# Daily standardization of milk mid-infrared spectra in a comprehensive regression model framework considering animal related data

*A. Mensching[1], J. Braunleder[1], E. Bohlsen[2], S. Schierenbeck[1] and R. Reents[1]*

*[1]Vereinigte Informationssysteme Tierhaltung w.V.(vit), IT Solutions for Animal Production, Heinrich-Schröder-Weg 1, 27283 Verden, Germany*
*[2]Landeskontrollverband Niedersachsen, Großstr. 30, 26789 Leer, Germany*
*Corresponding Author: andre.mensching@vit.de*

## Abstract

Milk mid-infrared (MIR) spectrometry has been utilized worldwide and for decades to analyze the components of milk. In a routine use, the method demonstrates a very high precision and repeatability, particularly for the main milk components. This is substantially supported by repeated analyses of standard milk samples with known reference values, whereof a slope-intercept (S/I) correction is derived for regular samples. However, this does not apply to routinely collected spectral data, where deviations and drift can be observed both between different instruments and within an instrument over time. The aim of this study was to demonstrate a new approach for standardization of MIR spectra using a framework of regression models considering results of laboratory analyses and information on the animal, such as days in milk (DIM) or parity, to estimate daily and instrument-wise standardization coefficients for the individual wavelengths.

The data were provided by the Landeskontrollverband Niedersachsen (LKV Niedersachsen, Leer, Germany) and included spectral data from 5 spectrometers (FOSS, Hillerød, Denmark) as well as the corresponding data from the dairy herd improvement (DHI) testing, which were routinely collected in the first half of 2022 (dataset I, DS-I). In addition to the total of 2.3 M spectra from routine DHI testing, triple analyses of the same 5.3 k DHI milk samples on 3 of the 5 spectrometers were carried out on samples of 7 different farms during the same period (DS-II). Furthermore, 61.0 k spectra of standard milk samples were available (DS-III). In the daily laboratory routine, these samples of the weekly changing North German Standard Milk (NGSM) were analyzed 3 times in a row every 200 regular samples (reference analysis by LUFA Nord-West, Oldenburg, Germany).

In a first step, the DHI spectra of DS-I were used to quantify and eliminate day-specific instrument effects in a complex framework of regression models, considering information on the animal as well as data obtained from the lab analyses to finally estimate instrument-, day- and wavelength-wise standardization coefficients. With the aim of demonstrating the effect of standardization, the dataset with triple analyses (DS-II) was utilized in a second step to develop calibration models for both raw and the standardized spectra using the S/I-corrected fat values obtained from the laboratory. Based on this, two separate analyses were performed: first, the dataset with the triply analyzed DHI samples (DS-II) was used for principal component analyses (PCA) and a comparison of the estimability of fat in each case for raw and standardized spectra. Second, the dataset with the analyses of standard milk samples (DS-III) was used to compare the S/I-corrected laboratory fat values with estimates from the developed

fat models based on both raw and standardized spectra across all 5 instruments and over time.

It could be shown that the standardization led to a harmonization of the spectra between instruments as well as over time and thus corrected both general and temporary instrument effects. In addition, the estimability of milk fat, which was used as an example trait for the validation of the methodology, was optimized by the standardization of the MIR spectra. Results showed that the root mean squared error (RMSE) in a leave-one-instrument-out cross-validation (LOIO-CV) could be reduced from 0.110 to 0.032% fat during calibration and from 0.045 to 0.020% fat during validation. Regarding the standard milk analyses, the RMSE was also reduced from 0.038 to 0.019% fat and thus closely approximates the RMSE of 0.014% fat of the S/I-corrected laboratory values. This study showed that there is not only a high demand for standardization across instruments, but also within instruments over time. Therefore, vit-standardization as a statistical procedure featuring a daily standardization seems to be a promising novel approach for future estimation of traits that are not covered by standard milk samples.

