Meat Science 200 (2023) 109169

Contents lists available at ScienceDirect

Meat Science

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

Across countries implementation of handheld near-infrared spectrometer for the *on-line* prediction of beef marbling in slaughterhouse

Moïse Kombolo-Ngah [a,1], Arianna Goi [b,*,1], Matteo Santinello [b], Nicola Rampado [b], Stefka Atanassova [c], Jingjing Liu [a], Pascal Faure [d], Laure Thoumy [e], Alix Neveu [f], Donato Andueza [a], Massimo De Marchi [b], Jean-François Hocquette [a]

[a] INRAE, Clermont Auvergne, Université Clermont Auvergne, VetAgro Sup, UMR1213, Recherches sur les Herbivores, 63122 Saint Genès Champanelle, France
[b] Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, PD, Italy
[c] Department of Agrarian Engineering, Faculty of Agriculture, Trakia University, 6000 Stara Zagora, Bulgaria
[d] INRAE, Herbipôle, Theix 63122, France
[e] IDELE, Boulevard des Arcades, 87000 Limoges, France
[f] IMR3GF International Meat Research 3G Foundation, Smulikowskiego St. 4/217, 00-389 Warsaw, Poland

A R T I C L E I N F O

Keywords:
MSA
Beef
NIRS
Meat quality
Longissimus thoracis
Marbling

A B S T R A C T

Only few studies have used Near-Infrared (NIR) spectroscopy to assess meat quality traits directly in the chiller. The aim of this study was therefore to investigate the ability of a handheld NIR spectrometer to predict marbling scores on intact meat muscles in the chiller. A total of 829 animals from 2 slaughterhouses in France and Italy were involved. Marbling was assessed according to the 3G (Global Grading Guaranteed) protocol using 2 different scores. NIR measurements were collected by performing 5 scans at different points of the *Longissimus thoracis*. An average MSA marbling score of 330–340 was obtained in the two countries. The prediction models provided a R^2 in external validation between 0.46 and 0.59 and a standard error of prediction between 83.1 and 105.5. Results did provide a moderate prediction of the marbling scores but can be useful in the European industry context to predict classes of MSA marbling.

1. Introduction

Beef production and consumption have a number of benefits on human health, food security, socioeconomic well-being of rural communities, and gastronomic pleasure of consumers (Hocquette et al., 2018). Despite all these, the beef production system faces a number of challenges due to the high variability of its palatability and a lack of a guarantee system for eating quality (Chriki et al., 2013; Legrand, Hocquette, Polkinghorne, & Pethick, 2013). As a result of these challenges, the Meat Standards Australia (MSA) grading scheme has been developed as an useful tool to assess beef palatability in commercial conditions from the 1990s onwards (Watson, Polkinghorne, & Thompson, 2008). The MSA grading system has become more popular and recognized by many countries as a pertinent innovation to predict beef eating quality (Bonny et al., 2018). The MSA system has investigated a set of critical control points to assess beef palatability through extensive consumer testing protocols which include factors as sex, carcass hanging method, cooking method and fat depth etc. (Bonny et al., 2018). The MSA protocol is set to be adapted to the European context through the 3G (Global Guaranteed Grading) protocol that uses the same control points. Among these control points, one of the most important components is marbling.

Marbling is usually seen as intramuscular deposit of fat within the muscle and it is assessed between the 5th and the 13th rib of the carcass according to the MSA and 3G protocols (Meat, Livestock, Australia & Meat Standards, Australia, 2001). In France and in Italy, it is measured at the 5th rib instead of the 10th rib in Australia (Liu et al., 2021). Marbling plays a great role in beef palatability as it has a positive relationship with flavor, tenderness, and juiciness even though its effect on these is not very clear (Calkins & Hodgen, 2007; Chriki et al., 2013; Guillemin et al., 2009). The French national food conference recently recommended the meat sector, represented by INTERBEV, to introduce marbling in the French beef grading scheme (Etats Generaux de l'Alimentation EGA, 2018). Actually, the MSA grading scheme is highly human-led with trained graders evaluating carcasses for different traits

* Corresponding author.
E-mail address: arianna.goi@unipd.it (A. Goi).

[1] Moise Kombolo Ngah and Arianna Goi contributed equally to this work.

https://doi.org/10.1016/j.meatsci.2023.109169
Received 10 September 2022; Received in revised form 14 March 2023; Accepted 16 March 2023
Available online 20 March 2023

(such as marbling). The MSA system therefore requires a grader in each slaughterhouse with the accreditation being expensive and requiring frequent trainings. It is therefore relevant to move towards a faster and more economic method which allows assessment on several slaughterhouses even in the absence of MSA accredited personnel. Indeed, the beef industry needs to look for the most effective and the less constraining methods to assess the quality of their product.

Near-infrared spectroscopy (NIRS) is a rapid, non-invasive, cost-effective, and environmentally friendly analytical technique which permits to obtain a complete picture of the organic composition of the analyzed matrix using an electromagnetic radiation in a known wavelength range (Van Kempen, 2001). Its principle is based on the absorption of the radiation which spans the wavelength range between 750 nm and 2500 nm by different chemical bonds (O—H, C—H, N—H, and S—H) in organic matter (Pasquini, 2003). When recorded, the light absorbed produces a spectrum which includes information related to chemical and physical properties of molecules of the sample and therefore information on sample composition (Prieto, Pawluczyk, Russell Dugan, & Aalhus, 2017). It has largely been studied and used in food industry as a rapid and convenient tool for the evaluation of quality (Fan, Liao, & Cheng, 2018), to assess fat, moisture, and protein content in ground beef (Su et al., 2014), as well as to predict tenderness and intramuscular fat in beef, pork and lamb (Fan et al., 2018; Prieto et al., 2014; Ripoll, Albertí, Panea, Olleta, & Sañudo, 2008). Near-infrared spectroscopy presents drawbacks due to its dependence on a reference method, its low sensitivity to minor constituents, limited transfer of calibration between different instruments and complicated spectral data (Büning-Pfaue, 2003; Givens, Boever, & Deaville, 1997). In addition, the spectral data analyses often require time-consuming and laborious calibration procedures along with the complexity of the choice of data treatment (Büning-Pfaue, 2003). It is therefore a challenge to tackle these drawbacks for its effective usage by building a strong calibration model constructed with data from on-line process to better reflect industrial variability. However, the recent development and diffusion of portable handheld devices has facilitated the application of this technology on a large scale, with rapid acquisition of spectra and avoiding the need for complex sampling protocols and transport of samples to a laboratory (Goi, Hocquette, Pellattiero, & De Marchi, 2022; Kademi, Ulusoy, & Hecer, 2018). Moreover, handheld spectrometers are very cheap, web-based with online storage, and thus easy to use in different slaughterhouses. Among NIRS' advantages, it allows simultaneous assessment of several traits, and it is suitable for on-line application within a processing plant without requiring sample preparation (Dixit et al., 2017). Indeed, NIRS shows to be a practical and low-cost solution easy to implement. A lot of studies have investigated the use of NIRS on ground samples of meat compared to intact samples where there are still a few of them dedicated to marbling. However, even less studies have investigated the application of NIRS technology to predict beef marbling traits in the slaughterhouse.

The purpose of this study was to investigate the use of a pocket-sized handheld near-infrared spectrometer in the slaughterhouse to predict marbling scores as they have been shown to have a relationship with eating quality and to discriminate animals according to their marbling class (Meat, and Livestock Australia, Meat Standards Australia, 2018).

2. Materials and methods

2.1. Experimental design

The study involved a total of 829 beef carcasses selected for their variability in terms of breed, age and weight to depict part of the intrinsic commercial variability on beef carcasses. They were provided by 2 commercial slaughterhouses located in Italy and France. From these animals, 400 carcasses were young Charolais bulls and heifers from different Italian fattening farms who have been processed in Italy (AZoVe, Cittadella PD). Four-hundred-twenty-nine carcasses were obtained from animals reared and processed in France (CV Plainemaison- Beauvallet, Limoges), among them there were 277 Limousin cows or heifers and 152 animals randomly selected to increase variability in animal type and breed (Charolais, Holstein, Aubrac, crossed).

2.2. Marbling measurement

All the processed carcasses were assessed after 24 h of *post-mortem* chilling in a refrigerated room with an average temperature between 0 and 4 °C. Marbling was assessed according to specifications of the ABCAS Reference Standards (Meat, Livestock, Australia & Meat Standards, Australia, 2001) and a benchmark for the measurement of the main quality characteristics of the bovine carcasses adopted by the UNECE Bovine Language standards through 2 scores (AUS-MEAT and MSA marbling scores). The chiller assessment standards are adapted for European cattle and consumers through extensive collaborative research in Europe with data being stored and utilized through the IMR3GF. The AUS-MEAT marbling score reflects the amount of marbling with score ranging from 0 to 9 in increments of one. The MSA marbling provides a more precise scale (from 100 to 1190 in increments of 10) indicating the amount, the size, the fineness, and the distribution of fat inclusions in muscles. The AUS-MEAT and the MSA marbling scores were assessed at the 5th rib of the carcass as described by (Liu et al., 2021). The assessment was done by two 3G chiller assessors (one for Italy and one for France) accredited by the International Research Meat 3G Foundation. They followed the same training session and have been always updated according to ABCAS (Australian Beef Chiller Assessment System) standards, to reduce the technical variability of their evaluations. The rib eye was exposed to air for at least 20 min up to 3 h after cutting, allowing the meat to bloom before the MSA marbling assessment carried out using a visual standard. Blooming refers to the color change due to oxygenation of myoglobin that occurs when a meat surface is exposed to oxygen (Jacob, 2020). The grading is done with visual standard cards (provided by ABCAS) to assess AUS-MEAT and MSA scores.

2.3. Near-infrared spectroscopy measurements

Spectra were recorded using the SCiO™ molecular sensor (Consumer Physics Inc., Tel Aviv, Israel), a handheld web-based wireless spectrometer that operates in reflectance mode in the NIR region between 740 and 1070 nm (13,514 and 9346 cm^{-1}) of wavelength at intervals of 1 nm (13 cm^{-1}) which is the default mode of the spectrometer. The size of the area analyzed by the sensor is 1 mm^2. The instrument was calibrated every day before starting the spectra collection, then five scans per sample were taken in 5 different positions on the surface of *Longissimus thoracis* muscle by applying an adapter to the scanning head in order to maintain a fixed distance from the sample of 1 cm over the surface and avoid any external light source. The average time to scan each animal was roughly 1 min. An internet connection was needed for the scans to be saved. The two 3G chiller assessors followed the same approach choosing each position according to the dimension of the rib eye to be representative of the entire surface. Reflectance values were collected through the SCiO™ online application ("The Lab") (Consumer Physics Inc., Tel-Aviv, Israel), converted to absorbance as log(1/reflectance) and each final spectrum to be used for the development of prediction models was then calculated as the average of the 5 scans.

