Direct analysis of Volumetric Absorptive Micro Sampling (VAMS) devices by ATR-FT-MIR and chemometric analysis: a new challenge

Giulia Gorla[a], A. Pasamontes[b], L. Rampazzi[a], Jordi Riu[b], Barbara Giussani[a]

[a] Dipartimento di Scienza e Alta Tecnologia. Università degli Studi dell'Insubria. Via Valleggio, 9. 22100 Como, Italy

[b] Department of Analytical Chemistry and Organic Chemistry, Universitat Rovira i Virgili, C/ Marcel·lí Domingo s/n, 43007 Tarragona, Spain

corresponding author:
Barbara Giussani
Dipartimento di Scienza e Alta Tecnologia
Università degli Studi dell'Insubria
via Valleggio, 9 - 22100 Como – Italy
tel: +39 031 234 6434
email: barbara.giussani@uninsubria.it

HIGHLIGHTS:

- Volumetric Absorptive Micro Sampling (VAMS) is a promising strategy in liquid food analysis because of its ease of use and handling
- Direct analysis of VAMS devices using ATR-FT-MIR spectroscopy without the need of an extraction process
- PLS regression models of the VAMS spectra can be used to estimate major components in milk samples

ABSTRACT:

Volumetric Absorptive Micro Sampling (VAMS) strategy, in its simplicity, has made a major contribution to the development of at-home sampling strategies. Mainly used for blood analysis, it absorbs a fixed volume of sample. Folded into its cover, the VAMS device dries, and it can be sent to a lab via mail.

In this article, for the first time in our knowledge, we explored the possibility to use this sampling strategy to expand the scope of VAMS to other samples than clinical ones. In this way we used VAMS to sample and analyze milk, which is one of the most important and analyzed samples all over the world. VAMS devices were employed to sample commercial milk samples from Italy, Switzerland and Spain, and for the first time the device was directly analyzed by ATR-FT-IR to predict protein, carbohydrate and fat content in the milk samples. Samples were collected in different sessions from different persons and analyzed by different lab operators to include in the models these sources of variability. Multivariate regression was used to correlate ATR-FT-IR spectra with the investigated properties: models were validated with external validation.

KEYWORDS:

VAMS, Volumetric Absorptive Micro Sampling, ATR-FT-MIR, milk, milk major constituents, multivariate regression

INTRODUCTION

The development of modern microsampling technologies was initially born from the need to overcome the limits of biological fluids sampling. Sampling of biological fluids is, in fact, a main concern of modern bioanalytical chemistry. The conventional sampling techniques are often highly invasive, and samples have to be collected in dedicated laboratories, which means that often

patients need to travel frequently, especially for diseases that need to be recurrently monitored. Modern microsampling techniques enable at-home sampling and less invasive sampling strategies.

Different strategies were developed and carried out using Volumetric Absorptive Microsampling (VAMS) on blood samples and other biological matrices for application with bioanalytical purposes [1,2]. Studies were reported for plasma, urine and saliva [3,4].

VAMS holds several interesting advantages in comparison with any other analytical sampling system. It allows to collect small and accurate volumes (10, 20 or 30 μl) of the samples under study by using a porous hydrophilic tip that dries out in 1-2 hours, as reported by Protti et al. in [5] and by Hecht in [6], and they can be stored at room temperature [7,8]. VAMS microsamplers are lightweight and easily transportable with the included clamshells (2–4 samplers) or cartridges (2 samplers), therefore allowing for samples to be stored along time (allowing for a timescale of different days) and for logistic costs to be reduced [9].

In the applications found in the literature [10], all of them in the field of biological samples, an extraction phase follows the sampling of the fluid by the microsampler: the tip of the device is immersed in a suitable solvent to extract the analyte under investigation. After this extraction, a subsequent analysis of the extracted solution is carried out with a tandem chromatography-mass spectrometry system. In each case the optimal extraction conditions must be determined as well as the recovery factor.

In this study, we propose to expand the fields of application of VAMS using for the first time to the best of our knowledge the VAMS sampling system in cow milk analysis. We have chosen this application because cow milk analysis is an issue of worldwide interest and of main concern in dairy industry, and because milk is a complex matrix and thus an ideal sample to test new application of the microsampler [11]. Moreover, we propose the direct analysis of VAMS devices (after milk sampling) using ATR-FT-MIR without the need of a further extraction process as a way to significantly reduce the time of the analysis and also to reduce the use of chemicals involved in the process.

To test the procedure and its applicability, major milk components such as fats, proteins and carbohydrates were quantified in milk samples.

The overall method is thought for an in-situ sampling, and an easy storage and shipment of the sampling device to a laboratory, which would be able then to provide results analyzing the microsampler without any pretreatment and with a small necessity of reagents.

MATERIALS AND METHODS

Sampling devices and procedure

Mitra VAMS devices (volume capacity of 10 µL) were purchased from Neoteryx, LLC. (Torrance, CA, USA). For the sampling of milk, the tips of the VAMS devices were held in beakers containing 50 mL of milk at room temperature for 30 sec. After that, the devices were deposited in their clamshells. They were left to dry for 48 hours at ambient temperature (21°C controlled but not monitored) to simulate real conditions of use and then they were directly analyzed with the ATR-FTIR spectrometer. A scheme of the sampling protocol is presented in Figure 1. Each milk was sampled in triplicate. Samples were collected and analyzed during different seasons along 12 months to construct models containing also the variability related to temperature and humidity in the drying phase. They were sampled by different persons. No chemical was used during this phase.

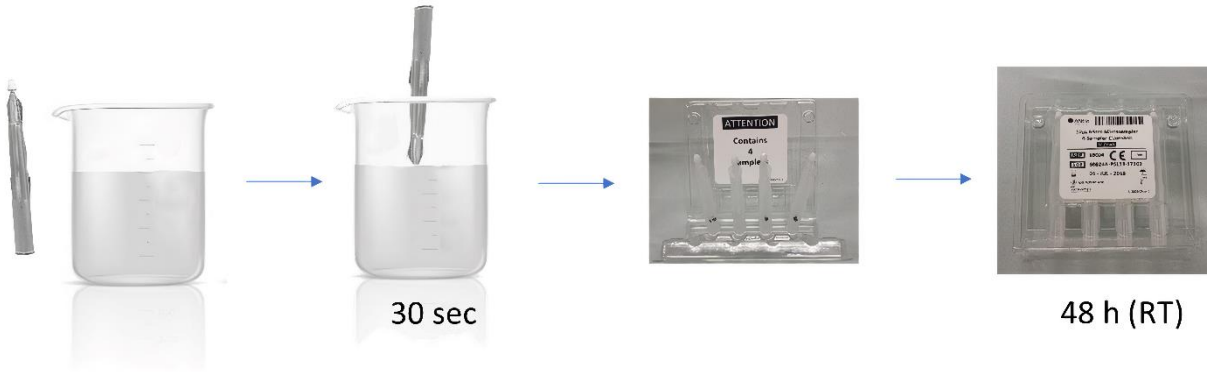


Figure 1 – Sampling protocol. Each VAMS sampler was immersed for 30 seconds in 50 mL of milk, being careful to not go over the edge of the tip. Then, it was stored in a clamshell box provided by the manufacturer and left to dry for 48 h before the analysis.

