J. Dairy Sci. 89:2817–2832

Calibration of Infrared Milk Analyzers: Modified Milk Versus Producer Milk[1]

K. E. Kaylegian,* G. E. Houghton,† J. M. Lynch,* J. R. Fleming,‡ and D. M. Barbano*[2]
*Northeast Dairy Foods Research Center Department of Food Science, Cornell University Ithaca, NY 14853
†Kestrel Software Consulting, Berkshire, NY 13736
‡USDA, Agricultural Marketing Service, Texas Milk Marketing Area, P. O. Box 110939 Carrollton 75011

ABSTRACT

Mid-infrared (MIR) milk analyzers are traditionally calibrated using sets of preserved raw individual producer milk samples. The goal of this study was to determine if the use of sets of preserved pasteurized modified milks improved calibration performance of MIR milk analyzers compared with calibration sets of producer milks. The preserved pasteurized modified milk sets exhibited more consistent day-to-day and set-to-set calibration slope and intercept values for all components compared with the preserved raw producer milk calibration sets. Pasteurized modified milk calibration samples achieved smaller confidence interval (CI) around the regression line (i.e., calibration uncertainty). Use of modified milk calibration sets with a larger component range, more even distribution of component concentrations within the ranges, and the lower correlation of fat and protein concentrations than producer milk calibration sets produced a smaller 95% CI for the regression line due to the elimination of moderate and high leverage samples. The CI for the producer calibration sets were about 2 to 12 times greater than the CI for the modified milk calibration sets, depending on the component. Modified milk calibration samples have the potential to produce MIR milk analyzer calibrations that will perform better in validation checks than producer milk-based calibrations by reducing the mean difference and standard deviation of the difference between instrument values and reference chemistry.
Key words: infrared milk analyzer, calibration, modified milk

INTRODUCTION

Mid-infrared (**MIR**) milk analysis using the classical fat B, fat A, protein, and lactose measurement wavelengths combined with separate reference wavelengths is a method that provides rapid determination of milk composition (AOAC, 2000; method 972.16; 33.2.31; IDF, 2000). Analysis of milk by MIR is based on the principle that different functional groups absorb MIR energy at different wavelengths. The principles underlying the MIR analysis of milk are presented elsewhere (Biggs et al., 1987; Biggs and McKenna, 1989). Milk analysis by MIR is an indirect method, so instruments must be calibrated using milk samples with reference values established by reference methods.

Accuracy of MIR milk analysis is affected by instrumental factors such as signal to noise ratio, repeatability, linearity (Smith et al., 1993b), gain, homogenization efficiency (Biggs et al., 1987; Smith et al., 1993a, 1995), purging efficiency, and intercorrection response (Biggs et al., 1987; Barbano and Clark, 1989), and analytical factors such as the uncertainty of chemical reference values (AOAC, 2000). In addition, individual milk sample composition factors such as variation in fatty-acid chain length and degree of unsaturation (Biggs and McKenna, 1989), variation in NPN as a percentage of total nitrogen (Biggs et al., 1987; Barbano and Lynch, 1992), citrate and free fatty acid content (Biggs et al., 1987) will also influence testing accuracy. Instrumental factors encompass mechanical and electronic aspects that are kept within operational tolerances by regular precalibration of the instrument (Barbano and Clark, 1989). Analytical factors include well-defined and performance-validated reference methods (AOAC, 2000) for measurement of fat by ether extraction (method 989.05; 33.2.26), protein by Kjeldahl [method 991.22; 33.2.13 (true protein) or 991.20; 22.2.11 (total N)], lactose by enzymatic method (method 984.15; 33.2.24), and TS by oven drying (method 990.20; 33.2.44). The variation in the results that can be expected for these chemical reference methods is given in the method validation statistics within each method for within-lab repeatability (s_r) and between-lab reproducibility (s_R) (Lynch, 1998; AOAC, 2000; Appendix D). Finally, deterioration of preserved, refrigerated calibration samples during storage due to lipolysis and proteolysis may cause infra-

Received October 2, 2005.
Accepted January 4, 2006.
[1]Use of names, names of ingredients, and identification of specific models of equipment is for scientific clarity and does not constitute any endorsement of product by authors, Cornell University, or the Northeast Dairy Foods Research Center.
[2]Corresponding author: dmb37@cornell.edu

red uncorrected readings to change, resulting in incorrect calibration adjustments.

Once the precalibration performance of the instrument has been controlled, there are 2 fundamentally different calibration approaches in filter-based MIR milk analyses (Barbano and Clark, 1989). The first approach uses previously determined fixed intercorrection factors (Barbano and Clark, 1989) and a secondary slope and intercept calculation by performing a linear regression of reference chemistry for each milk component as a function of instrument intercorrected values for each milk component. The second approach uses a multiple linear regression of uncorrected instrument values for each component to determine all intercorrection, slope, and intercept values based on the set of calibration samples. There are advantages and disadvantages to each approach (Barbano and Clark, 1989). A properly controlled fixed intercorrection approach has been recommended for best accuracy for raw milk testing (Lynch et al., 1995) because calibration sets with a narrow range of concentration of components, nonuniform distribution of concentrations within the range, and positive correlation between fat and protein content produce conditions where individual samples have too much influence on the determination of slope and intercept by regression analysis (Cook, 1977; Cook and Weisberg, 1980). In the present study the fixed intercorrection approach was used.

