**Multivariate Calibration of Spectroscopic Sensors for Postharvest Quality Evaluation: A Review**

Wouter Saeys1\*, Nghia Nguyen Do Trong1, Robbe Van Beers1

1KU Leuven, Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001, Leuven, Belgium

* To whom correspondence should be sent. Email: wouter.saeys@kuleuven.be (W. Saeys)

# Abstract

Vibrational spectroscopy methods are widely investigated as fast and non-destructive alternatives for postharvest quality evaluation. As these methods measure spectral responses at a large number of wavebands correlated to the quality traits of interest, multivariate calibration equations have to be built to estimate the quality traits from the acquired spectra. This paper provides an overview of the most important multivariate data analysis techniques for exploring spectral data, detecting outliers and building calibration models for predicting the quality traits of interest. Both linear and non-linear calibration methods are discussed for quantitative (continuous) and qualitative (discrete) quality traits. For each of the presented methods the theory is explained, followed by illustration of an example case from the postharvest domain and a discussion of applications of this technique for postharvest quality evaluation based on spectral sensors. As spectral preprocessing, careful validation and calibration transfer are crucial aspects for successful implementation of spectral sensors for postharvest quality evaluation, special attention is given to these aspects. Finally, conclusions are drawn and recommendations are made with respect to the steps to take and points of attention for successful calibration.

**Keywords:** multivariate data analysis; spectra; classification; preprocessing; validation; transfer

# Introduction

The fast and non-destructive character of NIR spectroscopy makes it an appealing alternative for the traditional destructive and cumbersome postharvest quality evaluation methods. As a result, its potential has been investigated for a wide range of applications in horticulture ([Nicolaï et al., 2007a](#_ENREF_106); Nguyen-Do-Trong et al., 2018). Recently, other vibrational spectroscopy techniques such as FTIR spectroscopy ([Downey, 1998](#_ENREF_13); [Wilson and Tapp, 1999](#_ENREF_64); [Karoui et al., 2010](#_ENREF_29); [Rodriguez-Saona and Allendorf, 2011](#_ENREF_57); [Lu and Rasco, 2012](#_ENREF_38); [De Marchi et al., 2014](#_ENREF_10); [Cozzolino, 2015](#_ENREF_6)), fluorescence spectroscopy (Sádecká and Tóthová, 2007; [Karoui and Blecker, 2011](#_ENREF_28); [Gorbe and Calatayud, 2012](#_ENREF_19)), Raman spectroscopy ([Herrero, 2008a](#_ENREF_23), [b](#_ENREF_24); [Yang and Ying, 2011](#_ENREF_70); [Li and Church, 2014](#_ENREF_34); [Zheng and He, 2014](#_ENREF_72); [Dhakal et al., 2016](#_ENREF_11); [Xu et al., 2017](#_ENREF_69)) and hyperspectral imaging ([Gowen et al., 2007](#_ENREF_20); [Elmasry et al., 2012](#_ENREF_15); [Lorente et al., 2012](#_ENREF_37); [Dale et al., 2013](#_ENREF_8); [Wu and Sun, 2013a](#_ENREF_65), [b](#_ENREF_66); Xiong et al., 2014; [Pu et al., 2015](#_ENREF_55); [ElMasry and Nakauchi, 2016](#_ENREF_41); [Wang et al., 2016](#_ENREF_62)) are also gaining attention. As these are indirect methods which measure spectral responses correlated to the quality traits of interest (e.g. soluble solids content, dry matter content, acidity or firmness), calibration equations have to be built which estimate the quality traits as a function of the acquired spectral variables. This is typically done by estimating the parameters of a data-based model from a set of samples for which both the spectral variables and the quality traits have been measured, a so-called calibration or training set.

Naes et al. (2002) indicated that this poses several challenges. Due to the broad, overlapping absorption or emission peaks, there is typically no spectral variable which is only influenced by the quality trait of interest. To overcome this **non-selectivity problem** the relevant information contained by multiple variables has to be extracted and combined through the use of multivariate data analysis techniques. The broad, overlapping peaks also lead to a situation where many spectral variables relate to the same absorption peaks and to different absorption peaks for the same component (e.g. overtone and combination bands). As a result, the information contained in the different spectral variables may be highly similar. These high correlations between the different variables complicate the estimation of the multivariate calibration models. This is known as the **collinearity problem**. Each calibration model contains unknown parameters that need to be estimated using data acquired for a number of calibration samples. The quality of the resulting calibration can depend strongly on the number of calibration samples used and on how these were selected. To obtain a calibration model which will perform well on future samples (e.g. from a different orchard, cultivar or season), the calibration samples should span the natural range of variability as well as the possible combinations of influencing factors which can be expected in the future. This requires careful **calibration data selection**. While extreme samples from the target population can be highly informative in the model building phase, erroneous samples and samples from outside the target population (e.g. pear in a set of apples) may have a serious negative impact on the resulting calibration model. Therefore, **outliers** have to be detected and handled with care. The hierarchical structure of biological tissues with many interfaces between materials with different refractive indices (e.g. cytoplasm with cell organelles, cell walls, air pores,…) make biological produce highly turbid. As the light is **scatter**ed, the path travelled in the tissue will change, thus complicating the relation between the acquired spectra and the quality traits of interest. As most multivariate calibration methods assume that the relation between the independent and dependent variables is linear, this should be verified. If the relation turns out to be **non-linear**, it is recommended to preprocess the variables to linearize the relationship (e.g. logarithmic transform to linearize the exponential relation between absorbance and concentration according to Beer’s law) or to replace the linear calibration methods by non-linear calibration methods. These challenges make the calibration of spectroscopic sensors non-trivial, requiring great care to avoid over-optimistic results which cannot be reproduced for future samples. Therefore, the objective of this article is to give a comprehensive overview of the most important steps and methods in multivariate calibration of spectroscopic sensors for non-destructive postharvest quality evaluation with guidelines on their use and the points of attention. For each method, the theory will be described first, followed by an illustration of an example data set and discussion of their application in postharvest quality evaluation. Finally, conclusions will be drawn and recommendations will be made.

The example data set used in this study contains Vis-NIR spectra of Braeburn and Jonagold apples acquired in diffuse reflectance mode on a diode array spectrometer (Corona Vis-NIR fiber, Zeiss, Jena, Germany) in the 380-1690 nm range during the maturation period pre-harvest. Analogous to the data presented in Alamar et al. (2007), this involved a weekly measurement of apple samples before the commercial harvesting period and destructive measurement of the soluble solids content (SSC) measured in % with a digital refractometer (PR-101α, Atago, Tokyo, Japan).

# Data exploration

Modern spectroscopic sensors acquire optical signals (transmittance, reflectance, fluorescence emission, Raman scattering,…) at many different wavelengths for each sample presented to the instrument. The first step in any multivariate data analysis should be to visualize the spectral responses for different samples to detect clear differences (e.g. obvious outliers) through visual inspection. However, it is more difficult to distinguish samples with more subtle differences in their quality traits, especially when this quality trait is not the dominant source of spectral variation. In Figure 1, this is illustrated for the Vis-NIR spectra of Braeburn apples, where the spectra are colored according to their SSC measured in %, the main quality trait of interest in this case. The large number of variables and the large variation not related to the quality trait of interest complicate the interpretation of these spectra considerably. Moreover, there is a clear signature in the spectra which indicates that the responses at different variables are strongly correlated. Therefore, it is desirable to have a lower dimensional representation of these data which captures the largest part of the variation contained in these spectral variables with a small number of uncorrelated variables. This can be obtained by means of principal component analysis (PCA).



Figure 1 Visualisation of a PCA model built on the Vis-NIR spectra of Braeburn apples from two harvest locations (Merdorp and Rotselaar): a) Vis-NIR spectra of apples harvested in Merdorp colored according to the SSC content (%); b) Biplot visualizing the samples from Merdorp (red diamonds) and Rotselaar (green squares), and the variables (blue triangles) on the PC1-PC2 score plot. The dashed black ellipse indicates the 95% confidence limit; c) Q residuals vs. Hotelling T2 plot with indication of the 95% confidence limits (dashed lines) and the two harvest locations: Merdorp (red diamonds) and Rotselaar (green squares).The colored areas indicate zones with candidate outliers, while the black circle indicates a strongly outlying sample. d) Loadings plot which visualizes how the different variables are combined to calculate the scores for the first 5 PCs.

## 2.1. Theory of PCA

PCA starts from an (*n* x *p*) matrix **X,** with *n* rows corresponding to the samples for which the spectral responses at *p* possibly correlated spectral variables have been acquired. The aim is to replace the coordinate system with *p* correlated variables by a new coordinate system with smaller number *a* of uncorrelated variables which are linear combinations of the original variables. These new, uncorrelated variables are referred to as principal components (PCs). To this end, the matrix **X** is decomposed as a sum of bilinear components or vector products, according to the following model:

**X**= + E (1)

Where the (*n* x *a*) scores matrix **T** contains the coordinates for the *n* samples in the new coordinate system with *a* uncorrelated variables, while the (*p* x *a*) loadings matrix **P** contains thecontributions of the *p* original variables to the *a* new variables in its columns; the superscript T indicates the transpose operator. Mathematically, PCA performs an eigenvector decomposition of the cross-product matrix **X**T**X** to find an estimate for the loadings matrix **P**. The estimate for the scores matrix is then found by regressing **X** onto .The principal components are ranked in order of decreasing amount of variance captured. The first *a* principal components capturing the largest part of the variation in the **X** matrix are assumed to contain the relevant information (signal), while the remaining part of the variation is considered noise. By only retaining the principal components which capture a substantial fraction of the variation in the data, the noise is transferred to the residual error matrix **E**.

## 2.2 Effect of centering and scaling

When the **X** matrix is centered by subtracting the mean value for every variable to obtain the centered matrix **Xc**, the cross-product matrix **XcTXc** becomes identical to *n-1* times the empirical covariance matrix (Naes et al., 2002). Subtracting the mean value for each variable from all the observed values for that variable places the origin of the new coordinate system in the center of the data cloud. This simplifies the interpretation of the scores and allows for drawing inferences on confidence intervals for the principal component (PC) scores. Therefore, it is highly recommended to mean center the spectral data prior to performing PCA and other multivariate data analysis methods.

It should also be noted that PCA is sensitive to the scale of the original variables. Variables with larger variance will have a higher impact on the PCA model than variables with smaller variance. Therefore, it is important to consider whether this larger variance is indeed informative or rather the result of measurements in different units (e.g. mm vs. m). If the variables have been measured in different units, it is recommended to bring these on a similar scale to give them an equal opportunity to contribute to the model. This can be done by dividing the mean-centered values for each variable by the standard deviation calculated over all samples in the training set. This process is known as variable standardization or autoscaling. A drawback of this scaling procedure is, however, that variables with a low signal to noise ratio will contribute the same amount to the total variance as variables with a high signal to noise ratio. Therefore, autoscaling is not recommended for spectral variables which have been measured with the same instrument and are expressed in the same units.

## 2.3 The biplot

As PCA captures the largest part of the variation in the data in the first few PCs, inspection of the sample scores for these first few PCs can already provide a good view of the grouping in the data and potential outliers. To this end, PC scores plots can be inspected where the scores of the samples for one PC are plotted against those for another PC (e.g. PC1 vs. PC2). When some structure (e.g. grouping, clear trend or outlier) is observed in the data, it would be interesting to know which original variables are contributing to this principal component. To this end, the contribution of the original variables to these principal components, the loadings, can also be included on the same plot. This type of plot is known as a biplot and is illustrated in Figure 1b for the Vis-NIR spectra of Braeburn apples harvested in one orchard (Merdorp) illustrated in Figure 1a (red diamonds) together with those from another orchard (Rotselaar). The black dashed ellipse indicates the 95% confidence limit. As 5% of the samples would be expected to fall outside these confidence limits, the few points just outside the dashed ellipse are most likely not outliers. This will be investigated further in section 2.4.

Some of the variables are marked, showing their corresponding wavelengths. It can be noticed that the variables around 550 nm, related to the absorption peak by anthocyanins causing the red color in Braeburn apples, are contributing most to PC2. A more clear view on the contribution of the different variables to the PCs scores will be provided by the PC loadings, which will be discussed in section 2.5.

## 2.4 Outlier detection

Thanks to the dimension reduction provided by PCA, a considerably smaller number of variables has to be inspected to identify candidate outliers. Moreover, under the assumption of multivariate normality, confidence intervals can be defined for the individual PC scores and confidence ellipses can be drawn on the scores plots thanks to the orthogonality between the PC scores. This can be very useful to identify extreme outliers, but may be less useful for samples which are outlying for one PC, but not for another. Therefore, it is more appropriate to evaluate the outliers in a sample set based on a metric which combines the different PC score values to calculate the Mahalanobis distance from the origin within the model space. This metric is known as the Hotelling *T*2 distance and can be calculated as follows for every individual sample with centered spectral vector **xi**:

(2)

Where *T*2 is the Mahalanobis distance from the origin of the PCA model calculated based on the first *a* PCs; *t*j is thescore value and *λ*j is the eigenvalue both for the *j*th principal component (Naes et al., 2002).

While the Hotelling T2 promotes the detection of samples which are extreme within the model space (green region in Figure 1b), samples which are poorly described by the model would not be identified in this way if they are inlying after projection onto the model. Therefore, it is recommended to also inspect the distance in the original variable space from the sample to its projection onto the model hyperplane. This metric is known as the Q residual and is calculated as follows:

(3)

Where is the PCA estimate for the *i*th sample with centered spectral vector **x***i* which contains the values measured at *k* spectral variables (Naes et al., 2002).

In Figure 1c, the Q residuals vs. Hotelling T2 values are plotted against each other for the Vis-NIR spectra of the Braeburn apples harvested in Merdorp illustrated in Figure 1a together with the samples harvested in Rotselaar. The horizontal and vertical dashed lines indicate the 95% confidence limits for both metrics. A sample can be identified as an outlier within the model (Hotelling T2 value above the confidence limit, green zone in Figure 1c) or with respect to the model (Q residual above the confidence limit, blue zone in Figure 1c). Samples showing both an extreme Hotelling T2 and Q residual value (red zone in Figure 1c) are flagged as candidate outliers by both metrics. For example, the most probable outlier in this dataset is marked with a black circle in Figure 1c. The reason for flagging it as an outlier can be investigated by inspecting which variables contribute the most to this extreme Hotelling T2 or Q residual value. If the measurement is obviously wrong, the error should be corrected or the measurement should be repeated. However, it is also possible that the measurement is correct, but that this sample belongs to a sub-population which is underrepresented in the data set. In this case, it may be recommended to measure additional samples from this sub-population. It should be noted that in theory 5% of the samples will fall outside the 95% confidence limits. So, samples just outside but close to the limit may still be acceptable and could make the model more robust when it is confronted with similar samples. On the other hand, it should also be noted that the PCA model and these metrics are influenced by the outliers included in the calibration set, such that outliers might not be detected. To avoid these problems the use of robust PCA algorithms such as RobPCA (Hubert et al. 2005) is recommended.

