J. Dairy Sci. 106:5288–5297
https://doi.org/10.3168/jds.2023-23226

Effectiveness of mid-infrared spectroscopy for the prediction of cow milk metabolites

M. Franzoi,[1]* G. Niero,[1]*† G. Meoni,[2,3] L. Tenori,[2,3] C. Luchinat,[2,3] M. Penasa,[1] M. Cassandro,[1,4] and M. De Marchi[1]

[1]Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy
[2]Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff," University of Florence, 50019 Sesto Fiorentino, Italy
[3]Consorzio Interuniversitario Risonanze Magnetiche Metallo Proteine (CIRMMP), 50019 Sesto Fiorentino, Italy
[4]Italian Holstein, Brown Swiss and Jersey Association (ANAFIBJ), Via Bergamo 292, 26100 Cremona, Italy

ABSTRACT

Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy is acknowledged as one of the most powerful analytical methods with cross-cutting applications in dairy foods. To date, the use of ^{1}H NMR spectroscopy for the collection of milk metabolic profile is hindered by costly and time-consuming sample preparation and analysis. The present study aimed at evaluating the accuracy of mid-infrared spectroscopy (MIRS) as a rapid method for the prediction of cow milk metabolites determined through ^{1}H NMR spectroscopy. Bulk milk (n = 72) and individual milk samples (n = 482) were analyzed through one-dimensional ^{1}H NMR spectroscopy and MIRS. Nuclear magnetic resonance spectroscopy identified 35 milk metabolites, which were quantified in terms of relative abundance, and MIRS prediction models were developed on the same 35 milk metabolites, using partial least squares regression analysis. The best MIRS prediction models were developed for galactose-1-phosphate, glycerophosphocholine, orotate, choline, galactose, lecithin, glutamate, and lactose, with coefficient of determination in external validation from 0.58 to 0.85, and ratio of performance to deviation in external validation from 1.50 to 2.64. The remaining 27 metabolites were poorly predicted. This study represents a first attempt to predict milk metabolome. Further research is needed to specifically address whether developed prediction models may find practical application in the dairy sector, with particular regard to the screening of dairy cows' metabolic status, the quality control of dairy foods, and the identification of processed milk or incorrectly stored milk.

Key words: dairy cattle, metabolome, predictive ability, chemometrics

INTRODUCTION

Proton nuclear magnetic resonance (^{1}**H NMR**) spectroscopy is widely adopted to study the composition of the metabolome in biological matrices and tissues. In the field of human medicine, ^{1}H NMR spectroscopy has been supportive in the identification of biomarkers able to discriminate between healthy and diseased tissues or organs, and consequently in the development of diagnostic clinical tests (Emwas et al., 2013). In food science, ^{1}H NMR spectroscopy is a high-throughput analytical method for the analysis of the metabolome, in both liquid and solid matrices. High-resolution ^{1}H NMR has been effectively used to obtain spectrum-based and metabolite-based food fingerprints, with important applications in quality and safety controls (Hatzakis, 2019; Dourou et al., 2020), as well as in traceability and authenticity validation purposes (Monakhova et al., 2013; Meoni et al., 2020).

As regards the dairy sector, ^{1}H NMR spectroscopy has several applications with various purposes along the entire production chain. At dairy farm level, milk ^{1}H NMR spectra and their metabolic profiles have been effectively used to discriminate between healthy udder quarters and udder quarters with subclinical mastitis (Bobbo et al., 2022), and between cows with subclinical and clinical mastitis (Luangwilai et al., 2021). The in-depth analysis of milk with high and low SCC is a promising approach in the livestock sector to elucidate the biochemical basis of mastitis (Sundekilde et al., 2013). At dairy industry level, ^{1}H NMR spectroscopy has been proposed as an anticounterfeiting analytical method to detect traces of bovine milk in caprine milk (Rysova et al., 2021), to discriminate between cow milk produced according to indoor farming and outdoor grazing conditions (Niero et al., 2022), to distinguish

Received January 3, 2023.
Accepted February 13, 2023.
*These authors contributed equally to this work.
†Corresponding author: g.niero@unipd.it

milk produced according to different feeding systems (Tenori et al., 2018), and to distinguish cheeses manufactured in compliance with different disciplinaries (Segato et al., 2019). Nowadays, the collection of milk metabolomic fingerprint and the application of related protocols for safety or traceability purposes are hampered by high costs and time of analysis of ^{1}H NMR spectroscopy.

Mid-infrared spectroscopy (**MIRS**) is one of the most rapid and cost-effective technologies to collect phenotypes on a large scale (De Marchi et al., 2014). Milk spectra obtained from MIRS analysis can be electronically stored as records of energy absorbed by the sample hit by an electromagnetic radiation in the range between 900 and 5,000 cm^{-1}. The same MIRS spectra can be processed through different chemometrics approaches to predict phenotypes at no additional costs. Currently, in national milk recording systems, milk samples from individual cows are periodically collected and analyzed through MIRS to determine fat, protein, casein, lactose, and urea. In recent years, much research has been carried out to develop MIRS prediction models for novel phenotypes, including detailed mineral (Soyeurt et al., 2009), fatty acid (Gottardo et al., 2017), and protein profile (Niero et al., 2021). The possibility to predict fine milk traits such as vitamins (Revilla et al., 2017), antioxidants (Niero et al., 2019), and lactoferrin (Soyeurt et al., 2020) represents the last frontier of this research field.

To authors' knowledge, no previous research has investigated the possibility to predict milk metabolites through MIRS. Therefore, the aim of the present study was to assess the accuracy of MIRS to predict cow milk metabolites determined through ^{1}H NMR spectroscopy. In the light of the accuracy of the developed models and based on available scientific literature, discussion on the best predicted metabolites have been appended to provide some potential implication for the dairy sector.

MATERIALS AND METHODS

Samples Collection

A total of 554 raw cow milk samples (72 bulk and 482 individual milks) were collected in dairy herds located in Northern Italy, between June 2019 and June 2021. Bulk milks were collected from conventional and organic multibreed herds, with Holstein being the most represented breed. Individual milks were collected from Simmental cows farmed according to indoor barn or outdoor pasture conditions. Detailed description of farm management, animal feeding, and sampling procedures can be retrieved from Niero et al. (2022). Immediately after sampling, 200 μL of Azidiol preservative (chloramphenicol 1.5 g/L, anhydrous trisodium citrate 34 g/L, sodium azide 36 g/L, bromophenol blue 0.35 g/L, ethanol 1%) were added to 100 mL of milk. Each sample was divided into 2 aliquots for subsequent analyses.

