Food Chemistry 205 (2016) 14–22

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Detection and characterisation of frauds in bovine meat *in natura* by non-meat ingredient additions using data fusion of chemical parameters and ATR-FTIR spectroscopy

Karen M. Nunes [a], Marcus Vinícius O. Andrade [b], Antônio M.P. Santos Filho [b], Marcelo C. Lasmar [b], Marcelo M. Sena [a,c,*]

[a] *Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil*
[b] *Polícia Federal, Superintendência Regional de Minas Gerais, 30430-340 Belo Horizonte, MG, Brazil*
[c] *Instituto Nacional de Ciência e Tecnologia em Bioanalítica, 13083-970 Campinas, SP, Brazil*

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 25 February 2016
Accepted 27 February 2016
Available online 3 March 2016

Keywords:
Data fusion
Meat adulteration
Forensic analysis
Mid-infrared spectroscopy
PLS-DA

ABSTRACT

Concerns about meat authenticity are increasing recently, due to great fraud scandals. This paper analysed real samples (43 adulterated and 12 controls) originated from criminal networks dismantled by the Brazilian Police. This fraud consisted of injecting solutions of non-meat ingredients (NaCl, phosphates, carrageenan, maltodextrin) in bovine meat, aiming to increase its water holding capacity. Five physico-chemical variables were determined, protein, ash, chloride, sodium, phosphate. Additionally, infrared spectra were recorded. Supervised classification PLS-DA models were built with each data set individually, but the best model was obtained with data fusion, correctly detecting 91% of the adulterated samples. From this model, a variable selection based on the highest VIPscores was performed and a new data fusion model was built with only one chemical variable, providing slightly lower predictions, but a good cost/performance ratio. Finally, some of the selected infrared bands were specifically associated to the presence of adulterants NaCl, tripolyphosphate and carrageenan.

1. Introduction

Food authenticity and integrity are steadily increasing concerns throughout the world for both the consumer protection authorities and the producers and dealers. Considering our food is under constant threat from adulteration, its detection provides a great challenge to the analytical chemists due to the increasing product diversity, the continuous development of new production technologies, and the emergence of new types of fraud. As compared to the classical authenticity assessment based on the analysis of specific marker compounds (target analysis), food fingerprinting, the non-target chemical analysis of food, has clear advantages for investigating multiple objectives with only one analytical method (Esslinger, Riedl, & Fauhl-Hassek, 2014). This usually involves spectroscopic techniques and multivariate data analysis. The capacity of detecting food frauds and extracting information can be even increased by the combination of data from different analytical techniques or origins, using the chemometric strategy of data fusion (Borràs et al., 2015).

Food fraud decreases the quality of the products, misleads the consumers and may imply health risks. It is considered food crime when it becomes an organized activity by groups which knowingly set out to deceive those purchasing food, with potential for generating huge negative impacts both on consumer confidence, and on the reputation and finances of food businesses (Her Majesty's Government, 2014). In the last years, several food adulteration scandals had great repercussion worldwide, particularly involving meat adulteration, such as the recent horsemeat scandal in EU (FSAI, 2013). In Brazil, a major scandal was reported by the Federal Police in 2012. Several slaughterhouses in the metropolitan region of Belo Horizonte, capital of Minas Gerais State, were assessed for frauds in bovine meat *in natura* by non-meat ingredient addition (DPF, 2012).

Authentication problems with respect to meat can be categorized into four major areas where fraud is most likely to occur: meat origin, meat substitution, meat processing and non-meat ingredient addition (Ballin, 2010). Substitution of one species with another is the most reported type of meat fraud, such as in the horsemeat scandal (FSAI, 2013). Most common methods for

---

* Corresponding author at: Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil.
E-mail address: marcsen@ufmg.br (M.M. Sena).

http://dx.doi.org/10.1016/j.foodchem.2016.02.158
0308-8146/© 2016 Elsevier Ltd. All rights reserved.

detecting this type of adulteration are based on the analysis of DNA, such as PCR (polymerase chain reaction) (Iwobi et al., 2015). Nevertheless, these methods are laborious and time-consuming, and in the last years alternative methods based on vibrational spectroscopy, in the majority of cases associated with chemometric tools, have been developed. Thus, adulterations involving bovine, pork, horse, turkey and llama meats have been detected using Raman (Zajac, Hanuza, & Dyminska, 2014), near infrared (NIR) (Mamani-Linares, Gallo, & Alomar, 2012), mid infrared (MIR) (Rohman, Sismindari, Erwanto, & Che Man, 2011) and data fusion with different spectroscopic techniques (Alamprese, Casale, Sinelli, Lanteri, & Casiraghi, 2013).

In contrast, addition of non-meat ingredients is the least common type of fraud described in the literature (Ballin, 2010). One of the few examples is the development of a method based on HPLC-MS for detecting enzymes, which take part in the blood clotting, in pork meat products (Grundy et al., 2008). Water can also be fraudulently added to meat and regulations establish the allowed amount of exogenous water in meat. A standard method to determine this type of meat fraud is based on the water/protein ratio (EC, 2005). If water is added to meat, this ratio may be too high and can serve as a clear indication of fraud. Nevertheless, proteins from other sources and salts, such as phosphate, can be added to meat products aiming to increase water binding and leaving the water/protein ratio close to its natural value. In that case, detection of proteins and/or exogenous salts is necessary in order to confirm the meat fraud (Ballin, 2010).

This work studied frauds by non-meat ingredient addition in a real case of adulteration (DPF, 2012). Due to complaints the Brazilian Federal Police impounded in the denounced slaughterhouses mixtures of adulterant substances. Suspicious substances include salts, such as sodium chloride, phosphate, tripolyphosphate and acid pyrophosphate, carrageenan (a linear sulphated polysaccharide extracted from red edible seaweeds), maltodextrin (a complex carbohydrate obtained from starch) and collagen (a structural protein naturally present in meat tissues). Aqueous solutions of these substances were injected in bovine meat using specific equipments also found in these slaughterhouses. Adulteration by addition of salts aims at increasing the water holding capacity (WHC) of meat (Cheng & Sun, 2008), providing an economic fraud related to weight gain, since the meat product will absorb more water. The increase in WHC enhances the meat's succulence due to the relaxation of muscle fibers, resulting in more tenderness. The addition of salts reduces the water/protein ratio, however, the higher WHC restores this ratio close to the normal values. The effect of the addition of salts, such as NaCl, KCl and $MgSO_4$, in the increase of WHC of meats is very dependent on the salt concentrations. This effect has been studied in the range of 1.5–9.0% of salt concentration (not in situations of fraud) by NIR and MIR spectroscopy jointly with chemometric methods, such as principal component analysis (PCA) (Perisic, Afseth, Ofstad, Narum, & Kohler, 2011, 2013). The WHC increase is related to changes in proteins conformation due to the addition of salts. Carrageenan can also increase the WHC of meats (Ayadi, Kechaou, Makni, & Attia, 2009). Nevertheless, at present there are no studies about meat frauds with the addition of these substances.

