Photogrammetric Week '15 Dieter Fritsch (Ed.) Wichmann/VDE Verlag, Belin & Offenbach, 2015

Early Identification of Plant Stress in Hyperspectral Images

Jan Behmann, Anne-Katrin Mahlein, Lutz Plümer, Bonn

ABSTRACT

In recent years, remarkable results have been achieved in the early detection of weeds, plant diseases and insect pests in crops. These achievements are related both to the development of non-invasive, high resolution optical sensors and data analysis methods that are able to cope with the resolution, size and complexity of the signals from these sensors. Especially hyperspectral cameras are capable sensors for the early detection of stress even before visible sympotoms become apparent. Their interpretation with regard to data amount, noise factors and their unknown effects is challenging.
In this study, the the focus lies on the early detection of drought stress in barley plants based on hyperspectral images. For the specific task of representing and predicting the development of drought stress different model types are compared. It turns out that the linear ordinal classification combines both, high precision and low model complexity. Prediction results for an drought stress experiment over time are presented.

1. INTRODUCTION

''Sustainable intensification'' – producing more food from the same area of land while at the same time reducing the environmental impacts – demands innovative agronomic practices (FAO, 2011). Up to 40 % of the world food production is lost through stress induced by weeds, diseases and insects (Oerke & Dehne, 2004). Thus the early identification of primary weed and disease foci as well as locating areas differing in weed or disease severity in the field is of high relevance. Weed and pest monitoring in addition to the decision making-process are fundamental for site-specific management. Such site-specific crop management requires a high density of spatial and temporal information.
Sensor systems, such as multi- and hyperspectral point sensors and images can provide high resolution data concerning the health state of agricultural crop stands. Progress in these technologies along with advances in computer science and analysis methods offers new opportunities for precision agriculture, and constitutes the basis for early detection of stress symptoms. Current sensor systems are highly sensitive to biological changes. However, they do not measure plant physiological parameters directly. Instead, they record a spectrum that is a sum of reflectance contributions attributed to various plant characteristics and the measurement conditions (e.g. leaf inclination, illumination and surface texture) (Gitelson et al. 2002).
These observations that have high spectral, spatial and temporal resolution require the use of advanced methods of data analysis. Traditional methods for data analysis, such as linear regression and linear discriminant analysis, are based on predefined distributions and model assumptions. For data that complies with these demands, these methods are applicable without adverse effects. Therefore, the field of application is limited for these kinds of analysis methods.
Advanced methods of machine learning, such as k-means, support vector machines (SVMs) and support vector regression (SVR), require less prior information and are applicable to a wider range of tasks as they derive the underlying distributions and model assumptions implicitly from training data. This capacity to adapt to almost any kind of data makes them well suited for tasks with limited prior knowledge about a suitable interpretation model or complex data characteristics like non-

This paper is based on Behmann et al. (2014a) and revives aspects of Behmann et al. (2014b) and Behmann et al. (2015).

linearity, non-Gaussian noise and outliers. These methods are particularly suitable for the interpretation of data from optical sensors monitoring vegetation because the included noise factors may be compensated by a sufficient amount of representative training data.
Methods from the field of machine learning have been successfully applied in numerous disciplines. In recent years, the precision crop protection community has started to benefit from the capabilities of machine learning. Therefore flexible machine learning methods can help to gain new insights into the plant – and find new relationships within data for the detection stress effects. This exposition of the problem statement and this discussion is based on Behmann et al. (2015).

2. INTERPRETATION OF HYPERSPECTRAL IMAGES

The interpretation of hyperspectral images is accompanied by a number of aggravating factors. Some of them are related to the sensor technology and the resulting bad signal-to-noise-ratio compared to multispectral images or RGB images. Others are related to the investigated phenomena and the possibility to characterize, quantize and parameterize changes in the observations. In this study we observe hyperspectral images of drought stressed barley plants. Specific problems related to the description of the drought state of a plant will be discussed after the experimental design and a short introduction of drought stress characteristics.

