

Applications

Florian Becker*, Andreas Backhaus, Felix Johrden and Merle Flitter

Optimal multispectral sensor configurations through machine learning for cognitive agriculture

Optimale multispektrale Sensorkonfigurationen mittels maschineller Lernverfahren für die kognitive Landwirtschaft

<https://doi.org/10.1515/auto-2020-0069>

Received April 28, 2020; accepted January 22, 2021

Abstract: Hyperspectral sensor systems play a key role in the automation of work processes in the farming industry. Non-invasive measurements of plants allow for an assessment of the vitality and health state and can also be used to classify weeds or infected parts of a plant. However, one major downside of hyperspectral cameras is that they are not very cost-effective. In this paper, we show, that for specific tasks, multispectral systems with only a fraction of the wavelength bands and costs of a hyperspectral system can lead to promising results for regression and classification tasks. We conclude that for the ongoing automation efforts in the context of cognitive agriculture reduced multispectral systems are a viable alternative.

Keywords: cognitive agriculture, machine learning, spectral band selection

Zusammenfassung: Hyperspektrale Sensorsysteme spielen eine Schlüsselrolle bei der Automatisierung von Arbeitsprozessen in der Landwirtschaft. Nicht-invasive Messungen von Pflanzen ermöglichen eine Beurteilung des Vitalitäts- und Gesundheitszustands und können auch zur Klassifizierung von Unkraut oder infizierten Pflanzenteilen verwendet werden. Ein großer Nachteil von Hyperspektralkameras ist jedoch, dass sie nicht sehr kosteneffektiv sind. In diesem Beitrag zeigen wir, dass für bestimmte Aufgaben multispektrale Systeme mit nur einem Bruchteil der Wellenlängenbänder und Kosten eines Hyperspektralsystems zu vielversprechenden Ergebnissen bei Regressions-

und Klassifikationsaufgaben führen können. Wir kommen zu dem Schluss, dass für die laufenden Automatisierungsbemühungen im Rahmen der kognitiven Landwirtschaft reduzierte multispektrale Systeme eine praktikable Alternative sind.

Schlagwörter: Kognitive Landwirtschaft, Maschinelles Lernen, Auswahl von spektralen Bändern

1 Introduction

1.1 Motivation

The automation of field machinery and work processes has continuously transformed the farming industry in the last decades. In recent years, the automation of seeding, crop maintenance and harvesting has been accompanied with the task of real-time monitoring of crop state. Central for this task is the development of sensor systems, that are able to non-invasively measure the biochemical proposition of the crop plant online and inline in order to generate information for selective crop treatment and general farm monitoring. Artificial intelligence and particular Machine Learning approaches play a key role in translating the raw sensor data into usable information. The potential of cognitive agriculture lies in the reduction of physically demanding labour as well as increased quality of fresh produce.

Optical sensor based systems are able to meet these demands. Hyperspectral imaging plays a central role in developing optical sensor systems for cognitive agriculture. This camera technology goes beyond color or selective multispectral imaging systems and provides a systematically sampled measurement of the reflectance properties of a material. Commercially available system focus on the wavelength range of roughly 400 to 2500 nm.

In order to translate the acquired physical property into meaningful information, pattern recognition utilizing machine learning is increasingly used. This requires the

*Corresponding author: Florian Becker, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany, e-mail: florian.becker@iosb.fraunhofer.de

Andreas Backhaus, Felix Johrden, Fraunhofer Institute for Factory Operation and Automation IFF, Magdeburg, Germany, e-mails: Andreas.Backhaus@iff.fraunhofer.de, felix.johrden@iff.fraunhofer.de

Merle Flitter, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany, e-mail: merle.flitter@iosb.fraunhofer.de

systematical acquisition of reference data along with accompanying field ratings and lab measurement for ground truth generation. Consequently a part of the sensor system is adapted or learned towards the monitoring task at hand. This paradigm for generating a sensor system is often referred to as *soft-sensor* [8]. However, hyperspectral imaging systems are still very cost intensive sensor hardware. In order for productive use in farm equipment, this system has to be scaled down to a cost effective system. Therefore, a development process for reducing the hyperspectral imaging system to a multispectral system with task relevant wavelength is desirable.

