

Early drought stress detection in cereals: Simplex Volume Maximization for hyperspectral image analysis

Christoph Römer¹, Mirwaes Wahabzada², Agim Ballvora³, Francisco Pinto⁴, Micol Rossini⁵, Cinzia Panigada⁵, Jan Behmann¹, Jens Léon³, Christian Thureau², Christian Bauckhage², Kristian Kersting², Uwe Rascher⁴, Lutz Plümer¹

- [1] Institute of Geodesy and Geoinformation, Geoinformation, University of Bonn, Meckenheimer Allee 172, 53115 Bonn, Germany
- [2] Institute for Intelligent Analysis and Information Systems, Fraunhofer, Schloss Birlinghoven, 53754 Sankt Augustin, Germany
- [3] Institute of Crop Science and Resource Conservation, Plant Breeding and Biotechnology, University of Bonn, Katzenburgweg 5, 53115 Bonn, Germany
- [4] Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich, Leo-Brandt-Str., 52425 Jülich, Germany
- [5] Laboratorio di Telerilevamento delle Dinamiche Ambientali (LTDA), Dip. di Scienze dell'Ambiente e del Territorio (DISAT), Università degli Studi di Milano Bicocca (UNIMIB), Piazza della Scienza, 1, 20126 Milano, Italy.

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 datasets which are difficult to get. Therefore, new approaches for unsupervised learning of relevant patterns are needed. We apply for the first time a recent matrix factorization technique, Simplex Volume Maximization (SiVM), to hyperspectral data. It is an unsupervised classification approach, optimized for fast computation of massive datasets. It allows calculating how similar each spectrum is to observed typical spectra. This provides the means to express how likely one plant is suffering from stress. It 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

Keywords: canopy, imaging spectroscopy, matrix factorisation, non-invasive, pattern recognition, plant phenotyping, unsupervised learning, water stress.

1 Introduction

Water scarcity is a principle global problem that causes aridity and serious crop losses in agriculture. Climate changes and a growing human population in parallel call for attention to advance research on understanding of plant adaptation under drought to improve management practices and breeding strategies.

The mechanistic basis underlying drought tolerance is complex as it is contributed by a number of related traits that are controlled mostly by polygenic inheritance. Genetic and biochemical approaches are time consuming and still fail to fully predict the performance of new lines in the field. In recent years it is discussed that phenomic approaches measuring the structural and functional status of plants may overcome the limited predictability and some authors have attributed this lack of high throughput phenomic data as the ‘phenomic bottleneck’ (Richards et al. 2010).

In the past years, hyperspectral camera systems became affordable and more widely used in plant sciences. In hyperspectral imaging, radiative properties of plant leaves or canopies are measured. Those can be used to determine structural and physiological traits of vegetation (Malenovský et al. 2009, Rascher et al. 2010, Ustin & Gamon 2010). For instance, the spectral reflectance of a leaf is characterized by a low reflectivity in the visible part of the spectrum (400-700 nm) due to a strong absorption by photosynthetic pigments, while a high reflectivity in the near infrared (700-1100 nm) is produced by a high scattering of light by the leaf mesophyll tissues (Knipling 1970; Rascher et al. 2010). In addition, in the shortwave infrared part of the spectrum (1100-2500 nm) the reflectance intensity is affected by water content, cellulose, proteins, and lignins of plant tissues (Rascher et al. 2010). At the canopy level, spectral reflectance is a combination of leaf and soil reflectance, canopy structure and illumination geometry. Despite that several studies in laboratory have shown a relationship between the amount of water in the leaf and the spectral reflectance in the optical region (Hunt and Barrett 1989, Danson et al. 1992), at canopy level the determination of water content presents some difficulties due to huge reflectance variation among leaves with the same water status (Cohen 1991), structural changes associated with loss of turgor (Kimes et al. 1984) or small reflectance differences at different levels of water stress. The biggest issue in estimating water content at canopy level, however, is the decoupling of the contributions of water content and LAI; in fact LAI variability may cancel out water-related features and hence preclude the retrieval of water content from single vegetation indices developed for leaf water content estimation. Better results can be obtained normalizing water indices by greenness indices (Colombo et al. 2008).

Remote sensing has been used in precision agriculture as a tool to provide timely information on crop conditions during the growing season. In the optical domain, the most widely investigated approach is based on the use of vegetation indices (VIs) to detect crop conditions (e.g. water and nitrogen status). Most approaches based on hyperspectral data aim to quantify plant traits by calculating VIs that quantify specific changes in plant structure and composition (Jackson and Huete 1991, Fiorani et al. 2012). While VIs have been successfully used to detect advanced stages of crop stress affecting for example crop biomass (Thenkabail et al. 2000), leaf nitrogen and/or chlorophyll content (Tilling et al. 2007, Haboudane et al.

2008), or vegetation moisture content (Yilmaz et al. 2008), the use of VIs for early stress detection is challenging. Furthermore, different crop stresses can affect the spectral bands involved in VI computation in a similar way, limiting the possibility to use VIs for this purpose.

For drought stress detection, several VIs have been developed. Although water status can primarily be observed in shortwave infrared wavebands, drought also has an effect on the visible and near infrared range, for instance on chlorophyll content and internal structure changes in the leaves due to wilting. Among many VIs for drought stress detection, the NDVI (Penuelas et al. 1995), PRI (Gamon et al. 1992), REIP (Penuelas and Filella 1998), and CRI (Gitelson et al. 2006) are well established and proved to be able to detect plant drought stress.

Recently, sun-induced fluorescence has been evaluated as a novel remote sensing measure to detect changes in the functional status of photosynthesis because of environmental constraints (Meroni et al. 2008, 2009a, 2009b). A first test case could already show that diurnal changes with afternoon drought induced downregulation of photosynthesis can be detected by sun-induced fluorescence (Damm et al. 2010). However, tests of this method have been limited to few studies and consequently, the potential of this approach has not yet been fully evaluated.

VIs are normally processed using two or three single wavelengths which are correlated with specific physical traits such as chlorophyll or water content. However, the complex physiological effects of drought stress, cause changes in the reflectance in most spectral regions (Aldakheel and Danson 1997, Penuelas et al. 1997). Thus, vegetation indices discard a lot of information available in the observed spectra.

Taking the full spectrum into account, on the other hand, involves the identification of relevant patterns in huge datasets with a demanding signal to noise ratio, although Römer et al. (2011) showed that using the information of the whole reflectance curve through polynomial approximations and machine learning leads to increased classification accuracies at very early stages of biotic stress. Additionally, for a better understanding of plant stress reaction, observations with high spatial and temporal resolution promise new insights. They enable the calculation of two-, or even three-dimensional stress dispersion models, which - apart from the better insight into plant reactions to biotic and abiotic stress - may become valuable for guided sampling of invasive, molecular measurements at the right time and the right spot.

Both factors pose a challenge for computationally fast and efficient data mining methods. This, however, is a problem as the runtime of many data mining algorithms grows quadratically with the number of samples. As hyperspectral imaging data could easily grow into several Giga- or even Terabytes, linear runtime would be preferable. In addition, supervised classification need classes well defined from training samples normally taken from the data set. However, in early stages of drought stress, symptoms are not visible yet and therefore detection of pixels with this information is not straightforward. Hence, data analysis of hyperspectral images is mainly an unsupervised task, where we know that certain plants have a higher probability of having pixels with first signs of drought stress, but we do not know which pixels and therefore are not able to determine sufficient training samples. Also most supervised and unsupervised classification methods result in discrete classes. This does not reflect the nature of the continuous change between healthy and diseased plant tissues.

Archetypal analysis as introduced by Cutler and Breiman (1994) computes low-rank approximations where data vectors are represented as a convex combination of extreme data points, such as very healthy and very senescent leaves. This allows the description of how similar an observed sample is to those extreme values and allows for an intuitive data interpretation. In this way, it is also similar to the idea of endmember selection (Somers et al. 2011, Bateson et al. 2000).

Traditionally, archetypal decompositions are achieved using methods such as gradient descend, multiplicative updates, or alternating least squares procedures which minimize a given objective function. However, memory and runtime requirements of such procedures also scale at least quadratically with the number of data. A deterministic, more efficient alternative consists in Simplex Volume Maximization (SiVM) which determines latent factors among extreme data vectors in linear time (Thureau et al. 2010a). This is akin to the idea of end-member detection (Winter 1999; Keshava and Mustard 2002) or sub-matrix maximization (Civril and Magdon-Ismail 2009) but is based on principles of distance geometry (Thureau et al. 2010b) and allows for the use of greedy search strategies.

