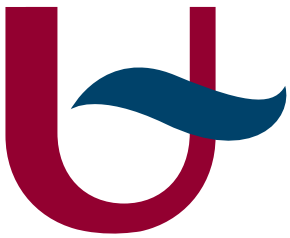


Faculteit Wetenschappen

Departement Natuurkunde

Information Extraction from Hyperspectral Images

Applied to Vegetation

Informatiewinning uit Hyperspectrale Beelden

Toegepast op Vegetatie

Proefschrift voorgelegd tot het behalen van de graad van **Doctor in de Wetenschappen** *aan de Universiteit Antwerpen te verdedigen door*

Pieter Kempeneers

Promotor:
Prof. Dr. Paul Scheunders,
Co-promotor:
Dr. Walter Debruyn.

Antwerpen, 2007

Information Extraction from Hyperspectral Images

Applied to Vegetation

Pieter Kempeneers

PhD Dissertation

2007

University of Antwerp

Promotor:
Prof. Dr. Paul Scheunders,
Co-promotor:
Dr. Walter Debruyn.

— To my wife Helga and daughters Pauline and Charlotte —

— To my parents—

Thanks

This thesis could not have been produced without the support and assistance of numerous friends and colleagues, for whom I would like to include this acknowledgment.

First and foremost I wish to thank my supervisor, Paul Scheunders. Thank you for your excellent guidance and stimulating ideas. Thanks for reading all the papers and this thesis with much care and for giving relevant comments. Thanks to all staff of the university of Antwerp for making this work possible, not in the least Dirk Van Dyck, as head of the Vision Lab department. My greatest gratitude to Steve De Backer. You have been my mentor throughout these five years. Thank you for your bright ideas and your help with numerous problems I encountered. You have become a friend to me.

I also thank my co-promotor and colleague at VITO, Walter Debruyn, for supporting me in pursuing this PhD and for your guidance as a team leader in our group. Thanks to all other colleagues at VITO, related in some way to this thesis: Dirk Van Speybroeck, our head of department, for supporting my research, and all members of the hyperspectral group. Marcel and Johan, thank you for your excellent work in the field and special thanks to the UAV group for preprocessing the datacubes. My gratitude to all reviewers, notably Frank, Sindy, Els, Luc and Bart. For your advice on C++ and scripts, many thanks to Walter (Heyns), Tim and Tom. I would also like to thank my co-authors and colleagues outside of VITO. In particular, but not only, Sam Provoost, Stephanie Delalieux and Pablo Zarco-Tejada and Wout Verhoef for their work in the field, reviews and helpful ideas.

This work greatly depends on the generosity of free software authors. I would like to express my support to them in this acknowledgment. I wrote this thesis in LaTeX 2_ε (Donald Knuth and Leslie Lamport). My computer runs the GNU/Linux operating system (Richard Stallman, Linus Torvalds and many others). I use

XEmacs (forked from GNU Emacs) as an editor, gnuplot for creating plots and the GNU C++ compiler for compiling my own code. In addition, I am especially grateful to: Andres Kuusk (MCRM), Mat Disney (C++ wrapper for MCRM), Peter North (FLIGHT), Lester Ingber (Adapted Simulated Annealing), Robert Davies (Newmat), Christian Holm (C++ wrapper for GNU scientific library and option–).

Special thanks to Carine Petit and Joost Vandenabeele from the federal science policy office. This work has been entirely funded by the STEREO program.

Combining a PhD thesis with a professional career and two children at home requires some personal sacrifices, but most of all a lot of support. Therefore, I would like to thank all my friends and family. My wife, Helga, thank you for your support, love and patience, I contribute this thesis first to you. My lovely daughters, Pauline en Charlotte, thank you for your love and patience. For your inspiration, I contribute this thesis to you as well. My parents, for your constant support. I owe everything to you, and for that, I contribute this thesis to you too. My parents in law, Christiane and Hugo, thank you for taking care of Pauline en Charlotte for so many hours and for your assistance at our home.

Balen
March, 2007 *Pieter Kempeneers*

Contents

Summary

This thesis deals with processing of hyperspectral remote sensing data, to optimize information extraction. Hyperspectral remote sensing, also referred to as imaging spectroscopy, is still a relatively new area of remote sensing. Hyperspectral data are both interesting and challenging. Classification problems can better be solved using the complete spectral signatures of the respective classes than if only a few broad spectral bands are available. Simultaneously, the dramatic increase of spectral bands in sensor technology has pushed the limits of spectral feature extraction from remote sensing imagery. Traditional techniques that use all available spectral bands, often fail on hyperspectral data due to the *curse of dimensionality* that comes along. Alternatively, index based features do not take full advantage of the information content in the available spectrum. They are either too general and fail for complex problems hyperspectral applications are typically faced with. Or, application specific indices are derived empirically, which often lack robustness.

The first part follows a data-driven approach, based on statistical techniques. We focus on feature extraction and classification. The objective is twofold. First, we try to find optimal techniques that perform well (in terms of classification accuracy) for high dimensional data. They have to take full advantage of the available spectrum. At the same time they must be generic, i.e., applicable to a broad range of classification problems. The second objective is to make a significant contribution to the current knowledge about the application of hyperspectral data on vegetation.

Our first contribution is a generic procedure for the classification of hyperspectral data, applied to vegetation stress detection. A traditional approach for this kind of applications is to use index based indicators. However, most part of the spectrum is not used in this way. Not so in the proposed procedure, where features are based on the discrete wavelet transform. To solve the problem of dimensionality, sequential forward floating search is used as a feature selection scheme. It is

compared to single best features, based on spectral bands and on some common vegetation stress indices found in the literature. The advantage of our approach is not only a better performance in stress detection accuracy, it also provides information of the discriminating spectral positions and bandwidths and the required spectral resolution.

Next, we propose a new band selection method. The band settings, defined by their spectral location and width, are optimized to minimize classification error. The method is applied as a dimensionality reduction procedure for hyperspectral classification. As an experiment, we classify 13 dune vegetation types in an hyperspectral data set. Results show an improvement in classification performance over feature selection and other band selection techniques. Furthermore, the selected bands mark locations in the spectrum that are important for classification. This makes the technique not only useful for dimensionality reduction, but also for interpretation and an aid for the tuning of sensor settings.

Our last contribution of part I constructs a classification framework for multiclass problems, based on a combination of binary classifiers. The classification framework is built on proven techniques, with some modifications. We use the posterior class probabilities for smoothing and unmixing. A case study is performed on the surveillance of dune vegetation.

The second part of this thesis has a different approach. It is a physical one. Reflectance is no longer just a digital signal from which features can be extracted in a machine learning process. It is related to a physical process, dealing with solar irradiation and its interaction (scattering and absorption) with different materials. This is the subject of radiative transfer models. They allow to simulate reflectance from a parameter set, describing radiative transfer in a scattering/absorptive medium together with sun and view geometry. When the model is inverted, these parameters can be estimated from the observed reflectance. In particular, part II of this thesis focuses on the retrieval of biochemical parameters of vegetation (chlorophyll) via inversion of leaf and canopy reflectance models. Our main contribution in part II is a study on the effect of viewing conditions on the accuracy of parameter estimation for different estimation techniques. A final case study covers a peach orchard in Zaragoza, Spain.

Nederlandse samenvatting

Deze thesis behandelt de verwerking van hyperspectrale data, met het oog op een optimale extractie van informatie. Hyperspectrale sensoren, of beeldspectrometers, zijn nog relatief nieuw in aardobservatie. Ze leveren zeer interessante data, maar vereisen ook speciale verwerkingstechnieken. Een typische toepassing is beeldklassificatie. In tegenstelling tot klassieke sensoren, met slechts een beperkt aantal brede spectrale banden, nemen hyperspectrale sensoren het volledige reflectantiespectrum op. Hierdoor zijn we beter in staat om verschillende klassen van elkaar te onderscheiden. Tegelijkertijd brengt de stijging van het aantal spectrale banden een uitdaging met zich mee voor het afleiden van kenmerken. Traditionele technieken die alle beschikbare spectrale banden gebruiken, zijn voor hyperspectrale data meestal ongeschikt. Een alternatief zijn de kenmerken op basis van indices, maar deze benutten slechts een fractie van de beschikbare informatie in hyperspectrale data. Hierdoor zijn ze ofwel te algemeen, waardoor ze niet in staat zijn om de complexe problemen, typisch voor hyperspsectrale toepassingen, op te lossen. Ofwel zijn ze op een empirische manier zo specifiek voor een bepaalde toepassing ontworpen, dat ze te weinig robuust zijn voor gewijzigde randvoorwaarden.

Het eerste deel heeft een data-gedreven aanpak, waarbij we vooral gebruik maken van statistische technieken. Het zwaartepunt ligt in het afleiden van kenmerken en klassificatie. We stellen twee grote objectieven tot doel. Ten eerste zoeken we naar optimale technieken die geschikt zijn (met het oog op een accurate klassificatie) voor de hoge dimensies van hyperspectrale data. Ze moeten het beschikbare spectrum ten volle benutten en daarnaast generisch zijn, i.e., toepasbaar op een brede waaier van problemen. Ten tweede willen we merkbaar bijdragen aan de huidige kennis over de toepassing van hyperspectrale data op vegetatie.

Onze eerste bijdrage is een generische procedure voor de klassering van hyperspectrale data, toegepast op stressdetectie in vegetatie. Traditioneel worden hier-

voor indices gebruikt. Een nadeel van deze stressindicatoren is dat ze het grootste deel van het beschikbare spectrum niet benutten. In de voorgestelde procedure kijken we wel naar het volledge spectrum, door gebruik te maken van kenmerken die gebaseerd zijn op de discrete wavelet transformatie. Het probleem van hoge dimensionaliteit lossen we op door kenmerken te selecteren volgens een sequentiëel zoekschema. Op een vlottende manier zoeken we naar de best presterende kenmerken, waarbij we ook kenmerken verwijderen om niet in een lokaal minimum verzeild te raken. We vergelijken met andere technieken zoals de univariate beste kenmerken, gebaseerd op spectrale banden, en enkele bekende stressindicatoren uit de literatuur. Het voordeel van onze aanpak is niet enkel een nauwkeurige stressdetectie. Het geeft ook inzicht in de relevante spectrale posities en bijhorende bandbreedte die nodig zijn voor deze applicatie.

Daarnaast stellen we een nieuwe methode voor bandselectie voor. De centrale golflengtes en hun bandbreedte worden geoptimaliseerd voor klassificatieproblemen. We passen deze methode toe op hyperspectrale data om de hoge dimensie te reduceren. Als experiment klasseren we hyperspectrale data in 13 klassen van duinvegetatie. Uit dit experiment blijkt dat de resultaten beter zijn met de voorgestelde techniek dan met andere schema's voor band- en kenmerkenselectie. Daarenboven markeren de geselecteerde banden de plaatsen in het spectrum die van belang zijn voor de klassificatie. Dit maakt de techniek niet alleen bruikbaar voor de kenmerkenreductie, maar ook voor interpretatie en als hulpmiddel voor het afstellen van sensoren.

Onze laatste bijdrage in deel I is het creëren van een kader voor klassificatie van hyperspectrale beelden. We lossen multi-klasse problemen op door binaire klasse-problemen met elkaar te combineren. We maken gebruik van bestaande technieken, op enkele aanpassingen na. We gebruiken de posterior probabiliteit van een klasse om het klassificatieresultaat op te kuisen en te ontmengen. Als casestudy kiezen we voor het toezicht op duinvegetatie.

Het tweede deel in deze thesis heeft een verschillende, fysische, aanpak. Reflectantie is meer dan een digitaal signaal waarvan we statistische kenmerken kunnen afleiden. Het is gerelateerd met het fysische proces van zonnestraling en haar interactie (verstrooiing en absorptie) met verschillende materialen. Dit kunnen we modelleren via speciale stralingsmodellen. Ze simuleren de reflectantie, uitgaande van de observatie- en zonnehoeken en een aantal parameters die de verstrooiing en absorptie in het medium beschrijven. Omgekeerd kunnen we, gegeven een reflectantiesignaal, deze parameters schatten. Hiervoor dienen we wel het model te inverteren. Specifiek handelt deel twee in deze thesis over de schatting van biochemische parameters van vegetatie (meer bepaald chlorofyl). We inverteren hiervoor modellen die de straling beschrijven in zowel blad als bladerdek.

Onze voornaamste bijdrage in deel II is een studie over het effect van observaties met verschillende hoeken op de schatting van parameters. We onderzoeken dit effect aan de hand van hyperspectrale vliegtuigbeelden over een perzikenboomgaard in Zaragoza in Spanje.

Thesis outline

In **chapter 1**, we start with some physical concepts of remote sensing. We give a brief overview of sensor types and then focus on imaging spectroscopy. We present some of the main applications, but our primary interest in this thesis is vegetation.

Chapter 2 presents the general processing scheme for hyperspectral imagery and covers a literature overview. Some of the steps are beyond the scope of this thesis and are only discussed briefly or not at all. The main focus is on feature extraction and classification of hyperspectral images. We introduce the feature space and discuss how features can be generated from hyperspectral data. Different features are presented, including vegetation indices and statistical features. Transformations, often used for feature extraction, are also covered. We introduce time versus frequency representations, with a focus on the discrete wavelet transform. We then continue with some fundamental concepts of supervised classification. Unsupervised classification is touched briefly in our overview. Some commonly used classifiers and architectural designs are also presented. We conclude the literature overview with a section on validation and give some examples of accuracy measures for both classification maps and continuous parameters.

The remaining sections of part I deal with original work. **Chapter 3** introduces a generic wavelet based feature extraction scheme. The narrow contiguous bands of hyperspectral sensors are able to detect subtle changes in spectral signatures. Most techniques in the literature concentrate on single indices (typically based on band ratios). In this thesis, a generic technique is proposed that fully exploits the available bandwidth provided by hyperspectral data, avoiding its lurking *curse of dimensionality*. A first case study deals with stress detection in apple orchards, from a binary classification approach (stress and no stress).

In **chapter 4**, we introduce a new technique for optimal band selection, optimizing the separation between the different spectral classes. As a prerequisite for this technique, we need spectra with narrow contiguous band, which is typically

the case for hyperspectral data. The method is applicable as a band reduction technique, but it can as well serve the purpose of data interpretation or be an aid in sensor design. The proposed band selection technique is compared to other existing techniques in a case study on dune vegetation mapping. We used hyperspectral data from the CASI sensor from a flight campaign in 2002. The large number of spectral bands (48 bands between 450 and 950 nm) provided a good data set for testing band selection.

Chapter 5 presents a generic hyperspectral classification framework. It is built on proven techniques, with some modifications. Binary and multi-class problems are dealt with first, introducing the one-against-one ensemble classifier. A technique for coupling probabilities is also introduced, retaining the class probabilities for the ensemble classifier. They are used for unmixing and smoothing the classification output. We tested the presented classification framework in another case study on dune vegetation mapping. We used a new hyperspectral data set from a flight campaign in July 2004. The data was acquired with the AISA sensor, with 32 bands and 1 m ground resolution. The classification of the dune vegetation is important in two perspectives. Apart from a diverse habitat, dune vegetation is also a natural seawall protecting the Belgian coastline from storms and floods.

In contrast to classification problems where a discrete label is assigned to the data observed, part II deals with real valued parameters. In this thesis, we are interested in vegetation parameters and status, i.e., the estimation of biophysical (e.g., leaf area index) and biochemical parameters (e.g., chlorophyll content). Parameter estimation is discussed within the context of radiative transfer models. This is because we focus on (inverse) modeling techniques. Statistical (regression) methods are also used, but often lack robustness. In the case of vegetation, the predictive algorithm obtained from regression, trained on a specific site and crop, is not reliable for other conditions. The selected bands depend on the dataset at hand, influenced by species, canopy structure, and viewing conditions. Radiative models have gained a lot of interest for the purpose of parameter estimation. In **chapter 6** we give an overview of leaf and canopy models. These models estimate reflectance from leaf and canopy parameters respectively. To be of any practical use for parameter estimation, they must be inverted, which is not a trivial task.

The original work in part II is on model inversion applied to multi-angular hyperspectral data. In **chapter 7** we first deal with spectral matching of modeled and observed reflectance spectra. Then, we focus on multi-angular observation data, with a case study on stress detection in peach orchards. We estimated leaf chlorophyll concentration through model inversion, from hyperspectral imagery over an artificially treated orchard. The objectives were to examine model inversion robustness under changing viewing conditions, and the potential of multi-angle hyperspectral data to improve the accuracy of chlorophyll estimation. Two canopy models, linked to a leaf model, were compared. A simple linear regression was also performed as a reference.

List of acronyms

Notation	Description
4SAIL2	SAIL canopy model including non-Lambertian soils and improved for numerical robustness
ACRM	A two-layer Canopy Reflectance Model (successor of MCRM)
AHS	Airborne Hyperspectral Scanner (sensor)
AISA	Airborne Imaging Spectrometer for Applications (sensor)
ASD	Analytical Spectral Devices
ATCOR-4	Atmospheric/Topographic CORrection for airborne imagery v.4
ATSR-2	Along Track Scanning Radiometer (sensor)
AVHRR	Advanced Very High Resolution Radiometer (sensor)
AVIRIS	Airborne Visible-Infra Red Imaging Spectrometer (sensor)
BHC	Binary Hierarchical Classifiers
BRDF	Bi-directional Reflectance Distribution Function
CASI	Compact Airborne Spectrographic Imager (sensor)
CCD	Charged Coupled Device
CD	Compact Disk

Notation	Description
CHRIS	Compact High Resolution Imaging Spectrometer (sensor)
CMOS	Complementary Metal Oxide Semiconductor
DBFE	Decision Boundary Feature Extraction
DCA	Detrended Correspondence Analysis
DT	Dark (reference) Target
DTC	Decision Tree Classifiers
DWT	Discrete Wavelet Transform
EGFN	Edge-Green First Derivative Normalized difference
EGFR	Edge-Green First Derivative Ratio
EM	ElectroMagnetic (spectrum)
Emacs	(GNU) Editor MACroS
FLIGHT	Forest LIGHT interaction (canopy) model
FWHM	Full Width at Half Maximum
GeoSAIL	SAIL canopy model adding second layer to mimic vertical leaf color gradient
GNU	Gnu's Not Unix
GO	Geometric-Optic (model)
GPS	Global Positioning System
ICZM	Integrated Coastal Zone Management
IFOV	Instantaneous Field Of View
IMU	Inertial Measurement Unit
INTA	Instituto Nacional de Técnica Aeroespacial
LAD	Leaf Angle Distribution
LAI	Leaf Area Index
LDA	Linear Discriminant Analysis
LIDAR	LIght Detection and Ranging
LUT	LookUp Table
LWIR	Long Wave InfraRed (part of the electromagnetic spectrum)
MCRM	Markov chain Canopy Reflectance Model
MIR	Middle-InfraRed (part of the electromagnetic spectrum)
MISR	Multi-angle Imaging SpectroRadiometer (sensor)
MNF	Minimum Noise Fraction

Notation	Description
MODIS	Moderate Resolution Imaging Spectroradiometer (sensor)
MODTRAN	MODerate resolution model for lowTRAN7
MWIR	Mid-Wave InfraRed (see MIR)
NDPI	Normalized Difference Pigment Index
NDVI	Normalized Difference Vegetation Index
NIR	Near InfraRed (part of the electromagnetic spectrum)
PARGE	PARametric GEocoding
PCA	Principal Component Analysis
PDF	Probability Density Function
POLDER	POLarization and Directionality of the Earth's Reflectances (sensor)
PROBA	PRoject for On-Board Autonomy (satellite platform)
PROSPECT	(leaf optical) PROperties SPECTra (leaf model)
PSF	Point Spread Function
RADAR	RAdio Detection And Ranging
RAMI	RAdiation transfer Model Intercomparison
RMSE	Root Mean Square Error
ROC	Receiver's Operating Characteristic (curve)
ROI	Region Of Interest
RRMSE	Relative Root Mean Square Error
RT	Radiative Transfer
SA	Simulated Annealing
SAIL	Scattering by Arbitrarily Inclined Leaves (canopy model)
SAILH	SAIL canopy model including Hot spot
SFFS	Sequential Floating Forward Search
SFS	Sequential Forward Search
SIPI	Structure Independent Pigment Index
SPECTRA	Surface Processes and Ecosystem Changes Through Response Analysis (sensor)
SPOT	Satellite Pour l'Observation de la Terre
SRF	Spectral Response Function
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
SWIR	Short Wave InfraRed (part of the electromagnetic spectrum)

Notation	Description
TM	(Landsat) Thematic Mapper (sensor)
TM	Turbid Medium
TWINSPAN	Two Way INdicator SPecies ANalysis
UAV	Unmanned Aerial Vehicle
UTM	Universal Transverse Mercator (projection system)
VITO	Vlaamse Instelling voor Technologisch Onderzoek
VNIR	Visual + Near InfraRed (part of the electromagnetic spectrum)
WT	White (reference) Target
XEmacs	fork of the (GNU) Emacs editor with focus on support for X (Window system)

Chapter 1

Introduction

A picture is worth a thousand words

Confucius

(... a hyperspectral image a ten thousand)

—

1.1 Fundamentals of remote sensing

The idea of obtaining information (sensing) about an object, area, or phenomenon without physical contact (remote) is old. In 1903, small light weight cameras were attached to pigeons of the Bavarian Pigeon Corps (Fig. 1.1). A timer was set to take pictures every 30 seconds as it flew. The aerospace industry took remote sensing to the next level. The real break through had still to come with the exploration of space in the sixties. It seems like a paradox, but the space industry, with its excessive cost, has made remote sensing available for the average researcher. Spaceborne images can be seen on television every day during weather forecasts. Google earth is (to date) a free Internet application of Google Inc. that allows you to search and view any place on earth, making use of the latest available high resolution satellite or airborne imagery. With improving technology (telescopes, RADAR, digital electronics), remote sensing applications increased exponentially. Nevertheless, it is the department of defense, who has been the

Fig. 1.1: *Left: pigeon of the Bavarian Pigeon Corps, carrying a light weight camera. Right: photograph acquired by Bavarian Pigeon Corps (wingtips of the pigeon are visible near the photo edges)*

driving force behind new remote sensing technology, recognizing the unlimited applications for military use. It still is today.

The early days of remote sensing consisted of airborne photography. Film is still common in some fields as in analog photogrammetry, but most sensors are digital now. We only consider digital images in this thesis. The art of data acquisition, including the choice of sensor or platform (space- or airborne) is only one aspect of remote sensing. This thesis focuses on another: information extraction. In this general overview, we first introduce some physical principles of remote sensing. We then focus on hyperspectral remote sensing, or imaging spectroscopy, and present some applications. The remainder of the thesis is on the digital processing of hyperspectral remote sensing data.

1.1.1 Radiation and reflectance

Looking at the bright Sun you observe the electromagnetic (EM) radiation emitted. This form of energy (Q) is transferred through space in the form of an oscillating sine wave, composed of different wavelengths (λ), corresponding to the EM spectrum (Fig. 1.2). The energy transfer occurs within a certain period of time (t) until it reaches an intercepting surface (A). You can feel this energy by the heat transfer to your body. The EM radiation arrives at the surface with some orientation, represented by a solid angle (Ω_s).

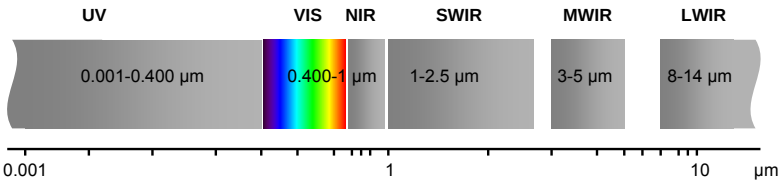


Fig. 1.2: *The electromagnetic spectrum from small to larger wavelengths: Ultra-Violet (UV), visual spectrum (VIS), Near-Infrared (NIR), Short Wave Infrared (SWIR), Mid-Wave Infrared (MWIR) and Long Wave Infrared (LWIR)*

The fundamental optical quantity *spectral radiance*, L_λ, is defined as [63]:

$$L_\lambda = \frac{d^6Q}{dt\ dA \cos\theta_s d\Omega_s\ d\lambda}\ W/m^2/sr/\mu m, \tag{1.1}$$

with θ_s the illumination zenith angle. The relevant angles for illumination (subscript s) and observation (subscript v) are shown in Fig. 1.3. The solid angles Ω_s and Ω_v are implicitly assumed to be centered around the directions (θ_s, ϕ_s) and (θ_v, ϕ_v), also referred to as the illumation and observation *conditions*. The definition of spectral radiance remains if the surface dA acts as a source of radiation instead of a receptor. The spectral radiance is thus the amount of radiant energy, which is received (emitted) by a surface dA over the period $t + dt$ and the wavelength band $\mu + d\mu$, from within (into) the solid angle $\Omega_{s(v)}$. Real sensors

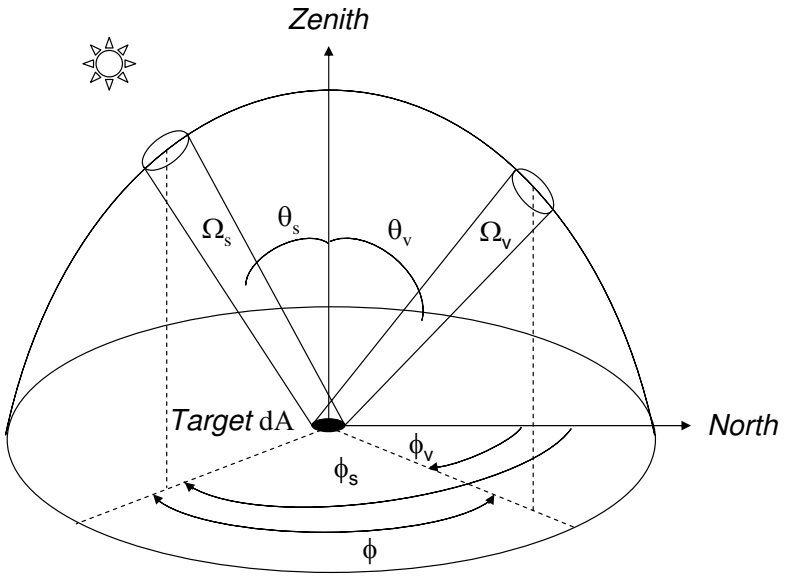


Fig. 1.3: *Definition of the sun and view angles: sun zenith angle (θ_s), view zenith angle (θ_v), sun azimuth angle (ϕ_s), view azimuth angle (ϕ_v), relative azimuth angle (ϕ), solid sun angle (Ω_s), and solid view angle (Ω_v).*

can not measure infenitesimally small bandwidths and measure the *Radiance*, L. It can be mathematically obtained by the *spectral integration* of L_λ over some sensor specific finite wavelength band $[\lambda_1, \lambda_2]$:

$$L = \int_{\lambda_1}^{\lambda_2} L_\lambda d\lambda = \frac{d^5Q}{dtdA \cos\theta_v d\Omega_v}\ W/m^2/sr \tag{1.2}$$

Another optical quantity that is important in remote sensing is *Irradiance*, the

directional integration of the *incoming* radiance:

$$\mathrm{E} = \int_{\Omega_s} \mathrm{L} \cos\theta_s \mathrm{d}\Omega_s \ \mathrm{W/m^2} \tag{1.3}$$

Notice that in this case, a distinction in terminology is made between the source and receptor situations. If dA is a radiant source instead of receptor, we speak of *Exitance*, M.

Any physical body, whose temperature (T) is above absolute zero, radiates (transmits) electromagnetic energy. In 1879, Boltzmann discovered that a body emits thermal radiation at a rate that is proportional to the temperature of the surface of the body raised to the fourth power:

$$\mathrm{M} = \sigma \mathrm{T}^4 \ \mathrm{W/m^2} \tag{1.4}$$

He found the proportionality constant $\sigma = 5.6693 \times 10^{-8}$ $\mathrm{W/m^2/K^4}$. A *black body* is a perfect absorber of radiation. It also emits electromagnetic energy at specific wavelengths λ depending on its temperature. Planck published his law of black body radiation in 1901 that predicts the spectral intensity of electromagnetic radiation in function of its temperature:

$$M_\lambda = \frac{\epsilon_\lambda c_1}{\lambda^5 (\mathrm{e}^{c_2/\lambda \mathrm{T}} - 1)} \ W/\mathrm{m^2}/\mu\mathrm{m}, \tag{1.5}$$

with constants $c_1 = 3.7415 \times 10^8$ $W\mu\mathrm{m}^4/\mathrm{m}^2$ and $c_2 = 1.4388 \times 10^4$ μm K. The factor ϵ_λ is the *emissivity* of the body. Most bodies only emit a fraction of the energy emitted by the perfect black body, determined by ϵ_λ. For gray bodies the emissivity is a constant between 0 and 1. Other bodies have emissivities that are wavelength dependent. By integrating (1.5) over all wavelengths, the Boltzmann law can be derived again (obtaining the exitance M). Typical energy radiation curves, for blackbodies at various temperatures, are shown in Fig. 1.4. The maximum energy is reached at shorter wavelengths as temperature increases (Wien's Displacement law):

$$\lambda_p = a/\mathrm{T} \ \mu\mathrm{m}, \tag{1.6}$$

with the constant $a = 2898$ μm/K. This explains why the human eye observes the color of a heating body as dark (cold) to bright red, yellow and blue (hot). The Sun emits energy with a spectral distribution similar to that of a blackbody at a temperature of 5800 K. Our visible spectrum ranges from 390 nm (blue) to 700 nm (red), coinciding with the peak radiation wavelengths radiated by the Sun. The light from a high temperature Tungsten filament of a vacuum incandescent lamp at 3000 K appears as white. At low temperatures (1800 K) it appears reddish yellow.

An overview of some important radiometric terminology in remote sensing is given in Table 1.1.

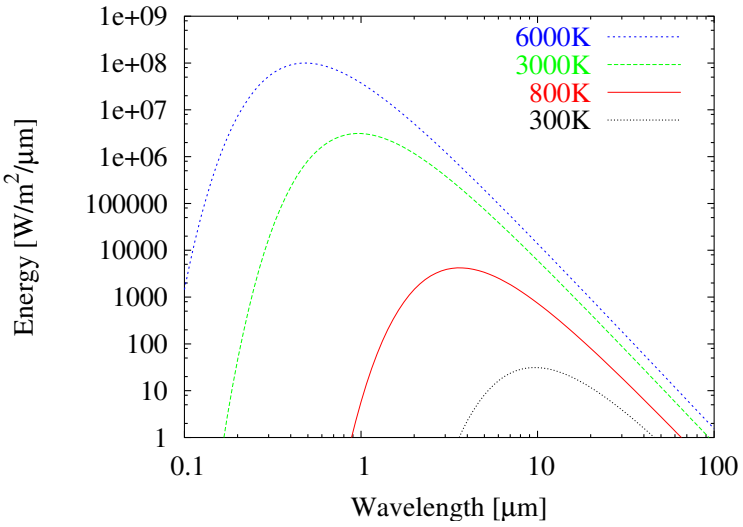


Fig. 1.4: *Energy emitted by blackbody with different temperatures: 6000 K (Sun), 3000 K (Tungsten filament), 800 K (red hot object), 300 K (Earth)*

1.1.2 Interaction with surfaces

Bodies that are not black only absorb a proportion of the incident radiation, defined by the *absorptance*. Part of the radiation is reflected, which is why we can see bodies with lower temperatures. The absolute amount of radiation, reflected by a surface dA, obviously depends on the incoming irradiance E, and hence can not be considered as an intrinsic feature of the surface. The nature of reflection depends on sizes of surface roughness or smoothness in relation to the wavelength of the radiation considered. A *Lambertian surface* behaves as a perfect diffuse reflector, emitting equal radiance in each direction. Because of its simplicity, many (rough) natural surfaces are assumed as Lambertian for visible radiation. However, reflectance of real objects depends on the direction from illumination (Sun) and viewing angles. These reflectance characteristics are described by the *Bi-directional Reflectance Distribution Function* (BRDF):

$$\mathrm{f}(\Omega_v, \Omega_s) = \frac{\mathrm{d}^2\mathrm{L}(\Omega_s, \Omega_v)}{\mathrm{L}(\Omega_s)\cos\theta_s \mathrm{d}\Omega_s} \ \mathrm{sr}^{-1} \tag{1.7}$$

The denominator represents the irradiance in dA (1.3) from within the solid angle $\mathrm{d}\Omega_s$. The numerator indicates how much of this incoming radiation is reflected by the target surface dA as a radiance into $\mathrm{d}\Omega_v$.

However, the BRDF is not feasible to measure outside laboratory conditions, where incidence orientation can not be controlled. In practice, another relative quantity is used, *reflectance factor* (R). It is defined as the radiance, reflected by

Tab. 1.1: *Radiometric terms with their expressions and units*

Symbol	Description	Expression	Unit
Q	Radiant Energy		J
Φ	Radiant flux (power)	$\frac{dQ}{dt}$	W
E	(incoming) Irradiance	$\frac{d\Phi}{dA}$	W/m^2
M	(outgoing) Radiant exitance	$\frac{d\Phi}{dA}$	W/m^2
E_λ	(incoming) spectral Irradiance	$\frac{dE}{d\lambda}$	$W/m^2/\mu m$
M_λ	(outgoing) spectral Radiant Exitance	$\frac{dM}{d\lambda}$	$W/m^2/\mu m$
L	Radiance	$\frac{d^2\Phi}{\cos\theta dA d\Omega}$	$W/m^2/sr$
L_λ	spectral Radiance	$\frac{dL}{d\lambda}$	$W/m^2/sr/\mu m$

the target surface dA in a certain solid angle $d\Omega_v$, divided by the radiance reflected by a white Lambertian panel in the same solid angle and in the same conditions of illumination. The BRDF of a white Lambertian panel is $1/\pi$, so we can write the latter radiance as E/π. We then obtain for the dimensionless reflectance factor:

$$R = \frac{\pi L(\Omega_v)}{E} \tag{1.8}$$

Next to absorptance and reflectance, we define *transmittance* as the radiation that passes through a substance. A typical example of a body able to transmit significant amount of radiation is water. But vegetation (leaves) also transmits radiation.

Finally we briefly mention *fluorescence* as another interaction of an object to incident radiation. It occurs when an object emits radiation at a different wavelength then it was illuminated with. Fluorescence is involved in vegetation, and can be used to discriminate healthy and stressed leaves [151, 245]. However, fluorescence effects are small and can only be measured with accurate instruments.

1.1.3 Atmospheric effects

The energy transfer into and out of the earth-atmosphere system affects the way in which we can remotely sense properties of this system. Fig. 1.5 shows a schematic overview of the principle components of *at sensor* radiance.

- Path radiance, I_O, is the radiation scattered by the atmosphere into the viewing direction
- Reflected irradiance from the target, I_S, and (directly or diffusely) transmitted to the sensor
- Adjacency, I_D, is caused by radiation reflected by the background (surroundings) and (directly or diffusely) transmitted to the sensor

Scattering and absorption are the two main atmospheric effects on electromagnetic radiation. Absorption converts the incident radiation energy to heat at molecular

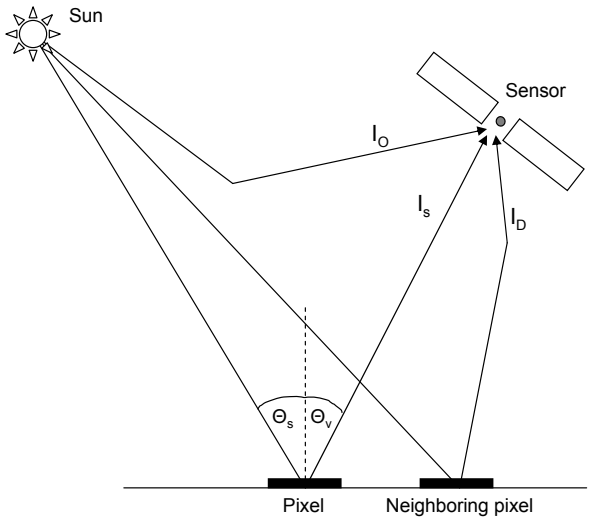


Fig. 1.5: *Principle components of at sensor radiance. I_S represents radiation reflected from ground pixel, I_O is obtained from direct scatter of the atmosphere, and I_D is the radiation reflected by a neighboring ground pixel (adjacency effect). θ_s and θ_v are the observation and sun zenith angles respectively*

level. It occurs at specific wavelengths, referred to as *absorption bands*. Particles such as gas molecules, dust and water molecules redirect the incident energy, resulting in scattering [176]. If particles are small (gas molecules of about 10^{-4} μm), we deal with *Rayleigh scattering*. The amount of Rayleigh scattered radiation is inversely proportional to λ^4. Sunlight is scattered first in the shorter wavelengths, which is why a (clear) sky appears blue. Sunsets are red, because for low sun angles, sun-rays must travel through a long atmospheric path before reaching us. In their journey, most of the shorter wavelengths are scattered. First blue is missing (Sun appears as yellow), then also green (Sun appears as red).

In case of haze, small cloud droplets are of comparable size to the radiation wavelength and *Mie scattering* [150] is involved. Finally, *non-selective scattering* occurs when particles are much larger than the wavelength (large aerosols, cloud droplets or ice crystals). In this case, the scattering is not very dependent on wavelength which results in the white or gray appearance of clouds.

1.1.4 Sensor types

We define a sensor as an instrument capable of measuring electromagnetic radiation. There are hand held sensors (e.g., field spectrometer) that can be used in the laboratory or during field work for reference measurements. Most sensors we are dealing with in remote sensing are mounted on an air- or spaceborne platform. Depending on the orbit height, spaceborne platforms can be geostationary (geosynchronous) or sun-synchronous. There is a lot of power needed to launch a geostationary satellite 36000 km above equator. Once in orbit, it stays vertically above the same spot on earth and rotates with the earth. With its large spatial and temporal coverage it is ideal for weather images, e.g., GOES (Geostationary Operational Environment Satellite series). Currently, ground resolution of geostationary sensors is rather coarse –within the range of kilometers– but that might be improved in the future.

For more detailed terrestrial applications, we need imagery with more detail. A high ground resolution[1] can be obtained with platforms at lower altitudes (700 to 900 km). The acquired images are stored onboard and sent to a receiving station on earth. Because of the near polar orbits (81 to 82 degree angle to equator), receiving stations are ideally located near the poles, e.g, Kiruna, Sweden. The satellites are also launched in sun-synchronous orbits, to produce similar daytime lighting conditions for each track.

Active sensors have a built-in source of radiation. Most common are RADAR (radio detection and ranging) and LIDAR (light detection and ranging). In this thesis, we only consider passive sensors, detecting electromagnetic (EM) radiation from an external energy source (Sun). *Panchromatic* sensors are sensitive to all wavelengths of visible light, producing a single plane (*band*) image of a scene.

[1] the term high resolution is a relative term and depends on state of the art. When the first AVHRR sensor was launched in 1978, its ground resolution of 1.1km was considered as very high (Advanced Very High Resolution Radiometer). Today, sub-meter ground resolution satellites are commercially available (Quickbird: 60cm, Ikonos: 1m).

Multispectral sensors contain multiple detectors that are sensitive to specific ranges of the EM spectrum (*spectral bands*). Typically, multispectral sensors have a few bands (< 10), *superspectral* sensors have many (> 10) bands, and *hyperspectral* sensors, our main focus in this thesis, have up to a few hundred spectral bands. Spectrometers are a special type of hyperspectral sensors, where spectral bands are contiguous. Some hyperspectral sensors have tunable bands that allow gaps in the spectral domain and thus are not spectrometers. However, this terminology is not strictly used and both terms are often used as synonyms (also in this thesis). The precise spectral information contained in a hyperspectral image enables better characterization and identification of targets.

To complete our sensor overview, we define two types of scanning sensors: whiskbroom and pushbroom scanners. Their introduction is important to understand some of the causes of errors in the image data we have to deal with in data processing. The whiskbroom scanner (Fig. 1.6) uses a rotating mirror to scan the target surface. It directs a narrow beam of energy onto the detector. An important factor of the whiskbroom is the instantaneous field of view (IFOV) of the scanner. It is the field of view (or cone angle) of the mirror at the instant that the energy is sensed on the detector. Because there is only one detector for each spectral image, we do not have to deal with inter-calibration problems between neighboring pixels. As a drawback, there is the mechanical part of the rotating mirror, limiting resolution due to slower data rates.

The pushbroom (or line detector, Fig. 1.7) scanner uses a wide-angle optical system that focuses on a strip across the whole of the scene onto a linear array of CCD (Charged Coupled Device) or CMOS (Complementary Metal Oxide Semiconductor) detectors. The signal from each detector is sampled to create a record for the across track pixels. Each column of each spectral image is acquired with a different sensor, which can lead to striping artifacts. On the other hand, high data rates can be obtained, which make this technique popular for high resolution sensors.

