Talanta 223 (2021) 121733

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit

Puneet Mishra [a,*], Federico Marini [b], Bastiaan Brouwer [a], Jean Michel Roger [c,d], Alessandra Biancolillo [e], Ernst Woltering [a,f], Esther Hogeveen-van Echtelt [a]

[a] Wageningen Food and Biobased Research, Bornse Weilanden 9, P.O. Box 17, 6700AA, Wageningen, the Netherlands
[b] Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185, Rome, Italy
[c] ITAP, INRAE, Institut Agro, University Montpellier, Montpellier, France
[d] ChemHouse Research Group, Montpellier, France
[e] Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, 67100, Coppito, L'Aquila, Italy
[f] Horticulture and Product Physiology Group, Wageningen University, Droevendaalsesteeg 1, P.O. Box 630, 6700AP, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Multi-block data analysis
Sequential data fusion
Chemometrics
Multivariate analysis
Miniature near infrared (NIR) spectrometers
Pear (*Pyrus communis* L.)

ABSTRACT

Near infrared (NIR) spectroscopy allows rapid estimation of quality traits in fresh fruit. Several portable spectrometers are available in the market as a low-cost solution to perform NIR spectroscopy. However, portable spectrometers, being lower in cost than a benchtop counterpart, do not cover the complete near infrared (NIR) spectral range. Often portable sensors either use silicon-based visible and NIR detector to cover 400–1000 nm, or InGaAs-based short wave infrared (SWIR) detector covering the 900–1700 nm. However, these two spectral regions carry complementary information, since the 400–1000 nm interval captures the color and 3rd overtones of most functional group vibrations, while the 1st and the 2nd overtones of the same transitions fall in the 1000–1700 nm range. To exploit such complementarity, sequential data fusion strategies were used to fuse the data from two portable spectrometers, i.e., Felix F750 (~400–1000 nm) and the DLP NIR Scan Nano (~900–1700 nm). In particular, two different sequential fusion approaches were used, namely sequential orthogonalized partial-least squares (SO-PLS) regression and sequential orthogonalized covariate selection (SO-CovSel). SO-PLS improved the prediction of moisture content (MC) and soluble solids content (SSC) in pear fruit, leading to an accuracy which was not obtainable with models built on any of the two spectral data set individually. Instead, SO-CovSel was used to select the key wavelengths from both the spectral ranges mostly correlated to quality parameters of pear fruit. Sequential fusion of the data from the two portable spectrometers led to an improved model prediction (higher R^2 and lower RMSEP) of MC and SSC in pear fruit: compared to the models built with the DLP NIR Scan Nano (the worst individual block) where SO-PLS showed an increase in R^2_p up to 56% and a corresponding 47% decrease in RMSEP. Differences were less pronounced to the use of Felix data alone, but still the R^2_p was increased by 2.5% and the RMSEP was reduced by 6.5%. Sequential data fusion is not limited to NIR data but it can be considered as a general tool for integrating information from multiple sensors.

1. Introduction

Pear (*Pyrus communis* L.) fruit are widely grown across the world. Key quality parameters such as moisture content (MC) and soluble solids content (SSC) are used to decide on an optimal harvest timing and to monitor fruit quality during the storage operations [1–4]. Similarly, for a variety of quality parameters in fresh fruit, destructive as well as non-destructive measurements (usually with instruments employing near infrared (NIR) spectroscopy) are widely used [5–12]. A drawback to the destructive measurements is that they are time consuming and involve a multi-step procedure, requiring the use of multiple instruments, such as refractometer, hot air oven dryer, weighing balance etc. In addition, destructive measurements should be done on selected samples of harvested fruit and as such lead to production losses as an economic factor and food loss as a sustainability/societal problem. NIR spectroscopy is a non-destructive technique that can be applied to

* Corresponding author.
E-mail address: puneet.mishra@wur.nl (P. Mishra).

https://doi.org/10.1016/j.talanta.2020.121733
Received 24 August 2020; Received in revised form 28 September 2020; Accepted 1 October 2020
Available online 13 October 2020

