UNIVERSITY OF SOUTHAMPTON

Band selection using hyperspectral data from airborne and satellite sensors

Michael Riedmann, Dipl.-Ing., M.Sc. (distinction)

DOCTOR OF PHILOSOPHY

Faculty of Engineering, Science and Mathematics
School of Geography
September 2003

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
GEOGRAPHY

Doctor of Philosophy

BAND SELECTION USING HYPERSPECTRAL DATA FROM AIRBORNE AND SATELLITE SENSORS

by Michael Riedmann

Hyperspectral data offer refined spectral discrimination of ground targets, but come at a substantial cost. For some sensors, the number of spatial pixels (swath width) needs to be reduced to acquire a large number of bands. In addition, coarser spatial resolution is required to achieve enough signal from narrow bands. This study aimed to investigate whether it was possible to reduce the number of bands and broaden their widths, while achieving the same or higher application accuracy as with hyperspectral data.

Three innovative band selection methods were developed as tools for this investigation. They were designed primarily for Maximum Likelihood Classification (MLC) applications, but their use with respect to other applications was discussed. All algorithms aimed to optimise the band location, width and number with respect to the MLC accuracy for the given classification task. The supervised band selection (SBS) algorithm is based on conventional feature selection techniques, while the unsupervised band selection (UBS) method aims to decorrelate the band set. The unsupervised clustering-based SBS (CSBS) algorithm uses the SBS, but with classes being defined by clustering.

The three approaches were evaluated on real data sets. All algorithms gave physically meaningful band sets, which achieved similar or higher MLC accuracies than band sets of current airborne and satellite sensors. The sub-optimality of the SBS bands was found to be least (7%) for sets with maximum three bands. The band number criteria were shown to be effective estimates of the intrinsic data dimensionality, although some subjectivity remains. Only SBS may be used to test whether narrow band data have a significant advantage over broad band data. UBS depends on dark image data for band expansion and requires each band to be normally distributed, which is only justified if the scene is made up of a single material type. CSBS has the drawback of producing inconsistent results depending on the initialisation and parameter settings of the clustering routine.

The methods can be applied with *programmable* sensors in a repeat-pass fashion: Band selection may be performed on hyperspectral data acquired over a representative part of the scene. Then, multispectral data may be collected over the same scene with the optimised band set under similar solar and atmospheric conditions. Both UBS and CSBS may be employed in-flight. For *non-programmable* sensors, a more generic band set is sought for a given classification scheme, which needs to be optimised to a large number of scenes.

The author believes that current data acquisition is inefficient in that spectrally redundant data are collected with most imaging spectrometers, often using narrow band data where this is not necessary. Application of the above band selection methods to real data sets showed that for all three classification tasks, the number of bands to acquire could be reduced dramatically with a maximum loss of 5% in MLC accuracy, and for two out of the three tasks, coarsening the spectral sensor resolution may be justified. This would allow collecting supplementary data and refining the sensor's spatial resolution. Coupled with algorithms to optimise other acquisition parameters, the band selection methods developed in this thesis lead the way towards an intelligent remote sensing expert system for data acquisition.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisor Prof. Edward Milton, for his unerring support, guidance and inspiration at all times during my work on this thesis. 'Ted' gave me the opportunity to work part-time in the NERC Equipment Pool for Field Spectroscopy (EPFS), which allowed me to receive a full grant as a European student. At this point, I would like to thank the University of Southampton, which funded the main part of my research.

I am also grateful to Prof. Peter Atkinson and Dr. Jonathan Forster for checking on statistics and formulae, and for all the constructive comments they made. I am also thankful to NERC for supplying the HyMAP and CASI-2 data acquired over the New Forest and the Tregaron bog, respectively. In addition, I thank the Environment Agency for supplying the CASI data acquired over the River Severn.

I also would like to thank Karen Anderson for her assistance during field work, and her advice for the New Forest training site selection. Also a great thank you to Eloise Peters for her assistance during field work, to Jenny Schulz for her advice in the training site selection for the Tregaron site, and to Bill Damon for his technical support.

I am also grateful to the following people for their general assistance at the School of Geography, University of Southampton: Dr. Charles Kerr, Dr. Jim Milne, Sally Kelday, Dr. Liz Rollin, John Hurst, Tim Aspen, Andy Vowles, Roz Campbell, and Prof. David Martin. Thanks, too, to Ilse Steyl, Jana Fried, Matthew Wilson, Dr. Sally Priest, Reno Choi, Nicholas Odoni, Marie Cribb and all other members of the Graduate School, School of Geography, University of Southampton.

Finally, thanks to my parents, family and friends for their continuous support.

ABBREVIATIONS

ACORN	Atmospheric Correction Now
AISA	Airborne Imaging Spectrometer
AMD	Absolute mean difference
ANN	Artificial neural network
ARSF	NERC Airborne Remote Sensing Facility
ARVI	Atmospherically resistant VI
ATREM	Atmosphere Removal Program
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BNSC	British National Space Centre
CA	Canonical analysis
CASI / CASI-2	Itres Instruments Compact Airborne Spectrographic Imager
CCD	Charge-coupled device
CHRIS	Compact High Resolution Imaging Spectrometer
CSBS	Clustering-based SBS
DA / DAFE	Discriminant analysis (feature extraction)
DBFE	Decision boundary feature extraction
DN	Digital number
DP	Discriminant power
EA	Environment Agency
EL	Empirical line
EM	Endmember
EPFS	NERC Equipment Pool For Field Spectroscopy
EVI	Enhanced VI
FA	Factor analysis
FOV	Field-of-view
FWHM	Full-width-half-maximum
GA	Genetic algorithm
GERIS	Geophysical Environmental Research Imaging Spectrometer
GIFOV	Ground-projected IFOV
GMES	Global Monitoring of Environment and Security
GVI	Green VI
HyMAP	Hyperspectral Mapper
IARR	Internal average relative reflectance
ICA	Independent component analysis
ID	Intrinsic dimensionality
IDD	Intrinsic discriminant dimensionality
IFOV	Instantaneous FOV
ISODATA	Iterative Self-Organising Data Analysis

KL	Karhunen-Loève
LAI	Leaf-area index
LC	Least correlated band pair first (UBS algorithm option)
LN	Least noisy band first (UBS algorithm option)
LOWTRAN	Low Resolution Atmospheric Radiance And Transmittance
LSMM	Linear spectral mixture model
LSU	Linear spectral unmixing
MAF	Minimum/maximum autocorrelation factors
MDS	Multidimensional scaling
MEIS	Multi-detector Electro-optical Imaging Scanner
MLC	Maximum likelihood classifier / classification
MNF	Maximum noise fraction
MODTRAN	Moderate resolution atmospheric radiance and transmittance model
MSPCA	Modified stepwise PCA
MTF	Modulation transfer function
NERC	Natural Environment Research Council, UK
NIR	Near infrared
NPL	National Physical Laboratory
OIF	Optimal index factor
OMBVI	Optimum multiple-band VI
OSP	Orthogonal subspace projection
PC	Principal component
PCA	Principal components analysis
PCVI	Principal component VI
PI	Projection index
pixel	Picture element
PMAMA	Proportion of maximum achievable MLC accuracy
PMATD	Proportion of maximum achievable Transformed Divergence
PN	Priority number
PP	Projection pursuit
PPDA	PP pre-processing and DAFE combined
PPDB	PP pre-processing and DBFE combined
PPFS	Projection pursuit feature selection
PSF	Point spread function
RBD	Relative absorption band-depth
RGB	Red-green-blue
RMSE	Root mean square error
ROI	Region of interest
ROSIS	Reflective Optics System Imaging Spectrometer
RTGC	Radiative transfer ground calibration
RTM	Radiative transfer model

SA	Simulated annealing
SAC	Special area of conservation
SAM	Spectral angle mapper
SBaFS	Sequential backward floating selection
SBaS	Sequential backward selection
SBS	Supervised band selection algorithm
SFF	Spectral feature fitting
SFFS	Sequential forward floating search
SFS	Sequential forward selection
SI	Système International
SNR	Signal-to-noise ratio
SPA	Special protection area
SSSI	Site of special scientific interest
SWIR	Short-wave infrared
TBVI	Two-band VI
TD	Transformed divergence
TOAVI	Top-of-atmosphere VI
UBS	Unsupervised band selection algorithm
ValDEOS	Validation of data for Earth observation services
VI	Vegetation index
VIS	Visible

1 Introduction

In this research project, new band selection methods for imaging spectrometers were developed and evaluated. In this introductory chapter, the rationale behind the research is explained, followed by a statement of the specific research objectives. The chapter closes with a description of the structure of the thesis.

1.1 Rationale

Remote sensing denotes the process of measuring physical characteristics of distant objects. In a typical remote sensing system, electromagnetic energy reflected and emitted from remote objects on the Earth's surface is recorded at different wavelengths by a sensor onboard an aircraft or satellite. From these measurements information may be inferred to identify or characterise these objects, for example, in terms of their physical or chemical properties.

In this thesis, the author will limit himself to optical solar-reflective passive remote sensing with the Sun as illumination source and the Earth's surface as target. The optical solar-reflective wavelength range includes the visible (VIS, 0.38 - 0.76 µm), the near infrared (NIR, 0.76 - 1.5 µm) and the short-wavelength infrared (SWIR, 1.5 – 3 µm) (Chen, 1997).

Remotely sensed radiation is, after atmospheric correction, a function of the location, time, wavelength and viewing geometry of a given resolution element (Verstraete and Pinty, 1992). In this thesis, the analysis was based exclusively on extracting information from detectable changes in the wavelength parameter of the measured radiance.

Until recently, many subtle changes in radiation with wavelength were not detectable with remote sensing (Curran, 1994). Available sensors sampled the spectrum with broad bands of around 0.1 - 0.2 µm in width (e.g. Landsat MSS, TM and SPOT), while most terrestrial materials are characterised by narrow absorption features typically 0.02 - 0.04 µm in width (Hunt, 1980).

Imaging Spectrometry

In the last decades, advances in sensor technology have produced a new generation of airborne and satellite sensors, called imaging spectrometers. Imaging spectrometers record a continuous radiance spectrum for each image pixel via numerous contiguous and narrow bands (see figure 1.1). Their spectral resolution is approximating closely the one in spectroscopy, typically 0.01 to 0.02 µm. That is, *in contrast to broadband sensors, imaging spectrometers allow the detection of narrow*

material-specific absorption features, resulting in increased mapping capacities. In addition, imaging spectrometer data enable the extraction of reflectance spectra at a pixel scale, which are directly comparable to reflectance spectra gathered in the field or laboratory. For a historical review of imaging spectrometry, the reader is referred to Kruse *et al.* (1990) and Van der Meer *et al.* (2001).

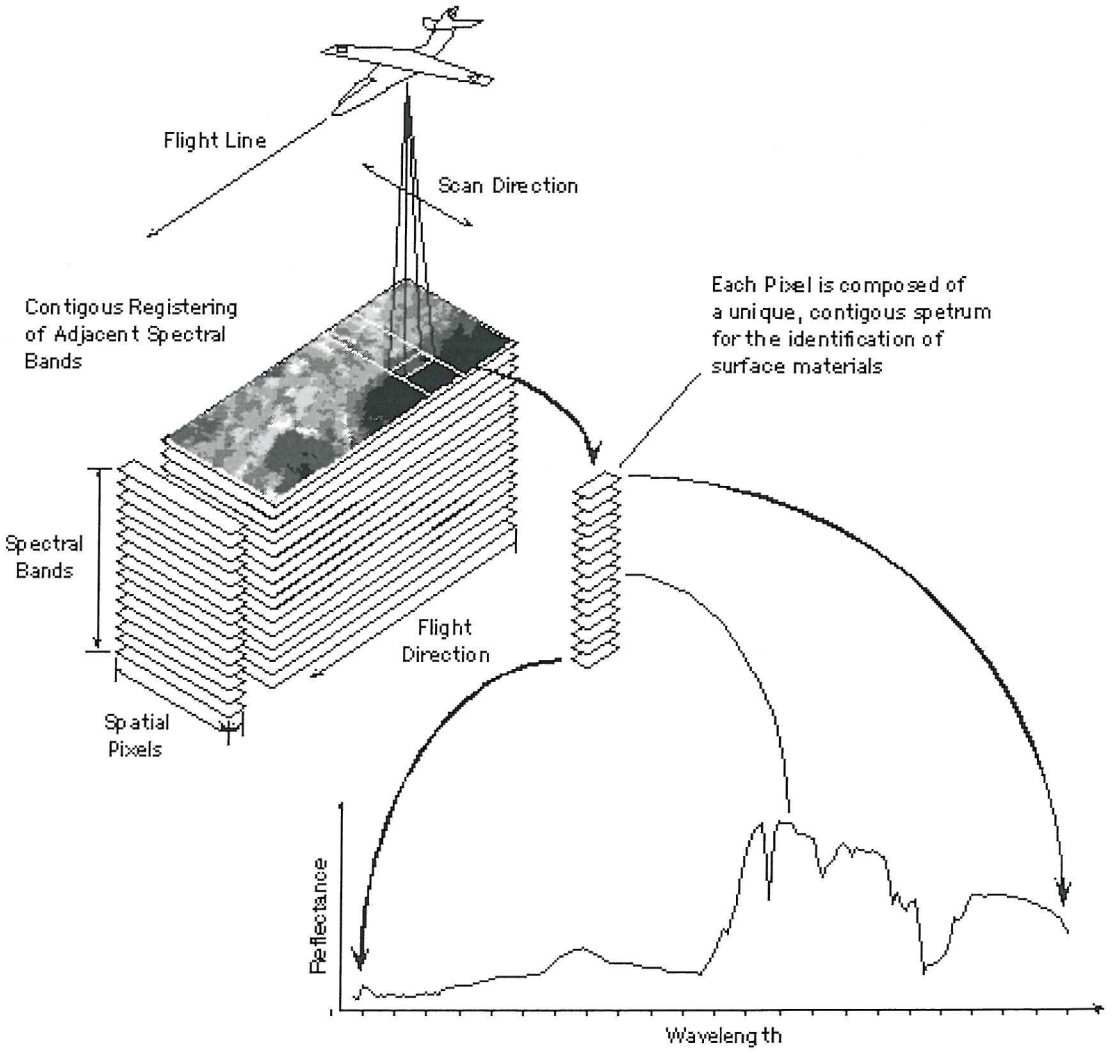


Figure 1.1: Imaging spectrometry data acquisition (source: http://www.apex-esa.org).

As with non-imaging spectrometers, the incoming radiation is usually dispersed by optical elements such as prisms, diffraction gratings, or in a special case, interferometers onto a detection device (usually a charge-coupled device, CCD, array). The spatial extent of the image is acquired using across-track scanning techniques (e.g. pushbroom, whiskbroom) and their inherent movement along-track.

In contrast to field and laboratory spectrometers which are usually employed on the ground, imaging spectrometers are typically airborne or spaceborne. That is, their data have a lower signal-to-noise

ratio (SNR) due to the distant radiation sources and the limited available dwell time of the sensor over the target. Since their field of view (FOV) is typically in the order of several meters for airborne sensors, the measurement sample within it is most likely to be heterogeneous. Being remotely-sensed image data, imaging spectrometer data may also be affected by varying topography, viewing geometry and atmospheric noise (Curran, 1994; Tsai and Philpot, 1998). In addition, depending on the sensor design, the bandwidth may vary over the spectrum.

Different names have been created to describe this new area of remote sensing: imaging spectrometry ('measuring'), imaging spectroscopy ('seeing') and hyperspectral remote sensing ('many bands') (Van der Meer and De Jong, 2001). Usually, an imaging spectrometer is calibrated to perform absolute measurements in one of the internationally recognised Système International (SI) base units, such as spectral radiance. It should then be referred to as 'imaging spectroradiometer' (see definition of 'spectroradiometer' in Wyatt, 1978).

A calibration to spectral radiance units is especially important when comparing measurements between different instruments or from different acquisition times (Mather, 1999). In addition, some radiative transfer models need physical quantities such as upwelling radiance as input variables to calculate an atmospheric correction that relates apparent at-sensor radiance to the radiance of the ground feature (Slater, 1984).

Imaging Spectrometers

After the commercialisation of airborne imaging spectrometers in the early 1980s, numerous instrument types have been developed. A good compilation of airborne imaging spectrometers can be found in Curran (1994) and Van der Meer *et al.* (2001). Table 1.1 and 1.2 list some of the airborne and spaceborne VNIR sensors that have been operated or planned since 2002.

Table 1.3 shows specifications of two airborne imaging spectrometer systems, the CASI-2 and the HyMAP, the data of which will be used in this thesis. Unlike the HyMAP sensor, the CASI-2 is programmable with respect to band location, width and number. Due to data recording rate limits of the CASI-2 hard drive subsystem, a trade-off exists between a maximal swath width and high spectral resolution. The full swath width (512 imaging pixels) is recorded in CASI-2's spatial mode, however at the cost of a reduced number of maximum 19 spectral bands (ITRES, 2001). In contrast, all 288 bands may be recorded in CASI-2's enhanced spectral mode, but with a much narrower swath (101 imaging pixels). For a more detailed summary see Riedmann (2003).

While airborne imaging spectrometers have a very fine spatial resolution (up to 1 m, e.g. CASI-2), their spaceborne counterparts possess coarser spatial resolution (typically 30 - 500 m), but similar spectral resolution. Table 1.4 presents some examples of currently orbiting spaceborne imaging spectrometers.

Table 1.1: Airborne VNIR Sensors operating in year 2002 and beyond. Hyperlinks were last verified on 10/10/2002.

Acronym	Name	Manufacturer	Operator	Instrument/Calibration Hyperlink
AAHIS-3	Advanced Airborne Hyperspectral Imaging System	STI Government Systems		http://www.sti-government.com/Optical_Imaging_Systems/AAHIS.html
AirMISR	Airborne Multi-angle Imaging Spectro-Radiometer	JPL	NASA	http://www-misr.jpl.nasa.gov/mission/air.html Calibration http://www-misr.jpl.nasa.gov/mission/valwork/mivcalres.html
AISA	Airborne Imaging Spectrometer for Applications	Specim Ltd.	Specim Ltd.	CaliGeo post-processing software http://www.specim.fi/products-aisa.html Calibration technique (CARSTAD) http://carstad.gsfc.nasa.gov/topics/erim99web.html
AMSS	Airborne Multispectral Scanner	SpecTerra Systems Pty Ltd (sTs)		http://www.specterra.com.au/amss_frame.html
APEX	Airborne PRISM Experiment		ESA	http://www.apex-esa.org/ System > Scanner Specifications > System Recommendations
ASAS	Advanced Solid State Array Spectroradiometer	NASA	NASA	http://asas.gsfc.nasa.gov/
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer	NASA, JPL	NASA, Ames	http://aviris.jpl.nasa.gov/ AVIRIS calibration (JPL publication) http://popo.jpl.nasa.gov/docs/aviris87/I-VANE1.PDF http://popo.jpl.nasa.gov/docs/workshops/90_docs/1.PDF http://www.optics.arizona.edu/rsg/
CAESAR	CCD Airborne Experimental Scanner for Applications in Remote Sensing	NLR	NLR	http://www.itc.nl/~bakker/info/rs-data/platform.html
CASI-2	Compact Airborne Spectrographic Imager	Itres Research Ltd.	NERC ARSF, EA	http://www.itres.com For calibration see chapter 3 of this thesis
DAIS 7915	Digital Airborne Imaging Spectrometer	GER Corp.	DLR / JRC	http://www.op.dlr.de/DAIS/ Calibration http://www.op.dlr.de/dais/dais-cal.htm
EPS-H	Environmental Probe System	GER Corp.		http://www.ger.com/epshman.html Calibration http://www.ger.com/epshman.html#syscal
HRIS	High Resolution Imaging Spectrometer	ESA / DSS		http://www.gisdevelopment.net/aars/acrs/1999/ts10/ts10210pf.htm
HYDICE	Hyperspectral Digital Imagery Collection Experiment	Naval Research Laboratory	ERIM	http://rsd-www.nrl.navy.mil/hydice/ Calibration http://www.optics.arizona.edu/rsg/
HYMAP	Hyperspectral Mapper	Integrated Spectronics	HyVista Corp.	http://www.intspec.com http://www.hyvista.com/hymap.html Calibration: http://www.aigllc.com/pdf/EARSEL98_HyMap.pdf
ISM	Imaging Spectroscopic Mapper	IAS and DESPA (Paris Observatory)		http://www.lpi.usra.edu/science/kirkland/Workshop1/ERARD2.PDF http://www.ias.fr/cdp/ISM/welcome.html

Table 1.1 continued.

MAIS	Modular Airborne Imaging Spectrometer	Shanghai Institute of Technical Physics		
MAMS	Multispectral Atmospheric Mapping Sensor	Daedalus Enterprise Inc.	NASA, MSFC	http://asapdata.arc.nasa.gov/dscrptns.htm
MAS / AMS	MODIS Airborne Simulator	Daedalus Enterprise Inc.	NASA, GSFC	http://mas.arc.nasa.gov/ Calibration http://www.optics.arizona.edu/rsg/
MASTER	MODIS/ASTER Airborne Simulator	Daedalus Enterprise Inc.	NASA, JPL	http://masterweb.jpl.nasa.gov/ Calibration: http://www.optics.arizona.edu/rsg/
MEIS-II	Multispectral Electro-optical Imaging Sensor	CCRS	Innotech	
MISI	Modular Imaging Spectrometer Instrument	Rochester Institute of Technology (RIT)		http://www2.rit.edu/fic/pub205c0.jsp Calibration: http://www.cis.rit.edu/research/dirs/annualreport_98/research/vis_cal/
MIVIS	Multispectral Infrared and Visible Imaging Spectrometer	SenSyTech Inc.	CNR, Rome	http://www.sensytech.com/Imaging/MIVIS.html Calibration: http://ltpwww.gsfc.nasa.gov/ISSSR-95/mivisair.htm
PHILLS	Portable Hyperspectral Imager for Low -Light Spectroscopy	Naval Research Laboratory		http://rsd-www.nrl.navy.mil/7212/phills.htm Calibration: http://rsd-www.nrl.navy.mil/7212/pdf/20020225_OE.pdf
PROBE-1		Earth Search Sciences Inc.		http://www.earthsearch.com/technology/frame_about_probe1.html
ROSIS	Reflective Optics System Imaging Spectrometer	DLR, GKSS, MBB	DLR	http://www.op.dlr.de/ne-oe/fo/rosis/home.html
TRWIS-III	TRW Imaging Spectrometer	TRW Inc.		http://www.trw.com/marketplace/main/0,1151,39_1541_415_564_578^5^578^578,00.html
VIFIS	Variable Interference Filter Imaging Spectrometer	Univ. of Dundee		http://www.somis.dundee.ac.uk/staffprofiles/e/R_SE15RA001/
WIS	Wedge Imaging Spectrometer	Hughes Santa Barbara Research Center	NASA SSC	

Table 1.2: Satellite VNIR Sensors operating in year 2002 and beyond. Hyperlinks were last verified on 10/10/2002.

Acronym	Name	Agency	Platform	Launch	Instrument/Calibration Hyperlink
AATSR	Advanced Along Track Scanning Radiometer	ESA	ENVISAT	03/2002	http://envisat.esa.int/instruments/aatsr/ on-board visible calibration system http://www.le.ac.uk/physics/research/eos/aatsr/pheapr_6.html http://envisat.esa.int/calval/cr/12092002/AATSR/
ALI	Advanced Land Imager	NASA	EO-1	11/2000	http://eo1.gsfc.nasa.gov/Technology/ALIhome1.htm Calibration http://eo1.usgs.gov/documents/ALI/SPIE_07-99c10.pdf
AMODIS	Advanced Moderate-Resolution Imaging Spectroradiometer	NASA	EOS-AM2	2004?	http://pao.gsfc.nasa.gov/gsfc/service/gallery/fact_sheets/earthsci/fs-96(07)-13.htm
ARIES	Australian Resource Information and Environment Satellite	Australia	ARIES-1,2,3	2002?	http://www.auspace.com.au/projects/aries.htm Calibration http://www.sdl.usu.edu/conferences/smallsat/proceedings/11/tech03.pdf
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer	NASA	EOS AM-1 (TERRA)	1999	http://asterweb.jpl.nasa.gov http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/viewInstrument.php?instrument=ASTER Calibration http://www.optics.arizona.edu/rsg/ http://edcdaac.usgs.gov/aster/asterprocessing.html
ATSR	Along Track Scanning Radiometer	ESA	ERS-1 ERS-2	1991 1995	
AVHRR	Advanced Very High Resolution Radiometer	NASA	NOAA-J NOAA-K NOAA-L NOAA-M NOAA-N	1994 05/1998 2000 2002 2003?	http://fermi.jhuapl.edu/avhrr/index.html http://www.itc.nl/~bakker/noaa.html http://www.agrecon.canberra.edu.au/Products/Satellite_Imagery/NOAA/NOAA.htm Calibration http://www.optics.arizona.edu/rsg/ http://www.ccrs.nrcan.gc.ca/ccrs/rd/ana/calval/calhome_e.html http://www.pcigeomatics.com/cgi-bin/pcihlp/AVHRRAD
AVHRR / HRIS	Advanced Very High Resolution Radiometer		NOAA-N	2004/ 2008?	http://www.itc.nl/~bakker/noaa.html
AVNIR-2	Advanced Visible and Near Infrared Radiometer type 2	NASDA (Japan)	ALOS	2004?	http://www.eorc.nasda.go.jp/ALOS/about/2avnir2.htm http://www.eoc.nasda.go.jp/guide/satellite/sendata/avnir_e.html Calibration http://www.eorc.nasda.go.jp/ALOS/da/2example.htm
CCD	High Resolution CCD Camera	CAST / INPE	CBERS-1 CBERS-2	1999 2003?	http://www.dgi.inpe.br/html/eng/cbers.htm http://www.inpe.br/programas/cbers/english/satelite.html Calibration http://www.gisdevelopment.net/aars/acrs/1999/ts7/ts7201.shtml
CHRIS	Compact High Resolution Imaging Spectrometer	ESA	PROBA	10/2001	www.rsacl.co.uk/chris Calibration http://www.cossa.csiro.au/reports/prata/chris_meeting.html
CIS	Chinese Imaging Spectrometer	China	N/A	N/A	Shanghai Institute of Technical Physics, Shanghai

Table 1.2 continued.

COIS/PIC	Coastal Ocean Imaging Spectrometer / Panchromatic Imaging Camera	US	NEMO	2002	http://www.nrl.navy.mil/ http://www.tec.army.mil/tio/nemo.htm Calibration http://rsd-www.nrl.navy.mil/7212/pdf/20020225_OE.pdf
DAVID		ISA, DLR	DAVID	2002?	http://beta.most.gov.il/sela_dir/DAVID.html www.ohb-system.de
EROS-A1 EROS-B1	Earth Remote Sensing Observation Satellite	ImageSat Int.	EROS-A EROS-B	12/2000 2003?	http://www.imagesatintl.com/aboutus/satellites/satellites.shtml#
ETM+	Enhanced Thematic Mapper Plus	NASA	LANDSAT-7	1999	http://landsat.gsfc.nasa.gov/ Calibration: http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_htmls/chapter8/chapter8.html http://www.optics.arizona.edu/rsg/
FTHSI	Fourier Transform Hyperspectral Imager	US Air Force Phillips Lab	Mighty Sat II	05/1999	http://www.vs.afrl.af.mil/TechProgs/MightySatII/FTHSI.html http://www.vs.afrl.af.mil/factsheets/msat2.html
GLI	Global Imager	Japan	ADEOS-2	2002?	http://adeos2.hq.nasda.go.jp/shosai_gli_e.htm Calibration http://sharaku.eorc.nasda.go.jp/GLI/meet/2001/40.pdf
HIRS/3	High Resolution Infrared Radiation Sounder	NOAA	NOAA-K NOAA-L NOAA-M NOAA-N	05/1998 2000 2002 2003?	http://www.eumetsat.de/en/index.html?area=left4.html&body=/en/area4/aapp/hirs_3.html&a=420&b=1&c=400&d=400&e=0 http://margotte.univ-paris1.fr/cgms/en/ap10-09.htm Calibration http://www2.ncdc.noaa.gov/docs/klm/html/c3/sec3-2.htm
HRV HRVIR HRG	High Resolution Visible Sensor High Resolution Visible and Infrared sensor High Resolution Geometric Sensor	Spot Image, CNES	SPOT-2 SPOT-4 SPOT-5	01/1990 03/1998 05/2002	http://www.spotimage.fr/ Calibration http://www.optics.arizona.edu/rsg/ http://vegetation.cnes.fr/vgtprep/vgt2000/henry.pdf
HYPERION		NASA	EO-1	11/2000	http://eo1.gsfc.nasa.gov/Technology/Hyperion.html Calibration: http://www.trw.com/extlink/1,,,00.html?ExternalTRW=/images/hyperion_cal.pdf http://www.ece.arizona.edu/~djal/base_files/NewPage/eo1.html http://www.eoc.csiro.au/hswww/Hyperion.htm
IR-MSS	Infrared Multispectral Scanner	CAST / INPE	CBERS-1 CBERS-2	1999 2003?	http://www.dgi.inpe.br/html/eng/cbers.htm http://www.inpe.br/programas/cbers/english/satelite.html Calibration http://www.gisdevelopment.net/aars/acrs/1999/ps2/ps20034pf.htm
IKONOS		Space Imaging	IKONOS	09/1999	http://www.spaceimaging.com/products/ikonos/index.htm Calibration http://www.spaceimaging.com/aboutus/satellites/IKONOS/spectral.htm http://www.geosystems.de/atcor/sensors/IKONOS.html
KVR-1000 TK-350	2 m resolution 10 m resolution pan. Cameras	Russia	SPIN-2	02/1998	http://www.tec.army.mil/tio/SPIN2.htm

Table 1.2 continued.

LAC	Linear etalon imaging spectrometer array (LEISA) Atmospheric Corrector	NASA	EO-1	11/2000	http://eo1.gsfc.nasa.gov/overview/RTR/Sec-12-AC/ http://eo1.usgs.gov/instru/leisa.asp Calibration http://ldcm.usgs.gov/eo-1forum/Validation_Reports/LAC.pdf
LISS-I LISS-II LISS-III LISS-IV	Linear Self-Scanning Sensor	India	IRS-1B IRS-1C IRS-1D IRS-P6	08/1991 12/1995 09/1997 2002?	http://www.euromap.de/ Calibration http://www.euromap.de/
M5	Multispectral 5m resolution	DigitalGlobe™	N/A	2006?	http://www.digitalglobe.com
M10	Multispectral 10 m resolution	Resource21	Resource21	2004?	http://www.resource21.com/default.htm
MERIS	Medium Resolution Imaging Spectrometer	ESA	ENVISAT	03/2002	http://envisat.esa.int/instruments/meris/ Calibration http://envisat.esa.int/calval/cr/10092002/MERIS/
MISR	Multi-Angle Imaging Spectroradiometer	NASA	EOS AM-1 (TERRA)	12/1999	http://www-misr.jpl.nasa.gov Calibration http://www-misr.jpl.nasa.gov/mission/calib.html
MMRS	Mulitspectral Medium Resolution Scanner	CONAE Argentina	SAC-C	11/2000	http://www.gsfc.nasa.gov/gsfc/service/gallery/fact_sheets/spacesci/sac-c.htm
MODIS	Moderate- Resolution Imaging Spectroradiometer	NASA	EOS AM-1 (TERRA) EOS PM-1 (AQUA)	12/1999 05/2002	http://modis.gsfc.nasa.gov Calibration http://www.optics.arizona.edu/rsg/
MOS	Modular Optoelectronic Scanner	DLR	IRS-P3	03/1996	http://www.ba.dlr.de/NE-WS/ws5/mos_home.html Calibration http://www.ba.dlr.de/NE-WS/ws5/index_mos.html
MSU-SK / MSU-E	Medium-resolution / high-resolution multispectral scanner	Russia	RESURS-O1-3 O1-4	10/1994 07/1998	http://sputnik.infospace.ru/resurs/engl/resurs.htm http://ceos.cnes.fr:8100/cdrom-00b2/ceos1/satellit/scanex/resurs/resurs_o.htm http://www.scanex.ru/stations/resurs.htm#MSUSK
OCI	Ocean Colour Imager	Taiwan	ROCSAT-1	12/1998	http://www.nspo.gov.tw/e50/menu0504.html http://rocsat1.oci.ntou.edu.tw/en/oci/index.htm
OCM	Ocean Colour Monitor	India	IRS-P4 / Oceansat	1999	http://www.isro.org/irsp4.htm http://202.54.32.164/test/docu/irsp4/c231.html
OCTS	Ocean Color and Temperature Scanner			08/1996-06/1997	http://www.eorc.nasda.go.jp/ADEOS/Project/Octs.html
OrbView		Orbimage	Orbview-3	2003?	http://www.orbimage.com/corp/orbimage_system/ov3/ Calibration http://www.optics.arizona.edu/rsg/
OSMI	Ocean Scanning Multi-spectral Imager	Korea	KOMPSAT-1	12/1999	http://kompsat.kari.re.kr/english/index.asp (Site Map > OSMI)
PAN		India	IRS-1C IRS-1D IRS-P5	1995 09/1997 2003?	http://www.euromap.de/doc_000.htm Calibration http://www.ipi.uni-hannover.de/html/publikationen/1998/jacobsen/jac_98_calib_irs1c.pdf

Table 1.2 continued.

POLDER-2	Polarization and Directionality of the Earth's Reflectance	CNES NASDA	ADEOS-2	2002?	http://adeos2.hq.nasda.go.jp/shosai_polder_e.htm http://ceos.cnes.fr:8100/cdrom-00b2/ceos1/satellit/polder/index.htm Calibration http://smsc.cnes.fr/POLDER/A_calibration.htm
PRISM	Panchromatic Remote-sensing Instrument for Stereo Mapping	NASDA (Japan)	ALOS	2004	http://www.eorc.nasda.go.jp/ALOS/about/prism.htm Calibration http://www.eorc.nasda.go.jp/ALOS/da/2example.htm
Quickbird		DigitalGlobe™	Quickbird	10/2001	http://www.digitalglobe.com/products/quickbird.shtml
SeaWiFS	Sea-viewing Wide Field-of-view Sensor	Orbimage	OrbView-2 / SeaStar	08/1997	http://seawifs.gsfc.nasa.gov/SEAWIFS.html Calibration http://www.optics.arizona.edu/rsg/ http://www.seaspace.com/service/support/TeraScan_Docs/doc/man1/seawifscal.html
Sciamachy	SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY	ESA	ENVISAT	03/2002	http://envisat.esa.int/instruments/sciamachy/ Calibration http://envisat.esa.int/calval/cr/11092002/SCIAMACHY/
TM	Thematic Mapper	NASA	LANDSAT-5	01/1984	http://geo.arc.nasa.gov/sge/landsat/landsat.html Calibration http://www.nal.usda.gov/ttic/tektran/data/000007/14/0000071413.html http://www.ccrs.nrcan.gc.ca/ccrs/rd/ana/calval/effort_e.html http://www.bsrsi.msu.edu/trfic/data_portal/Landsat7doc/landsatch8.html
UVISI (SPIM)	Ultraviolet and Visible Imagers and Spectrographic Imager	US	MSX	04/1996	http://spider.ipac.caltech.edu/staff/mmm/msx-related/td1702/hefernan.pdf Calibration http://simbios.gsfc.nasa.gov/Info/sensor_table.html
VEGETATION-1 VEGETATION-2		Spotimage, CNES	SPOT-4 SPOT-5	03/1998 05/2002	http://www.spotimage.fr/home/system/introsat/payload/vegetati/welcome.htm Calibration http://vegetation.cnes.fr/vgtprep/vgt2000/henry.pdf
VIRS	Visible Infrared Scanner	NASA	TRMM	11/1997	http://trmm.gsfc.nasa.gov/overview_dir/virs.html Calibration http://trmm.gsfc.nasa.gov/1b01.html
VIRSR	Visible and Infra-red Scanning Radiometer	NOAA	NOAA-O	2005?	http://www.ccrs.nrcan.gc.ca/ccrs/data/satsens/sats/noaa_e.html#noaao
WAOSS	Wide Angle Optoelectronic Stereo Scanner	DLR	BIRD	1999	http://www.ba.dlr.de/ne/ws/projects/waoss/waoss.html Calibration http://www.ba.dlr.de/NE-WS/ws3/cg/ws-dt-cg.html
WiFS AWiFS	Wide Field Sensor	India	IRS-1C IRS-1D IRS-2A IRS-P6	1995 09/1997 2000 2002?	http://www.euromap.de/ Calibration http://www.euromap.de/
WFI	Wide Field Imager	CAST / INPE	CBERS-1 CBERS-2	1999 2003?	http://msowww.anu.edu.au/observing/wfi/intro.shtml http://www.inpe.br/programas/cbers/english/satelite.html Calibration http://www.cresda.com.cn/en/products_01.htm

Table 1.3: Specifications of the airborne imaging spectrometers CASI-2 and HyMAP(*).

	CASI-2	HyMAP
Full name	Compact Airborne Spectrographic Imager	Hyperspectral Mapper
Manufacturer	ITRES Research Ltd., Canada	Integrated Spectronics Pty. Ltd., Australia
Operator	Natural Environment Research Council Airborne Remote Sensing Facility (NERC ARSF), UK	HyVISTA Corp. Pty. Ltd., Australia
Type	Pushbroom	Whiskbroom
Spectral range and coverage	545 nm between 400 and 1000 nm, continuous	450 - 2500 nm, continuous except gaps at 1400 & 1900 nm
Number of bands	288 (maximum)	126
Bandwidth	2.2 - 7.8 nm	11 - 22 nm
Programmable wavebands	Yes	No
Number of spatial pixels	512 (maximum)	512
IFOV(**) across-track	1.8 mrad	2.0 mrad
FOV	54.4 ° (customised lens)	61.3 °
GIFOV(***)	1 - 7 m	3 - 10 m
Signal-to-noise ratio	420:1 peak in laboratory with integrating sphere	500:1 peak outside, 30° Sun angle, 50% reflective standard
On-board calibration	No	Yes (spectral and radiometric)

(*) Only the details of the solar reflective HyMAP sensor are shown in this table.
(**) Instantaneous FOV
(***) Ground-projected instantaneous FOV
(Source: http://www.itres.com/, http://www.hyvista.com/, verified last on 22/09/2003)

Table 1.4: Characteristics of the spaceborne imaging spectrometers ASTER, CHRIS, Hyperion, MERIS, MODIS and MOS for the VNIR and SWIR wavelength regions.

	ASTER(*)	CHRIS	Hyperion	MERIS	MODIS(*)	MOS
Sponsor	NASA	ESA	NASA	ESA	NASA	DLR
Platform	Terra	PROBA	EO-1	Envisat	Aqua	IRS-P3
Launch	Dec 1999	Oct 2001	Nov 2000	Mar 2002	May 2002	Mar 1996
Type	pushbroom	pushbroom	pushbroom	pushbroom	whiskbroom	pushbroom
Spectral range	0.5 - 2.5 µm	0.4 - 1.0 µm	0.4 - 2.5 µm	0.4 - 1.0 µm	0.4 - 2.2 µm	0.4 - 1.0 µm
No. of bands	3 (VNIR) 6 (SWIR)	19 or 62	220	15	20	18
Spectral resolution	40 - 100 nm	1.3 - 11 nm	10 nm	3.7 - 20 nm	10-50 nm	1.4, 10, 100 nm
Spectral coverage	discrete	discrete	continuous	discrete	discrete	discrete
GIFOV	15 m, VNIR 30 m, SWIR	25 or 50 m	30 m	300 m or 1.2 km	250, 500 m or 1 km	500 m
Swath width	60 km	19 km	7.5 km	1150 km	2330 km	200 km

(*) Only the details of the solar reflective ASTER and MODIS sensors (VNIR, SWIR) are shown in this table.
(Source: http://asterweb.jpl.nasa.gov/, http://www.rsacl.co.uk/chris/, http://modis.gsfc.nasa.gov, http://eo1.gsfc.nasa.gov/Technology/Hyperion.html, http://envisat.esa.int/instruments/meris, http://www.ba.dlr.de/NE-WS/ws5/mos_home.html, verified last on 22/09/2003)

In the remote sensing community, some spaceborne sensors are referred to as imaging spectrometers simply because of their narrow bands, ignoring the fact that they do not sample the spectrum continuously (see table 1.4). Until now, Hyperion seems to represent the only true spaceborne imaging spectrometer recording data contiguously in the VNIR and SWIR wavelength regions with very fine bands.

The Benefits and Costs of Hyperspectral Data

The benefits of hyperspectral data are widely acknowledged: relevant narrow spectral absorption and emission features may be detected within the image spectra, allowing an improved identification and discrimination of ground targets, and characterisation of their related physical or chemical properties. Many rocks and minerals, and some plant species have been remotely identified and mapped on the basis of imaging spectrometer data (Vane and Goetz, 1988; Kruse, 1988).

Though very beneficial, hyperspectral capability does not come without cost:

1) **Data handling cost**: Both increased storage space and transmission time are necessary to cope with greater data volumes. Conventional data compression techniques may be applied, but not without loss (Chen and Landgrebe, 1989).
2) **SNR reduction**: The imaging spectrometer distributes the incoming radiation from a given area per unit time to a large number of spectral bands achieving a lower band SNR than broad band sensors, even with the aperture wide open. Then, the spatial resolution has to be degraded or the sensor residence time increased to reach an acceptable signal level (Price, 1994a), which is often not possible due to uncompromising data requirements or dwell time limitations.
3) **Data redundancy**: Imaging spectrometer data are highly correlated between adjacent bands, due to the closely located and overlapping spectral sampling intervals, and the typically gradually changing nature of the reflectance of most terrestrial materials with wavelength (Curran *et al.*, 1998). That is, high-dimensional feature spaces will be mainly empty with the significant information-bearing structure existing in a lower dimensional space (Landgrebe, 2000).
4) **Data processing cost**: Some data processing methods encounter difficulties when confronted with the high dimensionality of the data set. Standard matrix-based methods including matrix inversions, such as Principal Component Analysis (PCA), may be computationally unstable since being subject to round-off and truncation errors in the higher order terms (Price, 1990, 1994a). For other methods, such as Artificial Neural Networks (ANN), high dimensional inputs require a large number of parameters to train, resulting in long training times and low generalisation capabilities (Benediktsson and Sveinsson, 1997; Kavzoglu and Mather, 2000).
5) **Sampling cost**: Given a limited number of training samples, the overall application performance initially improves as new features are added, but at a certain point inclusion of further features will result in performance degradation. This phenomenon referred to as 'curse of dimensionality',

'peaking phenomenon' or 'Hughes phenomenon', was discovered by Hughes (1968) (Kittler, 1986). That is, the number of training samples needed to characterise classes accurately increases with dimensionality, and since ground samples are often limited, this may cause a problem.

6) **Data visualisation cost**: The feature space cannot be visualised in its entirety as for two- or three-dimensional data.

The last four costs from the above list can be eliminated via data reduction methods at processing level after data acquisition. For example, feature selection or extraction algorithms may be employed to reduce the data to the most information bearing features. Features are defined here as input dimensions into a processing algorithm and may represent image bands or transforms of those bands.

However, a solution for the first two problems can only exist 'at sensor level', i.e. a smaller amount of data has to be acquired. Some airborne and spaceborne sensors are designed to record only a reduced number of discrete wavebands specific to certain research field applications, such as land cover classification, atmospheric water vapour retrieval, etc. For a list of typical vegetation and coastal band sets of some imaging spectrometers, the reader is referred to table C.3 in the appendix. A band is defined in this context as a wavelength interval in the electromagnetic (EM) spectrum that may correspond to the bandpass of a channel or channel assemblages. A channel is a physical CCD detector element that records signals of specific wavelengths of the EM spectrum.

However, these pre-determined band sets may not be optimal for a specific application task chosen by the data user. For instance, a user may want to classify different vegetation species, but the sensor's default 'vegetation' band set may only offer bands which discriminate vegetation classes from non-vegetation classes, and not between different vegetation species.

Application-specific band selection has the potential of practically overcoming all of the problems associated with hyperspectral data. However, until now, only few researchers have exploited this idea, many of them guided by traditional band sets from earlier sensors.

1.2 Research Objectives

The main objectives of this research project were

- to develop and evaluate supervised and unsupervised band selection approaches for imaging sensors with respect to the Maximum Likelihood classification (MLC) as data application, and

- to generate computationally efficient ENVI™ programs to be used mainly with hyperspectral data from airborne and satellite sensors.

Subsidiary research objectives were

- to review dimensionality reduction methods for hyperspectral data, such as feature selection and feature extraction,
- to apply some of the reviewed dimensionality reduction methods to real hyperspectral image data and evaluate their effectiveness with respect to classification accuracy,
- to estimate the intrinsic dimensionality of hyperspectral data sets,
- to compare the classification accuracy of the band sets defined by the band selection methods developed in this thesis with the one generated using the default band configurations of common airborne and satellite sensors,
- to test the effectiveness of band selection and the benefit of hyperspectral data for the given classification tasks,
- to generalise the band selection methods to other applications such as other hard classification methods, regression, linear spectral unmixing or spectral angle mapper, and
- to discuss data simulation in cases where hyperspectral input data could not be acquired with the target sensor (e.g. for sensor design studies).

1.3 Thesis Outline

The outline of the thesis is illustrated in a data flow diagram in figure 1.2. Initially, in chapter 1, the reader is introduced to the concepts of remote sensing and imaging spectrometry, and the rationale and outline of the thesis are exposed. Chapter 2 provides a review of dimensionality reduction methods, some of which are applied to real data sets in chapter 3 with Maximum Likelihood classification (MLC) as application. The study areas and data resources are fully described in the latter chapter. A supervised band selection algorithm for MLC is described in chapter 4, and unsupervised strategies are presented in chapter 5. Both chapters 4 and 5 include a distinct analysis of the proposed methodology, its evaluation with respect to real hyperspectral data sets, and the comparison of derived band sets to commonly used band sets of airborne and satellite sensors. Some issues of discussion are raised in chapter 6, including the comparison between methods, their generalisation to other applications, and the investigation whether or not hyperspectral data are of additional benefit for the given application task. Finally, based on the research findings, a concluding summary is given in chapter 7.

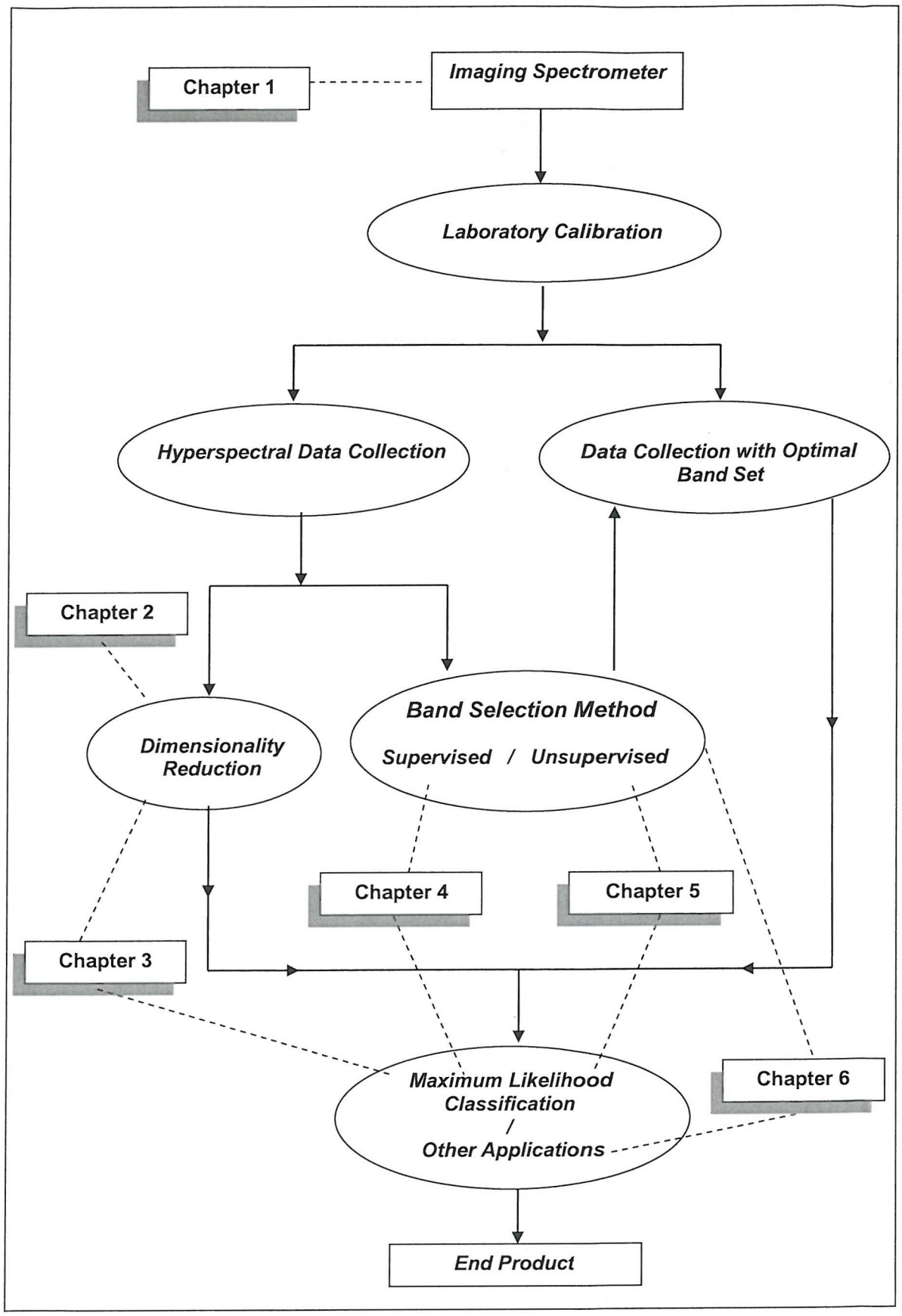


Figure 1.2: Data flow diagram illustrating the course of action (solid line) and the link of each data transformation to a chapter of the thesis (dashed line).

2 Dimensionality Reduction

2.1 Introduction

As explained in chapter 1, a reduction of hyperspectral data measurements at processing level is beneficial for the following reasons:

1) to reduce the overall cost of storage and processing (easier subsequent analysis and less computation time),
2) to improve application performance by discarding irrelevant, redundant, non-discriminatory and noisy bands,
3) to increase the stability of data processing algorithms and decrease their complexity (number of parameters) to avoid poor generalisation performance (Webb, 1999),
4) to gain a stable representation of classes for a limited number of training samples, and
5) to explore the underlying structure by obtaining a graphical representation (Webb, 1999).

Mathematical tools for reducing the data dimensionality are called ordination methods or geometrical methods in multivariate analysis (Webb, 1999), while in pattern recognition they are referred to as feature selection and feature extraction. In this thesis, the pattern recognition terminology is adopted. The word feature in this context refers to both sensor bands and any transforms of them. Features are arranged in vector form for each pixel. The number of vector elements determines the dimensionality of the feature space.

Feature selection is the process of selecting a subset of measurements out of the original set of measurements. A feature selection algorithm needs a criterion according to which it will select features, and a technique to search for the optimal feature subset in a systematic way. Section 2.2 introduces common unsupervised and supervised criterion functions (section 2.2.2) and the main optimal and suboptimal search algorithms (section 2.2.3) used for feature selection.

In contrast to feature selection, feature extraction methods transform the original high-dimensional measurement space into a new lower-dimensional feature space by optimising a certain criterion. The underlying type of transformation has to be determined, for example a linear or non-linear transformation. In this thesis, only linear transformations were considered, since they have the advantage of being simple, well defined, analytically tractable and computationally feasible (Kittler, 1986). Section 2.3 presents some unsupervised and supervised feature extraction techniques found in the literature. Non-linear frequency-based transformations, such as the discrete Fourier or wavelet transform, are not discussed in this thesis.

The methods presented in this chapter are aimed to select or extract features at the processing level after data acquisition. For a band selection review, the reader is referred to chapter 5.

Dimensionality reduction methods assume redundancy to be present in the data and that the main features can be described in terms of a tendency for the point cloud to concentrate into clusters (Jones and Sibson, 1987). The number of these features is commonly referred to as intrinsic dimensionality for representation. In contrast, the intrinsic dimensionality for classification (or intrinsic discriminant dimensionality) is the number of significantly reduced dimensions that still result in satisfactory classification accuracy. Section 2.4 presents both heuristic and statistical methods, which use the outcome of mainly feature extraction methods to determine the intrinsic dimensionality for either representation or classification.

Concept and notation

For the discussion of this chapter, the following concept and notation will be used. A statistical basis is assumed for both feature selection and feature extraction methods to explain the variability of the feature representation. A remotely sensed measurement set is understood to be generated by a state of nature, with one individual measurement representing one realisation of a continuous random variable defined over elements of the sensor sample space (Devijver and Kittler, 1982). The underlying model of the measurement process is a probability density function, which associates with each sample value the probability that this value will be assumed, and is usually approximated by a Gaussian function. This statistical framework will describe the extraneous (e.g. instrument noise) and intrinsic (e.g. pattern noise) variability of measurements, and may be applied to the entire data set or to specific measurements of user-defined classes.

The original pixel measurement for N bands is denoted by x_j, $j = 1,2,...,N$, or in vector form by $\mathbf{x}$, that is, $\mathbf{x} = [x_1 \quad ... \quad x_N]^T$. Each pixel vector $\mathbf{x}$ is supposed to belong to one of M possible classes ω_k, $k = 1,2,...,M$. The pixel measurement is assumed to be generated by a random process with a model characterised by class-conditional density functions $\mathrm{p}(\mathbf{x} \mid \omega_k)$ and *a priori* class probabilities $\mathrm{P}(\omega_k)$.

To analyse and summarise the data, descriptive measures will be used. As a measure of central tendency, indicating the centre of a density function, the (arithmetic) sample mean m_j is used (equation 2.1). It may be combined with the means of all other bands in the sample mean vector $\mathbf{m} = [m_1 \quad ... \quad m_N]^T$, with

$$m_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{2.1}$$

where m_j Sample mean for band j

x_{ij} Pixel measurement i for band j

n Number of pixel measurements

For band j, the dispersion or scatter of the density function around its centre m_j is estimated by the sample variance s_{jj}^2, defined by equation 2.2. The sample standard deviation s_{jj} is the square root of the sample variance.

$$s_{jj}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - m_j\right)^2 \tag{2.2}$$

where s_{jj}^2 Sample variance for band j

Often, the sample variances s_{jj}^2 of all bands are combined with the sample covariances s_{jl} between the bands (equation 2.3) to form the symmetric $N \times N$ sample covariance matrix $\mathbf{S}$. To avoid the singularity of $\mathbf{S}$, the number of samples n needs to be larger than the dimension N (Fukunaga, 1982). Kalayeh and Landgrebe (1983) suggest n being five times N to achieve a good estimate of $\mathbf{S}$.

$$s_{jl} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - m_j\right)\left(x_{il} - m_l\right) \tag{2.3}$$

where s_{jl} Sample covariance for bands j and l

The sample correlation coefficient r_{jl} ($r_{jl} \in [-1,1]$) is defined for bands j and l by equation 2.4 and may be used to form a symmetric $N \times N$ sample correlation matrix $\mathbf{R}$ for all bands. The square of the correlation coefficient r_{jl}^2 is termed the coefficient of determination ($r_{jl}^2 \in [0,1]$) and represents the proportion of the total variance of pixel values in band l which is due to the linear relationship between values of bands j and l.

$$r_{jl} = \frac{s_{jl}}{s_{jj}s_{ll}} \tag{2.4}$$

where r_{jl} Sample correlation coefficient of pixel values for bands j and l

To compare the variation in several sets of data, the sample variance may be an inappropriate measure, since it quantifies absolute variation and is therefore dependent on the measurement scale. The coefficient of variation is defined as a measure of relative variation (in %):

$$cv_j = \frac{s_{jj}}{m_j} \cdot 100 \tag{2.5}$$

where cv_j Coefficient of variation for band j (in %)

Further measures may be calculated to explain the shape of the probability density function. The skewness describes the deviation of the class density from symmetry (see equation 2.6), while the kurtosis quantifies its peakedness (equation 2.7). Both measures have a zero value for a normal probability density function (Press *et al.*, 1992).

$$skewness_j = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{ij} - m_j}{s_{jj}}\right)^3 \tag{2.6}$$

$$kurtosis_j = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{ij} - m_j}{s_{jj}}\right)^4\right] - 3 \tag{2.7}$$

Where $skewness_j$ Skewness for band j

$kurtosis_j$ Kurtosis for band j

2.2 Feature Selection

2.2.1 Introduction

Feature selection techniques choose a subset of bands of given size out of all image bands that contains the highest possible amount of information. They are based on the assumption that not all bands carry the same amount of information, which is particularly justified for hyperspectral data sets where adjacent bands are typically highly correlated (Tu *et al.*, 1998). Information is considered as any extracted data useful to the analyst.

A feature selection algorithm needs a criterion, which quantifies the amount of information of each band subset under investigation. One may distinguish between unsupervised (class-independent) and supervised (class-dependent) criterion functions, depending on whether or not class information is used within the criteria. Section 2.2.2 reviews common criterion functions used in feature selection. In addition, band subsets need to be searched for in a systematic manner via search algorithms. Section 2.2.3 presents common optimal and sub-optimal search algorithms.

Siedlecki and Sklansky (1988) perceive feature selection as 'an extremely difficult task', since it is charged with both theoretical and computational problems. The user has to trade-off between the

optimality of the resulting feature subset and the computational efficiency of the feature selection algorithm. This review focuses on feature selection techniques based on statistical pattern recognition techniques alone, excluding those techniques related to artificial neural networks (such as node pruning).

2.2.2 Criterion Functions

This section presents common supervised and unsupervised criterion functions that make use of class statistics or not, respectively.

Unsupervised criterion functions

Unsupervised criterion functions do not require *a priori* knowledge of class spectra. They may be used in data exploration techniques for classification.

1) Variance and standard deviation

Variance and standard deviation have been widely employed to measure information content in the context of feature selection for image visualisation. Examples are Chavez *et al.* (1982), Sheffield (1985), and Beauchemin and Fung (2001).

2) Coefficient of variation

The coefficient of variation (equation 2.5) is the sample standard deviation divided by the sample mean of the same data set. It is independent of the units of measurements and often used as a measure of noise (i.e. random noise in this context). Shaban and Dikshit (2001) used the multi-class average coefficient of variation as a criterion to select textural and spectral features for the classification of urban area. The measure has been found to decrease with increasing classification accuracy.

3) Priority number (PN) criterion

Lin and Chang (2001) introduced a criterion for band selection that is based on the relative mean ratio (band mean divided by the smallest mean of all bands) and the relative variance ratio (analogue to relative mean ratio):

$$PN_j = \frac{N^2}{\text{rank}(RMR_j) \cdot \text{rank}(RVR_j)} \tag{2.8}$$

$$RMR_j = \frac{m_j}{\min_j(m_j)} \tag{2.9}$$

$$RVR_j = \frac{s_j^2}{\min_j(s_j^2)} \tag{2.10}$$

Where

PN_j	Priority Number (PN) for band j
$\text{rank}(\cdot)$	Number assigned by priority in decreasing order
RMR_j	Relative mean ratio for band j
RVR_j	Relative variance ratio for band j
$\min_j(\cdot)$	Minimum over all bands j

The higher the value of the PN criterion for band j, the more important the band is considered. As equivalent measure to be maximised, Lin and Chang (2001) propose the simple product of RMR_j and RVR_j.

4) Band correlation coefficient

The correlation coefficient (equation 2.4) measures the strength of the linear association between two variables, here two sensor bands. It is not affected by changes in the unit of measurement of either or both variables. For the textural and spectral classification of urban areas, Shaban and Dikshit (2001) selected textural features which had a low correlation coefficient with the spectral bands and between themselves. Chavez *et al.* (1982), Sheffield (1985) and Beauchemin and Fung (2001) employed the correlation matrix in their feature selection criteria for image visualisation to achieve a low pairwise correlation between the selected features. Ebert (1987) uses the correlation coefficient in combination with the divergence measure to select features for cloud detection and classification.

5) Image signal-to-noise ratio (SNR)

The signal-to-noise ratio (SNR) may be defined as the ratio between signal and noise. The signal carries the information of interest to the user, while noise represents unwanted variations added or multiplied to the signal. As the definitions of signal and noise change with application, no single definition of SNR exists (Schowengerdt, 1997). For imaging spectrometry, the image signal may be described by the average measured at-sensor radiance or reflectance. Image noise is introduced by the atmosphere, the sensor and the platform (Smith and Curran, 1999). While its systematic components can be measured and removed, its random contributions may vary from image to image and may be estimated from the methods outlined in section 2.3.2. Lin and Chang (2001) chose the image SNR as a feature selection criterion to be maximised.

6) Spatial autocorrelation of ratioed bands

Warner *et al.* (1999) aimed to exclude noisy bands from the maximum likelihood classification of different plant communities. They assumed an H-resolution image scene (Strahler *et al.*, 1986) for the classes of interest and based their technique on the idea that under this assumption class information is spatially autocorrelated, while noise is not.

Spatial autocorrelation is quantified by local variance relative to the overall variance in the scene. Both local and global variations are similar for random noise (small spatial autocorrelation), while the local variation is much smaller than global variance for large homogeneous objects in the image scene (high spatial autocorrelation) (Warner and Shank, 1997).

Warner *et al.* (1999) chose the spatial autocorrelation of the ratio of a pair of bands as unsupervised feature selection criterion, which is maximised when the bands are not noisy. A band ratio is evaluated rather than a single band in order to suppress radiance changes due to variations in illumination across the scene. The pair of bands resulting in the highest spatial autocorrelation includes the optimal two bands. The next best feature then gives the highest spatial autocorrelation when ratioed with each one of the previously chosen features.

The authors concluded that the incorporation of spatial analysis in the feature selection process might reduce noisy bands being chosen that might otherwise decrease classification accuracy significantly. However, in cases where within-class variances are larger than between-class variances the criterion is not applicable.

Supervised criterion functions

All criteria presented in this section require information about the classes under investigation, which is usually gained from training and/or test pixels of the image to be classified. In classification, the analyst may be interested in an accurate or reliable classification result. The main objective of feature selection for classification is then to select a band subset of given size that maximises the classification accuracy. The classification accuracy may be estimated with measures derived from the classification error matrix, which represent the criterion functions to be optimised.

A surrogate method is to measure the overlap between class distributions for different features, since highly separated classes in feature space result in a small misclassification error. Penaloza and Welch (1996) quantified the class overlap by calculating the histograms for each feature and class pair. However, this procedure needs to store all samples in all dimensions and is very expensive to compute for high-dimensional data sets.

A computationally less costly way to quantify class overlap is to define a distance measure (also referred to as separability or dissimilarity measure) that estimates the separability between the class probability distributions under investigation. One may distinguish between criteria derived from heuristic reasoning (heuristic distance measures), and those based on information theory and statistics (probabilistic distance measures) (Fu, 1982).

In the following paragraphs, classification accuracy estimates are investigated, as well as heuristic and probabilistic distance measures for both the two-class and multi-class case. Probabilistic dependence and entropy measures are not considered here because of the difficulties in estimating their relatively complex expressions.

1) Classification accuracy estimates

The classification accuracy may be estimated by statistics derived from the error matrix (or confusion matrix or contingency table). Many accuracy measures exist (Foody, 2002), including the overall accuracy (or percentage agreement), the normalised accuracy, the TAU coefficient, and the KHAT estimate (KAPPA analysis).

The error matrix is made up of $M \times M$ elements, where each element n_{ij} refers the number of test samples that have been classified as class i although they belong to class j. The main diagonal elements n_{ii} represent the pixels that have been correctly classified. The error matrix is usually constructed on the basis of a test set that is independent of the training set used to form the classifier ('Holdout estimate', Webb, 1999). In case of low availability of training samples, a cross-validation method ('leave-one-out') can be employed, which is approximately unbiased (Webb, 1999). The latter method uses all samples except one as training samples and tests the classifier on the remaining test sample. This procedure is repeated for all samples being individually used as test samples. The 're-substitution' method employs all samples for both training and testing sets and is generally considered as biased. It will therefore be neglected in this study.

The overall accuracy is the sum of the total number of correctly classified pixels (sum of main diagonal matrix elements) divided by the total number of pixels in the error matrix:

$$P_O = \frac{1}{N_E} \sum_{i=1}^{M} n_{ii} \tag{2.11}$$

Where
P_O Overall accuracy
N_E Total number of pixels in the error matrix
n_{ii} Number of correctly classified pixels for class i

The normalised accuracy is the analogue criterion to P_O of a normalised or standardised error matrix obtained by forcing each row and column to sum to one via an iterative proportional fitting procedure (Bishop *et al.*, 1975 in Congalton, 1991). Standardisation makes error matrices directly comparable to other matrices since their differences in sample sizes are removed. Since the normalised accuracy indirectly contains information about off-diagonal matrix elements, it is said to be a better representation of the classification accuracy than the overall accuracy (Congalton, 1991).

However, P_O and the normalised accuracy do not consider the proportion of agreement between the training and test data set that is due to chance alone and therefore tend to overestimate classification accuracy (Ma and Redmond, 1995). The KAPPA analysis, defined by Cohen (1960), aims to account for the proportion of agreement between the training and test data set that is due to chance alone. It may be performed with the KHAT statistic, a maximum likelihood estimate of KAPPA. KHAT directly incorporates off-diagonal error matrix elements in form of the marginal row and column totals:

$$\hat{\kappa} = \frac{P_O - P_C}{1 - P_C} \tag{2.12}$$

$$P_C = \frac{1}{N_E^2} \sum_{i=1}^{M} n_{i+} n_{+i} \tag{2.13}$$

Where $\hat{\kappa}$ KHAT statistics
P_C Chance agreement
n_{i+} Marginal total of row i of the error matrix
n_{+i} Marginal total of column i of the error matrix

Nevertheless, KHAT assumes a multinomial sampling model, i.e. it requires pixel sampling with replacement, which is not applied in practice (Thomas and Allcock, 1984). This can, however, be ignored for a large number of samples. Foody (1992) showed that $\hat{\kappa}$ overestimates the chance agreement P_C without modifications. The TAU coefficient includes *a priori* probabilities apart from the marginal distributions of the reference data (Klecka, 1980):

$$\tau = \frac{P_O - P_R}{1 - P_R} \tag{2.14}$$

$$P_R = \frac{1}{N_E^2} \sum_{i=1}^{M} n_{i+} n_{xi} \tag{2.15}$$

Where τ Tau statistics
P_R Prior probability
n_{xi} A priori number of pixels belonging to class i out of N pixels; equals N/M for equal a priori class probabilities $1/M$

Ma and Redmond (1995) introduced a variation of the Tau statistics, τ_e, which measures the improvement of a classification over a random class assignment of pixels. The measure is calculated by replacing P_R with $1/M$ in equation 2.14, and gives, according to Ma and Redmond (1995), an accurate measure of classification accuracy.

2) Two-class heuristic distance measures

Heuristic distance measures have been defined on the basis of the experience that with decreasing distance between class distributions, the ability to separate between them lessens. Some popular distance measures are presented in table 2.1. Note that all pairwise criterion functions are not calculated for individual bands, but over all D candidate bands using the class mean vectors $\mathbf{m}$.

Table 2.1: Common heuristic distance measures to quantify the average pairwise separability between classes ω_k and ω_o using their mean vectors $\mathbf{m}$.

Distance measure	Mathematical form
Euclidean distance	$\mathrm{d_e}(k,o) = \sqrt{\sum_{j=1}^{D}\left(m_{j,k} - m_{j,o}\right)^2}$
City-block distance (Manhattan, box-car, absolute value)	$\mathrm{d_{cb}}(k,o) = \sum_{j=1}^{D}\left\lvert m_{j,k} - m_{j,o}\right\rvert$
Chebyshev distance (maximum value)	$\mathrm{d_{ch}}(k,o) = \max_j\left\lvert m_{j,k} - m_{j,o}\right\rvert$
Minkowski distance of order t	$\mathrm{d_M}(k,o) = \sqrt[t]{\sum_{j=1}^{D}\left(m_{j,k} - m_{j,o}\right)^t}$
Quadratic distance	$\mathrm{d_q}(k,o) = \sum_{l=1}^{D}\sum_{j=1}^{D}\left(m_{l,k} - m_{l,o}\right)Q_{lj}\left(m_{j,k} - m_{j,o}\right)$ where $\mathbf{Q}$ is a positive definite matrix
Canberra distance	$\mathrm{d_{ca}}(k,o) = \sum_{j=1}^{D}\frac{\left\lvert m_{j,k} - m_{j,o}\right\rvert}{m_{j,k} + m_{j,o}}$
Nonlinear distance	$\mathrm{d_n}(k,o) = \begin{cases} constant & \mathrm{d_e}(k,o) > threshold \\ 0 & \mathrm{d_e}(k,o) \le threshold \end{cases}$
Angular separation	$\mathrm{d_{as}}(k,o) = \cos(\alpha) = \frac{\sum_{j=1}^{D} m_{j,k} m_{j,o}}{\sqrt{\sum_{j=1}^{D} m_{j,k}^2 \sum_{j=1}^{D} m_{j,o}^2}}$

(Source: Webb, 1999)

The city-block metric is cheaper to compute than the Euclidean distance, while the Chebyshev distance is the cheapest to calculate out of the three measures. The Minkowski distance is a more general form of the city-block distance ($t = 1$) and the Euclidean distance ($t = 2$). The higher t is chosen, the more emphasis is placed onto larger distances, leading towards the Chebyshev distance for t tending towards infinity (Webb, 1999). The quadratic distance is similar to the Mahalanobis distance between two distributions (see below), and allows the simplification of the multiclass distance criterion (see below). The Canberra distance may be applied to variables with non-negative values. If both class means are zero, its ratio is taken zero. The angular separation measures the angle between the class mean vectors in feature space and may be applied when only the relative magnitudes are important (Webb, 1999). For example, Price (1994b) used the Euclidean distance measure and the angular separation measure to quantify within-class variability. Kittler (1975, in Kittler, 1986) gives guidelines for the selection of the constant and threshold of the nonlinear distance, which is likely to reflect class separability most reliably.

The heuristic measures presented so far exploit only the differences in class mean vectors. Kamp *et al.* (1997) used additionally the differences of the standard deviations of two class spectra to define spectral intervals where the discrimination between the given classes was possible.

The advantage of the above heuristic distance measures is that they do not involve the estimation of probability density functions and are therefore attractive for computational reasons and when the statistical model of the problem is not known. However, heuristic measures usually have no explicit relation to the classification accuracy: the extrapolation 'the larger the distance, the higher the classification accuracy' cannot be made. For example, let two class distributions be already perfectly (100%) separated and classified at small inter-class distance. While the distance between the distributions may be artificially increased, their separability and classification accuracy may not.

3) Two-class probabilistic distance measures

Probabilistic feature selection criteria are based on mathematical statistics and information theory, and require the estimation of the probability density functions of the classes under investigation. As stated earlier, the optimum criterion is the classification accuracy, i.e. the probability of error of the classifier. A direct calculation of the probability of error is often impractical partially due to the lack of general analytical expressions which are simple enough to be treated (Fukunaga, 1972).

Fukunaga (1972) describes the desirable characteristics of an alternative probability measure, which can be readily computed: it should

1) have a monotonic relationship with the probability of error, or with its lower or upper bounds,
2) be positive for unequal classes and zero for the same classes,
3) be monotonic, and
4) be additive for independent features.

A number of probability distance measures have been proposed in the pattern recognition literature which are indirectly related to the probability of error. These relatively complex measures are derived analytically from upper bounds to the error probability (Fu, 1982). They can be simplified under the assumption that the classes are normally distributed, and they are maximum when the classes are disjoint, i.e. the probability of error is small.

Table 2.2. Common pairwise probabilistic distance measures for normally distributed classes ω_k and ω_o using their mean vectors $\mathbf{m}$ and covariance matrices $\mathbf{S}$.

Distance measure	Parametric Form
Chernoff	$\mathrm{d_c}(k,o) = 0.5s(1-s)(\mathbf{m}_o - \mathbf{m}_k)^T[(1-s)\mathbf{S}_k + s\mathbf{S}_o]^{-1}(\mathbf{m}_o - \mathbf{m}_k) + 0.5\ln\left(\frac{\lvert(1-s)\mathbf{S}_k + s\mathbf{S}_o\rvert}{\lvert\mathbf{S}_k\rvert^{1-s}\lvert\mathbf{S}_o\rvert^{s}}\right)$, where $s \in [0,1]$
Bhattacharyya	$\mathrm{d_b}(k,o) = 0.25(\mathbf{m}_o - \mathbf{m}_k)^T[\mathbf{S}_k + \mathbf{S}_o]^{-1}(\mathbf{m}_o - \mathbf{m}_k) + 0.5\ln\left(\frac{\lvert\mathbf{S}_k + \mathbf{S}_o\rvert}{2(\lvert\mathbf{S}_k\rvert\lvert\mathbf{S}_o\rvert)^{0.5}}\right)$
Jeffries-Matusita	$\mathrm{d_{jm}}(k,o) = \sqrt{2\left[1 - \mathrm{e}^{-\mathrm{d_b}(k,o)}\right]}$
Divergence	$\mathrm{d_d}(k,o) = 0.5(\mathbf{m}_o - \mathbf{m}_k)^T\left[\mathbf{S}_k^{-1} + \mathbf{S}_o^{-1}\right](\mathbf{m}_o - \mathbf{m}_k) + 0.5\,\mathrm{trace}(\mathbf{S}_k^{-1}\mathbf{S}_o + \mathbf{S}_o^{-1}\mathbf{S}_k - 2\mathbf{I})$, where $\mathbf{I}$ Identity matrix, $\mathrm{trace}(\cdot)$ Trace of a matrix
Transformed Divergence	$\mathrm{d_{td}}(k,o) = a\left[1 - \mathrm{e}^{-\frac{\mathrm{d_d}(k,o)}{b}}\right]$ $\mathrm{d^*_{td}}(k,o) = a\left[1 - \mathrm{e}^{-\tan^2\left(\frac{\mathrm{d_d}(k,o)}{b}\right)}\right]$ a, b selected saturation and range values, e.g. 2000 and 8, respectively (Card and Angelici, 1983)
Mahalanobis	$\mathrm{d_m}(k,o) = (\mathbf{m}_o - \mathbf{m}_k)^T\left[\mathbf{S}_k^{-1} + \mathbf{S}_o^{-1}\right](\mathbf{m}_o - \mathbf{m}_k)$ $\mathrm{d^*_m}(k,o) = (\mathbf{m}_o - \mathbf{m}_k)^T\mathbf{S}_{ok}^{-1}(\mathbf{m}_o - \mathbf{m}_k)$ for $\mathbf{S}_o = \mathbf{S}_k = \mathbf{S}_{ok}$
Patrick-Fischer	$\mathrm{d_{pf}}(k,o) = (2\pi)^{-0.5D}\left[\lvert 2\mathbf{S}_k\rvert^{-0.5} + \lvert 2\mathbf{S}_o\rvert^{-0.5} - 2\lvert\mathbf{S}_k + \mathbf{S}_o\rvert^{-0.5}\,\mathrm{e}^{0.5(\mathbf{m}_o - \mathbf{m}_k)^T[\mathbf{S}_k + \mathbf{S}_o]^{-1}(\mathbf{m}_o - \mathbf{m}_k)}\right]$

(Sources: Webb, 1999; Kittler, 1986; Swain and Davis, 1978)

The most commonly used parametric probabilistic distance measures for normal class distributions are summarised in table 2.2. Recursive equations exist for these distance measures which allow a more rapid computation as one feature is removed or added to the subset (Goodenough *et al.*, 1978).

The Mahalanobis distance may be interpreted as the square of the distance between two class means expressed in terms of variances. It is identical to the quadratic distance (see table 2.1) with $\mathbf{Q}$ being the sum of the inverse class covariance matrices. However, it does not account for the differences between the class covariance matrices, which may provide further class discriminatory information, especially for hyperspectral data.

The Divergence measure combines the Mahalanobis distance with a term that exploits the differences between the class covariance matrices. But, both the Mahalanobis and the Divergence criterion suffer from the same unsaturating behaviour as the heuristic distance measures: they continue to increase for growing class mean distances, even after the classes are fully (100%) separated (Thomas *et al.*, 1987).

The Transformed Divergence takes the unsaturating nature of the Divergence measure into account and transforms it into a saturating function with a saturation value a and range value b. Different functions are proposed in Swain and Davis (1978), two of which are shown in table 2.2. Generally, the Transformed Divergence is regarded as superior to the Divergence measure.

The Jeffries-Matusita distance is the saturated form of the Bhattacharyya distance. It is computationally less efficient than the Transformed Divergence measure, since one additional matrix inversion has to be performed per class pair.

The analytical derivation of the above probabilistic class distance measures can be found in Swain and Davis (1978), Kittler (1986) or Thomas *et al.* (1987).

However, the classes may not be normally distributed for the given features and above distance measures may prove to be ineffective. In this case, Shaban and Dikshit (2001) used the skewness and kurtosis moment measures (equations 2.6 and 2.7) as criteria to select textural and spectral features for the classification of urban areas. They found that the absolute and averaged multi-class form of these measures were inverse proportional to the classification accuracy (KHAT coefficient).

4) Multi-class distance measures

The most common multi-class form of any of the above two-class distance measures is calculated as the sum of all pairwise distances ('maximum average', equation 2.16). In the case of probabilistic distance measures, the two-class distance value may be multiplied first by its corresponding class *a*

priori probabilities before being summed (equation 2.17). The use of the prior probability is important, especially when one or two classes are spatially dominant (Borak and Strahler, 1999). However, the latter criterion will not maintain a close relationship with the probability of error (Kittler, 1986) and may give sub-optimal results.

$$\mathrm{d}_{\mathrm{mh}} = \sum_{k=1}^{M} \sum_{o=k+1}^{M} \mathrm{d}(k,o) \tag{2.16}$$

$$\mathrm{d}_{\mathrm{mp}} = \sum_{k=1}^{M} \sum_{o=k+1}^{M} \mathrm{P}(\omega_k)\mathrm{P}(\omega_o)\mathrm{d}(k,o) \tag{2.17}$$

An alternative strategy is to select the feature subset that maximises the minimum distance value calculated from all class pairs. It will ensure that the selected feature set discriminates best between the pair of classes which is hardest to separate ('maximum minimum') (Swain and Davis, 1978).

Bruzzone *et al.* (1995) introduced an alternative to the weighted average multi-class form of the Jeffries-Matusita distance measure, which is equivalent to the Bhattacharyya bound (equation 2.18), but cannot be proved to always perform better than the weighted average criterion for the Jeffries-Matusita distance. However, it gives more importance to classes with low *a priori* probabilities.

$$\mathrm{d}_{\mathrm{mjm}} = \sum_{k=1}^{M} \sum_{o=k+1}^{M} \sqrt{\mathrm{P}(\omega_k)\mathrm{P}(\omega_o)}\,\mathrm{d}_{\mathrm{jm}}^{2}(k,o) \tag{2.18}$$

Bruzzone and Serpico (2000) proposed a criterion for multiclass cases that is related to the upper bound to the Bayes error probability and assumes Gaussian class distribution and equal class covariance matrices. The criterion was found to give greater weight to classes with low *a priori* probabilities as compared to the weighted average criterion. Its main drawback is that it does not use the class covariance matrices to measure the separability between classes.

Bruzzone (2000) presented an extension to the average weighted multi-class distance measure for the Bayes classifier for minimum cost, taking into account the cost associated with each confused class pair (equation 2.19). The latter was developed in the context of producing a land cover map to assess the risks of natural disasters, where it was critical not to confuse classes with highly different risks. A high cost corresponds to a situation in which the confusion of two classes is very critical.

$$\mathrm{d}_{\mathrm{mc}} = \sum_{k=1}^{M} \sum_{o=k+1}^{M} (c_{k,o} - c_{o,o})(c_{o,k} - c_{k,k})\mathrm{P}(\omega_k)\mathrm{P}(\omega_o)\mathrm{d}(k,o) \tag{2.19}$$

Where $c_{k,o}$ Cost of deciding $\mathbf{x} \in \omega_k$, when in reality $\mathbf{x} \in \omega_o$

Alternative heuristic multi-class criterion functions are based on sample based estimates of the within-class ($\mathbf{S}_w$) and between-class ($\mathbf{S}_b$) scatter matrices, which are defined as follows (Fukunaga, 1972):

$$\mathbf{S}_w = \sum_{k=1}^{M} P(\omega_k)\mathbf{S}_k \tag{2.20}$$

$$\mathbf{S}_b = \sum_{k=1}^{M} P(\omega_k)(\mathbf{m}_k - \mathbf{m}_0)(\mathbf{m}_k - \mathbf{m}_0)^T \tag{2.21}$$

$$\mathbf{m}_0 = \sum_{k=1}^{M} P(\omega_k)\mathbf{m}_k \tag{2.22}$$

The prior probability of class ω_k, $P(\omega_k)$, can be approximated by the ratio of the number class training samples to the total number of training samples (Webb, 1999).

Several class separability measures have been derived from these scatter matrices, which aim to minimise the within-class spread and maximise the between-class spread simultaneously (Fukunaga, 1972). For example, a possible criterion is given by Liu and Jernigan (1990):

$$d = \frac{|\mathbf{S}_w + \mathbf{S}_b|}{|\mathbf{S}_w|} \tag{2.23}$$

Where $|\cdot|$ Determinant of a matrix

Penaloza and Welch (1996) defined d_p as separability measure:

$$d_p = \text{trace}(\mathbf{S}_w^{-1}\mathbf{S}_b) \tag{2.24}$$

However, the latter type of measures based on scatter matrices presents the same disadvantage as other heuristic distance measures: they do not exhibit a saturating effect for large distance values. Moreover, within-class covariance matrices are not estimated reliably with a limited number of training samples (Jia and Richards, 1999). For more details, the reader is referred to Devijver and Kittler (1982).

San Miguel-Ayanz and Biging (1996) avoided using multi-class measures by considering only a class pair at each stage of a progressive multiple-stage feature selection and maximum likelihood classification approach. The procedure allows classifying each class pair on an individually selected best feature subset. Jia and Richards (1998) extended this binary method by allowing the classifier to adapt to the amount of training samples available for each class. The method chooses the maximum-likelihood classifier if sufficient training samples are available for the two classes, and the Euclidean distance classifier if not. Although increasing the computational cost, the multiple-stage

method gave an improved classification performance compared to the single-layer classification (San Miguel-Ayanz and Biging, 1996; Jia and Richards, 1998).

Choice of criterion function

Unless class samples are not available, feature selection for classification should be based on supervised criteria to approximate best the classification accuracy.

With respect to the classification accuracy estimates presented above, there are no clearcut rules as to when each measure should be used (Congalton, 1991). The feature selection process employing these measures then consists of a series of classifications using different band combinations, with the band subsets being ranked according to their resulting mapping accuracy. The advantage of this approach is that the ultimate accuracy of the actual classifier is directly optimised instead of a non-specific distance measure used for its approximation. However, since all classification accuracy estimates require the construction of the classifier to be used, the feature selection procedure may involve high computational cost. Despite this, some authors used this method for feature selection, e.g. Penaloza and Welch (1996).

The choice of a supervised distance measure should be guided by its ability to represent the probability of error accurately. Generally, probabilistic measures may outperform non-probabilistic measures, since, apart from the class mean vectors, they make use of the class covariance matrices, describing the entire shape of the class distribution. Landgrebe (1999, 2000) claimed that second-order statistics were more relevant than first order statistics in discriminating among classes for high dimensional data. Kittler (1986) stressed that first-order information could be entirely wiped out by large class variances and covariances. However, in cases where the execution speed of the feature selection algorithm is important, heuristic distance measures may be the preferred to the more accurate probabilistic measures.

Generally, the chosen feature selection criteria should reflect the complexity of the classification approach used. If, for example, a minimum-distance classifier is used for classification, the same heuristic distance measure should be applied also for feature selection.

If a probabilistic classifier is used, saturated probabilistic measures (such as the Transformed Divergence and the Jeffries-Matusita) are usually considered superior to their unsaturated counterparts (Divergence and Bhattacharyya) for representing error probability in most parts of the separability range. However, in some parts where the saturating function does not approximate well the behaviour of the error probability, they are inferior to the unsaturated measures. Thomas *et al.* (1987) suggested the following approach for choosing which measure to use:

1) If relatively homogeneous classes are used together with a reduced number of bands, the Divergence measure should be chosen.
2) If relatively homogeneous classes are used together with all bands, the Transformed Divergence measure should be chosen.
3) If less truly homogeneous classes are used together with a reduced number of bands, the Bhattacharyya Distance measure should be chosen.
4) If less truly homogeneous classes are used together with all bands, the Jeffries-Matusita Distance measure should be chosen.

Some researchers made experimental comparisons to evaluate the performance of different criteria for a specific data set. Typically, a classification accuracy estimate was used as performance measure, obtained by the feature subsets selected by the different criterion functions as input to the chosen classifier. Examples are Goodenough *et al.*(1978), San Miguel-Ayanz and Biging (1996), Mausel *et al.*(1990), Kavzoglu and Mather (2000), Bruzzone and Serpico (2000). However, to isolate the effects of the various criterion functions on the classification accuracy estimate, an optimal search algorithm should be used for feature selection, together with a cross-validation method for accuracy estimation to minimise the effect of the generalisation error (Bruzzone and Serpico, 2000). Chen (1973) pointed out that incorrect assumptions on class distribution or the limited number of available training samples might result in erroneous conclusions. He claimed that only a theoretical comparison could provide unique results.

To summarise, rigid recommendations as to which measures should be used cannot be made because of the choice being very much problem-specific and dependent on the classifier to be used. Clearly, classification accuracy estimates are the preferred choice, but may result in a computationally unfeasible task. In cases where the data are not normally distributed, Webb (1999) recommended, from a computational point of view, to use a measure that simplifies for normal distributions. However, the effect of non-normality is not known (Mather, 1999).

2.2.3 Search Algorithms

In feature selection, the search algorithm generates and compares possible feature subset solutions utilising the criterion function as a measure of the effectiveness of the feature subsets under consideration. The most effective feature subset optimises the criterion and is the result of the search algorithm. Commonly used selection methods employ either optimal or sub-optimal search algorithms.

Optimal search algorithms

Optimal methods consider all feature subsets resulting in a globally optimal solution. They include exhaustive search methods and accelerated search methods (Branch-and-Bound methods).

The exhaustive search method evaluates the criterion function for each candidate feature subset individually. To select D features out of a possible set of N features ($D < N$), the number A of possible subset combinations (without repetition) equals:

$$A(N,D) = \binom{N}{D} = \frac{N!}{D!(N-D)!} \tag{2.25}$$

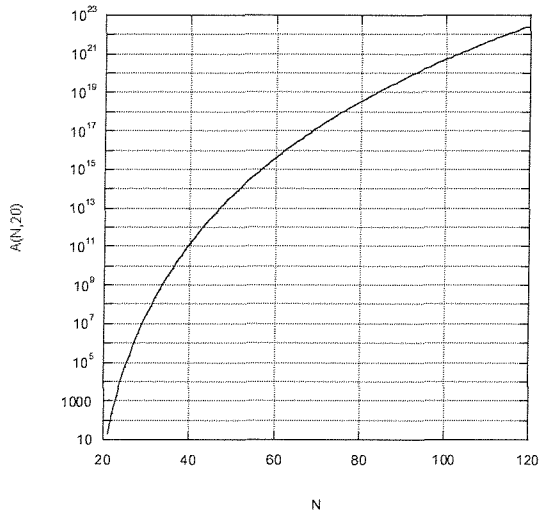


Figure 2.1: Number of possible subset combinations A(N,20).

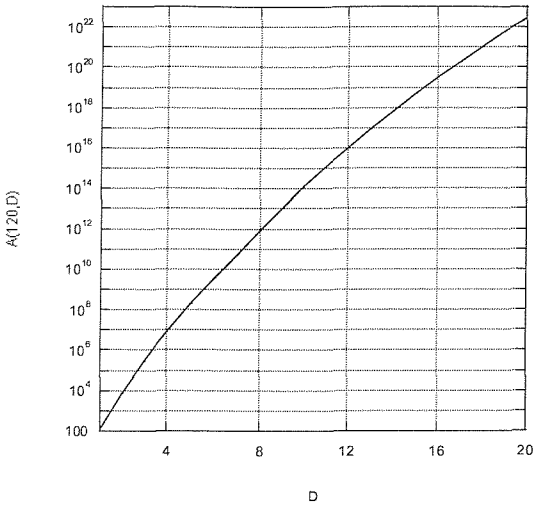


Figure 2.2: Number of possible subset combinations A(120,D).

Figures 2.1 and 2.2 illustrate A as a function of increasing D and N, respectively. It can be seen from these figures that A increases exponentially with increasing dimension N, and that A is excessive even for moderate values of D and N. That is, for a hyperspectral data set with around a hundred bands, the exhaustive search method would represent unacceptable computational cost and could very quickly cease to be computationally feasible, considering the computation effort involved in estimating the criterion function alone.

Consequently, an exhaustive search may be applied only to a data set of small dimensionality N, or to a very small subset dimension D. In the following paragraphs alternative search procedures are presented that find optimal and sub-optimal feature combinations at reduced computational cost.

The Branch-and-Bound algorithm (Narendra and Fukunaga, 1977) relies on the feature selection criterion being monotonic. The monotonicity property of a criterion implies that for a nested feature set X_i, related as $X_1 \subset X_2 \subset X_3 ... \subset X_p$, the criterion function satisfies

$d(X_1) \leq d(X_2) \leq d(X_3) ... \leq d(X_p)$ (Kittler, 1986). In other words, the criterion value of a feature subset is smaller than or equal to the criterion value calculated on the entire feature set. The procedure is a top-down search method, starting with the entire N -dimensional data set and constructing a tree structure by deleting features successively. First, the criterion value of the entire feature set is computed as an upper bound $d(X_p)$. Second, the optimal $(N-1)$-dimensional subset is found from $\begin{pmatrix} N \\ N-1 \end{pmatrix}$ possible subsets, giving a lower bound $d(X_{p-1})$ to $d(X_p)$. All other non-optimal subsets with smaller criterion values than $d(X_{p-1})$ are eliminated, since the deletion of any additional features of one of these subsets will result in a further decrease of its criterion value (according to the monotonicity criterion). In such a way, the subset is reduced consecutively until the desired feature subset size D is reached (Kittler, 1986).

The algorithm considers all feature subsets by either implicitly rejecting or explicitly evaluating them, and ensures that the globally optimal feature subset will be found (Goodenough *et al.*, 1978).

Sub-optimal search algorithms

In contrast to optimal methods, sub-optimal search algorithms do not evaluate all possible feature combinations. They can be grouped into deterministic methods that produce the same subset on a given problem every time, and stochastic methods that include a random element and may produce different subsets every run.

The deterministic methods typically construct the final feature subset incrementally by including ('bottom-up') or excluding ('top-down') features from preceding subsets. Bottom-up methods start with an empty set, while top-down strategies begin with the full feature set. Since the best subset of features is not necessarily made up of the features of preceding subsets (due to the correlation between features), the search result will not be the optimal one. The two stochastic methods presented here are Monte Carlo optimisation methods for combinatorial problems: genetic algorithms and simulated annealing. They use random elements to search the feature space for the sub-optimal subset.

1) Best individual

The 'best individual' method evaluates each feature of the original set individually with respect to the feature selection criterion. It is the simplest method, and may give some good results if the features in the original set are uncorrelated (Webb, 1999).

2) Sequential forward selection (SFS)

The SFS algorithm is a bottom-up method, which starts with an empty set. First, the best individual feature is selected. Then, at each subsequent stage, the feature from the remaining original set, that achieves the maximum criterion value in combination with all features already selected, is added to the preceding subset, until the desired subset dimension is reached or the best improvement makes the feature subset worse. Penaloza and Welch (1996) presented a fast version of the SFS, which deletes any features from the available original set if their performance together with the previously selected features they are added to, is worse than the performance of the already chosen features themselves. The generalised form of the SFS algorithm adds at each stage L features ($L > 1$) from the remaining feature set to the selected subset. All possible sets of size L are evaluated, and the set that maximises the criterion value in combination with the already selected subset, is included into the final subset. The sequential forward floating search (SFFS) algorithm is a modification of the generalised form, which allows L to 'float', i.e. L may change at different stages of the search. An example for floating search methods is given by Pudil *et al.* (1994).

The SFS results in a sub-optimal feature set, because selected features that were made redundant by additional ones, cannot be removed from the subset by the algorithm (Kittler, 1986). However, Zongker and Jain (1996, in Webb, 1999) found that the floating search method gave near optimal results close to the ones generated by the much slower Brach-and-Bound method.

3) Sequential backward selection (SBaS)

The SBaS is the top-down analogy to the SFS, starting from the complete feature set and deleting one feature at a time until the desired number of features remain. The feature selected for elimination is the one that results in the lowest reduction in the value of the criterion function. The fast version of the SBaS is similar to the one of the SFS (see above, Penaloza and Welch, 1996). The generalised form of the SBaS reduces the current feature set by R features at a time, and is the counterpart of the generalised form of the SFS. The sequential backward floating selection (SBaFS) algorithm is the counterpart to the SFFS, enabling R to vary during the search process.

The SBaS is a sub-optimal search algorithm, since it does not allow discarded features to be reconsidered (Kittler, 1986). Computationally, the SBaS is more costly than the SFS, because the criterion function is evaluated over larger feature subsets. Zongker and Jain (1996, in Webb, 1999) claim that the floating selection method results were near optimal.

4) Plus L – Take Away R

The 'Plus L – Take Away R' search algorithm combines the generalised forms of the SFS and SBaS methods to remove partially the problem of nesting of the feature set. Due to the correlation

between features, the best feature subset at a given stage does not necessarily include the selected features from a previous stage. After adding the L best features with the SFS, the worst R features are eliminated with the SBaS, etc. The net change in subset size equals to $L - R$. The method is of type bottom-up for $L > R$, and top-down for $L < R$.

5) Max-Min Feature Selection

The 'Max-Min Feature Selection' search algorithm is similar to the SFS method. But instead of adding the best available feature that maximises the criterion function in combination with *all* pre-selected features, it chooses the next available feature that maximises the minimum criterion function value that the new feature achieves with each of the pre-selected features *individually*. This ensures that the new feature adds significant information to all already selected features. The advantage of this search technique is that feature selection is performed on the individual and pairwise merit of features only, i.e. the criterion function needs to be evaluated in feature spaces of maximum two dimensions (Kittler, 1986).

6) Beam Search

The Beam Search technique (Siedlecki and Sklansky, 1988) is an extended version of the artificial intelligence best-first technique. The algorithm starts with the full feature set and creates a queue (of a maximal number) of best subsets with one feature less than the original set according to decreasing values of the criterion function. Next, some subsets of a few dimensions smaller than the current subset are drawn at random and evaluated with the criterion function. If one of the latter subsets has a higher criterion value than the first subset in the queue, then the best subset is chosen from the queue, which includes the subset from the lower level. Otherwise, the first subset from the queue will be chosen as the current subset. This process is repeated until a subset of given dimension is reached. The algorithm may be seen as a generalisation of the sequential feature selection algorithms.

7) Genetic Algorithm (GA)

GA (Holland, 1975) is a stochastic search technique, which helps to choose which candidate solutions should be tested at each stage of the search. The GA concept is inspired by the mechanics of evolution and will be explained here in the context of feature selection.

An initially (e.g. randomly or heuristically) selected collection of feature subsets of the desired dimension D ('population of chromosomes') will be filtered according to the feature selection criterion function ('fitness function'). The eliminated feature subsets will be replaced by new possible subsets ('offspring'), generated via 'crossover' and 'mutation'. Crossover randomly exchanges feature groups ('genetic material') between two subsets chosen by the fitness function

('chromosome parents'). In contrast, mutation randomly replaces a feature by another one chosen out of all possible features. The latter will ensure that the GA will not be limited by local maxima (Tso and Mather, 2001). A binary encoding technique may be used to represent a feature subset by a binary string. For example, a feature may have its fixed place within a sequence of features, and have a value of 1 or 0, depending on whether the feature belongs to the candidate subset or not, respectively. The GA may be iterated until the feature selection criterion converges.

An advantage of the GA is that multiple points in search space are used to start the search for the global maximum (rather than just one starting point). Additionally, the transition from one generation to the next is based on probabilistic rules, rather than on deterministic ones (Goldberg, 1989 in Tso and Mather, 2001). However, many decisions need to be made on the implementation of each step in the algorithm (e.g. initial populations, mutation rate, convergence limit of criterion function or number of iterations), and these might prove to be sub-optimal for complex problems (Prügel-Bennett and Shapiro, 1994).

For more details on GAs the reader is referred to Holland (1975), Mitchell (1996), Tso and Mather (2001), and Siedlecki and Sklansky (1988).

8) Simulated Annealing (SA)

As genetic algorithms, SA can be employed for combinatorial optimisation problems, where a criterion function is to be optimised over a discrete and large configuration space.

SA has its roots in thermodynamics, specifically in the processes of freezing liquids or cooling metals. If a liquid is cooled slowly enough, the freely moving atoms are often able to line themselves up to create a pure crystal, which represents the state of minimum energy for this system. This process of slowly cooling a system allowing ample time for the redistribution of the atoms is termed annealing and was introduced to numerical methods by Metropolis *et al.* (1953). Kirkpatrick *et al.* (1983) applied the same idea to search for solutions of other problems (simulated annealing).

The system starts at an initial configuration, which is updated repeatedly until the system is frozen. The selection of trial moves reachable from the current configuration can be random or informed (e.g. by downhill simplex method). If the trial move improves the search, it is always accepted. But if it makes it worse, it might be still accepted according to a probability given by a Boltzman factor in the change in the criterion function. This allows the algorithm to escape sub-optimal combinations and to continue searching for a better solution. The cooling rate has to be designed to be computationally efficient and to allow the system to find a good solution, i.e. neither too slow nor too fast. It can be determined empirically by running successively slower rates until the solution stops improving. For more details, the reader is directed to Press *et al.* (1992) and Siedlecki and Sklansky (1988).

Choice of search algorithm

If the features are statistically independent, sub-optimal strategies will yield optimal results. However, if the features are correlated with each other, sub-optimal search methods will result in sub-optimal feature subsets, whose departure from optimality is not bounded (Van Campenhout, 1980). This implies that the optimal feature subset will not consist of the optimal subsets of smaller dimension.

The degree of correlation is known to be high in hyperspectral remote sensing bands, and optimal methods that consider all possible feature subset combinations are the only protection against the bias of nesting. But sub-optimal search algorithms still may be preferred to optimal methods because of their computational efficiency, or the inappropriateness of the assumptions made by optimal methods (e.g. monotonicity criterion of the Branch-and-Bound algorithm).

The Monte Carlo optimisation methods (genetic algorithms and simulated annealing) were found to be extremely well suited for large-scale feature selection problems (Siedlecki and Sklansky, 1988). Kavzoglu and Mather (2000) compared the results of the SFS with the ones of the GA using a range of criterion functions. They found that the features selected by the GA algorithm performed better than those selected by the SFS.

Labovitz (1986) compared the result of the exhaustive search with the subset chosen by the sub-optimal SFS. An average decrease of 7-10% in classification accuracy was found when using the sub-optimal method instead of the optimal one. Labovitz (1986) predicted that this difference in classification accuracy would be more dramatic with hyperspectral data, where many more bands allow for a higher number of band combinations (less likelihood that the optimal band subset will be selected by chance) and the bands are typically highly correlated.

It is generally recognised that no perfect search algorithm exists and Press *et al.* (1992) recommended the use of more than one method in comparative fashion.

2.3 Feature Extraction

2.3.1 Introduction

Generally, feature extraction methods aim to map all of the original measurements into more informative new features for the purpose of data compression. The resulting feature vector $\mathbf{y} = [y_1 \quad \ldots \quad y_N]^T$ results from a transformation F of the original measurement vector $\mathbf{x} = [x_1 \quad \ldots \quad x_N]^T$, optimised by a criterion function:

$$\mathbf{y} = \mathrm{F}(\mathbf{x}) \tag{2.26}$$

where $\mathbf{x}$ Original pixel vector
$\mathbf{y}$ Transformed feature vector
F Mapping function

All supervised and unsupervised criterion functions used for feature selection (section 2.2.2) may be readily applied as criterion functions for feature extraction.

The natural form of the optimal mapping function F will be unknown. In this thesis, only linear transformations will be considered, since they have the advantage of being simple, well defined, analytically tractable and computationally feasible (Kittler, 1986). The problem of finding the optimal mapping function is then reduced to obtaining the coefficients $a_{j,l}$ of the linear function so as to maximise or minimise the criterion function:

$$\mathbf{y} = \mathbf{A}^T\mathbf{x} \tag{2.27}$$

$$y_j = a_{j,1}x_1 + a_{j,2}x_2 + \cdots + a_{j,N}x_N \tag{2.28}$$

where $a_{j,l}$ Linear mapping coefficients
$\mathbf{A}$ Transformation matrix (N,N) with elements $a_{j,l}$

The mapping coefficients may be found analytically for simple criteria by expressing the parameters of the criterion function in the mapped feature space via the transformation matrix, constructing the first derivative of the criterion and using numerical methods (Kittler, 1986). For more complex criteria, common multidimensional optimisation techniques for continuous functions can be drawn upon that require criterion function evaluations only, for example, the downhill simplex method, direction-set (Powell's) methods, GA and SA (Fukunaga, 1972).

The transformation scheme is closely linked with the criterion used (Kittler, 1986). In cases where important features are highly non-linear functions of original measurements, a suitable non-linear mapping function has to be found. However, a general theory to generate mapping functions systematically and to find the optimum one is not available and the selection of features becomes very much problem orientated (Fukunaga, 1972). In these cases, multi-layer neural networks may be used to perform non-linear feature extraction (Bishop, 1995).

Feature extraction methods may be used to generate and select features, and to determine the intrinsic dimensionality of a data set according to some criterion. The following sections investigate common unsupervised and supervised feature extraction methods used in pattern recognition and remote sensing.

2.3.2 Unsupervised Feature Extraction

When the classes of interest are unknown or poorly defined, the information content in the data may be measured via descriptive statistics, such as the data variance, correlation coefficient, etc. (see sections 2.1 and 2.2.2). In this section, frequently used feature extraction techniques using unsupervised criterion functions are presented.