*Keywords: standardization, milk mid-infrared spectrometry, regression model framework.*
*Presented at the ICAR Anual Conference 2024 in Bled at the Session 2: New tools to extend the horizon of milk mid-infrared spectrometry*

## Introduction

For decades, Fourier transform MIR spectrometry of milk has been used globally to determine milk components. Its accuracy for payment-relevant main milk components is essentially attributable to the check sample tests carried out regularly in the laboratory routine (Gengler *et al*., 2016; Nieuwoudt *et al*., 2021). By comparing target and actual values of standard milk samples with known composition, these tests are used to derive a S/I-correction for regular milk samples. The requirement for S/I corrections is well known, as recorded spectra the same sample are neither comparable between instruments nor over time (Young, 1978; Wang *et al*., 1991). In addition to fundamental differences between various manufacturers, there can be differences even within the same manufacturer between different instruments of the same model generation (Grelet *et al*., 2015; Nieuwoudt *et al*., 2021). Apart from this, environmental effects such as temperature and humidity, mechanical wear, sample handling and constructional differences of the instruments as well as mechanical and electronic effects, can result in spectral deviations or drift over time (Wang *et al*., 1991; Nieuwoudt *et al*., 2021).

General and temporary instrument effects are a particular hindering when estimating traits beyond those known from the standard milk samples, as no simple S/I correction is possible. To handle this, there are various approaches that focus on the standardization of MIR spectra. In theory, this shall allow the transfer of models between different instruments to obtain reliable estimates over time. The manufacturer FOSS, for example, has been using its own patented process for instrumental standardization, which should be carried out regularly, e.g. monthly, by analyzing a chemical liquid, the so-called ‘equalizer’, for which a reference spectrum is known (FOSS, 2014). In general, piecewise direct standardization (PDS) can be considered the gold standard of non-instrumental standardization methods (Wang *et al*., 1991). In this process, spectra from secondary instruments (“slaves”) are translated into spectra from primary instruments (“masters”) with the aid of statistical models. However, PDS requires the analysis of common milk samples. An example of the use of PDS is the standardization service offered by European Milk Recording (EMR, 2024, Ciney, Belgium), which is based on the procedure described by Grelet *et al*. (2015, 2017). In the EMR service, common standardization samples are analyzed in repetition on all instruments approximately every month. Based on the corresponding recorded spectra, standardization coefficients are derived in a centralized procedure, that can be applied

on regular milk samples. An alternative standardization method that does not require analysis of common samples is the procedure described by Bonfatti *et al*. (2017). In this so-called retroactive standardization of spectra, temporary homogeneous subsets are identified first. Using a PCA-based method, transformations are then determined which translate the various subsets in the form of “slave” datasets into a “master” by means of an S/I-based correction of the absorbance values.

It is important to emphasize that particularly short-term instrument effects can occur during spectrometric measurements. In this regard, Nieuwoudt *et al*. (2021) showed in their method for weekly monitoring of spectral data that there is a great need for continuous monitoring, as conspicuous drifts or deviations can already be detected between weekdays.

The objective of this study was to demonstrate a novel approach for the daily standardization of milk MIR spectra, in which general and temporal instrument effects are estimated and eliminated using a regression model framework to determine instrument-, day- and wavenumber-specific standardization coefficients.

