Institut für Geodäsie und Geoinformation
Geoinformation

Identification of Relevant Patterns in Spectral Signatures

Inaugural-Dissertation

zur

Erlangung des Grades

Doktor-Ingenieur

(Dr.-Ing.)

der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Dipl.-Ing. Christoph Römer

aus

Neuwied

Referent:	Prof. Dr. Lutz Plümer
Korreferent:	Prof. Dr. Uwe Rascher
Korreferent:	Prof. Dr. Stefan Wrobel
Tag der mündlichen Prüfung:	26.08.2015
Erscheinungsjahr:	2016

Abstract

For Plant Phenotyping, non-invasive measurements of early stress processes in plants and the quantification of soil parameters with hyperspectral sensors are of particular importance, but relevant information is often concealed in the data. While the relevant parameters are measured, only a combination of different, very application-specific methods may reveal them. Even though promising results have been achieved by using supervised machine learning methods, existing features are not optimal. In other cases, labels are not obtainable or, especially for soil measurements, the information about the searched parameters is masked in the spectrum itself. In this thesis, new methods are developed to quantify and visualize relevant processes for Plant Phenotyping from hyperspectral data. New features are constructed in order to deal with the bad signal to noise ratio of early stress processes. These features are describing the information about the whole spectrum by piece-wise polynomials. The new features enable an earlier and more accurate prediction of stress symptoms despite noisy measurements. A method is presented to extract labels from unlabeled hyperspectral images. For this, an unsupervised archetypal matrix factorization is used to construct a second order feature space. The new feature space enables the prediction of early drought stress on the plant level. A hierarchical classification approach is developed to deal with relevant parameters masked by undesirable influences on the spectrum. The hierarchical approach eliminates the variance resulting from these disturbing influences. This enables more accurate regression and classification models. The presented methods succeed in an earlier stress prediction and a significant improvement in soil parameter quantification from hyperspectral data.

Kurzfassung

Für die Phänotypisierung von Pflanzen sind nicht-invasive Messungen von frühen Stressprozessen und die Quantifizierung von Bodenparametern mit hyperspektralen Sensoren von hoher Bedeutung, aber die relevanten Informationen sind oftmals in den Daten verborgen. Während die relevanten Parameter zwar erfasst werden, können diese lediglich durch eine Kombination aus verschiedenen, sehr anwendungsspezifischen Methoden aufgedeckt werden. Obwohl sehr vielversprechende Erfolge mit überwachten Methoden des Maschinellen Lernens erzielt wurden, sind die bisher benutzen Merkmale nicht optimal. In anderen Fällen sind annotierte Daten nicht erhältlich oder, besonders für die Bestimmung von Bodenparametern, ist die Information über die gesuchten Prozesse in der hyperspektralen Signatur verborgen. In dieser Arbeit werden neue Methoden zur Quantifizierung und Visualisierung von relevanten Prozessen für die Phänotypisierung von Pflanzen aus hyperspektralen Daten entwickelt. Neue Merkmale werden konstruiert um mit dem schlechten Signal-Rausch-Verhältnis in der Früherkennung umzugehen. Diese Merkmale beschreiben die Information des gesamten Spektrums durch stückweise Polynome. Die neuen Merkmale erlauben eine frühere und präzisere Früherkennung von Stresssymptomen, trotz verrauschter Messdaten. Eine Methode wird präsentiert um annotierte Daten aus ungelabelten Hyperspektralbildern zu extrahieren. Dazu wird eine unüberwachte, archetypische Matrixfaktorisierung verwendet, um einen Merkmalsraum zweiter Ordnung zu konstruieren. Der neue Merkmalsraum erlaubt die Früherkennung von Trockenstress auf der Pflanzenebene. Ein hierarchisches Klassifikationsverfahren wird entwickelt um relevante Informationen aus der Signatur zu gewinnen, die durch unerwünschte Einflüsse verdeckt werden. Dies ermöglicht präzisere Regressions- und Klassifikationsmodelle. Die präsentierten Methoden erlauben eine frühere Stresserkennung und eine signifikante Verbesserung in der Quantifizierung von Bodenparametern aus hyperspektralen Daten.

Contents

1 Introduction

Expanding the food production for a rapidly increasing world population without increasing the amount of arable land and resources has been identified as a major challenge for the upcoming decades (Nature Editorial, 2010). One of the most limiting factors is the food production loss due to biotic and abiotic stressors like drought or fungi (Gaspar et al., 2002; Taiz and Zeiger, 2002), calling for crops with increased stress resistance traits optimized for specific climate conditions. However, as traditional breeding methods are too time consuming, considering rapidly changing environmental conditions, one scientific challenge is the optimization of the crop breeding process as well as the development of new methods for resource effective precision plant protection. A bottleneck in understanding plant traits, like yield, under varying environmental conditions is Plant Phenotyping (Furbank and Tester, 2011). Plant phenotypes are the result of the interaction between a plant's genotype and its environment. To understand this complex interaction, non-invasive and fast measurement techniques are essential (Furbank and Tester, 2011). In order to meet the challenges of new, non-invasive sensors, the Institute of Geodesy and Geoinformation in Bonn is cooperating with plant scientists in the context of the interdisciplinary CROP.SENSe.net network.

Of particular importance are hyperspectral sensors, which enable high-throughput, non-invasive measurements of radiative properties of plant tissue. However, the signals are influenced by many parameters, like the illumination, observation angle and the distance to the object, which have a strong, yet not fully researched impact. Regardless, hyperspectral sensors have successfully been used to observe plant characteristics. One important aspect of using hyperspectral sensors in Plant Phenotyping is the description of early stress processes in plants. In this context, early stress detection is defined as the stages of stress which symptoms are not visible to the naked eye. However, for very early stages of stress, classification problems become more demanding as the true probability distribution functions are unknown, labeled samples are often not available and differences between classes become minor compared to the measurement noise. In order to fully understand the interaction between genotypes and their environment, the measurement of soil parameters is required. While hyperspectral sensors have been used to quantify soil parameters, their usefulness is limited in the field, as important soil properties are masked by unwanted parameters, which typically have a high spatial and temporal inhomogeneity.

This thesis shows that advanced methods of machine learning can improve the extraction of information from hyperspectral data considerably. The methods are developed with the focus on plant phenotyping applications, in particular early stress prediction and soil parameter quantification. The thesis is embedded in the CROP.SENSe.net network. CROP.SENSe.net aims to support modern sensor technology for plant phenotyping in order to improve selection efficiency in plant breeding and field management, where, among others, hyperspectral sensors have been a major focus.

Hyperspectral sensors measure radiation reflected or emitted from objects. Hyperspectral reflectance is typically measured in the range from 400 nm to 2700 nm with a spectral resolution between 1 nm and 15 nm. As the electromagnetic energy which is reflected or emitted from objects depends on biological and chemical features, the structural and physiological traits of vegetation (Knipling, 1970; Malenovsky et al., 2009; Rascher et al., 2010) or mineral components of soil samples (Gholizadeh et al., 2013) are observable. Hyperspectral sensors either measure single spectra averaged over a spatial area or, in the case of imaging sensors, the spectrum is recorded for each pixel, resulting in tensors (data cubes) with millions of scalars.

Hyperspectral sensors allow to observe changes of radiative properties of objects like plants or soil non-invasively over time. In remote sensing, hyperspectral imaging has been used to classify vegetation since several decades. Consequently, most data analysis methods for hyperspectral imaging have been developed for remote sensing applications, in which the recorded spectrum is often the reflectance of an area of several square meters.

The most widely investigated approach for hyperspectral data analysis are Vegetation Indices. Vegetation Indices are ratios and / or linear combinations of mostly two different wavebands and aim at reducing the influence of irrelevant background information while maximizing the influence of a specific chemical compound. Vegetation Indices have been developed for the regression of biomass, chlorophyll content or water content, among many others. However, Vegetation Indices have not been developed with hyperspectral image classification in mind. While they are still used in combination with hypothesis tests or linear thresholds to classify images, unknown probability distribution functions and nonlinear classes are difficult to handle for these approaches.

Hence, alternative data analysis algorithms which make less assumptions on the data are needed. Especially kernel based machine learning methods have been used widely in remote sensing applications (Camps-Valls and Bruzzone, 2009; Waske and Braun, 2009; Roscher et al., 2012). Therefore, the question arose whether they would also improve the detection of early stress with hyperspectral images.

The first work for early stress detection with machine learning and hyperspectral cameras was based on Support Vector Machines combined with spectral Vegetation Indices (Rumpf et al., 2009, 2010). The use of linear and non-linear Support Vector Machines and specialized Vegetation Indices enabled more accurate and faster detection of biotic stress than traditional hypotheses tests or a linear discriminant analysis. Especially the high potential of Support Vector Machines for early stress detection has been proven. Due to the restriction to separating hyperplanes with low capacity, Support Vector Machines are comparatively robust against overfitting, even on high dimensional data with relatively few labeled training samples.

However, while the potential of Support Vector Machines for the specific application field was demonstrated, the choice of the right feature space is as important for the success of pattern recognition as the choice of the right learning algorithm (Cherkassky and Lari-Najafi, 1992; Radcliffe and Shurry, 1995; Piramuthu et al., 1998) and many problems have not been addressed yet, in particular:

1. *For early stress detection with noisy sensors, Vegetation Indices and feature selection algorithms are not feasible.*

   The number of labeled samples is normally small, as obtaining ground truth samples is time- and cost-intensive. However, in order to learn a robust classification model under very noisy conditions, either the amount of labeled samples has to increase or the robustness of the feature space against overfitting. A common approach is to use Vegetation Indices as features. Vegetation Indices operate on a very small subset of spectral features. This becomes problematic, as the variance within classes and noise is often higher than the margin between classes in the feature space. Feature selection methods, in addition, are prone to highly correlated feature spaces. For hyperspectral observations, large areas of neighboring wavebands are typically correlated with correlation coefficients above 0.95. Also, the complex physiological effects of stress cause reflectance changes in most spectral regions (Aldakheel and Danson, 1997; Peñuelas et al., 1997). Features based on one or two wavebands, however, have only a very narrow view on the overall changes happening in the spectral bands. Therefore, the question arises whether a feature space using the information of the whole spectrum improves the robustness of the model compared to features based on single wavebands or combinations of a small subset of selected wavebands.

2. *For early drought stress detection with hyperspectral images no labels are provided.*

   Plants typically react to drought with a premature leaf senescence (Lim and Nam, 2007). Leaf senescence is characterized by a redistribution of nutrients and a degradation of pigments. Leaf senescence first occurs in older leaves before affecting the younger, more productive ones (Lim and Nam, 2007; Guiboileau et al., 2010). The premature leaf senescence allows plants to prioritize more productive leaves for nutrient provision in times of drought. This process is continuous and invisible to the naked eye at early stages. Therefore, a manual annotation of pixels measuring early drought stress is not possible.

   The only reliable information is that one plant may have a higher probability of suffering from stress than another. This leads to mainly unsupervised tasks. Traditional unsupervised machine learning algorithms, like K-Means, partition the data into classes. In case of the continuous drought development process, the data is discretized into ordinal classes. However, the quality of the results is dependent on a good discretization. In order to control the quality, an interpretation of these classes is mostly mandatory. However, the interpretation of artificial means is not always easy and straightforward. An alternative method is an archetypal matrix factorization. Archetypal Analysis decomposes a matrix into basis vectors, which are existing data points (i.e. archetypes) and coefficients, which express each point as a linear combination of these archetypes. The question is if archetypal matrix factorization is able to describe stress development in plants efficiently with interpretable results.

3. *The impact of searched processes and parameters on the spectral signature is often masked by unwanted effects and parameters.*

   Irrelevant parameters and environmental influences may overshadow the effects of searched parameters. Often the influence of these irrelevant factors is so strong, that

they effectively mask the searched parameters. The problem is intensified the more parameters are influencing the signature. In order to improve hyperspectral data analysis, the question is if it is possible to reduce the influence of these parameters without the need for additional measurements.

In this thesis, the three problems described before are addressed in specific applications. In all presented applications, the searched processes are neither visible to the naked eye (even for biologically trained users) nor directly visible in the measurements themselves or directly extractable with established data analysis methods.

In the first application, wheat leaves were inoculated with leaf rust under very controlled conditions. Thus, it was exactly known where the symptoms will manifest at a later point in time. However, the task was to detect stress at the earliest possible day after inoculation (dai) with noisy fluorescence sensors. In this case, the mean difference between spectra of healthy tissue and spectra of inoculated tissue was smaller than the variance of both classes. This prevented a successful identification of leaf rust with Vegetation Indices.

In the second application, the task was to detect drought stress with hyperspectral images in barley. While it is known which plant was suffering from water stress, drought does not affect all leaves at the same time. As drought had to be classified before symptoms manifest, no labels were given. The typical approach would be to calculate a mean Vegetation Index for the whole plant and classify with this feature. However, this approach is neither very sensitive to changes in early stages nor does it allow a visualization of drought development in plants.

In the third application, soil organic carbon (SOC) content had to be quantified. SOC is an important environmental parameter for plants. However, moisture and roughness dominate the hyperspectral signature so strong that a direct regression of the SOC content does not meet the accuracy requirements for SOC quantification. Therefore, the need to quantify moisture and roughness from the same signature has been identified. However, until now, no approach was able to determine roughness and moisture simultaneously from hyperspectral signatures and use this information to improve SOC regression.