Quality control of bovine meat consumed in Brazil is based on the determination of physico-chemical and microbiological parameters by classical analytical methods (MAPA, 1999). These analyses generate a great amount of data, but individual parameters are not appropriate to characterize frauds, because they can vary as a function of sex, meat cuts, breed, feed intake, slaughter age, among others (Ballin, 2010). Thus, the use of chemometric tools is necessary. In this paper, detection and characterization of real adulterated bovine meat samples, seized by the Brazilian Federal Police, were performed by data fusion of five physico-chemical parameters (protein, sodium, chloride, phosphate and ash) and attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectra. Supervised classification models using partial least squares discriminant analysis (PLS-DA) were built for differentiating adulterated from control samples. The developed methods were compared and validated through the estimate of appropriate figures of merit (FOM).

2. Materials and methods

2.1. Police operation

This operation was carried out by the Brazilian Federal Police in 2012, and five slaughterhouses were assessed in the cities of Belo Horizonte, Contagem and Ibirité (DPF, 2012). In these places, packages and industrial products used for meat adulteration were found and seized. These frauds consisted in injecting in the meats aqueous solutions of these products, which contained NaCl, phosphate, tripolyphosphate, acid pyrophosphate, carrageenan, maltodextrin and/or collagen. Specific equipment used for these frauds were also found in the slaughterhouses, consisting of injection and tenderizing machines (Fig. 1S, Supplemental materials).

Samples of bovine meat *in natura* were confiscated, sealed and sent for analysis in order to characterize their composition. These samples presented anomalous aspects considered evidences of adulteration, such as accumulation of frozen liquid in the extremities of the pieces of meat, excessive exudation of viscous liquids forming foam, presence of symmetric marks of perforation in the meat surface, and excessive volume of exudated liquid after thawing (purge). Adulterated meat samples were also submitted to organoleptic analysis, such as color, smell, texture and visual identification of non-meat components, and determination of physico-chemical parameters. The identification of adulterations was not relied on individual observations/determinations, but on the entire group of them. As a result of this investigation, the Brazilian Federal Police identified all suspected samples as adulterated. In this paper, the focus was on the chemical analysis, including physico-chemical parameters and non-destructive ATR-FTIR measurements.

2.2. Samples

Fifty-five samples were provided by the Federal Police, 43 identified as adulterated by the Brazilian Federal Police and 12 control meat pieces. Control samples were guaranteed to be without adulteration, since they were sampled during the boning process and taken directly from the bovine carcass without any contact with adulteration machines. Samples were originated from five different seizures and obtained from six different cuts, top inside (*semimembranosus*), eye of the round (*semitendinosus*), knuckle (*rectus femoris*), outside round (*gluteus biceps*), shank (*extensor digitorum*) and top sirloin (*longissumus dorsi*). Adulterated and control samples were submitted to exactly the same treatments for analysis. All the samples were stored in a freezer below −5 °C immediately after sampling. They were stored for about one month prior to the analyses.

2.3. Physico-chemical parameters

All reagents used in this work were of analytical grade, purchased from certified suppliers, and used without further purification. Five parameters, protein, ash, sodium, chloride and phosphate, were determined in triplicate by classical methods according to the regulation of the Brazilian Ministry of Agriculture and Livestock (MAPA, 1999). Fat was not determined, because the natural variation among different meat cuts makes not feasible its use for evaluating frauds. Protein was determined by the Kjeldahl method. Ash was determined by calcination at 550 °C. The residues of calcination were dissolved in 5 mL of deionized water, heated in water bath for 15 min, and used for determining sodium, chloride and phosphate by ion exchange chromatography. A Metrohm 761, Compact IC, chromatograph (Metrohm AG, Herisau, Switzerland) was used, with electric conductivity detection, a Metrosep A Supp 5–100 column for anion analysis and a Metrosep Cation 1–2 column for cation analysis.

2.4. ATR-FTIR measurements and data processing

FTIR measurements were obtained in a Nicolet 380 FTIR spectrometer (Thermo Fisher Scientific Inc., Madison, USA), equipped with an attenuated total reflection accessory containing a diamond crystal of one reflection (Smart Orbit Diamond ATR, Thermo Fisher, USA). Small portions of thawed meat were put in the ATR accessory, covering the crystal. Spectra were recorded in triplicate from 4000 to 525 cm^{-1}, with a resolution of 2 cm^{-1} and 32 scans. Mean spectra of each sample were used for building the models. Data were processed using MATLAB software, version 8.4 (The MathWorks, Natick, USA) and PLS Toolbox, version 7.0 (Eigenvector Technologies, Manson, USA).

2.5. Chemometric methodology

Previous unsupervised PCA models were developed for both the data sets, physico-chemical parameters and ATR-FTIR spectra. Nevertheless, the main focus of this paper was to develop supervised PLS-DA models for the data obtained from individual techniques and also for the fused data. PLS-DA is one of the most used multivariate classification methods (Brereton & Lloyd, 2014). It is based on the traditional PLS regression, correlating a block of independent variables, X, with a block of dependent variables, Y. In this work, PLS1-DA was used and the $\boldsymbol{X}$ matrix was built containing physico-chemical parameters, FTIR spectra or fused data, while $\boldsymbol{y}$ is a vector containing the dummy variables 1 or 0, for adulterated or control samples, respectively. Bayesian threshold was adopted for the models' predictions (Botelho, Reis, Oliveira, & Sena, 2015; Wise et al., 2006). Samples with predicted y values above the threshold were classified as adulterated.