Wavelength selection has been investigated both from a theoretical point of view as well as for a plethora of different applications. Including uninformative or redundant wavelengths can have a negative effect for calibration and regression tasks [18, 6]. An excellent review and taxonomy of wavelength selection methods with focus on the field of food quality inspection is given in [11]. In [10] the feasibility of multispectral imaging in the context of predicting phenolics of tomatoes infield is discussed. Our work is similar in nature. However, the focus is to examine the feasibility of different classification and regression tasks with respect to the number of selected wavelengths while pointing out differences among the different methods used.

1.2 Contribution

In this paper, we will present a number of different monitoring tasks and show that they can be sufficiently performed using hyperspectral imaging data. We present a number of strategies to reduce the sensor to a multispectral sensor using data-driven feature selection and multispectral simulation. We then validate the multispectral selection by re-performing data modelling in order to test detection performance compared to using the full spectral reflectance property, constituting evidence that a reduced multispectral approach is a viable alternative.

2 Material and methods

2.1 Datasets

For the following study, hyperspectral image recording from two different crop plant varieties were considered.

1. Barley: Spectral images recorded on barley breeding plots using a custom build field measurement system in 2016–2018 for agronomic and nutritional state prediction [7]. For the image recording a Norsk Elektro Op-

tikk A/S HySpex SWIR 384 (288 spectral channels) was used.

2. Wine: Spectral images recorded on wine leaves measured in field campaigns in 2016-2018 using a wine harvester equipped with hyperspectral camera systems for detecting vine diseases [9]. For the image recording a Norsk Elektro Optikk A/S HySpex VNIR 1800 and HySpex SWIR 384 was used.

A VNIR camera covers the visual-near infrared range from 400 nm to 1000 nm while a SWIR camera covers the shortwave-infrared range from 1000 nm to 2500 nm. For each dataset, measurement data is calibrated towards a PTFE spectralon representing a near 100 % reflector. Calibration is performed by subtracting the dark current I_b^{DC} , measured at closed camera shutter, and dividing by the signal when the spectralon I_b^{WH} is in the camera view port, where b denotes the spectral band index.

$$x_b = \frac{I_b - I_b^{DC}}{I_b^{WH} - I_b^{DC}} \quad (1)$$

Therefore the value is restricted to the range of 0.0-1.0. Each dataset was subjected to image segmentation in order to keep only the pixels which represent plant vegetation e. g., plant leaves. All other materials like soil, wine grapes, plant stems, crop ears and field markings were discarded by image segmentation done by trained neural network classifier models, which is not the focus of this study. In this study, we will consider the following detection and estimation tasks. Throughout the paper, we will use the abbreviation in brackets to refer to the corresponding data set:

1. Classification: Detection of weed plants vs. barley plants (C1 barley)
2. Classification: Detection of leaves of virus infected grapes vs. control leaves (C2 wine)
3. Regression: Estimation of relative days till heading in barley plants (R1 developmental)
4. Regression: Estimation of potassium content ($\mu\text{g/g}$ dry weight) in barley plants (R2 nutrition)

The task are selected since when using a hyperspectral data, they yielded high accuracy in their respective research projects. In this study, we want to investigate how a simulated multispectral system with bands selected from machine learning and feature selection approaches perform in comparison.

In the following, we will use $\mathbf{X} \in \mathbb{R}^{N \times B}$ to denote a data or predictor matrix, where each row r contains one reflectance spectrum $\mathbf{x}_r = (x_{r1}, \dots, x_{rB})$ with B bands.

2.2 Preprocessing

Vegetation cover recorded under field condition does not represent itself as a smooth planar surface but has complex geometrical features. It is known that these geometrical properties modulate the reflectance measurement in offset and scale [17, 15]. In order to minimize this effect, preprocessing on the reflectance profile is performed. In this study we considered the L2-norm correction and standard-normal variate (SNV) as preprocessing variants:

$$\mathbf{x}_r^{\text{SNV}} = \frac{\mathbf{x}_r - \bar{\mathbf{x}}_r}{s_r}, \quad (2)$$

$$\mathbf{x}_r^{\text{L2}} = \frac{\mathbf{x}_r}{\|\mathbf{x}_r\|}, \quad (3)$$

where $\bar{\mathbf{x}}_r$ is a vector that contains the sample mean \bar{x}_r B times, s_r is the sample standard deviation of \mathbf{x}_r and $\|\mathbf{x}_r\|$ is the L_2 vector norm of \mathbf{x}_r .

2.3 Multispectral reduction and simulation

High-resolution hyperspectral imaging is not a cost-effective way for the real-time monitoring of crop parameters. For this reason, multispectral systems can be designed specifically for the application at hand. The presented multispectral reduction of the presented hyperspectral applications has four steps.

