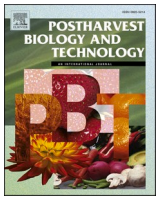
Postharvest Biology and Technology 204 (2023) 112477

Contents lists available at ScienceDirect

Postharvest Biology and Technology

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



Near infrared light penetration in watermelon related to internal quality evaluation

Miguel Vega-Castellote [a], María-Teresa Sánchez [a,*], Jens Petter Wold [b], Nils Kristian Afseth [b], Dolores Pérez-Marín [c,*]

[a] Department of Bromatology and Food Technology, University of Cordoba, Rabanales Campus, Córdoba 14071, Spain
[b] Nofima Ås – Norwegian Institute of Food, Fisheries and Aquaculture Research, PB 210, N-1431 Ås, Norway
[c] Department of Animal Production, University of Cordoba, Rabanales Campus, Córdoba 14071, Spain

ARTICLE INFO

Keywords:
Thick rind fruit
NIRS interactance
Light penetration
Sugar content
Multivariate Curve Resolution

ABSTRACT

The interaction between near infrared (NIR) light and the sample is a key factor to take into consideration when developing applications using NIR spectroscopy. The aim of this study was to assess the penetration depth of the NIR light through both intact watermelons and the red flesh of the cut fruit, using an interactance NIR system. The results showed a maximum penetration of between 9 and 11 mm in the assays carried out on intact watermelons and 11 mm in the watermelon flesh. The prediction performance of the different NIR system configurations used were compared by quantifying the SSC in intact watermelons in a non-destructive way. The results obtained for the assay in which the illuminated regions were separated by 28 mm (Distance 4) reported the best results for determination of SSC (%) obtaining an $R^2_{cv} = 0.73$ and RMSECV = 0.39 % and an $R^2_p = 0.81$ and RMSEP = 0.30 % when the set of data for that distance was split into calibration and internal validation sets.

1. Introduction

Over the past few years, rapid, non-destructive methods have emerged for measuring the degree of maturity and soluble solids content (SSC) in intact watermelons, in the quest for a fast, accurate, high throughput and continuous quality evaluation of these fruits. The most commonly used methods are those based on acoustic signals, machine vision systems, electrical and magnetic technologies, as well as vibrational spectroscopy techniques (Sun et al., 2010; Ali et al., 2017; Jie and Wei, 2018). The latter have been widely used in the quality assessment of agricultural products given their high speed, low cost, versatility and non-destructive nature (Cortés et al., 2019; Pandiselvam et al., 2022). Vibrational spectroscopy techniques rely on the characterization of a sample by means of its vibrational energy, which in turn is related to the sample's chemical bonds (Teixeira and Sousa, 2019). However, most of the reported studies on vibrational spectroscopy for assessing maturity in watermelons focus on the use of near infrared (NIR) spectroscopy (Arendse et al., 2018).

The viability of NIR techniques for evaluating maturity and for quantifying the SSC in intact and cut watermelons has been reported by different authors, including Flores et al. (2008), Tamburini et al. (2017), Ibrahim et al. (2022), Vega-Castellote et al. (2022) and Vega-Castellote et al. (2023), among others. In all cases, promising results were obtained as regards developing applications for measuring the maturity and SSC of intact fruit both in the field and on the industrial sorting lines, as well as for quantifying SSC in half-watermelons. The studies carried out with intact fruit rely on the correlation between NIR data acquired from the rind of the intact watermelons (usually sampling the first few millimeters of flesh below the rind) and the reference SSC values from different portions of the flesh. It is therefore important to assess the penetration depth of NIR light in intact watermelons to understand the potential and validity of NIR measurements of intact watermelons. Different authors have reported and assessed the penetration depth of NIR light into different fresh fruit tissues. For example, Guthrie and Walsh (1997) studied the interaction of NIR light with intact pineapple samples using a reflectance NIR instrument working in the 760–1300 nm range and concluded that most of the spectral data obtained were derived from within 5 mm of the fruit surface. Likewise, Lammertyn et al. (2000) assessed the penetration depth in intact apples using an NIR device working in reflectance mode in the 500–1900 nm range and reported that the light could reach a depth of 2–4 mm depending on the wavelength used, since, according to Fraser et al. (2001, 2003), the high

* Corresponding authors.
E-mail addresses: teresa.sanchez@uco.es (M.-T. Sánchez), dcperez@uco.es (D. Pérez-Marín).

https://doi.org/10.1016/j.postharvbio.2023.112477
Received 22 May 2023; Received in revised form 11 July 2023; Accepted 13 July 2023
Available online 18 July 2023

concentrations of water in fresh fruit makes it difficult for the light to penetrate the sample in regions above 1200 nm. Greensill and Walsh (2000) used a 'partial transmittance' NIR device, in which the light source illuminated an annular area of the sample, and the emerging radiation was collected from the non-illuminated center by the detector probe. These authors reported that the light received was mainly detected up to a depth of 10 mm in the intact rock melon samples. Nevertheless, no reports have been found which deal with the penetration of NIR light in thick rind watermelons and in the different watermelon tissues, i.e., the rind and the flesh, which highlights the need to clarify how NIR radiation interacts with these watermelon tissues, since one of the major barriers when measuring internal maturity in fruit is the limited penetration of NIR light, especially in fruit with a thick rind. This, in turn, depends on the wavelength range used, the distance between the sample and the light source, the characteristics of the product analyzed and the intensity of the light source, among others (Birth, 1978; Greensill and Walsh, 2000; Flores et al., 2008; Arendse et al., 2018). Only one study for a thin-skin watermelon cultivar (Qian et al., 2016) assessed the penetration depth of NIR in this type of fruit.

The aim of the present study was to assess the penetration of NIR light into intact watermelons and into the red flesh of the cut fruit, using a NIR prototype system working in interactance mode in the 761–1081 nm range and to compare the prediction performance of the different configurations used by quantifying the SSC in intact watermelons in a non-destructive way.

2. Material and methods

2.1. Samples

A total of 20 striped light green rind watermelons (*Citrullus lanatus* Thunb.) purchased from a Norwegian grocery market from August to November 2022 were used in this work. The fruits were taken to the spectroscopy laboratory at the Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima) in Ås, Norway, and kept at room temperature (ca. 20ºC) for a maximum of 5 days after purchase.

