Use of hyperspectral imaging devices for the measurement of small granular samples: Evaluation of grape seed protein concentrates

Julio Nogales-Bueno [a], Francisco José Rodríguez-Pulido [b], Francisco José Heredia [b], José Miguel Hernández-Hierro [b,*], Berta Baca-Bocanegra [a]

[a] *Department of Analytical Chemistry, Facultad de Farmacia, Universidad de Sevilla, 41012 Sevilla, Spain*
[b] *Food Colour and Quality Laboratory, Área de Nutrición y Bromatología, Facultad de Farmacia, Universidad de Sevilla, 41012 Sevilla, Spain*

ARTICLE INFO

ABSTRACT

Hyperspectral imaging is a well-known technique for quality control of food and agricultural products. This technique is often applied to the measurement of large and heterogeneous samples, where chemical imaging is extremely useful. In addition, when the amount of sample is limited by different factors (price, other analyses, sample production, etc.) hyperspectral imaging is an alternative to traditional spectroscopes for acquiring its spectral information.

In this study, a new and specific methodology to acquire hyperspectral information from small amounts of granular samples has been developed. For this purpose, two different hyperspectral devices (400–1100 nm, 900–1700 nm) have been used. A statistical procedure has been followed to test the proposed method. In grape seed protein concentrates, NIR radiation (900–1700 nm) penetrates deeper into the sample than VisNIR radiation (400–1100 nm). Therefore, the minimum amount of sample needed to measure in the NIR range is larger than that needed to measure in the VisNIR range. Finally, different calibration models have been developed for the control of protein content in the tested samples. Standard errors of prediction obtained in external validation are similar to those reported in the literature when sample amount is not an issue (9–10 %).

1. Introduction

Hyperspectral imaging adds vision techniques to classic spectroscopy and integrates spectroscopic and spatial information which otherwise cannot be achieved with either conventional imaging or spectroscopic techniques. Thus, hyperspectral imaging provides spatial and spectral information of samples and it allows to obtain multi-constituent information (Sun, 2010). Nowadays, hyperspectral imaging is a well-known technique for quality control of food and agricultural products. A large number of applications have already been developed, not only on a laboratory scale, but also at field or in the industrial sector (El-Mesery, Mao, & Abomohra, 2019; Su & Sun, 2018; Zhu et al., 2020). Hyperspectral imaging combines spectral and spatial information and allows combining advantages of traditional spectroscopy and image analysis (Sun, 2010). Therefore, this technique is particularly suitable for the measurement of large and heterogeneous samples, where the spatial dimension is of clear importance and hyperspectral images can be processed to develop chemical imaging. For example, good results have

been obtained with apples, kiwis or walnuts, among others (Guo, Zhao, & Dong, 2016; Nogales-Bueno et al., 2021; Tian, Li, Wang, Fan, & Huang, 2018; Wang & He, 2019).

Quality control of samples of intermediate size (e.g., 0.5–2 cm), such as berries, cereals or seeds, has also been successfully measured using hyperspectral imaging (Chakraborty et al., 2021; Chen et al., 2015; Fernandes et al., 2011). In these studies, spatial dimension has not been always taken into account due to the reduced size of the samples and hyperspectral devices have been usually applied as traditional spectroscopes. Even when the samples are topochemically homogeneous and the spatial dimension is not as useful as for other samples, hyperspectral imaging can yield really good results. In fact, hyperspectral imaging is also really useful for the identification and elimination of background, shadows or foreign bodies. Furthermore, these intermediate sample sizes are often not measurable on some traditional spectroscopes. Traditional spectroscopes usually have problems when the sample is similar in size to the measurement window (Baca-Bocanegra, Hernández-Hierro, Nogales-Bueno, & Heredia, 2019; Nogales-Bueno et al.,

2020).

Finally, hyperspectral imaging can be used in applications where the amount of sample is limited by different factors (price, other analyses, sample production, etc.) or for the measurement of small samples (less than 0.5 cm), such as small seeds or granular products and flours (Caporaso, Whitworth, & Fisk, 2018; Rodríguez-Pulido et al., 2014; Zhao et al., 2018). However, in most cases, it may be preferable to measure these samples with traditional spectroscopes, which are cheaper and have similar or better spectral performance than hyperspectral devices. These products can usually be packaged and easily presented in the measurement window of traditional spectroscopes. For this, a prerequisite is to have sufficient sample to cover the entire measurement window with a complete and thick layer of sample. In contrast, when the amount of sample available is small, measuring it on some dispersive spectroscopes may not be possible (Durig, Zunic, Costner, & Guirgis, 1993).

In these applications, hyperspectral imaging can be suitable alternative to traditional spectroscopes for the acquisition of its spectral information. The spatial dimension will allow the acquisition of their spectra, discarding regions without sample or even regions with a reduced amount of sample. Avoiding sample destruction and without any sample pretreatment, hyperspectral imaging is probably the spectral technique that requires the least amount of surface area to acquire a correct spectrum of a sample. Theoretically, hyperspectral imaging can acquire spectral information of a sample covering only a pixel. In practice, a larger number of pixels is necessary to reduce noise and obtain a representative spectrum. However, considering the spatial resolution of typical hyperspectral devices, this requirement is easily reachable. As for the sample quantity, another consideration must be taken into account: the penetration of electromagnetic radiation into the sample. This penetration depends on the nature of the sample and the spectral region measured (Siesler, Ozaky, Kawata, & Heise, 2002). In addition, for granular samples, scattering is the dominant contribution to the extinction of light in the NIR region. Scattering reduces the penetration depth of diffuse reflection-based measurements, as photons are ejected back out of the sample before they have a chance to be absorbed. Consequently, the light intensity drops exponentially within the sample (Scheibelhofer, Wahl, Larchevêque, Chauchard, & Khinast, 2018). In any case, the deeper the penetration, the larger the amount of sample necessary. Therefore, these aspects must be considered, and a specific and meticulous analysis methodology must be developed to acquire the spectra of these samples.

