired projections for $\boldsymbol{X}_i$ by computing the conjugate eigenvectors $\boldsymbol{a}_{i1}, \boldsymbol{a}_{i2}, \dots, \boldsymbol{a}_{im_1}$ corresponding to the first $m_1 = \min(m_1, m_2, \dots, m_n)$ eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{m_1}$ of the above eigensystem. In this case we assume that $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ to $\boldsymbol{\Sigma}_{nn}$ (and $\boldsymbol{\Sigma}$) are non-singular.

Letting δ_{ij} be the Kronecker delta ($\delta_{ij} = 1$ for $i = j$, $\delta_{ij} = 0$ otherwise) we have

$$\sum_{i=1}^{n} \mathrm{Cov}\{\boldsymbol{a}_{ik}^T\boldsymbol{X}_i, \boldsymbol{a}_{il}^T\boldsymbol{X}_i\} = \sum_{i=1}^{n} \boldsymbol{a}_{ik}^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_{il} = \delta_{kl}, \quad k = 1, \dots, m_i \;\text{ and }\; l = 1, \dots, m_j \tag{3.63}$$

$$\mathrm{Cov}\{\boldsymbol{a}_{ik}^T\boldsymbol{X}_i, \boldsymbol{a}_{jk}^T\boldsymbol{X}_j\} = \boldsymbol{a}_{ik}^T\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_{jk}. \tag{3.64}$$

We are now able to define the **generalized canonical variates**

$$\boldsymbol{U}_{1k} = \boldsymbol{a}_{1k}^T\boldsymbol{X}_1, \; k = 1, 2, \dots, m_1 \tag{3.65}$$

$$\boldsymbol{U}_{2k} = \boldsymbol{a}_{2k}^T\boldsymbol{X}_2, \; k = 1, 2, \dots, m_1 \tag{3.66}$$

$$\vdots$$

$$\boldsymbol{U}_{nk} = \boldsymbol{a}_{nk}^T \boldsymbol{X}_n, \;\; k = 1, 2, \ldots, m_1 \tag{3.67}$$

and with an obvious choice of notation

$$\boldsymbol{W}_1 = \boldsymbol{A}_1^T \boldsymbol{X}_1 \tag{3.68}$$

$$\boldsymbol{W}_2 = \boldsymbol{A}_2^T \boldsymbol{X}_2 \tag{3.69}$$

$$\vdots$$

$$\boldsymbol{W}_n = \boldsymbol{A}_n^T \boldsymbol{X}_n \tag{3.70}$$

where

$$\boldsymbol{A}_1 = [\boldsymbol{a}_{11}, \boldsymbol{a}_{12}, \ldots, \boldsymbol{a}_{1m_1}] \text{ is } m_1 \times m_1 \tag{3.71}$$

$$\boldsymbol{A}_2 = [\boldsymbol{a}_{21}, \boldsymbol{a}_{22}, \ldots, \boldsymbol{a}_{2m_1}] \text{ is } m_2 \times m_1 \tag{3.72}$$

$$\vdots$$

$$\boldsymbol{A}_n = [\boldsymbol{a}_{n1}, \boldsymbol{a}_{n2}, \ldots, \boldsymbol{a}_{nm_1}] \text{ is } m_n \times m_1. \tag{3.73}$$

To facilitate interpretation we calculate the covariance between the original variables and the transformed variables

$$\mathrm{Cov}\{\boldsymbol{X}_i, \boldsymbol{A}_j^T \boldsymbol{X}_j\} = \boldsymbol{\Sigma}_{ij} \boldsymbol{A}_j. \tag{3.74}$$

With constraint 4 if $n = 2$ we get

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{bmatrix}. \tag{3.75}$$

This leads to

$$\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} = (\lambda - 1)^2 \boldsymbol{\Sigma}_{11} \boldsymbol{a}_1 \tag{3.76}$$

$$\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} = (\lambda - 1)^2 \boldsymbol{\Sigma}_{22} \boldsymbol{a}_2 \tag{3.77}$$

which is the ordinary expression for canonical correlations analysis of two sets of variables.

As with the derivation of the eigensystem for canonical correlations analysis for $n = 2$ given in Section 3.1, it is easily shown that constraint 3 leads to the same eigensystem as constraint 4. Using the fact that when $\boldsymbol{a}_1^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}_1 = \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{22} \boldsymbol{a}_2 = 1$, $\boldsymbol{a}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{a}_2 = (\boldsymbol{a}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{a}_2)^T = \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{21} \boldsymbol{a}_1$ shows that $\lambda_1 = \lambda_2$.

3.2.2 Maximize Sum of Squared Covariances

To maximize the sum of squared covariances under constraints we use a Lagrange multiplier technique as in the previous section. Also, the same four constraints are examined. This treatment is given in Appendix B.

3.2.3 Maximize Largest Eigenvalue

Because $\boldsymbol{\Sigma}_{ii}$ is positive definite it can be Cholesky factorized

$$\boldsymbol{\Sigma}_{ii} = \boldsymbol{C}_i \boldsymbol{C}_i^T, \tag{3.78}$$

where $\boldsymbol{C}_i$ is lower triangular. Hence,

$$\boldsymbol{C}_i^{-1}\boldsymbol{\Sigma}_{ii}\boldsymbol{C}_i^{-T} = \boldsymbol{I}. \tag{3.79}$$

Instead of studying $\boldsymbol{X}$ we study a new variable $\boldsymbol{Y}^T = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T, \ldots, \boldsymbol{Y}_n^T]$ which we shall define shortly. We rewrite

$$\boldsymbol{U}_i = \boldsymbol{a}_i^T\boldsymbol{X}_i \quad \text{with} \quad \mathrm{Var}\{\boldsymbol{U}_i\} = \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i \tag{3.80}$$

to

$$\boldsymbol{U}_i = \boldsymbol{a}_i^T\boldsymbol{C}_i\boldsymbol{C}_i^{-1}\boldsymbol{X}_i \quad \text{with} \quad \mathrm{Var}\{\boldsymbol{U}_i\} = \boldsymbol{a}_i^T\boldsymbol{C}_i\boldsymbol{C}_i^T\boldsymbol{a}_i \tag{3.81}$$

$$\boldsymbol{U}_i = (\boldsymbol{C}_i^T\boldsymbol{a}_i)^T\boldsymbol{C}_i^{-1}\boldsymbol{X}_i \quad \text{with} \quad \mathrm{Var}\{\boldsymbol{U}_i\} = (\boldsymbol{C}_i^T\boldsymbol{a}_i)^T\boldsymbol{C}_i^T\boldsymbol{a}_i \tag{3.82}$$

$$\boldsymbol{U}_i = \boldsymbol{b}_i^T\boldsymbol{Y}_i \quad \text{with} \quad \mathrm{Var}\{\boldsymbol{U}_i\} = \boldsymbol{b}_i^T\boldsymbol{b}_i \tag{3.83}$$

where

$$\boldsymbol{Y}_i = \boldsymbol{C}_i^{-1}\boldsymbol{X}_i \quad \text{and} \quad \boldsymbol{b}_i = \boldsymbol{C}_i^T\boldsymbol{a}_i. \tag{3.84}$$

For the covariance we get

$$\mathrm{Cov}\{\boldsymbol{Y}_i, \boldsymbol{Y}_j\} = \boldsymbol{C}_i^{-1}\boldsymbol{\Sigma}_{ij}\boldsymbol{C}_j^{-T} \tag{3.85}$$

and for the dispersion of $\boldsymbol{Y}$

$$\boldsymbol{\Sigma}_Y = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{C}_1^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{C}_2^{-T} & \cdots & \boldsymbol{C}_1^{-1}\boldsymbol{\Sigma}_{1n}\boldsymbol{C}_n^{-T} \\ \boldsymbol{C}_2^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{C}_1^{-T} & \boldsymbol{I} & \cdots & \boldsymbol{C}_2^{-1}\boldsymbol{\Sigma}_{2n}\boldsymbol{C}_n^{-T} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_n^{-1}\boldsymbol{\Sigma}_{n1}\boldsymbol{C}_1^{-T} & \boldsymbol{C}_n^{-1}\boldsymbol{\Sigma}_{n2}\boldsymbol{C}_2^{-T} & \cdots & \boldsymbol{I} \end{bmatrix}. \tag{3.86}$$

Because of the diagonal block structure of $\boldsymbol{\Sigma}_D$ (Equation 3.50) this can be written in a more compact form. The eigenproblem

$$\boldsymbol{\Sigma}\boldsymbol{a} = \lambda\boldsymbol{\Sigma}_D\boldsymbol{a} \tag{3.87}$$

can be written as

$$\boldsymbol{\Sigma}\boldsymbol{a} = \lambda(\boldsymbol{C}_D\boldsymbol{C}_D^T)\boldsymbol{a} \tag{3.88}$$

where $\boldsymbol{C}_D\boldsymbol{C}_D^T$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_D$. We rewrite

$$\begin{aligned} \boldsymbol{C}_D^{-1}\boldsymbol{\Sigma}\boldsymbol{C}_D^{-T}(\boldsymbol{C}_D^T\boldsymbol{a}) &= \lambda(\boldsymbol{C}_D^T\boldsymbol{a}) \\ \boldsymbol{\Sigma}_Y\boldsymbol{b} &= \lambda\boldsymbol{b} \end{aligned} \tag{3.89}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_Y &= \boldsymbol{C}_D^{-1}\boldsymbol{\Sigma}\boldsymbol{C}_D^{-T} \quad \text{and} \\ \boldsymbol{a} &= \boldsymbol{C}_D^{-T}\boldsymbol{b}. \end{aligned} \tag{3.90}$$

Hence the eigenvalues of $\boldsymbol{\Sigma}_Y$ are equal to the eigenvalues of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Sigma}_D$ and the eigenvectors are related as indicated.

According to Kettenring (1971) results from this method with the above constraint 3 ($\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1 = 0$) can be obtained by a single eigenanalysis of $\boldsymbol{\Sigma}_Y$.

3.2.4 Minimize Smallest Eigenvalue

As in the previous section the original variables are sphered and according to Kettenring (1971) results with the above constraint 3 ($\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 = 0$) can be obtained by a single eigenanalysis of $\boldsymbol{\Sigma}_Y$.

3.2.5 Minimize Determinant

An important result in this connection is that because $\boldsymbol{\Sigma}_U$ is positive definite it can be Cholesky factorized

$$\boldsymbol{\Sigma}_U = \boldsymbol{C}\boldsymbol{C}^T, \tag{3.91}$$

where $\boldsymbol{C}$ is lower triangular. Hence, the eigenvalues of $\boldsymbol{C}$ are the diagonal elements c_{ii} and

$$\det \boldsymbol{\Sigma}_U = (\prod_{i=1}^{n} c_{ii})^2. \tag{3.92}$$

The above result is used in an iterative solution of the problem.

3.3 Case Studies

3.3.1 SPOT HRV Data in Agriculture (MAD)

Two 512×512 SPOT High Resolution Visible (HRV) multispectral (XS) subscenes from 5 February 1987 and 12 February 1989 are used to test the procedure. The selected study area contains economically important coffee and pineapple fields near Thika, Kiambu District, Kenya. The analysis takes place

on raw data (no atmospheric correction). This case is shown in Conradsen & Nielsen (1994) also. Simpson (1994) very successfully uses the MAD transformation on two channels of pre-processed NOAA AVHRR data (noise reduction in channel 3, correction for atmospheric water vapor attenuation, and cloud- and land-masking) in a more physically oriented study with good sea-truth data where principal components analysis of simple difference images fail. Because of its ability to detect change in many channels simultaneously, the MAD transformation is expected to be even more useful when applied to image data with more than three channels. This is supported by limited experience with Landsat TM data, see Section 3.3.3.

This case study is intended as an illustrative example showing how calculations are performed and how an interpretation of the canonical and MAD variates can be carried out. The case study is not meant as a careful assessment of the actual changes that occurred in the study area chosen.

Data and Univariate Change Detection

In Figures 3.2 and 3.3 we show false color composites of the multispectral SPOT HRV scenes acquired on 5 Feb 1987 and 12 Feb 1989, © SPOT Image Copyright 1987 and 1989 CNES. The area is dominated by large pineapple fields to the northeast and small coffee fields to the northwest. To the south is the town of Thika. This is sketched in Figure 3.1 which also shows the positions of fields with pineapple in different phenological stages.

