J. Dairy Sci. 106:9095–9104
https://doi.org/10.3168/jds.2023-23458

Imputation of missing milk Fourier transform mid-infrared spectra using existing milk spectral databases: A strategy to improve the reliability of breeding values and predictive models

H. Soyeurt,[1]* X.-L. Wu,[2,3] C. Grelet,[4] M. L. van Pelt,[5] N. Gengler,[1] F. Dehareng,[4] C. Bertozzi,[6] and J. Burchard[2]
[1]Research and Teaching Centre (TERRA), Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
[2]Council of Dairy Cattle Breeding, Bowie, MD 20716
[3]Department of Animal and Dairy Sciences, University of Wisconsin–Madison, Madison, WI 53706
[4]Walloon Agricultural Research Center, 5030 Gembloux, Belgium
[5]Cooperation CRV, Animal Evaluation Unit, PO Box 454, 6800 AL Arnhem, the Netherlands
[6]Walloon Breeders Association, 5590 Ciney, Belgium

ABSTRACT

The use of milk Fourier transform mid-infrared (FT-MIR) spectrometry to develop management and breeding tools for dairy farmers and industry is growing and supported by the availability of numerous new predicted phenotypes to assess the nutritional quality of milk and its technological properties, but also the animal health and welfare status and its environmental fingerprint. For genetic evaluations, having a long-term and representative spectral dairy herd improvement (DHI) database improves the reliabilities of estimated breeding values (EBV) from these phenotypes. Unfortunately, most of the time, the raw spectral data used to generate these estimations are not stored. Moreover, many reference measurements of those phenotypes, needed during the FT-MIR calibration step, are available from past research activities but lack spectra records. So, it is impossible to use them to improve the FT-MIR models. Consequently, there is a strong interest in imputing those missing spectra. The innovative objective of this study was to use the existing large spectral DHI database to estimate missing spectra by selecting probable spectra using, as the match criteria, common dairy traits recorded for a long time by DHI organizations. We tested 4 match criteria combinations. Combination 1 required to have equal fat and protein contents between the sample for which a spectrum was to be estimated and the reference samples in the DHI database. Combination 2 also required an equal urea content. Combination 3 requested equal fat, protein, and lactose contents. Finally, combination 4 included all criteria. When more than one spectrum was found during the search, their average was the estimated spectrum for the query sample. Concretely, this study estimated missing spectra for 1,700 samples using 2,000,000 spectral DHI records. For assessing the effect of this spectral estimation on the prediction quality, FT-MIR equations were used to predict 11 phenotypes, selected as their quantification used different FT-MIR regions. They were related to the milk fat and mineral composition, lactoferrin content, quantity of eructed methane, body weight (BW), and dry matter intake. The accuracy between predictions obtained from actual and estimated spectra was evaluated by calculating the mean absolute error (MAE). The criteria in the fourth and second combinations were too strict to estimate a spectrum for most samples. Indeed, for many samples, no spectra with the same values for those matching criteria was found. The third match criteria combination had a poorer prediction performance for all studied traits and spectral absorptions than the first combination due to fewer matched samples available to compute the missing spectrum. By allowing a range for matching lactose content (±0.1 g/dL milk), we showed that this new combination increased the number of selected samples to compute missing spectra and predict better the infrared absorption at different wavenumbers, especially those related to the lactose quantification. The prediction performance was further improved by performing queries on the entire Walloon DHI spectral database (6,625,570 spectra), and it varied among the studied phenotypes. Without considering the traits used for the matching, the best predictions were obtained for the content of saturated fatty acids (MAE = 0.15 g/dL milk) and BW (MAE = 12.80 kg). Yet, the predictions for the unsaturated fatty acids were less accurate (MAE = 0.13 and 0.018 g/dL milk for monounsaturated and polyunsaturated fatty acids), likely because of the poorer predictions of spectral regions related to long-chain fatty acids. Similarly, poorer predictions

Received March 6, 2023.
Accepted July 7, 2023.
*Corresponding author: hsoyeurt@uliege.be

were observed for the amount of methane eructed by dairy cows (MAE = 47.02 g/d), likely because it is not directly related to fat content or composition. Prediction accuracies for the remaining traits were also low. In conclusion, we observed that increasing the number of relevant matching criteria helps improve the quality of FT-MIR predicted phenotypes and the number of spectra used during the search. So, it would be of great interest to test in the future the suitability of the developed methodology with large-scale international spectral databases to improve the reliability of EBV from these FT-MIR-based phenotypes and the robustness of FT-MIR predictive models.

Key words: mid-infrared spectrometry, milk, prediction

INTRODUCTION

The use of milk Fourier transform mid-infrared (**FT-MIR**) spectroscopy to develop management and breeding tools for dairy farmers and the dairy industry is becoming increasingly popular, thanks to the many phenotypes currently available. Indeed, with a moderate to high prediction accuracy (R^2 from 0.60 to 0.99), equations exist to assess the nutritional quality of milk by estimating, for instance, the contents of milk fatty acids, protein fractions (Franzoi et al., 2019), minerals (Christophe et al., 2021), or lactoferrin (Soyeurt et al., 2020). The milk composition and, therefore, its related spectrum are also useful for assessing the technological properties of milk (Bonfatti et al., 2016; Visentin et al., 2015). Moreover, even as milk is an important source of nutritive elements for humans, its fine composition—or more specifically composition changes—is vital to know, as it mirrors the metabolic and health status of the animal. Some indicators related to metabolism and animal health and welfare can be derived from the milk FT-MIR spectrum, including energy balance or intake (Ho et al., 2020; McParland et al., 2012; McParland and Berry, 2016), BW (Tedde et al., 2021a), DMI (Tedde et al., 2021b), acetonemia (Grelet et al., 2016), pregnancy status (Delhez et al., 2020), subacute ruminal acidosis (Mensching et al., 2021), fertility (Bastin et al., 2016; Ho et al., 2019), lameness (Bonfatti et al., 2020), and so on. Another critical topic is the environmental footprint of milk production. Some models have been developed to cover part of this topic, for example, equations allowing the estimation of methane eructed by dairy cows (Vanlierde et al., 2021) or the assessment of the nitrogen usage efficiency of dairy cows to improve feed efficiency (Grelet et al., 2020). The FT-MIR spectroscopy can also be used to detect abnormal milk (Hansen and Holroyd, 2019).