Principal components analysis (PCA)

PCA (Pearson, 1901) is an unsupervised feature extraction technique to reduce the number of features to a smaller number of uncorrelated indices called principal components (PCs). This is achieved by diagonalising the global data covariance matrix via eigenanalysis:

$$\mathbf{S}\mathbf{a}_l = \lambda_l \mathbf{a}_l \tag{2.29}$$

where $\mathbf{a}_l$ Eigenvector of covariance matrix $\mathbf{S}$

λ_l Eigenvalue of covariance matrix $\mathbf{S}$

The PCs are linear combinations of the original bands. The eigenvectors $\mathbf{a}_l$ form the columns of the transformation matrix $\mathbf{A}$ (equation 2.27). Geometrically, PCA can be seen as a rotation of the axes of the original coordinate system to a new set of orthogonal axes, the eigenvectors. The information can then be presented in component images, which are projections of the data onto the eigenvectors.

The eigenvector coefficients, or loadings, can be plotted as an eigenfunction of the original features. Given some knowledge of the ground scene, the shapes of these curves may be interpreted to some degree. Generally, the magnitude of the eigenvector loadings indicates the relative importance of their corresponding original features in representing the newly transformed 'optimal' feature (Wiersma and Landgrebe, 1980).

The eigenvalues λ_l of the PCs correspond to their ability to account for the data variance and are ordered by decreasing amount of variance. When the noise variance can be assumed to be the same in all bands, the amount of data variance captured in a PC can be interpreted as a measure of its information content (Manly, 1994).

Since PCA is based on the data covariance matrix $\mathbf{S}$, a scale-dependent bias may be introduced in subsequent analysis. A standardised form of the PCA may be achieved by using the data correlation matrix instead of the covariance matrix for eigen-decomposition, or simply by transforming the original data first to zero mean and unit variance.

Other variants include the independent component analysis (ICA)(Karhunen *et al.*, 1997), which does not apply the orthogonal constraint to the axes transformation, and the common PCA (Flury, 1988), which generalises to several populations within the data, assuming identical principal components for all groups, but different variances. Nirala and Venkatachalam (2000) examined PCA versions, which maximise the data covariances or the covariance-to-variance ratios instead of the data variances.

Jia and Richards (1999) proposed a segmented PCA method, in which the complete band set is ordered and partitioned into subgroups of highly correlated bands that are transformed individually with PCA. The most informative PCs of all subgroups are regrouped and compressed further via PCA, until the required data reduction ratio is achieved for classification or storage purposes. The method is based on the idea that PCA works most efficiently on highly correlated features.

Wiersma and Landgrebe (1980) suggested the use of a diagonal weight matrix $\mathbf{W}$ in equation 2.29, specifying a weight factor for each original band to mask out noisy spectral bands:

$$\mathbf{SW}\mathbf{a}_l = \lambda_l \mathbf{a}_l \tag{2.30}$$

where $\mathbf{W}$ Diagonal matrix of weight coefficients

Maximum noise fraction (MNF)

Green *et al.* (1988) introduced the MNF transformation, or noise adjusted PCA, which chooses the new components in order to maximise the image SNR as indicator for image quality. Instead of the covariance matrix, the noise covariance matrix is diagonalised to minimise the noise effects on the signal sources. The MNF was designed for data that have unequal noise variances in different bands, such as imaging spectrometer data, where different levels of noise may be present in different regions of the optical spectrum (e.g. non-uniform detector noise). Being dependent on only the SNR, the transformation is invariant under scale changes to any band.

The image noise covariance matrix may be estimated with the sensor noise covariance matrix derived from dark reference measurements during a laboratory or in-flight sensor calibration (e.g. Lee *et al.*, 1990; Smith and Curran, 1999). Dark current images are at present also routinely collected before or after data acquisition for most airborne sensors. However, these data are generally not made available to the end user. Although accurate, instrument noise estimates do not account for atmospheric noise contributions (i.e. absorption and scattering). The latter may be partly avoided by choosing bands from well within atmospheric windows only.

Random image noise may also be quantified over spatially and spectrally homogeneous image areas, by subtracting from the current pixel neighbouring pixels, the local mean or median (Nielsen,

1994), or by calculating the local standard deviation (e.g. Fujimoto *et al.*, 1989). Other approaches to noise estimation using the principle of homogeneous image areas include

- dividing the image into homogeneous blocks and using the mean block standard deviation as noise estimate of the entire image (Gao, 1993),
- deriving residuals from a linear regression which models the pixel's response as a function of its responses from adjacent wavebands and one adjacent pixel (Roger and Arnold, 1996),
- deriving residuals in a simultaneous autoregressive model or from a fitted quadratic surface in a neighbourhood (Nielsen, 1994),
- using image training pixels to infer a maximum-likelihood estimate of the noise covariance matrix (Settle and Drake, 1993),
- using the square root of the nugget variance as noise estimate (Curran and Dungan, 1989),
- using the spatial correlation between adjacent pixels (minimum/maximum autocorrelation factors, MAF; Switzer and Green, 1984; Green *et al.*, 1988)
- using the ratio of local variance to global variance (Warner and Shank, 1997; Warner *et al.*, 1999), where the local variance is estimated as the sum of the squared difference between data values of neighbouring pixels over the entire scene (Geary's c metric; Geary, 1954).

The homogeneous image areas may be defined with the analyst's scene knowledge as an area of one object class, or with a clustering algorithm. The accuracy of the noise estimate then depends on 1) the homogeneity of the defined regions, 2) their signal level (low-signal regions contain less noise than high-signal ones), and 3) them being representative of the entire image.

However, if the image areas under investigation include some spatial variation, the image noise will be over-estimated. Boardman and Goetz (1991) developed a method, which estimates this scene variation by extrapolating the noise standard deviation for an infinite number of homogeneous pixels. The method assumes that the instrument and atmospheric noise will be averaged out for an increasing number of pixels used.

Some of the 'homogeneous area'-based methods mentioned above were reviewed and compared by Smith and Curran (1999) for image SNR estimation, and the authors concluded that different methods achieved different SNR results. Image noise may therefore best be estimated by sensor noise alone, i.e., using dark reference measurements.

Projection pursuit (PP)

PP is the term used to describe a method of numerical optimisation of a criterion (projection index or PI) for finding the most 'interesting' low-dimensional projections of a high-dimensional data cloud (only linear and orthogonal transformations are considered here). Friedman and Tukey (1974) first successfully implemented the technique, whose basic strategy is to project the data first with an initial transformation matrix $\mathbf{A}_0$ into lower-dimensional feature space, in which $\mathbf{A}_0$ is recomputed

by optimising the PI. The method is repeated, until the transformation matrix does not change any more.

So far, many authors have chosen unsupervised projection indices that quantify the difference between the Gaussian and the estimated probability distribution of a projection, since they believed to detect spectral classes via 'anomalous', i.e. multimodal and skewed distributions. For example, Ifarraguerri and Chang (2000) used the information divergence as PI, while Chiang and Chang (1999, in Ifarraguerri and Chang, 2000) based their PI on skewness and kurtosis statistics.

PCA may be interpreted as a special case of the PP methods using the proportion of the total variance accounted for by the projected data as projection index (Jones and Sibson, 1987). Bachmann and Donato (2000) compared the performance of the unsupervised PP method, using the product of trimmed variance and compactness function as projection index, with PCA for separating land cover classes. They found that PCA works best for separating broad class clusters, but performs worse than PP for classifying narrower sub-clusters.

Ifarraguerri and Chang (2000) preprocessed their data by applying PCA and scaling the PC with their corresponding eigenvalue (sphering). They chose a suboptimal optimisation technique by taking every pixel spectrum as possible projection vector. However, the optimal projection may not be found along a given data point. Using the information divergence as PI, they compared the resulting PPs with the PCs, which showed that the PPs tend to correspond better to spectral objects than PCs.

Factor analysis (FA)

FA aims to represent a set of features in terms of a smaller underlying set of variables called factors. The factor analysis model may be formulated as follows (Harman, 1976):

$$\mathbf{x} = \mathbf{\Lambda}\xi + \boldsymbol{\varepsilon} \tag{2.31}$$

where ξ Common factor vector
$\mathbf{\Lambda}$ Matrix of factor loadings
$\boldsymbol{\varepsilon}$ Specific or unique factor vector

This approach differs from the general feature extraction methodology presented here in that the pixel measurements $\mathbf{x}$ are expressed as a linear combination of the unobserved factors ξ and $\boldsymbol{\varepsilon}$, which are chosen to account for the correlation between the measurements. In contrast to PCA, FA is based on a statistical model, requires a large number of assumptions, and has no unique solution due to the many different methods of obtaining the factor loadings $\mathbf{\Lambda}$. Factor analysis is a widely criticised method, mainly due to the subjectivity of the decisions involved (Webb, 1999) and will not be pursued any further here. For a detailed review, the reader is referred to Mather (1976) or Webb (1992).

2.3.3 Supervised Feature Extraction

If feature extraction is applied for classification processing, the classification accuracy has to be considered as the primary criterion function in the extraction process. All supervised criterion functions employed for feature selection (section 2.2.2) may be applied as surrogates for classification accuracy estimates.

Karhunen-Loève (KL) expansion

The KL transformation is in one of its most basic forms identical to PCA, but includes variants that use supervised criteria. The eigenanalysis may be performed on the within-class covariance matrix $\mathbf{S}_w$ (equation 2.15) as opposed to the global covariance matrix $\mathbf{S}$ in the case of PCA. The transformation matrix $\mathbf{A}$ is formed by the eigenvectors resulting from eigen-decomposition (as for PCA, section 2.3.2). The ordering of the new uncorrelated features may be carried out on the basis of the eigenvalues alone or in combination with a measure of discriminability. Webb (1999) gives a detailed review about the KL variants proposed by Chien and Fu (1967), Devijver and Kittler (1982) and Kittler and Young (1973).

Linear discriminant analysis (DA) / canonical analysis (CA)

In the context of feature extraction, linear DA is understood as a method of finding a transformation that maximises between-class separability and minimises within-class variability. It is based on Fischer's two-class criterion, which may be generalised to the multiclass situation by using the sample based estimates of within-class and between-class scatter matrices, $\mathbf{S}_w$ and $\mathbf{S}_b$ (equations 2.15 - 2.17)(Webb, 1999):

$$d_F = \frac{\mathbf{a}_k^T \mathbf{S}_b \mathbf{a}_k}{\mathbf{a}_k^T \mathbf{S}_w \mathbf{a}_k} \tag{2.32}$$

Generally, the columns $\mathbf{a}_k$ of the transformation matrix $\mathbf{A}$ are chosen to maximise equation 2.32. Canonical analysis is identical to linear discriminant analysis expect of being constraint by transforming $\mathbf{S}_w$ into the identity matrix $\mathbf{I}$, giving orthogonal and uncorrelated features in the transformed space (Chen and Landgrebe, 1989). It leads to the following eigenvector equation:

$$\mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{a}_k = \lambda_k \mathbf{a}_k \tag{2.33}$$

However, the matrix $\mathbf{S}_w^{-1}\mathbf{S}_b$ is not symmetric, but can be reduced to a symmetric eigenvector problem using Cholesky decomposition (Press *et al.*, 1992).

Foley and Sammon (1975) proposed an alternative approach to find orthogonal transformed axes progressively: the first vector $\mathbf{a}_1$ is to maximise equation 2.32. The second vector $\mathbf{a}_2$ is chosen to maximise equation 2.32 subject to the orthogonality constraint $\mathbf{a}_2\mathbf{a}_1 = 0$, etc. Merembeck and Turner (1980) proposed a directed form of DA, which allows to account for user-specified contrasts defining underlying relationships among categories known *a priori* to exist within the data. This technique can only be used for adequate sample sizes.

CA produces a transformed feature space of dimension $M-1$, where M is the number of classes. The method is only appropriate for classes with different mean vectors, since it does not exploit differences between class covariance matrices (Hsieh and Landgrebe, 1998). Additionally, a class mean vector that is very different from the mean vectors of the other classes will dominate the calculation of the between-class scatter matrix. Moreover, the method requires a sufficiently large amount of training samples to estimate the within-class covariance matrix, and does not have a direct relationship to the probability of classification error.

On the other hand, CA is simple, popular, fast and easy to implement (Hsieh and Landgrebe, 1998), and, in contrast to PCA, it does not demand any correlation between the bands (Csillag *et al.*, 1993).

Decision boundary feature extraction (DBFE)

Lee and Landgrebe (1993) introduced DBFE, a feature extraction method based on decision boundaries. The method classifies the training sample data directly with the Bayes' decision rule for minimum error to locate decision boundaries between the classes. Lee and Landgrebe (1993) called the normal vectors to the decision boundaries 'discriminantly informative' features and creates a 'decision boundary feature matrix' with them. The eigenvectors of the latter matrix represent the columns of the transformation matrix $\mathbf{A}$, which results in the desired new feature vectors with the eigenvalues being directly related to the usefulness of the corresponding features for discriminating between the given classes.

DBFE has the advantage of not making assumptions about the underlying probability distribution of the training classes. In addition, its performance does not get worse when the differences in mean or covariance are small. However, the procedure suffers from Hughes phenomenon (section 1.1) and requires a large number of training samples for good performance, at least the number of dimensions plus one to avoid singularity. DBFE is suboptimal for more than two classes and very time-consuming (Hsieh and Landgrebe, 1998; Tadjudin and Landgrebe, 1998).

Projection pursuit (PP)

The PP method has been introduced earlier in section 2.3.2. In this paragraph, only the supervised PP variants are presented, i.e. with 'interestingness' defined as ability to differentiate between classes and the PI based on class statistics. All of the supervised feature selection criteria (section 2.2.2) may be promptly used for PP.

Petraskos *et al.* (1999) used the 'maximum minimum' multiclass version of the Mahalanobis, Bhattacharyya and Jeffries-Matusita distance as PIs to find the best PP features for three-dimensional visualisation of the classes under investigation.

Jimenez and Landgrebe (1999) proposed a sequential version of the PP where the original feature space is partitioned into groups of adjacent bands that are projected individually in turn to maximise the minimum Bhattacharyya distance among classes, while keeping the other transformation vectors constant and orthogonal to each other. The iteration stops when the initial (or updated) transformation matrix $\mathbf{A}_0$ no longer changes significantly.

The grouping of bands is based on a hybrid binary decision-tree technique. The latter method keeps dividing the groups of bands symmetrically into two halves (top-down approach), as long as the improvement of the PI relative to the previous separation exceeds a predefined threshold value. If no more improvement can be achieved, the corresponding bottom-up method is applied, where adjacent groups of bands are merged when increasing the PI significantly with respect to the previous grouping, etc.

This sequential method ensures that the extracted features relate to certain parts of the spectrum and can therefore be interpreted by the user. But only the best feature out of one group is chosen and any additional dimensions are neglected. In addition, the final feature set is not optimal in a global sense.

Jimenez and Landgrebe (1999) also suggested a feature selection procedure based on the above grouping method by choosing the best band out of each group for maximising the PI. Nevertheless, the bands are only selected best from adjacent features, but not globally from the entire band set.

Generally, PP has the advantage of estimating the class statistics more accurately in the projected low- rather than in the original high-dimensional space for limited training samples, especially true for hyperspectral data sets. In addition, Jimenez and Landgrebe (1999) point out that the assumption of a normal distribution of class data, mostly used in projection indices, is more justified in a subspace emerging from a linear projection of a higher dimensional space, as the dimensionality tends to infinity.

Multidimensional scaling (MDS)

MDS refers to a variety of feature extraction techniques that use supervised two-class distance measures as criterion functions (Webb, 1999). Only linear (metric) scaling techniques are investigated in this paragraph. MDS for feature extraction is termed MDS by transformation.

MDS by transformation aims to minimise the discrepancy between the criterion function value d_{ko}, measured in the original feature space between pixels k and o, and the distance value D_{ko} between pixels k and o in the new feature space to be found. An objective function O can be formulated as a function of this discrepancy, which is to be minimised with respect to the transformation parameters. Many forms exist for the objective function O. An example is given in equations (2.34 - 2.36) (Webb, 1999):

$$\mathrm{d}_{ko} = \left|\mathbf{x}_k - \mathbf{x}_o\right| \tag{2.34}$$

$$\mathrm{D}_{ko} = \left|\mathbf{A}^T\mathbf{x}_k - \mathbf{A}^T\mathbf{x}_o\right| \tag{2.35}$$

$$\mathrm{O}(\mathbf{A}) = \sum_{1\le o<k\le n}\left(\mathrm{d}_{ko} - \mathrm{D}_{ko}\right) \xrightarrow[\mathbf{A}]{} \min \tag{2.36}$$

where

d_{ko}	Supervised class distance (or separability) measure, measured in original feature space
D_{ko}	Distance measure, measured in transformed feature space
O	Objective or stress function

If d_{ko} corresponds exactly to the Euclidean distance measure (as above), the classical form of MDS results, which gives an identical transformation as PCA. Usually standard gradient methods are used for optimisation (Webb, 1999). The main problem for the application of MDS to remote sensing data is that all n image pixels need to be fed into the optimisation algorithm (2.36) and the storage of an (n, n) matrix may be problematic for large n.

2.3.4 Choice of Feature Extraction Technique

If no ground data are available, unsupervised feature extraction techniques may help to reduce the dimensionality of the data set. From the unsupervised methods presented above, the author recommends the PCA, mainly due to its simplicity and ease of implementation. If the features are measured in different scales, the standardised version of the PCA should be applied. However, this is unlikely for optical remote sensing data. The MNF is advised only when the detectors of the sensor are known to produce different noise variances and dark reference measurements are available. However, techniques for the estimation of the noise covariance matrix are not available for including structured noise such as edge effects or image striping. The latter should be dealt with via common image enhancement techniques. Unsupervised PP does not seem to offer any particular

benefit in comparison to PCA, and is more difficult to implement. Factor analysis is generally considered as an ambiguous and subjective method, not recommended for dimensionality reduction.

In feature extraction for classification, the new features should be chosen so as to discriminate best between the user-defined classes (Kittler, 1986). Since unsupervised methods are not based on class statistics, they may produce features that are not optimal with respect to class separability by merging different classes or neglecting some. For example, in the case of PCA, some valuable information for class discrimination may be contained in eigenvectors of low eigenvalues, generally considered as noisy and removed from the feature set.

If some class objects are small relative to the size of the image scene, they will contribute only a small amount to the overall variance. These small objects will not be captured by PCA unless their spectra are nearly orthogonal to the background spectra (Ifarraguerri and Chang, 2000). In addition, Jones and Sibson (1987) observed that the information related to the segmentation of different objects and background types does not generally align itself with the main PC axes.

Consequently, if ground data are available, they should be used in conjunction with supervised feature extraction methods. PCA may be used as no more than a suboptimal comparison, since classes can sometimes be distributed in the direction of maximum scatter for remotely sensed data (Richards and Jia, 1999).

Hsieh and Landgrebe (1998) stressed the use of both first and second order statistics for feature extraction to adequately discriminate between the classes. The KL expansion uses only the within-class covariance matrix as compared to CA, which also employs the between-class scatter matrix. However, both KL and CA have no relationship to the probability of error. In addition, a reliable estimation of the class covariance matrix requires a large number of training samples for hyperspectral data sets (Hughes phenomenon), which are not always available.

DBFE directly estimates the decision boundaries from the training samples without any assumptions about class distributions. Benediktsson and Sveinsson (1997) showed empirically that DBFE produced features giving lower classification variance than both PCA and CA. However, the method is considered as slow and requires many training samples, which are normally limited in availability. In addition, it is considered as suboptimal for more than two classes.

PP offers a framework to incorporate any separability measure with direct relation to the probability of error using both first- and second-order statistics. Since PP estimates the statistics at lower dimensional space, less training samples are necessary for reliable estimates. Nonetheless, the implementation of the PI optimisation including an orthogonality constraint may prove very difficult (Siedlecki *et al.*, 1988).

The structure offered by MDS is convenient insofar as it allows the use of any suitable separability measure. However, it is only optimal for the entire image scene if all image pixels feed into the optimisation procedure, which may cause problems for large images with respect to computer storage.

As a consequence, the author recommends DBFE as supervised feature extraction method if sufficient training samples are available to the user. If this is not the case, PP should be considered as second choice in conjunction with a separability measure with a strong relation to the probability of misclassification employing first- and second-order statistics.

2.3.5 Feature Extraction for Band Selection

Some authors perform band selection not directly via feature selection, but indirectly via eigenanalysis-based feature extraction. In this case, not the transformed features, but the results of the eigen-decomposition, i.e. the eigenvectors and eigenvalues, are of interest to the user. Both unsupervised and supervised feature extraction criteria can be used for the eigenanalysis, such as variance, image signal-to-noise ratio or DA criteria.

Eigenvector loadings

One technique is founded on the ranking of the original bands according to the loadings of the eigenvector(s) with extreme eigenvalue(s) indicating their contribution (or importance) to the resulting feature. For example, if the most informative bands are to be chosen and information content is to be represented by data variance, an eigenanalysis may be performed on the data covariance matrix (PCA). The bands contributing most to the eigenvectors with the highest eigenvalues may be selected, where the contribution of each band is indicated by the loadings of this eigenvector.

However, the analysis of the eigenvectors loadings may be very subjective, since it is unclear how many eigenvectors and bands need to be considered (Mausel *et al.*, 1990). The following two approaches tackle this problem.

Tu *et al.* (1998) created a (N , N) loading factor matrix $\mathbf{F}$, the columns of which can be calculated by multiplying the normalised eigenvectors by the square root of their corresponding eigenvalues λ_l:

$$\mathbf{f}_l = \sqrt{\lambda_l}\,\frac{\mathbf{a}_l}{\|\mathbf{a}_l\|} \tag{2.37}$$

Where $\mathbf{f}_l$ Column of the loading factor matrix, $\mathbf{f}_l = [f_{1l} \quad f_{2l} \quad \cdots \quad f_{Nl}]^T$

$\|\cdot\|$ Euclidean vector norm

To prioritise the original bands, they defined a measure for each band j, the discriminant power ρ_j, which is the sum of the squared loading factors f_{jl}^2 over all N eigenvectors:

$$\rho_j = \sum_{l=1}^{N} f_{jl}^2 \tag{2.38}$$

Where ρ_j Discriminant power of band j

The larger value of ρ_j, the more significance is implied for the corresponding band. On the other hand, the discriminant power measure does not consider the spectral correlation between the prioritised bands (Chang *et al.*, 1999).

Csillag *et al.* (1993) proposed a top-down band selection method based on both PCA and DA. First, the dimension of the band set is reduced iteratively. In each step, a PCA is performed on the data, and the band that contributes most to the eigenvector with the smallest eigenvalue will be considered noisiest and will be discarded. This step is repeated until a stopping criterion is reached, such as the desired number of bands or the minimum classification accuracy (also referred to as modified stepwise PCA, MSPCA). Second, DA is performed on the reduced set of PC features using a predefined classification scheme. For each discriminant feature, the bands are assigned a rank (1 to number of bands) in order of importance to the feature. The overall grade of a band is calculated by multiplying its rank in all discriminant features by the eigenvalue of the corresponding feature and summing the latter products for all features. Both variance and class separation measures are integrated in this approach. However, bands are discarded only on the basis of their data variance.

Eigenvector axis crossings

In pattern recognition, the axis crossing information of the eigenvector loadings has been shown to hold a substantial proportion of the information contained in the transformed feature (Chen and Landgrebe, 1989).

Chen and Landgrebe (1989) produced hard limited eigenvectors by converting their values into their signed form, i.e. +1, -1 or 0 for positive, negative or zero values, respectively. That is, the transformed features result from summing, subtracting and omitting the original features. The authors suggested the transformation to be carried out on-board the sensor in order to reduce the number of features to be transmitted. However, the procedure does not reduce the actual number of bands to be collected.

The summation of adjacent bands may be interpreted as the grouping of narrow bands to broad bands, the edges of which are defined by the axis crossings. It is unclear whether the optimal feature set should be defined on the basis of the broad bands of the first eigenvector alone, or if the bands of other eigenvectors should be considered as well.

Henderson *et al.* (1989) averaged the first few eigenvectors, and defined the band edges as the features where a transition in sign occurred. He determined the number of eigenvectors to be averaged by choosing the mean-square error of the representation to be 0.4% in order to limit the error's effect on the image SNR. But, averaging the eigenvectors as suggested does not consider the relative importance of the eigenvectors and may alter the original locations of the axis crossing, resulting in suboptimal bands.

Other examples based on one of the methods presented in this section include Spanner *et al.* (1984, PCA eigenvector loadings), Wu and Linders (2000, MSPCA), Chang *et al.* (1999) and Pu and Gong (2000, discriminant power).

2.4 Intrinsic Dimensionality Estimation

In pattern recognition, the term intrinsic dimensionality (for representation) refers to the minimum number of features required to capture the structure within the data (Webb, 1999). The geometric interpretation is that the whole data lie on a topological hypersurface of dimension equal to the intrinsic dimensionality (Fukunaga, 1982). In remote sensing, the data structure is usually represented by the data variance, and the intrinsic dimensionality is understood as the minimum number of features needed to account for most of the image variance. Eigenanalysis-based unsupervised feature extraction methods (such as PCA or MNF) are typically employed to create uncorrelated features, whose eigenvalues represent their contribution to the overall variance (in the case of PCA).

In contrast, Lee and Landgrebe (1993) defined the intrinsic *discriminant* dimensions (for classification) "as the smallest dimensional subspace wherein the same classification accuracy can be obtained as could be obtained in the original space". In this case, the dimensionality is dependent on the number of classes M . The optimum features of the Bayes classifier are the M class posterior probabilities, and since they all sum to one, only $M-1$ dimensions are linearly independent (Fukunaga, 1982). Thus, the intrinsic dimensionality for classification cannot exceed $M-1$.

Supervised feature extraction methods using eigenanalysis (such as DBFE, CA or KL expansion) may be employed to derive the intrinsic discriminant dimensionality. Alternatively, the best subsets of uncorrelated features of dimensions 1 to $M-1$, derived via feature selection or feature

extraction, may be classified directly. Hypothesis tests may then be performed to see whether differences between the performances of two feature subsets (with one subset having one more feature than the other) are significant or not.

Many heuristic and statistical methods exist for estimating the intrinsic (discriminant) dimensionality of a data set, but only the most common and practical are presented in the following sections. The problem of the estimation of the intrinsic dimensionality corresponds to finding the number of significant features or eigenvalues.

2.4.1 Heuristic Methods

The heuristic methods presented here use criteria that are all based on eigenanalysis.

Eigenimage analysis

The eigenimages, i.e. the two-dimensional display of the projection of the image data points onto the eigenvectors, can be analysed with respect to the spatial information content they carry. This requires an extreme familiarity of the analyst with the ground to identify possible spatial patterns of the prevalent classes. The method is very subjective and depends entirely on the knowledge of the analyst.

Scree plot

The term scree plot refers to a graph where the values of consecutive eigenvalues are plotted against the rank order. It visualises the importance of each eigenvalue with respect to the others and its contribution to the variance. The latter might be emphasised with a graph printing the cumulative percentage of variance accounted for by the eigenvalues as a function of rank order.

Cattell and Vogelmann (1977) suggested dividing the plot into a horizontal line of small eigenvalues representing random variation, and relatively large eigenvalues that clearly leave that line. They estimated the intrinsic dimensionality as the number of the line-departing eigenvalues plus one. However, frequently the plot does not present any obvious break points, and the identification of the horizontal line becomes completely subjective.

Proportion of the total variance

Many researchers chose the intrinsic dimensionality as the number of the first eigenvalues that represent together a certain proportion of the total variance, usually around 90% to 95% (Biwas *et al.*, 1981). Nonetheless, the proportion threshold is set completely arbitrarily, and the suggestion of a

fixed value cannot be vital due to strong data dependence of the threshold value (Ferré, 1995). Henderson *et al.* (1989) considered the maximal noise level of a remote sensing system (around 2%) to guide their choice of threshold value.

Kaiser's criteria

The Kaiser-Guttman criterion (Guttman, 1954) estimates the intrinsic dimensionality as the number of eigenvalues that are greater than the average eigenvalue. Other methods compare the eigenvalues directly with an arbitrary value, e.g. $1/N$, where N is the number of original dimensions (Kaiser, 1960). In all cases, the value threshold eigenvalue is very subjective and cannot be justified (Ferré, 1995).

Broken-stick method

Horn (1965) suggested a modification to the scree plot for more objective dimensionality estimation by creating various matrices of random data with all the new uncorrelated features. The eigenvalues of these 'random' matrices are calculated, averaged and plotted on the original scree plot. The point where the two lines cross represents the threshold for significant eigenvalues.

Frontier (1976) provides a model for the distribution of eigenvalues, if the total data variance is dispersed randomly amongst all components: the broken-stick distribution. The eigenvalues distributed under the broken-stick model can be calculated as follows:

$$\lambda_l^B = \sum_{j=l}^{N} \frac{1}{jN} \tag{2.39}$$

Where λ_l^B Eigenvalue of the l th component under the broken-stick model

Eigenvalues are regarded as significant if their value is larger than that of the eigenvalues generated by the broken-stick model. However, the method does not account for cases where multiple intersection points exist between the scree plot and the broken-stick distribution. In addition, Ferré (1995) argues that it cannot be justified why the given data should be compared with a given fixed distribution.

2.4.2 Statistical Methods

Statistical methods require the knowledge of the distribution of the statistics under consideration to perform hypothesis or significance tests, such as whether eigenvalues are significantly different or whether the classification performance of a feature subsets differs significantly from that of another subset. They may help to determine significant features and thus the intrinsic data dimensionality.

Bartlett's test for sphericity

Bartlett's test for sphericity is based on the idea that some of the eigenvalues bearing random noise are equal, and the test aims to investigate whether each sequential eigenvalue, λ_l, differs significantly from the remaining eigenvalues. The test statistic for λ_l given in equation 2.40 is χ^2 distributed with $0.5(N-l-1)(N-l+2)$ degrees of freedom (Jackson, 1993).

$$\chi^2 = (n-l)(N-l)\ln\left(\sum_{j=l+1}^{N}\frac{\lambda_j}{N-l}\right) - (n-l)\sum_{j=l+1}^{N}\lambda_j \qquad (2.40)$$

Where χ^2 Test statistic to test hypothesis that λ_l differs significantly from the remaining eigenvalues

The intrinsic dimensionality is chosen as the value of l for which the eigenvalue λ_l was tested to be different from following eigenvalues for the last time.

Other statistics have been proposed, including tests whether the first or second eigenvalues are equal to the remaining set of eigenvalues (Bartlett's test of equality and Lawley's test, respectively)(Jackson, 1993). Nevertheless, these tests for sphericity are made up of nested and dependent hypothesis tests, which may result in an unknown overall level of significance (Ferré, 1995).

Bootstrap eigenvalue-eigenvector

To sample the distribution of the eigenvectors and eigenvalues, bootstrap techniques may be employed which generate sets of bootstrap observations and provide nonparametric approximations of the eigenvalues and eigenvectors (Webb, 1999). Jackson (1993) bootstrapped the PCA 100 times and calculated average eigenvalues and eigenvectors with 95% confidence limits. Where the confidence intervals between pairs of successive eigenvalues did not overlap, the eigenvalues were assumed to be significant. In addition, only eigenvectors with two or more coefficients differing significantly from zero were regarded as informative.

Classification accuracy

The intrinsic discriminant dimensionality of a data set can be estimated by testing the significance of the difference in classification performance between a feature subset and all the remaining subsets of higher dimensionality. Successively, subsets of different size are tested with increasing dimensionality, until the difference in performance between the current subset and the remaining subsets is no longer significant. Then, the number of features in the current subset represents the

intrinsic discriminant dimensionality. The subset performances may be quantified by a measure of classification accuracy (see section 2.2.2).

In practical terms, a statistical test of the null hypothesis 'Two feature subsets have the same classification accuracy estimate' will be performed on the assumption of independent random samples. Given the classification accuracy estimates of the two feature subsets, the sample variances of the accuracy measures will be needed to test the significance between the accuracy estimates via the z-statistic:

$$z = \frac{C_1 - C_2}{\sqrt{s^2(C_1) + s^2(C_2)}} \tag{2.41}$$

Where z z-statistic
C_i Accuracy estimate of feature subset i, $i \in \{1,2\}$
$s^2(C_i)$ Sample variance of C_i

The z-statistic is generally used to test the difference between the means of two random variables, which are estimated on the basis of known variances and a large number of independent random samples (Miller and Freund, 1965).

If one wishes to test the null hypotheses (equality of the population means) at the 5% (1%) significance level, the critical values are –1.96 (-2.58) and 1.96 (2.58). In other words, if the statistic gives a value between these critical values, the null hypothesis cannot be rejected and the observed difference between the sample means can well be attributed to chance. The critical value for a one-sided test is 1.645 (2.33).

The variances of the accuracy estimates from section 2.2.2 are presented hereafter. P_O (equation 2.11) follows a discrete binomial probability distribution and, according to the Central Limit Theorem, the discrete binomial distribution converges to the continuous normal distribution for the sample size N_E tending towards infinity (Bishop *et al.*, 1975 in Congalton, 1991). If the total population has both a finite mean and standard deviation for large N_E, then the large sample (or asymptotic) variance may be described as follows (Thomas and Allcock, 1984):

$$s_\infty^2(P_O) = \frac{1}{N_E} P_O(1 - P_O) \tag{2.42}$$

Where $s_\infty^2(P_O)$ Asymptotic sample variance of P_O

Thus, for hypothesis testing on P_O, one can use the normal distribution, given that a large number of test samples are available to build the error matrix (greater than 100, Ma and Redmond, 1995). If

the distribution of the P_O statistic is asymptotically normally distributed, so are the KHAT and TAU statistics. It can then be shown that

$$s_\infty^2(\tau_e) = \frac{P_O(1-P_O)}{N_E(1-P_R)^2} \tag{2.43}$$

$$s_\infty^2(\hat{\kappa}) \approx \frac{P_O(1-P_O)}{N_E(1-P_C)^2} \tag{2.44}$$

Where $s_\infty^2(\tau_e)$ Asymptotic sample variance of τ_e

$s_\infty^2(\hat{\kappa})$ Asymptotic sample variance of $\hat{\kappa}$

(Ma and Redmond, 1995; Cohen, 1960).

Shaban and Dikshit (2001) used the z-statistic to test the equality of the classification results of two feature subsets measured by the KHAT statistic. Dutra and Huber (1999) applied the τ_e statistic to quantify classification accuracy and tested the statistical significance of the difference between error matrices obtained from different classifiers.

McNemar's test

McNemars's test (e.g. Dietterich, 1998) involves a statistic to test the classification error of two feature subsets, A and B. The criterion is not based on the error matrix but on the number of samples that were classified incorrectly exclusively by one of the two feature subsets (see equation 2.45). A table including all misclassified samples from one subset needs to be stored and compared with the classification results of the other subset.

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \tag{2.45}$$

Where n_{01} Number of samples misclassified by A but not by B

n_{10} Number of samples misclassified by B but not by A

The statistic is approximately χ^2 distributed with one degree of freedom. The null hypothesis, that the feature subsets have the same accuracy, can be rejected on a significance level of 95% if $|\chi^2| > 3.841459$.

Separability measures

Instead of classification accuracy measures, class separability criteria from section 2.2.2 may also be employed to approximate classification performance, as long as their distribution is known. For

example, Penaloza and Welch (1996) used the separability measure d_p (equation 2.19), which follows the F-distribution, as an indicator for class discrimination. They tested the class discrimination ability of a reduced feature subset with respect to the one of the original feature set.

2.4.3 Choice of Methods

There is no 'ideal' solution to the problem of intrinsic dimensionality determination. The most important deficiencies have already been highlighted for each method. But how can the users be guided to pick out the method which suites their problem best?

First of all, the users have to decide whether they would like to determine the intrinsic data dimensionality or its supervised discriminant form.

If they decide themselves for the former, they have a range of heuristic and statistical eigenanalysis-based methods to choose from (all above methods except classification accuracy, classification error, or separability measures). Jackson (1993) and Ferré (1995) performed comparative studies of the most of the unsupervised methods presented above on simulated data of known dimensionality.

From the methods Jackson (1993) compared (scree plot, proportion of the total variance, Kaiser's criteria, bocken-stick, Bartlett's test, bootstrap eigenvalue-eigenvector), the broken-stick model and the bootstrapped eigenvalues-eigenvector gave the most consistent results.

Ferré (1995) advised methods, where failure of the method can be controlled (e.g. percentage of total variance, Kaiser's methods). For example, the scree plot after Cattell and Vogelmann (1977) relies completely on the scale of the display and the perception of a straight line, and may lead to uncontrolled failure.

Although statistical tests are less subjective than heuristic methods, caution has to be taken with respect to the overall level of significance of the procedure. They should be considered as approximate heuristic tests rather than rigorously correct statistical tests. If the distribution of some statistic is not known, it has to be resampled with the bootstrap technique or similar methods (Monte Carlo methods, or infinitesimal Jack-Knife approaches), which may require considerable computational time and be therefore limited in its practical use.

If the users intend to approximate the instrinsic discriminant dimensionality, they can choose from all of the methods presented above. The ones based on classification accuracy measures may be interpreted as the most direct and objective ones. All statistics depending on the error matrix require the matrix to be constructed from a large number of samples, so that the statistic's asymptotic variance can be estimated with confidence. The leave-one-out cross-validation method is

recommended for the error matrix estimation, since it makes most use of the available samples and gives an almost unbiased estimate. In addition, for intrinsic dimensionality estimation, the user is not interested in the best approximation of absolute classification accuracy (i.e. the generalisation error), but in the relative loss of classification accuracy when applying dimensionality reduction methods.

Linearly extracted features used for intrinsic (discriminant) dimensionality estimation are prone to both noise and surface convolutions, which tend to enlarge the dimensionality estimate (Fukunaga, 1982). It is recommended to remove any noise from the data before the dimensionality is estimated. Still, the effect of surface convolutions may only be tackled with non-linear transformations. For example, a curved line in feature space has only one intrinsic dimension, which could be revealed via non-linear transformations. In contrast, linear feature extraction will always produce more than one significant feature. For intrinsic dimensionality determination using nonlinear mapping algorithms, see Fukunaga (1972) or Fukunaga (1982).

2.5 Summary

Chapter 2 introduced common feature selection and extraction techniques, and highlighted their advantages and disadvantages. Dimensionality reduction is not an easy task and relies heavily on the type of classifier to be used and the available data set. The user has to trade-off between computational efficiency and optimality of the result when choosing the dimensionality reduction algorithm. Therefore, a universally best feature subset does not exist.

Feature extraction methods should be preferred to feature selection techniques when the user is interested in a set of most informative and uncorrelated features, as the latter can be found only by transfoms of the original features (Benediktsson and Sveinsson, 1997). However, the transformed features may not have a physical significance and may not be interpreted easily. Moreover, they rely on all sensor output bands, that is, the complexity of the data acquisition system will not be reduced. That is, with respect to band selection, feature selection techniques are more applicable than feature extraction methods.

This chapter also presented techniques to estimate the intrinsic dimensionality of a data set on the basis of feature selection or extraction methodologies for both the supervised and unsupervised case. Again, it is difficult to give precise rules for the selection of a particular method or its reliability for a given data set and application. Practically, lower and upper value limits for the intrinsic dimensionality may be calculated by deriving multiple estimates from different techniques.

In the next chapter, some of the dimensionality reduction methods presented in this chapter were applied to real hyperspectral data sets acquired with the HyMAP and CASI-2 sensors, mainly with the intention to derive an optimal sensor band set for the given classification schemes.

3 Pilot Study

3.1 Introduction

This pilot study aimed to familiarise the author with the study areas and data sets available, to define classes and to explore some of the dimensionality reduction methods reviewed in chapter 2, for classification as application task. The objective was to see whether any major differences result from different dimensionality reduction and intrinsic dimensionality estimation techniques for the given class and data sets.

Section 3.2 introduces the two study areas used for subsequent analysis: the Mid Severn Estuary and the New Forest, Hampshire (both in the UK, figure 3.1). Both the intertidal and the heathland area represent semi-natural environments that exhibit only subtle spectral variation between some of their classes and are spatially complex. One of the aims of chapter 4 was to investigate whether the 'established' vegetation band sets (see chapter 4) would also be applicable to map these types of environments.

In this chapter, the land category of each study area is presented and specific details are given about the data sets acquired: remotely sensed imagery (with CASI and HyMAP) and *in situ* reference data. All data pre-processing is described. The general methodology for this pilot study is outlined in section 3.3, including guidelines for class definition and class training (section 3.3.1), and a plan of dimensionality reduction experiments with some of the methods reviewed in chapter 2 (section 3.3.2). The selected land cover classes are portrayed in section 3.4 for both study areas, and the results of the dimensionality reduction experiments are presented and discussed. Finally, some general conclusions are drawn in a closing summary.

3.2 Study Areas and Data

3.2.1 Mid Severn Estuary, UK

The study area consists of a 5.5 km long coastline north of the Mid Severn Estuary, UK, between Summerleaze and the Severn M4 motorway crossing (see figure 3.1). It is dominated by saltmarsh vegetation and includes the sites Magor Pill, Collister Pill and West Pill.

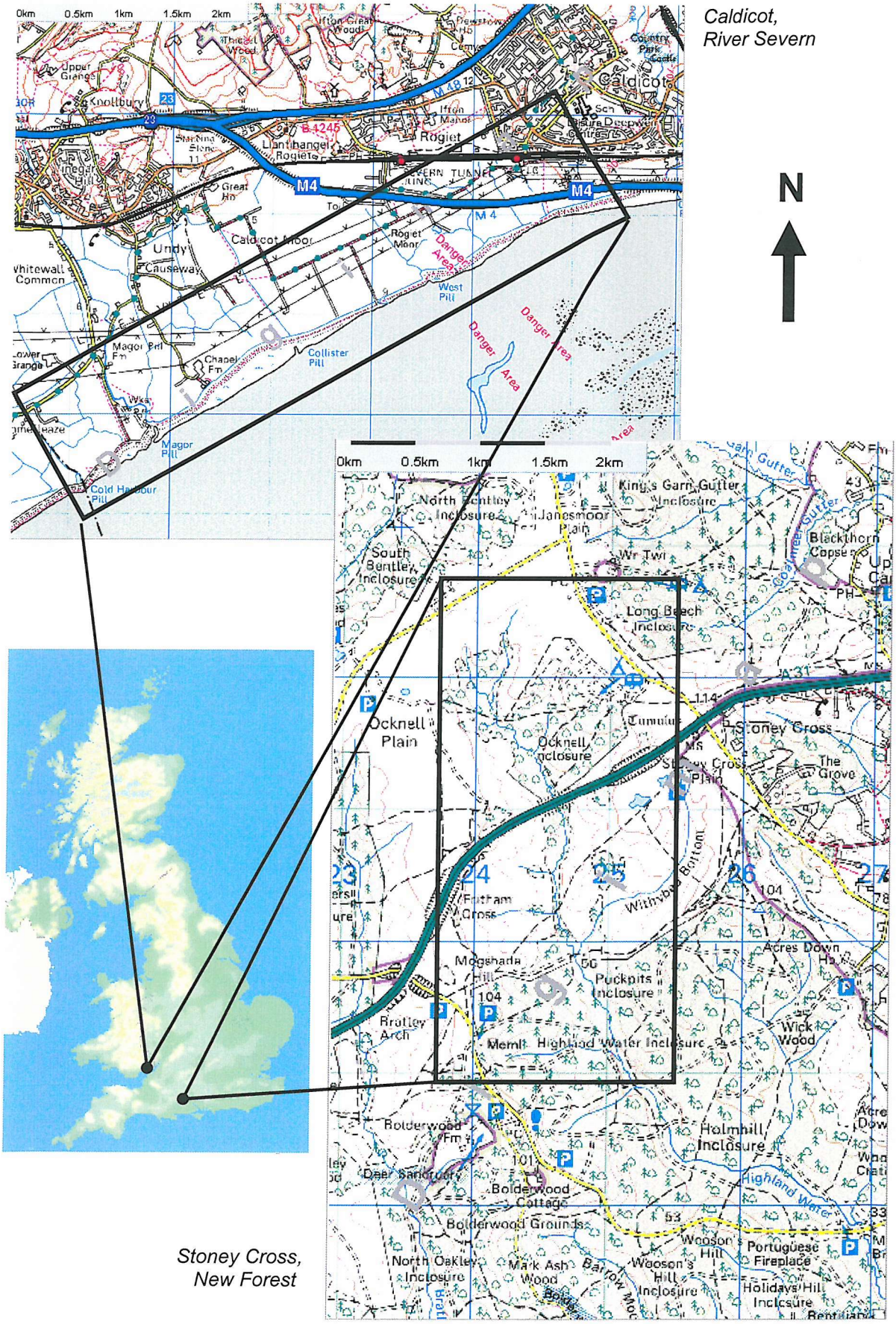


Figure 3.1: Ordnance Survey maps of the Mid Severn Estuary (above) and the New Forest in Hampshire (all maps © Crown copyright).

Land category

Coastal saltmarshes are marshland areas vegetated by herbs, grasses or low shrubs and bordering saline water bodies by which they are regularly inundated due to tidal action. Saltmarshes are highly productive and sensitive ecosystems acting as a buffer zone between the sea and the upland. Usually they are protected areas of conservation as they

- support a great diversity of life forms,
- provide a nursery habitat for fish and invertebrates,
- offer major breeding, nesting and migration staging areas for waterfowl and shorebirds,
- protect shorelines from erosion and flooding, as the water is slowed as it passes through shallow, plant-filled areas, facilitating the sediments to deposit and stabilise and water to be stored, and
- act as filters and help to neutralise and detoxify substances in the water.

The Mid Severn Estuary is an area of major nature conservation interest and contains three sites of special scientific interest (SSSI), a special protection area (SPA), a Ramsar site (a wetland of international importance), and a possible special area of conservation (SAC) under the EU Wildlife and Habitats Directive (Dargie, 1999).

The monitoring, management and protection of these fragile habitats may be realised via detailed and periodically updated maps that delineate the marshland from the upland and identify major vegetation species or communities and their distribution within the saltmarsh. The maps could be most rapidly and effectively produced using remotely sensed imagery, especially for a wide area of interest. Ground survey does not offer a practical solution because of the difficulties in gaining access to and manoeuvring in this potentially hazardous environment and locating oneself in the field.

Vegetation Classes

Table 3.1 presents the scientific and common names of some frequent saltmarsh plants found in the intertidal zone around the British Isles. These vascular plants are terrestrial in origin, but can withstand periodical water logging in a saline environment (also called 'halophytes').

Saltmarsh surfaces are very complex and heterogeneous. Individual plant species rarely form homogeneous communities, and usually mix with other plant species at a scale of centimetres to metres. As a consequence, airborne imagery with a spatial resolution of about 3 m may not be able to resolve individual plant species, but may distinguish between different plant communities extending between tens to hundreds of metres.

Table 3.1: Scientific and common names of some ordinary saltmarsh plants within the UK.

Scientific Name	Common Name
Ameria maritima	Thrift, sea pink
Aster tripolium	Sea aster
Elymus pycnanthus	Sea couch-grass
Festuca rubra	Red fescue
Glaux maritima	Sea milkwort, black saltwort
Halimione portulacoides	Sea purslane
Juncus gerardi	Mud rush
Leontodon autumnalis	Autumn Hawkbit
Limonium vulgare	Sea lavender
Phragmites australis	Reed
Plantago maritima	Sea plantain
Puccinellia maritima	Sea meadow grass
Puccinellia distans	Alkali-grass
Salicornia sp.	Glasswort
Spartina anglica	Cord-grass
Spergularia marina	Salt sandspurry
Sueda maritima	Sea blite
Triglochin maritima	Sea arrow grass
Zostera spp.	Eel grass

Table 3.2: Height levels averaged over several years (after Long and Mason, 1983).

Height level	Description
MHWS	Mean levels of High Water of Spring tides (*)
MHW	Mean levels of High Water of all tides
MHWN	Mean levels of High Water of Neap tides (*)
MLW	Mean levels of Low Water of all tides

(*) Spring tides are the largest tides and occur when Sun, Moon and Earth are aligned 1.5 days after new and full moons. Neap tides represent the minimum high water level between two consecutive spring tides.

According to Long and Mason (1983), three broad salt marsh zones, arranged in belts parallel to the shoreline, may be recognised according to their relative periods of seawater immersion and exposure to air:

- **Low marsh**: also referred to as 'pioneer zone', this zone falls typically within the MHWN and MHW height levels (see table 3.2). It consists of significant bare areas with few species, mainly cord-grasses (*Spartina anglica* in the study area) and algae. Its lower limit can be defined unambiguously as the seaward margin of vascular plant communities, excluding those composed of sea-grasses or other permanently submerged species.
- **Mid marsh**: this zone occurs between levels MHW and MHWS (see table 3.2) and is dominated by saltmarsh grasses such as *Puccinellia maritima* (as in this study area) or mixed communities including *Ameria maritima* and *Plantago maritima*.
- **High marsh**: this zone may be defined as the area above the MHWS level (see table 3.2) and resembles terrestrial vegetation in being predominantly composed of higher plants (angiosperms) which are also found in inland vegetation. Generally a great diversity in species

composition can be found in northern Europe. In the study area, high marsh is dominated by the *Festuca rubra* species.

In addition to this spatial variability, saltmarsh vegetation is subjected to a number of environmental and inherent factors, which may cause variations in the canopy's spectral signature:

- Apart from controlling soil salinity and the degree of water logging, tides carry sediments into marshes making tidal waters turbid. As a result, remote sensing of saltmarshes should be carried out at low tides, in order to be able to 'see' low and mid marsh vegetation in the optical spectrum. As the tide recedes, sediment may be deposited on the foliage of the plants and may remain there for longer in the absence of rain (Budd and Milton, 1982). In addition, plants may be covered with a thin salt water film much of the time. Strong tidal water flows may also have a physical impact on the canopy geometry, for example, grasses may be flattened.
- Depending on its speed and direction, wind has been found to cause spectral signal variation for certain plants. For example for cereal crops, radiance fluctuated 10-20% more in the 450-650 nm than in the 650-750 nm range (Rao *et al.*, 1979). Wind is particularly present on coastal saltmarsh areas, and may flatten grasses especially in the high marsh region.
- Grazing of domestic stock (such as cattle, sheep, geese, ducks) and wild animals (e.g. rabbits or hares) on high saltmarshes is a major factor controlling the diversity in flora and vegetation between sites (Adam, 1993).
- Canopy reflectance is also dependent on the plant phenology, i.e. the timing of flowering and seed release. The latter is different from plant to plant and there may be some time throughout the year when the spectral discrimination between saltmarsh species or communities is best. Hardisky *et al.* (1983) and Drake (1976) examined seasonal changes in the reflectance of some salt marsh communities and found that spectral discrimination of canopy characteristics was effective during the majority of the growing season, and that reflectance could be used for the detection of seasonal changes in community green biomass during this period.
- Finally, the viewing and illumination geometry between Sun, target and sensor may affect the spectral signal of the saltmarsh vegetation canopy. Bartlett *et al.* (1986) showed that changing the zenith angle of observation produces significant variability in measured red and infrared canopy reflectance for salt-marsh cord grass. Directional sun angle effects were found to cause significant changes in nadir-sensed reflectance from leafless canopies of salt marsh vegetation, while they have little effect on reflectance from broadleaf canopies (Gross *et al.*,1988).

In situ data

Dargie (1999) conducted a survey of vegetation of saltmarsh habitats in the Severn Estuary between July and mid-October 1998, using National Vegetation Classification (NVC) methods and description. The study area is part of this survey and figures 3.2 to 3.4 show the corresponding survey maps in detail. The NVC code for the saltmarsh communities is presented in table 3.3.

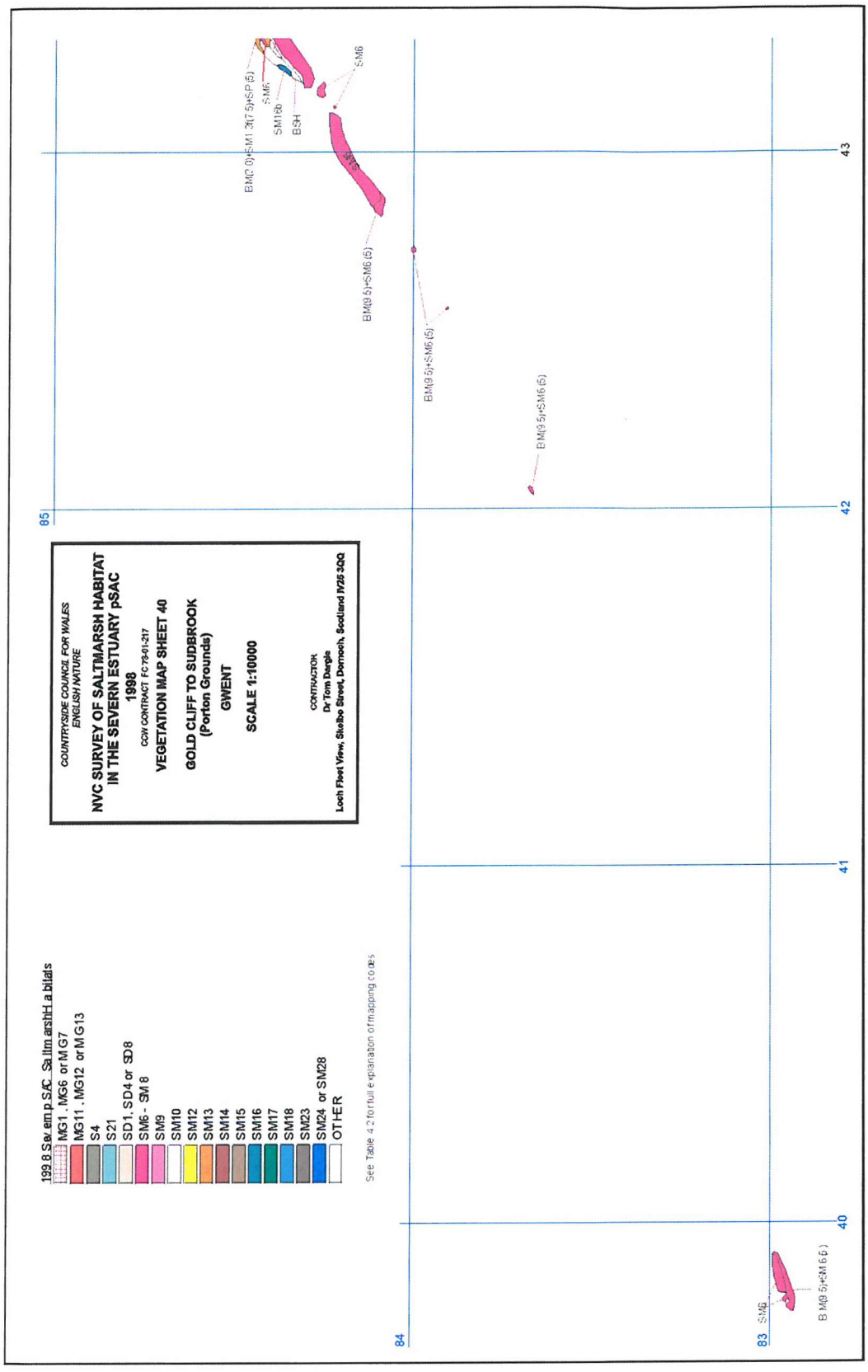


Figure 3.2: Dargie's vegetation map (Dargie, 1999) for the west part of the Mid Severn Estuary study area. The NVC codes for the saltmarsh communities are described in table 3.3.

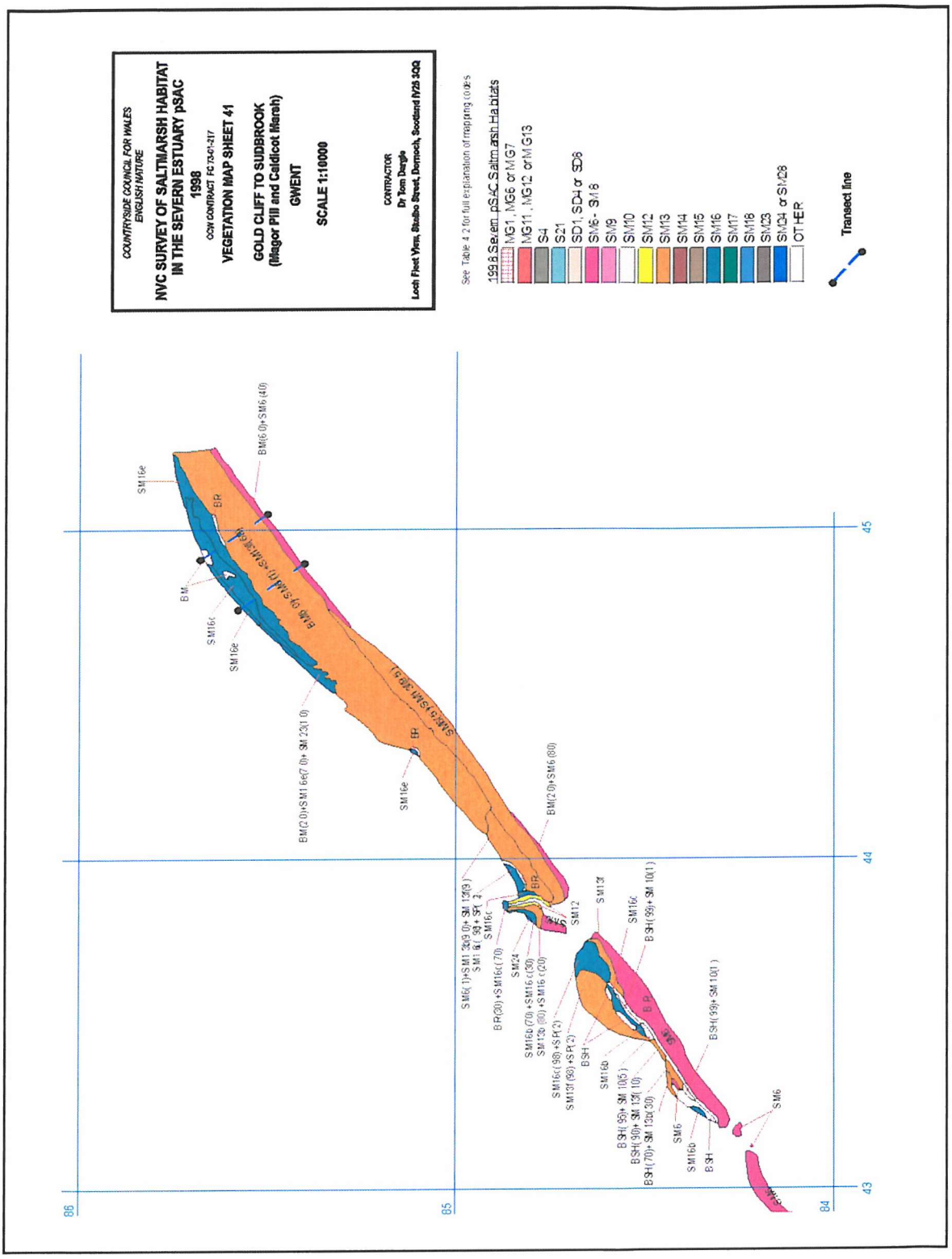


Figure 3.3: Dargie's vegetation map (Dargie, 1999) for the middle part of the Mid Severn Estuary study area. The NVC codes for the saltmarsh communities are described in table 3.3.

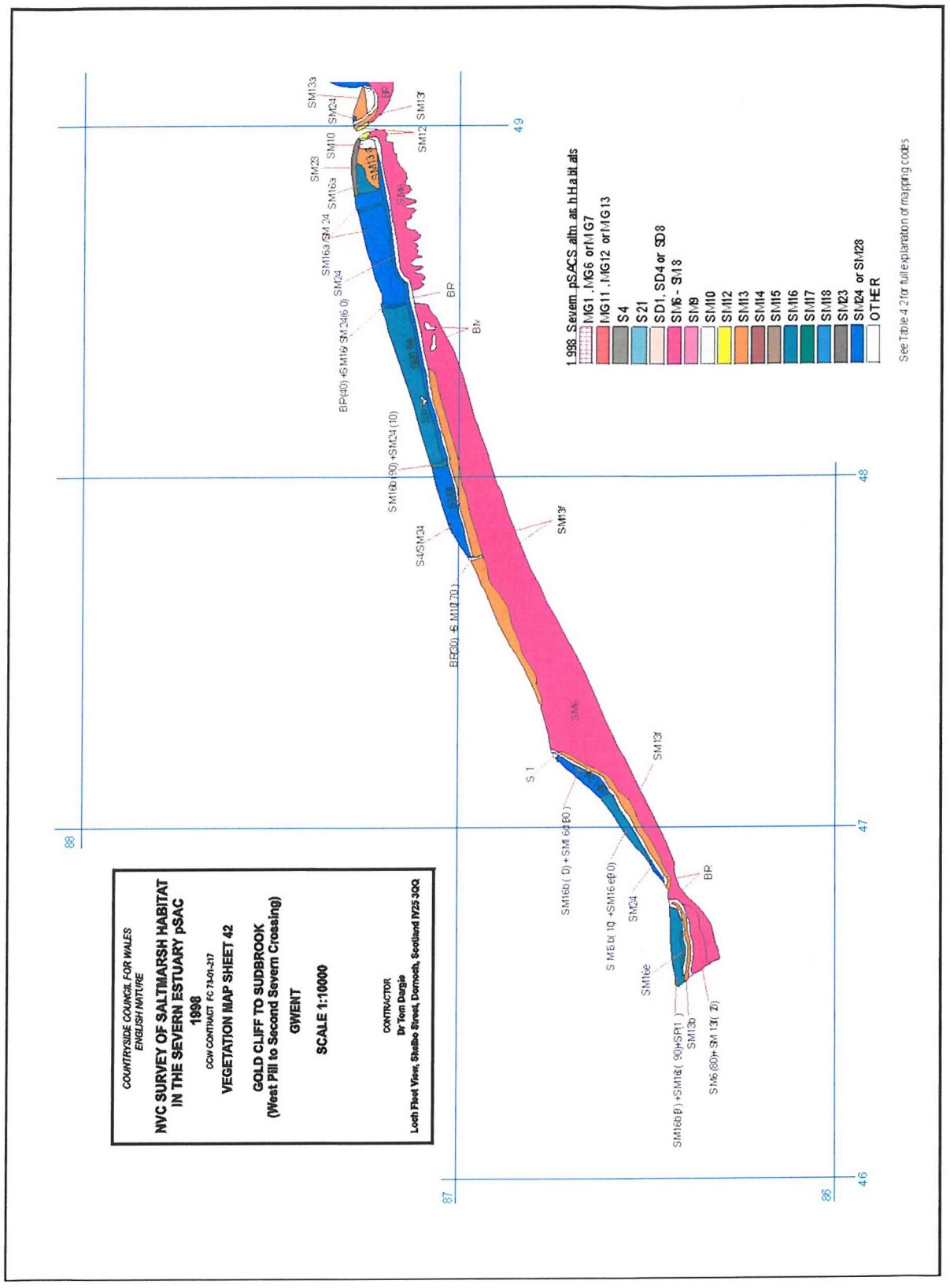


Figure 3.4: Dargie's vegetation map (Dargie, 1999) for the east part of the Mid Severn Estuary study area. The NVC codes for the saltmarsh communities are described in table 3.3.

Table 3.3: Description of the National Vegetation Classification (NVC) saltmarsh categories.

NVC code	Saltmarsh	Type	Comment
SM6	*Spartina anglica* saltmarsh	Pioneer Marsh	only lightly grazed by stock
SM10	Transitional low saltmarsh with *Puccinellia maritima*, annual *Salicornia* and *Suaeda maritime*		
SM12	*Aster tripolium* (rayed)	Middle Marsh	
SM13a	*Puccinellia maritima* saltmarsh, *P. maritima* sub-community		
SM13b	*Puccinellia maritima* saltmarsh, *Glaux maritima* sub-community		at well-drained locations, often in grazed marshes
SM13f	*Puccinellia maritima* saltmarsh, *Spartina anglica* sub-community		probably former SM6 converted to mid-marsh by stock grazing
SM16	*Festuca rubra* saltmarsh	High Marsh	
SM16a	*Festuca rubra* saltmarsh, *Puccinellia maritima* sub-community		lowest levels for SM16 types, above or in mosaics with SM13b vegetation
SM16b	*Festuca rubra* saltmarsh, *Juncus gerardi* sub-community		*J. gerardi* abundant in poorly drained patches of ground
SM16c	*Festuca rubra* saltmarsh, *Glaux maritima* sub-community		
SM16e	*Festuca rubra* saltmarsh, *Leontodon autumnalis* sub-community		
SM23	*Spergularia marina - Puccinellia distans* saltmarsh	High Marsh (disturbed)	Disturbances, e.g. by stock trampling or re-alignment of seawall
SM24	*Elymus pycnanthus* saltmarsh	Strandline	characteristic of the drift line, esp. on ungrazed marshes, 100% cover
S4	*Phragmites australis*, *P. australis* sub-community	Transitions to swamp	Freshwater seepage at the rear of saltmarsh, species-poor
BM	Bare mud	Non-NVC categories	Pioneer saltmarsh or product of poaching
BR	Exposed hard seawall (stone gabion, boulder riprap, concrete)		mostly sterile habitats with little plant growth
BSH	Bare shingle		Mostly a sea defence measure to retard erosion
SP	Saltpan(s)		uncommon upon saltmarsh in the Severn Estuary
ST	Strandline debris		mainly wood and plastic

Airborne hyperspectral imagery

CASI image data were collected in October 1997 by the Environment Agency (EA) over the Mid Severn Estuary study area. Details of the data acquisition are presented in table 3.4. The EA corrected the data for geometric distortions associated with the attitude of the aircraft platform (roll and pitch). A false colour image is displayed in figure 3.5. The 72 bands with their associated centre wavelengths and bandwidths are listed in table A.1 in the appendix.

According to the EA (K. Brown, 2002, personal communication) the data had been calibrated to spectral radiance to minimise detector responsivity variations. Nonetheless, some bands in the blue (bands 1 to 9) and NIR (bands 70 to 72) suffer from severe along-track striping, which is typical for pushbroom scanners due to the different response sensitivities of each detector. Besides, bands 1 to 13 and 70 to 72 show some across-track striping. Destriping algorithms exist on most image processing systems (e.g. ENVI©) or can be found in papers such as Hutsinpiller (1988). However, destriping is a cosmetic procedure which may have an unknown effect on the data. As a result, destriping algorithms were not applied to the data and bands 1 to 9 and 70 to 72 were excluded from further processing. The spectral calibration accuracy was evaluated by detecting the oxygen absorption feature at 762 nm, which is present in all radiance spectra. The feature fell onto band 48 (761.4 - 770.2 nm) suggesting a satisfactory wavelength calibration.

The southern image edge was affected by Sun glint during data acquisition. The mudflats acted as specular reflectors due to their flatness and the presence of a water film. The middle part of the image swath containing saltmarsh vegetation was not influenced by it.

Ideally, *in situ* data are collected at the time of image acquisition to allow for accurate class definition and mapping. In this case, the data were acquired about a year after image recording. Any changes in class composition may have been mainly introduced by grazing of stock and wild animals, and river diggings. However, a comparison between the imagery and the *in situ* map does not suggest any major changes in the spatial distribution and identity of the classes. Severe differences between image and map would lead to a potentially faulty class definition, classification result and feature selection.

Table 3.4: Characteristics of the CASI image acquisition over Mid Severn Estuary, UK.

Acquisition parameter	CASI imagery
Date of acquisition	28 October 1997
Time of acquisition (hrs GMT)	12:03 - 12:07 p.m.
Type of aircraft	twin engine Cessna 402
Altitude (m) (above ground)	1220
Ground speed (knots)	105
Number of scan lines	1252
Direction of flightline	South-West to North-East (60°)
Sensor mode	Enhanced Spectral Mode
Spatial resolution (m)	4
Number of spectral bands	72
Spectral resolution (nm)	8.4 - 8.8
Spectral range (nm)	408 – 954
Data format (bit)	12
View angle (°)	Nadir
Field of view (°)	77.3
Swath width (km)	1.54 (405 out of 512 pixels)
Solar elevation angle (°)	25.19
Solar azimuth angle (°)	-2.17
Status of the atmosphere	Clear blue sky

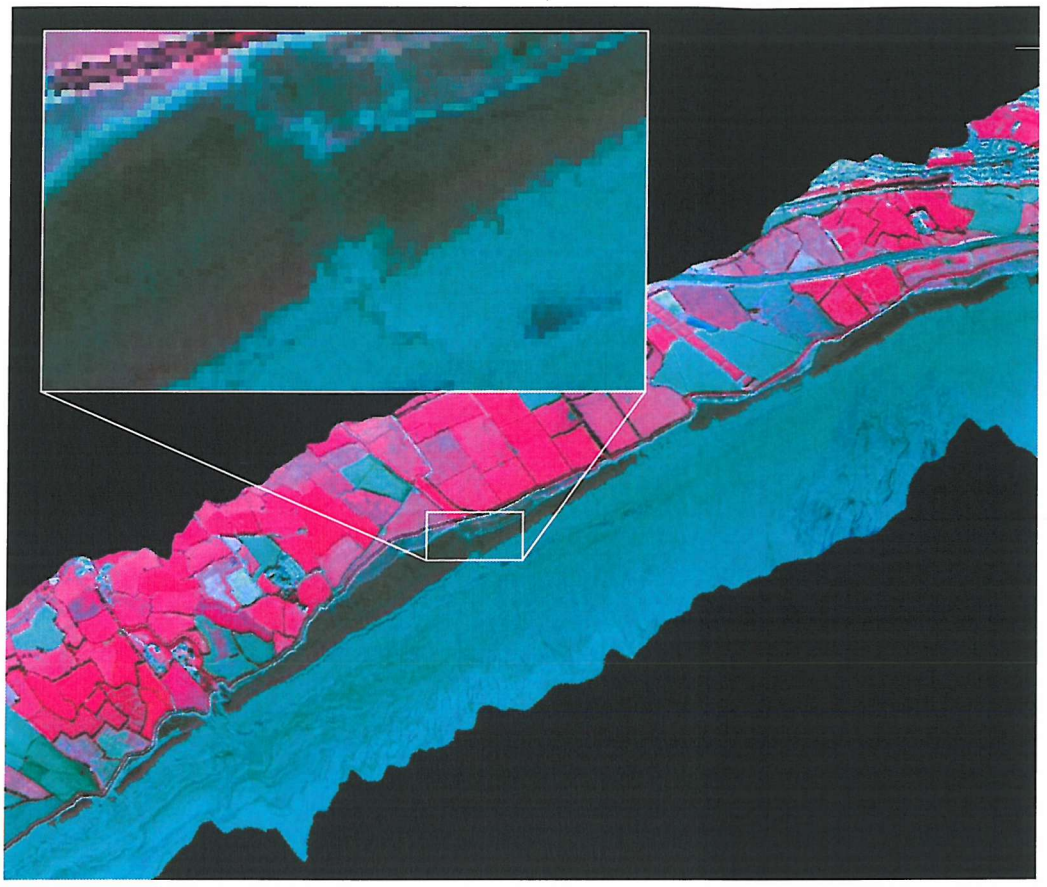


Figure 3.5: Geometrically corrected false colour composite of CASI data using bands 53 (R), 33 (G) and 20 (B) overlaid with a zoom area. © UK Environment Agency, 1997.

3.2.2 New Forest, Hampshire, UK

The study area is situated in the New Forest, Hampshire, UK, near the village of Stoney Cross, northwest of the town of Lyndhurst (see figure 3.1). It is dominated by enclosed forest areas and semi-natural heathland vegetation and includes the northern section of the Highland Water catchment.

Land category

The New Forest in Hampshire, UK, comprises the largest area of semi-natural vegetation of its kind in lowland Britain and consists of a unique combination of habitats once widespread in Western Europe: heathland, mire and pasture woodland (Westerhoff, 1992). The ecosystem comprises almost 20,000 ha of unenclosed forest, of which 15,000 ha represent lowland heath, mires, and acidic grassland, creating a multitude of complex transitional plant communities (Westerhoff, 1992). In addition to the open forest, the New Forest contains enclosed areas of conifer plantations and pasture woodland.

The New Forest is an area of international importance to nature conservation and biological science, protected by several European and national directives (JNCC, 2002; Westerhoff, 1992):

- National park status (1994),
- SSSI under the Wildlife and Countryside Act (1981),
- SAC designated under the EC Habitats and Species Directive for the protection of habitats and (non-bird) species,
- All wetlands protected by the International Convention on Wetlands of International Importance especially as Waterfowl Habitat (the Ramsar Convention),
- SPA due to the number of rare birds present.

Vegetation mapping and monitoring for this important and extensive area of Southern England is best achieved by remote sensing. Two of the reasons to monitor the New Forest at present are:

- the potential threat to vegetation communities (such as the valley mire and wet heath) from climate change due to increasing water tables, and
- the reintroduction of heathland on some plantation woodland areas in the next 20 years under the current management plan.

Vegetation classes

Table 3.5 presents the scientific and common names of some frequent plant species found in the study area.

Semi-natural land tends to be very heterogeneous, and mixing between plant species occurs at a scale of centimetres to metres. Airborne imagery with a spatial resolution of about 3 m may not resolve individual plant species. Westerhoff (1992) defined the following categories to classify open forest vegetation types of the New Forest from aerial photographs:

- **Dry heath (DH):** Dry heath is found mainly on permeable humus-iron podsols deposited on well-drained slopes and high ground. *Calluna vulgaris* is dominant, while *Erica cinerea* is consistently present. *Molinea caerulea* may be locally present, but *Erica tetralix* is always absent. Some *Ulex europaeus* and *Pteridium aquilinum* may be present.
- **Humid heath (HH):** Humid heath is the intermediate stage between dry heath and wet heath and is characteristic of slowly permeable and seasonally waterlogged ferric or humic-gley podsols in gravels, loams and clays. It is the most widespread heathland plant community in the New Forest. *Calluna vulgaris* and *Erica tetralix* are present and either may be dominant. *Molinea caerulea* may be abundant. Where *Calluna vulgaris, Erica tetralix* and *Molinea caerulea* are co-dominant the classic 'triple heath' is apparent. *Erica cinerea, Ulex europaeus* and *Pteridium aquilinum* may be locally present.

Table 3.5: Scientific and common names of common plant species of the open plant communities within the New Forest, UK (Westerhoff, 1992).

Scientific Name	Common Name
Agrostis sp.	Bent
Alnus glutinosa	Alder
Bellis perennis	Daisy
Betula sp.	Birch
Calluna vulgaris	Heather
Crataegus sp.	Hawthorn
Drosera sp.	Sundew
Erica cinerea	Bell heather
Erica tetralix	Cross-leaved heath
Eriophorium sp.	Cotton grass
Fagus sylvatica	Beech
Festuca rubra	Red fescue
Galium saxatile	Heath bedstraw
Ilex aquifolium	Holly
Juncus squarrosus	Heath rush
Lolium perenne	Perennial rye-grass
Molinea caerulea	Purple moor grass
Myrica gale	Bog myrtle
Pinus sylvestris	Scot's pine
Potentilla erecta	Tormentil
Pteridium aquilinum	Bracken
Quercus sp.	Oak
Salix sp.	Willow
Sphagnum sp.	Sphagnum moss
Tricophorum cespitosum	Deer grass
Trifolium repens	White clover
Ulex europaeus	Gorse

- **Wet heath (WH):** Wet Heath is dominated by an overstorey of *Erica tetralix* and *Molinea caerulea* tussocks most usually with a *Sphagnum* moss understorey. *Juncus squarrosus, Tricophorum cespitosum* and *Myrica gale* are also common.
- **Valley mire (VM):** Valley mire can be found in valleys with impeded drainage and downslope of hillside seepage steps at the junction between permeable and impermeable soils. *Sphagnum sp.* is dominant, and *Calluna vulgaris* and *Tricophorum cespitosum* are always absent. *Molinea caerulea* is often abundant together with some rare species including *Drosera sp.* and *Eriophorium sp.*
- **Carr (C):** Carr are residual alluvial forests, grading into bog woodland (JNCC, 2002). Carr habitat is characterised by *Alnus glutinosa* and *Salix sp.* This woodland often has a rich understorey and supports a number of rare epiphytes.
- **Broadleaved Woodland and Scrub (BWS):** This semi-natural woodland community comprises *Quercus sp., Fagus sylvatica, Ilex aquifolium, Betula sp.* and *Crataegus sp.*
- **Bracken (B):** *Pteridium aquilinum* is dominant, usually with an ericaceous understorey.
- **Bracken/Gorse (BG)**: Mixed class comprising *Pteridium aquilinum* and *Ulex europaeus*

- **Gorse (G):** *Ulex europaeus* is dominant, often forming a closed canopy. *Pteridium aquilinum* may be present.
- **Acid grassland (AG):** Grassland dominated by *Agrostis sp., Molinea caerulea, Potentilla erecta and Galium saxatile. Pteridium aquilinum* and *Ulex europaeus* may be present.
- **Scot's Pine (SP):** Area where *Pinus sylvestris* forms a closed canopy.
- **Lawn (L):** Neutral grassland dominated by *Agrostis sp., Festuca rubra, Lolium perenne, Trifolium repens* and many rosette-forming herbs.
- **Reseeded grassland (RSG):** Grasses such as *Agrostis sp., Bellis perennis* and a rich assemblage of prostrate herbs.

In situ data

Between 1986 and 1988 an extensive vegetation survey was carried out on behalf of English Nature by Clarke and Westerhoff across the unenclosed regions of the New Forest (Westerhoff, 1992). Vegetation class boundaries were mapped according to the defined categories above onto 1:10,560 scale Ordnance Survey maps of the area. At a later date, the maps were distributed in digital format. Figure 3.6 presents the survey data for the New Forest study area.

Airborne hyperspectral imagery

During the BNSC/NERC SHAC-2000 campaign, hyperspectral high-resolution HyMap image data were acquired over the New Forest study area on 19 June 2000. Details of the data acquisition are presented in table 3.6.

No simultaneous ground measurements were taken, impeding atmospheric image correction. Atmospheric water vapour is assumed to be uniformly distributed across the site, since the area is small and presents no extremes in topography or large water bodies.

The data were calibrated from radiance to reflectance by the survey operators (HyVISTA Corp. Pty. Ltd.) with HyCorr, a modified version of the Atmospheric Removal (ATREM) algorithm (Gao *et al.*, 1993) and the Empirical Flat Field Optimal Reflectance Transformation (EFFORT) program (personal communication with Anthony Dennis, Infoterra, 2001). ATREM uses radiative transfer modelling to calculate apparent reflectance, while EFFORT improves the resulting accuracy by using band statistics to remove calibration and atmospheric correction errors.

Geometric correction, based on aircraft ephemeris data, was performed with the ENVI© plug-in provided with the data. As the resulting image was still displaced from British national grid, it was registered to the digital ground data map (see above) via a simple linear transformation. A georectified false colour composite image can be seen in figure 3.7.

Figure 3.6: Part of Clarke and Westerhoff's vegetation map (Westerhoff, 1992) for the New Forest study area. The categories are explained in section 3.2.2.

Visual inspection of all 126 bands revealed 9 noisy bands (1, 2, 63, 64, 65, 95, 124, 125, 126), which have been excluded from further processing. The enclosed woodland area present on the imagery was masked out by manual on-screen digitising.

The image data were acquired fourteen years after *in situ* data collection. Within this period, changes in class distributions and compositions may have been caused by the grazing of wild animals and any human induced modifications, such as tree plantations or the reintroduction of heathland. However, the *in situ* map and the hyperspectral imagery do not exhibit any major differences in the spatial patterns and the identity of the given classes. This was confirmed by K. Anderson (2000, personal communication), who conducted extensive field measurements in the study area at the time of image acquisition. The *in situ* map was therefore used in conjunction with the delineation of classes within the hyperspectral image.

Table 3.6: Characteristics of the HyMAP image acquisition over New Forest, UK.

Acquisition parameter	HyMAP imagery
Date of acquisition	19 June 2000
Time of acquisition (hrs UTC)	11:56 a.m.
Type of aircraft	Dornier 228
Altitude (m) (above ground)	1500
Ground Speed (knots)	122
Number of scan lines	2038
Direction of flightline	North to South (180°)
Spatial resolution (m)	3
Number of spectral bands	126
Spectral resolution (nm)	11 – 22
Spectral range (nm)	430 – 2500
Data format (bit)	8
View angle (°)	Nadir
Field of view (°)	61.3 (512 pixels)
Swath width (km)	1.8
Solar elevation angle (°)	62.44
Solar azimuth angle (°)	5.95
Status of the atmosphere	Low level cloud

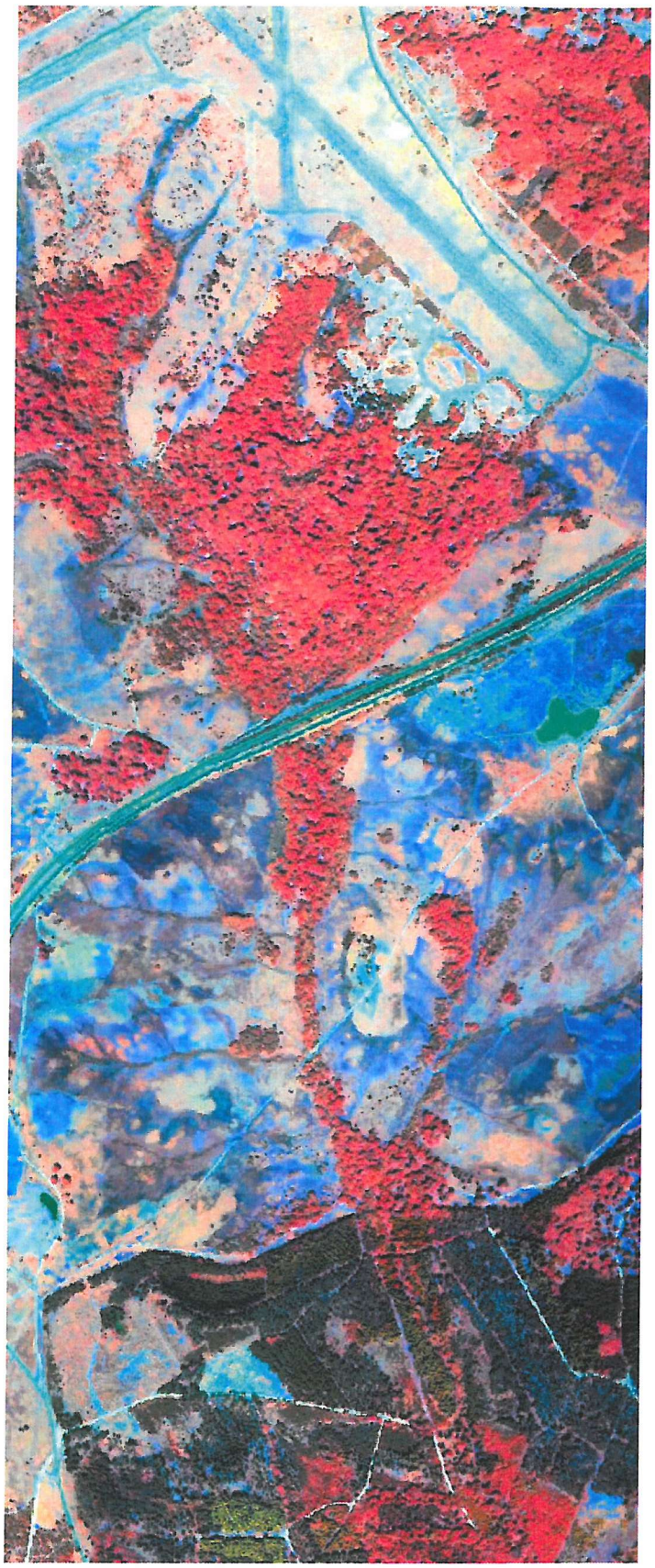


Figure 3.7: Geometrically corrected false colour composite of HyMAP data using bands 107 (R), 22 (G) and 7 (B). © UK Natural Environment Research Council, 2000.

3.3 Dimensionality Reduction for Classification

Section 3.3.1 describes the final data application, the maximum likelihood classification procedure. In addition, guidelines for class definition and training are outlined. Subsequently, some of dimensionality reduction methods reviewed in chapter 2 and applied in this study are presented in section 3.3.2.

3.3.1 Classification, Class Definition and Training

Classification

Generally, the aim of the data analyst is to transform raw image data into valuable information, such as thematic maps of object classes of interest. For the latter example, the most common image transformation is classification, the process of assigning class membership labels to each picture element (pixel). This procedure is based on a classifier, which relies on the class discriminative information content of the image data. In this project we limit ourselves to spectral or 'point' classifiers, where each pixel is considered as point observation and classified on the basis of its spectral information alone.

Within this group of classifiers, one may distinguish between supervised and unsupervised approaches, dependent on whether or not the analyst makes use of readily locatable training data (*in situ* data) for class characterisation. In addition, a distinction between statistical and non-statistical methods can be made, depending on whether or not a particular class statistical distribution is assumed (commonly the Gaussian distribution).

Some of the most common supervised spectral classifiers in remote sensing are:

- **Parallelepiped classifier** (also multivariate level slicing): a non-parametric supervised method, partitioning the feature space into multi-dimensional boxes around the spectral clusters on the basis of the class maximum and minimum data values. The pixel is assigned to the class in the box of which it falls. Some pixels may lie outside all defined parallelepipeds and will not be classified.
- **Minimum distance classifier**: a statistical supervised method, which calculates class means from training data and assigns a pixel to the class with the nearest mean. Results are not optimal, if the class covariance matrices are not equal.
- **Gaussian maximum likelihood classifier (MLC)**: a statistical supervised classifier, most commonly used in remote sensing. The Gaussian model is assumed for the class probability density function (pdf), which is calculated from the training data. The pixel is assigned to the most probable class according to the maximum value of the posterior class pdf. The normality assumption can only be justified with sufficient training samples being available to adequately

describe the classes of interest. In practice, the classes tend to have Gaussian distributions, and the classifier is relatively tolerant of deviations from normality (Swain, 1986). The classifier also allows for the incorporation of class prior probabilities, so that a minimum risk strategy can be adopted.

- **Spectral angle mapper (SAM)**: a statistical supervised classifier using the angular separation measure (see table 2.1 in section 2.2.2) to calculate the average angle between a reference and a sample spectrum, which indicates spectral similarity between the two. The spectra are treated as vectors in high dimensionality space.
- **Linear spectral unmixing (LSU)**: a sub-pixel ('fuzzy') supervised classification technique, which models the spectral reflectance of a pixel as the linear sum of cover proportions of known class spectral signatures and estimates class proportions via multiple linear regression. This classifier accounts for the 'mixed pixel' phenomenon occurring when two or more objects from different classes share the same pixel area.

In this chapter, the MLC was chosen for supervised classification, as it is the most commonly used classifier of all classification methods mentioned above.

Class definition and training

In general, two notions of 'classes' need to be distinguished: informational classes, i.e. the categories of interest to the data user, and spectral classes, which reflect groups of pixels that are uniform with respect to their values in each spectral band (Campbell, 1996). A direct link needs to be established between the spectral and informational classes to derive information from the data that is valuable to the data user. Ideally, a spectral class corresponds uniquely to an informational class.

If the informational classes have not been defined yet, the user should choose classes according to the following criteria (Wiersma and Landgrebe, 1980):

1) each class should be real and of interest to the user,
2) for each class sufficient training samples should be available in the imagery,
3) the class list should be exhaustive, so that all major classes on the image are accounted for,
4) each class should correspond to one or more spectral classes, as the latter can be mapped with high accuracy.

With respect to point 2), the question arises of how many training samples are sufficient for a certain application task. According to Hughes phenomenon (see section 1.1), the ratio of the number of training pixels to the number of spectral bands needs to be sufficiently high to ensure reliable estimates of class statistics. Labovitz (1986) recommended a ratio value of at least 2. Webb (1999) suggested the value to be at least around 5 to 10. Hsieh and Landgrebe (1998) chose a value of around 15, while Mather (1999) recommended a value of at least 10 to 30.

In the context of the application of MLC, a training sample needs to be relatively pure, i.e. its corresponding GIFOV has to contain only the objects belonging to the sampled class. If a class does not provide any pure training samples (i.e. the class cannot be resolved with the imagery's spatial resolution), or if the class does not provide a sufficient number of training pixels (see above), it may either be ignored or merged with other sub-classes to a larger category.

Regarding point 4), informational classes need to be spectrally distinct from each other in terms of features available, in order to assure high classification accuracy. The spectral classes of a data set represent separated clusters in high dimensional space and may be identified with clustering algorithms such as ISODATA (Iterative Self-Organising Data Analysis, Duda and Hart, 1973). The analyst may match them to informational classes using available ground data. This may help to identify spectrally separate informational classes, especially in cases where the classes are heterogeneously distributed over the entire scene and it is difficult to assign a certain pixel to a particular informational class.

Generally, class separability can be quantified with two-class heuristic or probabilistic distance measures (defined in section 2.2.2) and statistical tests (Labovitz, 1986; Penaloza and Welch, 1996). In this study, the general spectral separability between the defined classes was measured with an ENVI© (version 3.5) separability index that makes use of the transformed divergence measure. The index may take values between 0 and 2.0, with values greater than 1.9 indicating good separability. Classes may be removed from the class list, if they are not significantly different from all classes previously included.

When selecting training data, care should be taken to choose training pixels that are truly representative of the corresponding class population. Statistical estimates of population parameters are only adequate if the training data are random samples of the class population (Labovitz, 1986). By definition, a random sample must consist of individual pixels that are identically and independently distributed (Hogg and Craig, 1978).

Labovitz and Masuoka (1984) discovered spatial autocorrelation between adjacent pixels. Therefore, Labovitz (1986) recommended the use of more widely spaced training samples to produce a more robust estimate of training statistics for the purposes of signature extension. This may result in a classifier that is 20 to 25 % more accurate than a classifier with contiguous pixels in the training set (Labovitz, 1986). Craig (1979) suggested a sampling grid spacing of 10 instead of 1. Campbell (1996) advises the use of a larger number (at least 5 to 10) of small training areas per class instead of a few large ones.

Some sample pixels within the training data may not be representative with respect to their parent class. The inclusion of such outliers (aberrant, hybrid or mixed pixels) can seriously distort the sample statistics and hence the performance of the classifier (Mather, 1987). They can be removed

by an algorithm, which measures the distance of a sample pixel from the Gaussian curve of the parent class, and removes it if the distance is found to be large (Campbell, 1980).

Moreover, it is important to confirm the adequacy of the Gaussian model for each class population. This may be checked visually by displaying single band frequency histograms for the class training data and comparing its shape to a Gaussian curve. In addition, measures such as skewness or kurtosis may be calculated to quantify any deviations of the population from a normal curve. Skewness is a measure of the deviation of the distribution from symmetry and has a value of zero for the Gaussian distribution (see equation 2.6). Kurtosis measures the peakedness of the distribution and has a value of zero for the normal distribution (see equation 2.7). If the statistical model is found to be non-representative (for example for a bimodal frequency distribution), the class under investigation may be redefined or re-trained to fit a normal population, or a non-statistical algorithm may be employed for classification.

It is often difficult to judge whether other studies have adhered to the above training rules, as some authors do not explicitly describe the training sample selection process. For example, in studies that use hyperspectral data for image classification it is often not reported how many training samples per class were selected (for example, Thomson *et al.,* 1998b, Alberotanza *et al.*, 1999, Held *et al.*, 2003). However, the latter is important for judging the validity of class statistics and the associated classification result. A detailed description of the training selection process would be of benefit for the comparison of different classification studies.

Classification Methodology and Accuracy Assessment

After having defined the classes of interest according to the guidelines outlined above, the two data sets were initially classified with a maximum likelihood classification (MLC). This classification was performed on the entire band set to obtain an indication of the expected approximate level of accuracy and running time of the procedure for each data set. The classification performance was assessed with the overall classification accuracy estimate (see equation 2.11, section 2.2.2).

For the given class and data sets, the contingency table was created with the 'leave-one-out' cross-validation method rather than the 'holdout' estimate. The latter method splits the data into two mutually exclusive sets, a training set for classifier design and a test set for accuracy estimation. However, it makes inefficient use of the data by using only part of the data to train the classifier and gives a pessimistically biased error estimate (Devijver and Kittler, 1982). The 'Leave-one-out' cross-validation method uses n-1 out of n samples in the design of the classifier and tests on the remaining sample. This is repeated for all n subsets of size n-1 (Webb, 1999). Although this method is positively biased as the same samples are used in both the training and testing stages, the estimate uses all available *in situ* data for classifier design. This is particularly advantageous for hyperspectral data, as training samples are usually scarce, and a sufficient number of training pixels

are needed to calculate accurate class statistics. That is, if training samples are scarce, using the leave-one-out method may reduce the distorting effect of Hughes phenomenon on the accuracy estimate.

To isolate the effects of the various dimensionality reduction methods on the classification accuracy estimate, the same classification and accuracy assessment parameters (such as training data, MLC classification method, 'leave-one-out' cross-validation method) were applied to different band subsets. As only relative accuracy comparisons were undertaken for the same classification task and absolute accuracy estimates were not required, the positive bias of the 'leave-one-out' accuracy estimate was not of importance for this study. Throughout the thesis, differences between accuracy estimates were tested for significance at a 5% significance level using the z-statistic (see equation 2.41).

The results of the initial classification are shown in figures 3.8 and 3.9. The overall classification performance was estimated to 94.8% for the River Severn and 99.1% for the New Forest data set. The classification procedure was carried out with Multispec© software (version 2.5) running under Windows NT on a 600 MHz Intel Pentium III processor with 128 MB RAM. The River Severn data set took about 3 minutes of processing, while the New Forest classification result was completed after 22 minutes.

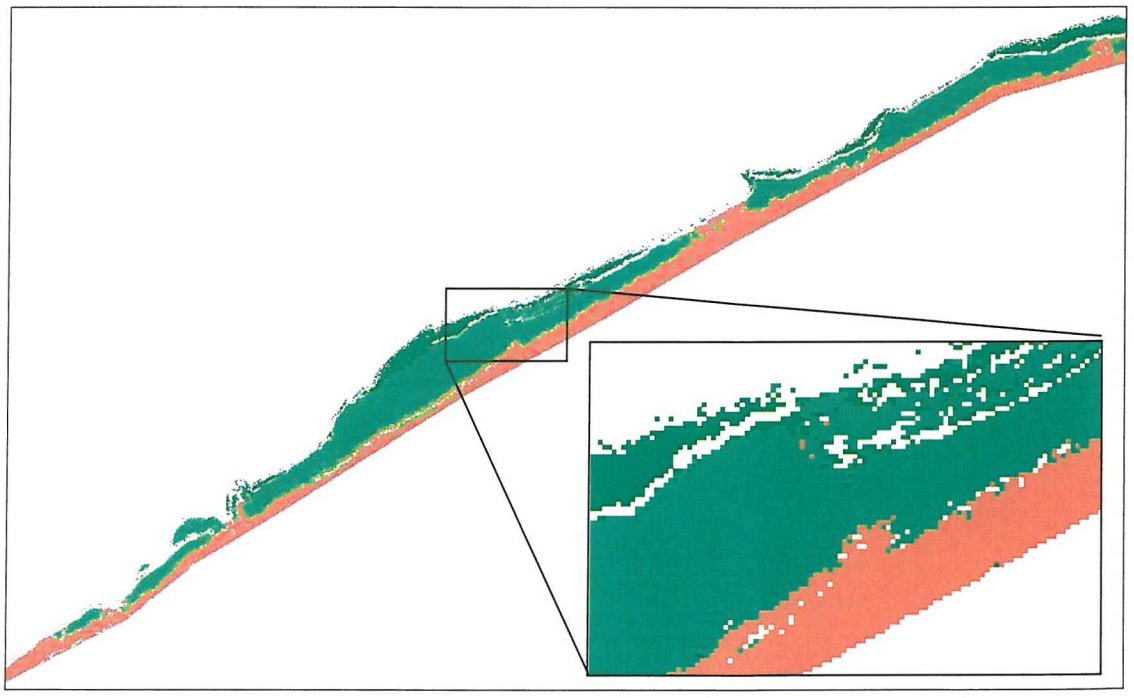


Figure 3.8: Masked River Severn Maximum Likelihood Classification result for all bands and the defined class set (Bare Rock, white; Pioneer Marsh, bright green; Mid Marsh, green; High Marsh, dark green; Bare Mud, brown).

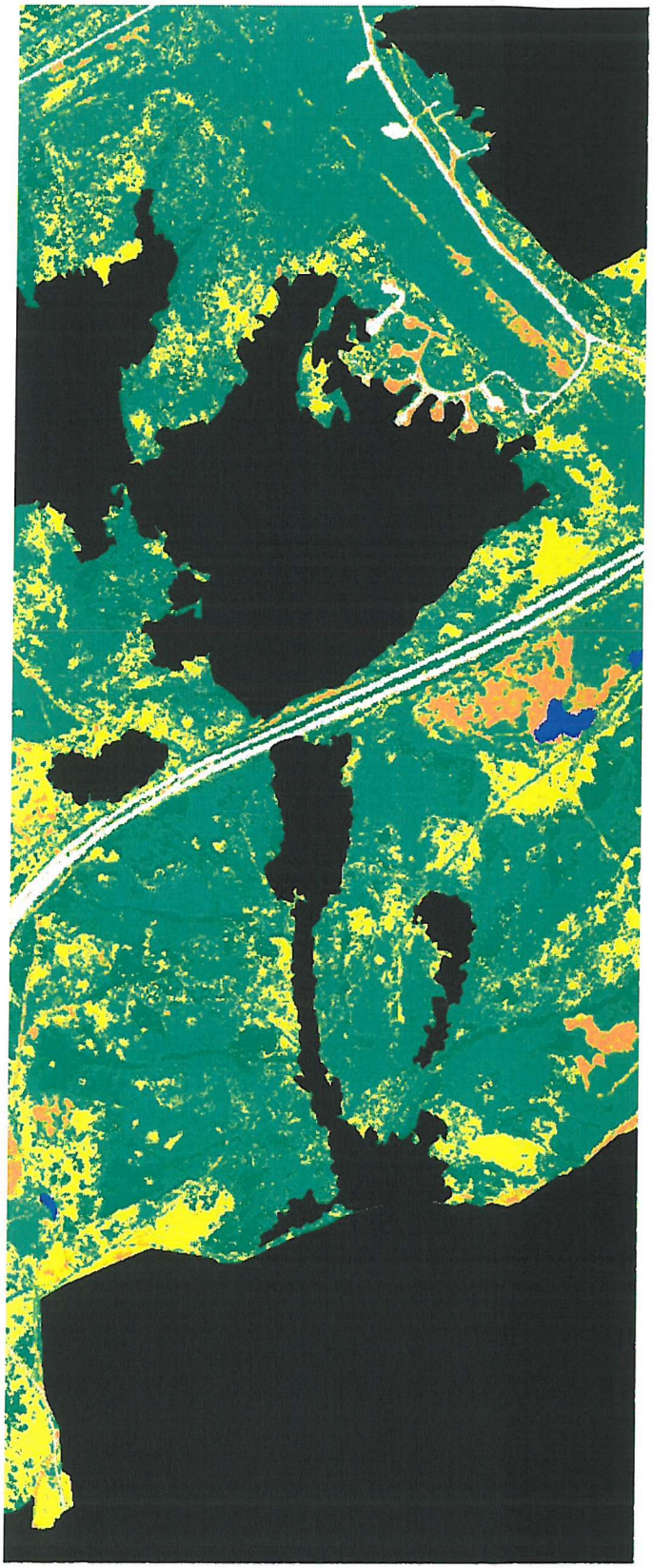


Figure 3.9: Masked New Forest Maximum Likelihood Classification result for all bands and the defined class set (Lake, blue; Asphalt, white; Bracken, yellow; Dry Heath, orange; Grassland, brightest green; Humid Heath, bright green; Wet Heath, green; Valley Mire, dark green).

3.3.2 Dimensionality Reduction

Hyperspectral imagery offers narrow spectral bands for the exploitation of minor differences in the spectral response between individual classes of interest, and thus, promises high classification accuracy. However, as stated in section 2.1, dimensionality reduction of hyperspectral data may improve greatly the overall classification performance. With less, but more information-bearing features, the redundancy in the hyperspectral data set, as well as the complexity of the resulting classifier, will be substantially reduced. In addition, the statistical estimates of the class population parameters will be more accurate for a limited amount of available training samples.

Figures 3.10 and 3.11 illustrate the data redundancy for the two data sets available via the band correlation matrix using the coefficient of determination (the colour white represents a value of 1, black one of 0). For both matrices, the auto-correlation value for the first band is located in the top left corner of the figure. The correlation matrix displays a high correlation value for bands of similar value, and a small correlation value for bands of dissimilar value.

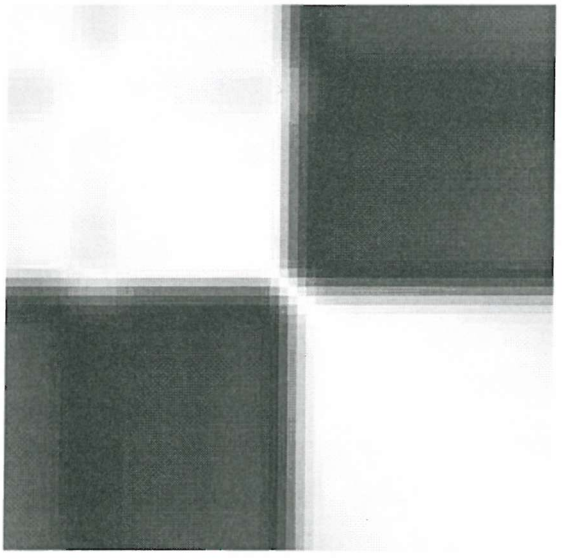


Figure 3.10: Correlation matrix of the 60 bands of the CASI River Severn data (12 bands have been masked out).