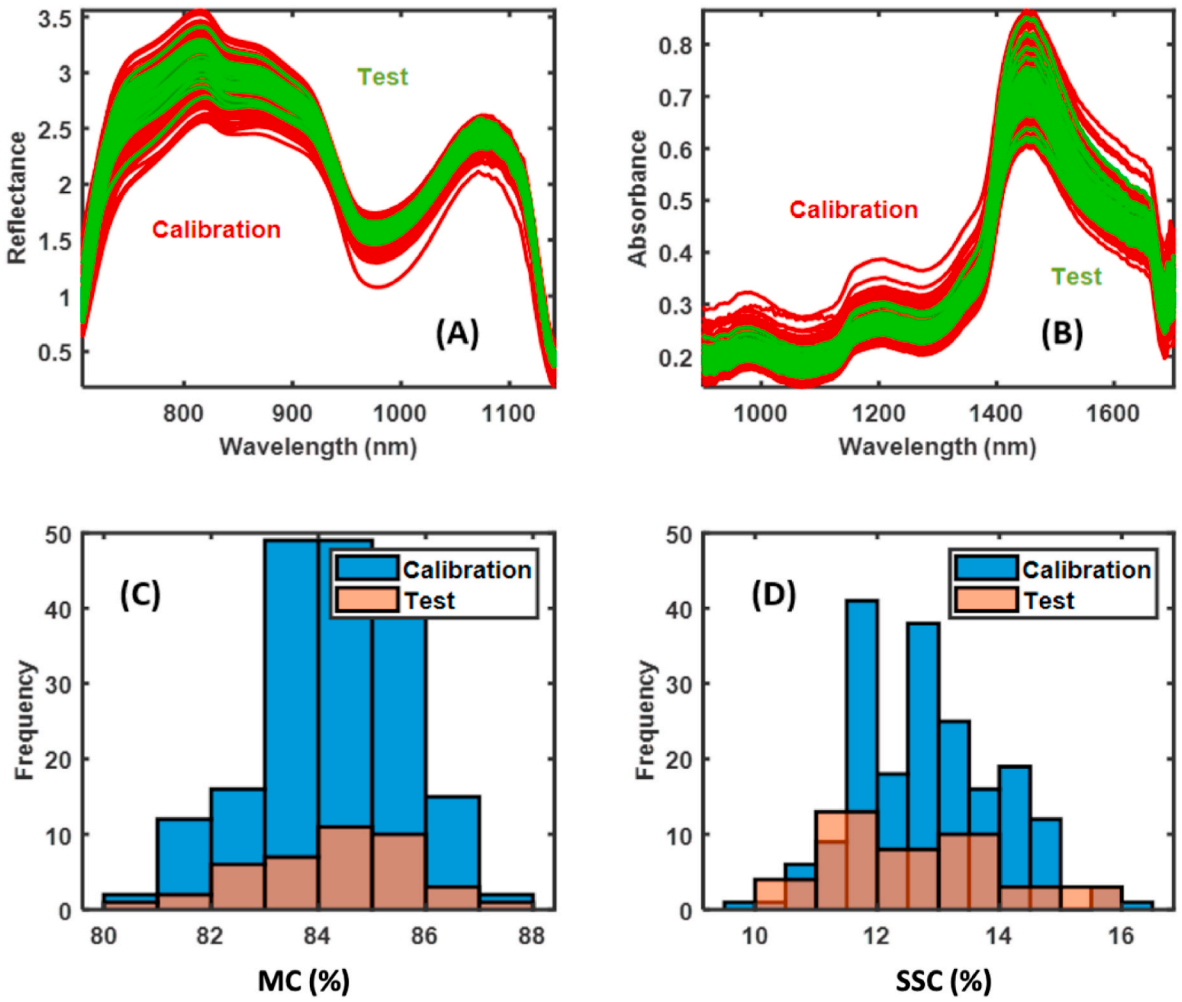

Fig. 1. Graphical representation of the spectroscopic data and of the distribution of the reference measurements for the training and test samples. (A) Spectra for calibration (red) and test set (green) from Felix, (B) spectra for calibration (red) and test set (green) from DLP NIR scan Nano, (C) histogram representing calibration (blue) and test set (orange) for moisture content (MC %), and (D) histogram representing calibration (blue) and test set (orange) for soluble solids content (SSC %). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

plant-attached fruit [13]. However, conventional diffuse NIR spectroscopy does not provide an absolute value and a calibration step is required to quantify the property of interest, which is always associated with a prediction error [14–16]. Prediction error can be high or low depending on the sensitivity of the instrument, which in turn is related to the spectral range investigated, and on the calibration model. Furthermore, in the case of fresh fruit several external factors such as cultivars, season and temperature of the sensor may also affect the model performances.

Portable NIR spectrometers are commonly used to analyze the quality properties of fresh fruit [3,17–22]. However, a main limitation with portable NIR spectrometers is that almost all of them are not designed to cover a wide NIR spectral range. Indeed, the portable spectrometers most commonly available on the market operate either in the 400–1000 nm or in the 900–1700 nm spectral ranges. This is because different optical detectors are required to measure the different spectral regions efficiently. In the case of the 400–1000 nm range (visible and near-infrared, VIS-NIR), silicon (Si)-based detectors are commonly used while InGaAs detectors covers the 900–1700 nm (short wave infrared, SWIR) [23]. Typically, since InGaAs detectors are far more expensive, Si-based detectors are the preferred choice for the portable spectrometers as they are lower cost. Si-based detectors can be used to calculate spectral indices to show changes in plant pigments in the VIS region while in the 700–1000 nm spectral range the MC, sugar, fat or protein content in fresh produce can be correlated to the 3rd overtones of the OH, CH and NH bonds [24]. This is because the 700–1000 nm spectral range captures the 3rd overtones related to the OH, CH and NH bonds which can be correlated to different chemical constituents [1,25]. Moreover, radiation in the 400–1000 nm range has a greater penetration depth than that of higher wavelengths [26]. However, a drawback of using the VIS-NIR region is that the NIR part (700–1000 nm) is characterized by the presence of weak and highly overlapping signals (3rd overtones) which makes model building and model optimization/selection (for instance, to choose the optimal complexity in partial least-squares regression) a challenging task. On the other hand, the spectral range >1000 nm contains the signals corresponding to the 1st and the 2nd overtones, which appear as comparatively less-overlapped and stronger bands, making model optimization easier. However, at wavelengths >1000 nm the penetration depth is less compared to the VIS-NIR range and this could be a limitation when analyzing fresh fruit [26]. Each of the two spectral ranges (400–1000 nm and 900–1700 nm) therefore has its own distinct benefits as well as disadvantages, but in general, they carry complementary information. Accordingly, the best way to exploit such complementarity (e.g. to achieve improved predictions) would be to integrate/combine information from both the spectral ranges.

Integration of multiple sensors to predict fruit quality parameters is gaining attention [27–29] as fruit quality has many different dimensions

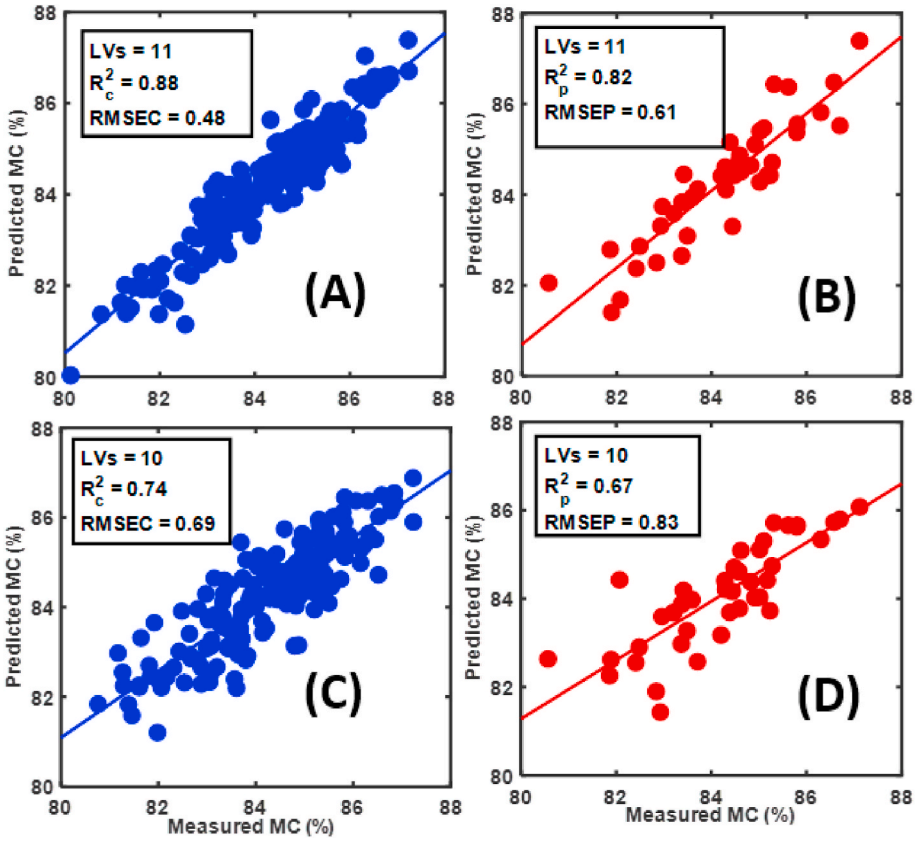

Fig. 2. Partial least-squares (PLS) modelling for the calibration of moisture content (MC %) based on individual spectral matrices. Model built on data from Felix handheld spectrometer: Results on the calibration (A) and test (B) sets. Models built on data from DLP NIR Scan Nano device; Results on the calibration (C) and test (D) sets.