## 2.5 PC loadings

As indicated in section 2.1, the loadings matrix **P** contains thecontributions of the *p* original variables to the *a* new variables in its columns. In Figure 1d, this is visualized for the PCA model with 5 PCs built on the Vis-NIR spectra of Braeburn apples illustrated in Figure 1a. PC1 has positive contributions for most wavelengths, which suggests that it mostly captures the variation in the reflectance spectra due to light scattering effects. The PC2 and PC3 loadings show a clear positive peak around 550 nm which can be related to the absorption peak by anthocyanins causing the red skin color in Braeburn apples. The PC2 loadings also show a negative peak around 680 nm which can be related to the absorption peak of chlorophyll, while the PC4 loadings have a positive peak here. The PC3 and 4 loadings also show relative positive peaks (local increase) at the water absorption bands around 980 nm, 1200 nm and 1450 nm, while PC2 shows negative relative peaks (local decrease) at these bands. It should be noted that while these PC loadings clearly contain information on the spectrally active components in apple tissue, they cannot be interpreted as pure components. Therefore, alternative decomposition methods such as multivariate curve resolution (MCR) have been proposed which relax the requirement for orthogonality of the components and replace it by spectroscopic constraints such as non-negativity of contributions and absorption peaks (de Juan & Tauler, 2006).

## 2.6 Applications of PCA on spectroscopic data for postharvest quality evaluation

In postharvest research, PCA has been widely used for data exploration in an initial phase or as a data reduction technique for classification purposes. Some researchers also used the PC loadings for identifying important variables contributing to the grouping. [Xing et al. (2007](#_ENREF_67)) applied PCA to hyperspectral reflectance images of apples measured in the 400-1000 nm range to construct the PC scores images. While PC1 score images accounted for the effect of variation in the reflected intensity from the curved apple surfaces, the combination of PC2 and PC3 scores images with image processing made it possible to detect the bruised apples with an accuracy of 86.36%. Based on the PC loadings plots they identified four wavebands as important for bruise detection in ‘Golden Delicious’ apples. [López-Maestresalas et al. (2016](#_ENREF_36)) showed the possibility to detect bruises in potatoes before they became visible using the PC1-PC4 and PC1-PC6 scores plots obtained from a PCA on hyperspectral images acquired respectively in the 400-1000 nm and 1000-2400 nm range. [Eisenstecken et al. (2019](#_ENREF_14)) built Quadratic Discriminant Analysis (QDA) models using the PC scores obtained from applying PCA to NIR diffuse reflectance spectra of apples in the 1000-2500 nm range to classify 3 different orchard elevation levels and 9 cultivars. The PCA-QDA models provided correct classification rates of 87.5% for orchard elevation and 86.3% for cultivar using independent validation sets. [Jacobs et al. (2016](#_ENREF_27)) used PCA to explore the spectral differences between spectra measured on the adaxial and abaxial side of lamb’s lettuce leaves. [Szymanska-Chargot et al. (2015](#_ENREF_59)) applied PCA to FT-IR spectra in the 1500-800 cm-1 range acquired for two apple cultivars during development. They found that PC1 was dominated by the development stages which could be associated with changes in Galacturonic acid and cellulose contents based on the spectral signature of the PC1 loadings, while the two cultivars grouped along the PC2 score axis.

While PCA is a highly valuable technique for data exploration, it does not provide predictions for the quality traits of interest. To this end, we should make a link between spectral data and the quality traits of interest. This will be the topic of the following section.

# Multivariate regression methods

The link between the acquired spectra and the quality traits of interest is typically modelled in a direct, statistical way. From a statistical point of view, the problem of predicting the quality traits of interest from the acquired spectra is defined as the search for a mathematical relation (function) between the predictor or independent variables X and the predicted or dependent variables Y:

(4)

The parameters of this function are then estimated from a set of calibration or training samples for which spectra have been acquired and the quality traits of interest have been measured with reference methods (e.g. refractometer measurements for SSC, Magness Taylor compression test for firmness…).

3.1 Linear regression methods

### 3.1.1. Multiple linear Regression (MLR)

To simplify the estimation process, a linear function is typically assumed for every predicted variable Y. By mean-centering both the predictor matrix **X** and the predicted vector **y** we couple the mean concentration to the mean spectrum and avoid that we have to estimate an intercept:

(5)

where **yc** is the (*n* x 1) vector with the mean-centered reference values for the considered quality trait for the *n* samples in the calibration set; **x**c1, **x**c2, …, **x**cp are the (*n* x 1) vectors with the mean-centered spectral readings acquired at the different wavelength variables for the *n* samples, which form the columns in the (*n* x *p*) matrix **Xc**; 1, 2, … , *p* are the regression coefficients, which form the elements of the (*p* x 1) column vector ; is the (*n* x 1) vector with the residuals. The set of regression coefficients **b** which minimizes the sum of squared errors (SSE) on the training set can be calculated as:

**b** (6)

Where the superscript -1 indicates that the inverse of the square cross-product matrix **XcTXc** is calculated, while the superscript + indicates that the pseudoinverse of the matrix **Xc** is calculated. For a new sample, the quality trait can be predicted as follows based on its mean-centered spectrum **xc,new**:

(7)

This method to estimate the linear relation between multiple predictor variables X and a predicted variable Y is known as **Multiple Linear Regression** (MLR).

While it is a very intuitive and straightforward method providing the best linear unbiased estimator for the calibration set, it is not the preferred choice for the calibration of spectral sensors. In the case of spectral sensors, the number of predictor variables *p* may be larger than the number of samples *n* in the calibration set. In this case, the mean-centered spectral matrix **Xc** has more columns than rows, which makes the covariance matrix **Xc**T**Xc** singular and non-invertible. This problem is known as **exact multicollinearity** and prevents the least squares estimation from finding a unique solution. This problem may be overcome by increasing the number of samples *n* in the calibration set or selecting a subset of *k* variables such that *k < n*. While this would result in a mathematically unique solution, the prediction performance may still be very poor due to the high correlation between the different spectral variables. These high correlations make it possible to approximate a variable by a linear combination of the other variables. As a consequence, very similar prediction performance can be obtained by reducing the regression coefficient for this variable and increasing the regression coefficients for the other variables. This phenomenon is known as **near multicollinearity** and results in large variation in the estimates for the regression coefficients obtained from different subsets of the training set, because the covariance matrix **Xc**T**Xc** becomesnearly singular and poorly invertible. As the estimated set of regression coefficients becomes very specific for the used calibration set, the model will generalize poorly to new data. We say that the model is ‘over-fitting’ the calibration data.

Three different approaches have been proposed to overcome this problem. The near multicollinearity can be avoided by carefully selecting a small number of variables *k* compared to the number of samples in the calibration set *n.* Alternatively, an extra term can be added to the sum of squared errors which is minimized in the estimation of the regression coefficients by ordinary least squares to regularize the estimation towards a certain target. For example, by adding a term proportional to **b**T**b** to the sum of squared errors,large regression coefficient values are penalized. This method is known as ridge regression. A third option is to replace the original, (highly) correlated spectral variables by a smaller number of uncorrelated variables which have been defined as linear combinations of the original variables. As this approach is by far the most commonly used, it will be discussed in more detail in the following sections.

### 3.1.2 Principal Component Regression (PCR)

As discussed in section 2.1, PCA captures the largest part of the variation present in a set of *p* correlated variables with a smaller number *a* of uncorrelated (orthogonal) linear combinations of the original variables (principal components). This provides an interesting way to avoid the multicollinearity problems in the multiple linear regression by replacing the centered matrix **Xc** with the *p* original, correlated spectral variables in Eqn.(6) by the matrix **T** from Eqn.(1) with the scores for the first *a* principal components:

**c** (8)

Where **c** is the (*a* x 1) vector of regression coefficients in the PC score space. For a new sample, the quality trait can then be predicted by multiplying the (1 x *p*) vector **xc,new** for the mean-centered, acquired spectrum by the (*p* x *a*) transpose of the loadings matrix **P**T from Eqn.(1) to obtain the (1 x *a*) scores vector **t**new and multiplying this with **c**:

(9)

Where is the (*p* x 1) regression vector in the original variable space obtained from PCR. Thanks to the use of the principal components in the estimation, the regression coefficients obtained with PCR typically have smaller standard errors and provide better prediction performance for new data than those obtained with MLR. To maximize the prediction performance on new data, a trade-off has to be made between including enough PCs to capture the relation between the spectral variables and the quality trait of interest (signal), and including principal components which are dominated by the noise in the calibration data (noise). To this end, the number of PCs is typically selected based on an evaluation of the prediction performance of models with different numbers of PCs on a validation set which has not been used for training the model. Some guidelines with respect to this validation will be given in section 6.

It should also be noted that thanks to the calculation of a PCA model prior to the regression, all plots and tools available in PCA (e.g. scores plots, loadings plots, biplot, Q-residual vs.  Hotelling T2) are also available in PCR and can be used for inspecting the data.

### 3.1.3. Partial Least Squares Regression (PLSR)

While the estimation of the regression coefficients based on the PCs in PCR provides a large improvement with respect to MLR, there is no guarantee that PCs capturing a large part of the variance in **X** are also important for predicting **y**. If the quality trait of interest has a large impact on the acquired spectra, it will most likely be captured in the first few PCs. However, if the spectral variation is dominated by other sources, the first few PCs may not be the most informative in predicting this quality trait. Therefore, it could be more efficient if the new variables would be defined based on the relation between X and Y. Partial Least Squares Regression (PLSR) has been proposed to this end. This method defines the new variables as orthogonal (uncorrelated) linear combinations of the original variables which maximally capture the covariance between X and Y (Martens & Naes, 1989; Wold, Sjöström, & Eriksson, 2001; Mevik & Wehrens, 2007). As the new variables are strictly speaking no longer principal components, they are typically referred to as latent variables (LVs) or PLS components. Mathematically, this is obtained by performing a singular value decomposition of the cross-product matrix **Xc**T**Yc** instead of **Xc**T**Xc** in PCA. This results in a bilinear decomposition of both the centered matrices **Xc** and **Yc** in scores and loadings matrices analogous to the decomposition obtained in PCA. Different algorithms have been proposed to obtain this decomposition such as the iterative NIPALS algorithm (Martens & Naes, 1989) and the optimization based SIMPLS algorithm (De Jong, 1993). In the case of the NIPALS algorithm, the following decomposition is obtained:

(10)

(11)

Where **T**PLSR, **P**PLSR and **E**PLSR,X respectively refer to the scores, loadings and residual matrices obtained from the bilinear decomposition of the **X** matrix by PLSR; while **Q**PLSR and **E**PLSR,Y are the Y loadings and residual matrices from the bilinear decomposition of the **Y** matrix by PLSR. It should be noted that in Eqn.(11) **Y** is defined as a matrix, because PLSR can handle multiple Y variables at the same time and define one set of latent variables which capture maximum covariance between the matrix of X variables and the matrix of Y variables. This is especially interesting when the Y variables are highly correlated but noisy, because the joint estimation could provide more stable estimates for the latent variables. This simultaneous estimation of the LVs for multiple Y variables is often referred to as PLS2. When the correlation between the Y variables is rather low, it is recommended to build separate PLS models for each of them (PLS1), because this typically gives better prediction performance (Naes et al., 2002). It should also be noted that in Eqn.(11) the scores from the X loadings **T**PLSR are multiplied with the Y loadings **.** This Y loadings matrix has some similarity with the vector **c** of regression coefficients in the score spacein Eqn.(8). It can also be used to derive a matrix of regression coefficients which can directly be applied to the centered spectral vector acquired for a new sample **x**c,new to estimate the quality trait(s) of interest from it similar to Eqn.(9):

(12)

Where **R**PLSR is a weighting matrix which is defined based on the X loadings matrix **P**PLSR. For more details on the algorithmic aspects of PLSR, the reader is referred to Martes & Naes (1989), De Jong (1993) and Mevik & Wehrens (2007).

Similar to the case of PCR, the number of latent variables to retain in a PLSR model to find the optimal trade-off between underfitting and overfitting is typically determined based on the prediction performance on a validation set which has not been used when building the model. It has been shown that in most situations PCR and PLSR can obtain similar prediction performance. However, as the latent variables have been defined to capture maximal covariance between X and Y rather than maximal variance in X, it is expected that PLSR will need fewer latent variables than PCR needs principal components to reach the same prediction accuracy. One could also expect that for the same number of components, PLSR will capture more variance in the quality traits, while PCR will capture more spectral variance (Mevik & Wehrens, 2007).

### 3.1.4. Quantification of model performance

In order to decide on the model complexity, it is important to quantify the model performance. As the linear regression methods aim at minimizing the sum of squared errors for predicting the Y variable based on the X variables, a performance metric related to this prediction sum of squares gives a good idea of the model performance. To obtain a performance metric in the same units as the quality trait of interest, the root mean squared error (RMSE) is typically used:

(13)

Where *n* is the number of samples, and and *yi* are respectively the estimated and measured Y value for sample *i* with spectral vector **xi***.* As will be explained in section 6, this can be calculated based on the calibration set (RMSEC), the validation set (RMSEV), a cross-validation (RMSECV) or a separate test or prediction set (RMSEP). The RMSE gives us a good idea on the absolute error, but it does not tell us whether the error is systematic or random. The systematic or average error is known as Bias and can be calculated as the average difference between the predicted values and the reference values:

(14)

The random or standard error on top of this bias is known as the standard error of prediction (SEP):

(15)

It should be noted that we divide by *n-1,* because one degree of freedom has been consumed for calculating the Bias. Due to the close relation between RMSE, Bias and SEP, it is considered sufficient to report two of the three.

As the absolute error depends on the units of the Y variable and the variation contained in the dataset, it can also be interesting to place the prediction error in perspective of the total variation in the Y data. The ratio of standard error of performance (RMSE) to standard deviation (STD), also referred to as ratio of prediction to deviation (RPD), quantifies how many times the model is more accurate in estimating the Y variable than the average value :

(16)

With SSTO the total sum of squares and SSE the error sum of squares. The RPD can be related to the coefficient of determination *R2,* which expresses which fraction of the variation in the Y data is captured by the model:

(17)

For a large number of samples *n* the ratio of *n* over *n-1* can be approximated as 1.

### 3.1.5. Application of MLR, PCR and PLSR to the apple example

To illustrate the performance of the presented multivariate calibration models, these models were tested on the apple dataset for predicting SSC from the Vis-NIR spectra. The Braeburn data presented in Figure 1a were augmented with data for a second cultivar (Jonagold) harvested at the same two locations. The Kennard-Stone algorithm was used to divide the dataset into a calibration set (on which cross-validation is performed) and a separate validation set. To allow a good comparison between the different modelling techniques, only a mean centering was applied in the preprocessing step. As a cross-validation strategy, venetian blinds with 10 data splits was performed. This involves 10 iterations of calibration and validation. Each time, 1 out of 10 samples is set aside in a validation set (e.g. 1st, 11th, 21st,…) and the other samples are used to build calibration models with different model complexities. These calibration models are then tested on the samples which have been set aside in the validation set. For more information on the concept of cross-validation the reader is referred to Section 6.2. Table 1 summarizes the results for the different multivariate calibration models.