In this communication we test for the first time Simplex Volume Maximization for the applicability in plant sciences. The main aim of this paper is the application of SiVM to hyperspectral images for the non-invasive quantification of temporal development of drought stress and the separation of subtle physiological effects of water and nitrogen availability. We used two independently designed data sets of hyperspectral images. Firstly, barley under controlled drought conditions within a rainout shelter. Barley (*Hordeum vulgare* L.) was chosen because due to its genetic diversity it underlies a huge variation with respect to drought adaptation (Schulte et al. 2009). Secondly, corn in a field experiment in order to test if the chosen method is applicable to controlled and field conditions. In addition corn was subject to a two-factorial treatment (drought and nutrient) and the method was used completely unsupervised, without any further information apart from the raw data. Water stress and high temperature are regarded as severe constraints to corn (*Zea mays* L.) production even under conditions where the soil profile is fully recharged at the beginning of the growing season.

We choose water stress because of its complex effects on plant traits and tested if SiVM is suitable for early stress detection in hyperspectral images. The different experiment designs were intentionally chosen to demonstrate that the proposed method can handle completely different conditions. Hyperspectral images were recorded in the visible to near infrared spectrum. We focused on those wavebands as those cameras are relatively cheap compared to infrared sensors or even thermography. Even though we mainly looked at water stress and the visible spectrum, the evaluation method is transferable to other treatments, plants, and spectral ranges.

We aim for a proof-of-concept description if automatically computed archetypes allow for an easy, explorative data interpretation, a visualization of stress dispersion, and a early prediction of water stress symptoms in barley and corn grown in a greenhouse and in a field, respectively. The working hypothesis of the paper is that SiVM, as a robust clustering technique, can handle the boundary conditions of early stress detection in hyperspectral images.

2 Material and Methods

2.1. Plant species and experimental conditions

We selected two sets of hyperspectral imaging data for this study. The first data set was barley recorded under semi-natural conditions in a rain-out shelter of the University of Bonn, Germany; the second was corn investigated under field conditions in an experimental agricultural site in northern Italy.

Imaging spectroscopy and plant traits in the rainout shelter - Experimental set-up and data acquisition for barley under controlled drought conditions

For the controlled water stress in the rainout shelter three barley summer cultivars (Scarlett, Wiebke, and Barke) were chosen. The seeds were sown in 11.5 liter pots filled with 17.5 kg substrate (Terrasoil, Cordel & Sohn, Salm, Germany). In year 2010 the genotype Scarlett was used in two treatments (well-watered and with reduced water) with six pots per treatment. Soil water potential of reduced water plants remained the same for the first seven days as for well-watered plants and decreased for the following eight days until it reached 50% compared to well-watered plants. In year 2011 the genotypes Wiebke and Barke were used in pot experiments arranged in a randomized complete block design with three treatments (well-watered, reduced-watered and drought stressed) with four pots per treatment. The drought stress was induced either by reducing the total amount of water or by complete water withholding. In both cases the stress was started at developmental stage BBCH31 (Biologische Bundesanstalt, Bundessortenamt and Chemical Industry), corresponding to the end of tillering and beginning of main shoot formation. By irrigation reduction the soil water potential of substrate remained at the same level as in well-watered pots for the first seven days but decreased rapidly in the following ten days reaching 40 % compared to reduced-watered plants.

For hyperspectral measurements the plants were transferred in the laboratory and illumination was provided by six halogen lamps (400 W ECO, OSRAM, Munich, Germany) from a distance of 1.6 meter. Hyperspectral images were acquired using the SOC-700 (Surface Optics, Corp., San Diego, California), which records images of 640 pixels x 640 pixels with a spectral resolution of approximately 4 nm with 120 equally distributed bands in the range between 400 and 900 nm. The system operates in push broom mode and uses a rotating mirror to scan the target (for further details see: Rascher et al., 2007). The hyperspectral camera was mounted at the same level as the lamps in nadir position. In year 2010 images were taken twice per week starting from day eight of water-stress (a ten time series). In year 2011 images were taken every consecutive day starting at the second day of watering reduction (a 21 days time series). Measurements of drought stressed plants were stopped at day ten as it was possible to clearly distinguish them from well-watered and reduced-watered plants by naked eye. Figure 1 shows some sample spectra for three different stages of senescence.

Additionally, leaf water potential was determined at midday time at days 1, 5, 9, 14, 17, and 20 after water stress was imposed. Leaf water potential was measured on one plant stem per plant using the Scholander pressure chamber (Scholander et al., 1965) on four plants per treatment grown under the same conditions as the ones measured with the hyperspectral camera. Soil water potential was calculated as the percentage of measured water content compared to the water-saturation point of the substrate. The soil moisture was measured using an HOBO weather station (ONSET, Massachusetts, USA).

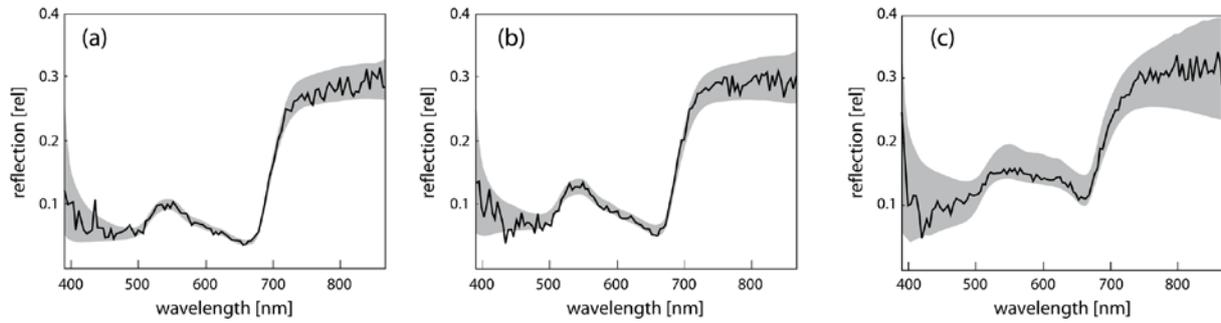


Figure 1: Sample spectra recorded in the 2011 barley experiment with the SOC 700. Shown are a typical green spectrum (a), a bright green spectrum (b), and a spectrum of a yellow, withered pixel (c) together with the standard deviation of the five hundred nearest neighbors (grey). It is clearly visible that measurement noise is very high until 470 nm and from around 800 nm.

Imaging spectroscopy and plant traits in the field - Experimental set-up and data acquisition

Field data of hyperspectral images were acquired on July 22nd 2010 from four meters above a corn canopy grown at the “Vittorio Tadini” experimental farm in Northern Italy (Gariga di Podenzano, PC - 44° 59' 00" N, 9° 41' 01" E). Corn plants were grown in combined treatments of two nitrogen fertilization levels (0 and 100 kg/ha N labeled as N0 and N1) and two irrigation regimes (rain-fed and full irrigation labeled as Irr 0 and Irr 1) during the growing season. Seedlings were planted on May 3rd 2010 in 15 m x 16.5 m plots and treatments were assigned randomly and repeated four times. An intensive field campaign was conducted on July 19th and 20th 2010, 42 days after seeding during stem elongation (12 leaves fully emerged), with the aim to measure leaf (i.e. chlorophyll fluorescence, PRI, SPAD, water content, and CO₂ assimilation) and canopy traits (plant height, density and leaf area index).

