Determining mandatory nutritional parameters for Iberian meat products using a new method based on near infra-red reflectance spectroscopy and data mining

Daniel Caballero[1,2,3], Maria Asensio[1], Carlos Fernández[1], Noelia Martín[1], and Antonio Silva[1]

A b s t r a c t: *The new regulation about mandatory labelling on nutrition requires the declaration of specific parameters: protein, lipid, salt and carbohydrate contents. This study reports a fast, accurate method to determine the values of these mandatory nutritional parameters based on near infra-red reflectance spectroscopy (NIRs) technology and data mining techniques, used in an automatic way. For that, two batches of different Iberian pork meat products (dry-cured ham, dry-cured loin, dry-cured shoulder, dry-fermented Salchichón sausage, and dry-fermented Chorizo sausage were used. One batch of each product was used to train the method and the remaining batch was used for validation. To develop the method, prediction equations were obtained from the NIRs, while nutritional data for the training batches were obtained by applying data mining techniques, and the prediction equations were evaluated against the NIRs data from the validation batch. The prediction equations achieved from very good to excellent degrees of relationship (R > 0.75) and accurate results (MAE < 1, RMSEC < 1, RMSEP < 1) from the training batch. These prediction equations were corroborated using the validation batch, which showed very good to excellent correlation coefficients (R > 0.75). This new method is rapid, as it takes around 10 minutes in comparison with traditional methods that take around 6 days.*
 Keywords: pork, NIRs, data mining, labelling, Iberian meat products, nutritional information.

Introduction

In the European Union (EU), mandatory labelling of nutritional information requires specific parameters (protein, lipid, saturated lipid, salt, carbohydrate and sugar contents) to be declared (*European Union*, 2011). This regulation aims to protect consumers' health by informing them of the food's nutritional content. Thus, it moderates the provision of useful, understandable and uniform information to consumers, allowing them to make coherent decisions and safe food choices (*Benoit et al.*, 2016).

In this regulation, the detection limits for the main nutritional parameters are 0.1 %, but if the value detected is less than this percentage, the amount can be declared as not detectable (*European Union*, 2011). In addition, this regulation aims that allergens or substances causing intolerances be declared, updating the guidance document under regulation 2000/13/EC (*European Union*, 2000). This regulation was active until 2014.

Consumers rate meat and meat products derived from Iberian pigs highly because of the products' unique sensory traits, which are a consequence of both the raw materials' characteristics, especially lipid-related ones, and their particular processing conditions. Among these products, the most appreciated are the expensive, dry-cured Iberian hams. These are derived from Iberian pigs reared outdoors and fed on natural resources, mainly acorn and grass (*Cava et al.*, 2000). The main nutritional parameters of protein, lipid, salt and carbohydrate contents of Iberian pork meat products derived from Iberian pigs were widely studied previously (*Cruz and Vieira*, 2017; *Muriel et al.*, 2004; *Utrilla et al.*, 2010; *Ventanas*, 2012).

Traditional physico-chemical analysis methods for nutritional information are tedious, time- and solvent-consuming and require the destruction of the samples. These analyses take around 6 days for results to be produced, delaying the response time. In this sense, the use of alternative techniques, such

[1]*Animal Source Foodstuffs Innovation Service (SiPA), University of Extremadura, Av. Ciencias S/N, ES- 10003, Cáceres, Spain.*
[2]*Media Engineering Group, School of Technology, University of Extremadura, Av. Ciencias S/N, ES-10003, Cáceres, Spain.*
[3]*Chemometrics and Analytical Technology, Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 26, DK-1958, Frederiksberg C, Denmark.*

Corresponding author: Daniel Caballero, *dcaballero@unex.es*

as computerized tomography (CT), magnetic resonance imaging (MRI) and near infra-red reflectance spectroscopy (NIRs) were proposed for determining some nutritional parameters in these products.

CT was applied by some authors to characterize some nutritional parameters of meat products (*Fulladosa et al.*, 2010; *Picouet et al.,* 2013). MRI was proposed to monitor the ripening process of hams (*Caballero et al.*, 2016) and to determine some quality parameters of Iberian dry-cured loins (Ávila et al., 2018; *Caballero et al.*, 2017a; *Caballero et al.*, 2017b; *Pérez-Palacios et al.*, 2017).

Collell *et al.* (2011) applied NIRs to predict data on moisture, water activity and NaCl content at the surface of dry-cured ham during processing. The fatty acid compositions of dry-cured ham subcutaneous fat were predicted by NIRs (*Pérez-Juan et al.*, 2010). In lamb or rabbit meat products (*Cifuni* et al., 2016; *Cozzolino et al.*, 2000), NIRs was applied in order to determine some nutritional parameters. Moreover, in other meat products, NIRs was applied to obtain some physico-chemical parameters (*González-Mohino et al.*, 2018; *Pérez-Palacios et al.*, 2019; *Zamora-Rojas et al.*, 2011).

Data mining is an important part of a larger process known as Knowledge Discovery in Databases (KDD) (*Fayyad et al.*, 1996). The main goal of data mining consists of extracting hidden information from a data set. This can be achieved by the automatic analysis of large amounts of data, which allows the extraction of interesting and previously unknown patterns (*Sayad,* 2011). The development of robust and efficient algorithms to process data and the increase in computing power has enabled the use of intensive computational methods for data analysis (*Hastie et al.,* 2001). The development and interest in data mining have recently grown because of the rapidly decreasing cost of large storage devices and increasing ease of data collection over networks (*Mitchell,* 1999).

There are some examples of data mining techniques applied to determine quality traits of different meats, like beef (*Song et al.*, 2002) or lamb (*Cortez et al.*, 2006). In the case of pork, several examples of application of data mining techniques have been published (*Caballero et al.*, 2016; *Caballero et al.*, 2017a; *Caballero et al.*, 2017b; *Caballero et al.*, 2019; *Pérez-Palacios et al.*, 2014; *Pérez-Palacios et al.*, 2017; *Silva et al.*, 2013).

