Measurement 256 (2025) 118481

Contents lists available at ScienceDirect

Measurement

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

Understanding variability and calibration challenges in NIR miniaturized sensors: the impact of particle size and analytical session in almond powder analysis

Barbara Giussani [a], Soumaya Ouzakar El Ghouch [b], Manuel Monti [a], Jordi Riu [b,*]

[a] Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio 9, 22100 Como, Italy
[b] Universitat Rovira i Virgili, Department of Analytical Chemistry and Organic Chemistry. Carrer Marcel·lí Domingo 1, 43007 Tarragona, Spain

ARTICLE INFO

Keywords:
Almond powder
Particle size
NIR miniaturized sensors
PLS regression

ABSTRACT

The increasing application of miniaturized Near-Infrared (NIR) sensors highlights their potential for rapid, non-destructive, and cost-effective analysis, particularly in food industry. These portable instruments are often marketed as easy-to-use solutions, intended for use by non-specialists rather than analytical chemistry experts, which has contributed to their widespread adoption. This study investigates the contamination of bitter almond in almond powder using various low-cost miniaturized NIR sensors, including the SCiO sensor, two NeoSpectra Micro Development Kits, and the NeoSpectra Scanner, with and without the Rotator accessory. Almond powders with different levels of contamination of bitter almond (0–100 wt%) were analysed, and Principal Component Analysis (PCA) was used as an initial data screening step, showing the importance of particle size, thus providing a valuable quality control in this type of measurements. Partial Least Squares (PLS) regression models were developed to predict the percentage of contamination of bitter almonds and to evaluate the performance of each NIR sensor. The best regression models were obtained using the NeoSpectra Scanner spectrometrer being to predict concentration values with an error around 2.5% and a limit of detection around 4.5% of bitter almond in almond powder. Performance discrepancies were observed between sensors of the same type and model, as well as across different experimental sessions. These results emphasize the importance of understanding the limitations of miniaturized NIR sensors, while also highlighting their effectiveness, affordability, and portability, which make them a valuable and reliable tool for on-site food safety applications.

1. Introduction

NIR spectroscopy using miniaturized sensors is a constantly developing field that has demonstrated its effectiveness in numerous applications and has addressed significant analytical challenges in various sectors, including food analysis [1–4]. The use of miniaturized NIR instruments is rapid, cost-effective, non-destructive, and requires minimal chemicals, thus aligning with most of the principles of green chemistry [5]. NIR spectroscopy data are typically studied in combination with multivariate analysis due to their multivariate nature [6].

The field of miniaturized NIR spectroscopy instrumentation is actively under development, with analytical measurements and analytical strategies varying significantly between instruments and types of samples. Most currently available miniaturized NIR sensors are not optimally designed for powder samples [1], posing a challenge in applications like food analysis. In fact, it is crucial to select the appropriate type of miniaturized NIR instrument based on the specific characteristics of the sample and the intrinsic properties of the instrument, as not all instruments are suitable for every type of sample. An adequate selection of the proper instruments ensures an optimal analytical performance [7]. Reflectance and transflectance are typical measurement modes in NIR spectroscopy, both of which are influenced by particle size due to its effect on light scattering. The effect of particle size in NIR reflectance spectroscopy has been evaluated in various studies and across different fields using benchtop instrumentation (e.g. [8–11] to name a few studies), and various signal treatments and transformations have been assessed to correct for such effects [12–14]. However, the situation changes when dealing with portable NIR instrumentation, and there are very few studies in the literature addressing the impact of particle size on powder analysis, discussing the relationship between

* Corresponding author.
E-mail address: jordi.riu@urv.cat (J. Riu).

https://doi.org/10.1016/j.measurement.2025.118481
Received 3 March 2025; Received in revised form 14 July 2025; Accepted 18 July 2025
Available online 20 July 2025

particle size and the spectroscopic data in the NIR region [15], when the particle size of the powder may produce multivariate models of different quality, either for classification or prediction purposes. This issue is further exacerbated when considering the increasing number of individuals working with portable NIR instruments who are not spectroscopists but rather are drawn to portable NIR spectroscopy due to its low-cost measurement instruments, which are accessible to a much broader audience. However, this low cost does not imply that it is a simple or easy-to-use technique; on the contrary. It should also be noted that, due to the relative novelty of many of these instruments, even different units of the same brand and model may introduce measurement errors with varying structures [16], which could further complicate their use by personnel who are not properly trained.

Among nuts, almonds are one of the most used products with also a wide range of applications in the food industry. Consumption of nuts is increasingly recognized for its numerous health benefits, making them an essential component of a balanced diet. Nuts are nutrient-dense foods, rich in unsaturated fatty acids, high-quality protein, fiber, vitamins (such as vitamin E and B vitamins), minerals (including magnesium, potassium, and calcium), and various bioactive compounds [17–19]. Regular consumption of nuts has been associated with a reduced risk of chronic diseases, including cardiovascular disease, type 2 diabetes, and certain cancers [20–22]. The high content of antioxidants and anti-inflammatory properties in nuts contribute to improved cardiovascular health by enhancing endothelial function, reducing oxidative stress, and lowering serum cholesterol levels [23–26]. In particular, almonds are used extensively in various forms, such as whole nuts, almond butter, almond milk, and almond powder. Almond powder is valued for its use in baking, confectionery, and as a gluten-free alternative to wheat flour.

The global almond market has seen substantial growth over the years. According to the US Department of Agriculture [27] and the International Nut & Dried Fruit [28], the annual global production of almonds reached approximately 1.51 million metric tons in 2023/24. The United States is the largest producer, contributing almost with 80 % of the world's almonds, primarily from California. Spain and Australia are also significant producers. The worldwide consumption of almonds has mirrored this growth, driven by increasing consumer awareness of the health benefits associated with almond consumption, which includes, among others, fiber, protein, vitamin E and low levels of saturated fatty acids [29].

However, the presence of bitter almonds, which contain amygdalin, a compound that can produce toxic cyanide when metabolized [30], poses a potential risk to consumers. Cyanide can be harmful, leading to symptoms such as headache, dizziness, and in severe cases, respiratory failure, or death. From the organoleptic point of view, amygdalin possesses a characteristic bitter taste that is typically unpleasant for the consumer, and bitter almonds in almond powder may lead to the rejection of the product. Therefore, accurate detection of the percentage of bitter almond in almond powder is crucial for food safety and quality control. The distinct sweet or bitter taste of almond kernels is a hereditary trait controlled by a single gene, with bitterness being a recessive characteristic. Globally, sweet almonds are the most widely cultivated, and breeding programs consistently aim to develop trees that produce only sweet kernels, gradually eliminating the genetic factors responsible for bitterness. However, certain long-established and commercially viable almond varieties still carry these alleles, which can result in seedlings with bitter kernels when combined [31].

This article aims to address the problem of determination of bitter almond in almond powder, proposing a rapid and low-cost methodology based on NIR spectroscopy using miniaturized sensors, together with multivariate calibration, for the modern food industry. It is crucial to ensure that bitter almonds are not inadvertently mixed with sweet almonds intended for consumption, as this can pose significant health risks and economical losses. Contaminating the almond powder with bitter almond alters its NIR spectrum by incorporating components responsible for the bitter taste, such as amygdalin (which is present in bitter almonds, in some cases probably in concentrations not detectable by portable NIR instruments), as well as other structural and physiological characteristics of bitter almonds that differ from sweet almonds.

Vibrational spectroscopy was already applied to the chemistry of nuts [32], and for instance FT-IR and FT-NIR instrumentation together with classification techniques were successfully used in the classification of almond powder adulterated with components such as apricot and peanut powder [33]. Portable instrumentation (with different degrees of portability) was used for the detection of bitter almond kernels in sweet almond batches [34], for the classification and prediction of amygdalin content and the classification by bitterness of in-shell and shelled intact almonds [35], for the detection of bitter almonds in batches of sweet almonds [36], and in the detection of adulteration of almond powder with low-cost flours [37].

This study was designed with several key objectives in mind. First, the primary aim was to evaluate the ability of portable Near-Infrared (NIR) sensors to detect the presence of bitter almond in almond powder. To do so, we compared five different measurement systems representing three categories of portable NIR sensors, in order to assess their relative performance and suitability for this specific analytical task. A crucial element of the investigation involved analysing the influence of particle size on the measurements, since the reflectance spectra obtained from powdered samples can be significantly affected by variations in granularity. Understanding this effect is essential for building robust predictive models. Similar attention was devoted to assessing how the analytical session influenced the measurements.

Sweet almond powder (or simply, almond powder) of three different particle sizes (5, 2 and 1 mm approximately) was added (percentages from 0 % to 100 %) with bitter almond and analysed in two independent analytical sessions with all the NIR sensors. Before proceeding with quantitative predictions, we conducted a thorough preliminary data analysis using Principal Component Analysis (PCA), a standard approach that allows for the qualitative exploration of patterns within the data. This step is fundamental in any chemometric workflow, as it helps to verify whether sample characteristics—such as particle size or analytical session—can be clearly distinguished in the spectral data, thus ensuring that any predictive modelling is grounded in solid data structure. To achieve reliable quantification of bitter almond content, multivariate prediction models based on PLS (Partial Least Square) regression were subsequently developed to quantify the percentage of change relative to pure almond powder due to the presence of bitter almond.

Beyond the technical findings, this work also supports broader goals. One of them is to demonstrate the potential of portable, miniaturized NIR sensors as practical tools for on-site, green analytical chemistry, thanks to their ability to analyse samples in their original form without the need for reagents or sample preparation. Moreover, the study highlights that, despite the great potential of these sensors, their practical application is more complex than commonly assumed. Achieving reliable performance requires careful optimization of measurement conditions and a thorough evaluation of experimental variables.

2. Materials and methods

2.1. Instruments and materials

2.1.1. Spectrometers

Five different measurement strategies including four different miniaturized NIR instruments were used for the analysis of almond powder.