Chlorophyll fluorescence measurements have been conducted on the level of single leaves using the miniaturized Fluorescence Yield Analyzer (Mini-PAM) (H. Walz, Effeltrich, Germany). The maximal photochemical efficiency of PS II was measured during night on ten leaves per treatment as $F_v/F_m = (F_m - F_o)/F_m$ where F_m and F_o are the maximum and minimum dark-adapted fluorescence yield when a saturating light pulse (800 ms duration, intensity $\sim 4000 \mu\text{mol m}^{-2} \text{s}^{-1}$) is applied. Light-adapted measurements were conducted with a leaf clip holder (Bilger et al. 1995) on 20 leaves per treatment three times during the day: at 10:30, 13:30 and 16:30 local solar time, taking care not to change the ambient conditions. For analysis, the morning measurements were used as they are closest in time with the time of image recording. The effective quantum yield of PS II was determined as $\Delta F/F_m' = (F_m' - F_t)/F_m'$, where F_m' is the maximum light-adapted fluorescence yield when a saturating light pulse (as above) was superimposed on the environmental light (Schreiber and Bilger 1993) and F_t is fluorescence yield of the light-adapted sample. Apparent rate of photosynthetic electron transport of PS II (ETR) was obtained as $\text{ETR} = \Delta F/F_m' \times \text{PPFD} \times 0.5 \times 0.84$, where PPFD is incident photosynthetic photon flux density measured with a miniaturized calibrated quantum sensor, the factor 0.5 is the fraction of excitation energy distributed to PSII and 0.84 is the fraction of the incoming light absorbed by leaves. Finally, the Non-Photochemical Quenching (NPQ) was computed as $\text{NPQ} = (F_m - F_m')/F_m'$ (Bilger and Björkman 1990), where F_m is the average F_m value of the ten measurements acquired for each treatment during night-time.

Leaf PRI calculated as $(R_{531}-R_{570})/(R_{531}+R_{570})$ was measured with the PlantPen PRI device (Photon Systems Instruments, Brno, Czech Republic) on 15 leaves (three measurements per leaf) per treatment at 10:30 local solar time. Measurements relative to leaf chlorophyll content were done with a SPAD 502 (Minolta, Tokyo - Japan) leaf chlorophyll meter on ten leaves (three measurements per leaf).

Leaf CO₂ exchanges were measured on rain-fed plots (Irr 0) with an open infrared gas exchange system (CIRAS-1, PP-Systems, Stotfold, UK) equipped with a Parkinson leaf chamber, able to clamp single leaves. Measurements were performed at ambient CO₂ concentrations and illumination. Net assimilation rate ($\mu\text{mol m}^{-2}\text{s}^{-1}$), stomatal conductance ($\text{mol m}^{-2}\text{s}^{-1}$), transpiration rate ($\text{mmol m}^{-2}\text{s}^{-1}$) and intercellular CO₂ (ppm) were measured three times during the day on three plants per treatment. Each plant value represents the average of two leaves.

Finally, in the N1 plots four leaves per treatments were collected and enclosed in a sealed plastic bag at 5 °C. Leaf disks having an area of 3.80 cm² (A) were cut from each leaf using a cork borer. Fresh weight (FW) was recorded using an analytical balance, after which the leaf disks were immersed for 72 hours in distilled water, blotted and weighed (turgid weight, TW). Leaf disks were then dried at 80 °C in an oven, until constant weight (dry weight, DW) was reached. Finally, leaf EWT was calculated as: $(FW-DW)/A$ (gcm^{-2}) and RWC as: $(FW-DW)/(TW-DW)$ (%).

Furthermore, plant density, plant height and Leaf Area Index (LAI, $\text{m}^2 \text{m}^{-2}$) were measured in each plot. LAI was computed from measurements conducted with a SunScan Plant Canopy Analyzer (Delta-T Devices). A single LAI value for each plot was calculated by averaging ten samples collected on a diagonal transect across two crop rows at approximately the plot centre.

The imaging system, consisting in a Spectral Camera PS V10E (Spectral Imaging Ltd., Oulu, Finland) which operates in a linear push broom mode, was moved horizontally using a one meter long scanning bar (BiSlide™ 40", Velmex Inc. Bloomfield, NY, USA) and mounted in a boom lift to measure in nadir position at four meters from top-of-canopy level. At that height, the scanned area was of two meters cross 1.5 meters approximately, with an acquisition time of one minute per image and a spatial resolution of one mm/ pixel. Spectral measurements were performed on 22nd July 2010 between 10:00 and 14:00 local time under cloudless sky in order to avoid big variations of light conditions. Due to the time factor and also to the distances within the experiment site, just one plot per treatment could be measured with hyperspectral images, however, three images of different areas within each plot were acquired.

The Spectral Camera PS V10E is an imaging spectrograph for the visible/ near infrared part of the spectrum (400 – 1000 nm). It has a sensitive high speed interlaced CCD detector that provides spatio-spectral images of 1392 x 1040 pixels with a dynamic range of twelve bits. The spectral resolution is 2.8 nm (FWHM) and the spectral sampling ranges from 0.63 to 5.06 nm. The fore lens (23 mm/f2.4) images a column of data on to a horizontal slit of 30 μm width at the entrance of the spectrometer, obtaining the spatial information across a line only. Light is then spectrally spread in the y- axis by a diffraction grating and projected on the detector. The second spatial dimension is then built by the sequentially recording of lines images while the camera moves at constant speed across the scanning bar and therefore the resolution at this dimension is defined by the speed of the scan and the frequency of images acquisition. In our case, images had a size of 1392 x 840 pixels with 1040 spectral bands for

each pixel resulting in three-dimensional data tensors having 2.4 Gb each. The acquisition was managed by the SpectralDAQ Software (Specim, Oulo, Finland).

2.2. Data pre-processing of hyperspectral cubes

Hyperspectral data from both devices were linearly corrected by subtracting a dark image in order to remove the instrument noise. In the case of corn, where the measurements were done under changing natural light conditions, a new dark image was acquired each time the integration time was modified to avoid oversaturation of the sensor. For the barley measurements, a new dark image was acquired for each image taken.

To calibrate the images, the spectral reflectance was calculated for each pixel. For that, the spectrum of a pixel was divided by the spectrum of the incoming radiation estimated from a white reference panel that exhibits Lambertian reflectance located in each scene. In the case of measurements under artificial light, a 99% calibrated reflectance standard (Spectralon®, Labsphere Inc., NH, USA) was placed close to the plants, while under high sun light conditions (corn), a 20% calibrated reflectance panel (Zenith Alucore, Sphere Optics GmbH, Uhldingen, Germany) was used.

Pre-processing of the images from the SOC-700 was done using the HS-Analysis Software provided with the device. In the case of images from the spectral camera PS, the pre-processing was done using ENVI Software (ITT VIS, Boulder, CO, USA).

Although the SOC-700 measured from 394nm to 890nm, the wavelengths below 470nm and above 750nm were discarded because they appeared to be very noisy. The reason for this is most likely an unstable source of illumination for these frequencies. Therefore, only the bandwidths from 470nm to 750nm were used in the 2010 and 2011 barley experiment.

K-Means (Bishop 2006) was used for the last pre-processing step. As only the foreground pixels, i.e. the plant itself, is of interest, a K-Means clustering was done beforehand to remove background pixels. This is useful as it further reduces the computational complexity of the task and allows an easier interpretation of the spectra as most non biotic spectra were excluded from the data. The K-Means was calculated for 15 classes. The results were plotted in a false-color image, where each color represented a different class. Those clusters which only contained background pixels were chosen manually, whereas clusters which had foreground and background pixels were kept within the dataset. By this procedure roughly 95% of the background was eliminated. For the field experiment no previous background elimination was done.

2.3. Vegetation Indices for Drought Stress Detection

Among many available vegetation indices for drought stress detection, we chose four well established indices in this communication to compare SiVM against them. The NDVI (Rouse et al. 1974) is the normalized difference between the near infrared and visible red reflectance. It is responsive to changes in chlorophyll content and changes in the leaves internal structure due to wilting. It is calculated as $\rho_{800} - \rho_{670} / \rho_{800} + \rho_{670}$ where ρ notates the used waveband. The PRI, or Photochemical Reflectance Index (Penuelas et al. 1995), is derived from $\rho_{570} - \rho_{531} / \rho_{570} + \rho_{531}$. Drought causes the PRI-efficiency relationship to diverge from that of well irrigated plants, possibly owing to severe canopy wilting and to increased use of reductant by photorespiration (Gamon et al. 1992). REIP (Red Edge Inflection Point) index

(Penuelas and Filella 1998) is calculated with $700 + 20 \frac{(\rho_{670} + \rho_{780}) - 2\rho_{700}}{\rho_{740} - \rho_{700}}$ and like NDVI indirectly influenced by changes in plant water status. Finally, the CRI (Carotenoid Reflectance Index) (Gitelson et al. 2006) through $\left(\frac{1}{\rho_{510 \rightarrow 552}} - \frac{1}{\rho_{560 \rightarrow 570}} \right) \rho_{760 \rightarrow 780}$.