2.1. Experimental design

A series of hyperspectral images was measured in drought stress experiments on single barley plants. Twelve barley plants were cultivated under controlled conditions. The plants were divided into three groups (each of four plants) with differing irrigation intensities (well-watered, reduced watered, and unwatered). The hyperspectral images were recorded in the laboratory under controlled illumination provided by six 400 W halogen lamps. As sensor, a hyperspectral imager SOC-700 from Surface Optics was used. The imager has a spatial resolution of 640 x 640 pixels and a spectral resolution of approximately 4 nm with 120 equally distributed bands in the range of 430 to 890 nm (visible and near-infrared spectrum). The lamps and the sensor were arranged above the imaging position of a single pot. All images were radiometrically normalized by subtracting the dark frame and calculating the absolute reflectance using the ratio to a white reference panel. The measurements started one day after water reduction and were continued daily for a period of twenty days. The measurement of the unwatered plants was stopped after eleven days, when differences between the unwatered and the reduced watered plants could be clearly perceived by the naked eye.

2.2. The chacteristics of drought stress

In this paper, we focus on the detection of early drought stress, which is not yet visible to the naked eye. Drought stress, induced by water shortage, is one of the biggest challenges in global crop production. It occurs, if a plant's transpiration rate exceeds its water uptake and is closely linked to the basic processes of absorption of light energy and the production of biomass through photosynthesis (Taiz, 2010). This process requires that the plant assimilates carbon from the atmosphere. The assimilation is regulated by the aperture of stomata, which are microscopic pores in the epidermis. Opening of stomata, however, is responsible for significant loss of water via transpiration (Taiz, 2010). Hence, transpiration and photosynthesis are inseparably linked and a plant's need for carbon uptake can only be satisfied at risk of water loss. Thus, increasing yield under drought conditions is a complex optimization problem for breeders and management, which is termed “more crop per drop”.

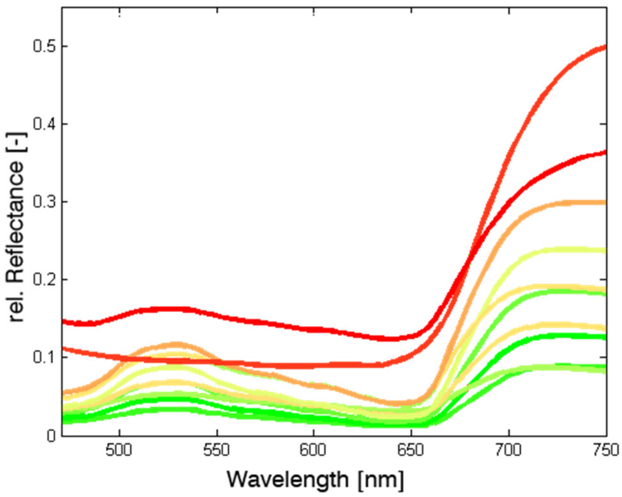

Figure 1: The ordinal order of the discrete centroids of the drought stress process (from green to red). The centroids result from a K-Means and the following order by the criterion defined in section 3.1.

2.3. Detection of drought stress

In contrast to most plant diseases, drought stress does not manifest itself in local symptoms. The reallocation of resources involves the entire plant – and occurs in all plants to a specific degree, even in the well-watered ones. Drought stressed plants are characterized by early and accelerated leaf senescence (Munné-Bosch et al., 2004). The detection of this creeping process and distinguishing it from normal variations requires spectral information with high degrees of temporal and spatial resolution.

The leaf senescence results in adaptions in the reflectance characteristics. Merzlyak et al. (1999) showed that the observed spectra of a proceeding senescence form an ordinal order (Fig. 1). This is mainly related with the degradation of chlorophylls and the reallocation of nutrients and resources. Under drought, the plant dismounts specific plant pigments and uses the released resources to support leaves with higher potential.

We hypothesize that series of hyperspectral images provide information which allows a description and detection of drought stress processes before changes occur which are visible to the naked eye. Hyperspectral imagers permit the capturing of spatio-temporal processes within entire plants and facilitate the observation of the spatial distribution of senescence – even within single leaves (Fig. 2). In contrast to biochemical analysis, their non-invasive characteristics enable the monitoring of stress related processes without affecting them. As the senescence process occurs in all plants, the differentiation between the treatments is based on different distributions, i.e. histograms, of plants' vitality. These histograms are derived from the frequencies of the senescence classes at pixel scale.

