The impact of the spectral dimension of hyperspectral datasets on plant disease detection

Dissertation

zur

Erlangung des Doktorgrades (Dr.rer.nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universiät Bonn

vorgelegt von

Dipl. Geogr. Thorsten Mewes

aus Einbeck

Bonn, 2010

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter
Prof. Dr. Gunter Menz
Department of Geography
Remote Sensing Research Group (RSRG)
University of Bonn

2. Gutachter
Prof. Dr. Sebastian Schmidtlein
Department of Geography
Vegetation Geography
University of Bonn

Tag der mündlichen Prüfung: 18.03.2011
Erscheinungsjahr: 2011

Acknowledgments

This thesis is originated at the Center for Remote Sensing of Land Surfaces (ZFL) and the Remote Sensing Research Group (RSRG) of the Rheinische Friedrich-Wilhelms-University of Bonn. All studies were carried out under sponsorship of the Research Training Group 722 'Information Techniques for Precision Crop Protection', which is funded by the German Research Foundation (DFG).

I would like to thank everyone who has taken part in the successful completion of this work. A lot of people ranging from family, friends and colleagues at the ZFL to colleagues abroad helped me a lot throughout this time with numerous discussions, co-operations, thoughts or just warm words. An entire list would exceed the limits of the acknowledgments, but some special words to some special people have to be spoken.

Warm thanks go to the people at the experimental farm 'Klein Altendorf'. Two different experiments were successfully carried out in cooperation with Mr. Markus Huober, who helped me a lot in planning and realization of the field experiments. We always could easily find a common denominator regarding cultivation and treatments. Thanks for that.

Two greenhouse experiments were carried out in collaboration with people from the Institute for Phytomedicine at the University of Bonn. Rooms, materials and know-how were provided in an unproblematic and communicative way. My special thanks thereby go to PD Dr. E.-C. Oerke and HD Dr. Ulrike Steiner.

All experiments would not have been run successfully without the help of my colleagues from the Research Training Group 722. Whether helping in the field on hot summer days, preparing and milling soil for greenhouse trials in cold winter days or discussions about theoretical or technical problems were anyway problematic. I really would like to thank my colleagues, especially Anne-Kathrin Mahlein and Christian Hillnhütter.

I would like to show my gratitude to my former and present colleagues at the ZFL for their help and appreciation throughout the three years. I really enjoyed the cooperations, the exchange of ideas, the teamwork, numerous discussions at coffee breaks and especially the meetings off work. My special thanks thereby go to Dr. Jan Jacobi for his professional know-how in agricultural questions and Ellen Götz for her great assistance at any time.

I would like to thank my supervisor Prof. Dr. Gunter Menz for his guidance and

encouragement. He supported me in a number of ways and I could always count on professional guidance, especially regarding the three month of airborne hyperspectral training in the US. I would like to thank Michael Frank from Galileo Group for the great introduction into airborne data acquisition and collaboration in hyperspectral remote sensing. I appreciate that Prof. Dr. Sebastian Schmidtlein joined the dissertation committee.

Special thanks go to my friend and colleague Dr. Jonas Franke. He showed me how to survive at the university. I hope we can still work together in future projects, even if it is not the Sunset Boulevard. Thank Jonas, Friederike and Knut for checking through the manuscript. You really helped me a lot.

Warm thanks go to my beloved family, especially to my parents. Without their belief and backup in an unselfish way I would have never come this far. First and foremost, this thesis would not have been possible without Friederike and her incredible amount of patience with me in the last three years. Thank you so much.

Contents

Abstract

Precision Agriculture as an information based approach requires explicit spatial information about the within field heterogeneities for site-specific applications. Thus, the usage of cost-intensive agrochemicals and the impact on the environment can be significantly reduced. Spectroscopic approaches are thereby a promising tool for providing fast and precise information on a local to regional level. In this thesis, the impact of hyperspectral near-range and remote sensing data for crop stress detection will be observed since spectroscopic approaches are of great interest for Precision Agriculture. Two greenhouse experiments and three field experiments were conducted with spectroscopic measurements to examine possibilities and limitations of hyperspectral data. The data were acquired using a near-range non-imaging spectrometer (ASD Fieldspec 3) and a near-range imaging spectrometer (ImSpec V10E) in the greenhouse, or were acquired by the airborne sensor systems HyMap™, ROSIS or AISA for the field experiments. The methodical foci thereby are the improvement of binary detection approaches, discriminating 'vital' and 'infected' wheat stands or parts of wheat stands, and quantification approaches to estimate disease severities at canopy level.

This thesis examines the spectral dimension of hyperspectral data for crop stress detection by assessing data redundancy and defining spectral necessities. Different feature selection methods were tested for their suitability in reducing the high amount of spectral data without losing significant information. Conventional classification approaches and recent developments, such as support vector machines for classification (SVM), were thereby tested based on the identified spectral subsets to assess the status of different wheat stands. By focusing on phenomenon-specific spectral bands, stressed wheat stands could successfully be identified with high accuracies. Using optimal band combinations could even increase classification accuracies. The results showed that not the entire spectrum of hyperspectral data is necessary for the detection of fungal infections in wheat. These findings are particularly interesting for future spectral sensor design and remote sensing missions that are aiming at the provision of spatial information for agricultural practice.

The ability of hyperspectral data in quantifying the severity of fungal diseases was observed. Site-specific fungicide treatments based on application maps are technically possible and doses can be adjusted if the maps provide information about the health status of the crops. Crop growth anomalies caused by fungal infections were observed, which differed significantly within one field. The derivation of disease severities based on hyperspectral near-range and remote sensing data were examined using classification approaches and support vector machines for regression (SVR). Fungal infections of

wheat stands in the greenhouse and wheat stands in the field could be quantified with a high level of certainty. The results are very promising and the findings may be of great interest for agricultural questionnaires and automatic phenotyping approaches, since the presented approaches are fast and non-destructive. Spatial maps with continual disease severity data could be derived, which can be used to generate application maps for agricultural practice.

Since the study shows that a reduction of hyperspectral data to a few but specifically selected spectral bands can improve the classification accuracies or regression analyses, a preliminary feature selection should be considered when working with hyperspectral remote sensing data. Agricultural and geographical approaches that are based on spatial-spectral information may thus profit from a faster and more reliable extraction of information. The study shows great advantages of the usage of hyperspectral imaging data but also the necessity of advanced and innovative analyzing methods.

List of Figures

List of Tables

1. Introduction and objectives

1.1. General concept and outline

The development of hyperspectral sensor systems, whether near-range, airborne or spaceborn, and specific analyzing techniques is a fast-growing topic within the remote sensing community. New analyzing possibilities provided by hyperspectral remote sensing data with its detailed spectral information are of great interest for many applications. Spectroscopy data showed to be useful for many different questionnaires for example in geology, atmospherical studies, medicine, biology, coastal studies or agriculture, and the demand for such data is still growing. This thesis examines and discusses the use of hyperspectral remote sensing data within the framework of precision agriculture. Possibilities, limitations and requirements of recent and future hyperspectral sensor systems for the detection and quantification of diseased wheat stands will thereby be demonstrated and critically observed. The first chapter 1 *Introduction and objectives* gives a short description of hyperspectral remote sensing. The concept of Precision Agriculture will be presented with a focus on the sensing of crop stress. Spectroscopy data have been shown to be suitable for the detection of inhomogeneities caused by different vitality of plants. Remote sensing is a promising tool for information extraction over large spatial areas and may support agricultural practice in decision making. The spectral properties of crops will be highlighted and the influence of stress on the spectral behavior as a prerequisite for optical detection will be shown. This chapter concludes with an outline of the thesis, summarizing their intentions. The single studies discussed in this thesis from chapter 3 to chapter 6 examine the use of hyperspectral near-range and remote sensing data for crop stress detection. Benefits, but also limitations are discussed regarding relevant and redundant spectral information and the minimal needed spectral resolution. Thus, the thesis may deliver a fundamental contribution to future sensor developments specified on the use in the context of Precision Agriculture and crop stress detection.

1.2. Hyperspectral Remote Sensing

When talking about hyperspectral data, the term spectroscopy is inevitable. Hyperspectral remote sensing usually refers to image data, whereas spectroscopy data usually refers to point measurements. But, since the term 'imaging spectroscopy' is commonly used within the remote sensing community, their differences or rather their similarities need to be clarified. Field and laboratory spectrometers measure the proportion of solar or artificial light as reflected by an observed surface at many narrow, contiguous spec-

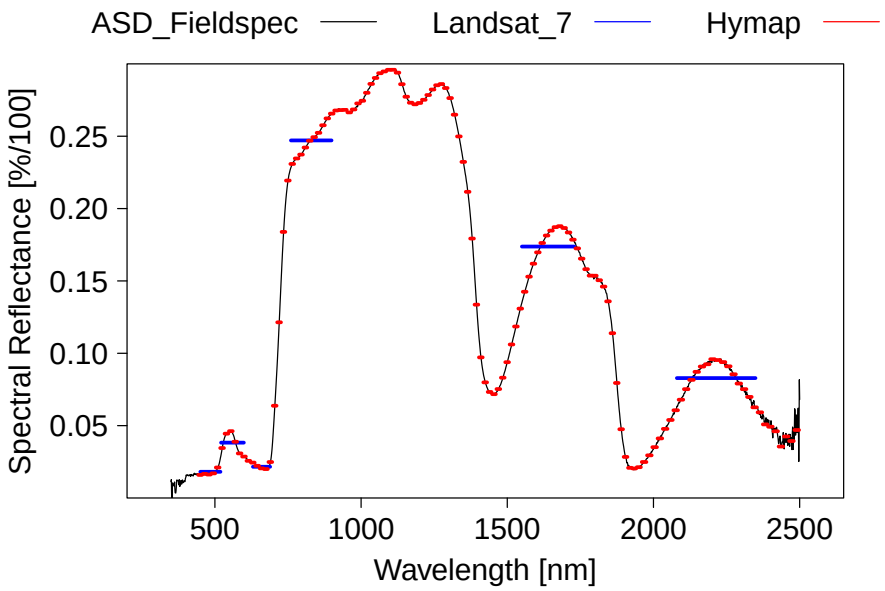


Figure 1.1.: Comparison of the spectral resolution of three different sensor systems on the basis of a typical vegetation spectrum

tral bands. Thus, the resulting spectra appear to be continuous curves over the entire sensor specific spectral range. When a spectrometer is used in an imaging system, the resulting images record a reflectance spectrum for each pixel in the image. In contrast to hyperspectral sensors, most multispectral systems (e.g., Landsat, Quickbird) measure the reflected radiation from a surface at a few broad and separated spectral bands. For example, a typical vegetation spectrum of a vital plant is displayed in figure 1.1. Three different sensors are compared, i.e. ASD Fieldspec® (as a spectrometer for laboratory or ground measurements with a very high spectral resolution), HyMap™ (an airborne hyperspectral imager with lower spectral resolution but contiguousness) and Landsat 7 ETM+ (a commonly used multispectral sensor with a few broad and separated spectral bands).

The underlying spectral response curve of the field spectrometer in black displays the spectrum between 450nm and 2500nm in a continual manner. Thus, all spectral features, like the high absorption in the blue (480nm) and red (660nm) range within the visible part of the electromagnetic spectrum and the relative maximum in the green range (550nm) are captured in detail. The blue lines in figure 1.1 represent the spectral resolution of the Landsat 7 sensor. The multispectral sensor also measures the spectral reflectance in the blue, green and red, but only with one band respectively. Some information, like the position of the relative maximum within the green range or the

inflection point between the red and the NIR plateau can not be located. Only a few spectral information within higher wavelength areas are captured by Landsat, as there is no continual scan of the surface reflectance. The red dots in the figure represent the spectral resolution of HyMap™. The spectral response of the vegetation surface is captured in a nearly continual way, but, the spectral information is not as detailed as the spectrum obtained by the ASD Fieldspec®. It is not the number of wavelength bands that define a sensor as hyperspectral, it is rather the narrowness and contiguousness.
Spectroscopy is based on the interaction of electromagnetic radiation with matter [Green et al., 1998]. If solar or artificial light hits any material, the radiance is modified in some ways. Parts of the radiance will be absorbed, parts will be transmitted and parts will be reflected. The proportion of the reflected radiance strongly depends on the matter and is wavelength dependent. Spectroscopy analyses these interactions in a quantitative and qualitative way, by capturing the reflected radiance of a matter and analyzing specific features or differences to known spectra. Thus, spectroscopy is used in many different subjects like chemistry, pharmacy or soils sciences. These measurements are usually point measurements, so the spectral response of a matter is captured in one value for every wavelength or wavelength range. If spectra are measured as images, every Pixel can be seen as a single spectroscopic measurement with an adherent spatial information, i.e. imaging spectroscopy. But, the term imaging spectroscopy is interchangeable with the term hyperspectral imaging, since both mean spectroscopic measurements on a pixel by pixel basis. Thus, both terms are treated equally for this thesis. The hyperspectral sensors used for this study will be discussed in chapter 2 and its complex preprocessing chain which is very crucial for the data quality will thereby be highlighted.

1.3. Plant - Signal interactions

As this study focuses on the spectral requirements of hyperspectral remote sensing data for the detection of crop stress in wheat plants, a short introduction into the meaning of remote sensing data for agricultural use is given. This section outlines the concept of Precision Agriculture, which is an information-based agricultural approach, with a strong focus on sustainability. After a definition of this concept and of crop stress in general, past and recent studies about the use of spectroscopy data for crop stress detection will be shown.

1.3.1. Precision Agriculture

In many EU member states, the agriculturally used area clearly decreased in the past few years, while the yields remained stable or could even be increased (European Commission 2008). This was possible due to the cultivation of high-yielding varieties, and the generally intensive usage of agrochemicals like herbicides and fungicides, whose uniform application over the entire field with significant environmental impact is still the current practice. At the same time, however, the concept of sustainability has become more and more important and economic, ecological and social pressure force farmers to change traditional practices, and to reduce their environmental impacts. Precision Agriculture as an information-based approach for crop production aims at a more sustainable management with constant or increasing yields, while reducing the applications of cost-intensive and polluting agrochemicals [Bongiovanni and Lowenberg-Deboer, 2004]. Wheat is one of the most cultivated crops in the European Union (21,935,000 ha in 2007) and is of high relevance for food security worldwide (European Commission 2008). In the 1990s, yields of wheat could constantly be increased while the agricultural area in use was decreased. This resulted in an intensive usage of farmland with systematic applications of fertilizers, fungicides and herbicides. Yet, agricultural treatments like the application of pesticides can be realized in a site-specific way if spatially explicit information on the within-field heterogeneities is given. Geographic Information Systems (GIS), Global Positioning System (GPS), and near-range or remote sensing systems may be used for information extraction and a more precise cultivation [Kühbauch and Hawlitschka, 2003, Pinter et al., 2003].

Today, the uniform usage of pesticides over the entire field is still the current practice. Thus for example plant without any fungal infections are treated with fungicides resulting in negative ecological and economical drawbacks. But, the costs of a uniform fungicide treatment are usually less than the financial loss due to decreasing yields and quality caused by fungal infections. Site-specific applications for example of nutrients is possible using recent near-range sensing technologies (e.g. Yara N-Sensor, Agri Con GmbH, Ostrau, Germany). A spectral sensor is therefore mounted at the top of a tractor and while crossing the field it permanently measures the spectral reflectance of plants within defined spectral ranges. The data is then processed and the nitrogen status of the plants is rated. After translating the data into application recommendation the fertilizes can be applied site-specifically. To realize a site-specific application of fungicides the farmer needs to observe his field visually which is cost and time intensive since no sensing system is available. The use of near-range and remote sensing data to visualize

spatial heterogeneities in the field caused by fungal infections would be a great benefit. Several works within the framework of this research training group 722 have shown that a detection of diseased crops is possible using spectral data [Jacobi, 2005, Voss, 2005, Franke, 2007]. But, the derivation of application maps is still a great challenge. The use of near-range and airborne hyperspectral data may be of great profit for this task.

1.3.2. Crop stress

Precision Agriculture can be seen from many different viewing angles. Fro example, the development of new and more precise nozzles, more effective fungicides, enhanced mechanical weeding techniques take part in the framework of Precision Agriculture. Another major objective, and this is the focus of this thesis, is the detection and the monitoring of spatial and temporal differences of plant vitality. Besides phenological issues, main causes of those differences are nutrient deficiencies and plant stress. There are several causes of plant stress. Jackson [1986] stated that all biotic and abiotic factors that adversely influence crop growth, can be defined as 'stress'. These factors may result in physiological and anatomical changes of plants. Stress factors can be categorized into different classes:

- Water stress - Water deficits cause a reduction of plants evapotranspiration, resulting in increased temperatures which can be analyzed by thermographic approaches [Carter, 1991].
- Salinity stress - Salts in soil and irrigation water limit productivity. Identification of contaminated soils, which usually have high surface reflectances [Pinter et al., 2003].
- Disease-caused stress - Bacterial, viral or fungal infections mainly reduce chlorophyll concentration of the host plant causing spectral changes in the visible and near-infrared spectrum [Carter and Knapp, 2001].
- Weeds - Reducing crop yield and quality by competing with crops for water, nutrients and light [Pinter et al., 2003].
- Nutrient deficiency - Reflectance changes in the visible spectrum caused by changed chlorophyll contents [Blackmer et al., 1995].

This thesis focuses on the detection of cropstress caused by fungal infections, i.e. disease-caused stress. There are several preconditions required for the occurrence of

a fungal infection. First, the host plant must be generally vulnerable to a pathogen, e.g. *Blumeria graminis* causing powdery mildew or *Puccinia recondita* causing leaf rust. Secondly, the phenological phase of the host plant and the environmental conditions are convenient and thirdly, the appropriate dispersal mechanisms like wind, water or animals [Börner et al., 1997] are given. If the conditions are suitable for a fungal infection, an infection of the leaves may occur. At the first stage, the pathogen is placed for example on the leave surface, which is called inoculation. At the second stage, the initial infection stage, the pathogen infiltrates into the host plant through the stomata or by using enzymes, toxins or mechanical processes [Hoffmann and Schmutterer, 1999]. Depending on the pathogen, a potential germination and sporulation is given and the pathogen infection spreads. Symptoms caused by a pathogen infection are manifold, from visually detectable symptoms like yellowness to metabolic changes. Fungal diseases can cause defoliation of plants or even the death of the entire host plant. Depending on the cultivar and on the pathogen, significant yield losses up to 30 to 50% are thus possible [Rodemann, 2009]. The focus of this thesis is the detection of the fungal diseases powdery mildew and leaf rust in wheat. The two pathogens, i.e. *Blumeria graminis* causing the disease powdery mildew and *Puccinia recondita* causing leaf rust, are among the most common in the EU.

Blumeria graminis *Blumeria graminis f. sp. tritici* is a very common pathogen and widely spread over the world. It can affect a wide range of plants, e.g. wheat, barley and rye. The pathogen belongs to the ascomycota or sac fungi, a phylum of the kingdom fungi. In wheat it can cause the plant disease powdery mildew which may cause defoliation of plants or even the death of the host plant with yield losses up to 20% [Carver, 2001, Rodemann, 2009]. The symptoms are, as exemplarily displayed in figure 1.2, powdery white or bright gray spots on the leaves and stems, i.e. mycelia. First mycelia usually appear on the upper side of leaves and the lower leaves are the most affected, but, any above-ground part of the plant can be infected by the disease. The mycelia are an aggregation of several conidia for asexual dispersal and haustoria that penetrate the tissue of the host plant and supply the fungus with necessary nutrients. *Blumeria graminis* has polycyclic characteristics, which means that it repeatedly infects plants during the growing season by wind dispersal of mature conidia, i.e. sporulation [Carver, 2001]. At the end of a pathogen cycle, cleistothecia are generated within the mycelia. A cleistothecium is a special form of an ascocarp or fruit body which is generated by some ascomycota. On hot summer days the pathogen outlast as ascospores

within the black bodies of the cleistothecia. In winter time the pathogen outlasts as cleistothecia on yield residua or as mycelia on winter grain. Under climatic conditions between 5 and 30 degrees Celsius and a relative humidity about 95% the ascospores germinate on new host plants and generate new mycelia. If temperatures are too high, no mycelia are generated but chlorosis and necrosis may occur on the host plants. Wheat is particularly susceptible for an infection with *Blumeria graminis* between the growth stages at the end of tillering until milk development [Rodemann, 2009]. The meaning of powdery mildew increased with the intensification of crop production. In Germany the common practice to avoid infections by this pathogen are the usage of cultivars with a high resistance and the treatment with fungicides. Each pathogen has typical spatiotemporal characteristics which have to be considered when working with remote sesnig data for crop stress detection. These facors have been studied in detail during the second phase of the research training group 722 (see Franke [2007], Franke et al. [2009a]) and are discussed in the context of remote sensing in Voss et al. [2010]. However, both pathogens have polycylcic characteristics which have an influence on the sensoral detectability of diseased wheat stands.

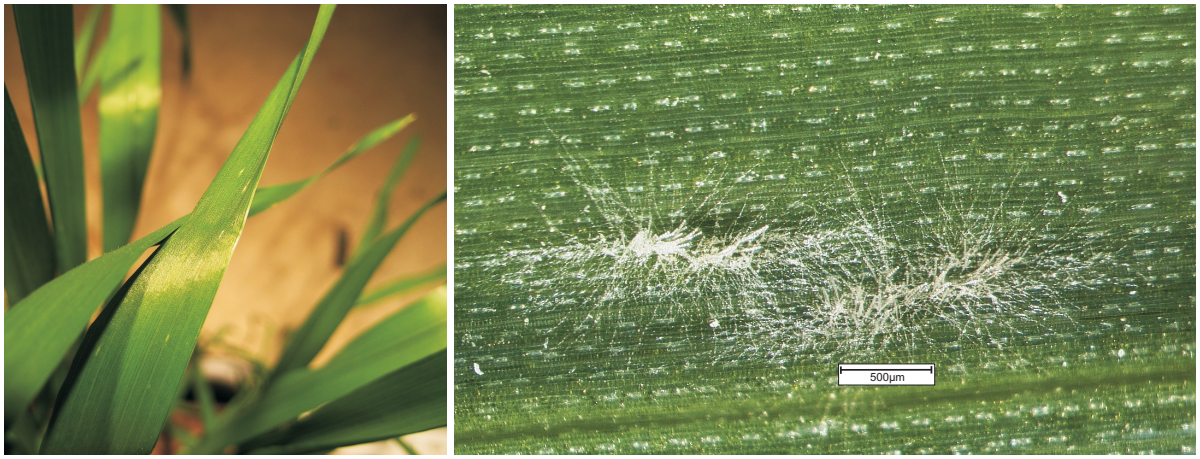


Figure 1.2.: Photograph of a minor infected wheat leaf (left) and a stereoscopic photo of a mycel from the same leaf (right).

Puccinia recondita f.sp. tritici *Puccinia recondita* is another very common pathogen in countries where wheat is cultivated. The fungus can infect wheat, weeds, rye and triticale and may cause the disease leaf rust. Typical symptoms are yellow-orange to reddish-brown pustules on the upper sides of leaves, on the leaf sheath, ear and awn, but, infection can occur on any above-ground plant part [Marsalis and Goldberg, 2006,

Rodemann, 2009]. The approximately 2mm large pustules, as exemplarily displayed in figure 1.3, produce a large number of spores which may infect other parts of the leave or other plants by wind dispersal. The pathogen belongs to the basidiomycota, a phylum of the kingdom fungi. Leaf rust on wheat is worldwide economically one of the most critical fungal infections. Leaf rust can cause defoliation of plants or even the death of the host plant, resulting in considerable yield losses in quantity and quality [Strange, 2003]. Yield losses up to 30% can arise and quality can be significantly affected in a negative way.

Puccinia recondita needs living host plants all year around. It outlasts the winter period as ascospores on alternate plants like weed or as uredospores on yield residua. The uredospores germinate on wheat and a hypha penetrates into the stomata. Mycelia are generated between the plant cells and haustoria supply the fungus with nutrients. New uredospores are generated within the mycelia that can be dispersed by wind and infect other leaves or plants. A potential infection with leaf rust need warm and humid weather conditions, i.e. temperatures between 15 and 20 degrees celcius and 100% relative humidity, for sporulation and germination [Rodemann, 2009]. Wheat is particularly susceptible for an infection with *Puccinia recondita* within the growth stages of heading and flowering. The meaning of leaf rust increased with the intensification of crop production. The common practice to avoid an infection is the usage of fungicides and cultivars with high resistance.

Figure 1.3.: Leaf (left) and stereoscopic photo (right) of wheat infected with leaf rust.

1.3.3. Sensing of crops and cropstress

The reflectance or rather the spectral characteristics of vegetation using near-range spectroscopy or remote sensing devices have been studied since the early 1960s. First examinations were made in the laboratory using spectroradiometers to document the reflectance of plant leaves or rather of plants at leaf level. These experiments were carried out in controlled environments using artificial light as radiation source at a defined illumination angle. Thus, even at this early stage of spectral analysis, important spectral phenomena could be documented and spectral differences could be identified [Gates et al., 1965, Knipling, 1970]. The typical absorption and reflectance features of a vegetation spectrum (see figure 1.4) and its causes were described for example in Knipling [1970] or Curran [1989]).

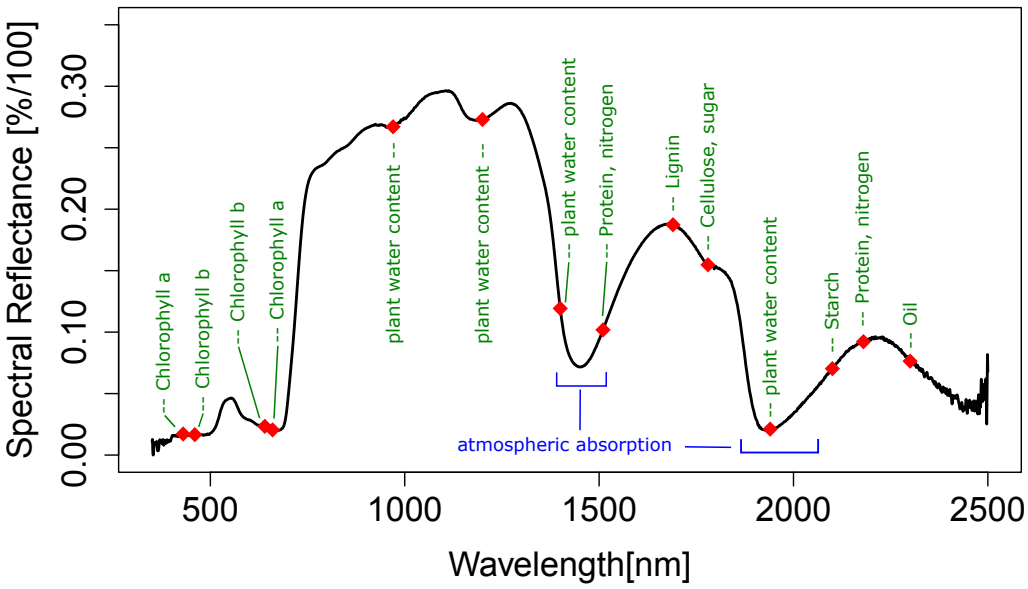


Figure 1.4.: Main absorption features in the VNIR range of a plant reflectance spectrum, feature positions based on Curran [1989].

Gathered knowledge, like the high spectral absorption in the wavelength range of the visible red caused by pigments and the comparably high reflectance in the near-infrared range, led to the first development of vegetation indices like NDVI (normalized difference vegetation index) by Rouse et al. [1973]. Common vegetation indices make profit from

those typical spectral feature by calculating ratios between local maximum reflectance and local minim reflectance. Many relationships between index values and ground truth parameters have been found since then. Also studies examining the direct relationship between single wavelength values and biochemical parameters or foliar chemistry were conducted successfully and many dependencies could be identified. For example, the concentration of chlorophyll-a modifies the reflectance at 430nm and 660nm, cellulose content at 1780nm and Starch at 2100nm (see figure 1.4 [Curran, 1989]). It has been demonstrated that several organic compounds, e.g. chlorophyll or starch, can be estimated by reflectance measurements [Yoder and Pettigrew-Crosby, 1995].

In the 1980s first examinations were made on the influence of stress on the spectral reflectance of plants. Stress factors of biotic and abiotic nature were thereby observed at leaf level [Thomas and Gausman, 1977, Jackson, 1986, Lorenzen and Jensen, 1989, Gitelson and Merzlyak, 1994]. Besides the spectral properties of soil and plants at canopy level were analyzed. Jackson [1986] stated that a detection of stress is possible using near-range spectroscopy and remote sensing data since the spectral reflectance is useful to assess the effect of stress on plant biomass involving architectural, physiological and anatomical changes. Good results could be achieved at leaf level using spectroradiometer data for the detection of fungal infections in barley [Lorenzen and Jensen, 1989]. Several studies using spectroscopy data for the detection of plant stress at canopy level were conducted in the 1990s. Blackmer et al. [1995] could successfully detect nitrogen deficiencies in corn using an optical sensor system. Carter [1994] used several narrow spectral bands of an imaging system to identify stress in soybean at canopy level. In the late 1990s and at the beginning of the new century some studies focused on early detection of plant stress using spectroscopy data. Bravo et al. [2003] successfully detected yellow-rust infestations on winter wheat at canopy level using an imaging spectrometer in the field. Carter [1994] examined the use of hyperspectral imaging for an early detection of nutrient deficiencies in soybeans. With the ongoing development of multispectral and hyperspectral sensor systems at near-range, airborne and space level, more and more studies focus on the use of remote sensing data for the detection of crop stress, e.g. Apan et al. [2004], Shaw and Kelley [2005], Franke and Menz [2007], proving their usage for applications in the context of Precision Agriculture.

A major goal of Precision Agriculture is the reduction of cost-intensive agrochemicals and thus decreasing the impact on the environment. Site-specific applications of

pesticides and fertilizers can be realized if spatially explicit information is given [Pinter et al., 2003]. At this point remote sensing applications can play a major role by delivering relevant information for decision support [Kühbauch and Hawlitschka, 2003]. Remote sensing data may support agricultural practice in identifying the right time and the right place for the right application. The detection of nitrogen deficiencies in cereal stands or site-specific herbicide applications have been studied in detail and are already partially integrated in practice, for example the N-Sensor® for fertilizer adjustment [Jain et al., 2007, West et al., 2003]. In contrast less attention has been paid to the detection of diseases via near-range or remote sensing data in the past. West et al. [2003] provided an overview of developments in the use of sensor-based detection of stress caused by diseases. Moshou et al. [2004] and Bravo et al. [2003] showed the possibilities for an automatic detection of yellow rust using near-range sensors. The potential of using airborne or satellite-based multi- or hyperspectral remote sensing data for disease detection has been shown for example by Apan et al. [2004], Shaw and Kelley [2005], Jacobi and Kühbauch [2005], Huang [2007] and Franke and Menz [2007]. Recent progress in hyperspectral sensor development and growing availability of airborne spectroscopy data provide the opportunity for upscaling methods and findings made with near-range hyperspectral data to airborne or space-borne level. Open questions within the field of spectroscopy for crop stress detection are following, whereby the points printed in bold are addressed in this thesis:

- Derivation of biochemical parameters with near-range and airborne spectrometry at canopy scale
- **Quantification of disease severities**
- **Discrimination of stress factors**
- Early detection of crop stress using airborne hyperspectral imaging
- **Relevance and redundancy of wavelength areas within the spectrum**

Several experiments in the laboratory and in the field were carried out for this study to describe and to make use of spectral changes due to a fungal infection of wheat. A pathogen infection affects the spectrum of leaves and canopies at different wavelength. Figure 1.5 displays a wheat stand infected with Powdery Mildew in the greenhouse as recorded with the hyperspectral imaging system ImSpector. The white mycelia spots can be identified easily even with a true color band combination. Thus, the influence

of a fungal infection on wheat can be visualized pixel by pixel. Figure 1.6 shows the spectra of seven consecutive pixel, going from visually not infected (Inf+0 in green) to the center of a Powdery Mildew mycelium (Inf+6 in red). The green peak at 550nm seems to vanish since the reflectance in the red around 650nm increases due to the loss of chlorophyll. The inflection point around 720nm shifts to shorter wavelength as shown in figure 1.5 on the right. The reflectance in the near-infrared region increases, but not as significantly as the changes in the visible part. As visualized, a fungal infection influences the spectral behavior of entire wavelength ranges, but a focus on the most affected bands may be of profit for the spectral detectability and detection accuracy.

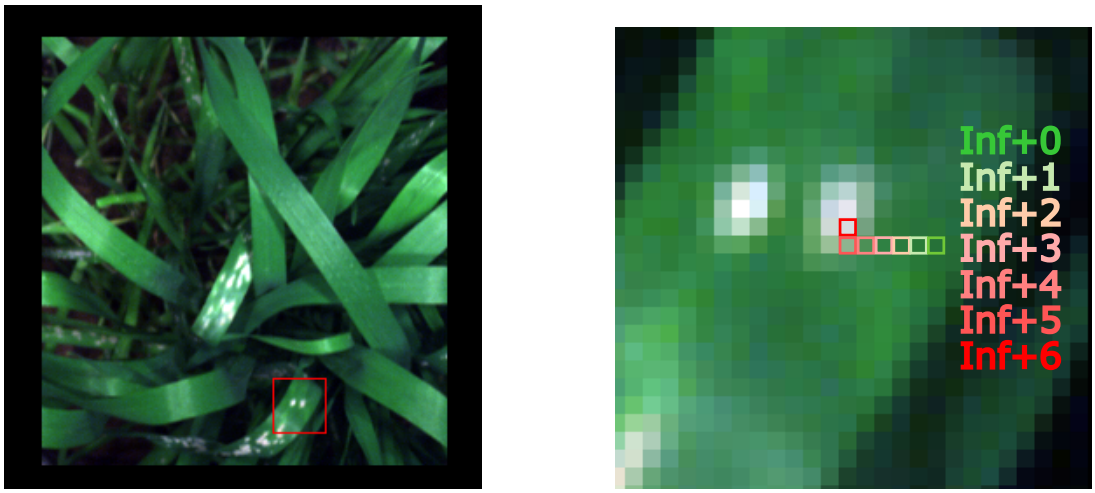


Figure 1.5.: RGB color image of a fungal infected wheat stand as recorded by Imspector V10E in the greenhouse (left) and the corresponding zoom image of a powdery mildew mycelium (right).