2.4 Unsupervised Classification: Simplex Volume Maximization

The idea of archetypal analysis is to find typical samples and express each other data point as a convex combination of these archetypes. This way, we can say how similar a sample is to a certain archetype and track the process of, for instance, senescence over time when samples become more and more similar to a signature of senescent samples (Figure 2).

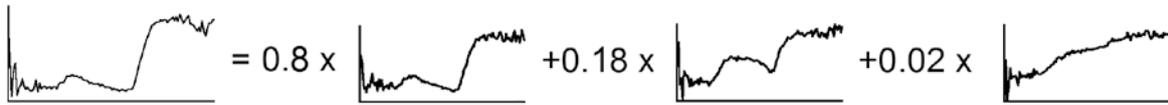


Figure 2: Any hyperspectral signature within the dataset can be expressed as a convex combination of archetypes. On the left is an arbitrary spectrum. The coefficients sum up to one and give a measure of similarity to the three archetypes chosen for this example: very healthy leaves (left), leaves already lightly changed by drought stress (middle), and senescent leaves (right).

If we have an input matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, where m is the number of features (i.e. wavelengths) and n the number of samples we have to select c columns in \mathbf{X} and construct a matrix of archetypal spectra $\mathbf{W} \in \mathbb{R}^{m \times c}$. Then we can construct the reconstruction matrix $\mathbf{H} \in \mathbb{R}^{c \times n}$ such that the Frobenius Norm $\|\mathbf{X} - \mathbf{H}\mathbf{W}\|$ becomes minimal with respect to the condition that all rows in the reconstruction matrix \mathbf{H} sum to one. Now the original data can be approximated as $\mathbf{X} \approx \mathbf{W}\mathbf{H}$. As all rows in \mathbf{H} sum to one, the coefficients in \mathbf{H} give a measure of similarity to the corresponding archetype in \mathbf{W} .

The problem is only how to select the archetypal samples in \mathbf{X} for \mathbf{W} . It was shown that a good subset of columns maximize their volume. However, this results in an NP hard problem (Civril and Magdon-Ismail 2009) and therefore this method is not applicable for large hyperspectral datasets. Thureau et al. (2010a) have developed a successful greedy algorithm, Simplex Volume Maximization (SiVM), which runs linear in time. The idea is to take the first two samples which are most likely furthest away from each other. The other points are selected in a way, such that the newly taken points maximize the volume of the previous selected points and the new sample (see Figure 3 for an example).

However, doing so ignores the distribution of the data. Therefore, in a next step, we subselected the most informative archetypes with respect to the data distribution, the ones involved in most reconstructions. This is measured by the entropy H of the Dirichlet spanned by these selected archetypes. In each iteration the archetype y with highest gain $H(A \cup \{y\}) - H(A)$ is selected, where A is the set of selected archetypes so far. The resulting signatures are the final set of archetypal spectra used for the following evaluation in section 3.1.

Once the archetypes are calculated, it is possible to express each other spectrum as a convex combination of those archetypes (Figure 2). The coefficients can be used as a measure of similarity to a certain archetype and, as they are a convex combination, interpreted as the probability that one signature belongs to a certain archetype. Thus, the results of SiVM quantify the contribution of each archetype for every pixel of a scene.

3 Results

3.1. Test case 1 – controlled rainout shelter: detection of temporal drought development in barley

Time course of drought development in barley plants was gradual. In 2011 experiment, soil moisture remained constant for the first seven days for reduced-watered plants, slightly decreased on day eight and then rapidly dropped for the following days until it reached 40% compared to the well watered soil.

Midday leaf water potential measured with the Scholander pressure chamber showed a strong increase for drought-stressed plants between day 5 and day nine after drought stress application, for reduced-watered plants water potential increased after day sixteen. Data of the biomass showed a similar time-course.

For the 2011 experiment plants without any water supply and fully watered plants were separable from day nine on by naked eyesight. Between plants with reduced water supply and fully watered plants from day 14 on (estimated by the team conducting the experiment).

For comparison, the vegetation indices for drought stress detection in the visible spectrum were calculated and applied to the 2011 dataset as it had the highest measurement frequency (daily). Thus, the NDVI, PRI, REIP, and CRI Green index were chosen. For each image, the mean vegetation indices were computed after complete background removal. To see whether reduced watered plants and plants with complete water withholding significantly differ from well watered plants at a significance level of 5% all four indices were evaluated together with a one-sided MANOVA test (conducted with the statistical toolbox of MatLab).

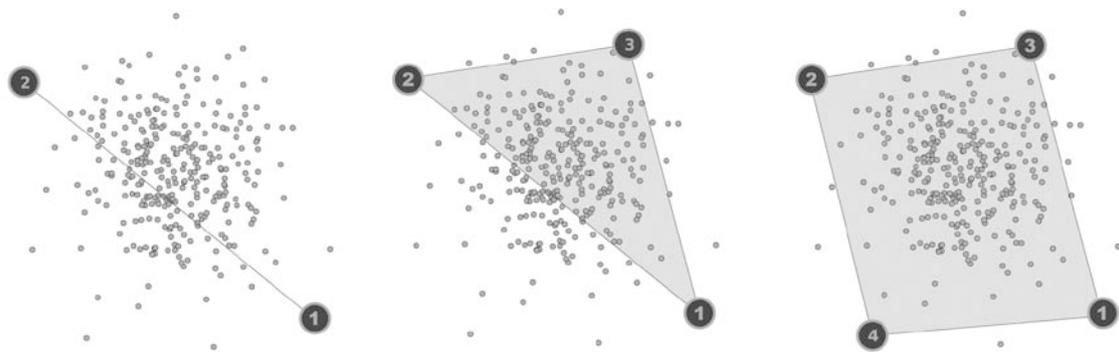


Figure 3: SiVM (Simplex Volume Maximization) for an instructive two-dimensional example. At first the two points which are furthest away from each other are chosen. Afterwards, points 3 and 4 were added as they maximize the volume spanned by all points.

The results are presented in Table 1. We can see that separation between reduced watered and fully watered plants is possible from day 13 on. This is on the same day or one day earlier than any single vegetation index could achieve with an ANOVA test (data not shown). Plants without any water supply are clearly identified on day eight, although the test failed for the days nine to ten. On day eleven it was possible to separate both classes again.

Day	Reduced / fully watered (p-values)	No water / 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 1: Results of a one-way MANOVA test. The mean values of four vegetation indices were used for a four dimensional data matrix. Then we tested if the four dimensional means differ significantly between plants with reduced water and fully watered plants, respectively between fully watered plants and plants without any water supply.

Now SiVM was applied to the whole dataset consisting of all spectra from all plants and time points to calculate the archetypes. Although subsampling would be possible for faster computation time, it is recommended to use the whole dataset for the lowest reconstruction error. The resulting archetypes from SiVM were visually inspected by an expert to exclude biologically implausible signatures. Afterwards ten bases were chosen by entropy, as described in section 3.1. The number of chosen bases depends on the number of clusters within the dataset. As this number is unknown, the number has to be estimated or the outcome of several choices has to be validated by an expert until a satisfactory result is achieved. This was possible as the results are existing signatures within the dataset which could be intuitively interpreted. Afterwards, the archetypes were categorized into spectra already showing first signs of drought stress and spectra normal for fully watered plants. Using the coefficients returned by the SiVM for each spectrum within an image, the probability that a randomly chosen signature is represented by archetypes categorized as stressed is calculated.

Figure 4 shows the calculated archetypes of barley that were manually classified as *healthy* and *stressed* pixels and the corresponding probabilities that a randomly chosen signature within the scene can be explained by *stressed* archetypes. In both experiments (Figure 4b and 4c) the probability that a pixel is more similar to *stressed* signatures increases during the experimental run for all treatments with the drought-stressed plants showing the fastest time kinetics. The general increase of *stressed* pixels is due to the fact that the plants were in their flowering stadium, and increasing leaf senescence is characteristic for the developmental stage (in 2011 plants were only measured until stressed plants could clearly be identified by naked eye, thus this effect is less visible). However, drought-stressed plants could significantly ($\alpha = 0.05$) be separated from well-watered plants 14 days after start of drought stress application for the 2010 experiment. In 2011 reduced-watered and plants without any water could both be distinguished from well-watered plants nine days after stress application (Figure 4b and c). It is noticeable that in the 2011 experiment (Figure 4c) the three

treatments showed a different time course of drought development that in principle was confirmed by visual inspection, however, SiVM detected the differences five days earlier than visual classification with the naked eye and four days faster than the tested vegetation indices.

