16/10/2024 17:22

(Article begins on next page)

Iodine Value and Fatty Acids Determination on Pig Fat Samples by FT-NIR Spectroscopy: Benefits of Variable Selection in the Perspective of Industrial Applications

Giorgia   Foca [1,*,2]

Phone +39 0522 522042
Email giorgia.foca@unimore.it

Carlotta   Ferrari [1]

Alessandro   Ulrici [1,2]

Maria Cristina   Ielo [2]

Giovanna   Minelli [1,2]

Domenico Pietro   Lo Fiego [1,2]

[1] Department of Life Sciences, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola 2, 42122 Reggio Emilia, Italy

[2] Interdipartimental Research Centre for Agri-Food Biological Resources Improvement and Valorisation, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola 2, 42122 Reggio Emilia, Italy

Abstract

In this work, FT-NIR spectroscopy was employed to determine iodine value (IV) and fatty acids (FA) content of pig fat samples, through the combined use of signal preprocessing, multivariate calibration, and variable selection methods. In particular, the main focus was on the use of variable selection methods, both in order improve the predictive performance of the calibration models, and to identify relevant wavelengths that could be subsequently used for the development of simple, fast, and cheap hand-held devices, able to measure IV and FA content directly on the fat without the need of any sample pretreatment. Firstly, for each property of interest, partial least squares (PLS) multivariate calibration models were calculated considering the whole spectral range and testing different signal preprocessing methods. Then, once chosen the optimal signal preprocessing method, a two-step variable selection procedure was applied. In the first step, the interval-PLS variable selection algorithm was used to calculate a set of calibration models, whose outcomes were considered altogether in the second step, in order to select the optimal calibration model. The variable selection procedure allowed to lower the number of spectral variables retained by the model, and often led to an increase of the performance in prediction of the external test set samples.

Keywords

Fatty acids
Iodine value
FT-NIR spectroscopy
Multivariate calibration
Variable selection

Electronic supplementary material

Introduction

Pig fat is generally intended for different end uses depending on its composition. In fact, the fat taken from the adipose layer immediately below the rind contains a greater amount of unsaturated fatty acids and connective tissue and it has a harder consistency, therefore in the Italian food industry it is diced to be used for the production of sausages. On the contrary, the fat coming from the deeper layer contains a higher amount of saturated fatty acids and it has a softer consistency, hence, it is generally melted for the preparation of semi-finished products (Santoro 1983). For the Italian PDO (protected designation of origin) products, the main methods of control are based on direct evaluation of the fatty acid composition of lipids by gas-chromatographic analysis or on indirect assessment of degree of lipids unsaturation by iodine value evaluation determined by using the Wijs method (Lo Fiego et al. 2005). These procedures are time-consuming and have a bad environmental impact.

In the last years a number of techniques have been proposed as an alternative to classical methods for measuring iodine value and fatty acids. Among them, results comparable to those of the conventional methods have been achieved by means of analytical methodologies based on NMR spectroscopy (Dais et al. 2007). Other non-invasive methods based on vibrational spectroscopic techniques, such as Raman (Berhe et al. 2016) and micro-Raman (Giarola et al. 2011), gave good correlations, together with the more diffuse mid-IR spectroscopy (Afseth et al. 2010). All these methods require a minimal—or none—sample preparation, a reduced analysis time and a very limited consumption of chemical reagents.

In this context, FT-NIR spectroscopy represents a valuable alternative, since it is a fast and non-destructive technique, free from chemical reagents. With respect to other techniques, NIR spectroscopy offers the possibility to implement automated monitoring systems or hand-held devices. This allows to carry out extensive monitoring, virtually on the whole production chain, without requiring sporadic sampling or sample destruction.

The performance of NIR-based systems can be further improved by using proper data processing methods; in particular, great advantages can be gained through the combined use of proper signal preprocessing and multivariate variable selection strategies. In this way, is possible to extract a small number of relevant wavelengths from a signal composed by a wide range of spectral variables, where the information useful to the problem at hand is mixed with other sources of

variability, including noise and non-pertinent information. The selected wavelengths can be subsequently used to implement cheaper and faster instruments (like, e.g., those based on optical filters) properly engineered for the specific application, thus further speeding up the monitoring process.

In a previous work we have already verified that NIR spectroscopy is a technique able to effectively discriminate between fat samples taken from two different subcutaneous layers (Foca et al. 2013). In the present work, we have evaluated the fat samples from a compositional point of view. Calibration models were built in order to predict the amount of different fatty acids contained in the samples and to estimate the iodine value (IV). IV is an index of the degree of unsaturation of a fatty tissue: the more the tissue is rich in saturated fatty acids, the lower is the IV value, and vice versa.

A literature survey revealed that IV of animal fat samples has been investigated by means of NIR spectroscopy in a limited number of papers. In a recent paper concerning animal fat blends destined for biodiesel production (Adewale et al. 2014), the IV values used as the response variable in the calibration models were measured using the Wijs analytical method. In other works (Prieto et al. 2014; Gjerlaug-Enger et al. 2011; Sørensen et al. 2012), the IV reference values were calculated by an equation that takes into account the composition in fatty acids expressed in percentages (AOCS 1998; Pétursson 2002).

The ability to predict the content of different fatty acids by NIR has instead been extensively investigated: in addition to the works previously mentioned about IV, in recent years other papers dealt with this topic. The results obtained in the various research works are often divergent. These differences may be due to variations in the homogeneity of the considered samples or to the different instrumental approaches (Ripoche and Guillard 2001). For instance, in some studies the fat samples were analyzed in transmission or transflectance providing excellent results (Adewale et al. 2014; Gjerlaug-Enger et al. 2011; Ripoche and Guillard 2001; Fernández-Cabanás et al. 2007; Zamora-Rojas et al. 2013). Unfortunately, this analytical method requires fat melting before analysis, therefore the signal acquisition procedure is not directly implementable in rapid monitoring systems.

Conversely, in other works the fat samples were analyzed without any preparative step, using different methods: diffuse

reflectance (Ripoche and Guillard 2001), transmission (Sørensen et al. 2012) and fiber optic probes (Zamora-Rojas et al. 2013; González-Martin et al. 2005; Pérez-Juan et al. 2010). Müller and Scheeder (2008) also measured the samples in diffuse reflectance but they homogenized the fat samples before spectra acquisition.

In the present work we have addressed the issue of "direct" fat characterization (i.e., without performing sample pretreatment steps like melting or homogenization) by means of NIR spectroscopy. With respect to the existing literature, the main concern of our investigation regards the usefulness of applying variable selection in order to identify the most relevant wavelengths for calibration purposes.

Concerning the prediction of iodine value and fatty acids based on NIR spectra, only in Sørensen et al. (2012) the interval-PLS (iPLS) variable selection was applied in addition to the classical partial least squares (PLS) calibration method. However, variable selection should be often considered when dealing with FT-NIR datasets to identify and select the spectral regions in which the information of interest is located (Ferrari et al. 2011; Foca et al. 2009; Ulrici et al. 2008; Cocchi et al. 2005, 2006). In fact, NIR spectra are notoriously composed of very correlated variables and they contain redundant information which is spread out over different spectral regions (Lee et al. 2012).

The variable selection procedure used in the present work consists of two phases. In the first phase, for each property to be predicted, the iPLS algorithm was applied using different interval widths in order to verify the possible—and desirable—convergence of the different models obtained on the same spectral regions. In the second phase, that can be considered as a "selection of the selection", new PLS models were iteratively built using the more frequently selected regions in the corresponding iPLS models. This procedure allows to make the most of the results of the first variable selection phase, putting together the results obtained considering different interval widths in order to gain a comprehensive model. Moreover, the possible convergence of the different models on the same regions brings to the identification of relevant wavelengths for the prediction of fat quality related properties. The wavelength selection has been regarded as an important subject by many authors before (Panford and deMan 1990; Wu et al. 2009; Balabin and Smirnov 2011; Chen et al. 2002), because it constitutes a critical step in the perspective of the development of hand-held devices. In fact, the selection of limited spectral regions for the calculation of multivariate calibration models

could allow the development of simple, fast, and cheap systems that are required for effective industrial applications.

Experimental

Samples

Sixty-five heavy pigs from *Italian Landrace × Large White* crossbreeds, reared on the same farm and fed with the same diet, were slaughtered at about 160 kg of live weight. After slaughter, 1 h post mortem, 205 fat samples were taken at the last rib level by means of the following sampling procedure.

The subcutaneous adipose tissue was hand-slashed by an expert operator in a way to obtain disks of fat tissue having diameter of about 3 cm and thickness ranging from 3 mm to 2 cm. These fat samples consisted in two adjacent layers, lying at different depths with respect to the rind. The layer close to the rind (that was previously removed) was labeled as Outer (*Out*) and the layer far from the rind as Inner (*In*). The two layers were then separated by means of a manual cut, after a visual assessment of the line of demarcation of the layers, to gain the corresponding *Out* and *In* samples. It has to be noticed that for each pig specimen from 1 to 4 *Out* and *In* samples have been collected depending on extent and thickness of the fat layer; which means that for each original fat disk the *Out* and *In* parts were not necessarily all kept.

The samples were stored in dark conditions at −20 °C before analyses. The whole procedure followed for samples collection and analysis is represented in detail in Scheme 1.