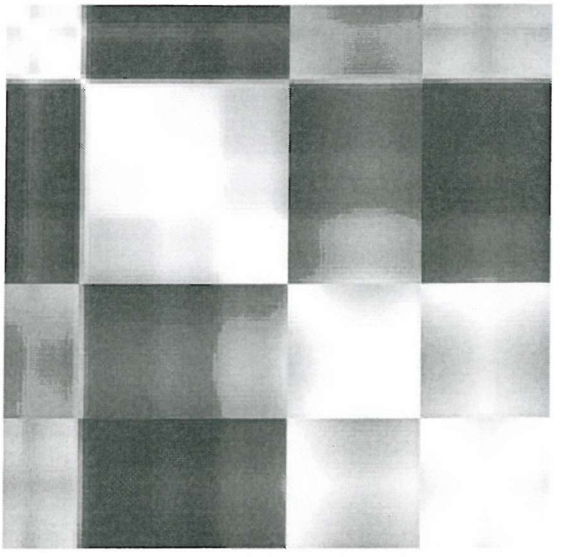


Figure 3.11: Correlation matrix of the 117 bands of the HyMAP New Forest data (9 bands have been masked out).

Both the CASI and HyMAP imagery are dominated by vegetation, which is reflected in the correlation matrices. In the VIS (CASI bands 10 to 39, HyMAP bands 3 to 15), reflectance values of a typical vegetation spectrum (e.g. see figure 4.1) are at a similar magnitude due to the blue and red absorption features resulting in high correlation between bands. The green reflectance peak gives slightly higher reflectance values and is therefore less correlated to neighbouring VIS bands. It shows as a darker stripe (CASI bands 18 to 21, HyMAP bands 7 to 10) in the VIS range of the correlation matrix. The red edge (CASI bands 40 to 44, HyMAP bands 16 to 18) forms the transition to the NIR (CASI bands 45 to 69, HyMAP bands 19 to 60), which represents the region of maximum

vegetation reflectance. Bands on the NIR plateau have similar vegetation reflectance values and are highly correlated with each other. As HyMAP imagery measures up to the SWIR wavelengths, it shows a slight decrease in correlation at longer NIR wavelengths (bands 34 to 60). The latter is connected to the drop in reflectance values partly caused by the water vapour absorption features at 942 and 1135 nm. HyMAP's two SWIR detectors measure between and beyond the two water vapour absorption features at 1379 and 1865 nm (SWIR-1 bands 66 to 94 from 1440 to 1812 nm, SWIR-2 bands 96 to 123 from 1960 to 2448 nm). As typical vegetation reflectance values are of similar magnitude within each SWIR region, bands are highly correlated within the two regions. As the vegetation reflectance values in the SWIR-2 are close to the ones in the VIS, an increased correlation is found between bands of these two regions.

Feature selection for classification

The main objective of feature selection for classification is to select a band subset of given size that maximises the classification accuracy, estimated with measures derived from the classification error matrix. Theoretically, a classifier may be designed on each possible feature subset and the subset resulting in the highest classification accuracy is chosen. However, it can be shown that a single classification accuracy estimation for the New Forest data set takes over 1.5 minutes of running time (using 3768 testing samples for 8 classes without class statistics calculation). Multiplied with the large number of combinations to test (see figure 2.2), the procedure becomes computationally infeasible (using a 600 MHz Intel Pentium III processor with 128 MB RAM).

A computationally cheaper alternative is to approximate the misclassification error by the overlap between class distributions, which may be quantified with heuristic or probabilistic distance measures. This method does not consider each testing image pixel, but only the data class statistics. In this pilot study, supervised probabilistic distance measures for the multi-class case will be investigated, that is the Bhattacharyya and Transformed Divergence distance measures sorted by minimum and average class distance. The latter assume a normal class distribution and use first- and second-order statistics to separate between the classes. Therefore, they are more suited than their heuristic counterparts as pre-processors for an MLC.

Nevertheless, using an exhaustive search method to examine all possible band subsets represents still a computational problem. For the New Forest data set with 121 bands, the search through all possible combinations of 6 bands would take over 22 days of processing with Multispec© (version 2.5 available from http://dynamo.ecn.purdue.edu/~biehl/MultiSpec) running on a 600 MHz Intel Pentium III processor with 128 MB RAM (see table 3.7).

Optimal and sub-optimal search algorithms (section 2.2.3) have been introduced to ease the computational load and have been shown to be effective for high-dimensional data sets. In this pilot study, the sub-optimal sequential forward selection (SFS) search procedure is used. The latter

method first selects the best band out of all bands, and then uses this band in the search for the best group of two bands. That is, for a data set of 121 bands, 121 possibilities exist for the search of one best band, 121 plus 120 combinations exist for the best subset of two bands, etc.

Table 3.7: Number of possible subset combinations and time needed for computational search for both exhaustive and sub-optimal feature selection algorithms (for Multispec© software version 2.5 under Windows NT on a 600 MHz Intel Pentium III processor, 128 MB RAM).

Number of selected features	Number of combinations for exhaustive FS algorithm	Search time for exhaustive FS algorithm	Number of combinations for sub-optimal SFS algorithm	Search time for sub-optimal SFS algorithm
1	121	0.06 s	121	0.06 s
2	7,260	3.63 s	241	0.12 s
3	287,980	2.4 m	360	0.18 s
4	8,495,410	1.2 h	478	0.24 s
5	198,792,594	1.15 d	595	0.30 s
6	3,843,323,484	22.2 d	711	0.35 s

In addition, the Projection Pursuit Feature Selection (PPFS) by Jimenez and Landgrebe (1999) (see section 2.3.3), which uses the Bhattacharyya class distance measure, was explored with the data sets. The processing of all feature selection techniques was performed with the Multispec© Windows version 2.5 and Macintosh version 3.0 (the latter was used for PPFS).

Feature extraction

Feature extraction methods transform the original high-dimensional measurement space into a new lower-dimensional feature space by optimising a certain criterion for the purpose of data compression. In this thesis only linear transformations were considered, which include unsupervised and supervised methods. Unsupervised methods, such as PCA or MNF, use a criterion that does not consider class statistics (e.g. data variance or SNR). In contrast, supervised methods, such as DAFE, DBFE and PP, use criteria that aim to discriminate between the classes under investigation.

In this pilot study, PCA, MNF, DAFE, DBFE and PP techniques were applied to the hyperspectral data sets introduced earlier. PP was also used as a pre-processing algorithm for following feature extraction with DAFE and DBFE. The performances of the feature extraction techniques were evaluated with the overall classification accuracy measured with the leave-one-out method, and compared with each other.

The processing of all feature extraction techniques was performed with the Multispec© Windows version 2.5 and Macintosh version 3.0 (the latter was used for PP), except for MNF, which was run with ENVI© version 3.5.

Intrinsic dimensionality (ID) estimation

ID estimates depend entirely upon the criteria to be optimised in the feature extraction process. For example, PCA maximises image variance, whereas MNF aims to optimise the image SNR. Therefore, discrepancy in ID estimates from different feature extraction methods is expected. The validity of the estimates, however, relies on how the user defines information to be captured by the features. For example, information in an unsupervised sense may be described by data variance, while in a supervised sense by the potential to discriminate between the defined classes.

Two notions of dimensionality have been introduced in chapter 2: intrinsic dimensionality (ID), defined as the minimum number of features required to represent most of the data variance, and intrinsic *discriminant* dimensionality (IDD), described as the number of significantly reduced dimensions that still result in satisfactory classification accuracy. The IDD measure may theoretically not exceed $M-1$ if features are statistically independent, where M represents the number of classes defined by the user (see section 2.4).

The results of the eigenanalysis-based unsupervised feature extraction methods PCA and MNF may be used to approximate the ID of the data sets, as the minimum number of features for data representation may be attained best via statistically independent features. In contrast, the eigenvalues of the supervised feature extraction methods DAFE and DBFE may be used to derive IDD estimates, as both methods transform the data into a minimum number of features that discriminate best between the given classes. The ID and IDD estimates may coincide, but as the best class-discriminating axes may not form an orthogonal basis, the IDD is expected to have a larger value than the ID.

Heuristic methods such as the scree plot (Catell-Vogelmann), 95% proportion of the total variance, Kaiser's criterion ($1/N$, where N is the number of original dimensions) and the broken-stick method were directly applied to the resultant eigenvalues of the above feature extraction methods.

In addition, the ID was evaluated by testing the significance of the difference in classification performance between a feature subset and all the remaining subsets of higher dimensionality. The classification accuracy was measured with the overall accuracy measure. To perform the significance test, the asymptotic sample variance of accuracy measure was calculated via equation 2.43. The z-statistic was evaluated for each pair made up of the actual feature and one of the remaining features. The rank of the feature achieving an absolute z-value below 1.96 with each one of the remaining features approximates the ID.

3.4 Results and Discussion

3.4.1 Mid Severn Estuary, UK

Training sites

The hyperspectral imagery of the River Severn Estuary demonstrates an expected high spatial and spectral variability within the plant communities outlined by Dargie's 1998 vegetation map (Dargie, 1999). This suggests that the National Vegetation Classification (NVC) plant categories used by Dargie to describe the make-up of individual communities are not appropriate for mapping with 3 m airborne imagery. Individual plant species are heterogeneously distributed over the entire community and it is impossible for the analyst to assign any particular pixel or region to any one of them. All areas other than saltmarsh were masked out with the exception of some bare mud regions, as it also appears throughout the marsh.

An unsupervised classification was performed with ISODATA under ENVI© version 3.5 to discover the true number of spectral classes of the image data. After several attempts with different parameters the algorithm found five major spectral classes roughly coinciding with the three broad saltmarsh zones (pioneer marsh, mid marsh and high marsh, after Long and Mason, 1983) and some bare rock and bare mud areas, which dominate parts of the scene. This class list was considered exhaustive, since all other 'uninteresting' class areas were masked out from the imagery.

In order to achieve a sufficiently high ratio of number of training pixels to number of bands (around 10 to 15), 600 to 900 training samples per class were necessary for 60 spectral bands to produce reliable class statistics.

The samples were chosen to be truly representative of the corresponding class population and were distributed across the entire scene. Class transitional areas were exempt from training sites. In order to achieve a high number of training samples, a wide sample spacing was not applied. Due to the inherent heterogeneity of the scene, the neighbouring training pixels were assumed to be independent.

Table 3.8 lists the classes defined for the Severn Estuary area, together with the number of samples and fields designed for each class. Class separability was tested and quantified with the divergence measure (using ENVI©). All pairwise distances were sufficiently large to separate between the defined classes (value greater than 2.0).

Table 3.8: Informational classes, their number of samples and fields defined for the River Severn study area.

CLASSES NAMES	ABBREVIATION	NUMBER OF SAMPLES	NUMBER OF FIELDS
Bare Mud	BM	2567	17
Bare Rock	BR	1027	102
High Marsh	HM	1011	8
Mid Marsh	MM	1149	7
Pioneer Marsh	PM	2124	7

To ensure that the Gaussian model was appropriate as underlying statistical model for each class population, frequency histograms were created for four representative channels (21, 33, 45 and 57) and overlaid with their corresponding Gaussian curve. In addition, skewness or kurtosis values were calculated to quantify any deviations of the population from the normal curve. Table B.1 in the appendix shows that the Gaussian curve was found to be representative for all classes. A perfect fit of the normal curve to the data is not expected and very unlikely to occur for all bands. Deviations generally happened in form of very peaked or skewed distributions. A severe bimodal frequency distribution was not observed. Figure B.1 and B.2 in the appendix display the corresponding skewness and kurtosis values, respectively. It may be observed that the fit of the normal curve to the class population data varies between bands chosen.

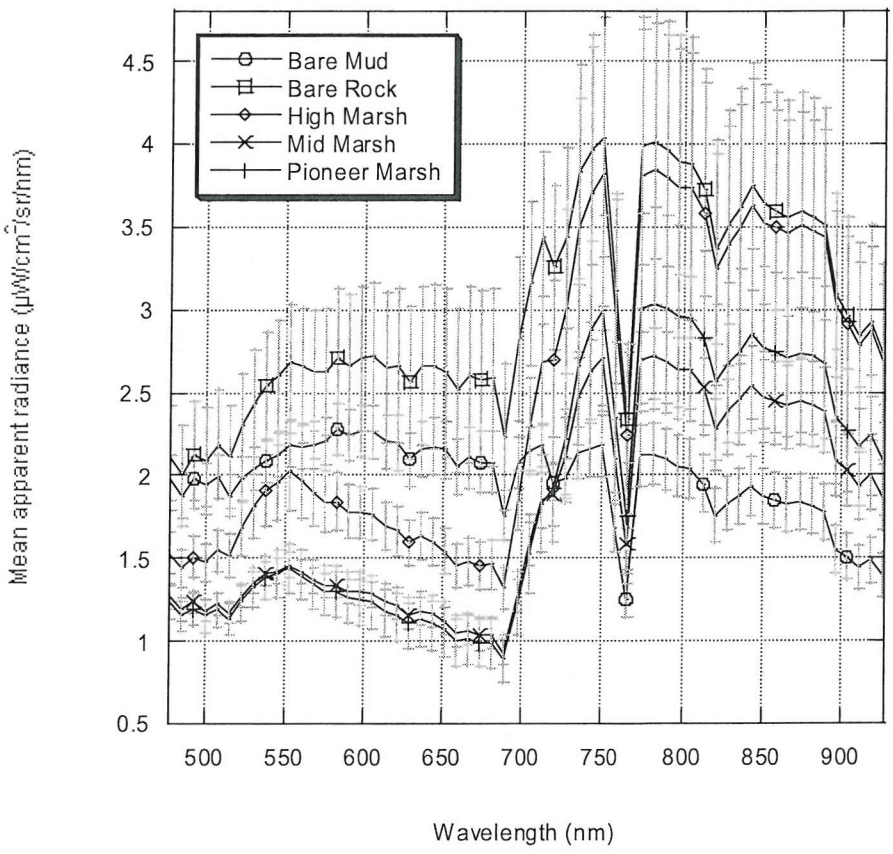


Figure 3.12: Mean spectral radiance curves of the class training areas defined over the River Severn data set. The mean curve is plotted with ± 1 standard deviation as grey error bar.

Figure 3.12 displays the class mean radiance values for the selected training areas along with the corresponding standard deviation plotted as grey vertical error bar. The mean spectral curves may be used to differentiate between the classes. Although the Mid and Pioneer Marsh curves are nearly identical in the VIS, as are the High Marsh and Bare Rock mean curves in the NIR, good separability between the mean values exists in the NIR and VIS, respectively. The standard deviation values are high as expected, due to the large spatial and spectral variability within each category. The Bare Rock class mean is very similar to the High Marsh curve in the NIR, as the Bare Rock class represents a spectral mixture of rock and vegetation classes. Pure Bare Rock samples in sufficient numbers were impossible to locate due to the large sensor GIFOV and the heterogeneity of the surface. The strong downward peak in band 48 (761.4 - 770.2 nm) represents the oxygen absorption feature at 762 nm, confirming the adequacy of the spectral calibration of the CASI sensor.

Feature selection

Figures 3.13 to 3.16 display the nine best feature sets selected with the SFS search method using the minimum and average Bhattacharyya Distance, the minimum and average Transformed Divergence. Since the SFS selects bands sequentially in a bottom-up manner, a feature set always includes all feature sets of smaller dimensionality. For this reason, only the band numbers and their frequency of appearance in the 9 sets are shown in figures 3.13 to 3.16. The band with the highest frequency appears in all 9 band sets. The band number is displayed next to the frequency column. A clover leaf spectrum is overlaid as wavelength reference. Wavelengths not available for feature selection are indicated by grey horizontal bars along the wavelength axis.

In contrast, PPFS does not select bands sequentially, but divides the available wavelength range into a number of sectors, which total the number of features to be selected. The band maximising the Bhattacharyya distance in each sector is part of the final band set. Therefore, the band sets are not necessarily nested, and figure 3.17 presents the wavelength locations of each band for each feature set. The exact band numbers for PPFS can be obtained from table B.2 in the appendix.

All feature selection methods applied showed similar results in terms of how often a certain wavelength region was selected. Practically no bands were chosen from the blue wavelength range which is partly due to the fact that most of these bands were excluded from processing due to insufficient SNR characteristics. Green wavelengths are equally under-represented. The majority of selected bands stem from the red, the red edge and NIR wavelengths, where the red and red-edge bands are related to chlorophyll content and the red-edge and NIR bands to canopy structure. As the defined salt-marsh vegetation classes may be distinguished from each other and from Bare Rock and Mud classes mainly by their canopy structure and chlorophyll content, the feature selection result was believed to be sensible. Nesting of bands occurred with all methods, suggesting that the intrinsic discriminant dimensionality for the River Severn data set is less than 9.

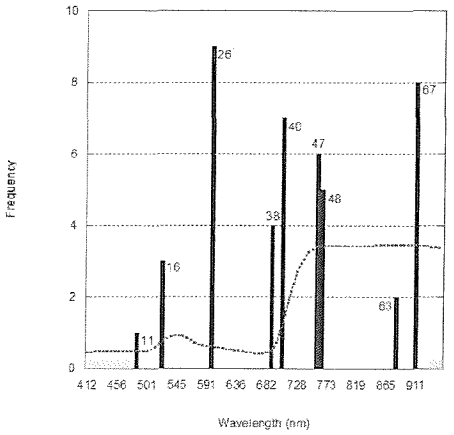


Figure 3.13: 9 best bands for the River Severn data selected via minimum Bhattacharyya distance and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

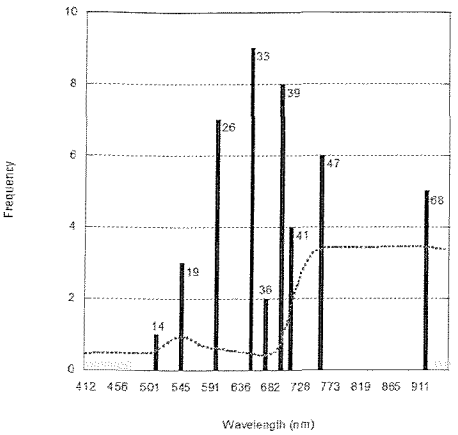


Figure 3.14: 9 best bands for the River Severn data selected via average Bhattacharyya distance and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

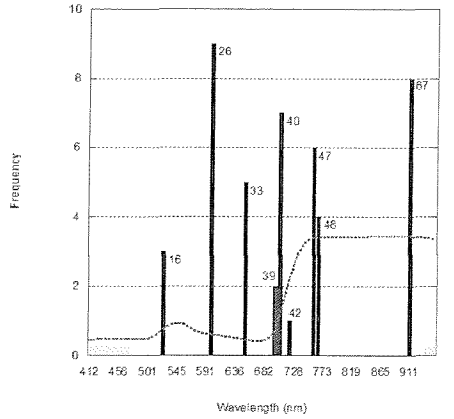


Figure 3.15: 9 best bands for the River Severn data selected via minimum Transformed Divergence and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

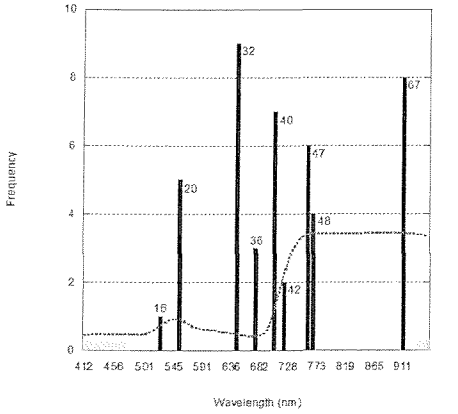


Figure 3.16: 9 best bands for the River Severn data selected via average Transformed Divergence and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

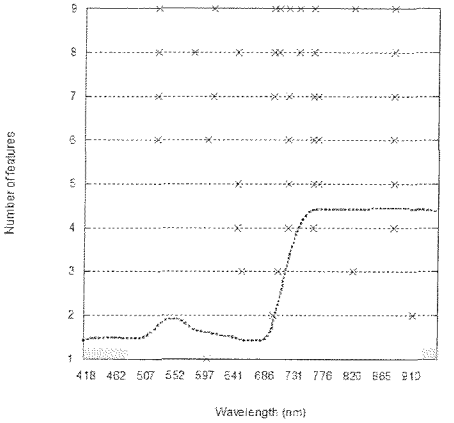


Figure 3.17: Best subsets with up to 9 bands for the River Severn data selected via Projection Pursuit Feature Selection.

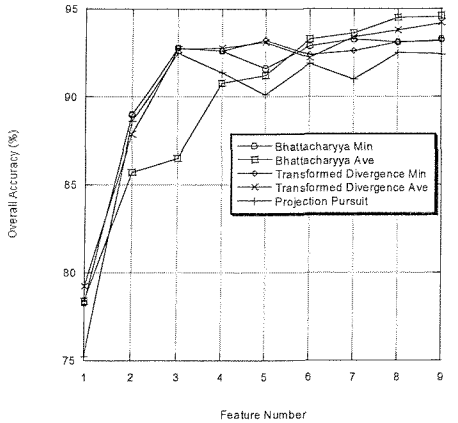


Figure 3.18: Overall MLC accuracy (leave-one-out method) as a function of number of bands chosen via the above techniques.

Surprisingly, the differences between band sets derived from the multi-class forms (minimum or average) of a distance measure seem to be larger than the ones derived from different distance measures (Bhattacharyya or Transformed Divergence) with the same multi-class form. This indicates that some thought may be given to the general class separability status of the data set when deciding the subset ordering.

The individual performances of the different feature selection algorithms on the MLC were evaluated with the overall MLC accuracy of their chosen band sets, which is plotted against the subset dimension in figure 3.18. In summary, all of the feature selection approaches performed well on the data set, with the average Bhattacharyya distance being distinctively less accurate than the rest of the methods for subsets of dimensionality smaller than six. The Projection Pursuit feature selection, which also uses the Bhattacharyya distance, may be considered as the second worst method, deviating from the average accuracy level for subsets with more than 3 bands. These differences may be explained by the unsaturated nature of the Bhattacharyya distance when compared to the saturated and therefore superior Transformed Divergence measure. The accuracies from different algorithms can be shown to be significantly different (5% level) from each other for the same dimension, with the exception of statistically insignificant differences in accuracy between the minimum Bhattacharyya and minimum Transformed Divergence methods for the first four dimensions.

The general level of optimality of the selected feature sets was compromised for computational feasibility by choosing sub-optimal search algorithms (SFS, PPFS) and class distance measures. Sub-optimal search strategies only yield optimal results when applied to statistically independent features. In addition, an incorrect Gaussian model assumption for the class statistics of some of the bands may have distorted their potential for class discrimination.

Feature extraction

Figures 3.19 to 3.21 present the scree plots resulting from the PCA, MNF, DAFE, PPDA (PP pre-processing and DAFE combined), DBFE and PPDB (PP pre-processing and DBFE combined).

The image noise covariance matrix of the River Severn CASI imagery was needed as input for the MNF, and since no dark data were available, it was estimated with the 'shift difference' technique (ENVI© version 3.5) from a selected image area. The latter method computes for each pixel the differences to its right and above neighbour, and assigns the average of both differences to the pixel. For the average difference value to represent image noise and not scene variability, the image pixels should be chosen from a relatively homogenous area. An area of intertidal mud was selected here for this purpose.

The PCA and MNF percentage eigenvalues are displayed in figure 3.19, and it may be evidenced that the PCA clearly outperforms the MNF with respect to generating highly informative features. Although the MNF may be applicable to the River Severn data due to the existence of unequal noise variances in the bands of the data set, the estimation of the noise covariance matrix was realised from image data rather than from dark data. The author believes that the poor MNF performance may be primarily attributed to an incorrectly estimated noise covariance matrix, even though great care was taken in the selection of a homogeneous area for evaluation.

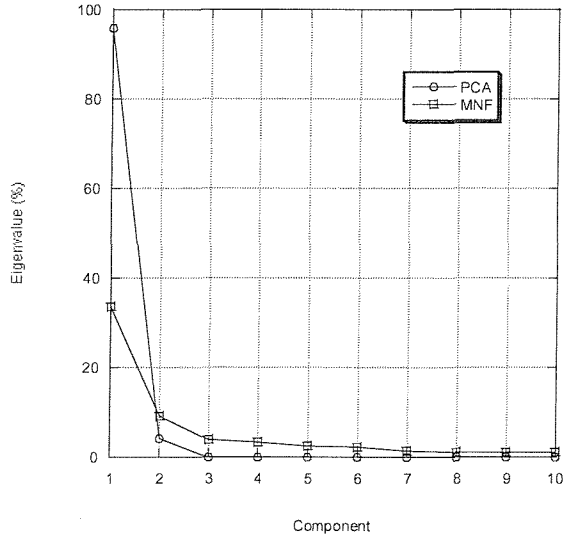


Figure 3.19: Eigenvalue (%) of the first 10 features extracted from the River Severn data via PCA and MNF.

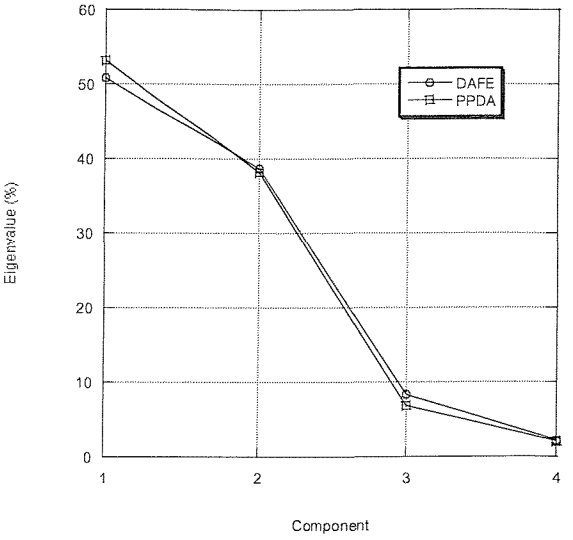


Figure 3.20: Eigenvalue (%) of the 4 features extracted from the River Severn data via DAFE and PPDA.

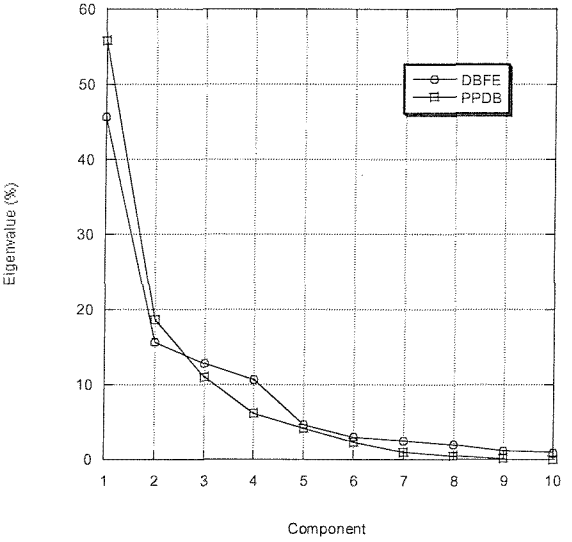


Figure 3.21: Eigenvalue (%) of the first 10 features extracted from the River Severn data via DBFE and PPDB.

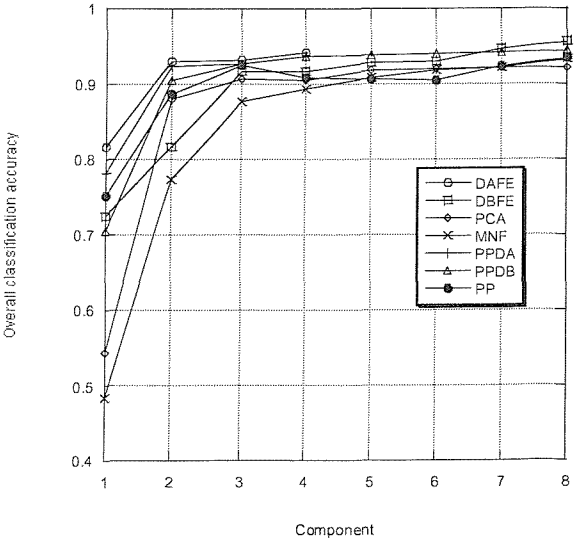


Figure 3.22. Overall MLC accuracy (leave-one-out method) as a function of number of features created via the above feature extraction techniques.

The percentage eigenvalues generated by DAFE and PPDA are very similar (see figure 3.20), suggesting that the Projection Pursuit pre-processing did not substantially contribute to the creation of highly informative DAFE features. In contrast, the DBFE and PPDB scree plots in figure 3.21 clearly demonstrate the improvement of the DBFE method with a prior Projection Pursuit pre-processing stage. DBFE features are created directly from class samples and a large number of samples are needed for a good performance of the algorithm. However, this number decreases for feature spaces that have already been transformed into a class-discriminatory space, for example with Projection Pursuit.

The performances of the feature extraction techniques were evaluated with the overall classification accuracy measured by the leave-one-out method, and are presented in figure 3.22. DAFE produced the most class-discriminatory features as evidenced by its significantly superior classification accuracy. The Projection Pursuit pre-processing for DAFE reduced the classification ability of its features only slightly. PP and DBFE features performed well, but were both outperformed by PPDB features, which favours the general use of PP pre-processing for DBFE. The PCA features were inferior to the ones created by most supervised methods, but the first four features produced by MNF noticeably achieved the worst classification result. The latter undoubtedly suggests that class-discriminatory information was not distributed along the orthogonal axes of maximum data variance.

Intrinsic dimensionality estimation

The unsupervised PCA and MNF were used to approximate the non-discriminant dimensionality, while the supervised methods DAFE, PPDA, DBFE, PPDB were chosen to determine the number of significantly discriminant features, which, in theory, cannot exceed 4 for the defined class set.

Figures 3.23 to 3.25 display the results of three heuristic ID estimation methods: 95% proportion of the total variance, Kaiser's criterion (1/N, here 1/60) and the broken-stick method. The Catell-Vogelmann estimate was read directly from the scree plots in figures 3.19 to 3.21. Tables B.3 to B.8 in the appendix list the results of the z-statistic for each feature extraction method. The latter was used to test the significance of the difference in overall classification accuracy between a feature and each one of the remaining features of lower rank.

All ID estimates are summarised in figure 3.26. Generally, for the unsupervised methods, the heuristic methods gave lower ID estimates (1 to 6) than for the supervised methods (3 to 8). This may suggest that the first few features of the PCA and MNF do not contain all class discriminative information. This is further evidenced by the relatively high ID value for the PCA and MNF derived from the statistical test of the classification result: nine PCA features and 8 MNF features were found to hold significant information for class discrimination. Two of the MNF based ID estimates are relatively high (6 from Kaiser, 53 from 95% total variance), implying that the MNF did not perform

well with the noise covariance estimate used. Theoretically, the IDD has a maximum value of 4, and DAFE and PPDA reflect this value by consistently producing exactly the same number of features.

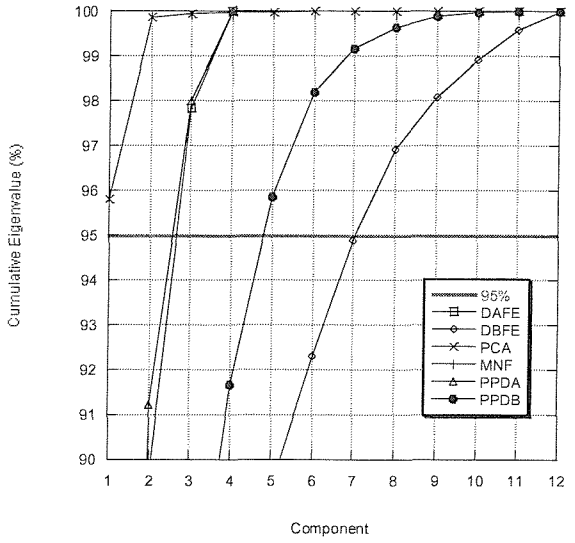


Figure 3.23: Cumulative eigenvalues (%) of the features extracted by DAFE, DBFE, PCA, MNF, PPDA, PPDB. The 95% mark is overlaid.

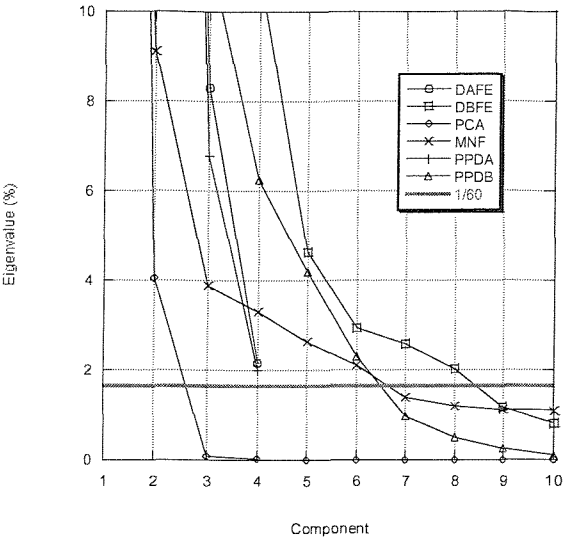


Figure 3.24: Eigenvalues (%) of the features extracted by DAFE, DBFE, PCA, MNF, PPDA and PPDB. Kaiser's 1/60 mark is overlaid.

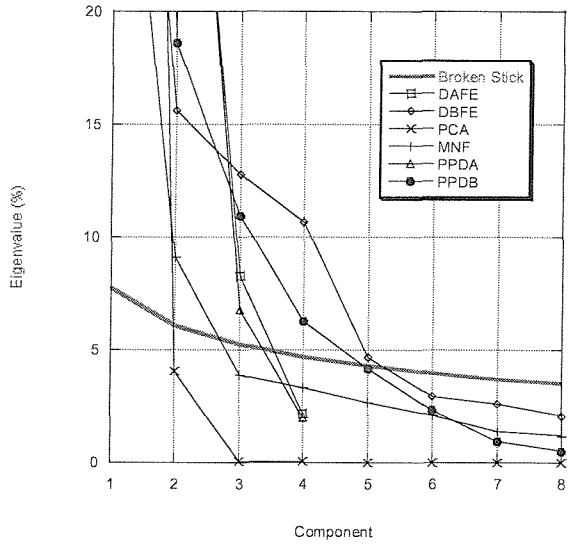


Figure 3.25: Eigenvalues (%) of the features extracted by DAFE, DBFE, PCA, MNF, PPDA and PPDB. Broken Stick values are overlaid.

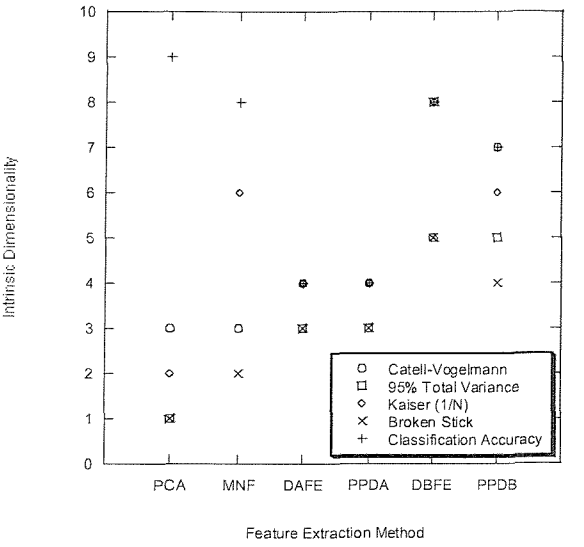


Figure 3.26: Summary of intrinsic dimensionality values for the River Severn data set estimated by Catell-Vogelmann, 95% total variance, Kaiser (1/N), Broken Stick and Classification Accuracy for the feature extraction methods PCA, MNF, DAFE, PPDA, DBFE and PPDB.

Their ID estimates range from 3 to 4. In contrast, the DBFE based methods exceed this theoretical limit in all cases (ID values from 5 to 8). This may reflect the inadequacy of the DBFE technique for

ID estimation when only a limited amount of training samples is available. Using the PP in a pre-processing stage shifted the ID estimates down by only one (ID values from 4 to 7).

As described in section 2.4.3, no ideal technique for ID estimation exists and the reader is referred to section 2.4.1 and 2.4.2 for a discussion of the deficiencies of individual methods. As a consequence, no single best ID and IDD estimate will be provided as final result, instead ID and IDD ranges. The ID may take values between 1 and 3 inclusive for the River Severn data, while its discriminant counterpart may fall between 3 and 4 for the class statistics defined on the River Severn data. MNF, DBFE and PPDB results were ignored for the final ranges as the author believed them to be not suitable for ID/IDD estimation for this data set.

3.4.2 New Forest, Hampshire, UK

Training sites

The New Forest study area consists of heterogeneous heathland areas, where individual plant species mix consistently at a scale of centimetres to metres. Using 3 m airborne imagery to classify different species the assignment of a pixel to a particular species is almost impossible, as it is to find pure training pixels on the ground. For this reason, vegetation classes were defined according to Clarke and Westerhoff's vegetation map of the area (Westerhoff, 1992)(see section 3.2.2), here used as *in situ* map.

To guide training site selection, the spectral classes of the data set were identified with the ISODATA clustering algorithm implemented under ENVI© version 3.5. The spectral classes matched surprisingly well some of the Clarke and Westerhoff's class categories. In one case, two spectral classes were merged to create an informational class. However, other categories were underrepresented in the scene and were omitted from the class table. To achieve a sufficiently high ratio of number of training pixels to number of bands (around 5 to 10; Webb, 1999), 585 to 1170 pure training samples per class had to be available to generate representative class statistics.

Altogether, eight categories were chosen as the dominant land cover classes for the study area: Grassland, Asphalt, Lake, Bracken, Dry Heath, Humid Heath, Wet Heath and Valley Mire. The enclosed woodland area was masked out manually, as it was not of interest in the mapping of semi-natural vegetation and was found to be spectrally similar to heathland shrubs in an unsupervised classification. Training samples were carefully selected on the basis of both the imagery and the unsupervised classification result, avoiding boundary pixels and using no sample spacing. Table 3.9 presents the classes defined for the New Forest area, together with the number of samples and fields designed for each class. All class pairs gave an ENVI© separability index value of 2.0, indicating that the defined classes are pairwise highly separable.

Table 3.9: Informational classes, their number of samples and fields defined for the New Forest study area.

CLASSES NAMES	ABBREVIATION	NUMBER OF SAMPLES	NUMBER OF FIELDS
Grassland	G	1315	6
Asphalt	A	641	19
Lake	L	816	3
Bracken	B	989	8
Valley Mire	VM	1391	12
Dry Heath	DH	667	16
Humid Heath	HH	2484	10
Wet Heath	WH	1098	10

Table B.9 in the appendix displays the frequency histograms of each class population for four bands representative of the visible (band 10), near infrared (band 40) and short-wave infrared (bands 80 and 100). All class distributions were generally well represented by the superimposed normal curve, suggesting that the Gaussian model was appropriate as underlying statistical model. No critical bimodal frequency distributions were observed, and some deviations from the normal curve occurred for some bands in skewed or very peaked distributions, for example for the Asphalt and Lake class (see also the skewness and kurtosis values in figures B.3 and B.4, respectively, in the appendix).

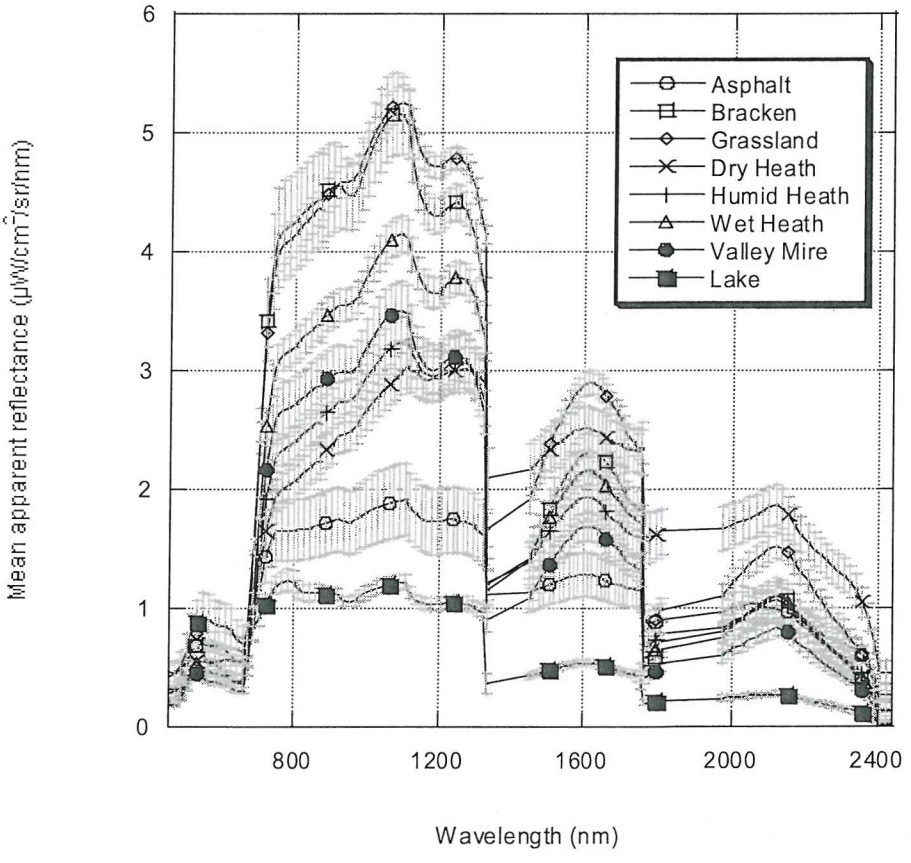


Figure 3.27: Mean spectral radiance curves of the class training areas defined over the New Forest data set. The mean curve is plotted with ± 1 standard deviation as grey error bar.

Figure 3.27 shows the class mean reflectance spectrum with grey error bars of one standard deviation. In general, the figure suggests a high spectral separation between all classes, due to well spaced mean spectra and relatively small standard deviations, especially in the near infrared and short wave infrared regions. No data were available for the spectral regions between the HyMAP detectors, indicated by missing error bars in figure 3.27.

Feature selection

Figures 3.28 to 3.31 display the frequencies of appearance of a feature in best 9 band sets selected with the SFS search method using the minimum and average Bhattacharyya Distance, and the minimum and average Transformed Divergence. The frequency of appearance may be interpreted as the band's importance with respect to discrimination ability between the given classes. The wavelength locations of the bands from the best nine PPFS subsets are plotted in figure 3.32 (for the exact band numbers see table B.10 in the appendix). For figures 3.28 to 3.32, wavelengths not available for feature selection are indicated by grey horizontal bars along the wavelength axis.

Although different multi-class forms and criteria produced different results in all cases, bands from all wavelength regions (VIS, NIR, SWIR), with exception of the blue region, were selected by each one of the individual feature selection methods. The resulting features seem to sample the entire spectral region fairly uniformly, with a higher sample density in the NIR wavelength region. The most frequent bands came to fall consistently into either the NIR or SWIR, especially near water vapour absorption features, and are sensitive to canopy structure and moisture content. This is consistent with the fact that most of the defined heathland vegetation classes differ more in their structure and water content than in their chlorophyll content.

Figure 3.33 shows the overall MLC accuracy of the nine best feature sets selected with the five methods (SFS with minimum/average Bhattacharyya Distance/Transformed Divergence, PPFS). To sum up, all feature selection approaches performed well on the data set, with the Transformed Divergence measures being superior to all other criteria. The average Bhattacharyya distance method produced the worst feature subset. The Projection Pursuit and the minimum Bhattacharyya feature selection gave the second worst single feature. As with the River Severn data set, the worse performance of the Bhattacharyya-based methods may be explained by the unsaturated nature of the distance measure. For the same dimension, accuracies from different algorithms were found to be significantly different (5% level) from each other, except when they were equal, e.g. for the four last dimensions for the average Bhattacharyya and average Transformed Divergence methods.

Feature extraction

Figures 3.34 to 3.36 present the eigenvalue plots of the following transformations: PCA, MNF, DAFE, PPDA (PP pre-processing and DAFE combined), DBFE and PPDB (PP pre-processing and

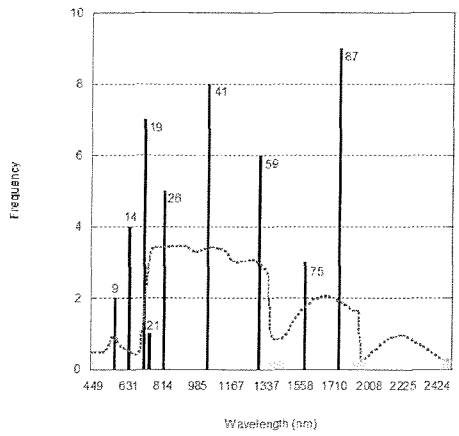


Figure 3.28: 9 best bands for the New Forest data selected via minimum Bhattacharyya distance and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

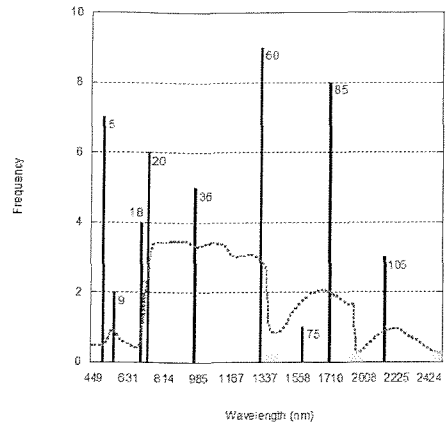


Figure 3.29: 9 best bands for the New Forest data selected via average Bhattacharyya distance and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

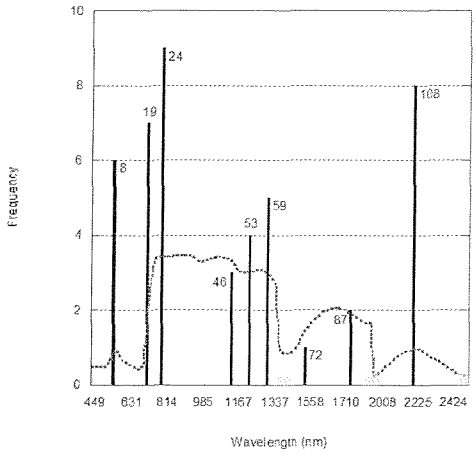


Figure 3.30: 9 best bands for the New Forest data selected via minimum Transformed Divergence and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

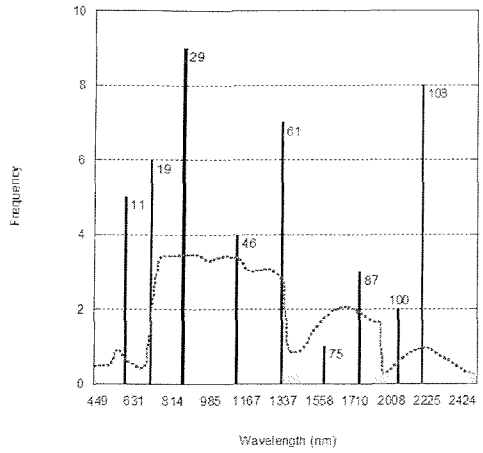


Figure 3.31: 9 best bands for the New Forest data selected via average Transformed Divergence and SFS. The height of the bars indicates the frequency of appearance of a band in the 9 subsets.

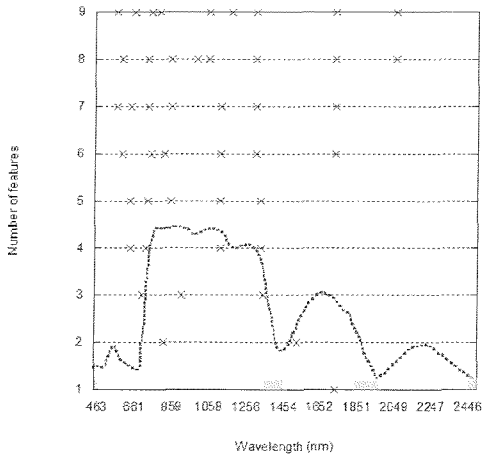


Figure 3.32: Best subsets with up to 9 bands for the New Forest data selected via Projection Pursuit Feature Selection.

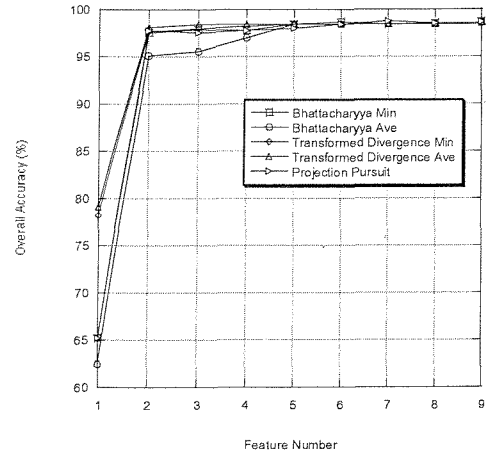


Figure 3.33. Overall MLC accuracy (leave-one-out method) as a function of number of bands chosen via the above techniques.

DBFE combined). The image noise covariance matrix of the New Forest HyMAP imagery was directly estimated from dark image data. The latter was used for the MNF transformation. The PCA and MNF percentage eigenvalues are displayed in figure 3.34. Both methods perform with similar high efficiency. Projection Pursuit pre-processing for DAFE increased the effectiveness of the features with respect to data representation (figure 3.35), while it seemed to substantially damage the ability of the DBFE to create highly informative features (figure 3.36).

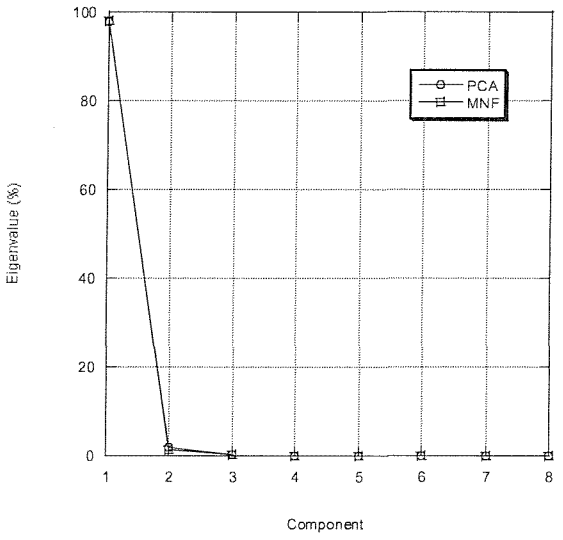


Figure 3.34: Eigenvalue (%) of the first 8 features extracted from the New Forest data via PCA and MNF.

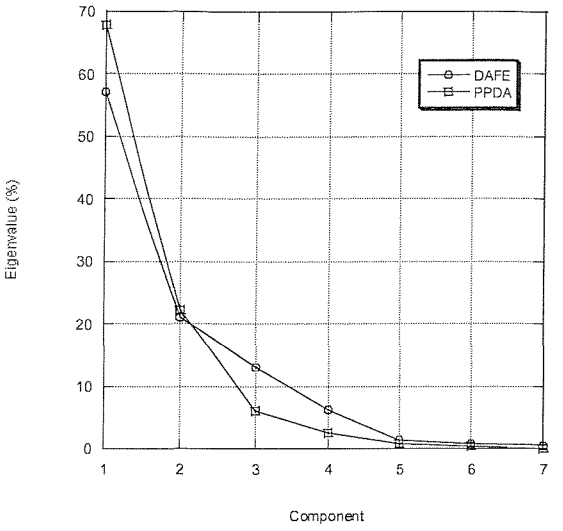


Figure 3.35: Eigenvalue (%) of the 7 features extracted from the New Forest data via DAFE and PPDA.

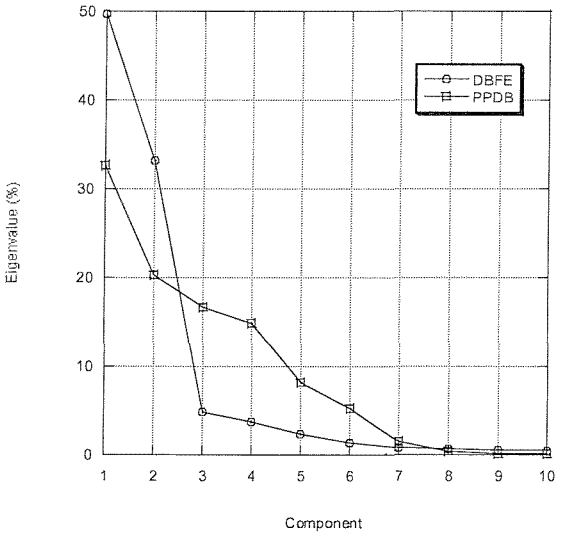


Figure 3.36: Eigenvalue (%) of the first 10 features extracted from the New Forest data via DBFE and PPDB.

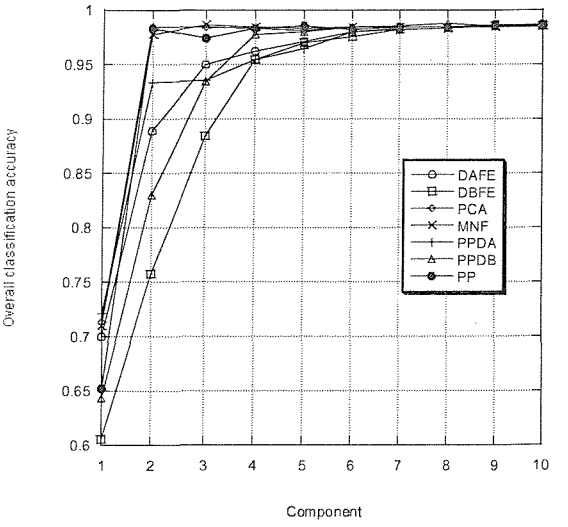


Figure 3.37: Overall MLC accuracy (leave-one-out method) as a function of number of features created via the above feature extraction techniques.

The overall MLC accuracy levels of the subsets created by the above feature extraction techniques were evaluated and are presented in figure 3.37. Surprisingly, PCA and MNF produced the most class-discriminatory features as evidenced by their superior classification accuracies. This suggests that class-discriminatory information was distributed along the axes of maximum data variance for the New Forest data set. Projection Pursuit feature extraction follows suit at third place, giving a nearly constant and high accuracy level for nine out of ten of its feature subsets. DBFE features perform worst, but their accuracy is improved with PP pre-processing. PP pre-processing for DAFE increases the accuracy of its first two features, but worsens the performance of its larger subsets. This result implies that PP pre-processing is beneficial for creating class-discriminatory DBFE features, but should be avoided when creating DAFE features for MLC.

For the first four dimensions, most FE methods produce feature sets that are significantly different in accuracy from each other. An exception to this are the PCA and MNF methods, which give features sets that, from the third dimension onwards, are statistically insignificant from each other in terms of MLC accuracy. From the fifth dimension onwards, most methods produce feature sets that do not significantly differ in the accuracy levels they achieve.

Intrinsic dimensionality estimation

The results of the PCA and MNF eigenanalysis were used to estimate the non-discriminant dimensionality, while those of the DAFE, PPDA, DBFE, PPDB methods were chosen to determine the number of significantly class-discriminant features, which, theoretically, cannot exceed 7 for the defined class set.

Figures 3.38 to 3.40 present the results of the 95% proportion of the total variance, Kaiser's criterion (1/N, here 1/117) and the broken-stick method. The Catell-Vogelmann estimate was determined directly from the scree plots in figures 3.34 to 3.36. Tables B.11 to B.16 in the appendix list the z-statistic for each feature extraction method, which was used to test the significance of the difference in overall classification accuracy between a feature and each one of the remaining features of lower rank.

All ID estimates are summarised in figure 3.41. Generally, the heuristic methods resulted in a relatively low ID (1 to 2) of the unsupervised methods and higher IDD estimate (3 to 7) of the supervised methods. The statistical IDD estimate for the PCA was two, suggesting that the first two PCA features contained all class discriminative information and that the spectral classes could be separated along the axes of maximum data variance. The heuristic and statistical IDD estimates of the DAFE methods give a minimum of three (PPDA, 95% total variance or broken-stick), while those of the DBFE methods reach as low as four (DBFE, broken-stick). With the classification accuracy of the PCA features being superior to the ones of all supervised features, the ID and IDD were estimated to be 1 and 2, respectively.

This result may reflect the necessity to judge the extracted features used for IDD estimation according their classification performance. For certain data sets, some supervised feature extraction methods may spread out the class-discriminatory information onto many features instead of a few, distorting the resulting IDD estimate.

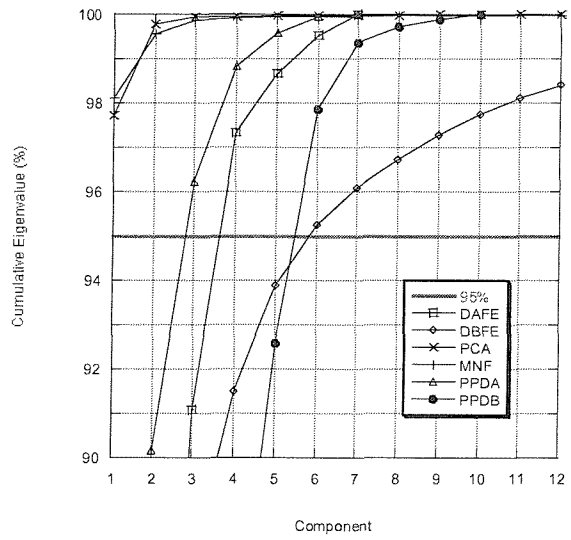


Figure 3.38: Cumulative eigenvalues (%) of the features extracted by DAFE, DBFE, PCA, MNF, PPDA, PPDB. The 95% mark is overlaid.

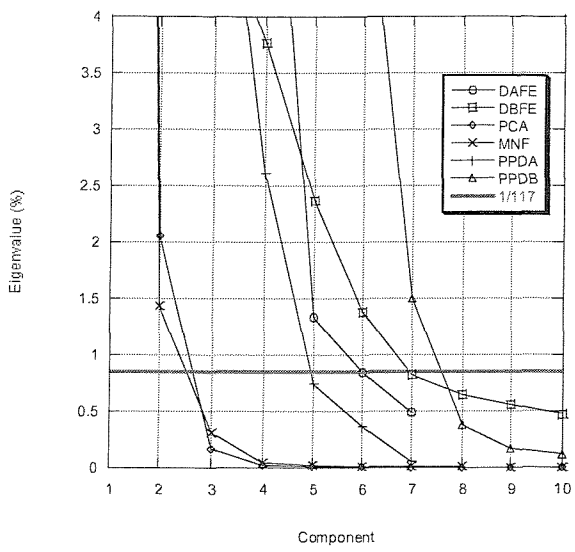


Figure 3.39: Eigenvalues (%) of the features extracted by DAFE, DBFE, PCA, MNF, PPDA and PPDB. Kaiser's 1/117 mark is overlaid.

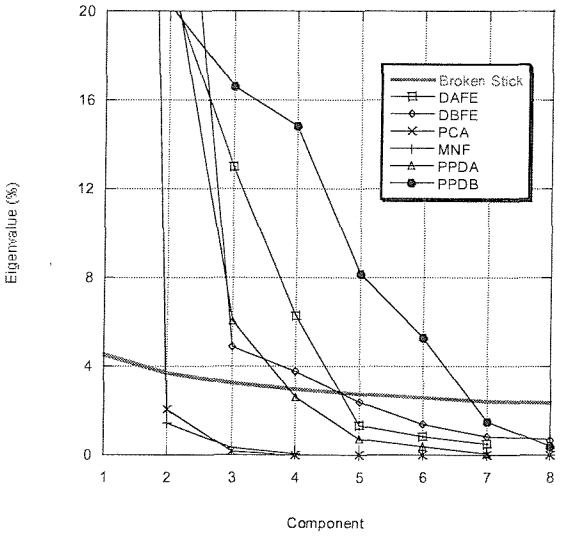


Figure 3.40: Eigenvalues (%) of the features extracted by DAFE, DBFE, PCA, MNF, PPDA and PPDB. Broken Stick values are overlaid.

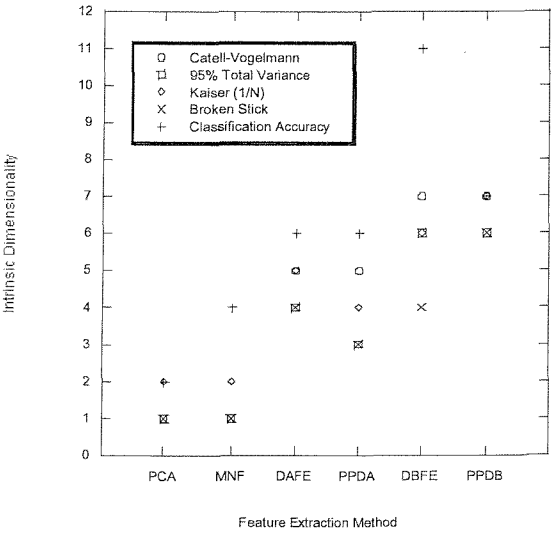


Figure 3.41: Summary of intrinsic dimensionality values for the New Forest data set estimated by Catell-Vogelmann, 95% total variance, Kaiser (1/N), Broken Stick and Classification Accuracy for the feature extraction methods PCA, MNF, DAFE, PPDA, DBFE and PPDB.

3.5 Summary

In this pilot study, some of the dimensionality reduction methods presented in chapter 2 were compared with respect to their classification performance and their ability to produce features that bear most class-discriminatory information. The two study areas of the thesis, the River Severn Estuary and the New Forest, both UK, were introduced in detail, as well as their available airborne and *in situ* data sets. General guidelines for class definition were described. Class training was performed using the result of ISODATA clustering and the vegetation map of the study areas.

Feature selection (FS) with the minimum and average multi-class forms of the Bhattacharyya and Transformed Divergence distance measures was performed to select the best spectral bands for class discrimination. The sub-optimal sequential forward selection search procedure was employed to make the selection process computationally feasible. In addition, the Projection Pursuit Feature Selection (PPFS) was applied, which uses the Bhattacharyya class distance measure.

The results of the FS methods were compared in terms of overall classification accuracy derived from the feature subsets and estimated with the leave-one-out cross-validation method. For both study areas all feature selection methods performed well, with the bands selected by Transformed Divergence distance measure achieving significantly higher accuracy than the ones chosen with the Bhattacharyya-based methods (including PPFS). The unsaturated nature of the Bhattacharyya distance measure, i.e. its misrepresentation of the classification accuracy, may explain this result. Also for both study areas, bands in all wavelength ranges except in the 'blue' were equally important for discriminating between the classes.

The classification performance of the unsupervised feature extraction techniques, Principal Components Analysis (PCA) and Minimum Noise Fraction (MNF), and that of the supervised methods, Discriminant Analysis Feature Extraction (DAFE), Decision Boundary Feature Extraction (DBFE) and Projection Pursuit (PP) was estimated for the data sets of both study areas. Additionally, a PP pre-processing stage was evaluated for both DAFE and DBFE, referred to as PPDA and PPDB, respectively.

For both data sets, DAFE features were consistently superior to the ones generated from the supervised PPDA, DBFE and PPDB method. Although pre-processing with PP improved the MLC performance of the DBFE features in both cases, it had a harmful effect on the one of the DAFE features. The generally worse accuracy of the DBFE features may be due to the fact, that not sufficient training samples were available for a good performance of the algorithm. However, the necessary sample number decreases for feature spaces that have already been transformed into a class-discriminatory space with Projection Pursuit, allowing for a higher DBFE feature performance.

PP features represented well the class-discriminant structure of the River Severn data set (60 bands), and outperformed all other supervised feature sets for the New Forest data set (117 bands). In the latter case, PP's success may be explained by the fact that PP does not perform the parameter estimation at full dimensionality, but at a linearly projected lower dimensional subspace, where the assumption of the Gaussian normal distribution is more justified. It may be possible that the training samples were not sufficient to adequately define the parameters of the assumed Gaussian class distribution for the 117 bands of the New Forest data set.

The unsupervised PCA and MNF produced highly class-discriminatory features for the New Forest data, but not so for the River Severn data set. This suggests that, only for the New Forest data set, class-discriminatory information was distributed along the axes of maximum data variance. The inferior performance of the MNF for the River Severn data set was attributed to a poor estimation of the noise covariance matrix from image data (rather than dark data as for the New Forest data set).

The results of the PCA and MNF eigenanalysis were used to estimate the intrinsic dimensionality (ID), the minimum number of features to represent the data variance. Those of the DAFE, PPDA, DBFE, PPDB methods were chosen to determine the intrinsic discriminant dimensionality (IDD), i.e. the minimum number of significant class-discriminating features. Heuristic and statistical dimensionality estimation methods were applied, including the 95% proportion of the total variance, Kaiser's criterion (1/N), Catell-Vogelmann, broken-stick and MLC accuracy hypothesis testing.

Generally, the PCA was preferred to the MNF as basis for estimating the ID, since, by definition, the data variance, and not the image SNR, was defined as the information to be represented by the ID features. The resulting ID estimated with the above methods for the New Forest data set ranged between one and two, and between one and three for the River Severn data set. Exceptionally, the PCA could also be used for determining the IDD of the New Forest data set, as its features performed superior to all those derived from the supervised methods. The IDD for this data set was evaluated as two. DAFE achieved the highest MLC accuracy on the River Severn data set and estimated its IDD to fall between three and four. These relatively low ID and IDD results for both data sets are in accordance of the findings of other authors (e.g. Milton, 1999, Curran *et al.*, 1998), that the intrinsic dimensionality is of low order for imaging spectrometer data of vegetated scenes in the Visible and Near Infrared (VNIR).

In short, the results of this pilot study showed that

1) the Transformed Divergence performed superior to the Bhattacharyya distance as surrogate measure for the MLC accuracy,
2) various feature selection and extraction methods resulted in different feature subsets that all gave sufficiently high classification accuracy, and
3) various heuristic and statistical IDD estimation methods gave a range of solutions, rather than a single common estimate.

From the experience with the given two data sets, the author recommends to principally base the IDD estimation on DAFE, but to check whether principal components achieve higher classification accuracy with less features than DAFE, as class-discrimination may be possible along the axis of maximum variance.

The second point from the above list implies that a multitude of acceptable feature subsets are possible for a given class- and data set. At this point it is important to recognise that the feature subsets of the above dimensionality reduction methods cannot be generalised (Dutra and Huber, 1999) and that they are specific to

- the feature selection or extraction algorithm employed,
- the data set,
- the class definition and training performed by the user,
- the data or class model assumptions, and
- the final data application.

That is, no universal best feature or band set exists, questioning the pre-defined or 'established' band sets of some of the imaging spectrometers introduced in chapter 1. The remainder of this study investigated whether the 'established' band sets could be used for vegetation mapping in semi-natural environments and whether hyperspectral data was indeed necessary to achieve high accuracy for the same classification task.

Band selection algorithms were developed in the following two chapters as a tool for answering the above questions. The algorithm extends the capability of feature selection algorithms presented above by considering bandwidth and SNR issues, and allows the creation of optimal band sets for a given data set and application task.

4 Supervised Band Selection for Classification

4.1 Introduction

Although high spectral resolution sensor capability is desirable as it allows the exploitation of characteristic narrow absorption or emission features of target classes of interest, imaging spectrometer data are often redundant and have a large volume that is difficult to store and process efficiently. In addition, a lack of sufficient training samples for high dimensional data will result in poor classification results (Hughes, 1968). Therefore, dimensionality reduction techniques (see chapters 2 and 3) are commonly applied at processing level after data acquisition.

However, hyperspectral capability also creates problems at sensor level. The huge data volume to record, store and transmit often forces the sensor designers and operators to make compromises. High spectral resolution capabilities result in a reduced SNR, which is often compensated by coarser spatial resolution, which may not be in the interest of the user. As for the CASI-2, a trade-off exists between maximal swath width and high spectral resolution due to data recording rate limits of the removable hard drive subsystem (see chapter 1).

Regarding these drawbacks of hyperspectral data collection, one might ask whether the full band set needs to be acquired in order to achieve a highly accurate classification result, and whether such fine spectral resolution generally leads to an improved classification performance. In short, what bands and bandwidth can be chosen for data acquisition without losing the full class-discriminating power of the hyperspectral band set?

The latter question has already become reality for sensor system designers and users of imaging spectrometers with data recording limits such as the CASI-2. In the latter case, if the data users decide to acquire a full-swath image in the CASI-2's spatial mode, they can choose up to 19 out of 288 non-overlapping bands of arbitrary width (see chapter 1). Some users select 'established' band sets recommended by the instrument operators and the literature, or simulate common satellite sensor bands. However, these 'default' band sets may not be optimal for a specific classification task at hand involving particular scene data acquired at a certain point in time.

This chapter focuses on band selection for imaging spectrometers with Maximum Likelihood classification as final data application. A new feature selection based method is presented for supervised band selection, which uses both class means and variances to discriminate between user-specified classes and introduces the bandwidth as a variable into the band selection process. The method provides a set of most class-discriminating bands of different widths for a specific data scene and class definition. The new band selection technique is applied to the CASI and HyMAP

data sets introduced in chapter 3 and results are evaluated and discussed. In chapter 6, a new data set is introduced in as a further test of this new algorithm.

Although an extensive feature selection review is presented in chapter 2, the author believed it was necessary to introduce, in an additional review, what factors needed to be accounted for when selecting bands for a sensor as opposed to features for a processing algorithm. The review also describes some other independently developed band selection methods and points out the shortcomings of these approaches.

4.2 Band Selection Review

In general, sensor band selection for remote sensing of target reflectance should take the following factors into account:

1) target reflectance properties,
2) solar spectrum, atmospheric absorption and scattering,
3) sensor characteristics (e.g. spatial and spectral resolution, point-spread function PSF, wavelength range, instrumental noise, SNR), and
4) data application (e.g. for classification, regression model).

4.2.1 Target Reflectance Properties

Usually, the data users have some knowledge about the target classes under investigation. As a first step they may examine their physical reflectance properties and search for specific spectral features (e.g. absorption features or reflectance peaks) unique to each target class. Absorption features in particular may be used to distinguish between different surface types and conditions, as they may evidence the presence or lack of certain primary chemical components unique to a specific surface type or condition. For example, the absorption features of vegetation include those of photosynthetically active leaf pigments such as chlorophyll, bound and unbound water, cellulose, lignin, starch and proteins (see for example Kumar *et al.*, 2001). In addition to absorption features, structural properties of the canopy, such as leaf thickness and number of stacked leaves, influence the spectral properties of vegetation.

Sensor bands and bandwidths may then be chosen according to individual spectral features of the target classes in order to extract target-specific information from the band measurements (Thomson *et al.*, 1998a). A typical set of spectral bands to sample a vegetation spectrum includes:

- a blue band centred on the absorption peak of chlorophyll at around 443 nm,

- a green band centred on the vegetation reflectance peak (chlorophyll absorption minimum) at around 555 nm,
- a red band centred on the absorption peak of chlorophyll at around 665 nm,
- a NIR band corresponding to a region of maximum vegetation reflectance between 750 and 1350 nm, typically centred at 865 nm and primarily related to the structural property of the canopy and the percentage of soil covered by vegetation,
- a SWIR band centred at around 1650 nm, where reflectance is related to water content of the canopy components and to its structure.

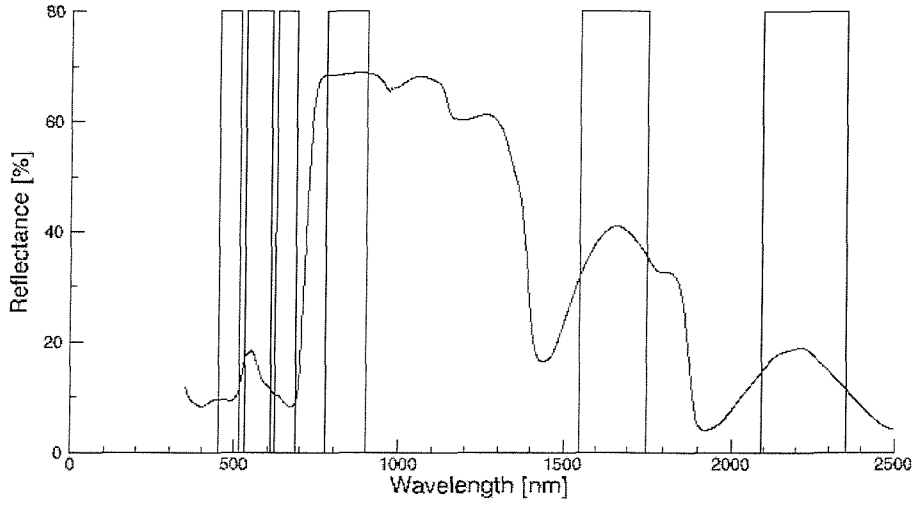


Figure 4.1: The six VNIR channels of Landsat ETM+ overlaid over a clover leaf spectrum.

Most of these bands are included in the band sets of satellite sensors designed for vegetation mapping, such as low to medium spatial resolution multispectral sensors (e.g. Landsat TM / ETM+ see figure 4.1, AATSR, ALI, ASTER, AVNIR), high spectral resolution sensors (e.g. MISR, MERIS, MODIS) and commercial high spatial resolution sensors (e.g. IKONOS, SPOT HRVIR, Orbview-3, Quickbird). The reader is referred to tables 1.1 and 1.2 for a list of airborne and spaceborne VNIR sensors, respectively, which have been in operation since 2002. The tables include Internet hyperlinks to most of the instruments and their calibration.

Price (1990, 1994a, 1997, 1998) showed another way of selecting bands to represent target spectra. He calculated spectral basis functions to approximate successively the narrow-band target spectra and selected few broad-band spectral intervals to determine the coefficients of these basis functions. However, the chosen bands have no physical significance and the validity of the basis functions is restricted to spectra that can be represented by the original target spectra the basis functions were derived from. In addition, the data dimensionality is reduced only at sensor level, not at processing level and the method is conceptually complex and difficult to implement. For these reasons, Price's method is not pursued any further in this study.

Some common target reflectance curves (such as water, leaf or mineral spectra) can be found in the literature. Hunt (1977, 1979) and Hunt and Ashley (1979) described spectral curves of minerals and rocks for the VNIR wavelength range (see also Hutsinpiller, 1988). A compiled list of absorption features of vegetation compounds can be found in Kumar *et al.* (2001). Curran (1994) presented a list of wavelength regions of interest for different fields of imaging spectrometer data applications, including geological, aquatic, ecological, and atmospheric studies.

If the target spectra are unknown to the user, or the targets represent complex mixtures of different elements, the user may acquire spectral data of the target classes from laboratory or field measurements or from training sets of known ground surface in imaging spectrometer data. For example, Dekker *et al.* (1992) reconstructed the spectral signature of specific lakes using field spectral measurements.

The target spectra may be searched for class-characteristic absorption features simply via visual inspection or by using techniques developed to detect them. For this task it is advisable to transform the spectral measurements to reflectance, and therefore remove solar irradiance and atmospheric effects, in order not to distort or hide class absorption features.

Absorption Feature Enhancement

Absorption features of a target spectrum may be enhanced by calibrating radiance data to reflectance. Reflectance calibration is performed by removing the effects of both the atmosphere and the solar curve. For ground or laboratory spectral measurements a preferably simultaneous irradiance measurement at target location is needed, while for aircraft-based measurements, the atmosphere between the target and the sensor has also to be accounted for.

A widely accepted method for reflectance calibration is to measure atmospheric conditions during the flight overpass with specialised on-site sensors and then use a detailed atmospheric radiative transfer model (RTM) to convert at-sensor radiance to surface reflectance. Examples of RTMs are

- LOWTRAN 7 (Low Resolution Atmospheric Radiance and Transmittance; Kneizys *et al.*, 1989),
- 6S (Second Simulation of the Satellite Signal in the Solar Spectrum; Vermote *et al.*, 1997),
- MODTRAN 3 (The Moderate Resolution Atmospheric Radiance and Transmittance Model; Anderson *et al.*, 1995),
- ACORN – Atmospheric CORrection Now (Analytical Imaging and Geophysics LLC, 2003).

Though accurate, such models require a well-calibrated sensor and *in situ* measurements or assumed values of atmospheric variables (such as water vapour or aerosol content) at the time of data acquisition, which may be difficult to obtain. Consequently, some RTMs have been simplified to minimise the number of atmospheric inputs. Teillet and Fedosejevs (1995) proposed an approach in

which aerosol optical depth was derived from the RTM using a dark target in the image itself, while Gao and Goetz (1990) retrieved water vapour amounts from water vapour image bands (ATREM, ATmosphere REMoval program; Gao *et al.*, 1992).

Other reflectance calibration methods have been created to avoid atmospheric *in situ* measurements altogether. Some authors proposed that an empirical linear relationship between at-sensor radiance (or DN) and reflectance based on within-image targets of known reflectance could be used for reflectance calibration ('Empirical Line', EL, method; Roberts *et al.*, 1985; Smith and Milton, 1999). The method relies on field or laboratory reflectance data of at least two spatially and spectrally uniform calibration targets, preferably one dark and one bright target, for which both image radiance and ground reflectance data are available. Spectral Mixture Modelling may also be employed for reflectance calibration using a reference spectral library (Farrand *et al.*, 1994). Clark *et al.* (1993, 1995) introduced a hybrid approach that first applies an RTM to correct for path radiance and solar spectral response, and then eliminates residual artefacts with the empirical line method using *in situ* measurements of one or more calibration targets. The U.S. Geological Survey has termed the latter method 'radiative-transfer-ground-calibration' (RTGC).

Other methods that do not require any *in situ* measurements at all include:

- 'Internal average relative reflectance' (IARR) method (Kruse, 1988),
- Flat-Field correction method (Roberts *et al.*, 1986; Hutsinpiller, 1988),
- Residual method (Marsh and McKeon, 1983; Green and Craig, 1985), and
- Dark-object subtraction.

The IARR is the ratio between the pixel spectrum and average spectrum of all scene pixels. Usually, the data are normalised by scaling the sum of the DN values in each pixel to a constant value ('equal-area' or 'equal-energy' normalisation).

The 'Flat Field' method depends on the existence of large spectrally flat and uniform areas in the image, the average radiance spectrum of which is divided into each pixel spectrum to calculate apparent reflectance. The method assumes that solar and atmospheric effects alone cause the shape and features of this average spectrum. If the latter assumption does not hold, target features will show up less intense in the residual spectral plot ('feature fading').

Marsh and McKeon (1983) created residuals by subtracting the average spectrum for the entire scene from individual pixel spectra. The residual method assumes that the average scene signal contains irradiance and atmospheric absorption features, but no target spectral features. Before creating residuals, all pixel spectra are scaled to a constant value in a reference channel relatively free of target absorption features to account for illumination variations between radiance spectra. Green and Craig (1985) calculated logarithmic residuals for specific pixel-channel combinations by dividing raw data values by their geometric mean over all bands and their geometric mean over all

pixels. The logarithm is applied to replace the geometric means by arithmetic ones. Hutsinpiller (1988) advised to use the flat-field log residual technique, using average values from spectrally flat scene segments, when absorption feature fading occurs.

Comparisons between some of these reflectance calibration methods were performed by Roberts *et al.* (1986), Conel *et al.* (1987), Crowley (1990), Farrand *et al.* (1994), Dwyer *et al.* (1995), Ferrier (1995), Ben-Dor and Levin (2000) and Perry *et al.* (2000). Generally, the RTMs performed most accurately, closely followed by the empirical line method, which produces usually more accurate reflectance results than the other normalisation procedures that use no additional information. The U.S. Geological Survey found that the hybrid RTGC method gave the best results as it corrects for the artefacts introduced by imperfect RTMs (Clark *et al.*, 2002).

Apart from reflectance calibration methods, techniques have been developed especially to enhance absorption features. Clark and Roush (1984) suggested calculating a continuum over each pixel spectrum by fitting a second-order polynomial as upper convex hull to selected channels without known absorption features. Dividing each pixel spectrum by its corresponding continuum spectrum may then reveal all absorption features, including atmospheric ones ('continuum correction' or 'hull-quotient').

Crowley *et al.* (1989) proposed the calculation of 'Relative Absorption Band-Depth' (RBD) images to discern particular diagnostic absorption features known to occur for a specific scene. The method does not require the data to be calibrated to reflectance (only to radiance) nor to be normalised. Several pre-defined channels near an absorption feature shoulder are summed and then divided by the sum of several pre-selected channels near the feature minimum. RBD corresponds to a local continuum correction reducing any radiance variations related to topographic slope and albedo differences. Nevertheless, the choice of channels used to define an absorption band is crucial for the success of the method, and indicative scene absorption features have to be known in advance.

If absorption features of particular target materials are known to exist in certain wavelength regions of the spectrum, Curran *et al.* (1998) suggested the use of PCA within those regions for the enhancement of very narrow absorption bands. The absorption bands discriminate best between the target materials and background and account for the greatest amount of variance in these regions (Feldman and Taranik, 1988; Hutsinpiller, 1988), which is reflected in high eigenvector loadings of the first PC feature(s) (see also section 2.3.5).

Absorption Feature Detection

All absorption feature detection methods presented in the following paragraphs require reflectance data as input, or at least data normalised with respect to illumination or topographic effects.

1) Waveform analysis

Marsh and McKeon (1983) performed a waveform analysis outlined by Collins *et al.* (1981), which is based on a best fit of Chebyshev polynomials to the spectral curve. Absorption features are reflected in the variations of the polynomial coefficients. The authors plotted the ratios of two coefficients against wavelength to discriminate between different target materials.

2) Absorption feature depth

Another way to discover absorption features is to calculate the depth of potential spectral features and threshold it. However, this requires the definition of two absorption feature parameters, namely absorption wavelength position and depth. In general, the wavelength position of an absorption feature may be defined as the wavelength of maximum absorption, i.e. the wavelength of minimum reflectance value in each absorption feature. The absorption depth is the distance of the spectrum at the absorption feature position from a reference line characterising the absorption-free background signal. The latter reference line was chosen differently by several authors. For example, Kruse (1988) used the continuum defined by Clark and Roush (1984) to derive an absorption depth measure, while Okada and Iwashita (1992) made use of the upper convex hull as reference described by Green and Craig (1985). Rubin (1993) defined the band depth as the deviation from the spectrum average. The existence of an absorption channel may then be declared whenever the absorption depth exceeds a certain threshold. The threshold value is dependent on the data and the techniques applied and needs to be chosen give meaningful absorption features.

3) Derivative analysis

Generally, the derivative of a spectrum is its rate of change with respect to wavelength. A derivative spectrum may be used to emphasise changes, such as spectral absorption features, and suppress the mean level. With increasing order of differentiation, the more low frequency background signals will be suppressed, but simultaneously, the more high frequency noise will dominate the derivative spectrum. The latter results in an SNR decrease, which may hide absorption peaks instead of exposing them. In order to be able to perform repeated differentiation with experimental data, random noise needs to be reduced by some smoothing technique beforehand (Sonka *et al.*, 1993).

Demetriades-Shah *et al.* (1990) gave a general introduction to the use of derivative spectra in remote sensing, while Tsai and Philpot (1998) presented a detailed review about numerical derivative calculation methods and smoothing algorithms.

A spectral absorption feature may be considered as a valley contained between two edges. The edge positions of the spectral feature occur where the slope of the spectral curve has a local

extremum, i.e. where its first derivative is maximum (or minimum), or equivalently, its second derivative equals zero. The points of extreme slope are referred to as points of inflection.

Holden and LeDrew (1998, 1999) made use of the first- and second order derivatives to manually select bands ideal for remote identification of healthy and non-healthy corals using *in situ* spectroradiometer data. They plotted corresponding derivative spectra of the classes involved in one graph, and used opposing strong peaks of the first derivative, as well as opposite signs of the second derivative as indicators for potential class-discriminatory bands.

Piech and Piech (1987, 1989) used a Gaussian filter to smooth the spectral curve and computed the second derivative ('Laplacian' operator) of the smoothed curve ('Laplacian of a Gaussian'). This process was repeated for progressively increasing filter widths (standard deviation values of the Gaussian filter). The wavelength positions of the resulting points of inflection were then plotted against their corresponding filter width values, resulting in a 'scale space image' (Witkin, 1983). Since the number of zero crossings does not grow with increasing smoothing scale, the plot of inflection points within the scale space image results in a 'fingerprint'. The latter is then used to locate the points of inflection of a spectral curve and to rank the spectral features according to the estimated relative area contained between their inflection points.

Butler and Hopkins (1970) showed that second and higher order derivatives can be used to determine the peak positions of contributing absorption bands. Huguenin and Jones (1986) suggested three criteria to be met simultaneously for the detection of component absorption features:

- the 2^{nd} derivative must be negative,
- the 4^{th} derivative must be positive, and
- the 5^{th} derivative must equal zero.

Data normalisation is not necessary for the latter technique, since higher order derivatives are relatively insensitive to variations in illumination intensity (whether caused by sun angle, cloud cover or topography) and spectral variations of sunlight and skylight (Tsai and Philpot, 1998).

Absorption Feature Identification

After the detection of absorption features, the analyst may be interested in associating them individually with the classes of interest.

If sufficient ground knowledge is available, for example in form of detailed maps of the distribution of classes on the ground, an image display of a detected absorption band may highlight the spatial distribution of the target material the absorption feature belongs to. Alternatively, the ratio or difference of an absorption band and an adjacent absorption-free background channel may be used

to display the relative absorption depth (e.g. Marsh and McKeon, 1983; Feldman and Taranik, 1988). Up to three absorption features (or differences or ratios) may be compressed into a single colour-composite image (Kruse, 1988) in case multiple features belong to the same target class.

If reference spectra of potential target classes are available in a spectral library (from previous imaging spectrometer, laboratory or field data) image absorption spectra may be directly compared to the ones in the library. This may be realised visually by the analyst, or automatically via several matching techniques.

Since full spectral matching is not efficient with respect to the high spectral redundancy in imaging spectrometer data, codes may be used as simple representation of a pixel spectrum, allowing for fast automatic library searching and matching (Richards and Jia, 1999).

Mazer *et al.* (1988) introduced the binary encoding technique, which encodes a spectrum by setting each channel to 1 or 0, depending on whether its value falls above or below a pre-defined threshold. Usually, the threshold is chosen as the mean value of the spectrum. Each image spectrum corresponds to an integer value, which is compared to the ones of the library spectra via a bit-wise exclusive OR operator ('Hamming distance' measure). The number of bit-matches needed for identification may be adjusted to allow for target spectral variability and noise. However, the method does not consider the absorption depth, and therefore the significance of the spectral features.

Spectral feature fitting (SFF, Crowley *et al.*, 1989) is an alternative spectral matching technique, which fits in a least-squares sense continuum-removed image spectra to continuum-removed reference spectra. The identity of two spectra is measured by the root mean square error (RMSE) of the fit. If all image spectra are compared to one reference spectrum, the RSME image may then reveal the pixels very similar to the selected target class. The SFF technique requires both library and image spectra to be calibrated to reflectance.

If a reflectance calibration is not feasible, the spectral derivative ratio technique (Philpot, 1991) may be applied to compare image radiance spectra to library reflectance data. The method is based on the fact, that for target-specific wavelength ranges, the order of any-order derivative of the at-sensor radiance data at two wavelengths approximately equals the ratio of the same-order derivative of the spectral reflectance.

In summary, sensor bands may be placed over certain target-specific reflectance features in order to derive information about the target class of interest. If the target class is known and consists a single substance (e.g. minerals), spectral features may be found in the literature. If the target class is unknown or consists of a complex mixture of material components (e.g. vegetation canopy) or has no published reflectance curve, laboratory, field or image measurements may then be used to detect class-specific spectral features. In this case, the measurements should be calibrated to reflectance

first, or at least normalised, to compensate for the effects of both the atmosphere and the solar irradiance. This will enhance the target spectral features. The features may then be extracted via curve fitting methods or derivative analysis of the spectral curve. If the target is unknown, known material classes may be associated with the spectral features via display or matching techniques.

4.2.2 Solar Spectrum, Atmospheric Absorption and Scattering

In this thesis, the author limited himself to the optical solar-reflective wavelength range (0.3 to 2.5 μm) with the Sun as the only source of illumination. Figure 4.2 shows the extraterrestrial solar spectrum (Wehrli, 1985), which has an irradiance peak in the visible wavelength region (around 450 nm) and falls off rapidly, providing only little signal for the short-wave infrared wavelength region.

Atmospheric absorption and scattering processes lessen the energy of the reflected upwelling solar signal measured by the sensor. Unlike for ground measurements, the atmospheric effect is significant for aircraft-based measurements (Richards and Jia, 1999), and together with the nature of the solar spectrum, it needs to be taken into account for band selection in order to achieve an adequate image SNR. The convolution of the solar irradiance spectrum with the atmospheric transmission and the target reflectance represents the amount of signal that may be received by a sensor from the target surface. Usually, a sensor is configured to achieve a sufficiently high image SNR for all bands (see section 4.2.3).

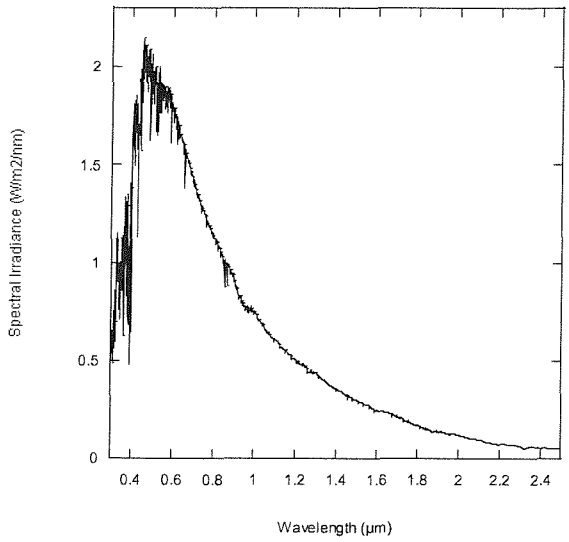


Figure 4.2: Extraterrestrial solar irradiance spectrum (Wehrli, 1985).

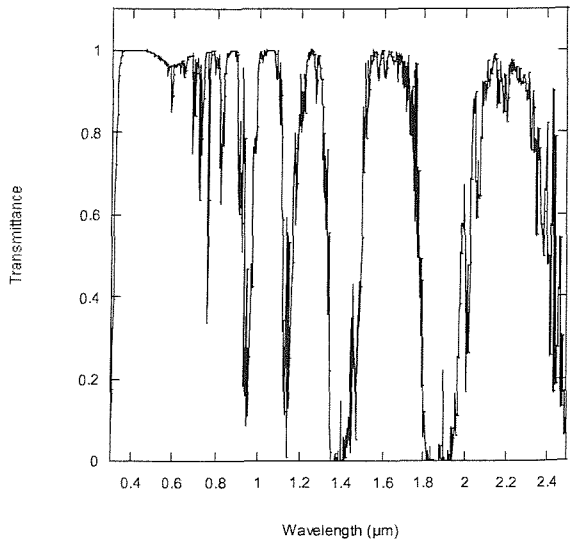


Figure 4.3: Total transmittance of the standard midlatitude summer atmosphere computed with the 6S code (Sun at zenith, 1.2 km aircraft altitude, no aerosols).

However, some bands with high levels of atmospheric scattering and absorption may be excluded prior to the band selection procedure. Wavelengths below 450 nm are substantially influenced by atmospheric scattering and absorption (Jensen, 2000). The total transmittance through a standard midlatitude summer atmosphere was approximated with the 'Second Simulation of the Satellite Signal in the Solar Spectrum', or '6S' code (Vermote *et al.*, 1997) and is displayed in figure 4.3.

The troughs in figure 4.3 represent absorption features of the main perturbing atmospheric gases in the solar-reflective region, that is, water vapour, ozone, oxygen and carbon dioxide. The wavelength centres of the major absorption bands are given in table 4.1. From these, water vapour bands clearly dominate the solar-reflective region from 0.7 µm onwards with respect to absorption depth and width.

Table 4.1: Absorption centre wavelengths of atmospheric gases for the solar-reflective wavelength range (Smith, 1985; Goetz *et al.*, 1995).

Atmospheric gas	Symbol	Absorption centre wavelengths (µm)
Ozone	O_3	0.32 - 0.36 and 0.44 - 0.74 (series of bands)
Molecular oxygen	O_2	0.5384, 0.7621, 1.07, 1.27
Water vapour	H_2O	a few weak bands in the visible, 0.72, 0.81, 0.942, 1.135, 1.379, 1.865
Carbon dioxide	CO_2	0.78 - 1.24 (a series of weak bands), 1.4, 1.6, 2.0

Table 4.2: Estimated absorption width of the major water vapour bands in the solar-reflective wavelength region.

Absorption centre wavelength (nm)	Band starting wavelength (nm)	Band ending wavelength (nm)	Absorption width (nm)
720	715	722.5	7.5
810	810	825	15
942	927.5	967.5	40
1135	1097.5	1170	72.5
1379	1340	1500	160
1865	1767.5	1992.5	225

Table 4.2 gives the widths of the water vapour bands as estimated from the data in figure 4.3. The width of an absorption feature was defined here by the difference between the wavelength of the starting and ending point. The starting point of the feature corresponds to the wavelength from where transmission begins to fall continuously until maximum absorption is reached. The ending point of the feature equals to the wavelength where approximately 100% transmission of the

atmosphere is reached again. Alternatives to the definition of absorption width are the full width at half maximum absorption depth (Kruse *et al.*, 1990), and the area of absorption between the hull quotient and unity divided by half of the absorption depth (Okada and Iwashita, 1992). However, the latter two descriptions require the determination of the absorption depth. All three definitions include subjective judgement of the analyst.

For terrestrial mapping purposes, so-called 'atmospheric windows' may be defined, which are wavelength regions where atmospheric gases only slightly absorb radiation. Wavebands outside a window are excluded from data collection or processing to avoid classification degradation from 'noisy' bands (Schowengerdt, 1997). According to the widths of major absorption bands (see figure 4.3 and table 4.2) atmospheric windows may be identified as the following wavelength intervals: 0.3 to 0.92 µm, 0.97 to 1.1 µm, 1.17 to 1.34 µm, 1.5 to 1.77 µm and 2.0 to 2.4 µm.

Experiments were performed with the 6S radiative transfer model to see whether the atmospheric windows inferred above remained unchanged for different amounts of water vapour (1.0 – 4.0 g/cm^2) in a US 1962 standard atmosphere. In addition, the effects of increasing solar zenith angle (up to 60°) and of different aircraft heights (0.8 – 1.6 km) for a midlatitude summer atmosphere were investigated. Figures C.1 to C.6 in appendix C show that the position and width of the absorption features stayed the same, while their absorption depths varied only slightly for the different scenarios. Consequently, the atmospheric windows as defined above were confirmed and accepted by the author.

The aerosol absorption effect was not considered here, since aerosols modify the solar spectrum over a large number of contiguous wavebands (Curran, 1994) and with less effect than certain atmospheric gases. Yet, bands may be selected to estimate atmospheric parameters (aerosol, water vapour or oxygen concentration) for atmospheric correction. CHRIS (mode 3) uses the Cimel CE 318™ sun photometer channels at 440, 670, 870 and 1020 nm to retrieve aerosol estimates (J. Settle, 2003, personal communication). The traditional blue (443 nm) and NIR (865 nm) bands of most satellite sensors designed for vegetation mapping (see section 4.2.1) may be employed to estimate the size distribution of aerosol particulates (for example, see ESA, 2002b). MODIS provides four channels centred at 905, 936, 940 and 1375 nm for atmospheric water vapour estimation, while Polder-2 offers two channels with centres at 763 and 910 nm to approximate oxygen and water vapour contents, respectively.

Very narrow atmospheric absorption features, e.g. the oxygen absorption band at 762 nm, may be used for the in-flight spectral calibration of hyperspectral sensors. A possible wavelength shift may be detected that arises mainly from temperature effects during aircraft operations.

4.2.3 Sensor Characteristics

The image SNR for every sensor band is a necessary component to consider for band selection as it measures the quality of each band with respect to noise. A high SNR is valuable for detecting minor but potentially important differences between class spectra and therefore securing high accuracy and reliability of the classification result. The image SNR depends on sensor-independent factors (e.g. solar irradiance, solar zenith angle, target reflectance, atmospheric state) and sensor-dependent ones (e.g. CCD response, sensor view angle and configuration parameters during data acquisition). Radiometric, spatial and spectral sensor characteristics are investigated in the following paragraphs with respect to their potential for increasing the sensor SNR.

Generally, the radiometric resolution is chosen by the system designers, and cannot be changed by the user of the system (e.g. for the CASI-2 1:4096 or 12 bit). A high SNR is normally assured by selecting the quantisation interval at least twice as large as the detector noise standard deviation (Schowengerdt, 1997).

A large GIFOV provides more signal to the sensor than a smaller one, increasing the resulting SNR. The size of a GIFOV of a pushbroom sensor, such as the CASI-2, is mainly dependent on the aircraft altitude, but also on the aircraft speed and integration time in along-track direction (ITRES, 2001). Consequently, a larger SNR may be achieved by flying higher, increasing the integration time or decreasing the aircraft speed. However, the quest for a high SNR needs to be compromised with feasible values for the latter parameters, and with the user's need for a specific swath width and GIFOV size to provide pure and sufficient training samples for his classification task (see section 3.3.1).

With respect to the spectral domain, a higher instrument SNR may be achieved by coarsening the spectral resolution, i.e. increasing the bandwidths. Some imaging spectrometers (such as the CASI-2) are programmable, allowing on-board summation of spectral channels. It can be shown that when the signals of C adjacent channels are being added together, the SNR of the resulting channel will increase by about $\sqrt{C}$ if signal and noise conditions are identical for all channels being summed. Apart from increasing the SNR, choosing a wider bandwidth tackles both data redundancy and banding noise, i.e. detector response changes from scan to scan.

However, the selection of a certain bandwidth should be keyed to the data application at hand. The precise location and width of a band may be chosen to optimally discriminate between user-defined classes. For example, GERIS bands were selected to provide maximum SNR while retaining sufficient spectral resolution to identify key minerals. The bandwidths reflected the widths of the corresponding mineral absorption features (Kruse *et al.*, 1990). In addition, care must be taken that all bands in the set obtain at least a certain percentage of the maximum achievable signal level (20% according to T. Wittebrood (ITRES), 2003, personal communication) to obtain an acceptable

SNR. For the CASI-2, aperture and integration time are usually adjusted by the instrument operator during data acquisition to produce a maximum achievable signal level of 80% to 90% for the band with the highest response.

Apart from the image SNR, the sensor's data recording rate and storage capacity limits need to be taken into account for band selection. For example, for the CASI-2, swath width may be traded against spectral resolution and number of bands in order to respect the data recording rate limits of the removable hard drive subsystem (see section 3.2.4). Similar applies for example to the CHRIS satellite sensor. That is, the more channels are chosen for recording, the fewer spatial pixels may be recorded.

4.2.4 Data Application

Some application techniques need certain wavebands in order to be applicable, while others may benefit from a band set that was specifically designed for them, resulting in a more accurate and reliable end product. In this section, the author shows examples of band requirements for the following land cover mapping techniques: vegetation indices, the red-edge index, and land cover classification.

Vegetation Indices

Vegetation indices (VIs) have been developed for vegetation detection and generally depend on the red-edge feature of the vegetation spectrum (sharp reflectance increase between 690 and 720 nm). Most broadband VIs are based on two bands sampling the bottom (reflectance minimum) and the top (reflectance maximum) of the red edge using the established red (665 nm) and NIR (865 nm) satellite sensor bands (see 4.2.1).

Table 4.3 presents the most common vegetation indices found in the literature and their input bands required. Some VIs in this table include also a blue band to correct for atmospheric effects (e.g. ARVI, TOAVI, EVI). The increased availability of hyperspectral data has led to the creation of narrow-band indices, such as the PRI using bands centred at 531 and 570 nm. Other examples include the DGVI which uses all sensor bands between 626 and 795 nm, or those which employ all sensor bands available (e.g. PCVI, GVI, TBVI, or OMBVI).

Table 4.3: List of most common vegetation indices with references and input band requirements.

Acronym	Name	Reference	Input Bands
RVI	Ratio VI	Jordan, 1969	Red, NIR (e.g. TM 3, 4)
DVI	Difference VI	Lillesand and Kiefer, 1987	
WDVI	Weighted Difference VI	Clevers, 1988	
NDVI	Normalised Difference VI	Tucker, 1979	
IPVI	Infrared Percentage VI	Crippen, 1990	
PVI	Perpendicular VI	Richardson and Wiegand, 1977	
SAVI	Soil-Adjusted VI	Huete, 1988	
MSAVI, MSAVI2	Modified SAVI	Qi *et al.*, 1994 Rondeaux *et al.*, 1996 Lyon *et al.*, 1998	
TSAVI	Transformed SAVI	Baret *et al.*, 1989	
RSVI	Red-edge Vegetation Stress Index	Merton, 1998	
GEMI	Global Environment Monitoring Index	Pinty and Verstraete, 1992	
PRI	Photochemical Reflectance Index	Gamon *et al.*, 1997	531 and 570 nm
MIVI	Middle Infrared based VI	Thenkabail *et al.*, 1995	Red, MIR (e.g. TM 3, 5)
ARVI	Atmospherically Resistant VI	Kaufman and Tanré, 1992	Blue, Red, NIR (e.g. TM 1, 3, 4)
TOAVI	Top-Of-Atmosphere VI	ESA, 2002a	
EVI	Enhanced VI	Huete *et al.*, 2002	
PCVI	Principal Component VI	Jensen, 1986	All bands (e.g. TM 1,2,3,4,5,7)
GVI	Green VI	Jackson, 1983 Crist and Cicone, 1984	
DGVI	First/second Derivative VI	Elvidge and Chen, 1995	All bands from 626 to 795 nm
TBVI	Two-Band VI	Thenkabail *et al.*, 2002	All bands
OMBVI	Optimum Multiple-Band VI		

Red-edge index

The red-edge index is commonly defined as the point of maximum slope of the red-edge feature and has been shown to be mainly related to biophysical canopy parameters such as the leaf-area index (LAI) or leaf chlorophyll content (Clevers and Büker, 1991). A high and fine sampling of the red-edge between 670 to 780 nm is necessary for an accurate index determination with derivative-based methods (Horler *et al.*, 1983; Demetriades-Shah *et al.*, 1990) or curve-fitting techniques (Bonham-Carter, 1988; Clevers and Büker, 1991). For more sparsely sampled data, Dawson and Curran (1998) suggested a three-point Lagrangian interpolation technique for available red-edge bands, while Guyot and Baret (1988) proposed the use of reflectance bands centred at 670, 700, 740 and 780 nm to approximate the red edge via linear interpolation.

Land cover classification

In contrast to vegetation or red-edge indices, classification is intended to map various land cover classes, and not simply vegetation and non-vegetation. Complex interactions between multiple target class spectra may occur, excluding the instant selection of class-specific spectral features. For example, some classes may represent different types of the same material and share most of the physically-based spectral features. As a consequence, in the absence of clear class-diagnostic spectral features, an empirical band selection procedure aimed to maximise the final classification accuracy may be advisable. Mather (1999, page 33) specified the three band set parameters to be optimised in such a procedure: "The position in the spectrum, width and number of spectral bands will determine the degree to which individual targets (vegetation species, crop or rock types) can be discriminated".

