

Compositional method for measuring the nutritional label components of industrial pastries and biscuits based on Vis/NIR spectroscopy

José A. Cayuela-Sánchez^{*1}, Javier Palarea-Albaladejo², Tatiana P. Zira³, Elena Moriana-Correro⁴

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

²Biomathematics and Statistics Scotland, JCMB, The King's Buildings, EH9 3FD, Edinburgh, U.K.

³Department Food Science and Human Nutrition, Agricultural University of Athens, 75 Iera Odos, 11855 Athens.

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

*Correspondence author: jacayuela@ig.csic.es

Journal Pre-proof

Highlights

- The basis to a rapid technique for measuring nutritional label components is presented.
- The technique base on Vis-NIRS taking into account the compositional data
- Nutritional components and types of fat were separately modelled
- Nutritional labeling or its verification would be advantageously performed

Abstract

Food nutritional labeling is compulsory in the European Union since 13 December 2016. The analysis of all batches of foodstuff using the official methods is expensive and the food industry seeks cost-efficient solutions for measuring the components of nutritional labels. The information is commonly provided in the form of fractions or parts of a number of nutritional categories, typically expressed in percentage units. This corresponds with so-called compositional data for which specific statistical methods and models have been developed. Compositional data analysis focuses on log-ratios between the parts, so that their relative scale and inherent interplay are accounted for. This study set up and validates a new rapid technique to measure the nutritional makeup of industrial pastries and biscuits for labelling purposes. Nutritional components and types of fat were separately modeled. Compositional partial least squares (PLS) calibration models, formulated through adequate log-ratio coordinates, were fitted based on visible and near infrared spectroscopy (NIRS) and applied to estimate the composition of (1) moisture, total fat, protein, and carbohydrates and (2) saturated, polyunsaturated and monounsaturated fats from intact and ground samples of a variety of types of industrial pastry and biscuit present in the market. This investigation proposes a useful and methodologically well-founded approach for measuring the components of the nutritional label or its verification, which may be set-up to other foodstuffs.

Keywords: nutritional label; pastry; biscuits; compositional data; monounsaturated fat; polyunsaturated fat; saturated fat.

Abbreviations: PLS, partial least squares; PUFA, polyunsaturated fatty acids; SFA, saturated fatty acids; TUFA, total unsaturated fatty acids; Vis/NIRS, visible and near infrared spectroscopy.

Introduction

Consumption of foods with a high glycemic index is hypothesized to contribute to insulin resistance, which is associated with increased risk of diabetes mellitus, obesity, cardiovascular disease, and some cancers (Neuhouser et al., 2006). Besides, there is an increasing prevalence of diabetes mellitus worldwide (Wild, Roglic, Green, Sicree, & King, 2004). Carbohydrates are the major part both in bakery, pastry and biscuit products, while total fat is variable. Their protein content is in general not relevant. The fat composition of the diet has also important nutritional implications (Schwingshackl & Hoffmann, 2014). The unfavorable effects of SFA on cardiovascular health and the positive effects of MUFA, are widely proven (Hernández et al., 2017; Schwingshackl & Hoffmann, 2014). Thus, some pastry producers have replaced SFA with MUFA, often by using high oleic sunflower.

Saturated fatty acids are those without any unsaturation within their chain. Pastry includes usually as major SFA palmitic acid (C16:0), stearic acid (C18:0), and small quantities of margaric acid (C17:0), arachidic acid (C20:0), and behenic acid (C22:0) (Yolci-Omeroglu & Ozdal, 2020). MUFA are those fatty acids which carbon chain has a single unsaturation. Palmitoleic acid (C16:1) and oleic acid (C18:1) are the major MUFA in pastry products, according to the last referred authors. Oleic acid unsaturation locates after the number 9 carbon, commonly named ω -9. PUFA are those fatty acids containing more than one double bond in their backbone. Good human health requires diets with small quantities of these compounds, such as the essential fatty acids linoleic (C18:2), ω -6, and linolenic (C18:3), this last called ω -3. Linoleic acid is often present in pastry products (Yolci-Omeroglu & Ozdal, 2020).

The standards of the European Union (CE, 2011), applicable since 13 December 2016, settles the duty to include nutritional information as part of the compulsory food information to the consumer. This includes energy value, total fat contents, saturated fatty acids (SFA), carbohydrates, sugars, proteins and salt. The rule considers as voluntary nutritional information other nutrients such as mono-unsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), trans-fatty acids, vitamins, healthy compounds, and other possible relevant compounds, as cholesterol. Therefore, pastry and biscuit products must include nutritional information on their label. The most common information

included up to date is total fat and saturated fat, although some products also provide information on monounsaturated and polyunsaturated fat.

Verifying the nutritional composition of different batches of pastry using fast and reliable techniques is a target both for the industry and the administration, responsible for controlling food quality. Among the various non-destructive techniques able to provide solutions to these needs, near infrared spectroscopy stands out for its important achievements so far. There is extensive literature on NIRS techniques for assessing food quality (Haiyan & Yong, 2007; Manley, 2014; Baeten et al., 2016; among others). Food authentication is also important aim of NIRS (v. g. Reid, O'Donnell, Downey, & 2006; Callao & Ruisanchez, 2018; Mendez, Mendoza, Cruz-Tirado, Quevedo, & Siche, 2019).

One of the first studies about near infrared spectroscopy (NIRS) on pastry was intended to determine fat, protein and water content (Kaffka, K. J., Norris, K. H., & Roszakiss, 1982). A later study revealed the possibility of measuring fiber content (Kays, Barton, Windham, & Himmelsbach, 1997; Kays, Windham, & Barton, 1998). Vines, Kays, & Koehler (2005) reported measurement of total fat by NIRS on cereal food. A study on rapid analysis of chemical composition in intact and milled rice biscuits is found in Wimonsiri et al. (2017). Despite the facts above, using NIRS for the specific purpose of determining the components of a food nutritional label has been reported in one only paper (Fernández-Cabanas, Polvillo, Rodríguez-Acuña, Botella, & Horcada, 2011). This was on the determination of the fatty acid profile in pork dry-cured sausages. Additionally, one more study refers to Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs (Pérez-Marín, Garrido-Varo, Guerrero-Ginel, & Gómez-Cabrera, 2004). These studies did not mention the need to adequately treat reference compositional data. In fact, the integration of specific methods for compositional data in NIRS techniques has been absent until not long ago. A novel technique of NIRS in combination with compositional statistical methods was introduced recently for nutritional labelling of olive oil (Cayuela-Sánchez, Palarea-Albaladejo, García-Martín, & Pérez-Camino, 2019).

