Janina S. Wulf and Manuela Zude, Potsdam-Bornim, as well as Dieter Treutter and Susanne Rühmann, Freising

Measuring Uncertainties in Non-invasive Phenolic Compound Detection in Fruit Tissue

The quantitative analyses of health promoting fruit compounds are becoming more and more important. Fluorescence spectroscopy presents a potential possibility for detecting even small amounts of valuable fruit compounds, such as vitamins and polyphenols, in a fast and non-destructive way. Particularly for quantitatively analysing the substance building calibration models, based on the fruit fluorescence spectra, multivariate data processing is suited. Appropriate pre-processing methods are decisive in handling fluorescence quenching and reabsorption effects that can arise in the fruit tissue.

Dr. Janina S. Wulf and PD Dr. habil. Manuela Zude are scientists at the department „Technik im Gartenbau" of Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V., Max-Eyth-Allee 100, 14469 Potsdam; e-mail: *jwulf@atb-potsdam.de*
Dr. Susanne Rühmann and Prof Dr. Dieter Treutter are members of the staff of Technische Universität München, Fachgebiet Obstbau, Dürnast 2, 85350 Freising.

Keywords

Fluorescence, fruit pigments, data pre-processing, calibration

The consumer's interest and demand for health-promoting food products increases continuously, due to actual discussions about the relationship between nutrition and healthiness. Fruits and vegetables provide high content of vitamins, phenolic compounds, and other products of the secondary plant metabolism, whose contribution to a healthy nutrition was investigated in several clinical studies. These native compounds occur in marginal contents in the fruit tissue. Therefore, at the present state of the art, quantitative data can only be detected after chemical extraction using optical methods such as fluorescence spectroscopy. The content of health-promoting compounds can vary due to factors such as genome, the fruit development conditions on the tree and postharvest treatments. A non-destructive method would be a valuable tool in product quality assessment and the fluorescence spectroscopy provides a high potential, since some of the valuable compounds are strong fluorophors.

However, the apparent fluorescence signal is influenced by various effects. Besides, instrumental drift and measuring uncertainties, the anisotropic fruit material, scattering effects, and also fluorescence quenching and reabsorption effects affect the measurable fluorescence spectrum [1]. Suitable methods for multivariate calibrations were already developed for the near-infrared spectroscopy to predict a specific fruit quality attribute such as the soluble solids content. The high sensitivity of the fluorescence spectroscopy may add information on the nutritional interesting compounds [2]. The aim of the multivariate analysis is to correlate the independent variables (here: spectral intensities) with the dependent properties e.g. a fluorescent fruit compound content. With the help of the calibration model the content of this specific fruit compound should be predicted, based on the measured fruit spectra.

In the present study the fluorescence spectroscopy was used to non-destructively determine the content of fruit phenolic compounds in strawberries (*Fragaria x ananassa* 'Elsanta'). The content of p-cumaroyl glucose was used as an example. The aim was to test different data pre-processing methods for a robust calibration model to quantitatively determine the fruit phenolic compounds content with the help of the emitted fruit fluorescence signal. The content of the single phenolic compounds was evaluated using high-pressure liquid chromatography (HPLC). For the identification in the chemical analysis, the specific absorption spectra and retention times were compared with standards. The strawberry fluorescence emission spectra were excited with 337 nm and the fluorescence signal was non-destructively detected on the fruit with a laser fluorescence spectrometer (LF 401, IOM GmbH, Berlin) equipped with glass fibres in a wavelength range from 400 nm to 820 nm and a photomultiplier tube serving as detector.

Adaptation of suitable data pre-processing methods

The calibration models were calculated using partial least square (PLS) regression. Pre-processing methods such as smoothing, derivation and standardisation are basically used to improve the signal-to-noise-ratio in the fluorescence spectra, to correct for scattering effects, baseline shifts, and to normalize different valences of the variables (here:

Table 1: Contents of p-coumaryol-glucose and anthocyanins measured in three ripeness stages of strawberry fruits [mg/g dry weight]

Phenole / *phenols*	**unreif / *unripe***	**reif / *ripe***	**überreif / *overripe***
p-Cumaroyl-Glucose / *p-coumaroyl glucose*	0.211	0.212	0.346
Pelargonidin-glucosid / *pelargonidin glucoside*	0.183	5.582	7.699
Pelargonidin-malonyl-glucosid / *pelargonidin-malonyl glucoside*	0.203	0.574	0.855

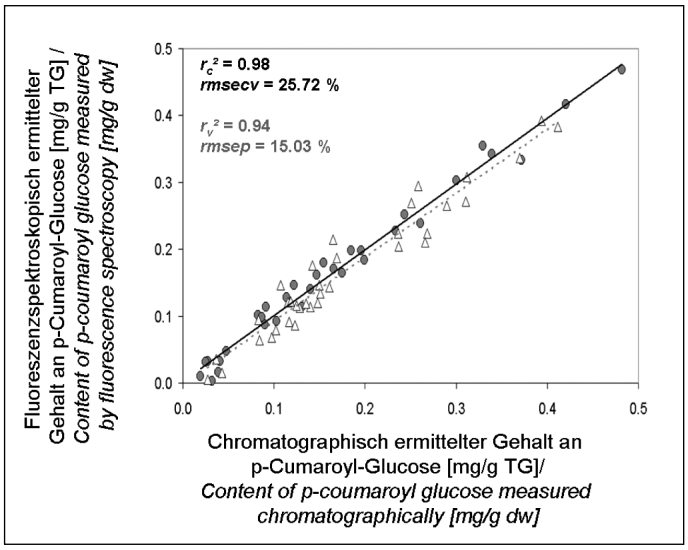

Fig. 1: Calibration model built on the spectral data of strawberry fluorescence and the fruit content of p-cumaroyl-glucose [mg/g dry weight] after pre-processing the fluorescence spectra by autoscaling and DOSC (o, gray) and validation on a independent dataset of fruit (Δ, white).