However, currently, there is not a fast and specified method to obtain the required nutritional information about Iberian pork products according to the EU regulation. Judprasong *et al.* (2013) studied the performance of laboratories in Thailand that reported nutritional analyses according to ISO 13528 regulation. The analytical method proposed in the current paper is faster than the proposed method by Judprasong *et al.* (2013) and is faster than the official methods proposed by AOAC (2000). In fact, some industrial meat companies have applied this novel NIRs method of analysis successfully on their dry-cured meat products, and have confirmed it takes less time and money to obtain the nutritional parameters.

Therefore, the main objective of this study was to report a fast and accurate method to determine the values of the main nutritional parameters mandatory on the labelling of Iberian dry-cured and dry-fermented meat products based on NIRs technology and data mining techniques used in an automatic way.

Materials and Methods

Experimental design

Two batches of some Iberian meat products (dry-fermented Chorizo sausage, dry-cured ham, dry-cured loin, dry-fermented Salchichón sausage and dry-cured shoulders) were used.

The first batch was composed of 45 samples of each Iberian meat product for training the new method, while the second batch was composed of 50 samples of each product used to validate the new method. All these Iberian meat products were acquired in different supermarkets to maximise the sample variability for each product.

Firstly, the samples from the first training batch were analysed in two ways: i) by physico-chemical analyses to obtain the content of the main mandatory nutritional parameters and ii) by using a NIRs (FOSS FOODSCAN lab, FOSS analytics, Hillerod, Denmark) to acquire NIRs spectra. All these data were gathered in a database. In this database, Multiple Linear Regression (MLR) was applied to obtain prediction equations for the mandatory nutritional parameters for each Iberian dry-cured meat product as a function of the bandwidth values. Then, the samples from second, validation batch were analysed in the same two ways as were used for the training batch. Thus, the prediction equations obtained in the first, training batch were applied using the NIRs data obtained from second, validation batch, and for validating the prediction equations, the calculated values were evaluated with the data from physico-chemical analyses from samples comprising the second, validation batch.

Figure 1 shows the experimental design followed in this study.
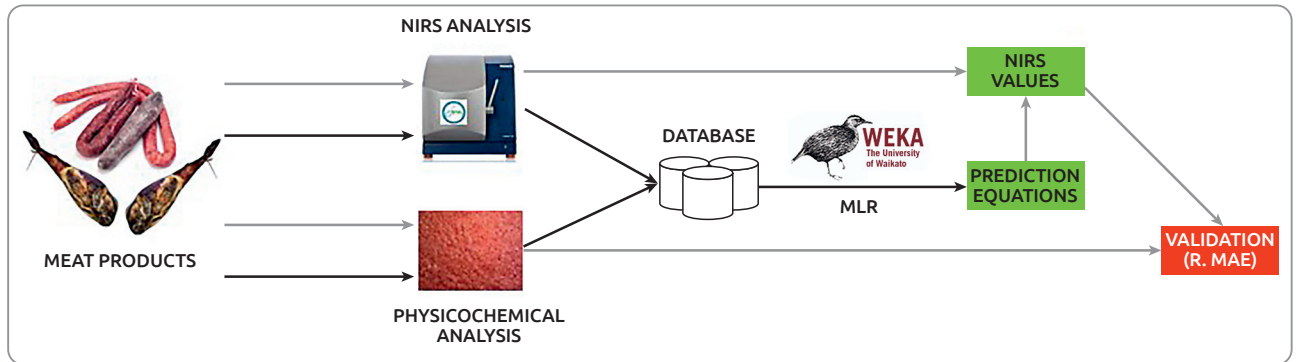
Figure 1. Experimental design. The training batch (n=45) was analysed by the black line (---) to obtain the prediction equations. The validation batch (n=50) was analysed by the grey line (---) to validate the prediction equations.

Physico-chemical analyses

The following physico-chemical analyses were performed in quintuplicate to determine the main nutritional parameters of each sample.

Lipid extraction of Iberian meat products was performed using the original extraction ratio of 20 parts of chloroform:methanol (2:1 v/v) to 1 part of sample. Briefly, 5 g of Iberian meat product sample were mixed with 100 ml of chloroform:methanol (2:1 v/v). The mixture was homogenised, centrifuged (10 min, 3000 rpm) and filtered. Subsequently, 5 ml of distilled water was added to the filtrate and the new mixture was shaken vigorously. The final biphasic system was separated by centrifugation (10 min, 3000 rpm). The upper aqueous phase was eliminated. The lower chloroform phase was filtered through anhydrous sodium sulphate and collected. The lipid content was then gravimetrically determined after chloroform was evaporated with a rotary evaporator under vacuum and the solvent was further evaporated under nitrogen (*Pérez-Palacios et al.*, 2008).

Salt content was determined by the official method for meat and meat products (*AOAC*, 2000, ref. 971.19). It consists of mixing the sample with water and ethyl alcohol. After successive centrifugations, the final extract is obtained and further measured using volumetric analysis by precipitation.

Protein content was determined by the official method (*AOAC*, 2000, ref. 981.10). It consists of determining the nitrogen content by the Kjeldahl digestion method based on volumetric analysis, and then deriving the protein content by multiplying the nitrogen content by a factor of 6.25.

Carbohydrate content was determined by difference (Equation 1), according to the official method (*AOAC*, 2000, Item 85).

$$Carbohydrate = 100 - (proteins + moisture + lipids + ash + salt + fibres) \qquad (1)$$

NIRs analyses

Approximately 20 g of each sample of Iberian meat product were minced using a commercial meat mincer (Moulinex A327R1, Moulinex, Alençon, France) and analysed by NIRs spectrometer (FOSS FOODSCAN lab, FOSS analytics, Hillerod, Denmark) using a wavelength range from 850 nm to 1048 nm and taking 45 s for each spectra. This spectrometer has a wavelength accuracy of 0.5 nm and a wavelength precision of 0.1 nm. For each sample, five spectra were acquired.

