

**CHEMOMETRIC TOOLS AND FTIR-ATR SPECTROSCOPY APPLIED IN MILK ADULTERATED WITH CHEESE WHEY****Layane L. Vinciguerra<sup>a</sup>, Marcelo C. A. Marcelo<sup>b</sup>, Tanara M. C. Motta<sup>b</sup>, Leonardo Z. Meneghini<sup>a</sup>, Ana M. Bergold<sup>a</sup> and Marco F. Ferrão<sup>b,\*</sup>**<sup>a</sup>Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, 90610-000 Porto Alegre – RS, Brasil<sup>b</sup>Instituto de Química, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre – RS, Brasil

Recebido em 02/08/2018; aceito em 07/01/2019; publicado na web em 18/02/2019

Brazilian law forbids the addition of cheese whey in milk. However, adulteration with cheese whey is one of the most applied fraud due to its low cost. The detection of this fraud is the quantification of Caseinomacropeptide (CMP). The CMP is a constituent of the whey that can be used as adulteration marker. Thus, an analytical method capable of identifying CMP by Fourier Transform Infrared Spectra (FTIR) was developed using chemometrics methods. Firstly, we attempted to develop an exploratory analysis model by Hierarchical Grouping Analysis (HCA) and Principal Component Analysis (PCA) that indicated similarity between samples of raw milk and semi-skimmed milk. Moreover, in the PCA scores, it was possible to observe a tendency of separation between samples with different concentrations of CMP. Afterwards, multivariate regression models were used for Partial Least Squares (PLS), Partial Least Square with Interval Synergism (siPLS) and Supporting Machines with Least Squares (LS-SVM) to quantify the adulteration in different types of milk by Cheese serum through the CMP. All the models were then compared to each other and the results of the official method with Liquid Chromatography Tandem mass spectrometry (LCMS/MS) analysis used by the Ministry of Livestock and Supply (MAPA). The model LS-SVM, employing the full spectrum, obtained the best result compared to the other models (PLS and siPLS) to quantify the CMP in the milk samples.

Keywords: milk; adulteration; CMP; cheese whey; FTIR-ATR spectroscopy; multivariate regression.

**INTRODUCTION**

In recent years in Brazil, it were discovered several frauds in bovine milk production and in its transportation. One of the most profitable milk frauds recently discovered was the cheese whey addition, which is prohibited by Brazilian law. This milk adulteration leads to a fat content dilution and a reduction of its nutritional value, since each 10% volume of cheese whey added reduces 8% the milk protein content. The cheese whey is a milk by-product and, consequently, it is one of the most troubling frauds to detect.<sup>1</sup> Moreover, the milk adulteration by cheese whey became easy to be performed in, since the milk industry does not have a rapid way to detect the fraud.

Nowadays, the Brazilian official methods for detect cheese whey addition in milk are based on caseinomacropeptide (CMP) detection through liquid chromatography coupled with mass spectrometry (LC-MS) or capillary electrophoresis.<sup>2</sup> The CMP is formed from the  $\kappa$ -casein cleavage through the chymosin action and it is present only in the cheese way and, therefore, can be used as adulteration marker.<sup>3</sup>

Several approaches to CMP analysis in milk have been reported in the literature, such as high-performance liquid chromatography (HPLC) used to determine the CMP index and detect cheese adulteration of whey,<sup>4</sup> Reversed-phase high-performance liquid chromatography (RP-HPLC) for determination of CMP and monitoring of the addition of rennet whey to powdered milk,<sup>5</sup> high-performance liquid chromatography- mass spectrometry (LCMS-MS) for determination of CMP in brazilian bovine milk,<sup>6</sup> induced breakdown spectroscopy (LIBS) to detect and quantify adulterated milk powder through adding whey powder by using laser,<sup>7</sup> near infrared spectroscopy (NIR) for quantification of common adulterants including whey cheese in powdered milk for example.<sup>8</sup>

Fourier Transformed Infrared Spectroscopy by attenuated total reflectance (FTIR-ATR) associated with chemometrics has been applied in researches as an alternative method for food matrices with low cost, rapid analysis and little or none sample pre-treatment, non-destructive and/or non-invasive avoid, high reagent consumption and has potential for portability. Moreover, the application of multivariate techniques in infrared spectroscopy enabled quantitative approach.<sup>9,10</sup> The aim of this work was to develop a FTIR-ATR method associated with chemometrics to quantify CMP in different types of milk and to compare the results with the Brazilian official method.<sup>11</sup>

**EXPERIMENTAL****Samples and reagents**

The ultra-high temperature (UHT) milk samples (skimmed and semi-skimmed) were acquired in local markets as well as raw milk. The water used was purified through Milli-Q system (Millipore). It was also used in this work acetonitrile (Tedia), methanol (J. T. Baker), glycine (Vetec), acetic acid (J. T. Baker), porcine gastric mucous pepsin (Sigma-Aldrich), trichloroacetic acid (Merck) and formic acid (Merck). All reagents had purity grade for residue analysis. The CMP standard (Davisco Foods, EUA) and a synthetic pepsin digestion peptide (MAIPPKKNQDKTEIPTINT, Mimotopes, Australia) were used as analytical standard.