1.4. Data reduction of hyperspectral data

Hyperspectral sensors deliver spectral data with contiguous bands resulting in a huge amount of data. On the one hand narrow bands give the opportunity to measure spectral differences within small spectral ranges, but on the other hand, those circumstances require advanced processing methods to handle the large and complex data amount. Besides, since the spectral bands of spectroscopy data are quite narrow, their information might be related in some dimensions. For example, if the spectral information of two neighboring bands of a ASD Fieldspec® with a spectral resolution of 1.43nm is compared, the difference or rather the gained information is low in many cases. This

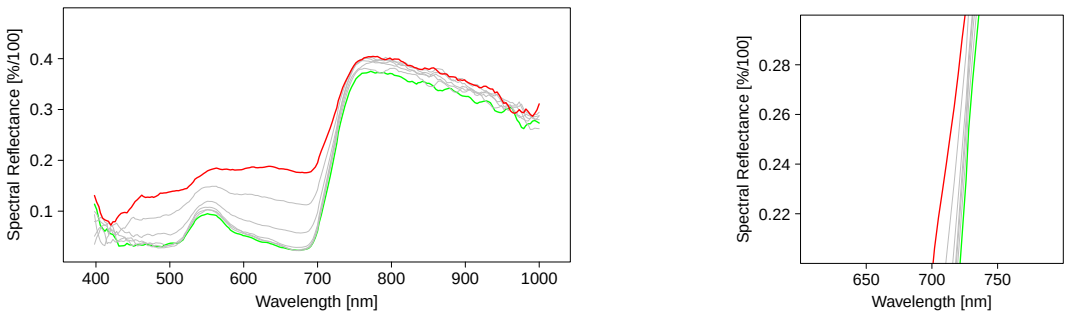


Figure 1.6.: Spectral changes due to different intensities of fungal infection from visually no infection (Inf+0 in green) to strong infection (Inf+6 in red); full spectrum (left) and inflection point at 720nm (right).

phenomenon known as autocorrelation needs to be taken into account when working with hyperspectral data. In other words, the methods and techniques for hyperspectral data analysis need to be able to handle data with an high amount of autocorrelation. Another way to cope with this problem is the removal of redundant data within a hyperspectral spectrum. Goetz [2009] stated that the modifier 'hyper' means 'too much' and that no single material requires hundreds of spectral bands to be identified uniquely. Several studies have shown that the large number of spectral bands increase the complexity of the analysis and analyzing techniques may profit by focusing on relevant spectral bands or wavelength areas [De Backer et al., 2005, Serpico and Bruzzone, 2001]. Thenkabail [2001] compared the use of narrow bands with the use of multispectral broad bands for the discrimination of six different crop classes. The study has shown an increase of detection accuracy by using hyperspectral data and, as a crucial point, significant improvements by using specific selected bands for vegetation analysis. This leads to the assumption, that the use and anlysis of hyperspectral data may profit by a preceding data reduction or feature selection. The following paragraph lists different feature selection approaches and distinguishes these from feature extraction. Its focus lies on methods that are available for and applicable on hyperspectral remote sensing data.

Band selection vs. feature extraction There are different approaches to reduce the dimensionality of hyperspectral data, which can generally be divided into two groups. First, feature extraction techniques that rearrange the existing feature space or generate new features, e.g. principal component analysis (PCA) or maximum noise fraction (MNF) (e.g., Lu et al. [2007], Chen and Qian [2008]). Secondly, feature selection ap-

proaches which reduce the feature space by identifying and selecting suitable bands, e.g. for class separation. Feature selection or band selection methods can furthermore be differentiated into filter and wrapper approaches [Kohavi and John, 1998]. For example, filter approaches use decision trees [Breiman et al., 1984, DeFries and Chan, 2000] or separability measures such as the Bhattacharyya distance [Bhattacharyya, 1943, Jimenez and Landgrebe, 1999]. These can independently be used for feature selection of the subsequent classification algorithm by searching for the feature combination which produces a high class differentiation. An overview of some band selection methods is given in Bajcsy and Groves [2004].

Wrapper approaches use the classification algorithm and a specific accuracy measure directly for feature identification by maximizing the classification accuracy [Waske et al., 2009]. If a classification into predefined discrete classes is wanted, a wrapper approach may increase the classification accuracy by using a few spectral bands specifically selected for the used classification algorithm [Keshava, 2004]. For every band combination a classification is performed. The results are compared and the spectral subset with the highest classification accuracy is then selected for further processing. Depending on the questionnaire any classification algorithm can be used for feature selection by implementing a wrapper approach. However, a classification for every band combination is computationally intensive and thus time-consuming.
Vegetation indices or ratio approaches are simple data reduction methods. Usually two to four spectral bands are used to calculate a ratio between a spectral area of high reflectance and an area of high absorbance, like the NDVI by [Rouse et al., 1973]. But since the chosen wavelength bands are predefined it is neither a selection nor extraction method.
As this study examines the impact of the spectral dimension of hyperspectral remote sensing data, the main focus lies on the identification of redundant and relevant spectral data for crop stress detection. Filter and wrapper approaches for feature selection will thereby be tested to increase the significance of the extracted information.

1.5. Thesis objectives

This thesis is the third remote sensing focused part within the three periods project *'Information Techniques for Precision Plant Protection'* of the Research Training Group 722, which is funded by the DFG (German Research Foundation). During the first phase

(2001-2004) the focus of remote sensing based studies lied on the spatial dimension of crop stress detection. The effect of a differing spatial resolution on the detectability of crop stress was thereby examined to define a minimal pixel size for a successful differentiation of healthy and stressed wheat stands (see Voss [2005]). The second phase (2004-2007) was focused on the temporal dimension of crop stress detection. The spatio-temporal dynamics of different fungal infections in wheat were thereby examined to define sensor requirements at the temporal scale [Franke, 2007].

To complete these coordinated phases by the assessment of the spectral dimension of crop stress detection, the present thesis, as phase three (2007-2010), focuses on the use of hyperspectral data. The overall aims of this phase were to examine the general use of hyperspectral spectroscopy data for Precision Agriculture questionnaires, the identification of significant wavelength areas for crop stress detection and the impact of a differing spectral resolution. Figure 1.7 shows the consecutive three phases of the project in a three dimensional space.

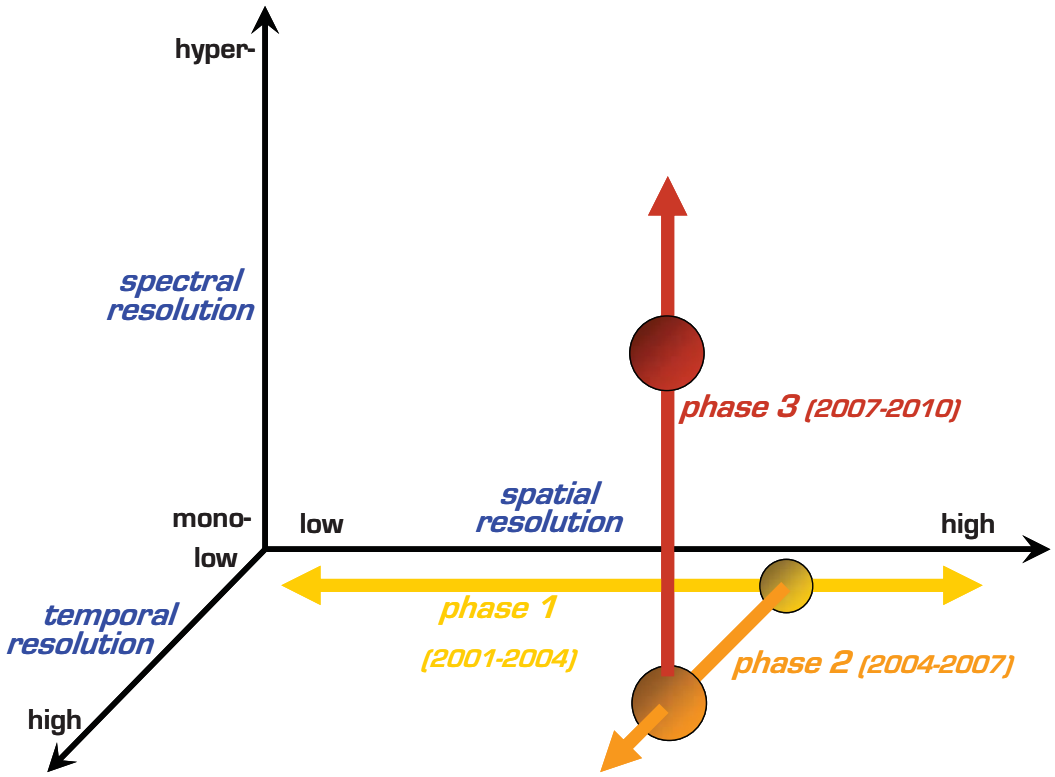


Figure 1.7.: Coordinated periods of the remote sensing projects of the DFG-Research Training Group 722 [Franke, 2007].

For this study, a number of experiments were carried out in field and laboratory regarding the impact of the spectral dimension of hyperspectral data for crop stress detection. The following chapters are separated by the used data type, i.e. near-range or airborne spectroscopy, or by the focus of the analysis, i.e. feature selection or disease quantification. Figure 1.8 gives an overview of the arrangement of the different chapters with its focus and the used hyperspectral datasets. In the center of the figure the approaches for analysis, which iteratively base on each other, are displayed. Since every chapter uses data derived from different study sites a short introduction will be given in each chapter respectively.

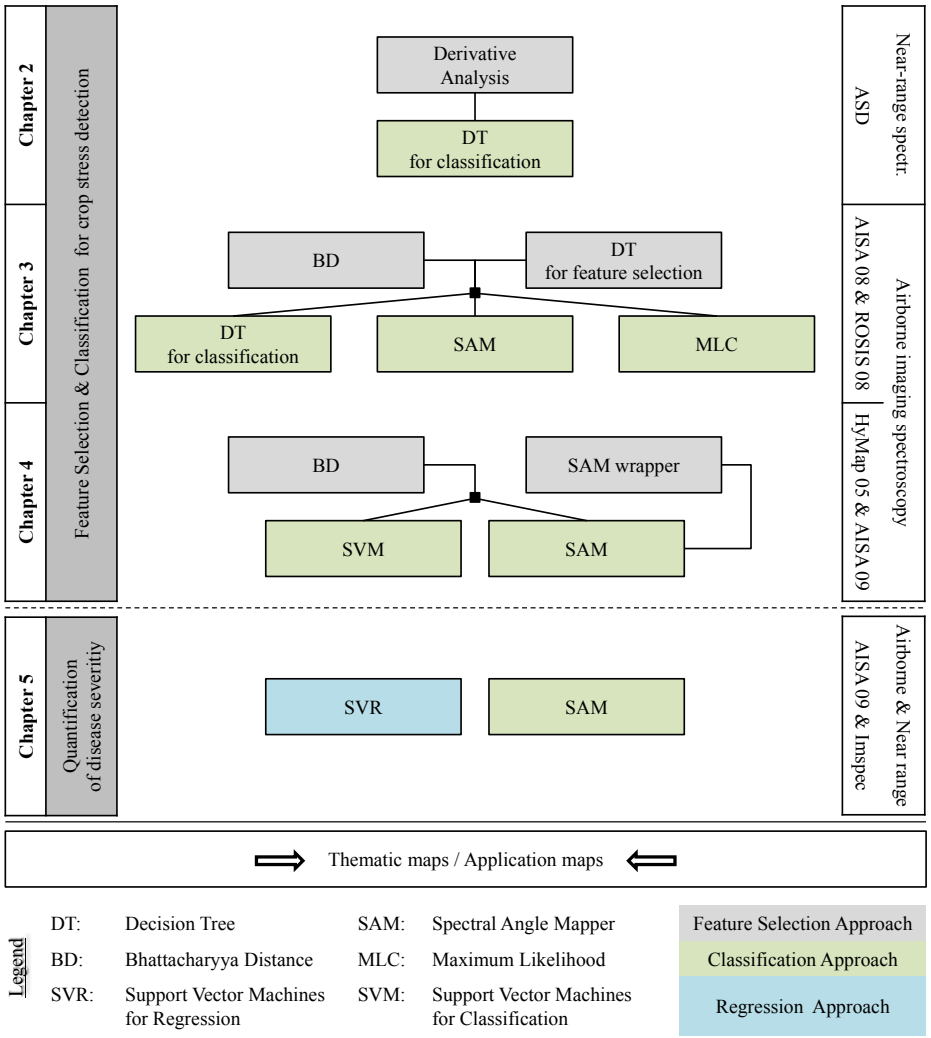


Figure 1.8.: Workflow of the entire thesis. The different foci are displayed on the left in gray, the different approaches for analysis in the center in light gray and light green and the different used datasets on the right.

Chapter 2 *Hyperspectral data - sensors and general preprocessing* provides a detailed overview to recent hyperspectral imaging and non-imaging sensor systems used for this study. The preprocessing chain which is very crucial for the quality of the data product will thereby be highlighted.

Chapter 3 *Near-range spectroscopy for crop stress detection* examines the use these data for the detection of wheat stands infected with pathogens. In a greenhouse trial, different canopy measurements of wheat with a non-imaging field spectrometer were conducted. The wheat stands differed in pathogen infected stands, stands with water deficiency and 'vital' stands. The analyses are focused on the spectral detectability of wheat stands, the differentiation of stress caused either by a pathogen infection or by water deficiency, and on the improvement of the prediction accuracy by developing specific indices. These indices were derived from a preliminary band selection, which is, in this case, a semi-automatic selection approach based on the calculation of local extrema within the spectral curve.

Chapter 4 *Airborne hyperspectral remote sensing for crop stress detection* focuses on the general detection of fungal infections in wheat with hyperspectral remote sensing data. Two different airborne hyperspectral images were compared in their use to discriminate between healthy and infected wheat stands. First examinations of the impact of different band combinations as selected by band selection methods on the detectability of cropstress were thereby additionally examined.

Chapter 5 *Reduction of the spectral dimension of hyperspectral data* focuses on the impact of the spectral dimension for the detection of fungal infections in wheat. Band selection methods were applied on different airborne spectroscopy data to identify a optimal number of spectral bands for the detection of a fungal infection. The influence of a differing spectral resolution by iteratively merging neighboring bands was additionally analyzed to define minimal spectral requirements of hyperspectral sensor systems in the context of Precision Agriculture.

Chapter 6 *Derivation of disease severities using hyperspectral data* is focused on the derivation of disease severities using either near-range or airborne spectroscopy data. In a greenhouse experiment several wheat stand were spectrally recorded over a defined time period using an imaging spectrometer. Different disease severities of the same plot

could thereby be examined. The visual rated severities could be modeled by using a pixel by pixel classification. In a second experiment using airborne hyperspectral data disease severities were modeled using a regression approach.

Chapter 7 *Synopsis* summarizes the findings of the conducted studies and concludes the thesis. An outlook on ongoing and future studies is given in this summary and new possibilities using future sensors are outlined. A short summary to the practicability of the usage of hyperspectral remote sensing data for Precision Agriculture will be given.

2. Hyperspectral data - sensors and general preprocessing

2.1. Introduction

Hyperspectral data are very complex and require specific preprocessing techniques. There are several sensor systems available for near-range spectroscopy and for airborne imaging. The following sections consider spectroscopy data in two different scales. First, the use of near-range spectroscopy data will be discussed with a presentation of the used spectrometer and crucial considerations about the measuring setup. Secondly, the scale is enlarged by going from ground spectral measurements to local mapping of entire areas by giving an introduction into airborne imaging spectroscopy.

2.2. Near-range spectroscopy

The use of spectroscopy data has been proven to be suitable for many questionnaires regarding earth observation sciences and applications. The role of field spectroscopy has significantly increased over the last twenty years due to its large area of applicability by providing the possibility of obtaining optical measurements of surfaces or elements directly in the field. Milton et al. [2009] stated that the applicability of field spectrometers can be categorized into three classes:

- Measuring the spectral properties of individual elements, which can be realized by measuring an object using a contact probe or a leaf clip.
- Measuring the reflectance of areas like soil surfaces or canopies, which usually results in a somehow mixed or more heterogeneous spectral signal.
- Measuring homogeneous surfaces for calibration and/or atmospheric correction purposes of airborne or space-borne remote sensing data.

Furthermore, depending on the questionnaire, field spectrometers can be used for a detailed analysis of specific spectral bands under more or less controlled conditions. Thus, the data can be useful for the development of enhanced processing or analyzing techniques and is suitable for upscaling a measurement, for example going from leaf scale to canopy scale.

2.2.1. Sensors at near-range

In this thesis the ASD Fieldspec® by Analytical Spectral Devices (ASD Inc., Boulder, CO, USA, see figure 2.1) was used for ground spectral measurements in the field and in the laboratory. It is a non-imaging spectrometer which measures the radiance as reflected by an element or surface. The ASD Fieldspec® covers the spectral range from 350nm to 2500nm and is therefore called a full range sensor. A silicon photo-diode array covers the visible and near-infrared range and two cooled InGaAs photo-diode arrays cover the spectral range of the short-wave-infrared (SWIR). The following table 2.1 lists the main specifications of the spectrometer.

Radiometric quantization	16bit
Total field of view (FOV)	25°
Spectral range	350 - 1000nm (VIS/NIR) 1000nm - 1830nm (SWIR-1) 1830nm - 2500nm (SWIR-2)
Spectral resolution (FWHM)	3nm (VIS), 10nm (SWIR1+2)
Spectral sampling interval	1.4nm (VIS), 2nm (SWIR1+2)
Number of spectral bands	2150 (resampled to 1nm steps)

Table 2.1.: Specifications of ASD Fieldspec® 3 Portable Spectroradiometer

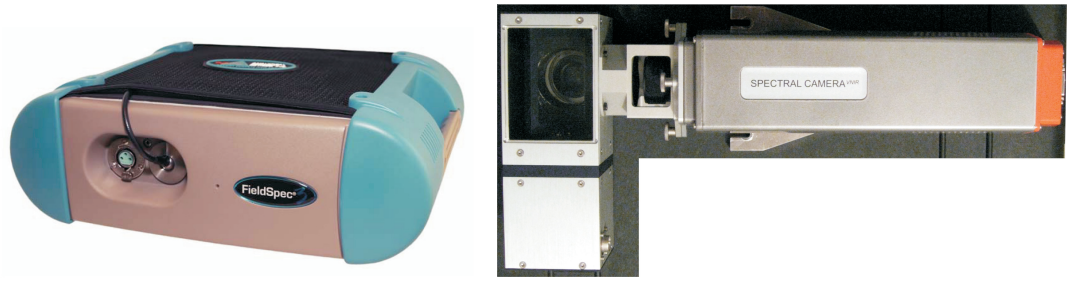


Figure 2.1.: Image of the ASD Fieldspec 3 FR (Analytical Spectral Devices Inc., Boulder, CO, USA) [left]; ImSpector V10E with scanning mirror (Specim, Spectral Imaging Ltd., Oulu, Finland) [right].

In the laboratory the imaging spectrometer ImSpector V10E by Specim (Oulu, Finland, see figure 2.1) was used to obtain spectral image data under controlled conditions. The sensor is a half-range sensor, covering the VIS/NIR region of the electromagnetic spectrum between 400nm and 1000nm. Following table 2.2 lists the main specifications of the imaging spectrometer. The spatial resolution of the sensor strongly depends on the size of the CCD (charge coupled device), the focal length of the used optics and the

height of the scanner in relation to the measured object. Simple equations can be used to calculate the spatial resolution (spatial size of a pixel). The sensor uses a CCD with a size of 11.84mm, resulting in 1600pixel per line. The used optics are defined with a focal length of 23mm. With these information, the FOV (field of view) of the sensor system can be calculated.

$$FOV = 2 * arctan\frac{x}{2f} \tag{2.1}$$

, whereby x is defined by the size of the CCD and f by the focal length of the optics. Using the information given above, the total field of view of the sensor setup is 28.8682°. The IFOV (instantaneous field of view), i.e. the angle seen by a single detector element, can be calculated by dividing the FOV angle with the number of spatial pixel. For the calculation of the spatial resolution, the sensor height above the object of interest, the size of a single detector element and the focal length need to be considered.

$$imagepixelsize[mm] = \frac{sensorheight[mm] * sizeofdetectorelement[mm]}{focallength[mm]} \tag{2.2}$$

Example:

$$imagepixelsize[mm] = \frac{1260mm * \frac{11.84mm}{1600pixel}}{23mm} = 0.4054mm \tag{2.3}$$

Since the ImSpector is a line scanner the y-dimension has to be captured either by shifting the sensor position in a steady movement, by shifting the object in a steady movement, or by using a rotating mirror device, which was used for this thesis. To gain pixel sizes with the same length and width, the scanning speed need to be adjusted. The scanning speed is determined by the IFOV, which defines the size of the pixel across-track, and the frame rate, which depends on the computational power. For example, using a frame rate of 20Hz with a IFOV of 0.0184°, leads to a scanning speed of 0.3687deg/s, which is quite slow. To increase the speed, and thus increasing the data acquisition time, the frame rate needs to be adjusted or the spatial resolution should be lowered. Figure 2.2 and figure 1.5 display images as acquired by the ImSpec V10E in the greenhouse for this thesis.

Total field of view (FOV)	28.87° (23mm focal length)
Spectral range	400 - 1000nm (VIS/NIR)
Spectral resolution (FWHM)	2.83nm (435.8nm) 3.15nm (696.5nm) 3.24nm (912.3nm)
Spectral sampling interval	2.73nm
Number of spectral bands	210 (at spectral binning 4)

Table 2.2.: Specifications of ImSpector V10E spectrograph

2.2.2. Experimental setup for near-range spectroscopy

The experimental setup for data acquisition with non-imaging and imaging spectroscopy is very crucial. To gather reliable data, several factors regarding the illumination, the positioning of the white reference and the measurement itself have to be considered, especially if a target will be spectrally examined multitemporarily. For this thesis, two experiments with multitemporal data acquisition were conducted in the laboratory, one with a non-imaging spectrometer (ASD Fieldspec®) and another one with an imaging spectrometer (ImSpector). The setup was kept constantly in both cases and is discussed in detail in the appropriate chapter. But, any differences in the illumination conditions should be avoided since their influence is tremendous. Detailed discussions within the context of experimental setup and artificial illumination for spectroscopy are for example given in Delalieux et al. [2008] or Milton et al. [2009].

2.2.3. Preprocessing of near-range spectroscopy data

There are several factors that have to be considered when working with near-range spectroscopy data. If the sensor system is recalibrated frequently, the analyses can be based on at-sensor-radiances. But, to work with spectral data which is comparable over longer time periods and under different environmental conditions, the data acquisition usually bases on reflectance data. Reflectance, in this case, means the proportion of outgoing radiance from an element or a surface to incoming radiance. In the context of remote sensing, the incoming radiance is solar induced and the outgoing is the reflectance of a surface. The transformation of at-sensor-radiance data to reflectance data can be expressed as a normalization approach. Three different point measurements or images, depending on the sensor type, have to be therefore acquired, i.e. dark current data, reference data and object data. But, beforehand, the integration time has to be specifically adjusted to the illumination conditions. On the one hand this has to be done to avoid

overexposing and, on the other hand, to exploit the maximal range of radiance differences. For both of the above mentioned sensor types, this can be realized by measuring the reflectance of a surface which is brighter than any element of a target, i.e. a white surface or a white reference.

The second step is the dark current measurement. The temperature of a sensor system varies over time, which has an influence on the spectral signal. To compensate the influence, or rather to correct the spectral signal, the inherent irradiance of the sensor has to be measured by closing a shutter to exclude external light. For non-imaging spectrometers like the ASD Fieldspec® one dark current signal is used per wavelength. For imaging spectrometers like the ImSpector one dark current signal per wavelength and per pixel is acquired to assess the differences in the spectral range as well as in the spatial range.

The third step before the target will be measured is the white reference measurement, which is literally the transformation process from at-sensor-radiance to at-surface-reflectance. For non-imaging spectrometers a white or gray panel has to be positioned at surface level. The panel has to comprise the entire light cone as determined by the field of view of the sensor system. For imaging spectrometers a white panel should be used which comprises the entire spatial area in x- and y-direction. Thus, every pixel can be transformed into reflectance by using its appropriate reference pixel. But, since the area along-track might be very large, smaller panels can be used. For this thesis, a reference panel for the ImSpector was used which covered the entire across-track spatial axis, but not the along-track axis. The following figure 2.3 shows three images as acquired by the ImSpector and the resulting calibrated and normalized image product.

2.3. Airborne Imaging Spectroscopy

Since the last decade the market for hyperspectral imaging or imaging spectroscopy systems, is rapidly growing. This concerns the demand for hyperspectral data for various applications, as well as the choice of imaging and non-imaging system providers. Most users just want to have preprocessed data with a maximal spatial resolution, highest spectral resolution and a good signal-to-noise-ratio (SNR) at low cost. The delivery of such a product is a cost-intensive and time-consuming task. This section gives a short overview of data providers in Europe and worldwide, flight planning and prepro-

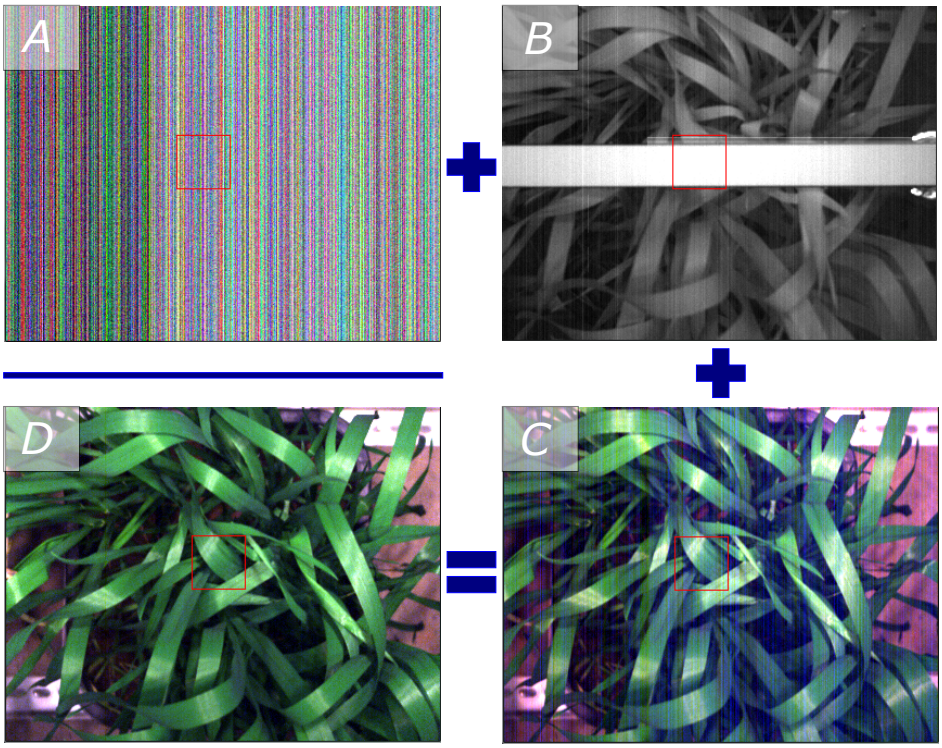


Figure 2.2.: Normalization of imaging spectroscopy data. Dark current image (A), white reference (B), target (C) and normalized image (D).

cessing techniques (to give an impression of the requirements),the processing chain and limitations of airborne hyperspectral data.

2.3.1. Hyperspectral sensors and data providers

In contrast for example to the United States, the ordering or rather the purchase of hyperspectral imaging data in Europe is a challenging task, since nearly no commercial data providers are available. In Germany hyperspectral data acquisition is mostly done by the German Space Agency (DLR, Deutsches Luft- und Raumfahrtzentrum). From 2007 to 2010 the DLR offered hyperspectral image acquisition via two different sensor systems, the *Reflective Optics System Imaging Spectrometer* (ROSIS, DLR, Germany) and the *Hyperspectral Mapper* (HyMap™, Integrated Spectronics Pty Ltd., Australia). The ROSIS sensor was developed under cooperation with the DLR and is thus located in Germany with a good availability. The HyMap™ sensor system is provided by the private company HyVista Corporation Pty Ltd. who cooperates with the sensor development company Integrated Spectronics Pty Ltd., both located in Australia. For German flight campaigns the HyMap™ sensor system is only available on a temporary basis since the system has to be rented from HyVista. Another availabe sensor system in Europe is the *Airborne Imaging Spectroradiometer for Applications* (AISA Dual, Specim, Oulu, Finland). This hyperspectral imaging system is deveolped by Specim, a private sensor development company in Finland, and is used for airborne data acquisition across Europe and worldwide. But, since most of the owners are universities or public research institutions, the purchase of such data is limited. Different datasets of the three mentioned sensor systems were used for analysis and will be presented in the following subsections.

ROSIS The hyperspectral sensor ROSIS (Reflective Optics System Imaging Spectrometer, DLR, Germany) is a compact airborne imaging spectrometer, which had been developed by the German Space Agency (DLR) in cooperation with private companies. The pushbroom scanner acquires spectral data within the range of 430-860nm with a bandwidth of 4.0nm resulting in 115 spectral bands in total. ROSIS primary function is the detection of spectral fine structures especially in coastal and inland waters. However, the spectral range of the sensor system and its high spectral resolution makes it applicable for many hyperspectral remote sensing questionnaires and applications. A short overview to basic sensor specifications of ROSIS is listed in the following table 2.3.

Radiometric quantization	14bit
Total field of view (FOV)	8°
Instantaneous field of view (IFOV)	0.56mrad
Pixel size at 3km flight altitude	1.6m
Spectral range	430 - 837nm
Spectral sampling interval	4.0nm
Number of spectral bands	115
Signal-to-noise ratio (SNR)	>500:1

Table 2.3.: Specifications of ROSIS

HyMap™ The hyperspectral sensor HyMap™ (Hyperspectral Mapper, HyVista, Australia) is a whiskbroom scanner with an ax-head double mirror. Compared to pushbroom sensors like ROSIS which acquire spectral data within one detector line at the same time (along track), whiskbroom systems use opto-mechanical devices like mirrors to collect spectral data within one line (across track). For both systems, the data in flight direction is imaged as the airplane or spacecraft is moving forward. HyMap™ has been developed by the private company Integrated Spectronics Pty Ltd (Sydney, Australia) and is operated by HyVista Pty Ltd (Sydney, Australia). The imaging sensor was developed for airborne hyperspectral data acquisition on a commercial basis and is operational since 1996. A short overview to basic sensor specifications of HyMap™ is listed in table 2.4.

Radiometric quantization	12-16bit
Total field of view (FOV)	30-65°
Instantaneous field of view (IFOV)	1-3mrad
Spatial resolution	2-10m
Spectral range	450 - 2500nm
Spectral sampling interval	10-20nm
Number of spectral bands	100-200
Signal-to-noise ratio (SNR)	>500:1

Table 2.4.: Specifications of HyMap™

AISA The hyperspectral sensor AISA (Airborne Imaging Spectroradiometer for Applications, Specim, Oulu, Finland) is a pushbroom or line scanner. There are different core products available, mainly differing in the spectral range. The AISA-EAGLE is a half-range sensor system working in the spectral range of VIS (visible) and NIR (near-infrared) between 400nm and 970nm. Another system, called AISA-HAWK, is working in the spectral range of SWIR (shortwave-infrared) between 970-2500nm. Next to the

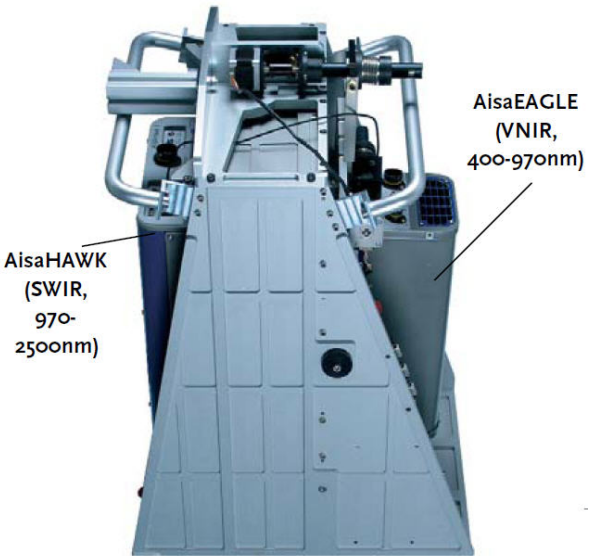


Figure 2.3.: Image of the AISA-DUAL sensor system (Specim, Spectral Imaging Ltd., Oulu, Finland)

spectral differences, the HAWK works with a CCD (charge- coupled device) which can collect 320 spatial pixel in one line (across track), whereas the EAGLE can collect 1024 spatial pixel in one line. Both sensors are united in the hyperspectral imaging system AISA-DUAL, see figure 2.3. This system is a full range sensor with a spectral range between 400nm to 2500nm. The specifications are listed in table 2.5. All sensors mentioned above are high performance airborne hyperspectral systems and thus tested and ready for airborne hyperspectral data acquisition.

Radiometric quantization	12-14bit
Total field of view (FOV)	24°
Instantaneous field of view (IFOV)	1.309mrad
Spatial resolution	0.43 * flight altitude
Spectral range	400-970nm (VNIR), 970-2500nm (SWIR)
Spectral sampling interval	3.3nm (VNIR), 12nm (SWIR)
Number of spectral bands	244 (VNIR), 254 (SWIR)
Signal-to-noise ratio (SNR)	>500:1

Table 2.5.: Specifications of AISA-DUAL

2.3.2. Flight planning

The flight planning is a very crucial and sensible point when working with airborne hyperspectral data. Different interests and requirements have to be considered and combined for a targeted flight campaign with successful data acquisition. At first stage the project or client needs to define the area, the date and the time of the flight. Furthermore, factors like flight direction, surface type, sun angle or water level affect the data acquisition and need to be considered. Thus, general requirements of the project must be combined with possibilities and limitations of the hyperspectral sensor system. In most cases, high to maximal spectral and spatial resolutions are favored, which strongly depend on the image rate. The image rate, usually indicated as FPS (frames per second), is defined by the sensor system and means how much data can be recorded per second. If a maximal spatial resolution is required, the spectral resolution usually has to be decreased, otherwise the sensor system is not able to collect consistent data due to limitations of the image rate. Figure 2.4 shows an example for the preparation of an area survey. An agricultural area was covered by an AISA-DUAL sensor system with a spatial resolution of 1.5m. The flight lines are arranged in northern to southern direction to avoid spectral effects caused by different illumination conditions, e.g. BRDF (bidirectional reflectance distribution function), shadow. The BRDF is a function that defines how light is reflected at different surfaces and considers the angles of both light source and sensor to surface which has a tremendous influence on the measured reflectance signal. BRDF effects can be minimized by a north-to-south or vise versa arrangement of the flight lines and by keeping the flight direction constantly for every single line. To cover the desired area with a south-to-north flight direction seven flight lines were needed.