Data were preprocessed by class centroid centering (Wise et al., 2006) in order to compensate the unequal sizes of the classes (Brereton & Lloyd, 2014). Other specific preprocessing will be mentioned in the discussion for each data set. The number of latent variables (LV) was chosen by venetian blinds (4 splits) cross validation, based on the smallest cross validation classification error (CVCE). In order to assure the development of robust and unbiased models, data should be split in training and test sets. This split should be based on systematic criteria, which assure that the samples used in the calibration/training step are representative and homogeneously distributed in the whole analytical range of the multivariate space. With this aim, the Kennard and Stone (1969) algorithm was applied for the selection of training samples separately in both the classes. Thus, the samples were split in 38 (30 adulterated and 8 control) for the training set and 17 (13 adulterated and 4 control) for the test set. This ratio of the number of training/test samples corresponds to two thirds/one third, typically used in multivariate calibration (ASTM & Annual Book of ASTM Standards, 2012) and supervised classification.

Once the models were built, they were validated and compared through the estimate of figures of merit (FOM) related to the presence of errors in the results, such as the numbers of false positives (FP) and false negatives (FN), sensitivity (SEN), specificity (SPE), reliability or efficiency rate (EFR) and area under the receiver operating characteristic (AUROC) (Botelho et al., 2015; Lopez, Colomer, Ruisanchez, & Callao, 2014). SEN is the rate of true positives (the ratio between the number of true positives and the sum of true positives and FN) and SPE is the rate of true negatives. EFR is a parameter encompassing both the types of errors, which is calculated as the difference between the total of results (100%) and the sum of the rates of FP and FN. AUROC is estimated from the ROC curves, which plot both SEN and SPE as functions of the threshold (y predicted). AUROC values range between 1.0, which means perfect discrimination between classes, and 0.0. Values lower than 0.5 indicate no discriminant models.

Models can be interpreted in order to identify the most predictive variables through the plot of variable importance in the projection (VIP) scores (Chong & Jun, 2005). VIP scores are weighted sums of squares of the PLS weights, taking into account the amount of explained y variance in each dimension. The utility of VIP scores lies in its intrinsic parsimony, since for a given model there will always be only one VIP-vector, summarizing all components and y variables. As VIP scores are a squared function of the PLS weights, only positive values will be found. Predictors with a large VIP, usually higher than 1.0, are the most influential for the model and can be used for variable selection.

2.6. Data fusion

Data fusion is a strategy that combines the outputs of different analytical instruments (such as NIR, MIR, Raman, UV/vis, mass spectrometry, nuclear magnetic resonance, fluorescence spectra or other types), sensors and/or physico-chemical variables. The large amount of generated data demands the use of chemometric tools for modeling and interpretation. As compared to a single technique, data fusion from complementary instruments can provide classifications with less error rate or predictions with less uncertainty. Evidently, data fusion is useful when complementary information is modeled, since its main goal is to increase the synergy between the fused techniques by merging complementary inputs (Borràs et al., 2015).

Data fusion can occur at three levels: low, medium and high. Low-level (measurements level) fusion consists in simply concatenating original variables after the preprocessing steps. Medium-level (feature level) fusion extracts relevant features from each data source individually and then combines them into a single matrix, which will be treated by classification or calibration methods. Commonly, the most significant scores from PCA, PLS or PLS-DA are fused. High-level (decision level) fusion constructs a separated multivariate model for each technique and combines the individual outputs to produce the final result. This level of fusion has often provided worse results than the other two levels (Borràs et al., 2015).

Data fusion is a relatively recent subject in analytical chemistry and its applications have been increasing in the last years, mainly in the development of multivariate classification or calibration methods for analysis of complex food samples, such as beer (Biancolillo, Bucci, Magri, Magri, & Marini, 2014; Vera et al., 2011), olive oil (Casale et al., 2012; Pizarro, Rodríguez-Tecedor, Pérez-del-Notario, Esteban-Díez, & González-Sáiz, 2013), cheese (Cozzi, Ferlito, Pasini, Contiero, & Gottardo, 2009), meat (Alamprese et al., 2013) and fish meat (Zotte et al., 2014), but also to other matrices, such as biological samples in metabolomic studies (Bro et al., 2013), pigment determination in works of art (Ramos, Ruisánchez, & Andrikopoulos, 2008) and quality control of transformer insulating oils (Godinho et al., 2014). In most of these papers fusion has been limited to spectral or sensor data continuously concatenated. In the present paper, low- and medium-level approaches were compared for fusion of continuous spectra (FTIR) and chemical parameters obtained individually, the latter discretely concatenated. Though less common, this strategy has been used in other papers (Cozzi et al., 2009; Pizarro et al., 2013; Zotte et al., 2014).

3. Results and discussion

3.1. Unsupervised PCA models

A PCA model was built with the five physico-chemical parameters determined for meat samples. Data were previously autoscaled (mean centered and divided by the standard deviation of each variable). The first two PC accounted for 90.81% of the total variance. By observing the scores of PC1 × PC2 biplot (Fig. 1), all the control samples were clearly grouped on PC1 negative values, while about two thirds of the adulterated samples were projected on PC1 positive values. Thus, PC1 provided a fairly discrimination between control and adulterated samples. The loadings (Fig. 1) indicated that control samples have in general lower contents of ash, sodium, chloride and phosphate and higher content of protein. This is coherent with the adulteration by addition of salts to meat samples, as was observed by the Federal Police in this operation.

Analytical methods employed for determining physico-chemical parameters are destructive, laborious and slow, consuming reagents and generating chemical waste. Thus, ATR-FTIR is a simple, rapid, low cost and environmentally friendly alternative for detecting frauds in bovine meat. The obtained spectra were pre-processed by Savitzky-Golay smoothing (Savitzky & Golay, 1964) (15 points in filter and second order polynomial fit), in order to increase the signal-to-noise ratio, multiplicative scatter correction (MSC) (Rinnan, van den Berg, & Engelsen, 2009), in order to eliminate non-linear baseline deviations (drifts), and mean centering. Resultant spectra are shown in Fig. 2. No significant visual differences between control and adulterated samples were noted. Two spectral regions were deleted before building the models. In the region between 2400 and 1800 cm^{-1}, interfering absorptions due to atmospheric CO_2 and diamond ATR crystal (Küpper, Heise, & Butvina, 2001) were present. The region above 3700 cm^{-1} presented only instrumental noise.

