Mapping of fatty acids composition in shelled almonds analysed in bulk using a Hyperspectral Imaging System

Irina Torres[a], Dolores Pérez-Marín[b,*], Miguel Vega-Castellote[a], María-Teresa Sánchez[a,*]

[a] *Department of Bromatology and Food Technology, ETSIAM, University of Cordoba, Rabanales Campus, 14071 Córdoba, Spain.*

[b] *Department of Animal Production, ETSIAM, University of Cordoba, Rabanales Campus, 14071 Córdoba, Spain.*

**Corresponding authors*. Tel.: +34 957 212576; fax: 34 957 212000

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

Abstract

The determination of the fatty acid profile in food products is an important issue, as it serves as a guide to consumers who wish to follow healthy diets. Hyperspectral Imaging (HSI), permits rapid, non-destructive quality evaluation in foods by integrating the spatial dimension of the composition distribution. The aim of this research was to measure the fatty acid profile of almonds using HSI in 149 samples of shelled sweet and bitter almonds. In addition, we analysed the inter- and intra-kernel distribution of fatty acids for both type of almonds. Shelled sweet and bitter almonds were scanned in bulk by reflectance HSI (946.6-1648.0 nm) and then analysed by gas chromatography to determine their fatty acid composition. Next, we built quantitative prediction models using Partial Least Squares (PLS) regression and tested two validation strategies — mean spectrum and pixel-by-pixel. The developed HSI calibration models showed a good performance when quantifying oleic and linoleic acids, while the models developed could be used for screening purposes for the rest of the fatty acids analysed and for the oleic to linoleic ratio. The results obtained confirm that HSI can be considered a promising approach for estimating fatty acids and their inter- and intra-kernel distribution.

Keywords: Shelled almonds; Hyperspectral Imaging; Fatty acid mapping; Non-destructive analysis in bulk

1. Introduction

The almond, the fruit of the almond tree (*Prunus dulcis*), is grown mainly for its seed, whose nutritional characteristics bring major health benefits due mainly to its high content of healthy fats (60 % of the total kernel mass), fibre, proteins, minerals and vitamins B and E (Askin, Balta, Tekintas, Kazankayab, & Balta, 2007; Yildirim, Akinci-Yildirim, Şan, & Sesli, 2016).

Almond lipid content is mainly composed, in decreasing order, of oleic (C18:1), linoleic (C18:2), palmitic (C16:0), stearic (C18:0) and palmitoleic (C16:1) acids (Yada, Lapsley, & Huang, 2011). Its lipid fraction is therefore mainly made up of unsaturated fatty acids, whose quantification determines the nutritional qualities and health benefits associated with consumption of this fruit (Chen, Lapsley, & Blumberg, 2006). Thus, due to its high oleic content, the intake of almonds has a similar effect on cardiovascular health to that of olive oil (Hernández & Zacconi, 2009). Another important aspect to take into account is the content of linoleic acid (omega-6), an essential fatty acid which is not synthesized by the body, and therefore must be obtained via the diet (Agunbiade & Olanlokun, 2006). Linoleic acid plays an important role in pro-inflammatory reactions, blood clots and allergic reactions (Mzimbiri, Shi, Liu, & Wang, 2014). Likewise, pamitoleic acid (omega-7), obtained by way of stearoyl CoA desaturase-1 in the synthesis of cis-vaccenic acid, is an omega-7 fatty acid associated with a lower risk of ischemic heart attack (Djoussé et al., 2014).

In addition, measuring the oleic and linoleic acid content in almonds provides information on the state of the fruits and their shelf-life as a product. A high content of oleic acid guarantees oxidative stability, which prevents the fruits from going rancid (Venkatachalam & Sathe, 2006), while high levels of linoleic acid could indicate almond

63 spoilage (Kodad & Socias i Company, 2008), and the oleic/linoleic acid ratio provides an
64 index of physical-chemical quality and a way of evaluating the shelf-life of the fruits
65 (Kodad, Estopañán, Juan, & Socias i Company, 2013). Thus, a high oleic to linoleic ratio
66 has the potential to greatly enhance the marketability of almonds.

67 The study of the saturated fatty acids (SFA) present in almonds - composed
68 mainly of palmitic and stearic acids - is also of interest, since these acids act to protect
69 the fruits from lipid oxidation during the postharvest period, which also lengthens their
70 shelf-life (Pleasance et al., 2018). In addition, measuring this content helps to control the
71 negative effect that SFA can have on human health caused by increased low-density
72 lipoprotein cholesterol (LDL-c) (Zock, 2006; Kodad & Socias i Company, 2008).

73 Almond kernels typically present wide variations in fatty acid composition, oil
74 content, rate of rancidity, oxidative stability, due to the influence of the cultivar (Yildrim
75 et al., 2016), agronomic practices and the environmental conditions during the growing
76 season (Gama, Wallace, Trueman, & Hosseini-Bai, 2018), the stage of maturity and time
77 of harvest (Piscopo, Romeo, Petrovicova, & Poiana, 2010) and the postharvest storage
78 (Kazantzis, Nanos, & Stavroulakis, 2003). This requires the use of real time methods to
79 assess the almond quality along the supply chain in order to identify low quality kernels
80 and batches, and thus guarantee the correct nutritional labelling of product batches
81 received and processed by the industry.

82 At present, the traditional analytical methods to determine fatty acid profile in
83 almonds are generally destructive, time-consuming and high-cost (Yada et al., 2011), and
84 therefore do not meet the requirements for real-time control on industrial production lines.
85 Hyperspectral imaging (HSI) is an emerging technique which can be used for this purpose
86 (Dale et al., 2013). HSI combines the advantages of spectroscopy and artificial vision,
87 providing both spectral and spatial information that can reflect internal physic-chemical

88 characteristics of food products. Thus, while conventional near infrared spectroscopy
89 (NIRS) only provides the mean value of the parameters measured in the product but not
90 their spatial distribution, HSI provides a complete spectrum at each pixel location in the
91 product analysed, facilitating the mapping of the spatial distribution of the different
92 physic-chemical components in the sample (Boldrini, Kessler, Rebnera, & Kessler, 2012;
93 Huang, Liu, & Ngadi, 2014; Pu, Feng, & Sun, 2015; Lu, Huang, & Lu, 2017). However,
94 the signal pre-processing and data management stage is even more complex than with
95 NIRS, with huge amounts of data needing to be processed (Riccioli, Pérez-Marín, &
96 Garrido-Varo, 2018). Similarly, although scientific literature exists on HSI using the short
97 wavelength infrared region — up to 1100 nm — there is a conspicuous gap of using the
98 extended NIR region up to 1700 nm or even to 2500 nm (Qin, Chao, Kim, Lu, & Burks,
99 2013; Liu, Zeng, & Sun, 2015).