1. A wavelength weighting is calculated with different methods in order to gain a relevance profile across the cameras wavelength range.
2. The wavelength weighting profile is used as probability density function (pdf) and a random generator generates 10,000 values according to this pdf.
3. A Neural Gas algorithm [12] is used to perform vector quantization on this generated data, naturally, regions of high weighting e. g., high sample density are covered more dense than regions of low weighting and also avoid selecting adjacent bands.
4. Response of Multispectral camera bands modelled as Gaussian functions of typical FWHM centered at the selected wavelengths are simulated.

2.4 Multispectral wavelength selection

The band selection approaches evaluated here are based on correlation, mutual information [5], class discriminance [4], radial basis function relevance [1] and PLS coefficients [16].

2.4.1 Band selection: PLS

Partial least squares regression can handle collinear data and is a suitable tool for selecting relevant predictors. The underlying description is given by

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (4)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}, \quad (5)$$

where \mathbf{X} is the $N \times B$ matrix of predictors and \mathbf{Y} contains the responses [19]. In general, \mathbf{Y} is a $N \times M$ matrix with M columns containing the multivariate responses. For the tasks considered in this paper, the responses are univariate, i. e., $M = 1$. Matrices \mathbf{E} and \mathbf{F} contain residuals that are assumed to be normally distributed. The scores matrix \mathbf{T} can be expressed in terms of the weighted original variables

$$\mathbf{T} = \mathbf{XW}, \quad (6)$$

with $\mathbf{W} \in \mathbb{R}^{B \times \ell}$ and ℓ being the number of components of the PLS model.

PLS simultaneously decomposes \mathbf{X} and \mathbf{Y} in terms of a scores and a loadings matrix by finding a set of latent vectors that explain the covariance between \mathbf{X} and \mathbf{Y} such that the responses can also be approximated in terms of the score matrix of \mathbf{X} , i. e.,

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{G} \quad (7)$$

$$= \mathbf{XWQ}^T + \mathbf{G}. \quad (8)$$

Finally, the regression coefficients that are used as a basis for the band selection in this paper are given by

$$\mathbf{C} = \mathbf{WQ}^T. \quad (9)$$

Model selection needs to be performed in order to decide on the number of columns of \mathbf{W} in Equation 6. Model selection for a PLSR means finding the suitable number of latent variables ℓ , such that the final model does not overfit the data. The coefficient of determination (R^2) and the root mean squared error (RMSE) are important metrics to evaluate a model and select an optimal model size. Figure 1 (a) plots both these metrics. The best predictive model is determined by cross-validation. PLS coefficients for variable selection have been successfully used to select important predictors in omics-type data [16]. A ranking of the hyperspectral bands can be easily obtained by the magnitude of the absolute coefficient values.

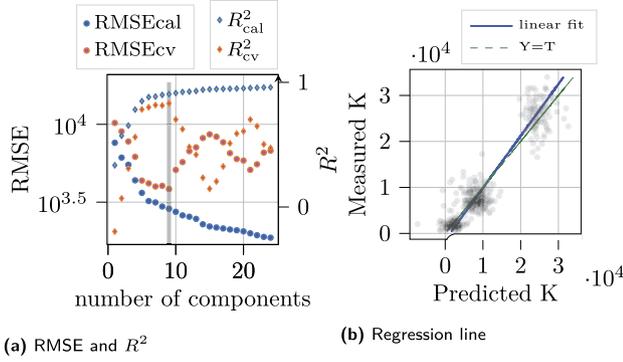


Figure 1: Model selection (i. e., selection of an appropriate number of latent variables) with 5-fold cross-validation (cv) and model calibration (cal) for potassium regression (R2 nutrition data set). The number of selected components is $\ell = 9$, after which R_{cv}^2 decreases and RMSEcv increases.

2.4.2 Band selection: Correlation

Assuming a mean-centred predictor matrix \mathbf{X} , correlation of a band $\mathbf{X}_{:,b}$ with the quantitative target vector \mathbf{y} (regression problems) is computed as

$$r_{\mathbf{X}_{:,b}, \mathbf{y}} = \frac{\sum_{i=1}^n \mathbf{X}_{i,b} \mathbf{y}_i}{\sqrt{\sum_{i=1}^n \mathbf{X}_{i,b}^2} \sqrt{\sum_{i=1}^n \mathbf{y}_i^2}}, \quad (10)$$

yielding a weighting profile for the cameras wavelength range.