Scheme 1

Representation of the procedure used to extract and analyze the pig fat samples
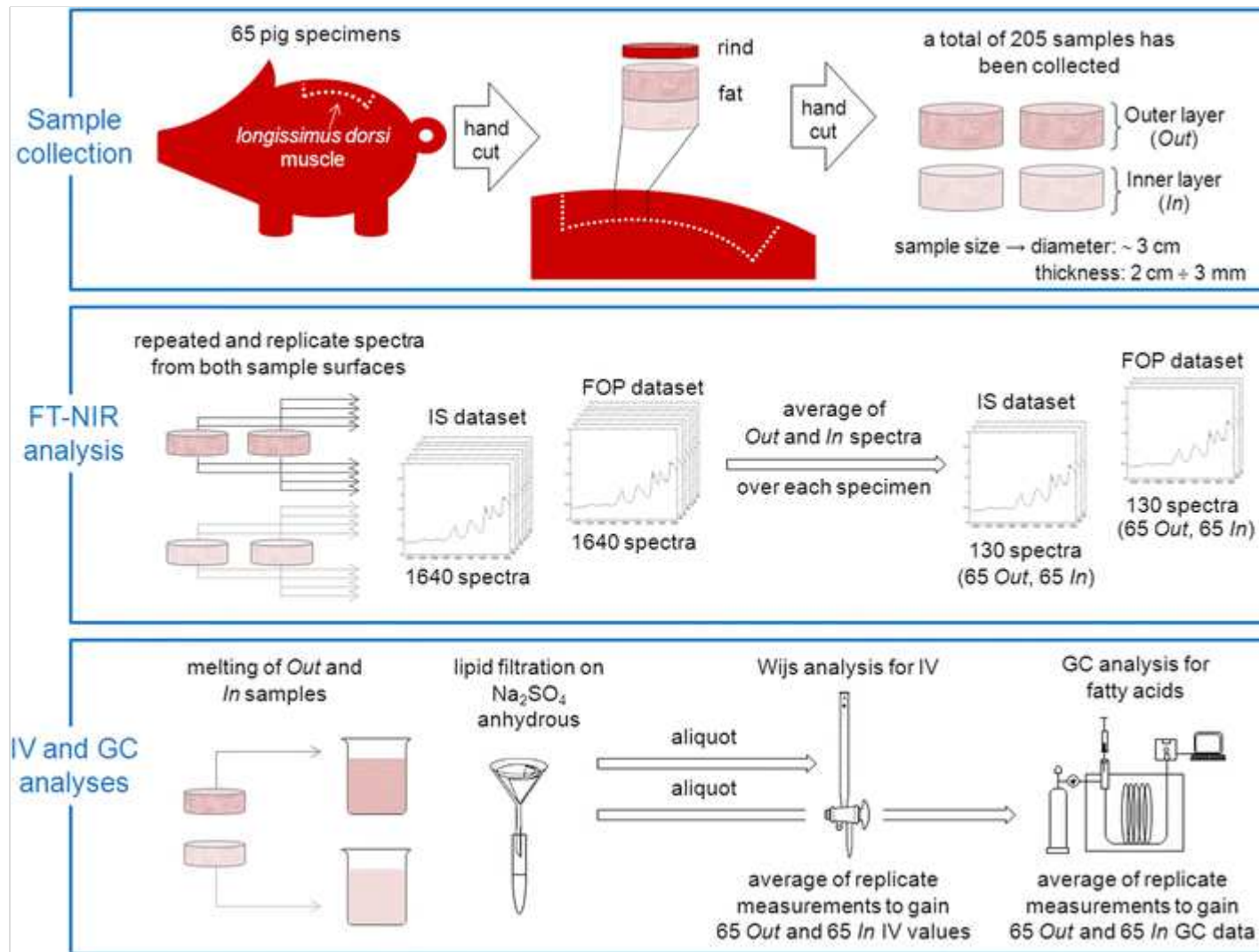
Analyses on Fat Samples

FT-NIR Spectroscopy

In each day of measurement the samples to be analyzed were randomly chosen and defrosted at 4 °C for 1 h and then at

room temperature for 30 min. All the measurements were performed at room temperature. Then, the sample order was shuffled and repeated measurements were performed. At the end of the daily measurement session, the samples were stored again at −20 ℃. This whole procedure was repeated twice in two acquisition sessions.

FT-NIR analysis was performed by means of a Bruker Optics MPA FT-NIR spectrophotometer equipped with Integrating Sphere (IS) and Fiber Optic Probe (FOP). All the spectra were acquired in reflectance mode at 2 cm^{-1} resolution by averaging 64 scans in the 3800–12500 cm^{-1} spectral region for IS, using a glass Petri dish as sample holder, and in the 4000–12500 cm^{-1} spectral region for FOP. A total of 1640 spectra have been acquired for each sampling tool. In particular, for each one of the 205 samples, 8 spectra have been collected as the result of (2 acquisitions on the upper and lower faces of the disk-shaped sample) × (2 acquisition sessions) × (2 repeated spectra in each session).

Iodine Value Determination

For each pig specimen the *In* and *Out* layers were melted separately and the lipid fraction of each sample has been extracted in accordance with IUPAC method II.A.1 (IUPAC 1979).

Iodine value was determined using the Wijs method (AOAC 1984). In detail, an amount of 0.3 g of extracted lipid mixture is weighed and then it is dissolved in 15 ml of $CHCl_3$ by stirring. Afterwards, 25 ml of Wijs reagent are added to the sample and the solution is kept in the dark for 1 h. After that, 20 ml of KI (10 % water solution) and 100 ml of water are added to the sample solution. The analytical sample solution is titrated with standard $Na_2S_2O_3$ 0.1 N using some drops of starch solution as indicator, since it gives an intensely blue complex with iodine; the end point is marked by the disappearance of the color indicator.

The blank corrected iodine value of the sample is finally calculated using Eq. 1:

$$IV = \left[\frac{(ml_{Na_2S_2O_3 blank} - ml_{Na_2S_2O_3 sample}) \times 0.1}{SW} \right] \times 12.69 \qquad 1$$

where 0.1 is the concentration of $Na_2S_2O_3$ expressed as equivalent/L, SW is the sample weight, and 12.69 is a constant related to the equivalent weight of iodine. The resulting IV value is expressed as the mass of iodine in grams that can be consumed by 100 grams of the fat (g I_2/100 g fat).

From the considered fat aliquots, belonging to 65 pig specimens, 158 IV determinations were obtained (65 *In* layer plus 13 replications and 65 *Out* layer plus 15 replications, so as to replicate at least 20 % of the samples).

Gas-Chromatographic Analysis

Fatty acid (FA) composition of lipids was determined using a TRACE™ GC Ultra (Thermo Electron Corporation, Rodano, Milano, Italy) equipped with the Ultra Fast Module (UFM), a Fast Flame Ionization Detector, a PTV injector, and a UFM-Carbowax column, 5 m long, 0.1 mm i.d., 0.2 μm film thickness. In detail, as reported by Minelli et al. (2013), 50 mg of extracted lipids were subjected to methylation by means of a methanolic solution of potassium hydroxide (KOH 2 N) according to Ficarra et al. (2010), adding 100 μl of methyl nonadecanoate (C19:0) (Larodan Fine Chemicals AB, Malmö, Sweden) as internal standard.

The injection of the fatty acid methyl ester sample (1 μl) was performed in split mode with a splitting degree equal to 1:150, operating at a constant flow of 0.5 mLmin^{-1} of helium as carrier gas. The temperature of injector and detector was kept at 240 °C. The temperature program used for the analysis started from 150 °C, was maintained for 10 s, then increased to 240 °C, at a rate of 102 °Cmin^{-1}, and kept at this temperature for 2.5 min. The peaks of the fatty acids were recorded and integrated using Chrom-Card software (vers. 2.3.3, Thermo Electron Corporation, Rodano, Milano, Italy) and identified by comparison with the retention times of standard solutions with known quantities of various methyl esters (Supelco® 37 Component FAME mix and PUFA standard n.2, Animal Source, Supelco, Bellafonte, PA, USA). For quantification purposes, the response factor was calculated and the method of the internal standard was used. The amount of each FA in the sample is expressed as FA relative percentage with respect to the total amount of FAs.

As it is widely known in the research field regarding meat products (Alonso et al. 2009; Monziols et al. 2007; Wood et

al. 2003), the most abundant FAs in pig subcutaneous fat are palmitic (C16) and stearic (C18) acids as saturated FAs, and oleic (C18:1) and linoleic (C18:2) acids as unsaturated FAs. For this reason, in this work, only the calibration of these specific FAs has been faced. In addition, starting from the single FA relative percentages, three further parameters have been calculated, to be used as response variables for calibration aims, i.e., saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), and polyunsaturated fatty acids (PUFA). These parameters are often considered crucial when studying the overall chemical characteristics of swine meat and fat (Zamora-Rojas et al. 2013; Lo Fiego et al. 2010; Piasentier et al. 2009).

SFA, MUFA, and PUFA have been calculated as the sums of the percentages of the FAs belonging to the corresponding category:

$$\text{SFA} = (C_{10} + C_{12} + C_{14} + C_{16} + C_{17} + C_{18} + C_{20}) \ \%$$

2

$$\text{MUFA} = (C_{16:1} + C_{17:1} + C_{18:1} + C_{20:1}) \ \%$$

3

$$\text{PUFA} = (C_{18:2} + C_{18:3n-3} + C_{18:3n-6} + C_{20:3} + C_{20:4}) \ \%$$

4

Also in this case, for the fat aliquots belonging to 65 pig specimens, 158 GC analyses were executed (65 *In* layer plus 13 replications and 65 *Out* layer plus 15 replications, so as to replicate at least 20 % of the samples).

Data Processing and Analysis

Statistical Survey on IV and GC Data

The values of IV, C16, C18, C18:1, C18:2, SFA, MUFA, and PUFA were subjected to a statistical survey, considering both the whole sample set and *In* and *Out* samples separately. In particular, the mean, the range and the standard deviation (S.D.) have been calculated, in addition to the experimental root mean square error (RMSE_{Exp}), that furnishes

an estimate of the reproducibility of duplicate measurements made on similar samples. The RMSE$_{Exp}$ is calculated using Eq. 5 (IUPAC 1998):

$$\text{RMSE}_{Exp} = \sqrt{\frac{\sum \left(x_i{'} - x_i{''}\right)^2}{2m}}$$

5

where *xi'* and *xi''* are the results of replicate measurements for the *i*th sample and *m* is the number of paired values.

This statistical survey was firstly conducted in order to verify the expected differences between *In* and *Out* samples based on IV and FAs values and, secondly, to compare the uncertainty associated to the calibration models with that associated to the reference measurements.

PCA and Data Organization

The replicate measurements collected by IV and GC analyses were averaged. Then, the average IV and GC values were combined together into a unique dataset, named IV-GC dataset, composed by 130 independent samples (65 corresponding to class *Out* and 65 to class *In*) and 8 variables (IV, C16, C18, C18:1, C18:2, SFA, MUFA, and PUFA) to be used for multivariate data analysis. In a similar way, the replicate and repeated spectra acquired using each sampling tool (IS and FOP) on the separate *In* and *Out* layers have been averaged over the 65 single specimens. In this way, two new datasets composed by 130 mean spectra each (65 corresponding to class *Out* and 65 to class *In*) have been obtained, as described in Scheme 1. In Figure 1-S (supplementary material) the average spectra of *In* and *Out* samples are reported both for IS (Figure 1-S.a) and for FOP (Figure 1-S.b) datasets.

Therefore, on the whole, the averaging procedure of lab and spectral data led to the creation of three datasets, composed of 130 independent objects, where each object corresponds to a single fat sample. This procedure was necessary to match the number of spectra with the number of laboratory determinations.

Principal component analysis (PCA) (Bro and Smilde 2014) was then used as unsupervised exploratory technique for

both IV-GC dataset and FT-NIR datasets with the aim of detecting the presence of possible outliers. The outliers were identified as the samples lying outside the 99.7 % confidence limits in the Q residuals vs. Hotelling T^2 plot and were removed from the datasets.

For validation purposes, before performing PLS, the samples were split into a training set (TRN) and a test set (TST), using a random selection procedure.