Pineapple is a triennial crop and therefore we observe changes from one year to another. In Figure 3.4 we show the simple change detection image (differences between bands 3, bands 2 and bands 1 in red, green and blue). The major differences are due to the changes primarily in the pineapple fields. Since the changes are connected to change in vegetation, it seems natural to study the change using the normalized difference vegetation index

$$\text{NDVI} = \frac{\text{NIR}-\text{R}}{\text{NIR}+\text{R}+1} \tag{3.93}$$

thika.eps

Figure 3.1: Sketch of areas of interest

S87.eps

Figure 3.2: False color composite of SPOT HRV XS, 5 Feb 1987

S89.eps

Figure 3.3: False color composite of SPOT HRV XS, 12 Feb 1989

Ssimpledif.eps

Figure 3.4: False color composite of simple difference image

Sndvi.eps

Figure 3.5: 1989 NDVI as red and 1987 NDVI as cyan

where NIR is the near-infrared channel (XS3) and R is the red channel (XS2). The philosophy behind the NDVI is that healthy green matter reflects the near-infrared light strongly and absorbs the red light. Therefore the NDVI will be large in vegetated areas and small in non-vegetated areas. An interesting study on NDVI change detection based on NOAA AVHRR decade (10 day) GAC data from Sudan covering a period of nearly 7 years was presented as a video by Stern (1990). In Figure 3.5 we show the 1989 NDVI as red and 1987 NDVI as cyan (causing no change to be represented by a grey scale). This image enhances the differences between fields in a much clearer way than the simple change detection image. This enhancement is not necessarily due to changes from 1987 to 1989 but may also be explained by differences between, say, crops with no seasonal change at all.

Multivariate Change Detection

In Figures 3.6 and 3.3 we show the canonical variates for the 1987 and the 1989 data (CV3, 2 and 1 in red, green and blue). In Figure 3.8 we show all three MADs (MAD1, 2 and 3 in red, green and blue). Areas with very high and very low values in MAD1 are the areas of maximal change, and the sign of MAD1 indicates the "direction" of change. Note that as with any technique based on eigenanalysis of covariance structures the sign of the transformed variables is arbitrary. An inspection of this image and a comparison with the simple change detection image shows that there is a much better distinction between different types of changes. In the simple change detection image red and cyan are dominating but in the MAD image we see that a much better discrimination has been achieved. In Figure 3.9 we show the absolute value of MAD1 with high values shown in red. This image outlines the areas where large changes occurred irrespective of the nature of the change (irrespective of change e.g. from vegetated to bare soil or *vice versa*, and irrespective of dominating wavelength of change).

Below we give an interpretation of the numerical results from the computations of the MADs and a brief discussion. We discuss (1) correlations between original variables, (2) canonical correlations which are measures of similarity between

Scv87.eps

Figure 3.6: Canonical variates of SPOT HRV XS, 5 Feb 1987

Scv89.eps

Figure 3.7: Canonical variates of SPOT HRV XS, 12 Feb 1989

Smad.eps

Figure 3.8: MAD1, 2 and 3 in red, green and blue

Smad1abs.eps

Figure 3.9: Absolute value of MAD1, high values in red

	1987		1989	
	Mean	Std Dev	Mean	Std Dev
XS1	45.00	5.40	32.27	4.79
XS2	36.86	7.12	22.88	4.87
XS3	74.15	12.55	62.33	10.66

Table 3.1: Simple statistics for 1987 and 1989 SPOT HRV XS data

the linear combinations found, (3) correlations between canonical variates and original variables in order to facilitate interpretation of the canonical variates, (4) correlations between MAD variates and original variables in order to facilitate interpretation of the MAD variates, (5) degrees of redundancy between the two sets of canonical variates, i.e. how much variance in either original data set is explained by the canonical variates, and (6) squared multiple correlations between one set of data and the canonical variates of the opposite set of data. Measures (5) and (6) assess other degrees of overlap or redundancy between the two sets of data than the canonical correlations themselves.

Basic Statistics In any interpretation of statistical analysis of multivariate data it is of course important to look at the basic statistics such as means, standard deviations and correlations. The means and standard deviations are shown in Table 3.1.

The values from 1989 are considerably lower than the values from 1987. Whether this is due to calibration problems in the sensors or to actual changes in albedo is not known.

The correlations among the original variables are shown in Table 3.2.

Despite the differences in means and standard deviations it is noted that the correlation structure is remarkably similar in the two years considered. The

		1987			1989		
		XS1	XS2	XS3	XS1	XS2	XS3
	XS1	1.0000	0.9057	–0.3336	0.5116	0.3955	–0.0082
1987	XS2	0.9057	1.0000	–0.4196	0.4352	0.4140	–0.0381
	XS3	–0.3336	–0.4196	1.0000	–0.3477	–0.2644	0.2492
	XS1	0.5116	0.4352	–0.3477	1.0000	0.8866	–0.2609
1989	XS2	0.3955	0.4140	–0.2644	0.8866	1.0000	–0.4191
	XS3	–0.0082	–0.0381	0.2492	–0.2609	–0.4191	1.0000

Table 3.2: Correlations among original variables

	Canonical Correlation (ρ)	Squared Canonical Correlation (ρ^2)
1	0.6505	0.4232
2	0.4024	0.1619
3	0.2403	0.0577

Table 3.3: Canonical correlations

crosscorrelations between years are less similar and decreasing with increasing wavelength, in this case indicating that changes in vegetation are the most important ones.

Canonical Correlation Analysis The magnitude of the canonical correlation coefficients shown in Table 3.3 can be used in assessing the degree of change in the bi-temporal imagery.

We see from the canonical correlations that only 6% of the variation in canonical variate 3 from one year may be explained by the variation in the other canonical variate 3. This indicates a considerable degree of change. For canonical variates

	1987			1989		
	CV1	CV2	CV3	CV1	CV2	CV3
XS1	0.3487	–0.1272	0.2370	0.4269	–0.1702	0.0887
XS2	–0.2154	0.2374	–0.1323	–0.3103	0.3669	–0.0909
XS3	–0.0473	0.0325	0.0672	–0.0245	0.0603	0.0850

Table 3.4: Raw canonical coefficients

2 the number is 16%, still a rather small number. Finally, canonical variates 1 show a common predictability of 42%.

The raw canonical coefficients are shown in Table 3.4. Thus the canonical variates for the 1987 XS data are

$$\begin{bmatrix} \mathrm{CV1} \\ \mathrm{CV2} \\ \mathrm{CV3} \end{bmatrix} = \begin{bmatrix} 0.3487 & -0.2154 & -0.0473 \\ -0.1272 & 0.2374 & 0.0325 \\ 0.2370 & -0.1323 & 0.0672 \end{bmatrix} \begin{bmatrix} \mathrm{XS1} - 45.00 \\ \mathrm{XS2} - 36.86 \\ \mathrm{XS3} - 74.15 \end{bmatrix}$$

and the canonical variates for the 1989 XS data are

$$\begin{bmatrix} \mathrm{CV1} \\ \mathrm{CV2} \\ \mathrm{CV3} \end{bmatrix} = \begin{bmatrix} 0.4269 & -0.3103 & -0.0245 \\ -0.1702 & 0.3669 & 0.0603 \\ 0.0887 & -0.0909 & 0.0850 \end{bmatrix} \begin{bmatrix} \mathrm{XS1} - 32.27 \\ \mathrm{XS2} - 22.88 \\ \mathrm{XS3} - 62.33 \end{bmatrix}.$$

The coefficients for computing the canonical variates are hard to interpret directly. The correlations between the original variables and the canonical variates are better for interpretation, cf. below.

Canonical Structure The correlations between the original variables and the canonical variables may be used in the interpretation of the canonical variables.

		1987			1989		
		CV1	CV2	CV3	CV1	CV2	CV3
	XS1	0.6915	0.7078	0.1442	0.4499	0.2848	0.0347
1987	XS2	0.4206	0.8967	–0.1377	0.2736	0.3609	–0.0331
	XS3	–0.5784	–0.0719	0.8126	–0.3763	–0.0289	0.1952
	XS1	0.5021	0.2423	–0.0491	0.7718	0.6021	–0.2045
1989	XS2	0.2667	0.3201	–0.1072	0.4099	0.7955	–0.4462
	XS3	–0.1050	0.0429	0.2357	–0.1613	0.1067	0.9811

Table 3.5: Correlations between original variables and canonical variables

The correlations between original variables and canonical variables are shown in Table 3.5.

In both years we see that canonical variate 2 is strongly correlated with the visible channels, i.e. MAD2 measures changes in the visible part of the spectrum. In both years canonical variate 3 is positively correlated with the near-infrared channel and negatively correlated with or at least almost not correlated with the visible channels. This conforms with a vegetation index. Therefore, in this case MAD1 measures vegetation changes. A similar pattern but with less emphasis on the near-infrared channel is seen for canonical variates 1 and MAD3.

MAD Structure In order to interpret the MADs we give the correlations between the original variables and the MADs. These values will not be supplied by a canned canonical correlations computer program. The values are computed by means of the expressions given in Equations 3.41 and 3.42. It is easier—and more CPU time consuming—to use an ordinary correlation program on the estimated MAD image. The correlations between original variables and MADs are shown in Table 3.6.

The most dominant correlations are MAD1 with 1987 XS3 (–0.50) and with 1989 XS3 (0.60). Pixels showing extreme values of MAD1 will predominantly

	MAD1	MAD2	MAD3
XS1	–0.0889	–0.3868	–0.2890
1987 XS2	0.0849	–0.4901	–0.1757
XS3	–0.5008	0.0393	0.2418
XS1	–0.1260	0.3292	0.3227
1989 XS2	–0.2750	0.4349	0.1714
XS3	0.6047	0.0583	–0.0674

Table 3.6: Correlations between original variables and MADs

have high values of 1987 XS3 and low values of 1989 XS3 or *vice versa*. Thus MAD1 basically describes changes in XS3, the photo-infrared channel, which again is strongly related to vegetation. Areas showing extreme values in MAD1 will then most likely have very different vegetation cover in 1987 and 1989. Changes orthogonal to (i.e. uncorrelated with) these changes are described by MAD2 and MAD3. Similar considerations on magnitudes of correlations show that MAD2 and MAD3 describe changes in the shorter wavelengths, MAD2 with the emphasis on XS2 and MAD3 with the emphasis on XS1.

At this point it should be emphasized again that the analysis presented is scene dependent. In other scenes the interpretations of the MADs will very likely be different. Where a technique as the NDVI change detection "looks for" changes in vegetation cover the present method detects general alterations in the scene no matter the source of the alteration. Once established the MADs may be interpreted by means of the correlations between the original and the transformed variables as presented above.

To illustrate the concept further we shall examine the pineapple fields north of Thika somewhat closer. In Figure 3.1 some pineapple fields are outlined along with the center of Thika. In Table 3.7 we have indicated the relative level of the three MAD variables mapped as red, green and blue in Figure 3.8.

First we consider area 3, bare soil in 1987 and healthy pineapple in 1989. This is an area that shows extreme deviation between the two scenes. The area is

Channel Area	Red MAD1	Green MAD2	Blue MAD3	MAD
1	High	High	High	Light Gray
2	Low	High	High	Cyan
3	High	Low	Low	Red
Town	Low	High	Low	Green

Table 3.7: Levels of MADs in three pineapple areas and in the town

strongly outlined in all change schemes used, a.o. simple difference change detection, NDVI change detection, decorrelated simple difference change detection (not shown), principle components and rotated factors of simple difference images (not shown) and MAD. Area 2 shows the opposite pattern, pineapple in 1987 and bare soil in 1989. The values related with these patterns are consistent with the general interpretation of the MADs given before. Area 1 is covered with pineapple in different phenological stages in 1987 and 1989. The alterations are strongly related to vegetation change and are therefore clearly visible in the NDVI change image. In the NDVI change image we see a totally black area in the center of Thika. This is very consistent with the notion of a vegetation index. In the same area the MAD change image reveals a considerable alteration. No information is available to us on the probable causes for these changes and we shall not speculate on their nature. Whatever the causes, the differences described illustrate the fact that the MADs may be used in general detection of alterations irrespective of the nature of the alterations.

As a concluding remark we therefore suggest the usage of the MAD transformation in the analysis of multispectral, bi-temporal imagery. The MADs give an optimal (in the sense of maximal variance) detection of alterations from one scene to the other, and also it provides a statistical analysis and interpretation of the nature of the alterations.

1987 Canonical Variables			1989 Canonical Variables		
		Cumulative			Cumulative
	Proportion	Proportion	ρ^2	Proportion	Proportion
CV1	0.3299	0.3299	0.4232	0.1396	0.1396
CV2	0.4368	0.7667	0.1619	0.0707	0.2103
CV3	0.2333	1.0000	0.0577	0.0135	0.2238

Table 3.8: Variance of 1987 XS explained by the individual canonical variates for 1987 and 1989

1989 Canonical Variables			1987 Canonical Variables		
		Cumulative			Cumulative
	Proportion	Proportion	ρ^2	Proportion	Proportion
CV1	0.2632	0.2632	0.4232	0.1114	0.1114
CV2	0.3356	0.5988	0.1619	0.0543	0.1658
CV3	0.4012	1.0000	0.0577	0.0232	0.1889

Table 3.9: Variance of 1989 XS explained by the individual canonical variates for 1989 and 1987

Canonical Redundancy Analysis A more detailed assessment of the degree of change may be obtained from a deeper study of the correlations between the variates involved. The standardized variance of 1987 XS explained by the individual canonical variates for 1987 and 1989 are shown in Table 3.8.

The standardized variance of 1989 XS explained by the individual canonical variates for 1989 and 1987 are shown in Table 3.9.

The squared multiple correlations (R^2) between 1987 XS and the first M canonical variates of 1989 XS, and squared multiple correlations (R^2) between 1989 XS and the first M canonical variates of 1987 XS are shown in Table 3.10.