2.4.3 Band selection: Mutual information

Mutual information (MI) quantifies the dependence of two random variables and can be used for band selection for regression and classification problems. Let b denote a band index, then

$$I(\mathbf{X}_{:,b}; \mathbf{y}) = \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{X}_{:,b}, \mathbf{y}) \log \left(\frac{p(\mathbf{X}_{:,b}, \mathbf{y})}{p(\mathbf{X}_{:,b}) p(\mathbf{y})} \right), \quad (11)$$

measures the joint mutual information between the selected band b and the target vector \mathbf{y} . By computing $I(\mathbf{X}_{:,b}; \mathbf{y})$ for all bands a ranking can be obtained. In contrast to the selection based on a PLS approach, MI is purely information-theoretic and does not require to first find a suitable model.

2.4.4 Band selection: Discrimination

The discrimination value quantifies the ratio of between-class scatter to the sum of within-class scatter. The ratio

therefore attributes how separable classes are and is derived from Fisher's linear discriminant [4]. The ratio is calculated as

$$D(\mathbf{X}_{:,b}) = \frac{\sum_{c \in \mathcal{C}} n_c (\mu_b^c - \mu_b)^2}{\sum_{c \in \mathcal{C}} \sum_{r \in \mathcal{R}} (x_{rb}^c - \mu_b^c)^2}, \quad (12)$$

where μ_b^c is the mean reflectance value of class c in band b , n_c the number samples in class c and μ_b the overall mean reflectance in band b across all C classes.

2.4.5 Band selection: Relevance

For determine the wavelength relevance profile, a radial basis function network with relevance [1] is trained on the dataset \mathbf{X} using the following loss function for the network output \mathbf{y} and target \mathbf{t} :

$$E(\mathbf{X}, \mathbf{W}, \boldsymbol{\lambda}) = \frac{1}{2} \sum_r \sum_k \{y_k(\mathbf{x}_r) - \mathbf{t}_k^r\}^2 \quad (13)$$

$$y_k(\mathbf{x}_r) = \sum_n u_{nk} \phi(d(\mathbf{v}_r, \mathbf{w}_n, \boldsymbol{\lambda})) \quad (14)$$

$$d(\mathbf{x}_r, \mathbf{w}_n, \boldsymbol{\lambda}) = \sum_b \lambda_b (x_{rb} - w_{bn})^2 \quad (15)$$

A radial basis function network is a two layer neural network. In the first layer prototypical samples \mathbf{W} are placed in the input space and an activation is calculated using the Euclidean distance and a Gaussian function ϕ . In a second layer, a weighted sum is generating the output values. For regression the target is set to the quantitative value, for classification, a 1-out-of-N coding scheme is applied to generate the target pattern. Since the network also optimizes the weighting factors λ_b , these can be interpreted as band relevance and are used as weighting profile.

2.4.6 Band placement and multispectral simulation

Each method described above generates a weighting profile across the wavelength range. In order to assess the hypothetical performance of a multi-band multispectral system, we need to place typical wavelength band centered around highly weighted wavelengths. The placement should cover highly weighted segments but avoid merely selecting neighbouring wavelength since due to the high correlation of reflectance in neighbouring wavelengths. For this purpose, we sampled from a probability density function, that matches the weighting profile, 10,000 wavelength candidates and trained a standard neural gas algorithm [12] on this data. The neural gas places a set of prototypes in the data space in order to minimize the quantization error calculated by the Euclidean distance. After

convergence was achieved, we used these prototypes as mid-wavelengths for Gaussian kernels set to a Full Width at Half Maximum (FWHM) of 30 nm, a typical value for multispectral camera systems. We generated ten wavelength candidates. These wavelength *prototypes* are then ordered due to the highest weighting and two to ten bands are selected for simulation in descending order of their weighting. Subsequent modelling of this simulated data is done to check detection performance.

The hypothetical response of a multispectral imaging system for a reflectance spectrum \mathbf{x}_r of a band centred at λ_m is calculated by

$$r_{mm}^{MSI} = \frac{1}{\sum_b g_m(\lambda_b)} \sum_b x_{rb} g_m(\lambda_b) \quad (16)$$

$$g_m(\lambda_b) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-0.5 \left(\frac{\lambda_b - \mu_m}{\sigma}\right)^2\right) \quad (17)$$

$$\sigma = \frac{FWHM}{2\sqrt{2 \ln 2}} \quad (18)$$

where λ_b denotes the spectral bands wavelength.