2.3.3. Preprocessing of airborne hyperspectral data

The preprocessing of airborne hyperspectral data is a very complex process with high computational requirements. It includes the entire processing chain to calculate an image product from hyperspectral raw data acquired during a flight campaign. The raw data has to be radiometrically calibrated and geometrically corrected. According to the product requirements it can additionally be atmospherically corrected. ROSIS and HyMap™ data acquired by the DLR, as well as the used AISA data are usually already preprocessed and ready for analysis. But, since the preprocessing of the data is highly complex with several optional parameters and has a strong influence on the data

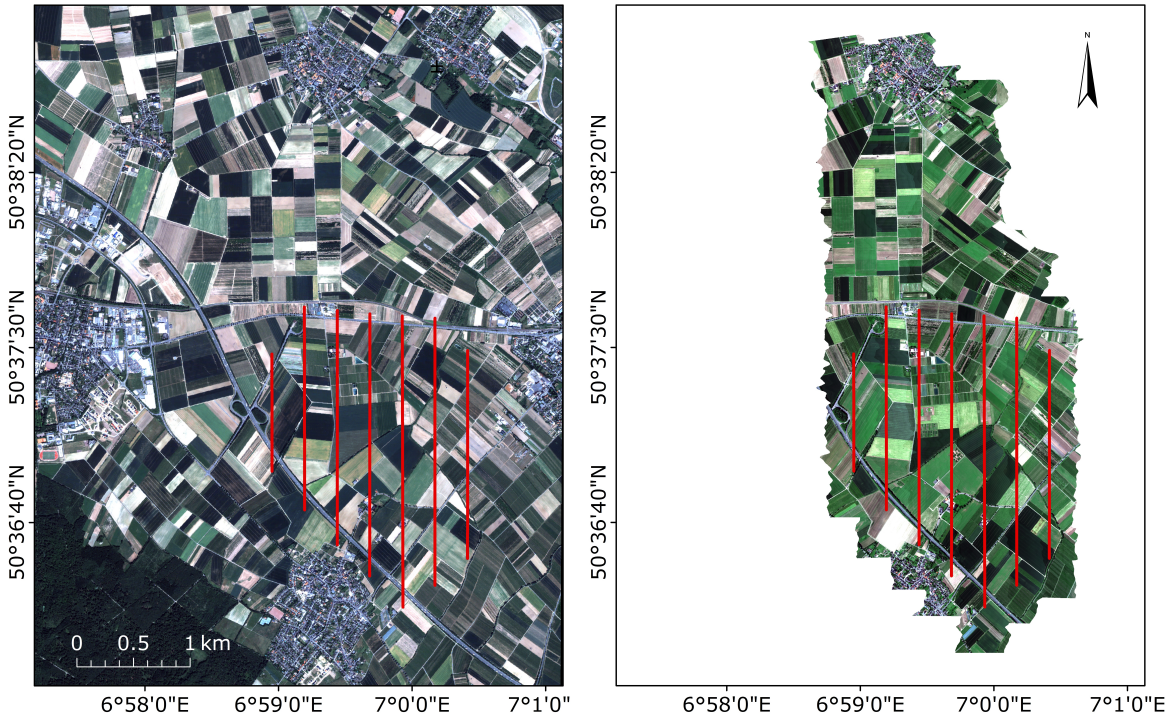


Figure 2.4.: Arrangement of flight lines within a Quickbird image from 05.06.2005 (left) and result of a data acquisition with AISA-Dual from 17.06.2009 (right) in the area of 'Klein Altendorf' near the city of Bonn in Germany.

quality, the single processing steps have to be considered carefully. The complexity of the entire preprocessing of airborne hyperspectral data will not be discussed in detail, but a short introduction into the single steps, i.e. radiometric calibration, geometric correction, atmospheric correction, will be given on the basis of one AISA dataset to show the importance and the sensibility of this task.

The AISA sensor systems comes with its own calibration software regarding the radiometric calibration and geometric correction. After successful data acquisition there are usually three different data types available per flight line.

- *.raw - In this file the raw-image-data is stored. RAW in this case means that the data is whether radiometrical corrected nor geometrical corrected.
- *.hdr - This is the header-file, associated to the raw-file, as generated during the survey. At this point it only contains some few information about day and time of the data acquisition, some sensor specifications and length and width of the flight line.
- *.nav - This file is as important as the image file itself. It contains information about the position of the airplane during data recording. This is usually done by using an IMU (inertial measurement unit), which measures the roll, pitch and yaw of the airplane, as well as the GPS data.

Radiometric Calibration The first step in converting raw data is the radiometric which transforms irradiance data into physical units. The raw DNs (digital numbers) are thereby converted to generate useful and reliable information. The range of the DNs depends on the radiometric resolution of the sensor system defined by the sensitivity of the CCD (charge coupled device). The radiometric resolution of the AISA is 12bit for the VIS and NIR spectral range and 14bit for the SWIR. 12bit means that 2^{12}, i.e. 4092, information can be stored and thus differentiated (14bit = 16.384). Every sensor system whether hyperspectral, multispectral or panchromatic, has beside the radiometric sensitivity, a spectral sensitivity. In the laboratory this sensitivity has to be measured and quantified for every single pixel of a line scanner like AISA. This is usually done by using an integrating sphere, also known as Ulbricht sphere, which is an optical component consisting of a hollow cavity with a coated interior for high diffuse reflectivity. By illuminating the sphere with a defined spectral range and simultaneous spectral data recording the spectral sensitivity of each pixel can be quantified, since the

incoming spectral radiance of each pixel is equal. This information is stored in a sensor specific calibration file, extension *.cal for AISA. During a flight campaign a dark current image is recorded, whether at the beginning of a flight line, the end of a flight line or separately. The dark current image or sample lines are necessary to record the inherent sensor noise which has an effect on the spectral signal. This sensor noise varies during a flight campaign due to changing temperatures within the system and has to be measured frequently. Using the spectral calibration file and the dark current image, the raw DN values can be converted into at-sensor-radiance and the spectral information acquired by AISA-EAGLE and AISA-HAWK are adapted. Figure 2.5 displays the processing step from raw DNs to at-sensor-radiance data, showing a spectrum as example. The spectrum in the upper part of the figure shows the raw DNs from a pixel of a vegetation surface. A significant offset near the wavelength of 1000nm is obvious. This offset is caused by the two different spectrometers used inside the AISA-DUAL sensor system, i.e. AISA-EAGLE and AISA-HAWK. Due to different radiometric resolution, the wavelength areas need to be matched during the radiometric calibration.

In some cases it is useful to convert the data into at-sensor-reflectance. This can be realized if a so called FODIS unit (Fiber Optics Downwelling Irradiance Sensor) is available, which measures the incoming irradiance from the sun. The AISA-DUAL system for example is able to collect the incoming irradiance as reflected from the surface and the incoming irradiance as emitted by the sun by reserving some pixel within the across-track line. Figure 2.6 displays the irradiance as measured by the FODIS unit and its related spectral data converted into at-sensor-reflectance.

Geometric Correction After a successful radiometric calibration usually the next step is to correct the data geometrically. The aim of this is to assign a spatial location to every single pixel. For airborne image acquisition a continual measurement of aircraft movement and position is essential. A IMU (inertial measurement unit) is therefor needed, which measures the roll (x-axis rotation), pitch (y-axis rotation) and yaw (z-axis rotation), as well as the position. Every measurement has its own timestamp. By comparing the timestamps of the IMU with those measured by the image sensor system, each pixel will be assigned with a position. If a ground-based GPS reference station lies within the survey area, their data can be used to increase the position accuracy of the in-airplane measurements by postprocessing techniques. However, an IMU or a GPS in an aircraft can not measure topographic differences within one flightline. Topographic effects have to be estimated and integrated into the geometric preprocessing

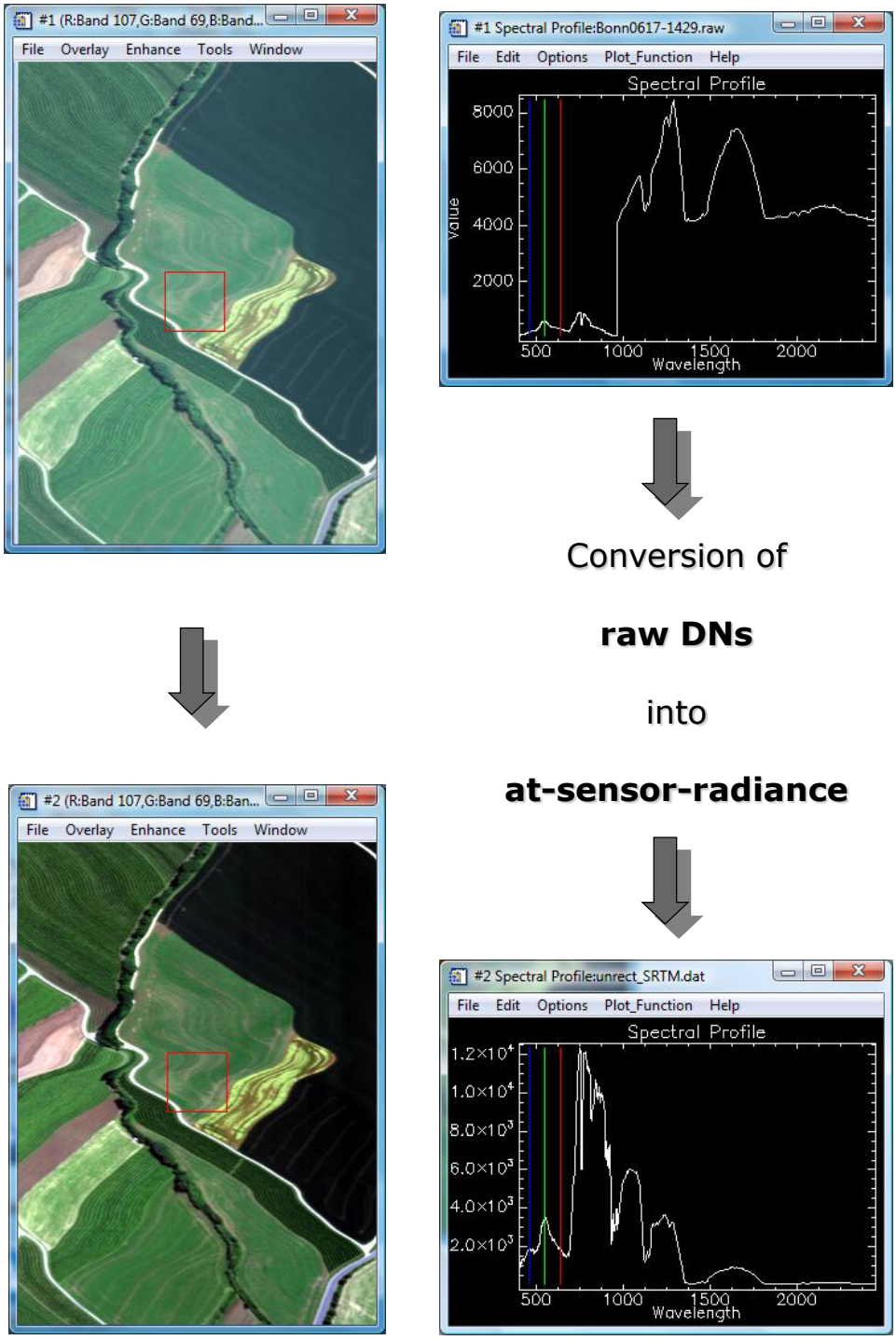


Figure 2.5.: Workflow from raw DNs to at-sensor-radiance (R: band 107 at 639nm, G: band 69 at 551nm, B: band 29 at 461nm)

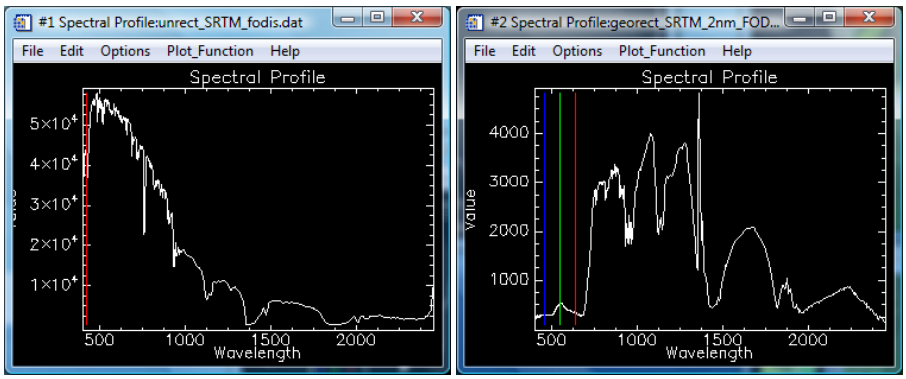


Figure 2.6.: FODIS spectrum (left) and spectral data converted to at-sensor-reflectance

of airborne imaging data, since they show significant effects on the position accuracy of the data. Depending on the surface roughness different DEMs (digital elevation model) may therefor be used. The image data can be topographically corrected by using freely available DEMs like SRTM (Shuttle Radar Topography Mission) data with a spatial resolution of 90m or ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) data with a spatial resolution of 30m. But, if a higher accuracy is preferred or needed, additional DEMs may be purchased and used with spatial resolutions and position accuracies in the range of centimeters, as, for example, derived by LIDAR (Light Detection And Ranging) remote sensing data. An example for a geometric correction is displayed in figure 2.7 showing the same location as figure 2.5.

Atmospheric Correction The third crucial step in preprocessing is the atmospheric correction. As shown in the preceding paragraphs the raw DNs can be converted into at-sensor-reflectance by using FODIS data. Since airborne hyperspectral data are acquired in altitudes of usually several hundred meters and above, the atmospheric influence on the spectral signal between aircraft and surface needs to be considered, if surface reflectances are desired.

There are several methods and options available to transform airborne hyperspectral data from at-sensor-radiance to ground reflectance. All methods try to estimate the influence of the atmosphere, e.g. water vapor or aerosols, between the aircraft and the surface on the spectral signal. Radiative transfer models like MODTRAN (Moderate resolution atmospheric transmission, Air Force Research Lab, USA) model atmospheric conditions at given time and location. It thereby calculates atmospheric transmittance and radiance, which can be used to correct spectral data. Inevitable inputs for the atmospheric correction are the exact location of the test site, the average surface elevation,

the acquisition altitude, the date and time, and the general visibility at the time of the overflight. To enhance the correction, spectral measurements on the ground should be taken simultaneously. Therefore, homogeneous surfaces have to be found and spectrally measured within the flight line. The usage of a white tarp with known homogeneous spectral reflectance is a benefit.

For this thesis airborne hyperspectral data were acquired with the ROSIS sensor, the HyMap™ sensor and the AISA sensor (see description above). The atmospheric correction of the ROSIS and HyMap™ data was done by the data provider the German Space Agency (DLR, Deutsches Luft- und Raumfahrtzentrum, Oberpfaffenhofen, Germany). Therefore, the software ATCOR-4 (Atmospheric and topographic correction, ReSe Applications Schläpfer, Wil, Switzerland) was used. The AISA data were atmospherical corrected using the software ACORN (Atmospheric Correction Now, ImSpec LLC, California, USA). Both software packages are based on MODTRAN. Figure 2.7 shows an atmospherical corrected pixel spectrum as produced with ACORN. Some spectral information within the spectral ranges with high influence of water vapor, i.e. 1.4μm and 1.9μm, were removed. Due to the high spectral resolution with 481 spectral bands and the remaining influence of water vapor in some spectral bands, some wavelength ranges seem still quite noisy.

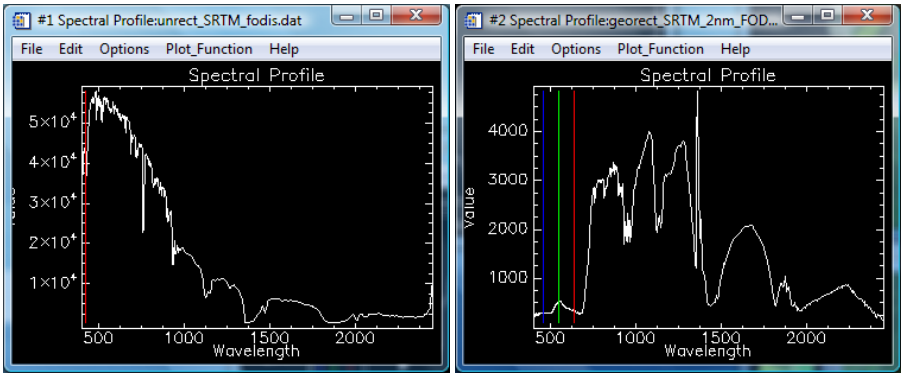


Figure 2.7.: Geometric corrected hyperspectral image as acquired by AISA on June 17, 2009 (left) and a pixel spectrum atmospherical corrected using ACORN (right)

3. Near-range spectroscopy for crop stress detection

3.1. Introduction

As described in chapter 1 *Introduction and objectives*, vegetation indices or spectral ratios, which are often used for the estimation of crop vitality [Apan et al., 2004, Jacobi and Kühbauch, 2005], can be seen as one kind of data reduction. Frequently applied ratio calculations such as NDVI use only one band in the red part of the electromagnetic spectrum and another band in the near-infrared (NIR) to transform parts of the spectrum into one band information with continual values giving the status of a specific phenomenon such as crop vitality.
However, hyperspectral data provide many contiguous bands along the spectrum, whereby not all information is needed to detect physiological plant stress symptoms. Therefore it is necessary to know which wavelengths are the most suitable to calculate such phenomenon-specific indices, since many vegetation indices are originally designed using broadband multispectral data. The identification of relevant bands for crop stress detection and thus a reduction of redundant information without loss of class separability discrimination ability is essential for a rather phenomenon-focused use of spectral data [Jimenez and Landgrebe, 1999].
Data mining or band selection techniques are useful approaches for a reduction of spectral data dimensionality and were already applied in remote sensing studies [Becker et al., 2005, Lu et al., 2007, Miao et al., 2007]. This chapter discusses a statistical feature selection of relevant bands for index calculation in regard of the detection of stress symptoms in wheat. In addition indices were tested that use most representative bands as identified during the band selection approach. These indices were applied on a decision tree for classification of crop stress to discriminate vital wheat stands from pathogen infected wheat and samples with water deficiency. Thus, new phenomenon-specific indices were designed and their cooperative usage for crop stress detection was examined.
The single questions within this chapter are (1) can different stress factors be spectrally differentiated, (2) does a statistically based data reduction affect the separability in a positive or negative way and (3) are the selected bands suitable for index calculation or development.

3.2. Data - greenhouse campaign 2008

Spring wheat (Triticum aestivum) of cultivar Triso was seeded on December 18, 2007, into five 1 m^2 boxes. The cultivar Triso is only medium resistant to the focused disease powdery mildew (*Blumeria graminis*). With an intensity of 430 seeds per m^2, the seedbed was built up in seven rows with 12cm gaps. The stands were watered every two days. On January 7, 2008, three stands were treated with fungicide. Fortress, a systemic fungicide with the agent Quinoxyfen, was applied with a dose of 30μl per stand. It has no curative effect but suppresses fungal developement up to eight weeks. Two of the treated stands were used as vital control variants (C-type) and one to cause water deficiency (W-type). The watering for the W-type was skipped after the first fungicide treatment. On January 17, 2008, two stands were inoculated with powdery mildew (P-type). The growth stage of the plants was BBCH 20 for all stands. From January 18, 2008, to February 7, 2008, i.e. 20 days, the spectral reflectance of the wheat stands was acquired by daily measurements with an ASD Fieldspec® 3 FR spectroradiometer. Three tungsten-quartz halogen lamps (ASD Pro Lamp) were used to illuminate the entire part of a box which was captured by the sensor. A distance of 90cm between lamps and object was kept constantly as well as the illumination angle and a distance of 52cm between fore optic and object. All the measurements were taken in a dark environment with the lamps as the only light source. The experimental setup is displayed in figure 3.1.

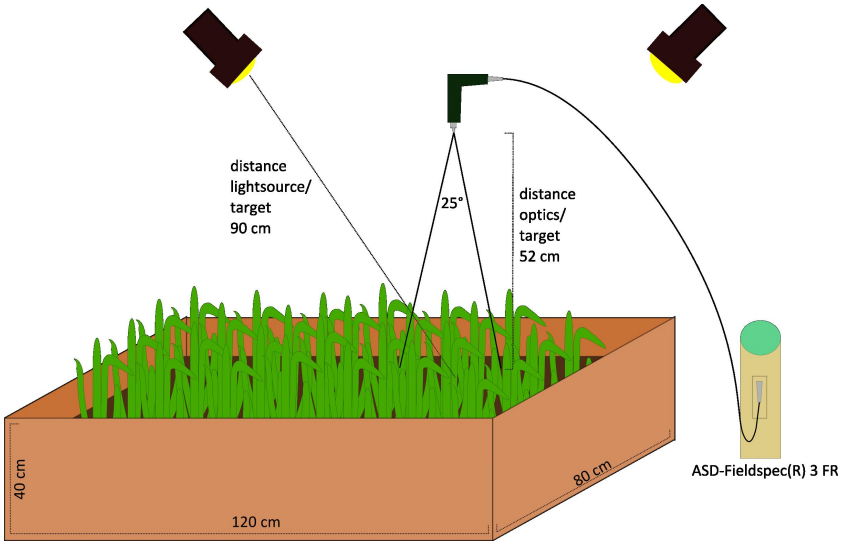


Figure 3.1.: Scheme of the experimental setup.

Every box was divided into two plots (Plot A and Plot B). This results in four rep-

etitions of the C-type stands, four repetitions of the P-type stands and two repetitions of the W-type. On January 21, 2008, the W-type and C-type boxes were treated with the fungicide Fortress a second time. First mycelia, as earliest evidence of pathogen infection could be found on P-type stands on the same day. On January 29, 2008, 12 days after inoculation, another fungicide application was necessary for the C-type and W-type stands because of high infection pressure based on the neighboring P-type stands. The data collection ended on February 7, 2008, with 80 (4x20) spectra of both C-type and P-type and 40 (2x20) of the W-type. Besides the data collection with ASD Fieldspec® digital photos of the stands were taken to record the disease progress. On day four, first chlorosis induced by an infection with powdery mildew could be seen on the P-type stands. So after day three, all measurements of the P-type stands could be visually classified as infected variants.

Spectrometry data are very sensitive to surrounding conditions. Especially when small spectral phenomena need to be recorded, the experimental setup must be kept constantly. Even under a controlled environment, some variables like water vapor still have effects on the spectral response of an object and need to be considered. Those variables produce spectral artifacts within the spectrum that may result in misinterpretations. An adaptive filtering method as a data preprocessing step has been therefore applied to every single spectral measurement. The main advantages of using the method presented by Savitzky and Golay [1964] on spectral data are given by Erasmi [2003] and were also proven in this study. Best results could be obtained using 11 points, that is, five adjacent bands upwards and five downwards from the central wavelength and assuming a fifth degree polynomial.

3.3. Methods for non-imaging datasets

In this case study, three different methods were applied stepwise trying to build up a fast and precise decision support analysis for the identification and differentiation of two stress factors in wheat. After identifying local extrema within the spectra, different ratios were calculated based on these bands. These ratio-values were then used as input in a decision tree analysis for classification to differentiate between three given classes, that is, 'vital' for the C-type stands, 'infected' for the P-type and 'water-deficiency' for the W-type. The complete work flow is shown in figure 3.2.

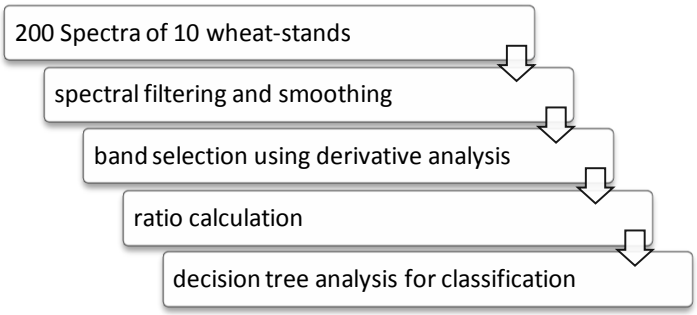


Figure 3.2.: Work flow showing single processing steps.

3.3.1. Derivative Analysis

Derivative analyses, as part of the differential calculus, have frequently been used for the identification of relevant and nonrelevant parts of hyperspectral data [Becker et al., 2005, Bajcsy and Groves, 2004, Becker et al., 2007, Smith et al., 2004]. Imaging spectrometers produce a high amount of data which requires a lot of disk space and time-consuming processing steps. The idea of band selection techniques is to reduce the data to relevant wavelengths which express the main spectral features of the spectra [Bajcsy and Groves, 2004]. By calculating the differential quotient (first derivative) of a spectrum, the slope gradient of every position will be estimated. In case of an extremum, that is, the spectral values reach a local maximum or minimum, the value for the first derivative gets null. Therefore, the spectral value r and the wavelength λ of the central band and the next adjacent band are needed.

$$d^{1st} = \frac{r_{n+1} - r_n}{\lambda_{n+1} - \lambda_n} \tag{3.1}$$

To decide whether the local extremum behaves like a maximum or a minimum, the second derivation was calculated. High positive values are associated with local minima, high negative values with local maxima.

$$d^{2nd} = \frac{d^{1st}_{n+1} - d^{1st}_{n}}{0.5 + (\lambda_{n+2} - \lambda_n)} \tag{3.2}$$

The usage of data reduction for remote sensing applications via derivative analysis is for instance proven by Becker et al. [2005] and Becker et al. [2007]. After calculating the 2^{nd} derivative values for different spectra, the author built up a spreadsheet to visualize the magnitudes and frequency of occurrence of the highest positive and the highest negative values. Eight bands could thereby be selected to differentiate the data without

a significant information loss. In the present study, the identification of high absorption and reflectance features within the spectra was processed based on the method presented by Erasmi [2003]. The 1^{st} (r'), 2^{nd} (r"), 4^{th} (r"") and 5^{th} (r""') derivatives were taken into account. By following predefined decision rules, local maxima (maxloc) and minima (minloc) could be detected.

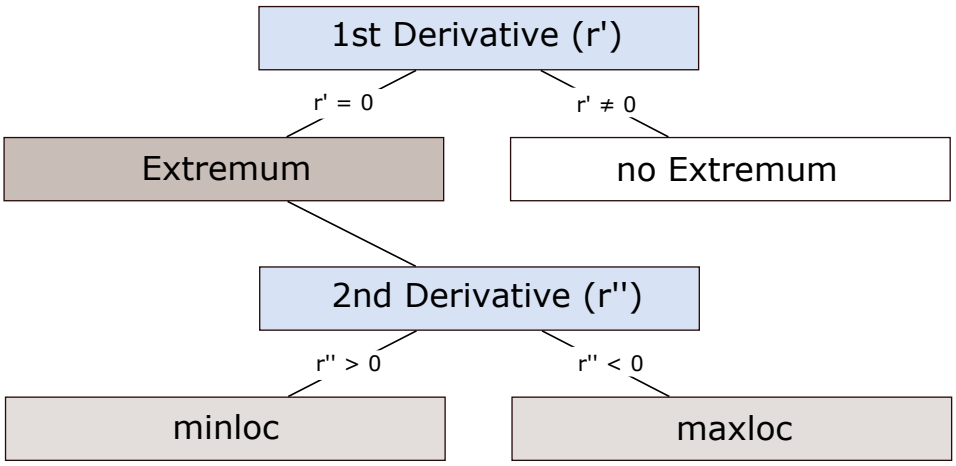


Figure 3.3.: Decision tree for the identification of local minima and local maxima

3.3.2. Ratio calculation

Frequently used vegetation indices such as the simple ratio or the NDVI use merely two or slightly more spectral bands. The advantage of ratio calculations in general is that only a few reflectance values of relevant spectral ranges are taken into account, whereby a new relative measure is created. The NDVI, for example, uses the relatively high reflectance of plants in the NIR and the high absorbance in the red to calculate a single value which highly correlates with plant vitality. Those calculations reduce processing time and enhance the comparability of different objects. Regarding vegetative objects, most indices use one band centered in the red and another one in the NIR, like the NDVI. Besides the use of these traditional bands for ratio calculations other bands were found to be suitable for vegetation analyses as proven by Smith et al. [2004].

In the present study, 12 normalized ratios were calculated. Based on every single selected local minimum, a ratio with the adjacent maxima (down- and upwards the spectrum) was calculated. For example, for the minimum at a wavelength position 626nm (A) a ratio was created using the maximum at 550nm (A_left) and another with

the maximum at 633nm (A_right).

$$Band - ratio = \frac{loc_{max} - loc_{min}}{loc_{max} + loc_{min}} \tag{3.3}$$

3.3.3. Decision Tree Analysis

Since the 1990s different decision tree algorithms have been tested on different remote sensing data, for example by Friedl and Brodley [1997]. These algorithms define class thresholds for an automatic supervised classification of the data. Within one dataset, the optimal attribute is identified using statistical calculations that enable splitting the data into predefined classes. Through an automatic threshold setting of the attributes, the classification can be processed. Different studies have shown that decision tree classification algorithms, like C5 by Quinlan or CART, have a significant potential for remote sensing applications like land-cover mapping [Friedl and Brodley, 1997, Lawrence and Wright, 2001, DeFries and Chan, 2000] or vegetation mapping [de Colstoun et al., 2003, Hansen et al., 2000].

In this study the open source software Rapidminer (Rapid-I GmbH, Dortmund, Germany) was used for the automatic decision tree building. The decision tree algorithm used in this package works similarly to C5 or CART. For the model calculation a training dataset with predefined classes is needed. In this case, the 200 spectra collected with the ASD Fieldspec® were split several times to produce different data subsets for the model calculation. Three different predefined classes were used according to the three variants vital stands (C-type), infected stands (P-type) and the stands with water deficiency (W-type). The algorithm identifies optimal attributes or nodes and thresholds to classify every single spectrum into one of the three given classes. To verify the model, a leave-one-out cross-validation was applied. The different models were then used for classification of the entire dataset.

Besides, the decision tree analyses were carried out on all collected spectra without ratio calculation. This was done in order to examine the method of band selection and ratio calculation with regard to their advantages and limitations. Finally, a decision tree analysis using single repetitions of the unrationed data, that is repetitions 1-4, was processed for comparison.

3.4. Results

The semi-automatic derivative analysis was applied on all 200 recorded spectra. About 1000 maxima and minima could be localized. Afterwards, those wavelengths were selected, at which in all 200 analyzed spectra extrema occurred at least 30 times. Hence, 19 wavelengths with maxima and 32 wavelengths with minima remained. In some cases, two adjacent extrema, however, described the same spectral phenomenon. To reduce this redundant information, further data reduction was applied by picking only those extrema within adjacent bands with the highest frequency, whereby additionally it was accounted for that a local maximum is followed by a local minimum. Thirteen bands were thus selected as displayed in figure 3.4, all located in the visible and near-infrared part of the spectrum. Twelve different band ratios (A_left, A_right, , F_left, F_right) were calculated based on the 13 identified bands.

band center λ	max [freq.]	min [freq.]	ID
550	52		
626		46	A
633	44		
666		70	B
774	55		
791		49	C
801	44		
815		34	D
818	35		
822		35	E
825	39		
846		31	F
881	31		

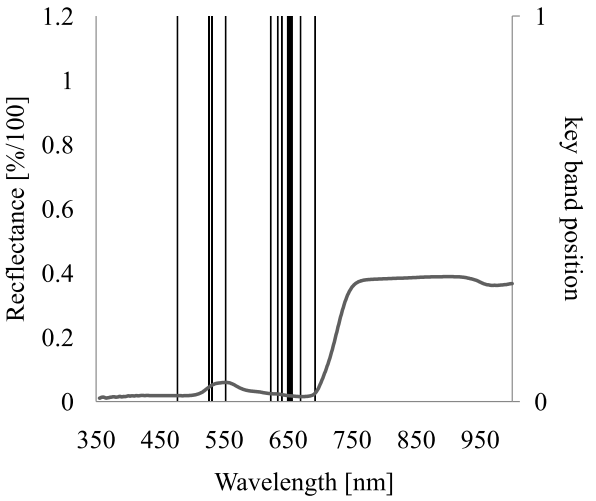


Figure 3.4.: Key bands based on frequency of occurrence (left); key band positions within the spectrum (right)

These ratios were then used as variables for the decision tree model building. In a first step the 12 variables were tested if accurate class separability was given. The spectral class separability was tested by using all 170 spectra after day three as training data for the decision tree analysis. The data were separated into three classes: 'vital', 'infected' and 'water-deficiency'. A cross-validation using leave-one-out as a sampling method was applied to verify the model. The results in table 3.1 show that all three classes could be

well separated.

	true 'vital'	true 'infected'	true 'w.d.'	class precision
prediction 'vital'	55	8	0	87.30%
prediction 'infected'	12	57	2	80.28%
prediction 'water-def.'	1	3	32	88.89%
class recall	88.88%	83.82%	94.12%	
				classification error 15.29%

Table 3.1.: Cross-validation using leave-one-out as sampling method.

The resulting model as shown in figure 3.5 classifies the data using 9 nodes, that is, 9 thresholds for 9 attributes out of 12. Because of high variability within the data, the model appears to be very complex. But regarding only the first three nodes, over 70% of the training data could already be classified successfully with only two misinterpretations, i.e. P-Type stands classified as W-Type stands.

After the model calculation and verification via cross-validation, the model was applied to the entire dataset of 200 spectra. The results are shown in table 3.2. The spectra of the first three days are difficult to classify as they produce high uncertainties. Disregarding the first three days, an overall accuracy of over 98% could be achieved.

	true 'vital'	true 'infected'	true 'w.d.'	class precision
prediction 'vital'	70	9	0	88.61%
prediction 'infected'	6	70	0	92.11%
prediction 'water-def.'	4	1	40	88.89%
class recall	87.50%	87.50%	100.00%	
				overall accuracy: 90.00%

Table 3.2.: Classification results showing separability of all 200 spectra1; 70 definable spectra were used as training data.

Rapidminer provides the possibility to interconnect an automatic data sampling into the processing chain. It is possible to repeat the entire process automatically multiple times so the analysis can be done testing multiple different models. This iteration chain was run several times using different sample dimensions and different repetition rates. For instance a decision tree analysis was run 20 times based on 85 randomly selected samples out of 170 definable spectra. As results, an overall accuracy between 83.50% and 90.00% was achieved.

