J. Dairy Sci. 107:9404–9414
https://doi.org/10.3168/jds.2023-24472

Standardization method, testing scenario, and accuracy of the infrared prediction model affect the standardization accuracy of milk mid-infrared spectra

W. Lou,[1] H. Lu,[2] X. Ren,[3] X. Zhao,[4] Y. Wang,[1]* and V. Bonfatti[5]
[1]State Key Laboratory of Animal Biotech Breeding, National Engineering Laboratory of Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
[2]Beijing Consortium for Innovative Bio-Breeding, Beijing 101206, China
[3]Henan Dairy Herd Improvement Center, Zhengzhou 450046, China
[4]Shandong Ox Livestock Breeding Co. Ltd., Jinan 250100, China
[5]Department of Comparative Biomedicine and Food Science, University of Padova, 35020 Legnaro (PD), Italy

ABSTRACT

The widespread use of milk mid-infrared (MIR) spectroscopy for phenotype prediction has urged the application of prediction models across regions and countries. Spectra standardization is the most effective way to reduce the variability in the spectral signal provided by different instruments and labs. This study aimed to develop different standardization models for MIR spectra collected by multiple instruments, across 2 provinces of China, and investigate whether the standardization method (piecewise direct standardization, PDS, and direct standardization, DS), testing scenario (standardization of spectra collected on the same day or after 7 mo), infrared prediction model accuracy (high or low), and instrument (6 instruments from 2 brands) affect the performance of the standardization model. The results showed that the determination coefficient (R^2) between absorbance values at each wavenumber provided by the primary and the secondary instruments increased from less than 0.90 to nearly 1.00 after standardization. Both PDS and DS successfully reduced spectra variation among instruments, and performed significantly better than nonstandardization. However, DS was more prone to overfitting than PDS. Standardization accuracy was higher when tested using spectra collected on the same day compared with those collected 7 mo after, but great improvement in model transferability was obtained for both scenarios compared with the nonstandardized spectra. The less accurate infrared prediction model (for C8:0 and C10:0 content) benefited the most from spectra standardization compared with the more accurate model (for total fat and protein content). For spectra collected 7 mo after standardization, after PDS the root mean square error between predictions obtained by different machines decreased on average by 86% and 94% compared with the values before standardization for C8:0 and C10:0, respectively. The secondary instrument had no significant effect on the R^2 between predictions. The variation in the spectral signal provided by different instruments was successfully reduced by standardization across 2 provinces in China. This study lays the foundations for developing a national MIR spectra database to provide consistent predictions across provinces to be used in dairy farm management and breeding programs in China. Additionally, this provides opportunities for data exchange and cooperation at international levels.

Key words: milk mid-infrared spectra, standardization, model transfer, Holstein dairy cow

INTRODUCTION

Mid-infrared (**MIR**) spectroscopy has proved to be a fundamental tool for milk quality control, showcasing its potential to generate high-throughput phenotypes that can be exploited in milk payment systems, herd management, and animal genetic evaluations (De Marchi et al., 2014; Tiplady et al., 2020). Significant insights have been obtained beyond the traditional milk components (De Marchi et al., 2014), such as fatty acids (Fleming et al., 2017b), fat globule size (Fleming et al., 2017a), coagulation and acidity (Manuelian et al., 2017; El Jabri et al., 2019), as well as some functional traits such as cows' methane emissions (Dehareng et al., 2012; Shetty et al., 2017a), feed intake (Shetty et al., 2017b; Wallén et al., 2018; Shadpour et al., 2022), blood metabolites (Benedet et al., 2019; Bonfatti et al., 2019; Luke et al., 2019), energy status (Smith et al., 2019), BW (Soyeurt et al., 2019; Frizzarin et al., 2023a), lameness (Bonfatti et al., 2020), ruminal acidosis risk (Mensching et al.,

Received November 26, 2023.
Accepted April 12, 2024.
*Corresponding author: wangyachun@cau.edu.cn

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-24. Nonstandard abbreviations are available in the Notes.

2021), nitrogen use efficiency (Grelet et al., 2020; Shi et al., 2023), pregnancy status (Ho et al., 2019; Delhez et al., 2020; Tiplady et al., 2022), and diet (Frizzarin et al., 2023b).

Developing new prediction models is a time-consuming and costly process because it requires the collection of numerous samples that must be analyzed following the gold standard methods (e.g., wet laboratory analysis, veterinary diagnosis). This is especially true for models directed toward predicting fine milk composition, a cow's metabolic profile, or other functional traits. Hence, there is an increasing interest in sharing mature prediction models among different laboratories to extend the practical value of such models beyond their original development settings. However, instruments used for milk testing may yield significantly different spectral responses due to variations in instrument brand, physical wear, laboratory environment (e.g., changes in temperature, humidity, experimental operation), and sensor setup (Bouveresse et al., 1998; Feudale et al., 2002; Grelet et al., 2015). Consequently, applying an infrared prediction model developed on one instrument (referred to as the "primary" or "master" instrument) to another (referred to as "secondary" or "slave" instrument) can result in significant errors.

Standardization methods that enable transferring prediction models from the primary instrument to one or more secondary instruments have been proposed to address the poor portability of MIR spectra prediction models and enable their large-scale adoption. These methods allow the pre-established spectral models from the primary instrument to be directly applied to the secondary instruments and guarantee consistent predictions across instruments. Algorithms such as direct standardization (**DS**; Wang et al., 1991) or piecewise direct standardization (**PDS**; Wang et al., 1991; Bouveresse and Massart, 1996) have been applied successfully to MIR spectra (Grelet et al., 2015, 2017). These methods use a set of milk samples, analyzed by the primary and secondary instruments. Spectra from the different instruments are then compared, and a set of standardization parameters are derived to "convert" the spectra from the secondary instrument into that of the primary instrument. A transfer is considered to be successful when the predictions obtained using the spectra collected by a secondary instrument are not significantly different to those obtained using the spectra from the primary spectrometer (Grelet et al., 2015, 2017). In the absence of samples analyzed simultaneously by the primary and secondary instrument (e.g., for datasets of unstandardized spectra collected before a calibration model was developed), a retroactive approach, which uses percentiles of spectral responses for each wavenumber to map the relationships between primary and secondary instruments, can be used (Bonfatti et al., 2017). However, collecting milk samples to be analyzed by the primary and secondary instruments is the most appropriate strategy to reduce prediction errors across time (Tiplady et al., 2019) and should be the method of choice.