Unfortunately, even if those FT-MIR-based traits are of interest to be considered in genetic evaluations, the reliabilities of their EBV are often low due to the lack of long-term spectral data acquisition strategies (Kandel et al., 2017). Therefore, imputing milk FT-MIR spectra from historical records could be relevant to improve the reliability of EBV. This spectral estimation could also be useful in developing or improving prediction models themselves if the imputation error is low enough. Indeed, developing a predictive model requires the creation of a data set containing the spectral data as well as the reference trait of interest measured using a certified methodology. Creating such a database is a crucial point for the future robustness of the developed equation. To enhance the model's robustness, the data sets need to be as diverse as possible in terms of spectral information and the reference traits (Grelet et al., 2021). Sometimes, the acquisition of reference samples is rare or costly, as is the case, for the amount of methane eructed by dairy cows or the quantity of DMI. Thus, we are forced to use as much as possible the number of reference measurements that we can gather from past experiments or through international collaborations. Unfortunately, those reference measurements, especially from past experiments, are often unrelated with milk spectral data, even if a milk sample was analyzed. In reality, spectral data storage is not automatized everywhere, precluding the use of an important amount of reference data due to missing spectra. Hence, estimating missing milk FT-MIR spectra could be a cost-effective way to use these available past reference records without spectral data.

The working hypothesis of this research is that a missing spectrum can be estimated by a spectrum that enables the quantification of the same or comparable FT-MIR-based phenotypes routinely recorded by DHI organization. The contents of protein, fat, and sometimes lactose and urea predicted by milk FT-MIR spectrometry have been collected for a long time and recorded in databases managed by the DHI organizations or its associated data record processing centers. These traits are predicted using spectral data, often not stored after milk spectrometric analyses. As the predictions of fat, protein, lactose, and urea traits used different FT-MIR regions, it seems logical to reversely estimate the spectral data by considering various combinations of those 4 traits as the criteria. In this study, we developed and evaluated a strategy that leverages the vast amount of spectral data recorded by DHI organizations to identify the spectra that most closely align with queries of fat, protein, urea, and lactose as the FT-MIR predictions. This approach can be characterized by finding the nearest FT-MIR spectrum based on the predicted contents of fat, protein, urea, and lactose.

Table 1. Descriptive statistics of the experimental data sets used in the present study

Item	Variable	N	Mean	SD	Minimum	Maximum
Full data set	Fat (g/dL milk)	6,625,570	4.10	0.76	0.10	18.46
	Protein (g/dL milk)	6,625,570	3.45	0.42	0.21	15.50
	Lactose (g/dL milk)	6,358,998	4.71	0.24	0.01	8.94
	Urea (mg/L)	6,621,432	248.50	86.92	1.00	2,000.00
Subset	Fat (g/dL milk)	2,000,000	4.14	0.76	0.13	18.46
	Protein (g/dL milk)	2,000,000	3.48	0.41	0.50	13.05
	Lactose (g/dL milk)	1,985,770	4.75	0.22	0.69	5.67
	Urea (mg/L)	1,998,759	257.98	88.07	1.00	1,870.00
Samples to predict	Fat (g/dL milk)	1,700	4.08	0.64	1.94	6.64
	Protein (g/dL milk)	1,700	3.46	0.41	2.41	7.14
	Lactose (g/dL milk)	1,700	4.72	0.22	2.90	5.25
	Urea (mg/L)	1,699	206.46	71.34	20.00	570.00

MATERIALS AND METHODS

No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Data

This study used the milk FT-MIR spectral database managed by the Walloon Breeders Association (Elevéo, AWÉ group, Ciney, Belgium), which included 6,625,570 records from 381,102 cows and 2,010 farms located in Southern Belgium that participated in milk recording between January 2007 and November 2021. All samples were analyzed using Foss spectrometers (Hillerod, Denmark). Milk analysis was performed using various models of Foss spectrometers (from MilkoScan FT4000 to MilkScan FT+ and FT7). To ensure comparability, the spectral data recorded after 2012 were standardized using the methodology proposed by Grelet et al. (2015). The contents of fat, protein, lactose, and urea were predicted by the equations developed by the manufacturer for each spectrometer model. To reduce the required computation resources for comparing the different searching methods, we selected randomly a subset of 2,000,000 records collected from December 2016 to November 2021 from 152,793 cows on 947 farms. We assumed that this subset had a good representation of the spectral data in the Walloon part of Belgium because both data sets had closely comparable means and standard deviations of the 4 querying traits (Table 1). We chose to impute the spectrum of the first 1,700 samples of this subset. The spectral data of those samples were considered missing and were estimated using the methodology proposed in this study. Then, the estimated missing spectra were compared with the actual spectra in the database to assess the methodology accuracy.