2.4.7 Machine learning setup

In order to make the modelling performance on hyper- and multispectral data comparable, the same model types are used. Among the model types, the best performing type in combination with the listed pre-processing given by Equation 2 and 3 is reported. The used methods and hyperparameters are given in Table 1.

Table 1: Machine learning methods and hyperparameters used in this study.

Method	Citation	Hyper-Parameter	Used for
MLP	[2, 14]	Hidden layers: 3 Optimization method: Scaled Conjugate Gradient Back-Propagation Hidden layer: 50, 25, 10	hyperspectral multispectral
PLS(R/DA)	[20]	number of components (optimized for each task separately)	hyperspectral multispectral band selection
rRBF	[13, 3, 1]	Base functions: 30 Optimization method: Scaled Non-linear Conjugate Gradient	band selection
LDA	[4]	no hyper-parameters	band selection

3 Results

Before we proceed with the performance results of the simulated multispectral system, we show the regression and classification performances of the above mentioned tasks given the full spectrum.

3.1 Hyperspectral performance

The performance for the hyperspectral data was assessed by means of a PLS(R/-DA) as well as a neural network based approach and is summarized in Tables 2 and 3. For the evaluation with PLS and Neural Networks (NN), the number of latent variables was determined by 5-fold cross-validation, equivalent to the band selection procedure described above. For simplification we state the best performance for the used neural network models under NN.

Table 2: Full-spectrum PLSR and NN 5-fold cross-validated results for the barley data sets. The R2 nutrition data set is about the estimation of the nutrition content (potassium in $\mu\text{g/g}$ dry weight), the R1 developmental data is about predicting the relative number of day to heading (HEA).

Location	Target	Preproc.	PLSR	NN	PLSR	NN
			R_{cv}^2	R_{cv}^2	MAE _{cv}	MAE _{cv}
R1 develop.	HEA	SNV	0.92	0.96	2.53	1.71
R2 nutrition	K	SNV	0.93	0.94	2179.70	1483.93

Our results for the multispectral reduction that is detailed in the following, indicate that for the classification and regression tasks considered in this paper multispectral systems are a feasible option.

Table 3: Full-spectrum PLS-DA and NN 5-fold cross-validated results for the classification data sets C1 barley and C2 wine. Performance measure: Accuracy (ACC), True-Positive Rate (TP) and False-Positive Rate (FN).

Type	Preproc.	PLS-DA	NN	PLS-DA	NN	PLS-DA	NN
		ACC	ACC	TP	FN	TP	FN
C1 barley	L_2	0.86	0.953	0.880	0.161	0.935	0.030
C2 wine	SNV	0.90	0.925	0.940	0.161	0.939	0.088

3.2 Multispectral performance: Classification tasks

We will first discuss the results on the detection tasks, i. e., the task of classifying weed plants vs. barley plants and the task of detecting virus infected grape leaves vs. control leaves. Figures 3 and 2 show the spectra and the ten bands that were selected by the different methods described above. The plots also include the normalized weighting profiles for the different methods. It is important to note that the selected bands do not necessarily match the local maxima of the weighting profiles. This is due to the fact that the neural gas algorithm quantizes the profile such that highly weighted regions are selected. There are some regions in the spectrum that all three methods found to be necessary for the multispectral reduction, e. g., for the detection of virus infected grape leaves, the wavelength band around 1500 nm seems to convey crucial information.

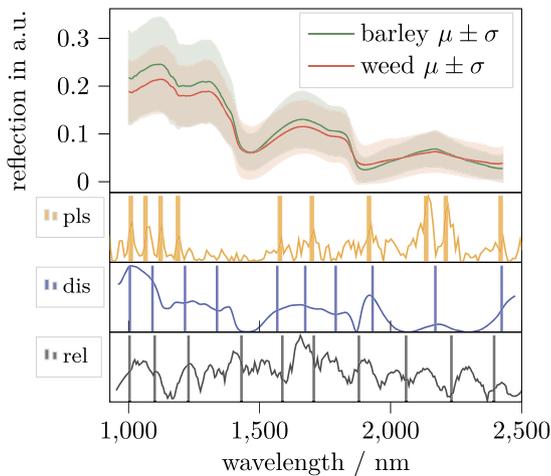


Figure 2: C1 barley results. Reflection spectra of barley vs. weed. Below the spectra are the different band selection positions based on PLSR, discrimination and relevance.

