Michael Kopf, Robin Gruna*, Thomas Längle, and Jürgen Beyerer

Evaluation and comparison of NIR multi-product calibration methods for Brix prediction

Evaluierung und Vergleich von NIR-Multiprodukt-Kalibrierungsverfahren zur Brix-Bestimmung

https://doi.org/10.1515/teme-2017-0069
Received July 3, 2017; revised January 23, 2018; accepted January 24, 2018

Abstract: Near-infrared (NIR) spectroscopy is a widespread technology for fruit and vegetable quality assessment. New fields of application of this technology, like mobile food analysis with handheld low-cost spectrometers, increase the demand for chemometric calibration models that are able to deal with multiple products and varieties thereof at once (so-called *multi-product calibration models*). While there are well studied methods for single-product calibration as *partial least squares regression* (PLSR), multi-product calibration is still challenging. Conventional approaches that work well for single-product calibration can lead to high errors for multi-product calibration. However, nonlinear methods as local regression and artificial neural networks were found to be suitable[1,2]. Preliminary studies in multi-product calibration for quantitative analysis of food with near-infrared spectroscopy showed good results for *memory-based learning* (MBL) and a *classification prediction hierarchy* (CPH)[3]. In this study, three varieties of apples, pears and tomatoes with known sugar content (in °Brix) are analysed with NIR hyperspectral imaging spectroscopy in the range from 900 nm to 2400 nm. Predictive performance of a linear PLSR model, two nonlinear models (CPH and MBL) and different preprocessing techniques are tested and evaluated. For error estimation, *leave-one-product-out* and *leave-one-out* cross-validation are used.

Keywords: NIR, chemometrics, foodstuff, multi-product calibration.

Zusammenfassung: Nahinfrarotspektroskopie ist eine etablierte Methode zur Qualitätsbestimmung von Obst und Gemüse. Neue Anwendungsgebiete, wie z. B. die mobile Lebensmittelanalyse mittels handgetragener und preisgünstiger Mikrospektrometer, verlangen nach neuen Ansätzen zur Multiprodukt-Kalibrierung. Zur produktspezifischen Kalibrierung existieren bereits geeignete Methoden wie *partial least squares regression* (PLSR). Der Versuch von Micklander et al.[1,2] zeigt jedoch auf, dass die Multiprodukt-Kalibrierung noch eine ungelöste Herausforderung darstellt. Nichtlineare Ansätze wie neuronale Netze und lokale Regression erzielten hier bessere Ergebnisse als konventionelle Methoden wie PLSR. Vorläufige Untersuchungen zur Multiprodukt-Kalibrierung zur quantitativen Analyse von Lebensmitteln mittels NIR Spektroskopie lieferten vielversprechende Ergebnisse durch *Memory-Based Learning* (MBL) und *Classification-Prediction-Hierarchy* (CPH)[3]. In dieser Arbeit werden drei Ansätze zur Multiprodukt-Kalibrierung untersucht. Hierzu werden drei unterschiedliche Apfelsorten, Birnen und Tomaten mit bekanntem Zuckergehalt (in °Brix) mittels bildgebender NIR Spektroskopie im Bereich von 900 nm bis 2400 nm analysiert. Die Genauigkeit eines linearen PLSR-Modells und zweier nichtlinearer Modelle (CPH und MBL) sowie unterschiedliche Vorverarbeitungsmethoden werden untersucht und evaluiert. Zur Bestimmung von Fehlermaßen dienen *Leave-One-Out*- und *Leave-One-Product-Out*-Kreuzvalidierungen.

Schlüsselwörter: NIR, Multiprodukt-Kalibrierung, Lebensmittelanalyse, Chemometrie.

1 E. Micklander, K. Kjeldahl, M. Egebo, and L. Norgaard. Multiproduct calibration models of near-infrared spectra of foods. *Journal of Near Infrared Spectroscopy*, 14:395–402, 2006.

2 L. R. Lopez, T. Behrens, K. Schmidt, A. Stevens, J. A. M. Dematte, and T. Scholten. The spectrum-based learner: A new local approach for modeling soil vis-NIR spectra of complex datasets. *Geoderma*, 195–196:268–279, 2013.