**SAMPLE PREPARATION**

Eighteen spiked milk samples were prepared by addition of 75 mg of CMP standard in 25 mL of milk and subsequent dilution to the final concentration of 60, 120, 180, 240, 300 and 360 mg L<sup>-1</sup> to 10 mL completed with respective milk (Figure 1). The sample preparation procedure for LC-MS/MS analysis was as follows: 1 mL of sample

\*e-mail: mfferrao@gmail.com

and 500  $\mu\text{L}$  of trichloroacetic acid were added to a centrifuge tube and mixed in a vortex for 1 min. These mixed samples were placed in an ultrasonic bath for 30 min followed by centrifugation in 1200 rpm for 10 min. A 200  $\mu\text{L}$  of the supernatant solution, a 200  $\mu\text{L}$  of 1 mol  $\text{L}^{-1}$  glycine solution and 50  $\mu\text{L}$  of pepsin solution (10  $\mu\text{g mL}^{-1}$ ) were added in 1.5 mL vial and left for 24 h at 37  $^{\circ}\text{C}$ . After this period, the solution was direct analysed. External calibration was used in liquid chromatography.

### Instrumentation

For milk analysis and separation of constituent digestion products were performed in a LCMS/MS system API 5000 AB Sciex (Foster City, CA, USA) coupled with a liquid chromatography 1100 Series (Agilent). The column was a PLRP-S (polystyrene-divinylbenzene), 150 x 4.6 mm<sup>2</sup>, 300 Å (Polymer Technologies, Varian). Quantitative analysis was obtained in MRM mode, using at least two transitions for each molecular ion. Mass spectrometer parameters for ionization and fragmentation were optimized using synthetic peptides standards injection by infusion followed by flow injection analysis (FIA). Mobile phase was composed by ultra-pure water (A) and acetonitrile (B), both with 0.1% of formic acid. Mobile phase flow was 600 mL  $\text{min}^{-1}$  and a gradient mode was used. Initial conditions were 10% of B in A, increasing to 60% from 2 to 5 min, holding for 5 min and returning to original composition in 2 min, for a total analysis time of 15 min. Equilibrium time was 2 min. A Triple quadrupole mass detector with electrospray ionization source (ESI) in positive mode was used for detection and quantification of targeted fragments. Turbo ion spray voltage was optimized at 5500 V and temperature source was 650  $^{\circ}\text{C}$ . Other optimized parameters are: EP=10 V; CAD=12 V; CUR=10 V; GS1=45 psi; GS2=55 psi Multiple reaction monitoring (MRM) conditions, typical retention time and optimal declustering potential (DP), collision energies (CE), collision cell exit potential (CXP) in the MS/MS mode for the product ions generated.<sup>3</sup> The fragmentation parameters were defined through a theoretical digestion by the Skyline software.

The infrared spectra were obtained in a FTIR spectrophotometer Cary 630 (Agilent Technologies) coupled with an attenuated total reflectance accessory with ZnSe crystal and helium/neon laser. The

spectral range was 4000 to 800  $\text{cm}^{-1}$ , 4  $\text{cm}^{-1}$  resolution and 32 scans. All samples were analyzed in duplicate, at room temperature, in random order and with a background between each duplicate.

### Multivariate Analysis

The software Matlab 8.1 (MathWorks Inc. Natick, USA), Chemostat,<sup>12</sup> the iToolbox<sup>13</sup> and the PLS\_toolbox were used in the multivariate analysis. Before the multivariate analysis, the infrared spectra were smoothed by Savitzky-Golay algorithm (1st degree polynomial, 10 points per window) for noise reduction, normalized to maximum value equal to unity to equalize the weight of each sample in the posterior regression. At last, the smoothed and normalized data were mean-centered.

The pre-processed infrared spectra as input for principal component analysis (PCA) and hierarchical component analysis (HCA) for exploratory analysis were used. The HCA was performed using the Euclidean distance and the Ward method. Partial least square regression (PLSR), synergy interval partial least square regression (siPLSR) and least square support vector machine regression (LS-SVM) algorithms were used for multivariate regression. The pre-processed spectra were used as input for PLSR. Moreover, the infrared spectra were divided in 8, 16, 32, 64 and 128 intervals with equal number of variables and the PLSR was performed in each interval. In addition, combination of these intervals two by two, three by three and four by four were made and the PLSR was performed in each combination. This approach (divide, combine) is called siPLSR. The LS-SVM was performed only in the best regression variables defined by siPLSR as well as the entire spectra. The best regression variables were defined as the variables with smaller root mean square error for cross validation (RMSECV). The meta-parameters of LS-SVM and the latent variables of PLSR and siPLSR were selected through leave-one-out cross validation method in order to minimize RMSECV. The root means square error for calibration (RMSEC), root mean square error for prediction (RMSEP) and the determination coefficient ( $R^2$ ) were also evaluated. The concentrations determined by LC-MS/MS were used as reference values for the multivariate regression. The Kennard-Stone algorithm was used to separate the samples between a calibration set (two-thirds) and a validation set (one-third).<sup>14</sup>

