Estimating Fat Components of Potato Chips Using Visible and Near-Infrared Spectroscopy and a Compositional Calibration Model

Javier Palarea-Albaladejo^{1,2} · José Antonio Cayuela-Sánchez³ · Elena Moriana-Correro⁴

Received: 11 November 2020 / Accepted: 23 July 2021 / Published online: 19 August 2021
© The Author(s) 2021, corrected publication 2021

Abstract

When aiming to assess the fat composition of commercial potato chip products, their diversity and the difficulties to verify the nutritional label of batches of chips by official methods are main challenges. Thus, the possibility of using alternative technologies is of great interest for both the industry and the public administration. Near-infrared spectroscopy (NIRS) is a rapid and non-destructive technique that has been proven useful in different applications in the food industry. However, suitable specific treatments of compositional references with NIRS methods have been until now very scarce in the literature. The nutritional label information is commonly given as percentage content values across several nutritional categories. This formally corresponds with the class of so-called compositional data, for which there are specific statistical methods. This study contributes to ongoing research on the feasibility of Vis/NIR spectroscopy for food nutritional labelling. In particular, a calibration model is formulated to estimate the relative content of fat in potato chips products based on NIR spectral signal that integrates a consistent statistical treatment of the nutritional reference data. The method provides accurate estimates of the fat composition, with this including saturated, monounsaturated, and polyunsaturated types of fat, as well as their total fat percentage (cross-validated overall $R^2=0.88$ and $R^2=0.82$ from ground and fragmented samples respectively) and shows its potential for both nutritional labelling and verification in a rapid and inexpensive manner.

Keywords Potato chips · Compositional data · Monounsaturated fat · Nutritional labelling · Polyunsaturated fat · Saturated fat

Introduction

Nutritional label information typically comprises numerical values referring to the contents across a number of classes of nutritional components. In the European Union, it is compulsory since December 2016 and the published standard (CE 2011) settles the duty for the producers to provide nutritional information as part of the food information to the consumer. Likewise, it is important to verify the compliance of imported products with the established nutritional labelling rules. The nutritional information includes total fat contents, saturated fatty acids (SFA), carbohydrates, sugars, proteins and salt; as well as energy value. The information on other nutrients, such as monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), trans-fatty acids, vitamins, and other healthy compounds, is voluntary according to the standard. Other possible relevant compounds, like cholesterol, can be also voluntarily provided.

✉ José Antonio Cayuela-Sánchez
jacayuela@ig.csic.es

Javier Palarea-Albaladejo
javier.palarea@udg.edu

¹ Biomathematics and Statistics Scotland, JCMB, The King's Buildings, Edinburgh EH9 3FD, UK

² Department of Computer Science, Applied Mathematics and Statistics, University of Girona, 17003 Girona, Spain

³ Instituto de La Grasa, CSIC, Campus de la Universidad Pablo de Olavide, Ed. No 46, Crtra. De Utrera, km 1, 41013 Seville, Spain

⁴ Department Chemical Engineering, University of Sevilla, C/ Virgen de África, 7, 41011 Sevilla, Spain

Total fat content and the distinction between SFA, PUFA and MUFA types of fat are relevant in potato chips products. In relation to this, the most common information included up to date by the producers is total fat and saturated fat, although some products also provide information about MUFA and PUFA. SFA are those without any unsaturation within their chain, whereas MUFA are those fatty acids (FA) with a single unsaturation in their carbon chain. Oleic acid (C18:1) and palmitoleic acid (C16:1) are the major MUFA in chips. Oleic acid unsaturation locates after carbon 9, which are commonly called ω -9. PUFA contains more than one double bond in its backbone. The unfavourable effects of SFA on cardiovascular health are widely proven, as well as the positive effects of MUFA (Hernández et al. 2017; Schwingshackl and Hoffmann 2014). Good human health requires a minimum quantity of these compounds in the diet, such as the essential FA linoleic (C18:2), ω -6, and linolenic (C18:3), this latter called ω -3. Linoleic acid is often present in vegetable oils, which may include also palmitic acid (C16:0), stearic acid (C18:0), and small quantities of other fatty acids as arachidonic (C20:0), and behenic (C22:0).

In practice, analysing all batches of foodstuff by the official methods currently applied is challenging for different reasons. Therefore, both industry and public administration can greatly benefit from the use of rapid technologies for nutritional labelling and verification. Near-infrared spectroscopy (NIRS) is a fast and non-destructive technique widely used in different industries such as food, pharmaceutical, petrochemical industries, amongst others (Flinn 2005; Norris 2005). NIRS is respectful of the environment since it does not need solvents or reagents (Cayuela and García 2017). The suitability of NIRS for quality and composition analysis of potato chips has been reported by several authors (Shiroma and Rodríguez-Solona 2009; Pedreschi et al. 2010; Suthatta et al. 2017). Fernández-Cabañas et al. (2011) report the determination of fatty acids profiles for nutritional labelling of pork dry-cured sausages. However, using NIRS for the specific purpose of food nutritional labelling is almost absent in the literature.