Compositional data are multivariate data representing fractions or parts of some whole and, thus, carry only relative information. This implies that values in each part have meaning only in relation to the other parts and specialized statistical analysis methods accounting for these features are required. Nutritional compositions, consisting of fractions measured across mutually exclusive categories commonly scaled as percentages, fall into this definition. Amongst others, compositional data bring

some difficulties to the use of ordinary correlations and distances, since these ignore the relative scale and the constraints they are subject to. Well-known issues include negative bias in correlation analysis, singularity of covariance matrices in regression analysis, predictions beyond the range of possible values (e.g. values outwith the [0, 100] interval when working in percentage units) or results which depend on the scale of measurement. Obviously, these issues can potentially lead to misleading scientific conclusions. A solid methodological framework based on using log-ratios between parts of a composition was presented in the seminal work by Aitchison (1986) and has been further elaborated and successfully applied in different scientific fields over the past 30+ years. Working with log-ratios focuses the data analysis on the relative information, guarantees that results do not depend on the scale of measurement (i.e. on whether it is percentages, proportions, ppm, mg/g, etc.) and, conveniently, facilitates the use of ordinary statistical methods which results can be transferred back in terms of the original compositions if required (Aitchison, 1986; Van den Boogaart & Tolosana-Delgado, 2013; Pawlowsky-Glahn et al., 2015).

NIRS offers several important advantages, as it is a rapid, non-destructive and potentially multi-parametric method. In addition, NIRS does not need solvents or reagents, avoiding a significant expense and protecting the environment. In this study we present a new approach to determine the nutritional makeup of industrial pastries and biscuits for measuring their nutritional label components. It is based on fitting PLS calibration models on Vis/NIR spectra which explicitly account for the compositional nature of the reference data and avoids analytical artifacts.

2. Material and Methods

2.1. Pastry and Biscuit Samples

The robustness of NIRS calibrations depends on the statistic range of the analyzed features. Therefore, the choice of the commercial product types for which predictive models are implemented is a critical factor.

Two separate tests were conducted with commercial pastry and biscuit products. Two independent batches of samples were selected by the authors in different superstore markets of Seville (Spain) from different manufacturers for both types of products. The total number of samples of pastry and biscuit products were 24 and 14, respectively.

The pastries and biscuits selected for the study are shown in Table 1. As can be seen, pastry products integrated a wide diversity including croissants (plain and chocolate), milk breads and muffins. The information provided by the manufacturers indicated that some of them included butter, olive oil, honey, high oleic sunflower oil, chocolate filling or nuggets. On the other hand, the biscuits assay included a variety of types of biscuit products present in the market. These were mainly 'golden', 'high oleic' biscuits, and toasted crackers. A high oleic biscuit product, specially oriented to children, and a biscuit product without added sugars were both also included to investigate their fit in relation to the other sampled products.

Table 1

2.2. Spectra Acquisition

Replicates of the samples were used, as indicated below, and two spectra replicates were registered from each of them. The spectra were generated from a Labspec spectrometer (Analytical Spectral Devices Inc., Boulder) equipped with three detectors. The detector for the visible range (350-1000 nm) is a fixed reflective holographic diode array with a sensitivity of 512 pixels. A holographic fast scanner InGaAs detector cooled at -25 °C covers the wavelength range of 1000-1800 nm. This coupled with a high order blocking filter runs for the 1800-2500 nm interval. The instrument is equipped with internal shutter and automatic adjustment offset, the scanning speed is 100 ms. The repeatability of the instrument, expressed as standard deviation on the average absorbance of five measures of a white tile between 350 and 2500 nm, is $6.00 \cdot 10^{-4} \text{ cm}^{-1} \text{ mol}^{-1}$. Two alternative methods for spectra acquisition from intact (a) and ground samples (b) were evaluated: (a) spectra were registered by reflectance directly, without any treatment, from two or three food intact samples depending on the type of product; (b) the product sample units were mixed and ground into powder before registering spectra similarly to the intact samples. In the case of the pastry products with chocolate filling, the samples were ground and manually homogenized with a stainless steel spatula. Two scan replicates on each sample were registered for both methods.

Registration of spectra was performed using the 'Sample Turn Table' accessory of the spectrometer (Fig. 1) with standard SMA 905 optical fiber connectors. A spectrum average was acquired in reflectance mode from each intact unit, contained in a quartz Petri dish, whilst the 'turn table' was turning. The software was configured for continuous acquisition of 10 scans that were automatically

averaged to form a single spectrum from each sample. The acquisition was carried out using Indico Pro software (Analytical Spectral Devices Inc., Boulder). The resolution of the spectra corresponded to an interval of 1 nm. Registration took less than a minute for each replicate sample, all steps included.

Figure 1

2.3. Reference Analysis

All the reference analyses were common both for intact and ground pastry and biscuit products.

2.3.1. *Moisture Content*

After registering the spectra of the two replicate samples of the ground products, 30-40 g of them were exactly weighed and separately placed into an oven at 105 °C in ceramic capsules for 2.5 h drying time. The drying time was experimentally defined for pastry products as the time corresponding to a maximum weight loss. Humidity content was determined by gravimetric analysis and expressed as weight percentage of the fresh commercial product.

2.3.2. *Fat Content*

The oil contents from each replicate sample of pastry or biscuit products were extracted by the Soxhlet method, using hexane as solvent. For this purpose, paper cartridges were prepared containing the powder of the pastry and biscuit products. Then they were introduced into 250 cm³ Soxhlet extractor units. The cartridge diameter was 35 mm and its length was approximately 75 mm, for all the samples. After 5.5 h extraction, the hexane remaining in the fat was removed by vacuum extraction on a rotary evaporator, water bath at 40 °C and tap water circulating through the refrigerant circuit. The elimination of possible remaining traces of solvent was completed using a stream of nitrogen gas. Once at room temperature, the flasks containing each sample were weighed to gravimetrically determine their fat content. It was expressed as percentage regarding the fresh commercial product. Then each fat sample was disposed into 4 mL vials. They were hermetically closed and stored at 4 °C for subsequent fatty acids analysis. The fatty acids compositions were analyzed within a week after fat sample extraction.

2.3.3. *Fatty Acids Composition*

The fatty acids compositions were analyzed by gas chromatography (GC), as fatty acid methyl esters (FAME), according to the IUPAC Standard Method (IUPAC, 1987). Briefly, 100-130 mg of the fat of pastry samples were dissolved in 2 mL heptane and then the sample was transesterified using 500 μL 2 M methanolic potassium hydroxide solution. After decanting, the supernatant was collected. GC analysis was carried out using an Agilent 7697A gas chromatograph (Agilent Technologies, Santa Clara) equipped with a capillary column (poly (90% biscyanopropyl–10% cyanopropylphenyl) siloxane, 60 m \AA , 0.25 mm Φ , and 0.20 μm film thickness). Automatic split injection and a flame ionization detector (FID) were used. The carrier gas was hydrogen at a flow rate of 1 mL min^{-1} . The temperatures of the injector and detector were 225 and 250 $^{\circ}\text{C}$, respectively. The oven was programmed at a temperature of 180 $^{\circ}\text{C}$ (10 min), which was then increased 3 $^{\circ}\text{C min}^{-1}$ up to 220 $^{\circ}\text{C}$ (10 min). The injection volume was 1 μL . Fatty acid contents were expressed in percentages.