1.2 Imaging spectroscopy

Beginning in the 1980s, Goetz [89] and his colleagues at the Jet Propulsion Laboratory began a revolution in remote sensing by developing a powerful new type of instruments, among which the AVIRIS (Airborne Visible-Infra Red Imaging Spectrometer). This hyperspectral sensor, or imaging spectrometer, is a special kind of radiometer[2]. Its sensor includes a dispersive element such as a prism or diffraction grating, that can break radiation extending over a part of the spectrum into discrete wavelengths and disperse (or separate) them at different angles to detectors[3]. The term imaging refers to the end result, which is an image or a raster

[2] A radiometer is a general term for an instrument that quantitatively measures the EM radiation in some interval of the EM spectrum.

[3] The correct term would be spectroradiometer, which implies that the dispersed radiation is in bands rather than discrete wavelengths. Both terms are used in in this thesis and most literature however.

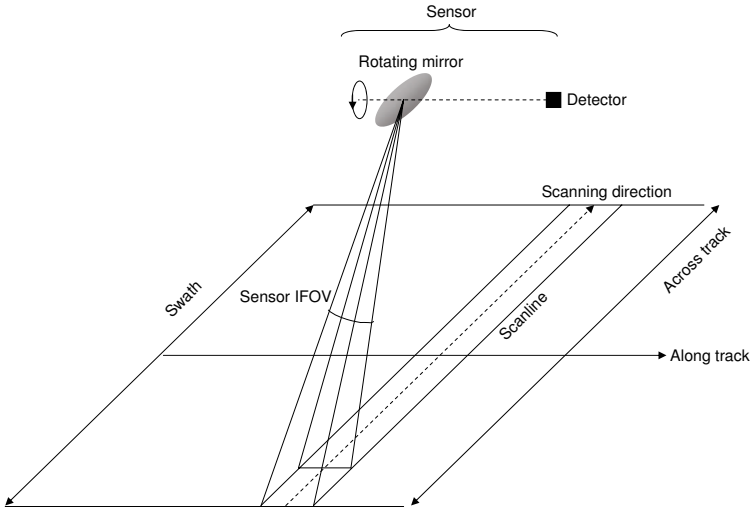


Fig. 1.6: *Whiskbroom scanner, showing swath, IFOV, across and along track*

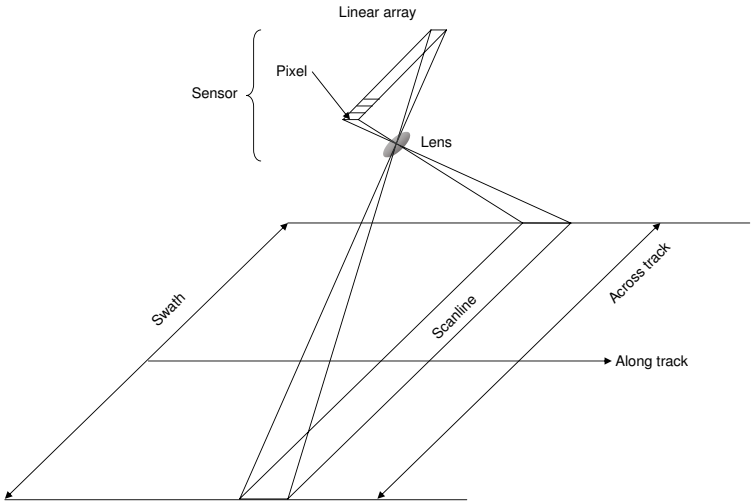


Fig. 1.7: *Pushbroom scanner, showing swath, IFOV, across and along track*

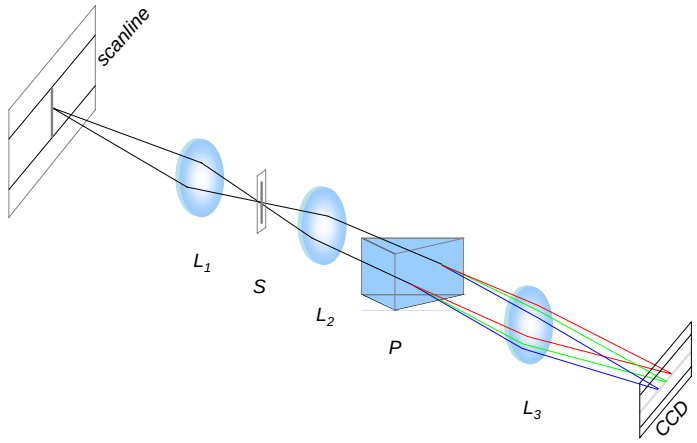


Fig. 1.8: *Basic layout of the prism spectrometer, with front lens (L_1), slit (S), collimating lens (L_2), prism (P), camera lens (L_3) and CCD.*

display. It relates the radiation to specific points in the target. A non-imaging spectroradiometer integrates the radiation received from all points in the sensed target. The hand held FieldSpec® from Analytical Spectral Devices, Inc. (ASD) is an example of a non-imaging spectroradiometer. Most imaging spectrometers, or hyperspectral sensors, operate in the visible to near infrared (NIR) spectral region, with some sensing into the short wave infrared (SWIR) wavebands.

1.2.1 Principles of imaging spectroscopy

New generation imaging spectrometers use a two-dimensional array such as a CCD to reconstruct the image of the target object with high spectral and spatial resolution (pushbroom). We illustrate some of the key features of a spectrometer starting from the prism spectrometer. Fig. 1.8 shows the optical components of a typical imaging spectrometer: front lens (L_1), entrance slit (S), collimating lens (L_2), dispersive element (P), camera lens (L_3) and detector (CCD). Images of the entrance slit are generated on the CCD. The light passes through a dispersive medium such as a prism. *Refraction* causes the incoming beam to bend in different angles for different wavelengths.

The incoming radiation that is mapped onto the focal plane is thus displaced according to the wavelength. This *spatially dispersed* spectrum will however be blurred to some extend by the finite width (w) of the slit S. The spectral resolution of the spectrometer $\delta\lambda$ is therefore limited by the magnitude of this smoothing effect. In addition, it is determined by the dispersive power of the glass of which

the prism is made and the focal length of the camera lens L_3. As a consequence, the effective resolution of an instrument can only be improved for a given prism material, by reducing the slit width or by increasing the focal length of the camera lens. As the slit width determines the instrument aperture (A), the first option has a serious drawback regarding the instrument throughput, which is directly proportional to A. Decreasing the slit width while maintaining the throughput can only be obtained by increasing the area of the camera lens, making the instrument bigger.

For gratings, the dispersion of the source spectrum is obtained by *diffraction* of light. This is either obtained if light passes through a large number of parallel slits, or if light is incident upon a periodic reflective surface. If the distance between the slits (or periodicity of the surface) is small enough, light will interfere. The effect is also shown in an audio compact disk[4], separating the colors of white light. Gratings have better resolving power than prisms of considerable size, but are subject to scattering, resulting in a lower signal to noise ratio. Some spectrometers use a combination of a prism with a grating (grism) for the dispersive element.

1.2.2 Applications of imaging spectroscopy

Since the benchmark paper of Goetz [89], hyperspectral remote sensing has gained a lot of interest. Initially, applications were dominated by mineral exploration [229, 39]. Now, hyperspectral remote sensing has found its way in all sorts of applications: land and soil monitoring [54, 17], atmosphere (water vapor, cloud properties, aerosols), environmental monitoring, and snow and ice (snow cover fraction, grain size, melting). For coastal management, hyperspectral remote sensing addresses retrieval of water constituents from ocean color such as chlorophyll concentration [196], suspended sediments and organic matter [126, 56, 257]. If shallow enough to be optically detected, bottom depth can be derived from remote sensing as well [141]. More general, the coastal zone (beaches, dune vegetation) can be addressed with hyperspectral remote sensing [201, 55, 256]. Finally, vegetation is another major application area of hyperspectral remote sensing. Vegetation plays a key role in our biosphere. It provides a primary mechanism for the bio-geochemical cycles as well as surface-atmosphere interactions (carbon and hydrological cycle, exchange of O_2 and CO_2). Remote sensing provides a tool to better understand these mechanisms. High resolution hyperspectral sensors are applied on a local to regional scale for landcover mapping and estimation of vegetation parameters and status. On a global scale, the quantification of the carbon cycle and the global budget for energy and nutrients is addressed by coarse resolution satellites (SPOT VEGETATION, MODIS).

[4] The nominal track separation of a CD is 1.6 μm. This is in the range of ordinary laboratory diffraction gratings

Part I

HYPERSPECTRAL DATA PROCESSING

Introduction

In this first part, we mostly deal with feature extraction and classification. As an introduction, we present some existing techniques in chapter 2. Our first original contribution is a generic hyperspectral wavelet based feature extraction scheme, discussed in chapter 3. It combines wavelet features with a state of the art feature selection scheme. We applied it to a first case study on stress detection in apple orchards. Chapter 4 presents a new hyperspectral band selection technique with a case study on dune vegetation mapping. Finally, we propose a generic hyperspectral classification framework in chapter 5, also tested on the classification of dune vegetation.

Chapter 2

Processing hyperspectral imagery

A journey of a thousand miles
begins with a single step

Confucius

2.1 Introduction

Remote sensing involves data acquisition and data processing. A schematic overview of the different steps involved is shown in Fig. 2.1. It is a modified version of the work flow presented in [30]. Other work flows have been proposed in the literature, e.g., the one by Landgrebe, referred to as the *data analysis paradigm*[1] [138].

We start with a brief discussion on data acquisition and preprocessing. We then focus on further processing techniques. In particular, feature extraction (section 2.4) and (supervised) classification (section 2.5) are the main subjects of part I. We present some common techniques, but mostly concentrate on the ones we use in the following chapters. Validation is also presented in some more detail in section 2.6, not only for its relevance, but also to introduce some of the accuracy measures we use in the case studies later on.

[1] "The major question that the analyst must deal with is how to choose and implement a suitable sequence of algorithms by which to accomplish the desired analysis"

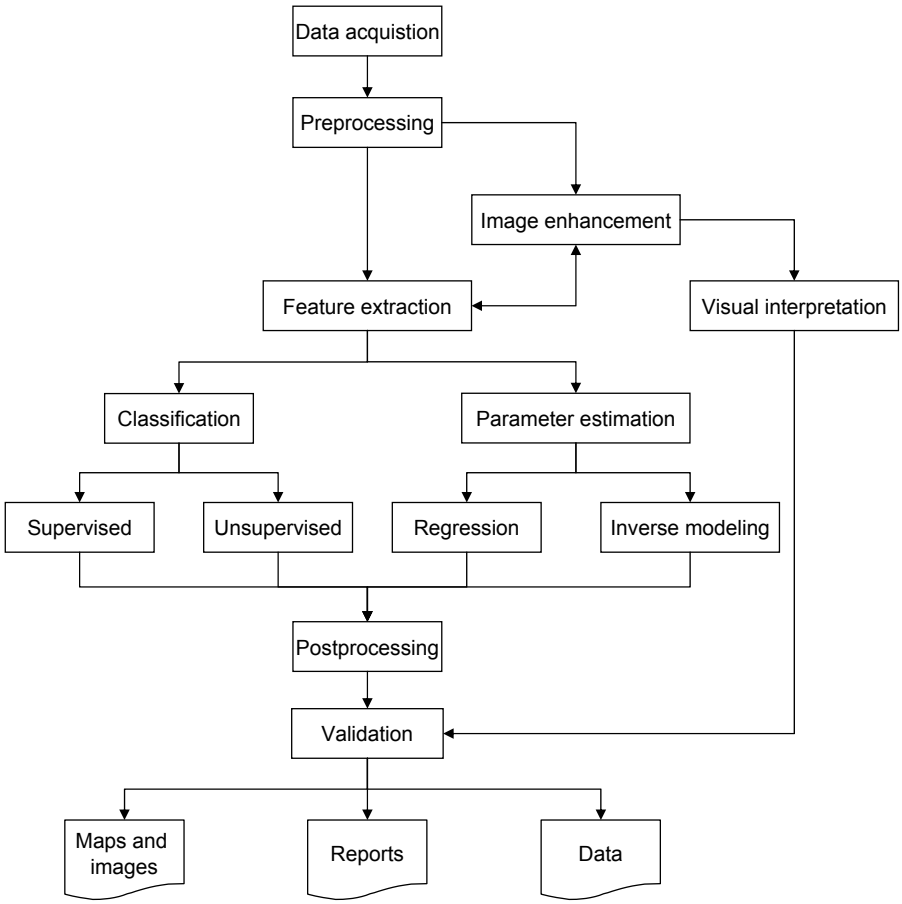


Fig. 2.1: *digital data processing work flow for remote sensing*

2.2 Data acquisition

Remote sensing data can be acquired from airborne or satellite sensors. Both platforms have their benefits and shortcomings. Once operational, satellite sensors provide low cost data at regular basis. Especially for low resolution sensors, they are ideal for monitoring (time series). Military users appreciate satellites for their discrete and safe operation, as no pilots are required. But also for scientific users the advantage is clear: there are no problems of air traffic and no need for official clearance of air space. Last year, we had to cancel a project on a historical site in Sagalassos, Turkey, due to such problems. The satellite platform is also a very stable one, which eases the task of preprocessing (geometric correction). On the other hand, larger and more expensive optical lenses are required to obtain similar ground resolution to airborne sensors. For local applications, where there is a need for hyperspectral data with very high spatial resolutions ($1 - 4$ m), the only option today is airborne imagery. This is one of the reasons why we only deal with airborne hyperspectral data in the case studies in this thesis. Airplanes are also very flexible. The flight plan can be optimized either for viewing conditions (principal plane), or maximum coverage with a minimum number of tracks. Flight plans allow to follow estuaries or other complex patterns, which results in a reduced image data set. For satellites with their fixed orbits, the number of images required to cover the same area of interest is much larger. The time of the flight can also be scheduled for cloud free observations or other requirements such as low tide (coastal applications) or optimal sun angle. A new type of platform that will certainly have a significant impact on future applications, is the Unmanned Aerial Vehicle (UAV). At tropospheric altitudes, it combines the benefits of both airborne and spaceborne platforms.

Apart from image data, measurements collected on the field (in situ data, point or localized data) are very important. A field spectrometer can be used during an acquisition flight for calibration of the image data (preprocessing). Spectral data is also collected on the ground to investigate surface conditions, to develop models for parameter estimation or to create a spectral library used in some classification techniques. It also allows to optimize band settings for sensors such as the CASI-2 (Compact Airborne Spectrographic Imager) that can be configured with different band settings.

Two other common instruments used in the field are the sunphotometer and a GPS. The first measures aerosol optical thickness and water vapour content, as input for atmospheric correction (preprocessing). A GPS can link the field location to the image data or can be applied for geometric correction (preprocessing).

Other than for preprocessing purposes, a field campaign is merely organized to collect ground reference data. This data is used for: training classifiers and regression methods, calibration of models and validation of the final products. McCloy [156] wrote a comprehensive chapter on the use of field data.

2.3 Preprocessing

Preprocessing involves radiometric and geometric corrections. Processing of radiometric errors are performed in two steps. First, errors in the optics or the electronics of the sensor are corrected by calibration. Next, atmospheric scattering and absorption are addressed by atmospheric correction.

A detector converts the incident energy into an analog electrical signal. After an analog to digital (A/D) conversion, a digital number is obtained with some bit depth (12 bits are commonly used for each spectral band). The digital numbers must be corrected for differences among sensors (relative calibration) which cause *striping* and other systematic effects in the resulting image. Absolute calibration relates the digital numbers to radiance, for which a band specific linear combination is often assumed. However, the detector sensitivity varies with temperature and age. Sensor calibration thus requires up to date calibration data, obtained from calibration sites on earth (White Sands, New Mexico), deep space, or the moon.

The radiometric distortions, introduced by the atmosphere between the surface and the sensor, are function of relative humidity, atmospheric pressure, temperature and visibility. They also depend on wavelength and the effect is larger at the edges of the swath. State of the art correction methods such as MODTRAN [18] or 6S [234] are based on detailed radiative transfer modeling. They take input from ancillary information that can be measured in the field. As an example, sun photometry measures the solar irradiance at the ground in a number of channels. It allows estimation of the amount of water vapor, ozone, and aerosols, as well as determination of the aerosol particle size distribution [198]. On the other hand, for image based methods, some of the ancillary data can be extracted from the image. A simple approach uses dark targets or flat fields for an estimate of the atmospherically scattered radiance within the image.

Sources of geometric errors are twofold: systematic and non-systematic distortions. Systematic errors are independent of flight or terrain conditions. In case of a whiskbroom sensor, the forward motion of the platform during each mirror sweep, results in *skew*, a ground swath not being normal to the along track. *Panoramic distortion* results in slightly larger pixels at the detector edge (pushbroom) or extreme positions of the rotating mirror (whiskbroom). With the whiskbroom, the effect of variable mirror-scan velocity comes to it. Especially for satellite platforms, the effect of earth rotation results also in an offset of subsequent rows. Geometric distortions can also result from optical aberrations in the sensor.

Non-systematic errors involve altitude and attitude (roll, pitch and yaw) variations of the platform (Fig. 2.2) and effects of topographic elevation.

Geometric rectification (correction) is a process by which points in an image are registered to corresponding points on a map or another image that has already been rectified. The goal is to put image elements in their proper planimetric (x and y) positions. Polynomial interpolation methods are satisfactory for registering single-band images to a rectified grid and have been used for multispectral broad-band data as well. For hyperspectral data, however, simpler nearest-neighbor resampling

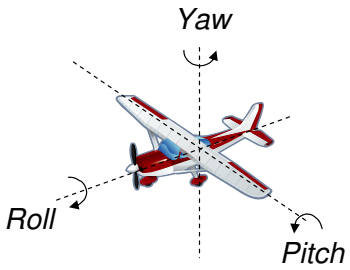


Fig. 2.2: *Attitude variations of the platform: roll, pitch and yaw.*

schemes may be preferred because these do not distort spectral characteristics. PARGE [199] and RECTIFY [164] are ground control point based procedures that have been developed to recalibrate the offsets of the attitude data, since they usually are given as relative angles. They exactly reconstruct the scanning geometry for each image pixel using position, attitude, and terrain elevation angle.

Remote sensing data products are categorized according to the level of (pre-)processing performed. Table 2.1 shows a naming convention that is commonly followed, although small deviations are possible.

Tab. 2.1: *Naming convention for remote sensing products, describing the different levels of processing. Preprocessing steps are involved up to level 2.*

Level 0	Raw image data (digital numbers)
Level 1A	Raw image data at full resolution, time-referenced, and annotated with ancillary information. Radiometric and geometric calibration coefficients are included but no correction is applied.
Level 1B	Platform/sensor exterior orientation is enhanced using GPS base station information, allowing for a full radiometric, atmospheric and geometric correction.
Level 1C	Includes in-situ measurements.
Level 2	Geometric and atmospheric corrected data of individual scenes (tracks)
Level 3+	User end product, typically the result of combining multiple scenes (mosaic) to cover the user's region of interest. Products can be classification maps or model parameters. According to the level of development, end user products are referred to as level 3 or 4.

At VITO, we developed the Central Data Processing Centre (CDPC), a processing chain to archive and process raw image data to products of higher lev-

els [20]. Geometric correction is performed by *direct georeferencing*. The positions and attitude data are measured in real time during data acquisition, using a GPS and Inertial Measurement Unit (IMU) onboard. Additionally, corrections for geographic relief are applied if a digital terrain model is available. Atmospheric correction is based on MODTRAN. The integrated atmospheric correction is equivalent to the ATCOR theory [191], with some improvements. Instead of using a single LUT (LookUp Table), we create a customized LUT for each image, which is adapted to the specific spectral bands of the sensor and the relevant geometry space obtained during geometric correction.

Commercial flight operators also have their own processing chains and provide preprocessed data products. Although most of them are based on identical principles, we prefer to preprocess the data in house. In this way we obtain a conform data set regardless of the raw data provider (flight operator). Furthermore, subsequent levels of data are available at any time, which is an advantage for some scientific applications. It also allows to reprocess the data if an improved preprocessing algorithm is available. This processing chain has only been available since the AHS airborne hyperspectral campaign in 2005, which covered the orchard plot in Zaragoza (case study in chapter 7). The data acquired during previous campaigns were obtained in geometric corrected form from the flight operator. Atmospheric correction was performed at VITO using ATCOR-4. This was the case for the CASI data in 2003, covering the apple orchard in Gorsem (case study in chapter 3), and the AISA data in 2004, covering the the dune vegetation along the Belgian coast (case study in chapter 5).

2.4 Feature Extraction

2.4.1 Introduction

Why is it, that sensors are designed with an increasing number of bands, but we speak of the *curse of dimensionality*? Indeed, more spectral bands include more information. However, an important concept in the theory of learning is *generalization*. Given a finite set of training samples, a well designed classifier must be able to classify an unseen test sample. The success of a classifier (expressed by its accuracy) is directly related to its ability to generalize what it has been trained for. As pointed out by many authors, including Duda [61] and Landgrebe [139], the performance of a classifier diminishes with a higher number of features. This problem is very prominent for hyperspectral data. In 1968, Hughes put a theoretical basis for the relationship between the number of features and the number of training samples with respect to generalization [104]. There are two important observations to be made. First, even with an infinite training set, there is a saturation effect after a certain number of features. Second, the accuracy reaches a maximum, after which it decreases. An intuitive explanation of this curse of dimensionality (or Hughes effect) might be as follows. As the dimensionality increases, so does the complexity of the statistics that govern the classes. Training of a classifier involves the estimation of these statistics. As only a finite number of

training samples is available, the more complex the statistics are, the less accurate the estimation will be. The combination of a less accurate estimate of the statistics with a saturating separability of the classes, is the reason why the number of features must be limited. Adding the fact that, neighboring bands in hyperspectral data are generally strongly correlated, it seems at first sight that multispectral data is sufficient for most applications. However, the required spectral bands for classification purposes should be specifically adapted to the problem at hand, and rarely correspond to the band settings of any multispectral sensor.

With the appropriate feature extraction technique, a reduced data set can be obtained that still contains the actual information we are interested in. Although different terminologies can be found in the literature, the following definitions apply here. In case of *feature selection*, the number of features is reduced by selecting a subset to some criterion like error of reconstruction or discrimination between classes. *Feature extraction* is a general term to create a feature (sub)set from the vector of available spectral bands. It creates new features by transforming the data in a new feature space, for example as a linear combination of the spectral bands. Finally, we refer to *band selection*, as a special case of feature extraction where a reduced set of new features is created by merging consecutive spectral bands in some way. Band selection is discussed in chapter 4, where we introduce a novel methodology.

In section 2.4.2, we first give a brief overview of some feature selection techniques and how the optimization process can be performed. Transformations, an important tool for feature extraction, are discussed in section 2.4.3. We explain the difference between transforms for representation and discrimination. We zoom in on the wavelet transform, the basis for our generic hyperspectral feature extraction scheme in chapter 3.

2.4.2 Feature selection

A selection scheme needs some measure to assess the performance of a selection[2]. The quality of a feature set is expressed by its ability to separate the classes. This can be done directly, by evaluating the performance of the investigated classifier on a test set[3]. The direct method is used in a *wrapper* selection model [114]. The *filter* selection model [114] uses an indirect method. Without actually implementing the classifier first, the filter model estimates a separability measure (distance) from the statistics of the training set.

A simple straightforward measure is the Euclidean distance (2.1) , where only

[2] Several terms are introduced, depending on the application or selection scheme: performance, cost, criterion, objective function, fitness (Genetic Algorithm) and energy (Simulated Annealing). We use the different terminologies throughout this thesis.

[3] Notice that this test set becomes invalid for testing. After completion of the feature selection, it is no longer "unseen" for the obtained classifier.

mean differences are used, neglecting the covariance of the classes.

$$\text{Euclidean Distance} \qquad D_E = (m_i - m_j)^T (m_i - m_j) \tag{2.1}$$

$$\text{Mahalanobis Distance} \qquad D_M = (m_i - m_j)^T \left(\frac{\sum_i + \sum_j}{2} \right)^{-1} (m_i - m_j) \tag{2.2}$$

$$\text{Bhattacharyya Distance} \qquad D_B = \frac{1}{8} D_M + \frac{1}{2} \ln \frac{\left| \frac{\sum_i + \sum_j}{2} \right|}{\sqrt{|\sum_i||\sum_j|}} \tag{2.3}$$

$$\text{Jeffries-Matusita Distance} \qquad D_{JM} = 2\left(1 - e^{D_B}\right) \tag{2.4}$$

The Mahalanobis distance uses a mean covariance matrix, whereas Bhattacharyya and Jeffreys-Matusita do account for the different class covariances. Equations (2.2)–(2.4) are the respective distances in their Gaussian form. Other measures for calculating the class separability are the Receiver's Operating Characteristic curve (ROC curve, [61]) or Fisher's criterion [77].

A naive search for a set of N features results in a univariate set of M best features. But ordering all the features by some class separability measure and selecting the M best, does not necessarily lead to the best multidimensional feature set. A more successful approach is a multivariate one, in which we try to find the optimal *combination* of available features. The best subset of M variables out of N may be found by evaluating the class separability measure (J) for all possible combinations of M variables. However, the number of all possible combinations, $\binom{N}{M}$, becomes huge even for modest values of N and M. Therefore, several approaches have been suggested to avoid the exhaustive search. We divide them into three groups: sequential, randomized and exponential algorithms. In chapter 3 we introduce a very efficient sequential algorithm in combination with wavelet features: the sequential floating forward search by Pudil [187].

An example of a randomized algorithm is the the *tabu search*, first presented by Glover [84]. Zang used it for feature selection [247]. The tabu search starts with an initial feature set of the desired length M. A random number i from 1 to M determines which feature is scanned. All unused features are tried in position i of the feature set to improve the performance of the feature set. The best performing feature replaces the old feature on that position. Notice that the old feature is always replaced, even if this means a loss in performance of the new feature set. If we would only replace the old feature in case of a performance gain, we were bound to be trapped in the first local minimum encountered.

Exponential algorithms evaluate a number of subsets that grow exponentially with the dimensionality of the search space, e.g., exhaustive search, branch and bound [169]. Sequential algorithms add and remove features sequentially, e.g. sequential forward and backward selection. Simulated annealing and genetic algorithms are randomized algorithms. Simulated annealing (SA) is inspired by a natural process, the annealing process of cooling glass (or recrystallizing metals). At high temperatures, a melt is disordered and is in a high energy state. If cooled

down slowly, the system is in a thermodynamic equilibrium, while the system becomes more ordered (a crystal is formed) and a minimal energy is reached at freezing temperature. If, however, the system is cooled down too quickly, the system gets trapped in a metastable state, reaching a local minimal energy. The link with a mathematical minimization was first made by Pincus in 1970 [179], but it was Kirkpatrick who first proposed the basis of an optimization technique in 1983 [127]. We have selected SA as the optimization scheme in this thesis. It was used in our hyperspectral band selection technique (chapter 4) and for parameter estimation through model inversion (chapter 7). We therefore discuss it in some more detail.

The objective function to be optimized, e.g., measure for class separability, plays the role of the energy state E. The algorithm employs a random search which not only accepts changes that decrease the objective function, but also some changes that increase it. An increase is accepted according to a probability p:

$$p = e^{\frac{-\delta E}{kT}} \tag{2.5}$$

where δE is the increase of the objective function and kT is a control parameter, known as the effective temperature[4]. The outline of a simulated annealing run is presented in algorithm 2.1.

The algorithm is initialized with a high temperature and some initial guess for x and bounded by a possible range. Initially, with high temperatures T, chances to climb up the hill are high. This is desirable, because we are likely to have started near one of the local minima and we do not want to get stuck there. As the system cools down, we keep getting downhill, in order not to overshoot the global minimum.

During the iteration step, the temperature is reduced according to the *annealing schedule*. It defines the number of temperature changes (N_T) and the initial and final temperatures. The initial temperature (kT_0) and final temperature (kT_{N_T-1}) are determined by the choice of the acceptance probability of a given increase in the cost at the start and end of the algorithm respectively, using equation (2.5).

2.4.3 Transformations for feature extraction

Principal component analysis

Principal Component Analysis (PCA [116]) or Karhunen-Loeve transform finds components (features) that are optimal (in a least-squares sense) for representing the data with a reduced feature set. In PCA, the features are constrained to be linear functions of the input variables (Kramer introduced a non-linear variant of PCA [131]). The components are ordered according to the amount of variance explained in the original signal. The first principal component is the linear combination that accounts for the largest amount of variance. Furthermore, all the

[4] k is the Boltzmann's constant and T is the absolute temperature (in Kelvin) in the annealing process

Algorithm 2.1 Simulated Annealing

Current solution ← random solution
Determine the cost E_i for the current solution
best-to-date ← current solution and cost
$kT_i \leftarrow kT_0$
repeat {Metropolis Monte Carlo [159] simulation at this temperature}
  Generate a new solution by replacing an object that is used in the current solution with one that is not.
  Calculate the cost of this new solution, E_{i+1}.
  if $E_{i+1} \leq E_i$ **then** {Determine if this new solution is kept or rejected}
    Accept the new solution, replacing the current solution and E_i.
    if $E_{i+1} < E_{\min}$ **then** {update best-to-date solution and cost}
      best-to-date ← new solution and cost
    end if
  else if The new Boltzmann acceptance probability $p >$ random number $[0-1]$ **then**
    Accept the new solution, replacing the current solution and E_i.
  else
    Reject the new solution, keeping the current
  end if
until New effective temperature $kT_{i+1} \geq kT_{N_T-1}$
Report the best-to-date solution and cost

components are uncorrelated with each other. This makes PCA a suitable transform for data compression. By representing the N dimensional signal by only the first M principal components, most of the variation in the signal is retained.

Related to PCA is the Minimum Noise Fraction (MNF [92]). It performs a principal component transform after noise-whitening of the data. The noise-whitening process results in noise-decorrelated bands each with unit variance. The PCA transform then yields a dataset where the MNF components are ranked in terms of the ratio of the signal to noise rather than the band variance. A prerequisite is the knowledge of the noise covariance matrix. In practice, it is estimated from homogeneous areas in the remote sensing image itself. Both PCA and MNF transforms are often used for hyperspectral imagery [118, 145]. They can be very useful for different applications and are available as a toolbox in most commercial remote sensing processing software. However, to avoid the curse of dimensionality for classification problems in hyperspectral data, these transforms have their shortcomings. They do not take into account class information and feature selection schemes are often a better alternative. At the end of this chapter, we present feature extraction techniques that are based on class discrimination.

The wavelet transform

The wavelet transform is used in many fields today. It is also very well suited for compression of imagery, because the information is statistically concentrated in just a few wavelet coefficients. For this reason, the wavelet transform was chosen for the new image compression standard JPEG 2000. Within the context of this thesis, we are interested in its ability to separate local details at different resolutions, with a highly efficient and easy implementation. We used the wavelet transform to create new features [255]. The discussion of the wavelet transform here aims for an intuitive understanding, without going into too much mathematical details. An in-depth discussion can be found in [50, 154]. Before discussing wavelets, we shortly review the more familiar Fourier transform. We start with the continuous transforms, although the discrete wavelet transform (DWT) is of more practical use for digital signals. But the continuous transforms provide a theoretical basis on which the discrete versions are based.

Time varying signals can be represented in frequency space. Fourier analysis represents a signal in function of time into its sinusoidal components [186]. These components can be resolved by the Fourier series for periodic signals. For non-periodic signals of finite energy, we can analyze frequency components by the Fourier transform $F(\omega)$. The original signal $f(t)$ can also be reconstructed by inverting the Fourier transform.

$$\text{Fourier Transform} \qquad \mathcal{F}(f(t)) = F(\omega) = \int f(t) e^{-i\omega t} \mathrm{d}t \tag{2.6}$$

$$\text{Inverse Fourier Transform} \qquad \mathcal{F}^{-1}(F(\omega)) = f(t) = \frac{1}{2\pi} \int F(\omega) e^{i\omega t} \mathrm{d}\omega \tag{2.7}$$

The problem with the Fourier transform for analyzing hyperspectral (and other) data, is that only frequency information is obtained, while time information is lost. Using Fourier analysis (forward transform), we know all the frequencies present in a signal, but we do not know where they are present. To solve this problem, we could multiply the signal $f(t)$ with a time window $g(t)$:

$$\text{Short Time Fourier Transform} \qquad F(\omega, \tau) = \int f(t) g(t - \tau) e^{i\omega t} \mathrm{d}t \tag{2.8}$$

This windowed Fourier transform is also referred to as the Short Time Fourier Transform (STFT) [80], because the signal function is cut in short time slices. The width of the time slices is defined by the window function $g(t - \tau)$. The STFT, $F(\omega, \tau)$, is localized in time (parameter τ). The problem remains how to choose the window function g. If we are interested in all frequency components at a certain moment in time, we could try to use the Dirac pulse $g(t - \tau) = \delta(t - \tau)$. The time sliced function $f(t)\delta(t - \tau)$ provides a perfect localization in time (everywhere 0 except for $t = \tau$, where it is $f(\tau)$). However, we have lost frequency resolution, since the Fourier transform of such a thin time slice (dirac pulse) contains all possible frequencies. The product of the window function and

the signal function in time, corresponds to a convolution product in the Fourier domain. Convolving $F(\omega)$ with the Fourier transform of the Dirac pulse spreads the spectrum of the signal function over the entire frequency domain. We have the opposite situation in the common Fourier transform, which is the asymptotic result of a STFT with an infinite time window. There, the frequency resolution is perfect, but there is no time resolution. Similar to the uncertainty principle in quantum mechanics concerning momentum and location of moving particles (Heisenberg's uncertainty), we have the situation that it is impossible to know simultaneously the exact frequency and the exact time of occurrence of this frequency in a signal. We only know the time *interval* in which a certain *band* of frequencies exists. The narrower we make the window function in time, the better the time resolution. However, a narrow window introduces new frequencies, with a poor frequency as a result. A more intuitive explanation of this resolution problem is the following. For the STFT analysis, we want to represent the time sliced signal function with a number of waveform (sine/cosine) base functions. If the time slice gets too narrow, it can not contain a waveform base function of meaningful wavelength (with limited bandwidth).

Several attempts were made to solve this dilemma. Gabor [80] used a Gaussian as a window function, with a constant width in time. This means that the same time resolution is obtained for each frequency component of the signal. The wavelet transform provides an elegant solution how to choose the optimal time window function g that is well suited for practical signals. Bearing in mind the intuitive explanation of the resolution problem, it seems that for high frequencies, a better time resolution should be feasible (short waveforms fit into thin time slices). The wavelet transform uses a fully scalable, modulated window function. In the context of the wavelet transform, we use scale s instead of frequency. The forward and inverse continuous wavelet transform are defined as ($s = 1/\text{frequency}$):

$$\text{Wavelet Transform} \qquad \gamma(s,\tau) = \frac{1}{\sqrt{s}} \int f(t) \Psi^* \left(\frac{t-\tau}{s} \right) \mathrm{d}t \tag{2.9}$$

$$\text{Inverse Wavelet Transform} \qquad f(t) = \frac{1}{\sqrt{s}} \iint \gamma(s,\tau) \Psi \left(\frac{t-\tau}{s} \right) \mathrm{d}\tau \mathrm{d}s \tag{2.10}$$

where Ψ^* denotes complex conjugation of Ψ. Similar to the Gabor STFT, a time window is shifted along the signal, and for every position τ and frequency (scale), components (wavelet coefficients) are calculated. Nevertheless, some major differences exist. First, the window function is scaled with a new scaling factor, s, for each iteration. The time window gets narrower for smaller scales (larger frequencies). As a result, the wavelet transform has a good time and poor frequency resolution at high frequencies (low scales), and good frequency and poor time resolution at low frequencies (high scales). This is a desired property in practice. For high frequencies, we need good time resolution, as signals vary a lot at these locations. If signals are very smooth (low frequency content) it is sufficient to have a coarse time resolution, because variation in time is little.

Another difference with the STFT is that the function Ψ, referred to as the

mother wavelet, takes the role of both the window function g and the basis functions. For the wavelet transform, these basis function are just the scaled and translated versions of the mother wavelet Ψ. Other than some *admissibility condition* [160] for the mother wavelet, there are little constraints to the basis functions, which make the wavelet transform versatile for different applications.

The advantage of the new time parameter τ with respect to the normal Fourier transform is clear. We gained some time resolution. Yet, we also introduced more redundancy. For a single time parameter t in the original space, we have obtained two parameters after translation: scale (s) and translation (τ). This is a serious problem for hyperspectral data, that is already highly redundant. Hence, the introduction of *discrete wavelets*, which can only be scaled and translated in discrete steps ($s = s_0^j,\ \tau = k\tau_0 s_0^j$, with j and k integers and $s_0 > 1$). In this case, the wavelet transform is referred to as the *wavelet series decomposition*, because it results in a series of wavelet coefficients.

The usual sampling scheme, that also allows for a perfect reconstruction [50], is the dyadic sampling grid presented in Fig. 2.3 ($s_0 = 2$ and $\tau_0 = 1$).

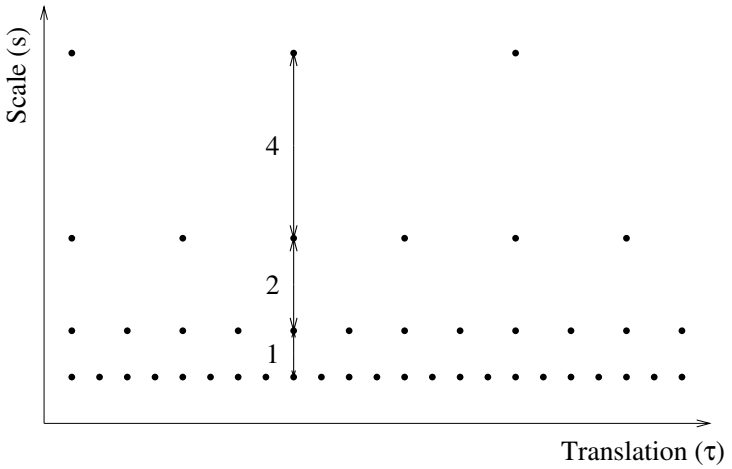


Fig. 2.3: *A typical sampling scheme for scale and translation in the discrete wavelet transform: the dyadic grid.*

For feature extraction, our main interest is class discrimination and signal reconstruction is not really an issue. However, for signal representation, the mother wavelet is chosen to make the discrete wavelets orthogonal to their own dilations and translations. In this case, an arbitrary signal can be reconstructed by summing the orthogonal wavelet basis functions, weighted by the wavelet transform coefficients.

The introduction of discrete scale and translation steps created discrete wavelets.

Nevertheless, they are still (piecewise) continuous in time, t. The wavelet transform must be made discrete to obtain the Discrete Wavelet Transform (DWT) and to be applicable to the digital domain [236]. A practical implementation of the DWT is based on filter banks [153]. The scaling operation halves the bandwidth at every subsequent scale s of the dyadic sampling grid, allowing for a full spectral coverage of the input signal function $f(s)$. However, as we can keep on splitting the bandwidth in two, we would end up with an infinite number of scales. This is avoided by using the scaling function for filtering the lowest level of the transform. A series of convolutions of the signal function with scaled and translated versions of the mother wavelet is thus equivalent to separating the signal function into complementary frequency bands. This can be performed with a combination of quadrature mirror filters [65], or filter bank, including a *scaling filter* (low-pass) and *wavelet filter* (high-pass).

Because we are dealing with spectra (hyperspectral signatures), we now use the discrete wavelength x instead of a discrete time t[5] and the discrete translation z instead of τ for our discussion on filter banks. The filter bank decomposes a signal $f(x)$ at different scales $s = 2^j$ into a lower resolution (approximation) signal $A_{2^j} f$ and a detail signal $D_{2^j} f$. This is done with a uniform sampling of convolution products. At each scale[6], $j = 1 \leq j \leq J$; and discrete translation $z \in \mathbb{Z}$:

$$\begin{aligned} A_{2^j} f(z) &= (f(x) \otimes h_{2^j})\,(2^j z) \\ D_{2^j} f(z) &= (f(x) \otimes g_{2^j})\,(2^j z) \end{aligned} \tag{2.11}$$

Here, h and g are the low and bandpass filters for analysis, operated on discrete signals ($x = 1, \ldots, N$) with a pyramidal algorithm, making use of quadrature mirror filters. The process consists out of convolutions and subsampling, by skipping one sample out of two. $D_{2^j} f(x)$ are called the detail coefficients at scale j.