NIRS quantitative methods generally aim to correlate the acquired spectra with reference data. In the context of nutritional labelling, the latter are typically expressed as percentages, mg g^{-1} , ppm, mg kg^{-1} or analogous relative units. Focusing on percentage data, accounting for relative content across nutritional classes (parts of a composition), these live in the numerical space of positive values adding up to 100 that is formally known as the simplex. This fact characterises them as so-called compositional data (Aitchison 1986; Pawlowsky-Glahn et al. 2015) and their key feature is that they carry relative information which does not depend on an arbitrary scale of measurement. Compositional data are naturally constrained (percentages range between 0 and 100) and inter-dependent (increasing percent content in one part

of the composition is necessarily compensated numerically by decreasing it in at least one other part). These special features make the direct use of conventional statistical and chemometrics methods prone to introduce distortions in the data analysis that may lead to deceiving results and scientific conclusions. These include negative bias in correlation measures, singularity of the covariance matrix, predictions outside the possible range, and results dependent on the measurement units (Aitchison 1986; Pawlowsky-Glahn et al. 2015). Aitchison (1986) introduced log-ratio analysis (LRA) as a consistent and well-principled statistical approach for the analysis of compositional data that deals with the above issues. This methodological approach has been further developed in the last decades and successfully introduced in varied scientific fields such as geochemistry, molecular biology or time-use epidemiology where analogous data formats are used.

In brief, the LRA approach focuses on (log-)ratios between the parts of the composition. Log-ratio transformations allow to represent the original (constrained) data on the simplex as (un-constrained) coordinates on the ordinary real space, where standard statistical methods and models can be applied. Once the analysis is conducted on coordinates, the results can be translated back in terms of the original composition by inverse log-ratio transformation. The compositional treatment of food nutritional data in NIRS-based chemometric calibration has only been considered recently (Cayuela-Sánchez et al. 2019; Cayuela-Sánchez et al. 2020).

The present study entails the use of Vis/NIR spectroscopy together with a tailored compositional statistical treatment of reference data for the estimation of fat components in the nutritional label of potato chips products based on fragmented and ground preparations. In the following sections, we provide details about the study setting, sample collection, NIRS processing, reference analysis, chemometric calibration, results and model performance, and conclude with some final remarks.

Material and methods

Study samples

The nutritional composition of potato chips products depends much more on the oils used in frying than on the raw material itself. Therefore, from a certain number on, the samples are repetitive regarding the type of product. The potato chips samples used in this study were purchased from two local supermarkets in Seville (Spain) on two different dates. The range of frying oils showed on the labels of the collected samples was limited, including vegetable oils of regular sunflower, high oleic sunflower, olive

oil, and ‘seeds oil’. This latter category corresponds to non-explicit seed vegetable oils, according to the technical sanitary regulations for edible vegetable oils for the Spanish market (Presidencia del Gobierno 1893). Considering the above, the samples used were deemed representative of the diversity amongst the most common commercial types of potato chips available in the local market. The final set comprised fifteen potato chips samples from different Spanish manufacturers. Two replicates of each of them were processed, hence leading to a final reference data set consisting of 30 sample replicates.

Spectra acquisition

The spectra from the sample replicates were registered with a spectrometer Labspec (Analytical Spectral Devices Inc., Boulder, USA). Labspec has separate detectors for different wavelength ranges. The range 350–1000 nm, including a short interval in the ultraviolet (350–380 nm) and the visible (400–780 nm) ranges, as well as a part of the near-infrared (NIR) wavelengths (781–1000 nm), was analysed by a fixed reflective holographic diode array with a sensitivity of 512 pixels. A holographic fast scanner InGaAs detector cooled at $-25\text{ }^{\circ}\text{C}$ covered the NIR wavelength range of 1000–1800 nm. The same detector coupled with a high order blocking filter analysed the NIR interval 1800–2500 nm. The spectrometer had an internal shutter and automatic offsets correction. The scanning speed was 100 ms, and the repeatability was $6.00 \times 10^{-4} \text{ cm}^{-1} \text{ mol}^{-1}$ (standard deviation on the average absorbance, between 350 and 2500 nm, of five measures of a white tile).

The spectra were registered in reflectance mode. Two alternative settings were prepared: (1) fragmented and (2) ground chips products. For this, the samples were firstly fragmented to slightly homogenize its size and shape and then two replicate scans from each sample replicate were registered. Afterwards, the samples were ground into powder by a manual mortar and the corresponding spectra from two scans were registered in the same manner.