Mid-infrared Spectra and Chemical Composition

One of the milk aliquots was transported at 4°C to the laboratory of the Breeders Association of Veneto Region (Vicenza, Italy). Within 24 h, milk samples were warmed, gently mixed by inversion to promote solid homogenization, and analyzed for gross composition (fat, protein, casein, and lactose, %), and urea (mg/dL) using MilkoScan 7DC (Foss) according to ISO 21543:2020 and ICAR guidelines (ICAR, 2018; ISO, 2020). Somatic cell count was determined using Fossmatic (Foss) according to ISO 13366–2:2006 and ICAR guidelines (ISO, 2006; ICAR, 2018). Mid-infrared spectra of all samples were collected and stored electronically.

Proton Nuclear Magnetic Resonance

The second aliquot of milk was transported at 4°C to the Magnetic Resonance Center (Sesto Fiorentino, Italy) for ^{1}H NMR analysis. Within 24 h, samples were dissolved in dichloromethane (CH_2Cl_2) at a ratio of 1:1 (vol/vol; Tenori et al., 2018). The mixture was homogenized by vortexing and then incubated at room temperature for 10 min. After centrifuging the mixture at 5,000 × *g* for 30 min at 4°C, 350 μL of the supernatant were added to 350 μL of sodium phosphate buffer [70 m*M* Na_2HPO_4; 20% (vol/vol) H_2O, 6.1 m*M* NaN_3; 4.6 m*M* sodium trimethylsilyl (2,2,3,3-H_4)-propionate; pH 7.4]. Then, 600 μL of this mixture were transferred to a 5-mm ^{1}H NMR tube (Bruker BioSpin) and stored at 80°C for the ^{1}H NMR analysis. Proton nuclear magnetic resonance spectra of milk extracts were obtained from a Bruker spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5-mm PATXI ^{1}H-13C-15N probe with a z-axis gradient coil, automatic tuning-matching, and an automatic refrigerated sample changer. For temperature equilibration, samples were maintained inside the ^{1}H NMR probe for at least 5 min prior measurement (310K). For each sample, 3 one-dimensional ^{1}H NMR spectra were acquired with water peak suppression and different pulse sequences that allowed the selective observation of different molecular components: (1) a standard NOESY 1Dpresat (noesygppr1d.comp; Bruker BioSpin) pulse sequence (using 64 scans, 98,304 data points, a spectral width of 18,028 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mix-

ing time of 0.01 s). This pulse sequence is designed to obtain a spectrum in which both signals of metabolites and high molecular weight molecules (lipids and lipoproteins) are visible; (2) a standard CPMG (cpmgpr1d.comp; Bruker BioSpin) pulse sequence (using 64 scans, 73,728 data points, a spectral width of 12,019 Hz, and a relaxation delay of 4 s). This pulse sequence is designed for the selective observation of small molecules in solutions containing macromolecules; (3) a standard diffusion-edited (ledbgppr2s1d.comp; Bruker BioSpin) pulse sequence (using 64 scans, 98,304 data points, a spectral width of 18,028 Hz, and a relaxation delay of 4 s) for the selective observation of macromolecular components in solutions containing small molecules. Free induction decays were amplified by an exponential function comparable to a 0.3 Hz line-broadening factor before applying the Fourier transform. TopSpin 3.6.2 (Bruker BioSpin) was used to automatically adjust for phase and baseline aberrations and to calibrate transformed NOESY and CPMG spectra to the α-lactose doublet (5.24 δ^1H ppm). Each one-dimensional spectra in the range of 0.02 to 10.00 ^{1}H ppm was segmented into 0.02 δ^1H ppm chemical shift bins for multivariate analysis (buckets). The portions of NOESY and CPMG spectra containing water (4.61–4.77 ^{1}H ppm) and dichloromethane (5.30–5.33 and 5.42–5.65 ^{1}H ppm) were eliminated. Normalization using X was applied to the bins before any statistical analysis.

A total of 35 metabolites were identified in the ^{1}H NMR spectra. Signal identification was performed using a library of ^{1}H NMR spectra of pure organic compounds (Assure NMR 2.2 software, Bruker BioSpin), public databases (FooDB, https://foodb.ca/, and Milk Composition Database, http://www.mcdb.ca/) storing references, and literature data (Tenori et al., 2018; Meoni et al., 2020; Bobbo et al., 2022).

Partial Least Squares Regression Models

Samples were randomly split in calibration and validation sets, in a 5:1 ratio, stratified by sample type (bulk and individual), to maintain the same representativeness of the 2 different type of samples. Descriptive statistics of metabolites in the whole data set and in calibration and validation sets are reported in Supplemental Table S1 (https://figshare.com/articles/journal_contribution/Franzoi_M_et_al_2023_-_Supplementary_Table_S1_-_JDS_R2/23093252; Niero, 2023). Prediction models were built on the calibration set using modified partial least squares regression analysis (WinISI III v. 1.60; Foss and Infrasoft International LLC) through 10-fold cross-validation. Several combinations both in terms of scattering correction (no correction, detrend, standard normal variate, standard normal variate + detrend, standard multiplicative scatter correction) and mathematical treatment (0,0,1,1; 1,4,4,1; 1,8,8,1; 2,5,5,1; 2,10,10,1) were tested. For the mathematical treatment, the 4 digits indicate the number of the derivative, gap used for derivative calculation, data points in the first smoothing, and data points in the second smoothing, respectively. The number of latent variables included in the model were selected according to van der Voet (1994).