For a PCA model built with spectral data, the first two PC accounted for 79.93% of the variance. Nevertheless, the scores plot of PC1 × PC2 (not shown) provided no clear discrimination between adulterated and control samples. Some samples were grouped as a function of their meat cut, but this was also not a clear trend. PC3 (14.24%) did not also provide any discrimination. By comparing both the PCA models, physico-chemical data have clearly more discriminant power. This is already expected, because these variables were specifically chosen searching for detecting adulterations. On the other hand, vibrational spectra provide non-selective information related to the total chemical composition of the samples, which depends on other factors, such as meat cut and animal origin. In the sequence, supervised PLS-DA models will be developed for both the data sets trying to increase the samples' discrimination.

3.2. PLS-DA model for physico-chemical data

Adulterated samples (their pre-classification was based on a body of evidence described in Section 2.1) were arbitrarily defined as class 1 and control samples as class 0. As already mentioned in Section 2.5, Kennard-Stone algorithm was used for selecting training samples. However, this algorithm was applied separately for

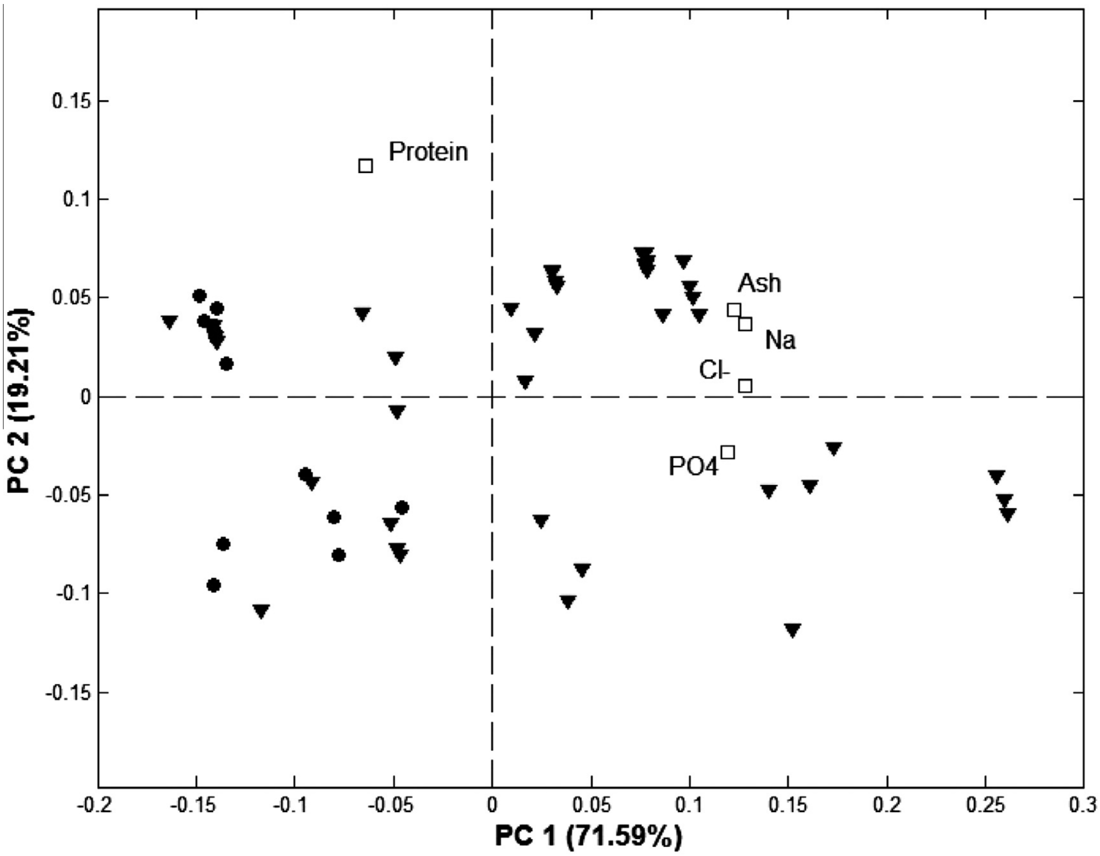

Fig. 1. Biplot of PC1 × PC2 for physico-chemical parameters. Full triangles indicate the scores of the adulterated samples, full circles the scores of the control samples, and empty squares the loadings.

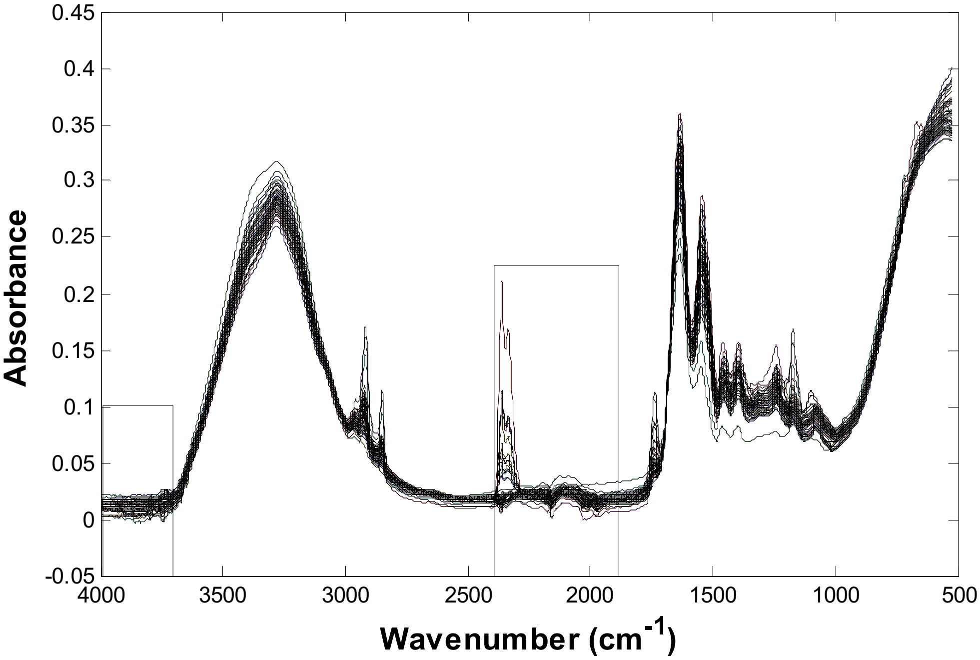

Fig. 2. ATR-FTIR spectra of the 55 analysed meat samples preprocessed by Savitzky-Golay smoothing and MSC. Deleted spectral regions are marked in rectangles.

each data set: physico-chemical, FTIR and fused data. Thus, the samples in the training sets were not necessarily the same in each of these three sets. Data were previously class centroid centered and divided by the standard deviation of each variable (equivalent to autoscaling). The best model was chosen by cross-validation with 3 LV, accounting for 94.16% and 36.58% of variance in *X* and *Y* blocks, respectively.