The relative abundances of the MUFA, PUFA, TUFA and SFA categories were expressed as percentages from the analyzed fatty acid values. Thus, MUFA was the sum of percentages of the palmitoleic (C16:1), heptadecenoic (C17:1), oleic (C18:1) and eicosenoic (C20:1) fatty acids. PUFA was the sum of percentages of the linoleic (C18:2) and linolenic (C18:3) fatty acids. TUFA was the sum of MUFA and PUFA. SFA was the sum of percentages of the caprylic (C8:0), capric or decanoic (C10:0), lauric (C12:0), miristic (C14:0), palmitic (C16:0), stearic (C18:0), araquidic (C20:0), behenic (C22:0), and lignoceric (C23:0) fatty acids. All the reference analysis was carried out at the Instituto de la Grasa (CSIC).

2.3.4. Protein

The protein content of the pastry products was determined by analyzing the nitrogen percentage of their dry powder. It was carried out by using an elemental analyzer LECO-932 CHNS (LECO Corp., St. Joseph, USA). Briefly, approximately 0.5 mg of the homogenized powder samples were exactly weighed into tin capsules, which were closed and placed into the analyzer holder, expressing the results in N percentages. The protein content of the pastry products was calculated from N by using a conversion factor 5.83, suitable for wheat based products. Note that protein content was not analyzed in biscuits.

2.4. Principal Component Analysis of Vis/NIR Data

The Vis/NIR spectral data from intact and ground products were processed in the same way for both industrial pastry and biscuit products. First, the spectrum replicates of each sample were averaged.

Then, the spectral variables were reduced at 8 nm intervals. The reflectance data of the whole spectra were mean normalized, transformed to absorbance and pre-treated by the Savitzky-Golay first derivative method, with polynomial order 2 and smoothing point 3. The suitability of these treatments has been previously reported for olive oil (Cayuela, García, Moreda, & Pérez, 2015) and were applied using The Unscrambler 9.7 (CAMO Software AS, Norway).

Principal component analysis (PCA) was applied to examine the overall consistency of the spectral data set. PCA computes new variables (the principal components; PCs) which are uncorrelated linear combinations of the spectral intensities across wavelengths. By construction, successive principal components summarize decreasing portions of the total variability in the original spectral variables. In particular, the two first principal components account for the highest percentage and are commonly used to represent multivariate data in a standard Cartesian coordinate system. Hence, for our purpose, PCA facilitated the visualization of the spectral information from pastry and biscuit samples in an ordinary scatter plot. A 95% confidence ellipse was calculated to identify potential outlier values. PCA was conducted on the R system for statistical computing (R Core Team, 2017).

2.5. PLS Calibration on Log-Ratio Coordinates

PLS modelling was based on log-ratio coordinates of the percentage nutritional composition of moisture, fat, protein and carbohydrates. The latter was determined as the remainder to 100% of the sum of the other three components. In particular, we employed an isometric log-ratio (ILR) representation (Egozcue et al., 2003) by which the information in the original 4-part composition was projected onto real space by way of three orthogonal real-valued coordinates ILR_1 , ILR_2 , and ILR_3 , according to the following equations:

$$ILR_1 = \sqrt{\frac{3}{4}} \ln \frac{\text{Moisture}}{(\text{Fat} \cdot \text{Protein} \cdot \text{Carbs})^{1/3}} \quad [1]$$

$$ILR_2 = \sqrt{\frac{2}{3}} \ln \frac{\text{Fat}}{(\text{Protein} \cdot \text{Carbs})^{1/2}} \quad [2]$$

$$ILR_3 = \sqrt{\frac{1}{2}} \ln \frac{\text{Protein}}{\text{Carbs}} \quad [3]$$

Moreover, fatty acids were amalgamated into the three fat categories (MUFA, PUFA, and SFA) and analogously represented by two ILR coordinates ILR_{FA1} and ILR_{FA2} according to the following equations:

$$ILR_{FA1} = \sqrt{\frac{2}{3}} \ln \frac{MUFA}{(PUFA \cdot SFA)^{1/2}} \quad [4]$$

$$ILR_{FA2} = \sqrt{\frac{1}{2}} \ln \frac{PUFA}{SFA} \quad [5]$$

Note that it is possible to define alternative ILR representations, but they all are orthogonal rotations of each other and would lead to the same results in terms of the original composition.

Multivariate PLS regression models were separately fitted by the kernel algorithm to the nutritional composition (for pastry products) and the fatty acids composition (for both pastry and biscuit products) using the respective ILR-coordinates as response and the Vis/NIR spectral variables as predictors (scaled by standard deviation). Note that for biscuits information about percentage total fat (TF) was treated as a separate 2-part composition (TF, 100 - TF). The corresponding ILR-coordinate, computed as $\sqrt{1/2} \ln(TF/(100 - TF))$, was included in the multivariate PLS regression model along with the ILR-coordinates from the biscuit fatty acids composition (Eqs. [4] and [5] above). The optimal number of PLS latent components was determined for each model by cross-validation aiming to reach the most parsimonious model among those reaching comparable highest performance in terms of coefficient of determination and root mean square error following the one-standard error rule (Kuhn & Johnson, 2013). Predictions obtained on ILR-coordinates were transformed back into the corresponding predicted percentage compositions by inverse ILR transformation. Results for total unsaturated fatty acids (TUFA) were calculated *a posteriori* by adding PUFA and MUFA. These models and the associated compositional operations were conducted on the R system for statistical computing (R Core Team, 2017).

2.6. Model Performance Assessment

Model performance was evaluated on ILR-coordinates based on the coefficient of determination (R^2), mean squared error of prediction (MSEP), and relative prediction deviation (RPD, ratio between standard deviation and cross-validated MSEP) using calibration (training) data and 5-time repeated 10-fold cross-validation (CV). The ability of the models to predict the nutritional and fat compositions

as a whole was assessed by computing compositional R^2 and metric standard deviation (MSD) as described in Cayuela-Sánchez, Palarea-Albaladejo, García-Martín, & Pérez-Camino (2019) based on training and cross-validation data. Finally, cross-validated correlations (r) between predicted (P) and observed (O) values and percentage relative deviations (RD), computed as $100|P-O|/O$ and averaged across samples, were calculated for each component.

3. Results

3.1. Pastry and Biscuit Products Spectra

Near-infrared spectra show various overlapping bands, because of their first and second overtones and a combination of fundamental vibrations, mainly carbon–hydrogen (Shenk, Workman, & Westerhaus, 2001). The large peak observed at 1900 nm corresponds to water. According to Salgó and Gergely (2012), carbohydrates peaks are detected in three wavelength regions: (i) between 1585 and 1595 nm, (ii) from 2270 to 2280 nm, and (iii) from 2325 to 2335 nm. This latter probably relates to the combination of the bond vibration of the C-H stretch and the C-H 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 stretching and C-C stretching vibrations (Osborne and Fearn, 1986). Protein absorption bands have been described in wheat at 2040-2160 nm and at 2160-2220 nm. Fat absorption bands in pastry products depend mainly on the type of plant oil, or animal fat, used in their manufacturing. The pastry 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 or butter. Main near-infrared absorption bands of plant oils have been described by Hourant, Baeten, Morales, Meurens, & Aparicio (2000). A broad absorbance band exists around 1220 nm, probably due to second overtones of C–H and CH=CH– stretching vibrations from oil. A high intensity absorbance peak occurs about 2300 nm, caused by a combination of fundamental vibrations from the C-H groups present in fatty acids (Hourant, Baeten, Morales, Meurens, & Aparicio, 2000). The bands cited above were present approximately at the same wavelengths in the spectra of pastry and biscuit products (see Fig. 2).

