

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Field Spectroradiometer Data:
Acquisition, Organisation, Processing and Analysis
on the Example of New Zealand Native Plants

A thesis presented in fulfilment of
the requirements for the degree of

Master of Philosophy

in

Earth Science

at Massey University, Palmerston North,
New Zealand.



Andreas Hueni

2006

Abstract

The purpose of this research was to investigate the acquisition, storage, processing and analysis of hyperspectral data for vegetation applications on the example of New Zealand native plants. Data covering the spectral range 350nm-2500nm were collected with a portable spectroradiometer.

Hyperspectral data collection results in large datasets that need pre-processing before any analysis can be carried out. A review of the techniques used since the advent of hyperspectral field data showed the following general procedures were followed:

1. Removal of noisy or uncalibrated bands
2. Data smoothing
3. Reduction of dimensionality
4. Transformation into feature space
5. Analysis techniques

Steps 1 to 4 which are concerned with the pre-processing of data were found to be repetitive procedures and thus had a high potential for automation. The pre-processing had a major impact on the results gained in the analysis stage. Finding the ideal pre-processing parameters involved repeated processing of the data.

Hyperspectral field data should be stored in a structured way. The utilization of a relational database seemed a logical approach. A hierarchical data structure that reflected the real world and the setup of sampling campaigns was designed. This structure was transformed into a logical data model. Furthermore the database also held information needed for pre-processing and statistical analysis. This enabled the calculation of separability measurements such as the JM (Jeffries Matusita) distance or the application of discriminant analysis.

Software was written to provide a graphical user interface to the database and implement pre-processing and analysis functionality.

The acquisition, processing and analysis steps were applied to New Zealand native vegetation. A high degree of separability between species was achieved and using independent data a classification accuracy of 87.87% was reached. This outcome required smoothing, Hyperion synthesizing and principal components transformation to be applied to the data prior to the classification which used a generalized squared distance discriminant function.

The mixed signature problem was addressed in experiments under controlled laboratory conditions and revealed that certain combinations of plants could not be unmixed successfully while mixtures of vegetation and artificial materials resulted in very good abundance estimations.

The combination of a relational database with associated software for data processing was found to be highly efficient when dealing with hyperspectral field data.

Acknowledgements

I would firstly like to thank my supervisor Mike Tuohy for his time and thoughtful advice throughout the preparation of this thesis.

I would also like to thank Mike Hedley and Bambang H. Kusumo for their valuable input in terms of end user requirements for spectral processing software.

Table of Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
2 Literature Review	3
2.1 Hyperspectral Remote Sensing	3
2.2 Hyperspectral Sensors	4
2.2.1 Field Spectroradiometers	4
2.2.2 Airborne Hyperspectral Sensors	6
2.2.3 Spaceborne Hyperspectral Sensors	7
2.3 Hyperspectral Data	7
2.3.1 Overview and Principles	7
2.3.2 Data Processing	11
2.3.3 Analysis	16
2.4 Spectral Libraries and Spectral Databases	24
2.4.1 Spectral Libraries	24
2.4.2 Spectral Databases	25
2.5 Intermediate Conclusions	26
3 Methods	27
3.1 Acquisition and Storage of Field Data	27
3.1.1 Dataflow Overview	27
3.1.2 ASD FieldSpecPro	27
3.1.3 Study Sites	27
3.1.4 Structure of Field Data	28
3.1.5 Acquisition of Field Data	28
3.1.6 Species	29
3.2 Spectral Database	30
3.2.1 Spectral Database Model	30
3.2.2 Spectral Database Implementation	40
3.3 A Spectral Data Management and Processing Software	40
3.3.1 Programming Language, Libraries and Environment	41
3.3.2 Software Architecture	41
3.4 Data Processing	46
3.4.1 Waveband Filtering	46
3.4.2 Smoothing	46
3.4.3 Synthesizing of other Sensors	48
3.4.4 Derivative Calculation	52
3.4.5 Feature Space Transformation	53
3.5 Statistical Analysis	56
3.5.1 Classification	56
3.5.2 Discriminant Analysis	57