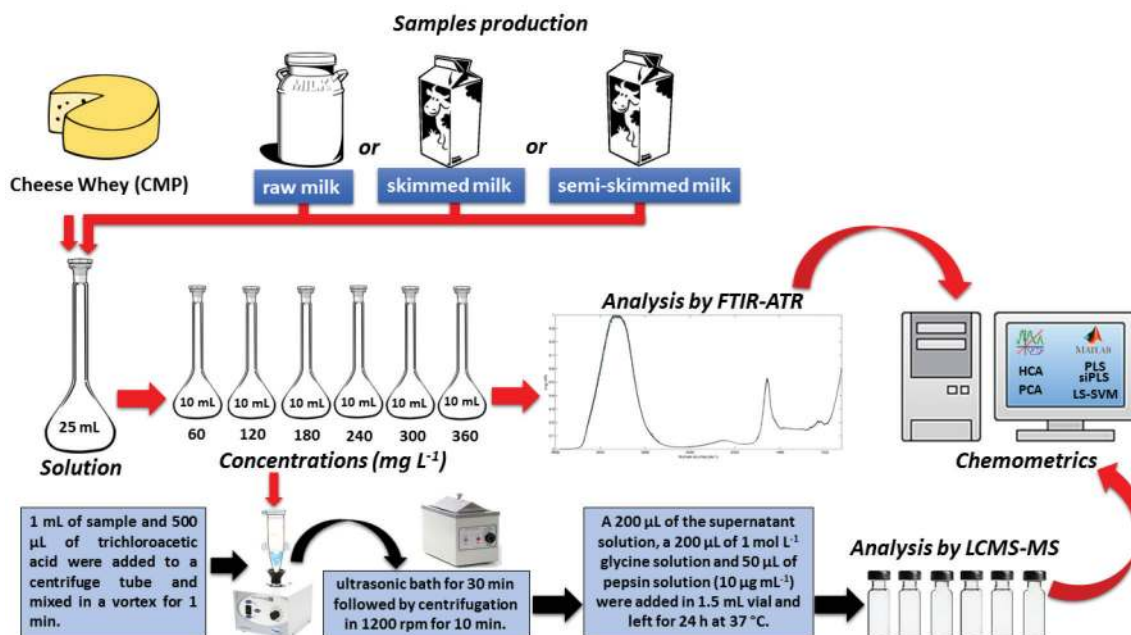


Figure 1. Graphical abstract of experimental steps of production and analysis of studied milk samples by FTIR-ATR

## RESULTS AND DISCUSSION

### Exploratory analyses

Figure 2 illustrates the infrared spectrum of the adulterated milk samples in the range of 4000-800  $\text{cm}^{-1}$ . The absorption bands in the regions of 3700-3000  $\text{cm}^{-1}$  and 1700-1500  $\text{cm}^{-1}$  are related to vibrational combination and angular deformation, respectively, of water hydroxyl group, since it is the major milk component.<sup>15</sup> The absorption band in the 2200-2000  $\text{cm}^{-1}$  region is related to the asymmetric  $\text{CO}_2$  stretching.<sup>16</sup> The amide absorption bands that could be associated with the presence of CMP are supposed to be in the region 1200-1000  $\text{cm}^{-1}$  and 1690-1670  $\text{cm}^{-1}$ . However, their bands were overlapped by the water bands.

The PCA and HCA were carried out in the pre-processed spectra to evaluate the milk types differences (skimmed (LD), semi-skimmed (LS) and raw (LC)) in the adulterated samples before the

quantification step. The dendrogram and the scores of PCA were depicted in the Figure 3, respectively. The three milk types are well defined in the dendrogram: the group 1 (red) is the skimmed milk, the group 2 (green) is the raw milk and the group 3 (blue) is the semi-skimmed milk. There are two exceptions: a skimmed sample (LD30), which has a large dissimilarity with the other samples, and a raw milk sample (LC5), which is more similar with the semi-skimmed than with the other raw milk samples, probably due to the change in the coordinates in which the distance is calculated after the new group is formed, since the raw and semi-skimmed samples are very close to each other. No pattern related to the CMP concentration was noticed in the dendrogram.

