Fast Verification of Buffalo’s Milk Authenticity by Mid-Infrared Spectroscopy, Analytical Measurements and Multivariate Calibration

Ben-Hur R. F. Gonçalves,∗a Grazielly J. Silva,a Josane C. de Jesus,a Daniele G. Conceição,a Leandro S. Santosa and Sibelli P. B. Ferrão a,b

∗Programa de Pós-Graduação em Engenharia e Ciência de Alimentos (PPGECAI), Universidade Estadual do Sudoeste da Bahia (UESB), 45700-000 Itapetinga-BA, Brazil
bCentro de Formação em Tecnoc-Ciências e Inovação (CFTCI), Universidade Federal do Sul da Bahia (UFSB), 45613-204 Itabuna-BA, Brazil

Mid-infrared spectroscopy (MID), chemical composition and physicochemical characteristics associated with chemometrics were used to rapidly detect and quantify the amount of cow’s milk added in buffalo’s milk. A total of 165 samples, divided into buffalo’s milk, buffalo’s milk added with cow’s milk (10 to 90%) and cow’s milk were evaluated to obtain fat, protein, lactose, total and defatted solids, urea, pH, acidity, cryoscopic index and band absorbances in the spectra associated with principal component analysis (PCA), multiple linear regression (MLR) and partial least squares regression (PLS). The treatments were separated into groups by PCA, allowing the classification of samples. MLR and PLS models were able to predict cow’s milk contents in buffalo’s milk. MID and results of the analytical measures studied when associated with chemometrics are efficient in the rapid quantitative detection of adulteration in buffalo’s milk.

Keywords: adulteration, bovine, buffaloes, chemometrics, fraud

Introduction

The reproductive seasonality of the buffalo species associated with the high price of buffalo’s milk when compared to cow’s milk are the main factors responsible for the common occurrence of adulteration by partial addition of buffalo’s milk.1 For that, ensuring the authenticity of buffalo’s milk in order to minimize this economic fraud is a concern for regulatory agencies, research laboratories and industries, in which researchers2–12 have demonstrated the importance of development and/or adaptation of analytical methods for detections and quantifications of the type of milk present in the samples evaluated.

Verification of buffalo’s milk authenticity is generally based on the identification of single markers or groups of markers, such as proteins and their fractions, peptides and amino acids by electrophoretic, chromatographic and immunological techniques.13–20 Although these techniques have good sensitivity and generate accurate and reliable results, they are expensive, time consuming, need specialists trained in conducting laboratory activities, use chemical reagents harmful to handlers and the environment, making it difficult to implement them in monitoring programs of industries.15

Thus, it is necessary to adapt techniques, especially non-destructive, through appropriate control methods, to be used in association with traditional/official methodologies to identify the type of milk used by industries.16 Mid-infrared spectroscopy (MID) from numerical analytical data of constituents and their functional groups in the spectra may be a viable alternative.5

The MID stands out for being a fast technique, easy to perform, which requires little preparation of the samples to be evaluated, besides presenting good sensitivity and low operating cost. In the study of milk, characteristic spectra are generated from the vibration of the functional groups that are part of the structure of its chemical constituents. The absorbance data of the bands in the spectra and the amount of the milk constituents, when associated with multivariate statistical analyses, produce accurate answers in the detection of adulteration,12,17,18 enabling the development of mathematical models capable of predicting with reliable results the levels of tampering in the material under study.19
A recent work used MID study to verify the authenticity of buffalo’s milk, experiments were performed with lyophilized milk samples and it was not possible to obtain a model capable of predicting the concentration of cow’s milk in buffalo’s milk, serving only as a screening technique. For fluid milk, MID has not been observed so far to quantify the presence of cow’s milk in buffalo’s milk samples.

A relatively simple alternative, that has not yet been performed to verify the authenticity of buffalo’s milk, is the association of physicochemical characteristics and chemical composition that quantitatively differentiate buffalo’s milk from cow’s milk with multivariate analysis, allowing application of these data, routinely obtained by laboratories of dairy industries that process this type of milk, to the statistical models generated.

In this regard, the objective was to use the MID, chemical composition and physicochemical characteristics associated with multivariate statistical methods for rapid detection and quantification of cow’s milk in buffalo’s milk.