Registration of spectra was conducted by using a ‘Sample Turn Table’ accessory (Analytical Spectral Devices Inc., Boulder, USA) with standard SMA 905 optical fibre connectors, while the samples were held in a quartz Petri dish. The software of the spectrometer was set up to register 10 scans with continuous acquisition to automatically averaging and form a single spectrum. The registering process was controlled by Indico Pro software (Analytical Spectral Devices Inc., Boulder, USA) with a spectral resolution of 1 nm. Two replicate spectra from each sample were registered. The registering process was the same for both fragmented and ground preparations and it took less than a minute for each chip product, all steps included.

Reference analysis

Reference analysis was the same for the fragmented and ground preparations and was applied separately to the 30 final sample replicates in each case (15 samples $\times$ 2 replicates each).

Fat content

The extraction of the oil content of each ground replicate was conducted by the Soxhlet method, using hexane as the solvent and paper cartridges of diameter 35 mm and length 75 mm. The extraction time was 5.5 h, using 250 mL Soxhlet extractor units. Vacuum extraction on a rotary evaporator removed the hexane remaining in the fat. A water bath at $40\text{ }^{\circ}\text{C}$ and tap water circulating through the refrigerant circuit allowed a suitable temperature gradient. Fat content was measured gravimetrically by weighing the flasks containing each sample once at room temperature. It was expressed as total fat percentage (TF) of the product. Then, each fat sample was held in 4 mL vials, hermetically closed, and stored at $4\text{ }^{\circ}\text{C}$ in preparation for FA composition analysis. This latter was performed within a week after the extraction of the fat samples.

Fatty acids composition

The FA compositions were analysed by gas chromatography (GC) according to the IUPAC Standard Method (IUPAC 1987) as fatty acid methyl esters (FAME). An aliquot of 100 to 130 mg of the sample of fat from the potato chips was dissolved in 2 mL heptane. Then, the sample was trans-esterified using 500 μL methanolic potassium hydroxide 2 N. The supernatant was collected after decanting and hosted in 1.5 mL vials for analysis. GC analysis was performed using an Agilent 7697A gas chromatograph (Agilent Technologies, Santa Clara, USA). A capillary column was employed: poly (90% biscyanopropyl-10% cyanopropylphenyl) siloxane, 60 m $\text{\AA}$, 0.25 mm Φ i, and 0.20 μm film thickness. A flame ionization detector (FID) with automatic split injection was used (injection volume equal to 1 μL). Hydrogen was the carrier gas at a flow rate of 1 mL min^{-1} . The injector and detector were fixed at a temperature of $225\text{ }^{\circ}\text{C}$ and $250\text{ }^{\circ}\text{C}$ respectively. The oven was programmed at a beginning temperature of $180\text{ }^{\circ}\text{C}$ (10 min), with a rate of increase of $3\text{ }^{\circ}\text{C min}^{-1}$ up to $220\text{ }^{\circ}\text{C}$ (10 min).

The fractions of SFA, MUFA, PUFA, and total unsaturated fatty acid (TUFA) were calculated and expressed in percentage unit. Thus, SFA was the sum of percentages of FA caprylic (C8:0), capric or decanoic (C10:0), lauric (C12:0), miristic (C14:0), palmitic (C16:0), stearic (C18:0), arachidic (C20:0), behenic (C22:0), and lignoceric

(C23:0). MUFA was the sum of percentages of FA palmitoleic (C16:1), heptadecenoic (C17:1), oleic (C18:1), and eicosenoic (C20:1). PUFA was the sum of percentages of FA linoleic (C18:2), and linolenic (C18:3). TUFA was the sum of percentages of MUFA and PUFA. All the reference analysis was carried out at the Instituto de la Grasa (CSIC; Seville, Spain).

Chemometric pre-processing

The replicate spectra from each sample replicate were firstly averaged, then, the wavelength resolution of the spectra was reduced to a window width of 8 nm. This procedure was the same for fragmented and ground samples. The resulting reflectance data were normalized by applying standard mean normalization. Subsequently, they were transformed into absorbance and treated by the Savitzsky-Golay's first derivative method with polynomial order 2 and smoothing point 3. The adequacy of these treatments has been previously reported (Cayuela et al. 2015). Spectral data pre-processing was carried out using The Unscrambler 9.7 (CAMO Software AS, Norway).