In China, the Holstein population consists of more than 5 million dairy cows and is widely distributed throughout the country. Dairy herd improvement organizations comprise many milk quality control laboratories, each using multiple instruments. This urges the introduction of standardization methods to establish a multiregional or national MIR network. Because various infrared prediction equations, each characterized by a different accuracy, are expected to be shared in the future within a network, the effect of the accuracies of prediction models on the accuracy of model transfer should be explored. Therefore, the objectives of this study were to: (a) develop different standardization procedures for MIR spectra collected by multiple instruments, across 2 provinces of China and (b) investigate the effects of standardization method, testing scenario, infrared prediction model accuracy, and instrument on the accuracy of standardization. This study provides the preliminary knowledge necessary for the establishment of a nationwide MIR database for farm management and breeding programs in China.

MATERIALS AND METHODS

Instruments, Sampling, and Data Collection

Six instruments, routinely employed by 2 DHI centers (respectively in Henan and Shandong Provinces, China), were used in this study (Table 1). The instruments from the DHI center in Henan (Zhengzhou City) were 2 CombiFoss 7 DC (Foss, Hillerød, Denmark), whereas the instruments from the DHI center in Shandong (Jinan City) were one Foss FT+, one Foss FT6000, and 2 FTS (Bentley, Chaska, MN). Because several prediction models have been recently developed using one of the 2 FTS instruments (Zhao et al., 2023), this was defined as the primary instrument (B-SD-1, Table 1). The wavenumber ranges of the Foss and Bentley instruments were 925.66 to 5,010.15 cm^{-1} (1,060 wavenumbers) and 649.03 to 3,998.59 cm^{-1} (899 wavenumbers), respectively, and the default output was expressed in transmittance and absorbance units, respectively.

Milk samples (approximately 240 mL each) to be used for the development of standardization models were all collected on the same day in March 2023, using Afimilk's milk equipment (Afimilk Ltd., Israel), from a total of 30 Holstein cows in one commercial farm in Henan. To maximize spectra variation among samples, cows were selected based on milk fat (2.48%–6.74%) and protein content (3.01%–4.31%), parity (1–5), and DIM (46–433

Table 1. Characteristics of the instruments used in the study

Laboratory	Instrument brand and model	Spectral range (cm^{-1})	Acronym	Instrument type in standardization
DHI center Henan	Foss, CombiFoss 7 DC	925.66–5,010.15	F-HN-1	Secondary
DHI center Henan	Foss, CombiFoss 7 DC	925.66–5,010.15	F-HN-2	Secondary
DHI center Shandong	Foss, FT+	925.66–5,010.15	F-SD-3	Secondary
DHI center Shandong	Foss, FT6000	925.66–5,010.15	F-SD-4	Secondary
DHI center Shandong	Bentely, FTS	649.03–3,998.59	B-SD-1	Primary
DHI center Shandong	Bentley, FTS	649.03–3,998.59	B-SD-2	Secondary

d; Supplemental Table S1; see Notes). Each sample was thoroughly mixed and split into 6 aliquots, one for each instrument. To prevent the milk from spoilage, 2 tablets of preservative (bronopol; 2-bromo-2-nitropropan-1, 3-diol; 0.028 g per tablet) were added to each aliquot, which was gently shaken for uniform mixing. Two milk aliquots were tested at the Henan laboratory, within 24 h of sample collection, and the other 4 aliquots were transferred at 4°C to the Shandong laboratory and tested within 36 h of sample collection. As in previous analyses spectra from technical replicates on the same instrument exhibited minimal variation; spectra acquisition was performed once for each sample. Hence, a total of 30 spectra were obtained from each instrument.

Standardization models were then validated on the same batch of spectra used for the development of the standardization models (internal validation) and on spectra collected after 7 mo, in October 2023 (external validation), from 5 of the original 6 instruments, using milk samples from the same farm, but from different cows (Supplemental Table S1).

Spectra Standardization Methods

Before standardization, the wavenumbers from instruments of different brands were harmonized. First, transmittance values from Foss instruments were transformed into absorbance as $\log_{10}(1/\text{transmittance})$, and then a linear interpolation was performed to align the spectral range of the different brands. The absorbance of Foss instruments was projected onto the range of the Bentley instrument (i.e., the brand of the primary instrument). After the linear interpolation, the number of wavenumbers of all spectra was 824, ranging from 928.77 to 3,998.59 cm^{-1}. Spectra water absorption regions (1,600.18–1,689.70 cm^{-1} and 3,010.13–3,998.59 cm^{-1}) were removed before the analysis (Karoui et al., 2010; Grelet et al., 2021). All MIR spectra were then centered.

Two standardization methods (PDS and DS) were compared. In PDS, each wavenumber j of the primary instrument spectra (**P** matrix) is reconstructed based on measurements from a small window (s, length = n) of the secondary instrument spectra (**S** matrix). For each wavenumber of the primary instrument (P_j), a partial least square (**PLS**) multivariate regression model is built assuming a linear relationship between **P** and **S**, with the following equation:

$$P_j = sb_j + e_j,$$

where b_j is the regression coefficient and e_j is the intercept term for the jth wavenumber. The calculation of these 2 standardization parameters with a moving spectral window leads to 2 banded diagonal matrixes (**B** and **E**, for the regression coefficients and intercept terms, respectively; Bouveresse and Massart, 1996). Finally, the standardized matrix of the secondary instrument ($\mathbf{S_{std}}$) can be obtained as follows:

$$\mathbf{S_{std}} = \mathbf{SB} + \mathbf{E}.$$

The optimal parameters (number of PLS components and window length) in PDS were defined as those maximizing the standardization accuracy based on the determination coefficient and root mean square error ($\mathbf{R^2_{Pred}}$ and $\mathbf{RMSE_{Pred}}$, respectively) between predictions obtained from each combination of primary and secondary instruments. The number of PLS components tested in PDS ranged from 1 to 29 (equal to the number of milk samples − 1) and the window length n ranged from 1 to 30 wavenumbers in the 2 scenarios, respectively. It is worth noting that an "edge effect" can occur at the highest and lowest wavenumbers because insufficient data are used to form a complete window in PDS. In this study, asymmetric modeling was used on those edge wavenumbers.

In DS, the complete spectrum of a secondary instrument undergoes transformation to mirror the spectra of the primary instrument, based on the linear relationship observed in spectra collected by different instruments. And each wavenumbers' absorbance of the secondary instrument was calculated using the whole primary spectra, as described in the following equation:

$$\mathbf{P} = \mathbf{SF},$$

where **F** is the transformation matrix. The secondary instrument (**S**) is standardized using the response matrix for the primary instrument as follows:

$$\mathbf{S_{std}} = \mathbf{SS}^{-1}\mathbf{P}.$$

A set of standardization matrixes (**B** and **E** for PDS, and **F** for DS) were calculated for each combination of primary and secondary instrument, obtaining a total of 5 sets of parameters for each standardization method.