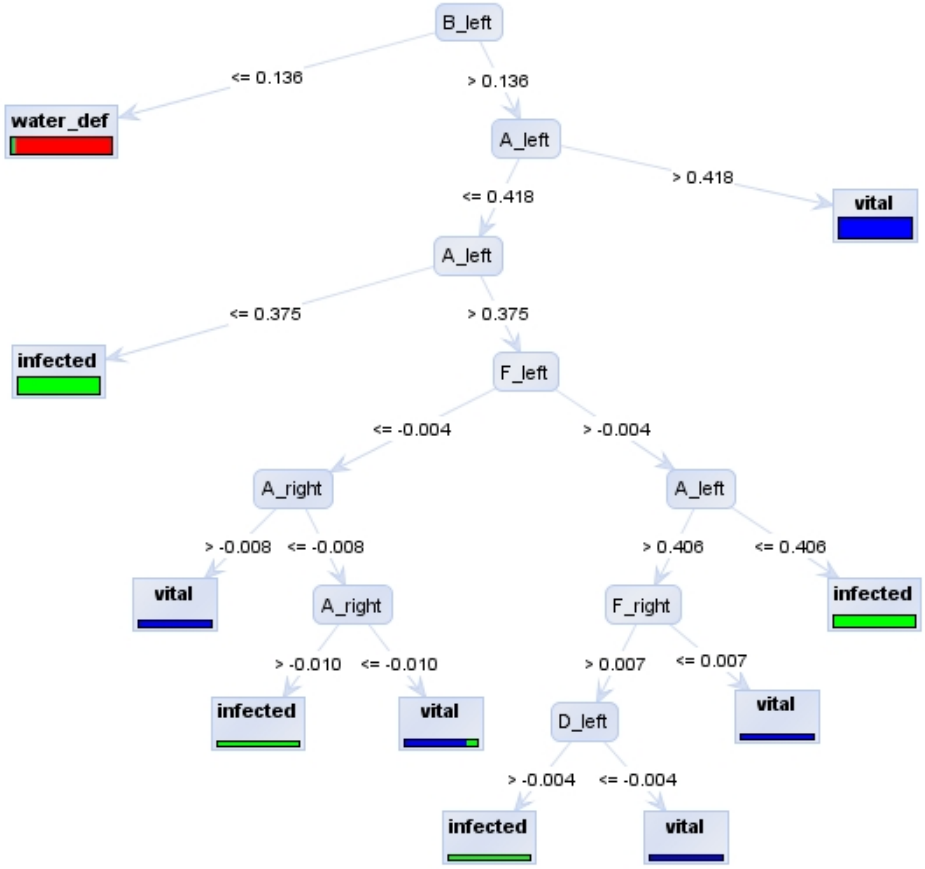


Figure 3.5.: Resulting Decision Tree using nine nodes for separation.

Besides the automatic selection of data subsets a manual splitting was done. This was necessary because the automatic sampling selects the spectra randomly. The data were manually divided into the four repetitions of the C-type and P-type stands. Both variants of the W-type were randomly distributed over the four subsets. All spectra of one repetition after day three were used as training data, that is, 51 spectra. This provided data sets with a minimal general variability. The remaining 149 spectra of the other three repetitions were then used for validation.
The entire processing chain as described above was afterwards run through every repetition. The results listed in table 3.3 show that the overall accuracy lowered but was still significant. One outlier was produced in repetition 3 for the class 'vital'.

training-data	class 'vital' [%]	class 'infected' [%]	class 'w.d.' [%]	accuracy
repetition 1, $\geq$ day 4	78.33	76.67	100.00	**80.71%**
repetition 2, $\geq$ day 4	68.33	63.33	100.00	**70.71%**
repetition 3, $\geq$ day 4	45.00	88.33	100.00	**71.43%**
repetition 4, $\geq$ day 4	80.00	71.46	100.00	**79.29%**

Table 3.3.: Results of different models based on subsets of the data for training. The values for class recall are shown.

The models resulting from only one repetition (51 spectra) are less complex than the one described above. Less variability is covered by the training data, so just four attributes with four thresholds are used to classify the data into the predefined classes. One repetition could be classified by spectral ratios with an accuracy over 98% and maximal one misinterpretation. Figure 3.6 shows the model for repetition 1 as example.

For reasons of comparison, the decision tree classifier was also applied directly on the original spectral data. All spectral bands as features without any preselection or ratio calculation were thereby used as input. Due to time constraints, the cross-validation had to be limited to 10-folds, since the leave-one-out criterium highly increases processing time when all bands or features are involved. The resulting decision tree is quite more complex. To achieve a discrimination of 52.94% of the class 'infected' from the other samples, four nodes were needed. Choosing five nodes enhances the discrimination up to 80.81% and seven nodes lead to 91.18%. Nine nodes were needed to classify all infected samples.

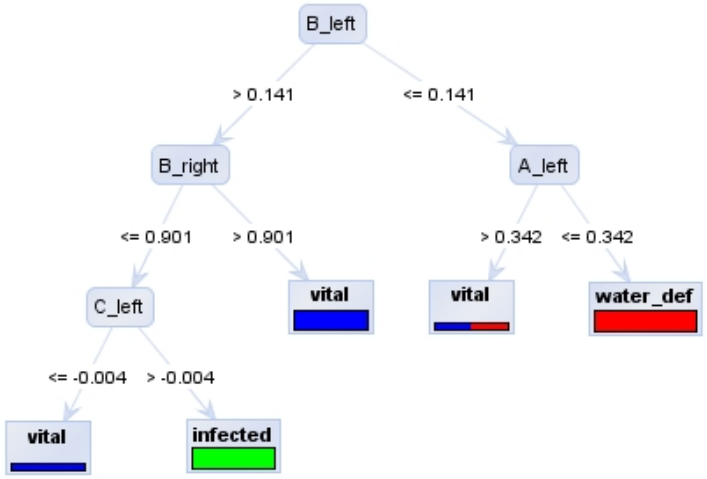


Figure 3.6.: Model example based on band-ratios of one repetition (manual data splitting) with 51 spectra for training.

3.5. Discussion and Conclusions

The results of the presented method of feature selection and index development with a following decision tree for classification are very promising. The model built up on the generated ratios based on data of stands with visible symptoms showed satisfactory class separability with a classification error of 15.29% and an overall accuracy of 90%. Even the spectra of the P-type stands with first evidence of chlorosis could be identified with minimal misinterpretation. The W-type stands could be clearly classified and separated from the P-type and C-type stands. Spectral differences between both stress factors could be assessed. It is expected that the method has potential for the detection of plant diseases at early stages. To validate the usage of the ratio calculation, the decision tree analyses were also carried out with unrationed spectra.

First, all spectra three days after inoculation with all bands, i.e. the entire spectrum without any preselection or ratio calculation, were used as training data for estimating the decision tree with a 10-fold cross-validation. The model is more complex and more than six nodes were needed to separate most of the training data. The classification error of 20% is significantly higher than the one produced by using the twelve designed ratios, but the classification accuracy of the estimated model is with 91% comparable. Thus, the advantage of band selection and ratio calculation is a lower complexity. In addition, decision tree analyses were performed using only one repetition of the raw spectra.

The second step verified the usage of ratio building. Different decision tree analyses were carried out using the original spectral data of the identified bands from the derivative analysis. Thus, the information of 13 spectral bands without any rationing was used to construct the decision tree. The resulting model is very complex. Eighteen nodes were used to separate the entire training data. The primary nodes described only 43%. Compared to the decision tree analysis using the ratios, whereby overall nine nodes were used and the primary three described over 70% of the data, this is a great disadvantage. Similar results were obtained using only one repetition for training.

The band selection technique has shown that only 13 selected bands within the VIS/NIR spectrum are significant for classification using the described method. In this respect wavelengths above 1000nm, which are covered by full-range sensors, showed to be not relevant for crop stress detection in this case. Identified wavelength positions holding a minimum or maximum above 1000nm did not pass a frequency of occurrence greater than 30. In this case, half-range sensors would have been adequate to analyze vegetation stress. However, this could be caused by a higher autocorrelation of the bands at higher wavelengths which results in extrema that are spread over various wavelengths and thus not satisfy the 30 times occurrence of the selection criteria.

The different key bands as shown in figure 3.4 have high potential for spectral classification. The identified bands are all between 550nm and 881nm. Most relevant bands within known wavelength ranges like green, red and NIR could be found. The chosen band selection method allows an identification of significant bands of hyperspectral data and thus a reduction of the spectral dimensionality of data. These bands can be used for ratio calculations as shown or as input to calculate existing indices.

Most wavelength positions with an extremum could be found in the NIR. Especially the area called red edge within this wavelength range is associated with relative high uncertainties. To cope with spectral noise, a selection limitation was set using the frequency of occurrence. Further studies need to be carried out focusing on that part of the spectrum and its uncertainties for band selection. Decision trees for classification just use specifically selected bands as nodes with a defined threshold to split given unknown samples into predefined classes. The algorithm seems to be a promising tool for feature selection, since the analysis presented here performed very well. The resulting

classification accuracies are comparable to the results obtained with the rationed data. The algorithm will thus be tested for its suitability on airborne imaging data to validate these promising results.

4. Airborne hyperspectral remote sensing for crop stress detection

4.1. Introduction

As demonstrated in the preliminary chapter 3 foliar diseases like powdery mildew affect the spectral characteristics of crop leaves and canopies and can thus be detected by spectral sensors at near-range. The potential of multi- or hyperspectral remote sensing data for disease detection was shown for example in Apan et al. [2004], Franke and Menz [2007], Franke et al. [2008]. Within the framework of Precision Agriculture, imaging spectroscopy data can thus play a major role by delivering relevant information for decision support [Kühbauch and Hawlitschka, 2003, Moran et al., 1997].
Chapter 3 showed that crop stress, or rather plant stress caused by a fungal infection, is detectable even at early stages by using near range spectroscopy data. This chapter examines the general detectability of diseased wheat stands in an agricultural field plot using airborne hyperspectral data. The focus lies on the detection accuracy of two different hyperspectral remote sensing systems for the discrimination of 'vital' wheat stands from areas 'infected' by a fungal disease. The results of a binary classification using the different datasets of the same spatial area will thereby be observed to examine differences in detection accuracy.
As discussed in chapter 1 stress impacts such as fungal infections affect reflectance and absorption features of leaves and canopies [Lorenzen and Jensen, 1989, Bravo et al., 2003], but a determination of spectral phenomena does not require hundreds of spectral bands over the electromagnetic spectrum [Goetz, 2009] since neighboring spectral bands are highly correlated. The high yield of data is furthermore inexpedient with regard to practical farming systems [West et al., 2003]. The reduction of redundant data by selecting relevant spectral features is important both to speed up the process of decision support for Precision Agriculture and to define requirements on sensor systems. As a preliminary study, the second focus of this chapter lies on the comparison of two different band selection methods tested on hyperspectral image data for their ability to reduce the amount of data with a focus on detecting fungal infections in wheat stands.
The single objectives and analyzing steps within this chapter are (1) the examination of the potential of airborne hyperspectral data for fungal disease detection in wheat, (2) the comparison of two different sensor systems in their detection accuracy, (3) the examination of the general potential of feature selection for this purpose and (4) the comparison of two different feature selection approaches as a preliminary study for chapter 5.

4.2. Study site and data of flight campaign 2008

The test site is located at 50°37'N, 6°59'E at an altitude of 175m ASL, near the city of Bonn in Germany. It is part of the experimental farm 'Klein Altendorf' of the University of Bonn. The chosen field plot, cultivated with a wheat cultivar vulnerable to leaf rust, has a size of 4.5 hectare with a flat topography. The field was divided into 12 subplots with 40*60m in size. To ensure different disease severities, variants 1 and 2 were treated with fungicides, whereas variants 3 and 4 received no fungicide application. The fertilization also differed to maximize the severity dispersion: Variants 1 and 3 received common doses and stands of variants 2 and 4 were treated with reduced doses. Each variant was repeated three times, resulting in 12 subplots. On July 1, 2008, the area was covered by the Airborne Imaging Spectroradiometer for Applications (AISA Dual, Specim, Oulu, Finland) and by the Reflective Optics System Imaging Spectrometer (ROSIS, DLR, Germany). An overview of the test site and the arrangement of the subdivisions are given in figure 4.1.

The AISA flight campaign was realized in cooperation with the Department of Water and Environmental Management, University of Debrecen, Hungary. As explained in chapter 1, the AISA Dual system combines two stand-alone sensors, AISA Eagle and AISA Hawk. As described in chapter 1 the AISA-DUAL covers the spectral range between 400 and 2500nm with a spectral resolution of 2.5nm in the VIS/NIR and 5.8nm in the SWIR resulting in 498 spectral bands over the entire covered spectrum. The data were collected at a flight altitude of 2300m, resulting in 1.5m spatial resolution.
The ROSIS flight campaign was conducted on the same day. ROSIS is a so called half range sensor system, which means that it is sensitive only to the spectral range of the VIS and NIR from 383nm up to 839nm with 115 spectral bands. The ROSIS data were acquired at a flight height about 2880m which resulted in a spatial resolution of about 2m per pixel. Both sensors covered the same experimental field plot between 12pm and 2pm on July 1,2008. Thus, the spectra of both sensors and the results produced by both sensor systems can directly be combined and evaluated. Figure 4.2 shows the images of the field plot as produced by ROSIS (left) with a spatial resolution of 2m and AISA (right) with a spatial resolution of 1.5m. In figure 4.3 a vegetation spectrum as acquired by ROSIS (left) can directly be visually compared to a spectrum of the full-range AISA sensor (right) of the same location within the experimental field.

Both sensor systems record the electromagnetic spectrum with many contiguous bands,

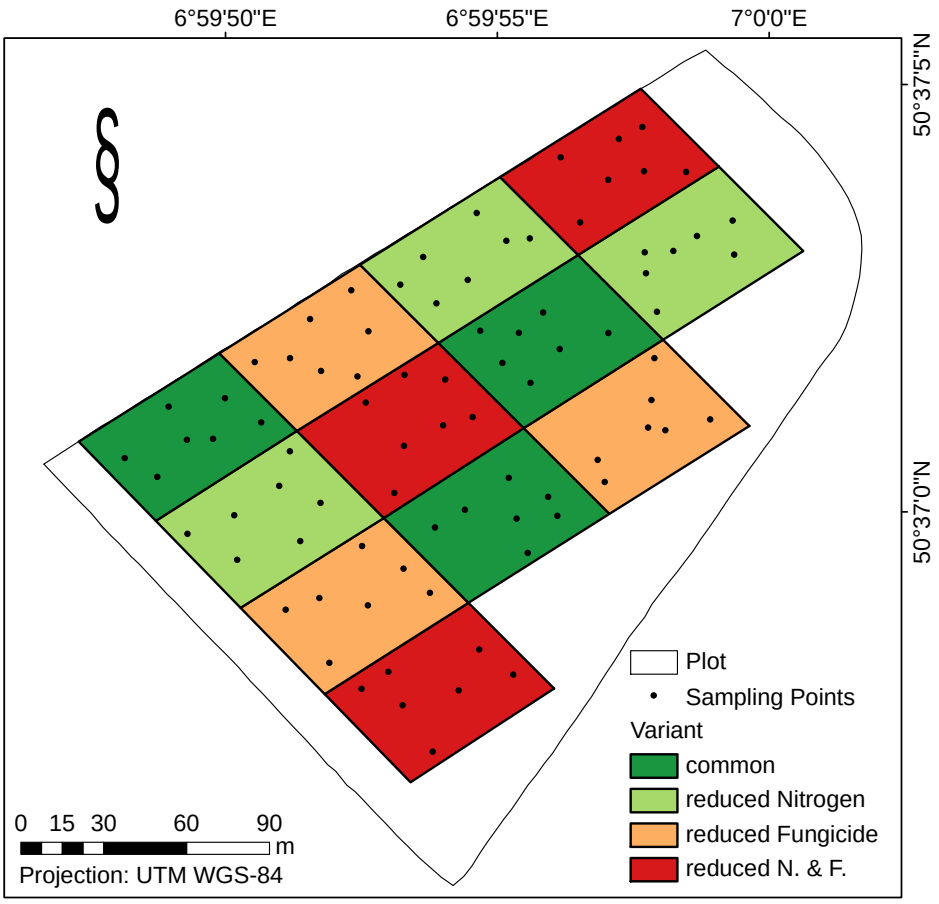


Figure 4.1.: Map of the test site showing the four sub-plots and the random disease rating points used for training and test.

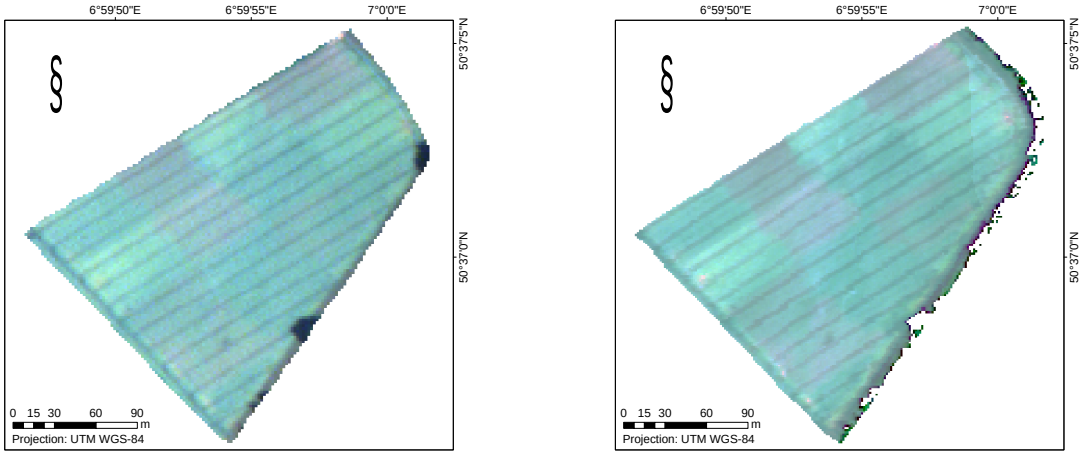


Figure 4.2.: RGB color image as produced by ROSIS (left) and AISA (right) on July 1st, 2008 (R:675nm, G:559nm, B:483nm)

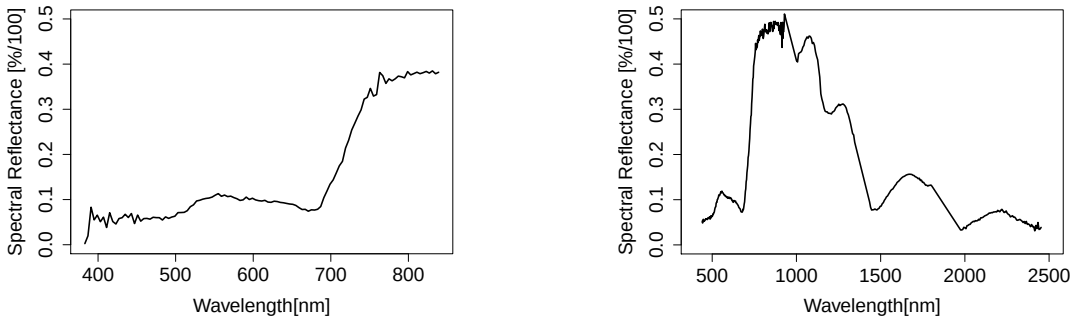


Figure 4.3.: Comparison of two vegetation spectra from the same location at N50°36'59.82" and E6°59'54.86"; ROSIS (left) and AISA (right).

resulting in high and complex amounts of spectral data. Relevant spectral phenomena and the effect of a pathogen infection may thus be examined in detail. The accuracies of both datasets are compared using the spectral angle mapper for classification with all spectral bands as input. However, a fast and precise data processing method is essential for Precision Agriculture [West et al., 2003]. Hence, different band selection methods were tested to reduce the amount of data by focusing on relevant spectral bands.
Additional ground truth data were collected at 84 randomly distributed sample points (7 per subplot). The disease severity (percentage of infected leaf area) and growth stage were rated. *Puccinia recondita* causing leaf rust occurred in the investigated field with severities between 3 and 80% of infected leaf area. For all classification and validation purposes, the variable 'disease severity' was used as the dependent variable. The rating data was therefore divided into test and training datasets.

4.3. Methods for airborne crop stress detection

To avoid misinterpretation caused by spectral noise, a manual removal of bands showing a high level of noise was conducted to the AISA dataset. Thus, the total amount of 498 spectral bands was reduced to 405 bands. The ROSIS dataset was used with the total amount of 115 spectral bands. The spectra derived from 3x3 image pixels covering the 84 sample points were then randomly divided into training (70%) and test (30%) data and afterwards binarily classified into healthy (class 'vital') or infected wheat stands (class 'infected'). Sampling points with a disease severity beyond 5 percent were thereby

classified as 'infected', and the remaining samples as 'vital'. An extra class showing nitrogen deficiencies was thereby ignored since the focus of this study is the detection of fungal diseases. However, the variants with less fertilization increased the variety of disease severity.

Training (70%, 59 samples)		**Test (30%, 25 samples)**	
'vital'	'infected'	'vital'	'infected'
26	33	11	14

Table 4.1.: Number of sample data in percent and absolute values.

4.3.1. Bhattacharyya distance

For classification purposes, appropriate differentiation ability between two or more classes using one or more spectral features or bands is necessary. The Bhattacharyya distance (BD) is a useful measure of separability between two classes [Jimenez and Landgrebe, 1999, Miao et al., 2007, Herold et al., 2003, DeFries and Chan, 2000]. The BD is defined as [Herold et al., 2003]:

$$BD = \frac{1}{8}[\mu_1 - \mu_2]^T [\frac{\sum_1 + \sum_2}{2}]^{-1} [\mu_1 - \mu_2] + \frac{1}{2} Ln \frac{|\frac{1}{2}[\sum_1 + \sum_2]|}{\sqrt{|\sum_1| + |\sum_2|}} \quad (4.1)$$

, whereby the mean vector is μ_i and the covariance matrix of class I is $\sum_i$. If two classes are given, the BD can be calculated to quantify the difference between certain features. The BD can be used as a band selection method by changing the number of input variables and selecting stepwise different band combinations, respectively. For this study a feature forward search strategy (FFS) was used as search algorithm, which means that the best single band for classification as identified by the BD analysis will be selected. Afterwards the complemental second band for the best 2-band-combination and so on will be added stepwise. The freely available software Multispec© (Landgrebe and Biehl, West Lafayette, IN, USA) was used for band selection using the BD with a FFS.

4.3.2. Decision Tree Analysis

Additional band selection techniques constitute decision tree estimation algorithms. The decision tree algorithm implemented into the Software Rapidminer (Rapid-I GmbH, Dortmund, Germany), which is similar to C5, was used to select suitable bands for the differentiation between the classes 'vital' and 'infected'. The 'information gain' as splitting criterion was therefore used.
For evaluation purposes, different supervised classifications were applied by the use of the selected bands. Common classifiers like SAM (Spectral Angle Mapper), maximum likelihood (MLC) and decision tree (DT) were compared and their accuracy was calculated by using the test data. The classifications were processed using ENVI/IDL 4.7© (ITT, White Plains, NY, USA).

4.4. Results

First, a classification using SAM with all spectral bands was performed. SAM is suitable for n-dimensional hyperspectral data and is a very common classification algorithm for hyperspectral data analysis. Both datasets, i.e. the AISA and the ROSIS datasets, were classified using SAM with a preselected training dataset. The spectra of both sensor systems reduced to the spectral range of ROSIS are exemplarily displayed in 4.4. The results of the classification were validated using a corresponding test dataset. The entire classification and validation process was repeated five times with randomly selected training and test datasets to prove the robustness of the data and of the classification. The results are listed in table 4.2 and the result of one classification is exemplarily mapped in figure 4.5 for each sensor system respectively.

Sensor			1	2	3	4	5	mean	stdv
ROSIS	30% test	OAA[%]	78.05	82.11	78.86	84.55	78.05	**80.33**	2.90
AISA	30% test	OAA[%]	85.46	82.82	84.58	83.70	85.02	**84.32**	1.06
ROSIS	30% test	Kappa	0.56	0.64	0.58	0.69	0.56	**0.61**	0.06
AISA	30% test	Kappa	0.71	0.66	0.69	0.68	0.70	**0.69**	0.02

Table 4.2.: Results of the SAM classification on the ROSIS and AISA datasets. Overall accuracies and Kappa coefficients for all five repitions.

Visually, the infected and non-infected areas nearly correspond to the variants with less fungicide treatment (see figure 4.1). A confusion matrix calculated on the test

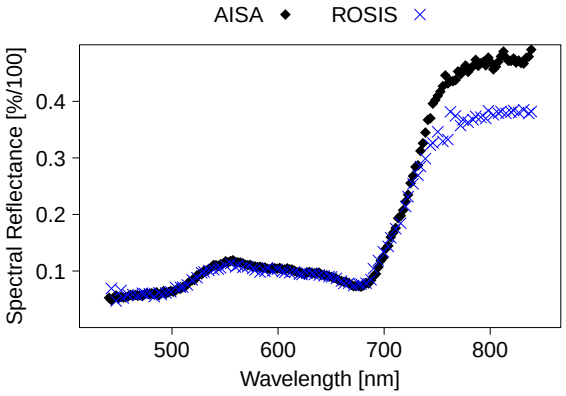


Figure 4.4.: Comparison of spectral values of a vegetation spectrum derived by ROSIS and AISA, spectrally sub-divided to the specifications of ROSIS.

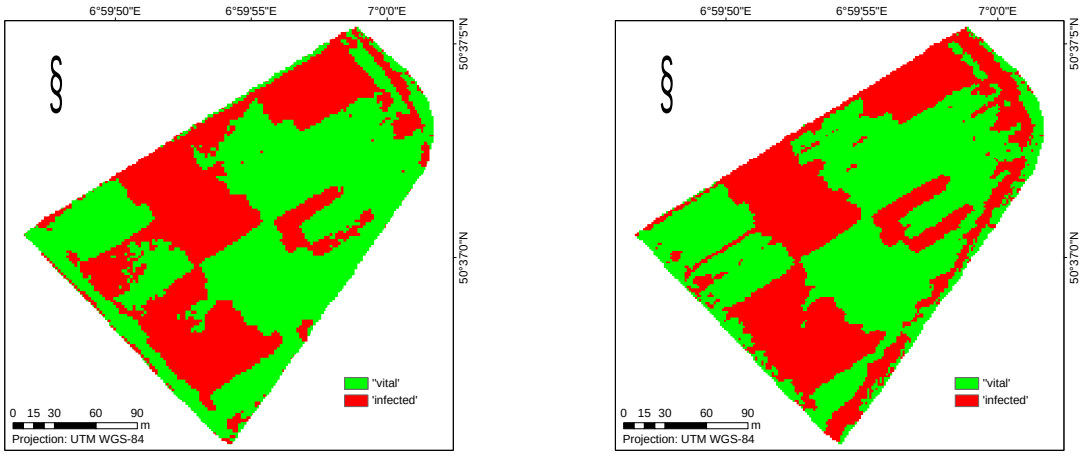


Figure 4.5.: Results of the classification using SAM on ROSIS with all 115 bands (left) and on AISA with all 405 spectral bands (right)

datasets proves the class estimations (see table 4.2). Since the results produced on the AISA datasets generally outperform the results of the ROSIS sensor system and due to the higher spectral resolution of AISA, following observations are focused on the AISA data.

4.4.1. Decision Tree Analysis

The decision tree analysis was performed using the information gain as splitting rule and a leave-one-out cross validation for accuracy assessment. The information gain splits the data where the reduction in entropy in the data is maximized [DeFries and Chan, 2000]. A decision tree was calculated using a training dataset. The performance of the model was tested using a cross validation with leave-one-out, resulting in a classification error of 32%. Six bands were used for this model (center wavelength): 440.68nm, 447.40nm, 585.39nm, 587.71nm, 727.23nm and 2433.77nm. Nearly 90% of the 'infected' samples could be classified using two bands, i.e. 587.71nm in the VIS and 727.23nm in the NIR. Three bands were needed to classify approximately 73% of the 'vital' samples, i.e. 440.68nm in the VIS, 727.23nm in the NIR and 2433.77nm in the SWIR.
The model was realized in ENVI/IDL 4.7© to classify the entire image and to validate the result with the test dataset (see figure 4.6). An overall accuracy of 78.7% and a Kappa of 0.56 were obtained. The six bands were also used to sub-divide the original AISA image and to perform a SAM and a maximum likelihood classification on the reduced dataset. The best result could be achieved using the six bands of the model with a MLC (maximum likelihood classifier).

	DT	**SAM**	**MLC**
OAA[%]	78.7	88.4	91.1
Kappa	0.56	0.77	0.82

Table 4.3.: Overall accuracies (OOA) and Kappa coefficients of different classification methods based on the six bands as selected by the decision tree analysis.

4.4.2. Bhattacharyya distance

A set of distance calculations was performed to estimate the optimal number of bands and their combinations. The BD for all combinations from 2-bands to x-bands was calculated, which resulted in a tremendous amount of data. Therefore it was merely possible

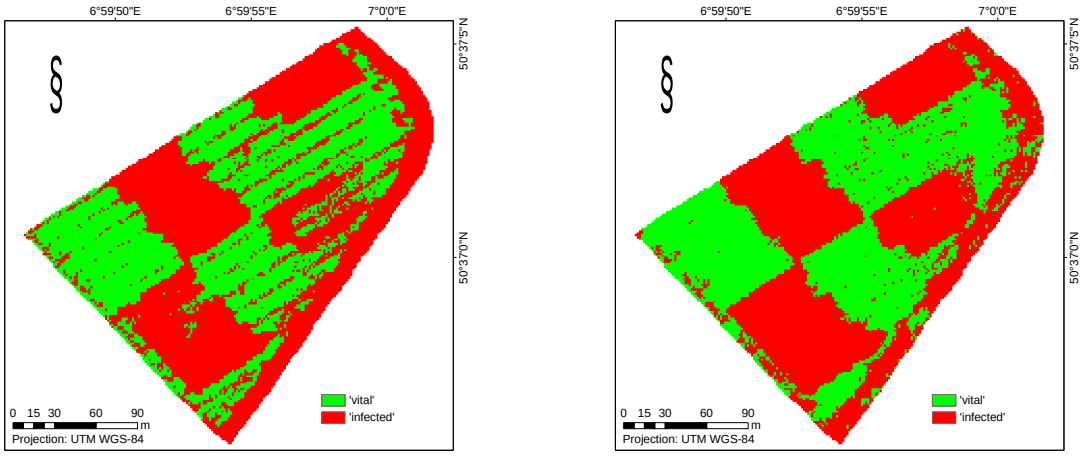


Figure 4.6.: Map of the classification results using MLC on the 6-band-combinations selected by decision tree (left) and Bhattacharyya distance (right).

to calculate up to the best 4-band-combinations, i.e. 1,101,475,905 search combinations. The calculation of all 5-band-combinations out of 405 bands exceeds the maximal number of search combinations of Multispec©, i.e. 88,578,967,581 to 2,147,483,647 maximal search combinations. Therefore, the step procedure provided by the software was used, which is a feature forward search strategy. This procedure first searches for the best single band for separation, takes this into the next step of finding the best 2- band-combination and so on.
Compared to the results of the DT analysis the best 6-band-combination was used for classification, i.e. 741.23nm, 764.86nm, 993.98nm, 1094.57nm, 1320.92nm and 2069.11nm (see figure 4.6). The AISA image was classified with SAM, with the maximum likelihood classifier and with a decision tree approach.

	DT	**SAM**	**MLC**
OAA[%]	72.9	94.2	94.7
Kappa	0.46	0.88	0.89

Table 4.4.: Overall accuracies and Kappa coefficients of different classification methods based on the six bands selected by BD.

Several classifications and performance tests were carried out to estimate the optimal number of bands. The highest classification accuracy using SAM could be obtained using the best combination of 36 bands, i.e. 96% overall accuracy (OAA) with a Kappa of 0.92.

For MLC best results could be achieved using three bands, i.e. 741.23nm, 993.98nm and 1320nm. An overall accuracy of 96.4% and a Kappa of 0.93 using the test dataset showed the lowest classification error of all procedures. More general classification accuracies with high values were calculated using the best combination of 16 bands, i.e. 95.6% OAA with a Kappa of 0.91 for SAM and 92.9% OAA with a Kappa of 0.86 for the maximum likelihood classification. Figure 4.7 gives an overview of band positions of the 6 bands derived from the decision tree analysis compared to the best 6 bands in combination derived from the BD analysis.

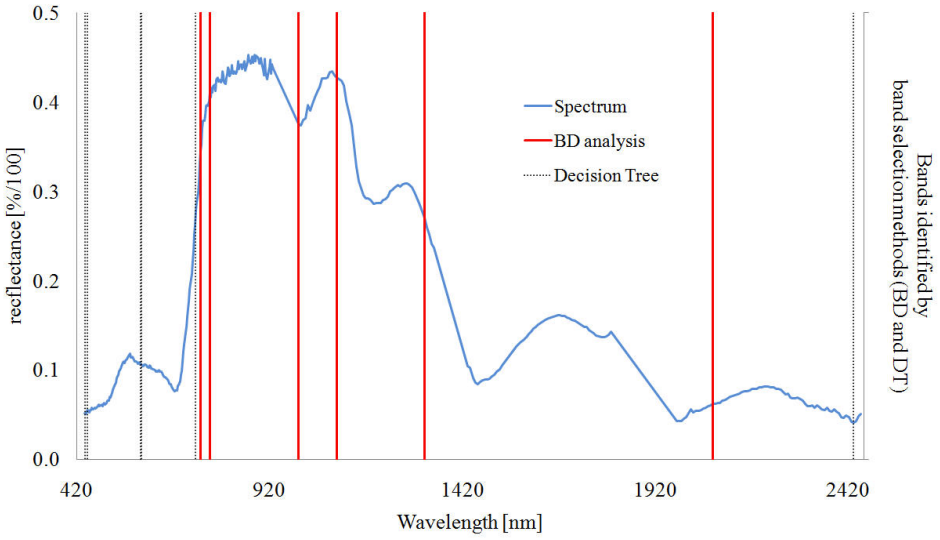


Figure 4.7.: Band Positions of the 6 bands identified by band selection on the AISA 2008 dataset.