Next, spectral data were imported using WinIsi III (FOSS analytics, Hillerod, Denmark) to extract the numerical data from each NIRs spectrum. Then, the noise was eliminated by MSC filter (*Martens and Naes*, 1989). This correction allows a measured spectrum to be compared against a reference spectrum, the spectrum can be corrected using the slope of this fit, and, consequently, any outlier spectra can be removed.

Finally, values for 2 nm bandwidths were calculated from the NIRs spectra. Therefore, for each NIRs spectrum, two hundred values were calculated, extracted and gathered into the database.

Data mining analyses

Predictive data mining techniques were applied to the database constructed with results from the physico-chemical and NIRs analyses. Future models can be predicted from current data using trend analysis (*Wu et al.*, 2008). Therefore, for each main nutritional parameter of each Iberian meat product, predictive equations as a function of NIRs data were obtained. Thus, the main nutritional parameters could be calculated as a function of NIRS data.

The free software WEKA (Waikato Environment for Knowledge Analysis v. 3.8.1, University of Waikato, Hamilton, New Zealand) was used to perform the predictive analyses. MLR models the linear relationship between a target variable and multiple independent variables. It produces a linear regression equation that can be used to predict future values (*Hastie et al.*, 2001). For the selection of attributes, the M5 method was applied. This method is based on stepping though the attributes, the one with the smallest standardized coefficient being removed until no further improvement is observed in the error estimation (*Kira and Rendell*, 1992). A ridge value of 1×10^{-4} was applied. The estimation procedure was a 10-fold cross validation (*Dietterich*, 1998), where the data were divided into 10 subsets of equal size. One subset was tested each time and the remaining data were used for fitting the model. The process was repeated sequentially until all subsets were tested. Therefore, all data subsets were used for both training and testing. However, since this method requires 10 analyses (i.e. with the 10 different data subsets), it is a robust method (*Grossman et al.*, 2010).

The correlation coefficient (R; equation 2) was used for evaluating the goodness of the prediction and for its validation. According to the Colton rules (*Colton*, 1974), R from 0 to 0.25 is considered as little to no association, from 0.25 to 0.50 indicates a weak degree of relationship, from 0.50 to 0.75 indicates a moderate to good degree of relationship and from 0.75 to 1 indicates a very good to excellent degree of relationship.

$$R = \sqrt{\frac{\sum_{i=1}^{n}(f_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (2)$$

where fi is the predicted value, yi is the real value and mean (y) is the average value.

Moreover, the mean absolute error (MAE; equation 3), root mean square error of calibration (RMSEC; equation 4) and root mean square error of prediction (RMSEP; equation 5) (*Hyndman*, 2006) were also used to validate the prediction results. The MAE measures the difference between real values and predicted ones. Values of MAE lower than 2 are appropriate (*Hartemink and Minasny*, 2016).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \quad (3)$$

where fi is the predicted value and yi is the real value.

The RMSEP measures the relative difference between real values and predicted ones. This measure is commonly used to assess the predictive ability of the models, since is a constant measure for prediction. Values of RMSEP lower than 5 are appropriate (*Hartemink and Minasny*, 2016).

$$RMSEP\ (\%) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \times 100 \quad (4)$$

where fi is the predicted value and yi is the real value.

The RMSEC measures the goodness of fit between real data and the data from the calibration model. Depending on the type of data, models and their application can be subject to huge optimistic bias due to over-fitting compared to the results when applying the calibration (*Austin and Tu*, 2004). Values of RMSEC lower than 5 are appropriate (*Hartemink and Minasny*, 2016).

$$RMSEC\ (\%) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \times 100 \quad (5)$$

where fi is the real value and yi is the value obtained by the calibration model.

Comparison of analytical methods

The compatibility index (En) (*Beilby*, 1972) was applied to evaluate the compatibility of the official methods of analyses and the new NIRs method. This index was calculated by the following equation (Equation 6):

$$E_n = \frac{|X_1 - X_2|}{\sqrt{(\frac{S_1}{N_1})^2 + (\frac{S_2}{N_2})^2}} \quad (6)$$

where X1 is the average value according to the official method and X2 is the average value according to the new proposed method. S1 and S2 are the standard deviation for the official method and the new method, respectively. N1 is the number of samples for the official method and N2 is the number of samples for the new method. Values of En lower than 2 indicate the methods are compatible (*Golnick et al.*, 2016).

Statistical analyses

Groups of values of the main nutritional parameters calculated by the official and proposed methods from training and validation batches were compared using one-way analysis of variance (ANOVA) using the general linear model procedure. In addition, the effect between batches was compared by using ANOVA. Statistical analyses were performed using the SPSS package software (IBM SPSS v. 20.0, IBM Co, New York, New York, USA, 2011).

Results and Discussion

Results from training batch

The main nutritional parameters of the Iberian meat products were obtained by applying the traditional physico-chemical analyses on the training batch. Table 1 shows the results from these analyses.

The values for the main nutritional parameters were quite in concordance with previous studies on Iberian meat products (*Caballero et al.*, 2016; *Caballero et al.*, 2018; *Cruz and Vieira*, 2017; *Lorenzo et al.*, 2000; *Muriel et al.*, 2004; *Utrilla et al.*, 2010; *Ventanas*, 2012).

MLR was applied to the NIRs data from the training batch of Iberian meat products to obtain prediction equations for the main nutritional parameters (protein, lipid, salt and carbohydrate contents) of these meat products. Table 1 shows the predicted values based on the NIRs method and the correlation coefficient, p-value, RMSEP, RMSEC and MAE of the prediction equations for the main nutritional parameters of each Iberian meat product.