Testing Scenarios for Spectra Standardization

The accuracy of the standardization methods was tested under 2 scenarios. In the first scenario, referred to as internal validation, all the available spectra collected in March 2023 were used to obtain the standardization matrixes. Then, the matrixes were applied to the spectra, allowing for the evaluation of the accuracy of the standardization model but not its robustness. In the second scenario, referred to as external validation, standardization parameters were obtained from internal validation and applied to the spectra collected in October 2023. The external-validation scenario can be used to simulate the accuracy of a standardization method on unknown samples, thus enabling the evaluation of the robustness of the standardization method.

Infrared Prediction Models

To test the effectiveness of the standardization process across infrared prediction models, 4 MIR prediction models with different accuracies were developed on the primary instrument and applied to all secondary instruments' MIR spectra, before and after standardization. Descriptive statistics of the reference values used for model development and accuracies of prediction models are reported in Supplemental Table S2 (see Notes). To obtain models with high prediction accuracy, a total of 5,277 records of milk fat content (**FP**, %) and protein content (**PP**, %) were predicted from the corresponding spectra and FP and PP MIR predictions, respectively, obtained from the primary instrument (B-SD-1) in March 2023. Previously developed prediction models for C8:0 (g/L) and C10:0 (g/L) were selected as the low-accuracy models. The development of the prediction models was described in Lou et al. (2024). Briefly, a milk sample was collected from 155 cows (6–305 d after calving) in July 2020. The content of C8:0 and C10:0 in milk was assessed by ultrahigh performance liquid chromatography – high resolution MS (Chen et al., 2021), and the corresponding MIR spectra were obtained from the primary instrument. Spectra water absorbance regions were removed before the analysis and, after interpolation, the spectra included the regions from 928.77 to 1,596.45 cm^{-1}, and from 1,693.43 to 3,006.40 cm^{-1} (533 variables). Spectra variables were then centered.

The models to predict FP, PP, C8:0, and C10:0 were developed using PLS regression and a 10-fold random cross-validation procedure, repeated 10 times. The optimal number of PLS components was chosen to maximize the determination coefficient ($\mathbf{R^2_{CV}}$) and minimize the root mean square error ($\mathbf{RMSE_{CV}}$) in cross-validation. The mean R^2_{CV} was 0.9997 ± 0.00001 for FP, 0.9976 ± 0.00002 for PP, 0.6630 ± 0.0192 for C8:0, and 0.5569 ± 0.0139 for C10:0, with values of $RMSE_{CV}$ equal to 0.0126 ± 0.0002, 0.0113 ± 0.00005, 6.9714 ± 0.2223, and 21.8815 ± 0.4059, respectively (see Supplemental Table S2).

Evaluation of the Accuracy of the Standardization Methods

The determination coefficient between the absorbance obtained at each wavenumber from each combination of primary and secondary instrument ($\mathbf{R^2_{WAV}}$) was calculated before and after standardization to investigate how the variation across instruments affects individual wavenumbers. The accuracy of PDS and DS standardization models was evaluated by calculating R^2_{Pred} and $RMSE_{Pred}$ between predictions of FP, PP, C8:0, and C10:0 obtained from each combination of primary and secondary instruments. A Hotelling-Williams *t*-test (Dunn and Clark, 1971; Steiger, 1980) and ANOVA (Cederkvist et al., 2005) were used respectively to determine the difference in R^2_{Pred} and absolute prediction errors obtained for the different standardization methods, testing scenarios, infrared prediction models, and instruments. The PROC MIXED procedure implemented in the SAS software (ver. 9.20, SAS Institute Cary, NC) was performed on the absolute prediction error (δ_{ijklmn}) using the following model:

$$\delta_{ijklmn} = \mu + M_i + T_j + Eq_k + Inst_l + Cow_m + \varepsilon_{ijklmn},$$

where μ was the overall mean; M_i was the fixed effects of the standardization method (i = 3 levels: nonstandardization, PDS, and DS); T_j was the fixed effects of the testing scenario (j = 2 levels: internal and external validation); Eq_k was the fixed effect of the prediction model (k = 4 levels: FP, PP, C8:0, and C10:0); $Inst_l$ was the fixed effect of the secondary instrument (l = 5 levels: F-HN-1, F-HN-2, F-SD-3, F-SD-4, and B-SD-2); Cow_m was the random effect of the milk sample (m = 60 levels); and ε_{ijklmn} was the residual error. It was assumed that $Cow_m \sim N\left(0, \mathbf{I}\sigma^2_a\right)$ and $\varepsilon_{ijklmn} \sim N\left(0, \mathbf{I}\sigma^2_e\right)$, where σ^2_a is the variance of the random effect and σ^2_e is the residual variance. All calculations were carried out using programs developed from MATLAB 2016a (https://www.mathworks.com), R soft-

ware (ver. 4.10; https://www.R-project.org/), and SAS (ver. 9.20; SAS Institute, Cary, NC).

RESULTS AND DISCUSSION

Instrument Response Before Standardization

To visualize the effects of standardization on the MIR spectra obtained by the secondary instrument, the R^2_{WAV} calculated between the primary and each of the secondary instruments were compared before and after standardization. As shown in Figure 1a, the average R^2_{WAV} across wavenumbers in the secondary instruments ranged from 0.7283 (F-HN-2) to 0.9051 (B-SD-1), with values varying across spectral regions. For example, the absorbance in the spectral region between 2,293.97 and 2,734.11 cm^{-1} was markedly affected by instrumental variation ($R^2_{WAV} < 0.85$). Because reproducibility between machines is particularly low for some regions, we can hypothesize that the signal in these regions might be more susceptible to random environmental perturbations. In this regard, a R^2_{WAV} close to 0 was found