2.4. Chemometrics analysis

Chemometric analysis was performed to obtain the prediction models for MSA and AUS-MEAT marbling scores using spectral and reference data. Several processes have been used to develop these models. The first process used WinISI 4 software (Infrasoft International, Port Matilda, PA, USA) where the calibrations were developed through modified partial least square (mPLS) regression analysis. The second process used the Pirouette 4.5 software (Infometrix, Inc., Woodinville, WA, USA) where calibrations were developed with a classical PLS regression analysis. Both processes used spectral information as independent variables, and correlated with the corresponding reference values to predict MSA marbling and AUS-MEAT scores. With the first process using the WinISI 4 software, the scores were predicted for the French and Italian datasets individually; then, the Italian and French datasets were also merged to get a wider variability. However the results from the merged dataset must be interpreted with caution since differences could arise from carcass graders, instruments or slaughterhouses conditions as in the study performed by Fowler, Wheeler, Morris, Mortimer, and Hopkins (2021) and Stewart et al. (2021).

Firstly, spectral outliers were eliminated using the Mahalanobis distance (Global H > 3.0). Then, different scatter corrections such as detrending (D), standard normal variate (SNV), and multiplicative scatter correction (MSC) were applied to the raw spectra as pretreatment to reduce noise effect and variability due to the scatter (De Marchi, Costa, Goi, Penasa, & Manuelian, 2019). Afterwards, each of these pre-processing techniques was combined with derivative mathematical treatments: 0,0,1,1; 1,4,4,1; 1,8,8,1; and 2,10,10,1. Here, the first digit represents the number of the derivative, the second represents the gap over which the derivative is calculated, the third is the number of data points in the first smoothing and the fourth is the number of data points in the second smoothing (Shenk and Westerhaus, 1991). After the mPLS regression analysis, 2 rounds of outliers' elimination were performed, and the critical T-statistic value was set to ±2.5 standard error to define chemical outliers. Two rounds of outlier elimination was the best compromise to keep sample variability. Fernández-Cabanás, Polvillo, Rodríguez-Acuña, Botella, and Horcada (2011) also used two rounds of outlier removal and stated that, according to their previous works, these consistent anomalies can be detected for different products with the first two runs. Hence, when predicted values were greater than ±2.5 standard error from the reference values, the sample was removed from the dataset to be used for the final calibration.

The prediction models were tested within and across countries performing both a 15-fold cross-validation and an external validation. For the former, the dataset of each country individually and the merged dataset were split into 15 equal groups with similar mean and standard deviation, one part was used to validate the model created using the remaining groups and this process was repeated until each group was used as validation set once. For the latter, the division into calibration and validation set was done in 2 ways: (i) the full French, Italian, and merged datasets were randomly divided, ensuring a similar mean and standard deviation for both MSA and AUS-MEAT scores, in a subset of 64% of samples to produce a prediction equation to be tested on the remaining 34% of samples; (ii) each country was used alternatively as calibration and test set for the other, i.e. French dataset was used to create the model to be tested on the Italian dataset and vice versa.

With the second process using Pirouette 4.5 (Infometrix, Inc., Woodinville, WA, USA), the samples were also divided into calibration and validation sets. Two-third of the samples were selected randomly and used as calibration data set. The sample with the largest MSA value was included in the calibration set and the rest of the samples were used as validation data set.

Orthogonal signal correction (OSC) factor was used during the development of the regression equations and this preprocessing technique is used to remove information unrelated to the target variables based on constrained principal component analysis. Orthogonal signal correction has been shown to provide better results when compared to traditional signal correction as well as with those of no pre-processing (Wold, Antti, Lindgren, & Öhman, 1998).

Different spectral data transformed were tested – smoothing, first and second derivatives. The number of PLS factors in the equations was determined by leave-one-out cross validation. The leave-one-out cross-validation procedure works by omitting one observation, recalculating the equation using the remaining data, and then predicting the omitted observation. This routine is repeated until each observation in the data set is used once as validation data.

Thus, for both processes, the best model was assessed according to the optimal number of latent variables (LV) chosen based on the lowest root mean square error, the bias, the coefficient of determination of calibration (R^2_C), of cross-validation (R^2_{CV}), and of external validation (R^2_{Ext}), the standard error of calibration (SE_c), of cross validation (SE_{cv}), and of external validation (SE_p), and the residual predictive deviation of cross-validation (RPD_{CV}) and of external validation (RPD_{EXT}), calculated as the ratio between the standard deviation (SD) of the validation dataset and SE_{CV} and SE_P, respectively. Hence, we had one goodness-of-fit statistics output for each approach (i.e. within country and across countries). The normality of the residuals for the predicted models was assessed and a *t*-test was performed to determine whether the bias differed statistically from zero.

2.5. Discriminant analysis

In addition, another approach conducted was to apply a partial least square discriminant analysis (PLS-DA) on 2 datasets. The first dataset was made of 677 samples, of which the sex of the animal was known. The second dataset was made of all of the 829 samples, with unknown information about breed and sex of the animals. The NIR data were analyzed according to their spectrum after segregating the carcasses in 3 MSA marbling homogeneous classes (low, medium, high) according to their frequencies.

Using the first dataset, each class was created within sex and breed to avoid their effect on the MSA score, obtaining classes containing about 33% of samples. In the second dataset, the classes were created by taking the average of the thresholds for Limousine and Charolais classes.

Discriminant analysis on AUS-MEAT could not provide satisfactory results and were therefore not shown. Subsequently, meat samples were randomly split into a training and a validation set to build the model in absence of the validation set and therefore obtain a prediction which is independent of the model optimization (Westerhuis et al., 2008). Calibration set created from the first dataset included 166, 172, and 170 samples with low, medium, and high MSA marbling value, respectively. Then 54, 57, and 58 samples belonging to low, medium, and high class respectively were assigned to the validation set. On the other hand, 231, 189, and 202 samples from the second dataset, with low, medium, and high MSA marbling value respectively were used for calibration set; the validation set was created with 63, 67, and 77 samples belonging to low, medium, and high class, respectively. The calculated thresholds for each MSA class created within breed and sex were reported in Table 1.

Using SAS software v. 9.4 (SAS Institute Inc., Cary, NC, USA), the

Table 1
Thresholds of MSA marbling classes according to the category of animals.

Dataset 1	Charolais			Limousin		
MSA Class	Low	Medium	High	Low	Medium	High
Male	MSA ≤ 270	270 < MSA ≥ 340	MSA > 340			
Female	MSA ≤ 310	310 < MSA ≥370	MSA > 370	MSA ≤ 290	290 < MSA ≥ 370	MSA > 370
Dataset 2[a]	All animals					
MSA Class	Low	Medium	High			
	MSA ≤ 290	290 < MSA ≥360	MSA > 360			

[a] Dataset 2 includes data from all animals without considering any information about breed and sex

elimination of spectral outliers on the training set was performed using the Mahalanobis distance and the parameters of the PLS-DA model were optimized using leave-one-out cross-validation on the training dataset. The process was repeated until the addition of more components to the model did not produce a decrease in the cross-validation error rate. The PLS algorithm was followed by the adoption of the DISCRIM procedure using the PLS score as predictors to classify samples of the validation set. The confusion table which describes the classification performance of the discrimination model was then evaluated with three outcomes: sensitivity, specificity, and overall accuracy for each class.

$$Sensitivity\ (\%) = \frac{TP}{TP + FN}\ X\ 100$$

$$Specificity(\%) = \frac{TN}{FP + TN}\ X\ 100$$

$$Accuracy\ (\%) = \frac{TP + TN}{TP + TN + FP + FN}\ X\ 100$$

where FP, FN, TP and TN, are the number of false positives, false negatives, true positives and true negatives, predicted in external validation, respectively.

Sensitivity and specificity are used to calculate the true rates, in fact the sensitivity is the model ability to correctly associate the sample to the real class to which it belongs, whereas specificity is the ability to correctly identify which class the sample does not belong to (Almeida, Fidelis, Barata, & Poppi, 2013). Accuracy is the overall proportion of samples correctly classified. The overall accuracy of the confusion matrix was also assessed with receiver operating characteristic (ROC) curves where AUC-ROC corresponds to the area under a ROC curve and a single value measures the overall performance of a binary classifier (Hanley & McNeil, 1982). It ranges from 0.5 to 1 where the lowest value represents a random classifier and the maximum value represents a perfect classifier (Hanley & McNeil, 1982). Three AUC-ROC curves were measured to evaluate the ability of the model to discriminate between the different classes and this was done for both models. The two PLS-DA were then compared to see which one provides the best results. The quality of the PLS-DA models was evaluated with the following parameters (R^2 and Q^2) representing the explained variance and the predictive capability of the model, respectively. R^2X and R^2Y represent the fraction of variance of the X and Y matrix, respectively. Q^2Y represents the predictive accuracy of the model, with the cumulative (cum) values of R^2X, R^2Y and Q^2Y equating to ~1 indicating an effective model (Kong et al., 2015). Permutation tests were performed with SIMCA 17.0 (Sartorius, 2020) on the validation set. Hundred permutations were done for each model and to reduce the risk of overfitting. Plots showing the correlation coefficients between the original Y and the permuted Y versus the R^2Y and Q^2 were obtained.

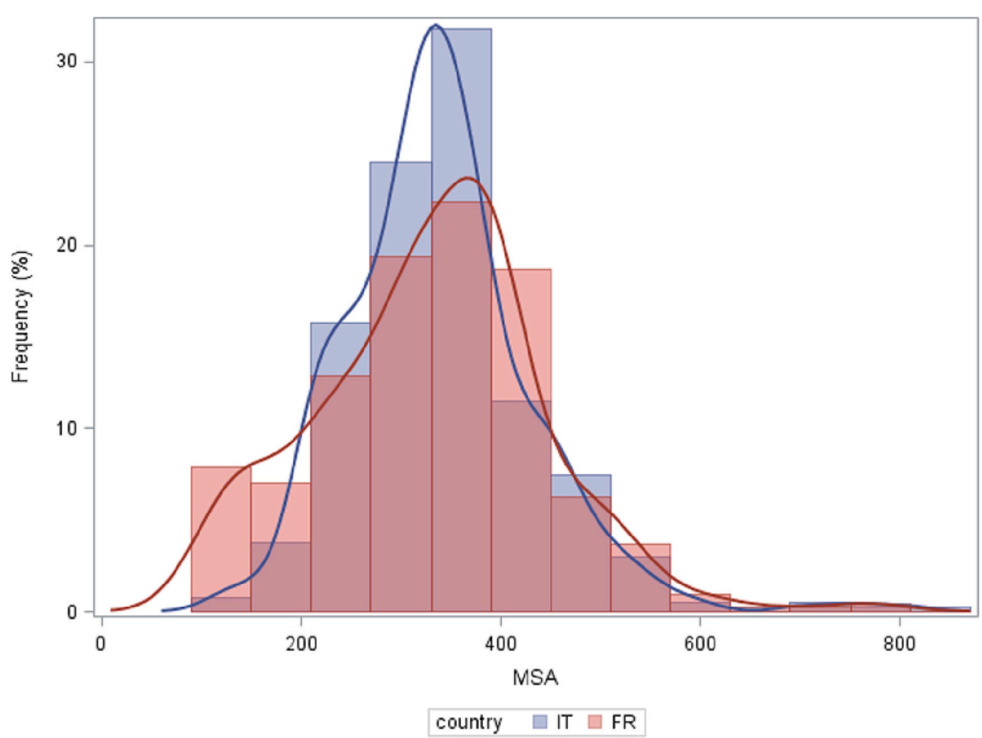
The fitted regression lines were displayed, connecting the observed Q^2 to the centroid of permuted Q^2 cluster. The model is considered valid if all Q^2 values from the permuted data set to the left are lower than the Q^2 value on the actual data set to the right and if the regression line has a negative value of intercept on the y-axis.