and no single non-destructive sensor can perfectly explain a range of fruit properties. In combination, multiple sensors can enhance model performance and in the same way combing multiple spectral ranges can lead to better and more accurate models [30]. To take maximum advantage from the combination of data from multiple spectral sensors, it is necessary to use processing strategies collectively referred to as multi-block data analysis [31–36]. In particular, a highly effective technique to integrate data from multiple sources that is increasingly used is sequential and orthogonalized partial-least squares regression (SO-PLS) [37]. SO-PLS is based on the sequential extraction of the information from the data of different sensors. SO-PLS was successfully applied to fuse spectroscopic data from different instruments to provide models with significantly better performance than models based on the data from a single sensor [38–40]. The concept of sequential modelling coupled to orthogonalization can also be used to identify a minimum set of variables (wavelengths) within each block, that are relevant to accurately predict the response(s) of interest. This approach is called sequential and orthogonalized covariance selection (SO-CovSel) as recently proposed in the literature [41]. SO-CovSel is a multi-block variable selection technique [41], which extends the covariance selection (CovSel) approach for feature reduction [42] to a multi-block scenario, borrowing the concepts of sequential inclusion of the matrices to be fused after orthogonalization from SO-PLS.

Thus, the aim of the present work was to demonstrate the use of the sequential data fusion strategies SO-PLS and SO-CovSel to integrate spectra from two portable spectrometers operating in different NIR ranges, in order to predict two quality attributes (MC and SSC) of pear fruit. The two portable spectrometers were the Felix handheld spectrometer and the DLP NIR Scan Nano.

2. Materials and methods

2.1. Samples and reference measurements

Two hundred forty pear samples coming from 11 different orchards (20 fruits each from 10 orchards, and 40 fruits from the remaining one) were used in the study. The pear fruits were measured immediately after harvest. Additional description of the samples and the measurements can be found in [4]. NIR spectroscopy measurements were recorded on the intact pear; afterwards, a cylindrical disc was cut at the position of the largest diameter of the fruit and it was further divided into four equal quadrants. The two sub-samples withdrawn from the same side of the pear on which the NIR spectra were collected, were used for the quantification of MC and SSC, using the reference methods. In particular, the determination of MC was conducted by weighing (XS10001L, Mettler-Toledo GmbH, Giessen, Germany) the subsample before and after hot-air oven drying (at 80 °C for 24 h with FP 720, Binder GmbH, Tuttlingen, Germany), while SSC was assessed measuring the refractive index of the juice extracted from the other sub portion, with a handheld refractometer (HI 96801, Hanna Instruments Inc, Woonsocket, RI, USA).

2.2. Data acquisition

2.2.1. Felix handheld spectrometer

The VIS-NIR spectral measurements were carried out with a portable spectrometer Felix F-750 (Camas, WA, USA), which utilizes a Carl Zeiss MMS-1 spectrometer to record the reflected light in the spectral range 310–1135 nm. Illumination is provided by a xenon tungsten lamp, with a built in white painted reference standard used to calibrate each scan. Data acquisition was performed by placing the fruit belly directly on the optical window support of the instrument and manually pressing the

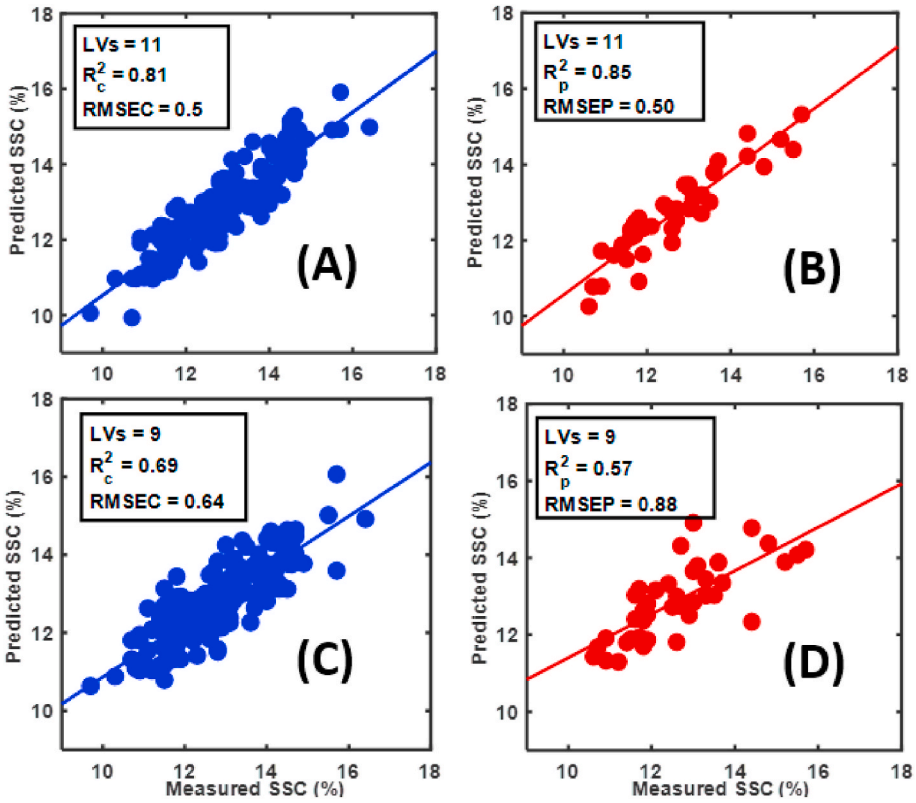

Fig. 3. Partial least-squares (PLS) modelling for the calibration of soluble solids content (SSC %) based on individual spectral matrices. Model built on data from Felix handheld spectrometer: Results on the calibration (A) and test(B) sets. Models built on data from DLP NIR Scan Nano device; Results on the calibration (C) and test (D) sets.

scan button on the Felix device: each recorded spectrum was the result of the average of 6 consecutive scans. The data were radiometrically calibrated with the in-built white reference standard. The data were automatically normalized by the built-in software of the Felix device and the raw reflectance spectra were extracted as excel files utilizing the "Data-Viewer" software from Felix (Camas, WA, USA). Due to the presence of high noise at extreme wavelengths, the spectral range was reduced from 310-1135 nm to 400–1135 nm.

2.2.2. DLP NIR Scan Nano

The SWIR measurements (900–1700 nm) were performed with a DLP NIR Scan Nano (Texas Instrument, USA). On each fruit, the SWIR spectrum was acquired immediately after collecting the VIS-NIR spectrum and exactly on the same position, by directly placing the sample on the sensor head of the DLP NIR Scan Nano. The measurements performed were in diffuse reflectance mode. The data were radiometrically calibrated with the in-built mathematical transformation for estimating the absorbance. The output of the sensor were the absorbance spectra. Spectral sampling resolution was 6.35 nm and the integration time was 8 ms. Each recorded spectrum was the result of the average of 5 consecutive scans. The measurements were controlled using the PC software provided, which automatically corrects for the dark and white references and outputs the absorbance measurements.