3.5.3	Separability Analysis	57
3.5.4	Most Discriminating Bands	57
3.6	Mixed Spectral Signatures.....	58
4	Results	63
4.1	Spectral Properties of New Zealand Native Plants.....	63
4.1.1	Smoothing.....	64
4.1.2	Sensor Synthesizing.....	68
4.1.3	Derivative Calculation	73
4.1.4	Feature Space Transformation	74
4.1.5	Statistical Analysis.....	81
4.2	Mixed Spectral Signatures.....	92
4.2.1	Paper/Plant Mixture	92
4.2.2	Paper/Plastic/Plant Mixture	93
4.2.3	Three plant mixture.....	95
4.2.4	Positional Dependence of Paper/Plastic Mixtures	97
4.2.5	Probe Rotation	98
5	Discussion.....	99
5.1	Collection of Spectral Data of New Zealand Native Plants.....	99
5.2	Spectral Databases.....	99
5.3	Spectral Processing Chain	99
5.4	Processing Speed of Smoothing Operations.....	100
5.5	Data Reduction.....	100
5.6	Discriminative Power of Feature Spaces.....	100
5.7	Discrimination and Classification	101
5.8	Principal Component Analysis	101
5.9	Linear Transformations	101
5.10	Most Discriminating Bands	102
5.11	Separability Analysis and Discriminant Analysis.....	102
5.12	Spectral Unmixing.....	102
5.13	Atmospheric Correction of Hyperion Imagery	103
6	Conclusion	105
7	Bibliography	107
8	Appendix	113
8.1	SpectraProc Graphical User Interface	113

List of Figures

Figure 1: Interaction between energy source, object and sensor	3
Figure 2: Examples of spectral signatures acquired in the preliminary stage of the project	4
Figure 3: Examples for spectral space and feature space (Data from a preliminary stage of this study)	8
Figure 4: Probability distributions in a 2d feature space (Richards, 1993)	9
Figure 5: An example of a data distribution in a 2d feature space, showing independent samples of a class and their mean	9
Figure 6: Two dimensional data with little correlation (a) and high correlation (b) (Richards, 1993).....	10
Figure 7: Data reduced to mean values (left) and data including 2nd order statistics information and showing regression lines for each class (right).....	10
Figure 8: An example of a spectrum showing water band noise in 3 wavelength ranges	11
Figure 9: An example of a mixed pixel (linear mixture model)	22
Figure 10: Dataflow and involved hardware	27
Figure 11: Hierarchical directory structure	28
Figure 12: Database model overview at entity level	31
Figure 13: ERD of the entities study, species, site and spectrum.....	31
Figure 14: ERD of the entities study and waveband_filter and waveband_filter_range	33
Figure 15: ERD of library, statistic, feature_space, sensor and associated entities.....	37
Figure 16: ERD of the entities species and mixture	40
Figure 17: File system interfaces.....	41
Figure 18: Non-MFC classes.....	43
Figure 19: Spectral data processing cascade	45
Figure 20: An example of pre and post filtering of noise bands	46
Figure 21: A smoothed signature of <i>Pittosporum eugenoides</i> before and after the removal of smoothing artefacts	48
Figure 22: Ratios for Landsat7 ETM+ band 1.....	49
Figure 23: Gaussian curve illustrating the FWHM measure	50
Figure 24: Sensor response functions for Hyperion sensor elements 8 and 9 and the FWHM of band 8 ..	52
Figure 25: Illustration of discrete reflectance values ρ and interpolated linear curves to form a continuous reflectance curve.....	54
Figure 26: General mixing setup	59
Figure 27: Mixture segments.....	59
Figure 28: Rotational positions of the bare fibre	62
Figure 29: Features of a vegetation curve	63
Figure 30: Mean Hyperion synthesized spectra of NZ native plants.....	64
Figure 31: Effects of variations of smoothing filter size and polynomial order on smoothed spectra	65
Figure 32: RMSE of raw minus smoothed spectra.....	66
Figure 33: Noise spectra of different Savitzky-Golay filter settings (raw minus filtered spectra).....	67
Figure 34: Red-NIR region of the noise spectra after filtering with order 3 smoothing filters (raw minus filtered spectra).....	67