Figure 4 shows the performance of a reduced multispectral system with an increasing number of channels. For all evaluated methods, the wavelength channels are sorted according to their importance, beginning with the channel yielding the highest discriminability. For the weed vs. barley task, the six channels determined by the discrimination and the relevance approach are sufficient for a classification accuracy of above 90%.

The results indicate that a carefully designed multispectral system with ten channels could reach the performance of a hyperspectral system for certain tasks.

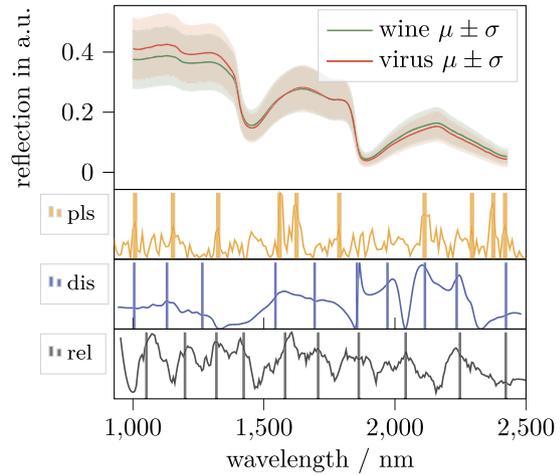


Figure 3: C2 wine results. Reflection spectra of virus infected grape leaves and control leaves. Below the spectra are the different band selection positions based on PLSR, discrimination and relevance.

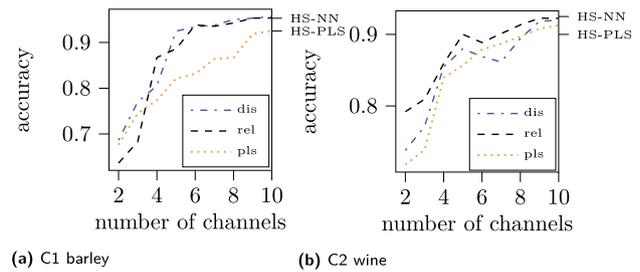


Figure 4: Classification performance of the simulated multispectral system with an increasing number of channels, starting with the channels contributing most to the classification task. The performance of the full-spectrum NN and PLS-DA is given by HS-NN and HS-PLS.

3.3 Multispectral performance: Regression tasks

Figures 6 and 5 show the bands that were selected for the two different regression tasks, i. e., the task of estimating the nutrition content (potassium) and the developmental parameter. For the former task, there is a cluster of selected bands in the region from 1000 nm to 1350 nm, which is in contrast to the other approaches that select bands that are more spread out across the spectrum. Figure 7 shows the cumulative coefficients of determination for the different methods and for an increasing number of channels. For the regression of potassium the coefficient of determination is around 0.94 using only the most important channel and it is not increasing with additional ones.

It is important to note that in nearly all cases above, the wavelengths at the boundary of the spectrum were se-

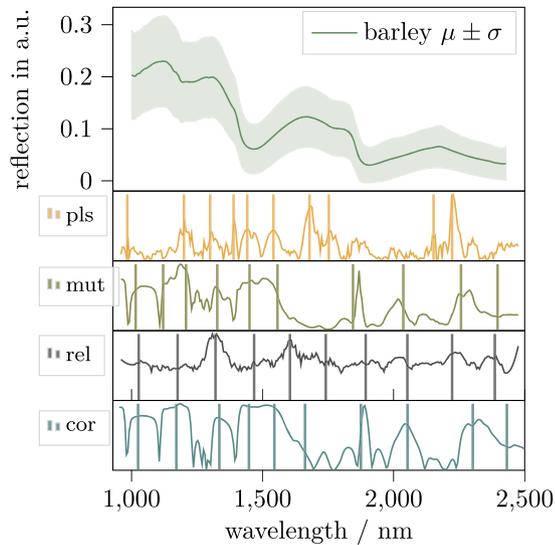


Figure 5: R1 developmental results. Reflection spectrum of barley. Below the spectra are the different band selection positions for the regression of the agronomic parameter based on PLSR, mutual information, discrimination and correlation.