Milk samples

43 commercial cow milks were acquired in supermarkets in Spain, Italy and Switzerland (see supplementary material for the composition). Ranges for the major components are reported in

Table 1 (data taken from the commercial labels). The samples were selected to obtain a representative population of the commercial samples, including the maximum variation range of protein, fat and carbohydrate contents. It is worthwhile to note that, according to the European regulation [12] these ranges are quite restricted. All samples were cow milk in liquid form, most of them UHT (Ultra-High-Temperature processing) treated and, to include more variability and enlarge the dataset, some infant formula milk samples were included in the analysis.

Observing the nutrient values, we noticed that the range of high proteins and low carbohydrates were not well represented in the commercial samples. This is probably due to the restriction of the European Legislation on milk production. To increase the number of samples in those regions, we constructed five other milk samples (identified in the models with the label "M") by mixing two commercial milks.

Macronutrients	Range (g/100mL)
Proteins	1.3 – 7.2
Fats	0.1 - 3.9
Carbohydrates	2.5 - 9.3

Table 1 - Range of variation of macronutrients in the milk samples under study

Instrumentation: ATR-FT-MIR spectroscopy

Spectra were collected using an ATR-FT-MIR instrument (Nicolet™ iS10 FTIR Spectrometer, Thermo Scientific™ equipped with the Smart iTR™ accessory with a diamond crystal). The VAMS sticks were used as a guideline to lay and align the tips on the diamond crystal. The needle hubs were maintained in position using the pressure tower of the Smart iTR™ Accessory.

Spectra were recorded in absorbance mode from 600 to 4000 cm^{-1} with 32 scans and at 4 cm^{-1} resolution (7053 data points) in order to obtain well detailed spectra. A background with the crystal empty (air background) was acquired before each measurement. Ethanol was used after each measurement to remove any residue from the equipment. It was the only chemical used throughout the study and only for cleaning. Spectra used to construct the multivariate models were obtained by different lab operators.

Chemometric tools and data processing

Preprocessing and data elaboration were conducted using the open source software environment R (3.6.3) with the IDE RStudio (1.2.5001) [13]. A list of the packages used is available in the supplementary information.

As previously said, each milk sample was sampled in triplicate. The average spectrum of the three experimental replicates was calculated and used for the multivariate models: thus, each milk sample was represented by one average spectrum in all the computed models.

After some preliminary investigations (data not reported), the wavenumbers range was restricted between 700 and 3700 cm^{-1} to avoid high noise regions with low information (spectra were described with 6223 data points). To further reduce the dataset dimensionality and speed up the computation, one wavelength over five was considered obtaining a final data matrix with 1245 columns (wavelengths), without significant loss of information.

Savitzky-Golay filter [14] with a window size of eleven points and a second-order polynomial degree gave the best results as data pretreatment method.

The spectra were then organized in an **X** matrix with samples in the rows and variables in the columns. PLS regression models were calculated to establish the relationship between the spectra and the major components, organized in the **y** vectors. Subsequently, some filter-based [15] variable selection methods were used to find subsets of wavenumbers that allowed for better prediction capabilities with lower computation weight. Loadings weights (LW), regression coefficients (RC), variable importance in projection (VIP) and a combination of them were tested as variable selection methods.

Samples were divided in calibration (36 samples) and validation sets (12 samples) using two selection algorithms: CADEX [16, 17], also known as Kennard Stone algorithm, and HONIGS [18,19,20]. These algorithms aimed at selecting a chosen number of samples for the calibration set to uniformly cover the multidimensional space of **X**.

The variable selection was carried out on the calibration set samples to improve the models, which models were validated using samples not used in the variable selection steps (the validation set).

The model performances were evaluated in terms of Root-Mean-Square-Error (RMSE) in venetian blinds (CV) cross-validation (after the variable selection step), Root-Mean-Square-Error of prediction (RMSEP – final models) and goodness of fit (r^2) of the predicted versus the measured values.

RESULTS & DISCUSSION

Method optimization: from sampling to spectra

Milk, as a well-known sample, was chosen to evaluate the possibility of the use of VAMS in food analysis coupled with direct spectroscopy. All the spectra acquisitions were performed about 48 hours after the sampling procedure in order to obtain information simulating a possible shipment.

Several optimization tests were conducted to investigate the sampling variability. Beakers with different capacities (50, 100, 200 mL) were used as sample containers and no significant variation in the spectra were detected making possible to infer the independence from the initial volume as expected.

Once the VAMS samplers had dried, the position of the tip on the ATR crystal was evaluated. Holding the tip vertically was found to compromise the tip by breaking it. The horizontal positioning, shown in Figure 2, was found more convenient and reproducible.

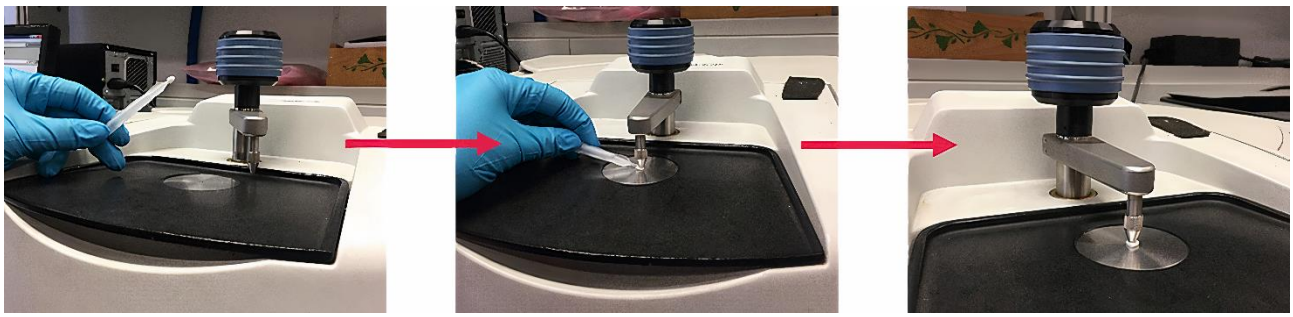


Figure 2 – Photos taken during the tip horizontal positioning on the ATR-FTIR instrument. The VAMS sticks were used as a guideline to place the tips in position on the diamond crystal. The pressure tower of the Smart iTR™ accessory were used to maintain the needle hubs in position during the spectra acquisition.