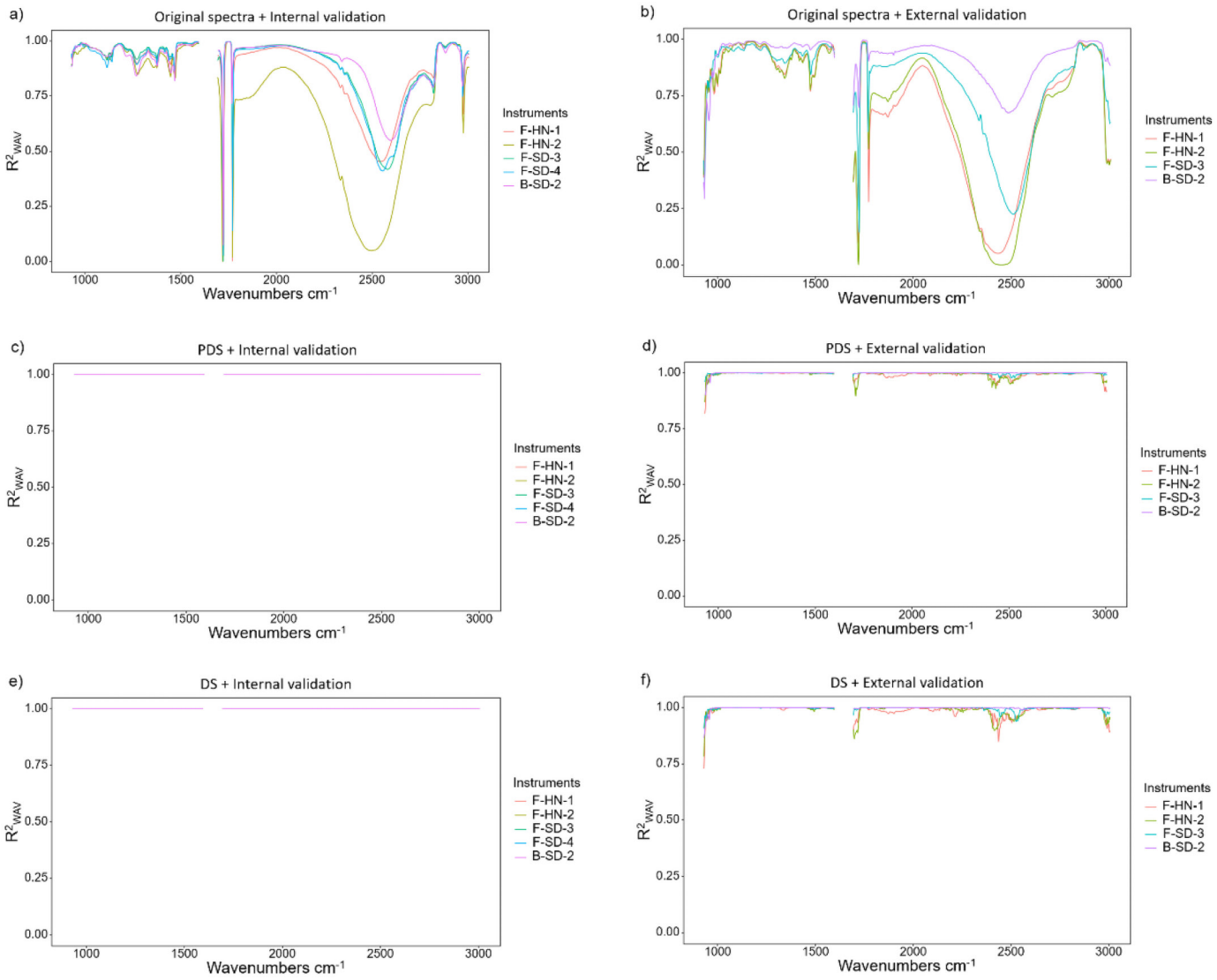

Figure 1. Coefficient of determination (R^2_{WAV}) of the absorbance obtained between the primary and each secondary instrument before (plot a and b) and after standardization (c–f). The correlation between the signal provided by the primary and secondary instruments increased after standardization, depending on the standardization method (PDS and DS) and scenario (internal and external validation). PDS = piecewise direct standardization. DS = direct standardization. Internal validation: standardization was tested on the spectra used for development of standardization models. External validation: standardization was tested on spectra collected 7 mo after the development of standardization models.

in the water absorbance regions (1,600.18 to 1,689.70 and 3,010.13 to 3,998.59 cm^{-1}) of the spectra (results not shown), consistent with the findings of Grelet et al. (2021). These results indicate that a prediction model developed excluding the noise regions or regions that exhibit a low reproducibility among spectrometers may be better suited for transfer to a new instrument than a model built using the whole spectrum. Results also indicate that the lack of standardization may affect some prediction models more than others, depending on which spectral regions are the most influential for the predictions.

The comparison of R^2_{WAV} also indicated that the secondary Foss instruments deviated from the primary instrument to a greater extent, whereas the same brand's instrument used as the primary instrument (B-SD-2) provided consistent responses (Figure 1a). This seems to indicate that standardization is particularly important among instruments of different brands. Variability in the spectral signal between milk laboratories was also observed. This might be due to the different instrument models, but also to the different time lag between milk collection and analysis and laboratory conditions (i.e., heating of samples, process, and workload) in Henan and Shandong laboratories. In China, all DHI laboratories adjust the infrared predictions (e.g., milk fat and protein content) from instruments of different brands (e.g., Foss and Bentley) monthly, using standard milk samples, but ignore spectra variations. Spectra are typically preprocessed before analyses to correct for baseline and scaling effects from light scattering (Tiplady et al., 2020), which helps adjust short-term variations on instruments, but more is needed to correct for long-term or multi-instrument variations in the spectral signal.

Accuracy of MIR Prediction Models Before Standardization

Table 2 shows the R^2_{Pred} and $RMSE_{Pred}$ between predictions of the same milk samples provided by the primary and secondary instruments for FP, PP, C8:0, and C10:0 prediction models before standardization. Due to the differences in instruments' response, the R^2_{Pred} between predictions from primary and secondary instruments was less than 1.00, and the $RMSE_{Pred}$ varied across prediction modes, instruments, and validation scenarios. On unstandardized spectra, the highly accurate FP and PP models seem to perform relatively well (R^2_{Pred} ranged between 0.9571 and 0.9965; $RMSE_{Pred}$ ranged between 0.065% and 0.505%), even though their accuracy was below ICAR recommendations (ICAR, 2023), and they performed better than prediction models for C8:0 and C10:0 (R^2_{Pred} ranged from 0.7064 to 0.9451; $RMSE_{Pred}$ ranged between 13.92 g/L and 57.61 g/L), as expected. Results indicate that variations in instrumental response might affect the prediction of the most common milk components (e.g., fat, protein, and lactose content), but are more likely to reduce considerably the consistency between predictions provided by different instruments for fine milk compositions or functional traits, because values of R^2_{CV} of those prediction models are frequently lower than 0.8 (De Marchi et al., 2014). The accuracy of these models relies highly on the correlations between the traits to be predicted and the concentration of other milk components or other "easy-to-measure" traits (Grelet et al., 2021), and often requires suitable regression methods, large samples, and accurate golden reference values (De Marchi et al., 2014; Grelet et al., 2021). Most of these models are still "immature" and have not yet

Table 2. Coefficient of determination (R^2_{Pred}) and root mean square error ($RMSE_{Pred}$, %) between predictions obtained from the primary instrument and each of the secondary instruments before standardization for FP, PP, C8:0, and C10:0 for the spectra used in internal and external validations

	FP (%)		PP (%)		C8:0 (g/L)		C10:0 (g/L)	
Secondary instrument	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$
Internal validation[1]								
F-HN-1	0.9924	0.4163	0.9575	0.3134	0.9191	21.2984	0.7966	55.5428
F-HN-2	0.9928	0.3566	0.9680	0.3316	0.9113	28.6549	0.7960	47.7732
F-SD-3	0.9902	0.3694	0.9707	0.3294	0.9451	16.3846	0.8213	56.7872
F-SD-4	0.9939	0.3062	0.9653	0.2980	0.8886	16.9127	0.8000	50.1634
B-SD-2	0.9965	0.0667	0.9768	0.0648	0.7846	13.9158	0.8128	21.7054
External validation[2]								
F-HN-1	0.9715	0.5045	0.9577	0.3257	0.8661	26.3714	0.7142	57.6060
F-HN-2	0.9730	0.4139	0.9571	0.4184	0.8973	31.8572	0.7064	50.4844
F-SD-3	0.9835	0.3912	0.9697	0.3958	0.8181	19.8612	0.7225	51.8017
B-SD-2	0.9913	0.1077	0.9668	0.1408	0.8277	15.2051	0.7622	25.6731

[1]Internal validation: spectra used for development of standardization models.