A preliminary assessment of possible patterns and discrepancies in the Vis/NIR spectral data was conducted by principal component analysis (PCA). PCA reduces the spectral variables by finding optimal linear combinations (principal components) of them that successively account for decreasing amounts of total variability in the original data. In particular, the values of the first two principal components (PCA scores, accounting for the highest portion of the original data variability) were used to represent the spectral data as points in an ordinary 2-dimensional scatterplot (PCA scores plot). This included a 95% confidence ellipse to help identifying any potential outlying spectral profiles.

Compositional calibration modelling

To maximize the use of the available data and provide more realistic measures of the calibration model performance, the data (for both fragmented and ground preparations separately) were sequentially split into train and validation data sets following an ordinary cross-validation procedure (see e.g. Kuhn and Johnson 2013). In particular, we considered tenfold cross-validation, so that the sample replicates were randomly allocated to 10 equal size subsets and then nine of them were used to fit the model (training set) and the remaining one was used to test the performance of the model independently (validation set). This was repeated 10 times with each of the subsets used in turn exactly once as validation set. That is, at each of the 10 rounds, 27 sample replicates were used for training and 3 for validation.

Calibration models were fitted to estimate the basic 3-part fat composition (SFA, MUFA, and PUFA) and the TF percentage content separately from the processed Vis/NIR spectral data. Following the LRA approach, the fat composition data were transformed into log-ratio coordinates. These were used as multivariate response in a partial least squares regression model (PLS2) fitted using the ordinary kernel algorithm. In particular, we considered isometric log-ratio (ILR) coordinates (Pawlowsky-Glahn et al. 2015), by which the 3-part composition was represented by two real-valued coordinates defined as follows:

$$\text{ILR}_{\text{FA1}} = \sqrt{\frac{2}{3}} \ln \frac{\text{SFA}}{(\text{PUFA} \cdot \text{MUFA})^{1/2}} \quad (1)$$

$$\text{ILR}_{\text{FA2}} = \sqrt{\frac{1}{2}} \ln \frac{\text{PUFA}}{\text{MUFA}} \quad (2)$$

In this case, the first ILR coordinate (1) represents the balance or trade-off of SFA against PUFA and MUFA contents in the FA composition, whereas the second ILR coordinate (2) refers to the balance between PUFA and MUFA.

The total fat percentage (TF) was treated as a separate 2-part composition (TF, 100 – TF), i.e. a composition formed by the percentage of TF along with the complementary percentage of the remaining contents, and thus represented as a single ILR coordinate given by

$$\text{ILR}_{\text{TF}} = \sqrt{\frac{1}{2}} \ln \frac{\text{TF}}{(100 - \text{TF})} \quad (3)$$

Another univariate response partial least squares calibration model (PLS1) was then fitted to this ILR coordinate on the spectral data.

In brief, the PLS models projected the data onto low dimensions by finding a few linear combinations (PLS components) of the spectral data with the highest correlation to the ILR coordinates. The optimal number of PLS components for each model was determined internally by cross-validation based on the coefficient of determination (R^2) and root mean square error (RMSE), choosing the most parsimonious model amongst those reaching comparable best predictive performance according to the one standard error rule (Kuhn and Johnson 2013). These calculations were conducted using the packages *compositions*, *caret* and *pls* on the R system for statistical computing (R Core Team 2020).

The predictions from the fitted PLS models in ILR coordinates were extracted and transformed back onto the simplex by applying inverse ILR transformation and multiplying by 100 to express them in percentages as originally. Note that alternative ILR coordinate systems could be used representing alternative balances between the FA

components. However, they all are orthogonal rotations of each other and lead to the same final predicted compositions after ILR back-transformation. Hence, the choice of ILR coordinates is not a relevant question for our purpose here. Finally, the results for the class total unsaturated fatty acids (TUFA) were obtained by arithmetically adding the results for PUFA and MUFA.

Evaluation of calibration model performance

Model performance assessment was conducted on ILR coordinates as ordinarily with PLS calibration models based on R^2 and RMSE measures. Values based on the entire reference data sets (training data estimates) and values resulting from the cross-validation process (CV data estimates) were calculated. Additionally, compositional R^2 and metric standard deviation (MSD) were computed as counterparts of the above measures to evaluate the fit for the fat composition as a whole after inverse ILR transformation (Cayuela-Sánchez et al. 2019; Cayuela-Sánchez et al. 2020). Finally, as customary in similar studies in the literature, the correlations between predicted and reference values for the individual components and their percentage relative deviations were computed.

Results and discussion

Potato chips products spectra

The ultraviolet (350–380 nm) and visible (380–780 nm) regions of the analysed spectra were not relevant, even

though an absorbance maximum is showed around 360–380 nm (Fig. 1). They could, in part, represent aromatic compounds or, according to absorbance data described elsewhere (Taniguchi and Lindsey 2018), may relate to compounds produced in the frying.