The data analysis of all applications require new methods in order to quantify and visualize the relevant processes. In particular, this thesis provides:

1. *A method to extract features representing the information of the whole spectrum.*
   For very early stress detection with fluorescence sensors, the low signal to noise ratio prevents the use of Vegetation Indices and filter algorithms. Instead, the signature is mathematically described by curve fitting algorithms. However, global fitting algorithms like splines result either in large fitting errors at the knots or require a high number of coefficients. This often causes errors in areas of high importance for classification, or results in high feature space dimensions, which are not feasible due to the low number of labeled samples available. Instead, the signature is split into smaller areas of high relevance for classification. Then, each area is described by a separate polynomial of low order and the polynomial coefficients are used for classification. This approach enables a classification of leaf rust at the second dai and is more efficient than filter selection methods or low rank approximations with principal components. The project and results are discussed in more detail in section 3.1 and summarize the results of Römer et al. (2011).

2. *A method to extract labels from unlabeled hyperspectral images to describe stress processes.*

   For drought stress detection in hyperspectral images, labeled pixels are not available. Therefore, an unsupervised archetypal matrix decomposition is used. As the optimization process of finding good archetypes is computationally inefficient, traditional archetypal factorization techniques are not suitable for large hyperspectral data sets. Instead, Simplex Volume Maximization (Thurau et al., 2010), a new, computationally fast algorithm for Archetypal Analysis, is used for hyperspectral image analysis for the first time. The archetypes found by the matrix factorization are easily interpretable, as they are existing samples in each image. This enables an easy and fast manual annotation of each archetype. However, this still does not provide enough labeled samples to train a supervised classification model. Instead, the similarities of each sample to the labeled archetypes can be used in order to construct a second order feature space for each plant, expressing the mean similarity to each archetype. This step enables to learn a classification model on the plant level. In addition, the similarities of each pixel to each archetype are used to visualize stress development. The results are described in Römer et al. (2012) and are summarized in section 3.3.

3. *A hierarchical classification approach to deal with unwanted parameters masking the searched, relevant parameters.*

   The influence of moisture and roughness on the signature masks the SOC content. In addition, roughness influences nearly all wavebands. This prevents the search for wavebands which are only affected by SOC. However, moisture has a significant stronger impact than roughness and the influence of roughness is much higher than the influence of SOC. Therefore, the idea is to use a hierarchical tree of Support Vector Machine models. On the first node, moisture is classified. Now, for each moisture class, roughness is determined and subsequently for each moisture and roughness combination the SOC content. This way, first a strong part of the variance from moisture is eliminated from the signature and then the variance from roughness. However, the classification of roughness from the spectral signature is very challenging an has not been achieved before. Therefore, an ordinal classification approach is used instead of a direct multi-class classification model. This enables a quantification of roughness from spectral signatures for the first time. The results summarized in section 3.2 show that the regression error was nearly halved this way.

The remaining chapters of this thesis are structured as follows: In section 2.1, the most important aspects of hyperspectral sensors and hyperspectral sensor evaluation are presented. Furthermore, the boundary conditions for hyperspectral data evaluation and the special challenges offered by hyperspectral sensors are introduced. In section 2.2, the theoretic foundation of Support Vector Machine classification is presented and it is explained why they are well suited for hyperspectral signature evaluation. Archetypal matrix factorization techniques for large data sets and Simplex Volume Maximization are introduced in section 2.3. Conclusions and perspectives are presented in chapter 4.

2 Sensors and Methods

This chapter introduces hyperspectral sensors and the methods most relevant for the thesis.

Section 2.1 shortly introduces hyperspectral sensors and their use for stress detection in plants as well as for soil parameter assessment. The challenges of hyperspectral data evaluation are described and a brief insight into the current state of the art in hyperspectral data evaluation is presented.

Section 2.2 introduces Support Vector Machines, which were chosen for supervised classification in Römer et al. (2011) and Römer et al. (2014). The section explains why Support Vector Machines were chosen for supervised classification of hyperspectral data in both papers.

Section 2.3 introduces Archetypal Matrix Factorization and Simplex Volume Maximization, a variant of Archetypal Analysis specialized for big data applications. Archetypal Analysis with Simplex Volume Maximization was used in Römer et al. (2012) and is fundamental for the understanding of section 3.3.

2.1 Hyperspectral Sensors for Plant Stress Detection and Soil Parameter Quantification

High-throughput and optimally non-invasive sensors capable of detecting and describing plant stress and soil parameters are needed for effective plant phenotyping. Soil has a major influence on plant development as it is one of the most important environmental factors for plants and therefore, in order to fully understand the phenotype-environment interaction, it is mandatory to measure soil parameters effectively. Hyperspectral sensors are capable to fulfill this role. The recorded spectrum is correlated with chemical compounds which can be correlated with the stress factors themselves or the effects stress has on plants. Hyperspectral sensors are divided into active and passive sensors.

Passive hyperspectral sensors acquire narrow, contiguous spectral bands in the visible (400 nm - 750 nm), near-infrared (750 nm - 1400 nm) and shortwave-infrared (1400 nm - 3000 mm) spectrum (Govender et al., 2007) with a bandwidth of typically between 1 nm to 15 nm (Borengasser et al., 2008). They detect the electromagnetic energy reflected or emitted from biological, geological and hydrological features from the earth's surface (Knipling, 1970), respectively from leaves and canopy (Jones et al., 2003; Malenovsky et al., 2009; Ustin and Gamon, 2010).

As figure 2.1 illustrates, the spectrum of solar radiation reflected, absorbed or transmitted from leaves or canopy is influenced by leaf level absorption and scattering (Rascher et al.,

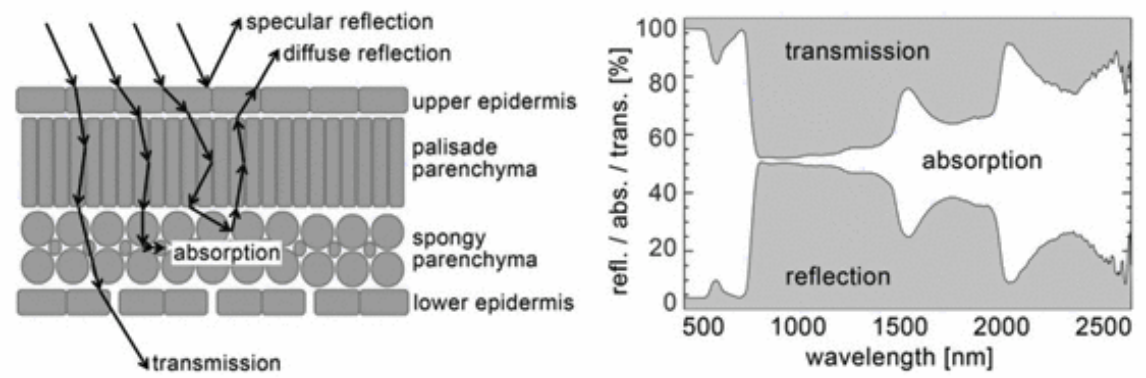

Figure 2.1: Light is either absorbed, transmitted of reflected in plant tissues (left). The resulting spectral characteristics for a healthy winter wheat leaf are shown on the right (Figure from Rascher et al. (2010)).

2010). In addition, canopy architecture and external effects like illumination and observation geometry have a high impact. Most important for stress description is the absorption, which is the result of electron transitions in plant pigments like chlorophyll (Curran, 1989). For vegetation, figure 2.2 gives an overview over some of the most important biochemical compounds which can be measured with passive hyperspectral sensors.

Active hyperspectral sensors measure the light re-emitted from molecules excited by fluorophores. For plant stress measurements, UV light excitation and Chlorophyll a and b are of importance (Malenovsky et al., 2009). Chlorophyll fluorescence was, for example, successfully used to determine the chlorophyll content in plant leaves (Lichtenthaler and Buschmann, 1987; Gitelson et al., 1999).

Hyperspectral sensors are either imaging or non-imaging. Non-imaging sensors measure a single spectrum averaged over a defined area (Mahlein, 2011). Imaging sensors record a full spectrum for each pixel, i.e. the x- and y-axis are referring to the spatial information while the z-axis contains the spectral information (see Figure 2.3).

Spectral measurements of changes in pigments and chemical compounds like water are able to reveal biotic and abiotic stress in plants (Ustin and Gamon, 2010) as stress constrains plant functions like photosynthesis, nutrition and water acquisition (West et al., 1997; Wright et al., 2004).

However, measuring the chemical concentration of compounds is demanding (Banninger, 1989; Curran, 1989). Reasons for this are interference between absorption features due to scattering effects in the plant tissue and the fact that wavebands are never uniquely related to only one compound. Another limiting factor is that absorption at the most correlated wavelengths is quickly reaching saturation at relatively low concentrations for hyperspectral measurements. Furthermore, while absorption is correlated with pigments and certain chemicals, the reflectance is influenced by other factors like canopy architecture, inclination, illumination and background (Suárez et al., 2008). Due to all these interference factors, hyperspectral measurements are noisy and may significantly change between different measurement setups even when observing the same object.

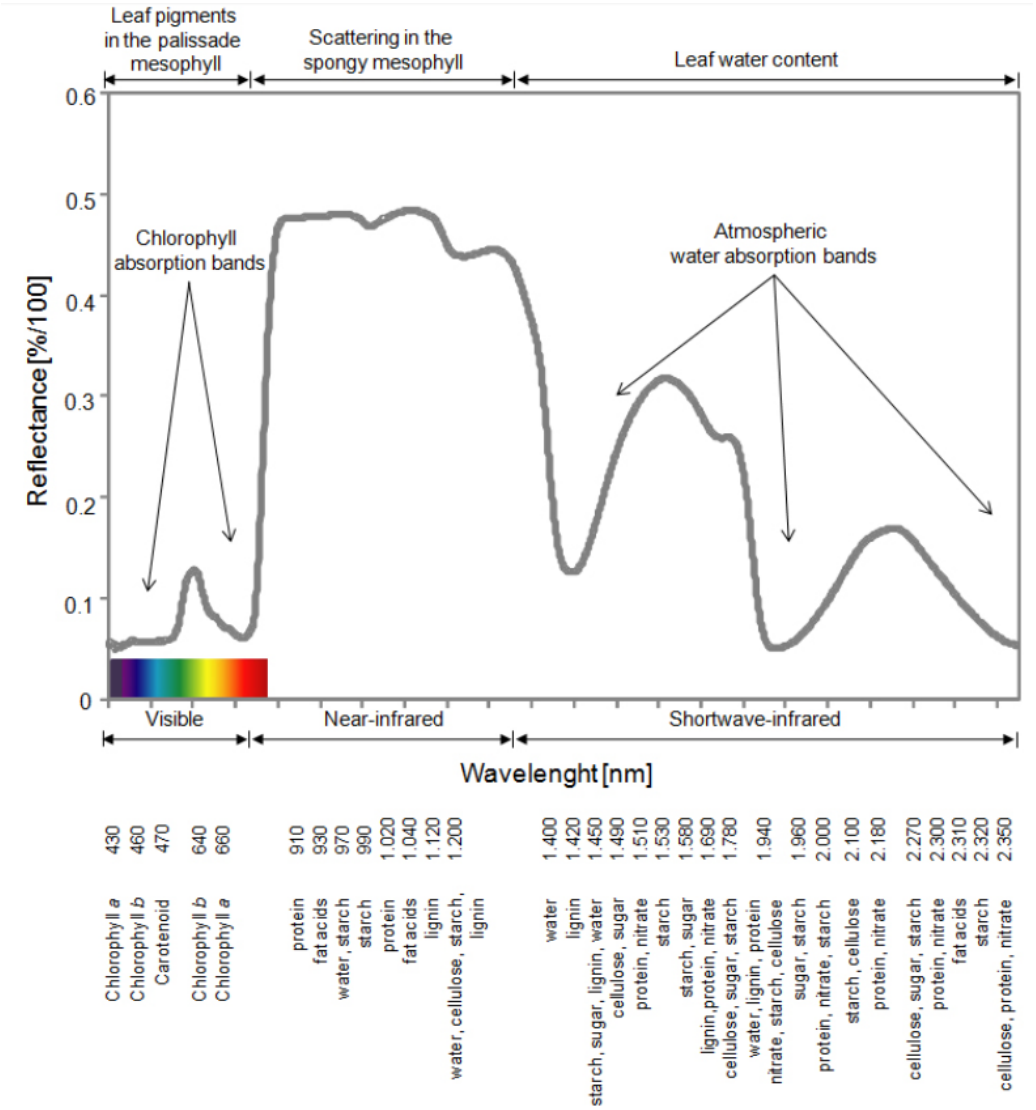

Figure 2.2: A typical hyperspectral reflectance spectrum of vegetation is shown in grey from the visible to the shortwave-infrared spectrum. In the visible spectrum, die colors corresponding to the different bands are depicted. Below, the biochemical plant compounds mostly influencing the specific area of the spectrum are listed (Figure from Mahlein (2011)).

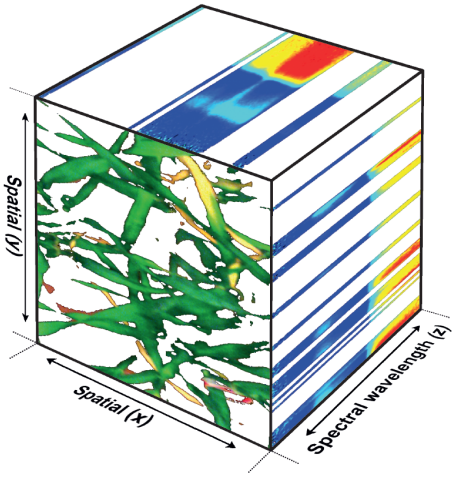

Figure 2.3: Hyperspectral tensor (data cube) resulting from an imaging sensor. The x- and y-axis correspond to the spatial information, whereas the z-axis contains the spectral information. Here, white pixels are removed background information (Figure from Behmann et al. (2014)).