In these cases, the acquisition of the sample spectrum is not always trivial: the sample must properly cover a sufficient area and the thickness of the sample must be greater than the penetration of the electromagnetic radiation. Whereas other dispersive spectroscopes need to completely cover a measurement window, hyperspectral imaging only needs a small amount of sample to cover a few pixels. Regarding the thickness of the sample, the minimum thickness depends on the penetration of the applied electromagnetic radiation. Moreover, this penetration depends on the energy of the radiation, the chemical and physical the nature of the sample (Siesler et al., 2002). Typical depths cover a wide range, from fractions of a millimeter to 10 mm (Qin & Lu, 2008). However, in powdered organic samples, radiation penetration has been found to typically range from 1 to 5 mm for both the visible and NIR regions (Huang et al., 2016). This is most likely due to the scattering processes that granular samples undergo. This scattering reduces the penetration depth of measurements based on diffuse reflection, as the photons are ejected from the sample again, before they have a chance to be absorbed (Scheibelhofer et al., 2018).

An important and current example of this type of samples is protein extracts from food and agricultural products. Today, the food industry is looking for plant-based proteins that can supply and complement society's broad protein demand (Pojić, Mišan, & Tiwari, 2018). In addition to the best-known protein sources (soybean, pea, etc.), other lesser-known vegetables are being studied as protein sources. In the early stages of the research process, the available amount of protein concentrates can be really small, which creates the need for methods for the control of small samples. For example, the potential of grape seed proteins is being studied (Baca-Bocanegra, Nogales-Bueno, Hernández-Hierro, & Heredia, 2021; Gazzola, Vincenzi, Marangon, Pasini, & Curioni, 2017; Zhou, Zhang, Liu, & Zhao, 2011). Grape seeds are an important by-product of the wine industry, with an annual production of thousands of tons (Ruggieri et al., 2009). In addition, grape seed proteins have a large number of technological and biological functionalities, which increases the interest in their exploitation. For example, they can be revalorized by the own wine industry for modulating quality features such as the appearance, color and stability of red wines (Gazzola et al., 2017; Zhou et al., 2011).

Therefore, the main aim of this study is to develop a suitable methodology to acquire spectral information from small amounts of granular samples using hyperspectral imaging. For this purpose, two different hyperspectral devices will be used, different spectral regions and spatial and spectral resolutions will be considered, and a statistical procedure will be rigorously followed to ensure the feasibility of the proposed method. In addition, using this novel methodology, different calibration models will be developed for monitoring the protein content in grape seed protein concentrates.

2. Materials and methods

2.1. Hyperspectral imaging

Two different hyperspectral imaging devices have been used in this study. The first device operates mainly in the visible region, although its spectral range also includes a small part of the near infrared region (VisNIR). This device is described in detail at Rodríguez-Pulido, Mora-Garrido, González-Miret, and Heredia (2022). The second one is a near infrared system (NIR-HSI) described in detail at Nogales-Bueno, Hernández-Hierro, Rodríguez-Pulido, and Heredia (2014). Briefly.

- VisNIR-HSI device is a Specim IQ hyperspectral camera (Spectral Imaging Ltd., Oulu, Finland). It is a compact camera including all necessary parts (sensor, spectrograph, and optics). Its sensor (CMOS technology) and optical system allow the acquisition of 512×512 pixel images and record 204 spectral bands, from 400 to 1000 nm, with a spectral resolution of 3 nm. The image is built up in a scanning process thanks to an internal scanner with a motor for moving the optics. The camera includes a certified plate for image calibration. When a measurement session is started, an image must be acquired from this plate and the system automatically corrects the next acquired hyperspectral images. As a result, the camera stores on a replaceable memory card not only the raw data, but also the corrected image in reflectance units.
- NIR-HSI device was provided by Infaimon S.L. (Barcelona, Spain) and comprises the optics (mirror scanner and lens), spectrograph, camera and computer. The camera (Xenics® XEVA-USB) uses a matrix InGaAs sensor with 320×256 pixels (Xenics Infrared Solutions, Inc., Leuven, Belgium). The spectrograph (Specim ImSpector N17E Enhanced; Spectral Imaging Ltd., Oulu, Finland) has a spectral resolution of 3.25 nm and covers the spectral range between 900 and 1700 nm (949–1648 nm after removal of noisy bands). This device is a push-broom system mounting a mirror scanner (Spectral Imaging Ltd., Oulu, Finland) that allows scanning the whole region under the camera. In this case, image correction was performed manually: two reference images were acquired, one acquisition of a reference white ceramic tile (Labsphere Inc., North Sutton, NH, USA) and one covering the camera lens. The raw images were then corrected using these two reference images and the reflectance data were obtained.

Both cameras are part of a single setup and use the same illumination: two 70 W tungsten-iodine halogen lamps (Prilux®, Barcelona,

Spain). In fact, the camera lenses are placed side by side. The samples were placed on a polyethylene plastic tray, where both devices could simultaneously record the hyperspectral image.