2.2. Instrumentation, experimental design and acquisition of the spectral information

2.2.1. NIR interactance assays

The NIR spectral information was taken using a NIR prototype system designed for non-contact interaction measurements in the 761–1081 nm range, taking data every 16 nm. The system featured an InGaAs array detector. The acquisition time was set at 2 s, in which 100 scans were collected, which were then averaged to obtain one mean spectrum per sample. The measurements were taken at a distance of approximately 20 cm from the sample. As described by Wold et al. (2021), the NIR light source (50 W halogen light source) featured an illumination chopper that enabled to obtain 5 pairwise illuminated regions termed 'Distance 1' to 'Distance 5' (Fig. 1A) separated by 10, 16, 22, 28 and 34 mm, respectively. The dimensions of each of the illuminated regions were 2 × 28 mm, and the interactance signals between each pair were recorded consecutively. The light penetrated the sample through the illuminated regions, and after interacting with the sample it was detected in the instrument's field of view (2 × 8 mm), which was positioned centrally between the paired regions. The spectrometer and illumination were electronically synchronized. The NIR data were acquired using the Matlab software version R2018a (The Mathworks, Inc., MA, USA).

The NIR interactance instrument was used to assess the penetration depth of the light through the watermelon's rind and through its red flesh. To achieve this, one of the 20 watermelons was used to take a spectrum from two squares, one of rind and red flesh and the other one only of red flesh, which measured 80 × 80 mm and 15 mm deep. After that, these squares were placed upon a block of coconut fat measuring 80 × 80 × 40 mm, and spectra were recorded slicing the squares gradually thinner from the underside, i.e., from a depth of 15–1 mm, taking off a 2 mm slice each time. Finally, a spectrum of the coconut fat was also recorded. Coconut fat was used as a contrast in these studies given the great difference between watermelon and coconut fat spectral signatures, as proposed in an experiment carried out by Wold et al. (2021).

Next, NIR spectra were acquired from 5 different measurement points on the surface of the intact rind of the remaining 19 watermelons to predict the internal SSC of the fruits (19 watermelons × 5 measurement points = 95 measurements). The first three points were taken by dividing the peduncle–pistil axis into four segments, so three equidistant points could be obtained in that axis, with the second measurement placed in the exact equator of the fruit. This distance was used to measure the fourth and fifth points on both sides of the second point in the equatorial line (Fig. 1B).

2.3. Reference analysis

To obtain the reference data to predict the internal SSC of the watermelons using the NIR system, based on the results obtained in the

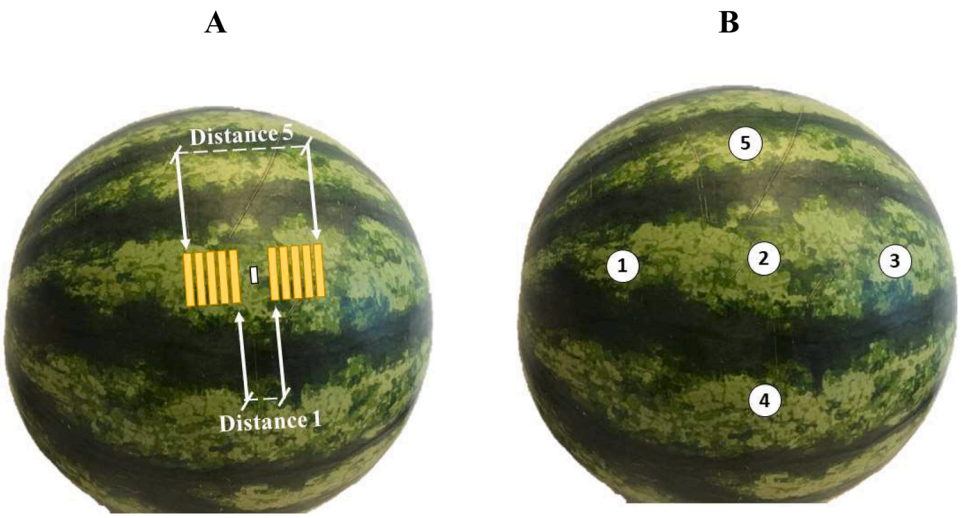


Fig. 1. Representation of the illuminated regions (Distance 1 to Distance 5) on the watermelon surface used to acquire the NIR interactance data, and position of the field of view of the system (white rectangle between the illuminated areas) (A). Position of the areas measured to predict the internal SSC of the fruits (B).

light penetration depth test, a 10 mm-deep portion was taken using a precision calliper right under the surface illuminated during the assay. The hard green skin was removed from that portion using a potato peeler and then it was liquidized using a blender. The juice extracted was measured by refractometry using a digital refractometer TS Meter-D (Reichert Inc., NY, USA).

2.4. Multivariate data analysis

The chemometric treatment of the data was carried out using Matlab software. The spectral information obtained with the NIR interactance instrument was first linearized and transformed into absorption data as $\log_{10}(1/I)$, where 'I' represents the intensity values of the detected signal, and then normalized using standard normal variate (SNV) treatment (Barnes et al., 1989).

For the NIR light penetration study, the multivariate curve resolution (MCR) method (Tauler, 1995; De Juan and Tauler, 2021) was used, together with the alternating least squares (ALS) algorithm, to estimate the relative contributions of watermelon and coconut fat as a function of the watermelon slice thickness for the 5 distances used by the system (Distance 1 to Distance 5). This method is based on a bilinear decomposition of the data that can be mathematically written as (De Juan and Tauler, 2021):

$$D = CS^T + E,$$

where 'D' is the matrix containing the raw data, 'C' is the matrix of the relative intensities, 'S^T' the matrix of pure spectra and E represents the variation unexplained by the model. The dimensions of these matrices were D ($m \times n$), C ($m \times k$), S ($n \times k$) and E ($m \times n$), where 'm' was the number of spectra, 'n' the number of variables or spectroscopic wavelengths and 'k' the number of factors. The matrix of relative intensities and the matrix of pure spectra were plotted and interpreted. In addition, the matrix of relative intensities was used to assess the contribution of the coconut fat and the watermelon in the signal detected by the sensor. The percentage of fat signal detected was calculated as a function of the scores of the relative intensities' matrix, taking the score value of the pure coconut fat sample and the factor representing the coconut fat component as 100 % of fat signal detected. Then, these percentage values were plotted for interpretation.