[2]External validation: spectra collected 7 mo after the development of standardization models.

reached good accuracy in prediction, and large prediction errors are expected when transferred to new, unstandardized instruments. Hence, increasing the model accuracy as much as possible (e.g., increasing the reference sample size, or maximizing the variation in the calibration set) before sharing the model within a network of instruments is recommended.

Furthermore, Wang and Bovenhuis (2019) reported that modeling based on the unstandardized spectra (e.g., using cumulative historical spectra) may yield an overoptimistic accuracy, because physical factors, including instrument wear and variations in the milk testing environment, lead to differences in the response among spectrometers over time. Such differences might be confounded with the sample batch (e.g., when samples from different farms are collected on different days) and lead to an overestimation of MIR prediction accuracy. Wang and Bovenhuis (2019) demonstrated the existence of the issue and proposed to estimate the prediction accuracy of water regions in MIR spectra as a negative control. If the prediction accuracy of water regions is far from zero, standardization is required to limit the impact of the instrumental variation.

Instrument Response After Standardization

The optimal number of PLS components and window length were chosen to maximize R^2_{Pred} and $RMSE_{Pred}$ between the primary and secondary instruments. In the internal-validation scenario, the optimal number of PLS components for PDS was 3, whereas the window length had no clear effect on the subsequent results and was set to 5. The same values were used in the external validation. After standardization, the distance among spectra collected on the same milk sample, based on the increase in R^2_{WAV} (Figure 2c–f), was significantly reduced. The average R^2_{WAV} across secondary instruments ranged from 0.9872 to 0.9934 in external validation, and approached 1.00 in internal validation. Although PDS and DS led to enhanced R^2_{WAV} in spectra between secondary and primary instruments during validations, DS exhibited poorer performance in external validation compared with its performance in internal validation and to PDS under both validation scenarios. This might indicate that DS is more prone to overfitting than PDS.

Accuracy of MIR Prediction Models After Standardization

Tables 3 and 4 show the R^2_{Pred} and $RMSE_{Pred}$ obtained for FP, PP, C8:0, and C10:0 after PDS and DS for the 2 testing scenarios. The R^2_{Pred} increased after PDS and DS, with a simultaneous decrease of the $RMSE_{Pred}$, with variations that were more evident in internal validation. Piecewise direct standardization performed slightly worse than DS, but even in the models with the lowest prediction accuracy (i.e., the C8:0 and C10:0 models) the R^2_{Pred} still remained above 0.99 and $RMSE_{Pred}$ were lower than 0.3 g/L (Table 3). Conversely, DS performed worse than PDS in external validation (Table 4), but the performance was better than nonstandardization (Table 2). For example, the R^2_{Pred} and $RMSE_{Pred}$ for FP predictions in the F-HN-1 instrument was 0.9715 and 0.5045%, respectively, before standardization, 0.9801 and 0.1819% after PDS, and 0.9757 and 0.2383% after DS in external validation.

This study is the first to compare the results from PDS and DS applied to milk MIR spectra. Results indicate

Table 3. Coefficient of determination (R^2_{Pred}) and root mean square error ($RMSE_{Pred}$, %) between predictions obtained from the primary instrument and each of the secondary instruments, after standardization, for prediction models for FP, PP, C8:0, and C10:0 in internal validation

Secondary instrument and standardization method[1]	Fat content (%)		Protein content (%)		C8:0 (g/L)		C10:0 (g/L)	
	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$
F-HN-1								
PDS	0.9999	0.0466	0.9999	0.0416	0.9999	0.0457	0.9999	0.2071
DS	0.9999	0.0025	0.9999	0.0019	0.9999	0.0005	0.9999	0.0037
F-HN-2								
PDS	0.9999	0.0387	0.9999	0.0399	0.9999	0.0184	0.9999	0.1836
DS	0.9999	0.0026	0.9999	0.0020	0.9999	0.0005	0.9999	0.0037
F-SD-3								
PDS	0.9999	0.0382	0.9999	0.0368	0.9999	0.0118	0.9999	0.1315
DS	0.9999	0.0026	0.9999	0.0020	0.9999	0.0005	0.9999	0.0035
F-SD-4								
PDS	0.9999	0.0185	0.9999	0.0154	0.9999	0.0308	0.9999	0.1892
DS	0.9999	0.0026	0.9999	0.0019	0.9999	0.0005	0.9999	0.0037
B-SD-2								
PDS	0.9999	0.0028	0.9999	0.0074	0.9999	0.0009	0.9999	0.0094
DS	0.9999	0.0026	0.9999	0.0019	0.9999	0.0005	0.9999	0.0036

[1]Internal validation: spectra used for development of standardization models.

Table 4. Coefficient of determination (R^2_{Pred}) and root mean square error ($RMSE_{Pred}$, %) between predictions obtained from the primary instrument and each of the secondary instruments, after standardization, for prediction models for FP, PP, C8:0, and C10:0 in external validation

Secondary instrument and standardization method[1]	Fat content (%)		Protein content (%)		C8:0 (g/L)		C10:0 (g/L)	
	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$	R^2_{Pred}	$RMSE_{Pred}$
F-HN-1								
PDS	0.9801	0.1819	0.9736	0.1449	0.9568	3.3346	0.9561	3.1719
DS	0.9757	0.2383	0.9666	0.2970	0.9452	4.9062	0.9418	4.6233
F-HN-2								
PDS	0.9788	0.1910	0.9737	0.1540	0.9475	4.0378	0.9618	3.0685
DS	0.9770	0.2087	0.9717	0.3249	0.9230	4.8975	0.9398	3.9511
F-SD-3								
PDS	0.9899	0.1949	0.9899	0.1687	0.9393	3.1612	0.9404	3.1450
DS	0.9865	0.2144	0.9752	0.3439	0.9271	5.7315	0.9293	5.5161
B-SD-2								
PDS	0.9948	0.0562	0.9890	0.0791	0.9607	2.7544	0.9636	2.0584
DS	0.9924	0.0851	0.9876	0.0927	0.9510	3.9285	0.9583	4.5973

[1]External validation: spectra collected 7 mo after the development of standardization models.

that both methods can effectively reduce the transfer errors. However, overfitting can possibly affect model robustness when the number of standardization samples is much smaller than the number of wavenumbers to be standardized, especially because DS assumes that there is correspondence between the absorbance at one specific wavenumber on the primary and secondary instruments. This might explain the inferior performance of DS compared with PDS in external validation. Overfitting can be avoided by adopting PDS, which uses a multivariate analysis to estimate the relationship between each wavenumber on the primary instrument and a corresponding limited number of variables (i.e., the window) on the secondary instrument. The advantages of PDS over DS can be attributed to its local character and multivariate nature, enabling simultaneous correction of intensity differences, wavelength shifts, and peak broadening (Feudale et al., 2002). However, the optimal window size and number of PLS components must be identified in advance and derived from the extent of the residuals between the unstandardized and standardized spectra. Generally, DS is more effective when extensive standardization spectra are available to estimate **F**, and PDS is more appropriate for small datasets (Morais et al., 2019).

