

THE UNIVERSITY OF CALGARY

**HYPERSPECTRAL IMAGE ANALYSIS: A STUDY ON BAND SELECTION  
METHODS AND THE SENSITIVITY OF LINEAR SPECTRAL UNMIXING**

by

AIRONG ZHANG

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

JUNE, 1994

© Airong Zhang 1994



National Library  
of Canada

Acquisitions and  
Bibliographic Services Branch

395 Wellington Street  
Ottawa, Ontario  
K1A 0N4

Bibliothèque nationale  
du Canada

Direction des acquisitions et  
des services bibliographiques

395, rue Wellington  
Ottawa (Ontario)  
K1A 0N4

*Your file    Votre référence*

*Our file    Notre référence*

THE AUTHOR HAS GRANTED AN  
IRREVOCABLE NON-EXCLUSIVE  
LICENCE ALLOWING THE NATIONAL  
LIBRARY OF CANADA TO  
REPRODUCE, LOAN, DISTRIBUTE OR  
SELL COPIES OF HIS/HER THESIS BY  
ANY MEANS AND IN ANY FORM OR  
FORMAT, MAKING THIS THESIS  
AVAILABLE TO INTERESTED  
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE  
IRREVOCABLE ET NON EXCLUSIVE  
PERMETTANT A LA BIBLIOTHEQUE  
NATIONALE DU CANADA DE  
REPRODUIRE, PRETER, DISTRIBUER  
OU VENDRE DES COPIES DE SA  
THESE DE QUELQUE MANIERE ET  
SOUS QUELQUE FORME QUE CE SOIT  
POUR METTRE DES EXEMPLAIRES DE  
CETTE THESE A LA DISPOSITION DES  
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP  
OF THE COPYRIGHT IN HIS/HER  
THESIS. NEITHER THE THESIS NOR  
SUBSTANTIAL EXTRACTS FROM IT  
MAY BE PRINTED OR OTHERWISE  
REPRODUCED WITHOUT HIS/HER  
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE  
DU DROIT D'AUTEUR QUI PROTEGE  
SA THESE. NI LA THESE NI DES  
EXTRAITS SUBSTANTIELS DE CELLE-  
CI NE DOIVENT ETRE IMPRIMES OU  
AUTREMENT REPRODUITS SANS SON  
AUTORISATION.

ISBN 0-315-99532-7

Name AIRONG ZHANG

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

*General*

SUBJECT TERM

0537

SUBJECT CODE

U·M·I

## Subject Categories

### THE HUMANITIES AND SOCIAL SCIENCES

#### COMMUNICATIONS AND THE ARTS

Architecture ..... 0729  
Art History ..... 0377  
Cinema ..... 0900  
Dance ..... 0378  
Fine Arts ..... 0357  
Information Science ..... 0723  
Journalism ..... 0391  
Library Science ..... 0399  
Mass Communications ..... 0708  
Music ..... 0413  
Speech Communication ..... 0459  
Theater ..... 0465

#### EDUCATION

General ..... 0515  
Administration ..... 0514  
Adult and Continuing ..... 0516  
Agricultural ..... 0517  
Art ..... 0273  
Bilingual and Multicultural ..... 0282  
Business ..... 0688  
Community College ..... 0275  
Curriculum and Instruction ..... 0727  
Early Childhood ..... 0518  
Elementary ..... 0524  
Finance ..... 0277  
Guidance and Counseling ..... 0519  
Health ..... 0680  
Higher ..... 0745  
History of ..... 0520  
Home Economics ..... 0278  
Industrial ..... 0521  
Language and Literature ..... 0279  
Mathematics ..... 0280  
Music ..... 0522  
Philosophy of ..... 0998  
Physical ..... 0523

Psychology ..... 0525  
Reading ..... 0535  
Religious ..... 0527  
Sciences ..... 0714  
Secondary ..... 0533  
Social Sciences ..... 0534  
Sociology of ..... 0340  
Special ..... 0529  
Teacher Training ..... 0530  
Technology ..... 0710  
Tests and Measurements ..... 0288  
Vocational ..... 0747

#### LANGUAGE, LITERATURE AND LINGUISTICS

Language ..... 0679  
General ..... 0289  
Ancient ..... 0290  
Linguistics ..... 0291  
Modern ..... 0401  
Literature ..... 0294  
Classical ..... 0295  
Comparative ..... 0297  
Medieval ..... 0298  
Modern ..... 0316  
African ..... 0591  
American ..... 0305  
Asian ..... 0352  
Canadian (English) ..... 0355  
Canadian (French) ..... 0593  
English ..... 0311  
Germanic ..... 0312  
Latin American ..... 0315  
Middle Eastern ..... 0313  
Romance ..... 0314  
Slavic and East European ..... 0370

#### PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy ..... 0422  
Religion ..... 0318  
General ..... 0321  
Biblical Studies ..... 0319  
Clergy ..... 0320  
History of ..... 0322  
Philosophy of ..... 0469  
Theology ..... 0323

#### SOCIAL SCIENCES

American Studies ..... 0324  
Anthropology ..... 0326  
Archaeology ..... 0327  
Cultural ..... 0310  
Physical ..... 0272  
Business Administration ..... 0770  
General ..... 0454  
Accounting ..... 0338  
Marketing ..... 0385  
Canadian Studies ..... 0501  
Economics ..... 0503  
General ..... 0505  
Agricultural ..... 0508  
Commerce-Business ..... 0509  
Finance ..... 0510  
History ..... 0511  
Labor ..... 0358  
Theory ..... 0366  
Folklore ..... 0351  
Geography ..... 0578  
Gerontology ..... 0579  
History ..... 0579  
General ..... 0579

Ancient ..... 0579  
Medieval ..... 0581  
Modern ..... 0582  
Black ..... 0328  
African ..... 0331  
Asia, Australia and Oceania ..... 0332  
Canadian ..... 0334  
European ..... 0335  
Latin American ..... 0336  
Middle Eastern ..... 0337  
United States ..... 0585  
History of Science ..... 0398  
Law ..... 0615  
Political Science ..... 0616  
General ..... 0617  
International Law and Relations ..... 0814  
Public Administration ..... 0452  
Recreation ..... 0626  
Social Work ..... 0627  
Sociology ..... 0938  
General ..... 0631  
Criminology and Penology ..... 0628  
Demography ..... 0629  
Ethnic and Racial Studies ..... 0630  
Individual and Family Studies ..... 0700  
Industrial and Labor Relations ..... 0344  
Public and Social Welfare ..... 0709  
Social Structure and Development ..... 0999  
Theory and Methods ..... 0453  
Transportation ..... 0453  
Urban and Regional Planning ..... 0453  
Women's Studies ..... 0453

### THE SCIENCES AND ENGINEERING

#### BIOLOGICAL SCIENCES

Agriculture ..... 0473  
General ..... 0285  
Agronomy ..... 0475  
Animal Culture and Nutrition ..... 0476  
Animal Pathology ..... 0359  
Food Science and Technology ..... 0478  
Forestry and Wildlife ..... 0479  
Plant Culture ..... 0480  
Plant Pathology ..... 0817  
Plant Physiology ..... 0777  
Range Management ..... 0746  
Wood Technology ..... 0306  
Biology ..... 0287  
General ..... 0308  
Anatomy ..... 0309  
Biostatistics ..... 0379  
Botany ..... 0329  
Cell ..... 0353  
Ecology ..... 0369  
Entomology ..... 0793  
Genetics ..... 0410  
Limnology ..... 0307  
Microbiology ..... 0317  
Molecular ..... 0416  
Neuroscience ..... 0433  
Oceanography ..... 0821  
Physiology ..... 0778  
Radiation ..... 0472  
Veterinary Science ..... 0786  
Zoology ..... 0760  
Biophysics ..... 0425  
General ..... 0996  
Medical ..... 0996

#### EARTH SCIENCES

Biogeochemistry ..... 0425  
Geochemistry ..... 0996

Geodesy ..... 0370  
Geology ..... 0372  
Geophysics ..... 0373  
Hydrology ..... 0388  
Mineralogy ..... 0411  
Paleobotany ..... 0345  
Paleoecology ..... 0426  
Paleontology ..... 0418  
Paleozoology ..... 0985  
Palynology ..... 0427  
Physical Geography ..... 0368  
Physical Oceanography ..... 0415

#### HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences ..... 0768  
Health Sciences ..... 0566  
General ..... 0300  
Audiology ..... 0992  
Chemotherapy ..... 0567  
Dentistry ..... 0350  
Education ..... 0769  
Hospital Management ..... 0758  
Human Development ..... 0982  
Immunology ..... 0564  
Medicine and Surgery ..... 0347  
Mental Health ..... 0569  
Nursing ..... 0570  
Nutrition ..... 0380  
Obstetrics and Gynecology ..... 0354  
Occupational Health and Therapy ..... 0381  
Ophthalmology ..... 0571  
Pathology ..... 0419  
Pharmacology ..... 0572  
Pharmacy ..... 0382  
Physical Therapy ..... 0573  
Public Health ..... 0574  
Radiology ..... 0575  
Recreation ..... 0460  
Speech Pathology ..... 0383  
Toxicology ..... 0386  
Home Economics ..... 0386

#### PHYSICAL SCIENCES

Pure Sciences ..... 0485  
Chemistry ..... 0749  
General ..... 0486  
Agricultural ..... 0487  
Analytical ..... 0488  
Biochemistry ..... 0738  
Inorganic ..... 0490  
Nuclear ..... 0491  
Organic ..... 0494  
Physical ..... 0495  
Pharmaceutical ..... 0754  
Polymer ..... 0405  
Radiation ..... 0605  
Mathematics ..... 0986  
Physics ..... 0606  
General ..... 0608  
Acoustics ..... 0748  
Astronomy and Astrophysics ..... 0607  
Elementary Particles and High Energy ..... 0798  
Fluid and Plasma ..... 0759  
Molecular ..... 0609  
Nuclear ..... 0610  
Optics ..... 0752  
Radiation ..... 0756  
Solid State ..... 0611  
Statistics ..... 0463

#### Applied Sciences

Applied Mechanics ..... 0346  
Computer Science ..... 0984

Engineering ..... 0537  
General ..... 0538  
Aerospace ..... 0539  
Agricultural ..... 0540  
Automotive ..... 0541  
Biomedical ..... 0542  
Chemical ..... 0543  
Civil ..... 0544  
Electronics and Electrical ..... 0348  
Heat and Thermodynamics ..... 0545  
Hydraulic ..... 0546  
Industrial ..... 0547  
Marine ..... 0794  
Materials Science ..... 0548  
Mechanical ..... 0743  
Metallurgy ..... 0551  
Mining ..... 0552  
Nuclear ..... 0549  
Packaging ..... 0765  
Petroleum ..... 0554  
Sanitary and Municipal ..... 0790  
System Science ..... 0428  
Geotechnology ..... 0796  
Operations Research ..... 0795  
Plastics Technology ..... 0994  
Textile Technology ..... 0621

#### PSYCHOLOGY

General ..... 0621  
Behavioral ..... 0384  
Clinical ..... 0622  
Developmental ..... 0620  
Experimental ..... 0623  
Industrial ..... 0624  
Personality ..... 0625  
Physiological ..... 0989  
Psychobiology ..... 0349  
Psychometrics ..... 0632  
Social ..... 0451



Nom \_\_\_\_\_

*Dissertation Abstracts International* est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.



SUJET

CODE DE SUJET

## Catégories par sujets

### HUMANITÉS ET SCIENCES SOCIALES

#### COMMUNICATIONS ET LES ARTS

Architecture .....	0729
Beaux-arts .....	0357
Bibliothéconomie .....	0399
Cinéma .....	0900
Communication verbale .....	0459
Communications .....	0708
Danse .....	0378
Histoire de l'art .....	0377
Journalisme .....	0391
Musique .....	0413
Sciences de l'information .....	0723
Théâtre .....	0465

#### ÉDUCATION

Généralités .....	515
Administration .....	0514
Art .....	0273
Collèges communautaires .....	0275
Commerce .....	0688
Économie domestique .....	0278
Éducation permanente .....	0516
Éducation préscolaire .....	0518
Éducation sanitaire .....	0680
Enseignement agricole .....	0517
Enseignement bilingue et multiculturel .....	0282
Enseignement industriel .....	0521
Enseignement primaire .....	0524
Enseignement professionnel .....	0747
Enseignement religieux .....	0527
Enseignement secondaire .....	0533
Enseignement spécial .....	0529
Enseignement supérieur .....	0745
Évaluation .....	0288
Finances .....	0277
Formation des enseignants .....	0530
Histoire de l'éducation .....	0520
Langues et littérature .....	0279

Lecture .....	0535
Mathématiques .....	0280
Musique .....	0522
Orientation et consultation .....	0519
Philosophie de l'éducation .....	0998
Physique .....	0523
Programmes d'études et enseignement .....	0727
Psychologie .....	0525
Sciences .....	0714
Sciences sociales .....	0534
Sociologie de l'éducation .....	0340
Technologie .....	0710

#### LANGUE, LITTÉRATURE ET LINGUISTIQUE

Langues	
Généralités .....	0679
Anciennes .....	0289
Linguistique .....	0290
Modernes .....	0291
Littérature	
Généralités .....	0401
Anciennes .....	0294
Comparée .....	0295
Médiévale .....	0297
Moderne .....	0298
Africaine .....	0316
Américaine .....	0591
Anglaise .....	0593
Asiatique .....	0305
Canadienne (Anglaise) .....	0352
Canadienne (Française) .....	0355
Germanique .....	0311
Latino-américaine .....	0312
Moyen-orientale .....	0315
Romane .....	0313
Slave et est-européenne .....	0314

#### PHILOSOPHIE, RELIGION ET THÉOLOGIE

Philosophie .....	0422
Religion	
Généralités .....	0318
Clergé .....	0319
Études bibliques .....	0321
Histoire des religions .....	0320
Philosophie de la religion .....	0322
Théologie .....	0469

#### SCIENCES SOCIALES

Anthropologie	
Archéologie .....	0324
Culturelle .....	0326
Physique .....	0327
Droit .....	0398
Économie	
Généralités .....	0501
Commerce-Affaires .....	0505
Économie agricole .....	0503
Économie du travail .....	0510
Finances .....	0508
Histoire .....	0509
Théorie .....	0511
Études américaines .....	0323
Études canadiennes .....	0385
Études féministes .....	0453
Folklore .....	0358
Géographie .....	0366
Gérontologie .....	0351
Gestion des affaires	
Généralités .....	0310
Administration .....	0454
Banques .....	0770
Comptabilité .....	0272
Marketing .....	0338
Histoire	
Histoire générale .....	0578

Ancienne .....	0579
Médiévale .....	0581
Moderne .....	0582
Histoire des noirs .....	0328
Africaine .....	0331
Canadienne .....	0334
États-Unis .....	0337
Européenne .....	0335
Moyen-orientale .....	0333
Latino-américaine .....	0336
Asie, Australie et Océanie .....	0332
Histoire des sciences .....	0585
Loisirs .....	0814
Planification urbaine et régionale .....	0999
Science politique	
Généralités .....	0615
Administration publique .....	0617
Droit et relations internationales .....	0616
Sociologie	
Généralités .....	0626
Aide et bien-être social .....	0630
Criminologie et établissements pénitentiaires .....	0627
Démographie .....	0938
Études de l'individu et de la famille .....	0628
Études des relations interethniques et des relations raciales .....	0631
Structure et développement social .....	0700
Théorie et méthodes .....	0344
Travail et relations industrielles .....	0629
Transports .....	0709
Travail social .....	0452

### SCIENCES ET INGÉNIERIE

#### SCIENCES BIOLOGIQUES

Agriculture	
Généralités .....	0473
Agronomie .....	0285
Alimentation et technologie alimentaire .....	0359
Culture .....	0479
Élevage et alimentation .....	0475
Exploitation des péturages .....	0777
Pathologie animale .....	0476
Pathologie végétale .....	0480
Physiologie végétale .....	0817
Sylviculture et faune .....	0478
Technologie du bois .....	0746
Biologie	
Généralités .....	0306
Anatomie .....	0287
Biologie (Statistiques) .....	0308
Biologie moléculaire .....	0307
Botanique .....	0309
Cellule .....	0379
Ecologie .....	0329
Entomologie .....	0353
Génétique .....	0369
Limnologie .....	0793
Microbiologie .....	0410
Neurologie .....	0317
Océanographie .....	0416
Physiologie .....	0433
Radiation .....	0821
Science vétérinaire .....	0778
Zoologie .....	0472
Biophysique	
Généralités .....	0786
Médicale .....	0760

#### SCIENCES DE LA TERRE

Biogéochimie .....	0425
Géochimie .....	0996
Géodésie .....	0370
Géographie physique .....	0368

Géologie .....	0372
Géophysique .....	0373
Hydrologie .....	0388
Minéralogie .....	0411
Océanographie physique .....	0415
Paléobotanique .....	0345
Paléoécologie .....	0426
Paléontologie .....	0418
Paléozoologie .....	0985
Palynologie .....	0427

#### SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT

Économie domestique .....	0386
Sciences de l'environnement .....	0768
Sciences de la santé	
Généralités .....	0566
Administration des hôpitaux .....	0769
Alimentation et nutrition .....	0570
Audiologie .....	0300
Chimiothérapie .....	0992
Dentisterie .....	0567
Développement humain .....	0758
Enseignement .....	0350
Immunologie .....	0982
Loisirs .....	0575
Médecine du travail et thérapie .....	0354
Médecine et chirurgie .....	0564
Obstétrique et gynécologie .....	0380
Ophtalmologie .....	0381
Orthophonie .....	0460
Pathologie .....	0571
Pharmacie .....	0572
Pharmacologie .....	0419
Physiothérapie .....	0382
Radiologie .....	0574
Santé mentale .....	0347
Santé publique .....	0573
Soins infirmiers .....	0569
Toxicologie .....	0383

#### SCIENCES PHYSIQUES

Sciences Pures	
Chimie	
Généralités .....	0485
Biochimie .....	0487
Chimie agricole .....	0749
Chimie analytique .....	0486
Chimie minérale .....	0488
Chimie nucléaire .....	0738
Chimie organique .....	0490
Chimie pharmaceutique .....	0491
Physique .....	0494
Polymères .....	0495
Radiation .....	0754
Mathématiques .....	0405
Physique	
Généralités .....	0605
Acoustique .....	0986
Astronomie et astrophysique .....	0606
Électronique et électricité .....	0607
Fluides et plasma .....	0759
Météorologie .....	0608
Optique .....	0752
Particules (Physique nucléaire) .....	0798
Physique atomique .....	0748
Physique de l'état solide .....	0611
Physique moléculaire .....	0609
Physique nucléaire .....	0610
Radiation .....	0756
Statistiques .....	0463

#### Sciences Appliqués Et Technologie

Informatique .....	0984
Ingénierie	
Généralités .....	0537
Agriculture .....	0539
Automobile .....	0540

Biomédicale .....	0541
Chaleur et ther modynamique .....	0348
Conditionnement (Emballage) .....	0549
Génie aérospatial .....	0538
Génie chimique .....	0542
Génie civil .....	0543
Génie électronique et électrique .....	0544
Génie industriel .....	0546
Génie mécanique .....	0548
Génie nucléaire .....	0552
Ingénierie des systèmes .....	0790
Mécanique navale .....	0547
Métallurgie .....	0743
Science des matériaux .....	0794
Technique du pétrole .....	0765
Technique minière .....	0551
Techniques sanitaires et municipales .....	0554
Technologie hydraulique .....	0545
Mécanique appliquée .....	0346
Géotechnologie .....	0428
Matériaux plastiques (Technologie) .....	0795
Recherche opérationnelle .....	0796
Textiles et tissus (Technologie) .....	0794

#### PSYCHOLOGIE

Généralités .....	0621
Personnalité .....	0625
Psychobiologie .....	0349
Psychologie clinique .....	0622
Psychologie du comportement .....	0384
Psychologie du développement .....	0620
Psychologie expérimentale .....	0623
Psychologie industrielle .....	0624
Psychologie physiologique .....	0989
Psychologie sociale .....	0451
Psychométrie .....	0632



THE UNIVERSITY OF CALGARY  
FACULTY OF GRADUATE STUDIES

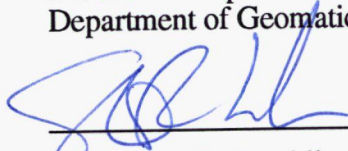
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "Hyperspectral Image Analysis: a Study on Band Selection Methods and the Sensitivity of Linear Spectral Unmixing" submitted by Airong Zhang in partial fulfillment of the requirements for the degree of Master of Science.



Supervisor, Dr. Peng Gong  
Department of Geomatics Engineering



Dr. M.A. Chapman  
Department of Geomatics Engineering



Dr. Steven E. Franklin  
Department of Geography



Dr. Deren Li  
Department of Photogrammetry and Remote Sensing  
Wuhan Technical University of Surveying and Mapping

Date: June 13, 1994

## ABSTRACT

An imaging spectrometer records spectral signatures with hundreds of spectral bands in very narrow intervals allowing subtle spectral difference among different materials to be detected. One major technique for quantitative analysis of imaging spectrometer data is linear spectral mixture modeling. In a linear mixing model, a number of scene components with relatively pure spectral properties, called endmembers, are often used. Two least squares methods, unconstrained and constrained, are used to invert the linear mixing model to extract quantitative information such as the proportion of each endmember in a pixel. Although hundreds of spectral bands are available for the purpose of spectral decomposition, only a small number of bands showing relatively large spectral separabilities among the endmembers need to be used. Therefore, the amount of computation can be reduced with the reduction of the number of bands. In decomposition analysis, the existence of noise may affect the accuracy of pixel decomposition results.

In this study, some simulations were made to test the sensitivity of the inversion of a linear mixing model. Atmospheric noise and the effect of undefined endmember were artificially generated. Five methods were tested for the purpose of selecting spectral bands for image decomposition. These included the successive correlation examination method, averaged correlation analysis, interband minimum difference analysis, the combination of minimum difference and successive correlation examination method, and the step-wise principal component analysis. Efficiencies of these methods were evaluated based on the spectral separabilities among endmembers in selected bands, the reference decomposition results, and the ground truth. The effect of the atmosphere on pixel decomposition using selected bands was also evaluated through simulation. Those band selection methods with the consideration of spectral separability and correlation as well as the step-wise principal component analysis method are effective for band reduction and pixel decomposition.

## ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Dr. Peng Gong for his continuous support, encouragement and constructive criticism throughout the course of my graduate studies.

All my colleagues are gratefully acknowledged for their serving the time in discussions and assistance. Special thanks are extended to Mr. Xiaoming Zheng and Mr. Gongchen Shi for their strong support in the field data acquisition.

My deepest thanks go to my husband, Penggang Zhang, for sharing with me the troubles and triumphs faced by a foreign student.

This research is partially supported by an NSERC research grant to Dr. P. Gong and an NSERC strategic grant to Drs. K.P. Schwarz, M.E. Cannon and M.A. Chapman.

## **Dedication**

*To my parents!*



## TABLE OF CONTENTS

APPROVAL PAGE.....	ii
ABSTRACT .....	iii
ACKNOWLEDGMENTS.....	iv
DEDICATION.....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES .....	ix
LIST OF FIGURES .....	x

CHAPTER	Page
<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 Research Background.....	1
1.2 Research Objectives.....	3
1.3 Thesis Outline and Contributions .....	4
<b>2 LINEAR MIXING MODELING .....</b>	<b>6</b>
2.1 Principles of Linear Mixing Model .....	6
2.2 Solutions to the Linear Spectral Mixing Problem.....	8
2.2.1 Singular value decomposition .....	10
2.2.2 Nonnegative least squares (NNLS) method.....	12
2.3 Applications of Linear Mixing Model .....	16
<b>3 DATA ACQUISITION AND ANALYSIS.....</b>	<b>19</b>
3.1 Compact Airborne Spectrographic Imager (CASI).....	19
3.2 Filed Data Acquisition .....	22
3.3 Characteristics of Hyperspectral Imagery .....	25
3.4 Noise Behaviour .....	26

<b>4</b>	<b>BAND SELECTION.....</b>	<b>28</b>
4.1	Band Selection Based on the Successive Correlation Examination .....	29
4.2	Averaged Correlation Analysis .....	33
4.3	Interband Minimum Difference Analysis .....	36
4.4	Combining the Minimum Difference and Successive Correlation Methods.....	37
4.5	Step-Wise Principal Component Analysis .....	38
<b>5</b>	<b>IMPLEMENTATION AND EXPERIMENTS .....</b>	<b>41</b>
5.1	Data Preprocessing.....	41
5.1.1	CASI data .....	41
5.1.2	Field Measurements.....	43
5.2	Experiments.....	44
5.2.1	The sensitivity of linear mixing model.....	46
5.2.1.1	Undefined endmember inside study area.....	46
5.2.1.2	Atmospheric effect .....	47
5.2.2	Evaluation of band selection methods .....	49
5.2.2.1	The first evaluation criterion .....	51
5.2.2.2	The second evaluation criterion .....	55
5.2.2.3	The third evaluation criterion .....	57
5.2.2.4	Result evaluation based on RMS .....	60
5.2.3	The sensitivities of band selection methods to atmospheric noise .....	64
5.3	Summary .....	67
<b>6</b>	<b>CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>69</b>
6.1	Conclusions.....	69
6.2	Recommendations.....	71
	<b>REFERENCES .....</b>	<b>72</b>

## LIST OF TABLES

No.	Page
3.1 Atmospheric scattering models for different atmospheric conditions .....	27
5.1 Decomposition results with an undefined endmember.....	46
5.2 Decomposition results from data contaminated by atmospheric effect.....	49
5.3 The decomposition results obtained through SVD and NNLS procedure from field measurements.....	59
5.4 The atmospheric effect on linear decomposition by using SVD with 50 selected bands.....	66
5.5 The atmospheric effect on linear decomposition by using NNLS with 50 selected bands.....	67

## LIST OF FIGURES

No.	Page
2.1	Algorithm of nonnegative least squares (NNLS)..... 15
3.1	Operation of CASI ..... 20
3.2	An example of CASI data from 419 nm to 789 nm ..... 22
3.3	Measurement devices ..... 24
3.4	An example of field spectrometer data from 334 nm to 1062 nm..... 25
4.1	Correlation between two successive bands of CASI data ..... 30
4.2	Lags along a transect of spectral bands ..... 31
4.3	The mean correlation of all bands with different lags..... 32
4.4	Row-wise (or column-wise) sum of correlation for each band ..... 33
4.5	Histogram of selected bands using the empirical formula with different coefficients ..... 36
4.6	The distribution of minimum difference in each band..... 37
4.7	Sequence of original band reduction by MSPCA..... 40
5.1	Smoothed CASI reflectance curves for six endmembers using a 1x7 filter ..... 42
5.2	Smoothed ASD spectrometer reflectance curves for four endmembers using a 1x7 filter..... 44
5.3	Experimental design..... 45
5.4	Simulated Rayleigh scattering added onto the artificially created spectra..... 48
5.5(a)	Separability for 20 selected bands obtained by the successive correlation examination method (the first method)..... 52
5.5(b)	Separability for 20 selected bands obtained by the averaged correlation method (the second method)..... 53
5.5(c)	Separability for 20 selected bands obtained by the interband minimum difference analysis method (the third method)..... 53

5.5(d) Separability for 20 selected bands obtained by combining the minimum difference and successive correlation method (the fourth method).....	54
5.5(e) Separability for 20 selected bands obtained with the step-wise principal components method (the fifth method) .....	54
5.6(a) The correlation between the results obtained from all the bands and from 50 selected bands using the SVD method .....	56
5.6(b) The correlation between the results obtained from all the bands and from 50 selected bands using the NNLS method.....	57
5.7(a) The RMSs of the decomposition results obtained from all the bands and from 20 selected bands using the SVD method .....	61
5.7(b) The RMSs of the decomposition results obtained from all the bands and from 50 selected bands using the SVD method .....	62
5.7(c) The RMSs of the decomposition results obtained from all the bands and from 20 selected bands using the NNLS method.....	63
5.7(d) The RMSs of the decomposition results obtained from all the bands and from 50 selected bands using the NNLS method.....	64

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 Research Background**

A widely used method for extracting information from remotely sensed data is image classification. In image classification, each image pixel is usually assumed to belong to one pure class. The result is desirable when the amount of boundary pixels is small and the scene is nicely partitioned into regions of homogeneous cover types. Many agricultural and forestry areas satisfy this requirement when the resolution and classification scheme matches. However, a lot of areas do not satisfy this requirement. Classification may not be suitable because of the existence of a large amount of mixed pixels. In this case, it only can provide an estimate of various classes for the study area. When the size of image pixel is comparable to, or much larger than the natural size of ground cover unit, each image pixel may contain more than one ground component. This is the common situation in remotely sensed data.

Unlike laboratory spectral reflectances which are usually measured from pure materials, a large portion of remotely sensed data are spectrally mixed [Settle et al, 1993]. This is due to the fact that the scales of spatial variation of natural phenomena are often smaller than the spatial resolutions of the sensors. Spectral reflectances from different materials within each instantaneous field of view (IFOV) of the sensor are recorded as one spectral response for each band. This causes a problem known as spectral mixing. Each

observation is spectrally mixed from several different materials in each band. In this case, traditional classification is far from an ideal tool for information extraction.

A more crucial objective for estimating different components inside the study area is the extraction of quantitative information from each image pixel. This is the inverse process of spectral mixing, which is called spectral unmixing or decomposition analysis. The proportions of different materials contributing to one image pixel, which are called endmembers, are expected to be obtained. The imaging spectrometer, as a new tool, provides the opportunity for us to capture very narrow spectral features of various targets. Unlike traditional multispectral sensors such as Landsat Multispectral Scanner and Thematic Mapper, which have only a few relatively wide spectral bands, imaging spectrometers record spectral signatures with hundreds of spectral bands, such as the Compact Airborne Spectrographic Imager (casi) has 288 bands in approximately 1.8 nm spectral intervals. An Analytical Spectral Devices field spectrometer used in this project has 512 bands in approximately 1.4 nm spectral intervals. High spectral resolutions of imaging spectrometers allow direct identification of ground targets and quantitative analysis of subtle spectral changes down to sub-pixel level.

One major technique for quantitative analysis of imaging spectrometer data is the spectral mixing modeling. Two types of models have been developed to describe the spectral mixing process: macroscopic spectral mixing [Gong et al, 1991; Adams et al, 1989; Boardman, 1989; Singer et al, 1979] and intimate spectral mixing [Mustard et al, 1989; 1987; Shipman et al, 1987; Johnson et al, 1983; Hapke, 1981]. A macroscopic mixture is characterized by large homogeneous patches of different materials where electromagnetic energy reflected from a single material dominates. In contrast, intimate spectral mixing results from electromagnetic energy being multiply transmitted and scattered by more than one material. Spectra of macroscopic mixtures are linear

combinations of reflectance spectra from different materials, whereas intimate mixtures are nonlinear spectral mixing. Both linear and nonlinear spectral mixing models are simple tools used to describe spectral mixing processes. It has been reported that nonlinear spectral mixing can be linearized by converting reflectance to single scattering albedo [Johnson et al, 1983].

The basic physical assumption underlying the linear mixing model is that there is no significant amount of multiple scattering among the different cover types. Each photon that reaches the sensor has interacted with just one cover type. Under these conditions the received energy can be considered as a simple sum of the energy received from each cover component. In a real situation, the reflectance from neighboring areas contributes to the received signal. If the material near the target was not defined in the reflectance matrix, what will happen during the decomposition analysis? On the other hand, the reflectance matrix is generally obtained inside the laboratory. However, the data for analysis is obtained from remotely sensed imagery. Atmospheric effects exist in each band of the remotely sensed data. How does it affect the results obtained through linear decomposition analysis? For hyperspectral image, hundreds of narrow bands have very high correlation which provides redundant information. How can we reduce the redundancy and improve the computational efficiency without losing decomposition accuracy? These are the problems which are discussed in this study.

## **1.2 Research Objectives**

In this study, linear mixing modeling will be adopted as the tool to decompose hyperspectral data for the estimation of spatial proportions or abundance of various component materials from their composites. It has only been used to determine the spatial proportions of known endmembers. The primary objectives are:



- to evaluate the sensitivity of linear unmixing analysis in situations where undefined endmember exists in the study area;
- to simulate and evaluate the effect of atmosphere on the linear unmixing analysis;
- to develop band selection methods for hyperspectral data analysis;
- to assess the effectiveness of the band selection methods in linear spectral unmixing analysis.

### **1.3 Thesis Outline and Contributions**

In Chapter 2, the fundamental concepts of the linear mixing modeling are first reviewed. Some basic definitions, such as endmember, reflectance matrix, observations, are introduced. The unknowns in the linear mixing process are the endmember proportions. Two methods for finding the unknown proportions are presented.

Chapter 3 describes the characteristics of the data sets used in this thesis. Data acquisition methods are introduced. Two data sets, one extracted from a casi image and the other obtained from field measurements, have been used and analyzed.

In Chapter 4, five band selection methods are developed and implemented. Threshold determination for each method is described.

Chapter 5 introduces the experiment designed for two purposes of this project. One is to test the consistency of the linear mixing model either when undefined endmember exists in the study area, or when atmospheric effect is considered. Simulated data is used for these tests. The other is to evaluate results obtained through different band selection methods. The correlation of extracted proportions with reference proportions or with

ground truth, and the root-mean-squares (RMS) error obtained are used to evaluate the band selection effectiveness. The atmospheric effect on the decomposition results obtained using selected bands through different band selection methods are also presented.

The last chapter, Chapter 6, gives conclusions of this research work, and presents recommendations for future research.

The primary contributions of this thesis are as following:

1. For the first time, the sensitivities of linear unmixing analysis to an undefined endmember inside a study area and to the atmospheric noise in the observations have been examined.
2. Five band selection methods have been developed. Some of these methods can be used to select bands automatically. The results obtained using selected bands through most of these methods are consistent no matter whether the atmospheric noise is included or not. This consistency implies that the effect of atmosphere on the results obtained with selected bands is the same as those obtained when all the spectral bands are used.

## **CHAPTER 2**

### **LINEAR MIXTURE MODELING**

#### **2.1 Principles of Linear Mixing Model**

In laboratory spectral measurement, a small area with pure materials can be measured under controlled conditions. For remotely sensed data, however, different components may be detected in one measurement. This is due to the fact that the natural phenomena vary spatially. The spatial resolutions of sensors are generally lower than the scale of natural variation. For each band, the observed spectral reflectance is combined by those of different materials within an instantaneous field of view (IFOV) of the sensor.

Accurate estimation of spatial components and spectral properties of various materials from remotely sensed data is of great importance to many research and application fields. In order to identify various original materials and to determine their spatial proportions from remotely sensed data, the spectral mixing process has to be properly modeled.

The most popular model is the linear spectral mixing modeling. In a linear spectral mixing model, we assume that only a small number of pure materials mix together with various areal proportions in the observed spectra. These materials are called endmembers, components, or factors.

### Linear mixing model:

Suppose there are  $m$  bands in a remotely sensed data set, and  $p$  endmembers.  $r_{jk}$  represents the spectral reflectance of  $k$ th endmember at  $j$ th band. All the reflectance can be arranged in an  $m \times p$  matrix  $\mathbf{R}$  as follows:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mp} \end{bmatrix}$$

The linear mixture model can be expressed as:

$$\begin{bmatrix} d_{i1} \\ \vdots \\ d_{im} \end{bmatrix} = \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mp} \end{bmatrix} \begin{bmatrix} f_{i1} \\ \vdots \\ f_{ip} \end{bmatrix} \quad (2.1)$$

$$\text{or} \quad \mathbf{d}_i = \mathbf{R} \cdot \mathbf{f}_i$$

with the following constraints:

$$f_{ik} \geq 0 \quad \text{and} \quad \sum_{k=1}^p f_{ik} = 1, \text{ for } i=1,2,\dots,n, \quad (2.2)$$

where  $n$  is the total number of image pixels.  $\mathbf{d}_i$  is the spectral responses measured from the  $i$ th pixel which is recorded by sensors in  $m$  bands.  $f_{ik}$  denotes the fractional area of the  $k$ th endmember in the  $i$ th pixel. All the fractions of endmembers in pixel  $i$  compose a  $p \times 1$  vector  $\mathbf{f}_i$  and they should sum to one.

For the decomposition of a mixed pixel, three mixture parameters in Equation (2.1) are of interest:

- (a) the total number of endmembers in the mixture pixel,  $p$ ;
- (b) the spectral identity of endmember  $k$  in  $m$  bands,  $r_{ik}$ ,  $i = 1, 2, \dots, m$ ;
- (c) the proportion of each endmember in the pixel,  $f_{ik}$ .

The solution to Equation (2.1) is limited by the knowledge of these parameters. If parameters (a) and (b) are known, it is possible to determine (c) pixel by pixel. Actually, this is a typical situation used to derive spatial proportions of various endmembers from remotely sensed data. It is easy to get the proportional solution by inverting matrix  $\mathbf{R}$ . If (a) and (c) are known, (b) can be obtained from  $\mathbf{d}_i$  and  $\mathbf{f}_i$  [Hanan, et al, 1991]. This method is applicable to situations where available ground measurements of  $\mathbf{f}_i$  are used to derive  $\mathbf{R}$ . There are circumstances when only (a) is known, or none of the three types of parameters is known. However, it is still possible to estimate these mixture parameters from a simultaneous analysis of a number of pixels rather than by analysis of each sample individually. The mixing proportions should vary from one pixel to the other. Obviously, in order to determine the solution from Equation (2.1), enough information has to be obtained to define some parameters. A number of methods have been proposed to achieve the expected solution by using principal component analysis [Smith et al, 1985], factor analysis [Klovan, 1975], and the use of mathematical programming [Liang et al, 1991].

In this research, the identity spectral matrix  $\mathbf{R}$  is known from either a Compact Airborne Spectrographic Imagery (casi) or a field spectrometer. Therefore, the expected solution to Equation(2.1) is to derive the proportion vector  $\mathbf{f}_i$  in a pixel by pixel manner.

## 2.2 Solutions to the Linear Spectral Problem

When the number and the spectral identities of endmembers are known, there are

three possible cases existing to extract proportions of each endmember: (a)  $m+1 < p$ , (b)  $m+1 = p$ , (c)  $m+1 > p$ . For case (b), it is a simple case to solve Equation(2.1) with the constraint Equation(2.2). When remotely sensed data were acquired from sensors with limited spectral bands, such as Landsat Multispectral Scanner, Thematic Mapper, and SPOT High Resolution Visible, the number of desired endmembers may be more than the number of spectral bands plus one which is in case (a). Under such circumstances, Equation(2.1) becomes undetermined. In order to obtain the spatial proportions for each image pixel, an optimal solution has to be explored. Some researchers recover the proportions  $f$  using the geometrical structure of data scattering in multispectral space in combination with evidence such as tasseled cap behavior of vegetation on a red-infrared scatter plot [Jasinski et al, 1989; 1990].

For the data from the imaging spectrometer containing hundreds of bands for each image pixel, Equation (2.1) is almost always overdetermined. That means the number of endmember  $m$  plus one is smaller than the number of bands. This is case (c).

Combining Equation (2.2) with Equation (2.1), the following overdetermined Equation (2.3) can be solved by using the least squares method.

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mp} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \bullet \begin{bmatrix} f_{i1} \\ f_{i2} \\ \vdots \\ f_{ip} \end{bmatrix} = \begin{bmatrix} d_{i1} \\ d_{i2} \\ \vdots \\ d_{ip} \\ 1 \end{bmatrix} \quad (2.3)$$

in a vector form:  $\begin{pmatrix} \mathbf{R} \\ 1 \end{pmatrix} \bullet \mathbf{f} = \begin{pmatrix} \mathbf{d}_i \\ 1 \end{pmatrix}$

The constraints to the solution are  $f_{ik} \geq 0$  (2.4)

There are a number of methods which may be used to solve Equation (2.3). The

commonly used one is the direct inversion of this equation through

$$\mathbf{f} = \left( \begin{bmatrix} \mathbf{R}^T & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{R}^T & \mathbf{1} \end{bmatrix} \bullet \mathbf{d} \quad (2.5)$$

If the singular results are caused by the direct inversion, other methods are recommended for solving Equation(2.3). Two other algorithms are the unconstrained singular value decomposition method and the constrained nonnegative least squares method. In the later method, we obtain the solution of proportions  $\mathbf{f}_i$  by taking Equation(2.4) into consideration.

### 2.2.1 Singular value decomposition

A Singular Value Decomposition (SVD) algorithm is recommended to deal with sets of equations or matrices that are either singular or numerically very close to singular such as Equation (2.3) [Boardman, 1989; Press et al, 1992]. With this method, however, the constraint of nonnegative proportions is not considered.

Based on the theorem of linear algebra, any  $m \times n$  matrix  $\mathbf{A}$ , whose number of rows  $m$  is greater than or equal to its number of columns  $n$ , can be written as the product of an  $m \times m$  column-orthogonal matrix  $\mathbf{U}$ , an  $m \times n$  diagonal matrix  $\mathbf{W}$  with positive or zero elements, and the transpose of an  $n \times n$  orthogonal matrix  $\mathbf{V}$ . This can be represented in the following form:

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^T \quad (2.6)$$

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}_m \quad \mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}_n \quad (2.7)$$

Because  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrix, for square matrix, their inverses are equal to their transposes. Because  $\mathbf{W}$  is diagonal, its inverse is the diagonal matrix whose

elements are the reciprocals of the element  $w_j$ . Therefore, the inverse of  $\mathbf{A}$  is

$$\mathbf{A}^+ = \mathbf{V} [\text{diag}(1/w_j)] \mathbf{U}^T \quad (2.8)$$

Where  $\mathbf{A} = \begin{bmatrix} \mathbf{R}^T & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ 1 \end{bmatrix}$ . If more than one of the  $w_j$  have a zero value or very small values

which are dominated by roundoff error, a solution may not be obtained at all. In order to obtain one particular number of solution set  $\mathbf{f}$  for the linear equation

$$\mathbf{R} \cdot \mathbf{f} = \mathbf{b} \quad (2.9)$$

Where  $\mathbf{b} = [\mathbf{R}^T \mathbf{1}] \cdot \mathbf{d}$ .  $\mathbf{f}$  has to be found so that it minimizes the norm of the residual of the solution  $|\mathbf{R} \cdot \mathbf{f} - \mathbf{b}|$ . It has been proven [Press et al, 1988] that the solution of  $\mathbf{f}$  can be obtained by simply replacing  $1/w_j$  by zero if  $w_j$  is zero or close to zero. Zeroing a singular value corresponds to throwing away one linear combination of the set of equations that is being solved. It may seem paradoxical to reduce the number of equations to make the linear system determined [Press et al, 1988]. This also means that precise combination of equations, which is not so reliable by roundoff error as to be at best useless, are thrown away. Then the solution of Equation (2.9) which is the proportion in Equation (2.3) is

$$\mathbf{f} = \mathbf{V} [\text{diag} (1/w_j)] (\mathbf{U}^T \mathbf{b}) \quad (2.10)$$

When using the SVD algorithm, a threshold has to be specified for deciding how small  $w_j$  may be before it is treated as zero. This threshold is determined based on experiments.

When  $\mathbf{f}$  is obtained, the appropriateness of the least squares estimation of  $\mathbf{f}$  can be judged by the root-mean-squared error (RMS),

$$\text{RMS} = \sqrt{\frac{1}{m} \sum_{j=1}^m (d_j - \sum_{k=1}^p r_{jk} \cdot f_k)^2} \quad (2.11)$$



However, solutions of  $\mathbf{f}$  in Equation(2.10) do not guarantee  $0 \leq f_k \leq 1$ . When situations such as  $f_k > 1$  or  $f_k < 0$  happen, there are three things that need to be considered [Adams, et al, 1989]. First, the image endmembers should be examined to make sure that each image endmember is exactly what it is supposed to be. If image endmembers are appropriate and solutions like  $f_k > 1$  or  $f_k < 0$  still occur, then second, ignore those  $f_k$ 's close to 0 or 1 which may be caused by roundoff errors during computer processing. Thirdly, those  $f_k$ 's significantly smaller than zero or larger than one imply that there are other possible image endmembers that have not been identified. When the RMS is high for certain pixels, it suggests that either the linear mixing model is inappropriate or a new image endmember has not been identified.