2.3. Data processing

Data processing involved four main steps, i.e., spectral pre-processing, building calibration models for the prediction of MC and SSC with the separate spectral matrices using standard PLS regression, verifying whether the joint processing of both spectral matrices through sequential and orthogonalized PLS (SO-PLS) regression could lead to better results than those obtained with single data blocks and identifying which spectral variables in each wavelength range supplied relevant and complementary information for the prediction of the two quality attributes through the use of SO-CovSel. All the analyses were performed using MATLAB 2017b (The Mathworks, Natick, MA, USA).

2.3.1. Spectral pre-processing

In the case of VIS-NIR data, the spectral range was reduced and restricted to the NIR bands (700–1135 nm), in order to remove the possible influence of fruit color, but retaining the wavelength interval corresponding to the 3rd overtones of OH and CH bonds, likely to be related to MC and SSC. The full range (900–1700 nm) provided by the DLP NIR Scan Nano was used. Then, the same combination of pre-processing was applied to both spectral data sets. Firstly, spectra were smoothed with a Savitzky-Golay filter (2nd order interpolating polynomial and 15 points window). Later, to remove the effects of light scattering, variable sorting for normalization (VSN) technique [43], which can be considered as a weighted analogue of the standard normal variate (SNV) transform, was used.

2.3.2. Partial least-squares regression

Partial least-squares (PLS) regression is probably the most commonly used chemometric technique for calibration, especially when dealing with NIR spectroscopic data [44,45]. This is at least partly due to its effectiveness in dealing with multi-collinearities or ill-conditioned data matrices in general, which derives from its use of latent variables (LVs) instead of original spectral intensities as predictors. Indeed, PLS compresses the relevant spectroscopic variation into a reduced number of orthogonal scores, obtained by projecting the data onto directions of maximum covariance with the response variables. In the present work, the quality of the regression model was evaluated through the calculation of the coefficient of determination (R2) and root mean squared error (RMSE), while the optimal number of LVs were using 10-fold cross

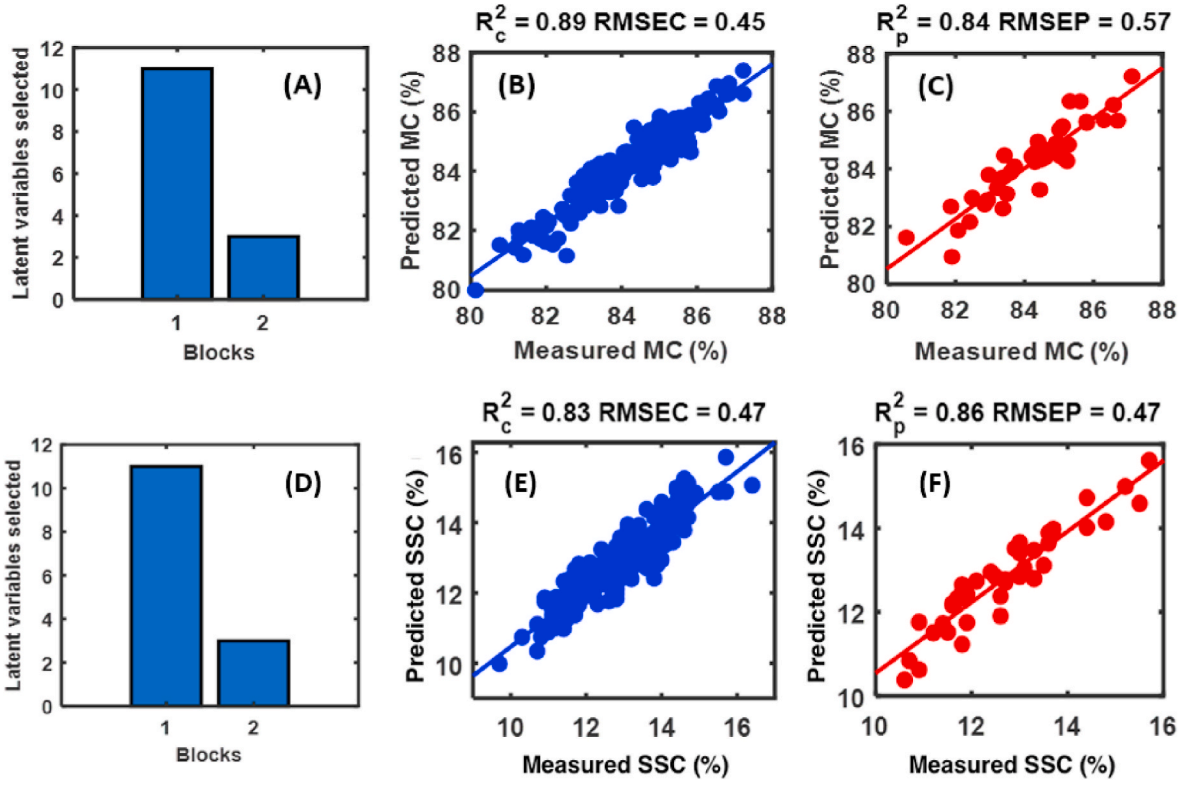

Fig. 4. SO-PLS fusion of Felix and DLP NIR Scan Nano data for the prediction of moisture content (MC %) and soluble solids content (%). MC: (A). Number of LVs extracted from each data block. Results on the calibration (B) and validation (C) sets. SSC: (D). Number of LVs extracted from each data block. Results on the calibration (E) and validation (F) sets. R^2 explain the coefficient of determination and RMSEC and RMSEP are the root mean squared error for calibration and prediction respectively.

validation procedure with LVs carrying minimum RMSE. The PLS regression analysis was implemented on the free open-source MATLAB based chemometrics toolbox called MBA-GUI [46].

2.3.3. Sequential orthogonalized partial least-squares regression

Sequential and orthogonalized PLS (SO-PLS) belongs to the family of multi-block PLS-based approaches. The main ideas behind the method are the sequential incorporation of the predictor matrices. SO-PLS assesses the incremental contributions of the different blocks, and the orthogonalization step removes redundant information [37]. In the simplest multi-block scenario, with two predictor matrices, SO-PLS involves the following steps: at first, a PLS regression model between the data from the first sensor ($\mathbf{X}_1$) and the responses $\mathbf{Y}$ is calculated, yielding, among other outcomes, the scores $\mathbf{T}_1$. Then, the data from the second sensor ($\mathbf{X}_2$) and the $\mathbf{Y}$ block are orthogonalized with respect to $\mathbf{T}_1$ and a second PLS model is fitted between the Y-residuals and the orthogonalized second block of predictors (if the number of blocks is larger than two, these steps are iterated until the last predictor matrix is incorporated). Eventually, all the block scores (here $\mathbf{T}_1$ and $\mathbf{T}_2$) are concatenated and a final ordinary least-squares regression model is calculated between the concatenated scores and the response(s). Here it is worth stressing once again that one of the main benefits of SO-PLS is its sequential nature which, together with the orthogonalization step, allows a straightforward evaluation of whether incorporating a new block really brings additional information, to significantly improve the model quality, or not. The SO-PLS regression analysis was implemented on the free open-source MATLAB based chemometrics toolbox called MBA-GUI [46].