Generally one may distinguish between band selection methods for classification that are based on test statistics, eigenanalysis and distance measures. In all methods, the classification accuracy of a band set is approximated by measures of class separability.

1) Band selection based on test statistics

Some authors have used test statistics to see whether class spectra are significantly different within certain wavelength regions. Thomson *et al.* (1998a) made use of the Student's t statistic as band index of dissimilarity between the reflectance responses of several intertidal surface types. The entire VNIR spectral range under investigation was found to be useful for class discrimination after considering the t spectra of all class pairs. Leckie *et al.* (1988) employed the t - and Mann-Whitney hypothesis tests to find bands that significantly separate between different levels of defoliation caused by the spruce budworm (1 and 5% significance level). However, a hypothesis test cannot prioritise between bands found to be significant and, consequently, cannot indicate how well each band set performed. In addition, the above test statistics do not exploit the differences of second-order statistics.

In contrast, Kamp *et al.* (1997) calculated the absolute difference between the double standard deviations (2σ) of the class spectra of interest, which account individually for 95% of all sample spectra belonging to one class. An absolute difference of greater than zero indicates that the classes can be discriminated to 95% on the basis of their second-order statistics alone. For the multiclass case, all spectral regions derived from the two-class differences are condensed. Obviously, the selected band set may be used only with classification methods based on second-order statistics, such as MLC.

Nevertheless, the test statistics presented above are single-band measures and do not allow for the quantification of class separation performance of a specified band set.

2) Band selection based on eigenanalysis

Section 2.3.5 presented methods of band selection using the results of eigenanalysis. In the context of supervised band selection, the supervised DA criteria and the unsupervised PCA may be employed to extract features that discriminate most between the predefined classes.

The two approaches based on eigenvector loadings, discriminant power (Tu *et al.*, 1998) and MSPCA (Csillag *et al.*, 1993), are preferred to the methods using eigenvector axis crossings, as they incorporate the contribution of each band towards the eigenvector.

3) Band selection based on distance measures

Some authors have employed feature selection methods (section 2.2) for band selection. Both supervised heuristic and probabilistic distance measures were used as surrogates for MLC accuracy. Richter and Lehmann (1989) guided the selection of the MOMS-02 spectral bands by identifying high difference regions between the mean (field) spectra of seven common surface types. They used the Euclidean distance measure to quantify the separability between the given classes. Mausel *et al.* (1990) used the Transformed Divergence, Jeffries-Matusita, Bhattacharyya Distance and the Divergence measures for band selection for MLC of agricultural area. In addition a PCA was performed on all channels, with the highest eigenvector loadings (over 50%) indicating the bands most likely to discriminate between classes of interest. However, the band selection approaches discussed so far do not account for a bandwidth increase within the algorithm.

Petrie *et al.* (1998) introduced a band selection algorithm that incorporates a band expansion routine. First, the best available band in terms of maximum divergence is found. Second, the best band is enlarged by adding neighbouring bands as long as the divergence measure increases. In a forward feature selection manner, the latter two steps are repeated until a desired number of bands plus two has been reached or all bands have been assigned. Finally, the worst two bands in terms of divergence are discarded. The proposed band selection algorithm is mainly based on a single-band measure with the band set performance only being considered when the worst two bands are deleted. In addition, the band expansion routine was not specified in great detail by the authors. For example, rules about the direction of band merger were not described. Furthermore, bands were first selected and then expanded, assuming that the expanded best band will equal to the best expanded band.

Similarly, Withagen *et al.* (2001) used a forward feature selection algorithm together with the Mahalanobis distance as criterion function. In contrast to Petrie *et al.* (1998), the authors chose and broadened the best available band on the basis of the band set performance. A band expansion constraint is that the resulting band needs to be centred on the original narrow band. The merged band is calculated as the simple average of the intensity values of the bands involved. The authors

assumed that the expanded best band will equal to the best expanded band. Although the band selection procedure was class- and data-specific, the resulting band set was recommended to be generally optimal for the detection of military vehicles and land mines in a natural background.

In the following section, an empirical band selection algorithm similar to one proposed by Withagen *et al.* (2001) was developed to define optimum band set parameters (band number, location and width) with respect to the accuracy of the actual classification task at hand. But unlike the approaches of Petrie *et al.* (1998) and Withagen *et al.* (2001), the method uses a saturated distance measure instead of a non-saturated one as a more representative surrogate of the MLC accuracy. In addition, band expansion was designed to take place before, not after, individual bands are selected. This is because an increase in width changes the statistics of a band. What is more, a new band number determination routine is employed to give estimates of intrinsic dimensionality. And finally, SNR issues are taken into account, which prevents the method from choosing bands that result in an unacceptable signal level.

The new method recognises the need for band selection to be data-, class- and application-specific. For the method to take also the sensor's spatial and spectral characteristics into account, an example employment is described for operational use with the CASI-2. It is based on the sensor's capability to be programmed in spatial and spectral modes. The algorithm to be presented is computationally efficient, allows pre-specified bands to be incorporated into or omitted from the final band set.

4.3 Supervised Band Selection Algorithm

The aim of this section is to find a feature selection method from the ones presented in section 2.2 that is best suited to band selection for MLC and to extend its capacity to bandwidth increase and band number determination. Feature selection is the process of selecting a band subset out of the original band set, using a criterion to rank features and a technique to search for the optimal band subset in a systematic way.

The design phase of the algorithm is presented in the following paragraphs and considers the reliability, discriminability and computational cost of the algorithm. While the reliability assures that the algorithm is based on accurate assumptions, the discriminability guarantees that the algorithm routine chosen results in a band set with the highest class-discriminatory power.

Computational feasibility and speed are important factors in the design phase, as the aim is to provide a band selection procedure that is computationally efficient. All algorithms were programmed in IDL (version 5.5 Win 32 x86) within the ENVI (version 3.5) processing environment and were run

in Windows NT on a 1 GHz Intel Pentium III processor with 256 MB RAM. Any quantifications of computational speed made in this section refer to these specifications.

The two hyperspectral data sets used to evaluate the band selection algorithm were introduced in the pilot study (chapter 3) as the New Forest (HyMAP) and the River Severn (CASI) data sets.

4.3.1 Algorithm Input and Output

The band selection algorithm presented in this chapter requires hyperspectral image data recorded over the scene or parts of it that include all the informative classes of interest to the user. As described in section 4.2, band selection should also take characteristics of the Sun, the atmosphere and the sensor into account. These data attributes would be considered if hyperspectral data were acquired under the same or similar conditions as the target data to be flown. Otherwise these conditions need to be simulated (see section 6.6).

A programmable imaging spectrometer such as the CASI-2 would allow using the same sensor for hyperspectral and final data recording. With the CASI-2's enhanced spectral mode (see chapter 1), 288 bands may be obtained at the cost of a reduced swath width, while the spatial mode records the full image swath with a reduced set of up to 19 bands. A practical way for data collection would be to acquire first an image with a reduced FOV in hyperspectral mode over a representative part of the scene, which is then input into the band selection algorithm. The hyperspectral image should represent all target classes of interest with enough training samples to achieve adequate class statistics for the given number of bands. After that, the CASI-2 is programmed in the selected band set to record the full-swath multispectral data over the same scene. Both the hyperspectral and multispectral data acquisitions are to be performed with identical spatial resolution to ensure that similar class and data statistics result. The time between the two image flights should be minimised to guarantee that target class spectral characteristics and atmospheric and illumination conditions have not changed. For the remainder of this chapter, this image acquisition technique was assumed by the author.

Apart from the hyperspectral image, the band selection algorithm expects regions of interest (ROIs), depicting class training areas defined by the analyst (supervised version), and their statistics (the class mean vectors and covariance matrices) as input. The algorithm expects that all noisy and faulty bands were removed from the original data set by the analyst.

The algorithm outputs a list of the resulting band set parameters for each band, including the sensor channels to merge for a particular band, the (new) band centre and full-width-half-maximum (FWHM). If the number of bands in the set was fixed by the user, the algorithm may result in a

smaller band set if this target was not feasible, or may recommend a smaller number of bands according to the results of the band number determination procedure.

4.3.2 Choice of Criterion Function

The optimal feature selection criterion is the application accuracy, in this case the MLC overall accuracy. As stated in section 3.3.1, the 'leave-one-out' cross-validation method is best suited to generating the contingency table for feature selection with hyperspectral data, as it makes maximum use of the *in situ* data for defining the class statistics. As it constructs a classifier for each class training sample, the calculation speed is a function of the number of training samples. As the latter number ought to be relatively high for hyperspectral data sets to circumvent the Hughes phenomenon, the MLC accuracy estimate using the 'leave-one-out' method may become computationally very expensive.

As a consequence, a surrogate measure for the MLC accuracy estimate is needed and possible candidates include the supervised distance measures reviewed in section 2.2.2, or the discriminant power measure determined from eigenanalysis via DAFE and presented in section 2.3.5 (equation 2.38). All of these measures take first- and/or second-order class statistics into account. As the performance of the supervised feature selection criteria depends strongly on the accuracy of the class statistics, the most appropriate criterion was selected on a purely theoretical basis.
In order to represent correctly the accuracy estimate of the probabilistic ML classifier, a probabilistic distance measure, which also makes use of the class covariance matrix (see section 2.2.2), should be preferred to a heuristic one, only employing the class mean vector. In addition, saturated measures are more likely to be justified than unsaturated ones for this algorithm, as the entire distance range will be evaluated.

Regarding the multi-class form of the distance measure, the 'maximum average' was favoured to the 'maximum minimum' type, as the latter may introduce a bias by focussing only on the class pair hardest to separate. Thus, from a theoretical point-of-view, either the Transformed Divergence or the Jeffries-Matusita distance measure should be employed as feature selection criterion.

An experimental analysis was performed on the two data sets available for this study (New Forest and River Severn) to approve the theoretical choice of criterion. The value for some of the surrogate measures from section 2.2.2 was calculated for both multiple-band (table 2.2) and single-band versions (table 4.4) and linearly regressed against the corresponding MLC overall accuracy estimated with the leave-one-out method.

Table 4.4: Probabilistic distance measures for one band and normally distributed classes ω_k and ω_o, using the class mean and variance values m and s^2, respectively. The equations were derived from their corresponding counterparts in table 2.2.

Measure	Parametric Form
Bhattacharyya	$\mathrm{d_b}(k,o) = \frac{(m_o - m_k)^2}{4(s_{oo}^2 + s_{kk}^2)} + 0.5 \ln\left(\frac{s_{oo}^2 + s_{kk}^2}{2 s_{oo} s_{kk}}\right)$
Jeffries-Matusita	$\mathrm{d_{jm}}(k,o) = \sqrt{2\left[1 - \mathrm{e}^{-\mathrm{d_b}(k,o)}\right]}$
Divergence	$\mathrm{d_d}(k,o) = 0.5(m_o - m_k)^2\left(\frac{1}{s_{oo}^2} + \frac{1}{s_{kk}^2}\right) + 0.5\left(\frac{s_{oo}^2}{s_{kk}^2} + \frac{s_{kk}^2}{s_{oo}^2} - 2\right)$
Transformed Divergence	$\mathrm{d_{td}}(k,o) = a\left[1 - \mathrm{e}^{-\frac{\mathrm{d_d}(k,o)}{b}}\right]$ a, b selected saturation and range values, e.g. 2000 and 8, respectively (Card and Angelici, 1983)
Mahalanobis	$\mathrm{d_m}(k,o) = (m_o - m_k)^2\left(\frac{1}{s_{oo}^2} + \frac{1}{s_{kk}^2}\right)$
Patrick-Fischer	$\mathrm{d_{pf}}(k,o) = (2\pi)^{-\frac{D}{2}}\left[\frac{1}{\sqrt{2}s_{oo}} + \frac{1}{\sqrt{2}s_{kk}} - \frac{2}{\sqrt{s_{oo}^2 + s_{kk}^2}} \times \mathrm{e}^{\frac{(m_o - m_k)^2}{2(s_{oo}^2 + s_{kk}^2)}}\right]$

Tables C.1 and C.2 in the appendix show the regression results and the correlation coefficient for single-band and multiple-band surrogate measures, respectively, for both the New Forest and River Severn data sets. For the multiple-band measures, only two-band combinations with the best and the worst bands from the single-band results were evaluated in order to cover most of the value range of the MLC accuracy estimate. The regressions show that both the Jeffries-Matusita distance and Transformed Divergence measure were most correlated to the MLC accuracy estimate regarding all other distance measures (see table 4.5) and therefore confirm the theoretical choice.

Table 4.5: Correlation coefficient between the single-/multiple-band distance measures and the MLC overall accuracy estimated with the leave-one-out method for the New Forest and River Severn data set.

	Single-Band New Forest	Multiple-Band New Forest	Single-Band River Severn	Multiple-Band River Severn
Bhattacharyya	0.562	0.711	0.961	0.566
Jeffries-Matusita	**0.889**	0.971	**0.981**	0.858
Divergence	0.415	0.175	0.944	0.563
Transformed Divergence	0.876	**0.978**	0.977	**0.907**
Mahalanobis	0.418	0.183	0.951	0.554
Euclidean	0.832	0.639	0.451	0.418
City-Block	0.832	0.704	0.451	0.168
Chebyschev	0.832	0.867	0.451	0.572
Canberra	0.272	0.523	0.702	0.193
Discriminant Power (DAFE)	0.331	-	0.341	-

In addition, the same experimental analysis was performed for the discriminant power (DP, single-band) measure derived from the DAFE feature extraction method. The result showed a very poor correlation between the DP measure and MLC accuracy for both data sets. As multiple-band measures were needed, the DP measure was not considered any further.

According to these theoretical and experimental results, the author chose the Transformed Divergence multi-class distance measure as surrogate measure for the MLC accuracy estimate, as it performs slightly better than the Jeffries-Matusita distance for the (more frequent) multiple-band case.

4.3.3 Choice of Search Method

The optimal search method is an exhaustive one. However, when the final band set is to include, for example, 19 bands (the maximum band set size for the CASI-2 spatial mode), the number of possible band set combinations amounts to more than 3.46×10^{21} and 2.04×10^{15} combinations for the New Forest and the River Severn data set, respectively (see equation 2.25). The computation time of an exhaustive search with the MLC accuracy estimate (leave-one-out method using all samples) as criterion would total about 4.3×10^{15} years for the New Forest data, while using the Transformed Divergence as criterion function would still need more than 1.2×10^{12} years of computation time. Both cases are computationally impossible tasks (see also table 3.6 in chapter 3).

The Branch-and-bound algorithm is classified as an optimal method under the assumption that the criterion function always increases monotonically with addition of new features to the set. However, the Transformed Divergence function may drop with increasing number of bands for hyperspectral

data sets, when the ratio of the number of training pixels to the number of spectral bands is not high enough to ensure reliable class statistics (Hughes phenomenon). Consequently, this method was not pursued any further.

Among the sub-optimal search methods, the 'best individual' method only delivers acceptable results when the bands within the original set are uncorrelated, which is an unusual case for hyperspectral data and therefore not applicable.

The 'sequential forward selection' (SFS, 'bottom-up') procedure results in less computational effort than its backward ('top-down') counterpart for most band selection tasks, as the number of features to be selected for the final set (e.g. less than 20 for CASI-2 spatial mode) is much smaller than the number of original bands in the hyperspectral data set (e.g. up to 288 bands for the CASI-2 spectral mode). As the SFS successively adds features to the set, the bands may be removed in reverse order, allowing for flexibility in the dimensionality of the final band set. Table 3.6 in chapter 3 gives an example of the computation time involved with the SFS algorithm.

The 'Max-Min' search method focuses on the optimization of the minimum criterion function value that the new feature achieves with each of the pre-selected features individually. While the computational cost is reduced as only a two-dimensional evaluation of the criterion is needed, the performance of the overall set, and therefore the redundancy within the set, is neglected.

Suboptimal search methods such as beam search, genetic algorithms, and simulated annealing were not considered for band selection. Although they search the entire feature space by randomly altering some of the bands within the set to escape locally optimal solutions, they do not select features in order, and therefore do not allow the reduction of the dimensionality of final band set. The latter three methods are also more costly to implement.
According to these theoretical results, the author chose the SFS as the search algorithm for feature selection based band selection.

4.3.4 Bandwidth Increase

As mentioned earlier, a bandwidth increase of sensor bands is beneficial

- to reduce the amount of data to record and allow for more spatial pixels (wider swath width),
- to suppress instrumental spectral noise increasing the signal-to-noise ratio,
- to decrease the integration time and thus refine along-track spatial resolution, and
- to reduce the effect of the spectral overlap among adjacent pixels on the total band signal.

The aim of this subroutine is to merge two neighbouring bands that are redundant with respect to class discrimination. Physically, the increase in bandwidth is performed as a simple sum of the

signals (or values) of two or more bands. This simulates the CASI-2 on-board and in-memory summation of individual bands before data recording. In case two neighbouring bands are to be merged, the class statistics for the resulting band need to be recalculated for criterion function the merger is based on. Since a computation based on the training data itself is very costly, the following formulae based on simple additions and multiplications are applied.

The mean of a new band merged from bands x and y is calculated using equation 4.1.

$$m_{x+y} = m_x + m_y \tag{4.1}$$

The variance of the new band may be computed as (Weisstein, 2003):

$$s^2_{x+y} = s^2_x + s^2_y + 2s_{x,y} \tag{4.2}$$

The covariance between the new band, merged from bands x and y, and another band within the set, for example z, can be computed as (Weisstein, 2003):

$$s_{x+y,z} = s_{x,z} + s_{y,z} \tag{4.3}$$

If the band z was itself a band merged from bands s and t equation 4.3 could be expanded using the symmetry property of the covariance:

$$s_{x+y,s+t} = s_{x,s+t} + s_{y,s+t} = s_{s+t,x} + s_{s+t,y} = s_{s,x} + s_{t,x} + s_{s,y} + s_{t,y} \tag{4.4}$$

If in equation 4.3 another band w was merged with band $x+y$ the new band covariance could be computed as follows:

$$s_{(x+y)+w,z} = s_{x+y,z} + s_{w,z} = s_{x,z} + s_{y,z} + s_{w,z} \tag{4.5}$$

The centre wavelength and the FWHM of the merged band are calculated with equations 4.6 and 4.7.

$$\lambda_{new} = \frac{\lambda_{startrow} + \lambda_{endrow}}{2} \tag{4.6}$$

$$FWHM_{new} = \lambda_{endrow} - \lambda_{startrow} + \frac{FWHM_{startrow} + FWHM_{endrow}}{2} \tag{4.7}$$

where

λ_{new}, $FWHM_{new}$	Centre wavelength and FWHM of the new merged band
$\lambda_{startrow}$, $FWHM_{startrow}$	Centre wavelength and FWHM of the start row of the merger
λ_{endrow}, $FWHM_{endrow}$	Centre wavelength and FWHM of the end row of the merger

The criterion function for bandwidth increase may be based on the increase in classification performance of either the single band or the entire set after band merger, using the classification accuracy measure or a surrogate, such as the Transformed Divergence. Petrie *et al.* (1998) and Csillag *et al.* (1993) employed the single-band measures as the basis for merging adjacent bands,

while Warner *et al.* (1999) and Withagen *et al.* (2001) took the set performance into account. Here, the percentage increase in Transformed Divergence for the entire band set was preferred as criterion function to the corresponding individual band criterion. The ultimate aim of the band selection process is to apply the selected band set in its entirety for MLC, rather than extracting individual bands from it for single-band use.

The bandwidth increase and ranking routines are applied alternately in the procedure to guarantee that maximum band expansion is possible at any given set dimension. In contrast to the band selection procedures of Petrie *et al.* (1998) and Withagen *et al.* (2001), bandwidth increase in this algorithm takes place before a new centre band location is found. Bandwidth increase alters the performance of the band set, and all bands are expanded independently and compared first, before the best expanded band is chosen. In addition, adjacent narrow bands chosen for the final band set should not be merged by default, as they may, for example, exploit the top and bottom of a narrow spectral feature.

With respect to the direction of band broadening, both mergers of a band with its left and right neighbours are compared in terms of overall set performance, and the merger leading to the better performance is carried out. The process is repeated until both left and right mergers decrease the overall set accuracy. This procedure assumes that a band expansion in a direction of performance decrease will lead to added deterioration in the discrimination ability of the set.

The user is presented with three algorithm options:

1) to force all bands to be merged with an equal number of neighbours, where the number is determined by the user,
2) to force all bands to be merged with an equal number of neighbours, where the number is determined by the algorithm,
3) to allow bands to be merged with an unequal number of neighbours, where the number is determined by the algorithm.

Option 1 and 2 allow the user to choose equal bandwidths and even to fix the bandwidth in option 1 setting in order to gain further control over the resulting band set. For example, an equal bandwidth across all bands may be useful if further data are to be acquired in CASI-2's enhanced spectral mode (see chapter 1).

In option 3 bands are allowed to expand independently, allowing for the most optimal band set to emerge. However, to ensure that all bands in the set obtain a sufficiently high signal level (at least about 20%, see section 4.2.3), a constraint to algorithm 3 was introduced. A band may not be increased further in width if the smallest of all band means in the set becomes a user-defined percentage (lower limit 20%) smaller than its mean. If an expanded band does not achieve a sufficiently high signal level, it is not considered for the final band set. This percentage constraint

does not apply to the first band as it may not be compared with subsequently selected bands. If the signal level of the first band is too high and the user-defined number of bands cannot be reached, its width is reduced by one row in an iterative fashion until the desired number of bands may be obtained for the set.

4.3.5 Band Number Determination

In general, classification accuracy is a saturating function of the number of input bands, and beyond a certain threshold the increase in the number of bands does not produce any significant improvement in classification accuracy. This section aims to determine this threshold, i.e. the optimum band number.

Clearly, the upper limit of the number of bands is dictated by the sensor type and its operational mode. For example, for the acquisition of a CASI-2 image data in spatial mode, a maximum of 19 out of 288 bands may be selected. With respect to the maximum achievable classification accuracy, two cases may then be distinguished. The best 19-band subset may result in significantly less classification accuracy than the complete band set does, and all 19 bands will be required to achieve the best possible result. Or, the accuracy values achieved with the 19-band and the full band set may not differ significantly from each other, as additional dimensions may not add any extra discriminatory power to the band set.

In the latter case, the question may be asked whether even 19 bands are indeed necessary to achieve the maximum classification accuracy of the full band set, or whether fewer bands may produce results of similar accuracy. This section aims to provide an answer to this problem, which is of particular importance when dealing with highly redundant imaging spectrometry data. Hughes phenomenon may occur and the optimum classification accuracy may be obtained with much less than 19 bands.

In chapters 2 and 3, intrinsic dimensionality (ID) estimation methods were discussed and applied, respectively. As the classification was chosen as final data application, only methods for the determination of the supervised ID (the intrinsic discriminant dimensionality, IDD) are considered here. In section 2.4, the IDD was defined as the minimum number of features needed to obtain the same classification accuracy as could be obtained in the original feature space.

As original bands, and not transformed features, are to be selected by the algorithm, IDD methods based on eigenanalysis are less appropriate. Consequently, the aim is to find an upper bound to the real IDD estimate which may be defined as the minimum number of bands needed to obtain the same classification accuracy as could be obtained in the original feature space. The constraint that the IDD estimate should not exceed the maximum number of linearly independent features (that is,

$M-1$, where M is the number of classes, see section 2.4), is not valid for band selection, as bands are in general highly correlated features.

All other non-eigenanalysis-based IDD methods were investigated if they could be used to determine the optimal band number. The statistical methods that test whether the difference between the classification accuracy of subsequent band subsets is significant or not, require knowledge of the distribution of the classification accuracy measure (see section 2.4.2). However, in this band selection algorithm, the classification accuracy was approximated by the Transformed Divergence measure, the statistical distribution of which is unknown. The latter may be estimated under high computational cost via density estimation methods, such as Monte Carlo simulations, assuming a normal and symmetric distribution (Bressan and Vitrià, 2002). The latter was not attempted in this study due to computational reasons, as for each band set dimension, a density estimation needed to be performed.

The band number determination technique should satisfy the following needs of the band selection algorithm:

- to find the optimal band set dimension which achieves maximum classification accuracy,
- to consider Hughes phenomenon, and
- to allow for statistically dependent bands and a sub-optimal best band set.

The latter requirement reflects the correlated nature of the bands and the sub-optimality of the search procedure in the feature selection process.

Inspired by the heuristic ID approximation method 'Proportion of total variance' (section 2.4.1), the author defined the 'proportion of maximum achievable Transformed Divergence', PMATD, which is calculated by dividing the TD value of a band subset by the maximum achievable TD value of the entire band set. The PMATD is based on the linear relationship between the TD and the MLC accuracy, that is, the PMATD approximates the proportion of maximum achievable MLC accuracy (PMAMA). Then, the optimal number of bands equals to the dimension of the smallest band set, which achieves a certain PMATD, for example 95% or 99%.

Assuming that the TD measure is monotonically increasing with growing number of bands, the maximum achievable TD value is reached using all bands of the set. However, Hughes phenomenon may occur and the maximum achievable TD value may be reached by a band subset of much smaller dimension. Therefore, the algorithm compares the TD values of the best subsets of smaller dimension than the 'allowable' number of dimensions (e.g. 19 bands for the CASI-2 spatial mode, or user-defined) with the TD performance of the full set, and the highest value is chosen as the maximum achievable TD value.

If any subset of dimension smaller than or equal to the maximum allowable number of bands (19 for CASI-2 spatial mode) does not achieve 95% or 99% of the maximum TD value, the maximum allowable number will be chosen as the optimal band number.

This subroutine gives inaccurate proportions if the maximum achievable TD value occurs for a dimension between the 'allowable' and the total number of bands. However, for most hyperspectral data sets saturation occurs at a low number of dimensions, and choosing the 'allowable' number of bands high enough (e.g. between 10 and 20) may circumvent this problem. In addition, supplying an adequate number of training samples (see section 3.3.1) may prevent the event of Hughes phenomenon altogether.

An advantage of this routine is that it is application-oriented, using an approximation of the proportion of maximum achievable MLC accuracy. It is not based on absolute values of TD, as the latter depend on the data and the definition of training sites and classes. The error of this technique is controlled, as a certain percentage of TD, and therefore classification accuracy, will be guaranteed, if the number of maximum allowable features is high enough and all assumptions made earlier apply.

4.3.6 Algorithm Implementation

A computer program SBS (Supervised Band Selection) has been written using this algorithm and is described in appendix G. A flowchart of the algorithm is presented in figure 4.4, which is based on the routines described above.

The algorithm has been written in IDL (version 5.5 Win 32 x86) and ENVI (version 3.5), and reads the image file, the class ROIs, and class statistics. The user has the option to select certain bands to be either used in the final band set or excluded from it (e.g. noisy bands, bands outside atmospheric windows, absorption bands, bands from a specific wavelength range).

In general, the SBS program is computationally highly efficient. The total execution time amounts to about 40 seconds if 19 optimal bands are to be selected from the 117 bands of the New Forest data set. The computation time mainly depends on the number of classes and output bands specified by the user. If an output image was to be created from the optimal band set, the running time would be significantly increased. The latter is dependent on the size of the data set, but also on the number and width of the output bands. Creating an output image of the first three optimal bands of the New Forest data set in the above example adds 25 seconds to the total computation time (5 bands to read).

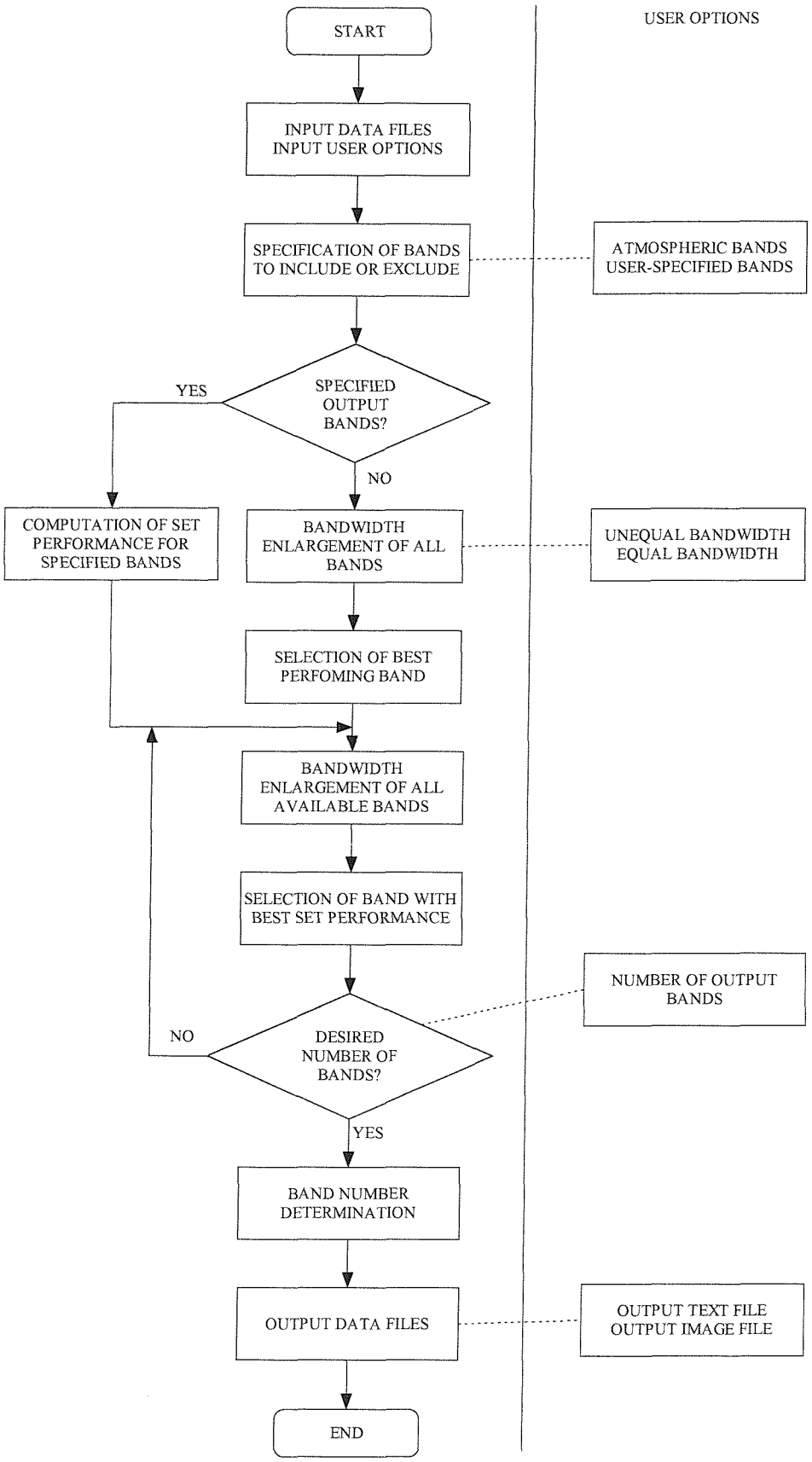


Figure 4.4: Data flowchart of the SBS (Supervised Band Selection) program with algorithm options displayed on the right side.

4.4 Algorithm Evaluation

This section aims to assess the value of the proposed algorithm and its results qualitatively and quantitatively with respect to its reliability, consistency, and effectiveness.

4.4.1 Qualitative Evaluation

The first six 'optimal' SBS bands derived for the New Forest and the River Severn data sets and their corresponding class scheme are listed in table 4.6 and plotted against a vegetation spectrum in figures 4.5 and 4.6, respectively. The band sets were selected with the constraint that the minimum signal mean was equal to at least 30% of the maximum band mean.

For the New Forest band set, the best six bands chosen by the algorithm stem from the NIR and SWIR wavelength range. The first band is located on the NIR plateau, and corresponds to the traditional vegetation band centred on 865 nm. Bands 3 and 6 are two adjacent bands positioned at the upper end of the red-edge feature. The latter three bands are sensitive to the structure of the vegetation canopy and the percentage soil covered. Bands 2 and 4 are centred on 1580 nm and 2110 nm, respectively, and are narrow-band equivalents to Landsat ETM+ bands 5 (1550 - 1750 nm) and 7 (2090 - 2350 nm), which are both sensitive to vegetation moisture content. Band 5, placed at 1330 nm between the two water absorption features at 1135 and 1379 nm, is sensitive to canopy structure and water content. As most of the image consists of vegetation classes (grassland, bracken, valley mire, dry, humid and wet heath) that differ more in their canopy structure and moisture content than in their chlorophyll content, the band selection result is accepted as a sensible choice.

For the River Severn band set, the blue CASI bands were excluded from band selection as they were perceived as noisy (see section 3.2.1). The first band chosen by the algorithm is a relatively broad band situated on the green slope near the red chlorophyll absorption peak. The wavelength region around the red band is generally known to potentially differentiate between different vegetation types. The second best band is found on the NIR plateau, which is sensitive to canopy structure and helps to delineate the land-water interface. The third and fourth band sample the bottom and top end of the red-edge feature and are both responsive to chlorophyll variations in the canopy. Band 5 and 6 further sample the green slope and the high NIR response of vegetation, respectively. Salt-marshes are highly complex mixtures of different intertidal vegetation types (here high, mid, and pioneer marsh classes), bare rock and mud. The salt-marsh vegetation classes are distinguishable from each other and the Bare Rock and Mud classes principally by their canopy structure and chlorophyll content, which is reflected in the choice of the bands.

Table 4.6: First 6 optimal bands selected by the SBS algorithm for the New Forest and River Severn data set with respect to their corresponding classification task (see table 3.9 and 3.8).

Band number	New Forest				River Severn			
	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]
1	867.4	31.25	851.775	883.025	639.9	31.5	624.15	655.65
2	1583.5	16.2	1575.4	1591.6	910.9	8.8	906.5	915.3
3	738.1	15.7	730.25	745.95	704.7	8.8	700.3	709.1
4	2109.75	38.15	2090.675	2128.825	758.2	8.8	753.8	762.6
5	1330.2	29.9	1315.25	1345.15	594.45	16.1	586.4	602.5
6	760.95	30.4	745.75	776.15	796.4	8.8	792	800.8

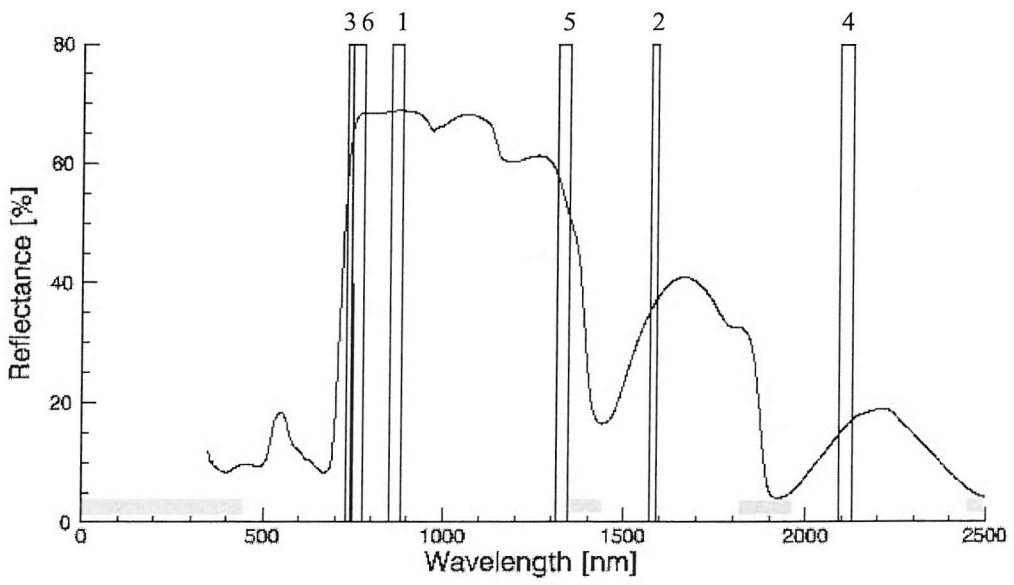


Figure 4.5: First six bands selected by SBS for the New Forest classification task (see table 3.9 and 4.6). Wavelengths not available for band selection are indicated by grey bars.

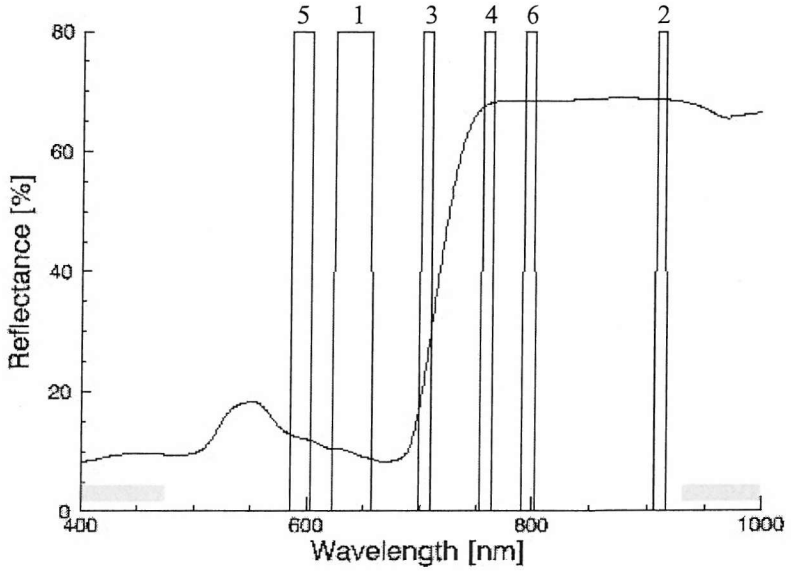


Figure 4.6: First six bands selected by SBS for the River Severn classification task (see table 3.8 and 4.6). Wavelengths not available for band selection are indicated by grey bars.

When generating colour composites from the best band triplet, six RGB combinations are possible, and the question arises which RGB colour should be assigned to which bands. The eye can discriminate colour best in the yellow, the combination of red and green light ('hue discrimination curve'; Gregory, 1977). Therefore, the two most informative bands could be assigned to the red and green channels, the third most informative band to the blue channel. Sheffield (1985) and Drury (1993) both agreed with this allocation. Sheffield (1985) recommended further to assign the band with maximum information to the green channel, as the eye was more sensitive to green than to red. In contrast, Drury (1993) advised the most informative band to be displayed in red, but does not provide a reason for this recommendation.

Apart from the pure physiology of the human eye, the psychology and sense of aesthetics of an individual also have an influence on how much information is read from a colour image, and whether it pleases the eye or not. The latter effects can easily be underestimated, and the author believed that a colour composite should be chosen to associate familiar objects with their natural colours. For this reason, the RGB colour assignment for the New Forest best band triplet was chosen to display grass in a greenish colour (figure 4.7), while that of the River Severn data set was chosen to show the bare mud class in blue (the latter class is regularly inundated by water; figure 4.9).

Both figures 4.7 and 4.9 show a high contrast in colour between the different vegetation classes involved. For the New Forest data and the given RGB band combination, the asphalt road and the lake appear black in colour, whereas bracken occupies the cyan and dry heath the crimson colour. Wet and humid heath and valley mire may be distinguished from different tones ranging from salmon red to dark green (see also classified image in figure 3.5). Regarding the River Severn colour composite of the optimal three bands, pioneer, mid and high marsh vegetation is coloured in bright orange - red, red – blue, and yellow, respectively. Bare rock appears white in the image and bare mud blue (see classified image in figure 3.4). Figures 4.8 and 4.10 display the resulting map of the New Forest and River Severn classification task, respectively, using the six bands selected by the SBS algorithm (see table 4.6).

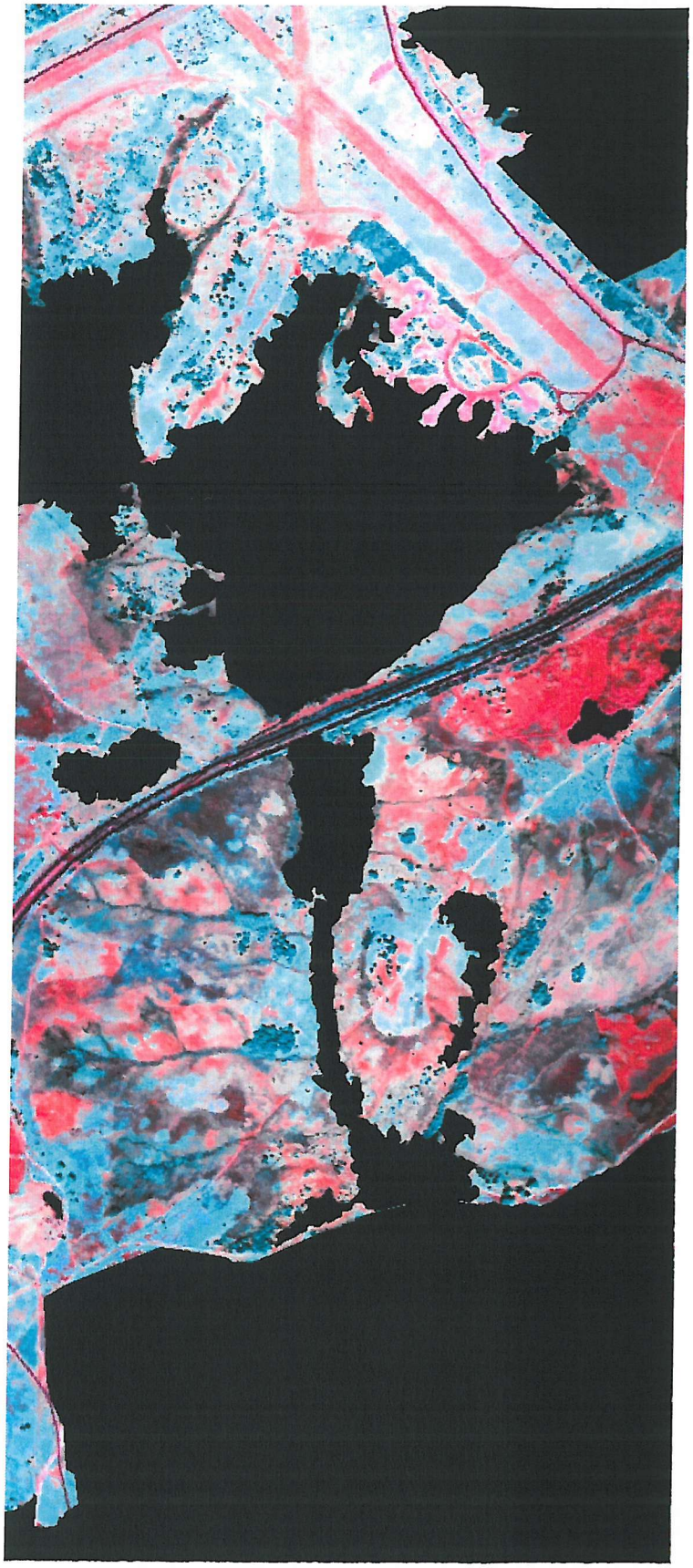


Figure 4.7: Masked HyMAP New Forest data displayed using the first three optimal bands output by the SBS algorithm (see table 4.6, R = band 2, G = band 1, B = band 3).

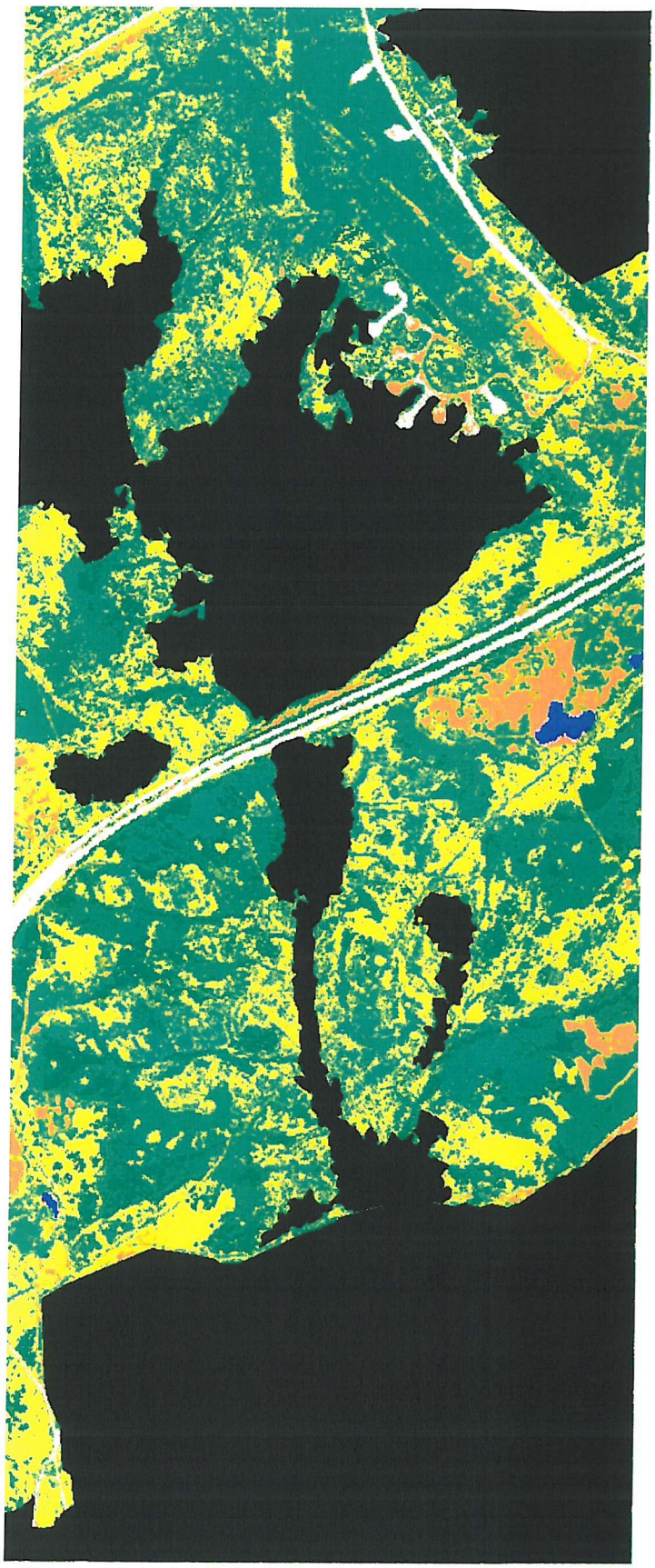


Figure 4.8: New Forest Maximum Likelihood Classification result using the first six bands selected by the SBS algorithm (see table 4.6; Lake, blue; Asphalt, white; Bracken, yellow; Dry Heath, orange; Grassland, brightest green; Humid Heath, bright green; Wet Heath, green; Valley Mire, dark green).

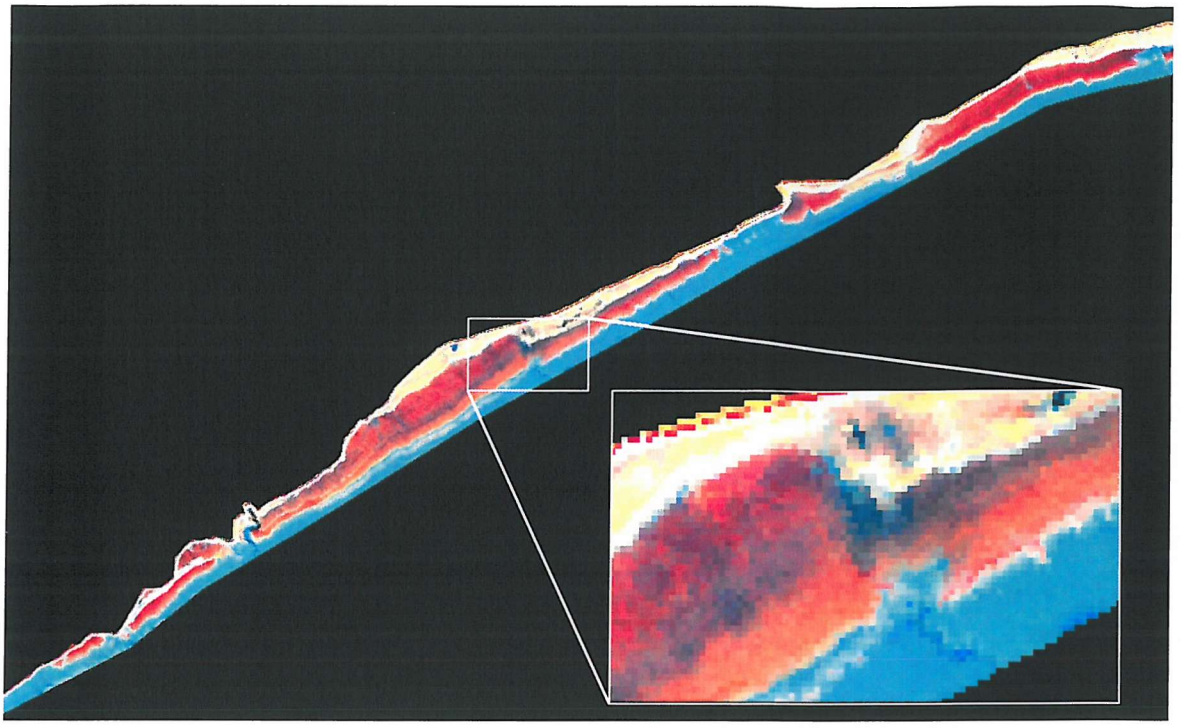


Figure 4.9: Masked CASI River Severn data displayed using the first three optimal bands output by the SBS algorithm (see table 4.6, R = band 2, G = band 3, B = band 1).

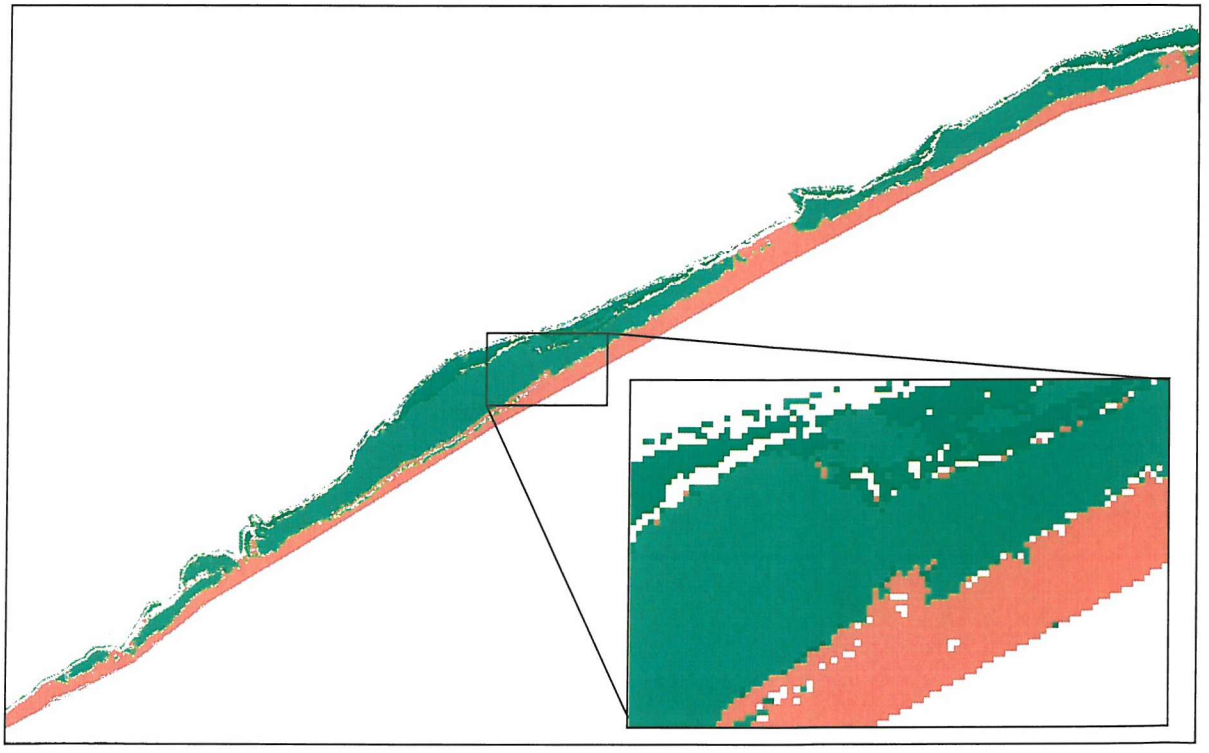


Figure 4.10: River Severn Maximum Likelihood Classification result using the first six bands selected by the SBS algorithm (see table 4.6; Bare Rock, white; Pioneer Marsh, bright green; Mid Marsh, green; High Marsh, dark green; Bare Mud, brown).

4.4.2 Reliability

The reliability of the band selection algorithm was judged by questioning the appropriateness or accuracy of its assumptions and routines.

- The band selection method assumes that image data recorded in multiple flight-lines over the same scene with the same image acquisition parameters will result in similar image pixels. In general, this assumption is appropriate as long as the H-resolution case applies (and all other parameters remain the same). This is true most of the time when using high spatial resolution airborne data for classification purposes.
- The method further assumes that the separation between the spectra of user-defined classes remains the same between the hyperspectral and multispectral image acquisitions. To ensure the suitability of this assumption, the time lag between the two data recordings needs to be minimised.
- The band selection algorithm assumes the class samples to be normally distributed. The user of the algorithm is advised to test the normality of the class data during the class definition and training process (see section 3.3.1).
- The band location, width and number estimation routines are all based on the linear relationship between the Transformed Divergence (TD) and the MLC accuracy. The accuracy of these subroutines then depends partly on the appropriateness of this relationship. For the given data sets, the multiple-band TD was shown to be highly correlated to MLC accuracy, more than any other distance measure (see table 4.5). Generally this assumption is valid as the TD was derived directly from an upper bound to the error probability (Fu, 1982).
- The sub-optimal SFS search procedure was employed for the band location determination, where the sub-optimality of the results is computationally impossible to quantify for the given data sets and classification schemes (see above). However, the SFS procedure is optimal for uncorrelated features. That is, for a correlated set of hyperspectral bands, the first few relatively uncorrelated bands chosen by the SFS are more likely to be optimal than the highly redundant remainder of the selected band set.
- The band number determination routine assumes that the maximum achievable MLC accuracy is not produced by a band set of dimension between the allowable and the total number of bands. The error of this assumption may only be quantified by performing an extensive search with the MLC accuracy measure as criterion function, which is impossible to perform due to the immense computational cost involved. The inappropriateness of the assumption may be reduced by choosing a higher number of training samples or allowable bands (see above).

4.4.3 Consistency

The algorithm is consistent in that it is repeatable, assuming no two TD estimates are equal for the band location and number determination, and band broadening sub-routines. However, the ordering of bands within the band location determination subroutine depends on the bands already selected for the set. That is, the output band set may only be reduced by repeatedly discarding the feature added last to it. Removing the first band from the 'optimal' set destroys the validity of the order of subsequent bands.

4.4.4 Effectiveness

Sub-optimality of the algorithm

The sub-optimality of the band selection algorithm was quantified in terms of MLC accuracy by deriving the best sets of one, two and three bands with an exhaustive search using the MLC accuracy measure and the SBS band selection algorithm (no bandwidth increase). Only band sets up to three dimensions were selected for the exhaustive search to be computationally feasible. Figures 4.11 and 4.12 display the MLC accuracy of the best band sets for the New Forest and River Severn data set, respectively.

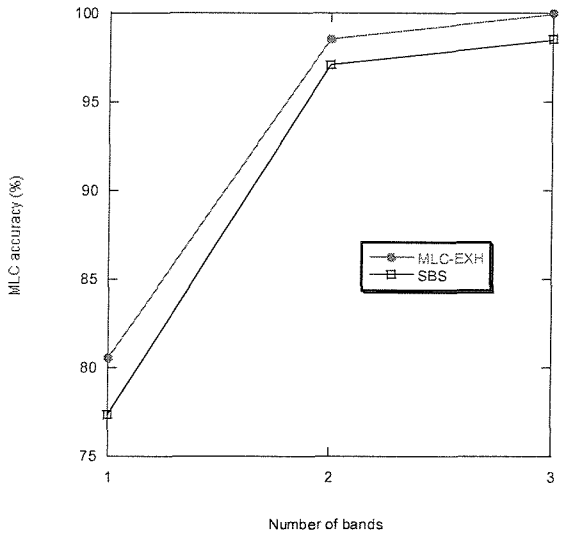


Figure 4.11: MLC accuracy of optimal band sets derived using an exhaustive search with MLC accuracy (MLC-EXH) and the SBS algorithm for the New Forest data set.

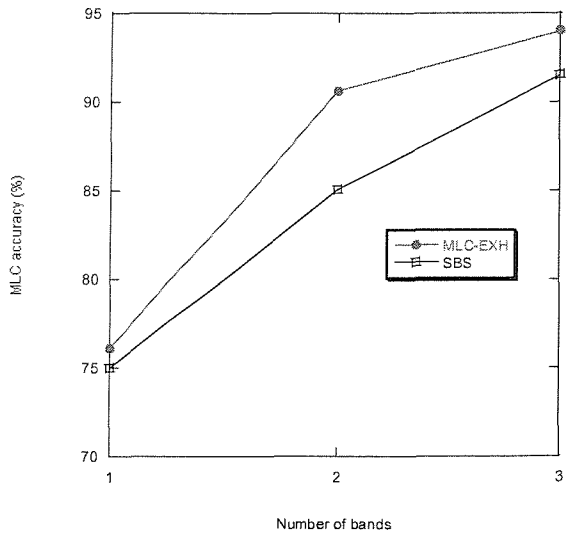


Figure 4.12: MLC accuracy of optimal band sets derived using an exhaustive search with MLC accuracy (MLC-EXH) and the SBS algorithm for the River Severn data set.

From figure 4.11 it can be seen that the band subsets selected with the SBS algorithm for the New Forest data set results in a maximum 4% loss in accuracy. For the River Severn data set, this accuracy loss reaches a maximum of 6%. The percentage difference for the New Forest data set is statistically insignificant, mainly due to the fact that the number of testing samples were reduced from 9400 to 130 to make the method computationally feasible. However, the largest percentage loss is significant for the River Severn data set, where 780 testing pixels were used for accuracy estimation.

Comparison of 'optimal' SBS band set to 'established' band sets

The 'optimal' band set output by the SBS algorithm ('unequal bandwidth' option) was evaluated regarding the established 'vegetation' and 'coastal' band sets used in current satellite and airborne sensors. As for some of the latter band sets no band ordering was available, the comparison was only performed for the full band sets. This was necessary, as otherwise ambiguity would have been introduced. In general, water vapour and oxygen absorption bands as well as specific bands for aerosol determination were excluded from the comparison as these serve principally the purpose of atmospheric correction. An exact sensor band simulation was generally not possible, as the bands of the given hyperspectral data were sometimes equal or larger in width, or did not cover the entire spectral range of some of the bands to be simulated.

Table C.3 presents the satellite and airborne band sets used for comparison and their simulation with the available HyMAP and CASI data. The band sets to be simulated include the vegetation bands from CHRIS (mode 3), Landsat ETM+, MERIS, MISR and MODIS, as well as vegetation bands specifically designed for the NERC CASI-2 sensor (NERC ARSF, 2002) and the Environment Agency (EA) CASI-2 (K. Brown, 2002, personal communication). In addition, 'coastal' band sets such as the one designed for the EA CASI-2 (K. Brown, 2002, personal communication) and the BIOTA band set, designed to discriminate between intertidal and marine suspended sediments, chlorophyll and plant tissue of intertidal vegetation communities (Thomson *et al.*, 1998a), are compared to the band set recommended by the algorithm with respect to their classification performance for the River Severn CASI data.

Figures 4.13 and 4.14 show the MLC accuracy of different simulated satellite and airborne sensor band sets in comparison with the SBS band set for the New Forest and the River Severn data, respectively. For the New Forest data set, the SBS band set performed superior to any of the simulated band sets in corresponding dimensions. The differences in classification accuracy between the SBS subset and the comparable band sets were significant in all cases and range from 1.7% for MISR to 0.15% for MODIS. The largest difference occurred for a small band set dimension. For the River Severn data set, the latter pattern was repeated. However, the SBS band set appears to have superior performance only for dimension three and four, with a maximum difference in classification accuracy of 4.6% for the MODIS band set. Above ten bands the available band sets

performed similar or better than the SBS band set. For example, the classification accuracy of the EA coastal band set was about 0.7% higher than that of the SBS set.

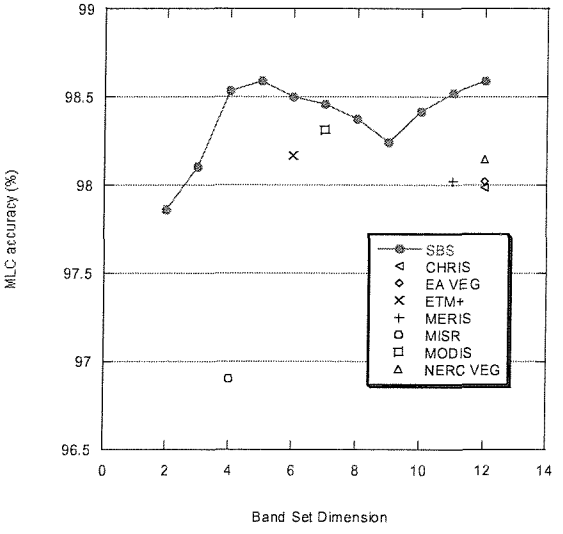


Figure 4.13: MLC accuracy of the SBS band set compared with the one of satellite and airborne band sets for the New Forest data set.

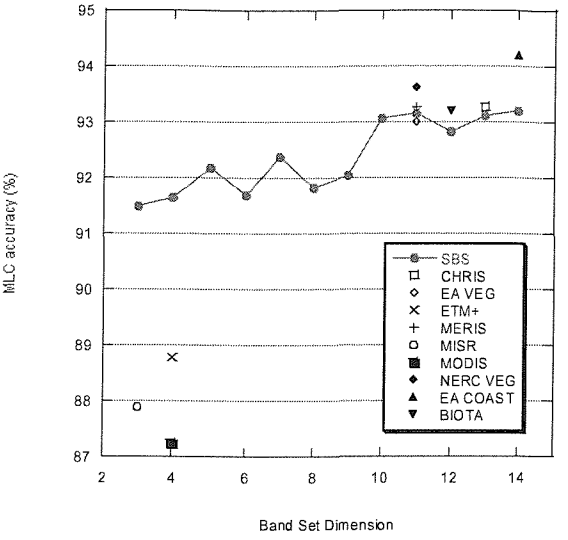


Figure 4.14: MLC accuracy of the SBS band set compared with the one of satellite and airborne band sets for the River Severn data set.

Band number evaluation

Regarding the evaluation of the band number criterion (the proportion of maximum achievable Transformed Divergence, PMATD), the validity of the proportion of maximum achievable MLC accuracy (PMAMA) as IDD estimation measure had to be tested. The PMATD is an approximation of the PMAMA. Table 4.7 displays the PMAMA and the PMATD for the first five bands of the New Forest and River Severn data set.

For the SBS algorithm, the optimal number of bands was defined as the dimension of the smallest band set that achieves a PMATD of at least 95%. If 95% is chosen as the threshold for the PMAMA, the optimal band number can be determined as 2 and 3 for the New Forest and River Severn data set, respectively (see table 4.7). These values coincide with the lower values of the estimated IDD ranges in chapter 3. The corresponding PMATD criterion gives an IDD estimate of 2 and 4 for the New Forest and River Severn data set, respectively, using the 95% threshold. The estimate is exact for the New Forest data set, but overestimates the lowest IDD value of the River Severn data set by one. However, the latter still falls in the IDD range estimated in chapter 3.

Table 4.7: Proportion of maximum achievable Transformed Divergence (PMATD) and MLC accuracy (PMAMA) for the first five dimensions of the SBS band set selected for the New Forest and River Severn data sets.

Band number	New Forest		River Severn	
	PMAMA	PMATD	PMAMA	PMATD
1	79.18245	89.74730	79.72142	82.39539
2	98.74942	98.96224	90.71178	90.75644
3	98.99402	99.44225	96.51071	93.44345
4	99.43002	99.68375	96.66257	96.40755
5	99.48319	99.80777	97.22867	97.24103

4.5 Summary

Hyperspectral imagery offers many narrow spectral bands to exploit subtle spectral differences between target classes of interest, potentially increasing application accuracy (in this chapter the author focussed on classification as the application task). However, reduced classification accuracy could result from high-dimensional imagery if an insufficient number of training samples were available. What is more, hyperspectral data are often redundant and large in volume that is difficult to record, store, transmit or process efficiently.

Band selection is a way of reducing data dimensionality at sensor level and is necessary for sensor design studies and the use of imaging spectrometers that are limited in the amount of data they can record (e.g. the CASI-2). This chapter focuses on band selection for imaging spectrometers with Maximum Likelihood classification as the final data application.

A supervised band selection (SBS) algorithm is proposed that aims to optimise band configuration parameters with respect to the accuracy of the classification task at hand. It is based on conventional feature selection methods that select the most class-discriminant bands on the basis of the user-defined class definition and the given data set. By doing so, the routine automatically reduces the high correlation between spectrometer bands and maximises the signal, that is, the class-discriminant information. In addition, it introduces the bandwidth and band number as variables into the feature selection process. The technique assumes the class samples to be normally distributed, and a linear relationship between the MLC accuracy and its surrogate, the Transformed Divergence (TD).

The SBS procedure is similar to the band selection approaches developed independently by Petrie *et al.* (1998) and Withagen *et al.* (2001). But contrary to the latter approaches, SBS utilises a saturated criterion function as opposed to a non-saturated one, which has a stronger linear relationship with the MLC application accuracy. In addition, band expansion takes place before, not after, band location determination. This is because bandwidth increase will alter the performance of

the band set. SBS introduced also a band expansion constraint based on the band SNR to assure that all bands achieve a sufficient signal level. What is more, a band number determination procedure was added to SBS to give an estimate of the intrinsic discriminant dimensionality of the data set.

The algorithm requires hyperspectral data of the target scene and a user-defined class set. The hyperspectral data should be acquired under similar conditions (sensor, scene, illumination) as the target data, or conditions need to be simulated. To take sensor characteristics (e.g. spatial and spectral resolution, point-spread function, instrumental noise) as well as the scene of interest into account, the programmable imaging spectrometer CASI-2 may be used for the collection of both the hyperspectral and multispectral data. However, a trade-off exists between number of spatial pixels and that of spectral channels that can be recorded. A reduced-swath hyperspectral CASI-2 image over a representative part of the scene may be acquired first, on which band selection is performed. A full-swath multispectral CASI-2 image may then be recorded over the same scene using the selected band set. The approach assumes the H-resolution case, which is generally applicable to remote sensing for classification, and requires the time gap between hyperspectral and multispectral acquisitions to be minimised to avoid changes in the class spectral responses and illumination conditions.

The band selection algorithm was implemented in an efficient IDL™ (version 5.5 Win 32 x86) and ENVI™ (version 3.5) program called SBS (Supervised Band Selection), that can be run within ENVI™. The program options include to either force the bands to be of equal width (can be used for creating a band set for CASI-2's enhanced spectral mode) or letting the bands expand freely to achieve the highest criterion performance (optimal band set). In addition, specific bands or classes may be included in the evaluation or excluded from it.

The algorithm was evaluated by applying it to hyperspectral data from the CASI and HyMAP sensors. The resulting optimal band configurations were found to be physically meaningful with respect to the classes under investigation. The sub-optimality of the feature selection part of the algorithm was quantified by performing an exhaustive search using the MLC accuracy measure and comparing the 'exhaustive' and SBS band set in terms of MLC accuracy. For the first three dimensions, the use of the SBS band set resulted in a maximum loss in accuracy of 4% and 6% for the New Forest and River Severn data set, respectively. This sub-optimality of the algorithm is the cost to pay for its computational feasibility and efficiency.

Furthermore, the MLC accuracy of the SBS band set was compared with the one of established 'vegetation' and 'coastal' band sets used in current satellite and airborne sensors and simulated with the given data sets. For both data sets, the SBS set performed superior to the traditional band sets at least for the first few dimensions. This may be explained by the fact that traditional band sets do not necessarily discriminate between the user-defined classes but sample the most important

spectral features of a general material class. The SBS band set for the CASI data set underperformed slightly for higher dimensions, exposing the sub-optimality of the algorithm. The algorithm therefore seems to be more applicable for deriving band subsets of small dimension for the given data sets. In other words, it achieves the aim of identifying the most computationally-efficient solution to the problem of each specific classification task.

The band number criterion, that is, the proportion of maximum achievable Transformed Divergence PMATD with a 95% threshold, was shown to be an effective measure of the intrinsic discriminant dimensionality (IDD) of the given data sets. The PMATD coincided with the lower value of the IDD range estimated for the New Forest data set in chapter 3, while it overestimated the lower IDD value of the River Severn data set by one, but falling still within the estimated IDD range.

Being a supervised method, SBS depends on accurate class definition. As the latter is an elaborate and often non-trivial task, the SBS cannot be employed in-flight between data acquisitions. To be able to perform band selection in-flight, an unsupervised band selection routine was developed, and this is described in the following chapter.

5 Unsupervised Band Selection for Classification

5.1 Introduction

This chapter focuses on the development of an automatic unsupervised band selection procedure with MLC as final data application. The ideal output is a band set optimised in band number, width and location with respect to the MLC accuracy. Generally, an unsupervised algorithm does not require any class information, and may rely entirely on band statistics. That is, statistics are no longer calculated for homogeneous class areas, but for entire band images, which may be inhomogeneous. In order to ensure the validity of the band statistics for a given data set, univariate assumptions need to be verified for each band within the set (see section 5.2).

The MLC accuracy may no longer be approximated via multiple-band class distance measures as in the supervised SBS algorithm described in chapter 4, but indirectly via band image quality measures instead. Section 5.3 discusses the use of unsupervised feature selection criteria (see section 2.2.2) as possible measures of image quality with respect to MLC as application task. The latter single-band measures do not account for the redundancy between hyperspectral bands, and section 5.4 presents possible measures of redundancy.

A correlation-based band selection algorithm, UBS, is then described in section 5.5, which addresses both image quality and redundancy problems mentioned above. The latter requires band variables to be normally distributed, which may not be the case for each data set. As an alternative to UBS, CSBS, an unsupervised version of the SBS algorithm, is presented in section 5.6, which uses a clustering procedure to define the classes within the scene. Both algorithms were programmed in IDL^{TM} (version 5.5 Win 32 x86) within the ENVI^{TM} (version 3.5) processing environment and were run in Windows NT on a 1 GHz Intel Pentium III processor with 256 MB RAM. Any quantifications of computational speed made in this chapter refer to these specifications, if not otherwise stated.

One objective of this chapter was to create an algorithm that can be used in-flight for the CASI-2 imaging spectrometer, where the data collection followed the procedure described in section 4.3.1. Therefore, computational efficiency and speed needed to be considered in the design phase of the band selection programs to allow for their real-time and in-flight executions.

The two hyperspectral data sets introduced in the pilot study (chapter 3) as the New Forest and the River Severn data sets were used to evaluate the band selection algorithms (section 5.7). In section 5.8, a concluding summary is presented.

5.2 Testing Univariate Assumptions

Most measurement processes are assumed to sample a random variable, i.e. a variable with an uncertain outcome (Hoel, 1984). In addition, data are assumed to come from a single process that can be represented by a single distribution with both a fixed location and a fixed variation. Most classical methods assume a normal data distribution. The validity of conclusions drawn from any scientific method is intrinsically linked to the validity of the latter univariate assumptions, and may be tested via graphical and quantitative techniques.

5.2.1 Graphical Techniques

Graphical techniques for testing univariate assumptions include:

1) the run sequence plot,
2) the lag plot, and
3) the histogram.

The run sequence plot displays the sequence of data points (and rows) for the entire image. A flat and non-drifting graph suggests the presence of a fixed location, while a constant vertical spread over all samples implies a fixed variation of the underlying distribution. A drifting location and variation result in a poor and biased single estimate of central tendency and spread, respectively.

The lag plot with lag equals one shows direct neighbouring values plotted against each other (here only horizontal and vertical neighbours). It tests the spatial correlation of neighbouring data pixels. The plot shows no structure for random data. If non-randomness is detected, all of the usual parameter estimates and their uncertainties may become meaningless and statistical tests invalid.

The histogram is a widely used frequency plot for data value intervals of the same width (bins), used to summarise the distribution of a univariate data set. It shows the centre, spread and the skewness of data, as well as the presence of outliers and multiple modes. Its shape may help to decide whether the data follow a certain type of distribution. In this study, the author was interested in assessing the normality of the underlying data. The corresponding normal probability density function was plotted over the histogram to illustrate any similarities in shape. Non-normality can manifest itself in a double peak (measurements are being drawn from two or more distributions), in long tails (indicating outliers in the process), in a flat pattern or a pattern with two peaks at either end (process is not in control or not properly specified).

The histogram is sensitive to the number, width and placement of bins, and any change in the latter parameters may indicate a different result as to whether the data are normally distributed or not. An optimal choice of bin width does not exist, as this depends on the data distribution. Most reasonable

choices should produce similar but not identical results. Dataplot™, a public-domain multi-platform software system for statistical analysis, calculates the bin width by multiplying the sample standard deviation with a factor of 0.3 (Filliben and Heckert, 2002).

5.2.2 Quantitative Techniques

Quantitative tests for normal distributional adequacy include the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Chi-Square Test of Independence. While the former tests are both limited to continuous distributions, the latter test of independence can be applied to any univariate distribution. Remotely-sensed data have a discrete distribution, as only integer values are possible as a result of data quantisation.

The Chi-square (χ^2) test of independence is applied to nominal (binned) data and compares observed with expected (normal) frequencies. The χ^2 statistic is calculated with equation 5.1. Its distribution is well known and tabulated.

$$\chi^2 = \sum_{i=1}^{b} \frac{(O_i - E_i)^2}{E_i} \tag{5.1}$$

Where
- χ^2 Chi-squared test statistic
- O_i Observed frequency for bin i
- E_i Expected frequency for bin i
- b Number of non-empty (> 5) bins

For χ^2 to be valid for discrete distributions, the expected frequencies need to be at least 5 for each bin, otherwise bins need to be combined to achieve this minimum frequency. The null hypothesis H_0 of the test assumes that the observed and expected frequencies are independent. The alternative hypothesis is that a relationship exists between the data. The level of significance α is commonly set to 1% or 5%, so that one has one or five chances in 100 of making a type I error. The critical value $\chi^2_{\alpha,df}$ is the upper alpha percentile point of the chi-squared distribution with df degrees of freedom. df equals $b-1$ minus the number of estimated parameters (2 for a normal distribution). The decision rule is: Reject H_0 if $\chi^2 < \chi^2_{\alpha,df}$, otherwise accept H_0. For a large number of samples (df > 30), the χ^2 distribution approximates the normal distribution closely (Hinton, 1995). The χ^2 test of independence is sensitive to the choice of bins.

In addition to the quantitative methods described above, a correlation coefficient may be computed between the histogram data and a theoretical normal distribution. The correlation coefficient r can

be compared to a table of critical values $r_{\alpha,N}$ to provide a formal test of the null hypothesis H_0 that the data do not come from a normal distribution, with N being the number of data pairs. The decision rule is: Reject H_0 if $r > r_{\alpha,N}$, otherwise accept H_0. This test is a lower one-tailed test, as perfect normality implies the maximum correlation value of 1 and the interest in this test lies in rejecting normality for correlation values that are too low. The table of critical values may be found in NIST/SEMATECH (2003). This test of normality is also susceptible to the selection of bins.

Most methods in pattern recognition and statistics require that the data variables follow a normal distribution, or one that is derived from the normal distribution (such as the t, F, or χ^2 distribution). If the variables are found not to be normally distributed, 'non-parametric' methods, equivalently referred to as 'parameter-free' or 'distribution-free' methods, can be used. These techniques require few assumptions about the underlying populations from which the data are obtained, in particular forgoing the traditional normality assumption.

5.2.3 Application to Remotely-Sensed Data

Sensor bands may be considered as random variables belonging to a single normal distribution, when measured over pixels belonging to the same homogeneous class. However, for the unsupervised band selection program to be described in section 5.5, the knowledge of certain class areas is not available, and pixels of the entire scene are now considered as measurements of one band. However, a remotely-sensed image scene hardly ever consists of one single homogeneous class, but of several distinct classes. That is, data arise from multiple distributions rather than a single one, resulting in several data clusters in feature space. The assumption of a single underlying distribution for a band measured over an entire image scene may only hold if most of the scene was made up of the same material. That is, if an algorithm uses band statistics, such as band mean and variance, the normality assumption needs to be verified for each image band.

The above graphical techniques were applied to some bands of the New Forest and River Severn imagery, and results are displayed for some of the bands in appendix D.

The run sequence plots for the bands of the New Forest data (see table D.1) clearly expose a strongly drifting location and variation for the data. Bands 18 to 60 show a less pronounced change, but the variations in location and spread are still significant. The run sequence plots for the bands of the River Severn data (see table D.2) start with a relatively fixed mean and spread for the first few bands. However, the differences of location and variation increase for the remainder of the bands.

The lag plot for the bands of the New Forest and River Severn data are presented in tables D.3 and D.4, respectively. The lag plots of the New Forest data clearly show the data points distributed along

the identity line, revealing a strong spatial correlation between neighbouring pixels. The shape transforms slowly from a 'spinning top' (band 1) to a diagonal ellipse (band 19), and maintains the elliptical shape until it transforms back to a 'spinning top' (band 88). In contrast, the lag plots of the River Severn data maintain an elliptical shape surrounded by a cloud of data points that even transforms into a near-circular shape for bands 31 to 34, suggesting a smaller spatial autocorrelation than for the New Forest data. For MLC-type data application tasks that use remotely-sensed imagery, usually high spatial autocorrelation is required in order to obtain the desired H-resolution case. Data points should therefore fall close to the identity line.

The histogram of the New Forest bands (table D.5) show mostly a single centre with a relatively symmetrical spread, suggesting that the data have a single underlying distribution. A long tail to large DN (digital number) values indicates the presence of outliers. The histograms are generally too peaked and skewed to fit the normal probability distribution function, however a relatively good fit may be observed for bands 19 to 30. Table D.6 displays the histograms of some of the bands of the River Severn imagery, which clearly exhibit a bimodal distribution for some bands. For bands 34 to 38, the second peak is hidden, but re-emerges slightly for the remaining bands. The single-peaked histograms fail to match the corresponding normal curve due to their strong peakedness and skewness.

The results of the chi-squared test and the significance test of the correlation coefficient are shown for the New Forest and River Severn data in tables D.7 and D.8. Only a few bands of the New Forest data were shown to be approximately normally distributed by the significance test of the correlation coefficient (for bands 19 to 26 and 29 to 31, the independence hypothesis was rejected at a significance level of 1%, for bands 17 and 32 at a significance level of 5%). The chi-squared test did not suggest a normal distribution of any of the bands of the two data sets.

5.3 Image Quality Measures

Clearly, the definition of 'image quality' is dependent of the application of the image data. For MLC as data application, a band image may be considered of higher quality if it conveys more discriminant information about the classes of interest than a noisy image does. As classes of interest are unknown in the unsupervised case, information need to be quantified in a class-independent way.

Some unsupervised measures equate image variance with information, assuming that the noise within the data is small relative to the signal (Richards and Jia, 1999). In short, the higher the scene variance, the more class discrimination is possible. These measures include the band variance itself, the discriminant power measure determined from eigenanalysis via PCA (section 2.3.5, equation 2.38), the Priority Number criterion and the coefficient of variation (section 2.2.2).

However, if the noise proportion becomes larger with respect to the signal, the variance image statistic also measures noise. Then, a noisy scene image with low class discriminatory power may equally result in a high scene variance.

The signal-to-noise ratio (SNR) is one of the most common measures of image quality (Green *et al.*, 1988) and may be defined as ratio between signal and noise (see section 2.2.2). The information of interest to the user is inherent in the signal and may be estimated with the image band mean to be representative for the entire image scene. Different methods for image noise estimation exist and are presented in section 2.3.2. In general, the author preferred noise estimates derived from dark data to those from scene-based data.

SNR differences between bands may be due to solar, atmospheric and instrumental influences on the radiance signal, as well as the scene reflectance. One might assume that the SNR and MLC accuracy are positively correlated, as an image high in noise (and low in SNR) would lead to a lower MLC accuracy and vice versa. However, if SNR increase is achieved, for example, via band expansion, the MLC accuracy may suffer, as broader bands may discriminate less between the given classes. That is, only a very loose relationship exists between the SNR and MLC accuracy.

Spatial autocorrelation, e.g. calculated using Geary's c-metric (Geary, 1954; see section 2.3.2), may also be used as image quality measure. The higher correlated neighbouring pixels are, the less noise may be present in the image.

In an experimental analysis, the appropriateness of the above image quality measures as relative estimates for the MLC accuracy was tested. Table 5.1 shows the correlation coefficient between the MLC accuracy estimated with the 'leave-one-out' method (see section 3.3.1) and the unsupervised criterion functions for the New Forest and River Severn data and class sets. Their corresponding scatter plots may be found in tables F.1 and F.2 in the appendix. Regarding the band SNR, the standard deviation of a homogeneous area was used as noise estimate for the River Severn data, as dark current data were only available for the New Forest data set.

Table 5.1: Correlation coefficient between unsupervised criterion functions and the MLC overall accuracy estimated with the leave-one-out method for the New Forest and River Severn data set.

	New Forest	River Severn
Variance	0.824	-0.766
Discriminant power (PCA)	0.767	-0.734
Priority number criterion	0.779	-0.780
Band SNR	0.732	-0.844
Coefficient of variation	-0.799	-0.168
Geary's c metric	-0.353	-0.084

The variance-based criterion functions, including band variance, discriminant power and priority number criterion, as well as the band SNR showed a relatively high correlation with MLC accuracy estimation for both data sets. However, in case of the New Forest data set, this relation was positive, while it was negative for the River Severn data set. This suggests that all these measures are not reliable estimates for ML classification accuracy, being able to represent either information or noise, depending on the quality and heterogeneity of the data set. In this case, the River Severn data set was on the whole much more heterogeneous than the New Forest one.

Both the band coefficient of variation (here defined as the ratio between band standard deviation and mean) and Geary's c metric are, as measures of image noise, negatively correlated to the MLC accuracy, which is reflected in the correlation coefficient for both data sets. However, for both criteria, the correlation to MLC accuracy is very low for at least one of the two data sets.

As a result from the experiment, none of the unsupervised image quality measures under investigation represents a consistent estimator of MLC accuracy. However, if a choice was to be made, the author would prefer the SNR as quality measure as it is computationally cheap and scales the signal component with the noise component. The latter is important, as a band with a small noise estimate may only be considered 'information-bearing' if the signal of the band is relatively high. The dark current standard deviation should be preferred to the homogeneous-area standard deviation as detector noise estimate. It is more accurate and particularly important for imaging spectrometers using separate detectors for different wavelength ranges (such as HyMAP), as the noise originating from different detectors may vary.

Figures 5.1 and 5.2 display the result of the SNR value calculation for the New Forest and River Severn data set, respectively. In both cases, the signal for each band was estimated with the image band mean. For the New Forest image, the noise covariance matrix was derived directly from HyMAP's dark current measurements. As dark data were not available for the River Severn CASI image, the noise covariance matrix was derived with the 'shift difference' technique from a spatially and spectrally homogeneous Bare Mud region within the image (see section 3.4.1). To note is that the calculated SNR of the New Forest imagery (figure 5.1) does not represent realistic values, as the dark data were not scaled to the same unit as the band mean, that is, reflectance.

It may be noticed that in both cases the SNR curve resembled roughly the mean spectrum, a vegetation spectrum. However, in both cases the NIR plateau includes a fall of SNR suggesting a noise increase between detectors (HyMAP data) or at the end of a detector (CASI imagery). In addition, it may be seen from figure 5.1 that the HyMAP sensor noise is higher in the SWIR region than in the VIS or NIR.

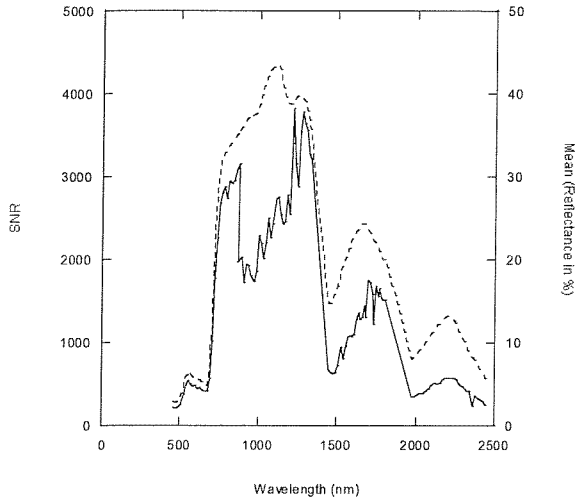


Figure 5.1: Band Signal-to Noise Ratio (solid line) and mean (dashed line) for the New Forest data set with the noise estimated from dark current data.

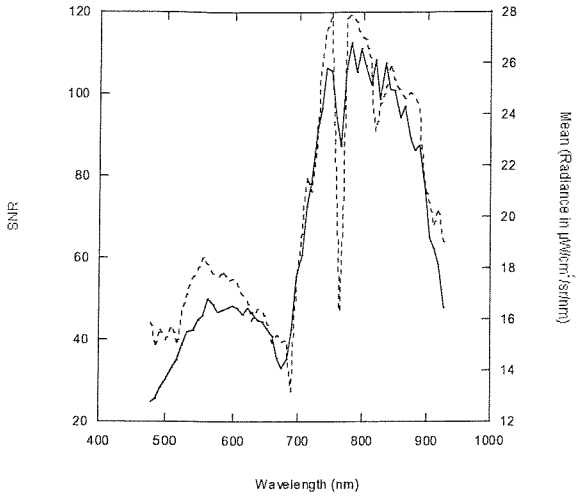


Figure 5.2: Band Signal-to Noise Ratio (solid line) and mean (dashed line) for the River Severn data set with the noise estimated from a bare mud image region.

Image quality estimates including the band SNR are single-band measures only and do not explain the redundancy between hyperspectral bands. The next section introduces potential measures of band redundancy.

5.4 Redundancy Measures

The algorithm in section 5.5 is designed to find a set of bands that are least redundant with each other for a given scene. In the context of MLC, a band in a band set may be called redundant if it does not convey any more information for class discrimination than the rest of the bands in the set do. That is, redundant bands within a band set do not contribute to a significant increase in the band set classification accuracy.

For the unsupervised algorithm in section 5.5, an unsupervised measure is needed to estimate the redundancy between bands with respect to the MLC accuracy. Chang *et al.* (1999) used the divergence to measure the overlap between band distributions in order to eliminate redundant bands. If the distance measure was below an empirically chosen threshold, the bands were assumed to be highly redundant and were removed. However, the divergence only quantifies the 'shape' difference between two band histograms, not the redundancy of the bands. Two bands with identical distributions may result from highly non-redundant band images that have a different spatial distribution of DN values.

5.4.1 Correlation

In general, correlation is used to describe a relationship between two variables j and l. To measure the strength and the direction of a linear correlation, the Pearson (or product-moment) correlation coefficient, r_{jl} (or r), introduced in sections 2.1, 2.2.2 and 3.3.2 may be calculated. Equation 2.4 illustrates that r corresponds to the covariance of standardised variables, where the covariance is a measure of how much the deviations of two variables match. r ranges from -1 (perfect negative correlation) over 0 (lack of correlation) to 1 (perfect positive correlation). If only the strength of a relationship is of interest, the coefficient of determination, r^2, is reported, which corresponds to the proportion of common variation in the two variables.

The calculation of r assumes

- data to be measured on an interval or ratio scale,
- variables to be normally distributed,
- measurement errors to be small,
- the underlying relationship between variables to be linear, and
- the relationship between the two variables to remain constant at all points (homoscedasticity).

Examining the scatter plot of the two variables may check both the homoscedasticity and linearity assumptions. Points should be evenly spread around the regression line. Any outliers or isolated points and clusters can strongly influence the strength of the correlation. Unfortunately, a widely accepted method to remove outliers does not exist. Attempts to deal with nonlinear relationships include the removal of nonlinearity by some logarithmic or square root data transformation, or the use of a less sensitive nonparametric correlation coefficient (see section 5.4.2).

The reliability or significance of a correlation, i.e. the likelihood of achieving the same correlation if another sample from the same population was drawn, may be tested with significance tests. Generally, the stronger the correlation, the more reliable it is. In addition, the significance of a correlation is a function of the sample size. A small correlation can only be proven to be significant in large samples, whereas a strong one may be found significant even in a small sample size. For example, in a coin tossing experiment with 10 samples, a heads to tails frequency ratio of 49:51 is harder to prove than a ratio of 0:100. That is, if the sample size is small, the probability of a random deviation (noise) is high, while it decreases with increasing sample size. A highly significant correlation does not automatically imply that the correlation is strong.

Significance tests in general aim to check whether a relation of a given magnitude or larger from a sample of given size is likely or not. They usually assume no such relation between the variables in the population (null hypothesis), and quantify the probability of this hypothesis. If this probability is very small, conventionally below 5% or 1% (significance level α), the null hypothesis can be

rejected with 95% or 99% confidence, respectively. The significance level is the probability of making a type I error, i.e. rejecting the null hypothesis although it is true.

To test the significance of a correlation, first the null hypothesis 'No correlation exists between the variables' is stated ('test for zero correlation'). It can be shown that, assuming a zero population correlation, the distribution of the sample correlation coefficient r_{jl} based on N pairs of observations follows a distribution, which is related to the Student's t-distribution with $N-2$ degrees of freedom for small N according to equation 5.2 (Owen, 1962).

$$t = r\sqrt{\frac{N-2}{1-r^2}} \tag{5.2}$$

For large N, the sampling distribution of r_{jl} is approximately normal, but bounded at -1 and 1. Any value of r may be transformed to a standard variable according to equation 5.3 assuming zero population correlation (Edwards, 1984):

$$z_r = \frac{\sqrt{N-3}}{2}\left[\ln(1+r) - \ln(1-r)\right] \tag{5.3}$$

With the knowledge of the distribution of the test statistic under the assumption of zero correlation, a critical value t_α (or z_α for large N) at a chosen significance level α may be calculated for the statistic. Standard tables of critical values exist for the t- and z-distributions for certain significance levels and a certain range of degrees of freedom. If a table was calculated for two-tailed tests, corresponding values for one-tailed tests may be found in the column for twice the significance level.

The following decision rules for rejecting the null hypothesis (for small N) are possible, depending on whether a one-tailed or two-tailed test for zero correlation has been chosen:

- One-sided upper-tail test with alternative hypothesis 'Variables are positively correlated':

  Reject null hypothesis, if $t \geq t_\alpha$.

- One-sided lower-tail test with alternative hypothesis 'Variables are negatively correlated':

  Reject null hypothesis, if $t \leq -t_\alpha$.

- Two-sided test with alternative hypothesis 'Variables are correlated':

  Reject null hypothesis, if $|t| \geq t_{\alpha/2}$.

For large N, use z_r and z_α instead of t and t_α, respectively. If the null hypothesis is rejected, a significant correlation is said to exist between the variables for the given experiment at the chosen level of significance. In other words, the correlation is big enough to reject the possibility that it arose by chance. One-tailed tests are statistically more powerful than two-tailed ones, as they decrease the probability of making a type II error, i.e. accepting the null hypothesis although it is not true.

The matrices of the coefficient of determination for the River Severn and New Forest data sets are displayed in figures 3.10 and 3.11, respectively, and discussed in section 3.3.2. As for both data sets, N is large (379,169 and 43,720 samples for the New Forest and the River Severn data set, respectively), the large sample approximation of the critical value z_α may be applied. For a two-sided test and a significance level of 1% (5%), $z_{\alpha/2}$ equals to 2.576 (1.960). Figures 5.3 and 5.4 show the significance of the correlation coefficients at a significance level of 1% in form of a binary image for the New Forest and River Severn data sets, respectively. The significant correlation coefficients are marked white and the insignificant ones black. For both data sets, the majority of all correlations between bands is statistically significant.

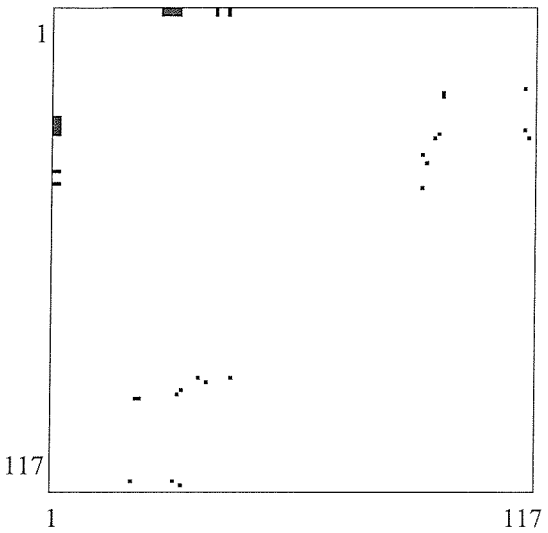


Figure 5.3: Significance (1%) of the correlation coefficients for the New Forest data set (significant = white, insignificant = black; axes = band number).

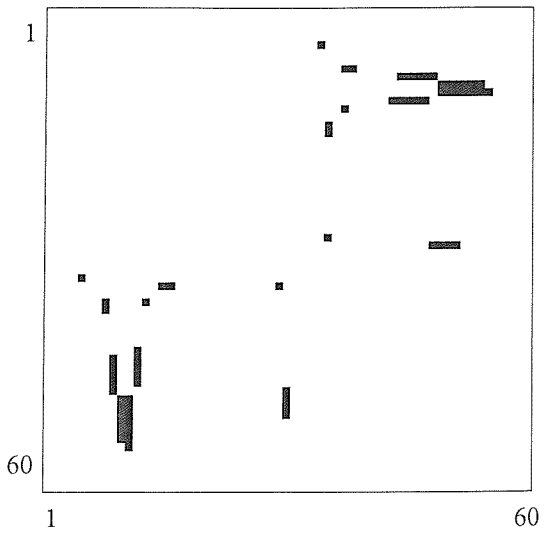


Figure 5.4: Significance (1%) of the correlation coefficients for the River Severn data set (significant = white, insignificant = black; axes = band number).

5.4.2 Nonparametric Correlation Measures

If the measured samples are not normally distributed, the calculation of the correlation as indicator of a relationship between two variables may be misleading. Nonparametric equivalents to the Pearson correlation coefficient include Spearman's and Kendall's rank correlation coefficient, which are only based on the ranks of the data. The former is described further as an example. The Spearman rank correlation coefficient, r_S, is defined by equation 5.4 (Owen, 1962).

$$r_S = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} \tag{5.4}$$

Where r_S Spearman's rank correlation coefficient

d_i Difference between the ranks of the i th pair of item

N Number of pairs of items

The distribution for r_S assuming a zero population correlation is known and critical values $r_{S,\alpha}$ are tabulated (Owen, 1962, table A.31). The same decision rules apply as for the Pearson correlation coefficient replacing t and t_α with r_S and $r_{S,\alpha}$, respectively. For large N, the standardised form of r_S, r_S^*, may be calculated (with equation 5.5), which has an asymptotic z -distribution if N tends to infinity and the null hypothesis is true.

$$r_S^* = r_S\sqrt{N-1} \tag{5.5}$$

For the large-sample approximation, the same decision rules apply as for the Pearson correlation coefficient, replacing t and t_α with r_S^* and z_α, respectively.

Although r_S requires the variable only to be measured on an ordinal scale and forgoes the normality assumption of the variables, it is generally considered statistically less efficient and powerful as its parametric counterpart, that is, it may not detect a correlation when that correlation actually exists.

Normal distribution based tests may still be used with non-normally distributed variables for large sample sizes, as the consequences of violating the normality assumption have been shown to be less severe than previously thought (Statsoft, Inc., 2002). However, the user has to be aware that some conclusions may not be valid.

5.4.3 Multiple Correlation

The algorithm in this chapter is designed to find a set of bands that are least redundant with each other and with any combination of other bands within the set. The multiple correlation coefficient $R_{Y.12}$ may be used as a measure of redundancy of a new band Y with the entire band set (bands 1 and 2). $R_{Y.12}$ takes values between 0 and 1, and its square form, the coefficient of determination for multiple correlation, $R^2_{Y.12}$, represents the variability of band Y that is explained by bands 1 and 2. Usually, $R^2_{Y.12}$ is computed directly via regression using original data. However, for large data sets, such as hyperspectral image data with a high number of bands (about 100 or more), each containing 100,000 samples or more, computer memory problems for standard regression software packages (such as ENVI™) may occur. Therefore, the author sought an alternative and more efficient way of calculating $R^2_{Y.12}$ from the band correlation matrix alone.

If all bands were mutually uncorrelated, the variability of Y explained by both bands 1 and 2, $R^2_{Y.12}$, may be computed by the simple sum of the variability of Y explained by band 1, r^2_{Y1}, and that explained by band 2, r^2_{Y2} (see equation 5.6). This is because the contribution of band 1 to the value of $R^2_{Y.12}$ is unique and independent of the contribution of band 2 (Edwards, 1984). The generalisation of equation 5.6 to any number of mutually orthogonal bands, k, is shown in equation 5.7.

$$R^2_{Y.12} = r^2_{Y1} + r^2_{Y2} \tag{5.6}$$

$$R^2_{Y.123\cdots k} = r^2_{Y1} + r^2_{Y2} + r^2_{Y3} + \cdots + r^2_{Yk} \tag{5.7}$$

However, hyperspectral image bands are usually highly correlated, and the above formulae result in an incorrect estimate of the coefficient of multiple determination. In order to correct equation 5.6 for correlated bands, the variability in r^2_{Y2} that is explained by band 1 needs to be removed. This is achieved by first calculating the correlation of bands Y and 2 having removed the effects of band 1, that is, the partial correlation coefficient $r_{Y2.1}$ (equation 5.8; Hinton, 1995). Equation 5.9 gives the formula to partial out the effects of more than one band (here bands 1 and 2) on the correlation of Y and 3.

$$r_{Y2.1} = \frac{r_{Y2} - r_{Y1}r_{21}}{\sqrt{1-r^2_{Y1}}\sqrt{1-r^2_{21}}} \tag{5.8}$$

$$r_{Y3.12} = \frac{r_{Y3.1} - r_{Y2.1}r_{32.1}}{\sqrt{1-r^2_{Y2.1}}\sqrt{1-r^2_{32.1}}} \tag{5.9}$$

But $r_{Y2.1}$ does no longer apply to the entire variability of Y, but to the one that remains after the variability of Y explained by band 1 has been removed, that is to $(1-r^2_{Y1})$. The correct formula for $R^2_{Y.12}$ when bands 1 and 2 are correlated is given in equation 5.10 (Hinton, 1995).

$$R^2_{Y.12} = r^2_{Y1} + r^2_{Y2.1}\left(1-r^2_{Y1}\right) \tag{5.10}$$

$$R^2_{Y.123} = R^2_{Y.12} + r^2_{Y3.12}\left(1-R^2_{Y.12}\right) \tag{5.11}$$

Similarly, the coefficient of multiple determination may be computed for any number of bands in a recursive fashion. An example is given in equation 5.11 for $R^2_{Y.123}$. Using formulae 5.6 to 5.11, the multiple correlation coefficient may be determined entirely on the basis of the band correlation matrix.

The significance of a multiple correlation may be tested with the F -test (equation 5.12), with degrees of freedom $(k, N-k-1)$, where N is the number of samples and k the number of independent variables (number of bands already in the set). The F -statistic (or variance ratio)

divides the estimated variance of the 'explained variability' by the one of the 'unexplained variability' (Hinton, 1995).

$$F = \frac{R^2 / k}{\left(1 - R^2\right)/\left(N - k - 1\right)} \tag{5.12}$$

The F -distributions are known and critical values F_α for a certain significance level α (usually 1 and 5%) and any combination of degrees of freedom (even up to infinity) are tabulated. The null hypothesis assumes no systematic difference between the variances of 'explained' and 'unexplained variability' ($F = 1$), i.e. no significant multiple correlation. The decision rule for rejecting the null hypothesis is given as follows:

- One-sided upper-tail test with alternative hypothesis 'Multiple correlation is significant':

  Reject null hypothesis, if $F \geq F_\alpha$.

From equation 5.12 and the latter decision rule, a lower limit of R^2 for its significance at a given level may be calculated for the given data sets (sample number N) and the number of bands already in the set (k) (see table 5.2).

Table 5.2: Lower limits of the coefficient of multiple determination R^2 for statistical significance at a level of 1% and 5% for the New Forest and River Severn data sets. k is the number of bands already in the band set when the new band is added.

k	New Forest		River Severn	
	1%	5%	1%	5%
2	0.002485	0.001628	0.021096	0.013736
3	0.003004	0.002081	0.025929	0.017854
4	0.003466	0.002468	0.030368	0.021696
5	0.004051	0.002996	0.03453	0.02534
6	0.004463	0.00335	0.038495	0.028842
7	0.00485	0.00368	0.042229	0.032138
8	0.005221	0.004001	0.045944	0.035461
9	0.005765	0.004512	0.049579	0.038721
10	0.006111	0.004817	0.053124	0.041928
11	0.006449	0.005114	0.056518	0.044965
12	0.006985	0.005617	0.059956	0.048082
13	0.007314	0.005905	0.06335	0.051173
14	0.007627	0.006185	0.066698	0.054236
15	0.008155	0.006696	0.069904	0.05713
16	0.008468	0.006968	0.073181	0.06014
17	0.008773	0.007232	0.076432	0.063132

From table 5.2 it may be seen that in order to become insignificant R^2 values have to reach below 0.01 and 0.1 in magnitude for the New Forest and River Severn data set, respectively, and a maximum of 19 output bands.

5.5 Correlation-Based Algorithm

Both the problem of redundancy and image quality stand at the heart of this algorithm (see sections 5.3 and 5.4). Selecting bands entirely on the basis of a high SNR value may result in a band set of high image quality but also of high redundancy. For hyperspectral remote sensing data, the author perceived redundancy as the greater dilemma, as sensor noise from current airborne sensors is usually very low (e.g. high SNRs for CASI-2 and HyMAP sensors of up to 500:1 in the solar reflective range, see table 1.3) and the few noisy bands usually at the edges of detectors may be singled out by the analyst.