### 2.2.2 Nonnegative least squares (NNLS) method

Since Equation (2.3) is a set of linear functions with nonnegative constraint Equation (2.4), the proportions  $f_i$  can be solved through a nonnegative least squares strategy.

For a general least squares problem with linear inequality constraints (LSI), the problem is defined to minimize  $\| \mathbf{R}\mathbf{f} - \mathbf{d} \|$  subject to  $\mathbf{G}\hat{\mathbf{f}} \geq \mathbf{h}$ , where  $\mathbf{h}$  is the constraint of  $\mathbf{f}$ . The following theorem characterizes the solution vector of the LSI problem [Lawson, 1974].

#### THEOREM

An  $n$ -vector  $\hat{\mathbf{f}}$  is a solution for problem LSI if and only if there exists an  $m$ -vector  $\hat{\mathbf{y}}$  and a partitioning of the integers 1 through  $m$  into subsets  $\mathcal{E}$  and  $\mathcal{\delta}$  such that

$$\mathbf{G}^T \hat{\mathbf{y}} = \mathbf{R}^T (\mathbf{R} \hat{\mathbf{f}} - \mathbf{d}) \quad (2.12)$$

$$\hat{y}_i \geq 0 \text{ for } i \in \mathcal{E}, \quad \hat{y}_i = 0 \text{ for } i \in \mathcal{\delta}$$

Let  $g_i^T$  denotes the  $i$ th row vector of the matrix  $G$ . The  $i$ th constraint,  $g_i^T \hat{f} \geq h_i$ , defines a feasible half-space,  $\{f: g_i^T f \geq h_i\}$ . The vector  $g_i$  is orthogonal (normal) to the bounding hyperplane of this halfspace and is directed into the feasible halfspace. The point  $\hat{f}$  is interior to the halfspaces indexed in  $\mathcal{\delta}$  and on the boundary of the halfspaces indexed in  $\mathcal{E}$ . The vector

$$p = R^T(R\hat{f} - d) \quad (2.13)$$

is the gradient vector of  $\varphi(f) = \frac{1}{2} \|Rf - d\|^2$  at  $f = \hat{f}$ . Since  $\hat{y}_i = 0$  for  $i \notin \mathcal{E}$ , Equation (2.12) can be written as

$$\sum_{i \in \mathcal{E}} \hat{y}_i (-g_i) = -p \quad (2.14)$$

which states that the negative gradient vector of  $\varphi$  at  $\hat{f}$  is expressible as a nonnegative ( $\hat{y}_i \geq 0$ ) linear combination of outward-pointing normal ( $-g_i$ ) to the constraint hyperplanes on which  $\hat{f}$  lies ( $i \in \mathcal{E}$ ). Geometrically, this means that the negative gradient vector  $-p$  lies in the convex cone based at the point  $\hat{f}$  and is generated by the outward-pointing normal  $-g_i$ .

The nonnegative least squares method is a special case of least squares problem with linear inequality constraints on the solution. This problem is defined as

$$\text{Minimize } \|Rf - d\| \text{ subject to } 0 \leq f \leq 1. \quad (2.15)$$

Figure 2.1 shows the algorithm for solving NNLS. In this algorithm,  $R$  is an  $(m+1) \times n$  matrix and  $f$  is a  $(m+1)$ -vector. The  $n$ -vectors  $w$  and  $z$  provide working space. Two index sets  $p$  and  $Z$  are defined and modified in the course of the execution of the algorithm. The set  $Z$  identifies the components of the current vector  $f$  that are zero. The

components of  $f$  indexed in  $\rho$  are positive. The index  $t$  selects a coefficient not presently in set  $\rho$  that will introduce the positive solution. This coefficient is saved in the tentative solution vector  $z$ . If all other components of  $z$  indexed in set  $\rho$  remain positive, then the components of  $z$  are transferred to the vector  $f$  and returns to the beginning. In this process, set  $\rho$  is augmented and set  $Z$  is diminished by the transfer of the index  $t$ . This sequence of events simply repeats with the addition of one more positive coefficient on iterations until  $Z$  is empty or  $W_j \leq 0$  for all  $j \in Z$ . However, if some coefficient indexed in set  $\rho$  becomes zero or negative in the vector  $z$ , then the algorithm remains performing a move that replaces  $x$  by  $f + \alpha(z - x)$ ,  $0 < \alpha \leq 1$ .  $\alpha$  should be chosen as large as possible subject to keeping the new  $f$  nonnegative. This iteration is repeated until it eventually satisfies that all the components in  $z$  are positive.

Upon termination, the solution vector  $f$  satisfies  $f_j > 0$ ,  $j \in \rho$ , and  $f_j = 0$ ,  $j \in Z$  and is a solution vector for the least squares problem  $R_p f \equiv d$ .

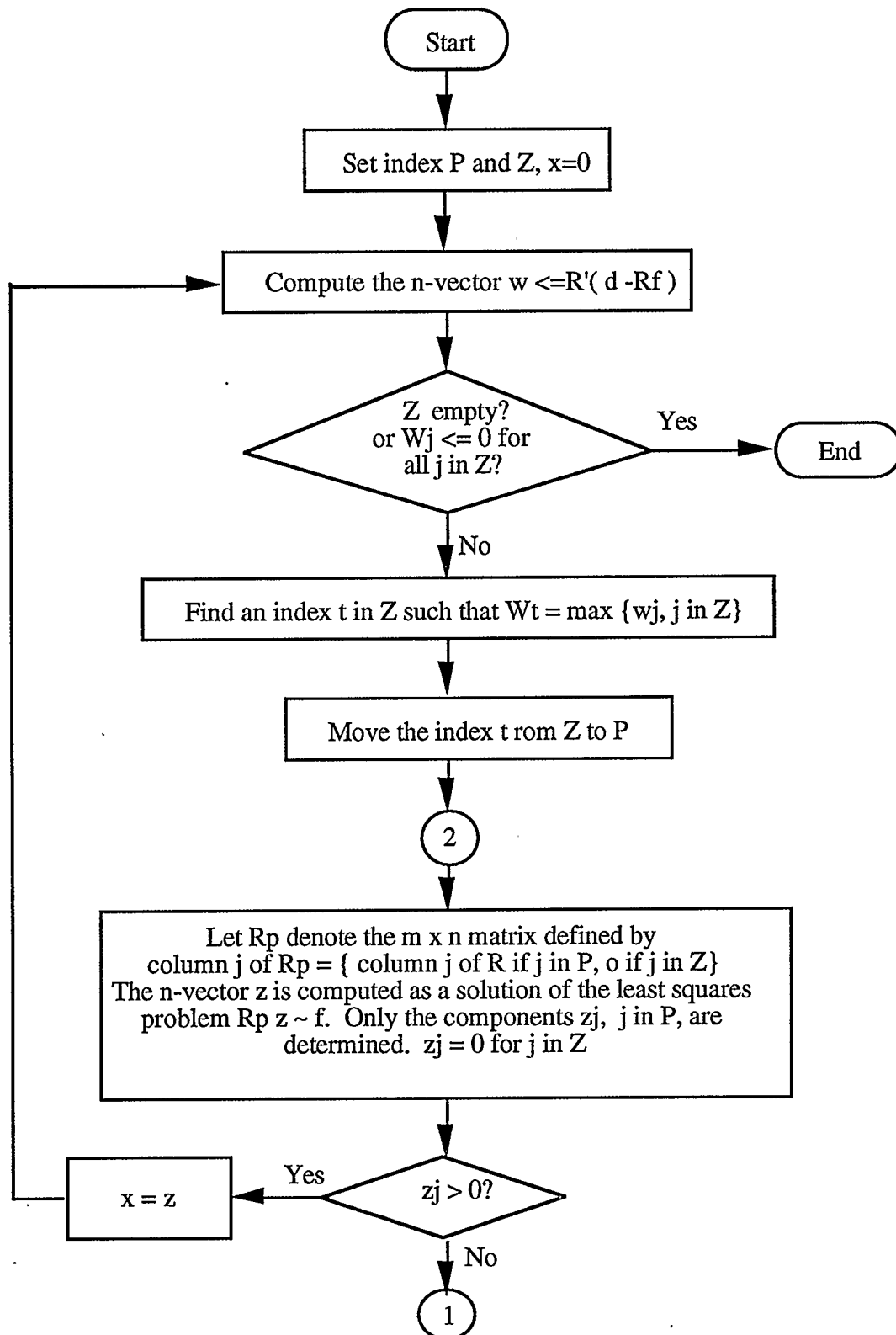


Figure 2.1 Algorithm of nonnegative least squares (NNLS)

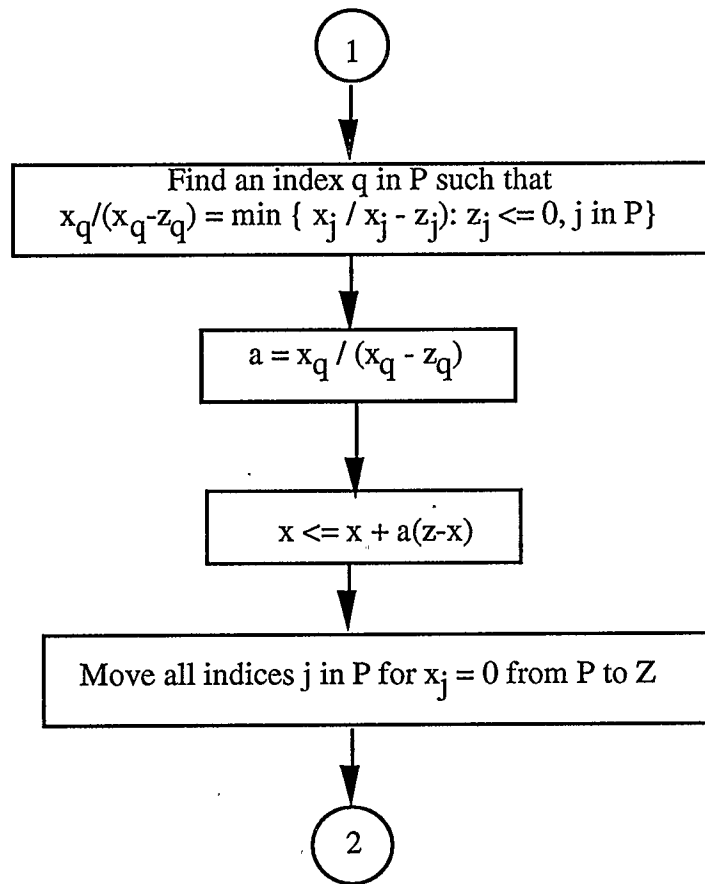


Figure 2.1 (Continued) Algorithm of nonnegative least squares (NNLS)

### 2.3 Applications of Linear Mixing Model

Techniques for spectral mixing analysis have been developed for more than twenty years in a number of disciplines, particularly in geology and chemistry. A lot of applications of this model have been implemented in remote sensing. Determination of mineral types and abundance from mixtures of minerals is an obvious example in geological and planetary studies [Bierwirth, 1990]. Smith et al [1990] applied these techniques to Landsat Thematic Mapper (TM) data acquired over a desert area in Owens Valley, California. In their study area, individual shrubs distribute discretely. They

derived seasonal changes of scrub abundance and analyzed the relationships between these changes and some environmental factors. Blunt et al [1990] used linear mixing technique to study regional variations of sand composition in Mexico with Landsat TM data. Duncan et al [1990] employed linear mixing technique to enhance differences between rock types in TM images for structural mapping in Saudi Arabia. Lithological image endmembers were selected for the linear mixing analysis. The same technique has been used in mapping the ophiolitic melanges of the Central Eastern Desert of Egypt with Landsat TM imagery [Rivard, 1989]. Adams et al [1986] applied these techniques to multispectral data obtained from Mars by Viking Lander 1 with 6 spectral bands. Four image endmembers were defined: shade with zero digital numbers, secondary illuminating effects, rock and soil. By isolating the shade and the secondary illumination effects, they found the rock and soil having close analogs in a library of laboratory spectra. Sabol et al [1990] studied material detectability using linear spectral mixing modeling by taking spectral contrast of materials, limitations of imaging systems, illumination conditions into consideration. They concluded that the contrast and sensor spectral resolution are important factors in material detection. Roberts et al [1990] studied the applicability of linear mixing analysis for leaf identification using a leaf radiative transfer model. They reported that at 450 to 650 nm, where leaves are relatively opaque, leaf spectra mix linearly. This implies that linear spectral mixing analysis at the scale of vegetation leaves, shorter wavelength may provide better results. Additional examples can be found in applications of this method in urban and forest environments [Cross et al, 1991; Gong et al, 1991].

There has been few research results published on linear mixing analysis of imaging spectrometer data. Boardman [1990] conducted linear spectral mixing analysis using Geophysical and Environmental Research Imaging Spectrometer data (63 channels ranging from 0.4 to 2.5 micrometers). He introduced an unconstrained version and a constrained

version of the least squares solution put forward by Lawson and Hanson [1974]. Because noise exist in the spectrometer data, the sensitivity of unconstrained method in linear decomposition analysis has been tested by Zhang et al. [1993] using simulated white noise.

## **CHAPTER 3**

### **DATA ACQUISITION AND ANALYSIS**

Two sets of data were used in this research. One was the Compact Airborne Spectrographic Imager (casi) data which were acquired on board a small aircraft. The other was the field spectral data which were obtained using Analytical Spectral Device's (ASD) high spectral resolution Personal Spectroradiometer.

#### **3.1 Compact Airborne Spectrographic Imager (CAST)**

casi is an airborne pushbroom sensor sensitive in the visible and near-infrared portion of the spectrum. A line perpendicular to the aircraft flight path is imaged along one dimension with the two-dimensional sensors (CCD). 512 pixels of spatial resolution across the flight path is obtained. The spectrum is dispersed along the flight path. This is analogous to having an array of spectrographs simultaneously imaging adjacent points across the line in the scene. Figure 3.1 shows the operation of casi operation.

Casi can be used onboard aircraft or in the laboratory. In order to achieve acceptable ground resolution during aircraft operation, data are normally collected in one of two modes, the spatial mode and the spectral mode. The spatial mode digitizes and records up to 19 spectral bands of information while maintaining full spatial resolution of 612 pixels of which the imaging field of view comprises 512 pixels. Each band is summed on-chip and may include any number of adjacent, non-overlapping spectral rows [Babey et al, 1992].



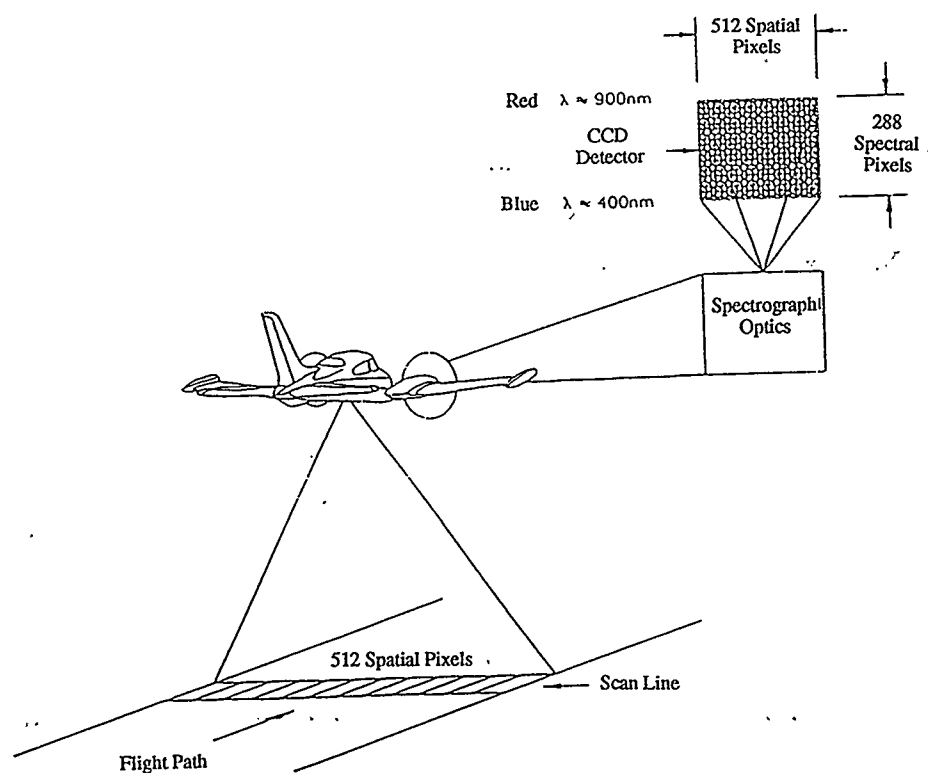


Figure 3.1 Operation of CASI [Babey et al, 1992]

In spectral mode, full spectral resolution is maintained in which 288 elements encompassing the wavelength range specification of 417 nm to 917nm at 1.8 nm spacing. This kind of high spectral resolution reduced the spatial resolution due to data logging limitations. In this mode, each spectral band contains 40 CCD detectors acquiring different view directions and covering the whole scene [Staenz, 1992]. Recently, a third mode (full-frame mode) has been added to the casi instrument software. This mode digitizes and records the entire data frame of 612 by 288 pixels. The recording of this amount of data

takes approximately 2s/frame [Babey, et al, 1992].

The original airborne casi data have to be calibrated in order to obtain reliable radiance with high spectral and spatial resolution. The data set used in this study were calibrated using the casi manufacturers parameters which reduced the noise caused by system design. There were still other noise sources. Two major kinds of noise were the atmospheric scattering and the radiometric error. In order to reduce the effects of these factors, the atmospheric effects may be suppressed using radiance to reflectance conversion either based on the ground truth data [Freemantle et al, 1992], or using radiative transfer model [Williams et al, 1992]. The radiance data used in our study were converted into reflectance based on the data obtained from field surveying. The radiometric calibration can be performed using a single radiance standard optical source to characterize the radiance sensitivity of the CCD and to indicate the wavelength of the sensor [Babey et al, 1992].

In this study, casi data were acquired during the period of May, 1991 in the middle of Oregon. The radiance values at wavelengths longer than 790 nm, in which the radiometric effects are stronger, were not used. Six endmembers are selected based on the ground truth investigated in the study area. Gravel at three sites are treated as three kinds of endmembers. Gravel\_1 is the relative pure gravel. Gravel\_2 and Gravel\_3 have some tuft grass cover. Two types of grass are tuft grass and snow grass. The other one is ponderosa pine. Figure 3.2 is the example of reflectance spectra extracted from the casi image.

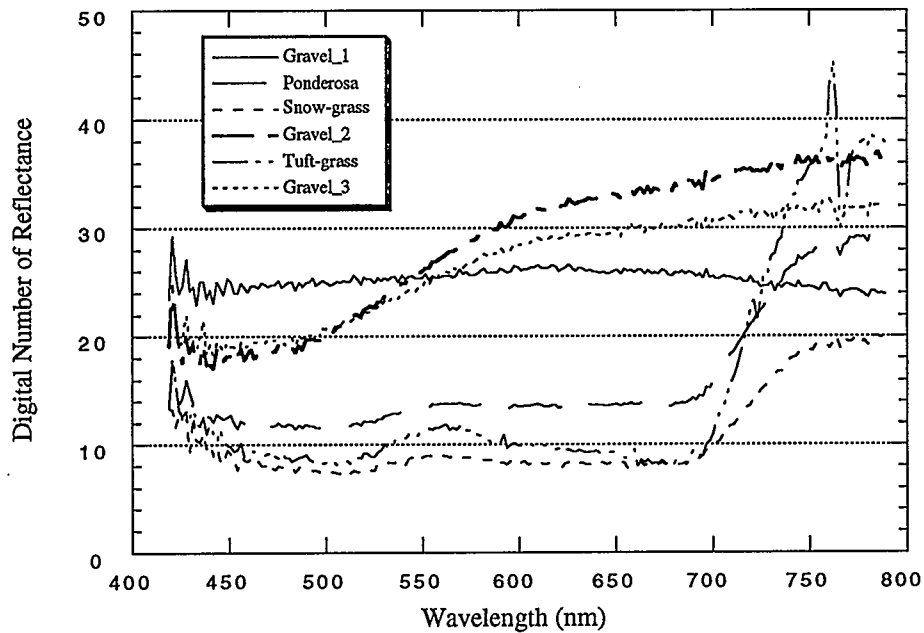


Figure 3.2 An example of CASI data from 419 nm to 789 nm

### 3.2 Field Data Acquisition

An ASD high spectral resolution Personal Spectroradiometer II was used to take measurements in the field. This spectrometer contains a 100 mm focal length, holographic grating spectrometer designed to collect light from an external source through a bundle of 19 optical fibers. The captured light travel through the fibers and strikes a grating which diffracts the light into its respective wavelengths. The diffracted light falls on the surface of a silicon photo diode array detector. The detector contains 512 elements, each of which records the number of photons striking on it. The number of photon are accumulated in each detector element and are recorded by incrementing counters which exist in each element. The number of photons striking over a period of time is converted into a voltage

to represent the amount of radiance of the target. In the silicon diode array, there is an undercurrent existing even if no photons strikes the array. This is referred to as the dark current. Any measured current is a combination of the signal current and the dark current. To eliminate the dark current effect, a mechanical shutter is built in the spectrometer to differentiate the dark current from signal current. When the shutter is closed only dark current is measured. When a signal is measured, the dark current should be subtracted from the measured current to produce the raw spectrum. The maximum number of photons accumulated by the array is 4095. To avoid achieving this saturation, the integration time for collecting photon should be adjusted based on the light conditions. The integration time is suitable when signal current has the dominant proportion in the measurement while it is not in saturated. It could be a shorter time in sunshine day, and longer in the laboratory. Whenever the integration time is adjusted, the dark current needs to be detected again. A reflectance spectra can be generated using a white panel whose radiance is assumed as all the incident light reaching to the target. The radiance of the target divided by the radiance of the white reference is the reflectance spectra of the target. This suppresses the effect of atmospheric distortion.