**Experimental**

**Sampling and formulation**

Buffalo and cow’s milks were collected from May to August 2017, during mornings. Bovine milk was obtained by mechanical milking of crossbred cows (Dutch × Zebu) fed on pasture (Brachiaria decumbens), while buffalo’s milk was acquired from manual milking of crossbred buffalo (Jafarabadi × Murrah) females fed on pasture (Brachiaria decumbens), both in adequate hygienic-sanitary conditions. After the collections, the samples were made in 15 repetitions, totaling 165 experimental units, by adding variable and increasing amounts of cow’s milk (10, 20, 30, 40, 50, 60, 70, 80 and 90%) to the buffalo’s, as well as formulations made exclusively with buffalo and cow’s milk.

**Chemical composition, physicochemical characteristics and MID**

A 40 mL aliquot of each milk formulation was heated in water bath of 40 ± 2 °C and analyzed with MID (DairySpec FT, Bentley Instruments, Inc., Chaska, Minnesota, USA) from 1000 to 3000 cm⁻¹, obtaining fat, protein, lactose, total solids (TS), defatted solids (DS), urea and cryoscopic index, as well as acidity measurements were made using 10 mL of samples and 1 mL of phenolphthalein indicator and titrating with NaOH (0.1 mol L⁻¹) and pH determinations were made using the glass electrode method, by a direct reading pH meter calibrated with solutions of pH 7.00 and 4.01. The results of the analytical measurements and the absorbance data of the obtained spectra were used to calibrate the content of cow’s milk added to the buffalo’s milk.

**Chemometrics**

**Experimental design**

The sample data set was organized in matrices formed by m × n elements (m rows corresponding to the treatments and n columns corresponding to the variables), constructed for the measurements of chemical composition and physicochemical characteristics (matrix A₁), absorbance data of specific bands of MID (matrix A₂) and absorbance of the full MID spectra (matrix A₃). The Statistical Analysis System software was used to conduct multivariate statistical analyses and to remove outliers from the Cook’s distance, in which samples with a distance ≥ 2 were discarded individually for each matrix.

**Principal component analysis (PCA)**

PCA was performed with the transformation of the original data matrices (matrices A₁ and A₂) into covariance matrices, which express the individual variances and linear combinations (covariances) between two variables. The eigenvalues and their respective normalized eigenvectors were used to construct the principal components (PC). To set the number of PC’s, the interpretability criteria was considered and the eigenvalue diagram, therefore the PC that had the highest proportions of variance of the original attributes (above 70% of variance) were selected along with the PC’s related to eigenvalues that caused significant changes in variance.

**Multiple linear regression (MLR)**

The original data from the matrix A₁ were standardized (mean value = 0 and standard deviation = 1), eliminating the differences between the measurement units of the variables. Multicollinearity analysis was performed between the variables, so that low multicollinearity values did not influence the estimation of β values of the regression equations, minimizing the errors. Data were randomized and separated into two sets (70% for training and 30% for validation) by the Kennard-Stone algorithm. Three forms of model selection were tested: backward, forward, and stepwise, and the model was chosen based on the highest coefficient of determination (R²), root mean square error of training (RMSET) and root mean square error of validation (RMSEV), along with the evaluation of the number of parameters, the performance/deviation ratio (RPD) and the error rate (RER).
Partial least squares regression (PLS)

Data from matrix A were randomized and divided into training (70%) and validation (30%), evaluated by PLS in order to develop a multivariate calibration model. The number of latent variables in the model was tested according to the method proposed in the software, in which the T² test was used to determine the significant differences of two predicted residual sum of squares (PRESS) values. If there were no significant differences, the models were chosen and only those having fewer factors than the minimum PRESS model were compared. After selecting the number of latent variables, the PLS model was used to predict the concentration of cow’s milk in buffalo’s milk. The values of R², RMSET, RMSEV, correlation coefficient, RPD and RER were obtained to explain the prediction capacity of the model.

Results and Discussion

Chemical composition and physicochemical characteristics

The results of chemical composition and physicochemical characteristics of the evaluated milk samples presented variations when compared to the studied treatments (Table 1).

The milk consists mainly of water, where solid components are dissolved or dispersed, constituting the TS formed by proteins, lipids, carbohydrates, vitamins and minerals. From the total solids, the lipid content is subtracted, obtaining the defatted solids (DS). Due to the peculiar characteristics of the buffalo species, associated with non-genetic factors such as lactation stage, feeding strategies and udder health, among others, buffalo’s milk, when compared to cow’s milk, usually presents higher contents of fat, lactose, protein, TS, DS, vitamins, minerals and acidity due to the amount of proteins with acid characteristics and lower values for water content, cryoscopic index, facilitating the differentiation between the present treatments study.