## Material and method

All the data was provided by the LKV Niedersachsen (Leer, Germany) and can be differentiated in 3 sub-datasets. The first dataset (DS-I) included data of 2.3 M samples from routine DHI testing in the period from 01/01/2022 to 31/06/2022. The related milk samples were preserved with Bronopol (Georg Hansen e.K., Wrestedt, Germany) and were taken according to the ICAR guidelines (ICAR, 2022). The laboratory of the LKV Niedersachsen in Leer was equipped with a total of 5 FOSS instruments, of which 2 were MilcoScanTM 7 RM (A and B) and 3 were MilcoScanTM FT+ instruments (C, D, and E). Besides the MIR spectra, fat, protein, and lactose contents of the milk determined by the laboratory were available. These MIR spectra-based values were determined by the manufacturer’s equations but were adjusted in a further step as part of the laboratory routine using analyses of NGSM and a thereof derived S/I correction. Furthermore, animal identification and the affiliation to the farm, the date of calving and therefore the DIM, parity, the information about the milk testing scheme including milking time as well as the milk performance were provided. The second data (DS-II) set also contained DHI data, but the associated samples were analyzed sequentially on 3 different instruments (A, B, and C). The 5.3 k samples were taken from 7 different farms whose DHI tests covered the same period from 01/01/2022 to 31/06/2022. The third dataset (DS-III) consisted of 61.0 k records from routine check sample tests of the weekly changed NGSM as well as their laboratory reference measurements analyzed, which were analyzed 3 times in a row every 200 regular samples. The reference analysis for the main milk components of the NGSM samples, which were also preserved with Bronopol (Georg Hansen e.K.), was carried out at LUFA Nord-West (Oldenburg, Germany). Depending on the dataset, plausibility checks and outlier removals were carried out at the spectral level with the global H-value (Soyeurt *et al*., 2019), regarding the agreement of the S/I-corrected fat values in the triplicate analyses of the DHI samples or in terms of completeness of the control samples (3 analyses of NGSM per test).

The dataset DS-I was used to determine instrument-specific, daily, and wavenumber-related standardization coefficients of the MIR spectra. To quantify and eliminate both general and temporal instrument effects from raw spectra, a regression model framework was developed to obtain the standardized spectra. In this context, wavenumber-wise regression models were used to differentiate the observed variance of MIR spectral absorbance values into fixed effects associated with the milk sample and random effects related to the respective instruments. For the milk sample-associated effects, S/I-corrected laboratory values and information on the sample origin, such as

DIM and parity of the cow, were considered. In a further step, the absorbance values corrected for instrument effects were regressed on the raw absorbance values by simple linear regression to obtain slope and intercept correction factors as standardization coefficients. This was done by day, instrument, and wavenumber and thus allowed a translation of raw spectra into the standardized spectra.

Furthermore, fat models were developed using dataset DS-II for both raw and standardized spectra to evaluate the effect of spectral standardization. In each case, partial least squares (PLS) regression models were used, considering n = 6 latent variables. The 1st gap derivative according to Soyeurt *et al*. (2011) was used as spectral pretreatment. In addition, the spectral ranges were reduced to 516 of 1,060 WN according to the selection by Grelet *et al*. (2015). The dataset was split so that 80% of the data was used for calibration and 20% for validation.

In the first part of analysis, the data of repeated analyzed DHI samples (DS-II) were used to investigate the effects of standardization. On spectral level, PCA was used similar as in the work of Grelet *et al*. (2017). Separate PCA were performed both for untreated spectra and for spectra pretreated with a 1st gap derivation. The scores of the standardized spectra were projected into the same vector space as that spanned by the raw spectra. For comparison, the results were graphically displayed with score plots. To evaluate the estimability of fat, common statistics such as the RMSE at calibration, LOIO-CV and validation were determined for both raw and standardized spectra-based fat calibration models. For graphical investigation, the observed S/I-corrected laboratory fat values were compared with the MIR-based estimates in scatter plots. In the second part of the analysis, the dataset DS-III of analyzed check samples (NGSM) was utilized. For visualization, aggregated mean values of the differences between fat reference values of the NGSM and the S/I-corrected laboratory fat values as well as the fat estimates based on raw and standardized spectra were calculated at the daily level and displayed individually for each instrument over time using line plots.

## Results and discussion

PCA was utilized to assess the impact of standardization on MIR spectra. Figures 1A and C display variations in the spectra data across different instruments, notably in the 1st and 2nd principal components (PC) for non-derived raw spectra and particularly in the 5th and 6th PC of gap-derived raw spectra. In contrast, standardized spectra (Figures 1B and 1D) exhibit greater alignment between instrument scores, indicating a successful harmonization through standardization. These results are therefore basically comparable with those of Grelet *et al*. (2017). Instrument similarities are notably higher between instruments A and B (both MilcoScan™ 7 RM, diamond cuvette) compared to instrument C (MilcoScanTM FT+, $CaF_2$ cuvette), possibly due to the different material of the cuvette or other marginal constructional distinctions. An influence of the cuvette material, for example, was also described by Nieuwoudt *et al*. (2021). A closer look at the data revealed that the separate point clouds of the analyses with instrument B (see Figure 1C) were each carried out on the same days and thus reveal temporary instrument effects on the spectra. These were significantly minimized by standardization, as can be seen in Figure 1D. Thus, these results underline the effectiveness of standardization for harmonization of spectra between different instruments and over time.

