Infrared Physics & Technology 111 (2020) 103547

Contents lists available at ScienceDirect

Infrared Physics and Technology

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

Regular article

Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit

Puneet Mishra [a,*], Ramin Nikzad-Langerodi [b]

[a] Wageningen Food and Biobased Research, Bornse weilanden 9, P.O. Box 17, 6700AA Wageningen, The Netherlands
[b] Software Competence Center Hagenberg (SCCH) GmbH, Softwarepark 21, 4232 Hagenberg, Austria

ARTICLE INFO

Keywords:
di-PLS
Robust
Calibration transfer
Near-infrared
Domain adaptation

ABSTRACT

Calibration models required for near-infrared (NIR) spectroscopy-based analysis of fresh fruit frequently fail to extrapolate adequately to conditions not encountered during initial data acquisition. Such different conditions can be due to physical, chemical or environmental effects and might be encountered for instance when measurements are carried out on a new instrument, at a different sensor operating temperature or if the model is applied to samples harvested under different seasonal conditions. To cope with such changes efficiently, for the first time, this study investigates the application of domain-invariant partial least square (di-PLS) regression to obtain calibration models that maintain the performance when used on a new condition. In particular, di-PLS allows unsupervised adaptation of a calibration model to a new condition, i.e. without the need to have access to reference measurements (e.g. dry matter contents) for the samples analyzed under the new condition. The potential of di-PLS for compensation of instrumental/seasonal and sensor temperature changes is demonstrated on four different use cases in the realm of NIR-based fruit quality assessment. The results showed that di-PLS regression outperformed standard PLS regression when tested on data affected by the aforementioned factors. The prediction R^2 increased by up to 67 % with a 46 % and 80 % decrease in RMSEP and prediction bias, respectively. The main limitation of di-PLS is that, to operate efficiently, it requires that the distribution of the response variables to be similar in the data from the different conditions.

1. Introduction

Near-infrared (NIR) spectroscopy has been the key technique for fruit quality analysis and has gained wide acceptance in different stages of the fruit supply chain [1,2]. NIR spectroscopy is of high importance being a rapid and non-destructive technique and providing access to key chemical components and physical properties of agricultural produce [3–5]. An accurate estimation of properties such as dry matter (DM) and soluble solids content (SSC) with NIR spectroscopy provides real-time access to fruit quality [1,2,6]. For detailed information regarding application of NIR spectroscopy to fruit quality analysis, readers are referred to the following references [7,2]. In recent years, NIR spectroscopy is being increasingly deployed in portable modes such as handheld or pocket devices [8,9]. Such a portable nature of new sensors and their commercial availability has made NIR spectroscopy popular not only in research laboratories but also among consumers.

NIR spectroscopy data are multivariate and consist of highly overlapping peaks. Thus, latent variables (LVs) based methods, such as partial least square (PLS) regression, are usually used to establish calibrations to predict quality attributes of fruit [10–12]. PLS regression extracts the underlying peaks related to the property of interest as LVs exhibiting maximal covariance with the property of interest [13]. The LVs can be understood as the resolved peaks which were previously hidden and were extracted by the PLS regression. However, in the domain of fruit quality analysis, PLS regression models often fail to perform well when tested on a new condition, e.g. a different instrument [14], on a sample set collected from a different season [15–17], a new cultivar [18] and a different sensor operating temperature [19]. A reason can be understood as the modelling to be suboptimal and the extracted LVs are highly specific to the first condition, and thus fails to generalize to a new condition [20]. In other words, the measurements from the new condition have some variability which is not yet modeled by the model made on the data form the first condition.

Several works have been conducted to make specific models dedicated to the conditions, for example, a segmented model for pears based on their firmness levels [21], DM and SSC models for a single pear

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

https://doi.org/10.1016/j.infrared.2020.103547
Received 5 September 2020; Received in revised form 6 October 2020; Accepted 8 October 2020
Available online 13 October 2020

cultivar [22], DM and SSC models for a single mango cultivar [23], SSC models for a single pear cultivar [24], firmness models for a single mango cultivar [25], internal browning detection models for a single mango cultivar [26], and DM and SSC for a single pear cultivar [27]. However, developing specific models limits the generalizability of the NIR models. Other works have tried to update calibration models by incorporating new reference samples measured under the new conditions and recalibrating the old models. For example, the effects of temperature were removed with the help of an external parameter orthogonalization approach to improving SSC prediction in apples [28], the external influences during the on-line implementation of NIR spectroscopy for wine fermentation monitoring were removed with the use of dynamic orthogonal projections [29] and the fruit models were recalibrated by incorporating information from new conditions [16]. Model updating works well, but a drawback of model updating is that it requires new reference samples to be measured which from a portable spectroscopy point of view is inconvenient, leading to wastage of time and money [8]. A potential solution for portable spectroscopy could be the development of generalized modelling approaches that do not require new reference measurements at the user end and can be used directly on a new condition.