100 In this context, no references have been found in the literature related to measuring
101 the fatty acid profile in intact shelled almonds using HSI technology, and only two studies
102 have been published in legumes: one, a review highlighting the importance of using HSI
103 technology to study the fatty acid profile in peanuts (Mzimbiri et al., 2014) and another,
104 in which single soybeans of different varieties were classified according to their oleic and
105 linoleic contents (Fu, Zhou, & Scaboo, 2019).

106 The aim of this study was therefore to quantify and map, at the laboratory scale,
107 the main fatty acids of shelled almonds analysed in bulk using a line-scan hyperspectral
108 reflectance imaging system working in the NIR (946.6 to 1648.0 nm) range to establish
109 the nutritional quality of the product on receipt and in the sorting lines in the industry.

110

111 **2. Material and methods**

112

113 *2.1. Sampling and reference analysis*

114

115 A total of 149 samples of shelled almonds — consisting of 89 samples of sweet
116 almonds (*Prunus dulcis* Mill., cv. 'Antoñeta', 'Belona', 'Guara', 'Lauranne', 'Soleta', and
117 'Vairon') and 60 samples of bitter almonds of non-specific varieties were analysed. The
118 samples were collected during the 2018-2019 season. On arrival at the laboratory, the
119 almonds were immediately placed in dark, refrigerated storage. Prior to measurement,
120 each sample was left to stabilize at the laboratory temperature of 20 ºC.

121 The fatty acid (FA) profile, used as reference data to develop the prediction
122 models, was determined using gas chromatography. The methyl esters of the fatty acids
123 with hexane were extracted, using a PerkinElmer Sigma 3D chromatograph with an FID
124 detector and an automatic injection system (OJEC, 1991). Values were expressed as g per
125 100 g of the total FA content. The five main fatty acids (palmitic (C16:0), palmitoleic
126 (C16:1), stearic (C18:0), oleic (C18:1), and linoleic (C18:2)), as well as the ratio between
127 the oleic and linoleic acids (O/L), were used for the calibration development. All the
128 analytical measurements were performed in duplicate and the standard error of laboratory
129 (SEL) was calculated from these replicates (Table 1).

130

131 *2.2 Hyperspectral imaging acquisition*

132

133 Spectral images were acquired in reflectance mode using a laboratory-based push-
134 broom HSI system. The HSI system consisted of the following parts: 1. a charge-coupled
135 device (CCD) camera with a spatial resolution of 320 × 256 pixels (model Xeva-FPA-
136 1.7-320, Xenics, Leuven, Belgium); 2. a C-mount objective lens (F1.4 25-mm compact
137 lens, Schneider Optics, Hauppauge, NY, USA); 3. a line scan spectrograph (Specim

138 ImSpector V10E, Oulu, Finland) working in the range of 946.6 to 1648.0 nm with a
139 spectral resolution of 3.3 nm; 4. two lamps of 250 W located at a 45° angle to the sample;
140 5. a conveyor belt system (Velmex, Inc., Bloomfield, NY, USA), which moved the sample
141 across the camera's field of view.

142 From each sample, about 100 g were uniformly distributed on a black plastic plate
143 (12.5 x 17.5 cm). were uniformly distributed on a black plastic plate (12.5 x 17.5 cm). To
144 obtain square pixels, the conveyor belt was set up to move at 0.39 mm/scan and the
145 number of lines was 450, obtaining a hypercube of dimensions 450 x 320 x 212. A dark
146 current image was obtained each hour by covering the camera lens and a white reference
147 was collected immediately after the dark current image, using a 99 % reflectance standard
148 (Spectralon^{TM}, SRS-99-10, Labsphere, Inc., North Sutton, NH, USA).

149

150 *2.3. Hyperspectral image processing. Spectral profile extraction*

151

152 Data analysis was performed using Matlab v. 2015a, equipped with the PLS and
153 the Image Processing toolboxes (The Mathworks, Inc., Natick, MA, USA).

154 The reflectance value of each sample was initially calculated as the difference
155 between the intensity of the sample and dark reference divided by the difference between
156 the white and dark references (Kim, Chen, & Mehl, 2001). Once the images were
157 corrected, segmentation was applied to remove the background and to extract the Region
158 of Interest (ROI) in each sample. To do this, the difference between the images obtained
159 at 1009.80 and 1541.60 nm was calculated and then applied a threshold value of 0.08 to
160 the resultant image. From this procedure, a binary image (mask) was obtained for each
161 sample, with 0 value for the background and 1 for the pixels corresponding to almonds.
162 To find the mean spectra of each sample, the mask was applied to the reflectance spectral

data for each image; next, all the spectra extracted from the pixels not identified as background were averaged to obtain a mean spectrum per sample, producing a total of 149 spectra.

2.4. Model building and evaluation

Before developing the models, a Principal Component Analysis (PCA) was carried out to study the structure and variability of the population. The Q residuals and Hotelling's T^2 statistics, which represent how far each sample is from the center of the population, were used to detect the possible spectral outliers (Biancolillo & Marini, 2018). After that, the dataset composed of the average spectrum for each sample was split into calibration and validation samples using the Kennard-Stone method, which selects the samples for calibration set based on the Euclidean distance (Kennard & Stone, 1969; Naes, Isaksson, Fearn, & Davies, 2002). Thus, the calibration set was made up of 104 samples and the remaining 43 samples constituted the validation set.

Partial Least Squares (PLS) regression was used to develop the calibration models, applying venetian blinds for cross validation (10 splits). For each analytical parameter, Standard Normal Variate (SNV) was used as spectral pre-processing for scatter correction (Barnes, Dhanoa, & Lister, 1989), and the first and second Savitsky-Golay derivatives treatments were also tested. The best calibration models for each parameter were selected by statistical criteria, using the coefficient of determination for calibration (R^2_c), the standard error of calibration (SEC), the coefficient of determination for cross validation (R^2_{cv}), the standard error of cross validation (SECV) and the residual predictive deviation for cross validation (RPD_{cv}), calculated as ratio of the standard deviation (SD) of the reference data for calibration to the SECV.