2.3.4. Sequential and orthogonalized covariance selection

In this case, explanation of the basic steps of the So-CovSel algorithm will refer to the simplest multi-block scenario (which corresponds to the data discussed in the present paper), i.e. the situation where two blocks of predictors are used to model the response(s) $\mathbf{Y}$. CovSel is first used to extract the relevant variables from the first block; then the second block and the responses are orthogonalized with respect to the selected predictors and a second round of CovSel is operated between the orthogonalized second matrix and the Y residuals to identify another subset of relevant variables. Eventually, all the selected wavelengths are combined into a single predictor matrix $\mathbf{X}_{sel}$, and the final predictive model is built by correlating the response matrix with $\mathbf{X}_{sel}$ through ordinary least-squares regression. The SO-CovSel regression analysis was implemented on the free open-source MATLAB based chemometrics toolbox called MBA-GUI [46].

3. Results

3.1. Data description

For the Felix device, data are output as reflectance (Fig. 1A), while the DLP NIR Scan Nano directly returns absorbance values (Fig. 1B). The spectra were partitioned into calibration and test set, which are highlighted in the Figure in red and green, respectively, using the Kennard-Stone algorithm [47]. At both wavelength ranges, the profiles present broad but identifiable bands, ascribable to the contributions of the main constituents of the food matrix such as water and sugar. For both properties, the distributions of the reference values in the calibration and the test set were comparable: indeed; in the case of MC (Fig. 1C), the calibration and test set reference values were 84.26 ± 2.74% and 84.22 ± 2.86%, respectively; for SSC (Fig. 1D), the calibration and test set reference values were 12.76 ± 2.56% and 12.65 ± 2.29%, respectively.

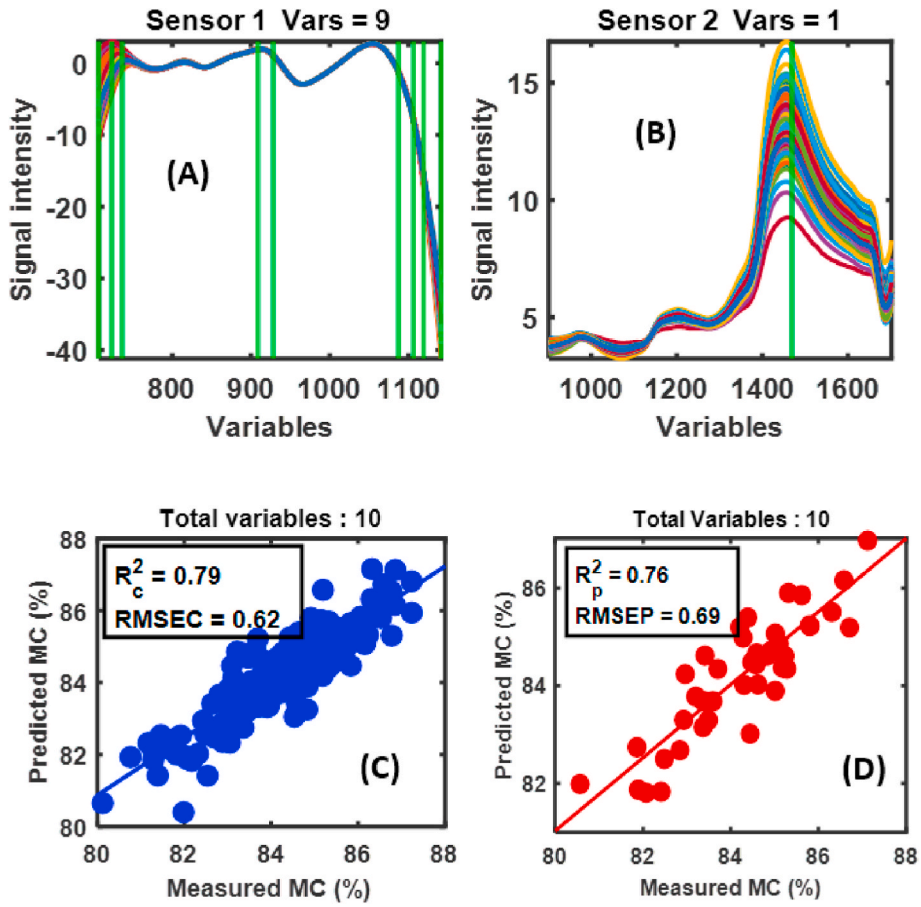

Fig. 5. SO-CovSel multi-block wavelength selection for the prediction of moisture content (MC %). Wavelengths selected from (A). Felix, (B). DLP NIR Scan Nano. Prediction results on the calibration (C) and test (D) sets. R^2 explain the coefficient of determination and RMSEC and RMSEP are the root mean squared error for calibration and prediction respectively.

3.2. Partial least-squares regression on individual spectroscopic matrices

PLS models were first separately built on the individual data matrices resulting from each of the two spectroscopic sensors (Felix and DLP NIR Scan Nano). The models for the prediction of MC had an optimal complexity of 11 LVs in the case of Felix data and of 10 LVs when built on spectra recorded by DLP NIR Scan Nano; the corresponding results are graphically displayed in Fig. 2. Both models provide good MC prediction results, the one built on the Felix data shows a 20.5% higher R^2_p and a 25% lower RMSEP when compared to those obtained from the DLP NIR Scan Nano. For the prediction of SSC, the models built on the spectra recorded on the Felix or on the DLP NIR Scan Nano devices had an optimal complexity of 11 and 9 LVs, respectively; the corresponding results are shown in Fig. 3. The SSC model obtained using the Felix spectra showed a 56% higher R2p and a 44% lower RMSEP when compared to those based on the DLP NIR Scan Nano data.