3. MODELLING THE DROUGHT STRESS DEVELOPMENT IN HYPERSPECTRAL IMAGES

3.1. Unsupervised labeling

One major problem in the automatic interpretation of hyperspectral images is that the drought stress processes which are to be analyzed cannot be visualized for a validation by a human being. Thus, supervised algorithms, which need labeled data, are not applicable directly. The limited possibilities for visualizing high dimensional data require an automatic mechanism without human contribution.

Our approach of unsupervised labeling utilizes a randomly initialized k-Means to group plant pixels with similar spectral properties; if time series of images are used as input, all images are processed at once. As pointed out by Merzlyak et al. (1999), the senescence process forms an ordinal order mainly related to chlorophyll degradations. Accordingly, the resulting centroids of k-Means represent this order and can be labeled in ascending order from 1 to k (Fig. 1). The labels of the centroids represent a health state and are transferred to the single pixels.
The resulting centroids for $k = n_{class} = 10$ are depicted in Fig 1. In order to avoid impairments from noisy data in the spectral border regions, a reduced spectrum from 470 nm to 750 nm was used. The figure reveals that the order is partly disturbed, presumably by effects of shadowing and leaf inclination. Therefore we applied $Crit_{order} = mRENDVI - PSRI$ to restore the desired order of the centroids by a function of two Vegetation Indices (VI) which are related to chlorophyll content and senescence processes. The combination of these evaluates the spectra at widespread wavelengths (445nm, 500nm, 680nm, 705nm, 750nm).

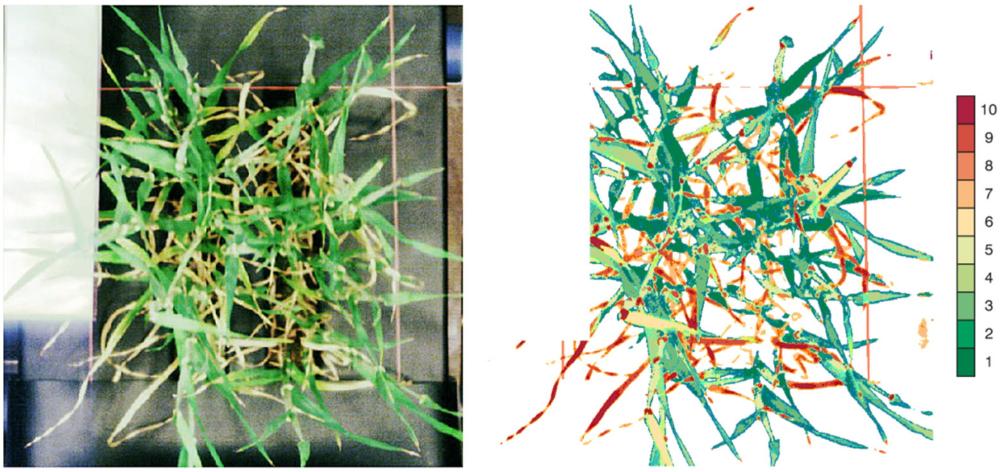

Figure 2: The result of the unsupervised labeling that incorporates the k-means algorithm and the ordinal order. Left is a RGB visualization of the hyperspectral image and right are the labels with 10 classes.

3.2. Model selection

In this section, we compare the performance of different prediction algorithms. The data set consists of five VIs derived from hyperspectral images of barley plants under drought stress.

3.2.1. Model types

The description of the prediction model types is based on Behmann et al. (2014a). The evaluated model types cover a broad range of prediction approaches. From very general multi-class SVMs via very specific ordinal prediction models to the Support Vector Regression (SVR) which predicts continuous target variables.

3.2.1.1. Support Vector Machine

The Support Vector Machine (SVM; Cortes et al., 1995) is an established classification method that determines the optimal, linear discriminant function between two classes based on the maximum margin principle. Extensions of this method handle overlapping classes and even non-linear discriminant functions. Multi-class tasks are handled in general by decomposing the multi-class problem in multiple binary class problems.