For plant phenotyping, the knowledge about the quality of soil is essential to understand crop yield, the need for nutrients or a plant's ability to withstand stress factors (Pätzold et al., 2008). The quality of soil is dependent on its chemical and biological characteristics like soil moisture, structure or soil organic matter. However, assessment of these parameters for soil is very time and cost intensive, as soil parameters have a high variance in time and space, even on small scales within a single field. Hyperspectral sensors in the visual and near infrared spectrum, as well as in the mid infrared spectrum, provide the opportunity to assess a broad range of parameters from soil. In soil sciences, hyperspectral sensors have successfully been used to quantify these parameters (Viscarra-Rossel et al., 2006).

In order to analyze hyperspecral data Vegetation Indices are normally used (Jackson and Huete, 1991; Bannari et al., 1995; Fiorani et al., 2012). The aim of Vegetation Indices is to enhance the signal of the searched chemical compounds while minimizing the influence of background signals like unwanted compounds or illumination. According to Jackson and Huete (1991), Vegetation Indices can be differentiated into ratios and linear combinations. Ratios are relatively simple functions of mostly two wavebands and may consist of differences, sums or products of any number of observed spectral bands. Linear combinations are orthogonal sets of linear equations calculated from spectral bands. A typical example is the use of Principal Component Analysis (PCA) to transform the spectral dataset (Pearson, 1901; Jolliffe, 2002).

For classification, Vegetation Indices normally have been used with multivariate hypothesis tests. However, hypothesis tests are restricted to deal with linearly separable classes. Therefore, in remote sensing, machine learning methods like Neural Networks or Support Vector Machines have lead to superior classification results in recent years (Camps-Valls and Bruzzone, 2009). For crop stress detection with hyperspectral sensors the use of machine

learning for classification, regression and unsupervised learning is scarce. Few papers have revealed their potential (Karimi et al., 2006; Wu et al., 2008; Liu et al., 2010; Behmann et al., 2015). Especially discriminative methods like Support Vector Machines, which make no assumptions on the data distribution, can deal with linear and non-linear applications and usually have lead to better classification performances than generative methods. The enhanced performance of Support Vector Machines lead to the prediction of stress before any symptoms were visible yet to the naked eye (Rumpf et al., 2010). For soil parameter estimation, machine learning methods have successfully been used in several studies and it was concluded that Support Vector Regression is promising to improve the quality of data evaluation significantly (Gholizadeh et al., 2013).

2.2 Support Vector Machines

The relevance of Support Vector Machines (SVMs) for hyperspectral stress detection and soil parameter estimation has been shown by several authors (Rumpf et al., 2010; Römer et al., 2011; Gholizadeh et al., 2013; Behmann et al., 2014; Römer et al., 2014; Behmann et al., 2015). In this section, the most important aspects of SVMs and why they were chosen for hyperspectral classification in this thesis are discussed.

SVMs belong to the group of supervised learning algorithms. For supervised learning, a part of the data is labeled, i.e. the class membership of the instances of the training data is known beforehand. In this section, all classification problems are binary to keep the notations simple. The notation of the equations are based on Schölkopf and Smola (2001).

For binary pattern recognition, a training dataset

$$(\vec{x}_1, y_1), \ldots, (\vec{x}_m, y_m) \in \chi \times \{\pm 1\}, \tag{2.1}$$

is given, where χ is the domain from which the instances $\vec{x}_i$ are taken and y_i is referred to as labels which are either $+1$ or -1. The decision to which class a sample $\vec{x}_i$ belongs is given by the unknown joint probability function $P(\vec{x}, y)$, also known as the generator function.

In order to solve the classification problem, generative and discriminative approaches are distinguished here.

The *generative* approach is to estimate the parameters of the joint probability function $P(\vec{x}, y)$ to predict y for new values of $\vec{x}$ (Bishop and Laserre, 2007). In order to perform well on unseen data, the training data has to be a good representation of the underlying distribution function $P(\vec{x}, y)$. However, for high-dimensional data like spectral signatures, the amount of labeled data needed for a significant statistical representation grows exponentially with increasing dimension (Bellman, 1961). Or, as Vapnik (1995) states, in order to estimate high-dimensional co-variance matrices sufficiently, an unpredictable number of observations is needed.

For stress detection, especially at an early stage and for close range measurements, obtaining this critical mass of training data is problematic. Measuring, growing and inoculating plants, as well as manual labeling in hyperspectral images, is very time and therefore cost intensive. This typically leads to classification problems with relatively few samples compared to the high amount of features recorded by hyperspectral sensors.

Instead of estimating the unknown joint probability function $P(\vec{x}, y)$ it is easier to construct a good predictor function f directly from the training data. This leads to the so called *discriminative* models (Figure 2.4). The task is now to find a classification function

$$f : \chi \to \{\pm 1\}, \tag{2.2}$$

minimizing the empirical risk of misclassification on the training dataset

$$R_{emp}(f) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} |f(\vec{x}_i) - y_i|, \tag{2.3}$$

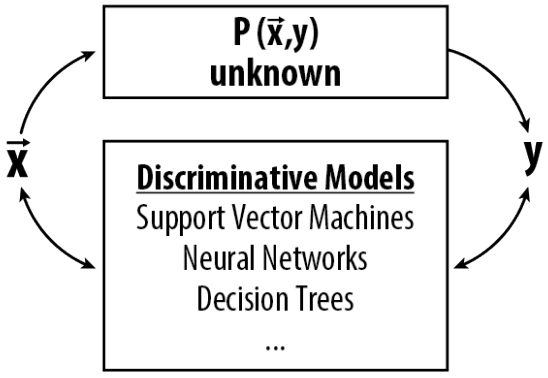

Figure 2.4: The class membership of a vector $\vec{x}$ is generated by the unknown joint probability function $P(\vec{x}, y)$. While generative models try to estimate the parameters of the probability function and predict y in this way, discriminative models learn a new decision function $f(\vec{x})$, which is a black box and does not need $P(\vec{x}, y)$. Modified from Breiman (2001).

in order to approximate the real risk function based on the underlying distribution $P(\vec{x}, y)$,

$$R(f) = \int \frac{1}{2}|f(\vec{x}) - y| dP(\vec{x}, y). \tag{2.4}$$

However, estimating f solely based on the empirical risk function 2.3 induces the risk of overfitting, i.e. prediction on random patterns and noise instead of estimating a good predictor function for $P(\vec{x}, y)$.

For hyperspectral stress detection, the number of features is typically larger than the number of samples. This means that the risk of overfitting is large for hyperspectral data evaluation.

Therefore, following the *Occams Razor* principle, stating that among many possible solutions, the one with the fewest assumptions should be taken, the empirical risk function is extended by a measurement of the complexity of the set of classification functions allowed to solve the task. This is done by a penalty term in order to increase robustness with regard to overfitting in equation 2.3

$$R(f) \leq R_{emp}(f) + \phi(F, m), \tag{2.5}$$

where ϕ is the capacity term of the set of functions F from which the predictor function f is taken. The capacity of a set of indicator functions is measured with the VC-dimension h (Vapnik and Chervonenkis, 1971):

The VC-dimension h is defined as the maximum number h of vectors $\vec{x}_1 \ldots \vec{x}_h$ that can be separated into all 2^h possible ways using functions of the set F.

The capacity term can now be defined as

$$\phi(h, m, \delta) = \sqrt{\frac{1}{m}\left(h\left(ln\frac{2m}{h}+1\right)+ln\frac{4}{\delta}\right)}, \tag{2.6}$$

where δ is the confidence bound of the probability $1-\delta$ with which the bound 2.5 should hold (Vapnik, 1995, 1998; Schölkopf and Smola, 2001).

This so called structural risk minimization balances the empirical risk minimization (equation 2.3) with the capacity of the set of indicator functions and the size of the training data set. To reduce the structural risk $R(f)$ either the size m of the sample set has to increase or F has to be restricted to a set of smaller capacity h.

Therefore, in order to control the statistical effectiveness of the learning algorithm, SVMs restrict the set F to separating hyperplanes whose VC-dimension can easily be computed (Schölkopf and Smola, 2001).

The restriction to hyperplanes limits the possible number of class separations with SVMs. However, this greatly enhances the probability that the results are not overfitting if they work well on the training data. This leads to the superior generalization ability typical for Support Vector Machines and discriminative classifiers in general (Breiman, 2001).

The good generalization ability is one of the reasons SVMs are performing well on hyperspectral data. The restriction of the decision functions enables to deal with problems where the number of labeled samples and the number of features is imbalanced, as they are statistically less prone to overfitting, and the resulting high generalization ability suits SVMs to ignore the large variability and noise typical for hyperspectral measurements to some degree.

The hyperplanes are defined in the dot product space κ as

$$\langle \vec{w}, \vec{x} \rangle + b = 0, \vec{w} \in \kappa, b \in \mathbb{R}. \tag{2.7}$$

The indicator function is now

$$f(\vec{x}) = sgn(\langle \vec{w}, \vec{x} \rangle + b). \tag{2.8}$$

The optimization problem is to find the optimal large margin hyperplane on the training data set:

$$\text{minimize } \frac{1}{2}||\vec{w}||^2, \tag{2.9}$$

$$\text{subject to } y_i(\langle \vec{w}, \mathbf{x} \rangle + b) \geq 1, \forall i. \tag{2.10}$$

To solve the optimization problem, the dual Lagrangian problem is solved using the Karush-Kuhn-Tucker conditions

$$\text{minimize } W(\vec{\alpha}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle \vec{x}_i, \vec{x}_j \rangle, \tag{2.11}$$

$$\text{subject to } \alpha_i \geq 0 \ \forall i \text{ and } \sum_{i=1}^{m} \alpha_i y_i = 0. \tag{2.12}$$

As the dot product $\langle \vec{x}_i, \vec{x}_j \rangle$ in equation 2.11 is a kernel, both linear hyperplanes and nonlinear kernel decision functions can be realized with Support Vector Machines:

$$\text{minimize } W(\vec{\alpha}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j k(\vec{x}_i, \vec{x}_j), \tag{2.13}$$

where k is a function fulfilling the kernel conditions (Shawe-Taylor and Cristianini, 2004).

The kernel trick enables an implicit transformation in a higher dimensional feature space. As higher dimensional feature spaces are less dense, each data set becomes linear separable with increasing feature space dimension. While this allows SVMs to classify non-linear data, controlling the capacity term becomes harder and SVMs loose some of their generalization power. Hence, while for example radial basis function kernels nearly always exceed the classification performance of linear kernels on the training data set, they have to be used with care to keep the good generalization abilities of SVMs which are an important factor for the use of SVMs for hyperspectral data evaluation.

Summarized, hyperspectral measurements often result in datasets with an imbalanced ratio between samples and features. At the same time, the measurement noise is high especially compared to the number of training samples. This poses problems for generative models, as the estimation of the probability density function is very challenging for high dimensional classification tasks. Discriminative models are better suited to deal with these tasks, as they do not try to estimate the generator function. Instead they learn an indicator function to predict the correct labels for each data point. Therefore, the results of discriminative models lack interpretability, but are able to cope with less labeled samples to achieve the same prediction accuracy. As hyperspectral measurements are influenced by many not yet fully explained factors, they usually tend to be noisy in addition. The structural risk minimization principle of Support Vector Machines strongly restricts the set of possible indicator functions to separating hyperplanes for linear classification problems. This greatly increases reliability and robustness of the learned models. It further reduces the amount of labeled samples needed in order to explain high dimensional data sets.

2.3 Archetypal Matrix Factorization

A popular method to approximate a high-dimensional feature space to a lower dimension is a matrix factorization of the form

$$X \approx WH, \tag{2.14}$$

where $X \in R^{m \times n}$ is the $m \times n$ input matrix. The $c \in \mathbb{N}$ columns of $W \in R^{m \times c}$ are called basis vectors and the reconstruction matrix $H \in R^{c \times n}$ contains the coefficients forming a linear combination with the basis vectors to reconstruct the input matrix.

PCA is a commonly used matrix factorization method for the construction of vegetation indices and low-rank approximation of hyperspectral images (chapter 2.1). The basis vectors can statistically be interpreted as the direction of the largest variance in feature space (Lee and Seung, 1999). The singular value decomposition underlying the PCA makes no restriction on the sign of H and W and the feature matrix X is unconstrained (Ding et al., 2010), i.e.

$$X_{\pm} \approx W_{\pm} H_{\pm}. \tag{2.15}$$

However, due to the arbitrary sign of the entries in H and W, the decomposition mostly allows for no obvious visual interpretation of the coefficients and the basis vectors are not corresponding to any data points (Lee and Seung, 1999). This lack of interpretability is a serious drawback for the low rank approximation of high dimensional spectral signatures. An interpretable decomposition would allow for better spectral unmixing models. Furthermore, the communication of results is much easier if decompositions are intuitively understandable by non-experts.

A variant of PCA is the Archetypal Analysis (Cutler and Breiman, 1994). For Archetypal Analysis, the basis vectors are restricted to be mixtures of samples in the input matrix or to be samples themselves (Stone and Cutler, 1996). As each basis vector is either a positive, weighted sum of samples or a sample itself, it is easier to interpret the decomposition and assign a physical meaning to the coefficients of W.