Figure 35: Raw and Hyperion synthesized spectra of <i>Pittosporum eugenioides</i>	69
Figure 36: Raw and Hyperion synthesized in NIR and SWIR2 parts of the spectrum	69
Figure 37: Noise spectrum of <i>Pittosporum eugenioides</i> (raw minus Hyperion synthesized)	69
Figure 38: Raw and decimated by factor 10 and 5 spectra of <i>Pittosporum eugenioides</i> (offset for clarity)	70
Figure 39: Raw and decimated by factor 10 and 5 (NIR part of the spectrum)	70
Figure 40: Raw and decimated by factor 10 and 5 (SWIR2 part of the spectrum)	71
Figure 41: Noise spectrum of <i>Pittosporum eugenioides</i> (Raw minus decimated by factor 5)	71
Figure 42: Noise spectrum of <i>Pittosporum eugenioides</i> (Raw minus decimated by factor 10)	71
Figure 43: RMSE of Hyperion synthesizing and Decimation 10 and 5.....	72
Figure 44: Landsat7 ETM+ and Hyperion signatures.....	72
Figure 45: Derivatives based on different pre-processing and derivative calculations.....	73
Figure 46: DGVI regions overlaid with a typical plant spectrum (<i>Pittosporum eugenioides</i>).....	74
Figure 47: Frequency of statistically significant differences of DGVI and their dependence on pre- processing	76
Figure 48: Example of the discrimination of species by DGVI (calculation based on Hyperion synthesized data).....	77
Figure 49: Frequency of statistically significant differences of NDVI and their dependence on pre- processing	78
Figure 50: Scree plots of eigenvalues (Hyperion synthesized and Decimation 5 data)	79
Figure 51: Histogram of statistically significant differences between species pairs for PC transformed Hyperion synthesized data	80
Figure 52: Average PC Factor Loadings for the first five components	80
Figure 53: PC factor loadings for PC1 and PC2. The mean reflectance of <i>Pittosporum eugenioides</i> is displayed to relate the factors to typical vegetation reflectance features	81
Figure 54: Graphical comparison of the number of bands with frequencies higher than the threshold (mean + standard deviation) per spectrum segment.....	89
Figure 55: Histogram of the statistically significant differences in reflectance calculated using raw data of all library relevant species. The mean reflectance of <i>Pittosporum eugenioides</i> is displayed to relate the frequency to typical vegetation reflectance features.	90
Figure 56: Histogram of the statistically significant differences in reflectance calculated using Hyperion synthesized data of all library relevant species. The mean reflectance of <i>Pittosporum eugenioides</i> is displayed to relate the frequency to typical vegetation reflectance features....	90
Figure 57: Histogram of the statistically significant differences in the first derivative of reflectance calculated using Hyperion synthesized data of all library relevant species. The mean reflectance of <i>Pittosporum eugenioides</i> is displayed to relate the frequency to typical vegetation reflectance features.....	91
Figure 58: Spectral curves of mixtures of vegetation and paper.....	92
Figure 59: Scatterplot of reflectance values for simulated Landsat bands 1 vs 7 for mixtures of paper and vegetation.....	93

Figure 60: Spectral curves of mixtures of vegetation, paper and plastic. Endmembers are plotted in thick lines	94
Figure 61: Scatterplot of reflectances for simulated Landsat bands 1 vs 7 for mixtures of vegetation, paper and plastic. Endmembers are plotted in bigger symbols	94
Figure 62: Spectral curves for mixtures of three plants. Endmembers are plotted as dashed lines.	95
Figure 63: Scatterplot of reflectances for simulated Landsat bands 1 vs 7 for mixtures of 3 plants	96
Figure 64: Positional dependence for paper/plastic mixtures.....	97
Figure 65: First derivative of 50% paper/plastic mixtures	98
Figure 66: Probe Rotation on paper/plastic mixture.....	98
Figure 67: Spectral curves for the endmembers of the 3 plant mixing experiment.....	103
Figure 68: Screen capture of SpectraProc	113