3.3. Data fusion with SO-PLS

The results reported in the previous section show that good models for the prediction of MC and SSC can be obtained by calibrating NIR data (particularly the Felix ones) with single-block PLS regression. A multi-block approach based on sequential and orthogonalized partial least-squares regression (SO-PLS) was used to integrate the complementary data from the two spectroscopic sensors to try to achieve better calibration performance and also a deeper understanding/interpretation of the models. When fusing the data matrices, one of the advantages of the SO-PLS method is that it is scale-invariant, meaning that no scaling of the blocks is needed prior to integration. The order in which the blocks are introduced into the model can be relevant and, in the present case, the optimal choice was found to be Felix data first and DLP NIR Scan Nano afterwards. The results of SO-PLS multi-block analysis for the calibration of MC and SSC are shown in Fig. 4, respectively. In both predictive models, most of the scores are extracted from the Felix block, which was the one giving the better performance when used individually. The sequential multi-block approach showed better predictive ability with higher R^2_p and lower RSMEP values when compared to building calibration models on individual data matrices, thus confirming that integrating the information from the two spectral matrices substantially improves the regression performance. For MC prediction, compared to the models built on the Felix or DLP NIR Scan Nano data individually, R^2_p were increased by 2.5% and 23%, respectively and RMSEP were reduced by 6.5 and 30%, respectively. Similarly, for SSC prediction, compared to the models built on the Felix or DLP NIR Scan Nano data individually, the RMSEP were reduced by 6% and 47%, respectively. R^2_p improved in the case of the DLP NIR scan Nano (by 56%) but not in the case of Felix.

3.4. Wavelength selection with SO-CovSel

The use of SO-PLS demonstrated that the integration of the information from two spectroscopic ranges could lead to an improvement in the accuracy of the predictions of quality attributes in pears. To gain further insight into the variables mostly correlated with the responses, the recently proposed SO-CovSel technique was employed. The main results of SO-CovSel in the prediction of MC and SSC are graphically summarized in Figs. 5 and 6. In general, the predictive accuracy is lower than that obtained with SO-PLS (i.e. using all the spectral variables), but

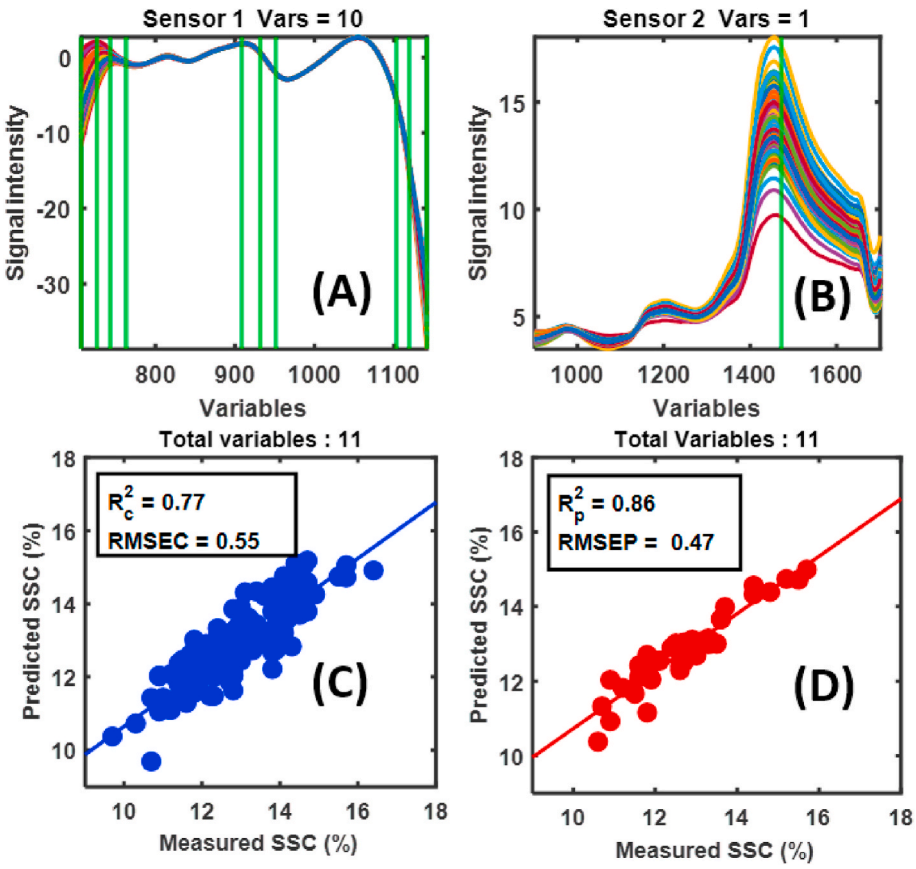

Fig. 6. SO-CovSel multi-block wavelength selection for the prediction of soluble solids content (SSC %). Wavelengths selected from (A). Felix, (B). DLP NIR Scan Nano. Prediction results on the calibration (C) and test (D) sets. R^2 explain the coefficient of determination and RMSEC and RMSEP are the root mean squared error for calibration and prediction respectively.

still quite good. In both cases the number of selected variables is rather low, confirming the effectiveness of the SO-CovSel approach in identifying parsimonious subsets of relevant and non-redundant predictors. In particular, 10 wavelengths (9 from Felix + 1 from DLP NIR Scan Nano) were selected for the prediction of MC, and 11 wavelengths (10 from Felix + 1 from DLP NIR Scan Nano) for calibrating SSC. These findings are highly consistent with the results discussed in the previous sections, indicating that the Felix block is the one carrying a higher amount of information correlated with the responses. The fact that at least 1 variable was selected from the DLP NIR Scan Nano block indicates that the corresponding spectral range carries some complementary information which can be integrated with the Felix data. Indeed, the results highlight that the prediction of MC can be improved by adding the absorbance at around 1450 nm collected with the DLP NIR Scan Nano (corresponding to the 2nd overtone of water), to the 9 variables selected in the VIS-NIR range (Fig. 5). Similar considerations can be made for the prediction of SSC, where the SO-CovSel model selected a single predictor corresponding to the signal at around 1500 nm (interpreted as the 2nd overtone of *C*–H bonds) from the DLP NIR Scan Nano block and directly correlated to the SSC. Moreover, since SO-CovSel provides a highly parsimonious wavelength selection, retaining at the same time a good predictive ability, its results could be exploited to design and build lower cost multispectral systems as it clearly indicates a discrete set of the most informative wavelengths.

4. Discussion

In the present work, data from two portable spectrometers (Felix and DLP NIR Scan Nano) were sequentially fused for improving the prediction of MC and SSC in individual pear fruit. The results show that both spectral ranges carry complementary information in terms of functional group overtones (OH and CH) responsible for predicting MC and SSC content in pear fruit. SO-PLS was able to extract complementary LVs from each spectral range and resulted in model with improved prediction performance (higher R^2_p and lower RMSEP) compared to those obtained on the individual blocks. A major strength of the approach is that the sequential strategy first modelled the data from the Felix spectrometer and then, once it could not find any further improvement, it explored the DLP NIR Scan Nano data. In the real case scenarios, this approach is practical as it allows users to understand if there is a clear need of an additional sensor or if a single sensor is sufficient to predict the traits of interest. Moreover, since the sequential approach requires the order of the sensors to be defined [37], this means that the user can first test all the lower cost sensor options available and keep the costlier or more sophisticated sensors for the end. The sequential approach will explore all the sensors one by one (based on the defined order) and stop if no further improvement is observed.