For the assay to predict the SSC in intact watermelon, a principal component analysis (PCA) was first carried out to study the structure and variability of the NIR spectra using the Mahalanobis Global distance (GH), calculated as $GH = D^2/f$, where 'D' is the calculated Mahalanobis distance calculated on principal components and 'f' the number of factors used (Walsh et al., 2000). To detect any potential spectral outlier samples, the spectra showing GH values over 4 were studied in detail and excluded from their respective sets if their removal was justified. After that, a characterization of the calibration sets was carried out considering the number of samples (N), range, mean, standard deviation (SD), and the coefficient of variation (CV), which relates the SD to the mean value of the population. Then, in order to evaluate the calibration performance of the different distances assessed with the NIR instrument, partial least squares regression (PLSR) models were developed (Martens and Naes, 1989), including all the samples in the calibration set and using segmented cross validation, which is ideal when there are structures in the data set (Naes et al., 2002) –i.e., in each of the cross validation passes the SSC of one watermelon (N = 5 measurements) was predicted using a model developed with the remaining watermelons in the set of samples. The results were assessed considering the coefficient of determination for cross validation (R^2_{cv}) and the standard error of cross validation (RMSECV) which is defined by:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_{CV,i} - y_i)^2}{N}}$$

where N is the number of samples included in the calibration model, i the sample from 1 to N, $\hat{y}_{CV,i}$ the cross validation predicted value for the i sample and y_i the measured reference value for the i sample. The RMSECV values obtained using the different distances tested were compared to find potential differences between them using Fisher's F test (Mark and Workman, 2003; Fearn, 2009). The values for the F statistic were calculated as:

$$F = \frac{RMSECV_2^2}{RMSECV_1^2}$$

where $RMSECV_1$ and $RMSECV_2$ are the standard errors of cross validation of two different models being $RMSECV_1 < RMSECV_2$. F is compared to $F_{critical}$ (1 - P, n_1 - 1, n_2 - 1) read from the table with $P = 0.05$ and n - 1 degrees of freedom, where n_1 stands for the number of times the measurement is repeated with method 1 and n_2 the number of times the measurement is repeated with method 2. If F is higher than $F_{critical}$, the two RMSECV values are significantly different. Once the best distance was identified, the original set of samples was divided into calibration and internal validation sets, including 2 randomly selected watermelons (N = 10 measurements) in the internal validation set, with the rest used to make the calibration set. A new PLSR calibration model was carried out, which was evaluated using the samples included in the internal validation set, following the protocol outlined by Windham et al. (1989).

3. Results and discussion

3.1. Characterization of the NIR data collected

The NIR spectra collected from the dark and light green parts of the watermelon rind and from the red flesh are shown in Fig. 2. Differences can be observed between the spectra taken for each of the distances tested, with Distance 1 showing the lowest attenuations and thus, the highest intensities captured by the sensor. This was to be expected, given the shorter path the light had to travel through the watermelon. The signal intensity decreased from Distance 1 to Distance 5 and, consequently, greater attenuations can be observed between the different distances. It is important to note the weak attenuation increase from Distance 4 to Distance 5 in the assays carried out in the watermelon rind (Fig. 2A and B) and red flesh (Fig. 2C). This could be due to non-linear changes in the attenuation related to the long distance the light travels from the illuminated areas in Distance 5 to the field of view (FOV) of the instrument (Wold et al. 2021). It is well known that deviations from linearity in Beer-Lambert-Bouguer's law can occur as, for example, in the cases in which the concentrations of the analytes are extremely high, the medium is highly scattering or the deviations depending on the thickness of the length travelled by the light (Mayerhöfer et al. 2020; Mamouei et al. 2021). In this sense, Dahm and Dahm (2001) indicated that the linear absorption coefficient of the sample is dependent on the actual distance that the light travels through a sample. Moreover, Mayerhöfer et al. (2020) reported that the received reflectance signal after the light interaction with the matter is dependent on the layer thickness; they highlight that absorbance may not only increase linearly with thickness, but it could even decrease, as appears to occur in this case.

No differences can be seen between the spectral curves of the two assays carried out in the dark and light green areas of watermelon rind (Fig. 2A and B). Nevertheless, when the measurements were carried out in the red flesh of the fruit, a greater signal intensity was collected compared to the assays carried out in the rind, as can be seen in Fig. 2C, where the attenuation values were lower for the 5 distances tested. The light in the assays carried out in the rind travelled through the rind twice: first in the illuminated area and back again in the FOV to reach the detector of the instrument, which involves a greater loss of light when compared with the assays carried out in the red flesh.