188

189 *2.5. Validation strategies*

190

191 Two strategies were considered for validation purposes for the parameters
192 analysed. For Strategy I, the best calibration models were subjected to external validation
193 using the mean spectrum for each sample extracted from the ROI and the predictive
194 ability of the models was evaluated following the protocol outlined by Windham,
195 Mertens, and Barton II (1989), based on the following statistics: standard error of
196 prediction (SEP), standard error of prediction corrected for bias (SEP(c)), bias, coefficient
197 of determination for external validation (R^2_p) and slope. Generally, for validation groups
198 containing nine or more samples, the following control limits are assumed: Limit Control
199 SEP(c) = 1.30 x SEC, Limit Control bias = ± 0.60 x SEC and minimum value of 0.6 for
200 r2p and slope value between 0.9 – 1.1.

201 For strategy II, we performed the pixel-by-pixel prediction (mapping) of the
202 images corresponding to the validation set using 12 samples, randomly selected images
203 from the validation set. Initially, the validation of all the pixels in the images
204 corresponding to the ROIs was carried out. After that, taking into account that the
205 calibration models had been developed using the mean spectrum extracted from the ROI
206 and the average reference value obtained for that sample, we assumed that when the PLS
207 model was applied to an image for its pixel-by-pixel prediction, some pixel values would
208 not be included within the available calibration range (Chaudhry et al., 2020). For this
209 purpose, we excluded from the prediction map those pixels whose predicted values fell
210 outside the calibration range ± 2 x SECV (Westerhaus, 1989; Williams, 2001). Finally,
211 we compared the results obtained using both pixel-by-pixel validation strategies — total

of pixels and pixels within calibration range ± 2 x SECV—in terms of the standard error of prediction (SEP).

3. Results and discussion

3.1. Main features of the calibration and validation sets

After applying the PCA and before the calibration and validation sets were selected, a total of 2 samples, which presented Hotelling's T^2 values greater than the limit established for a confidence level of 95 %, were identified as outliers. A detailed analysis of the outliers showed that these samples had a high content in stearic and oleic acids. We therefore removed these two samples, thus leaving a set composed of 147 available samples.

Table 1 shows the number of samples, range, mean, standard deviation (SD) and coefficient of variation (CV) of the calibration and validation sets for the different parameters analysed. The Kennard-Stone method proved to be suitable for selecting the calibration and validation sets. It can be appreciated that both sets displayed similar statistics for all parameters, and that the validation set ranges laid within those of the calibration set, which allows to confirm that the validation set is representative of the whole range of variance.

The parameter with the greatest variability is the O/L ratio, with a CV of 26.90 % and 23.57 % for the calibration and validation sets, respectively. The great variability showed for this parameter is due to the fact that it includes a wide variation of linoleic acid ($CV_{calibration}$ = 16.94 % and $CV_{validation}$ = 20.24 %) among the different almond cultivars included in this work (Kodad & Socias i Company, 2008). Likewise, the stearic

and palmitoleic acid also showed considerable variability ($CV_{calibration}$ = 20.74 % and 18.95 %; $CV_{validation}$ = 17.10 % and 21.02 %, respectively). However, for the oleic ($CV_{calibration}$ = 6.05 % and $CV_{validation}$ = 6.40 %) and palmitic ($CV_{calibration}$ = 9.22 % and $CV_{validation}$ = 8.30 %) acids, the sets showed the lowest variability. These results also agree with those reported in a study on the fat composition of 20 almond cultivars by Zamany, Samadi, Kim, Keum, and Saini (2017), who stated that stearic and linoleic acids showed the greatest variability in the fatty acid profile. However, it must be highlighted that it is important to have sets not only with high variability, but also with a uniform distribution of the samples along the variation range. Thus, for stearic acid, despite the wide variability displayed for both the calibration and validation sets, the samples from the calibration set were not uniformly distributed throughout the entire available range (Fig. 1), with certain areas underrepresented, which can affect the robustness of the models developed and their validation.

3.2. Development and validation of models for the prediction of the fatty acid composition in almonds

Table 2 shows the statistics of the best calibration models obtained for the characterization of whole shelled almonds in terms of their fatty acid profile using an HSI system. For all the fatty acids analysed, the best prediction models were obtained using SNV and the first derivative as spectral pre-treatments.

To predict the SFA analysed (palmitic and stearic acids), the models enabled to discriminate between high, medium and low values of the two acids tested. The models developed to predict oleic and linoleic acids showed a good predictive capacity. In addition, the models devised to predict palmitoleic acid and the O/L ratio showed a

predictive capacity which allowed almonds to be classified as high, medium and low values of these parameters when interpreting the coefficient of determination of cross validation, as proposed by Shenk and Westerhaus (1996) and Williams (2001).

Predicting the fatty acid profile in almonds is of great importance due to its correlation with kernel quality during the storage, processing and transportation of almonds, and to their nutritional value (Kodad et al., 2013; Yildirim et al., 2016; Oliveira et al., 2019). The results obtained here are therefore of great interest to the industry and consumers, since they confirm the feasibility of using HSI as a non-destructive analytical tool which enables not only to monitor the product's shelf life and the oxidative variations during the transportation and storage, but also to specify on food labels the nutritional properties of almonds processed in bulk**.** Despite the importance of measuring the fatty acid profile in almonds in a non-destructive way, no references have been found related to the use of an HSI system for this purpose.

After that, the best models were subjected to external validation using firstly the mean spectrum extracted for each sample (Table 2). Following the protocol outlined by Windham et al. (1989), the SFA models developed met the validation requirements in terms of the standard error of prediction corrected for the bias ($SEP_{(c)}$) for stearic acid and the bias for palmitic acid. However, the models did not meet the validation requirements in terms of R^2_p ($R^2_p > 0.6$), although palmitic acid came close ($R^2_p = 0.58$), as was the case for the slope (0.9–1.1).