Some studies were conducted to gain information about the possibility of making spectral replicates from the same tip. Results showed that after one measurement the deformation of each tip significantly hindered the possibility of rotating it to acquire further spectra.

After this first optimization step, sample spectra were acquired and three blank tips also were analyzed in order to investigate the influence of the VAMS tip in the spectra.

A driving force in the use of this sampling system is that the water signal does not emerge in the spectra because water has been completely removed from the sample. This represents an interesting aspect for the analysis, because dealing with the strong signal of the O-H stretching usually obtained for this kind of samples can be difficult, as repeatedly reported in literature [20]. Figure 3 shows the signal of the porous hydrophilic tip with and without milk.

The IR absorbance bands mainly due to the tip of the microsampler are those of the CH_2 asymmetric stretching around 2860 cm^{-1}, the aldehyde CH stretching at 2800 cm^{-1}, the O-H bending at 1420 cm^{-1}, the S=O stretching at 1410 cm^{-1} and the fingerprint region around 680 and 670 cm^{-1}. The main features concerning the milk contents are the C=O stretching at 1690 cm^{-1}, the peak from the interaction of the C=O stretching and the NH bending around 1600 cm^{-1}. Other important absorbance bands are the broad band of the O–H stretching around 3300 cm^{-1}. At 1490 the O-H, C-O-H and C-C-H deformation due to carbohydrates are partially covered by the tip spectra. Between 993 cm^{-1} and 1110 cm^{-1} there are some overlapped peaks due to C–O, C–C, C–H stretching and C–O–C ether stretching [21–23]. It is worthwhile to note that the quality of spectra in the 1800-2700 cm^{-1} range is noticeably influenced by a strong diamond absorption by the diamond ATR crystal, but in this range neither the microsampler tip nor the milk show their more intense absorption bands.

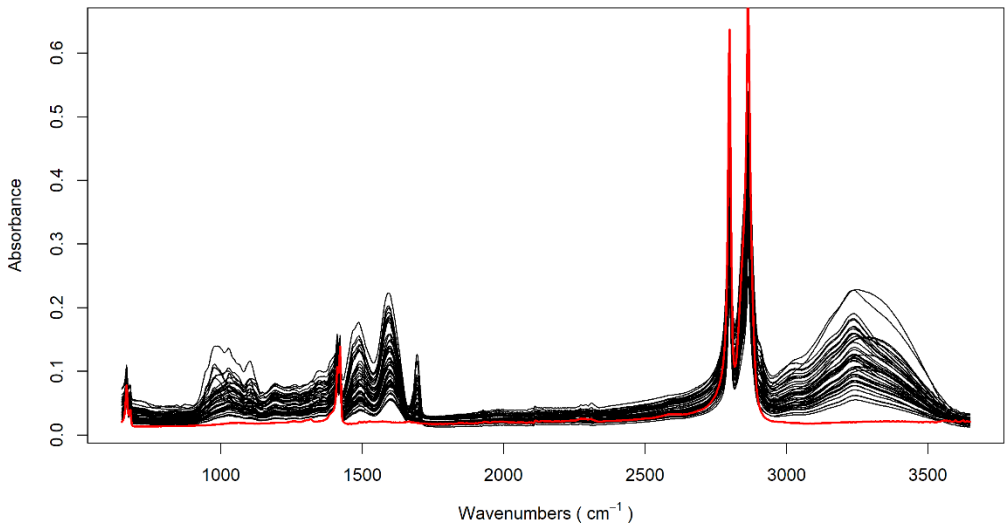


Figure 3 – Spectra registered for milk samples dried over the sampling system (black) and spectra of the blank VAMS (red).

Prediction of major nutritional components

Calibrations for each one of the main components of milk samples were developed after a preliminary investigation related to the exclusion of grossly outlying samples (bad measurements or similar) and the best preprocessing method. Only two samples, labelled as L31 and L34, were identified as outliers mainly because of spectral intensity anomalies. Another sample (L24) presented ~~or~~ peculiar peak~~s~~ shoulders. We tried to perform the average of only two replicates and to use these samples in the model but, unfortunately, they showed anomalous residual values. We tried to use them in the validation phase, but again they showed high residual values, confirming the need of three replicates to describe the milk samples. A similar conclusion is obtained constructing the entire models using only one replicate with even worse results (Table S2 in Supporting Material).

It is worthwhile to note that considering the high number of Mitra VAMS analyzed (147 devices) the fact that only 2 sampling procedures went wrong (less than 1.5%) can be considered as a symptom of the easiness and user-friendliness of the technique proposed. In addition, no further evident outliers were identified during the calculation of the regression models and this low number of outliers reinforce the possibility of using VAMS as a sampling method for this type of samples.

No tendencies were observed according to the sampling time, thus spectra collected over different seasons were safely included in the models.

As data preprocessing methods Standard Normal Variate, Multiplicative Scatter Correction and Savitsky Golay filter with different window widths (7-9-11-13-15 data points) and derivative order (first and second order) were considered. The best preprocessing method was found as the second Savitsky Golay filter with a window width of 11 data points and second polynomial order. This data pretreatment allowed the emphasis of the signal variations useful for the predictions of the macronutrients in milk samples, and led to a reduction of the signal scattering. The application of scattering correction methods did not improve the modeling performances.

Three blank VAMS were also measured, and their spectra were averaged to obtain a blank VAMS spectrum. The subtraction of the blank VAMS spectrum to the sampled VAMS spectra was also considered. The model capabilities did not improve, so this possible additional preprocessing step was discarded. Because of the high intensity of the blank VAMS spectra in the region over 2500 cm^{-}

[1], models were calculated with and without this spectral region. No significant improvement of the models was observed, and thus models were optimized starting from the whole VAMS spectra.

Data were mean centered before the PLS modeling. The split of the data into the calibration and validation sets with two algorithms led, as expected, to different sample selection depending on the algorithm, with most of the samples chosen for the calibration sets in common.