Signals are represented with variable resolutions when we recursively apply the low-pass filter h and high-pass filter g on the output of the analysis filter bank (Fig. 2.4). As scale increases at each iteration, we move to coarser representations of the input signal. The DWT coefficients consist of the output of the high pass filter at each scale. In Fig. 2.5, only the low-pass filter output is considered for the next iteration. However, other variants consider both low-pass and high-pass filter output (complete tree or wavelet packets).

Wavelet transforms have been applied to many signal types, including cardiograms, seismographic and audio signals. The two dimensional wavelet transform is very popular for image processing. Typically, the one dimensional DWT is applied on rows and columns sequentially. This can easily be extended to three dimensions, covering the spectral domain [117]. However, in this thesis, we focus on the spectral domain only. In chapter 3, we therefore apply the one dimensional wavelet

[5] Terminology is confusing when analyzing spectra instead of time varying signals. For time varying signals, the term frequency refers to the periodicity of the base functions (sines and cosines in case of Fourier transform). However, analyzing spectral signatures, time is replaced with wavelength, which is equivalent to frequency, referring to the electromagnetic spectrum.

[6] From now, we abuse the terminology for scale when we refer to j. The correct definition for the dyadic scale remains $s = 2^j$.

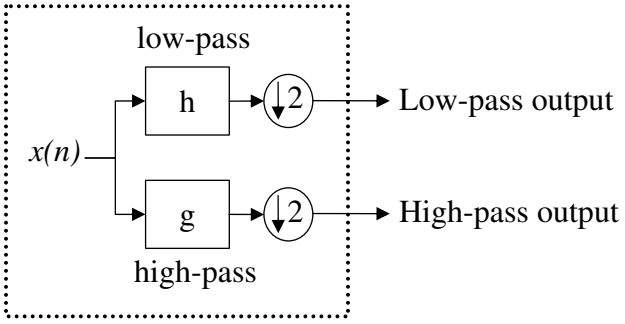


Fig. 2.4: *Basic element of the filter bank for implementing the DWT. The low- and high-pass filters are represented by h and g respectively. After filtering, the output is downsampled with a factor of 2*

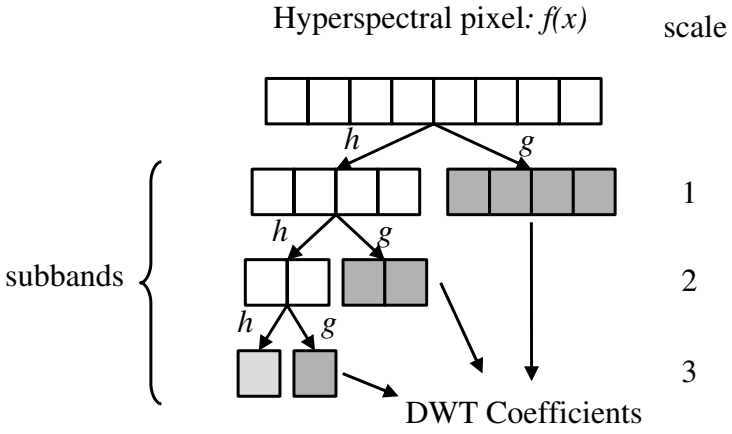


Fig. 2.5: *Iterated filter bank. The hyperspectral vector is the discrete input signal $f(x)$. The DWT coefficients consist of the high-pass filter output (dark-shaded). The low-pass filter outputs are recursively filtered by the basic element in Fig. 2.4. At each iteration, the scale is increased until a single pixel element is obtained (light-shaded low pass filter output at bottom left).*

transform, as illustrated in Fig. 2.5. The discrete wavelet coefficients derived from the DWT of a spectral signature can then be used as a new feature set for each pixel.

The multi-resolution representation of the wavelet transform allows to exploit the available spectrum, even with a limited set of selected features. Wavelet coefficients from lower scales can be selected for complex classification problems, where classes can only be distinguished with detailed spectral information. If combined with coefficients from other scales, a powerful feature set can be obtained, as illustrated in chapter 3.

Linear discriminant analysis

Linear Discriminant Analysis (LDA [68]) projects a d-dimensional space to $C-1$ features, with C the number of classes. Typically, $d \gg C$ and thus we can look at LDA as a transform for data reduction. Unlike the previous transformations, LDA seeks directions that are efficient for discriminating classes. Obviously, for classification purposes, this is exactly what we are interested in. However, two limitations apply. First, by definition, the dimension of the new feature space depends on the number of classes. Second, without prior feature selection, the dimensionality of the original data can be too high for a direct application of LDA. LDA is well described in [61], and we only summarize the basic idea below.

The linear projection involved in LDA is represented by $C-1$ discriminant functions $y_i - \boldsymbol{w}_i^t \boldsymbol{x}$, where $\boldsymbol{x}$ is a vector representing the input spectrum and the vector $\boldsymbol{w}_i$ projects $\boldsymbol{x}$ to new features y_i. If we order the vectors $\boldsymbol{w}_i$ as the columns of a matrix $\mathbf{W}$, we can rewrite the projection in matrix form:

$$\boldsymbol{y} = \mathbf{W}^t \boldsymbol{x} \tag{2.12}$$

Because optimal class discrimination is aimed for, we need to introduce some class information. This is done by defining the *within-class scatter* $\mathbf{S_W}$ and the *between-class scatter* $\mathbf{S_B}$:

$$\mathbf{S_B} = \sum_{i=1}^{C} n_i (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^t. \tag{2.13}$$

$$\mathbf{S_W} = \sum_{i=1}^{C} \sum_{j=1}^{n_i} (\boldsymbol{x}_j - \boldsymbol{m}_i)(\boldsymbol{x}_j - \boldsymbol{m}_i)^t, \tag{2.14}$$

where $\boldsymbol{m}_i$ and $\boldsymbol{m}$ are the mean spectra of each class and the mean spectrum overall, and n_i represent the number of spectra in class ω_i. The within-class scatter is the sum of scatter matrices $\mathbf{S_i}$ in each class ω_i. It shows the scatter of samples around their respective class expected vectors and thus should be minimized after transformation. The between-class scatter refers to the scatter of the expected vectors around the mixture mean, which should be maximized.

We therefore look for a projection maximizing the function:

$$J(\mathbf{W}) = \frac{|\mathbf{W}^t \mathbf{S_B} \mathbf{W}|}{|\mathbf{W^t} \mathbf{S_W} \mathbf{W}|}. \tag{2.15}$$

The optimal $\mathbf{W}$ can be found by solving the generalized eigenvector problem:

$$\mathbf{S_B}\,\boldsymbol{w}_i = \lambda_i\,\mathbf{S_W}\,\boldsymbol{w}_i, \tag{2.16}$$

where λ_i represent the eigenvalues and $\boldsymbol{w}_i$ the eigenvectors. In the special case of two classes, it is unnecessary to solve for the eigenvalues and eigenvectors of ${\mathbf{S_W}}^{-1}\,\mathbf{S_B}$. In this case, the matrix $\mathbf{W}$ consists of a single vector ($\boldsymbol{w}_1$), and is immediately obtained from [61]:

$$\boldsymbol{w} = {\mathbf{S_W}}^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2) \tag{2.17}$$

The idea of LDA is old, but the methodology is still successfully used in pattern recognition problems. There are a few shortcomings though:

- It does not take into account class covariance differences
- It is suitable for linearly separable classes only
- As a feature reduction method, it only produces optimal features up to one less than the number of classes.

In [140], a procedure called Decision Boundary Feature Extraction (DBFE) overcomes the first two shortcomings. From the decision boundary, it derives the minimum number of features required to achieve the same accuracy as in the original space (defined as *discriminantly informative features*). It also suggests a method to calculate such features.

In chapter 5, we propose two techniques to overcome the third shortcoming. The first technique is to apply feature selection before LDA. The other is the use of classifier ensembles. It solves the multi-class problem as a number of binary class problems, projecting the feature space on a single variable.

2.5 Classification

2.5.1 Introduction

A classifier divides the feature space into a set of non-overlapping regions, one for each class. The classes are mapped to unique *labels*, describing the classes. Within this context, the input of a classification problem is a remote sensing image and the objective is to create a thematic map. Each image element (pixel or segment) obtains a label according to some theme. The labels are represented by a (color) code, explained in the *legend*.

Classification can be supervised or unsupervised. In this thesis, we focus on the supervised approach, requiring a labeled training set. The objective of unsupervised classification, or clustering, is to construct decision boundaries based on unlabeled training data. In this case, the boundaries are based on natural groupings, or clusters, in a multidimensional dataset. Some common unsupervised classification schemes are: k-means [61], Iterative Self Organizing Data Analysis Technique

(ISODATA [112]) and hierarchical clustering [115]. An overview of the state of the art unsupervised techniques can be found in the dissertation of De Backer [53]. Semi-supervised algorithms also exist. They learn from both labeled and unlabeled data. To illustrate the idea, we give the example of co-training [11]. An initial training set is used to learn a (weak) classifier. This is then applied to all unlabeled samples, and co-training detects those samples on which the classifier makes most confident predictions. These high-confident samples are labeled with the estimated class labels and added to the initial training set. Based on the new training set, a new classifier is trained, and the whole process is repeated for several iterations. Xu [241] and Bruzzone [28] used a similar approach in on transductive support vector machines.

Classifiers can be ordered in parametric and nonparametric classifiers. The parametric classifiers assume that the form of the underlying density function that characterizes the class distribution is known. The probability density function $p(x)$ is then represented in terms of a specific functional form which contains a number of adjustable parameters. During training, the values of the parameters are optimized to give the best fit to the data. Nonparametric classifiers do not assume any form of the underlying probability density of the classes. Some examples of nonparametric classifiers are Parzen windows [177], k-nearest neighbor [178], and artificial neural networks (NN [22]). The Support Vector Machine (SVM [230]) is a relatively new type of classifier with growing popularity. Although a linear machine, it can solve nonlinear problems by first projecting the original feature space in a higher dimension using a nonlinear mapping function $\boldsymbol{\phi}(\boldsymbol{x})$. Then, the SVM finds a linear separating hyperplane in the higher dimensional space. SVM makes use of the *kernel trick* [4] to replace the expensive computation of the dot product $(\boldsymbol{\phi}(\boldsymbol{x}) \cdot \boldsymbol{\phi}(\boldsymbol{x}'))$ in the high dimensional space with an equivalent kernel function $k(\boldsymbol{x}, \boldsymbol{x}')$ in the original feature space.

Finally, classifiers can be grouped in pixel based and object oriented classifiers. The latter use contextual information to segment the image into meaningful areas, which are then classified. They have gained a lot of interest. This is related to the ever increasing spatial resolution of remote sensing data on one hand, and the availability of commercial of the shelf software packages on the other. Contextual information can also be used to improve the results after the classification step in a pixel based classifier, which is the approach we use in this thesis.

Parametric classifiers are first introduced in section 2.5.2, assuming a Gaussian distribution. To increase the performance of such a classifier, some techniques are presented in section 2.5.3. We adopted this approach in chapters 3 and 5. By combining an efficient feature selection scheme with an architecture of classifiers with low complexity, a powerful classification framework can be obtained.

2.5.2 Statistical classifiers and Bayes decision rule

We follow the mathematical notation presented in [61], describing a classifier as a number of (discriminant) functions g_i of an observed feature vector $\boldsymbol{x}$. The

function g_i that maximizes x, defines the class ω_i to which x belongs

$$\boldsymbol{x} \in \omega_i \iff g_i(\boldsymbol{x}) \geq g_j(\boldsymbol{x}), \forall j = 1, \ldots, C \tag{2.18}$$

Using Bayes decision rule [61], we can write the discriminant function $g_i(\boldsymbol{x})$ as:

$$g_i(\boldsymbol{x}) = p(\boldsymbol{x}|\omega_i)P(\omega_i), \tag{2.19}$$

where $p(x|\omega_i)$ represent the class-conditional densities and $P(\omega_i)$ are the class prior probabilities. For the ease of notation, we assume from now on equal prior probabilities. In this case, we can leave out $P(\omega_i)$ in (2.19). Such a classifier is referred to as the *Bayes classifier* and is the best classifier we can get (in terms of minimum error). However, its implementation for real world problems can be problematic, because the class probability density functions must be estimated from a training set of finite size. Assuming normal distributions for all classes simplifies this problem a lot, with the only parameters to estimate, the elements of $\boldsymbol{\mu}_i$ and the independent elements of the covariance matrix $\boldsymbol{\Sigma_i}$. We thus obtain the *quadratic classifier*[7]:

$$g_i(\boldsymbol{x}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma_i}| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) \tag{2.20}$$

Still, this leaves us with $d + d(d-1)/2$ parameters to estimate for each class (i.e., the elements of $\boldsymbol{\mu}$ and the independent elements of the covariance matrix $\boldsymbol{\Sigma}$). In other words, for a hyperspectral sensor with a moderate number of 32 bands, we need to estimate 528 parameters for each class. For most practical problems, this is too much.

A further simplification is to assume equal covariance matrices for all classes, obtaining *Fisher's Linear Discriminant*:

$$g_i(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i \tag{2.21}$$

If, in addition, no correlation exists and all features have equal variance, we can simply implement the Bayes classifier as the *Minimum Distance to Means*:

$$g_i(\boldsymbol{x}) = -(\boldsymbol{x}-\boldsymbol{\mu}_i)^T(\boldsymbol{x}-\boldsymbol{\mu}_i) \tag{2.22}$$

Complex problems will not be solved by such a simple classifier in general. However, as less parameters must be estimated, a simple classifier is more likely to generalize a classification problem if only a reduced training set is available. From our experience on classification of hyperspectral images, the (linear) parametric classifier, assuming a Gaussian distribution of the classes, offers a good balance between complexity and the required sample size [253]. The case studies discussed in chapters 3, 4 and 5 will be based on this classifier.

[7] Replacing $g_i(\boldsymbol{x})$ with $\ln(g_i(\boldsymbol{x}))$ does not change the classification result.

2.5.3 Ensemble classification and decision trees

By making intelligent use of training samples, and combining multiple classifiers, the performance of the final classifier can be boosted. We already mentioned how unlabeled samples can be introduced in our training scheme (as in co-training). Ensemble classifiers on the other hand, create individual classifiers based on a subset of the existing training samples. By selecting the subsets in an intelligent way and combining the votes of the individual classifier components, an improved ensemble classifier can be designed [35]. A first example of an ensemble classifier is obtained via bootstrap aggregation, or *bagging* [26]. A training dataset is randomly resampled using a bootstrap mechanism[8]. The classifier is trained for each new training set obtained, resulting in different components of the classifier. The final (ensemble) classifier uses the vote of each component for its classification decision. Bagging can work for larger training datasets. Only then the random subset can have an acceptable training size n while still offering diversity from the other subsets [223].

Boosting [197] is similar to bagging, but resamples the n training data samples without replacement. A first subset of $n_1 < n$ samples is selected, training a first classifier g_1. A second subset of $n_2 < n - n_1$ samples is created to train g_2. The n_2 samples are chosen randomly, with the constraint that, on average, half of the samples are classified incorrectly by g_1. Finally, a third classifier is trained for a remaining subset of those $n_3 \leq n - n_2 - n_1$ samples where the first two classifiers disagree. Hence, the second classifier g_2 is complementary to g_1 and g_3 is complementary to the combination of g_1 and g_2. The final decision is of the ensemble classifier g is based on g_1 and g_2 if an agreement is reached. Else, the third classifier g_3 is used.

A variation of boosting is AdaBoost [75], which iteratively creates a user defined number of individual classifiers. During the iteration step, a new classifier g_{i+1} is trained with samples of the training dataset, selected with a probability that depends on the performance of the previous classifier g_i for that sample. If a sample is (un)correctly classified by g_i, the probability of being selected for training g_{i+1} is low (high). As a result, the subsequent classifiers focus on the difficult samples. The ensemble classifier is a weighted sum of the outputs of the individual classifiers g_i.

Another technique to improve the diversity of the classifier components, is the random subspace method [102]. In this case, each classifier component uses a reduced feature set, randomly selected from the input feature space.

Decision Tree Classifiers (DTC [214, 194]) break down a complex decision-making process into a collection of simpler decisions. A DTC is composed of nodes and edges, corresponding to decision rules and their results respectively. A node without outgoing edges is called a leaf and reflects a class label. The root has no incoming edges. The classification algorithm starts at the decision rule in the root. Depending on the results of the rules in the nodes, a path is followed in the tree until a leaf is reached and a class label is known.

[8] randomly selecting n points from the training set with replacement

One of the strongest points of DTC for hyperspectral classification is the flexibility of choosing different reduced subsets of features at different nodes of the tree, such that a minimal feature subset is selected for optimal class discrimination in that node. For this reason, the DTC can outperform the single-stage classifier [226]. Savavian discussed the design of a decision tree classifier [194].

Binary Hierarchical Classifiers (BHC) are a special type of DTC. Each node has no more than two (binary) outgoing edges, i.e., a decision between only two classes must be made. The decision at intermediate nodes allows rejection of class labels (hierarchical). BHC have been proposed to decompose a C-class problem into a binary hierarchy of $(C-1)$ simpler two-class problems [133].

2.6 Validation

2.6.1 Introduction

Any classification result or estimated parameter is worthless without knowing its accuracy. We need to know how reliable the results are. This is the objective of validation, or accuracy assessment. We could define accuracy as the closeness of the estimated label or value to the truth. However, there is a fundamental problem with this definition. The *truth*, even if it exists, can never be measured. Many authors do not want to use this term in the context of accuracy assessment [21, 121]. Reference measurement techniques are far from perfect. In most cases, only a relative accuracy is obtained, comparing the results obtained from remote sensing to reference methods (in situ measurements). Needless to say, the reference methods must have high accuracy for an assessment to make sense [43]. A more careful definition for accuracy measure is used by McCloy: "estimate of closeness of the estimated value to the best estimate that can be made of that value" [156].

Another aspect of accuracy assessment is how representative a measure is. In remote sensing, we are interested in the results for an entire area, whilst only a few reference measurements are obtained. By knowing the few reference measurements, the true value for the entire area can be estimated, but is not known. This problem has been studied thoroughly in statistics, where we need to estimate an unknown parameter from a large population, for which only a limited sample is available. Evidently, we use tools from statistics for the accuracy assessment [207]. Strahler has given a comprehensive overview for validating (global) landcover maps [211].

According to Stehman, a statistically rigorous accuracy assessment consists of three basic components: sampling design, labeling (or response design), and the actual calculation of the accuracy measure (analysis and estimation) [209]. Although the theory was developed for thematic maps, most of the ideas are generally applicable. Sampling design is discussed in section 2.6.2. We also introduce cross validation as a technique to avoid biased validation results. Some common accuracy measures are presented in section 2.6.3.

2.6.2 Sampling and response design

The first step in the experimental design is the *sampling design*. It creates the rules (protocol) how to select a sample from a population. The sample consists of *sampling units*, the fundamental units on which the accuracy assessment is based[9]. Point and areal sample units are distinguished. Most familiar areal sample units for classification results are pixels and vector based polygons of specific area. Points are continuous sampling units, which discard the representation of a pixel in terms of coverage. Selecting a sampling unit is part of the sampling design. Spatial resolution of the image and georeference accuracy must be taken into account.

The next step in the sampling design is to define the *sampling protocol*, describing the rules how the sample is collected. Smith highly recommends a *probability sampling design*, because of its objectivity [206]. This means in the first place that no sample units have zero probability. Otherwise, the validation is not representative for the entire map. A common error is to exclude areas with difficult access or to choose only *nice* sample units which are homogeneous. In addition, the probability of selecting a particular sample unit must be known. This is important for the design of a consistent estimate of the accuracy measure. For non-uniform distributions, each sample unit must have its own weight in the accuracy measure.

Systematic sampling and random sampling are both probability designs. The systematic sampling uses a regular grid, whereas a random sampling selects a random sample unit. A simple random sampling corresponds to a uniform distribution where each sample unit has equal probability of being selected. For both designs a stratified variant exists, where a sample is obtained in each stratum. A common approach for a stratified sampling is to allocate an equal sample size for each class label, treating all classes as equally important. If high priority classes are identified by the objectives, an unbalanced sample allocation might be an option.

Ground reference is limited in practice due to the expensive field work involved. Careful sampling design can reduce cost, without major quality loss. As an example, spatially clustering the reference sample units reduces travel time. The cost for collecting clusters of 5×5 pixels is usually less than if the 25 pixels were collected with a simple random sampling. However, the information per pixel is typically less due to spatial correlation between the pixels within a cluster. Moreover, ad hoc based clustering, inspired by low cost practices or easy access (sampling near roads or pathways) is in conflict with randomization and must be avoided. In the case study on dune vegetation mapping, the sample was collected in a clustered sampling of 5×5 pixels. However, only the center 9 pixels were retained as sample units, because geometrical accuracy was limited. The clusters were randomly selected from the entire mapped area. All classes were adequately represented because of the large sample size. The sample size was unbalanced for the final classes obtained from vegetation analysis.

In case of a supervised technique, both a training sample and a test (validation) sample must be available. For practical reasons, both samples are often collected

[9] within this context, sample is used for a representative set of sample units from some population. However, sample and sample unit are sometimes used interchangeably, also in this thesis

during the same field campaign. However, accuracy assessment must be performed with a test set, that is different from the training set to avoid a biased accuracy result [213, 209, 208]. Hence, the test set must be hold out from the training set. Holding out a large subset for testing significantly reduces the available training sample, which is in conflict with the need for abundant training data. If the training sample left is too small, we can not obtain a reliable estimate of the class statistics. Conversely, if the hold out test sample is too small, the estimate of the true accuracy measure is not reliable.

A random split of the reference samples in a training and test sample is referred to as an *m-fold cross-validation* [137]. m disjoint (sub)samples of equal size are selected from the original total amount of reference sample units. Each disjoint test sample provides a new classification result, using a classifier that is trained with the remaining reference sample units. An estimate of the true probability of error is the mean of the m errors. A special case of this approach is the *leave-one-out* or *Jackknife* estimation. In this case, a single test sample unit is used until all individual reference sample units have been hold out once. An advantage of this approach is that all classifiers only differ to the actual classifier tested by a single training sample unit. The *bootstrap* procedure also uses different test samples for estimating the true classification accuracy, but test samples are generated randomly and need not to be disjoint.

Cross validation is very common, but implies a serious risk for violating the probability sampling conditions. For some classifiers, homogeneous sample units corresponding to *pure pixels* are preferred. If field work is performed with this in mind, a serious bias will be introduced by not taking into account *mixed pixels* (section 5.4).

The final step in the experimental design, the *response design*, consists of labeling the sample units collected during the sampling design. In case of a thematic map, the sampling units must be assigned to the reference landcover. For parameter estimation, reference values for the continuous variables are typically obtained using an independent reference method.

2.6.3 Accuracy measures

Accuracy measures can be ordered in three categories. The most subjective is according to a "looks good" standard. The problem with such a subjective measure, is that it is difficult to quantify. As a result, more objective measures have been introduced by Congalton [44] and Foody [71]. The next category is based on comparisons of the areal extent of the classes. In this case, the proportion of the classes are checked with those of the ground reference. Although a more objective approach, it does not take into account the locations at which class labels have been assigned.

The confusion matrix, also referred to as error or contingency matrix, is a very useful tool for accuracy assessment. In fact, most recent accuracy measures have been derived from its information content [30, 32, 71]. Fig. 2.6 is a schematic representation of the confusion matrix for 4 classes with labels (A, B, C, D). The

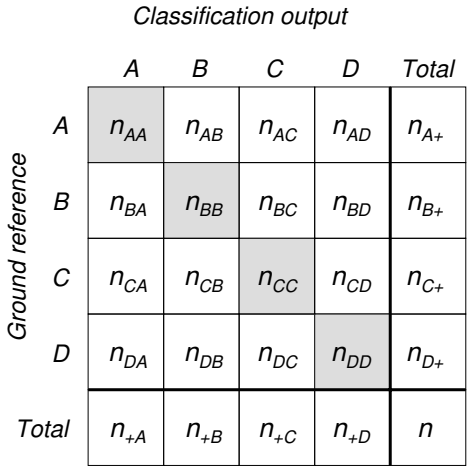


Fig. 2.6: *Schematic representation of the confusion matrix for 4 classes (labels A, B, C, D).*

ground reference class labels are shown in the rows 1–4. The corresponding labels from the classification output are shown in columns 1–4[10]. The matrix elements, $n_{XY}, (X, Y = A, B, C, D)$, represent the number of sample units from class X that has been assigned to class Y. The total amount of sample units in the ground reference is n, from which $n_{A+} = \sum_{k=1}^{4} n_{Ak}$ are in class A. The integer k represents the class labels A to D ($1 = A, \ldots, 4 = D$). The shaded diagonal matrix elements are the number of sample units that have been assigned to the correct label. The off-diagonal elements are the number of misclassifications. For example, n_{AB} is the number of ground reference sample units in class A that have been assigned to class B.

A number of commonly used accuracy measures can be derived directly from the confusion matrix. The equations (2.23)–(2.26) assume a simple random sampling design (for a total of C classes). Alternative expressions must be applied if

[10] Some authors use a different convention where rows represent classification output and columns represent ground reference

sample units are selected with non-uniform probability distributions.

$$\text{Overall accuracy} = \frac{\sum_{k=1}^{C} n_{kk}}{n} \tag{2.23}$$

$$\text{Producer's accuracy} = \frac{n_{XX}}{n_{X+}} \tag{2.24}$$

$$\text{User's accuracy} = \frac{n_{XX}}{n_{+X}} \tag{2.25}$$

$$\text{Kappa coefficient} = \frac{n\sum_{k=1}^{C} n_{kk} - \sum_{k=1}^{C} n_{k+}n_{+k}}{n^2 - \sum_{k=1}^{C} n_{k+}n_{+k}} \tag{2.26}$$

The *overall accuracy* is the proportion of correctly classified sample units over all classes. Two measures are introduced on a class level. We take class A as an example. The *producer's accuracy* of class A is defined as the proportion of correctly classified *ground reference* sample units (calculated on elements in row A). All disagreements ($n_{AX}, X \neq A$) represent errors of *omission*. A similar reasoning defines the *user's accuracy* of a class A. It is the proportion of correctly classified sample units *labeled* as A (calculated on elements in column A). All disagreements ($n_{XA}, X \neq A$) represent errors of *commission*. The user's accuracy is relevant from the user's perspective of a thematic map: what is the chance for a mapped pixel to be correct? Conversely, the producer of the map wants to know what proportion of the available ground reference was correctly classified. However, one can argue that user's and producer's accuracies are equally important to both, because no single measure provides sufficient information on its own. This can easily be illustrated from an unbalanced sampling design. Suppose the number of reference sample units in A is ten times larger than in B. If only 10% of the sample units in A are omitted and misclassified as B, the user's accuracy of class B drops below 50%, regardless of the performance of the classifier for reference sample units in class B, because at least half of the sample units labeled as B were wrong.

Some of the labels may have been allocated correctly purely by chance. For two classes, the random classifier has a 50% chance for agreement. The Kappa coefficient of Cohen [40] takes into account this expected agreements and could therefore be considered as a better single accuracy measure [224, 43, 181]. However, no single accuracy measure can provide all the information contained in the confusion matrix. Moreover, all errors are treated as equally wrong. For most applications, this is not the case. For example, confusing different vegetation species within the same family can possibly considered as minor errors, while confusion between vegetation and water is not. The complete confusion matrix can reveal interclass confusion and might suggest to add additional discriminatory information or, if not feasible, to aggregate non-separable classes. Several authors even developed an approach to use the confusion matrix for non-crisp (fuzzy) classification results [143].

Even though the confusion matrix is very useful, it can not characterize all errors. The spatial distribution of error must be addressed by complementary

tools. One of them is the confidence-based quality assessment [70]. It provides a measure of classification quality at each pixel.

The allocation of the discrete labels is either right or wrong[11]. At the most we could give more weight to some errors than others. To assess estimates $\hat{x}$ of continuous parameters x, appropriate measures must be used. Common are the Root Mean Square Error (RMSE) and its relative variant (RRMSE). A graphical representation is obtained with the *scatter plot*, where each sample unit is represented as a dot. A high scattering indicates a low accuracy, while a perfect estimation results in all dots on the diagonal $\hat{x} = x$. The correlation coefficient, R^2 measures the linearity of the scatter plot.

A different type of error is related to the accuracy of location. This is not only important to delineate boundaries of landcover classes, but it also has an impact on the product accuracy. Accurate georeferenced image data is crucial to link the image to the reference data. This can be ground measurements collected in the field with positioning equipment (GPS). If the corresponding pixels in the remote sensing image can not be retrieved without error, the calculated accuracy is biased. Even worse, in a supervised approach, the classifier is trained with the wrong data.

Finally, we define the confidence limits (lower and upper boundaries) for the estimate $\hat{p}$ of the true classification accuracy p, according to a 95% confidence level as:

$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}}, \tag{2.27}$$

where the distribution of $\hat{p}$ is approximated by a Gaussian with mean k/n and standard deviation $\sqrt{p(1-p)/n}$.

[11] except for fuzzy classifier and unmixing results, which are continuous parameters

Chapter 3

A generic hyperspectral wavelet based feature extraction scheme

> Things should be made as simple as possible, but not simpler
>
> Albert Einstein

3.1 Introduction

In this section, we present a generic feature extraction scheme for supervised classification problems dealing with hyperspectral data. Rather than a novel technique, it is based on wavelet coefficients, combined to an efficient state of the art feature selection scheme. We present a new approach to vegetation stress detection, taking full advantage of the valuable information contained in the spectral signature of vegetation. As a case study, we apply the presented approach on stress detection in apple orchards. Nevertheless, the presented feature extraction scheme is *generic*, in a sense that it performs well, regardless of the nature and complexity of the application and sensor characteristics (i.e., number and width of spectral bands) Ad hoc features such as narrow band indices are designed with specific applications in mind and are clearly not generic. They also depend on sensor characteristics. Moreover, valuable information of the available spectrum is lost. This is also the case if a limited number of spectral bands is selected. On the other hand, applying all the available spectral bands is often not feasible either, due to the curse of dimensionality, as explained in chapter 2.

In section 3.2, we present different features for vegetation. We use the vegetation spectrum to demonstrate the difference between ad hoc and generic features. Some common vegetation indices are proposed first. Then we show how wavelet based features can be derived by analyzing the vegetation spectrum. The feature selection scheme is presented in section 3.3. As a case study, we apply the proposed methodology to the problem of stress detection in apple orchards in section 3.4.

3.2 Features for vegetation

3.2.1 Vegetation indices

The most straightforward feature set is given by the reflectance in each spectral band. The high correlation between them and the problem of high dimensional data, suggest to concentrate only on a selected number of bands. Often, a linear combination of bands or a normalized difference of bands is retained. When appropriately chosen, this has the additional advantage of cancellation of noise (e.g., atmospheric effects). The problem is converted into finding the most discriminating bands or combination of bands for the application at hand.

Features can also be dictated from physical indicators. A close look at the vegetation signature in Fig. 3.1, reveals a sharp rise in reflectance between 670 and 780 nm, the *red-edge* [42, 103] (a more detailed discussion on the spectral signature of vegetation is given in section 6). Vegetation stress can sometimes be monitored via a shift of the red edge toward shorter wavelengths, also referred to as the *blue shift*. Therefore, all kind of indices characterizing the red-edge have been proposed in the literature for detecting vegetation stress [33, 24, 162]. Some

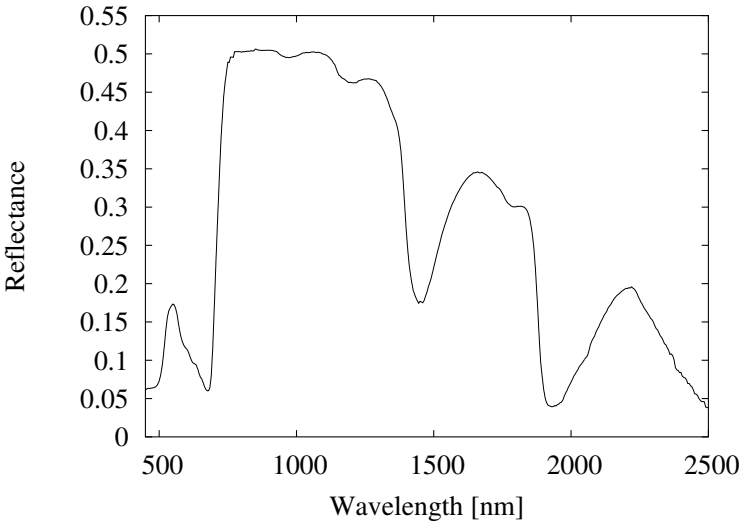


Fig. 3.1: *Example of a leaf reflectance signature.*

are shown in Table 3.1. This kind of feature is very popular thanks to the direct relationship to physical parameters, easing the interpretation. As a drawback for each application, specific indicators must be derived by experts in that matter. In practice, one will revert to the features found in the literature which may be outdated or not customized to the application at hand. Some of the stress indicators have been designed for multispectral sensors, and might be not optimal for higher number of spectral bands.

Tab. 3.1: *Vegetation indices*

narrow band reflectance ratios	$R740/R720$
	$R750/R550$
	$R750/R700$
	$R605/R760$
	$R695/R760$
	$R695/R670$
	$R710/R760$
NDVI (norm. diff. vegetation index)	$(R777 - R665) \,/\, (R777 + R665)$
NDPI (norm. diff. pigment index)	$(R445 - R680) \,/\, (R445 + R680)$
SIPI (structure independent pigment index)	$(R800 - R445) \,/\, (R800 - R680)$
EGFN (edge-green first deriv. norm. diff.)	$(dRE - dG) \,/\, (dRE + dG)$
EGFR (edge-green first deriv. ratio)	(dRE/dG)
ratio of first derivatives	$dR715/dR705$
Red edge Gaussian fitting	$\lambda_0, \lambda_p, R_0, R_s, \sigma$

dG: maximum of the first derivative of reflectance in the green
dRE: maximum of the first derivative of reflectance in the red edge

3.2.2 Wavelet based features

The desire to generate generic features, independent of the spectral resolution justifies the use of multiresolution representations such as the ones generated from wavelet transforms. Wavelet features have been used frequently for classification purposes. Recently, it has been shown that wavelet coefficients of hyperspectral data are useful features, e.g., for detecting weeds [129].

Wavelets represent the reflectance spectrum at different scales. Hence a representation is obtained that provides a detailed and in the same time a global view of the signal function. This is a useful property for the analysis of hyperspectral reflectance signatures, where too much detailed information can obscure some features that are obvious at some particular scale. In chapter 3.4, we show that wavelet analysis can improve feature extraction for classification problems [255].

To demonstrate the features of the discrete wavelet transform, we calculated the DWT coefficients for the vegetation signature of Fig. 3.1 using the Haar wavelet.

The wavelet coefficients corresponding to the DWT of the vegetation signature in Fig. 3.1 are presented in Fig. 3.2. We used a logarithmic scale because of the

exponential decrease of energy content for lower scales. Coefficients from each scale correspond to specific spectral resolutions. If the coefficients are used as features, the scales from which features are selected, reveal information about relevant spectral resolution levels. This allows, for example, to study the required spectral resolution for vegetation stress.

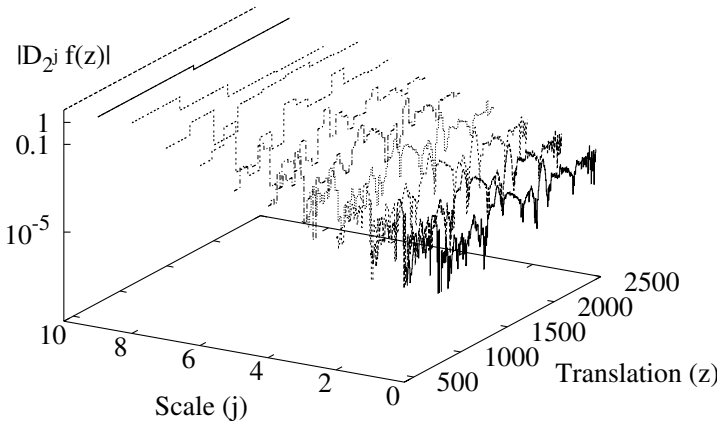


Fig. 3.2: *Absolute values of the discrete wavelet transform coefficients for the vegetation signature of Fig. 3.1.*

In Fig. 3.3, we focus on scales corresponding to $j = 6$ and $j = 7$ (with respective spectral resolutions of 64 and 128 nm). The high pass filter corresponding to the HAAR wavelet identifies the sharp edges in vegetation spectral signature. For a scale $j = 6$, three edges can be clearly distinguished: the red-edge [42, 103] near 700 nm, and the water absorption features around 1400 nm and 1900 nm. The water absorption features are blurred in the coarser signal at scale $j = 7$, due to the additional low pass filter operation. Notice that for this discussion, we only considered a single spectrum. If we are interested in the different behavior of two spectral signatures (as in classification), other scales can become more important.

The wavelet transform itself does not reduce the dimensionality of the feature space. The total number of detail coefficients (together with the residual low pass signal) equals N, the number of hyperspectral bands. For most applications, the large number of features in hyperspectral data leads to dimensionality problems. Using the complete set of detail coefficients as new features will not change this. One option is to derive statistical features from the detail spectra. Bruce applied energy features for vegetation monitoring [27]. The energy features are calculated

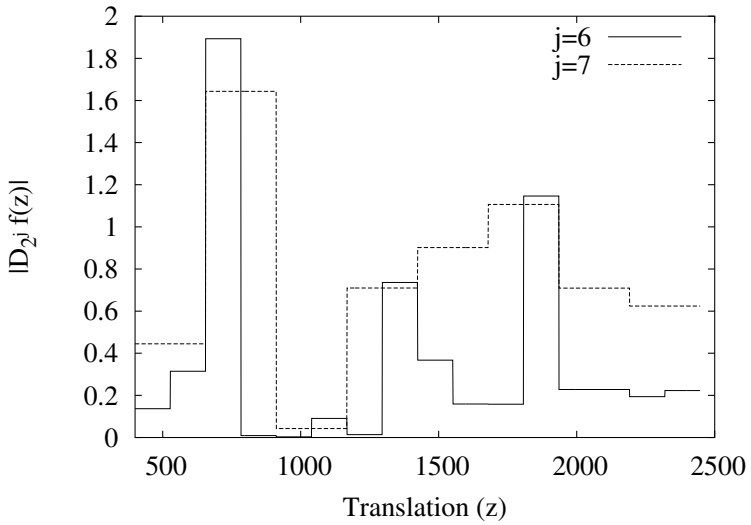


Fig. 3.3: *Absolute values of the discrete wavelet transform coefficients for the vegetation signature of Fig. 3.1, focused on scales 6 and 7.*

at each scale j by taking the root mean squared coefficients values:

$$E_j = \sqrt{\frac{1}{N_j}\sum_{x=0}^{N_j}[D_{2^j} f(x)]^2}, \tag{3.1}$$

with $N_j = N/2^j$. The energies reveal first order statistics of the detail signals, since they denote the variance of the coefficient's histogram. By using the wavelet energies as features, the total number of features is reduced to $\log_2(N)$.

Since the discrete wavelet transform is equivalent to applying a separable filter bank to the original signal [153] and the wavelet coefficients result from the high-pass filter, their mean is zero. Consequently, the energy equals the variance of the detail histogram. Employing energy as a feature, characterizes the detail histogram by a Gaussian.

Fig. 3.4 shows the energy features of the DWT for the same vegetation signature of Fig. 3.1. The energy content decreases exponentially for larger scales, indicated by the near linear behavior on the logarithmic scale. In agreement with the observation made above, scale $j = 6$ is of particular interest, with an energy content standing out of the regular curve. However, retaining only a single energy feature for each scale can be insufficient for complex classification problems. The energy features did not work for our case study in chapter 3.4. Another approach using more detailed information was needed.