Fundamental vibrations, mainly involving carbon-hydrogen combinations, induce various overlapping bands in near-infrared spectra because of their first and second overtones (Shenk et al. 2001). A large peak around 1900 nm corresponds to water. According to Salgó and Gergely (2012), carbohydrates peaks are found in three wavelength regions: (i) between 1585 and 1595 nm, (ii) from 2270 to 2280 nm, and (iii) from 2325 to 2335 nm. This is probably associated with overlapping of the C-H bonds vibration and their stretch and deformation. Water-soluble carbohydrates such as fructose, glucose and sucrose show distinct absorption bands around 2275 nm (Chung and Arnold 2000) due to combinations of O-H and C-C stretching (Osborne and Fearn 1986). Fat absorption bands in products from potatoes depend mainly on the plant oil, or animal fat, used in their elaboration. The potato chips products included in the present study had a fatty acid composition mainly associated with sunflower oil. Sunflower oil of high oleic acid content was present in some samples, as well as olive oil and a few cases of palm oil. Hourant et al. (2000) described main near-infrared absorption bands of vegetable oils. A broad absorbance band exists around 1220 nm, probably due to second overtones of C-H and CH=CH- stretching vibrations from oil. There is a high-intensity absorbance peak about 2300 nm, caused by a combination of fundamental vibrations from the C-H groups present in FA (Hourant et al. 2000). The bands

Fig. 1 Absorbance spectra of potato chips ground samples

Table 1 Univariate summary statistics of total fat and fat composition in potato chips samples

	Range	Quartiles			Mean $\pm$ SE*
		Q ₂₅	Q ₅₀	Q ₇₅	
Total fat	31.12–39.83	34.17	36.97	38.74	36.54 $\pm$ 0.49
SFA	7.22–14.92	9.63	10.02	12.40	10.89 $\pm$ 0.40
MUFA	31.27–77.30	32.81	34.06	67.97	46.48 $\pm$ 3.36
PUFA	8.31–59.05	22.65	55.43	56.98	42.63 $\pm$ 3.60
TUFA	85.08–92.78	87.60	89.98	90.37	89.11 $\pm$ 0.40

SFA, saturated fatty acids; PUFA, polyunsaturated fatty acids; MUFA, monounsaturated fatty acids; TUFA, total unsaturated fatty acids (values expressed as percentages)

* Arithmetic mean and standard error (SE) of the mean

mentioned above were confirmed at approximately the same wavelengths in the potato chips products.

Descriptive statistics

Ordinary univariate summary statistics of the reference fat composition and total fat are shown in Table 1. Note that using statistics such as geometric mean and standard deviation (or more generally compositional statistics that consider the fat composition as a whole; see e.g. Pawlowsky-Glahn et al. 2015) would be alternatives consistent with the relative scale of percentage data. However, ordinary (arithmetic) means and geometric means (not shown) were very similar in this case, and hence the most typical measures are displayed in Table 1 to facilitate comparability with previous literature. The percentages of total fat in the data set ranged

between 31.12 and 39.83. As to the fat composition, SFA ranged from 7.22 to 14.92, MUFA was between 31.27 and 77.30, PUFA ranged from 8.31 to 59.05, and TUFA did so between 85.08 and 92.78. Total unsaturated fats represented the largest fraction of the reference fat composition (89.11% vs. 10.89% of SFA in mean) and also showed the highest variability (standard error of the mean equal to 3.60 and 3.36 for PUFA and MUFA respectively). These results agree with the fact that most of the potato chip products analysed used high oleic sunflower or normal sunflower oils in the frying process according to their labels.

Principal component analysis

Once registered, the spectra generated from both preparations (fragmented and ground) were run through PCA. Score plots based on the first and second principal components (PC1 and PC2) are shown in Figs. 2a (fragmented samples) and 2b (ground samples). The total data variability accounted for by these plots was 74.9% and 78.1% for fragmented and ground samples respectively. No groupings were obvious and only a few pairs of replicates showed a borderline profile, with the corresponding points laying around the boundaries of the 95% confidence limits. However, no clear outlying cases were apparent, and thus the entire data set was used in subsequent data processing.

Compositional PLS calibration modelling

The calibration models formulated in ILR coordinates as detailed in Eqs. (1)–(2) for fat composition were fit on spectral data from fragmented and ground preparations

Fig. 2 Principal component analysis of potato chips samples: fragmented (a) and ground (b)

Fig. 3 Fatty acids composition calibration model on NIRS profiles from fragmented potato chips samples: standardised regression coefficients and scatterplot of predicted against actual values from PLS2 model fitted to ILR coordinates ILR_{FA1} (a, c) and ILR_{FA2} (b, d)

separately. The optimal numbers of PLS components were 9 and 14 for the fragmented and ground cases respectively. Figures 3 and 4 show the respective standardised PLS regression coefficients and scatterplot of predicted against actual values from the fitted PLS2 models. Note that interpretation of these coefficients cannot be done in terms of the individual fat components, but in relation to the contrast between components represented by the ILR coordinates. In any case, we here fundamentally use these models as a device for prediction. In this sense, we observe that the scatterplots of predicted versus actual values show good agreement.