Spectral Prediction

The method used to estimate the missing milk FT-MIR spectrum involved searching for a matched record or records between a query sample and reference samples in the DHI database based on one of the following 4 match criteria combinations (**MCC**):

- MCC1: equal contents of fat and protein with a precision of 2 decimals;
- MCC2: equal contents of fat, protein, and urea;
- MCC3: equal contents of fat, protein, and lactose;
- MCC4: equal contents of fat, protein, lactose, and urea.

If a match was found, the predicted spectrum for the query sample was set to be the spectrum of the matched reference sample. If more than one match was identified, the predicted spectrum for the query sample was calculated by averaging the absorbance values for all wavenumbers of all the matched reference samples. The number of the FT-MIR spectra used to calculate the predicted spectra was recorded to assess the amount of information used by each match criteria combination to estimate the spectrum.

To evaluate the spectral imputation quality, we employed 2 approaches. First, for assessing the effect of this spectral estimation on the prediction of specific FT-MIR phenotypes, the FT-MIR equations predicting the contents of fat, SFA, MUFA, PUFA, calcium (Ca), sodium (Na), phosphorus (P), lactoferrin (**LACTOF**) as well as the BW, the quantity of eructed methane (CH_4) and DMI were applied to the estimated and referenced spectra. Table 2 describes the characteristics of the equations used. The prediction accuracy was assessed by calculating the correlation and mean absolute error (**MAE**) estimated between the predicted and actual FT-MIR values of each predicted phenotypes. MAE was computed as follows:

Table 2. Characteristics of the used FT-MIR equations[1]

Trait	Unit	n	R^2	Reference
Fat content	g/dL milk	1,799	1.00	NP[2]
DMI	kg/d	10,711	0.46	Tedde et al., 2021b
BW	kg	1,849	0.61	Tedde et al., 2021a
Methane	g/d	1,089	0.68	Vanlierde et al., 2021
Calcium	mg/kg	1,106	0.81	Christophe et al., 2021
Sodium	mg/kg	997	0.43	Christophe et al., 2021
Phosphorus	mg/kg	1,126	0.72	Christophe et al., 2021
Lactoferrin	mg/L milk	2,442	0.53	Soyeurt et al., 2020
SFA	g/dL milk	1,790	0.99	Grelet et al., 2014
MUFA	g/dL milk	1,793	0.97	Grelet et al., 2014
PUFA	g/dL milk	1,788	0.97	Grelet et al., 2014

[1]n = number of reference samples used to build the equation; R^2 = cross-validation coefficient of determination; Reference = the reference for the published equation if available.
[2]NP = not published.

$$\text{MAE} = \frac{\sum_{i=1}^{nsample} |predicted_i - observed_i|}{n},$$

where n is the number of observations. The second approach involved visualizing the correlation values estimated between the predicted and reference infrared absorbance values to assess the part of the FT-MIR regions that were more accurately predicted.

RESULTS AND DISCUSSION

The performance of the 4 match criteria combinations was evaluated and compared based on the correlation and MAE between the actual and predicted phenotypes, as shown in Table 3. The number of predicted samples that met all desired conditions varied with the combination used. As the number of match conditions increased, the number of predicted samples decreased drastically. MCC1 with only 2 match conditions (i.e., equal fat and protein contents) yielded the highest percentage (99.5%) of predicted samples, followed by MCC2 (57.2%) and MCC3 (55.6%), each with 3 match conditions. MCC4, which used all 4 match conditions, had only 29 (1.7%) predicted samples.

The correlation and MAE obtained for each FT-MIR-based phenotype also varied depending on the match criteria combination used (Table 3). The highest correlations were consistently obtained with MCC1. The prediction error was the lowest or close to the lowest for MCC1, although the MAE values were not directly comparable due to the varying number of predicted samples for each match criteria combination. MCC4 yielded the smallest corrections and mostly the largest MAE (except MUFA, PUFA, and Na). However, when we only kept the samples selected by MCC4 (n = 29), the MAE value obtained using the estimated spectra from MCC1 was still lower. Therefore, enforcing the 4

Table 3. Correlation and mean absolute error (MAE) between the actual and the predicted phenotypes based on the 4 match criteria combinations[1]

	Correlation				MAE			
Item	C1	C2	C3	C4	C1	C2	C3	C4
Number of matched samples	1,692	946	972	29				
% Loss of samples					0.47	44.35	42.82	98.29
%Fat	0.95	0.89	0.93	0.73	0.097	0.081	0.109	0.135
Methane	0.66	0.31	0.45	0.20	61.98	61.83	71.45	84.12
SFA	0.89	0.77	0.81	0.76	0.20	0.19	0.21	0.20
MUFA	0.63	0.30	0.35	0.21	0.14	0.16	0.15	0.15
PUFA	0.68	0.40	0.39	0.18	0.024	0.019	0.026	0.023
BW	0.94	0.91	0.92	0.84	13.50	16.37	15.17	20.60
DMI	0.63	0.42	0.41	0.37	1.84	2.24	2.21	2.92
Calcium	0.66	0.36	0.40	0.02	68.57	81.95	79.98	86.94
Sodium	0.31	0.07	0.60	0.17	39.26	44.59	25.49	22.89
Phosphorus	0.51	0.27	0.34	0.10	75.75	84.64	82.49	88.60
Lactoferrin	0.37	0.11	0.23	0.20	80.41	99.06	81.67	110.51

[1]Combination 1 (C1): equal contents of fat and protein; Combination 2 (C2): equal contents of fat, protein, and urea; Combination 3 (C3): equal contents of fat, protein, and lactose; Combination 4: equal contents of fat, protein, urea, and lactose.

match conditions in MCC4 did not seem appropriate for estimating missing milk FT-MIR spectra. MCC2 and MCC3 had roughly the same number of predicted samples, yet the former had smaller correlations than the latter for most of the traits. Overall, MAE was also larger with MCC2 than MCC3. Hence, considering the urea content as a criterion did not seem to yield as accurate results as considering lactose. The 2 promising match criteria combinations are MCC1 and MCC3.