	R^2(1987 XS, 1989 CAN)			R^2(1989 XS, 1987 CAN)		
M	1	2	3	1	2	3
XS1	0.2024	0.2835	0.2847	0.2521	0.3108	0.3132
XS2	0.0749	0.2051	0.2062	0.0711	0.1736	0.1851
XS3	0.1416	0.1424	0.1805	0.0110	0.0129	0.0684

Table 3.10: Squared multiple correlations (R^2) between 1987 (1989) XS and the first M canonical variates of 1989 (1987) XS

The canonical redundancy analysis confirms that we have considerable changes between the two years. The degrees of explanation of one set of original variables by the opposite canonical variates range from 1% to 14%, very low numbers. Similarly, we see from the squared multiple correlations between the original 1987 variables and the first M 1989 canonical variates and the squared multiple correlations between the original 1989 variables and the first M 1987 canonical variates that the degree of explanation is poorest in the near-infrared band, again indicating that vegetation changes are dominating.

Geometric Illustration of Canonical Variates

To hopefully give a better feel for what canonical variates are and to illustrate geometrically the solution to the real, symmetric, generalized (RSG) eigenproblem involved in finding them, two bivariate sets of data were generated. The data consist of every 50'th row and every 50'th column of the image data analyzed above. The first set of variables are bands 1 and 2 from the 1987 data and the second set of variables are bands 2 and 3 from the 1989 data. The 1987 (1989) data are estimated from the 1989 (1987) data by regression.

The two top plots in Figure 3.10 show scatterplots and ellipses corresponding to $\chi^2_{0.95}(2) = 5.991$ contours for the 1987 and 1989 data. These contour ellipses are (top-left) $a^T\hat{\Sigma}_{11}^{-1}a = 5.991$ (for the data) and $a^T(\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21})^{-1}a = 5.991$ (for the regressions), and (top-right) $b^T\hat{\Sigma}_{22}^{-1}b = 5.991$ (for the data) and ellipses.eps

Figure 3.10: Canonical variates geometrically

$b^T(\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12})^{-1}b = 5.991$ (for the regressions). The open circles symbolize observations and the crosses symbolize regressions made from the opposite set of variables.

The two bottom plots show the solution to the eigenproblem. The ellipses shown are contours for (bottom-left) $a^T\hat\Sigma_{11}a = 1$, $a^T\hat\Sigma_{12}\hat\Sigma_{22}^{-1}\hat\Sigma_{21}a = 1$, $a^T\hat\Sigma_{12}\hat\Sigma_{22}^{-1}\hat\Sigma_{21}a = \rho_1^2$ and $a^T\hat\Sigma_{12}\hat\Sigma_{22}^{-1}\hat\Sigma_{21}a = \rho_2^2$, and (bottom-right) $b^T\hat\Sigma_{22}b = 1$, $b^T\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12}b = 1$, $b^T\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12}b = \rho_1^2$ and $b^T\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12}b = \rho_2^2$. In the bottom-right plot the contour lines are identified; here N means the matrix in the numerator of the Rayleigh coefficient identifying the canonical correlation problem and D means the matrix in the denominator. In this case $\rho_1^2 = 0.2730$ and $\rho_2^2 = 0.05147$ corresponding to canonical correlations 0.5199 and 0.2269.

In the two bottom plots the eigenvectors to the canonical correlation problem are vectors with end points in the center of the ellipses and the points where the ellipses have a common tangent (indicated with short lines). The square root of the eigenvalues (the canonical correlations) ρ_i are the ratios of the the lengths of the major (or minor) axes in the ellipses corresponding to $a^T\hat\Sigma_{12}\hat\Sigma_{22}^{-1}\hat\Sigma_{21}a = 1$ and $a^T\hat\Sigma_{12}\hat\Sigma_{22}^{-1}\hat\Sigma_{21}a = \rho_i^2$ (bottom-left). The same is true for $b^T\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12}b = 1$ and $b^T\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12}b = \rho_i^2$ (bottom-right). The major axes of $a^T\hat\Sigma_{12}\hat\Sigma_{22}^{-1}\hat\Sigma_{21}a$ and $b^T\hat\Sigma_{21}\hat\Sigma_{11}^{-1}\hat\Sigma_{12}b$ (the matrices in the numerators of the Rayleigh coefficient) are indicated with long lines.

3.3.2 Landsat TM Data in Forestry (MUSECC)

The applicability of multiset canonical correlations analysis to multivariate and truly multitemporal change detection studies is demonstrated in a case study using Landsat-5 Thematic Mapper (TM) data covering a small forested area approximately 20 kilometers north of Umeå in northern Sweden (data from the Swedish Space Corporation). The data consist of six times six spectral bands with 512×512 20 meter pixels rectified to the Swedish national grid from the summers 1984–89. The acquisition dates are 1 August 1984, 26 June 1985, 6 June 1986, 12 August 1987, 27 June 1988 and 21 June 1989. Results from such analyses are linear combinations that transform the original bands into new variables that show decreasing similarity over six points in time. The minimum similarity variables are measures of change in all bands simultaneously. This analysis of correlations between variables where observations are considered as repetitions is termed R-mode analysis. In this case, in R-mode analysis we consider Landsat TM bands 1, 2, 3, 4, 5 and 7 for each of the years 1984–1989 as one set of variables. In Q-mode analysis of correlations between observations where variables are considered as repetitions we consider TM bands 1 for all years 1984–1989 as one set of variables, TM bands 2 for all years 1984–1989 as another set of variables, etc. For a sketch of R- and Q-mode analysis set-up see Figures 3.11 and 3.12. In both figures the sets of variables indicated on the top are transformed into new variables on the bottom.

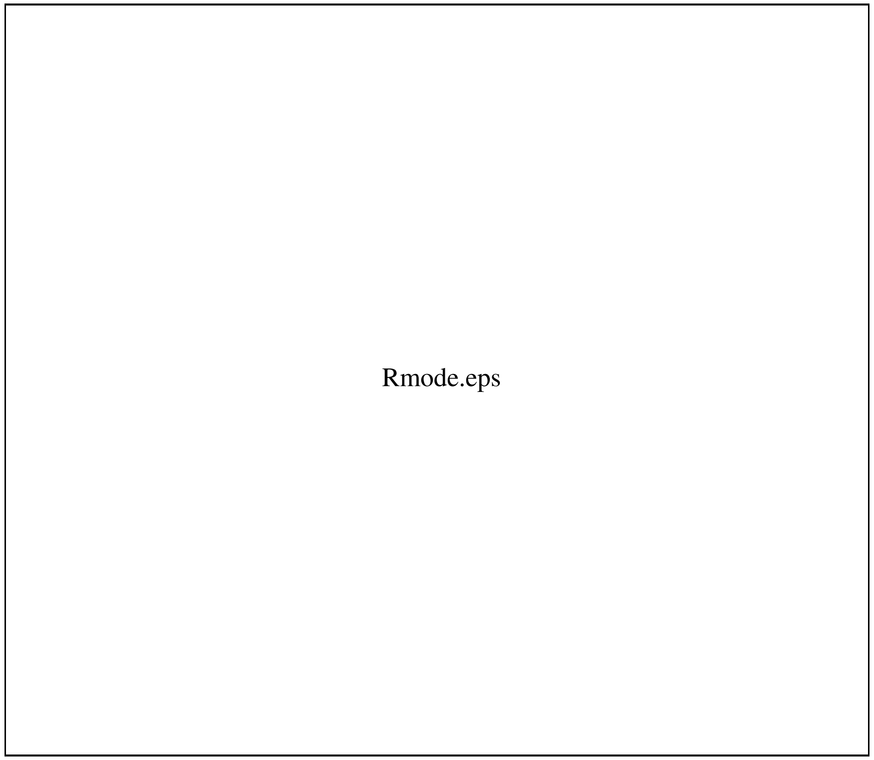

Figure 3.11: Sketch of R-mode multiset canonical correlations analysis

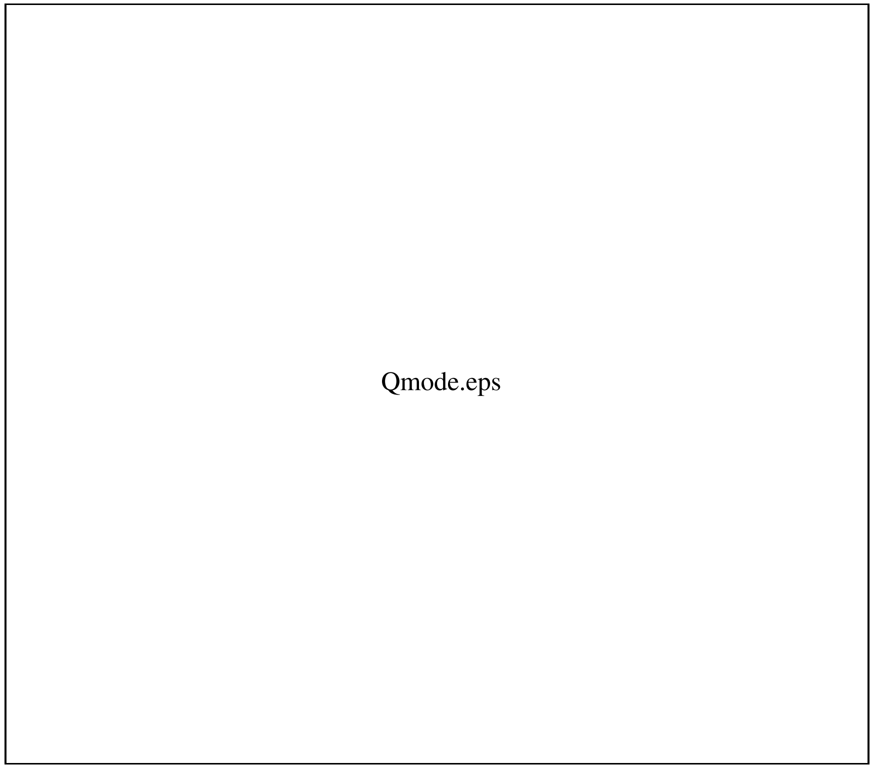

Figure 3.12: Sketch of Q-mode multiset canonical correlations analysis

cvfigs.eps

Figure 3.13: Order of variables in following images, left: False color and R-mode, right: Q-mode

This case study is described in lesser detail than the above MAD analysis of SPOT HRV XS data. As with the above MAD analysis case, this case is intended as an illustrative example showing how calculations are performed and how an interpretation of the canonical variates can be carried out. The case study is not meant as a careful assessment of the actual changes that occurred in the study area chosen.

Figures 3.14 to 3.24 are to be viewed with the paper in landscape mode. The order of the variables is shown in Figure 3.13.

Figure 3.14 shows Landsat TM channels 4, 5 and 3 as red, green and blue respectively.

Figure 3.15 shows R-mode canonical variates 1, 2 and 3 as red, green and blue respectively. We see that we have indeed obtained a high degree of similarity over years. Figure 3.16 shows R-mode canonical variates 6, 5 and 4 as red, green and blue respectively. This is the RGB combination that shows minimum similarity over years. We see that noise (striping and drop-outs) is depicted well

umeaa.eps

Figure 3.14: Landsat TM channels 4, 5 and 3 as red, green and blue

as is to be expected: if data from one year is noisy and data from another year is not then certainly the largest difference could be that noise. As for the MAD transformation (see Section 3.1.1) this observation inspires an iterative use of the procedure: first identify noise, restore data or exclude areas with noise from further analysis, and carry out the analysis once more.

The minimum similarity variables are measures of change in all bands simultaneously. To find areas of minimum similarity with high autocorrelation we use the absolute value of minimum/maximum autocorrelation factors (Switzer & Green (1984) and Section 2.2) of the highest order canonical variates. Figures 3.17, 3.18, 3.19 and 3.20 show R-mode canonical variates 6, their absolute values, their MAFs and absolute values of their MAFs. Figures 3.21, 3.22, 3.23 and 3.24 show Q-mode canonical variates 6, their absolute values, their MAFs and absolute values of their MAFs. In the Q-mode case MAF analysis concentrates the information in two components. Q-mode analysis also reveals that striping and drop-outs occur basically in bands 1, 2 and 3. Another good impression of overall change that includes lower order CVs also, is achieved by inspecting (absolute values of) MAFs of Q-mode canonical variates (CVs) 5 and 6 (not shown).