Several PLS models were performed to predict the fat, protein and carbohydrate content of the samples, combining data pretreatments and variable selection algorithms as above mentioned. The best results are presented in Table 2. They are the best compromise between the model complexity (number of factors required), the r^2 and the prediction error.

The plots of the predicted versus measured data for the best models are depicted in Figure 4.

Major component	Variable Selection Method	Number of variables selected	LVs	r^2_{cv}	RMSECV	Validation Algorithm	$r^2_{prediction}$	RMSEP
Fats	LW	87	6	0.838	0.49	CADEX	0.972	0.27
	LW	89	6	0.833	0.50	HONIGS	0.965	0.32
Proteins	RC	306	5	0.788	0.68	CADEX	0.806	0.51
	RC	125	5	0.793	0.68	HONIGS	0.788	0.45
Carbohydrates	RC	117	4	0.797	0.71	CADEX	0.806	0.41
	RC	81	3	0.802	0.73	HONIGS	0.748	0.36

Table 2- Best prediction models obtained for the macronutrients quantification in milk samples. LW = loading weights, RC = regression coefficients.

Concerning the fat prediction, the most promising models are the ones obtained when the loadings weights filter was applied to select 87 or 89 variables. The models showed a bias equal to -0.0217 and -0.0483 in the cross-validation phase respectively. This low bias is indicative ~~for~~ of the absence of systematic errors. The selected variables are distributed along the whole spectrum, and they belong to both milk and VAMS spectroscopic signals.

The best model to predict proteins was found when applying the regression coefficient filter as the variable selection method, obtaining an RMSEP of 0.45 and a bias in the cross-validation phase of -0.0447 which is indicative of the absence of systematic errors. The selected variables come from peaks around 1100 cm^{-1} and mainly from the VAMS signals.

The best model for carbohydrates prediction is the one obtained with the use of 81 variables selected with the regression coefficient filters. 3 factors were needed to obtain an RMSEP of 0.36 and an r^2 of 0.802. Even in this prediction, the variables selected belonged to both milk and VAMS spectroscopic signals. The bias calculated was -0.069 in the cross-validation phase. Also in this case, this is indicative of the absence of systematic errors.

The obtained error values are very satisfactory considering that the reference values for the samples under study, according to the European Regulation 1169/2011/EC, include a tolerance of 2 g/100 g for proteins and carbohydrates and 1.5 g/100 g for fat [24]. It is worthwhile to recall that the contents of macronutrients declared by the producers were used as the reference values, and this influenced the quantification errors.

Under these premises, it is worthy to note that two algorithms for the selection of the calibration and the validation sets were used, obtaining results completely comparable as showed in table 2, proving the robustness of the models presented.

Moreover, the models include the variability between the different sampling conditions over the seasons, as samples were sampled in different sampling campaigns. The obtained error of prediction can be considered a robust estimation.

The obtained results for the fat content are even better to those obtained in a previous work [25] on a similar dataset (commercial milk samples) using the direct application of a drop of milk on the same ATR-FT-MIR instrument. Concerning protein and carbohydrate contents, the models appear to be slightly better when using the direct drop analysis with the ATR-FT-MIR instruments. Iñón et al. [26] reported very similar results on commercial milk samples.

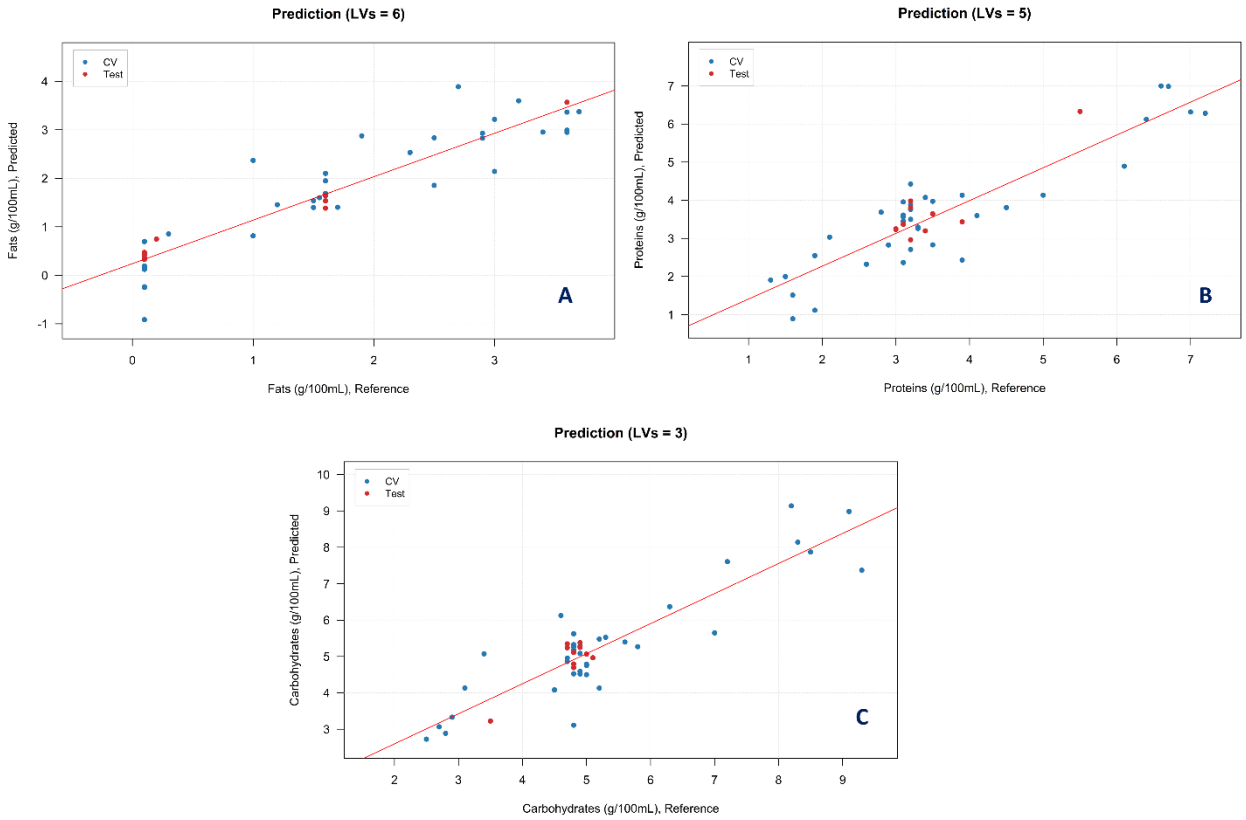


Figure 4 - Predicted vs measured values for the fats (A), proteins (B) and carbohydrates(C) contents.