Note the correlation coefficient of the carbohydrate contents of dry-cured ham, dry-cured loin and dry-cured shoulder were not calculated because the values of the carbohydrates in these Iberian dry-cured

Table 1. Mean ± standard error value, p-value, correlation coefficient (R), root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and mean absolute error (MAE) between values obtained using the official and new NIRs methods to produce the parameters (lipid content (%), protein content (%), salt content (%) and carbohydrate content), the main mandatory items on nutritional labels, for the training batch of different Iberian dry-cured and dry-fermented meat products (chorizo, ham, loin, Salchichón and shoulder).

	Official method	NIRs	p-value	R	RMSEC	RMSEP	MAE
LIPID CONTENT (%)							
Chorizo	44.0 ± 0.17	44.2 ± 0.14	0.996	0.923	0.81	1.22	0.67
Ham	19.5 ± 0.21	19.4 ± 0.12	0.969	0.996	0.10	0.22	0.08
Loin	20.1 ± 0.32	20.0 ± 0.35	0.998	0.957	0.95	1.21	0.82
Salchichón	39.9 ± 0.28	39.8 ± 0.16	0.881	0.932	0.95	1.28	0.81
Shoulder	27.7 ± 0.23	27.6 ± 0.19	0.999	0.913	0.98	1.39	0.86
PROTEIN CONTENT (%)							
Chorizo	26.5 ± 0.21	26.6 ± 0.07	0.999	0.909	0.42	0.76	0.34
Ham	28.8 ± 0.18	28.7 ± 0.10	0.979	0.993	0.12	0.25	0.10
Loin	30.9 ± 0.39	32.9 ± 0.50	0.999	0.973	0.92	1.38	0.74
Salchichón	29.3 ± 0.14	29.4 ± 0.23	0.893	0.860	0.38	0.52	0.30
Shoulder	30.0 ± 0.49	29.8 ± 0.62	0.999	0.924	0.86	1.31	0.68
SALT CONTENT (%)							
Chorizo	3.76 ± 0.14	3.77 ± 0.02	0.992	0.910	0.02	0.04	0.02
Ham	4.06 ± 0.10	4.04 ± 0.19	0.948	0.972	0.05	0.08	0.03
Loin	4.42 ± 0.30	4.41 ± 0.42	0.996	0.989	0.13	0.23	0.09
Salchichón	3.42 ± 0.08	3.43 ± 0.02	0.789	0.875	0.05	0.07	0.04
Shoulder	3.06 ± 0.48	3.05 ± 0.55	0.999	0.981	0.30	0.42	0.23
CARBOHYDRATE CONTENT (%)							
Chorizo	2.17 ± 0.04	2.19 ± 0.04	0.997	0.934	0.21	0.39	0.16
Ham	< 0.10	--------------	----------	-------	----------	----------	---------
Loin	< 0.10	--------------	----------	-------	----------	----------	---------
Salchichón	2.19 ± 0.05	2.17 ± 0.04	0.862	0.912	0.27	0.35	0.22
Shoulder	< 0.10	--------------	----------	-------	----------	----------	---------

Table 2. Mean ± standard error value, the p-value, the correlation coefficient (R), root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and mean absolute error (MAE) between values obtained using the official and new NIRs methods to produce the parameters (lipid content (%), protein content (%), salt content (%) and carbohydrate content), the main mandatory items on nutritional labels, for the validation batch of different Iberian dry-cured and dry-fermented meat products (chorizo, ham, loin, Salchichón and shoulder).

	Official method	NIRs	p-value	R	RMSEC	RMSEP	MAE
LIPIDS CONTENT (%)							
Chorizo	43.6 ± 0.12	43.8 ± 0.14	0.459	0.808	1.13	1.27	1.09
Ham	20.1 ± 0.24	20.4 ± 0.27	0.635	0.929	0.96	1.23	0.88
Loin	18.8 ± 0.22	19.1 ± 0.23	0.525	0.905	1.17	1.31	1.05
Salchichón	38.5 ± 0.09	38.6 ± 0.19	0.961	0.896	0.96	1.02	0.93
Shoulder	28.2 ± 0.15	28.3 ± 0.23	0.197	0.864	1.41	1.53	1.27
PROTEINS (%)							
Chorizo	25.7 ± 0.14	25.6 ± 0.16	0.826	0.843	0.72	0.81	0.66
Ham	27.7 ± 0.12	27.6 ± 0.13	0.983	0.783	1.16	1.40	1.10
Loin	32.6 ± 0.26	32.8 ± 0.27	0.774	0.956	0.97	1.11	0.85
Salchichón	29.3 ± 0.09	29.2 ± 0.19	0.782	0.847	0.71	0.78	0.64
Shoulder	27.3 ± 0.22	28.1 ± 0.24	0.695	0.870	1.47	1.64	1.40
SALT CONTENT (%)							
Chorizo	3.76 ± 0.14	3.77 ± 0.02	0.321	0.865	0.11	0.15	0.08
Ham	4.06 ± 0.10	4.04 ± 0.19	0.203	0.828	0.14	0.18	0.12
Loin	4.42 ± 0.40	4.41 ± 0.62	0.989	0.927	0.26	0.35	0.21
Salchichón	3.42 ± 0.08	3.43 ± 0.02	0.987	0.816	0.18	0.23	0.14
Shoulder	3.06 ± 0.48	3.05 ± 0.55	0.725	0.955	0.35	0.45	0.31
CARBOHYDRATES (%)							
Chorizo	2.21 ± 0.03	2.29 ± 0.04	0.314	0.779	0.32	0.47	0.28
Ham	< 0.10	--------------	----------	-------	----------	----------	---------
Loin	< 0.10	--------------	----------	-------	----------	----------	---------
Salchichón	2.00 ± 0.03	2.04 ± 0.04	0.665	0.818	0.32	0.39	0.27
Shoulder	< 0.10	--------------	----------	-------	----------	----------	---------

meat products is considered minimum (<0.100%) and at undetectable levels according to the current regulation (*European Union*, 2012; *Ventanas*, 2012). Therefore, for the new NIRs method proposed in this study, the values of carbohydrates for these meat products is shown as lower than 0.100 % (not detectable) in all samples of dry-cured ham, dry-cured loin and dry-cured shoulder.