Before each regression, spectra were evaluated for global Mahalanobis distance (**GH**) and those with GH >3 were excluded. Hereafter, potential outliers were removed using the T-outlier test (Soyeurt et al., 2012) in Winisi software (Foss), setting to 3 the critical value. Modified partial least squares regression and outlier determination were iterated 3 times. The best prediction equation for each trait was chosen based on the standard error of prediction in cross-validation. Standard error of prediction was calculated also for calibration and validation. Relative standard error, coefficient of determination, and ratio performance to deviation were calculated for calibration ($\mathbf{RSE_{Cal}}$, $\mathbf{R^2_{Cal}}$, and $\mathbf{RPD_{Cal}}$, respectively), cross-validation ($\mathbf{RSE_{CVal}}$, $\mathbf{R^2_{CVal}}$, and $\mathbf{RPD_{CVal}}$), and external validation ($\mathbf{RSE_{EVal}}$, $\mathbf{R^2_{EVal}}$, and $\mathbf{RPD_{EVal}}$).

RESULTS AND DISCUSSION

Milk Composition, Proton Nuclear Magnetic Spectra, and Mid-Infrared Spectra

Descriptive statistics of milk composition, urea, and SCC are summarized in Table 1. Fat, protein, casein, and lactose averaged 3.73, 3.47, 2.75, and 4.84%, respectively. Fat had the greatest variability with coefficient of variation (**CV**) of 21.59%. Milk urea and SCC averaged 21.05 mg/dL and 190.09 cells/μL, and had CV of 20.95 and 125.84%, respectively. Such a great variability was somewhat expected and is likely due to the experimental design of the present study, which indeed included (1) individual cow and bulk milk, (2) different cow breeds, (3) different farming systems (i.e., organic and conventional), (4) different farming conditions (i.e., indoor and outdoor) and related feeding strategies (i.e., silages and forages). Broadly speaking, the possibility to capture great variability is favorable for the development of MIRS prediction models. Indeed, great variability is associated with great spectra variability, which in turn, is an effective way to improve the accuracy of MIRS prediction models (Soyeurt et al., 2011; De Marchi et al., 2014).

Average CMPG ^{1}H NMR spectrum, average mid-infrared spectrum, and hetero-spectral synchronous 2D correlation spectrum are represented in Figure 1.

Table 1. Descriptive statistics of milk composition, urea, and SCC

Trait	n	Mean	CV (%)	Minimum	Maximum
Milk composition					
Fat (%)	548	3.73	21.59	1.22	6.98
Protein (%)	550	3.47	9.43	2.42	4.44
Casein (%)	554	2.75	10.29	1.86	3.69
Lactose (%)	546	4.84	3.35	4.34	5.30
Urea (mg/dL)	553	21.05	20.95	9.30	33.20
SCC (cells/μL)	541	190.09	125.84	6.00	1,731.00

To the best of our knowledge, this is the first time a hetero-spectral synchronous 2D correlation spectrum was produced between ^{1}H NMR and MIRS milk spectra. Most of the 2D cross-signals appeared in specific regions of the mid-infrared spectrum. A complex pattern of both positive and negative cross-signals appears between 1,000 and 1,100 cm^{-1}, commonly assigned to lactose C–O, C–C, and C–H stretching vibrations (Grelet et al., 2015). Such region showed the strongest positive cross-signals with ^{1}H NMR peaks typical of *N*-acetyl carbohydrates (2.07 ppm), methanol (3.37 ppm), lactose (4.47 ppm), and galactose (5.28 ppm). Correlation between lactose NMR peak and infrared lactose region was expected. At the same time, methanol and galactose derive from bacterial fermentation of lactose that could explain the observed correlation. Moreover, galactose and *N*-acetyl carbohydrates share several functional groups with lactose, contributing to samples absorbance between 1,000 and 1,100 cm^{-1}. The same infrared region highlighted a strong negative cross-peak with glycerophosphocholine (3.23 ppm), whereas the same metabolite showed an intense positive cross-signal with the near-infrared spectral region around 1,157 cm^{-1}, typical of lactose C–O–C stretching (Grelet et al., 2015).

Lactose NMR *m.* $-CH_2$ peak also showed a positive cross-peak with infrared region around 1,400 cm^{-1}, typical of the bending of CH_n in fatty acids aliphatic chains (Grelet et al., 2015). Infrared regions associated with protein C–C and C–N stretching, around 1,550 cm^{-1}, are correlated with several ^{1}H NMR peaks such as *N*-acetyl carbohydrates, glycerophosphocholine, methanol, lactose, and galactose. Infrared regions associated with fat C=O stretching (1,743 cm^{-1}) and fatty acids CH_n stretching (between 2,800 and 3,000 cm^{-1}) showed in general an opposite trend of cross-peaks compared with both lactose and fat CH_n bending regions and similar to C–O–C lactose stretching, and weaker cross-peaks with metabolites such as 3-hydroxybutyrate (2.326 ppm), citrate (3.55 ppm), and sarcosine + dimethylamine (2.736 ppm).

Overall, hetero-spectral synchronous 2D correlation spectrum highlighted a complex relationship between mid-infrared spectrum and NMR-determined metabolites. Despite such complexity, the presence of specific mid-infrared regions associated with metabolites could be used for wavelengths selection in the development of mid-infrared predictions (Forouzangohar et al., 2013).

