Journal of Food Composition and Analysis 115 (2023) 105017

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

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

Near infrared spectroscopy coupled to chemometrics for the authentication of donkey milk

Francesca Di Donato [a,*], Alessandra Biancolillo [a], Alessandra Ferretti [b], Angelo Antonio D'Archivio [a], Federico Marini [b]

[a] Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, 67100, Coppito, L'Aquila, Italy
[b] Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro, 5, 00185 Roma, Italy

ARTICLE INFO

Keywords:
Donkey milk
Near Infrared (NIR) spectroscopy
Authentication
Partial Least Square–Discriminant Analysis (PLS-DA)
Variable Sorting for Normalization (VSN)
ANOVA-Simultaneous Component Analysis (ASCA)

ABSTRACT

Donkey milk (DM) is an emerging foodstuff, gaining increasing attention due to its nutritional properties. As a high value-added product, it can be subjected to adulteration with cheaper milks, e.g., cow milk. The present work addresses the possibility of developing a fast and economic method for the authentication of pure DM using Near Infrared (NIR) Spectroscopy. For this purpose, 147 samples (67 pure DM and 80 mixtures) were analyzed fresh; additionally, all samples underwent successive freezing-thawing cycles. Partial Least Squares-Discriminant Analysis (PLS-DA) was used to differentiate pure and adulterated samples. PLS-DA model built on the fresh milk correctly classified all training and test samples (100% accuracy). When model was applied to the freeze-thawed individuals, it showed high accuracy (79.7% after 3 cycles), suggesting that the spectroscopic signature of adulteration prevails on that of freezing-thawing, though the impact of the latter was proved to be significant by ANOVA-Simultaneous Component Analysis (ASCA).

1. Introduction

Donkey milk is attracting growing interest in human nutrition as a valuable alternative in feeding infants because of its similarity with the human one and its hypoallergenic properties (Souroullas et al., 2018). Several authors (Guo et al., 2007; Salimei et al., 2004; Vincenzetti et al., 2008) have proved that pH, protein content, ashes and total solids in donkey milk are more similar to mare and human milk than those of all the other mammals. In addition, it stimulates the immune system, regulates the gastrointestinal flora, and prevents inflammatory diseases (Derdak et al., 2020). Coppola et al. suggested also the use of donkey milk for probiotic purposes (Coppola et al., 2002), providing a good growth medium for probiotic lactobacilli strains thanks to its high content of lysozyme and lactose. Despite its unique nutrient profile and economic potential, in Europe donkey milk can be commercialized only since 2004, when EC Regulation n. 853/2004, included it among the "other milk species" category (European Commission, 2004). Nowadays, donkey breeding is circumscribed to Asia, Africa, Eastern and some countries of Western Europe, and it is so scattered that the milk yield is very low. Thanks to its properties and due to its niche-product status, donkey milk is a high value-added product, with a relatively high market value (around 15 € per liter); as such, it can be subjected to adulteration with milk from other mammalian species, especially with cheaper cow milk whose most recent average European price at the barn has been reported to be 35.69€/100 kg (European Commission, 2021). This is turning into a common illicit practice, involving also other valuable kinds of milk, such as buffalo, goat and sheep milk; as a consequence, several approaches have been proposed in literature in order to detect and prevent these kinds of fraud (Agrimonti et al., 2015; Pesic et al., 2011). Nevertheless, there is still a lack of methods aimed at the authentication of donkey milk. These fraudulent practices result in an authenticity problem, but also in a real health issue. Indeed, recent clinical studies confirmed feeding with donkey milk as the safest and the most valid alternative for the nutrition of infants affected by cow milk protein intolerance (CMPI), an Ig-E mediated severe pathology affecting 3% of infants under the age of 12 months, often misdiagnosed (Ewing and Allen, 2005).

The adulteration-detection study proposed here lays on these considerations and aims at developing a non-destructive approach for authenticating donkey milk and detecting its adulteration with cow

* Corresponding author.
E-mail addresses: francesca.didonato3@graduate.univaq.it, didonatofrancesc@gmail.com (F. Di Donato), alessandra.biancolillo@univaq.it (A. Biancolillo), aleferretti1996@gmail.com (A. Ferretti), angeloantonio.darchivio@univaq.it (A.A. D'Archivio), federico.marini@uniroma1.it (F. Marini).

https://doi.org/10.1016/j.jfca.2022.105017
Received 15 May 2022; Received in revised form 15 October 2022; Accepted 30 October 2022
Available online 1 November 2022

milk. Near Infrared (NIR) Spectroscopy was employed to achieve this goal, because it had proved to perform effectively in such authentication issues (Dos Santos Pereira et al., 2021; Teixeira et al., 2021). NIR was used to analyze 67 samples of genuine donkey milk and 80 mixtures of donkey and cow milk (in different proportions), prepared to mimic the adulteration. Moreover, both pure and adulterated samples were subjected to successive freezing-thawing (F/T) cycles prior to being re-analyzed by NIR. This physical sample preprocessing was included in order to investigate whether this pretreatment alters milk specimens. Freezing products and selling them as fresh food is a further illegal action commonly applied, which has to be prevented (Hassoun et al., 2020).

Partial Least Squares - Discriminant Analysis (PLS-DA) was used to distinguish pure and adulterated donkey milk samples, before and after freezing-thawing procedure. The choice of the classifier fell on PLS-DA because it had proved to be a suitable tool for the authentication of dairies (Aït-Kaddour et al., 2021; Di Donato et al., 2021; Genis et al., 2021; Mabood et al., 2017) or in similar situations in combination with NIR spectroscopy (Firmani et al., 2020; Haughey et al., 2015; Ríos-Reina et al., 2018).

Eventually, in order to further verify whether freezing-thawing treatment could have a significant effect on the spectroscopic signature, data were further processed by Analysis of Variance (ANOVA)-Simultaneous Component Analysis (ASCA) (Smilde et al., 2005).

2. Material and methods

2.1. Samples

Donkey milk samples were supplied directly from breeders settled in Abruzzo (Central Italy). Eighty samples were artificially adulterated mixing pure donkey milk and cow milk (purchased in local supermarkets). Milk samples were stored in a refrigerator at 4 °C before the analysis, which was performed not later than 5 days from the collection. The sophistications were performed by gradually increasing the concentration of adulterant (from 2% up to 17%). Four samples were prepared at each level of adulteration. Data were also used to attempt the quantification of cow milk content in donkey milk. One hundred and forty-seven samples were available in total, sixty-seven pure and eighty mixtures. Pure and adulterated samples were analyzed before (t0) and after undergoing one (t1), two (t2) and three (t3) freezing-thawing (F/T) cycles. In particular, forty pure and 80 adulterated samples, were analyzed at t1, and 40 pure and 20 adulterated mixtures at both t2 and t3.