A recent study using the deviation data fusion strategy and a spectral range of 590–1091 nm, attained a R^2_p and RMSEP of 0.81 and 0.59% respectively for predicting SSC in pears [5]. In the present work, the sequential fusion of data from two spectral ranges clearly obtained a higher prediction accuracy for SSC with an R^2_p and RMSEP of 0.86 and 0.47%, respectively. Another related study, performed to predict dry matter (DM) (in our study, MC was predicted) and SSC in pear fruit using the spectral ranges of 600–1000 nm and 1000–2500 nm separately, showed that a maximum $R^2_p = 0.77$ (DM) and $R^2_p = 0.81$ (SSC) could be obtained. In the case of a model built on the 1000–2500 nm spectral range minimum errors were obtained for pears with a RMSEP of 0.8% and 0.47% for DM and SSC, respectively [10]. Although these authors predicted DM and SSC for pear fruit using two spectral ranges

(600–1000 nm and 1000–2500 nm) by using different spectrometers, they did not explore the possibility of fusing the complementary information from the two wavelength intervals. One reason could be the lack of sequential fusion methods available at the time when the study was performed. The use of a sequential data fusion strategy employed in the present study, effectively exploited information from multiple data matrices, leading to a better predictive model performance, compared to those reported by Ref. [5,10].

One of the main benefits of the sequential data modelling approaches, such as SO-PLS and SO-CovSel, is that they extract complementary information from the different data blocks: even if they may carry relevant information, if some predictors in the second block are redundant with respect to the variables already extracted from the first block, they won't be selected. This is also evident from the results of the present study: even if DLP NIR Scan nano covers a spectral range partly overlapped with that of Felix, none of the variables extracted from the Felix device is "duplicated" when the selection is operated on the spectra from DLP NIR Scan Nano.

5. Conclusions

The study showed that fusion of NIR and the SWIR data can improve model performance when the spectroscopic profiles are used to predict key quality parameters (MC and SSC) in pear fruit. In particular, by selecting the VIS-NIR block as the one to be modelled first, complementary information related to 2nd overtones of functional groups present in the SWIR could be integrated to substantially improve the predictive ability of the MC and SSC models. Moreover, the introduction of a further variable selection step selection step identified a minimum set of predictors that still showed good model performance. This minimum set of predictors could be used to build lower cost multispectral sensors or be integrated into existing portable devices such as the Felix. In general, the sequential fusion of spectral data from the Felix and the DLP NIR Scan Nano instruments increased the R^2_p by 2.5% and reduced the RMSEP by 6.5% compared to the best results obtained on data from a single device (in this case the Felix VIS-NIR spectrometer). Since most portable spectrometers currently available in the market work with a limited spectral range, their synergistic fusion should be explored to take advantage from the complementary information present in the different wavelength intervals. To this purpose, specifically designed multi-block chemometric techniques should be used.

Author contribution

Puneet Mishra: Conceptualization, Data curation, Investigation. Federico Marini: Formal analysis, Software, Visualization. Bastiaan Brouwe: Formal analysis, Writing - review & editing. Jean Michel Roger: Software, Writing - review & editing. Alessandra Biancolillo: Formal analysis, Methodology, Software. Ernst Woltering: Writing - review & editing. Esther Hogeveen-van Echtelt: Formal analysis, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Mariska Nijenhuis, Manon Mensink, Najim El Harchioui and Marcel Staal for their help in sample preparation and analysis for MC and SSC. Pear samples measurements with Felix sensor were part of a bigger study funded through Foundation TKI, The Netherlands Horticulture and Starting Materials and other private partners in the project (TU-16025 (1605-043) Humistatus).

References

[1] B.M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review, Postharvest Biol. Technol. 46 (2007) 99–118.
[2] R.F. Lu, R. Van Beers, W. Saeys, C.Y. Li, H.Y. Cen, Measurement of optical properties of fruits and vegetables: a review, Postharvest Biol. Technol. 159 (2020), 111003.
[3] K.B. Walsh, V.A. McGlone, D.H. Han, The uses of near infra-red spectroscopy in postharvest decision support: a review, Postharvest Biol. Technol. 163 (2020) 111139.
[4] P. Mishra, et al., Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach, Postharvest Biol. Technol. 171 (2021), 111348, https://doi.org/10.1016/j.postharvbio.2020.111348.
[5] L.M. Yuan, F. Mao, X.J. Chen, L.M. Li, G.Z. Huang, Non-invasive measurements of 'Yunhe' pears by vis-NIRS technology coupled with deviation fusion modeling approach, Postharvest Biol. Technol. 160 (2020), 111067.
[6] X.J. Yu, H.D. Lu, D. Wu, Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging, Postharvest Biol. Technol. 141 (2018) 39–49.
[7] X.M. He, X. Jiang, X.P. Fu, Y.W. Gao, X.Q. Rao, Least squares support vector machine regression combined with Monte Carlo simulation based on the spatial frequency domain imaging for the detection of optical properties of pear, Postharvest Biol. Technol. 145 (2018) 1–9.
[8] J.H. Wang, J. Wang, Z. Chen, D.H. Han, Development of multi-cultivar models for predicting the soluble solid content and firmness of European pear (Pyrus communis L.) using portable vis-NIR spectroscopy, Postharvest Biol. Technol. 129 (2017) 143–151.
[9] S.E. Adebayo, N. Hashim, R. Hass, O. Reich, C. Regen, M. Munzberg, K. Abdan, M. Hanafi, M. Zude-Sasse, Using absorption and reduced scattering coefficients for non-destructive analyses of fruit flesh firmness and soluble solids content in pear (Pyrus communis 'Conference')-An update when using diffusion theory, Postharvest Biol. Technol. 130 (2017) 56–63.
[10] S. Travers, M.G. Bertelsen, K.K. Petersen, S.V. Kucheryavskiy, Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy, Lwt-Food Science and Technology 59 (2014) 1107–1113.
[11] A.M. Cavaco, P. Pinto, M.D. Antunes, J.M. da Silva, R. Guerra, Rocha' pear firmness predicted by a Vis/NIR segmented model, Postharvest Biol. Technol. 51 (2009) 311–319.
[12] T. Sun, H.J. Lin, H.R. Xu, Y.B. Ying, Effect of fruit moving speed on predicting soluble solids content of 'Cuiguan' pears (Pomaceae pyrifolia Nakai cv. Cuiguan) using PLS and LS-SVM regression, Postharvest Biol. Technol. 51 (2009) 86–90.
[13] P. Mishra, et al., Close-range hyperspectral imaging of whole plants for digital phenotyping: recent applications and illumination correction approaches, Comput. Electron. Agric. 178 (2020), 105780, https://doi.org/10.1016/j.compag.2020.105780.
[14] P. Mishra, et al., SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials, Postharvest Biol. Technol. 168 (2020), 111271, https://doi.org/10.1016/j.postharvbio.2020.111271.
[15] P. Mishra, et al., Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable, Postharvest Biol. Technol. 170 (2020), 111326, https://doi.org/10.1016/j.postharvbio.2020.111326.
[16] P. Mishra, et al., Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit, Infrared Phys. Technol. (2020), 103547, https://doi.org/10.1016/j.infrared.2020.103547.
[17] P.P. Subedi, K.B. Walsh, Assessment of avocado fruit dry matter content using portable near infrared spectroscopy: method and instrumentation optimisation, Postharvest Biol. Technol. (2020), 111078.
[18] X.D. Sun, P. Subedi, K.B. Walsh, Achieving robustness to temperature change of a NIRS-PLSR model for intact mango fruit dry matter content, Postharvest Biol. Technol. (2020), 111117.
[19] X. Sun, P. Subedi, R. Walker, K.B. Walsh, NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment, Postharvest Biol. Technol. 163 (2020) 111140.
[20] C.A.T. dos Santos, M. Lopo, R.N.M.J. Pascoa, J.A. Lopes, A review on the applications of portable near-infrared spectrometers in the agro-food industry, Appl. Spectrosc. 67 (2013) 1215–1233.
[21] S. Gabriëls, Non-destructive measurement of internal browning in mangoes using visibleand near-infrared spectroscopy supported by artificial neural network analysis, Postharvest Biol. Technol. (2020), 111206, https://doi.org/10.1016/j.postharvbio.2020.111206.
[22] P. Mishra, et al., Improved prediction of 'Kent' mango firmness during ripening by near-infrared spectroscopy supported by interval partial least square regression, Infrared Phys. Technol. 110 (2020), 103459, https://doi.org/10.1016/j.infrared.2020.103459.
[23] J. Qin, Chapter 5 - hyperspectral imaging instruments, in: D.-W. Sun (Ed.), Hyperspectral Imaging for Food Quality Analysis and Control, Academic Press, San Diego, 2010, pp. 129–172.
[24] P. Mishra, M.S.M. Asaari, A. Herrero-Langreo, S. Lohumi, B. Diezma, P. Scheunders, Close range hyperspectral imaging of plants: a review, Biosyst. Eng. 164 (2017) 49–67.
[25] K.B. Walsh, J. Blasco, M. Zude-Sasse, X. Sun, Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use, Postharvest Biol. Technol. 168 (2020) 111246.