PLS Calibration and Performance Parameters

PLS was applied as multivariate calibration method to predict the value of each IV-GC variable. The a priori choice of a specific preprocessing method of NIR spectra, irrespective of the matrix under examination, may not be the preferable choice (Rinnan et al. 2009). Therefore, while the response variables were individually autoscaled, the FT-NIR datasets were subjected to the following 24 signal preprocessing methods: none (N), meancentering (m), first order derivative (d1, Savitzky-Golay with 15 points filter and second order polynomial), second order derivative (d2, Savitzky-Golay with 15 points filter and second order polynomial), linear detrend (det1), quadratic detrend (det2), smoothing (S, Savitzky-Golay with 15 points filter), standard normal variate (SNV), and multiplicative scatter correction (MSC), that were tested both separately and in the following combinations: d1 + m, d2 + m, det1 + m, det2 + m, S + m, SNV + m, MSC + m, d1 + S + m, d2 + S + m, det1 + S + m, det2 + S + m, SNV + S + m, MSC + S + m, SNV + d1 + m, SNV + d2 + m.

The performance of the obtained PLS calibration models were expressed in terms of coefficient of determination (R^2), root mean square error (RMSE), and residual predictive deviation (RPD).

R^2 is particularly useful to compare directly models calculated on different response variables, since it does not depend on the scale of the Y variable. For each model, three R^2 values were calculated, i.e., one for the results in calibration of the training set (R^2_{Cal}), one for the results in cross-validation (R^2_{CV}), and one for the prediction of the external test set (R^2_{Pred}). While R^2_{Cal} corresponds to the squared value of the Pearson correlation coefficient (r) between the experimentally measured Y values and the corresponding values calculated by the calibration model, R^2_{CV} and R^2_{Pred} are defined by the following equation:

$$R^2 = 1 - (\text{PRESS/SS})$$

6

where PRESS is the prediction error sum of squares, defined as the sum of the squared differences between the experimental and the predicted y values, while SS is the sum of squares of the experimental Y values (of the training set and of the test set for R^2_{CV} and R^2_{Pred}, respectively).

Another useful parameter for the estimation of the predictive capability of the model is the root mean squared error, which reports the prediction error in the same units of the Y variable. Also for RMSE three values were calculated for each model, corresponding to the calibration of the training set (RMSEC), to the results in cross-validation (RMSECV), and to the prediction of the external test set (RMSEP). RMSE is defined as (Mevik and Cederkvist 2004):

$$\text{RMSE} = \sqrt{\frac{\Sigma\left(Y - \hat{Y}\right)^2}{n}}$$

7

where Y refers to the experimentally measured values, $\hat{Y}$ are the corresponding values calculated (for RMSEC) or predicted (for RMSECV and RMSEP) by the model, and n is the number of samples of the training set (for RMSEC and RMSECV) or of the test set (for RMSEP). In particular, RMSECV (calculated using a random subsets cross-validation method with 5 deletion groups and 20 iterations) was used both to select the best signal preprocessing method for the calibration of each response variable, and to define the optimal number of latent variables (LVs) of each PLS model (up to a maximum value of 12 LVs).

Moreover, the values of residual predictive deviation (RPD) were also calculated for each model, referring to the prediction of the test set samples. RPD is defined as the standard deviation of the Y test set values divided by the standard error of prediction, SEP (Wu et al. 2010), where the SEP value is derived by RMSEP by means of the following equation:

8

$$SEP = \sqrt{\frac{(RMSEP^2 - BIAS^2) \cdot n}{n - 1}}$$

where BIAS is the arithmetic mean of the prediction errors for the test set samples.

Since the RPD is defined as the standard deviation of reference validation data divided by the prediction error, it follows that if the RPD value is relatively small; the obtained NIR calibration model is not robust, while if the RPD value is relatively high, the obtained model has greater predictive ability. According to (Cozzolino et al. 2004), an RPD greater than three could be considered very good for prediction purposes.

All the parameters defined in this Section were employed to estimate the performance of the final calibration models, paying particular attention to the prediction of the external test set, evaluated by means of R^2_{Pred}, RMSEP, and RPD.

Variable Selection

Variable selection methods are often used on NIR data to reduce the computational load and to obtain more robust models in prediction (Gosselin et al. 2010). To these aims, in this work a two-step procedure has been used (Scheme 2).

Scheme 2

Representation of the two-step procedure used for variable selection

---

Phase 1: building of the iPLS models (*forward* mode)

Phase 2: "selection of the selection"

New PLS models are built considering:
- most frequently selected spectral regions (e.g. 8 times)
- second most frequently selected spectral regions (e.g. 7 times)
- ...
  so on till the minimum RMSECV
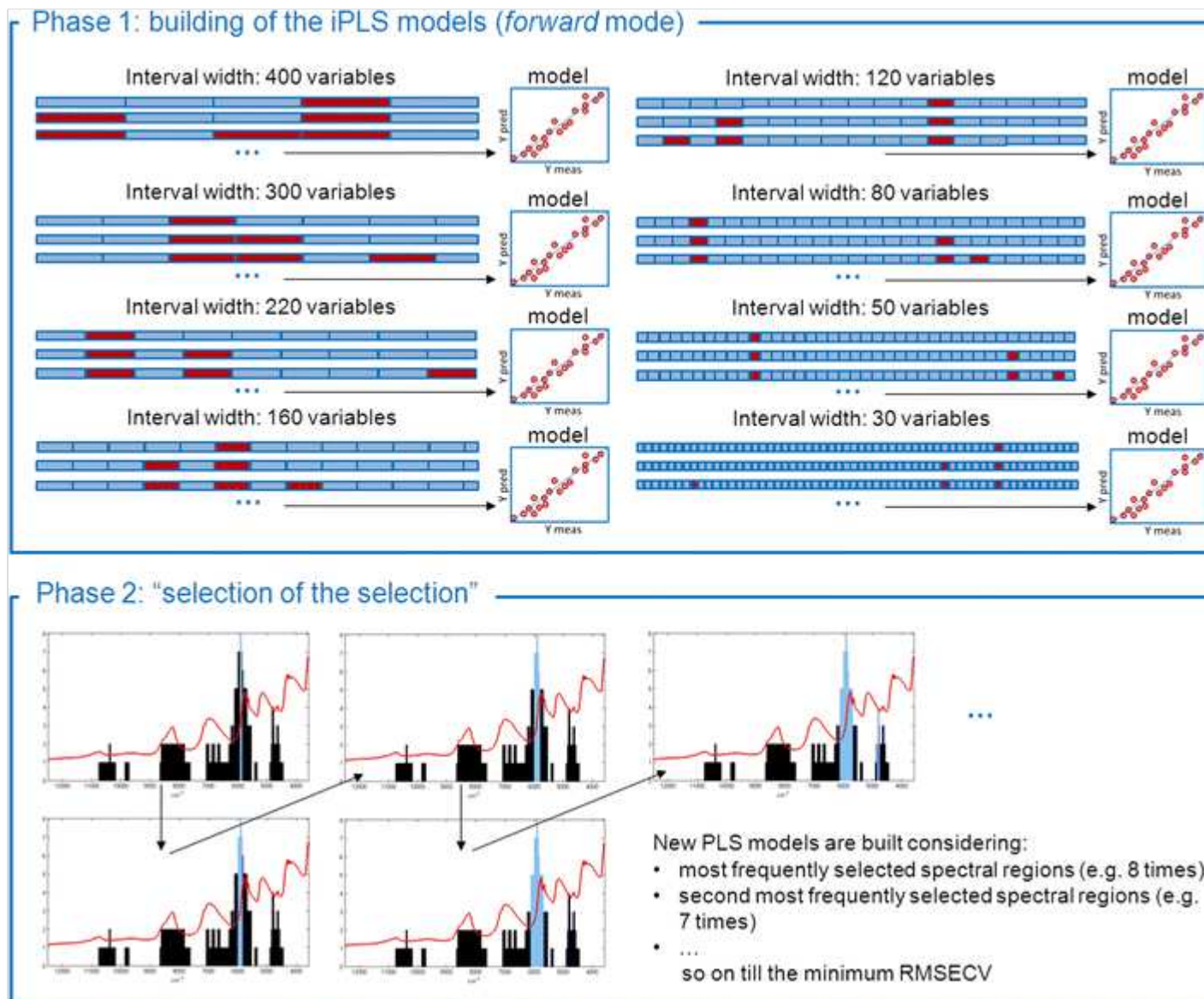
In the first phase, for each response variable Y, the interval-PLS (iPLS) algorithm was applied. Basically, it consists in dividing the whole spectral range in a user-defined number of intervals of equal width, then in selecting the intervals

most useful for calibration by an iterative procedure (Nørgaard et al. 2000; Leardi and Nørgaard 2004). In particular, iPLS was applied in the *forward* mode, i.e., the intervals were iteratively added until a significant decrease of RMSECV is no longer observed (Xiaobo et al. 2010). For both the FT-NIR datasets, eight different interval widths were considered for splitting the whole spectral range, consisting of 400, 300, 220, 160, 120, 80, 50, and 30 spectral variables. To model each response variable by iPLS, the FT-NIR datasets were pretreated using the preprocessing methods that gave the best results for the corresponding PLS model. Also in this case, a random subsets cross-validation (5 deletion groups, 20 iterations) was performed.

In the second phase, after obtaining all the iPLS models as described above, we have chosen to focus on the most frequently selected variables in the models obtained considering different interval widths. To do this, for each response variable, further PLS models were calculated according to the following procedure: first only the most frequently selected spectral regions were considered (for instance, those corresponding to the spectral variables retained by all the eight iPLS models resulting from the eight different interval width values); subsequently, the second most frequently selected regions (i.e., seven times) were also added, and so on until including all the spectral variables that have been retained at least once. Finally, the model with the minimum RMSECV value was retained, and the corresponding selected regions were considered as the most informative ones from the statistical and chemical points of view.

PCA, PLS, and iPLS models were elaborated by means of PLS Toolbox ver. 7.8.2 (Eigenvector Research Inc) for Matlab© platform 7.11 R2010b (The MathWorks Inc), also using some Matlab functions written *ad hoc* for the final selection of the optimal spectral regions based on the results of iPLS.

Results and Discussion

IV and GC Data

Statistical Survey Results

The results of the statistical survey on IV values and GC data are reported in Table 1. As expected, and in agreement

with the literature (Minelli et al. 2013; Alonso et al. 2009), the samples taken from the *In* layer are mainly characterized by higher amounts of C16 and C18 and, consequently, of SFA, while the ones taken from the *Out* layer are more rich in C18:1 and C18:2 and, consequently, show higher values for MUFA, PUFA, and IV.