Correlations between R-mode CVs 6 and the original data given in Figure 3.26 show that changes over years are associated with TM bands 1 especially from 1984 to 1987. This is probably because of differences in atmospheric conditions. Therefore analysis of atmospherically corrected data would be interesting. Correlations between Q-mode CVs 6 and the original variables given in Figure 3.28, for TM bands 1, 2, 3, 5 and 7 reveal a pattern of positive correlation with 1984, negative correlation with 1985, and again positive correlation with 1986 (but not as high as with 1984) combined with (nearly) no correlation with 1987, 1988 and 1989. Q-mode CV6 for TM4 is positively correlated with TM4 in 1984, 1985 and 1986, uncorrelated with TM4 in 1987, and negatively correlated with TM4 in 1988 and 1989. This could indicate that vegetation related changes occurred from 1986 to 1988. Correlations between Q-mode CVs 1 and TM4 given in Figure 3.27 are (except for TM4 CV1) lower than correlations between Q-mode CVs 1 and the other bands. Again, this indicates changes that are

cv123rgb.eps

Figure 3.15: R-mode canonical variates 1, 2 and 3 as red, green and blue

cv654rgb.eps

Figure 3.16: R-mode canonical variates 6, 5 and 4 as red, green and blue

cv6r.eps

Figure 3.17: R-mode canonical variates 6

abs-rcv6.eps

Figure 3.18: Absolute values of R-mode canonical variates 6

mafcv6r.eps

Figure 3.19: MAFs of R-mode canonical variates 6

absmafcv6r.eps

Figure 3.20: Absolute values of MAFs of R-mode canonical variates 6

cv6q.eps

Figure 3.21: Q-mode canonical variates 6

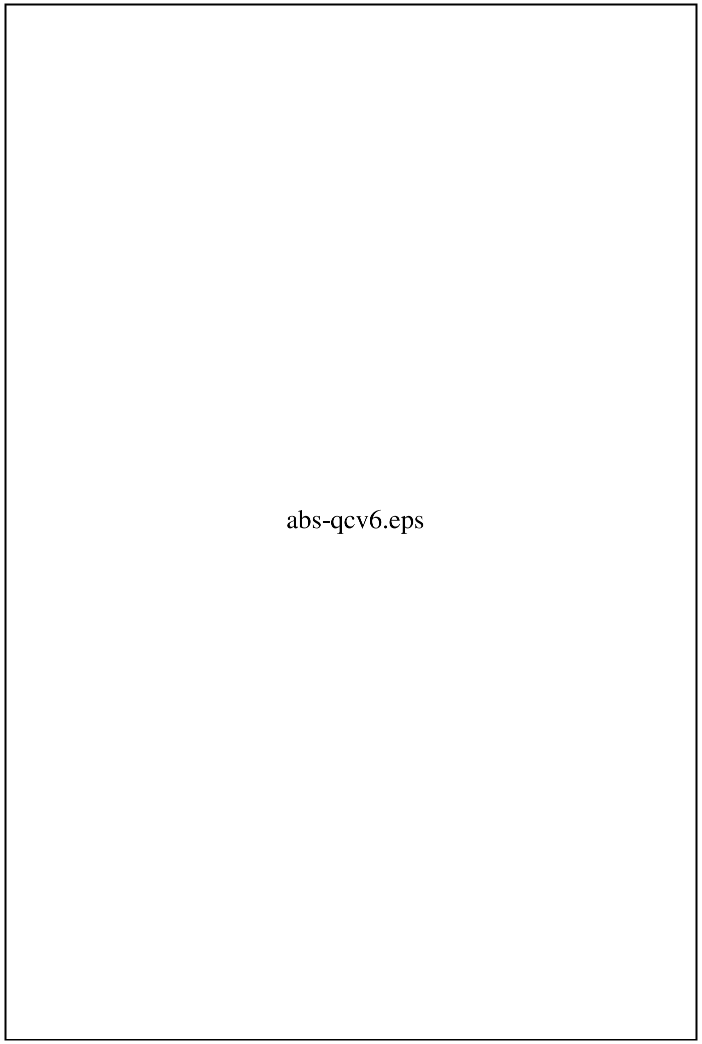

Figure 3.22: Absolute values of Q-mode canonical variates 6

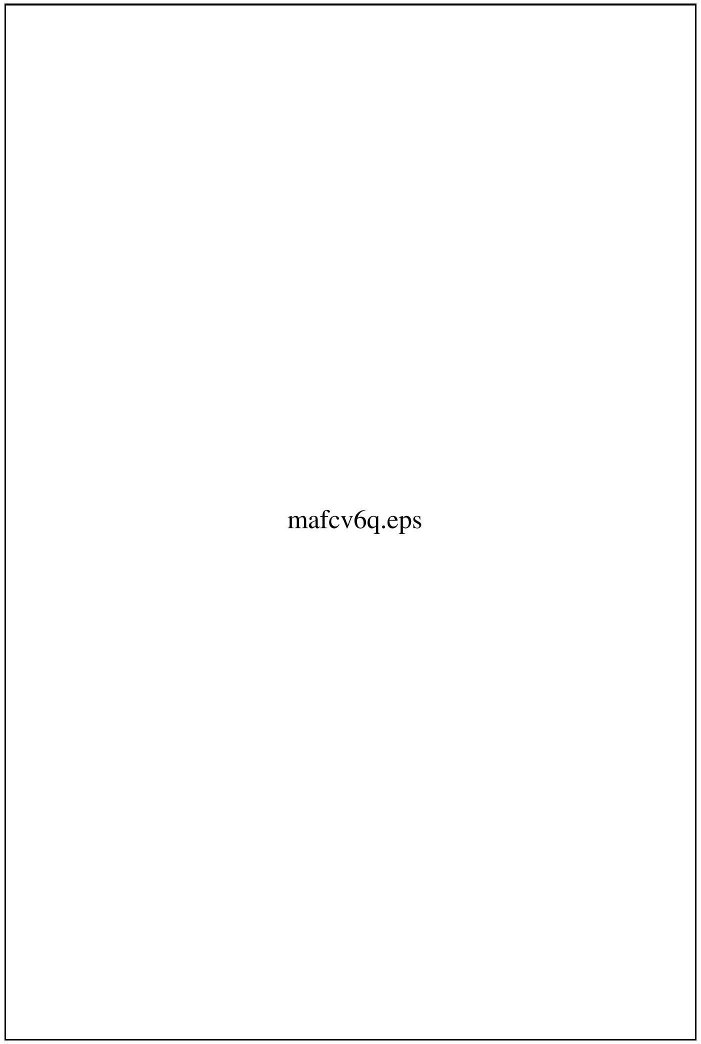

Figure 3.23: MAFs of Q-mode canonical variates 6

absmafcv6q.eps

Figure 3.24: Absolute values of MAFs of Q-mode canonical variates 6

related with TM4, possibly vegetation changes. For completeness Figure 3.25 gives correlations between R-mode CVs 1 and the original data.

For reasons given in the above paragraph Section 3.3.3 contains a brief report on a MAD analysis of the bi-temporal data from 1986 and 1988.

Figure 3.29 shows the sum of the absolute values of MAFs 1 and 2 of the Q-mode canonical variates 6. The dark areas in this one image are areas of maximum change in all years and all bands regardless of what caused the change and regardless of the "direction" of change.

The following comparisons between R- and Q-mode canonical variates of the above data all refer to constraint and orthogonality criterion $\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = 1$, i.e. each canonical variate has unit variance (constraint 3 above). Table 3.11 shows correlations between R-mode canonical variates 1 ($\boldsymbol{\Sigma}_U$) for all methods investigated. The same correlations for Q-mode analysis is shown in Table 3.12. Again, we see a special behavior for TM4 indicating vegetation changes.

In these comparisons, Sumcor, Ssqcor and Maxvar seem to perform much in the same fashion. Minvar and Genvar seem to perform differently and not in the same fashion. Gnanadesikan (1977) observes a similar different behavior for Minvar. This is understandable when contemplating the design criteria behind the individual methods. Sumcor and Ssqcor both focus on all correlations between CVs. Maxvar maximizes the largest eigenvalue, again a focus on all elements in $\boldsymbol{\Sigma}_U$. Minvar relies heavily on the smallest eigenvalue, whereas Genvar minimizes the determinant of $\boldsymbol{\Sigma}_U$ and therefore relies on several small eigenvalues. Due to lack of ground truth data it has not been possible to determine empirically which of the five methods (if any) perform best in this context.

Tables 3.13 and 3.14 show comparisons of the actual values of the optimization criteria for the five methods discussed for R- and Q-mode canonical variates 1. The optimization criteria are not contradicted, e.g. for Minvar λ_{min} is smaller than for the other methods. Also in this comparison, Sumcor, Ssqcor and Maxvar seem to perform much in the same fashion, and Minvar and Genvar seem to perform differently and not in the same fashion.

corrRCV1ori.eps

Figure 3.25: Correlations between R-mode CVs 1 and original data

corrRCV6ori.eps

Figure 3.26: Correlations between R-mode CVs 6 and original data

corrQCV1ori.eps

Figure 3.27: Correlations between Q-mode CVs 1 and original data

corrQCV6ori.eps

Figure 3.28: Correlations between Q-mode CVs 6 and original data

Sumcor					
1.0000	0.9114	0.8534	0.8768	0.8893	0.8852
0.9114	1.0000	0.9327	0.9248	0.8939	0.9037
0.8534	0.9327	1.0000	0.9202	0.8861	0.9048
0.8768	0.9248	0.9202	1.0000	0.8868	0.9127
0.8893	0.8939	0.8861	0.8868	1.0000	0.9544
0.8852	0.9037	0.9048	0.9127	0.9544	1.0000
Ssqcor					
1.0000	0.9114	0.8532	0.8765	0.8893	0.8851
0.9114	1.0000	0.9330	0.9250	0.8939	0.9038
0.8532	0.9330	1.0000	0.9204	0.8860	0.9049
0.8765	0.9250	0.9204	1.0000	0.8867	0.9126
0.8893	0.8939	0.8860	0.8867	1.0000	0.9545
0.8851	0.9038	0.9049	0.9126	0.9545	1.0000
Maxvar					
1.0000	0.9114	0.8534	0.8767	0.8893	0.8852
0.9114	1.0000	0.9328	0.9249	0.8939	0.9037
0.8534	0.9328	1.0000	0.9202	0.8860	0.9048
0.8767	0.9249	0.9202	1.0000	0.8867	0.9127
0.8893	0.8939	0.8860	0.8867	1.0000	0.9544
0.8852	0.9037	0.9048	0.9127	0.9544	1.0000
Minvar					
1.0000	0.8334	0.7259	0.7709	0.7651	0.7797
0.8334	1.0000	0.9195	0.9027	0.8472	0.8709
0.7259	0.9195	1.0000	0.8595	0.7692	0.8246
0.7709	0.9027	0.8595	1.0000	0.8667	0.9023
0.7651	0.8472	0.7692	0.8667	1.0000	0.9564
0.7797	0.8709	0.8246	0.9023	0.9564	1.0000
Genvar					
1.0000	0.9067	0.8390	0.8645	0.8896	0.8792
0.9067	1.0000	0.9412	0.9276	0.8903	0.9022
0.8390	0.9412	1.0000	0.9241	0.8772	0.9031
0.8645	0.9276	0.9241	1.0000	0.8795	0.9076
0.8896	0.8903	0.8772	0.8795	1.0000	0.9577
0.8792	0.9022	0.9031	0.9076	0.9577	1.0000

Table 3.11: Correlations between R-mode canonical variates 1 for all five methods

Sumcor					
1.0000	0.9420	0.9548	0.6414	0.8919	0.9275
0.9420	1.0000	0.9531	0.7571	0.9059	0.9021
0.9548	0.9531	1.0000	0.6989	0.9038	0.9219
0.6414	0.7571	0.6989	1.0000	0.7366	0.6392
0.8919	0.9059	0.9038	0.7366	1.0000	0.9673
0.9275	0.9021	0.9219	0.6392	0.9673	1.0000
Ssqcor					
1.0000	0.9442	0.9574	0.6385	0.8935	0.9301
0.9442	1.0000	0.9547	0.7547	0.9064	0.9038
0.9574	0.9547	1.0000	0.6942	0.9049	0.9241
0.6385	0.7547	0.6942	1.0000	0.7326	0.6337
0.8935	0.9064	0.9049	0.7326	1.0000	0.9678
0.9301	0.9038	0.9241	0.6337	0.9678	1.0000
Maxvar					
1.0000	0.9437	0.9566	0.6396	0.8931	0.9293
0.9437	1.0000	0.9543	0.7549	0.9064	0.9034
0.9566	0.9543	1.0000	0.6958	0.9047	0.9235
0.6396	0.7549	0.6958	1.0000	0.7333	0.6358
0.8931	0.9064	0.9047	0.7333	1.0000	0.9677
0.9293	0.9034	0.9235	0.6358	0.9677	1.0000
Minvar					
1.0000	0.9451	0.8939	0.4978	0.8767	0.9316
0.9451	1.0000	0.8535	0.6725	0.8988	0.9055
0.8939	0.8535	1.0000	0.3950	0.7683	0.8438
0.4978	0.6725	0.3950	1.0000	0.6609	0.5046
0.8767	0.8988	0.7683	0.6609	1.0000	0.9666
0.9316	0.9055	0.8438	0.5046	0.9666	1.0000
Genvar					
1.0000	0.9488	0.9666	0.5350	0.8903	0.9369
0.9488	1.0000	0.9566	0.6985	0.9036	0.9058
0.9666	0.9566	1.0000	0.5687	0.8953	0.9302
0.5350	0.6985	0.5687	1.0000	0.6778	0.5261
0.8903	0.9036	0.8953	0.6778	1.0000	0.9676
0.9369	0.9058	0.9302	0.5261	0.9676	1.0000

Table 3.12: Correlations between Q-mode canonical variates 1 for all five methods