2.2. NIR measurements

A Nicolet 6700 (Thermo Scientific Inc., Madison, WI) FT-NIR instrument equipped with an integrating sphere allowed the direct analysis of all the available samples without any further pre-treatment. A suitable volume of each sample was introduced into a glass vial placed on the window of the sphere and two replicate spectra were acquired. The measurements were performed in reflectance mode in the spectral range between 4000 cm^{-1} and 10,000 cm^{-1}, at a nominal resolution of 4 cm^{-1}. The signals were exported from the OMNIC software (Thermo Scientific Inc., Madison, WI) to the MATLAB environment (R2020b; The Mathworks, Natick, MA) and converted to pseudo-absorbance (log(1/R)) to be further processed.

2.3. Partial Least Squares (PLS) and Partial Least Squares-Discriminant Analysis (PLS-DA)

Dealing with spectroscopic data, Partial Least Squares (PLS) (Wold et al., 2001) is one of the most widely used regression methods.

In the present study, a PLS model was built on NIR data to quantify the amount of cow milk in the adulterated donkey milk samples, using Variable Sorting for Normalization (VSN) as data pre-processing technique. The reader is addressed to (Rabatel et al., 2020) for more details on the VSN algorithm.

The PLS algorithm can also be used to build classification models, by encoding class-belonging into a binary-valued response variable **y**. The resulting discriminant classifier is called Partial Least Squares-Discriminant Analysis (PLS-DA) (Sjöström et al., 1986). Since the present study involved a two-class problem (pure/adulterated), the dummy ***y*** was a vector, whose generic element y_i was equal to 1 when corresponding to a sample belonging to Class "*pure*" or to 0, for Class "*adulterated*".

PLS-DA relies on building a PLS regression model between the predictors and the dummy **y**. Since the predicted response $\hat{y}$ is real-valued, a further step is needed to achieve classification. In the present work, the threshold was calculated by applying Linear Discriminant Analysis (LDA) on the predicted responses (Indahl et al., 2007).

2.4. ANOVA-Simultaneous Component Analysis (ASCA)

ANOVA-Simultaneous Component Analysis (Jansen et al., 2005; Smilde et al., 2005) was applied in order to investigate the effect of the time factor (as F/T cycles) on the NIR-fingerprint of the pure donkey milk and its significance. ASCA operates by partitioning the total variance of the experimental data into the individual contributions induced by the effect of the factors under control and their interactions. The corresponding effect matrices derived are then analyzed by Simultaneous Component Analysis (SCA) (Timmerman and Kiers, 2003).

In the present study, the matrix $\boldsymbol{X}_c$ (of dimensions NxM), containing the NIR experimental data after mean-centering, was partitioned into the matrices accounting for the effect of time ($\boldsymbol{X}_{time}$) and for the residuals ($\boldsymbol{X}_{res}$), associated to the random experimental error, according to Eq. (1):

$$\boldsymbol{X}_c = \boldsymbol{X} - 1\boldsymbol{m}^T = \boldsymbol{X}_{time} + \boldsymbol{X}_{res} \quad (1)$$

where $\boldsymbol{m}$ is the grand mean, a row vector collecting the overall average spectrum.

By calculation, $\boldsymbol{X}_{time}$ contains the average profiles corresponding to the four levels of factor time, i.e., t0, t1, t2 and t3. More in detail, $\boldsymbol{X}_{time}$ is a NxM matrix in which all the rows corresponding to a particular level of the controlled factor are filled with identical copies of the mean spectrum of the samples collected at that level. Accordingly, to assess the extent of the contribution of the factor, the sum of squares of the elements of the effect matrix, i.e., its Frobenius' norm, is calculated as:

$$SSQ_{time} = \|\boldsymbol{X}_{time}\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(x_{ij}^{time}\right)^2 \quad (2)$$

To verify whether the effect of the factor is significant or not, the value of SSQ_{time} obtained according to Eq. (2) is compared with its null distribution, which is estimated non-parametrically by means of a permutation test (Vis et al., 2007). In a permutation test, the sample is randomly reassigned to one of the factor levels (t0, t1, t2 and t3); then, the corresponding effect matrix is recalculated and, accordingly, the sum of squares of the matrix elements is computed as:

$$SSQ_{time}^{perm} = \|\boldsymbol{X}_{time}^{perm}\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(x_{ij}^{time,perm}\right)^2 \quad (3)$$

This procedure is repeated for a sufficient number of times (here, 10,000) providing an empirical estimate of distribution of the SSQ values under the null hypothesis.

Eventually, Simultaneous Component Analysis is applied to the effect matrix to model the variability associated to the effect of the freezing/thawing cycles on the spectroscopic signal. $\boldsymbol{X}_{time}$ is therefore decomposed as:

$$X_{time} = T_{time}P_{time}^{T} + E_{time} \quad (4)$$

where T_{time}, P_{time} and E_{time} are the scores, the loadings and the residual matrices, respectively.

3. Results and discussion

Prior to further investigations, data were imported in MATLAB (R2020b; The Mathworks, Natick, MA), and the two replicates acquired on each sample were averaged, leading to the NIR profiles displayed in Fig. 1.

As above-mentioned, the aim of the present work was to develop a NIR-based method to identify possible adulterations made by the addition of cow milk to donkey milk, and, successively, to quantify the amount of adulterant. In order to build and validate the classification and regression models, t0-data were split into two (training and test) representative sets of samples by the application of the Duplex (Snee, 1977) algorithm on each class separately. Since the model optimization stage also involved the selection of the optimal data pre-processing, to avoid any bias related to data splitting the division of the samples was carried out as follows. Six matrices were obtained by pre-processing the experimental NIR data with any of the combinations of pre-treatments to be evaluated (mean centering (MC), SNV+MC, 1st derivative (19 points, 2nd order polynomial)+MC, 2nd derivative (33 points, 3rd order polynomial)+MC, SNV+ 1st derivative+MC, SNV+ 2nd derivative+MC). A multi-block exploratory model was built on these matrices to extract common components which could account for the relevant variability shared by the differently pre-treated data, using ComDim (Qannari et al., 2001). Three common components (CCs) were calculated and the projection of the pure and adulterated samples onto the corresponding sub-space resulted in the two score-matrices T_{pure} and $T_{adulterated}$, respectively. Eventually, Duplex algorithm was run on each of these score matrices individually to split each category in a representative way. Of the 147 t0-milk samples, 112 (25 pure and 60 adulterated) were selected as training set and 35 (15 pure and 20 adulterated) were employed as test set for the validation of the models. Furthermore, with the idea of verifying whether the model built on the fresh samples could be also applied for the identification of adulteration in the freezed-thawed ones, all the spectra collected at t1, t2 and t3 were treated as further validation samples.