Clear differences in the results obtained with the different modelling techniques can be observed in Table 1. MLR on the full spectrum displays an extremely low RMSEC value, but much higher RMSECV and RMSEP values, which indicate that this method was not able to obtain reliable estimates for the large number of regression coefficients. We can conclude that the model has overfitted the calibration data. To overcome this problem, PCR and PLSR can be used. From Table 1, it becomes clear that these models show less ‘over-fitting’, with RMSECV and RMSEP values close to the RMSEC. PLSR needs fewer latent variables (10 LVs) in comparison to PCR (14 PCs) to reach a similar prediction performance. Figure 2 shows a comparison between PCR and PLSR. Figure 2a shows the increase in the captured variance for both modelling techniques, with an overall higher Y variance captured by the PLSR model.

Table 1 Performance metrics of different multivariate calibration methods for predicting SSC values (%) from the Vis-NIR spectra of Jonagold and Braeburn apples. A separate test set was used to validate the models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | RMSEC (%) | RMSECV  (%) | RMSEP  (%) | Pred. bias  (%) | Extra |
|  | (n = 408) | (n = 408) | (n = 212) | (n = 212) |  |
| MLR | 1.14e-11 | 1.27 | 3.31 | -0.60 | All WL |
| MLR | 0.64 | 0.65 | 0.62 | 0.08 | 5 WL |
| PCR | 0.61 | 0.65 | 0.60 | 0.11 | # PCs = 14 |
| PLSR | 0.56 | 0.60 | 0.55 | 0.11 | # LVs = 10 |
| SVM-R | 0.51 | 0.63 | 0.54 | 0.02 | / |
| LWR | 0.68 | 0.78 | 0.69 | 0.09 | # PCs = 4 |



Figure 2 (a) the cumulative X/Y variance captured for both the PCR and PLSR models; (b) Regression coefficients for both the PCR and PLSR models

On the other hand, Figure 2b gives an overview of the regression coefficients, indicating which variables (wavelengths) have a high contribution to the SSC prediction. The absorption regions in the spectrum around 680 nm for chlorophyll and around 980, 1200 nm and 1450 nm for water clearly contribute to the models for SSC prediction. This suggests that the model exploits the correlation of SSC with moisture content and chlorophyll content. The former may be attributed to the water replacement effect, where a higher SSC corresponds to a lower volume concentration of water. The latter may be attributed to variation in the maturity stages of the apples , because during maturation chlorophyll is broken down, while the SSC increases. It should be noted that the latter correlation may not be stable over different crop years or varieties. Therefore, the robustness of the SSC prediction might be improved by removing the visible part of the spectrum.

If we build an MLR model with the variables corresponding to 5 dominant peaks in the plot, we obtain predictions which are only slightly worse than those obtained with PCR and PLSR on the full spectrum (RMSEP = 0.62% vs. 0.60% and 0.55%). This could be interesting when one wants to develop a low cost sensor measuring only 5 wavebands.

### 3.1.6. Applications of MLR, PCR and PLSR in postharvest quality evaluation

[Park et al. (2003](#_ENREF_52)) applied PCR to NIR spectra (800-1100 nm) of Gala and Red Delicious apples to predict SSC and firmness. The prediction performances for SSC were good for both cultivars (*R2* = 0.934, *SEP* = 0.28 % for 10 PCs), while the results for firmness were poor for Gala (*R2* = 0.667, *SEP* = 7.33 N). The very large number of 24 PCs suggests that this model was overfitting the data. [Tian et al. (2007](#_ENREF_60)) reported the prediction of SSC of two watermelon varieties Qilin (QL), Zaochunhongyu (ZC) by applying PCR on their diffuse transmittance in the range 350-1100 nm (*R2* = 0.5, *RMSEP* = 0.78 % for QL and *R2* = 0.32, *RMSEP* = 0.94 % for ZC). [Gowen et al. (2008](#_ENREF_21)) predicted quality attributes of sliced mushrooms during storage by applying PCR to 20 selected wavelengths from hyperspectral data in the range 400-1000 nm. They obtained good prediction results for: moisture content (*R2* = 0.75, RMSEP = 0.74 %), L\* values (*R2* = 0.95, RMSEP = 0.47), b\* values (*R2* = 0.75, RMSEP = 0.66), hardness (*R2* = 0.77, RMSEP = 0.49 N) and chewiness (*R2* = 0.72, RMSEP = 0.27 N). The obtained PCR models were also able to provide the mapping images of moisture content at pixel level on the sliced mushroom surfaces during storage. [Gaston et al. (2010](#_ENREF_18)) predicted the activity of polyphenol oxidase (PPO) in mushroom caps applying PCR to the mushroom hyperspectral images obtained in the range 445-945 nm with reasonable accuracy (*R2* = 0.78, RMSEP = 0.30 [ln(EAU/g)]). They also showed the possibility to acquire the predicted images of the distribution of PPO enzyme activity using the PCR model for mushroom cap quality monitoring: higher PPO activity was observed over the sample surface of damaged mushroom caps as compared to intact ones.

PLSR has by far been the most widely used method to predict quality attributes of fruit and vegetables from spectroscopic data and the obtained results have been reported in many reviewed publications ([Nicolaï et al., 2007a](#_ENREF_48); [Elmasry et al., 2012](#_ENREF_15); [Lorente et al., 2012](#_ENREF_37); [Magwaza et al., 2012](#_ENREF_42); [Pu et al., 2015](#_ENREF_55); [Wang et al., 2016](#_ENREF_62); [Arendse et al., 2018](#_ENREF_1); [Nguyen-Do-Trong; et al., 2018](#_ENREF_47)). Many researchers have reported on the use of PLSR for the prediction of SSC in various cultivars of apples and pears with good accuracy (*R2* ≈ 0.9, RMSEP ≈ 0.4 %) from NIR spectra in the 800-1700 nm range ([McGlone et al., 2003](#_ENREF_44); [Nicolaï et al., 2007b](#_ENREF_49); [Nicolaï et al., 2008](#_ENREF_50); [Fan et al., 2009](#_ENREF_16)). Dry matter content for apple could also be predicted accurately (*R2* = 0.95, RMSEP = 0.32%) from NIR spectra using PLSR ([McGlone et al., 2003](#_ENREF_44)). However, the results reported for the prediction of texture properties (e.g. firmness) and titratable acidity (TA) or pH are more variable. PLSR has also been used for predicting the optimal harvest date (*R2* = 0.81, SEP = 7.4 d) and Streif index (*R2* = 0.71, SEP = 0.18 100 . kg) for apples ([Peirs et al., 2001](#_ENREF_53); [Peirs et al., 2005](#_ENREF_54); Van Beers et al., 2015). It should be noted that the streif index is an indicator for maturity which is calculated by combining a firmness measurement expressed in kg, the SSC expressed in % and the dimensionless starch index. Good prediction results (*R*2 = 0.83, RMSEP = 0.63) were also reported for prediction of the level of internal flesh browning as evaluated through visual inspection on a 5 point scale (Khatiwada et al., 2016).

## 3.2 Nonlinear regression methods

The methods presented in section 3.1 assume that the relation between the spectral variables (X) and the quality traits of interest (Y) is linear. While it has been shown that PCR and PLSR can also cope with some mild nonlinearities by including more components in the model, they cannot cope with strong nonlinearities with an acceptable model complexity (Naes et al., 2002). Such nonlinearities can be detected as deviations from a straight line in a plot of the predicted values for the quality trait of interest against the corresponding measured values. However, they are even more obvious in a plot where the residuals (difference between predicted and measured values) are plotted as a function of the measured values.

One way to cope with such strong nonlinear relations is by applying a nonlinear transformation to the data to make the relation sufficiently linear for modelling with the above mentioned linear regression methods. As there are typically more spectral variables (X) than there are quality traits (Y), it is most convenient to transform the Y data. Commonly used nonlinear transformations are logarithms, exponential functions and power functions. If it turns out that it is not possible to find a transformation which makes the relation sufficiently linear, nonlinear regression techniques have to be used. While many different nonlinear regression techniques have been proposed, we will discuss two categories which are widely used in the analysis of spectral data: local methods and kernel methods.

### 3.2.1. Local methods

This class of methods target a particular type of nonlinearity by assuming that the complex, non-linear relation can be approximated by a linear model in the vicinity of the spectrum to be predicted. Most local methods develop a specific calibration model for every new sample based on a subset of training samples from a larger library which are considered sufficiently similar.

In **local PLS** (LPLS) a subset of samples with spectra similar to that of the sample to be predicted is selected for building a local PLSR model. The LOCAL algorithm proposed by Shenk et al. (1997) selects the spectra from the calibration set which have the highest correlation with the spectrum acquired for the new sample. As a Modified PLS algorithm is trained on this subset, a sufficiently large number of samples has to be selected to include enough variation to obtain a good calibration model. Moreover, the number of latent variables to include in the model should also be tuned. Alternative algorithms have been proposed which select the training samples for the local PLSR based on Euclidean or Mahalanobis distances instead of correlation coefficients (Shen et al., 2019).

The **locally weighted regression** (LWR) method proposed by Naes et al. (1990) also searches for spectra in the calibration set which are most similar to the spectrum of the sample to be predicted. However, in this case a PCA model is built on the calibration data and the similarity between the new spectrum and all spectra in the calibration set is calculated based on the Mahalanobis distance in the PCA space. Instead of building a PLSR model based on the selected samples, a weighted regression is performed where the samples with the smallest distance to the selected sample get the largest weight in predicting the Y value for the new sample. In this method, two parameters have to be optimized: the number of PCs to retain in the PCA model and the number of samples to be selected based on their spectral distance.

The **local central algorithm (LCA)** proposed by Zamorra-Rojas et al. (2010) also performs a PCA on the calibration set and selects samples based on their Mahalanobis distance calculated based on the selected PCs. However, it is different from LWR in the calculation of the quality trait from the selected samples. Instead of performing a weighted regression with the selected samples, the mean, median or mode of their reference values is used.

### 3.2.2 Kernel methods

The local methods discussed in section 3.2.1 have been shown to provide good predictions for large sample sets where linear models failed. However, they require that for every new sample the most similar samples are found. As this involves calculating the distance or correlation from this sample to all other samples in the library, this can become very computationally intensive for large datasets. While a large computation time may be acceptable in the training phase, this is not desirable in prediction mode. Another limitation of these methods is that the number of samples included in the local model can have a large impact on the predicted value, especially in LPLS and LCA where all selected samples get an equal weight in the local calibration. Therefore, it would be more desirable to train a calibration model which is able to capture the nonlinear relation between the acquired spectra and the quality trait of interest. Such a calibration model could then be applied to all new samples, similar to the methods discussed in section 3.1. This can be achieved by applying the Kernel trick, which involves a transformation from the original variable space to a feature space where the relation with the quality trait of interest is (more) linear. While many possible kernels exist, the most frequently used ones are the linear kernel and the radial basis function (RBF) kernel, which both calculate spectral distances between the different samples in the calibration set to define their similarity in the feature space. As the resulting feature matrix is of size *n* x *n*, the regression between this matrix **K** and the quality trait of interest suffers from similar multicollinearity problems as the original **X** matrix. Therefore, the regression coefficients either have to be estimated using a projection method such as PCR or PLSR (Kernel PCR, Kernel PLSR) or through regularization (Kernel ridge regression). The latter method is equivalent to Least Squares Support Vector Machines (LS-SVM - Suykens et al., 2002). This method was developed through simplification of the Support Vector Machines (SVM) proposed by Vapnik (1995) with 3 tuning parameters requiring quadratic programming to a least squares problem with two tuning parameters. For more details on these methods, the reader is referred to Cristianini & Shawe-Taylor (2000) and Suykens et al. (2002).

### 3.2.4. Application of non-linear regression methods on the apple example data

### The non-linear regression methods SVM regression and locally weighted regression (LWR) were also tested on the Braeburn and Jonagold apple data. Table 1 shows that the prediction performance of the LWR model is slightly worse in comparison to both PCR and PLSR, with an RMSEP value of 0.69 %. However, only four principal components were used in this model, reducing model complexity. The SVM regression model showed the best overall performance on the test set with low values for the RMSEP (0.54 %) and prediction bias (0.02 %). However, the improvement compared to PLSR is limited and might not be significant, especially because it performed slightly worse in cross-validation. This suggests that this calibration problem is fairly linear and that the weak nonlinearity might also be removed through spectral preprocessing. This will be investigated in Section 5.

### 3.2.5. Applications of non-linear regression methods in postharvest quality evaluation

The use of local methods in postharvest quality evaluation so far has been quite limited. [Luo et al. (2018](#_ENREF_40)) compared global vs. local PLSR models for predicting SSC of apples using spectroscopic data in the 667-1111 nm range. They used validation sets containing ‘extrapolation’ apples which had SSC or maturity outside the calibration sets. Similar samples for LPLS could be selected either based on their Mahalanobis distances in the PLS scores space or using correlation levels between the unknown samples with those in the calibration set. They reported that the first and second approach for similarity selection respectively improved the prediction accuracy for SSC significantly as compared to that of the global PLSR calibration model: RMSEPlocal = 1.17 % vs RMSEPglobal = 1.38 % and RMSEPlocal = 1.34 % vs RMSEPglobal = 1.52 %. Shen et al. (2019) proposed a local PLSR algorithm with preliminary data compression using the scores from a global PLS model (LPLS-S) and compared its performance to that of LPLS for the prediction of the total β-carotene content (TBC) of cassava roots. A global PLSR model trained on samples harvested from 2009 to 2012 performed well in cross-validation (RMSECV = 1.089), but gave poor predictions on the test set with samples harvested in 2013 (RMSECV = RMSEP = 1.7169, RPD = 2.39). LPLS improved the prediction performance on the test set significantly (RMSEP = 1.1962, RPD = 3.43). S-LPLS gave a similar prediction performance (RMSEP = 1.1602, RPD = 3.53), but was 20 times faster than the LPLS algorithm.

The Kernel methods have been evaluated more widely in postharvest research. Nicolaï et al. (2007b) applied kernel PLS for the prediction of sugar content (%) in Golden Delicious apples from NIR spectra in the 800-1690 nm range. They tested different types of kernel functions and optimized the tuning parameters by computer experiments. As they obtained similar prediction accuracies (RMSEP = 0.44 %) for all different kernels as those obtained with ordinary PLSR, they concluded that the use of kernel PLS was not advantageous in this case. [Chauchard et al. (2004](#_ENREF_5)) investigated NIR spectroscopy in the 680-1100 nm range to predict total acidity (TA) in grape berries. As non-linearity was identified in the calibration plots of predicted versus measured TA values at high TA values, LS-SVM regression was used to compare the prediction performances with PLSR and MLR. LS-SVM regression outperformed PLSR and MLR on the test set (*R2* = 0.83 for LS-SVM vs. 0.77 for PLSR and 0.68 for MLR). Mukarev and Walsh (2012) compared PLSR and LS-SVM regression with model pruning for SSC prediction in intact peaches from second derivative interactance spectra in the 731 to 926 nm range. The pruned LS-SVM regression model decreased the prediction error on the independent test set by 9% to an RMSEP of 14% compared to PLSR. [Dong and Guo (2015](#_ENREF_12)) applied PLSR and LS-SVM to SWIR hyperspectral data (900-1700 nm) of ‘Fuji’ apples to predict SSC, firmness, moisture content (MC), and pH values. While all models failed to predict firmness, LS-SVM outperformed others in predicting SSC, MC, and pH. [Li et al. (2013](#_ENREF_33)) predicted SSC, pH, and firmness of pears using PLSR and LS-SVM models on Vis-NIR spectra in the 400-1800 nm, 400-781 nm and 781-1800 nm ranges. The prediction performances of the two models were comparably good for all quality attributes with all the wavelength ranges, since there was little non-linear information in the spectral data and a small number of samples used. [Fernández-Ahumada et al. (2008](#_ENREF_17)) indicated that the benefits of using LS-SVM appeared when there was strong non-linearity and a high number of samples involved.