A typical Archetypal Analysis decomposition is the convex-non-negative matrix factorization (Convex-NMF). For Convex-NMF the basis vectors need to lie within the column space of X, i.e.

$$W = XG. \tag{2.16}$$

Equation 2.16 is restricted to convex combinations only, i.e. they follow the constraint

$$\|g_i\|_1 = 1, g_i \geq 0. \tag{2.17}$$

Therefore, each basis vector fulfills the conditions for Archetypal Analysis, i.e. they are either samples or positive sums of samples themselves (Ding et al., 2010). Since additionally only positive combinations of basis vectors are allowed in H, no subtractions can occur and

Convex-NMF follows the intuitive combination of parts of basis vectors to form a whole (Stone and Cutler, 1996), i.e.

$$X_{\pm} \approx X_{\pm} G_{+} H_{+}^{T}. \tag{2.18}$$

The resulting optimization problem of minimizing $X - X_{\pm} G_{+} H_{+}^{T}$ with respect to the constraint in equation 2.17, is solvable by an iterative update rule. The starting solution can, for example, be given by a prior K-Means (Bishop, 2006). However, according to Thurau et al. (2009), while this matrix decomposition improves interpretability, it has some problems coming with large data sets (which are typical for hyperspectral imaging sensors):

- the update rules have a complexity of $O(n^2)$,
- the update rule needs a matrix multiplication of $X^T X$, which is not feasible for large sample matrices.

For large scale data analysis, Convex Hull NMF (CH-NMF) was introduced by Thurau et al. (2010). CH-NMF restricts basis vectors in H to lie on the convex hull of X. Thus, beside the constraint 2.17 and the requirement for H to be positive, an additional constraint is added to the minimization problem of C-NMF:

$$\|h_i\|_1 = 1, h_i \geq 0. \tag{2.19}$$

However, while this approach may strongly reduce the number of candidates for basis vectors, calculating the convex hull has a high complexity of $O(n^{\frac{m}{2}})$ and the number of samples on the convex hull is large for high dimensional data. To solve this problem, Thurau et al. (2009) approximate the convex hull by pairwise combinations of eigenvectors of the covariance matrix.

As a faster alternative to the CH-NMF computation of the convex hull, Simplex Volume Maximization (SiVM) aims at finding the basis vectors through spanning the simplex with maximal volume over the data points in X (Thurau et al., 2010). For this, the distance between each basis vector and each data point has to be computed. However, as the number c of basis vectors is normally significantly smaller than the number of samples n ($c << n$), the computational complexity can be considered as scaling linearly with the number of samples and, in addition, the pairwise distance computation is easily parallelized.

The linear runtime of SiVM makes Archetypal Analysis feasible even for large hyperspectral measurement campaigns. Another interesting property is that all archetypes lie on the convex hull of the dataset. This is useful for the understanding of stress development, as each sample can now be unmixed in, for example, very stressed and very healthy pixels. As the coefficients have to sum up to one, they are good measures of similarity. Both have the benefit that similarities to extrema are very intuitive to understand and interprete, even for non-experts, which is an important property for an interdisciplinary research field.

As the coefficients have to be positive, only samples within the simplex spanned by the archetypes can be described correctly. All points outside of the simplex are unmixed with an error proportional to their distance to the simplex. This means that subsampling is not recommended for simplex calculation. Another critical point are outliers. For the algorithms to perform well, it is mandatory to either deal with outliers beforehand or choose the archetypes in a more robust way.

3 Most Important Findings

This chapter summarizes and discusses the most important findings of the manuscripts appended to this thesis. Each of these manuscripts is presented in a separate section. Section 3.1 outlines the mathematical description of the whole spectrum through low order polynomials to improve the robustness of feature space. This leads to better classification results compared to Vegetation Indices and classical feature selection algorithms. The method is presented for fluorescence measurements of wheat leaf rust. Section 3.2 presents a hierarchical classification tree used to deal with unwanted parameters masking the searched, relevant parameters. The method was developed for soil organic carbon detection under varying moisture and roughness. Section 3.3 shows in which way Archetypal Analysis can be used to extract information from unlabeled hyperspectral images in order to quantify stress and to visualize early stress processes. Archetypal Analysis was used to detect drought stress in barley at an early stage of development.

3.1 Polynomial Features for Early Stress Prediction

Recently machine learning methods like Support Vector Machines have been used in combination with Vegetation Indices for an improved classification accuracy (Rumpf et al., 2009, 2010). However, for stress prediction at the earliest possible date, the differences between spectra caused by healthy tissue and stressed tissue of plants become marginal. In addition to the normal measurement noise, hyperspectral sensors are influenced by many different factors (section 2.1). This leads to situations where the difference between the means of both classes are smaller than the standard deviation of all samples within one class due to noise and the natural biologic variability of plants (see figure 3.1). As Vegetation Indices only use typically 2-3 different wavebands, they are very prone to noise. Therefore, the idea presented in this section is to use the information of the whole hyperspectral signature by polynomial approximation. The polynomial coefficients are then used as features for a Support Vector Machine model to predict stress at an early stage.

This section presents the main findings of Römer et al. (2011). The underlying dataset consists of 36 wheat leaves (*Puccinia triticina*) inoculated with leaf rust and 36 healthy control leaves. Spectral fluorescence measurements were conducted on dai 2, 3 and 4. Measurements were done under laboratory conditions with a point measurement sensor. The sensor's spectral range was from 370 nm to 800 nm with a spectral resolution of about 2 nm, leading to 215 spectral features recorded per measurement[1].

[1] The data discussed in this section is from Kathrin Bürling, Mauricio Hunsche and Georg Noga, INRES, University of Bonn and was not measured by myself. For more details regarding the measurement setup, please see Römer et al. (2011).

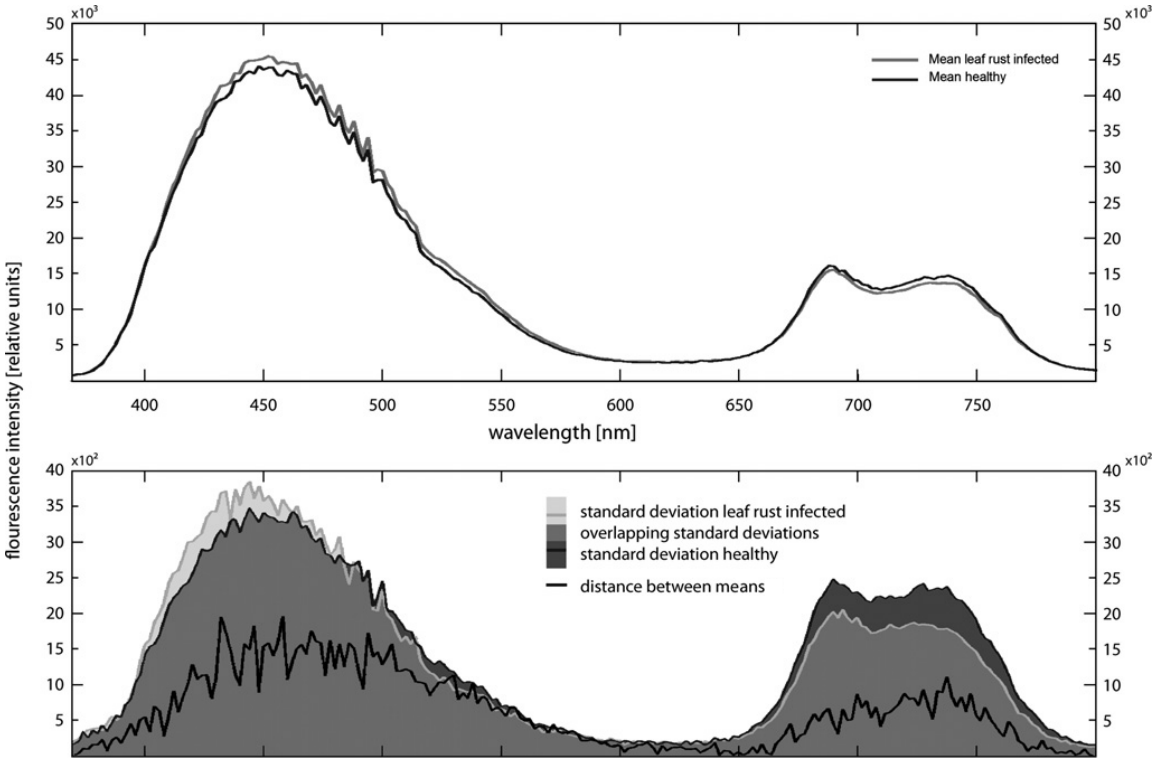

Figure 3.1: Top: Shown is the mean spectrum of healthy leaves and the mean spectrum of leaf rust infected leaves. Bottom: The black line shows the difference between the means of healthy and inoculated leaves. Shadowed are the standard deviations of both classes.

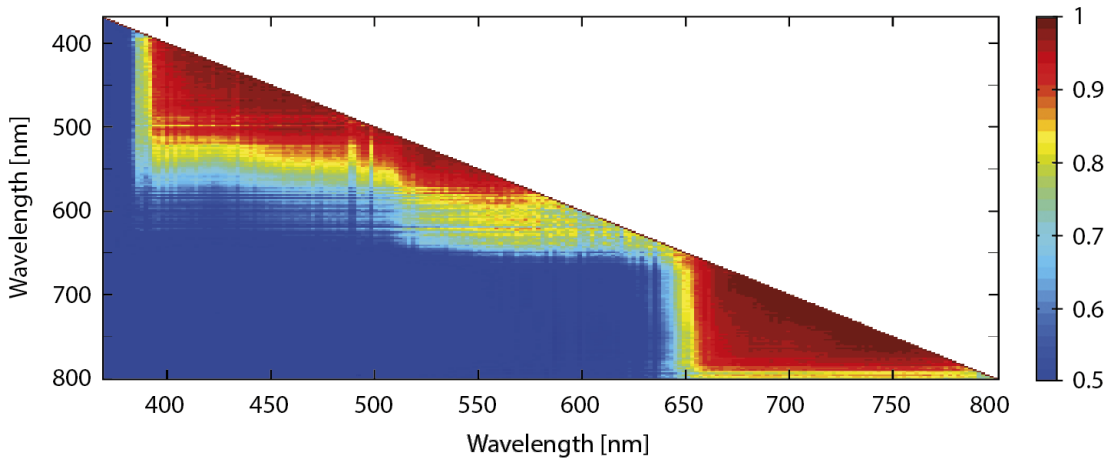

Figure 3.2: Correlation coefficients between wavebands of the fluorescence data. Two large areas of highly correlated wavebands can be observed, with correlations above 0.95.

Using all features for classification poses a high risk for overfitting. Typically, the number of features is reduced either by using Vegetation Indices, wrapper or filter algorithms (Guyon and Elisseeff, 2003). In the following filter and wrapper algorithms are briefly discussed and it is outlined why they are not an optimal choice for hyperspectral signatures.

Filter algorithms like Relief (Kononenko, 1994) are a popular choice to reduce feature space dimension to a subset of most relevant features. They use general characteristics of the training set to select a feature subset and therefore are independent from the classifier that will use their output (Blum and Langley, 1997).

However, filter algorithms have problems dealing with highly correlated feature spaces (Guyon and Elisseeff, 2003). Chemical compounds like chlorophyll have an impact on a broad area of different wavebands. This results in the problem of spectral sensors that large areas of neighboring wavebands are strongly correlated with correlation coefficients of typically above 0.95 (Figure 3.2). As filter algorithms normally weight each feature independently, they do not take correlation into account (Kohavi and John, 1997). While two neighboring and dependent features may achieve high rankings with the filter algorithms, the redundancy of the information contained reduces their combined relevance for a classification model. In the wrapper approach, the feature subset is determined using the classifier itself. The best subset is selected due to the classifiers performance on it (Kohavi and John, 1997; Guyon and Elisseeff, 2003). However, the selection of single features bears the risk that the good separability in those features may be random due to the bad ratio of labeled samples to features compared to the noise.

Instead of using wrapper and filter algorithms, the evaluation of the shape of the whole signature may improve classification results. While it is not advisable to simply use all recorded wavebands, the idea is the mathematical description of the shape of the whole

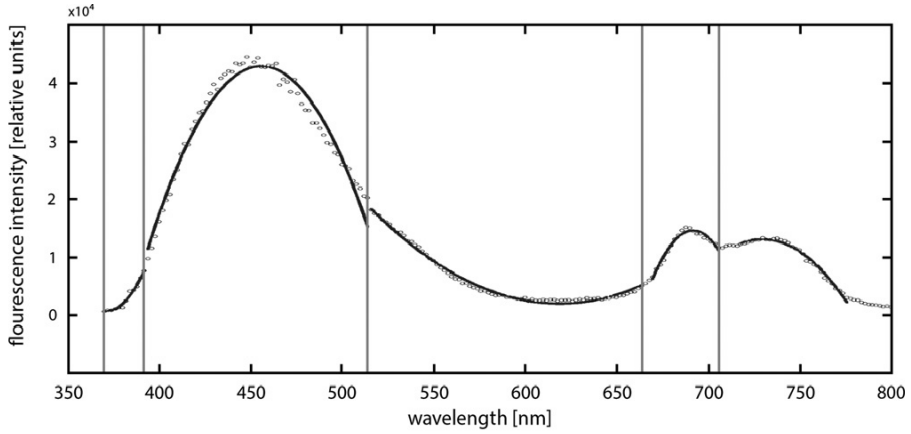

Figure 3.3: Approximation by piece-wise polynomials. The grey lines segment the spectrum into areas approximated by different polynomials into areas of high relevance for classification. Each area is then approximated by a separate polynomial of low order.

spectral signature. The construction of polynomial features describing the shape of a certain area of the spectrum has a number of advantages:

1. Polynomial coefficients describe the local properties of the area they approximate and compress them into a comparably small number of features. In this way, less information is dismissed.
2. Polynomial coefficients are less sensitive to random noise deviations, as they are smoothed over a broader area.
3. Highly correlated areas are summarized into a small subset of features which are less correlated.