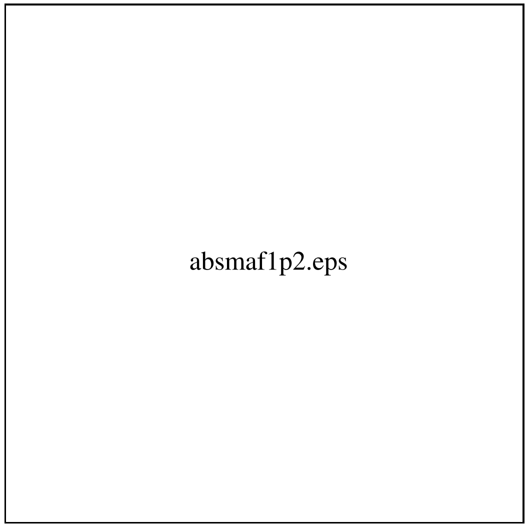

Figure 3.29: Sum of absolute value of MAFs 1 and 2 of Q-mode CVs 6

	$\sum\sum \Sigma_{Uij}$	$\sum\sum (\Sigma_{Uij})^2$	λ_{max}	λ_{min}	det $\mathbf{\Sigma}_U$
Sumcor	33.0725	30.4481	5.5125	0.0415	$2.3488\ 10^{-5}$
Ssqcor	33.0724	30.4482	5.5125	0.0414	$2.3347\ 10^{-5}$
Maxvar	33.0725	30.4482	5.5125	0.0415	$2.3448\ 10^{-5}$
Minvar	31.1882	27.2723	5.2038	0.0336	$1.0191\ 10^{-4}$
Genvar	32.9787	30.2877	5.4971	0.0371	$2.0069\ 10^{-5}$

Table 3.13: Optimization criteria for all five methods, R-mode

	$\sum\sum \Sigma_{Uij}$	$\sum\sum (\Sigma_{Uij})^2$	λ_{max}	λ_{min}	det $\mathbf{\Sigma}_U$
Sumcor	31.4870	28.0484	5.2730	0.0177	$1.2683\ 10^{-5}$
Ssqcor	31.4812	28.0566	5.2732	0.0167	$1.0872\ 10^{-5}$
Maxvar	31.4842	28.0562	5.2734	0.0171	$1.1480\ 10^{-5}$
Minvar	29.2292	24.9382	4.9373	0.0073	$1.6008\ 10^{-5}$
Genvar	30.6156	26.9877	5.1558	0.0078	$3.4272\ 10^{-6}$

Table 3.14: Optimization criteria for all five methods, Q-mode

3.3.3 Landsat TM Data in Forestry (MAD revisited)

The correlation structures shown in Figure 3.28 and described in page 125 indicate that vegetation related changes occurred from 1986 to 1988. Therefore this section gives a brief report of a MAD analysis of the Landsat TM data from 6 June 1986 and 27 June 1988. Figure 3.30 shows all six MADs (view with paper in landscape mode, row-wise from top-left is MAD1, MAD2, etc.). Again, we see that noise is a major difference between the two points in time but also the areas we have seen in the above multiset analysis stand out clearly in this analysis. This is also evident in the absolute values of the MADs, Figure 3.31.

Correlations between the MADs and the original variables given in Table 3.15 are generally quite low. However, the pattern revealed shows that MAD1 is associated with TM1, i.e. probably differences in atmospheric conditions. MAD4 is positively correlated with 1986 TM4 and negatively correlated with 1986 TM1, 2 and 3. The opposite correlation structure is true for MAD4 and the 1988 data. Therefore MAD4 is a sort of vegetation index change detector. With reverse signs for the correlations this is true for MAD5 also. MAD6 is a change detector of the (weighted) overall level.

A MAF transformation of the MADs is shown in Figure 3.32 and the correlations between the MADs and their MAFs are shown in Table 3.16. We see that low order MAFs (signal) are associated with high order MADs, i.e. maximum mad.eps

Figure 3.30: MADs of TM bands from 1986 and 1988

abs-mad.eps

Figure 3.31: Absolute values of MADs of TM bands from 1986 and 1988

	MAD1	MAD2	MAD3	MAD4	MAD5	MAD6
Can.corr.	0.0464	0.0978	0.3199	0.5957	0.7255	0.8966
1986 TM1	0.2583	–0.0010	–0.0789	–0.0759	0.1078	0.1941
1986 TM2	–0.0389	0.1994	–0.1150	–0.0512	0.0313	0.2096
1986 TM3	–0.0280	0.0202	0.0634	–0.0896	0.1008	0.2123
1986 TM4	0.0061	0.0781	–0.0103	0.1105	–0.2109	0.1764
1986 TM5	–0.0167	0.0355	–0.0431	0.1380	0.1089	0.2047
1986 TM7	–0.0592	–0.0676	–0.1041	0.0638	0.1493	0.1993
1988 TM1	–0.1453	0.0239	0.0778	0.1197	–0.1621	–0.1867
1988 TM2	0.0986	–0.0480	0.0399	0.1069	–0.0978	–0.2089
1988 TM3	0.0068	0.0922	–0.0676	0.1583	–0.1567	–0.1854
1988 TM4	–0.0089	–0.0218	–0.0211	–0.1216	0.1930	–0.1838
1988 TM5	0.0122	0.0567	–0.0765	–0.0662	–0.1588	–0.1995
1988 TM7	0.0390	0.1630	0.0069	0.0072	–0.2069	–0.1798

Table 3.15: Correlations between MADs and original variables

	MAF1	MAF2	MAF3	MAF4	MAF5	MAF6
Autocorr.	0.9130	0.8769	0.6822	0.5402	0.4770	0.4696
MAD1	0.0403	–0.0784	–0.1402	–0.1820	0.2033	0.9477
MAD2	0.2501	–0.1095	–0.2162	0.7958	0.4954	–0.0051
MAD3	0.0331	0.0074	–0.0456	0.4942	–0.8262	0.2646
MAD4	0.4769	0.1545	0.8523	0.0701	0.0540	0.1205
MAD5	–0.7485	0.4915	0.2780	0.2759	0.1661	0.1309
MAD6	0.3834	0.8464	–0.3576	–0.0913	–0.0162	–0.0133

Table 3.16: Correlations between MAFs of MADs and MADs

similarity CVs and *vice versa*. Absolute values of the MAFs are shown in Figure 3.33.

Table 3.17 shows correlations between MAFs of MADs and the original TM bands. MAF1 is positively correlated with 1986 TM4 and negatively correlated with 1986 TM1 and 3. The opposite correlation structure is true for MAF1 and the 1988 data. Therefore MAF1 is a sort of vegetation index change detector and it concentrates the information from MADs 4 and 5. MAF2 is a change detector of the weighted mean of all bands except TM4, i.e. a change detector of the non-vegetation related level. MAF6 measures change in TM1 which is likely to represent changes in atmospheric conditions. We see that the MAF analysis of the MADs has isolated the changes related to TM1 (presumably atmospheric conditions) and changes related to TM4 (presumably vegetation) in each end of the autocorrelation "spectrum".

Figure 3.34 shows the sum of the absolute values of MAFs 1 and 2 of MADs. This one graytone image shows the location and strength of most of the changes in the area. Figure 3.35 shows MAFs 1, 2 and 3 of MADs as red, green and blue. Figure 3.36 shows the absolute values of MAFs 1, 2 and 3 of MADs as red, green and blue. This image beautifully depicts locations and strength of change. Different types of change are indicated by different colors. The noise related changes are isolated in the higher order MAFs of MADs.

	MAF1	MAF2	MAF3	MAF4	MAF5	MAF6
Autocorr.	0.9130	0.8769	0.6822	0.5402	0.4770	0.4696
1986 TM1	–0.0349	0.1848	–0.1365	–0.0801	0.1278	0.2263
1986 TM2	0.0771	0.1653	–0.1423	0.0949	0.1849	–0.0731
1986 TM3	–0.0307	0.2159	–0.1276	0.0547	–0.0396	–0.0103
1986 TM4	0.2976	0.0536	–0.0448	–0.0105	0.0166	–0.0140
1986 TM5	0.0696	0.2452	0.0713	0.0310	0.0720	0.0007
1986 TM7	–0.0277	0.2632	0.0522	–0.0670	0.0655	–0.0587
1988 TM1	0.1095	–0.2099	0.1354	0.0646	–0.0994	–0.1215
1988 TM2	0.0374	–0.2105	0.1334	–0.0368	–0.0438	0.1071
1988 TM3	0.1428	–0.2206	0.1398	0.0236	0.0884	–0.0109
1988 TM4	–0.2794	–0.0765	0.0227	0.0353	0.0333	–0.0009
1988 TM5	0.0229	–0.2649	–0.0397	–0.0251	0.0671	–0.0351
1988 TM7	0.1319	–0.2737	–0.0281	0.0858	0.0519	0.0141

Table 3.17: Correlations between MAFs of MADs and original variables

maf-mad.eps

Figure 3.32: MAFs of MADs of TM bands from 1986 and 1988

abs-maf-mad.eps

Figure 3.33: Absolute values of MAFs of MADs of TM bands from 1986 and 1988

amm1p2.eps

Figure 3.34: Sum of absolute values of MAFs 1 and 2 of MADs of TM bands from 1986 and 1988

mafmadrgb.eps

Figure 3.35: MAFs 1, 2 and 3 of MADs of TM bands from 1986 and 1988 as RGB

absmafmadrgb.eps

Figure 3.36: Absolute values of MAFs 1, 2 and 3 of MADs of TM bands from 1986 and 1988 as RGB

This case study shows that because of the smart extension to univariate differences obtained by MAD analysis, all MAD components—also the high order MADs that contain information on maximum similarity as opposed to information on minimum similarity (i.e. change) contained in the low order MADs—are important in interpreting multivariate changes. This conclusion is supported by a case with simulated changes not shown here.

Appendix A

Redundancy Analysis

This appendix describes some further ways of analyzing the degree of overlap between the ordinary two-set canonical variates (cf. Section 3.1). In order to clarify matters let us recall a fundamental property of the correlation coefficient ρ, namely that the squared correlation between two variables equals the fraction of the variation in one variable Y that may be explained by an affine expression in the other variable X. If we call this predicted value $\hat{Y}$ we have

$$\mathrm{Var}\{\hat{Y}\} = \rho^2_{YX}\ \mathrm{Var}\{Y\}. \tag{A.1}$$

If the correlation equals 1 we have the same variance of the predicted value as of the original, i.e. we can make a perfect prediction of Y based on X. If on the other hand the correlation equals 0 the predictor X contains no information on Y.

If we have k predictors $X_1, \ldots, X_k$ the expression is still valid if we replace the correlation with the multiple correlation coefficient

$$\mathrm{Var}\{\hat{Y}\} = \rho^2_{Y|X_1,\ldots,X_k}\ \mathrm{Var}\{Y\}. \tag{A.2}$$

The squared multiple correlation coefficient is thus the degree of variation in Y that can be explained by $X_1, \ldots, X_k$.

After these more general remarks we shall investigate some properties of the correlation structure of canonical variates. From the definitions we get

$$\mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{U}\} = \mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{A}'\boldsymbol{X}\} = \boldsymbol{\Sigma}_{11}\boldsymbol{A} \tag{A.3}$$

$$\mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{V}\} = \mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{B}'\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{12}\boldsymbol{B} \tag{A.4}$$

$$\mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{V}\} = \mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{B}'\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{22}\boldsymbol{B} \tag{A.5}$$

$$\mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{U}\} = \mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{A}'\boldsymbol{X}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{A} \tag{A.6}$$

We shall for simplicity and without lack of generality assume that the X's an the Y's are standardized, i.e. they have variance 1. Then the above matrices are correlation matrices.

The expression

$$f_j = \frac{1}{p}\sum_{i=1}^{p}[\mathrm{Corr}\{X_i, U_j\}]^2 = \frac{1}{p}\,\boldsymbol{a}_j'\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}\boldsymbol{a}_j \tag{A.7}$$

equals the fraction of the (standardized) variance of the original X variables that are explained by canonical variate U_j. It follows that

$$\sum_{j=1}^{p} f_j = \sum_{j=1}^{p}\frac{1}{p}\,\boldsymbol{a}_j'\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}\boldsymbol{a}_j = \frac{1}{p}\,\mathrm{tr}\,(\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}\boldsymbol{A}), \tag{A.8}$$

where tr denotes the trace of a matrix. From $\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{A} = \boldsymbol{I}$ we obtain $\boldsymbol{A}' = \boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{11}$ and therefore

$$\sum_{j=1}^{p} f_j = \frac{1}{p}\,\mathrm{tr}\,(\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{A}) = \frac{1}{p}\,\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}\boldsymbol{A}\boldsymbol{A}^{-1}) = 1. \tag{A.9}$$

This corresponds to the fact that all of the variation in the original variables $X_1, \ldots, X_p$ may be explained by the whole set of canonical variates $U_1, \ldots, U_p$.