According to the Figure 4 A), the first principal component (PC1), with 46.35% of the explained variance, separates the samples between two PCs: the raw milk samples in the negative side of PC1 and the skimmed and semi-skimmed milk samples in the positive side. The samples distribution along the second principal component (PC2),

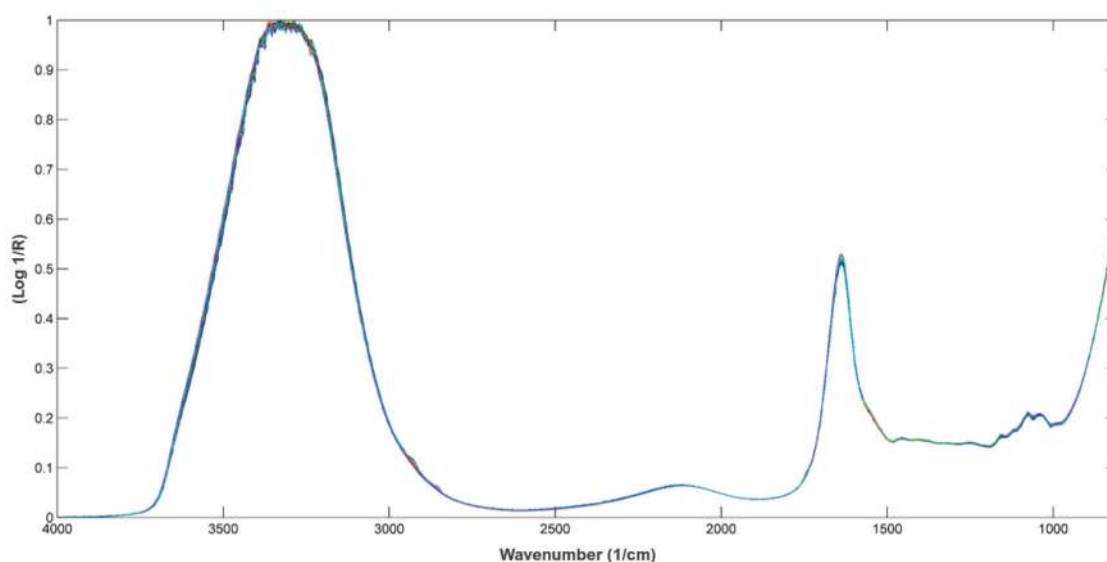


Figure 2. Spectra of the raw milk samples adulterated with CMP

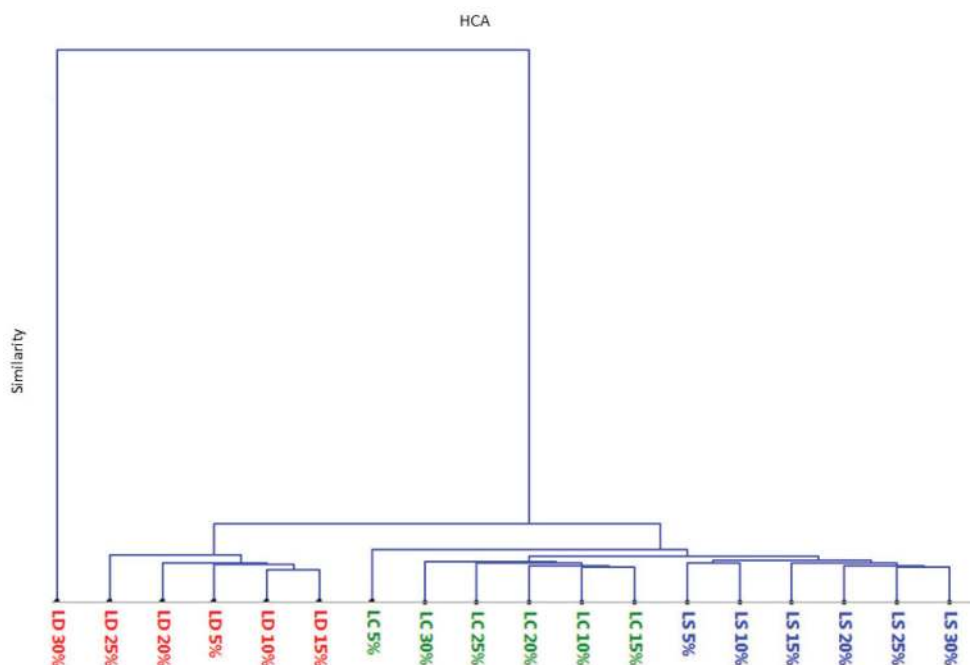
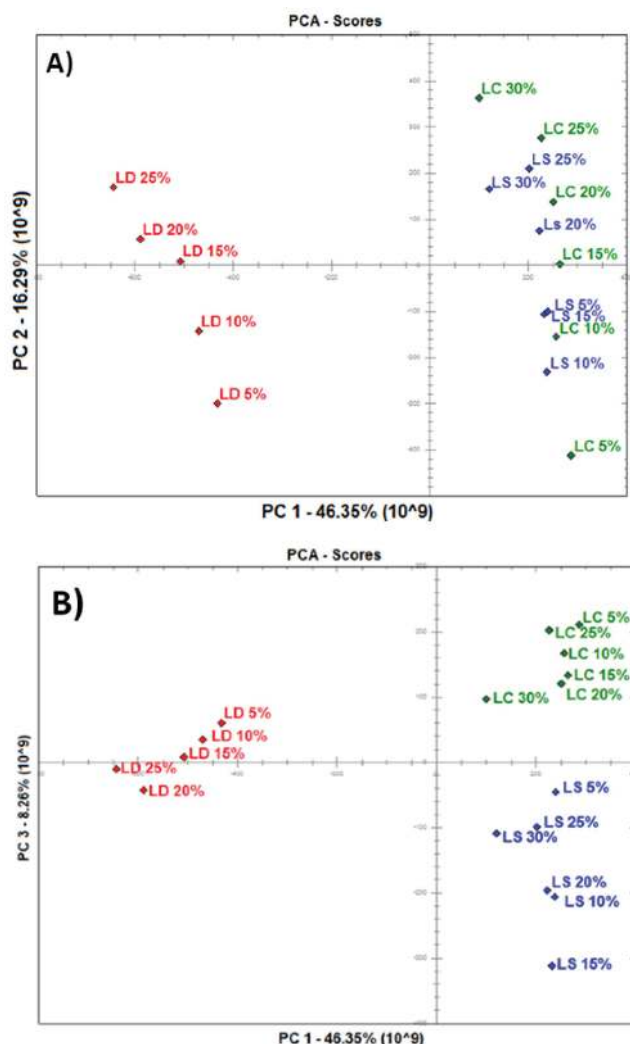
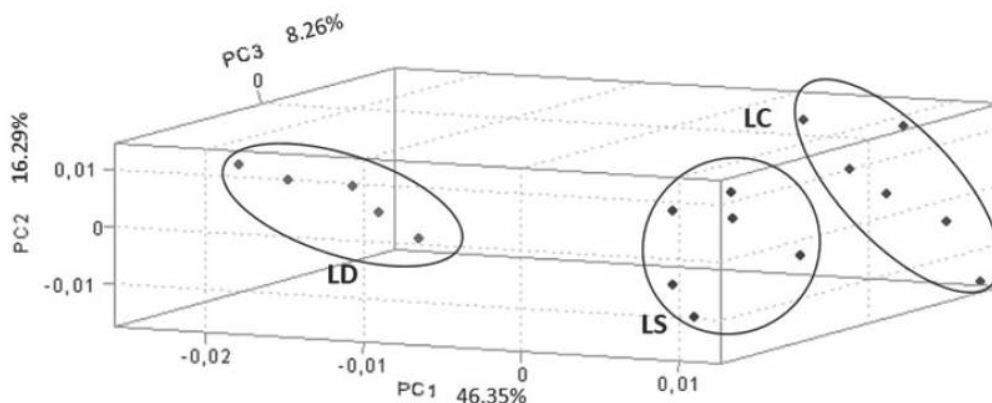