2.2. Background and description of the samples

Eighty one grape seed protein concentrates (GSPC) were obtained as part of a process to maximize the protein content of the concentrate under practical operating conditions of pH, temperature, flour/water ratio and extraction time (Baca-Bocanegra et al., 2021). To do that, grape seed meals were subjected to alkaline extraction and isoelectric precipitation, applying different operating conditions of pH, temperature, flour/water ratio and extraction time ranged respectively from 8.5, 25 °C, 1:6 and 1 h to 10.5, 45 °C, 1:12 and 3 h. After the isoelectric precipitation, proteins were recovered by centrifugation and freeze-drying and GSPC were stored in airtight containers until hyperspectral image acquisition. Sample weights ranged from 0.02 to 1.28 g, with an average value of 0.32 g and a standard deviation of 0.30 g.

For each GSPC, nitrogen content was determined using a manual Kjeldahl distiller (J.P. Selecta, Barcelona, Spain). Then, protein content was calculated using the nitrogen-to-protein conversion factor of 5.75 (Cejudo-Bastante, Oliva-Sobrado, González-Miret, & Heredia, 2022). Protein content was reported as the ratio of protein weight to GSPC weight, expressed as a percentage. In the present study, protein content data will be used as the reference parameter for the calibration of the spectral methods.

2.3. Sample capsule

A sample holder capsule was designed and fabricated to meet the requirements for acquiring spectral information from small granular samples. These requirements are to contain granular samples, maximize sample depth, have a characteristic spectrum, easily discriminable from the sample, and be easily cleaned. Therefore, to fulfill all these requirements, the sample capsule was fabricated from a polypropylene centrifuge tube (Nerbe plus GmbH & Co., Winsen, Germany). This tube had a capacity of 50 ml and a skirted conical bottom. Using a knife and a rasp, the bottom of this tube was cut, and it was used as a sample holder capsule (Fig. 1a). The internal dimensions of the fabricated capsule are shown in Fig. 1b. This geometry allows maximizing the sample depth even when the amount of sample available is reduced.

In order to determine the penetration depth of the electromagnetic radiation in GSPC, samples were measured in this sample holder capsule. The geometry of this capsule and the spatial resolution of the cameras make it possible to determine the minimum depth under each pixel in the hyperspectral images (Fig. 2). Hyperspectral images of three samples were acquired with both HSI devices. The surface of these samples was flattened. These images were used to determine the penetration depth of electromagnetic radiation. They were selected because they had a sufficient amount of sample to cover a considerable region of the capsule. Thus, it can be assumed that, in the central region of the capsule, the sample depth is much greater than the radiation penetration depth. Consequently, there is certainty that the spectra of the central pixels belong entirely to GSPC or, in other words, that the capsule material does not affect the spectra of the central pixels.

2.4. Statistical tools

Different statistical tools were applied to discriminate between GSPC and non-GSPC pixels of hyperspectral images, to assess the reliability of the spectral information contained in the different GSPC pixels and to develop calibration methods to control the protein contents in GSPC.

- The discrimination between GSPC and non-GSPC pixels of hyperspectral images, or segmentation procedure, was carried out by a linear discriminant analysis (LDA). Using Matlab R2018a

Fig. 1. Description of the sample holder capsule. (a) Initial polypropylene centrifuge tube and sample holder capsule made from it. (b) Dimensions of the sample holder capsule.

(TheMathWorks, Inc., MA, USA), 70 GSPC spectra, 70 capsule spectra and 70 background spectra were manually extracted from the hypercubes. These spectral data were used as dependent variables for the development of the LDA and the sample or non-sample membership of the spectra was used as a factor. LDA was carried out using SPSS 27.0 (SPSS, Inc., Chicago, IL, USA). The obtained discriminant models were input on a Matlab R2018a script that evaluates each pixel of the hyperspectral images and identifies them as sample or non-sample. As a result, a segmentation mask was created for each image.
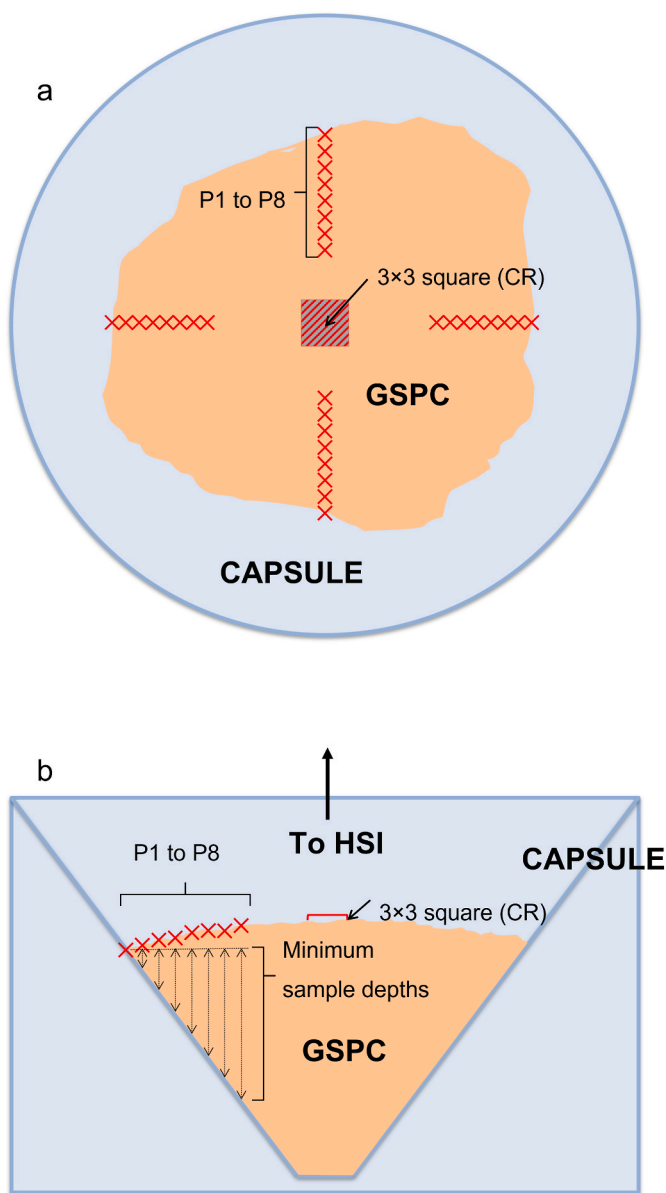
Fig. 2. Schematic representation of the arrangement of the GSPC in the sample capsule. (a) Top view. (b) Cross section.