All the MATLAB functions employed in the present work can be freely downloaded from the RomeChemometrics web-site (*https://www.chem.uniroma1.it/romechemometrics/research/algorithms/; last accessed March 2022)*.

3.1. Authentication of donkey milk: PLS-DA analysis of milk samples

The most suitable data-preprocessing approach and the optimal number of latent variables (LVs) were chosen as those leading to the highest average correct classification rate (CCR) in a 7-fold cross-validation procedure. The results are shown in Table 1.

Standard Normal Variate (SNV, (Barnes et al., 1989)), first and second derivatives (Savitzky and Golay, 1964) were tested alone or in combination. Mean-centering was additionally performed prior to the

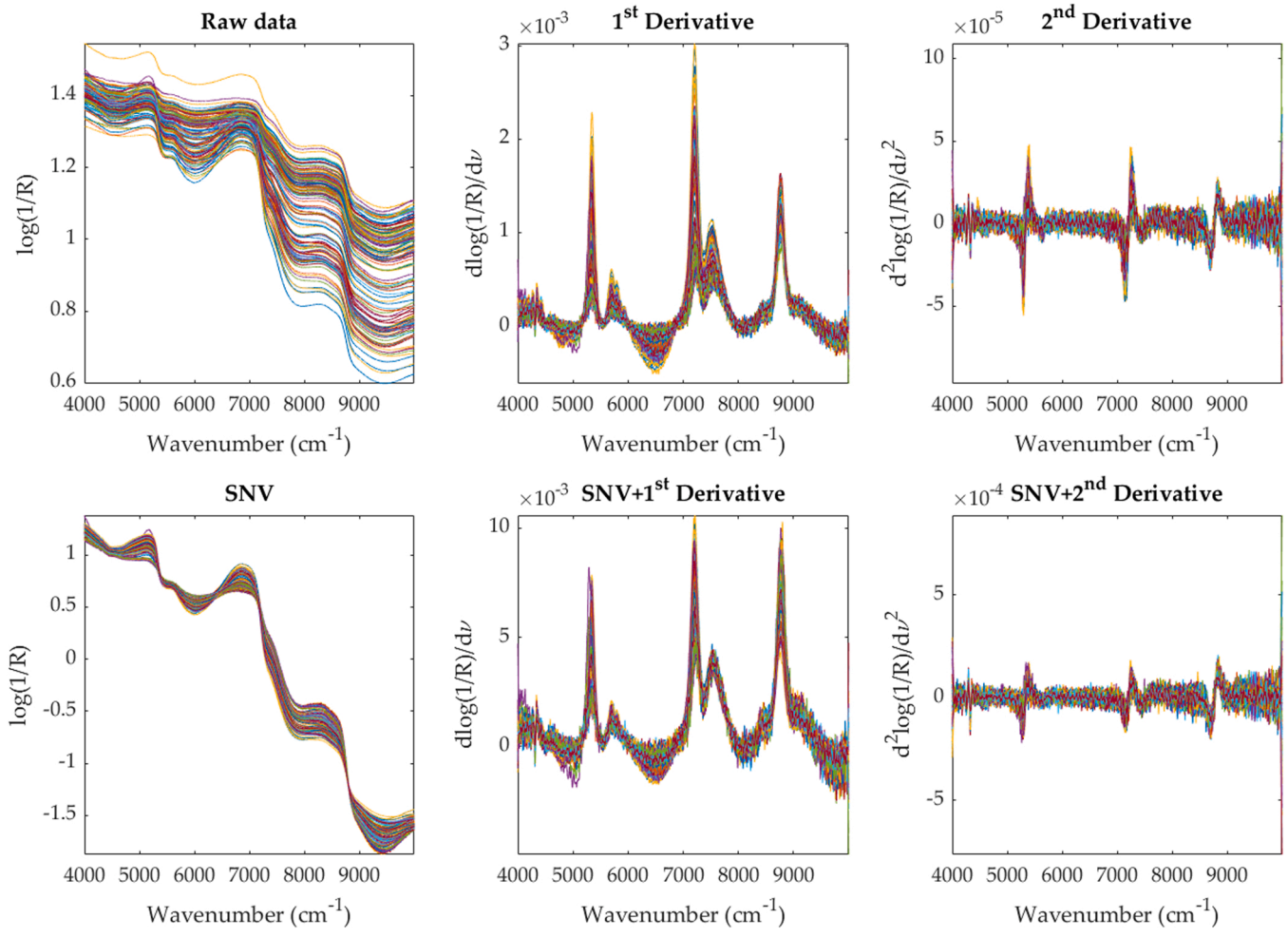

Fig. 1. Raw and differently pre-processed (1st derivative, 2nd derivative, SNV, SNV+1st derivative, SNV+2nd derivative) Near infrared spectra collected on the milk samples.

Table 1
Partial Least Squares-Discriminant Analysis (PLS-DA) model selection: Cross-validated results obtained on the NIR data at t0 (training set). All the figures of merit are expressed as percentages.

Pre-treatment	LVs	Accuracy	Mean CCR	Adulterated		Pure	
				Sensitivity	Specificity	Sensitivity	Specificity
Mean-centering (MC)	9	100.00	100.00	100.00	100.00	100.00	100.00
SNV+MC	**6**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
1st Derivative+MC	5	99.11	99.17	98.33	100.00	100.00	98.33
2nd Derivative+MC	6	92.86	93.21	88.33	98.08	98.08	88.33
SNV+ 1st Derivative+MC	5	99.11	99.17	98.33	100.00	100.00	98.33
SNV+ 2nd Derivative+MC	4	91.96	92.24	88.33	96.15	96.15	88.33

creation of each model. Looking at Table 1, it seems that SNV (followed by MC) is the most appropriate data pre-treatment because it guarantees the highest mean Correct Classification Rate (CCR) (and accuracy) in cross-validation (100%) with the minimum number of LVs (six). The application of the chosen PLS-DA model to the test set provided 100% of CCR for both pure and adulterated classes. In Fig. 2 the results of PLS-DA analysis are graphically represented through the display of the predicted *y* values as a function of the training/test sample index.

Variables Importance in Projection (VIP) indices (Wold et al., 1993) were calculated to investigate of the relevance of each spectral variable in the definition of the latent variables space. Fig. 3.