Figure 3. Dendrogram of the milk samples adulterated with CMP. LC – raw milk (green); LD – skimmed milk (red); and LS – semi-skimmed milk (blue)

with 16.62% of explained variance, describes a pattern in relation of CMP concentration, since the concentration increase with the PC2 scores, with exception of LS5. The third principal component (PC3), with 8.26% of total variance explained, separates the raw, in the positive side, and semi-skimmed milk samples, in the negative side, Figure 4 B). Figure 5 shows the scores of the first three main



**Figure 4.** A) Milk samples adulterated with CMP in the PCA scores of PC1 and PC2. LC – raw milk; LD – skimmed milk; and LS – semi-skimmed milk and B) Milk samples adulterated with CMP in the PCA scores of PC1 and PC3. LC – raw milk; LD – skimmed milk; and LS – semi-skimmed milk



**Figure 5.** Scores of 3D graph of PCA for samples of milk adulterated with CMP

components where we can observe the clear separation of the three types of milk used.

The PC1 loadings (Figure 6 (a)) showed a major influence of water related bands, which explained the separation between the skimmed milk to the others. On the other hand, the PC3 loadings (Figure 6 (b)) showed a major influence of the absorption bands in the 1700-1500 cm<sup>-1</sup>.

## Multivariate Regression

The pre-processed spectra were submitted to multivariate regression algorithms with the intent of develop a CMP quantification methodology with infrared spectra as input. The CMP concentrations determined by LC-MS/MS were used as reference. The CMP values determined by LC-MS/MS were 0.0053-0.4076 g L<sup>-1</sup> and 0.0166-0.3976 g L<sup>-1</sup> for the set of calibration and prediction samples, respectively. The samples were divided in two groups: a calibration set, in which the spectra and concentrations were used for model development, and a prediction set, in which the spectra and concentrations were used to validate the model. The PLSR, siPLSR and LS-SVM were used as multivariate algorithms. The regression parameters and the latent variables used for PLSR (M1), the best intervals and interval combinations (siPLSR) (M2-M8) as well as the regression parameters for LS-SVM (S1-S6) were depicted in Table 1. Employing paired t test with 95% confidence, for the prediction set, all models by PLS presented significantly different results, except (M1 and M4) and (M5 and M8) that presented equivalent results among themselves. For the models by LS-SVM all presented significantly different results.