For stearic acid, the lower predictive capacity obtained after validating the model using the mean spectrum strategy could be a result of the final distribution of the samples from the calibration and validation groups. As can be seen in the frequency histogram (Fig. 1), around 28 % of the samples from the validation group (12 of the 43 samples) correspond to a range of values (between 1.40 g/100 g and 1.70 g/100 g) with a low

representativeness within the calibration group. This indicates that the calibration model developed for stearic acid does not satisfactorily cover this range of values, and therefore the prediction of samples showing values within this range will be less accurate. This fact can be verified when calculating the SEP values in the different ranges of the parameter interval, which were 0.57 g/100 g and 0.15 g/100 g for the intervals between 1.46–1.70 g/100 g and between 1.70–3.03 g/100 g, respectively, thus showing higher SEPs values in the less representative area of the calibration model chosen.

In addition, for palmitoleic, oleic and linoleic acids, as well as for the O/L ratio, the models complied the validation requirements established by Windham et al. (1989) in terms of R^2_p and bias, whereas the $SEP_{(c)}$ did not lie within the confidence limits and the slope values did not attain the recommended value for palmitoleic, oleic and linoleic acids.

Fig. 2 shows two random sweet and bitter almond samples used for the external validation to visualize the mapping for the different fatty acids analysed. Each sample is accompanied by the reference and mean predicted value for each parameter. In sweet almond kernels, there is clearly a more homogeneous distribution of the different fatty acids than in bitter almond kernels. This may be due to the fact that the samples of sweet almonds analysed belonged to a certain variety which is perfectly suited to later commercialization, while the bitter almond samples consisted of mixtures of different varieties. In turn, a greater difference was detected between the reference and predicted values using the mean spectrum in the samples of bitter almonds, which could be a result of the difficulty in obtaining representative samples of variability, as they were in this case a mixture of varieties obtained for later wet analysis.

3.3. Comparison between the different validation strategies

Table 3 shows the SEP values obtained for the different prediction strategies: the mean spectrum for each sample, the prediction of the total number of pixels in the images, and the prediction of only those pixels within the established range of the calibration set. In the pixel-by-pixel prediction, it is common to find pixels which, when predicted, presented higher or lower values than the range of the calibration group. This could be due to the fact that the PLS model is usually developed by using the mean spectrum of each sample, which might not explain all the variability of each individual almond and/or pixel in the image. Therefore, in order to consider only those pixels whose fatty acid content was represented in the model, as discussed in the Material and Methods section, we discarded all the pixels whose predicted value was outside the ± 2 x SECV calibration range, thus taking into account the predictive uncertainty of the model. The percentage of pixels removed for all the parameters tested is also shown in Table 3.

For the parameters tested, it can be observed that the SEP values obtained when the mean spectrum of the samples was used in validation were lower than those obtained when pixel-by-pixel validation was carried out using all the pixels available (Table 3). This lower degree of error may be due to the closer correlation between the spectral information used to develop the models and that used to perform the validation. We should also add that by averaging the spectra of all the pixels, the sources of error due to the possible existence of extreme pixels and outliers are minimized. Despite the fact that the SEP values obtained when carrying out the prediction pixel by pixel were higher, it should be noted that in this validation strategy, not only did we obtain the predicted value of all the pixels in the sample, but we acquired important information, which makes it possible to establish both the average content of fatty acids of each sample and the spatial distribution of this composition within a batch of almonds, which enables to evaluate the

337 homogeneity or heterogeneity of the batch. It must also be considered that in the pixel-
338 by-pixel validation, the SEP values were calculated using the mean reference data of the
339 samples, and so it can therefore be assumed that this value is the same for all the pixels
340 (Torres & Amigo, 2019). Moreover, the SEPs values for the pixel-by-pixel validation
341 were minimized by deleting non-representative pixels from the image, thus avoiding the
342 extrapolation of the models.

343 For all the parameters analysed, the number of pixels deleted was higher in the
344 case of the bitter almond samples, which may be due to their greater heterogeneity, as
345 mentioned above. Additionally, and with the exception of stearic acid, the mean
346 percentage of deleted pixels ranged between 33.73 % and 41.07 %, thus reducing the SEP
347 values between 15.12 % and 39.13 %. For stearic acid, only 5.27 % of pixels were
348 removed, lowering the SEP value by 0.73 %. For this acid, the existence of a smaller
349 number of pixels out of range may be due to the fact that this parameter could present a
350 lower variability between almonds of the same sample, in line with Kodac and Socias I
351 Company (2008), who found that there were no significant differences for stearic acid
352 content between different genotypes, regions or year of production. This would indicate
353 a greater representativeness of the reference value for the entire sample analysed and, in
354 turn, a lower sampling error.

355 Fig. 3 shows the predicted pixel frequency histogram and the prediction map of
356 one sample for oleic acid for all the pixels available, and for the pixels within the
357 calibration range. Before removing the pixels, the range of predicted values oscillated
358 between 0.98 g/100 g and 127.32 g/100 g, while the calibration range was 59.32–76.70
359 g/100 g. In the distribution map obtained after removing the pixels, it can be seen that
360 most of the pixels removed for presenting values outside the calibration range
361 corresponded to the outer edge of the almonds. This means we could be dealing with areas

in shadow, which are extremely difficult to delete when performing image segmentation. What is more, pixels from the central area of the almonds were also removed, which may be linked to the texture; since it has a curved surface, this area could present higher intensity levels than the rest.

4. Conclusions

The results obtained demonstrated the viability of HSI for predicting the fatty acid profile in intact shelled almonds analysed in bulk, thus enabling almonds to be characterised by their lipid composition without any previous grinding or extraction process. HSI technology can be applied at a single pixel level, with the potential to provide mapping information on the distribution of the fatty composition in both the batch and the individual kernels. This approach can be extremely useful at an industrial level for detecting batches with high heterogeneity in fatty acid composition, especially when dealing with a mixture of different qualities or varieties. It is also of special interest as it can provide precise nutritional labelling of the product when sold in packaged formats.

CRediT authorship contribution statement

Irina Torres: Data acquisition, Methodology, Formal analysis, Investigation, Software, Data curation, Validation, Writing - original draft, Writing - review & editing, Visualization. **Dolores Pérez-Marín**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Miguel Vega-Castellote**: Data acquisition, Formal analysis, Software, Data curation, Writing - original

draft, Writing - review & editing, Visualization. **María-Teresa Sánchez**; Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgements

This research was carried out as part of the research project P-12018024 'Measuring the quality of almonds grown in the Guadalquivir Valley (Cordoba)', funded by Desarrollo y Aplicaciones Fitotécnicas, DAFISA. The funding source was not involved in any stage of writing of this article. The authors are grateful to Mrs. María Carmen Fernández from the Animal Production Department for her technical assistance.