Samples measured with different instruments, under different seasonal conditions, or sensor operating temperatures have intrinsic differences [20,25]. In the case of different instruments, the differences could be due to different detector sensitivities, light source and calibration of the detectors [30,31]. In the case of multi cultivar/seasons, the differences in physicochemical properties of the fruit peel and flesh could introduce variation in data set [32]. In the case of temperature, the difference could be due to the differences in the packing of molecules due to temperature variations [28]. In all the cases, there is some variation which is left unmodeled and limits the generality of models when tested on a new condition. A possible solution could be to develop models that are invariant to minor differences caused by different conditions. In the framework of PLS regression modelling, the aim should be to learn LVs that are invariant to different conditions. To this end a new method called domain-invariant partial least square (di-PLS) regression has recently been proposed [33,34]. di-PLS is a semi-supervised distribution alignment technique that aims at identifying invariant LVs across different domains [33,34] with respect to mean and co-variance. Motivated by the theory of learning from different domains, such distribution alignment techniques aim at finding representations where some source and target domain data appear to be sampled from the same underlying distribution with the goal to derive models that generalize over the domains [35]. Most importantly, di-PLS is capable of extracting invariant LVs from labeled data (spectra and reference values) from the source domain and unlabeled data (only spectra) from the target domain. More details on the di-PLS method are provided in the materials and methods section. Regarding NIR-based fruit analysis, different domains can be considered as the different instruments, seasons, cultivars and sensor operating temperatures. Hence, di-PLS might be employed to compensate for such changes.

The objective of this study is to demonstrate the potential of di-PLS regression for obtaining generalized predictive models for fruit quality. As a baseline, di-PLS regression was compared with the standard PLS regression. The demonstration is presented with four different cases of fruit quality prediction i.e. standards free calibration transfer between two instruments (one case), compensation of seasonal differences (two cases) and sensor temperature variability (one case). It is worth highlighting that di-PLS was implemented without the need for reference measurements from the new conditions. In all the cases, the aim was to predict DM in fresh fruit. To demonstrate the applicability of the method for different fruit types, we used data sets from apple, mango and olive. To the best of our knowledge, there are no existing methods in the chemometrics domain, which allow to compensate the external influences without requiring any reference measurement from the new condition. There are methods available such as EPO [28] and DOP [29], which removes the external influences from the NIR measurements, however, they require reference measurements from a different condition to model and remove those external influences.

2. Materials and methods

Data sets

Four different cases related to NIR -based fruit quality prediction were considered in this study. A description of data set is provided in Table 1. All the cases were related to the prediction of DM (%) in individual fruit. All the spectra were 2nd derivative pre-processed (Savitzky-Golay window size = 15, polynomial order = 2) to reveal the underlying peaks related to –OH bond overtones related to water. In all the cases, the instrument used was a Felix handheld spectrometer, Felix instruments, Camas, WA, USA. The Felix spectrometer generates visible and near-infrared (VNIR) spectroscopy data in the spectral range of ~ 400–1100 nm. Since, the DM has direct correlation with water bands that are present in the NIR region (>705 nm) of the spectrum. Therefore, only the NIR region was considered for multivariate analysis. Specifically, the olive data set [19] was used to demonstrate model transfer between two instruments. The data set consists of olive fruit measured with two spectrometers (Felix handheld spectrometers). The original data set consisted of NIR spectroscopy measurement on two sides of each olive fruit which was averaged in this analysis to have a single spectrum for each fruit. The mango temperature data set [18] used in this study consists of mango fruit measured at two different sensor temperature levels i.e., medium (~25 °C) and high (~30 °C). The mango season data set [18] used in this study consists of multi-season mango fruit NIR data from the years 2016 and 2018. The apple season data set [17] consists of NIR spectroscopy measurements on individual apples acquired during two seasons, i.e. 2015 and 2016.

3. Partial least square regression

PLS regression is a common chemometric technique for calibration on NIR data [13]. PLS regression deals with the multi co-linearity in multivariate signals by extracting the underlying peaks in terms of LVs that explain most of the variability in some response variable(s) $\mathbf{y}$. To this end, PLS regression first extracts a set of A latent variables $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_A]$ (known as scores) from an $N \times K$ measurement matrix $\mathbf{X}$ (of N samples and K variables), that exhibit large covariance with the response. $\mathbf{y}$ is subsequently regressed against the $N \times A$ matrix $\mathbf{T}$ in order to establish the functional relationship between measurements and response. The NIPALS (non-linear iterative partial least-squares) algorithm for PLS regression starts by using the response variable (in case of a single response variable) to estimate the weights $\mathbf{w}$ for the $\mathbf{X}$ matrix such that the covariance between $\mathbf{Xw}$ and $\mathbf{y}$ is maximized. The weight vector is further normalized to unit norm, i.e. $||\mathbf{w}|| = 1$. The $\mathbf{X}$-scores are then estimated as $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{y}$ subsequently regressed against $\mathbf{t}$. Finally, $\mathbf{X}$ and $\mathbf{y}$ are deflated in order to remove the variation explained by the current LV. The process is repeated e.g. until some cross-validation statistic indicates that there is no increase in model performance when extracting additional LVs [13].