3 M. C. Kopf and R. Gruna. Examination of multiproduct calibration approaches for quantitative analysis of food with near infrared spectroscopy. Bachelor's thesis, Karlsruhe Institute of Technology KIT, 2016.

***Corresponding author: Robin Gruna,** Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Karlsruhe, Germany, e-mail: robin.gruna@iosb.fraunhofer.de
Michael Kopf, Thomas Längle, Jürgen Beyerer: Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Karlsruhe, Germany

1 Introduction

Near-infrared spectroscopy as well as hyperspectral imaging became widespread technologies for quality inspection and optical sorting issues. Due to their ability for non-destructive quantitative and qualitative analysis, they are increasingly being used in the agricultural and food sector [15]. By contrast with mid-infrared spectroscopy, quantitative information cannot be read directly from the spectrum. Mathematical methods and models, called *chemometric methods*, must be applied for extracting information concerning relevant quality parameters. For quantitative analysis in NIR spectroscopy *partial least squares regression* (PLSR) and *principal component regression* (PCR) are considered as gold standard methods [13]. With increasing interest in new areas of application, such as the handheld food scanners for consumers (e.g., SCiO or TellSpec [6, 12]), the requirements for chemometric methods changed. In contrast to laboratory tasks, there is less or no prior knowledge about the samples being scanned. Conventional methods work well for samples that are all of one product class and acquired under constant measurement conditions. Those models cannot or only hardly be transferred to other measurement conditions or other product classes. Preliminary studies in *memory-based learning* (MBL) and a *classification prediction hierarchy* (CPH) showed promising results for multi-product calibration [7]. In this work, those two nonlinear methods and a linear PLSR model were tested and evaluated. The task was to predict the °Brix value, correlating strongly with sugar content and therefore ripeness, to multiple agricultural products. In this study, three apple varieties, tomatoes and pears, which are the most popular fruit and vegetable in Germany [14], were utilized for experiments. For those, hyperspectral near-infrared images were captured and used for a chemometric analysis. Five pre-processing techniques and combinations of them were tested: first and second derivative, *standard normal variate transformation* (SNV), *multiplicative scatter correction* (MSC) and absorbance transformation. These methods aim to enhance spectral features and to compensate the effects of non-uniform light scattering and other physical interferences. A comprehensive introduction to these methods can be found in [11].

2 Related work

Since NIR spectroscopy has been used for decades in qualitative and quantitative analysis, there are several well-established chemometric methods for pre-processing spectral data and calculating single-product calibrations. Partial least squares regression and principal component regression are the two calibration procedures most frequently used for single-product calibration [3]. To model nonlinear correlations, e.g. caused by multiplicative effects arising from non-uniform particle sizes in the samples, *artificial neural networks* (ANN) show better capabilities [3, 4]. In the experiment of Micklander et al. [9], linear models reach unsatisfactory accuracy for multi-product calibration for foodstuff. There is a noticable improvement in prediction with ANNs and local regression methods. Berzaghi et al. [2] also improved multi-product calibration models by use of a local regression method called LOCAL that is similar to memory-based learning. The memory-based learning approach used in our work was originally designed for the chemometric analysis of soils. The algorithm outperformed partial least squares regression, support vector regression machines, locally weighted regression and LOCAL [8].

3 Material and methods

3.1 Data acquisition

Three varieties of apples (Jonagold, Gala and Elstar) were used to get product separation on different levels. For a separation on a higher product level, tomatoes and pears were added to the dataset. For hyperspectral image acquisition each fruit was cut into two halves, the ripest and the most unripe half. A near-infrared line-scanning hyperspectral imaging camera was used to obtain hyperspectral reflectance images in the range of 900 nm to 2400 nm (see

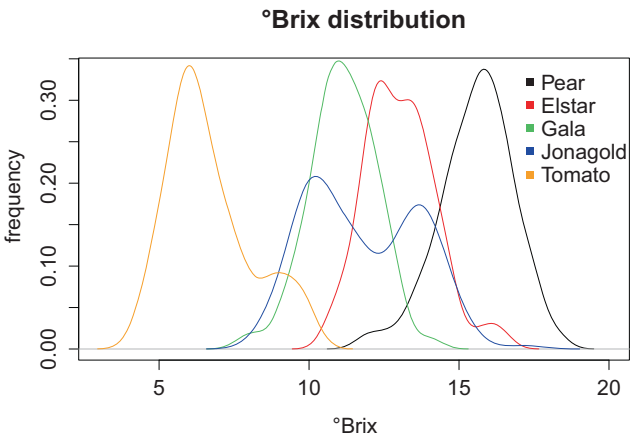

Figure 1: Distribution of °Brix values for the investigated product classes (Kernel density estimation with Gaussian kernel). The distributions have overlapping regions.