Predictions for this model are shown in Supplemental material (Fig. 2S). Estimated FOM are shown in Table 1. This model provided 5 (16.7%) FN, adulterated samples predicted as control ones, in the training set, and 2 (15.4%) FN in the test set, while no FP was observed. VIP scores are shown in Fig. 3, in which the importance of chloride (1.86) as the main discriminant variable can be noted. Sodium and ash also presented VIP scores near or above 1.00, indicating that these are significant variables. Considering NaCl was one of the main adulterants used in this fraud, the higher discriminant power of chloride in the model can be associated to the natural presence of sodium in bovine meats.

3.3. PLS-DA model for ATR-FTIR data

This data set were sequentially preprocessed by Savitzky-Golay smoothing, MSC and class centroid centering. As for the previous PCA model, spectral regions between 2400–1800 cm^{-1} and above 3700 cm^{-1} were deleted after preprocessing. The best model was obtained with 4 LV, accounting for 95.70% and 26.71% of variance in *X* and *Y* blocks, respectively. This model, whose predictions are shown in Fig. 3S, provided 5 (16.7%) FN and 3 FP (37.5%) in the training set, and 3 (23.1%) FN and 1 (25%) FP in the test set. As can be observed in Table 1 for all the FOM, this model is clearly worse than the PLS-DA model built with physico-chemical data. This was already expected, since FTIR spectra contain less selective information, which are more susceptible to the influence of other chemical constituents not related to the adulteration. Nevertheless, the results for this model were considered reasonable. Discussion about the most discriminant wavenumbers will be carried out in the next section, for the data fusion model.

3.4. Low-level data fusion

Physico-chemical and ATR-FTIR data were simply concatenated and the same preprocessing previously mentioned was used for each data set. In addition, FTIR spectra were autoscaled. The same spectral regions previously cited were deleted. Low-level data fusion provided the best PLS-DA model. Predictions are shown in Fig. 4, and only 2 (6.7%) FN were observed in the training set and also (15.4%) in the test set. No FP was observed. Estimated FOM shown in Table 1 corroborates the superiority of this model, which was built with 5 LV, accounting for 97.45% and 58.48% of variance in X and Y blocks, respectively.

VIP scores vector was used for model interpretation and variable selection. When using VIP scores for variable selection, the threshold value of 1.0 has been suggested (Chong & Jun, 2005).

Table 1
Comparison of the five developed PLS-DA models through the estimate of FOM.

Model	Training Set					Test Set					AUROC
	FN	FP	SEN (%)	SPE (%)	EFR (%)	FN	FP	SEN (%)	SPE (%)	EFR (%)	
Physico-chemical data	5	0	83.3	100	83.3	2	0	84.6	100	84.6	0.846
ATR-FTIR data	5	3	83.3	62.5	45.8	3	1	76.9	75.0	51.9	0.750
Data fusion – low-level	2	0	93.3	100	93.3	2	0	84.6	100	84.6	0.923
Data fusion – mid-level	3	0	90.0	100	90.0	3	0	76.9	100	76.9	0.923
Data fusion – variable selection	5	1	83.3	87.5	70.8	2	0	84.6	100	84.6	0.907

FN = number of false negatives; FP = number of false positives; SEN = sensitivity rate; SPE = specificity rate; EFR = efficiency rate; AUROC = area under the ROC curve.

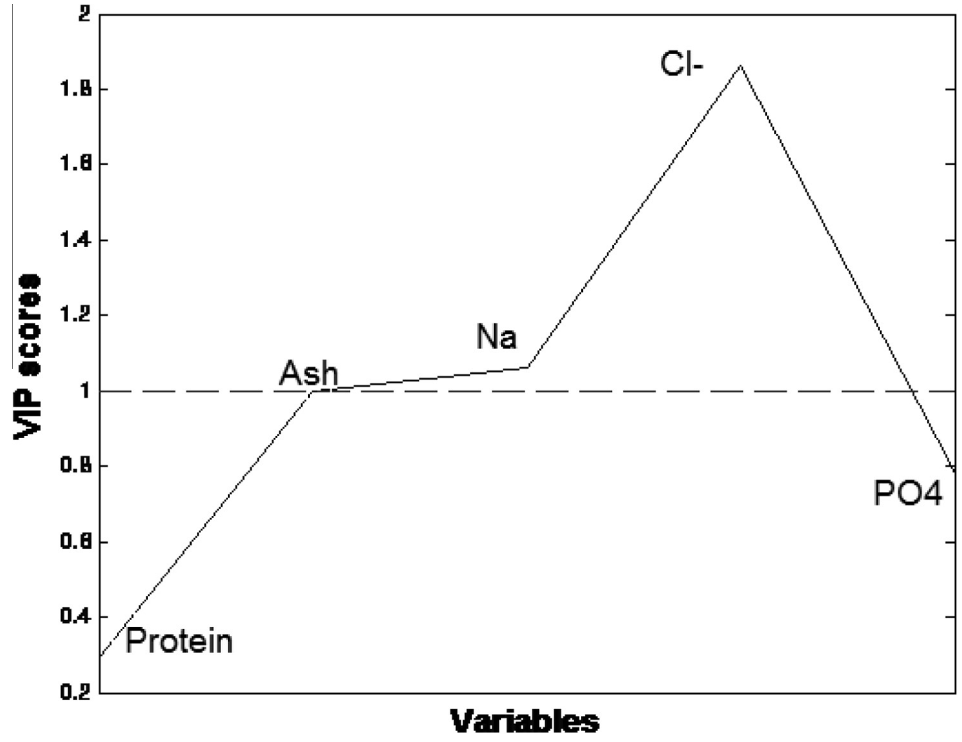

Fig. 3. VIP scores of the PLS-DA model for physico-chemical data.