Characteristics of the calibration sample set that affect the calibration include the number of samples, the range of component concentration and distribution within the range, natural correlation of fat and protein concentrations, and changes in these characteristics from set to set. In general, increasing the number of samples in a calibration set (i.e., number of points in the linear regression analysis) has the potential to reduce both the overall width of the 95% confidence interval (CI) around the regression line and the shape of the CI (i.e., width at the midpoint vs. the ends of the calibration range). The width of the CI at the midpoint decreases geometrically with increasing number of samples with most of the reduction being achieved with a set containing from 12 to 16 samples. The ratio of the width at the ends of the concentration range to the midpoint is influenced by the number and distribution of samples within the range. A uniform distribution of individual samples across the concentration range minimizes the influence of single samples (i.e., leverage). Correlation between change among components (e.g., fat and protein) can cause errors in slope and intercept determination particularly when intercorrection factors are not set correctly or when there is residual nonlinearity in the uncorrected signals. Having an orthogonal matrix of fat, protein, and lactose concentrations within the calibration samples would be ideal. Typically sets of >8 individual producer milk samples are used for calibration of infrared milk analyzers (AOAC, 2000; method 972.16; 33.2.31; IDF, 2000). However, a set with 10 to 14 samples with an orthogonal matrix of component concentration would be preferable (IDF, 2000), but the improvement that could be gained by this approach has not been quantified. One approach to achieve an orthogonal calibration set is to manufacture calibration samples using combinations of pasteurized cream, UF skim milk retentate, and permeate, as outlined by the International Dairy Federation Standard 141C (2000). A modification of this approach was used in the current study.

The objectives of the current study were to determine if the use of pasteurized preserved modified milk calibration samples could reduce the width of the 95% confidence interval (CI) of the calibration linear regression for each component measured using filter-based MIR milk analyzers and improve the consistency of regression slope and intercept between calibration sets compared with the current industry practice of calibration with preserved raw milk individual producer calibration samples.

MATERIALS AND METHODS

Experimental Design

Survey of Producer Milk Calibration Samples. Four laboratories (USDA Federal Milk Market and commercial) provided information about the sample sets that they distributed to other laboratories for IR calibration over a 2-yr period. Each laboratory submitted a description of how they prepared their calibration samples and their reference chemistry values for each set. The purpose of this survey was to determine the characteristics of typical producer milk calibration sets currently being made and to select a single source of producer milk calibration samples for use in Experiments 1 and 2.

Experiment 1. The purpose of this experiment was to compare the characteristics of the calibrations derived using either producer or modified milk calibration samples. A single source of producer milk calibration samples was selected for use based on survey results. The modified milk calibration samples were manufactured at Cornell University, by the method described later in this paper.

Calibrations were performed twice each week over a 102-d period using a single MIR milk analyzer. On each calibration day, separate calibration equations were derived in duplicate for both producer and modified milk calibration samples. The modified and producer milk calibration sets had shelf-lives of 4 and 2 wk, respec-

Table 1. Mid-infrared milk analysis schedule of calibration sets in Experiment 1

		Calibration set	
Study week	Day of analysis	Producer milk	Modified milk
1	1	1	1
	4	1	1
2	8	1	1
	11	1	1
3	15	2	1
	18	2	1
4	22	2	1
	25	2	1
5	29	3	2
	32	3	2
6	37	3	2
	40	3	2
7	45	4	2
	48	4	2
8	53	4	2
	56	4	2
9	60	5	3
	63	5	3
10	68	5	3
	71	5	3
11	76	6	3
	79	6	3
12	84	6	3
	87	6	3
13	92	7	4
	95	7	4
14	99	7	4
	102	7	4

Table 2. Mid-infrared milk analysis schedule of calibration sets in Experiment 2

		Calibration set	
Study week	Day of analysis	Producer milk	Modified milk
1	1	1	1
	4	1	1
2	8	1	1
	11	1	1
3	15		1
	18		1
4	22		1
	25		1
5	29	2	2
	32	2	2
6	37	2	2
	40	2	2
7	45		2
	48		2
8	53		2
	56		2
9	60	3	3
	63	3	3
10	68	3	3
	71	3	3
11	76		3
	79		3
12	84		3
	87		3

tively. Four consecutive sets (i.e., a new set every 4 wk) of modified milk calibration samples and 7 consecutive sets of producer milk samples (i.e., a new set every 2 wk) were run over the experimental period as shown in Table 1. The linear regression slope and intercept values, 95% CI for the linear regressions, and the leverage values for each sample and component in each calibration set were determined for producer and modified milk calibration sets.