During the field measurements, the spectrometer was mounted on a survey tripod. The sensor gun was pointed vertically to the ground target. A tube with a field of view of  $18^\circ$  was screwed on the sensor in order to achieve accurate estimation of spatial resolution. A plumbing device was mounted along the tube. It was used to assure that the direction was vertical and also to locate the target. The distance between the sensor and the ground target was set at approximately 120 cm. The position of the sensor and the field of view determined the spatial resolution which was approximately 38 cm in diameter. The configuration of these devices is shown in the Figure 3.3.

In our field measurement, the atmospheric distortion on the reflectance was

ignored. The reflectance values with stronger radiometric effect at the shorter or longer wavelength regions since the weak response of the detectors were not used. Four types of materials, concrete road, grass, soil, and aspen leaf, were measured to obtain the reflectances for identification. Some mixing spectrum were taken from the mixed area combined by any two of these materials. Figure 3.4 is an example of the reflectance spectra measured by the ASD filed spectrometer. Four endmembers are represented.

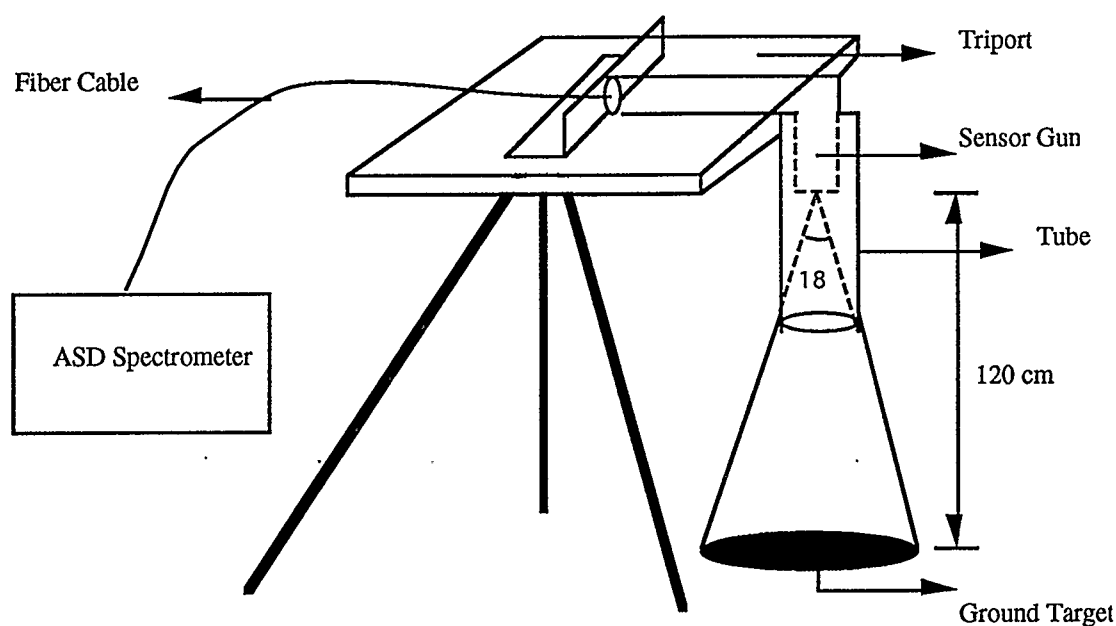


Figure 3.3 Measurement devices

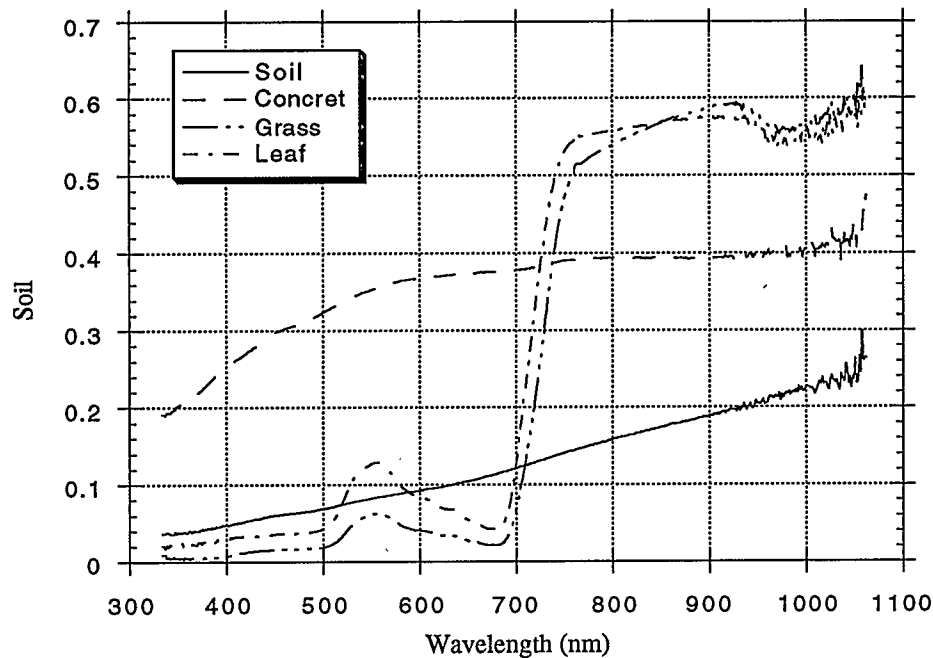


Figure 3.4 An example of field spectrometer data from 334 nm to 1062 nm

### 3.3 Characteristics of Hyperspectral Imagery

Hyperspectral data have some unique characteristics in comparison with other multispectral data. The main advantage of hyperspectral data is the hundreds of bands designed in a certain spectral range. The more bands used, the greater the spectral resolution that can usually be achieved.

casi has 288 bands covering a spectral range from 417 nm to 917 nm with a spectral resolution of approximately 3.5 nm. The field spectrometer used in this study has 512 bands within spectral range of 350 nm to 1067 nm at 1.4 nm spacing. This kind of high spectral resolution supplies sufficient amount of information on spectral features. A

little spectral variation of the different features may be recorded by the sensor. It provides a possibility to distinguish features that are spectrally similar. Because every material usually has a different reflectance spectra, even the materials with very similar spectra, such as different vegetation species, have some spectral differences at certain wavelengths. Only hyperspectral data with high spectral resolution can be used to distinguish these subtle spectral differences.

On the other hand, hundreds of bands contain a lot of data with redundant information. They occupy large computer memory and cost a lot of computational time when they are used for information extraction. This is the disadvantage of hyperspectral data. For example, some vegetation species have almost the same reflectance at most spectral regions. Only a few spectral bands can be used to distinguish between those species. These bands may be sufficient in the feature extraction analysis. In order to improve the computational efficiency, the number of redundant bands need to be reduced.

### **3.4 Noise Behaviour**

Any remote sensing image contains noise. In most cases, the noise can not be removed completely. The noise comes from two principal sources. One is the radiometric error. The other is the atmospheric interference. The radiometric noise is caused by hardware design of the sensor, and it is difficult to remove. The effect of atmospheric interference on remotely sensed data is critical in the image analysis. The amount of effect on the data may affect the accuracy of ground cover estimation and analysis. The atmosphere influences the amount of electromagnetic energy that is sensed by the detectors of an imaging system, and these effects are wavelength dependent [Curcio, 1961; Chavez, 1989]. The atmospheric effect is strong for systems, such as CASI, that record data in the visible and near infrared parts of the spectrum. The atmosphere affects images by

scattering, absorbing, and refracting light. The dominant source of these effects is scattering [Siegal *et al.*, 1980; Slater *et al.*, 1983].

Table 3.1 Atmospheric scattering models for different atmospheric conditions

Atmospheric Conditions	Relative Scattering Model
Very Clear	$\lambda^{-4.0}$
Clear	$\lambda^{-2.0}$
Moderate	$\lambda^{-1.0}$
Hazy	$\lambda^{-0.7}$
Very Hazy	$\lambda^{-0.5}$

One possible set of relative scattering models are listed in Table 3.1 [Chavez, 1989]. These models were selected based on the fact that very clear atmosphere is characterized by Rayleigh scattering, moderate atmosphere by Rayleigh and Mie scattering, while very hazy atmosphere is influenced by both Rayleigh and Mie scattering, but with Mie being more important than in the moderate atmosphere case [Slater *et al.*, 1983].

The existence of the noise affects the accuracy of ground information estimation and analysis. In a previous paper [Zhang *et al.*, 1993], random white noise was added as the atmospheric effect for testing the sensitivity of the linear decomposition analysis. In this study, the Rayleigh scattering model was adopted instead.



## CHAPTER 4

### BAND SELECTION

Band selection is traditionally treated as an important subject in image classification. Groups of bands are selected based on the calculation of class signature separability [Richard, 1986]. This involves not only class means but also class covariance matrix that are difficult to obtain for the case of hyperspectral data because of the large number of statistically meaningful samples required. However, in linear spectral unmixing the covariance matrix is not required for each endmember. Therefore, it is not possible to use the traditional separability measures.

Because hyperspectral images have hundreds of bands, it requires a lot of computation time for information extraction. It is desirable to use only those bands containing the necessary information for a particular task to improve the computational efficiency.

This is one of the major challenges for the use of the imaging spectrometer data. In a general sense, it primarily means the reduction of data dimensionality. Those spectral bands contain redundant information can be ignored. Several methods were developed in this research. Our emphasis was put on choosing some of valuable bands from the hyperspectral data set based on the relationship among the reflectance of different endmembers. The spectral bands were selected based on their capability to differentiate between different endmembers. More specifically, in the context of linear spectral

unmixing, band selection is done after the spectral identities of various endmembers are determined. The task of band selection is to examine which band contributes more to the discrimination of these different endmembers.

#### **4.1 Band Selection Based on the Successive Correlation Examination**

The hyperspectral image has hundreds of bands. The data in each band have a very high correlation with their neighboring bands. The effective channels in the pixel decomposition analysis are those which represent significant properties of spectral features. Therefore, the neighboring bands with high correlation do not need to be kept. Only one of them is sufficient in the subsequent analysis. The intention of this method is to drop some bands whose correlation is high compared to its neighboring bands. High correlation coefficient of two channel represents these two bands provide similar information.

In this method, the correlation coefficients between two successive neighboring bands are calculated. A threshold is specified for determining which band can be dropped. At the first step, the correlation of the first two neighbor bands is calculated. It is then compared with the predefined threshold, the second band will not be taken into consideration if the correlation is greater than the threshold. Otherwise, the second band is kept for further decomposition analysis. In the second step, if the second band was dropped in the first step, then the correlation between the first and the third bands is calculated and compared with the threshold. If the second band was kept in the first step, the correlation between the second band and the third band is calculated and compared with the threshold. If the correlation is larger than the threshold, the third band is dropped. If not, it will be maintained. Going through all the bands in this manner, some bands will be dropped from the decomposition consideration, some will be kept for subsequent analysis. Ideally, after this process a small number of bands without losing any important

information should be obtained.

The threshold is important in determining how many bands will be kept. It sets a criterion for how high the correlation between two neighboring bands cannot be accepted. Because hyperspectral data have many narrow bands, the correlation between two originally adjacent bands is very high. Usually the correlation exceeds 0.99. The correlation among any two original neighboring bands based on the six endmembers selected from the casi data is shown in Figure 4.1.

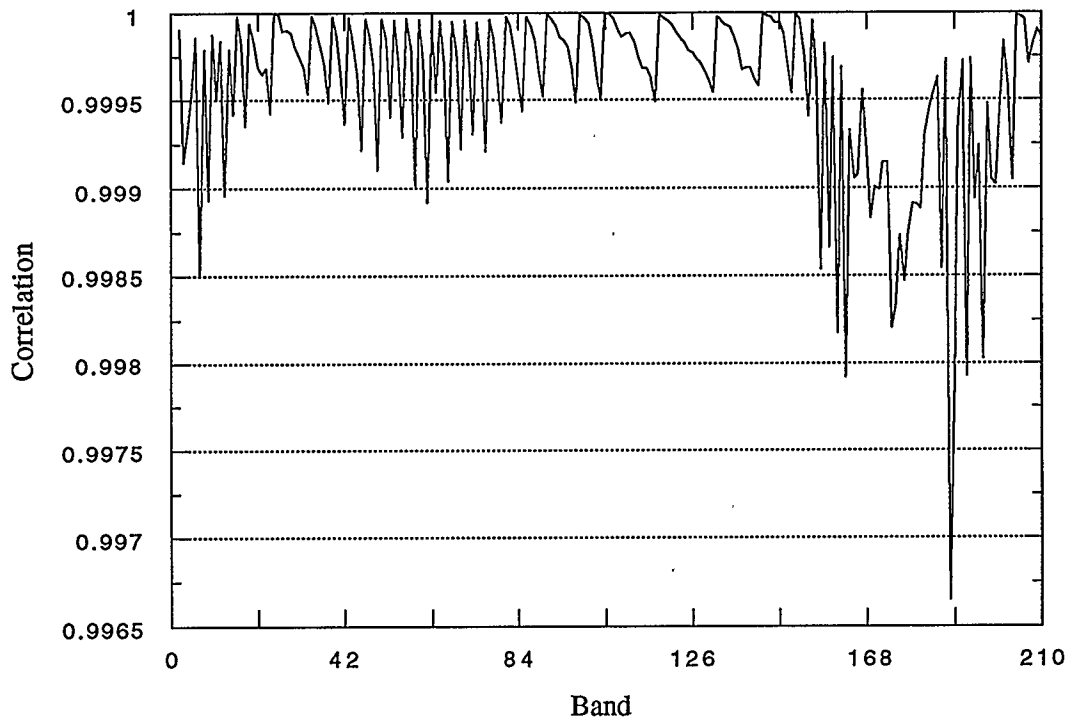


Figure 4.1 Correlation between two successive bands of CASI data

It is obvious that the hyperspectral data contain a lot of redundant information. The threshold determination is critical. If a large threshold is selected, some redundant bands may be kept. If a small threshold is selected, a small number of bands may be reserved.

Those remaining bands, for the small threshold case, may not be sufficient for the subsequent linear unmixing analysis. Because the correlation are all very high, the variance among them is very small. It is difficult to have a suitable threshold for discarding some of the redundant bands.

Besides using a heuristic procedure, the threshold determination can be made based on the relationship between correlation and the lag distance which is very similar to the semivariogram estimation used in geostatistics [Curran, 1988, Curran and Dungan, 1989]. The difference is that in semivariogram one examines the spatial variability along a profile while in this method the average lagged spectral correlation is examined along a spectral transact. The basic procedure is shown in Figure 4.2.

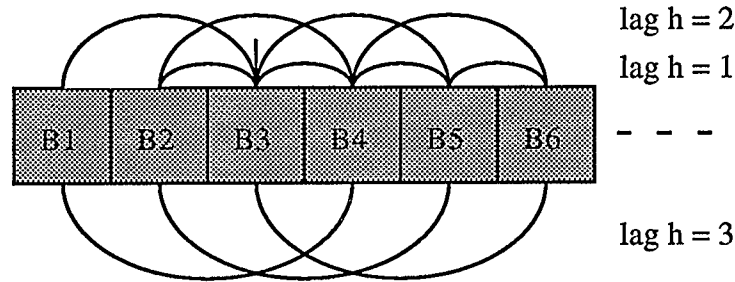


Figure 4.2. Lags along a transact of spectral bands

The correlation between a pair of bands,  $h$  intervals apart (the lag distance), was calculated by

$$\text{cor}(h) = \frac{1}{NB - h} \sum_{i=1}^{NB-h} \frac{\sum_{j=1}^n (X_{i,j} - \bar{x}_i)(X_{(i+h),j} - \bar{x}_{(i+h)})}{\sqrt{\sum_{j=1}^n (X_{i,j} - \bar{x}_i)^2 \sum_{j=1}^n (X_{(i+h),j} - \bar{x}_{(i+h)})^2}} \quad (4.1)$$

where  $NB$  is the total number of original bands,  $m$  is the pairs of bands separated by the same lag,  $n$  is the number of endmembers in one band. Figure 4.3 shows the relationship

between correlation and lag using the casi data with 210 bands. The threshold can be estimated from a curve such as Figure 4.3 and the number of bands desirable. Then the threshold will be adjusted through experiment in order to reach the exact number of bands desired. When 50 bands are specified to be selected, the minimum lag distance is about 4 or 5 which is obtained from  $210/50$ . From the curve, we can see the minimum threshold should be set larger than 0.996. Through experimentation, 0.999 was set and 51 bands were selected. It is difficult to control the procedure to get the exact number of bands. But the number of bands may be chosen as close as expected.

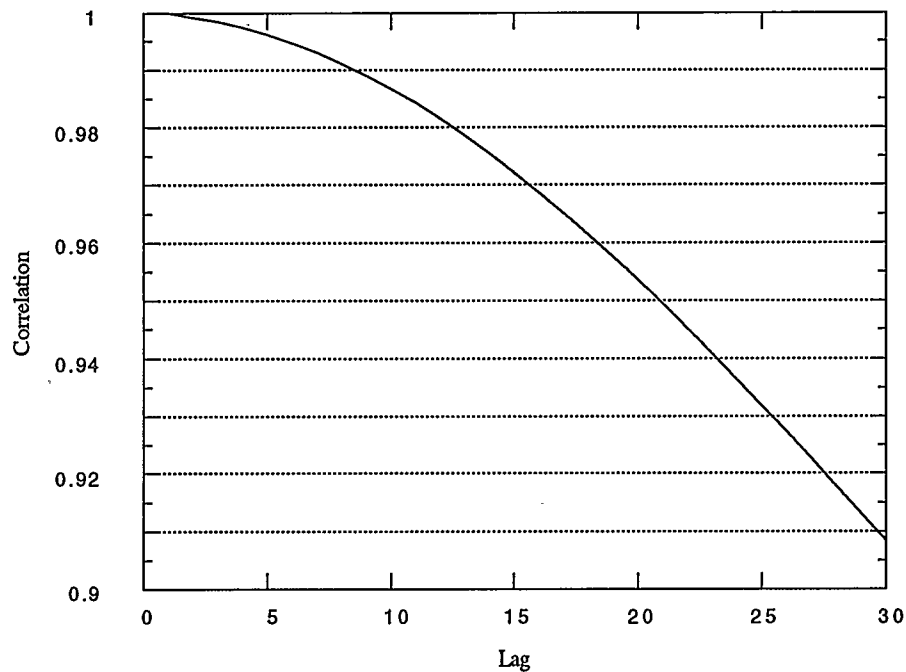


Figure 4.3 The mean correlation of all bands with different lags

## 4.2 Averaged Correlation Analysis

In this method, the correlation coefficient between any two bands is calculated, and a correlation matrix is generated. Because the endmember reflectance matrix has dimensions  $m \times p$ , the correlation matrix is a symmetrical  $m \times m$  matrix with diagonal elements as one. In a previous study, the mean value of all the correlation coefficients in the correlation matrix was used as the threshold [Zhang *et al.* 1993]. A band with row-wise (or column-wise) mean correlation exceeding this threshold was dropped. In this study, an empirical formula was developed for determining the threshold.

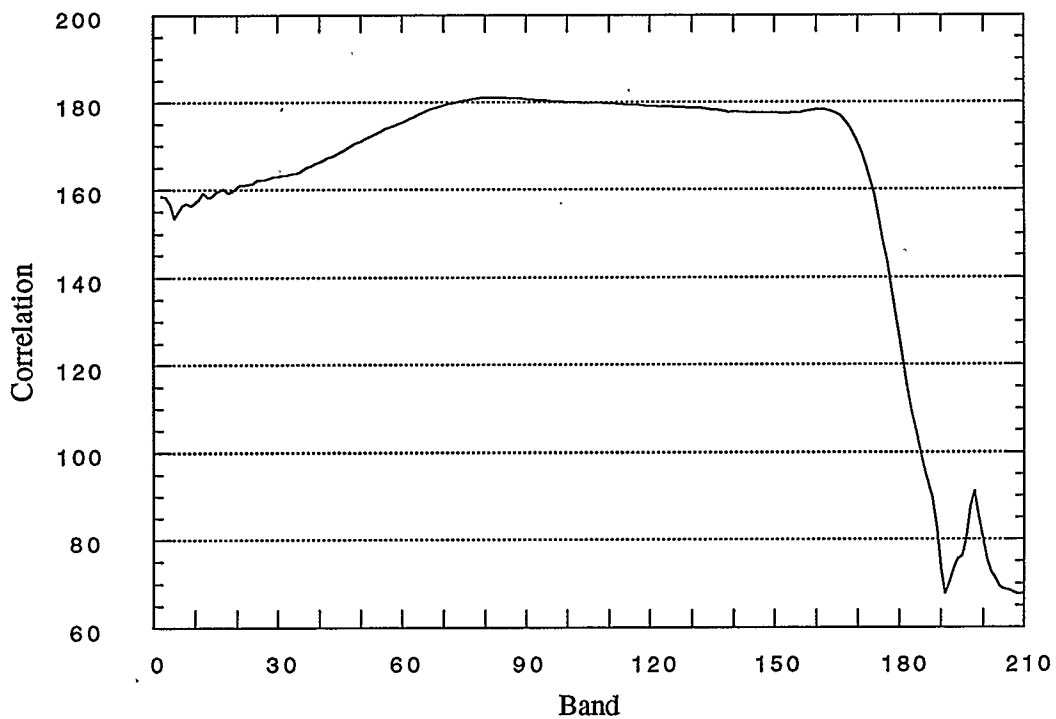


Figure 4.4 Row-wise ( or column-wise) sum of correlations for each band

Figure 4.4 shows the row-wise (or column-wise) mean correlation of each band using casi data with 210 bands. It is hard to determine the threshold from this curve. It

cannot be said that only the bands after 175 contains higher separability than at the rest bands. For instance, from Figure 3.1, the separability in some other bands are higher than band 180.