The PLSR model (M1) with 7 latent variables presented a satisfactory RMSEC, RMSECV and RMSEP as well as a high R<sup>2</sup>. In the siPLSR two models of combinations two by two (M2 and M3), two models of combination three by three (M4 and M5) and three models of combinations four by four (M6 and M7) due to its good regression parameters were selected. However, not all interval combinations improved the parameters in relation of M1. Figure 7a shows the pure CMP spectrum. The variance importance index scores (VIP) (Figure 7 B) showed a major influence of three regions: 3300-3200 cm<sup>-1</sup>, 1700-1600 cm<sup>-1</sup> and 1200-1000 cm<sup>-1</sup>.<sup>17</sup> These regions are in agreement with the main signals present in the pure CMP spectrum, being the most representative ones for the construction of the calibration models employing the infrared spectra.

Figures 8 and 9 show the explained variance per latent variable and the regression model for M1, respectively.

The LS-SVM models (S1-S6) were, in general, better in relation regression parameters than the PLSR models. The best LS-SVM regression was also the entire spectra model (S1) (Figure 10) that was even better than the M1 model. The S1 presented a similar



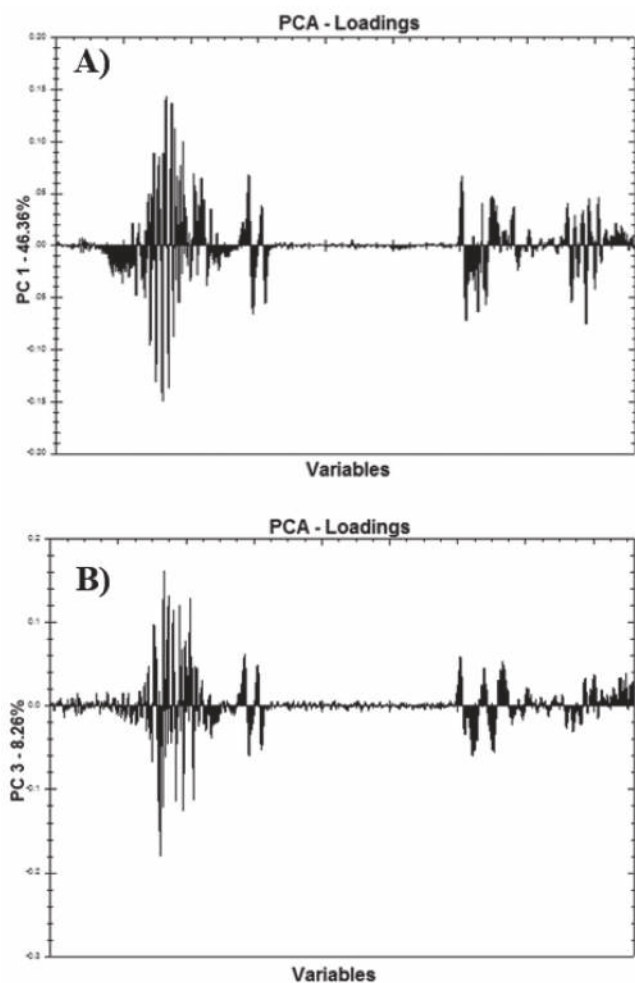


Figure 6. Graph loadings of PC1 (A) and PC3 (B)

RMSEC, RMSECV and  $R^2$ , but a better RMSEP in relation to the others. Moreover, the variables reduction for quantification of CMP in milk did not improve the model, probably because the information that can be related to the adulteration was overlapped in all spectra.

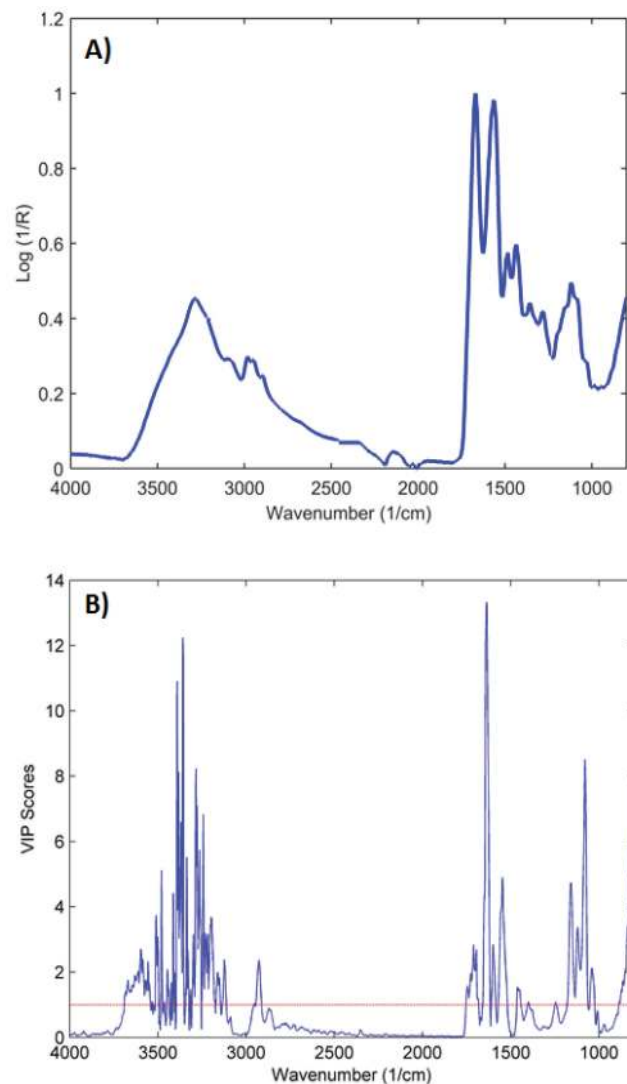


Figure 7. A) CMP FTIR spectrum and B) Variance importance index for PLSR with entire spectra. The values above the red line have more influence in the model