4. Domain invariant partial least square regression

di-PLS regression extends ordinary PLS regression by a domain regularization term in order to minimize between-domain variability across two matrices $\mathbf{X}_S$ ($N_S \times K$) and $\mathbf{X}_T$ ($N_T \times K$) while maximizing the covariance between $\mathbf{X}_S$ and the corresponding response $\mathbf{y}$. $\mathbf{X}_S$ and $\mathbf{X}_T$ stands for the source and the target domain matrices, respectively. The first step of di-PLS regression is the mean centering of the inputs ($\mathbf{X}_S$, $\mathbf{X}_T$) and the outputs ($\mathbf{y}$). Subsequently, the NIPALS algorithm is employed to extract the domain-invariant latent variables across the source and target domains, i.e. by minimizing the function as explained in Eq. (1).

Table 1
Summary of data sets used for comparing the predictive performance of partial least-squares (PLS) and domain invariant partial least-squares (di-PLS).

Dataset	Spectral range (nm)	Source (Samples × Wavelengths)	Target (Samples × Wavelengths)	Reference measurement
Olive instrument transfer (Sun et al. 2020)	705–1115	186 × 135	96 × 135 (New instrument)	Dry matter (%)
Mango temperature correction [18]	705–1128	1003 × 142	996 × 142 (Different temperature)	Dry matter (%)
Mango season correction [18]	705–1115	455 × 135	483 × 135 (New season)	Dry matter (%)
Apple season correction [17]	729–975	1219 × 83	1007 × 83 (New season)	Dry matter (%)

$$\min_{\mathbf{w}} \left\| \mathbf{X}_s - \mathbf{y}\mathbf{w}^{\mathrm{T}} \right\|_{\mathrm{F}}^2 + \gamma \mathbf{w}^{\mathrm{T}} \Lambda \mathbf{w} \tag{1}$$

where $||.||\mathbf{F}$ refers to the Frobenius norm, γ is the domain regularization parameter, $\mathbf{w}$ is the weight vector and $\Lambda = \mathbf{K}\mathrm{diag}(|\lambda_1|, \cdots, |\lambda_K|)\mathbf{K}^{\mathrm{T}}$ is the matrix obtained by taking the absolute value of all eigenvalues $\lambda_1, \cdots, \lambda_K$ in the Eigen decomposition as explained in Eq. (2).

$$\mathbf{K}\mathrm{diag}(\lambda_1, \ldots, \lambda_K)\mathbf{K}^{\mathrm{T}} = \frac{1}{N_s - 1}\mathbf{X}_S^{\mathrm{T}}\mathbf{X}_S - \frac{1}{N_T - 1}\mathbf{X}_T^{\mathrm{T}}\mathbf{X}_T. \tag{2}$$

K in Eq. (2) is the eigenvector matrix of the difference between the domain-specific covariance matrices. The first term in eq. (1) corresponds to the ordinary NIPALS objective and its minimum is obtained by the direction **w** (weight vector) where $\mathbf{X}_S$ has maximum squared sample covariance with the response vector **y**. The second term in eq. (1) represents an upper bound on the absolute difference between the source sample variance and the target sample variance in the direction **w**. The (unique) solution of eq. (1) is attained by the weight vector obtained as Eq. (3) divided by its length $\mathbf{w}^{\mathrm{T}}\mathbf{w}$.

$$\mathbf{w}^{\mathrm{T}} = \frac{\mathbf{y}^{\mathrm{T}}\mathbf{X}_S}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\left(\mathbf{I} + \frac{\gamma}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\Lambda\right)^{-1} \tag{3}$$

The coordinates (scores) $\mathbf{t}_S$ and $\mathbf{t}_T$ of the (domain-invariant) projections corresponding to the direction **w** can be computed by

$$\mathbf{t}_s = \mathbf{X}_s\mathbf{w} \text{ and } \mathbf{t}_T = \mathbf{X}_T\mathbf{w}. \tag{4}$$

Similar to PLS regression, di-PLS regression also involves an orthogonalization step to remove the variation from the data which is already explained by the current LV. The orthogonalization is performed such that

$$\mathbf{X}_S := \mathbf{X}_S - \mathbf{t}_S(\mathbf{t}_S^{\mathrm{T}}\mathbf{t}_S)^{-1}\mathbf{t}_S^{\mathrm{T}}\mathbf{X}_S \tag{5}$$

and analogously for $\mathbf{X}_T$. The remaining steps in di-PLS are equal to the standard PLS regression algorithm [34]. All experiments were conducted using in-house implementations of PLS and di-PLS in python 3.7.

The regularization parameter γ in Eq. (1) was fixed for each LV according to the heuristic described in literature [36]. In brief, γ was set such that equal weight was assigned to both terms in Eq. (1). The number of LVs for PLS regression and di-PLS regression models were chosen based on the inflection point of the cross-validation error plot.