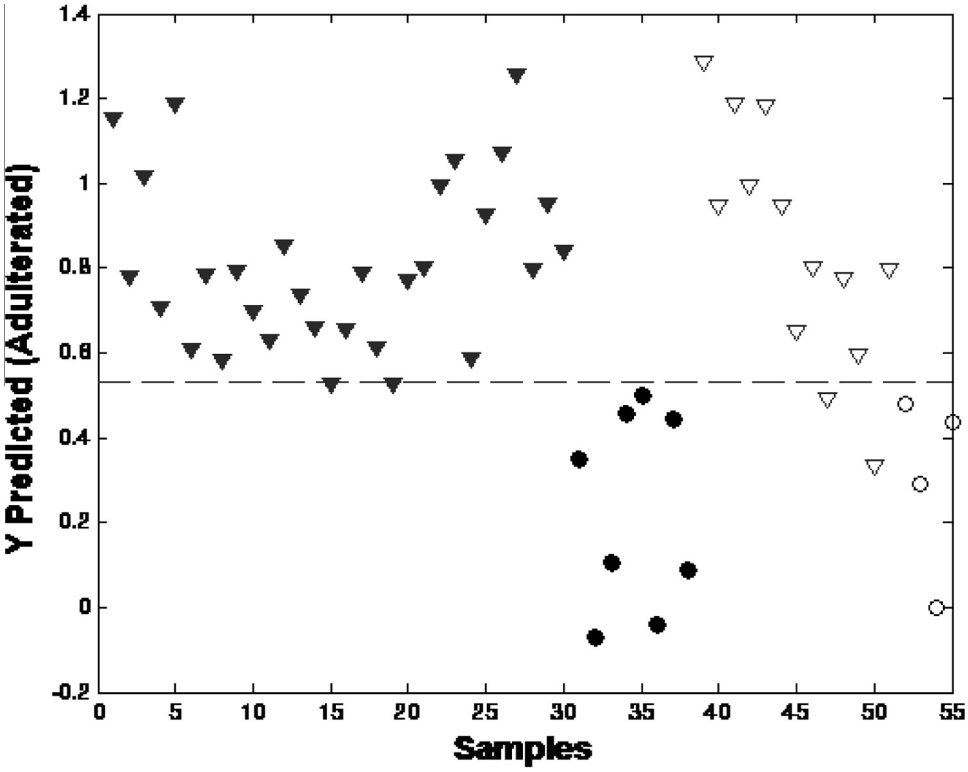

Fig. 4. PLS-DA predictions for low-level data fusion. Threshold is indicated by the dashed line. Adulterated and control samples are indicated by down triangles and circles, respectively. Full symbols represent the training set and empty symbols represent the test set.

Nevertheless, the more stringent value of 1.5 was adopted in this paper in order to select a smaller number of variables and spectral regions. VIP scores for physico-chemical data were similar to the previous model (Fig. 3) and only chloride presented a value above 1.5. VIP scores for spectral variables are shown in Fig. 5, in which the eight most important regions are marked with letters a–h. These spectral regions jointly with chloride were used to build a new PLS-DA model described in Section 3.6.

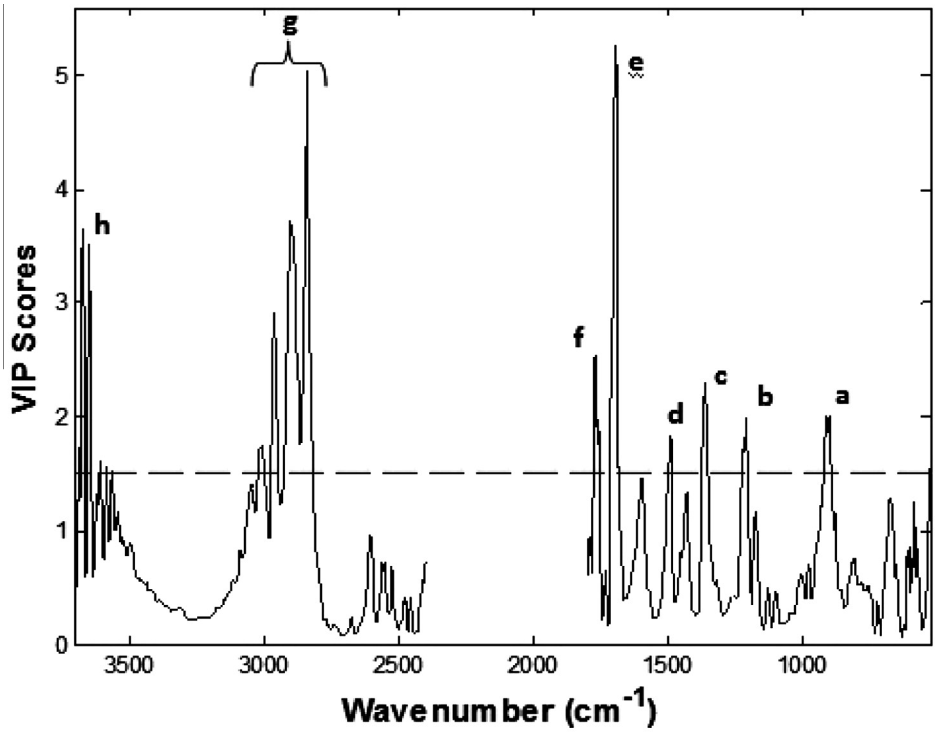

Fig. 5. VIP scores (spectral variables) for low-level data fusion PLS-DA model. Letters a–h indicate the most discriminant spectral bands.

Selected spectral regions were in the wavenumbers between 918–893 cm^{-1} (a), 1236–1210 cm^{-1} (b), 1373–1347 cm^{-1} (c), 1500–1481 cm^{-1} (d), 1717–1677 cm^{-1} (e), 1775–1752 cm^{-1} (f), 2972–2815 cm^{-1} (g) and 3690–3638 cm^{-1} (h). The most intense VIP scores are associated to bands e and g. A peak at 2916 cm^{-1} can be attributed to CH_2 asymmetric bending (Pavia, Lampman, Kriz, & Vyvyan, 2008). However, some peaks can be specifically associated to the adulterants used in this meat fraud. FTIR vibrations of aggregated β-sheets of proteins (C═O stretchings in the amide I region) at 1690 and at 1235 cm^{-1} are highly correlated to the additions of NaCl and/or KCl to bovine meat (Perisic, Afseth, Ofstad, Narum, & Kohler, 2013). These vibrations were in the selected spectral regions e and b, respectively, and the peak *e* (1717–1677 cm^{-1}) is the most discriminant for the model, demonstrating the effect of NaCl addition in this fraud. Region b also includes a specific vibration assigned to the stretching of ester sulphate group of carrageenan, at 1220 cm^{-1} (Volery, Besson, & Schaffer-Lequart, 2004). Region e partly coincides with the spectral range of 1700–1600 cm^{-1}, which has been the most used mid infrared region for the analysis of the secondary structure of collagen due to amide I vibrations (Cao & Xu, 2008). Finally, region a is included in the range between 930 and 890 cm^{-1} that has been related to sodium tripolyphosphate vibrations (Khanmohammadi, Ashori, Kargosha, & Garmarudi, 2007). For the other selected spectral regions specific attributions related to adulterations cannot be provided. However, general attributions can be suggested: region c is characteristic of CH_3 bending and/or the combination between OH bending and C–O stretching; region d could be associated to amine NH bending; region f to carbonyl stretching; and region h to water OH stretching (Pavia et al., 2008).