The corresponding estimates from the calibration models for total fat based on fragmented and ground samples are shown in Figs. 5 and 6 (using ILR coordinate given in Eq. (3); optimal numbers of PLS components were 9 and 4 respectively). Good agreement between predicted and actual values is also observed here, particularly in the fragmented case.

Note that the total fat replicates of one reference sample were identified as highly influential in the PLS regression

fit when comparing actual and predicted values using the Cook's distance method (Cook 1977). These values most probably corresponded to defective reference analysis. Omitting them notably improved the fit of the total fat models, increasing R^2 from 0.514 to 0.967 for fragmented samples and from 0.539 to 0.800 for ground samples.

The performance statistics based on train and cross-validated data for all the models above are summarised in Table 2.

Both fragmented and ground preparations provided similar calibration results, with train R^2 values over 0.9 for all ILR coordinates except for ILR_{TF} in the ground preparation. The corresponding cross-validated estimates also reached values regarded acceptable. The lowest performance was observed for total fat. This might be related to the Soxhlet reference method, which presented a larger deviation between replicates than gas chromatography. Additionally, the RMSEs of the models can be compared with the dispersion (standard deviation) observed in the reference data as baseline to assess any improvement in the error of prediction (all consistently computed in ILR coordinates). The

Fig. 4 Fatty acids composition calibration model on NIRS profiles from ground potato chips samples: standardised regression coefficients and scatterplot of predicted against actual values from PLS2 model fitted to ILR coordinates ILR_{FA1} (a, c) and ILR_{FA2} (b, d)

Fig. 5 Total fat calibration model on NIRS profiles from fragmented potato chips samples: standardised regression coefficients (a) and scatterplot of predicted against actual values (b) from PLS1 model fitted to ILR coordinate ILR_{TF}

standard deviations of the ILR coordinates from the reference data were 0.540, 0.595 and 0.082 for coordinates (1), (2) and (3) respectively. These were then markedly larger than the corresponding RMSE values of the models (either based on train or CV data as reported

in Table 2) and further supported the good performance of the fitted calibration models.

Fig. 6 Total fat calibration model on NIRS profiles from ground potato chips samples: standardised regression coefficients (a) and scatterplot of predicted against actual values (b) from PLS1 model fitted to ILR coordinate ILR_{TF}

Table 2 Performance statistics of compositional calibration models based on ILR coordinates of fatty acids composition and total fat percentage

	Fragmented			Ground		
	ILR_{FA1}	ILR_{FA2}	ILR_{TF}	ILR_{FA1}	ILR_{FA2}	ILR_{TF}
	R^2					
Train	0.946	0.964	0.967	0.991	0.994	0.800
CV	0.771	0.856	0.832	0.827	0.871	0.744
	$RMSE$					
Train	0.123	0.111	0.013	0.049	0.044	0.032
CV	0.254	0.222	0.035	0.221	0.211	0.042

ILR , isometric log-ratio coordinates; R^2 , coefficient of determination; CV , cross validated; $RMSE$, root mean squared error

Table 3 Compositional performance statistics of calibration models for the estimation of fat components

	Fragmented				Ground			
	SFA	MUFA	PUFA	TUFA	SFA	MUFA	PUFA	TUFA
R^2	0.952				0.993			
R^2_{CV}	0.820				0.879			
MSD	0.125				0.047			
MSD_{CV}	0.242				0.201			
r_{CV}	0.479	0.919	0.921	0.479	0.656	0.948	0.946	0.656
RD_{CV} (%)	15.590	15.877	19.679	1.757	15.082	12.335	14.871	1.761

R^2 , coefficient of determination; R^2_{CV} , cross validated coefficient of determination; MSD , metric standard deviation; r_{CV} , cross-validated coefficient of correlation between observed and predicted values; RD_{CV} , percentage relative deviations

Estimation of fat composition

Estimates of (SFA, MUFA, PUFA) expressed in percentage based on the Vis/NIR spectral data were obtained from the fitted calibration models by inverse ILR transformation. The compositional performance measures are summarised in Table 3. Cross-validated coefficients of correlation between actual and predicted values for each component (r_{CV}) along with percentage relative deviations (RD_{CV}), as commonly used in component-wise assessment, are also displayed in Table 3.