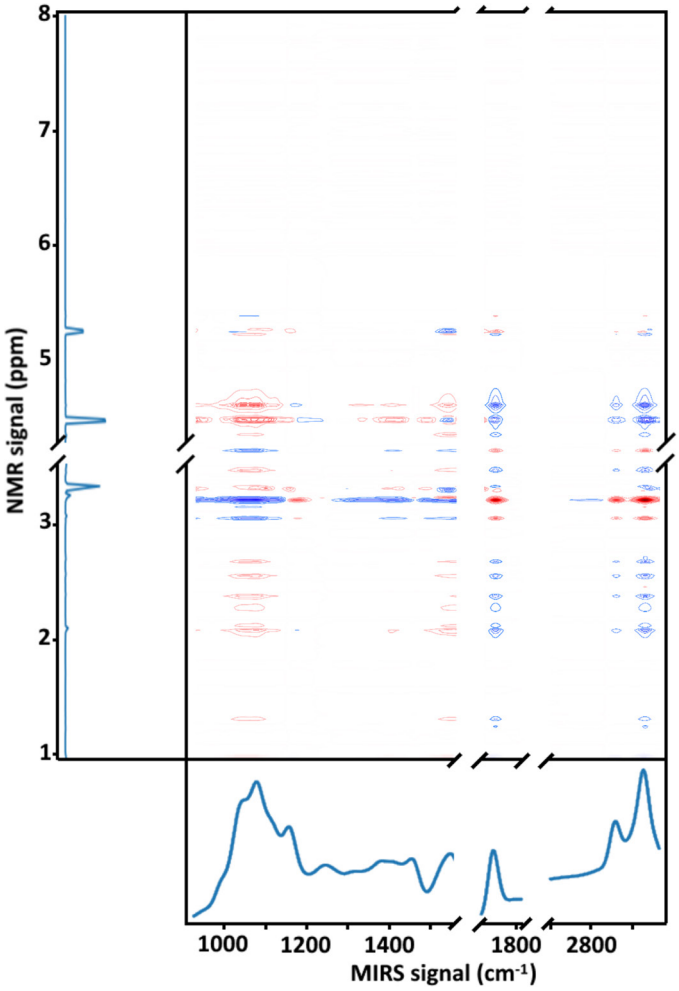

Figure 1. Two-dimension correlation between mid-infrared spectra (MIRS) and proton nuclear magnetic resonance (^{1}H NMR) spectra. Red and blue lines represent positive and negative cross-peaks, respectively.

Infrared Prediction Models

Among the 35 identified metabolites, 8 exhibited $RPD_{EVal} \geq 1.5$ and the remaining 27 were poorly predicted, with $RPD_{EVal} < 1.5$. Fitting statistics of the prediction models for cow milk metabolites with $RPD_{EVal} \geq 1.5$ are reported in Table 2, and scatter plots of the measured metabolites (x-axis) versus predicted metabolites (y-axis) are in Figure 2. Among the studied metabolites, MIRS prediction models for galactose-1-phosphate and glycerophosphocholine exhibited the best prediction performances, with R^2_{EVal} of 0.85 and 0.84, and RPD_{EVal} of 2.64 and 2.56, respectively. To the best of authors' knowledge, this is the first study that has attempted to predict galactose-1-phosphate and glycerophosphocholine concentration in milk and thus comparison with the literature was somewhat difficult. Broadly speaking and according to Grelet et al. (2021), R^2_{EVal} between 0.74 and 0.89, and RPD_{EVal} between 2 and 3 are sufficiently reliable and allow for rough screening of samples to estimate the quantity of the predicted phenotype.

Prediction models for orotate, choline, galactose, and lecithin concentration were slightly less accurate than those for galactose-1-phosphate and glycerophosphocholine, with R^2_{EVal} from 0.70 to 0.79, and RPD_{EVal} from 1.72 to 1.88. Zaalberg et al. (2020) assessed the effectiveness of MIRS for the prediction of milk orotate determined through ^{1}H NMR spectroscopy and reported R^2_{EVal} from 0.60 in milk of Danish Jersey cows to 0.79 in milk of Danish Holstein cows. Franzoi et al. (2018) developed MIRS prediction models for the prediction of galactose in different kinds of delactosated milk; the excellent accuracy of the models developed by Franzoi et al. (2018; $R^2_{CVal} = 0.98$, $RPD_{CVal} = 7.42$) is likely due to the great amount of galactose in delactosated milk, which is not the case of the present study.

Moderate R^2_{EVal} and RPD_{EVal} were obtained for glutamate (0.63 and 1.65, respectively) and lactose (0.58 and 1.50, respectively). In the present study glutamate was predicted with slightly greater accuracy compared with McDermott et al. (2016), who obtained R^2_{EVal} of 0.59 and RPD_{EVal} of 1.20. In the light of the fitting statistics obtained for lactose, it is likely that ^{1}H NMR spectroscopy is not the best reference method to be used for the development of MIRS prediction models. Indeed, a correlation of 0.996 in validation has been declared between lactose measured through HPLC as reference method and lactose predicted through MIRS (application note 5373 Rev. 3, MilkoScan 7RM/FT+/6000; Foss).

Fitting statistics of prediction models for cow milk metabolites with $RPD_{EVal} < 1.5$ are reported in Table 3. According to Grelet et al. (2021) the use of models with such RPD_{EVal} is uncertain. McDermott et al. (2016) ob-

Table 2. Fitting statistics[1] of prediction models for milk metabolites with ratio of performance to deviation in external validation ≥1.5

Trait (AU)[2]	Calibration				Cross-validation				External validation			
	SEP_{Cal}	RSE_{Cal}	R^2_{Cal}	RPD_{Cal}	SEP_{CVal}	RSE_{CVal}	R^2_{CVal}	RPD_{CVal}	SEP_{EVal}	RSE_{EVal}	R^2_{EVal}	RPD_{EVal}
Galactose-1-phosphate	89.03	0.17	0.88	2.88	98.80	0.18	0.85	2.59	97.16	0.18	0.85	2.64
Glycerophosphocholine	1,362.51	0.08	0.90	3.17	1,557.77	0.09	0.87	2.77	1,686.93	0.10	0.84	2.56
Orotate	100.23	0.14	0.82	2.38	110.58	0.16	0.78	2.15	126.54	0.18	0.79	1.88
Choline	1,211.48	0.23	0.80	2.24	1,358.27	0.26	0.75	1.99	1,541.74	0.29	0.72	1.76
Galactose	46.39	0.17	0.80	2.21	49.49	0.18	0.77	2.08	58.72	0.22	0.70	1.75
Lecithin	56.94	0.14	0.81	2.31	60.64	0.15	0.79	2.17	76.46	0.19	0.70	1.72
Glutamate	60.01	0.20	0.69	1.79	65.95	0.22	0.62	1.62	64.85	0.22	0.63	1.65
Lactose	7,705.43	0.02	0.73	1.93	7,811.57	0.02	0.72	1.90	9,891.96	0.03	0.58	1.50

[1]SEP_{Cal} = standard error of prediction in calibration; RSE_{Cal} = relative SE in calibration; R^2_{Cal} = coefficient of determination in calibration; RPD_{Cal} = ratio of performance to deviation in calibration; SEP_{CVal} = standard error of prediction in cross-validation; RSE_{CVal} = relative standard error in cross-validation; R^2_{CVal} = coefficient of determination in cross-validation; RPD_{CVal} = ratio of performance to deviation in cross-validation; SEP_{EVal} = standard error of prediction in external validation; RSE_{EVal} = relative standard error in external validation; R^2_{EVal} = coefficient of determination in external validation; RPD_{EVal} = ratio of performance to deviation in external validation.
[2]AU = arbitrary units.