If we multiply f_j with the j'th squared canonical correlation λ_j we obtain the so called redundancy factor R_j. It may be instructive to note that

$$\begin{aligned} R_j &= \lambda_j f_j \\ &= \lambda_j \frac{1}{p}\, \boldsymbol{a}_j' \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{11} \boldsymbol{a}_j \\ &= \frac{1}{\lambda_j} \frac{1}{p}\, \boldsymbol{a}_j' \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{a}_j \\ &= \frac{1}{p}\, \boldsymbol{b}_j' \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{12} \boldsymbol{b}_j \\ &= \frac{1}{p} \sum_{i=1}^{p} [\mathrm{Corr}\{X_i, V_j\}]^2 \end{aligned} \tag{A.10}$$

Introducing the mnemotechnical expressions CVX_i for U_i and CVY_i for V_i and we obtain

$$\begin{aligned} R_j &= \frac{1}{p}\, \mathrm{cancor}_j^2 \sum_{i=1}^{p} [\mathrm{Corr}\{X_i, CVX_j\}]^2 \\ &= \frac{1}{p} \sum_{i=1}^{p} [\mathrm{Corr}\{X_i, CVY_j\}]^2. \end{aligned} \tag{A.11}$$

The squared canonical correlation is the shared variability between the two sets of canonical variates. Consequently we have the following interpretation of the

redundancy factor of the first set of variables given the availability of the second set of variables: R_j expresses the amount of variation in the original variables that is explained by the j'th canonical variate adjusted with the shared variation between the j'th canonical variates. This equals the amount of variation of variation explained by the opposite j'th canonical variate.

Related to these considerations are properties of multiple squared correlations between the original variables and the set of opposite canonical variates. If $\boldsymbol{\sigma}_i'$ denotes the i'th row in $\boldsymbol{\Sigma}_{12}$ the squared multiple correlation between X_i and $CVY_1, \ldots, CVY_k$ is

$$\begin{aligned} \rho_{X_i|CVY_1,\ldots,CVY_k} &= \boldsymbol{\sigma}_i' [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k] \begin{bmatrix} \boldsymbol{b}_1' \\ \vdots \\ \boldsymbol{b}_k' \end{bmatrix} \boldsymbol{\sigma}_i \\ &= \sum_{j=1}^{k} [\boldsymbol{\sigma}_i \boldsymbol{b}_j]^2 \\ &= \sum_{j=1}^{k} [\mathrm{Corr}\{X_i, CVY_j\}]^2, \end{aligned} \tag{A.12}$$

a quantity that is also useful in assessing overlap between the two sets of variables.

Appendix B

Multiset Canonical Correlations

Multiset canonical correlations analysis is dealt with in Section 3.2. This appendix deals with the maximization of the sum of squared covariances of the transformed variables only.

B.1 Maximize Sum of Squared Covariances

To maximize the sum of squared covariances under constraints in the multiset canonical correlations analysis problem we use a Lagrange multiplier technique as in Section 3.2.1. The same four constraints are examined.

Constraint 1: $\boldsymbol{a}_i^T \boldsymbol{a}_i - 1 = 0$

Introduce

$$F = W - 2\sum_{i=1}^{n} \lambda_i(\boldsymbol{a}_i^T \boldsymbol{a}_i - 1). \tag{B.1}$$

By setting $\partial F/\partial \boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n} (\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j) \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j = \lambda_i \boldsymbol{a}_i, \; i = 1, \ldots, n \tag{B.2}$$

or

$$\begin{bmatrix} \rho_{11}\boldsymbol{\Sigma}_{11} & \rho_{12}\boldsymbol{\Sigma}_{12} & \cdots & \rho_{1n}\boldsymbol{\Sigma}_{1n} \\ \rho_{21}\boldsymbol{\Sigma}_{21} & \rho_{22}\boldsymbol{\Sigma}_{22} & \cdots & \rho_{2n}\boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\boldsymbol{\Sigma}_{n1} & \rho_{n2}\boldsymbol{\Sigma}_{n2} & \cdots & \rho_{nn}\boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \boldsymbol{a}_1 \\ \lambda_2 \boldsymbol{a}_2 \\ \vdots \\ \lambda_n \boldsymbol{a}_n \end{bmatrix} \tag{B.3}$$

with $\rho_{ij} = \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j$ (in this case ρ_{ij} is not a correlation). This is not a normal eigensystem.

Constraint 2: $\boldsymbol{a}^T \boldsymbol{a} - 1 = \sum_{i=1}^n \boldsymbol{a}_i^T \boldsymbol{a}_i - 1 = 0$

Introduce

$$F = W - 2\lambda(\sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{a}_i - 1). \tag{B.4}$$

By setting $\partial F/\partial \boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n}(\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j)\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j = \lambda\boldsymbol{a}_i,\ i = 1,\ldots,n \tag{B.5}$$

or

$$\begin{bmatrix} \rho_{11}\boldsymbol{\Sigma}_{11} & \rho_{12}\boldsymbol{\Sigma}_{12} & \cdots & \rho_{1n}\boldsymbol{\Sigma}_{1n} \\ \rho_{21}\boldsymbol{\Sigma}_{21} & \rho_{22}\boldsymbol{\Sigma}_{22} & \cdots & \rho_{2n}\boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\boldsymbol{\Sigma}_{n1} & \rho_{n2}\boldsymbol{\Sigma}_{n2} & \cdots & \rho_{nn}\boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} \tag{B.6}$$

with $\rho_{ij} = \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j$ (in this case ρ_{ij} is not a correlation). This is not a normal eigensystem as ρ_{ij} depends on λ_i so the matrix on the left-hand-side is not constant.

Constraint 3: $\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1 = 0$

Introduce

$$F = W - 2\sum_{i=1}^{n}\lambda_i(\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1). \tag{B.7}$$

By setting $\partial F/\partial\boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n}(\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j)\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j = \lambda_i\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i,\ i = 1,\ldots,n \tag{B.8}$$

or

$$\begin{bmatrix} \rho_{11}\boldsymbol{\Sigma}_{11} & \rho_{12}\boldsymbol{\Sigma}_{12} & \cdots & \rho_{1n}\boldsymbol{\Sigma}_{1n} \\ \rho_{21}\boldsymbol{\Sigma}_{21} & \rho_{22}\boldsymbol{\Sigma}_{22} & \cdots & \rho_{2n}\boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\boldsymbol{\Sigma}_{n1} & \rho_{n2}\boldsymbol{\Sigma}_{n2} & \cdots & \rho_{nn}\boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \tag{B.9}$$
$$\begin{bmatrix} \lambda_1\boldsymbol{\Sigma}_{11} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2\boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \lambda_n\boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}$$

with $\rho_{ij} = \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j$ (in this case ρ_{ij} is a correlation). This is not a normal generalized eigensystem.

Constraint 4: $\boldsymbol{a}^T\boldsymbol{\Sigma}_D\boldsymbol{a} - 1 = \sum_{i=1}^n \boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1 = 0$

Introduce

$$F = W - 2\lambda(\sum_{i=1}^{n}\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i - 1). \tag{B.10}$$

By setting $\partial F/\partial\boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n}(\boldsymbol{a}_i^T\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j)\boldsymbol{\Sigma}_{ij}\boldsymbol{a}_j = \lambda\boldsymbol{\Sigma}_{ii}\boldsymbol{a}_i,\ i = 1,\ldots,n \tag{B.11}$$

or

$$
\begin{bmatrix}
\rho_{11}\boldsymbol{\Sigma}_{11} & \rho_{12}\boldsymbol{\Sigma}_{12} & \cdots & \rho_{1n}\boldsymbol{\Sigma}_{1n} \\
\rho_{21}\boldsymbol{\Sigma}_{21} & \rho_{22}\boldsymbol{\Sigma}_{22} & \cdots & \rho_{2n}\boldsymbol{\Sigma}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{n1}\boldsymbol{\Sigma}_{n1} & \rho_{n2}\boldsymbol{\Sigma}_{n2} & \cdots & \rho_{nn}\boldsymbol{\Sigma}_{nn}
\end{bmatrix}
\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} =
\lambda
\begin{bmatrix}
\boldsymbol{\Sigma}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \boldsymbol{\Sigma}_{22} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{nn}
\end{bmatrix}
\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}
\tag{B.12}
$$

with $\rho_{ij} = \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j$ (in this case ρ_{ij} is not a correlation). This is not a normal generalized eigensystem as ρ_{ij} depends on λ_k so the matrix on the left-hand-side is not constant.

Other Natural Constraints

Other natural constraints in this case are

1. $(\boldsymbol{a}_i^T \boldsymbol{a}_i)^2 = 1$,
2. $\sum_{i=1}^n (\boldsymbol{a}_i^T \boldsymbol{a}_i)^2 = 1$,
3. $(\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i)^2 = 1$, or
4. $\sum_{i=1}^n (\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i)^2 = 1$

All of these constraints lead to non-ordinary eigenvalue-type systems of this shape

$$
\begin{bmatrix}
\rho_{11}\boldsymbol{\Sigma}_{11} & \rho_{12}\boldsymbol{\Sigma}_{12} & \cdots & \rho_{1n}\boldsymbol{\Sigma}_{1n} \\
\rho_{21}\boldsymbol{\Sigma}_{21} & \rho_{22}\boldsymbol{\Sigma}_{22} & \cdots & \rho_{2n}\boldsymbol{\Sigma}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{n1}\boldsymbol{\Sigma}_{n1} & \rho_{n2}\boldsymbol{\Sigma}_{n2} & \cdots & \rho_{nn}\boldsymbol{\Sigma}_{nn}
\end{bmatrix}
\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} =
\tag{B.13}
$$

$$
\begin{bmatrix}
\lambda_1\rho_{11}\boldsymbol{\Sigma}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \lambda_2\rho_{22}\boldsymbol{\Sigma}_{22} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \lambda_n\rho_{nn}\boldsymbol{\Sigma}_{nn}
\end{bmatrix}
\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}
$$

with $\rho_{ij} = \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j$.

Appendix C

Computer Implementations

The construction and maintenance of a large toolbox of computer programs for (non-trivial) analysis of (spatial, multivariate, and multi-temporal) data is a never ending task. In the early and in the mid-1980s the software development in the IMSOR Image Group took place in a local environment, the Picture Processing System (PPS), developed by Jan Gunulf, Gert Nilsson and Bjarne Kjær Ersbøll under Knut Conradsen's supervision. Programs were written in Fortran and run on (then) large IBM main frames such as systems 360 and 370, later on systems 3081 running MVS/TSO under OS and 3033 running CMS under VM.

In 1985 a dedicated image processor, a GOP-300 from ContextVision AB, Sweden, was purchased. This engine was equipped with a powerful software package and also the group wrote its own software for the GOP. Bjarne Kjær Ersbøll and under his supervision Jan Pedersen were instrumental in this effort. The GOP has later been updated and it is still a strong machine.

To be compatible with some of our partners in large research projects, an ERDAS/PC system (updated in 1991, ERDAS (1990)) was purchased in 1988. I wrote a selection of computer programs (in Fortran, Nielsen (1990)) for the ERDAS/PC system running under DOS.

Presently our software development takes place on a network of UNIX workstations comprising HP 9000/7xx, Sun SPARC IPC, Sun SPARC 10, IBM RS/6000, Sony NeWS, SiliconGraphics Indigo 2, and powerful PCs (486s and Pentiums) running Linux. We are basing our toolbox of computer programs on the C programming language, the X Window System and the HIPS image processing system (Cohen & Landy, 1991; Landy, 1991). The HIPS system at IMM including the developments of the group itself (Nielsen, 1991), is maintained by Jens Michael Carstensen and myself. We intend to update a description of this software regularly and to make the description (along with a lot of other information from IMM, for instance this thesis) available on the World Wide Web via Mosaic.

The remainder of this appendix describes HIPS software developed at IMM that relates directly to the work described in this thesis.

C.1 Geostatistics

Based on Lee & Schachter (1980) Kristian Windfeld wrote `delaunay` to establish a Delaunay triangulation of a set of irregularly spaced points in 2-D.

On my initiative and under my supervision Karsten Hartelius wrote `crossv` to estimate 1- and 2-D cross-variograms, cross-covariance and cova functions, and `cokrig` to perform point cokriging. On my initiative and under my supervision Henrik Juul Hansen wrote `krig` to perform point and block (simple, ordinary and universal) kriging. `cokrig` and `krig` perform other types of interpolation also (such as inverse distance and inverse distance squared) and they provide different local characteristics such as local variance. Also, `cokrig` and `krig` are prepared for ancillary data such as digital elevation models, geological maps or maps of catchment areas. This type of information can then be included in the search for neighbors in the estimation process.

Also, additional formatting and plotting software (in S-PLUS, Statitistcal Sciences (1993)), and software to estimate 1- and 2-D semivariogram models (in SAS, SAS Institute (1990)) was written.

C.2 Dimensionality Reduction

On my initiative and under my supervision, several methods for orthogonalization and dimensionality reduction are implemented in a computer program, `maf` programmed by Rasmus Larsen. `maf` finds principal components, (rotated) principal factors, minimum/maximum autocorrelation factors, maximum noise fractions, (multiset) canonical variates (cf. Chapter 3) and linear combinations that give maximal multivariate differences of two sets of variables (MAD, Conradsen & Nielsen (1994), cf. Section 3.1.1).

Based on the Delaunay triangulation `sigma_n` (written with Karsten Hartelius) finds an estimate of the noise dispersion matrix as described above.