As reported above based on ILR coordinates, the results from the ground preparation were better than those from fragmented preparation. The predictions of MUFA and PUFA using ground samples were satisfactory, with r_{CV} equal to 0.948 and 0.946 respectively. However, the estimation of SFA was more problematic, with a moderate r_{CV} equal to 0.656. This might be in part related with the limited size of the reference data set, which affects the partition into train and validation sets in the cross-validation procedure. Using more data would contribute to reduce the variability between

Fig. 7 Ternary diagrams of observed versus predicted fat compositions for fragmented (left) and ground (right) potato chips samples

cross-validation runs. Moreover, note that these individual measures of performance are not actually very adequate for compositional variables as they ignore the intrinsic interplay between the parts of the composition as fractions of 100. For example, the correlations coincide by construction for SFA and PUFA. Recall that the basic percentage composition consists of SFA, MUFA and PUFA; therefore, SFA is $100 - (\text{MUFA} + \text{PUFA})$ and then SFA is just $100 - \text{PUFA}$. This linear relationship implies the correlation values will be always the same for both components. This stresses again the pitfalls of considering measures on the individual parts of a composition alone, and the convenience of using compositional measures to summarise the data and assess model performance. With this regard, the ground samples provided compositional R^2 and R^2_{CV} equal to 0.993 and 0.879 respectively, with associated low deviations in prediction as pointed out by the MSD measures (Table 3). These same statistics for the fragmented samples were 0.952 and 0.820. As a compositional counterpart to the common scatterplots of observed versus predicted values, Fig. 7 shows these for the entire percentage fat compositions using ternary plots (for fragmented samples on the left-hand side and ground samples on the right-hand side).

Conclusions

This study adds further evidence about the feasibility of using Vis/NIR spectroscopy for the assessment of nutritional contents in nutritional labels of commercial foodstuff products. Importantly, the method integrates a tailored statistical treatment of compositional reference data that avoids possible incoherencies associated with ordinary statistical procedures. Despite the somewhat limited number of samples

available in this study to train the calibration models, these already showed acceptable ability to estimate the relative contents of fat classes (SFA, MUFA, and PUFA) in potato chips products according to common model performance measures used in the field, based on both train and cross-validation sets. The results for ground potato chips samples were generally somewhat better than for fragmented samples. In particular, the compositional R^2 and R^2_{CV} (from train and cross-validation fat composition data respectively) were 0.993 and 0.879 for ground samples and 0.952 and 0.820 for fragmented samples. This demonstrates that the proposed method is a practicable alternative to current official methods and shows its potential to be used for nutritional labelling and verification in a rapid and inexpensive manner. In any case, this method will require the calibration of specific instruments as well as a periodic validation protocol, as it the case for NIR spectroscopy techniques in general.