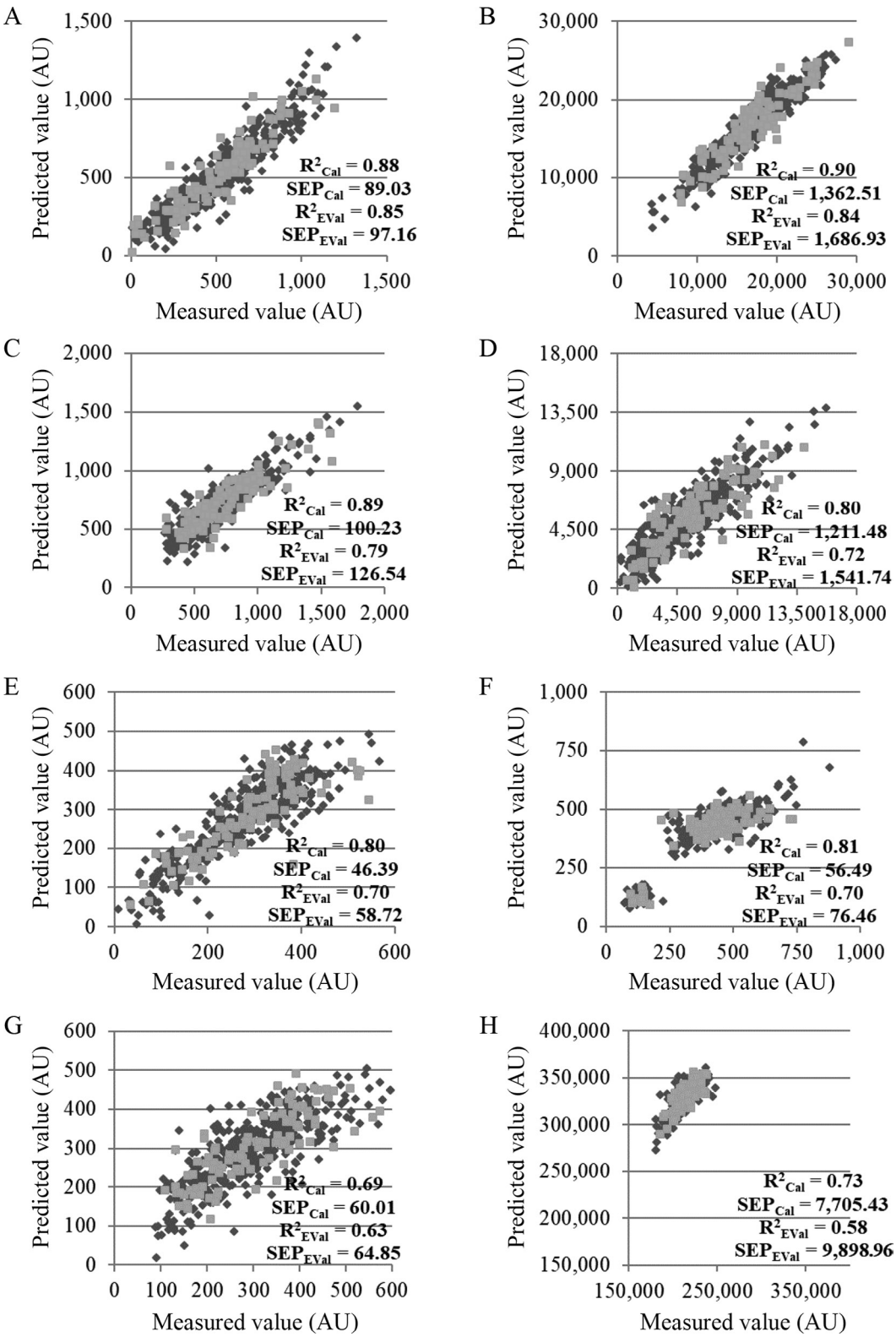

Figure 2. Scatter plots of predicted (y-axis) versus measured (x-axis) (A) galactose-1-phosphate, (B) glycerophosphocholine, (C) orotate, (D) choline, (E) galactose, (F) lecithin, (G) glutamate, and (H) lactose, in arbitrary units (AU). Dark and light gray represent calibration and validation data sets, respectively. Fitting statistics are the coefficient of determination in calibration (R^2_{Cal}), the standard error of prediction in calibration (SEP_{Cal}), the coefficient of determination in external validation (R^2_{Eval}), and the standard error of prediction in external validation (SEP_{EVal}).

Table 3. Fitting statistics[1] of prediction models for milk metabolites with ratio of performance to deviation in external validation <1.5