3. Results

3.1. Marbling measurement and near-infrared spectra

The distribution of MSA and AUS-MEAT scores within country and relative descriptive statistics are shown in Figs. 1 and 2. French data showed a slighter lower mean but a greater coefficient of variation (CV) than Italian ones (Fig. 1). Fig. 2 shows samples from each country grouped by classes of AUS-MEAT, specifically class 0, 1, 2 and > 2.

The average raw absorbance spectrum for French and Italian



Country	n	Mean	SD	Minimum	Maximum	CV
France	429	330.42	114.41	100.00	780.00	34.63
Italy	400	339.40	93.45	120.00	810.00	27.53

Fig. 1. Distribution of MSA marbling scores and descriptive statistics within country.

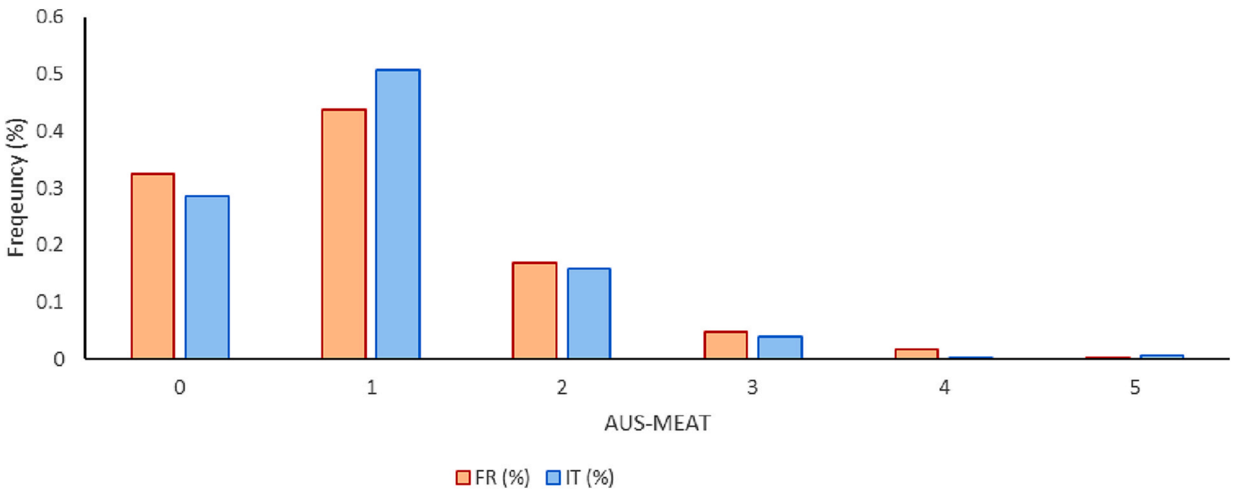


Fig. 2. Distribution of AUS-MEAT marbling scores within country.

carcasses are depicted in Fig. 3, showing the same trend of the curves with a narrow peak observed at 762 nm (13,123.36 cm^{-1}) and a wider one around 984 nm (10,162.60 cm^{-1}), but a difference in absorbance. The total Italian (a) and French (b) spectra colored according to the level of marbling are shown in Fig. 4.

3.2. Near-infrared spectroscopy prediction models

Statistics of the best prediction for each country and for the 2 datasets merged are shown in Table 2 with the first process using WinISI 4 and Table 3 with the second process using Pirouette 4.5.

With the first process using WinISI 4, the model was developed using a 15-fold cross-validation. No scatter correction was applied for MSA and AUS-MEAT scores except for the prediction of the latter in the French dataset, where SNV was used. Latent variables ranged from 5 to 7 for MSA and from 7 to 9 for AUS-MEAT. Overall, the equations for MSA marbling were developed by applying a first derivative, over a gap of four data points for the individual countries and 8 data points for the merged dataset. On the other hand, the best models for AUS-MEAT were obtained without calculating derivative for the French and Italian dataset but using the first order of derivatization applied over a gap of 8 data points for the complete dataset with both countries merged. The outliers detected were < 11.2% for all the cases. The best prediction model for MSA scores was obtained using data from France ($R^2_{CV} = 0.58$, RPD = 1.69), whereas the best one for AUS-MEAT scores was developed using Italian samples ($R^2_{CV} = 0.47$, RPD = 1.33).

The external validation performed within and across countries confirmed the results obtained from the cross-validation (Table 4). The best R^2_{CV} and R^2_{Ext} for MSA marbling were reached using the French dataset but overall, final prediction accuracies did not really differ between countries nor using the merged dataset. Low prediction accuracy ($R^2_{Ext} < 0.3$) was obtained for AUS-MEAT scores when performing external validation using all the datasets, therefore results were not reported in the table and no other calibrations were developed for the AUS-MEAT parameter. Lastly, when external validation was carried out across countries, the LV used for the development of MSA model were 5 for France and 7 for Italy and the results differed significantly. In particular, the Italian data allowed the development of the best prediction model ($RPD_{Ext} = 1.14$) once used as calibration set and then the equation has been validated on the French dataset. The linear regression of MSA marbling reference versus predicted values when performing the external validation within country is shown in Fig. 5. Residual of prediction equations were normally distributed, and bias did not differ statistically from zero.

With the second process using Pirouette 4.5, the spectra from both countries were pre-processed by using a first derivative. Both datasets provided similar R^2_{CV} (0.59 for France and 0.58 for Italy). Detailed results are shown in Table 3. These results are similar from the results of the first process.

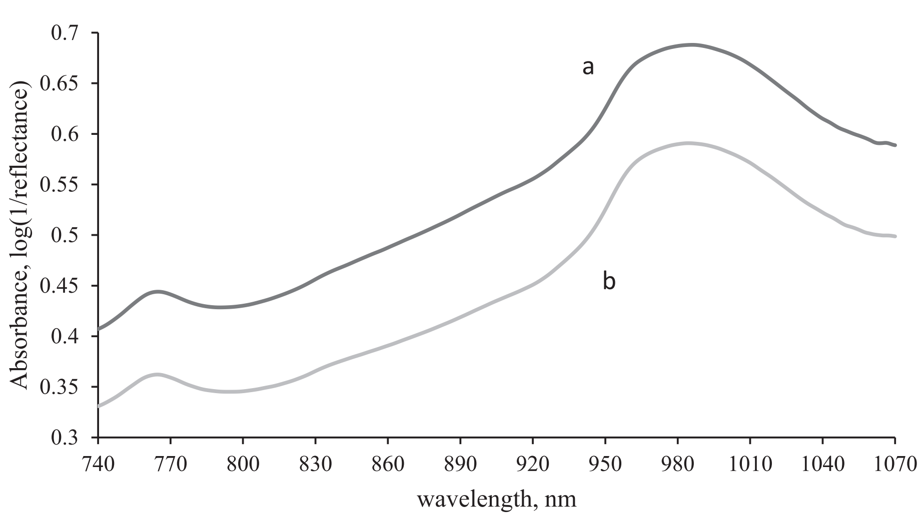


Fig. 3. Average raw NIR spectra of French (a) and Italian (b) datasets.

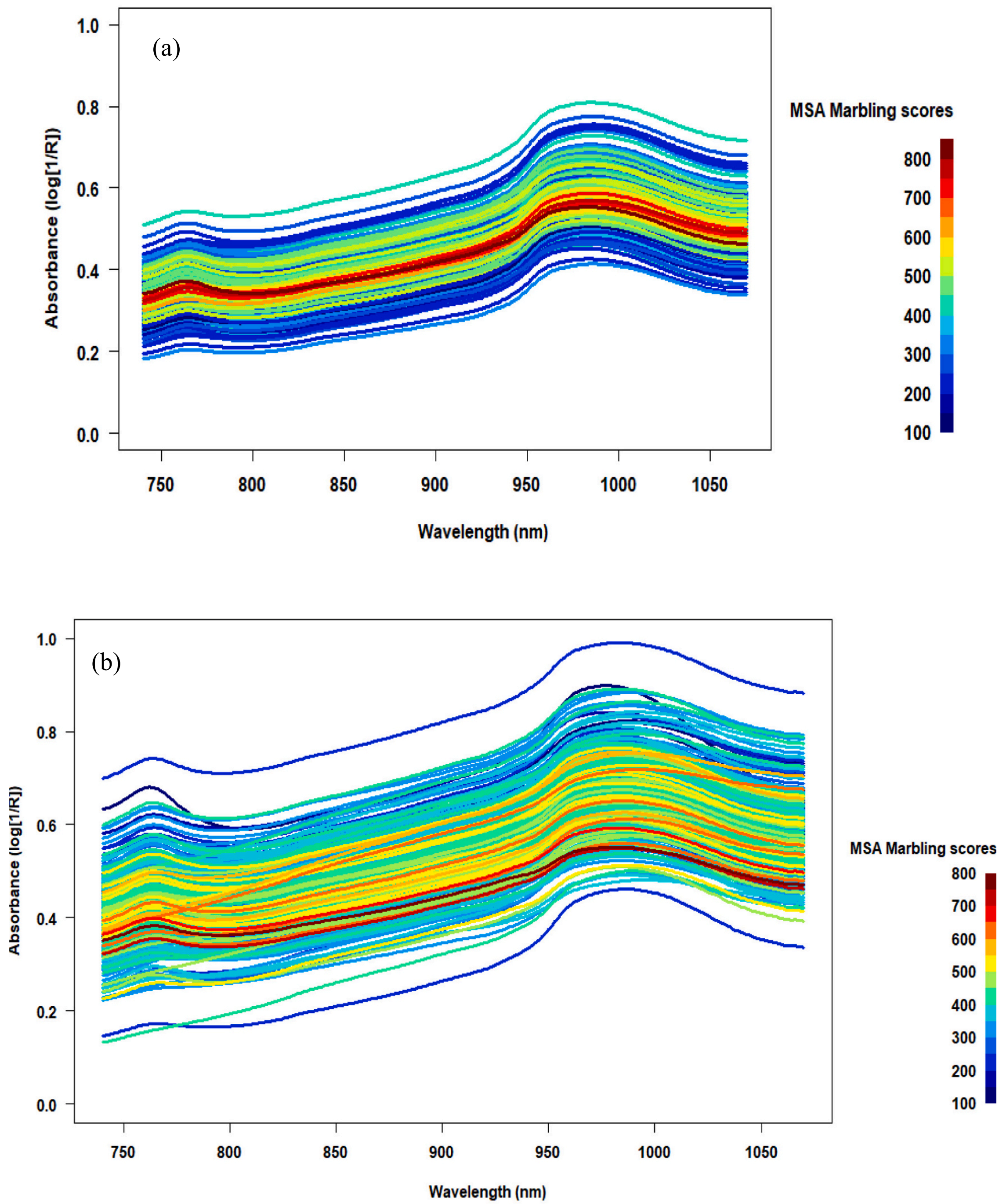


Fig. 4. Raw NIR spectra of (a) Italian samples and (b) French samples, colored according to the MSA marbling score.