- Pearson's similarity index was applied to determine the minimum sample depth that produces reliable GSPC spectra, i.e., to assess the reliability of the spectra. That index was used to measure the similarity between different spectral signals in the same hyperspectral image. An adaptation of the methodology described elsewhere in (Baca-Bocanegra et al., 2019) has been applied. First, a region belonging to the GSPC was identified and a representative spectrum of this region was obtained. Second, different regions of uncertain origin or 'problem' regions were selected and their spectral signals were obtained. Third, a Pearson linear regression was performed between each 'problem' spectrum and the representative spectrum of the sample captured in the hyperspectral image. Lastly, Pearson's similarity indexes (SI) were calculated as $SI = 1/ (1 - R^2)$. The identification of the different spectral regions and the extraction of the corresponding spectra were performed using Matlab R2018a. Pearson's similarity indexes were calculated using Win ISI® (v1.50) (Infrasoft International, LLC, Port. Matilda, PA, USA).

- Protein content calibrations at GSPC spectra were performed using the reliable spectral data obtained from VisNIR-HSI and NIR-HSI

devices and using fusion of these data for each sample. For each data set, the following procedure were carried out: GSPC samples were randomly divided into calibration and validation sets (75% and 25 % of the samples, respectively). A principal component analysis (PCA) was applied to spectra allocated into the calibration set. This PCA allows to identify the spectral outliers, to distribute samples in the created space and to detect their possible separations in different spectral groups. Next, different spectral pretreatments were applied to the spectra allocated into calibration set: multiplicative scatter correction (MSC), standard normal variate (SNV) and detrending were applied to reduce scattering effects that granulated samples produce (Dhanoa, Lister, & Barnes, 1995; Geladi, MacDougall, & Martens, 1985). Moreover, the effect of differentiation and variations in spectral ranges were tested in the development of the calibrations. Then, modified partial least square regression (MPLS) was applied (Shenk & Westerhaus, 1995). This chemometric method divides the calibration set into a number of subsets to perform a cross-validation to establish the number of PLS factors. In that way, the possibility of overfitting is reduced and chemical outliers are identified (T ≥ 2.5 criterion) and removed. Finally, the most promising calibration (according to its standard error in cross-validation, SECV) was selected and the validation set was used to assess the reliability of the selected method by obtaining the standard error of prediction (SEP) in external validation. This entire process was carried out with Win ISI® (v1.50).

3. Results and discussion

3.1. Acquisition of spectral information from small quantities of granular samples

For each one of the three images selected for the determination of the penetration depth of the electromagnetic radiation in GSPC sample, a 3 × 3-pixel square was selected in the central region of the sample, using Matlab R2018a software, and the average reflectance spectrum of this region (CR) was extracted. In addition, the sample-capsule interface was identified and 8 consecutive pixels were selected from this interface towards the center of the sample (P1 to P8). In this way, the thickness of the sample represented in each pixel increases from pixel P1 to pixel P8. This procedure was performed four times, on the right, left, top and bottom of the sample (Fig. 2). The reflectance spectra of all these pixels were extracted and the average and the standard deviation spectra of the 4 pixels P1, the 4 pixels P2, the 4 pixels P3 and so on were obtained. Then, for each standard deviation spectrum, the average standard deviation was obtained as the average of each spectrum.

Fig. 3a shows a comparative of the average standard deviation for each group of pixels from P1 to P8 for the hyperspectral images acquired with the NIR-HSI device. It can be seen that, for all three samples considered, the average standard deviations of the outer pixels are larger than those of the inner pixels. However, for samples T04A and T26C, initial standard deviation (P1) is quite reduced. This can be explained by an important influence of the capsule spectrum on these pixels, as the composition of the capsule and consequently its spectrum is homogeneous. The standard deviations of pixels P2 and P3 are high for all samples, indicating that these pixels probably represent a heterogeneous mixture of sample and capsule spectra. Then, the standard deviation decreases for pixels P4 and P5, which may indicate a drop in the influence of the capsule spectra on these pixels. Finally, the behavior of the average standard deviation of the inner pixels is rather chaotic. This may be due to the intrinsic spectral variability of the samples.