4.5. Conclusions

The present study focused on the application of two different airborne hyperspectral imaging systems for the detection of fungal diseases in wheat. A classification with spectral angle mapper was carried out on each dataset using the same ground truth data for classification and validation. The ground truth dataset, built-on visual in-situ disease rating values, was therefore randomly divided into training and test subsets. The random selection was repeated five times to estimate the robustness of the procedure. A mean overall accuracy of 80.33% and a mean Kappa coefficient of 0.61 could be obtained with the ROSIS dataset, a mean OAA of 84.32% and a mean Kappa of 0.69 with the AISA dataset. The results show that the AISA datasets generally outperform the

results obtained with the ROSIS sensor system. This is caused mainly by the higher spectral range of the AISA sensor system as displayed in figure 4.3. But also the spectral resolution may cause the higher accuracies even within the smaller spectral range. The flight height may be crucial as well. The ROSIS data were acquired within an altitude of about 2880m, whereas the AISA data were recorded in an altitude about 2300m. On the one hand, this results in a higher spatial resolution of the AISA sensor system. On the other hand it may also cause less signal intensity available for ROSIS due to higher atmospheric absorption and scattering of the reflected surface signal. A direct comparison of the same spectral range of both sensor system is displayed in figure 4.4. The signal as obtained with the AISA systems has a higher intensity especially within the range of the NIR. The AISA image data was reduced to the same spectral range between 383nm and 839nm, resulting in 173 spectral bands. Five classifications with SAM and randomly selected samples for training and test were conducted and compared to the results obtained with ROSIS. A mean OAA of 79.91% and a mean Kappa of 0.60 were calculated. These values are slightly lower than those obtained with ROSIS, i.e. 80.33% OAA and a Kappa of 0.61. But the standard deviations of 1.75 for the OAA and 0.03 for the Kappa are significantly lower. This shows that the classification using SAM on the AISA seems to be more robust. Since the main objective of this chapter was the general detection ability of fungal infection in wheat using airborne hyperspectral data, the results show that airborne spectroscopy data is suitable for this purpose.

Since the results obtained with the AISA datasets generally outperform the results obtained with the ROSIS sensor system, the band selection techniques for finding optimal hyperspectral bands for a binary classification of infected and vital wheat stands were only performed on the AISA data. Two different selection methods were tested accordingly. The Bhattacharyya distance performed better than the applied decision tree estimation algorithm. The BD takes the variance of the data into account. Therefore, the identified bands are spread more widely over the entire spectrum as shown in figure 4.7. The results of SAM and MLC using the bands detected by the Bhattacharyya distance analysis were slightly better than the results of SAM using all 405 bands. This indicates that a reduction of redundant data or rather a focus on relevant bands can enhance both processing time and classification accuracy.

The relatively high disease severities at late growth stages result in absorption features that rather affect the NIR and SWIR region, so lower disease severities would have been desirable. For other growth stages it might be more efficient to choose different bands for a binary discrimination. In addition, a classification into different severity classes

would be an advantage for Precision Agriculture. Both aspects will be the main focuses during the following studies.

5. Reduction of the spectral dimension of hyperspectral data

5.1. Introduction

In the previous chapter 4, the ability and general the potential of feature selection was proven on an AISA dataset (Airborne Imaging Spectroradiometer for Applications, Specim, Oulu, Finland) for the detection of fungal infections in wheat stands. Two different band selection methods, a decision tree estimation and the Bhattacharyya distance with a feature forward search strategy, were successfully applied on hyperspectral image data within this pretest. The Bhattacharyya distance as a separability measure generally showed promising results as relevant bands for a differentiation of healthy and stressed wheat stands could be identified. Since the objectives of chapter 3 and 4 were more focused on the general stress detection ability, this chapter examines the spectral dimension of hyperspectral data for the discrimination of vital and fungal infected wheat stands in detail by both decreasing the spectral resolution and selecting suitable bands.

Hyperspectral sensors have many contiguous bands along the electromagnetic spectrum, mostly between 0.4 and 2.5μm, which is known as full-range, and are thus more sensitive to spectral changes in plant reflectance induced by stress factors. The spectral bands are narrow and contiguous, and the data of adjacent bands are often correlated, depending on the observed phenomenon. This leads to the assumption that a subset of the entire spectrum is needed to observe a phenomenon, hence a removal of redundant information according to the phenomenon-specific characteristics may be valid and useful. Goetz [2009] stated that the modifier 'hyper' means 'too much' and that no single material requires hundreds of spectral bands to be identified uniquely. Several studies have shown that the large number of spectral bands increases the complexity of the analysis and that the classification accuracy may profit by focusing on relevant spectral bands or wavelength areas [De Backer et al., 2005, Serpico and Bruzzone, 2001]. A selection of relevant bands may thus maximize the detection accuracy of infected wheat stands and, for example, more reliable application maps could be derived. Furthermore, a reduction of redundant data by selecting relevant spectral features is important to speed up the process of decision support for agricultural applications and to define the requirements for new sensors [West et al., 2003].

As shown in chapter 4 and for example by Apan et al. [2004], Franke et al. [2008] remote sensing data are suitable to detect crop stress by the use of all hyperspectral bands. However, the objective of the present study is to reduce the spectral complexity

of hyperspectral datasets so that the requirements of the phenomenon of interest are addressed, based on: (1) the detailed analysis of the suitability of feature selection for the detection of fungal infections in wheat, (2) the identification of an optimal number of features for this purpose, and (3) an assessment of the spectral scale at which crop stress induced by a pathogen infection is detectable (by decreasing spectral resolutions). As a result, a subset of hyperspectral data with an optimal number of bands and optimal spectral resolution for a binary classification of 'healthy' and 'infected' wheat stands is validated. An accurate localization of infected wheat within an agricultural field may allow the extraction of more reliable information and thus, for example, the derivation of application maps for effective treatments.

5.2. Study site and data of flight campaign 2005

The study area is located at the same position as the field plot described in chapter 4. In this chapter datasets acquired in 2005 were used for analysis, referring to a different experimental arrangement than in 2008 (see figure 5.1). The chosen field plot with a total size of six hectare and a flat topography was cultivated with a wheat cultivar vulnerable to powdery mildew caused by the pathogen *Blumeria graminis*. The field was divided into three subplots of different sizes (see figure 5.1). Plot 1 was treated with common doses of fungicides adapted to farming standards in this region. On May 11, 2005, at growth stage GS 34, Plot 1 was treated with a fungicide mixture of 'Opus Top' (BASF, Ludwigshafen, Germany) at a dose rate of 0,8l/ha and 'Flexity' (BASF, Ludwigshafen, Germany) at a dose rate of 0.4l/ha. On June 9, 2005, at GS 57, the plot was additionally treated with the fungicide 'Juwel Top' (BASF, Ludwigshafen, Germany) at a dose rate of 1,0l/ha. Plot 2 only received a single treatment on May 11, 2005, to cause a reduced infection. The treatment was analogical to the first one in Plot 1. The largest plot 3 received no fungicide treatment at all, in order to give a potential pathogen infection the possibility to spread in a nearly undisturbed environment. Other treatments such as the use of fertilizers or herbicides were carried out applying conventional doses to reduce the impact of other stress factors. Additional EM38 measurements of the apparent electrical conductivity (ECa) (Geonics Limited, Ontario, Canada) showed only slight variability in the soil condition, i.e. no significant differences in clay content or soil moisture. Thus, it could be assumed that the occurring crop stress was primarily caused by a pathogen infection.

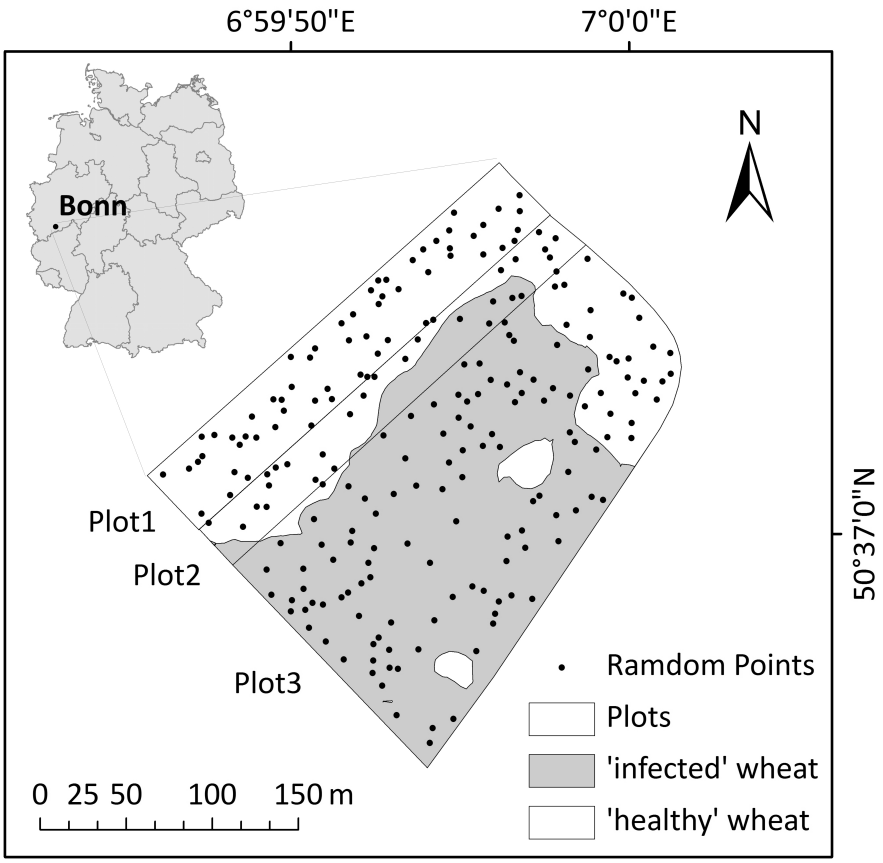


Figure 5.1.: Map of the test site showing the three sub-plots, the random points used for training and the disease map whose pixels were used for validation (based on interpolated severities as observed in field.

On May 28, 2005 (at GS 45), a flight campaign with the airborne Hyperspectral Mapper (HyMap™, HyVista, Australia) was realized by the German Aerospace Center (DLR) as part of the HyEurope campaign. The sensor system is a whisk-broom scanner with an ax-head double mirror and the collected data consist of 126 bands over the spectral range between 0.4 and 2.45μm. Depending on the wavelengths, bandwidths have full widths at half maximum (FWHM) between 0.013μm and 0.020μm. The flight height of approximately 2000m above ground level resulted in a nominal spatial resolution IFOV of 4m*4m per pixel. The hyperspectral image was radiometrically calibrated and atmospherically corrected by the data providers (DLR, Germany).

Additional ground truth data were collected at 54 sample points in the field at the day of the overflight with DGPS readings. The disease severity (i.e. the percentage of infected leaf area) and growth stage were rated by visual assessment. The powdery mildew causing pathogen *Blumeria graminis* appeared in the investigated wheat field and caused disease severities between 0% and 50% of infected leaf area on the same day of the HyMap™ campaign. Based on these observed severities, the binary disease map as shown in figure 5.1 was generated by categorizing the interpolated values. The growth stage of BBCH 45 was equal over the entire field and it can be expected that it did not cause any heterogeneities in the vegetation status.

An independent hyperspectral dataset was used for evaluation purposes to test the validness of the selected features. A comparable experiment was conducted in 2009. On June 16, 2009, the same agricultural area was covered by the Airborne Imaging Spectroradiometer for Applications (AISA Dual, Specim, Oulu, Finland). Disease occurrence with severities up to 50% was estimated at 96 sample points in the field. The ground truth data was divided into a 38 points dataset for training and a 58 points dataset for validation. A detailed description of the experimental setup can be found in chapter 6.

5.3. Methods - data reduction and evaluation

First, HyMap™ bands with lower signal-to-noise ratio, particularly in the blue range and around the water absorption bands around 1.4μm and 1.9μm were omitted, resulting in a reduced dataset with 109 bands. To extract sufficient samples for training and validation, the observed disease severities as observed at 54 sample points in field were interpolated

using ordinary Kriging. Afterwards, the interpolated severities were categorized into the binary classes 'healthy', which enfolds disease severities below five percent, and 'infected', which enfolds areas with severities of five percent or higher. This results in 4243 pixel for the class 'healthy wheat' and 5432 pixel for 'infected wheat'. For training purposes 110 randomly distributed points per class (220 in total) were selected and used as training data for the feature selection and for the classification (see figure 5.1). The remaining points, i.e. 4133 pixel for 'healthy' and 5322 pixel for 'infected', were used to generate the test dataset. Thus, the in-situ observations are partly used for generating the training data and partly used for validation purposes.

5.3.1. Feature selection

In the present study a distance measure, i.e. the Bhattacharyya distance (BD) was tested for its suitability to identify bands within the feature space to define an appropriate data base for classification. It is not a feature selection algorithm itself, but in combination with an adequate search algorithm it can be used to evaluate the reliability of selected features. The BD has been frequently used with remote sensing data and showed to be a useful measure of separability between two classes [Jimenez and Landgrebe, 1999, Herold et al., 2003, Miao et al., 2007]. Since the BD as shown in chapter 4 seems to be superior to other feature selection approaches like the decision tree analysis, this chapter is focused on the BD analysis for feature selection. If two classes are given, the BD can be calculated to quantify the difference between certain features. The BD can be used as a band selection method by changing the number of input variables and selecting stepwise different band combinations. In a first step, the distance between two classes was calculated for every single feature. The feature or band with the highest separability value was then selected. There are two possibilities for further band selection: either using 1) a step-wise search or 2) an all-possibility search by calculating every combination. The step-wise approach, also called feature forward search strategy (FFS) iteratively searches for a feature which produces the highest distance value in combination with one or more pre-selected features. This feature or band is then selected before the next adequate feature is evaluated in combination. In contrast to an all-possibility or exhaustive search, the number of search steps can thus be significantly reduced, but the feature combination may not be absolutely optimal. Calculating all possible combinations for finding the optimal feature subset requires high computational time and is thus restricted to a maximum number of search combinations determined by computational limitations. However, both possibilities were tested up to these limitations and

the selected features were used to reduce or sub-divide the hyperspectral data.

As an additional feature selection method, a wrapper approach based on the spectral angle mapper (SAM) with a feature forward search strategy was tested. The results were compared to the results obtained by the comparable classification based features as selected by the BD analysis.

5.3.2. Classification of the data subsets

In the present study two supervised classification methods particularly suitable for hyperspectral data were applied, i.e. the Spectral Angle Mapper [Kruse et al., 1993] and the Support Vector Machines [Vapnick, 1998]. The Spectral Angle Mapper is widely used for various applications by the remote sensing community. It is based on a measure of similarity, which calculates the angle between two spectra (described as vectors) by processing the vector direction. At least two bands or features are needed to calculate the angle between the spectra. The number of bands is not limited, thus the spectral angle can easily be calculated using n-dimensional datasets. Reference spectra, which are compared to pixel spectra, may base on measurements taken in the field or in the laboratory. They can also be derived from radiative transfer models or other images, as well as directly from the current image. A great advantage of SAM is that the length of a spectrum vector can be disregarded and the algorithm is therefore insensitive to gain factors (illumination) [Kruse et al., 1993]. An unknown pixel spectrum is classified to a predefined class if the angle between its vector and a class vector is equal or smaller than a defined threshold size. For this study all classifications based on SAM were conducted using a threshold of '0.1'.

Support Vector Machines (SVM) proposed by Vapnick [1998] have been introduced into the machine learning community in the late 90's as a binary classification algorithm and have been widely used for image classification purposes in the past few years (e.g., Waske et al. [2009], Dalponte et al. [2009]). Major benefits of this classifier are its high generalization ability, the ability to model complex, non-linear boundaries - by using kernel functions - and the requirement of a comparatively small training dataset [Van der Linden et al., 2009]. To solve a binary classification problem, an optimal linear separating hyperplane is constructed on the basis of the training data to discriminate two classes. It is the aim to maximize the margins between the hyperplane and the training samples that define the hyperplane, the support vectors, by minimizing the

cost function. In this study the free IDL tool 'imageSVM' by Rabe et al. [2009] based on LIBSVM by Chang and Lin [2001] was used for classification. A grid-search using a 3-fold cross validation determined the best combination of the kernel parameter 'g' and the regularization parameter 'C' which both are used to estimate the function for the SVM classification. Thus, the problem of overfitting was minimized.

5.3.3. Spectral resampling

Franke et al. [2009b] examined the requirements of the spectral resolution in remote sensing data for crop stress detection by applying spectral mixture analyses to stepwise resampled hyperspectral image data. The results have shown that the high spectral resolution of HyMap™ is needless to estimate disease severities satisfactorily. In the present study the interdependencies of selected features and spectral resampling were examined. To analyze the effect of decreasing spectral resolution on the classification accuracy, in this case the accuracy of the detection of fungal disease in wheat, the HyMap™ data with 109 bands was spectrally resampled by stepwise merging adjacent bands, i.e. summing up their FWHM. The FWHM of the original HyMap™ data slightly differs over the spectral range with values between 0.013μm and 0.020μm and thus differs for each resampled dataset (figure 5.2).

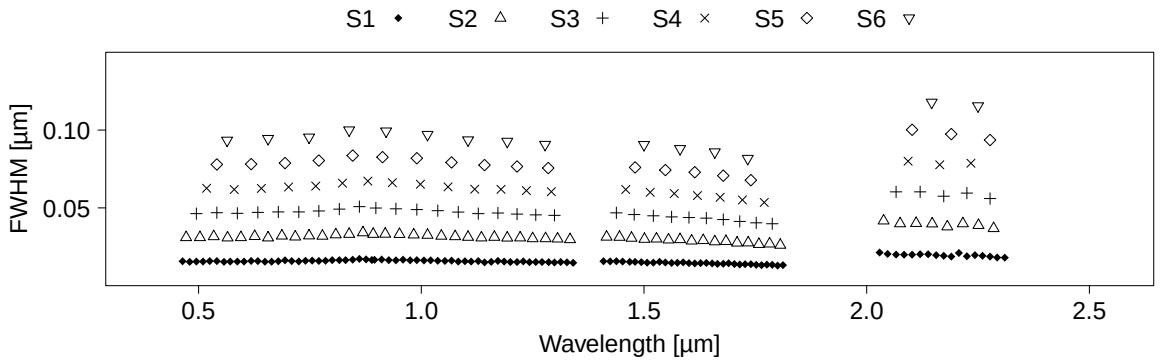


Figure 5.2.: Spectral characteristics of the 6 datasets from S1 (original HyMap™ bands) to S6 with

In addition to the original dataset (labeled as S1), 5 image datasets were thereby generated with 54 (S2), 34 (S3), 24 (S4), 19 (S5) and 15 (S6) spectral bands and a mean FWHM of 0.016μm (S1), 0.033μm (S2), 0.049μm (S3), 0.065μm (S4), 0.081μm (S5) and 0.098μm (S6) respectively. For comparison, six spectra of a typical vegetation surface corresponding to the single steps (S1 to S6), are displayed in figure 5.3. Each dataset

was classified by SAM and SVM. In addition, a feature selection was applied to each dataset and also classified afterwards.

5.3.4. Evaluation of the spectral dimension

For validation purposes, the Kappa coefficient was calculated for every classified dataset. This statistical measure takes the agreement occurring by chance into account, thus it is a more robust validation parameter than the overall accuracy (OAA) for example. Landis and Koch [1977] have proposed that Kappa values between 0.41 and 0.60 can be seen as a moderate agreement and values between 0.61 and 0.80 as a substantial agreement. The Kappa coefficient was calculated on the one hand for the training data to quantify the training error and for direct comparison of both classifiers. On the other hand it was calculated for the validation data, i.e. 9455 pixel based on the interpolated data to prove the classification results on test data.

This study bases on 3 different analyses in order to determine the spectral requirements on hyperspectral remote sensing data for crop stress detection. First, the classification accuracies of SAM and SVM by using 109 HyMap™ bands as well as using selected HyMap™ bands are determined. Secondly, the effect of decreasing spectral resolution on the classification accuracy is evaluated by using spectrally resampled datasets. The third analysis is a combined analysis that defines the potential of selected features of spectrally resampled datasets for crop stress detection. The following section 'result' is structured according to the three analyses.

5.4. Results

5.4.1. Effects of feature selection on the classification accuracy

First, the effect of feature extraction on the classification accuracy was observed by increasing the number of selected HyMap™ bands iteratively. The results of the step-wise band selection approach using the Bhattacharyya distance for evaluation applied to the HyMap™ dataset with 109 bands were therefore analyzed. Thus, the best single band, as well as the best 2-band-combination, the best 3-band-combination and so on, were defined to reduce the hyperspectral data. All subsets were then classified using SAM and SVM. The results were validated by calculating the confusion matrix using either the training data or the test data. The calculation of the training error is particularly

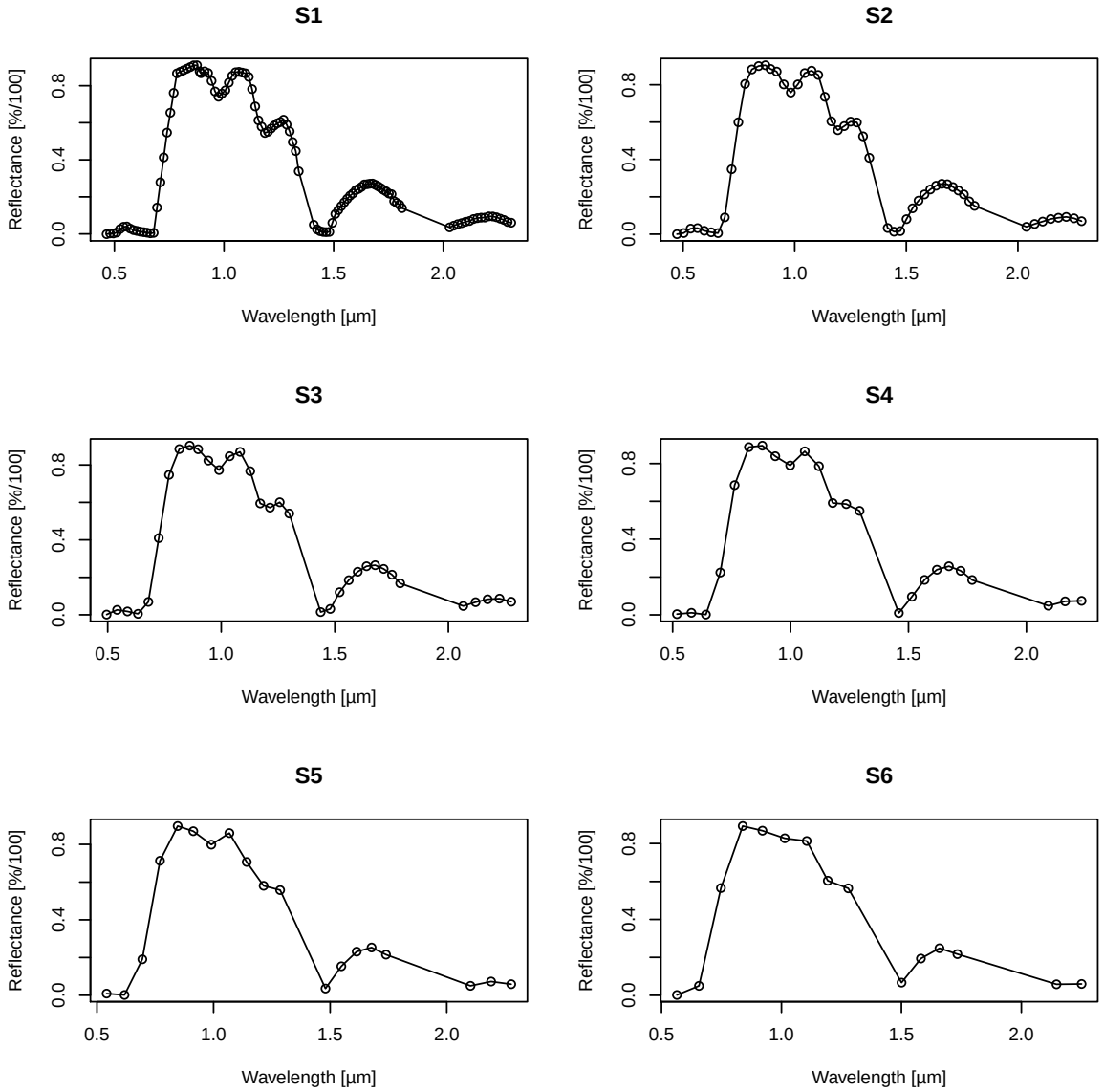


Figure 5.3.: Comparison of the six main data sets showing the resampled spectra (S1 - 109 bands, S2 - 54 bands, S3 - 34 bands, S4 - 24 bands, S5 - 19 bands, S6 - 15 bands).

useful for a direct comparison of the classifier performance. The results are displayed in figure 5.4.

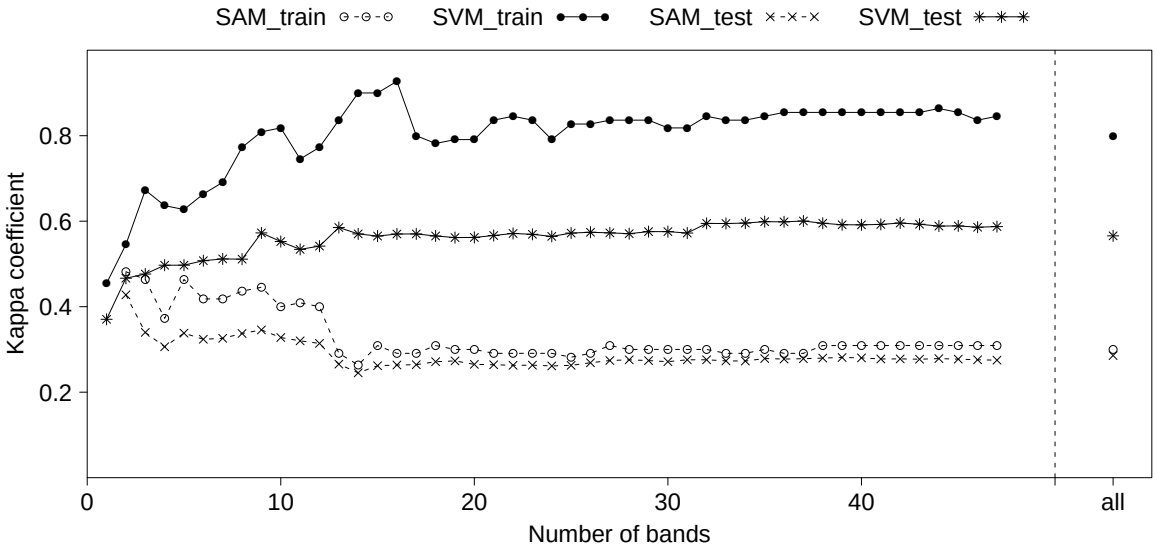


Figure 5.4.: Classification accuracies of SAM and SVM classifications based on band combinations as selected by BD with increasing number of bands. Accuracies are shown for training and test data.

The results of the SVM classifier start with a Kappa value of 0.45 based on only one spectral band. The general trend chart of the accuracies suggests a gain of classification accuracy by the using more spectral information. Regarding the training error, the highest accuracies are produced by using 16 spectral bands in combination with a corresponding Kappa value of 0.93. The 16 band positions displayed in table 5.1 and figure 5.5 are spread over the entire spectrum.

By further increasing the number of spectral bands the values begin to oscillate around a of Kappa 0.83. The values of the accuracy assessment using the test data have a similar trend to the training data with generally lower accuracies. The highest accuracies on the test data could be achieved by a 13-band-combination with a Kappa of 0.59. From 14 bands in combination, the trend of the Kappa coefficient stagnates around a Kappa value of 0.57. The results of SAM almost show the opposite trend. By putting more spectral information into the classification process the accuracy decreases in both cases. The Kappa value of the classification result based on the best 2-band-combination is comparable with the results of SVM, i.e. 0.48 (SAM) to 0.55 (SVM). By involving more

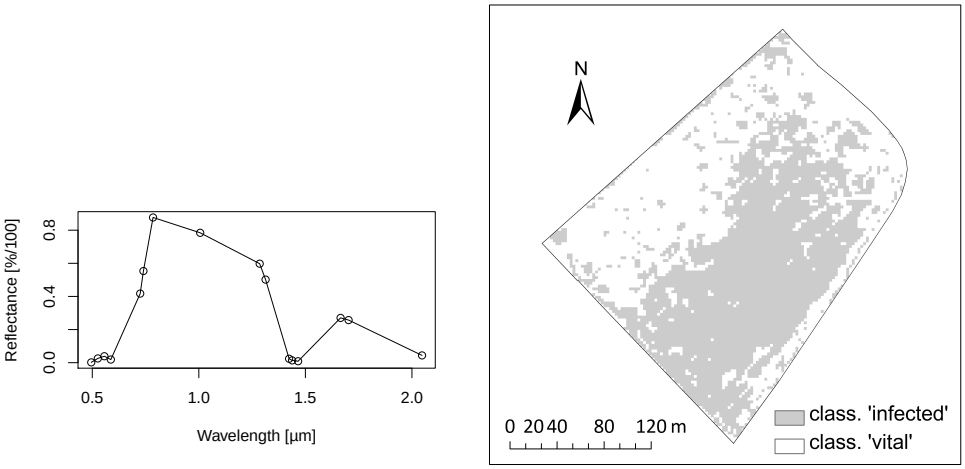


Figure 5.5.: Vegetation spectrum of the selected 16 spectral bands (left) and the SVM classification result based on the 16 selected bands (right) with lowest training error.

spectral bands the accuracy even decreases to a level around Kappa 0.30 from 13 bands in combination. Using all 109 bands, SVM achieved a Kappa of 0.57 and SAM a Kappa of 0.29 regarding the test data.

No.	Band	Wavelength [μm]	No.	Band	Wavelength [μm]
1	65	1.4378	9	58	1.2855
2	11	0.5871	10	85	1.7027
3	21	0.7400	11	5	0.4952
4	7	0.5259	12	9	0.5568
5	64	1.4236	13	39	1.0063
6	100	2.0474	14	24	0.7852
7	20	0.7249	15	60	1.3132
8	67	1.4658	16	82	1.6651

Table 5.1.: Band numbers and wavelength positions of the most significant spectral bands. 13 bands (gray) selected by the step-wise search produced thereby the best result for the test data and 16 spectral bands produced highest accuracies for the training data.

The selected features were additionally evaluated with the independent AISA 2009 dataset. The test site was classified using, first, all 481 bands of the sensor, secondly a dataset reduced to 13 spectral bands and thirdly a 16 bands subset according to the features as identified by using the HyMap™ dataset (see table 5.1). The classification results in a Kappa of 0.71 by using all 481 AISA-DUAL bands. The 13 bands subset

achieved a Kappa of 0.62 and the 16 bands subset a Kappa of 0.79.

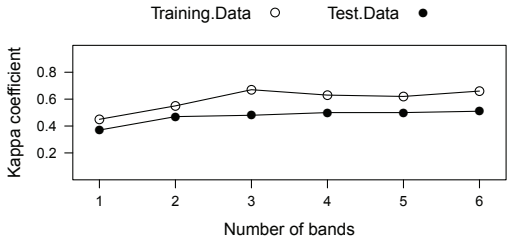
The results of the band selection using the step-wise approach were then compared to the results of the approach using all possible band combinations. Because of computational capacity, this procedure could take into account merely up to 6 bands of the original 109 hyperspectral bands. For the identification of the best 6 out of 109 bands, 2.025.023.364 combinations had to be calculated. The search for the best 7 bands in combination and beyond exceeds the computational limitation and could thus not be realized. Selection steps 1 to 6 were compared to the results of the step-wise search. Figure 5.6 demonstrates the SVM classification results of the first six selection steps of the step-wise search and the approach using all possible band combinations.

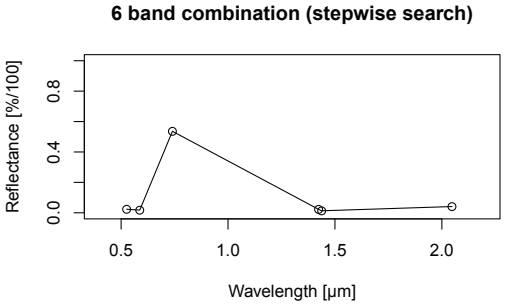
The results show that the Kappa values of the all-possibility approach exceed the results of the stepwise search with a Kappa coefficient of 0.56 compared to 0.51 for the 6-band-combination. Example spectra and their associated band positions for both selections are also displayed in figure 5.6. It can be stated that the band positions of the all-possibility search are spread over the entire spectrum. The first band is positioned in the blue range of the electromagnetic spectrum (0.48μm), the second in the red (0.65μm), followed by two bands in the near infrared (1.07μm and 1.10μm) and two bands in the SWIR (1.52μm and 1.67μm).

Since the resulting Kappa coefficients achieved by SAM are very poor, an additional examination of SAM for crop stress detection was performed. As outlined in chapter 1, band or feature selection can be realized by using either filter or wrapper approaches. In this case it would be of great interest how SAM performs if the feature selection and the resulting feature subsets are specifically adapted to SAM. A wrapper approach using SAM as selection and classification algorithm was therefore implemented into ENVI/IDL 4.7©. Parts or rather the core of the wrapper code and an image of the GUI are given in the first appendix A of this thesis. To examine the performance of SAM with specifically selected bands using the wrapper, the HyMap™ data with the same training dataset was used. For reasons of comparison the performance of the feature subsets were evaluated using the training dataset additionally. The results of the SAM wrapper are significantly higher than those using the features selected with the Bhattacharyya distance. Figure 5.7 shows the results of the SAM wrapper using a feature forward search strategy. Best results with a Kappa coefficient of 0.59 and an overall accuracy of 79.55% for the training

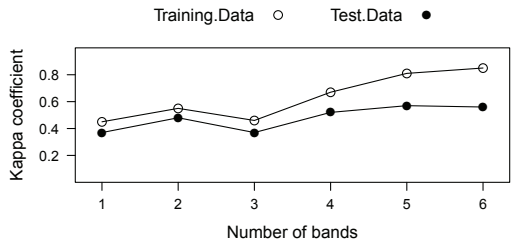
STEP-WISE SEARCH

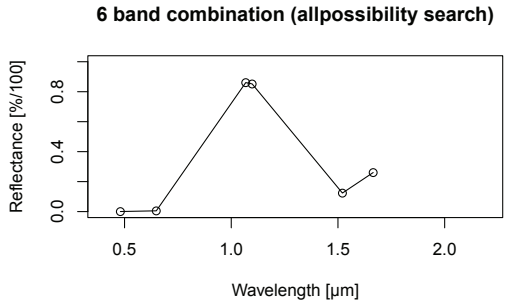
No. of bands	HyMap band No.	Pos. [µm]	SVM Kappa (train. data)	SVM Kappa (test data)
1	65	1.4378	0.45	**0.37**
2	11 65	0.5871 1.4378	0.55	**0.47**
3	11 21 65	0.5871 0.7400 1.4378	0.67	**0.48**
4	7 11 21 65	0.5259 0.5871 0.7400 1.4378	0.63	**0.50**
5	7 11 21 64 65	0.5259 0.5871 0.7400 1.4236 1.4378	0.62	**0.50**
6	7 11 21 64 65 100	0.5259 0.5871 0.7400 1.4236 1.4378 2.0474	0.66	**0.51**





ALL-POSSIBILITY SEARCH





No. of bands	HyMap band No.	Pos. [µm]	SVM Kappa (train. data)	SVM Kappa (test data)
1	65	1.4378	0.45	**0.37**
2	11 68	0.5871 1.4800	0.55	**0.48**
3	35 36 56	0.9431 0.9746 1.2574	0.46	**0.37**
4	7 11 68 85	0.5259 0.5871 1.4800 1.7027	0.67	**0.52**
5	4 16 63 91 104	0.4804 0.6638 1.4093 1.7759 2.1198	0.81	**0.57**
6	4 15 43 45 71 82	0.4804 0.6487 1.0679 1.0979 1.5210 1.6651	0.85	**0.56**

Figure 5.6.: Comparison of the step-wise band selection and the all-possibility (exhaustive) search.

data could be achieved by using only three spectral bands. All three bands are within the spectral range of the NIR between 0.79μm and 1.13μm, i.e. band 38, 45 and 22 from the original 109 bands HyMap™ image ordered by the selection step. Using four spectral bands, i.e. 38, 45, 22 and 14 did not improve the classification accuracy. Except for the best 22-band-combination all other band combinations produce lower accuracies.