In order to avoid losing some useful information, bands should be chosen fairly along the spectral wavelength. The 210 band casi data were analyzed for estimating the threshold. Based on the knowledge of the reflectance spectrum, the separabilities between band 80 to band 130 are high enough to be used for the purpose of decomposition. The mean correlation in these bands are, however, relatively high. The number of desirable bands should be inversely proportional to the level of mean correlation. In other words, we can select a small number of bands from a group of bands whose row-wise mean correlation are high. On the other hand, a relatively large number of bands from a set of bands whose row-wise mean correlation are low should be selected. Therefore, we need to rank all of the bands based on their row-wise mean correlation and divide them into different groups based on their different ranges of row-wise mean correlation. When the total number of spectral bands to be selected is determined or specified, the distribution of bands to be selected in each different group is determined by a non-linear empirical function. The empirical non-linear function for selecting bands is defined as

$$NB_i = w_i \cdot B \quad (4.2)$$

$$w_i = \frac{nb_i}{cor_i^a} / \sum_{i=1}^m \frac{nb_i}{cor_i^a} \quad (4.3)$$

where  $NB_i$  is the number of bands to be selected in the  $i$ th range if we rank the range from small correlation to large.  $B$  is the total number of desired bands.  $w_i$  is percentage of the number of bands in the  $i$ th range with respect to the total number of desired bands.  $nb_i$  is the number of bands in the  $i$ th range.  $cor_i$  is the correlation of the  $i$ th range.  $a$  is the

parameter to be determined through experiment. The parameter  $a$  is used to balance the effects of the correlation and the number of bands with this correlation.

In our experiment, the correlation was divided into eight ranges. The number of bands in each range were counted. Then, a histogram was generated and shown in Figure 4.5. The power parameter  $a$  was set to 1.0, 2.0, 3.0, 4.0, respectively. 50 bands was set as the desirable number of bands. The number of selected bands in each range is shown in Figure 4.5. The greater the power  $a$ , the more effect of correlation has on the distribution of desirable bands. The smaller the power  $a$ , the more effect of the number of bands in each range has. By comparing the effects of the different powers in Figure 4.5, it is easy to see the distribution of desirable bands is much more suitable when  $a$  is set to 2. In subsequent experiments, the empirical formula

$$NB_i = \left( \frac{nb_i}{cor_i^2} / \sum_{j=1}^m \frac{nb_j}{cor_j^2} \right) \cdot B \quad (4.4)$$

was used to determine the distribution of desirable bands. After the number of bands to be selected in each correlation range was determined, the bands were selected randomly within each range. In this procedure, it is hard to obtain the exact number of bands specified at the beginning. In order to compare each band selection method, the same number of bands has to be chosen. Additional bands were added manually to reach the desirable number.



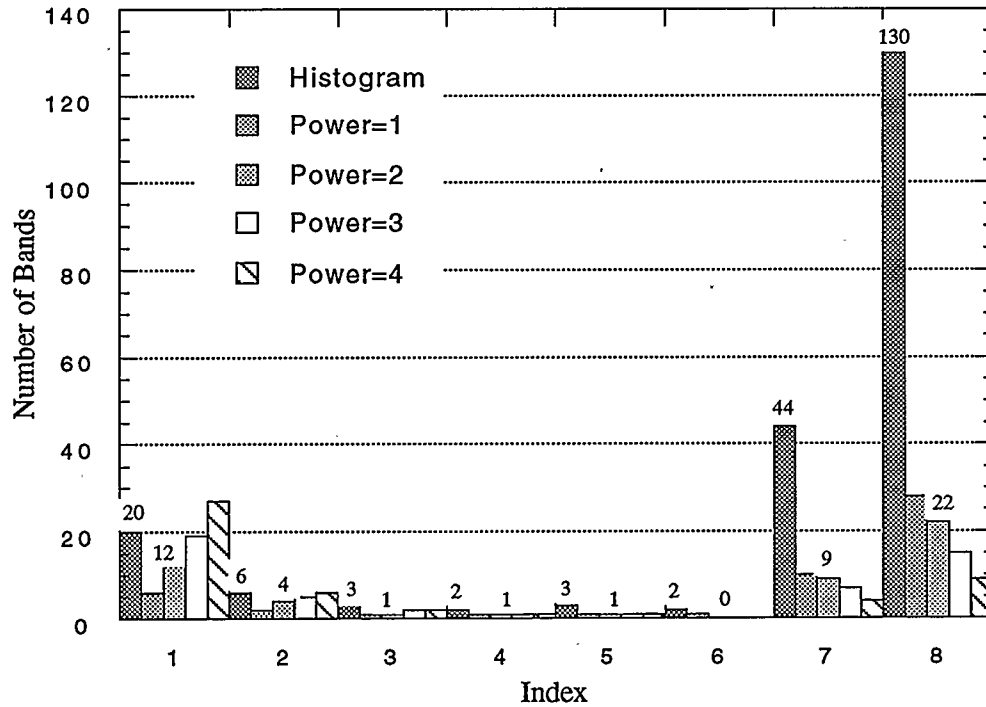


Figure 4.5 Histogram of selected bands using the empirical formula with different coefficients

### 4.3 Interband Minimum Difference Analysis

We can assess the separability between endmembers in each band based on how different their spectral reflectances are. It is obvious that larger differences show higher separability. Based on this kind of separability, bands can be selected.

The minimum distance of spectral reflectance among various endmembers is used to describe the separabilities of different endmembers in each spectral band. Along the spectral wavelength range, some of the endmembers may have the same or very similar reflectances at certain bands. The desirable bands are those with relatively large minimum

reflectance differences. The minimum difference among all endmembers in each band can be calculated. Figure 4.6 shows the distribution of the minimum difference along the bands by using the 210-band CASI data. Then these minimum differences can be ranked from large to small. When the number of desirable bands is specified, those bands with larger minimum reflectance differences will be selected. The drawback of this method is that the selected bands still have high correlation between them. The other problem with this method is the selected bands may be distributed in a small spectral range if the number of desirable bands is small.

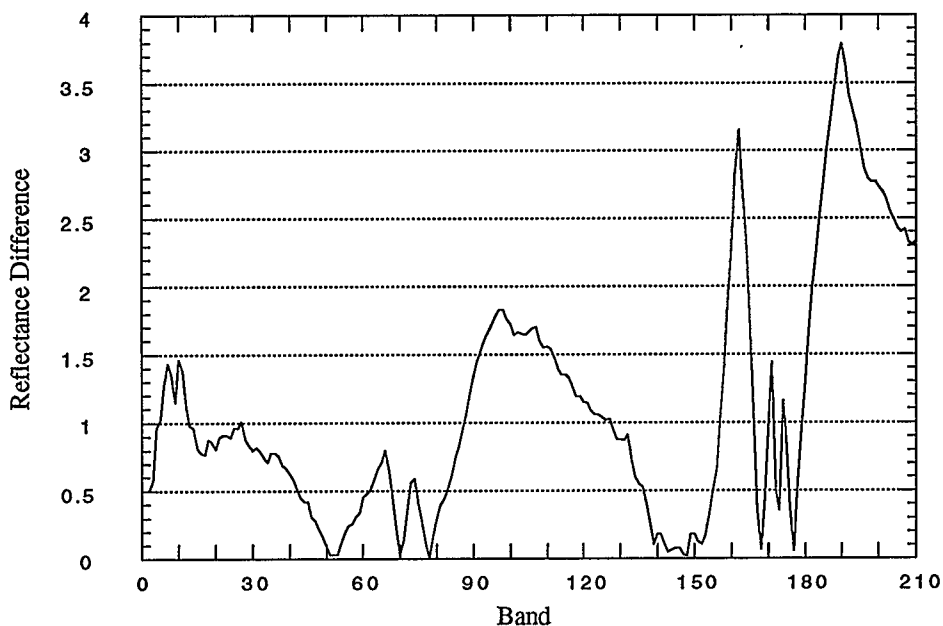


Figure 4.6 The distribution of minimum difference in each band

#### 4.4 Combining the Minimum Difference and Successive Correlation Methods

Considering the problems existed in the method presented in Section 4.3, a method

combining the minimum difference and the correlation is implemented. The initial band selection can be done using the minimum difference method. The successive correlation method described in Section 4.1 can then be applied to those selected bands. The thresholds for minimum difference and for correlation of neighboring bands need to be adjusted through experimentation. In order to preserve as much information as possible after the first selection, more bands need to be selected with the minimum difference method. The threshold for correlation can be adjusted to reach the number of desirable bands.

#### **4.5 Step-Wise Principal Component Analysis**

Principal component analysis is widely applied in remote sensing for reducing data dimensionality [Mather, 1987]. It is often used to determine a linear combination of spectral bands, resulting in a smaller number of uncorrelated dimensions in the feature space. This approach determines the importance of each original band in each resultant feature and finds some optimum features without a significant loss of information in terms of the proportion of total variance of the original data set.

A modified stepwise principal component analysis (MSPCA) procedure is applied for spectral band selection. The MSPCA can be efficiently used to select particular subset of original bands to decrease the dimensionality of the original data. It does not use linear combinations of all the input bands as has been done in Chen and Landgrebe [1989]. In MSPCA we select a subset of the entire input bands by preserving most information.

Before performing MSPCA, the correlation of any two bands is calculated to generate a correlation matrix,  $A$ . This correlation matrix is used as the input to MSPCA. Eigenvalues and eigenvectors can be extracted from  $A$ . They will satisfy the following

Equation (4.5)

$$(A - UI)V = 0 \quad (4.5)$$

$U$  is the eigenvalue of  $A$ , and  $V$  is the eigenvector of  $A$ . Larger eigenvalues correspond to higher magnitudes of variances which implies more information for discrimination purposes. Smaller eigenvalues correspond to lower variances which corresponding to less amount of information.

In contrast to the general PCA strategy, the feature accounting for the smallest amount of variance is considered first in MSPCA. The smallest amount of variance corresponds to the feature with the lowest eigenvalue. The original band having the highest weight in the eigenvector corresponding to the lowest eigenvalue is considered to contain the least amount of information. This band is dropped from further processing.

The MSPCA procedure iteratively drops bands by finding the lowest eigenvalue with the highest weight until the desirable number of bands is reached. It will not decrease the correlation among the original bands as efficiently as the original PCA, but it preserves the original physical meaning of each spectral band.

Since the computation involved in the MSPCA is relatively large, the band selection is actually done through making a log file to record the number of iterations and the corresponding dropped band. Figure 4.7 shows the sequence of initial dimensionality reduction through the MSPCA procedure. The horizontal axis is the band index. The vertical axis shows the number of iteration. The horizontal bars represent the bands dropped at a particular iteration. When the number of desirable bands is specified, the bands will be chosen in a reverse order beginning from the end of the file. The number of bands is easily controlled to reach the expected number.

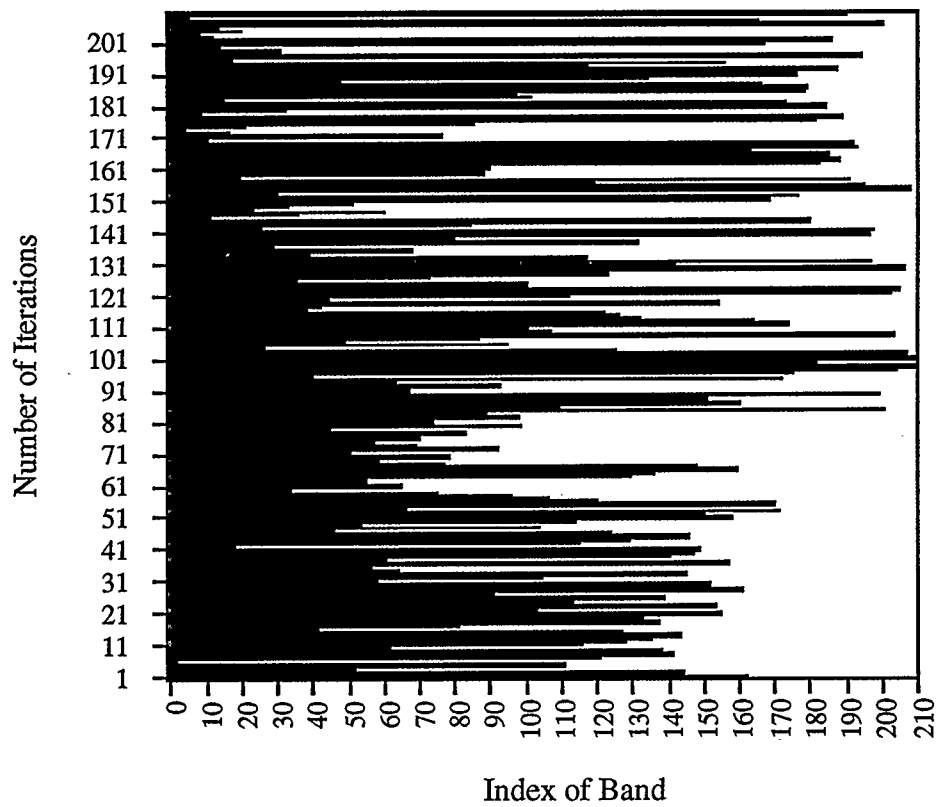


Figure 4.7 Sequence of original band reduction by MSPCA

## **CHAPTER 5**

### **IMPLEMENTATION AND EXPERIMENTS**

#### **5.1 Data Preprocessing**

Two data sets obtained from different sources were used to test the sensitivity of the linear unmixing and band selection methods. One data set was the Compact Airborne Spectrographic Imager (casi) data. The other was the field spectrometer data.

##### **5.1.1 CASI Data**

The casi data used in this research were obtained on May 20, 1991 from Metolius River (44°23' N, 121°40' W) in middle Oregon. In this experiment, the spectral mode of the casi data which provides reflectance spectra from 417 nm to 927 nm in 288 bands was used.

Although the casi spectral-mode data were calibrated using the manufacturer's parameters, the radiance values for each image pixel were noisy which can be seen from Figure 3.1. The noise came from atmospheric interference and the radiometric effect [Gong et al. 1992]. Because the radiometric noise has a high effect in longer wavelengths, the casi reflectance imagery between 417 and 788 nm were used in this experiment. In order to reduce radiometric noise effect on the separability analysis, a 1 x 7 smoothing filter has been used to smooth the reflectance curves. To suppress the atmospheric effect, the

radiance was converted to reflectance using the field-measured gravel spectra as a pseudo-invariant reflectance target [Freemantle et al., 1992].

The reflectance spectra used in this project were extracted from the calibrated spectral-mode casi image. Some relatively pure areas were selected as endmembers for generating the reflectance matrix. These endmembers are three kinds of gravel in three sites, two kinds of grass which are tuft grass and snow grass, and ponderosa. These areas were chosen based on the knowledge of study sites. The spectra displayed in Figure 5.1 is a smoothed version of Figure 3.2. They have been smoothed by a  $1 \times 7$  mean filter. Other 45 samples were extracted from the casi image as the mixtures which were used as observation vectors  $\mathbf{b}_i$  in Eq.(2.3). Same processing was applied to these observation vectors as those endmember spectra.

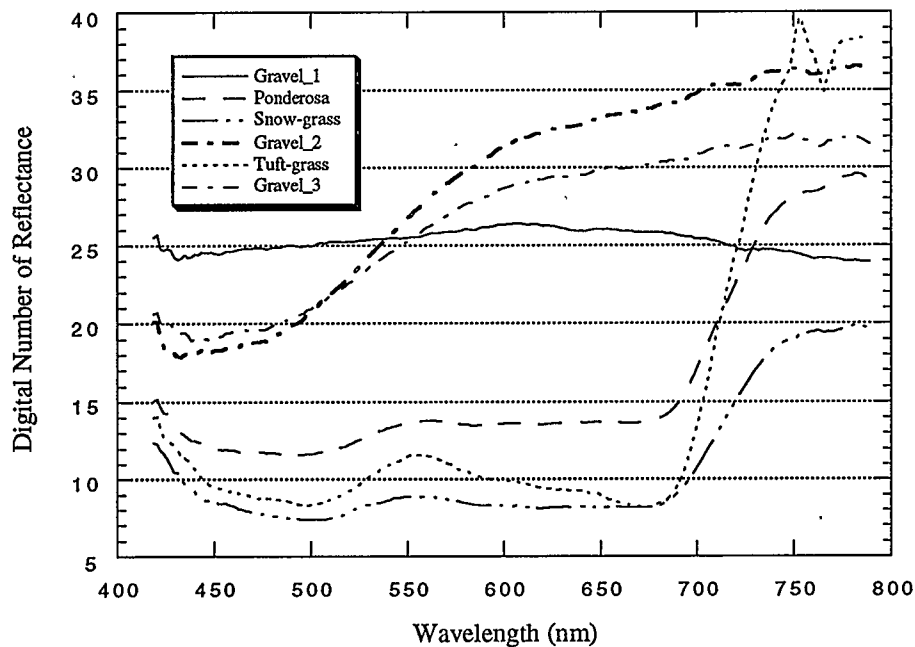


Figure 5.1 Smoothed CASI reflectance curves for six endmembers using a  $1 \times 7$  filter

### 5.1.2 Field Measurements

Another set of reflectance spectra used in this study was measured on the campus of The University of Calgary in September, 1993. The Analytical Spectral Device's (ASD) personal spectrometer II was used with a field of view of  $18^\circ$ . These spectra were collected for the purpose of linear unmixing analysis and evaluating the efficiency of the band selection methods.

The reflectance spectra of four kinds of targets, grass, soil, aspen leaf and concrete, were collected. These targets were very easy to find, and was relatively easy to take reflectance spectra from relatively pure targets thanks to the limited field of view of the spectrometer. The spectra of these pure targets constituted a reflectance matrix. Figure 5.2 shows the reflectance spectra of these four endmembers. The mixing targets were artificially generated. Any two of those four pure materials were mixed to generate a mixing spectra in specific proportions. On the other hand, three kinds of mixtures were measured on the campus. These mixtures are soil and grass, soil and spruce leaf, concrete and grass. In each mixture, we selected mixtures such that the proportion of each endmember was approximately 50%. However, because the surveying conditions were hard to control, it cannot be guaranteed that the measured spectra is mixed by two pure endmembers with exactly equal proportions. Because the field spectrometer used in our experiment is less sensitive in the longer wavelength and the shorter wavelength regions, the reflectances shorter than 378 nm and longer than 952 nm were discarded. Only 430 bands out of the 512 bands from the raw data were preserved for further analysis.



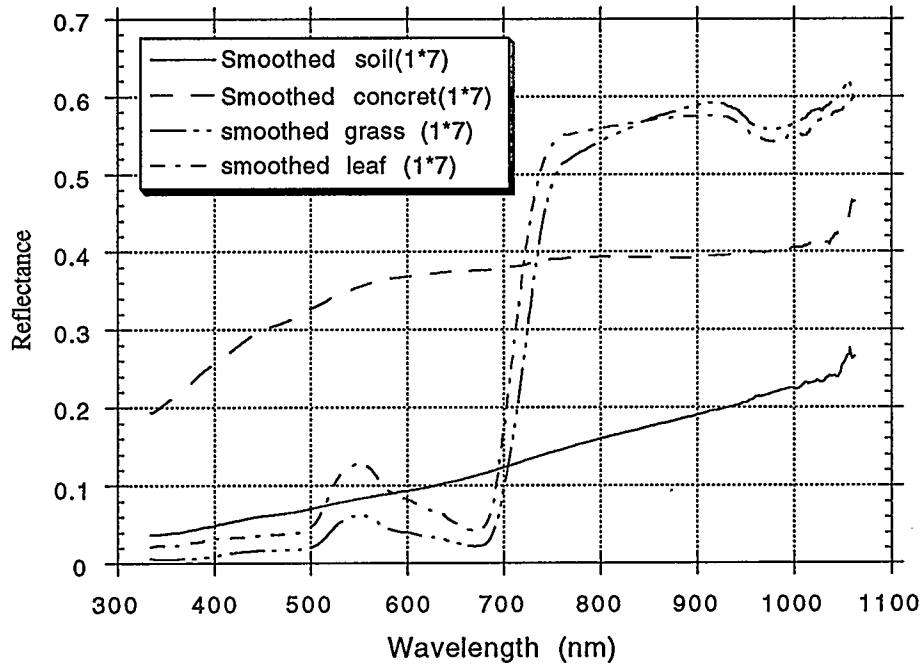


Figure 5.2 Smoothed ASD spectrometer reflectance curves for four endmembers using a 1x7 filter

## 5.2 Experiments

In this research, the sensitivity of linear spectral unmixing was tested under two conditions. The first condition was when there was undefined endmember inside a study area. The second condition was when atmospheric effect was considered existing in the observation data. On the other hand, the five band selection methods were implemented and were used to reduce the number of spectral bands. Figure 5.3 shows the experimental design.