To clarify these hypotheses, Pearson linear regressions were performed between the average spectra of each pixel (P1 to P8) and the CR spectrum. Coefficients of determination were obtained and the corresponding similarity indices (SI) were calculated. Fig. 3b and c shows the evolution of these variables in the considered pixels. It can be appreciated that average spectra of the selected pixels show a better correlation
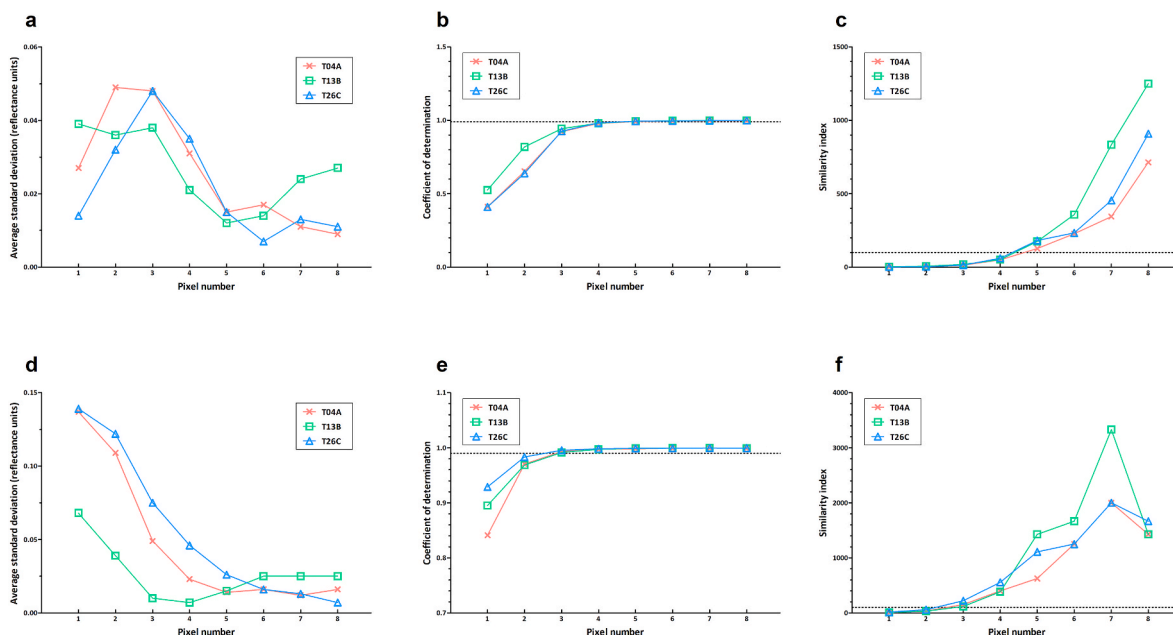
Fig. 3. Comparison of the main statistics calculated for the determination of the penetration of electromagnetic radiation in the samples. (a), (b) and (c) Respectively, comparisons of the average standard deviation, the coefficient of determination and the similarity index for the NIR-HSI camera. (d), (e) and (f) Respectively, comparisons of the average standard deviation, the coefficient of determination and the similarity index for the VisNIR-HSI camera. The threshold values of the coefficient of determination (0.99) and the similarity index (100) are marked with a dotted line in figures (b), (e), (c) and (f).

with the central region spectrum for inner pixels than for outer pixels. Indeed, pixels P5 to P8 present a SI higher than 100 units (or a coefficient of determination higher than 0.99). When these indexes exceed the above-mentioned thresholds, the evaluated spectra can be considered similar spectra (i.e., the acquired spectra can be considered entirely as sample and not as capsule). Therefore, it has been obtained that for pixel P5 the thickness of the sample is greater than the penetration of the electromagnetic radiation. Taking into account the geometry of the sample holder capsule and the spatial resolution for the NIR camera at the fixed measurement distance (0.61 mm pixel^{-1}), it was stablished that the penetration of the NIR radiation (900–1700 nm) in the GSPC is of 4.35 mm.

A similar procedure was carried out for the VisNIR-HSI device. Hyperspectral images of the same GSPC were used for stablishing the penetration of the VisNIR radiation (400–1100 nm) in the GSPC. The evolution of the average standard deviations is shown at Fig. 3d. Again, for all three samples considered, the average standard deviations of the outer pixels are larger than those of the inner pixels. Fig. 3e and f shows the evolution of the coefficients of determination and the similarity indexes in the considered pixels. The trends of these parameters are similar than those explained for the NIR-HSI device: the inner pixels show a better correlation with the central region spectrum than the outer pixels. However, in that case, pixel P3 is the first that present a SI higher than 100. Therefore, the penetration of the electromagnetic radiation for the VisNIR-HSI device is lower than for the NIR-HSI device. Concretely, it was stablished that the penetration of the VisNIR radiation (400–1100 nm) in the GSPC is of 1.45 mm (the spatial resolution for VisNIR camera is 0.41 mm pixel^{-1}).

In summary, the experimental procedure described using a conical sample holder capsule has made it possible to determine the penetration of electromagnetic radiation for the spectral range measured by two hyperspectral devices. It has been found that, for GSPC, the penetration of electromagnetic radiation is deeper for the spectral range measured by the NIR-HSI device than for that measured by the VisNIR-HSI device. In other words, NIR radiation penetrates deeper than visible radiation into these samples. The literature describes that high spectral absorption intensities lead to lower penetration of radiation and that this is even

more evident in the case of diffuse reflection measurements of powders or granular samples, where the actual pathlengths of radiation are much greater than the penetration depth (Siesler et al., 2002). It corresponds with the findings described elsewhere in Qin and Lu (2008) and in Laborde et al. (2020) where the lowest radiation penetration correspond to spectral regions of high spectral absorption intensities. Fig. 4 shows the GSPC average absorbance spectra ($log[1/R]$) of the GSPC for the both HSI devices. It can be appreciated that the absorbance for the VisNIR spectrum is on average higher than for the NIR spectrum, which corresponds to the radiation penetrations obtained for these spectral ranges.