Prediction by means of NIRs data and values for nutritional analyses for the remaining cases achieved very good to excellent degrees of relationship according to the rules given by Colton (*Colton*, 1974) (R >0.75). R values were R >0.97 for the dry-cured

ham and R >0.95 for the dry-cured loin, for all nutritional parameters. For MAE, for all prediction equations, the values obtained were lower than 1, which is an appropriate value for MAE (*Hyndman*, 2006). For RMSEP and RMSEC, for all prediction equations, the values obtained were lower than 1.50, which are appropriate values for RMSEP and RMSEC (*Hartemink and Minasny*, 2016). However, the most accurate results were obtained for the dry-cured ham (MAE < 0.15, RMSEC < 0.15, RMSEP < 0.30). Moreover, no significant differences were found between values from official methods and the new method based on prediction equations.

These results indicate the ability of the proposed model to calculate the main nutritional parameters based on NIRs and data mining are comparable with the results obtained by the official methods of analysis. Thus, the NIRs method could be proposed as an alternative method to estimate the values of these nutritional parameters with similar accuracy as the official methods. Previous studies obtained similar results (R > 0.75) for the nutritional parameters salt and lipid contents (*Collell et al.*, 2011; *González-Mohino et al.*, 2018; *Pérez-Juan et al.*, 2010; *Pérez-Palacios et al.*, 2019; *Zamora-Rojas et al.*, 2011). MLR was evaluated to predict some quality parameters of Iberian dry-cured meat products, dry-cured ham or dry-cured loin, in previous studies (*Caballero et al.*, 2016; *Caballero et al.*, 2017a; *Caballero et al.*, 2017b; *Caballero et al.*, 2018; *Pérez-Palacios et al.*, 2014; *Pérez-Palacios et al.*, 2017).

The ability of this proposed method to calculate the main nutritional parameters of different Iberian dry-cured meat products was next evaluated using the validation batch.

Results from validation batch

Taking a step forward, for validating the models obtained from the training batch, the prediction equations obtained from the training batch were applied to the NIRs data from validation batch of the different Iberian meat products.

Table 2 shows the values from the official methods of analyses and the values from the new method based on NIRs, the correlation coefficients, RMSEC, RMSEP and MAE of the results obtained from NIRs data and data from the official analytical methods for the main nutritional parameters (lipid, protein, salt and carbohydrate contents) of the studied Iberian meat products. As previously stated, the values of carbohydrates for some dry-cured products (ham, loin and shoulder) were labelled as not detectable according to the current regulation (*European Union*, 2012).
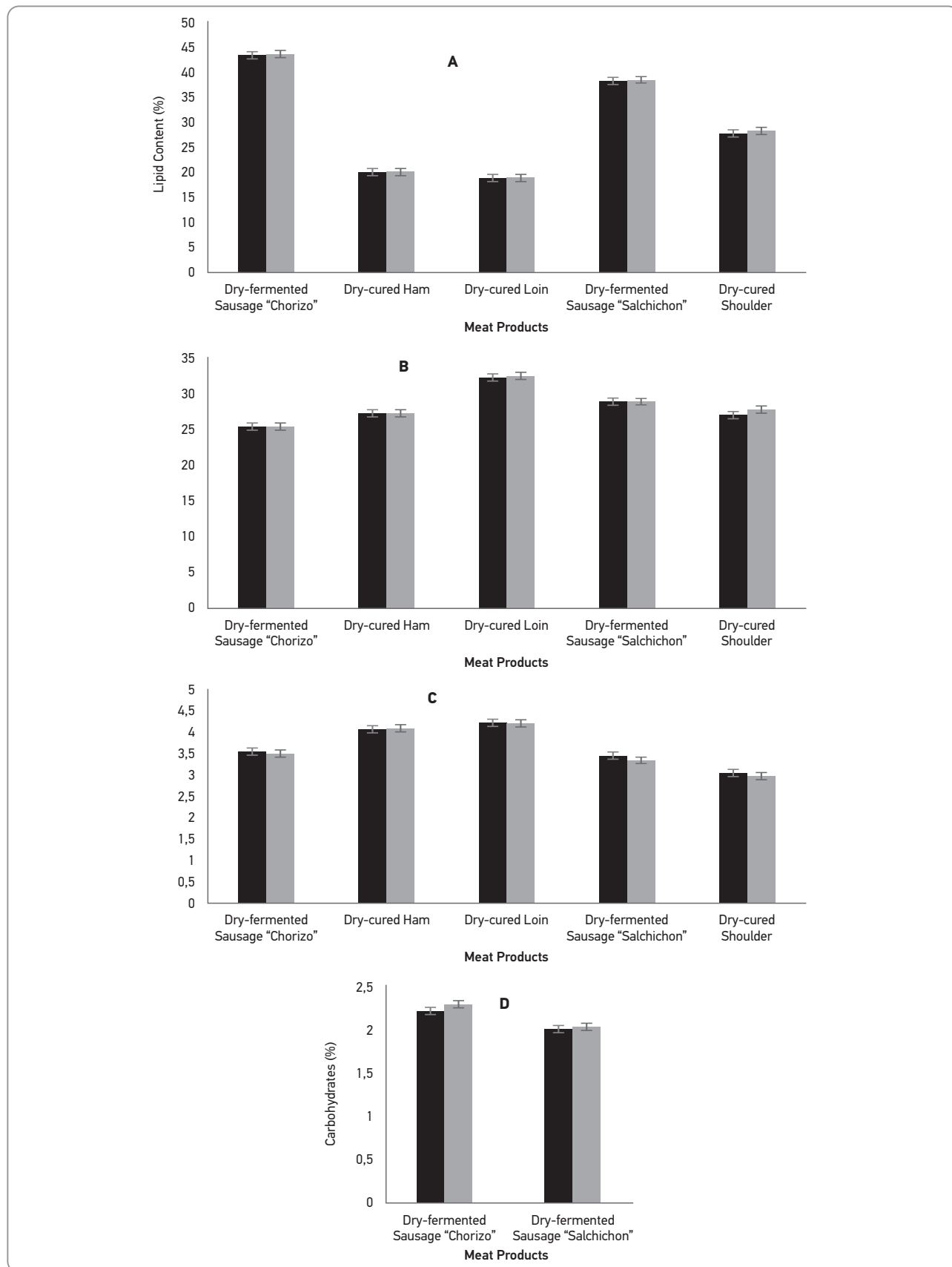
The values obtained from the official methods and from the new method of analysis (Table 2) were similar to the results obtained from the training batch (Table 1), and they are in agreement with the results obtained in previous studies (*Caballero et al.*, 2016; *Caballero et al.*, 2018; *Cruz and Vieira*, 2017; *Lorenzo et al.*, 2000; *Muriel et al.*, 2004; *Utrilla et al.*, 2010; *Ventanas*, 2012).