According to the "greater-than-one" criterion, variables with VIP value higher than 1 (highlighted in Fig. 3) are considered to be relevantly contributing to the model. Moreover, even if their interpretation can suffer from the presence of several overlapping (non-orthogonal) contributions, the regression coefficients were also examined. Indeed, variables with a positive regression coefficient (red in Fig. 3) can be assumed to be higher in the adulterated class, while those with a negative coefficient (blue in Fig. 3) to be higher in the pure samples. By inspection of Fig. 3, the spectral ranges which appear as the most discriminant are mainly associated to moisture and fat content (Ayvaz et al., 2021):

1. the spectral range between 4000 cm^{-1} and 4200 cm^{-1}, likely due to the combination bands of C–H and C–O stretching vibrations in fats;.

2. the variables between 5050 and 5250 cm^{-1} ascribable to moisture, probably as a consequence of the different water content in donkey and cow milk;.

3. a large area centered at 7020 cm^{-1} mostly associated to the first overtone of O-H stretching;.

4. a large area centered at 8400 cm^{-1} where the absorptions are mainly due to the second overtone of the C–H stretching in fats;.

5. around 9200 and 10,000 cm^{-1}, ascribable to the second overtones of O-H and C-H stretching modes.

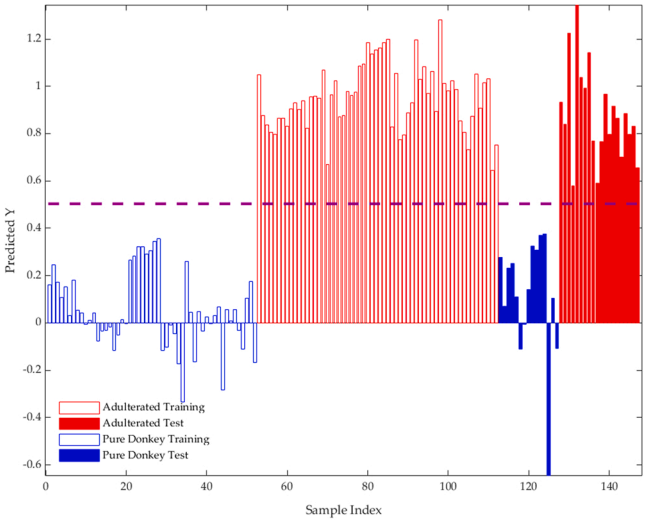

Fig. 2. PLS-DA analysis: values of the predicted response calculated on the training (empty bars) and test (full bars) samples. The violet dashed line is the classification threshold.

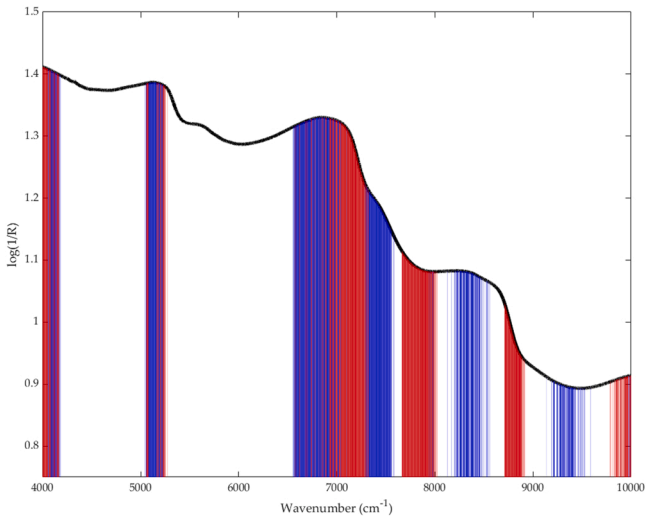

Fig. 3. PLS-DA analysis: graphical representation of the spectral variables identified as significant based on their corresponding VIP index. The average NIR spectrum is represented in black, the variables with VIP > 1 and regression coefficient > 0 are highlighted in red while those with VIP > 1 and regression coefficient < 0 in blue.

Eventually, the optimal PLS-DA model (SNV+MC, 6 LVs) was used to predict the class-membership of the samples after the different F/T cycles. The results are displayed in Table 2, in terms of total classification accuracy, mean correct classification rate and class-sensitivities and specificities, and graphically in Fig. 4.

Among these further predictions, the highest accuracy was obtained for the samples which only underwent a single F/T cycle (t1). When the model was applied to the spectra collected after more F/T cycles (t2 and t3), results were identical and slightly worse than those for t1, the total accuracy was 79.66%, especially due to a significantly lower sensitivity for the pure class. In general, it is possible to observe that the developed model efficiently classified pure and adulterated samples, keeping a very good accuracy also after up to three freezing-thawing cycles, revealing a good robustness with respect to the F/T process. Moreover, even after three F/T cycles, the correct classification rate on the adulterated samples is 100.0%.

3.2. *Quantification of cow milk content in donkey milk samples*

After having assessed the potential of the NIR spectroscopic fingerprint for the detection of the adulteration of donkey milk samples, the possibility of building a model for the quantification of cow milk added as adulterant was investigated. For this purpose, a PLS model was built on the t0-spectroscopic data of the training set (which was the same as for PLS-DA) preprocessed by VSN followed by mean centering. The

Table 2
Predictions of the optimal PLS-DA model on the validation samples (at t0, t1, t2 and t3). All the figures of merit are expressed as percentages.

	LVs	Accuracy	Mean CCR	Adulterated		Pure	
				Sensitivity	Specificity	Sensitivity	Specificity
t0	6	100.00	100.00	100.00	100.00	100.00	100.00
t1	6	85.71	84.78	87.50	82.05	82.05	87.50
t2	6	79.66	84.62	100.00	69.23	69.23	100.00
t3	6	79.66	84.62	100.00	69.23	69.23	100.00