Bruce [27] also suggested to reduce the dimensionality of hyperspectral reflectance spectra by a few, say n, features rendering the best one-dimensional

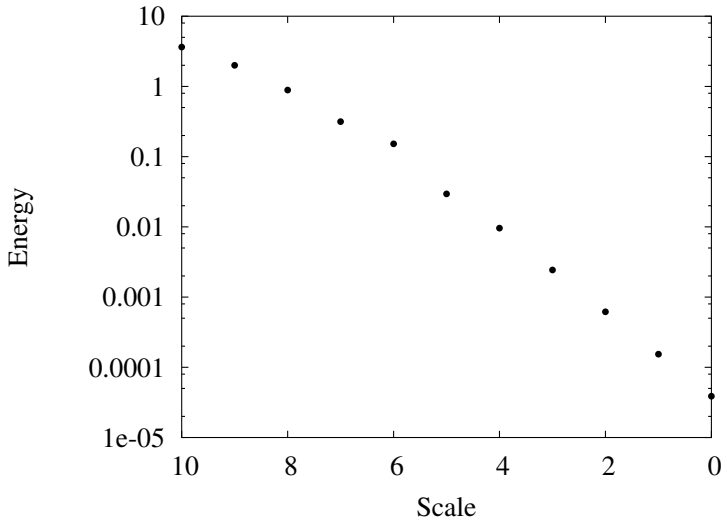


Fig. 3.4: *Energy features, derived from the discrete wavelet transform coefficients for the vegetation signature of Fig. 3.1.*

classification. But these n features are not necessarily the optimal combination, i.e., adding the second best single feature to the best might not add information for class separation. Therefore, we used the multivariate selection scheme, sequential floating search, introduced by Pudil [187].

3.3 Sequential floating selection

Sequential forward selection (SFS) first picks the variable with the best score for the criterion J. Next it adds a second variable for which the combination of both give the best score for the criterion, and so on. It is clear that it requires just $\frac{(N+M)(N-M+1)}{2}$ evaluations. Drawback of this simple selection scheme is that it can become stuck in a local minimum, because it can not correct for previously added features. The counterpart of SFS is sequential backward selection (SBS), which starts from the full set of features and removes the worst performing features sequentially. It has the same weakness as SFS, while the computational cost is much larger for the usual case where $M << N$.

An improved search strategy is proposed by Pudil, avoiding the local minimum using the *sequential floating forward selection* (SFFS [187]). After each forward step, one or more backward steps are taken, i.e., removing a previously selected variable to see if the separability measure can be increased at that level. The SFFS algorithm can be described as in in algorithm 3.1:

Algorithm 3.1 Sequential Floating Forward Search (SFFS)

Input: features to select from, e.g., Spectral Bands, Wavelet Coefficients,...
$Y = \{y_j | j = 1, \ldots, N\}$
Output: selected subsets
$X_k = \{x_j | j = 1, \ldots, k, x_j \in Y\}$, $k = 0, 1, \ldots, M$
Initialization:
$X_0 = \emptyset$; $k = 0$
repeat
  Inclusion {Add the most significant feature with respect to X_k}
  $x^+ = \arg\max_{x \in Y - X_k} J(X_k + x)$
  $X_{k+1} = X_k + x^+$; $k = k + 1$
  repeat
    Conditional exclusion {Try to remove features}
    $x^- = \arg\max_{x \in X_k} J(X_k - x)$ {the least significant feature in X_k}
    $X_{k-1} = X_k - x^-; k = k - 1$
  until $J(X_k - x^-) < J(X_{k-1})$
until X_k has the required dimension ($k = M$)

3.4 Case study on stress detection in apple orchards

3.4.1 Introduction

Hyperspectral remote sensing has gained a lot of interest, also for agricultural applications [227]. In particular, detection of crop stress in an early stage is important for precision agriculture practices.

It is possible to estimate the biochemical constituents of a leaf using leaf optical models, such as PROSPECT [108]. This is a physically-based model that describes the radiative transfer within the leaves by using absorption and scattering coefficients. It can be inverted by fitting it to the measured reflectance spectra, mostly acquired from individual leaves or small agricultural crops. This allows to estimate the chlorophyll, dry matter and water concentration together with a leaf structure parameter [110]. This is the subject of part II of this thesis.

However, indication of stress might as well be obtained directly from the obtained reflectance spectrum. Several stress indices have been proposed for stress detection. They are based on ratios of bands, or properties of the *red-edge*. Under stress conditions, leaf chlorophyll content typically decreases. As a consequence, the reflection of incident radiation from within the leaf interior increases, providing an optical indicator of stress. Both the position and the slope of the red-edge change [103].

These stress indices have their shortcomings and can not groundlessly be adopted for any sensor. Ratios are typically computed from leaf reflectances that are measured in the laboratory using a scanning spectroradiometer with high spectral resolution. It is not obvious to adopt the indices to airborne sensors with different specifications [33, 34, 216]. The precision at which the shift in the red-

edge position can be determined from airborne spectra is much lower than from ground-based spectra. A red-edge based index does not only depend on the spectral, but also on the spatial resolution of the sensor [49]. On top of that, the measured red-edge position can be ambiguous if the reflectance spectrum exhibits more than one maximum in its first derivative [103, 246].

The specific definition of a stress index highly depends on the specific nature of the problem at hand (particular type of vegetation, stress and/or sensor). Adopting an existing index is not guaranteed to work for the specific problem at hand, and deriving a new index requires specific prior knowledge of the problem that might not be available.

In this case study, we rather followed a more generic approach, by employing the reflectance spectrum without deriving features from it a priori, as is done with the radiative transfer models or vegetation stress indices. This approach has the advantage that no prior information of the specific vegetation stress problem or applied sensor is required. This can be done by applying all the available reflectance bands of a spectrum as features in a supervised classification procedure. Such a system is trained by using ground truth data, containing a representative sample of reference and stressed reflectance spectra. Based on these spectra, a classifier subdivides the feature space into a reference and a stressed class, so that an unknown spectrum can be classified a posteriori as reference or stressed.

It is well known that the performance of a classifier diminishes with higher number of features [61, 139], a problem that is very prominent for hyperspectral data. Bruce [27] suggested to reduce the dimensionality of hyperspectral reflectance spectra by a few, say n, features rendering the best one-dimensional classification. But these n features are not necessarily the optimal combination, i.e., adding the second best single feature to the best might not add information for class separation. Therefore, we followed the approach of Pudil [187], described in algorithm 3.1. Each time a feature is added, by looking at the best combined performance, hence optimizing the multivariate class separability. In this case study, we compared the proposed selection procedure to the combination of the best single features for the problem of vegetation stress detection.

Another issue that was studied, is the representation in which to present the reflectance spectra for feature extraction. The reflectance values themselves might not be the most appropriate representation. The desire to generate generic features, independent of the spectral resolution justifies the use of multiresolution representations such as the ones generated from wavelet transforms. Wavelet features have been used frequently for classification purposes [129].

Wavelets represent the reflectance spectrum at different scales. Features from particular scales correspond to specific spectral resolutions. The scales from which features are selected, reveal information about relevant spectral resolution levels. This allows, for example, to study the required spectral resolution for the application at hand. Also, this aspect was studied.

To validate the proposed method, the detection of stress in fruit orchards was considered. An experimental test plot was created at the Royal Research Station of Gorsem, Sint-Truiden (Belgium). Apple trees were put under Nitrogen stress. This

was simulated on a Jonagold stand which did not receive any nitrogen fertilizer since three years. A reference control plot was created where trees were kept at normal conditions, free from any stress or infections. From both the stressed trees plots and reference plots, shadow free leaves were sampled randomly. Foliar reflectance measurements were done with a portable field spectroradiometer on the fresh leaves. Stress indices calculated on the leave spectra revealed most of the stressed samples, confirming that the trees were indeed under stress. But the generic approach did a better job, detecting all the stressed leaves from the reference leaves. The experiments were repeated for other stress types and fruit variants. Examples are shown where the stress indices can no longer detect stress, whereas the generic method still can.

For the application of stress detection in fruit orchards, the wavelet-based features are shown to outperform the pure spectral reflectances. Also, the method of floating search selection is shown to outperform selecting the best single features. Finally, the acquired spectra are subsampled, hereby simulating a hyperspectral sensor, to be used for generating airborne and spaceborne data. It is shown that the available spectral resolution is sufficient for detecting stress.

The experimental setup is discussed in section 3.4.2. In section 3.4.3 we briefly present vegetation stress indices as a reference. We then use the full reflectance spectrum instead of a single index (section 3.4.4). The wavelet features are introduced in section 3.4.5, with a discussion on the selected features in section 3.4.6. We show some results of additional experiments in section 3.4.7.

3.4.2 Experimental setup

An experimental test plot was created at the Royal Research Station of Gorsem, Sint-Truiden (Belgium). Fruit trees were put under Nitrogen stress. This was simulated on a Jonagold stand that did not receive any Nitrogen fertilizer since three years. A reference control plot was created where trees were kept at normal conditions, free from any stress or infections. From both the stressed trees plots and healthy reference plots, shadow free leaves were sampled randomly. The leaf spectra were collected in July 2003 using a ASD FieldSpec Pro FR spectroradiometer. Measurements were taken on fresh leaves in VNIR and SWIR from 350 to 2500 nm with a spectral resolution of 3 nm in the VNIR and 10 nm in the SWIR [98]. The spectra were then resampled to 1 nm with the accompanying software. We concentrated on the bands between 420 nm and 2467 nm. The bands between 350 nm and 420 nm proved to be very noisy and were removed. In Fig. 3.5, the mean leaf reflectance curves are plotted for reference and stressed trees. Notice that differences are very subtle. Where differences are maximum, the signal variance is also high, making the detection task even more difficult. Moreover, there is no red-edge shift or increase of reflection in the visible spectrum (see also section 3.4.3). For each tree, 48 spectra were sampled around the crown. Five stressed trees and another five reference trees were selected. With 10 bad measurements, we obtained 470 samples. The difference between the two classes is most distinct in the near infrared to short wave infrared. However the

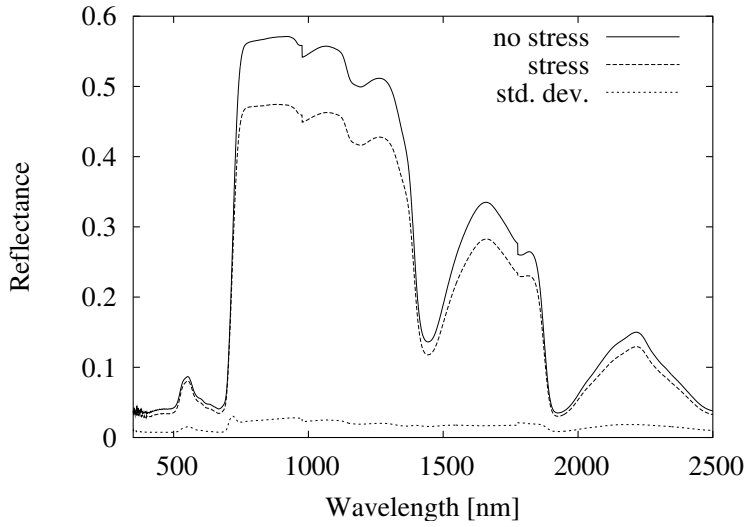


Fig. 3.5: *Hyperspectral reflectance curves of reference (solid line) and stressed (dashed line) vegetation on leaf acquired with ASD FieldSpec Pro FR*

reflectance standard deviation is also maximal in this area as shown in Fig. 3.5.

The experiments were conducted using *cross validation*. The available set of samples (470) was split into a training set and a hold-out set used for testing [99]. Here, we used the leave-one-out method [137]. By holding out one data point, a discriminant function was computed with the remaining set and the held out point was classified by use of this function. Each spectrum was used only once for testing. Hence, the instances were independent and identically distributed. We estimated the error of classification with confidence limits according to equation (2.27). The 95% confidence limits are indicated as an error bar in Fig. 3.6–3.9 and in Table 3.2.

3.4.3 Stress detection using vegetation stress indices

Parameters of the PROSPECT model, found by model inversion, were used for differentiating between stress and reference. The model parameters were obtained by minimizing the squared difference between the model and measured reflectance spectrum. We applied the publicly available implementation of the PROSPECT model. The four parameters obtained are: the structure parameter, the equivalent water thickness, concentration of chlorophyll a+b and dry matter. For this experiment, the structure parameter was best to differentiate between stress and reference, with 8% error rate. In the model this parameter is responsible for the reflectance of the near-infrared.

Chlorophyll concentration was observed to be not significantly different for stress and reference leaves. The expected decrease in chlorophyll concentration

due to the nitrogen deficiency did not take place. As a result, the shift of the red-edge was absent and most of the related stress indices were bound to fail. From all 18 stress indices tested, R_s [33] lead to the best stress detection result. This index corresponds to the reflectance at the red-edge shoulder, derived from an inverted Gaussian, which more or less corresponds to the structure parameter of the PROSPECT model. Using R_s, 9% of the samples were misclassified.

3.4.4 Stress classification using spectral bands

In the following experiments, the performance of the proposed generic procedure was elaborated. In a first experiment, we investigated the use of the complete reflectance spectra by applying all reflectance values as features. Two different approaches for the selection procedure were compared: the univariate combination of the best features (n best) and the multivariate SFFS procedure. In Fig. 3.6, the classification errors together with the confidence intervals are shown in function of the numbers of selected features from 1 to 9. To compare, the classification error using the single best vegetation stress index R_s is also shown. From the figure,

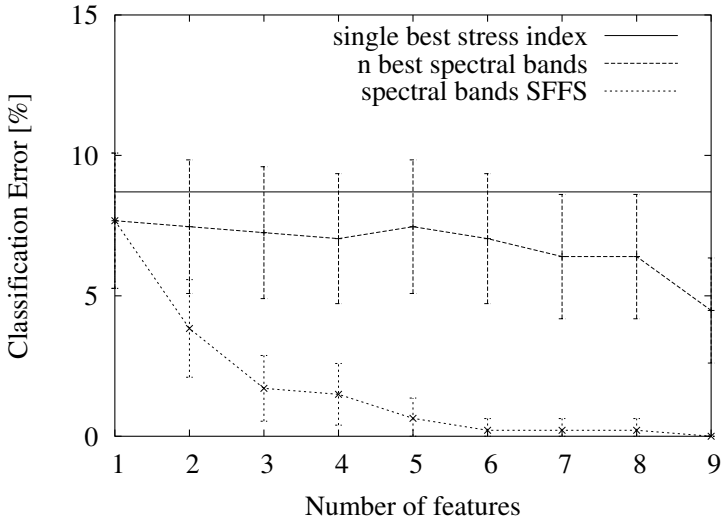


Fig. 3.6: *Classification error for reflectance features, selected with n best (dashed) and SFFS (dotted), compared to the best stress index R_s (solid)*

it is clear that the use of the complete spectrum improved results, compared to the use of a single stress index. The more reflectance features selected, the better the classification result. When the spectral features were selected independently, using the n best technique, results did not improve much with higher number

of selected features. This is due to the high correlation that exists between the selected features. When selecting using the SFFS, results improved dramatically by adding new features, so that with 6 features, the classification result is about 100%.

3.4.5 Stress classification using wavelet features

In the next experiment, the reflectance spectra were wavelet transformed and the performance of the wavelet features was investigated. Again the two feature selection procedures: n best and SFFS were compared. The results are shown in Fig. 3.7. The result for the single best vegetation stress index is also shown. Results are similar to the results for the spectral features: applying n best selection

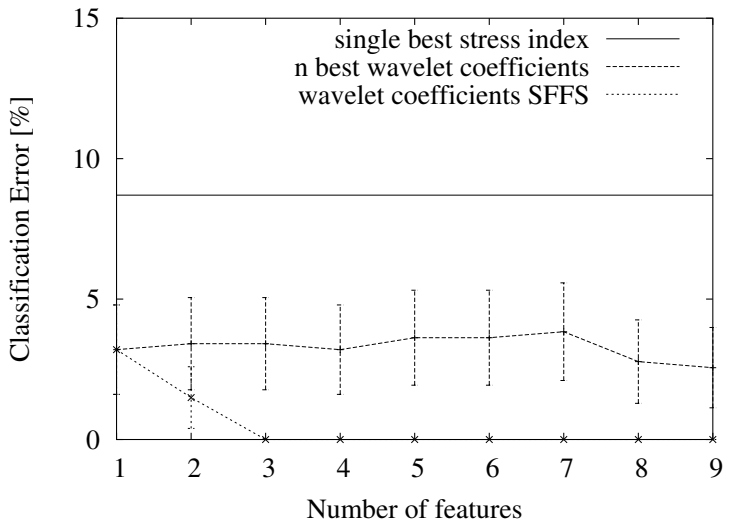


Fig. 3.7: *Classification error using wavelet features, selected with n best (dashed) and SFFS (dotted), compared to the best stress index R_s (solid)*

did not improve results with higher number of selected features, while applying SFFS did. This time, the use of 3 selected features was sufficient to obtain 100% accuracy.

When comparing Fig. 3.6 and Fig. 3.7, it is clear that the wavelet coefficients outperformed the reflectance features. The wavelet features reflect differences between reflectance bands at different scales. These relative differences appear to have better discriminative power with respect to the stress than the absolute reflectance values. Differences are also shown in some of the vegetation stress indices, and are optimally employed in a generic multiresolution framework.

3.4.6 Discussion on selected features

Apart from the improved classification performance, the use of the proposed generic technique has extra advantages. First, the feature selection procedure reveals information about the specific wavelengths that have discriminative power. On top of that, the use of the wavelet features reveals information about the specific spectral resolution scales with discriminative power. Such information can be useful for biophysical interpretation of the results. Moreover, this information can be used to determine specific sensor requirements, e.g., for an airborne or spaceborne upscaling of the ground measurements to sensors with lower spectral resolution and limited spatial resolution.

Selected reflectance bands

Both the univariate and multivariate schemes selected reflectance bands at the near infrared plateau around 1200 nm. With the univariate scheme, all bands were selected contiguously within a narrow spectral bandwidth between 1203 nm–1211 nm. Within such a small bandwidth, the features are much correlated. SFFS spread out the selected bands within the same plateau over a spectral bandwidth that was 10 times broader.

When applying the wavelet features, the specific scales at which features are selected determine the corresponding bandwidth of the feature. Bandwidths become broader as higher scales are selected. Using the univariate selection, all coefficients corresponded to wavelengths within a bandwidth between 1548 nm and 1571 nm. Using the multivariate selection, coefficients were selected from the entire spectrum from the visible to the short wave infrared. Within such a broad range, less correlated features can be more easily selected.

Selected scales

When we look at the scales from which the wavelet coefficients were selected, there was no difference between the univariate and multivariate selection. There was a preference for the lower scales corresponding to a high spectral resolution. From the first scale, 5 coefficients were selected, while 2 were selected from each of the two next scales.

Experiments with reduced spectral resolution

For most airborne sensors the spectral range is limited. For the CASI-2 (Compact Airborne Spectrographic Imager) this is from 420 nm to 931 nm. To simulate airborne spectra, the original ASD spectra were truncated before and after these wavelengths and then Gaussian filtered. We approximated the true transfer function of the sensor by centering the Gaussian functions at the exact locations of the CASI-2 sensor and identical full width half max at these locations. In Fig. 3.8, this is done for the the average reference and stressed spectra.

To acquire information on the required spectral resolution, we filtered the ASD spectra with different Gaussian functions, simulating spectral resolutions from 1 nm to 256 nm. The classification errors for these respective spectra are shown in Fig. 3.9. From this experiment, we can see that a spectral resolution of 32 nm is needed to detect stress in leaves with an accuracy above 95%.

Similarly, we determined the discriminating power of the resolution scales, by successively removing the coefficients of the higher scales. In this way, a sensor with diminishing spectral resolution was simulated. When all scales were applied, we disposed of the full resolution of 1 nm. A spectral resolution of 256 nm was simulated when retaining only the lowest 3 scales. The dashed curve in Fig. 3.9 corresponds to this experiment.

To determine the effect of truncating the ASD spectra, the solid curve in Fig. 3.9 represents the error of the complete spectral range of the ASD sensor at the different resolutions. There is useful discriminating power in the missing part of the spectrum, although not significant at resolutions better than 32 nm.

The results of the Gaussian filtered spectrum and the (truncated) spectrum at reduced scales are very similar. The discrete wavelet transform provides us a tool to quickly determine the required resolution, without having to Gaussian filter the original spectrum.

Some airborne sensors such as the CASI-2 can be configured with different settings. There is a trade-off between the number of bands and the spatial resolution, due to a limited integration time of the pixels. From the knowledge of the selected wavelet scales, the spatial resolution can be optimized with respect to minimally required spectral resolution for the problem at hand. The spectral resolution of 6 nm in Fig. 3.8 corresponds to a spatial resolution of about 1 m to 2 m.

3.4.7 Other experiments

To further validate the proposed technique for its generic properties, more experiments were conducted on other data sets, covering different time periods, stress types and fruit variants. The apple variants were Jonagold and Golden Delicious. The stress types we concentrated on are Nitrogen deficiency (only on Jonagold) and fungus (on Jonagold and Golden Delicious).

Measurements were done in June, August and September 2002 and June 2003, using a Li-Cor 1800-12S integrating sphere. This instrument measures both reflectance and transmittance, but is very time consuming in use. Only 30 samples for each class were measured. The results from the data, acquired with the integrating sphere should merely serve as an indication, since the small sample sizes may lead to inaccurate predictions. This is reflected in large confidence intervals. In July and August 2003, a contact probe was used. It is much faster in use. More samples per class (460–475) were measured. This contact probe data was only available for the Jonagold variety. Also, the Li-Cor measurements of August 2002 were missing for Jonagold.

In table 3.2, the classification errors and the confidence intervals are shown for the different data sets. Each time, five results are shown: using the best

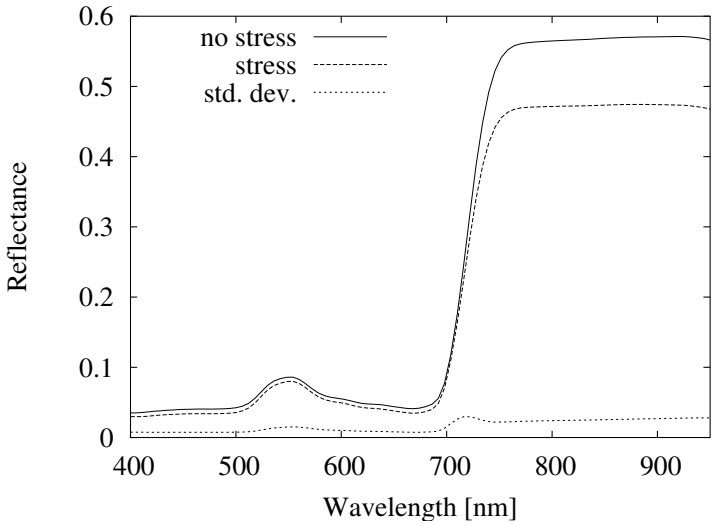


Fig. 3.8: *Filtered and resampled hyperspectral reflectance curves of reference (solid) and stressed (dashed) vegetation on leaf conform the CASI-2 sensor*

vegetation stress index, using spectral features with n best and SFFS selection and using wavelet features with n best and SFFS selection. The best result per data set is indicated in bold. The number of features was increased, until no further improvement was observed. The optimal number is case specific, with a minimum of 2 and a maximum of 9 features.

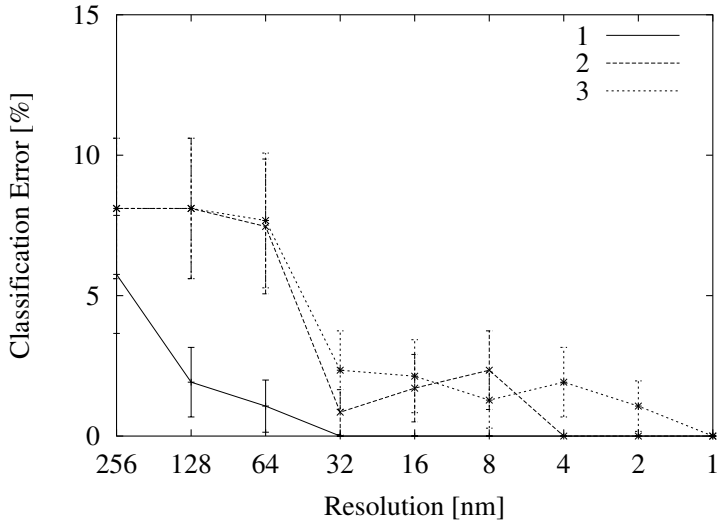


Fig. 3.9: *Classification errors using wavelet coefficients, simulating an airborne sensor; 1: by removing scales on the complete spectra; 2: by removing scales on the truncated spectra; 3: by filtering and subsampling the truncated spectra*

Tab. 3.2: *Comparison of classification error (in %) on different fruit variants and stress types. Best results are indicated in bold. The number of samples is indicated in column 2 (#)*

	Fungus Golden					
Period	#	indices	spectral bands		wavelet coefficients	
			nbest	SFFS	nbest	SFFS
06/2002	58	36±12	26±11	2±4	4±5	**0±0**
08/2002	58	10±8	10±8	2±4	**0±0**	**0±0**
09/2002	59	29±12	10±8	9±8	16±9	**3±4**
06/2003	60	17±9	10±8	14±8	**0±0**	2±4
	Fungus Jonagold					
Period	#	indices	spectral bands		wavelet coefficients	
			nbest	SFFS	nbest	SFFS
06/2002	60	33±12	24±11	**5±6**	**5±6**	8±7
09/2002	60	35±12	25±11	**0±0**	**0±0**	**0±0**
06/2003	60	35±12	17±10	**0±0**	**0±0**	**0±0**
07/2003	461	39±4	33±4	8±3	6±2	**2±2**
08/2003	460	43±4	40±4	25±4	24±4	**20±4**
	Nitrogen deficiency Jonagold					
Period	#	indices	spectral bands		wavelet coefficients	
			nbest	SFFS	nbest	SFFS
06/2002	60	8±7	**2±4**	5±6	3±4	3±4
09/2002	59	**10±8**	12±8	**10±8**	17±10	16±10
06/2003	60	27±11	31±12	**0±0**	**0±0**	**0±0**
07/2003	470	9±3	4±2	**0±0**	3±1	**0±0**
08/2003	475	32±4	25±4	20±3	27±4	**12±3**

Chapter 4

A hyperspectral band selection technique

If you are out to describe the
truth, leave elegance to the tailor.

Albert Einstein

4.1 Introduction

We will present a general approach for feature extraction, by treating the hyperspectrum as a continuous function and select complete waveband settings. In this way a sensor with new characteristics is imitated. The objective of this new band selection procedure is to perform a dimensionality reduction for classification purposes. We propose to use "local" continuous functions, as weight-functions for combining bands. These functions have 2 degrees of freedom, central wavelength and width (FWHM). The band selection problem then results into a $2M$ dimensional continuous optimization problem, with M the number of requested features. This allows the use of continuous optimization strategies rather than the discrete search methods used with band selection approaches. The band settings are optimized, using a criterion that maximizes class separation of the specific classification problem. The prime objective is to use the technique as a dimensionality reduction step preceding classification. But, the technique can serve other purposes. The selected wavebands provide useful information to the user in the interpretation

of the classification problem. Also, the technique can be a valuable tool for the selection of a sensor suitable for the problem at hand.

In section 4.2, a comparison with other selection procedures is made in a brief literature review. We then elaborate the construction of the wavebands (section 4.3), introduce the optimization criterion (section 4.4) and explain the optimization process (section 4.5). To demonstrate and evaluate the proposed technique, a specific classification experiment of dune vegetation is performed in section 4.6.

4.2 Comparison with other selection procedures

In classification, feature selection methods select a subset of features that are relevant to the target concept. Here, a search over all combination of bands is necessary. Since this is an NP-complete[1] problem, one has to resort to sub-optimal techniques, such as floating search. By definition, feature selection techniques see the neighboring spectral bands as independent, which in hyperspectral data is not necessary the case. Therefore, techniques merging local bands were proposed, i.e., band selection. These techniques reduce the number of features by combining neighboring bands. As this is also NP-complete, suboptimal search strategies are again used, such as top-down splitting of the spectrum.

Different efforts have already been made for suitable band selection. The problem is often defined as a selection of a new basis to describe the spectra by optimizing some criterion like error of reconstruction or discrimination between classes. A straight forward way to generate a new basis is to perform a *uniform sampling*. A reduced feature set $d' < d$ is obtained by smoothing the d dimensional input spectrum, followed by subsampling. The smoothing is typically performed using a Gaussian filter with d' means, uniformly distributed over the original bandwidth and a fixed FWHM that allows some overlap.

Price [185] proposed an iterative method to create a local basis with minimal reconstruction error. Karlholm [119] applied the basis matching technique to the band selection problem for target detection, whereas Wiersma [240] used PCA. Kumar [132] proposed a top-down band selection method, merging neighboring bands. Forward feature selection was used. The bands are merged using the mean reflectance values as features. This process is repeated for the remaining bands recursively, until the increase in discrimination is no longer significant. A variation is also proposed using a bottom-up approach, merging pairs of classes. It builds an agglomerative tree by merging highly correlated adjacent bands and projecting them onto their Fisher direction. Similarly, Riedmann [193] proposed an iterative method for merging neighboring bands.

[1] The worst-case time to solve such a problem can never be expressed as a polynomial.

4.3 Continuous band settings

The effect of a set of M filters $R(\lambda; c_m, w_m)$, $m = 1, \ldots, M$, with central wavelength c_m and width w_m, on a spectrum $x(\lambda)$ is described as a linear transformation:

$$y[m] = \int x(\lambda) R(\lambda; c_m, w_m) d\lambda \qquad \text{for } m = 1, \ldots, M \tag{4.1}$$

which corresponds to a projection of the signal on an M-dimensional vector space, describing the measuring process of the sensor. The measurement of spectrum $x(\lambda)$ is represented as an array of digital numbers, contained in a vector $\boldsymbol{y}$.

The problem of band selection corresponds to the selection of M filters from a redundant dictionary of functions. The possible filter functions mimic response functions of a spectral sensor. Here, we allow the response functions to be continuous functions defined by a central wavelength and a width. We continue with using the normal shape for this function, but any continuous local function describing a sensor response can be used. Each response function $R(\lambda; c_m, w_m)$ is then uniquely defined by the value of its central wavelength c_m and width w_m:

$$R(\lambda; c_m, w_m) = \exp\left(-\frac{1}{2}\frac{(c_m - \lambda)^2}{w_m^2}\right), \tag{4.2}$$

The task is now to define an optimization criterion to optimize the setting of the filters, i.e., the values c_m and w_m. In this manner, the band selection problem becomes an optimization problem of dimension $2M$.

4.4 Classification criterion

When classification is the goal, one could select $R(\lambda; c_m, w_m)$ such that, for the obtained values of $\boldsymbol{y}$, the classification performance is optimal. Another way would be to maximize the separation between the spectral classes. We will implement the filter approach. Within a supervised classification framework, we start from example spectra of the different classes with known class labels. From these, the wavebands giving maximal class separation are then obtained by maximizing a criterion that represents the class separation.

For classification, a lower bound on the expected classification error is indicated by the Bayes error. For a Gaussian assumption of the class distributions, this error can not be calculated exactly for a dimension higher than one. The Bhattacharyya bound, however, is an upper bound on the the Bayes error [77], and is calculated in the two class case as $\varepsilon_{i,j} = \sqrt{P_i P_j}\exp(-\mu_{i,j})$, with $\mu_{i,j}$ the Bhattacharyya distance from (2.3), and P_i, P_j the prior class probabilities. This distance renders a measure of the separability between two classes i and j. Here, we are considering the multi-class case, for which no analytical expression of the Bhattacharyya bound is available. Therefore, we consider the sum of all binary combinations of classes:

$$J_b = \frac{2}{C(C-1)} \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} \varepsilon_{i,j} \tag{4.3}$$

with C the number of classes.

4.5 Optimization

The optimization process consists of optimizing the parameters c_m and w_m in (4.2) to find the best projection $\boldsymbol{y}$, using (4.1). The criterion is the class separation criterion of (4.3). In order to allow for sub-band optimization and to be able to apply gradient-based optimization techniques, these values remain continuous. However, in practice, the actual continuous spectrum $x(\lambda)$, as required in (4.1), is not known. Therefore, we will use hyperspectral measurements instead. These can be obtained by spectrometer measurements or with airborne/spaceborne hyperspectral instruments. From these devices, a finely sampled spectrum is obtained, producing an N-dimensional array $\boldsymbol{x}$, with $N \gg M$. Assuming that the wavelength of the nth hyperspectral band is denoted by λ_n, the sampled version of the response function is $R(\lambda_n; c_m, w_m)$. We can combine the discrete version of the M different response functions into a matrix of size $N \times M$, with the elements $\mathbf{R}_{nm} = R(\lambda_n; c_m, w_m)$. Equation (4.1) for discrete hyperspectral data then simplifies to $\boldsymbol{y} = \mathbf{R}^T\boldsymbol{x}$. Using this linear relationship, we can relate the mean and covariance of the hyperspectral training set with those of the projected set. If the mean class value for the hyperspectral data set is $\boldsymbol{m}_k^x$ and the covariance $\boldsymbol{\Sigma}_k^x$, the corresponding values for the projected parameters are

$$\boldsymbol{m}_k^y = \mathbf{R}^T\boldsymbol{m}_k^x \text{ and } \boldsymbol{\Sigma}_k^y = \mathbf{R}^T\boldsymbol{\Sigma}_k^x\mathbf{R}. \tag{4.4}$$

These values are subsequently used in (4.3) to evaluate the criterion function.

Although no analytical expression for the minimum of (4.3) is available, the derivatives of the criterion function can be easily obtained. This allows the use of gradient based minimization methods. Unfortunately, we observed that the criterion function contains many local minima, making this approach fairly useless. Therefore, we resorted to global optimization procedures. A whole range of global optimizers have been suggested, like pattern search [144], genetic algorithms [90], simulated annealing [127], differential evolution [210], etc. Comparison of these different approaches showed that they all render similar results. Therefore, we restricted ourselves to the adaptive simulated annealing approach introduced in chapter 2, using a fast annealing schedule [105, 106].

4.6 Case study on optimal bands for dune vegetation mapping

4.6.1 Introduction

To demonstrate the technique of optimal bands, we applied it to the problem of classification of dune vegetation. In the experiments, the proposed method was compared to two feature selection and two band selection techniques. For feature selection, we chose for the simple M best feature selection technique and SFFS. The simple M best approach does not look at complementary information in the

different bands whereas SFFS does. For band selection, we first compared to the straight forward uniform resampling of the spectrum. A reduced spectrum is obtained by smoothing and subsampling the spectrum. Finally, the top-down band selection method of Kumar [132] was applied. This method splits the spectrum iteratively, and takes the mean reflectance as features. The criterion (4.3) was used to find the optimal split.

Section 4.6.2 introduces the test site and the airborne hyperspectral data. The classification problem is described together with the feature/band selection experiments. Results and discussion are presented in section 4.6.3.

4.6.2 Experimental setup

A test area at the west coast of Belgium has been selected for which hyperspectral image data was obtained in October 2002, using the Compact Airborne Spectrographic Imager (CASI-2 [1]) sensor. The data was acquired using 48 spectral bands between 400 nm and 950 nm. It was corrected for atmospheric absorption and geometric distortion. Around the same time, reference samples were taken for 13 different plant species during field work. The positions were identified using a differential GPS, and linked to 2098 spectra in the hyperspectral image. In this way a labeled training set was obtained. For both the Bhattacharyya distance and the classification, the prior probability of the classes was taken equal.

In a first experiment, we demonstrated the problem of the local minima using gradient based optimization on the dune data set. For this, $M = 5$ bands were uniformly distributed over the spectral range, their widths were set to $\sigma = 35$ nm, as shown in Fig. 4.3(a). This figure shows the five response functions, a typical spectrum for the data set and the spectrum resulting form the projection. Fig. 4.1 plots the criterion value of (4.3) in function of the central wavelength of the first band, keeping all other parameters fixed. One can clearly see the local minima in the criterion surface, a problem which gets more complicated when optimizing the parameters of all bands simultaneously. The peaks in the energy function are the consequence of overlapping bands, which result in singularities in the criterion value. This is because then the covariance matrix is no longer of full rank, which makes the inverse, required in (4.3), undefined.

In the next experiment, we performed the classification task using the proposed band selection technique. The goal was to select a limited number of bands that still contained the information to distinguish between the 13 different plant species. The proposed method was compared to other feature and band selection techniques.

4.6.3 Results and discussion

Fig. 4.2 shows the actual classification results on the labeled set in function of the number of selected bands. The error is estimated using leave-one-out bootstrapping with a multi-class classifier. This classifier is constructed by combining one-against-one combinations of binary classifiers. For these, LDA was applied.

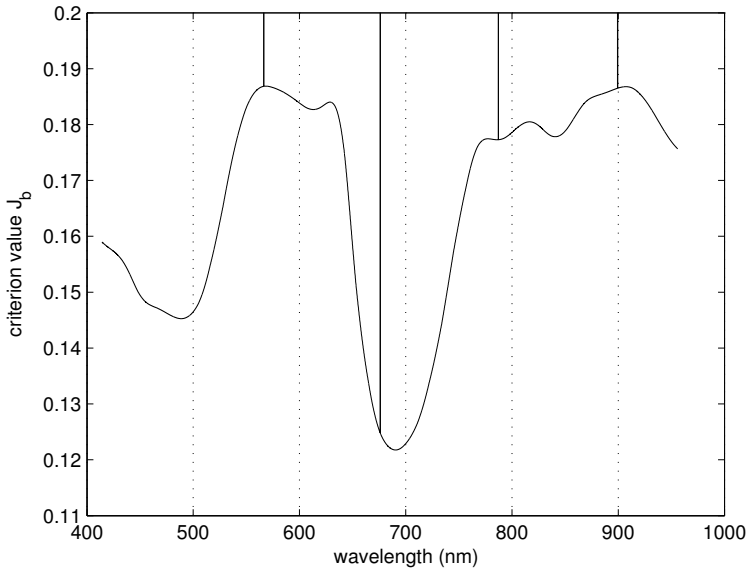


Fig. 4.1: *Criterion value in function of central wavelength of one band.*

The poor performance of the M best feature selection is clearly shown in the figure. This can be explained by the strong correlations between the M best separating bands. In Fig. 4.3(b) the selected bands are shown. The bands were all selected from within the same region. Applying SFFS improved the classification performance dramatically. Fig. 4.3(c) shows the selected bands for $M = 5$, which are clearly more spread out over the spectrum. When applying band selection, a simple uniform resampling of the spectrum as in Fig. 4.3(a), already provided satisfying results. Classification performance is only a few percent worse than with SFFS. This clearly indicates that important information is contained in the entire spectral range. The top-down approach is a recursive scheme that stops when no further improvements can be made. Here, the algorithm halted at dimension 10. The performance for low dimensions is good, but it deteriorates compared to the SFFS for higher dimensions. In Fig. 4.3(d) the selected areas are shown. When applying the proposed band selection technique, extra information on the class separability was used. We observed an important decrease in the expected classification error. Results improved over the use of SFFS, which indicates that both, the spreading out over the entire spectrum and the class separability, are important criteria for optimal classification. In Fig. 4.3(e), the selected bands are shown for $M = 5$.

Besides the advantage of the improved classification performance, the proposed technique is also useful for interpretation purposes. As indicated, Figs. 4.3(a)–4.3(e), show the selected response functions or bands for the different techniques

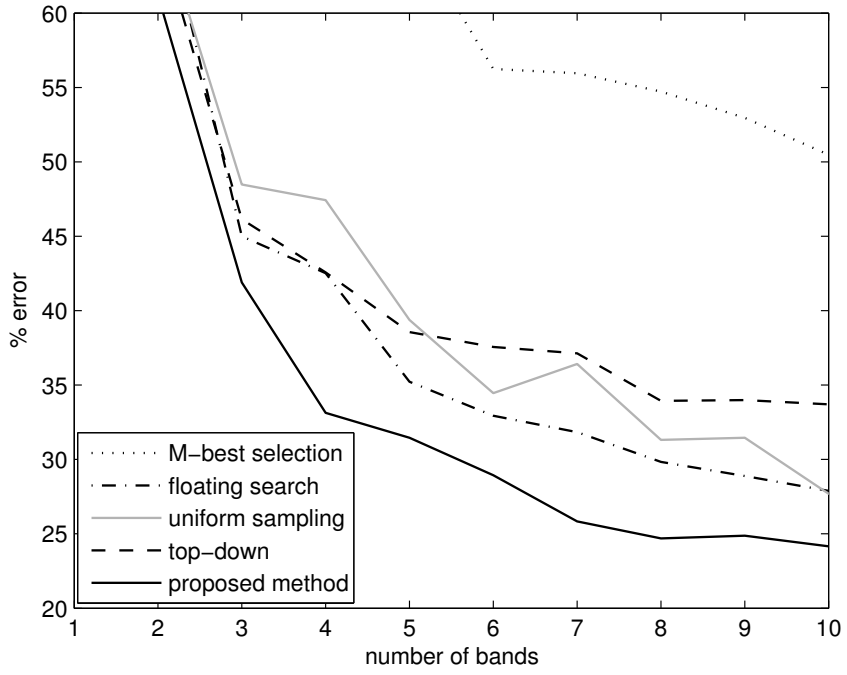
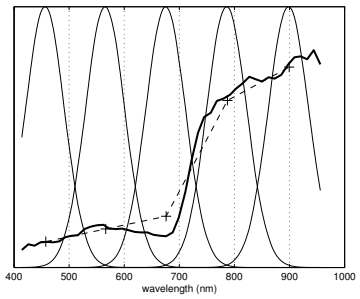


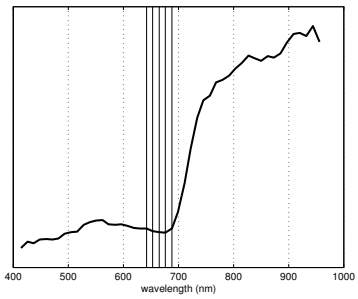
Fig. 4.2: *Classification error in function of number of selected bands for the various selection schemes.*

with M set to 5. The figures also show a typical vegetation spectrum originating from the data set and, for the band selection procedures, the resulting spectrum. The obtained selection can be interpreted within the framework of vegetation classification. Two major contributions in a vegetation reflectance spectrum are the leaf optical properties and the canopy structure. The visible region of the vegetation reflectance spectrum is characterized by low reflectance due to strong absorptions by pigments, like chlorophyll. The transition from visual (from 400 nm–650 nm) to NIR is characterized by a steep increase in reflectance, referred to as the red-edge. This edge and the subsequent plateau (from 750 nm–900 nm) are related to the leaf chlorophyll content and to the canopy structure [227]. One can expect the discriminating information between vegetation types to be characterized by the height of the visual part, the location of the red-edge (at about 700 nm), and the height of the NIR plateau. The M best selection procedure selected all bands right before the transition. Using SFFS, the bands were selected at green (550 nm) and before, at, and behind the red-edge. The top-down method selected for the visual part only one band, while sampling the red-edge and the NIR more finely. Using the proposed method, wide bands were selected in the visual range (400 nm–650 nm), while the red-edge (around 700 nm) is more finely sampled. On top of that, one band was selected at the NIR plateau.