The values obtained for the NIRs method and the official methods for all studied cases achieved very good to excellent correlation coefficients according to the rules given by Colton (1974) (R > 0.75). The strongest degrees of relationships were achieved for dry-cured loin (R > 0.90) and dry-cured shoulder (R > 0.85). Regarding the nutritional parameters, the highest correlation coefficients were found for the lipid and salt contents (R > 0.80) from all Iberian meat products. In relation to the MAE, in all cases, the values obtained were lower than 1.50, which is a good value for MAE (*Hyndman*, 2006), and for RMSEP and RMSEC, all values were lower than 2, which is a very good value for RMSEP and RMSEC (*Hartemink and Minasny*, 2016). However, the most accurate results were obtained for the Salchichón dry-fermented sausage (MAE < 1, RMSEC < 1 and RMSEP < 1.10).

Evaluating the values obtained for the compatibility index (Table 3), all values obtained were lower than 2 (En < 2), indicating the analytical methods used in this study were compatible (*Golnick et al.*, 2016). For some products, their compatibility indices were lower than 1: dry-cured ham (En < 0.75), dry-fermented Chorizo sausage (En < 0.85) and dry-cured loin (En < 0.95). These results support the

Table 3. Compatibility index (En) results, comparing the official methods and the new NIRs method of analysis analysing the main parameters on mandatory nutritional labels for the validation batch of different Iberian dry-fermented and dry-cured meat products.

	Chorizo	**Ham**	**Loin**	**Salchichón**	**Shoulder**
Lipid content (%)	0.816	0.727	0.910	0.077	0.535
Protein content (%)	0.263	0.125	0.438	0.214	1.898
Salt content (%)	0.463	0.482	0.011	1.089	0.341
Carbohydrate content (%)	0.583	--------------	--------------	0.241	--------------

Figure 2. Results for the main mandatory nutritional parameters (mean value ± standard error) of the validation batch determined by official methods (black) and estimated from the new method based on NIRs data (grey) for a) lipid content (%), b) protein content (%), c) salt content (%) and d) carbohydrate content (%).

compatibility of the newly-developed method based on NIRs and data mining with the official methods of analysis.

The batch effect on the main mandatory nutritional parameters of the studied Iberian meat products (dry-fermented Chorizo sausage, dry-cured ham, dry-cured loin, dry-fermented Salchichón sausage and dry-cured shoulder) was studied. No significant differences were found between training and validation batches for the main mandatory nutritional parameters. These results indicated the assignment of samples into batches had no influence on the results obtained, and therefore, the results are batch-independent. In a previous study, the batch effect acted like a random effect (*Herbert et al.*, 1974), but in our case, the batch did not influence the resul ts.

Average values for the main mandatory nutritional parameters determined by official methods and our new method based on NIRs are shown in Figure 2. Thus, the accuracy of the new method based on NIRs was corroborated for lipid (Figure 2A), protein (Figure 2B), salt (Figure 2C) and carbohydrate (Figure 2D) contents. No significant differences were found between values from official methods and the new method based on NIRs. These results are in accordance with the results showed in Table 3.

Conclusion

A new, fast and accurate analytical method was studied, which should be suitable to analyse the content of the main nutritional parameters required for mandatory labelling according to the new EU regulation. The new method is based on prediction equations obtained from the combination of NIRs spectra and data mining techniques. In other proposed methods in ISO 13528, the lead-in time to produce results is around 5 or 6 days, with similar accuracy to the official methods (although these analyses were conducted on rice, not on meat and meat products). However, the new NIRs method proposed in this study is faster and more accurate than the previous methods for the required nutritional analyses in comparison with the traditional methods. Additionally, only a small number of samples is required before the results obtained are accurate. Thus, the new NIRs method produces the main nutritional parameters for each dry-cured meat product in around 10 minutes per sample, using a NIRs spectrometer and a conventional laptop in an automated way. This method could be of interest to inspection agencies in order to evaluate the nutritional labelling of Iberian meat products in a timely manner.