As expected, standardization models performed consistently better in internal validation than in external validation, with R^2_{Pred} values around 0.9999 and $RMSE_{Pred}$ ranging between 0% and 0.05% for FP and PP and 0 to 0.30 g/L for C8:0 and C10:0. However, external validation is important to ensure robust prediction for routine use. Furthermore, the performances of the standardization models after 7 mo led to relatively high $RMSE_{Pred}$ for FP and PP, above the threshold recommended by ICAR (2023). This indicates that instrument standardization should be repeated more often, that is, testing every 2 mo (Grelet, et al., 2015), to obtain stable effects. Small variations in temperature, complexity of the sample composition, and instrumental disturbances can jeopardize standardization stability over time (Zeaiter et al., 2004). Hence, standardization should be regularly checked and adjusted to maintain its robustness during process monitoring, for example by carrying out ring tests once or twice a month and repeating the standardization using a small number of milk samples across instruments.

The accuracy of the FP and PP prediction after standardization improved: in the internal-validation scenario, the average R^2_{Pred} increased from 0.9932 (Table 2) to 0.9999 (Table 3) for FP, and from 0.9677 (Table 2) to 0.9999 (Table 3) for PP, whereas the average $RMSE_{Pred}$ decreased from 0.303% (Table 2) to 0.016% (Table 3) for FP and from 0.267% (Table 2) to 0.015% (Table 3) for PP across the secondary instruments. It is noteworthy that both the FP and PP model accuracy after standardization falls within the threshold recommended by ICAR (2023). This suggests that high-accuracy models can sustain their performance through standardization. Furthermore, standardization greatly improved the accuracy of C8:0 predictions in secondary instruments. In internal validation, the average R^2_{Pred} increased from 0.8897 (Table 2) to 0.9999 (Table 3) for C8:0 and from 0.8053 (Table 2) to 0.9999 (Table 3) for C10:0, and the average $RMSE_{Pred}$ decreased from 19.433 g/L (Table 2) to 0.011 g/L (Table 3) for C8:0 and from 46.394 g/L (Table 2) to 0.074 g/L (Table 3) for C10:0 across the secondary instruments. Hence, the model with low accuracy benefited the most from spectra standardization. In particular, for spectra collected 7 mo after standardization, after PDS the RMSE among predictions obtained by different machines decreased on average by 86% and 94% compared with the values before standardization, for C8:0 and C10:0 respectively.

Results were confirmed by the Hotelling-Williams *t*-test and ANOVA, which indicated that the standardiza-

Table 5. Hotelling-Williams *t*-test and ANOVA for the difference, respectively, in the coefficient of determination (R^2_{Pred}) and absolute prediction errors between predictions obtained from the primary instrument and each of the secondary instruments[1]

Effect	R^2_{Pred}		Absolute error, *P*-value
	t-value	*P*-value	
Standardization method[2]			
PDS vs. DS	0.0599	>0.05	>0.05
PDS vs. nonstandardization	7.1299	<0.01	<0.01
DS vs. nonstandardization	5.5334	<0.05	<0.05
Application scenarios			
Internal vs. external validation[3]	29.2090	<0.01	<0.01
Prediction model			
Fat content vs. protein content	9.2425	<0.05	<0.05
Fat content vs. C8:0	27.5190	<0.01	<0.01
Fat content vs. C10:0	36.4990	<0.01	<0.01
Protein content vs. C8:0	19.2310	<0.01	<0.01
Protein content vs. C10:0	30.195	<0.01	<0.01
C8:0 vs. C10:0	5.0099	<0.05	<0.05
Secondary instruments	0.0015	>0.05	>0.05

[1]Differences across standardization methods, testing scenarios, and infrared prediction models.

[2]DS = direct standardization; PDS = piecewise direct standardization.

[3]Internal validation: standardization was tested on the spectra used for development of standardization models. External validation: standardization was tested on spectra collected 7 mo after the development of standardization models.

tion methods, testing scenarios, and infrared prediction model accuracy significantly affected standardization performances ($P < 0.05$; Table 5). Although the secondary instrument had no significant effects on R^2_{Pred} ($P > 0.05$), the lower $RMSE_{Pred}$ values were observed between the primary instrument (Bentley) and the secondary instrument B-SD-2, of the same brand before and after standardization, with all the investigated traits. For example, $RMSE_{Pred}$ for FP was 4.5 times lower than the one generated by the other secondary instruments; the $RMSE_{Pred}$ of the FP model for B-SD-2 after PDS was lower (0.0486%) than for other instruments (0.0610% to 0.0873%). Despite the very high R^2_{Pred} and low $RMSE_{Pred}$ of the prediction model with high accuracy, standardization is still expected to improve model transferability in unfamiliar instruments.

Limitation and Prospects

Overall, this study is the first to investigate the effects of standardization methods, testing scenarios, and accuracies of the infrared prediction models on the standardization accuracy of milk MIR spectra. These findings have important implications for developing and deploying prediction models across diverse scenarios and instruments. They highlight the benefits of standardization techniques, and underscore the critical importance of model accuracy. These insights can guide researchers in optimizing prediction models for various real-world applications, ensuring robust and accurate performance across different contexts. However, more instruments should be incorporated to enable an accurate estimate of the effect of the brand and model. In addition, the stability of the standardization models over time (i.e., how often standardization should be repeated) or the effectiveness of other standardization methods (e.g., positive matrix factorization and artificial neural networks drift correction; Paatero and Tapper, 1994; Goodacre et al., 1997) deserve exploration. This would provide a solid foundation to develop and maintain a national MIR spectra database over the years, and for the in-depth investigation of applications in dairy cow phenotyping and breeding.

CONCLUSIONS

The variation in the spectral signal provided by different instruments was successfully reduced by standardization across 2 provinces in China. Standardization accuracy was significantly affected by the standardization method, testing scenario, and accuracy of the infrared prediction model to be transferred across instruments. This study lays the foundations for developing a national MIR spectra database to provide consistent predictions across provinces to be used in dairy farm management and breeding programs in China. Furthermore, it provides opportunities for data exchange and cooperation at international levels.