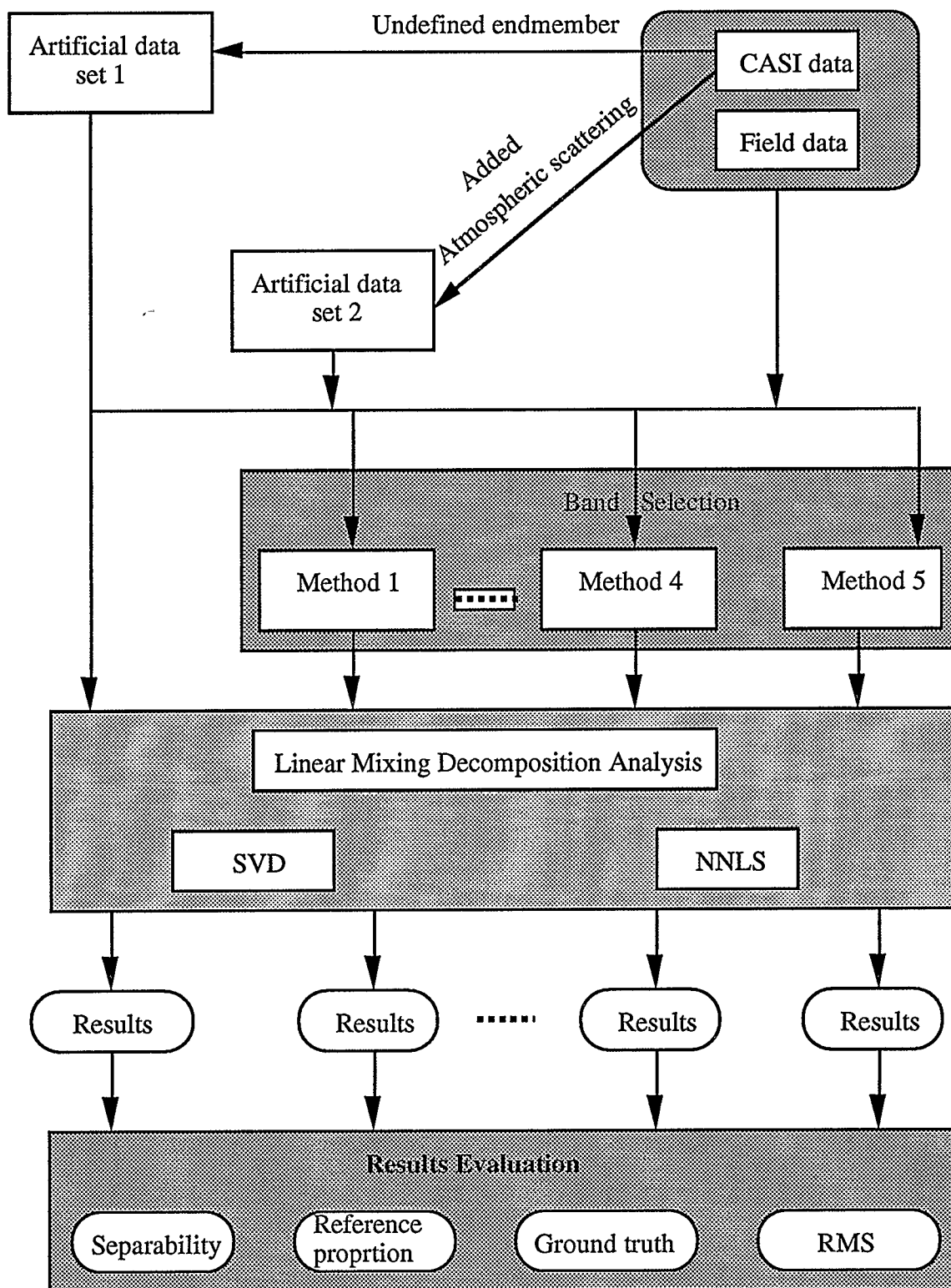


Figure 5.3 Experimental design

### 5.2.1 The Sensitivity of Linear Mixing Model

#### 5.2.1.1 Undefined endmember inside study area

Sometimes, some endmembers existing in a study area may not be identified and thus will not be included in the endmember matrix. They may contribute to the observation vector. This experiment was designed for testing the performance of the linear spectral unmixing under these conditions.

An artificial observation data set was created by proportionally mixing the six endmembers from the casi image, with the proportions of 0.005, 0.27, 0.11, 0.025, 0.44 and 0.15 for the six endmembers, respectively. The observation vector was linearly combined by these six endmembers, but one of these endmembers was not included in the endmember matrix. This one was assumed to be the neighboring effect, and the proportion of its contribution in the mixing spectra could be small. Table 5.1 shows the results obtained from linear spectral unmixing when one of the combination endmember does not exist in the reflectance matrix.

Table 5.1 Decomposition results with an undefined endmember

Endmember	Gravel_1	Ponderosa	Grass_1	Gravel_2	Grass_2	Gravel_3
Real Proportion	0.005	0.27	0.11	0.025	0.44	0.15
Unconstrained Method (SVD)						
Fraction1	*	0.27	0.11	0.01	0.44	0.17
Fraction2	0.00	0.28	0.10	*	0.44	0.19
Constrained Method (NNLS)						
Fraction1	*	0.27	0.11	0.01	0.44	0.17
Fraction2	0.00	0.27	0.09	*	0.45	0.18

The “\*”s in Table 5.1 represents the endmember which had certain proportion in

creating the mixed pixel but was not included in the endmember matrix. All the bands were used in the linear decomposition process. Two methods, which were the unconstrained singular value decomposition (SVD) and the nonnegative least square (NNLS) method, were used to decompose the mixture. From the results shown in the table, the endmember with very small proportion did not affect the decomposition analysis very much. The linear mixing model still provided reasonable estimation when the material near the target had little contribution to the spectra of the desirable target.

#### 5.2.1.2 Atmospheric effect

This test was designed to test the sensitivity of the linear spectral unmixing when atmospheric scattering exists in the observation data set. Generally, the reflectance of endmembers are measured in the field or obtained in the laboratory. The atmospheric effect on these measurements may be ignored. However, the data obtained from remotely sensed image are always affected by the atmospheric scattering.

In this experiment, the atmospheric effects were simulated. With the assumption of clear atmosphere, Rayleigh scattering was used to generate the artificial atmospheric component to test the sensitivity of linear decomposition. Rayleigh scattering can be simply represented by

$$S_r = k / \lambda^4 \quad (5.1)$$

$k$  is the parameter changing with wavelength. A completely dark object on the ground has zero reflectance should be used to estimate  $k$ . Actually, this kind of dark object does not exist in the image. Chavez [1989] suggested that a one percent minimum reflectance be used to represent atmospheric scattering. Because  $k$  only has a role on modifying the magnitude of the atmospheric effect, it was assumed as a constant in our study in order to

simplify the problem. It was estimated from the minimum reflectance in wavelength near 420 nm that is corresponding to blue light. Light at the shorter wavelength region has stronger scattering. One percent of the minimum reflectance in that wavelength was assumed as the minimum atmospheric scattering. Then  $k$  can be obtained by

$$k = S_r \lambda^4 \quad (5.2)$$

For the casi data,  $k$  was calculated as  $7.67 \times 10^9$ . The atmospheric effect on other wavelengths was simulated based on Eq.(5.1).

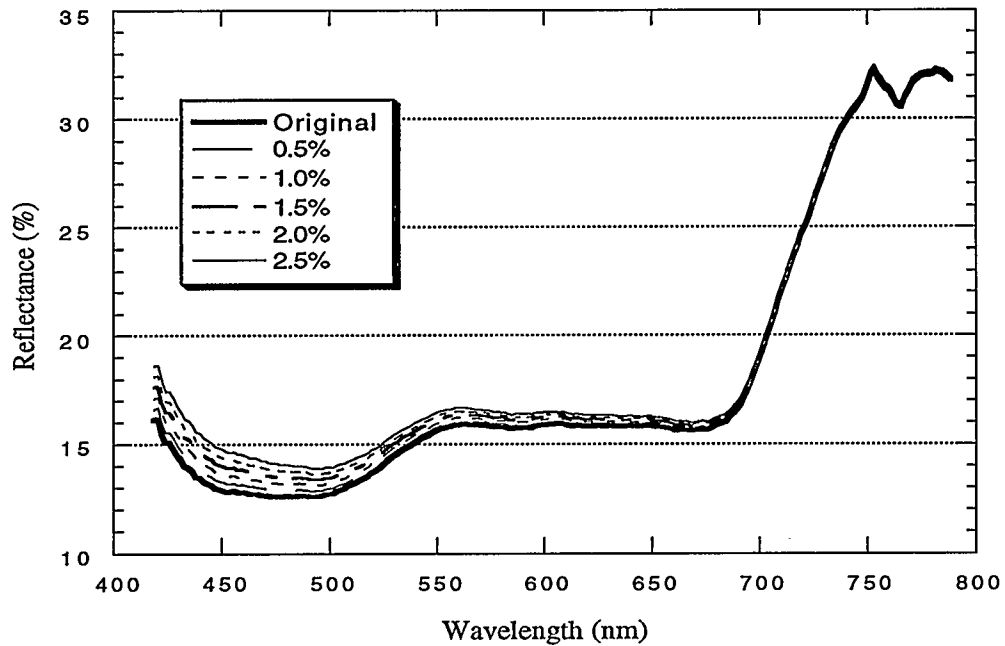


Figure 5.4 Simulated Rayleigh scattering added onto the artificially created spectra

The simulated atmospheric noise was added on the artificially mixed observation data. Figure 5.4 shows the atmospheric effect on the observation data along the spectra. The artificial observation was mixed by six endmembers using proportions of 0.05, 0.27, 0.11, 0.07, 0.35, 0.15, respectively. The magnitude of the Rayleigh scattering increases

from index one to five with the percentage of minimum reflectance at 420 nm from 0.5 - 2.5%. Table 5.2 shows the atmospheric effect on the pixel decomposition analysis. Two decomposition methods were used.

From Table 5.2, it can be seen that as the magnitude of Rayleigh scattering increases, the reliability of the decomposition results decreases. The results are poorer for the constrained method than for the unconstrained method. When the atmospheric effect is low, the results show that the linear decomposition analysis may provide reliable estimation. However, the results from the linear decomposition analysis may not be reasonable if the atmospheric effect is strong. Therefore, atmosphere calibration is required before the remotely sensed image is used for decomposition analysis.

Table 5.2 Decomposition results from data contaminated by atmospheric effect

Endmember	Gravel1	Ponderosa	Grass1	Gravel2	Grass2	Gravel3	RMS
Real Proportion	0.05	0.27	0.11	0.07	0.35	0.15	0.0
Index	Unconstrained Method (SVD)						
1	0.06	0.23	0.19	0.04	0.34	0.17	0.0169
2	0.08	0.18	0.28	0.02	0.33	0.19	0.0339
3	0.09	0.14	0.36	-0.01	0.32	0.21	0.0508
4	0.11	0.10	0.45	-0.04	0.31	0.23	0.0678
5	0.12	0.06	0.53	-0.07	0.30	0.24	0.0847
Index	Constrained Method (NNLS)						
1	0.06	0.23	0.19	0.04	0.34	0.17	0.0169
2	0.08	0.18	0.28	0.02	0.33	0.19	0.0339
3	0.10	0.14	0.37	0.00	0.32	0.19	0.0509
4	0.12	0.09	0.46	0.00	0.31	0.17	0.0680
5	0.15	0.04	0.55	0.00	0.31	0.15	0.0852

### 5.2.2 Evaluation of band selection methods

The purpose of optimal band selection is to reduce the dimensionality of the original data for further pixel decomposition analysis. Ideally, we expect that the reduction of data

dimension does not affect the accuracy of the decomposition analysis. Theoretically, any band reduction should not affect the linear decomposition result if the observed reflectance is linearly mixed by the reflectances from components in the study area. However, this assumption may not hold true in reality. If this assumption does not hold, band reduction affects the decomposition results. To evaluate the effectiveness of different band selection methods, the following four criteria were used in this study [Csillag et al.1993]:

1. the separability of reflectance curves of different endmembers,
2. the reference proportion of same kinds of components in each pixel,
3. the ground-truth proportion of different components in one measurement,
4. the root mean squared error (RMS).

The first criterion is based on expert experience in manual band selection. With the second criterion, the reference proportion for an endmember was assumed to be the one obtained from the data with all the bands being used. The correlation between the reference proportions and the proportions obtained from the selected bands were calculated for evaluation. It is a relative evaluation criterion that does not take the non-linear mixing into consideration. The third criterion is more reliable. The RMS can be used to evaluate the appropriateness of proportions extracted from selected bands.

We used the *casi* and field spectrometer data for evaluating the five band selection methods. The band selection results obtained from the *casi* data will be evaluated by of all the methods except the third one because the ground-truth is not available. Because of the shortage of ground samples, the band selection results obtained from the field spectrometer data will not be evaluated using the second criterion..

In order to compare the effectiveness of different band selection methods, the number of selected bands has to be the same in each band reduction method. In this study, the number of spectral bands was specified as either 20 or 50.

#### **5.2.2.1 The first evaluation criterion**

The spectral separability of different endmembers is an obvious way to evaluate the selected bands. It provides us with the preliminary evaluation of different band selection methods. Figure 5.5 (a) – (e) shows the positions of selected bands obtained from the five band selection methods when 20 bands were specified. The *casi* data were used here. Figure 5.5 (a) shows the bands selected using the successive correlation method. Some selected bands are not at wavelength positions where different endmembers have high separability. The reason for this phenomenon is that this band selection method is not based on the reflectance separability of endmembers. Only the correlation between pairs of neighboring bands is considered. Nevertheless, the selected bands are well distributed along the wavelength axis. The smaller the change of the reflectance curves, the fewer bands are selected. The larger the change of the reflectances, the more bands are selected. From Figure 5.5 (a), most bands selected based on the neighborhood correlation have acceptable spectral separability. Figure 5.5 (b) shows the selected bands obtained through the correlation of each band with all the others. Some of the selected bands still have low reflectance separability. Similar to the first band selection method, this method is only based on the overall correlation with other bands. The largest reflectance separability is not guaranteed. Comparing with Figure 5.5 (a), the selected bands in Figure 5.5 (b) is not as well distributed as in (a). No band was selected in the shorter wavelength region since it might affect the subsequent linear decomposition analysis. Figure 5.5 (c) shows the selected bands obtained based on a reflectance separability analysis. Most of the bands



selected are centralized in longer wavelength region. The result is caused by the threshold determined for this method. Only these bands with higher reflectance separability were kept without considering the correlation among those bands. The selected bands shown in Figure 5.5 (d) were obtained from the improved method which selected bands based on both reflectance separability and the correlation of the reserved bands. This band selection method is much better than the one considering reflectance separability only. Figure 5.5 (e) shows the bands selected with the step-wise PCA method. The distribution of those bands are very similar to those in Figure 5.5 (a). Whether or not the successive correlation method consistently produces similar band selection results to those from the step-wise PCA method needs further study.

Bands selected using different band selection methods from the spectrometer data have similar characteristics to those obtained from the casi data.

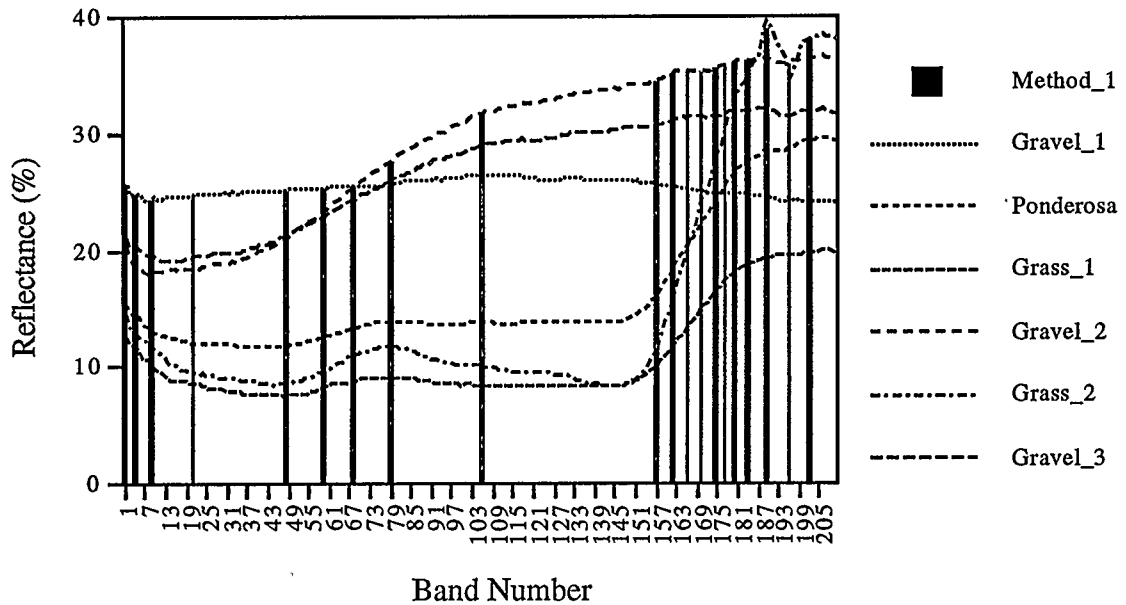


Figure 5.5 (a) Separability for 20 selected bands obtained by the successive correlation examination method (the first method)

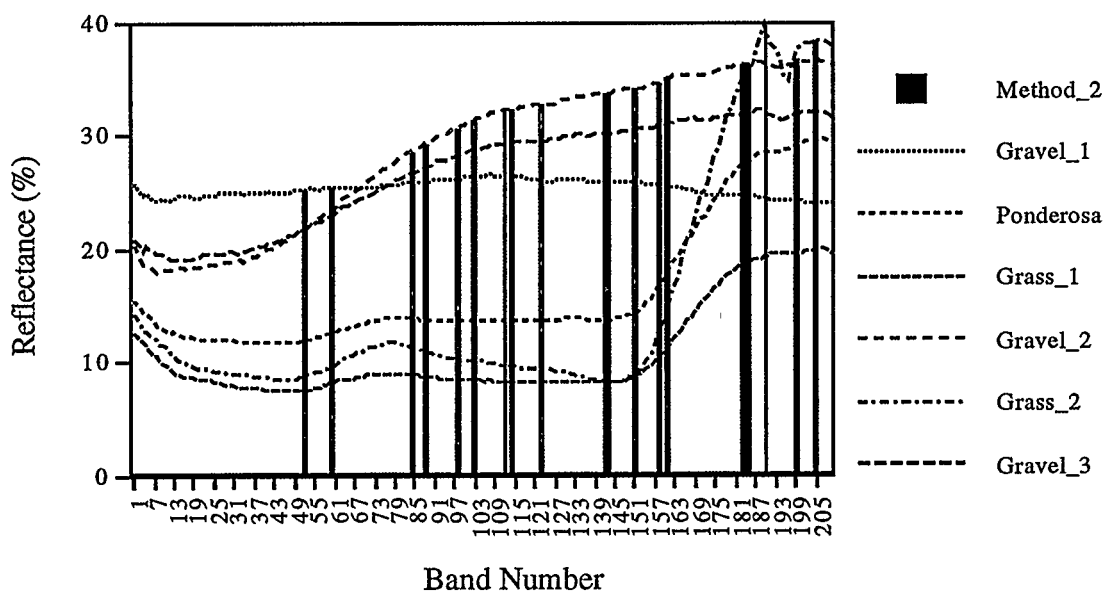


Figure 5.5 (b) Separability for 20 selected bands obtained by the averaged correlation method (the second method)

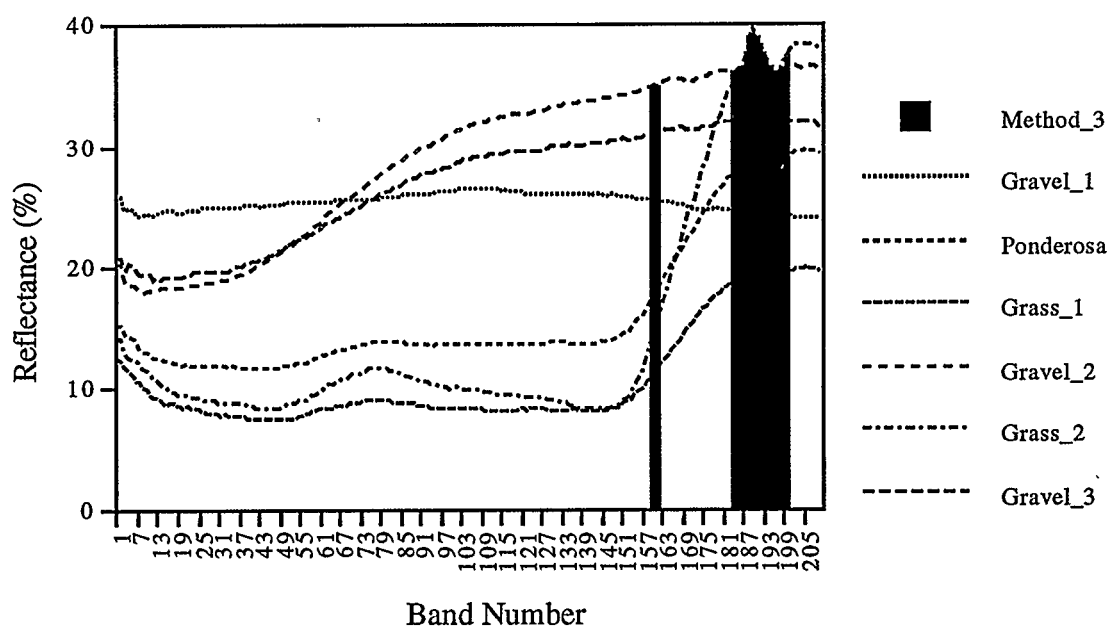


Figure 5.5 (c) Separability for 20 selected bands obtained by the interband minimum difference analysis method (the third method)

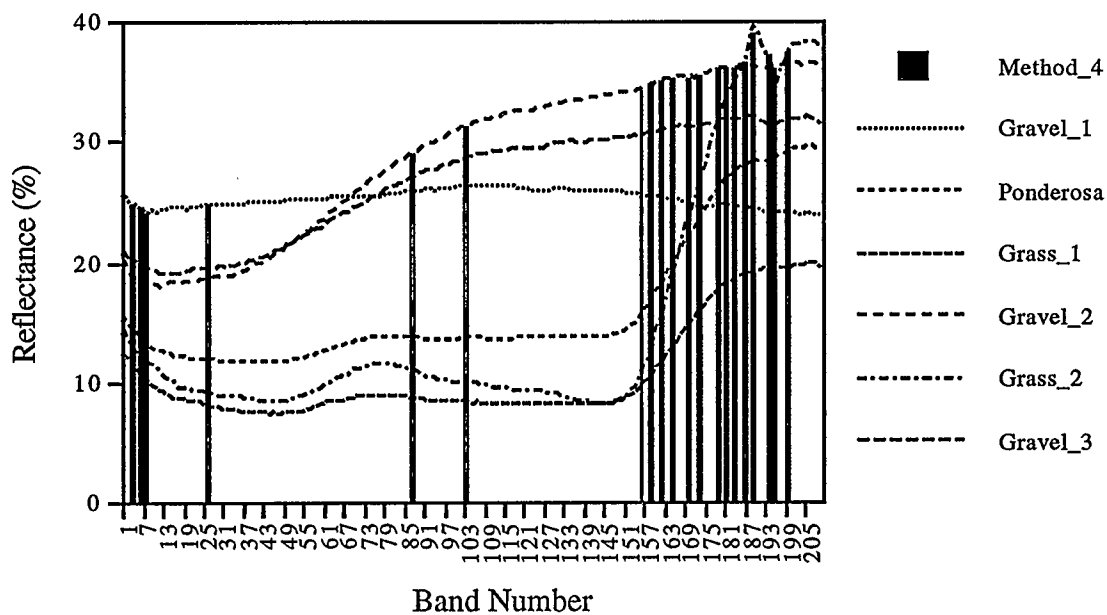


Figure 5.5 (d) Separability for 20 selected bands obtained by combining the minimum difference and successive correlation method (the fourth method)