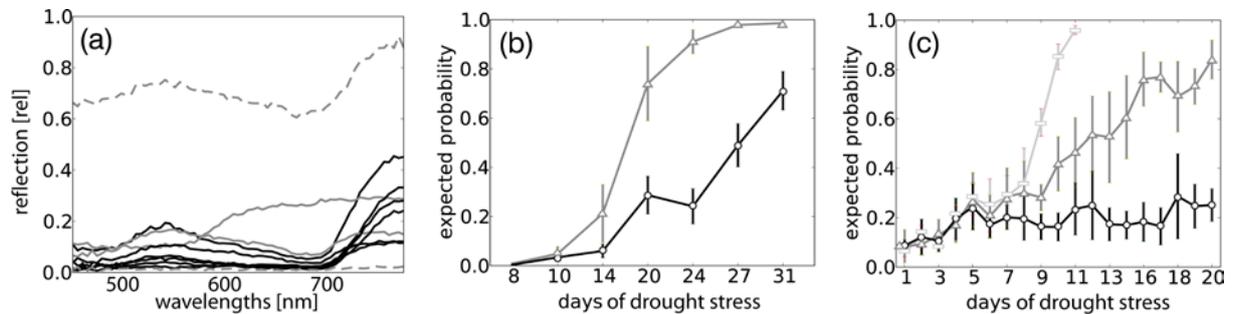


Figure 4 Archetypes and pixel classification of the controlled rainout shelter experiment with drought in barley. **(a)** Archetypes as selected from the 2010 experimental run with grey signatures being labeled as spectra representing “drought-stressed” pixels and black signatures being labeled as spectra for “healthy” pixels. Dashed grey spectra are typical for background pixels. **(b, c)** Time course of probability distribution of pixels being classified as “drought-stressed” or “healthy” according to the archetypes of (a). Circles and black lines give the probability that a randomly chosen signature of a “well-watered” plant can be explained by “stressed” archetypes, triangles and grey lines give this probability for the “reduced-watered” plants and rectangles and grey lines correspond to the “drought-stressed” plants. Error bars indicate the standard deviation between the pots. **(b)** shows the results for the 2010 experiment. Well-watered and reduced-watered plants could be separated significantly ($\alpha = 0.05$) at day 14 using a t-test. **(c)** shows the results from the 2011 experiment. Separation of dry plants and reduced-watered plants from well-watered plants was significant at day nine ($\alpha = 0.05$).

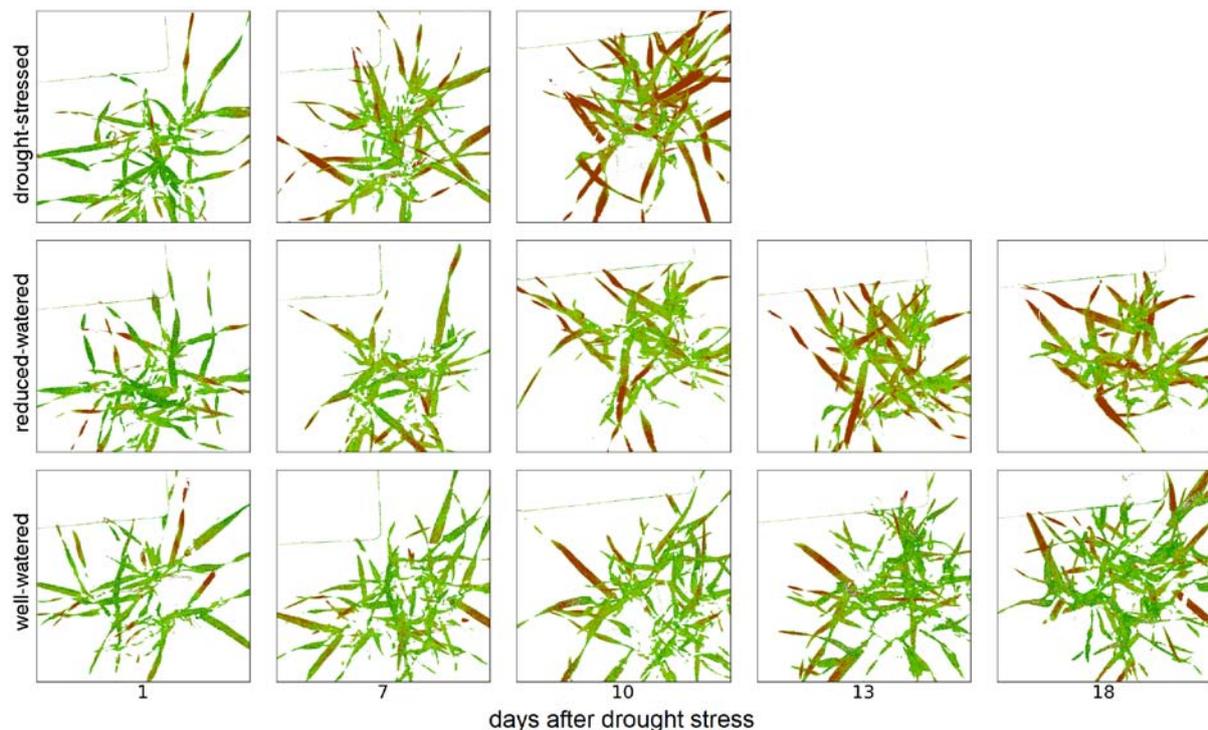


Figure 5: Spatio-temporal dynamics of drought stress dispersion in barley as classified by SiVM. Visualized are the coefficients of the SiVM. A dark green color indicated a high probability that the signature corresponds to a pixel belonging to the healthy archetypes, while a dark red color indicated a high probability of being associated with the stressed archetype. Images for drought-stressed plants stop at day ten as plants were only observed until drought stress was visible by naked eye.

Using the coefficient matrix returned by the SiVM calculation, it was then possible to visualize the spatio-temporal dynamics of drought effects using false color images. Figure 5 shows the results for the three treatments of the 2011 experiment, i.e. well-watered, reduced-watered and drought-stressed plants. Results of the SiVM indicate that even for severely stressed plants single leaves may be unaffected while other leaves were greatly affected by drought stress. Additionally, SiVM approach was able to detect the accelerated senescence of plant induced pigment degradation. The change in the well-watered plants is due to the fact that the plants were in their flowering season. The data suggests that an accelerated senescence can be observed.

3.2. Test case 2 – field experiment: distinguishing subtle differences of crop traits in the field by hyperspectral imaging and SiVM

The corn plants grown in the experimental field site having two different watering regimes and nitrogen availability showed manifold and greatly interlinked variations in leaf and canopy traits (see Table 2 for an overview). For instance, effective quantum efficiency ($\Delta F/F_m'$) was significantly increased by higher irrigation regimes while photosynthetic electron transport rate (ETR) was mainly stimulated by N treatment. On the canopy level, canopy height was significantly affected by irrigation, while leaf area index (LAI) was unaffected. These results show a non-trivial effect of nitrogen and water availability in plant functional and structural parameters on various levels, which is in accordance with other studies.

Parameter	N0		N1	
	Irr 0	Irr 1	Irr 0	Irr 1
F_v/F_m	0.83 ± 0.01^a	0.82 ± 0.01^a	0.83 ± 0.01^a	0.82 ± 0.01^a
$\Delta F/F_m'$	0.37 ± 0.07^b	0.43 ± 0.07^a	0.35 ± 0.05^b	0.40 ± 0.07^a
ETR ($\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$)	249 ± 35^a	253 ± 38^a	251 ± 34^a	253 ± 43^a
NPQ	1.92 ± 0.59^a	1.45 ± 0.49^b	2.09 ± 0.39^a	1.59 ± 0.56^b
Hmax (m)	1.72 ± 0.05^b	1.93 ± 0.07^a	1.71 ± 0.11^b	1.98 ± 0.12^a
Plant density ($n^\circ\text{plant}\cdot\text{m}^{-2}$)	6.80	7.91	7.05	7.44
SPAD (A.u)	47.8 ± 5.2^a	49.4 ± 3.0^a	45.9 ± 2.1^b	54.1 ± 3.9^a
PRI	0.0031 ± 0.0057^b	0.0088 ± 0.0063^a	0.0217 ± 0.0054^b	0.0263 ± 0.0067^a
Ai	35.8 ± 3.0^b	-----	41.0 ± 2.5^a	-----
Ci	90 ± 37^b	-----	126 ± 8^a	-----
Ti	3.48 ± 0.84^b	-----	4.74 ± 0.39^a	-----
Gi	261 ± 66^b	-----	363 ± 40^a	-----
EWT (g/cm^2)	-----	-----	0.012 ± 0.001^a	0.014 ± 0.001^a
RWC (%)	-----	-----	0.925 ± 0.024^a	0.959 ± 0.026^a
LAI	3.2	2.77	2.59	3.0