The **one-vs.-one (ovo) SVM** is the most common multiclass approach. It is based on pairwise classification, separating all classes from each other. In the learning step, a discrimination function is optimized for each class pair resulting in $\frac{n*(n-1)}{2}$ discrimination functions for n classes. Each optimization uses only the training samples of the regarded pair of classes.
The optimization is quite efficient because the amount of training data for a single optimization is small (Duan & Keerthi, 2005). However, the number of optimization procedures increases with the square of the number of classes. A high number of classes result in many, potential unnecessary, discriminate function.

The **one-vs.-all (ova) SVM** consists of discrimination functions which separate one class from all other classes, wherefore it is also called one-vs.-the-rest approach. The discriminant functions are determined by separating the training samples of one class from the aggregated training samples of all other classes. The model is more compact as only n discriminant functions are needed to separate n classes (Duan & Keerthi, 2005).

The one-vs.-all multiclass approach is less common than the one-vs.-one. It is less robust against outliers because a single misleading discriminant function can impair the result quality significantly (Duan & Keerthi, 2005). However, using well-tuned SVM classifiers comparable result qualities are achievable. Each of the binary discriminant functions suffers from a class imbalance since one class is separated from all the others. Moreover, the usage of all training data for each optimization can reduce the performance in the training step. Whereas in the one-vs.-one approach only two classes contribute to a discriminant function, in the one-vs.-all approach all classes contribute to all discriminant functions. However, the number of SVM evaluations is lower than in the one-vs.-one approach: each of the discriminant functions has to be evaluated for a prediction but the number of functions is lower, especially for higher numbers of classes.

3.2.1.2. Ordinal Support Vector Machine

The ordinal classification is applicable in scenarios with discrete labels and known class order. As it is a special case of the general multi-class scenarios, the introduced general prediction methods can be used. Prediction methods which are more adapted to the ordinal structure utilize the additional knowledge about the data set (Behmann et al., 2014b) and achieve higher performance measurements. In general, the information about the ordinal data structure is used to reduce the model size by omitting model parts which are not required Regression (Chu & Keerthi, 2007). Different approaches were developed which differ in specific assumptions on data characteristics and the robustness against non-ordinal aspects.

Support Vector Ordinal Regression (SVORIM) was developed by Chu & Keerthi (2007) in an explicit and an implicit formulation. Both formulations determine $c-1$ parallel hyperplanes that separate c classes and preserve the natural ordinal ordering. The parallelism of the hyperplanes reduces model size and complexity significantly. The linear model comprises only a single weight vector w for the whole model and a threshold b_i for each of the $c-1$ hyperplanes.
The model is limited regarding non-linear ordinal processes or non-ordinal aspects. However, the SVORIM prediction step is extremely efficient and comprises only a multiplication with the weight vector and the application of the $c-1$ thresholds. The Support Vector Ordinal Regression represents the most compact model with the lowest model complexity but the low complexity is accompanied by a low adaptability to deviating data characteristics. This may reduce the prediction accuracy on real-world data sets.

The **Linear Ordinal SVM (LOSVM)** is defined by discriminant functions between classes which are neighboring on the ordinal scale like at the Support Vector Ordinal Regression (Chu & Keerthi, 2007). Deviating from this approach, the hyperplanes are not forced to be parallel but are optimized locally. The number of discriminant function remains at $c - 1$ but the number of model parameter is significantly higher because each discriminant function has an individual weight vector w_i (Behmann et al., 2014b).
The improved flexibility of the model is apparent but in the regions without training instances, the discriminant functions may intersect each other. Without additional information, these regions of intersections are undefined. Therefore, a tree structure is introduced to unambiguously assign a class for each part of the feature space (Behmann et al., 2014b). In the tree structure a hierarchy of classes is established by an interval bisection approach. The discriminant functions can be represented by various classification approaches, e.g. SVM, random forests, logistic regression or naive Bayes.
In this study, the Linear Ordinal SVM classification is represented by the aggregate of all discriminant functions and the tree structure is used for class prediction. In the prediction, the number of evaluation steps is reduced by using the tree structure.