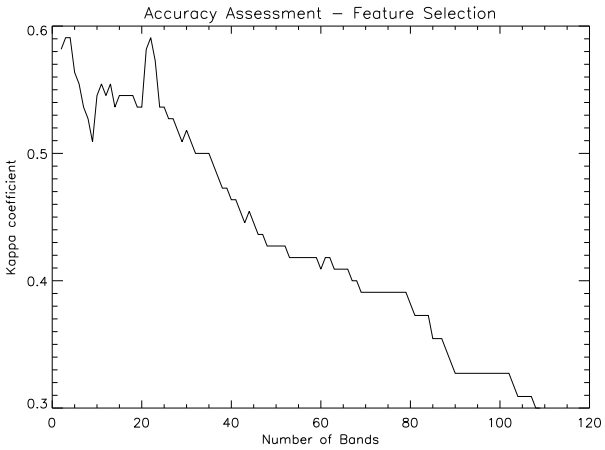


Figure 5.7.: Results of the SAM wrapper approach using a feature forward selection strategy. Graph is showing the Kappa coefficients based on the training data.

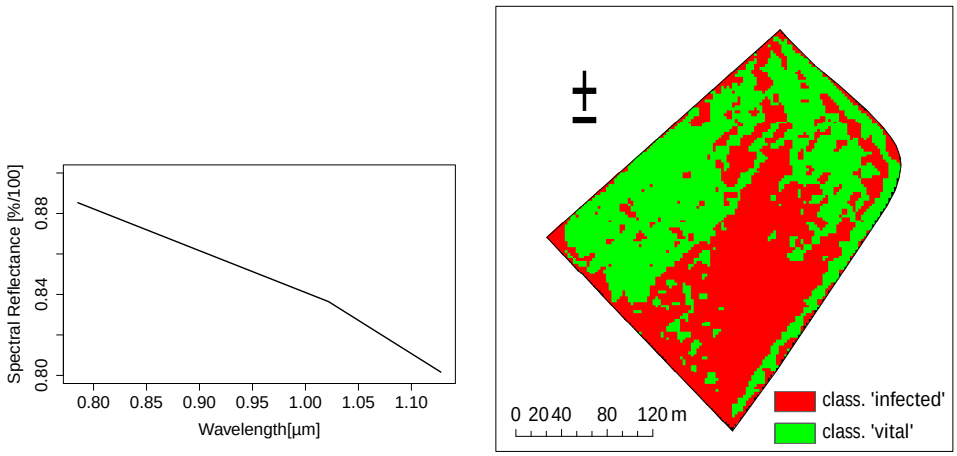


Figure 5.8.: Vegetation spectrum of the 3 spectral bands (left) selected by the SVM wrapper and the corresponding classification result (right).

5.4.2. Effects of decreasing spectral resolution of hyperspectral data on the classification accuracy

To evaluate the role of the spectral resolution in terms of classification accuracies of stress detection, the six datasets S1, S2, S3, S4, S5 and S6 with decreasing spectral resolution were classified by applying SAM and SVM. The accuracies of the classifications of every single step were calculated as shown in figure 5.9. The SVM classifier produced higher Kappa values than SAM for both training and test data. Therefore the following examinations focus primarily on the results of the SVM classifier. In case that all bands were included in the classification process, the detection accuracy showed a Kappa value of 0.57 for dataset S1. The binary classification could be satisfactorily performed by using the training data of the S1, S2, S3 and S4 datasets. Even dataset S4, with spectral resolutions four times lower than the original HyMap™ dataset (S1), showed an appropriate classification accuracy of Kappa 0.55 for the test data. A slight decrease of accuracies from dataset S2/S3 is obvious.

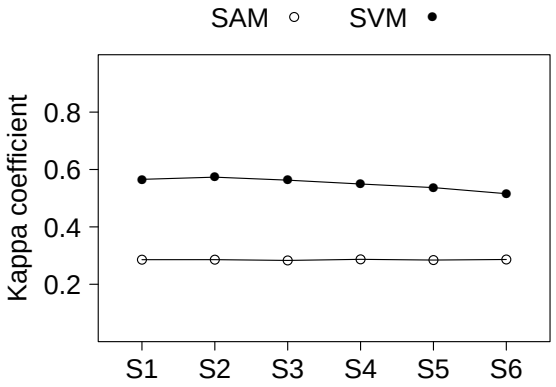


Figure 5.9.: Resulting Kappa coefficients from SAM and SVM classification for each dataset with different spectral resolutions. Kappa coefficients of the test accuracy.

The results of SVM clearly outperform the results obtained with SAM. Using all bands of each dataset, SAM produces comparatively low classification accuracies with no significant changes of the Kappa values.

5.4.3. The role of feature selection with different spectral resolution

This chapter examines the role of feature selection of spectrally resampled bands. The feature selection approach described above was therefore carried out on the six derived datasets S1 to S6. To reduce the high computational volume, only the step-wise search approach was applied and solely the best single band, the best 2 bands, 3 bands, 5 bands, 7 bands, 10 bands and 15 bands in comparison to all bands were used for classification. For an overview, the results are shown in two boxplots in figure 5.10. The boxplots show the highest and the lowest classification accuracies of each dataset, but also the median and the lower and upper quartile. The highest accuracies regarding the training data could be achieved by selecting specific bands of the original HyMap^{TM} dataset (S1). For example, the best 15 bands in combination result in a Kappa value of 0.90 for the training data and in a Kappa of 0.57 for the test data. Datasets S2, S3 and S4 produce similar results, whereas the overall trend for datasets S5 and S6 decreases. Regarding the test data, the highest Kappa value could be produced using seven selected bands from dataset S4, i.e. a Kappa coefficient of 0.61 compared to a Kappa of 0.57 using 15 features of dataset S1 (see table 5.2).

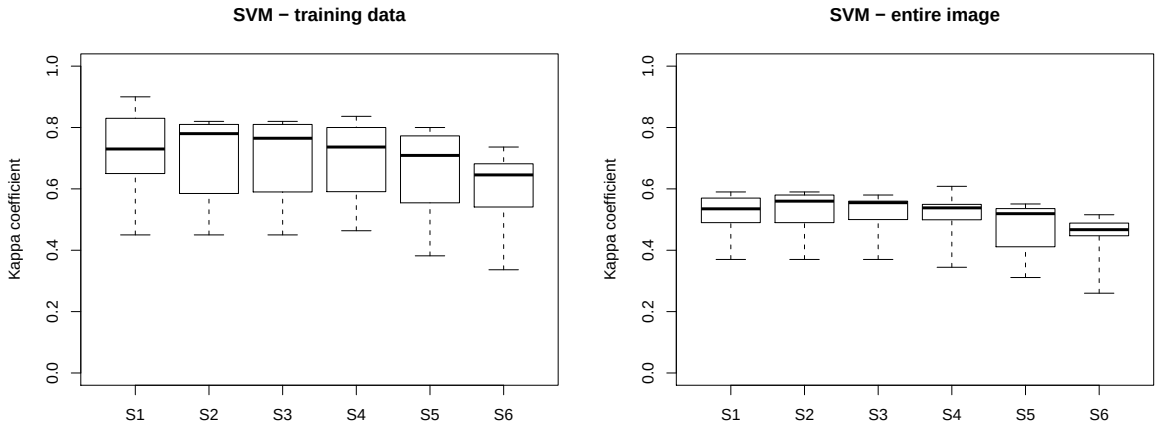


Figure 5.10.: Boxplots showing the trend of Kappa accuracies of both training and test data. Each boxplot consists of accuracy values of eight single band selection steps, i.e. 1, 2, 3, 5, 7, 10, 15 and all bands in combination.

In total, the difference between the accuracies of the six datasets is not very high, even if not all band combinations were compared. The comparison also shows how many bands each dataset needs for achieving the highest classification accuracy. The following table lists the best results for each step.

		Test Data
Data set	Selection	Kappa
S1	15	0.57
S2	7	0.58
S3	7	0.58
S4	7	0.61
S5	7	0.55
S6	all (15)	0.52

Table 5.2.: Highest derived classification accuracies of each single dataset. Best result regarding the test data.

5.5. Discussion

This chapter focused on the interdependencies of feature selection and spectral resolution of hyperspectral image data. It was examined which spectral characteristics are necessary for the detection of pathogen infected wheat stands. For example, De Backer et al. [2005] as well as Serpico and Bruzzone [2001] stated that the accuracy of a classification task may profit by focusing on relevant spectral bands or wavelength ranges. The results of the first analysis validate that the classification, or rather the detectability of wheat stands infected by powdery mildew, profits from a preceding feature selection. Regarding the test data, highest accuracies could be obtained by using 13 spectral bands, i.e. Kappa 0.59 compared to Kappa 0.57 using all 109 HyMap™ bands. The 13 spectral bands are thereby spread over the entire spectrum as obvious in table 5.1. There were selected one band in the blue range of the electromagnetic spectrum (0.495μm), two bands in the green range (0.526μm and 0.557μm), and one band in the red range (0.587μm). In the spectral range of the red edge two bands were identified (0.725μm and 0.740μm), as well as two in the SWIR-I (1.006μm and 1.286μm) and one in the SWIR-II at 1.703μm. In addition, four bands adjacent to wavelength areas around 1.4μm and 1.9μm were selected, that are affected by atmospheric water, i.e. 1.434μm, 1.438μm, 1.466μm and 2.077μm. Since bands with low signal-to-noise ratios in these spectral ranges were already removed during preprocessing, the selected bands are likely to be suitable for describing the health status of a plant particularly because they are also affected by the plant water status. Thus, the atmospheric correction of the data should be as accurate as possible to maximize the information within the spectral ranges mentioned above. However, only a few but specific bands are necessary for a successful differentiation between 'healthy' and 'infected' areas within the wheat plot. Including more spectral information than 16 bands did not yield higher classification accuracies

due to information redundancy. The 16 spectral features which produced the lowest training error and the 13 spectral features which produced the highest classification accuracy using the test data were additionally evaluated using the independent AISA 2009 dataset. By using all 481 AISA bands a Kappa value of 0.71 could be achieved. After reducing the data to 16 spectral bands according to the features as selected for the HyMap™ 2005 data by using the same wavelengths, the classification accuracy could be increased to a Kappa coefficient of 0.79, which is a substantial agreement. On the one hand this underlines the importance of a feature selection for the detection of wheat infected by powdery mildew. On the other hand, the identified spectral features as selected for the HyMap™ 2005 dataset seem to be significant also for other sensor systems since the obatined classification accuracies based on the AISA data could be increased. However, the identified wavelengths need to be validated with different data as well as other cultivars and stress factors to prove their general relevance.

In every single classification process conducted in the present study, the SVM classifier outperforms the SAM classifier. The SVM classifier is able to use only one spectral band for a classification, whereas SAM needs at last two for the calculation of the angle between two samples. In a direct comparison of both classifiers SVM achieves a Kappa of 0.55 by using merely two spectral bands, whereas SAM produces a Kappa of 0.48. The accuracy achieved by SVM increases up to selection step 16 regarding the training error or selection step 13 regarding the test data (see figure 5.4). The results of SAM compared with those of SVM nearly show the opposite trend. By putting more spectral information into the classification process, the classification accuracy decreases. This may result from the decoupled use of feature selection and classification approach. The Bhattacharyya distance is a distance measure, which calculates the separability of two class vectors. Less relevant data concerning the separability between the two classes can thus be filtered. The calculation of the hyperplane of the SVM classifier is able to profit from spectral bands that were selected with distance metrics. However, these selected bands do not increase the spectral angle between the two classes, they rather decrease the angle between the two class vectors. The SAM classifier may thus profit from a coupled feature selection as presented by Keshava [2004] which directly uses the spectral angle values for feature selection. It is thus comparable to recent wrapper approaches, which use the classification accuracy as selection criteria [Waske et al., 2009]. In the present study, SAM seems to necessarily require a feature selection to achieve moderate classification accuracies. If all bands were used for classification the results were signif-

icantly lower, i.e. a Kappa of 0.48 using just two spectral bands compared to a Kappa of 0.29 using all 109 spectral bands was achieved (see figure 5.4).

It could be demonstrated that the band selection approach would benefit from a calculation of all possible band combinations. The total number of band combinations that can be calculated is restricted by computational limitations. A comparison of the step-wise band selection and the all-possibility band selection was performed by using the original HyMap™ bands, but the process was limited to a maximum number of 6 bands in combination. Except for the 2-band-combination, the results demonstrate that the selected bands of the step-wise approach are outperformed by those of the all-possibility approach, i.e. a Kappa value of 0.51 against a Kappa value of 0.56 for the best six bands in combination. Sporadically decreasing Kappa values, like the results produced by the 2- band-combination of the all-possibility approach, are possible since the feature extraction is decoupled from the classification algorithm. It can be assumed that the highest classification accuracy of this study using 16 spectral bands would be exceeded by using the same number of features or fewer, if all possible feature combinations could be calculated.

The spectral bands as selected by the BD analysis were compared to the bands as selected by the SAM wrapper. For the SAM wrapper approach highest accuracies were produced by using only three specifically selected bands, i.e. band 38 at wavelength 1.02μm, band 45 at 1.13μm and band 22 at 0.79μm. A classification accuracy of 79.55% and a Kappa of 0.59 were thereby produced using the same data for training and testing. As described in chapter 1 the NIR wavelength area marks a spectral range with a relatively high reflection factor if vegetation is present. Since the selected wavelengths are all located in the NIR (see figure 5.8), SAM seens to profit from the use of a varying slope of the NIR to classify the spectra into 'healthy' and 'infected' wheat stands. The BD analysis selected three different bands, i.e. band 63 at wavlength 1.44μm, band 11 at 0.59μm and band 21 at 0.74μm. In comparison to the wrapper approach, the use of the three BD bands for a classification with SAM resulted in lower accuracies with an OAA of 73.18% and a Kappa of 0.46. So specifically selected wavelength enhance the classification accuracy. SAM seems to be very sensitive to the amount of spectral bands used as input for the classification. As displayed in figure 5.7, the classification accuracies are decreasing by involving 23 bands and more. This refers to the Hughes phenomenon which describes the degradation of the classifier performance with a fixed sample size

when increasing the number of features or bands [Hughes, 1968]. The sensitivity to this phenomenon depends on the used classifier. Enhanced classification algorithms like SVM are less sensitive. The usage of SVM significantly increased classification accuracies in this case. SVM also outperforms the SAM wrapper approach. An overall accuracy of 83.64% and a Kappa coeffiecient of 0.67 could be obtained by using the SVM classifier on the three bands as selected by the BD analysis.

In the second analysis, the role of the spectral resolution for the detection of infected wheat stands was examined. Six datasets differing in band width were therefore created. The original HyMap™ dataset (S1) with 109 spectral bands was spectrally resampled down to dataset S6 with only 15 bands. Every dataset was classified by using both SAM and SVM. The results show that the influence of the spectral resolution seems to be non-significant for this classification task, even though an 'objective' resampling strategy was used. 'Objective' means that the resampling process did not take into account known spectral features of vegetation such as the green peak or the red edge. Thus, the green peak is not detectable with the datasets S5 and S6 as shown in figure 5.2, since the FWHM exceed the spectral width of the green peak. For S1, S2, S3 and S4 the calculated accuracies using the SVM classifier are higher than or equal to Kappa 0.80 for the training data and 0.55 for the test data. Only the results produced with datasets S5 and S6 are below these values, although the Kappa of 0.74 (training data) and 0.52 (test data) did not drop as much as expected. The trend is comparable to the results of Dalponte et al. [2009]. The outcome of SAM is quite low. Regarding the test data, the Kappa values of all six datasets oscillate around 0.29 with no significant changes. Due to the fact that spectral characteristics of vegetation do not exhibit narrow absorption features and crop stress factors cause changes in broader wavelength areas, spectral bands with lower spectral resolution than the original hyperspectral bands can be used for vegetation analyses. They still allow an identification of changes in vegetation status.

It was shown that the SVM classifier is more suitable for this binary classification and is able to distinguish wheat stands infected with powdery mildew from vital ones using all spectral bands at highest spectral resolution, i.e. 109 bands at S1, but also with a decreased FWHM at S2, S3 and S4 without any significant impairment. The influence of the spectral resolution on the classification seems quite low. This underlines the robustness of the SVM classifier [Dalponte et al., 2009]. However, it has to be mentioned that a spectral resampling does not necessarily mean a loss of information, since resampled

bands represent an integration of spectral information which can be considered, at the very most, as spectral blurring. Another reason may be the simplicity of the classification task. The used binary decision into the classes 'vital' and 'infected' may be a relatively easy task for such a powerful classification algorithm and narrow bands may not be needed in this case.

The aim of this study was to maximize the classification accuracy for a reliable derivation of information regarding the occurrence of diseased wheat in the field. The suitability of the SVM classifier to derive different disease severity classes in wheat should be analyzed further. In the third combined analysis, the feature selection approach was applied to each resampled dataset, in order to compare the suitability of the selected features with different spectral resolution for classification. To reduce the computational effort, the feature selection approach was limited to the best 1-, 2-, 3-, 5-, 7-, 10-, 15-band-combinations compared to all bands per dataset, respectively. Regarding the test data, the highest performance could be achieved using 7 spectral bands of dataset S4 (table 5.2). For the other datasets such high classification accuracies could not be reproduced, but the mean performance of the spectrally resampled data was for S3 higher than for the original data with the highest spectral resolution (figure 5.10). However, the influence of the feature selection seems to be quite more significant than the influence of a decreasing spectral resolution. This can be seen in the range of each single boxplot in figure 5.10 which is significantly higher than the trend in the level of the boxes due to the decreasing resolution.

As listed in table 5.2, a 7-band-combination produced the highest accuracies for nearly all datasets. Figure 5.11 shows the corresponding band positions of all 7-band-combinations. The results show that it is useful to select spectral information at specific wavelengths for satisfactory information extraction. In all cases bands in the visual part of the spectrum, in the red edge area and in the SWIR, were selected whereas in some wavelengths no bands were selected (figure 5.11). Thus, a specific combination of bands within selected spectral ranges seems preferable for the discrimination of 'healthy' and 'infected' wheat plants.

It was proven that the Bhattacharyya distance is an adequate measure for feature selection and identifies the suitable bands for this classification task, especially if the highest spectral resolution is given. The higher the number of spectral bands, the more useful is a feature selection. Relevant information can be filtered and thus, maybe more

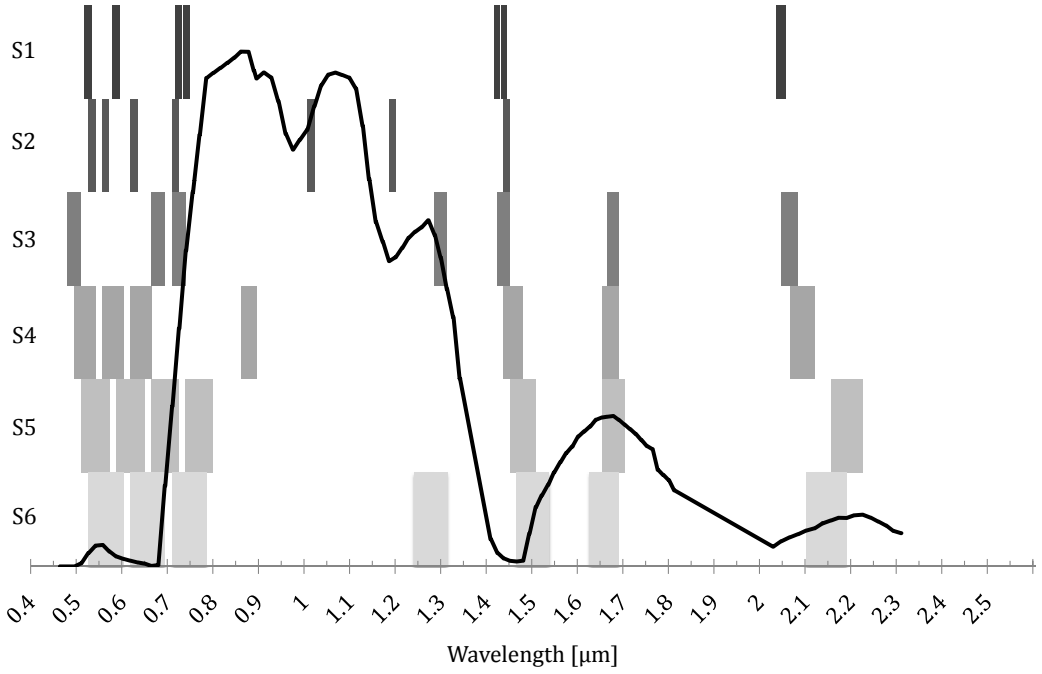


Figure 5.11.: Band positions and FWHM of the best 7 bands of each single dataset.

important, masses of not relevant information can be omitted. Amongst others, reliable information extraction of remotely sensed images strongly depends on a high signal-to-noise ratio, which is directly linked with the spectral resolution [Kruse et al., 2003]. Hence, the exposed results can be of high relevance for future sensor designs and/or data product definitions.

5.6. Conclusion

In the present study, the role of feature selection and decreasing spectral resolution on the classification accuracy of healthy and infected wheat plants was examined. Plant stress was induced by a pathogen infestation with powdery mildew. The disease severity was visually rated and separated into the binary classes 'healthy' and 'infected' at 54 sampling points within the field. This study underlines the important role of the spectral characteristics of remote sensing data for vegetation analyses. The results show that only a few but specific spectral bands are necessary for a successful identification of wheat stands infected by powdery mildew. However, other spectral phenomena or other stress factors may require other spectral bands to remain comparable to the use of

the entire spectrum. The influence of the spectral resolution on discrimination accuracy was also examined by iteratively decreasing the full width at half maximum (FWHM) of HyMap™ bands. It could be demonstrated that a feature selection approach applied to each dataset with decreasing spectral resolution has a stronger influence on the classification accuracy than the decreasing FWHM itself. By the use of only 7 selected bands with FWHM four times lower than the original HyMap™ data accurate crop stress detection was still possible. In general, the results show on the one hand the high potential of hyperspectral data for the identification of diseased wheat stands and thus the possibility to produce reliable information for application maps. On the other hand the influence of a feature selection on the classification accuracy is underlined. Most reliable information could be extracted by focusing on specific spectral bands of the original HyMap™ dataset. The derived results are of high importance for future data analysis and sensor design with a focus on information extraction for agricultural questionnaires.

6. Derivation of disease severities from hyperspectral data

6.1. Introduction

The previous chapters focused on the general detectability of cropstress in wheat and the improvement of this detectability by data reduction techniques for a binary classification into 'vital' and 'infected' wheat stands. It was shown that a determination of crops infected by fungal diseases and thus a classification into the binary classes 'healthy' and 'infected' is possible with a high certainty using hyperspectral data.

The potential of imaging spectroscopy data in deriving for example several disease severity classes or continual disease severity data was not proven yet. On the one hand, agricultural practice may profit from more detailed information about the within field heterogeneities of the disease severity since doses of applied agrochemicals can be adjusted and treatments may thus be realized in a site-specific way. On the other hand, plant breeding practice would profit from a fast and precise phenotyping approach. An automatic and non-destructive spectral sensing of crop stands may therefore provide an innovative technique for greenhouse as well for field applications.

In this chapter the potential of hyperspectral data for the derivation of disease severities in different wheat stands will be examined. A classification approach was tested to estimate the amount of pixel at canopy level classified as 'infected' using a near-range imaging spectrometer in the greenhouse. Plant breeders estimate the vulnerability of new crop varieties to fungal infections by visual rating. An automatic sensing approach at near-range may therefore be of great profit due to its objectiveness and gain of time.

For airborne data, different approaches are required since one single pixel comprises a spectral mixture of several wheat plants with possibly both 'vital' and 'infected' parts. A regression approach using support vector machines for regression (SVR) was therefore applied to hyperspectral image data in order to test its ability to derive continual severity data of fungal infections in wheat. A quantification of real in-field disease severity may thus be realized. If such results would be transferred into application maps, site-specific fungicide treatments with varying doses would be theoretical possible.

6.2. Data - greenhouse and flight campaign 2009

Two different hyperspectral datasets were used within this study. To estimate disease severities at near-range the hyperspectral imaging spectrometer 'ImSpec V10E', as described in chapter 1, was used to acquire spectral reflectance data of wheat stands. For the assessment of disease severities by remote sensing, the airborne sensor AISA-DUAL was used.

6.2.1. Near-range spectroscopy

On January 16, 2009, three pots (P01, P02, P03), 23cm*23cm*20cm in size filled with 9 liters substratum mixture, were cultivated with a wheat cultivar vulnerable to the disease powdery mildew in the greenhouse. On Feburary 24, 2009, at a mean growth height of 30cm, all pots were inoculated with *Blumeria graminis*. Spectral data of the wheat stands were acquired with the hyperspectral imaging sensor ImSpec V10E at a constant height of 152cm above ground. The daily measurements of the four pots started on February 24, 2009, and ended on March 12, 2009, resulting in 17 spectral measurements per plot over the period of time. One additional measurement was conducted on February 17, 2009, to gather spectral data from wheat stands with high disease severities. The sensor acquired spectral data in approximately 126cm above ground resulting an a spatial resolution about 0.41mm per pixel (see chapter 1 for the equation). The illumination was kept constantly using six 'ASD Prolamps' at fixed positions. Disease severity was visually rated between 0 and 100 percent of infected leaf area compared to non-infected leaf area.

6.2.2. Airborne spectroscopy

The study area of the field experiment is located at the same position as the field plot described in chapter 4. For this chapter, datasets acquired in 2009 were used for analysis with a different experimental arrangement than in 2005 or 2008 (figure 6.1). The chosen field plot, cultivated with a wheat cultivar vulnerable to leaf rust, had a size of 4 hectare with flat topography. The field was divided into 6 subplots with 30m*90m up to 30m*270m in size, alternating with no fungicide application and common fungicide application to cause a variety in severities (figure 6.1). The occurrence of other stress factors such as nitrogen deficiency or weed infestation was suppressed by appropriate uniform treatments. Disease severity was estimated at 96 sample points in the field,

whereby DGPS readings were collected at each sample point.

Puccinia recondita causing Leaf Rust occurred in the investigated field with severities up to 50% of infected leaf area. On June 17, 2009, the area was covered by the Airborne Imaging Spectroradiometer for Applications (AISA Dual, Specim, Oulu, Finland), which is described in chapter 1. The data, consisting of 481 spectral bands per pixel, were collected at a flight altitude of 1300m, resulting in a spatial resolution of 1.5m. After radiometric and geometric calibration using CaliGeo 4.9.7 (Specim, Oulu, Finland), as described in chapter 1, the data were atmospherically corrected using ACORN (Atmospheric Correction Now, ImSpec LLC, California, USA). An overview of the test site and the arrangement of the subplots is given in figure 6.1.

6.3. Methods for disease quantification

6.3.1. Counting affected areas

The hyperspectral data acquired by the 'ImSpec V10E' sensor were transformed into reflectance data as described in chapter 1. Since no symptoms could be visually identified before dai-6 (day after inoculation 6) only the spectral measurements from dai-6 to dai-17 and those of the additional high severity day were used for analysis, resulting in 13 images per plot. Every image was further preprocessed by focusing exclusively on pixel that are within the pot area of approximately 23*23cm. Pixel outside this area were masked and set to zero as shown in figure 6.2.

Support vector machines for classification (SVM) was tested for its ability to differentiate pixel representing either parts infected with powdery mildew or non-infected leaf areas. A dataset to train the classifier is obligatory needed. Therefore, pixel values from visually detectable infected parts, healthy parts and soil parts were located to extract their spectra. The pixel values of the class 'infected' were derived from the center pixel of a visually identifiable mycelium of powdery mildew. For every class, i.e. 'infected', 'vital' and 'soil', a training dataset of 200 pixel was thereby generated. Their mean spectrum is plotted in figure 6.3 on the left.

Depending on the classification algorithm, the training dataset must contain one or more spectral samples per class. The spectral angle mapper (SAM) for example needs only one spectrum per class and compares unknown pixel spectra with this reference spectrum, which can be a mean spectrum of several samples. Other algorithms like the

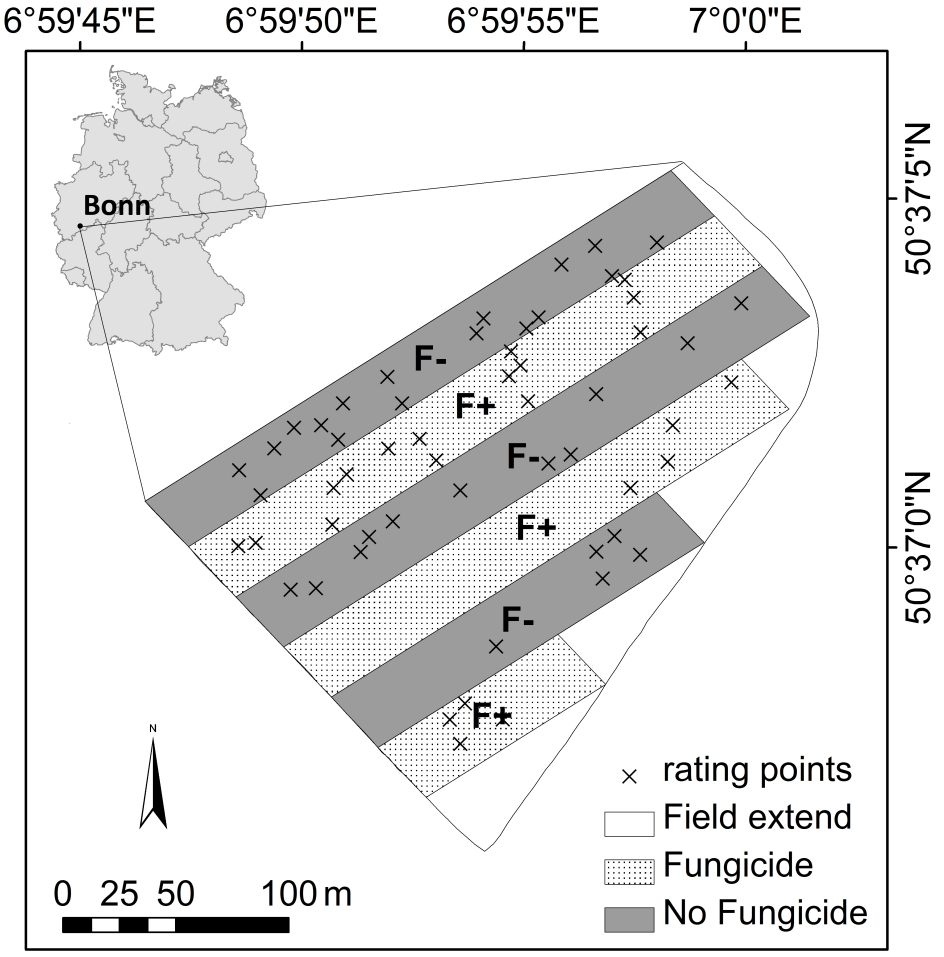


Figure 6.1.: Map of the study site displays the different subplots divided into common Fungicide application (F+) and no fungicide application (F−), as well as the location of the disease rating points.

Figure 6.2.: Masking image pixel outside the pot for reasons of comparison; Wheat stand P02 at dai-6.

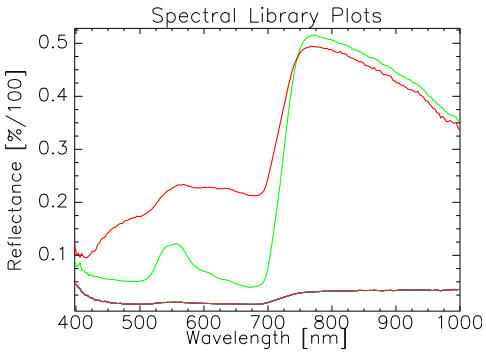


Figure 6.3.: Mean spectrum of the three classes 'infected' (red), 'vital' (green) and 'soil' (brown) as derived from one hyperspectral image (left); Position of the training samples appended to one masked sample image (right).

maximum likelihood classifier (MLC) need at least as much spectral samples per class as there are spectral bands to generate the required covariance matrix. In this study, SVM for classification was examined with regards to its usage in deriving reliable disease severity values from hyperspectral images. To ensure that every classification is based on the same training dataset, the generated training dataset or rather its pixel values were appended to every image and used like a spectral library. A standard image of 500*400*210 (columns, rows, bands) was therefore generated and the spectral library containing 200 pixel per class as well as the image data of every sample was implemented respectively (see figure 6.3 (right)), resulting in 3*13 images with the same extend and the same 600 reference pixel at the top left. The images were then classified using support vector machines for classification (SVM) with a Gaussian kernel for non-linear discrimination problems and a gamma of 0.001, which was found to be adequate in this case. The pixel classified as infected were afterwards counted and compared to the values of the visual rating in percent.