	Calibration				Cross-validation				External validation			
Trait (AU)[2]	SEP_{Cal}	RSE_{Cal}	R^2_{Cal}	RPD_{Cal}	SEP_{CVal}	RSE_{CVal}	R^2_{CVal}	RPD_{CVal}	SEP_{EVal}	RSE_{EVal}	R^2_{EVal}	RPD_{EVal}
N-Acetyl carbohydrates	1,041.35	0.08	0.74	1.95	1,139.58	0.09	0.68	1.78	1,387.24	0.11	0.60	1.46
Creatine	480.46	0.10	0.62	1.61	524.61	0.11	0.54	1.48	550.19	0.12	0.53	1.41
Glucose	14.09	0.19	0.61	1.60	15.04	0.20	0.55	1.50	17.38	0.23	0.56	1.30
Mannose	16.66	0.27	0.34	1.23	17.12	0.28	0.30	1.20	16.12	0.27	0.35	1.27
2-Oxoglutarate	28.87	0.16	0.45	1.34	30.40	0.16	0.38	1.28	31.15	0.17	0.39	1.24
Dimethyl sulfone	171.01	0.21	0.48	1.39	184.30	0.23	0.40	1.29	194.49	0.24	0.41	1.22
Fumarate	24.37	0.29	0.57	1.52	25.64	0.30	0.52	1.44	30.40	0.36	0.48	1.22
Sarcosine + dimethylamine	113.44	0.31	0.55	1.50	121.28	0.33	0.49	1.40	143.08	0.39	0.43	1.19
Valine	20.80	0.26	0.52	1.44	22.61	0.28	0.43	1.33	25.59	0.32	0.44	1.17
cis-Aconitate	76.20	0.28	0.51	1.43	80.06	0.29	0.46	1.36	94.47	0.34	0.34	1.15
Butyrate	16.08	0.53	0.65	1.68	17.84	0.58	0.56	1.52	23.66	0.78	0.60	1.14
Carnitine	4.21	0.53	0.07	1.04	4.24	0.53	0.05	1.03	4.18	0.53	0.02	1.04
Alanine	36.91	0.19	0.40	1.29	39.74	0.21	0.30	1.19	45.98	0.24	0.28	1.03
Cytidine	4.46	0.51	0.09	1.05	4.54	0.52	0.06	1.03	4.61	0.53	0.11	1.02
O-Acetylcarnitine	11.52	0.34	0.24	1.15	11.95	0.36	0.18	1.11	13.53	0.40	0.13	0.98
Riboflavin	7.90	0.57	0.27	1.17	7.98	0.58	0.25	1.16	9.64	0.70	0.16	0.96
Formate	10.71	0.39	0.01	1.01	10.73	0.39	0.01	1.00	11.54	0.42	0.00	0.93
Methanol	65.76	0.56	0.55	1.48	70.10	0.60	0.48	1.39	107.73	0.92	0.38	0.91
Isoleucine	10.67	0.58	0.11	1.06	10.95	0.59	0.06	1.03	12.61	0.68	0.02	0.90
N,*N*-Dimethylglycine	16.70	0.51	0.28	1.18	17.87	0.54	0.18	1.10	22.36	0.68	0.12	0.88
Citrate	1,669.08	0.37	0.24	1.15	1,700.55	0.38	0.21	1.13	2,660.21	0.59	0.10	0.72
Succinate	23.35	0.16	0.40	1.29	24.81	0.17	0.32	1.22	42.30	0.28	0.19	0.71
Ethanol	501.76	0.54	0.09	1.05	499.18	0.54	0.08	1.06	828.03	0.90	0.01	0.64
Acetate	109.01	0.38	0.51	1.42	120.01	0.42	0.40	1.29	274.50	0.97	0.41	0.57
Acetone	14.45	0.27	0.42	1.31	14.99	0.28	0.37	1.26	52.97	0.98	0.15	0.36
3-Hydroxybutyrate	7.01	0.52	0.42	1.31	7.26	0.53	0.37	1.26	34.55	2.54	0.04	0.27
Lactate	235.10	0.49	0.21	1.12	239.19	0.49	0.17	1.10	1,057.72	2.19	0.24	0.25

[1]SEP_{Cal} = standard error of prediction in calibration; RSE_{Cal} = relative standard error in calibration; R^2_{Cal} = coefficient of determination in calibration; RPD_{Cal} = ratio of performance to deviation in calibration; SEP_{CVal} = standard error of prediction in cross-validation; RSE_{CVal} = relative standard error in cross-validation; R^2_{CVal} = coefficient of determination in cross-validation; RPD_{CVal} = ratio of performance to deviation in cross-validation; SEP_{EVal} = standard error of prediction in external validation; RSE_{EVal} = relative standard error in external validation; R^2_{EVal} = coefficient of determination in external validation; RPD_{EVal} = ratio of performance to deviation in external validation.

[2]AU = arbitrary units.

tained slightly greater accuracy when predicting valine determined through HPLC reference method (R^2_{EVal} of 0.59 and RPD_{EVal} of 1.14). Grelet et al. (2016) reported considerably greater prediction accuracy in a study aimed to predict citrate, acetone, and 3-hydroxybutyric acid by the combination of data from different cow breeds farmed in different countries and according to different production systems.

Discussion on the Best Predicted Metabolites

Metabolites discussed in this chapter have been selected among those with $RPD_{EVal} \geq 1.5$ (Table 2) and, at the same time, documented technological, biological, or nutritional implications. Based on this approach, 5 metabolites have been considered, namely galactose-1-phosphate, glycerophosphocholine, choline, orotate, and galactose. Given this, it is worth reporting that the use of MIRS-predicted metabolites must be deepened through further specific research trials.

Galactose-1-Phosphate

Results by Lu et al. (2013) suggest that milk galactose-1-phosphate can be used as a reliable and specific indicator to determine the energy status of dairy cows. Those authors reported increased levels of galactose-1-phosphate in milk of cows experiencing negative energy balance, whereas very low amount (or even an absence) of this compound was observed in milk of cows with positive energy balance. On this background, accurate prediction of galactose-1-phosphate through MIRS may find application for the screening of metabolic status of dairy cows, with particular regard to the energy balance during the onset of lactation, and the enhancement of animal resistance to metabolic disorders.

Glycerophosphocholine and Choline

In ruminants, dietary choline is largely degraded by rumen bacteria (Neill et al., 1979); therefore, endogenous synthesis becomes a critical source of choline (Baldi and Pinotti, 2006). Still, the supplementation of rumen-protected choline has been demonstrated to reduce the extent of hepatic fatty infiltration and to increase the expression of genes involved in very low-density lipoprotein transport (Goselink et al., 2013). Such supplementation may be particularly important in the periparturient period, which is characterized by high incidence of fatty liver disease (Shahsavari et al., 2016). The possibility to monitor the content of these milk compounds may be of interest in the view of cow health and dietary balance.

Orotate

Organic acids in milk such as orotic acid are known to be important for the flavor of fermented dairy products (Guzel-Seydim et al., 2000). Therefore, the prediction of milk orotic acid before fermentation could be used to select milk with desired orotic acid concentration or even to select suitable starter cultures.

Galactose

Rapid and cost-effective quantification of milk galactose can find applications in identifying raw milk exposed to appropriate storage temperature. Ajmal et al. (2018), demonstrated lactose was partially degraded when milk was chilled 2 h after collection, compared with milk immediately refrigerated. This effect was likely linked to higher microbial activity in milk not stored at appropriate temperatures. Therefore, the rapid quantification of milk galactose may be of interest for the dairy industry to assess whether milk batches have been properly stored after collection.