The SCiO (Consumer Physics, Herzliya, Israel) is a miniaturized NIR device featuring a molecular sensor with dimensions of 67.7 mm × 40.2 mm x 18.8 mm and a weight of 35 g. It operates within a wavelength range of 740 to 1070 nm and was controlled via an Android smartphone using the 'SCiO Lab' application over Bluetooth. SCiO features a

dispersive element, a Osram broadband IR led and a photodiode array. The default scan time for SCiO measurements is between 2 and 5 s, and this duration cannot be adjusted manually. The spectra are stored in the cloud and can then be downloaded from 'The Lab' website (https://thelab.consumerphysics.com). The resolution of the SCiO device is not disclosed by the manufacturer [4]. Calibration of the SCiO is required before the first measurement of each session using the specific reference standard located on the back side of the cover of the instrument.

Two different NeoSpectra microdevelopment kits (MDK) (Si-Ware, Cairo, Egypt) were also used. The NeoSpectra MDK features a monolithic microelectromechanical system (MEMS) Michelson interferometer, a single InGaAs photodetector and three halogen lamps. The device measures 32 mm x 32 mm x 22 mm and weighs 17 g. It operates within a wavelength range of 1350 to 2558 nm with a resolution of 16 nm. The NeoSpectra MDK is connected to a Raspberry Pi single-board computer, which acts as a host and enables connection to a laptop via a universal serial bus (USB). The software, compatible with Windows and Linux, allows users to configure a limited number of parameters such as scan time, run mode (single or continuous), and data interpolation for each collected spectrum. Calibration of the NeoSpectra using a reference is required each time the software is initiated. The scan time has been optimized to 5 s without data interpolation. It is worth mentioning that the two instruments, purchased at different times, have different resolutions although operating in the same wavelength range. In other words, the wavelengths recorded by each instrument are not exactly the same.

The NeoSpectra Scanner (Si-Ware, Cairo, Egypt) measures 180 mm x 45 mm x 8 mm and weighs 730 g and operates within a wavelength range of 1351 to 2559 nm. The NeoSpectra Scanner features a monolithic MEMS Michelson interferometer, a single InGaAs photodetector and seven halogen lamps. The sensor was operated by the proprietary mobile application on an Android phone using Bluetooth connection. The spectrometer allowed the acquisition under charge and standing on its own battery. A scan time of 5 s without data interpolation was used. A Rotator accessory was also used in the measurements with the NeoSpectra Scanner. This accessory automatically rotates the sample on top of the device, enabling the spatial averaging of non-homogeneous samples [38] in a way that seeks to improve the representativeness of the area exposed for analysis, as claimed by the producer.

The wavelength selector in Si-Ware Systems instruments (Scanner and MDK) is a monolithic MEMS Michelson Interferometer, which means that the equipment is designed to collect an interferogram that is then Fourier-transformed to obtain the spectrum. In the case of the SCiO, its wavelength selector utilizes bandpass filters. Therefore, the different instruments are based on different operating principles, which may influence their performance, quality parameters, and applications. A recent review describes in detail the technical characteristics of the main miniaturized NIR spectrometers [4].

Reflectance mode analysis was utilized for all the instruments. Before any sample, a background measurement using a 99 % reflectance material with a Spectralon® standard is performed. For the SCiO and the NeoSpectra Scanner, the reflectance material is integrated into the back of the cover. Analytical sessions with the two NeoSpectra MDK sensors and the NeoSpectra Scanner with and without the Rotator accessory began after a 20-minute warm-up period. Analytical sessions with SCiO started immediately without any warm-up period. The analytical sessions were conducted aiming to reproduce actual laboratory working conditions. The two analytical sessions were performed by the same analyst, in the same specific location within the laboratory, under the same ambient temperature and lighting conditions, but these were not strictly controlled.

All the measurements were made in contact mode between the instrument and the sample holder except measurements performed with the SCiO sensor. For the two NeoSpectra MDK sensors, a custom-made cell served as the sample holder [15]. The sample holder contained about 1.3 g of almond powder. The lateral sides of the holder were made using a MakerBot Replicator® 2 Desktop 3D Printer (MakerBot Industries, New York, NY, USA), while interchangeable coverslips made of borosilicate glass, measuring 22 mm x 22 mm and with a thickness ranging from 0.13 to 0.17 mm (Knittel Glass, Bielefield, Germany), were employed at the intermediate bottom of the cell. For the measurements conducted with the SCiO the distance between the sensor and the sample was 7 mm. Fig. S1 in the Supplementary Materials shows the experimental configuration for these measurements.

2.1.2. Other material

A coffee grinder (Black + Decker BXCG150e, Oliana, Spain) was employed to grind the almonds and the different particle sizes were obtained using commercial sieves (Lacor 68342, Bergara, Spain).

2.2. Reference samples

Whole sweet almonds without skin were purchased in local markets in Tarragona (Spain). Whole bitter almonds with skin were purchased from Schmuetz Naturkost (Malente, Germany). To remove the skin from bitter almonds, water was brought to a boil and then bitter almonds were immersed for 1 min. Subsequently, the skin was manually removed, and the almonds were air-dried for one day. The almonds did not lose their hardness, and they were physically unaltered.

Three different particle sizes (named large, medium and small in this article) were used in the production of reference samples. The particle sizes (Fig. S2 in Supplementary Material) were obtained using a coffee grinder and sieving the ground material using a commercial sieve through three different mesh sizes (5 holes/inch, 12 holes/inch, 22 holes/inch: 5, 2 and 1 mm approximatively). It is important to recall that the distribution of individual particles within each particle size was not the focus of this study, as the objective was to determine whether differentiating the three particle sizes (which are macroscopically different from one another) affected the prediction of the percentage of bitter almond in almond powder and the performance of each miniaturized instrument. The grinding process was carried out at room temperature and has been optimized to prevent the temperature of the sample from increasing, at least not to a level detectable by a food thermometer used during the whole process. In this way, we should have ensured chemical stability within the samples. The samples were ground and sieved in a single session so that all the almonds underwent the same treatment. For each particle size, 20 reference samples were prepared with the following concentrations of contamination of bitter almond: 0 %, 1 %, 2 %, 4 %, 6 %, 8 %, 10 %, 12 %, 14 %, 16 %, 18 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 75 %, 100 % in weight. In all cases the samples were prepared with sweet and bitter almonds of the same particle size. The reference samples were then labelled and stored in the refrigerator at a temperature of 5 °C in zip-lock bags. Before taking measurements, the samples were kept at room temperature.

Each experiment was replicated twice (in independent sessions), resulting in two calibration models for each sensor and powder size. For the NeoSpectra Scanner equipped with the Rotator, only one calibration model was created for each sample size.

2.3. Statistical data analysis

Calculations were performed using PLS Toolbox 9.3.1 (Eigenvector Inc., Manson,WA, USA) running under the Matlab 2024a (Mathworks Inc., Natick, MA, USA) environment. 5 analytical replicates (measurements with resampling) were made for each reference sample, particle size and sensor. The average of these analytical replicates was used for the calculations. Principal Component Analysis (PCA) was used for the first qualitative study of the data. A model was built for the data collected from each spectroscopic instrument to verify if there were any peculiar samples and, in general, to identify any issues with the spectroscopic data. PLS regression was employed to predict the percentage of bitter almonds in almond powder. Ten reference samples were used in

the calibration step, and the remaining samples were used to externally validate the model. The reference samples used in the calibration step are 0 %, 1 %, 4 %, 8 %, 12 %, 16 %, 20 %, 35 %, 75 % and 100 %, and the test samples used are 2 %, 6 %, 10 %, 14 %, 18 %, 25 %, 30 %, 40 %, 45 %, and 50 %. The samples were selected trying to span all the calibration range and focussing the predictions in the lower-middle part of the range, where contaminations are more probably to take place. This split avoids the possibility of predicting validation samples outside the training interval (a scenario that could have arisen with a random division of the training and test sets), which would effectively mean performing extrapolation and potentially lead to unreliable outcomes.

Several spectral preprocessing techniques were evaluated to improve prediction model performances. Standard normal variate (SNV), multiplicative scatter correction (MSC), baseline (weighted least squares) to remove baseline offsets, Savitzky-Golay smoothing (from 7 to 25 smoothing points), and derivatives (first and second order) were tested. A recent paper dealing also with almond powder identified these preprocessing techniques as the most adequate to deal with this type of samples [39]. Prior to any model calculation, data were mean-centered. PLS regression models were evaluated looking at the RMSEC (root-mean square error of calibration), the RMSEP (root-mean square error of prediction), the RMSECV (root-mean square error of cross-validation), the r^2 value of the regression line between the predicted and the measured values, the LOD (limit of detection) and the number of latent variables used. The best PLS models were selected as the best compromise between the lower RMSEP, the higher r^2 value and the smaller number of latent variables needed. The Cross-Validation (CV) used was a random CV.

It is worth recalling that two independent experiments were conducted for each sensor (except for the one using the NeoSpectra Scanner with the Rotator accessory) in two different sessions, resulting in two models for each instrument. To verify the robustness and reliability of the method, the following were performed:

- the model built during the first experimental session was validated using the validation samples from the second experimental session, and vice versa. This was done to assess whether the prediction error varied significantly, indicating whether it was dependent on the analytical session.
- a linear approach using the joint confidence interval for the intercept and slope [40] was used to assess if the predictions of the two experimental sessions were comparable between them.

The multivariate detection limits were determined using the approximate formula for the sample-specific standard error of prediction, as proposed by Faber and Bro [41].