# Qualitative calibration

In section 3, the focus was on calibration methods which make it possible to predict a quality trait with continuous values (e.g. SSC, firmness). However, there are also many quality traits which are qualitative and expressed in terms of quality classes (e.g. sweet vs. sour; healthy vs. diseased; sound tissue vs. bruise vs. scald vs. puncture;…). In this case, we need to find a way to discriminate between different classes or to classify samples or pixels within a hyperspectral image in the right group based on the acquired spectra. By acquiring spectra for a set of calibration samples for which the quality class is known or can be determined destructively after the spectral measurement, we can construct a discrimination or classification model which can be applied to predict the quality class for future samples. This procedure is known as supervised classification or qualitative calibration.

## 4.1. Principal Components and Partial Least Squares Discriminant Analysis (PC-DA & PLS-DA)

Since PCR and PLSR have proven their value as quantitative calibration methods for the prediction of continuous quality traits from acquired spectra, they have also been extended to qualitative calibration. In the case of discrimination between two classes (e.g. sound vs. bruised) this can be achieved by defining a dummy variable which is set to 1 if the sample belongs to the target class and set to 0 if it belongs to the other class (Naes et al., 2002). Next, a PCR or PLSR model is built to predict these dummy variables Y based on the acquired spectra. By applying these regression models to new samples, the probability that this sample belongs to the corresponding class is predicted. If the predicted probability exceeds a pre-defined threshold value, the sample is assigned to the class. If it is below the threshold, it is assigned to the other class. This threshold can be set based on the prior probabilities (e.g. 0.5 for equal probabilities) or can be tuned to give the lowest number of misclassifications in (cross-)validation.

In the case of multiple classes, multiple dummy variables and classification rules have to be defined. This can be done in a pairwise way using a separate dummy variable to train a regression equation for every combination of two classes. Alternatively, one can define one dummy variable for every class indicating whether the sample belongs to this class (1) or not (0). This enables the development of a PLS2 regression model that uses the same set of latent variables to predict these dummy variables from the acquired spectra (Naes et al., 2002). While it is possible to classify the samples by thresholding these class memberships, better classification results may be obtained by performing linear discriminant analysis (LDA), a probabilistic approach for classification, on the predicted values for the dummy variables or the PLS2 scores (Indahl et al., 1999; Barker and Rayens, 2003).

In Figure 3, this is illustrated for the classification of apples from four cultivars (Braeburn, Gala, Golden and Jonagold) based on their Vis-NIR spectra acquired in the 380-1690 nm range. It is clear that the classification is not perfect as some samples belonging to the Braeburn class have a predicted value below the threshold, while some samples belonging to the other classes are above the threshold line. Table 2 presents the corresponding confusion matrix which indicates that 91% of the 89 Braeburn apples have been correctly classified as Braeburn. From the 8 misclassified apples, 1 has been erroneously classified as Gala, 1 as Golden and 6 as Jonagold. In the case of Gala and Golden, the accuracy is considerably lower with true positive rates (TPR) of 67 and 70.4%, respectively, because a large fraction of these apples have been misclassified in the Jonagold class.



Figure 3. PLS-DA output for the classification as Braeburn of apples from the Braeburn, Gala, Golden and Jonagold cultivars based on the acquired Vis-NIR spectra; the dashed line indicates the threshold for assigning the samples to the Braeburn class.

As PLS-DA builds a bilinear regression model to predict the dummy variables, this method assumes that the relation between the spectra and the dummy variables is linear and that the spectra for the different classes are sampled from the same multivariate normal distribution. Although PCR and PLSR are quite robust against deviations from this assumption, strong deviations may lead to poor classification performance. Therefore, alternative methods have been proposed which don’t rely on this assumption.

## 4.2. Soft Independent Modelling of Class Analogies (SIMCA)

The classification problem has some similarity with outlier detection in the sense that all samples belonging to a class are expected to be inliers for that class, while all samples not belonging to that class are expected to be outliers. In section 2.1, we have shown that PCA is very efficient in capturing the variation in a dataset of highly correlated variables and that outlying samples can be detected easily based on their Hotelling T2and Q residuals. Soft Independent Modelling of Class Analogies (SIMCA) makes use of this by defining a PCA model for every class. The spectrum acquired for a new sample **x**i is then projected onto the PCA model of every class *j* to investigate whether it belongs to this class (inlier) or not (outlier). In Figure 4, this concept is illustrated for the example with the 4 apple cultivars.



Figure 4. Q residual vs. Hotelling T2 plot for a PCA Model built for the class Braeburn as part of a SIMCA model for classifying Braeburn, Gala, Golden and Jonagold apples based on their Vis-NIR spectra. It can be seen that many samples from the other cultivars are outlying, while most Braeburn samples fall within the confidence limits (dashed lines).

The evaluation whether a sample belongs to the class or not is done based on the following distance metric which combines the Hotelling T2and Q residual values:

(18)

Where the reduced Q residual *Qr* and reduced Hotelling T2 value *Tr2* are calculated by dividing the *Q* and *T*2values for the *i*th sample by the 95 % confidence limits *Q*0.95 and *T*20.95. As the classification is done for the different classes independently, this allows for a soft classification where a sample can be classified in multiple classes for which the criterion is met or in no class at all.

In Table 2, the confusion matrix is also given for the SIMCA model built for the example case with apples from 4 cultivars. While 88 out of the 89 Braeburn apples are correctly classified, 73 Gala apples are misclassified as Braeburn, resulting in a low TPR of 31.1 %. The classification of the Jonagold apples is even worse with a TPR of 3.6% So, in this case, SIMCA is considerably less successful than PLS-DA. This may be explained by the fact that the variation in the Vis/NIR spectra for apples from the same cultivar (within group variation) is rather large compared to the differences between the cultivars (between group variation). While the cultivar-specific PCA models in SIMCA focus on capturing the variation within each class, PLS-DA is more successful in finding the subtle differences. In this case, spectral pre-processing (see section 5) might help to reduce the within-class variability and improve the performance of the SIMCA model.

## 4.3. Classification based on similarity

While SIMCA does not make any assumptions on the distribution of the data over all classes, it still builds on the assumption that the spectral data for each class follow a multivariate normal distribution. While this assumption is more likely to be fulfilled, it may also be violated. To cope with these situations, alternative methods have been proposed which do not rely on these assumptions. These typically classify samples based on their similarity with samples with known class membership. This similarity can be defined in terms of a distance metric (Euclidean distance or Mahalanobis distance calculated on original variables or PC scores) in the original variable space (e.g. k nearest neighbors – kNN) or in the feature space (e.g. SVM for classification).

In Table 2, the confusion matrices obtained with kNN and SVM classification of the Braeburn and Jonagold apples are given as well. In this case, kNN with 3 neighbors based on Euclidean distance calculated on the original variables performs better than PLS-DA and SIMCA for the Golden apples (TPR of 74.8% vs. 70.4% and 48.9%), but worse than PLS-DA for the other cultivars. SVM with a radial basis function on the original variables performs well for all three cultivars with TPR values around 90%, which makes it the best performing method for this application.

Table 2 Confusion matrices obtained with PLS-DA, SIMCA, kNN and SVM for the classification of apples based on their Vis-NIR spectra according to their cultivars: Braeburn (BR), Gala (GA), Golden (GD) and Jonagold (JG).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PLS-DA | | | | SIMCA | | | |
|  | BR | GA | GD | JG | BR | GA | GD | JG |
| Predicted as BR | **81** | 8 | 2 | 2 | **88** | 73 | 22 | 68 |
| Predicted as GA | 1 | **71** | 1 | 2 | 1 | **33** | 46 | 19 |
| Predicted as GD | 1 | 5 | **95** | 9 | 0 | 0 | **66** | 20 |
| Predicted as JG | 6 | 22 | 37 | **98** | 0 | 0 | 1 | **4** |
| TPR (%) | **91** | **67** | **70.4** | **88.3** | **98.9** | **31.1** | **48.9** | **3.6** |
|  | kNN | | | | SVM | | | |
|  | BR | GA | GD | JG | BR | GA | GD | JG |
| Predicted as BR | **78** | 23 | 8 | 16 | **83** | 2 | 0 | 0 |
| Predicted as GA | 6 | **67** | 5 | 10 | 3 | **95** | 5 | 3 |
| Predicted as GD | 3 | 9 | **101** | 18 | 0 | 4 | **125** | 7 |
| Predicted as JG | 2 | 7 | 21 | **67** | 3 | 5 | 5 | **101** |
| TPR (%) | **87.6** | **63.2** | **74.8** | **60.4** | **93.3** | **89.6** | **92.6** | **91** |

TPR = true positive rate, the fraction of samples belonging to a class that has also been assigned to that class.

## 4.4. Applications of qualitative calibration in postharvest quality evaluation

[Ariana and Lu (2008](#_ENREF_2)) built a PLSDA model on hyperspectral diffuse reflectance and transmittance images of cucumbers in the 400-1000 nm range to detect internal defects (formation of water-soaked lesions in the endocarp (seed cavity) and/or hollow at the center). They obtained overall classification accuracies of 99.7 % when using the transmittance data in the 675-1000 nm range. The PLS-DA model built on four selected wavelengths (754, 764, 951 and 993 nm) could provide overall classification accuracy of 95.8 %. [Keresztes et al. (2016](#_ENREF_31)) built a PLSDA model to detect bruises in Jonagold apples using their SWIR hyperspectral reflectance images in the 1000-2500 nm range. They obtained overall prediction accuracies at pixel level for 2 class classification (sound vs. bruised) of 88.7 % and for 3 class classification (sound vs. bruised vs. glare) of 94.3 %. [Keresztes et al. (2017](#_ENREF_30)) acquired SWIR hyperspectral reflectance images in the 1000-2500 nm range of ‘Jonagold’, ‘Joly Red’, and ‘Kanzi’ apples subjected to bruising impact. While glare regions are commonly considered as anomalies and discarded from further analysis, their spectra could be employed to predict the cultivars with PLS-DA. The reported cultivar prediction accuracies for ‘Jonagold’, ‘Joly Red’, and ‘Kanzi’ were 73.2, 80.1, and 96.5 %, respectively. They also obtained a prediction accuracy for bruise detection at pixel level of 90.1 % using PLS-DA. Post-processing of the binary image by exploiting spatial information further improved the bruise detection accuracy to 94.4 %. [López-Maestresalas et al. (2016](#_ENREF_36)) obtained better classification performance using PLSDA than using SIMCA in the detection of subsurface blackspot defects in potatoes using hyperspectral reflectance images in the VNIR (400-1000 nm) and SWIR (1000-2500 nm) range, with an overall classification rate above 94 % for both setups.

[Manley et al. (2007](#_ENREF_43)) used SIMCA to predict internal breakdown of South African Bulida apricots during cold storage using NIR diffuse reflectance in the 645-1201 nm range. They obtained classification accuracies with sensitivity of 47.9 % and specificity of 60.9 %. They hypothesized that these moderately good accuracies could have been caused by a non-linear relationship between the spectral data and the post-storage quality. [Baranowski et al. (2012](#_ENREF_3)) built SIMCA models to classify bruised apples using hyperspectral images in the VNIR (400-1000 nm), SWIR (1000-2500 nm) and MWIR (3500-5000 nm) ranges and combinations of these. They achieved prediction accuracies on separate test sets up to 67 % when combining VNIR-SWIR-MWIR to classify bruised vs. sound apples and up to 68 % when using the VNIR range to classify shallow vs. deep bruises. [Wei et al. (2014](#_ENREF_63)) classified persimmon ripeness levels (unripe, mid-ripe, ripe and overripe) by building a SIMCA model on their hyperspectral images in the 400-1000 nm range. They obtained a classification accuracy up to 73.4 % when averaging spectra from two sides (stem-calyx-end and stylar-end sides). By using 3 selected wavelengths: 518, 711 and 980 nm, the classification accuracy was slightly reduced to 68.8 % and the misclassifications occurred mostly between unripe vs. mid-ripe or ripe vs. overripe classes.

[Miller et al. (1998](#_ENREF_45)) used kNN to detect blemishes on apples using diffuse reflectance spectra in the 460-1130 nm range. They obtained good prediction accuracies on the test set for two consecutive years with sensitivity and specificity respectively ranging from 85.4 to 99 % and from 66.7 to 95.2 %. [Luo et al. (2011](#_ENREF_39)) built kNN models using NIR spectra in the 1000-2500 nm range to classify origins of Fuji apples, grades of New Jonagold apples and different apple varieties: Fuji, New Jonagold, Red Star and Ralls Janet. The obtained classification accuracies were 72.2, 80.0, and 86.5 %, respectively for origin, grade and variety. [Dang et al. (2012](#_ENREF_9)) indicated the possibility to use kNN on hyperspectral images of pears in the 949 to 1666 nm range to detect bruised pears 1 hour after bruising, when the bruises were still invisible to human eyes. [Rady et al. (2015](#_ENREF_56)) used kNN to classify glucose and sucrose contents of two potato varieties (Frito Lay and Russet Norkotah) using the selected wavelengths obtained from iPLS models for glucose quantification using potato hyperspectral scatter imaging data in the 400-1000 nm range. The threshold value for glucose content to define the two groups (low vs. high glucose content) was 0.035 %, while those for sucrose content were 0.03 and 0.10 %, respectively for the two varieties. The obtained misclassification rates for glucose content were 14 and 22 %, respectively for Frito Lay and Russet Norkotah, while the classification of sucrose content was not accurate enough. [Cen et al. (2016](#_ENREF_4)) detected chilling injury in cucumbers using hyperspectral images in the 500-675 nm (reflectance) and 675-1000 nm (transmittance) ranges using kNN on either spectral features or image texture features from two band ratio images. They reported that the SVM models outperformed kNN models in detecting cucumber chilling injury in both cases of using spectral features or image textural features for two-class or three-class classifications.