The calibration and validation metrics determined during model development as well as the comparison of the generated estimates with the S/I-corrected laboratory fat values were used to evaluate the effect of standardization on the estimability of fat. In Figure 2, the S/I-corrected laboratory values are plotted against the estimates from the calibration models based on raw (Figure 2A) and standardized spectra (Figure 2B). In general, the values lie close to the identity in both cases, but the estimates based

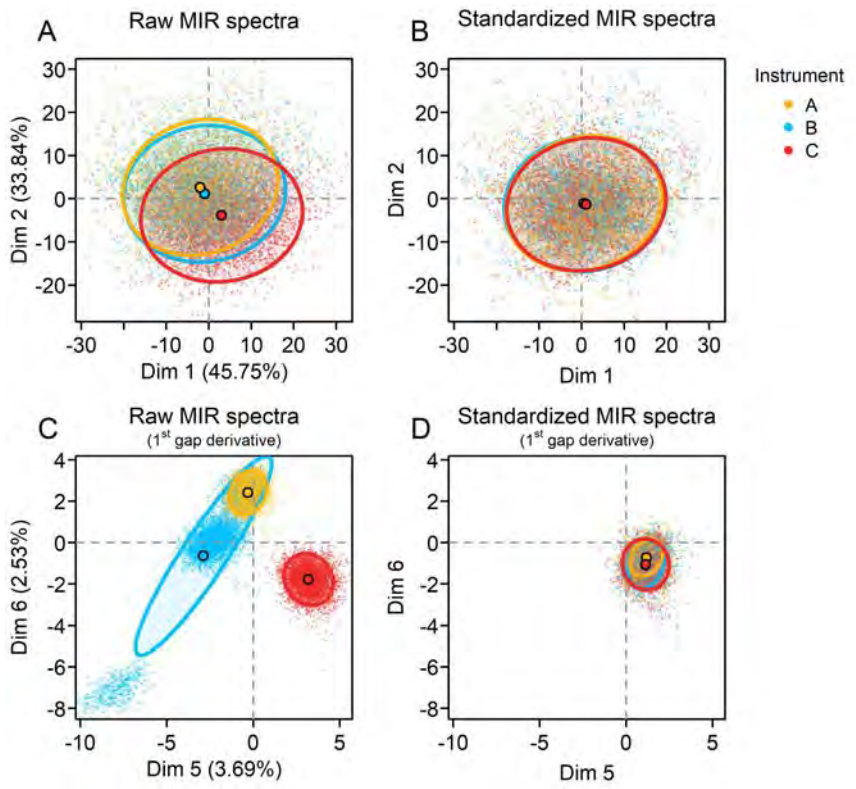


*Figure 1. Results of principal component analyses of the spectra of triple analyzed milk samples before (Figures 1A and 1C) and after (Figures 1B and 1D) standardization (n = 4,471 per instrument) on 212 of 1060 wavenumbers. The instrument-wise points correspond to the means in the respective principal components, where the respective ellipses represent the 95% two-dimensional density contour. Dim = Dimension, MIR = mid-infrared.*

on the raw spectra reveal larger scatter and deviations. As in the previous part of the analysis, instrument B in particular shows considerable outliers in the estimated values based on raw spectra, but not in the estimated values based on standardized spectra. Thus, the standardization shows a substantial contribution to the reduction of overall and temporary instrument effects for the MIR-based fat estimates, which is also reflected in a significant reduction of the RMSE values. The RMSE of LOIO-CV during calibration decreased from 0.110 to 0.032% fat and from 0.045 to 0.020% fat at validation. These results can thus be compared with those of Grelet *et al*. (2015), who also observed a notable reduction in the RMSE in pre- and post-standardization comparisons of fat.