An obvious approach would be to use splines (DeBoor, 1978) to approximate the whole spectrum. However, the disadvantage of splines is the significantly increasing approximation error at the knots. When knots are placed in areas with a high relevance for the classification problem at hand, important shape features are not described precisely enough. At the same time a higher order would increase the number of features while a higher knot density would have decreased the smoothing effect of polynomials as the approximated areas become smaller.

Instead, the focus for the approximation is not a minimized global approximation error, but an accurate representation of a spectrum's characteristics in areas relevant for classification. Therefore, the hyperspectral signature is split at areas of low relevance for classification. This enables an easy approximation with piece-wise polynomials of low order (Figure 3.3), where each polynomial represents an area of high relevance for the classifier.

In order to find good areas for polynomial approximation, areas of high and low relevance have to be determined. As the weight of a whole area is of importance, a filter algorithm is usable again, as not the relevance of a single feature is the aim, but the sum of the feature

Dai	3rd order (%)	4th order (%)	5th order (%)	6th order (%)
2	84.22	93.05	88.89	91.67
3	87.50	88.89	84.72	90.28
4	91.67	90.28	87.50	79.17

Table 3.1: Classification accuracy in dependence of different polynomial orders using SVMs for the first 4 dais. Polynomials of 4th order performed best on dai 2. Higher orders were not used, as they would result in both unstable feature spaces and high feature space dimensions.

weights in a certain range of wavebands. This was done with Relief, as it is computationally fast and suitable of ranking linear and non-linear classification problems. Now polynomials can be optimized in a way that they do not split areas of high importance, while minimizing the geometric approximation error and the polynomial order at the same time.

For classification SVMs are used with a linear kernel as it performs better than a RBF kernel. In order to train the SVM model, a ten-fold cross validation with the LibSVM (Chang and Lin, 2011) toolbox is used. As a further pre-processing step polynomial coefficients are normalized to a mean of zero and a standard deviation of one. In addition, as the coefficients are significantly lower correlated now than the original features, Relief is used in order to weight the polynomial coefficient features. The polynomial order is optimized during the classification process.

Table 3.1 shows the results for the SVM models on dai two, three and four with differing polynomial orders. The results show that it is possible to detect leaf rust on the 2nd dai reliably with 4th order polynomials with 93% accuracy. For comparison, a SVM model was learned with features extracted from a PCA and with the best features filtered by Relief. Table 3.2 reveals that polynomial coefficients significantly improve the classification accuracy compared to both algorithms.

The results demonstrate that features describing the information of the shape of the whole spectrum, combined with SVMs, lead to a very early leaf rust prediction only two days after inoculation. They perform significantly better than low rank approximations with PCA or a filter algorithms (Relief) to select the best features (Table 3.2). However, important is not a perfect geometric approximation of the whole hyperspectral signature, but to describe those areas optimal which are relevant for classification. Therefore, piece-wise polynomials are used, where each polynomial is placed in a segment of high relevance. The polynomial order is also not optimized with regard to minimizing the geometric approximation error, but instead chosen to maximize classification accuracy on the one hand and to keep the number of coefficient features as low as possible on the other hand.

	Polynomial (%)	Principal components (%)	Best 20 features (%)
Dai 2	93	72	72

Table 3.2: Comparison of different feature space representations. The features are based on piece-wise polynomials of 4th order are compared to a low rank approximation with PCA and the best 20 features from a Relief algorithm. 20 features are selected as it equals the sum of polynomial coefficients to enable a better comparison between both approaches.

3.2 Hierarchical Classification Models for Soil Parameter Quantification under varying Moisture and Surface Roughness

Soil heterogeneity causes a large part of the heterogeneity of plant phenotypes. High organic matter, for example, positively influences plant stress resistance towards diseases (Pätzold et al., 2008). The measurement of soil parameters is therefore important to understand the performance of different plant phenotypes. According to O' Rourke and Holden (2011), hyperspectral imaging has the potential to significantly reduce the monetary cost compared to traditional chemical analyzing methods. This enables a higher spatial and temporal measurement resolution and explains the high interest in hyperspectral soil parameter assessment.

However, spectral signatures are influenced by numerous factors. Beside the strong influence of the measurement setup (e.g. illumination, observation angle, distance to the object), physical and chemical properties of the observed object have a significant impact on the reflectance. As discussed in section 3.1 and section 2.1, wavebands are highly correlated over large areas of neighboring wavebands and most often related to many different chemical compounds. If the unwanted parameters have a higher impact on the spectral signature and influence similar wavebands, the searched parameters are effectively masked. A practical solution would be to use a very controlled measurement setup with comparable soil samples in order to rule out the influence of unwanted parameters.

For field measurements, this becomes even more problematic. For soil, the most important properties are soil texture, soil moisture, roughness and vegetation cover (Stevens et al., 2006), where soil moisture and vegetation cover are the most limiting factors (Cécillon et al., 2009). Kooistra et al. (2003) concluded that a multivariate method is needed, which extracts the relevant information from soil spectra and that simultaneously is able to deal with these interfering factors.

In this context, surface roughness classification is especially challenging. While Wu et al. (2009) were able to correlate spectral reflectances with surface roughness, they also concluded that changing moisture was influencing the features needed for roughness classification more than roughness itself, and therefore effectively masking the roughness parameter.

In this thesis, the problem of masking parameters is solved with a hierarchical approach. A hierarchical classification tree is build with SVMs in order to classify moisture and roughness and use the information to learn a moisture and roughness specific Support Vector

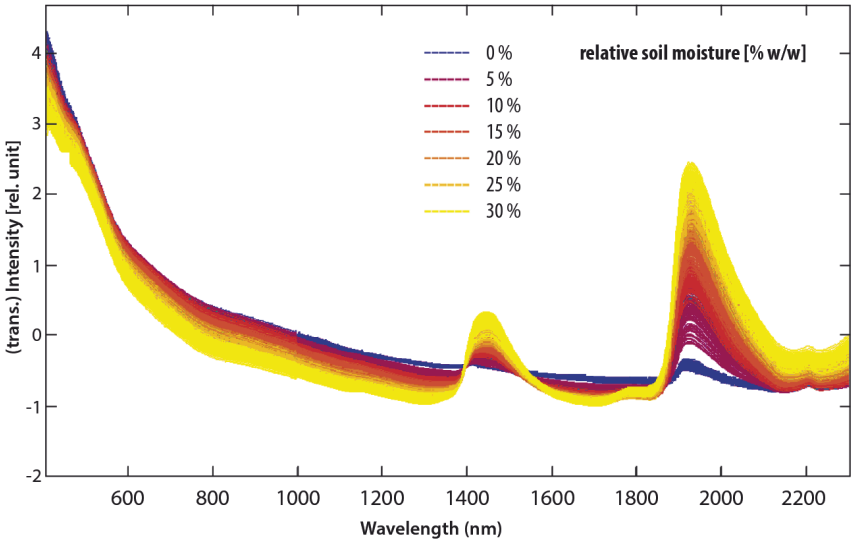

Figure 3.4: Shown is the influence of moisture on the hyperspectral signature. While mainly the areas from 1440 nm to 1550 nm and 1900 nm to 2100 nm are influenced by the relative water content, moisture has clearly an impact on the whole spectrum.

Regression model to estimate SOC. Due to this method, it is possible to classify surface roughness under varying moisture with hyperspectral sensors for the first time, as described in Römer et al. (2014).

The method is applied on a dataset[2] of 1046 samples in total. Air-dried samples were sieved into aggregations of sizes <2 mm, 2-16 mm and 16-25 mm aggregate sizes. The samples were stepwise wetted to 5, 10, 15, 20, 25 and 30% w/w (water / weight) (figure 3.4). The samples were recorded with a full range VIS-NIR spectrometer from 410 nm to 2300 nm. The SOC content varied between 7 g/kg and 12 g/kg.

The spectrum is approximated by polynomials as described in section 3.1. The resulting polynomials were placed between the area of 410 nm -1440 nm, 1450 nm - 1550 nm, 1560 nm - 1900 nm, and 1910 nm - 2300 nm, with polynomial order ranging from 4 to 8.

For reference, SOC is first quantified without a previous moisture or roughness classification. For regression, a Support Vector Regression model with RBF kernel is used. The model parameters are learned using a ten-fold cross validation. The resulting root mean square error after cross validation (RMSECV) is 0.91 g/kg.

To analyze the influence the different factors (SOC, moisture, and roughness) have on the signature, the correlation coefficients between the wavebands and the three factors are calculated first. The correlation coefficients for moisture are between 0.65 and 0.85, with the maximum in the waterbands (Figure 2.2). For roughness, nearly the whole spectrum

[2]The data is from Andrei Rodionov, Stefan Pätzold and Gerhard Welp, INRES, University of Bonn and was not measured by myself. For more details regarding the measurements please see Römer et al. (2014).

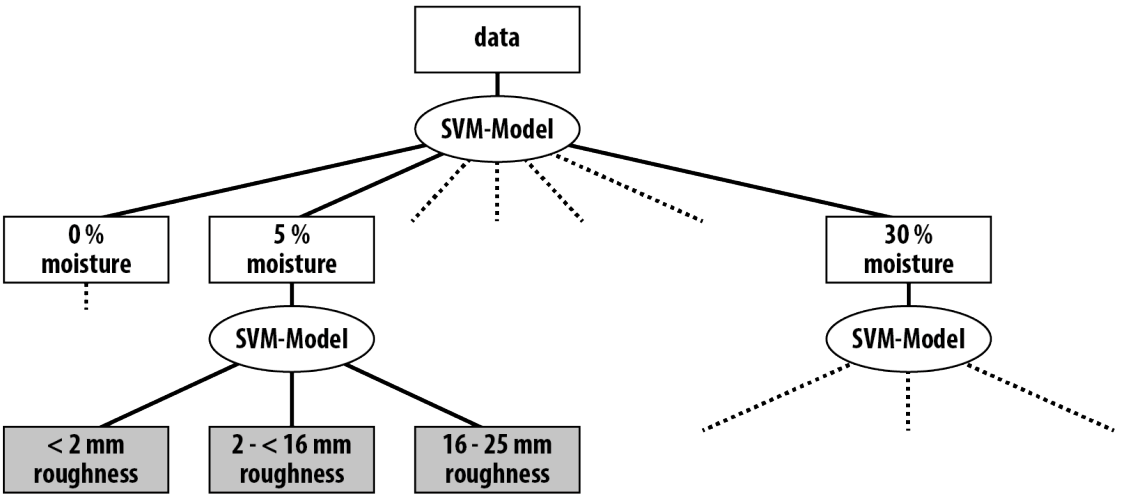

Figure 3.5: Ilustration of the hierarchical classification tree. First, a SVM model is learned to quantify moisture from the whole dataset. Second, for each moisture class a seperate SVM model classifies roughness.

reveals a correlation coefficient of 0.6. Firstly, there are no wavebands specifically suitable for roughness quantification, as roughness affects the whole spectrum equally. Secondly, moisture has a more dominating influence than roughness, especially in those bands typical for water. For SOC prediction, the correlations within the visual wavebands are two times higher than in the near-infrared, but about 50% smaller than moisture.

On the one hand, this means that there is no single waveband where SOC or roughness have a stronger impact on the signature than moisture, and that roughness has always a larger influence on the spectrum than SOC.

On the other hand, it can be concluded that moisture, roughness and SOC follow some sort of taxonomy. As moisture dominates the signature, the moisture content can be approximated relatively well despite the varying roughness and carbon. If the moisture content is then known, the signature is mostly influenced by roughness and carbon, eliminating the variance from moisture in the signature. Thus, as roughness again has more influence on the signature than the remaining SOC, the process can be repeated for roughness. SOC is then classified on predicted moisture / roughness combinations (see figure 3.5). This effectively eliminates the influence of moisture and roughness from the spectrum. This approach is related to hierarchical classification (Silla and Freitas, 2011).

Hierarchical classification uses the known natural taxonomy of classes to decompose the problem into a set of smaller problems. This corresponds to splits into a tree of classifiers (Dumais and Chen, 2000). At each leaf of the tree an independent classifier is learned specified for the smaller sub-problem. Each sub-problem can now be solved more efficiently than a flat all-in-one multi-class approach (Dumais and Chen, 2000; Dekel et al., 2004). For the detection of soil carbon, instead of using a natural taxonomy defined by related classes, the order of classification is given by the need to eliminate dominating influences from the signature.

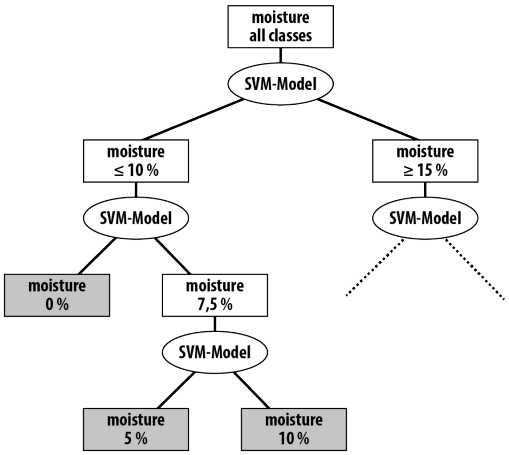

Figure 3.6: Illustration of the decomposition of an ordinal classification problem into a series of binary classifiers. Each independently learned classifier splits the data at the given threshold. Final classes are shaded in grey.