C.3 Multiset Data Analysis

On my initiative and under my supervision, the traditional method for performing two-set canonical correlations analysis is implemented in two computer programs, `maf` programmed by Rasmus Larsen and `cancorr` programmed by Anders Rosholm. `maf` also finds principal components, (rotated) principal factors, minimum/maximum autocorrelation factors, and linear combinations that give maximal multivariate differences (MAD, Conradsen & Nielsen (1994), see Section 3.1.1).

Also on my initiative and under my supervision, the multiset canonical correlations analysis methods of maximizing the sum of covariances under constraints 2 ($\sum \boldsymbol{a}_i^T \boldsymbol{a}_i = 1$) and 4 ($\sum \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = 1$) are implemented in `maf`.

All covariance matrices are found by the method of provisional means (Dixon, 1985). To find inverse covariance matrices, LINPACK routines `dpofa` and `dpodi` are used (Dongarra, Bunch, Moler, & Stewart, 1979)

- `dpofa` performs a Cholesky factorization of a positive definite matrix, $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$. $\boldsymbol{L}$ is lower triangular.

- `dpodi` uses the Cholesky factorization to find the determinant and/or the inverse of $\boldsymbol{\Sigma}$.

To solve the real, symmetric, generalized eigenproblem (RSG), EISPACK routines are used (Wilkinson & Reinsch, 1971; Garbow, Dongarra, Boyle, & Moler, 1977). The recommended EISPACK path to find all eigenvalues with all corresponding eigenvectors is to use routines `reduc`, `tred2`, `tql2` and `rebak`

- `reduc` reduces the real, symmetric, generalized eigenproblem $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x}$ where $\boldsymbol{B}$ is positive definite, to the standard real, symmetric eigenproblem $\boldsymbol{C}\boldsymbol{y} = \lambda\boldsymbol{y}$ using Cholesky factorization of $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}^T$. $\boldsymbol{L}$ is lower triangular. Output is $\boldsymbol{C} = \boldsymbol{L}^{-1}\boldsymbol{A}\boldsymbol{L}^{-T}$ with the same eigenvalues as the original RSG (eigenvectors $\boldsymbol{x} = \boldsymbol{L}^{-T}\boldsymbol{y}$ can be found by `rebak`).
- `tred2` reduces a real, symmetric matrix (in casu $\boldsymbol{C} = \boldsymbol{L}^{-1}\boldsymbol{A}\boldsymbol{L}^{-T}$) to a real, symmetric, tridiagonal matrix (with same eigenvalues) using the Householder method in which a series of orthogonal similarity transformations are accumulated.
- `tql2` determines eigenvalues and -vectors of a real, symmetric, tridiagonal matrix; the eigenvalues are computed by means of the QL algorithm (with shifting to accelerate convergence) which in turn involves successive orthogonal similarity transformations, resulting in convergence to a diagonal matrix; the eigenvectors are computed from the accumulated QL transformations.
- `rebak` forms eigenvectors of the RSG from the eigenvectors of the derived symmetric matrix (from `reduc`, $\boldsymbol{x} = \boldsymbol{L}^{-T}\boldsymbol{y}$).

To solve the real, symmetric eigenproblem (RS), only `tred2` and `tql2` are used. Good general descriptions of the methods used in the above computer programs are given in e.g. Strang (1980), Hansen (1987) and Press, Teukolsky, Vetterling, & Flannery (1992).

The remaining optimization problems described above (in fact all of them, including the eigenvalue problems) are solved by means of the GAMS (**G**eneral

Algebraic **M**odeling **S**ystem) (Brooke, Kendrick, & Meeraus, 1992) NLP solver CONOPT (Drud, 1985). A computer program `musecc` that writes the needed GAMS code, calls GAMS, reads GAMS output and performs the remaining analysis is implemented. The generic GAMS code was written by Dr. Arne Drud. Much of the remainder code for `musecc` comes from `maf`. The optimization problems involved could be solved by means of other algorithms also.

References

Andersen, J. S. (1994). Flerdimensionale rumligt korrelerede forureningsdata. Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby. In Danish.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (second edition). John Wiley, New York. 675 pp.

Armstrong, M. (1984). Problems with universal kriging. *Mathematical Geology*, **16**(1), 101–108.

Berman, M. (1994). Automated smoothing of image and other regularly spaced data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(5), 460–468.

Breiman, L. & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, **80**(391), 580–619.

Brooke, A., Kendrick, D., & Meeraus, A. (1992). *Release 2.25 GAMS: A User's Guide*. The Scientific Press, South San Francisco. 289 pp.

Buja, A. (1990). Remarks on functional canonical variates, alternating least squares methods and ACE. *The Annals of Statistics*, **18**, 1032–1069.

Clark, I. (1979). *Practical Geostatistics*. Elsevier Applied Science, London. 129 pp.

Cohen, Y. & Landy, M. S. (1991). The HIPS image processing software. Tech. rep. SharpImage Software. 30 pp.

Conradsen, K., Ersbøll, B. K., Nielsen, A. A., Pedersen, J. L., Stern, M. & Windfeld, K. (1991). *Development and Testing of New Techniques for Mineral-Exploration Based on Remote Sensing, Image Processing Methods and Multivariate Analysis. Final Report*. The Commission of the European Communities, Contract No. MA1M-0015-DK(B). 196 pp.

Conradsen, K., Nielsen, A. A., Windfeld, K., Ersbøll, B. K., Larsen, R., Hartelius, K. & Olsson, C. K. (1993). *Application and Development of New Techniques Based on Remote Sensing, Data Integration and Multivariate Analysis for Mineral Exploration. Final Report. Technical Annex*. The Commission of the European Communities, Contract No. MA2M-CT90-0010. Institute of Mathematical Modelling, Technical University of Denmark. 96 pp.

Conradsen, K. & Ersbøll, B. K. (1991). Data dependent orthogonal transformations of multichannel image data. Tech. rep. IMSOR, Technical University of Denmark. 35 pp.

Conradsen, K. & Nielsen, A. A. (1991). *Remote Sensing in Forecasting Agricultural Statistics in Kenya*. Danida, the Danish International Development Agency, Contract No. 104.Dan.8/410. 191 pp.

Conradsen, K. & Nielsen, A. A. (1994). Multivariate alteration detection (MAD) in multispectral, bi-temporal image data: a new approach to change detection studies. Submitted to *Remote Sensing of Environment*.

Conradsen, K., Nielsen, B. K., & Thyrsted, T. (1985). A comparison of min/max autocorrelation factor analysis and ordinary factor analysis. In *Proceedings from Symposium in Applied Statistics*, pp. 47–56. Lyngby, Denmark.

Conradsen, K., Nielsen, B. K., & Nielsen, A. A. (1991a). Noise removal in multichannel image data by a parametric maximum noise fractions estimator. In Environmental Research Institute of Michigan (Ed.), *Proceedings of* *the 24th International Symposium on Remote Sensing of Environment*, pp. 403–416. Rio de Janeiro, Brazil.

Conradsen, K., Nielsen, B. K., & Nielsen, A. A. (1991b). Integration of multi-source data in mineral exploration. In Environmental Research Institute of Michigan (Ed.), *Proceedings of the Eighth Thematic Conference on Geologic Remote Sensing*, pp. 1053–1066. Denver, Colorado, USA.

Conradsen, K., Nielsen, A. A., & Windfeld, K. (1992). Analysis of geochemical data sampled on a regional scale. In Walden, A. & Guttorp, P. (Eds.), *Statistics in the Environmental and Earth Sciences*, pp. 283–300. Griffin.

Cooley, W. W. & Lohnes, P. R. (1971). *Multivariate Data Analysis*. John Wiley and Sons, New York.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, **17**(5).

Cressie, N. A. C. (1991). *Statistics for Spatial Data*. Wiley & Sons, New York. 900 pp.

David, M. (1977). *Geostatistical Ore Reserve Estimation. Developments in Geomathematics 2*. Elsevier, Amsterdam. 364 pp.

David, M. (1988). *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Developments in Geomathematics 6*. Elsevier, Amsterdam. 216 pp.

Dixon, W. J. (Ed.). (1985). *BMDP Statistical Software*. University of California Press. 734 pp.

Dongarra, J. J., Bunch, J. R., Moler, C. B., & Stewart, G. W. (1979). *LINPACK Users' Guide*. Society for Industrial and Applied Mathematics.

Drud, A. (1985). CONOPT – a GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming*, **31**, 153–191.

ERDAS, Inc. (1990). *Field Guide*. ERDAS, Inc., Atlanta. 410 pp.

Ersbøll, B. K. (1989). *Transformations and classifications of remotely sensed data*. Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby. 297 pp.

Friedman, J. H. & Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE, Trans. Comput., Ser. C*, **23**, 881–889.

Fung, T. & LeDrew, E. (1987). Application of principal components analysis to change detection. *Photogrammetric Engineering and Remote Sensing*, **53**(12), 1649–1658.

GAF, MAYASA, IMSOR, & DLR (1993). *Application and Development of New Techniques Based on Remote Sensing, Data Integration and Multivariate Analysis for Mineral Exploration. Final Report*. The Commission of the European Communities, Contract No. MA2M-CT90-0010. 117 pp.

Garbow, B. S., Dongarra, J. J., Boyle, J. M., & Moler, C. B. (1977). *Lecture Notes in Computer Science: Matrix Eigensystem Routines – EISPACK Guide Extension*. Springer-Verlag (G. Goos and J. Hartmanis, ed.). 343 pp.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley and Sons. 311 pp.

Green, A. 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.

Grunsky, E. C. & Agterberg, F. P. (1988). Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis area, Ontario. *Mathematical Geology*, **20**(7), 825–861.

Grunsky, E. C. & Agterberg, F. P. (1991). SPFAC: a Fortran program for spatial factor analysis of multivariate data. *Computers & Geosciences*, **17**(1), 133–160.

Hanaizumi, H. & Fujimura, S. (1992). Change detection from remotely sensed multi-temporal images using multiple regression. In *Proceedings from the 1992 International Geoscience and Remote Sensing Symposium*, pp. 564–566.

Hanaizumi, H., Chino, S., & Fujimura, S. (1994). A method for change analysis with weight of significance using multi-temporal, multi-spectral images. To appear in *Proceedings from the European Symposium on Satellite Remote Sensing*.

Hansen, P. S. (1987). *Lineær algebra – datamatorienteret*. Matematetisk Institut og Numerisk Institut, Danmarks Tekniske Højskole, Lyngby.

Horst, P. (1961). Relations among m sets of measures. *Psychometrika*, **26**, 129–149.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, **24**, 417–441.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **XXVIII**, 321–377.

Isaaks, E. H. & Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York. 561 pp.

Journel, A. G. & Froidevaux, R. (1982). Anisotropic hole-effect modeling. *Mathematical Geology*, **14**(3), 217–239.

Journel, A. G. & Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press, London. 600 pp.

Journel, A. G. & Rossi, M. E. (1989). When do we need a trend model in kriging?. *Mathematical Geology*, **21**(7), 715–739.

Journel, A. G. (1989). *Fundamentals of Geostatistics in Five Lessons. Short Course in Geology: Volume 8*. American Geophysical Union, Washington DC. 40 pp.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, **58**, 433–451.

Landy, M. S. (1991). A programmer's guide to the HIPS software. Tech. rep. Department of Psychology and Center for Neural Science, New York University. 24 pp.

Larsen, R. (1991). MAF and other transformations applied in remote sensing. Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby. In Danish. 130 pp.

Lee, D. T. & Schachter, B. J. (1980). Two alogrithms for constructing a Delaunay triangulation. *Int. J. Comp. and Info. Sci.*, **9**(3), 219–242.

Lee, J. B., Woodyatt, A. S., & Berman, M. (1990). Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, **28**(3), 295–304.

Nielsen, A. A. & Larsen, R. (1994). Restoration of GERIS data using the maximum noise fractions transform. In Environmental Research Institute of Michigan (Ed.), *Proceedings from the First International Airborne Remote Sensing Conference and Exhibition, Volume II*, pp. 557–568. Strasbourg, France.

Nielsen, A. A. (1990). Computer programs for inclusion in the ERDAS. In ERDAS, Inc. (Ed.), *Proceedings of the Third European User Group Meeting*. 7 pp.

Nielsen, A. A. (1991). HIPS programs from the IMSOR image group. Tech. rep. Institute of Mathematical Modelling, Technical University of Denmark. 39 pp.

Nielsen, A. A. (1993). 2D semivariograms. In Cilliers, P. (Ed.), *Proceedings of the Fourth South African Workshop on Pattern Recognition*, pp. 25–35. Simon's Town, South Africa.

Nielsen, A. A. (1994a). Geochemistry in Eastern Erzgebirge: data report. Tech. rep. Institute of Mathematical Modelling, Technical University of Denmark. 38 pp.

Nielsen, A. A. (1994b). Geophysics and integration with geochemistry in Eastern Erzgebirge: data report. Tech. rep. Institute of Mathematical Modelling, Technical University of Denmark. 25 pp.

Olesen, B. L. (1984). *Geochemical Mapping of South Greenland*. Ph.D. thesis, Department of Mineral Industry, Technical University of Denmark, Lyngby.