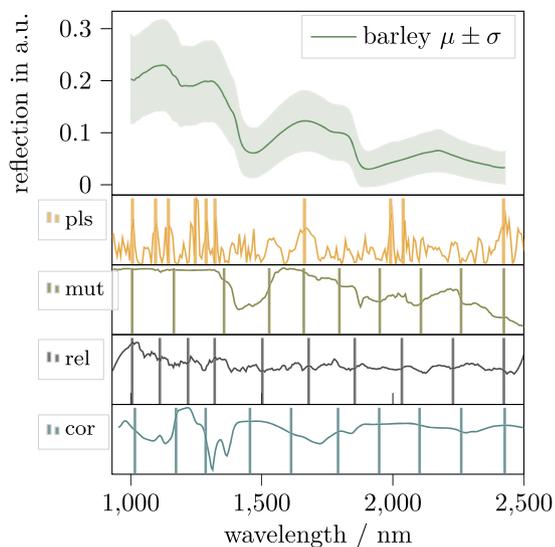
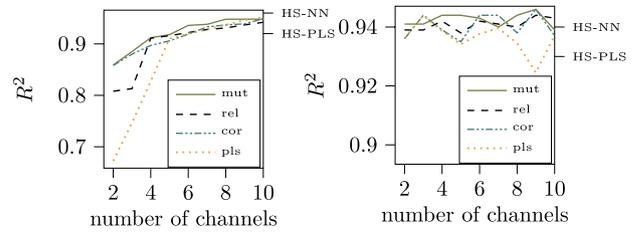


Figure 6: R2 nutrition results. Reflection spectrum of barley. Below the spectra are the different band selection positions for the regression of the nutrition parameter potassium based on PLSR, mutual information, discrimination and correlation.

lected by the different band selection schemes. However, as this regions are usually prone to noise, they should be discarded.



(a) R1 developmental

(b) R2 nutrition

Figure 7: Coefficient of determination of the simulated multispectral system with an increasing number of channels, starting with the channels contributing most to the regression task. The R^2 of the full-spectrum NN and PLSR is given by HS-NN and HS-PLS.

4 Conclusion

Hyperspectral sensor systems are vital for the detection of crop state. It is obvious that certain wavelength bands contribute more to the task at hand than others. This study has shown that by selecting up to ten most important bands a hyperspectral sensor system can be reduced to a multispectral one with comparable performance for the various tasks considered in this paper. The position of the selected bands is highly task-specific. Different wavelength selection approaches, even when they converged to similar results at ten wavelengths, showed considerable difference when selecting smaller numbers of wavelength. Notably the widely used method PLS performed the worst in most applications in selecting the proper wavelengths.

Since the biochemical processes underlying a detection task are not known or hard to evaluate, the wavelength selection is based on data analytics and only the performance of a model trained on selected data shows how suitable the selection is. For this purpose, a simulation of the multispectral camera system with realistic wavebands is necessary. This work constitutes evidence that different wavelength selection approaches lead to different results regarding the selected positions. Therefore, for a practical application, where the minimal number of necessary wavebands is aimed for, a number of selection methods should be considered and tested.

This study also shows the advantage of performing a hyperspectral measurement campaign first, since from such a dataset, possible multispectral systems can be evaluated. A practical approach would be to simulate a number of off-the shelf cameras and compare them with a custom build camera system based on optimal wavelength and then decide for candidate systems for the following validation campaign.

Acknowledgment: We like to acknowledge the group of Prof. Klaus Pillen from the Plant Breeding department at the University of Halle-Wittenberg, Germany, and the group of Prof. Reinhard Töpfer at the Julius Kühn Institute, Siebeldingen, Germany, who were performing the field trials and measurements in earlier research projects in partnership with the Fraunhofer IFF, that generated the application datasets used in this study.