3. Results and discussion

3.1. Spectral data and figures of precision

Fig. 1 shows the average spectra of almond powder (for the sake of clarity, in figures a-j only the percentages of adulteration with bitter almond for the calibration samples are shown, and only results for the first experimental session are presented), the relative standard deviation and the root mean square (RMS) of the replicate spectra corresponding to the large particle size. Figs. 2 and 3 show the same pictures but for the medium and the small particle size, respectively. The RMS is calculated following the procedure described in different manuscripts [42–44] and used to evaluate the spectral variability of spectroscopic measurements. Briefly, the RMS for an individual spectrum within a set of replicates is defined as the square root of the mean of the squared differences between the reflectance values of that specific spectrum and the mean spectrum, calculated over the entire wavelength range. Then, the individual RMS values for all replicate spectra of a given sample (e.g., a specific percentage of adulteration, particle size, and instrument) are averaged to obtain a single, representative RMS value for that sample. For the measurements conducted with the SCiO, which does not require direct contact with the sample, the distance between the sensor and the sample was optimized to maximize the signal in quantitative terms while ensuring the best possible precision (standard deviation of the spectrum). The instrument was held vertically in relation to the sample, as in our experience any tilt of the sensor can alter the angle of light incidence, leading to substantial variations in the intensity of the

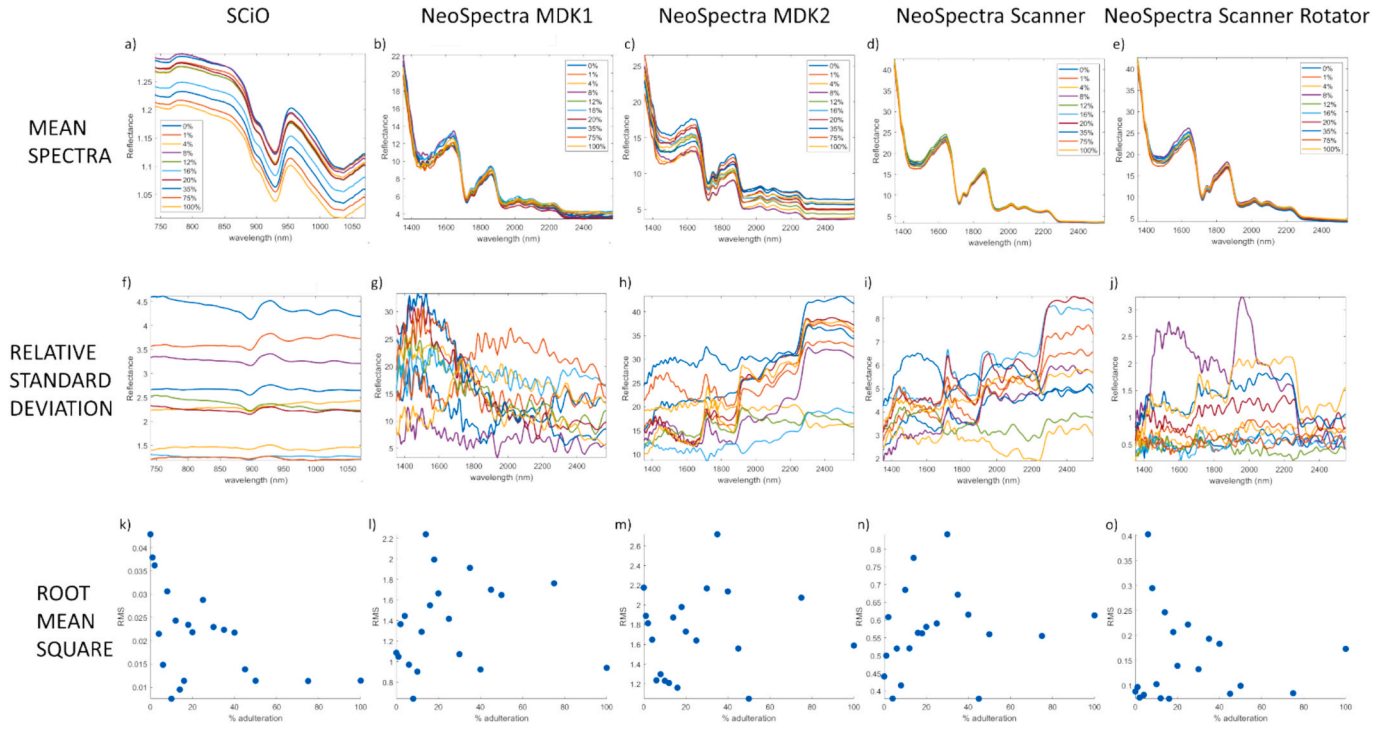

Fig. 1. Spectra and figures of precision for the large particle size. The legends in figures f–j have been removed for clarity, and it is the same than in figures a–e.

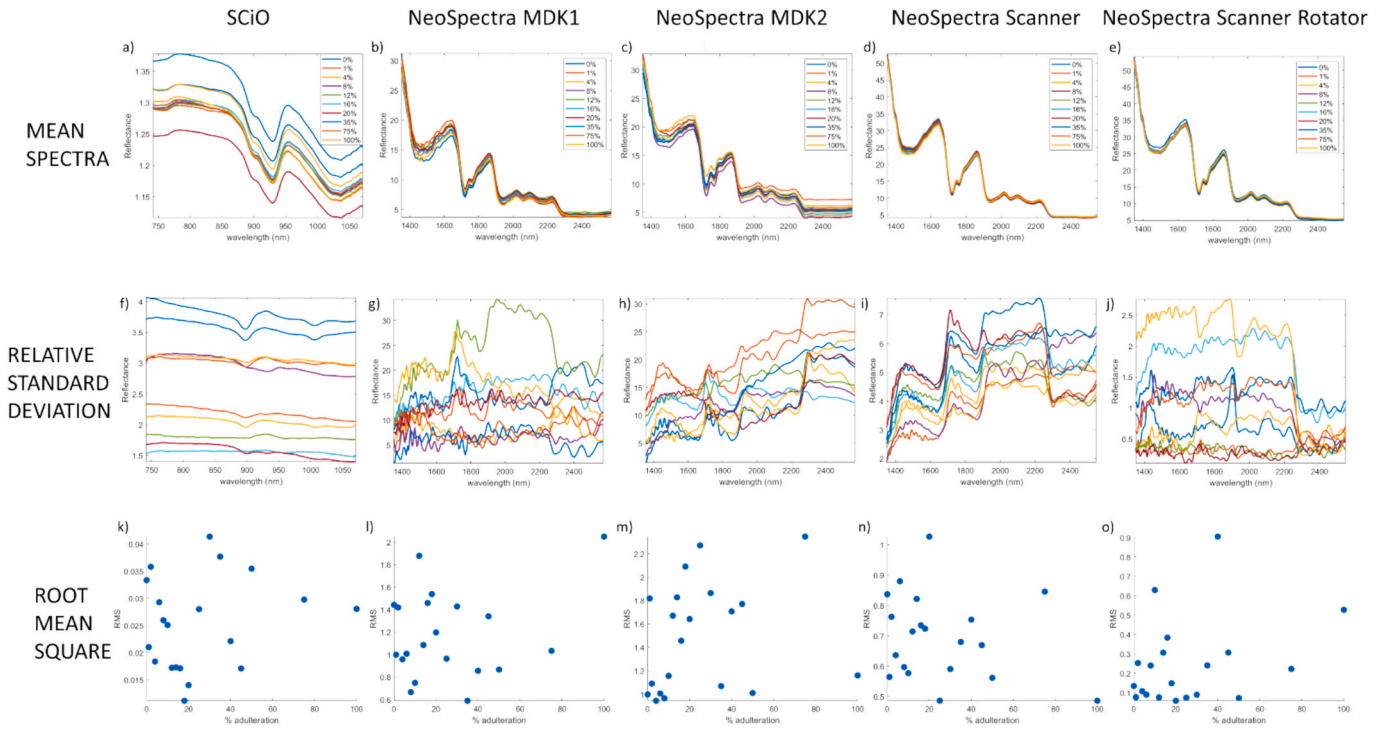

Fig. 2. Spectra and figures of precision for the medium particle size. The legends in figures f–j have been removed for clarity, and it is the same than in figures a–e.

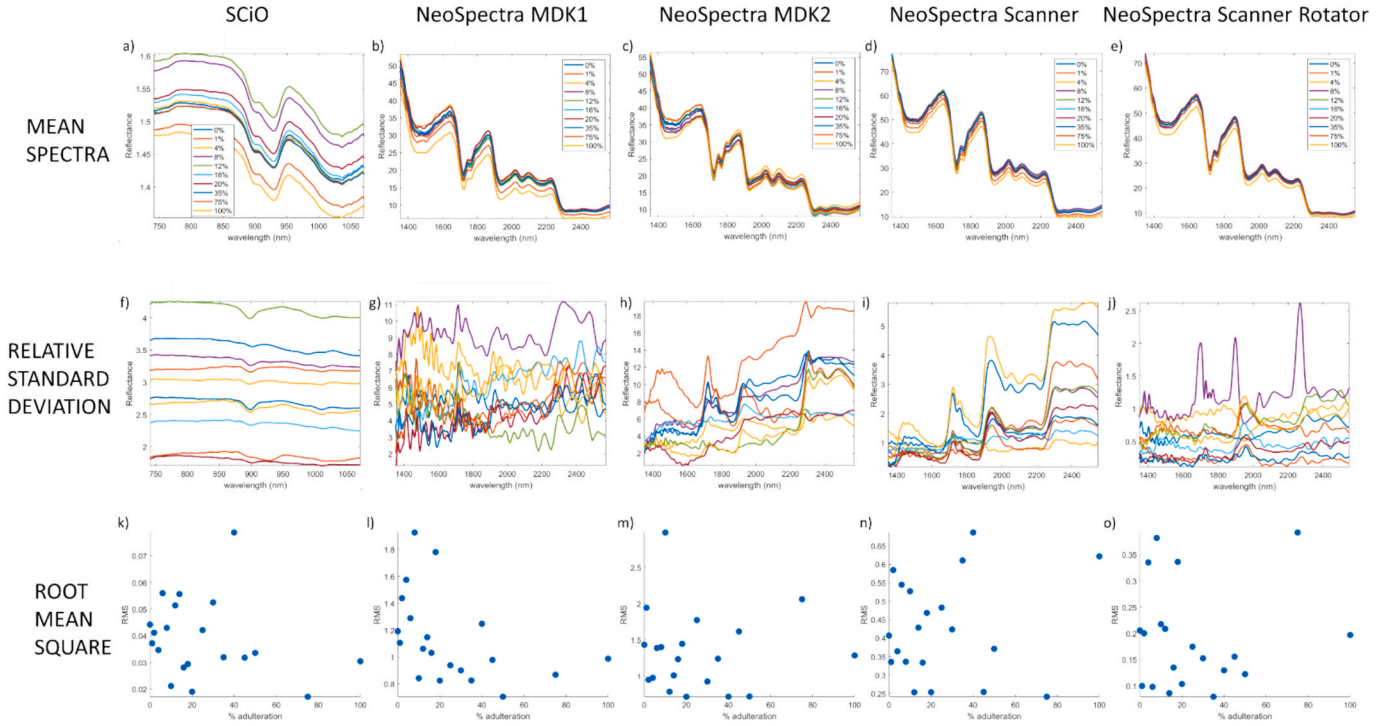

Fig. 3. Spectra and figures of precision for the small particle size. The legends in figures f–j have been removed for clarity, and it is the same than in figures a–e.

collected signal [15,45]. An optimum distance between sample and sensor of 7 mm was found. From Figs. 1 to 3 it is evident that it is challenging to find a trend relating the percentage of adulteration with reflectance values. This finding is not surprising, since in reflectance measurements of powder samples, light scattering is a dominant factor influencing the measurements, and light scattering is a random process among samples of the same particle size. No qualitative changes are apparent in the spectral shape, which remains qualitatively similar across all particle-size fractions within measurements made with the same sensor.