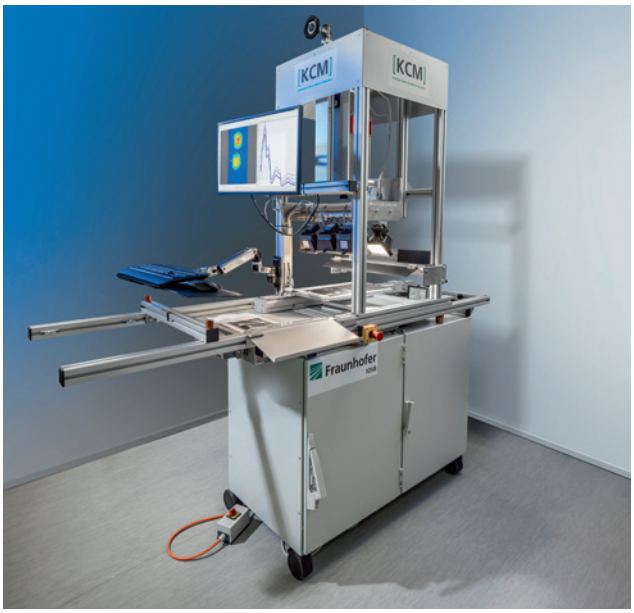

Figure 2: Near-infrared hyperspectral imaging system used in this study. Hyperspectral images (256 bands, spectral resolution FWHM ≈ 10 nm) in the wavelength region from 900 nm to 2400 nm were captured to analyse the sugar content of apples, tomatoes and pears.

Figure 2). Six halogen lamps in bright field constellation were used as a light source. Dark and white reference images were acquired at the beginning of each measurement and used for reflectance calculation. For the white reference image, a white teflon bar was used. A total amount of 124 pear, 454 apple (146 Elstar, 146 Gala, 162 Jonagold) and 90 tomato images were taken. After the hyperspectral image acquisition of each half, its middle third part was extracted and pressed to juice. The °Brix value for each half was determined with a bench-top refractometer. The distribution of the °Brix values overlap partially (see Figure 1).

3.2 Data processing

After defective pixel elimination, reflectance calculation and segmentation, the median spectrum from each segmented object was calculated. The median compensates for interfering influences (e.g., specular highlights or shadows) during image acquisition and provides a robust measurement of the object's spectral reflectance. The median spectra serve as the basis for further pre-processing and analysis. Pre-processing was used to remove scatter effects or to extract different features. The first two derivatives, SNV and MSC are used as well as absorbance transformation and combinations of those pre-processing techniques.

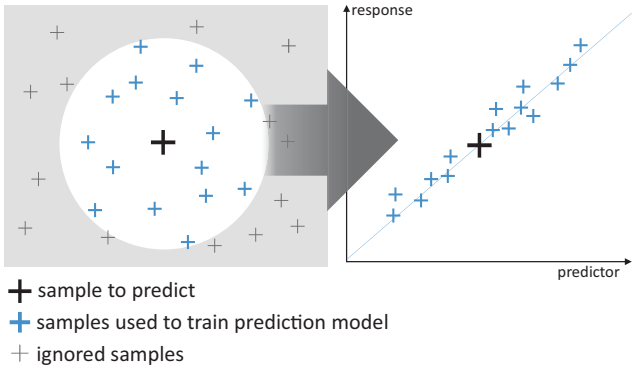

Figure 3: Memory-based learning: First step is to select nearest neighbour samples in the database, then a local regression model is trained with those selected samples.