In a first step, moisture is classified by a SVM model according to figure 3.5. As can be seen in figure 3.4, the influence of water on the signature follows a continuous function which is discretized into classes in the experiment setup. Hence, the moisture classes have an ordinal regression structure. This ordinal regression can be decomposed into a set of binary classifiers with an ordinal classification (Frank and Hall, 2001). The advantages of this ordinal classification process is similar to the advantages of hierarchical classification. As the multi-class classification is simplified to a series of sub-tasks, the classifier is able to specialize the feature space and its parameters to the current problem, instead of finding the best trade-off for several classes. The data is always split in half at each step of the ordinal classification (see figure 3.6). Thus, the number of samples is approximately the same for both meta-classes at each split. The data has to be subsampled as otherwise, SVMs would favor the larger class during the risk minimization in equation 2.5. In addition, the computational time is minimized in this way.

For the ordinal classification of moisture a linear Support Vector Machine is learned, using polynomial features as described above with a ten-fold cross validation. The mean accuracy for moisture classification is 89% (see table 3.3). The model is very accurate for the air-dried samples. This is reasonable, as air-dried samples are clearly visually separable in the spectrum from all other classes. For the moist samples, the accuracy is decreased by the factor that samples dry faster at the surface and do not reliably have the same moisture as the mean of the whole sample. Additionally, the spectra are, of course, also influenced by the SOC and roughness. Although this influence is minor compared to the influence from moisture, this still drops the accuracy of the moisture prediction model. If the SOC content is quantified for each single moisture class, the root mean square error drops to 0.64 g/kg (table 3.3), which is a major improvement compared to the previous result without moisture classification.

Roughness is then predicted as outlined in figure 3.5 for all samples in each moisture class.

tb	Moisture, % w/w							
	Air-Dried	5	10	15	20	25	30	**All**
Accuracy	99%	90%	85%	84 %	90%	81%	92%	**89%**
SOC RMSECV	0.63	0.6	0.76	0.78	0.6	0.62	0.42	**0.64**

Table 3.3: Upper row: Accuracy of moisture prediction with an ordinal Support Vector Machine model. Lower row: Root mean square error of soil organic carbon quantification using a Support Vector Regression model for each predicted moisture class.

	Prediction, mm							Precision
True Class	grounded	<2	2-5	5-8	8-16	16-20	20-25	
grounded	168	1	0	0	0	0	0	99%
<2mm	0	167	0	0	0	1	0	99%
2-5mm	0	0	47	16	8	3	3	61%
5-8mm	0	0	108	137	105	15	3	37%
8-16mm	0	0	10	11	44	5	2	61%
16-20mm	0	0	2	0	6	135	12	87%
20-25mm	0	0	0	0	0	7	148	95%
Accuracy	100%	99%	28%	84%	64%	81%	88%	

Table 3.4: Accuracy of soil roughness classification with an ordinal SVM. The matrix shows the number of samples which were correctly predicted and the number of samples which were wrongly assigned to different classes.

The ordinal multi-class SVM is learned in the same way as described for moisture. The confusion matrix in table 3.4 shows the classification results for all roughness classes as well as the prediction of each wrongly classified sample. The confusion matrix reveals that the prediction accuracy varies strongly between the roughness classes. For sample sizes smaller than 2 mm the prediction accuracy is extremely good with 99%, respectively 100%. For 16-20 mm, respectively 20-25 mm, the classification accuracy is also sufficient with over 80% and 88%. For the sample sizes between 2 nm and 16 mm, however, the prediction accuracy drops very low if the precision is also taken into consideration. For these classes no accurate and reliable estimation of roughness was possible. However, the confusion matrix reveals that most wrong predictions have been between neighboring classes. This means that the model is able to correctly estimate between smooth, rough and moderate surface roughness. If the roughness classes are grouped into these three approximations, the prediction accuracy becomes very high (Table 3.5) with a mean of 96%. If these estimations are used in order to learn a new SOC regression model, the RMSECV drops to 0.5 g/kg, which is a significant improvement to the original RMSECV of 0.91 g/kg and the 0.61 g/kg for SOC quantification with moisture knowledge. This means that even the estimation of roughness is sufficient for a notable improvement in SOC quantification.

The presented method enables a quantification of roughness under varying moisture for the first time. The improvement in SOC prediction accuracy demonstrates the need to

	Roughness, aggregate size diameter, mm			
	<2	2-16	16-25	**All**
Accuracy	100%	98 %	90%	**96%**
SOC RMSECV	0.4	0.51	0.58	**0.5**

Table 3.5: Upper row: Accuracy of roughness prediction with an ordinal Support Vector Machine model. Lower row: Root mean square error of soil organic carbon quantification using a Support Vector Regression model for each predicted combination of moisture and roughness.

reduce the influence of these parameters. Although the experiment was concluded with laboratory measurements, it is a first step that demonstrates the possibilities of hierarchical classification in order to reduce the variance from moisture and roughness from the spectrum for SOC regression. The method reduced the SOC regression accuracy from 0.91 g/kg to 0.5 g/kg. For reference, destructive in-situ measurements for SOC typically achieve an accuracy of about 0.25 g/kg. For the prediction of SOC, the ISO Norm DIN ISO 10694 (1996) regulates that the SOC has to be quantified with an accuracy of at least 0.7 g/kg for the data discussed here. This means that the prediction accuracy of SOC with hyperspectral imaging sensors comes significantly closer to replacing in-situ measurements using this approach.

3.3 Unsupervised Archetypal Drought Detection

Under laboratory conditions with spectral point measurements (section 3.1 & 3.2), classes are well defined and training samples for supervised classification models exist. For early drought stress prediction with hyperspectral images before symptoms become visible, no labels exist. Therefore, an unsupervised approach is needed.

Cluster algorithms group the data unsupervised into a finite number of different clusters which are similar in feature space. As symptoms of drought naturally develops continuously, the clustering algorithm will discretize the process into ordinal clusters. Now stressed plants have a higher probability of having a high percentage of pixels belonging to clusters of stressed spectra. The percentage of pixels per image belonging to the clusters can now be used as a second order feature space for a supervised classification, as the labels for the whole plant are known.

However, discretization of a continuous process always comes with a loss of information and can lead to wrong conclusions if the discretization is wrongly chosen. A more desirable alternative may be to unmix the signals into their similarity to stressed and healthy spectra through an archetypal matrix factorization. While the archetypes themselves are still a discretization of the process, the spectrum of each pixel is continuously described by the archetypes (Figure 3.7).

However, finding the optimal archetypes runs quadratic with the number of samples. For hyperspectral images the algorithmic complexity is an important restriction. Simplex Volume Maximization (Section 2.3) finds a set of archetypes with nearly linear algorithmic complexity.

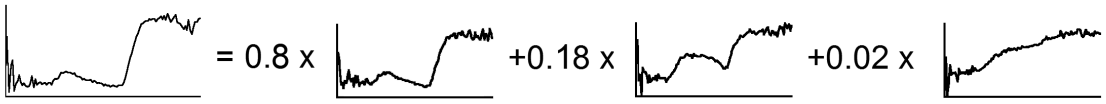

Figure 3.7: Any spectral signature within the data set can be expressed as a convex combination of archetypes. On the left is an arbitrary spectrum. The coefficients sum up to one and give a measure of similarity to the three archetypes chosen for this example: very healthy (left), leaves already lightly affected by drought (middle) and senescent leaves (right).

In Römer et al. (2012) Archetypal Analysis is used with Simplex Volume Maximization in order to predict stress unsupervised at a very early stage for the first time. The results are visualized to enable an observation of stress development.

Two experiments[3] are evaluated. For both experiments barley was used in a rainout-shelter. In 2010 the cultivar Scarlett was exposed to controlled water stress, where six pots where well watered and six pots were reduced watered. Images were taken twice every week starting on day 8 after drought stress application. In 2011 the genotypes Wiebke and Barke were used with the additional treatment of extreme drought stress, where no water at all was given to the plants. This results in three treatments (well watered, reduced watered, drought stressed) with 4 pots per treatment. Measurements were taken each day starting at day 1 after stress application. In both experiments plants where at growing stage BBCH31, i.e. also the group of control plants naturally show signs of senescence due to their development stage. The hyperspectral camera recorded images of 640x640 pixels with a spectral resolution of 4 nm in the range between 400 nm and 900 nm. In addition, control measurements were done for the soil moisture for the 2011 experiment.

The archetypes are calculated using the runtime efficient Simplex Volume Maximization (SiVM) algorithm (Section 2.3). For SiVM, the archetypes are restricted to lie on the convex hull. This is not very restricting for stress description, as this allows to unmix the other spectra into their similarity to the most stressed and the healthiest plant pigments.

The drawback is that outliers heavily influence the convex hull and can lead to an unwanted data decomposition. Therefore, to reduce the probability of outliers, the background is removed beforehand. A large percentage of outliers are eliminated by simple thresholds on certain wavebands (the green peak for example) and a previous clustering algorithm.

Afterwards, the set of archetypes is visually inspected by an expert to exclude biologically implausible signatures. Subsequently the remaining archetypes are manually annotated into spectra already showing first signs of drought stress and spectra expected for fully watered plants.

Figure 3.8 shows the archetypes calculated with SiVM on the 2011 data set which are manually classified as healthy and stressed. The mean probabilities of the similarities of all

[3]The experiment setup was done by Agim Ballvora from the INRES, University of Bonn. Spectral measurements were done by myself for the barley data. The measurements and experimental setup for the maize data in Römer et al. (2012) were done by Francisco Pinto from the FZ Jülich.

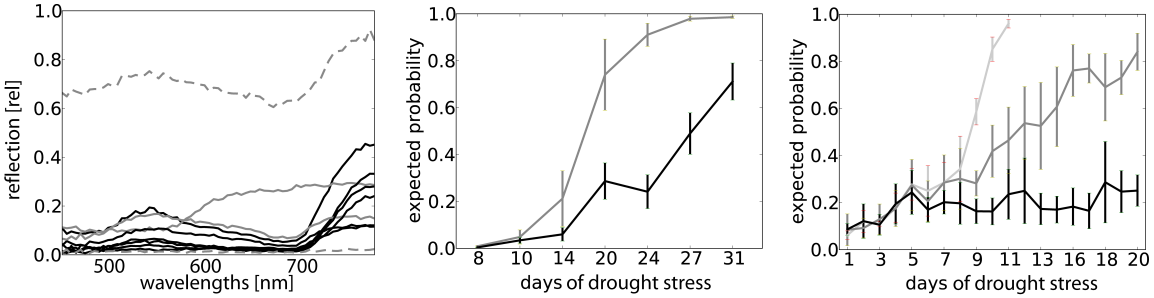

Figure 3.8: Archetypes of the controlled rainout shelter experiment with drought in barley. On the left: Archetypes as selected from the 2010 experimental run with grey signatures being labeled as spectra representing 'drought-stressed' pixels and black signatures being labeled as spectra for 'healthy' pixels. Dashed grey spectra are typical for the remaining background pixels. Middle, Right: Time course of expected probability of pixels being classified as 'drought-stressed' or 'healthy' according to the archetype. Black lines give the probability that a randomly chosen signature of a 'well-watered' plant can be explained by 'stressed' archetypes, grey lines give this probability for the 'reduced-watered' plants. Error bars indicate the standard deviation between the plants. In the middle: Results from the 2010 experiment: well-watered and reduced-watered plants could be separated significantly (α=0.05) at day 14 using a t-test. On the right: Results from the 2011 experiment: separation of dry plants and reduced-watered plants from well-watered plants is significant at day 9 (α=0.05).

pixels for each plant are then calculated. These probabilities are used as a second order feature space for each plant. Figure 3.8b and 3.8c show the expected probability of drought stress for the 2010 and 2011 data. The general increase of stressed pixels is due to the fact that plants are in their flowering stadium and increasing leaf senescence is characteristic for the development stage. Drought stressed plants could significantly (α=0.05) be separated from well-watered plants 14 days after start of drought stress application in 2010 and 9 days after stress application in 2011 using a t-test (Koch, 1997).

For the 2011 experiment the day of drought stress detection was only 1 day after a slight decrease in soil moisture could be measured and 5 days earlier than visual classification is possible. For comparison of the detection date with common evaluation methods, the results are compared with Vegetation Indices for drought stress detection in the visible area. The comparison is done based on the 2011 data set, as the frequency of measurements is much higher here and therefore delivers more exact results. The four most established drought stress indices are chosen, namely NDVI (Rouse et al., 1974), PRI (Peñuelas et al., 1995), REIP (Peñuelas and Filella, 1998) and CRI green (Gitelson et al., 2006). For each image, the Vegetation Indices are computed after background removal. All four indices are evaluated together using a one sided MANOVA test to see whether the treatments can be distinguished with a 5% significance level. Table 3.6 presents the results. Separation between fully watered and reduced-watered plants are possible from day 13 on. Plants without any

Day	Reduced-watered vs. fully watered (P-values)	No water vs. fully watered (P-values)
7	0.46	0.68
8	0.58	0.01
9	0.48	0.09
10	0.22	0.11
11	0.57	0.01
12	0.34	-
13	0.02	-
14	0.03	-

Table 3.6: The mean values of four vegetation indices are used for a four dimensional data matrix. Then it is tested if the four dimensional means differ significantly between plants with reduced-watered and fully watered plants, respectively, between fully watered plants and plants without any water supply.

water supply are clearly identified on day 8, although the hypothesis test fails for the days 9 and 10. On day 11 it is possible to separate both classes.

In order to visualize stress development on plants the weights calculated with SiVM can be used for a continuous plot. For discretization each pixel was clustered to the archetype with the highest individual coefficient (see figure 3.9).

In conclusion, archetypal matrix factorization has a high potential for unsupervised stress detection. However, the large hyperspectral datasets prevents an efficient use due to the quadratic algorithmic complexity. Simplex Volume Maximization allows the computation of archetypes in linear time and therefore enables the application of archetypal matrix decomposition to hyperspectral data. Here, for the first time, SiVM is used for unsupervised stress detection.