CONCLUSIONS

The present study evaluated the robustness of MIRS to predict cow milk metabolites determined through ^{1}H NMR spectroscopy. A total of 35 milk metabolites were identified through ^{1}H NMR spectroscopy analysis. Among them, galactose-1-phosphate, glycerophosphocholine, orotate, choline, galactose, lecithin, glutamate, and lactose exhibited $RPD_{EVal} \geq 1.5$ and R^2_{EVal} from 0.58 to 0.85. Further research is advisable to understand whether such prediction models may find any application in the dairy sector, with particular regard to predicted galactose-1-phosphate (for the screening of metabolic status of dairy cows), predicted orotate (for quality control in dairy products), and predicted galactose (as milk quality reporter).

ACKNOWLEDGMENTS

This research was funded by Fitoche project (Veneto Region, DGR n. 376/2018), NIP project (Veneto Region, DGR n. 1139 of 19/07/2017), and Organic-PLUS project (European Union's Horizon 2020 research and innovation program, grant agreement No. 774340). The authors have not stated any conflicts of interest.

REFERENCES

Ajmal, M., M. Nadeem, M. Imran, M. Abid, M. Batool, I. T. Khan, N. Gulzar, and M. Tayyab. 2018. Impact of immediate and delayed chilling of raw milk on chemical changes in lipid fraction of pasteurized milk. Lipids Health Dis. 17:190. https://doi.org/10.1186/s12944-018-0843-0.

Baldi, A., and L. Pinotti. 2006. Choline metabolism in high-producing dairy cows: Metabolic and nutritional basis. Can. J. Anim. Sci. 86:207–212. https://doi.org/10.4141/A05-061.

Bobbo, T., G. Meoni, G. Niero, L. Tenori, C. Luchinat, M. Cassandro, and M. Penasa. 2022. Nuclear magnetic resonance spectroscopy to investigate the association between milk metabolites and udder quarter health status in dairy cows. J. Dairy Sci. 105:535–548. https://doi.org/10.3168/jds.2021-20906.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. *Invited review:* Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. https://doi.org/10.3168/jds.2013-6799.

Dourou, A. M., S. Brizzolara, G. Meoni, L. Tenori, F. Famiani, C. Luchinat, and P. Tonutti. 2020. The inner temperature of the olives (cv. Leccino) before processing affects the volatile profile and the composition of the oil. Food Res. Int. 129:108861. https://doi.org/10.1016/j.foodres.2019.108861.

Emwas, A. H. M., R. M. Salek, J. L. Griffin, and J. Merzaban. 2013. NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations. Metabolomics 9:1048–1072. https://doi.org/10.1007/s11306-013-0524-y.

Forouzangohar, M., D. Cozzolino, R. J. Smernik, J. A. Baldock, S. T. Forrester, D. J. Chittleborough, and R. S. Kookana. 2013. Using the power of C-13 NMR to interpret infrared spectra of soil organic matter: A two-dimensional correlation spectroscopy approach. Vib. Spectrosc. 66:76–82. https://doi.org/10.1016/j.vibspec.2013.02.002.

Franzoi, M., C. L. Manuelian, L. Rovigatti, E. Donati, and M. De Marchi. 2018. Development of Fourier-transformed mid-infrared spectroscopy prediction models for major constituents of fractions of delactosated, defatted milk obtained through ultra- and nanofiltration. J. Dairy Sci. 101:6835–6841. https://doi.org/10.3168/jds.2017-14343.

Goselink, R. M. A., J. Van Baal, H. C. A. Widjaja, R. A. Dekker, R. L. G. Zom, M. J. De Veth, and A. M. Van Vuuren. 2013. Effect of rumen-protected choline supplementation on liver and adipose gene expression during the transition period in dairy cattle. J. Dairy Sci. 96:1102–1116. https://doi.org/10.3168/jds.2012-5396.

Gottardo, P., M. Penasa, F. Righi, N. Lopez-Villalobos, M. Cassandro, and M. De Marchi. 2017. Fatty acid composition of milk from Holstein-Friesian, Brown Swiss, Simmental and Alpine Grey cows predicted by mid-infrared spectroscopy. Ital. J. Anim. Sci. 16:380–389. https://doi.org/10.1080/1828051X.2017.1298411.

Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci. 99:4816–4825. https://doi.org/10.3168/jds.2015-10477.

Grelet, C., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Stevens, N. Gengler, and F. Dehareng. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. Methods 186:97–111. https://doi.org/10.1016/j.ymeth.2020.07.012.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci. 98:2150–2160. https://doi.org/10.3168/jds.2014-8764.

Guzel-Seydim, Z., A. C. Seydim, and A. K. Greene. 2000. Organic acids and volatile flavor components evolved during refrigerated storage of kefir. J. Dairy Sci. 83:275–277. https://doi.org/10.3168/jds.S0022-0302(00)74874-0.

Hatzakis, E. 2019. Nuclear magnetic resonance (NMR) spectroscopy in food science: A comprehensive review. Compr. Rev. Food Sci. Food Saf. 18:189–220. https://doi.org/10.1111/1541-4337.12408.

ICAR. 2018. ICAR guidelines. Accessed Dec. 20, 2022. https://www.icar.org/index.php/icar-recording-guidelines/.

ISO. 2006. Milk—Enumeration of somatic cells—Part 2: Guidance on the operation of fluoro-opto-electronic counters. International Organization for Standardization. Accessed Dec. 20, 2022. https://www.iso.org/standard/40260.html.

ISO. 2020. Milk and milk products—Guidelines for the application of near infrared spectrometry. International Organization for Standardization. Accessed Dec. 20, 2022. https://www.iso.org/obp/ui/#iso:std:iso:21543:ed-2:v1:en.

Lu, J., E. Antunes Fernandes, A. E. Páez Cano, J. Vinitwatanakhun, S. Boeren, T. van Hooijdonk, A. van Knegsel, J. Vervoort, and K. A. Hettinga. 2013. Changes in milk proteome and metabolome associated with dry period length, energy balance, and lactation stage in postparturient dairy cows. J. Proteome Res. 12:3288–3296. https://doi.org/10.1021/pr4001306.

Luangwilai, M., K. Duangmal, N. Chantaprasarn, and S. Settachaimongkon. 2021. Comparative metabolite profiling of raw milk from subclinical and clinical mastitis cows using 1H-NMR combined with chemometric analysis. Int. J. Food Sci. Technol. 56:493–503. https://doi.org/10.1111/ijfs.14665.