3.3. MSA marbling classes discrimination

Tables 5 and 6 show the results obtained from the PLS-DA models for the validation set created selecting randomly 25% of the samples in both datasets. Similar results were obtained for both datasets using 11 LV. The greatest sensitivity was reached for samples correctly predicted to have low MSA scores (72.2% in the dataset with known sex and breed [Table 5] and 66.7% in the dataset with random sex and breed [Table 6]). Lower sensitivity was obtained for samples belonging to medium (52.6% in first dataset [Table 5] and 52.2% in the second dataset [Table 6])) and high MSA classes (58.6% in the first dataset [Table 5] and 63.6% in the second dataset [Table 6]). On the other hand, specificity ranged from 74.4% to 83.1% for all the classes. The confusion matrix for both datasets had an overall accuracy of 61%.

The AUC-ROC curves for both datasets also provided similar results. The AUC values for the prediction for the low and high classes were above 0.8 in both datasets. Results for the AUC values are shown in Fig. 7. The prediction for the medium classes had values above 0.5 in both datasets but <0.8 (Fig. 7).

The permuted R^2 and Q^2 for the low and high classes were lower than the original values on the right (Fig. 8(a); Fig. 8(c) respectively). This suggests that the model fitting was valid, and this was unlikely to be built randomly. The permuted R^2 and Q^2 for the medium classes had similar values to the original showing that the prediction for the medium class could be built randomly (Fig. 8(b)).

Table 2
Fitting statistics of modified partial least squares regression models (mPLS) in 15-fold cross-validation for MSA marbling score and AUS-MEAT score within and across countries, developed using near-infrared reflectance spectroscopy on beef *Longissimus thoracis*.

Trait	Country	Scatter correction[a]	MT[b]	n[c]	outliers	LV[d]	SE_C[e]	R^2_C[f]	SE_{CV}[g]	R^2_{CV}[h]	RPD_{Ext}[i]
MSA	Italy	None[j,k]	1,4,4,1	400	19	7	54.49	0.57	57.92	0.51	1.61
	France	None	1,4,4,1	429	48	5	66.05	0.60	67.84	0.58	1.69
	Italy + France	None	1,8,8,1	829	78	7	64.46	0.56	65.64	0.54	1.59
AUS-MEAT	Italy	None	0,0,1,1	400	20	8	0.55	0.50	0.57	0.47	1.33
	France	SNV[k]	0,0,1,1	173	19	9	0.62	0.48	0.69	0.36	1.02
	Italy + France	None	1,8,8,1	573	61	7	0.57	0.47	0.58	0.44	1.27

[a] Scatter correction = pre-processing technique to reduce noise.
[b] MT = mathematical treatment (first digit indicates the derivative treatment).
[c] n = total number of samples.
[d] LV = optimal number of latent variables.
[e] SE_C = standard error of calibration.
[f] R^2_C = coefficient of determination of calibration.
[g] SE_{CV} = standard error of cross-validation.
[h] R^2_{CV} = coefficient of determination of cross-validation.
[i] RPD = ratio of performance to deviation.
[j] none = no correction.
[k] SNV = standard normal variate.

Table 3
Statistics for PLS models for prediction of meat MSA marbling with Pirouette 4.5.

Trait	Country	Scatter correction[a]	MT[b]	n[c]	outliers	LV[d]	SE_C[e]	R^2_C[f]	SE_{CV}[g]	R^2_{CV}[h]
MSA	Italy	OSC	1D	400	4	4	56.76	0.51	54.09	0.58
	France	OSC	1D	429	0	9	76.15	0.53	70.71	0.59

[a] Scatter correction = pre-processing technique to reduce noise.
[b] MT = mathematical treatment (first digit indicates the derivative treatment).
[c] n = total number of samples.
[d] LV = optimal number of latent variables.
[e] SE_C = standard error of calibration.
[f] R^2_C = coefficient of determination of calibration.
[g] SE_{CV} = standard error of cross-validation.
[h] R^2_{CV} = coefficient of determination of cross-validation.

Table 4
Fitting statistics of modified partial least square regression models using external validation for MSA marbling within and across countries.

		Within countries								
		Calibration set (66%)				Validation set (34%)				
Trait	Country	LV[a]	SE_{CV}[b]	R^2_{CV}[c]		Bias	Slope	SE_P[d]	R^2_{Ext}[e]	RPD_{Ext}[f]
MSA	Italy	7	57.27	0.51		10.64	0.97	72.37	0.46	1.34
	France	10	65.96	0.60		−2.65	0.88	83.1	0.46	1.36
	Italy + France	10	64.49	0.57		−4.72	0.96	79.89	0.44	1.23
Across countries										
MSA	France	5	67.84	0.58	Italy	78.5	0.86	105.47	0.44	0.89
	Italy	7	57.92	0.51	France	−26.44	0.89	100.11	0.29	1.14

[a] LV = optimal number of latent variables.
[b] SE_{CV} = standard error of cross-validation.
[c] R^2_{CV} = coefficient of determination of cross-validation.
[d] SE_P = standard error of external validation.
[e] R^2_{Ext} = coefficient of determination of external validation.
[f] RPD_{Ext} = ratio of performance to deviation in external validation.

4. Discussion

The objective of this study was to develop NIRS prediction models for MSA and AUS-MEAT marbling scores with spectra collected using a portable and handheld NIR spectrometer. The spectra were collected in the slaughterhouse and validated within and across two countries mainly for the two major French beef breeds (Limousine and Charolais). It was ensured that the variability among the instruments used in the two countries did not make the comparison flawed since NIR spectrometers often require standardization methods (Bouveresse & Massart, 1996; Shenk & Westerhaus, 1991). We therefore carried out the measurements on the same carcasses with the same SCiO instruments. Inter and intra instrument variability was negligible and therefore we concluded that the instruments perform similarly.

The spectra were collected in reflectance mode and then recorded as absorbance, which is calculated as $\log_{10}(1/\text{Reflectance})$. In the literature, NIR reflectance spectroscopy showed a higher accuracy in predicting fat content using both benchtop (Alomar, Gallo, Castañeda, &

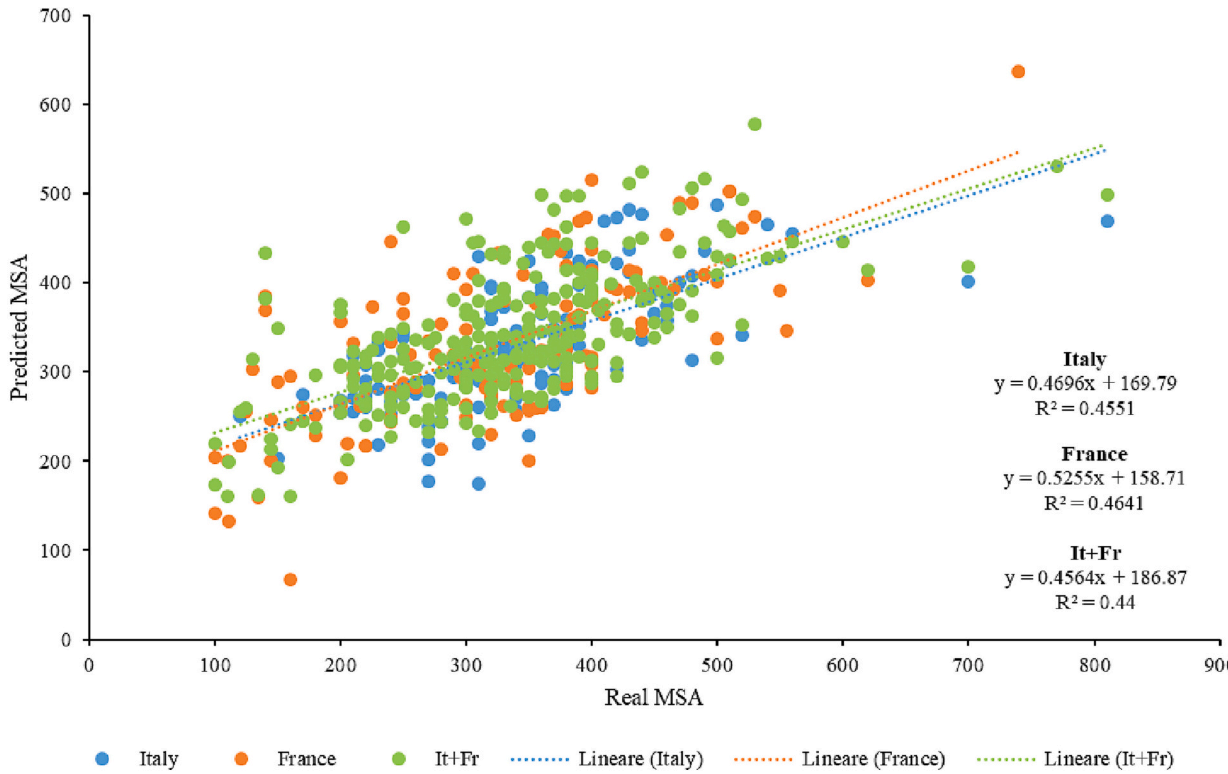


Fig. 5. Real vs. Predicted MSA marbling score values when performing external validation within country (data from Table 3).