5. Results and discussion

Instrument transfer case for olive fruit data

The results of PLS regression and di-PLS regression for standard-free calibration transfer between two Felix NIR handheld spectrometers for prediction of DM in fresh olive fruit are shown in Fig. 1. Most of the spectral differences between the devices that are visible at wavelengths beyond 1000 nm could be because of differences in the sensitivities of the silicon detectors above 1000 nm. However, that is not a problem with di-PLS as it captured the major source of useful variability and in the present case the NIR part that correlated most with the DM. Even though the variability in DM was similar for both instruments, application of a 3 LVs PLS regression model established on instrument 1 showed poor predictive performance (RMSEP = 2.98 % and $R^2_p = 0.28$) on the data collected with instrument 2. The bias obtained with the standard PLS was 0.4 %. On the other hand, di-PLS regression with the same number of LVs successfully recovered the functional relationship between NIR spectra and DM and improved the R^2 and RMSEP and to 0.85 and 1.61 %, respectively while increasing the bias. The improvements can be mostly attributed to the alignment of the (marginal) distributions of the domain-specific samples in the LVs space (see Fig. 5A and 5E) .

Temperature correction case for mango fruit data

For the sensor temperature correction experiment in the case of mango fruit samples, where calibration and test sets were recorded at ~ 15 and ~ 25 °C, respectively, the similarity between the source and target domains in terms of the NIR spectra was comparably high (Fig. 2A). Consequently, the PLS regression (baseline to di-PLS regression) yielded good prediction results for DM based on the data acquired

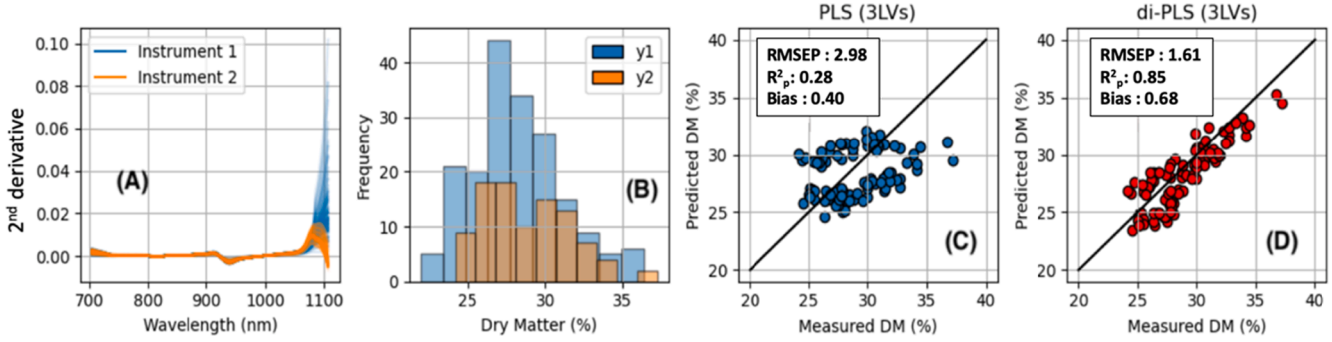

Fig. 1. Olive calibration transfer case. (A) Spectra from instrument 1 (blue) and instrument 2 (orange), (B) histograms explaining the distribution of the reference dry matter for samples measured on different instruments (Y1 for instrument 1 and Y2 for instrument 2), (C) testing the PLS regression calibration made on instrument 1 on instrument 2, and (D) testing the di-PLS regression calibration made on instrument 1 on instrument 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

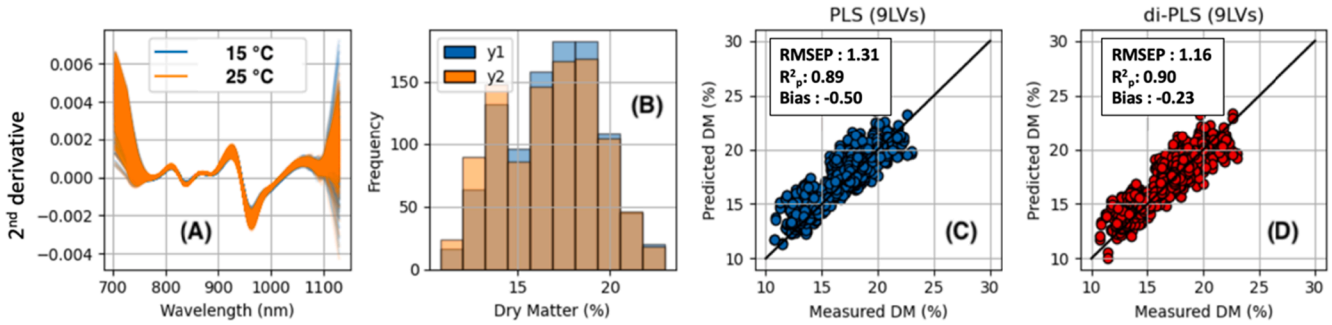

Fig. 2. Mango temperature correction case. (A) Spectra from 15 °C (blue) and 25 °C (orange), (B) histograms explaining the distribution of the reference dry matter for samples measured at different temperatures (Y1 for 15 °C and Y2 for 25 °C), (C) testing the PLS regression calibration made on data from 15 °C and tested on data from 25 °C, and (D) testing the di-PLS regression calibration made on data from 15 °C and tested on data from 25 °C. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

at the higher sensor temperature (Fig. 2C). However, application of di-PLS regression further improved the accuracy. The RMSEP improved from 1.31 % to 1.16 %, and the prediction bias was reduced from −0.50 % to −0.23 % compared to the baseline PLS model.