Figs. 1–3a, corresponding to the SCiO spectra, show two peaks around 930 nm and 1038 nm ascribable to the third overtone C–H stretch vibrations of triglycerides [46]. In the remaining figures, corresponding to the NeoSpectra spectra, the aliphatic hydrocarbon part of the lipids seemed to appear at various peaks: at about 1725 nm is located the C–H stretch first overtone of methylene while the symmetric CH_2

bonds vibration appears at around 1760 nm. At around 2047 nm protein vibrations are located, possibly due to the amide combination band of $CONH_2$. The broad band in the range 1400–1640 nm could be ascribed to the N–H stretching (first overtone) of proteins and the water association absorbance. A combination band of water also appears around 1940 nm possibly due to the O–H stretching and H-O–H bending combination. The area between 2100–2500 nm corresponds to the CH_2 combination bands, associable to fatty acids. Finally, between around 1870 nm and 2030 nm, the different bands correspond to O–H first overtone, O–H stretching and N–H amide and C=O of amide stretching and combinations, given by peptides and proteins [47,48]. Fig. S3 in the Supplementary Materials shows graphically these bands. Similar peaks, but within a somewhat different wavelength range and for whole almonds, have also been reported [49]. A paper by Borràs et al. [50] identified the wavelengths responsible for class separation when measuring whole sweet and bitter almonds. The authors did not attribute these wavelengths to specific compounds such as amygdalin, likely reinforcing the idea that, although the primary component responsible for the bitterness of bitter almonds is amygdalin, the differences in the spectra between bitter and sweet almonds may also be due to other compounds or structural and physiological characteristics of bitter almonds.

Due to the different technological characteristics of the various instruments tested, such as acquisition window (about 1 mm^2 for NeoSpectra MDK, about 48 mm^2 for SCiO and about 10 mm^2 for NeoSpectra Scanner), acquisition configuration and also spectroscopic range, that is different for one of the instruments, the spectra obtained for the samples are different in shape and related characteristics.

It is worth noting that while the NeoSpectra instruments, which share the same manufacturer, express spectra using the same units of measurement, the data produced by the SCiO are not directly comparable: each manufacturer uses its own standards to describe the signal produced by its instrument. Lower reflectance values were obtained with the NeoSpectra MDK (both sensors) compared to the NeoSpectra Scanner. These results could be due to the differences in the spot and detector sizes (NeoSpectra Scanner holds a larger spot size up to 10 mm to cope better with non-homogenous material) and to the intensity of light sources [51]. Looking at the variation of reflectance values with the particle size, it is possible to see that it mainly affects the NeoSpectra instruments, with highest reflectance values for the small particle size (Fig. 3b and e) and lowest reflectance values for the large particle size (Fig. 1b and e). Despite the lowest reflectance values obtained with SCiO, their relative standard deviation profiles (Figs. 1-3f) present the lowest values, only outperformed by NeoSpectra Scanner equipped with Rotator (Figs. 1–3j). In the case of NeoSpectra Scanner equipped with Rotator, the use of this accessory enables automatic rotation of the sample on top of the device and probably enhance representativeness of the area exposed for analysis, what helps in increasing representativity and enhancing the associated precision. The highest relative standard deviation values are obtained with NeoSpectra MDK1 (Figs. 1-3g) and NeoSpectra MDK2 (Figs. 1-3h). In all cases (except for SCiO), the relative standard deviation values decrease with the particle size: higher values for large particle size and lower values for small particle size.

Figs. 1-3k and o show the RMS valued for the different particle sizes. The RMS statistics computes the agreement between spectra from different replicates. As we have mentioned before, the RMS values produced by SCiO are not directly comparable to those from the other instruments, since the spectra are expressed in different measurement units. In the case of the SCiO, the highest RMS values are obtained for the small particle size and the lowest values for the large particle size, although the difference between the large and medium sizes does not appear to be significant. For the other instruments, the best (i.e., lowest) RMS values are found for the small particle size, with no clear trend observed between the large and medium sizes. The two NeoSpectra MDKs produce similar RMS values, although the NeoSpectra MDK2 generally yields slightly higher values. The NeoSpectra Scanner provides better (i.e., lower) RMS values than the NeoSpectra MDKs, and the NeoSpectra Scanner equipped with the Rotator accessory improves upon the values of the NeoSpectra Scanner alone.

Regarding the trend between the RMS values and the percentage of bitter almond adulteration, no overall trend is observed. This seems to reinforce the aforementioned point that light scattering is a dominant factor influencing the reflectance measurements, and that it is a random process among samples of the same particle size.

To better study the precision considering the different percentages of bitter almond in almond powder, Fig. 4 shows the average relative standard deviation for each value for all the instruments and particle sizes. Only results of the first experiment (first session) are presented in Fig. 4 (the results of the second experiment agreed with those presented). Each individual value has been obtained finding the standard deviation of the analytical replicates at all wavelengths, and then averaging all these standard deviations.

Fig. 4a, corresponding to the SCiO results, does not appear to indicate specific trends either in terms of particle size or percentage of bitter almonds. The mean relative standard deviation for all the particle sizes, excluding the largest individual values, are 2.27 %, 2.59 % and 3.07 % respectively for large, medium and small particle sizes. The unusually high values are attributed to experimental errors (e.g. when within the replicates, one replicate shows a value completely different from the others). These values have not been removed to give the reader an idea of how many measurements might be affected by errors, which, in the regular use of these instruments, should either be repeated or at least excluded from calculations. They have, however, been removed for the calculation of the mean relative standard deviations reported in the following paragraph.

NeoSpectra MDK1 and MDK2 show a clear trend with the particle size. The mean relative standard deviations for all the particle sizes (excluding again the largest individual values), respectively for large, medium and small particle sizes, are 17.72 %, 11.91 % and 5.32 % for NeoSpectra MDK1, and 19.70 %, 12.78 % and 5.86 % for NeoSpectra MDK2. The data between the two sensors agree and show the same trend. It is possible to clearly see how the relative precision is worse as the particle size increases. For the NeoSpectra Scanner and the NeoSpectra Scanner equipped with Rotator, there seems to be a slight difference between the results for the large and the medium particle size, and a larger difference compared to the small particle size. Excluding again the largest individual values, the mean relative standard deviations for NeoSpectra Scanner are 4.90 % (large particle size) and 4.74 % (medium particle size) and for NeoSpectra Scanner equipped with Rotator these values are 1.17 % (large particle size) and 0.93 % (medium particle size). The values for the small particle size seem to be significantly smaller: 1.41 % for the NeoSpectra Scanner and 0.52 % for the NeoSpectra equipped with Rotator. It is important to mention that, although the NeoSpectra Scanner equipped with Rotator is the configuration providing the lowest relative standard deviations for all the particle sizes, is also the instrument with a higher number of peculiar large values, as it is possible to see in Fig. 4e.

For all the instruments and particle sizes, there seems not to be a clear relationship between the percentage of bitter almond and the relative standard deviation.

3.2. Principal component analysis

For each sensor studied (with all the configurations), a PCA model was created using all measured reference samples with all the concentrations of contamination of bitter almond, utilizing the average spectrum of each sample. The goal was to identify any differences related to the particle size distribution of the samples, or other spectral properties of the reference materials that would later be used for constructing the PLS models. Fig. 5 shows the first two principal components for each of the models performed.