To ensure a fair comparison based on the same number of samples, we calculated correlation and MAE values for the predicted phenotypes for all the 4 match criteria combinations but using the samples with matched spectra based on MCC3 (Table 4). Still, MCC1 performed better than MCC3 because the former had more spectra used to create the final estimated spectrum. On average, MCC1 used 108 ± 60 spectra, whereas MCC3 used 4 ± 3 samples. Because matched samples with equal querying milk contents can vary with their spectra, by taking the average spectra value of the matched samples as a predicted spectrum, MCC1, with far more matched samples per query, tended to yield more precise and accurate estimates of spectra than MCC3. The theory behind this hypothesis could be formulated as follows. Milk FT-MIR spectra (denoted by X) are often used to predict milk quality traits or composition (denoted by Y): $Y = f(X)$, where f is a function that maps from an input X to the output Y. However, we are interested in predicting missing X, given Y in the data repository as the input. This task can be viewed as a reverse mapping. Not to lose generality, we let $X = g(Y)$, where $g(Y)$ is a predictive function of X, now given Y as the predictor. For the prediction purpose, $g(Y)$ can take any form, which does not have to be precisely an inverse function of f. In this research, we used a straightforward approach to predict the milk FT-MIR spectrum by matching samples based on each match criteria combination. Statistically, applying each match criteria combination leads to a form of $g(Y)$, which is a degenerate distribution with the point mass equaling the actual FT-MIR spectra (x_i) of the matched sample, conditional on that the selected phenotypes of both samples matched precisely or approximately. When multiple matches ($n > 1$) exist, the FT-MIR estimate $\left(\hat{x}^*\right)$ is given by the corresponding mean of the degenerate points: $\hat{x}^* = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\nabla\left|y^* - y_i\right| \le \delta$, where $\delta = 0$ if precise matches are required. Note that we later relaxed this zero restraint to allow an acceptable range of errors (e.g., for lactose content). Assume that each matched point x_i deviates from the actual value x^* by a quantity ε_i, which represents an error term. Hence, the above becomes: $\hat{x}^* = x^* + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$, $\nabla\left|y^* - y_i\right| \le \delta$. The standard error of $\hat{x}^*$ is given by $se\left(\hat{x}^*\right) = \frac{1}{n}\sqrt{\sum_{j=1}^{n}\varepsilon_i^2}$. Here, the prediction accuracy depends on 2 facts, the number of "matched samples" and the mapping accuracy. First, if the deviations occur randomly by nature, the estimation error will decrease or diminish as the number of matched records increases because $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \to 0$ when $n \to \infty$. Hence, the more matched samples, the higher accuracy (i.e., higher correlation and smaller MAE). Second, we observed a link between the forward and reverse mapping of phenotypes. Assume that the forward mapping from X to Y has an extremely high coefficient of determination. Then, the coefficient of determination for the reverse mapping from Y to X is also high. For example, consider fat content. The determination coefficient using the existing FT-MIR for predicting fat content, as evaluated by cross-validations, was extremely high ($R^2 = 1.00$). Hence, the correlation between the predicted and actual fat content was also high (r = 0.93) based on MCC1. Conversely, if the determination coefficient for forward mapping is low, then the accuracy for the reverse mapping will also be low. For the simplicity of illustration, assume a linear relationship between X and Y, say $y_i = bx_i + e_i$. Then, we show that the reverse mapping is the same coefficient of determination as the forward mapping because:

$$R^2_{Y\to X} = 1 - \frac{\left(\frac{1}{b}\right)^2 \sum_i^n \varepsilon_i^2}{\left(\frac{1}{b}\right)^2 \sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} =$$

$$1 - \frac{\sum_i^n \varepsilon_i^2}{\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} = R^2_{X\to Y}.$$

Although the reverse mapping from Y to X is high for fat and protein, including addition traits, such as lactose and urea, did not increase the accuracy further. Instead, it matched considerably fewer records, leading to drastically increased estimation errors (Tables 3 and 4).

The lactose content is not always corrected by a laboratory because it is not included in milk pricing. Therefore, enforcing precisely equal lactose contents, as with MCC3, can be too strict to find any matched sample. To address this issue, we proposed a modification to MCC3 by allowing a difference of lactose content within a range of, say, ±0.1 g/dL milk between the query sample and a reference sample, in addition to

Table 4. Correlation and mean absolute error (MAE) calculated from the actual and predicted phenotypes using match criteria combinations 1, 3, and 3b, respectively[1]