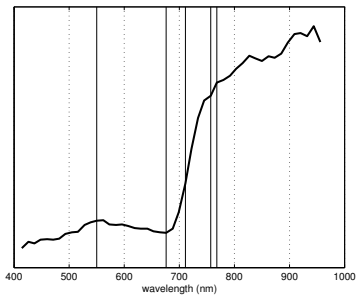
Interpretation possibilities such as these can be useful for a better understanding of the classification problems, and they can even be useful as an indication for sensor band settings. Nowadays, sensors are being developed with limited degrees of freedom with respect to the filter settings. A priori knowledge of the required band settings may help for the configuration of sensors.



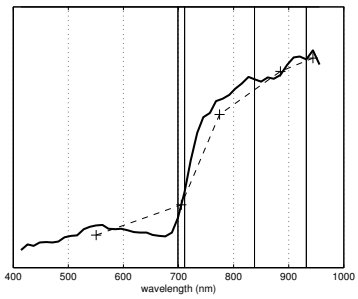
(a) Uniform sampling



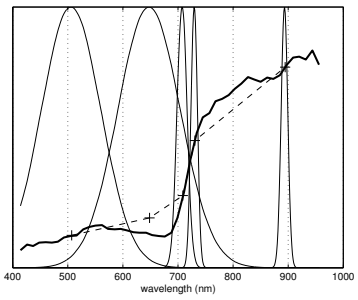
(b) M best selection



(c) floating search



(d) top-down method



(e) proposed method

Fig. 4.3: *5 bands selection with different techniques. A typical spectrum is also shown, before (bold line) and after (dashed line) band selection*

Chapter 5

A classification framework for hyperspectral imagery

> The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny ...
>
> Isaac Asimov

5.1 Introduction

In the previous sections we proposed some feature selection techniques for classifying hyperspectral data. We will now present a generic classification framework that will cover the classification of hyperspectral data for multi-class problems. This classification framework combines binary classifiers in a one-against-one approach, by coupling the posterior class probabilities. These probabilities are then used to smooth and unmix the obtained classification result.

In section 5.2, we first show how the multi-class problem can be solved by an architecture of binary classifiers. Contextual smoothing and unmixing are discussed in sections 5.3 and 5.4. The techniques have been implemented and applied to a new case study on dune vegetation mapping in section 5.5.

5.2 Solving multi-class problems

Due to the complexity of multi-class classifiers, a common approach is to combine the output of several binary ones. For the binary classifier, we can adopt a simple linear discriminant classifier. A projection weight vector $\boldsymbol{\beta}$ and bias β_0 are the parameters to estimate in the two class problem. They are calculated by:

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \qquad \beta_0 = -\frac{\boldsymbol{\beta}^T}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \tag{5.1}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the means of each class, and $\boldsymbol{\Sigma}$ is the estimated class covariance matrix. Test samples ($\boldsymbol{x}$) are then classified by the simple rule:

$$\boldsymbol{\beta}^T\boldsymbol{x} + \beta_0 \quad \begin{cases} \leq 0 & : \text{sample assigned to class 1} \\ > 0 & : \text{sample assigned to class 2.} \end{cases} \tag{5.2}$$

This method is very fast to train and to calculate the classification. In case the training set is not sufficiently large, $\boldsymbol{\Sigma}$ can become singular. We then can calculate the pseudo-inverse to obtain $\boldsymbol{\beta}$ and β_0 [189].

One-against-all or one-against-one [76] approaches are commonly used. With the one-against-all strategy, each classifier is trained to differentiate one class from all the others, which requires a number of classifiers equal to the number of classes C. In the one-against-one approach, all possible pairs of classes are compared, requiring $\frac{C(C-1)}{2}$ classifiers. Different methods defining other codings of the classes were also suggested [57, 7].

Based on the conclusions of Furnkranz [78], we will concentrate on the one-against-one scheme and propose some techniques how to combine the binary classifiers. Maximum voting [76] is a simple yet powerful technique. For each binary classification, a vote is given to the winning class. The class with the maximum number of votes is assigned to the sample to be classified.

A problem with most techniques that combine classifiers, is that no class probability is retained. Hastie and Tibshirani [97] proposed a classification scheme by pairwise coupling that did. We present the basic idea and will extend the original algorithm to improve the classification accuracy.

If we assume a Gaussian class distribution, we obtain the class probability given the observed variable $\boldsymbol{x}$:

$$\begin{aligned} p(\omega_i|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|\omega_i)p(\omega_i)}{\sum_{j=1,2} p(\boldsymbol{x}|\omega_j)p(\omega_j)} \qquad i = 1, 2 \\ &\text{where} \\ p(\boldsymbol{x}|\omega_i) &= \frac{1}{\sqrt{2\pi\boldsymbol{\beta}^T\Sigma_i\boldsymbol{\beta}}} \exp\left(-\frac{\left(\boldsymbol{\beta}^T(\mu_i - \boldsymbol{x})\right)^2}{2(\boldsymbol{\beta}^T\Sigma_i\boldsymbol{\beta})}\right). \end{aligned} \tag{5.3}$$

being the probability of the projected point $\boldsymbol{\beta}^T\boldsymbol{x}$ in (5.1).

For the C-class case we have to look for C p_i's ($i = 1, \ldots, C$) which satisfy

$$r_{ij} = \frac{p_i}{p_i + p_j} \text{ and } \sum_{i=1}^{C} p_i = 1, \tag{5.4}$$

This set of equations, to be solved for p_i, has $C - 1$ free parameters and $\frac{C(C-1)}{2}$ constraints, so it is generally impossible to find $\hat{p}_i$'s that will meet all equations. However, we can find the best approximation $\hat{r}_{ij} = \frac{\hat{p}_i}{\hat{p}_i + \hat{p}_j}$ by minimizing the Kullback-Leibler distance between r_{ij} and $\hat{r}_{ij}$ [97]:

$$l(p) = -\sum_{i \neq j} n_{ij} \left[r_{ij} \log \frac{r_{ij}}{\hat{r}_{ij}} + (1 - r_{ij}) \log \frac{1 - r_{ij}}{1 - \hat{r}_{ij}} \right] \tag{5.5}$$

where n_{ij} is the sum of the number of training points in class i and j. They also suggest an iterative scheme to minimize this distance (algorithm 5.1).

Algorithm 5.1 Coupling probabilities

start with initial guess for $\hat{p}_i$, and calculate $\hat{r}_{ij}$
repeat
  for all $i = 1, \ldots, C$ **do**
    $\hat{p}_i \leftarrow \hat{p}_i \frac{\sum_{j \neq i} n_{ij} r_{ij}}{\sum_{j \neq i} n_{ij} \hat{r}_{ij}}$
    normalize $\hat{p}_i$, and calculate $\hat{r}_{ij}$
  end for
until convergence

Hastie and Tibshirani [97] proved that the distance between r_{ij} and $\hat{r}_{ij}$ decreases at each step, and since the distance is bound above zero, the procedure converges. This procedure is repeated for all points in the test set. Now, classification is obtained by selecting the class with maximum posterior probability.

Coupling probabilities is not guaranteed to produce better classification accuracies than maximum voting. In fact, our experience is that maximum voting outperforms coupling probabilities when the number of classes increases. Although we do not have a theoretical basis for this, one explanation could be that the small contributions of the probabilities of the "losing" classes can become significant if more classes are competing. For this reason, we modified the binary class probabilities r_{ij} to:

$$r_{ij} = \begin{cases} 0 & \text{if } r_{ij} < 0.5 \\ 1 & \text{if } r_{ij} \geq 0.5 \end{cases} \tag{5.6}$$

Using the modified probabilities from (5.6), the benefits of both techniques can be combined, i.e., posterior class probabilities can be obtained, while producing an identical classification accuracy as maximum voting.

5.3 Contextual smoothing

Pixel based classification procedures consider the pixels as spatially independent. No contextual information is applied to make a decision on the class label. Building a classification image based only on the spectral information often results in poor classification performance [163] and in class images with a noisy appearance, containing many single pixel classes. This "salt and pepper" noise is related to a combination of factors. First, *misclassified* pixels due to the intra-class spectral variability occur at random locations in the image. Second, the geometric outline of classes produced by a pixel based classifier does not follow natural spatial entities such as fields, forests and streams or man made objects such as roads and buildings. Third, noise is not restricted to classification errors alone. Isolated, but correctly classified pixels, that are much smaller than the size of the object under study, also contribute to the noisy appearance of a classification map. This problem is getting more significant with the fine resolution of the current sensors.

What is missing in the pixel based classification schemes using only spectral information, is spatial contextual information. This is characterized by the spatial distribution of pixels of the objects being studied. Texture features are one option to include spatial information. In this case, the spatial information is included prior to classification. Alternatively, spatial information can be included in the classification scheme itself (contextual classifiers). For example, Markov random fields [125] can be used to model how the probability of presence of one class is affected by its neighbors. Markov random fields have been successfully applied for classification purposes [221, 158], but depend on how well the contextual information can be modeled [171]. In practice, information on neighborhood configurations is hard to obtain from (sparse) ground truth and the modeling is often limited to trivial situations (see also [222]). Another type of contextual classifier is the object based classifier that uses a segmentation step to group neighboring pixels into meaningful areas or *objects*. Segmentation is a well known technique in machine vision, which partitions an image into homogeneous regions on the basis of some homogeneity measure [170, 96]. The classification is then performed on the segmented image instead of the pixels. Object based classification has gained a lot of interest, not in the least by the success of some commercial software products such as eCognition® (Definiens Imaging) and Feature Analyst® (Visual Learning Systems, Inc.). However, the segmentation step is not trivial, though crucial, because errors are propagated to the classification step. Moreover, the impact of a classification error of a single object is severe, as many pixels can be involved.

Another way to get rid of noisy classification results is to re-allocate noisy pixels by including spatial contextual information as a postprocessing step, i.e., after the classification. Although spectral analysis can produce noisy classification results, the majority of pixels in an area of interest of the image is expected to be classified correctly. A wrong classified pixel that is inconsistent to its correctly classified neighborhood can then be reassigned to its most related class in a spatial context.

Several filtering techniques have been proposed for post classification proce-

dures since 1980 [217, 219]. One of the simplest techniques is the *majority filter*. It labels the center pixel within a structuring element as the majority class of all neighboring pixels in the element. The minority class will thus be removed as noise. The majority filter is easy to implement, but alters the boundaries between classes for larger structuring elements. The median filter can be a better alternative for this reason. Some more refined methods have been proposed by Barnsley [14], Kim [122] and Groom [79].

Jensen [113] "disassembled" the classification output into individual class layers ω_i such that a binary image is obtained for each class. All pixels that do not belong to the particular class ω_i obtain a value 0 (unclassified). A morphological operator then cleans the individual layers from "exterior" and "interior" noise. Exterior noise consists of small groups of pixels that exist outside the major regions of a class and is expressed as speckle. Interior noise consists of small groups of unclassified pixels within major regions of a class and appears as holes in homogeneously classified regions. A user defined threshold decides on the size of a group to be considered as noise. Qian [188] further improved this technique by using graphs operations instead of pixel operations on the classified image.

In most of the post classification techniques, the probability or confidence with which a pixel is classified is discarded. Not so in relaxation techniques, where probabilities of neighboring pixels are used iteratively to update the probability of a given pixel [190].

Iterative procedures are time consuming. A simple alternative is to include contextual information by using an ordinary low pass or median filter on the posterior probabilities before labeling the classes. However, the statistical correctness is lost. The filtered version is a mixture of posterior class probabilities over several pixels and is no longer a posterior class probability itself.

A new post classification technique is presented here, requiring little extra computational effort while retaining the meaning of a posterior class probability for each pixel. We call $p_i(k,l)$, the posterior probability for class i, for the pixel at location (k,l) in the image. Normally, to assign a label to the pixel, the label of the class with maximum posterior probability is taken. Define $c(k,l)$ as the class with the maximum posterior probability at location (k,l):

$$c(k,l) = \max_{\arg i} p_i(k,l). \tag{5.7}$$

One can assume neighboring pixels to have similar posterior probabilities. This information can be used as prior knowledge for defining a new prior probability for a pixel, based on the posterior probability from classification in the neighborhood of the pixel. Define this new prior probability of a pixel as the average over the posterior probabilities of neighborhood Ω:

$$p_i^{\text{prior}}(k,l) = \frac{1}{N} \sum_{(a,b)\in\Omega} p_i(a,b) \tag{5.8}$$

where N is the number of points in Ω. When looking at $p_i(k,l)$ as an image, the new prior $p_i^{\text{prior}}(k,l)$ is in fact a smoothed version of this image. A new posterior

probability is obtained by using Bayes' rule:

$$p_i^{\text{post}}(k,l) = \frac{p_i^{\text{prior}}(k,l)p_i(k,l)}{\sum_j p_j^{\text{prior}}(k,l)p_j(k,l)} \tag{5.9}$$

Classifying using these p_i^{post} will result in smoother classification image maps containing less single pixel classes.

5.4 Unmixing

The pixels in a remotely sensed image often represent an area on the earth's surface that contains multiple classes [67, 46], a problem that is referred to as mixed pixels (mixels). Mixing is mainly related to either discrete (crisp) boundaries or gradual (fuzzy) transition between thematic classes. But also the presence of small elements within a pixel or the contributions of a target outside the pixel area but influencing it by the point spread function (PSF) and the adjacency effect result in spectral mixture [72]. The proportion of these mixed pixels in an image thus depends both on the properties of the sensor (point spread function, spatial resolution) and the landcover mosaic on the ground (class composition, spatial arrangement). In relation to vegetation mapping, the problem of mixed pixels at crisp boundaries is a matter of spatial resolution of representing this boundary. The mixels will appear as a linear feature in the classified image and the error made will depend on which feature these pixels are attributed to in the post processing of spatial smoothing. In the case of fuzzy transitions between vegetation classes (so called ecotones), attributing the mixels to one of the involved classes would result in substantial loss of information.

To estimate sub-pixel class compositions, several approaches can be used [62]. Linear spectral unmixing uses a set of end members, defined as pure spectra of a class [36]. The basic assumption underlying the linear mixture model is that there is no multiple scattering between the different cover types, so that each photon that reaches the sensor has interacted with just one cover type. Under these conditions, the energy received at the sensor can be considered as the simple linear sum of the energy received from each cover component. Unfortunately, it was found that the mixing models tend to be reliable under conditions where only one (or a few) vegetation components exists, which is often not the case. The unmixing model assumes that all image pixels are formed by simple mixing of the end members. This can be written as product of matrices $\mathbf{X} = \mathbf{AE}$, where $\mathbf{X}$ is a matrix containing the spectra of the image pixels, $\mathbf{A}$ the matrix of abundances, and $\mathbf{E}$ the matrix containing the end members. Generally, this linear model is constrained by 1) positivity of the abundances, and 2) the abundances per pixel should add up to one. The solution of the unmixing model is obtained by least squares solution of the constrained mixing model, with a given a set of end members.

Many studies have found that the posterior class probabilities can be used as an indicator of sub-pixel proportions [69, 16]. The posterior class probabilities obtained in this classification framework can be directly used as class abundances

for the unmixing. Both models result in C gray scale images from 0 to 100%, representing the proportion coverage of each class 1 to C. In case of a large number of classes, a lot of disk space is required and the result is hard to interpret.

Rarely all class combinations are meaningful, so a limited number of classes can be chosen to unmix. An example is elaborated in the next case study on dune vegetation mapping, where only 4 out of 16 classes are selected for unmixing.

5.5 Case study on surveillance of dune vegetation

5.5.1 Introduction

The surveillance of vegetation is an important issue with respect to Integrated Coastal Zone Management (ICZM). Vegetation maps are in the first place required for planning and evaluation of nature conservation. Vegetation units are the basic constituents of habitat types used in regional, national or international conservation policy [100]. A key obligation upon member states under the European Habitats Directive (Council directive 92/43/EEC) for instance, is reporting on the state of the components, i.e., habitats, of the Natura 2000 network every 6 years. But also from a health and safety point of view, coastal ecosystems in general and their vegetation in particular, deserve special attention. Mainly in areas where large parts of the coastal plain are reclaimed and valuable urban and agricultural land is situated below sea level, dunes are carefully managed as a natural coastal defense structure. From this point of view, detailed mapping of topography and dune fixation is an essential tool for estimating the strength of the foredunes. So both nature conservation and coastal defense acknowledge the benefit of accurate and up to date vegetation maps, although they have a different approach to vegetation management. The challenge in this case study was to develop a mapping instrument using airborne hyperspectral imagery, which integrates both requirements and therefore contributes to a more efficient ICZM [60, 228].

The integration of an ecological and a functional approach requires an appropriate definition of vegetation types. Subsequently, these types should be confronted with the limitations related to classification of remotely sensed images. Response design was one of the challenges in this test case. The mapped area is one of the most ecologically diverse. Reference data consisted of multi-species plots, including 303 plant species (vascular plants, mosses and lichens). Moreover, the separate species could often only be distinct on sub pixel level. We collected the abundances for each plot in a database. With this information, we labeled the plots with TWINSPAN (Two Way INdicator SPecies ANalysis), a commonly used clustering algorithm for vegetation data [100]. Labels were only defined after vegetation analysis of the collected field data, which we refer to as a bottom-up approach. This avoided a biased typology where labels are defined a priori. The sampling strategy consisted of recording the entire species composition as well as the structure of the vegetation.

Two main ways of analyzing such a data set are clustering and ordination [120]. Vegetation clustering can be used for determination and subsequent mapping of

discrete habitat types while ordination is more likely used for modeling the general relationship between species composition and remotely sensed data. The latter results in a continuous representation on abstract axes, which are related to the major components of variation in species composition [202, 142].

Several authors have used ordination techniques to link reflectance to floristic composition of vegetation. Schmidtlein [203] for example, used Detrended Correspondence Analysis (DCA) to determine the major floristic gradients in a grassland complex and found a high correlation between the first two DCA axes and the reflectance values of the considered plots ($R^2 = 0.71$ and 0.66 respectively). Trodd [220] found even better correlations for a more limited vegetation gradient in heathlands. Armitage [8] showed clear species responses to the direct ordination axis representing the joint floristic/spectral variation in sample plots of British semi-natural upland vegetation.

However, the resulting ordination axes remain rather abstract. Furthermore, the ordination results are uniquely associated with the combination of floristic and spectral data of the research considered. This can be highly disadvantageous in monitoring programs where repeatability is a major requirement. A main goal of this case study is the development of a monitoring tool, so we need to ensure that subsequent maps will provide the same information. This implies an unequivocal definition of vegetation types as endmembers in an image classification.

In order to obtain discrete vegetation types, the recorded vegetation samples can be clustered. TWINSPAN is commonly used with vegetation data [100]. This algorithm carries out a hierarchical, dichotomous division of the dataset, based on the presence and abundance of each species within the samples. An advantage of this divisive clustering approach is that it tends to retain the overall structure of the data.

Zak [242] successfully applied TWINSPAN on a large scale for the classification of Chaco vegetation in central Argentina using Landsat TM imagery. Their study resulted in an overall accuracy of 85% for 11 classes. Mehner [157] also used it for classifying vegetation in British uplands. In their classification of very high spatial resolution satellite imagery (IKONOS), using 26 a priori defined classes (including 18 vegetation classes), an overall accuracy of about 75% was obtained. However, from experiences in Canadian peatland, Thomas [218] conclude that TWINSPAN is not always a suitable method for clustering species abundance data to optimize image classification. These authors used very high spatial resolution airborne multispectral data and obtained an overall accuracy of 41% for 8 vegetation classes.

As already stated in the review article by Nagendra [168], results of image classification based on floristic composition are highly divergent in terms of accuracy because spatial patterns and spectral properties of vegetation can differ substantially between ecosystems. Furthermore, the differences in approach make it hard to compare results of various studies. One of the key elements seems to be the structural homogeneity of the reference data within a single training class, as spectral reflectance is highly influenced by vegetation structure [132]. Therefore, in this work, a step-by-step definition of vegetation classes is proposed. Firstly,

woody vegetation, dominated by a single *species*[1] can be assigned to a certain a priori defined class (top-down). Secondly, for more complex herbaceous vegetation characterized by the co-occurrence of several species, a preliminary vegetation survey and analysis is required in order to define the major vegetation types. This was achieved using TWINSPAN. Finally, only structurally homogeneous representatives of these types were retained to label the ground reference for the supervised classification. Hence, for these types, final class labels were not known a priori but were derived after analyzing the ground reference data (bottom-up).

Hyperspectral sensors have gained a lot of interest with their large number of contiguous narrow bands. The introduction of new algorithms and wide access to increased processing power have increased the potential of hyperspectral remote sensing data.

Schmidt [201] and de Lange [55] introduce an expert system, based on prior knowledge and ancillary data. Major input for these expert systems is a digital elevation model derived from LIDAR data. Schmidt [201] realized an increase in overall accuracy from 40 to 66 percent. The study area, located in the Netherlands, mainly consists of coastal salt marsh. In this type of ecosystem, topography, expressed as inundation frequency, is the key factor determining the distribution of plant communities. In coastal dunes however, this dominant factor is not present and vegetation development is generally determined by a larger number of environmental factors such as groundwater level, microclimate and sand mobility. This probably explains why de Lange [55] was not able to improve results with an expert system. In our work, we only used data from an airborne hyperspectral sensor.

In section 5.5.2, we first present the vegetation survey and analysis. We then discuss the available data in section 5.5.3. The results of the TWINSPAN analysis are presented in section 5.5.4, and section 5.5.5 show how ground reference data were labeled. The classification results are presented in section 5.5.6.

5.5.2 Vegetation survey and analysis

Dune vegetation is often characterized by several co-occurring species, resulting in a certain mixed *vegetation type*[2]. These vegetation types were sampled in plots with a diameter of 5 m. A total of 1113 multi-species plots were recorded, including 303 plant species (vascular plants, mosses and lichens). The plots were randomly distributed over all dune sites along the Belgian coast, respecting an equal distribution over the different sites. In each plot, vegetation structure was recorded by estimating vegetation height and attributing a percent membership-value to 5 structure types: sand, moss, herbaceous, scrub and litter. Furthermore, the entire plant species composition was noted and cover was estimated.

A vegetation analysis was then performed on these multi species plots. They mainly consisted of herbaceous vegetation in which separate species can often only

[1] The term species is adopted even if sub-species taxonomical levels are used.

[2] We will use the term vegetation type in relation to the ecological vegetation characterization. In contrast, the term *vegetation class* will be related to the image classification procedure.

be distinct on sub pixel level. Therefore, herbaceous vegetation types needed to be defined before they could be used as ground truth for the image classification. A number of scrub types was included in the analysis to ensure the representation of the entire dry dune succession series. However, most of the woody vegetation consisted of a single species dominating areas larger than one pixel and hence they were not treated in the analysis of mixed vegetation types.

The analysis was performed using the TWINSPAN version of Turboveg for Windows (©2005 Stefaan Hennekens, Alterra). We refer to Hill [101] for more details on the TWINSPAN algorithm. Nine abundance classes (cut levels) were used, which is more than is common in phytosociology, where species composition is the major concern. However, if vegetation types are to be used for ground truth, it is important to emphasis the abundance of species because this determines the structural resemblance of vegetation types.

5.5.3 Available data

Hyperspectral data on the Belgian coastline were acquired with the AISA Eagle sensor on July 6th 2004. The VNIR range of 400 nm–970 nm was covered by 32 spectral bands with a ground resolution of 1 m. The interesting spectral bands for vegetation were located around the green peak and red-edge (680 nm–740 nm), where the dynamics of the vegetation spectrum is the most important [238]. To maximize the information content of the acquired data, a narrow bandwidth was chosen in these regions (2.2 nm to 2.3 nm), at the cost of broader bandwidths in the blue and near infrared part of the spectrum (25 nm). This choice was motivated by the lower signal to noise ratio in blue, that can be compensated by the larger bandwidth. Internal leaf structure is the dominant factor controlling the spectral response of plants in the near-infrared [204]. The typical NIR plateau contains less detailed information, motivating a larger bandwidth in this region [238].

Nine study areas were covered by 16 flight lines (Fig. 5.1). All images were corrected for atmosphere using ATCOR-4 [192], based on the radiometric transfer model MODTRAN [19]. For the geometric correction, an integrated GPS/IMU was used, monitoring the aircraft position and attitude. However, the accuracy obtained was not sufficient. Errors of 2 to 5 pixels were found by:

- visually comparing with infrared imagery of superior spatial resolution over the same area
- checking overlay of overlapping tracks
- identification of ground control points

A ground campaign was set up around the time of the flight. The ground reference was carefully mapped onto the georeferenced hyperspectral images, resulting in more than 120,000 pixels. They represented all vegetation types defined and were distributed over the available tracks. The position of each plot center was determined with differential GPS. For the multi species plots, a minimum surface area of 25 m^2 was respected. Due to the geometric inaccuracy of the hyperspectral

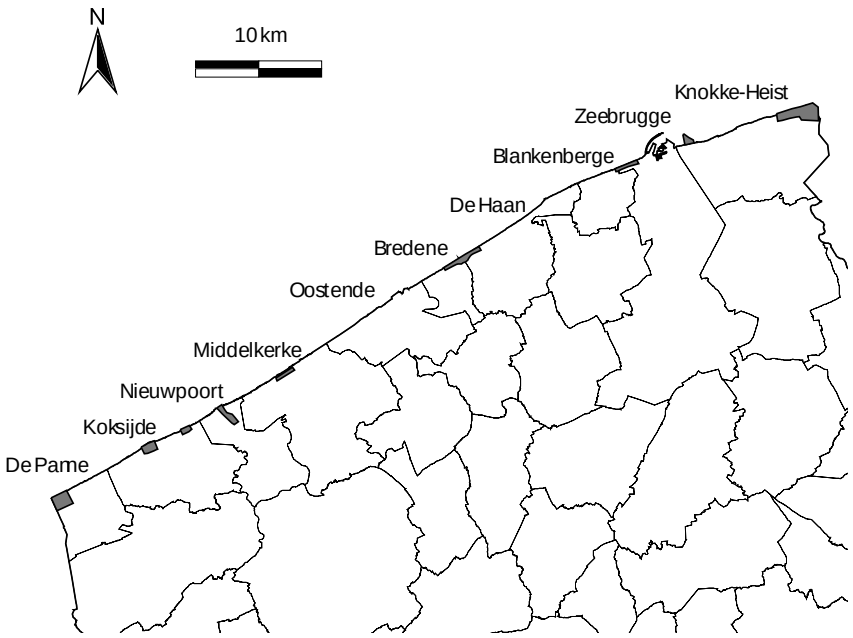


Fig. 5.1: *Nine study areas along the Belgian coastline were covered. The Belgian coastline stretches from De Panne near the French border to Knokke-Heist, close to the Netherlands.*

image, the location of the ground truth polygons had to be manually corrected, making use of 30 cm resolution orthophotos. To be on the safe side, only nine (3x3) central pixels of the multi-species plots were used for training and validation.

5.5.4 TWINSPAN results

The TWINSPAN clustering of the set of 25 m^2 vegetation plots resulted in eight vegetation types (Fig. 5.2).

Pioneer vegetation in embryonic dunes is characterized by its sparse vegetation cover (about 50%), dominated by *Elymus farctus*. An average plot only contains 3.6 species. *Cakile maritima* is present in half of the plots.

Vital *Ammophila* dune is dominated by marram (*Ammophila arenaria*). On average, this species covers 60% of the plot surface, with one third appearing as dead plant material. Another 20% is covered by sand whereas the remaining area is covered mainly by forbs such as *Cirsium arvense*, *Senecio jacobaea*, *Sonchus arvensis* and *Hypochaeris radicata*. *Festuca rubra* (mainly ssp. *arenaria*) is present in 75% of the plots. The average number of species per plot is 7.3.

Semi-fixed *Ammophila* dune structurally differs from the vital marram dunes in its higher moss cover (on average 15%) at the expense of litter and herbaceous plant cover. Average number of species per plot is 10.5, which is clearly higher than in vital marram dunes. The species composition is a combination of marram dunes and moss dunes. Only the high presence and cover of the moss *Brachythecium albicans* can be considered as a characteristic feature.

Moss dune vegetation consists of about 20% sand, 40% moss and 40% herbaceous cover. *Syntrichia ruraliformis* is the most abundant moss species, often accompanied by *Hypnum cupressiforme*, *Ceratodon purpureus* or *Brachythecium albicans*. Besides, the vegetation is characterized by the high presence of the annuals *Phleum arenarium*, *Erodium glutinosum*, *Cerastium semidecandrum* and others.

Moss rich dune grassland holds an intermediate position between moss dune and dune grassland, both in species composition and vegetation structure. The average number of species per plot is 16.8, which is not significantly different from the dune grasslands (16.5). This vegetation type is differentiated by the dominance of *Hypnum cupressiforme* in the moss layer and the high presence of *Cladonia rangiformis*.

Herbaceous dune grassland is rather species rich and variable in terms of species composition. Structurally it consists of about 75% of graminoids and forbs. Abundant in terms of presence and/or cover are *Carex arenaria*,

Galium verum, *Festuca rubra*, *Poa pratensis* and *Plantago lanceolata*. Characteristic species are, e.g., *Koeleria albescens*, *Ononis repens* and *Avenula pubescens*.

***Rubus caesius* grassland** is characterized by the dominance of dewberry (*Rubus caesius*). Within the TWINSPAN group, this species has an average cover of 60%. Due to the presence of this competitor, the average number of species drops to 11.3, which corresponds to the level of moss dunes and semi fixed marram dunes.

***Hippophae* scrub** is dominated by sea-buckthorn (*Hippophae rhamnoides*). The species composition of the buckthorn scrub in this study is mostly related to the *Ammophila* dunes because of the high presence of *Ammophila arenaria* and *Festuca rubra*. *Galium aparine* and *Anthriscus caucalis* are characteristic for this vegetation type. The average number of species per plot is rather low (9.2).

5.5.5 Labeling of ground reference data

The TWINSPAN output was used as a basis for the final class labels for image classification (Table 5.1). However, some modifications have been carried out, because not all of these vegetation types represented clear-cut classes, suitable for image classification. First, in terms of species composition, some types represented a transitional stage between more unambiguously definable plant associations. Second, the TWINSPAN groups were determined based on species composition so they could still be heterogeneous in structure (as illustrated in Fig. 5.2). For example, semi-fixed *Ammophila* dune represents a transitional stage between vital *Ammophila* and moss dune, both with respect to species composition and vegetation structure. As mixed classes can contain various proportions of the constituents, their definition is susceptible to ambiguity. Therefore we chose to map these classes using spectral unmixing. This implied that the heterogeneous TWINSPAN groups **pioneer vegetation**, **semi-fixed *Ammophila* dune** and **moss rich dune grassland** were not retained as labels for image classification. Furthermore, extra criteria were added in order to obtain structurally homogeneous classes for Marram (MA), Moss dune (MO) and Dune grassland (DG) as indicated.

The TWINSPAN groups ***Rubus caesius* grassland** and ***Hippophae* scrub** fit in with the second type of ground truth, which is determined by a single, dominant species. Labeling of these types was mainly based on the digitized polygons in stead of the 25 m^2 multi species vegetation plots. In addition to Dewberry (*Rubus Caesius*) and Sea-buckthorn (*Hippophae scrub*), we defined the most dominant scrub types Traveller's-Joy (*Clematis vitalba*), Creeping willow (*Salix repens*) and Wild privet (*Ligustrum vulgare*). Some less abundant scrub types were joined in a single class OS (Other Scrub). They covered a significantly smaller fraction of the scrub covered area. Furthermore, their spectral similarity made the classification task more arduous.

Tab. 5.1: *Definition of the classification labels. English plant nomenclature accords to [38].*

Label	Description
BS	Bare Sand
MA	Marram dune (training samples must contain at least 60% marram and less than 10% sand)
MO	Moss dune (training samples must contain at least 75% moss, less than 15% sand and less than 20% herbaceous cover)
DG	Dune Grassland (training samples must contain at least 80% grass or herbaceous cover, less than 10% sand and their height must be smaller than 30 cm).
TJ	Traveller's Joy: *Clematis vitalba* scrub
DB	Dewberry: *Rubus caesius* scrub
SB	Sea-buckthorn: *Hippophae rhamnoides* scrub
CR	Creeping Willow: *Salix repens* scrub
WP	Wild Privet: *Ligustrum vulgare* scrub
OS	Other Scrub: collection of less abundant scrub types (*Crataegus monogyna*, *Eleagnus angustifolia*, *Euonymus europaeus*, *Lycium barbarum*, *Rosa rubiginosa*, *Rosa rugosa*, *Salix cinerea* and *Sambucus nigra*)
WS	Mixed type of mostly Woodland and Scrub: *Acer pseudoplatanus*, *Fraxinus excelsior*, *Populus alba*, *Populus canescens*, *Salix alba* and *Prunus spinosa*.
BW	Broadleaf Woodland (*Alnus glutinosa*, *Betula*) spp. and *Quercus robur*
CW	Coniferous Woodland (*Picea abies*, *Pinus nigra* and *Pinus pinaster*)
SH	Shadow
WA	Water
UA	Urban Area

A mixed type WS (Woodland and Scrub) was defined from species with similar spectra, which are therefore difficult to classify. Moreover, this mixed class often showed an intermediate structure in the field (heights of 3 to 6 m), which was particularly striking in the case of clonal stands of several poplar species. Broadleaf Woodland (BW) and Coniferous Woodland (CW) completed the vegetation classes. Finally we ended up with 16 class labels, including the non-vegetation classes BS (Bare Sand), SH (Shadow), WA (Water) and UA (Urban Area).

5.5.6 Classification results

An extract of the classification result is shown in Fig. 5.3. The length of the strip is just over 1 km, covering only 2% of the entire classification result. Nevertheless, it shows a typical Belgian dune site, situated between the urban areas of Oostduinkerke and Nieuwpoort and fragmented by a major road in Zeebermduinen

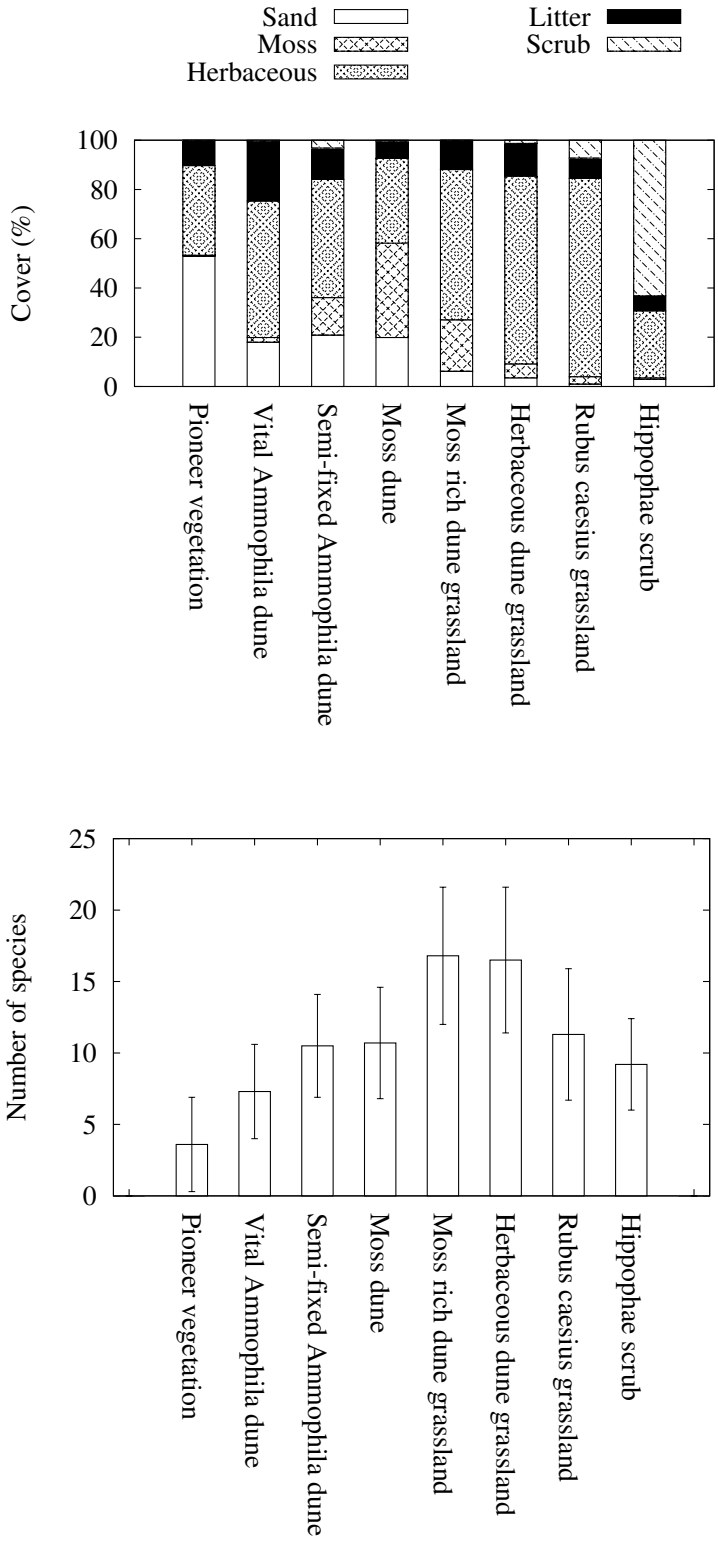


Fig. 5.2: *Average vegetation structure of the plots (top) and average number of species per plot with standard deviations (bottom) within the* 8 *retained TWINSPAN types.*

(north) and Ter Yde (south). The classification was performed by combining binary classifiers in a one-against-one approach. Because of a limited number of bands (32), combined with a large training set, all bands were used as features. There was no additional value by using any of the proposed feature selection schemes, proposed in chapters 3 and 4.

The classified image in Fig. 5.3 shows the final class labels after classification and smoothing. This extract illustrates the geographical detail of the classification and the kind of processes which can be monitored. In general, the dynamics of the area can be derived from the presence of large areas of bare sand, which are mainly related to aeolian sand drift. But also (over)recreation contributes to vegetation degradation, resulting in the presence of (mostly smaller) patches of unvegetated sand. Sand fixation is a more complicated process to which several vegetation types contribute. Dune fixing vegetation development is mainly represented by the series sand - (open) marram vegetation - moss dune and finally dune grassland. On the classified AISA-eagle hyperspectral data, we were largely able to distinguish the components of this series and further development into scrub or woodland.

The effect of smoothing and unmixing is shown in Fig. 5.4. The result of the pixel based classification is represented in Fig. 5.4 a. A prior probability was constructed as the average over the posterior class probabilities of some neighborhood. A new posterior probability was obtained by using Bayes' rule (5.9), resulting in the smoother classification image in Fig. 5.4 b.