5.5.1 Decorrelation

The first main aim of this algorithm is to produce an approximately decorrelated set of bands where the bands are ranked according to increasing correlation. The first two bands are the least correlated, and added bands become more redundant with respect to the band set with increasing band number.

Decorrelation of a set of redundant features may be defined as the process of transforming original features to a reduced set of mutually statistically independent features that explain the majority of the variance in the data set. Generally, decorrelation is performed with PCA, as during this transformation the covariance matrix (and the correlation matrix) will be diagonalised, and therefore resulting features orthogonalised. However, PCA produces new features that usually do not relate to the original sensor bands on a one-to-one basis, but are linear combinations of all original bands, destroying the integrity of the data. That is, without modifications PCA cannot be employed directly for decorrelating a band set, when original bands should result.

Some authors use PCA results indirectly to select least correlated bands. Thomson *et al.* (1998a) used the eigenvector loadings of a PC feature as indicator of the domination of bands in the linear combination to eliminate highly redundant bands. Higher order PCs may be ignored as they provide little significant information for separating different classes in an image (Richards and Jia, 1999). This procedure is similar to calculating the discriminant power measure from PC features (section 2.3.5). However, this methodology involves subjective judgement of the user about the level of noise within a certain band, that is, when to ignore a PC and when not. Whether or not to include certain bands within an eigenvector is also not clear.

Gruninger *et al.* (2001) presented a sequential unsupervised band selection method, which aims to eliminate redundant bands using a modified Gram-Schmidt procedure. First, the brightest band is chosen and removed from the remaining bands via convex projection. As each band mean represents a vector in feature space, the latter projection practically subtracts the longest vector

from all other vectors. Then, the second band is selected as the brightest, i.e. the longest, remaining vector and is in turn removed from all other vectors not yet in the band set. This process is repeated for all bands. When terminated, highly correlated neighbouring bands may be merged using a correlation threshold. However, this procedure relies on the vector length as a measure of redundancy between the removed vector and the remaining vectors. This may be misleading as vectors (bands) may be of different length (brightness). As a result, a brighter but more correlated band may be chosen over a darker but less correlated one.

Some decorrelating band selection procedures were created mainly to produce most informative colour composites. Chavez *et al.* (1982) developed the Optimal Index Factor (OIF) measure, which ratios the sum of the standard deviations of all three bands with the sum of the absolute correlation coefficients between any two of the three bands. All possible band combinations need to be investigated before the one with a maximum OIF is chosen. Sheffield (1985) selected the band triplet which maximised the volume of the ellipsoid defined by the 3 by 3 covariance submatrix. The volume may be calculated by the determinant of the covariance submatrix. This method discourages the selection of highly correlated bands, as in such cases one eigenvalue will be close to zero giving a small ellipsoid volume. Crippen (1987) advised to use the band correlation matrix instead of the covariance matrix in the latter method to eliminate band scaling effects on the result. Both of the methods described above are applicable to higher dimensional output band sets. However, Chavez's method exploits only pairwise and not multiple correlations, and uses variance as measure of information. Sheffield's method relies on the calculation of the determinant of a matrix, involving higher computational cost for higher-dimensional matrices.

In this thesis, the author chose the coefficient of multiple determination, R^2, to decorrelate the band set. The latter reflects the percentage variance of a new band that is explained by the bands already within the set, i.e. by any possible linear combination of them. Then, a band set is decorrelated by first choosing the bands with the least correlation between them. Any additional band to be included into the set needs to have the least R^2 with respect to the bands already in the 'decorrelated' set. As sensor bands are by nature statistically dependent features, this decorrelation process will not result in an orthogonal band set, but in a set of approximately least redundant bands.

For the decorrelation result to be appropriate, noisy bands need to be either excluded from the process or smoothed with some spatial filter beforehand. Noise can result in a low correlation coefficient.

5.5.2 Bandwidth Increase

The second aim of this algorithm is to increase the width of individual bands within the set in order to both improve the band image quality (and hopefully the application accuracy) and allow for finer spatial resolution or a larger swath width (as for the CASI-2) in the sensor configuration.

However, there is a general trade-off between bandwidth increase and achieving minimum band correlation in the set. The author demonstrated in a mathematical proof in appendix E that once the least correlated bands have been chosen within a set, any increase in width of one of the bands will result in a higher correlation coefficient. Consequently, the user should have the option to either fix the bandwidth to a certain number of neighbouring bands to be summed, or let the algorithm expand the bands freely without a bandwidth constraint.

As mentioned earlier in section 4.3.4, the merger of adjacent bands may improve the image quality (here the SNR) and possibly the MLC performance of the band set. However, broader bands may discriminate less between the given classes. The SNR is therefore not the ideal criterion to be used for MLC accuracy increase with band expansion, but it has been chosen from other image quality measures due to the reasons described in section 5.3.

The width of a band may then be increased as follows. The adjacent band leading to a better SNR performance of the band is merged. If an expansion into a certain direction drops the SNR, the bandwidth is no longer increased in this direction. A drop in SNR may occur when a band is merged to another one with a smaller SNR. This process is repeated until both left and right band mergers decrease the SNR of the expanded band. The actual merger of bands involves simply summing their DN values (see section 4.3.4), and the signal and noise statistics are updated according to equations 4.1 to 4.5. By summing the DN signals between neighbouring bands, the bandpass is automatically taken care of.

If the user lets the algorithm to broaden bands freely in order to maximise the band SNR, the same constraint to the 'unequal bandwidth' option for the supervised band selection algorithm needs to be applied (see section 4.3.4). This is to ensure that all bands in the set obtain at least a certain percentage (20%) of the maximum achievable signal level.

5.5.3 Band Number Determination

This section aims to determine the optimal band number for the final set. As in the supervised case, the question may be asked whether the user- or sensor-defined output band number is actually necessary to achieve the same MLC accuracy as with the full band set, or whether fewer bands may produce similar results.

For the algorithm described here, class information is not available to produce an estimate for the intrinsic discriminant dimensionality (IDD). In addition, intrinsic dimensionality (ID) estimation methods are not applicable, as original bands are to be selected by the algorithm, and not linearly independent features. In the following paragraph, a measure is described which represents an upper limit to the ID estimate.

In the algorithm described above, the band with the minimum coefficient of (multiple) determination regarding previously selected bands is chosen as next member of the set. It is the band among all remaining bands with the highest percentage variance that is not explained by the bands already within the set. If this unexplained variance decreases below 5% or 1% for the actual band (above an R^2 value of 95% or 99%, respectively), the addition of the actual or any further band can be regarded as an insignificant contribution of variance to the set. The optimum band number then equals to the size of the band set including the last band that does not exceed the given R^2 threshold (e.g. 95% or 99%).

5.5.4 Algorithm Implementation

As increasing the width of a band changes its correlation with any other band, band expansion has to precede any decorrelation procedure. Hence, the following unsupervised band selection approach is proposed.

1) Bandwidth increase of all bands in turn: First, the bandwidth of all bands is increased independently under the constraints that the SNR is improved by the band merger, and that only spectrally adjacent bands may be merged.
2) Selection of the first band(s): Second, according to the user's choice the band with the highest SNR or the band pair with the least coefficient of determination is chosen as the first band(s) in the final set.
3) Bandwidth increase of available bands in turn: Third, the bandwidth of all remaining bands is increased under the same constraints as in 1).
4) Select additional band: Fourth, the band with the least coefficient of (multiple) determination with respect to already selected bands is added to the set.
5) Repetition of step 3) and 4) until the user-specified number of final bands is reached.

A list of selected bands results, showing least correlated bands on top of the list, and the most redundant ones at the bottom.

The user may limit the input data by excluding certain bands from the algorithm either via direct selection or rejecting any bands outside the atmospheric windows determined in section 4.2.2. In addition, the user has the option to influence parts of the algorithm. In point 2, the selection of the

first band (pair) may be based on the highest SNR band value or the least coefficient of determination. The user may also guide the band expansion process (points 1 and 3) by introducing an 'equal bandwidth' constraint, which forces the bands in the set to be merged with an equal number of neighbours. What is more, the number of bands to be merged may be fixed to a certain value. In case of an 'unequal bandwidth' setting, bands are allowed to expand independently. Here the user has the option to set the maximum bandwidth to a specific value.

If a high number of output bands is chosen, together with a large maximum bandwidth, the number of available bands may be reduced quickly as wide bands may be selected for the first bands of the set. If no rows are left to fill the remaining places of the set, the algorithm iterates reducing the width of the first band by one row (if possible).

The algorithm options described above are illustrated in figure G.5 and may be used to trade between the influences of decorrelation and noise reduction on the final band set.
The algorithm was extended to include user-specified bands in the final band set (for example certain material absorption features), defined either by row number or wavelength interval. The algorithm then chooses the remaining bands with respect to these pre-selected bands and the user-specified algorithm options.

A computer program UBS (Unsupervised Band Selection) was written in IDL^{TM} (version 5.5 Win 32 x86) and ENVI^{TM} (version 3.5) using this algorithm and is described in Appendix section G.2. A flowchart of the algorithm is presented in figure 5.5, which is based on the routine described above.

Altogether, the UBS program is computationally highly efficient. The total execution time amounts to about 8 seconds if 19 optimal bands are to be selected from the 117 bands of the New Forest data set. The computation time mainly depends on the number of output bands specified by the user. If an output image was to be created from the optimal band set, the running time would be significantly increased. The latter is dependent on the size of the data set, but also on the number and width of the output bands. Reading and writing an output image of one of the bands of the New Forest data set adds about 5 seconds to the total computation time.

The UBS algorithm assumes that the band variables are normally distributed. However, as shown in section 5.2.3, this assumption is inappropriate for most of the New Forest and River Severn bands. To circumvent the normality assumption, an alternative unsupervised band selection method may be employed which is introduced in the next section.

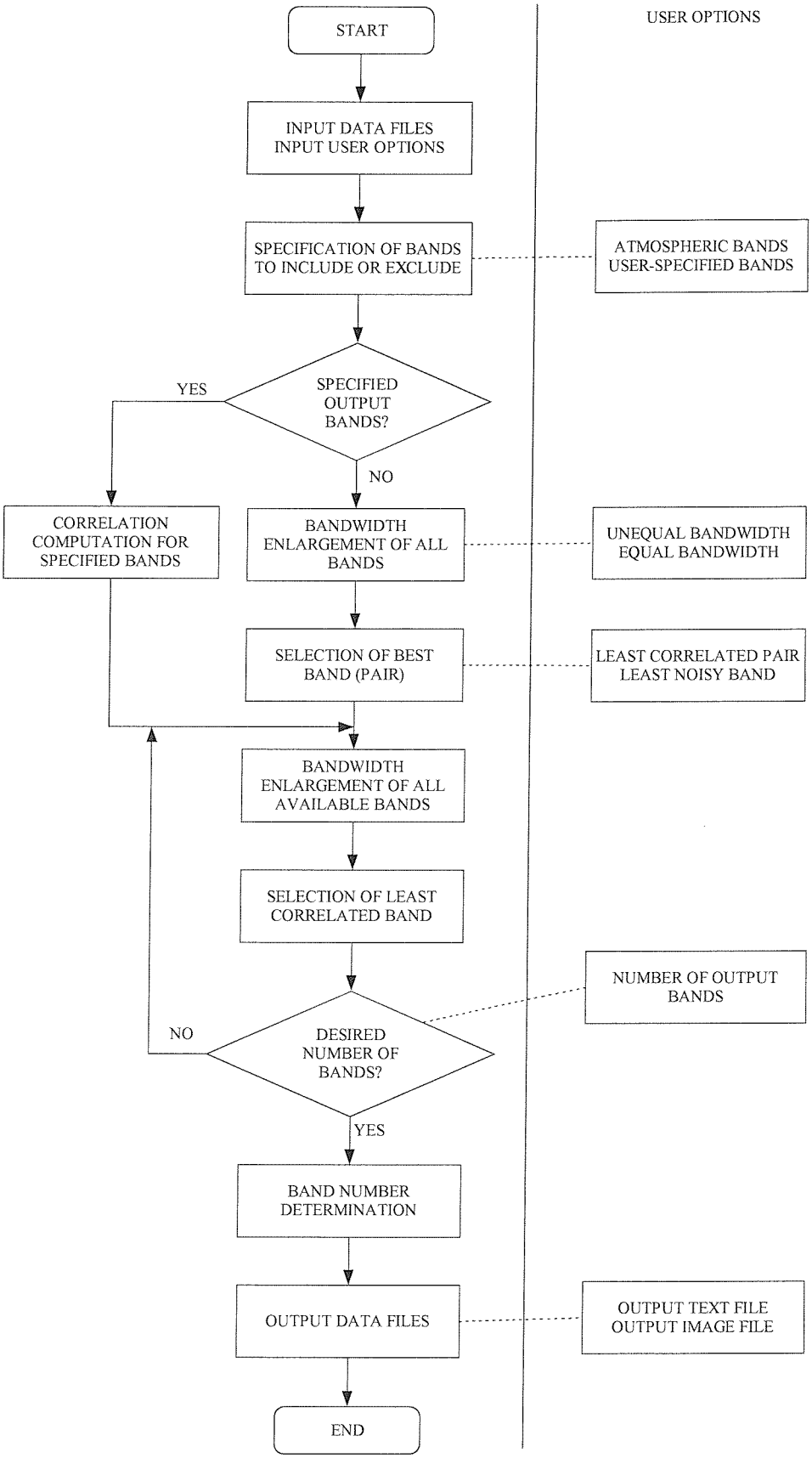


Figure 5.5: Flowchart of the UBS (Unsupervised Band Selection) program with algorithm options displayed on the right side.

5.6 Clustering-Based SBS (CSBS)

The supervised band selection algorithm presented in section 4.3 may be applied in an unsupervised way by using a clustering algorithm to define spectral classes from which class statistics may be computed. As described in section 3.3.1, informational classes may be chosen by the user to correspond to one or more spectral classes increasing the accuracy of the classification task. Two commonly used clustering algorithms are the K-Means and ISODATA (Iterative Self-Organising Data Analysis) procedures.

K-Means clustering assigns pixels iteratively to K clusters on a nearest-to-centre basis and updates the cluster centres because of this assignment at each iteration. The initial cluster centres are usually positioned randomly in feature space. For the K-Means program version available under ENVI™ (version 3.5), the user needs to supply the number of clusters and iterations. In addition, a pixel change threshold may be supplied that terminates the iterative process when the number of pixels in each cluster changes by less than the specified threshold. Other parameters include the maximum standard deviation and maximum distance error, which allow pixels to be classified only when they fall within the given limit. If the latter are not specified, all pixels are allocated to nearest clusters.

Many variants of the K-means algorithm exist to improve its efficiency, especially regarding the validity of the generated clusters. For example, the ISODATA algorithm allows new clusters to be created and existing ones to be merged or deleted between iterations, considering more feature space partitions than the basic K-Means. However, this improvement comes at the cost of needing to specify additional parameters. For the ISODATA algorithm available under ENVI™ (version 3.5), a minimum and maximum number of clusters have to be supplied apart from the maximum number of iterations and the pixel change threshold (see above). In addition, the minimum number of pixels needed to form a cluster (for cluster deletion), as well as the maximum class standard deviation (for cluster splitting), and the minimum distance between class means and maximum number of merge pairs (for cluster merger) must be defined.

The setting of the parameters of both the K-Means and the ISODATA algorithms is a non-trivial task, and no universal guidelines exist to the author's knowledge.

The optimal number of classes or clusters to choose for clustering depends entirely on the data and the class scheme defined by the user. However, the latter is unknown for unsupervised band selection. One approach would be to choose a relatively high number of classes in order to account for as many material subclasses as possible. The final user-defined informational classes may then consist of combinations of these basic subclasses. For adequate data representation, Jia and Richards (2002) used a cluster number of around 20 or higher in their experiments. However, with respect to the supervised band selection algorithm developed in chapter 4, bands are chosen so as

to differentiate between all given input subclasses, that is, also between those that form a larger information class. This may be counterproductive and unwanted by the user, as bands should discriminate only between information classes. From this point of view, the class number chosen should allow only for the most dominant material classes to be distinguished.

The number of iterations can be set high enough to ensure the generation of natural clusters, i.e. when practically no pixel re-assignments occur between subsequent iterations. Pixel change thresholds may be applied to stop the iterative process if the number of pixel re-allocations between iterations becomes insignificant (e.g. at about 5%).

The ISODATA algorithm allows for the deletion of clusters that do not contain a minimum number of pixels. This may be useful to control the number of pixels per cluster, which should be no less than 10 to 15 times the number of spectral bands in order to guarantee acceptable class statistics (see section 3.3.1). On the other hand, ISODATA further requires distance measures to be provided in data-specific units of DN for cluster merger and splitting, as well as the maximum number of merge pairs. These class- and data-dependent parameters are difficult to optimise without user interaction. For the unsupervised band selection method described in this section, a clustering procedure is required that works with a minimum amount of user interaction, i.e. a minimum number of parameters, and that is efficient with respect to computation time and MLC accuracy. Although the ISODATA has the advantage over the K-Means to allow for creating more representative clusters via cluster merging, deleting and splitting, more class- and data-specific algorithm thresholds have to be supplied. For that reason, the author chose the K-Means algorithm for the determination of class statistics being aware that the feature space partitioning may not be an optimal one.

To quantify the effect of number of iterations for the K-Means procedure on the accuracy of the resulting clusters in representing the feature space structure, the MLC accuracy was calculated for both the New Forest and River Severn data sets using all bands and the various cluster sets as class definitions. Figure 5.6 and 5.8 display the MLC accuracy as a function of number of clusters for a varying number of iterations. They show a general improvement of MLC accuracy with increasing number of iterations (for the New Forest data set only for a small cluster number). In addition, it may also be seen that for an increasing number of clusters, the classification performance tends to decrease steadily. This may be explained by the fact that the separation of a large number of small and overlapping clusters will be less accurate than the partition of the dominant clusters in feature space.

To see the effect of varying class number for the K-Means algorithm on the MLC accuracy of the given classification tasks, different band sets were generated with the SBS algorithm on the basis of the clustering results. The New Forest and River Severn scenes were then classified using these band sets and their class definitions from chapter 3. In figures 5.7 and 5.9, MLC accuracy is plotted against the number of bands for different class numbers used during clustering with one iteration

cycle. In figure 5.7 (New Forest data), the MLC accuracy seemed to be noticeably increased for low band numbers when using a larger class number. But this difference becomes less pronounced as the number of output bands increases. For the River Severn data set (figure 5.9), this did not seem the case, as nearly all band sets performed equally well for different cluster numbers.

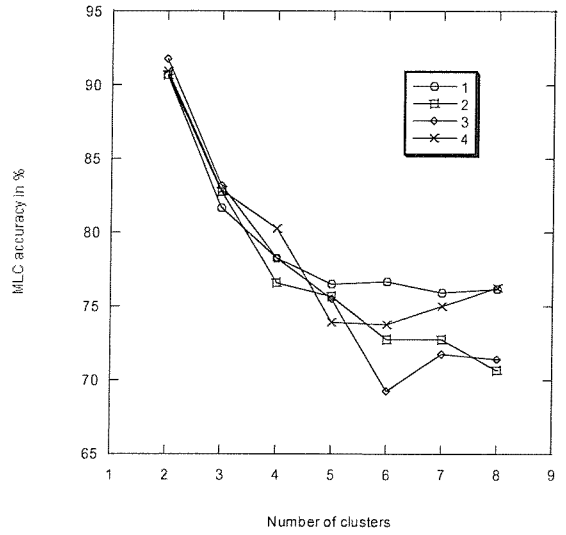


Figure 5.6: MLC accuracy plotted against the number of clusters for varying number of iterations (1 to 4) for the New Forest data set.

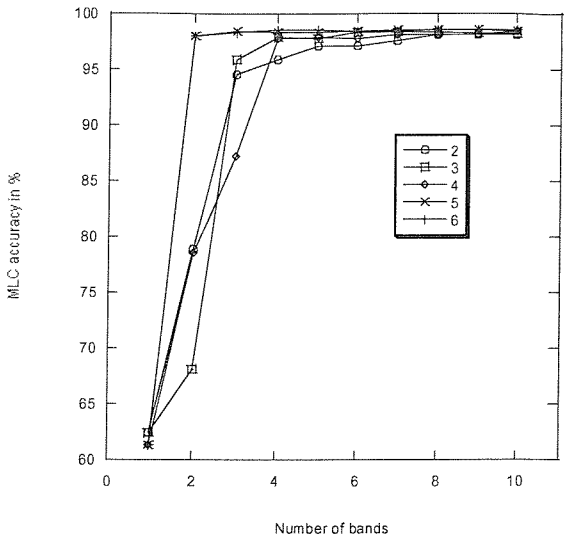


Figure 5.7: MLC accuracy plotted against the number of bands selected by the clustering-based SBS algorithm for varying number of clusters (2 to 6), one iteration cycle and the New Forest class and data set.

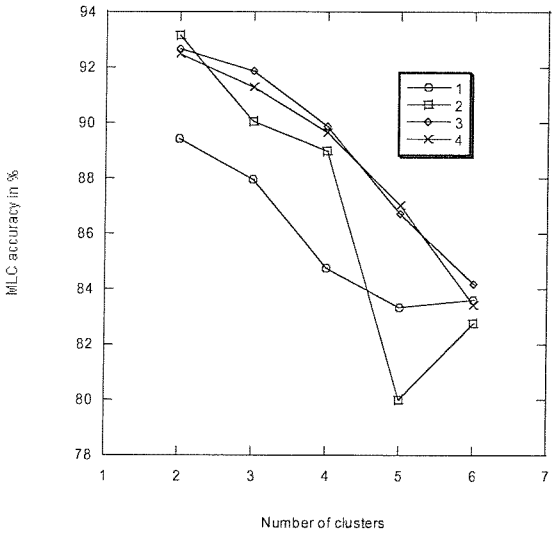


Figure 5.8: MLC accuracy plotted against the number of clusters for varying number of iterations (1 to 4) for the River Severn data set.

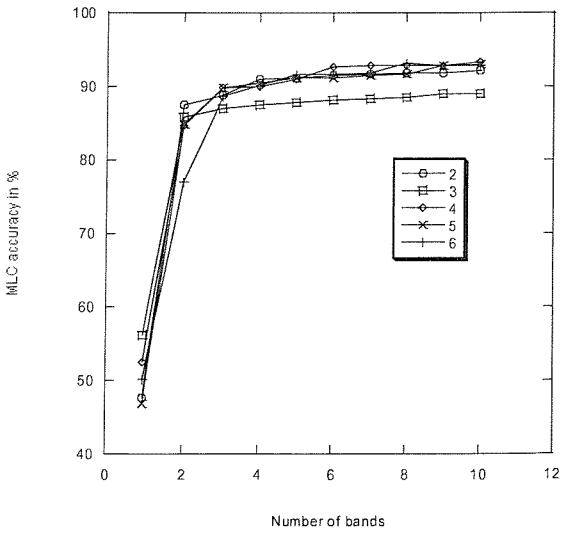


Figure 5.9: MLC accuracy plotted against the number of bands selected by the clustering-based SBS algorithm for varying number of clusters (2 to 6), one iteration cycle and the River Severn class and data set.

From a computational point-of-view, the execution time of the clustering algorithm in combination with the class statistics and SBS computation should be minimised in order to be suitable for in-flight application. This implies that user interaction has to be either reduced to a minimum or completely eliminated. Regarding the parameter settings for the clustering algorithms, the data analyst may not be able to determine the optimal configuration in a multiple-run trial-and-error procedure by visual inspection of the classified output image. Consequently, default parameter settings need to be provided that are efficient in both computation time and accuracy of the MLC task at hand.

A computer program was written in IDL™ (version 5.5 Win 32 x86) and ENVI™ (version 3.5) that creates clusters using the K-Means clustering routine in ENVI™ and calculates their statistics (see details of use in Appendix G.5). The program classifies the entire image into a pre-specified number of clusters, and samples classes randomly over the entire image until a sufficient number of samples (10 times the number of spectral bands to be used) are found to calculate representative class statistics.

The execution time of the clustering part of the program may be decreased considerably by resizing the original imagery by a factor so as to have just enough sample pixels per class for adequate class statistics. The factor may be computed as the square root of the ratio between the number of samples needed and the total number of image samples, where the number of samples needed may be estimated as the product of the number of classes, the number of bands times 10, and a security factor of 2. That is, for example for the New Forest data set, the number of samples needed may amount to 20 (chosen class number) times 117 (band number) times 10 times 2 (security factor), which is equal to 46800. The total number of samples of the New Forest imagery is the product of the number of rows (1280) times the number of columns (512), or 655360, assuming no masking. The resize factor then equals to 0.3. The sub-sampling is performed using the nearest-neighbour method in ENVI™. This resizing routine is acceptable when the H-resolution case applies for the given classes and when at least 1 over the resize factor number of adjacent samples (8 for the example above) are available for each class in row or column direction. It also assumes that classes are equal in size and uniformly distributed over the entire imagery.

It can be shown that the execution times of the clustering and the cluster statistics calculations depend exponentially on both the number of clusters and the number of iterations. As, for both data sets examined here, the increase in cluster number or number of iterations did not improve considerably the class data representation by the resulting clusters, a low cluster number of 2 to 5 (corresponding to the number of dominant material classes in the scene) coupled with one iteration cycle were chosen as the recommended settings for the clustering routine. For the New Forest data set, choosing five clusters achieved high class discrimination ability also for small band numbers, while still giving a computation time less than a minute (with resized New Forest imagery, factor 0.3).

The histograms for the resulting clusters for some of the bands of the River Severn and New Forest data set are displayed in table F.3 and F.4, respectively. Two clusters were generated for the River

Severn and five for the New Forest data set (one iteration cycle each). Table F.3 shows that some bi-modality appeared for the first cluster in some bands of the River Severn data. The histograms of the clusters generated for the New Forest data generally seem to fit the normal curve better than those of the River Severn clusters.

5.7 Algorithm Evaluation

This section aims to assess the value of the proposed unsupervised algorithms and its results 1) quantitatively with respect to its reliability, consistency, and effectiveness, 2) qualitatively, and 3) in terms of its usefulness for real-time in-flight applications.

5.7.1 Reliability

The reliability of the UBS and CSBS band selection algorithms was judged by questioning the appropriateness or accuracy of its assumptions and routines.
Both the UBS and CSBS algorithms share the same image acquisition method as described in section 4.3.1 for the SBS method, and therefore equally assume the H-resolution case, as the image data recorded in repeated acquisitions for the same scene should result in similar pixels. In addition the scene is assumed not to change between acquisitions, which may be realised by reducing the time gap to a minimum.

For a successful application of the UBS algorithm, which is based on the correlation coefficient, data need to meet the following criteria:

- The sensor bands follow a normal distribution. Generally, the assumption of a single distribution for a remotely-sensed band image is hard to achieve unless the image is depicting a single material type only. For example, in case of the New Forest data set, vegetation was the dominant target class, and the normality assumption was shown to be more appropriate than for the River Severn data set, which consisted of two spectrally different materials, vegetation and water-covered mud (see section 5.2). Users should therefore aim to collect data from one material type only, e.g. vegetation, soil or water.
- Band correlations are significant, that is, enough image samples are provided. This is usually the case for high-resolution remotely-sensed images from current airborne imaging spectrometers such as HyMAP or CASI-2.
- If bands are dependent on each other, then only in a linear fashion. The correlation coefficient is a measure of linear dependency only and fails to quantify, for example, quadratic relationships. As detectors of sensors such as HyMAP or CASI-2 typically respond linearly to the incoming signal, the nature of the relationship between bands depends entirely on the incoming spectral radiance signal. If an image scene contains mainly target

materials with smooth spectral curves and relatively low standard deviation, other than linear dependencies among bands are highly unlikely.

- The data set does not include very noisy bands. Noisy bands need to be either excluded from the band selection procedure or smoothed with some spatial filter beforehand. This is because noisy bands can result in a low correlation coefficient with other bands and may therefore be chosen as optimal bands in the decorrelation process.

For the clustering-based CSBS algorithm, the same assumptions and comments apply as for the SBS algorithm (see section 4.4.2). In addition, it is assumed that the parameter settings that were shown to be effective for the New Forest and River Severn data set are also valid for other data sets. Further tests with other data sets and classification schemes need to be performed to verify this result.

5.7.2 Consistency

The UBS algorithm is repeatable, assuming no two correlations are equal in the band location and number determination, and band broadening sub-routines. The ordering of bands within the band location determination subroutine depends on the bands already selected for the set. Removing one of the first selected bands puts an end to the validity of the order of subsequent bands.

The clustering part of the unsupervised version of the SBS algorithm may result in different cluster distributions according to the parameters (number of clusters and iterations) chosen by the user. If default parameter settings were to be chosen repeatedly, the outcomes might still differ from each other, as the initial cluster means were randomly selected by the K-Means procedure.

5.7.3 Effectiveness

Sub-optimality of the algorithm

The sub-optimality of the unsupervised band selection algorithms UBS and CSBS was estimated by comparing the MLC accuracy of their band sets with that of a band set derived from an exhaustive search with the MLC accuracy measure as criterion function. The comparison was limited to the first three dimensions for the exhaustive search to be still computationally feasible. As the UBS offers the possibility to start the set with the least correlated or least noisy band (pair) (see section 5.5.4), the selected sets from both options were considered. Bands were not increased in width for both the UBS and CSBS methods to be comparable to those of the exhaustive procedure. Figures 5.10 and 5.11 present the MLC accuracy of the different band sets for the New Forest and River Severn data set, respectively.

For the New Forest data set (figure 5.10), the band subset selected with the UBS LC (least correlated band pair first) algorithm results in a 50% loss in accuracy for the first dimension, while the corresponding loss is limited to around 23% for both the UBS LN (least noisy band first) and the CSBS band set. For two set dimensions, the UBS LN bands give a 15% accuracy loss, while the UBS LC and CSBS bands are with about 6% accuracy loss not statistically different from the exhaustive band set. In the third dimension, all three methods produce band sets of similar insignificant accuracy loss (around 3.6%).

For the River Severn data set (see figure 5.11), the UBS LC band set clearly outperforms the sets of both the UBS LN and CSBS in the first and third dimension, where its accuracy loss is insignificant at around 7% and 2%, respectively. In the second dimension, all band sets gave a similar, but significant accuracy loss of around 6.5%.

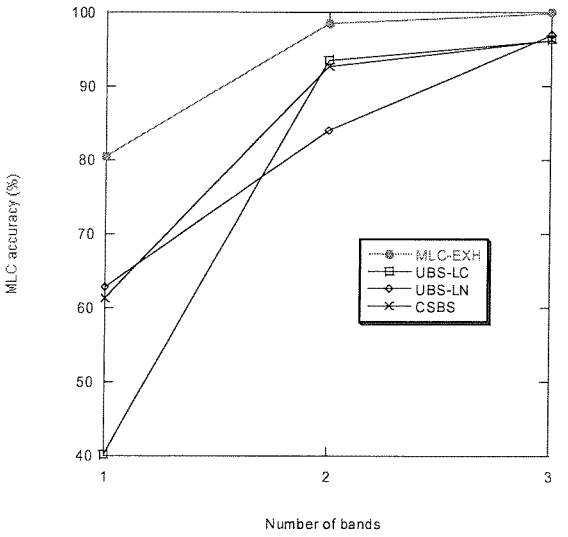


Figure 5.10: MLC accuracy of optimal band sets derived using an exhaustive search with MLC accuracy (MLC-EXH), the UBS (start with least correlated, LC, and noisy, LN, bands), and CSBS algorithm for the New Forest data.

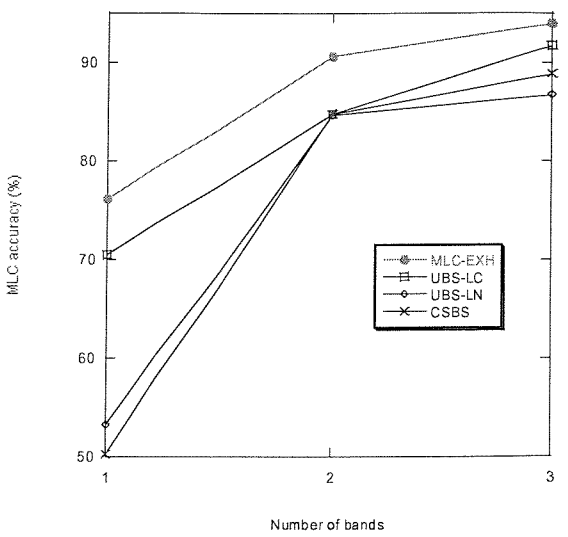


Figure 5.11: MLC accuracy of optimal band sets derived using an exhaustive search with MLC accuracy (MLC-EXH), the UBS (start with least correlated, LC, and noisy, LN, bands) and CSBS algorithm for the River Severn data.

The results from these two data sets demonstrate that the UBS LC method gives optimal band selection results, except for the first dimension regarding the New Forest data set. This may be explained by the fact that UBS LC selects the least correlated band pair first, and not the single best band. For the first band, both the UBS LN and CSBS methods result in acceptable accuracies for the New Forest data set, but fail to do so for the River Severn data set. The suboptimal performance of the UBS LN in the latter case may be explained by the use of a less accurate noise estimation method.

Comparison of 'optimal' UBS and CSBS band sets with 'established' band sets

The performance of the 'optimal' band sets output by the UBS ('unequal bandwidth - maximum bandwidth 4 rows) and CSBS (unequal bandwidth) algorithms were compared with the one of different simulated 'vegetation' and 'coastal' band sets from current satellite and airborne sensors (for details see section 4.4.4). The maximum bandwidth value was chosen lower for the UBS than for the CSBS algorithm, as in contrast to the MLC criterion based CSBS, the expansion of bands increases the SNR criterion of the UBS in most cases, resulting in very wide bands. Figures 5.12 and 5.13 display the band set performances for the New Forest and the River Severn class and data sets, respectively.

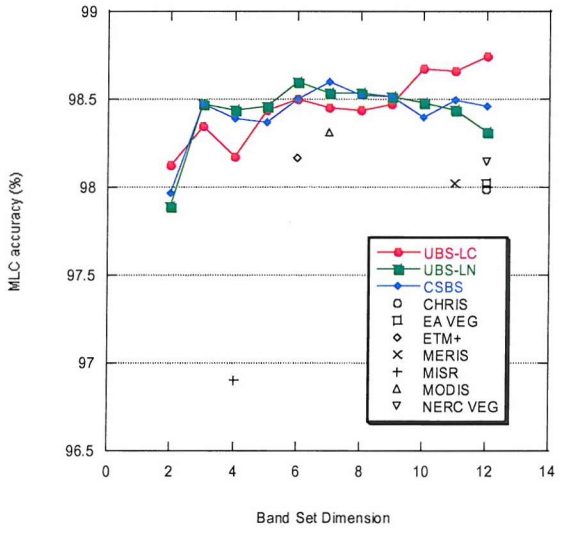


Figure 5.12: MLC accuracy of the 'optimal' UBS (start with least correlated, LC, and noisy, LN, bands) and CSBS band sets compared with the one of satellite and airborne band sets for the New Forest data.

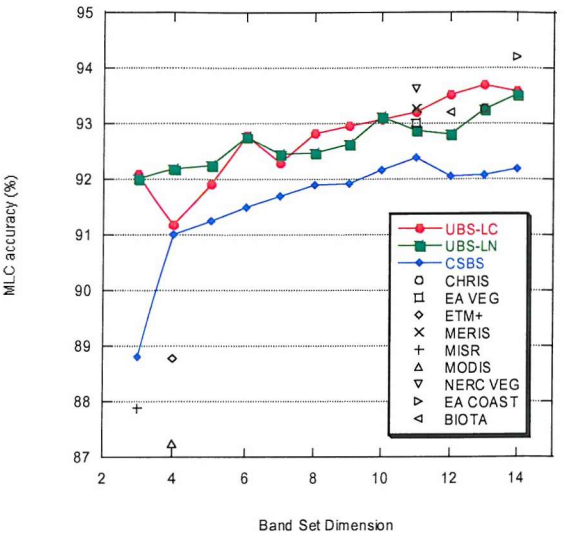


Figure 5.13: MLC accuracy of the optimal UBS (start with least correlated, LC, and noisy, LN, bands) and SBS band sets compared with the one of satellite and airborne band sets for the River Severn data.

For the New Forest data set, the UBS and CSBS band sets achieved higher MLC accuracy than any of the simulated band sets for corresponding dimensions. However, only for the UBS LC band set, most differences in classification accuracy can be shown to be statistically significant (with the exception of the ETM+ and MODIS band set) and to range from 1.3% for MISR to 0.6% for the NERC vegetation band set. Both the UBS LN and CSBS band set gave a 1.5% significantly higher accuracy than the MISR band set.

For the River Severn data set, the UBS and CSBS band sets resulted in higher MLC accuracy than the 'established' band sets for small band set dimensions (3 to 4), but in similar or less accuracy for larger band set dimensions (11 to 14). For the UBS band sets, the differences in accuracy for only

small dimensions are significant (3 to 4), ranging from 5.4% for MODIS to 2.6% for ETM+. For the CSBS band set, most differences in accuracy were significant (except for EA vegetation, MERIS and MISR), including its underperformance with respect to the CHRIS, NERC Vegetation, EA Coastal and BIOTA band sets (maximum difference 2.2%).

Band number evaluation

The optimal band number was estimated for the UBS algorithm as the number of bands already in the set which achieve an explained variance or coefficient of (multiple) determination of higher than 95% for the actual band to be added to the set. It represents an upper limit to the intrinsic dimensionality (ID) of the data set, which was estimated to range between one and two, and one and three for the New Forest and River Severn data set, respectively (see section 3.5).

Table 5.3 displays the coefficient of (multiple) determination for the first 5 selected bands (least correlated bands first, LC, and least noisy band first, LN, UBS options). Using a 95% threshold on the coefficient of determination, the optimal band number may be obtained for the New Forest data set as 4 and 3 for the LC and LN options, respectively, while for the River Severn data set it may be determined as 3 for both algorithm options. This result overestimates the upper derived ID limit of the New Forest data set by one and two, but gives an exact reflection of the upper estimated ID value of the River Severn data set. That is, the band measure could be used as an upper limit estimate to the ID of the given data sets.

Table 5.3: Coefficient of determination for the first five bands of the UBS (LC and LN) set, and the PMATD for the CSBS set for the New Forest and River Severn data sets.

Band number	Coefficient of determination (%)				PMATD (%)	
	UBS LC		UBS LN		CSBS	
	New Forest	River Severn	New Forest	River Severn	New Forest	River Severn
1	0	0	0	0	80.13538	96.77774
2	0.000023	0.000005	18.363762	0.000121	88.36081	98.43816
3	67.20255	92.72588	72.390701	91.415230	92.81153	99.99999
4	91.32729	96.33556	97.102341	98.054764	94.02414	100
5	97.83753	97.93714	97.651962	98.270760	96.44305	100

The definition of the optimal number of bands for the CSBS algorithm is identical to the one of the SBS algorithm, that is, the dimension of the smallest set of bands that achieves a PMATD of at least 95%. The PMATD values of the first five CSBS bands are displayed for the New Forest and River Severn data sets in table 5.3, and an IDD estimate of 5 and 1, respectively, can be read from the table. These estimates cannot be compared directly with the IDD approximations made in chapter 3, as the IDD is intrinsically dependent on the defined class set and the extracted clusters do not correspond in number and spectral characteristics to the user defined classes. The IDD estimate is therefore only valid for the actual set of classes being used for PMATD estimation.

5.7.4 Qualitative Evaluation

The first six bands of the optimal band sets derived with the UBS algorithm for the New Forest and the River Severn data sets are listed in table 5.4 and plotted against a vegetation spectrum in figures 5.14 and 5.15, respectively. Only the band set derived from the least-correlated-first option was investigated as an example of the correlation-based algorithm.

For the New Forest UBS band set, the best three bands chosen by the algorithm stem from the visible, NIR and SWIR wavelength ranges. The first band is located on the NIR plateau including parts of the 942 nm water absorption feature. The second band is positioned right next to the 1865 nm water absorption feature. According to Kumar *et al.* (2001) increased water leaf content also decreases reflectance in wavelength regions adjacent to water absorption features. That is, both first and second band are sensitive to canopy moisture content. Band 3 is centred on the peak of the green reflectance feature, responding to the amount of chlorophyll within the vegetation canopy. So does band 5, a relatively broad band placed over the entire red-edge feature. Band 4, a narrow-band equivalent to the Landsat ETM+ band 5 (1550 – 1750 nm), and band 6, a narrow band positioned at 1330 nm between the two water absorption features at 1135 and 1379 nm, are both sensitive to canopy structure and water content. Most of the image classes (grassland, bracken, valley mire, dry, humid and wet heath) differ mainly in their canopy structure and moisture content. This is reflected in the UBS selected bands.

With regard to the River Severn data set, the first band chosen by the UBS algorithm is centred on the green reflectance peak, responding to the chlorophyll amount in the canopy. The second best band is located on the NIR plateau, which is sensitive to canopy structure and helps to delineate the land-water interface. Band 3 is positioned in the middle of the red-edge feature, while band 5 lies on the bottom of it. Both bands respond to chlorophyll variations in the canopy and help to differentiate between different vegetation types. Band 4 is located on the NIR plateau adjacent to the 942 water absorption feature, and responds to both canopy structure and water content. Band 6 is situated over the blue vegetation absorption feature, potentially used for delineating water and vegetation surfaces. The main salt-marsh classes in the River Severn data set include high, mid, and pioneer marsh classes, bare rock and mud. The vegetation classes may be differentiated from each other and the Bare Rock and Mud classes mainly by their structure and chlorophyll content, which is mirrored in the selection of the first six bands.

Figures 5.16 and 5.18 display colour composites of the first three bands chosen for the New Forest and River Severn data set, respectively. Both images show a high contrast in colour between the different vegetation classes involved. For the New Forest data and the given RGB band combination, the asphalt road stands out as a purple colour, whereas bracken occupies the light green and dry heath the pink-red colour. Wet and humid heath and valley mire may be distinguished from different tones of green (see classified image in figure 3.5). Regarding the River Severn colour

Table 5.4: First 6 optimal bands selected by the UBS algorithm (least correlated first – unequal bandwidth option with maximum 4 rows width) for the New Forest and River Severn data set.

Band number	New Forest				River Severn			
	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]
1	930.1	34.35	912.925	947.275	556.8	31.2	541.2	572.4
2	2017.7	77.35	1979.025	2056.375	807.9	31.8	792	823.8
3	546.45	61.6	515.65	577.25	708.55	31.7	692.7	724.4
4	1672.7	40.55	1652.425	1692.975	914.7	31.5	898.95	930.45
5	699.9	61.4	669.2	730.6	685.6	16.4	677.4	693.8
6	1330.2	29.9	1315.25	1345.15	489.55	30.9	474.1	505

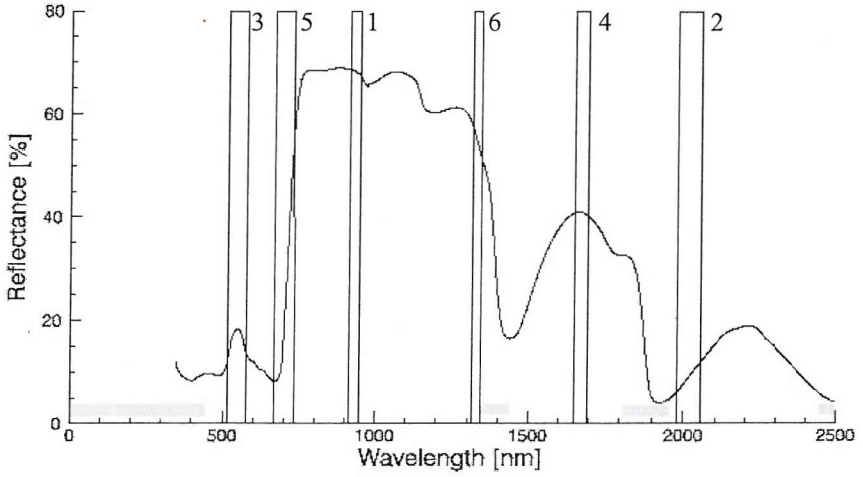


Figure 5.14: First six bands selected by UBS for the New Forest data set (see table 5.4). Wavelengths not available for band selection are indicated by grey bars.

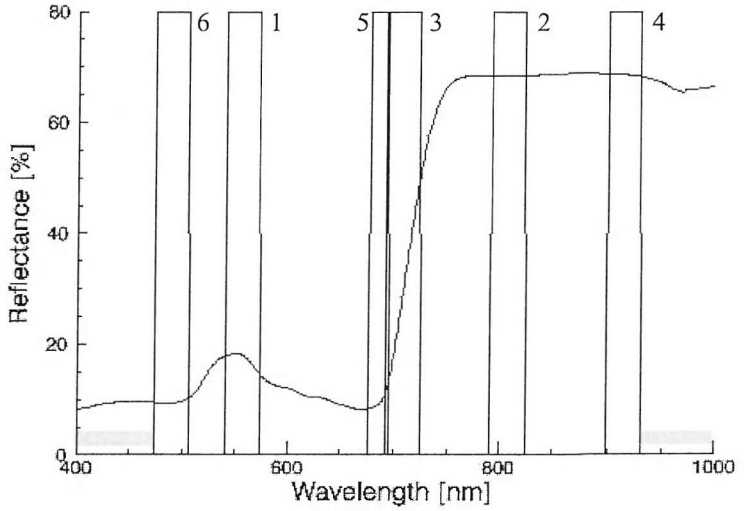


Figure 5.15: First six bands selected by UBS for the River Severn data set (see table 5.4). Wavelengths not available for band selection are indicated by grey bars.

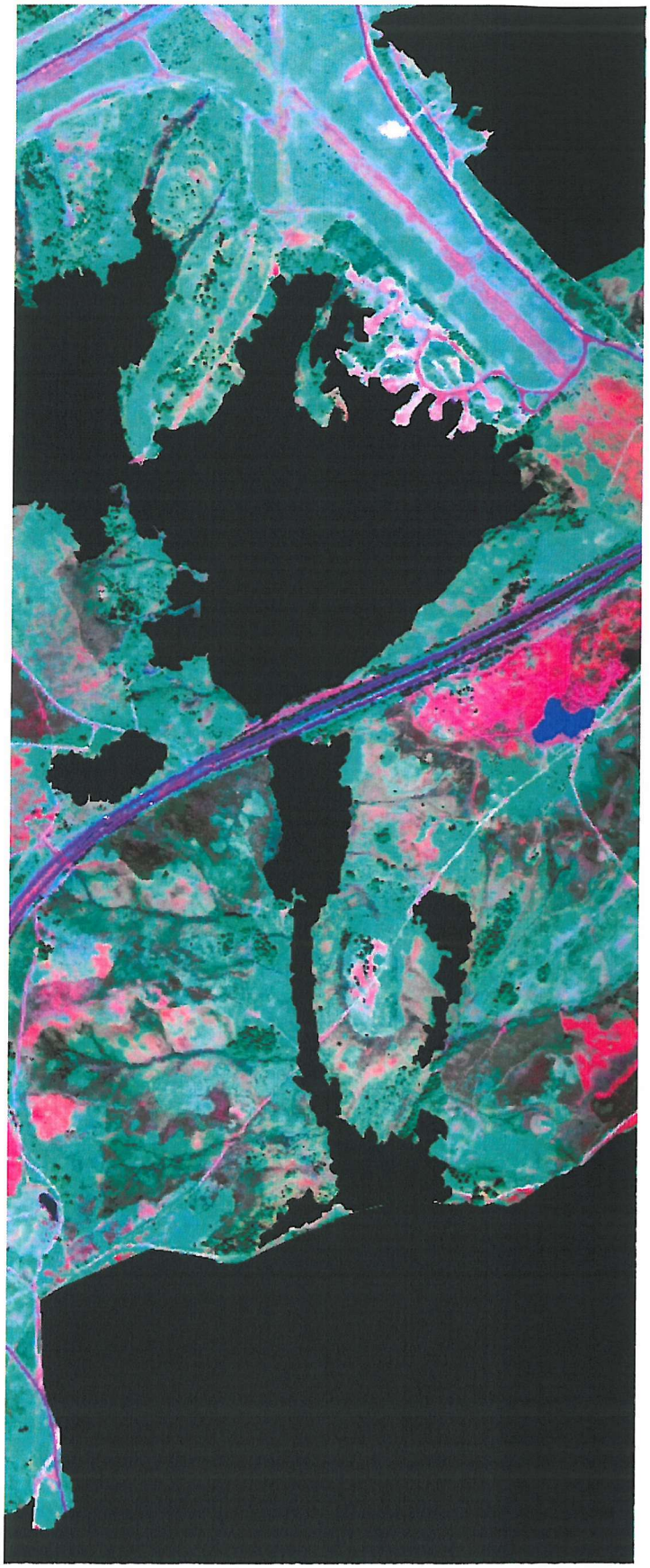


Figure 5.16: Masked HyMAP New Forest data displayed using the first three optimal bands output by the UBS algorithm (see table 5.4, R = band 2, G = band 1, B = band 3).

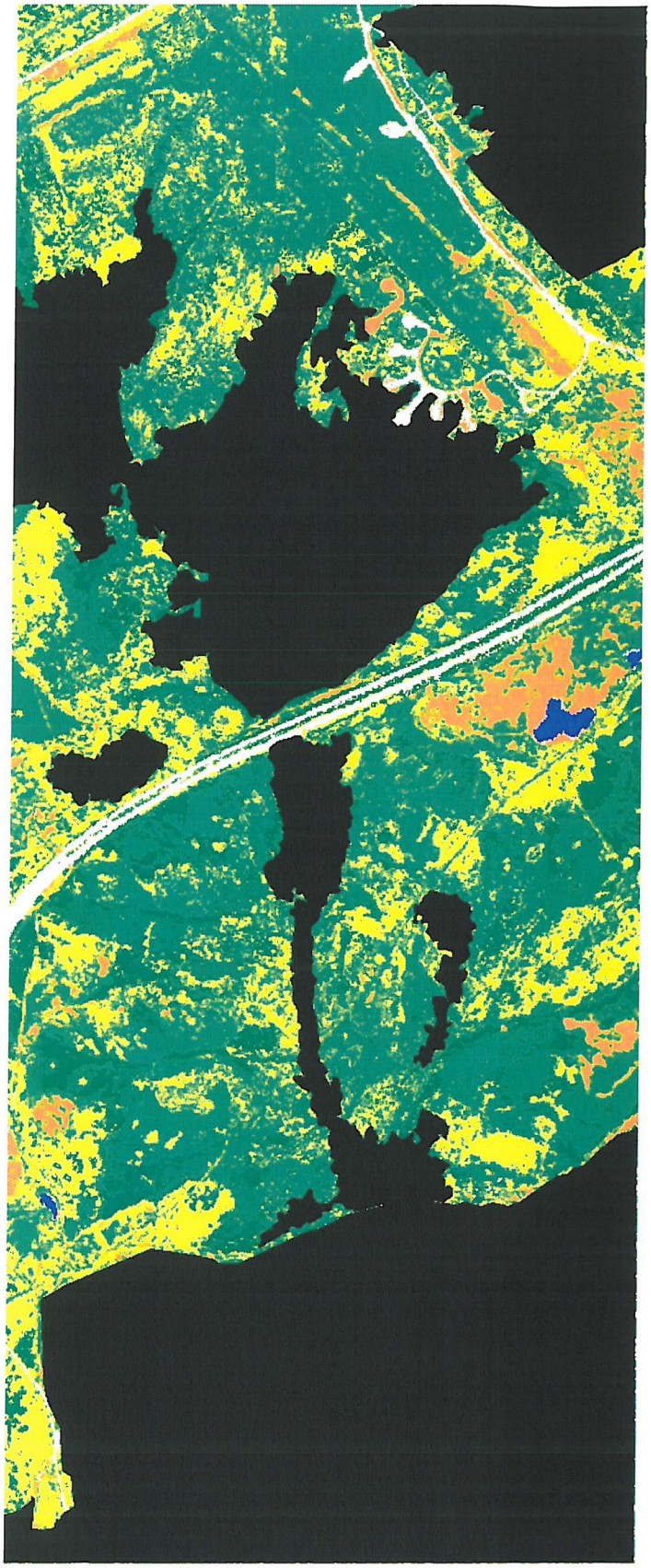


Figure 5.17: New Forest Maximum Likelihood Classification result using the first six bands selected by the UBS algorithm (see table 5.4; Lake, blue; Asphalt, white; Bracken, yellow; Dry Heath, orange; Grassland, brightest green; Humid Heath, bright green; Wet Heath, green; Valley Mire, dark green).

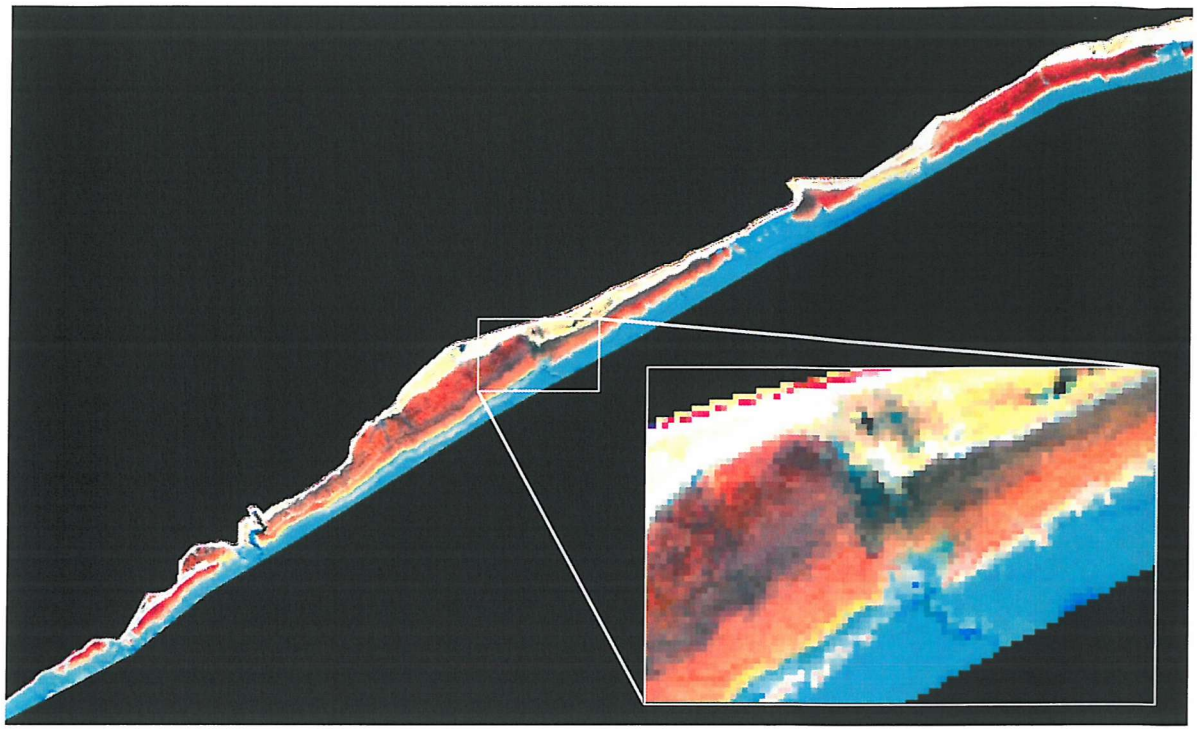


Figure 5.18: Masked CASI River Severn data displayed using the first three optimal bands output by the UBS algorithm (see table 5.4, R = band 2, G = band 3, B = band 1).

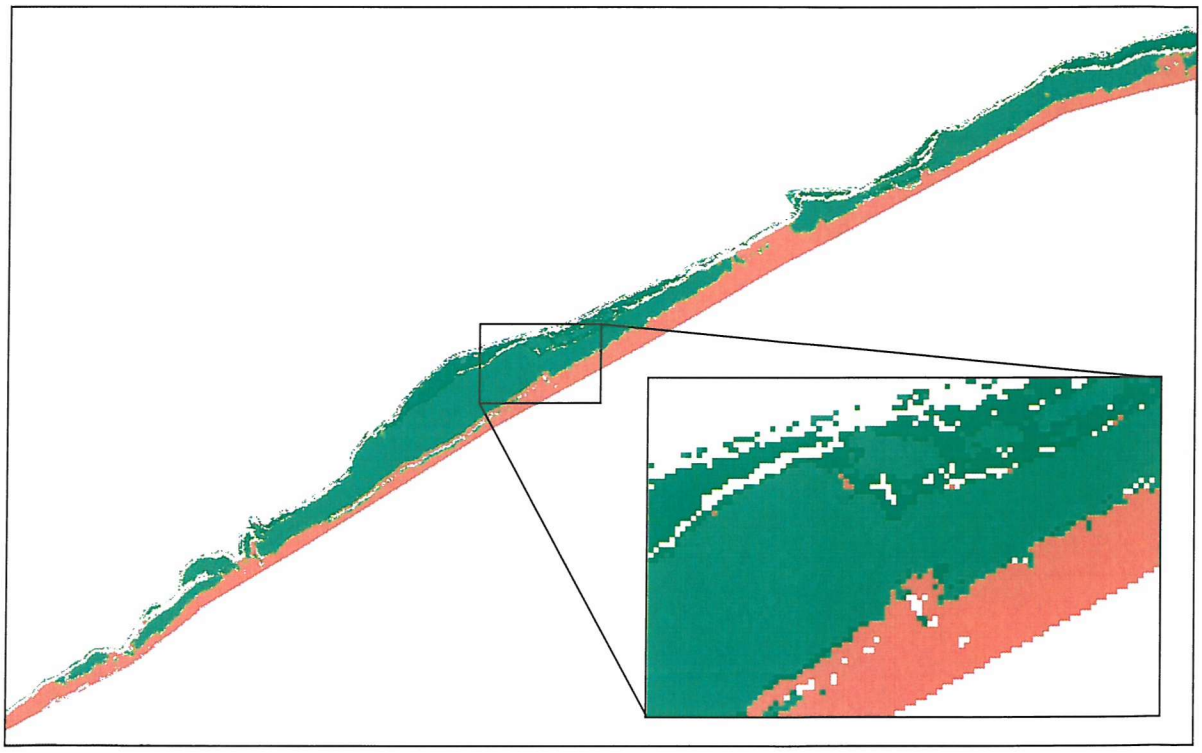


Figure 5.19: River Severn Maximum Likelihood Classification result using the first six bands selected by the UBS algorithm (see table 5.4; Bare Rock, white; Pioneer Marsh, bright green; Mid Marsh, green; High Marsh, dark green; Bare Mud, brown).

composite of the optimal three bands, pioneer, mid and high marsh vegetation is coloured in bright orange - red, red – blue, and yellow, respectively. Bare rock appears white in the image and bare mud blue (see classified image in figure 3.4). Figures 5.17 and 5.19 display the resulting map of the New Forest and River Severn classification task, respectively, using the six bands selected by the UBS algorithm (see table 5.4).

The band sets listed in table 5.5 have been generated by the CSBS algorithm for the New Forest and River Severn data sets. They are plotted against a vegetation spectrum in figures 5.20 and 5.21 for the two data sets. For the New Forest data set, the CSBS band set reflects more the variations in canopy structure and water content (bands 1 to 3, 5 and 6) than in the chlorophyll content (band 4), similarly to the band set selected by UBS. For the River Severn data set, the CSBS gives a band set similar to the UBS result, with bands potentially exploiting variations in canopy chlorophyll (bands 1 and 4 to 6) and canopy structure (bands 2, 3). The corresponding classification maps for the River Severn and New Forest data sets, derived by using the six bands selected by the CSBS algorithm (see table 5.5), are shown in figures 5.22 and 5.23, respectively.

Table 5.5: First 6 optimal bands selected by the CSBS algorithm (1 iteration, unequal bandwidth) for the New Forest (5 clusters) and River Severn (2 clusters) data set.

Band number	New Forest				River Severn			
	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]
1	1196.1	18.1	1187.05	1205.15	742.9	115.8	685	800.8
2	2136.7	20.4	2126.5	2146.9	807.9	16.4	799.7	816.1
3	844.4	15.8	836.5	852.3	888.1	8.8	883.7	892.5
4	700.1	30.95	684.625	715.575	482.15	16.1	474.1	490.2
5	2172.2	19.6	2162.4	2182	497	16	489	505
6	1517.2	16.2	1509.1	1525.3	511.85	16.1	503.8	519.9

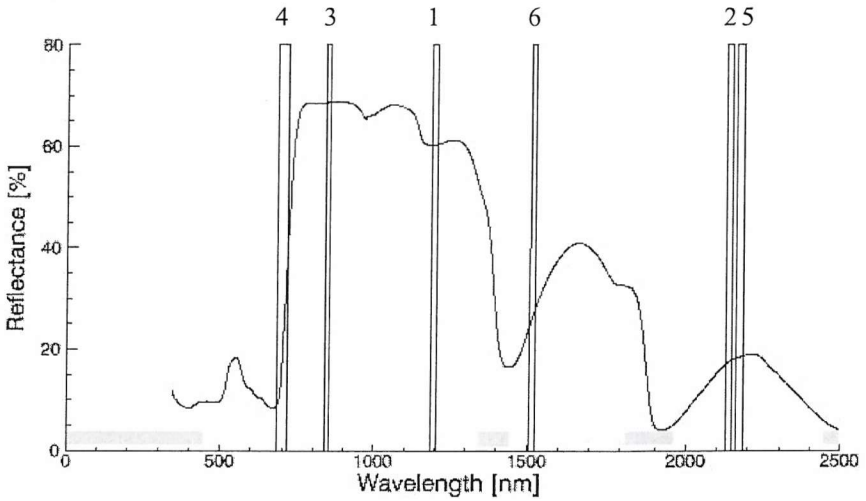


Figure 5.20: First six bands selected by CSBS for the New Forest data set (see table 5.5). Wavelengths not available for band selection are indicated by grey bars.

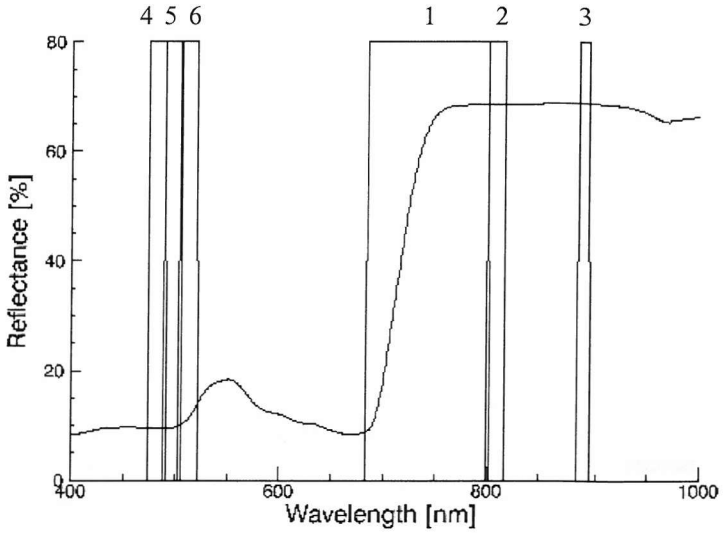


Figure 5.21: First six bands selected by CSBS for the River Severn data set (see table 5.5). Wavelengths not available for band selection are indicated by grey bars.

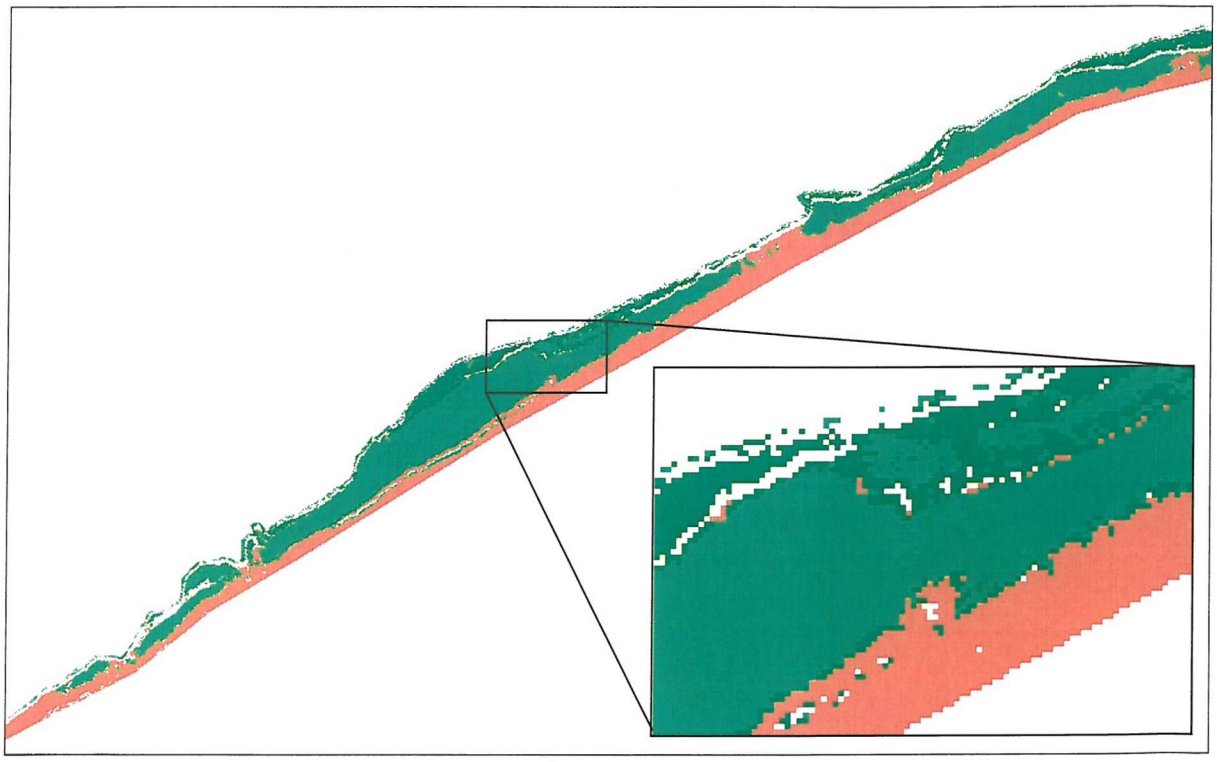


Figure 5.22: River Severn Maximum Likelihood Classification result using the first six bands selected by the CSBS algorithm (see table 5.5; Bare Rock, white; Pioneer Marsh, bright green; Mid Marsh, green; High Marsh, dark green; Bare Mud, brown).

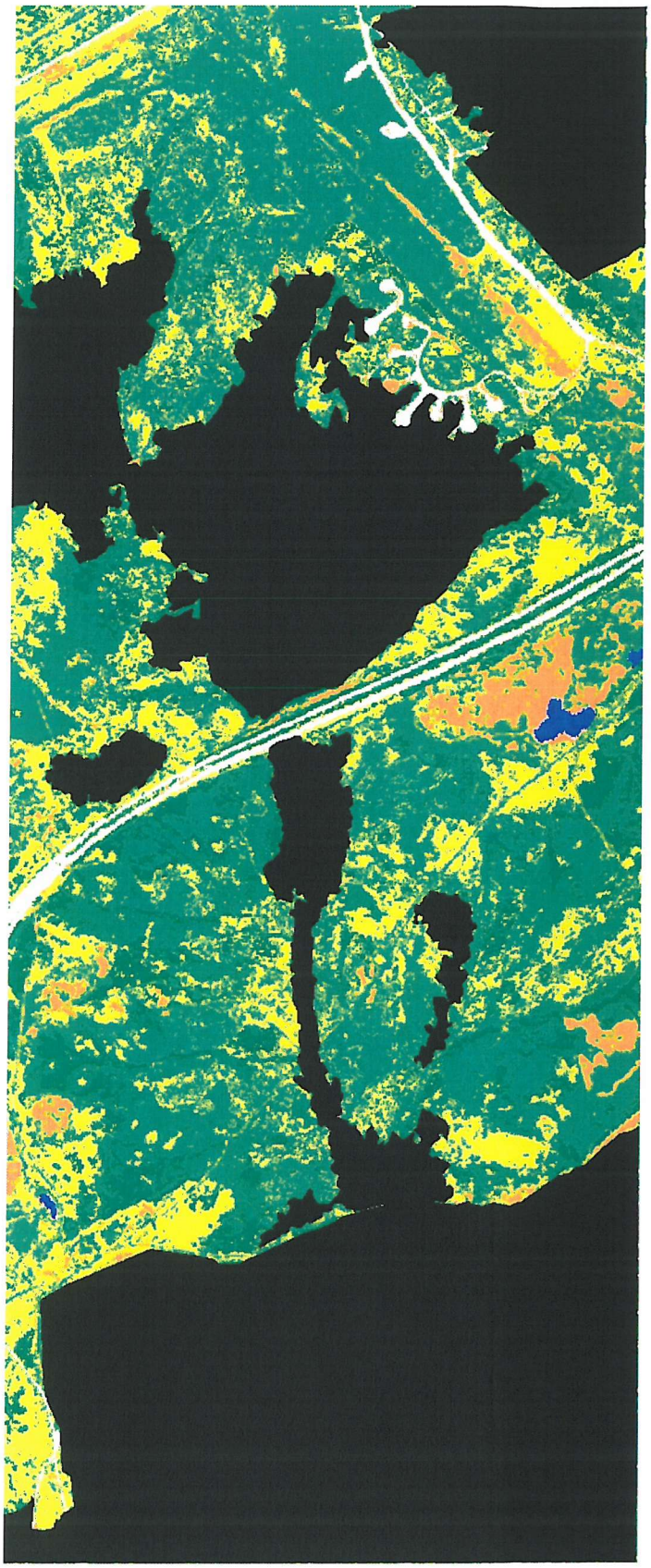


Figure 5.23: New Forest Maximum Likelihood Classification result using the first six bands selected by the CSBS algorithm (see table 5.5; Lake, blue; Asphalt, white; Bracken, yellow; Dry Heath, orange; Grassland, brightest green; Humid Heath, bright green; Wet Heath, green; Valley Mire, dark green).

5.7.5 In-flight Application

The unsupervised band selection methods presented in this chapter were specifically designed for a real-time in-flight execution between the acquisitions of successive flight-lines. As the aircraft may only need 3 to 5 minutes to turn around before re-starting to scan the scene in reverse flight direction (F. Tadina, NERC ARSF, 2003, personal communication), the speed of the algorithm is crucial. The following paragraphs aim to give an overview of the execution time involved in performing band selection with UBS.

In contrast to the supervised SBS algorithm presented in chapter 4, UBS does not depend on a time-consuming identification of the scene classes on the imagery. However, as the UBS routine is implemented in ENVI™, the raw binary image has to be transformed into an ENVI™ supported file format. The program for radiometric correction of CASI-2 raw data, Radcorr (Version 2.2 for RedHat Linux 6.2), generates files in the ENVI™ supported pix-format. The program, which is provided by the instrument manufacturer, ITRES Research Ltd., also corrects the image data for dark current, scattered light, frame shift smear and electronic offset, and then transforms the data DN values into spectral radiance (ITRES, 2001). The computation time amounted to about 16 seconds for a typical CASI-2 15-band full-swath image of 1,500 lines using a 266 MHz AMD-K6 processor with 64 MB RAM. Linearly extrapolating this execution time to a data set of equal number of lines but with 288 bands and a smaller swath width 101 pixels (see chapter 1), a value of about 61 seconds would result. Using a higher speed computer, for example, with a 2 GHz processing speed, the duration of the program run could be reduced to about 8 seconds.

The Radcorr routine may also be used to extract about 50 lines of dark data from the raw image file by spatially subsetting the image from line 50 to line 100. That is, dark data will then be calibrated to spectral radiance as well. This process should take less than 5 seconds with a 2 GHz processor.

After the image has been radiometrically corrected, image data statistics need to be calculated with the Datastats program (see section G.4) as these are input into the UBS algorithm. The execution time of the program was calculated for the New Forest data set and extrapolated to 288 bands. To achieve a computation time of under a minute, the number of samples in the image should be smaller than 60,000 samples. A typical image of 1,500 lines by 101 columns would give 151,500 samples which equals to 3.5 minutes.

To reduce the computation time, a quicker processor could be employed (here 1 GHz Intel Pentium III processor with 256 MB RAM) or the image may be sub-sampled by the 'Resize' function in ENVI™. The resizing process involves the application of a factor to both the number of lines and columns to reduce the number of samples. In the above example, each image dimension needs to be multiplied by a factor of 0.6 to achieve 60,000 samples. As mentioned in section 5.6, resizing may be applied when the H-resolution case applies for most image classes, that is, no significant

information loss occurs during the resizing process. The latter takes about 15 seconds for an output image with 60,000 samples (2 GHz processor).

Figures 5.24 and 5.25 display the amount of image samples of a 288-band data set against the execution time needed for the Datastats and the Resizing ENVI™ routines, respectively.

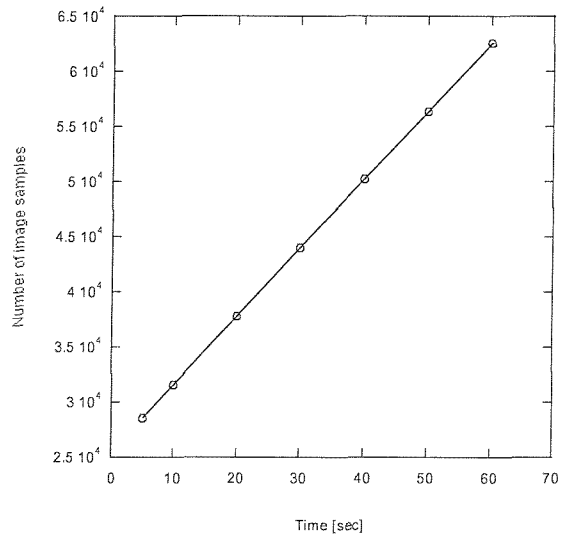


Figure 5.24: Estimate of amount of image samples against the execution time of the Datastats routine (see section G.4) for a 288-band data set and a 1 GHz processor.

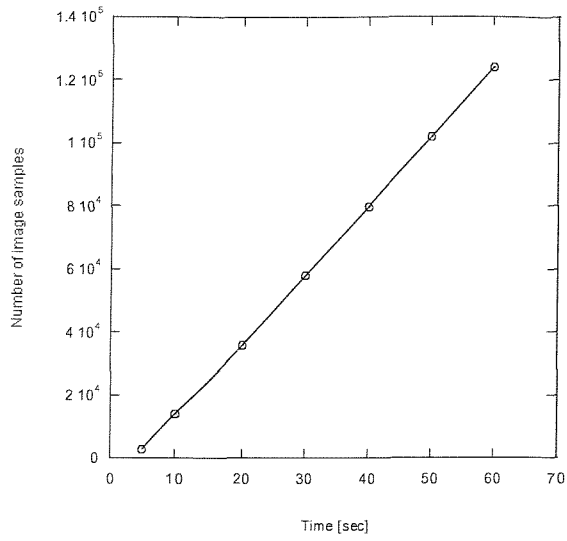


Figure 5.25: Estimate of amount of output image samples against the execution time of the Resizing ENVI™ routine for a 288-band data set and a 1 GHz processor.

The execution time of the UBS program only depends on the number of image bands and the parameter settings. For an output set of 15 bands, the maximum allowable bandwidth should be chosen relatively low (e.g. 4 to 5 rows), in order to avoid an iteration of the routine (no more bands are available to fill the remaining places as first bands are too broad) and therefore an increase in the overall running time. The execution time was extrapolated to 7 seconds for a 2 GHz processor using a 288-band image and the following UBS parameter settings: unequal bandwidth, least correlated first, output 10 bands, maximum bandwidth 4 rows, 30% minimum signal level, no pre-specified bands or output image.

In summary, for the UBS method, the net execution time amounts to 8 (radiometric correction) plus 5 (dark data) plus 15 (resizing) plus 30 (data statistics) plus 7 (UBS) equals to 65 seconds using a 2 GHz processor and a 288-band image with 1500 lines and 101 columns. This leaves the instrument operator 2 to 4 minutes to switch between routines and program the CASI-2 with respect to the selected band set.

In contrast to the UBS method, the CSBS method requires the clustering and the calculation of cluster statistics on top of the possible sub-sampling, data statistics computation, and SBS band

selection routine. CSBS would take about two minutes longer than the UBS method for an image of 60,000 samples and 288 bands using 5 clusters and 1 iteration as parameters for the clustering procedure (2 GHz processor). CSBS is therefore less applicable for in-flight use with the given time constraints and computer processing speeds.

5.8 Summary

The purpose of this chapter was to develop an unsupervised band selection routine that optimises band configuration parameters with respect to the accuracy of the classification task and that could be used in-flight for the acquisition of CASI-2 multispectral data. For the in-flight procedure, it was assumed that a reduced-swath hyperspectral CASI-2 image would be acquired over a representative part of the scene as algorithm input. Then, the CASI-2 would be programmed with the selected band set, and a full-swath multispectral image would be collected. The hyperspectral data to be acquired should include all target areas of interest to a similar proportion as they appear in the final multispectral imagery.

An unsupervised band selection (UBS) algorithm was developed, which is based on the assumption that the most class-informative band set consists of bands that are least redundant, that is, least correlated. It incorporates the increase of bands in width as long as the SNR as measure of image quality improves, the relative band signal levels are above a user-specified threshold, and the bandwidth is below a user-defined upper limit. The algorithm was implemented in an IDL^{TM} (version 5.5 Win 32 x86) and ENVI^{TM} (version 3.5) program that can be run in ENVI^{TM}. Computational efficiency of the routine was especially considered to allow the routine to be used in-flight. The program options include either to force the bands to be of equal width or to let the bands expand to achieve maximum possible SNR (optimal band set). An algorithm option was included that allowed to start the band set with the band of highest SNR (UBS LN), instead of the least correlated band pair (UBS LC). UBS also allows specific bands to be included in the selection process or excluded from it.

Similar to PCA, UBS aims to decorrelate the band set. But instead of transforming the data into new orthogonal features, UBS picks the bands that are most different from each other. That is, UBS aims to explain the overall variance with as few original bands as possible, while PCA aims to represent data variance with as few orthogonal features as possible. UBS has the disadvantage of being based on a sub-optimal sequential search procedure, while PCA is not. However, for band selection, UBS has four major advantages over PCA. First, UBS requires less computational effort than PCA. Second, it preserves data integrity, as its selected bands are directly physically interpretable. Third, it does not try and fit an orthogonal basis to the overall data variance, which may conceal unique band information that is small in variance. And fourth, it allows for a band expansion routine similar

to the one in the SBS method, but purely based on the band SNR. UBS provides an upper bound to the intrinsic dimensionality of the data set.

The use of correlation measures in the UBS routine requires each band to be normally distributed. However, the latter assumption was shown to be inappropriate for most of the New Forest and River Severn bands. CSBS, an unsupervised version of the application-specific SBS algorithm discussed in chapter 4, was introduced that uses clustering to define the classes within the scene, circumventing the normality assumption of band variables. A computer program was written in IDL™ (version 5.5 Win 32 x86) and ENVI™ (version 3.5) that creates clusters using the K-Means clustering routine in ENVI™ and calculates their statistics. Regarding the parameter settings of the K-Means clustering routine, changes in the parameters had only little effect on the discrimination ability of the resulting band set for the given data and class sets.

The UBS and CSBS algorithms were evaluated by applying them to the hyperspectral data introduced in chapter 3. The sub-optimality of the algorithms was quantified by comparing the MLC accuracy of the derived band sets with that of a band set obtained from an exhaustive search using the MLC accuracy measure as criterion. For all three set dimensions, the UBS LC band set performed consistently well with a maximum accuracy loss of 7% (with one exception). UBS LC is more applicable to band sets of at least two dimensions, as the least correlated band pair is chosen at the start of the algorithm. The sub-optimality of the UBS LN and CSBS band sets was generally greater than that of the UBS LC band set for the two data sets.

In comparison to 'established' vegetation and coastal band sets, the UBS and CSBS band sets performed superior for the first six dimensions for the given data sets. For higher dimensions (11 to 14), the 'established' band sets gave a similar MLC accuracy than the selected band sets. That is, the unsupervised band selection methods were very effective for low-dimensional output band sets. The resulting optimal band configurations were also found to be physically meaningful with respect to the classes under investigation.

The band number criterion used in the UBS algorithms was shown to be an effective upper bound to intrinsic dimensionality (ID) of the given data sets. It equals to the number of bands already in the set which achieve a coefficient of determination of higher than 95% with the least correlated of the remaining bands. The CSBS IDD estimate is based on the same principles as the SBS IDD estimate and was not evaluated here.

The UBS and CSBS algorithm may both be applied operationally and in-flight to select a band set for multispectral image acquisition, although the UBS was shown to be computationally more efficient. To decide which of the band selection methods to choose, a comparison between all band selection methods introduced so far is presented in the following chapter.

6 Discussion

This discussion

- compares the band selection algorithms described in chapter 4 and 5 with respect to their band set performance, underlying assumptions, consistency, effectiveness and computational efficiency (section 6.1),
- investigates the effectiveness of band selection (section 6.2),
- tests the benefit of narrow band data for the classification tasks introduced in chapter 3 (section 6.3),
- introduces a new data set to test the SBS algorithm and the benefit of hyperspectral data for the given classification task (section 6.4),
- examines whether the band selection methods developed in this thesis could be applied to other hyperspectral remote sensing applications (section 6.5),
- discusses data simulation as an option for cases where the hyperspectral input data to the algorithm could not be acquired with the target sensor, such as for sensor design studies (section 6.6), and
- investigates the need for calibration of data to radiance or apparent reflectance in the context of band selection (section 6.7).

A concluding summary is presented in section 6.8.

6.1 Comparison of the Band Selection Algorithms Developed for this Thesis

The MLC accuracies of the band sets resulting from the SBS, UBS LC (least correlated band pair first), UBS LN (least noisy band first) and the CSBS algorithms were compared with each other for the New Forest and River Severn data set. All algorithms employed the 'unequal bandwidth' option to potentially achieve an optimal band set. The UBS algorithm had the maximum bandwidth value set to 4 rows to be merged. CSBS clustering was performed with 5 and 2 clusters for the New Forest and River Severn data set, and one iteration cycle.

Regarding the New Forest data (see figure 6.1), all four algorithms performed equally well for at least two bands with insignificant differences of below 0.4%. Significant differences occurred only for the single best bands, where the SBS band achieved 78.5% MLC accuracy, the UBS LC band 76.3%, the UBS LN band 61.4% and the CSBS band 61.3%. Similar results were obtained for the River Severn data set, as shown in figure 6.2. Differences between the SBS and UBS algorithms were insignificant (below 1.2%) for band sets with at least two bands. The CSBS band set gave a significant accuracy difference (maximum 3.6%) with respect to the band sets of other algorithms for

the first three dimensions. It achieved less accuracy than other band sets for set dimensions one and three. The order of the first bands in terms of accuracy is equivalent to that of the New Forest data set: First the SBS band (75.6%), followed by the UBS LC band (71.8%), the UBS LN band (57.2%) and the CSBS band (47.6%). That is, the SBS, UBS and CSBS band sets achieved comparable MLC accuracy (at least for two bands in the set), indicating that all four algorithms may be equally applied for the given data and class set for band selection.

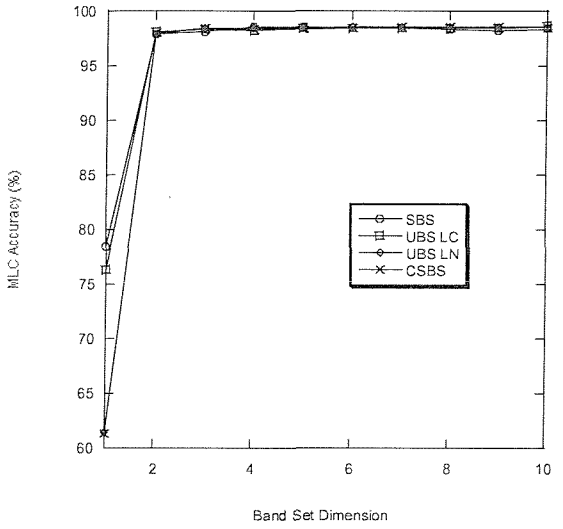


Figure 6.1: MLC accuracy of band sets output by the SBS, UBS LC, UBS LN and CSBS algorithms for the New Forest data set.

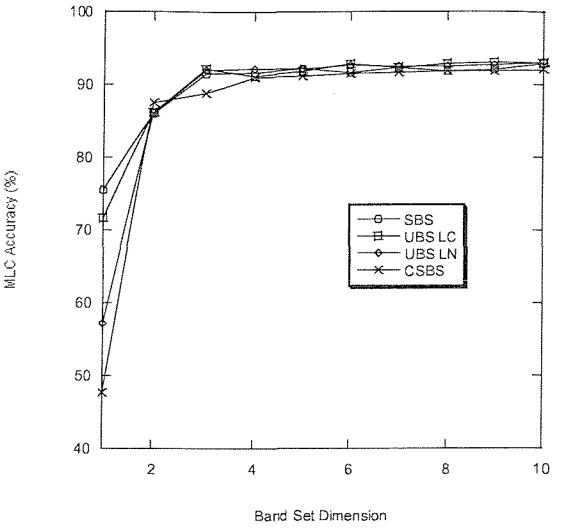


Figure 6.2: MLC accuracy of band sets output by the SBS, UBS LC, UBS LN and CSBS algorithms for the River Severn data set.

Apart from their band set performance, the algorithms may be compared with respect to their underlying assumptions, consistency, effectiveness and computational efficiency.

Both the SBS and CSBS assume the class training samples to be normally distributed. For the SBS, the accurateness of this assumption lies in the hand of the analyst, and it depends on the clustering parameters for the CSBS algorithm. In contrast, the justification of the assumption of normally distributed bands in the UBS algorithm relies on the data itself, which may only be changed by elaborate data masking. Although some of these normality assumptions were not completely met for the test data and class sets used in this thesis, all methods were shown to be robust to deviations from normality by giving a band set that achieves high accuracy for the given classification task.

The assumption of the SBS and CSBS, that the Transformed Divergence (TD) measure and the MLC accuracy are linearly dependent for the given data and class set, may be generally met, because the TD was derived as an upper bound to the MLC error probability. The UBS assumption that bands depend only in a linear fashion onto each other generally holds for remotely sensed data of Earth surface targets.

The sub-optimality of the different algorithms was quantified in terms of accuracy loss of their band sets with respect to the optimal band set for up to three set dimensions (see sections 4.4.4 and 5.7.3). For the SBS band set, this accuracy loss amounted to a maximum of 6% for the given data sets. The UBS and CSBS band sets resulted on average in the same amount of accuracy loss (6%) for two and three bands in the set, but with deviations of up to 15%. For the single best band, however, UBS and CSBS bands were up to 50% less accurate. That is, for the given data sets, SBS resulted in more optimal bands, especially in the search for the single best bands.

Regarding the consistency of the algorithms, both the SBS and UBS versions are repeatable, while the CSBS may result in different band sets with changing initialisation and clustering parameter settings. This is a major drawback for the CSBS algorithm, especially as it is unclear what settings are to be chosen in each case. This makes the algorithm less suitable for band selection than the SBS algorithm.

A major disadvantage of the UBS algorithm is that it cannot be used to find the optimal maximum bandwidth for MLC. This is due to the loose relationship between its bandwidth criterion, the SNR, and the MLC accuracy. The algorithm typically expands some of the output bands to the maximum allowable bandwidth defined by the user. In contrast, the latter limit has little effect on the choice of optimal maximum bandwidth within the SBS or CSBS algorithms. That is, UBS can be used for bandwidth selection for MLC, if a reasonable maximum bandwidth limit is specified.

Another shortcoming of the UBS algorithm is that it is dependent on sensor noise to be estimated in order to calculate the SNR for band expansion. Although dark data are routinely collected with most image sensors at the present time, they are not normally delivered to end users (A. Wilson, 2003, personal communication).

With regard to the band number determination, all algorithms delivered upper bound estimates to the ID or IDD for the given data sets that were consistent with those found using traditional estimation methods. However, for all methods, the users have to decide whether they employ a 95% or 99% threshold to derive the band number estimate.

The UBS and CSBS algorithms were shown to be computationally more efficient than the SBS algorithm, and therefore more applicable to real-time in-flight execution between the acquisitions of hyperspectral and multispectral image data using the current computer processing speeds.

Furthermore, the UBS, SBS and CSBS can all be potentially used for applications other than MLC (see section 6.4). However, the UBS has the advantage over the SBS and CSBS that it can be applied without modifying the algorithm.

The choice of band selection algorithm depends very much on the requests of the data user. The SBS, UBS and CSBS algorithms presented in this chapter all incorporate benefits and costs, and the user has to decide if the benefits of a method outweigh its costs for a given application. But if a band selection method was to be sought for a supervised MLC, the SBS method would be the preferred one. Table 6.1 presents a summary of the above comparison between the methods.

Table 6.1: Comparison of the SBS, UBS and CSBS algorithms.

	SBS	UBS	CSBS
Assumptions	- Normal class distributions - Linear relationship between TD and MLC accuracy	- Normal band distributions - Only linear dependencies between bands	- Normal cluster distributions - Linear relationship between TD and MLC accuracy
Sub-optimality	Low	High for single bands, low for at least two bands in the set	
Consistency	High	High	Low
Optimal maximum bandwidth estimation	Yes	No	Yes
Sensor noise estimation necessary?	No	Yes, for bandwidth increase	No
Optimal band number estimation	IDD estimate	Upper bound estimate to ID	IDD estimate
In-flight use	No	Yes	Yes
Generalisation to other hyperspectral applications	Yes, but with modification	Yes	Yes, but with modification

6.2 Effectiveness of Band Selection

From figures 6.1 and 6.2 it may be seen that the band sets chosen by the SBS, UBS and CSBS algorithms achieved similar accuracy after a certain number of set dimensions, although the chosen bands differed in placement and width between methods. At this point, one may ask at which dimension it becomes irrelevant which band is added to the set, as its addition will not significantly change the classification accuracy of the set. That is, when does band selection become ineffective?

To find the limit of the effectiveness of band selection for the given data and class sets, the SBS band set was compared with sets of randomly and equally spaced bands with respect to the

achieved classification accuracy. The uniformly spaced band sets were created separately for each dimension to have maximum equal spacing between bands and minimum spacing towards the edges of the original band set. Table C.4 lists randomly and uniformly spaced band sets used in this evaluation and their performance is displayed in figures 6.3 and 6.4 for the New Forest and River Severn, respectively. For the uniformly-spaced band sets, only the accuracy of the entire set was calculated, and sets of all possible dimensions within the optimal set were generated.

The SBS band set achieves significantly higher accuracy than the random or equally spaced band sets for the first three dimensions for the New Forest data set. Above four dimensions, the band sets reach similar accuracy levels with mostly insignificant differences of less than 1% from the SBS set. The River Severn data set gives a comparable result, with discrepancies between SBS and random or equally spaced band set performances of less than 3.2% above the fourth dimension. However, in the latter case, differences in accuracy are significant between the SBS and equally and randomly spaced band sets for most dimensions.

These results imply that, for the given class and data sets, band selection is principally effective only for the first few bands of the selected set, corresponding to the intrinsic discriminant dimensions of the data sets. The latter dimensions equate to 3 and 4 for the New Forest and River Severn data sets, respectively. However, as for the River Severn band set, the SBS selected bands also perform significantly better for higher dimensions, suggesting that they should be preferred to any equally or randomly spaced band set.

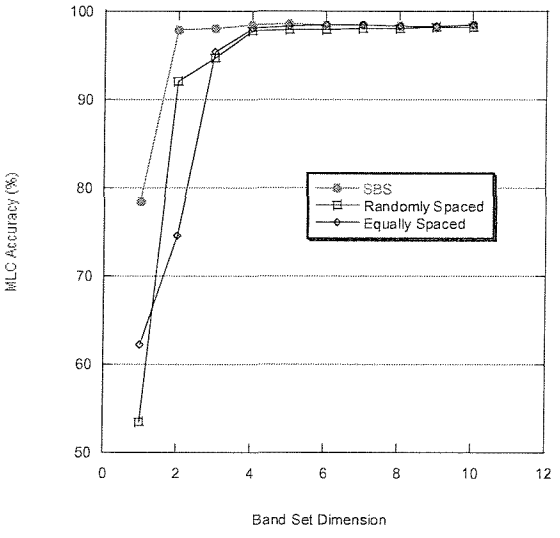


Figure 6.3: MLC accuracy of the SBS band set compared with the one of randomly and uniformly spaced band sets for the New Forest data set.

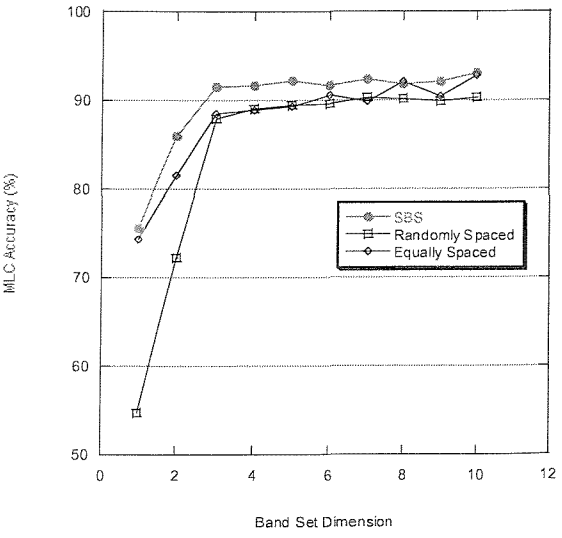


Figure 6.4: MLC accuracy of the SBS band set compared with the one of randomly and uniformly spaced band sets for the River Severn data set.

6.3 Testing the Benefit of Narrow Band Data

In sections 4.4.4 and 5.7.3, broad band sensors such as the ETM+ (bandwidths from 60 to 260 nm) were shown to result in an equally high MLC accuracy than the narrow band sets selected with the corresponding band selection algorithms. The question may then be asked whether or not narrow bands actually achieve a significantly higher MLC accuracy than broad bands for a given classification task.

To answer this question, the 'equal and fixed width' SBS algorithm option was chosen to produce sets of bands of increasing width, with a (minimum) upper bandwidth of 370 and 220 nm for the New Forest and River Severn data set, respectively. As discussed in section 6.1, the UBS method cannot be employed for bandwidth determination, while the CSBS algorithm is less suitable than the SBS method for this task due to its dependency on clustering parameters.

Figure 6.5 and 6.6 display the accuracy of the sets of expanded bands for each set dimension for the New Forest and River Severn data set, respectively. For the New Forest data set, the SBS algorithm produces consistently optimal band sets that achieve an MLC accuracy that lies, for at least two dimensions, within insignificant 0.5% of the accuracy of the narrow-bandwidth set. In contrast, the expanded band sets for the River Severn data set are within 5.1% significantly different in accuracy from the narrow-band set for at least three dimensions.

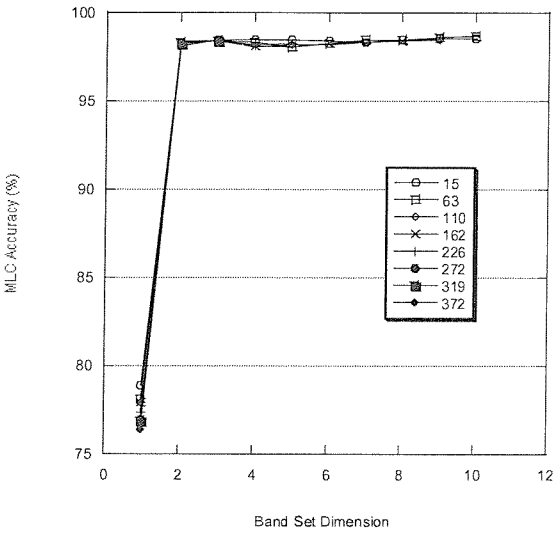


Figure 6.5: Accuracy of SBS band sets for increasing bandwidth for the New Forest data set. The width was fixed to 1, 5, 9, 13, 17, 21, 25 and 29 rows to be merged. The number in the legend is the corresponding minimum width (in nm) of the bands in each set.

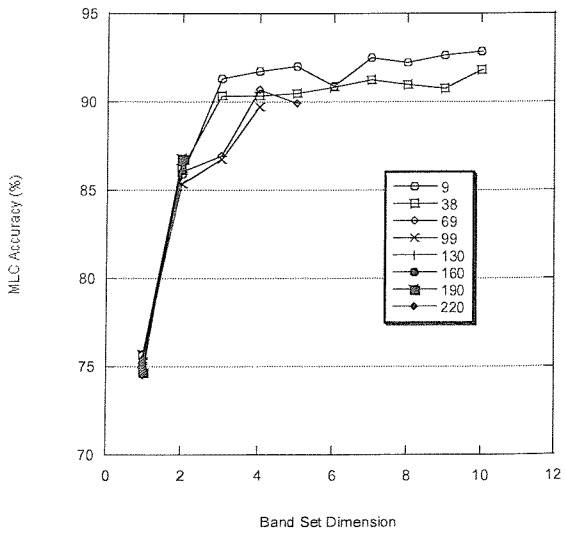


Figure 6.6: Accuracy of SBS band sets for increasing bandwidth for the River Severn data set. The width was fixed to 1, 5, 9, 13, 17, 21, 25 and 29 rows to be merged. The number in the legend is the corresponding minimum width (in nm) of the bands in each set.

With increasing bandwidths, fewer bands are available for the final set. This may result in sub-optimal set accuracy if the number of bands falls below the IDD estimate of the data set. The IDD for the New Forest and River Severn data set was estimated to 2 and 3, respectively (see section 3.5). As, for the New Forest data set, the accuracy difference to the narrow band set was insignificant for all expanded band sets at the IDD dimension, two bands of 370 nm in width may be employed to achieve a similar classification accuracy as two bands of 15 nm in width. That is, narrow band data are not of significance for the New Forest classification task. However, they are of advantage for the River Severn classification task, as they give a significantly higher accuracy than expanded band sets do at the IDD dimension of 3.

The classes for the New Forest data set reflect partly vegetated areas of different hydrological conditions (dry, humid and wet heath), and multispectral data were shown to be sufficient to distinguish between them. However, the author's interest was also to examine whether hyperspectral data was of advantage for the spectral differentiation between vegetation classes from areas of mostly the same hydrological regime, e.g. a bog surface. In the next section, a new data set of a bog surface is introduced for this purpose. The new data set was also used to further test the SBS algorithm developed in chapter 4.

6.4 Tregaron Bog Case Study

This case study aims to see what band set is the most optimal in separating bog condition classes, and whether hyperspectral resolution data have a significant advantage over multispectral data for this classification task. Section 6.4.1 introduces the study area, the hyperspectral data and the classes of interest. The method is described in section 6.4.2, while the results and conclusions of the experiment are presented in section 6.4.3.

6.4.1 Study Area, Data Set and Class Definition

Study area

Cors Caron, also known as Tregaron Bog, is located north of Tregaron village in the Teifi valley in Ceredigion, west Wales (see figure 6.7). It is an extensive lowland raised bog complex of 816 ha, consisting of three hydrologically independent peat domes. All three domes suffer from marginal peat cutting at different scales, but their central part is still structurally intact. The least disturbed bog is the largest one on the western side of the Teifi river. Cors Caron became a National Nature Reserve in 1955 and a Ramsar site in 1993. It supports a range of rare plant species (e.g. the bog moss *Sphagnum pulchrum*) and rare animals (e.g. the rosy marsh moth; Milton *et al.*, 2003).

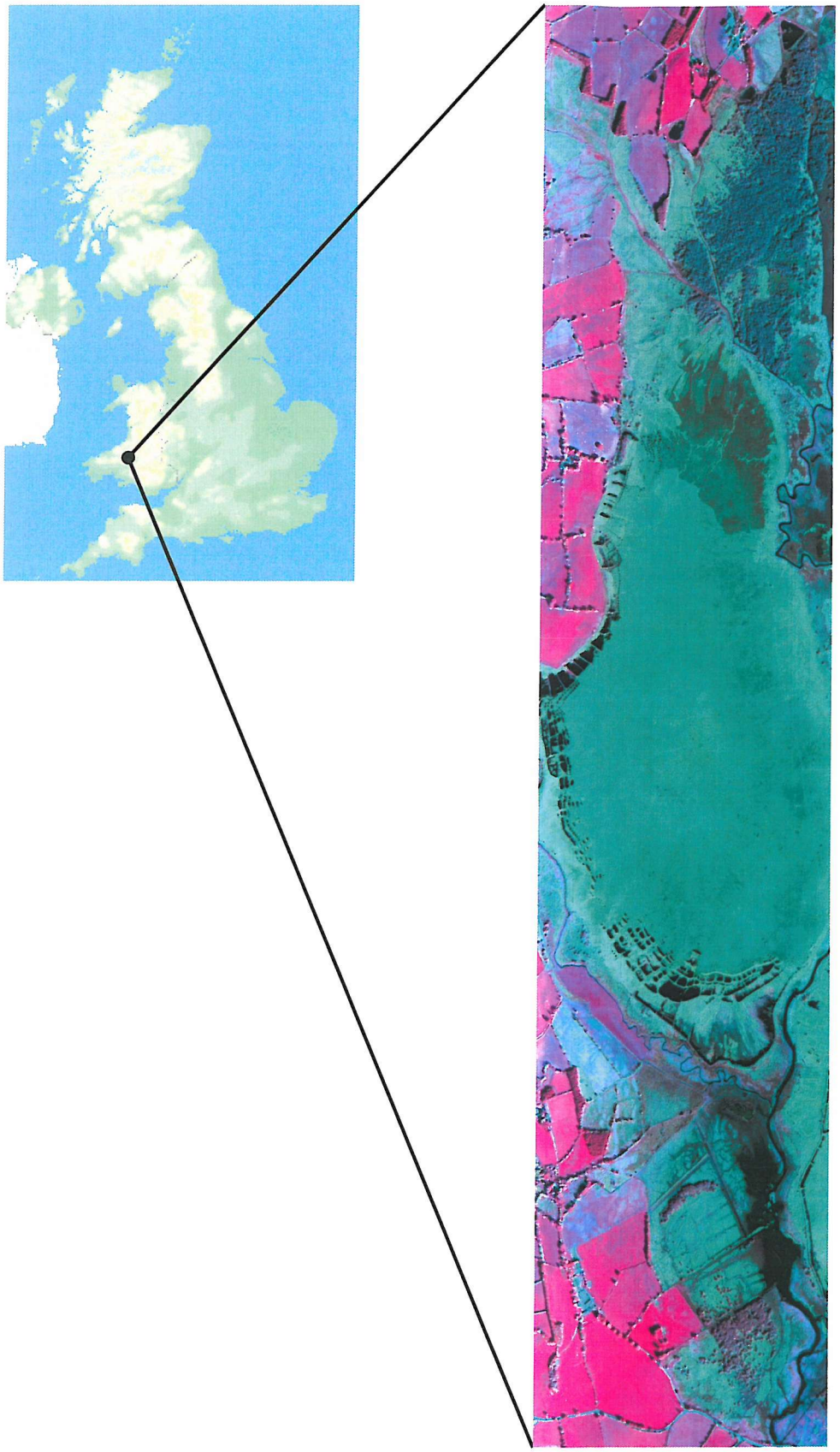


Figure 6.7: False colour CASI-2 image of the Tregaron bog (R = band 40, G = band 22, B = band 14). © UK Natural Environment Research Council, 2001.