The model performed on the raw data (using mean centering before

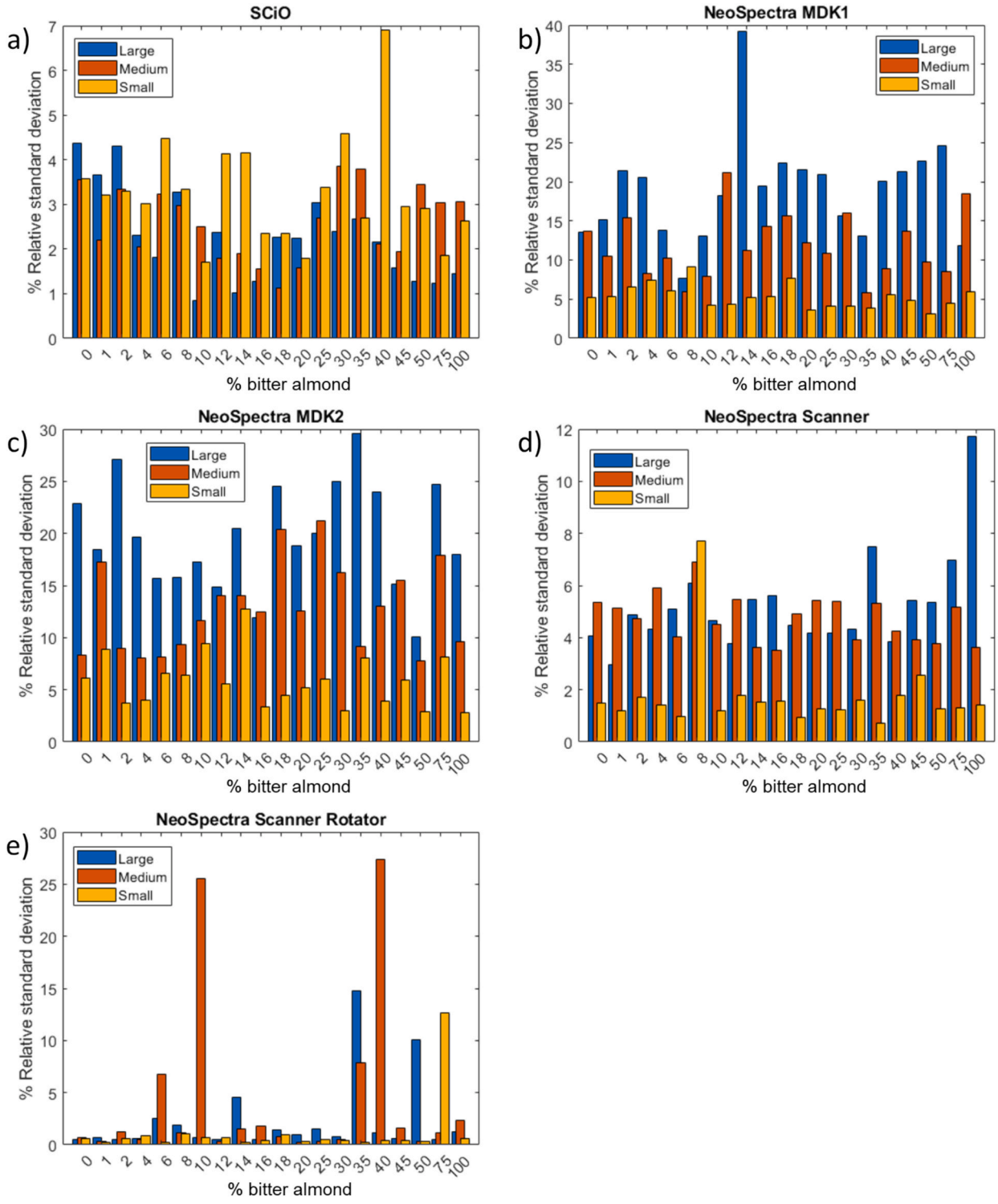

Fig. 4. Average spectra relative standard deviation for the different measurement configurations.

analysis, although other data preprocessing techniques were used obtaining similar results) for the NeoSpectra Scanner (Fig. 5d) shows a clear distinction between the samples according to their particle size, as expected. The model built on the data collected with the accessory Rotator is almost the same (Fig. 5e). The largest difference is observed along PC1 (very high % of information for all the scores plot depicted) between the small particle size and the others, while the difference between medium and large particle sizes is significantly smaller (accounted for PC2). Regarding analytical session, no substantial differences are observed for this sensor between the two analytical sessions in this model. The same conclusions can be drawn from the models built on the MDK1 (Fig. 5b) and MDK2 sensors (Fig. 5c) and the SCiO sensor, except that in the latter case, the two independent analytical sessions appear to be more separated along PC2 (Fig. 5a).

A key finding of this study is that this category of portable sensors can effectively differentiate powder samples based on their particle size. Fig. S4 in the Supplementary Material shows the loadings for the first PC of each of the PLS models in Fig. 5. In all cases, the first PC accounts for a

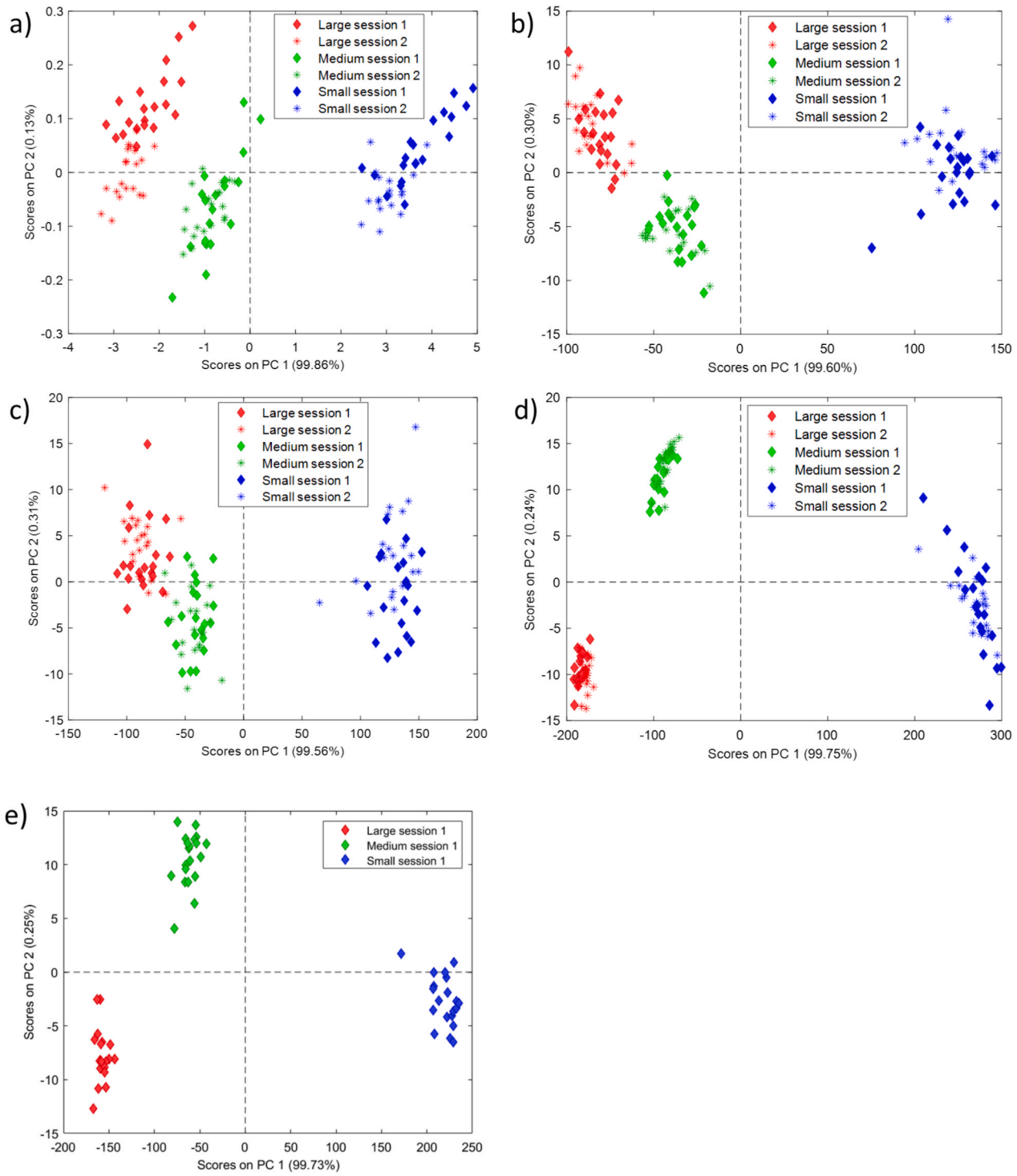

Fig. 5. PCA score plots of the first two PCs on mean centered data – a) SCiO b) NeoSpectra MDK1 c) NeoSpectra MDK2 d) NeoSpectra Scanner e) NeoSpectra Scanner with the Rotator accessory.

very high percentage of the total variance (i.e., over 99 %), and the sign of the different loading values for all variables is the same (whether positive or negative), which is indicative of the presence of scattering. This pattern indicates that the primary variation captured by PC1 is not due to chemical differences between samples, but rather to global changes in spectral intensity or multiplicative scaling effects across the entire wavelength range. Such effects are typically induced by physical sample variations (like particle size) which heavily influence how light scatters. The study of multivariate error associated to spectroscopic measurements of almond powder, has been able to identify the primary effects associated to these measurements, and globally, they include offset noise, shot noise, and multiplicative noise [39].

While this result is specific to the type of samples used in this study, it is likely applicable to all similar samples with comparable characteristics. This may be a problem to bear in mind in a study such as the one presented in this article, but also a strength of this technique. In fact, results obtained strongly indicate that a classification model based on particle size could be easily developed, making it useful for internal

quality control in companies producing powder samples. However, we did not pursue this goal in our study, as our research objectives were different.

3.3. Prediction of bitter almond content in powder blends

PLS (partial least squares) was used to predict the percentage of bitter almond in the almond powder for the different sensors and particle sizes.

Ten reference samples were used to build the model, and ten were used to perform external validation. Table 1 summarizes the obtained results. The regression plots (showing both calibration and validation data) for the best PLS models for the three particle sizes (in all cases, these models correspond to those obtained using the NeoSpectra Scanner equipped with the Rotator accessory) are shown in Fig. S5 in the Supplementary Material. The Limit Of Detection (LOD) in Table 1 was estimated according to the following expression [41]:

$$\mathrm{LOD} = \Delta(\alpha,\beta,\upsilon)\cdot\mathrm{RMSEC}\cdot\sqrt{1+h_0} \quad (1)$$

All models have been analysed by checking the RMSEC/RMSEP against the number of factors, aiming in all cases to avoid model overfitting. Consequently, the majority of the PLS models described in Table 1 of the manuscript (22 out of 27 models) are built with 3 latent variables, and only 5 are built with 2 latent variables. When models from the same instrument type or for a specific particle size do not exhibit good performance characteristics, consistency in the number of latent variables across these models cannot generally be expected. However, for the different models we obtained, we consider the consistency achieved in the number of latent variables to be reasonably good. Table 1 also shows the presence of low bias in the predicted models, with the only exception of the NeoSpectra MDKs and the large particle size.

The RMSEC is obtained from the best model for each instrument and particle size (Table 1), and the leverage h_0 quantifies the distance of the predicted sample at 0 % concentration of bitter almond to the mean of the calibration set. We have calculated h_0 as the leverage value in the calibration model corresponding to the sample having 0 % contamination of bitter almond. The term $\Delta(\alpha,\beta,\upsilon)$ considers the α and β probabilities of wrongly concluding the presence/absence of analyte. When the degrees of freedom υ is high ($\upsilon > 25$), the term $\Delta(\alpha,\beta,\upsilon)$ can be approached to 2.

To make the interpretation of the results clearer, Fig. S6 in the Supplementary Material shows the values of the RMSEP and LOD listed in Table 1. In the framework of this work, the LOD can be defined as the minimum amount of bitter almond in almond powder required to differentiate between the spectra of sweet almond and sweet almond contaminated with bitter almond. It is worth to recall that the LOD is, therefore, not only related to amygdalin but also to all the structural and physiological characteristics that differentiate bitter from sweet

Table 1
Comparison of the best Partial Least Squares (PLS) models obtained for the different instruments and particle sizes. RMSEC, RMSEP, bias and LOD are expressed in % of bitter almond. Apart from the listed preprocessing methods, all the models were finally mean centred. LVs = Latent Variables. MSC = multiplicative scatter correction. 1st der = first derivative. 2nd der = second derivative. Smoothing = second-order polynomial smoothing (in parenthesis, the number of points of the second-order polynomial). Significant figures in Table 1 are shown according to [54].