	Correlation			MAE		
Item	C1	C3	C3b	C1	C3	C3b
%Fat	0.94	0.93	0.94	0.093	0.109	0.095
Methane	0.63	0.45	0.57	55.23	71.45	56.79
SFA	0.88	0.81	0.87	0.18	0.21	0.19
MUFA	0.53	0.35	0.51	0.12	0.15	0.12
PUFA	0.61	0.39	0.59	0.022	0.026	0.023
BW	0.94	0.92	0.95	12.63	15.17	12.48
DMI	0.60	0.41	0.59	1.72	2.21	1.77
Calcium	0.59	0.40	0.53	63.79	79.98	66.55
Sodium	0.30	0.60	0.72	28.74	25.49	20.41
Phosphorus	0.44	0.34	0.46	75.87	82.49	74.89
Lactoferrin	0.41	0.23	0.49	59.99	81.67	60.36

[1]Combination 1 (C1): equal contents of fat and protein; Combination 3 (C3): equal contents of fat, protein, and lactose; Combination 3b (C3b): equal contents of fat and protein plus a comparable lactose content within a range of ±0.1 g/dL milk between the query sample and a reference sample. n = 972.

enforcing precisely the same fat and protein contents. This new match criteria combination is referred to as MCC3b in Table 4 and throughout this manuscript. With this modification, the number of matched samples increased to 35 ± 23. As a result, MCC3b had a significantly higher correlation and lower MAE between the prediction and actual phenotypes than MCC3 (Table 4). Moreover, MCC3b matched 1,682 samples whereas MCC3 matched only 972 samples. Still, these numbers were smaller than the 1,692 matched samples by MCC1.

MCC3b did not outperform MCC1 because MCC1 had far more matched samples. Nevertheless, we detected exceptions. MCC3b gave more accurate predictions for BW, Na, and P (i.e., higher correlation and lower MAE) than MCC1. These results did not agree with the hypothetical theory presented earlier. Indeed, we observed an additional dimension of the problem to determine the prediction accuracy, which the above hypothetical theory did not consider: all FT-MIR regions do not have the same importance for predicting all traits. For instance, fat or protein contents are not predicted using the same FT-MIR regions. Therefore, adding new traits using a different part of the spectrum for its prediction in the match criteria can be relevant. The accuracy variability between the predicted phenotypes could be explained by the precision in the spectral estimation with different wavelengths. Figure 1 shows the correlation values obtained for each spectral point after using MCC1 and MCC3b, respectively, to select the spectra. For both match criteria combinations, we can observe that the accuracy of predicting the absorbance value at a specific wavenumber varied. In some regions, the correlation value was close to 0. These regions corresponded to those located between 1,600 to 1,689 cm^{-1} and 3,008 to 5,010 cm^{-1}, which are related to the noisy regions related to water absorption, according to Grelet et al. (2015). Overall, the correlations calculated based on MCC3b were higher than those based on MCC1, suggesting a benefit in keeping the lactose content in the match criteria set. For instance, the region around 1,550 cm^{-1} had high correlations between the predicted and actual absorptions for MCC1 and MCC3b. This region was related to C–N and N–N stretching, required to quantify the protein content (Grelet et al., 2015). This is expected as the match criteria considered the protein content. Grelet et al. (2015) mentioned that the prediction of lactose requires absorbances estimated around 1,045 cm^{-1} with C–O stretching vibration of alcohols functions, 1,076 cm^{-1} with C–O, C–C, and C–H stretching vibration, and 1,157 and 1,250 cm^{-1} with C–O–C ether stretching. Those regions were well predicted using MCC3b but not using MCC1 because the former included the lactose content as a match criterion. These results again confirmed the benefit of including lactose as a match criterion. We will discuss the case of the fat content a little bit later in this discussion about the phenotype performances. The above results pinpointed that the performance of match criteria can vary enormously at different spectrum locations or wavenumbers (Figure 1).

To confirm our findings, we implemented new predictions based on MCC3b in the entire spectral database generated from the routine milk recording (Table 5), which consisted of 6,625,570 spectral records (Table 1). Even though the number of matched samples based on MCC3b increased from 972 (Table 4) to 1,682 (Table 5), the accuracy as measured by the correlation between the predicted and actual phenotypes of the 11 traits increased substantially from 0.46 (P) and 0.93 (%Fat) to between 0.67 (P) and 0.97 (%Fat). Meanwhile, MAE decreased for all 11 traits. On average, the