Figure 2

3.2. Summary Statistics

Univariate summary statistics of the nutritional and fat components of the pastry and biscuit samples in this study are shown in Table 2 (values in percentages). In accordance with the relative scale of the reference data, the geometric mean (including 95% confidence interval) and standard deviation were used as summary statistics of central position and dispersion respectively, along with ordinary 25%, 50% and 75% quartiles and range.

In the case of pastry samples, the percentages of moisture ranged from 11.7 to 23.9, fat between 12.2 and 46.6, protein from 5.8 to 13.3, and carbohydrates from 30.9 to 57.6. As to fat composition, SFA ranged from 12.2 to 61.4. MUFA between 27.1 and 69.1, PUFA from 9.9 to 54.4, and TUFA from 38.5 to 88.9. Average percentage content of total unsaturated fat was larger than saturated fat (60.3 vs. 31.5), with MUFA being the main contributor (35.84). The highest relative variation was shown by PUFA ($SD_g = 1.9$).

The biscuit products had an average percentage total fat content of 14.8, with SFA ranging from 8.1 to 46.1, MUFA from 40.7 and 82.1, PUFA from 9.6 to 18.0, and TUFA from 53.9 to 91.9. As with pastries, unsaturated fats were more abundant than saturated fats in biscuits (73.5 vs. 19.1), with MUFA content standing out (60.1). The highest relative variation was found in SFA ($SD_g = 2.1$).

Table 2

3.3. Principal Component Analysis

PCA of the spectral data was conducted for intact and ground pastry samples. For intact samples, an atypical case falling beyond the 95% confidence ellipse limits was identified. It corresponded with one replicate of a commercial product of muffins based on olive oil. Moreover, four outliers were identified amongst ground pastry samples. Two of them were the two replicates of a type of cake type using butter. Another one was a replicate of a chocolate cake, and the last one was a replicate of a chocolate croissant. Interestingly, the outlying samples did not coincide between intact and ground samples. The variability explained by the first two PCA dimensions, after discarding the outliers, was moderate for both intact and ground products (45.7% and 62.8% respectively). The resulting PCA plot for ground pastry samples is shown in Fig. 3a, where no particular patterns, apart from the expected higher proximity, i.e. similarity, between pairs of replicates from the same sample, or evidence of additional atypical samples are observed.

Similarly, the PCA of the spectral data of the biscuits products provided different outcomes from intact and ground samples. Outliers were not detected for either intact or ground biscuits. The variance explained by the first two principal components was 69.9% and 66.6% for intact and ground biscuits respectively. The PCA plot for intact biscuits is shown in Fig. 3b, and a consistent structure is also observed here.

Figure 3

3.4. PLS Models on Log-Ratio Coordinates

The 4-part composition formed by moisture, fat, protein and carbohydrates led to a set of three ilr-coordinates as defined in formulas [1], [2] and [3] for industrial pastry products, intact and ground. Table 3 summarizes the performance statistics of the multivariate PLS regression fitted to them. Although intact samples provided somewhat better results based on training data, the cross-validated statistics resulted fairly comparable between both intact and ground samples. The best fit was obtained for the coordinate ILR_2 , which represents a contrast between fat, protein and carbohydrates contents.

Table 3

Observed versus predicted values plots (by ILR-coordinate and based on cross-validation) from ground pastry samples are shown in Fig. 4. The results for the third ILR-coordinate [3] reflect clearly a partition in two subgroups, due to the balance between protein and carbohydrates content. This partition corresponds to different types of pastry products. Analogous results were obtained from intact pastry products.

Figure 4

Table 4 shows performance statistics of the compositional PLS models for the fat composition based on the ILR-coordinates [4] and [5] set up from the 3-part composition formed by MUFA, PUFA, and SFA, for intact and ground pastry and biscuits products. For pastry products, ground samples provided better overall results than intact samples. This may be due to the homogenization effect from grinding, since some of the products had fillings and fat was an important part. The models performed worse for the ilr-coordinate ILR_{FA1} , which represented the balance between MUFA and the other two types of fatty acids (Eq. [4]).

Table 4

Results for biscuits show overall better performance statistics from intact samples, except for total fat percentage. Figure 5 shows observed versus predicted values plots (by ILR-coordinate and based on cross-validation) for intact biscuits.

Figure 5

3.5. Overall Model Performance

3.5.1. Nutritional Compositions

Predictions from the fitted PLS models on ILR-coordinates were conveniently transformed back to be expressed in terms of the original 4-part nutritional composition. Compositional measures of the overall performance of the models to predict the nutritional compositions were computed and are summarized in Table 5. The cross-validated R^2_{cv} were 0.81 and 0.86 respectively for intact and ground samples which, along with a slightly lower MSD, indicates better expected performance for ground samples. This agrees with what was observed in the results based on the ILR-coordinates. Cross-validated correlations between predicted and observed values (r_{cv}) were slightly higher (except for carbohydrates) for ground samples, and percentage relative deviations (RD_{cv}) were slightly lower for ground samples as well. These results are logical, since the reflectance spectrum of the ground samples better integrates the physicochemical features of the entire product.

Table 5

3.5.2. Fat Compositions

As shown in Table 6, compositional R^2 coefficients from fat composition PLS model fitting for intact and ground pastry products were 0.93 and 0.96 respectively based on training data (decreasing to 0.70 and 0.82 when based in cross-validation). The training-based MSD obtained were 0.17 and 0.13 respectively (0.28 and 0.22 from cross-validation). The prediction of the fat composition of pastry products was overall satisfactory, although best results were obtained from ground samples in agreement with the results shown above for the nutritional composition.

The bottom section of Table 6 shows the same performance measures for biscuits samples. In this case, ground biscuits gave somewhat worse model performance results, which may be related to variability on the apparent density of the different products. Note that correlations (r) for SFA and

TUFA are exactly the same. This is simply due to the fact that TUFA is computed as MUFA + PUFA and, given the compositional structure of the data being parts of 100%, this implies that SFA is equal to $100 - \text{TUFA}$ and the correlations then coincide. The results for intact biscuits are illustrated in Fig. 6, showing predicted (triangles) and observed (circles) values of the sampled fat compositions on a ternary diagram.

Table 6.

Figure 6

4. Conclusions

This study introduces a methodological basis to measure the nutritional composition of industrial pastries and biscuits for labelling purposes, using Vis-NIRS and taking into account the compositional nature of the reference data. The data modelling was conducted so that it took into account the intrinsic relative and inter-dependent nature of percentage nutritional compositions and this was incorporated within the well-known framework of PLS calibration. In spite of the limited number of samples available to train the models in this initial investigation, the overall performance of the proposed procedure using compositional Vis-NIRS models was satisfactory for estimating the percentage composition of moisture, total fat, protein and carbohydrates and of fat types (SFA, MUFA and PUFA). This spectroscopy method could be used without requiring reference analysis. The results obtained are promising and support the idea that the method proposed is an advantageous alternative for the nutritional labeling or its verification. In conclusion, this study proposes a useful approach which may be extended to other foodstuffs. This method would require a periodic validation protocol as usual in Vis/NIR spectroscopy techniques.