Table 1

Descriptive statistics of the samples belonging to the different fat layers

	Min.	Max.	Mean	S.D.	$RMSE_{Exp}$
All samples					
C16	21.63	29.67	25.05	1.45	0.44
C18	9.46	19.55	14.38	1.97	0.31
C18:1	37.23	47.97	43.17	2.32	0.48
C18:2	6.45	17.16	10.01	1.86	0.20
SFA	35.43	48.33	41.87	2.97	0.50
MUFA	40.06	52.06	46.79	2.45	0.50
PUFA	7.41	19.09	11.34	2.02	0.24
IV	53.83	73.20	63.32	4.13	1.75
Out samples					
C16	21.63	27.15	24.35	1.21	0.48
C18	9.46	16.38	13.10	1.41	0.36
C18:1	40.09	47.97	44.02	2.04	0.55

FA values are expressed as % of total FAs analyzed. IV is expressed as g I_2/100 g fat

	Min.	Max.	Mean	S.D.	$RMSE_{Exp}$
C18:2	8.56	17.16	10.79	1.65	0.19
SFA	35.43	43.96	39.91	2.08	0.50
MUFA	43.33	52.06	47.87	2.06	0.50
PUFA	9.64	19.09	12.22	1.80	0.26
IV	60.20	73.20	65.93	3.03	1.29
In samples					
C16	23.33	29.67	25.76	1.33	0.39
C18	12.16	19.55	15.69	1.56	0.24
C18:1	37.23	46.71	42.30	2.28	0.39
C18:2	6.46	15.56	9.21	1.73	0.21
SFA	38.62	48.33	43.88	2.35	0.50
MUFA	40.06	50.60	45.67	2.33	0.50
PUFA	7.41	16.75	10.45	1.84	0.21
IV	53.83	70.17	60.64	3.32	2.17
FA values are expressed as % of total FAs analyzed. IV is expressed as g I_2/100 g fat					

The results obtained for *In* and *Out* samples have been compared by means of statistical hypothesis tests. Firstly, for each variable, the *F* test was used to compare the variances of *In* and *Out* samples: only for C16 the variances resulted statistically different at the 95 % confidence level. Then, the two-tailed *t* test was performed in order to compare the mean *In* and *Out* values of each parameter. In all cases, the calculated *t* values led to the rejection of the null hypothesis

(i.e., equivalence of *In* and *Out* samples) at the 95 % confidence level. Despite this observation, we preferred to consider the entire dataset for the construction of calibration models. This was done in order to obtain more robust models of general use, also considering that the values of the measured properties vary with continuity over the entire range covered by the two categories of samples.

Table 1 reports also the values of $RMSE_{Exp}$, representing the degree of variability of replicate measurements. For each measured parameter, the $RMSE_{Exp}$ value resulted lower than the corresponding standard deviation, confirming the reproducibility of these measurements. Among all the measured parameters, the IV is the one affected by the lowest reproducibility, especially for the *In* samples.

Explorative Data Analysis

PCA was performed on the autoscaled IV-GC dataset (3 PCs, 97.2 % cumulative variance). In the PC1-PC2 biplot, shown in Fig. 1, the samples belonging to *Out* and *In* classes are somehow separated along PC1, that seems to contain the information about the different chemical composition which characterize the two subcutaneous fat layers.

Fig. 1

PC1-PC2 biplot obtained by performing PCA on the IV-GC dataset. Class *Out*: *circles*; class *In*: *diamonds*
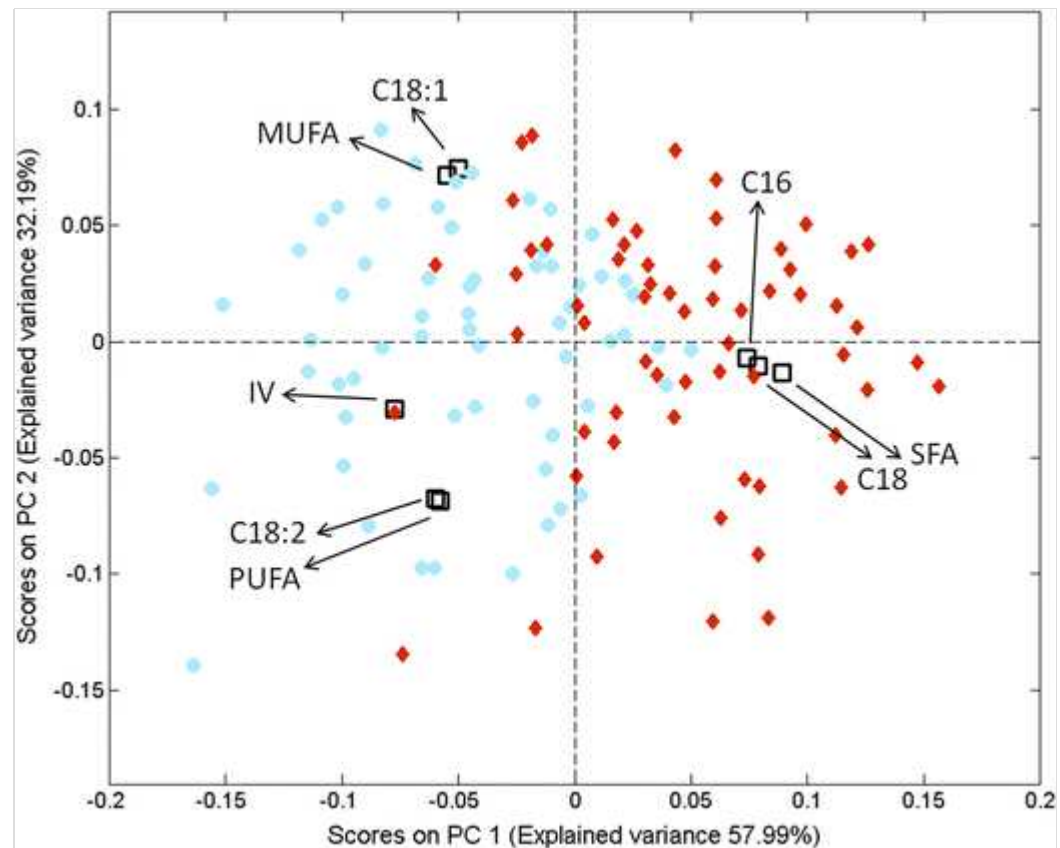
In agreement with what was observed using classical statistical tools, PCA also highlights that the *Out* samples generally present higher values for the variables C18:1, C18:2, PUFA, MUFA, and IV, located at high negative values of PC1 in the PCs space. Conversely, the *In* samples show lower values for the variables C16, C18, and SFA, located at positive values of PC1.

The biplot in Fig. 1 also gives clear indications about the correlations among the *Y* variables. Looking at the data, correlations are observed in the *Y* variables based on the number of double bonds: the amounts of saturated fatty acids are correlated with each other and with SFA, linoleic acid is strongly correlated with IV and PUFA, while oleic acid is strongly correlated with MUFA.

PCA allowed to highlight the presence of four outlier samples, lying outside the 99.7 % confidence limits in the Q residuals vs. Hotelling T^2 plot.

FT-NIR Data

Explorative Data Analysis on FT-NIR Datasets

A preliminary screening by PCA was also done on the mean-centered spectra of IS and FOP datasets; for both datasets, three PCs were selected, explaining 97.1 and 98.9 % of the cumulative variance, respectively. The PC1-PC2 score plot obtained for the IS dataset is reported in Fig. 2. Similarly to what was observed along PC1 for the IV-GC dataset, the score plot of the IS dataset shows that the two subcutaneous fat layers are separated each other along PC2, though the clusters are slightly superimposed. Analogous results (not shown) have been obtained for the FOP dataset.

Fig. 2

PC1-PC2 scores plot obtained for the IS dataset. Class *Out*: *circles*; class *In*: *diamonds*
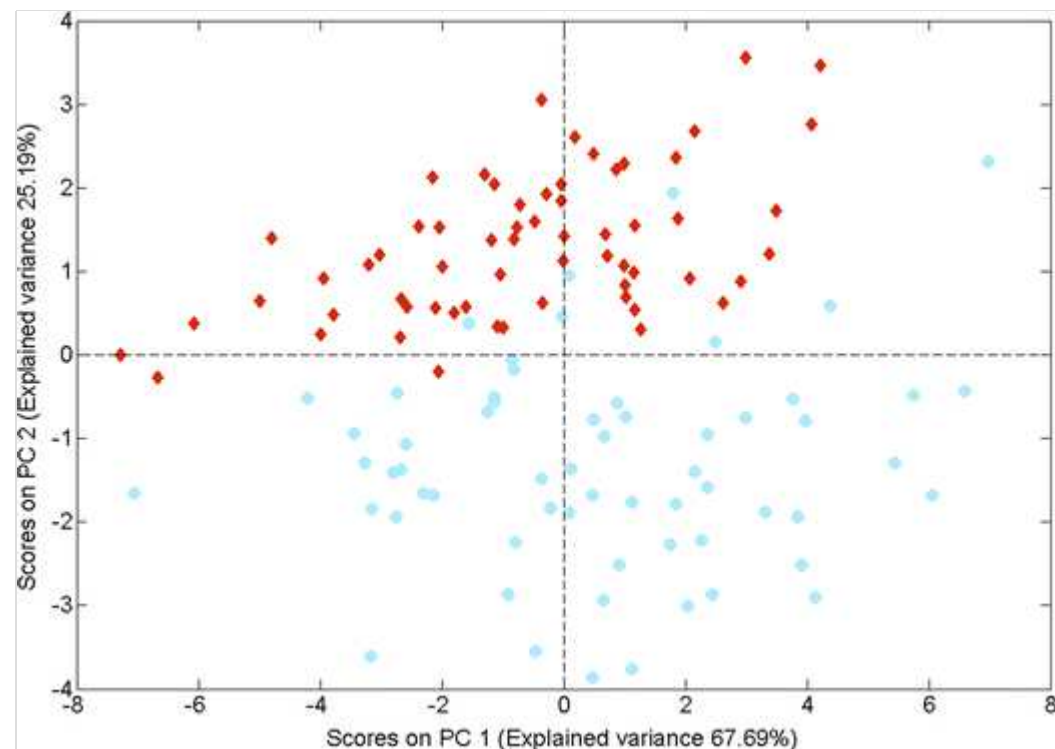
In both models, the presence of outlier samples have been evidenced (one outlier in the IS dataset and two outliers in the FOP dataset). Considering also the outliers found in the IV-GC dataset, on the whole seven samples have been removed from all the datasets. Of the remaining 123 samples (61 *Out* and 62 *In* samples), 88 were randomly assigned to the training set and 35 to the test set. Basic statistics calculated on the values of the *Y* variables separately for the training set and for the test set objects have been included in the supplementary material.

PLS Calibration Models

Table 2 reports an overview on the best PLS models obtained for each *Y* variable using both IS and FOP datasets, including selected signal preprocessing method, model dimensionality, and the performance parameters (R^2, RMSE, and RPD) described in Section 2.3.