Table 1. Regression parameters of the PLSR, siPLSR and LS-SVM models to quantify CMP in milk

Model <sup>a</sup>	Interval (cm <sup>-1</sup> )	LVs <sup>a</sup>	RMSECV (g L <sup>-1</sup> )	RMSEC (g L <sup>-1</sup> )	R <sup>2</sup> <sub>cal</sub>	RMSEP (g L <sup>-1</sup> )
M1	4000-800	7	0.0250	0.0359	0.9910	0.0542
M2	1997-1598, 3599-3200	9	0.0214	0.0070	0.9955	0.0568
M3	3296-3271, 3348-3323	12	0.0197	0.0092	0.9923	0.0466
M4	1997-1598, 2798-2399, 3599-3200	8	0.0204	0.0081	0.9940	0.0540
M5	1795-1598, 2792-2595, 3396-3197	11	0.0186	0.0037	0.9980	0.0626
M6	1997-1598, 2798-2399, 3599-3200, 4000-3601	9	0.0204	0.0068	0.9958	0.0565
M7	1795-1598, 2792-2595, 2994-2794, 3396-3197	11	0.0181	0.0036	0.9988	0.0618
M8	1786-1737, 1887-1838, 3296-3247, 3346-3297	10	0.0149	0.0044	0.9983	0.0628
S1	4000-800	-	0.0260	0.0056	0.9981	0.0409
S2	1997-1598, 2798-2399, 3599-3200	-	0.0328	0.0010	0.9999	0.0522
S3	1795-1598, 2792-2595, 3396-3197	-	0.0314	0.0048	0.9986	0.0463
S4	1997-1598, 2798-2399, 3599-3200, 4000-3601	-	0.0327	0.0009	0.9999	0.0542
S5	1795-1598, 2792-2595, 2994-2794, 3396-3197	-	0.0306	0.0052	0.9981	0.0415
S6	1786-1737, 1887-1838, 3296-3247, 3346-3297	-	0.0313	0.0043	0.9992	0.0660

<sup>a</sup>LVs – Latent variables.

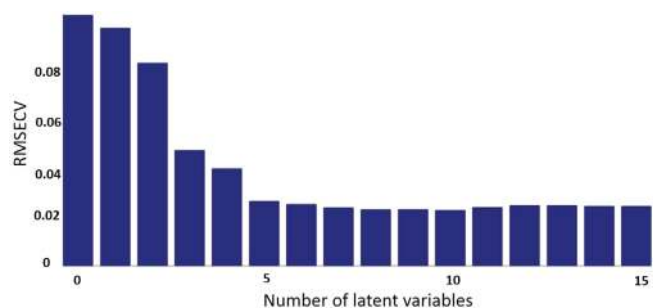


Figure 8. Explained variance per latent variable in cross validation for PLSR of entire spectra

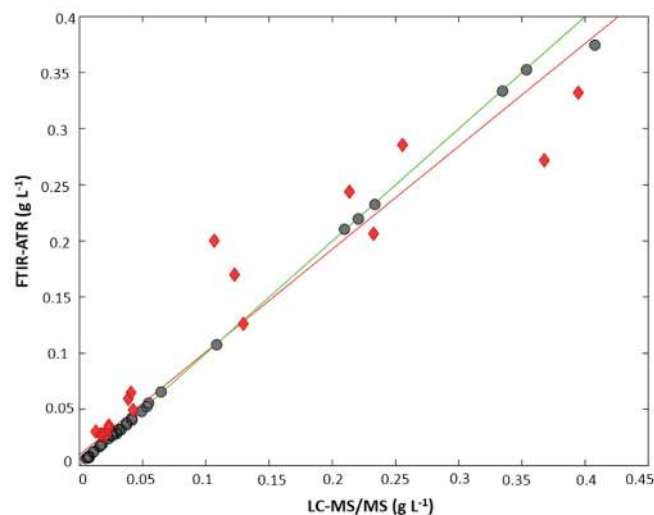


Figure 9. LC-MS/MS versus the FTIR-ATR/PLS CMP concentrations with entire spectra. Spheres are the calibration samples; rhombus are the predicted samples