3.2.1 Partial least squares regression

A conventional PLSR from the *R*-Package *PLS* [1] was used as linear calibration. The number of components was chosen via *leave-one-product-out* (LOPO) and *leave-one-out* (LOO) cross-validation. All samples of all classes were merged into one pooled dataset for training and validation.

3.2.2 Memory-based learning

In contrast to the so called *eager learning* like PLSR, where an explicit model is built, there is no offline training in memory-based learning (also referred to as *lazy learning*). To predict the response variable(s) for a sample, a distance metric is used to find nearest neighbours in the training data. A regression model is trained with those nearest neighbours on demand (see Figure 3). Parameters to set are among others the distance metric to find the nearest neighbours, the number of neighbours, the regression algorithm and the use of the dissimilarity matrix. The used *R*-package *resemble* [10] offers three kinds of distance metrics and four regression methods (see [10] for more details). It is also possible to use the distance information as additional predictor variables or as weights for weighted regressions.

3.2.3 Classification prediction hierarchy

Another approach for multi-product calibrations is classification prediction hierarchy (CPH). For each product class or subclass, a specialized regression model is trained. For prediction, the optimal model is chosen by a classifier.

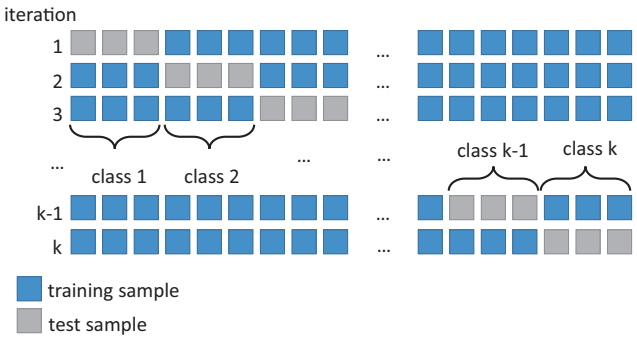

Figure 4: Leave-one-product-out cross-validation with k classes. For each iteration, all samples of one class are used for testing while all samples of other classes are used for training.

Then the regression model predicts the response variable. Due to the bad results when applying a specialized model to another class than it was trained on, the models are evaluated only with LOO cross-validation [7].

3.3 Evaluation method

To evaluate the calibration models, LOO and LOPO cross-validation were used. In contrast to LOO cross-validation, where single samples are left out while training, for LOPO cross-validation complete classes are left out. To estimate the ability of prediction on materials that are not in the training set, the model of an iteration is trained with all classes but one. The class left out is then used for validation (Figure 4). The *root mean square error* of cross-validation (RMSECV) was used as the error function:

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} (p_i - r_i)^2}{n}},$$

where p_i is the i-th predicted °Brix value, r_i is the i-th observed reference value, $i = 1, \ldots, n$ and n is the number of test samples.

4 Results

The combination of different pre-processing steps and model parameters results in a large number of possible calibration models. Due to space limitations, only the best calibration models of each method are presented in this section. SNV, MSC, first and second derivative and absorbance transformation followed by one of the other pre-processing steps were used for pre-processing. The number of components in PLSR is chosen to minimize the error of cross-validation. The CPH uses different pre-processing for each regression model and the classifier. Every regression model and the classifier were optimized separately. As MBL has many parameters, such as the distance metric, the regression method, the number of neighbours or the use of the dissimilarity matrix, the task to find the best MBL calibration is more effortful. For each regression type in MBL, all the other parameters and pre-processing methods were varied and evaluated with LOPO cross-validation.

4.1 Partial least squares regression

The best linear calibration model for a LOO cross-validation was trained with 16 components and absorbance transformation as pre-processing (Figure 5). The RMSECV is 0.78 °Brix. With a LOPO cross-validation the best PLSR model was obtained with absorbance transformation and MSC as pre-processing. The RMSECV is 1.13 °Brix with 6 components (see Figure 6).