Table 2

Results of the best PLS models

Y variable	Processing	No. of LVs	RMSEC	RMSECV	RMSEP	RPD	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
IS dataset									
C16	S + m	2	1.13	1.18	1.12	1.31	0.32	0.25	0.42
C18	det1 + S + m	8	0.90	1.09	1.57	1.55	0.75	0.64	0.45
C18:1	det1	12	1.32	2.06	1.39	1.83	0.68	0.22	0.63
C18:2	S + m	12	0.87	1.19	0.93	2.01	0.80	0.62	0.74
SFA	d2 + S + m	7	1.11	1.68	1.38	2.33	0.84	0.64	0.79
MUFA	d2 + S + m	10	1.03	1.91	1.20	2.17	0.82	0.37	0.77
PUFA	S + m	12	0.90	1.23	1.05	1.95	0.81	0.65	0.73
IV	d2 + S + m	10	1.10	2.03	1.76	2.24	0.93	0.76	0.80
FOP dataset									
C16	det2 + S + m	5	1.02	1.13	1.20	1.24	0.45	0.32	0.33
C18	det1 + S + m	7	0.95	1.10	1.37	1.77	0.72	0.62	0.59
C18:1	MSC + S + m	12	1.14	1.93	1.69	1.39	0.76	0.32	0.45
C18:2	det1 + S + m	9	0.90	1.22	0.95	2.04	0.78	0.60	0.72
SFA	det1 + m	7	1.32	1.68	1.24	2.65	0.78	0.64	0.83
MUFA	MSC + S + m	12	1.15	1.93	1.78	1.47	0.77	0.36	0.49

RMSE are expressed in the same units as the Y variables (% of total FAs analyzed for each FA and g I_2/100 g of fat for IV)

Y variable	Processing	No. of LVs	RMSEC	RMSECV	RMSEP	RPD	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
PUFA	MSC + S + m	10	0.86	1.28	0.85	2.39	0.83	0.62	0.82
IV	det2 + S + m	9	1.51	2.00	1.67	2.38	0.87	0.77	0.82

RMSE are expressed in the same units as the Y variables (% of total FAs analyzed for each FA and g I$_2$/100 g of fat for IV)

In general, the predictive ability of the models was poor for the variables C16, C18, and C18:1; for the FOP dataset also MUFA was not predicted satisfactorily. The variables that were better predicted are IV, PUFA, SFA, and C18:2. Acceptable results in prediction of the test set were obtained also for MUFA when using the IS dataset (RMSEP = 1.20, RPD = 2.17, R^2_{Pred} = 0.77); however, in this case the performance in cross-validation resulted very scarce (RMSECV = 1.91, R^2_{CV} = 0.37), suggesting the low stability of this calibration model.

Only for IV the RMSEP values of the calibration models are comparable with the RMSE$_{Exp}$ values of the reference data (Table 1), which means that there is a little margin for further model improvements. Considering the single FAs, the RMSEP values vary in the range 0.93–1.57 for IS and 0.95–1.78 for FOP. The FA that is estimated with the lowest RMSEP (and higher R^2_{Pred}) values for both the datasets is C18:2 that is also the parameter with the lower RMSE$_{Exp}$ value.

Interestingly, the comparison between RMSEP and RMSECV (and between R^2_{Pred} and R^2_{CV}) shows that in most of the cases the performance in prediction of the test set samples is better than the performance in cross-validation, notwithstanding the random selection of the external validation samples.

In general, recalling that the higher is the RPD value, the greater is predictive ability of the corresponding model, and that good performances in prediction are observed when RPD is greater than 3 (Cozzolino et al. 2004), the RPD values of Table 2 suggest that the present PLS models are suitable for rough estimates or at least for screening purposes.

Other authors in the literature reported calibration models with higher performances in prediction. For instance, Müller

and Scheeder (2008) have obtained good calibration models: they predict IV with an $R^2_{Pred} = 0.98$ and RMSEP = 0.5, therefore such a calibration is adequate for any application. Also Fernández-Cabanás et al. (2007) have obtained good calibration models, with RMSEP always lower than 1 % for the analyzed fatty acids. However, it must be underlined that the effectiveness of fat characterization by FT-NIR spectroscopy greatly varies depending on sample pretreatment, as proved by Zamora-Rojas et al. (2013). They collected the spectral datasets both using a fiber optic probe on intact adipose tissue and in transmission on melted fat. Their results highlighted that the models based on intact samples show prediction errors always higher (even double or triple) than those obtained on melted fat. Therefore, recalling that in the present work we were focused on "direct" fat characterization (i.e., without performing sample pretreatment steps like melting or homogenization), a fair comparison should be made considering results obtained using similar methods for sample preparation and signal acquisition. Considering the use of diffuse reflectance, we have obtained results similar to those of Ripoche and Guillard (2001). As for the use of fiber optic probe, our results are definitely comparable, in terms of predictive ability of the external validation samples, to those reported by Pérez-Juan et al. (2010).

According to the results of Table 2, the performances of IS and FOP are comparable, since none of the two devices has given systematically better results. Therefore, since the results are not influential as to the choice between the two sampling tools, we believe that the use of fiber optic probe is preferable. In fact, with respect to the integrating sphere, the fiber optic probe can be more easily implemented for fast measurements performed *in loco*.

The results in Table 2 also suggest that the use of a unique signal preprocessing method is not able to lead to the optimal solution for all the different *Y* variables. Similar outcomes were obtained by Fernández-Cabanás et al. (2007), who pretreated fat spectra in many different ways and observed that there is not a preprocessing which performs systematically better than the others. In any case, from our results we can draw some general remarks. In almost all cases, the spectra have to be mean-centered and also the smoothing operation has proved to be very often useful. For the spectra acquired with the fiber optic probe, the derivative never resulted an effective preprocessing, while detrend worked well.

With regard to the model dimensionality, it can be observed that the variables C16, C18, and SFA, corresponding to

saturated fatty acids, have generally required a smaller number of latent variables for building the corresponding models, although with different outcomes.

Variable Selection

Phase One   As explained in the Methods section, iPLS has been used in the first phase of variable selection. Table 3 reports the details of the best iPLS models (selected in cross-validation) obtained after the first step of the variable selection procedure; it can be observed that there is not an optimum interval width for all the *Y* variables and that often the prediction results are not improved compared to the PLS models in Table 2.

Table 3

Results of the best iPLS models obtained in the first phase of variable selection

| *Y* variable	Interval width	No. of selected variables	No. of LVs	RMSEC	RMSECV	RMSEP	RPD	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
IS dataset										
C16	50	200	11	0.74	0.97	1.43	1.05	0.71	0.50	0.06
C18	50	100	8	0.76	0.91	0.80	2.74	0.82	0.75	0.86
C18:1	30	210	12	0.72	1.34	1.51	1.55	0.91	0.67	0.56
C18:2	300	1200	12	0.52	0.95	0.90	2.03	0.93	0.76	0.75
SFA	400	1200	8	0.95	1.39	1.34	2.25	0.89	0.75	0.80
MUFA	80	400	11	0.95	1.43	1.38	1.82	0.84	0.64	0.69
PUFA	400	800	11	0.74	1.00	0.83	2.43	0.88	0.77	0.83
IV	50	200	7	1.56	1.83	1.96	2.03	0.86	0.80	0.75

RMSE are expressed in the same units as the *Y* variables (% of total FAs analyzed for each FA and g I_2/100 g of fat for IV)

Y variable	Interval width	No. of selected variables	No. of LVs	RMSEC	RMSECV	RMSEP	RPD	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
FOP dataset										
C16	160	320	12	0.54	0.79	1.03	1.48	0.84	0.67	0.51
C18	80	80	4	0.88	0.94	1.02	2.26	0.76	0.73	0.77
C18:1	50	200	11	1.05	1.45	1.58	1.45	0.80	0.62	0.52
C18:2	400	400	7	0.78	1.00	0.78	2.33	0.84	0.73	0.82
SFA	220	220	4	1.26	1.39	1.31	2.51	0.80	0.76	0.81
MUFA	80	240	10	1.07	1.32	1.39	1.83	0.80	0.70	0.69
PUFA	30	150	11	0.70	0.94	1.03	1.98	0.89	0.80	0.74
IV	80	160	5	1.68	1.81	1.92	2.05	0.83	0.81	0.76
RMSE are expressed in the same units as the Y variables (% of total FAs analyzed for each FA and g I$_2$/100 g of fat for IV)										

Figure 3 shows the results for all the models obtained by considering different interval widths (from 400 to 30 spectral variables), calculated using the signals acquired by means of the fiber optic probe. In particular, for each modeled variable the corresponding histogram represents the frequency of selection of each spectral region. For example, the top left plot in Fig. 3 (variable C16) shows that seven out of eight models have led to the selection of the region 6010–5850 cm^{-1}; this means that the spectral information related to the presence of palmitic acid seems to be located particularly in that region. In the same histogram, six out of eight models have led to the selection of the region 6160–5840 cm^{-1} (all the wavenumbers except for a small spectral window), five out of eight models have led to the selection of the region 6210–5780 cm^{-1}, and so on. Hence, this sort of representation shows the degree of convergence of the various models computed for each variable.