Season effect correction for mango and apple data

For the experiments involving seasonal effects correction on the NIR spectra of mango and apple, it was found that adapting the corresponding calibration models by means of di-PLS regression yielded significant improvements in terms of the RMSEP (Figs. 3 and 4). For the former, the RMSEP improved from 2.57 % (PLS) to 1.66 % di-PLS). However, it was found that the di-PLS regression model systematically underestimated the DM which can be explained by the fact that the average DM of the 2019 sample is slightly higher compared to the 2016 sample (Fig. 3B) for which the reference values are included when fitting the model. This is because di-PLS regression assumes that the distribution of the response **y** is similar in the source and target domains, which is an important limitation of the method. di-PLS also reduced the prediction bias from −1.91 % to 1.10 % for the mango season correction data set (Fig. 3). Finally, for correcting the seasonal variability between apples harvested in 2015 and 2016, di-PLS regression increased the model accuracy on the target domain samples (Fig. 4). In particular, di-PLS improved the RMSEP from 0.76 % to 0.56 % and the prediction bias from 0.46 % to 0.09 %. A complete summary of improvement in model accuracies for all data sets can be found in Table 2. In all the cases, di-PLS regression outperformed standard PLS regression in terms of higher prediction R^2 and lower RMSEP.

Comparison of PLS and di-PLS based on scores and regression vectors

Fig. 5 shows the projections of the 4 data sets on the first 2 LVs of the corresponding PLS regression and di-PLS regression models. As expected, the distributions of source and target domain differ the most for the olive instrument transfer experiment, where the differences in the NIR spectra were most pronounced. Whereas, di-PLS regression successfully aligns the two distributions, which explains the notable increase in prediction accuracy over the PLS regression model tested on the samples recorded on instrument 2 (Fig. 1). In contrast, the distributional differences are more subtle for the temperature and seasonal change experiments, where standard PLS regression models generalized reasonably well on the target domain samples. For the mango season experiment, alignment of the distributions is not optimal as can be seen from the change of structure of the target domain samples (Fig. 5C and 5G). In particular, the two clusters of the target domain data (orange) seen in Fig. 5C indicates a bi-modal distribution of the spectra which disappears after domain regularization.

The regression vectors from PLS regression and di-PLS regression for all four data sets are shown in Fig. 6. The overall shape of the regression vectors (PLS regression and di-PLS regression) was similar (with the major peaks related to moisture). However, notable differences between the regression vectors of PLS regression and di-PLS regression are the higher weights and well resolved peaks in the regression vector of di-PLS. The regression vector of di-PLS regression has higher weights at the similar spectral regions where the PLS regression vector showed significant peaks but with relatively less weights. In addition, there were some spectral regions where the peaks get well resolved with the di-PLS regression; for instance, around 850 nm in the olive data set (Fig. 6A), around 800 nm in mango temperature correction data set (Fig. 6B), around 800 nm and 950 nm in mango season correction data set, around 920 nm in apple season correction dataset. Such resolved peaks and higher regression weights at important wavelengths obtained by di-PLS regression compared to the PLS regression could be the reason for better performance of the di-PLS regression over the PLS regression for dealing with crucial and practically relevant tasks such as instrument transfer, temperature correction and seasons effects correction.

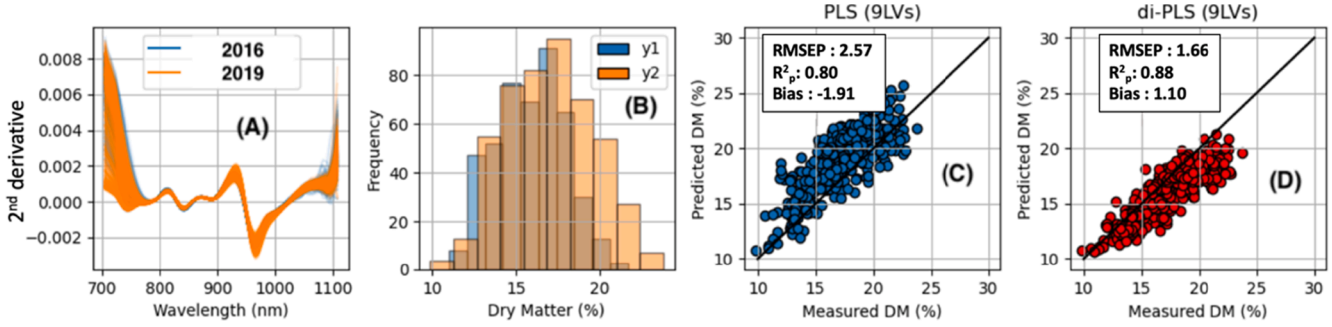

Fig. 3. Mangoes seasonal effects correction case. (A) Spectra from year 2016 (blue) and year 2019 (orange), (B) histograms explaining the distribution of the reference dry matter for samples measured at different seasons (Y1 for year 2016 and Y2 for year 2019), (C) testing the PLS regression calibration made on data from year 2016 and tested on data from year 2019, and (D) testing the di-PLS regression calibration made on data from year 2016 and tested on data from year 2019. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