NOTES

This study was supported by the National Natural Science Foundation of China (No. 32150610478, Beijing, China); the Key Research Project of Henan Province (No. 221111111100, Zhengzhou, China); the Special Fund for Henan Agriculture Research System (HARS-22-14-S, Zhengzhou, China); Shandong Provincial Natural Science Foundation (ZR2021MC070 & ZR2021QC209, Jinan, Shandong, China). China Scholarship Council (No. 201913043, Beijing, China) and Hainan University (Haikou, China) are acknowledged for funding the PhD study of Wenqi Lou. Supplemental material for this article is available at https://doi.org/10.6084/m9.figshare.25048385.v1 (Table S1) and https://doi.org/10.6084/m9.figshare.25048397.v1 (Table S2). Milk sample collection at the farms was conducted under the Dairy Herd Improvement program; therefore, no ethical approval (Institutional Animal Care and Use Committee) was required for this study. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: DS = direct standardization; FP = milk fat content; MIR = mid-infrared; PDS = piecewise direct standardization; PLS = partial

least square; PP = milk protein content; R^2_{CV} = determination coefficient of cross-validation; R^2_{Pred} = determination coefficient of predictions; R^2_{WAV} = determination coefficient at each wavenumber; $RMSE_{CV}$ = root mean square error of cross-validation; $RMSE_{Pred}$ = root mean square error of predictions.

REFERENCES

Benedet, A., M. Franzoi, M. Penasa, E. Pellattiero, and M. De Marchi. 2019. Prediction of blood metabolites from milk mid-infrared spectra in early-lactation cows. J. Dairy Sci. 102:11298–11307. https://doi.org/10.3168/jds.2019-16937.

Bonfatti, V., A. Fleming, A. Koeck, and F. Miglior. 2017. Standardization of milk infrared spectra for the retroactive application of calibration models. J. Dairy Sci. 100:2032–2041. https://doi.org//10.3168/jds.2016-11837.

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.

Bonfatti, V., S.-A. Turner, B. Kuhn-Sherlock, T. D. W. Luke, P. N. Ho, C. V. C. Phyn, and J. E. Pryce. 2019. Prediction of blood β-hydroxybutyrate content and occurrence of hyperketonemia in early-lactation, pasture-grazed dairy cows using milk infrared spectra. J. Dairy Sci. 102:6466–6476. https://doi.org/10.3168/jds.2018-15988.

Bouveresse, E., C. Casolino, and C. de la Pezuela. 1998. Application of standardisation methods to correct the spectral differences induced by a fibre optic probe used for the near-infrared analysis of pharmaceutical tablets. Presented at the 7th Meeting on Recent Developments in Pharmaceutical Analysis, Island of Elba, Italy, September 16–20, 1997. J. Pharmaceut. Biomed. 18:35–42. https://doi.org/10.1016/S0731-7085(98)00169-1.

Bouveresse, E., and D. L. Massart. 1996. Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. Chemometr. Intell. Lab Syst. 32:201–213. https://doi.org/10.1016/0169-7439(95)00074-7.

Cederkvist, H. R., A H. Aastveit, and T. Næs. 2005. A comparison of methods for testing differences in predictive ability. J. Chemometr. 19:500–509. https://doi.org/10.1002/cem.956.

Chen, L., X. Sun, A. S. Khalsa, M. T. Bailey, K. Kelleher, C. Spees, and J. Zhu. 2021. Accurate and reliable quantitation of short chain fatty acids from human feces by ultra high-performance liquid chromatography-high resolution mass spectrometry (UPLC-HRMS). J. Pharmaceut. Biomed. 200:114066 https://doi.org/10.1016/j.jpba.2021.114066.

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.

Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde, and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal 6:1694–1701. https://doi.org/10.1017/S1751731112000456.

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.

Dunn, O. J., and V. Clark. 1971. Comparison of tests of the equality of dependent correlation coefficients. J. Am. Stat. Assoc. 66:904–908. https://doi.org/10.2307/2284252.

El Jabri, M., M.-P. Sanchez, P. Trossat, C. Laithier, V. Wolf, P. Grosperrin, E. Beuvier, O. Rolet-Répécaud, S. Gavoye, Y. Gaüzère, O. Belysheva, E. Notz, D. Boichard, and A. Delacroix-Buchet. 2019. Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. J. Dairy Sci. 102:6943–6958. https://doi.org/10.3168/jds.2019-16320.

Feudale, R. N., N. A. Woody, H. Tan, A. J. Myles, S. D. Brown, and J. Ferré. 2002. Transfer of multivariate calibration models: A review. Chemometr. Intell. Lab. Syst. 64:181–192. https://doi.org/10.1016/S0169-7439(02)00085-0.

Fleming, A., F. S. Schenkel, J. Chen, F. Malchiodi, R. A. Ali, B. Mallard, M. Sargolzaei, M. Corredig, and F. Miglior. 2017a. Variation in fat globule size in bovine milk and its prediction using mid-infrared spectroscopy. J. Dairy Sci. 100:1640–1649. https://doi.org/10.3168/jds.2016-11427.

Fleming, A., F. S. Schenkel, J. Chen, F. Malchiodi, V. Bonfatti, R. A. Ali, B. Mallard, M. Corredig, and F. Miglior. 2017b. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. J. Dairy Sci. 100:5073–5081. https://doi.org/10.3168/jds.2016-12102.

Frizzarin, M., I. C. Gormley, D. P. Berry, and S. McParland. 2023a. Estimation of body condition score change in dairy cows in a seasonal calving pasture-based system using routinely available milk mid-infrared spectra and machine learning techniques. J. Dairy Sci. 106:4232–4244. https://doi.org/10.3168/jds.2022-22394.

Frizzarin, M., G. Visentin, A. Ferragina, E. Hayes, A. Bevilacqua, B. Dhariyal, K. Domijan, H. Khan, G. Ifrim, T. L. Nguyen, J. Meagher, L. Menchetti, A. Singh, S. Whoriskey, R. Williamson, M. Zappaterra, and A. Casa. 2023b. Classification of cow diet based on milk mid infrared spectra: A data analysis competition at the "International Workshop on Spectroscopy and Chemometrics 2022." Chemometr. Intell. Lab. Syst. 234:104755. https://doi.org/10.1016/j.chemolab.2023.104755.

Goodacre, R., E. M. Timmins, A. Jones, D. B. Kell, J. Maddock, M. L. Heginbothom, and J. T. Magee. 1997. On mass spectrometer instrument standardization and interlaboratory calibration transfer using neural networks. Anal. Chim. Acta 348:511–532. https://doi.org/10.1016/S0003-2670(97)00062-7.