Author Statement

The first author of this study states the individual author contributions was as follows:

The corresponding author ensures that the descriptions are accurate and agreed by all authors.

The roles of the authors in this study were as follows:

Conceptualization	Ideas, conceptualization and experimental design was by J. A. Cayuela-Sánchez
Methodology	Development and design of methodology concerning compositional data as well as models development was by J. Palarea-Albaladejo. Methodology concerning reference analysis was provided by J. A. Cayuela-Sánchez
Validation	Model validation was carried out by J. Palarea-Albaladejo
Formal analysis	Application of statistical, mathematical, and software conducted techniques were carried out by J. Palarea-Albaladejo and J. A. Cayuela-Sánchez
Investigation	Conducting the research process, specifically performing the experiments, were carried out by T. P. Zira and E. Moriana-Correro
Writing - Original Draft	Writing the original manuscript was performed by J. A. Cayuela-Sánchez
Writing - Review & Editing	Review and editing were carried out by J. Palarea-Albaladejo and J. A. Cayuela-Sánchez
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team was by J. A.
Project administration	Management and coordination responsibility for the research activity planning and execution was by J. A. Cayuela-Sánchez

Conflict of interests

The authors declare no competing interests.

Acknowledgements

This work has been done without specific funding. J. A. Cayuela-Sánchez thank the Spanish Council for Scientific Research for supporting his activity as staff. 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].

References

- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London, UK.
- Baeten, V., Fernández-Pierna, J. A., Lecler, B., Abbas, O.; Vincke, D., Minet, O., Vermeulen, P., & Dardenne, P. (2016). Near infrared spectroscopy for food and feed: a mature technique. *NIR news*, 27, 4-6.
- Callao, P. M., & Ruisanchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control* 86, 283-293.
- Cayuela-Sánchez, J. A., Palarea-Albaladejo, J., García-Martín, J. F., & Pérez-Camino, M. C. (2019). Olive oil nutritional labeling by using Vis/NIR spectroscopy and compositional statistical methods. *Innovative Food Science and Emerging Technologies* 51, 139–147.
- 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 consumers. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011R1169&from=EN>. Last accessed June 2017.
- Chung, H., & Arnold, M.A. (2000). Near-infrared spectroscopy for monitoring starch hydrolysis. *Applied Spectroscopy* 54, 277-283.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barceló-Vidal, C. (2003). Isometric Log-ratio Transformations for Compositional Data Analysis. *Mathematical Geology*, 35(3), 279-300.

Fernández-Cabanas, V. M., 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 Chemistry*, *124*, 373-378.

Hernández, A., Castañer, O., Goday, A., Ros, E., Pintó, X., Estruch, R., Salas-Salvadó, J., Corella, D., Arós, F., Serra-Majem, L., Martínez-González, M. A., Fiol, M., Lapetra, J., De la Torre, R., López-Sabater, M.C., & Fitó, M. (2017). The Mediterranean Diet decreases LDL atherogenicity in high cardiovascular risk individuals: a randomized controlled trial. *Molecular Nutrition & Food Research*, *61*(9), 1601015, 9 pp. <http://dx.doi.org/10.1002/mnfr.201601015>

Haiyan, C., Yong, H. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology* *18*(2),72-83.

Hourant, P., Baeten, V., Morales, M. T., Meurens, M., & Aparicio, R. (2000). Oil and fat classification by selected bands of near-infrared spectroscopy. *Applied Spectroscopy*, *54*, 1168–1174.

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, Great Britain.

Kaffka, K. J., Norris, K. H., & Roszakiss, M. (1982). Determining fat, protein and water content of pastry products by the NIR technique. *Acta Alimentaria* *11*(2), 199-217.

Kays, S. E., Barton, F. E. Windham, W. R., & Himmelsbach, D. S. (1997). Prediction of total dietary fiber by near-infrared reflectance spectroscopy in cereal products containing high sugar and crystalline sugar. *Journal of Agricultural and Food Chemistry* *45*, 3944-3951.

Kays, S. E., Windham, W. R., & Barton, F. E. (1998). Prediction of total dietary fiber by near-infrared reflectance spectroscopy in high-fat- and high-sugar-containing cereal products. *Journal of Agricultural and Food Chemistry* *46*, 854-861.

Kuhn, M., & Johnson, K. (2013). Applied predictive modelling. New York: Springer <http://dx.doi.org/10.1007/978-1-4614-6849-3>.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews* *43*(24), 8200-8214.

Mendez, J., Mendoza, L., Cruz-Tirado, J. P., Quevedo, R., Siche, R. (2019). Trends in application of NIR and hyperspectral imaging for food authentication. *Sientia Agropecuaria* *10*(1), 143-161.

- Neuhouser, M. L., Tinker, L. F., Thomson, C., Caan, B., Van Horn, L., Snetselaar, L., Parker, L. M., Patterson, R. E., Robinson-O'Brien, R., Beresford, S. A. A., Shikany, J. M. (2006). Development of a glycemic index database for food frequency questionnaires used in epidemiologic studies. *Journal of Nutrition* 136(6), 1604-1609.
- Osborne, B.G., & Fearn, T. (1986). Theory of near infrared spectrophotometry. In: Near Infrared Spectroscopy in Food Analysis, first ed. Longman Scientific & Technical, Harlow, pp. 20-42.
- Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*. Wiley & Sons, Chichester, UK. <https://doi.org/10.1002/9781119003144.ch1>
- Pérez-Marín, D. C., Garrido-Varo, A., Guerrero-Ginel, J. E., Gómez-Cabrera, A. (2004). Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration. *Animal Feed Science and Technology*, 116, 333-349.
- R Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>
- Reid, L. M.; O'Donnell, C. P., Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science and Technology* 17(7), 344-353.
- Salgó, A., & Gergely, S. (2012). Analysis of wheat grain development using NIR spectroscopy. *Journal of Cereal Science* 56, 31-38.
- Schwingshackl, L., & Hoffmann, G. (2014). Monounsaturated fatty acids, olive oil and health status: a systematic review and meta-analysis of cohort studies. *Lipids in Health and Disease*, 13, 154. <https://doi.org/10.1186/1476-511X-13-154>
- Shenk, J. S., Workman, J. J., & Westerhaus, M. O. (2001). Application of NIR spectroscopy to agricultural products. In: D. A. Burns, and C. W. Ciurcak (Eds.), *Handbook of Near Infrared Analysis*, 2nd Edition (pp. 419–474). New York: Marcel Dekker.
- Van den Boogaart, K. G., & Tolosana-Delgado, R. (2013). *Analyzing compositional data with R*. Heidelberg, Germany: Springer <http://dx.doi.org/10.1007/978-3-642-36809-7>.
- Vines, L. L., Kays, S. E., & Koehler, P. E. (2005). Near-infrared reflectance model for the rapid prediction of total fat in cereal foods. *Journal of Agricultural and Food Chemistry* 53, 1550-1555.

Wild, S., Roglic, G., Green, A., Sicree, R., King, H. (2004). Global prevalence of diabetes - Estimates for the year 2000 and projections for 2030. *Diabetes Care* 27(5), 1047-1053.