Taking into account these findings, VisNIR and NIR hyperspectral images can be processed in order to identify GSPC in the image and then, to erode 2 and 4 external pixels, respectively. In that way, the segmentation masks obtained will only content pixels of GSPC without any influence of the capsule polymer. Furthermore, the information gathered can be utilized to estimate the minimum amount of sample
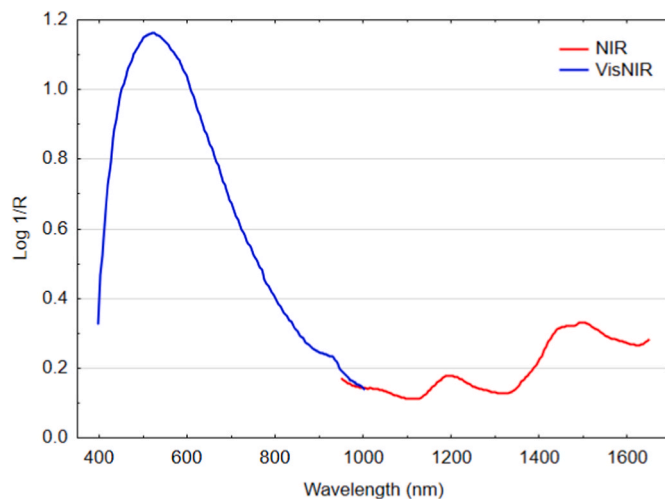

Fig. 4. Average absorbance spectra (log [1/R]) of the GSPC for the VisNIR-HSI and NIR-HSI devices.

necessary for a correct measure in both devices. Considering their spatial resolutions, the obtained penetrations of the radiation for both spectral ranges and estimating a minimum number of 12 GSPC pixels in the images for obtaining a correct average spectrum, without appreciable spectral noise, it can be calculated the minimum amount of sample for each device. This amount is the contained in a cylindrical region with a section of 4.68 mm^2 and with a volume of 20.34 mm^3 and a section of 2.11 mm^2 and with a volume of 3.06 mm^3 for the NIR and Vis-NIR devices, respectively. Therefore, as a consequence of the better spatial resolution and lower radiation penetration for the VisNIR device than for the NIR device, the amount of sample required to acquire an adequate spectrum with the VisNIR device is considerably less than that required for the NIR device.

3.2. Modified partial least square regressions

After the erosion of the external pixels (i.e. 2 and 4 pixels for VisNIR-HSI and NIR-HSI, respectively), the segmentation masks of 7 of the 81 hyperspectral images acquired with each device were completely eroded due to the small amount of sample available and these images were discarded. The remaining 74 images were processed and a representative average spectrum was extracted for each image. Then 3 spectral matrixes were constructed, one NIR matrix, one VisNIR matrix and one Fusion matrix as the fusion of the two previous datasets. The Fusion matrix was constructed following a level 1 procedure (Castanedo, 2013) in which absorbance signals were fused with an alignment consisting of a simple interpolation in the shared spectral range (between 949 and 1000 nm).

Each spectral matrix was divided into calibration and validation sets and calibration sets were subjected to PCA and MPLS analyses. After PCA, one NIR spectrum was discarded for being a spectral outlier with a Mahalanobis distance above the 3-unit threshold. This was the only spectral outlier in the whole process. Analysis of this outlier showed that the erosion process had produced a segmentation mask with only 3 pixels. Thus, it appears that such a small number of pixels produces an average spectrum with a signal-to-noise ratio that is too low, resulting in an invalid spectrum. Remaining segmentation masks were checked and their number of pixels was higher than 10 in all cases.

SNV and MSC spectral pretreatment produced the best results of SECV in the regression processes for NIR and VisNIR datasets, respectively (Table 1). The best MPLS regression for the Fusion dataset was also obtained after a MSC pretreatment. However, in this case, it was also convenient to apply a second derivative to the dataset. Afterwards, the validations sets were used to obtain the SEPs in external validation for the three regression models obtained. All these statistical parameters are summarized in Table 1. It can be appreciated that both the NIR and Fusion datasets produce the best results, with errors in the external validation process of less than 10%. However, the low coefficient of determination of the VisNIR regression model is holding the result of the Fusion model back.

The results obtained for the NIR matrix are in line with those obtained with the same or different HSI devices, for applications where the amount of sample or the penetration of the electromagnetic radiation is not a problem. For example, in our laboratory, the VisNIR-HSI device has been used to calibrate sugar, total phenols and total flavanols contents in grape, producing errors of 9.49%, 14.69% and 15.50% in external validation, respectively (Rodríguez-Pulido et al., 2022). The NIR-HSI device has also been used with a similar aim, and similar results in external validation have been obtained (10.25% and 24.11% for sugar and total phenols contents in grape, respectively) (Nogales-Bueno et al., 2014). Several attempts have been made in our laboratory to calibrate the protein content in whole grapes, but the prediction errors have been too high. However, other authors have obtained adequate calibrations of proteins in different matrices using different hyperspectral or conventional spectrometers. For example, Berzaghi, Cherney, and Casler (2021) got errors in external validation comprised between 3.96 and 14.65% and between 3.6 and 10.3% for crude protein at grass and alfalfa, respectively, using different portable spectrometers. Furthermore, similar results have been found in the prediction of the protein content of kernel or bulk samples of wheat using hyperspectral imaging (Caporaso et al., 2018; Mahesh, Jayas, Paliwal, & White, 2015). Consequently, the proposed methodology for the acquisition of spectral information from small amounts of granular samples by hyperspectral image analysis produces adequate results, and comparable to those obtained when the amount of sample available is large enough not to be considered a problem.

4. Conclusions

In the present study it has been developed and tested a methodology for the acquisition of spectral information from small amounts of granular samples by hyperspectral image analysis. The proposed methodology has allowed to determine the penetration of the electromagnetic radiation into the samples. For the same samples, different penetrations have been obtained for the different spectral ranges considered. These results are in agreement with those described in the literature, which indicate that the penetration of radiation into a sample depends not only on the energy of the radiation, but also on the absorption intensity of the sample. In the case of GSPC samples, NIR radiation (900–1700 nm) penetrates deeper into the sample than VisNIR radiation (400–1100 nm), which implies that the minimum amount of sample needed to measure in the NIR range is larger than that needed to measure in the VisNIR range.