Odtređivanje obaveznih nutritivnih parametara Iberijskih mesnih proizvoda korišćenjem nove metode zasnovane na NIR i pretraživanju podataka

Daniel Caballero, Maria Asensio, Carlos Fernández, Noelia Martín, Antonio Silva

A p s t r a k t: Nova uredba o obaveznom označavanju nutritivnih sastojaka zahteva deklarisanje određenih parametara, kao što su proteini, sadržaj lipida, sadržaj soli ili ugljenih hidrata. Ovo istraživanje ima za cilj da razvije brz i tačan metod da se na automatski način utvrde vrednosti obaveznih nutritivnih parametara na osnovu NIR tehnologije (bliska infracrvena spektroskopija) i tehnika pretraživanja podataka. U ovu svrhu su korišćene dve serije različitih proizvoda od svinjskog mesa (suva pršuta, suva slabina, suva fermentisana kobasica „Salchichon" i suva fermentisana kobasica „Chorizo"). Jedna serija je korišćena kako bi se trenirala metoda, a druga serija za svrhu validacije. Da bi se razvio postupak, jednačine predviđanja dobijene su iz NIR i nutritivnih podataka prve serije primenom tehnika pretraživanja podataka, a zatim su jednačinee procenjivane sa NIR podacima iz serije validacije. Štaviše, jednačine predviđanja postigle su vrlo dobar do odličan stepen korelacije (R> 0,75) i tačnih rezultata (MAE < 1, RMSEC < 1, RMSEP < 1) sa prvom, trenažnom serijom. Ove jednačine predviđanja potkrepljene se u grupi za validaciju, koja je pokazala vrlo dobre do odlične koeficijente korelacije (R> 0,75). Prema tome, ova nova metoda traje oko 10 minuta u poređenju sa tradicionalnom metodom koja traje oko 6 dana.

Ključne reči: *svinjsko meso, NIR, pretraživanje podataka, deklarisanje, Iberijski mesni proizvodi, nutritivne informacije.*

References

Austin, P.C. & Tu, J.V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction morality. *Journal of Clinical Epidemiology, 57*, 1138–1146.

Association of Official Analytical Chemistry (AOAC) (2000). Official Methods of analysis of AOAC International, AOAC International, Gaithersburg, Maryland, USA.

Ávila, M.M., Caballero, D., **Antequera, T., Durán, M.L., Caro, A. & Pérez-Palacios, T. (2018).** Applying 3D texture algorithms on MRI to evaluate quality traits of loin. *Journal of Food Engineering, 222*, 258–266.

Beilby, A. L. (1972). Comparison of analytical methods. *Journal of Chemical Education*, *49*(10), 679–681.

Benoit, M., Grass, F., Demartines, N., Coti-Bertrand, P., Schäfer, M. & Hübner, M. (2016). Use of nutritional risk score by surgeons and nutritionist. *Clinical Nutrition*, *35*(1), 230–233.

Caballero, D., Caro, A., Rodríguez, P.G., Durán, M.L., Ávila, M.M., Palacios, R., Antequera, T. & Pérez-Palacios, T. (2016). Modeling salt diffusion in Iberian ham by applying MRI and data mining. *Journal of Food Engineering*, *189*, 115–122.

Caballero, D., Pérez-Palacios, T., Caro, A., Amigo, J.M., Dahl, A.B., Ersboll, B.K. & Antequera, T. (2017a). Prediction of pork quality parameters by applying fractals and data mining on MRI. *Food Research International*, *99*, 739–747.

Caballero, D., Antequera, T., Caro, A., Ávila, M.M., Rodríguez, P.G., & Pérez-Palacios, T. (2017b). Non-destructive analysis of sensory traits of dry-cured loins by MRI-computer vision techniques and data mining. *Journal of the Science of Food and Agriculture*, *97*, 2942–2952.

Caballero, D., Asensio, M., Fernández, C., Reina, R., García-Casco, J., Martín, N. & Silva, A. (2019). Chemical-instrumental-sensory traits and data mining for classifying dry-cured Iberian shoulders from pigs with different diets. *Journal of Food Measurement and Characterization*, *13*, 2935–2950.

Cava, R., Ventanas, J., Ruiz, J., Andrés, A.I. & Antequera, T. (2000). Sensory characteristics of Iberian ham: Influence of rearing system and muscle location. *Food Science and Technology International*, *6*, 235–242.

Cifuni, G. F., Cantó, M., & Failla, S. (2016). Potential use of visible reflectance spectra to predict lipid oxidation of rabbit meat. *Journal of Food Engineering*, *169*, 85–90.

Collell, C., Gou, P., Arnau, J. & Comaposada, J. (2011). Non-destructive estimation of moisture, water activity and NaCl at ham surface during resting and drying using NIR spectroscopy. *Food Chemistry*, *129*, 601–607.

Colton, T. (1974). Statistical in Medicine, Little Brown and Co., New York, New York, USA.

Cortez, P., Portelinha, S., Rodrigues, S., Cadavez V. & Teixeira, A. (2006). Lamb meat quality assessment by support vector machine. *Neural Processing Letters*, *24*, 41–51.

Cozzolino, D., Murray, I., Scaife, J. R. & Paterson, R. (2000). Study of dissected lamb muscle by visible and near infrared reflectance spectroscopy for composition assessment. *Animal Science Journal*, *70*(3), 417–423.

Cruz, R. M. S. & Vieira, M. C. (2017). Mediterranean Foods. Composition and Processing, CRC Press, Boca Ratón, Florida, USA.

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895–1923.

European Union (2000). Commission regulation (EU) No. 2000/13/EC of 20 March 2000. Laws of the member states relating to the labelling presentation and advertising foodstuffs.

European Union (2011). Commission regulation (EU) No. 1169/2011 of 25 October 2011. Regulation on food information provided to the consumer.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery in databases. *Artificial Intelligence Magazine*, *17*, 37–54.

Fulladosa, E., Santos-Garcés, E., Picouet, P. & Gou, P. (2010). Prediction of salt and water content in dry-cured hams by computed tomography. *Journal of Food Engineering*, *96*, 80–85.

Golnick, P. C., Chaffin, J. D., Bridgeman, T. B., Zellner, B. C. & Simons, V. E. (2016). A comparison of water sampling and analytical methods in western Lake Erie. *Journal of Great Lakes Research*, *42*(5), 965–971.