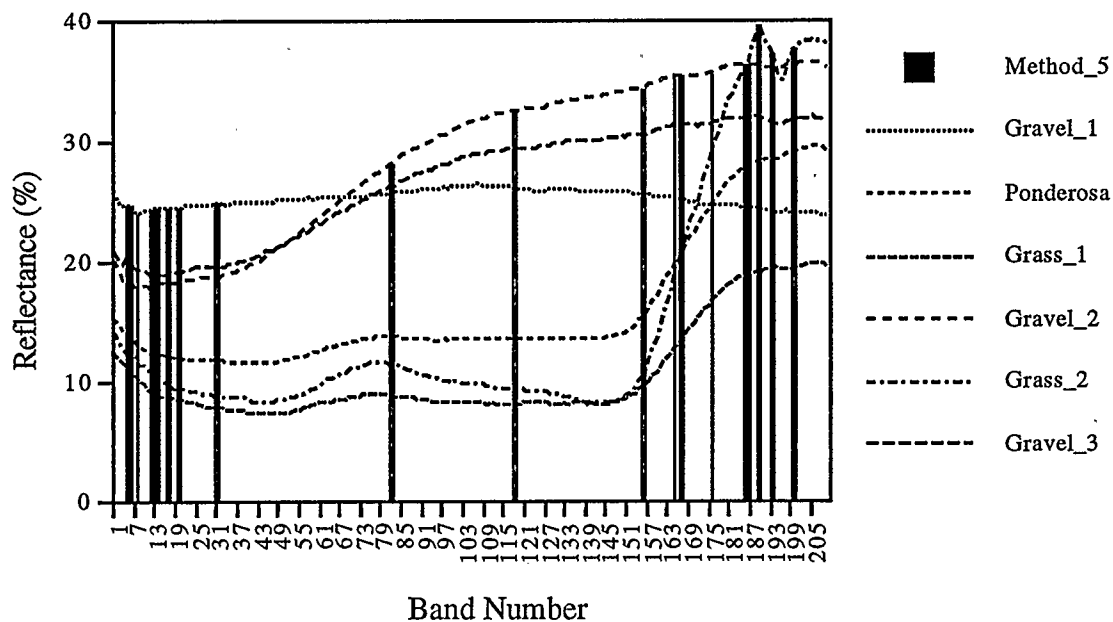


Figure 5.5 (e) Separability for 20 selected bands obtained with the step-wise principal components method (the fifth method)

### 5.2.2.2 The second evaluation criterion

This method is used to compare the decomposition results obtained from different methods when the ground truth is not available. A total of 45 samples extracted from the *casi* image were used to test the different band selection methods. The results obtained using all the bands were assumed to be the true proportions of endmembers and were used as a reference. The correlation between the reference results and the results obtained from the selected bands through different band selection methods were calculated for comparison. A correlation close to one is desirable.

Figure 5.6 (a) shows the correlation coefficients between the unmixing results obtained from the unconstrained decomposition method using 50 selected bands through each band selection method and the results obtained using all the bands. Figure 5.6 (b) shows the results obtained with the constrained decomposition method. The correlation shown in Figures 5.6 (a) and (b) indicate that the third band selection method which only considers the reflectance separability is the worst one among the five methods. The inconsistent result may be caused by the threshold determination for this method. All of the selected bands obtained from this method closely distribute in a narrow spectral range. It implies that the linear spectral unmixing is sensitive to the distribution of selected bands along the wavelength. If the selected bands cannot provide the sufficient information for separating different materials, the results will not be reliable for quantifying the ground cover at the subpixel level. The fourth method is not stable in decomposing different mixing samples especially when the constrained unmixing algorithm (NNLS) was used. The decomposition results obtained from the selected bands using other methods are relatively stable with both the unconstrained (SVD) and the constrained (NNLS) linear

unmixing procedure, and the correlation coefficients are close to the one with most testing samples except for one case. The explanation for this is that the linear mixing model itself does not perfectly describe the spectral mixing process. If the reflectance of the mixture is linearly combined, the band reduction should not affect the decomposition results. However, even the linear decomposition results are obtained from all the bands may not provide accurate proportions of different endmembers in a particular mixture pixel. When the results are used as the reference to evaluate the other results obtained from a selected number of bands, the correlation between these results and those others may not be reliable. But it still provides a means to evaluate the decomposition results obtained from selected bands when no other information is available.

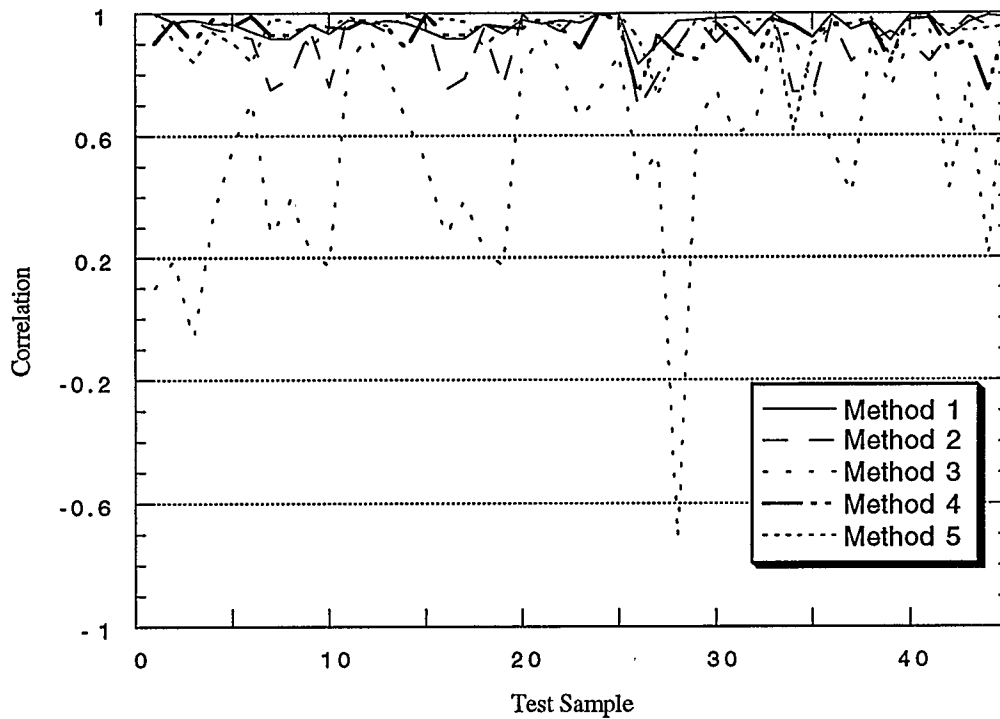


Figure 5.6 (a) The correlation between the results obtained from all the bands and from 50 selected bands using the SVD method

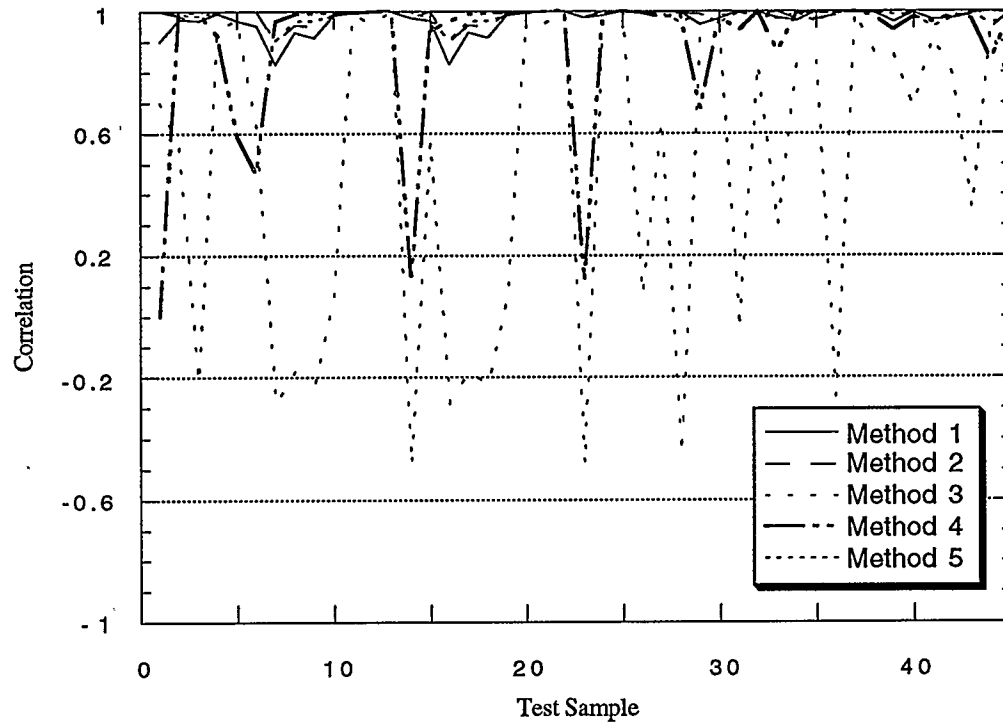


Figure 5.6 (b) The correlation between the results obtained from all the bands and from 50 selected bands using the NNLS method

### 5.2.2.3 The third evaluation criterion

When data is collected with well known conditions, the data provides us with an opportunity to use ground-truth for evaluating the decomposition results. Three kinds of mixture were measured and tested. Table 5.3 shows the results obtained from the SVD and the NNLS linear decomposition procedures with all the bands and 20 selected bands. Because the field measured target is not combined by two types of materials in exactly half proportions, these results do not show 50% of each combined endmember. The

decomposition results obtained from those selected bands except those from the third one are reasonable. Their root mean squares (RMS) are very low. However, the proportions extracted from different subsets of bands are not the same. This inconsistent proportions of endmembers may be caused by the assumption of the linear mixing process itself. The results obtained from bands selected with the third band selection method are not reliable. Its RMSs are much larger than those from other band selection methods. This implies that optimal band selection methods should select bands whose wavelengths are well distributed along a spectral range. The effect of band reduction on the decomposition analysis may be caused by the least square solution of the overdetermined equations.

Table 5.3 The decomposition results obtained through SVD and NNLS procedure from field measurements

	Mixture	Soil	Concrete	Grass	Leaf	RMS
		Unconstrained Method (SVD)				
All the bands	soil - grass	0.45	0.00	0.37	0.15	0.0048
	soil - leaf	0.54	-0.06	-0.05	0.55	0.0033
	con - grass	0.06	0.34	0.52	0.09	0.0024
Method_1	soil - grass	0.48	0.00	0.47	0.06	0.0068
	soil - leaf	0.56	-0.05	0.01	0.48	0.0046
	con - grass	0.04	0.34	0.54	0.07	0.0025
Method_2	soil - grass	0.49	-0.01	0.37	0.14	0.0051
	soil - leaf	0.57	-0.06	-0.06	0.55	0.0037
	con - grass	0.05	0.33	0.52	0.10	0.0030
Method_3	soil - grass	-0.31	1.75	0.98	-1.42	0.3314
	soil - leaf	0.08	1.01	0.07	-0.15	0.2086
	con - grass	0.40	-0.43	0.13	0.91	0.1421
Method_4	soil - grass	0.47	0.01	0.48	0.04	0.0072
	soil - leaf	0.56	-0.05	0.01	0.48	0.0048
	con - grass	0.04	0.34	0.54	0.07	0.0025
Method_5	soil - grass	0.47	-0.01	0.49	0.05	0.0090
	soil - leaf	0.55	-0.04	0.04	0.46	0.0060
	con - grass	0.05	0.32	0.52	0.11	0.0054
		Constrained Method (NNLS)				
All the bands	soil - grass	0.45	0.00	0.38	0.14	0.0048
	soil - leaf	0.41	0.00	0.08	0.42	0.0061
	con - grass	0.06	0.34	0.52	0.09	0.0024
Method_1	soil - grass	0.48	0.00	0.47	0.06	0.0068
	soil - leaf	0.51	0.00	0.14	0.35	0.0111
	con - grass	0.04	0.34	0.54	0.08	0.0025
Method_2	soil - grass	0.49	0.00	0.38	0.13	0.0054
	soil - leaf	0.53	0.00	0.07	0.39	0.0098
	con - grass	0.05	0.33	0.52	0.10	0.0029
Method_3	soil - grass	0.17	0.68	0.15	0.00	0.1429
	soil - leaf	0.15	0.85	0.00	0.00	0.1786
	con - grass	0.20	0.00	0.28	0.52	0.0608
Method_4	soil - grass	0.47	0.01	0.48	0.04	0.0072
	soil - leaf	0.51	0.00	0.12	0.36	0.0106
	con - grass	0.04	0.34	0.54	0.07	0.0025
Method_5	soil - grass	0.46	0.00	0.50	0.04	0.0092
	soil - leaf	0.51	0.00	0.13	0.36	0.011
	con - grass	0.05	0.32	0.52	0.11	0.0054



#### 5.2.2.4 Result evaluation based on RMS

The RMS is calculated based on Eq.(2.9). The proportions obtained from different methods are used to reconstruct an observation for each band. The root mean squared errors are used to evaluate the accuracies of those proportions. Figure 5.7 (a) shows the RMS results obtained from the SVD procedure using all the bands and 20 selected bands. Figure 5.7 (b) shows the RMSs obtained with the SVD method using all the bands and 50 selected bands. Since the error magnitudes for Method 3 are much greater than the rest of the band selection methods, the RMS results for Method 3 are not shown in Figures 5.7(a) and (b). Figures 5.7 (c) and (d) show the RMSs obtained from the NNLS method using all the bands or 20 selected bands, and all the bands or 50 selected bands, respectively. Comparing these four figures, the third band selection method is very sensitive to the band reduction. It cannot provide stable result when the number of bands is reduced. All of the other methods give relative consistent results although the RMS increased slightly when the number of selected bands is small. The results obtained from NNLS have larger RMS relative to those obtained from SVD when Figures 5.7(a) and (c) or Figures 5.7(b) and (d) are compared. This is reasonable because the NNLS method has more and stronger constraints in the search of optimal solution than the SVD method.

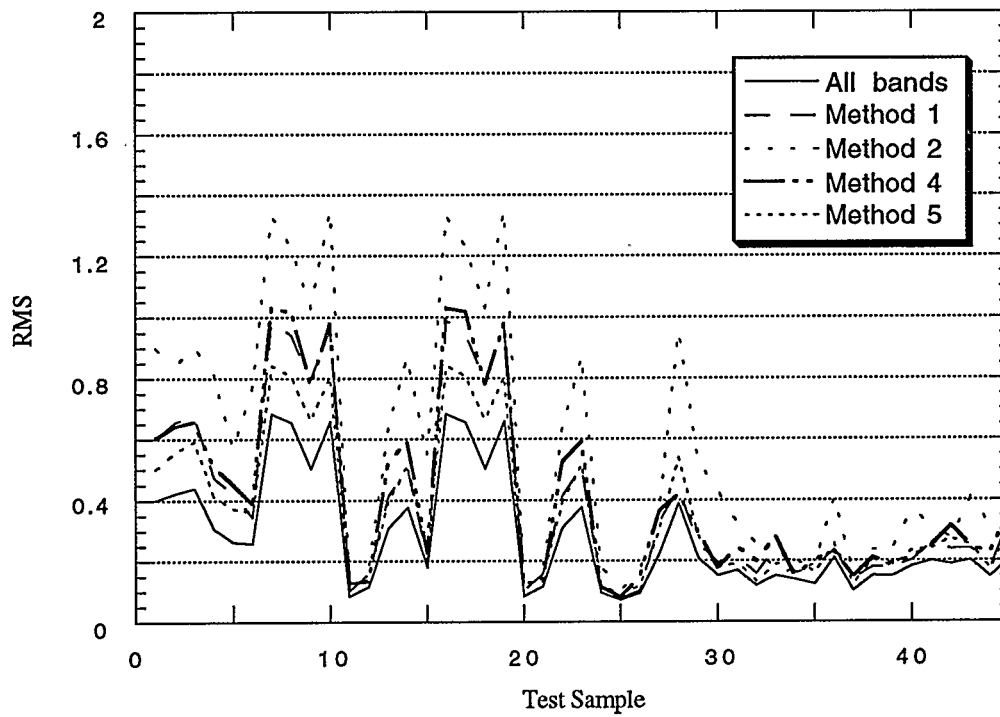


Figure 5.7 (a) The RMSs of the decomposition results obtained from all the bands and from 20 selected bands using the SVD method

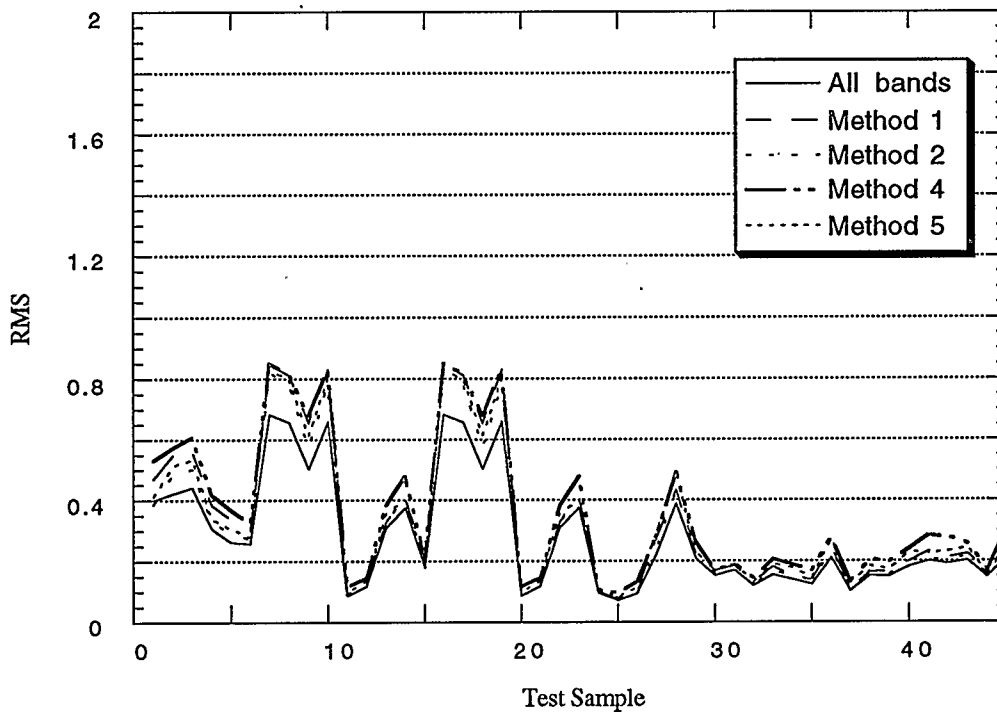


Figure 5.7 (b) The RMSs of the decomposition results obtained from all the bands and from 50 selected bands using the SVD method

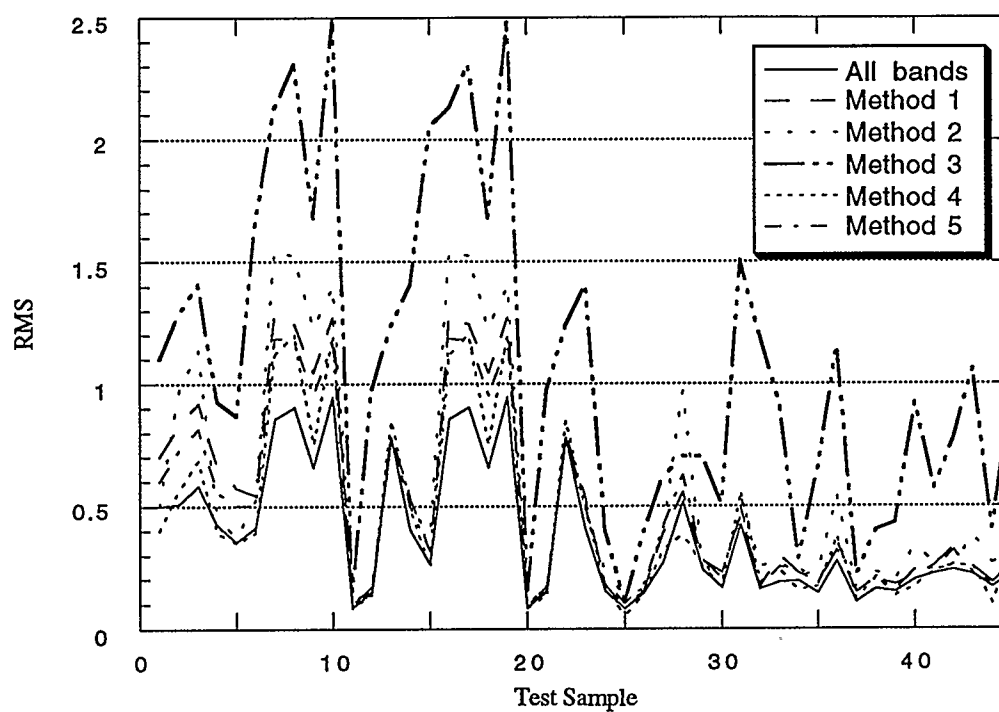


Figure 5.7 (c) The RMSs of the decomposition results obtained from all the bands and from 20 selected bands using the NNLS method