	SCiOsession 1	SCiO session 2	NeoSpectra MDK1session 1	NeoSpectra MDK1 session 2	NeoSpectra MDK2 session 1	NeoSpectra MDK2 session 2	NeoSpectra Scanner session 1	NeoSpectra Scanner session 2	NeoSpectra Scanner Rotator
Large particle size									
RMSEC	5.7	5.6	6.0	8.6	5.0	3.8	2.9	1.6	1.4
RMSEP	7.6	11	20	18	13	10	7.2	6.7	4.5
r^2 Cal	0.97	0.97	0.98	0.91	0.98	0.99	0.99	0.99	0.99
r^2 Pred	0.77	0.64	0.19	0.22	0.66	0.66	0.88	0.88	0.92
bias Cal	0	0	0	0	0	0	0	0	0
bias Pred	7.6	2.7	−14	−13	2.7	−3.7	1.5	3.2	−0.14
LVs	2	3	2	3	3	3	3	3	3
LOD	13	13	13	19	11	9.3	6.4	3.5	3.2
Preprocessing	1st der	1st der	MSC	MSC	Smoothing (15 points) + MSC + 2nd der	Smoothing (15 points) + MSC + 2nd der	MSC + 2nd der	MSC + 2n der	Smoothing (13 points) + MSC + 1st der
Medium particle size									
RMSEC	7.2	6.1	9.7	8.0	3.5	5.0	1.5	1.9	2.8
RMSEP	8.0	6.7	8.3	10	8.8	8.4	4.1	3.3	2.5
r^2 Cal	0.95	0.97	0.91	0.94	0.99	0.98	0.99	0.99	0.99
r^2 Pred	0.75	0.87	0.75	0.76	0.75	0.80	0.94	0.97	0.97
bias Cal	0	0	0	0	0	0	0	0	0
bias Pred	1.7	−3.2	−3.4	6.2	−1.0	−2.7	1.1	1.7	0.52
LVs	2	3	2	3	3	3	3	3	3
LOD	15	15	22	16	7.8	11	3.3	4.4	6.9
Preprocessing	MSC + 1st der	MSC + 1st der	Smoothing (15 points) + MSC + 1st der	Smoothing (15 points) + MSC + 1st der	Smoothing (11 points) + MSC	Smoothing (11 points) + MSC	Smoothing (15 points) + 2nd der	Smoothing (15 points) + 2nd der	Smooting (15 points) + MSC + 2nd der
Small particle size									
RMSEC	7.0	3.9	4.1	5.4	5.4	2.5	1.9	2.0	0.99
RMSEP	7.9	9.8	4.5	9.8	5.7	4.2	2.7	2.0	2.3
r^2 Cal	0.96	0.99	0.98	0.97	0.97	0.99	0.99	0.99	0.99
r^2 Pred	0.77	0.62	0.90	0.63	0.90	0.93	0.98	0.98	0.99
bias Cal	0	0	0	0	0	0	0	0	0
bias Pred	2.2	3.2	−2.9	1.6	2.5	0.30	−0.90	0.16	1.1
LVs	3	3	3	2	3	3	3	3	3
LOD	18	9.2	9.2	11	13	5.9	4.0	4.9	2.5
Preprocessing	MSC + 1st der	MSC + 1st der	Smoothing (17 points) + MSC + 1st der	Smoothing (17 points) + MSC + 1st der	Smoothing (15 points) + MSC + 2nd der	Smoothing (15 points) + MSC + 2nd der	Smoothing (15 points) + 2nd der	Smoothing (15 points) + 2nd der	Smoothing (23 points) + MSC + 1st der

almonds.

The RMSEP and LOD results follow similar trends, but not identical since the LOD is based on the RMSEC (Eq. (1)), and the ratio RMSEP/RMSEC is not constant for all the instruments and particle sizes.

At first glance, the results from the two independent experiments seem to be fairly consistent (showing the same trend) across all instruments and particle sizes, particularly regarding the prediction error. This error estimates the inaccuracies the model incurs when applied to a dataset that is independent of the one used to develop the model. The number of latent variables used in the different models is also quite consistent.

For the NeoSpectra instruments, the best results in terms of prediction error were obtained for the small particle size, for both NeoSpectra Scanner and NeoSpectra Scanner equipped with Rotator accessory. NeoSpectra Scanner and NeoSpectra Scanner equipped with Rotator also provide the best results for the medium and large particle sizes. Observing the models of the SCiO sensor, a certain consistency in the results can be noted. This sensor appears to be the least affected by the different particle sizes, which can be explained by the fact that it operates in a different spectral region and utilizes a different technology, being the only sensor among the investigated that does not require contact with the sample during the measurements. The NeoSpectra MDK sensors produce poor and highly variable results when analysing large particle size samples. This is probably due to the small size of the detector they are equipped with, making these instruments more suitable for analysing samples that are more homogeneous than those under investigation. It is worth noting that the RMSEP values of the models built with the two NeoSpectra MDK sensors are not identical: in fact, they show quite different values. Specifically, the MDK2 sensor delivers better performance than the MDK1 sensor. It is challenging to determine the reason for this occurrence, because this is most likely due to a technological/instrumental limitation. However, the fact that it is observed in both independent experiments indicates that it is likely not a random result, but rather due to variations in the precision of the data collection. In the case of the NeoSpectra Scanner equipped with the Rotator accessory, the use of an automated device for the measurements does not seem to significantly enhance the performance of the instrument when the particle size is small. However, the performance improves for medium and large particle sizes compared to the sensor used without the accessory. Therefore, the use of the accessory makes the analyses more reproducible, improving performance as the sample heterogeneity increases. As an internal validation indicator, the Cross-Validation (CV) results are provided in Table S1 of the Supplementary Material. For the best models—particularly those from the NeoSpectra Scanner (with and without the Rotator accessory) for the small particle size—the RMSECV and RMSEP values are similar, which indicates good model robustness. In contrast, discrepancies between RMSECV and RMSEP can be observed in certain cases, suggesting that some experimental conditions are more favourable for building more robust calibration models than others. It is worth to mention that the RMSECV values can be higher than the RMSEP values, especially in cases where a random validation segment contains samples from the extremes of the calibration range (e.g., the highest or lowest adulteration levels). The temporary models built during such CV folds must then extrapolate to predict these values, leading to larger errors for that segment and inflating the overall RMSECV. This artifact is not expected in the real-world application of the final model, as future samples should fall within the established calibration range.

Focussing on the limits of detection (Table 1 and Fig. S6b), the best results are obtained for the NeoSpectra Scanner and the NeoSpectra Scanner equipped with the Rotator accessory analysing small particle size, with detection limits around 4.0 % of bitter almond in almond powder, and even 2.5 % in the configuration with Rotator. For comparison purposes, the literature reports the detection of the presence of 5 % of bitter almonds in batches of sweet almonds [36]. The Rotator accessory appeared to offer a further, albeit marginal, sensitivity enhancement compared to the NeoSpectra Scanner alone, particularly notable for large and small particle sizes.

We can see that the main variation in these values occurs between instruments, rather than being strongly dependent on particle size. The best (i.e., lowest) LODs are those obtained with the NeoSpectra Scanner (both with and without the Rotator accessory). Variation is also observed in the LODs from different analytical sessions, even for the same instruments. This suggests that the LOD, as determined here, may not be a consistently robust indicator, with the possible exception of the values obtained for the NeoSpectra Scanner.

As can be seen in Table 1, there are some cases where the LOD values are similar to the RMSEP values. In these cases, this implies that the typical uncertainty of the quantitative predictions (RMSEP) is of the same order of magnitude as the smallest concentration that can be detected (LOD), bearing in mind that RMSEP is an average value and is not necessarily constant across the entire calibration range. This can mean that at concentrations close to the LOD, the precision of quantification is lower; the measurement error is comparable to the measured value. This is, to some extent, expected when attempting to quantify at levels so close to the limit of detection. If the main objective of the method is screening at these concentration levels, for example in product quality control tasks, the LOD is the key parameter. If precise quantification at low levels is required, and the RMSEP is similar to the LOD, then the method might not be as reliable for this precise quantification at low levels.

As previously mentioned, the model built during the first analytical session was validated using both the validation samples from the first session and those from the second session. The same approach was applied to the model calibrated during the second analytical session. The RMSEP values obtained by predicting both validation sets using the models built during the first experimental session are reported in Table 2 (a similar table for the model built in the second experimental session exists, but it is not included as it does not add any further information to the discussion). In Table 2 the prediction errors are indicated as:

RMSEP1: Prediction error calculated using the model calibrated during the first experimental session, applied to the validation set of samples measured during the first experimental session.

RMSEP2: Prediction error calculated using the model calibrated during the first experimental session, applied to the validation set of samples measured during the second experimental session.

Looking at the values in Table 2, there is a general good agreement with the RMSEP values using the validation sets of session 1 and session 2 (all the models used in Table 1 were the models calibrated during the first experimental session) for NeoSpectra Scanner and SCiO sensors. The situation changes for NeoSpectra MDK sensors, and in particular for MDK1 that shows high discrepancy between the two RMSEP values in small and large particle sizes. MDK2 shows differences in the prediction for the medium particle size. This may indicate that in these instruments (MDK1 and MDK2) the analytical session may have a significant influence in the measured experimental data and therefore caution should be taken when using data from different analytical sessions.

This information was rationalized through the calculation of the joint confidence interval for the intercept and slope of the regression line between different prediction results [40]. This approach assumes that if the analytical sessions were comparable, the prediction set from analytical session 1 should yield the same result when predicted by the calibration model of session 1 as when predicted by the calibration model built in session 2. Similarly, the prediction set from analytical session 2 should yield the same result when predicted by the calibration model of session 1 and when predicted by the model of session 2. If the results were comparable, then by plotting on one axis the prediction values from analytical session 1 and on the other axis the prediction values from analytical session 2 for each prediction data set, the points should align on a line with an intercept of 0 and a slope of 1.