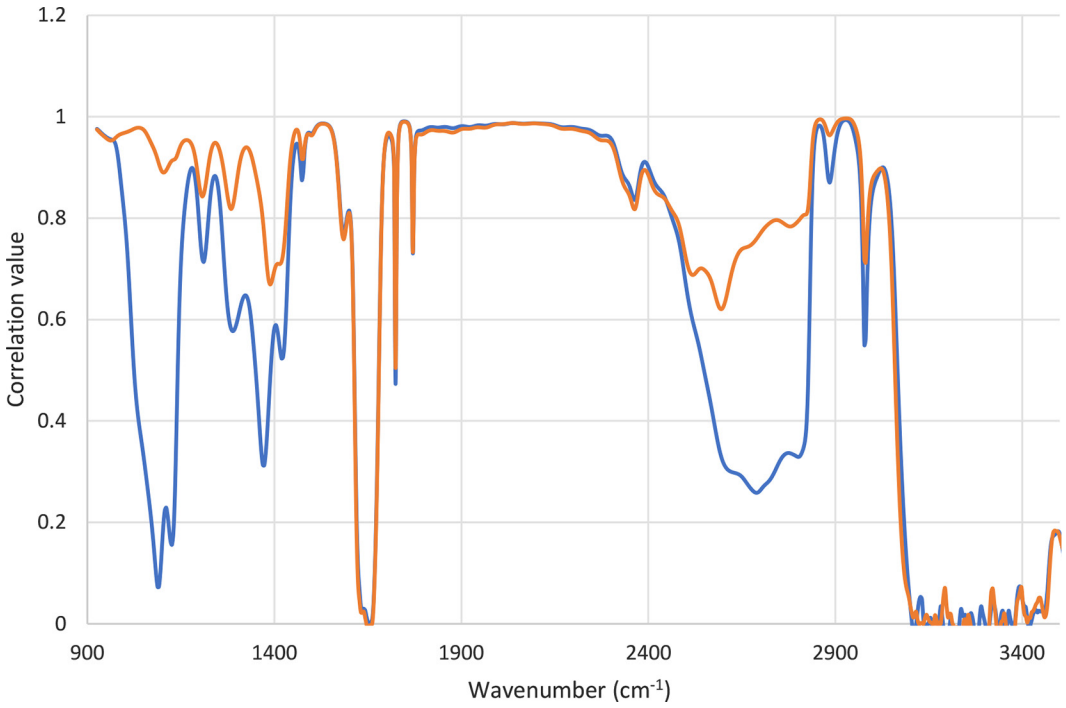

Figure 1. Correlation values between actual and predicted absorptions at different wavenumbers based on match criteria combinations 1 (blue line) and 3b (orange line) in 972 samples. Combination 1: equal contents of fat and protein; Combination 3b: equal contents of fat and protein plus a comparable lactose content within a range of ± 0.1 g/dL milk between the query sample and a reference sample.

increment rate of the correlations for the 11 traits was 19.6%, and the decrement rate of MAE was 8.8%. In the entire data set, MCC3b generated a comparable number (1,682) of predicted samples compared with the previous number (1,692) of predicted samples for MCC1. Yet, MCC3b had higher correlations than those previously obtained based on MCC1, and MAE were also lower for most of the 11 traits except MUFA, BW, and Ca. The performance improvement with MCC3b was related to the number of selected spectra increased from 35 (±23) when applied to the 2 million records to 84 (±64) when applied to the entire Walloon milk recording spectral database. This is still lower than the number of spectra used by MCC1 on the initial 2 million records. Therefore, these results showed that the number of samples mattered but was not the sole decisive factor. Indeed, and this is an important point, to have a better-predicted spectrum, one also needs to consider using match criteria in different milk FT-MIR regions.

The final performances in Table 5, and other tables, indicate that prediction performance differed between the studied traits. We start with the comparison between phenotypes prediction performances from the predicted spectrum for fat content. Despite the matching criteria requiring an equal content between the fat content estimated by the milk laboratory for the reference and the selected spectra, we never obtained a correlation between the fat contents equal to 1. The reason is that this correlation was calculated with the predicted fat contents using the equation in Table 1 from the actual and estimated spectra. Therefore, the traits used were not identical as the fat and protein records provided by the milk analysis laboratory were corrected. This correction, based on ISO 9622:2013–IDF 141:2013, is for the bias and slope, which is carried out by analyz-

Table 5. Correlation and mean absolute error (MAE) calculated from the predicted and observed phenotypes (n = 1,682) after applying match criteria combination 3b[1] on the entire Walloon milk recording database (n = 6,625,570)

Item	Correlation	MAE
%Fat	0.97	0.048
Methane	0.66	47.02
SFA	0.91	0.15
MUFA	0.69	0.13
PUFA	0.72	0.018
BW	0.92	12.80
DMI	0.70	1.71
Calcium	0.71	63.83
Sodium	0.83	20.43
Phosphorus	0.62	67.50
Lactoferrin	0.66	71.73

[1]Combination 3b: equal contents of fat and protein plus a comparable lactose content within a range of ±0.1 g/dL milk between the query sample and a reference sample.

ing samples with known fat and protein content and then applying the slope and bias to the fat and protein predictions provided by the spectrometer to account for the variability of the spectral signal through time. This correction is not applied to the spectral data. Even if the spectral data were standardized, this standardization is not carried out as often as the bias and slope corrections. Indeed, bias is checked every 40 samples, whereas spectral standardization is done only once monthly.

For a trait not used as the match criterion, the prediction accuracy depends on its correlation with the traits used as match criteria. For example, consider MCC1. The prediction accuracy for a nonmatch-criterion trait will be high (or low) if it has a high (or low) correlation with fat and protein contents. This was the reason for the high prediction accuracy (correlation = 0.89) for SAT content. Indeed, the correlation between fat and SAT predictions was around 0.98 from all our data sets. In contrast, the predictions for the unsaturated fatty acids (MUFA and PUFA) were less accurate (correlation = 0.63–0.68) than the predictions for SAT. The large discrepancies in prediction accuracies among these traits lead us to postulate that important wavenumbers responsible for the latter traits with poor prediction accuracies were not accurately predicted by our method. Indeed, Grelet et al. (2015) mentioned that the informative signals related to the fat chains appeared around 1,390 and 1,454 cm^{-1} with C–H bending of −CH3 and −CH2, and around 2,862 and 2,927 cm^{-1} with C–H stretching of −CH3 and −CH2. This first region located around 1,400 cm^{-1} was not well represented by our method (Figure 1). This could lead to the insufficient prediction of long-chain fatty acids, which are mainly present in the quantification of MUFA and PUFA. The region around 1,743 cm^{-1} is also interesting in quantifying the fat content because of the C = O ester stretching (Grelet et al., 2015). This region was, more or less, well predicted. Still, the predictions were poor for the quantity of methane eructed by dairy cows because it was not directly linked to fat content or composition (Vanlierde et al., 2021). Possibly, including additional traits for a more balanced FT-MIR coverage is appealing to improve the phenotypes prediction performances, yet subject to having an adequate number of matched samples.