The concept of Linear Ordinal SVM lies between the flexible one-vs.-one multi-class approach and the extreme compact but inflexible Support Vector Ordinal Regression (Chu & Keerthi., 2007). It is able to represent also non-linear ordinal processes but it still relies on the ordinal data characteristics. Non-ordinal aspects cannot be represented due to the reduced number of discriminant functions compared to the generic multi-class approaches. The Linear Ordinal SVM classification results in much more compact models compared to one-vs.-one classification and may adapt to real-world data sets with only slight losses in accuracy.

3.2.1.3. Support Vector Regression

The main difference between Support Vector Regression (SVR) and SVM is the type of target variable. Regression algorithms predict continuous, real-valued labels in contrast to the discrete classes of classification models (Smola & Schölkopf, 2004). This is reflected in the optimization algorithm which adapts the basic principle of the binary SVM and results in similar formulas. The SVR was designed to find a regression function based on training instances which are continuously distributed in the feature values as well as the labels.

The SVR is able to model also ordinal data sets but the approximation errors will reduce the prediction quality for ordinal classification data sets. The kernels provide linear and non-linear model types. The linear SVR is an extremely compact model but achieves an inferior accuracy and, therefore, we applied the SVR with a radial basis function (rbf) kernel. This model is able to represent the ordinal transition with a competitive accuracy. The increased accuracy is accompanied by a higher model complexity due to the non-linear kernel function.

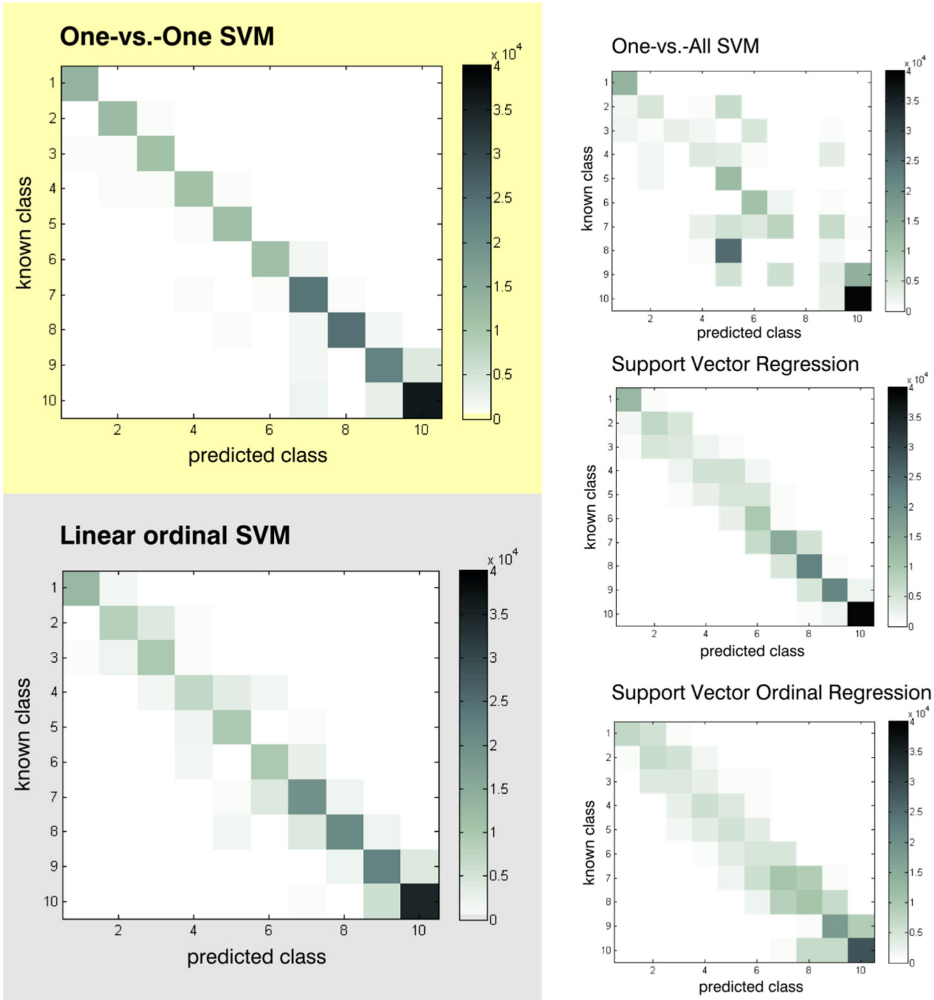

Figure 3: Confusion matrices of the investigated prediction algorithms. The one-vs.-one SVM is used as the golden standard because it represents the capable and flexible model. The Linear Ordinal SVM achieves comparable results with much lower model complexity and out-performs the three other models.