List of Tables

Table 1: Widely used airborne hyperspectral sensor systems	7
Table 2: Collected species.....	29
Table 3: Short description of MFC derived classes	42
Table 4: Short description of non-MFC derived classes	43
Table 5: Mixtures of paper and kawakawa	59
Table 6: Mixtures of paper, plastic and kawakawa	60
Table 7: Mixtures of kawakawa, lemonwood and karaka.....	61
Table 8: Paper/plastic mixtures and positions.....	61
Table 9: Mean frequencies of statistically significant differences in species pairs for DGVI's calculated for differing pre-processing parameters.....	75
Table 10: NTBI and mean frequencies of statistically significant differences in species pairs	77
Table 11: First 18 components of the eigenanalysis of Hyperion-synthesized and Decimation by 5 data ...	78
Table 12: The 10 principal components with the highest significances (according to the Wilcoxon test) ordered by significance	79
Table 13: Classification results for calibration and independent datasets (accuracy in percentage).....	82
Table 14: Error matrix for DGVI's of smoothed Hyperion synthesized data classified using the quadratic distance discriminant function	83
Table 15: Producer and user accuracy statistics for DGVI's of smoothed Hyperion synthesized data classified by the quadratic distance discriminant function (accuracy in percentage).....	84
Table 16: Statistics of separability analysis	85
Table 17: JM distances (upper triangle) and B distances (lower triangle) between species in DGVI feature space.....	86
Table 18: Significance statistics.....	88
Table 19: Number of bands with frequencies higher than mean plus one standard deviation	88
Table 20: Unmixing results for vegetation/paper mixtures.....	93
Table 21: Unmixing results for vegetation/paper/plastic mixtures	95
Table 22: Unmixing results for 3 plant mixtures	96

1 Introduction

Spectroradiometry has become increasingly popular in the last few years. The technology has advantages over conventional techniques, allowing the non destructive sampling of objects and enabling users to gain critical information more quickly and cheaply. The operation of the equipment tends to be relatively easy and data are collected quickly. However, the interpretation of these data is not dealt with quite as easily. The main issue when dealing with hyperspectral data is their dimensionality. Hyperspectral data are more complex than previous multispectral data and different approaches for data handling and information extraction are needed (Vane and Goetz, 1988; Landgrebe, 1997).

The Institute of Natural Resources, Massey University, had acquired a spectroradiometer built by ASD (Analytical Spectral Devices) and a study utilizing this instrument was considered to be of interest.

The goals of this study were: Enhance the knowledge of the Institute in the field of hyperspectral remote sensing utilizing the recently acquired FieldSpecPro spectroradiometer; study the processes of field data acquisition, data processing and analysis; create a spectral database of New Zealand native vegetation; analyze the spectral separability of New Zealand native vegetation; investigate the problem of mixed signatures; suggest a basis for the classification of land cover using Hyperion data

While the main focus of this research was on hyperspectral data, the simulation of Landsat7 ETM+ was also undertaken, mainly to provide a basis for further investigation of the problem of atmospheric correction. Landsat7 imagery of New Zealand has been successfully corrected for atmospheric influences by Landcare Research, Palmerston North.

During the project, support was given to a Soil Science PhD study at Massey University and to a study on soils and pastures at Landcare Research, Palmerston North, in terms of sharing expertise, collecting data and subsequent processing. These collaborations led to further development of the database and processing requirements and widened the focus of this study to include data from soil and pasture studies. As a result of this, a section on correlation of spectral data with other physical properties was added to the literature review. It serves to complete the picture of the analysis that can be applied to hyperspectral data. The above mentioned collaborations also supported the hypothesis that tools for efficient data handling, organisation and processing were of high interest to scientists.