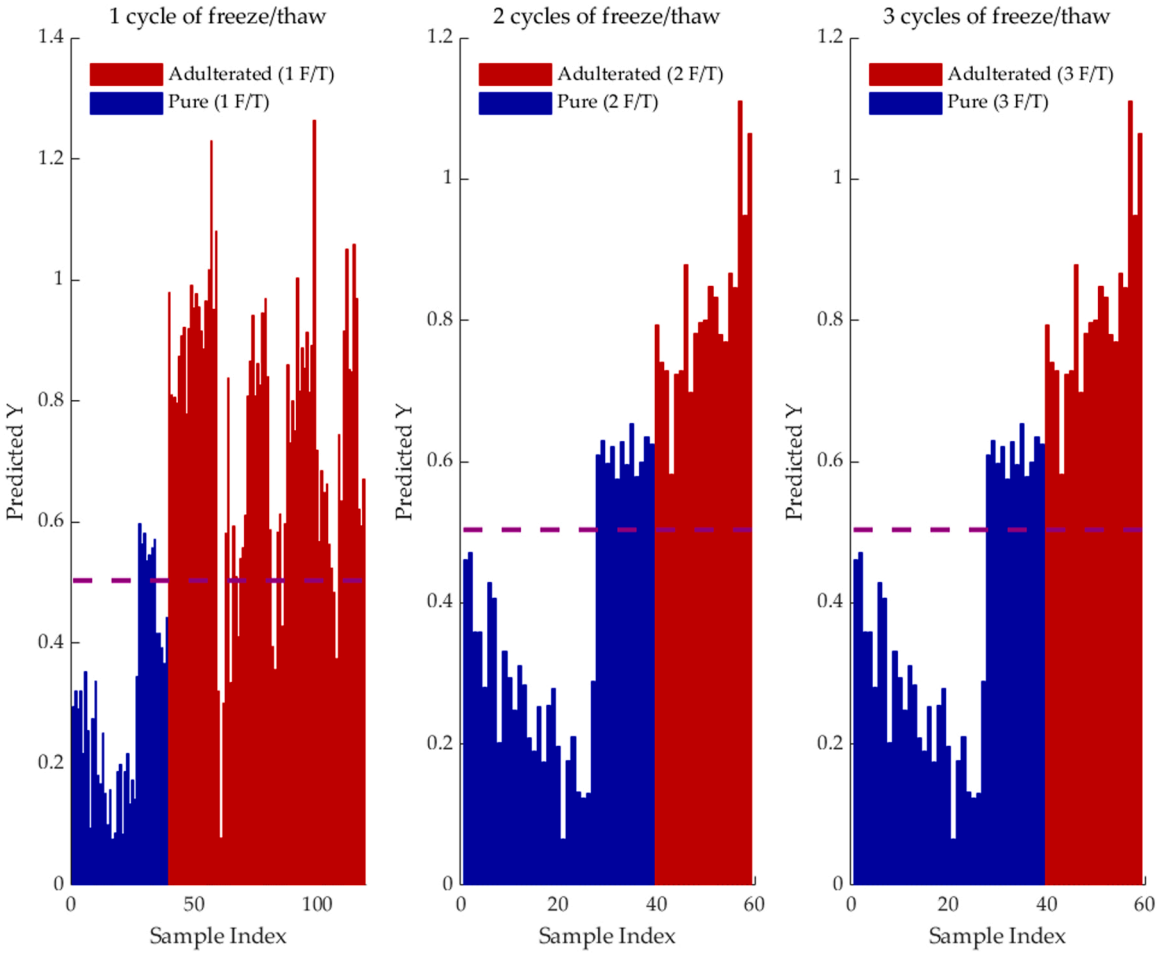

Fig. 4. PLS-DA predictions on the freeze/thawed samples. The violet dashed line represents the classification threshold.

optimal model complexity was identified as the one leading to the lowest value of the root mean square error in a 7-fold cross-validation (RMSECV) and it was found to be 5 LVs (RMSECV=3.9%). When the model was applied to the test set, the results presented a negligible bias (0.2%) and a comparable value of the RMSEP (4.0%). Fig. 5 shows the plot of the predicted vs actual concentration (%) of cow milk in donkey milk, both for calibration (red) and validation samples (black). The inspection of the figure reveals that, for the majority of the samples, the absolute quantification error is lower than 1.5%; nevertheless, for few samples it exceeds 10%.

3.3. ASCA

The ANOVA-like decomposition was applied on the NIR spectra of the pure samples at consecutive F/T cycles (adulterated milk samples are excluded from ASCA calculation) as described in Eq.2 after SNV or VSN+weighted-SNV correction and mean-centering. The design is balanced because the different levels (t0, t1, t2 and t3) of the considered factor (time as freezing/thawing cycles) have the same number of replicates (40 samples).

3.3.1. Estimation of the effect and its significance

In the condition of mutual orthogonality of the effect estimates, the variation in $\|\boldsymbol{X}_c\|^2$ can be split in independent parts ($\|\boldsymbol{X}_c\|^2 = \|\boldsymbol{X}_{time}\|^2 + \|\boldsymbol{X}_{res}\|^2$). The different contributions are given by the term accounting for the freezing/thawing transformation ($\boldsymbol{X}_{time}$) and the residual variation ($\boldsymbol{X}_{res}$), i.e., the random variation that is not accounted for by the mean contribution of the factor levels. In particular, the amount of variance explained by the terms $\boldsymbol{X}_{time}$ and $\boldsymbol{X}_{res}$ were 61% and 39% for the SNV-model and 6% and 94% for the VSN ones, respectively, revealing that the use of the VSN pretreatment can efficiently remove differences in the signals due to the effect of the F/T cycles, while this factor still has a relevant impact on the spectra pre-treated by SNV. To assess if the effect of the factor of interest is statistically significant, a strategy based on permutation test was adopted, as described in Section 2.4. Fig. 6 shows the comparison between the value of the SSQ of the

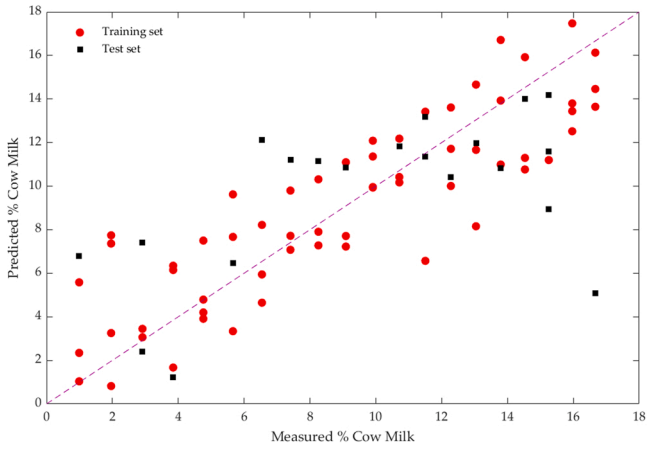

Fig. 5. PLS regression. Predicted percentage of cow milk in the different mixtures with donkey milk. See Section 2.1 for more details of composition.

effect matrix for the experimental data pre-treated either with SNV or VSN and the corresponding null distribution estimated by permutation testing.

Looking at Fig. 6, it is apparent how the effect of freezing/thawing cycles is statistically significant ($p(SNV) < 0.0001$, $p(VSN) = 0.0066$) irrespectively of the pre-treatment adopted, i.e., the F/T procedure affects the NIR spectral fingerprint of pure donkey milk.