Fig. 3

Distribution histogram of the spectral regions selected by iPLS considering different interval widths: results for the FOP dataset
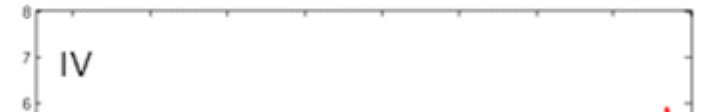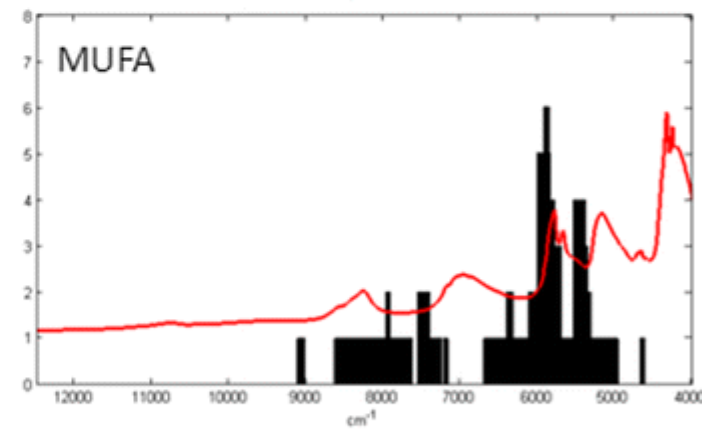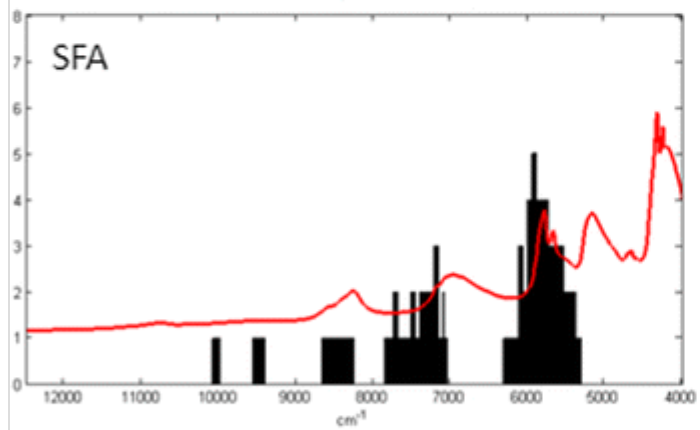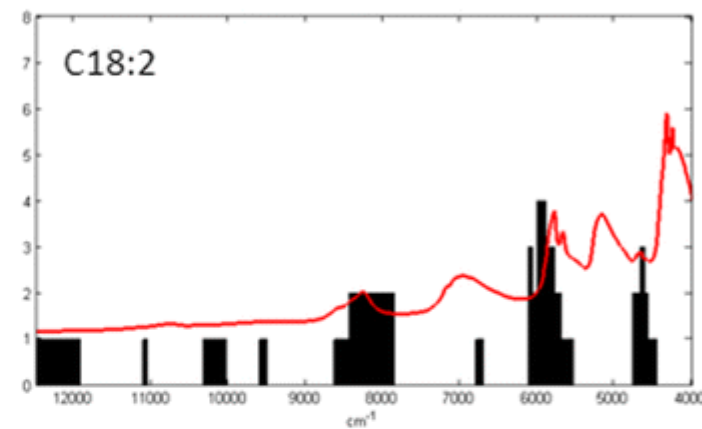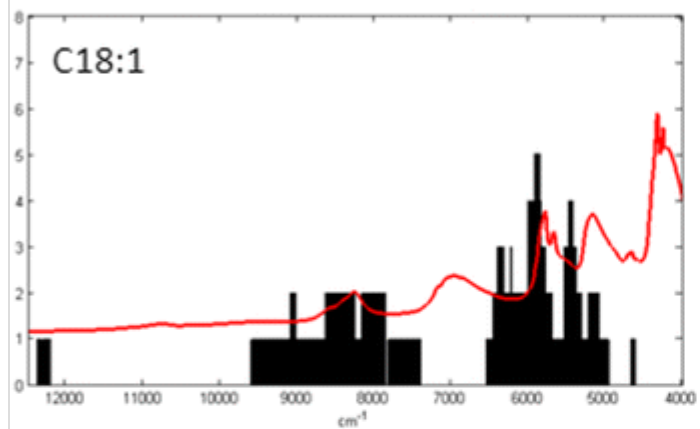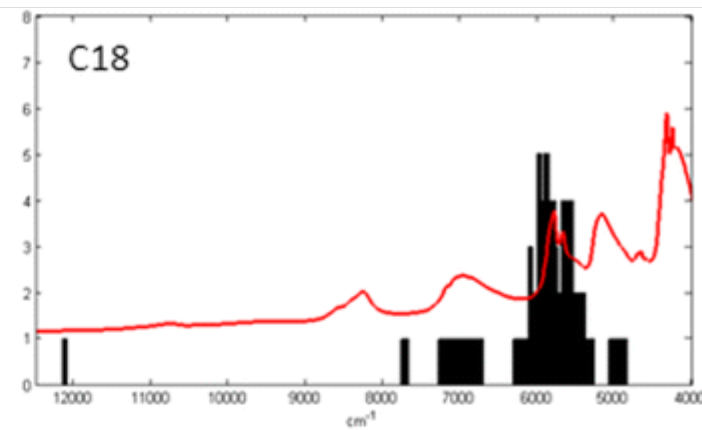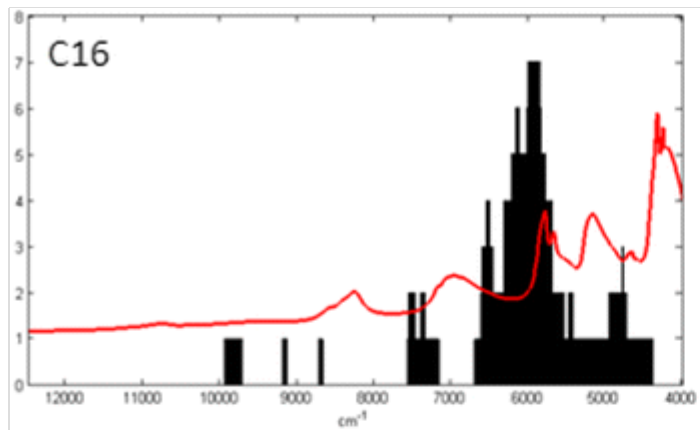
It can be observed that, for all the modeled chemical variables, the most frequently selected region is the one corresponding to the spectral bands centered at 5785 and 5670 cm^{-1}, in which the absorption bands related to the C–H stretching first overtone of CH_2 and CH_3 groups are located (Sørensen et al. 2012; González-Martin et al. 2005; Pérez-Juan et al. 2010; Li et al. 1999).

For C16, C18, and SFA the region approximately in the range 7400–6900 cm^{-1} is generally selected, where a peak at about 7175 cm^{-1}, partly superimposed to the absorption band of water, can be seen; this peak is attributed to the C–H combination band of CH_2 (Shenk et al. 2008). For C18:1 and MUFA, a different region was frequently selected, i.e., the one in the 5620–5460 cm^{-1} range, where a water related band (O–H combination) is located, maybe due to the tissue structure (Shenk et al. 2008). At variance, for C18:2, PUFA, and IV the recurring selection of the region in the interval at about 8600–8100 cm^{-1} is observed, in addition to a region at lower wavenumbers in which two sharper peaks centered at 4710 and 4660 cm^{-1} are observed. Many works attribute the bands present in the 8600–8100 cm^{-1} region to the C–H stretching second overtone (Ripoche and Guillard 2001; Pérez-Juan et al. 2010; Li et al. 1999). In the same spectral region, Cox et al. (2000) observed the increase of the peak at 8550 cm^{-1}, according to the increase of the iodine value of different oil samples. Finally, in the work by Li et al. (1999) the bands near 8475 and 4675 cm^{-1} are attributed to strong —CH=CH—(*cis* double bond) overtone and combination band vibrations, respectively. Interestingly, most of the bands selected for calibrating IV were the same selected by iPLS in the work by Sørensen et al. (2012).

Summarizing, for chemically related variables a good convergence of the models is observed. These considerations, together with those made about the correlations among the *Y* variables discussed in the IV and GC data section, may suggest that these models are mainly based on "family-specific" features, rather than on features specific for the single fatty acids. Although a detailed discussion on this issue is beyond the scope of the work, we point out the recent work by Eskildsen et al. (2014), which deals with this topic and opens an interesting field of investigation. Basically, in this work the authors state that predicted concentrations of individual fatty acids in milk rely on covariance structures with total fat content rather than absorption bands directly associated with individual fatty acids, i.e., prediction results are good if calibration and test sets have the same covariance structure between individual fatty acids and total fat content.

Similar results to those reported in Fig. 3 were obtained also for the IS dataset (Figure 2-S in the supplementary material). The same considerations made above are still valid, although the convergence of the selection on the same regions for the variables that are chemically related is a little less noticeable.

Phase Two   During the phase which we have called "selection of the selection", new PLS models were iteratively built starting from the spectral regions that were more frequently selected in the previous iPLS models. It has to be underlined that the main concern of the present work is not to merely compare the performance of our models with those of other works dealing with the same fat properties, but rather to compare the models before and after variable selection. The results of the comprehensive models that were obtained are reported in Table 4.

Table 4

Model performances after applying the two-step variable selection procedure

Y variable	No. of selected variables	No. of LVs	RMSEC	RMSECV	RMSEP	RPD	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
IS dataset									
C16	250	9	0.84	1.01	1.13	1.36	0.62	0.45	0.41
C18	120	7	0.80	0.90	0.95	2.48	0.80	0.75	0.80
C18:1	250	12	0.76	1.53	1.22	1.87	0.90	0.57	0.71
C18:2	280	9	0.76	0.97	0.72	2.52	0.84	0.75	0.84
SFA	60	7	1.25	1.39	1.24	2.45	0.80	0.75	0.83
MUFA	180	9	1.12	1.46	1.50	1.67	0.78	0.63	0.64
PUFA	60	4	0.91	0.98	0.78	2.58	0.81	0.78	0.85

RMSE are expressed in the same units as the *Y* variables (% of total FAs analyzed for each FA and g I$_2$/100 g of fat for IV)

Y variable	No. of selected variables	No. of LVs	RMSEC	RMSECV	RMSEP	RPD	R^2_{Cal}	R^2_{CV}	R^2_{Pred}
IV	1100	8	1.54	1.90	1.70	2.38	0.86	0.79	0.81
FOP dataset									
C16	340	12	0.57	0.80	1.09	1.40	0.83	0.66	0.45
C18	70	5	0.90	0.97	0.92	2.45	0.75	0.71	0.81
C18:1	260	12	1.09	1.47	1.37	1.67	0.78	0.61	0.64
C18:2	60	7	0.92	1.08	0.75	2.44	0.77	0.69	0.83
SFA	130	5	1.33	1.46	1.16	2.62	0.77	0.73	0.85
MUFA	90	7	1.29	1.44	1.36	1.84	0.71	0.64	0.71
PUFA	620	8	0.89	1.11	0.81	2.49	0.82	0.72	0.84
IV	540	6	1.69	1.88	1.73	2.30	0.83	0.79	0.81
RMSE are expressed in the same units as the Y variables (% of total FAs analyzed for each FA and g I$_2$/100 g of fat for IV)									

The first aspect that is worth to comment is the generally enhanced predictive ability of the models obtained at the end of the variable selection procedure, compared with that of PLS and iPLS models reported in Tables 2 and 3, respectively. On the whole, the best performances were obtained for the prediction of C18:2, SFA, and PUFA.

In particular, the predictive ability of the models related to the response variable C16 is still poor, while for the variables SFA and IV, which were already predicted fairly successfully, the R^2_{Pred}, RMSEP, and RPD values are substantially unchanged. Actually, a similar result for the iodine value was expected, since the prediction error of the PLS models before variable selection was already comparable with the $RMSE_{Exp}$ of the reference data.

Slight improvements were obtained for the prediction of PUFA using the FOP dataset, where the results of the PLS

model calculated on the whole signal was already satisfactory (RMSEP decreased from 0.85 to 0.81, R^2_{Pred} increased from 0.82 to 0.84, and RPD increased from 2.39 to 2.49); conversely, variable selection led to a significant decrease of the prediction error of PUFA values when using the IS dataset: RMSEP decreased from 1.05 to 0.78, R^2_{Pred} increased from 0.73 to 0.85, and RPD increased from 1.95 to 2.58.