Mid-infrared spectroscopy (MID) of milk samples

Significant differences were observed in the infrared (IR) spectrum of both bovine and buffalo samples (Figure 1), generated from the identifications of the functional groups of each band in the spectra, in association with the constituents (lipids, proteins or carbohydrates) of the milk.

Thirteen bands were identified (2 bands in the 2977-2814 cm⁻¹ regions for lipids’ C–H vibrations, 2 bands in 1788-1605 cm⁻¹ related to lipids’ C=O, 3 bands in 1574-1214 cm⁻¹ originated due to proteins’ H–N–C=O, 2 bands at 1417-1285 cm⁻¹ of C=O/C–H/C=C vibrations of lipids, 1 band at 1173-1148 cm⁻¹ for C–H/C=O/C of lipids, 1 band at 1127-1097 cm⁻¹ of C–C–O/C–O representing carbohydrates, 1 band at 1102-1057 cm⁻¹ of the vibration C–H/C=O of lipids and 1 band at 1061-1026 cm⁻¹ of the carbohydrates C–C–O/C–O vibrations) from available literature data.

The milk spectra of the two species did not present distinct bands, but differences were observed regarding the absorption intensity. The cow’s milk samples presented bands with lower intensity than the buffalo’s milk samples. This was due to the higher percentages of water in bovine milk samples, thus, lower levels of lipids, proteins and carbohydrates, for the intensity of the vibrational modes of the functional groups is related to the contents of the present constituents, since the intensity of the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum/maximum</th>
<th>Average ± standard deviation</th>
<th>Regression equation</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat / %</td>
<td>3.95/8.84</td>
<td>5.28 ± 0.59</td>
<td>Ŷ = Y = 5.28</td>
<td>b</td>
</tr>
<tr>
<td>Protein / %</td>
<td>2.82/4.27</td>
<td>3.53 ± 0.27</td>
<td>Ŷ = Y = 3.53</td>
<td>b</td>
</tr>
<tr>
<td>Lactose / %</td>
<td>3.89/5.37</td>
<td>4.75 ± 0.32</td>
<td>Ŷ = −0.09248x + 5.30627</td>
<td>0.84</td>
</tr>
<tr>
<td>TS / %</td>
<td>12.47/16.90</td>
<td>14.48 ± 1.01</td>
<td>Ŷ = −0.3055x + 16.30974</td>
<td>0.92</td>
</tr>
<tr>
<td>DS / %</td>
<td>7.91/10.36</td>
<td>9.20 ± 0.55</td>
<td>Ŷ = −0.16526x + 10.1915</td>
<td>0.91</td>
</tr>
<tr>
<td>Urea / %</td>
<td>10.70/35.10</td>
<td>21.75 ± 6.44</td>
<td>Ŷ = 0.04525x² − 2.49756x + 34.65387</td>
<td>0.93</td>
</tr>
<tr>
<td>Cryoscopic index</td>
<td>−0.630/−0.580</td>
<td>−0.610 ± 0.0069</td>
<td>Ŷ = Y = −0.61</td>
<td>b</td>
</tr>
<tr>
<td>pH</td>
<td>6.56/6.80</td>
<td>6.69 ± 0.05</td>
<td>Ŷ = Y = 6.69</td>
<td>b</td>
</tr>
<tr>
<td>Acidity / °D</td>
<td>10.00/20.00</td>
<td>14.05 ± 1.53</td>
<td>Ŷ = Y = 14.05</td>
<td>b</td>
</tr>
</tbody>
</table>

aCoefficient of determination; b could not adjust linear or quadratic model regression equations (p > 0.05); c total solids; d defatted solids; e cryoscopic index expressed in °Hortvet (°H); f acidity expressed in °Dornic (°D). Ŷ: response variable; Y: mean value of the respective analytical measure; x: values obtained in the analytical measures to generate the response variable.
vibrational modes is proportional to the concentration of the constituents.

The spectra of buffalo and bovine treatments were overlapped on 9 levels of tampering (10 to 90%) (Figure 2).

Spectra with the lowest levels of adulteration showed similar behavior to buffalo’s and, as cow’s milk was added to buffalo’s milk, the treatments had characteristics similar to bovine milk. The peaks are caused due to the vibration of functional groups present mainly in the structure of milk fat, protein and lactose, so that any changes in the contents of these components caused variations in the behavior of the milk spectrum bands, due to peculiar characteristics in the milk of the different species.