4.2 Memory-based learning

Memory-based learning calibrations were tested with LOPO cross-validation to estimate the ability for predicting unknown materials. Only little improvement to the PLSR model is possible with a MBL calibration that uses

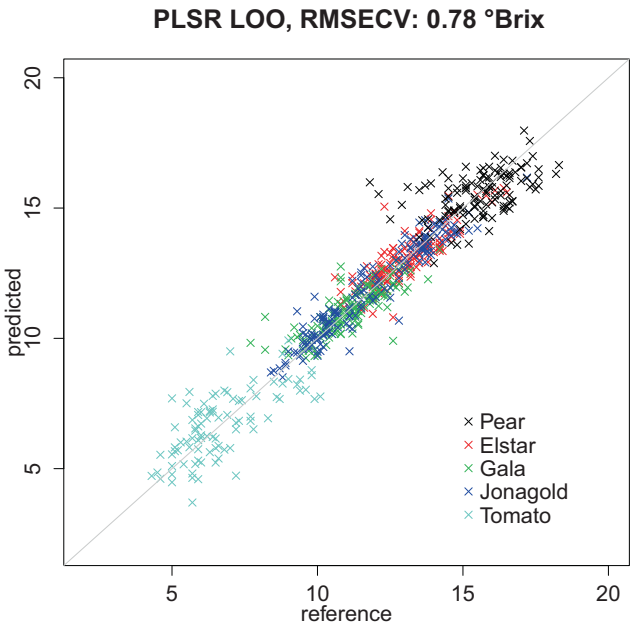

Figure 5: Best PLSR °Brix calibration model, validated with LOO cross-validation, absorbance transformation as pre-processing. The results are satisfying as all samples are well predicted.

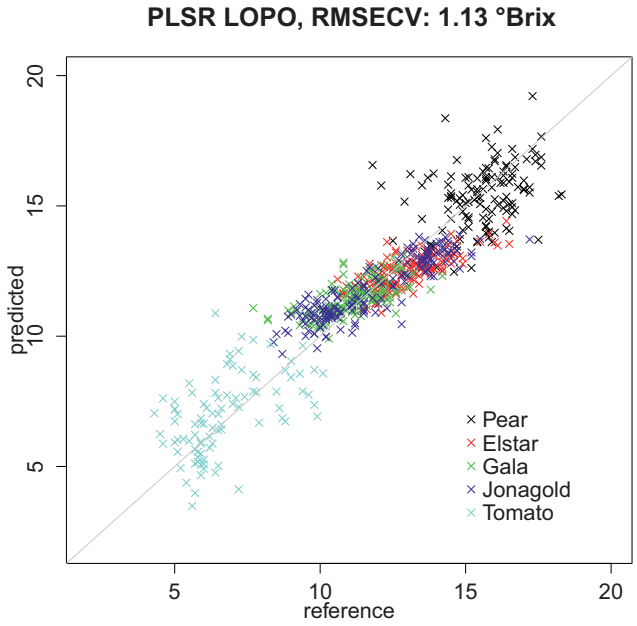

Figure 6: Best PLSR °Brix model, validated with LOPO cross-validation, absorbance transformation and MSC as pre-processing. Outer classes (pear and tomato) scatter, apple classes show a bias.

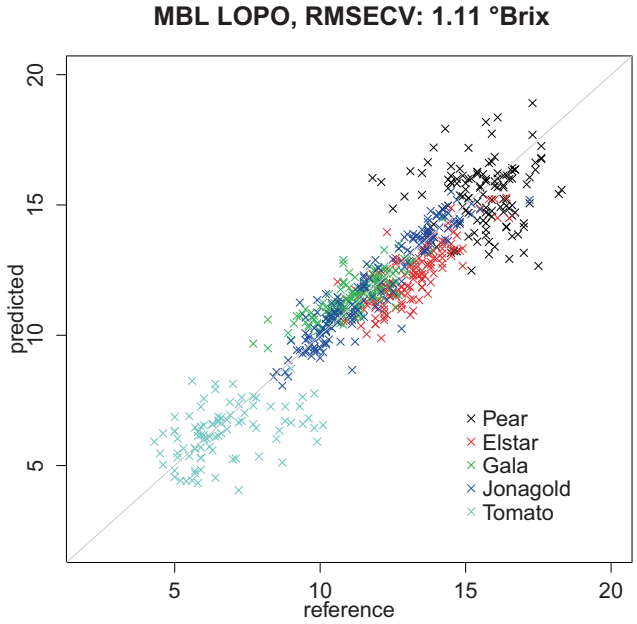

Figure 7: Best MBL °Brix calibration for LOPO with 300 nearest neighbours and absorbance transformation as pre-processing. Outer classes (pear and tomato) scatter, prediction shows no bias.