3.2.2. Evaluation

The results for the barley data set represent the performance in real world applications and are based on a study presented in Behmann et al. (2014a) (Fig. 3 and Table 1). The prediction algorithms can be separated in two groups: a good accuracy is achieved by the on-vs.-one SVM (83%), the SVR (66%) and the Linear Ordinal SVM (70%); an inferior accuracy is achieved by the on-vs.-all SVM (46%) and the Support Vector Ordinal Regression (47%).

Table 1: Performance comparison of the presented prediction algorithms for the task of the ordinal labeled hyperspectral image. The accuracy and the mean square error (MSE) are opposed to the quantity of functions and parameters which represents the complexity of the model.

	ovo-SVM	ova-SVM	SVORIM	LOSVM	SVR
Accuracy	83	46	47	70	66
MSE	0.80	1.90	1.02	0.82	0.72
#functions	45	10	9	9	1
#parameters	270	60	14	54	43986

The loss of accuracy of the SVORIM compared to the remaining methods is significant and is related to the data characteristics. The real-world data incorporating non-linear development of features over the ordinal scale cannot be described by parallel five-dimensional hyperplanes.
The lowest accuracy is achieved by the one-vs.-all SVM. This effect is most probably related to the low number of features (five features and ten classes). Linear discriminant functions seem not to be able to separate one of the ten senescence classes from the others. This effect can be faced by using more features but this would increase data volume as well as model complexity.

The rbf SVR reaches competitive accuracy comparable to the one-vs.-one SVM and the Linear Ordinal SVM. The MSE value is the lowest of all methods related to the continuous predictions. Such output enables further evaluations like probability extraction and a more detailed visualization. However, its non-linear kernel increases the model size up to a factor of 800. Such a model size prevents a high-throughput prediction as it is required for the efficient evaluation of hyperspectral images. Therefore, it is not suited to be applied for the introduced phenotyping scenario.

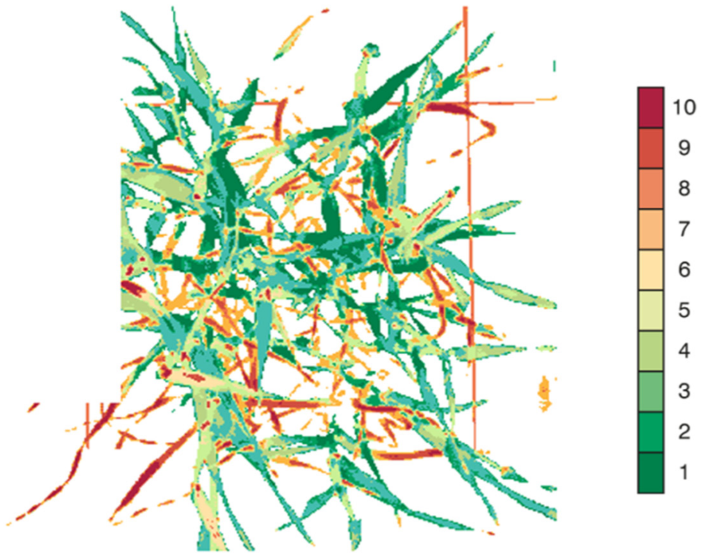

Figure 4: Pixelwise predictions based on the Linear Ordinal SVM model. Compared to the one-vs.-one SVM with a much higher model complexity a similar accuracy is reached. It becomes clear that a drought stressed barley plant is composed of senescent and vital leaves and that the class distribution is capable to separate the different treatment groups.