We analysed each instrument and particle size across the two analytical sessions. For certain sensors, the level of disagreement

Table 2
RMSEP values using the PLS model build with the calibration values of session 1 and the validation values of session 1 (RMSEP1) and the validation values of session 2 (RMSEP2).

Sensor	Small particle size		Medium particle size		Large particle size	
	RMSEP1	RMSEP2	RMSEP1	RMSEP2	RMSEP1	RMSEP2
NeoSpectra Scanner	2.7	3.4	3.8	3.7	7.2	7.3
NeoSpectra MDK1	4.5	37	8.3	12	20	34
NeoSpectra MDK2	5.7	6.3	8.8	17	13	16
SCiO	7.9	8.1	8.0	8.3	7.6	10

between the predicted data from the two sessions was so pronounced that the r^2 value of the line fell short of the statistical significance threshold for 10 points, which is $r^2 = 0.40$ ($\alpha = 5$ %) [52]. In these instances, constructing the joint confidence interval for the intercept and slope is not worthwhile, as the values of the predictions of the two analytical sessions are excessively divergent. This was the case of the NeoSpectra MDK1 and MDK2 for the medium and large particle sizes, and of the SCiO sensor for the large particle size. For these sensors and particle sizes, this indicates a clear and significant discrepancy between the data from the two analytical sessions.

In the other cases, the results from the two analytical sessions were consistent for the NeoSpectra Scanner across all particle sizes investigated (as shown in Fig. S7a in the Supplementary Material, with an example of the joint confidence interval for the small particle size in which the theoretical point (0,1) falls inside to the joint confidence interval for the intercept and the slope). In contrast, for the MDK1 and MDK2 (small size), and for SCiO (small and medium size), the results did not align across in any of these particle sizes (an example is shown in Fig. S7b in the Supplementary Material, depicting the calculated joint confidence interval for the medium particle size analysis for the SCiO, for which the theoretical point (0,1) is outside the joint confidence interval for the intercept and the slope).

4. Conclusions

A systematic study involving five measurement systems across three categories of portable NIR sensors to predict the contamination of almond powder with bitter almond is presented. The study systematically investigates three distinct particle sizes, emphasizing that particle size can significantly affect reflectance measurements. This influence is an important consideration for developing multivariate models based on external reflection data. The experimental methodology presented here can be applied in any context where particle size is a key attribute of the sample.

The PCA models developed to qualitatively analyse the data confirmed the significance of particle size. They successfully differentiated the samples based on their particle size, providing an important quality control tool for companies dealing with powder samples that have varying physical characteristics such as particle size.

Concerning the prediction of the amount of bitter almond in almond powder, the best prediction models were achieved using the NeoSpectra Scanner spectrometer when analysing small particle size almond powder, with performance further enhanced using the Rotator accessory. This result highlights that minimizing sampling error due to heterogeneity—such as by moving the samples or averaging over a large sample probing area—consistently improves quantitative results. Models were able to predict concentration values with an error around 2.5 % and a LOD around 4.5 % of bitter almond in almond powder. Notably, this affordable and portable sensor provided results consistent with those obtained using more complex and expensive techniques, making them well-suited for field applications. The quantification of the percentage of contamination with bitter almond powder is carried out based on the difference between the spectra of almond powder and almond powder contaminated with bitter almonds. This difference may be attributed to amygdalin, the compound responsible for the bitter taste of almonds, but also to other structural and physiological characteristics of bitter almonds, which are not present in sweet almonds.

The different experimental sessions were not comparable for the majority of the sensors examined. This issue definitely requires further investigation to understand the reasons behind it and to assess whether it can be corrected. Sensors marketed as identical actually demonstrate in the literature different performance, and this is also a point that needs additional investigation [16,53] with a view to facilitating a more extensive use of this type of instruments.

This study, therefore, demonstrates the potential of the proposed technique. It is important to mention that in non-destructive techniques such as NIR spectroscopy, which require minimal or no sample preparation, chemical and physical sources of variability (such as particle size or moisture) may directly affect the spectral response. Therefore, this variability must be properly represented in the calibration set to ensure the model is robust. For its application for instance in the food industry, a broader validation would be required to reflect the physicochemical variability inherent to a complex matrix like a granulated food product, ensuring its performance under real-world conditions.

Miniaturized and, more in general, portable sensors, represent the future of on-site measurements. These sensors meet the requirements of green analytical chemistry by analysing samples in their original form without the use of any reagents, thus minimizing waste, energy and costs. For these reasons, they are particularly well-suited for applications industry. However, the experimental conditions for using them need to be optimized in detail, and the limitations should be understood and identified to overcome them, if possible.

CRediT authorship contribution statement

Barbara Giussani: Writing – original draft, Supervision, Methodology, Conceptualization. **Soumaya Ouzakar El Ghouch:** Investigation, Formal analysis. **Manuel Monti:** Investigation, Formal analysis. **Jordi Riu:** Writing – review & editing, Visualization, Software, Resources, Methodology, Investigation, Formal analysis.

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.

Acknowledgements

JR acknowledge the financial support from the Spanish Ministry of Science, Innovation and Universities (MICIU), the State Research Agency (AEI) and the European Regional Development Fund (ERDF), EU: PID2022-136649OB-I00. The authors acknowledge Optoprim S.R.L. for providing the Rotator accessory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.measurement.2025.118481.

Data availability

Data will be made available on request.

References

[1] B. Giussani, G. Gorla, J. Riu, Analytical chemistry strategies in the use of miniaturised NIR instruments: an overview, Crit. Rev. Anal. Chem. 54 (2024) 11–43, https://doi.org/10.1080/10408347.2022.2047607.

[2] L. Rodriguez-Saona, D.P. Aykas, K.R. Borba, A. Urtubia, Miniaturization of optical sensors and their potential for high-throughput screening of foods, Curr. Opin. Food Sci. 31 (2020) 136–150, https://doi.org/10.1016/j.cofs.2020.04.008.

[3] J. Müller-Maatsch, F.R. Bertani, A. Mencattini, A. Gerardino, E. Martinelli, Y. Weesepoel, S. van Ruth, The spectral treasure house of miniaturized instruments for food safety, quality and authenticity applications: a perspective, Trends Food Sci. Technol. 110 (2021) 841–848, https://doi.org/10.1016/j.tifs.2021.01.091.

[4] K.B. Beć, J. Grabska, C.W. Huck, Principles and applications of miniaturized near-infrared (NIR) spectrometers, Chem. – A Eur. J. 27 (2021) 1514–1532, https://doi.org/10.1002/chem.202002838.

[5] P. Anastas, N. Eghbali, Green chemistry: principles and practice, Chem. Soc. Rev. 39 (2009) 301–312, https://doi.org/10.1039/B918763B.

[6] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – a review, Anal. Chim. Acta 1026 (2018) 8–36, https://doi.org/10.1016/j.aca.2018.04.004.

[7] G. Gorla, P. Taborelli, B. Giussani, A multivariate analysis-driven workflow to tackle uncertainties in miniaturized NIR data, Molecules 28 (2023) 7999, https://doi.org/10.3390/molecules28247999.

[8] M.C. Pasikatan, J.L. Steele, C.K. Spillman, E. Haque, Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials, J. Near Infrared Spectrosc. 9 (2001) 153–164, https://doi.org/10.1255/jnirs.303.

[9] J. Rantanen, E. Räsänen, J. Tenhunen, M. Känsäkoski, J.-P. Mannermaa, J. Yliruusi, In-line moisture measurement during granulation with a four-wavelength near infrared sensor: an evaluation of particle size and binder effects, Eur. J. Pharm. Biopharm. 50 (2000) 271–276, https://doi.org/10.1016/S0939-6411(00)00096-5.

[10] E. Tamburini, F. Vincenzi, S. Costa, P. Mantovi, P. Pedrini, G. Castaldelli, Effects of moisture and particle size on quantitative determination of total organic carbon (TOC) in soils using near-infrared spectroscopy, Sensors 17 (2017) 2366, https://doi.org/10.3390/s17102366.

[11] D.R. Ely, M. Thommes, M.T. Carvajal, Analysis of the effects of particle size and densification on NIR spectra, Colloids Surfaces A Physicochem. Eng. Asp. 331 (2008) 63–67, https://doi.org/10.1016/j.colsurfa.2008.07.017.

[12] A. Szalay, I. Antal, Z. Zsigmond, S. Marton, I. Erós, G. Regdon jr, K. Pintye-Hódi, Technical note : study on the relationship between particle size and near infrared diffuse reflectance spectroscopic data, Part. Part. Syst. Charact. 22 (2005) 219–222, https://doi.org/10.1002/ppsc.200500870.

[13] L.S. Aucott, P.H. Garthwaite, S.T. Buckland, Transformations to reduce the effect of particle size in near-infrared spectra, Analyst 113 (1988) 1849–1854, https://doi.org/10.1039/an9881301849.

[14] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends Anal. Chem. 28 (2009) 1201–1222, https://doi.org/10.1016/j.trac.2009.07.007.

[15] J. Riu, A. Vega, R. Boqué, B. Giussani, Exploring the analytical complexities in insect powder analysis using miniaturized NIR spectroscopy, Foods 11 (2022) 1–16, https://doi.org/10.3390/foods11213524.

[16] J. Ezenarro, D. Schorn-García, M. Plans, O. Busto, R. Boqué, Quantification of spectral measurement errors to guide preprocessing method selection: a case study on cannabinoid prediction across multiple NIR instruments, Anal. Chim. Acta 1343 (2025) 343705, https://doi.org/10.1016/j.aca.2025.343705.

[17] M. Garcia-Aloy, P.J.M. Hulshof, S. Estruel-Amades, M.C.J. Osté, M. Lankinen, J. M. Geleijnse, J. De Goede, M. Ulaszewska, F. Mattivi, S.J.L. Bakker, U. Schwab, C. Andres-Lacueva, Biomarkers of food intake for nuts and vegetable oils: an extensive literature search, Genes Nutr. 14 (2019) 1–21, https://doi.org/10.1186/S12263-019-0628-8.