In any case, the methodology presented here allows to correctly acquire spectral information from small amounts of sample using HSI. Moreover, this spectral information can be easily extracted and used for the calibration of spectral methods for the prediction of parameters of interest in the sample, such as protein content in GSPC. This methodology can be easily extrapolated to other areas where the amount of sample available for spectral analysis is small. For example, the

Table 1

Main statistical descriptors for the MPLS models developed for protein content prediction at GSPC using three different hyperspectral datasets: NIR, VisNIR and the level 1 fusion of these two sets.

Set	Pretreatment[a]	N[b]	Mean	SD[c]	SEC[d]	RSQ[e]	SECV[f]	SEP[g]	SEP %
NIR	SNV +0,0,1	51	52.57	5.42	2.46	0.79	4.42	4.75	9.04
VisNIR	MSC +0,0,1	55	51.58	6.35	4.81	0.43	4.97	5.18	10.04
Fusion	MSC +2,10,10	55	51.58	6.35	4.39	0.52	4.74	4.84	9.38

[a] Pretreatment: SNV - Standard Normal Variate. MSC - Multiplicative Scattering Correction. First digit - number of the derivative. Second digit - gap over which the derivative is calculated. Third digit - number of data points in a smoothing.

[b] N: number of samples (calibration set).

[c] SD: standard deviation.

[d] SEC: standard error of calibration.

[e] RSQ: coefficient of determination (calibration set).

[f] SECV: standard error of cross-validation.

[g] SEP: standard error of prediction in the external validation (also expressed in percentages with respect to the mean value).

proposed methodology may be of great interest for forensic drug analysis or for the analysis of health samples, among others.

Fundings

CRediT authorship contribution statement

Julio Nogales-Bueno: Data curation, Methodology, Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis. **Francisco José Rodríguez-Pulido:** Data curation, Investigation, Methodology, Writing – review & editing. **Francisco José Heredia:** Funding acquisition, Supervision, Writing – review & editing. **José Miguel Hernández-Hierro:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Berta Baca-Bocanegra:** Conceptualization, Data curation, Methodology, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

References

Baca-Bocanegra, B., Hernández-Hierro, J. M., Nogales-Bueno, J., & Heredia, F. J. (2019). Feasibility study on the use of a portable micro near infrared spectroscopy device for the "*in vineyard*" screening of extractable polyphenols in red grape skins. *Talanta, 192*, 353–359. https://doi.org/10.1016/j.talanta.2018.09.057

Baca-Bocanegra, B., Nogales-Bueno, J., Hernández-Hierro, J. M., & Heredia, F. J. (2021). Optimization of protein extraction of oenological interest from grape seed meal using design of experiments and response surface methodology. *Foods, 10*(1), 79.

Berzaghi, P., Cherney, J. H., & Casler, M. D. (2021). Prediction performance of portable near infrared reflectance instruments using preprocessed dried, ground forage samples. *Computers and Electronics in Agriculture, 182*, Article 106013. https://doi.org/10.1016/j.compag.2021.106013

Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2018). Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chemistry, 240*, 32–42. https://doi.org/10.1016/j.foodchem.2017.07.048

Castanedo, F. (2013). A review of data fusion techniques. *The Scientific World Journal, 2013*, Article 704504. https://doi.org/10.1155/2013/704504

Cejudo-Bastante, M. J., Oliva-Sobrado, M., González-Miret, M. L., & Heredia, F. J. (2022). Optimisation of the methodology for obtaining enzymatic protein hydrolysates from an industrial grape seed meal residue. *Food Chemistry, 370*, Article 131078. https://doi.org/10.1016/j.foodchem.2021.131078

Chakraborty, S. K., Mahanti, N. K., Mansuri, S. M., Tripathi, M. K., Kotwaliwale, N., & Jayas, D. S. (2021). Non-destructive classification and prediction of aflatoxin-B1 concentration in maize kernels using Vis-NIR (400-1000 nm) hyperspectral imaging. *Journal of Food Science and Technology-Mysore, 58*(2), 437–450. https://doi.org/10.1007/s13197-020-04552-w

Chen, S., Zhang, F., Ning, J., Liu, X., Zhang, Z., & Yang, S. (2015). Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chemistry, 172*(0), 788–793.

Dhanoa, M. S., Lister, S. J., & Barnes, R. J. (1995). On the scales associated with near-infrared reflectance difference spectra. *Applied Spectroscopy, 49*(6), 765–772.

Durig, J. R., Zunic, W. M., Costner, T. G., & Guirgis, G. A. (1993). Fourier transform Raman spectroscopy of brightly colored commercially available dyestuffs and pigments. *Journal of Raman Spectroscopy, 24*(5), 281–285. https://doi.org/10.1002/jrs.1250240504

El-Mesery, H. S., Mao, H. P., & Abomohra, A. (2019). Applications of non-destructive technologies for agricultural and food products quality inspection. *Sensors, 19*(4). https://doi.org/10.3390/s19040846

Fernandes, A. M., Oliveira, P., Moura, J. P., Oliveira, A. A., Falco, V., Correia, M. J., et al. (2011). Determination of anthocyanin concentration in whole grape skins using hyperspectral imaging and adaptive boosting neural networks. *Journal of Food Engineering, 105*(2), 216–226.

Gazzola, D., Vincenzi, S., Marangon, M., Pasini, G., & Curioni, A. (2017). Grape seed extract: The first protein-based fining agent endogenous to grapes. *Australian Journal of Grape and Wine Research, 23*(2), 215–225. https://doi.org/10.1111/ajgw.12268

Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy, 39*(3), 491–500.