Table 5
Confusion matrix for MSA marbling classes with Model performance using samples with known information of animal sex and breed.

MSA class	Low (predicted)	Medium (predicted)	High (predicted)	Total
Low (Actual)	**39**	13	2	54
Medium (Actual)	15	**30**	12	57
High (Actual)	7	17	**34**	58
Model performance, %				
Sensitivity	72.2	52.6	58.6	
Specificity	74.4	70.9	83.1	
Overall Accuracy		61		

Table 6
Confusion matrix for MSA marbling with samples of different animals without any information of breed and sex.

MSA class	Low (predicted)	Medium (predicted)	High (predicted)	Total
Low (Actual)	**42**	18	3	63
Medium (Actual)	21	**35**	11	67
High (Actual)	3	25	**49**	77
Model performance, %				
Sensitivity	66.7	52.2	63.6	
Specificity	77.8	67.9	84.6	
Overall Accuracy		60.9		

Fuchslocher, 2003; Ripoll et al., 2008) and handheld (Goi, Hocquette, Pellattiero, & De Marchi, 2022) devices in beef and also in other species (Dixit et al., 2017).

This study had a relatively large sample size of 829 animals compared to other studies mostly ranging between 30 and 150 (Prevolnik, Čandek-Potokar, & Škorjanc, 2011). A large sample size accounting for high variance for the predicted traits is crucial to develop robust models.

4.1. Near-infrared spectra and prediction models

4.1.1. Spectra specificities

The average NIR spectra reported in Fig. 3 ranges between 740 and 1070 nm, which corresponds to the Herschel region. According to Beć, Grabska, and Huck (2020) and Schrieve, Melish, and Ullman (1991) this spectral region is characterized by a narrow range, a low to moderate chemical specificity, a low sensitivity, an overlapping contributions in the spectra, and by difficulties in spectral interpretation. Beć et al. (2020) stated that meat includes a certain number of chromophores that are oxymyoglobin, water, fat, and protein which have an absorption band in the NIR region. In our study, the VIP profile indicated the important regions of the spectra, considering 1 as a threshold for MSA values (Fig. 6). VIP scores indicate the importance of each variable in the projection used in a PLS (Banerjee et al., 2013) model and is often used for variable selection. The most important parts of the spectra were: 740–772 nm, 826–843 nm, 862–867 nm, 918–948 nm, 963–969 nm, and 1054–1070 nm. The peak observed at 762 nm (13,123.36 cm^{-1}) is assumed to be related to OH third overtone or produced by the oxidation of myoglobin or deoxymyoglobin as stated by Cozzolino and Murray (2002), while according to Allen, Hall, Dhillon, Owen, and Beard (2012) the band nearby 970 nm (10,309.28 cm^{-1}) nm is related with OH second overtone associated to the water content in the samples. The absorption peak at 920 nm (10,869.57 cm^{-1}) has been associated with the 2nd overtone of the stretching vibrational mode of the C—H bond (Wilson, Nadeau, Jaworski, Tromberg, & Durkin, 2015). Lipids and water consist

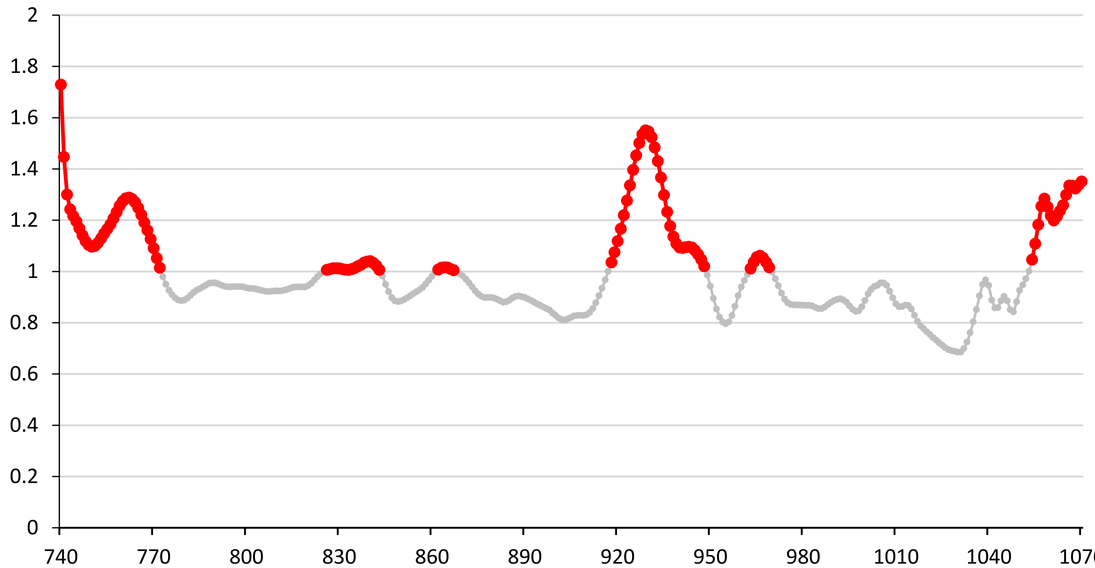


Fig. 6. Variable Importance in the Projection (VIP) profile of the selected Partial Least Square Regression (PLSR) model predicting MSA marbling scores. X axis : Wavelength (nm) Y axis : Variable Importance in the projection.

of several overlapping peaks as a result of C—H bond vibrations within the different structural groups (single bond CH3, >CH2, triple bond CH and > CH (aromatic)) of the lipid molecules (Jansen, Wu, van der Steen, & van Soest, 2014). Therefore, our data will be mostly influenced by the level of oxidation of myoglobin and water content of each carcasses as demonstrated by Wold et al. (2022). It has been shown that the concentration of myoglobin as well as its level oxygenation in meat has a strong impact on the NIR spectra and can give deviations in the estimated marbling scores (Wold et al., 2022). This impact of myoglobin oxidation should therefore be considered for the industrial implementation of NIR spectrometer.

4.1.2. Model validation

With the first process, when performing the MSA calibration, the deleted outliers ranged between 4.8 and 11.2% and 5 LV were used for the French dataset which was slightly lower than that used to obtain the best model for the Italian and the merged dataset (Table 2).

With the second process, 9 LV were used for French dataset and 4 LV for Italy, while 4 outliers were removed from the Italian dataset and no outlier from the French one.

In order to validate the model, the interpretation of R^2_{CV} was based on the work by Karoui et al. (2006). This study reported that a prediction model with R^2_{CV} between 0.66 and 0.81 could give an approximate quantitative estimation of the reference value, a R^2_{CV} between 0.82 and 0.90 could give a good estimation, and above 0.91 could give an excellent estimation. With the first process and when performing cross-validation, samples from France provided the greatest accuracy with both processes. However, these samples had the highest SE_{CV}. This higher SE_{CV} could be explained by the fact that samples from France included animals from different sexes and breeds. The Italian samples had the lowest SE_{CV} since the animals all had the same characteristics. Sample uniformity can therefore help to reduce SE_{CV}.

The interpretation of RPD was based on the criteria established by Williams (2014) who indicated that prediction model with an RPD below 1.9 is not recommended to be used, an RPD between 2.5 and 2.9 could be applied for screening, an RPD above 3.0 could be adequate for quality control, and a RPD above 4.0 could be adequate for any application. This parameter provides a metric of model validity which is more objective since it is a non-dimensional statistic. When performing cross-validation, RPD value was at most 1.69. Thus, we can observe that the prediction models for MSA marbling provided better results than for AUS-MEAT marbling which had RPD of <1.4 in all datasets. When considering the RPD for MSA marbling, no difference was detected between the Italian and the merged dataset.

When performing the external validation within country (Table 4), R^2_{Ext} was slightly lower than using cross-validation as expected, and all the datasets achieved an almost equal value between them as reported in Fig. 5. The same occurred with the RPD. When performing the validation across countries, the more accurate model when considering the R^2_{Ext} was developed using the French dataset and validated on the Italian one, however due to the greater SE_P it resulted being unsatisfactory with reference to RPD_{Ext}. Furthermore, the small variation among the samples (standard deviation of 0.15) most likely decreased the predictive performance and robustness of the calibration model, generating a low R^2.

Overall, according to the criteria of estimation of goodness of the models, similar results were obtained from both processes. OSC did provide better results as discussed by Wold et al., 1998. These models cannot be considered good enough to be applied in place of marbling evaluation by graders. Nevertheless, low or moderate phenotypic correlations between the reference and its prediction may not reflect the genetic association that exists between them (Costa, Visentin, De Marchi, Cassandro, & Penasa, 2019). Moreover, studies on dairy cows have demonstrated that infrared predictions could be weakly associated with the measured trait at either phenotypic and genetic level, but still their association can be considered good for selection purposes (Benedet, Costa, De Marchi, & Penasa, 2020; Costa et al., 2021). Therefore, although the prediction accuracy is not sufficient for a punctual determination, the prediction can be used as a proxy of the real target trait and be used to discriminate between low and high values of MSA as well as for genetic purposes.

For a chiller assessor to gain/retain ABCAS accreditation, he must be able to classify correctly 70% of carcasses within ±50 marbling units from the reference score and 100% between 100 units of the reference score (Rosenvold, 2012). From this prediction model, we therefore investigated what percentage of the data was correctly predicted in the range ± 50. According to our results, only about 50% of marbling scores were correctly predicted within ±50 units and 83.4% within ±100. Therefore, SCiO spectrometer cannot be accredited for the assessment of marbling based on the current prediction model.