Data set

Airborne hyperspectral CASI-2 image data were collected in October 2001 by NERC ARSF over the western dome of the Tregaron bog. Details of the data acquisition are presented in table 6.2. The data were radiometrically corrected to spectral radiance. The 48 bands with their associated centre wavelengths and bandwidths are listed in table A.5. Some bands in the blue (bands 1 to 9) and NIR (bands 47 to 48) suffer from severe striping and were excluded from further processing. The data were not geometrically corrected as this was not necessary for the actual application task. A false colour image is displayed in figure 6.7.

Table 6.2: Characteristics of the CASI-2 image acquisition over the Tregaron bog.

Acquisition parameter	CASI-2 imagery
Date of acquisition	12 October 2001
Time of acquisition (hrs GMT)	10:49 - 10:53 a.m.
Type of aircraft	Dornier 228
Altitude (km) (above ground)	1.47
Ground speed (knots)	125
Number of scan lines	6731
Sensor mode	Enhanced Spectral Mode
Spatial resolution (m)	2.9
Number of spectral bands	48
Spectral resolution (nm)	11.4 - 11.8
Spectral range (nm)	409 – 945
Data format (bit)	16
View angle (°)	Nadir
Field of view (°)	53.2
Swath width (km)	1.5 (511 pixels)
Status of the atmosphere	Hazy with clear sky above

Class definition

Between August and September 2002 a field survey was carried out on behalf of English Nature by J. Schulz (Milton *et al.*, 2003). An *in situ* map of surface condition classes was created (see figure 6.8) according to the categories defined in Milton *et al.* (2003). The airborne imagery available only depicts the western dome of the Tregaron bog, and the corresponding surface condition classes are described in table 6.3.

As the defined bog condition classes appeared very homogeneous in the CASI-2 imagery, enough samples were available to randomly select training and testing pixels for each class. Regarding the number of training pixels, a conservative value of 30 was used for the ratio of the number of training pixels to the number of spectral bands (i.e. 1110 training pixels for each class).

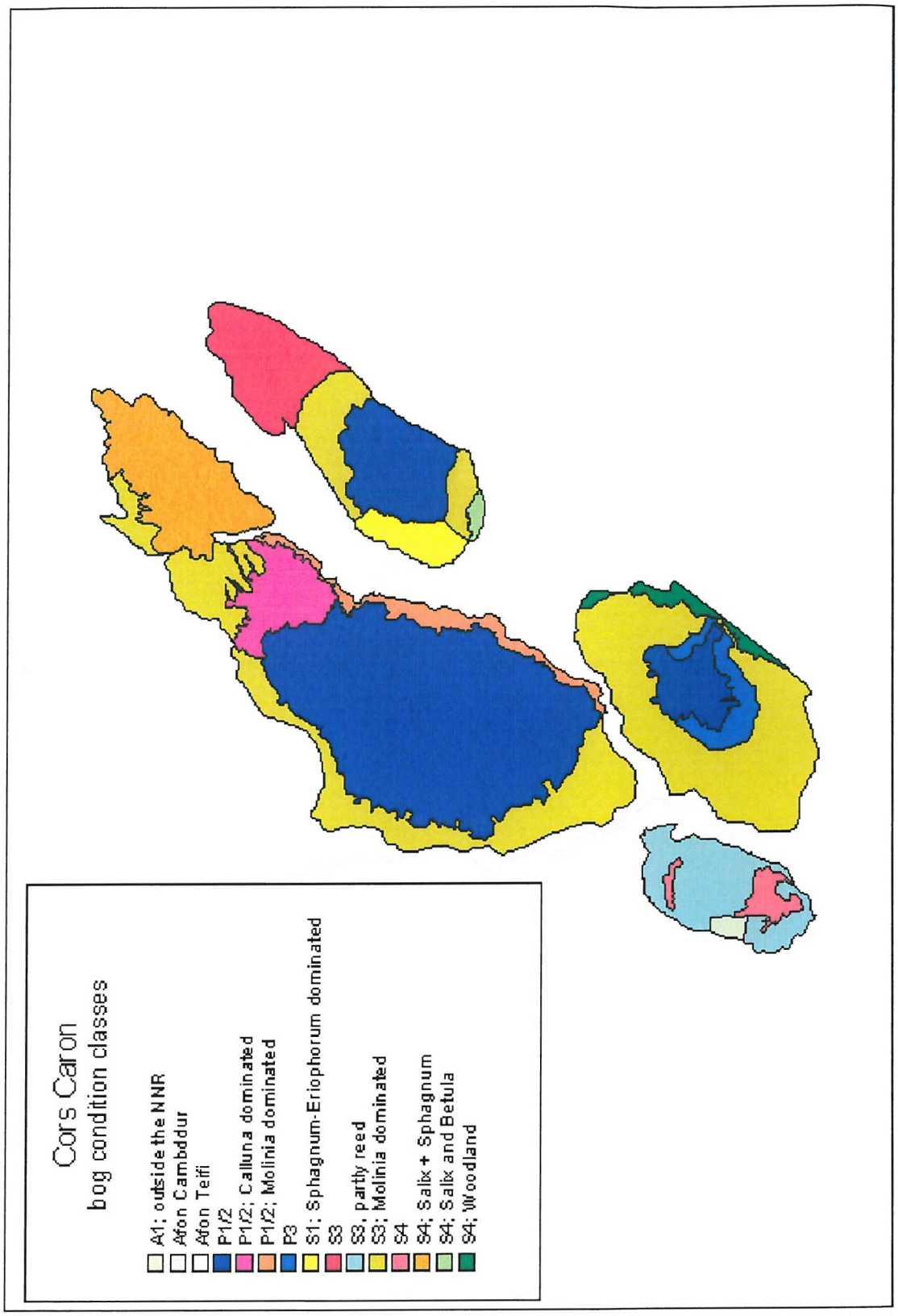


Figure 6.8: J. Schulz's map of surface condition classes for the Tregaron bog (Milton *et al.*, 2003). The classes used are explained in table 6.3.

Table 6.3: Surface condition classes used for the classification of the Tregaron bog (after Milton *et al.*, 2003).

Class	Description
P1/2	Primary (near-) natural or degraded raised bog with a substantial cover of colourful *Sphagna* and the ability to accumulate peat. If degraded, then by factors other than drainage (e.g. burning or grazing).
P1/2 *Molinia* dominated	As P1/2, but chiefly covered by *Molinia caerulea.*
P1/2 *Calluna* dominated	As P1/2, but dominated by *Calluna.*
S3 *Molinia* dominated	Secondary re-vegetated degraded bog. Dry peat cuttings (non-peat forming). Dominated by *Molinia caerulea.*
S3 partly reed	As S3 above, but partly covered with reed.
Carr	Wooded areas on cut-over peat (willow carr).
Standing Water	Bodies of standing water formed behind dams around the crowns of the bogs (dams were built to prevent surface runoff).

6.4.2 Method

The supervised SBS algorithm was chosen for the band selection task at hand, as specific bog condition classes were given. The assumptions of the SBS method, that is, the normal distribution of the defined classes and the linear relationship between the Transformed Divergence and the MLC accuracy, needed to be tested. In addition, the sub-optimality of the method was quantified in terms of MLC accuracy by comparing the SBS band sets (up to three dimensions, no bandwidth increase) with the sets of corresponding dimension derived using an exhaustive search with the MLC accuracy as criterion. As sufficient training and testing pixels were available, the holdout method was used to estimate the overall MLC accuracy (see section 2.2.2).

The SBS band set was derived using the following algorithm options: unequal bandwidth, maximum bandwidth of 20 spectral rows and minimum band mean of at least 30% of the maximum band mean in the set. The SBS band number criterion was tested by comparing it to ID and IDD estimates of the data set. The latter were both based on the PCA, where the total PC variance was used to determine the ID, and the scree plot method (see section 2.4.1) was applied to the MLC accuracy of sets of consecutive PCA features to give an estimate for the IDD. The effectiveness of the SBS band selection result was quantified by the difference in accuracy between the SBS set and equally and randomly spaced band sets.

Finally, it was tested whether hyperspectral data have an extra benefit over multispectral data in discriminating between the bog condition classes for the given data set. Sets of bands of increasing width were generated with the SBS 'equal and fixed width' algorithm option, which were compared in terms of MLC accuracy with the SBS set having no expanded bands.

6.4.3 Results and Conclusion

Testing the assumptions and sub-optimality of the SBS algorithm

The main assumptions of the SBS method are:

- normally distributed class training data, and
- a linear relationship between the TD measure and the MLC accuracy.

The histograms of the 7 classes were calculated for evenly spaced bands (13, 23, 33 and 43), and it could be shown that for most bands and classes, the normal curve fitted the histogram relatively well. Regarding the second assumption, a significant correlation coefficient of 0.99 was estimated between the TD and MLC accuracy for the given sample points (see figure 6.9). These results indicate that the SBS method is applicable to given class and data set, as both assumptions are met.

The comparison of the MLC accuracy of the SBS band sets with the set derived from the exhaustive search method revealed an (insignificant) accuracy loss of less than 2.8% for the first three set dimensions (see figure 6.10). That is, the sub-optimality of the SBS algorithm is relatively small and consistent for the first three band set dimensions.

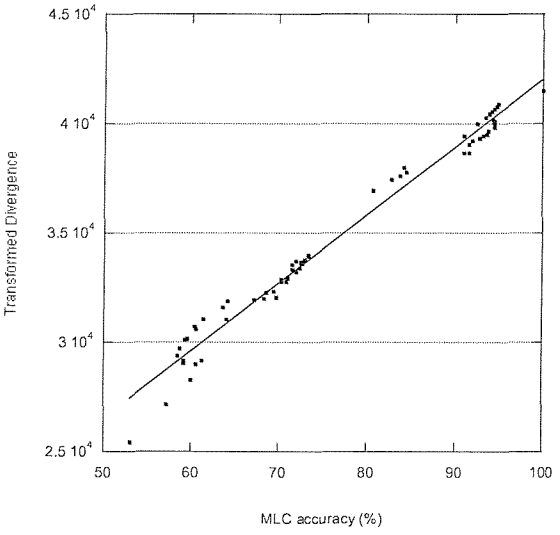


Figure 6.9: Scatter plots of the Transformed Divergence measure against MLC overall accuracy estimated with the holdout method for the Tregaron data set. The regression line is displayed (correlation coefficient r = 0.99).

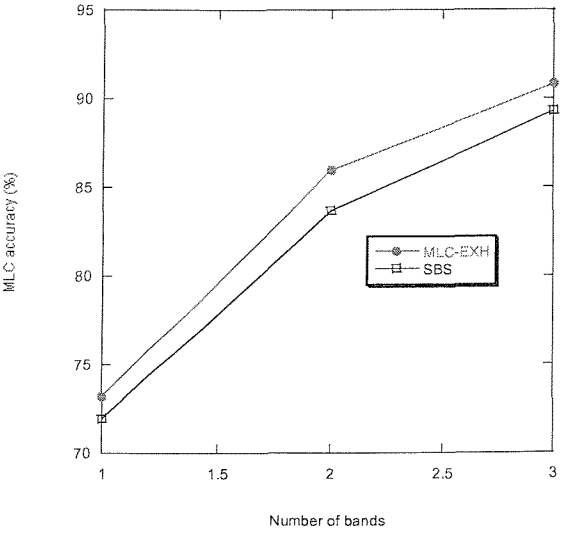


Figure 6.10: MLC accuracy of optimal band sets derived using an exhaustive search with MLC accuracy (MLC-EXH) and SBS algorithm for the Tregaron data set.

The SBS band set

Table 6.4 displays the first six bands selected by the SBS 'unequal bandwidth' option, which are overlaid onto a leaf spectrum in figure 6.11. The first band is centred on the NIR plateau, which is sensitive to canopy structure and which may be used to delineate carr and the standing water surfaces from the bog surface. Bands 2 and 4 in the VIS are relatively broad bands located over the red absorption feature and the green reflectance peak, respectively. Both bands are related to the broad spectral absorption by chlorophyll and respond to chlorophyll amount in the canopy. Therefore, they may help to differentiate between different plant species, such as *Sphagnum*, *Molinia*, *Calluna* and reed. The same is true for bands 3, 5, and 6, which sample the red edge feature.

Table 6.4: First 6 optimal bands selected by the SBS algorithm (unequal bandwidth, maximum bandwidth of 20 spectral rows, minimum band mean of at least 30% of the maximum band mean) for the Tregaron data set.

Band number	Tregaron			
	Band centre [nm]	Band width [nm]	Band start [nm]	Band end [nm]
1	881.718	23.213	870.1115	893.3245
2	675.031	46.129	651.9665	698.0955
3	715.173	11.768	709.289	721.057
4	566.729	102.451	515.5035	617.9545
5	755.38	23.269	743.7455	767.0145
6	732.4	23.26	720.77	744.03

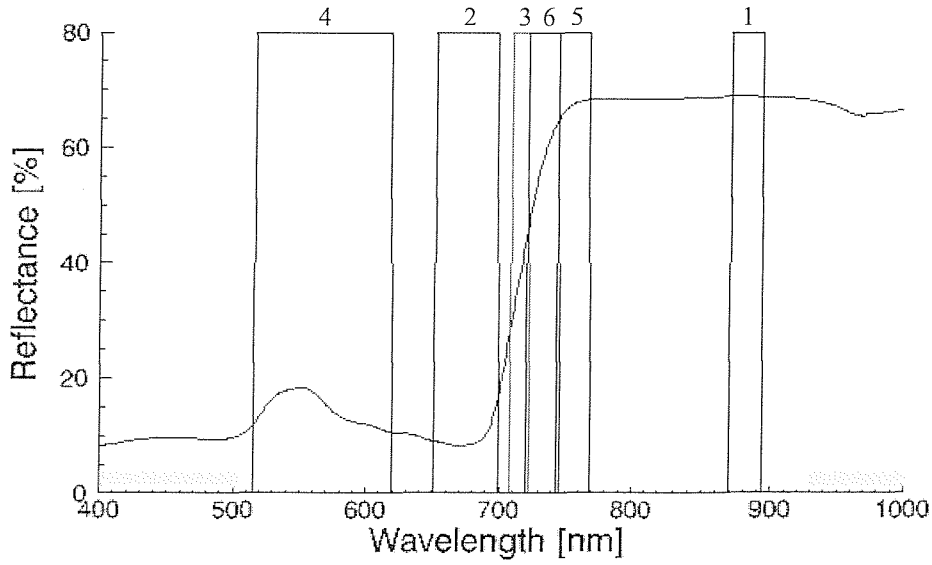


Figure 6.11: First six bands selected by SBS for the Tregaron classification task (see table 6.3 and 6.4). Wavelengths not available for band selection are indicated by grey bars.

Estimated optimal band number

The ID of the data set was estimated to 3 as nearly 99% of the variance is explained by the first three PCs. This can be seen from the cumulative PC eigenvalues listed together with the MLC accuracy of the first ten PCs in table 6.5. The scree plot method was then applied to the MLC accuracy of successive PCs to give an IDD estimate of 5. The same result is achieved when using a 99% threshold value on the PMATD criterion (see table 6.5).

Table 6.5: PMATD band number criterion for the first 10 SBS bands, and cumulative eigenvalue and MLC accuracy for the first 10 PCs of the Tregaron data set.

Band / PC	PMATD	Cumulative PC eigenvalue (%)	MLC accuracy of PC (%)
1	81.45	93.42	72.03
2	89.79	97.91	78.33
3	97.19	98.97	85.80
4	98.79	99.37	91.97
5	99.59	99.50	93.94
6	99.85	99.57	94.38
7	99.996	99.62	94.21
8	99.30	99.66	94.48
9	99.81	99.70	94.62
10	100	99.72	94.83

The optimal band number estimate of 5 is also reflected in figure 6.12, where the SBS band set is compared in terms of MLC accuracy with equally and randomly spaced band sets. SBS band selection seems to be especially effective for the first five dimensions, where differences to the randomly and equally spaced band sets are as high as 12%. From 6 to 10 dimensions, the differences get reduced to below 3%, but remain statistically significant in most cases. Although the PC features give a higher MLC accuracy than the SBS band set from 5 dimension onwards (see figure 6.12), PCs do not have a physical meaning attached to them and do not consider a possible expansion of bands.

In summary, the optimal band number criterion was confirmed using a threshold value of 99% rather than 95%. The users are advised to judge in each case which threshold they apply. Using the 99% threshold will result in a more accurate end result but at the cost of a potentially much greater number of features.

Benefit of hyperspectral data for the classification task

The widths of the best six SBS bands vary from 11 to 103 nm (see table 6.4). The fact that the broad bands were picked out around the red and green wavelengths may suggest that an aerial photograph may probably be sufficient for differentiating between the given bog condition classes. An experiment was conducted to see whether hyperspectral data have an additional advantage over

multispectral data for the given classification task. Sets of bands of equal width were created for increasing bandwidths (12 to 194 nm) using the SBS 'equal and fixed width' algorithm option. Figure 6.13 shows the MLC accuracy estimated for the given sets. Only sets up to 57 nm in bandwidth could be considered as they provide enough bands to reach the optimal band number of 5 dimensions. As the bands of 57 nm gave an MLC accuracy which was insignificantly smaller (below 1%) than that of the 12 nm bands, the author concluded that five carefully placed multispectral bands of about 57 nm could be used to achieve the same accuracy for this classification task than five bands of around 12 nm.

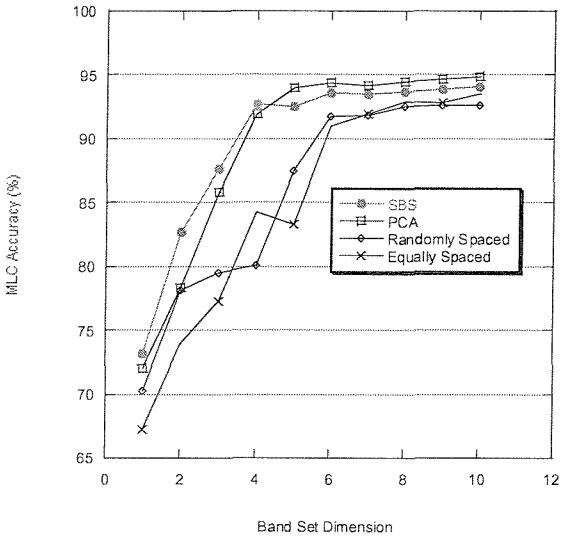


Figure 6.12: MLC accuracy of the SBS band set compared with the one of randomly and uniformly spaced band sets for the Tregaron data set.

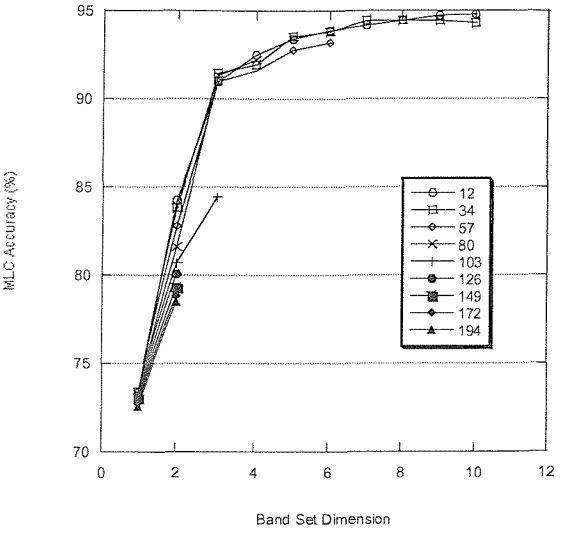


Figure 6.13: Accuracy of SBS band sets for increasing bandwidth for the Tregaron data set. The width was fixed to 1, 3, 5, 7, 9, 11, 13, 15 and 17 rows to be merged. The number in the legend corresponds to the minimum width (in nm) of the bands in each set.

6.5 Generalisation of the Use of Band Selection Methods

In this section, the band selection methods SBS and UBS presented in chapters 4 and 5, respectively, are scrutinised whether they can be used with other hyperspectral data applications, such as regression, linear spectral unmixing and spectral angle mapper. Likely changes of the algorithms to adapt to the different techniques are investigated. Furthermore, the employment of the two methods for other scene, atmospheric and illumination conditions, and for other imaging spectrometers is discussed.

6.5.1 Use with other Hyperspectral Applications

Other Hard Classification Methods

Classification methods such as minimum distance, k-nearest neighbour or the parallelepiped classifier would also benefit from a reduced band set that discriminates most between the given classes. However, in contrast to the probabilistic Maximum Likelihood classification, the methods mentioned above usually rely only on first-order class statistics and typically measure class distance in the Euclidean space. If SBS was to be employed for supervised band selection for these methods, the criterion function used in the SBS algorithm, the Transformed Divergence, would need to be replaced by the Euclidean distance measure. The unsupervised UBS algorithm has the advantage over the SBS method that it was not constructed to optimise a particular criterion function, but simply to decorrelate the band set. It is therefore applicable to any of the above classification methods.

Regression

Regression has been widely used in remote sensing, mainly to relate physical earth surface variables to remotely sensed measurements. In general, multiple regression allows determining whether values of a (dependent) variable x are related to values of c independent variables μ_i ($i \in [1, c]$). For a simple regression, c equals one. Equation 6.1 shows the linear form of a multiple regression model with matrix notation. Generally, the independent variables are selected by the analyst and are assumed to be without error. For each of their values, n observations of the dependent variable x are obtained. The aim is to invert the model and find regression coefficients f_i with the least squares method that result in the highest correlation (r^2 or R^2 see section 5.4) between the observed and modelled values of x, i.e., in minimal prediction errors e_j (or residuals). The latter are assumed to be independent and normally distributed random variables with zero mean and constant variance.

$$x_j = f_0 + \sum_{i=1}^{c} f_i \mu_{i,j} + e_j \qquad , i \in [1, c],\ j \in [1, n]$$
$$\mathbf{x} = \mathbf{M}\mathbf{f} + \mathbf{e} \tag{6.1}$$

where

x_j	Value of the dependent variable in observation j
f_i	Regression coefficient for independent variable i
$\mu_{i,j}$	Value of the independent variable i in observation j
c	Number of independent variables
n	Number of observations
$\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$	Observation vector, dimension (n,1)
$\mathbf{f} = (f_0, f_1, \cdots, f_c)^T$	Vector of coefficients, dimension (c,1)
$\mathbf{M} = (\mathbf{1}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_c)$	Matrix of independent variables, dimension (n, c)

$\mathbf{1} = (1,1,\cdots,1)^T$	Unit vector, dimension here (n,1)
e_j	Error for observation j of the fit of independent variables
$\mathbf{e} = (e_1, e_2, \cdots, e_n)^T$	Vector of random errors, dimension (n,1)

A band selection method for regression may aim to select a band set, which minimises the absolute error of the regression result. The root-mean-square-error (RMSE) calculated between the modelled and known inversion result for a given validation data set, is an absolute error measure for any type of model inversion (Settle and Drake, 1993; Townshend *et al.*, 2000). For example, Weiss *et al.* (2000) inverted the SAIL radiative transfer model to estimate canopy biophysical variables from remotely sensed reflectance data. For each canopy variable, they selected a band set that gave the best estimation performance in terms of minimum absolute RMSE.

The absolute RMSE includes the model fit as well as the data dependent generalisation error of the model, that is, the model's performance with respect to unseen data. For supervised band selection, however, the main interest lies in the fit of a certain band (subset) to the model, which is captured in a goodness-of-fit measure such as the coefficient of determination (r^2 or R^2) between the observed and modelled values of the dependent variable.

An example for band selection for simple regression models is Leckie *et al.* (1988), who aimed to find airborne scanner bands which can be used to assess defoliation caused by the spruce budworm. They performed linear and quadratic regressions to relate band reflectance data and damage symptom quantities, and the bands with a high and significant correlation coefficient (0.1% and 1% significance level) were considered as highly suitable for the task. Another example is Thenkabail *et al.* (2002), who selected optimal sensor bands for characterising biophysical variables of agricultural crops and yield. They related *in situ* measurements of these variables directly to vegetation indices (and not bands) calculated from both narrow and broad wavebands using field spectroradiometer and Landsat-5 TM reflectance data, respectively. The best vegetation indices were the ones that gave the highest r^2 for the exponential regression model, and the best bands were selected as the ones used in the three best vegetation index models.

The latter two examples represent single-band selections, which do not consider the suitability of the entire band subset for the regression model. Hyperspectral bands of a remotely sensed data set are usually correlated, so the best band subset does not necessarily consist of the best individual bands. Feature selection search techniques (see section 2.2.3) may be applied to find the best band subset at any given dimension. As an exhaustive search may quickly become computationally infeasible, a sub-optimal method such as the sequential forward or backward selection may be employed. To note is that the minimum number of bands required for obtaining a regression solution equals to $c+1$, and in case of the forward selection, an exhaustive search needs to be performed for the band set dimension $c+1$ before further bands may be added via forward selection search. The

backward elimination method may in total be computationally less expensive than the modified forward selection search. It performs a regression on all bands and deletes the band iteratively that contributes least to the coefficient of determination R^2 when entered last.

Nevertheless, band selection methods using the coefficient of determination as criterion function require a regression calculation for each criterion value and therefore come at a high computational cost. Unfortunately, a surrogate for a goodness-of-fit measure entirely based on the given values of the independent variables does not exist.

One of the assumptions in regression is that the independent variables are linearly independent, i.e., they have no perfect correlation between them. In case variables were collinear, the regression model could not be inverted. That is, for the linear example in equation 6.1, the matrix $\mathbf{M}$ (or $\mathbf{M}^T\mathbf{M}$) becomes singular and its inverse does not exist. Although *near* collinearity between independent variables allows for model inversion, it results in a very large standard error of their regression coefficients (Edwards, 1984). That is, the coefficients become unstable and may vary substantially depending on which other independent variables are included. Consequently, an accurate regression solution relies on the independence of the model variables, and a supervised band selection routine may be designed to eliminate those bands from the set that result in an increased variable correlation. The correlation between the variable vectors for each band may be measured via the angle between them. That is, the SBS algorithm framework could be applied to band selection for regression if the criterion function was replaced by an angle measure between variable vectors. As the latter is not a single-band measure, the band pair resulting in the largest angle between the given variables is selected first.

The unsupervised UBS algorithm decorrelates the observations. This may be of advantage for regression, as collinear observations prevent regression model inversion and near-collinear ones result in increased variances of the estimated coefficients. The regression model assumes observations to be independent and vary randomly around the regression line.

As a result, both the modified SBS and the UBS algorithm may be applied for band selection for regression. However, in both cases, the procedure would not select bands the model fits best, but those that potentially reduce the standard error of the regression coefficients. In both cases, the band number determination procedure is not applicable, as for regression the number of independent random observations should be considerably larger than the number of variables to reduce the prediction variance (Miller, 1990).

Linear Spectral Unmixing

Linear spectral unmixing (LSU, e.g. Horwitz *et al.*, 1971) is an estimation method which inverts the linear spectral mixture model (LSMM). The latter formulates the radiance captured by each instantaneous field of view (IFOV) as a linear sum of the individual material radiances ('spectral endmembers', EMs) received from the scene model elements. The weight coefficients of the sum correspond to the area proportions of the EMs within the ground-projected instantaneous field of view (GIFOV). The LSMM corresponds to regression equation 6.1, where the EMs represent the independent variables, the sensor radiance (or reflectance) the dependent variable and the area proportions the regression coefficients. The constant regression coefficient may stand for effects not explained by the EMs (Nielsen, 1999). Sometimes the column of ones in $\mathbf{M}$ is replaced by a column of zeros to represent the 'total shade' EM with 0 % reflectance in all bands (Tompkins *et al.*, 1997), but more often than not the constant coefficient is removed from the equation. The regression equation may be constrained by forcing the coefficient values to sum to one or restricting them to be greater than zero.

With respect to band selection, Lévesque *et al.* (1998) investigated the effects of varying bandwidth and band number on (constrained) spectral unmixing results for a 68-band CASI data acquired over a mine tailings site with five EM spectra. The absolute mean difference (AMD) between the all-bands unmixing result and the band-expansion results increased for growing bandwidths, while the AMD for band reduction was below 2 and 9%, when using only 8 and 4 uniformly-spaced bands, respectively. This illustrated that broadband sensors may be limited in their ability to separate spectrally between EM, and that using the minimum number instead of all observations may result in comparable results (AMD below 10%).

Chang *et al.* (1999) described a supervised band selection method for linear spectral unmixing with a regression model that separates between target and background EMs. In this method, pixels are first projected into the subspace orthogonal to the background (orthogonal subspace projection, OSP, Harsanyi and Chang, 1994). Second, target signals are separated from noise maximising the SNR matrix via a matched filter, which is equivalent to solving a generalised eigenvalue problem. From the eigenvalues and –vectors, the discriminant power (see section 2.3.5) of each band is calculated to prioritise the bands. That is, bands are ranked according to their ability to transmit only the part of the target EM signal that is orthogonal to the background EMs. A similar approach was presented by Karlholm and Renhorn (2002), who projected the signal onto the subspace orthogonal to the background, and then estimated the SNR by decomposing the projected signal into orthogonal components in a target and noise subspace. The bands were selected that transmit the undistorted part of the target signal orthogonal to background. However, the latter two examples represent single-band selections, which do not take the band subset performance into account.

Regarding subset selection for LSU and the suitability of the SBS and UBS algorithms, the same comments apply as for regression (see above). These results are also valid for extensions of LSU, such as OSP, constrained energy minimisation (Farrand and Harsanyi, 1997) and support vector machines (Brown *et al.*, 1999).

Spectral Angle Mapper

The spectral angle mapper (SAM, e.g. Mather, 1999) allows the mapping of the spectral similarity of each image spectrum to a given reference spectrum. The similarity is measured by the angle between an image pixel vector and the reference pixel vector, which is invariant to the lengths of the vectors, and therefore to illumination effects. The SBS algorithm is applicable if the criterion function is replaced by an angle measure (see table 2.1), which has to be minimised to achieve high class separation. In addition, the algorithm needs to be modified and start by selecting the band pair resulting in the largest angle between the given classes, as the criterion is not a single-band measure. The UBS can be employed as it reduces the redundancy between bands and therefore the need to estimate vector similarity for equivalent bands.

6.5.2 Use with other Scene, Atmospheric and Illumination Conditions

The band selection methods described in chapter 4 and 5 assume that the scene, atmospheric and illumination conditions do not change significantly between the hyperspectral image flight and the following multispectral image flight. The multispectral band configuration corresponds to the output of the band selection algorithm which uses the hyperspectral image as input. If, however, significant changes in atmospheric and illumination conditions occur between data acquisitions, hyperspectral data measured for the target conditions need to be either simulated or calibrated to reflectance (see section 6.6).

6.5.3 Use with other Imaging Spectrometers

The band selection methods are generally directly applicable to all programmable imaging spectrometers that measure a continuous spectrum for each pixel and possess an on-board channel summation capability. Examples include the CASI-2, the Airborne Imaging Spectrometer (AISA) and the Reflective Optics System Imaging Spectrometer 03 (ROSIS-03). If bands were to be selected for a sensor different to the one with which the hyperspectral input data were acquired, the hyperspectral data set would need to be modified to simulate measurements of the target sensor. Data simulation is described in the next section.

6.6 Data Simulation

The selection of a band set for future multispectral airborne or satellite sensors or unavailable programmable imaging spectrometers may also be accomplished via the SBS and UBS band selection methods presented in chapters 4 and 5, respectively. The latter algorithms require high spectral resolution sensor data as input, but as these may not be available for the given target sensors, they need to be simulated with data from other hyperspectral imaging or non-imaging sensors.

For correct data simulation, sensor characteristics as well as atmospheric and illumination conditions need to be taken into account for both the target and the simulating sensor. In this section, the author focused exclusively on sensor attributes such as spectral, spatial and signal response. Transforming the available hyperspectral data into radiance will correct for dark current noise and calibrating them to apparent reflectance reduces atmospheric and illumination effects. That is, simulating data in reflectance units eliminates the need to consider prevailing atmospheric and illumination conditions. If at-sensor radiance was the desired output unit of the simulation, a radiative transfer model could be used (e.g. Isaacs and Vogelmann, 1988; Kerekes and Landgrebe, 1989; Fischer and Fell, 1999). For both target and simulating sensor, nadir viewing was assumed, reducing the effects of oblique viewing on the band selection result. For the use of orbital, platform, or attitude models the reader is referred to O'Neill and Dowman (1993) and Schowengerdt (1997).

6.6.1 Sensor Spectral Response

Two cases of band selection for different sensors may be distinguished:

1) band selection for a non-existing sensor (e.g. in sensor design studies), and
2) band selection for an existing but unavailable programmable imaging spectrometer.

In both cases, it is essential that the entire wavelength range of the target sensor is covered by the hyperspectral data set used. In the first case, data simulation will not be necessary as none of the target sensor bands are specified. Nonetheless, the finer the spectral resolution of the base data set, the more optimal the bands will be selected with respect to their width and placement.

In the second case, the spectral resolution of the base data needs to be finer than the finest resolution of the target sensor in order to allow for an accurate band simulation. A target sensor band may be simulated by the simple average of all narrow bands that fall within its spectral response. For example, Thenkabail *et al.* (2000) simulated spaceborne sensor bands from field spectrometer data by averaging the narrow band data for the corresponding band ranges.

If the target sensor's spectral response function (also referred to as bandpass or transmittance function) was known a more accurate band simulation would result. The simulated band response would then be calculated as the spectrally weighted average of the narrow band responses over the spectral bandpass. For example, Wetzel (1995) simulated spectral bands of two shortwave-infrared AVHRR channels with hyperspectral AVIRIS data by using a spectral transmittance function similar to the existing AVHRR shortwave channel transmittance.

However, for some imaging spectrometers, such as the CASI-2, the spectral response function is not available (Riedmann, 2003). Normally, it is considered to be Gaussian in shape but with varying bandwidth across the spectral range (Schowengerdt, 1997).

6.6.2 Sensor Spatial Response

When simulating the spectral response of a sensor, it is important that the size of the simulating sensor's GIFOV matches the one of the target sensor, so that both sensors capture the same scale of scene variation. If this was not the case one sensor would measure small-scale scene variation, the other large-scale variation, leading to fundamentally different sensor responses. That is, the simulating sensor's GIFOV measurements need to be regularised to the GIFOV size of the target sensor, if the GIFOVs were different. This implies that the spatial resolution of the simulating sensor needs to be equal or finer than the resolution of the target sensor, as it is more difficult to scale down, i.e. simulate small-scale from large-scale variation. In addition, the base data set should cover the entire spatial extent of the scene of interest.

Generally, two types of simulating sensor may be distinguished:
1) imaging sensor (e.g. an airborne imaging spectrometer), and
2) non-imaging sensor (e.g. a field spectroradiometer).

The following paragraphs discuss data simulation using data from these two types of sensors.

Simulation with Data from Imaging Sensors

An imaging sensor may simulate the spatial response of another one flown or orbiting at higher altitude. Up-scaling of image data may be performed in two consecutive steps:
1) discrete convolution (filtering), and
2) resampling.

Convolution filters are moving windows that operate on a relatively small neighbourhood of the windows centre pixel. A linear convolution filter is a weighted sum of the pixels within the moving window. The moving window needs to be at least the size of the target sensor's GIFOV. The image

coordinate of the output value is the same as the one of the current centre pixel of the window. The window is moved over every possible image pixel position, blurring the image according to the nature of the filter function. To note is that the sampling between pixels during the convolution remains constant. In the second step, the image is resampled to the desired sampling interval. Usually the spacing between pixels is chosen to equal the GIFOV size of the target sensor. Resampling is achieved by interpolating new pixel values between the convoluted pixels using nearest neighbour, bilinear, or cubic interpolation.

In this context, the convolution filter corresponds to the point-spread function (PSF) of the target sensor, which describes the spatial response of the sensor due to its optics, image motion, detector and electronics (Schowengerdt, 1997). The PSF represents the spatial irradiance of a point source on the detector in the sensor focal plane. The modulation transfer function (MTF) is the equivalent of the PSF in the spatial frequency domain. To be precise, the MTF equals to the magnitude of the Fourier transform of the PSF (Bretschneider, 2002).

If the PSF of the target sensor is unknown, an idealised square wave response is often assumed and the centre pixel value is calculated as the average of all pixels within the moving window. In this case, the GIFOV of the target sensor generally equals multiple times the pixel size of the simulating sensor. For example, Woodcock and Strahler (1987) degraded Landsat Thematic Mapper data using this simple averaging technique. Marceau *et al.* (1994) degraded Multi-detector Electro-optical Imaging Scanner (MEIS) II airborne imagery and then resampled to four progressively coarser spatial resolutions (5, 10, 20, and 30 m). Wetzel (1995) simulated the spatial response of two AVHRR channels with 20 m AVIRIS image data by averaging 55 by 55 pixels to give the AVHRR footprint of 1.1km. The blurred image was successively down-sampled by the same window size.

If the PSF or MTF of the target sensor is given, a more realistic sensor response may be estimated. For example, Schowengerdt (1997) simulated Landsat Thematic Mapper data by scanning an aerial photograph with a 2-meter GIFOV, blurring the image with the known PSF and subsampling the filtered image from 2 to 30 m pixel spacing. If the PSF or MTF of the simulating sensor is also known, it may be equally accounted for. Justice *et al.* (1989) simulated lower resolution data from Landsat MSS imagery using both the original MTF and the target MTF. The target MTF was based on the original one, but using different settings for ground-projected sensor blur and target detector dimension. The final spatial filter for image degradation was then created as the ratio between the target MTF and the original MTF.

Simulation with Data from Non-Imaging Sensors

In the literature, field spectrometer data have been used for the selection of airborne and satellite sensor bands, e.g. Thomson *et al.* (1998a) and Dekker *et al.* (1992), respectively. However, in the latter examples, the authors did not scale-up field measurements at ground to match the

hypothetical measurements to be made with the target sensor at much coarser spatial resolution. In general, the larger GIFOV of the target sensor spatially averages small-scale class variation measured with the field spectrometer.

To simulate the large-scale class variation seen by the target sensor, field spectral measurements need to be sampled and averaged over an area of the size of the target GIFOV. A carefully designed sampling strategy is needed to define both sampling scheme and intensity. With respect to the sampling scheme, systematic sampling has been shown to be more efficient than random sampling (Atkinson, 1997), as it ensures a uniform coverage across the pixel area. A square grid is therefore recommended.

The sampling interval depends on the GIFOV size of the two sensors and the spatial homogeneity of the surface under investigation. Two cases may be distinguished: oversampling and undersampling. In the former case, the GIFOV of the spectroradiometer is larger than the ground sampling interval, while it is smaller in the latter case. Ideally, the pixel area would be slightly oversampled to ensure a continuous surface coverage. However, as the resources and time available to any given field work project are generally scarce, the sampling effort could be considerably reduced by increasing the GIFOV of the spectrometer for a better approximation of the GIFOV of the target sensor (preferably an increase in height above ground, as a larger FOV averages bidirectional effects). Undersampling is possible for surfaces that are homogeneous at sub-pixel scale. To quantify large-scale class variation, pixels from other scene locations need to be sampled in a similar way. An example of undersampling is given by Thenkabail *et al.* (2000). They simulated 30 m Landsat TM bands of agricultural crops (cotton, potato, soybeans, corn and sunflower) with data from a field spectrometer with a GIFOV-diameter of 38 cm. Spectral measurements were taken every 10 m along transacts of 30 to 100 m in length.

In general, high-resolution airborne imaging data should be preferred to non-imaging field spectrometer data for the simulation of hyperspectral airborne or spaceborne imagery. The amount of fieldwork needed to adequately sample class areas may be substantial, especially for heterogeneous surfaces with many classes. In addition, extrapolation from non-image to image data represents a weak approximation of the real data. The SBS and UBS algorithms were developed for image data only, but can be programmed to accommodate field spectral data stored in ENVI spectral library format (.spl).

6.6.3 Sensor Signal Characteristics

Theoretically, the radiometric resolution of the simulating sensor needs to be equal to or finer than the one of the target sensor for an accurate data simulation. However, due to advances in senor technology, the new generation of sensors possess very high radiometric resolution equal to or

beyond 12-bit (CASI-2 or Hyperion 12 bit; HyMAP 16 bit). Differences between a 12-bit and 16-bit version of the same image seem negligible, as the signal will be still sufficiently quantised with 12-bit. Obviously, 16-bit data differ significantly from 6-bit data (e.g. Landsat MSS sensor).

Apart from the radiometric resolution, the noise characteristics of the target sensor need to be taken into consideration. If hyperspectral airborne or spaceborne data are simulated from higher spatial and spectral resolution data, the averaging during the degradation process reduces the noise levels of the original data to an unrealistically low level. Justice *et al.* (1989) reported an 80% decrease in sensor noise due to digital filtering. In addition, high-resolution data from field measurements or low-altitude airborne acquisitions generally possess a higher SNR than comparable data collected at higher altitudes from airborne or spaceborne sensors.

Noise levels of the target sensor bands may be restored to approximately realistic levels by adding Gaussian random noise of zero mean and one noise standard deviation. The noise standard deviation may be derived from dark current measurements of the target sensor. As the simulated data are most practically measured in apparent reflectance, the noise standard deviation needs to be scaled from the target sensor's DN value to reflectance. The scaling factor may be calculated as the ratio between the maximum possible response values of reflectance (100%) and that of the sensor's DN value (4095 DN for CASI-2).

In this section, data simulation was based on data calibrated to apparent reflectance units in order to circumvent atmospheric and illumination modelling. The next section discusses the need for data calibration to radiance or apparent reflectance in the context of band selection.

6.7 The Need for Calibration

During the period 1999 to 2002 the author was responsible for the laboratory calibration of the NERC ARSF CASI-2 imaging spectrometer and the procedure adopted for this has been published by Riedmann and Rollin (2000) and Riedmann (2003). Section 4.2.1 gave a short overview of common reflectance calibration methods. In this section, the benefits of both radiance and reflectance calibration are considered for remote sensing, and the relevance of data calibration for band selection will be discussed.

Radiometric or *sensor* calibration aims to transform the sensor response (in DN) to Système International (SI) units (at-sensor spectral radiance, $\mu Wcm^{-2}sr^{-1}nm^{-1}$) to make the measurements independent of the instrument. It is especially important for

- the comparison of data acquired with different sensors (for across- or multi-sensor products)

- the comparison of data acquired at different times as the sensor response may drift over time (for multi-temporal products such as image mosaics or land cover change)
- quantitative techniques that depend on inter-band relationships (such as band ratios), as detectors within a sensor may respond differently to a spectrally uniform signal,
- the use of radiance data for physical models (e.g. radiative transfer models).

However, at-sensor radiance is also a function of atmospheric conditions, solar illumination, sensor-sun viewing-illumination geometry, topographic slope and aspect. Accurate multi-sensor and multi-temporal data products should account for these factors, too.

Reflectance or *scene* calibration eliminates the influence of the atmosphere and illumination on the data signal by converting it from at-sensor radiance to units of surface reflectance. As surface reflectance is independent of the sensor, the reasons for radiometric calibration are also valid for reflectance calibration. In addition, reflectance calibration is vital for

- the comparison of data to other reflectance data, e.g. acquired in the laboratory or in the field,
- quantitative techniques that exploit spectral features of the reflectance curve, e.g. precise absorption band-depth measurements or absorption feature detection, and
- the use of reflectance data for physical models.

The effects of sensor viewing and solar illumination geometry, as well as topography, on remote sensing data are significant but will not be addressed here. The reader is referred to Mather (1999) for a discussion of correction methods.

With regards to the band selection algorithms developed in this thesis, neither radiance nor reflectance calibration are strictly necessary if it can be assumed that the sensor response and atmospheric and illumination conditions will remain constant between the hyperspectral and multispectral data acquisitions. However, if significant changes in sensor response or atmospheric and illumination conditions occurred between the two acquisitions, for example, due to a large time lag or multispectral data being collected with a different sensor, reflectance calibration would be beneficial to guarantee optimal band selection. Obviously, if the application requires radiance or reflectance data as input, band selection should be performed on calibrated data.

The author believes that, in principle, surface reflectance calibration should be part of any data pre-processing routine, if the interest of the analyst lies in surface properties, for example as for land cover classification. However, so far, different sensor and scene calibration procedures of different accuracy have been applied across the remote sensing research community. To make data as comparable as possible calibration methods need to be standardised. Validation of data for Earth observation services (ValDEOS) is a UK initiative that aims to integrate the existing capability for data calibration and validation across Europe in a network of excellence. It plans to develop a

uniform approach towards validation and calibration under which nationally funded programmes are likely to align (see BNSC/NPL on-line e-survey; NPL, 2003).

6.8 Summary

When applied to the given real data sets, the SBS, UBS and CSBS algorithms achieved similar MLC accuracies, indicating that all four algorithms may be equally applied for the given data and class set for band selection. However, the SBS resulted in more optimal bands, especially in the search for the single best bands. SBS is the preferred band selection method for a supervised MLC, as UBS cannot determine the optimal maximum bandwidth and depends on the availability of dark image data for band expansion. CSBS has the drawback of producing inconsistent results depending on the initialisation and parameter settings of the clustering routine. However, the ultimate choice of method is very much dependent on the requests of the data user, and the UBS method may be chosen simply because it may be applied in-flight.

When applied to the given data sets, band selection is principally effective only for the first few bands of the selected set. The number of effective bands corresponds to the intrinsic discriminant dimensions of the data sets. In addition, a band expansion experiment on the given data sets showed that for two out of three data sets hyperspectral data were found not to be of significance for the corresponding classification tasks. That is, multispectral data with less and broader bands would achieve similar accuracy as the full hyperspectral data set.

With respect to other hyperspectral data applications, UBS may be used without modification for all cases examined. In contrast, SBS can only be employed to other hard classification methods if the criterion function used reflects the complexity of the classifier. For example, the Euclidean distance measure should be employed when classifying with the Minimum Distance method. For regression, linear spectral unmixing or spectral angle mapper, the SBS criterion function needs to be replaced by the spectral angle between variable or class vectors.

SBS and UBS may be employed with any other programmable imaging spectrometers such as AISA or ROSIS-03. If band selection was carried out for a sensor different to the one with which the hyperspectral input data were acquired, for example for sensor design studies, the hyperspectral data of the target sensor would need to be simulated. If the spectral response of target sensor was known, data of spectral resolution finer than the one of the target sensor should be used for simulation. Likewise, the spatial resolution of the simulating sensor needs to be equal or finer than the resolution of the target sensor. If the spatial sensor response was simulated with data from an imaging sensor, data would first need to be convoluted with the PSF of the sensor and then resampled. If non-imaging data from field spectrometers was used for spatial data simulation, the GIFOV of the target sensor would need to be sampled systematically in a square grid with the

sampling interval depending on the GIFOV size of the two sensors and the spatial uniformity of the target classes. The closer the GIFOVs are in size, and the more homogeneous the classes, the less sampling is needed. In addition to the spatial and spectral response, the target sensor noise has to be considered as well for data simulation. Noise levels may be derived from the target sensor's dark current and added as Gaussian noise with zero mean and one standard deviation.

If it cannot be assumed that the sensor response and atmospheric and illumination conditions will remain constant between hyperspectral and multispectral data acquisitions, data should be calibrated to radiance and apparent reflectance, respectively, to guarantee optimal band selection. This is also valid if the data application requires calibrated data as input. With respect to data simulation, calibrating hyperspectral data to apparent reflectance allows to circumvent modelling of the existing atmospheric and illumination conditions at the target sensor.

7 Summary and Conclusion

7.1 Rationale and Objectives

During the last twenty years advances in sensor technology have produced a new generation of airborne and satellite sensors that can acquire data in many narrow bands. With hyperspectral data from these sensors, an improved discrimination of ground targets may be possible. However, this increased capacity comes at a substantial cost:

- First, some sensors, such as the CASI-2 or CHRIS, are limited in the amount of data they can record. In order to acquire a large number of bands, the amount of spatial pixels (i.e. the swath width) needs to be reduced.
- Second, very fine bands give a low signal-to-noise ratio, unless coarser spatial resolution is used. For example, the first spaceborne imaging spectrometer, Hyperion, records a continuous spectrum with 10 nm wide bands, but offers a spatial resolution of 30 m. The latter is relatively coarse compared to a possible 1 m resolution of the Ikonos satellite sensor (see table 1.2).

The question may then be asked whether hyperspectral data are of benefit for a given application task. In other words:

Is it possible to reduce the number of hyperspectral bands and broaden their width, while achieving the same or higher application accuracy as with the original hyperspectral data set?

This study aimed to answer this question by developing band selection methodologies using hyperspectral data from airborne and satellite sensors as input. So far, band sets for airborne and spaceborne sensors have been designed mainly on the basis of established band sets of other sensors. Alternatively, they have been chosen using feature selection or extraction methods. But the latter methods are unable to provide an answer to the above question.

7.2 Innovative Methods

To answer the above question, all band selection methods developed in this thesis introduce both bandwidth and band number as variables into the band selection process. That is, band location, width and number are optimised with respect to the accuracy of the application task at hand. Maximum Likelihood classification was chosen as application in the design of the algorithms, but the use of the band selection methods with respect to other applications was discussed.

The supervised band selection (SBS) procedure is based on conventional feature selection techniques using user-specified classes. However, it outperforms other published feature selection-based methods in the choice of criterion function and algorithm structure. Both the band expansion and band number determination procedures are based on the criterion function.

For the case that *in situ* data or classification schemes are unavailable, an unsupervised band selection (UBS) method was designed, which aims to find the least redundant bands within the band set on the basis of the correlation coefficient. While the optimal number determination method uses also the correlation coefficient, the expansion procedure is based on the band signal-to-noise ratio. Unlike PCA, UBS decorrelates the band set without data transformation, preserving the physical meaning of the bands. The clustering-based SBS (CSBS) algorithm is an alternative to UBS, which uses clustering to define the classes needed as input for the SBS.

All algorithms were implemented in computationally efficient IDL™ (version 5.5 Win 32 x86) and ENVI™ (version 3.5) programs that can be run in ENVI™. Besides for band selection, they may also be used for

- feature selection processing,
- determining the best three bands for colour composites to visually discriminate between the given classes or display maximum information,
- visualising discriminant or uncorrelated spaces (e.g. for endmember derivation),
- ordering a given band set according to information content or discrimination ability, and
- adding least redundant or most discriminative bands to a pre-defined band set.

7.3 Method Evaluation

The band selection approaches developed were evaluated with two real data sets from saltmarsh areas in the Mid Severn Estuary and heathland areas in the New Forest, UK. The SBS was further tested on a data set from the Tregaron bog, UK. All algorithms gave physically meaningful band sets, which achieved similar or higher MLC accuracies than vegetation and coastal band sets from current airborne and satellite sensors. The sub-optimality of the SBS band set was found to be no more than 6% for sets with maximum three bands, while the UBS band set achieved a maximum accuracy loss of 7% for two and three-dimensional band sets. The CSBS set was found to be less optimal. The band number criteria were shown to be effective estimates of the intrinsic dimensionality of the data sets. However, some subjectivity is still present in the latter criteria regarding the threshold being used (95% or 99%).

Generally, SBS is the preferred band selection method for supervised MLC as data application. Only SBS may be used to test whether narrow band data have a significant advantage over broad band data for the given classification tasks at hand. UBS uses the SNR for band expansion, which is not

related to MLC accuracy and depends on the availability of dark image data. Furthermore, UBS (as PCA) requires each band to be normally distributed, which is only justified if the scene is made up of a single material type. CSBS has the drawback of producing inconsistent results depending on the initialisation and parameter settings of the clustering routine. If no classes are available, UBS should be preferred to CSBS due to the inconsistency and sub-optimality of the latter.

7.4 Results

A band expansion experiment was performed with the SBS algorithm to test whether hyperspectral data gave a significantly higher MLC accuracy than multispectral data for three classification tasks: heathland vegetation in the New Forest, salt marsh vegetation on the River Severn and bog condition classes on the Tregaron bog.

- As for the New Forest classification task, two carefully placed bands of 370 nm seem to be achieving at least 95% of the classification accuracy that 117 15 nm-wide bands reach.
- Regarding the Tregaron classification task, five bands of 57 nm seem to result in at least 99% of the accuracy of 37 12 nm-wide bands.
- With respect to the River Severn classification task, three bands seem to achieve at least 95% of the classification accuracy of 60 bands. However, narrow bands were found to give a significantly higher accuracy than broader bands.

That is, for all three classification tasks, the number of bands to acquire could be reduced dramatically, enabling an increase in swath width. For two out of three tasks, coarsening the spectral sensor resolution may be justified e.g. in favour of a refinement in the sensor's spatial resolution.

7.5 Use of Methods

Band selection is specific to the characteristics of 1) the scene, 2) the sun, 3) the atmosphere, 4) the sensor, and 5) the final data application. To find a band set that is valid for the target data acquisition, all dependencies have to be considered.

According to whether target sensor is programmable or not, the band selection methods may be applied in two different ways. Using *programmable* airborne sensors, such as the CASI-2, AISA or ROSIS-03, the first four attributes may be taken into account by using hyperspectral data of a representative part of the scene acquired with the sensor of interest under similar solar and atmospheric conditions. Application-specific band selection may then be performed on the hyperspectral data, and multispectral data may be collected subsequently using the optimised band set. The approach requires the time gap between hyperspectral and multispectral acquisitions to be

minimised to avoid changes in the class spectral responses, atmospheric and illumination conditions. Both the UBS and CSBS may be readily employed in-flight. The SBS is inappropriate for in-flight application as it depends on a user-defined class set. Reflectance calibration of the hyperspectral data set may be necessary, if atmospheric and illumination conditions changed between acquisitions.

For *non-programmable* airborne or satellite sensors, a more generic band set may be sought for a given classification scheme (e.g. land cover mapping), which needs to be optimised to a large number of scenes. On the search of such a band set, long-term temporal changes in the scene, sensor, atmosphere, illumination need to be taken into account. As hyperspectral data of the target sensor are often not available, in particular for sensor design studies, they need to be simulated in units of apparent reflectance to consider changes in the sensor and atmospheric and illumination conditions. With respect to the validity of the band set regarding short-term atmospheric and illumination changes, a sensitivity analysis may be performed that systematically varies atmospheric or illumination parameter settings for a given scene. The SBS would be the most suitable band selection method for this case.

7.6 Conclusion

The author believes that current data acquisition is inefficient in that spectrally redundant data are collected with most imaging spectrometers, often using narrow band data where this is not necessary. A more efficient data acquisition with respect to the number of bands collected would allow obtaining supplementary data, e.g. additional bands for other data applications, or more spatial pixels for an increased swath width or angular data. If narrow bands were found to result in similar accuracy as broad bands, the spatial resolution of the sensor could be refined.

This research project has demonstrated the potential of three innovative band selection methods for imaging airborne and satellite sensors. They may be used as tools to optimise a sensor band set regarding the number, width and location of bands, and therefore, increase the efficiency of data acquisition. Coupled with algorithms to optimise other acquisition parameters, e.g. spatial resolution, they lead the way towards an intelligent Remote Sensing expert system for data acquisition.

APPENDIX A

SENSOR BAND SETS

Table A.1: Environment Agency (EA) CASI band specifications (October 1997, enhanced spectral mode) (Source: EA, 1997).

Band	Centre Wavelength (nm)	FWHM (nm)	Band start	Band End	Region
1	412.4	8.4	408.2	416.6	VIS
2	419.7	8.4	415.5	423.9	VIS
3	427	8.4	422.8	431.2	VIS
4	434.3	8.4	430.1	438.5	VIS
5	441.6	8.4	437.4	445.8	VIS
6	449	8.6	444.7	453.3	VIS
7	456.3	8.6	452	460.6	VIS
8	463.7	8.6	459.4	468	VIS
9	471.1	8.6	466.8	475.4	VIS
10	478.4	8.6	474.1	482.7	VIS
11	485.9	8.6	481.6	490.2	VIS
12	493.3	8.6	489	497.6	VIS
13	500.7	8.6	496.4	505	VIS
14	508.1	8.6	503.8	512.4	VIS
15	515.6	8.6	511.3	519.9	VIS
16	523.1	8.6	518.8	527.4	VIS
17	530.5	8.6	526.2	534.8	VIS
18	538	8.6	533.7	542.3	VIS
19	545.5	8.6	541.2	549.8	VIS
20	553	8.6	548.7	557.3	VIS
21	560.5	8.6	556.2	564.8	VIS
22	568.1	8.6	563.8	572.4	VIS
23	575.6	8.6	571.3	579.9	VIS
24	583.1	8.6	578.8	587.4	VIS
25	590.7	8.6	586.4	595	VIS
26	598.2	8.6	593.9	602.5	VIS
27	605.8	8.6	601.5	610.1	VIS
28	613.4	8.6	609.1	617.7	VIS
29	621	8.6	616.7	625.3	VIS
30	628.5	8.6	624.2	632.8	VIS
31	636.1	8.6	631.8	640.4	VIS
32	643.7	8.8	639.3	648.1	VIS
33	651.3	8.8	646.9	655.7	VIS
34	658.9	8.8	654.5	663.3	VIS
35	666.6	8.8	662.2	671	VIS
36	674.2	8.8	669.8	678.6	VIS
37	681.8	8.8	677.4	686.2	VIS
38	689.4	8.8	685	693.8	VIS
39	697.1	8.8	692.7	701.5	VIS
40	704.7	8.8	700.3	709.1	VIS
41	712.3	8.8	707.9	716.7	VIS
42	720	8.8	715.6	724.4	VIS
43	727.6	8.8	723.2	732	VIS
44	735.3	8.8	730.9	739.7	VIS
45	742.9	8.8	738.5	747.3	VIS
46	750.5	8.8	746.1	754.9	VIS
47	758.2	8.8	753.8	762.6	VNIR
48	765.8	8.8	761.4	770.2	NIR
49	773.5	8.8	769.1	777.9	NIR
50	781.1	8.8	776.7	785.5	NIR

Table A.1 continued.

Band	Centre Wavelength (nm)	FWHM (nm)	Band start	Band End	Region
51	788.8	8.8	784.4	793.2	NIR
52	796.4	8.8	792	800.8	NIR
53	804.1	8.8	799.7	808.5	NIR
54	811.7	8.8	807.3	816.1	NIR
55	819.4	8.8	815	823.8	NIR
56	827	8.8	822.6	831.4	NIR
57	834.7	8.8	830.3	839.1	NIR
58	842.3	8.8	837.9	846.7	NIR
59	850	8.8	845.6	854.4	NIR
60	857.6	8.8	853.2	862	NIR
61	865.2	8.8	860.8	869.6	NIR
62	872.9	8.8	868.5	877.3	NIR
63	880.5	8.8	876.1	884.9	NIR
64	888.1	8.8	883.7	892.5	NIR
65	895.7	8.8	891.3	900.1	NIR
66	903.3	8.8	898.9	907.7	NIR
67	910.9	8.8	906.5	915.3	NIR
68	918.5	8.6	914.2	922.8	NIR
69	926.1	8.6	921.8	930.4	NIR
70	933.7	8.6	929.4	938	NIR
71	941.3	8.6	937	945.6	NIR
72	948.9	8.6	944.6	953.2	NIR

Table A.2: HyMAP band specifications (June 2000) (Source: HyVISTA Corp. Pty. Ltd., 2000).

Band	Centre Wavelength (nm)	FWHM (nm)	Band start	Band End	Region
1	437	15	429.5	444.5	VIS
2	448.9	11.2	443.3	454.5	VIS
3	461.3	15.8	453.4	469.2	VIS
4	477.3	15.5	469.55	485.05	VIS
5	492.3	15.5	484.55	500.05	VIS
6	507.4	15.8	499.5	515.3	VIS
7	523.4	15.6	515.6	531.2	VIS
8	538.8	15.9	530.85	546.75	VIS
9	554.2	15.3	546.55	561.85	VIS
10	569.5	15.4	561.8	577.2	VIS
11	584.7	15.4	577	592.4	VIS
12	600.1	15.5	592.35	607.85	VIS
13	615.7	15.7	607.85	623.55	VIS
14	631.3	15.6	623.5	639.1	VIS
15	646.6	15.3	638.95	654.25	VIS
16	661.7	15.1	654.15	669.25	VIS
17	676.9	15.5	669.15	684.65	VIS
18	692.4	15.8	684.5	700.3	VIS
19	707.8	15.3	700.15	715.45	VIS
20	722.9	15.3	715.25	730.55	VIS
21	738.1	15.7	730.25	745.95	VIS
22	753.4	15.3	745.75	761.05	VIS
23	768.5	15.3	760.85	776.15	VIS
24	783.5	15.3	775.85	791.15	NIR

Table A.2 continued.

Band	Centre Wavelength (nm)	FWHM (nm)	Band start	Band End	Region
25	798.8	15.7	790.95	806.65	NIR
26	814.1	15.5	806.35	821.85	NIR
27	829.2	15.5	821.45	836.95	NIR
28	844.4	15.8	836.5	852.3	NIR
29	859.8	16.2	851.7	867.9	NIR
30	875	15.9	867.05	882.95	NIR
31	872.6	18.1	863.55	881.65	NIR
32	890.3	20	880.3	900.3	NIR
33	905.6	19	896.1	915.1	NIR
34	922.2	18.8	912.8	931.6	NIR
35	938	18.3	928.85	947.15	NIR
36	953.6	18.5	944.35	962.85	NIR
37	969.3	18.5	960.05	978.55	NIR
38	984.8	18.7	975.45	994.15	NIR
39	1000.9	18.7	991.55	1010.25	NIR
40	1016.5	18.8	1007.1	1025.9	NIR
41	1032	18.9	1022.55	1041.45	NIR
42	1047.5	18.8	1038.1	1056.9	NIR
43	1063	18.3	1053.85	1072.15	NIR
44	1078	18.6	1068.7	1087.3	NIR
45	1093.1	18.5	1083.85	1102.35	NIR
46	1108	18.3	1098.85	1117.15	NIR
47	1122.9	18.7	1113.55	1132.25	NIR
48	1137.6	18	1128.6	1146.6	NIR
49	1152.2	18.1	1143.15	1161.25	NIR
50	1166.7	17.9	1157.75	1175.65	NIR
51	1181.4	18.3	1172.25	1190.55	NIR
52	1196.1	18.1	1187.05	1205.15	NIR
53	1210.3	17.4	1201.6	1219	NIR
54	1224.5	17.5	1215.75	1233.25	NIR
55	1238.7	17.9	1229.75	1247.65	NIR
56	1253.2	17.7	1244.35	1262.05	NIR
57	1267.3	17.3	1258.65	1275.95	NIR
58	1281.4	17.1	1272.85	1289.95	NIR
59	1295.4	17.3	1286.75	1304.05	NIR
60	1309.4	18.1	1300.35	1318.45	NIR
61	1323.6	16.5	1315.35	1331.85	NIR
62	1336.8	16.9	1328.35	1345.25	NIR
63	1406.4	17.2	1397.8	1415	NIR
64	1420.6	15.9	1412.65	1428.55	NIR
65	1434.4	16.4	1426.2	1442.6	NIR
66	1448.5	16.5	1440.25	1456.75	NIR
67	1462.6	16.1	1454.55	1470.65	NIR
68	1476.6	16.2	1468.5	1484.7	NIR
69	1490.3	16.5	1482.05	1498.55	NIR
70	1503.9	16.2	1495.8	1512	NIR
71	1517.2	16.2	1509.1	1525.3	SWIR
72	1530.7	16.2	1522.6	1538.8	SWIR
73	1544.1	16.6	1535.8	1552.4	SWIR
74	1557.6	16.5	1549.35	1565.85	SWIR
75	1570.6	16.1	1562.55	1578.65	SWIR

Table A.2 continued.

Band	Centre Wavelength (nm)	FWHM (nm)	Band start	Band End	Region
76	1583.5	16.2	1575.4	1591.6	SWIR
77	1596.5	16.6	1588.2	1604.8	SWIR
78	1609.5	16.2	1601.4	1617.6	SWIR
79	1622.3	15.9	1614.35	1630.25	SWIR
80	1635.1	16.1	1627.05	1643.15	SWIR
81	1647.8	16.1	1639.75	1655.85	SWIR
82	1660.3	15.8	1652.4	1668.2	SWIR
83	1672.7	15.5	1664.95	1680.45	SWIR
84	1685.1	15.7	1677.25	1692.95	SWIR
85	1697.5	15.6	1689.7	1705.3	SWIR
86	1709.9	15.3	1702.25	1717.55	SWIR
87	1722.1	14.9	1714.65	1729.55	SWIR
88	1734.1	15.1	1726.55	1741.65	SWIR
89	1746.5	15	1739	1754	SWIR
90	1758.4	14.5	1751.15	1765.65	SWIR
91	1770.3	14.2	1763.2	1777.4	SWIR
92	1782.2	14.4	1775	1789.4	SWIR
93	1794	14.4	1786.8	1801.2	SWIR
94	1805.7	13.6	1798.9	1812.5	SWIR
95	1951.6	20.9	1941.15	1962.05	SWIR
96	1970.7	21.1	1960.15	1981.25	SWIR
97	1989.5	21	1979	2000	SWIR
98	2008.2	20.7	1997.85	2018.55	SWIR
99	2027.1	20.9	2016.65	2037.55	SWIR
100	2045.9	20.9	2035.45	2056.35	SWIR
101	2064.5	20.7	2054.15	2074.85	SWIR
102	2082.7	20.4	2072.5	2092.9	SWIR
103	2100.8	20.2	2090.7	2110.9	SWIR
104	2118.7	20.3	2108.55	2128.85	SWIR
105	2136.7	20.4	2126.5	2146.9	SWIR
106	2154.7	20.3	2144.55	2164.85	SWIR
107	2172.2	19.6	2162.4	2182	SWIR
108	2189.2	19	2179.7	2198.7	SWIR
109	2206.6	20.7	2196.25	2216.95	SWIR
110	2224.8	18.8	2215.4	2234.2	SWIR
111	2241.7	19.2	2232.1	2251.3	SWIR
112	2259.4	19.4	2249.7	2269.1	SWIR
113	2276.6	18.2	2267.5	2285.7	SWIR
114	2293.5	18	2284.5	2302.5	SWIR
115	2310	17.8	2301.1	2318.9	SWIR
116	2326.3	17.8	2317.4	2335.2	SWIR
117	2342.8	17.9	2333.85	2351.75	SWIR
118	2359.7	18	2350.7	2368.7	SWIR
119	2375.8	17.7	2366.95	2384.65	SWIR
120	2391.8	17.4	2383.1	2400.5	SWIR
121	2407.6	17.3	2398.95	2416.25	SWIR
122	2423.6	16.5	2415.35	2431.85	SWIR
123	2439.3	16.3	2431.15	2447.45	SWIR
124	2455	16.8	2446.6	2463.4	SWIR
125	2470.5	17	2462	2479	SWIR
126	2485.9	16.5	2477.65	2494.15	SWIR

Table A.3: NERC CASI-2 band specifications for the Tregaron bog overflight in enhanced spectral mode (12 October 2001) (Source: NERC, 2001).

Band	Centre Wavelength (nm)	FWHM (nm)	Band start	Band End	Region
1	409.56	11.41	403.85	415.26	VIS
2	420.63	11.44	414.91	426.35	VIS
3	431.72	11.46	426.00	437.45	VIS
4	442.85	11.48	437.11	448.58	VIS
5	453.99	11.50	448.24	459.74	VIS
6	465.16	11.52	459.40	470.92	VIS
7	476.35	11.54	470.58	482.12	VIS
8	487.57	11.55	481.79	493.34	VIS
9	498.80	11.57	493.02	504.59	VIS
10	510.06	11.59	504.26	515.85	VIS
11	521.33	11.60	515.53	527.13	VIS
12	532.62	11.62	526.81	538.43	VIS
13	543.94	11.63	538.12	549.75	VIS
14	555.26	11.65	549.44	561.09	VIS
15	566.61	11.66	560.78	572.44	VIS
16	577.97	11.67	572.13	583.80	VIS
17	589.34	11.68	583.50	595.18	VIS
18	600.73	11.70	594.88	606.58	VIS
19	612.13	11.71	606.28	617.98	VIS
20	623.54	11.72	617.68	629.40	VIS
21	634.96	11.72	629.10	640.83	VIS
22	646.40	11.73	640.53	652.26	VIS
23	657.84	11.74	651.97	663.71	VIS
24	669.29	11.75	663.42	675.17	VIS
25	680.75	11.75	674.88	686.63	VIS
26	692.22	11.76	686.34	698.10	VIS
27	703.69	11.76	697.81	709.58	VIS
28	715.17	11.77	709.29	721.06	VIS
29	726.66	11.77	720.77	732.54	VIS
30	738.14	11.77	732.26	744.03	VIS
31	749.63	11.78	743.75	755.52	VIS
32	761.13	11.78	755.24	767.02	NIR
33	772.62	11.78	766.73	778.51	NIR
34	784.11	11.78	778.22	790.00	NIR
35	795.61	11.78	789.72	801.50	NIR
36	807.10	11.78	801.21	812.99	NIR
37	818.59	11.77	812.70	824.48	NIR
38	830.08	11.77	824.19	835.97	NIR
39	841.56	11.77	835.68	847.45	NIR
40	853.04	11.76	847.16	858.93	NIR
41	864.52	11.76	858.64	870.40	NIR
42	875.99	11.75	870.11	881.86	NIR
43	887.45	11.75	881.57	893.32	NIR
44	898.90	11.74	893.03	904.77	NIR
45	910.35	11.73	904.48	916.21	NIR
46	921.78	11.73	915.92	927.65	NIR
47	933.21	11.72	927.35	939.07	NIR
48	944.62	11.71	938.77	950.48	NIR

APPENDIX B

PILOT STUDY – SUPPLEMENTARY DATA

Table B.1: Frequency histograms for the classes defined over the Mid Severn Estuary study area, calculated for bands 21, 33, 45 and 57. The normal curve is overlaid.

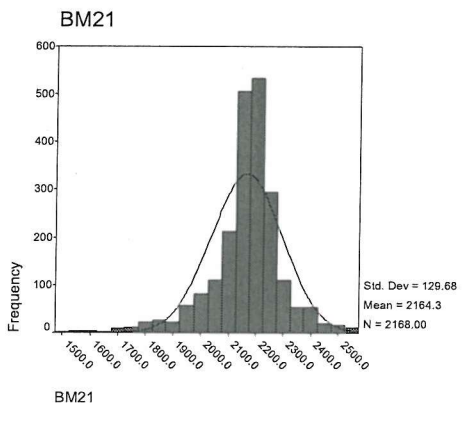
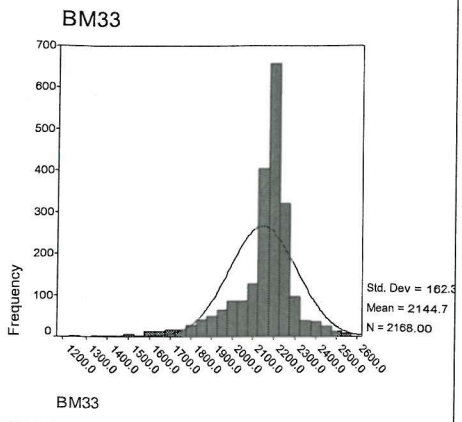
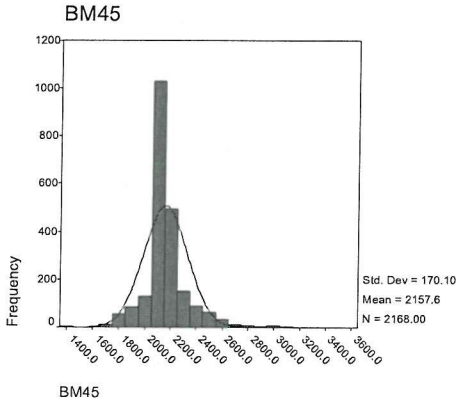
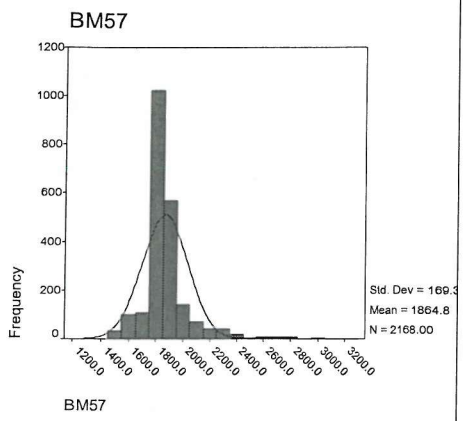
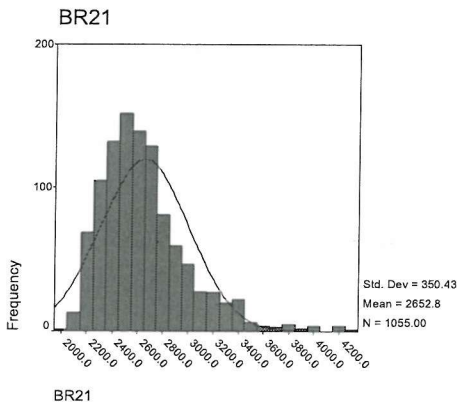
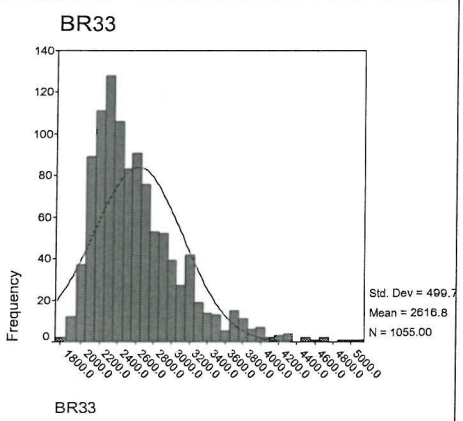
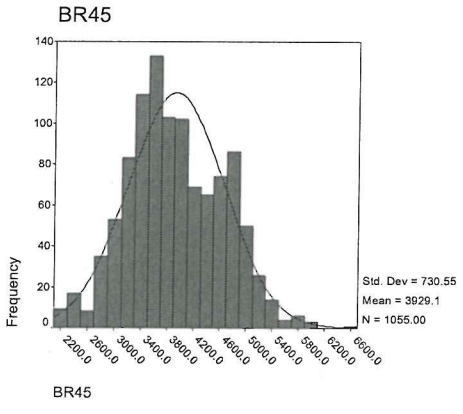
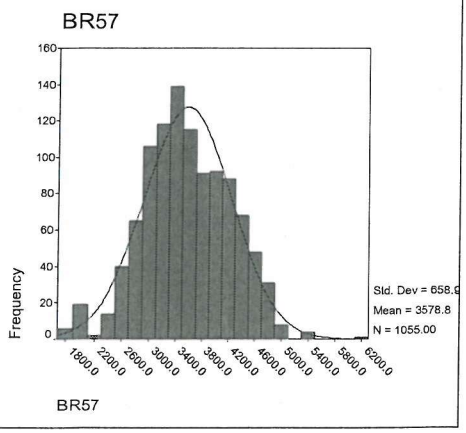
Classes	Band 21,45	Band 33, 57
Bare Mud		
		
Bare Rock		
		

Table B.1 continued.

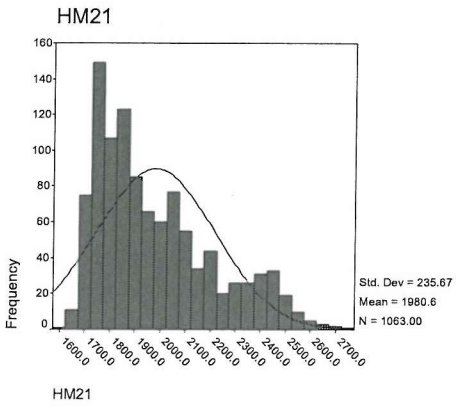
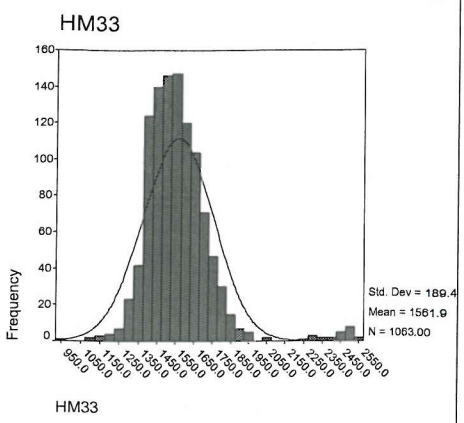
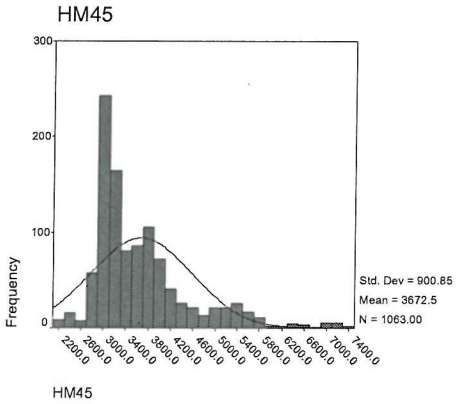
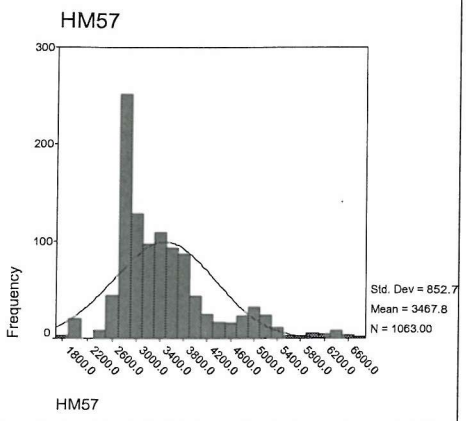
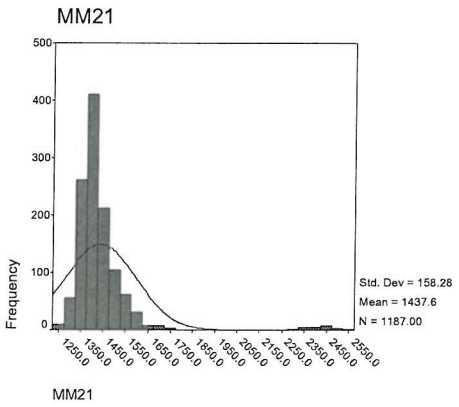
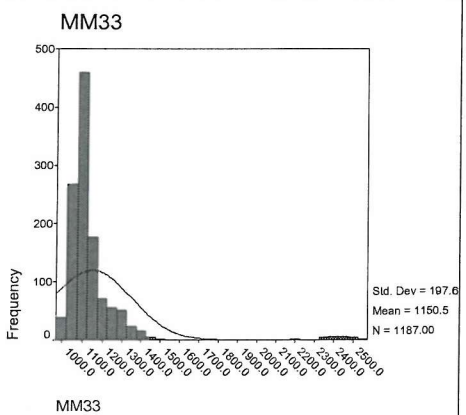
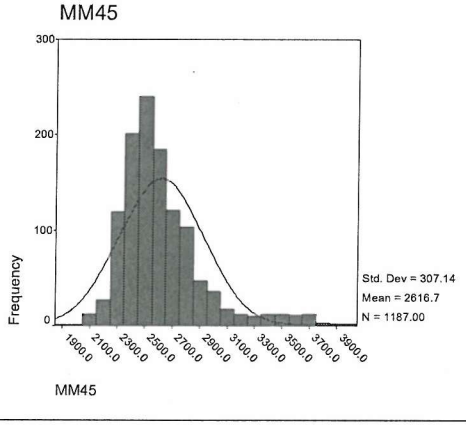
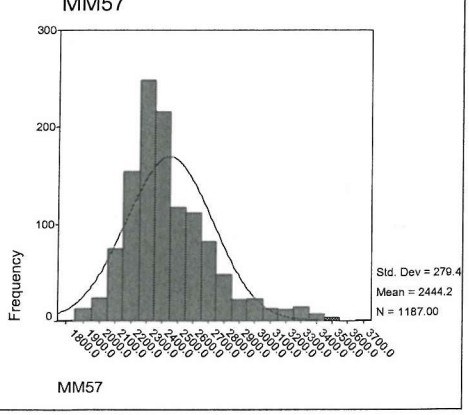
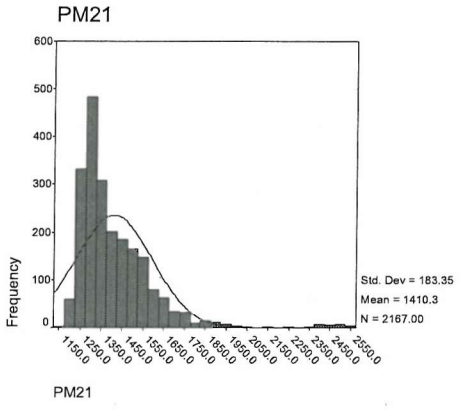
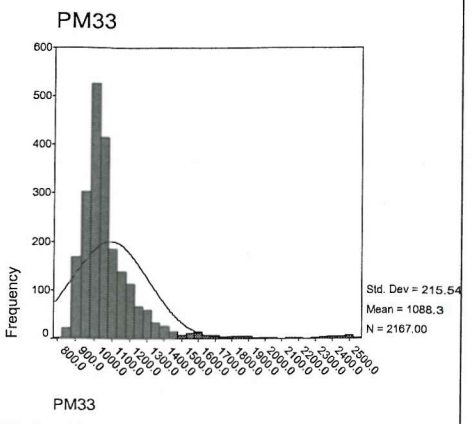
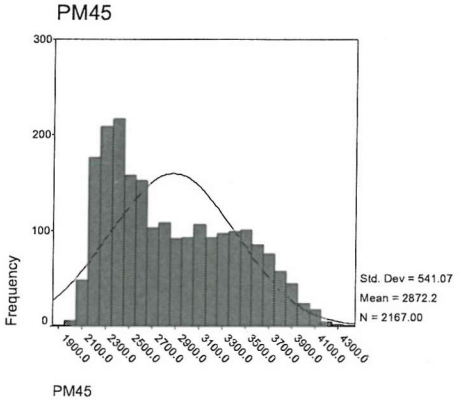
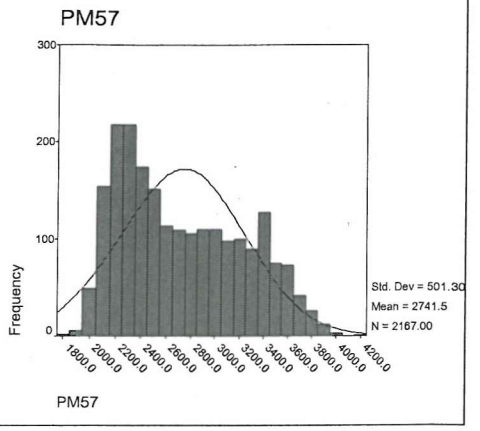
Classes	Band 21,45	Band 33, 57
High Marsh		
		
Mid Marsh		
		

Table B.1 continued.

Classes	Band 21,45	Band 33, 57
Pioneer Marsh		
		

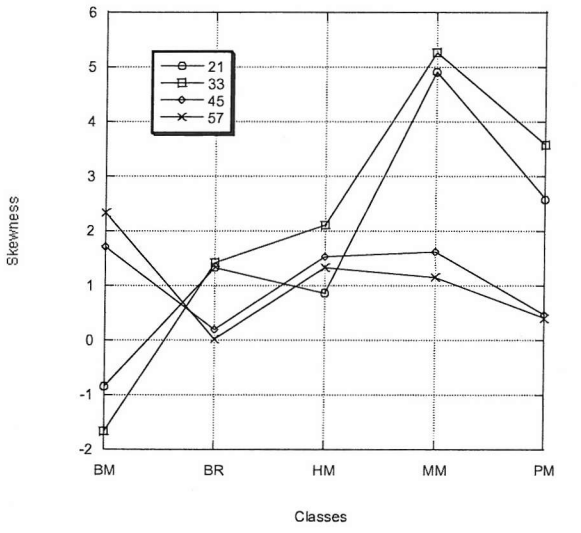


Figure B.1: Skewness calculated for all classes of the Mid Severn Estuary study area for bands 21, 33, 45 and 57.

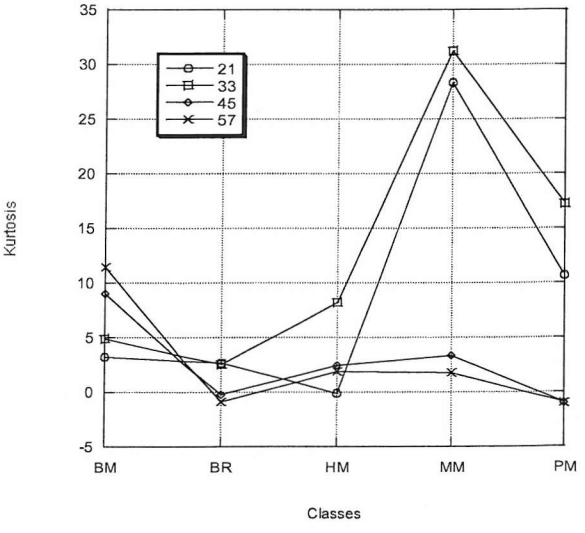


Figure B.2: Kurtosis calculated for all classes of the Mid Severn Estuary study area for bands 21, 33, 45 and 57.

Table B.2: CASI bands selected with the Projection Pursuit Feature Selection (PPFS) algorithm for the classes of the Mid Severn Estuary study area.

Feature set	Selected Bands										
1	26										
2	39	67									
3	33	40	55								
4	32	42	47	63							
5	32	42	47	48	63						
6	16	26	42	47	48	63					
7	16	27	39	42	47	48	63				
8	16	23	32	39	40	44	47	63			
9	16	27	39	40	42	44	47	55	63		
10	11	16	23	27	39	40	42	46	47	63	
11	16	26	24	36	39	40	42	44	47	48	63

Table B.3: Z-statistic for testing the significance between PCA features with respect to the overall MLC accuracy for the Mid Severn Estuary study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	50.29									
3	56.03	5.50								
4	55.59	5.07	**0.43**							
5	58.53	7.93	2.44	2.87						
6	58.99	8.39	2.90	3.33	**0.46**					
7	59.23	8.62	3.13	3.56	**0.69**	**0.23**				
8	59.46	8.85	3.37	3.80	**0.93**	**0.47**	**0.23**			
9	60.65	10.02	4.55	4.98	2.11	**1.65**	**1.42**	**1.19**		
10	61.37	10.74	5.28	5.71	2.84	2.38	2.15	**1.92**	**0.73**	
11	61.62	10.98	5.53	5.95	3.09	2.63	2.40	2.16	**0.98**	**0.25**
12	61.37	10.74	5.28	5.71	2.84	2.38	2.15	1.92	**0.73**	**0.00**

Table B.4: Z-statistic for testing the significance between MNF features with respect to the overall MLC accuracy for the Mid Severn Estuary study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	39.47									
3	58.27	17.15								
4	62.15	20.66	3.54							
5	65.35	23.54	6.48	2.94						
6	67.73	25.70	8.69	5.17	2.23					
7	68.46	26.37	9.38	5.86	2.92	**0.69**				
8	71.48	29.11	12.23	8.73	5.81	3.59	2.90			
9	72.26	29.82	12.97	9.48	6.56	4.35	3.66	**0.76**		
10	72.26	29.82	12.97	9.48	6.56	4.35	3.66	**0.76**	**0.00**	
11	72.52	30.06	13.22	9.73	6.82	4.60	3.91	**1.02**	**0.26**	**0.26**
12	72.26	29.82	12.97	9.48	6.56	4.35	3.66	**0.76**	**0.00**	**0.00**

Table B.5: Z-statistic for testing the significance between DAFE features with respect to the overall MLC accuracy for the Mid Severn Estuary study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3
2	21.65		
3	22.12	**0.50**	
4	24.31	2.81	2.32

Table B.6: Z-statistic for testing the significance between PPDA features with respect to the overall MLC accuracy for the Mid Severn Estuary study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3
2	25.82		
3	26.27	**0.48**	
4	28.60	2.95	2.47

Table B.7: Z-statistic for testing the significance between DBFE features with respect to the overall MLC accuracy for the Mid Severn Estuary study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	13.82									
3	32.47	18.69								
4	32.47	18.69	0.00							
5	35.15	21.42	2.83	2.83						
6	35.38	21.65	3.07	3.07	**0.25**					
7	39.20	25.56	7.22	7.22	4.41	4.17				
8	41.73	28.19	10.07	10.07	7.29	7.05	2.91			
9	41.99	28.46	10.37	10.37	7.59	7.35	3.21	**0.31**		
10	42.78	29.28	11.28	11.28	8.52	8.28	4.15	**1.25**	**0.94**	
11	41.99	28.46	10.37	10.37	7.59	7.35	3.21	**0.31**	**0.00**	**0.94**
12	41.47	27.92	9.77	9.77	6.99	6.75	2.60	**0.30**	**0.61**	**1.56**

Table B.8: Z-statistic for testing the significance between PPDB features with respect to the overall MLC accuracy for the Mid Severn Estuary study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	32.71									
3	37.27	4.74								
4	39.82	7.45	2.73							
5	40.29	7.97	3.25	**0.52**						
6	40.53	8.23	3.51	**0.78**	**0.26**					
7	41.01	8.75	4.04	**1.32**	**0.80**	**0.53**				
8	41.75	9.55	4.85	2.13	**1.61**	**1.35**	**0.81**			
9	42.49	10.37	5.68	2.96	2.44	2.18	**1.65**	**0.84**		
10	41.75	9.55	4.85	2.13	**1.61**	**1.35**	**0.81**	**0.00**	**0.84**	
11	41.26	9.02	4.31	**1.58**	**1.06**	**0.80**	**0.27**	**0.55**	**1.38**	**0.55**
12	41.01	8.75	4.04	**1.32**	**0.80**	**0.53**	**0.00**	**0.81**	**1.65**	**0.81**

Table B.9: Frequency histograms for the classes defined over the New Forest study area, calculated for bands 10, 40, 80 and 100. The normal curve is overlaid.

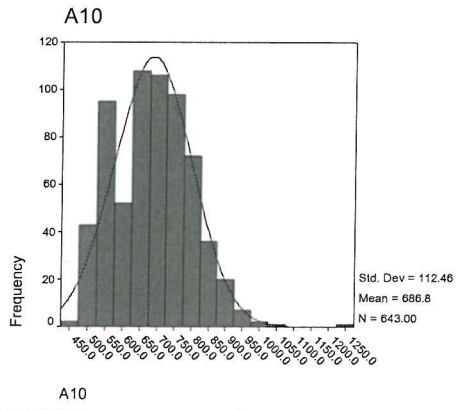
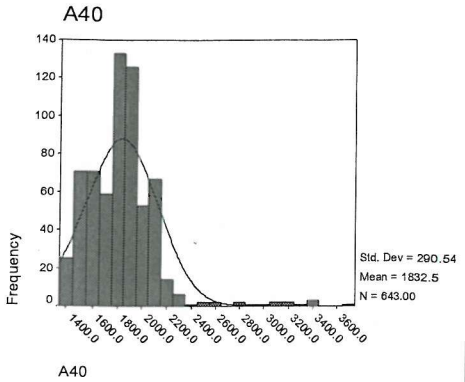
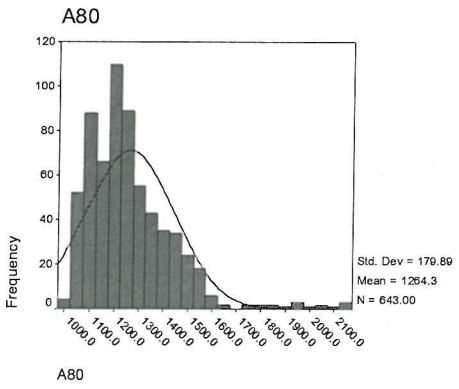
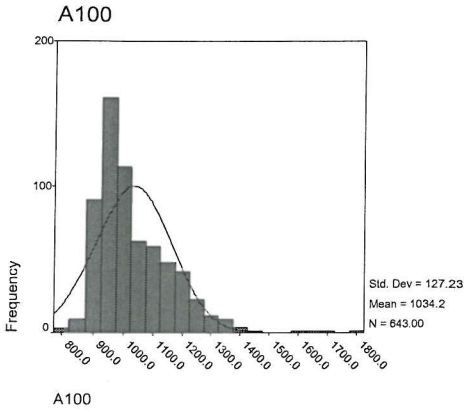
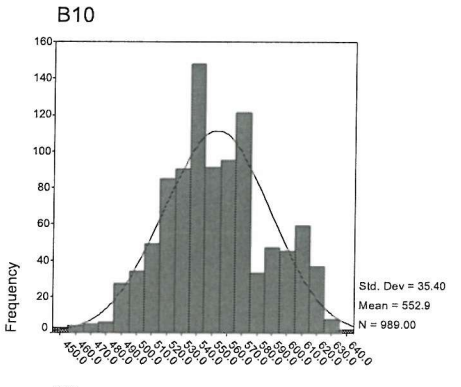
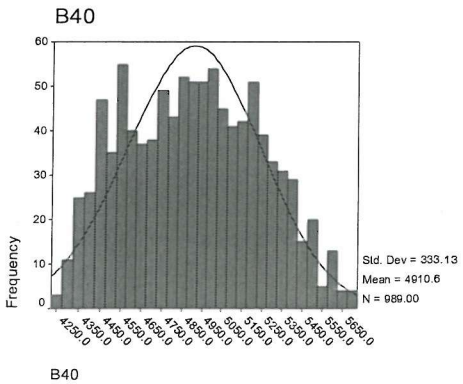
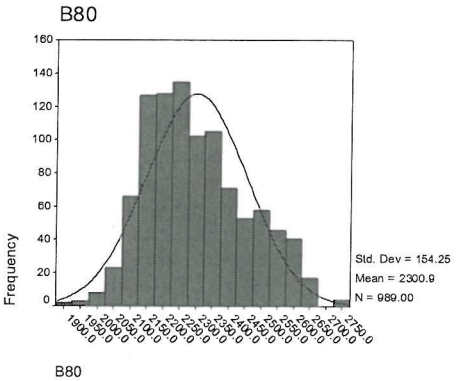
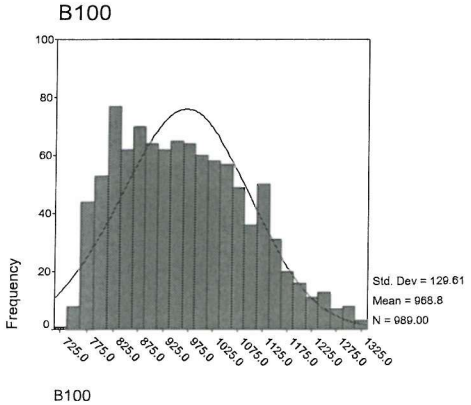
Classes	Band 10,80	Band 40, 100
Asphalt		
		
Bracken		
		

Table B.9 continued.

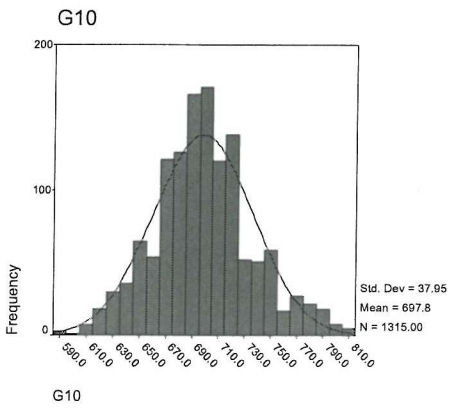
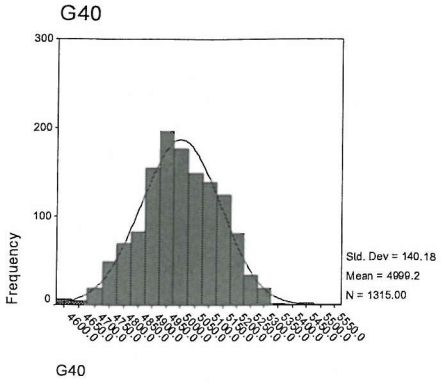
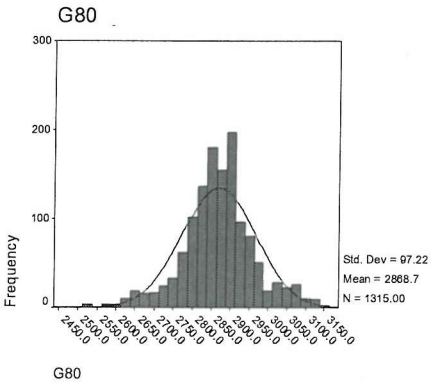
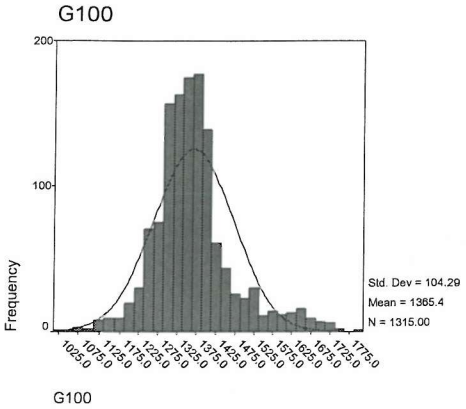
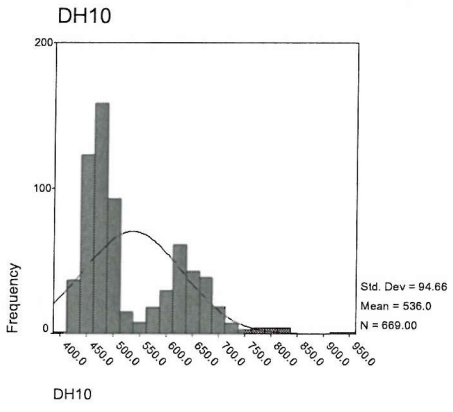
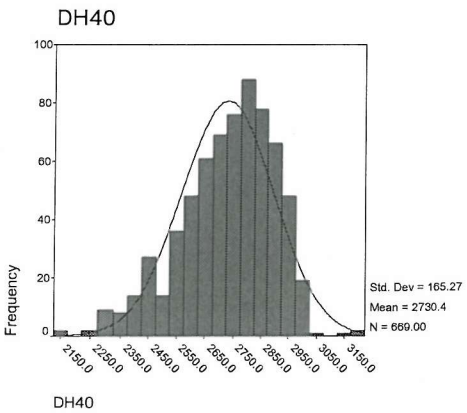
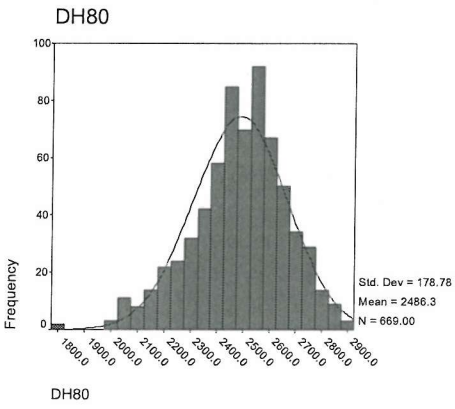
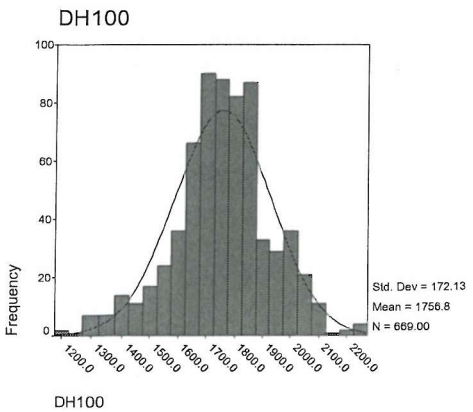
Classes	Band 10,80	Band 40, 100
Grassland		
		
Dry Heath		
		

Table B.9 continued.