Wimonsiri, L., Ritthiruangdej, P., Kasemsumran, S., Therdthai, N., Chanput, W., & Ozaki, Y. (2017). Rapid analysis of chemical composition in intact and milled rice biscuits using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 25(5), 330-337.

Yolci-Omeroglu, P. Y., Ozdalb, T. (2020). Fatty acid composition of sweet bakery goods and chocolate products and evaluation of overall nutritional quality in relation to the food label information. *Journal of Food Composition and Analysis* 88, 103438.

Journal Pre-proof

Figure captions



Figure 1. Labspec Vis/NIR spectrometer and sample turn table accessory.

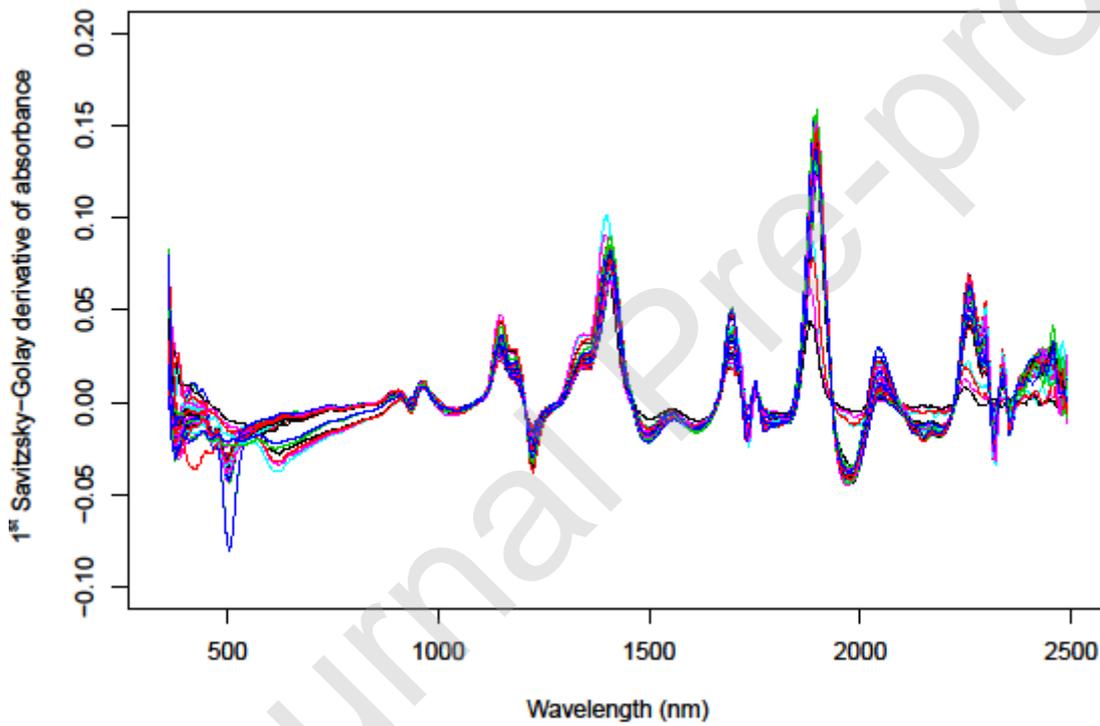


Figure 2. Vis-NIR spectra of pastry and biscuit products.

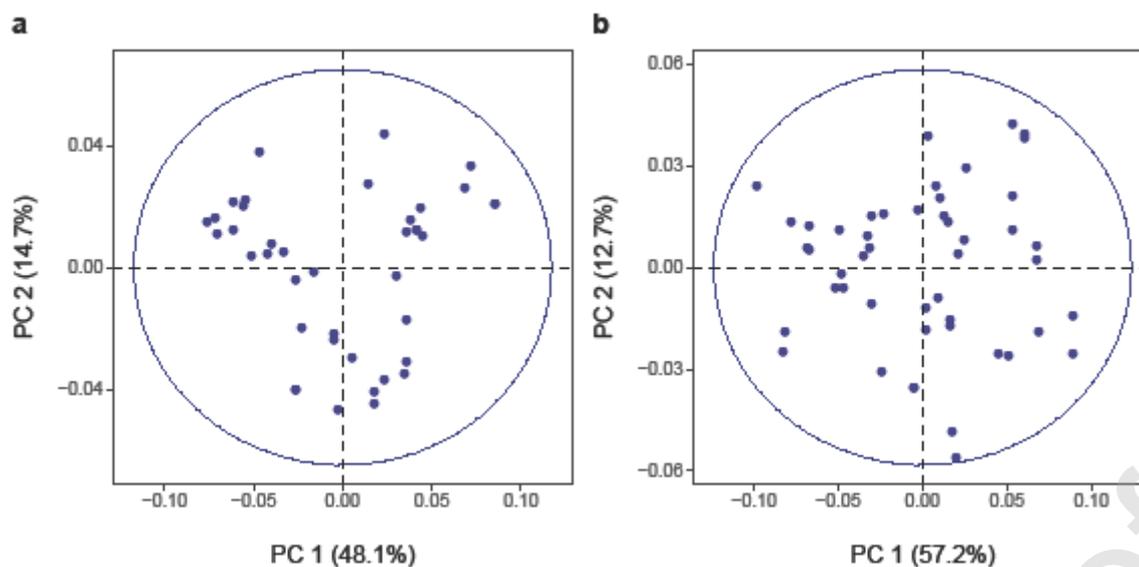


Figure 3. PCA including 95% confidence ellipse (first PC on the horizontal axis and second PC on the vertical axis). a) ground pastry products b) biscuit products.

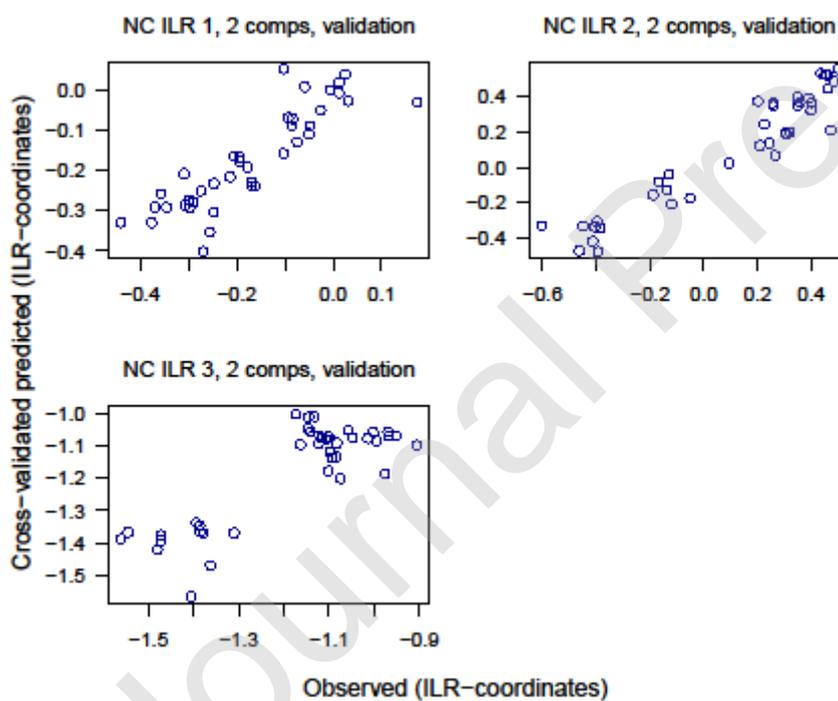


Figure 4. Nutritional composition of ground pastry products: observed versus predicted values plots from compositional PLS regression model fit on Vis/NIR spectral data.