4.2. Technical limitations

As shown by different studies, the efficiency of prediction on meat

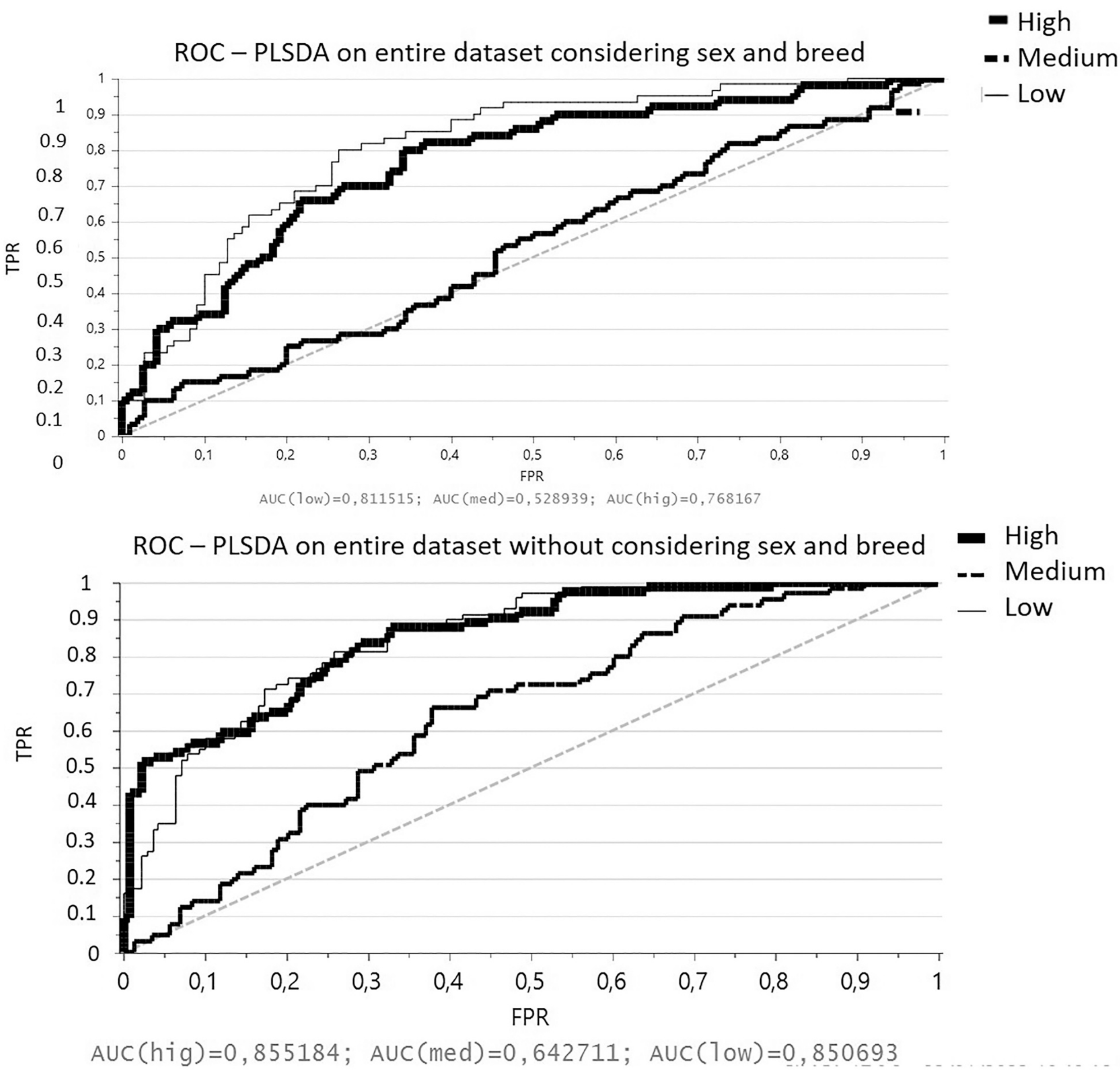


Fig. 7. ROC curves for the prediction for each class (a) considering sex and breed' (b) without considering sex and breed. TPR = True positive rate, FPR = False Positive rate.

depends on the state of the sample, whether minced or intact. Better predictions are made with minced samples due to a more homogeneous mixture (Barlocco, Vadell, Ballesteros, Galietta, & Cozzolino, 2006; Prevolnik, Čandek-Potokar, & Škorjanc, 2004). On intact sample, the NIR signal is therefore affected by the heterogeneity of the sampling site (Przybylski & Hopkins, 2015). Homogenization gives better results because fiber arrangement has been destroyed and randomized thereby averaging the effect of scattering but this is a strong limitation for on-line applications (Tøgersen, Arnesen, Nilsen, & Hildrum, 2003). Intact muscles, on the other hand, can act as optical fibers that conduct light along their lengths and reduce reflectance (Przybylski & Hopkins, 2015). This can therefore explain why our prediction models could not give totally satisfactory results since the spectra were collected on intact samples. If the contribution of muscle heterogeneity to the variability in NIR predictions was known, it would be possible to estimate how much of this variation in predictions is due to other factors and how much NIR needs to be enhanced to improve its accuracy. As it was reported by Rosenvold (2012), when collecting the spectra, even whether efforts are done to adjust the probe in order to capture as much signal from the fat as possible, it is inevitable that information will be captured from other tissues. We experienced the same situation when collecting spectra with SCiO spectrometer where we eventually had some absorbance on the muscle and not on the intramuscular fat. This can therefore increase noise and decrease predictivity.

Furthermore, MSA marbling AUS-MEAT scores (considered as the reference value) are visual parameters with technical variability, and this introduces additional error into the model of prediction despite its positive correlation with intramuscular fat levels obtained in the lab (Frank et al., 2016). The different pre-processing methods used to reduce the effect of heterogeneity cannot increase NIR sensitivity.

The NIR sensitivity could be enhanced by increasing the number of scan replication (Rødbotten, Nilsen, & Hildrum, 2000; Rosenvold, 2012). For this reason, we collected 5 scans per carcass despite NIR measurement was done on the production chain, which is fast and cannot be slowed down. Rosenvold (2012) could reach the targeted standard (70% in the range ± 50) with 40 scans. This will unfortunately increase the time of scanning for each carcass and limit online application since, in practice, the number of scans could not be increased

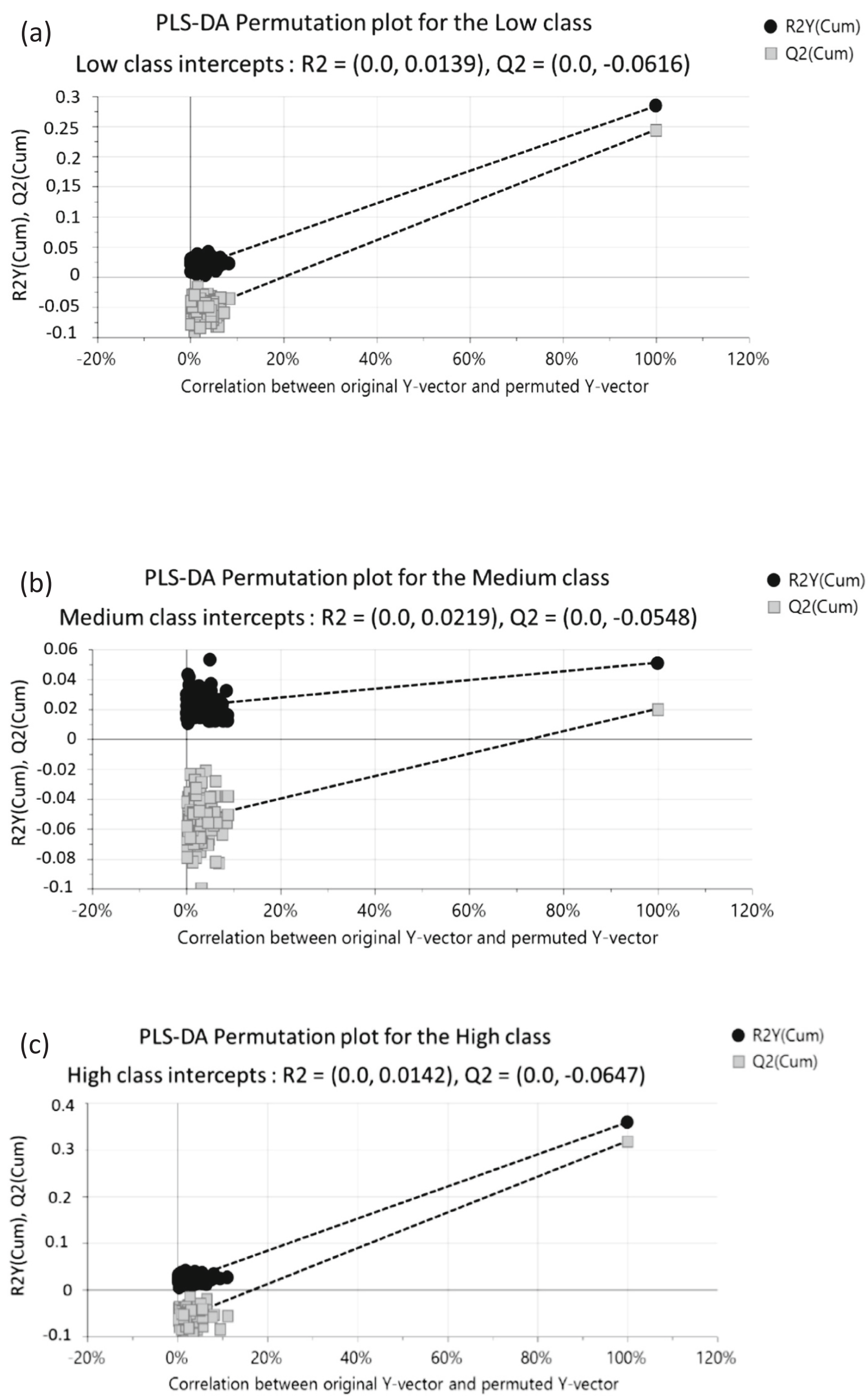


Fig. 8. Permutation plots for the (a) low, (b) medium, and (c) high class.

indefinitely. When collecting the spectra, we used a stand-off to avoid the meat touching the probe glass window and limit interference with the surrounding light. This stand-off could be removed to increase sensitivity and this could increase the signal captured from the sample (Rady, Fischer, Reeves, Logan, & James Watson, 2020). However, if this approach was proven as useful, it might require more frequent cleaning of the probe window (Rosenvold, 2012).

Near-infrared models lacked robustness and did not gave completely satisfactory predictions when used on a new batch. Also, a similar type of sensor (900–1700 nm) could not give satisfactory results as shown in a study by Coombs, Fajardo, and González (2021). We should also consider that industrial conditions are exposed to fluctuations in humidity and temperature (Przybylski & Hopkins, 2015). One solution could be the use of fiber optic probes. Better results could be achieved by acquiring adding spatial information to spectral information as Hypectral imaging does. Hyperspectral imaging integrates spectroscopy and computer vision techniques and this provided better results to predict marbling score for the Japanese standard classification (Velásquez, Cruz-Tirado, Siche, & Quevedo, 2017).Other types of statistical approaches may be suggested in order to increase accuracy, such as neural networks (Prevolnik et al., 2011), decision tree (Velásquez et al., 2017) and a model updating method based on just-in-time learning (Zhang et al., 2022).

4.3. MSA class prediction

Beef industry does not need to accurately predict marbling score but it rather aims to ensure good eating quality of beef to the consumer, thus reaching a minimum level of marbling accordingly. In other words, classifying carcasses in low-, medium- and high- marbling groups could be sufficient from a commercial point of view. Previous studies have demonstrated that the intramuscular fat deposition, and therefore marbling, are influenced by a wider range of factors, such as breed, sex, age, and diet (Nguyen, Nguyen, & Malau-Aduli, 2021; Park et al., 2018). Differences were found in the accuracies of prediction of fatty acids for Aberdeen Angus and cross bred Limousine (Prieto et al., 2011). One reason could be the differences in the adipocyte size between breeds that can influence the NIR spectra collected. These results therefore suggest the use of specific prediction equations for individual breeds and sexes (Prieto et al., 2011).