Experiment 2. Based on results observed in Experiment 1, the modified milk calibration sample set was redesigned to increase the fat concentration range and make the matrix of component concentrations more orthogonal to determine if performance improvements could be achieved in the modified milk calibration set. The experimental design was similar to Experiment 1 except that the study duration was shorter (87 d) and only 3 modified milk and producer calibration sets were run. The timing was such that the 3 modified milk calibration sets were run consecutively (i.e., a new set every 4 wk) and a producer calibration set was run for first 2 wk at the start of each new modified milk calibration set (Table 2). Thus, no producer sets were run during the last 2 wk of shelf life of each of the modified milk calibration sets. This was done so that both sets (modified and producer) were approximately the same age at the time of analysis. The linear regression slope and intercept values, 95% CI for the linear regressions, and the leverage values for each sample and component in each calibration set were determined for producer and modified milk calibration sets.

Producer Calibration Samples Used in Experiments 1 and 2

Producer calibration samples (12 samples per set) used in both experiments of this study were obtained from a USDA Federal Milk Market laboratory that normally assembles and distributes calibration samples for IR milk analyzers. This laboratory was selected, based on the survey of 4 laboratories, as the source for the producer calibration samples in our experiments because the samples from this laboratory consistently had a wider component range and a better distribution of samples within the component ranges than the other laboratories. Samples were split into 90-mL vials (Capitol Vial, Auburn, AL), preserved with potassium dichromate, and refrigerated (4°C).

Modified Milk Calibration Samples Used in Experiments 1 and 2

Modified milk calibration samples were produced every 4 wk at the pilot plant facilities at Cornell Univer-

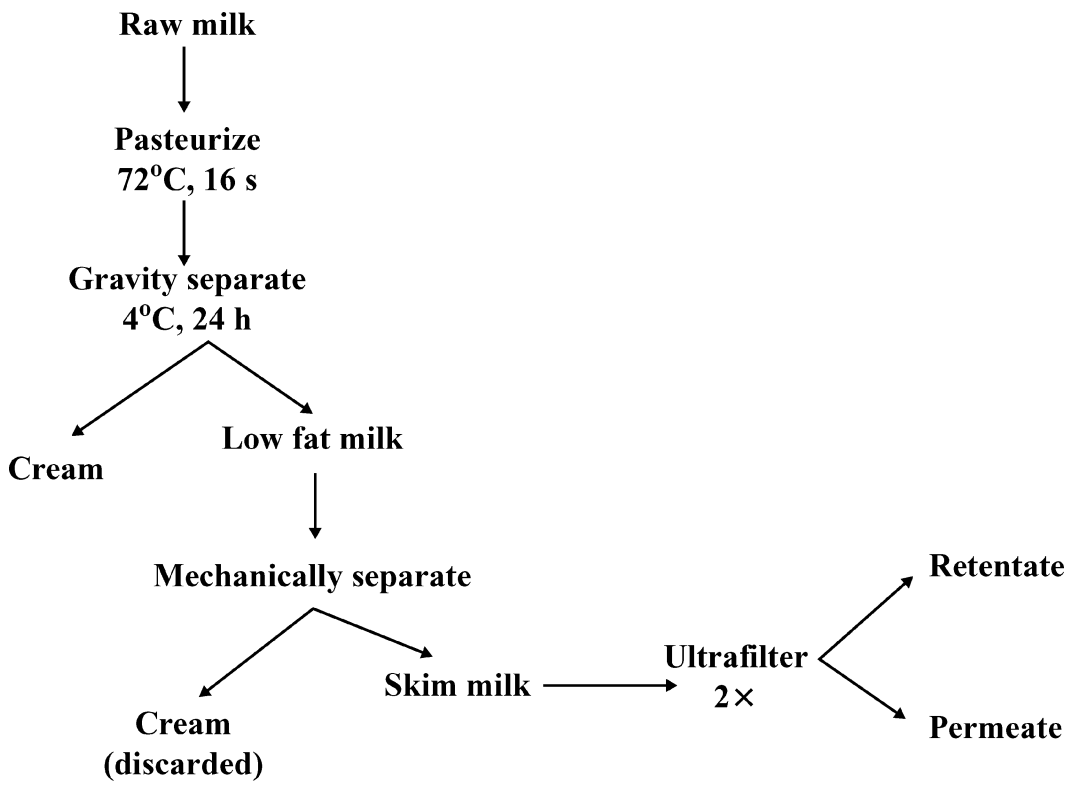

Figure 1. Flow chart of milk processing steps for manufacture of modified milk calibration samples. The gravity-separated cream, the 2× UF retentate, and the UF permeate are the end products that are used in combinations with water and lactose to formulate the modified milk calibration samples.

sity using a variation of the protocol outlined by the International Dairy Federation (2000). The variation was the addition of anhydrous lactose and water to extend the lactose range of the calibration sample sets.