Since chapter 4 and 5 showed that a data reduction to relevant bands can improve the discrimination accuracy, the near-range spectroscopy data were spectrally sub-divided to those wavelength identified as most relevant for the detection of fungal infections in chapter 5. The 'ImSpec V10E' used for this study is a half-range sensor, so only bands with a center wavelength of less than 1000nm could be used. Table 6.1 lists the seven bands from the identified 16 spectral bands (see 5.1, chapter 5) to reduce the hyperspectral 'ImSpec V10E' data. The SVM with the same parameters as described above was then applied to the reduced datasets to classify the images into 'infected' and 'vital' parts.

No.	Band	Wavelength [μm]	No.	Band	Wavelength [μm]
1	65	1.4378	9	58	1.2855
2	11	0.5871	10	85	1.7027
3	21	0.7400	11	5	0.4952
4	7	0.5259	12	9	0.5568
5	64	1.4236	13	39	1.0063
6	100	2.0474	14	24	0.7852
7	20	0.7249	15	60	1.3132
8	67	1.4658	16	82	1.6651

Table 6.1.: Band numbers and wavelength positions of the best 16 spectral bands as identified in chapter 5. The bands used to sub-divide the near-range data are printed in gray.

6.3.2. Support Vector Machines for regression

Support Vector Machines for Regression (SVR) has been shown to be superior to other estimation approaches. As Support Vector Machines (SVM) for classification, the approach seems interesting for remote sensing applications, due to its ability to handle complex, non-linear dependencies in high dimensional feature spaces. As SVM, SVR is a universal learning machine for solving binary classification problems and can be considered as an approximate implementation of Vapnik's Structural Risk Minimization principle, which has been shown to be superior to traditional Empirical Risk Minimization principle [Chang and Lin, 2001]. Let, for the regression problem in a dimensional feature space,

$$x_i \in \Re^d, i = 1, 2, ..., L \tag{6.1}$$

, be a training set of L samples with their corresponding outputs $y_i \in \Re_1$ (i.e., infection rate). The aim is to estimate a function $f : \Re^d \rightarrow \Re_1$ using the training x_i vector and y_i as output. SVR estimate a linear dependency by fitting an optimal approximating hyperplane to the training samples in a multidimensional feature space. In cases of inestimable linearity, the input data are mapped into a high dimensional space, which enables the fitting of a more adequate linear hyperplane. A widely used kernel in remote sensing applications is the Gaussian radial basis function kernel

$$K(x, x_i) = exp(-\gamma \left| x - x_i \right| 2) \tag{6.2}$$

K(x,xi) = exp(-g—x-xi—2), with γ as width of the Gaussian kernel function. A regularization parameter 'C' is used to handle difficult training samples. Effectively, it controls the shape of the solution and thus affects the generalization capability of the SVR. For classification problems both parameters, 'γ' and 'C', depend on the data set and are consequently case specific. Thus, the training process requires the estimation of the kernel parameter γ and the regularization parameter C. In literature, several approaches for an automatic model selection have been introduced, which employ a leave-one-out cross-validation procedure [Chapelle, 2002, Chung et al., 2003].

Due to the high spectral sample frequency of the chosen AISA configuration, several bands were affected by atmospheric influences or achieved a merely low signal-to-noise ratio. Thus, the data had to be reduced to 343 spectral bands. The sample points were rasterized by locating their DGPS positions within the image and the rated severity

values were assigned to each identified pixel. The training of the SVR with Gaussian kernel and the generation of the map containing the estimated values was performed using imageSVM (software available at http://www.hu-geomatics.de) [Van der Linden et al., 2009]. ImageSVM is a freely available ENVI/IDL 4.7© implementation that uses the LIBSVM approach by Chang and Lin [2001] for the training of the SVM. The kernel parameters of 'C' and 'γ' are determined by a grid-search using a x-fold cross validation. Possible combinations of 'C' and 'γ' are tested in a user-defined range and the best combinations for 'γ' and 'C' were selected based on 3-fold cross validation.
The main difference between the two approaches presented here is that the classification applied on the near-range data discriminates unknown pixel spectra into discrete classes, i.e. 'vital', 'infected' or 'soil', and the amount of pixel identified as 'infected' was counted afterwards to assess the disease severity of one entire sample. The regression approach instead assigns a continual value to each pixel of the image scene, which can literally be seen as one sample in the near-range experiment. These values represent the amount of diseased wheat plant compared to healthy wheat plants.

6.4. Results

6.4.1. Disease quantification at near-range

The disease severities of the three different samples differed significantly over the period of time. Highest severities were estimated for P03 with a infection rate of 70%, whereas P01 was estimated with 35% disease severity. All values as visually rated are displayed in figure 6.4. The total amount of pixel classified as 'infected' was compared to the visual rating of disease severity for each image respectively. The highest correlation value could be obtained for the sample P02 with r=0.94. A correlation of r=0.89 was calculated for P03 and r=0.56 for P01. The resulting classification image of dai-17 of the sample P03 with a visually rated disease severity of 27% and 2877 pixel classified as 'infected' is exemplarily displayed in figure 6.5. A correlation analysis based on all estimated values, i.e. the results achieved on P01, P02 and P03, resulted in r=0.82.
The classification applied on the 7-bands-images resulted in a correlation coefficient of r=0.77 for P01, r=0.89 for P02 and r=0.94 for P03. A correlation coefficient of r=0.88 could be obtained using all results of the 7-bands-images in the analysis. All results are listed in table 6.2.

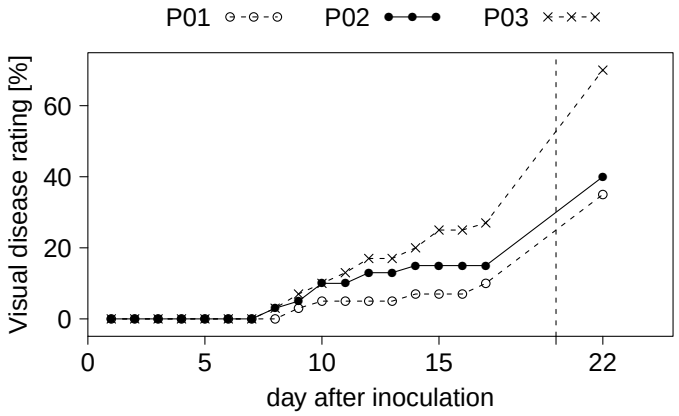


Figure 6.4.: Development of disease severities as visually estimated and rated in percent.

	all-bands-images	7-bands-images
P01	0.56	0.77
P02	0.94	0.89
P03	0.89	0.94
all 3 samples	0.82	0.88

Table 6.2.: Results of the correlation analysis based on the amount of pixel classified as 'infected' by SVM classification compared to visual rated disease severity.

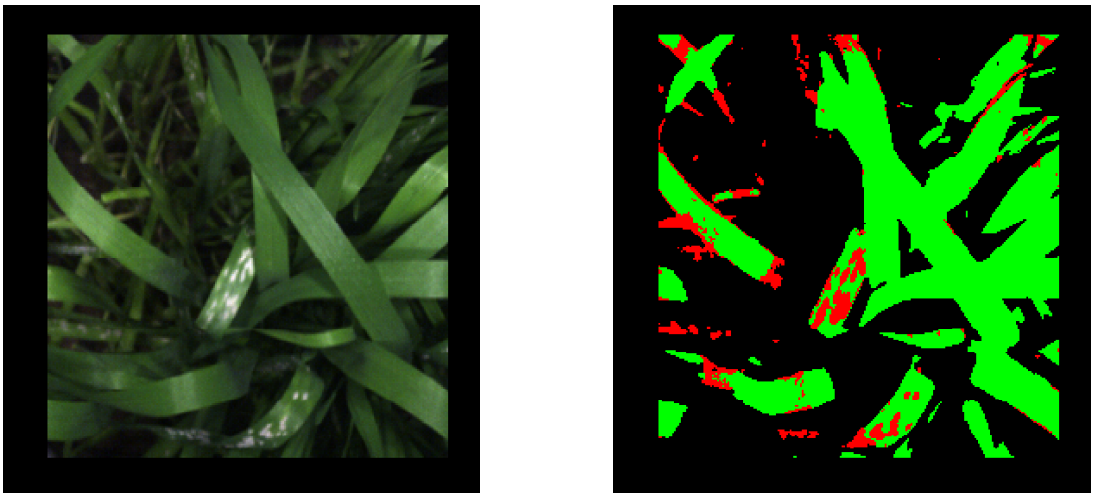


Figure 6.5.: RGB (left) and results of the binary classification applied on P03 at dai17 using SVM with 'infected' leaf parts (red) and 'vital' leaf parts (green) based on the entire spectrum (right). A disease severity of 27% was estimated by the visual rating.

6.4.2. Airborne remote sensing for disease quantification

To test the general performance of the SVR, all 96 reference points that are within the field plot were used to train the support vectors. In this case all reference points were also used for validation. An accuracy of 0.61 was thereby produced. Some pixel spectra at the rating points are affected by structural differences caused by tractor lines resulting in lower reflectance values. To avoid pixel spectra mixed with those spectral influences, a 3m buffer was applied to the vector of each tractor line. Only rating points beyond this 3m buffer were used as reference points. Thus, the total amount of reference points was reduced from 96 to 56. By using only 56 points of the reference data, an accuracy of r=0.86 for the training data could be obtained (see figure 6.6).

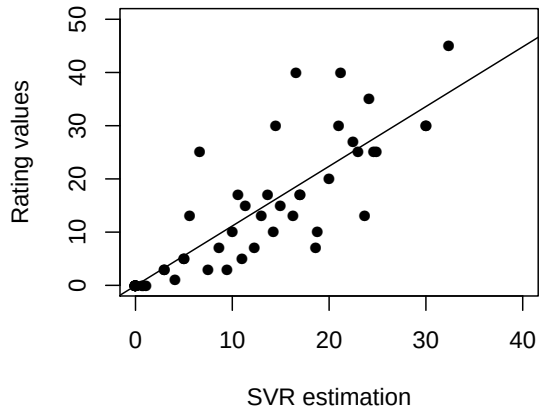


Figure 6.6.: Regression analysis of the disease estimation with SVR and the ground truth disease rating values (r=0.86, r^2=0.74).

The mapped result of the regression using support vector machines can be seen in figure 6.7. On the left side, a grayscale image of the disease estimation with continual values is displayed. Areas which are strongly infected with leaf rust are obvious, as well as 'healthy' areas and tractor lines. The right figure shows the estimation values manually classified into five classes, ranging from severities between 5 and 10 percent disease severity, to strongly infected wheat stands with values above 30 percent. To evaluate the general performance, the reference dataset was divided into training and test subsets. Two different percentages were used to build up the training data, i.e. 30 percent and 50 percent. Each definition of the training dataset was repeated five times using a random sampling strategy resulting in ten datasets with differing training and

test data substance. The performance of each training dataset was examined using the according test data. Using 30 percent of the reference data as training data, a mean accuracy of r=0.97 and a mean correlation coefficient of r=0.62 for the test data was obtained. If 50 percent were used for training, a mean accuracy of r=0.83 for the training data and a mean correlation of r=0.69 for the test data were achieved. The results of each step are shown in table 6.3.

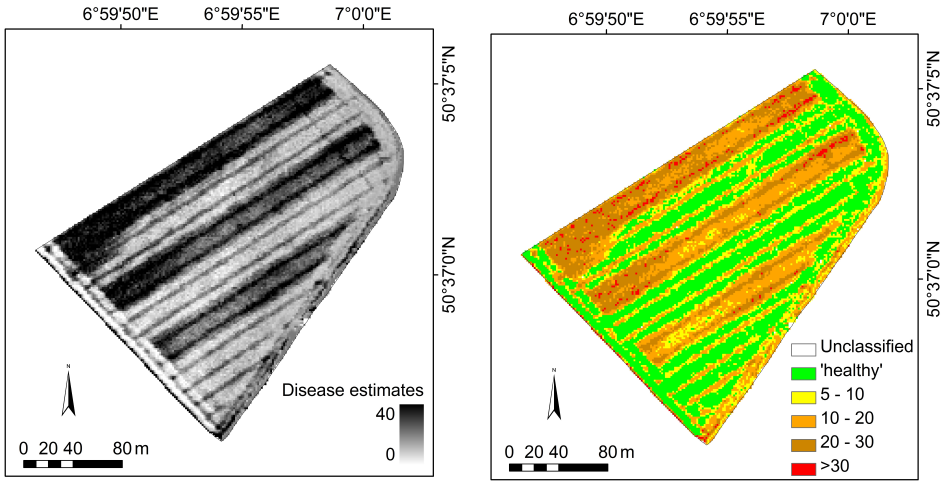


Figure 6.7.: Map of the SVR result using 56 reference points for training (left); Classified SVR map with five disease severity classes (right).

The results of the linear regression models based on the SVR estimations were compared to the results obtained by a linear regression model based on NDVI values. The hyperspectral image was transformed using one band in the red at 667nm and one band in the near-infrared at 808nm. The NDVI values of the 56 reference points were used for regression analysis, which produced a Pearson correlation coefficient of r=-0.77.

		1	2	3	4	5	mean	stdv	variance
30%	train	1.00	0.98	0.91	0.98	0.99	0.97	0.04	0.00
30%	test	0.68	0.41	0.69	0.67	0.65	0.62	0.12	0.01
50%	train	0.89	0.93	0.72	0.81	0.81	0.83	0.08	0.01
50%	test	0.71	0.65	0.66	0.74	0.70	0.69	0.03	0.00
100%	train	-	-	-	-	-	0.86	-	-

Table 6.3.: Pearson Correlation coefficients obtained by different randomly selected training/test data combinations.

6.5. Discussion

This study examined two different disease severity quantification methods based on hyperspectral spectroscopy data. The presented approaches are different in regard to their spatial resolution, i.e. near-range vs. airborne, as well as their methodology. Nevertheless, both are practical to estimate the same phenomenon in this case. Whereas the near-range approach is classification-based and therefore has discrete values, the airborne approach estimates continuous severity values. The results of the classification approach are very promising. An overall correlation of r=0.82 could be achieved for all three wheat stands over the 13 time steps. But the correlation of P01 (r=0.56) differs significantly from the correlation based on P02 (r=0.94). Especially the results for the wheat stands with no or only slight disease severity have a relatively high level of uncertainty. The problem is that due to the spatial resolution of the data, pixel at the edge of the leaves are mixed with soil signals. In figure 6.5 this phenomenon can be observed especially for the leave in the left center. In addition, senescent leaves appear first at the lower levels of the wheat stands. The spectrum of a senescent leave can be misclassified as an infected leaf since senescence causes comparable effects on the spectrum as a fungal infection. This effect can be significantly minimized by using the reduced datasets. Since these datasets are reduced to bands that deliver high information to differentiate 'vital' and 'infected' leaf parts, the amount of misclassification can be significantly reduced. Especially for wheat stand P01 the correlation coefficient could thus be increased from r=0.56 to r=0.77. However, there are still some uncertainties for the wheat stands with low or no disease severities. A feature selection or band selection approach designed for this purpose or rather a pathogen specific feature selection would be a benefit.

The regression analysis based on airborne hyperspectral data examined the suitability of support vector regression for the derivation of continuous disease severities of wheat. Using 50% of the reference data for training and test respectively, a mean training accuracy of r=0.83 and a mean correlation of r=0.69 for the test data were achieved. Using all 56 reference points for training and test, an accuracy of r=0.86 for the training data could be obtained. This low training error with a high correlation coefficient shows the general suitability of the algorithm for deriving disease severities, but since the approach is sensitive to overfitting the training error has to be seen critically. The regression model using the NDVI values compared to within-field disease severities produced a correlation coefficient of r=-0.77. It has to be mentioned that the comparably good results strongly

depend on the high amount of reference points with low disease severities. If rating points with disease severities less than 5% were removed, i.e. 33 points, the correlation coefficient of the regression model using the NDVI values drops to r=-0.62, which is equal to the results obtained by SVR using 30% of the data for training and which is outperformed by using 50% of the data for training (table 6.3).

Vegetation indices such as the NDVI are useful in evaluating the consequences of stress, but they are not sensitive to the onset of stress. Another disadvantage of the NDVI is the scaling of the data. A threshold needs to be set to decide at which NDVI value a crop stand is infected, whereas the SVR directly delivers disease severities. However, if the reference data is divided into training and validation parts, the performance of the regression analyses shows low but significant correlation values (table 6.3). A mean correlation coefficient for the test data of 0.62 could be achieved using just 30% of the reference data for training. If 50% of the reference points were used, the mean correlation increased to r=0.69. Obvious misclassifications along the tractor lines appear as high severities, which indicates that an increase in soil fraction has similar spectral effects as crop diseases, i.e. both factors cause a reduced chlorophyll content of the canopy covered by a pixel. However, these errors can be reduced by context-based corrections. General results show that the regression analysis using support vectors can achieve satisfactory results, but it needs a good variety of the training data. If 50% of the reference data were used, the results of the five repetitions are robust with a standard deviation of 0.03.

However, for both studies, whether near-range or airborne spectroscopy, it has to be noted that the results include some uncertainties due to the variance inherited in visually rated disease severity data, as there is no standardized method to estimate disease severity. The visual disease rating thus varies and strongly depends on the field-observer. In addition, the spatial resolution of the airborne data was 1.5m, whereas the disease severity was rated on single points and may vary within the range of 1.5m pixel size. These facts influence training performance and results, and have to be considered in further studies.

Nevertheless, both studies achieved promising results and may be of relevance for future observations as their methods are both fully automatic and non-destructive. The findings may thus be of great interest for phenotyping in plant breeding or for adjusted fungicide treatment in the context of Precision Agriculture. Plant breeder prove the

vulnerability of new crop varieties to fungal diseases. An automatic sensing approach may therefore be of great profit since the visual rating is time consuming and highly subjective.

7. Conclusion

7.1. Summary

This thesis focused on the use of hyperspectral spectroscopy data at near-range and airborne level to detect and discriminate healthy wheat from wheat stands infected with fungal diseases. In this thesis infections with the pathogens *Blumeria graminis* causing powdery mildew and/or *Puccinia recondita* causing leaf rust were spectrally observed. Several experiments were carried out to examine the potential of hyperspectral data and to examine the impact of the spectral dimension on the detectability of crop stress, or rather analyzing the spectral requirements on plant disease detection as an semi- to fully-automatic and non-destructive sensing technique. Site-specific crop growth anomalies caused by fungal infections could thereby be identified in the laboratory and in the field, which allows a derivation of thematical maps showing areas in need of specific management actions (see chapter 4, 5 and 6).

Chapter 3 *Near-range spectroscopy for crop stress detection* examines the use of non-imaging spectroscopy data at near-range for the detection of wheat stands infected with *Blumeria graminis*. The results of the presented method of feature selection and index development with a following decision tree for classification are promising. Spectral bands representing spectral phenomena like reflectance maxima and minima were selected using derivative analysis and ratios or rather 'new' indices were designed. Even wheat stands with first evidence of mycelia and/or chlorosis could be identified with minimal misinterpretations and were classified or separated from 'vital' wheat stands and wheat stands with 'water deficiency' as a different factor of plant stress. The different key bands as shown in figure 3.4 have high potential for spectral classification. Bands within known wavelength ranges like green, red and NIR were found to be most relevant in this case study. These bands can be used for ratio calculations as presented or as input for the calculation of known indices such as NDVI. Many vegetation indices are originally designed for broadband multispectral data. This feature selecting method can be used to identify an optimal spectral band within the range of a multispectral broadband. The used band selection technique has shown that only few selected bands within the VIS/NIR spectrum are significant for a successful discrimination at near-range, whereas wavelengths above 1000nm covered by full-range sensors were not relevant. In this case study a half-range sensor would have been adequate to analyze vegetation stress. In the context of Precision Agriculture these findings show that a differentiation of biotic, i.e. in this case a fungal infection, and abiotic, i.e. in this case water deficiency, is possible using non-imaging spectroscopy data at near-range.

Since the approach as presented in chapter 3 performed very well, the decision tree analysis was tested as a feature selection technique in chapter 4 *Airborne hyperspectral remote sensing for crop stress detection*. It was thereby tested for its suitability in selecting relevant bands for the identification of fungal infections using airborne hyperspectral remote sensing data. Another main object of chapter 4 was a comparison of two different airborne hyperspectral imaging systems for the detection of fungal diseases in wheat. The results show that the AISA datasets generally outperformed the results obtained with the ROSIS sensor system due to its higher spectral and spatial resolution. As a preliminary study, two different feature selection approaches, i.e. decision trees for feature selection and the Bhattacharyya distance analysis with a feature forward search strategy, were examined. Generally the classifications based on the reduced datasets produced comparable or higher accuracies than those based on the entire spectrum. This indicates that a reduction of redundant data or rather a focusing on relevant bands can enhance the processing time of the detection approach on the one hand and the classification accuracy on the other hand.

The results of the feature selection presented in chapter 4 showed that the Bhattacharyya distance analysis with a feature forward selection is a promising tool for data reduction. The approach outperformed the results obtained with the decision tree for feature selection and was thus tested in detail in a consecutive study. In chapter 5 *Reduction of the spectral dimension of hyperspectral data* the Bhattacharyya distance with a feature forward selection was applied on a HyMap™ dataset and compared to the performance of features selected by a wrapper approach based on the spectral angle mapper. It was examined which spectral characteristics are important for the detection of pathogen infected wheat stands, in this case an infection with the pathogen *Blumeria graminis*. Lowest training error was produced using 16 spectral features and highest classification accuracies could be obtained by using 13 spectral bands (see table 5.1). It was shown that only a few but specific bands are necessary for a successful differentiation between 'healthy' and 'infected' areas within the wheat plot. For all conducted classification processes, the SVM classifier outperformed the SAM classifier, also if the spectral bands were specifically selected for SAM by a wrapper approach. The features as selected by the Bhattacharyya distance with a feature forward search strategy were additionally compared to a selection based on an exhaustive search. It can be assumed that the highest classification accuracy of this study using 13 spectral bands would be

exceeded by using the same number of features or fewer if all possible feature combinations were calculated.

In the second main analysis of chapter 5 the role of the spectral resolution for the detection of infected wheat stands was examined. The results show that the influence of the spectral resolution seems to be non-significant for this classification task. It can be assumed that a spectral resampling does not necessarily mean a loss of information, since resampled bands represent an integration of spectral information. Due to the fact that spectral characteristics of vegetation do not exhibit narrow absorption features and crop stress factors at this stage cause changes in broader wavelength areas, spectral bands with lower spectral resolution can be used for vegetation analyses since they still allow an identification of changes in vegetation status. However, the studies conducted in this chapter show that the influence of the feature selection seems to be quite more significant than the influence of a decreasing spectral resolution. In all cases the selected bands were part of the visual spectrum, or located in the red edge area and in the SWIR, whereas in some wavelengths no bands were selected (see figure 5.11). Thus, a specific combination of bands within selected spectral ranges seems preferable for the discrimination of 'healthy' and 'infected' wheat plants. It was proven that the Bhattacharyya distance is an adequate measure for feature selection and identifies the suitable bands for this classification task, especially if the highest spectral resolution is given.

Chapter 6 *Derivation of disease severities using hyperspectral data* proves the usage of hyperspectral near-range and airborne spectroscopy data for the derivation of continual disease severity values. The two approaches differed both in scale, i.e. near-range and airborne, and methodology, i.e. classification and regression. However, both approaches are in this case specifically used to examine the same phenomenon. The obtained results based on the pixel by pixel classification of hyperspectral near-range images are highly promising. Different disease severities could be derived with a high level of certainty. But, especially for the wheat stands with no or minor infection rates, the class 'infection' was overestimated. Anyway, for agricultural practice this would result in a reduced usage of fungicide treatments. A reduction to the spectral bands as identified in chapter 5 could decrease the misinterpretation and thus increase the correlation between the total number of pixel classified as 'infected' and the visual rated disease severity. The regression analysis based on airborne hyperspectral data examined the suitability of support vector machines for regression in deriving continual disease severity rates in wheat. The

performance of the regression analyses showed low but significant correlation values (see table 6.3) and continual disease severity data could be derived. Due to the findings of the previous chapters, it can be assumed that the regression approach would benefit from a preliminary feature selection. Both studies, whether near-range or airborne, achieved promising results in deriving continual disease severity data and may be of relevance for future observations, as both of them are fully automatic and non-destructive methods. The findings may thus be of great interest for phenotyping in plant breeding or for adjusted fungicide treatment in the context of Precision Agriculture.

7.2. Discussion and Outlook

Within the last nine years the remote sensing subproject of the DFG research training group 722 'Information Techniques for Precision Crop Protection' examined the three dimensions of spectral data for crop stress detection, i.e. the spatial, the temporal and the spectral dimension. A minimal spatial resolution could be defined Voss [2005] and temporal requirements of fungal disease detection were assessed Franke [2007]. As a conclusion of the three phases, the results of this study show the impact of the spectral dimension of hyperspectral data for plant disease detection. Fungal infections in wheat could thereby be identified with a high level of certainty. The accuracy of a binary classification could be significantly improved by feature selection techniques. Minimal spectral requirements for the identification of diseased wheat stands were defined at near-range level in chapter 3. 13 spectral bands within the visible and near-infrared part of the electromagnetic spectrum between 550nm and 880nm were found to be highly suitable for disease detection (figure 3.4). Spectral bands above 1000nm were not needed in this case. For airborne hyperspectral data different bands showed to be most relevant as presented in chapter 4 (figure 4.7). In this case 16 spectral bands in the near-infrared part of the spectrum between 740nm and 2070nm were selected for a successful binary classification into 'vital' and 'infected' wheat stands. A comparison of the two datasets shows significant differences in scale, cultivar, growth stage and environmental conditions. Thus a comparison of the selected bands of the airborne hyperspectral dataset to the features as selected in chapter 5 for a different airborne dataset seems preferable. Also 16 spectral bands were identified to be most suitable for the binary classification, but the location of the spectral bands (figure 5.5) significantly differ from the location of the 16 bands as selected in chapter 4. These differences may

be caused by the disparity of the two datasets, as for example,

- the hyperspectral datasets differ in spectral and spatial resolution,
- the wheat cultivar used in 2005 is different from the cultivar used in 2008,
- the wheat stands in 2005 were infected by powedery mildew, the stands of 2008 by leaf rust,
- the growth stage differed, since the day of data acquisition differed by more than a month.

It has thus to be mentioned that no general optimal spectral bands for stress detection were defined within this thesis, since those differences could not be avoided. But, it was shown that a focus on relevant data significantly improves the detectability of the considered phenomenon. It can be stated that a spectral sensor used for the detection on fungal diseases should cover the spectrum at certain spectral areas in the visible, the near-infrared and the shortwave-infrared, regardless of its spectral resolution (figure 5.11). Regarding existing sensor systems, the higher the number of spectral bands, the more useful is a feature selection. Relevant information can be filtered with the presented methods and thus, maybe more important, masses of irrelevant information can be omitted. Within this thesis, most reliable information could be extracted by focusing on specific spectral bands of the original datasets. It can be stated that a preceding feature extraction should be considered every time when working with non-imaging and imaging spectroscopy data.

A more detailed extraction of information by providing maps showing different severity classes or rather continual severity values would be a benefit for farming practice. A site-specific application of fungicides is technically possible, so spatial information about the within field heterogeneities like the maps presented in chapter 6 are particularly useful for Precision Agriculture. But some considerations, especially regarding the imaging spectroscopy data at near-range have to be made. A near-range sensing approach based on the used sensor system is difficult to realize in the field. For example, the tractor movement steadily modifies the conditions of the data acquisition in viewing angle and lightning conditions. A lot of information is needed to be able to correct and normalize the images.

Many possibilities using hyperspectral data and the high potential in deriving significant information were shown within this thesis. As presented especially in chapter 4 and 5 a binary discrimination of wheat stands into 'vital' and 'infected' is possible using airborne hyperspectral data. But, regarding the applicability of hyperspectral imaging for questionnaires in the context of Precision Agriculture, some considerations have to be made. Until now, airborne hyperspectral data are very cost intensive and a single farmer may not be able to afford them. To reduce the costs, the data may be bought by a farming consortium or similar, or by external data providers. Another problem is the availability of airborne sensor systems and the sensitivity of optical data to weather and lightning conditions. A sensor may be available when the weather conditions might be critical and vice versa. Thus, the data can possibly be not acquired at the optimal period of time. But, if hyperspectral data can be acquired when demanded, a farmer may profit from the extractable information and application maps may be derived.

As shown within this thesis the reliability of the information is very sensible to the used analyzing process. Since the presented results are phenomenon-specific at a specific time stamp, no general optimal spectral bands were defined. To extract information based on monotemporal data, reference or ground truth data have to be acquired in the field. A field observer for example trained in fungal disease rating is therefore required which increases the time and cost investments tremendously and a monitoring of entire regions is not possible. If multitemporal data are available the change or the difference of plant conditions within the field can be derived by comparing image scenes of different time stamps. By implementing this information into forecasting models based on weather data, significant conclusions on the cause of the plant stress may be possible and the field observer may be directed to the most critical locations.

For all studies, whether near-range or airborne spectroscopy, it has to be noted that the achieved results include some uncertainties due to a variance of visually rated disease severity data, as there is no standardized method to estimate disease severity. The visual disease rating thus varies and strongly depends on the field-observer. But in general, the conducted studies have shown, that it is not the entire spectrum at highest spectral resolution that is needed to remotely assess a spectral phenomenon like a fungal infection. A reduction of redundant data within hyperspectral datasets can increase the discrimination accuracy and thus the reliability of the generated or extracted information. The concluding results of this thesis are:

- The detection of fungal diseases in wheat using airborne hyperspectral data is possible with a high level of certainty, but the accuracy depends on the used sensor system and rating process. Advanced classification algorithms like support vector machines (SVM) significantly outperform conventional approaches.

- The detectability significantly profits from a preliminary data reduction or feature selection. Data reduction methods should be considered when working with hyperspectral data.

- The selected features strongly differ depending on the used sensor system, the observed phenomenon, the spatial resolution, the growth stage, the cultivar and the environmental conditions.

- A derivation of disease severity values using hyperspectral data is possible at both near-range and airborne level. Advanced approaches like support vector machines for regression (SVR) showed good results.

The results of this thesis are of great interest for future sensor design or missions, but also for different approaches of hyperspectral imagery. With the upcoming satellite-based hyperspectral sensor systems like EnMAP (Environmental Mapping and Analysis Program, DLR, Germany), PRISMA (PRecursore IperSpettrale of the application mission, ASI, Italy) or HyspIRI (Hyperspectral Infrared Imager, NASA, USA) innovative approaches for the analysis of hyperspectral data are needed. Hyperspectral data will be available to and achievable for a broader community, and multitemporal hyperspectral data for example within one phenological period will be available. Thus, the temporal dimension can be implemented into the analyzing process, resulting in an increased detectability and differentiability of species, but also for example of fungal infections. Multitemporal hyperspectral remote sensing data will certainly improve applications which are based on spatial-spectral information. Adapted and innovative methods are therefore demanded.

Bibliography

Apan, A., Held, A., Phinn, S., and Markley, J. (2004). Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. *International Journal of Remote Sensing*, 25(2):489–498.

Bajcsy, P. and Groves, P. (2004). Methodology for Hyperspectral Band Selection. *Photogrammetric Engineering and Remote Sensing*, 70(7):793–802.

Becker, B. L., Lusch, D. P., and Qi, J. G. (2005). Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis. *Remote Sensing of Environment*, 97(2):238–248.

Becker, B. L., Lusch, D. P., and Qi, J. G. (2007). A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery. *Remote Sensing of Environment*, 108(1):111–120.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109.

Blackmer, T. M., Schepers, J. S., and Meyer, G. E. (1995). Remote sensing to detect nitrogen deficiency in corn. In *Proceedings of site-specific management for Agricultural Systems*, pages 505–512, Minneapolis, USA. American Society of Agronomy.

Bongiovanni, R. and Lowenberg-Deboer, J. (2004). Precision Agriculture and Sustainability. *Precision Agriculture*, 5(4):359–387.

Börner, H., Aumann, J., and Schlüter, K. (1997). *Pflanzenkrankheiten und Pflanzenschutz (in german)*. Springer, Berlin, 7 edition.

Bravo, C., Moshou, D., West, J. S., McCartney, A., and Ramon, H. (2003). Early disease detection in wheat fields using spectral reflectance. *J.Biosyst.Eng.*, 84(2):137–145.

Breiman, L., Friedman, J. H., Olshen, R. A., and J, S. C. (1984). *Classification and Regression Trees*. CRC Press LLC, Florida.

Carter, G. A. (1991). Primary and secondary effects of water content on the spectral reflectance of leaves. *American Journal of Botany*, 78:916–924.

Carter, G. A. (1994). Ratios of Leaf Reflectances in Narrow Wavebands As Indicators of Plant Stress. *International Journal of Remote Sensing*, 15(3):697–703.

Carter, G. A. and Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88:677–684.

Carver, T. (2001). Inhibition of Blumeria graminis germination and germling development within colonies of oat mildew. *Physiological and Molecular Plant Pathology*, 58(5):209–228.

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*.

Chapelle, O. (2002). Choosing Multiple Parameters for Support Vector Machines. *Machine Learning*, 46:131–159.

Chen, G. Y. and Qian, S. E. (2008). Evaluation and comparison of dimensionality reduction methods and band selection. *Canadian Journal of Remote Sensing*, 34(1):26–32.

Chung, K. M., Kao, W. C., Sun, C. L., Wang, L. L., and Lin, C. J. (2003). Radius margin bounds for support vector machines with the RBF kernel. *Neural Computation*, 15(11):2643–2681.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3):271–278.

Dalponte, M., Bruzzone, L., Vescovo, L., and Gianelle, D. (2009). The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, 113(11):2345–2355.

De Backer, S., Kempeneers, P., Debruyn, W., and Scheunders, P. (2005). A band selection technique for spectral classitication. *Ieee Geoscience and Remote Sensing Letters*, 2(3):319–323.

de Colstoun, E. C. B., Story, M. H., Thompson, C., Commisso, K., Smith, T. G., and Irons, J. R. (2003). National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sensing of Environment*, 85(3):316–327.

DeFries, R. S. and Chan, J. C. W. (2000). Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74(3):503–515.

Delalieux, S., Somers, B., Verstraeten, W. W., Keulemans, W., and Coppin, P. (2008). Hyperspectral canopy measurements under artificial illumination. *International Journal of Remote Sensing*, 29(20):6051–6058.

Erasmi, S. (2003). *Optische Fernerkundung zur teilflächenspezifischen Zustandsbeschreibung von Winterweizenbetänden (in german)*. Göttinger Geographische Abhandlungen, Göttingen.

Franke, J. (2007). *Spatiotemporal Dynamics of stress factors in wheat analysed by multisenoral remote sensing and geostatistics*. Online-Publikationen an deutschen Hochschulen, University of Bonn.