Huang et al. (2012) detected mealiness in Delicious apples using SVM and PLSDA models calibrated on apple hyperspectral scatter images in the 600-1000 nm range. They reported overall better classification accuracies (mealy vs. non-mealy apples) for SVM as compared to PLSDA models: 78.3% vs. 73.0% (using mean spectra) and 82.5% vs. 80.4% (using locally linear embedding algorithm). [Khanmohammadi et al. (2014](#_ENREF_32)) classified persimmon samples from 7 different regions in Spain according to their origin based on their FT-NIR diffuse reflectance spectra in the 740-2700 nm range using LS-SVM models. They reported classification errors up to 2 % for the validation set. Khatiwada et al. (2016) compared LDA, PLSDA, SVM and logistic regression for separation of acceptable apples from apples with internal flesh browning based on visible-short wave NIR spectra (500-975 nm) measured in partial or full transmission geometry. They obtained the best results on an independent validation population with PLSDA (classification accuracy > 95% and false discovery rate <2%). [Zhang et al. (2016](#_ENREF_71)) classified ripeness of strawberries (unripe, mid-ripe and ripe) by using SVM models built on the sample hyperspectral reflectance images in the VNIR range (380-1030 nm) or SWIR range (874-1734 nm) or textural features of the images at optimal wavelengths obtained from the loadings of PCA models. They reported overall classification accuracies on the validation set up to 95 % when using 6 selected optimal wavelengths in the VNIR range combined with textural features. Most of the misclassifications occurred between unripe vs. mid-ripe or mid-ripe vs. ripe classes.

# Spectral pre-processing

## 5.1 Spectroscopic transformations

### 5.1.1. (Pseudo)-Absorbance

As mentioned in Section 1 and Section 3.3, Beer’s law states that the ratio of the outgoing intensity *I* of a light beamover the incoming intensity *I0*, known as the transmittance *T,* in a transparent sample decreases exponentially with the path travelled through the sample *L* and the concentration *c* of the absorbing species with absorption coefficient *k*:

(19)

By taking the logarithm of the left and right side of Eqn.(19) and using the properties of the logarithm operator, the measured transmittance can be converted in an absorbance *A* which is proportional to the concentration of the absorbing species:

(20)

For transmittance measurements on a transparent liquid in a cuvette with constant thickness this transformation linearizes the relation between the predictor variables X, absorbance spectra **A**, and the predicted variables Y, concentration vector **c**,such that it can be modelled with the linear regression methods based on Eqn.(6). However, the hierarchical microstructure of biological products consisting of tissues formed by cells with cell organelles results in strong light scattering, which makes these samples highly turbid. This makes that the pathlength *L* travelled through the sample is no longer constant. Most fruit and vegetables are so turbid that the transmittance through an intact product is too low to be measured with a reasonable accuracy. In these cases, it is recommended to measure in diffuse reflectance or interactance mode where the pathlength is unknown and varies with the optical properties of the sample (Nicolaï et al., 2007). These effects of light scattering and the possibilities to cope with it will be discussed in Section 5.2. However, the transformation to absorbance may still be useful in these cases to make the relation between the acquired diffuse transmittance T or diffuse reflectance R and the quality traits of interest more linear. In these cases, Log(1/T) and Log(1/R) are referred to as ‘pseudo-absorbance’. The effect of the Log(1/R) transform on the Vis-NIR spectra of Braeburn apples is illustrated in Figure 5a, where the spectra are colored according to the corresponding SSC value similar to Figure 1a.



Figure 5 (a) Log(1/R) and; (b) Kubelka-Munck transform performed on the Vis-NIR spectra of Braeburn apples harvested in Merdorp. The preprocessed spectra are colored according to the SSC content (%).

### 5.1.2. Kubelka-Munck transform

Based on the theory proposed by Kubelka and Munck (1931) for homogeneous scattering materials with sufficient thickness, an alternative transformation may be more appropriate for spectra measured in diffuse reflectance mode:

(21)

Where *K* is the true absorbance, *S* is the scatter and *R* is the reflectance defined as the ratio of the intensity of the diffuse reflected light *I* over the incoming light *I0* (Naes et al., 2002). If the assumptions of their theory hold, this transformation makes the relation between the transformed reflectance and the absorbance linear, while the effect of light scattering becomes purely multiplicative. This makes it possible to remove the effect of differences in light scattering by multiplying all wavelength variables by the same constant. The effect of the Kubelka-Munck transform on the Vis-NIR spectra of Braeburn apples is illustrated in Figure 5b. It should be noted that this transformation has enhanced the absorption peaks related to the biological chromophores compared to the Log(1/R) transform in Figure 5a. The resulting spectra have high similarity with the absorption coefficient spectra of apple tissue estimated from double integrating spheres measurements, as reported by Saeys et al. (2008) and Van Beers et al. (2017).

## 5.2 Scatter correction

When the quality trait of interest is the concentration of a chemical component (e.g. sugar content), the variation in the tissue microstructure (e.g. porosity) may have a negative impact on the prediction of the quality trait based on the acquired spectra. In these cases, it may be beneficial to remove the effects of light scattering as much as possible from the spectra. To this end, a wide range of empirical scatter correction techniques have been proposed, of which the dominant ones will be discussed in this section.

### 5.2.1. Multiplicative signal correction (MSC) and Standard Normal Variates (SNV)

Kubelka and Munck (1931) indicated that after transformation with Eqn.(21) the effect of scattering would be purely additive. Geladi et al. (1985) observed a similar multiplicative effect of light scattering for the Log(1/R) transform. When plotting diffuse reflectance spectra for different samples against the mean spectrum, they also noted an additive effect which they attributed to the fact that not all diffusely reflected light is captured by the detectors. Therefore, they proposed a model with an additive term *ai* and a multiplicative term *bi* for the deviation of the spectrum **x**i of the *i*-th sample from its scatter free spectrum. As *ai* and *bi* are unknown, estimates and are obtained through linear regression of this spectrum against a reference spectrum (typically the mean spectrum). This method is known as multiplicative signal correction (MSC) and the corrected value at the *k*-th wavelength variable *x*ik\*MSC is calculated as follows:

(22)

If the baseline is not constant over the full wavelength range, it may be more interesting to perform it piecewise for different intervals (PMSC: Isaksson & Kowalski, 1993).

An alternative approach to remove the additive and multiplicative effects due to light scattering was proposed by Barnes et al. (1989). Their approach, known as standard normal variates (SNV) involves a spectrum-wise standardization where the mean value of the *i*-th spectrum *mi* is subtracted from every wavelength variable *xik*, and the mean-centered value is divided by the standard deviation *si* of this spectrum to obtain the scatter corrected value *xik\*SNV* :

(23)

The similarity between Eqns.(22) and (23) shows that although the correction terms are calculated in a different way, MSC and SNV are conceptually very similar. It has been shown that the results obtained with both methods are empirically equivalent apart from the scale of the transformed spectra (Helland et al., 1995). This can also be seen in Figure 6a and Figure 6b, which are visually identical apart from the Y axis scale. They both remove the variation (mainly due to scattering effects) in the spectral regions with low absorption, like around 800 nm. These preprocessing steps seem to improve the difference between low and high SSC apple spectra in the higher wavelength ranges above 1500 nm.

As MSC and SNV have been widely used for scatter correction of NIR spectra, several limitations of these methods have been identified for which solutions have been proposed. As MSC changes the mean spectrum, it has been suggested to perform it multiple times until the mean spectrum no longer changes (Loopy MSC: Windig et al., 2008). To extend the scatter correction to more complex baselines and safeguard the chemical information, the extended multiplicative signal correction (EMSC: Martens & Stark, 1991; Martens et al., 2003) and a more physics-based version of it (EMSCL: Thennadil et al., 2006) have also been proposed.

### 5.2.2. Baseline correction

The additive and multiplicative effect of light scattering and other background signals (e.g. autofluorescence in Raman spectroscopy) can also be interpreted as a baseline on which the chemical absorption peaks ‘sit’. Therefore, different methods have been proposed to remove this baseline. Barnes et al. (1989) suggested that instead of just subtracting the mean of each spectrum, it could also be useful to fit a first or second order polynomial to each spectrum to capture a linear or curved baseline and remove it by subtraction. This method is typically referred to as detrend. However, as this method does not make a distinction between the signal level due to the background signals and that due to the absorption, it may remove valuable information. To safeguard the chemical information, it has been proposed to fit the baseline in the Log(1/R) spectra through an automatic weighted least squares procedure where all variables above the baseline are considered to correspond to an absorption peak and are therefore excluded from the fitting procedure (Wise et al., 2006). This is illustrated for the Braeburn spectra in Figure 6c.

### 5.2.3. Derivatives

An alternative way to remove baseline effects and focus on the curvature in continuous spectra, is by calculating derivatives. The first derivative calculates the slope of the spectrum at every wavelength variable, thus removing an additive baseline. Two spectra which were initially parallel to each other will have the same first derivative spectrum, because the slope at every wavelength variable is identical (Naes et al., 2002). Analogously, the second derivative quantifies the slope in the first derivative spectra, which corresponds to the curvature in the original spectra. As derivative calculation through differencing would magnify the measurement noise present in the spectra, smoothed derivatives are calculated over an interval through averaging (Norris & Williams, 1984) or by fitting a polynomial (Savitzky and Golay, 1964). By changing the width of the interval the degree of smoothing can be tuned: a wider interval will make the derivative less sensitive to measurement noise at the price of removing the sharp features related to narrow absorption peaks.

While derivative spectra may be very interesting to remove baseline effects and enhance subtle curvature effects in the spectra, the interpretation of derivative spectra should be done with care. In the first derivative spectra, every single absorption peak in the original spectra will be replaced by a positive (upward slope) to the left of it and a negative (downward slope) peak to the right (Figure 6d). Calculation of the second derivative spectra results in additional multiplication and movement of the peaks (Figure 6e).



Figure 6. Illustration of the effect of scatter correction techniques on the Log(1/R) transformed Vis-NIR spectra of Braeburn apples harvested in Merdorp, Belgium: (a) MSC; (b) SNV; (c) WLS baseline correction; (d) 1st derivative and a (e) 2nd derivative. The preprocessed spectra are colored according to the corresponding SSC content (%).

## 5.3 Interference removal

Once the background signals due to light scattering or other baseline effects have been removed, there may still be substantial variation in the acquired spectra which is not caused by the quality trait of interest. This spectral variation may be attributed to other chemical components which are spectrally active or to external sources such as temperature effects or instrument drift. To minimize the effect of these interferents on the prediction of the quality trait of interest, different filtering methods have been proposed.

### 5.3.1. OSC and OPLS

As a counterpart to PLSR, orthogonal signal correction (OSC), defines latent variables which maximally capture the variation in the spectra which is orthogonal (not correlated) to the quality trait of interest. The contribution of these orthogonal latent variables is then removed from the X matrix to obtain a filtered matrix for which the relation to the quality trait of interest is easier to model (Wold et al., 1998). In principle, one would expect that the number of latent variables required by the PLSR model would decrease by one for every latent variable consumed in the OSC. However, empirical results have shown that in some cases OSC is more efficient, while in others it is not (Fearn et al., 2000).

Instead of performing two separate steps involving filtering by OSC and model building by PLSR, it could be more interesting to perform both steps at the same time. This would make it possible to obtain a model which decomposes the spectra into latent variables related to the quality trait of interest and latent variables orthogonal to the quality trait of interest. This method is known as O-PLS (Trygg & Wold, 2002). While this typically does not provide better prediction performance than PLSR, it can improve the interpretability. To further improve its interpretational functionality it has been extended to O2-PLS which also considers X-orthogonal variation in Y (Trygg, 2002).

### 5.3.2. EPO and Pre-whitening

A major drawback of OSC is that it estimates the orthogonal components capturing the contribution of the interferents to the spectra based on the training set which is also used for building the multivariate calibration model (PCR, PLSR, LS-SVM,…). This assumes that there is no correlation between the effect of the interferents and the quality trait of interest in the calibration set. As will be discussed in Section 6, this should be the aim when selecting the calibration set, but may be very hard to reach (and verify) in practice. Therefore, it could be more interesting to quantify the effect of known interferents independently and reduce or remove their effect prior to feeding these data to the multivariate calibration model.

Roger et al. (2003) proposed to estimate the part of the X space that is mostly influenced by the external parameter variations by computing a PCA model on a small set of spectra measured on the same objects while the external parameter is varying. By deflating the X matrix for the variation captured by this PCA model, the matrix can be orthogonalized for the effect of the external parameter. This method is known as External Parameter Orthogonalization (EPO).

Instead of subtracting the parasitic subspace through the orthogonalization step, Martens et al. (2003) proposed to make the model ‘blind’ to certain interferents while retaining its analyte sensitivity. This is done by scaling the input spectrum with a scaling matrix **G** which is constructed based on a noise covariance matrix **Σ**. This noise covariance matrix may be estimated in a similar way as the external parameter space in EPO, but it can also be defined based on prior knowledge of the interferent spectra. They showed that this weighting approach has similarity with weighted and generalized least squares (GLS) and can simplify and robustify the subsequent calibration model considerably.

## 5.4 Comparison of the different spectral pre-processing techniques on the apple example dataset

So far, all calibration models were built using a mean centering (MC) procedure as the only form of preprocessing. It subtracts the mean value per variable from all the data points to center the model around the center of the calibration set. Even if other preprocessing steps are used, this should be the last step prior to training the calibration model. The effect of the different preprocessing steps presented above on the performance of the subsequent PLSR models is quantified in Table 3. It can be noted that performing the log(1/R) and Kubelka-Munck transformation in combination with mean centering did not increase the model performance (compared to just MC) for this apple dataset. The visual improvement by MSC and SNV observed in Figure 6 did not result in more accurate prediction of the SSC. Instead, higher values for the RMSEP and prediction bias were obtained than with only MC as preprocessing. The first and second derivative also did not result in a higher prediction performance, but the second derivative obtained a similar prediction performance with a lower number of latent variables, thus decreasing the model complexity. While a weighted least squares baseline correction also had a positive effect on the apple spectra as illustrated in Figure 6c, it also did not result in an improvement in the prediction performance.

Finally, two interference removal techniques (OSC and EPO) were applied to the apple data. Their effect is illustrated in Figure 7.



Figure 7. Illustration of the effect of interference removal techniques on the Vis-NIR spectra of Braeburn apples harvested in Merdorp, Belgium: (a) OSC; (b) EPO; The preprocessed spectra are colored according to the SSC content (%).

From Table 3, we can conclude that especially the OSC preprocessing step improves the prediction performance of the PLSR model, resulting in an RMSEP value of 0.52 % with 1 latent variable in the OSC and 4 latent variables in the PLSR model. On the other hand, EPO did not improve the model performance, because the parasitic subspace had to be estimated from the calibration data.

So far, we have only evaluated the effect of individual preprocessing methods in combination with mean centering. However, it could be interesting to combine scatter correction and interference removal. This is illustrated in Table 3, where spectroscopic transformation to pseudo-absorbance has been combined with scatter correction, and both a 1st and 2nd derivative have been combined with OSC and MC. The spectroscopic transformation to pseudo-absorbance prior to scatter correction seems to have a detrimental effect, while the performance of the models combining a derivative with OSC is significantly better, with a strong reduction in the prediction bias and RMSEP values as low as 0.36 %. The larger gap between the RMSEC and RMSECV for 2nd Der – OSC – MC suggests that the model is overfitting on the training data. However, as it resulted in a lower RMSEP, this overfitting was not detrimental for the predictions on the test set. It should be noted that another combination might be more appropriate for other cases. While it may be valuable to test different combinations, it is recommended to visually inspect the actual effects of different techniques and combinations on the spectra to avoid selection of over-optimistic models which will not be robust in practice.