3.5. Mid-level data fusion

Mid-level data fusion PLS-DA was built by concatenating the scores of the LV used for constructing the individual physico-chemical and ATR-FTIR models. Thus, seven scores were combined, the first three from the physico-chemical model and the first four from the ATR-FTIR model. The best model was selected with 3 LV, accounting for 59.49% and 55.02% of variance in *X* and *Y* blocks, respectively. This model, whose predictions are shown in Fig. 4S, provided 3 (10.0%) FN in the training set and also (23.1%) in the test set. No FP was observed. These results and the estimated FOM indicated that mid-level data fusion was slightly worse than low-level data fusion.

3.6. Data fusion with variable selection

As already mentioned, a new low-level data fusion PLS-DA model was built based on a variable selection using the most significant VIP scores (above 1.50). Thus, the original number of variables was reduced from 1808 to 191, including the most discriminant physico-chemical variable, chloride, and the eight spectral bands marked in Fig. 5. This model was built with 4 LV, accounting for 97.31% and 46.87% of variance in *X* and *Y* blocks, respectively. Predictions are shown in Fig. 5S, in which the presence of 5 (16.7%) FN and 1 (12.5%) FP in the training set can be observed. For the test set, only 2 FN (15.4%) were detected.

Model based on variable selection presented higher error rates than models based on full data fusion. In general, low-level full data fusion presented the best results, providing only 4 FN (9.3%) and no FP. The best performance of this model (Table 1) can also be evaluated by EFR, a global FOM that takes into account both the FN and FP rates, and AUROC. EFR of 93.3% and 84.6% were obtained for training and test sets, respectively, while this model presented the higher AUROC, 0.923, indicating the best discriminating ability. On the other hand, the model based on variable selection was particularly cost-effective and provided reasonable SEN and SPE, above 83%. A method that utilized only one out of the five physico-chemical parameters presented the advantage of saving costs related to time and reagents spent in the other four determinations.

4. Conclusion

Considering the growing worldwide concern about food adulteration and the emergence of new types of fraud, one interesting aspect of this paper was to analyse real samples originated from a criminal network dismantled by the Brazilian Federal Police. This fraud consisted of injecting aqueous solutions of non-meat ingredients (NaCl, phosphates, carrageenan, maltodextrin, collagen) in bovine meat *in natura*, in order to increase its WHC and obtain an economic fraud.

For meat samples seized by the police, the determination of five specific physico-chemical parameters allowed to build multivariate classification models that differentiated them from control/unadulterated samples. Unsupervised PCA provided a fairly discrimination of about two thirds of the samples and supervised PLS-DA correctly detected 84% (36 out 43) of the adulterations. The lack of homogeneity in the adulterations, since not all the samples were injected with all the adulterants at the same levels, may justify the difficulty in detecting some samples.

On the other hand, ATR-FTIR spectroscopy was an advantageous alternative for detecting and characterising these adulterated samples, providing a simpler, more rapid, non-destructive, cleaner and low cost method. The best classification model was obtained by merging physico-chemical parameters and FTIR spectra. Low-level data fusion correctly detected 91% (39 out 43) of the adulterations, while mid-level data fusion provided a slightly worse model. Nevertheless, a low-level data fusion model based on variable selection using only variables with VIP scores above 1.5 provided good predictions, similar to the physico-chemical model. This model was considered particularly cost-effective, since only the most discriminant chemical variable, chloride, was used in addition to the eight more discriminant spectral bands selected by VIP scores. Adulterated samples were also characterised by the attribution of some of the IR bands to specific adulterants. Particularly, the most discriminant peak at about 1690 cm^{-1}, which showed the highest VIP scores, was associated to the effect of NaCl addition to bovine meat causing specific aggregated β-sheets vibrations of proteins (Perisic et al., 2013).

Conflict of interest

None of the authors have any conflicts of interest.

Acknowledgments

Programa Ciências Forenses (“Pró-Forenses”), Edital no. 25/2014, from CAPES – Brazil for financial support. K.M.N. acknowledges a fellowship from CAPES. We also thank Flávia S. Coelho and LANAGRO (Laboratório Nacional Agropecuário, Pedro Leopoldo/MG, Brazil) for supporting physico-chemical analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodchem.2016.02.158.

References

ASTM, Annual Book of ASTM Standards. (2012). *Standard Practices for Infrared Multivariate Quantitative Analysis – E1655–05*.West Conshohocken: ASTM International.

Alamprese, C., Casale, M., Sinelli, N., Lanteri, S., & Casiraghi, E. (2013). Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT – Food Science and Technology, 53*, 225–232.

Ayadi, M. A., Kechaou, A., Makni, I., & Attia, H. (2009). Influence of carrageenan addition on turkey meat sausages properties. *Journal of Food Engineering, 93*, 278–283.

Ballin, N. Z. (2010). Authentication of meat and meat products. *Meat Science, 86*, 577–587.

Biancolillo, A., Bucci, R., Magri, A. L., Magri, A. D., & Marini, F. (2014). Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Analytica Chimica Acta, 820*, 23–31.

Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Analytica Chimica Acta, 891*, 1–14.

Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chemistry, 181*, 31–37.

Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics, 28*, 213–225.

Bro, R., Nielsen, H. J., Savorani, F., Kjeldahl, K., Christensen, I. J., Brunner, N., & Lawaetz, A. J. (2013). Data fusion in metabolomic cancer diagnostics. *Metabolomics, 9*, 3–8.

Cao, H., & Xu, S. Y. (2008). Purification and characterization of type II collagen from chick sternal cartilage. *Food Chemistry, 108*, 439–445.