References

Agunbiade, S.O., & Olanlokun, J.O. (2006). Evaluation of some nutritional characteristics of Indian almond (*Prunus amygdalus*) nut. *Pakistan Journal of Nutrition*, 5, 316–318. https://doi.org/10.3923/pjn.2006.316.318.

411 Askin, M.A., Balta, M.F., Tekintas, F.E., Kazankayab, A., & Balta, F. (2007). Fatty acid
412 composition affected by kernel weight in almond [(*Prunus dulcis* Miller) D.A.
413 Webb.] genetic resources. *Journal of Food Composition and Analysis*, 20, 7–12.
414 https://doi.org/10.1016/j.jfca.2006.06.005.
415 Barnes, R.J., Dhanoa, M.S., & Lister, S.J. (1989). Standard normal variate transformation
416 and de-trending of near infrared diffuse reflectance spectra. *Applied Spectroscopy*,
417 43, 772–777. https://doi.org/10.1366/0003702894202201.
418 Biancolillo, A., & Marini, F. (2018). Chemometric methods for spectroscopy-based
419 pharmaceutical analysis. *Frontiers in Chemistry*, 6, 576, 1–14.
420 https://doi.org/10.3389/fchem.2018.00576.
421 Boldrini, B., Kessler, W., Rebnera, K., & Kessler, R.W. (2012). Hyperspectral imaging:
422 a review of best practice, performance and pitfalls for in-line and on-line
423 applications. *Journal of Near Infrared Spectroscopy*, 20, 483–508.
424 https://doi.org/10.1255/jnirs.1003.
425 Chaudhry, M.M.A., Amodio, M.L., Amigo, J.M., de Chiara, M.L.V., Babellahi, F., &
426 Colelli, G. (2020). Feasibility study for the surface prediction and mapping of
427 phytonutrients in minimally processed rocket leaves (*Diplotaxis tenuifolia*) during
428 storage by hyperspectral imaging. *Computers and Electronics in Agriculture*, 175,
429 105575, 1–9. https://doi.org/10.1016/j.compag.2020.105575.
430 Chen, C.Y., Lapsley, K., & Blumberg, J. (2006). A nutrition and health perspective on
431 almonds. *Journal of the Science of Food and Agriculture*, 86, 2245–2250.
432 https://doi.org/10.1002/jsfa.2659.
433 Dale, L.M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Fernández-
434 Pierna, J.A. (2013). Hyperspectral imaging applications in agriculture and agro-

435 food product quality and safety control: a review. *Applied Spectroscopy*, Rev. 48,
436 142–159. https://doi.org/10.1080/05704928.2012.705800.
437 Djoussé, L., Matsumoto, C., Hanson, N.Q., Weir, N.L., Tsai, M.Y., & Gaziano, J.M.
438 (2014). Plasma cis-vaccenic acid and risk of heart failure with antecedent
439 coronary heart disease in male physicians. *Clinical Nutrition*, 33, 478–482.
440 https://doi.org/10.1016/j.clnu.2013.07.001.
441 Elliot, C. (2014). Elliott Review into the Integrity and Assurance of Food Supply
442 Networks –Final Report. A National Food Crime Prevention Framework. HM
443 Government. Available at:
444 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/atta
445 chment_data/file/350726/elliotreview-final-report-july2014.pdf. Accessed 28
446 July 2020.
447 Fu, D., Zhou, J., & Scaboo, A. (2019). Fast measurement of fatty acid in soybean using
448 reflective hyperspectral imaging. ASABE 2019 Annual International Meeting.
449 Paper Number: 1900386. pp. 1–12. https://doi.org/10.13031/aim.201900386.
450 Gama, T., Wallace, H.M., Trueman, S.J., & Hosseini-Bai, S. (2018). Quality and shelf
451 life of tree nuts: A review. *Scientia Horticulturae*, 242, 116–126.
452 https://doi.org/10.1016/j.scienta.2018.07.036.
453 Hernández, S.A., & Zacconi, F. (2009). Sweet almond oil: extraction, characterization
454 and application. *Química Nova*, 32, 1342–1345. https://doi.org/10.1590/S0100-
455 40422009000500044.
456 Huang, H., Liu, L., & Ngadi, M. (2014). Recent developments in hyperspectral imaging
457 for assessment of food quality and safety. *Sensor*s, 14, 7248–7276.
458 https://doi.org/10.3390/s140407248.

459 Kazantzis, I., Nanos, G.D., & Stavroulakis, G.G. (2003). Effect of harvest time and
460 storage conditions on almond kernel oil and sugar composition. *Journal of the*
461 *Science of Food and Agriculture*, 83, 354–359. https://doi.org/10.1002/jsfa.1312.
462 Kennard, R.W., & Stone, L.A. (1969). Computer-aided Design of Experiments.
463 *Technometrics*, 11, 137–148. https://doi.org/10.1080/00401706.1969.10490666.
464 Kim, M.S., Chen, Y.R., & Mehl, P.M. (2001). Hyperspectral reflectance and fluorescence
465 imaging system for food quality and safety. *Transactions of the ASAE*, 44, 721–
466 729. https://doi.org/10.13031/2013.6099.
467 Kodad, O., & Socias i Company, R. (2008). Variability of oil content and of major fatty
468 acid composition in almond (*Prunus amygdalus* Batsch) and its relationship with
469 kernel quality. *Journal of Agricultural and Food Chemistry*, 56, 4096–4101.
470 https://doi.org/10.1021/jf8001679.
471 Kodad, O., Estopañán, G., Juan, T., & Socias i Company, R. (2013). Protein content and
472 oil composition of almond from Moroccan seedlings: genetic diversity, oil quality
473 and geographical origin. *Journal of the American Oil Chemists' Society,* 90, 243–
474 252. https://doi.org/10.1007/s11746-012-2166-z.
475 Liu, D., Zeng, X.A., & Sun, D.W. (2015). Recent developments and applications of
476 hyperspectral imaging for quality evaluation of agricultural products: a review.
477 *Critical Reviews in Food Science and Nutrition*, 55, 1744–1757. doi:
478 https://doi.org/10.1080/10408398.2013.777020.
479 Lu, Y., Huang, Y., & Lu, R. (2017). Innovative hyperspectral imaging-based techniques
480 for quality evaluation of fruits and vegetables: a review. *Applied Sciences*, 7, 189,
481 1–36. https://doi.org/10.3390/app7020189.