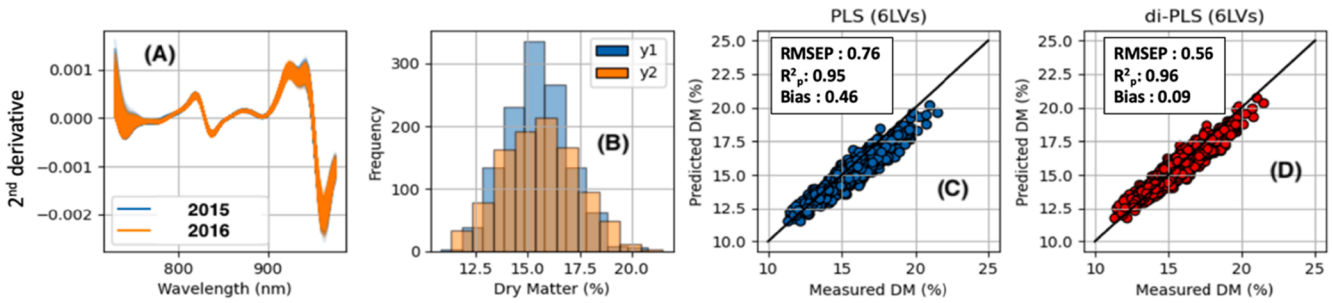

Fig. 4. Apples seasonal effects correction case. (A) Spectra from year 2015 (blue) and year 2016 (orange), (B) histograms explaining the distribution of the reference dry matter for samples measured at different seasons (Y1 for year 2015 and Y2 for year 2016), (C) testing the PLS regression calibration made on data from year 2015 and tested on data from year 2016, and (D) testing the di-PLS regression calibration made on data from year 2015 and tested on data from year 2016. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 2
Summary of improvements in model accuracies with the use of di-PLS regression compared to standard PLS regression.

Dataset	% increase in R^2 with di-PLS regression	% decrease in RMSEP with di-PLS regression	% decrease in Bias with di-PLS regression
Olive instrument transfer	67	46	No improvement
Mango temperature correction	1	11	54
Mango season correction	10	35	42
Apple season correction	1	26	80

Compared to the original studies related to the data sets used in this work, the di-PLS regression in the case of instrument transfer for olive fruit allows standard free calibration transfer which in the original study [19] was performed by performing extra measurements in the new instrument. Thus, it has demonstrated that with the application of di-PLS regression the new measurements may not be necessary. In the case of the mango temperature correction data set [19], the original work stated the best $R^2_p = 0.82$ obtained with the PLS regression, however, the di-PLS obtained a $R^2_p = 0.9$ with a similar error value. For the multi-season data set of mangoes [18] and apple [17], the di-PLS model performed similarly to the results presented in the original published studies. However, unlike those studies where calibration models were developed with the data of samples from all seasons, the di-PLS calibration model developed here used data of samples from one season and tested on data from a different season. Hence, the di-PLS model developed in this study can be considered more legitimate.

6. Conclusions

Failure of NIR models has been a long-existing problem in the domain of fresh fruit quality analysis. The PLS regression models developed for fruit quality analysis usually work well within the domain in which they were calibrated but fails when tested on a new domain corresponding to measurements from a different instrument, sensor operating temperatures and seasons. In this study, di-PLS regression has been proposed for modelling the NIR spectra of fresh fruit to deal with NIR model failure. In all the four cases presented related to instrument transfer/ temperature correction/ season effect correction, the di-PLS regression model showed superiority to the standard PLS regression modelling commonly performed in the NIR domain. The improvements were noticed as di-PLS regression was able to extract the generalized latent variables from multiple batches corresponding to a different