Chemometrics

Principal component analysis (PCA)

The classification of buffalo and bovine treatments by PCA (Figure 3) showed differences between the samples of the two species, as well as the relationships between the evaluated milk formulations and the analytical measures. By the interpretability criteria, associated with the eigenvalue diagram, 2 PC’s with the highest proportions of variance of the original attributes were selected.

For the matrix A1, the principal component 1 (PC1) explained 95.51% and the principal component 2 (PC2) 3.62% of the total data variance, sufficing to discriminate the treatments regarding the chemical composition and

![Figure 1. MID of buffalo and cow’s milk samples.](image1)

![Figure 2. MID in the regions between 3000 and 2800 cm\(^{-1}\) and between 1800 and 1000 cm\(^{-1}\).](image2)
physicochemical characteristics studied. The most relevant data information is contained in PC1, which behaved as a combination of fat, protein, lactose, TS, DS, urea and acidity (positive measures) against cryoscopic index and pH (negative). Thus, buffalo’s milk samples had superior results of positively correlated measurements when compared to cow’s milk samples, hence the spatial separation into two distinct groups by PCA.

For the matrix A₂, PC1 explained 87.72% of the total data variance, while PC2 6.54%, thus the samples also separated themselves in relation to PC1. All variables (13 bands) were significantly correlated with PC1. As cow’s milk was added to buffalo’s milk, the absorbance values of the bands decreased, causing changes in the location of samples with higher percentages of bovine milk to negative regions of PC1, since there were lower vibration intensities of the functional groups present in the sample constituents, due to lower presence of constituents in cow’s milk when compared to buffalo’s. Thus, buffalo’s and cow’s milk samples presented distinct characteristics. For the differentiation between the treatments, the 8 fat-related bands, 3 protein-related bands and the 2 lactose-related bands influenced.

The PC’s, both in relation to composition (Figure S1, Supplementary Information (SI) section) and absorbance data (Figure S2, SI section), were able to explain most of the total variance of the original data, with the highest explanation percentages being attributed to PC1 (significant positive correlation between fat, protein, lactose, TS, DS, urea and acidity levels for matrix A₁ and for all 13 absorption peaks of matrix A₂). As adulteration rates increased in buffalo’s milk, samples moved toward the negative PC1 scale, which displays a decrease in the constituent values (except pH and cryoscopic index) and, consequently, in the absorbance of the 13 bands of the spectra, allowing the separation/classification into distinct groups. By PCA (Figures S1 and S2, SI section), it can be stated that from 20% of adulteration it is possible to observe the separation of buffalo’s milk samples from adulterated samples and cow’s milk samples.

Multiple linear regressions (MLR)

The MLR model was chosen by the backward technique (Figure 4a). The multicollinearity between the variables was low (≤ 100), hence there was no high degree of correlation between the independent variables, thus the parameters were estimated with precision. The model was able to predict the amount of bovine milk in buffalo’s milk (Figure 4b) from the chemical composition and physicochemical characteristics data (matrix A₁), since there were no high errors in the comparisons between the values predicted by the models and the experimental values.

Among the composition variables and physicochemical characteristics only cryoscopic index, protein and TS were not significant, therefore they are not relevant for the quantification of cow’s milk content in buffalo’s milk. The model presented $R^2 = 0.9844$, RMSET = 4.73% and RMSEV = 4.04%, correlation = 0.9922 (for the validation data), besides low multicollinearity, RPD = 7.89 and RER = 24.78, thus, the model is able to predict a constituent with good practical applicability. This demonstrates that variables that are commonly obtained on a daily basis or are part of the routine of milk quality control laboratories can be used to verify the authenticity of buffalo’s milk. As the percentages of cow’s milk increased, the components of buffalo’s milk were diluted, allowing the quantification of cow’s milk in the formulations, enabling the verification of the authenticity of buffalo’s milk by the dairy industries that use buffalo’s milk as feedstock.
Partial least squares regression (PLS)

Spectral data (matrix $A_k$) were used to predict the levels of cow’s milk in buffalo’s milk by PLS, a statistical technique that allows the adjustment of highly correlated variables, a common fact in MID data. The adulteration levels studied were quantified due to the decrease in peak intensities of the MID spectra as cow’s milk was added to the formulations (Figure 2).

The $R^2$ of the training model was 0.9877, requiring 7 latent variables to explain the variance of the data with greater precision. The correlation between the predicted values by the model and the experimental values for the validation data was 0.9938 (Figure 5), with RMSET and RMSEV of 3.484 and 3.481%, respectively, and RPD = 9 and RER = 28.72, indicating good ability to detect and quantify the presence of cow’s milk in buffalo’s milk with good practical applicability.