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.

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.

Grelet, C., J. A. F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng. 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci. 100:7910–7921. https://doi.org/10.3168/jds.2017-12720.

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.

ICAR. 2023. Limits of error for milk yield and milk composition. Accessed Mar. 19, 2024. https://www.icar.org/Guidelines/11-Milk-recording-devices-Procedure-05.pdf.

Karoui, R., G. Downey, and C. Blecker. 2010. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure−quality relationships—A review. Chem. Rev. 110:6144–6168. https://doi.org/10.1021/cr100090k.

Lou, W., L. F. Brito, X. Zhao, V. Bonfatti, J. Li, and Y. Wang. 2024. Selection of the most informative wavenumbers to improve prediction accuracy of milk fatty acid profile based on milk mid-infrared spectra data. Animal Research and One Health 10.1002/aro2.72.

Luke, T. D. W., S. Rochfort, W. J. Wales, V. Bonfatti, L. Marett, and J. E. Pryce. 2019. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J. Dairy Sci. 102:1747–1760. https://doi.org/10.3168/jds.2018-15103.

Manuelian, C. L., G. Visentin, C. Boselli, G. Giangolini, M. Cassandro, and M. De Marchi. 2017. Short communication: Prediction of milk coagulation and acidity traits in Mediterranean buffalo milk using Fourier-transform mid-infrared spectroscopy. J. Dairy Sci. 100:7083–7087. https://doi.org/10.3168/jds.2017-12707.

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.

Morais, C. L. M., M. Paraskevaidi, L. Cui, N. J. Fullwood, M. Isabelle, K. M. G. Lima, P. L. Martin-Hirsch, H. Sreedhar, J. Trevisan, M. J. Walsh, D. Zhang, Y. G. Zhu, and F. L. Martin. 2019. Standardization of complex biologically derived spectrochemical datasets. Nat. Protoc. 14:1546–1577. https://doi.org/10.1038/s41596-019-0150-x.

Paatero, P., and U. Tapper. 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5:111–126. https://doi.org/10.1002/env.3170050203.

Shadpour, S., T. C. S. Chud, D. Hailemariam, H. R. Oliveira, G. Plastow, P. Stothard, J. Lassen, R. Baldwin, F. Miglior, C. F. Baes, D. Tulpan, and F. S. Schenkel. 2022. Predicting dry matter intake in Canadian Holstein dairy cattle using milk mid-infrared reflectance spectroscopy and other commonly available predictors via artificial neural networks. J. Dairy Sci. 105:8257–8271. https://doi.org/10.3168/jds.2021-21297.

Shetty, N., G. Difford, J. Lassen, P. Løvendahl, and A. J. Buitenhuis. 2017a. Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk. J. Dairy Sci. 100:9052–9060. https://doi.org/10.3168/jds.2017-13014.

Shetty, N., P. Løvendahl, M. S. Lund, and A. J. Buitenhuis. 2017b. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. J. Dairy Sci. 100:253–264. https://doi.org/10.3168/jds.2016-11609.

Shi, R., W. Lou, B. Ducro, A. van der Linden, H. A. Mulder, S. J. Oosting, S. Li, and Y. Wang. 2023. Predicting nitrogen use efficiency, nitrogen loss and dry matter intake of individual dairy cows in late lactation by including mid-infrared spectra of milk samples. J. Anim. Sci. Biotechnol. 14:8. https://doi.org/10.1186/s40104-022-00802-3.

Smith, S. L., S. J. Denholm, M. P. Coffey, and E. Wall. 2019. Energy profiling of dairy cows from routine milk mid-infrared analysis. J. Dairy Sci. 102:11169–11179. https://doi.org/10.3168/jds.2018-16112.

Soyeurt, H., E. Froidmont, I. Dufrasne, D. Hailemariam, Z. Wang, C. Bertozzi, F. G. Colinet, F. Dehareng, and N. Gengler. 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livest. Sci. 227:82–89. https://doi.org/10.1016/j.livsci.2019.07.007.

Steiger, J. H. 1980. Tests for comparing elements of a correlation matrix. Psychol. Bull. 87:245–251. https://doi.org/10.1037/0033-2909.87.2.245.

Tiplady, K. M., T. J. Lopdell, M. D. Littlejohn, and D. J. Garrick. 2020. The evolving role of Fourier-transform mid-infrared spectroscopy in genetic improvement of dairy cattle. J. Anim. Sci. Biotechnol. 11:39. https://doi.org/10.1186/s40104-020-00445-2.

Tiplady, K. M., R. G. Sherlock, M. D. Littlejohn, J. E. Pryce, S. R. Davis, D. J. Garrick, R. J. Spelman, and B. L. Harris. 2019. Strategies for noise reduction and standardization of milk mid-infrared spectra from dairy cattle. J. Dairy Sci. 102:6357–6372. https://doi.org/10.3168/jds.2018-16144.

Tiplady, K. M., M.-H. Trinh, S. R. Davis, R. G. Sherlock, R. J. Spelman, D. J. Garrick, and B. L. Harris. 2022. Pregnancy status predicted using milk mid-infrared spectra from dairy cattle. J. Dairy Sci. 105:3615–3632. https://doi.org/10.3168/jds.2021-21516.

Wallén, S. E., E. Prestløkken, T. H. E. Meuwissen, S. McParland, and D. P. Berry. 2018. Milk mid-infrared spectral data as a tool to predict feed intake in lactating Norwegian Red dairy cows. J. Dairy Sci. 101:6232–6243. https://doi.org/10.3168/jds.2017-13874.

Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle. J. Dairy Sci. 102:6288–6295. https://doi.org/10.3168/jds.2018-15684.

Wang, Y., D. J. Veltkamp, and B. R. Kowalski. 1991. Multivariate instrument standardization. Anal. Chem. 63:2750–2756. https://doi.org/10.1021/ac00023a016.

Zeaiter, M., J.-M. Roger, V. Bellon-Maurel, and D. N. Rutledge. 2004. Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. Trends Analyt. Chem. 23:157–170. https://doi.org/10.1016/S0165-9936(04)00307-3.

Zhao, X., Y. Song, Y. Zhang, G. Cai, G. Xue, Y. Liu, K. Chen, F. Zhang, K. Wang, M. Zhang, Y. Gao, D. Sun, X. Wang, and J. Li. 2023. Predictions of milk fatty acid contents by mid-infrared spectroscopy in Chinese Holstein cows. Molecules 28:666. https://doi.org/10.3390/molecules28020666.

ORCIDS

W. Lou https://orcid.org/0000-0002-0150-6068
H. Lu https://orcid.org/0000-0003-1155-8942
X. Zhao https://orcid.org/0000-0001-8537-9854
Y. Wang https://orcid.org/0000-0003-3629-2802
V. Bonfatti https://orcid.org/0000-0003-3970-5764