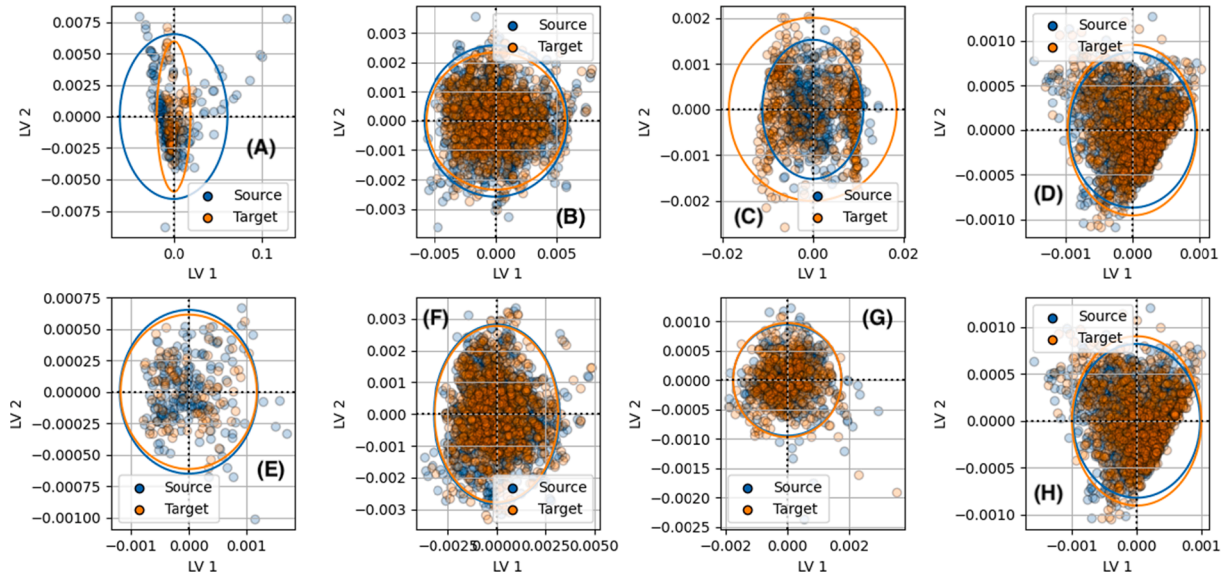

Fig. 5. Latent space representations. Projections of source and target domain samples onto the first 2 LVs of PLS regression (top row) and di-PLS regression (bottom row) models. Ellipses denote 95 % confidence intervals. PLS regression: LV1 vs LV2 (A) olive data set, (B) mango temperature data set, (C) mango season data set, and (D) apple season data set. di-PLS regression:LV1 vs LV2 (E) olive data set, (F) mango temperature data set, (G) mango season data set, and (H) apple season data set.

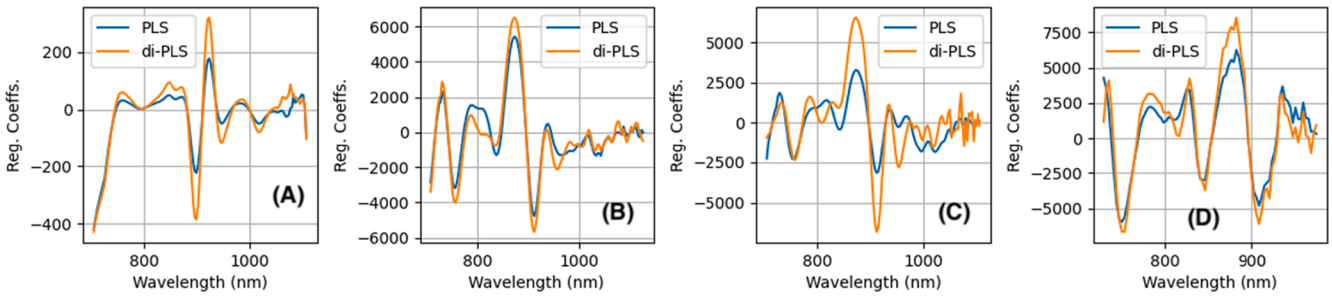

Fig. 6. Regression vector for PLS regression (blue) and di-PLS regression (orange). (A) olive instrument transfer, (B) mango temperature correction, (C) mango season correction, and (D) apple season correction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

instrument, temperature condition and season. In summary, advanced methods like di-PLS regression can facilitate the development of generalized fruit quality models, which work well on multiple instruments, temperature conditions and multi-season experiments. Domain regularization-based methods such as di-PLS regression can support in making NIR spectroscopy models scalable and widely applicable. However, a limitation of the di-PLS method is that it requires the distribution of response variables from two conditions to be similar. In addition, compared to standard PLS regression, domain invariant learning requires several parameters to be optimized.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgments

Prof. Kerry B. Walsh and Dr. Phul Subedi from Central Queensland University, Australia for sharing the Olive fruit dataset used in the study.

Dr. Soon Li The and Prof. Kate Evans from Washington State University for sharing the Apple fruit multi-season dataset used in the study.

RNL acknowledges funding by BMVIT, BMDW, and the Province of Upper Austria in the frame of the COMET Program managed by the FFG and the COMET Centre CHASE (project No 868615).