A flow chart of the production of the fresh milk ingredients used to manufacture the modified milk calibration samples is shown in Figure 1. On d 1, raw whole milk (470 kg) was obtained from the Cornell University farm and pasteurized (72°C, 16 s, plate heat exchanger system). The pasteurized milk was poured into 1 plastic (120-kg capacity, 86 cm × 44 cm) and 2 stainless steel (200-kg capacity, 102 cm × 56 cm) cone-bottom tanks, and left in a cold room (4°C) overnight (about 22 h) for gravity separation. On d 2, the gravity skim phase (approximately 425 kg total) was drained through the bottom valve of each tank into milk cans. The gravity skim milk (about 2.0% fat) was immediately heated to 50°C with a plate heat exchanger, and separated using centrifugal separation (model 619, De Laval, Poughkeepsie, NY) to reduce the fat content to about 0.07%. The centrifugally separated skim milk was transferred to the feed tank of a UF system, and maintained at 50°C. The skim milk was ultrafiltered to a 2× concentration using a Dorr-Oliver Series S plate and frame system (Stamford, CT), with an inlet pressure of 310 KPa, outlet pressure of 124 KPa, and total running time of about 3 h. The permeate and retentate were transferred to milk cans in ice and cooled to 4°C.

The gravity cream phase was drained from the bottom of the tanks in layers (total cream about 45 kg). The first layer was 1 L and the remaining layers were approximately 2 L each. The cream layers were subsampled and diluted to <6.0% fat by weight with UF permeate to reduce viscosity for IR analysis of fat content. The UF permeate used for cream dilution was saved from a previous production of modified milk samples (held frozen at –40°C and thawed in a 4°C cooler). The fat content for each diluted cream sample and the initial pasteurized milk was determined by IR analysis. The IR analyzer was calibrated using a previous batch of modified milk samples. The fat contents for the undiluted cream layers were calculated. The bottom cream layer was about 13% fat and the top layer was about 34% fat. Cream layers were chosen for blending so that the final cream ingredient had a fat content in the range of 22 to 27%. The selected cream layers were poured into a milk can and mixed. The protein and lactose content of the cream ingredient was estimated by calcu-

lation using the IR values for protein and lactose content of the initial milk and the determined fat content from the cream ingredient.

The formulations for the modified milk calibration samples were determined using an Excel (Microsoft Corp., Seattle, WA) linear solver function. The solver program used the composition of the ingredients and the desired minimum and maximum range values for fat, true protein, and anhydrous lactose as standard parameters. Design points were chosen to create sample formulations that met the target composition. Distilled water and α-lactose monohydrate (MultiPharm, EM Science, Gibbstown, NJ) were used in some samples to increase the range of lactose concentrations.

Ingredients (4°C) were weighed into 20-L plastic containers (model 50812YK, Rubbermaid, Fairlawn, OH). The UF permeate, and water if required, was weighed into the container, and then the lactose powder was added. These ingredients were stirred until the lactose dissolved before the remaining ingredients were added. Each batch of modified milk was preserved by adding an aqueous 6.7% potassium dichromate (ACS grade, Fisher Scientific, East Lawn, NJ) solution at a level of 3 mL per 1,000 g of modified milk (to achieve a final concentration in milk of 0.02% potassium dichromate). Samples were mixed and held overnight at 4°C. On d 3, each batch of modified milk was stirred continuously with a mixer (type RZR 50, Heidolph, Schwabach, Germany) while being pumped (Easy Load II model 77200-62, Masterflex, Cole-Parmer Ind., Vernon Hills, IL) at 900 mL/min into vials (60 mL for Experiment 1, 90 mL for Experiment 2; Capitol Vial), and refrigerated (4°C). The samples were shipped with wet ice by overnight carrier to the laboratories participating in reference chemical analysis. Chemical analysis was started on d 4.

In Experiment 1, the modified milk calibration sets consisted of 12 samples each, with the target sample compositions shown in Table 3. The sample sets were designed to provide a wide range of components that varied independently. In Experiment 2, the modified milk calibration sets consisted of 14 samples each, with the target sample compositions shown in Table 4. Compared with Experiment 1, the number of samples in the set was increased from 12 to 14 to extend the fat range and make the matrix of different component concentrations more orthogonal.

Reference Chemical Analyses

Reference chemistry values for all calibration samples were determined using the following validated methods (AOAC, 2000): fat by modified Mojonnier ether extraction (method 989.05; 33.2.26), true protein by Kjeldahl analysis (method 991.22; 33.2.13), and TS by oven drying (method 990.20; 33.2.44). Lactose was determined either by enzymatic analysis (method 984.15; 33.2.24) or by difference [lactose = TS − (fat + true protein + ash + 0.19)]. Ash was estimated using an updated version of the equation described by Lynch et al. (1990): ash = [(0.0596 × true protein) + 0.5379].