With regard to the variable MUFA, the results showed a discordant trend for the two spectral datasets acquired with the different tools. As for the IS, the model performance got worse after variable selection: notwithstanding the expected increase of the performance in cross-validation (due to the fact that feature selection is driven by the RMSECV values), R^2_{Pred} decreased from 0.77 to 0.64, RMSEP increased from 1.20 to 1.50 and RPD decreased from 2.17 to 1.67. Conversely, variable selection improved the predictive performance of the model based on FOP spectra: R^2_{Pred} increased from 0.49 to 0.71, RMSEP decreased from 1.78 to 1.36 and RPD increased from 1.47 to 1.84.

As far as the single FAs are concerned, the two-step variable selection allowed to increase the performance in prediction in particular for C18 and C18:2, which for the IS dataset reached R^2_{Pred} values equal to 0.80 and 0.84, respectively, while for the FOP dataset the R^2_{Pred} values resulted equal to 0.81 and 0.83, respectively. The corresponding RPD values resulted equal to 2.48 and 2.52 for the IS dataset and to 2.45 and 2.44 for the FOP dataset. The limited number of informative wavelengths that were selected when using the FOP dataset (70 for C18 and 60 for C18:2) suggests that these final models could constitute a starting point for the implementation of a hand-held device for the screening of these FAs in fat samples.

Concerning the number of latent variables used to build the models, the dimensionality is generally decreased or at least it remained unchanged after variable selection, except for C16 for which, however, the models are not good. Compared to the original number of spectral variables (more than 4000), some final models are very parsimonious (less than 100 variables), while still providing good performances. Even in cases where some hundreds of variables have been selected, the process of variable selection is still convenient, since it allows to obtain more parsimonious models without compromising at all their predictive ability.

In Fig. 4, for both IS (Fig. 4a) and for FOP (Fig. 4b), the variables selected for each comprehensive model are reported together with the corresponding average spectrum. For all the models, the variables corresponding to the C–H stretching first overtone have been selected as informative for the calibration of fat related quantities. Again, there is some consistency in the models related to response variables with chemically similar meaning. For example, the variables around 8500 cm^{-1}, corresponding to the—CH = CH—overtone (Li et al. 1999), have been selected for the prediction of IV and C18:2 for the IS dataset and for the prediction of IV and PUFA for the FOP dataset.

Fig. 4

Representation of the variables selected for the different properties by applying the two-step variable selection procedure to the IS (**a**) and to the FOP (**b**) datasets. In each subplot the corresponding average spectrum is reported as reference
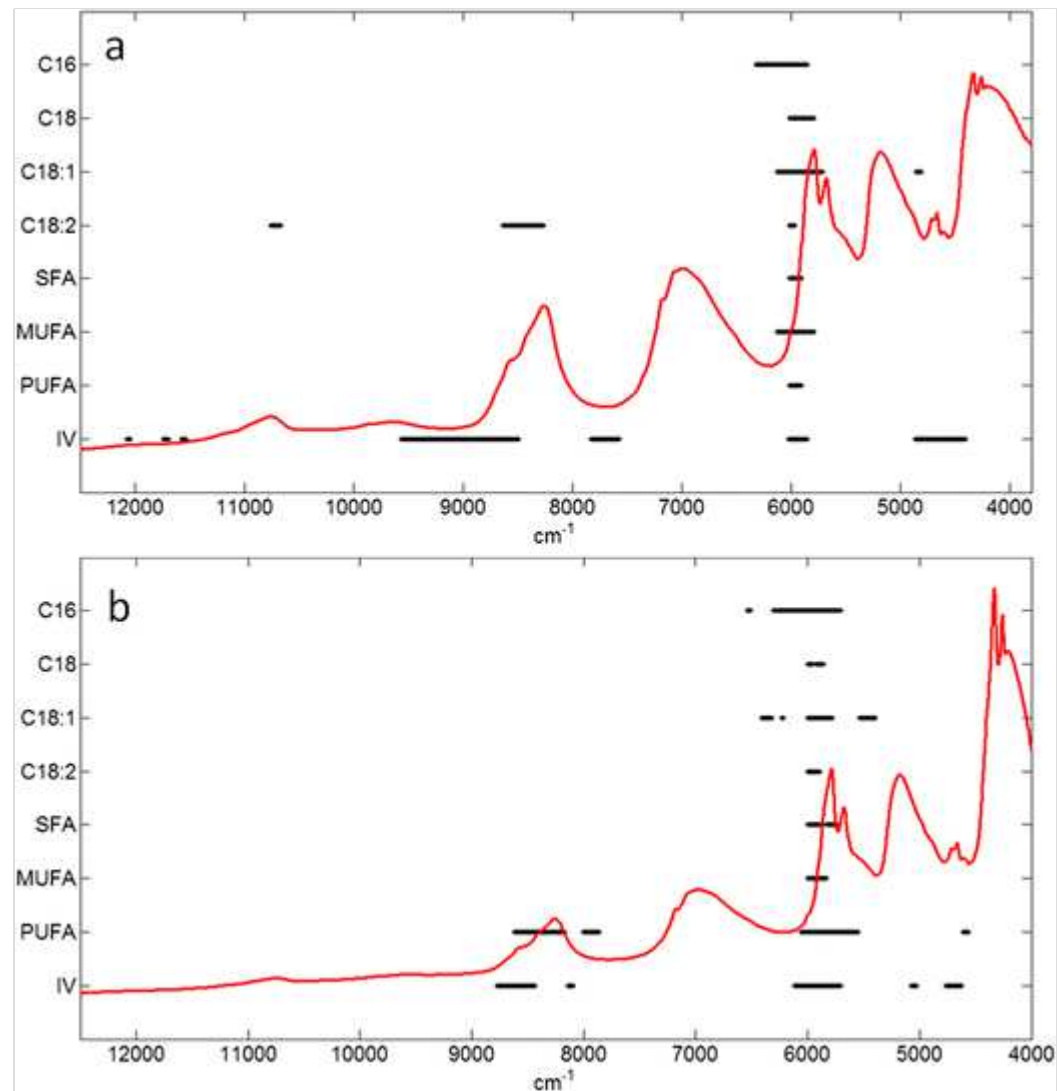
Conclusions

The results of the multivariate calibration models confirm that FT-NIR is able to predict IV and FA on intact pig fat samples, although the correlation coefficient values obtained in prediction are not higher than 0.85 and the RPD values

are always lower than 3. These findings suggest that NIR spectroscopy is more suitable to be used as a screening method, than to fully replace the time-consuming wet analyses currently used in the industry. As suggested by Bekiaris et al. (2015), the use of larger datasets could result in the decrease of RMSE and, consequently, in the increase of RPD, allowing further improvements of robustness, predictive power, and accuracy of the models.

Moreover, we found that IS and FOP datasets gave models with very similar performances. As a consequence, the use of FOP as sampling tool is recommended, since its on-line implementation is certainly more feasible.

The approach that we proposed in this work for variable selection is more effective than ordinary interval-PLS, that requires choosing a specific interval width. This two-step procedure has proved to be a beneficial ploy to condense the results of individual iPLS models obtained considering different interval widths in a comprehensive—and generally best performing—model.

The wavelengths selected as relevant for calibration purposes correspond to the C–H stretching first overtone for all the Y variables and to the—CH = CH—overtone for the prediction of IV and other compounds belonging to the group of polyunsaturated fatty acids. The selection of such a limited number of relevant wavelengths is favorable for the possible construction of a NIR-based low cost device for the screening of fat composition. On the other hand, since the selected variables are almost the same for the different Y variables, it follows that these absorption bands are not specific for each given fatty acid, but good predictions probably rely on covariance structures among individual fatty acids and total fat content. Hence, the study of the interdependence characteristics of the predictions of highly collinear Y variables is an interesting issue for future studies.

Compliance with Ethical Standards

All experimental procedures were carried out in accordance with UE and National legislation for the care and use of animals. This article does not contain any studies with human participants performed by any of the authors.

Conflict of Interest   Giorgia Foca declares that she has no conflict of interest. Carlotta Ferrari declares that she has no conflict of interest. Alessandro Ulrici declares that he has no conflict of interest. Maria Cristina Ielo declares that she has no conflict of interest. Giovanna Minelli declares that she has no conflict of interest. Domenico Pietro Lo Fiego declares that he has no conflict of interest.

Informed Consent   Not applicable.

Electronic supplementary material

Below is the link to the electronic supplementary material.

Figure 1-S

(DOCX 413 kb)

Figure 2-S

(DOCX 696 kb)

Table 1-S

(DOCX 16 kb)

References

Adewale P, Mba O, Dumont MJ, Ngadi M, Cocciardi R (2014) Determination of the iodine value and the free fatty acid content of waste animal fat blends using FT-NIR. Vib Spectrosc 72:72–78. doi: 10.1016/j.vibspec.2014.02.016

Afseth NK, Martens H, Randby A, Gidskehaug L, Narum B, Jørgensen K, Lien S, Kohler A (2010) Predicting the fatty acid composition of milk: a comparison of two Fourier transform infrared sampling techniques. Appl Spectrosc 64(7):700–707. doi: 10.1366/000370210791666200

Alonso V, Campo MM, Español S, Roncalés P, Beltrán JA (2009) Effect of crossbreeding and gender on meat quality and fatty acid composition in pork. Meat Sci 81:209–217. doi: 10.1016/j.meatsci.2008.07.021

AOAC (1984) Official methods of analysis AOAC International, Association of Official Analytical Chemists. Arlington. Official method 28023. Iodine Absorption number Wijs Method

AOCS (1998) Official methods and recommended practices, 5th ed. Edited by D. Firestone, AOCS Champaign, method Cd 1c-85

Balabin RM, Smirnov SV (2011) Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. Anal Chim Acta 692(1–2):63–72. doi: 10.1016/j.aca.2011.03.006

Bekiaris G, Triolo JM, Peltre C, Pedersen L, Jensen LS, Bruun S (2015) Rapid estimation of the biochemical methane potential of plant biomasses using Fourier transform mid-infrared photoacoustic spectroscopy. Bioresour Technol 197:475–481. doi: 10.1016/j.biortech.2015.08.050

Berhe DT, Eskildsen CE, Lametsch R, Hviid MS, van den Berg F, Engelsen SB (2016) Prediction of total fatty acid parameters and individual fatty acids in pork backfat using Raman spectroscopy and chemometrics: understanding the cage of covariance between highly correlated fat parameters. Meat Sci 111:18–26. doi: 10.1016/j.meatsci.2015.08.009

Bro R, Smilde AK (2014) Principal component analysis. Anal Methods-UK 6(9):2812–2831. doi: 10.1039/c3ay41907j