Euclidean distance for dissimilarity calculation, PLSR as regression algorithm with 8 components for prediction, dissimilarity matrix as additional predictor variables and 300 nearest neighbours. Samples were pre-processed by

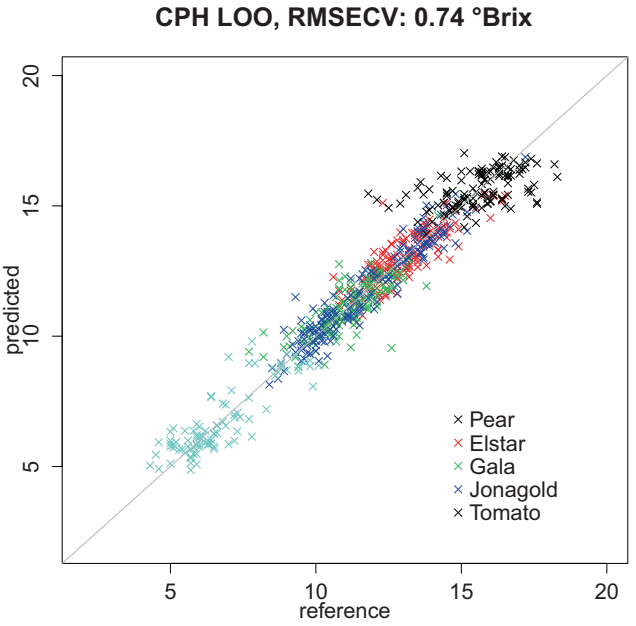

Figure 8: Best CPH °Brix model, validated with LOO cross-validation. Pre-Processing was adapted for each class. Results are satisfying as no class scatters and no bias is noticable.

an absorbance transformation. The root mean square error of cross-validation was 1.11 °Brix (see Figure 7).

4.3 Classification prediction hierarchy

In this study, linear discriminant analysis was used to classify and to choose prediction models. The best models were obtained with 6 components PLSR and MSC preprocessing for pears, 12 components PLSR and absorbance transformation for Elstar apples, 13 components PLSR and absorbance transformation for Gala apples, 17 components PLSR and SNV after absorbance transformation for Jonagold apples and 7 components and SNV after absorbance transformation for tomatoes. With those specialized models a RMSECV of 0.73 °Brix was reached in LOO cross-validation (see Figure 8).

5 Conclusion

This study showed that linear calibration models as PLSR can still obtain satisfying results for multi-product calibration. In contrast to the study of Micklander et al. [9], where PLSR calibrations showed weakness in prediction compared to nonlinear methods as local regressions and neuronal networks, a linear multi-product calibration for °Brix value on apples, pears and tomatoes shows compara-

ble accuracy to nonlinear models. MBL and CPH can only reach little improvement in predictive performance. A hierarchical calibration, such as the classification prediction hierarchy, can increase accuracy in °Brix prediction for apple varieties, pears and tomatoes by 6.4% from 0.78 °Brix to 0.73 °Brix.

Memory-based learning as a multi-product calibration approach reaches slightly better results for LOPO cross-validation than the pooled PLSR model. An RMSECV for the best PLSR model is 1.13 °Brix while MBL reaches 1.11 °Brix. It is an improvement by 1.8%. Especially memory-based learning has a high computational effort due to calculating big dissimilarity matrices for nearest neighbour search, which slows down prediction. As can be seen in Figure 6, the outer classes, pears and tomatoes, have a higher error and a higher deviation than the apple varieties in a LOPO cross-validation. When predicting for a LOPO cross-validation, the model has to extrapolate the response variable, which might cause the higher error, as conventional regression models are said not to be able to extrapolate without loss in accuracy [5]. Same effects are noticeable for MBL calibrations (Figure 7).