Finally, the smoothed posterior probabilities were used for unmixing (Fig. 5.4 c). Only three vegetation classes (Dune grassland, Marram dune and Moss dune) and Bare sand were selected for this process, defined as subset M. In other words, only the proportions of the pixels that were labeled as one of the 4 classes in M were retained. However, due to the spectral similarities of the classes, the posterior probabilities of the classes in $\overline{M}$ (not in M) are not guaranteed to be 0, even without actual mixing of those classes. We assumed those probabilities as errors and forced them to 0, as if the respective "prior" probabilities of the classes in $\overline{M}$ have been set to 0. Finally, the proportions of the classes in M were normalized to 100%.

New labels can be assigned to the mixed classes. We chose to add only 2: sparse marram and dry pioneer dune vegetation (see Fig. 5.4 c). The first is a combination of bare sand and marram. This class is important to characterize the dune formation process, which is characterized by sparse marram vegetation on the beach. The latter is a mixture of bare sand with moss dune and grassland species. The distinction between sparse moss dune and sparse grassland was not retained because of its minor ecological significance. Pioneer stages of both vegetation types tend to be characterized by a similar species composition. The mixed classes were obtained using the presence of bare sand. Because no ground truth was available for these classes, they are not included for the quantitative validation in Table 5.2. For the same reason, they are not included in the legend of 16 classes in Fig. 5.3 either.

A validation was performed using 20 replications of a 2-fold cross validation, i.e., half of the reference pixels were used for training and half of them for valida-

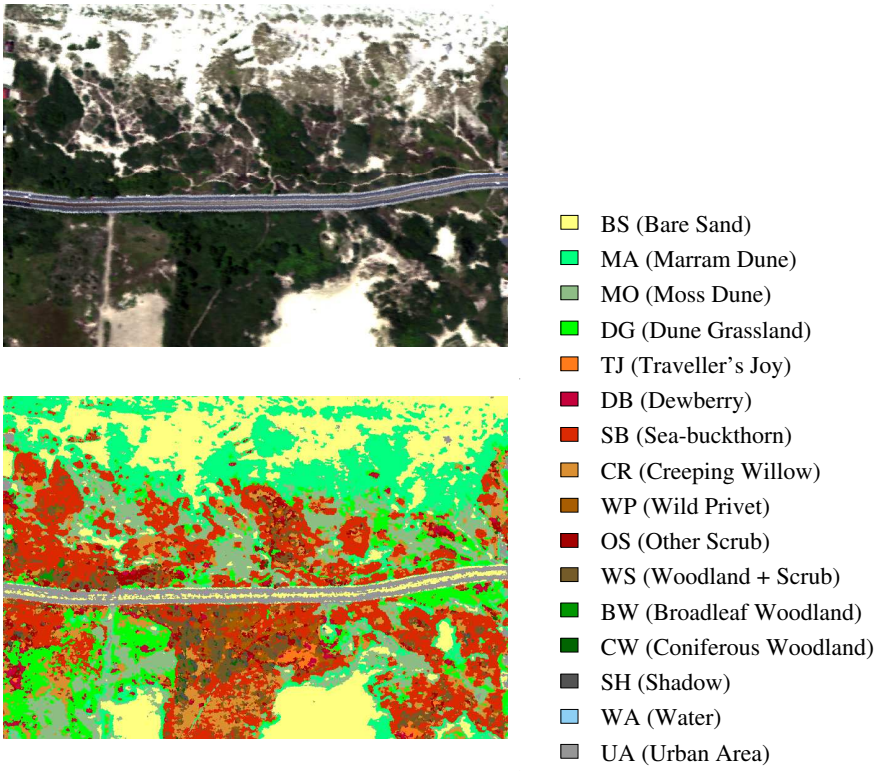


Fig. 5.3: *Above: true color image (Red = 634 nm, Green = 545 nm, Blue = 463 nm) of a typical Belgian dune site near Oostduinkerke, Zeeberm-duinen (North of the road) and Ter Yde (South of the road), generated from a hyperspectral image with 32 bands. Below: Classified vegetation map with 16 classes.*

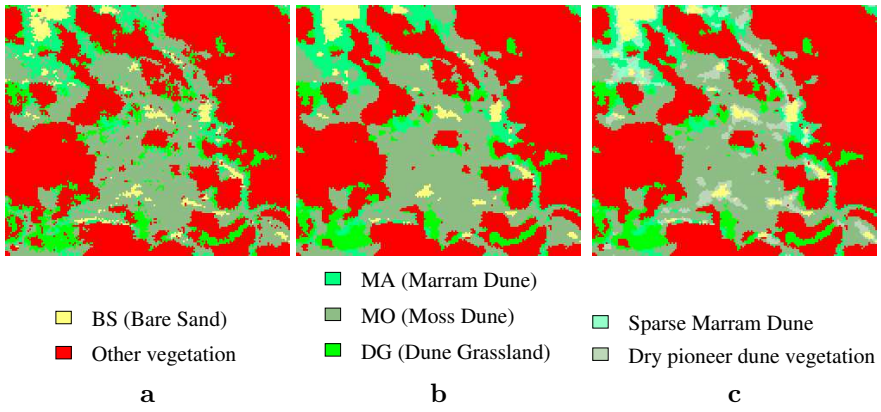


Fig. 5.4: *Detail of the classification result (a) and the effect of spatial smoothing (b). Two fuzzy classes were introduced after unmixing: sparse marram and dry pioneer dune vegetation (c). Vegetation classes other than grass, marram and moss dune are not taken into account for unmixing and are shown in red (other vegetation).*

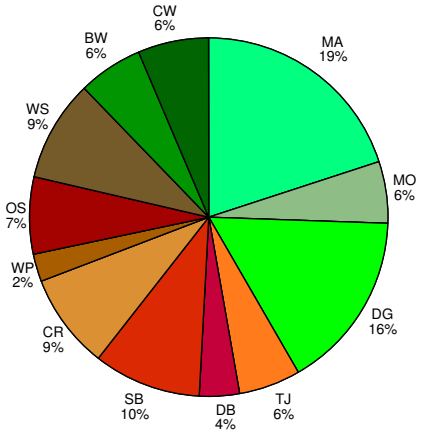


Fig. 5.5: *Pie chart showing the importance of the different dune vegetation classes throughout the Belgian coastline.*

tion. The split was randomly chosen and the validation result was averaged over all replications. The overall accuracy and Kappa coefficient for the entire sample of the dune vegetation along the Belgian coast is 81% and 0.78 respectively. To take into account the effect of imbalanced classes, the average accuracy (sum of producer's class accuracies divided by the number of classes) is a better measure than the overall accuracy, which is 75% in this case.

A confusion matrix [43] is shown in Table 5.2. The final column with the row total represents the number of test samples used for each class (50% of the sample size). The confusion matrix is based on the ground reference data only. In Fig. 5.5, the percentile coverage of each vegetation class is shown, based on the complete hyperspectral image data set. The pie chart thus indicates the importance of the different dune vegetation classes along the Belgian coastline. Bare sand, shadow, water and urban area are not included, because they were not the focus of this study. Furthermore, including them would not represent an actual cover, because the flight planning was optimized to merely cover the dune vegetation.

Marram dune

Marram covered 19% of the entire dune vegetation. It represents two vegetation associations. The association dominated by marram grass itself (*Ammophila arenaria*) is most abundant (TWINSPAN-group **Vital *Ammophila* dune**), while the association dominated by sand couch-grass (*Elymus farctus*) is quite rare. The latter is mostly found as an open vegetation and will therefore generally be mapped as the spectral unmixing-class 'sparse marram'. Due to its definition (see Table 5.1), the class Marram dune can consist of up to 40% of different species. This mixing is inherent to the ecology of the vegetation since closed vegetation with *Ammophila* only occurs in semi-fixed dunes and thus contains species characteristic for moss dunes and dune grasslands. Vegetation with only Marram grass on the other hand, unavoidably contains a significant portion of sand, which contaminates a pure end-member signature as well. Nevertheless, the accuracy for marram is high (79%) and the minor confusion with dune grassland is explicable in light of the class definition.

Moss dune

Moss dunes have a very distinct appearance with a relatively low representation (6%). It can quite clearly be separated from the other dune vegetation (91% accuracy). Moss dune vegetation is extremely vulnerable to trampling because it lacks robust sand fixing grasses such as marram. Therefore, its mapping is important with respect to the estimation of dune stability.

Grassland

Classifying grassland types can certainly be considered a major challenge for the integration of vegetation ecology and remote sensing [203]. However, within the dry dynamic dune system, the retained grassland class derived from the TWINSPAN

Tab. 5.2: *Confusion matrix of the classification results of the final vegetation classes:* $\kappa = 0.78$*, overall accuracy=81%.*

Label	BS	MA	MO	DG	TJ	DB	SB	CR	WP	OS	WS	BW	CW	SH	WA	UA	Total
BS	8032	5	4	0	0	0	0	0	0	0	0	0	0	0	0	62	8103
MA	5	251	18	30	0	0	0	1	0	0	0	0	0	0	0	14	319
MO	3	4	393	13	0	0	1	4	0	0	0	0	1	0	0	12	430
DG	0	38	15	512	3	41	2	38	10	25	1	0	0	0	0	0	684
TJ	0	0	0	1	127	11	0	0	26	8	1	2	0	0	0	0	176
DB	0	5	0	42	9	242	0	5	3	37	5	0	0	0	0	0	346
SB	1	30	21	34	5	12	5220	656	159	287	563	70	154	9	0	0	7222
CR	0	8	2	60	0	2	128	982	19	48	84	14	40	0	0	0	1387
WP	0	7	0	8	26	5	23	63	1033	141	118	32	40	1	0	0	1496
OS	0	40	1	133	91	116	282	280	688	2395	571	83	169	63	0	0	4912
WS	0	3	1	45	31	56	833	334	574	571	3876	958	342	115	0	0	7739
BW	0	0	0	2	0	13	17	25	18	14	76	982	47	31	0	0	1226
CW	0	0	7	25	4	6	32	147	8	58	66	14	11747	438	0	12	12564
SH	0	0	0	0	0	0	3	5	1	5	26	2	107	839	0	51	1038
WA	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2803	10	2817
UA	651	5	35	0	0	0	1	0	0	0	0	0	0	0	3	9746	10442
Users acc.	0.92	0.63	0.79	0.57	0.43	0.48	0.80	0.39	0.41	0.67	0.72	0.46	0.93	0.56	1.00	0.98	
Prod. acc.	0.99	0.79	0.91	0.75	0.72	0.70	0.72	0.71	0.69	0.49	0.50	0.80	0.93	0.81	0.99	0.93	

clustering forms an ecologically relevant unit and spectrally it is clearly separable from other herbaceous plant communities dominated by marram and dewberry. Although the producer's accuracy for the remaining grassland class is good (75%), the user's accuracy is low (57%), mainly because of confusion with scrub. The grassland class only covers 16% of the study area but is important for a large part of the diversity of vascular plant species. Dewberry (4% coverage) was classified with 70% accuracy. It is spectrally close to scrub but ecologically it is more related to dry dune grassland.

Scrub

Scrub represents 34% of the mapped dune vegetation. This figure should be increased with some percents derived from the mixed class scrub/woodland. Lumping several rare scrub types into the class 'Other Scrub' seems justified since it only covers about 7% of the studied area. The 4 major scrub species are classified with a producer's accuracy of about 70%. User's accuracies are less (about 40%), except for sea-buckthorn (80% accuracy), due to the discrepancy in numbers of test pixels. Nevertheless the maps give a good indication on the general type of scrub. Young scrub is mainly dominated by sea-buckthorn (10%) and creeping willow (9%). Both scrub types are accurately classified (72% and 71% respectively). Wild privet (69% accuracy) only occurs in older landscapes, representing only 2% of the dune vegetation coverage. Finally, Traveller's-Joy is a recent gain in the flora of the Belgian coastal dunes. Currently it covers 6% of the dune vegetation, but it is spreading rapidly and will probably cover a substantial part of our future dunes. With an accuracy of 72%, it is possible to follow up this process quite satisfactory.

Woodland

The spectral separability of broadleaf and coniferous woodlands has been demonstrated in many remote sensing applications [66]. However, in this study no further distinction of individual trees was made. Broadleaf and coniferous woodland cover each 6% of the dune vegetation and both can be mapped with good accuracy (80% and 93% respectively). The mixed class scrub/woodland was mapped with only 50% accuracy. It was introduced as a buffer for the pure scrub and woodland classes. In the image it is often found in the transition zone between scrub and woodland. In total, this mixed class covers 16% of the entire scrub and woodland area.

Other types

The non-vegetation classes soil, water and urban area were well classified, with accuracies exceeding 90%. The shadow class confused with coniferous woodland. Apart from the spectral overlap, there was an unavoidable presence of shadow pixels within the conifer training polygons.

Part II

RADIATIVE TRANSFER MODELING

Introduction

The second part of this thesis is on (continuous) parameter estimation using radiative transfer models. As an introduction, we first give an overview of leaf and canopy models. We then discuss some common techniques for parameter estimation, focusing on model inversion. Our main contribution is on the use of multi-angular hyperspectral data, discussed in chapter 7. In a final case study, we compare two canopy models for the retrieval of chlorophyll from an airborne hyperspectral image over an artificially treated peach orchard [260].

Chapter 6

Overview

> The only good model is Claudia Schiffer, the rest is just a mathematical approximation
>
> Guido Cosemans

6.1 Motivation for radiative transfer models

Radiative Transfer (RT) models simulate, amongst other, the reflectance when electromagnetic radiation interacts with objects in a medium, e.g., atmosphere, water and vegetation. Depending on the medium, different objects must be modeled. In case of vegetation, the objects can be leaves or canopies. Reflectance models provide an essential tool to interpret remote sensing imagery and help us to understand the radiometric behavior of objects. RT models are also used for various practical applications. Verhoef [233] applied *forward* modeling to prepare the user community of hyperspectral data for the kind of imagery data that could be expected from the candidate mission SPECTRA of the European Space Agency ESA. The SPECTRA mission was canceled in favor of the RADAR mission (EarthCARE) however. North [174] did the same by generating landscape scale BRDFs for the pre-launch preparation of the ATSR-2 instrument. We used forward modeling to simulate remote sensing data by upscaling of leaf reflectance field measurements, because airborne hyperspectral reference field data were sparse [252].

Reflectance models are most popular when *inverted*. Inversion has stimulated much of the development in optical canopy reflectance modeling over the last two

decades [212, 231, 167, 9, 134, 88]. The main focus in this thesis is on the modeling of leaf and canopy reflectance, although we used the techniques introduced in the following sections to water body applications as well [257]. More in particular, we are interested in quantifying the physiological state of canopies with observed canopy reflectance values. We therefore coupled a canopy reflectance model with a leaf optical model. By inverting the coupled models, leaf (biochemical) and canopy (biophysical) parameters can be estimated. Quantifying the physiological state of crops via remote sensing is relevant for precision agricultural practices. It allows timely and efficient agricultural management, which can improve yield and quality [45, 215].

It is important to notice that vegetation has an important *temporal* signature, driven by its phenological stage (see also discussion in section 6.2). The estimation of biochemical parameters is largely influenced by this temporal effect [91]. Changes in phenology can be included as additional features for characterizing the physiological state or detection of stress if time series of reflectance data are available. However, in this thesis. we mostly concentrate on the *spectral* signature of vegetation. In chapter 7, we also deal with its *angular* signature (bi-directional effects) for improving the estimation accuracy.

6.2 Leaf level

Reflectance of vegetation is a complex modeling exercise. Leaf reflectance is influenced by the concentration of leaf biochemicals, water content, and leaf structure [82, 6, 48, 73]. All these constituents are variable in time and space. Temporal change is induced by climate, catastrophic events (floods, fire, drought, disease) and anthropogenic activities. The expression of temporal change is elicited by phenology (annual cyclic process) and the diurnal cycle of the opening and closing of stomata in the leaf. The stomata regulate the exchange of moisture, CO_2 and O_2. Spatial differences in leaf characteristics result from species differences (needle leaf and broadleaf), but also with the same species, spatial stratification of leaves has a strong impact on canopy reflectance. The cell walls within the leaf cause multiple scattering in many directions, depending on the angle between incident light and the orientation of the cell walls. On the other hand, the waxy layer covering the leaf epidermis (*cuticle*) results in a strong specular component. In a review on a state of the art on leaf optical properties, Jacquemoud [109] concluded that "our understanding of leaf bidirectional properties is still in its infancy". Although a lot of research has been performed in this field, a thorough insight still lacks. In the next two sections, we focus more specifically on the optical properties of leaves and we present some well known leaf models.

6.2.1 Leaf optical properties

Remote sensing is based on the measurement of reflectance. With a spectrometer and an integrating sphere, transmittance can also be measured in the lab. The remaining incident light is absorbed. The proportions of these three types of

interactions of light with a leaf are wavelength dependent and change with the angle of incidence, surface roughness, chlorophyll, water content and biochemical composition of the leaves [227].

Visible spectrum (400 nm–700 nm)

Pigments within a leaf selectively absorb light depending on the wavelength. This results in a quite specific spectral reflectance signature. Chlorophyll-a is a primary photosynthetic pigment in plants. By absorbing light (with absorption peaks at 430 nm and 660 nm), it enables photosynthesis, the transformation of light into chemical energy. Chlorophyll has a relatively low absorption in the green part of the spectrum, giving vegetation its green color. Chlorophyll-b, carotenoids and xanthophylls are accessory pigments with complementary absorption bands in the visible part of the spectrum. Absorption by carotenoids and xanthophylls is dominant in the blue part of the spectrum only. Because they are degraded less early than chlorophylls [195], leaves turn yellow and red in autumn.

The spectral signature at the end of the visible spectrum is characterized by a *red-edge* [42, 103]. It is represented by a sharp transition between low reflectance in the red, to a high reflectance in the near-infrared region of the spectrum (800 nm). It corresponds with the wavelength interval at which chlorophyll ceases to absorb light close to the near-infrared.

Near-infrared spectrum (700 nm–1300 nm)

The quantum energy per photon in the near-infrared (NIR) region of the spectrum is not sufficiently high for photochemical reactions to take place and hence NIR is not absorbed by the foliar pigments [152]. Therefore, common to all vegetation species, is a high reflectance and transmittance in the NIR. For these wavelengths, leaf optical properties are dominated by the cells within a leaf, as well as by the distribution of inter-cellular air spaces in the *mesophyll* layer [29]. A large number of air spaces will induce more scattering, and a lower transmittance in the NIR. There are also a few water absorption bands related to leaf moisture content near 970 and 1200 nm. However, most of the water absorption bands have peaks in the MIR [47].

Middle-infrared spectrum (1300 nm–2500 nm)

The middle-infrared (MIR) is dominated by the absorption characteristics of water. The absorption is strong and increases with longer wavelengths. The water absorption peaks are centered at 1450 nm, 1940 nm, and 2500 nm [47] and extend over large bandwidths. Leaf (bio)chemicals such as lignin, cellulose, starch, proteins and nitrogen also induce absorption in the MIR region. However, the absorption is much weaker, and can best be analyzed with dry leaf spectra [227]. The absorption by dry matter is uniform up to around 1400 nm from where it starts to increase slowly with wavelength [225].

6.2.2 Leaf optical models

Physically based models describe the radiative transfer within leaves using absorption and scattering coefficients. Allen [6] introduced the first plate model, representing a leaf as a single absorbing plate with three input parameters: the refractive index, n, the incidence angle, α, and the absorption coefficient, a. The anisotropic reflectance of leaves was modeled by giving the plate a rough surface. Allen and Gausman [5] later improved the model by considering the leaf no longer as a single plate, but as a stack of N plates. The plates are separated by air spaces, representing inter-cellular air spaces.

The PROSPECT (Leaf Optical Properties Spectra) model of Jacquemoud and Baret [108, 110] is based on the N-plate model. It is the most widely used model for broadleaf canopies [107, 180, 233]. The authors modeled the anisotropic characteristics of leaves by giving the upper plate a different reflectance and transmittance coefficient and separating it from the N-1 other plates. Inside a leaf, the light flux is assumed to be isotropic because of multiple scattering. The structure parameter, N, was also extended from a discrete to a continuous number. The PROSPECT model calculates leaf hemispherical reflectance and transmittance in the spectral range of 400 nm to 2500 nm using three principal input parameters: chlorophyll content, C_{ab} (μg cm^{-2}) , equivalent water thickness, C_w (cm), and the number of plates, N. Equivalent water thickness is defined as the water mass per unit leaf area in cm. The number of plates, N, is also referred to as an internal structure parameter. It influences reflectance for the whole optical domain covered by the model. The effect is most strong expressed in regions of high reflectance such as the NIR region, where leaf structure is dominant. In the visible region of the spectrum, the internal structure effect is masked by the absorption of photosynthetic pigments. A larger value of N corresponds with a more complex internal structure, more scattering and a higher reflectance. Improved versions of PROSPECT include leaf biochemical coefficients. Initially, Jacquemoud [110] used two input parameters, leaf protein content, C_p (g cm^{-2}), and leaf cellulose + lignin content, C_c (g cm^{-2}). However, experiments showed that protein content could not be retrieved due to strong water absorption features [73, 110, 74]. Also cellulose and lignin could not be quantified successfully. Baret and Fourty [13] used a simplified model with only a dry matter, C_m (g cm^{-2}), as a lumped biochemical constituent. This is also the current version of PROSPECT to date. We have used both the version with 2 and only 1 biochemical coefficients.

6.3 Canopy level

6.3.1 Motivation for canopy models

Leaf radiative transfer models provide leaf reflectance and transmittance from leaf biochemical content (chlorophyll, water content and dry matter) and structural parameters (N in case of PROSPECT, or a description of the cells in case of ray tracing). However, with remote sensing and the use of airborne or spaceborne

spectrometers, it is canopy reflectance which is measured, not leaf reflectance. In that respect, the upscaling from a single leaf to canopy is not a trivial task. The transition from the leaf to the canopy level introduces effects due to, e.g., a variable solar illumination intensity and angles of observation, atmospheric conditions, vegetation canopy architecture and understorey [52]. Daughtry [51] found that variations in background reflectance and *Leaf Area Index* (LAI) complicated the detection of relatively subtle differences in canopy reflectance due to changes in leaf chlorophyll concentration. Conversely, high LAI canopies allow weak leaf-level biochemical information to be enhanced at the canopy scale via multiple scattering [12]. Also *Leaf Angle Distribution* (LAD) is an important canopy structural parameter controlling the coupling strength between leaf and canopy reflectance. Vertically (*erectophile*) oriented leaves will soften this coupling for (near) nadir observations, by increasing the influence of the understorey.

A generally applied approach is to use the output of a leaf RT model as the input of a canopy reflectance model. A canopy model thereby focuses on the effects of leaf structural parameters (LAD and leaf size) and canopy (LAI) on spectral reflectance. Depending on the complexity of the model, multiple scattering within and between canopies, as well as between canopy and soil, is taken into account. Geometric-optical models also account for the influence of 3D canopy structure.

Pinty [180] initiated and coordinate the RAMI (RAdiation Transfer Model Intercomparison) experiment to inter-compare canopy models. The goal of this RAMI model intercomparison is to provide a relative basis to compare a large variety of computer codes modeling the radiation fields at the top of canopies typically encountered in remote sensing. With RAMI, the model developer downloads benchmark cases and solutions to test a model. The overall benefit to the BRDF and remote sensing communities is a demonstration of the maturity and especially consistency with other models to foster a better understanding of how these models are to be used in the practical interpretation of remote sensing data. The first edition (RAMI-1) started in 1999. In the current version, (RAMI-3), 18 models are registered at this moment in time.

Similar to the leaf RT models, we can categorize the canopy models into physical models and computer simulation models. Increasingly popular are the semi-empirical models that combine the advantages of physical as well as empirical models. In the following section, we briefly present some fundamental concepts of canopy models, thereby focussing on the ones we have used in a case study. A more theoretical overview has been written by Goel [85], and was updated by Goel and Thompson [88]. Very comprehensive are the reviews by Miller [161], McCloy [156, chap.7.4], and Van der Meer [227]

6.3.2 Canopy reflectance models

A prime advantage of using physical models is that they are based on physical processes. As a result, their parameters have a physical meaning that can be estimated reasonably well. This is important when we want to estimate those parameters *a posteriori*, given some observed reflectance values. The efficiency

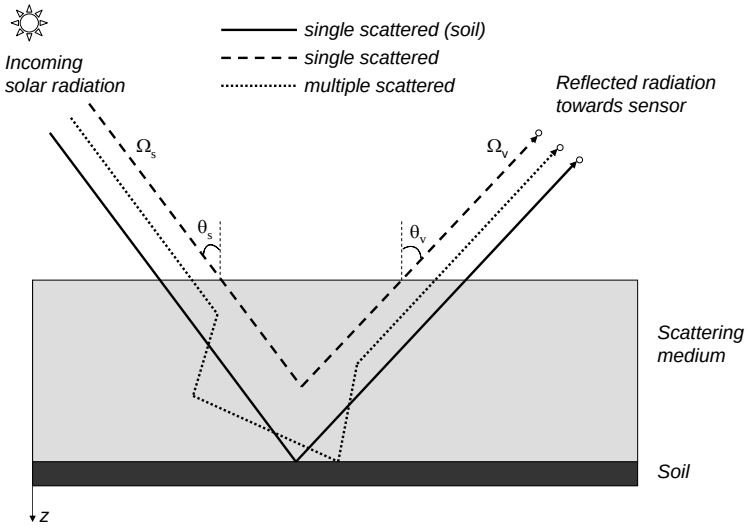


Fig. 6.1: *Schematic representation of the turbid medium model of vegetation canopy as for a single layer. The direct radiation from the Sun causes single and multiple scattering.*

of model inversion increases considerably if we constrain the model parameters with physically realistic values. Moreover, the risk that the optimization process involved in the inversion gets captured in a local minimum of the merit function can be significantly reduced for well chosen initial parameter values.

In Turbid Medium (TM) models, the canopy is modeled as (a number of) homogeneous layers above and plane-parallel to the background surface as in Fig. 6.1. These layers are assumed to be extended into infinity. Scattering and absorbing facets of the canopy are assumed to be infinitesimally small as well as randomly oriented and distributed in each layer.

The SAIL (Scattering by Arbitrarily Inclined Leaves) model [231] is one of the most frequently used RT models. Since its first introduction in 1984, a family of improved SAIL models has been developed since then [111]. SAILH is an extended version of SAIL including the *hotspot* effect. The hotspot is the retro-reflectance peak caused by the minimum amount of shadowed foliage and soil that is observed along this direction (the direction of the sunrays). Everything one observes in the hotspot is sunlit, so this causes a peak in the reflectance [232]. The assumption of very small scatterers and absorbers could not explain this effect in previous versions. In SAILH, the size of the leaves and the associated shadowing effects are taken into account for the calculation of the single scattering contribution to the bi-directional reflectance. As far as single scattering is concerned, it is not a TM model. Similar improvements have been elaborated by Myneni [166]. He

considered voids between finite dimensional canopy facets to model the dimensions of scattering elements explicitly. The next model in the SAIL family is GeoSAIL [233]. It adds a second layer in the canopy model to mimic a vertical leaf color gradient often seen in agricultural canopies. The structural properties (leaf angle distribution and leaf size) are assumed to be identical in both layers, but the LAI for green and brown leaves may differ. The division of the LAI (leaf area index) for both types of leaf over both layers is governed by the parameters fraction brown leaves f_B and the so-called dissociation factor D. The extreme values $D = 1$ and $D = 0$ correspond to complete dissociation and a homogeneous mixture respectively. Furthermore, GeoSAIL incorporates a sub-model for soil's reflectance [10]. It describes the influence of soil moisture in the top layer of the soil on the soil reflectance spectrum. The input is the reflectance spectrum under dry conditions and the expected soil moisture percentage. The latest improvements are included in 4SAIL2. The soil model is extended for non-Lambertian soils and has been improved for numerical robustness (singularity removal) and speed.

Another family of canopy models is developed by Kuusk. The first Kuusk model, MultiSpectral canopy Reflectance Model (MSRM [134]), combines the single scattering approximation of the Nilson-Kuusk model [173] with the diffuse scattering approximation of the SAIL model [231]. It assumes a homogeneous canopy and accounts for specular reflection of direct Sun radiation on leaves, the hotspot effect, and a two-parameter elliptical LAD. The soil reflectance spectrum is modeled by Price's base functions [184]. Good results can be obtained for homogeneous dense canopies. However, for clumped vegetation, the assumed Poisson random distribution for the stand geometry [172] is not realistic [135]. The successor model, MCRM [135], introduced a Markov chain to describe canopy architecture. It includes a Markov parameter corresponding with the level of clumping. A value smaller than 1.0 represents clumping and a value equal to 1.0 corresponds to a Poisson random distribution (no clumping). The last version of this model family, ACRM [136], introduces an additional canopy layer in the Markov chain canopy reflectance model.

Jacquemoud [107] compared four canopy reflectance models in forward and inverse mode, including MCRM and SAIL. The canopy models were coupled with the leaf RT model PROSPECT. A thorough model comparison covering all possible inputs is very complex. The authors concluded that there is no single best canopy model. Synthetic as well as field measurements were included in the test cases. The MCRM model was fast and produced the best estimates for chlorophyll concentration. The SAIL model performed best for LAI. However, the older version of the SAIL family was tested (not taking into account the hotspot, including only a single canopy layer and not optimized for speed). We selected the MCRM model in our case study on stress detection in peach orchards, thereby selecting foliar chlorophyll content as our main stress indicator.

The turbid medium models have been applied with success for relatively uniform and dense canopies. Although the clumping factors in the improved models have increased their scope, the modeling of open forests canopies with TM

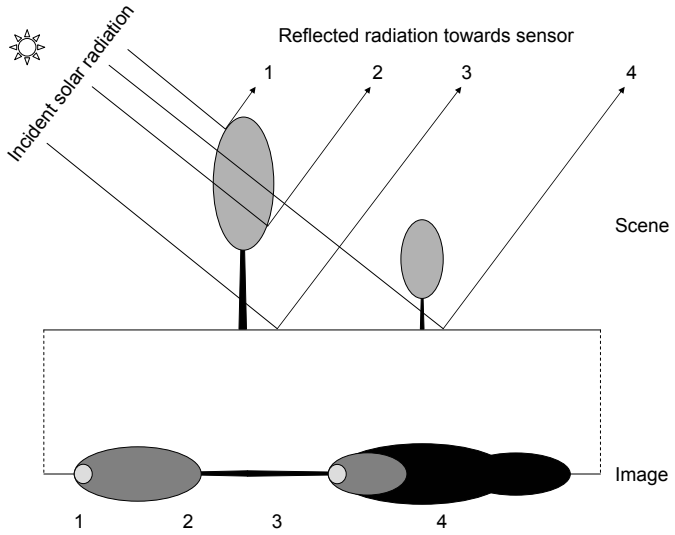


Fig. 6.2: *Two plant objects modeled as cones and spheres by a Geometric-Optical model*

models remains problematic. TM models do not take into account tree crown closure, tree density, tree height, shape and dimension of crowns. In contrast, Geometric-Optic (GO) models describe the canopy as a collection of discrete object (facets), accounting for the influence of canopy architecture on reflectance. Li and Strahler [147] represent forest canopies as cones and spheroids, arranged on a surface under the assumption of parallel-ray geometry (Fig. 6.2). They first considered non-overlapping cones only, but later mutual shadowing was accommodated for [148]. Canopy reflectance is calculated as a linear mixture of four components: sunlit and shadowed canopy (1 and 2), and sunlit and shadowed soil (3 and 4), as shown in Fig. 6.2. The weights are computed proportional to the projected areas of the respective components that contribute to the reflectance.

In pure GO models, the canopy consists of objects with an opaque Lambertian reflectance. This has been further developed by representing the canopy objects themselves with a TM model. This hybrid approach [149, 146] combines the ability of GO models to describe the discrete nature of a discontinuous canopy (shadowing, hotspot), with the accuracy of which RT models describe radiation absorption and multiple scattering. This powerful approach has been proved very successful if the actual canopy architecture can be described in a realistic way.

Now that computing power becomes available at a lower price, numerical approaches become feasible. As an example of a numerical approach, we focus on ray tracing. The basis for the selection of photon trajectories and various other aspects of ray tracing is the Monte Carlo method [167, 59]. In general, Monte Carlo meth-

ods are a simple, robust and powerful set of tools to solve large multi-dimensional systems. They characterize the behavior of these systems, by stochastically sampling a Probability Density Function (PDF) [95]. The problem of describing a vegetation canopy can indeed be considered as a complex and multi-dimensional problem area. The corresponding PDF defines the probability of scattering at particular angles. Hence, sampling of this PDF then characterizes the behavior of possible photon trajectories in the scattering canopy matrix. The Monte Carlo method is used to sample possible photon trajectories.

Within a given medium with constant density, a ray, describing a photon trajectory, will travel along a straight line. Consequently, the main computational issue is to test the intersection of a set of rays with a defined canopy geometric configuration. For canopy reflectance modeling, a photon is either scattered (reflected or transmitted) or absorbed, i.e., the respective probabilities P_r, P_t, P_a sum up to one. The integral over all conditions can be simulated, using the Monte Carlo method, by generating a random number, $\mathcal{P}$, over the interval $(0, 1]$. If $\mathcal{P}$ is less than or equal to P_a then the photon is absorbed. If $\mathcal{P}$ is between P_a and $P_a + P_r$ the photon is reflected. Otherwise, the photon is transmitted. In the case of reflection and transmission, the BRDF of the scattering facet (e.g., leaf reflectance function) is used as a probability density function to relate another random number to a scattering direction. The process is iterated until the photon is absorbed, or escapes from the canopy. Hence, this estimates reflectance as well as absorptance.

If one is only interested in reflectance, all rays that represent absorbed photons do not contribute to the final estimation and thus are a burden with respect to computing time. North [174] used an alternative approach in his FLIGHT model. Instead of a "hard" decision (scattering or absorption), the intensity and directionality of a ray are represented as a continuous function. This function is defined by the conditional probability of scattering from direction Ω_s to a unit solid angle about Ω_v, given that a collision occurs with scattering element at point $\boldsymbol{r}$ ($\Gamma(\boldsymbol{r}, \Omega_s \rightarrow \Omega_v)$). Since all trajectories followed contribute to the final reflectance estimate, the computing time burden is annihilated. The different steps involved are presented in algorithm 6.1.

Algorithm 6.1 Ray tracing canopy model FLIGHT

repeat
  Initialize photon position, direction and intensity: $\boldsymbol{r} = \boldsymbol{r}_0,\ \Omega_v = \Omega_v^0,\ I_\lambda^0 = 1$
  repeat
    Simulate the free path s to the next collision
    $\boldsymbol{r} \leftarrow \boldsymbol{r} + s\Omega_s$
    Simulate the new scattering direction Ω after collision at $\boldsymbol{r}$
    $I_\lambda \leftarrow \Gamma_\lambda(\boldsymbol{r}, \Omega_s \rightarrow \Omega_v) I_\lambda$
  until photon has left canopy
  Accumulate intensity I_λ in bin corresponding to angle Ω
until an acceptable number of photons is traced

Radiosity models [25, 86, 41] are based on a radiation energy balance model. Unlike the physical RT models, they assume a deterministic distribution of all scattering and absorbing elements in a scene. The radiant exitance (radiosity) leaving a surface then equals the sum of its own (thermal) emission with the contribution of all other facets in the scene that illuminate that surface. The initial computing time required for radiosity models grows exponentially with the number of elements in the scene. However, once the equations have been solved, canopy reflectance can be simulated for any viewing geometry angle.

6.4 Parameter estimation

Elvidge [64] measured the reflectance signatures for pure biochemicals in leaves. he found that their characteristic and pronounced absorption features fade substantially in the reflectance curve of a fresh leaf, since the biochemical constituents are bound in more complex organic molecules. In addition, the molecules and pigments are part of complex membraneous structures of cells in the leaf cell. Evidently, this increases the complexity of biochemical parameter estimation. We will present different methodologies to retrieve parameters from canopy reflectance. Empirical methods are briefly introduced first. However, our focus is on model inversion, the main subject of this second part.

6.4.1 Empirical methods

Empirical models simplify the image data, for instance by converting them into a vegetation index. This index is then related to a vegetation parameter of interest by an empirical function. Alternatively, regression analysis can also be applied directly to (a) spectral band(s), in combination with a feature selection technique.

Empirical methods are popular because of their simplicity and the desire to be able to derive estimates of canopy status without the use of field data. Several studies have introduced narrow-band spectral indices to this end [33, 175, 81]. More specifically, Daughtry [51], Zarco [244], Lichtenthaler [151], Gitelson [83] and Haboudane [94], derived indices for chlorophyll content estimation, allowing the detection of chlorosis in an early stage. One of the difficulties with these indices is that they are responsive to other vegetation and environmental parameters such as LAI, and underlying soil reflectance [51, 123, 243]. Hence, relationships obtained at leaf level cannot be transferred to the canopy level. Therefore, derivative indices have been proposed, especially related to the red-edge [162], to estimate chlorophyll at both the leaf and canopy levels. However, in this case, structure and canopy architecture are critical boundary conditions and may cause the empirical relationships to fail for different canopy structures or other phenological stages. Moreover, the relationships at canopy level must be calibrated for spectral and spatial sensor characteristics as well as for viewing geometry conditions [165, 235]. Due to the sensitivity of the indices to these boundary conditions, the precise estimation of biochemical constituents remains problematic.

Regression is a statistical method, based on a reference data set with a known set of parameter values. Its aim is to establish a functional description of the data [61]. Often, as in this context, the objective is to predict a parameter value, the *response value*, for new input (*explanatory variables*). The response value can be measured in the field (or lab), using a reference methodology. Most regression approaches are based on a linear function to describe the observed data in function of an index or reflectance (linear regression). The explanatory variable(s) in the regression analysis is (are) the reflectance value(s) in one (*simple linear regression*) or more bands (*multiple regression*). The linearity of the regression refers to the relation of the response to the explanatory variable(s). The explanatory variable itself can be a non-linear combination of spectral bands, as is the case for many spectral indices.

Wessman [239], Curran [48] and Martin [155] derived a predictive algorithm for chlorophyll from a training data set, by estimating the statistical relationship between canopy chlorophyll concentration and specific bands for leaf or canopy reflectance. Camps [31] suggested a robust regression, based on support vector machines that is particular useful in case of limited in situ measurements.

Regression can be performed both on leaf and canopy levels. However, Grossman [93] demonstrated that the predictive algorithm obtained from regression, trained on a specific site and crop, is not reliable for other conditions. The selected bands depend on the data set at hand, influenced by species, canopy structure and viewing conditions. Several techniques have been introduced to increase the robustness of biochemical parameter estimation, including continuum-removal [130], band ratioing [37, 23] and normalized difference vegetation indexing [123, 51].

6.4.2 Model inversion

Another technique to estimate (vegetation) parameters from reflectance data is by inversion of a coupled leaf and canopy reflectance model system. The study of model inversion was first described by Goel and Strebel [87]. In principle, every canopy model can be used, but the larger the number of model variables, the more difficult the inversion. The constraint on invertibility, where only a limited number of variables can be estimated with a sufficient degree of accuracy, limits the inversion of the complex canopy models [52].

To enable the measurement of the goodness of fit between an observed ($\mathbf{R}$) and modeled ($\hat{\mathbf{R}}$) reflectance spectrum, a *cost function* (also referred to as merit or error function) is needed. This is typically the sum of the squared differences for each spectral band ($i = 1 \dots d$):

$$J(\boldsymbol{p}) = \frac{1}{d} \sum_{i=1}^{d} \left(R_i - \hat{R}_i(\boldsymbol{p}) \right)^2 , \tag{6.1}$$

where the vector $\boldsymbol{p}$ represents the model parameters. Alternatively, the cost function can be expressed in matrix notation as:

$$J(\boldsymbol{p}) = \left(\mathbf{R} - \hat{\mathbf{R}}(\boldsymbol{p}) \right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{R} - \hat{\mathbf{R}}(\boldsymbol{p}) \right) , \tag{6.2}$$

where the covariance matrix, $\mathbf{\Sigma}$, of the signal noise has been taken into account. If the band specific signal to noise ratio is known for a specific sensor, this information can be used as a weighting criterion. Sometimes, a relative cost function is preferable, since it does not emphasize bands or directions with large absolute reflectance values [238]. Other modifications of the cost function (6.1) exist, for example by weighting the contributions of the individual wavelengths. Zarco [244] used a cost function based on an optical index R_{750}/R_{710}. Such a cost function focuses on a single band ratio, rather than an entire spectrum. The authors showed superior results, especially if the reflectance signal includes shadowed pixels.