In Experiment 1, reference chemistry for the calibration sets was calculated from the average of duplicate analyses by either 2 laboratories (modified milk calibration set) or 4 laboratories (producer milk calibration set). Lactose was determined by enzymatic analysis for the modified milk calibration sets and by difference for the producer milk calibration sets. In Experiment 2, reference chemistry values for the calibration sets were calculated from the average of duplicate analyses by 7

Table 3. Composition targets for modified milk calibration samples used in Experiment 1

Sample	Fat	True protein	Anhydrous lactose
	(%)		
1	2.00	4.30	4.65
2	2.50	2.28	4.65
3	3.50	2.90	4.65
4	4.00	3.74	4.65
5	4.50	3.46	4.65
6	5.00	2.62	4.65
7	5.50	4.02	4.65
8	6.00	2.06	4.65
9	3.00	3.01	5.30
10	3.00	3.01	4.98
11	3.00	3.01	4.33
12	3.00	3.01	4.00
Mean	3.75	3.12	4.65
Range	4.00	2.24	1.30

Table 4. Composition targets for modified milk calibration samples used in Experiment 2

Sample	Fat	True protein	Anhydrous lactose
	(%)		
1	0.20	4.30	4.00
2	0.62	2.27	4.55
3	1.05	3.96	5.10
4	1.47	2.61	4.96
5	1.89	3.62	4.28
6	2.32	2.95	4.55
7	2.74	3.28	4.55
8	3.16	3.12	4.42
9	3.58	3.45	4.68
10	4.01	2.78	4.14
11	4.43	3.79	4.82
12	4.85	2.44	4.00
13	5.28	4.13	4.55
14	5.70	2.10	5.10
Mean	2.95	3.20	4.55
Range	5.50	2.20	1.10

laboratories for both types of calibration sets, with the exception of lactose. Lactose was determined in duplicate by enzymatic analysis by 4 laboratories for both the modified and producer calibration sets.

MIR Analysis

Instrument Specifications. The MIR analysis was performed with a Milko-Scan 605 (Foss Electric, Hillerød, Denmark) using the following wavelengths: fat B 3.48 μm (3.6 μm reference), fat A 5.723 μm (5.6 μm reference), protein 6.465 μm (6.7 μm reference), and lactose 9.610 μm (7.7 μm reference; van de Voort et al., 1990; Smith et al., 1993a, 1995). Fat content was determined using 100% fat B. The zeroing solution used for analysis was a 0.01% Triton-X-100 solution (Foss Electric). The MIR analyzer was controlled and data were collected using the IR-QC software package that was developed at Cornell University. The calculations of regression slope, intercept, CI, and leverage by the IR-QC software are verified by using a test data set analyzed by both IR-QC and SAS.

Instrument Precalibration. Precalibration procedures were used to ensure that the instrument was performing within the mechanical and electronic tolerances. Precalibration procedures were conducted monthly and included mechanical flow and sample uptake volumes, homogenization efficiency, zero drift, water and milk repeatability, primary slope, and purging efficiency as described by Barbano and Clark (1989). Residual nonlinearity was evaluated as described by Smith et al. (1993b) and intercorrection factors were evaluated as described by Biggs et al. (1987) and IDF (2000) at the start of each experiment and kept constant during each experiment.

The precalibration performance of the instrument was kept within the following parameters on uncorrected data for all components throughout the study: water repeatability at <0.04%, zero shift at <0.02%, residual nonlinearity at <0.02%, primary slope between 0.95 and 1.05, raw and homogenized milk repeatability at <0.04%, and purging efficiency for both water to milk and milk to water of >99% (Barbano and Clark, 1989). The same intercorrection factors were used for both modified milk and producer milk calibrations (Table 5). The intercorrection factors used in Experiment 2 were adjusted to further improve their performance (Table 5).

Calibration. Calibration for this study is defined as the adjustment of secondary slope and intercept using linear regression of chemistry as a function of the intercorrected data for fat B, lactose, protein, and fat A. An example of the equation for fat is: $F_B = S_f[F_{Bu} + P_u(P/F) + L_u(L/F)] + B_f$, where F_B is the corrected fat B value, S_f is the secondary slope for fat B, F_{Bu}, P_u, and L_u are the uncorrected IR signals for fat B, protein, and lactose, respectively, (P/F) and (L/F) are the intercorrection factors for the influence of protein on fat B and lactose on fat B, respectively, and B_f is the intercept (bias) for fat B. Calibration samples were run as follows: the instrument was zeroed and the average of 3 uncorrected readings of zeroing solution was taken to check the initial zero. If the initial zero check was not within ± 0.02% on each component, then the instrument was rezeroed before continuing. Next, 3 readings for each calibration sample were taken. The first reading was discarded to exclude any carryover effect from the previous sample, and the average of the second and third readings was used. After the last calibration sample, 3 pumping cycles of zeroing solution were used to flush the flow system. Next, 3 uncorrected readings of zeroing solution were taken and averaged to check the final zero reading. The final zero reading was subtracted from the initial zero reading to obtain a zero shift. If the zero shift was >0.02 on any component, then the calibration was considered invalid, and the calibration run was repeated because of excessive zero shift of the instrument during the calibration run. Generally, this did not happen.