Casale, M., Olivieri, P., Casolino, C., Sinelli, N., Zunin, P., Armanino, C., ... Lanteri, S. (2012). Characterisation of PDO olive oil Chianti Classico by non-selective (UV–visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques. *Analytica Chimica Acta, 712*, 56–63.

Cheng, Q., & Sun, C. W. (2008). Factors affecting the water holding capacity of red meat products: A review of recent research advances. *Critical Reviews in Food Science and Nutrition, 48*, 137–159.

Chong, I.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems, 78*, 103–112.

Cozzi, G., Ferlito, J., Pasini, G., Contiero, B., & Gottardo, F. (2009). Application of near-infrared spectroscopy as an alternative to chemical and color analysis to discriminate the production chains of Asiago d'Allevo cheese. *Journal of Agricultural and Food Chemistry, 57*, 11449–11454.

DPF (2012). Brasília: Departamento de Polícia Federal, Available in http://goo.gl/adPvdR, accessed in November 2015.

EC (2005). Commission Recommendation concerning a coordinated programme for the official control of foodstuffs. Official Journal of the European Union, L59 (pp. 27–39), Brussels.

Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International, 60*, 189–204.

FSAI (2013). FSAI survey finds horse DNA in some beef burger products. Dublin: Food Safety Authority of Ireland. Available in http://www.fsai.ie/news_centre/press_releases/horseDNA15012013.html, accessed in November 2015.

Godinho, M. S., Blanco, M. R., Gambarra Neto, F. F., Lião, L. M., Sena, M. M., Tauler, R., & Oliveira, A. E. (2014). Evaluation of transformer insulating oil quality using NIR, fluorescence, and NMR spectroscopic data fusion. *Talanta, 129*, 143–149.

Grundy, H. H., Reece, P., Sykes, M. D., Clough, J. A., Audsley, N., & Stone, R. (2008). Method to screen for the addition of porcine blood-based binding products to foods using liquid chromatography/triple quadrupole mass spectrometry. *Rapid Communications in Mass Spectrometry, 22*, 2006–2008.

Her Majesty's Government (2014). Elliott Review into the Integrity and Assurance of Food Supply Networks – Final Report. A National Food Crime Prevention Framework, London, UK.

Iwobi, A., Sebah, D., Kraemer, I., Losher, C., Fisher, G., Busch, U., & Huber, I. (2015). A multiplex real-time PCR method for the quantification of beef and pork fractions in minced meat. *Food Chemistry, 169*. 305-303.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*, 137–148.

Khanmohammadi, M., Ashori, A., Kargosha, K., & Garmarudi, A. B. (2007). Simultaneous determination of sodium tripolyphosphate, sodium sulfate and linear alkylbenzensulfonate in washing powder by attenuated total reflectance: Fourier transform infrared spectrometry. *Journal of Surfactants and Detergents, 10*, 81–86.

Küpper, L., Heise, H. M., & Butvina, L. N. (2001). Novel developments in mid-IR fiber-optic spectroscopy for analytical applications. *Journal of Molecular Structure, 563*, 173–181.

Lopez, M. I., Colomer, N., Ruisanchez, I., & Callao, M. P. (2014). Validation of multivariate screening methodology. Case study: Detection of food fraud. *Analytica Chimica Acta, 827*, 28–33.

Mamani-Linares, L. W., Gallo, C., & Alomar, D. (2012). Identification of cattle, llama and horse meat by near infrared reflectance or transflectance spectroscopy. *Meat Science, 90*, 378–385.

MAPA (1999). *Instrução Normativa SDA 20/1999, Secretaria de Defesa Agropecuária.*. Brasília: Ministério da Agricultura, Pecuária e Abastecimento (MAPA).

Pavia, D., Lampman, G., Kriz, G., & Vyvyan, J. (2008). *Introduction to Spectroscopy* (4th ed.)Stanford: Cengage Learning.

Perisic, N., Afseth, N. K., Ofstad, R., Narum, B., & Kohler, A. (2011). Monitoring protein structural changes and hydration in bovine meat tissue due to salt substitutes by Fourier transform infrared (FTIR) microspectroscopy. *Journal of Agricultural and Food Chemistry, 59*, 10052–10061.

Perisic, N., Afseth, N. K., Ofstad, R., Narum, B., & Kohler, A. (2013). Characterizing salt substitution in beef meat processing by vibrational spectroscopy and sensory analysis. *Meat Science, 95*, 576–585.

Pizarro, C., Rodríguez-Tecedor, S., Pérez-del-Notario, N., Esteban-Díez, I., & González-Sáiz, J. M. (2013). Classification of Spanish extra virgin olive oils by data fusion of visible spectroscopic fingerprints and chemical descriptors. *Food Chemistry, 138*, 915–922.

Ramos, P. M., Ruisánchez, I., & Andrikopoulos, K. S. (2008). Micro Raman and X-ray fluorescence spectroscopy data fusion for the classification of ochre pigments. *Talanta, 75*, 926–936.

Rinnan, A., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC – Trends in Analytical Chemistry, 28*, 1201–1222.

Rohman, A., Sismindari Erwanto, Y., & Che Man, Y. B. (2011). Analysis of pork adulteration in beef meatball using Fourier transform infrared (FTIR) spectroscopy. *Meat Science, 88*, 91–95.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*, 1627–1639.

Vera, L., Aceña, L., Guasch, J., Boqué, R., Mestres, M., & Busto, O. (2011). Discrimination and sensory description of beers through data fusion. *Talanta, 87*, 136–142.

Volery, P., Besson, R., & Schaffer-Lequart, C. (2004). Characterization of commercial carrageenans by Fourier transform infrared spectroscopy using single-reflection attenuated total reflection. *Journal of Agricultural and Food Chemistry, 52*, 7457–7463.

Wise, B. M., Gallagher, N. B., Bro, R., Shaver, J. M., Windig, W., & Koch, R. S. (2006). *PLS_Toolbox 4.0 for use with MATLAB™ (Manual)*.Wenatchee: Eigenvector Research Inc..

Zajac, A., Hanuza, J., & Dyminska, L. (2014). Raman spectroscopy in determination of horse meat content in the mixture with other meats. *Food Chemistry, 156*, 333–338.

Zotte, A. D., Ottavian, M., Concollato, A., Serva, L., Martelli, R., & Parisi, G. (2014). Authentication of raw and cooked freeze-dried rainbow trout (*Oncorhynchus mykiss*) by means of near infrared spectroscopy and data fusion. *Food Research International, 60*, 180–188.