*Table 3 Performance of PLSR models using different combinations of preprocessing techniques*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Preprocessing | RMSEC  % | RMSECV  % | RMSEP  % | Pred. bias  % | #LVs |
| MC | 0.60 | 0.64 | 0.58 | 0.09 | 8 |
| Log(1/R) + MC | 0.61 | 0.64 | 0.62 | 0.13 | 8 |
| Kubelka-Munck + MC | 0.69 | 0.73 | 0.67 | 0.09 | 7 |
| SNV + MC | 0.62 | 0.68 | 0.60 | 0.14 | 8 |
| MSC + MC | 0.63 | 0.67 | 0.62 | 0.15 | 8 |
| 1st Der. + MC | 0.63 | 0.66 | 0.63 | 0.09 | 5 |
| 2nd Der. + MC | 0.60 | 0.63 | 0.58 | 0.08 | 5 |
| Baseline + MC | 0.81 | 0.83 | 0.78 | 0.06 | 3 |
| EPO + MC | 0.61 | 0.65 | 0.59 | 0.07 | 7 |
| OSC + MC | 0.54 | 0.58 | 0.52 | 0.11 | 4 |
| Log(1/R)+SNV+MC | 0.68 | 0.72 | 0.65 | 0.18 | 8 |
| Log(1/R)+MSC+MC | 0.66 | 0.69 | 0.64 | 0.16 | 8 |
| Log(1/R)+1st Der.+MC | 0.68 | 0.72 | 0.67 | 0.13 | 5 |
| Log(1/R)+2nd Der.+MC | 0.67 | 0.68 | 0.64 | 0.15 | 5 |
| Log(1/R)+Baseline+MC | 0.68 | 0.71 | 0.69 | 0.16 | 7 |
| 1st Der. + OSC + MC | 0.39 | 0.46 | 0.38 | 0.03 | 3 |
| 2nd Der. + OSC + MC | 0.30 | 0.48 | 0.36 | 0.01 | 1 |

## 5.6 Applications of spectral preprocessing in postharvest quality evaluation

Scatter correction techniques have been most frequently used in postharvest research for quality assessment of fruits and vegetables using spectroscopic sensors. [Bobelyn et al. (2010](#_ENREF_4)) predicted SSC and firmness of about 6000 apples covering different cultivars, seasons, shelf-life and origins by applying PLSR on the acquired diffuse reflectance spectra in the 800-1690 nm range. All the obtained reflectance spectra were converted to absorbance values (Abs) using the log(1/R) transformation before applying other pre-processing methods and PLSR modeling. For SSC, SNV outperformed second derivatives and raw Abs spectra on separate validation sets and provided RMSEP values of 0.6 – 0.8 % for the different cultivars. In the case of firmness, the best results (RMSEP = 5.9 – 8.8 N) were obtained without scatter correction.

[Cunha Junior et al. (2015](#_ENREF_7)) classified intact açaí (Euterpe oleracea Mart.) and juçara (Euterpe edulis Mart) fruits based on their dry matter content (DM) (lower vs. higher than 32 %) with PLSDA and PCA-DA models (combining PCA and LDA) and also quantified %DM with PLSR models using NIR spectra in the 1000-2500 nm range in combination with scatter correction methods: MSC or 2nd derivatives after Savitzky-Golay smoothing using 15 point windows and second order polynomial. The PLSR models for % dry matter prediction from the full 1000-2500 nm range built on a calibration set containing 371 samples obtained the best prediction accuracy in cross-validation and test set prediction with MSC preprocessing (*R2CV* = 0.81, RMSECV = 2.45 %; *R2P* = 0.54, RMSEP = 5.43 %). [Keresztes et al. (2016](#_ENREF_31)) used SWIR hyperspectral imaging in the 900-2500 nm range for early detection of bruises on Jonagold apples in real time on a moving conveyor using PLSDA. They indicated that first derivative pre-processing provided the best classification accuracy at the pixel level (up to 98 %) and a processing time per apple below 200 ms.

In contrast to the scatter correction techniques, the interference removal techniques have been used less frequently in postharvest research. Roger et al. (2003) demonstrated that the prediction of SSC (%) in intact apples from Vis-NIR spectra in the 300-1100 nm range could be made robust against a temperature variation of 20°C. To this end, they orthogonalized the spectra for the parasitic subspace estimated by applying PCA to the spectra acquired for 10 fruits measured at 8 different temperatures. Thanks to the EPO preprocessing the bias induced by a temperature variation of 20 °C was reduced from 8 % to 0.3 %. [Khanmohammadi et al. (2014](#_ENREF_32)) used LS-SVM to classify the origins of persimmon samples harvested from 7 different regions in Spain using their FT-NIR spectra in the 740-2700 nm range. They concluded that OSC preprocessing in combination with LS-SVM significantly reduced the overall classification error on a separate validation set from 24 % (LS-SVM) to 2 % (OSC-LS-SVM).

# Model calibration and validation

## 6.1 Selection of representative calibration data

The multivariate calibration methods discussed in section 3 and 4 require estimation of the model parameters based on a set of samples for which the quality traits are known and spectra have been acquired: the calibration set. To obtain good generalization of the model on future samples, it is important that it is trained on a calibration set which is representative for the population of samples on which this model will be used. To avoid that the model has to extrapolate, the range for the quality trait(s) of interest should be at least equally large and preferably slightly larger than the range of values that is expected in future samples. As this is quite intuitive, few researchers violate this rule. However, what many researchers ignore is that this calibration set should also cover the relevant variation in all other factors which contribute to the spectra and the correlations between these factors. For example, it has been shown that cultivar, season, country of origin, orchard and the exposure to shelf-life conditions have an impact on the NIR spectra of apples (Bobelyn et al., 2010) and the prediction of SSC and firmness from these spectra. So, if we want to build a calibration model for SSC prediction from the NIR spectrum acquired for any apple, the calibration set will have to include the relevant variation related to the effects of cultivar, season, country of origin and orchard together with other relevant sources of spectral variation such as temperature or different instruments over the full range of SSC values.

Different approaches can be followed to include the relevant variation in the calibration set. The most commonly used approach is random sampling from the population. According to the central limit theorem, this should provide a representative calibration or validation set if the number of samples is sufficiently large. However, it is hard to determine which number of samples is sufficient to cover all relevant variation. Moreover, truly random sampling may not be feasible in practice. Therefore, it is recommended to use a designed sampling scheme which aims at a proper representation of known sources of variation. For example, Bobelyn et al. (2010) acquired spectra for around 6000 apples from different combinations of 6 different cultivars and up to 5 different origins (countries or orchards). From each subset of about 200 apples per cultivar-origin combination, half of the samples were scanned fresh, while the other half was scanned after 1 week of shelf-life. This allowed them to actively include the effects of cultivar, origin and shelf-life as well as their interactions in the calibration set to make the calibration models robust against these. While this targeted approach guaranteed that variation from these different sources was included in the calibration set, the 200 samples per cultivar-origin combination were still sampled randomly. As the SSC and firmness values in a population of apples approximately follow a normal distribution, this resulted in the selection of a large number of samples with SSC and firmness values close to the average value. While these provide a lot of information on the mean value for this sub-population, they are not very informative for building the regression models which aim at predicting the deviation from the mean value for SSC and firmness based on the deviation of the acquired spectra from the mean spectrum. Therefore, one might consider to deliberately select samples which deviate more from the mean and thus provide more information for estimating the regression model. One option would be to deliberately select samples with high, intermediate and low values for the quality traits (e.g. firmness and SSC). However, as the reference measurements for these quality traits are more costly and time consuming than the spectral measurements, it would be more interesting to select the samples based on their spectra. Isaksson & Naes (1990) proposed to acquire spectral data for a large number of samples and perform a cluster analysis on the acquired spectra to decide which samples would be most informative for training the calibration model. As the reference analyses for the quality traits only have to be performed for the selected samples, this approach may considerably reduce the time and cost required for the reference analyses.

As it is often not feasible to acquire spectra for all possible combinations of the relevant sources of spectral variation, these effects may be estimated by measuring a limited number of samples under the different conditions (e.g. on different instruments, at different temperature or moisture contents). This matrix of perturbations or repeatability file can then be used to remove or downweight the effect of this source with EPO or pre-whitening, as discussed in section 5.3.2. However, it can also be used to augment or perturb the calibration set with these effects to ensure that the estimated regression model is robust against these perturbations (Westerhaus, 1990; Mevik et al., 2004; Perez-Marin et al., 2007). Acharaya et al. (2014) evaluated different approaches to robustify PLSR models for prediction of the dry matter content of tomatoes from NIR spectra against temperature variation. They concluded that augmentation of the calibration set with spectra of the same samples acquired at different temperatures only results in robust models if their proportion is sufficiently large (> 0.5%). Inclusion of the temperature effect based on a repeatability file gave the best prediction results on a test set of samples measured at a different temperature than the ones included in the calibration set.

## 6.2 Selection of a proper validation strategy

All calibration methods discussed in section 3 have one or more parameters which determine the model complexity and thus its ability to capture the relation between the spectra (X) and the quality traits (Y) in the calibration set. For example, in the case of PCR and PCR-DA this is the number of principal components (PCs) included in the model, while in the case of PLSR and PLS-DA this is the number of latent variables (LVs). In Section 5, spectral pre-processing methods have been discussed to simplify the relation between X and Y. Therefore, it is expected that a successful pre-processing method will reduce the model complexity to obtain similar prediction performance or reach better prediction performance for a similar model complexity. As the true dimensionality of the underlying relation is typically unknown, this trade-off has to be made by evaluating the model performance for different levels of complexity (e.g. number of PCs or LVs in the model) on a representative set of samples which has not been used for building the model. As this set is used for validating models with different complexity (combination of pre-processing and degrees of freedom in the model), it is often referred to as a ‘validation set’. However, some researchers refer to this set as a calibration test set, which should not be confused with the independent test set discussed in section 6.3 which is used for evaluation of the predictive performance of the final model and is sometimes referred to as ‘independent validation set’.

Ideally, the validation data have been acquired independently from the calibration data and do not contain any chance correlations which are also present in the calibration set. To obtain good generalization to future samples it is crucial that the validation set contains the challenges which the model will be confronted with during its future use. For example, Bobelyn et al. (2010) evaluated the performance of their calibration models for different apple cultivars on data from a different season, a different origin or a different storage condition to evaluate the robustness of their models against these sources of variation.

In practice, one is typically limited in the number of samples that can be collected and analyzed with the reference methods. Therefore, one would like to use as many samples as possible for building the calibration model, while on the other hand it is also desirable to use as many samples as possible for validating the model. When a sufficiently large number of samples is available (> 150 samples), it is recommended to use 2/3 of the data for calibration and 1/3 for validation. This selection can be done sequentially or randomly, but this poses the risk to select by chance a validation set which is ‘easy’ and thus leads to overfitting and over-optimistic results, or a validation set which is ‘difficult’ and thus leads to poor results. To avoid this problem, dedicated algorithms are available which aim at splitting the data in two sets which both span the multidimensional space almost equally well. Two widely-used examples are the Kennard-Stone (Kennard & Stone, 1969) and Duplex algorithms (Snee, 1977).

If the available data is rather limited, it is more interesting to use multiple calibration-validation splits and combine the results from these. This approach is known as cross-validation and different variants exist: leave-one-out, contiguous blocks, venetian blinds, random subsets and groupwise cross-validation. The latter is interesting when there is a known group structure in the data (e.g. samples measured on different days, from different orchards,…), because it allows for evaluating how well the models will be able to predict samples from a new group. This is especially important when there might be a correlation between the group and the quality trait to be predicted. Kemps et al. (2010) clearly demonstrated the importance of choosing the right validation strategy in multivariate calibration of a spectroscopic sensor for postharvest quality evaluation of eggs. As the quality trait of interest (pH) was highly correlated with the measurement day, leave-one-out and random cross-validation resulted in the selection of too complex models which were over-fitting the day-specific measurement noise. However, a groupwise cross-validation strategy keeping all samples measured on the same day together in the calibration or validation set, correctly indicated that the pH could not be predicted with an acceptable accuracy from the acquired spectra.

## 6.3 Test set evaluation

As different combinations of pre-processing and model complexity are evaluated based on the same validation set or cross-validation, there is a risk that the obtained prediction performance is somewhat over-optimistic. Therefore, it is recommended to evaluate the prediction performance of the final model (with the selected pre-processing and model complexity) on an independent test set representative of the population of samples on which the calibration model will be used in the future (e.g. samples from a new orchard or season). In practice, it is often recommended to optimize the model complexity based on a cross-validation strategy involving the anticipated challenges (e.g. season-wise cross-validation) and evaluate its performance on a separate test set (e.g. new season). If no truly independent test set of sufficient size is available, one might consider to perform a double cross-validation where the model complexity is optimized in the inner loop (cfr. multiple calibration-validation splits), while the prediction performance of the optimized model is evaluated in the outer loop (cfr. test set evaluation).

# Calibration transfer

## 7.1. Methods for calibration transfer

As establishing and validating a calibration model requires a large investment for every sample type and quality trait of interest, it is desirable to leverage this investment by transferring the calibration to other instruments. However, the variability between spectrophotometers, even those from the same type, makes that this transfer typically results in an increase in the prediction error (Workman, 2018). Different strategies have been proposed to overcome these limitations, which can according to Fearn (2001) be classified in three groups: 1) development of calibrations which are robust against the instrument variation, 2) modifying the acquired spectra, 3) modifying the calibration model. However, this classification is not clear cut, as some methods fit in the different classes.