Guo, W., Zhao, F., & Dong, J. (2016). Nondestructive measurement of soluble solids content of kiwifruits using near-infrared hyperspectral imaging. *Food Analytical Methods, 9*(1), 38–47. https://doi.org/10.1007/s12161-015-0165-z

Laborde, A., Jaillais, B., Bendoula, R., Roger, J.-M., Jouan-Rimbaud Bouveresse, D., Eveleigh, L., et al. (2020). A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy, 28*(1), 25–36. https://doi.org/10.1177/0967033519891594

Mahesh, S., Jayas, D. S., Paliwal, J., & White, N. D. G. (2015). Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of canadian wheat. *Food and Bioprocess Technology, 8*(1), 31–40. https://doi.org/10.1007/s11947-014-1381-z

Nogales-Bueno, J., Baca-Bocanegra, B., Hernández-Hierro, J. M., Garcia, R., Barroso, J. M., Heredia, F. J., et al. (2021). Assessment of total fat and fatty acids in walnuts using near-infrared hyperspectral imaging. *Frontiers of Plant Science, 12*, Article 729880. https://doi.org/10.3389/fpls.2021.729880

Nogales-Bueno, J., Feliz, L., Baca-Bocanegra, B., Hernández-Hierro, J. M., Heredia, F. J., Barroso, J. M., et al. (2020). Comparative study on the use of three different near infrared spectroscopy recording methodologies for varietal discrimination of walnuts. *Talanta, 206*, Article 120189. https://doi.org/10.1016/j.talanta.2019.120189

Nogales-Bueno, J., Hernández-Hierro, J. M., Rodríguez-Pulido, F. J., & Heredia, F. J. (2014). Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: A preliminary approach. *Food Chemistry, 152*, 586–591.

Pojić, M., Mišan, A., & Tiwari, B. (2018). Eco-innovative technologies for extraction of proteins for human consumption from renewable protein sources of plant origin. *Trends in Food Science & Technology, 75*, 93–104. https://doi.org/10.1016/j.tifs.2018.03.010

Qin, J., & Lu, R. (2008). Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest Biology and Technology, 49*(3), 355–365. https://doi.org/10.1016/j.postharvbio.2008.03.010

Rodríguez-Pulido, F. J., Hernández-Hierro, J. M., Nogales-Bueno, J., Gordillo, B., González-Miret, M. L., & Heredia, F. J. (2014). A novel method for evaluating flavanols in grape seeds by near infrared hyperspectral imaging. *Talanta, 122*(0), 145–150.

Rodríguez-Pulido, F. J., Mora-Garrido, A. B., González-Miret, M. L., & Heredia, F. J. (2022). Research progress in imaging technology for assessing quality in wine grapes and seeds. *Foods, 11*(3), 254.

Ruggieri, L., Cadena, E., Martínez-Blanco, J., Gasol, C. M., Rieradevall, J., Gabarrell, X., … Sánchez, A. (2009). Recovery of organic wastes in the Spanish wine industry. Technical, economic and environmental analyses of the composting process. *Journal of Cleaner Production, 17*(9), 830–838. https://doi.org/10.1016/j.jclepro.2008.12.005

Scheibelhofer, O., Wahl, P. R., Larchevêque, B., Chauchard, F., & Khinast, J. G. (2018). Spatially resolved spectral powder analysis: Experiments and modeling. *Applied Spectroscopy, 72*(4), 521–534. https://doi.org/10.1177/0003702817749839

Shenk, J. S., & Westerhaus, M. O. (1995). *Routine operation,calibration, development and network system management manual*. Silver Spring, Maryland: NIRSystems.

Siesler, H. W., Ozaky, Y., Kawata, S., & Heise, H. M. (2002). *Near infrared spectroscopy: Principles, instruments, applications*. Weinheim, Germany: Wiley-VCH.

Su, W. H., & Sun, D. W. (2018). Multispectral imaging for plant food quality analysis and visualization. *Comprehensive Reviews in Food Science and Food Safety, 17*(1), 220–239. https://doi.org/10.1111/1541-4337.12317

Sun, D. W. (2010). *Hyperspectral imaging for food quality analysis and control*. San Diego, California: Elsevier Science & Technology.

Tian, X., Li, J., Wang, Q., Fan, S., & Huang, W. (2018). A bi-layer model for nondestructive prediction of soluble solids content in apple based on reflectance spectra and peel pigments. *Food Chemistry, 239*, 1055–1063. https://doi.org/10.1016/j.foodchem.2017.07.045

Wang, Q., & He, Y. (2019). Rapid and nondestructive classification of Cantonese sausage degree using hyperspectral images. *Applied Sciences, 9*(5), 822.

Zhao, X., Wang, W., Ni, X., Chu, X., Li, Y.-F., & Sun, C. (2018). Evaluation of near-infrared hyperspectral imaging for detection of peanut and walnut powders in whole wheat flour. *Applied Sciences, 8*(7), 1076.

Zhou, T., Zhang, T., Liu, W., & Zhao, G. (2011). Physicochemical characteristics and functional properties of grape (*Vitis vinifera* L.) seeds protein. *International Journal of Food Science and Technology, 46*(3), 635–641. https://doi.org/10.1111/j.1365-2621.2010.02532.x

Zhu, M., Huang, D., Hu, X. J., Tong, W. H., Han, B. L., Tian, J. P., et al. (2020). Application of hyperspectral technology in detection of agricultural products and

food: A review. *Food Science and Nutrition, 8*(10), 5206–5214. https://doi.org/10.1002/fsn3.1852