3.3.2. SCA models

Simultaneous Component Analysis was applied to extract the systematic information from the effect matrix derived from the ANOVA-decomposition. The factor "time" was studied at four levels (t0, t1, t2 and t3); therefore, three components accounted for 100% of the effect variance. For the sake of a more straightforward visual interpretation of the significance of the overall factor effect and of the differences between factor level means, scores plot were built after projecting the residual matrix onto the effect subspace defined by the loadings $\boldsymbol{P}_{time}$, according to (Zwanenburg et al., 2011):

$$\boldsymbol{T}'_{time} = (\boldsymbol{X}_{time} + \boldsymbol{X}_{res})\boldsymbol{P}_{time} \quad (7)$$

$\boldsymbol{P}_{time}$ being the loading matrix calculated from $\boldsymbol{X}_{time}$ only, as in Eq. (4).

The resulting scores plots are displayed in Fig. 7, both for the SNV- and VSN-pretreated NIR spectra.

In both cases, the largest part of the effect variance is captured by the first component (98.06% in the case of SNV and 80.61%, for VSN). Inspection of the scores plot confirms how the impact of the F/T cycles on the spectroscopic data is higher in the case of the spectra pre-treated with SNV, while for VSN the differences among the factor levels are more subtle. To relate the observed trends with variations in the spectral signal, the loadings on SC1 are reported in Fig. 8.

A bootstrap procedure was employed to estimate their confidence intervals at 95%. Visual inspection of the figure shows that the whole NIR-spectrum is affected by the variation in the "freshness" of the milk, additional F/T cycles will result in an increase of the intensity of signals with positive loadings and in a decrease of spectral signals associated with negative ones.

4. Conclusions

In this work, NIR Spectroscopy coupled with PLS-DA classification was employed as a non-destructive, rapid, and green approach to authenticate donkey milk and to detect its possible adulteration with cow milk. After having verified that the proposed strategy could allow the discrimination between pure and adulterated fresh specimens with 100% accuracy on both the training and the test sets, the study was extended to investigate whether the approach could be suitable to identify the adulteration also when samples had undergone different F/T cycles. In this respect, the results have demonstrated that the model built on fresh sample can still discriminate with reasonable accuracy pure from adulterated samples after 1, 2 or 3 F/T cycles (85.71%, 79.66% and 79.66%, respectively). Moreover, when considering the individual categories, it was demonstrated that the class of adulterated samples can be recognized with 100% sensitivity even after repeated freezing/thawing.

Successively, PLS regression was used to quantify the percentage of cow milk used as adulterant, obtaining, for most of the samples, absolute prediction errors lower than 1.5%.

Finally, the effect of freezing/thawing on the spectroscopic signals of pure samples was further investigated by ASCA. The results indicated that the F/T treatment has a statistically significant effect on the spectroscopic fingerprint.

These results indicate that the proposed method is suitable to be applied for the quality control of donkey milk samples and potentially for the quantification of cow milk as adulterant, irrespectively of the

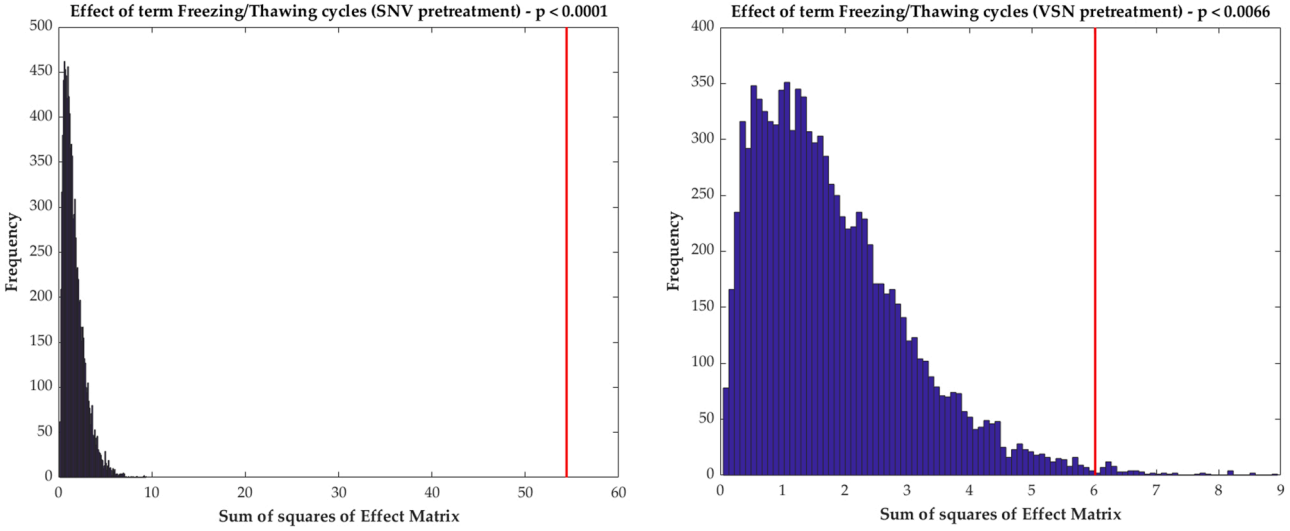

Fig. 6. ASCA. Validation of the main effect of the factor Freezing/Thawing Cycles on pure samples: comparison between the value of the sum of squares obtained for the experimental data (red vertical line) and the corresponding null distribution estimated by permutation test (blue histogram); left: SNV pre-treated data, right: VSN pre-treated data.