Mzimbiri, R., Shi, A.M., Liu, H., & Wang, Q. (2014). A Review: peanut fatty acids determination using hyperspectroscopy imaging and its significance on food quality and safety. *Food Science and Quality Management*, 28, 90–97.

Naes, T., Isaksson, T., Fearn, T., & Davies, A. (2002). *A user-friendly guide to multivariate calibration and classification*. Chichester, UK: NIR Publications.

Official Journal of the European Communities (OJEC). (1991). *Commission Regulation (ECC) No 2568/91 of 11 July 1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis*. Official Journal L 248, 5.9.1991, 1–128.

Oliveira, I., Meyer, A.S., Afonso, S., Aires, A., Goufo, P., Trindade, H., & Gonçalves, B. (2019). Phenolic and fatty acid profiles, α-tocopherol and sucrose contents, and antioxidant capacities of understudied Portuguese almond cultivars. *Journal of Food Biochemistry*, 43, e12887. https://doi.org/10.1111/jfbc.12887.

Piscopo, A., Romeo, F., Petrovicova, B., & Poiana, M. (2010). Effect of the harvest time on kernel quality of several almond varieties (*Prunus dulcis* (Mill.) DA Webb). *Scientia Horticulturae*, 125, 41–46. https://doi.org/10.1016/j.scienta.2010.02.015.

Pleasance, E. A., Kerr, W. L., Pegg, R. B., Swanson, R. B., Cheely, A. N., Huang, G., Parrish, D. R., & Kerrihard, A. L. (2018). Effects of storage conditions on consumer and chemical assessments of raw 'nonpareil' almonds over a two-year period. *Journal of Food Science*, 83, 822–830. https://doi.org/10.1111/1750-3841.14055.

Pu, Y.Y., Feng, Y.Z., & Sun, D.W. (2015). Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: a review. *Comprehensive*

506 *Reviews in Food Science and Food Safety*, 14, 176–188.
507 https://doi.org/10.1111/1541-4337.12123
508 Qin, J., Chao, K., Kim, M.S., Lu, R., & Burks, T.F. (2013). Review. Hyperspectral and
509 multispectral imaging for evaluating food safety and quality. *Journal of Food*
510 *Engineering*, 118, 157–171. http://dx.doi.org/10.1016/j.jfoodeng.2013.04.001.
511 Riccioli, C., Pérez-Marín, D., & Garrido-Varo, A. (2018). Identifying animal species in
512 NIR hyperspectral images of processed animal proteins (PAPs): Comparison of
513 multivariate techniques. *Chemometrics and Intelligent Laboratory Systems*, 172,
514 139–149. https://doi.org/10.1016/j.chemolab.2017.12.003.
515 Shenk, J.S., & Westerhaus, M.O. (1996). Calibration the ISI way. In A.M.C. Davies, &
516 P.C. Williams (Eds.), *Near Infrared Spectroscopy: the future waves* (pp. 198–
517 202). Chichester, UK: NIR Publications.
518 Torres, I., & Amigo, J.M. (2019). An overview of regression methods in hyperspectral
519 and multispectral imaging. In J.M. Amigo (Ed.), *Hyperspectral Imaging–data*
520 *handling in science and technology* (pp. 205–230). Amsterdam, The Netherlands:
521 Elsevier.
522 Venkatachalam, M., & Sathe, S.K. (2006). Chemical composition of selected edible nut
523 seeds. *Journal of Agricultural and Food Chemistry*, 54, 4705–4714.
524 https://doi.org/10.1021/jf0606959.
525 Westerhaus, M.O. (1989). Interpretation of regression statistics. In G.C. Marten, J.S.
526 Shenk, & F.E. Barton II (Eds.), *Near Infrared Spectroscopy (NIRS): analysis of*
527 *forage quality*. Agriculture Handbook vol. 643 (pp. 39). Washington, DC, USA:
528 USDA-ARS, Government Printing Office.

529 Williams, P.C. (2001). Implementation of near-infrared technology. In P.C. Williams, &
530 K.H. Norris (Eds.), *Near-infrared technology in the agricultural and food*
531 *industries* (pp. 145–171). St. Paul, MN, USA: AACC, Inc.
532 Windham, W.R., Mertens, D.R., & Barton II, F.E. (1989). Protocol for NIRS calibration:
533 sample selection and equation development and validation. In G.C. Martens, J.S.
534 Shenk, & F.E. Barton II (Eds.), *Near infrared spectroscopy (NIRS): analysis of*
535 *forage quality*. Agriculture Handbook vol. 643 (pp. 96–103). Washington, DC,
536 USA: USDA-ARS, Government Printing Office.
537 Yada, S., Lapsley, K., & Huang, G. (2011). A review of composition studies of cultivated
538 almonds: macronutrients and micronutrients. *Journal of Food Composition and*
539 *Analysis*, 24, 469–480. https://doi.org/10.1016/j.jfca.2011.01.007.
540 Yildirim, A.N., Akinci ‐ Yildirim, F., Şan, B., & Sesli, Y. (2016). Total oil content and
541 fatty acid profile of some almond (*Amygdalus communis* L.) cultivars. *Polish*
542 *Journal of Food and Nutrition Sciences*, 66, 173–178.
543 https://doi.org/10.1515/pjfns ‐ 2015 ‐ 0032.
544 Zamany, A.J., Samadi, G.R., Kim, D.H., Keum, Y.S., & Saini, R.K. (2017). Comparative
545 study of tocopherol contents and fatty acids composition in twenty almond
546 cultivars of Afghanistan. *Journal of the American Oil Chemists' Society*, 94, 805–
547 817. https://doi.org/10.1007/s11746-017-2989-8.
548 Zock, P.L. (2006). Health problems associated with saturated and trans fatty acids intake.
549 In C.M. Williams, & J. Buttriss (Eds.), *Improving the fat content of foods (*pp. 3–
550 24). Cambridge, UK: Woodhead Publishing Ltd.
551 https://doi.org/10.1533/9781845691073.1.3.
552