The above results have motivated us to consider including additional information as match criteria that could allow predicting each wavenumber better. We have tested the effect of adding information external to the milk spectrum per se, such as days in milk and test seasons, which could be related to the milk spectral variability. For example, Soyeurt et al. (2022) showed that it was possible to predict the intensity of grass in the diet of dairy cows which was indirectly affected by the seasons when the milk spectral data were collected in the Walloon Region of Belgium. However, adding the information for days in milk during the selection of spectra and test season did not improve the prediction performance compared with using MCC1 (data not shown). The example of MCC3b demonstrates that increasing the number of relevant matching parameters, enhancing the matching of each individual wavenumber, and additional matching criteria such as days in milk, milk yield, and parity, could be included as the reference data set expands. Then, it would be worthwhile to investigate in the future whether the developed methodology is applicable to a large-scale international spectral database. Indeed, by grouping spectral databases, the prediction accuracy could be higher.

CONCLUSIONS

We have proposed a strategy that leverages existing spectral databases to estimate missing spectra using common dairy traits as matching criteria. This strategy has the potential to improve FT-MIR prediction models and accuracies for FT-MIR predicted traits, particularly for traits driven by similar FT-MIR signals as those used as part of the matching criteria. The best imputations required more than 80 selected spectra and the match criteria including traits with predictions by FT-MIR involving different spectral regions. Therefore, the third match criteria combination allowing an appropriate range for the lactose of ±0.1 g/dL milk, in addition to enforcing equal fat and protein contents, yielded the best predictions. On the one hand, some spectral regions were well predicted based on the observed accuracy measures (i.e., correlation and MAE between actual and predicted phenotypes). In principle, applying an equation on those well-predicted FT-MIR regions could yield good predictions for relevant phenotypes. Thus, we hold that new equations should be developed on these well-predicted regions. The oldest equations could also be rebuilt by utilizing those well-predicted regions. If the proposed limitation of spectral points does not significantly decrease the prediction accuracy, then this solution is relevant. Nevertheless, further confirmation and investigations probing into this situation as well as of the interest of this method to improve also indirectly the reliabilities of EBV may be needed. More specifically, it is required to quantify the increase of R2 for a prediction equation and the increase in the reliability of EBV by adding imputed spectra. To achieve this first objective, it will be needed to collect the predictions of matching criteria done by the spectrometers when samples used to build equations are analyzed by FT-MIR spectrometry. Unfortu-

nately, even if we have many training sets, this job was not done leading to the impossibility of evaluating the effect of imputed spectra on the equation performances. For the reliability improvement of EBV, a genetic evaluation should be performed using the predictions given by the entire spectral data set. Then, the obtained reliability for EBV should be compared with the same predictions, except that a certain percentage would be obtained by a prediction from imputed spectra. The presence of predictions obtained from imputed spectra could reduce the variability, especially for the trait using the spectral wavenumbers having the poorer imputed prediction performances (e.g., methane). Therefore, this must have a very bad effect on the genetic evaluation of a specific cow if the entire spectral data used to evaluate it is imputed. So, a study about the maximum imputed spectra that could be added to a training set or in a genetic evaluation must be investigated. Another needed investigation could be related to the interest in including additional informative criteria to improve the overall predictive quality of the FT-MIR spectra, which can enhance the estimation of FT-MIR regions whose prediction performance is poorer based on the currently proposed match criteria. We tested external criteria such as test seasons and days in milk, but they did not improve the prediction performances. Nevertheless, the roles of other traits directly related to the milk FT-MIR variability need to be evaluated. For instance, if there are records for unsaturated fatty acid contents predicted by FT-MIR and calibrated by their respective laboratories, they may be a potential candidate for additional criteria in future studies. In conclusion, this article is the first piece of the puzzle about the use of imputed milk FT-MIR spectra in dairy farming and additional investigations are needed to evaluate its concrete interest in the field.

ACKNOWLEDGMENTS

The University of Liège–Gembloux Agro-Bio Tech (Liège, Belgium) supported computations through the technical platform Calcul et Modélisation Informatique of the TERRA Teaching and Research Centre, partly supported by the Fund for Scientific Research–FNRS under grant no. T.0095.19 (PDR "DEEPSELECT"). The authors acknowledge the milk laboratory of Comité du Lait (Battice, Belgium) for providing the milk spectra used in this study. The authors have not stated any conflicts of interest.

REFERENCES

Bastin, C., L. Théron, A. Lainé, and N. Gengler. 2016. On the role of mid-infrared predicted phenotypes in fertility and health dairy breeding programs. J. Dairy Sci. 99:4080–4094. https://doi.org/10.3168/jds.2015-10087.

Bonfatti, V., L. Degano, A. Menegoz, and P. Carnier. 2016. Short communication: Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. J. Dairy Sci. 99:8216–8221. https://doi.org/10.3168/jds.2016-10953.

Bonfatti, V., P. N. Ho, and J. E. Pryce. 2020. Usefulness of milk mid-infrared spectroscopy for predicting lameness score in dairy cows. J. Dairy Sci. 103:2534–2544. https://doi.org/10.3168/jds.2019-17551.

Christophe, O. S., C. Grelet, C. Bertozzi, D. Veselko, C. Lecomte, P. Höeckels, A. Werner, F.-J. Auer, N. Gengler, F. Dehareng, and H. Soyeurt. 2021. Multiple breeds and countries' predictions of mineral contents in milk from milk mid-infrared spectrometry. Foods 10:2235. https://doi.org/10.3390/foods10092235.