From the confusion matrix, the predictability for the low marbling levels is generally higher. Therefore, marbling scores lower than 270 are better predicted for both sexes and breeds. This is quite close to our average MSA values considering the range of ±50. The overall accuracy of the confusion matrix was 61%, meaning that 61% of all the carcasses were correctly classified. However, with this approach, we still could not reach the targeted range for MSA accreditation of carcass graders (70% of well classified samples within ±50).

The interpretation for the AUC-ROC curves was done based on a previous study done by (Yang & Berdine, 2017). It was suggested that a value of AUC = 0.5 allows no discrimination, AUC value between 0.5 and 0.6 gives a poor discrimination, AUC value between 0.6 and 0.7 gives an acceptable discrimination, AUC value between 0.7 and 0.8 gives an excellent discrimination and AUC value >0.9 gives an outstanding discrimination. From this, we can conclude that our model could predict the low and high class given that the AUC values were above 0.8 for both models.

As the two PLS-DA models with or without the inclusion of data related to animal breed and sex provided similar results, our study suggests that the prediction of MSA classes could be done independently of sex and breed information with a large dataset. Results could be improved by having more animals in each class.

5. Conclusion

The measurement of carcass quality traits in the chiller can be laborious since commercial slaughterhouses need to acquire chiller assessors, finance the acquisition and retention of the accreditation for assessors. This process can be expensive for small or large scale slaughterhouses. Therefore, in order to maximize economic benefits and increase grading accuracy as well as reliability, NIRS can offer a viable solution. This study was performed on a large sample size to evaluate if the spectra obtained from a handheld NIR spectrometer could help to predict MSA and AUS-MEAT marbling scores. Different procedures were applied using different software in order to develop the best models. The chemometrics procedures could not provide completely satisfactory results for prediction of individual scores for both measured traits, preventing the predicted values from being used as a substitute for the official evaluation of marbling by accredited operators. The PLS-DA model could predict extreme classes (low and high). In the context of this study, carcasses with a marbling score higher than 370 were considered high. This value can relatively be low compared to the whole MSA marbling scale which has a maximum value of 1190. This is because Australian breeds (especially English native breeds) have more intramuscular fat than European breeds. Our results are therefore more adapted to the European context. Better results can be achieved by performing the analysis on a dataset with large variability in each class. Moreover, there could be a genetic association between infrared predicted and reference traits which might be sufficient for selection purposes and it is worth noting that this device allows the reading of a sample rapidly, obtaining a very high number of phenotypes in many slaughterhouses when compared to the activity that the few official graders present in Europe can perform. The investigations in this study therefore led to the conclusion that, hand held spectrometers can be suitable to predict classes of marbling rather than individual marbling scores. For meat companies that wish to use such small, simple, and cheap NIR devices with narrow spectral range, further calibrations should be performed on a wide variety of animals and marbling levels, taking into account technical variability for both carcass graders and slaughtering conditions including constraints in commercial slaughterhouses.

Ethics statement

Ethical review and approval were not required for the animal study because Data from animals was used from commercial slaughterhouses at Limoges, France and Padova, Italy according to normal rules.

CRediT authorship contribution statement

Moïse Kombolo-Ngah: Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Arianna Goi:** Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Matteo Santinello:** Formal analysis, Writing – original draft. **Nicola Rampado:** Methodology, Validation, Data curation. **Stefka Atanassova:** Methodology, Formal analysis. **Jingjing Liu:** Methodology, Validation, Data curation. **Pascal Faure:** Methodology, Validation, Data curation. **Laure Thoumy:** Methodology, Validation, Data curation. **Alix Neveu:** Methodology, Validation. **Donato Andueza:** Methodology, Formal analysis. **Massimo De Marchi:** Methodology, Validation, Funding acquisition, Project administration, Supervision. **Jean-François Hocquette:** Methodology, Validation, Funding acquisition, Project administration, Supervision.

Declaration of Competing Interest

None.

Data availability

Data will be made available on request.

Acknowledgements

AZoVe (Cittadella, Italy), and Plainemaison Beauvallet (Limoges, France) are gratefully acknowledged for the provision of carcasses for sampling. This work was supported by the SustaIn4Food project (funded by Veneto region, Italy) and the INTAQT project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101000250

References

Allen, T. J., Hall, A., Dhillon, A. P., Owen, J. S., & Beard, P. C. (2012). Spectroscopic photoacoustic imaging of lipid-rich plaques in the human aorta in the 740 to 1400 nm wavelength range. *Journal of Biomedical Optics, 17*(6), Article 061209. https://doi.org/10.1117/1.JBO.17.6.061209

Almeida, M. R., Fidelis, C. H. V., Barata, L. E. S., & Poppi, R. J. (2013). Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. *Talanta, 117*, 305–311. https://doi.org/10.1016/j.talanta.2013.09.025

Alomar, D., Gallo, C., Castañeda, M., & Fuchslocher, R. (2003). Chemical and discriminant analysis of bovine meat by near infrared reflectance spectroscopy (NIRS). *Meat Science, 63*(4), 441–450.

Banerjee, P., Ghosh, S., Dutta, M., Subramani, E., Khalpada, J., Roy Choudhury, S., … Chaudhury, K. (2013). Identification of key contributory factors responsible for vascular dysfunction in idiopathic recurrent spontaneous miscarriage. *PLoS One, 8*(11), Article e80940. https://doi.org/10.1371/journal.pone.0080940

Barlocco, N., Vadell, A., Ballesteros, F., Galietta, G., & Cozzolino, D. (2006). Predicting intramuscular fat, moisture and Warner-Bratzler shear force in pork muscle using near infrared reflectance spectroscopy. *Animal Science, 82*, 111–116. https://doi.org/10.1079/ASC20055

Beć, K., Grabska, J., & Huck, C. (2020). Near-infrared spectroscopy in bio-applications. *Molecules, 25*, 2948. https://doi.org/10.3390/molecules25122948

Benedet, A., Costa, A., De Marchi, M., & Penasa, M. (2020). Heritability estimates of predicted blood β-hydroxybutyrate and nonesterified fatty acids and relationships with milk traits in early-lactation Holstein cows. *Journal of Dairy Science, 103*(7), 6354–6363. https://doi.org/10.3168/jds.2019-17916

Bonny, S. P. F., O'Reilly, R. A., Pethick, D. W., Gardner, G. E., Hocquette, J.-F., & Pannier, L.. (2018). Update of meat standards Australia and the cuts based grading scheme for beef and sheepmeat. *Journal of Integrative Agriculture, 17*, 1641–1654. https://doi.org/10.1016/S2095-3119(18)61924-0

Bouveresse, E., & Massart, D. L. (1996). Standardisation of near-infrared spectrometric instruments: A review. *Vibrational Spectroscopy, 11*(1), 3–15. https://doi.org/10.1016/0924-2031(95)00055-0

Büning-Pfaue, H. (2003). Analysis of water in food by near infrared spectroscopy. *Food Chemistry, 82*, 107–115. https://doi.org/10.1016/S0308-8146(02)00583-6, 2nd International Workshop on Water in Foods.

Calkins, C. R., & Hodgen, J. M. (2007). A fresh look at meat flavor. *Meat Science, 77*, 63–80. https://doi.org/10.1016/j.meatsci.2007.04.016, 53rd International Congress of Meat Science and Technology (53rd ICoMST).

Chriki, S., Renand, G., Picard, B., Micol, D., Journaux, L., & Hocquette, J. F. (2013). Meta-analysis of the relationships between beef tenderness and muscle characteristics. *Livestock Science, 155*, 424–434. https://doi.org/10.1016/j.livsci.2013.04.009

Coombs, C. E. O., Fajardo, M., & González, L. A. (2021). Comparison of smartphone and lab-grade NIR spectrometers to measure chemical composition of lamb and beef. *Animal Production Science, 61*(16), 1723–1733. https://doi.org/10.1071/AN21069

Costa, A., Franzoi, M., Visentin, G., Goi, A., De Marchi, M., & Penasa, M. (2021). The concentrations of immunoglobulins in bovine colostrum determined by the gold standard method are genetically correlated with their near-infrared prediction. *Genetics, Selection, Evolution, 53*, 87. https://doi.org/10.1186/s12711-021-00681-8

Costa, A., Visentin, G., De Marchi, M., Cassandro, M., & Penasa, M. (2019). Genetic relationships of lactose and freezing point with minerals and coagulation traits predicted from milk mid-infrared spectra in Holstein cows. *Journal of Dairy Science, 102*(8), 7217–7225. https://doi.org/10.3168/jds.2018-15378

Cozzolino, D., & Murray, I. (2002). Effect of sample presentation and animal muscle species on the analysis of meat by near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy, 10*(1), 37–44. https://doi.org/10.1255/jnirs.319

De Marchi, M., Costa, A., Goi, A., Penasa, M., & Manuelian, C. L. (2019). Novel applications of infrared technologies in dairy industry. *Advanced Technologies, 8*(2), 92–98. https://doi.org/10.5937/savteh1902092d

Dixit, Y., Casado-Gavalda, M. P., Cama-Moncunill, R., Cama-Moncunill, X., Markiewicz-Keszycka, M., Cullen, P. J., & Sullivan, C. (2017). Developments and challenges in online NIR spectroscopy for meat processing. *Comprehensive Reviews in Food Science and Food Safety, 16*, 1172–1187. https://doi.org/10.1111/1541-4337.12295

Etats Generaux de l''Alimentation EGA. (2018). https://agriculture.gouv.fr/alimagri-les-etats-generaux-de-lalimentation (accessed 6.15.22).