References

1. Andreas Backhaus, Praveen C. Ashok, Bavishna B. Praveen, Kishan Dholakia and Udo Seiffert. Classifying scotch whisky from near-infrared Raman spectra with a radial basis function network with relevance learning. In *ESANN 2012 proceedings, 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 411–416, 2012.
2. G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCS)*, 2(4):303–314, December 1989.
3. Reza Dehghani and Nezam Mahdavi-Amiri. Scaled nonlinear conjugate gradient methods for nonlinear least squares problems. *Numerical Algorithms*, pages 1–20, 2018.
4. Ronald Aylmer Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.
5. Baofeng Guo, Steve R Gunn, Robert I Damper and James DB Nelson. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4):522–526, 2006.
6. Qing-Juan Han, Hai-Long Wu, Chen-Bo Cai, Lu Xu and Ru-Qin Yu. An ensemble of Monte Carlo uninformative variable elimination for wavelength selection. *Analytica Chimica Acta*, 612(2):121–125, 2008.
7. Paul Herzig, Andreas Backhaus, Udo Seiffert, Nicolaus von Wirén, Klaus Pillen and Andreas Maurer. Genetic dissection of grain elements predicted by hyperspectral imaging associated with yield-related traits in a wild barley nam population. *Plant Science*, 285:151–164, 2019.
8. Petr Kadlec, Bogdan Gabrys and Sibylle Strandt. Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 33(4):795–814, 2009.
9. Anna Kicherer, Katja Herzog, Nele Bendel, Hans-Christian Klück, Andreas Backhaus, Markus Wieland, Johann Rose, Lasse Klingbeil, Thomas Läbe, Christian Hohl, et al. Phenoliner: A new field phenotyping platform for grapevine research. *Sensors*, 17(7):1625, Jul 2017.
10. Changhong Liu, Wei Liu, Wei Chen, Jianbo Yang and Lei Zheng. Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit. *Food Chemistry*, 173:482–488, 2015.
11. Dan Liu, Da-Wen Sun and Xin-An Zeng. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food and Bioprocess Technology*, 7(2):307–323, 2014.
12. T. M. Martinetz and K. J. Schulten. A Neural-Gas network learns topologies. In *Artificial Neural Networks*, pages 397–402. North-Holland, Amsterdam, 1991.
13. John Moody and Christian J. Darken. Fast learning in networks of locally tuned processing units. *Neural Computation*, 1:281–294, 1989.
14. Martin Foddslette Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533, 1993.
15. Masateru Nagata, Jasper G Tallada and Taiichi Kobayashi. Bruise detection using NIR hyperspectral imaging for strawberry (*fragaria × ananassa* Duch.). *Environmental Control in Biology*, 44(2):133–142, 2006.
16. Giuseppe Palermo, Paolo Piraino and Hans-Dieter Zucht. Performance of PLS regression coefficients in selecting variables for each response of a multivariate pls for omics-type data. *Advances and applications in bioinformatics and chemistry: AABC*, 2:57, 2009.
17. Gerrit Polder, Gerie WAM van der Heijden and Ian T Young. Spectral image analysis for measuring ripeness of tomatoes. *Transactions of the ASAE*, 45(4):1155, 2002.
18. Clifford H Spiegelman, Michael J McShane, Marcel J Goetz, Massoud Motamedi, Qin Li Yue and Gerard L Coté. Theoretical justification of wavelength selection in PLS calibration: development of a new algorithm. *Analytical Chemistry*, 70(1):35–44, 1998.
19. Svante Wold, Michael Sjöström and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, 2001.
20. Svante Wold, Michael Sjöström and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, 2001. PLS Methods.

Bionotes

Florian Becker

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany
florian.becker@iosb.fraunhofer.de

Florian Becker studied cognitive and computer science at Eberhard Karls University of Tübingen and Karlsruhe Institute of Technology (KIT). In 2018, he joined the Vision and Fusion Laboratory at KIT. His research activities are in close cooperation with the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB. In his research, he focuses on machine learning for hyperspectral image analysis.

Andreas Backhaus

Fraunhofer Institute for Factory Operation and Automation IFF,
Magdeburg, Germany

Andreas.Backhaus@iff.fraunhofer.de

Dr. Andreas Backhaus is a senior research engineer and has been associated with the Fraunhofer Institute for Factory Operation and Automation IFF in Magdeburg (Germany) since 2009. He holds a diploma degree in Technical Computer Science from the University of Ilmenau (Germany) and a Ph.D. in Computational Neuroscience from the University of Birmingham (United Kingdom). He has specialized on the use of machine learning techniques for sensor data analysis and fusion with an expertise in the acquisition and processing of hyperspectral sensor data. He is currently using his expertise in application areas like smart farming, plant breeding or food quality control. He has been leading or participating a numerous research projects funded by industry as well as the public sector.

Felix Jhrden

Fraunhofer Institute for Factory Operation and Automation IFF,
Magdeburg, Germany

felix.jhrden@iff.fraunhofer.de

Felix Jhrden is a research engineer and has been associated with Fraunhofer Institute for Factory Operation and Automation IFF in Magdeburg (Germany) since 2018. He holds a bachelor and master degree in information technology with specialization in image processing and machine learning from the University of Magdeburg.

Merle Flitter

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany

merle.flitter@iosb.fraunhofer.de

Merle Flitter is a mechatronics and research engineer at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB in Karlsruhe, Germany. She holds a bachelor degree in engineering from the Technical University of Berlin and a masters degree in mechatronics from the Karlsruhe Institute of Technology.