The one-vs.-one SVM and the Linear Ordinal SVM reach an almost identical MSE value. However, the accuracy of the one-vs.-one approach is 13\% higher. The combination of both result quality measurements reveals the classification characteristics. The one-vs.-one classifies more test samples correctly but if a test sample is misclassified it is more likely assigned to a more distant class. In contrast, the Linear Ordinal SVM assigns the misclassified samples in the most cases to one of the two neighboring classes. For the detection of disperse drought stress effects, the overall impression is most important (Fig. 4). It is not or only little affected by misclassifications to neighboring classes because these classes have nearly the same meaning with regard to the senescence level. Therefore, the higher accuracy of the one-vs.-one SVM approach has only slight positive effects on real-world applications but this advantage is at the expense of a five times higher number of model parameters.

4. DETECTION OF DROUGHT STRESS IN HYPERPSECTRAL IMAGES OF BARLEY

The health states of the individual plants are described by the relative frequency of the single classes within the set of plant pixels. The senescence descriptions of each plant are arranged on a scale, ranging from vital to senescent. The separability of the treatments (unwatered, reduced watered and well-watered) allows to draw conclusions about the temporal development of drought stress. The day at which treatments can be separated allows to evaluate the quality of data analysis. For the arrangement of a single plant at a single day on the senescence-scale, a further linear SVM model is derived, which analyzes the aggregated histogram features on plant scale instead of spectral features on pixel scale. We use the SVM score, which is defined as the normalized distance to the discrimination function, to evaluate the histograms of relative class frequencies. Figure 5 shows the development of the p-values of a two sample t-test. The black line represents a significance level of $\alpha = 0.05$. The premature and accelerated senescence of the drought stressed treatments becomes visible by the separability from the watered control group. A p-value below the significance border stands for successful separation at the corresponding day.

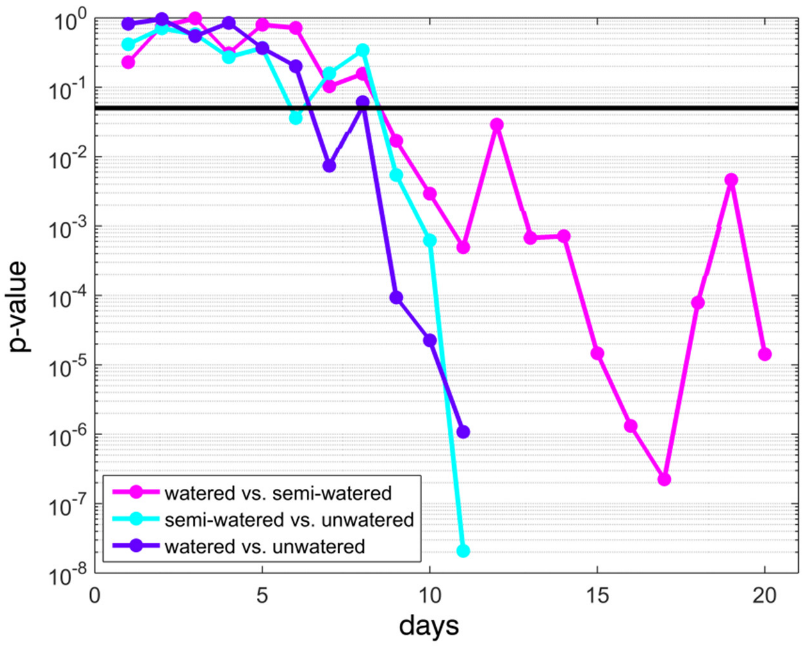

Figure 5: Result of the two sample t-test on the decision values of the three treatments. The black line at 0.05 indicates the significance level of 0.05. The three treatments are separable starting at day 8 but, as expected, the unwatered and watered plants are separable two days earlier.

Obviously, from day 1 to 5, the extracted vitality functions develop in a similar manner; at day 6, the functions of unwatered and semi-waters plants start to diverge, and at day 7 also the divergence of the function of reduced watered plants is visible. Significant differences between well-watered and unwatered plants are obtained, starting at day 8; differences between all treatments are obtained from day 9. In Behmann et al. (2014b) we showed that this approach allows a significantly earlier separation of the treatments than the application of singe VIs.