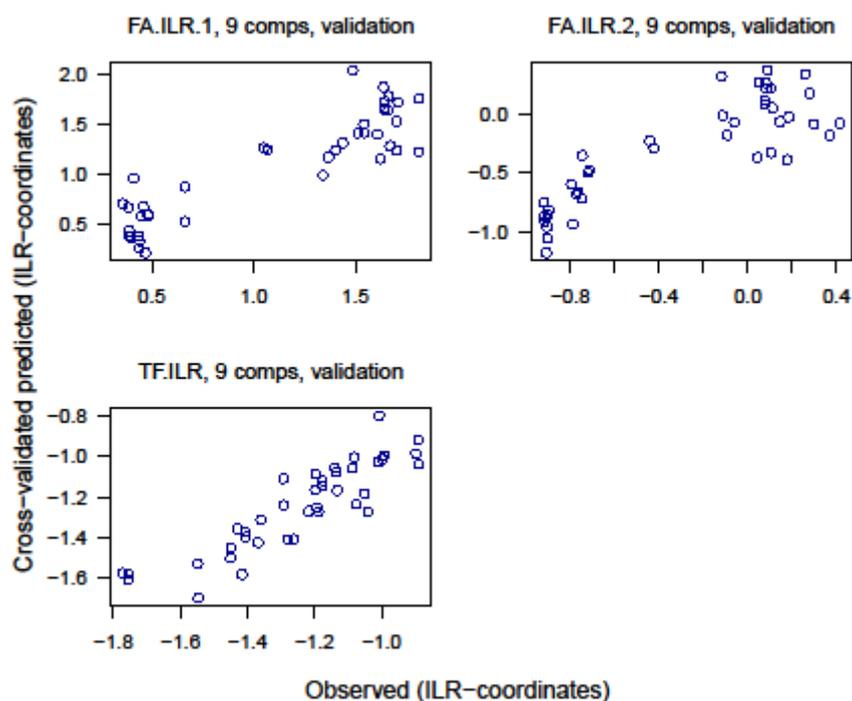


Figure 5. Fat composition of intact biscuits: observed versus predicted values plots from compositional PLS regression model fit on Vis/NIR spectral data.

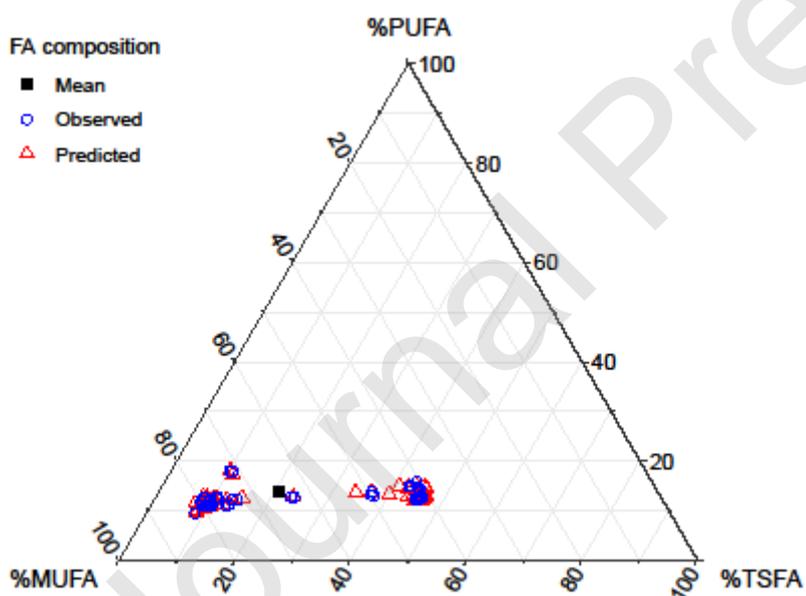


Figure 6. Ternary plot of predicted and observed percentage fat compositions for intact biscuits from fitted compositional PLS model.

Table 1. Samples of commercial pastry and biscuit products used for calibrations and validations.

(All components in %. Energy expressed in kj.)

Name	Type	Other	Energy	Fat	SFA	Carbs	Sugars	Fiber	Protein
Carrefour	Croissant		1974	28	14	47	12	1.3	6.9
La Bella Easo 1	Croissant		1787	24	12	44	13	2.7	7.6
La Bella Easo 2	Croissant	0% sugar	1774	27	13	40	<0.5	3.5	7.5
Bimbo Sonrisas	Croissant		1747	28	14	47	12	1.3	6.9
Casado Croissants	Croissant		1682	17	9	54	15	1.8	7.1
Martínez Croissant	Croissant	Chocolate	1917	26	13	49	23	2.9	5.3
Mildred Pasteleria	Croissant		1778	22	12	48	8	2.9	7.6
Brioche Pasquier	Croissant		1724	21	12	47	10	3.6	7.0
Alcampo	Croissant		1959	28	14	47	12	1.1	6.8
Auchan 1	Croissant		1946	28	14	46	12	2	6.6
Auchan 2	Croissant		2006	28	14	49	12	2.3	7.0
Auchan 3	Milk bread		1458	12	2	49	11	2	9.7
Dulcesol	Milk bread		1458	12	2	49	11	2	9.7
Carrefour	Milk bread		1530	13	6	53	11	1.5	8
Brioche Pasquier	Milk bread		1412	10	2	51	12	1.9	9.4
Martínez	'Sobaos'	Butter	1942	25	11	52	30	1.8	7.0
Serafina	'Sobaos'		1885	24	6	53	27	-	5.4
Dulcesol	'Valencianas'		1914	25	3	52	29	1.6	5.4
Auchan 4	Muffins	Olive oil	1806	22	2	53	28	1.1	4.3
Auchan 5	Muffins		1921	25	3	54	21	0.8	5.1
La Bella Easo 3	Muffins		1835	23	4	52	27	1.7	5.1
Auchan 6	Muffins	Chocolate chips	1920	26	5	50	26	2.7	5.1
La Boulangerie	'Napolitana'	Chocolate	1739	22	11	46	12	2.6	6.6
Martinez	Apple cake		-	21	11	52	-	3.1	3.4
Auchan 7	Biscuits		1926	15	2	74	20	2	5.7
Cuétera Maria Oro	Biscuits		2021	20	4	68	21	-	6.4
Fontaneda	Biscuits	Honey	1911	15	2	72	22	2.2	6.7
Tejedor Mini	Biscuits	High oleic	1776	10	1	75	17	1.5	7.0
Auchan 8	Biscuits	Toasted	1826	10	1	78	21	2.7	6.4
Gullón Dorada	Biscuits	High Oleic	1926	15	2	74	20	-	5.7
Flora	Biscuits	Sunfl. oil	1828	12	1	74	21	3.4	6.2
Auchan 9	Biscuits	0% sugar	1713	11	1	75	<0.5	4.5	7.0
Carrefour 1	Biscuits		468	17	1.7	72	21	1.8	5.8
Gullón	Biscuits		480	19	9.2	70	19	1.8	6.3
Carrefour 2	Biscuits		459	15	1.4	74	24	2.4	5.7
Carrefour 3	Biscuits		480	19	9.2	70	19	1.8	6.3
Carrefour	Biscuits	Toasted	419	8	3.9	78	22	2.7	7
Marbú	Biscuits		473	17	6.3	73	19	-	6

Table 2. Summary statistics of sampled pastry and biscuit products.