References

[1] 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. Technology 168 (2020) 111246.
[2] K.B. Walsh, V.A. McGlone, D.H. Han, The uses of near infra-red spectroscopy in postharvest decision support: a review, Postharvest Biology Technology 163 (2020) 111139.
[3] P. Mishra, A. Biancolillo, J.M. Roger, F. Marini, D.N. Rutledge, New data preprocessing trends based on ensemble of multiple preprocessing techniques, TrAC Trends Analytical Chemistry 116045 (2020).
[4] P. Mishra, S. Lohumi, H. Ahmad Khan, A. Nordon, Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches, Comput. Electron. Agriculture 178 (2020) 105780.
[5] Mishra Puneet, et al., Close range hyperspectral imaging of plants: A review, Biosyst Eng 164 (2017) 49–67, https://doi.org/10.1016/j.biosystemseng.2017.09.009.
[6] Mishra Puneet, et al., Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit, Talanta 223 (Part 2) (2021) 121733, https://doi.org/10.1016/j.talanta.2020.121733.
[7] 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. Technology 46 (2) (2007) 99–118.
[8] R.A. Crocombe, Portable spectroscopy, Appl. Spectroscopy 72 (12) (2018) 1701–1751.
[9] J. Yan, L. van Stuijvenberg, S.M. van Ruth, Handheld near-infrared spectroscopy for distinction of extra virgin olive oil from other olive oil grades substantiated by compositional data, European J. Lipid Science Technology 121 (12) (2019).
[10] W. Saeys, N.N. Do Trong, R. Van Beers, B.M. Nicolai, Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review, Postharvest Biology Technology 158 (2019).
[11] 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, Chemomet. Intelligent Laboratory Systems 104139 (2020).
[12] P. Mishra, J.M. Roger, D.N. Rutledge, E. Woltering, SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials, Postharvest Biology Technology 168 (2020) 111271.
[13] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intelligent Labor. Systems 58 (2) (2001) 109–130.
[14] 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 Biology Technology 163 (2020) 111140.
[15] B.B. Wedding, C. Wright, S. Grauf, R.D. White, B. Tilse, P. Gadek, Effects of seasonal variability on FT-NIR prediction of dry matter content for whole Hass avocado fruit, Postharvest Biol. Technol. 75 (2013) 9–16.
[16] P. Rungpichayapicheta, B. Mahayothee, M. Nagle, P. Khuwijitjaru, J. Mullera, Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango, Postharvest Biol. Technology 111 (2016) 31–40.
[17] S.L. Teh, J.L. Coggins, S.A. Kostick, K.M. Evans, Location, year, and tree age impact NIR-based postharvest prediction of dry matter concentration for 58 apple accessions, Postharvest Biol. Technology 166 (2020) 111125.
[18] N.T. Anderson, K.B. Walsh, P.P. Subedi, C.H. Hayes, Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content, Postharvest Biology Technol. 168 (2020) 111202.
[19] 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 Biology Technology 162 (2020).
[20] P. Mishra, J.M. Roger, D.N. Rutledge, E. Woltering, Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable, Postharvest Biology and Technology 170 (2020) 111326.
[21] 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 Biology Technology 51 (3) (2009) 311–319.
[22] 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 Sci. Technology 59 (2) (2014) 1107–1113.
[23] J.P.D. Neto, M.W.D. de Assis, I.P. Casagrande, L.C. Cunha, G.H.D. Teixeira, Determination of 'Palmer' mango maturity indices using portable near infrared (VIS-NIR) spectrometer, Postharvest Biol. Technology 130 (2017) 75–80.
[24] 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. Technology 160 (2020).
[25] P. Mishra, E. Woltering, N. El Harchioui, Improved prediction of 'Kent' mango firmness during ripening by near-infrared spectroscopy supported by interval partial least square regression, Infrared Phys. Technology 110 (2020) 103459.
[26] S.H.E.J. Gabriëls, P. Mishra, M.G.J. Mensink, P. Spoelstra, E.J. Woltering, Non-destructive measurement of internal browning in mangoes using visible and near-infrared spectroscopy supported by artificial neural network analysis, Postharvest Biol. Technology 166 (2020) 111206.
[27] P. Mishra, E. Woltering, B. Brouwer, E. Hogeveen-van Echtelt, Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach, Postharvest Biology Technology 171 (2021) 111348.
[28] J.-M. Roger, F. Chauchard, V. Bellon-Maurel, EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits, Chemometrics Intelligent Laboratory Systems 66 (2) (2003) 191–204.
[29] M. Zeaiter, J.M. Roger, V. Bellon-Maurel, Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations, Chemometrics Intelligent Laboratory Systems 80 (2) (2006) 227–235.

[30] T. Fearn, Standardisation and calibration transfer for near infrared instruments: a review, J. Near Infrared Spectroscopy 9 (4) (2001) 229–244.
[31] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, Chemometrics Intelligent Laboratory Systems 64 (2) (2002) 181–192.
[32] 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 Biology Technology 159 (2020).
[33] R. Nikzad-Langerodi, W. Zellinger, E. Lughofer, S. Saminger-Platz, Domain-invariant partial-least-squares regression, Anal. Chem. 90 (11) (2018) 6693–6701.
[34] R. Nikzad-Langerodi, W. Zellinger, S. Saminger-Platz, B. Moser, Domain-invariant regression under beer-lambert's law, 2019 18th IEEE International Conference On Machine Learning Applications (ICMLA) (2019).
[35] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan, A theory of learning from different domains, Machine Learning 79 (1) (2010) 151–175.
[36] R. Nikzad-Langerodi, W. Zellinger, S. Saminger-Platz, B.A. Moser, Domain adaptation for regression under Beer–Lambert's law, Knowledge-Based Systems 210 (2020) 106447.