A well known problem of model inversion is that it may be *ill posed* as different parameter sets may lead to the same solution. It is therefore important to constrain most of the parameter sets, or to fix them to their known (measured) values. Reducing the number of optimized parameters can also be required for performance reasons.

Finding the target parameters $\boldsymbol{p}$ which minimize $J(\boldsymbol{p})$ is the task of the optimization routine. *Analytical solutions* can easily be obtained for models that are linear for the parameters $\boldsymbol{p}$. For most of the models we deal with, this is not the case. Inversion is usually performed based on a numerical procedure with an iterative optimization methodology. It starts with an initial guess and iterates till an optimal parameter set for the observed reflectacne data is found with respect to the cost function.

Gradient descent algorithms are the most popular algorithms for nonlinear optimization. They include steepest-descent, conjugate-gradient, quasi-Newton and Levenberg-Marquardt algorithms [183]. Their main advantages are their relative simplicity and low computational cost. The challenge is to find the *global minimum* with avoidance of potential *local minima*. This is the objective of *global search methods* such as genetic algorithms and simulated annealing (discussed in chapter 2). Neural networks also have been shown to be efficient in the inversion of canopy RT models [205, 3, 124, 200]

Based on our positive experience for other optimization tasks (e.g., band selection), we opted for the use of simulated annealing in the inversion process. The forward model can be applied directly for each iteration step, though lookup tables (LUTs) might be a better option if performance is critical [128, 182]. The use of a LUT is very effective, but the size of the table increases dramatically with the number of parameters. To invert the FLIGHT model in our case study (chapter 7), we used a LUT with chlorophyll values in function of varying: relative azimuth angle, view zenith angle, fraction of ground cover and chlorophyll content.

Chapter 7

Model inversion for multi-angular hyperspectral data

It's easy to make a cake from a recipe, but can you write down the recipe from a cake?

—

7.1 Introduction

Quantitative remote sensing has gained a lot of interest in the last decade, thanks to improved methodologies and sensors. The large number of narrow bands in hyperspectral sensors allow to measure spectral signatures more precisely. On the other hand, more realistic and more accurate models become available.

Another trend that improved the accuracy on parameter retrieval is the use of multi-angular observation data. The angles are taken into account in the canopy models and thus can provide additional constraints. Some sensors provide multi-angular viewing (MISR, POLDER, CHRIS) and thus extract crucial information from this extra angular signature of vegetation. Alternatively, multi-angular observation data with single viewing airborne sensors can be obtained from multiple tracks with different headings.

In the next section, we first deal with spectral matching of a modeled and an observed reflectance. In section 7.3, we then focus on the special case of multi-angular observations and its impact on model inversion. A final case study is

presented section 7.4, on chlorophyll retrieval in peach orchards from multi-angular airborne hyperspectral data.

7.2 Spectral matching

It is clear that for model inversion, the modeled reflectance must be matched to the observed reflectance. More specifically, the cost function (6.1) must use a filtered version of the modeled reflectance, according to the sensor characteristics. However, few papers address this issue explicitly. Moreover, ready to use model software that include a mode for inversion, only resample the modeled reflectance without appropriate filtering. In our implementation for model inversion, we have included a spectral matching within the optimization scheme.

Most models provide reflectance at a defined spectral resolution. Typically, a resolution of 1 nm or 5 nm is obtained. However, a multitude of available sensors exists, each having different spectral characteristics. The modeled spectrum $R^{mod}(\lambda)$ must first be matched to the observed spectrum R for both spectral range and resolution (FWHM) before it can be used as an estimate $\hat{\mathbf{R}}$ in the cost function (6.1). This can be done by Gaussian filtering the model output $R^{mod}(\lambda)$ for each spectral band with central wavelength c_i and standard deviations σ_i:

$$\hat{R}_i = \int R^{mod}(\lambda) e^{-\frac{1}{2}(c_i-\lambda)^2/\sigma_i^2} \mathrm{d}\lambda, \tag{7.1}$$

The standard deviation σ_i can be calculated as $\sigma_i = w_i/(\sqrt{8\ln 2})$, with w_i the FWHM. As an example, Fig. 7.1 shows the modeled spectrum before and after filtering according to the AHS hyperspectral airborne sensor.

7.3 Multi-angular observations

If multi-angular data are available, the cost function (6.1) can be extended [238]

$$J(\boldsymbol{p}) = \sum_{j=1}^{n} \sum_{i=1}^{d} \frac{1}{n} \frac{1}{d} \left(R_{i,j} - \hat{R}_{i,j}(\boldsymbol{p}_j) \right)^2 \tag{7.2}$$

The ensemble of parameters $\boldsymbol{p} = \{\boldsymbol{p}_j \mid j = 1, \ldots, n\}$ now contains the viewing conditions for each observation j. Apart from the angles, biochemical and biophysical parameters remain constant for the different observations. With the multi-angular observations $R_{i,j}$, we thus have additional constraints for the optimization routine. This can be used to increase the accuracy of the parameter estimation. We refer to this as the *direct mode*, because multi-angular data are used directly within the optimization process. In contrast, we can use the obtained results for each observation in a *posterior* mode, by just averaging the individual results. The direct mode better constrains the optimization process and is probably preferred if more parameters must be estimated simultaneously. Likewise, if the model is not robust to different observation conditions, the individual results are not reliable and must

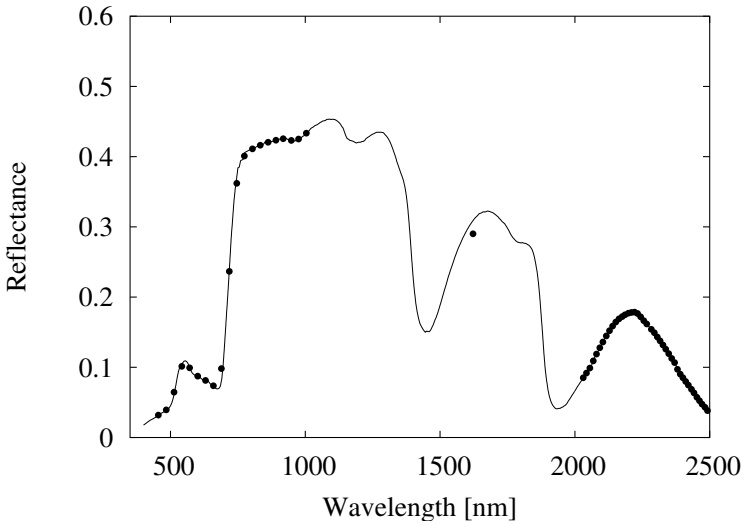


Fig. 7.1: *The modeled canopy spectrum before (solid) and after (dots) filtering according to the specifications of the AHS sensor*

not be combined. However, in case of a robust model, averaging the individual results might be a good alternative.

7.4 Case study on stress detection in peach orchards

7.4.1 Introduction

The objective of this case study is to demonstrate (and quantify) the dependence of the accuracy of parameter retrieval on viewing conditions. This is important if remote sensing is to be applied as an operational tool for precision agriculture. Satellite sensors with large swaths seldom meet the prerequisite of nadir viewing. On the other hand, high spatial resolution sensors suffer from sub-optimal viewing conditions as well. They must be pointed to off-nadir positions to obtain an acceptable revisit time. But even for near nadir observations, the effect of viewing geometry can become important. This is demonstrated for the estimation of chlorophyll from canopy reflectance of a peach orchard. Hyperspectral images were acquired with an airborne sensor (AHS) at near nadir, but with different relative azimuth angles, including forward and backward scatter.

Different methodologies exist to estimate parameters from canopy reflectance. Among these, model inversion is most promising, especially if we aim for a technique that is robust for different observation angles. Myneni [165] and Verstraete [235] showed that indices derived from empirical methods must be tuned for different viewing conditions. Similar conclusions were drawn by Grossman for

regression [93]. Conversely, canopy models are developed to compensate for canopy structure and viewing geometry. However, for the open crop canopy we are faced with, where soil background and shadows dominate the bidirectional reflectance signature, this is far from evident. Zarco [243] found PROSPECT-SAILH results derived for pure crowns to be inaccurate when applied to pixels with aggregated soil background and shadow scene components.

We compared two canopy models in this study. They were coupled to the PROSPECT leaf model [110]. The first, MCRM [135], assumes the canopy layer as a turbid medium. It does not take into account tree crown closure, tree density, height, shape and dimension of crowns. The second is a hybrid model, FLIGHT [174]. It is a more realistic model based on ray tracing, accounting for influences of canopy architecture on reflectance. Bi-directional reflectance effects are calculated as function of the illumination and viewing geometry and scene components such as shadow and soil effects. This is an advantage over the first type of models with respect to the robustness of viewing geometry, as is shown in this study.

We also investigated how multi-angular remote sensing data can increase the accuracy of chlorophyll estimation. Multi-angular viewing data are now available. For example, the MISR (Multi-angle Imaging SpectroRadiometer [58]) instrument onboard NASA's Terra satellite has nine sensors with different viewing angles and the CHRIS (Compact High Resolution Imaging Spectrometer [2, 15]) instrument is pointable off-nadir in both the along-track and across-track direction, by tilting the PROBA platform. For this study, the peach orchard was imaged in four tracks. Each track resulted in a different viewing geometry. We tested different techniques to combine the multi-angular data. The accuracy of chlorophyll retrieval was increased for both models. Still, the 3D model outperformed the turbid medium model.

In section 7.4.2, we present the experimental setup of the peach orchard. We then explain how data were collected in the field (section 7.4.3) and introduce the airborne hyperspectral data (section 7.4.4). Results are presented and discussed in section 7.4.5.

7.4.2 Experimental setup

A peach orchard in Zaragoza, Spain (Lat: 41.28 North, Long: 1.22 West) was chosen as a test case to experiment with the chlorophyll retrieval techniques. It was treated with iron chelates to recover from iron chlorosis conditions. The orchard, represented as a matrix in Fig. 7.2, consists of 35 rows and 6 columns. The total number of trees is only 205 instead of 210, because 5 trees are missing. Iron chlorosis is induced in 48 trees, located in the 2 rightmost columns of the orchard. Trees are represented, and treated, in blocks of 3. As an exception, the blocks at the upper end of the plot only contain two trees. The treatment consisted of an iron chelate with 4 different concentrations (Fig. 7.2): 0g/tree, 60g/tree, 90g/tree and 120g/tree.

The objective was to create a dynamic range of chlorophyll concentration.

Column 3 contains 12 trees (4 blocks of 3) that were grafted during the previous year (2004).

7.4.3 Field data collection

Fresh leaves were sampled for each tree and measured with the ASD spectrometer using a Leaf Clip. Leaf reflectance was obtained for 716 leaves.

For chlorophyll content, two reference methods were used. The first method consisted of a (destructive) chemical analysis and was performed in the lab. Chlorophyll was determined by spectrophotometry of tissue extracts (using ethanol as a solvent). Although this method is relatively reliable, it is tedious and slow. Moreover, special care had to be taken for transporting the picked leaves from the field to the lab in order to delay the degradation process. Leaves must be kept cold, preferably frozen, in a dark place. A faster alternative was to use a portable chlorophyll meter. The SPAD 502 (Minolta) determines chlorophyll content in a non-destructive way, based on measurements of light transmittance and absorptance of the leaf. This method is not as reliable, but allowed for a larger sample size. The SPAD values were calibrated by comparing the SPAD values to chlorophyll concentrations derived from the destructive chemical analysis. The result is shown in Fig. 7.3. A correlation coefficient of 0.82 was obtained with a RMSE of 6.0 μg/cm^2. The values were then averaged per tree to compare with parameters obtained on a canopy level. The plot average for chlorophyll content is 36.9 μg/cm^2, with a variance of 53.4 (σ = 7.3). Extreme values range from 13.7 to 45.7 μg/cm^2, showing a large dynamic range as intended.

In addition to chlorophyll, leaf area index was estimated with an LAI2000 instrument for 10 trees. The calculations for LAI were performed according to [237], resulting in a mean value of 2.

7.4.4 Airborne hyperspectral data

The AHS sensor has 63 bands covering the visual and near infrared part of the spectrum (450 nm-2500 nm). It was mounted on a CASA C-212 airplane and operated by INTA. The peach orchard was acquired on July 12, 2005 in cloud free conditions (Fig. 7.4). Four tracks were flown with different headings (Fig. 7.5). The tracks were headed with a 45° difference, with track 1 perpendicular to track 3 and track 2 perpendicular to track 4. The individual tracks fully covered the peach orchard, so that each pixel was acquired with 4 different view angles. Optimal flight conditions are for track 1, corresponding to a flight path in (and measurements perpendicular to) the solar plane (heading away from the sun). The small peach orchard is located in the center of each track, leading to near nadir observations. The pixels were thus selected from a narrow range in the image, where relative azimuth angles correspond to a continuous transition from the left/right part of each scene with respect to nadir. As an example, selected pixels for track 1 were not observed with the expected relative azimuth angles of 90° and 270°. This is illustrated by the scatterplot in Fig. 7.6, which schematically shows the view

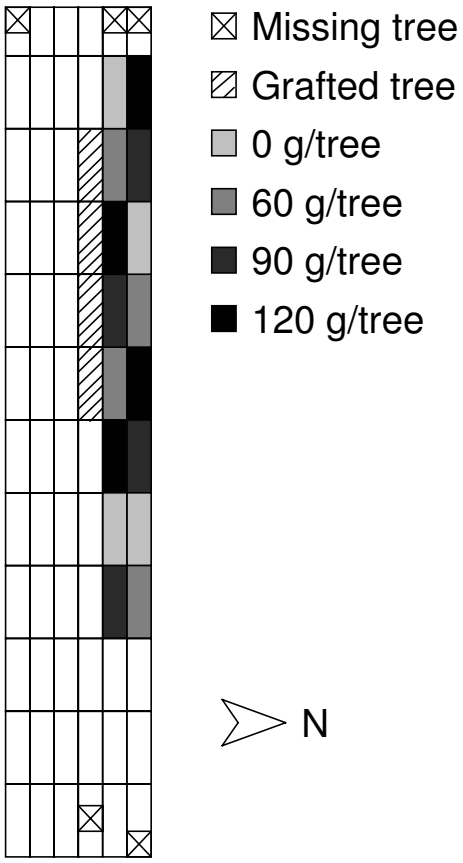


Fig. 7.2: *Schematic overview of the plot. Trees are grouped per 3 (except for upper row). In total, 48 trees in the 2 rightmost columns were treated with iron chelate with 4 different concentrations (from light to dark). Column 3 contains 12 trees that were grafted during the previous year.*

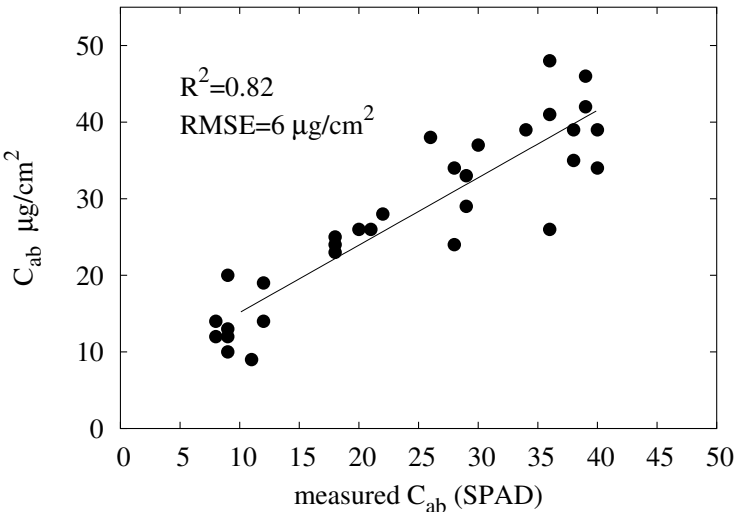


Fig. 7.3: *Calibration of the SPAD values using C_{ab} concentrations derived from chemical analysis in the laboratory.*

zenith and relative azimuth angles for each pixel and for the different tracks. View zenith is minimum for track 1 (between 4° and 5°), and maximum for track 3 (near to 10°). Tracks 1 and 4 are acquired in backward scatter, and tracks 2 and 3 in forward scatter.

Image data were processed to top of canopy level, using in house developed software [20]. Atmospheric correction, based on MODTRAN, is made column dependent in order to account for the different atmospheric paths across AHS images properly. If one single set of atmospheric parameters were used instead, meaning that all the columns in the image are processed assuming one single observation angle, angular variations detected in the chlorophyll estimations could be due to atmospheric effects. A customized LUT was calculated using a single platform elevation (983 m) and solar zenith angle (27.75°), 3 view zenith angles and 5 relative azimuth angles. The atmospheric correction used in situ measured parameters, e.g., sunphotometer and ASD spectral measurements over white and dark reference targets (see Fig. 7.4). The pixel positions were calculated using GPS measurements and orientation parameters (by means the Inertial Measurement Unit) onboard the airplane. The image was geocoded in a Universal Transverse Mercator projection system (UTM, Zone 30 North, datum WGS84). A sub-pixel accuracy was obtained after geometric correction with a digital elevation model of the site. The ground resolution of 2.5m was similar to the crown diameter, which complicated tree identification in the image. The peach trees have been planted according to the scheme in Fig. 7.2 every 4 m ($\frac{8}{5}$ pixels). To facilitate the tree extraction, the hyperspectral image was upsampled with a factor of 5 (with nearest

Fig. 7.4: *Image of the orchard acquired with the AHS airborne hyperspectral sensor. The white (WT) and dark (DT) reference targets are visible on the right hand side of the orchard.*

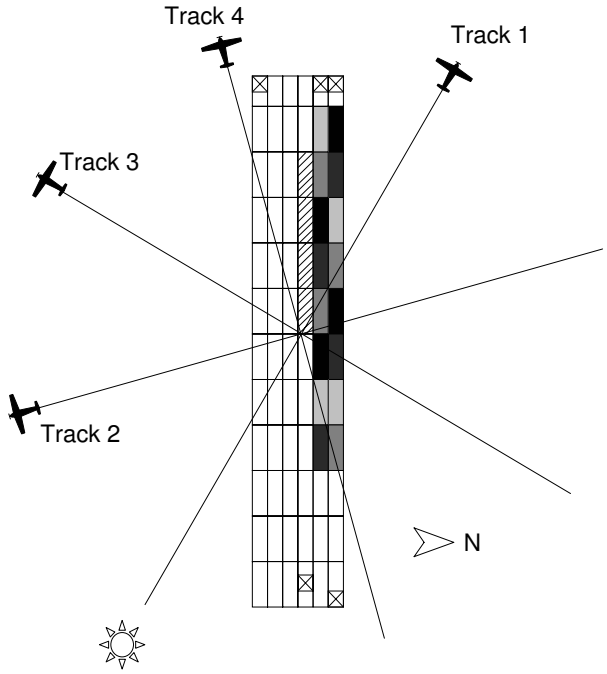


Fig. 7.5: *Schematic view of the peach orchard with headings for the 4 tracks.*

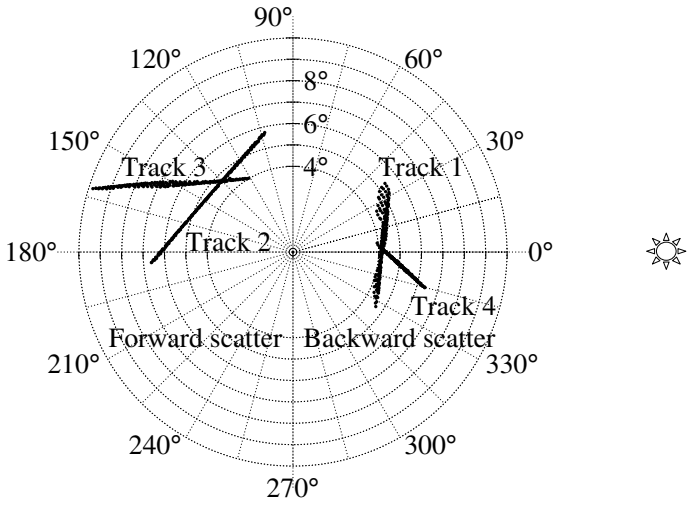


Fig. 7.6: *Polar plot of relative azimuth and zenith viewing angle for each track. The individual tracks fully covered the peach orchard, resulting in 4 different observations for each pixel (represented as dots).*

neighbor). In this way, a tree was covered in a region of interest (ROI) with an integer number of (sub)pixels. Using nearest neighbor, the subpixels contained exactly the same spectral information as the original pixels. To obtain the target canopy reflectance, we then took the median reflectance of the subpixels in the ROI. An alternative is to select the pixel in the ROI with maximum NDVI or maximum NIR value to minimize the influence of shadows and understorey (canopy openings between rows) as in [244]. However, in our case a ROI could contain sub-pixels of neighboring trees due to imperfect tree positioning. To reduce the effect of geometrical inaccuracy and the limited spatial resolution, we averaged the estimated values for a block of 3 neighboring trees. This choice was also motivated by the way in which the trees were treated (section 7.4.2). On canopy level, we estimated chlorophyll for these individual blocks. This has a severe impact on the variance of the chlorophyll estimation results. This should be taken into account when results are compared to similar studies in the literature. In most cases, larger areas of plots are aggregated, reducing noise and edge effects. With respect to the potential of the proposed techniques for precision agriculture, we chose to assess the crop status with considerable detail.

7.4.5 Results and discussion

We started from a leaf level before scaling up to canopy level. This step-wise approach allowed to assess the performance of the selected leaf radiative transfer model without entering the complexity of the canopy level. Second, by inverting the leaf model to extract foliar chlorophyll, other biochemical parameters were estimated as well, which could then be constrained on a canopy level. A simple linear regression served as a reference, rather than a final goal. It demonstrates the capability to derive chlorophyll with a good accuracy from leaf/canopy reflectance. Furthermore, it was shown that a single band contained sufficient information for this purpose.

In situ data: leaf level

A predictive algorithm was first derived from a simple regression applied to the ASD leaf spectra. As we were not interested in the results on leaf level as such, we averaged the estimated chlorophyll for all leaves in a single tree. We thus obtained a RMSE value of 1.78 μg/cm^2 on tree level, obtained from the leaf spectra. In a second step, we filtered the ASD spectra with the AHS spectral response function (SRF) to check if the AHS sensor was sufficiently conceived for the task of chlorophyll retrieval. In fact little degradation was observed due to the filtering process as shown in Fig. 7.7 (RMSE value increased from 1.78 to 2.03 μg/cm^2).

An interesting aspect of the filtering process was that the optimal band for simple linear regression shifted from the red-edge (702 nm) toward the green peak (571 nm). Indeed, for narrow bands, the red-edge provides valuable information on chlorophyll. However, for broader bands, the green peak was better suited for this experiment.

We obtained good results with the inversion of PROSPECT, although the RMSE was not as low as with regression (6.48 μg/cm^2 and $R^2 = 0.78$, Fig. 7.8).

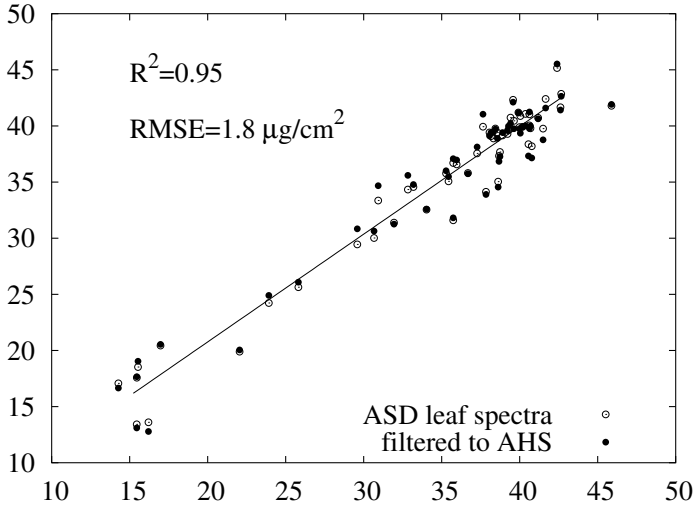


Fig. 7.7: *Chlorophyll estimation through simple linear regression of the ASD leaf spectra (solid dots are filtered with the AHS spectral response function).*

All 5 PROSPECT parameters were estimated simultaneously. Simulated annealing was used for optimizing the cost function. Table 7.1, shows the mean values for all leaves measured. Apart from chlorophyll, there was very little variation in biochemical parameters, so we fixed them for further analysis on canopy level.

Tab. 7.1: *Estimated PROSPECT parameters (mean over all leaves)*

Parameter	Value	Variance
Water equivalent thickness (cm)	0.0121	$9.98 \cdot 10^{-7}$
Leaf protein content (g/cm^2)	0.0002	$1.8 \cdot 10^{-20}$
Leaf cellulose and lignin content content (g/cm^2)	0.0031	$1.02 \cdot 10^{-7}$
Elementary layers inside a leaf	1.8	0.0047
Chlorophyll concentration (μg/cm^2)	39	98

The next step was to estimate the chlorophyll on a tree level, obtained from the airborne hyperspectral data.

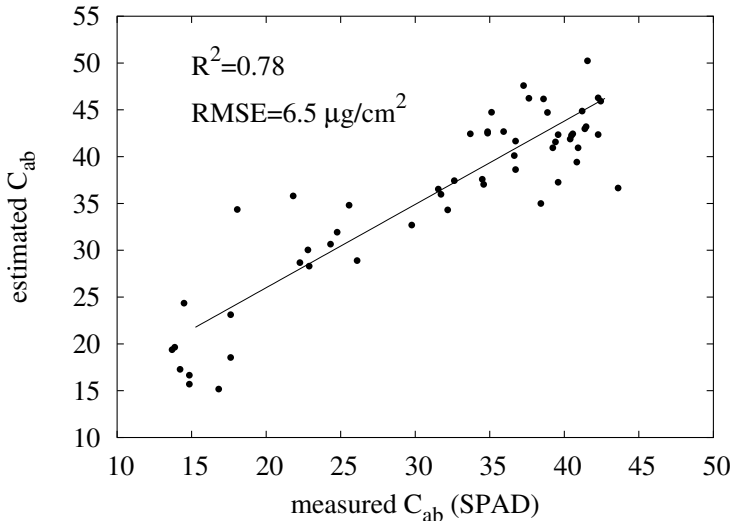


Fig. 7.8: *Chlorophyll estimation through inversion of the PROSPECT leaf radiative transfer model for ASD leaf spectra (filtered with the AHS spectral response function).*

Airborne hyperspectral data: canopy level

The canopy spectra for all 205 peach trees in the orchard were derived from the AHS image as explained in section 7.4.4. A single spectrum was obtained for each tree, by taking the median reflectance for the corresponding region of interest.

Similar to the chlorophyll retrieval from leaf reflectance, we performed a simple linear regression. Independent from the leaf level, the same band was selected on canopy level by the a feature selection routine, confirming the importance of this spectral region (571 nm). Other than just a reference result on canopy level, the motivation for this was to perform a sensitivity analysis on the viewing conditions. It is a generally known problem that the predictive algorithm is site and crop specific and is not reliable for other conditions [93]. In this experiment, crop and site (canopy architecture) was fixed. The time difference between the first and last track was 40 minutes. With clear sky conditions, atmospheric conditions and illumination geometry were nearly identical for the 4 tracks as well (difference in atmospheric path length was negligible because of small viewing zenith angles). By training the predictive algorithm on one track and testing on the other, we were able to assess the effect of viewing conditions on the regression technique. The results in Table 7.2 show there is an important dependence, as expressed by the variance in the last column for most tracks. The first column indicates the tracks used for training, while those for test are shown in the next columns. As an example, if track 4 was used for training (last row), the RMSE for estimating chlorophyll using data from track 3 was 9.7 μg/cm^2. This is more than double

the error obtained when testing on the same track (using a leave-one-out cross validation).

Tab. 7.2: *Assessment of the effect of viewing conditions on simple linear regression. Training was performed using a single track (column 1). RMSE ($\mu g/cm^2$) and R^2 values of C_{ab} are calculated on the test tracks in row 1 respectively.*

RMSE (R^2)	Track 1	Track 2	Track 3	Track 4	Mean	Variance
Track 1	4.5 (0.56)	5.6 (0.52)	6.5 (0.62)	5.3 (0.50)	5.5 (0.55)	0.7 (0.003)
Track 2	5.3 (0.56)	4.8 (0.52)	4.7 (0.62)	6.9 (0.50)	5.4 (0.55)	1.1 (0.003)
Track 3	7.7 (0.58)	5.1 (0.49)	4.4 (0.59)	9.9 (0.60)	6.8 (0.57)	6.4 (0.002)
Track 4	5.1 (0.58)	8.6 (0.49)	9.7 (0.59)	4.5 (0.60)	7.0 (0.57)	6.7 (0.002)

We then inverted the radiative transfer models PROSPECT and MCRM. We derived C_{ab} by first fixing the other four parameters for the PROSPECT leaf model to the mean values from Table 7.1. The canopy model parameters were fixed as well. The sun and view angles were logged for each pixel during the flight and thus could be set to their actual values. The leaf area index (LAI) was fixed to 2, the mean value of the measured subset. The remaining parameters were fixed to the center of the model range. The chlorophyll was loosely constrained (0 to 60 $\mu g/cm^2$). An overview of the fixed values or ranges for the MCRM canopy reflectance model is shown in Table 7.3.

Tab. 7.3: *Parameters used for the MCRM canopy reflectance model*

Parameter	Value
Solar zenith angle	$25° - 30°$
Solar azimuth angle	$121° - 132°$
View zenith angle	$4.4° - 6.6°$
Relative viewing azimuth	$0° - 360°$
Leaf area index	2
Leaf size	0.02
Mean leaf angle (elliptical LAD)	45
Eccentricity of LAD	0.91
Ångström turbidity factor	0.1
Clumping parameter	0.8
Ratio of refractive indices	1
Price weight 1 (soil)	0.2
Price weight 2 (soil)	0
Price weight 3 (soil)	0
Price weight 4 (soil)	0

Tab. 7.4: *Parameters used for the 3D canopy reflectance model FLIGHT*

Parameter	Value
Solar zenith angle	27.75°
Solar azimuth angle	126.25°
View zenith angle	5.35°
Relative viewing azimuth	$0° - 360°$
Total LAI	2
Leaf size	0.02
Leaf area distribution	spherical
Mode	reverse
Dimension	3
Fraction ground cover	$0.6 - 0.9$
Fraction of green cover	1
Number of photons	100000
Soil roughness index	0.1
Aerosol optical thickness (555 nm)	0.1
Crown radius (m)	1.2
Crown center to top distance (m)	1
Min height to first branch (m)	1
Max height to first branch (m)	1.5

The results in the first row of Table 7.5 show a severe impact of the viewing on the performance of the turbid medium model inversion. Best results are obtained for track 1, with a small value for RMSE (5.2 μg/cm^2) and a high correlation coefficient ($R^2 = 0.44$). This track corresponds to the optimal flight conditions. In contrast, track 3, with its sub-optimal viewing conditions, results in a high RMSE value of 12.5 μg/cm^2 ($R^2 = 0.5$). In general, chlorophyll is overestimated with forward scatter acquisitions (tracks 2 and 3). Most surprising is that even for the narrow range of view zenith angles (4°to 10°) chlorophyll estimations with the turbid medium model are that sensitive to relative azimuth angles. Because no reference data were available for pixels with larger zenith angles, we created synthetic data. Using the same turbid medium model in forward mode would be of little relevance due to circular reasoning. We therefore used the 3D FLIGHT model to create the synthetic data and the turbid medium model for inversion. We simulated view zenith angles from 5° to 10° and from 30° to 35° (the AHS sensor provides data with view zenith angles up to 45°). For backward scatter (tracks 1 and 4), there was little impact of larger view zenith angles on the accuracy of chlorophyll estimates. Values for R^2 decreased from 0.98 to 0.86, but RMSE also decreased from 7.7 to 5.8 μg/cm^2. In case of forward scattering though (corresponding to tracks 2 and 3), the estimates were influenced for the worst: RMSE values increased from 13 to 17 $\mu g/cm^2$, with less correlation between estimated and true chlorophyll content (R^2 decreased from 0.84 to 0.6). Similar to our experimental data, there was an overestimation for the chlorophyll content for the

forward scattered simulated data.

Finally, we inverted the 3D FLIGHT model coupled with PROSPECT. For performance reasons, a LUT was used. It was created for different values of chlorophyll ($10-60$ μg/cm^2), fraction cover ($0.6-0.9$) and view azimuth angle ($0°-360°$ in steps of $15°$). The sun angles and view zenith angle (Table 7.3) were approximated to their mean values and fixed in the model. The relevant values (ranges) of the parameter settings for FLIGHT are listed in Table 7.4. Again, best results were obtained for track 1, with RMSE = 5.4 μg/cm^2 and $R^2 = 0.46$. Most importantly, results were much more robust to viewing conditions as shown in row 2 of Table 7.5. Compared to the turbid medium model, variance for RMSE dropped from 10.5 to 0.4, with extreme RMSE values between 5.4 and 6.9 μg/cm^2 (5.2 and 12.5 μg/cm^2 for MCRM). Correlation variated from 0.39 (track 4) to 0.58 (track 3), and was always higher than results from MCRM inversion.

Tab. 7.5: *RMSE (μg/cm^2) and R^2 values of C_{ab} retrieval through model inversion. The 3D model (FLIGHT) is more robust to viewing conditions than turbid medium model (MCRM). Mean and variance values are calculated over RMSE values for tracks 1-4.*

RMSE (R^2)	Track 1	Track 2	Track 3	Track 4	Mean	Variance
MCRM	**5.2** (0.44)	9.9 (0.36)	12.5 (0.5)	6.7 (0.38)	8.6 (0.42)	10.5 (0.005)
FLIGHT	**5.4** (0.46)	6.1 (0.42)	6.4 (0.58)	6.9 (0.39)	6.2 (0.46)	0.4 (0.007)

Tab. 7.6: *RMSE (μg/cm^2) and R^2 values of C_{ab} retrieval through model inversion. Multi-angular information is either directly integrated in the model inversion using equation (7.2) (direct mode) or post-processed by averaging the individual results for each observation (posterior mode).*

RMSE (R^2)	direct mode	posterior mode
MCRM	5.8 (0.48)	6.1 (0.52)
FLIGHT	5.0 (0.52)	4.7 (0.56)

If multi-angular information is available, this information can be used to increase the accuracy of the parameter retrieval. The merit function was adapted according to (7.2). The obtained RMSE was 5.8 and 5.0 μg/cm^2, with a correlation R^2 of 0.48 and 0.52 for MCRM and FLIGHT respectively. Although the 3D model still outperformed the turbid medium model, the difference is small. It should be noted that for this experiment only chlorophyll was estimated. If the optimization scheme must deal with more parameters, the extra constraint of the multi-angular information is expected to be more beneficial. However, for the more robust 3D model, the individual results from the single observations are more re-

liable and contribute to the best estimation (RMSE=4.7 μg/cm^2 and $R^2 = 0.56$). The chlorophyll values, estimated using multi-angular information, are plot against the measured values in the scatterplot of Fig. 7.9. Linear regression, RMSE and R^2 are indicated for both models.

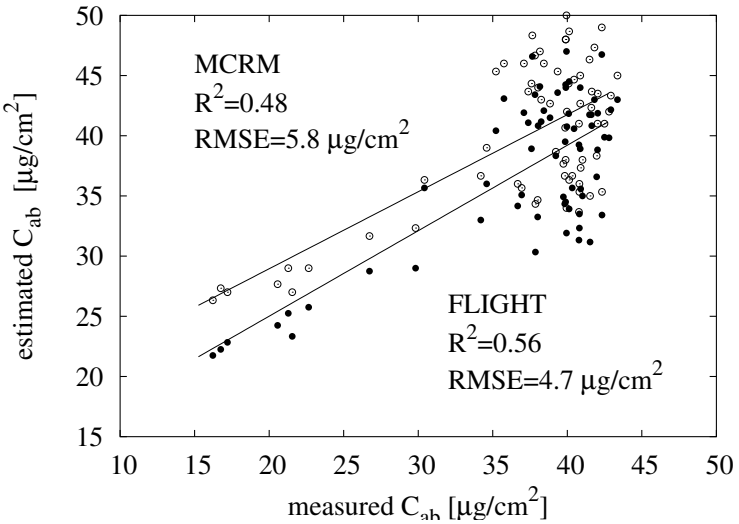


Fig. 7.9: *Scatterplot with measured v.s. estimated chlorophyll values using multi-angular information and model inversion. Correlation coefficients R^2 are 0.48 and 0.56 for MCRM and FLIGHT respectively*

Conclusion

A conclusion is the place where
you got tired of thinking

Arthur Bloch

The processing of hyperspectral remote sensing data, with the objective of optimal information extraction, is dealt with. A data- and a model-driven approach were followed in part I and II. The first is much related to pattern recognition and based on statistical techniques. The second is based on physical principles, explaining the observed reflectance spectrum via radiative transfer models. Which one is the better approach? The answer is application dependent. Classification problems are typically solved using machine learning techniques, which belong to the data-driven approach. For continuous parameters, both approaches are common, i.e., regression and model inversion. The need for abundant training data is an obvious drawback for the data-driven approach. Moreover, the predictive algorithm obtained from regression often lacks robustness. Within the last decade, the introduction of realistic models has increased the popularity of the modeling approach.

It remains a difficult question whether the full dimensionality of hyperspectral data is really needed. There are two aspects involved. First, there is the required spectral resolution (band width of each spectral band). The case study on stress detection in apple orchards in chapter 3 using wavelet based features, showed that 5 wavelet coefficients were selected from the first scale, corresponding to the full spectral resolution of the ASD spectroradiometer data (FWHM of 3 nm). Also for the second case study, illustrating the optimal bands in chapter 4, narrow bands (FWHM of 10 nm) were selected in the red-edge region of the spectrum. Does this prove the need for a very high spectral resolution? Not necessarily. The first case study concentrated on leaf spectra only. Probably, a different result would

have been obtained with real canopy spectra from an airborne sensor. Also for the second case study, the motivation for high spectral resolution is relative. When only 4 optimal bands were selected, the proposed approach showed a significant improvement compared to uniform sampling (resulting in contiguous bands with a FWHM of 100 nm). However, with the introduction of more bands, the difference in classification accuracy decreased to less than 5%. The second aspects is the spectral coverage of hyperspectral data. Clearly, not all applications need a complete coverage of the EM spectrum from 350 nm to 2500 nm. For the estimation of chlorophyll from canopy reflectance, a sensor covering the visible to near infrared region of the spectrum is sufficient. If, however, we focus on water content, the near to short wave infrared region becomes more important. Sensors are designed with specific applications in mind (e.g., MERIS for water applications), which reflects in their spectral bands. However, such a design is based on state of the art algorithms, which depends on current knowledge. Improved methods can require new bands, that are not covered by the original design. These also include preprocessing algorithms, e.g., cloud detection and atmospheric correction. Moreover, sensors with a large spectral coverage can be used for a broad range of applications, and thus can serve a larger user community. The proposed methods for feature extraction in part I of this thesis can then be used to select the relevant features or bands for an application at hand.

Future research includes the investigation of image enhancement techniques (superresolution and image fusion) and their effect on classification accuracy. In this thesis, the focus was on spectral information, with a small contribution on spatial features for smoothing the classification result. A lot of research still has to be performed on non-pixel based (object oriented) techniques. Finally, the data- and model-driven approach can be linked. As an example radiative transfer models can be introduced in classification problems.

Bibliography

References

[1] http://www.itres.com.

[2] http://www.chris-proba.org.uk.

[3] A. A. Abuelgasim, S. Gopal, and A. H. Strahler. Forward and inverse modeling of canopy directional reflectance using a neural network. *International Journal of Remote Sensing*, 19:453–471, 1998.

[4] M. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

[5] W. A. Allen, H. W. Gaussman, and A. J. Richardson. Mean effective constants of cotton leaves. *Journal of the Optical Society of America*, 60:542–547, 1970.

[6] W. A. Allen, A. J. Richardson, and J. R. Thomas. Interaction of isotropic light light with a compact leaf. *Journal of the Optical Society of America*, 59:1376–1379, 1969.

[7] E. L. Allwein, R. E. Schapire, and Y. Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113–141, 2000.

[8] R. P. Armitage, M. Kent, and R. E. Weaver. Identification of the spectral characteristics of British semi-natural upland vegetation using direct ordination: a case study from Dartmoor, UK. *International Journal of Remote Sensing*, 24:3369–3388, 2004.

[9] G. Asrar. *Theory and Applications of Optical Remote Sensing*. John Wiley and Sons, 1989.