It is remarkable that PLSR performs well for those different products as a multi-product calibration method. Despite different products, the correlation between the sugar content and the NIR spectrum seems to be linear. For future work it might be interesting to find out if this linearity is applicable for more agricultural products or if nonlinear approaches outperform conventional methods in the long term. In addition, a more comprehensive study on the parameter settings of the investigated methods is planned.

References

1. R. Wehrens B. Mevik and K. H. Liland. PLS: Partial least squares and principal component regression. CRAN repository.
2. P. Berzaghi, J, Shenk, and M. O. Westerhaus. LOCAL prediction with near infrared multi-product databases. *Journal of Near Infrared Spectroscopy*, 8(1):1–9, 2000.
3. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, and J. Pages. Calibration in non-linear near infrared reflectance spectroscopy: a comparison of several methods. *Analytica Chimica Acta*, 384(2):207–214, 1999.
4. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, and J. Pages. NIR calibration in non-linear systems: different PLS approaches and artificial neural networks. *Chemometrics and Intelligent Laboratory Systems*, 50(1):75–82, 2000.
5. E. W. Ciurczak and D. A. Burns, editors. *Handbook of Near Infrared Analysis*. Practical Spectroscopy, 3 edition, 2007.
6. K. Grifantini. Knowing what you eat: Researchers are looking for ways to help people cope with food allergies. *IEEE Pulse*, 7, 2016.
7. M. C. Kopf and R. Gruna. Examination of multiproduct calibration approaches for quantitative analysis of food with near infrared spectroscopy. Bachelor's thesis, Karlsruhe Institute of Technology KIT, 2016.
8. L. R. Lopez, T. Behrens, K. Schmidt, A. Stevens, J. A. M. Dematte, and T. Scholten. The spectrum-based learner: A new local approach for modeling soil vis-NIR spectra of complex datasets. *Geoderma*, 195–196:268–279, 2013.
9. E. Micklander, K. Kjeldahl, M. Egebo, and L. Norgaard. Multiproduct calibration models of near-infrared spectra of foods. *Journal of Near Infrared Spectroscopy*, 14:395–402, 2006.
10. L. Ramirez-Lopez and A. Stevens. resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics. CRAN repository, 2016.
11. Åsmund Rinnan, Frans van den Berg, and Søren Balling Engelsen. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10):1201–1222, 2009.
12. H. Schulte, G. Brink, R. Gruna, R. Herzog, and H. Grueger. Utilization of spectral signatures of food for daily use. In *OCM 2015 – Optical Characterization of Materials – conference proceedings*, 2015.
13. H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise, editors. *Near-Infrared Spectroscopy: Principles, Instruments, Applications (Chemistry)*. Wiley-VCH Verlag GmbH & Co. KG, 1 edition, 2002.
14. S. Stracke, B. Homann, and S. Gröning. Branchenanalyse Obst-, Gemüse- und Kartoffelverarbeitende Industrie: Entwicklung von Markt, Beschäftigung und Arbeitsbedingungen. Study der Hans-Böckler-Stiftung 351, Hans-Böckler-Stiftung, Düsseldorf, 2017.
15. P. Williams, K. Norris, et al. *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, Inc., 1987.

Bionotes

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

Michael Kopf is a student at the Karlsruhe Institute of Technology KIT and a research assistant at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB in Karlsruhe. His research interests include computer vision and machine learning.

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

Robin Gruna is research group manager at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB in Karlsruhe. His research interests include hyperspectral imaging, machine vision, computational imaging and machine learning.

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

Thomas Längle is adjunct professor at the Karlsruhe Institute of Technology KIT, Karlsruhe, and the head of the business unit “Vision Based Inspection Systems” at the Fraunhofer IOSB in Karlsruhe, Germany. His research interests include different aspects of image processing and real-time algorithms for inspection systems.

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

Jürgen Beyerer is full professor for informatics at the Institute for Anthropomatics and Robotics at the Karlsruhe Institute of Technology KIT since March 2004 and director of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB in Ettlingen, Karlsruhe, Ilmenau and Lemgo. Research interests include automated visual inspection, signal and image processing, variable image acquisition and processing, active vision, metrology, information theory, fusion of data and information from heterogeneous sources, system theory, autonomous systems and automation.