Fan, Y., Liao, Y., & Cheng, F. (2018). Predicting of intramuscular fat content in pork using near infrared spectroscopy and multivariate analysis. *International Journal of Food Properties, 21*, 1180–1189. https://doi.org/10.1080/10942912.2018.1460606

Fernández-Cabanás, V. M., Polvillo, O., Rodríguez-Acuña, R., Botella, B., & Horcada, A. (2011). Rapid determination of the fatty acid profile in pork dry-cured sausages by NIR spectroscopy. *Food Chemistry, 124*(1), 373–378. https://doi.org/10.1016/j.foodchem.2010.06.031

Fowler, S. M., Wheeler, D., Morris, S., Mortimer, S. I., & Hopkins, D. L. (2021). Partial least squares and machine learning for the prediction of intramuscular fat content of lamb loin. *Meat Science, 177*, Article 108505. https://doi.org/10.1016/j.meatsci.2021.108505

Frank, D., Ball, A., Hughes, J., Krishnamurthy, R., Piyasiri, U., Stark, J., … Warner, R. (2016). Sensory and flavor chemistry characteristics of Australian beef: Influence of intramuscular fat, feed, and breed. *Journal of Agricultural and Food Chemistry, 64*(21), 4299–4311. https://doi.org/10.1021/acs.jafc.6b00160

Givens, D. I., Boever, J. L. D., & Deaville, E. R. (1997). The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans. *Nutrition Research Reviews, 10*, 83–114. https://doi.org/10.1079/NRR19970006

Goi, A., Hocquette, J. F., Pellattiero, E., & De Marchi, M. (2022). Handheld near-infrared spectrometer allows on-line prediction of beef quality traits. *Meat Science, 184*, Article 108694. https://doi.org/10.1016/j.meatsci.2021.108694

Guillemin, N., Cassar-Malek, I., Hocquette, J. F., Jurie, C., Micol, D., Listrat, A., Leveziel, H., Renand, G., & Picard, B. (2009). La maîtrise de la tendreté de la viande bovine : identification de marqueurs biologiques. *INRAE Productions Animales, 22*, 331–344. https://doi.org/10.20870/productions-animales.2009.22.4.3358

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hocquette, J.-F., Ellies-Oury, M.-P., Lherm, M., Pineau, C., Deblitz, C., & Farmer, L. (2018). Current situation and future prospects for beef production in Europe — A review. Asian-Australas. *Journal of Animal Science, 31*, 1017–1035. https://doi.org/10.5713/ajas.18.0196

Jacob, R. (2020). Implications of the variation in bloom properties of red meat: A review. *Meat Science, 162*, Article 108040. https://doi.org/10.1016/j.meatsci.2019.108040

Jansen, K., Wu, M., van der Steen, & van Soest, G. (2014). Photoacoustic imaging of human coronary atherosclerosis in two spectral bands. *Photoacoustics, 2*(1), 12–20.

Kademi, H. I., Ulusoy, B. H., & Hecer, C. (2018). Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products. *Food Reviews International, 35*(3), 201–220. https://doi.org/10.1080/87559129.2018.1514624

Karoui, R., Mouazen, A. M., Dufour, É., Pillonel, L., Picque, D., Bosset, J.-O., & De Baerdemaeker, J. (2006). Mid-infrared spectrometry: A tool for the determination of chemical parameters in Emmental cheeses produced during winter. *Le Lait, 86*, 83–97. https://doi.org/10.1051/lait:2005040

Kong, X., Yang, X., Zhou, J., Chen, S., Li, X., Jian, F., Deng, P., & Li, W. (2015). Analysis of plasma metabolic biomarkers in the development of 4-nitroquinoline-1-oxide-induced oral carcinogenesis in rats. *Oncology Letters, 9*(1), 283. https://doi.org/10.3892/ol.2014.2619

Legrand, I., Hocquette, J.-F., Polkinghorne, R. J., & Pethick, D. W. (2013). Prediction of beef eating quality in France using the meat standards Australia system. *Animal, 7*, 524–529. https://doi.org/10.1017/S1751731112001553

Liu, J., Pogorzelski, G., Neveu, A., Legrand, I., Pethick, D., Ellies-Oury, M. P., & Hocquette, J. F. (2021). Are Marbling and the Prediction of Beef Eating Quality Affected by Different Grading Sites? *Frontiers in Veterinary Science, 8*, 611153.

Meat & Livestock Australia, Meat Standards Australia. (2018). *Meat standards Australia information kit*. North Sydney: Meat & Livestock Australia.

Meat, Livestock, Australia, & Meat Standards, Australia. (2001). *Meat Standards Australia information kit*. Meat & Livestock Australia.

Nguyen, D. V., Nguyen, O. C., & Malau-Aduli, A. E. O. (2021). Main regulatory factors of marbling level in beef cattle. *Veterinary and Animal Science, 14*, 100219.

Park, S. J., Beak, S.-H., Jung, D. J. S., Kim, S. Y., Jeong, I. H., Piao, M. Y., … Baik, M. (2018). Genetic, management, and nutritional factors affecting intramuscular fat deposition in beef cattle — A review. *Asian-Australasian Journal of Animal Sciences, 31*(7), 1043–1061.

Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society, 14*(2), 198–219. https://doi.org/10.1590/S0103-50532003000200006

Prevolnik, M., Čandek-Potokar, M., & Škorjanc, D. (2004). Ability of NIR spectroscopy to predict meat chemical composition and quality - a review. *Czech Journal of Animal Science, 49*, 500–515. https://doi.org/10.17221/4337-CJAS

Prevolnik, M., Čandek-Potokar, M., & Škorjanc, D. (2011). Ability of NIR spectroscopy to predict meat chemical composition and quality – a review. *Czech Journal of Animal Science, 49*, 500–510. https://doi.org/10.17221/4337-CJAS

Prieto, N., López-Campos, Ó., Aalhus, J. L., Dugan, M. E. R., Juárez, M., & Uttaro, B. (2014). Use of near infrared spectroscopy for estimating meat chemical composition, quality traits and fatty acid content from cattle fed sunflower or flaxseed. *Meat Science, 98*(2), 279–288. https://doi.org/10.1016/j.meatsci.2014.06.005

Prieto, N., Pawluczyk, O., Russell Dugan, M. E., & Aalhus, J. L. (2017). A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Applied Spectroscopy, 71*(7), 1403–1426. https://doi.org/10.1177/0003702817709299

Prieto, N., Ross, D. W., Navajas, E. A., Richardson, R. I., Hyslop, J. J., Simm, G., & Roehe, R. (2011). Online prediction of fatty acid profiles in crossbred Limousin and Aberdeen Angus beef cattle using near infrared reflectance spectroscopy. *Animal, 5*(1), 155–165.

Przybylski, P. W., & Hopkins, P. D. (Eds.). (2015). *Meat quality: Genetic and environmental factors* (1st ed.). CRC Press. https://doi.org/10.1201/b19250.

Rady, A., Fischer, J., Reeves, S., Logan, B., & James Watson, N. (2020). The effect of light intensity, sensor height, and spectral pre-processing methods when using NIR spectroscopy to identify different allergen-containing powdered foods. *Sensors, 20*(1), 1. https://doi.org/10.3390/s20010230

Ripoll, G., Albertí, P., Panea, B., Olleta, J. L., & Sañudo, C. (2008). Near-infrared reflectance spectroscopy for predicting chemical, instrumental and sensory quality of beef. *Meat Science, 80*, 697–702. https://doi.org/10.1016/j.meatsci.2008.03.009

Rødbotten, R., Nilsen, B. N., & Hildrum, K. I. (2000). Prediction of beef quality attributes from early post mortem near infrared reflectance spectra. *Food Chemistry, 69*, 427–436. https://doi.org/10.1016/S0308-8146(00)00059-5

Rosenvold, K. (2012). Accreditation of VISNIR spectroscopy to support MSA grading of beef. In *Published By Meat and Livestock Australia limited locked bag 991 North Sydney NSW 2059 23*.

Schrieve, G. D., Melish, G. G., & Ullman, A. H. (1991). The Herschel-Infrared—A useful part of the spectrum. *Applied Spectroscopy, 45*, 711–713. https://doi.org/10.1366/0003702914336903

Shenk, J. S., & Westerhaus, M. O. (1991). New standardization and calibration procedures for Nirs analytical systems. *Crop Science, 31*(6). https://doi.org/10.2135/cropsci1991.0011183X003100060064x. cropsci1991.0011183X003100060064x.

SIMCA® - Multivariate Data Analysis Software. (2020). Sartorius. https://www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/mvda-software/simca.

Stewart, S. M., Lauridsen, T., Toft, H., Pethick, D. W., Gardner, G. E., McGilchrist, P., & Christensen, M. (2021). Objective grading of eye muscle area, intramuscular fat and marbling in Australian beef and lamb. *Meat Science, 181*, Article 108358. https://doi.org/10.1016/j.meatsci.2020.108358

Su, H., Sha, K., Zhang, L., Zhang, Q., Xu, Y., Zhang, R., … Sun, B. (2014). Development of near infrared reflectance spectroscopy to predict chemical composition with a wide range of variability in beef. *Meat Science, 98*(2), 110–114. https://doi.org/10.1016/j.meatsci.2013.12.019

Tøgersen, G., Arnesen, J. F., Nilsen, B. N., & Hildrum, K. I. (2003). On-line prediction of chemical composition of semi-frozen ground beef by non-invasive NIR spectroscopy. *Meat Science, 63*, 515–523. https://doi.org/10.1016/S0309-1740(02)00113-4

Van Kempen, T. (2001). Infrared technology in animal production. *World's Poultry Science Journal, 57*, 29–48. https://doi.org/10.1079/WPS20010004

Velásquez, L., Cruz-Tirado, J. P., Siche, R., & Quevedo, R. (2017). An application based on the decision tree to classify the marbling of beef by hyperspectral imaging. *Meat Science, 133*, 43–50. https://doi.org/10.1016/j.meatsci.2017.06.002

Watson, R., Polkinghorne, R., & Thompson, J. M. (2008). Development of the meat standards Australia (MSA) prediction model for beef palatability. *Australian Journal of Experimental Agriculture, 48*, 1368. https://doi.org/10.1071/EA07184

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, … van Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics, 4*(1), 81–89.

Williams, P. (2014). The RPD statistic: A tutorial note. *NIR News, 25*, 22–26. https://doi.org/10.1255/nirn.1419

Wilson, R., Nadeau, K., Jaworski, F., Tromberg, B., & Durkin, A. (2015). Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization. *Journal of Biomedical Optics, 20*, 30901. https://doi.org/10.1117/1.JBO.20.3.030901

Wold, J. P., Solberg, L. E., Gaarder, M.Ø., Carlehøg, M., Sanden, K. W., & Rødbotten, R. (2022). In-line estimation of fat marbling in whole beef striploins (longissimus lumborum) by NIR hyperspectral imaging. A closer look at the role of myoglobin. *Foods, 11*(9), 9. https://doi.org/10.3390/foods11091219

Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems, 44*(1), 175–185. https://doi.org/10.1016/S0169-7439(98)00109-9

Yang, S., & Berdine, G. (2017). The receiver operating characteristic (ROC) curve. *The Southwest Respiratory and Critical Care Chronicles, 5*(19), 19. https://doi.org/10.12746/swrccc.v5i19.391

Zhang, F., Kang, T., Sun, J., Wang, J., Zhao, W., Gao, S., … Ma, Q. (2022). Improving TVB-N prediction in pork using portable spectroscopy with just-in-time learning model updating method. *Meat Science, 188*, Article 108801. https://doi.org/10.1016/j.meatsci.2022.108801