553 **Table 1.**

554 Statistics of main fatty acids and oleic to linoleic ratio measured in intact shelled almonds.

Parameter	Calibration set (N = 104 samples)				Validation set (N = 43 samples)				SEL
	Range	Mean	SD	CV (%)	Range	Mean	SD	CV (%)	
Palmitic acid (g/100 g of total fatty acids)	5.32 – 7.70	6.55	0.60	9.22	5.36 – 7.46	6.42	0.53	8.30	0.19
Palmitoleic acid (g/100 g of total fatty acids)	0.34 – 0.77	0.55	0.10	18.95	0.34 – 0.76	0.52	0.11	21.02	0.02
Stearic acid (g/100 g of total fatty acids)	1.46 – 3.39	2.17	0.45	20.74	1.46 – 3.03	1.93	0.33	17.10	0.04
Oleic acid (g/100 g of total fatty acids)	59.32 – 76.70	69.60	4.21	6.05	60.85 – 76.40	71.21	4.55	6.40	0.42
Linoleic acid (g/100 g of total fatty acids)	14.67 – 29.98	20.73	3.51	16.94	14.87 – 28.87	19.52	3.95	20.24	0.30
Oleic/Linoleic ratio	1.98 – 5.22	3.49	0.78	26.90	2.11 – 5.14	3.83	0.90	23.57	-

555 N = Number of samples; SD = standard deviation; CV = coefficient of variation; SEL = standard error of laboratory

Table 2.

Calibration, cross validation and prediction statistics for the best PLS regression models of the main fatty acids and the oleic to linoleic ratio measured on intact shelled almonds.

Parameter	Calibration					Prediction						Limits	
	N	LV	R^2_{cv}	SECV	RPD_{cv}	R^2_p	SEP	$SEP_{(c)}$	Bias	Slope	RPD_p	$SEP_{(c)}$	Bias
Palmitic acid (g/100 g of total fatty acids)	104	11	0.66	0.36	1.67	0.58	0.35	0.36	0.07	0.86	1.50	0.34	± 0.16
Palmitoleic acid (g/100 g of total fatty acids)	104	12	0.68	0.06	1.67	0.63	0.07	0.07	0.02	0.89	1.64	0.05	± 0.02
Stearic acid (g/100 g of total fatty acids)	102	7	0.51	0.32	1.42	0.32	0.27	0.33	0.19	0.86	1.21	0.35	± 0.16
Oleic acid (g/100 g of total fatty acids)	104	12	0.74	2.17	1.94	0.69	2.58	2.65	-0.59	1.13	1.76	1.96	± 0.91
Linoleic acid (g/100 g of total fatty acids)	104	12	0.73	1.83	1.92	0.70	2.22	2.26	0.42	1.16	1.78	1.61	± 0.74
Oleic/Linoleic ratio	104	11	0.67	0.45	1.73	0.68	0.52	0.54	-0.16	1.10	1.75	0.44	± 0.20

N = number of samples of calibration; SNV = standard normal variate; LV = latent variables; R^2_{cv} = coefficient of determination for cross validation; SECV = standard error of cross validation; RPD_{cv} = residual predictive deviation for cross validation; R^2_p = coefficient of determination for prediction; SEP = standard error of prediction; $SEP_{(c)}$ = standard error of prediction corrected for bias; RPD_p = residual predictive deviation for prediction. Limits = Control limits established in the protocol of Windham et al. (1989)

564 **Table 3.**

565 Comparison between the different validation strategies tested and percentage of pixels removed.

Parameter	Validation strategies				Pixels removed (%)		
	Mean spectrum	Pixel-by-pixel					
		Totality of pixels	Pixels within calibration range ± 2 x SECV	ΔSEP (%)			
	SEP	SEP	SEP		Sweet almonds	Bitter almonds	Mean
Palmitic acid (g/100 g of total fatty acids)	0.35	1.24	0.78	37.47	31.20	40.39	34.73
Palmitoleic acid (g/100 g of total fatty acids)	0.07	0.15	0.11	22.74	34.75	51.17	41.07
Stearic acid (g/100 g of total fatty acids)	0.27	0.48	0.48	0.73	2.79	7.76	5.27
Oleic acid (g/100 g of total fatty acids)	2.58	7.21	5.03	30.27	33.97	43.57	37.66
Linoleic acid (g/100 g of total fatty acids)	2.22	4.86	4.12	15.12	33.54	43.34	37.31
Oleic/Linoleic ratio	0.52	1.53	0.93	39.13	30.03	39.65	33.73

566 SEP = standard error of prediction.

567

Fig. 1. Frequency distribution of calibration (orange) and validation (green) sets for stearic acid.