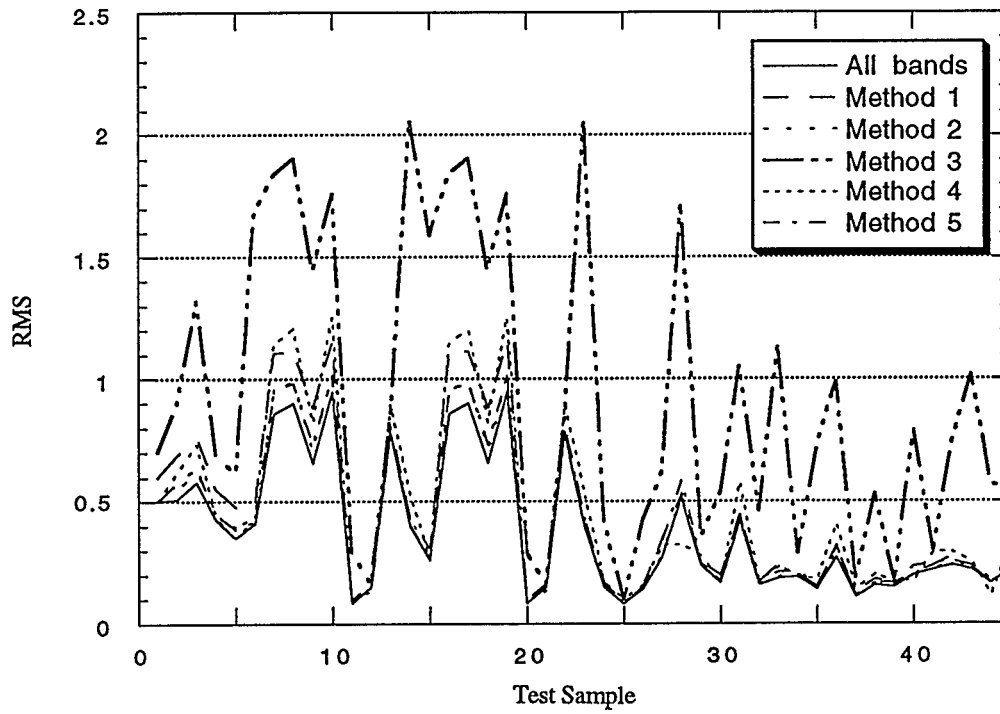


Figure 5.7 (d) The RMSs of the decomposition results obtained from all the bands and from 50 selected bands using the NNLS method

### 5.2.3 The sensitivities of band selection methods to atmospheric noise

The same artificial data used in testing the sensitivity of linear unmixing in Section 5.2.1.2 were used here. The atmospheric noise was added to the observation in the same manner as shown in Figure 5.3. A sum of squares (SSD) for the difference of each set of results with the real proportion is used to rank these five methods. Table 5.4 displays the decomposition results obtained using the SVD method derived from all of the bands or 50 selected bands. RMSs are also listed in the table. Table 5.5 shows the decomposition results obtained using the NNLS method. From the RMSs in both tables, it can be seen

that the effect of atmospheric scattering on the linear decomposition analysis is larger in the results obtained from selected bands than those obtained from all the bands. This is because the reflectances in certain wavelengths may contain more noise components. If only a small number of bands are selected for linear decomposition analysis, and some of the selected bands are in the shorter wavelengths, the linear unmixing may be affected more by the false spectral features exhibited through larger amount of noise in the shorter wavelengths. In this case, more bands may be helpful for the decomposition analysis. The third method presents much better results compared to other methods based on SSD. This is because most of the bands selected using this method are in the longer wavelength regions, where the atmospheric effect is small. Therefore, the results of linear spectral unmixing are better than all the other methods. Based on the results shown in Table 5.4 and Table 5.5, these band selection methods still can provide reliable estimation of proportions when the noise is relatively small. When the magnitude of noise increases, even the result obtained with all of the bands are not consistent. Therefore, some of those band selection methods are effective for the purpose of decomposition analysis when compared with the results obtained from all the bands. The linear mixing model itself is actually affected more by noise than band selection.

Table 5.4 The atmospheric effect on linear decomposition by using SVD with 50 selected bands

Endmember	A	B	C	D	E	F	SSD	RMS
Fraction	0.05	0.27	0.11	0.07	0.35	0.15	0.00	0.00
All the bands	0.06	0.23	0.19	0.04	0.34	0.17	0.010	0.017
	0.08	0.18	0.28	0.02	0.33	0.19	0.042	0.034
	0.09	0.14	0.36	-0.01	0.32	0.21	0.092	0.051
	0.11	0.1	0.45	-0.04	0.31	0.23	0.156	0.068
	0.12	0.06	0.53	-0.07	0.3	0.24	0.251	0.085
Method_1	0.06	0.23	0.19	0.03	0.34	0.19	0.011	0.019
	0.07	0.19	0.27	-0.02	0.33	0.23	0.047	0.038
	0.08	0.15	0.35	-0.06	0.33	0.27	0.105	0.057
	0.09	0.11	0.42	-0.1	0.32	0.31	0.179	0.076
	0.1	0.07	0.5	-0.14	0.31	0.35	0.250	0.095
Method_2	0.07	0.23	0.2	0.06	0.34	0.14	0.010	0.018
	0.09	0.2	0.28	0.05	0.32	0.13	0.037	0.037
	0.12	0.16	0.37	0.04	0.31	0.12	0.088	0.055
	0.14	0.12	0.45	0.03	0.3	0.11	0.152	0.072
	0.16	0.08	0.54	0.02	0.29	0.1	0.242	0.092
Method_3	0.07	0.28	0.11	0.05	0.35	0.15	0.001	0.060
	0.09	0.28	0.1	0.04	0.35	0.16	0.003	0.120
	0.1	0.29	0.1	0.02	0.35	0.16	0.006	0.180
	0.12	0.29	0.09	0.01	0.35	0.17	0.010	0.240
	0.14	0.3	0.09	-0.01	0.35	0.17	0.016	0.300
Method_4	0.06	0.25	0.16	0.02	0.34	0.19	0.007	0.027
	0.08	0.23	0.2	-0.03	0.34	0.23	0.027	0.054
	0.09	0.22	0.25	-0.08	0.33	0.28	0.064	0.081
	0.11	0.2	0.3	-0.13	0.32	0.32	0.114	0.108
	0.12	0.18	0.35	-0.18	0.32	0.36	0.178	0.135
Method_5	0.07	0.23	0.18	0.05	0.34	0.16	0.008	0.022
	0.09	0.2	0.25	0.02	0.34	0.17	0.029	0.044
	0.11	0.16	0.32	0	0.33	0.18	0.066	0.067
	0.13	0.13	0.39	-0.03	0.33	0.19	0.116	0.089
	0.16	0.09	0.45	-0.05	0.32	0.2	0.178	0.111

Table 5.5 The atmospheric effect on linear decomposition by using NNLS with 50 selected bands

Endmember	A	B	C	D	E	F	SSD	RMS
Fraction	0.05	0.27	0.11	0.07	0.35	0.15	0.00	0.00
All the bands	0.06	0.23	0.19	0.04	0.34	0.17	0.010	0.02
	0.08	0.18	0.28	0.02	0.33	0.19	0.042	0.03
	0.10	0.14	0.37	0	0.32	0.19	0.094	0.05
	0.12	0.09	0.46	0	0.31	0.17	0.167	0.07
	0.15	0.04	0.55	0	0.31	0.15	0.263	0.09
Method_1	0.06	0.23	0.19	0.03	0.34	0.19	0.011	0.02
	0.08	0.18	0.27	0	0.34	0.21	0.043	0.04
	0.10	0.13	0.36	0	0.33	0.19	0.092	0.06
	0.13	0.08	0.45	0	0.32	0.17	0.164	0.08
	0.16	0.03	0.54	0	0.32	0.15	0.260	0.09
Method_2	0.07	0.24	0.19	0.06	0.34	0.15	0.008	0.02
	0.09	0.20	0.28	0.04	0.32	0.15	0.037	0.04
	0.11	0.17	0.36	0.03	0.31	0.14	0.079	0.05
	0.13	0.13	0.44	0.01	0.30	0.14	0.141	0.07
	0.15	0.10	0.53	0	0.28	0.14	0.225	0.09
Method_3	0.07	0.28	0.11	0.05	0.35	0.15	0.001	0.06
	0.09	0.28	0.10	0.04	0.35	0.16	0.003	0.12
	0.10	0.29	0.10	0.02	0.35	0.16	0.006	0.18
	0.12	0.29	0.09	0.01	0.35	0.17	0.010	0.24
	0.14	0.30	0.09	0	0.35	0.16	0.014	0.30
Method_4	0.06	0.25	0.16	0.02	0.34	0.19	0.007	0.01
	0.09	0.22	0.22	0	0.34	0.20	0.024	0.02
	0.12	0.18	0.29	0	0.33	0.17	0.051	0.04
	0.15	0.14	0.36	0	0.33	0.15	0.095	0.05
	0.18	0.09	0.43	0	0.33	0.12	0.158	0.07
Method_5	0.07	0.23	0.18	0.05	0.34	0.16	0.008	0.02
	0.09	0.20	0.25	0.02	0.34	0.17	0.029	0.04
	0.11	0.16	0.32	0	0.33	0.18	0.066	0.07
	0.14	0.11	0.40	0	0.33	0.15	0.123	0.09
	0.17	0.06	0.49	0	0.32	0.13	0.209	0.11

### 5.3 Summary

In this chapter, artificial data were generated to test the sensitivity of the linear mixing model with the existence of undefined endmembers or the atmospheric effect. In



this experiment, all the bands were used in the linear unmixing process with the SVD and the NNLS methods. The decomposition results were compared with the real proportion to detect the sensitivity of the linear mixing model.

Five band selection methods were developed using *casi* data sets and field data sets. Through the SVD and the NNLS solutions, the decomposition results were evaluated using four criteria: the separability of selected bands, the correlation coefficients between the results obtained through selected bands and the reference result which was defined as the one obtained through all the bands, the comparison with the ground truth, and the RMS values. Artificial mixture was also used to test the sensitivity of those selected bands during the linear decomposition process.

## **CHAPTER 6**

### **CONCLUSIONS AND RECOMMENDATIONS**

#### **6.1 Conclusions**

In this thesis, all of the objectives listed in Section 1.2 were achieved. The first one was to test the linear mixing model using simulated data by considering the presence of an undefined endmember. The second one was to simulate and evaluate the atmospheric effect on the linear unmixing analysis. The third was to develop band selection methods for hyperspectral data. The fourth was to assess the effectiveness of the band selection methods. Based on the experimental results and analyses, the following conclusions can be drawn:

1. When a truly linear combination of endmembers is simulated, an undefined endmember with a relatively small proportion in the mixing spectra does not have a strong effect on the decomposition results obtained from both linear unmixing methods. Under such circumstances, the solution based on a linear mixing model can still provide reasonable estimation of proportions of other endmembers. This suggests that the spectral effect from the neighbor material near the target can be ignored in linear spectral decomposition.
2. Linear mixing model is sensitive to atmospheric effects regardless which decomposition method is used. When the amount of atmospheric noise is

small, inversion of this model can still generate reasonable results. Otherwise, it will provide unreliable estimates.

3. If a real mixture is simulated linearly by endmembers, the band reduction does not affect the decomposition results as long as the number of bands is greater than the number of endmembers. Based on the results obtained from the experiments with real data, band reduction does affect the decomposition results. This implies that the linear mixing model does not completely describe the actual spectral mixing process.
4. The selected band set for the decomposition analysis should be well distributed along the entire spectral range with low correlation to each other and high separability among different endmembers. By using the decomposition results obtained from all the bands as the reference, decomposition results from selected bands with various band selection methods were evaluated. The successive correlation examination, the averaged correlation analysis, the combining minimum difference and successive correlation method and the step-wise principal components analyses are more effective than the interband minimum difference analysis. Assuming that the spectral mixing is a linear one, the decomposition results obtained from selected bands through different band selection methods implemented in this research have similar response to the atmospheric effect as those obtained from all the bands.
5. The results obtained using the constrained decomposition method (NNLS) have larger RMS than those obtained using the unconstrained method (SVD). It is due to the procedure of constrained method in which NNLS forces the results into the range of  $[0,1]$ .

## **6.2 Recommendations**

More work needs to be done on testing the linear mixing model and the band selection methods proposed in this thesis. More ground truth data should be used to assess the decomposition results. The measurement design should be well controlled to measure the proportions of mixture components. Different atmospheric scattering models may be used by considering different kinds of weather condition. The endurance of the linear mixing model to noise effect needs to be tested further in a more quantitative manner. The reliability of the band selection methods needs to be tested under the situation of noise. The decomposition analysis with selected bands needs to be performed on the complete image to further test the linear mixing assumption and the effectiveness of various band selection methods.

## REFERENCES

- Adams, J.B., Smith, M.O., Gillespie, A.R., 1989, Simple models for complex natural surfaces: a strategy for the hyperspectral era of remote sensing, Proceedings of the IGARSS'89, Vancouver, British Columbia, Canada, pp.16-21.
- Adams, J.B., Smith, M.O., P.E. Johnson, 1986, Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site, Journal of Geophysical Research, 91: 8098-8112.
- Babey, S.K. R.J. Soffer, 1992, Radiometric calibration of the compact airborne spectrographic imager (CASI), Canadian Journal of Remote Sensing, Vol. 18, No.4, October 1992.
- Bierwirth, P.N., 1990, Minearl mapping and vegetation removal via data-calibrated pixel unmixing, using multispectral images, International Journal of Remote Sensing, 11: 1999-2017.
- Blunt, G., M.O. Smith, J.B. Adams, R. Greeley, P.R. Christensen, 1990, Regional aerolian dynamics and sand mixing in the Gran Desierto: evidence from Landsat Thematic Mapper images, Journal of Geophysical Research, 95: 15463-15482.
- Boardman, J.W., 1990, Inversion of high spectral resolution data, Proceedings of the IGARSS'90, College Park, Maryland, USA.
- Boardman, J.W., 1989, Inversion of imaging spectrometry data using singular value decomposition, Proceedings of the IGARSS'89, Vancouver, British Columbia, Canada, pp.2069-2072.
- Borstad, G.A, D.A. Hill, 1989, Using visible range imaging spectrometers to map ocean

phenomena, Proceedings from Advanced Optical Instrumentation for Remote Sensing of the Earth's Surface from Space, Vol. 1129.

Chavez, P.S., Jr., 1989, Radiometric calibration of landsat thematic mapper multispectral images, Photogrammetric Engineering & Remote Sensing, Vol. LV, No. 9, September, pp. 1285 - 1294.

Chen, C.C.T., D.A. Landgrebe, 1989, A spectral feature design system for the HIRIS/MODIS era, IEEE Transaction of Geoscience Remote Sensing, 27:891-894.

Cross, A.M., J.J. Settle, N.A. Drake, Paivinen, 1991, Subpixel measurement of tropical forest cover using AVHRR data. International Journal of Remote Sensing, 12: 1119-1129.

Csillag, F., L. Pasztor, L.L. Bièhl, 1993, Spectral band selection for the characterization of salinity status of soils, Remote Sensing Environment, Vol. 43, pp. 231-242.

Curcio, J.A., 1961, Evaluation of atmospheric aerosol particle size distribution from scattering measurement in the visible and infrared, Journal Optical Society of America, Vol. 51, pp. 548 - 551.

Curran, P.J., 1988, The semivariogram in remote sensing: An introduction, Remote Sensing of Environment, Vol.24, pp. 493 - 507.

Curran, P.J., J.L. Dungan, 1989, Estimation of signal-to-noise: A new procedure applied to AVIRIS data, IEEE Transactions on Geoscience and Remote Sensing, Vol. 27, No.5, September, pp. 620 - 628.

Duncan, I.J., B. Rivard, R.E. Arvidson, M. Sultan, 1990, Structural interpretation and tectonic evolution of a part of the Najd Shear Zone (Saudi Arabia) using Landsat

- Thematic Mapper data, *Tectonophysics*, 178: 309-335.
- Farebrother, R.W., 1988, *Linear least square computation*, Marcel Dekker Inc., New York.
- Freemantle, J.R., R. Pu, J.R. Miller, Calibration of imaging spectrometer data to reflectance using pseudo-invariant features, *Proceedings of the 15th Canadian Symposium on Remote Sensing*, June, 1992, Toronto, pp.452 - 455.
- Gong, P., J.R. Miller, J. Freemantle, B. Chen, 1991, Spectral decomposition of Landsat TM data for urban land-cover mapping, *14th Canadian Symposium on Remote Sensing*, pp. 458-461.
- Gong, P., R. Pu, J.R. Miller, 1992, Correlating leaf area index of ponderosa pine with hyperspectral CASI data, *Canadian Journal of Remote Sensing*, Vol. 18, No. 4, October, pp.275-282.
- Hanan, N.P., S.D. Prince, P.H.Y. Hiernaux, 1991, Spectral modelling of multicomponent landscapes in Sahel, *International Journal of Remote Sensing*, 12, pp. 1243-1258.
- Hapke, B., 1981, Bidirectional reflectance spectroscopy 1. Theory, *Journal of Geophysical Research*, 86: 3039-3054.
- Jasinski, M.F., P.S. Eagleson, 1989, The structure of red-infrared scattergrams of semivegetated landscapes, *IEEE Transactions on Geoscience and Remote Sensing*, 27, pp.441-451.
- Jansinski, M.F., P.S. Eagleson, 1990, Estimation of subpixel vegetation cover using red infrared scattergrams, *IEEE Transactions on Geoscience and Remote Sensing*, 28,

pp.153-267.

Johnson, P.E., M.O. Smith, S. Taylor-Georage, J.B. Adams, 1983, A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures, *Journal of Geophysical Research*, 88: 3557-3561.

Klován, J.E., A.T. Miesch, 1975, Extended CABFAC and QMODEL computer programs for Q-Mode factor analysis of compositional data, *Computer and Geosciences*, pp.161-178.

Kruse, F.A., K.S. Kierein-Young, J.W. Boardman, 1990, Mineral mapping at Caprite, Nevada with a 63-channel imaging spectrometer, *Photogrammetric Engineering & Remote Sensing*, 56(1): 83-92.

Lawson, C.L., R.J. Hanson, 1974, *Solving least squares problems*, Prentice-Hall, Inc., Englewood Cliffs, N.J.

Liang, S., A.H. Strahler, X. Li, 1991, Simultaneous inversion of subpixel proportions and signatures of mixed pixels: two components, *ACSM-ASPRS Fall convention*.

Mather, P.M., 1987, *Computer processing of remotely-sensed images*, Wiley Chichester.

Mustard, J.F., C.M. Pieters, 1989, Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra, *Journal of Geophysical Research*, 94: 13619-13634.

Mustard, J.F., C.M. Pieters, 1987, Quantitative abundance estimates from bidirectional reflectance measurements, *Journal of Geophysical Research*, 92: E617-E626.

Press, W.H., W.T. Vetterling, B.P. Flannery, S.A. Teukolsky, *Numerical recipes: the*



- art of scientific computing, Cambridge University press 1986.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Neumerical recipes in C: the art of scientific computing, Cambridge University press 1988, 1992.
- Rivard, B., 1989, Mapping ophiolitic melanges of the central eastern desert of Egypt using mixing model applied to Landsat Thematic Mapper data, 7th Thematic Conference on Remote Sensing fro Exploration Geology, Calgary, Alberta, Canada, pp. 847-859.
- Rubin, T.D., 1993, Spectral mapping with imaging spectrometers, Photogrammetric Engineering & Remote Sensing, Vol. 59, No.2, February 1993, pp.215-220.
- Sabol, Jr.D.E., J.B. Adams, M.O. Smith, 1990, Predicting the spectral detectability of surface materials using spectral mixture analysis, Proceedings of IGARSS'90, College Park, Maryland, USA, pp. 967-970.
- Settle, J.J., N.A. Drake, 1993, Linear mixing and the estimation of groundcover proportions, International Journal of Remote Sensing, Vol. 14, No. 6, pp. 1159-1177.
- Shipman, H., J.B. Adams, 1987, Detectability of minerals on desert alluvial fans using reflectance spectra, Journal of Geophysical Research, 92: 10391-10402.
- Siegal, B.S., A.R. Gillespie, J.E. Skaley, 1980, Remote Sensing in Geology, Wiley, New York, 120p.
- Singer, R.B., T.B. Mccord, 1979, Mars: large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance, Proceedings of the 10th Lunar and Planetary Science Conference, pp.1835-1848.

- Slater, P.N., F.J. Doyle, N.L. Fritz, R. Welch, 1983, Photographic systems for remote sensing, *Manual of Remote Sensing*, Second Edition, American Society of Photogrammetry, Vol. 1, Chap. 6, pp. 231 - 291.
- Smith, M.O., S.L. Ustin, J.B. Adams, A.R. Gillespie, 1990, Vegetation in deserts: 1. A regional measure of abundance from multispectral images, *Remote Sensing of Environment*, 31: 1-25.
- Smith, M.O., P.E. Johnson, J.B. Dams, 1985, Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis, *Journal of Geophysical Research*, 90:C797-C804.
- Staenz, K., 1992, A decade of imaging spectrometry in Canada, *Canadian Journal of Remote Sensing*, Vol. 18, No. 4, October, pp.187-197.
- Stoer, J. and R. Bulirsch, Introduction to numerical analysis, Springer-Verlag New York Inc., 1980.
- Williams, D.J., A. Royer, N.T. O'Neill, S. Achal, G. Weale, Reflectance extraction from CASI spectra using radiative transfer simulations and a rooftop irradiance collector, *Canadian Journal of Remote Sensing*, Vol. 18, No. 4, October, 1992, pp.251-261.
- Zacharias, M., O. Niemann, G. Borstad, An assessment and classification of a multispectral bandset for the remote sensing of intertidal seaweeds, *Canadian Journal of Remote Sensing*, Vol. 18, No. 4, Oct. 1992, pp. 263 - 274.
- Zhang, A., P. Gong, J.R. Miller, 1993, The effects of autocorrelation and noise on image pixel decomposition through spectral mixing analysis, 16th Canadian Symposium on Remote Sensing, pp. 847 - 850.