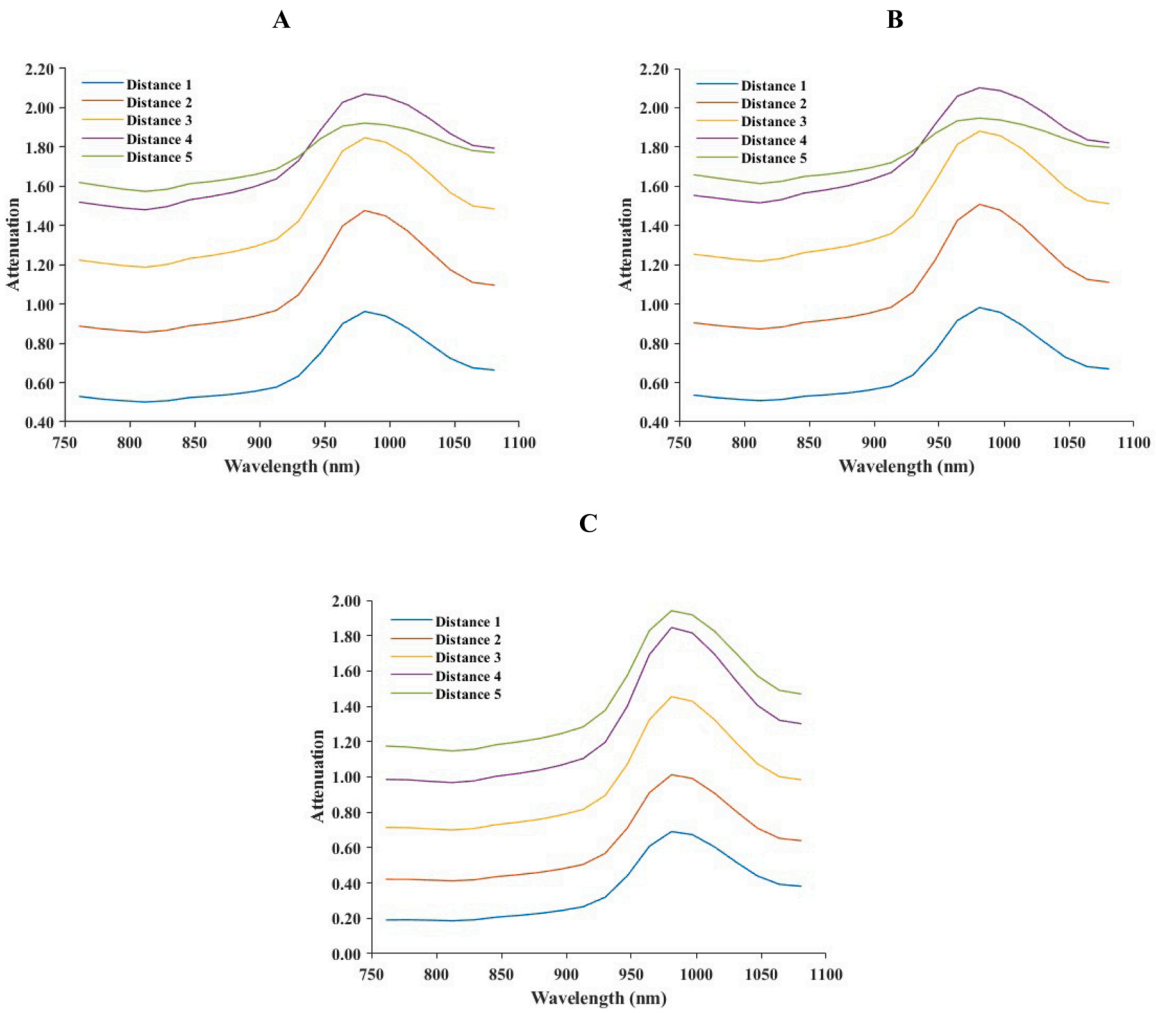


Fig. 2. Attenuation values of the spectra taken from the dark (A) and light (B) green parts of the watermelon rind and the flesh (C), with the 5 different distances used by the NIR instrument.

3.2. Assessment of the penetration of the NIR light through the watermelon rind and red flesh

The SNV normalized NIR spectra taken from the watermelon without the block of coconut fat, from the watermelon slices of different thicknesses placed on top of the coconut fat block, and from the coconut fat block itself, were studied to spot the differences in their spectral curves. An example is shown in Fig. 3 taking the data of the assay carried out using the red flesh at Distance 4 of the instrument. A distinctive peak of the coconut fat block can be seen at 930 nm in the coconut fat spectrum and in some of the spectra of the watermelon slices with different thicknesses: the thicker the watermelon slice, the more difficult it is to spot that peak.

Two factors were used in the MCR-ALS model, since the two components analyzed in the mixture were the watermelon and the coconut fat. The bilinear decomposition of the original data matrix enabled to obtain the matrix of relative intensities and the matrix of pure spectra, as can be seen in the example shown in Fig. 4A and B, respectively, obtained using the data taken from the light green area of watermelon rind at Distance 4 of the instrument. The two factors used in the MCR-ALS model can be identified in the loadings of the pure spectra matrix (Fig. 4B), showing factor 1 and factor 2 curves very close to the pure components' spectral signatures, i.e., watermelon and coconut fat, respectively. Consequently, the scores obtained in the matrix of relative intensities (Fig. 4A) show, for example, for the coconut fat sample, values close to 0 for factor 1 and higher values for factor 2.

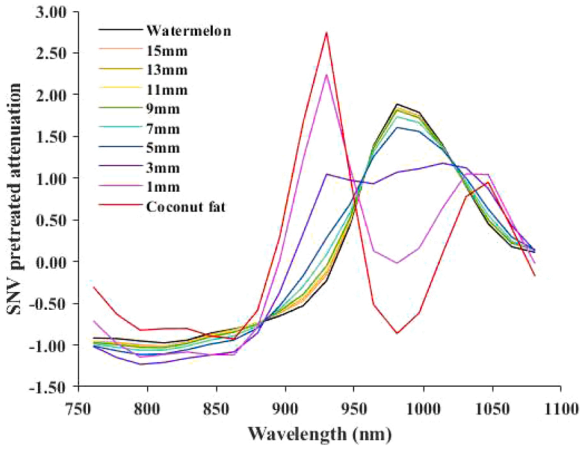


Fig. 3. SNV pre-treated data of the assay carried out using the watermelon red flesh, with Distance 4 of the NIR interactance instrument.

Fig. 5 A shows the percentage of fat signal, calculated as a function of the scores from the matrix of relative intensities, using the data of the assay with the spectral information from the light green area of the watermelon's rind. As can be seen, a total of around 60 % of the signal belonged to the coconut fat when the watermelon slice was 1 mm deep

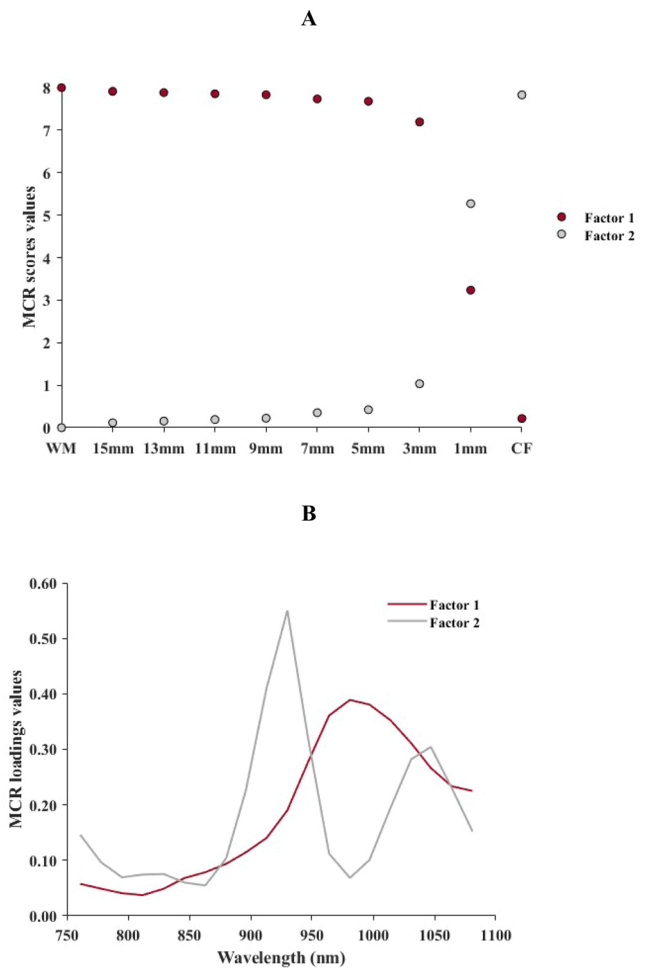


Fig. 4. Relative intensities (A) and pure spectra (B) matrices from the MCR-ALS model using the data taken from the light green area of the watermelon rind using Distance 4. WM = Watermelon; CF = coconut fat.