[26] J. Lammertyn, A. Peirs, J. De Baerdemaeker, B. NicolaÏ, Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment, Postharvest Biol. Technol. 18 (2000) 121–132.
[27] Q. Liu, K. Sun, N. Zhao, J. Yang, Y.R. Zhang, C. Ma, L.Q. Pan, K. Tu, Information fusion of hyperspectral imaging and electronic nose for evaluation of fungal contamination in strawberries during decay, Postharvest Biol. Technol. 153 (2019) 152–160.
[28] F. Mendoza, R.F. Lu, H.Y. Cen, Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content, Postharvest Biol. Technol. 73 (2012) 89–98.
[29] L. Zhou, C. Zhang, Z. Qiu, Y. He, Information fusion of emerging non-destructive analytical techniques for food quality authentication: a survey, Trac. Trends Anal. Chem. 127 (2020) 115901.
[30] V. Steinmetz, F. Sevila, V. Bellon-Maurel, A methodology for sensor fusion design: application to fruit quality assessment, J. Agric. Eng. Res. 74 (1999) 21–31.
[31] A.K. Smilde, I. Måge, T. Næs, T. Hankemeier, M.A. Lips, H.A.L. Kiers, E. Acar, R. Bro, Common and distinct components in data fusion, J. Chemometr. 31 (2017), e2900.
[32] M. Alinaghi, H.C. Bertram, A. Brunse, A.K. Smilde, J.A. Westerhuis, Common and distinct variation in data fusion of designed experimental data, Metabolomics 16 (2019) 2.
[33] Y. Song, J.A. Westerhuis, A.K. Smilde, Separating common (global and local) and distinct variation in multiple mixed types data sets, J. Chemometr. 34 (2020), e3197.
[34] P. Mishra, et al., New data preprocessing trends based on ensemble of multiple preprocessing techniques, TrAC Trends Anal. Chem. (2020), 116045, https://doi.org/10.1016/j.trac.2020.116045.
[35] P. Mishra, et al., Improved prediction of fuel properties with near-infrared spectroscopy using a complementary sequential fusion of scatter correction techniques, Talanta 223 (Part 1) (2020), 121693, https://doi.org/10.1016/j.talanta.2020.121693.
[36] P. Mishra, et al., Improved prediction of tablet properties with near-infrared spectroscopy by a fusion of scatter correction techniques, J. Pharmaceut. Biomed. Anal. (2020), 113684, https://doi.org/10.1016/j.jpba.2020.113684.
[37] A. Biancolillo, T. Næs, Chapter 6 - the sequential and orthogonalized PLS regression for multiblock regression: theory, examples, and extensions, in: M. Cocchi (Ed.), ata fusion methodology and applications, Data Handling in Science and Technology, 31, Elsevier, Oxford, UK, 2019, pp. 157–177.
[38] A. Biancolillo, R. Bucci, A.L. Magrì, A.D. Magrì, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, Anal. Chim. Acta 820 (2014) 23–31.
[39] A. Biancolillo, I. Måge, T. Næs, Combining SO-PLS and linear discriminant analysis for multi-block classification, Chemometr Intell Lab 141 (2015) 58–67.
[40] P. Firmani, A. Nardecchia, F. Nocente, L. Gazza, F. Marini, A. Biancolillo, Multi-block classification of Italian semolina based on Near Infrared Spectroscopy (NIR) analysis and alveographic indices, Food Chem. 309 (2020) 125677.
[41] A. Biancolillo, F. Marini, J.-M. Roger, So-CovSel, A novel method for variable selection in a multiblock framework, J. Chemometr. 34 (2020), e3120.
[42] J.M. Roger, B. Palagos, D. Bertrand, E. Fernandez-Ahumada, CovSel: variable selection for highly multivariate and multi-response calibration Application to IR spectroscopy, Chemometr. Intell. Lab. Syst. 106 (2011) 216–223.
[43] G. Rabatel, F. Marini, B. Walczak, J.-M. Roger, VSN: variable sorting for normalization, J. Chemometr. 34 (2020) e3164.
[44] W. Saeys, N.N. Do Trong, R. Van Beers, B.M. Nicolai, Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: a review, Postharvest Biol. Technol. (2019), 110981.
[45] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr Intell Lab 58 (2001) 109–130.
[46] P. Mishra, J.M. Roger, D.N. Rutledge, A. Biancolillo, F. Marini, A. Nordon, D. Jouan-Rimbaud-Bouveresse, MBA-GUI: A Chemometric Graphical User Interface for Multi-Block Data Visualisation, Regression, Classification, Variable Selection and Automated Pre-processing, Chemometrics and Intelligent Laboratory Systems (2020), 104139.
[47] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148.