Table 2: Functional and structural leaf and canopy traits of the experimental corn plots that were treated with two irrigation regimes (Irr 0: rainfed and Irr 1: full irrigation) and two nitrogen levels (N0: 0 kg N / ha and N1: 100 kg N / ha). Performing two-way ANOVA, no significant interaction between both factors was detected in all traits, except SPAD. Significant differences ($\alpha=0.05$) between irrigation levels within each of the Nitrogen treatments were detected using one-way ANOVA (for each nitrogen level, different letters indicate significant differences). When the assumptions of normality and variance homogeneity of the data were not met, Kruskal-Wallis Analysis was performed. The following data are presented in the table. F_v/F_m = potential quantum yield of photosystem II; $\Delta F/F_m'$ = effective quantum yield of photosystem II; ETR= photosynthetic electron transport rate; NPQ = Non photochemical quenching; Hmax = maximum height of canopy measured on the field; SPAD=

relative values of leaf chlorophyll measured by the SPAD meter; *PRI*= Photochemical Reflectance Index at leaf level obtained with the PlantPen PRI device (PSI, Czech Republic); *A_i* = photosynthetic CO₂ assimilation rate; *C_i* = leaf internal CO₂ concentration; *T_i* = leaf transpiration rate; *G_i* = stomata conductance at water vapor (*A_i*, *C_i*, *T_i*, and *G_i* were measured with CIRAS (PP-System); *EWT*= water leaf equivalent thickness (obtained for leaf discs of 3,8 cm²); *RWC*=Relative water content (obtained for leaf discs of 3,8 cm²); *LAI*= Leaf Area Index (measurements obtained by SunScan were acquired along a transect crossing two consecutive crop rows at approximate the parcel center.

Using non-invasive imaging spectroscopy, the challenge is to identify special features in the spectral reflectance that are characteristic for drought stress and/or nitrogen status in the plant. In our test case, hyperspectral reflectance cubes taken some meters from above the canopy were of high quality and may show different canopy density (Figure 6a), but hyperspectral reflectance signature of the plants appears very similar and only subtle differences between the treatments can be depicted (Figure 6b and c). Water and nitrogen availability are known to affect leaves differently depending of their developmental stage or position in the canopy (Niinemets 2007 and references therein), we thus separated our analyses to sun- and shade exposed leaves (Figure 6b and c, respectively).

As a consequence vegetation indices that rely on a few wavebands only also failed to separate the treatments. We tested a full suit of vegetation indices, including the Normalized Difference Vegetation Index (NDVI, Figure 6d), the Photochemical Reflectance Index (PRI, Figure 6e) and the Water Band Index (WBI, Figure 6f). None of the vegetation indices was able to significantly separate irrigation and nitrogen treatment. The PRI however showed significant differences when measured on the leaf level (Table 1), but failed on the canopy scale (Figure 6e). However, a multivariate analysis with PRI and NDVI was able to identify all four treatments successfully.

25 archetypes were calculated with the SiVM algorithm. For further processing, the ten archetypes (Figure 7a) with the highest entropy (as described in section 3.1) were selected. Again, the number of selected archetypes has to be estimated. As environmental influences as well as the influence on the signatures from combined treatments is more difficult to interpret as for the barley experiment, data evaluation was completely unsupervised, i.e. there was no interpretation of the signatures by any expert. Hence, evaluation was slightly different.

In order to compute the distances between the matrices the reconstructions from the simplex sample space were transformed into Euclidean space using the alr-approach (Aitchison 1986). In the transformed space then any standard multivariate method can be used, such as an estimated multivariate Gaussian distribution. Afterwards, the distances between the distributions on a simplex were computed. Similarity was measured with the Bhattacharyya distance for multivariate Gaussian distributions. The resulting distance matrix was embedded into a two dimensional space with IsoMap (Tenenbaum et al. 2000). IsoMap first creates a graph y connecting each object to k of its neighbors, and then uses Euclidean distances of paths in the graph for embedding using multidimensional scaling (Cox and Cox 1984). The two irrigation regimes were well separated and there was also a clear distinction between fully irrigated plants with fertilization and without fertilization. Rainfed plots were separated with respect to fertilization by the algorithm, although the gap is rather small compared to the distinction between the other treatments (Figure 7b).

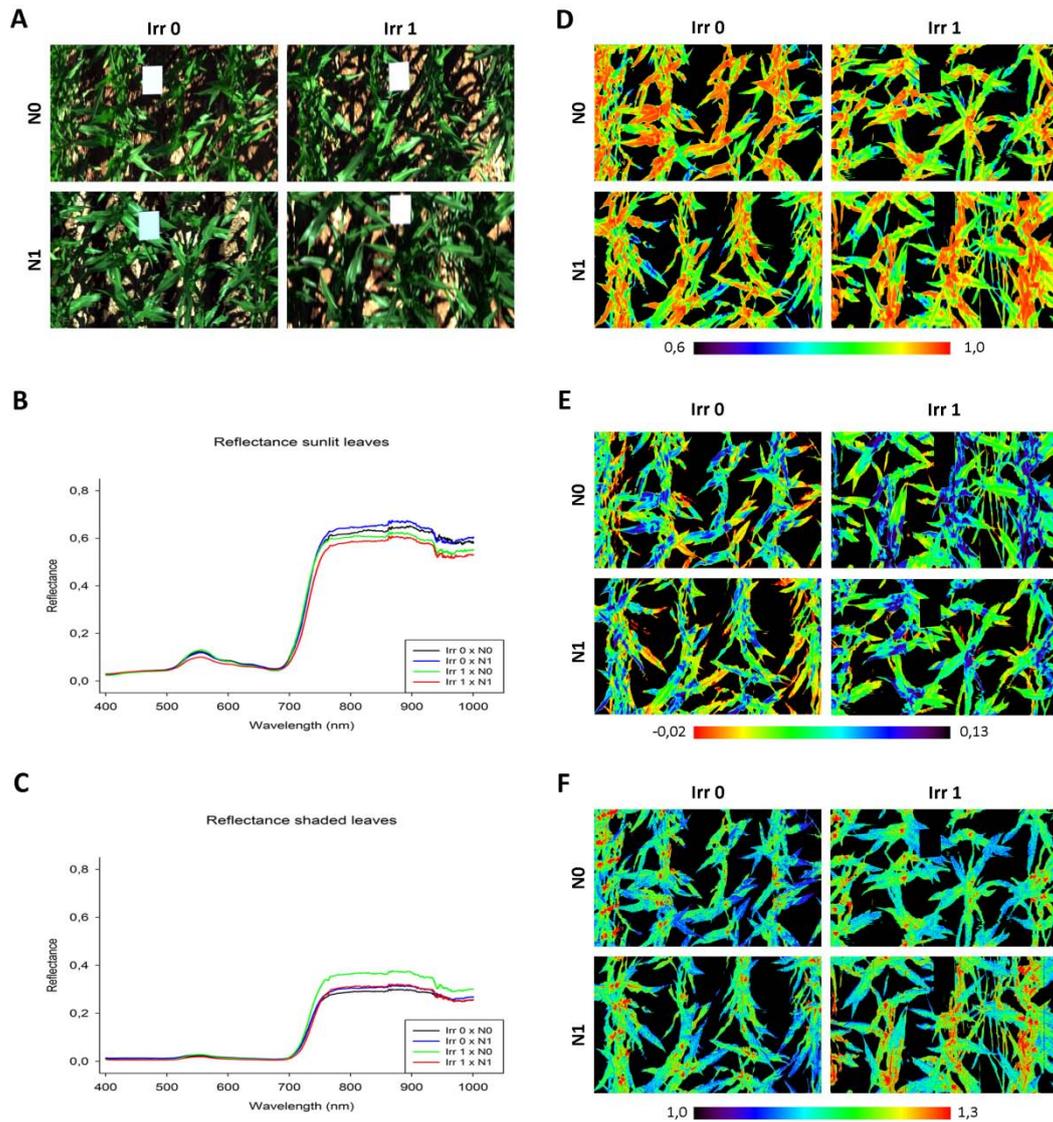


Figure 6 Imaging spectroscopy of a corn canopy that was treated with two irrigation regimes and nutrient availability. (A) RGB image of the canopy, (B) hyperspectral reflectance integrated over all sun-lit leaves and (C) integrated over all shaded leaves. Map of (D) the Normalized Difference Vegetation Index (NDVI), (E) the photochemical Reflectance Index (PRI), and (F) the Water Band Index (WBI). Reflectance and maps were calculated from hyperspectral data cubes that were acquired from 4 meters top-of-canopy using the Spectral Camera PS V10E (Spectral Imaging Ltd., Oulu, Finland). A 20% calibrated reflectance panel (Zenith Alucore, Sphere Optics GmbH, Uhlidingen, Germany) was placed at top-of-canopy level in each scene as reference for spectral reflectance estimation. Sun and shaded leaves surfaces were selected manually using ENVI software (ITT VIS, Boulder, CO, USA).