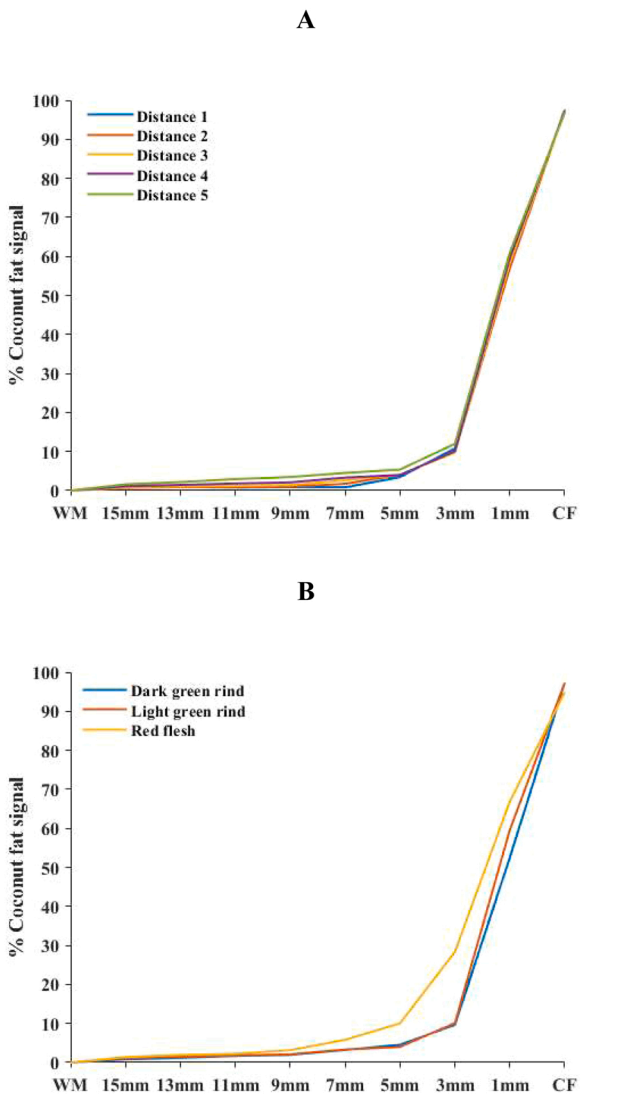


Fig. 5. Percentage of coconut fat signal captured in the assays carried out taking the information from the light green areas of watermelon rind considering the 5 distances tested (A), and from the light and dark green areas of the watermelon rind and red flesh using Distance 4 (B). WM = Watermelon; CF = coconut fat.

for all the distances. The thicker the slice, the lower the percentage of coconut fat detected by the sensor. For Distance 1, the percentage of coconut fat detected was close to 0 % at a depth of 7 mm, while for Distance 4 and Distance 5 the percentages of coconut fat detected were close to 0 % when the slice was thicker than 9 mm and 11 mm, respectively. Consequently, it is important to consider the position of the illuminated areas and the FOV of the instrument when carrying out NIR interactance tests, in order to adjust the intensity level of the signal captured and the amount of product sampled by the light travelling through it. When the assays were carried out using the spectral information from the light and dark green areas of the rind, no great differences could be seen in terms of the percentage of coconut fat captured (Fig. 5B). Nevertheless, a clear difference could be observed when the assays carried out in the rind were compared with those with the red flesh. In the latter case, there was a higher percentage of coconut fat signal captured in the slices from 1 mm to 9 mm (Fig. 5B). This could be explained by the greater amount of light absorbed by the watermelon rind compared to the red flesh, which makes it difficult to obtain information from deeper in the watermelon. However, as in the assays carried out in the rind, the percentage of fat signal captured was close to 0 % when the slice was thicker than 11 mm using Distance 4.

It is important to note that these results may differ if other experimental designs are considered or other instrument with different optical configurations are used. Consequently, the results obtained in this work were compared with similar studies aiming at the assessment of NIR light penetration in other thick rind fruit. Xu et al. (2018) measured the penetration of NIR light in the range 720–880 nm through the rind of 'Hami' melons, placing the light source and the detectors in contact with the sample at different distances (10, 20 and 30 mm). These authors also reported an increase in penetration depth, with an increased distance between the light source and the detector and indicated that the light can penetrate in melons from 11 mm (using a 10 mm distance between the light source and the detector) to 21 mm deep (using a 30 mm distance). It is important to note that the maximum penetration distances reported by these authors differ from that obtained in this study, due to the differences in the design of the experiments (i.e., the light source and the detector were not in contact with the sample), the spectral range used, and the characteristics of the samples analyzed.

3.3. Prediction of the SSC in intact watermelons using the PLSR method

The PCA of the spectral data acquired for each of the 5 distances assessed in this study revealed no spectral outliers from any of the 5 sets. Consequently, the 5 sets of samples used to develop the calibration models showed the same values for the SSC mean = 3.9 %, range = 2.5–5.9 %, SD = 0.8 % and CV = 20.5 %. Due to the fact that the

reference SSC values were taken from a portion of a 10 mm-deep sample (from the outer part of the watermelon removing the rind), the maximum value of SSC obtained was lower than 6 %. It must be highlighted that the sugar content in the pulp of this fruit varies from the rind to the center (Lammertyn et al., 2000; Magwaza and Opara, 2015), and the central part in general has a greater SSC compared to the outer parts of the pulp close to the rind.

Table 1 shows the segmented cross validation results for the models developed for each of the 5 distances tested. As can be seen from the RMSECV, the prediction ability of the models increased from Distance 1 to Distance 4. As the light traveled further, thus permitting a deeper optical sampling with Distance 4 compared to Distance 1, 2 and 3, a greater portion of the watermelon piece used to obtain the reference data was sampled. The results obtained with Distance 5 showed the greatest RMSECV, which is most likely due to the greater attenuation taking place when using that distance and thus the weak signal intensity received in that case. Fisher's F test only showed significant differences between RMSECV values for Distance 4 *vs* Distance 1 and 5 (Table 1). Distance 4 was selected as the optimal configuration for the quantification of the SSC content in intact watermelons using this NIR interactance instrument since the lowest RMSECV values were obtained for that distance. In addition, a lower number of latent variables (LVs) were needed for Distance 4 as compared to Distance 2 and Distance 5, which could indicate the greater difficulty involved in modeling the correlation between the NIR data and the reference SSC in the latter case. The lowest number of components was obtained for Distance 1, albeit with a poor correlation between the spectral and reference data, and therefore, the corresponding cross validation results were not good. Wold et al. (2021) also reported Distance 4 as the optimum configuration for monitoring the dry matter content of intact potatoes using the same NIR prototype system.