Delhez, P., P. N. Ho, N. Gengler, H. Soyeurt, and J. E. Pryce. 2020. Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? J. Dairy Sci. 103:3264–3274. https://doi.org/10.3168/jds.2019-17473.

Franzoi, M., G. Niero, G. Visentin, M. Penasa, M. Cassandro, and M. De Marchi. 2019. Variation of detailed protein composition of cow milk predicted from a large database of mid-infrared spectra. Animals (Basel) 9:176. https://doi.org/10.3390/ani9040176.

Grelet, C., C. Bastin, M. Gelé, J.-B. Davière, M. Johan, A. Werner, R. Reding, J. 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. Fernandez Pierna, H. Soyeurt, F. Dehareng, N. Gengler, and P. Dardenne. 2014. Creation of universal MIR calibration by standardization of milk spectra: Example of fatty acids. Page 108 in Book of Abstracts of the 65th Annual Meeting of the European Federation of Animal Science, 108; Copenhagen, Denmark. Wageningen Academic Publishers.

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.

Grelet, C., E. Froidmont, L. Foldager, M. Salavati, M. Hostens, C. P. Ferris, K. L. Ingvartsen, M. A. Crowe, M. T. Sorensen, J. A. Fernandez Pierna, A. Vanlierde, N. Gengler, and F. Dehareng. 2020. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. J. Dairy Sci. 103:4435–4445. https://doi.org/10.3168/jds.2019-17910.

Hansen, P. W., and S. E. Holroyd. 2019. Development and application of Fourier transform infrared spectroscopy for detection of milk adulteration in practice. Int. J. Dairy Technol. 72:321–331. https://doi.org/10.1111/1471-0307.12592.

Ho, P. N., V. Bonfatti, T. D. W. Luke, and J. E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102:10460–10470. https://doi.org/10.3168/jds.2019-16412.

Ho, P. N., L. C. Marett, W. J. Wales, M. Axford, E. M. Oakes, and J. E. Pryce. 2020. Predicting milk fatty acids and energy balance of dairy cows in Australia using milk mid-infrared spectroscopy. Anim. Prod. Sci. 60:164. https://doi.org/10.1071/AN18532.

Kandel, P. B., M.-L. Vanrobays, A. Vanlierde, F. Dehareng, E. Froidmont, N. Gengler, and H. Soyeurt. 2017. Genetic parameters of mid-infrared methane predictions and their relationships with milk production traits in Holstein cattle. J. Dairy Sci. 100:5578–5591. https://doi.org/10.3168/jds.2016-11954.

McParland, S., G. Banos, B. McCarthy, E. Lewis, M. P. Coffey, B. O'Neill, M. O'Donovan, E. Wall, and D. P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy

status in Holstein-Friesian cows. J. Dairy Sci. 95:7225–7235. https://doi.org/10.3168/jds.2012-5406.

McParland, S., and D. P. Berry. 2016. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. J. Dairy Sci. 99:4056–4070. https://doi.org/10.3168/jds.2015-10051.

Mensching, A., M. Zschiesche, J. Hummel, C. Grelet, N. Gengler, S. Dänicke, and A. R. Sharifi. 2021. Development of a subacute ruminal acidosis risk score and its prediction using milk mid-infrared spectra in early-lactation cows. J. Dairy Sci. 104:4615–4634. https://doi.org/10.3168/jds.2020-19516.

Soyeurt, H., C. Gerards, C. Nickmilder, J. Bindelle, S. Franceschini, F. Dehareng, D. Veselko, C. Bertozzi, N. Gengler, A. Marvuglia, A. Bayram, and A. Tedde. 2022. Prediction of indirect indicators of a grass-based diet by milk Fourier transform mid-infrared spectroscopy to assess the feeding typologies of dairy farms. Animals (Basel) 12:2663. https://doi.org/10.3390/ani12192663.

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.

Tedde, A., C. Grelet, P. Ho, J. Pryce, D. Hailemariam, Z. Wang, G. Plastow, N. Gengler, Y. Brostaux, E. Froidmont, F. Dehareng, C. Bertozzi, M. Crowe, I. Dufrasne, and H. Soyeurt. 2021a. Validation of dairy cow bodyweight prediction using traits easily recorded by dairy herd improvement organizations and its potential improvement using feature selection algorithms. Animals (Basel) 11:1288. https://doi.org/10.3390/ani11051288.

Tedde, A., C. Grelet, P. Ho, J. Pryce, D. Hailemariam, Z. Wang, G. Plastow, N. Gengler, E. Froidmont, F. Dehareng, C. Bertozzi, M. Crowe, and H. Soyeurt. 2021b. Multiple country approach to improve the test-day prediction of dairy cows' dry matter intake. Animals (Basel) 11:1316. https://doi.org/10.3390/ani11051316.

Vanlierde, A., F. Dehareng, N. Gengler, E. Froidmont, S. McParland, M. Kreuzer, M. Bell, P. Lund, C. Martin, B. Kuhla, and H. Soyeurt. 2021. Improving robustness and accuracy of predicted daily methane emissions of dairy cows using milk mid-infrared spectra. J. Sci. Food Agric. 101:3394–3403. https://doi.org/10.1002/jsfa.10969.

Visentin, G., A. McDermott, S. McParland, D. P. Berry, O. A. Kenny, A. Brodkorb, M. A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J. Dairy Sci. 98:6620–6629. https://doi.org/10.3168/jds.2015-9323.