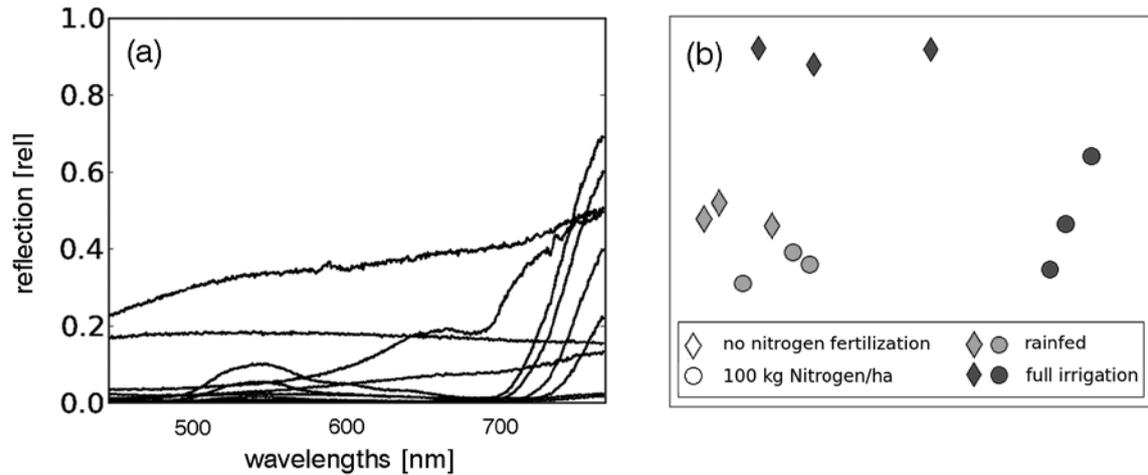


Figure 7 (a) ten most informative archetypes calculated by SiVM and selected by entropy for the field dataset. (b) the probability that any spectrum within an image can be explained by an archetype results in a ten dimensional feature space. For clustering and visualization, this is embedded into a two dimensional feature space with IsoMap. This method results in clear patterns for all treatments, allowing for a classification of each image.

4 Discussion

The working hypothesis of this manuscript is that the SiVM approaches are able to early detect stressed plants using hyperspectral imaging data. We expected that the results would enable an intuitive data description and visualization of stress dispersion within plants and canopies. For the evaluation of SiVM we used two contrasting data sets: the controlled drought experiments with barley in the rainout shelter that provided a clear time course of drought development and the corn field data that had a well-designed two factorial treatment but allowed only one measurement point. Despite the challenging boundary condition of large data volumes, unlabeled datasets and the fact that all barley plants had senescing leaves due to flowering, SiVM was able to track the temporal development of drought stress in barley four days before symptoms became visible by naked eye and shortly after reduction in soil moisture and leaf water potential (Figures. 4 and 5). For the field experiment SiVM proved to separate the four treatments into three clearly separable clusters (Figure 7). The fact that rain-fed plots were not separated as clearly as the other treatments by SiVM irrespectively of the nitrogen availability is noteworthy but may be reflected in leaf and canopy traits with water availability being the most limiting factor.

This is the first experimental test-case that uses SiVM on field data. We could show that the algorithm yields robust results that are supported by measured plant traits. Irrigation and nitrogen affect plant physiology on various levels. Water stress, for instance, affects photosynthesis on the level of light and dark reaction and alters functional traits such as stomatal density and mesophyll conductance (Chaves et al 2009). Thus, even under more demanding field conditions where illumination changes and variability between plots is inherent, SiVM gives robust results. This is especially vital, as the long term aim of this study is to develop methods and sensors able to be of use in applied plant science, breeding, and farming.

Instead of a labeled dataset, SiVM only needed manual input when the number of archetypes was selected and (in case of the barley dataset) which archetypes should be assigned to which senescence class. This is an enormous reduction of labor time for the user. In addition, an unsupervised approach may help to get new insights, as it lets the data speak for itself and therefore is more objective. In contrast to most other unsupervised approaches, the calculation of archetypes proved to be very helpful, as it was easier to interpret and discuss the results with experts, compared to, for instance, kMeans, where the resulting cluster means were artificial.

An alternative non-invasive, non-contact sensor for early stress detection is thermography. Beside their limitations (they are highly sensitive to environmental conditions and thermal response to plants largely lacks diagnostic potential (Mahlein et al. 2011)) they are undeniably fast and reliable stress detection sensors. The purpose of this paper was to demonstrate that widely available, cheaper, and well manageable hyperspectral sensors contain the necessary information for early drought stress detection. Both sensors are complementary and suitable for early stress detection. Sensor fusion may improve the detection ability and diagnostic potential. Further studies need to show that.

Comparison with vegetation indices was mainly done for the 2011 barley dataset. Here four established indices, NDVI, PRI, REIP, and CRI, were applied to the data after background removal with a one-sided multivariate ANOVA test on a 5% significance level. Results are presented in table 1. They could clearly separate reduced watered from well watered plants on day 13. Drought stress detection by SiVM was four days faster. Plants with complete water withholding were detected by VIs on day eight, which is one day faster than SiVM, although VIs fails for day nine and ten, whereas SiVM reliably detected drought stressed plants from day nine on.

For the Italy data PRI was able to detect difference in nutrient treatment, but failed at water detection, while NDVI was able to detect drought, but failed for nutrient. A combined evaluation with PRI and NDVI was able to detect all four clusters successfully. As only data from one day was available, comparison which method was faster was not possible. One advantage of SiVM is that the method worked completely unsupervised without any previous assumptions or knowledge about the data, while you have to know what you are looking for to choose the right index for your task. In addition, it is also possible to modify SiVM in a way to get the most important wavebands for clustering. This would allow identifying different groups and identifying possible treatments afterwards. Of course, also an unsupervised, quick screening for any stress followed by a detailed analysis with indices would be a possible scenario for a combination of SiVM with VIs.

To take full advantage of hyperspectral image sensors, it is indispensable to visualize the results (Figure 5), which allows better insights into the temporal and spatial dynamics of stress development in plants. They are vital for invasive methods, as they allow for guided sampling, and reduce the risk that leaves are taken as samples which are not affected by plant stress, as they are still fully supplied, despite clear stress developments in other leaves.