Stability

Investigations about variability in spectra depending on the day of analysis were carried out.

Three milk samples were chosen from the commercial milk population. Considering that the variability of available commercial milk is mainly due to fat levels (high or whole, medium or semi-skimmed and low or skimmed according to the EU regulation [12]), the three samples were selected to be one for each fats category with protein and carbohydrate content around the mean value for commercial samples.

Each milk sample was sampled with 15 VAMS devices simultaneously. The spectroscopic analysis was carried out then at 2h (D1), 24h (D2), 48h (D3), 72h (D4) and 96h (D5) from the time of sampling, according to the following scheme: for each milk sample, three sampled VAMS were randomly selected from the 15 devices and analyzed after 2 hours of sampling. Then, other three VAMS sampled devices were randomly selected and analyzed after 24 hours from the time of sampling, up to 96h after sampling (see Figure S5 - Supplementary Material). Replicates were averaged, obtaining

one mean spectrum for each milk and sample time. The spectra were then pretreated as required by the models and the fats, proteins and carbohydrates values were predicted to assess if some error in prediction was dependent on time.

The mean error in prediction obtained for fats was 0.34, that for the proteins was 0.32 and that for the carbohydrates 0.23. For all the compounds the range of the errors in prediction was found in agreement with that of the respective model and no trends were identified among the five different measures over time (Figure 5).

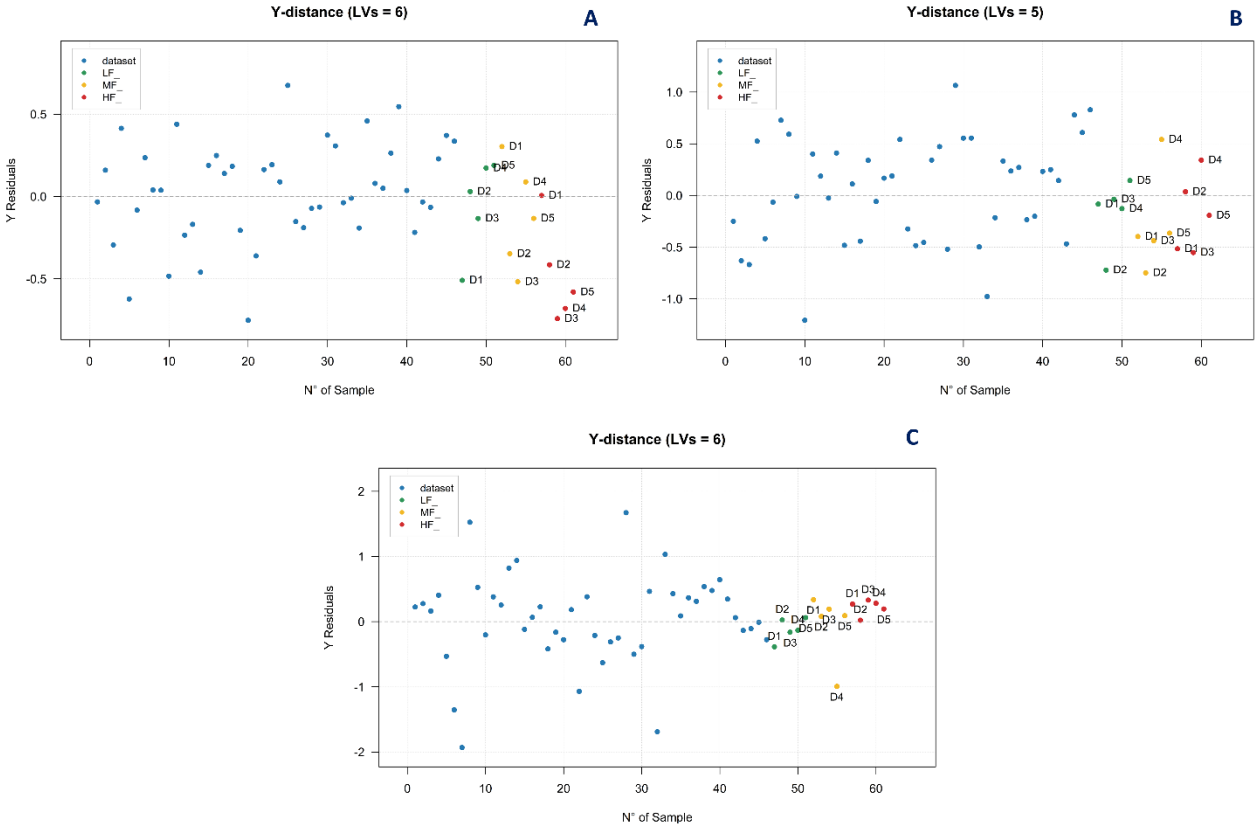


Figure 5 – y residuals plots for the fats (A), proteins (B), and carbohydrates (C) models. “dataset” represents the residuals on y of the samples used to construct and validate the models. “LF_” ,“MF_” and “HF_” are the residuals values when the y value of the samples used for the stability studies were predicted. They stand for low fat, medium fat and high fat, respectively.

From the data obtained it was demonstrated that, even if the sampler was delivered late to the laboratory, it can be correctly analyzed by the models. This is checked for a delivery time of up to 96 hours.

CONCLUSIONS

For the first time, major components in commercial milk samples were determined by means of a microsampling strategy. This strategy allows the direct analysis of the microsampler devices by ATR-FT-MIR without the need of further extraction processes, and therefore without the need of any chemical in the whole process of analysis. Multivariate regression was used to find the concentrations of major components in commercial cow milk samples. Results were very satisfactory and comparable with other portable technologies, as mentioned in the manuscript.

This application paves the way for a new logic of portable milk analysis. When it is not possible to reach the sampling point with an instrument of measure, or it is complicated to send the liquid sample to a laboratory, a microsampler device may be the solution. A very small quantity of sample is required, and after sampling VAMS proved to be stable for several days, making the application truly usable in real contexts.

Moreover, this research firstly suggests new possible applications of VAMS devices out from the clinical field, expanding therefore its scope of application. They are now mainly employed in biological and biomedical applications, but they can represent a valid alternative for sampling, sample storage and transportation of different kind of liquid materials, in many different fields of application. Secondly, analytes of interest have always been extracted from VAMS devices in previous works with solvents, making the whole procedure less green than it could be. In this article, we demonstrate that spectroscopic techniques and multivariate analysis prove to be powerful tools for the direct analysis of the microsampling devices, avoiding thus solvent use, manipulation, and wastes.