Table 5. Intercorrection factors used in Experiments 1 and 2

Intercorrected component	Experiment 1	Experiment 2
Lactose on fat B	–0.251	–0.201
Protein on fat B	–0.112	–0.100
Fat B on lactose	0.034	0.039
Protein on lactose	0.033	0.036
Fat B on protein	0.080	0.081
Lactose on protein	0.070	0.046
Lactose on fat A	0.059	0.061
Protein on fat A	0.035	0.055

Evaluation of Calibration Set Performance

Regression 95% CI and High Leverage Samples. The 95% CI for the regression line and high leverage of individual samples within a calibration set are 2 parameters of performance that were determined. These parameters are directly related to the component concentration range and the distribution of individual sample concentrations within the range of a calibration set. The calculation of both parameters was done for fat B, protein, lactose, and fat A for all calibration sets used in Experiments 1 and 2 with the IR-QC software.

The uncertainty of the linear regression was illustrated by a funnel curve that represented the 95% CI for the regression. If the component range was broad, samples were well distributed within the range, and

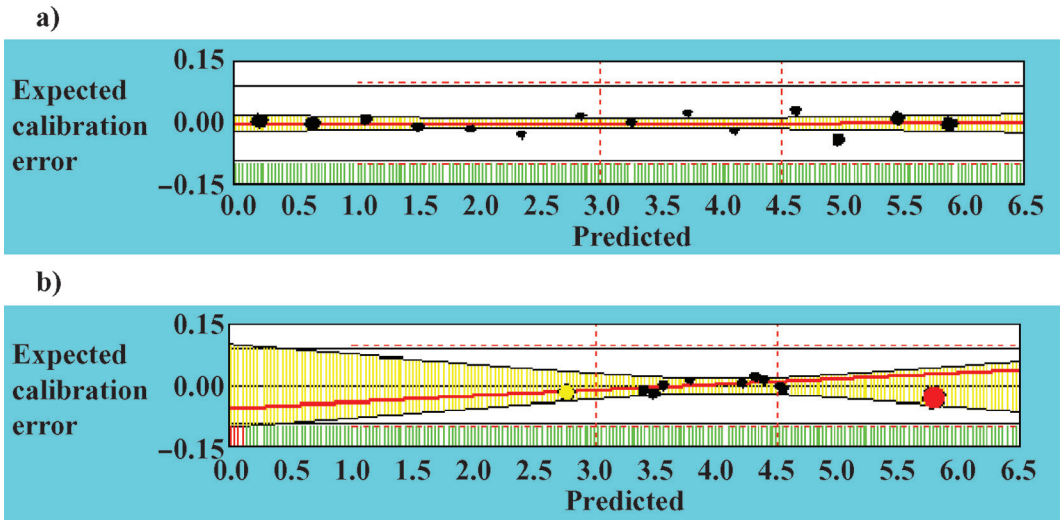

Figure 2. Residual plot of differences between instrument predictions and reference chemistry for Fat B for a) a modified milk calibration set expected calibration error, and b) a producer milk calibration set expected calibration error as a function of the predicted fat concentration.

variability of sample response small, then the 95% CI slope was represented by a narrower funnel curve, illustrated in Figure 2a for a modified milk calibration set. If the component range was narrow or there was an uneven sample distribution, the funnel curve was narrow in the middle and wider at the ends of the range (i.e., more hourglass-shaped), indicating an increased uncertainty at the ends of the chemical distribution range, illustrated in Figure 2b for a producer milk calibration set.

The funnel curve (i.e., CI for predicted values) was calculated as described below. The linear regression $\hat{y} = A + Bx$, where A = intercept and B = slope was calculated. The residual difference for each sample was calculated as for the ith sample, the residual is $r_i = y_i - \hat{y}$; $r_i = y_i - (A + Bx_i)$. From these residuals a variance and standard deviation was computed:

$$\text{Variance of residuals} = Var(r) = \sum (r_i^2)/n - 2)$$

$$\text{Standard deviation of residuals} = S_r = \sqrt{Var(r)}$$

where n = number of samples, y_i = value of y (chemistry) for the ith sample, x_i = value of x (instrument corrected reading) for the ith sample, $\hat{y}_i$ = predicted value of y for the ith sample, and r_i = residual for the ith sample.

Next, let $\bar{x}$ = mean of all x_i values and $\bar{y}$ = the value of the regression equation at the point where it crosses $\bar{x}$; i.e., $\bar{y} = A + B\bar{x}$. From the standard deviation of the residuals (S_r), the variance and standard error of the this mean value were calculated as follows:

$$\text{Variance of the mean predicted value} = Var(\bar{y}) = \frac{S_r^2}{n}$$

$$\text{Standard error of the mean} = S_{\bar{y}} = \sqrt{Var(\bar{y})}$$

From the standard deviation of the residuals, the variance and standard error of the slope were calculated as follows:

$$\text{Variance of the slope} = Var(B) = \frac{S_r}{\sum (x - \bar{x})^2}$$

$$\text{Standard error of the slope} = S_B = \sqrt{Var(B)}$$

The regression line predicts a value of y for any value of x. The uncertainty in this prediction results from combined uncertainty of the mean and the slope. In fact, the variance of the predicted value of y for any value of x is simply the sum of the variance of the mean and the variance of the slope multiplied by the square of the distance of x from the mean of x. The limits for all values of x were computed and connected to get the CI as shown in Figures 2a and b. The square in the last term makes the confidence interval hyperbolic.

$$\text{Variance of predicted value} = Var(\hat{y}) = Var(\bar{y}) + Var(B)(x - \bar{x})^2$$

$$\text{Standard error of the predicted value} = S\hat{y} = \sqrt{Var(\hat{y})}$$

$$\text{Confidence limits} = \pm tS_{\hat{y}}$$

The value of t is selected to give the desired level of confidence. For a large set of data, t = 2 gives approximately 95% confidence; we used t = 2.2 in our calculations.