5. CONCLUSIONS

“Phenomics” has been identified as “the next challenge” in Nature in 2010. It is expected that the availability of new sensor technology on the one hand and the ability to interpret their signals on the other hand provides new insights into the complex interactions between genotype and environment, augmenting and complementing molecular biological approaches which attracted so much attention in the last decade. In crop sciences hyperspectral sensors attained specific interest since they allow a noninvasive estimation of physiological parameters of crop. Since they are nondestructive they are predestined to the study of processes which are relevant for yield increase and stress tolerance. With regard to global challenges of world population growth on the one hand and climate change on the other hand, in short the open question of “how to feed the world in 2050” a better understanding of these processes is of utmost interest.

In Bonn we started to study these questions in the DFG Graduate School” Information Techniques on Precision Plant Protection in Agriculture and Horticulture” from 2001 to 2010. This interdisciplinary collaboration was continued in the BMBF network of Excellence Cropsense from 2010 to 2014. Its goal was to support the use of modern sensor technology for plant phenotyping to improve the selection efficiency in plant breeding and field management. This network constituted a novel cooperation between biologists and agronomists on one side and geodesists, computer scientists and mathematicians on the other hand. Agronomists realized that the heterogeneity of the observed processes and the complexity of the applied sensors overstrained methods of data interpretation which have been applied in this field so far. We contribute 3D /4D models representing the geometric aspect of the plants phenotype and apply advanced methods of supervised and unsupervised classification in order to provide relevant information on the physiological processes inside the plant. In fact, hyperspectral imaging, appropriately analyzed, opens a new window into the interior of the plant, making visible what has been not visible so far.

6. ACKNOWLEDGEMENTS

The authors further acknowledge the funding of the CROP.SENSe.net project in the context of Ziel 2-Programms NRW 2007–2013 “Regionale Wettbewerbsfähigkeit und Beschäftigung (EFRE)” under the aegis of the Ministry for Innovation, Science and Research (MIWF) of the North Rhine Westphalia (NRW) federal state as well as the receipt of European Union Funds for regional development (EFRE) (005-1103-0018) for the preparation of this manuscript.

7. REFERENCES

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

Behmann, J., Schmitter, P., Steinrücken, J. & Plümer, L. (2014a): Ordinal classification for efficient plant stress prediction in hyperspectral data. International Archives of Photogrammetry, Remote Sensing and Spatial Informaition Science XL-7, pp. 29-36.

Behmann, J., Steinrücken, J. & Plümer, L. (2014b): Detection of early plant stress responses in hyperspectral images. ISPRS Journal of Photogrammetry and Remote Sensing 93, pp. 98-111.

Chu, W. & Keerthi, S. (2007): Support vector ordinal regression. Neural Computation 19 (3), pp. 792-815.

Cortes, C. & Vapnik, V. (1995): Support-vector networks. Machine learning 20 (3), pp. 273-297.

Duan, K. & Keerthi, S. (2005): Which is the best multiclass SVM method? An empirical study. In: Proceedings of the 6th International Workshop on Multiple Classifier Systems, Springer, pp. 278-285.

FAO (2011): Declaration of the world summit on food security. Food and Agriculture Organisation (FAO) of the United Nations, 16-18 November, Rome. [last access: 12.11.2013]

Gitelson, A. A., Kaufman, Y. J., Stark R. & Rundquist, D. (2002): Novel algorithms for remote estimation of vegetation fraction. Remote Sensing of Environment, 80, pp. 76-87.

Munné-Bosch, S. & Alegre, L. (2004): Die and let live: leaf senescence contributes to plant survival under drought stress. Functional Plant Biology 31 (3), pp. 203-216.

Oerke, E.-C. & Dehne, H.-W. (2004): Safeguarding production – losses in major crops and the role of crop protection. Crop Protection, 23, pp. 275-285.

Smola, A. J. & Schölkopf, B., (2004): A tutorial on support vector regression. Statistics and Computing 14 (3), pp. 199-222.

Taiz, L. & Zeiger, E. (2010): Plant physiology. 5 edn, Sinauer Associates, Sunderland and Mass.

Tester, M. & Langridge, P. (2010): Breeding technologies to increase crop production in a changing world. Science 327 (5967), pp. 818-822.