The construction of the second order feature space using annotated archetypes enables drought stress classification on the plant level. The annotation is possible due to the easy interpretation of the archetypes. The similarities of each pixel to an archetype are used in order to visualize drought stress development.

The results demonstrate that the presented method are four days faster for reduced watered plants and two days faster for plants without any water supply than a data analysis with Vegetation Indices. In addition, drought stress is successfully predicted only one day after it could be measured by the destructive reference measurements (Römer et al., 2012).

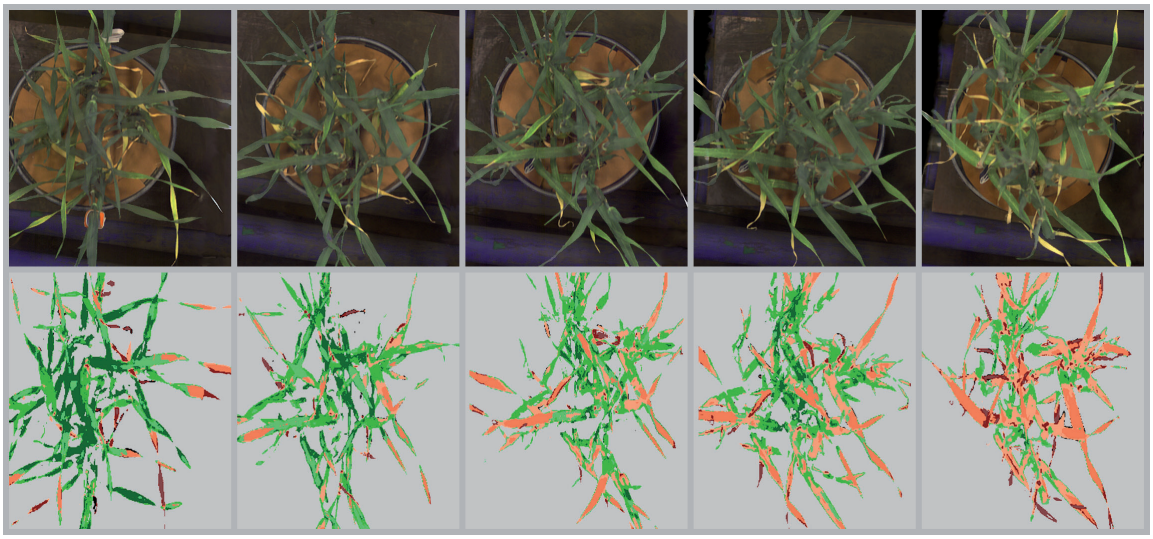

Figure 3.9: Stress dispersion in the time series of a reduced watered plant. On top are the original RGB images. On bottom the clusters derived from SiVM. Dark red are senescent pixels, orange are leaves affected by drought, light green are very healthy pixels and dark green pixels are lightly affected by stress.

4 Conclusions and Perspectives

The aim of this thesis was the development of new methods for the extraction of relevant patterns from hyperspectral signatures. The presented methods solve challenging problems, which are predominant in hyperspectral data analysis due to a low signal to noise ratio, the masking of important parameters in the spectrum and the absence of labeled data. The methods extracted an improved feature space and a hierarchical approach was presented in order to solve the three main problems identified in the introduction. The methods significantly improved the visualization and quantification of processes relevant for the description of early stress provesses in Plant Phenotyping and for soil parameter quantification.

The results of this thesis lead to a very early classification of stress, despite small differences between classes and strong noise. This was achieved with features describing the information of the whole spectrum by piece-wise polynomials. Important was not a global approximation of the whole spectrum with a single spline, but a partition of the spectrum into areas of high relevance for classification. With this approach, a small number of polynomial coefficients were sufficient to describe the shape of the most relevant areas. The resulting low dimensional feature space was used to predict wheat leaf rust at the second dai with a high classification accuracy, which has not been possible with Vegetation Indices before. The approach was also significantly better than a low rank approximation with PCA or a selection of the most important features with the filter algorithm Relief.

The thesis showed that the extraction of a second order feature space with Archetypal Analysis proved to be a very efficient method to predict early drought stress on unlabeled hyperspectral images. In contrary to artificial means, these archetypes were easily interpretable. This enabled a subsequent annotation of the archetypes by experts. A second order feature space was constructed with the mean similarity of each feature to the labeled archetypes. As the plants are labeled due to the experimental setup, classification on the plant level was now possible. This way, drought stress was detected significantly earlier than with Vegetation Indices and only one day after reduction of soil moisture was noticed with reference measurements. Once archetypes were annotated, it was also possible to visualize drought stress development in leaves.

The quantification of SOC despite the strong influence of moisture and roughness was improved with a hierarchical classification approach. As moisture had a significantly larger impact on the spectrum than roughness, a hierarchical classification tree was able to quantify moisture first, and then roughness for each moisture class. The classification accuracy was improved with ordinal classification contrary to a normal multi-class approach. This enabled roughness classification from spectral signatures for the first time. The hierarchical classification approach effectively removed the variance resulting from moisture and roughness. A SOC regression model was subsequently learned for each combination of moisture and roughness. This approach significantly reduced the SOC regression error.

The methods presented in this thesis resulted in a significantly improved analysis of hyperspectral images for Plant Phenotyping. However, in order to achieve further progress, further approaches may be considered, which have not been the focus of this thesis.

One way to get additional information is to include the information of neighboring pixels. For remote sensing applications, several papers have been published with the aim to smooth the classified image. This can be achieved by Markov Random Fields (Bishop, 2006), second order feature spaces as in stacked learning approaches (Cohen and Carvalho, 2005) or composite kernels for SVMs (Camps-Valls et al., 2006). However, for early stress detection, the problem becomes more challenging as the area where the stress occurs is naturally quite small at the beginning and therefore the challenge is to balance the trade-off between using the information from neighboring pixels without smoothing the early stress signals too strongly.

While this thesis showed the possibility to quantify early stress without the need for training data, supervised algorithms are generally still more precise and efficient for classification. For fungi quantification, for example, labeled samples are quite easy to obtain as the stressed pixels are easy to identify. However, the manual process of acquiring labeled pixels is still time consuming, especially as this process has to be repeated for every measurement campaign. Active Learning offers a fast and efficient alternative to solve this problem. Active Learning is an iterative approach where manual input is given regarding the most unsure points. With the improved training data an enhanced classification model is learned until the user is satisfied with the quality of classification. This way, only a fraction of the labeled samples are needed compared to a normal learning approach. For early stress detection the question remains whether the interactive communication between the learning algorithm and the phenotyping expert is able to learn and detect stress at stages where no symptoms are visible yet.

This thesis demonstrated the improvement achievable with new machine learning methods for plant phenotyping with hyperspectral sensors. However, these methods have to be available through free open source software tools. While there are free machine learning toolboxes available, these are not tailored for the exploration of hyperspectral images, where especially the visualization of the results, but also of the intermediate steps in classification, are of importance. Therefore, in order to support the use of advanced methods of data analysis in the Plant Phenotyping community and trigger new developments, new methods need to be available through free and optimally open source software products tailored for Plant Phenotyping with hyperspectral sensors.

Bibliography

Aldakheel, Y. Y., Danson, F. M., 1997. Spectral reflectance of dehydrating leaves: measurements and modelling. International Journal of Remote Sensing 18 (17), 3683–3690.

Bannari, A., Morin, D., Bonn, F., Huete, A. R., 1995. A review of vegetation indices. Remote Sensing Reviews 13 (1-2), 95–120.

Banninger, C., 1989. Changes in canopy leaf area index and biochemical constituents of a spruce forest ss measured by the ais-2 airborne imaging spectrometer. In: Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International. Vol. 4. pp. 2085 – 2089.

Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., Plümer, L., 2015. A review of advanced machine learnig methods for the detection of biotic stress in precision crop protection. Precision Agriculture 16 (3), 239 – 260.

Behmann, J., Steinrücken, J., Plümer, L., 2014. Detection of early plant stress responses in hyperspectral images. ISPRS Journal of Photogrammetry and Remote Sensing 93, 98–111.

Bellman, R. E., 1961. Adaptive control processes - A guided tour. Princeton University Press, Princeton, New Jersey, USA.

Bishop, C. M., 2006. Pattern recognition and machine learning. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Bishop, C. M., Laserre, J., 2007. Generative or discriminative? Getting the best of both worlds. Bayesian Statistics 8, 3 – 24.

Blum, A. L., Langley, P., 1997. Selection of relevant features and examples in machine learning. Artificial Intelligence 97 (1-2), 245 – 271.

Borengasser, M., S., H. W., Watkins, R., 2008. Hyperspectral remote sensing: principles and applications. CRC Press, Boca Raton, Florida, USA.

Breiman, L., 2001. Statistical modeling: the two cultures. Statistical Science 16 (3), 199 – 231.

Camps-Valls, G., Bruzzone, L. (Eds.), 2009. Kernel Methods for Remote Sensing Data Analysis, 1st Edition. John Wiley & Sons, Ltd, UK.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Frances, J., Calpe-Maravilla, J., 2006. Composite kernels for hyperspectral image classification. Geoscience and Remote Sensing Letters, IEEE 3 (1), 93 – 97.

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A., Brun, J. J., 2009. Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (nirs). European Journal of Soil Science 60 (5), 770 – 784.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27, software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Cherkassky, V., Lari-Najafi, H., 1992. Data representation for diagnostic neural networks. IEEE Expert: Intelligent Systems and Their Applications 7 (5), 43 – 53.

Cohen, W. W., Carvalho, V. R., 2005. Stacked sequential learning. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence. IJCAI'05. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 671 – 676.

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

Cutler, A., Breiman, L., 1994. Archetypal analysis. Technometrics 36, 338 – 347.

DeBoor, C. R., Dec. 1978. A practical guide to splines. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Dekel, O., Keshet, J., Singer, Y., 2004. Large margin hierarchical classification. In: Proceedings of the 21th International Conference on Machine Learning. ICML '04. ACM, New York, NY, USA, p. 27ff.

DIN ISO 10694, August 1996. Elemente und deren Verbindungen - Bodenbeschaffenheit. Bestimmung von organischem Kohlenstoff und Gesamtkohlenstoff nach trockener Verbrennung (Elementaranalyse) (ISO 10694:1996).

Ding, C. H. Q., Li, T., Jordan, M. I., 2010. Convex and semi-nonnegative matrix factorizations. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (1), 45 – 55.

Dumais, S., Chen, H., 2000. Hierarchical classification of web content. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '00. ACM, pp. 256 – 263.

Fiorani, F., Rascher, U., Jahnke, S., Schurr, U., 2012. Imaging plants dynamics in heterogenic environments. Current Opinion in Biotechnology 23, 227 – 235.

Frank, E., Hall, M., 2001. A simple approach to ordinal classification. In: Proceedings of the 12th European Conference on Machine Learning. EMCL '01. Springer-Verlag, London, UK, pp. 145 – 156.

Furbank, R. T., Tester, M., 2011. Phenomics - technologies to relieve the phenotyping bottleneck. Trends in Plant Science 16 (12), 635 – 644.

Gaspar, T., Franck, T., Bisbis, B., Kevers, C., Jouve, L., Hausman, J.-F., Dommes, J., 2002. Concepts in plant stress physiology. Application to plant tissue cultures. Plant Growth Regulation 37 (3), 263 – 285.

Gholizadeh, A., Borůvka, L., Saberioon, M., Vašát, R., 2013. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: state-of-the-art and key issues. Applied Spectroscopy 67, 1349 – 1362.

Gitelson, A. A., Buschmann, C., Lichtenthaler, H. K., 1999. Chlorophyll fluorescence ratio f735/f700 as an accurate measure of the chlorophyll content in plants: experiments with autumn chestnut and maple leaves. Remote Sensing of Environment 69 (3), 296 – 302.

Gitelson, J. A., Keydan, G. P., Merzylak, M. N., 2006. Three-band model for non-invasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical Research Letters 33.

Govender, M., Chetty, K., Bulcock, H., 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. Water SA 33 (2), 145 – 152.

Guiboileau, A., Sormani, R., Meyer, C., Masclaux-Daubresse, C., 2010. Senescence and death of plant organs: Nutrient recycling and developmental regulation. Comptes Rendus Biologies 333 (4), 382 – 391.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157 – 1182.

Jackson, R. D., Huete, A. R., 1991. Interpreting vegetation indices. Preventive Veterinary Medicine 11, 185 – 200.

Jolliffe, I. T., 2002. Principal component analysis, 2nd Edition. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Jones, H. G., Archer, N., Rotenberg, E., Casa, R., 2003. Radiation measurement for plant ecophysiology. Jouornal of Experimental Botany 54 (384), 879 – 889.

Karimi, Y., Prasher, S. O., Patel, R. M., Kim, S. H., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. Computers and Electronics in Agriculture 51 (1–2), 99 – 109.

Knipling, E. B., 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sensing of Environment 1 (3), 155 – 159.

Koch, K.-R., 1997. Parameterschätzung und Hypothesentests in linearen Modellen, 3rd Edition. Dümmler, Bonn, Germany.

Kohavi, R., John, G. H., 1997. Wrappers for feature subset selection. Artificial Intelligence 97 (1-2), 273 – 324.

Kononenko, I., 1994. Estimating attributes: analysis and extensions of relief. In: Proceedings of the European conference on Machine Learning. ECML-94. Springer-Verlag New York, Inc., Secaucus, NJ, USA, pp. 171 – 182.

Kooistra, L., Wanders, J., Epema, G. F., Leuven, R. S. E. W., Wehrens, R., Buydens, L. M. C., 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. Analytica Chimica Acta 484 (2), 189 – 200.