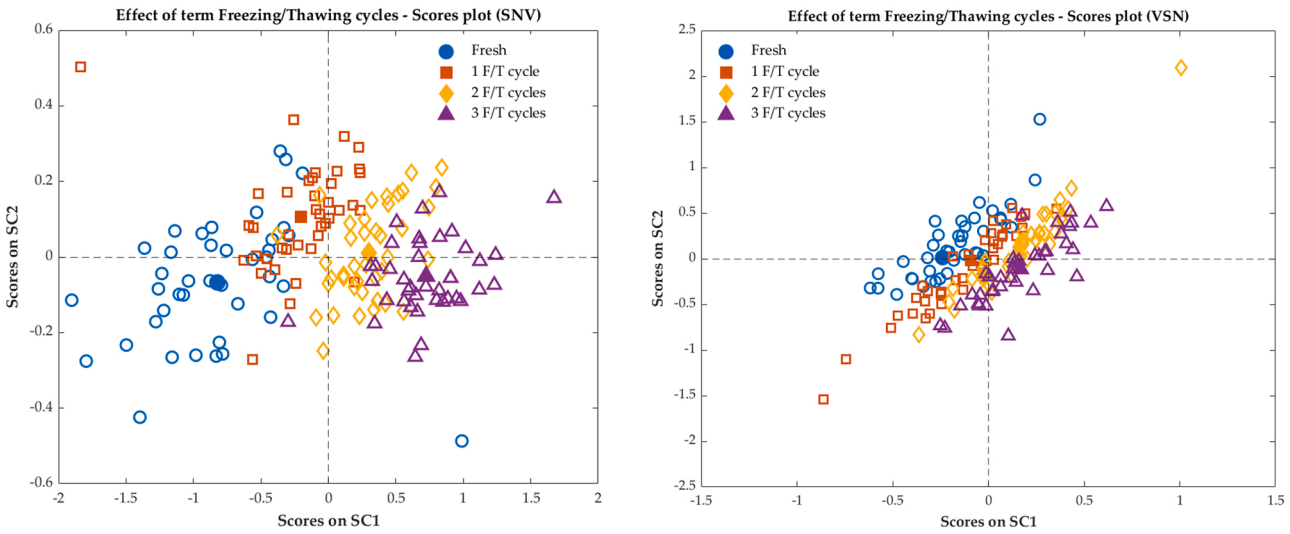

Fig. 7. SCA modeling of the effect of F/T cycles (left: SNV pre-treated data, right: VSN pre-treated data). Scores plots. Full symbols represent the level averages, empty symbols represent the scores after back-projection of the residuals.

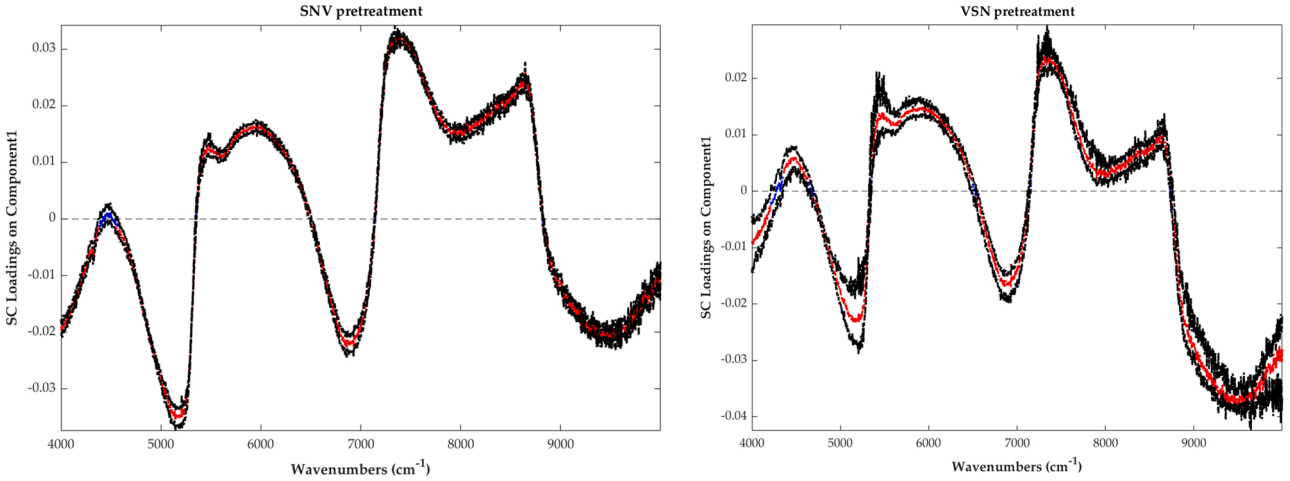

Fig. 8. SCA modeling of the effect of F/T cycles (left: SNV pre-treated data, right: VSN pre-treated data). Loadings of the spectral variables on SC1 together with their bootstrapped confidence intervals (dashed black lines). Loadings are highlighted in red when significantly different from the zero, otherwise they are highlighted in blue.

conditions used for its storage.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Francesca Di Donato: Conceptualization, Validation, Data curation, Writing – original draft, Writing – review & editing. **Alessandra Biancolillo**: Conceptualization, Methodology, Software, Validation, Data curation, Writing – review & editing. **Alessandra Ferretti**: Investigation, Formal analysis. **Federico Marini**: Conceptualization, Methodology, Software, Validation, Data curation, Writing – review & editing, Supervision. **Angelo Antonio D'Archivio**: Conceptualization, Validation, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Agrimonti, C., Pirondini, A., Marmiroli, M., Marmiroli, N., 2015. A quadruplex PCR (qxPCR) assay for adulteration in dairy products. Food Chem. 187, 58–64. https://doi.org/10.1016/J.FOODCHEM.2015.04.017.

Aït-Kaddour, A., Hassoun, A., Bord, C., Schmidt-Filgueras, R., Biancolillo, A., Di Donato, F., Temiz, H.T., Cozzolino, D., 2021. Application of spectroscopic techniques to evaluate heat treatments in milk and dairy products: an overview of the last decade. Food Bioprocess Technol. 14, 781–803. https://doi.org/10.1007/S11947-021-02607-0.

Ayvaz, H., Mortas, M., Dogan, M.A., Atan, M., Yildiz Tiryaki, G., Karagul Yuceer, Y., 2021. Near- and mid-infrared determination of some quality parameters of cheese