The values for the funnel curve for fat at 2.5, 4.0, and 5.5%, protein at 2.5, 3.25, and 4.0%, and lactose at 4.2, 4.6, and 5.0% were calculated and used for comparison of the uncertainty of the calibration slope for modified vs. producer milk calibration sets. The leverage of individual samples with respect to their influence on linear regression slope was calculated as follows (when a set of data was fitted to a regression equation, not all data points had equal influence). Although the equations are usually expressed in terms of slope and intercept; that is,

$$y = A + Bx \quad [1]$$

the equation was fit in terms of slope and mean:

$$y = \bar{y} + B(x - \bar{x}) \quad [2]$$

and then converted to form [1] by $A = \bar{y} - B(\bar{x})$. The location of $\bar{y}$ was simply the average of all y values and so all points contribute equally to its position. On the other hand, the further a point is from $\bar{x}$, the greater effect it has on the slope. In an extreme case, if 9 points are located at, say 2.4, and 1 at 4.5, the single point will have as much influence as the other 9 together. The slope, therefore, will be only as good as that point.

An example of this can be seen for the high protein sample in the producer calibration set shown in Figure 2b. If that sample was removed from the calibration set (shown as the large red data point), the relative change in slope of the calibration regression line (compared with the current horizontal line at 0.00 calibration error) is shown as the red line in Figure 2b. Thus, when a sample has high leverage, it can have a large impact on the slope of the calibration line.

A statistic called leverage is an index of the relative contribution of each point to the regression line. It was computed using matrix methods as follows (Cook, 1977; Cook and Weisberg, 1980): suppose we use a sample of 3 milks with protein levels at 2.0, 2.4, and 5.0%. The first 2 points are close together at one end and the third is off by itself at the other end of the range. We define the Design Matrix (X) for this equation as:

$$X = \begin{bmatrix} 1 & 2.0 \\ 1 & 2.5 \\ 1 & 5.0 \end{bmatrix} = \text{The Design Matrix}$$

The normal matrix for this design was the product of the transpose of X times X; i.e., X′X. The normal (or X′X) matrix for this example is

$$X'X = \begin{bmatrix} 3 & 9.5 \\ 9.5 & 35.25 \end{bmatrix}$$

where 3 = number of data points, 9.5 = sum of X, and 35.25 = sum of squares of X. The hat matrix was defined by the equation:

$$H = X(X'X)^{-1}X'$$

where $(X'X)^{-1}$ is the inverse of X′X. The hat matrix is a symmetrical matrix with a row and column for each row of the X matrix; i.e., for each data point. The diagonal elements of this matrix are the leverages of each data point. In this example,

$$H = \begin{bmatrix} 0.597 & 0.484 & -0.081 \\ 0.484 & 0.419 & 0.097 \\ 0.081 & 0.097 & 0.984 \end{bmatrix}.$$

So the leverages for this example are 0.597, 0.419, and 0.984, respectively. The first 2 points were close together and the third was far away from both; this is reflected in the fact that the third point has a larger leverage. Points near the mean will have very small leverages. This calculation was built into the IR-QC software to calculate the leverage of each sample in a calibration set for each component. A common rule of thumb is to call a leverage large if it is >2p/n and very large if it is >3p/n, where n = the number of data points (3 in the example) and p = the number of predictors (1, namely protein, in the example). In the example: 2p/n = 0.667 and 3p/n = 1.000, so the third point would be high leverage, whereas the other 2 are not.

Although the funnel curve is affected by the variability of the individual sample responses, the leverage of individual samples is strictly a function of the design of the calibration set and has nothing to do with the responses. Ideally, a calibration set should be designed in which all leverages are moderately low because, in such a design, the slope is not overly dependent on one or a few samples. Reduction in leverage of individual samples should not be achieved by putting all points near the mean because this will lead to a very unreliable estimate of the slope and produce a wide funnel curve as if there are high leverage samples in the set. Variability in the characteristics of these samples (e.g., variation in the background chemistry of minor milk components) will cause inconsistency in the calibration slope and intercept. In our study, samples with a calculated

Table 6. Characteristics of producer milk calibration sets surveyed over a 2-yr period

Laboratory	Calibration sets in survey	Calibration samples per set	Raw milk sources	Production frequency	Laboratories running reference chemistry	Preservative	Techniques used to increase component range
1	71	12	Individual farms	2 wk	1	Bronopol	Cream added back to gravity separated skim milk
2	41	12	Individual farms	2 wk	4	Potassium dichromate	Pasteurized skim milk added to raw whole milk; addition of water
3	29	10	Individual farms; commingled milk	2 wk	1	Potassium dichromate	Cream added back to gravity separated skim milk
4	53	24	Individual farms; commingled milk	1 wk[1]	1	Bronopol	Blending of multiple individual farms

[1]Each weekly set included 12 milk samples from the previous set and 12 new milk samples.

leverage >0.333 and ≤0.500 were identified as moderate leverage, and those that were >0.500 were identified as high leverage.