The Distance 4 set of samples was split into calibration and internal validation sets, whose characterization is shown in Table 2. The cross validation results shown in Fig. 6A ($R^2_{cv} = 0.73$; RMSECV = 0.39 %) showed a good fit to the PLSR model (Shenk and Westerhaus, 1996; Williams, 2001). The regression coefficients plot in the PLSR model (Fig. 7) shows peaks at around 778 nm and 980 nm, which can be related to the sugar OH stretching vibrations groups, while the peak at 913 nm corresponds to the third overtone associated to CH stretching, which, in turn, is related to the presence of carbohydrates (Osborne et al., 1993; Golic et al., 2003; Shenk et al., 2008; Bantadjan et al., 2020). When the model was validated using the samples not included in the calibration set, the SSC of those samples was predicted accurately (Fig. 6B), obtaining an $R^2_p = 0.81$ and an RMSEP of 0.30 %. The R^2_p value ($R^2_p > 0.6$), slope (0.9 < slope < 1.1), the RMSEP(c) (RMSEP(c) < 1.30 × RMSEC), the bias (bias < ± 0.6 × RMSEC) and the mean and SD values of the NIRS predictions (the mean and SD values corresponding to the reference and NIRS predicted values should not differ by more than 20 %) met the requirements established in the protocol established by Windham et al. (1989). It is important to note that good results were obtained despite the heterogeneity of watermelon included in this study, since the fruits were purchased from the Norwegian grocery market from August to November 2022, presented different shapes and sizes and were imported from different countries such as Brazil and Spain.

Similar studies have been carried out to measure the SSC in intact watermelons using NIR techniques, such as those performed by Flores et al. (2008) and Vega-Castellote et al. (2022), who worked with NIR instruments in the 400–1700 nm and 908–1676 nm ranges, respectively. These authors reported poorer prediction results than those obtained in this study. Nevertheless, when interpreting these results, it should be considered the fact that the spectral information in both research papers was taken in reflectance mode and from the outer part of the intact fruit (i.e., from the rind), whereas the SSC reference data were taken from different portions of the pulp of the fruit. This mismatch makes it difficult to find a good correlation between the NIRS data acquired from the outer part of the fruit and the reference SSC values, also considering the SSC gradients that exist in the watermelon pulp.

Table 1
Segmented cross validation results to predict the SSC (%) of intact watermelons using the 5 distances featured by the NIR interactance instrument.

Set	N	Mean	Range	[a]SD	[b]LVs	[c]RMSECV*	[d]R^2_{cv}
Distance 1	90	3.83	2.50–5.90	0.74	3	0.43[a]	0.66
Distance 2	89	3.81	2.50–5.90	0.73	9	0.42[ab]	0.69
Distance 3	86	3.79	2.50–5.90	0.72	5	0.39[ab]	0.70
Distance 4	89	3.81	2.50–5.90	0.73	5	0.36[b]	0.75
Distance 5	87	3.79	2.50–5.90	0.72	6	0.44[a]	0.63

* The coincidence of any of the superscript letters in the different RMSECV values indicates that no significant differences were found (P > 0.05) between those values.
[a] SD, standard deviation
[b] LVs, number of latent variables
[c] RMSECV, root mean squared error for the cross validation
[d] R^2_{cv}, coefficient of determination for the cross validation

Table 2
Number of samples, range, mean, standard deviation (SD) and coefficient of variation (CV) for the SSC (%) of the Distance 4 calibration and internal validation sets.

Set	N	Range	Mean	SD	CV
Calibration	85	2.5–5.9	3.9	0.8	20.5
Internal validation	10	3.1–4.9	3.9	0.6	15.4

4. Conclusions

The current study shows that the penetration depth of the NIR light is highly dependent on the chosen distance between the source of light and the FOV of the instrument, obtaining a maximum penetration of approximately 7 mm when the shortest distance is used (Distance 1), and approximately 9 and 11 mm when Distance 4 and Distance 5 are used, respectively. When the penetration of light in the red flesh of the fruit was assessed, a stronger signal emerging from the coconut fat placed underneath the watermelon flesh in the slices from 1 mm to 9 mm was observed, where a maximum penetration of approximately 11 mm was obtained.

When the predictive capacity of the models developed using the different distances tested were compared, it was clear that the optical sampling, when taking the spectral data, and the captured intensity were of maximum importance. The results obtained when Distance 4 was used reported the best results for the determination of the SSC (%) obtaining an $R^2_{cv} = 0.73$ and RMSECV = 0.39 % and an $R^2_p = 0.81$ and RMSEP = 0.30 % when the set of data for that distance was split into calibration and internal validation sets. Distance 4 was therefore selected as the optimum configuration for the NIR system when working with intact watermelons. In general, the results demonstrate that the accuracy of the calibration models depends significantly on the optical measurement geometry. Such instrument optimization is crucial when measuring complex heterogeneous samples using NIR-based systems.

Funding

This work was supported by the Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima Ås, Norway).

CRediT authorship contribution statement

Miguel Vega-Castellote: Data acquisition, Methodology, Formal analysis, Investigation, Software, Data curation, Validation, Writing - original draft, Writing - review & editing, Visualization. **María-Teresa Sánchez**: Conceptualization, Methodology, Validation, Investigation,