González-Mohino, A., Antequera, T., Ventanas, S., Caballero, D., Mir-Bel, J., & Pérez-Palacios, T. (2018). NIRs-based analysis to study sensory parameters of pork loins as affected by cooking methods and conditions. *Journal of the Science of Food and Agriculture*, *98*, 4227–4236.

Grossman, R., Seni, G., Elder, J., Agarwal, N. & Liu, H. (2010). Ensemble methods in data mining: Improving accuracy through combining predictions, Morgan & Claypool Publishers, Williston, Vermont, USA.

Hartemink, A. E. & Minasny, B. (2016). Progress in soil science. Digital soil morphometrics, Springer-Verlag, St. Gallen, Switzerland.

Hastie, T., Tibshirani, R. & Friedman, J. (2001). The elements of statistical learning. Data mining inference and prediction. Springer-Verlag, New York, New York, USA.

Herbert, L. S., Dillon, J. F., MacDonald, M. W. & Skurray, G. R. (1974). Batch dry rendering: influence of processing conditions on meat meal quality. *Journal of the Science of Food and Agriculture*, *25*, 1063–1070.

Hyndman, R. J. (2006). Another look at forecast accuracy metrics for intermittent demand. *International Journal of Applied Forecast*, *4*, 43–46.

IBM SPSS (Statistical Package Software) v. 20.0 (2012). IBM co. New York, New York, USA.

Judprasong, K., Puwastein, P., Boonpor, J. & Pinprapai, N. (2013). Laboratory performance on analysis of mandatory nutrients and preparation of nutrition labelling. *Food Chemistry*, *140*(3), 598–607.

Kira, K. & Rendell, R. A. (1992). A practical approach to feature selection. IX International workshop on machine learning. Aberdeen, United Kingdom.

Lorenzo, J. M., Michinel, M., López, M., & Carballo, J. (2000). Biochemical characteristics of two Spanish traditional dry-cured sausages varieties: Androlla y Botillo. *Journal of Food Composition and Analysis*, *13*(5), 809–817.

Martens, H. & Naes, T. (1989). Multivariate calibration, John Wiley, Chichester, United Kingdom.

Mitchell, T. M. (1999). Machine learning and data mining. *Communications of ACM*, *42*, 30–36.

Muriel, E., Ruiz, J., Martín, D., Petrón, M.J., & Antequera, T. (2004). Physico-chemical and sensory characteristics of dry-cured loin from different Iberian pig lines. *Food Science and Technology International*, *10*, 117–123.

Pérez-Juan, M., Afseth, N. K., González, J., Diaz, I., Gispert, M., Furnols, M. F., Oliver, M. A. & Realini, C. E. (2010). Prediction of fatty acid composition using a NIRs fibre optics probe at two different locations of ham subcutaneous fat. *Food Research International*, *43*, 1416–1422.

Pérez-Palacios, T., Ruiz, J., Martín, D., Muriel, E. & Antequera, T. (2008). Comparison of different methods for total lipid quantification. *Food Chemistry*, *110*, 1025–1029.

Pérez-Palacios, T., Caballero, D., Caro, A., Rodríguez, P.G. & Antequera, T. (2014). Applying data mining and computer vision techniques to MRI to estimate quality traits in Iberian hams. *Journal of Food Engineering*, *131*, 82–88.

Pérez-Palacios, T., Caballero, D., Antequera, T., Durán, M. L., Ávila, M. M. & Caro, A. (2017). Optimization of MRI acquisition and texture analysis to predict physico-chemical parameters of loins by data mining. *Food and Bioprocess Technology*, *10*, 750–758.

Pérez-Palacios, T., Caballero, D., González-Mohino, A., Mir-Bel, J., & Antequera, T. (2019). Near Infrared Reflectance spectroscopy to analyse texture related characteristics of sous vide pork loin. *Journal of Food Engineering*, *263*, 417–423.

Picouet, P.A., Gou, P., Fulladosa, E., Santos-Garcés, E., & Arnau, J. (2013). Estimation of NaCl diffusivity by computed tomography in the Semimembranosus muscle during salting of fresh and frozen/thawed hams. *LWT-Food Science and Technology*, *51*, 275–280.

Sayad, S. (2011). Real time data mining. Self-Help Publishers, Cambridge, Ontario, Canada.

Silva, A., Reina, R., García-Casco, J. & Ventanas, J. (2013). Chemical-instrumental-sensory parameters and chemometrics as tools to discriminate among the quality of dry-cured Iberian shoulder. *Grasas y Aceites*, *64*(2), 201–209.

Song, Y. H., Kim, S. J. & Lee, S. K. (2002). Evaluation of ultrasound for prediction of carcass meat yield and meat quality in Korean native cattle. *Asian-Australasian Journal of Animal Science*, *15*(4), 591–595.

Utrilla, M. C., Soriana, A., & García-Ruiz, A. (2010). Quality attributes of pork loin with different levels of marbling from Duroc and Iberian cross. *Journal of Food Quality*, *33*, 802–820.

Ventanas, J. (2012). Jamón Ibérico y Serrano: Fundamentos de la elaboración y la calidad, Mundi-Prensa, Madrid, Spain.

WEKA (Waikato Environment Knowledge Analysis) v.3.8.1 (2018). University of Waikato, Hamilton, New Zealand.

WinISI III software (2005). FOSS analytics, Hillerod, Denmark

Wu, X., Kumar, V., Ross-Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J. & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge Information Systems*, *14*, 1–37.

Zamora-Rojas, E., Garrido-Varo, A., De Pedro-Sanz, E., Guerrero-Ginel, J. E. & Pérez-Marín, D. (2011). Monitoring NIRs calibration for use in routine meat analysis as part of Iberian pig breeding programs. *Food Chemistry*, *129*(4), 1889–1897.