OPTIMAL WAVELENGTHS FOR AN EARLY IDENTIFICATION OF *CERCOSPORA BETICOLA* WITH SUPPORT VECTOR MACHINES BASED ON HYPERSPECTRAL REFLECTION DATA

Till Rumpf[1], Anne-Katrin Mahlein[2], Christoph Römer[1] and Lutz Plümer[1]

[1] Institute of Geodesy and Geoinformation, Department of Geoinformation, University of Bonn, Meckenheimer Allee 172, D-53115 Bonn, Germany
[2] Institute of Crop Science and Resource Conservation (INRES – Phytomedicine), University of Bonn, Nußallee 9, D-53115 Bonn, Germany

Automatic classification of plant diseases at an early stage is vital for precision crop protection. Our aim was to identify sugar beet leaves that were inoculated with *Cercospora beticola* based on hyperspectral reflection data between 400 and 1050 nm. In this early stage, the difference between the medians of the two classes (healthy leaves and inoculated leaves) is smaller than their respective standard deviations (Fig .1). Hence, the classifier has to be able to deal with highly overlapping classes and needs superior generalization abilities. Support Vector Machines (SVMs) have statistical properties fulfilling these requirements.

Relevant wavelengths have to be identified in order to implement practical sensor systems with reduced development costs. The main contribution of this study is the identification of a minimal subset which is sufficient for differentiation between both classes. Furthermore, using all wavelengths is critical as the classifier could randomly separate both classes due to their noise instead of finding general patterns. We applied the filter algorithm of Hall [1]. This heuristic analyses the relevance of a feature subset considering the intercorrelation among the features. However, an exhaustive search of all possible subset combinations of 462 wavelengths is not feasible. Therefore, in order to find a good subset in a reasonable amount of time a genetic algorithm was applied. The solution consists of six bands.

Using this subset as feature space for the SVM to identify low disease severity ($\leq 5\%$) a classification accuracy of 83% was achieved. Disease severity above 5% was classified with 99.7%. The selected wavelengths resulted in even better classification performance than a comparable approach based on vegetation indices as features on the same dataset [3]. To improve the performance the size of the subset would have to be considerably increased. However, using more wavelengths with regard to the number of available samples would possibly result in overfitting.

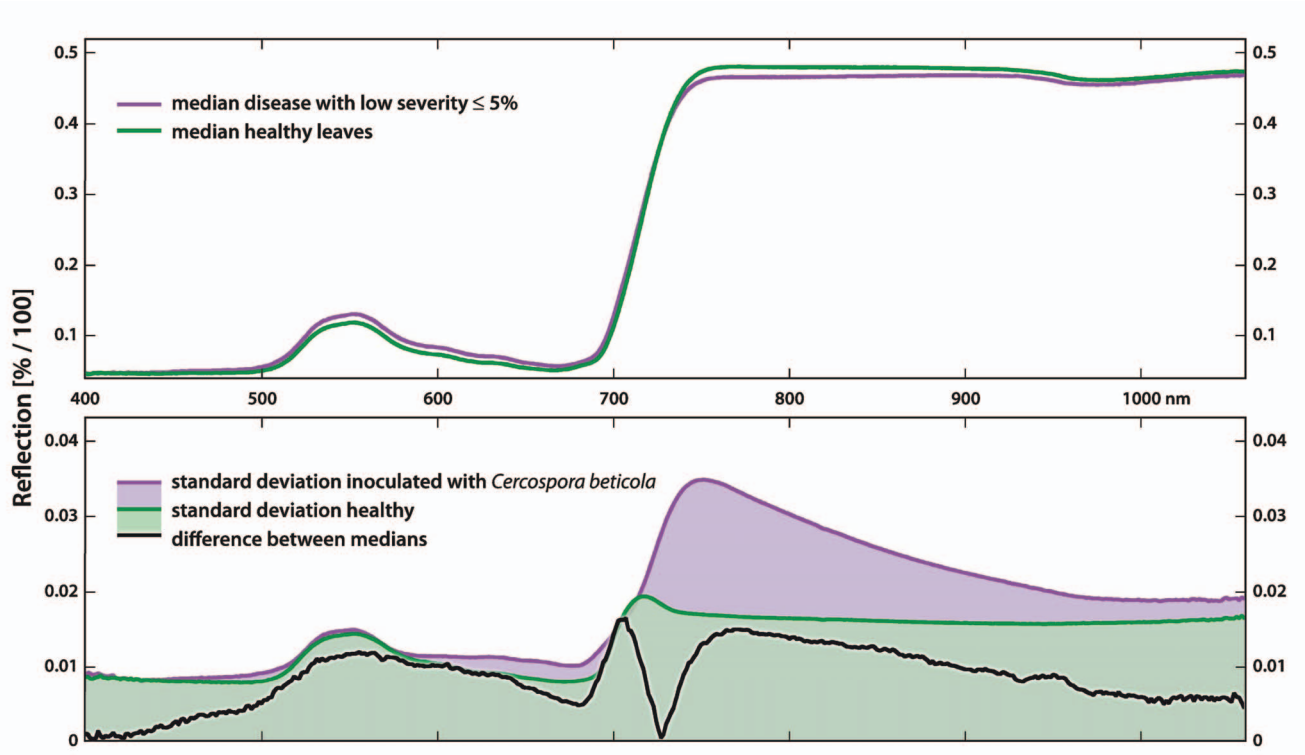

Fig. 1 **Reflection spectra healthy versus inoculated leaves: medians and intra-class deviations.**

A feasible way to avoid this would be to construct features which aggregate the information of certain wavelength ranges [2]. This, however, still needs an identification of relevant bands with feature selection algorithms and does not restrict the number of observations to just six wavelengths.

References:

[1] Hall, M.A., "Feature selection for discrete and numeric class machine learning", Proc. 7th International Conference on Machine Learning, San Francisco, USA, Morgan Kaufmann Publishers, 359-366, 2000.

[2] Römer, C., Rumpf, T., Bürling, K., Hunsche, M., Noga, G. and Plümer, L., „Early identification of leaf rust on wheat leaves with robust fitting of hyperspectral signatures", Proc. 10th International Conference Precision Agriculture, Denver, USA, 2010 (accepted).

[3] Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W. and Plümer, L., „Early Detection and Classification of Plant Diseases with Support Vector Machines based on Hyperspectral Reflectance", Computer and Electronics in Agriculture, 2010 (submitted).