Olsen, S. I. (1993). Estimation of noise in images: an evaluation. *Graphical Models and Image Processing*, **55**(4), 319–323.

Pendock, N. & Nielsen, A. A. (1993). Multispectral image enhancement neural networks and the maximum noise fraction transform. In Cilliers, P. (Ed.), *Proceedings of the Fourth South African Workshop on Pattern Recognition*, pp. 2–13. Simon's Town, South Africa.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. R. (1992). *Numerical Recipes in C: The Art of Scientific Computing* (second edition). Cambridge University Press. 994 pp.

Ripley, B. (1981). *Spatial Statistics*. J. Wiley & Sons. 252 pp.

Royer, J. J. & Mallet, J. L. (1982). Imagerie spatielle et cartographie géochemique. Étude des correlation entre géochemie en roches et télédétetion: application au massif de la Marche Orientale (Massif Central). Tech. rep. 80/CNES/279, CNES et CNRS. 125 pp.

SAS Institute, Inc. (1990). *SAS Language and Procedures: Introduction, Version 6*. SAS Institute, Inc., Cary. 124 pp.

Schneider, T., Petersen, O. H., Nielsen, A. A., & Windfeld, K. (1990). A geostatistical approach to indoor surface sampling strategies. *Journal of Aerosol Science*, **21**(4), 555–567.

Statitistcal Sciences (1993). *S-PLUS User's Manual, Version 3.2*. StatSci, a division of MathSoft, Inc., Seattle.

Shettigara, K. V. & McGilchrist, C. A. (1989). A principal component and canonical correlation hybrid technique for change detection in two-image sets. In Barrett, R. F. (Ed.), *ASSPA 89, Signal Processing, Theories, Implementations and Applications*, pp. 47–52.

Shi, S. G. & Taam, W. (1992). Non-linear canonical correlation analysis with a simulated annealing solution. *Journal of Applied Statistics*, **19**(2), 155–165.

Simpson, J. J. (1994). Accurate change detection during ENSO events using the multivariate alteration detection (MAD) transformation. Submitted to *Remote Sensing of Environment*.

Steel, R. G. D. (1951). Minimum generalized variance for a set of linear functions. *The Annals of Mathematical Statistics*, **22**, 456–460.

Stern, M. (1990). Video of NOAA AVHRR GAC change detection over a 7 year period in Sudan. Pers. comm.

Strang, G. (1980). *Linear Algebra and Its Applications* (second edition). Academic Press, New York. 414 pp.

Switzer, P. & Green, A. A. (1984). Min/max autocorrelation factors for multivariate spatial imagery. Tech. rep. 6, Stanford University. 10 pp.

van der Burg, E. & de Leeuw, J. (1983). Non-linear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, **36**, 54–80.

Vinograde, B. (1950). Canonical positive definite matrices under internal linear transformations. In *Proceedings of the American Mathematical Society*, pp. 159–161.

Wilkinson, J. H. & Reinsch, C. (1971). *Handbook for Automatic Computation, Volume II*. Springer-Verlag, New York.

Windfeld, K. (1992). *Application of Computer Intensive Data Analysis: Methods to the Analysis of Digital Images and Spatial Data*. Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby. 190 pp.

Ph.D. theses from IMSOR

1. **Sigvaldason, Helgi.** (1963). *Beslutningsproblemer ved et hydro-termisk elforsyningssystem.* 92 pp.

2. **Nygaard, Jørgen.** (1966). *Behandling af et dimensionerings problem i telefonien.* 157 pp.

3. **Krarup, Jakob.** (1967). *Fixed-cost and other network flow problems as related to plant location and to the design of transportation and computer systems.* 159 pp.

4. **Hansen, Niels Herman.** (1967). *Problemer ved forudsigelse af lydhastighed i danske farvande. Analyse af et stokastisk system, del 1, 104 pp. del 2, Figurer og tabeller.* 95 pp.

5. **Larsen, Mogens E.** (1968). *Statistisk analyse af elementære kybernetiske systemer.* 210 pp.

6. **Punhani, Amrit Lal.** (1968). *Decision problems in connection with atomic power plants.* 133 pp.

7. **Clausen, Svend.** (1969). *Kybernetik, systemer og modeller.* 205 pp.

8. **Vidal, R.V. Valqui.** (1970). *Operations research in production planning. Interconnections between production and demand, vol. 1-2.* 321 pp.

9. **Bilde, Ole.** (1970). *Nonlinear and discrete programming in transportation, location and road design, vol. 1-2.* 291 pp.

10. **Rasmusen, Hans Jørgen.** (1972). *En decentraliseret planlægningsmodel.* 185 pp.

11. **Dyrberg, Christian.** (1973). *Tilbudsgivning i en entreprenør virksomhed.* 158 pp.

12. **Madsen, Oli B.G.** (1973). *Dekomposition og matematisk programmering.* 271 pp.

13. **Dahlgaard, Peter.** (1973). *Statistical aspects of tide prediction, vol. 1, 202 pp. vol. 2, Figures and tables.* 170 pp.

14. **Spliid, Henrik.** (1973). *En statistisk model for stormflodsvarsling.* 205 pp.

15. **Pinochét, Mario.** (1973). *Operations research in strategic transportation planning. The decision process in a multiharbour system.* 374 pp.

16. **Christensen, Torben.** (1973). *Om semi-markov processer. Udvidelser og anvendelser inden for den sociale sektor.* 239 pp.

17. **Jacobsen, Søren Kruse.** (1973). *Om lokaliseringsproblemer, modeller og løsninger.* 355 pp.

18. **Marqvardsen, Hans.** (1973). *Skemalægning ved numerisk simulation.* 222 pp.

19. **Mortensen, Jens Hald.** (1974). *Interregionale godstransporter. Teoridannelser og modeller.* 223 pp.

20. **Severin, Juan Melo.** (1974). *Introduction to operations research in systems synthesis. A chemical process design synthesis application.* 249 pp.

21. 22. **Spliid, Iben & Uffe Bundgaard-Jørgensen.** (1974). *Skitse til en procedure for kommunalplanlægning.* 544 pp.

23. **Mosgaard, Christian.** (1975). *International planning in disaster situations.* 187 pp.

24. **Holm, Jan.** (1975). *En optimeringsmodel for kollektiv trafik.* 246 pp.

25. **Jensson, Pall.** (1975). *Stokastisk programmering, del 1, modeller, del 2, metodologiske overvejelser og anvendelser.* 333 pp.

26. **Iversen, Villy Bæk.** (1976). *On the accuracy in measurements of time intervals and traffic intensities with application to teletraffic and simulation.* 202 pp.

27. **Drud, Arne.** (1976). *Methods for control of complex dynamic systems. Illustrated by econometric models.* 209 pp.

28. **Togsverd, Tom.** (1976). *Koordinering af kommunernes ressourceforbrug.* 295 pp.

29. **Jensen, Olav Holst.** (1976). *Om planlægning af kollektiv trafik. Operationsanalytiske modeller og løsningsmetoder.* 321 pp.

30. **Beyer, Jan E.** (1976). *Ecosystems. An operational research approach.* 315 pp.

31. **Bille, Thomas Bastholm.** (1977). *Vurdering af egnsudviklingsprojekter. Samspil mellem benifit-cost analyse og den politiske vurdering i en tid under forandring.* 260 pp.

32. **Holst, Erik.** (1979). *En statistisk undersøgelse af tabletserier.* 316 pp.

33. **Aagaard-Svendsen, Rolf.** (1979). *Econometric methods and Kalman filtering.* 300 pp.

34. **Hansen, Steen.** (1979). *Project control by quantitative methods.* 230 pp.

35. **Scheufens, Ernst Edvard.** (1980). *Statistisk analyse og kontrol af tidsafhængige vandkvalitetsdata.* 152 pp.

36. **Lyngvig, Jytte.** (1981). *Samfundsøkonomisk planlægning.* 252 pp.

37. **Troelsgård, Birgitte.** (1981). *Statistisk bestemmelse af modeller for rumlufttemperatur.* 213 pp.

38. **Raft, Ole.** (1981). *Delivery planning by modular algorithms.* 220 pp.

39. **Jensen, Sigrid M.** (1981). *Analyse af interregionale togrejser. 212 pp. + Figurer og appendices.* 174 pp.

40. **Ravn, Hans.** (1982). *Technology and underdevelopment. The case of Mexico.* 376 pp.

41. **Hansen, Sten.** (1983). *Phase-type distributions in queueing theory.* 209 pp.

42. **Ferreira, José A.S.** (1984). *Optimal control of discrete-time systems with applications.* 252 pp.

43. **Behrens, Jens Christian.** (1985). *Mathematical modelling of aquatic ecosystems applied to biological waste water treatment.* 28 pp. + app. 1, 389 pp. + app. 2, 180 pp.

44. **Poulsen, Niels Kjølstad.** (1985). *Robust self tuning controllers.* 240 pp.

45. **Madsen, Henrik.** (1985). *Statistically determined dynamic models for climate processes, part 1-2.* 428 pp.

46. **Sørensen, Bo.** (1986). *Interactive distribution planning.* 253 pp.

47. **Lethan, Helge B.** (1986). *Løsning af store kombinatoriske problemer.* 173 pp

48. **Boelskifte, Søren.** (1988). *Dispersion and current measurements. An investigation based on time series analysis and turbulence models.* Risø-M-2566. 154 pp.

49. **Nielsen, Bo Friis.** (1988). *Modelling of multiple access systems with phase type distributions.* 253 pp.

50. **Christensen, John M.** (1988). *Project planning and analysis. Methods for assessment of rural energy projects in developing countries.* Risø-M-2706. 158 pp.

51. **Olsen, Klaus Juel.** (1988). *Texture analysis of ultrasound images of livers.* 162 pp.

52. **Holst, Helle.** (1988). *Statistisk behandling af nærinfrarøde refleksionsmålinger.* 309 pp. + app.

53. **Knudsen, Torben.** (1989). *Start/stop strategier for vind-diesel systemer.* 275 pp.

54. **Ersbøll, Bjarne Kjær.** (1989). *Transformations and classifications of remotely sensed data. Theory and geological cases.* 297 pp.

55. **Kragh, Anders Laage.** (1990). *Kø-netværksmodeller til analyse af FMS anlæg.* 205 pp.

56. **Hansen, Christian Kornerup.** (1990). *Statistical methods in the analysis of repairable systems reliability.* 56 pp.

57. **Parkum, Jens Ejnar.** (1992). *Recursive identification of time-varying systems.* 206 pp.

58. **Bilbo, Carl M.** (1992). *Statistical analysis of multivariate degradation models.* 167 pp.

59. **Carstensen, Jens Michael.** (1992). *Description and simulation of visual texture.* 234 pp.

60. **Halse, Karsten.** (1992). *Modeling and solving complex vehicle routing problems.* 372 pp.

61. **Hendricks, Elbert.** (1992). *Identification and estimation of nonlinear systems using physical modelling.* 273 pp.

62. **Windfeld, Kristian.** (l992). *Application of computer intensive data analysis methods to the analysis of digital images and spatial data.* 190 pp.

63. **Iwersen, Jørgen.** (1992). *Statistical control charts. Performance of Shewhart and Cusum charts.* 326 pp.

64. **Olsson, Carsten Kruse.** (1993). *Image processing methods in materials science.* 274 pp.

65. **Sejling, Ken.** (1993). *Modelling and prediction of load in heating systems.* 290 pp.

66. **Søgaard, Henning T.** (1993). *Stochastic systems with embedded parameter variations. Applications to district heating.* 286 pp.

67. **Grunkin, Michael.** (1993). *On the analysis of image data using simultaneous interaction models.* 223 pp.

68. **Pálsson, Ólafur Pétur.** (1993). *Stochastic modeling, control and optimization of district heating systems.* 282 pp.

69. **Sørensen, Lene.** (1993). *Environmental planning and uncertainty.* Risø-R-709(EN). 226 pp.

70. **Thygesen, Bent Georg.** (1993). *Modelling and control of systems in the delta domain.* 207 pp.

71. **Jensen, Jens Arnth.** (1993). *Digital image processing designed for evaluation of MR-examinations of the cardiovascular system.* 256 pp.

72. **Jørgensen, Jan Friis.** (1993). *Scanning probe microscopy image restoration and analysis.* ATV erhvervsforskerprojekt EF 349. l44 pp.

73. **Carstensen, Jacob.** (1994). *Identification of wastewater processes.* ATV erhvervsforskerprojekt EF 354. 242 pp.

Ph.D. theses from IMM

1. **Larsen, Rasmus.** (1994). *Estimation of visual motion in image sequences.* 143 pp.

2. **Rygard, Jens Moberg.** (1994). *Design and optimization of flexible manufacturing systems.* 232 pp.

3. **Lassen, Niels Christian Krieger.** (1994). *Automated determination of crystal orientations from electron backscattering patterns.* 136 pp.

4. **Melgaard, Henrik.** (1994). *Identification of physical models.* 246 pp.

5. **Wang, Chunyan.** (1994). *Stochastic differential equations and a biological system.* 153 pp.

6. **Nielsen, Allan Aasbjerg.** (1994). *Multivariate data-analysis of regularly and irregularly sampled spatial data.* 213 pp.