Chen JY, Iyo C, Terada F, Kawano S (2002) Effect of multiplicative scatter correction on wavelength selection for near infrared calibration to determine fat content in raw milk. J Near Infrared Spectrosc 10(4):301–307. doi: 10.1255/jnirs.346

Cocchi M, Corbellini M, Foca G, Lucisano M, Pagani MA, Tassi L, Ulrici A (2005) Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra. Anal Chim Acta 544:100–107. doi: 10.1016/j.aca.2005.02.075

Cocchi M, Durante C, Foca G, Marchetti A, Tassi L, Ulrici A (2006) Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. Talanta 68(5):1505–1511. doi: 10.1016/j.talanta.2005.08.005

Cox R, Lebrasseur J, Michiels E, Buijs H, Li H, Van de Voort FR, Ismail AA, Sedman J (2000) Determination of iodine value with a Fourier transform-near infrared based global calibration using disposable vials: an international collaborative study. JAOCS 77(12):1229–1234. doi: 10.1007/s11746-000-0192-4

Cozzolino D, Kwiatkowski MJ, Parker M, Cynkar WU, Dambergs RG, Gishen M, Herderich MJ (2004) Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy. Anal Chim Acta 513:73–80. doi: 10.1016/j.aca.2003.08.066

Dais P, Spyros A, Christophoridou S, Hatzakis E, Fragaki G, Agiomyrgianaki A, Salivaras E, Siragakis G, Daskalaki D, Tasioula-Margari M, Brenes M (2007) Comparison of analytical methodologies based on ^{1}H and ^{31}P NMR spectroscopy with conventional methods of analysis for the determination of some olive oil constituents. J Agric Food Chem 55:577–584. doi: 10.1021/jf061601y

Eskildsen CE, Rasmussen MA, Engelsen SB, Larsen LB, Poulsen NA, Skov T (2014) Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: understanding predictions of highly collinear reference variables. J Dairy Sci 97(12):7940–7951. doi: 10.3168/jds.2014-8337

Fernández-Cabanás VM, Garrido-Varo A, García Olmo J, De Pedro E, Dardenne P (2007) Optimisation of the spectral pre-treatments used for Iberian pig fat NIR calibrations. Chemom Intell Lab Syst 87:104–112. doi: 10.1016/j.chemolab.2006.10.005

Ferrari E, Foca G, Vignali M, Tassi L, Ulrici A (2011) Adulteration of the anthocyanin content of red wines: perspectives for authentication by Fourier transform-near infrared and ^{1}H NMR spectroscopies. Anal Chim Acta 701:139–151. doi: 10.1016/j.aca.2011.05.053

Ficarra A, Lo Fiego DP, Minelli G, Antonelli A (2010) Ultra fast analysis of subcutaneous pork fat. Food Chem 121:809–814. doi: 10.1016/j.foodchem.2010.01.003

Foca G, Cocchi M, Li Vigni M, Caramanico R, Corbellini M, Ulrici A (2009) Different feature selection strategies in the wavelet domain applied to NIR-based quality classification models of bread wheat flours. Chemom Intell Lab Syst 99:91–100. doi: 10.1016/j.chemolab.2009.07.013

Foca G, Salvo D, Cino A, Ferrari C, Lo Fiego DP, Minelli G, Ulrici A (2013) Classification of pig fat samples from different subcutaneous layers by means of fast and non-destructive analytical techniques. Food Res Int 52:185–197. doi: 10.1016/j.foodres.2013.03.022

Giarola M, Rossi B, Mosconi E, Fontanella M, Marzola P, Scambi I, Sbarbati A, Mariotto G (2011) Fast and minimally invasive determination of the unsaturation index of white fat depots by micro-raman spectroscopy. Lipids 46:659–667. doi: 10.1007/s11745-011-3567-8

Gjerlaug-Enger E, Kongsro J, Aass L, Ødegard J, Vangen O (2011) Prediction of fat quality in pig carcasses by near-infrared spectroscopy. Animal 5(11):1829–1841. doi: 10.1017/S1751731111000814

González-Martin I, González-Pérez C, Alvarez-García N, González-Cabrera JM (2005) On-line determination of fatty

acids composition in intramuscular fat of Iberian pork loin by NIRs with a remote reflectance fibre optic probe. Meat Sci 69:243–248. doi: 10.1016/j.meatsci.2004.07.003

Gosselin R, Rodrigue D, Duchesne C (2010) A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications. Chemom Intell Lab Syst 100:12–21. doi: 10.1016/j.chemolab.2009.09.005

IUPAC (1979) Method 1.122. Standard methods for the analysis of oils, fats and derivatives, 6th ed. Pergamon Press, New York

AQ1

IUPAC (1998) Compendium of analytical nomenclature, the orange book, 3rd ed. Blackwell Science, Oxford

Leardi R, Nørgaard L (2004) Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions. J Chemom 18(11):486–497. doi: 10.1002/cem.893

Lee HW, Bawn A, Yoon S (2012) Reproducibility, complementary measure of predictability for robustness improvement of multivariate calibration models via variable selection. Anal Chim Acta 757:11–18. doi: 10.1016/j.aca.2012.10.025

Li H, van de Voort FR, Sedman J, Ismail AA (1999) Rapid determination of cis and trans content, iodine value, and saponification number of edible oils by Fourier transform near-infrared spectroscopy. JAOCS 76(4):491–497. doi: 10.1007/s11746-999-0030-6

Lo Fiego DP, Santoro P, Macchioni P, De Leonibus E (2005) Influence of genetic type, live weight at slaughter and carcass fatness on fatty acid composition of subcutaneous adipose tissue of raw ham in the heavy pig. Meat Sci 69:107–114. doi: 10.1016/j.meatsci.2004.06.010

Lo Fiego DP, Macchioni P, Minelli G, Santoro P (2010) Lipid composition of covering and intramuscular fat in pigs at different slaughter age. Ital J Anim Sci 9(2):200–205. doi: 10.4081/ijas.2010.e39

Mevik BH, Cederkvist HR (2004) Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). J Chemom 18:422–429. doi: 10.1002/cem.887

Minelli G, Macchioni B, Ielo MC, Santoro P, Lo Fiego DP (2013) Effects of dietary level of pantothenic acid and sex on carcass, meat quality traits and fatty acid composition of thigh subcutaneous adipose tissue in Italian heavy pigs. Ital J Anim Sci 12:329–336. doi: 10.4081/ijas.2013.e52

Monziols M, Bonneau M, Davanel A, Kouba M (2007) Comparison of the lipid content and fatty acid composition of intramuscular and subcutaneous adipose tissues in pig carcasses. Meat Sci 76:54–60. doi: 10.1016/j.meatsci.2006.10.013

Müller M, Scheeder MRL (2008) Determination of fatty acid composition and consistency of raw pig fat with near infrared spectroscopy. J Near Infrared Spectrosc 16(3):305–309. doi: 10.1255/jnirs.792

Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB (2000) Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc 54:413–419. doi: 10.1366/0003702001949500

Panford JA, deMan JM (1990) Determination of oil content of seeds by NIR: influence of fatty acid composition on wavelength selection. JAOCS 67(8):473–482. doi: 10.1007/BF02540751

Pérez-Juan M, Afseth NK, González J, Díaz I, Gispert M, Font i Furnols M, Oliver MA, Realini CE (2010) Prediction of fatty acid composition using a NIRS fibre optics probe at two different locations of ham subcutaneous fat. Food Res Int 43:1416–1422. doi: 10.1016/j.foodres.2010.04.006

Pétursson S (2002) Clarification and expansion of formulas in AOCS recommended practice Cd 1c-85 for the calculation of iodine value from FA composition. JAOCS 79(6):621–622. doi: 10.1007/s11746-002-0551-1

Piasentier E, Di Bernardo N, Morgante M, Sepulcri A, Vitale M (2009) Fatty acids composition of heavy pig back fat in relationship to some animal factors. Ital J Anim Sci 8(2):531–533. doi: 10.4081/ijas.2009.s2.531

Prieto N, Uttaro B, Mapiye C, Turner TD, Dugan MER, Zamora V, Young M, Beltranena E (2014) Predicting fat quality from pigs fed reduced-oil corn dried distillers grains with soluble by near infrared reflectance spectroscopy: fatty acid composition and iodine value. Meat Sci 98(4):585–590. doi: 10.1016/j.meatsci.2014.06.009

Rinnan A, Van den Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. Trends Anal Chem 28:1201–1222. doi: 10.1016/j.trac.2009.07.007

Ripoche A, Guillard AS (2001) Determination of fatty acid composition of pork fat by Fourier transform infrared spectroscopy. Meat Sci 58:299–304. doi: 10.1016/S0309-1740(01)00031-6

Santoro P (1983) Fat quality in pig meat with special emphasis on cured and seasoned raw hams. In: Fat Quality in Lean Pigs, Workshop in the EEC programme, Brussels, pp 43–46

Shenk JS, Workman JJJ, Westerhaus MO (2008) Application of NIR spectroscopy to agricultural products. In: Burns DA, Ciurczak EW (eds) Handbook of near-infrared analysis, 3rd edn. CRC Press, Boca Raton, Chpt. 17

Sørensen KM, Petersen H, Engelsen SB (2012) An on-line near-infrared (NIR) transmission method for determining depth profiles of fatty acid composition and iodine value in porcine adipose fat tissue. Appl Spectrosc 66(2):218–226. doi: 10.1366/11-06396

Ulrici A, Li Vigni M, Durante C, Foca G, Belloni P, Brettagna B, De Marco T, Cocchi M (2008) At-line monitoring

of the leavening process in industrial bread making by near infrared spectroscopy. J Near Infrared Spectrosc 16(3):223–231. doi: 10.1255/jnirs.781

Wood JD, Richardson RI, Nute GR, Fisher AV, Campo MM, Kasapidou E, Sheard PR, Enser M (2003) Effects of fatty acids on meat quality: a review. Meat Sci 66:21–32. doi: 10.1016/S0309-1740(03)00022-6

Wu D, Chen X, Shi P, Wang S, Feng F, He Y (2009) Determination of α-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. Anal Chim Acta 634(2):166–171. doi: 10.1016/j.aca.2008.12.024

Wu D, Yong H, Pengcheng N, Fang C, Yidan B (2010) Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. Anal Chim Acta 659:229–237. doi: 10.1016/j.aca.2009.11.045

Xiaobo Z, Jiewen Z, Povey MJ, Holmes M, Hanpin M (2010) Variables selection methods in near-infrared spectroscopy. Anal Chim Acta 667:14–32. doi: 10.1016/j.aca.2010.03.048

Zamora-Rojas E, Garrido-Varo A, De Pedro-Sanz E, Guerrero-Ginel JE, Pérez-Marín D (2013) Prediction of fatty acids content in pig adipose tissue by near infrared spectroscopy: at-line versus in-situ analysis. Meat Sci 95:503–511. doi: 10.1016/j.meatsci.2013.05.020