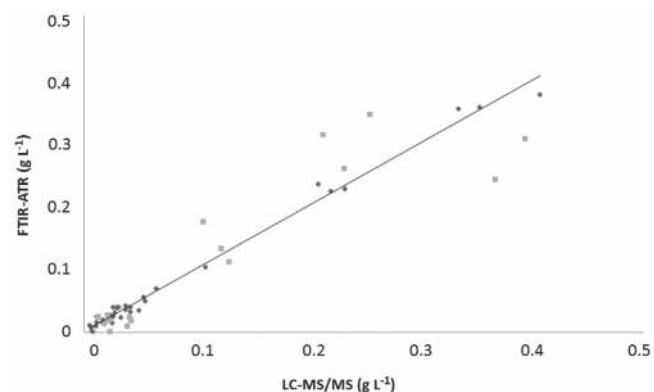


Figure 10. LC-MS/MS versus the FTIR-ATR/LS-SVM CMP concentrations with entire spectra. Rhombus are the calibration samples; squares are the predicted samples

At last, the two models S1 and M1 presented good results for CMP quantification in milk, with good RMSEC, RMSECV, RMSEP and  $R^2$ . In addition, the LS-SVM presented a better result mainly due to its generalization capacity avoiding the over fitting.<sup>18-20</sup>

## CONCLUSIONS

A methodology to quantify CMP in milk samples through FTIR-ATR associated with chemometrics was developed. The HCA and

PCA separated the three milk types according to its fat and water content. In addition, the PCA showed a separation trend related to CMP content. The PLSR, siPLSR and LS-SVM models were developed and had good regression parameters. The best models used all spectra range, which indicates that the CMP information was overlapped in all spectra. The methodology developed with LS-SVM algorithm showed a better result in relation to the PLSR algorithm. Lastly, the methodology developed was suitable to detect CMP adulteration and quantify it.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), (Projeto n°. 133642/2014-3) for their financial support and scholarships.

## REFERENCES

- Oliveira, G. B.; Gatti, M. D. S.; Valadão, R. C.; Martins, J. F. P.; Luchese, R. H.; *Rev. Inst. Laticínios Cândido Tostes* **2009**, *64*, 56.
- Ministério da Agricultura, Pecuária e Abastecimento (MAPA), *Instrução Normativa n° 07*, 2010, available at <https://www.diariodasleis.com.br/legislacao/federal/213634-determinacao-de-cmp-caseinomacropetideo-em-leite>, accessed in February 2019.
- Motta, T. M. C.; Hoff, R. B.; Barreto, F.; Andrade, R. B. S.; Meneghini, L. Z.; Pizzolato, T. M.; *Talanta* **2014**, *120*, 498.
- Pádua Alves, É.; de Alcântara, A. L. D.; Guimarães, A. J. K.; de Santana, E. H. W.; Botaro, B. G.; Fagnani, R.; *J. Sci. Food Agric.* **2018**, *98*, 3994.
- Ferreira, M. P. L. V. O. I.; Oliveira, M. B. P. P.; *J. Liq. Chromatogr. Relat. Technol.* **2003**, *26*, 99.
- Lenardon, L.; Meneghini, L. Z.; Hoff, R. B.; Motta, T. M.; Pizzolato, T. M.; Ferrão, M. F.; Bergold, A. M.; *Anal. Lett.* **2017**, *50*, 2068.
- Bilge, G.; Sezer, B.; Eseller, K. E.; Berberoglu, H.; Topcu, A.; Boyaci, I. H.; *Food Chemistry*, **2016**, *212*, 183.
- Borin, A.; Ferrão, M. F.; Mello, C.; Maretto, D. A.; Poppi, R. J.; *Anal. Chim. Acta* **2006**, *579*, 25.
- Luna, A. S.; Pinho, J. S. A.; Machado, L. C. *Anal. Methods* **2016**, *8*, 7204.
- Davis, R.; Mauer, L. J.; *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*, Formatex: Spain, 2010.
- Ministério da Agricultura, Pecuária e Abastecimento (MAPA), *Instrução Normativa n° 69*, institui critério da qualidade do leite in natura, concentrado e em pó, reconstituídos, com base no método analítico oficial físico-químico denominado "índice de CMP", 2006.
- Helfer, G. A.; Bock, F.; Marder, L.; Furtado, J. C.; Costa, A. B.; Ferrão, M. F.; *Quim. Nova* **2015**, *38*, 575.
- Noorgard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B.; *Appl. Spectrosc.* **2000**, *54*, 413.
- Kennard, R. W.; Stone, L. A.; *Technometrics* **1969**, *11*, 137.
- Jensen, R. G.; *Handbook of milk composition*, Academic Press: San Diego, 1995.
- Walker, N. R.; Waters, R. S.; Duncan, M. A.; *J. Chem. Phys.* **2004**, *120*, 10037.
- Farrés, M.; Platikanov, S.; Tsakovski, S.; Tauler, R.; *J. Chemom.* **2015**, *29*, 528.
- Balabin, R. M.; Smirnov, S. V.; *Talanta* **2011**, *85*, 562.
- Domingo, E.; Tirelli, A. A.; Nunes, C. A.; Guerreiro, M. C.; Pinto, S. M.; *Food Res. Int.* **2014**, *60*, 131.
- Hornik, K.; Stinchcombe, M.; White, H.; *Neural Networks* **1989**, *2*, 359.
- Laporte, M.; Paquin, P.; *J. Agric. Food Chem.* **1999**, *47*, 2600.