[10] H. Bach, W. Verhoef, and K. Schneider. Coupling remote sensing observation models and growth model for improved retrieval of (geo)-biophysical information from optical remote sensing data. In M. Owe, G. D'Urso, and E. Zilioli, editors, *Proceedings of SPIE, Remote Sensing for Agriculture, Ecosystems, and Hydrology II*, volume 4171, pages 1–11, Bellingham, WA, USA, 2000.

[11] M. F. Balcan, A. Blum, and K. Yang. *Advances in neural information processing systems*, chapter Co training and expansion: Towards bridging theory and practice. MIT press, Cambridge, MA, 2005.

[12] F. Baret. *Imaging Spectrometry in Agriculture, Comparison of Modellistic Approaches and Experimental Data*, page 20. Kluwer Academic, 1994.

[13] F. Baret and T. Fourty. Estimation of leaf water content and specific weight from reflectance and tranmittance measurements. *Agronomie*, 17:455–464, 1997.

[14] M. J. Barnsley and S. L. Barr. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric Engineering & Remote Sensing*, 62:949–958, 1996.

[15] M. J. Barnsley, J. J. Settle, M. Cutter, D. Lobb, and F F. Teston. The PROBA/CHRIS mission: a low-cost smallsat for hyperspectral, multi-angle, observations of the earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, 42:1512–1520, 2004.

[16] L. Bastin. Comparison of fuzzy c-means classification, linear mixture modelling and MLC probabilities as tools for unmixing coarse pixels. *International Journal of Remote Sensing*, 18:3629–3648, 1997.

[17] E. Ben-Dor, K. Patkin, A. Banin, and A. Karnieli. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data. a case study over soils in Israel. *International Journal of Remote Sensing*, 23:1043–1062, 2002.

[18] A. Berk, G. P. Anderson, P. K Acharya, J. H. Chetwynd, L. S. Bernstein, E. P. Shettle, M. W. Matthew, and S. M. Adler-Golden. MODTRAN4 user's manual. Technical report, Hanscom AFB: Air Force Research Laboratory, Space Vehicles Directorate, Air Force Material Command, MA 01731-3010, 2000.

[19] A. Berk, L. S. Bernstein, and D. C. Robertson. MODTRAN: A moderate resolution model for LOWTRAN7. Final report GL-TR-0122, AFGL, Hanscom AFB, MA, 1989.

[20] J. Biesemans and J. Everaerts. Image processing workflow for the PEGASUS HALE UAV payload. In *Proceedings of 2nd International Workshop "The Future of Remote Sensing"*. ISPRS Inter-Commission working group IV Autonomous Navigation, 2006.

[21] A. C. Bird, J. C. Taylor, and T. R. Brewer. Mapping national park landscape from ground, air and space. *International Journal of Remote Sensing*, 21:2719–2736, 2000.

[22] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon-Press, Oxford, 1995.

[23] T. M. Blackmer, J. S. Schepers, G. E. Varvel, and E. A. Walter-Shea. Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies. *Agron. J.*, 88:1–5, 1996.

[24] G. F. Bonham-Carter. Numerical procedures and computer program for fitting an inverted gaussian model to vegetation reflectance data. *Computers and Geosciences*, 14:339–356, 1988.

[25] C. C. Borel, S. A. W. Gerstl, and B. J. Powers. The radiosity method in optical remote sensing of structured 3d surfaces. *Remote Sensing of Environment*, 36:13–44, 1991.

[26] L. Breiman. Bagging predictors. *Machine learning*, 24:123–140, 1996.

[27] L. M. Bruce, C. H. Koger, and J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40:2331–2338, 2002.

[28] L. Bruzzone, M. Chi, and M. Marconcini. A novel transductive SVM for the semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 2006.

[29] C. Buschmann and E. Nagel. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*, 14:711–722, 1993.

[30] J. B. Campbell. *Introduction to remote sensing*. Taylor and Francis Ltd, 2 edition, 1996.

[31] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Rojo, and F. Melgani. Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters*, 3:339–343, 2006.

[32] F. Canters. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. *Photogrammetric Engineering & Remote Sensing*, 63:403–414, 1997.

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

[34] G. A. Carter and R. L. Miller. Early detection of plant stress by digital imaging with narrow stress-sensitive wavebands. *Remote Sensing of Environment*, 50:295–302, 1994.

[35] J. C-W Chan, C. Huang, and R. S. De Fries. Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting. *IEEE Transactions on Geoscience and Remote Sensing*, 39:693–695, 2001.

[36] C. I. Chang. *Hyperspectral Imaging, techniques for spectral detection and classification*. Kluwer Academic, New York, 2003.

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

[38] A. R. Clapham, T. G. Tutin, and D. M. Moore. *Flora of the British Isles*. Cambridge University, 1989.

[39] R. N. Clark. *Manual of Remote Sensing, Remote Sensing for the Earth Sciences*, volume 3, chapter Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy, pages 3–58. John Wiley and Sons, New York, 1999.

[40] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.

[41] M. F. Cohen and J. R. Wallace. *Radiosity and realistic image synthesis*. Academic Press Professional, 1993.

[42] W. Colins. Remote sensing of crop type and maturity. *Photogrammetric Engineering & Remote Sensing*, 44:43–55, 1978.

[43] R. G. Congalton. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, 37:35–46, 1991.

[44] R. G. Congalton and K. Green. *Assessing the accuracy of remotely sensed data: principles and practices*. Lewis Publishers, MI, 1999.

[45] A. M. Cordeiro, E. Alcantara, and D. Barranco. *Differences in tolerance to iron deficiency among olive cultivar*, pages 197–200. Kluwer Academic Publishers, Netherlands, 1995.

[46] A. P. Cracknell. Synergy in remote sensing-what's in a pixel? *International Journal of Remote Sensing*, 19:2025–2047, 1998.

[47] P. J. Curran. Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30:271–278, 1989.

[48] P. J. Curran, J. L. Dungan, B. A. Macler, S. E. Plummer, and D. L. Peterson. Reflectance spectroscopy of fresh whole leaves for the estimation of chemical concentration. *Remote Sensing of Environment*, pages 153–166, 1992.

[49] M. E. J. Cutler, P. M. Atkinson, and P. J. Curran. The effect of spatial resolution on the maximum detectable shift in the red edge position of plant reflectance. In *RSS'98 Developing International Connections, Remote Sensing Society, Nottingham*, volume RSS98, pages 511–517, 1998.

[50] I. Daubechies. Ten lectures on wavelets, second edition. In *CBMS-NSF regional conference series in applied mathematics 61.*, 1992.

[51] C. S. T. Daughtry, C. L. Walthall, M. S. Kim, E. Brown de Colstoun, and J. E. McMurtrey. Estimating corn leaf chlorophyll status from leaf and canopy reflectance. *Remote Sensing of Environment*, 74:229–239, 2000.

[52] T. P. Dawson, P. J. Curran, P. North, and S. E. Plummer. The propagation of foliar biochemical absorption features in forest canopy reflectance: a theoretical basis. *Remote Sensing of Environment*, 67:147–159, 1999.

[53] S. De Backer. *Unsupervised Pattern Recognition, Dimensionality Reduction and Classification*. PhD thesis, University of Antwerp, Antwerp, Belgium, 2002.

[54] S. M. de Jong and H. Th. Riezebos. *Conserving soil resources: European perspectives*, chapter Imaging spectroscopy, geostatistics and soil erosion modelling, pages 232–245. CAB Int., Oxon, UK, 1994.

[55] R. de Lange, M. van Til, and S. Dury. The use of hyperspectral data in coastal zone vegetation monitoring. In *EARSeL eProceedings 3 (2)*, pages 143–153, 2004.

[56] A. G. Dekker, R. J. Vos, and S. W. M. Peters. Comparison of remote sensing data, model results and in situ data for total suspended matter (TSM) in the southern Frisian lakes. *Science of the total environment*, 268:197–214, 2001.

[57] T. G. Dietterich and R. Bakiri. Solving multiclass learning problems using error correcting output codes. *Journal on Artificial Intelligence Research*, 2:263–286, 1995.

[58] D. J. Diner., J. C. Beckert, T. H. Reilly, C. J. Bruegge, J. E. Conel, R. Kahn, J. V. Martonchik, R. Davies T. P. Ackerman, S. A. W. Gerstl, H. R. Gordon, R. Myneni J-P. Muller, R. J. Sellers, B. Pinty, and M. M. Verstraete. Multi-angle imaging spectroradiometer (MISR) description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*, 39:1072–1087, 1998.

[59] M. I. Disney, P. Lewis, and P. North. Monte Carlo ray tracing in optical canopy reflectance modeling. *Remote Sensing Reviews*, 18:163–196, 2000.

[60] J. P. Doody. Information required for integrated coastal zone management: Conclusions from the European demonstration programme. *Coastal Management*, 31:163–173, 2003.

[61] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2 edition, 2001.

[62] J. R. Eastman and R. M. Laney. Bayesian soft classification for sub-pixel analysis: a critical evaluation. *Photogrammetric Engineering & Remote Sensing*, 68:1149–1154, 2002.

[63] H. Eerens. Course on remote sensing, 1994. Institute for land and water management, KU-Leuven.

[64] C. D. Elvidge. Visible and near infrared reflectance characteristics of dry plant materials. *International Journal of Remote Sensing*, 11:1775–1795, 1990.

[65] D. Esteban and C. Galand. Application of quadrature mirror filters to split-band voice coding schemes. In *Proceedings of the IEEE International Conference on Acoustics, Signal and Speech Processing*, pages 191–195, Hartford, USA, 1977.

[66] K. S. Fassnacht, W. B. Cohen, and T. A. Spies. Key issues in making and using satellite-based maps in ecology: a primer. *Forest Ecology and Management*, 222:167–181, 2006.

[67] P. Fisher. The pixel: a snare or a delusion. *International Journal of Remote Sensing*, 18:679–685, 1997.

[68] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[69] G. M. Foody. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17:1317–1340, 1996.

[70] G. M. Foody. Mapping land cover from remotely sensed data with a softened feedforward neural network classification. *Journal of Intelligent and Robotic Systems*, 29:433–449, 2000.

[71] G. M. Foody. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80:185–201, 2002.

[72] G. M. Foody. Remote sensing and digital image processing. In S. M. de Jong and F. D. van der Meer, editors, *Remote Sensing Image Analysis*, chapter 3, pages 37–49. Kluwer Academic Publishers, 2004.

[73] T. Fourty, F. Baret, S. Jacquemoud, G. Schmuck, and J. Verdebout. Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote Sensing of Environment*, 56:104–117, 1996.

[74] T. Fourty and f. Baret. On spectral estimates of fresh leaf biochemistry. *International Journal of Remote Sensing*, 19:1283–1297, 1998.

[75] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55:119–139, 1995.

[76] J. Friedman. Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University, 1996.

[77] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Boston, second edition, 1990.

[78] J. Furnkranz. Round robin classification. *Journal on Machine Learning Research*, 2:721–747, 2002.

[79] A. R. Jones G. B. Groom, R. M. Fuller. Contextual correction: techniques for improving land cover mapping from remotely sensed images. *International Journal of Remote Sensing*, 17:69–89, 1996.

[80] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.

[81] J. A. Gamon. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 41(1):35–44, 1992.

[82] D. M. Gates. Energy, plants and ecology. *Ecology*, 46:1–13, 1965.

[83] A. A. Gitelson and M. N. Merzlyak. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 148:494–500, 1996.

[84] F. Glover. Future paths for integer programming and links to artificial intelligence. *Computers and Operations Research*, 5:533–549, 1986.

[85] N. S. Goel. Models of vegetative canopy reflectance and their use in estimation of biophysical parameters from reflectance data. *Remote Sensing Reviews*, 4:1–212, 1988.

[86] N. S. Goel, I. Rozenhal, and R. L. Thompson. A computer graphics based model for scattering from objects of arbitrary shapes in the optical region. *Remote Sensing of Environment*, 36:73–104, 1991.

[87] N. S. Goel and D. Strebel. Inversion of vegetation canopy reflectance models for estimating agronomic variables i: Problem definition and initial results using the suits model. *Remote Sensing of Environment*, 13:487–507, 1983.

[88] N. S. Goel and R. L. Thompson. A snapshot of canopy reflectance models, and a universal model for the radiation regime. *Remote Sensing Reviews*, 18:197–225, 2000.

[89] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. Imaging spectroscopy for earth remote sensing. *Science*, 228:1147–1153, 1985.

[90] D. E. Goldberg. *Genetic Algorithms, in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[91] V. Gond, D. G. G. de Pury, F. Veroustraete, and R. Ceulemans. Seasonal variations in leaf area index, leaf chlorophyll, and water content; scaling-up to estimate fAPAR and carbon balance in a multilayer, multispecies temperate forest. *Tree Physiology*, 19:673–679, 1999.

[92] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. 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, 1988.

[93] Y. L. Grossman, S. L. Ustin, S. Jacquemoud, E. W. Sanderson, G. Schmuck, and J. Verdebout. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sensing of Environment*, 56:1–12, 1996.

[94] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81:416–426, 2002.

[95] J. H. Halton. A retrospective and prospective survey of the Monte Carlo method. *SIAM Reviews*, 12:1–63, 1970.

[96] R. M. Haralick and L. Shapiro. Survey: Image segmentation techniques. *Computer vision, graphics and image processing*, 29:100–132, 1985.

[97] T. Hastie and R. Tibshirani. Classification by pairwise coupling. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Proceedings of the 1997 conference on Advances in neural information processing systems*, volume 10, pages 507–513. The MIT Press, 1998.

[98] D. C. Hatchell. *ASD Technical Guide 3rd Edition*. ASD Inc., Boulder, 1999.

[99] W. Highleyman. The design and analysis of pattern recognition experiments. *Bell system Technical Journal*, 41:723–743, 1962.

[100] D. Hill, M. Fasham, G. Tucker, M. Shewry, and P. Shaw. *Handbook of biodiversity methods: survey, evaluation and monitoring*. Cambridge University Press, Cambridge, UK, 2005.

[101] M. Hill. TWINSPAN: a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. *Ecology and Systematics*, page 90, 1979.

[102] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:832–844, 1998.

[103] D. N. H. Horler, M. Dockray, and J. Barber. The red-edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4:273–288, 1983.

[104] G. F. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, IT-14, 1968.

[105] L. Ingber. Very fast simulated re-annealing. *Mathematical Computer Modelling*, 12:967–973, 1989.

[106] L. Ingber. Adaptive simulated annealing (ASA). Technical report, Caltech Alumni Association, 1993.

[107] S. Jacquemoud, C. Bacour, H. Poilvé, and J.-P Frangi. Comparison of four radiative transfer models to simulate plant canopies reflectance. *Remote Sensing of Environment*, 75:471–481, 2000.

[108] S. Jacquemoud and F. Baret. PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34:75–91, 1990.

[109] S. Jacquemoud and S. L. Ustin. Leaf optical properties: A state of the art. In *Proceedings of the 8th International Symposium of Physical Measurements & Signatures in Remote Sensing*, pages 223–232, 2001.

[110] S. Jacquemoud, S. L. Ustin, J. Verdebout, G. Schmuck, G. Andreoli, and B. Hosgood. Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment*, 56:194–202, 1996.

[111] S. Jacquemoud, W. Verhoef, F. Baret, P. J. Zarco-Tejada, G. P. Asner, C. Franois C, and S. L. Ustin. PROSPECT+SAIL: 15 years of use for land surface characterization. In *Proceedings of IEEE International Geoscience & Remote Sensing Symposium*, Denver, 2006.

[112] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, NJ, 1988.

[113] J. R. Jensen and F. Qiu. A neural network image interpretation system to extract rural and urban land use and land cover information from remote sensor data. *Geocarto International*, 16:1–10, 2001.

[114] G. H. John, R. Kohavi, and K. Pfleger. Irrelevant features and the subset selection problem. In M. Kaufmann, editor, *Proceedings of the eleventh international conference on machine learning*, pages 121–129, New Brunswick, NJ, 1994.

[115] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 2:241–254, 1967.

[116] I. T. Jollife. *Principal component analysis*. Springer Verlag, New York, 1986.

[117] A. Kaarna. *Multispectral Image Compression Using the Wavelet Transform*. PhD thesis, Lappeenranta University of Technology, Lappeenranta, Finland, 2000.

[118] A. Kaarna, P. Zemcik, H. Kälviäinen, and J. Parkkinen. Compression of multispectral remote sensing images using clustering and spectral reduction. *IEEE Transactions on Geoscience and Remote Sensing*, 38:1073–1082, 2000.

[119] J. Karlholm and I. Renhorn. Wavelength band selection method for multispectral target detection. *Applied Optics*, 41:6786–6795, 2002.

[120] M. Kent and P. Coker. *Vegetation description and analysis, a practical approach*. Belhaven, London, UK, 1992.

[121] S. Khorram, G. S. Biging, N. R. Chrisman, D. R. Colby, R. G. Congalton, J. E. Dobson, R. L. Ferguson, M. F. Goodchild, J. R. Jensen, and T. H. Mace. Accuracey assessment of remote sensing-derived change detection. *American Society for Photogrammetry and Remote Sensing Monograph Series*, page 64, 1999.

[122] K. E. Kim. Adaptive majority filtering for contextual classification of remote sensing data. *International Journal of Remote Sensing*, 17:1083–1087, 1996.

[123] M. S. Kim, C. S. T. Daughtry, E. W. Chappelle, J. E. Mcmurtrey III, and C. L. Walthall. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation. In *Proceedings of the 6th Symposium on Physical Measurements and Signatures in Remote Sensing*, pages 299–306, Val D'Isere, France, January 1994.

[124] D. Kimes, R. Nelson, M. Manry, and A. Fung. Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing*, 19:2639–2663, 1998.

[125] R. Kinderman and J. L. Snell. Markov random fields and their applications. *American Mathematics Society*, 1:1–142, 1980.

[126] J. T. O. Kirk. Characteristics of the light field in highly turbid waters: a Monte Carlo study. *Limnology and Oceanography*, 39:702–706, 1994.

[127] S. Kirkpatrick, C. D. Jr. Gerlatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[128] Y. Knyazikhin, J. Martonchik, R. Myneni, D. Diner, and S. Running. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research*, 103(D24):32,257–32,276, 1998.

[129] C. H. Koger, L. M. Bruce, D. R. Shaw, and K. N. Reddy. Wavelet analysis of hyperspectral reflectance data for detecting morningglory in soybean. *Remote Sensing of Environment*, 86:108–119, 2003.

[130] R. F. Kokaly, , and R. N. Clark. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise linear regression. *Remote Sensing of Environment*, 67:267–287, 1999.

[131] M. A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37:233–243, 1991.

[132] S. Kumar, J. Ghosh, and M. M. Crawford. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39:1368–1379, 2001.

[133] S. Kumar, J. Ghosh, and M. M. Crawford. Hierarchical fusion of multiple classifiers for hyperspectral data analysis. *International journal on pattern analysis applications*, 5:210–220, 2002.

[134] A. Kuusk. A multispectral canopy reflectance model. *Remote Sensing of Environment*, 50:75–82, 1994.

[135] A. Kuusk. A markov chain model of canopy reflectance. *Agricultural and Forest Meteorology*, 76:221–236, 1995.

[136] A. Kuusk. A two-layer canopy reflectance model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 71:1–9, 2001.

[137] P. Lachenbruch. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 23:639–645, 1967.

[138] D. A. Landgrebe. Hyperspectral image data analysis as a high dimensional signal processing problem. *IEEE signal processing magazine*, 19:17–28, 2002.

[139] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. Wiley, 2003.

[140] C. Lee and D. A. Landgrebe. Feature extraction based on decision boundaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:388–400, 1993.

[141] Z. P. Lee, K. L. Carder, R. F. Chen, and T. G. Peacock. Properties of the water column and bottom derived from airborne visible infrared imaging spectrometer (AVIRIS) data. *Journal of geophysical research*, 106:639–651, 2001.

[142] J. Lepš and P. Šmilauer. *Multivariate analysis of ecological data using CANOCO*. Cambridge University Press, 2003.

[143] H. G. Lewis and M. Brown. A generalised confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, 22:3223–3235, 2001.

[144] R. M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9:1082–1099, 1999.

[145] L. Li, S. L. Ustin, and M. Lay. Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. *International Journal of Remote Sensing*, 26:5193–5207, 2005.

[146] S. Z. Li. *Markov Random Field Modeling in Computer vision*. Springer-Verlag, Berlin, Germany, 1995.

[147] X. Li and A. H. Strahler. Geometric-optical modelling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing*, 23:705–721, 1985.

[148] X. Li and A. H. Strahler. Geometric-optical bidirectional reflectance modelling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, 30:276–292, 1992.

[149] X. Li, C. E. Woodcock, and R. Davis. A hybrid geometric optical and radiative transfer approach for modeling pyranometer measurements under a Jack Pine forest. *International Journal of Geographical Information Science*, 1:34–41, 1994.

[150] Shunlin Liang. *Quantitative Remote Sensing of Land Surfaces*. Wiley-Interscience, 2004.

[151] H. K. Lichtenthaler. Vegetation stress : an introduction to the stress concept in plants. *Journal of Plant Physiology*, 148:3–14, 1996.

[152] S. J. Maas and J. R. Dunlap. Reflectance, transmittance, and absorptance of light by normal, etiolated, and albino leaves. *Agronomy journal*, 81:105–110, 1989.

[153] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[154] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, USA, 1998.

[155] M. E. Martin and J. D. Aber. High spectral resolution remote sensing of forest canopy lignin, nitrogen and ecosystem process. *Ecological Applications*, pages 431–443, 1997.

[156] K. R. McCloy. *Resource management Information Systems: Remote Sensing, GIS and Modelling*. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2006.

[157] H. Mehner, M. Cutler, D. Fairbairn, and G. Thompson. Remote sensing of upland vegetation: the potential of high spatial resolution satellite sensors. *Global Ecology and Biogeography*, 13:359–369, 2004.

[158] F. Melgani and B. S. Serpico. A markov random field approach to spatio-temporal contextual information image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 2003.

[159] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of chemical physics*, 21:1087–1092, 1953.

[160] Y. Meyer. *Wavelets and Operators*. Cambridge University Press, New York, 1992.

[161] J. Miller, M. Berger, Y. Goulas, S. Jacquemoud, J. Louis, G. Mohammed, N. Moise, J. Moreno, I. Moya, R. Pedrós, W. Verhoef, and P. J. Zarco-Tejada. Development of a vegetation fluorescence canopy model. Technical report, ESA, Aril 2005. Final report.

[162] J. R. Miller, E. W. Hare, and J. Wu. Quantitative characterization of the vegetation red-edge reflectance 1. an inverted-gaussian reflectance model. *International Journal of Remote Sensing*, 11:1755–1773, 1990.

[163] E. Mohn, N. L. Hjort, and G. O. Storvik. A simulation study of some contextual classification methods for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 25:796–804, 1987.

[164] R. Müller, M. Lehner, P. Reinartz, M. Schroeder, and B. Vollmer. A program for direct georeferencing of airborne and spaceborne line scanner images. In *Proceedings of ISPRS, Commission I Mid-Term Symposium on Integrated Remote Sensing at the Global, Regional and Local Scale*, Denver, CO USA, 2002.

[165] R. Myneni, F. Hall, P. Sellers, and A. Marshak. The interpretation of spectral vegetation indices. *IEEE Transactions on Geoscience and Remote Sensing*, 33:481–486, 1995.

[166] R. B. Myneni, A. L. Marshak, and Y. Knyazikhin. Transport theory for a leaf canopy of finite-dimensional scattering centers. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 45:135–140, 1991.

[167] R. B. Myneni, J. Ross, and G. Asrar. A review on the theory of photon transport in leaf canopies. *Agricultural and Forest Meteorology*, 45:1–153, 1989.

[168] H. Nagendra. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22:2377–2400, 2001.

[169] P. M. Narendra and K. Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, C-31:917–922, 1977.

[170] P. M. Narendra and M. Goldberg. Image segmentation with directed trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:185–191, 1980.

[171] M. V. Ibá nez and A. Simó. Parameter estimation in markov random field image modeling with imperfect observations. a comparative study. *Pattern Recognition Letters*, 24:2377–2389, 2003.

[172] T. Nilson. A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology*, 8:25–38, 1971.

[173] T. Nilson and A. Kuusk. A reference model for the homogeneous plant canopy and its inversion. *Remote Sensing of Environment*, 27:157–167, 1989.

[174] P. North. Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Transactions on Geoscience and Remote Sensing*, 34:946–956, 1996.

[175] J. Pe nuelas, J. A. Gamon, A. L. Fredeen, J. Merino, and C. B. Field. Reflectance indices associated with physiological changes in nitogen- and water-limited sunflower leaves. *Remote Sensing of Environment*, 48:135–146, 1994.

[176] G. W. Paltridge and C. M. R. Platt. *Radiative Processes in Meteorology and Climatology*. Elsevier Scientific Pub Co., Amsterdam, The Netherlands, 1976.

[177] E. Parzen. On estimation of a probability density function and mode. *Annals of mathematical statistics*, 33:1065–1076, 1962.

[178] E. A. Patrick and F. P. Fischer. III, a generalized k-nearest neighbor rule. *Information and control*, 16:128–152, 1970.

[179] M. Pincus. A Monte Carlo method for the approximate solution of certain types of constrained optimization problems. *Operations Research*, pages 1225–1228, 1970.

[180] B. Pinty, N. Gobron, J. L. Widlowski, A. W. Gerstl, M. M. Versrtraete, M. I. Disney, P. Lewis, F. Gascon, J. P. Gastellu, L. Jiang, X. Li, L. Su, S. Tang, H. Wang, J. Wang, G. Yan, H. Zang, A. Kuusk, T. Nilson, W. Ni-Meister, P. North, W. Qin, R. Thompson, and W. Verhoef. Radiation transfer model intercomparison (RAMI) exercise. *Journal of Geophysical Research (Atmospheres)*, 106:11937–11956, 2001.

[181] R. G. Pontius. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering & Remote Sensing*, 66:1011–1016, 2000.

[182] A. Pragnère, F. Baret, M. Weiss, R. B. Myneni, Y. Knyazikhin, and L. B. Wang. Comparison of three radiative transfer model inversion techniques to estimate canopy biophysical variables from remote sensing data. In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'99*, Hamburg, Germany, 1999.

[183] W. H. Press, S. A. Teukolski, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C++*. Cambridge University Press, 2002.

[184] J. C. Price. On the information content of soil reflectance spectra. *Remote Sensing of Environment*, 33:113–121, 1990.

[185] J. C. Price. Band selection procedure for multispectral scanners. *Applied Optics*, 33:3281–3287, 1994.

[186] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Macmillan, New York, 1994.

[187] P. Pudil, J. Novovièová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.

[188] Yu Qian, Kang Zhang, and Fang Qiu. Spatial contextual noise removal for post classification smoothing of remotely sensed images. In *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, pages 524–528, New York, 2005. ACM Press.

[189] S. Raudys and R. P. W. Duin. Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. *Pattern Recognition Letters*, 19:385–392, 1998.

[190] J. A. Richards, D. A. Landgrebe, and P. H. Swain. Pixel labelling by supervised probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3:188–191, 1981.

[191] R. Richter. Atmospheric/topographic correction for airborne imagery. ATCOR-4 user guide version 3.1. Technical report, DLR, Wessling, Germany, 2004.

[192] R. Richter and D. Schläpfer. Geo-atmospheric processing of airborne imaging spectrometry data. part 2: Atmospheric/topographic correction. *International Journal of Remote Sensing*, 23(13):2631–2649, 2002.

[193] M. Riedmann and E. J. Milton. Supervised band selection for optimal use of data from airborne hyperspectral sensors. In *IEEE International Geoscience and Remote Sensing Symposium*, 2003.

[194] S. R. Safavian and D. A. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21:660–674, 1991.

[195] J. E. Sanger. Quantitative investigation of leaf pigments from their inception in buds through autumn coloration to decomposition in falling leaves. *Ecology*, 52:1075–1089, 1971.

[196] S. Sathyendranath, G. Cota, V. Stuart, H. Maas, and T. Platt. Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, 22:249–273, 2000.

[197] R. E. Schapire. Using output codes to boost multiclass learning problems. In *Machine learning: Proceedings of the forteenth international conference*, pages 313–321, San Mateo, CA, 1997.

[198] D. Schläpfer. *Differential absorption methodology for imaging spectroscopy of atmospheric water vapor*. PhD thesis, Remote Sensing Laboratories, Department of Geography, University of Zurich, Zurich, Switzerland, 1998.

[199] D. Schläpfer. PARGE: Parametric geocoding based on GCP-calibrated auxiliary data. In *Imaging Spectrometry IV, in Proceedings SPIE*, volume 3438, pages 334–344, 1998.

[200] M. Schlerf and C. Atzberger. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, 100:281–294, 2006.

[201] K. S. Schmidt, A. K. Skidmore, E. H. Kloosterman, H. van Oosten, L. Kumar, and J. A. M. Janssen. Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 70:703–715, 2004.

[202] S. Schmidtlein. Imaging spectroscopy as a tool for mapping ellenberg indicator values. *Journal of Applied Ecology*, 42:966–974, 2005.

[203] S. Schmidtlein and J. Sassin. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment*, 92:126–138, 2004.

[204] T. R. Sinclair, R. M. Hoffer, and M. M. Schreiber. Reflectance and internal structure of leaves from several crops during a growing season. *Agronomy Journal*, 63:864–868, 1965.

[205] J. A. Smith. LAI inversion using back propagation neural network trained with multiple scattering model. *IEEE Transactions on Geoscience and Remote Sensing*, 31:1102–1106, 1993.

[206] T. M. F. Smith. Comment on history and development of the theoretical foundations of survey based estimation and analysis. *Survey Methodology*, 16:26–29, 1990.

[207] S. V. Stehman. Thematic map accuracy assessment from the perspective of finite population sampling. *International Journal of Remote Sensing*, 66:589–593, 1995.

[208] S. V. Stehman. Basic probability sampling designs for thematic map accuracy assessment. *International Journal of Remote Sensing*, 20:2423–2441, 1999.

[209] S. V. Stehman and R. L. Czaplewski. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, 64:331–344, 1998.

[210] R. Storn and K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[211] A. Strahler, L. Boschetti, G. M. Foody, M. A. Friedl, M. C. Hansen, M. Herold, P. Mayaux, J. T. Morisette, S. V. Stehman, and C. E. Woodcock. Global land cover validation: recommendatoins for evaluation and accuracy assessment of global land cover maps. Technical report, Report of committee of Earth Observation Satellites (CEOS) -Working Group on Calibration and Validation, 2006.

[212] G. H. Suits. The calculation of directional reflectance of a vegetative canopy. *Remote Sensing of Environment*, 2:117–125, 1972.

[213] P. H. Swain. *Fundamentals of pattern recognition in remote sensing*, page 136=187. McGraw Hill, New York, USA, 1978.

[214] P. H. Swain and H. Hauska. The decision tree classifier design and potential. *IEEE Transactions on Geoscience Electronics*, GE-15:142–147, 1977.

[215] M. Tagliavini and A. D. Rombola. Iron deficiency and chlorosis in orchard and vineyard ecosystems. *European Journal of Agronomy*, 15:71–92, 2001.

[216] P. S. Thenkabail, R. B Smith, and E. De Pauw. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering & Remote Sensing*, 68:607–622, 2002.

[217] I. L. Thomas. Spatial postprocessing of spectrally classified Landsat data. *Photogrammetric Engineering & Remote Sensing*, 46:1201–1206, 1980.

[218] V. Thomas, P. Treitz, D. Jelinski, J. Miller, P. Lafleur, and J. H. McCaughey. Image classification of a northern peatland complex using spectral and plant community data. *Remote Sensing of Environment*, 84:83–99, 2002.

[219] F. E. Townsend. The enhancement of computer classifications by logical smoothing. *Photogrammetric Engineering & Remote Sensing*, 52:231–221, 1986.

[220] N. M. Trodd. Analysis and representation of heathland vegetation from near-ground level remotely-sensed data. *Global Ecology and Biogeography Letters*, 5:206–216, 1996.

[221] B. Tso and P. M. Mather. Classification of multisource remote sensing imagery using a genetic algorithm and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 37:1255–1260, 1999.

[222] B. Tso and R. C. Olsen. A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process. *Remote Sensing of Environment*, 97:127–136, 2005.

[223] Kagan Tumer and Joydeep Ghosh. Error correlation and error reduction in ensemble classifiers. *Connection Science*, 8:385–403, 1996.

[224] G. Turk. GT index: a measure of the success of prediction. *Remote Sensing of Environment*, 8:75–86, 1979.

[225] S. L. Ustin. *Manual of Remote Sensing: Volume 4 — remote Sensing for Natural Resource Management and Environmental Monitoring*. John Wiley & Sons, 2004.

[226] P. E. Utgoff. Incremental induction of decision trees. *Machine Learning*, 4:161–186, 1989.

[227] F. D. van der Meer and S. M. de Jong. *Imaging Spectrometry, Basic Principles and Prospective Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[228] F. Van der Meulen and H. A. Udo de Haes. Nature conservation and integrated coastal zone management in Europe: present and future. *Landscape and Urban planning*, pages 401–410, 1996.

[229] G. Vane, J. E. Duval, and J. B. Wellman. *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, chapter Imaging spectroscopy of the earth and other solar system bodies, pages 121–143. Cambridge University Press, Cambridge, 1993.

[230] V. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, New York, 1998.

[231] W. Verhoef. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, 16:125–141, 1984.

[232] W. Verhoef, 2006. Personal communication.

[233] W. Verhoef and H. Bach. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sensing of Environment*, 87:23–41, 2003.

[234] E. F. Vermote, D. Tanré, J. L. Deuzé, M. Herman, and J. J. Morcette. Second simulation of the satellite signal in the solar spectrum, 6s: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35:675–686, 1997.

[235] M. Verstraete, B. Pinty, and R. Myneni. Potential and limitations of inormation extraction on the terrestrial biosphere from satellite remote-sensing. *Remote Sensing of Environment*, 58:201–214, 1996.

[236] M. Vetterli and J. Kovačević. *Wavelets and Subband Coding*. Prentice-Hall, USA, 1995.

[237] F. J. Villalobos, F. Orgaz, and L. Mateos. Non-destructive measurement of leaf area in olive (olea europaea l.) trees using a gap inversion method. *Agricultural and Forest Meteorology*, 73(1):29–42, 1995.

[238] M. Weiss, F. Baret, R. B. Myneni, A. Pragnère, and Y. Knayazikhin. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20:3–22, 2000.

[239] C. A. Wessman, J. D. Aber, D. L. Peterson, and J. M. Melillo. Remote sensing of canopy chemistry and nitrogen cycling in temperate forest ecosystems. *Nature*, 335:154–156, 1988.

[240] D. J. Wiersma and D. A. Landgrebe. Analytical design of multispectral sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 18:180–189, 1980.

[241] L. Xu and D. Schuurmans. Unsupervised an semi supervised multi class support vector machines. In *AAAI G5, The Twentieth National Conference on Artificial Intelligence*, 2005.

[242] M. Zak and M. Cabido. Spatial patterns of the chaco vegetation of central Argentina: integration of remote sensing and phytosociology. *Applied Vegetation Science*, 5:213–226, 2002.

[243] P. J Zarco-Tejada, J. R. Miller, A. Morales, A. Berjón, and J. Agüera. Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. *Remote Sensing of Environment*, 90:463–476, 2004.

[244] P. J Zarco-Tejada, J. R. Miller, T. L. Noland, G. H. Mohammed, and P. H. Sampson. Scaling-up and model inversion methods with narrow band optical indices for chlorophyll content estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 39:1491–1507, 2001.

[245] P. J Zarco-Tejada, J. R. Miller G. H. Mohammed, T. L. Noland, and P. H. Sampson. Vegetation stress detection through chlorophyll a + b estimation and fluorescence effects on hyperspectral imagery. *Journal of Environmental Quality*, 31:1433–1441, 2002.

[246] P. J. Zarco-Tejada, J. C. Pushnik, S. Dobrowski, and S. L. Ustin. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment*, 31:1433–1441, 2002.

[247] H. Zhang and G. Sun. Feature selection using tabu search method. *Pattern Recognition*, 35:701–711, 2002.

Publication list of Pieter Kempeneers

[248] **P. Kempeneers**, S. De Backer, W. Debruyn, and P. Scheunders. Wavelet based feature extraction for hyperspectral vegetation monitoring. In L. Bruzzone, editor, *Proceedings of 10th SPIE Europe Symposium on Remote Sensing, Image and Signal Processing for Remote Sensing IX*, volume 5238, pages 297–305. 2003.

[249] S. Sterckx, P. Coppin, S. De Backer, W. Debruyn, **P. Kempeneers**, K. Meuleman, K. Nackaerts, I. Reusen, and P. Scheunders. Information extraction techniques for monitoring of stress symptoms in orchards. In *Proceedings of Earsel, Imaging Spectroscopy workshop*, pages 278–283. 2003.

[250] S. De Backer, **P. Kempeneers**, W. Debruyn, and P. Scheunders. Classification of dune vegetation from remotely sensed hyperspectral images. In *Proceedings of International Conference on Image Analysis and Recognition*, pages 497–503. 2004.

[251] S. De Backer, **P. Kempeneers**, W. Debruyn, and P. Scheunders. Wavelet based hyperspectral data analysis for vegetation stress classification. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, pages 387–391. 2004.

[252] **P. Kempeneers**, S. De Backer, S. Delalieux, S. Sterckx, W. Debruyn, P. Coppin, and P. Scheunders. Upscaling of spectroradiometer data for stress detection in orchards with remote sensing. In *Proceedings of 11th SPIE Europe Symposium on Remote Sensing, Remote Sensing for Agriculture, Ecosystems, and Hydrology VI*, volume 5568, pages 37–45. 2004.

[253] **P. Kempeneers**, S. De Backer, B. Deronde, L. Bertels, W. Debruyn, and P. Scheunders. Classifying hyperspectral airborne imagery for vegetation survey along coastlines. In *Proceedings of IEEE Geoscience and Remote Sensing Symposium*, volume 2, pages 1475–1478. 2004.

[254] S. De Backer, **P. Kempeneers**, W. Debruyn, and P. Scheunders. A band selection technique for spectral classification. *IEEE Geoscience and Remote Sensing Letters*, 2:319–323, 2005.

[255] **P. Kempeneers**, S. De Backer, W. Debruyn, and P. Scheunders. Generic wavelet-based hyperspectral classification applied to vegetation stress detection. *IEEE Transactions on Geoscience and Remote Sensing*, 43:610–614, March 2005.

[256] **P. Kempeneers**, S. De Backer, S. Provoost, W. Debruyn, and P. Scheunders. Hyperspectral classification applied to the Belgian coastline. In L. Bruzzone, editor, *Proceedings of 12th SPIE Europe Symposium on Remote Sensing*, volume 5982, pages 144–152. 2005.

[257] **P. Kempeneers**, Y. Park, S. De Backer, S. Sterckx, W. Debruyn, K. Ruddick, and P. Scheunders. Retrieval of oceanic constituents from ocean color using simulated annealing. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Seoul, Korea*, volume 8, pages 5651–5654. 2005.

[258] **P. Kempeneers**, S. De Backer, P. J. Zarco-Tejada, S. Delalieux, G. Sepulcre-Cantó, F. Morales, R. Sagardoy, J. van Aardt, P. Coppin, and P. Scheunders. Chlorophyll retrieval from canopy reflectance over orchards using hyperspectral techniques. In J. A. Sobrino, editor, *Proceedings of Second Symposium on Recent Advances in Quantitative Remote Sensing*. Servicio de Publicaciones. Universitat de Valencia, Valencia, 2006.

[259] **P. Kempeneers**, S. De Backer, P. J. Zarco-Tejada, S. Delalieux, G. Sepulere-Cantó, F. Morales-Iribas, J. V. Aardt, P. Coppin, and P. Scheunders. Stress detection in orchards with hyperspectral remote sensing data. In M. Owe, G. D'Urso, C. M. Neale, and B. T. Gouweleeuw, editors, *Proceedings of 13th SPIE Europe Symposium on Remote Sensing*. 2006.

[260] **P. Kempeneers**, S. Provoost, S. De Backer, W. Van Gompel, B. Deronde, and P. Scheunders. Surveillance of dune vegetation along the Belgian coast using hyperspectral remote sensing. *International Journal of Remote Sensing*, in review, 2007.

[261] **P. Kempeneers**, P. J. Zarco-Tejada, P. North, S. De Backer, S. Delalieux, G. Sepulcre-Cantó, F. Morales, J. A. N. van Aardt, R. Sagardoy, P. Coppin, and P. Scheunders. Model inversion and viewing conditions for chlorophyll retrieval from canopy reflectance over orchards. *International Journal of Remote Sensing*, accepted for publication, 2007.