REFERENCES

[1] P. Denniff, S. Parry, W. Dopson, N. Spooner, Quantitative bioanalysis of paracetamol in rats using volumetric absorptive microsampling (VAMS)tVolumetric absorptive microsampling, J. Pharm. Biomed. Anal. 108 (2015) 61–69. https://doi.org/10.1016/j.jpba.2015.01.052.

[2] X. Gao, C. Chen, D. Geng, K.P. Bateman, S. Shi, E.J. Woolf, Y. Xu, Volumetric absorptive microsampling

(VAMS®) in therapeutic protein quantification by LC-MS/MS: Investigation of anticoagulant impact on assay performance and recommendations for best practices in method development, J. Pharm. Biomed. Anal. 196 (2021) 113895. https://doi.org/10.1016/j.jpba.2021.113895.

[3] L. Mercolini, M. Protti, M.C. Catapano, J. Rudge, A.E. Sberna, LC-MS/MS and volumetric absorptive microsampling for quantitative bioanalysis of cathinone analogues in dried urine, plasma and oral fluid samples, J. Pharm. Biomed. Anal. 123 (2016) 186–194. https://doi.org/10.1016/j.jpba.2016.02.015.

[4] Y. Qu, K. Brady, R. Apilado, T. O'Malley, S. Reddy, P. Chitkara, C. Ibarra, R.V. Alexander, T. Dervieux, Capillary blood collected on volumetric absorptive microsampling (VAMS) device for monitoring hydroxychloroquine in rheumatoid arthritis patients, J. Pharm. Biomed. Anal. 140 (2017) 334–341. https://doi.org/10.1016/j.jpba.2017.03.047.

[5] M. Protti, M.C. Catapano, B.G. Samolsky Dekel, J. Rudge, G. Gerra, L. Somaini, R. Mandrioli, L. Mercolini, Determination of oxycodone and its major metabolites in haematic and urinary matrices: Comparison of traditional and miniaturised sampling approaches, J. Pharm. Biomed. Anal. 152 (2018) 204–214. https://doi.org/10.1016/j.jpba.2018.01.043.

[6] M. Hecht, H. Evard, K. Takkis, R. Veigure, R. Aro, R. Lohmus, K. Herodes, I. Leito, K. Kipper, Sponge sprayreaching new dimensions of direct sampling and analysis by MS, Anal. Chem. 89 (2017) 11592–11597. https://doi.org/10.1021/acs.analchem.7b02957.

[7] P. Denni, N. Spooner, Volumetric Absorptive Microsampling: A Dried Sample Collection Technique for Quantitative Bioanalysis, 86 (2014) 8489–8495. https://doi.org/10.1021/ac5022562.

[8] Y. Harahap, R. Diptasaadya, D.J. Purwanto, Volumetric absorptive microsampling as a sampling alternative in clinical trials and therapeutic drug monitoring during the covid-19 pandemic: A review, Drug Des. Devel. Ther. 14 (2020) 5757–5771. https://doi.org/10.2147/DDDT.S278892.

[9] M. Protti, R. Mandrioli, L. Mercolini, Tutorial: Volumetric absorptive microsampling (VAMS), Anal. Chim. Acta. 1046 (2019) 32–47. https://doi.org/10.1016/j.aca.2018.09.004.

[10] M.G.M. Kok, M. Fillet, Volumetric absorptive microsampling: Current advances and applications, J. Pharm. Biomed. Anal. 147 (2018) 288–296. https://doi.org/10.1016/j.jpba.2017.07.029.

[11] Robert G. Jensen, ed., Handbook of Milk Composition, 1st Editio, Elsevier - Academic Press, 1995.

[12] Communities., The Commission of the European, Council Regulation (EC) No 2597/97 of 18 December 1997 Laying Down Additional Rules on the Common Organization of the Market in Milk and Milk Products for Drinking Milk. Available online: https://eur-lex.europ, (1997).

[13] RStudio Team, RStudio: Integrated Development for R., (2018). http://www.rstudio.com/.

[14] A.S. and M.J.E.G. The, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639. https://doi.org/10.1109/ICASSP.2000.859059.

[15] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, Chemom. Intell. Lab. Syst. 118 (2012) 62–69. https://doi.org/10.1016/j.chemolab.2012.07.010.

[16] R.W.K. and L.A. Stone, Computer Aided Design of Experiments, Technometrics. 11 (1969) 137–148. https://doi.org/10.1007/978-1-349-95810-8_109.

[17] D.E. Honigs, G.M. Hieftje, H.L. Mark, T.B. Hirschfeld, Unique-Sample Selection via Near- Infrared Spectral Subtraction, (1985) 2299–2303. https://doi.org/10.1021/ac00289a029.

[18] S.M. and C. de la P. M. Blanco, J. Coello, H. Iturriaga, Near-infrared spectroscopy in the pharmaceutical industry, Analyst. 123 (1998) 135R-150R.

[19] A.S. and L. Ramirez–Lopez, An Introduction to the prospectr package, R Packag. Vignette. (2014). https://cran.r-project.org/package=prospectr (active at 23/03/2021).

[20] H. Mohamed, P. Nagy, J. Agbaba, A. Kamal-Eldin, Use of near and mid infra-red spectroscopy for analysis of protein, fat, lactose and total solids in raw cow and camel milk, Food Chem. 334 (2021) 127436. https://doi.org/10.1016/j.foodchem.2020.127436.

[21] B. Balan, A.S. Dhaulaniya, R. Jamwal, Amit, K.K. Sodhi, S. Kelly, A. Cannavan, D.K. Singh, Application of Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR) spectroscopy coupled with chemometrics for detection and quantification of formalin in cow milk, Vib. Spectrosc. 107 (2020) 103033. https://doi.org/10.1016/j.vibspec.2020.103033.

[22] M.P. Ye, R. Zhou, Y.R. Shi, H.C. Chen, Y. Du, Effects of heating on the secondary structure of proteins in milk powders using mid-infrared spectroscopy, J. Dairy Sci. 100 (2017) 89–95. https://doi.org/10.3168/jds.2016-11443.

[23] X. Wang, C. Esquerre, G. Downey, L. Henihan, D. O'Callaghan, C. O'Donnell, Assessment of infant formula quality and composition using Vis-NIR, MIR and Raman process analytical technologies, Talanta. 183 (2018) 320–328. https://doi.org/10.1016/j.talanta.2018.02.080.