Differences between reference values from the NGSM samples to S/I-corrected lab fat values as well as fat estimates based on raw and standardized spectra were aggregated per instrument into daily mean values and plotted over time in Figure 3. Minimal scatter can be observed for S/I-corrected lab fat values, while strong short-term instrument effects with temporal daily mean biases up to 0.2% fat occur in estimates based on raw spectra. The differences between the reference values and the fat estimates based on standardized spectra scatter only slightly larger around 0 than of the S/I-corrected fat values but much smaller than for the estimates based on non-standardized spectra. This is also confirmed by the calculated mean RMSE values. The standardization led

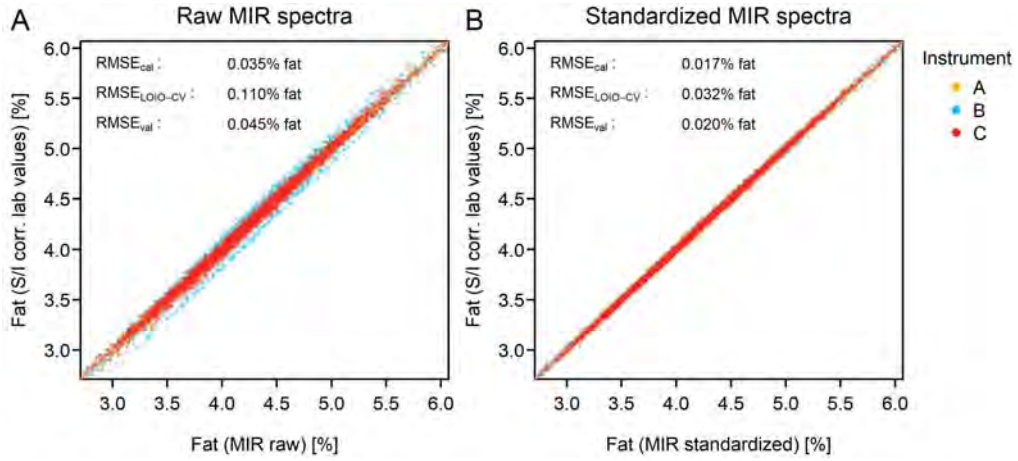


*Figure 2. Comparison of the slope-intercept corrected laboratory fat values of triple analyzed milk samples to the fat estimates based on raw (Figure 2A) and standardized (Figure 2B) spectra (n = 4,471 per instrument). MIR = mid-infrared, $RMSE_{val}$, $RMSE_{LOIO-CV}$, and $RMSE_{val}$ = root mean squared error of calibration, leave-one-instrument-out cross-validation and validation, S/I = slope-intercept.*