Acknowledgements J. Palarea-Albaladejo was partly supported by the Scottish Government's Rural and Environment Science and Analytical Services Division and the Spanish Ministry of Economy and Competitiveness [Ref: RTI2018-095518-B-C21]. J. A. Cayuela-Sánchez thanks the Spanish Council for Scientific Research for supporting their activity as staff.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Ethics approval and consent to participate This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest Javier Palarea-Albaladejo declares that he has no conflict of interest. José Antonio Cayuela-Sánchez declares that he has no conflict of interest. Elena Moriana-Correro declares that he has no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aitchison J (1986) *The Statistical Analysis of Compositional Data*. Chapman and Hall, London
- Cayuela JA, García JF, Moreda W, Pérez MC (2015) Characterization of some olive oil quality aspects by NIRS analysis of its fatty acids and triglycerides. In: Pulkrabová J, Tomaniová M, Nielsen M, Hajšlová J (eds) 7th International symposium on recent advances in food analysis: book of abstracts. Poster, Prague
- Cayuela JA, García JF (2017) Sorting olive oil based on alpha-tocopherol and total tocopherol content using Near-Infrared Spectroscopy (NIRS) analysis. *J Food Eng* 202:79–88. <https://doi.org/10.1016/j.jfoodeng.2017.01.015>
- Cayuela-Sánchez JA, Palarea-Albaladejo J, García-Martín JF, Pérez-Camino MC (2019) Olive oil nutritional labeling by using Vis/NIR spectroscopy and compositional statistical methods. *Innov Food Sci Emerg* 51:139–147. <https://doi.org/10.1016/j.ifset.2018.05.018>
- Cayuela-Sánchez JA, Palarea-Albaladejo J, Zira TP, Moriana-Correro E (2020) Compositional method for measuring the nutritional label components of industrial pastries and biscuits based on Vis/NIR spectroscopy. *J Food Comp Anal* 92:103572. <https://doi.org/10.1016/j.jfca.2020.103572>
- CE (2011) Regulation (EU) No 1169/2011 of the European Parliament and of the council of 25 October 2011 on the provision of food information to consumer
- Chung H, Arnold MA (2000) Near-infrared spectroscopy for monitoring starch hydrolysis. *Appl Spectrosc* 54:277–283. <https://doi.org/10.1366/0003702001949212>
- Cook RD (1977) Detection of influential observation in linear regression. *Technometrics* 19:15–18. <https://doi.org/10.2307/1268249>
- Fernández-Cabañas VM, Polvillo O, Rodríguez-Acuña R, Botella B, Horcada A (2011) Rapid determination of the fatty acid profile in pork dry-cured sausages by NIR spectroscopy. *Food Chem* 124:373–378. <https://doi.org/10.1016/j.foodchem.2010.06.031>
- Flinn P (2005) An average day, or how NIR affects daily live. *NIR News* 16:4–8. <https://doi.org/10.1255/nirn.852>
- Hernández A, Castañer O, Goday A, Ros E, Pintó X, Estruch R, Salas-Salvadó J, Corella D, Arós F, Serra-Majem L, Martínez-González MA, Fiol M, Lapetra J, De la Torre R, López-Sabater MC, Fitó M (2017) The Mediterranean Diet decreases LDL atherogenicity in high cardiovascular risk individuals: a randomized controlled trial. *Mol Nutr Food Res* 61(9):1601015. <https://doi.org/10.1002/mnfr.201601015>
- Hourant P, Baeten V, Morales MT, Meurens M, Aparicio R (2000) Oil and fat classification by selected bands of near-infrared spectroscopy. *Appl Spectrosc* 54(8):1168–1174. <https://doi.org/10.1366/0003702001950733>
- IUPAC (1987) Standard Method 2.302 Standard methods for the analysis of oils, fats and derivatives determination of FAMES by capillary GC. Blackwell Scientific, Oxford
- Kuhn M, Johnson K (2013) *Applied predictive modelling*. Springer, New York
- Norris K (2005) NIR is alive and growing. *NIR News* 16(7):12. <https://doi.org/10.1255/nirn.854>
- Osborne BG, Fearn T (1986) Theory of near infrared spectrophotometry. In: Osborne BG, Fearn T (eds) *Near infrared spectroscopy in food analysis*, 1st edn. Longman Scientific & Technical, Harlow
- Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) *Modeling and Analysis of Compositional Data*. Wiley & Sons, Chichester
- Pedreschi F, Segtnan VH, Knutsen SH (2010) On-line monitoring of fat, dry matter and acrylamide contents in potato chips using near infrared interactance and visual reflectance imaging. *FoodChem* 121:616–620. <https://doi.org/10.1016/j.foodchem.2009.12.075>
- Presidencia del Gobierno (1893) Real Decreto 308/1983 de 25 de enero, que aprueba la Reglamentación Técnico-Sanitaria de Aceites Comestibles. BOE 44:4853–4858. <https://www.boe.es/eli/es/rd/1983/01/25/308>. Accessed 8 Sept 2020
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, 2020. <https://www.r-project.org>
- Salgó A, Gergely S (2012) Analysis of wheat grain development using NIR spectroscopy. *J Cereal Sci* 56:31–38. <https://doi.org/10.1016/j.jcs.2012.04.011>
- Shenk JS, Workman JJ, Westerhaus MO (2001) Application of NIR spectroscopy to agricultural products. In: Burns DA, Ciurcak CW (eds) *Handbook of Near Infrared Analysis*, 2nd edn. Marcel Dekker, New York, pp 419–474
- Shiroma C, Rodríguez-Solana L (2009) Application of NIR and MIR spectroscopy in quality control of potato chips. *J Food Comp Anal* 22:596–605. <https://doi.org/10.1016/j.jfca.2008.09.003>
- Schwingshackl L, Hoffmann G (2014) Monounsaturated fatty acids, olive oil and health status: a systematic review and meta-analysis of cohort studies. *Lipids Health Dis* 13:154. <https://doi.org/10.1186/1476-511X-13-154>
- Suthatta A, Ritthiruangdej P, Kasemsumran S, Therdthai N, Haruthaithanasan V, Ozaki Y (2017) Rapid and nondestructive analysis of deep-fried taro chip qualities using near infrared spectroscopy. *J near Infrared Spec* 25:127–137. <https://doi.org/10.1177/0967033516686655>
- Taniguchi IM, Lindsey JS (2018) Database of absorption and fluorescence spectra of > 300 common compounds for use in PhotochemCAD. *Photochem Photobiol* 94:290–327. <https://doi.org/10.1111/php.12860>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.