One way to develop robust calibration models is by including samples from different instruments in the calibration set. In this way, the calibration algorithm is taught that this instrument variability should be ignored when predicting the quality trait of interest. Instead of letting the calibration model handle this ‘noise’, the spectral pre-processing methods discussed in section 5 can be used to remove the instrument-specific effects. Both the scatter correction methods (MSC, SNV, derivatives,…) and the methods for interference removal (EPO, OSC) have been shown to be beneficial for removing additive and multiplicative, instrument-specific effects. If the number of samples measured on the different instruments is unbalanced, the calibration set can be augmented or perturbed with the instrument differences obtained from measurements with different instruments on the sample, as suggested in section 6.1. However, these spectrum-wide pre-processing methods are not capable of removing wavelength shifts and are only likely to work when the differences between the instruments are fairly small (Fearn, 2001). To build calibration models which are robust against wavelength shifts, it has been proposed to carefully select a small number of wavelengths to be used in an MLR calibration (Swierenga et al., 2000)

When one wants to apply an established calibration model on a new instrument (e.g. replacement instrument or portable instrument), the instrument on which the calibration model has been established is referred to as the “master”, while the other instruments to which the calibration model has to be transferred are referred to as “slaves”. In this case, the most cost efficient way to perform a calibration transfer is by measuring a number of samples both on the master and slaves, and adjusting the spectra acquired on the slaves to make these look as if they have been measured on the master. This makes it possible to apply the calibration model(s) available for the master to the spectra acquired on the slaves. These methods are known as direct standardization methods. Different variants exist of which piecewise direct standardization (PDS) is the most commonly used (Wang et al., 1991). This method performs a PCR or PLSR to predict the adjusted absorbance values at each wavelength from the raw absorbance values in a region around the target wavelength. While this approach is fairly straightforward and does not require reference analyses for the transfer samples, these samples have to be selected carefully to obtain good quality estimates for the large number of model parameters required for the adjustment of an entire spectrum (Fearn, 2001). Moreover, the requirement that the same samples are scanned on different instruments in the same condition may not be trivial for fruit and vegetables if the instruments cannot be placed next to each other. Artificial samples or generic standards seem to be the most feasibly option in this case, but care has to be taken that these are sufficiently representative for the differences which can be expected when measuring the target samples on both instruments (e.g. wavelength shifts, intensity variation,…).

If it is not possible to measure the same samples on both instruments (e.g. when the master instrument is no longer available), the only option left is to adjust the predictions by the calibration model based on a limited number of transfer samples for which reference data is available. When the calibration model established on the old instrument is applied to the spectra acquired on the new instrument, the predicted values for the quality traits can be compared to the reference values. This allows to estimate a transformation equation. The most simple form is a bias correction to remove the systematic deviation between the predicted and reference values. However, a linear correction involving a slope and an intercept may be more appropriate. In this case, it is crucial that the transfer samples cover a sufficiently large range in the quality traits of interest to estimate the slope.

## 7.2. Applications of calibration transfer in postharvest quality evaluation

Greensill et al. (2001) and Greensill and Walsh (2002) investigated the transfer of calibration models between photodiode array spectrometers in the NIR range for SSC measurement respectively in rockmelons and mandarins. They obtained the best results with a direct standardization method based on a wavelet decomposition of the spectra, which were on par with predictive model updating based on selected representative spectra from the slave spectrometer.

Alamar et al. (2007) applied PDS to transfer a prediction model for SSC of apple fruit from a Fourier Transform NIR spectrometer to a photodiode array spectrometer, and between two photodiode array spectrometers. To obtain prediction errors (RMSEP) on the slave instruments comparable with those on the master instrument they required a transfer set of 15 spectra and a window size of 15 wavelengths. However, for transfer between the photodiode array spectrometers only a window size of 3 wavelengths was required. Hayes et al. (2016) also obtained good results with PDS in a similar transfer between photodiode array instruments for TSS measurement in intact apple fruit. However, they found that correction for the wavelength assignment errors in the photodiode array instruments in combination with difference spectrum adjustment and model updating based on the measurement of 10 fruits also performed well and was easier to implement. Barreiro et al. (2008) and Salguero-Chaparro et al. (2013) reported good results for PDS and with transfer by orthogonal projection (TOP), a transfer method related to EPO, in the transfer of calibration models for quality assessment of olives between a laboratory NIR equipment with a scanning monochromator and a portable prototype based on a photodiode array. PDS also outperformed other standardization methods in the transfer of calibration models for SSC in bananito flesh from a handheld microNIR spectrometer to a desktop hyperspectral imaging system (Pu et al., 2018).

# Conclusions and recommendations

The use of vibrational spectroscopy methods for fast and non-destructive evaluation of postharvest quality traits requires the elaboration of calibration equations which generalize well to new samples. As the spectral variables are typically non-specific and highly correlated, this requires multivariate data analysis techniques which are able to reduce the dimensionality of the data and extract the underlying relations. Thanks to its effectiveness in reducing the dimensionality, principal component analysis (PCA) has become the dominant technique for exploring spectral data for the presence of groupings and outliers. While it can also be used to overcome the multicollinearity problems in multivariate regression, regression based on the principal components (PCR) is considered suboptimal compared to partial least squares regression (PLSR) which defines the dimensionality reduction based on the relation between X and Y. Consequently, PLSR has become the dominant multivariate regression technique for the calibration of spectral sensors. As fruit and vegetables have a complex microstructure and contain many chemical components which are spectrally active, the relation between the acquired spectra and the quality traits of interest may be fairly complicated, requiring a high number of latent variables in the PLSR model. Spectral preprocessing techniques may help to simplify the relationship, thus reducing the model complexity and providing better generalization on new samples. Combination of a spectroscopic transformation, some scatter correction and interference removal may result in a substantial reduction in the required number of latent variables and/or improvement in the prediction performance. However, one should be aware that these preprocessing techniques also consume degrees of freedom and are only justified if their effect is substantial. When a non-linear relation between the spectra and the quality traits of interest is expected, non-linear calibration models can be built using local linear methods (LWR, LOCAL, LCA) or a transformation to kernel space (kernel PLS, SVM). However, this comes at the price of an increased model complexity and risk of over-fitting. Moreover, it is more difficult to verify if the spectral information used by these methods is indeed related to the quality trait of interest. In most reported cases, transformation of the Y-variable or spectral preprocessing allowed for making the relation between the spectra and the quality trait of interest fairly linear. Moreover, PLSR can cope with weak non-linearity. Therefore, preprocessing may be the better option. For both the linear and non-linear calibration methods, variants are available for predicting qualitative traits and many applications of this have been reported in postharvest research. Last but not least, the calibration data on which the model is built and the validation strategy to optimize the model complexity should be selected carefully to obtain a calibration model which will generalize well to new data. This prediction performance should always be evaluated on a test set which is as independent as possible and contains the challenges expected in future use. As most instrument manufacturers are still not able to produce identical instruments, a calibration transfer may be required before an established calibration can be applied on a new instrument. This will be especially important to avoid the need for complete recalibration when moving from off-line measurements in the laboratory to on-line and in-line measurements in the packhouse or the orchard.

# Acknowledgements

The authors gratefully acknowledge the Flanders Center for Postharvest Technology (VCBT) for providing the apple data, and the financial support from Flanders Innovation & Entrepreneurship (IWT 140992 – IFAST) and the Flemish Community’s Special Research Fund (BOF - research professorship Prof. Wouter Saeys and postdoctoral mandate Dr. Robbe Van Beers).

# References

Acharya, U.K., Walsh, K.B., Subedi, P.P., 2014. Robustness of partial least-squares models to change in sample temperature: II. Application to fruit attributes. Journal of Near Infrared Spectroscopy 22, 287-295.

Alamar, M. C., Bobelyn, E., Lammertyn, J., Nicolaï, B. M., & Moltó, E., 2007. Calibration transfer between NIR diode array and FT-NIR spectrophotometers for measuring the soluble solids contents of apple. Postharvest Biology and Technology 45, 38–45. https://doi.org/10.1016/j.postharvbio.2007.01.008

Arendse, E., Fawole, O.A., Magwaza, L.S., Opara, U.L., 2018. Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review. Journal of Food Engineering 217, 11-23.

Ariana, D.P., Lu, R., 2008. Quality evaluation of pickling cucumbers using hyperspectral reflectance and transmittance imaging: Part I. Development of a prototype. Sensing and Instrumentation for Food Quality and Safety 2, 144-151.

Baranowski, P., Mazurek, W., Wozniak, J., Majewska, U., 2012. Detection of early bruises in apples using hyperspectral data and thermal imaging. Journal of Food Engineering 110, 345-355.

Barker, M., Rayens, W., 2003. Partial least squares for discrimination. J. Chemometrics 17, 166-173.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard Normal Variate Transformation and Detrending of Near Infrared Diffuse Reflectance. Appl. Spectrosc. 43,772-777.

Barreiro, P., Herrero, D., Hernandez, N., Gracia, A., Leon, L., 2008. Calibration transfer between portable and laboratory NIR spectrophotometers. Acta Hort. 802, 373-378.

Bobelyn, E., Serban, A., Nicu, M., Lammertyn, J., Nicolai, B., Saeys, W., 2010. Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. Postharvest Biology and Technology 55(3), 133-143.

Cen, H., Lu, R., Zhu, Q., Mendoza, F., 2016. Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification. Postharvest Biology and Technology 111, 352-361.

Chauchard, F., Cogdill, R., Roussel, S., Roger, J.M., Bellon-Maurel, V., 2004. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. Chemometrics and Intelligent Laboratory Systems 71, 141-150.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines, Cambridge University Press (ISBN 0-521-78019-5).

Cozzolino, D., 2015. Infrared Spectroscopy as a Versatile Analytical Tool for the Quantitative Determination of Antioxidants in Agricultural Products, Foods and Plants. Antioxidants 4.

Cunha Junior, L.C., Nardini, V., Khatiwada, B.P., Teixeira, G.H.d.A., Walsh, K.B., 2015. Classification of intact açaí (Euterpe oleracea Mart.) and juçara (Euterpe edulis Mart) fruits based on dry matter content by means of near infrared spectroscopy. Food Control 50, 630-636.

Dale, L.M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., Pierna, J.A.F., 2013. Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. Applied Spectroscopy Reviews 48, 142-159.

Dang, H., Kim, I., Cho, B., Kim, M.S., 2012. Detection of bruise damage of pear using hyperspectral imagery, 2012 12th International Conference on Control, Automation and Systems, pp. 1258-1260.

De Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 18(3), 251-263. https://doi.org/10.1016/0169-7439(93)85002-X

De Juan, A., Tauler, R., 2006. Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications. Critical Reviews in Analytical Chemistry 36: 163-176.

De Marchi, M., Toffanin, V., Cassandro, M., Penasa, M., 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits1. Journal of Dairy Science 97, 1171-1186.

Dhakal, S., Chao, K., Qin, J., Kim, M., Chan, D., 2016. Raman spectral imaging for quantitative contaminant evaluation in skim milk powder. Journal of Food Measurement and Characterization 10, 374-386.

Dong, J., Guo, W., 2015. Nondestructive Determination of Apple Internal Qualities Using Near-Infrared Hyperspectral Reflectance Imaging. Food Analytical Methods 8, 2635-2646.

Downey, G., 1998. Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. TrAC Trends in Analytical Chemistry 17, 418-424.

Eisenstecken, D., Stürz, B., Robatscher, P., Lozano, L., Zanella, A., Oberhuber, M., 2019. The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations. Postharvest Biology and Technology 147, 123-131.

Elmasry, G., Kamruzzaman, M., Sun, D.-W., Allen, P., 2012. Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review. Critical Reviews in Food Science and Nutrition 52, 999-1023.

Fan, G., Zha, J., Du, R., Gao, L., 2009. Determination of soluble solids and firmness of apples by Vis/NIR transmittance. Journal of Food Engineering 93, 416-420.

Fearn, T., 2000. On Orthogonal Signal Correction. Chemometr. Intell. Lab. 50, 47-52.

Fearn, T., 2001. Standardisation and calibration transfer for near infrared instruments: a review. J. Near Infrared Spectrosc. 9, 229-244.

Fernández-Ahumada, E., Fearn, T., Gómez, A., Vallesquino, P., Guerrero, J.E., Pérez-Marín, D., Garrido-Varo, A., 2008. Reducing NIR prediction errors with nonlinear methods and large populations of intact compound feedstuffs. Measurement Science and Technology 19, 085601.

Gaston, E., Frías, J.M., Cullen, P.J., O'Donnell, C.P., Gowen, A.A., 2010. Prediction of Polyphenol Oxidase Activity Using Visible Near-Infrared Hyperspectral Imaging on Mushroom (Agaricus bisporus) Caps. Journal of Agricultural and Food Chemistry 58, 6226-6233.

Geladi, P., McDougall, D. and Martens, H., 1985. Linearisation and Scatter Correction for Near Infrared Reflectance Spectra of Meat. Appl. Spectrosc. 39, 491-500.

Gorbe, E., Calatayud, A., 2012. Applications of chlorophyll fluorescence imaging technique in horticultural research: A review. Scientia Horticulturae 138, 24-35.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. Trends in Food Science & Technology 18, 590-598.

Gowen, A.A., O’Donnell, C.P., Taghizadeh, M., Gaston, E., O’Gorman, A., Cullen, P.J., Frias, J.M., Esquerre, C., Downey, G., 2008. Hyperspectral imaging for the investigation of quality deterioration in sliced mushrooms (Agaricus bisporus) during storage. Sensing and Instrumentation for Food Quality and Safety 2, 133-143.

Greensill, C.V., Wolfs, P.J., Spiegelman, H., Walsh, K.B., 2001. Calibration transfer between PDA-based NIR spectrometers in the NIR assessment of melon soluble solids content. Appl. Spectrosc. 55, 647-653.

Greensill, C.V., Walsh, K.B., 2002. Calibration transfer between miniature photodiode array-based spectrometers in the near infrared assessment of mandarin soluble solids content. J. Near Infrared Spectrosc. 10, 27-35.

Hadi, A.S., Ling, R.F., 1998. Some Cautionary Notes on the Use of Principal Components Regression. The American Statistician 52, 15-19.

Hayes, C.J., Walsh, K.B., Greensill, C.V., 2016. Improving calibration transfer between shortwave near infrared silicon photodiode array instruments. J. Near Infrared Spectrosc. 24, 59-68.

Helland, I.S., Naes, T., Isaksson, T., 1995. Related Versions of the Multiplicative Scatter Correction Methods for Pre-Processing Spectroscopic Data. Chemometr. Intell. Lab. 29, 233-241.

Herrero, A.M., 2008a. Raman spectroscopy a promising technique for quality assessment of meat and fish: A review. Food Chemistry 107, 1642-1651.

Herrero, A.M., 2008b. Raman Spectroscopy for Monitoring Protein Structure in Muscle Food Systems. Critical Reviews in Food Science and Nutrition 48, 512-523.

Huang, M., Zhu, Q., Wang, B., Lu, R., 2012. Analysis of hyperspectral scattering images using locally linear embedding algorithm for apple mealiness classification. Computers and Electronics in Agriculture 89, 175-181.

Hubert, M., Rousseeuw, P.J., Vanden Branden, K. 2005. ROBPCA: a new approach to robust principal component analysis. Technometrics 47, 64-79.

Indahl, U.G., Sahni, N.S., Kirkhus, B., Naes, T., 1999. Multivariate strategies for classification based on NIR-spectra – with applications to mayonnaise. Chemometr. Intell. Lab. 49, 19-31.

Isaksson, T., Kowalski, B., 1993. Piece-Wise Multiplicative Scatter Correction Applied to Near-Infrared Diffuse Transmittance Data from Meat Products. Appl. Spectrosc. 47,702-709.

Isaksson, T., Naes, T., 1990. Selection of Samples for Calibration in Near-Infrared Spectroscopy. Part II: Selection based on Spectral Measurements. Appl. Spectrosc., 44, 1152-1158.