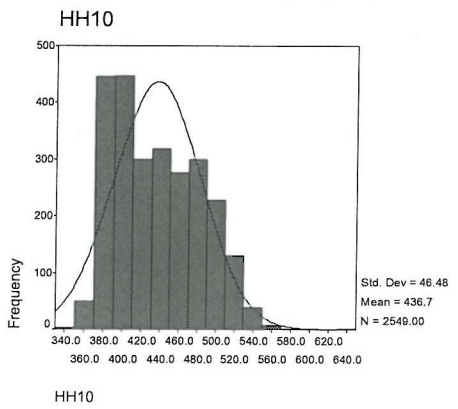
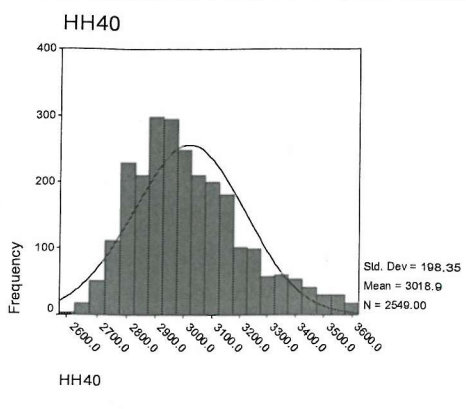
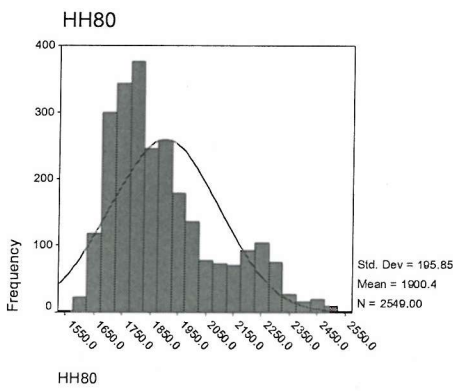
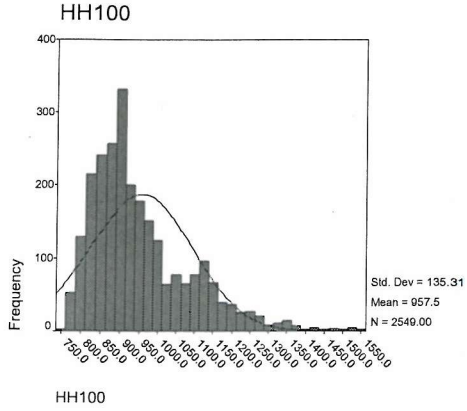
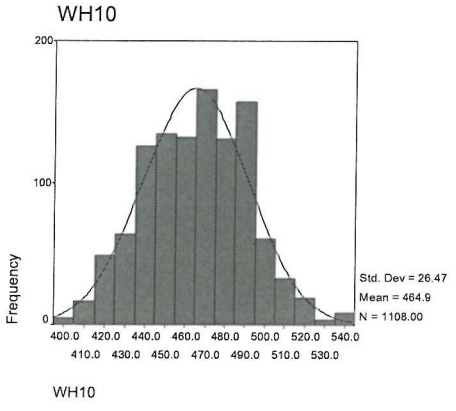
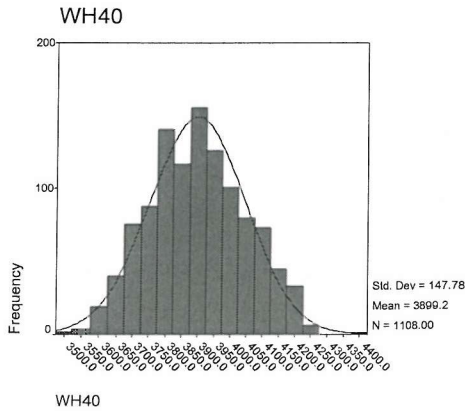
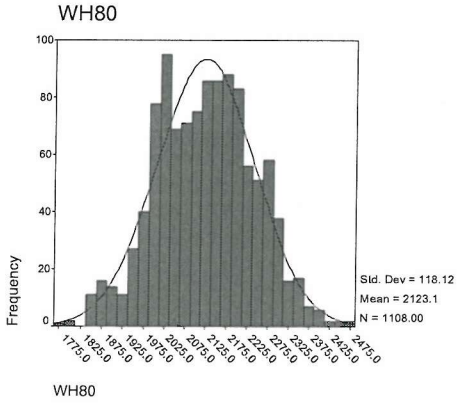
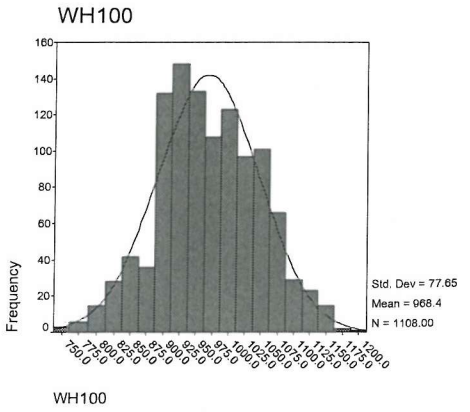
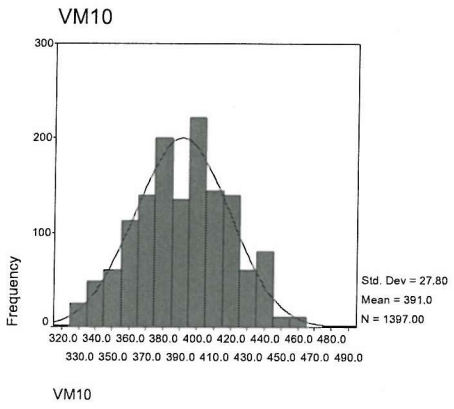
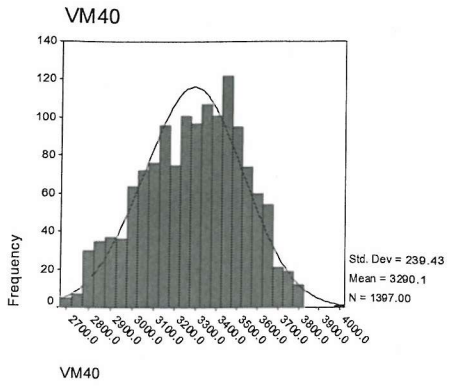
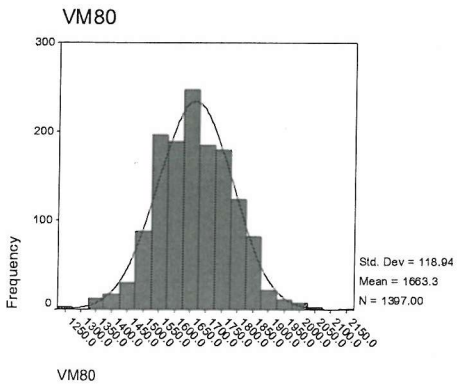
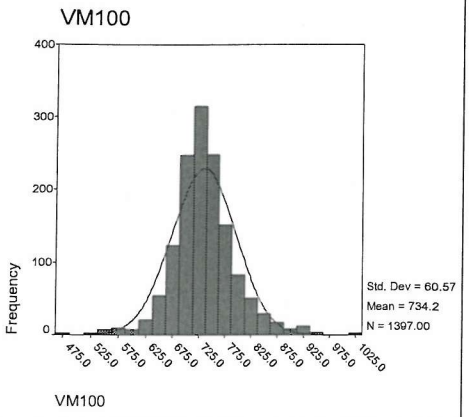
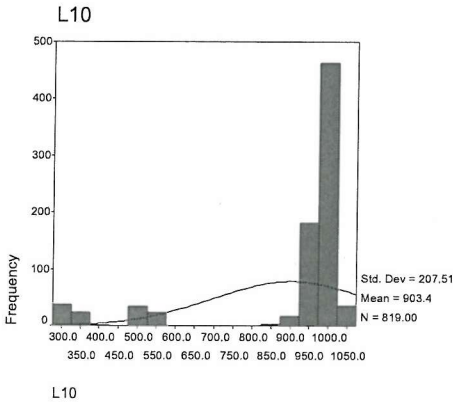
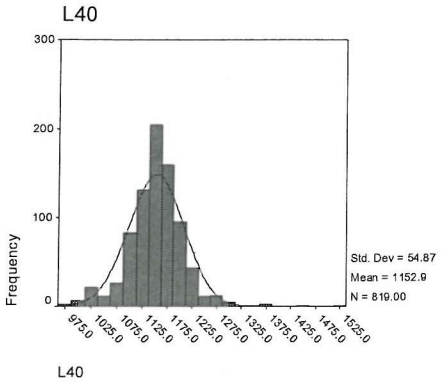
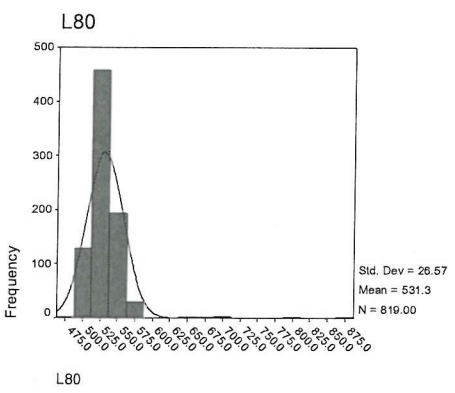
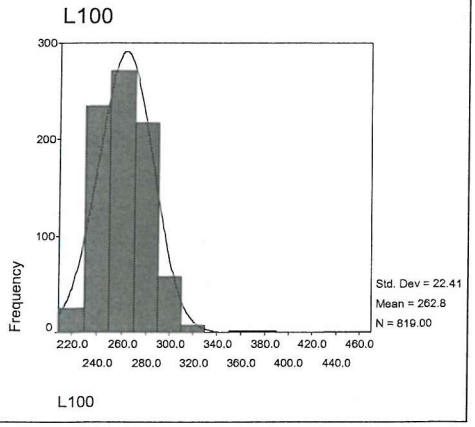
Classes	Band 10,80	Band 40, 100
Humid Heath		
		
Wet Heath		
		

Table B.9 continued.

Classes	Band 10,80	Band 40, 100
Valley Mire		
		
Lake		
		

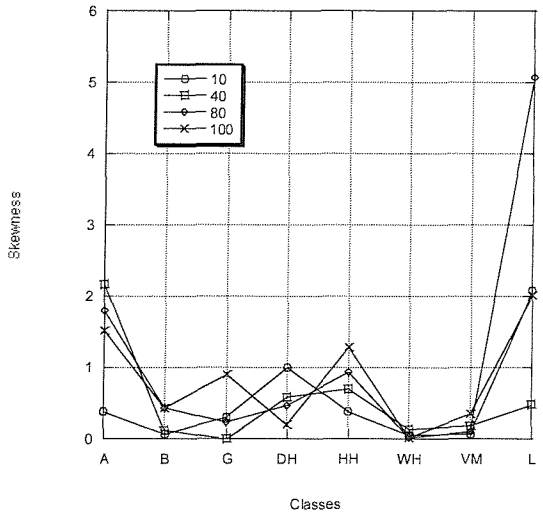


Figure B.3: Skewness calculated for all classes of the New Forest study area for bands 10, 40, 80 and 100.

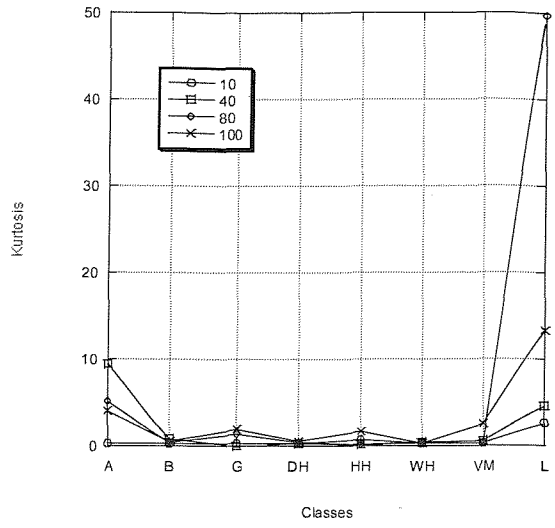


Figure B.4: Kurtosis calculated for all classes of the New Forest study area for bands 10, 40, 80 and 100.

Table B.10: HyMAP bands selected with the Projection Pursuit Feature Selection (PPFS) algorithm for the classes of the New Forest study area.

Feature set	Selected Bands									
1	87									
2	26	71								
3	18	33	62							
4	14	19	46	61						
5	14	20	28	46	61					
6	11	21	26	46	59	87				
7	9	14	20	28	46	59	87			
8	11	20	28	38	42	59	87	100		
9	9	15	21	24	42	50	59	87	100	
10	9	15	19	21	28	42	56	75	87	100

Table B.11: Z-statistic for testing the significance between PCA features with respect to the overall MLC accuracy for the New Forest study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	55.97									
3	55.97	**0.00**								
4	55.64	**0.54**	**0.54**							
5	54.99	**1.57**	**1.57**	**1.03**						
6	56.30	**0.56**	**0.56**	**1.09**	2.12					
7	55.97	**0.00**	**0.00**	**0.54**	**1.57**	**0.56**				
8	56.30	**0.56**	**0.56**	**1.09**	2.12	**0.00**	**0.56**			
9	56.30	**0.56**	**0.56**	**1.09**	2.12	**0.00**	**0.56**	**0.00**		
10	55.97	**0.00**	**0.00**	**0.54**	**1.57**	**0.56**	**0.00**	**0.56**	**0.56**	
11	56.64	**1.13**	**1.13**	**1.67**	2.69	**0.57**	**1.13**	**0.57**	**0.57**	**1.13**
12	56.64	**1.13**	**1.13**	**1.67**	2.69	**0.57**	**1.13**	**0.57**	**0.57**	**1.13**

Table B.12: Z-statistic for testing the significance between MNF features with respect to the overall MLC accuracy for the New Forest study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	54.49									
3	57.43	4.71								
4	56.76	3.56	**1.17**							
5	56.10	2.48	2.26	**1.09**						
6	56.76	3.56	**1.17**	**0.00**	**1.09**					
7	56.76	3.56	**1.17**	**0.00**	**1.09**	**0.00**				
8	56.43	3.01	**1.72**	**0.56**	**0.54**	**0.56**	**0.56**			
9	56.43	3.01	**1.72**	**0.56**	**0.54**	**0.56**	**0.56**	**0.00**		
10	57.09	4.13	**0.59**	**0.57**	**1.67**	**0.57**	**0.57**	**1.13**	**1.13**	
11	57.43	4.71	**0.00**	**1.17**	2.26	**1.17**	**1.17**	**1.72**	**1.72**	**0.59**
12	57.76	5.31	**0.62**	**1.78**	2.87	**1.78**	**1.78**	2.33	2.33	**1.21**

Table B.13: Z-statistic for testing the significance between DAFE features with respect to the overall MLC accuracy for the New Forest study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6
2	32.98					
3	47.77	15.47				
4	51.16	19.25	4.01			
5	53.84	22.32	7.40	3.43		
6	56.66	25.66	11.23	7.36	3.99	
7	57.63	26.83	12.63	8.82	5.49	**1.53**

Table B.14: Z-statistic for testing the significance between PPDA features with respect to the overall MLC accuracy for the New Forest study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6
2	40.04					
3	40.81	**0.83**				
4	45.64	6.24	5.42			
5	48.82	10.00	9.19	3.83		
6	53.46	15.91	15.13	10.01	6.30	
7	54.11	16.78	16.02	10.94	7.27	**1.00**

Table B.15: Z-statistic for testing the significance between DBFE features with respect to the overall MLC accuracy for the New Forest study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9	10
2	22.68									
3	46.36	23.06								
4	63.75	40.14	17.84							
5	68.19	44.59	22.78	5.42						
6	70.06	46.47	24.94	7.90	2.54					
7	71.99	48.43	27.21	10.63	5.38	2.87				
8	72.32	48.76	27.60	11.11	5.89	3.39	**0.52**			
9	73.31	49.78	28.81	12.61	7.49	5.03	2.19	**1.67**		
10	73.31	49.78	28.81	12.61	7.49	5.03	2.19	**1.67**	**0.00**	
11	73.65	50.12	29.21	13.13	8.05	5.60	2.78	2.26	**0.59**	**0.59**
12	73.65	50.12	29.21	13.13	8.05	5.60	2.78	2.26	**0.59**	**0.59**

Table B.16: Z-statistic for testing the significance between PPDB features with respect to the overall MLC accuracy for the New Forest study area. Values below the critical 5% significance level (1.96) are printed in bold.

Feature	1	2	3	4	5	6	7	8	9
2	29.79								
3	52.55	22.66							
4	64.83	35.59	14.54						
5	65.47	36.29	15.39	**0.96**					
6	66.77	37.71	17.18	3.01	2.06				
7	67.43	38.44	18.11	4.13	3.18	**1.13**			
8	68.09	39.18	19.07	5.31	4.37	2.33	**1.21**		
9	67.43	38.44	18.11	4.13	3.18	**1.13**	**0.00**	**1.21**	
10	67.76	38.81	18.59	4.71	3.77	**1.72**	**0.59**	**0.62**	**0.59**

APPENDIX C

SUPERVISED BAND SELECTION – SUPPLEMENTARY DATA

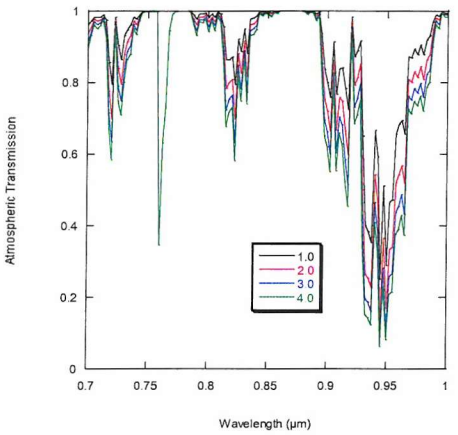


Figure C.1: Atmospheric transmission for 0.7 to 1.0 μm wavelength range modelled with 6S for different water vapour contents (g/cm^2) in a US 1962 standard atmosphere.

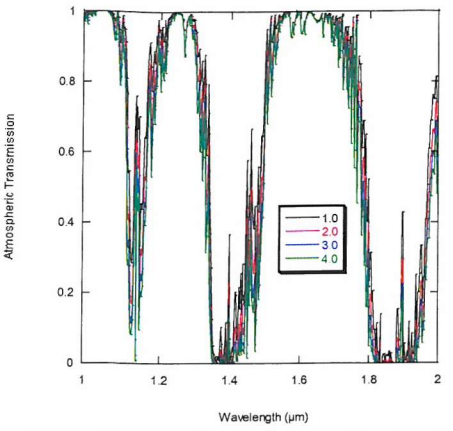


Figure C.2: Atmospheric transmission for 1.0 to 2.0 μm wavelength range modelled with 6S for different water vapour contents (g/cm^2) in a US 1962 standard atmosphere.

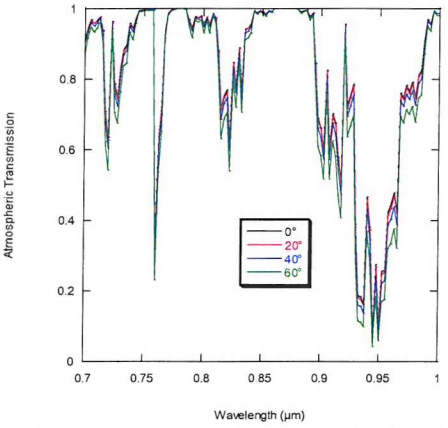


Figure C.3: Atmospheric transmission for 0.7 to 1.0 μm wavelength range modelled with 6S for different solar zenith angles (°) in a midlatitude summer atmosphere.

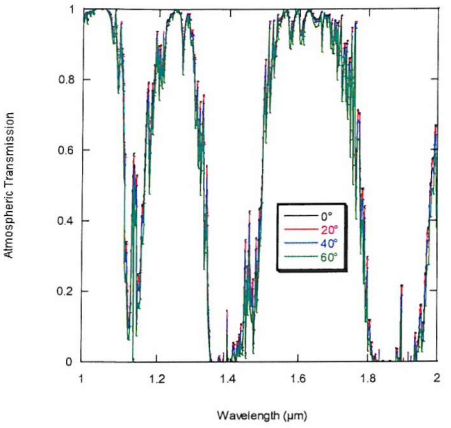


Figure C.4: Atmospheric transmission for 1.0 to 2.0 μm wavelength range modelled with 6S for different solar zenith angles (°) in a midlatitude summer atmosphere.

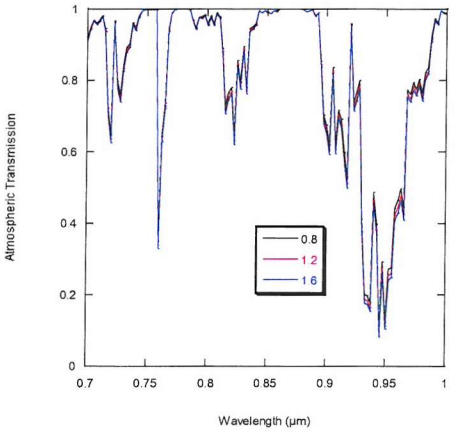


Figure C.5: Atmospheric transmission for 0.7 to 1.0 μm wavelength range modelled with 6S for different aircraft heights (km) in a midlatitude summer atmosphere.

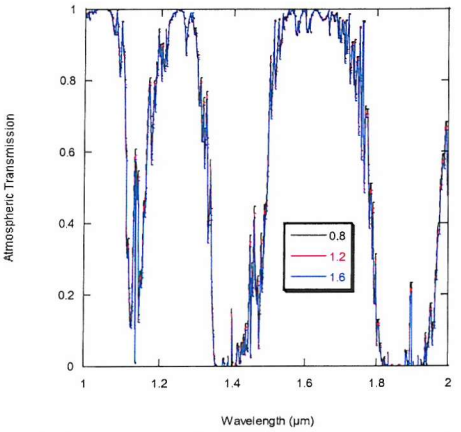
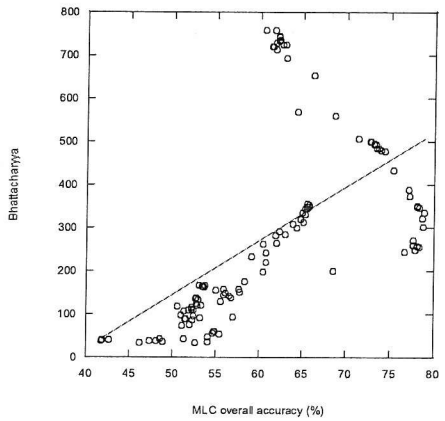
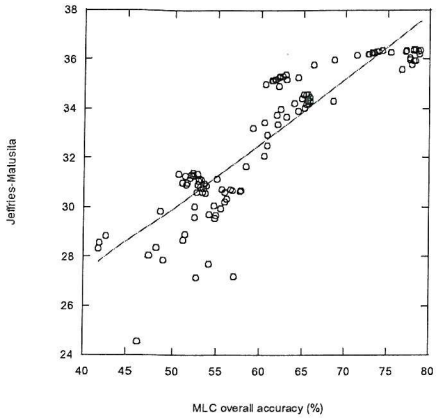


Figure C.6: Atmospheric transmission for 1.0 to 2.0 μm wavelength range modelled with 6S for different aircraft heights (km) in a midlatitude summer atmosphere.

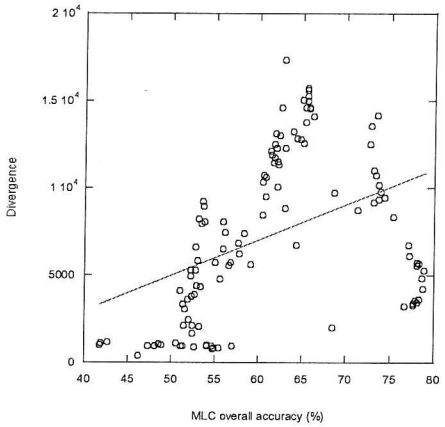
Table C.1: Scatter plots of single-band distance measures against MLC overall performance estimated with the leave-one-out method for the New Forest and River Severn data set. The regression line is displayed and the correlation coefficient r given.



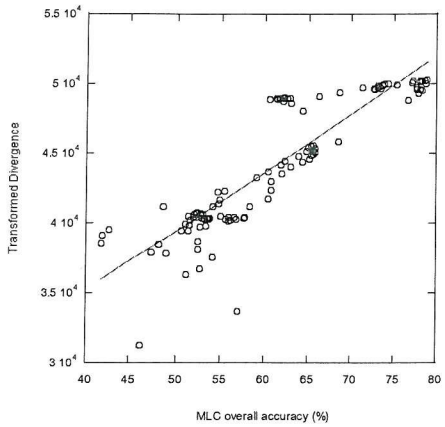
Bhattacharyya – New Forest – r=0.56178



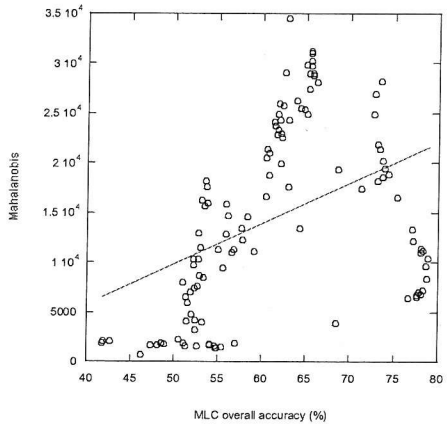
Jeffries-Matusita - New Forest – r=0.88941



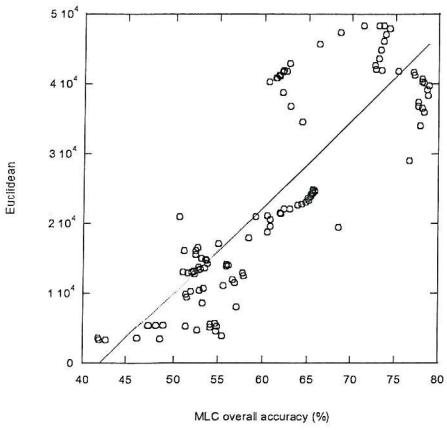
Divergence – New Forest – r=0.41462



Transformed Divergence – New Forest – r=0.8765

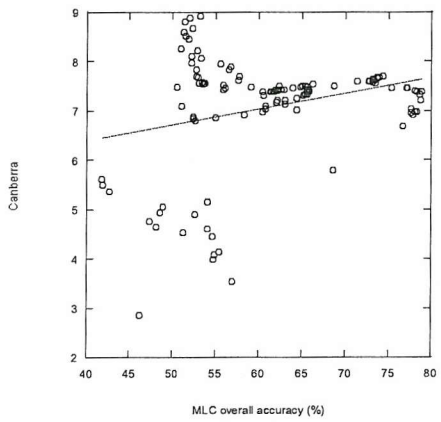


Mahalanobis Distance – New Forest – r=0.41835

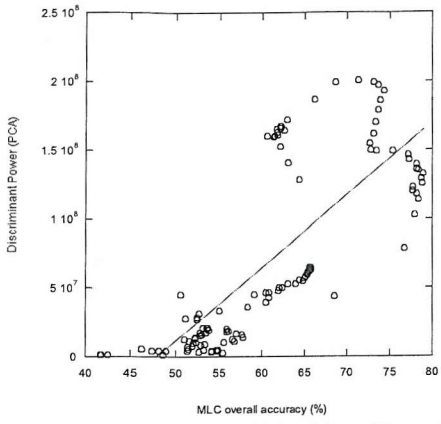


Euclidean Distance - New Forest – r=0.83178

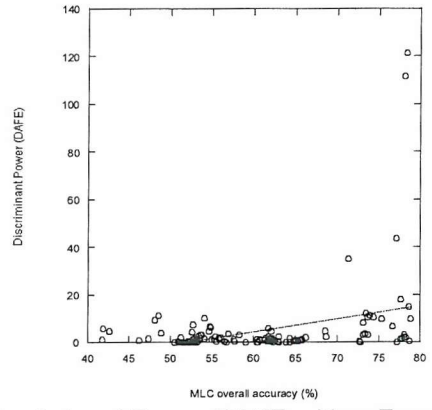
Table C.1 continued.



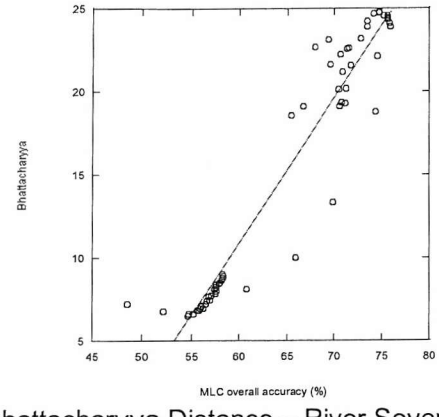
Canberra Distance – New Forest – r=0.27153



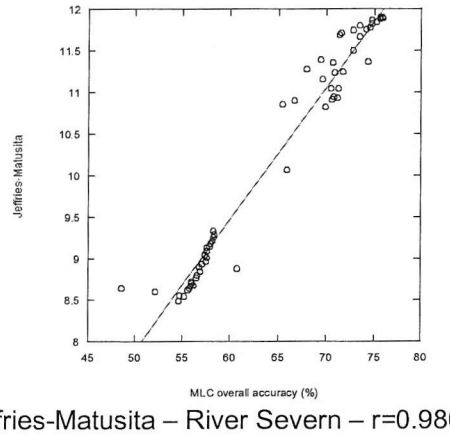
Discriminant Power PCA - New Forest – r=0.76679



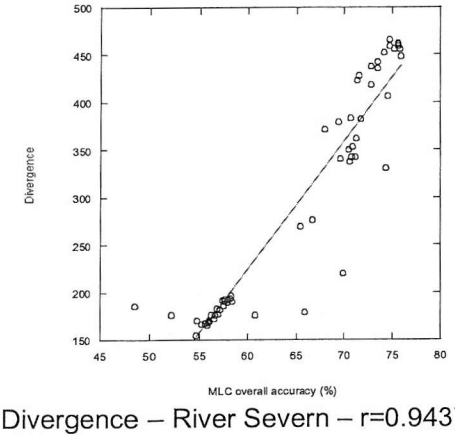
Discriminant Power DAFE – New Forest – r=0.331



Bhattacharyya Distance – River Severn – r=0.96114

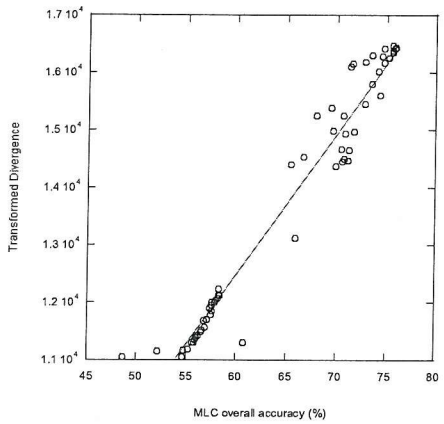


Jeffries-Matusita – River Severn – r=0.98079

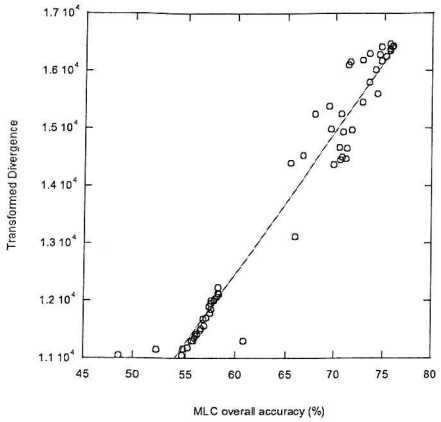


Divergence – River Severn – r=0.94373

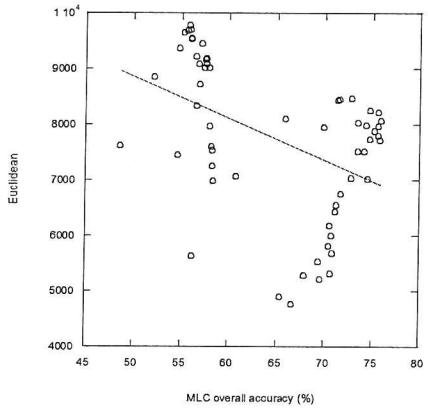
Table C.1 continued.



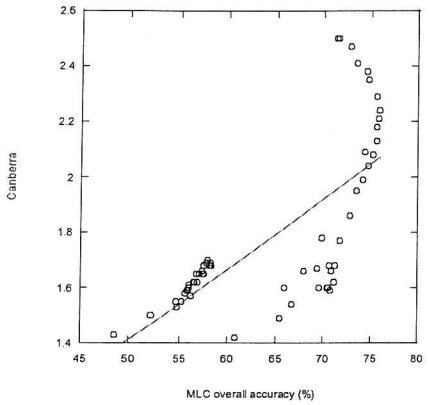
Transformed Divergence – River Severn – r=0.97688



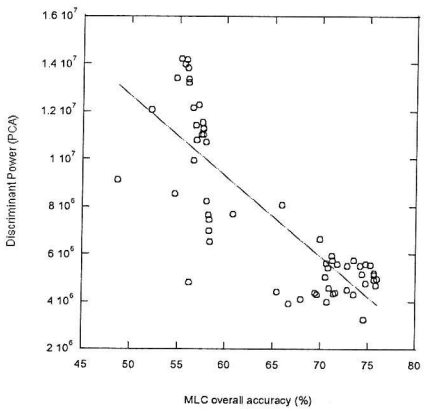
Mahalanobis – River Severn – r=0.95083



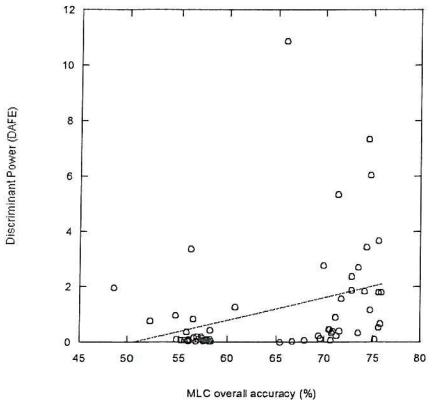
Euclidean Distance – River Severn – r=0.4506



Canberra Distance – River Severn – r=0.70197

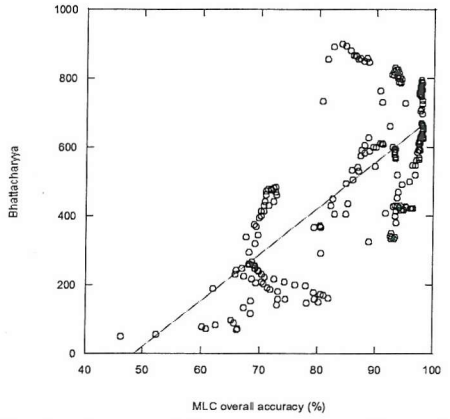


Discriminant Power PCA – River Severn – r=0.8214

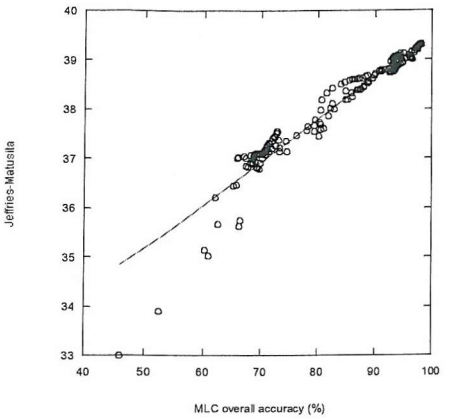


Discriminant Power DAFE – River Severn – r=0.34131

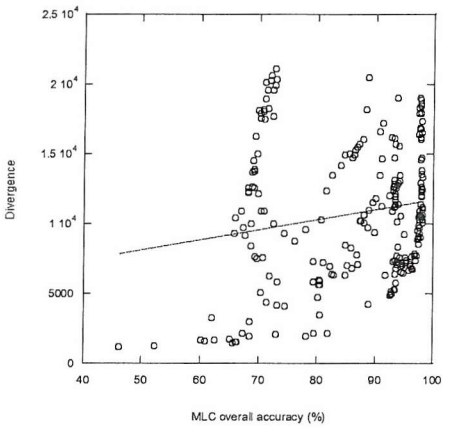
Table C.2: Scatter plots of multiple-band distance measures against MLC overall performance estimated with the leave-one-out method for the New Forest and River Severn data set. The regression line is displayed and the correlation coefficient r given.



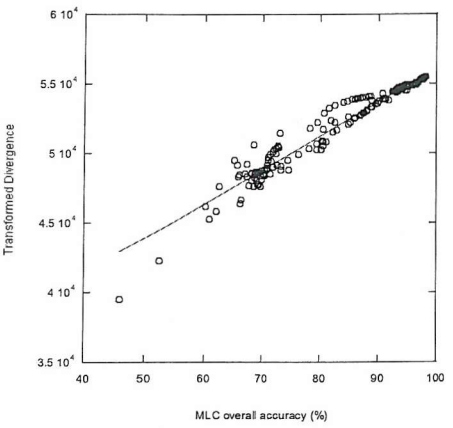
Bhattacharyya Distance – New Forest – r=0.7113



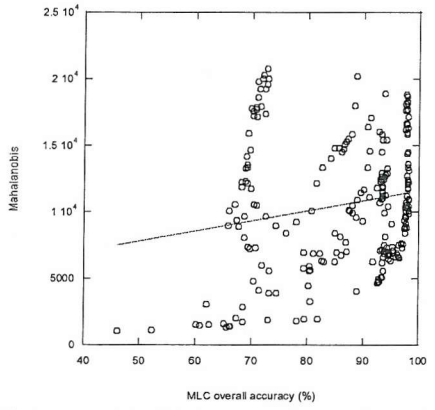
Jeffries Matusita - New Forest– r=0.97083



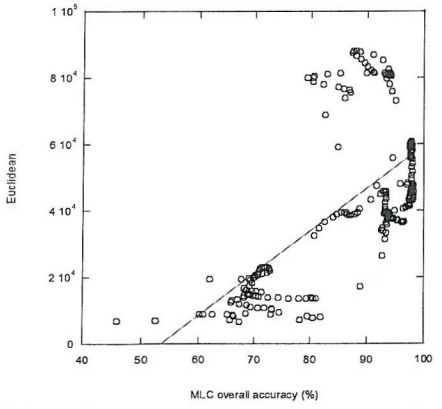
Divergence – New Forest – r=0.17501



Transformed Divergence – New Forest – r=0.97815

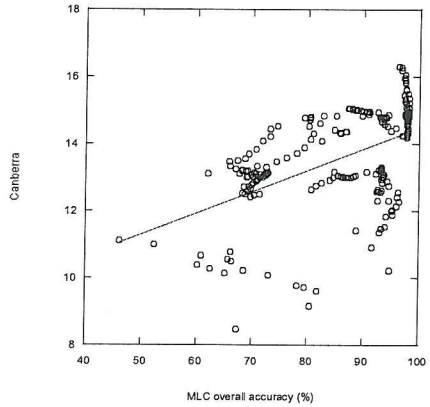


Mahalanobis Distance – New Forest – r=0.18344

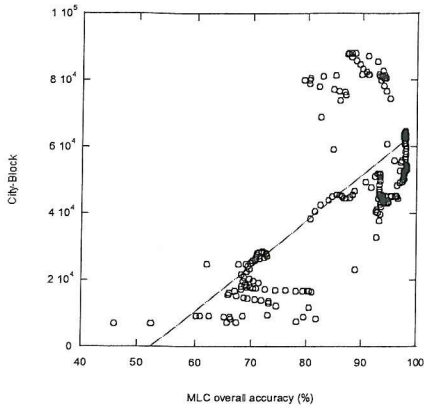


Euclidean Distance - New Forest – r=0.6394

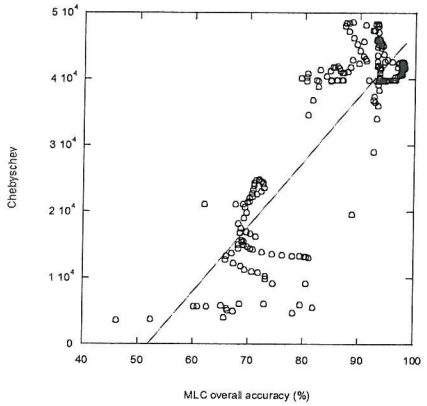
Table C.2 continued.



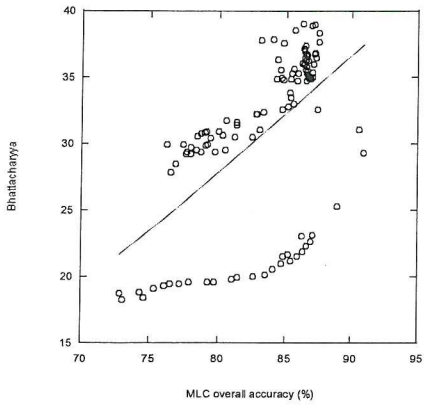
Canberra Distance – New Forest – r=0.52348



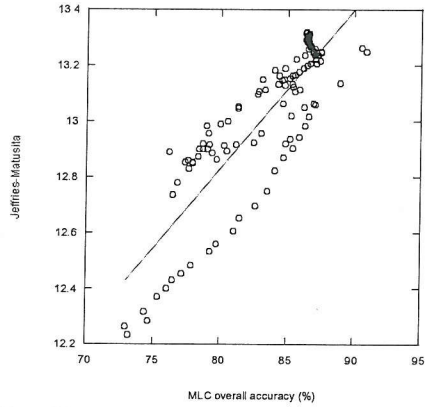
City-Block - New Forest – r=0.70412



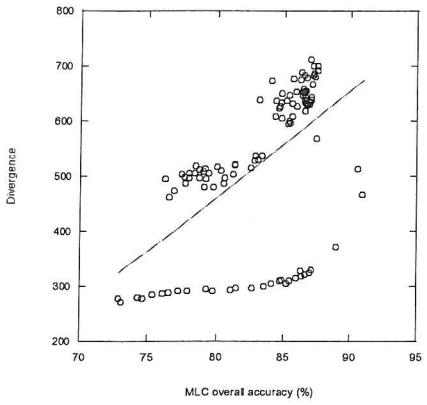
Chebyschev – New Forest – r=0.86719



Bhattacharyya Distance – River Severn – r=0.56617

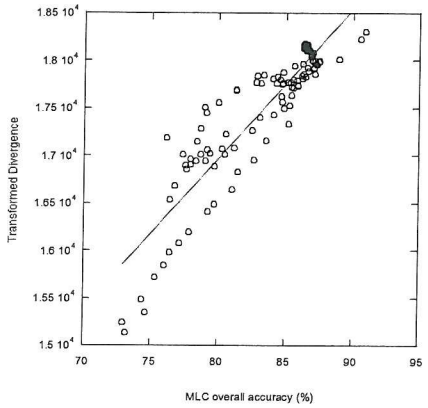


Jeffries-Matusita – River Severn – r=0.85805

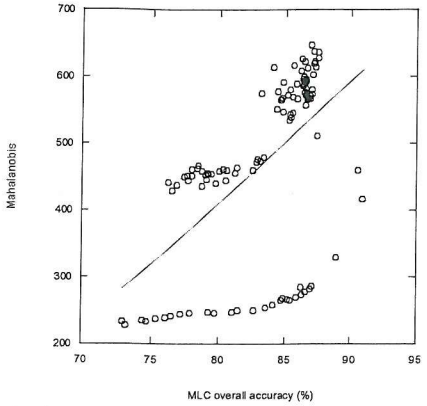


Divergence – River Severn – r=0.56269

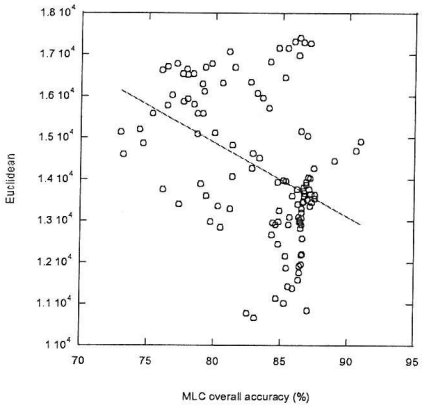
Table C.2 continued.



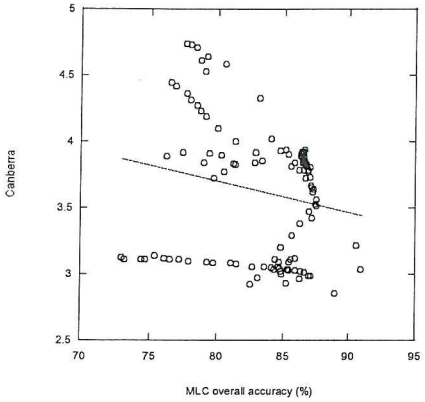
Transformed Divergence – River Severn – r=0.90708



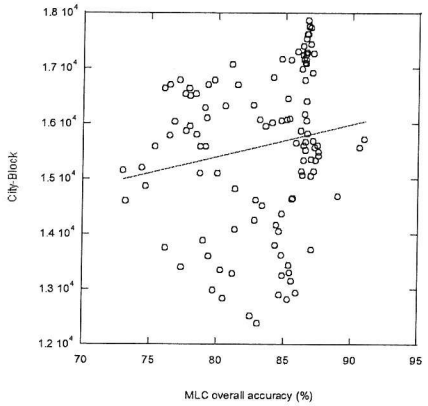
Mahalanobis – River Severn – r=0.55397



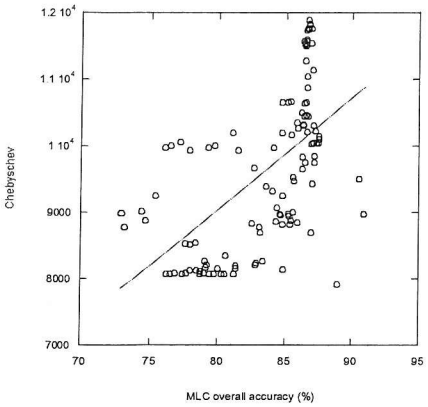
Euclidean Distance – River Severn – r=0.41751



Canberra Distance – River Severn – r=0.19266



City-Block – River Severn – r=0.16787



Chebyschev – River Severn – r=0.57154

Table C.3: Band sets from current airborne and satellite sensors and their simulation with HyMAP (New Forest) and CASI (River Severn) bands sets available (data sets introduced in chapter 3). Bands excluded or not available are marked by an X.

Sensor	Band	Centre [nm]	Width [nm]	HyMAP Centre [nm]	HyMAP Width [nm]	CASI Centre [nm]	CASI Width [nm]
CHRIS (mode 3)	1	442.4	8.8	X	X	X	X
	2	490.2	9.2	492.3	15.5	489.6	16
	3	529.9	8.6	523.4	15.6	530.5	8.6
	4	551.25	9.7	554.2	15.3	553	8.6
	5	569.85	7.1	569.5	15.4	568.1	8.6
	6	631.3	9.4	631.3	15.6	632.3	16.2
	7	661.05	10.5	661.7	15.1	658.9	8.8
	8	671.75	10.9	X	X	X	X
	9	697.2	5.8	692.4	15.8	697.1	8.8
	10	703.2	6	X	X	704.7	8.8
	11	709.3	6.2	707.8	15.3	712.3	8.8
	12	741.6	6.8	738.1	15.7	742.9	8.8
	13	748.45	6.9	753.4	15.3	750.5	8.8
	14	780.85	15.1	783.5	15.3	777.3	16.4
	15	872.2	18.2	X	X	X	X
	16	895.45	9.5	890.3	20	895.7	8.8
	17	905	9.6	X	X	X	X
	18	1018.5	33	X	X	X	X
ETM+	1	485	70	484.35	61.9	497	45.8
	2	570	80	569.45	77	568.15	83.9
	3	660	60	654.1	61.15	662.75	62
	4	840	120	836.9	124.45	838.4	123.4
	5	1650	200	1652.05	204.65	X	X
	7	2220	260	2221.8	261.05	X	X
MERIS	1	412.5	10	X	X	X	X
	2	442.5	10	X	X	X	X
	3	490	10	492.3	15.5	489.6	16
	4	510	10	507.4	15.8	508.1	8.6
	5	560	10	554.2	15.3	560.5	8.6
	6	620	10	615.7	15.7	621	8.6
	7	665	10	661.7	15.1	666.6	8.8
	8	681.25	7.5	676.9	15.5	681.8	8.8
	9	705	10	707.8	15.3	704.7	8.8
	10	753.75	7.5	753.4	15.3	754.35	16.5
	11	760	2.5	X	X	X	X
	12	765	5	X	X	X	X
	13	775	12.5	776	30.3	777.3	16.4
	14	865	10	872.6	18.1	865.2	8.8
	15	890	10	890.3	20	888.1	8.8
	16	900	10	X	X	X	X
MISR	1	446.4	41.9	461.3	15.8	X	X
	2	557.5	28.6	561.85	30.65	556.8	31.2
	3	671.7	21.9	669.3	30.5	674.2	24
	4	866.4	39.7	867.4	31.25	865.25	39.3
MODIS	1	645	50	646.5	45.75	643.8	54.3
	2	858.5	35	859.7	46.45	857.6	39.4
	3	469	20	469.3	31.65	478.4	8.6
	4	555	20	554.2	15.3	553	23.6
	5	1240	20	1245.95	32.3	X	X
	6	1640	24	1641.45	28.8	X	X
	7	2130	50	2136.7	56.3	X	X

Table C.3 continued.

NERC VEG	1	450	20	461.3	15.8	X	X
	2	490	20	492.3	15.5	489.6	16
	3	552	10	554.2	15.3	553	8.6
	4	608	10	607.9	31.2	605.8	8.6
	5	647	10	646.6	15.3	643.7	8.8
	6	670	10	661.7	15.1	670.4	16.4
	7	700	10	692.4	15.8	700.9	16.4
	8	710	10	707.8	15.3	712.3	8.8
	9	740	10	738.1	15.7	742.9	8.8
	10	750	7	753.4	15.3	750.5	8.8
	11	762	5	X	X	X	X
	12	780	10	783.5	15.3	781.1	8.8
	13	820	10	X	X	X	X
	14	865	10	859.8	16.2	865.2	8.8
	15	942	10	X	X	X	X
EA VEG	1	445.9	13.2	X	X	X	X
	2	469.9	13.4	461.3	15.8	X	X
	3	490.3	13.4	492.3	15.5	489.6	16
	4	550.1	13.4	554.2	15.3	549.25	16.1
	5	670.9	13.6	661.7	15.1	670.4	16.4
	6	683.3	8	676.9	15.5	681.8	8.8
	7	700.4	11.8	692.4	15.8	700.9	16.4
	8	710.9	9.8	707.8	15.3	712.3	8.8
	9	721.4	11.8	722.9	15.3	720	8.8
	10	751	13.6	753.4	15.3	750.5	8.8
	11	763.4	8	X	X	X	X
	12	780.6	11.8	783.5	15.3	781.1	8.8
	13	860.1	13.8	859.8	16.2	861.4	16.4
	14	880.2	23.2	881.45	36.75	880.5	24
EA COAST	1	443.1	22.4	X	X	X	X
	2	489.4	22.6	X	X	485.85	23.5
	3	510.8	20.8	X	X	508.15	23.5
	4	554.8	22.8	X	X	553	23.6
	5	599	13.6	X	X	602	16.2
	6	625.4	13.6	X	X	624.75	16.1
	7	662.4	7.8	X	X	658.9	8.8
	8	672.8	13.6	X	X	670.4	16.4
	9	683.3	8	X	X	681.8	8.8
	10	691.8	9.8	X	X	689.4	8.8
	11	702.3	8	X	X	704.7	8.8
	12	711.8	11.8	X	X	712.3	8.8
	13	751	13.6	X	X	750.5	8.8
	14	857.2	27.2	X	X	857.6	24
	15	881.2	21.4	X	X	880.5	24
BIOTA	1	442.5	15	X	X	X	X
	2	490	10	X	X	493.3	8.6
	3	540	10	X	X	538	8.6
	4	552	10	X	X	553	8.6
	5	608	10	X	X	605.8	8.6
	6	652	10	X	X	651.3	8.8
	7	670	10	X	X	666.6	8.8
	8	682.5	5	X	X	681.8	8.8
	9	710	10	X	X	712.3	8.8
	10	749.5	5	X	X	750.5	8.8
	11	761.5	5	X	X	X	X
	12	780	10	X	X	781.1	8.8
	13	820	10	X	X	819.4	8.8
	14	880	20	X	X	880.5	24

Table C.4: Randomly and uniformly spaced band sets for the HyMAP (New Forest) and CASI (River Severn) data set introduced in chapter 4. The band number refers to the index of available bands (117 and 60 for HyMAP and CASI, respectively), not to the original detector number shown in table A.1.

Set Dimension	HyMAP sensor bands	CASI sensor bands
RANDOMLY SPACED		
15	106, 77, 71, 52, 14, 92, 32, 51, 94, 22, 37, 96, 56, 34, 12	36, 40, 10, 45, 9, 14, 21, 57, 39, 7, 48, 58, 16, 53, 4
UNIFORMLY SPACED		
15	3, 11, 19, 27, 35, 43, 51, 59, 67, 75, 83, 91, 99, 107, 115	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58
14	7, 15, 23, 31, 39, 47, 55, 63, 71, 79, 87, 95, 103, 111	4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56
13	5, 14, 23, 32, 41, 50, 59, 68, 77, 86, 95, 104, 113	6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54
12	4, 14, 24, 34, 44, 54, 64, 74, 84, 94, 104, 114	3, 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58
11	4, 15, 26, 37, 48, 59, 70, 81, 92, 103, 114	5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55
10	7, 19, 31, 43, 55, 67, 79, 91, 103, 115	3, 9, 15, 21, 27, 33, 39, 45, 51, 57
9	3, 17, 31, 45, 59, 73, 87, 101, 115	2, 9, 16, 23, 30, 37, 44, 51, 58
8	3, 19, 35, 51, 67, 83, 99, 115	2, 10, 18, 26, 34, 42, 50, 58
7	2, 21, 40, 59, 78, 97, 116	3, 12, 21, 30, 39, 48, 57
6	1, 24, 47, 70, 93, 116	3, 14, 25, 36, 47, 58
5	3, 31, 59, 87, 115	2, 16, 30, 44, 58
4	2, 40, 78, 116	2, 21, 40, 59
3	1, 59, 117	1, 30, 59
2	1, 117	1, 60
1	58	30

APPENDIX D

TESTING UNIVARIATE ASSUMPTIONS

Table D.1: Run sequence plot of some bands for the New Forest imagery. Pixels were plotted in sequence of an image row.

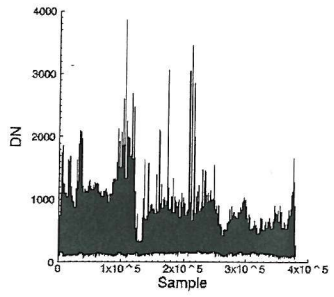
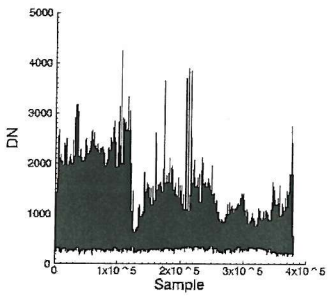
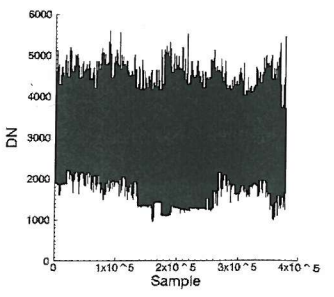
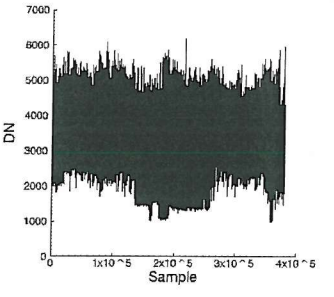
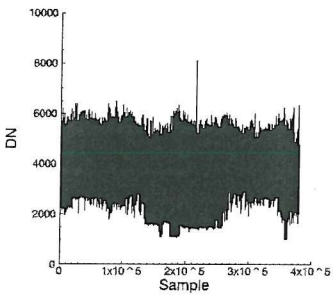
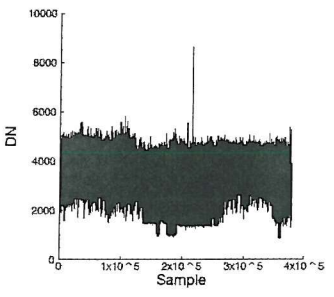
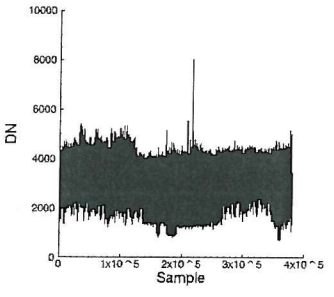
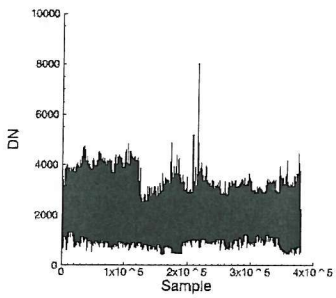
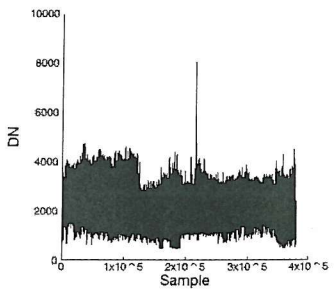
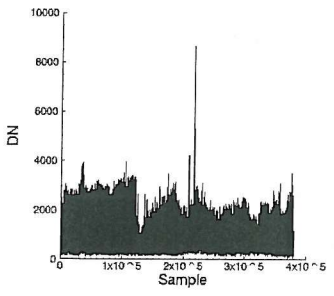
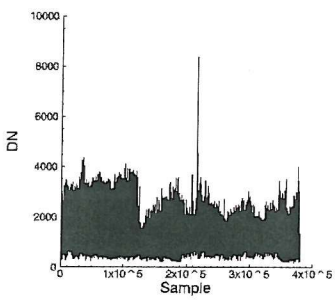
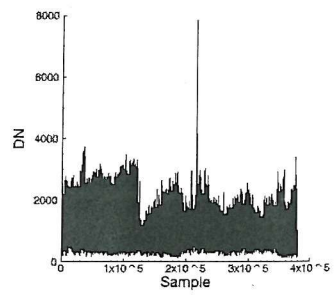
		
Band 1 - 461 nm	**Band 10 – 600 nm**	**Band 20 – 753 nm**
		
Band 30 – 890 nm	**Band 40 – 1048 nm**	**Band 50 – 1196 nm**
		
Band 60 – 1337 nm	**Band 70 – 1571 nm**	**Band 80 – 1698 nm**
		
Band 90 – 1971 nm	**Band 100 – 2155 nm**	**Band 110 – 2326 nm**

Table D.2: Run sequence plot of some bands for the River Severn imagery. Pixels were plotted in sequence of an image row.

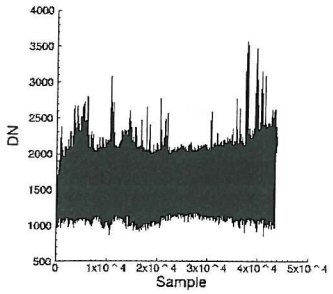
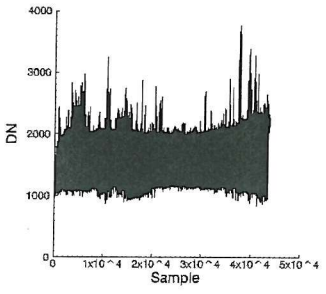
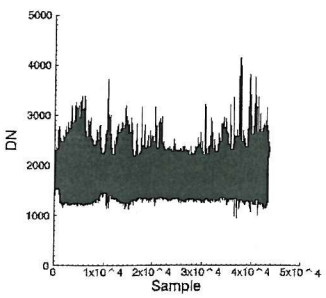
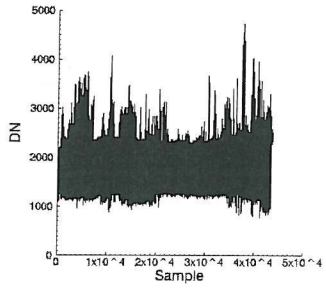
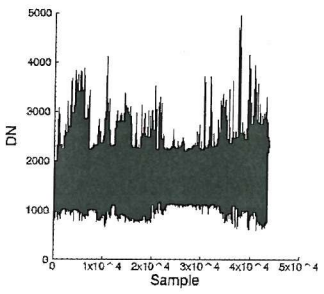
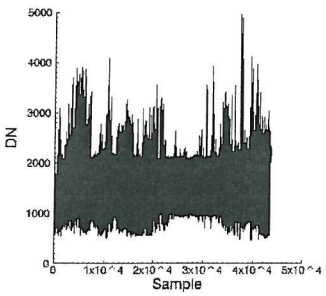
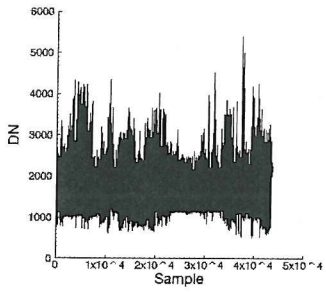
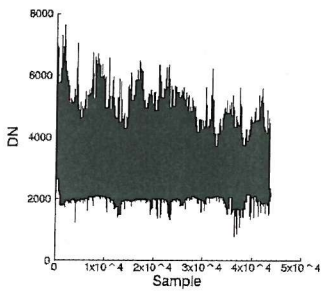
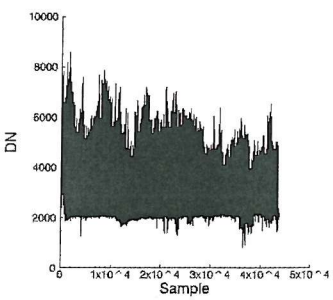
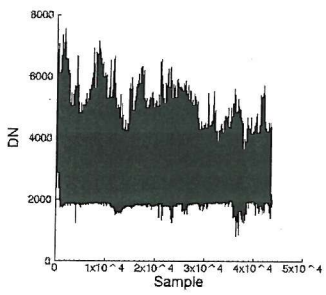
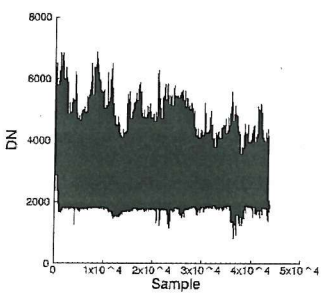
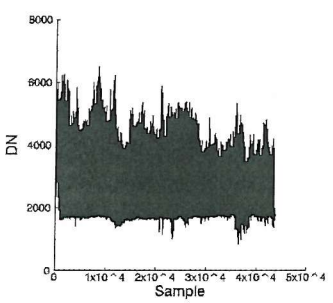
		
Band 1 – 478 nm	**Band 5 – 508 nm**	**Band 10 – 546 nm**
		
Band 15 – 583 nm	**Band 20 – 621 nm**	**Band 25 – 659 nm**
		
Band 30 – 697 nm	**Band 35 – 735 nm**	**Band 40 – 774 nm**
		
Band 45 – 812 nm	**Band 50 – 850 nm**	**Band 55 – 888 nm**

Table D.3: Lag plot of some bands for the New Forest imagery.

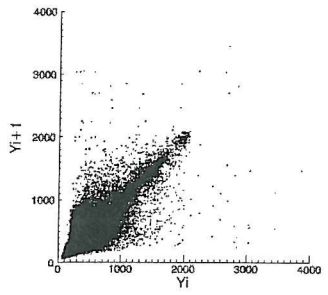
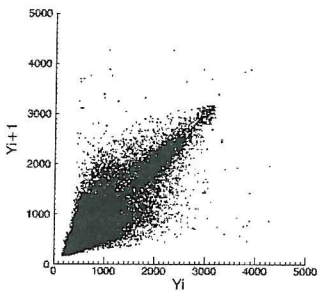
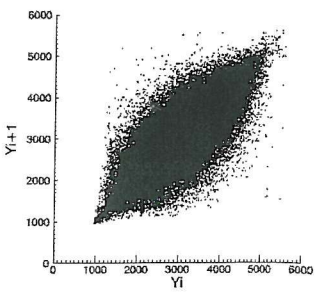
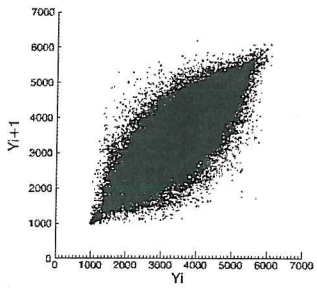
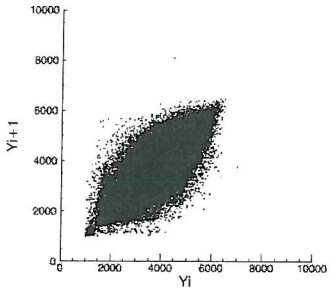
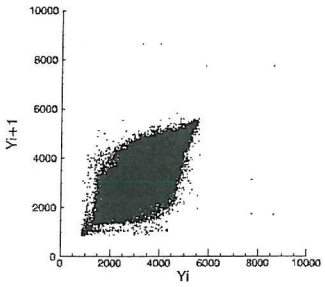
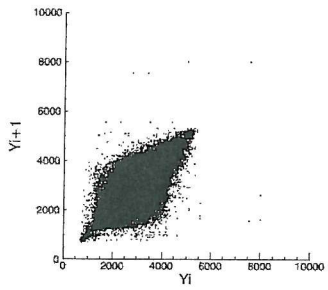
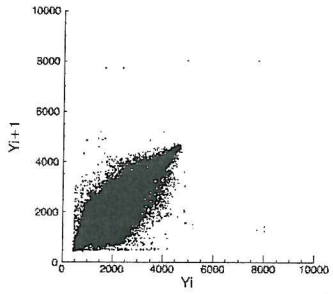
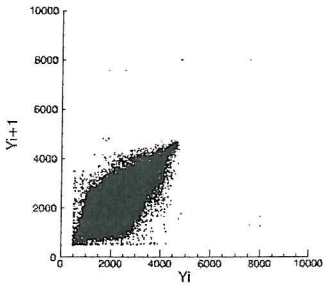
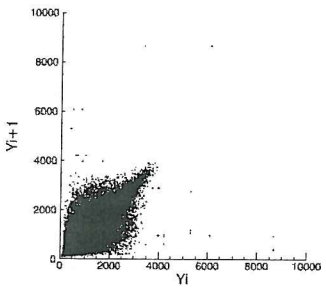
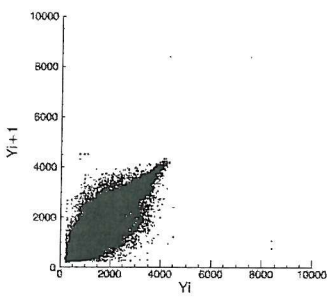
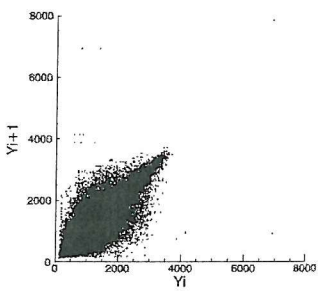
		
Band 1 - 461 nm	**Band 10 – 600 nm**	**Band 20 – 753 nm**
		
Band 30 – 890 nm	**Band 40 – 1048 nm**	**Band 50 – 1196 nm**
		
Band 60 – 1337 nm	**Band 70 – 1571 nm**	**Band 80 – 1698 nm**
		
Band 90 – 1971 nm	**Band 100 – 2155 nm**	**Band 110 – 2326 nm**

Table D.4: Lag plot of some bands for the River Severn imagery.

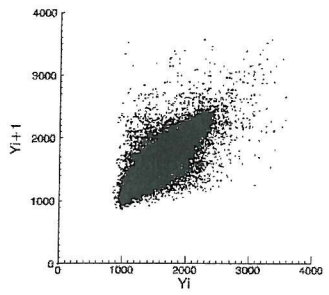
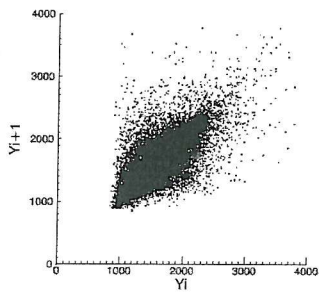
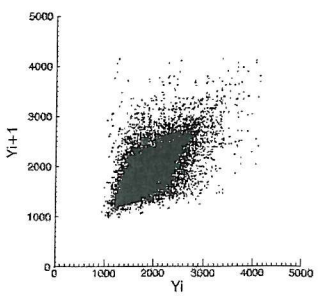
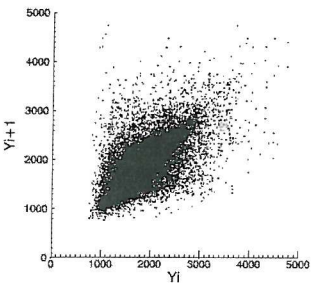
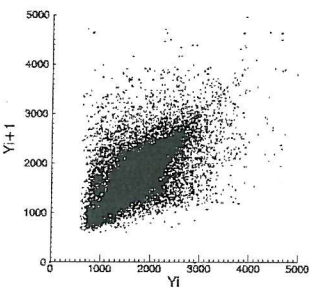
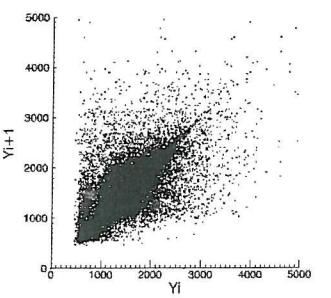
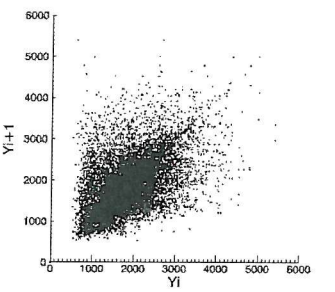
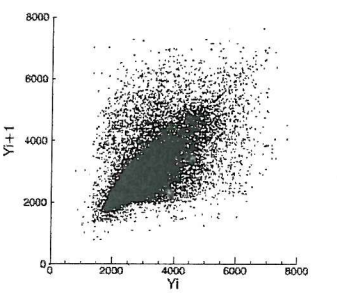
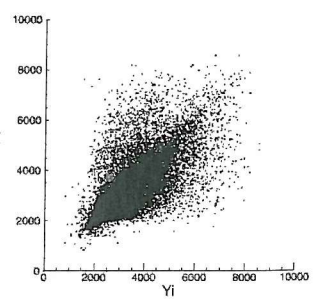
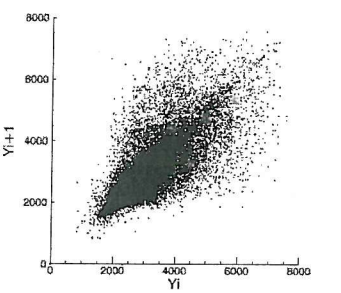
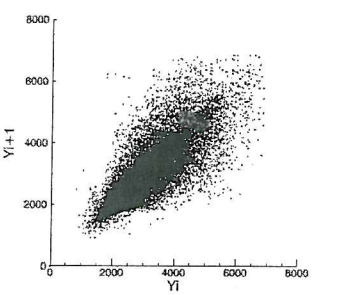
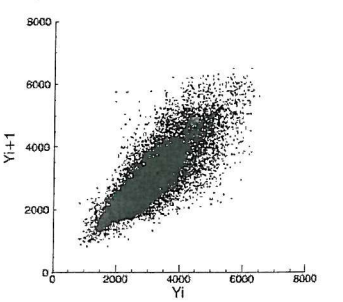
		
Band 1 – 478 nm	**Band 5 – 508 nm**	**Band 10 – 546 nm**
		
Band 15 – 583 nm	**Band 20 – 621 nm**	**Band 25 – 659 nm**
		
Band 30 – 697 nm	**Band 35 – 735 nm**	**Band 40 – 774 nm**
		
Band 45 – 812 nm	**Band 50 – 850 nm**	**Band 55 – 888 nm**

Table D.5: Histogram of some bands for the New Forest imagery with the normal density function overlaid. Bands 20 to 80 binsize 30, others binsize 20.

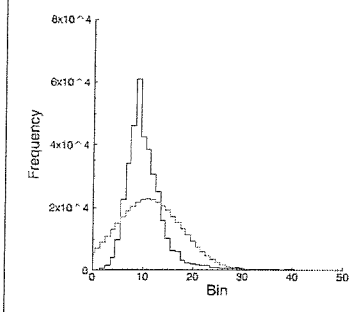
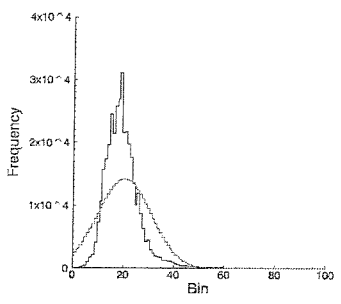
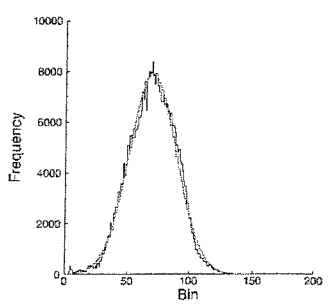
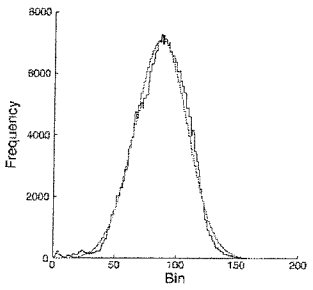
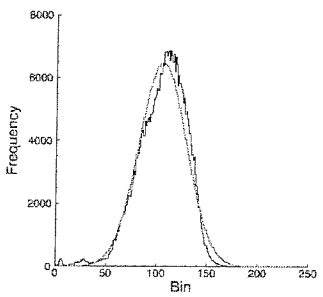
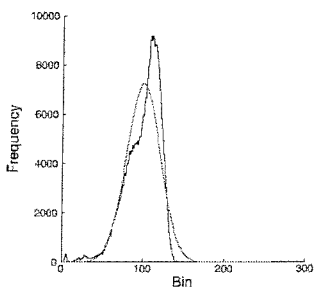
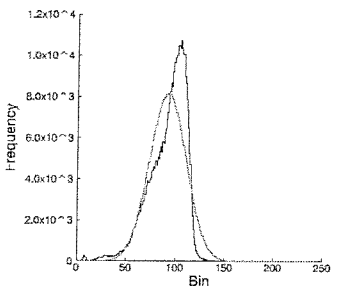
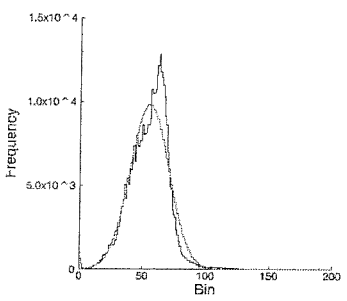
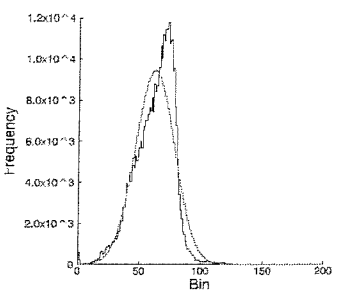
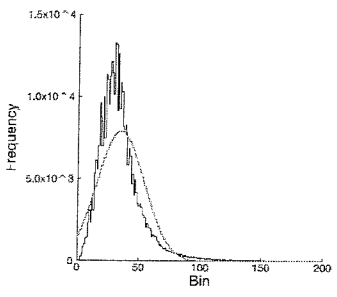
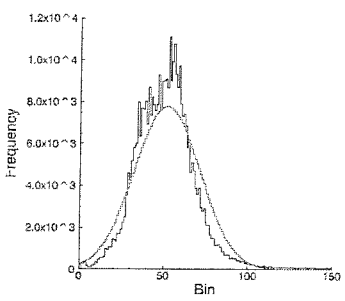
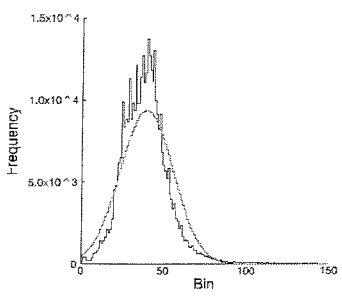
		
Band 1 - 461 nm	**Band 10 – 600 nm**	**Band 20 – 753 nm**
		
Band 30 – 890 nm	**Band 40 – 1048 nm**	**Band 50 – 1196 nm**
		
Band 60 – 1337 nm	**Band 70 – 1571 nm**	**Band 80 – 1698 nm**
		
Band 90 – 1971 nm	**Band 100 – 2155 nm**	**Band 110 – 2326 nm**

Table D.6: Histogram of some bands for the River Severn imagery with the normal density function overlaid.

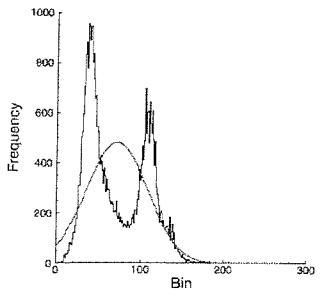
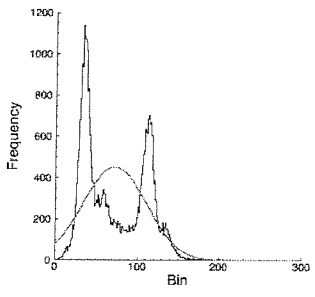
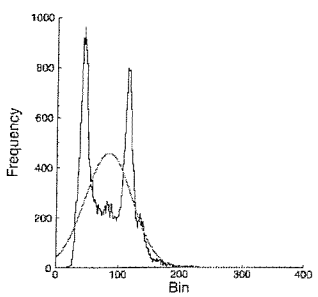
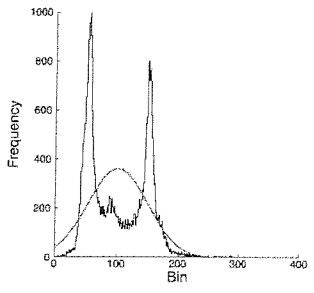
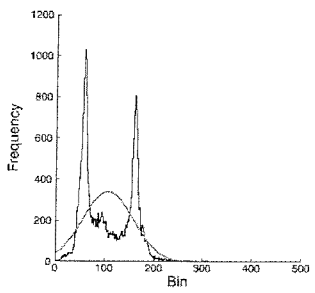
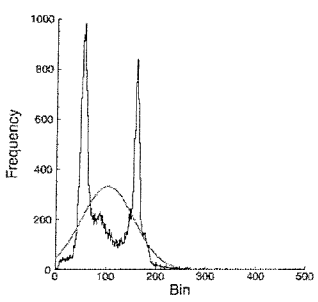
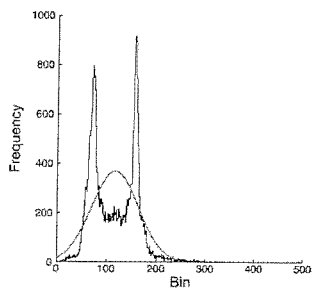
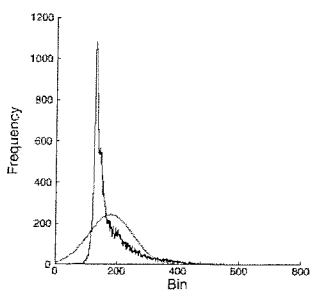
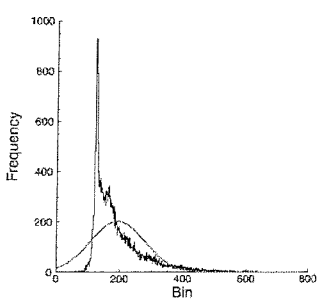
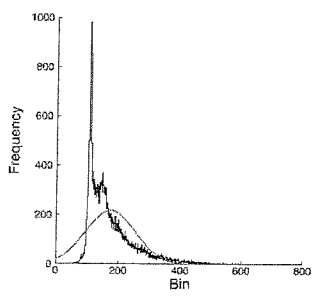
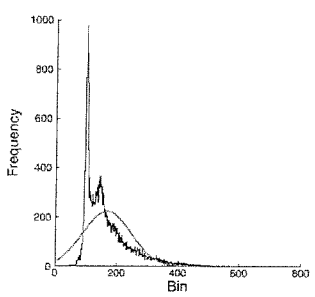
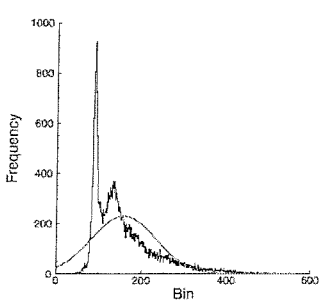
		
Band 1 – 478 nm	**Band 5 – 508 nm**	**Band 10 – 546 nm**
		
Band 15 – 583 nm	**Band 20 – 621 nm**	**Band 25 – 659 nm**
		
Band 30 – 697 nm	**Band 35 – 735 nm**	**Band 40 – 774 nm**
		
Band 45 – 812 nm	**Band 50 – 850 nm**	**Band 55 – 888 nm**

Table D.7: Results of the Chi-square test of independence and the correlation coefficient significance for all bands of the New Forest imagery. Bands that are tested normal receive a 'Y' in the corresponding test column.

Band	χ^2	df	$\chi^2_{5\%,df}$	χ^2 Test 5%	r	N	$r_{5\%,N}$	$r_{1\%,N}$	r Test 5%	r Test 1%
1	2695989	37	52.19231		0.828321	190	0.9897	0.9927		
2	2644475	34	48.60237		0.843645	168	0.9879	0.9915		
3	2750468	34	48.60237		0.847191	162	0.9879	0.9915		
4	3224491	37	52.19231		0.86866	163	0.9879	0.9915		
5	2852819	44	60.48087		0.923616	173	0.9887	0.9919		
6	1708802	52	69.83216		0.94362	190	0.9897	0.9927		
7	1898597	58	76.77777		0.94165	200	0.9903	0.993		
8	3423520	61	80.23212		0.917279	207	0.9903	0.993		
9	1981171	61	80.23212		0.893576	206	0.9903	0.993		
10	3144534	62	81.38103		0.880732	206	0.9903	0.993		
11	2962499	63	82.52869		0.873948	207	0.9903	0.993		
12	3354611	64	83.67525		0.871187	206	0.9903	0.993		
13	3516447	64	83.67525		0.861876	206	0.9903	0.993		
14	3979343	63	82.52869		0.851852	201	0.9903	0.993		
15	3153688	61	80.23212		0.857947	192	0.9897	0.9927		
16	1922378	64	83.67525		0.935533	173	0.9887	0.9919		
17	787887.3	89	112.022		0.989527	183	0.9891	0.9923	Y	
18	14326.12	137	165.3161		0.98914	197	0.9897	0.9927		
19	11348.88	187	219.9059		0.994775	216	0.9907	0.9933	Y	Y
20	9963.368	144	173.004		0.994925	155	0.9871	0.9909	Y	Y
21	13002.25	151	180.6755		0.994098	162	0.9879	0.9915	Y	Y
22	13355.16	153	182.8647		0.992827	164	0.9879	0.9915	Y	Y
23	12724.68	154	183.9587		0.994803	163	0.9879	0.9915	Y	Y
24	15647.93	156	186.1459		0.994722	164	0.9879	0.9915	Y	Y
25	26837.94	158	188.3317		0.994215	166	0.9879	0.9915	Y	Y
26	32499.55	159	189.4243		0.99403	167	0.9879	0.9915	Y	Y
27	42712.6	161	191.6084		0.985835	169	0.9879	0.9915		
28	55626.13	162	192.7		0.984447	170	0.9887	0.9919		
29	43663.6	161	191.6084		0.993998	169	0.9879	0.9915	Y	Y
30	47118.07	164	194.883		0.993882	174	0.9887	0.9919	Y	Y
31	61277.87	166	197.0636		0.992742	184	0.9891	0.9923	Y	Y
32	70804.59	168	199.2443		0.991905	187	0.9891	0.9923	Y	
33	94670.28	167	198.1539		0.988572	197	0.9897	0.9927		
34	193092.5	165	195.9735		0.983869	214	0.9907	0.9933		
35	202942.4	165	195.9735		0.981901	215	0.9907	0.9933		
36	227307.8	168	199.2443		0.981792	217	0.9907	0.9933		
37	227206.8	171	202.5128		0.983277	220	0.991	0.9936		
38	213941.6	175	206.8667		0.984877	226	0.991	0.9936		
39	206630.2	179	211.2171		0.984937	229	0.991	0.9936		
40	229234.9	181	213.3907		0.986571	237	0.9914	0.9939		
41	211780.9	185	217.735		0.986313	248	0.9917	0.9941		
42	241179.3	186	218.8204		0.985054	247	0.9917	0.9941		
43	258272.8	185	217.735		0.983352	243	0.9917	0.9941		
44	272712	185	217.735		0.979235	248	0.9917	0.9941		
45	294975.9	180	212.3041		0.97105	248	0.9917	0.9941		
46	297487.2	176	207.9546		0.959548	262	0.9924	0.9945		
47	341718.2	169	200.334		0.943158	270	0.9926	0.9947		
48	338909.5	166	197.0636		0.942332	270	0.9926	0.9947		
49	306793.8	165	195.9735		0.940475	256	0.9921	0.9943		
50	317038.4	164	194.883		0.939146	260	0.9924	0.9945		
51	330993.3	163	193.7914		0.938731	259	0.9921	0.9943		

Table D.7 continued.

52	335642.3	164	194.883		0.939026	257	0.9921	0.9943		
53	336281	165	195.9735		0.93931	254	0.9921	0.9943		
54	375318.6	164	194.883		0.939894	257	0.9921	0.9943		
55	368709.6	165	195.9735		0.939564	257	0.9921	0.9943		
56	406736.7	162	192.7		0.936341	255	0.9921	0.9943		
57	375566.8	160	190.5163		0.932151	254	0.9921	0.9943		
58	366459.8	156	186.1459		0.928928	252	0.9921	0.9943		
59	348618.2	151	180.6755		0.927204	239	0.9914	0.9939		
60	334315.6	146	175.1974		0.927355	244	0.9917	0.9941		
61	816814	106	131.0314		0.977097	313	0.9936	0.9954		
62	851467.3	101	125.4585		0.977169	283	0.9929	0.9949		
63	750741.6	101	125.4585		0.98119	279	0.9926	0.9947		
64	602108	101	125.4585		0.983095	283	0.9929	0.9949		
65	512463.8	103	127.6893		0.984662	275	0.9926	0.9947		
66	413800.3	105	129.9178		0.984756	265	0.9924	0.9945		
67	296859.4	107	132.1444		0.983588	263	0.9924	0.9945		
68	236510.4	110	135.48		0.980451	260	0.9924	0.9945		
69	204619.6	113	138.8114		0.977211	257	0.9921	0.9943		
70	146092.6	115	141.0297		0.973023	253	0.9921	0.9943		
71	117770.9	117	143.2461		0.970774	256	0.9921	0.9943		
72	110505.1	120	146.5674		0.967377	257	0.9921	0.9943		
73	116997.5	122	148.7792		0.962877	261	0.9924	0.9945		
74	111054.5	123	149.8844		0.960315	255	0.9921	0.9943		
75	128289.9	124	150.9894		0.957529	254	0.9921	0.9943		
76	119078.4	126	153.198		0.959037	254	0.9921	0.9943		
77	131919.7	126	153.198		0.95965	253	0.9921	0.9943		
78	127739.4	126	153.198		0.957962	252	0.9921	0.9943		
79	124697.3	125	152.0938		0.958497	257	0.9921	0.9943		
80	121224.1	124	150.9894		0.957068	252	0.9921	0.9943		
81	118217.8	124	150.9894		0.957001	260	0.9924	0.9945		
82	113880.9	122	148.7792		0.959196	256	0.9921	0.9943		
83	102525.2	120	146.5674		0.961157	261	0.9924	0.9945		
84	102326.8	118	144.3536		0.96496	257	0.9921	0.9943		
85	104394.3	116	142.1382		0.967015	256	0.9921	0.9943		
86	115654	114	139.9207		0.969785	249	0.9917	0.9941		
87	116903.3	113	138.8114		0.973321	254	0.9921	0.9943		
88	118059.1	115	141.0297		0.975285	252	0.9921	0.9943		
89	95838.63	122	148.7792		0.971323	278	0.9926	0.9947		
90	1454759	109	134.3687		0.926594	429	0.9951	0.9965		
91	1419165	108	133.2568		0.938037	381	0.9947	0.9961		
92	1926034	111	136.5912		0.938469	378	0.9945	0.996		
93	1320124	113	138.8114		0.952057	401	0.9949	0.9963		
94	1317003	113	138.8114		0.954097	374	0.9945	0.996		
95	2231611	117	143.2461		0.953261	396	0.9948	0.9962		
96	1525364	119	145.4608		0.956593	404	0.9949	0.9963		
97	2254609	122	148.7792		0.961835	412	0.995	0.9964		
98	2156259	125	152.0938		0.965696	413	0.995	0.9964		
99	1287918	126	153.198		0.969848	414	0.995	0.9964		
100	1081821	127	154.3015		0.974177	411	0.995	0.9964		
101	888410.6	127	154.3015		0.979143	403	0.9949	0.9963		
102	655780.4	129	156.5078		0.983158	400	0.9949	0.9963		
103	477414.4	128	155.4046		0.984236	409	0.9949	0.9963		
104	432058.7	126	153.198		0.982985	400	0.9949	0.9963		
105	662228.1	123	149.8844		0.982596	399	0.9948	0.9962		
106	992472.3	119	145.4608		0.974699	398	0.9948	0.9962		
107	983773.3	114	139.9207		0.96746	398	0.9948	0.9962		

Table D.7 continued.

108	1501190	111	136.5912		0.963958	389	0.9947	0.9961		
109	1370825	107	132.1444		0.963242	400	0.9949	0.9963		
110	2046046	103	127.6893		0.964275	386	0.9947	0.9961		
111	1708786	100	124.342		0.962271	378	0.9945	0.996		
112	1387989	97	120.9898		0.956849	377	0.9945	0.996		
113	1440751	92	115.3898		0.958351	353	0.9942	0.9958		
114	1905514	91	114.2679		0.953606	363	0.9944	0.9959		
115	1371047	84	106.3949		0.942887	328	0.9937	0.9955		
116	1596817	83	105.2672		0.926044	351	0.9942	0.9958		
117	1940270	75	96.21669		0.913432	318	0.9936	0.9954		

Table D.8: Results of the Chi-square test of independence and the correlation coefficient significance for all bands of the River Severn imagery. Bands that are tested normal receive a 'Y' in the corresponding 'Test' column.

Band	χ^2	df	$\chi^2_{5\%,df}$	χ^2 Test 5%	r	N	$r_{5\%,N}$	$r_{1\%,N}$	r Test 5%	r Test 1%
1	25745.97	187	219.9059		0.657105	271	0.9926	0.9947		
2	26434.38	182	214.4769		0.651317	263	0.9924	0.9945		
3	32865.63	195	228.5799		0.614998	287	0.9929	0.9949		
4	36362.29	192	225.3289		0.589005	280	0.9929	0.9949		
5	37097.46	194	227.4966		0.593531	294	0.9931	0.9951		
6	38120.86	186	218.8204		0.597599	280	0.9929	0.9949		
7	34670.22	185	217.735		0.622309	287	0.9929	0.9949		
8	32327.18	189	222.0759		0.650476	296	0.9931	0.9951		
9	29994.72	197	230.7463		0.675591	310	0.9936	0.9954		
10	30200.94	205	239.4034		0.679487	320	0.9937	0.9955		
11	30596.37	214	249.1278		0.677585	340	0.9941	0.9957		
12	33362.68	216	251.2863		0.652726	343	0.9941	0.9957		
13	37589.38	226	262.0704		0.623333	356	0.9942	0.9958		
14	42134.55	236	272.8355		0.592986	369	0.9944	0.9959		
15	43955.66	252	290.0283		0.583253	397	0.9948	0.9962		
16	45233.77	253	291.1018		0.579389	403	0.9949	0.9963		
17	45162.3	260	298.6103		0.580565	417	0.995	0.9964		
18	44974.8	263	301.8268		0.584151	431	0.9953	0.9966		
19	45771.46	260	298.6103		0.579957	424	0.9951	0.9965		
20	45955.73	265	303.9698		0.580431	436	0.9953	0.9966		
21	45768.54	257	295.3927		0.583787	423	0.9951	0.9965		
22	45821.81	268	307.184		0.585992	446	0.9954	0.9966		
23	47902.96	275	314.6792		0.574316	455	0.9954	0.9967		
24	49319.24	276	315.7484		0.567542	460	0.9955	0.9968		
25	50241.66	265	303.9698		0.566643	449	0.9954	0.9966		
26	50145.38	278	317.8887		0.572444	488	0.9957	0.9969		
27	50153.73	276	315.7484		0.57606	492	0.9958	0.9969		
28	50223.11	277	316.8181		0.578705	498	0.9958	0.9969		
29	47193.51	232	268.5312		0.60157	422	0.9951	0.9965		
30	35719.64	264	302.8982		0.693972	488	0.9957	0.9969		
31	35252.54	271	310.3962		0.802536	497	0.9958	0.9969		
32	58854.63	287	327.512		0.803959	497	0.9958	0.9969		
33	73466.04	289	329.6488		0.768421	445	0.9954	0.9966		
34	91360.2	330	373.3634		0.648499	544	0.9961	0.9972		
35	84804.27	395	442.3406		0.645151	685	0.9969	0.9977		

Table D.8 continued.

36	87534.04	429	478.291		0.664524	757	0.9972	0.998		
37	80842.16	445	495.1805		0.676309	794	0.9973	0.998		
38	77591.45	353	397.8122		0.695696	618	0.9965	0.9975		
39	83096.08	268	307.184		0.688495	460	0.9955	0.9968		
40	77848.88	446	496.2353		0.6853	774	0.9972	0.998		
41	75850.64	450	500.4561		0.686709	776	0.9973	0.998		
42	72577.86	443	493.0694		0.686652	755	0.9972	0.998		
43	72999.6	433	482.5141		0.680918	729	0.9971	0.9979		
44	71435.68	431	480.4034		0.685753	715	0.997	0.9978		
45	70597.85	412	460.325		0.682476	672	0.9968	0.9977		
46	65801.63	373	419.0344		0.681708	597	0.9964	0.9974		
47	64353.43	387	433.8703		0.673963	605	0.9965	0.9975		
48	69795.97	403	450.8068		0.672091	620	0.9965	0.9975		
49	62889.7	413	461.383		0.679534	630	0.9967	0.9976		
50	59953.13	400	447.6326		0.680637	603	0.9965	0.9975		
51	60684.27	393	440.2236		0.676683	594	0.9964	0.9974		
52	57643.12	387	433.8703		0.681928	579	0.9964	0.9974		
53	54928.68	394	441.2827		0.682458	590	0.9964	0.9974		
54	53439	390	437.048		0.686423	574	0.9963	0.9973		
55	51872.64	384	430.6916		0.688824	567	0.9963	0.9973		
56	48203.35	339	382.9361		0.690238	497	0.9958	0.9969		
57	45828.28	319	361.6525		0.694968	461	0.9955	0.9968		
58	44520.63	305	346.7295		0.706789	441	0.9954	0.9966		
59	42373.58	321	363.7822		0.70996	462	0.9955	0.9968		
60	37642.58	295	336.057		0.734216	422	0.9951	0.9965		

APPENDIX E

MATHEMATICAL PROOF

E. Mathematical Proof

This appendix demonstrates a mathematical proof that shows that once the least correlated band pair has been chosen of a set, any increase in the width of one of the bands will result in a higher correlation coefficient. Given are three bands X, Y and Z of identical length. The null hypothesis is then defined as follows:

$$H_0\text{: If } r_{X,Z}^2 < r_{Y,Z}^2 \text{ then } r_{X+Y,Z}^2 > r_{X,Z}^2$$

, where $r_{X,Y}$ is the Pearson correlation coefficient between bands X and Y, and $X+Y$ the sum of vectors X and Y. That is, if the correlation between bands X and Z is smaller than the one between bands Y and Z, the merger (or sum) of band X and Y would result in a higher correlation with band Z than band X would do on its own.

The proof was performed with the correlation coefficient r instead of the coefficient of determination r^2, and therefore two cases need to be distinguished:

1) Hypothesis 1: If $0 < r_{X,Z} < r_{Y,Z}$ then $r_{X+Y,Z} > r_{X,Z}$

2) Hypothesis 2: If $r_{Y,Z} < r_{X,Z} < 0$ then $r_{X+Y,Z} < r_{X,Z}$

Proof of Hypothesis 1 ($r_{X,Z} > 0$)

The condition in the hypothesis (equation E.2) may be re-written using equation 2.4:

$$0 < r_{X,Z} < r_{Y,Z} \tag{E.1}$$

$$r_{X,Z} = \frac{s_{X,Z}}{s_X s_Z} < \frac{s_{Y,Z}}{s_Y s_Z} = r_{Y,Z} \tag{E.2}$$

$$s_{X,Z}\frac{s_Y}{s_X} < s_{Y,Z} \tag{E.3}$$

where $s_{X,Y}$ Covariance between vectors X and Y

s_X Standard deviation of vector X

The hypothesis inequality that has to be proven (equation E.4) may be expanded on both sides resulting in equation E.6.

$$r_{X+Y,Z} > r_{X,Z} \tag{E.4}$$

$$r_{X+Y,Z} = \frac{s_{X+Y,Z}}{s_{X+Y}s_Z} = \frac{s_{X,Z}+s_{Y,Z}}{s_{X+Y}s_Z} > \frac{s_{X,Z}}{s_X s_Z} = r_{X,Z} \tag{E.5}$$

$$\frac{s_{X,Z}+s_{Y,Z}}{s_{X+Y}} > \frac{s_{X,Z}}{s_X} \tag{E.6}$$

The covariance $s_{Y,Z}$ on the left side of equation E.6 may be bounded using the hypothesis condition E.3:

$$\frac{s_{X,Z}+s_{Y,Z}}{s_{X+Y}} > \frac{s_{X,Z}+s_{X,Z}\dfrac{s_Y}{s_X}}{s_{X+Y}} \tag{E.7}$$

If the hypothesis inequality is true under its condition (E.7), inequality E.7 should be bounded to the right by the right side of inequality E.6:

$$\frac{s_{X,Z}+s_{Y,Z}}{s_{X+Y}} > \frac{s_{X,Z}\left(1+\dfrac{s_Y}{s_X}\right)}{s_{X+Y}} > \frac{s_{X,Z}}{s_X} \tag{E.8}$$

The middle and the right ratio of inequality E.8 may be expressed as:

$$s_{X,Z}\frac{s_X+s_Y}{s_{X+Y}} > s_{X,Z} \tag{E.9}$$

As according to the hypothesis condition, $r_{X,Z}$, and therefore $s_{X,Z}$, is positive, inequality E.9 can be rewritten as

$$s_X + s_Y > s_{X+Y} \tag{E.10}$$

, squared as the standard deviations are always positive, expanded and rewritten:

$$\left(s_X+s_Y\right)^2 > s_{X+Y}^2 \tag{E.11}$$

$$s_X^2+s_Y^2+2s_Xs_Y > s_X^2+s_Y^2+2s_{X,Y} \tag{E.12}$$

$$s_Xs_Y > s_{X,Y} \tag{E.13}$$

$$1 > r_{X,Y} \tag{E.14}$$

The latter inequality E.14 is always true except for when $r_{X,Y}$ equals to 1. Hypothesis 1 is therefore proven if bands if bands X and Y are not linearly dependent, which is mostly the case for remotely sensed data.

Proof of Hypothesis 2 ($r_{X,Z} < 0$)

The condition in the hypothesis (equation E.15) may be re-written using equation 2.4:

$$r_{Y,Z} < r_{X,Z} < 0 \tag{E.15}$$

$$r_{X,Z} = \frac{s_{X,Z}}{s_Xs_Z} > \frac{s_{Y,Z}}{s_Ys_Z} = r_{Y,Z} \tag{E.16}$$

$$s_{X,Z}\frac{s_Y}{s_X} > s_{Y,Z} \tag{E.17}$$

The hypothesis inequality that has to be proven (equation E.18) may be expanded on both sides resulting in equation E.20.

$$r_{X+Y,Z} < r_{X,Z} \tag{E.18}$$

$$r_{X+Y,Z} = \frac{s_{X+Y,Z}}{s_{X+Y}s_Z} = \frac{s_{X,Z} + s_{Y,Z}}{s_{X+Y}s_Z} < \frac{s_{X,Z}}{s_X s_Z} = r_{X,Z} \tag{E.19}$$

$$\frac{s_{X,Z} + s_{Y,Z}}{s_{X+Y}} < \frac{s_{X,Z}}{s_X} \tag{E.20}$$

The covariance $s_{Y,Z}$ on the left side of equation E.20 may be bounded using the hypothesis condition E.17:

$$\frac{s_{X,Z} + s_{Y,Z}}{s_{X+Y}} < \frac{s_{X,Z} + s_{X,Z}\frac{s_Y}{s_X}}{s_{X+Y}} \tag{E.21}$$

If the hypothesis inequality is true under its condition (E.21), inequality E.21 should be bounded to the right by the right side of inequality E.20:

$$\frac{s_{X,Z} + s_{Y,Z}}{s_{X+Y}} < \frac{s_{X,Z}\left(1 + \frac{s_Y}{s_X}\right)}{s_{X+Y}} < \frac{s_{X,Z}}{s_X} \tag{E.22}$$

The middle and the right ratio of inequality E.22 may be expressed as:

$$s_{X,Z}\frac{s_X + s_Y}{s_{X+Y}} < s_{X,Z} \tag{E.23}$$

As according to the hypothesis condition, $r_{X,Z}$, and therefore $s_{X,Z}$, is negative, inequality E.23 can be rewritten as

$$s_X + s_Y > s_{X+Y} \tag{E.24}$$

, resulting in inequality E.25 after similar transformations between E.11 and E.14:

$$1 > r_{X,Y} \tag{E.25}$$

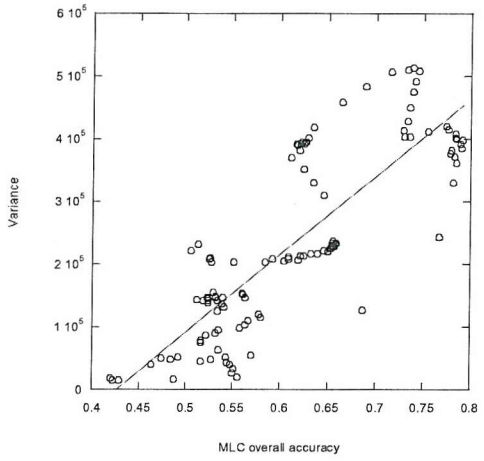
The latter inequality E.25 is always true except for when $r_{X,Y}$ equals to 1. Hypothesis 2 is therefore proven if bands X and Y are not linearly dependent, which is mostly the case for remotely sensed data.

As both hypothesis 1 and 2 have been proven, the null hypothesis has been proven.

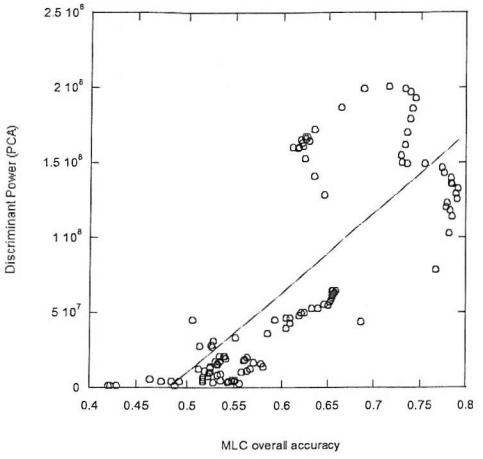
APPENDIX F

UNSUPERVISED BAND SELECTION – SUPPLEMENTARY DATA

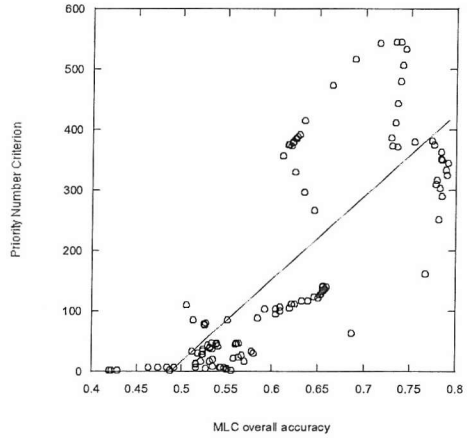
Table F.1: Scatter plots of image quality measures against MLC overall performance estimated with the leave-one-out method for the New Forest data set. The regression line is displayed and the correlation coefficient r given.