manufactured from the mixture of different milk species. J. Food Sci. Technol. 58, 3981–3992. https://doi.org/10.1007/s13197-020-04861-0.
Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–777. https://doi.org/10.1366/0003702894202201.
Coppola, R., Salimei, E., Succi, M., Sorrentino, E., Nanni, M., Ranieri, P., Blanes, R.B., Grazia, L., 2002. Behaviour of Lactobacillus rhamnosus strains in ass's milk. Ann. Microbiol. 52, 55–60.
Derdak, R., Sakoui, S., Pop, O.L., Muresan, C.I., Vodnar, D.C., Addoum, B., Vulturar, R., Chis, A., Suharoschi, R., Soukri, A., Khalfi, B.El, 2020. Insights on health and food applications of Equus asinus (Donkey) milk bioactive proteins and peptides—An overview. Foods. https://doi.org/10.3390/foods9091302.
Di Donato, F., Biancolillo, A., Mazzulli, D., Rossi, L., D'Archivio, A.A., 2021. HS-SPME/GC–MS volatile fraction determination and chemometrics for the discrimination of typical Italian Pecorino cheeses. Microchem. J. 165, 106133 https://doi.org/10.1016/J.MICROC.2021.106133.
Dos Santos Pereira, E.V., de Sousa Fernandes, D.D., de Araújo, M.C.U., Diniz, P.H.G.D., Maciel, M.I.S., 2021. In-situ authentication of goat milk in terms of its adulteration with cow milk using a low-cost portable NIR spectrophotometer. Microchem. J. 163, 105885 https://doi.org/10.1016/J.MICROC.2020.105885.
European Commission, 2021, Milk market observatory [WWW Document]. URL https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/markets/overviews/market-observatories/milk_en.
European Commission, 2004, Regulation 2004/853 - Specific hygiene rules for food of animal origin, in: EU Monitor.
Ewing, W.M., Allen, P.J., 2005. The diagnosis and management of cow milk protein intolerance in the primary care setting. Pediatr. Nurs. 31, 486–493.
Firmani, P., La Piscopia, G., Bucci, R., Marini, F., Biancolillo, A., 2020. Authentication of P.G.I. Gragnano pasta by near infrared (NIR) spectroscopy and chemometrics. Microchem. J. 152, 104339 https://doi.org/10.1016/j.microc.2019.104339.
Genis, D.O., Sezer, B., Durna, S., Boyaci, I.H., 2021. Determination of milk fat authenticity in ultra-filtered white cheese by using Raman spectroscopy with multivariate data analysis. Food Chem. 336, 127699 https://doi.org/10.1016/J.FOODCHEM.2020.127699.
Guo, H.Y., Pang, K., Zhang, X.Y., Zhao, L., Chen, S.W., Dong, M.L., Ren, F.Z., 2007. Composition, Physiochemical Properties, Nitrogen Fraction Distribution, and Amino Acid Profile of Donkey Milk. J. Dairy Sci. 90, 1635–1643. https://doi.org/10.3168/JDS.2006-600.
Hassoun, A., Shumilina, E., Di Donato, F., Foschi, M., Simal-Gandara, J., Biancolillo, A., 2020. Emerging techniques for differentiation of fresh and frozen-thawed seafoods: Highlighting the potential of spectroscopic techniques. Molecules 25, 4472. https://doi.org/10.3390/molecules25194472.
Haughey, S.A., Galvin-King, P., Ho, Y.-C., Bell, S.E.J., Elliott, C.T., 2015. The feasibility of using near infrared and Raman spectroscopic techniques to detect fraudulent adulteration of chili powders with Sudan dye. Food Control 48, 75–83. https://doi.org/10.1016/j.foodcont.2014.03.047.
Indahl, U.G., Martens, H., Næs, T., 2007. From dummy regression to prior probabilities in PLS-DA. J. Chemom. 21, 529–536. https://doi.org/10.1002/cem.1061.
Jansen, J.J., Hoefsloot, H.C.J., Van Der Greef, J., Timmerman, M.E., Westerhuis, J.A., Smilde, A.K., 2005. ASCA: Analysis of multivariate data obtained from an experimental design. J. Chemom. 19, 469–481. https://doi.org/10.1002/cem.952.
Mabood, F., Jabeen, F., Ahmed, M., Hussain, J., Al Mashaykhi, S.A.A., Al Rubaiey, Z.M. A., Farooq, S., Boqué, R., Ali, L., Hussain, Z., Idrees, M., Manzoor, S., 2017. Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk. Food Chem. 221, 746–750. https://doi.org/10.1016/j.foodchem.2016.11.109.
Pesic, M., Barac, M., Vrvic, M., Ristic, N., Macej, O., Stanojevic, S., 2011. Qualitative and quantitative analysis of bovine milk adulteration in caprine and ovine milks using native-PAGE. Food Chem. 125, 1443–1449. https://doi.org/10.1016/J.FOODCHEM.2010.10.045.
Qannari, E.M., Courcoux, P., Vigneau, E., 2001. Common components and specific weights analysis performed on preference data. Food Qual. Prefer 12, 365–368. https://doi.org/10.1016/S0950-3293(01)00026-X.
Rabatel, G., Marini, F., Walczak, B., Roger, J.-M., 2020. VSN: Variable sorting for normalization. J. Chemom. 34, e3164 https://doi.org/10.1002/CEM.3164.
Ríos-Reina, R., García-González, D.L., Callejón, R.M., Amigo, J.M., 2018. NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin. Food Control 89, 108–116. https://doi.org/10.1016/j.foodcont.2018.01.031.
Salimei, E., Fantuz, F., Coppola, R., Chiofalo, B., Polidori, P., Varisco, G., 2004. Composition and characteristics of ass's milk. Anim. Res 53, 67–78. https://doi.org/10.1051/animres:2003049.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. https://doi.org/10.1021/ac60214a047.
Sjöström, M., Wold, S., Söderström, B., 1986. Partial least squares discriminant plots. In: Gelsema, E.S., Kanal, L.N. (Eds.), Pattern Recognition in Practice II. Elsevier, Amsterdam, pp. 461–470. https://doi.org/10.1016/b978-0-444-87877-9.50042-x.
Smilde, A.K., Jansen, J.J., Hoefsloot, H.C.J., Lamers, R.J.A.N., van der Greef, J., Timmerman, M.E., 2005. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics 21, 3043–3048. https://doi.org/10.1093/BIOINFORMATICS/BTI476.
Snee, R.D., 1977. Validation of regression models: methods and examples. Technometrics 19, 415–428. https://doi.org/10.1080/00401706.1977.10489581.
Souroullas, K., Aspri, M., Papademas, P., 2018. Donkey milk as a supplement in infant formula: Benefits and technological challenges. Food Res. Int. https://doi.org/10.1016/j.foodres.2018.04.051.
Teixeira, J.L., da, P., Caramês, E.T., dos, S., Baptista, D.P., Gigante, M.L., Pallone, J.A.L., 2021. Rapid adulteration detection of yogurt and cheese made from goat milk by vibrational spectroscopy and chemometric tools. J. Food Compos. Anal. 96, 103712 https://doi.org/10.1016/J.JFCA.2020.103712.
Timmerman, M.E., Kiers, H.A.L., 2003. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. Psychometrika 68, 105–121. https://doi.org/10.1007/BF02296656.
Vincenzetti, S., Polidori, P., Mariani, P., Cammertoni, N., Fantuz, F., Vita, A., 2008. Donkey's milk protein fractions characterization. Food Chem. 106, 640–649. https://doi.org/10.1016/J.FOODCHEM.2007.06.026.
Vis, D.J., Westerhuis, J.A., Smilde, A.K., van der Greef, J., 2007. Statistical validation of megavariate effects in ASCA. BMC Bioinforma. 8, 1–8. https://doi.org/10.1186/1471-2105-8-322.
Wold, S., Johansson, E., Cocchi, M., 1993. PLS - Partial least squares projections to latent structures. In: Kubinyi, H. (Ed.), 3D QSAR in Drug Design. Theory, Methods, and Applications. ESCOM Science Publishers, Leiden, pp. 523–550.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: A basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.
Zwanenburg, G., Hoefsloot, H.C.J., Westerhuis, J.A., Jansen, J.J., Smilde, A.K., 2011. ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison. J. Chemom. 25, 561–567. https://doi.org/10.1002/CEM.1400.