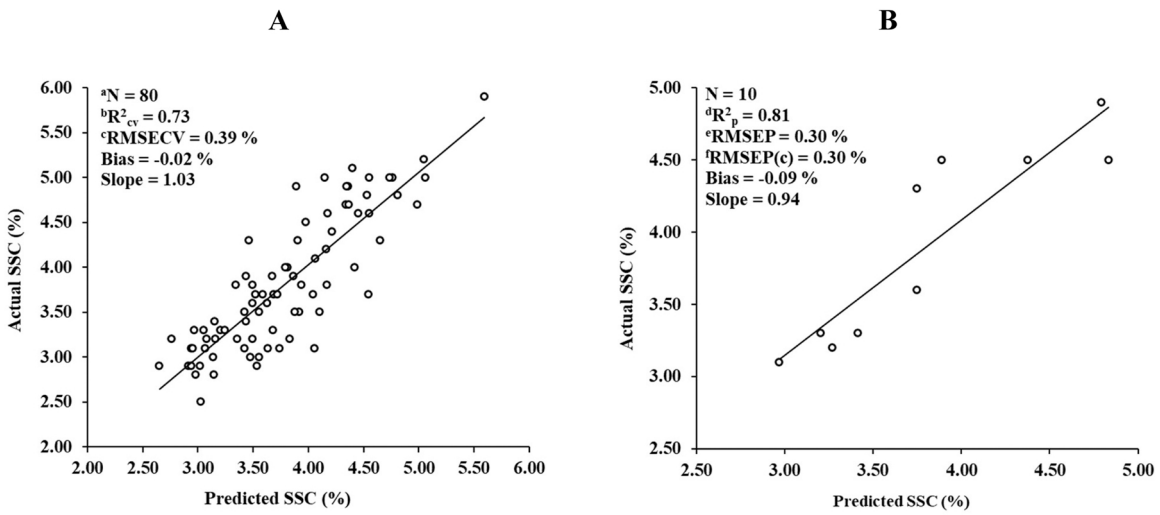


Fig. 6. Actual versus predicted data for the cross validation (A) and internal validation (B) assays for the prediction of soluble solid content (SSC) using the Distance 4 set of samples. [a]N, number of samples. [b]R^2_{cv}, coefficient of determination for the cross validation. [c]RMSECV, root mean squared error for the cross validation. [d]R^2_p, coefficient of determination for the internal validation. [e]RMSEP, root mean squared error for the internal validation. [f]RMSEP(c), root mean squared error for the internal validation corrected by bias.

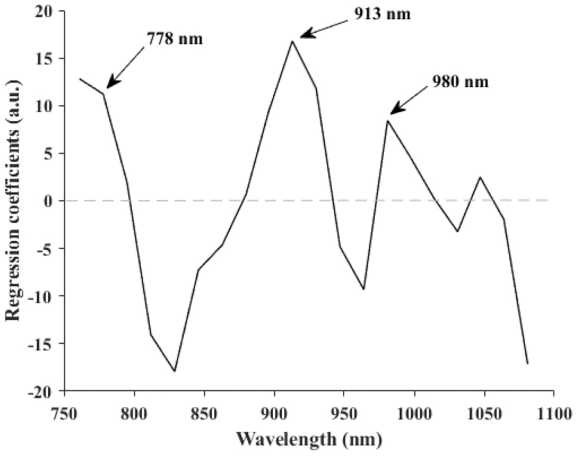


Fig. 7. Regression vector of the model developed to predict the soluble solid content (%) in intact watermelons using the set of samples included in the calibration set, for the assay carried out with Distance 4.

Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Funding acquisition. **Jens Petter Wold:** Conceptualization, Methodology, Investigation, Resources, Software, Data curation, Validation, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Nils Kristian Afseth:** Conceptualization, Methodology, Investigation, Resources, Software, Data curation, Validation, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Dolores Pérez-Marín**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper in any way.

Data Availability

Data will be made available on request.

Acknowledgements

The authors are grateful to the Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima Ås) for the financial support. Furthermore, the authors wish to express their gratitude to the Spanish Ministry of Universities for the support offered to Miguel Vega-Castellote through the Training Programme for Academic Staff (FPU).

References

Ali, M.M., Hashim, N., Bejo, S.K., Shamsudin, R., 2017. Rapid and nondestructive techniques for internal and external quality evaluation of watermelons: a review. Sci. Hortic. 225, 689–699. https://doi.org/10.1016/j.scienta.2017.08.012.

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. J. Food Eng. 217, 11–23. https://doi.org/10.1016/j.jfoodeng.2017.08.009.

Bantadjan, Y., Rittiron, R., Malithong, K., Narongwongwattana, S., 2020. Establishment of an accurate starch content analysis system for fresh cassava roots using short-wavelength near infrared spectroscopy. ACS Omega 5, 15468–15475. https://doi.org/10.1021/acsomega.0c01598.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43 (5), 772–777. https://doi.org/10.1366/0003702894202201.

Birth, G.S., 1978. The light scattering properties of food. J. Food Sci. 43, 916–925. https://doi.org/10.1111/j.1365-2621.1978.tb02455.x.

Cortés, V., Blasco, J., Aleixos, N., Cubero, S., Talens, P., 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: a review. Trends Food Sci. Technol. 85, 138–148. https://doi.org/10.1016/j.tifs.2019.01.015.

Dahm, D.J., Dahm, K.D., 2001. The physics of near-infrared scattering. In: Williams, P.C., Norris, K.H. (Eds.), Near-infrared Technology in the Agricultural and Food Industries. AACC Inc, St. Paul, MN, USA, pp. 1–17.

De Juan, A., Tauler, R., 2021. Multivariate curve resolution: 50 years addressing the mixture analysis problem – a review. Anal. Chim. Acta 1145, 59–78. https://doi.org/10.1016/j.aca.2020.10.051.

Fearn, T., 2009. Comparing standard deviations (continued). NIR News 20 (7), 24–25. https://doi.org/10.1255/nirn.1153.

Flores, K., Sánchez, M.T., Pérez-Marín, D.C., López, M.D., Guerrero, J.E., Garrido, A., 2008. Prediction of total soluble solid content in intact and cut melons and watermelons using NIR spectroscopy. J. Infrared Spectrosc. 16, 91–98. https://doi.org/10.1255/jnirs.771.

Fraser, D.G., Jordan, R.B., Künnemeyer, R., McGlone, V.A., 2003. Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy. Postharvest Biol. Technol. 27, 185–196. https://doi.org/10.1016/S0925-5214(02)00058-3.

Fraser, D.G., McGlone, V.A., Jordan, R.B., Künnemeyer, R., 2001. NIR (Near Infra-Red) light penetration into an apple. Postharvest Biol. Technol. 22 (3), 191–194. https://doi.org/10.1016/S0925-5214(01)00103-X.

Golic, M., Walsh, K., Lawson, P., 2003. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. Appl. Spectrosc. 57 (2), 139–145. https://doi.org/10.1366/000370203321535033.

Greensill, C.V., Walsh, K.B., 2000. A remote acceptance probe and illumination configuration for spectral assessment of internal attributes of intact fruit. Meas. Sci. Technol. 11, 1674. https://doi.org/10.1088/0957-0233/11/12/304.