McDermott, A., G. Visentin, M. De Marchi, D. P. Berry, M. A. Fenelon, P. M. O'Connor, O. A. Kenny, and S. McParland. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. J. Dairy Sci. 99:3171–3182. https://doi.org/10.3168/jds.2015-9747.

Meoni, G., L. Tenori, and C. Luchinat. 2020. Nuclear magnetic resonance-based metabolomic comparison of breast milk and organic and traditional formula milk brands for infants and toddlers. OMICS 24:424–436. https://doi.org/10.1089/omi.2019.0125.

Monakhova, Y. B., T. Kuballa, and D. W. Lachenmeier. 2013. Chemometric methods in NMR spectroscopic analysis of food products. J. Anal. Chem. 68:755–766. https://doi.org/10.1134/S1061934813090098.

Neill, A. R., D. W. Grime, A. M. Snoswell, A. J. Northrop, D. B. Lindsay, and R. M. C. Dawson. 1979. The low availability of dietary choline for the nutrition of the sheep. Biochem. J. 180:559–565. https://doi.org/10.1042/bj1800559.

Niero, G., M. Franzoi, C. L. Manuelian, G. Visentin, M. Penasa, and M. De Marchi. 2021. Protein profile of cow milk from multibreed herds and its relationship with milk coagulation properties. Ital. J. Anim. Sci. 20:2232–2242. https://doi.org/10.1080/1828051X.2021.1996288.

Niero, G. 2023. Franzoi M et al 2023 - Supplementary Table S1 - JDS R2. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.23093252.v1.

Niero, G., G. Meoni, L. Tenori, C. Luchinat, G. Visentin, S. Callegaro, E. Visentin, M. Cassandro, M. De Marchi, and M. Penasa. 2022. Grazing affects metabolic pattern of individual cow milk. J. Dairy Sci. 105:9702–9712. https://doi.org/10.3168/jds.2022-22072.

Niero, G., M. Penasa, A. Costa, S. Currò, G. Visentin, M. Cassandro, and M. De Marchi. 2019. Total antioxidant activity of bovine milk: Phenotypic variation and predictive ability of mid-infrared spectroscopy. Int. Dairy J. 89:105–110. https://doi.org/10.1016/j.idairyj.2018.08.014.

Revilla, I., O. Escuredo, M. I. González-Martín, and C. Palacios. 2017. Fatty acids and fat-soluble vitamins in ewe's milk predicted by near infrared reflectance spectroscopy. Determination of seasonality. Food Chem. 214:468–477. https://doi.org/10.1016/j.foodchem.2016.07.078.

Rysova, L., V. Legarova, Z. Pacakova, O. Hanus, I. Nemeckova, M. Klimesova, and J. Havlik. 2021. Detection of bovine milk adulteration in caprine milk with *N*-acetyl carbohydrate biomarkers by using ^{1}H nuclear magnetic resonance spectroscopy. J. Dairy Sci. 104:9583–9595. https://doi.org/10.3168/jds.2020-20077.

Segato, S., A. Caligiani, B. Contiero, G. Galaverna, V. Bisutti, and G. Cozzi. 2019. ^{1}H NMR metabolic profile to discriminate pasture based Alpine Asiago PDO cheeses. Animals (Basel) 9:722. https://doi.org/10.3390/ani9100722.

Shahsavari, A., M. J. D'Occhio, and R. Al Jassim. 2016. The role of rumen-protected choline in hepatic function and performance of transition dairy cows. Br. J. Nutr. 116:35–44. https://doi.org/10.1017/S0007114516001641.

Soyeurt, H., C. Bastin, F. G. Colinet, V. M. R. Arnould, D. P. Berry, E. Wall, F. Dehareng, H. N. Nguyen, P. Dardenne, J. Schefers, J. Vandenplas, K. Weigel, M. Coffey, L. Theron, J. Detilleux, E. Reding, N. Gengler, and S. McParland. 2012. Mid-infrared prediction of lactoferrin content in bovine milk: Potential indicator of mastitis. Animal 6:1830–1838. https://doi.org/10.1017/S1751731112000791.

Soyeurt, H., D. Bruwier, J. M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. J. Dairy Sci. 92:2444–2454. https://doi.org/10.3168/jds.2008-1734.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. https://doi.org/10.3168/jds.2010-3408.

Soyeurt, H., C. Grelet, S. McParland, M. Calmels, M. Coffey, A. Tedde, P. Delhez, F. Dehareng, and N. Gengler. 2020. A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. J. Dairy Sci. 103:11585–11596. https://doi.org/10.3168/jds.2020-18870.

Sundekilde, U. K., N. A. Poulsen, L. B. Larsen, and H. C. Bertram. 2013. Nuclear magnetic resonance metabolomics reveals strong association between milk metabolites and somatic cell count in bovine milk. J. Dairy Sci. 96:290–299. https://doi.org/10.3168/jds.2012-5819.

Tenori, L., C. Santucci, G. Meoni, V. Morrocchi, G. Matteucci, and C. Luchinat. 2018. NMR metabolomic fingerprinting distinguishes milk from different farms. Food Res. Int. 113:131–139. https://doi.org/10.1016/j.foodres.2018.06.066.

van der Voet, H. 1994. Comparing the predictive accuracy of models using a simple randomization test. Chemom. Intell. Lab. Syst. 25:313–323. https://doi.org/10.1016/0169-7439(94)85050-X.

Zaalberg, R. M., N. A. Poulsen, H. Bovenhuis, J. Sehested, L. B. Larsen, and A. J. Buitenhuis. 2021. Genetic analysis on infrared-predicted milk minerals for Danish dairy cattle. J. Dairy Sci. 104:8947–8958. https://doi.org/10.3168/jds.2020-19638.

ORCIDS

M. Franzoi https://orcid.org/0000-0001-8701-5632
G. Niero https://orcid.org/0000-0002-6169-1162
G. Meoni https://orcid.org/0000-0002-8608-4641
L. Tenori https://orcid.org/0000-0001-6438-059X
M. Penasa https://orcid.org/0000-0001-9984-8738
M. Cassandro https://orcid.org/0000-0002-8709-2870
M. De Marchi https://orcid.org/0000-0001-7814-2525