wavelengths). However, the information of the phenolic compounds content depends on the molecule-specific fluorescence emission and is, therefore, not included in every wavelength recorded. Resulting, calibration models built on the wrong variables will lead to low robustness of the calibration. Furthermore the fluorescence signal of the specific phenolic compound can be masked by the signals of other fluorophors and reabsorbed by fluorescent or non-fluorescent fruit compounds. The p-cumaroyl glucose belongs to the group of hydroxycinnamic acids, which can fluoresce in the blue-green wavelength range. In the same wavelength range different fruit pigments have strong absorption bands. One of the major pigment groups of strawberries are anthocyanins absorbing in the ultra-violet and visible wavelength range, while attenuating the apparent fluorescence signal of p-cumaroyl glucose. An approach was carried out to eliminate the variation in the fluorescence spectral data, which was not correlated to the chromatographically detected phenolic compound content by means of a mathematical data pre-processing method. The complexity of the data matrix was reduced by minimising the non-relevant variation. At the same time the variation in the fluorescence spectra, which was due to the differences in the fruit phenolic compound content, became more important.

The quality of a calibration model is defined by its coefficient of determination for the calibration data set (r^2_c), the number of latent variables (LV), and the root mean square error of calibration (rmsec). The optimal number of LV was determined with the help of the root mean square error of cross-validation (rmsecv) or the root mean square error of prediction (rmsep). The rmsep was calculated using an independent test-set that generally leads to a more robust calibration model. However, for calibration models of horticultural products the rmsecv is often used, because the measuring of several experimental series is cost and time consuming and the availability of fruits and vegetables, which are seasonal products, is limited.

The phenolic compounds content and the fruit fluorescence spectra were measured on strawberries (n = 32) of three different maturity stages. Fruits with green-white colour were defined as "unripe", "ripe" fruits were light red with white regions on the bottom, and "overripe" fruits had deep red fruit skin colour. An effect of the visual apparent fruit anthocyanins on the measured fluorescence intensities can therefore already be assumed.

Results

The chromatographically determined phenolic compounds content showed a lower content of p-cumaroyl glucose in the unripe fruits compared to the ripe and overripe strawberries, whose contents were not significantly different (*Table 1*). On the other hand, the content of anthocyanins increased during ripening (shown here for pelargonidin glucoside and pelargonidin-malonyl glucoside).

When calculating calibration models based on the strawberry fluorescence spectra and their content of p-cumaroyl glucose, the application of smoothing, derivation and standardisation on the fluorescence spectral data matrix led to high root mean square errors of cross validation and relatively small coefficients of determination. Resulting, these data pre-processing methods have only a limited potential to correct the spectra for varying inter and intra molecular interactions. Therefore, the direct orthogonal signal correction (DOSC) algorithm was tested as data pre-processing method to correct for fluorescence quenching and reabsorption effects, which can disturb the fluorescence signal detection of the p-cumaroyl glucose content. DOSC was used as filter for structural variations in the spectral data matrix to correct for sample material properties [3]. The non relevant information for the chemically analyzed phenolic compound content was calculated and the DOSC factors were subtracted from the spectral data matrix.

Using DOSC as pre-processing method before applying a linear PLS regression on the strawberry fluorescence spectra and their p-cumaroyl glucose content resulted in a less complex and more robust calibration model (r^2_c = 0.98, rmsecv = 25.72%) (*Fig. 1*). An independent data set of strawberry fruits (n=35) was used to validate the calibration model with the same data pre-processing methods leading to a low rmsep value (rmsep = 15.03 %) for the non-destructive prediction of the p-cumaroyl glucose content in strawberries.

In former studies, corrections for re-absorbance effects on the fluorescence spectra were carried out using simultaneously recorded reflectance spectra to correct for varying absorption properties of the fruit tissue. However, correction of the fluorescence spectra with the help of the reflectance intensities led to less robust calibration models, probably because not only reabsorption but also quenching effects disturbed the fluorescence signal detection.

Conclusions

The fluorescence spectroscopy has the potential to non-destructively detect the content of health-promoting compounds in fruits and vegetables. Therefore adequate data pre-processing methods have to be selected to handle the instrumental inaccuracies such as noise as well as fluorescence quenching and reabsorption effects in the complex sample tissue. In particular the DOSC algorithm is suitable to correct the fluorescence signal for the prediction of specific fruit compound contents by removing the non-relevant variation from the spectral data matrix.

Literature

[1] *Wulf, J. S., und M. Zude*: Fluorimetrische Pigmentanalyse im Obst – Einflussfaktoren auf das Fluoreszenzsignal. Landtechnik 63 (2008), H. 1, S. 24-25

[2] *Geladi, P.*: Some recent trends in the calibration literature. Chemometrics and Intelligent Laboratory Systems 60 (2002), p. 211-224

[3] *Westerhuis, J. A., S. De Jong and A. K. Smilde*: Direct Orthogonal Signal Correction. Chemometrics and Intelligent Laboratory Systems 56 (2001), p. 13-25