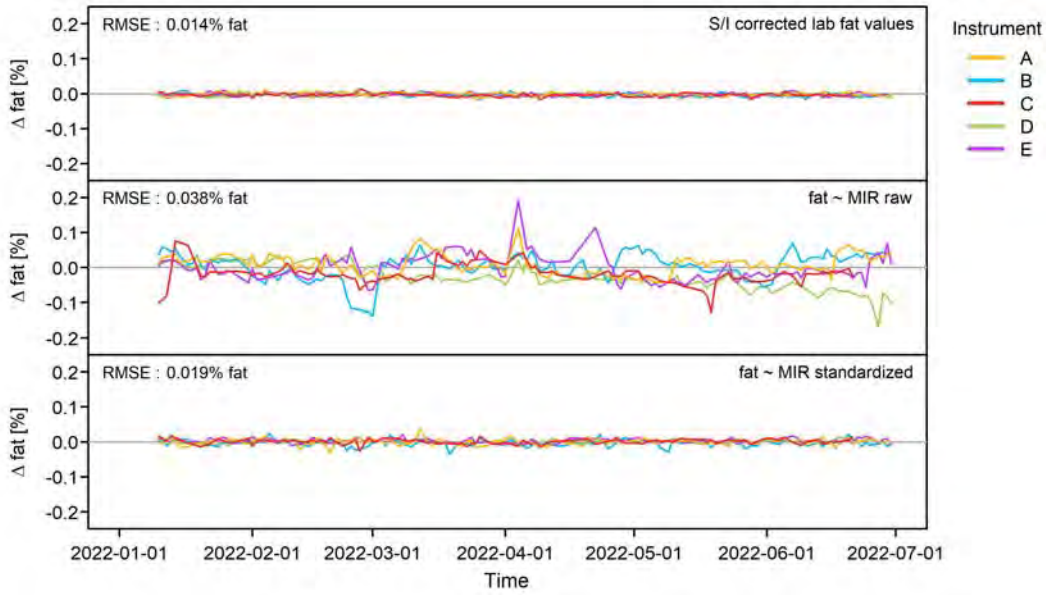


*Figure 3. Daily mean differences between reference values of the North German Standard Milk samples to the slope-intercept corrected laboratory fat values as well as the fat estimates based on raw and standardized mid-infrared spectra (n = 53.5 k). lab = Laboratory, MIR = mid-infrared, RMSE = root mean squared error.*

to a reduction from 0.038 to 0.019% fat and thus almost corresponds to the RMSE of 0.014% fat of S/I-corrected laboratory fat values. As the NGSM samples neither were used for standardization nor for model development, this part of the analysis served as external validation both for the models and for the standardization approach. The estimated standardization coefficients based on DHI samples thus also show a potential for application to bulk milk samples analyzed with the same instrument and on the same day.

## Conclusions and outlook

MIR spectrometric measurements are often influenced by general and temporal effects. This is why there is a high need for regular standardization to reduce drifts in both spectra and estimations of traits, that are not covered by standard milks and therefore cannot be adjusted via S/I correction. The presented procedure for daily standardization based on a regression model framework showed the ability to harmonize MIR spectra both across instruments and over time. This was confirmed by an improved estimability of milk fat, that was used as an example trait to validate the methodology. The underlying standardization method therefore has great potential to generate reliable MIR-based predictions of further phenotypes in the future, both to promote the development of herd monitoring tools for feeding and animal health and to serve as a data source for genomic evaluations.

The work presented here can be seen as the first part of a proof of concept for the “vit- standardization”. In the meantime, the procedure has been upgraded so that standardization can be carried out not only on a closed dataset but also with daily new incoming data. Furthermore, a first MIR-based tool for monitoring of ketosis on routine DHI data has been released for over 3000 farms of the LKV Niedersachsen.

## List of references

**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. http://dx.doi.org/10.3168/jds.2016-11837.

**EMR**, 2024. European milk recording. Accessed April 11, 2024. https://www.milkrecording.eu/emr.site/.

**FOSS**, 2014. Standardization of FT-IR instruments 1–7.

**Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont and P. Dardenne,** 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. http://dx.doi.org/10.3168/jds.2015-10140.

**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. http://dx.doi.org/10.3168/jds.2014-8764.

**Grelet, C., J.A. Fernández 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. http://dx.doi.org/10.3168/jds.2017-12720.

**ICAR, 2022. Section 2 - Guidelines for Dairy Cattle Milk Recording**. Accessed March 26, 2024. https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf.

**Nieuwoudt, M.K., C. Giglio, F. Marini, G. Scott and S.E. Holroyd**, 2021. Routine Monitoring of Instrument Stability in a Milk Testing Laboratory With ASCA: A Pilot Study. Front. Chem. 9:1–13. http://dx.doi.org/10.3389/fchem.2021.733331.

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

**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. http://dx.doi.org/10.1016/j.livsci.2019.07.007.

**Wang, Y., D.J. Veltkamp and B.R. Kowalski,** 1991. Multivariate Instrument Standardization. Anal. Chem. 63:2750–2756. http://dx.doi.org/10.1021/ac00023a016.

**Young, R.S**, 1978. Calibration and Standardization of the Infrared Milk Analyzer. The California Experience. J. Dairy Sci. 61:1279–1283. http://dx.doi.org/10.3168/jds.S0022-0302(78)83718-7.