Some studies used fast methods associated with multivariate statistics to verify the authenticity of buffalo’s milk. Silva et al.,9 in a study with lyophilized milk samples and ATR-FTIR (attenuated total reflectance-Fourier transform infrared spectroscopy) associated with PCA and artificial neural network (ANN), found differences between the milk of the species only from 40% of adulteration, and could not assess the content of cow’s milk added to the buffalo’s milk, only its presence. On the other hand, Velioglu et al.,15 when researching the authenticity of buffalo’s milk by fluorescence spectroscopy associated with PCA and PLS found models with $R^2$ of 0.98 and RMSET and RMSEV of 2 and 4%, respectively, results similar to the ones found in this search.

Regarding the effectiveness of the tested models, it was found that both models (MLR and PLS) presented compatible results. The PLS modeling for the spectra data and MLR from the results of chemical composition and physicochemical characteristics proved to be efficient to verify the authenticity of buffalo’s milk, with the capacity of quantification when cow’s milk is present in the samples assessed, since they have good RMSE ($\leq 4.73\%$), $R^2$ ($\geq 0.9844$) and correlation index ($\geq 0.9922$). The

---

**Figure 4.** MLR: (a) $\beta$ coefficients of the regression equation; (b) correlation between actual and predicted MLR values.

**Figure 5.** PLS: (a) $\beta$ coefficients of the regression equation; (b) correlation between actual and predicted values generated by PLS.
MID technique, in addition to being a reliable, fast and accurate technique, enabling its use in dairy industries and surveillance laboratories, can be applied directly to milk without the need for sample preparation and the use of reagents harmful to handlers and the environment. Data on chemical composition and physicochemical characteristics can be used by research laboratories that routinely analyze cow’s and buffalo’s milk.

Compared to cow’s milk, buffalo’s milk has higher contents and is marketed at high prices. Associated with this, there is the issue of decreased availability of buffalo’s milk at certain times of the year, usually in the months with spring-summer seasons. The use of chemometric combined with physicochemical characteristics and MID data can detect the adulteration in buffalo’s milk starting at 10% of addition of cow’s milk, offering alternative strategies that may provide greater power of inspection to the dairy industry that perform the processing of buffalo’s milk and the supervisory authorities. The authenticity assessment of buffalo’s milk in the samples collection places using portable devices can provide a fast and cheap option for the results obtention, as long as these devices can correctly quantify the value os the chemical composition and physicochemical characeristics and/or MID used in this study.

Conclusions

Techniques have been developed to verify the authenticity of buffalo’s milk, based on analyses of chemical composition and physicochemical characteristics performed routinely and from MID spectra. Both techniques, when associated with chemometrics, showed compatible results in the detection of tampering.

PCA allowed the differentiation between the treatments in different regions in the two-dimensional graphs, allowing the classification of the samples according to the levels of tampering. The MLR and PLS models applied to the chemical composition, physicochemical characteristics and MID data were effective in quantifying cow’s milk levels in buffalo’s milk.

Due to the accuracy and speed of MID analyses, the use of the technique becomes a viable alternative to verify the authenticity of buffalo’s milk.

Supplementary Information

The supplementary information (PCA of adulterated milk samples (10, 20, 30, 40, 50, 60, 70, 80 and 90%), buffalo’s and bovine spectra peak (MID)) is available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

The authors acknowledge the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for providing study scholarships.

Author Contributions

Ben-Hur R. F. Gonçalves was responsible for the conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, validation, visualization, writing original draft, review and editing; Grazielly J. Silva for the data curation, investigation, project administration, visualization and writing original draft; Josane C. de Jesus and Daniele G. Conceição for the data curation, investigation and project administration, visualization and writing original draft, review and editing; Sibelli P. B. Ferrão for the conceptualization, project administration, visualization and writing review and editing; Leandro S. Santos for the conceptualization, project administration, visualization and writing review and editing; Caira, S.; Nicolai, M. A.; Lilla, S.; Calabrese, M. G.; Pinto, G.; Scalon, A.; Chianese, L.; Addeo, F.; Pizzano, R.; Nicolai, M. A.; Cairsa, S.; Chianese, L.; J. Agric. Food Chem. 2009, 57, 10063.


Submitted: November 11, 2019
Published online: February 27, 2020