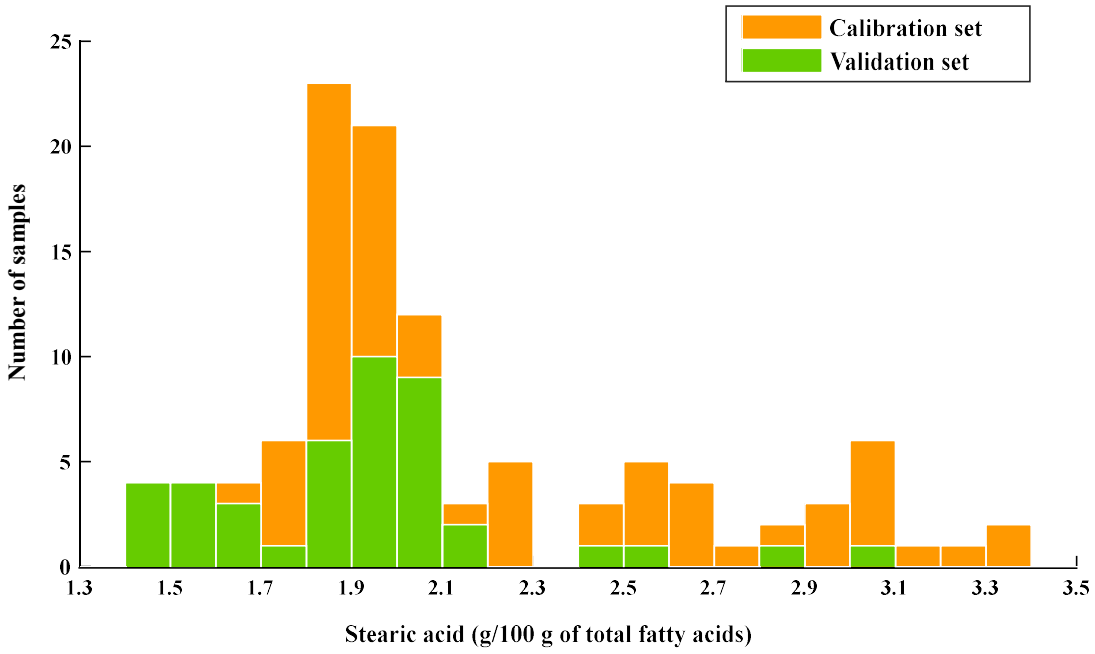
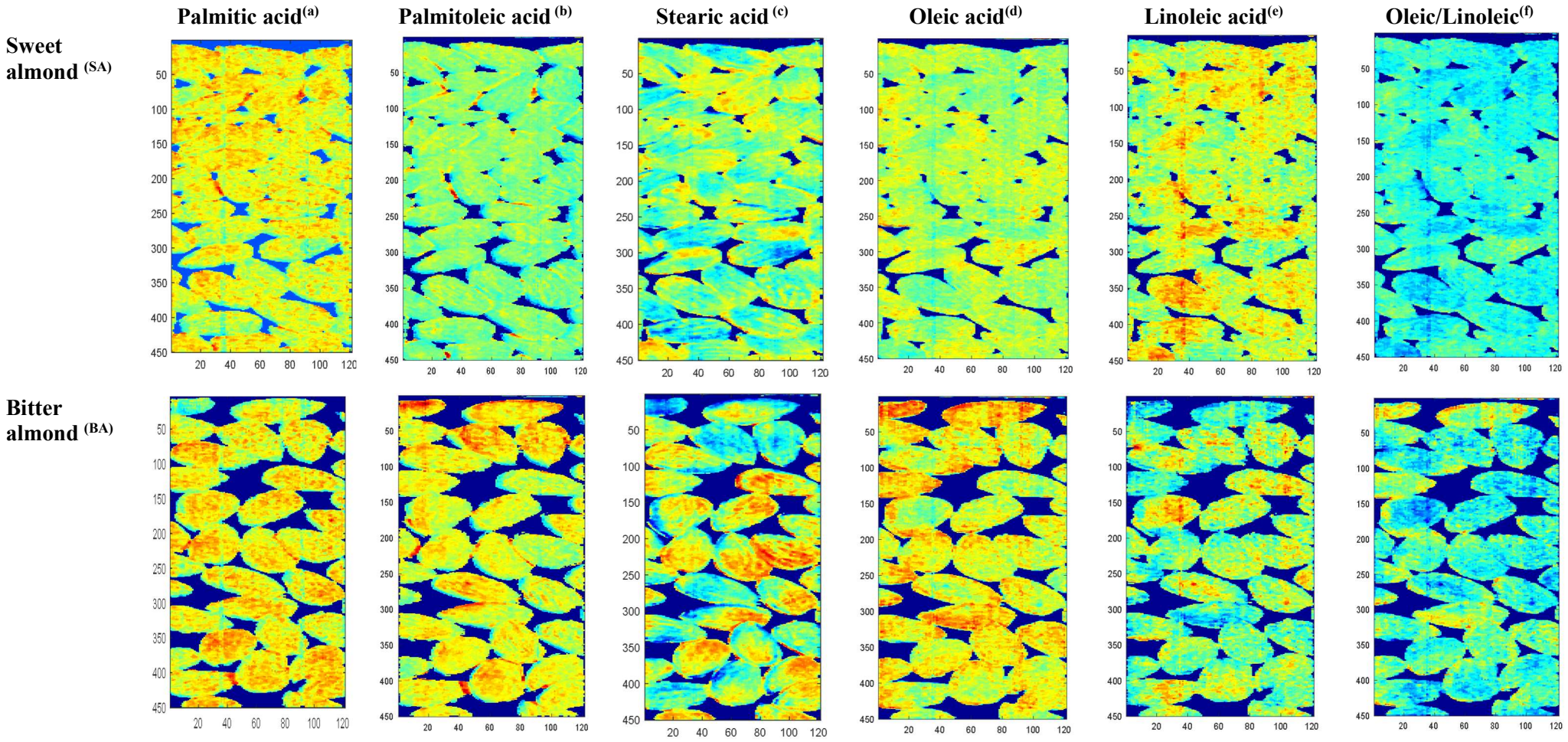


Fig. 2. Distribution maps for the fatty acids analysed and for the oleic to linoleic ratio for sweet and bitter almonds. Reference and predicted mean values are included for each parameter analysed.



SA: [a]Reference: 6.66 g/100 g TFA, Predicted: 6.29 g/100 g TFA; [b]Reference: 0.61 g/100 g TFA, Predicted: 0.53 g/100 g TFA; [c]Reference: 1.69 g/100 g TFA, Predicted: 2.47 g/100 g TFA; [d]Reference: 72.05 g/100 g TFA, Predicted: 72.38 g/100 g TFA; [e]Reference: 18.59 g/100 g TFA, Predicted: 18.54 g/100 g TFA; [f]Reference: 3.88 g/100 g TFA, Predicted: 4.24 g/100 g TFA.

BA: [a]Reference: 6.66 g/100 g TFA, Predicted: 6.29 g/100 g TFA; [b]Reference: 0.61 g/100 g TFA, Predicted: 0.53 g/100 g TFA; [c]Reference: 1.69 g/100 g TFA, Predicted: 2.47 g/100 g TFA; [d]Reference: 72.05 g/100 g TFA, Predicted: 72.38 g/100 g TFA; [e] Reference: 18.59 g/100 g TFA, Predicted: 18.54 g/100 g TFA; [f]Reference: 3.88 g/100 g TFA, Predicted: 4.24 g/100 g TFA.

585 **Fig. 3.** a) Histogram of frequencies for predicting oleic acid following pixel-by-pixel
586 strategy; b) Distribution maps for oleic acid for all the pixels and for the pixels within the
587 calibration range.
588