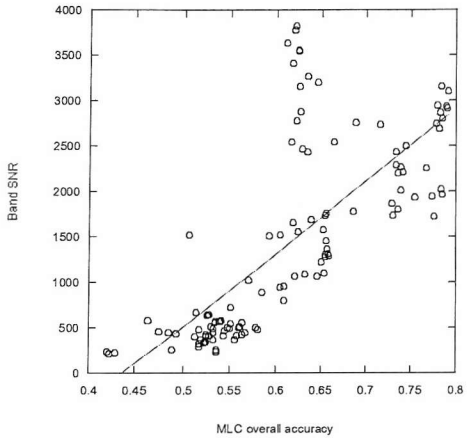
Variance – r=0.82364



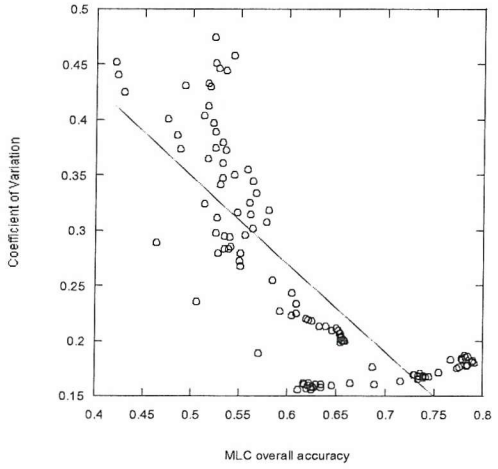
Discriminant Power PCA - r=0.76679



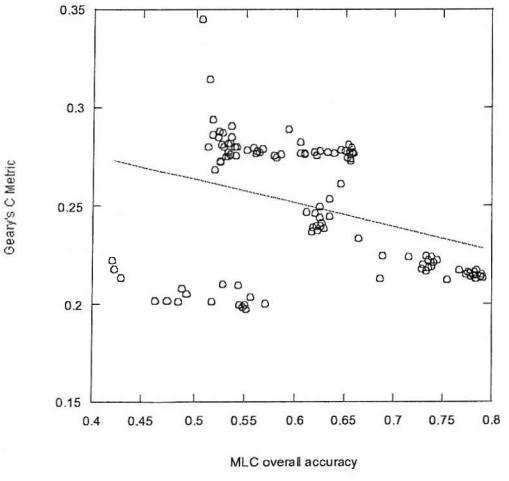
Priority number criterion – r=0.77901



Band SNR – r=0.73215

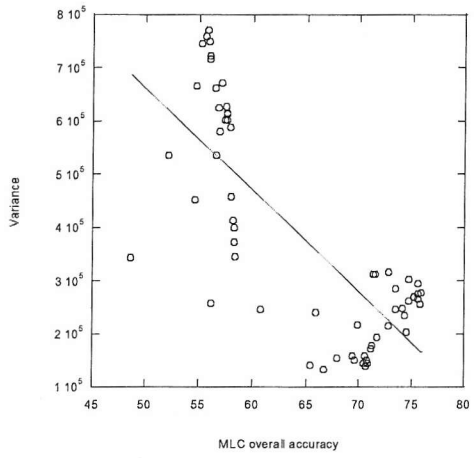


Coefficient of variation – r=-0.79911

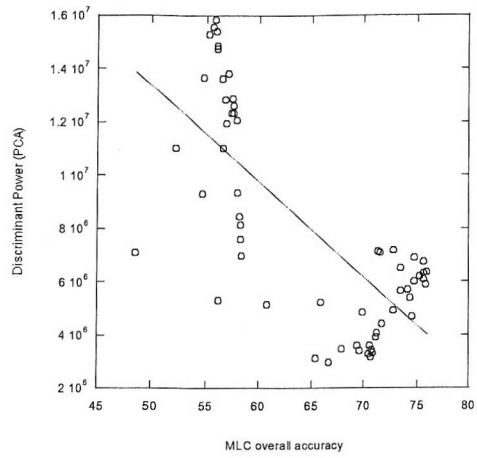


Geary's c metric - r=-0.35318

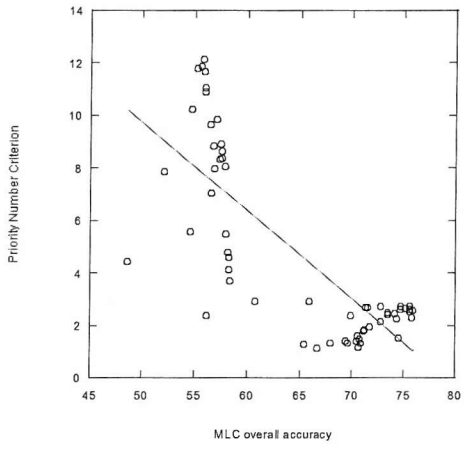
Table F.2: Scatter plots of image quality measures against MLC overall performance estimated with the leave-one-out method for the River Severn data set. The regression line is displayed and the correlation coefficient r given.



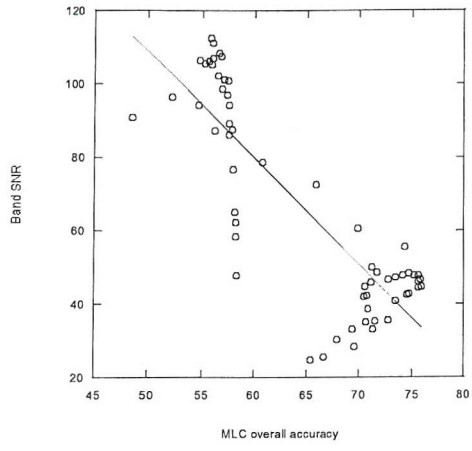
Variance – r=-0.76569



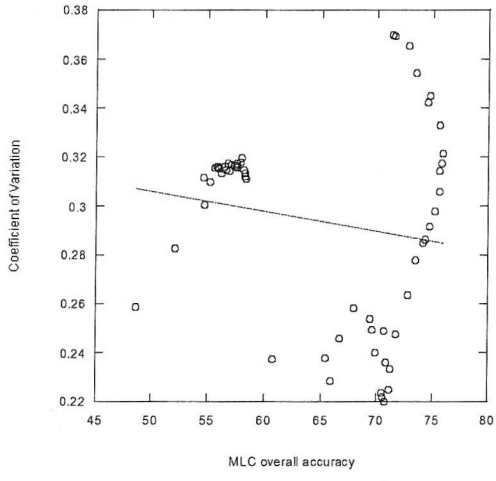
Discriminant Power PCA - r=-0.73442



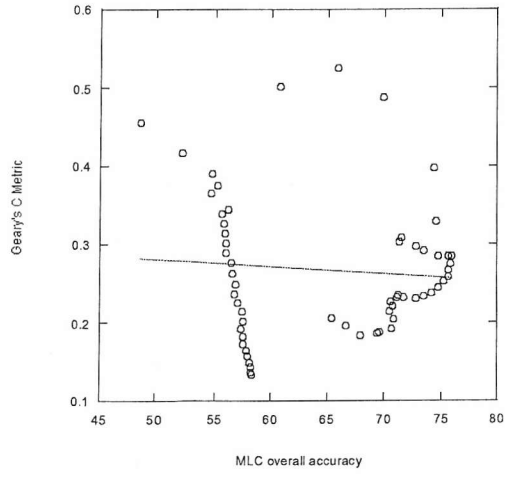
Priority number criterion – r=-0.77982



Band SNR – r=-0.84406



Coefficient of variation – r=-0.16767



Geary's c metric - r=-0.08399

Table F.3: Frequency histograms for the 2 clusters formed with the K-Means algorithm for the River Severn data set for bands 21, 33, 45 and 57. The normal curve is overlaid.

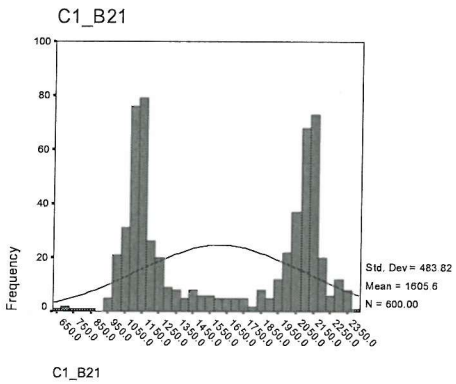
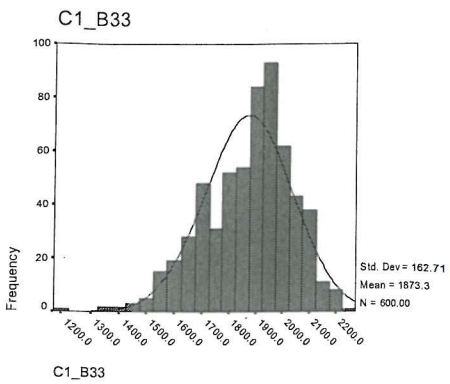
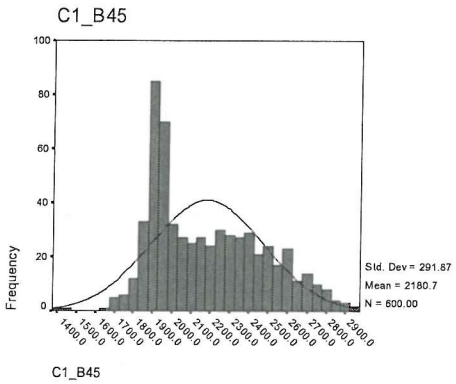
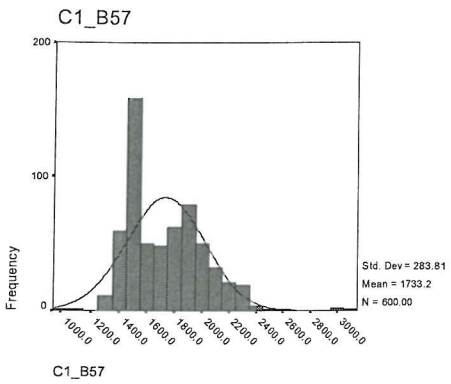
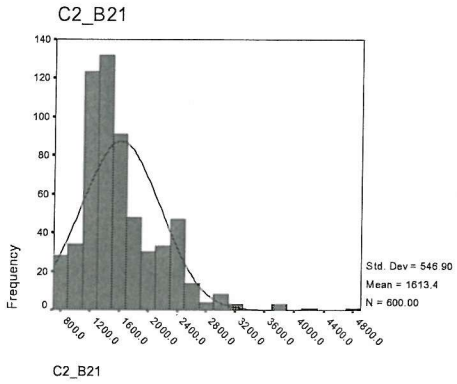
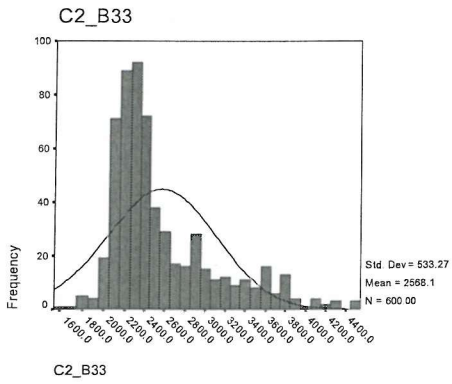
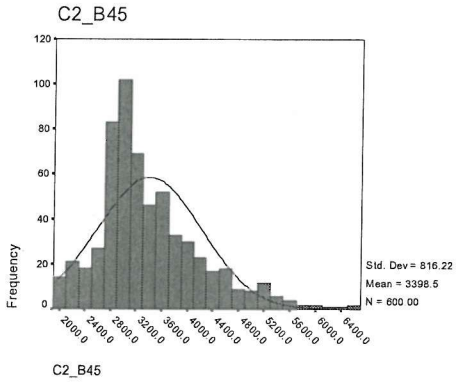
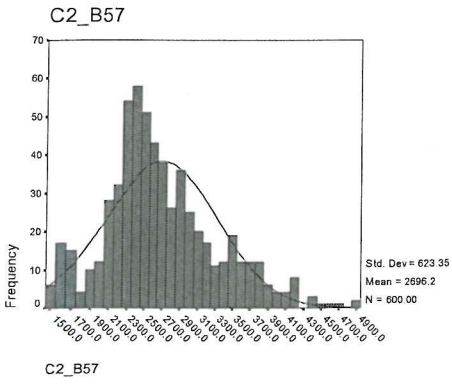
Classes	Band 21,45	Band 33, 57
Cluster 1		
		
Cluster 2		
		

Table F.4: Frequency histograms for the 5 clusters formed with the K-Means algorithm for the New Forest data set for bands 10, 40, 80 and 100. The normal curve is overlaid.

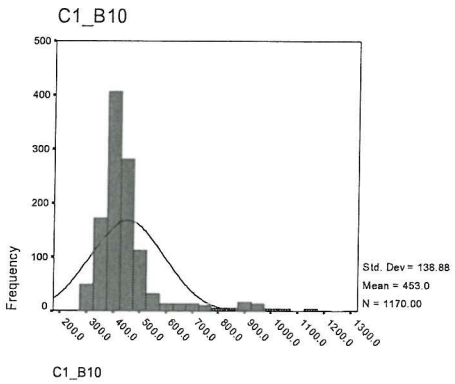
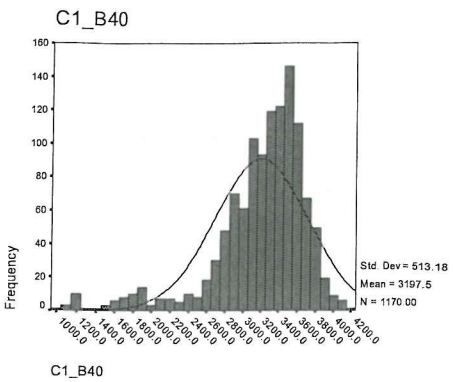
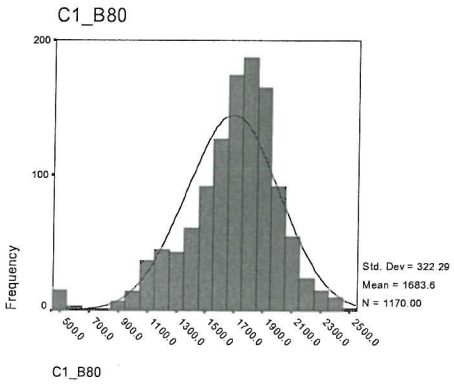
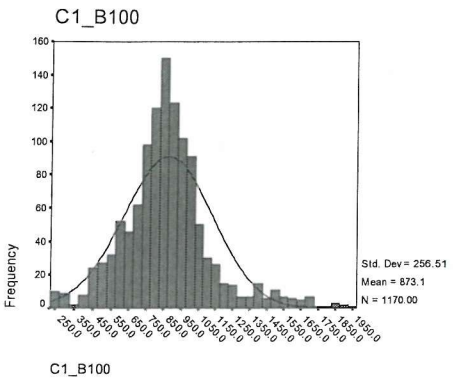
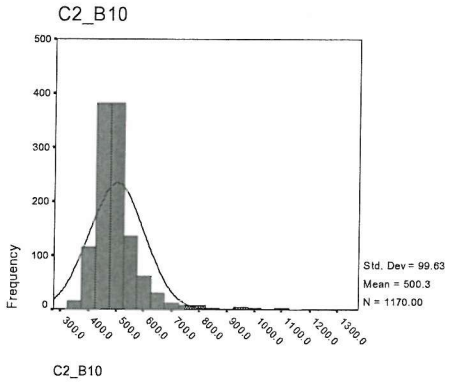
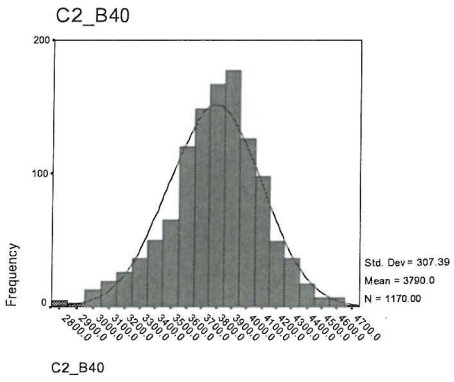
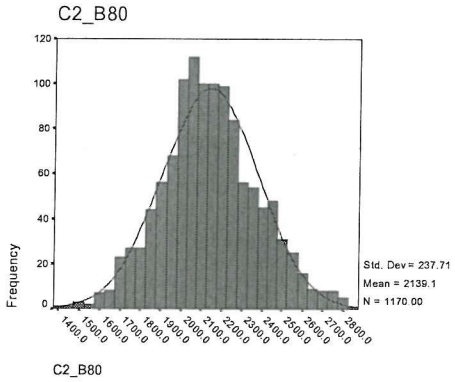
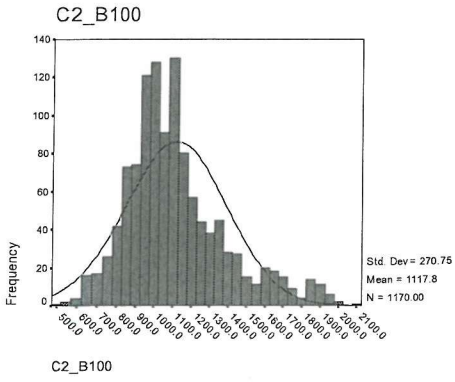
Classes	Band 10,80	Band 40, 100
Cluster 1		
		
Cluster 2		
		

Table F.4 continued.

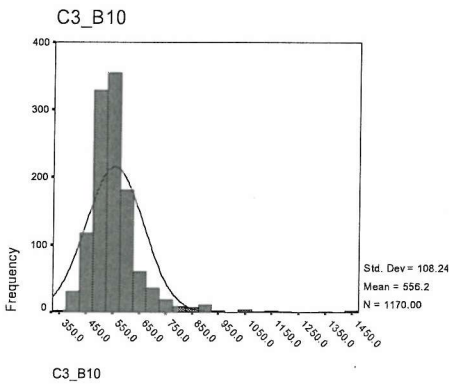
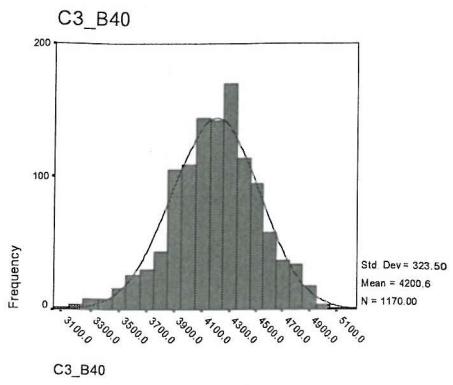
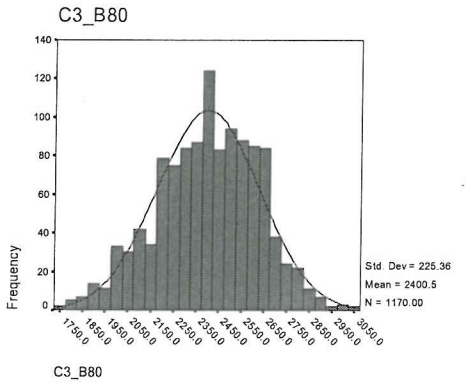
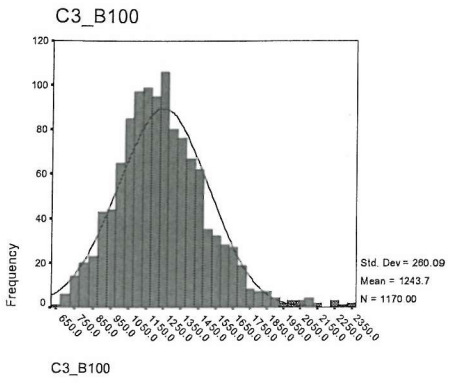
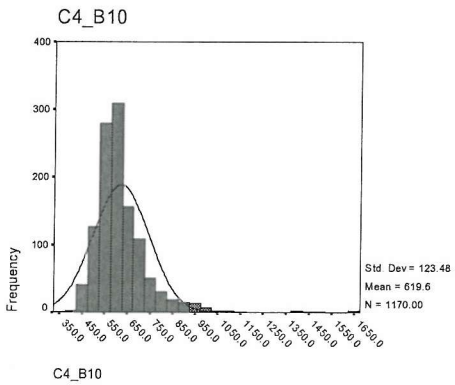
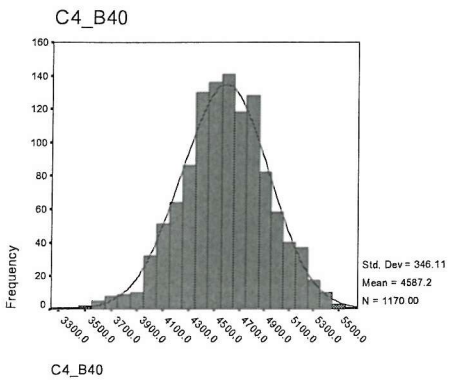
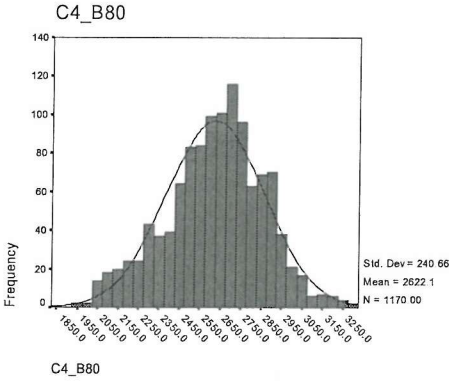
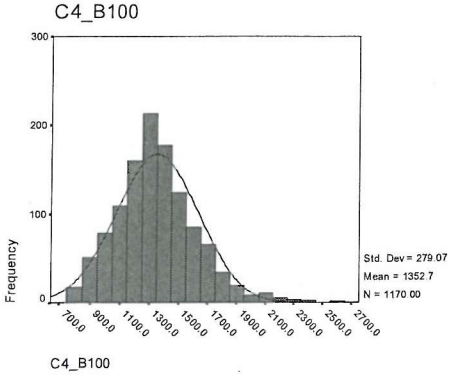
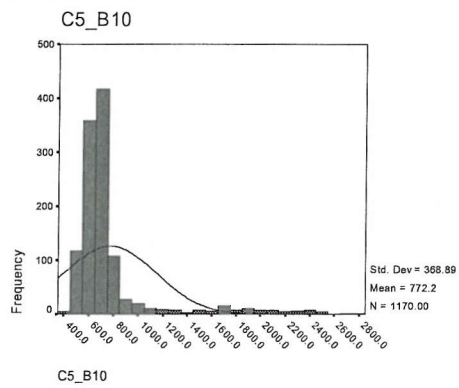
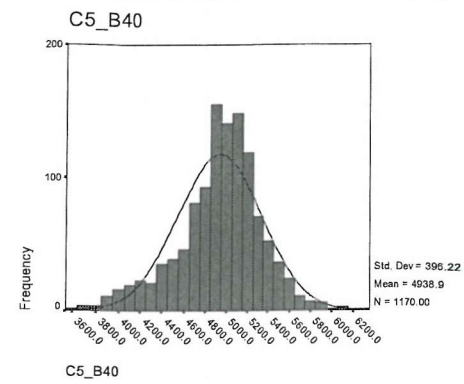
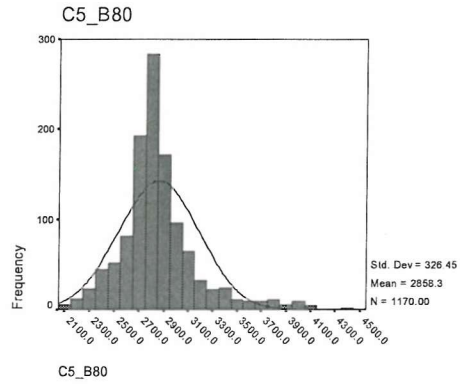
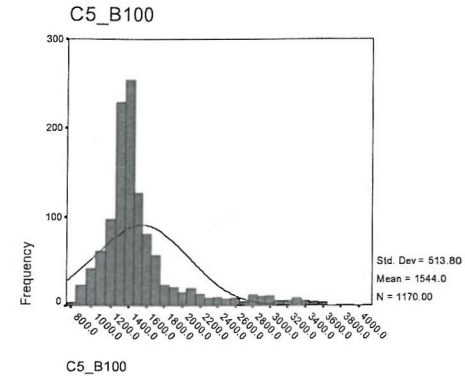
Classes	Band 10,80	Band 40, 100
Cluster 3		
		
Cluster 4		
		

Table F.4 continued.

Classes	Band 10,80	Band 40, 100
Cluster 5		
		

APPENDIX G

DESCRIPTION OF THE MAIN PROGRAMS DEVELOPED

G. Description of the Main Programs Developed

All algorithms were written in IDL™ (Version 5.5 Win 32) using ENVI™ (Version 3.5) sub-routines. In order to run the programs, at least the runtime licence of ENVI™ is required. The programs were implemented under the Windows™ NT operating system but may theoretically run under any ENVI™ compatible operating system.

The programs can be obtained from the author in binary, already compiled format (with a *.sav* extension), which protects them from accidental modifications. They were developed in a user-friendly environment, using buttons and pull-down menus, avoiding direct keyboard inputs by the user where possible. The programs were written to be computationally efficient and to handle large image data sets.

G.1 Description of the Program SBS (Supervised Band Selection)

Overview

A computer program called SBS (Supervised Band Selection) was developed to estimate an optimal band set for Maximum Likelihood classification (MLC) as application procedure. It uses hyperspectral airborne or satellite data collected over a representative part of the scene and the class definition by the user to estimate an optimal band set of user-specified dimension. Optimality refers to the achievable MLC accuracy. This section contains a description of the SBS computer program only. For the theory background, the reader is referred to chapter 4 of this thesis.

Input Data Description

Hyperspectral CASI-2 image

The hyperspectral image data may be provided in any ENVI™ supported format. However, the centre wavelength and width (Full-Width-Half-Maximum, FWHM) of the bands should be included in the image header file.

Class Region Of Interest (ROI) file

All the information classes of interest need to defined over the hyperspectral image in form of an ENVI™ Region of Interest (ROI) file. For class definition theory, the reader is referred to section 4.3.1.

Class statistics file

The SBS program expects the class statistics to be pre-calculated by an ENVI™ routine CLASSTATS (section G.3), written by the author. The latter program gives the class statistics (mean vector and covariance matrix for each class) in a format that is recognised by the program.

Data statistics file (optional)

The data statistics file is calculated with the ENVI™ routine DATASTATS (see section G.4) and only needs to be supplied when the unequal bandwidth algorithm option is chosen or bands have been pre-specified by the user. The file provides the band means to the algorithm, with which signal level comparisons are carried out.

Output Data Description

Output text file

The output text file provides a summary of all files used in the program (full pathnames), the options selected by the user and the resulting recommended band set, describing the starting and ending image row of the new band, its centre wavelength and FWHM (if the corresponding wavelength and FWHM of the image data were supplied), as well as the corresponding band set performance of the criterion measure. The latter value is compared with the performance of the entire band set, and the number of dimensions is displayed for the set that achieves at least 95% or 99% of this value first. The bands are listed in order of importance, with the first band being the most discriminant one.

Output image file (optional)

The user has the option to create an output image reflecting the band selection results. If the option is selected, an output image file is saved under the given name and opened in ENVI™'s 'Available Bands List', from where it can be displayed.

Algorithm choice

The user has the choice between three different algorithms inside the SBS program, which is displayed in figure G.1. First, a choice is made as to whether the width of the final bands (i.e. the number of rows to be merged for each band) is set equal for all bands or not. Second, for the 'equal bandwidth' case, the user can set the bandwidth to a specific value or leave it up to the program to decide on the optimum bandwidth.

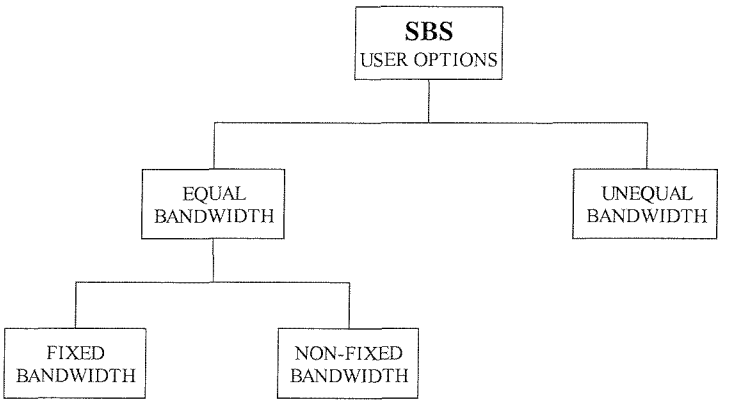


Figure G.1: Algorithm choices for the SBS (Supervised Band Selection) program.

Running the program

Definition of input and output filenames

First the user is asked to specify the filenames of

- the hyperspectral image file,
- the class ROI file,
- the class statistics file, and
- the output text file.

User Options

Second, the user is presented with an interactive menu of different choices (see figure G.2).

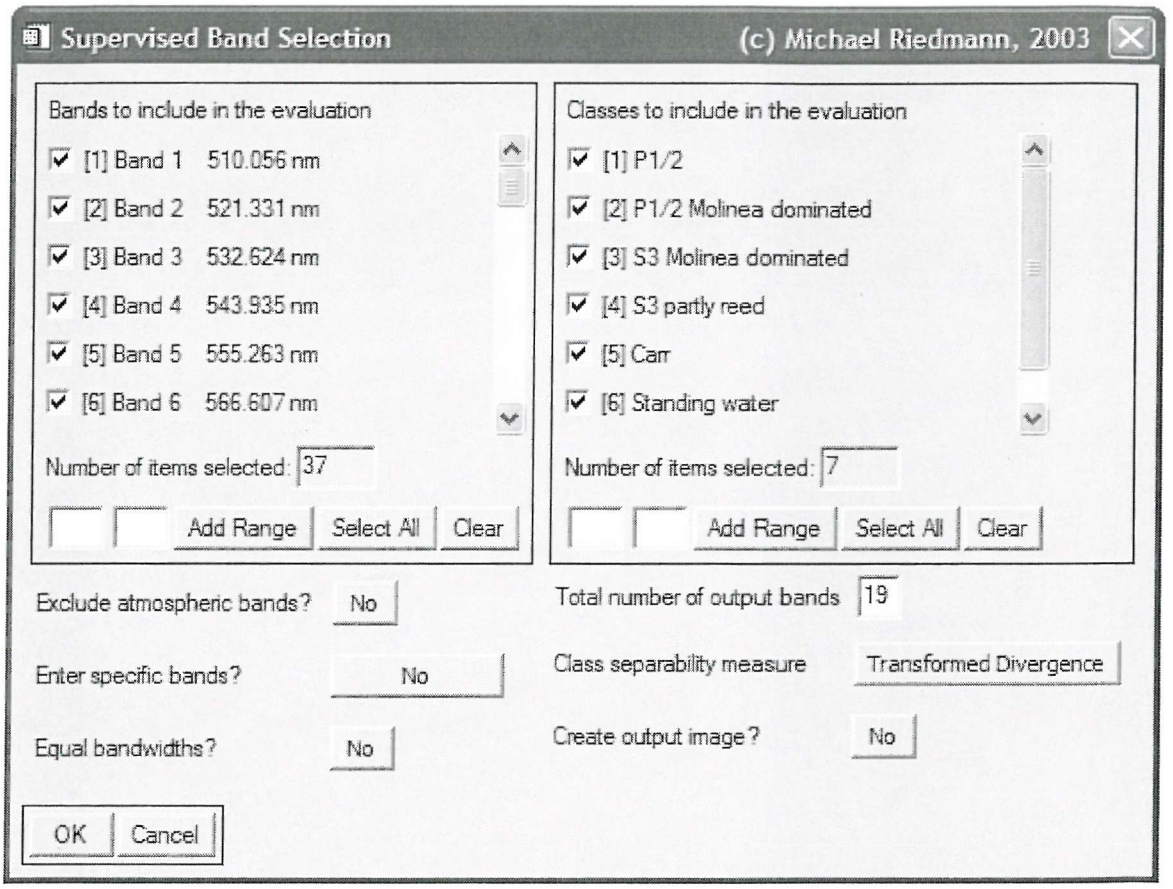


Figure G.2: ENVI™ widget allowing the user to choose between the given SBS program options.

The user is given the option to

- exclude specific bands from the evaluation (for example bands that are perceived as noisy),
- exclude atmospheric bands (default setting: 0.92-0.97, 1.1-1.17, 1.34-1.5, 1.77-2, and greater than 2.4 µm),
- decide whether or not specific bands (maximum 19) should be included in the final band set (for example mineral absorption bands), and if so whether they should be entered by wavelength or row number (note that for the former, the centre wavelength value of each band needs to be specified in the image header file),

- set equal widths for all bands (if this option is not selected rows may be merged under the constraint that the row merger improves the MLC accuracy of the band set),
- exclude specific classes from the evaluation (for example background classes),
- specify the maximum number of output bands (i.e. the dimension of the final band set; the default value is set to 10),
- create an output image (note that this part may increase the running time of the program considerably for large data sets).

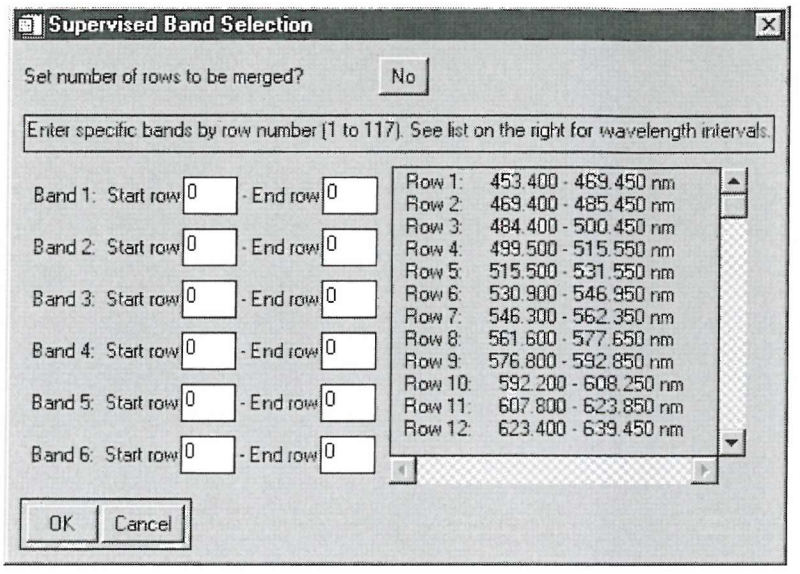


Figure G.3: ENVI™ widget for the option-dependent definition of further parameters.

Option-dependent definition of further parameters

Thirdly, depending on which set of options has been specified, the user may be asked to

- set the number of rows to be merged if the 'equal bandwidths' option was selected (if no fixed bandwidth is selected, the program is allowed to expand bands where possible but returns a set of bands of equal width; this option does not apply to user-specified bands; see figure G.3),
- set the maximum number of rows to be merged if the 'equal bandwidths' option was not selected (the default value is set to 16 rows),
- set the minimum relative band response if the 'equal bandwidths' option was not selected (the default value is set to 10% higher as the lower limit given by ITRES Research Ltd.),
- enter specific bands either by row number or wavelength, according to the option chosen before (a list of the available rows and their wavelength interval is displayed for information; specified bands should not overlap, and entered values should fall within the allowed range; if no values are specified the algorithm ignores this option; see figure G.3),
- specify the filename of the data statistics file (if unequal bandwidth option was selected or bands have been pre-specified), and

- give the name of the data statistics and output image files.

Program run

Finally the program executes displaying its current status within a status window in the middle of the screen. After program completion, a message window indicating successful program performance appears (figure G.4).

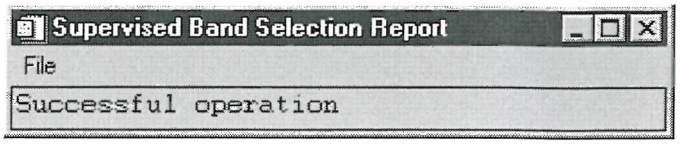


Figure G.4: ENVI™ widget report about successful program operation.

G.2 Description of the Program UBS (Unsupervised Band Selection)

Overview

A computer program called UBS (Unsupervised Band Selection) was developed to estimate a band set optimal in band number, width and location. It uses hyperspectral data from airborne and spaceborne sensors collected over a representative part of the scene. Bands are considered optimal when they are least correlated amongst each other and have a high Signal-to-Noise Ratio (SNR). This section contains a description of the UBS computer program only. For the theory background, the reader is referred to chapter 5 of this thesis.

Input Data Description

Hyperspectral image file

The hyperspectral image data may be provided in any ENVI™ supported format. However, the centre wavelength and width (Full-Width-Half-Maximum, FWHM) of the bands should be included in the image header file to allow for the consideration of atmospheric windows and the proper expansion of bands.

Data statistics file

The UBS program expects the data statistics to be pre-calculated by the ENVI™ routine DATASTATS, written by the author. The latter program gives the data statistics (band mean vector and covariance matrix) in a format that is recognised by the UBS program (see section G.4).

Dark image file (optional)

Dark current image data may be used to estimate the noise statistics for the SNR calculation. For most sensors, a dark image is collected routinely as part of both the data collection process and the instrument calibration. The dark current allows approximating instrument-induced noise only.

Region of interest for SNR calculation (optional)

If dark current image data is not available for the data set, the SNR may be estimated from an ENVI™ Region of Interest (ROI) file defined by the user over a spatially and spectrally homogeneous area of the hyperspectral image.

Output Data Description

Output text file

The output text file provides a summary of all files used in the program (full pathnames), the options selected by the user (if different from default) and the resulting recommended band set, describing the starting and ending image row of the new band, its centre wavelength and FWHM (if corresponding wavelength and FWHM of image data were supplied), as well as the corresponding band coefficient of multiple determination.

Output image file (optional)

The user has the option to create an output image reflecting the band selection results. If the option is selected, an output image file is saved under the given name and opened in ENVI™'s 'Available Bands List', from where it can be displayed.

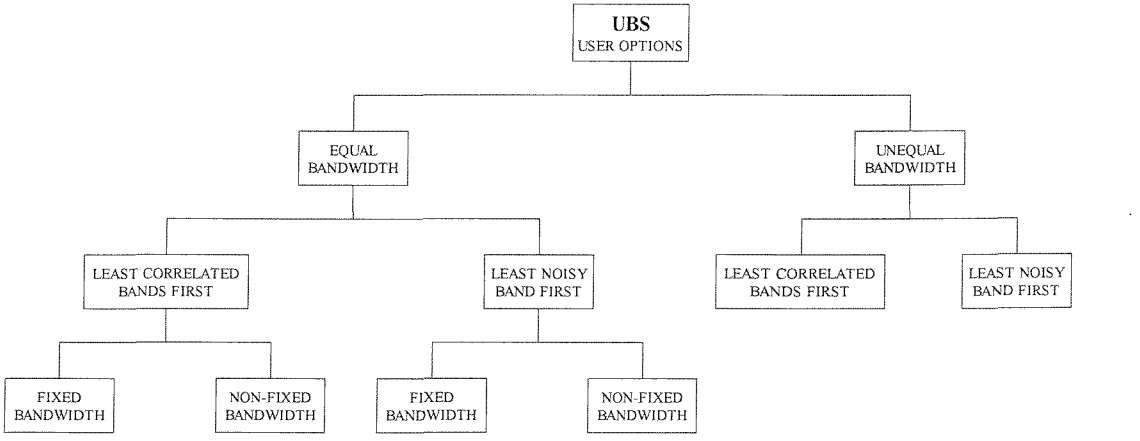


Figure G.5: Algorithm choices for the UBS (Unsupervised Band Selection) program.

Algorithm choice

The user has the choice of six different algorithms inside the UBS program. First, a choice is made as to whether the width of the final bands (i.e. the number of rows to be merged for each band) is set equal to all bands or not. Second, the option is given to start the final band set with either the least correlated pair of bands or the least noisy band. Third, for the 'equal bandwidth' case, the user can set the bandwidth to a specific value or leave it up to the program to decide on the optimum bandwidth. This choice of algorithms is displayed in figure G.5. Note that no noise estimation is required for the first algorithm, i.e. when all bands are chosen to be of a specified width and the set is to start with the least correlated band pair.

Running the program

Definition of input and output filenames

First the user is asked to specify the filenames of

- the hyperspectral image file,
- the data statistics file, and
- the output text file

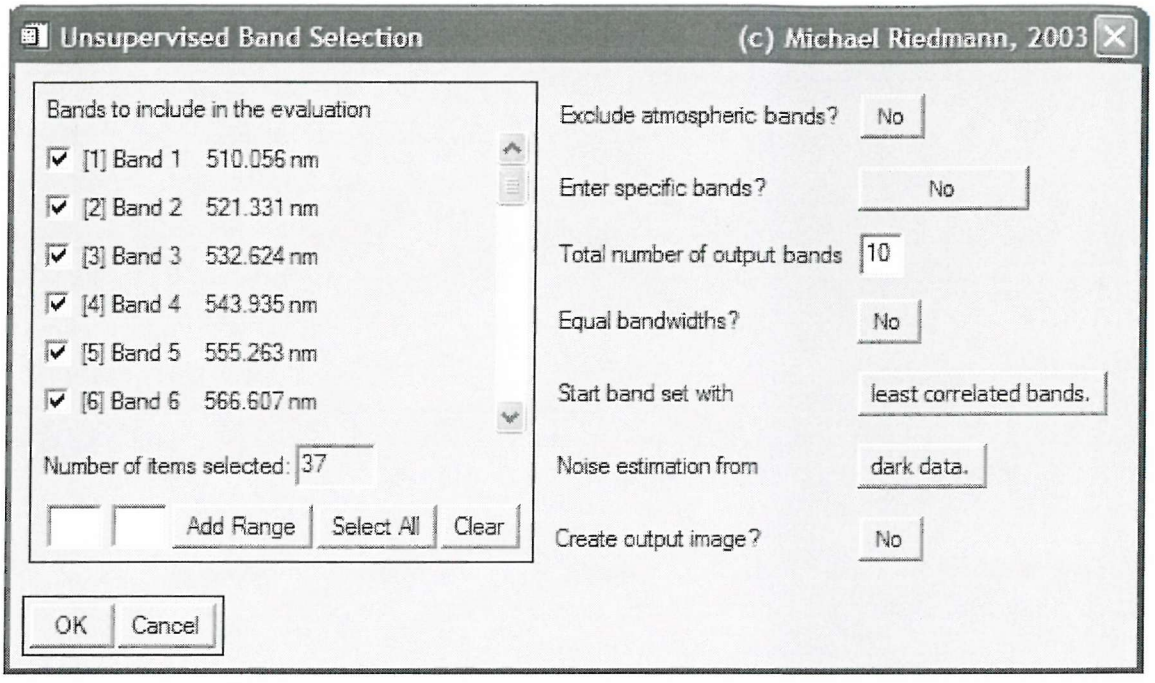


Figure G.6: ENVI™ widget allowing the user to choose between the given UBS program options.

User Options

Second, the user is presented with an interactive menu of different choices (see figure G.6). The user is given the option to

- exclude specific bands from the evaluation (for example bands that are perceived as noisy),
- exclude atmospheric bands (default setting: 0.92-0.97, 1.1-1.17, 1.34-1.5, 1.77-2, and greater than 2.4 μm),
- decide whether of not specific bands (maximum 6) should be included in the final band set (for example mineral absorption bands), and if so whether they should be entered by wavelength or row number (note that for the former, the centre wavelength value of each band needs to be specified in the image header file),
- specify the maximum number of output bands (i.e. the dimension of the final band set; the default value is set to 10),
- set equal widths for all bands (if this option is not selected rows may be merged under the constraint that the row merger improves the band SNR),
- select the option to start the band set with the least noisy band or the least correlated band pair,

- choose to estimate the noise statistics from a dark image file or an ROI file defined over a spatially and spectrally homogeneous image area, and
- create an output image (note that this part may increase the running time of the program considerably if a for large data sets).

Option-dependent definition of further parameters

Thirdly, depending on which set of options has been specified, the user may be asked to

- set the number of rows to be merged if the 'equal bandwidths' option was selected (if no fixed bandwidth is selected, the program is allowed to expand bands where possible but returns a set of bands of equal width; this option does not apply to user-specified bands; similar to figure G.3),
- set the maximum number of rows to be merged if the 'equal bandwidths' option was not selected (the default value is set to 2 rows),
- set the minimum relative band response if the 'equal bandwidths' option was not selected (the default value is set to 30%, 10% higher as the lower limit given by ITRES Research Ltd.),
- enter specific bands either by row number or wavelength, according to the option chosen before (a list of the available rows and their wavelength interval is displayed for information; specified bands should not overlap, and entered values should fall within the allowed range; if no values are specified the algorithm ignores this option; similar to figure G.3),
- specify the dark image file or the ROI file for SNR estimation, and
- give the name of the output image file.

Program run

Finally the program executes displaying its current status within a status window in the middle of the screen. After program completion, a message window indicating successful program performance appears (similar to figures G.4).

G.3 Description of the Program CLASSTATS

Overview

A computer program called CLASSTATS was developed to calculate the mean vector and covariance matrix of each class defined within an ENVI™ Region of Interest (ROI) file. It assumes the ROI file to be specified over available hyperspectral data collected over a representative part of the scene.

Input Data Description

Hyperspectral image file

The hyperspectral image data may be provided in any ENVI™ supported format. However, the centre wavelength and width (Full-Width-Half-Maximum, FWHM) of the bands should be included in the image header file to allow for the consideration of atmospheric windows and the proper expansion of bands.

Class ROI file

All the information classes of interest need to defined over the hyperspectral image in form of an ENVI™ Region of Interest (ROI) file. For class definition theory, the reader is referred to section 4.3.1.

Output Data Description

Output text file

The output text file provides a summary of the computed statistics (mean, variance, skewness, kurtosis, covariance matrix) for each class (ROI), together with all files used in the program (full pathnames). The output format is recognisable by the SBS program (see section G.1).

Running the program

First the user is asked to specify the filenames of

- the hyperspectral image file,
- the class ROI file, and
- the output text file.

Then, the program executes displaying its current status within a status window in the middle of the screen. After program completion, a message window appears indicating successful program performance (similar to figure G.4).

G.4 Description of the Program DATASTATS

Overview

A computer program called DATASTATS was developed to calculate the band mean vector and covariance matrix for the available hyperspectral data collected over a representative part of the scene.

Input Data Description

Image file

The hyperspectral image data may be provided in any ENVI™ supported format. However, the centre wavelength and width (Full-Width-Half-Maximum, FWHM) of the bands should be included in the image header file to allow for the consideration of atmospheric windows and the proper expansion of bands.

Mask file

The ENVI™ image file selection dialog allows for the specification of a mask band. It is important to define the mask file if existent, as otherwise false data statistics will result. Note that a mask band is necessary if the flight lines do not fill the entire image display window. That is, all non-image areas need to be masked out.

Output Data Description

Output text file

The output text file provides a summary of the computed statistics (band covariance and correlation matrix, eigenvalues and eigenvector matrix, as well as the band mean vector) for the data set. The output format is recognisable by the SBS and UBS program (see sections G.1 and G.2).

Running the program

First the user is asked to specify the filenames of

- the image file together with a binary mask band (if existent), and
- the output text file.

Then, the program executes displaying its current status within a status window in the middle of the screen. After program completion, a message window appears indicating successful program performance (similar to figure G.4).

G.5 Description of the Program CLASSTATSCLUSTER

Overview

The computer program CLASSTATSCLUSTER calculates the mean vector and covariance matrix of each class, defined by the ENVI™ K-Means clustering routine.

Input Data Description

Hyperspectral image file

The hyperspectral image data may be provided in any ENVI™ supported format.

Mask file

The ENVI™ image file selection dialog allows for the specification of a mask band. Note that a mask band is necessary if the flight lines do not fill the entire image display window. That is, all non-image areas need to be masked out.

Output Data Description

Output class ROI file

All the information classes of interest will be defined over the hyperspectral image in form of an ENVI™ Region of Interest (ROI) file.

Output text file

The output text file provides a summary of the computed statistics (mean, variance, skewness, kurtosis, covariance matrix) for each class (ROI), together with all files used in the program (full pathnames). The output format is recognisable by the SBS program (see section G.1).

Running the program

First, the user is asked to specify the filenames of

- the hyperspectral image file,
- the output class ROI file, and
- the output text file.

Second, the following parameters need to be provided for the K-Means clustering procedure:

- the number of output classes, that is the number of clusters (2-10) to be defined, and
- the number of iterations (1-4).

Then, the clustering program executes displaying its current status within a status window in the middle of the screen. An image of the clustering map is automatically opened in ENVI™'s 'Available Bands List', from where it can be displayed. The statistics of each class are then successively calculated. After program completion, a message window appears indicating successful program performance (similar to figure G.4).

G.6 Description of the program DATASTATSROI

Overview

This program is similar to the DATASTATS program described in section G.4, but allows the data statistics to be calculated from an ENVI™ Region of Interest (ROI) rather than from the entire image.

Input Data Description

Image file

The hyperspectral image data may be provided in any ENVI™ supported format.

ROI file

The ENVI™ Region of Interest (ROI) file should be defined over the image area from which data statistics are to be calculated. Non-image pixels should not be included.

Output Data Description

Output text file

The output text file provides a summary of the computed statistics (band covariance and correlation matrix, eigenvalues and eigenvector matrix, as well as the band mean vector) for the data set. The output format is recognisable by the SBS and UBS program (see sections G.1 and G.2).

Running the program

First the user is asked to specify the filenames of

- the image file,
- the ROI file, and
- the output text file.

Then, the program executes displaying its current status within a status window in the middle of the screen. After program completion, a message window appears indicating successful program performance (similar to figure G.4).

Glossary

Variables and constants are listed in order of their appearance in the thesis. Once defined, they are not defined a second time unless their meaning has changed.

Chapter 2

N	Number of bands	[-]
j, l	Band indices	[-]
x_j	Pixel measurement in band j	[DN]
$\mathbf{x}$	Pixel vector	[DN]
M	Number of classes	[-]
k, o	Class indices	[-]
ω_k	Class k	[-]
$\mathrm{p}(\mathbf{x} \mid \omega_k)$	Class-conditional density function for pixel vector $\mathbf{x}$	[-]
$\mathrm{P}(\omega_k)$	*A priori* probability of class k	[-]
m_j	Sample mean of band j	[DN]
$\mathbf{m}$	Sample mean vector	[DN]
n	Number of pixel measurements	[-]
i	Pixel index	
x_{ij}	Pixel measurement i for band j	[DN]
s_{jj}^2, s^2	Sample variance for band j	[DN^2]
s_{jl}	Sample covariance for bands j and l	[DN^2]
$\mathbf{S}$	Sample covariance matrix	[DN^2]
r_{jl}	Sample correlation coefficient for bands j and l	[-]
cv_j	Coefficient of variation for band j	[%]
$skewness_j$	Skewness for band j	[-]
$kurtosis_j$	Kurtosis for band j	[-]
PN_j	Priority number for band j	[-]
RMR_j	Relative mean ratio for band j	[-]
RVR_j	Relative variance ratio for band j	[-]
n_{ij}	Number of test samples classified as class i while belonging to class j	[-]
P_O	Overall accuracy	[-]
N_E	Total number of pixels in the error matrix	[-]
$\hat{\kappa}$	KHAT statistics	[-]
P_C	Chance agreement	[-]

n_{i+}	Marginal total of row i of the error matrix	[-]
n_{+i}	Marginal total of column i of the error matrix	[-]
τ	Tau statistics	[-]
P_R	Prior probability	[-]
n_{xi}	A priori number of pixels belonging to class i out of N pixels	[-]
D	Number of candidate bands for feature selection	[-]
$\mathrm{d}(k,o)$, d_{ko}	Measure of distance between classes (or pixels) ω_k and ω_o	[DN]
t	Order of Minkowski distance	[-]
$\mathbf{Q}$	Positive definite matrix	[-]
α	Angle between two class vectors	[rad]
s	Chernoff parameter ranging from 0 to 1	[-]
a	Saturation value for the Transformed Divergence	[-]
b	Range value for the Transformed Divergence	[-]
d_{m}	Multi-class distance measure	[DN]
$c_{k,o}$	Cost of deciding $\mathbf{x} \in \omega_k$, when in reality $\mathbf{x} \in \omega_o$	[-]
$\mathbf{S}_{\mathrm{w}}$	Within-class scatter matrix	[DN^2]
$\mathbf{S}_{\mathrm{b}}$	Between-class scatter matrix	[DN^2]
A	Number of possible subset combinations	[-]
X_i	Feature subset	[DN]
L	Number of features added to the set in the generalised form of SFS	[-]
R	Number of features deducted from the set in the generalised form of SBS	[-]
y_j	Pixel measurement in feature j	[DN]
$\mathbf{y}$	Transformed feature vector	[DN]
F	Mapping function	[-]
$a_{j,l}$	Linear mapping coefficients	[-]
$\mathbf{A}$	Transformation matrix with elements $a_{j,l}$	[-]
λ_l	Eigenvalue for feature l	[-]
$\mathbf{W}$	Diagonal matrix of weight coefficients	[-]
ξ	Common factor vector	[DN]
$\mathbf{\Lambda}$	Matrix of factor loadings	[-]
$\boldsymbol{\varepsilon}$	Specific or unique factor vector	[DN]
D_{ko}	Distance value between pixels k and o in transformed feature space	[DN]
O	Objective function	[-]
$\mathbf{F}$	Loading factor matrix	[-]
$\mathbf{f}_l$	Column of the loading factor matrix	[-]
ρ_j	Discriminant power of band j	[-]
λ_l^B	Eigenvalue of the lth component under the broken-stick model	[-]
χ^2	Chi-square test statistic	[-]

z	Z test-statistic	[-]
C_i	Accuracy estimate of feature subset i	[-]
s_∞^2	Asymptotic sample variance	[-]
n_{01}	Number of samples misclassified by A but not by B	[-]
n_{10}	Number of samples misclassified by B but not by A	[-]

Chapter 3

N	Number of original dimensions	[-]

Chapter 4

C	Number of channels	[-]
$FWHM$	FWHM	[nm]

Chapter 5

t	Student's statistic	[-]
N	Number of observations	[-]
r_S	Spearman rank correlation coefficient	[-]
d_i	Difference between the ranks of the i th pair of item	[-]
r_S^*	Standardised form of r_S	[-]
R	Multiple correlation coefficient	[-]
R^2	Coefficient of multiple determination	[-]
Y	Band index	[-]
F	F statistic	[-]
k	Number of independent variables	[-]

Chapter 6

x_j	Value of the dependent variable in observation j	[-]
f_i	Regression coefficient for independent variable i	[-]
$\mu_{i,j}$	Value of the independent variable i in observation j	[-]
c	Number of independent variables	[-]
n	Number of observations	[-]
$\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$	Observation vector, dimension (n,1)	[-]
$\mathbf{f} = (f_0, f_1, \cdots, f_c)^T$	Vector of coefficients, dimension (c,1)	[-]
$\mathbf{M} = (\mathbf{1}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_c)$	Matrix of independent variables, dimension (n, c)	[-]
$\mathbf{1} = (1,1,\cdots,1)^T$	Unit vector, dimension here (n,1)	[-]
e_j	Error for observation j of the fit of independent variables	[-]
$\mathbf{e} = (e_1, e_2, \cdots, e_n)^T$	Vector of random errors, dimension (n,1)	[-]

LIST OF REFERENCES

Adam, P., 1993. *Saltmarsh ecology*, Cambridge University Press, Cambridge, pp 473.

Alberotanza, L., Brando, V.E., Ravagnan, G., and Zandonella, A., 1999. Hyperspectral aerial images. A valuable tool for submerged vegetation recognition in the Orbetello Lagoons, Italy. *International Journal of Remote Sensing*, **20**, 523-533.

Anderson, G.P., Wang, J., and Chetwynd, J.H., 1995. MODTRAN3: An update and recent validations against airborne high resolution interferometer measurements. *5th Annual JPL Airborne Sciences Workshop*, 5-8.

Atkinson, P.M., 1997. Scale and spatial dependence. *In:* van Gardingen, P.R., Foody, G.M., Curran, P.J., *ed. Scaling-up: from cell to landscape*, Cambridge University Press, Cambridge, UK, 35-60.

Bachmann, C.M., and Donato, T.F., 2000. An information theoretic comparison of projection pursuit and principal component features for classification of Landsat TM imagery of central Colorado. *International Journal of Remote Sensing*, **21**, 2927-2935.

Baret, F., Guyot, G., and Major, D., 1989. TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. *12th Canadian Symposium on Remote Sensing and IGARSS'89*, Vancouver, Canada, IEEE, 1355-1358.

Bartlett, D.S., Johnson, R. W., Hardisky, M. A., and Klemas, V., 1986. Assessing Impacts of Off-Nadir Observation on Remote Sensing of Vegetation: Use of the Suits Model. *International Journal of Remote Sensing*, **7**, 247-264.

Beauchemin, M., and Fung, K.B., 2001. On statistical band selection for image visualization. *Photogrammetric Engineering and Remote Sensing*, **67**, 571-574.

Ben-Dor, E., and Levin, N., 2000. Determination of surface reflectance from raw hyperspectral data without simultaneous ground data measurements: a case study of the GER 63-channel sensor data acquired over Naan, Israel. *International Journal of Remote Sensing*, **21**, 2053-2074.

Benediktsson, J.A., and Sveinsson, J.R., 1997. Feature extraction for multisource data classification with artificial neural networks. *International Journal of Remote Sensing*, **18**, 727-740.

Bishop, C.M., 1995. *Neural networks for pattern recognition*, Clarendon Press, Oxford, pp 500.

Biswas, G., Jain, A.K., and Dubes, R.C., 1981. Evaluation of projection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **3**, 701-708.

Boardman, J.W., and Goetz, A.F.H., 1991. Sedimentary facies analysis using AVIRIS data: A geophysical inverse problem. *Proceedings of the Third AVIRIS Workshop*, Jet Propulsion Laboratory, Pasadena, CA, Publication 91-28, 4-13.

Bonham-Carter, G.F., 1988. Numerical procedures and computer program for fitting an inverted Gaussian model to vegetation reflectance data. *Computers & Geosciences*, **14**, 339-356.

Borak, J.S., and Strahler, A.H., 1999. Feature selection and land cover classification of a MODIS-like data set for a semiarid environment. *International Journal of Remote Sensing*, **20**, 919-938.

Bressan, M., and Vitrià, J., 2002. Independent component analysis and naive Bayes classification. *2nd IASTED International Conference, Visualization, Imaging and Image Processing, VIIP 2002*, Malaga, Spain, IASTED, CD-ROM.

Bretschneider, T., 2002. On the deconvolution of satellite imagery. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) and the 24th Canadian Symposium on Remote Sensing*, Toronto, Canada, IEEE, Piscataway, NJ, USA, CD-ROM.

Brown, M., Gunn, S.R., and Lewis, H.G., 1999. Support vector machines for optimal classification and spectral unmixing. *Ecological Modelling*, **120**, 167-179.

Bruzzone, L., 2000. An Approach to Feature Selection and Classification of Remote Sensing Images Based on the Bayes Rule for Minimum Cost. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 429-438.

Bruzzone, L., and Serpico, S.B., 2000. A technique for feature selection in multiclass problems. *International Journal of Remote Sensing*, **21**, 549-563.

Bruzzone, L., Roli, F., and Serpico, S.B., 1995. An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 1318-1321.

Budd, J.T.C., and Milton, E.J., 1982. Remote sensing of saltmarsh vegetation in the first four proposed thematic bands. *International Journal of Remote Sensing*, **3**, 147-161.

Butler, W.L., and Hopkins, D.W., 1970. Higher derivative analysis of complex absorption spectra. *Photochemistry and Photobiology*, **12**, 439-450.

Campbell, J.B., 1996. *Introduction to remote sensing*, 2nd Edition, Taylor & Francis, London, pp 622.

Campbell, N.A., 1980. Robust procedures in multivariate analysis. I: Robust covariance estimation. *Applied Statistics*, **29**, 231-237.

Card, D.H., and Angelici, G.L., 1983. A minicomputer based software system for the selection of optimal subsets of Thematic Mapper channels. *IEEE International Geoscience and Remote Sensing Symposium*, San Francisco, CA, TP-4 6.1-6.4.

Cattell, R.B., and Vogelmann, S., 1977. A comprehensive trial of the scree and the KG criteria for determining the number of factors. *Multivariate Behavioral Research*, **12**, 289-325.

Chang, C.-I., Du, Q., Sun, T.-L. and Althouse, M.L.G., 1999. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 2631-2641.

Chavez, P.S., Berlin, G.L., and Sowers, L.B., 1982. Statistical method for selecting Landsat MSS ratios. *Journal of Applied Photographic Engineering*, **8**, 23-30.

Chen, C.-C.T., and Landgrebe, D.A., 1989. A spectral feature design system for the HIRIS/MODIS era. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, 681-686.

Chen, C.-H., 1973. *Statistical Pattern Recognition*, Hayden Book Company, Rochelle Park, New Jersey, pp 236.

Chen, H.S., 1997. *Remote sensing calibration systems: an introduction*, A. Deepak, Hampton, USA, pp 238.

Chien, Y.T., and Fu, K.S., 1967. On the generalized Karhunen-Loève expansion. *IEEE Transactions on Information Theory*, **13**, 518-520.

Clark, R.N., and Roush, T.L., 1984. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, **89**, 6329-6340.

Clark, R.N., Swayze, G., Heidebrecht, K.,Goetz, A.F.H., and Green, R.O., 1993. Comparison of methods for calibrating AVIRIS data to ground reflectance. *In:* Green, R.O., *ed. Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, Pasadena, California, JPL Publication, 35-36.

Clark, R.N., Swayze, G.A., Heidebrecht, K., Green, R.O., and Goetz, A.F.H., 1995. Calibration to surface reflectance of terrestrial imaging spectrometry data: comparison of methods. *In:* Green, R.O., *ed. Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, JPL, Pasadena, California, JPL Publication, 41-42.

Clevers, J.G.P.W., 1988. The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sensing of Environment*, **35**, 53-70.

Clevers, J.G.P.W., and Büker, C., 1991. Feasibility of the red-edge index for the detection of nitrogen deficiency. *5th International Colloquium on Physical Measurements and Signatures in Remote Sensing*, Courchevel, France, ESA, 165-168.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37-46.

Collins, W., Chang, S.H., and Kuo, J.T., 1981. *Infrared airborne spectroradiometer survey results in the western Nevada area*, Final report to NASA, Contract JPL 955832, Columbia University, Aldridge Lab. Appl. Geophysics, pp 61.

Conel, J.E., Green, R.O., Vane, G., Bruegge, C.J., and Alley, R.E., 1987. AIS-2 radiometry and a comparison of methods for the recovery of ground reflectance. *In:* Vane, G., *ed. 3rd Airborne Imaging Spectrometer Data Analysis Workshop*, Jet Propulsion Laboratory, Pasadena, CA, JPL Publication, 18-47.

Congalton, R.G., 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing Environment*, **37**, 35-46.

Craig, R.G., 1979. Autocorrelation in Landsat data. *In:* Cook, J., *ed. Thirteenth International Symposium Remote Sensing of the Environment*, Ann Arbor, Michigan, Environmental Research Institute of Michigan, University of Michigan, 1517-1524.

Crippen, R.E., 1987. The regression intersection method of adjusting image data for band ratioing. *International Journal of Remote Sensing*, **8**, 137-155.

Crippen, R.E., 1990. Calculating the vegetation index faster. *Remote Sensing of Environment*, **34**, 71-73.

Crist, E.P., and Cicone, R.C., 1984. Application of tasselled cap concept to simulated Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, **52**, 81-86.

Crowley, J.K., 1990. Techniques for AVIRIS data normalization in areas with partial vegetation cover. *In:* Green, R.O., *ed. Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, Jet Propulsion Laboratory, Pasadena, CA, JPL Publication, 192-198.

Crowley, J.K., Brickey, D.W., and Rowan, L.C., 1989. Airborne imaging spectrometer data of the Ruby Mountains, Montana: mineral discrimination using relative absorption band-depth images. *Remote Sensing of Environment*, **29**, 121-134.

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

Curran, P.J., and Dungan, J.L., 1989. Estimation of signal-to-noise: A new procedure applied to AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, 620-628.

Curran, P.J., 1994. Imaging spectrometry. *Progress in Physical Geography*, **18**, 247-266.

Curran, P.J., Milton, E.J., Atkinson, P.M., and Foody G.M., 1998. Remote Sensing: from data to understanding. *In:* Longley, P., Brooks, S., Macmillan, W. and McDonall, R., *ed. Geocomputation: a primer*, Wiley and Sons, Chichester, 33-59.

Dargie, T., 1999. *NVC survey of saltmarsh habitat in the Severn Estuary 1998*, 341, Countryside Council for Wales and English Nature, Bangor, Gwynedd, UK, pp 166.

Dawson, T.P., and Curran, P.J., 1998. A new technique for interpolating the reflectance red edge position. *International Journal of Remote Sensing*, **19**, 2133-2139.

Dekker, A.G., Malthus, T.J., Wijnen, M.M., and Seyhan, E., 1992. The effect of spectral bandwidth and positioning on the spectral signature analysis of inland waters. *Remote Sensing of Environment*, **41**, 211-225.

Demetriades-Shah, T.H., Steven, M.D., and Clark, J.A., 1990. High resolution derivative spectra in remote sensing. *Remote Sensing of Environment*, **33**, 55-64.

Devijver, P., and Kittler, J., 1982. *Pattern recognition: a statistical approach*, Prentice-Hall, London, pp 448.

Dietterich, T.G., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, **10**, 1895-1923.

Drake, B.G., 1976. Seasonal changes in reflectance and standing crop biomass in three salt marsh communities. *Plant Physiology*, **58**, 696-699.

Drury, S.A., 1993. *Image interpretation in geology*, 3rd Edition, Chapman and Hall, London, pp 283.

Duda, R.O., and Hart, P.E., 1973. *Pattern Classification and Scene Analysis*, John Wiley and Sons Inc., New York, NY, pp 482.

Dutra, L.V., and Huber, R., 1999. Feature extraction and selection for ERS-1/2 InSAR classification. *International Journal of Remote Sensing*, **20**, 993-1016.

Dwyer, J.L., Kruse, F.A., and Lefkoff, A.B., 1995. Effects of empirical versus model-based reflectance calibration on automated analysis of imaging spectrometer data: a case study from the Drum mountains, Utah. *Photogrammetric Engineering and Remote Sensing*, **61**, 1247-54.

Ebert, E., 1987. A pattern recognition technique for distinguishing surface and cloud types in the polar regions. *Journal of Climate and Applied Meteorology*, **26**, 1412-1427.

Edwards, A.L., 1984. *An introduction to linear regression and correlation*, Second Edition, W.H. Freeman and Company, New York, pp 206.

Elvidge, C.D., and Chen, Z., 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, **54**, 38-48.

Farrand, W.H., and Harsanyi, J.C., 1997. Mapping the distribution of mine tailings in the Coeur d'Alene river valley, Idaho, through the use of a constrained energy minimization technique. *Remote Sensing of Environment*, **59**, 64-76.

Farrand, W.H., Singer, R.B., and Merényi, E., 1994. Retrieval of apparent surface reflectance from AVIRIS data: a comparison of empirical line, radiative transfer, and spectral mixture methods. *Remote Sensing of Environment*, **47**, 311-321.

Feldman, S.C., and Taranik, J.V., 1988. Comparison of techniques for discriminating hydrothermal alteration minerals with airborne imaging spectrometer data. *Remote Sensing of Environment*, **24**, 67-83.

Ferré, L., 1995. Selection of components in principal component analysis: a comparison of methods. *Computational Statistics and Data Analysis*, **19**, 669-682.

Ferrier, G., 1995. Evaluation of apparent surface reflectance estimation methodologies. *International Journal of Remote Sensing*, **16**, 2291-2297.

Fischer, J., and Fell, F., 1999. Simulation of MERIS measurements above selected ocean waters. *International Journal of Remote Sensing*, **20**, 1787-1807.

Flury, B., 1988. *Common principal components and related multivariate models*, John Wiley & Sons, New York, pp 258.

Foley, D.H., and Sammon, J.W., 1975. An optimal set of discriminant vectors. *IEEE Transactions on Computers*, **24**, 281-289.

Foody, G.M., 1992. On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, **58**, 1459-1460.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, **80**, 185-201.

Friedman, J.H., and Tukey, J.W., 1974. A projection algorithm for exploratory data analysis. *IEEE Transactions on Computers*, **23**, 881-889.

Frontier, S., 1976. Étude de la decroissance des valeurs propres dans une analyze en composantes principales: comparison avec le modèle de baton brisé. *Journal of Experimental Marine Biology and Ecology*, **25**, 67-75.

Fu, K.S., 1982. Application of pattern recognition to remote sensing. *In:* Fu, K.S., *ed. Applications of pattern recognition*, CRC Press, Boca Raton, Florida, 65-105.

Fujimoto, N., Takahashi, Y., Moriyama, T., Shimada, M., Wakabayashi, H., Nakatani, Y., and Obayashi, S., 1989. Evaluation of SPOT HRV image data received in Japan. *IGARSS'89*, IEEE, 467-470.

Fukunaga, K., 1972. *Introduction to statistical pattern recognition*, Academic Press, London, pp 369.

Fukunaga, K., 1982. Intrinsic dimensionality extraction. *In:* Krishnaiah, P.R., and Kanal, L.N., *ed. Handbook of Statistics*, North Holland, Amsterdam, 347-360.

Gamon, J.A., Serrano, L., and Surfus, J.S., 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, **112**, 492-501.

Gao, B.-C., 1993. An operational method for estimating signal to noise ratios from data acquired with imaging spectrometers. *Remote Sensing of Environment*, **43**, 23-33.

Gao, B.-C., and Goetz, A.F.H., 1990. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *Journal of Geophysical Research*, **95**, 3549-3564.

Gao, B.C., Heidebrecht, K.B., and Goetz, A.F.H., 1992. *ATmospheric REMoval program (ATREM) users guide*, Center for the Study of Earth from Space (CSES), University of Colorado, Boulder, pp 24.

Gao, B.C., Heidebrecht, K.B., and Goetz, A.F.H., 1993. Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing of Environment*, **44**, 165-178.

Geary, R., 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, **5**, 115-145.

Goetz, A.F.H., Curtiss, B., Kruse, F.A., and Boardman, J.W., 1995. Hyperspectral imaging and data analysis. *Workshop organised by the Center for the Study of Earth from Space (CSES)*, University of Colorado, Boulder, Notes.

Goodenough, D.G., Narendra, P.M., and O'Neill, K., 1978. Feature subset selection in remote sensing. *Canadian Journal of Remote Sensing*, **4**, 143-148.

Green, A.A., and Craig, M.D., 1985. Analysis of aircraft spectrometer data with logarithmic residuals. *In:* Vane, G., and Goetz, A., *ed. Airborne Imaging Spectrometer (AIS) Workshop*, Jet Propulsion Laboratory, Pasadena, California, JPL Publication, 111-119.

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

Gregory, R.L., 1977. *Eye and brain - the psychology of seeing*, Weidenfeld and Nicolson, London, pp 256.

Gross, M.F., Hardisky, M.A., and Klemas, V., 1988. Effects of Solar Angle on Reflectance from Wetland Vegetation. *Remote Sensing of Environment*, **26**, 195-212.

Gruninger, J., Sundberg, R., Fox, M., Levine, R., Mundkowsky, W., Salisbury, M.S., and Ratcliff, A.H., 2001. Automated optimal channel selection for spectral imaging sensors. *Proceedings of Algorithms for Multispectral, Hyperspectral and Ultraspectral Imagery VII*, Orlando, FL, SPIE, CD-ROM.

Guttman, L., 1954. Some necessary conditions for commonn factor analysis. *Psychometrika*, **19**, 149-161.

Guyot, G., and Baret, F., 1988. Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. *4th International Colloquium on Spectral Signatures of Objects in Remote Sensing*, Aussois, France, ESA, 279-286.

Hardisky, M.A., Smart, R.M. and Klemas, V., 1983. Seasonal spectral characteristics and aboveground biomass of the tidal marsh plant, *Spartina alterniflora*. *Photogrammetric Engineering and Remote Sensing*, **49**, 85-92.

Harman, H.H., 1976. *Modern Factor Analysis*, 3rd Edition, University of Chicago, Chicago, pp 488.

Harsanyi, J.C., and Chang, C.-I., 1994. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 779-785.

Held, A., Ticehurst, C., Lymburner, L., and Williams, N., 2003. High resolution mapping of tropical mangrove ecosystems using hyperspectral and radar remote sensing. *International Journal of Remote Sensing*, **24**, 2739-2759.

Henderson, T.L., Szilagyi, A., Baumgardner, M.F., Chen, C.T., and Landgrebe, D.A., 1989. Spectral band selection for classification of soil organic matter content. *Soil Science Society of America Journal*, **53**, 1778-1784.

Hinton, P.R., 1995. *Statistics explained - a guide for social science students*, Routledge, London, pp 322.

Hoel, P.G., 1984. *Introduction to mathematical statistics*, John Wiley & Sons, New York, pp 435.

Hogg, R.V., and Craig, A.T., 1978. *Introduction to mathematical statistics*, 4 Edition, MacMillan Publishing Co., Inc., New York, pp 438.

Holden, H., and LeDrew, E., 1998. Spectral discrimination of healthy and non-healthy corals based on cluster analysis, principal components analysis, and derivative spectroscopy. *Remote Sensing of Environment*, **65**, 217-224.

Holden, H., and Ledrew, E., 1999. Hyperspectral identification of coral reef features. *International Journal of Remote Sensing*, **20**, 2545-2563.

Holland, J.H., 1975. *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor.

Horler, D.N.H., Dockray, M. and Barber, J., 1983. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, **4**, 273-288.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30**, 179-185.

Horwitz, H.M., Nalepka, R.F., Hyde, P.D., and Morgenstern, J.P., 1971. Estimating the proportions of objects within a single resolution element of a multispectral scanner. *7th International Symposium on Remote Sensing of Environment*, Environmental Research Inst. of Michigan (ERIM), 1307-1320.

Hsieh, P.-F., and Landgrebe, D.A., 1998. Linear feature extraction for multiclass problems. *IEEE International Geoscience and Remote Sensing Symposium*, Seattle, WA, IEEE, CD-ROM.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195-213.

Huete, A.R., 1988. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, **25**, 295-309.

Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognition. *IEEE Transactions on Information Theory*, **14**, 55-63.

Huguenin, R.L., and Jones, J.L., 1986. Intelligent information extraction from reflectance spectra: absorption band position. *Journal of Geophysical Research*, **91**, 9585-9598.

Hunt, G.R., 1977. Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics*, **42**, 501-513.

Hunt, G.R., 1979. Near-infrared (1.3-2.4μm) spectra of alteration minerals- potential for use in remote sensing. *Geophysics*, **44**, 1974-1986.

Hunt, G.R., 1980. Electromagnetic radiation: the communication link in remote sensing. *In:* Siegal, B., and Gillespie, A., *ed. Remote Sensing in Geology*, Wiley, New York, 702.

Hunt, G.R. and Ashley, R.P., 1979. Spectra of altered rocks in the visible and near infrared. *Economic Geology*, **74**, 1613-1629.

Hutsinpiller, A., 1988. Discrimination of hydrothermal alteration mineral assemblages at Virginia City, Nevada, using the airborne imaging spectrometer. *Remote Sensing of Environment*, **24**, 53-66.

Ifarraguerri, A., and Chang, C.-I, 2000. Unsupervised Hyperspectral Image Analysis with Projection Pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 2529-2538.

Isaacs, R.G., and Vogelmann, A.M., 1988. Multispectral sensor data simulation modeling based on the multiple scattering LOWTRAN code. *Remote Sensing of Environment*, **26**, 75-99.

ITRES, 2001. *CASI-2 manual*, ITRES Research Ltd., PDF Files.

Jackson, D.A., 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, **74**, 2204-2214.

Jackson, R.D., 1983. Spectral indices in n-space. *Remote Sensing of Environment*, **13**, 409-421.

Jensen, J.R., 1986. *Introductory digital image processing - a remote sensing perspective*, Prentice-Hall, Englewood Cliffs, NJ, pp 379.

Jensen, J.R., 2000. *Remote Sensing of the Environment: an Earth resource perspective*, Prentice-Hall, London, pp 544.

Jia, X., and Richards, J.A., 1998. Progressive two-class decision classifier for optimization of class discriminations. *Remote Sensing of Environment*, **63**, 289-297.

Jia, X., and Richards, J.A., 1999. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 539-542.

Jia, X., and Richards, J.A., 2002. Cluster-space representation for hyperspectral data classification. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 593-598.

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

Jones, M.C., and Sibson, R., 1987. What is Projection Pursuit. *Journal of the Royal Statistical Society Series*, **150**, 1-36.

Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, **50**, 663-666.

Justice, C.O., Markham, B.L., Townshend, J.R.G., and Kennard, R.L., 1989. Spatial degradation of satellite data. *International Journal of Remote Sensing*, **10**, 1539-1561.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20**, 141-151.

Kalayeh, H.M., and Landgrebe, D.A., 1983. Predicting the required number of training samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**, 664-667.

Kamp, U., Olbert, C., Schaale, M., and Fischer, J., 1997. Umweltmonitoring durch flugzeuggestuetzte Fernerkundung und neuronale Netze. *Geooekodynamik*, **18**, 41-60.

Karhunen, J., Oja, E., Wang, L.Y., Vigario, R., and Joutsensalo, J., 1997. A class of neural networks for independent component analysis. *IEEE Transactions on Neural Networks*, **8**, 486-504.

Karlholm, J., and Renhorn, I., 2002. Wavelength band selection method for multispectral target detection. *Applied Optics*, **41**, 6786-6795.

Kaufman, Y.J., and Tanré, D., 1992. Atmospherically resistant vegetation index - ARVI for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 261-270.

Kavzoglu, T., and Mather, P.M., 2000. The Use of Feature Selection Techniques in the Context of Artificial Neural Networks. *RSS 2000*, Leicester, CD-ROM.

Kerekes, J.P., and Landgrebe, D.A., 1989. Simulation of optical remote sensing systems. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, 762-771.

Kirkpatrick, S., Gelatt Jr., C.D., and Vecchi, M.P., 1983. Optimization by simulated annealing. *Science*, **220**, 671-680.

Kittler, J., 1986. Feature Selection and Extraction. *In:* Young, T.Y., and Fu, K.-S., *ed. Handbook of pattern recognition and image processing*, Academic Press Inc. Ltd., London, 60-84.

Kittler, J., and Young, P.C., 1973. A new approach to feature selection based on the Karhunen-Loève expansion. *Pattern Recognition*, **5**, 335-352.

Klecka, W.R., 1980. *Discriminant Analysis*, SAGE Publication, Inc., Beverly Hills, California, pp 71.

Kneizys, F.X., Shettle, E.P., Anderson, G.P., Abrew, L.W., Chetwynd, J.H., Shelby, J.E.A., and Gallery, W.O., 1989. *Atmospheric Transmittance/ Radiance; Computer Code LOWTRAN 7*, AFGL Hanscom, AFB, Bedford, MA.

Kruse, F.A., 1988. Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the Northern Grapevine Mountains, Nevada, and California. *Remote Sensing of Environment*, **24**, 31-51.

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

Kumar, L., Schmidt, K., Dury, S., and Skidmore, A., 2001. Imaging spectrometry and vegetation science. *In:* Van der Meer, F.D., and De Jong, S.M., *ed. Imaging Spectrometry - Basic principles and prospective applications*, Kluwer Academic Publishers, Dordrecht, 111-155.

Labovitz, M.L., 1986. Issues arising from sampling designs and band selection in discriminating ground reference attributes using remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, **52**, 201-211.

Labovitz, M.L., and Masuoka, E.J., 1984. The influence of autocorrelation in signature extraction - an example from a geobotanical investigation of Cotter Basin, Montana. *International Journal of Remote Sensing*, **5**, 315-332.

Landgrebe, D.A., 1999. On information extraction principles for hyperspectral data. *4th International Conference on GeoComputation*, Fredericksburg, Virginia, USA, Mary Washington College.

Landgrebe, D.A., 2000. Information extraction principles and methods for multispectral and hyperspectral image data. *In:* Chen, C.-H., *ed. Information Processing for Remote Sensing*, World Scientific Publishing Co., River Edge, NJ, Chapter 1.

Leckie, D.G., Teillet, P.M., Ostaff, D.P., and Fedosejevs, G., 1988. Sensor band selection for detecting current defoliation caused by the spruce budworm. *Remote Sensing of Environment*, **26**, 31-50.

Lee, C., and Landgrebe, D.A., 1993. Feature extraction based on decision boundaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 388-400.

Lee, J.B., Woodyatt, A.S. and Berman, M., 1990. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 295-304.

Lévesque, J., Szeredi, T., Staenz, K., and Singhroy, V., 1998. Spectral band selection from CASI data for monitoring mine tailings site rehabilitation. *Proceedings of the 20th Canadian Symposium on Remote Sensing*, Calgary, Canada, 161-164.

Lillesand, T.M., and Kiefer, R.W., 1987. *Remote sensing and image interpretation*, 2nd edition Edition, John Wiley & Sons, New York, pp 721.

Lin, C., and Chang, C.-I., 2001. Dependence of land use and spectral information in the aspect of band selection of seashore and urban hyperspectral data. *Fifth International Airborne Remote Sensing Conference*, San Francisco, Veridian, USA, CD-ROM.

Liu, S.S., and Jernigan, M.E., 1990. Texture analysis and discrimination in additive noise. *Computer Vision and Image Processing*, **49**, 52-67.

Long, S.P., and Mason, C.F., 1983. *Saltmarsh ecology*, Blackie & Son Ltd., Glasgow, pp 160.

Lyon, J.G., Yuan, D., Lunetta, R.S., and Elvidge, C.D., 1998. A change detection experiment using vegetation indices. *Photogrammetric Engineering and Remote Sensing*, **62**, 143-150.

Ma, Z., and Redmond, R.L., 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, **61**, 435-439.

Manly, B.F.J., 1994. *Multivariate Statistical Methods - A primer*, 2nd Edition, Chapman & Hall, London, pp 225.

Marceau, D.J., Howarth, P.J., and Gratton, D.J., 1994. Remote Sensing and the Measurement of Geographical Entities in a Forested Environment. 1. The Scale and Spatial Aggregation Problem. *Remote Sensing Environment*, **49**, 93-104.

Marsh, S.E., and McKeon, J.B., 1983. Integrated analysis of high-resolution field and airborne spectroradiometer data for alteration mapping. *Economic Geology*, **78**, 618-632.

Mather, P.M., 1976. *Computational methods of multivariate analysis in physical geography*, John Wiley & Sons, London, pp 532.

Mather, P.M., 1987. Preprocessing of training data for multispectral image classification. *13th Annual Conference of the Remote Sensing Society*, University of Nottingham, Remote Sensing Society, 111-120.

Mather, P.M., 1999. *Computer processing of remotely-sensed images: an introduction*, 2nd Edition, John Wiley & Sons, Chichester, pp 292.

Mausel, P.W., Kramber, W.J., and Lee, J.K., 1990. Optimum band selection for supervised classification of multispectral data. *Photogrammetric Engineering and Remote Sensing*, **56**, 55-60.

Mazer, A.S., Martin, M., Lee, M., and Solomon, J.E., 1988. Image processing software for imaging spectrometry data analysis. *Remote Sensing of Environment*, **24**, 201-211.

Merembeck, B.F., and Turner, B.J., 1980. Directed canonical analysis and the performance of classifiers under its associated linear transformation. *IEEE Transactions on Geoscience and Remote Sensing*, **18**, 191-196.

Merton, R.N., 1998. Monitoring community hysteresis using spectral shift analysis and the Red-edge Vegetation Stress Index. *Seventh Annual JPL Airborne Earth Science Workshop*, Pasadena, California, USA, NASA, Jet Propulsion Laboratory.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087-1092.

Miller, A.J., 1990. *Subset selection in regression*, Chapman and Hall, London, pp 229.

Miller, I., and Freund, J.E., 1965. *Probability and statistics for engineers*, Prentice-Hall, Englewood Cliffs, New Jersey, pp 432.

Milton, E.J., 1999. Image endmembers and the scene model. *Canadian Journal of Remote Sensing*, **25**, 112-120.

Milton, E.J., Hughes, P.D., Anderson, K., Schulz, J., Lindsay, R., Kelday, S.B., and Hill, C.T., 2003. *Bog Surfaces - Remote Sensing: a report to English Nature*, EN-EIT33-01-006, GeoData Institute, University of Southampton, Southampton, pp 97.

Mitchell, M., 1996. *An introduction to genetic algorithms*, The Massachusetts Institute of Technology (MIT) Press, Cambridge, Massachusetts, pp 210.

Narendra, P.M., and Fukunaga, K., 1977. A Branch and Bound algorithm for feature subset selection. *IEEE Transactions on Computers*, **26**, 917-921.

Nielsen, A.A., 1994. *Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data*. Ph.D. Institute of Mathematical Modelling, Technical University of Denmark, pp 189.

Nielsen, A.A., 1999. Linear mixture models, full and partial unmixing in multi- and hyperspectral image data. *Öresyn'99, Öresund vision workshop*, Lund University, Lund, Sweden,

Nirala, M.L., and Venkatachalam, G., 2000. Rotational transformation of remotely sensed data for land use classification. *International Journal of Remote Sensing*, **21**, 2185-2202.

Okada, K., and Iwashita, A., 1992. Hyper-multispectral image analysis based on waveform characteristics of spectral curve. *Advances in Space Research*, **12**, (7)433-(7)442.

O'Neill, M.A., and Dowman, I.J., 1993. A simulation study of the ASTER sensor using a versatile general purpose rigid sensor modelling system. *International Journal of Remote Sensing*, **14**, 565-582.

Owen, D.B., 1962. *Handbook of statistical tables*, Addison-Wesley Publishing Company, London, pp 580.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, **2**, 559-572.

Penaloza, M.A., and Welch, R.M., 1996. Feature selection for classification of polar regions using a fuzzy expert system. *Remote Sensing of Environment*, **58**, 81-100.

Perry, E.M., Warner, T., and Foote, P., 2000. Comparison of atmospheric modelling versus empirical line fitting for mosaicking HYDICE imagery. *International Journal of Remote Sensing*, **21**, 799-803.

Petrakos, M., Di Carlo, W., and Kanellopoulos, I., 1999. Dimensionality reduction for the visualization of fuzzy multispectral data. *25th Annual Conference and Exhibition of the Remote Sensing Society*, Swansea, Remote Sensing Society, 55-62.

Petrie, G.M., Heasler, P.G., and Warner, T.A., 1998. Optimal band selection strategies for hyperspectral data sets. *IGARSS'98, Sensing and Managing the Environment*, Seattle, WA, IEEE, 1582-1584.

Philpot, W.D., 1991. The derivative ratio algorithm: avoiding atmospheric effects in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 350-357.

Piech, M.A., and Piech, K.R., 1987. Symbolic representation of hyperspectral data. *Applied Optics*, **26**, 4018-4026.

Piech, M.A., and Piech, K.R., 1989. Hyperspectral interactions: invariance and scaling. *Applied Optics*, **28**, 481-489.

Pinty, B., and Verstraete, M. M., 1992. GEMI: A non-linear index to monitor global vegetation from satellites. *Vegetatio*, **101**, 15-20.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., 1992. *Numerical Recipes in C*, 2nd edition Edition, Cambridge University Press, Cambridge, pp 994.

Price, J.C., 1990. On the information content of soil reflectance spectra. *Remote Sensing Environment*, **33**, 113-121.

Price, J.C., 1994a. Band selection procedure for multispectral scanners. *Applied Optics*, **33**, 3281-3288.

Price, J.C., 1994b. How unique are spectral signatures? *Remote Sensing of Environment*, **49**, 181-186.

Price, J.C., 1997. Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 1277-1285.

Price, J.C., 1998. An approach for analysis of reflectance spectra. *Remote Sensing of Environment*, **64**, 316-330.

Prügel-Bennett, A., and Shapiro, J.L., 1994. An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, **72**, 1305-1309.

Pu, R., and Gong, P., 2000. Band selection from hyperspectral data for conifer species identification. *Geographic Information Sciences*, **6**, 137-142.

Pudil, P., Novovicova, J., and Kittler, J., 1994. Floating search methods in feature selection. *Pattern Recognition Letters*, **15**, 1119-1125.

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

Rao, V.R., Brach, E.J., and Mack, A.R., 1979. Effect of surface winds on the spectral signatures of crops. *Agronomy Journal*, **71**, 515-518.

Richards, J.A., and Jia, X., 1999. *Remote sensing digital image analysis: an introduction*, 3rd Edition, Springer Verlag, Berlin, pp 363.

Richardson, A.J., and Wiegand, C.L., 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, **43**, 1541-1552.

Richter, R., and Lehmann, F., 1989. MOMS-02 sensor simulation and spectral band selection. *International Journal of Remote Sensing*, **10**, 1429-1435.

Roberts, D.A., Yamaguchi, Y., and Lyon, R.J.P., 1985. Calibration of airborne imaging spectrometer data to percent reflectance using field spectral measurements. *19th International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, ISPRS, 679-688.

Roberts, D.A., Yamaguchi, Y., and Lyon, R.J.P., 1986. Comparison of various techniques for calibration of AIS data. *2nd Airborne Imaging Spectrometer (AIS) Workshop*, Jet Propulsion Laboratory, Pasadena, California, JPL Publication, 21-30.

Roger, R.E., and Arnold, J.F., 1996. Reliably estimating the noise in AVIRIS hyperspectral images. *International Journal of Remote Sensing*, **17**, 1951-1962.

Rondeaux, G., Steven, M., and Baret, F., 1996. Optimisation of soil-adjusted vegetation indices. *Remote Sensing of Environment*, **55**, 95-107.

Rubin, T.D., 1993. Spectral mapping with imaging spectrometers. *Photogrammetric Engineering and Remote Sensing*, **59**, 215-220.

San Miguel-Ayanz, J., and Biging, G.S., 1996. An iterative classification approach for mapping natural resources from satellite imagery. *International Journal of Remote Sensing*, **17**, 957-981.

Schowengerdt, R.A., 1997. *Remote Sensing Models and Methods for Image Processing*, Academic Press, London, pp 525.

Settle, J.J., and Drake, N.A., 1993. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, **14**, 1159-1177.

Shaban, M.A., and Dikshit, O., 2001. Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh. *International Journal of Remote Sensing*, **22**, 565-593.

Sheffield, C., 1985. Selecting band combinations from multispectral data. *Photogrammetric Engineering and Remote Sensing*, **51**, 681-687.

Siedlecki, W., and Sklansky, J., 1988. On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence*, **2**, 197-220.

Siedlecki, W., Siedlecka, K., and Sklansky, J., 1988. An overview of mapping techniques for exploratory pattern analysis. *Pattern Recognition*, **21**, 411-429.

Slater, P.N., 1984. The importance and attainment of accurate absolute radiometric calibration. *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, 34-40.

Smith, W.L., 1985. Satellites. *In:* Houghton, D.D., *ed. Handbook of applied meteorology*, John Wiley and Sons, New York, 380-472.

Smith, G.M., and Curran, P.J., 1999. Methods for estimating image signal-to-noise ratio (SNR). *In:* Atkinson, P.M., and Tate, N.J., *ed. Advances in remote sensing and GIS analysis*, John Wiley & Sons Ltd, Chichester, 61-74.

Smith, G.M., and Milton, E.J., 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing*, **20**, 2653-2662.

Sonka, M., Hlavac, V., and Boyle, R., 1993. *Image processing, analysis and machine vision*, International Thomson Computer Press, London, pp 555.

Spanner, M.A., Brass, J.A., and Peterson, D.L., 1984. Feature selection and the information content of Thematic Mapper simulator data for forest structural assessment. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-22**, 482-489.

Strahler, A.H., Woodcock, C.E., and Smith, J.A., 1986. On the nature of models in remote sensing. *Remote Sensing of Environment*, **20**, 121-139.

Swain, P.H., 1986. Remote Sensing. *In:* Young, T.Y., and Fu, K.-S., *ed. Handbook of pattern recognition and image processing*, Academic Press Inc. Ltd., London, 613-627.

Swain, P.H., and Davis, S.M., 1978. *Remote Sensing: The quantitative approach*, McGraw-Hill, New York, pp 396.

Switzer, P., and Green, A., 1984. *Min/max autocorrelation factors for multivariate spatial imagery*, Tech. Rep. 6, Department of Statistics, Stanford University, Stanford, CA, pp 10.

Tadjudin, S., and Landgrebe, D., 1998. *Classification of high dimensional data with limited training samples*. Ph.D. School of Electrical and Computer Engineering, Purdue University, pp 123.

Teillet, P.M., and Fedosejevs, G., 1995. On the dark target approach to atmospheric correction of remotely sensed data. *Canadian Journal of Remote Sensing*, **21**, 374-387.

Thenkabail, P.S., Smith, R.B., and De Pauw, E., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, **71**, 158-182.

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

Thenkabail, P.S., Ward, A.D., and Lyon, J.G., 1995. Landsat-5 Thematic Mapper models of soybean and corn crop characteristics. *International Journal of Remote Sensing*, **15**, 49-61.

Thomas, I.L., and Allcock, G.McK., 1984. Determining the confidence level for a classification. *Photogrammetric Engineering and Remote Sensing*, **50**, 1491-1496.

Thomas, I.L., Ching, N.P., Benning, V.M., and D'Aguanno, J.A., 1987. A review of multi-channel indices of class separability. *International Journal of Remote Sensing*, **8**, 331-350.

Thomson, A.G., Fuller, R.M., Sparks, T.H., Yates, M.G. and Eastwood, J.A., 1998a. Ground and airborne radiometry over intertidal surfaces: waveband selection for cover classification. *International Journal of Remote Sensing*, **19**, 1189-1205.

Thomson, A.G., Fuller, R.M., and Eastwood, J.A., 1998b. Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing. *International Journal of Remote Sensing*, **19**, 3423-3431.

Tompkins, S., Mustard, J.F., Pieters, C.M., and Forsyth, D.W., 1997. Optimization of Endmembers for Spectral Mixture Analysis. *Remote Sensing of Environment*, **59**, 472-489.

Townshend, J.R.G., Huang, C., Kalluri, S.N.V., Defries, R.S., and Liang, S., 2000. Beware of per-pixel characterization of land cover. *International Journal of Remote Sensing*, **21**, 839-843.

Tsai, F., and Philpot, W., 1998. Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, **66**, 41-51.

Tso, B., and Mather, P.M., 2001. *Classification methods for remotely sensed data*, Taylor & Francis, London, pp 332.

Tu, T.-M., Chen, C.-H., Wu, J.-L. and Chang, C.-I., 1998. A fast two-stage classification method for high-dimensional remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 182-191.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127-150.

Van Campenhout, J.M., 1980. The arbitrary relation between probability of error and measurement subset. *Journal of the American Statistical Association*, **75**, 104-109.

Van der Meer, F.D., and De Jong, S.M., 2001. *Imaging Spectrometry - Basic principles and prospective applications*, Kluwer Academic Publishers, Dordrecht, pp 403.

Van der Meer, F.D., De Jong, S.M., and Bakker, W., 2001. Imaging spectrometry: basic analytical techniques. *In:* Van der Meer, F.D., and De Jong, S.M., *ed. Imaging Spectrometry - Basic principles and prospective applications*, Kluwer Academic Publishers, Dordrecht, 17-61.

Vane, G., and Goetz, A.F., 1988. Terrestrial Imaging Spectroscopy. *Remote Sensing Environment*, **24**, 1-29.

Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., and Morcrette, J.-J., 1997. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An Overview. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 675-686.

Verstraete, M.M., and Pinty, B., 1992. Extracting surface properties from satellite data in the visible and near-infrared wavelengths. *In:* Mather, P.M., *ed. TERRA-1: understanding the terrestrial environment*, Taylor & Francis, London, 203-209.

Warner, T.A., and Shank, M.C., 1997. Spatial autocorrelation analysis of hyperspectral imagery for feature selection. *Remote Sensing of Environment*, **60**, 58-70.

Warner, T.A., Steinmaus, K., and Foote, H., 1999. An evaluation of spatial autocorrelation feature selection. *International Journal of Remote Sensing*, **20**, 1601-1616.

Webb, A., 1999. *Statistical Pattern Recognition*, Arnold, London, pp 448.

Wehrli, C., 1985. *Extraterrestrial solar spectrum*, July, 615, Physikalisch-Meteorologisches Observatorium and World Radiation Center (PMO/WRC), Davos Dorf, Switzerland.

Weiss, M., Baret, F., Myneni, R.B., Pragnere, A., and Knyazikhin, Y., 2000. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, **20**, 3-22.

Westerhoff, D., 1992. *The New Forest heathlands, grasslands and mires: a management review and strategy*, English Nature, Lyndhurst, pp 75.

Wetzel, M.A., 1995. Simulation of radiances for future AVHRR platforms with the AVIRIS spectral radiometer. *International Journal of Remote Sensing*, **16**, 1167-1177.

Wiersma, D.J., and Landgrebe, D.A., 1980. Analytical design of multispectral sensors. *IEEE Transactions on Geoscience and Remote Sensing*, **18**, 181-189.

Withagen, P.J., den Breejen, E., Franken, E.M., de Jong, A.N., and Winkel, H., 2001. Band selection from a hyperspectral data-cube for a real-time multispectral 3CCD camera. *In:* Shen, S.S., and Descour, M.R., *ed. SPIE Aerosense, Algorithms for Multi-, Hyper-, and Ultraspectral Imagery VII*, Orlando, FL, SPIE - The International Society for Optical Engineering, 84-93.

Witkin, A.P., 1983. Scale-space filtering. *4th International Joint Conference on Artificial Intelligence*, Karlsruhe, IEEE, 1019-1022.

Woodcock, C.E., and Strahler, A.H., 1987. The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**, 311-316.

Wu, D., and Linders, J., 2000. Comparison of three different methods to select feature for discriminating forest cover types using SAR imagery. *International Journal of Remote Sensing*, **21**, 2089-2099.

Wyatt, C.L., 1978. *Radiometric calibration: Theory and methods*, Academic Press, London, pp 200.

INTERNET REFERENCES

Analytical Imaging and Geophysics LLC, 2003. *ACORN – Atmospheric CORrection Now.* http://www.aigllc.com/acorn/intro.asp (12 Mar. 2003).

Clark, R.N., Swayze, G.A., Livo, K.E., Kokaly, R.F., King, T.V.V., Dalton, J.B., Vance, J.S., Rockwell, B.W., Hoefen, T., and McDougal, R.R., 2002. *Surface reflectance calibration of terrestrial imaging spectroscopy data: a tutorial using AVIRIS.* http://speclab.cr.usgs.gov/PAPERS.calibration.tutorial/ (12 Mar. 2003).

ESA, 2002a. Top of Atmosphere Vegetation Index, *MERIS Level 2 algorithms.* http://envisat.esa.int/dataproducts/meris/ (12 Mar. 2003).

ESA, 2002b. MERIS Level 1b data products, *MERIS data products.* http://envisat.esa.int/instruments/meris/data-app/dataprod.html (12 Mar. 2003).

Filliben, J.J., and Heckert, A., 2002. *Dataplot Homepage.* http://www.itl.nist.gov/div898/software/dataplot (22 Sep. 2003).

JNCC, 2002. Introduction to International Designations, *Protected Sites*, http://www.jncc.gov.uk/idt (13 Aug. 2002).

NERC ARSF, 2002. *CASI-2 sensor user manual version 2.0.*

http://www.nerc.ac.uk/arsf/Contents/Instruments/PDF/casi2_v03.PDF (12 Mar. 2003).

NIST/SEMATECH, 2003. *E-Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook/ (23 Apr. 2003).

NPL, 2003. Questionnaire on data for GMES, http://www.npl.co.uk/optical_radiation/gmessurvey (24 Sep. 2003).

StatSoft Inc, 2002. *Electronic Statistics Textbook*. http://www.statsoft.com/textbook/stathome.html (24 Apr. 2003).

Weisstein, E.W., 2003. Covariance, *Eric Weisstein's World of Mathematics* (*MathWorld*™), http://mathworld.wolfram.com/Covariance.html (22 Sep. 2003).

AUTHOR'S REFERENCES

As part of my investigations which have lead to the thesis I have also published the following papers:

Anderson, K., Milton, E.J., Dargie, T., Riedmann, M., and Hill, C.T., 2000. *Evaluation of CASI and digital salt marsh survey information: a report to English Nature*, F14/01/592, GeoData Institute, University of Southampton, Southampton, pp 92.

Milton, E.J., Riedmann, M., and Lawrance, D.J., 2001. A novel scene recording spectroradiometer. *8th International Symposium Physical Measurements and Signatures in Remote Sensing*, Aussois, France, CNES, Toulouse, France, 401-406.

Riedmann, M., 2003. Laboratory calibration of the Compact Airborne Spectrographic Imager (CASI-2), *EPFS Research Monograph*, http://www.soton.ac.uk/~epfs/monographs/monographs.shtml (29 Sep. 2003).

Riedmann, M., and Milton, E.J., Development and evaluation of an in-flight band selection methodology for optimal use of data from airborne hyperspectral sensors. *Annals of Geophysics*, in press.

Riedmann, M., and Milton, E.J., 2003. Supervised band selection for optimal use of data from airborne hyperspectral sensors. *IGARSS 2003 IEEE International Geoscience and Remote Sensing Symposium*, Toulouse, France, IEEE, CD-ROM.

Riedmann, M., and Rollin, E.M., 2000. Laboratory calibration procedure of the Compact Airborne Spectrographic Imager (CASI-2) owned by NERC. *Activities of the NERC Equipment Pool for Field Spectroscopy (EPFS) in support of the NERC Airborne Remote Sensing Facility*, University of Southampton, British Geological Survey, Keyworth, Nottingham, UK, 21-31.