[24] E. Commission, I. Disclaimer, December 2012 GUIDANCE DOCUMENT FOR COMPETENT AUTHORITIES FOR THE CONTROL OF COMPLIANCE WITH EU LEGISLATION ON : Regulation (EU) No 1169 / 2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to cons, Eur. Comm. 1 (2012) 1–15. https://www.fsai.ie/uploadedFiles/guidance_tolerances_december_2012.pdf.

[25] G. Gorla, M. Mestres, R. Boqué, J. Riu, D. Spanu, B. Giussani, ATR-MIR spectroscopy to predict commercial milk major components: A comparison between a handheld and a benchtop instrument, Chemom. Intell. Lab. Syst. 200 (2020) 103995. https://doi.org/10.1016/j.chemolab.2020.103995.

[26] F.A. Iñón, S. Garrigues, M. De La Guardia, Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques, Anal. Chim. Acta. 513 (2004) 401–412. https://doi.org/10.1016/j.aca.2004.03.014.

[27] Ron Wehrens, Chemometrics With R: Multivariate Data Analysis in the Natural Sciences and Life Sciences, Springer, 2011. https://doi.org/10.1007/978-3-642-17841-2.

[28] S. Kucheryavskiy, mdatools – R package for chemometrics, Chemom. Intell. Lab. Syst. 198 (2020) 103937. https://doi.org/10.1016/j.chemolab.2020.103937.

ADDITIONAL INFORMATION

S1) Table S1. Milk samples major components composition expressed in g/100 mL. Infant formula is indicated by * in the table.

Label	Fats (g/100mL)	Carbohydrates (g/100mL)	Proteins (g/100mL)
L1	1.6	4.8	3.1
L2	3.6	4.7	3
L3	0.3	5.6	3.9
L4	1.7	4.8	3.1

L5	1.6	4.9	3
L6	3.6	4.7	3.9
L7	3.6	4.8	3.3
L8	0.1	5	3.5
L9	1.6	5.1	3.9
L10*	2.9	5.3	2.6
L11	2.3	5.2	3.1
L12	0.1	4.7	5
L13*	3.4	7.2	1.9
L14	1.6	5.8	4.1
L15*	3	7	2.9
L16	1.5	4.7	3.2
L17*	2.5	9.3	2.1
L18	1.55	4.9	3.4
L19	1.5	4.6	2.8
L20	0.2	4.8	3.3
L21*	3.2	8.2	1.3
L22*	3	9.1	1.6
L23*	1.9	6.3	1.5
L24	0.1	2.5	7
L25	0.1	4.5	3.5
L26*	2.5	8.3	1.6
L27	1.2	4.8	3.2
L28	1	4.9	3.1
L29	1.6	5.2	3.4
L30	2.7	4.9	3.2
L31	3.9	4.9	3.2
L32	3.7	5	3.5
L33*	2.9	8.5	1.9
L34	3.7	5.1	3.4
L35	1	4.9	3.1
L36	1.6	5	3.2
L37	0.1	3.4	4.5
L38	1.6	4.8	3.2
L39	0.1	5	3.2
L40	0.1	4.9	3.1
L41	3.6	4.8	3.2
L42	1.6	4.8	3.2
L43	0.1	4.8	7.2
M1	0.1	3.5	5.5
M2	0.1	3.1	6.1
M3	0.1	2.9	6.4
M4	0.1	2.8	6.6
M5	0.1	2.7	6.7

S2) Table S2. Best results obtained for the prediction model of the macronutrients in milk samples using one replicate instead of the mean of three. LW = loading weights, RC = regression coefficients.

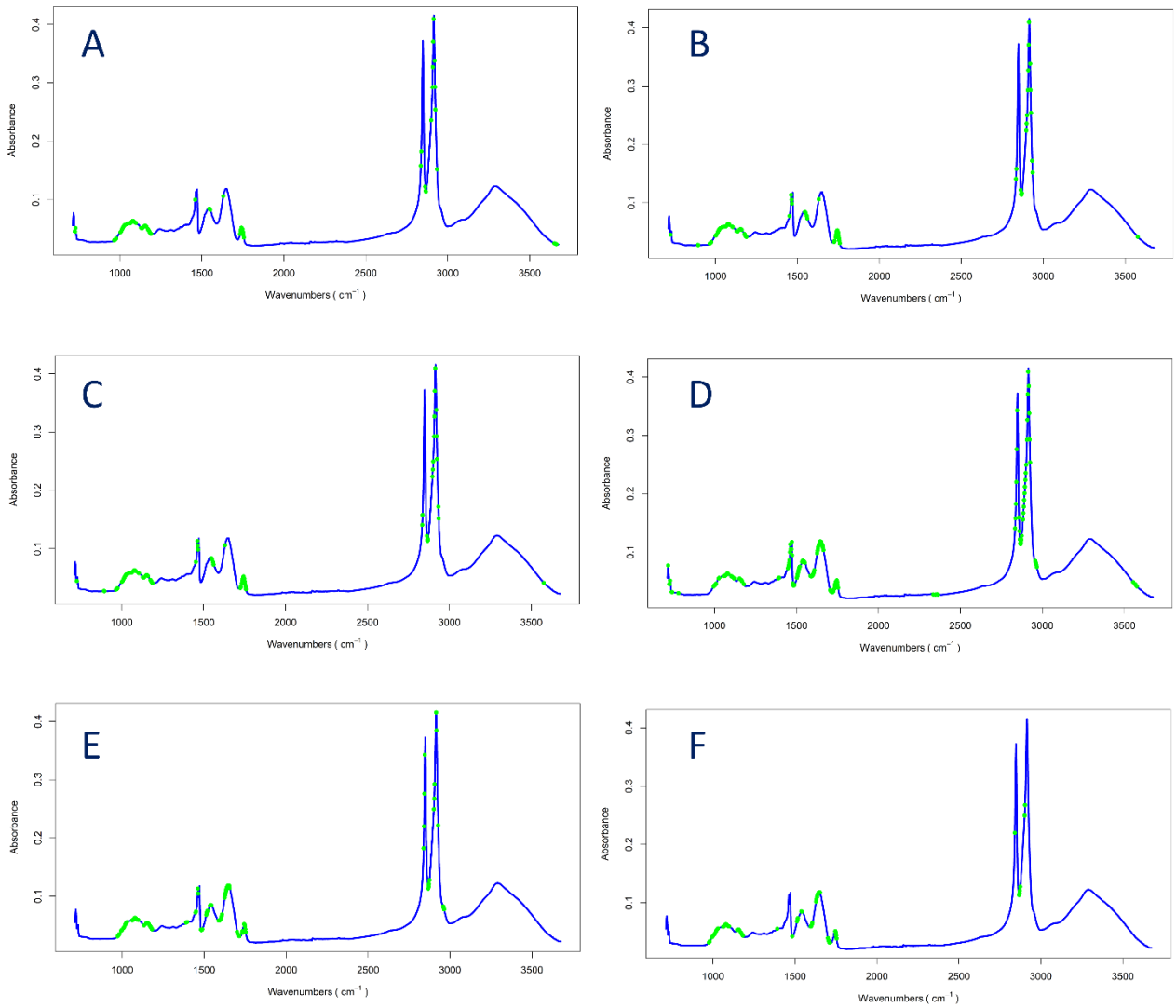
Major component	Variable Selection Method	Number of variables selected	LVs	r^2_{cv}	RMSECV	Algorithm	$r^2_{prediction}$	RMSEP
Fats	LW	91	4	0.808	0.56	CADEX	0.667	0.68
	LW	118	6	0.809	0.51	HONIGS	0.814	0.67
Proteins	LW	480	3	0.637	0.90	CADEX	0.668	0.67
	RC	154	3	0.615	0.93	HONIGS	0.692	0.61
Carbohydrates	RC	115	4	0.823	0.67	CADEX	0.947	0.39
	RC	357	4	0.850	0.63	HONIGS	0.704	0.42