Franke, J., Gebhardt, S., Menz, G., and Helfrich, H.-P. (2009a). Geostatistical analysis of the spatiotemporal dynamics of powdery mildew and leaf rust in wheat. *Phytopathology*, 99(8):974–84.

Franke, J. and Menz, G. (2007). Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture*, 8(3):161–172.

Franke, J., Mewes, T., and Menz, G. (2008). Airborne hyperspectral imaging for the detection of powdery mildew in wheat. In Shen, S. and Lewis, P., editors, *Proceedings of SPIE - Optics & Photonics*, San Diego, USA. SPIE.

Franke, J., Mewes, T., and Menz, G. (2009b). Requirements on spectral resolution of remote sensing data for crop stress detection. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 2–5, Cape Town, South Africa. IEEE.

Friedl, M. A. and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3):399–409.

Gates, D. M., Keegan, H. J., Schleter, J. C., and Weidner, V. R. (1965). Spectral Properties of Plants.

Gitelson, A. and Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of Aesculus-Hippocastanum l. and Acer-Platanoides l. leaves - spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143(3):286–292.

Goetz, A. F. H. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing of Environment*, 113:S5–S16.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., Olah, M. R., and Williams, O. (1998). Imaging spectroscopy and the Airborne Visible/Infrred Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, 65:227–248.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., and Sohlberg, R. (2000). Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21(6-7):1331–1364.

Herold, M., Gardner, M. E., and Roberts, D. A. (2003). Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1907–1919.

Hoffmann, G. M. and Schmutterer, H. (1999). *Parasitäre Krankheiten und Schädlinge an landwirtschaftlichen Kulturpflanzen. (Nutzpflanzen) (in german).* Ulmer, 2 edition.

Huang, W. (2007). Identification of yellow rust in wheat using in-situ spectral reflectance measurements ans airborne hyperspectral imaging.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63.

Jackson, R. D. (1986). Remote Sensing of biotic and abiotic plant stress. *Annual Review of Phytopathology*, 24:265–287.

Jacobi, J. (2005). Teilfl"chige Diagnose von Pilzbefall und Stickstoffmangel an Weizen mit Hilfe der optischen Fernerkundung.

Jacobi, J. and Kühbauch, W. (2005). Site-specific identification of fungal infection and nitrogen deficiency in wheat crop using remote sensing. *Proceedings of the 5th European Conference on Precision Agriculture*, pages 73–80.

Jain, N., Shibendu, S. R., Singh, J. P., and Panigraphy, S. (2007). Use of hyperspectral data to assess the effect of different nitrogen apllications on a potato crop. *Precision Agriculture*, 8(4-5):225–239.

Jimenez, L. O. and Landgrebe, D. A. (1999). Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667.

Keshava, N. (2004). Distance Metrics and Band Selection in Hyperspectral Processing With Applications to Material Identification and Spectral Libraries. *Geoscience and Remote Sensing, IEEE Transactions on*, (7):1552–1565.

Knipling, E. B. (1970). Physical and Physiological Basis for the Reflectance of Visible and Near-Infrared Radiation from Vegetation.

Kohavi, N. and John, G. H. (1998). Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324.

Kruse, F. A., Boardman, J. W., and Huntington, J. F. (2003). Comparison of Airborne Hyperspectral Data and EO-1 Hyperion for Mineral Mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1388–1400.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H. (1993). The Spectral Image-Processing System (Sips) - Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sensing of Environment*, 44(2-3):145–163.

Kühbauch, W. and Hawlitschka, S. (2003). Remote sensing - a future technology in precision farming. *Proceedings of the Workshop on POLinSAR - Applications of SAR Polarimetry and Polarimetric Interferometry (ESA SP-529)*.

Landis, J. R. and Koch, G. G. (1977). Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.

Lawrence, R. L. and Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing*, 67(10):1137–1142.

Lorenzen, B. and Jensen, A. (1989). Changes in Leaf Spectral Properties Induced in Barley by Cereal Powdery Mildew. *Remote Sensing of Environment*, 27(2):201–209.

Lu, S., Oki, K., Shimizu, Y., and Omasa, K. (2007). Comparison between several feature extraction/classification methods for mapping complicated agricultural land use patches using airborne hyperspectral data. *International Journal of Remote Sensing*, 28(5):963–984.

Marsalis, M. A. and Goldberg, N. P. (2006). Leaf , Stem and Stripe Rust Diseases of Wheat. URL: http://cahe.nmsu.edu/pubs/_a/A-415.pdf (Mar 2006).

Miao, X., Gong, P., Swope, S., Pu, R. L., Carruthers, R., and Anderson, G. L. (2007). Detection of yellow starthistle through band selection and feature extraction from hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing*, 73(9):1005–1015.

Milton, E. J., Schaepman, M. E., Anderson, K., Kneubühler, M., and Fox, N. (2009). Progress in field spectroscopy. *Remote Sensing of Environment*, 113:92–109.

Moran, M. S., Inoue, Y., and Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, 61(3):319–346.

Moshou, D., Bravo, C., West, J. S., Wahlen, S., McCartney, H. A., and Ramon, H. (2004). Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture*, 44:173–188.

Pinter, P. J., Hatfield, J. L., Schepers, J. S., Bames, E. M., Moran, M. S., Daughtry, C. S. T., and Upchurch, D. R. (2003). Remote Sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, 69(6):647–664.

Rabe, A., Van der Linden, S., and Hostert, P. (2009). *imageSVM, Version 2.0*.

Rodemann, B. (2009). Krankheiten durch pilzliche Pathogene im Weizen. In Christen, O., editor, *Winterweizen - Das Handbuch für Profis*, pages 177–236. DLG-Verlags-GmbH, Frankfurt.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. In *Third ERTS Symposium, NASA*, pages 309–317, USA. NASA.

Savitzky, A. and Golay, J. M. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639.

Serpico, S. B. and Bruzzone, L. (2001). A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1360–1367.

Shaw, D. and Kelley, F. (2005). Evaluating Remote Sensing for determining and classifying soybean anomalies. *Precision Agriculture*, 6(5):421–429.

Smith, K. L., Steven, M. D., and Colls, J. J. (2004). Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks. *Remote Sensing of Environment*, 92(2):207–217.

Strange, R. N. (2003). *Introduction to Plant Pathology.* John Wiley and Sons Ltd., England.

Thenkabail, P. S. (2001). Optimal hyperspectral narrowbands for discriminating agricultural crops. *Remote Sensing Reviews*, 20(4):257–291.

Thomas, J. R. and Gausman, H. W. (1977). Leaf reflectance vs leaf chlorophyll and carotenoid concentrations for eight crops. *Agronomy Journal*, 69(5):799–802.

Van der Linden, S., Rabe, A., Okujeni, A., and Hostert, P. (2009). *Image SVM classification, Application Manual: image SVM version 2.0.*

Vapnick, V. (1998). *Statistical learning theory.* John Wiley and Sons Inc.

Voss, K. (2005). *Beitrag zur skalenabhängigen Erfassung teilschlagspezifischer Pflanzenschäden mit Methoden der Fernerkundung und Geoinformation (in german).* Online-Publikationen an deutschen Hochschulen, University of Bonn.

Voss, K., Franke, J., Mewes, T., Menz, G., and Kühbauch, W. (2010). *Remote Sensing for Precision Crop Protection A Matter of Scale*, pages 101–118. Springer, Netherlands.

Waske, B., Van der Linden, S., Benediktsson, J. A., Rabe, A., and Hostert, P. (2009). Impact of different morphological profiles on the classification accuracy of urban hyperspectral data. In *IEEE Whispers - Workshop on Hyperspectral Image and Signal Processing*, Grenoble, France. IEEE.

West, J. S., Bravo, C., Oberti, R., Lemaire, D., Moshou, D., and McCartney, H. A. (2003). The potential of optical canopy measurement for targeted control of field crop diseases. *Ann.Rev.Phytopathol.*, 41:593–614.

Yoder, B. J. and Pettigrew-Crosby, R. E. (1995). Prediction nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. *Remote Sensing of Environment*, 53(3):199–211.

A. SAM wrapper

For this thesis a wrapper approach using the spectral angle mapper (SAM) for feature selection and evaluation was programed and implemented in ENVI/IDL 4.7©. The GUI (see figure A.1) of this implementation and the flow chart explaining the code are given below. Until now, there are only few functions or options implemented into the application, but, the code is easily extendible by adding other classification algorithms and selectable options. As the very first step, the user has to open an multispectral or hyperspectral image via the 'Open IMG' button (red Box 1). If the image is correctly loaded into ENVI/IDL 4.7© a thumbnail of it will be displayed in the main window (red Box 3). The program is based on the usage of regions of interest (ROI) directly derived from the image as training data for the classifier. Thus, these regions have to be preliminary defined, otherwise the classification cannot work and the program will print: 'No regions associated with the selected file. Process canceled!'. To associate predefined regions with the opened image, a *.roi file is to be selected after activating the ROI button (red Box 2). If the selected file fits to the image the different defined regions will be listed at the bottom of the GUI (red Box 7). Afterwards there are two buttons available, i.e. SAM_all (red Box 4) and SAM_FFS (red Box 5). The SAM_all button executes a classification of the image using SAM with all spectral bands. The SAM_FFS button executes the SAM wrapper approach using a feature forward search strategy. For both approaches different options are available. The result of SAM strongly depends on the used threshold, which is angular limitation for the class decision. The default setting is a threshold of 0.1. If a different threshold is desired, a fixed value can be set using the scroll bar. But, since the optimal threshold is usually unknown, the program is able to automatically search for an optimal threshold between 0.1 and 5. The threshold found for an optimal separation of the given classes using all bands is used for every following classification within the SAM_FFS approach. The user may decide whether to keep the data subsets of the different band combinations for further analysis or not . But, since the amount of combinations may be very large for hyperspectral data, the requirements for free disk space might be too high.

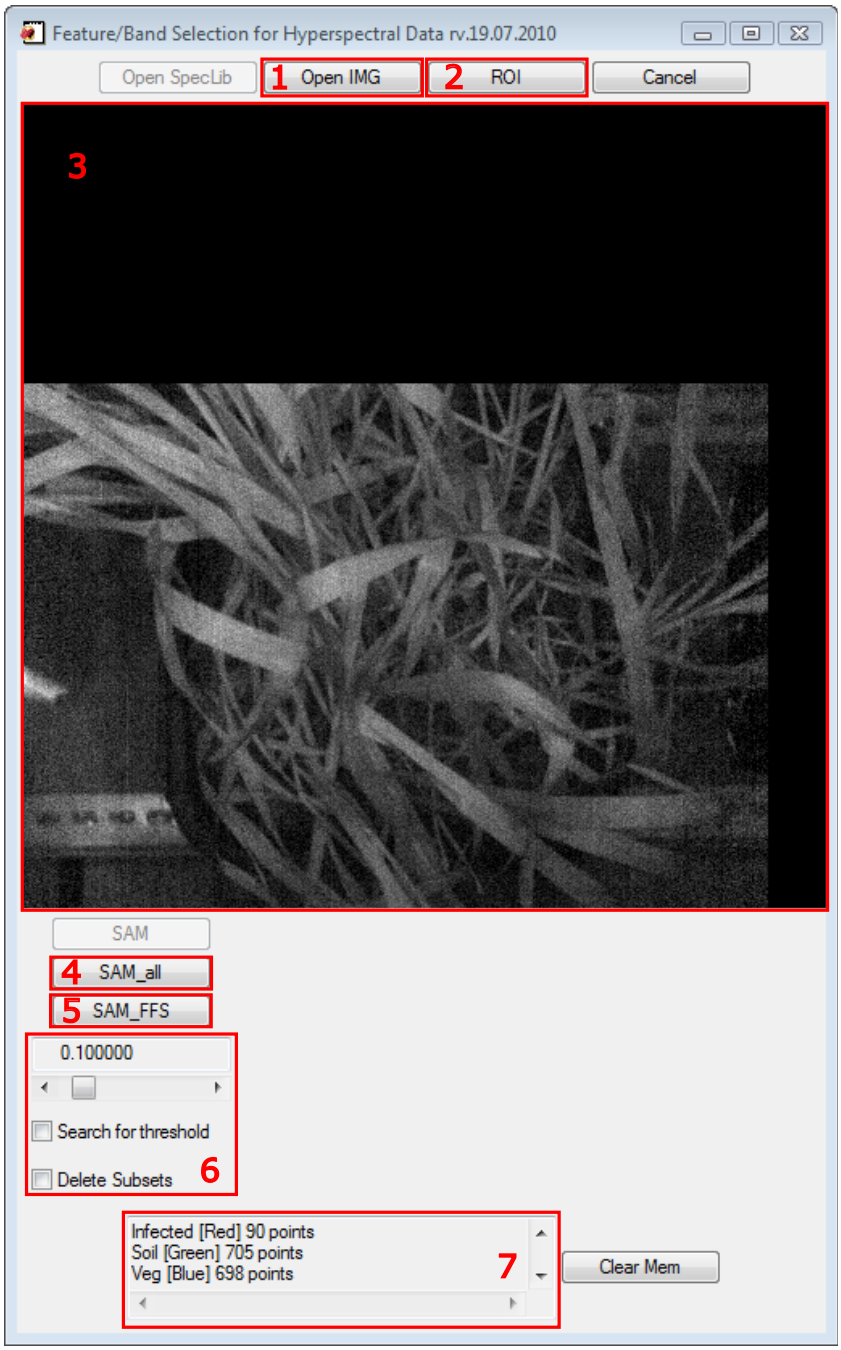


Figure A.1.: GUI of the SAM wrapper implementation.

```
pro sam_fcc, threshold, DelSub

; ------------------------------------------------------------------------------------
; Programm for the calculation of optimal band combinations for classification
; The input data can be a multi- or hyperspectral image with associated ROIS
; Each ROI represents spectral samples for one class
; The ROIs have to be defined before the calculation and can then be selected via GUI
; It can be used to simply classify an image using all bands
; or as an band selection approach using wrapper techniques.
; Until now the entire program focusses only on SAM classification
; ------------------------------------------------------------------------------------
; Author: Thorsten Mewes -> tmewes@uni-bonn.de
; rv.19.07.2010
; ------------------------------------------------------------------------------------

envi_select, title='Choose the image to classify by SAM', fid=fid, dims=dims, pos=pos
      if fid eq -1 then begin
        print, 'input cancelled'
        goto, quit
      endif

envi_file_query, fid, fname=fname, bname=bname, xstart=xstart, ystart=ystart, $
    data_type=data_type, file_type=file_type, interleave=interleave

map_info = envi_get_map_info(fid=fid)
num_cols = dims[2]-dims[1]+1
num_rows = dims[4]-dims[3]+1
num_pixels = (num_cols*num_rows)
num_bands = n_elements(pos)

image=fltarr(num_cols,num_rows,num_bands)

for i=0,num_bands-1 do $
  image[*,*,i]=envi_get_data(fid=fid,dims=dims,pos=pos[i])

pos  = lindgen(num_bands)

;;; ROIs  ;;;;

  ;; A -> Training ROIs

    roi_ids = ENVI_GET_ROI_IDS(fid=fid, roi_colors=roi_colors, roi_names=class_names)
    if (roi_ids[0] eq -1) then begin
      print, 'No regions associated with the selected file. Process cancelled!'
      return
    endif

    base = widget_auto_base(title='ROI Selection and Options')
    wm   = widget_multi(base, list=class_names, uvalue='list', $
      Prompt='Please select two ROIs for the Calculation', /NO_Range, /auto)
    resultwid = auto_wid_mng(base)
    if (resultwid.accept eq 0) then return
    ptr1 = where(resultwid.list eq 1)

    roi_id_dummy=roi_ids
    roi_colors_dummy=roi_colors
    roi_names_dummy=class_names
    num_ptrs=n_elements(ptr1)
    roi_ids=intarr(num_ptrs)
    roi_colors=intarr(3,num_ptrs)
    class_names=strarr(num_ptrs)
    for i=0, num_ptrs-1 do begin
      roi_ids[i]=roi_id_dummy[ptr1[i]]
      roi_colors[*,i]=roi_colors_dummy[*,ptr1[i]]
      class_names[i]=roi_names_dummy[ptr1[i]]
    endfor

    ;; Sorting the training ROIs

    roi_ids=roi_ids[sort(class_names)]
    roi_colors=roi_colors[*,sort(class_names)]
    class_names=class_names[sort(class_names)]

    ;; Loading additional information about the training ROIs

    num_rois=n_elements(roi_ids)
    roi_elem=intarr(n_elements(roi_ids))
    for i=0, n_elements(roi_ids)-1 do begin
        roi_addr = envi_get_roi(roi_ids[i], roi_name=name)
        roi_elem[i]=n_elements(roi_addr)
    endfor
```


Figure A.2.: Core part of the code of the SAM wrapper implementation.

sam_fcc_no_comments.pro

```
    ;; Lookup table for the training data

    num_classes = n_elements(roi_ids)
    lookup = bytarr(3, num_classes + 1)
    lookup = bytarr(3,num_classes+1)
    lookup[0,1] = roi_colors

    roi_names=strarr(num_classes+1)
    roi_names[0]='unclassified'
    roi_names[1]=class_names
    class_pt=sort(roi_names)


    ;; B -> Test ROIs

    val_roi_ids = ENVI_GET_ROI_IDS(fid=val_fid, roi_colors=val_roi_colors, roi_names=val_class_names)
    if (val_roi_ids[0] eq -1) then begin
      print, 'No regions associated with the selected file. Process cancelled!'
      return
    endif

    base = widget_auto_base(title='ROI Selection and Options')
    wm   = widget_multi(base, list=val_class_names, uvalue='list', Prompt='Please select at least two ROIs for
the Validation', /NO_Range, /auto)
    resultwid = auto_wid_mng(base)
    if (resultwid.accept eq 0) then return
    ptr2 = where(resultwid.list eq 1)

    val_roi_id_dummy=val_roi_ids
    val_roi_colors_dummy=val_roi_colors
    val_roi_names_dummy=val_class_names
    val_num_ptrs=n_elements(ptr2)
    val_roi_ids=intarr(val_num_ptrs)
    val_roi_colors=intarr(3,val_num_ptrs)
    val_class_names=strarr(val_num_ptrs)
    for i=0, val_num_ptrs-1 do begin
      val_roi_ids[i]=val_roi_id_dummy[ptr2[i]]
      val_roi_colors[*,i]=val_roi_colors_dummy[*,ptr2[i]]
      val_class_names[i]=val_roi_names_dummy[ptr2[i]]
    endfor

    ;; Sorting the test ROIs

    val_roi_ids=val_roi_ids[sort(val_class_names)]
    val_roi_colors=val_roi_colors[*,sort(val_class_names)]
    val_class_names=val_class_names[sort(val_class_names)]

    ;; Loading additional information about the test ROIs

    val_num_rois=n_elements(val_roi_ids)
    val_roi_elem=intarr(n_elements(val_roi_ids))
    for i=0, n_elements(val_roi_ids)-1 do begin
        val_roi_addr = envi_get_roi(val_roi_ids[i], roi_name=val_name)
        val_roi_elem[i]=n_elements(val_roi_addr)
    endfor

    ; Building an array which contains the Ground Truth Values

    roi_val=intarr(total(val_roi_elem))
    roi_val[0:val_roi_elem[0]-1]=1
    for i=1, val_num_rois-1 do begin
      roi_val[total(val_roi_elem[0:i-1]):total(val_roi_elem[0:i])-1]= [i]+1
    endfor

    print, val_class_names

; Mean for Classification

mean = fltarr(n_elements(pos), num_classes)
  for j=0, num_classes-1 do begin
    ; get the statistics for each selected class
    roi_dims=[envi_get_roi_dims_ptr(roi_ids[j]),0,0,0,0]
    if roi_elem[j] EQ 1 then begin
      roi_specdata=fltarr(num_bands, 1)
      roi_specdata[*,*]= ENVI_GET_ROI_DATA(roi_ids[j], ADDR=ADDR, fid=fid, pos=pos)
      mean[0,j]=roi_specdata
    endif else begin
      envi_doit, 'envi_stats_doit', fid=fid, pos=pos, $
        dims=roi_dims, comp_flag=4, mean=c_mean
      mean[0,j] = c_mean
```

```
    endelse
  endfor

; Searching for a good threshold! (OPTIONAL)

    if threshold EQ 99 then begin
      k=(findgen(200)/10)+0.1
      best=fltarr(200,2)
      for z=0, 49 do begin
        thresh=replicate(k[z], num_classes)
        envi_doit, 'class_doit', fid=fid, pos=pos, dims=dims, $
                   out_bname='SAM', method=3, $
                   mean=mean, r_fid=r_fid, $
                   lookup=lookup, class_names=roi_names, $
                   in_memory=1, thresh=thresh, /no_open, /NO_REALIZE
        ; Building an array which contains the Class Values
        roi_elem=val_roi_elem
        roi_data=strarr(total(roi_elem))
        roi_data[0:roi_elem[0]-1]= ENVI_GET_ROI_DATA(val_roi_ids[0], fid=r_fid, pos=[0])
        for i=1, num_rois-1 do begin
        roi_data[total(roi_elem[0:i-1]):total(roi_elem[0:i])-1]= &
          ENVI_GET_ROI_DATA(val_roi_ids[i], fid=r_fid, pos=pos[0])
        endfor

        ; Calculating the overall accuracy (OAA) & Kappa

        all=total(roi_elem)
        dummy=lonarr(all)
        dummy1=where(roi_val EQ roi_data)
        dummy[dummy1]=1
        true=total(dummy, 1)
        OAA= (true/total(roi_elem))*100
        true_classes=intarr(num_rois)
        dummy2=where(roi_val[0:roi_elem[0]-1] EQ roi_data[0:roi_elem[0]-1])
        true_classes[0]=n_elements(dummy2)
        ref_classes=intarr(num_rois)
        ref_classes[0]=roi_elem[0]
        pred_classes=intarr(num_rois)
        pred_classes[0]=n_elements((where(roi_data EQ 1)))
        for i=1, num_rois-1 do begin
          dummy3=where(roi_val[roi_elem[i-1]:total(roi_elem[0:i])-1] &
            EQ roi_data[roi_elem[i-1]:total(roi_elem[0:i])-1])
          true_classes[i]=n_elements(dummy3)
          ref_classes[i]=roi_elem[i]
          pred_classes[i]=n_elements((where(roi_data EQ i+1)))
        endfor
        dummy=intarr(num_rois)
        for j=0, num_rois-1 do begin
          dummy[j]=pred_classes[j]*ref_classes[j]
        endfor
        dummy2=total(dummy)
        Kappa=((all*true)-dummy2)/(all^2-dummy2)
        best[*,0]=k
        best[z,1]=Kappa
      endfor
      bestsort=best  ;selecting the threshold with the lowest angle (few)
                     ;and highest value (best_all)
      a=sort(best[*,1])
      b1=best[*,0]
      b2=best[*,1]
      b1s=b1[reverse(a)]
      b2s=b2[reverse(a)]
      bestsort[*,0]=b1s
      bestsort[*,1]=b2s
      best_transp=transpose(bestsort[0:z-1,*])
      best_val=max(best_transp[1,*])
      posi=where(best_transp[1,*] EQ best_val)
      posi_elem=n_elements(posi)
      best_all=fltarr(2,posi_elem)
      best_all[0,*]=best_transp[0,posi]
      best_all[1,*]=best_transp[1,posi]
      sort_best=sort(best_all[0,*])
      few=fltarr(2,posi_elem)
      few[0,*]=best_all[0,sort_best]
      few[1,*]=best_all[1,sort_best]
      threshold=few[0,0]
      clear_memory
    endif
    thresh=replicate(threshold, num_classes)
    envi_doit, 'class_doit', fid=fid, pos=pos, dims=dims, $
        out_bname='SAM', method=3, $
```

sam_fcc_no_comments.pro

```
          mean=mean, r_fid=r_fid, $
          lookup=lookup, class_names=roi_names, $
          in_memory=1, thresh=thresh, /no_open, /NO_REALIZE

; END OF CLASSIFICATION
;;;;
;;;;
;;;;
; START VALIDATION


; Building an array which contains the Class Values

roi_elem=val_roi_elem
roi_data=strarr(total(roi_elem))
roi_data[0:roi_elem[0]-1]= ENVI_GET_ROI_DATA(val_roi_ids[0], fid=r_fid, pos=[0])
for i=1, num_rois-1 do begin
  roi_data[total(roi_elem[0:i-1]):total(roi_elem[0:i])-1]= ENVI_GET_ROI_DATA(val_roi_ids[i], fid=r_fid,
pos=pos[0])
endfor

; Calculating the overall accuracy (OAA) & Kappa

all=total(roi_elem)
dummy=lonarr(all)
dummy1=where(roi_val EQ roi_data)
dummy[dummy1]=1
true=total(dummy, 1)
OAA= (true/total(roi_elem))*100
true_classes=intarr(num_rois)
dummy2=where(roi_val[0:roi_elem[0]-1] EQ roi_data[0:roi_elem[0]-1])
true_classes[0]=n_elements(dummy2)
ref_classes=intarr(num_rois)
ref_classes[0]=roi_elem[0]
pred_classes=intarr(num_rois)
pred_classes[0]=n_elements((where(roi_data EQ 1)))
for i=1, num_rois-1 do begin
  dummy3=where(roi_val[roi_elem[i-1]:total(roi_elem[0:i])-1] EQ roi_data[roi_elem[i-1]:total(roi_elem[0:i])-1])
  true_classes[i]=n_elements(dummy3)
  ref_classes[i]=roi_elem[i]
  pred_classes[i]=n_elements((where(roi_data EQ i+1)))
endfor
dummy=intarr(num_rois)
for j=0, num_rois-1 do begin
  dummy[j]=pred_classes[j]*ref_classes[j]
endfor
dummy2=total(dummy)
Kappa=((all*true)-dummy2)/(all^2-dummy2)

print, OAA
print, Kappa

; END OF USING ALL BANDS
;;;;
;;;;
; START WRAPPER

;; 2-bands-combinations - Counting

    n_bands=num_bands
    n_classes=2

    count_possibilities , n_bands, n_classes, possi
    combitot=n_elements(possi[0,*])

;; Subsets formating

    namedummy=strarr(combitot)
    filenamedummy=strarr(combitot)
    for i=0L, combitot-1 do begin
      namedummy[i]=string('SAM_' +strjoin(strtrim(string(possi[*,i]),2), '_'));+'.img')
      filenamedummy[i]=string('SAM_2er_' +strjoin(strtrim(string([i]),2), '_'));+'.img')
    endfor

    for i=0L, combitot-1 do begin
          datasub=image[*,*,possi[0:1,i]]
          file=namedummy[i]
          ENVI_WRITE_ENVI_FILE, datasub, out_name='\temp\temp\' &
            +filenamedummy[i], map_info=map_info, /no_open
    endfor

;; Preparing Result-Files
```

```
Accuracy=strarr(3+num_bands,combitot+1,2)
Accuracy[0,0,0]='Accuracy_all_bands'
Accuracy[1,0,0]='OAA [%]'
Accuracy[2,0,0]='Kappa'
Accuracy[3:num_bands+2,1,0]=indgen(num_bands)
Accuracy[0,1,0]='All bands'
Accuracy[1,1,0]=OAA
Accuracy[2,1,0]=Kappa

Accuracy3=strarr(3+num_bands,num_bands+1,num_bands)
Accuracy3[0,0,0]='Accuracy_all_bands'
Accuracy3[1,0,0]='OAA [%]'
Accuracy3[2,0,0]='Kappa'
Accuracy3[3:num_bands+2,1,0]=indgen(num_bands)
Accuracy3[0,1,0]='All bands'
Accuracy3[1,1,0]=OAA
Accuracy3[2,1,0]=Kappa

;; Loading 2-combi-Subset and classify

ENVI_CLOSE_DISPLAY, 0
for i=0L, combitot-1 do begin
  filename='\Temp\Temp\'+filenamedummy[i]
  sam_fcc_cl, filename, ptr1, ptr2, Accura, threshold
  Accuracy[0,0,1]='Accuracy_2_bands'
  Accuracy[1,0,1]='OAA[%]'
  Accuracy[2,0,1]='Kappa'
  Accuracy[0,i+1,1]=namedummy[i]
  Accuracy[1,i+1,1]=Accura[0,1]
  Accuracy[2,i+1,1]=Accura[1,1]
  Accuracy[3:4,i+1,1]=possi[*,i]
endfor

j=1
result=Accuracy
run=combitot
sort_data, result, j, run, num_bands
Accuracy=result

print, result

; delete 2-bands-combinations from disk

if DelSub EQ 1 then begin
  for k=0L, combitot-1 do begin
    filena='\Temp\Temp\'+filenamedummy[k]
    filehead='\Temp\Temp\'+filenamedummy[k]+'.hdr'
    file_delete, filena
    file_delete, filehead
  endfor
endif

; Wrapper with more than 2 bands

for a=3, num_bands-1 do begin
  combirun=intarr([a],n_bands)
  combirun[a-1,*]=indgen(n_bands)
  if a EQ 3 then begin
    combirun[0,*]=Accuracy[3,1,1]
    combirun[1,*]=Accuracy[4,1,1]
    for r=0, n_bands-1 do begin
      if (combirun[2,r] EQ Accuracy[3,1,1]) OR (combirun[2,r] &
          EQ Accuracy[4,1,1]) then combirun[*,r]=0
    endfor
  ENDIF ELSE BEGIN
    for w=0, n_bands-1 do begin
      combirun[0:a-2,w]=result[3:a+1,1,a-2]
    endfor
    for r=0, n_bands-1 do begin
      for t=0, a-2 do begin
        if (combirun[a-1,r] EQ combirun[t,r]) then combirun[*,r]=0
      endfor
    endfor
  ENDELSE


  combisort=combirun
  d=combirun[1,*]
  dim=size(d,/DIMENSIONS)
  s=sort(d)
```

sam_fcc_no_comments.pro

```
  h=histogram(s mod dim[0],REVERSE_INDICES=ri)
  sorted_inds=transpose(reform(s[ri[dim[0]+1L:*]],dim))
  for u=0, a-1 do begin
    e=combirun[u,*]
    combisort[u,*]=transpose(e[reverse(sorted_inds)])
  endfor

  namedummy=strarr(num_bands)
  filenamedummy=strarr(num_bands)
  for i=0L, num_bands-a do begin
      namedummy[i]=string('SAM_' &
        +strjoin(strtrim(string(combisort[*,i]),2), '_')+'.img')
      filenamedummy[i]=string('SAM_'+strjoin(strtrim(string([a]),2)) &
        +'er_' +strjoin(strtrim(string([i]),2), '_'));+'.img')
  endfor

  for i=0L, num_bands-a do begin
          datasub=image[*,*,combisort[0:[a]-1,i]]
          file=namedummy[i]
          ENVI_WRITE_ENVI_FILE, datasub, out_name='\temp\temp\' &
            +filenamedummy[i], map_info=map_info, /no_open
  endfor

  for i=0L, num_bands-a do begin
    filename='\Temp\Temp\'+filenamedummy[i]
    sam_fcc_cl, filename, ptr1, ptr2, Accura, threshold
    Accuracy3[0,0,a-1]='Accuracy_X_bands'
    Accuracy3[1,0,a-1]='OAA[%]'
    Accuracy3[2,0,a-1]='Kappa'
    Accuracy3[0,i+1,a-1]=namedummy[i]
    Accuracy3[1,i+1,a-1]=Accura[0,1]
    Accuracy3[2,i+1,a-1]=Accura[1,1]
    Accuracy3[3:a+2,i+1,a-1]=combisort[*,i]
  endfor
  j=a-1
  result=Accuracy3
  run=0
  run=num_bands
  sort_data, result, j, run, num_bands
  Accuracy3=result

  ; delete X-bands-combinations from disk

  if DelSub EQ 1 then begin
    for k=0L, num_bands-a do begin
      filena='\Temp\Temp\'+filenamedummy[k]
      filehead='\Temp\Temp\'+filenamedummy[k]+'.hdr'
      file_delete, filena
      file_delete, filehead
    endfor
  endif

endfor

; Printing results into files (physically)

file='\temp\Accuracy_ALL_n_2er.txt'
    print, file
    openw, unit, file, /get_lun
    printf, unit, Accuracy
    close, unit
    free_lun, unit

file='\temp\Accuracy_3er_n_up.txt'
    print, file
    openw, unit, file, /get_lun
    printf, unit, Accuracy3
    close, unit
    free_lun, unit

; Output Summary

accu_sum=strarr(3+num_bands,3,num_bands-1)
accu_sum[0,0,*]='Accuracy_all_bands'
accu_sum[1,0,*]='OAA[%]'
accu_sum[2,0,*]='Kappa'
accu_sum[3:num_bands+2,0,*]='Band'
accu_sum[*,1,0]=accuracy[*,1,0]


for i=1, num_bands-2 do begin
```

Page 6

```
  accu_sum[0,0,i]='Accuracy_using_'+strjoin(strtrim(string([i]+1),2))+'bands'
  accu_sum[*,1,i]=accuracy3[*,1,i]
endfor
accu_sum[*,1,1]=accuracy[*,1,1]

plot_kappa=fltarr(num_bands-1)
plot_kappa[0]=accuracy[2,1,1]
plot_kappa[1:num_bands-2]=accuracy3[2,1,2:num_bands-1]
plot_kappa[num_bands-2]=accuracy[2,1,0]
x_ax=intarr(num_bands-1)
x_ax=indgen(num_bands-1)+2

set_plot,'ps'
device, filename = '\temp\Accuracy.ps'
plot, x_ax, plot_kappa, TITLE='Accuracy Assessment - Feature Selection', $
         XTITLE='Number of Bands', YTITLE='Kappa coefficient', /YNOZERO
device, /close
set_plot,'WIN'


file='\temp\Accuracy_Summary.txt'
    print, file
    openw, unit, file, /get_lun
    printf, unit, accu_sum
    close, unit
    free_lun, unit


print, 'Used threshold:', threshold
thresh_inf=strarr(2,200+1)
thresh_inf[0,0]='Used threshold'
thresh_inf[1,0]=threshold
thresh_inf[0,1:200]=transpose(best[*,0])
thresh_inf[1,1:200]=transpose(best[*,1])
file='\temp\Infos_threshold.txt'
    print, file
    openw, unit, file, /get_lun
    printf, unit, thresh_inf
    close, unit
    free_lun, unit

 class_inf=strarr(2,n_elements(class_names))
 class_inf[0,*]=class_names
 class_inf[1,*]=val_class_names
 file='\temp\Infos_classes.txt'
    print, file
    openw, unit, file, /get_lun
    printf, unit, class_inf
    close, unit
    free_lun, unit

quit:

end
```