Jacobs, B.A.J.G., Verlinden, B.E., Bobelyn, E., Decombel, A., Bleyaert, P., Van Lommel, J., Vandevelde, I., Saeys, W., Nicolai, B.M., 2016. Estimation of the prior storage period of lamb’s lettuce based on visible/near infrared reflectance spectroscopy. Postharvest Biology and Technology 113, 95-105.

Khatiwada, B.P., Subedi, P.P., Hayes, C., Cunha Carlos, L.C., Walsh, K.B., 2016. Assessment of internal flesh browning in intact apple using visible-short wave near infrared spectroscopy. Postharvest Biology and Technology 120, 103-111.

Karoui, R., Blecker, C., 2011. Fluorescence Spectroscopy Measurement for Quality Assessment of Food Systems—a Review. Food and Bioprocess Technology 4, 364-386.

Karoui, R., Downey, G., Blecker, C., 2010. Mid-Infrared Spectroscopy Coupled with Chemometrics: A Tool for the Analysis of Intact Food Systems and the Exploration of Their Molecular Structure−Quality Relationships − A Review. Chemical Reviews 110, 6144-6168.

Kemps, B., Saeys, W., Mertens, K., Darius, P., De Baerdemaeker, J., De Ketelaere, B., 2010. The importance of choosing the right validation strategy in inverse modelling. Journal of Near Infrared Spectroscopy, 18(4), 231-237.

Kennard, R.W., Stone, L.A., 1969. Computer Aided Design of Experiments. Technometrics, 11, 137-148.

Keresztes, J.C., Diels, E., Goodarzi, M., Nguyen-Do-Trong, N., Goos, P., Nicolai, B., Saeys, W., 2017. Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging. Postharvest Biology and Technology 130, 103-115.

Keresztes, J.C., Goodarzi, M., Saeys, W., 2016. Real-time pixel based early apple bruise detection using short wave infrared hyperspectral imaging in combination with calibration and glare correction techniques. Food Control 66, 215-226.

Khanmohammadi, M., Karami, F., Mir-Marqués, A., Bagheri Garmarudi, A., Garrigues, S., de la Guardia, M., 2014. Classification of persimmon fruit origin by near infrared spectrometry and least squares-support vector machines. Journal of Food Engineering 142, 17-22.

Kubelka, P., Munck, F., 1931. Ein Beitrag zur Optik Farbanstriche. 2. technische Physik 12, 593-604.

Li, J., Huang, W., Zhao, C., Zhang, B., 2013. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. Journal of Food Engineering 116, 324-332.

Li, Y.-S., Church, J.S., 2014. Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. Journal of Food and Drug Analysis 22, 29-48.

Liu, Y., Sun, X., Ouyang, A., 2010. Nondestructive measurement of soluble solid content of navel orange fruit by visible–NIR spectrometric technique with PLSR and PCA-BPNN. LWT - Food Science and Technology 43, 602-607.

López-Maestresalas, A., Keresztes, J.C., Goodarzi, M., Arazuri, S., Jarén, C., Saeys, W., 2016. Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. Food Control 70, 229-241.

Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O.L., Blasco, J., 2012. Recent Advances and Applications of Hyperspectral Imaging for Fruit and Vegetable Quality Assessment. Food and Bioprocess Technology 5, 1121-1142.

Lu, X., Rasco, B.A., 2012. Determination of Antioxidant Content and Antioxidant Activity in Foods using Infrared Spectroscopy and Chemometrics: A Review. Critical Reviews in Food Science and Nutrition 52, 853-875.

Luo, W., Huan, S., Fu, H., Wen, G., Cheng, H., Zhou, J., Wu, H., Shen, G., Yu, R., 2011. Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apple samples. Food Chemistry 128, 555-561.

Luo, X., Ye, Z., Xu, H., Zhang, D., Bai, S., Ying, Y., 2018. Robustness improvement of NIR-based determination of soluble solids in apple fruit by local calibration. Postharvest Biology and Technology 139, 82-90.

M. ElMasry, G., Nakauchi, S., 2016. Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality – A comprehensive review. Biosystems Engineering 142, 53-82.

Magwaza, L.S., Opara, U.L., Nieuwoudt, H., Cronje, P.J.R., Saeys, W., Nicolaï, B., 2012. NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit—A Review. Food and Bioprocess Technology 5, 425-444.

Manley, M., Joubert, E., Myburgh, L., Lotz, E., Kidd, M., 2007. Prediction of Soluble Solids Content and Post-Storage Internal Quality of Bulida Apricots Using near Infrared Spectroscopy. Journal of Near Infrared Spectroscopy 15, 179-188.

Martens, H., Naes, T., 1989. Multivariate Calibration. Wiley, Chichester.

Martens, H., Stark, E.J., 1991. Extended Multiplicative Signal Correction and Spectral Interference Subtraction: New Preprocessing Methods for Near Infrared Spectroscopy. Pharmaceut. Biomed Anal. 9, 625-635.

Martens, H., Nielsen, J.P., Engelsen, S.B., 2003. Light scattering and light absorbance separated by extended multiplicative signal correction: Application to near-infrared transmission of power mixtures. Anal. Chem. 75, 394-404.

McGlone, V.A., Jordan, R.B., Seelye, R., Clark, C.J., 2003. Dry-matter—a better predictor of the post-storage soluble solids in apples? Postharvest Biology and Technology 28, 431-435.

Mevik, B.H., Segtnan, V.H., Naes, T., 2004. Ensemble methods and partial least squares regression. Journal of Chemometrics 18: 498-507

Mevik, B.-H., Wehrens, R., 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. Journal of Statistical Software, 18(2), 1-24.

Miller, W.M., Throop, J.A., Upchurch, B.L., 1998. Pattern recognition models for spectral reflectance evaluation of apple blemishes. Postharvest Biology and Technology 14, 11-20.

Mukarev, M.I., Walsh K.B., 2012. Prediction of brix values of intact peaches with least squares-support vector machine regression models. J. Near Infrared Spectrosc. 20, 647-655.

Naes, T.I., T.; Fearn, T.; and Davies, T., 2002. A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK.

Naes, T., Isaksson, T., Kowalski, B.R., 1990. Locally Weighted Regression and scatter correction for near-infrared reflectance data, Analytical Chemistry 62, 664–673.

Nguyen-Do-Trong;, N., Nicolaï;, B., Wouter Saeys, 2018. NIRS is ripe for use in horticulture, The Handbook of Near-Infrared Analysis, 4th ed. CRC Press.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007a. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46, 99-118.

Nicolaï, B.M., Theron, K.I., Lammertyn, J., 2007b. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. Chemometrics and Intelligent Laboratory Systems 85, 243-252.

Nicolaï, B.M., Verlinden, B.E., Desmet, M., Saevels, S., Saeys, W., Theron, K., Cubeddu, R., Pifferi, A., Torricelli, A., 2008. Time-resolved and continuous wave NIR reflectance spectroscopy to predict soluble solids content and firmness of pear. Postharvest Biology and Technology 47, 68-74.

Noh, H.K., Lu, R., 2007. Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. Postharvest Biology and Technology 43, 193-201.

Norris, K.H., Williams, P.C., 1984. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat: I. Influence of particle size. Cereal Chemistry, 61, 158-165.

Park, B., A. Abbott, J., J. Lee, K., H. Choi, C., H. Choi, K., 2003. Near-infrared Diffuse Reflectance for Quantitative and Qualitative Measurement of Soluble Solids and Firmness of Delicious and Gala Apples. Transactions of the ASAE 46, 1721-1731.

Peirs, A., Lammertyn, J., Ooms, K., Nicolaı̈, B.M., 2001. Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. Postharvest Biology and Technology 21, 189-199.

Peirs, A., Schenk, A., Nicolaı̈, B.M., 2005. Effect of natural variability among apples on the accuracy of VIS-NIR calibration models for optimal harvest date predictions. Postharvest Biology and Technology 35, 1-13.

Perez-Marin, D., Garrido-Varo, A., de Pedro, E., Guerrero-Ginel, J.E., 2007. Chemometric utilities to achieve robustness in liquid NIRS calibrations: Application to pig fat analysis.

Pu, Y.-Y., Feng, Y.-Z., Sun, D.-W., 2015. Recent Progress of Hyperspectral Imaging on Quality and Safety Inspection of Fruits and Vegetables: A Review. Comprehensive Reviews in Food Science and Food Safety 14, 176-188.

Rady, A., Guyer, D., Lu, R., 2015. Evaluation of Sugar Content of Potatoes using Hyperspectral Imaging. Food and Bioprocess Technology 8, 995-1010.

Rodriguez-Saona, L.E., Allendorf, M.E., 2011. Use of FTIR for Rapid Authentication and Detection of Adulteration of Food. Annual Review of Food Science and Technology 2, 467-483.

Roger, J.M., Chauchard, F., Bellon-Maurel, V., 2003. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemometrics and intelligent laboratory systems 66, 191-204.

Sádecká, J., Tóthová, J., 2007. Fluorescence spectroscopy and chemometrics in the food classification − a review. Czech J. Food Sci. 25, 159-174.

Saeys, W., Velazco-Roa, M.A., Thennadil, S.N., Ramon, H., Nicolaï, B.M., 2008. Optical properties of apple skin and flesh in the wavelength range from 350 to 2200 nm. Appl. Opt. 47, 908-919.

Salguero-Chaparro, L., Palagos, B., Pena-Rodriguez, F., Roger, J.M., 2013. Calibration transfer of intact olive NIR spectra between pre-dispersive instrument and portable spectrometer. Computers and Electronics in Agriculture 96, 202-208.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least-Squares Procedures. Anal. Chem. 36, 1627-1639.

Shen, G., Lesnoff, M., Baeten, V., Dardenne, P., Davrieux, F., Ceballos, H., Belalcazar, J., Dufour, D., Yang, Z., Han, L., Fernandez Pierna, J.A., 2019. Local partial least squares based on global PLS scores. Journal of Chemometrics 33:e3117.

Shenk, J.S., Westerhaus, M.O., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments, Journal of Near Infrared Spectroscopy, 5, 223–232.

Snee, R., 1977. Validation of regression models: methods and examples. Technometrics 19: 415-428.

Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. Least Squares Support Vector Machines. World Scientific Publishing Co.: Singapore.

Swierenga, H., Wülfert, F., de Noord, O.E., de Weijer, A.K., Buydens, L.M.C., 2000. Development of robust calibration models in near-infrared spectrometric applications. Anal. Chim. Acta 411, 121-135.

Szymanska-Chargot, M., Chylinska, M., Kruk, B., Zdunek, A., 2015. Combining FT-IR spectroscopy and multivariate analysis for qualitative and quantitative analysis of the cell wall composition changes during apples development. Carbohydrate Polymers 115, 93-103.

Thennadil, S.N., Martens, H., Kohler, A., 2006. Physics-based multiplicative scatter correction approaches for improving the performance of calibration models. Appl. Spectrosc. 60, 315-321.

Tian, H.-q., Ying, Y.-b., Lu, H.-s., Fu, X.-p., Yu, H.-y., 2007. Measurement of soluble solids content in watermelon by Vis/NIR diffuse transmittance technique. Journal of Zhejiang University SCIENCE B 8, 105-110.

Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). J Chemom. 16, 119–128.

Trygg, J., 2002. O2-PLS for qualitative and quantitative analysis in multivariate calibration. J. Chemometrics 16, 283-293.

Van Beers, R., Aernouts, B., Leon Gutierrez, L., Erkinbaev, C., Rutten, K., Schenk, A., Nicolai, B., Saeys, W., 2015. Optimal Illumination-Detection Distance and Detector Size for Predicting Braeburn Apple Maturity from Vis/NIR Laser Reflectance Measurements. Food and Bioprocess Technology 8 (10), 2123-2136. doi: 10.1007/s11947-015-1562-4.

Van Beers, R., Aernouts, B., Watté, R., Schenk, A., Nicolai, B., Saeys, W., 2017. Effect of maturation on the bulk optical properties of apple skin and cortex in the 500 to 1850 nm wavelength range. Journal of Food Engineering 214, 79-89.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc.

Wang, H., Peng, J., Xie, C., Bao, Y., He, Y., 2015. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. Sensors 15.

Wang, N.-N., Sun, D.-W., Yang, Y.-C., Pu, H., Zhu, Z., 2016. Recent Advances in the Application of Hyperspectral Imaging for Evaluating Fruit Quality. Food Analytical Methods 9, 178-191.

Wang, Y., Veltkamp, D.J., Kowalski, B., 1991. Multivariate instrument standardization. Anal. Chem. 63, 2750-2756.

Wei, X., Liu, F., Qiu, Z., Shao, Y., He, Y., 2014. Ripeness Classification of Astringent Persimmon Using Hyperspectral Imaging Technique. Food and Bioprocess Technology 7, 1371-1380.

Westerhaus, M.O., 1990. Improving the repeatability of NIR calibrations across instruments. In Third International Conference on Near Infrared Spectroscopy, Brussels, Belgium.

Wilson, R.H., Tapp, H.S., 1999. Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods. TrAC Trends in Analytical Chemistry 18, 85-93.

Windig, W., Shaver, J., Bro, R., 2008. Loopy MSC: A simple way to improve multiplicative scatter correction. Appl. Spectrosc. 62, 1153-1159.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S., 2006. PLS\_Toolbox Version 4.0 for use with MatlabTM, Eigenvector Research Inc., Wenatchee, WA. ISBN: 0-9761184-1-6.

Wold, S., Antti, H., Lindgren, F., Ohman, J., 1998. Orthogonal signal correction of near-infrared spectra. Chemometr Intell Lab. 44, 175–185.

Wold, S., Sjöström, M., & Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

Workman, J.J.Jr., 2018. A review of calibration transfer practices and instrument differences in spectroscopy. Appl. Spectrosc. 72, 340-365.

Wu, D., Sun, D.-W., 2013a. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals. Innovative Food Science & Emerging Technologies 19, 1-14.

Wu, D., Sun, D.-W., 2013b. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part II: Applications. Innovative Food Science & Emerging Technologies 19, 15-28.

Xing, J., Saeys, W., De Baerdemaeker, J., 2007. Combination of chemometric tools and image processing for bruise detection on apples. Computers and Electronics in Agriculture 56, 1-13.

Xiong, Z., Sun, D.-W., Zeng, X.-A., Xie, A., 2014. Recent developments of hyperspectral imaging systems and their applications in detecting quality attributes of red meats: A review. Journal of Food Engineering 132, 1-13.

Xu, M.-L., Gao, Y., Han, X.X., Zhao, B., 2017. Detection of Pesticide Residues in Food Using Surface-Enhanced Raman Spectroscopy: A Review. Journal of Agricultural and Food Chemistry 65, 6719-6726.

Yang, D., Ying, Y., 2011. Applications of Raman Spectroscopy in Agricultural Products and Food Analysis: A Review. Applied Spectroscopy Reviews 46, 539-560.

Zhang, C., Guo, C., Liu, F., Kong, W., He, Y., Lou, B., 2016. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. Journal of Food Engineering 179, 11-18.

Zheng, J., He, L., 2014. Surface-Enhanced Raman Spectroscopy for the Chemical Analysis of Food. Comprehensive Reviews in Food Science and Food Safety 13, 317-328.