Guthrie, J., Walsh, K., 1997. Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy. Aust. J. Exp. Agric. 37 (2), 253–263. https://doi.org/10.1071/EA96026.

Ibrahim, A., Daood, H.G., Égei, M., Takács, S., Helyes, L., 2022. A comparative study between Vis/NIR spectroradiometer and NIR spectroscopy for the non-destructive quality assay of different watermelon cultivars. Horticulturae 8 (509), 1–17. https://doi.org/10.3390/horticulturae8060509.

Jie, D., Wei, X., 2018. Review on the recent progress of non-destructive detection technology for internal quality of watermelon. Comput. Electron. Agric. 151, 156–164. https://doi.org/10.1016/j.compag.2018.05.031.

Lammertyn, J., Peirs, A., De Baerdemaeker, J., Nicolaï, B., 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Biol. Technol. 18, 121–132. https://doi.org/10.1016/S0925-5214(99)00071-X.

Mamouei, M., Budidha, K., Baishya, N., Qassem, M., Kyriacou, P.A., 2021. An empirical investigation of deviations from the Beer–Lambert law in optical estimation of lactate. Sci. Rep. Nat. Portf. 11, 13734. https://doi.org/10.1038/s41598-021-92850-4.

Mark, H., Workman, J., 2003. The F statistic. In: Mark, H., Workman, J. (Eds.), Statistics in Spectroscopy, 2nd ed. Academic Press, San Diego, CA, USA, pp. 205–211.

Martens, H., Naes, T., 1989. Multivariate Calibration. Wiley, New York, NY, USA.

Magwaza, L.S., Opara, U.L., 2015. Analytical methods for determination of sugars and sweetness of horticultural products–a review. Sci. Hortic. 184, 179–192. https://doi.org/10.1016/j.scienta.2015.01.001.

Mayerhöfer, T.G., Pahlow, S., Popp, J., 2020. The Bouguer-Beer-Lambert law: shining light on the obscure. ChemPhysChem 21, 2020–2046. https://doi.org/10.1002/cphc.202000464.

Naes, T., Isaksson, T., Fearn, T., Davies, A., 2002. A User-friendly Guide To Multivariate Calibration and Classification. NIR Publications, Chichester, UK.

Osborne, B.G., Fearn, T., Hindle, P.H., 1993. Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific and Technical, Harlow, Essex, England.

Pandiselvam, R., Prithviraj, V., Manikantan, M.R., Kothakota, A., Rusu, A.V., Trif, M., Khaneghah, A.M., 2022. Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: a comprehensive review. Front. Nutr. 9, 973457 https://doi.org/10.3389/fnut.2022.973457.

Qian, M., Wang, Q., Chen, L., Huang, W., Fan, S., Zhang, B., 2016. Penetration depth of near-inrfared light in small, thin-skin watermelon. In: 9th International Conference on Computer and Computing Technologies in Agriculture, CCTA, Part II, 479. IFIP Advances in Information and Communication Technology, Beijing, China, pp. 194–201.

Shenk, J.S., Westerhaus, M.O., 1996. Calibration the ISI way. In: Davis, A.M.C., Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, UK, pp. 198–202.

Shenk, J.S., Workman, J., Westerhaus, M., 2008. Application of NIR spectroscopy to agricultural products. In: Burns, D.A., Ciurczac, E.W. (Eds.), Handbook of Near Infrared Analysis, 3rd ed. Marcel Dekker, New York, NY, USA, pp. 347–386.

Sun, T., Huang, K., Xu, H., Ying, Y., 2010. Research advances in nondestructive determination of internal quality in watermelon/melon: a review. J. Food Eng. 100, 569–577. https://doi.org/10.1016/j.jfoodeng.2010.05.019.

Tamburini, E., Costa, S., Rugiero, I., Pedrini, P., Marchetti, M.G., 2017. Quantification of lycopene, β-carotene and total soluble solids in intact red-flesh watermelon (Citrullus lanatus) using on-line near-infrared spectroscopy. Sensors 17, 746. https://doi.org/10.3390/s17040746.

Tauler, R., 1995. Multivariate curve resolution applied to second order data. Chemom. Intell. Lab. Syst. 30, 133–146. https://doi.org/10.1016/0169-7439(95)00047-X.

Teixeira, A.M., Sousa, C., 2019. A review on the application of vibrational spectroscopy to the chemistry of nuts. Food Chem. 277, 713–724. https://doi.org/10.1016/j.foodchem.2018.11.030.

Vega-Castellote, M., Pérez-Marín, D., Torres-Rodríguez, I., Moreno-Rojas, J.M., Ordoñez-Díaz, J.L., Sánchez, M.T., 2023. Green, multivariate approach for obtaining a fingerprint of quality of watermelons at supermarket level using near infrared spectroscopy. LWT 182, 1–8. https://doi.org/10.1016/j.lwt.2023.114831, 114831.

Vega-Castellote, M., Sánchez, M.-T., Torres, I., De la Haba, M.-J., Pérez-Marín, D., 2022. Assessment of watermelon maturity using portable new generation NIR spectrophotometers. Sci. Hortic. 304, 1–7. https://doi.org/10.1016/j.scienta.2022.111328, 111328.

Walsh, K.B., Guthrie, J.A., Burney, J.W., 2000. Application of commercially available, low-cost, miniaturized NIR spectrometers to the assessment of the sugar content of intact fruit. Aust. J. Plant Physiol. 27, 1175–1186. https://doi.org/10.1071/PP99111.

Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P.C., Norris, K.H. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists Inc., St. Paul, MN, USA, pp. 145–169.

Windham, W.R., Mertens, D.R., Barton II, F.E., 1989. Protocol for NIRS calibration: sample selection and equation development and validation. In: Martens, G.C., Shenk, J.S., Barton,II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, 643. USDA-ARS. Government Printing Office, Washington DC; USA, pp. 96–103.

Wold, J.P., O'Farrell, M., Andersen, P.V., Tschudi, J., 2021. Optimization of instrument design for in-line monitoring of dry matter content in single potatoes by NIR interaction spectroscopy. Foods 10 (828), 1–14. https://doi.org/10.3390/foods10040828.

Xu, L., Li, J., Zhang, D., 2018. Near-infrared light penetration depth analysis inside melon with thick peel by a novel strategy of slicing combining with least square fitting method. J. Food Process. Eng. 41, 1–6. https://doi.org/10.1111/jfpe.12886, 12886.