References

- Aitichison J (1982) The statistical analysis of compositional data. Chapman and Hall, London
- Aldakheel YY, Danson FM (1997) Spectral reflectance of dehydrating leaves: measurements and modelling. *International Journal of Remote Sensing* **18**(17), 3683-3690
- Bateson CA, Asner GP, Wessmann CA (2000) Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE T*
- Bilger W, Björkman O (1990) Role of the xanthophylls cycle in photoprotection elucidated by measurements of light-induced absorbance changes, fluorescence and photosynthesis in leaves of *Hedera canariensis*. *Photosynthesis Research* **25**, 173-185
- Bilger W, Schreiber U, Bock M (1995) Determination of the quantum efficiency of photosystem II and of non-photochemical quenching of the chlorophyll fluorescence in the field. *Oecologia* **102**, 425-432
- Bishop CM. (2006) Pattern recognition and machine learning. Springer, New York, USA
- Chaves MM, Flexas J, Pinheiro C (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany* **103**(4), 551-50
- Civril A, Magdon-Ismael M (2009) On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Computer Science* **410**(47-49), 4801-4811
- Cohen WB (1991) Temporal versus spatial variation in leaf reflectance under changing water stress conditions. *International Journal of Remote Sensing* **9**, 1865-1876.
- Colombo, R., Meroni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C., Panigada, C. (2008). Remote sensing of leaf and canopy water content in poplar plantations by means of hyperspectral data. *Remote Sensing of Environment*, 112, 1820-1834.
- Cox T, Cox M (1984) Multidimensional scaling. Chapman and Hall, London
- Cutler A, Breiman L (1994) Archetypal analysis. *Technometrics* **36**(4), 338-347
- Damm A, Elbers J, Erler E, Giolo B, Hamdi K, Hutjes R, Kosyancoya M, Meroni M, Miglietta F, Moreno J, Schickling A, Sonnenschein R, Udelhoven T, van der Linden S, Hostert P, Rascher U (2010) Remote sensing of sun induced fluorescence to improve modeling of diurnal courses of gross primary production (GPP). *Global Change Biology* **16**, 171-186.
- Danson FM, Steven MD, Malthus TJ, Clark JA (1992) High-spectral resolution data for determining leaf water content. *International Journal of Remote Sensing* **13**(3), 461-470.
- Fiorani F, Rascher U, Jahnke S, Schurr U (2012) Imaging plants dynamics in heterogenic environments. *Current Opinion in Biotechnology*, doi: <http://dx.doi.org/10.1016/j.copbio.2011.12.010>.
- Gamon JA, Penuelas J, Field CB (1992) A narrow waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment* **41**, 35-44.

- Gitelson AA, Keydan GP, Merzylak MN (2006) Three-band model for non-invasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research letters* **33**, L11402.
- Haboudane D, Tremblay N, Miller JR, Vigneault P (2008) Remote Estimation of Crop Chlorophyll Content Using Spectral Indices Derived From Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing* **46**, 423-437
- Hunt ER Jr., Rock BN (1989) Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment* **30**, 43-54.
- Jackson RD, Huete AR (1991) Interpreting vegetation indices. *Preventive Veterinary Medicine* **11**,185-200
- Keshava N, Mustard J (2002) Spectral unmixing. *IEEE Signal Processing Magazine* **19**(1), 44-57
- Kimes DS, Newcomb WW, Shutt JB, Pinter PJ Jr., Jackson RD (1984) Directional reflectance factor distributions of a cotton row crop. *International Journal of Remote Sensing* **5**, 263-277
- Knipling EB (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment* **1**(3), 155-159.
- Mahlein AK, Oerke E-C, Steiner U, Dehne H-W (2012) Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology* **133**, 197-209.
- Malenovský Z, Mishra KB, 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**, 2987-3004
- Meroni M, Rossini M, Picchi V, Panigada C, Cogliati S, Nali C, Colombo R (2008) Assessing Steady-state Fluorescence and PRI from Hyperspectral Proximal Sensing as Early Indicators of Plant Stress: The Case of Ozone Exposure. *Sensors* **8**, 1740-1754
- Meroni, M., Panigada, C., Rossini, M., Picchi, V., Cogliati, S., Colombo, R. (2009a). Using optical remote sensing techniques to track the development of ozone-induced stress. *Environmental Pollution*, 157, 5, 1413-1420. ISSN: 0269-7491
- Meroni M, Rossini M, Guanter L, Alonso L, Rascher U, Colombo R, Moreno J (2009b) Remote sensing of solar induced chlorophyll fluorescence: review of methods and applications. *Remote Sensing of Environment* **113**, 2037-2051
- Niinemets U (2007) Photosynthesis and resource distribution through plant canopies. *Plant Cell and Environment* **30**, 1052-1071
- Penuelas J, Filella I, Gamon JA (1995) Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytologist* **131**, 291-296.
- Penuelas J, Pinol J, Ogaya R, Filella I (1997) Photochemical reflectance index and leaf photosynthetic radiation-use-efficiency assessment in Mediterranean trees. *International Journal of Remote Sensing* **18**(13), 2869-2875

- Penuelas J, Filella L (1998) Technical focus: visible and near-infrared reflectance techniques for diagnostic plant physiological status. *Trends in Plant Science* **3**, 151-156.
- Rascher U, Nichol CJ, Small C, Hendricks L (2007) Monitoring spatio-temporal dynamics of photosynthesis with a portable hyperspectral imaging system. *Photogrammetric Engineering and Remote Sensing* **73**, 45-56
- Rascher U, Damm A, van der Linden S, Okujeni A, Pieruschka R, Schickling A, Hostert P (2010) Sensing of photosynthetic activity of crops. In: *Precision Crop Protection - the Challenge and Use of Heterogeneity*. Oerke E.-C. et al. (eds.), Springer Science+Business Media B.V.
- Richards RA, Rebetzke GJ, Watt M, Condon AG, Spielmeyer W, Dolferus R (2010) Breeding for improved water productivity in temperate cereals: phenotyping, quantitative trait loci, markers and the selection environment. *Functional Plant Biology* **37**(2), 85-97
- Römer C, Bürling K, Rumpf T, Hunsche M, Noga G, Plümer L (2011) Robust fitting of fluorescence spectra for presymptomatic wheat leaf rust detection with support vector machines. *Computers and Electronics in Agriculture* **79**(1), 180-188
- Rouse JW, Haas RH, Schell JA, Deering DW, Harlan JC (1974) Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFC, Type III, Final Report, Greenbelt, MD, 371pp.
- Schreiber U, Bilger W (1993) Progress in chlorophyll fluorescence research: major development during the past years in retrospect. *Progress in Botany* **53**, 151-173
- Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waugh R, Wise RP, Stein N (2009). The international barley sequencing consortium - at the threshold of efficient access to the barley genome. *Plant Physiology* **149**, 142-147
- Scholander P, Bradstreet E, Hemmingsen E, Hammel H (1965) Sap pressure in vascular plants: negative hydrostatic pressure can be measured in plants. *Science* **148**(3668), 339-346
- Somers B, Asner GP, Tits L, Coppin P (2011) Endmember variability in spectral mixture analysis: a review. *Remote Sensing of Environment* **115**, 1603-1616
- Tenenbaum JB, De Silva V, Langford JC (2000) A global geometrix framework for nonlinear dimensionality reduction. *Science* **5500**(390), 2319-2323
- Thenkabail PS, Smith RB, De Pauw E (2000) Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. *Remote Sensing of Environment* **71**, 158-182
- Thureau C, Kersting K, Bauckhage C (2010a) Yes we can – simplex volume maximization for descriptive web-scale matrix factorization. In: *Proceedings of the conference on information and knowledge management*, 1785-1788
- Thureau C, Kersting K, Bauckhage C, Quadrianto N (2010b) Grappling with gigantic matrices: fast approximations based on distance geometry. In: *Proceedings of the. NIPSWorkshop on Low-rank Methods for Large-scale Machine Learning*

- Tilling AK, O'Leary GJ, Ferwerda JG, Jones SD, Fitzgerald GJ, Rodriguez D, Belford R (2007) Remote sensing of nitrogen and water stress in wheat. *Field Crops Research*, **104**, 77-85
- Ustin S, Gamon JA (2010) Remote sensing of plant functional types. *New Phytologist* **186**, 795–816
- Winter ME (1999) N-FINDR: an algorithm for fast and autonomous spectral endmember determination in hyperspectral data. In: *Proceedings of the International Conference on Applied Geologic Remote Sensing*
- Yilmaz M T, Hunt ERJr, Jackson TJ (2008) Remote Sensing of Vegetation Water Content from Equivalent Water Thickness using Satellite Imagery, *Remote Sensing of Environment* **112**, 2514-2522

Acknowledgements

This work has been carried out in frame of a research program funded by BMBF with project number 315309/ CROP.SENSE and the DAAD fellowship to Francisco Pinto. Kristian Kersting and Mirwaes Wahabzada were supported by the Fraunhofer ATTRACT fellowship STREAM. The Italian campaign has been conducted in the framework of the Axia (CRUI-Nestlé funded project for research and development) and EUFAR (European Facility For Airborne Research) Transnational Access project. The authors gratefully acknowledge the staff of DISAT-UNIMIB and Università cattolica di Piacenza for the support during the Italian field campaigns. The authors would like to thank Merle Noschinski for excellent technical assistance, Henrik Schumann, Melanie Herker, and Sfehan Teutsch for their support with hyperspectrometry.