Pastries						
	Range	25% Q	50% Q	75% Q	Mean _g [95% CI]	SD _g
Moisture	11.75-23.93	15.16	17.81	19.93	17.29 [16.45, 18.18]	1.18
Protein	5.80-13.29	7.08	9.18	10.30	8.86 [8.28, 9.49]	1.26
Carbohydrates	30.85-57.61	44.44	48.41	51.53	47.48 [45.90, 49.11]	1.12
Fat	12.25-46.60	19.47	27.17	29.78	24.55 [22.44, 26.87]	1.35
SFA	11.12-61.46	15.58	46.78	52.80	31.55 [26.22, 37.96]	1.86
MUFA	27.09-69.15	31.92	34.52	37.41	35.84 [33.88, 37.91]	1.21
PUFA	9.92-54.44	11.98	13.99	48.36	21.02 [17.25, 25.61]	1.95
TUFA	38.54-88.88	47.20	53.22	84.42	60.34 [55.41, 65.71]	1.33
Biscuits						
	Range	25% Q	50% Q	75% Q	Mean _g [95% CI]	SD _g
Fat	7.55-22.14	12.04	15.54	17.99	14.77 [13.66, 15.97]	1.31
SFA	8.14-46.07	9.88	13.30	44.15	19.11 [15.44, 23.64]	2.08
MUFA	40.69-82.11	41.95	72.74	78.88	60.10 [55.12, 65.53]	1.35
PUFA	9.60-18.03	11.24	12.59	12.92	12.61 [12.14, 13.10]	1.14
TUFA	53.93-91.86	55.85	86.70	90.12	73.48 [68.69, 78.60]	1.26

SFA, saturated fatty acids; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; TUFA, total unsaturated fatty acids; Q, quartile; Mean_g, geometric mean; CI, confidence interval; SD_g, geometric standard deviation.

Table 3. Performance statistics of compositional PLS regression models by ILR-coordinates of pastry nutritional composition.

	Intact pastry products			Ground pastry products		
	ILR ₁	ILR ₂	ILR ₃	ILR ₁	ILR ₂	ILR ₃
				<i>R</i> ²		
Train	0.859	0.951	0.947	0.803	0.923	0.765
CV	0.750	0.872	0.676	0.758	0.906	0.708
				<i>MSEP</i>		
Train	0.059	0.076	0.039	0.060	0.093	0.087
CV	0.078	0.123	0.098	0.067	0.102	0.096
				<i>RPD</i>		
	2.023	2.827	1.777	2.058	3.310	1.875

CV, cross validation; ILR, isometric logarithmic ratio; *R*², coefficient of determination; *MSEP*, mean squared error of prediction; *RPD*, relative prediction deviation.

Table 4. Performance statistics of compositional PLS regression models by ILR-coordinates of pastry and biscuit fat compositions.

	Intact pastry			Ground pastry		
	ILR _{FA1}	ILR _{FA2}		ILR _{FA1}	ILR _{FA2}	
			<i>R</i> ²			
Train	0.382	0.950		0.450	0.978	
CV	0.184	0.857		0.319	0.920	
			<i>MSEP</i>			
Train	0.137	0.194		0.114	0.136	
CV	0.158	0.329		0.127	0.262	
			<i>RPD</i>			
	1.120	2.673		1.228	3.580	
	Intact biscuits			Ground biscuits		
	ILR _{FA1}	ILR _{FA2}	ILR _{TF}	ILR _{FA1}	ILR _{FA2}	ILR _{TF}
			<i>R</i> ²			
Train	0.980	0.943	0.919	0.942	0.896	0.917
CV	0.803	0.728	0.784	0.754	0.712	0.842
			<i>MSEP</i>			
Train	0.078	0.112	0.065	0.135	0.152	0.066
CV	0.249	0.246	0.106	0.278	0.254	0.091
			<i>RPD</i>			
	2.280	1.943	2.181	2.043	1.886	2.551

CV, cross validation; ILR_{FA}, isometric logarithmic ratio of fatty acids; ILR_{TF}, isometric logarithmic ratio of total fat; *R*², coefficient of determination; *MSEP*, mean squared error of prediction; *RPD*, relative prediction deviation.

Table 5. Compositional performance statistics of PLS regression models for pastry nutritional compositions based on calibration data and cross-validation (CV).

	Intact pastry products				Ground pastry products			
R^2	0.937				0.878			
R^2_{cv}	0.811				0.861			
MSD	0.061				0.082			
MSD _{cv}	0.104				0.089			
	Moisture	Fat	Protein	Carbs	Moisture	Fat	Protein	Carbs
r_{cv}	0.871	0.929	0.871	0.874	0.909	0.954	0.900	0.842
RD _{cv} (%)	7.143	8.840	9.286	4.726	5.579	7.127	8.750	4.273

R^2 , coefficient of determination; R^2_{cv} , squared coefficient of cross validation; MSD, metric standard deviation; MSD_{cv}, cross validated metric standard deviation; r_{cv} , cross-validated correlation; RD_{cv}, cross validated relative deviation.

Table 6. Compositional performance statistics of PLS regression models for pastry and biscuit fat compositions based on calibration data and cross-validation (CV).

	Intact pastry				Ground pastry			
R^2	0.928				0.964			
R^2_{CV}	0.705				0.822			
MSD	0.170				0.127			
MSD_{CV}	0.281				0.217			
	SFA	MUFA	PUFA	TUFA	SFA	MUFA	PUFA	TUFA
r_{CV}	0.895	0.209	0.909	0.895	0.941	0.390	0.959	0.941
RD_{CV} (%)	20.238	12.833	24.325	10.928	18.466	11.358	14.783	7.946
	Intact biscuits				Ground biscuits			
R^2	0.965				0.924			
R^2_{CV}	0.702				0.611			
MSD	0.099				0.146			
MSD_{CV}	0.226				0.268			
	SFA	MUFA	PUFA	TUFA	SFA	MUFA	PUFA	TUFA
r_{CV}	0.929	0.931	0.757	0.929	0.897	0.902	0.693	0.897
RD_{CV} (%)	27.592	9.437	7.918	7.479	31.949	12.892	9.413	10.297

R^2 , coefficient of determination; R^2_{CV} , squared coefficient of cross validation; MSD, metric standard deviation; MSD_{CV} , cross validated metric standard deviation; r_{CV} , cross-validated correlation; RD_{CV} , cross validated relative deviation; SFA, saturated fatty acids; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; TUFA, total unsaturated fatty acids.