S3) List of R packages used for data preprocessing and analysis

Package	Version	Main use	References
ChemoSpec	5.3.2	Manipulation and data pretreatments	https://CRAN.R-project.org/package=ChemoSpec (available on 26.04.2021)
ChemoSpecUtil	0.4.51	Manipulation and data pretreatments	https://CRAN.R-project.org/package=ChemoSpecUtils (available on 26.04.2021)
signal	0.7-6	Data preprocessing	https://CRAN.R-project.org/package=signal (available on 26.04.2021)
Chemometrics	1.4.2	Exploratory analysis and prediction models	https://CRAN.R-project.org/package=chemometrics (available on 26.04.2021)
ChemometricswithR	0.1.13	Exploratory analysis and prediction models	[27]
pls	2.7-3	Exploratory analysis and	[27]

		prediction models	
mdatools	0.11.2	Predictive models	[28]
prospectr	0.2.1	Samples split into calibration and validation sets	https://CRAN.R-project.org/package=prospectr (available on 26.04.2021)
plsVarSel	0.9.6	Variable selection	[15]

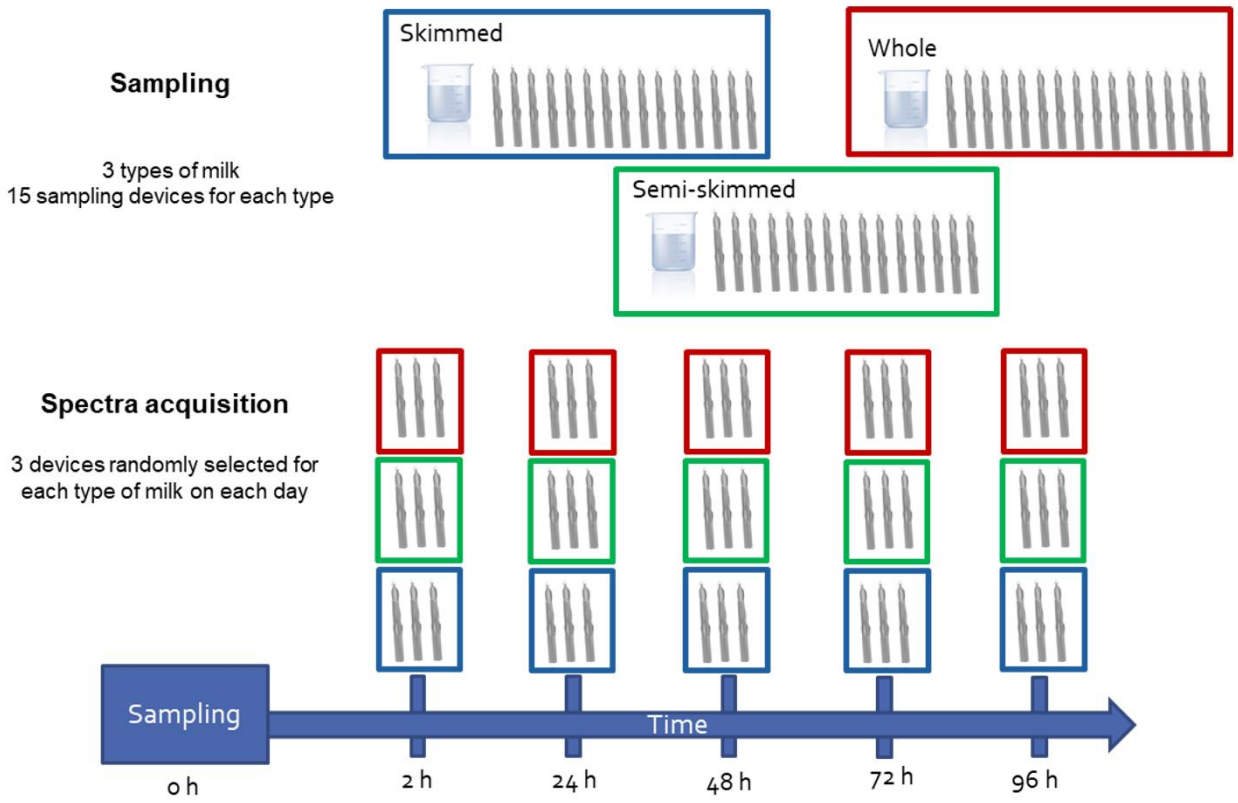
S4) Figure representing the selected variables for the best models described in the text and reported as Table 2. A) Variables selected in the model for fats prediction when using KS algorithm to select the calibration set. B) Variables selected in the model for fats prediction when using HO algorithm to select the calibration set. C) Variables selected in the model for proteins prediction when using KS algorithm to select the calibration set. D) Variables selected in the model for proteins prediction when using HO algorithm to select the calibration set. E) Variables selected in the model for carbohydrates prediction when using KS algorithm to select the calibration set. F) Variables selected in the model for carbohydrates prediction when using HO algorithm to select the calibration set.



457
458
459
460
461

462 S4) Figure representing the stability experimental sequence

463


Sampling
3 types of milk
15 sampling devices for each type
Skimmed
Whole
Semi-skimmed
Spectra acquisition
3 devices randomly selected for each type of milk on each day
Sampling
Time
0 h
2 h
24 h
48 h
72 h
96 h