Lee, D. D., Seung, H. S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401 (6755), 788 – 791.

Lichtenthaler, H. K., Buschmann, C., 1987. Chlorophyll fluorescence spectra of green bean leaves. Journal of Plant Physiology 129 (1-2), 137 – 147.

Lim, P. O., Nam, H. G., 2007. Aging and senescence of the leaf organ. Journal of Plant Biology 50 (3), 291–300.

Liu, Z.-Y., Wu, H.-F., Huang, J.-F., 2010. Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. Computers and Electronics in Agriculture 72 (2), 99 – 106.

Mahlein, A.-K., 2011. Detection, identification, and quantification of fungal diseases of sugar beet leaves using imaging and non-imaging hyperspectral techniques. Dissertation, University of Bonn.

Malenovsky, Z., Mishra, K. B., Zemek, F., Rascher, U., Nedbal, L., 2009. Scientific and technical challenges in remote sensing of plant canopy reflectance and fluorescence. Journal of Experimental Botany 60 (11), 2987 – 3004.

Nature Editorial, 2010. How to feed a hungry world. Nature 466 (7306), 531 – 532.

O' Rourke, S. M., Holden, N. M., 2011. Optical sensing and chemometric analysis of soil organic carbon - a cost effective alternative to conventional laboratory methods? Soil Use and Management 27 (2), 143 – 155.

Pätzold, S., Mertens, F., Bornemann, L., Koleczek, B., Franke, J., Feilhauer, H., Welp, G., 2008. Soil heterogeneity at the field scale: a challenge for precision crop protection. Precision Agriculture 9 (6), 367 – 390.

Peñuelas, J., Filella, I., Gamon, J. A., 1995. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. New Phytologist 131 (3), 291 – 296.

Peñuelas, J., Filella, L., 1998. Technical focus: visible and near-infrared reflectance techniques for diagnostic plant physiological status. Trends in Plant Science 3, 151 – 156.

Peñuelas, J., Pinol, J., Ogaya, R., Filella, I., 1997. Photochemical reflectance index and leaf photosynthetic radiation-use-efficiency assessment in mediterranean trees. International Journbal of Remote Sensing 18 (13), 2869 – 2875.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2 (11), 559 – 572.

Piramuthu, S., Ragavan, H., Shaw, H. J., 1998. Using feature construction to improve performance of neural networks. Management Science 44 (3), 416 – 430.

Radcliffe, N. J., Shurry, P. D., 1995. Fundamental limitation on search algorithms: evolutionary computing in perspective. In: Van Leeuwen, J. (Ed.), Computer Science Today: Recent Trends and Developments: Lecure Notes in Computer Science. Springer-Verlag, London, pp. 275 – 291.

Rascher, U., Damm, A., van der Linden, S., Okujeni, A., Pieruschka, R., Schickling, A., Hostert, P., 2010. Sensing of photosynthetic activity of crops. In: Oerke, E.-C. (Ed.), Precision crop protection - the challenge and use of heterogeneity. Springer Science and Business Media, pp. 878 – 899.

Römer, C., Bürling, K., Hunsche, M., Rumpf, T., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines. Computers and Electronics in Agriculture 79 (2), 180 – 188.

Römer, C., Rodionov, A., Behmann, J., Pätzold, S., Welp, G., Plümer, L., 2014. Quantifying moisture and roughness with support vector machines improves spectroscopic soil organic carbon prediction. Journal of Plant Nutrition and Soil Science 77 (6), 845 – 847.

Römer, C., Wahabzada, M., Kersting, K., Rascher, U., Plümer, L., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., Behmann, J., Leon, J., Thurau, C., Bauckhage, C., 2012. Early drought stress detection in cereals: simplex volume maximization for hyperspectral image analysis. Functional plant biology 39 (11), 878 – 890.

Roscher, R., Waske, B., Förstner, W., 2012. Incremental import vector machines for classifying hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 50 (9), 3463 – 3473.

Rouse, J. W., Jr, Deering, D. W., Schell, J. A., Harlan, J. C., 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Tech. rep., Texas A & M University, Remote Sensing Center.

Rumpf, T., Mahlein, A.-K., Dörschlag, D., Plümer, L., 2009. Identification of combined vegetation indices for the early detection of plant diseases. In: Nealy, C., Maltese, A. (Eds.), Proceedings of the SPIE Conference on Sensing for Agriculture, Ecosystems and Hydrology. Vol. 7472. pp. 1 – 10.

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture 74 (1), 91 – 99.

Schölkopf, B., Smola, A. J., 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning), 1st Edition. The MIT Press.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, New York, NY, USA.

Silla, Jr., C. N., Freitas, A. A., 2011. A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery 22 (1-2), 31 – 72.

Stevens, A., Wesemael, B. V., V, G., Touré, S., Tychon, B., 2006. Detection of carbon stock change in agricultural soil using spectroscopic techniques. Soil Science Society of America Journal 70 (3), 844 – 850.

Stone, E., Cutler, A., 1996. Introduction to archetypal analysis of spatio-temporal dynamics. Physica D: Nonlinear Phenomena 96 (1-4), 110 – 131.

Suárez, L., Zarco-Tejada, P., Sepulcre-Cantó, G., Pérez-Priego, O., Miller, J., noz, J. J.-M., Sobrino, J., 2008. Assessing canopy PRI for water stress detection with diurnal airborne imagery . Remote Sensing of Environment 112 (2), 560 – 575.

Taiz, L., Zeiger, E., 2002. Plant physiology. New York: Sinauer.

Thurau, C., Kersting, K., Bauckhage, C., 2009. Convex non-negative matrix factorization in the wild. In: Proceedings of the 2009 Ninth IEEE International Conference on Data Mining. ICDM '09. IEEE Computer Society, Washington, DC, USA, pp. 523 – 532.

Thurau, C., Kersting, K., Bauckhage, C., 2010. Yes we can: Simplex volume maximization for descriptive web-scale matrix factorization. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. CIKM '10. ACM, New York, NY, USA, pp. 1785 – 1788.

Ustin, S. L., Gamon, J. A., 2010. Remote sensing of plant functional types. New Phytologist 186 (4), 795 – 816.

Vapnik, V. N., 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA.

Vapnik, V. N., Sep. 1998. Statistical learning theory, 1st Edition. Wiley.

Vapnik, V. N., Chervonenkis, A. Y., 1971. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications 16 (2), 264 – 280.

Viscarra-Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., Skjemstad, J. O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131 (1–2), 59 – 75.

Waske, B., Braun, M., 2009. Mapping of hyperspectral aviris data using machine-learning algorithms. Canadian Journal of Remote Sensing 35 (1), 106 – 116.

West, G. B., Brown, J. H., Enquist, B. J., 1997. A general model for the origin of allometric scaling laws in biology. Science 276 (5309), 122 – 126.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M.-L., Niinemets, l., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., Villar, R., 2004. The worldwide leaf economics spectrum. Nature 428 (6985), 821 – 827.

Wu, C., Jacobson, A., Laba, M., Baceye, P., 2009. Accounting for surface roughness effects in the near-infrared reflectance sensing of soils. Geoderma 152, 171–180.

Wu, D., Feng, I., Zhang, C., He, Y., 2008. Early detection of botrytis cinerea on eggplant leaves based on visible and near-infrared spectroscopy. Transactions of the American Society of Agricultural and Biological Engineers 53 (3), 1133 – 1139.

5 List of Own Publications

5.1 List of publications appended to this thesis

The following list of publications is most relevant for this thesis and appended below.

Römer, C., Bürling, K., Hunsche, M., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with Support Vector Machines. Computers and Electronics in Agriculture 79(2), 180–188.

Römer, C., Wahabzada, M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., Behmann, J., Léon, J., Thurau, C., Bauckhage, C., Kersting, K., Rascher, U., Plümer, L., 2012. Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. Functional Plant Biology 39(11), 878 – 890.

Römer, C., Rodionov, A., Behmann, J., Pätzold, S., Welp, G., Plümer,L, 2014. Quantifying moisture and roughness with Support Vector Machines improves spectroscopic soil organic carbon prediction. Journal of Plant Nutrition and Soil Science, 77(6), 845 – 847.

5.2 List of relevant publications not appended to this thesis

Römer, C., Bürling, K., Rumpf, T., Hunsche, M., Noga, G., Plümer, L.„ 2010. Early identification of leaf rust on wheat leaves with robust fitting of hyperspectral signatures. In: Proceedings of 10th International Conference of Precision Agriculture. Denver, USA 2010.

Alsaaod, M., Römer, C., Kleinmanns, J., Hendriksen, K., Rose-Meierhöfer, S., Plümer, L., Büscher, W., 2013. Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. Applied Animal Behavior Science 142(3-4), 134–141.

Römer, C., Plümer, L., 2010. Identifying Architectural Style in 3D City Models with Support Vector Machines. PFG 5, 371–384.

Henn, A., Römer, C., Gröger, G., Plümer, L., 2012. Automatic classification of building types in 3D city models - Using SVMs for semantic enrichment of low resolution building data.Geoinformatica (16), 281-306.

Rumpf, T., Römer, C., Weis, M., Sökefeld, M., Gerhards, R., Plümer, L., 2012. Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Galium aparine*. Computers and Electronics in Agriculture, 80, 89–96.

Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., Plümer, L., 2014. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precision Agriculture - online first - doi://10.1007/s11119-014-9372-7.

Rumpf, T., Mahlein, A.-K., Römer, C., Plümer, L, 2010. Optimal wavelengths for an early identification of Cercospora beticola with Support Vector Machines based on hyperspectral reflection data. In: IEEE (Hg.): 2010 IEEE International Geoscience and Remote Sensing Symposium 2010.

Kersting, K., Wahabzada, M., Römer, C., Ballvora, A., Thurau, C., Rascher, U., Léon, J., Bauckhage, C., Plümer, L, 2012. Simplex Distributions for embedding data matrices over time. In: Proceesings of the 12th SIAM International Conference on Data Mining (SDM). Anaheim, CA, USA 2012, S.26-28.

A Appended Papers

A.1 Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with Support Vector Machines

Römer, C., Bürling, K., Hunsche, M., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with Support Vector Machines. Computers and Electronics in Agriculture 79(2), 180 – 188.

Abstract

Early recognition of pathogen infection is of great relevance in precision plant protection. Pre-symptomatic disease detection is of particular interest. By use of a laserfluoroscope, UV-light induced fluorescence data were collected from healthy and with leaf rust inoculated wheat leaves of the susceptible cultivar Ritmo 2-4 days after inoculation under controlled conditions. In order to evaluate pathogen impact on fluorescence spectra 215 wavelengths in the range of 370-800 nm were recorded. The medians of fluorescence signatures suggest that inoculated leaves may be separated from healthy ones, but high-frequency oscillations and individual reactions of leaves indicate that separability is difficult to achieve. The misbalance between the high number of measured wavelengths and the low number of training examples induces a high overfitting risk. For a pre-symptomatic pathogen identification a small number of robust features was desired which comprise most of the information relevant for the given classification task. Instead of choosing only the most relevant wavelengths, the coefficients of polynomials fitting the spectra were used for classification. They specify the global curve characteristics. Piecewise fitting by polynomials of fourth order led to high classification accuracy. Support Vector Machines were used for classification. Cross validation demonstrated that the achieved classification accuracy reached 93%. This result could be attained on the second day after inoculation, before any visible symptoms appeared. The described method is of general interest for pre-symptomatic pathogen detection based on fluorescence spectra.

A.2 Quantifying moisture and roughness with Support Vector Machines improves spectroscopic soil organic carbon prediction

Römer, C.,Rodionv, A., Behmann, J., Pätzold, S., Welp, G., Plümer, L., 2014, Quantifying moisture and roughness with Support Vector Machines improves spectroscopic soil organic carbon prediction, Journal of Plant Nutrition and Soil Science, 77(6), 845 – 847.

Abstract

The challenges of Vis-NIR spectroscopy are permanent soil surface variations of moisture and roughness. Both disturbance factors reduce the prediction accuracy of soil organic carbon (SOC) significantly. For improved SOC prediction, both disturbance effects have to be determined from Vis-NIR spectra, which is especially challenging for roughness. Thus, an approach for roughness quantification under varying moisture and its impact on SOC assessment using Support Vector Machines is presented here.

A.3 Early drought stress detection in cereals: Simplex Volume Maximisation for hyperspectral image analysis

Römer, C., Wahabzada, M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., Behmann, J., Léon, J., Thurau, C., Bauckhage, C., Kersting, K., Rascher, U., Plümer, L., 2012. Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. Functional Plant Biology 39(11), 878-890.

Abstract

Early water stress recognition is of great relevance in precision plant breeding and production. Hyperspectral imaging sensors can be a valuable tool for early stress detection with high spatio-temporal resolution. They gather large, high dimensional data cubes posing a significant challenge to data analysis. Classical supervised learning algorithms often fail in applied plant sciences due to their need of labeled data sets, which are difficult to obtain. Therefore, new approaches for unsupervised learning of relevant patterns are needed. We apply for the first time a recent matrix factorization technique, Simplex Volume Maximisation (SiVM), to hyperspectral data. It is an unsupervised classification approach, optimized for fast computation of massive data sets. It allows calculation of how similar each spectrum is to observed typical spectra. This provides the means to express how likely it is that one plant is suffering from stress. The method was tested for drought stress, applied to potted barley plants in a controlled rain-out shelter experiment and to agricultural corn plots subjected to a two factorial field setup altering water and nutrient availability. Both experiments were conducted on the canopy level. SiVM was significantly better than using a combination of established vegetation indices. In the corn plots, SiVM clearly separated the different treatments, even though the effects on leaf and canopy traits were subtle.