[18] R. Bodoira, D. Maestri, Phenolic compounds from nuts: extraction, chemical profiles, and bioactivity, J. Agric. Food Chem. 68 (2020) 927–942, https://doi.org/10.1021/acs.jafc.9b07160.

[19] S.K. Chang, C. Alasalvar, B.W. Bolling, F. Shahidi, Nuts and their co-products: the impact of processing (roasting) on phenolics, bioavailability, and health benefits – a comprehensive review, J. Funct. Foods 26 (2016) 88–122, https://doi.org/10.1016/J.JFF.2016.06.029.

[20] O. Eslami, F. Khorramrouz, M. Sohouli, N. Bagheri, F. Shidfar, M.L. Fernandez, Effect of nuts on components of metabolic syndrome in healthy adults with overweight/obesity: a systematic review and meta-analysis, Nutr Metab Cardiovasc Dis 32 (2022) 2459–2469, https://doi.org/10.1016/J.NUMECD.2022.07.015.

[21] D. Martini, J. Godos, S. Marventano, M. Tieri, F. Ghelfi, L. Titta, A. Lafranconi, H. Trigueiro, A. Gambera, E. Alonzo, S. Sciacca, S. Buscemi, S. Ray, F. Galvano, D. Del Rio, G. Grosso, Nut and legume consumption and human health: an umbrella review of observational studies, Int. J. Food Sci. Nutr. 72 (2021) 871–878, https://doi.org/10.1080/09637486.2021.1880554.

[22] E. Bitok, J. Sabaté, Nuts and cardiovascular disease, Prog. Cardiovasc. Dis. 61 (2018) 33–37, https://doi.org/10.1016/J.PCAD.2018.05.003.

[23] L.C. Mead, A.M. Hill, S. Carter, A.M. Coates, The effect of nut consumption on diet quality, cardiometabolic and gastrointestinal health in children: a systematic review of randomized controlled trials, Int. J. Environ. Res. Public Health 18 (2021) 454, https://doi.org/10.3390/IJERPH18020454.

[24] S.Y. Tan, S.L. Tey, R. Brown, Nuts and older adults' health: a narrative review, Int. J. Environ. Res. Public Health 18 (2021) 1848, https://doi.org/10.3390/IJERPH18041848.

[25] C. Alasalvar, J.S. Salvadó, E. Ros, Bioactives and health benefits of nuts and dried fruits, Food Chem. 314 (2020) 126192, https://doi.org/10.1016/J.FOODCHEM.2020.126192.

[26] L. Khalili, T.M.E. A-Elgadir, A.K. Mallick, H.A. El Enshasy, R.Z. Sayyed, Nuts as a part of dietary strategy to improve metabolic biomarkers: a narrative review, Front. Nutr. 9 (2022) 881843, https://doi.org/10.3389/FNUT.2022.881843.

[27] United States Department of Agriculture, 2023/2024 Almonds Production, (n.d.). https://fas.usda.gov/data/production/commodity/0577400 (accessed May 28, 2024).

[28] International Nut & Dried Fruit, Almonds Global Statistical Review, (n.d.). https://inc.nutfruit.org/almonds-global-statistical-review-2/ (accessed May 28, 2024).

[29] S. Yada, K. Lapsley, G. Huang, A review of composition studies of cultivated almonds: macronutrients and micronutrients, J. Food Compos. Anal. 24 (2011) 469–480, https://doi.org/10.1016/J.JFCA.2011.01.007.

[30] H. Barakat, T. Aljutaily, M.S. Almujaydil, R.M. Algheshairy, R.M. Alhomaid, A. S. Almutairi, S.I. Alshimali, A.A.H. Abdellatif, Amygdalin: a review on its characteristics, antioxidant potential, gastrointestinal microbiota intervention, anticancer therapeutic and mechanisms, toxicity, and encapsulation, Biomolecules 12 (2022) 1514, https://doi.org/10.3390/BIOM12101514.

[31] R. Sánchez-Pérez, F.S. Belmonte, J. Borch, F. Dicenta, B.L. Møller, K. Jørgensen, Prunasin hydrolases during fruit development in sweet and bitter almonds, Plant Physiol. 158 (2012) 1916–1932, https://doi.org/10.1104/pp.111.192021.

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

[33] M.A. Faqeerzada, S. Lohumi, R. Joshi, M.S. Kim, I. Baek, B.K. Cho, Non-targeted detection of adulterants in almond powder using spectroscopic techniques combined with chemometrics, Foods 9 (2020), https://doi.org/10.3390/foods9070876.

[34] M. Vega-Castellote, M.T. Sánchez, I. Torres, D. Pérez-Marín, An innovative non-targeted control system based on NIR spectral information for detecting non-compliant batches of sweet almonds, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 250 (2021) 119407, https://doi.org/10.1016/j.saa.2020.119407.

[35] M. Vega-Castellote, D. Pérez-Marín, I. Torres, J.M. Moreno-Rojas, M.T. Sánchez, Exploring the potential of NIRS technology for the in situ prediction of amygdalin content and classification by bitterness of in-shell and shelled intact almonds, J. Food Eng. 294 (2021), https://doi.org/10.1016/j.jfoodeng.2020.110406.

[36] I. Torres, M.-T. Sánchez, M. Vega-Castellote, D. Pérez-Marín, Fraud detection in batches of sweet almonds by portable near-infrared spectral devices, Foods 10 (2021) 1221, https://doi.org/10.3390/foods10061221.

[37] J.M. Netto, F.A. Honorato, P.G. Celso, M.F. Pimentel, Authenticity of almond flour using handheld near infrared instruments and one class classifiers, J. Food Compos. Anal. 115 (2023) 104981, https://doi.org/10.1016/j.jfca.2022.104981.

[38] O. Khater, A. Khater, A.S. Al-Nasr, S. Abozyd, B. Mortada, Y.M. Sabry, Advancing near-infrared spectroscopy: a synergistic approach through Bayesian optimization and model stacking, Spectrochim Acta Part A Mol. Biomol. Spectrosc. 318 (2024) 124492, https://doi.org/10.1016/j.saa.2024.124492.

[39] B. Giussani, M. Monti, J. Riu, From spectroscopic data variability to optimal preprocessing: leveraging multivariate error in almond powder adulteration of different grain size, Anal. Bioanal. Chem. (2024), https://doi.org/10.1007/s00216-024-05710-1.

[40] J. Mandel, F.J. Linnig, Study of accuracy in chemical analysis using linear calibration curves, Anal. Chem. 29 (1957) 743–749, https://doi.org/10.1021/ac60125a002.

[41] N. (Klaas) M. Faber, R. Bro, Standard error of prediction for multiway PLS, Chemom. Intell. Lab. Syst. 61 (2002) 133–149, https://doi.org/10.1016/S0169-7439(01)00204-0.

[42] L. Salguero-Chaparro, V. Baeten, O. Abbas, F. Peña-Rodríguez, On-line analysis of intact olive fruits by vis–NIR spectroscopy: optimisation of the acquisition parameters, J. Food Eng. 112 (2012) 152–157, https://doi.org/10.1016/j.jfoodeng.2012.03.034.

[43] M.L. Martínez, A. Garrido-varo, E. De Pedro, L. Sánchez, Effect of sample heterogeneity on near infrared meat analysis : the use of the RMS statistic, J. Near Infrared Spectrosc. 320 (1998) A313–A320.

[44] A.J. Gaitán-Jurado, M. García-Molina, F. Peña-Rodríguez, V. Ortiz-Somovilla, Near infrared applications in the quality control of seed cotton, J. Near Infrared Spectrosc. 16 (2008) 421–429, https://doi.org/10.1255/jnirs.806.

[45] B. Giussani, A.T. Escalante-Quiceno, R. Boqué, J. Riu, Measurement strategies for the classification of edible oils using low-cost miniaturised portable NIR instruments, Foods 10 (2021) 2856, https://doi.org/10.3390/foods10112856.

[46] S. Šašić, Y. Ozaki, Bands assignment of near-infrared spectra of milk by use of partial least-squares regression, Appl. Spectrosc. 54 (2000) 1327–1338, https://doi.org/10.1366/0003702001951002.

[47] M. Arndt, M. Rurik, A. Drees, K. Bigdowski, O. Kohlbacher, M. Fischer, Comparison of different sample preparation techniques for NIR screening and their influence on the geographical origin determination of almonds (Prunus dulcis MILL.), Food Control 115 (2020) 107302, https://doi.org/10.1016/j.foodcont.2020.107302.

[48] P. Firmani, R. Bucci, F. Marini, A. Biancolillo, Authentication of "Avola almonds" by near infrared (NIR) spectroscopy and chemometrics, J. Food Compos. Anal. 82 (2019) 103235, https://doi.org/10.1016/j.jfca.2019.103235.

[49] V. Cortés, P. Talens, J.M. Barat, M.J. Lerma-García, Potential of NIR spectroscopy to predict amygdalin content established by HPLC in intact almonds and classification based on almond bitterness, Food Control 91 (2018) 68–75, https://doi.org/10.1016/j.foodcont.2018.03.040.
[50] E. Borràs, J.M. Amigo, F. van den Berg, R. Boqué, O. Busto, Fast and robust discrimination of almonds (Prunus amygdalus) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis, Food Chem. 153 (2014) 15–19, https://doi.org/10.1016/j.foodchem.2013.12.032.
[51] G. Gorla, P. Taborelli, H.J. Ahmed, C. Alamprese, S. Grassi, R. Boqué, J. Riu, B. Giussani, Miniaturized NIR spectrometers in a nutshell: shining light over sources of variance, Chemosensors 11 (2023), https://doi.org/10.3390/chemosensors11030182.
[52] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1998. DOI: 10.1002/9781118625590.
[53] S.M. Mitu, C. Smith, J. Sanderman, R.R. Ferguson, K. Shepherd, Y. Ge, Evaluating consistency across multiple NeoSpectra (compact Fourier transform near-infrared) spectrometers for estimating common soil properties, Soil Sci. Soc. Am. J. 88 (2024) 1324–1339, https://doi.org/10.1002/saj2.20678.
[54] A.C. Olivieri, Practical guidelines for reporting results in single- and multi-component analytical calibration: a tutorial, Anal. Chim. Acta 868 (2015) 10–22, https://doi.org/10.1016/j.aca.2015.01.017.