Slope and Intercept Consistency. Slope and intercept consistency between calibration sets was evaluated by plotting the slope and intercept values for each calibration set type as a function of study day over the course of each experiment to determine the variation. Inherent factors that affected slope and intercept consistency included the component concentration range, sample distribution within that range, and sample leverage. Other factors that may have contributed to slope and intercept consistency were errors in chemical reference values, chemical deterioration (e.g., proteolysis or lipolysis) of a sample during its useful life, and an unusually high or low concentration of some component in a milk sample other than fat, protein, or lactose that absorbs light at the sample or reference wavelength for one or more components (e.g., citrate).

RESULTS AND DISCUSSION

Survey of Producer Milk Calibration Samples

The production characteristics of producer milk calibration sets prepared by 4 different laboratories over a 2-yr period are presented in Table 6. Calibration sets consisted of between 10 and 24 samples, produced once or twice every 2 wk. All laboratories reported difficulty in achieving a wide and even distribution of component concentrations. Techniques used to increase component range included gravity separation of the cream and recombination of milk and gravity cream or the addition of reduced-fat pasteurized homogenized milk, and the addition of lactose and water. The average chemical analysis from 4 laboratories was used to assign the reference chemical values in the calibration sets from laboratory 2, whereas single-laboratory reference chemical analysis was used by the other 3 laboratories.

The range (mean, largest, and smallest) of component concentration in the producer calibration sets distributed by the 4 laboratories over a 2-yr period is presented in Table 7. Both within and among laboratories, the range for individual components varied from calibration set to set. For fat, the mean range among laboratories varied from 1.68 to 2.67%, with laboratory 2 exhibiting the largest variation in range among calibration sets within a laboratory, from a low of 1.44 to a high of 2.78%. The mean range for true protein among laboratories varied from 0.49 to 1.15%, and the largest variation among calibration sets within a laboratory was observed for laboratory 1, with a low of 0.65% and a high of 2.92%. Lactose concentration had the smallest range and least variation in all calibration sets.

Based on this survey, laboratory 2 was chosen as the source of producer milk calibration sets for Experiments 1 and 2 because the producer calibration sets produced by laboratory 2 had a wider component range and a better distribution of samples within the component ranges than the other 3 laboratories.

Experiment 1

Component Range of the Calibration Sample Sets. The modified milk calibration sets were designed to increase the component concentration range compared with producer milk calibration sample sets. The mean, largest, and smallest concentration ranges for each component of the modified milk and producer milk calibration sets used in Experiment 1 are summarized in Table 8. The modified milk sets had a larger mean range of 3.98% fat compared with 2.29% for the producer milk sets. The mean range of true protein in the modified milk calibration sets was twice that of the producer milk sets (Table 8). The mean range of lactose concentration was 1.32% in the modified milk sets compared with a mean range of 0.47% for the producer milk calibration sets. The component concentration ranges in the sets of modified milk calibration samples were more consistent from set to set than the producer milk sample sets, which is shown by comparing the differ-

Table 7. Survey of reference chemistry ranges (mean, largest, and smallest) for fat, true protein, lactose, and total solids within producer calibration sample sets produced in 4 different laboratories over a 2-yr period

Calibration set source	Range of			
	Fat	True protein	Lactose[1]	Total Solids
	(%)			
Laboratory 1 (n = 71 sets of 12 samples each)				
Mean	2.67	1.15	0.30	3.63
Largest	3.02	2.92	0.63	4.51
Smallest	2.15	0.65	0.16	2.91
Laboratory 2 (n = 41 sets of 12 samples each)				
Mean	2.09	0.94	0.28	2.85
Largest	2.78	1.32	0.47	4.31
Smallest	1.44	0.54	0.17	1.94
Laboratory 3 (n = 29 sets of 10 samples each)				
Mean	1.72	0.49	0.23	1.77
Largest	2.02	0.82	0.39	2.43
Smallest	1.38	0.21	0.13	1.44
Laboratory 4 (n = 53 sets of 24 samples each)				
Mean	1.68	0.73	0.24	2.21
Largest	2.12	0.96	0.55	3.14
Smallest	1.30	0.29	0.10	1.53

[1]Lactose determined by difference method.

ence between the smallest to largest range for each component for the modified milk vs. producer milk sets (Table 8).

Correlation of Fat and Protein Concentrations Within Calibration Sets. In addition to a larger and more consistent component concentration range, the modified milk calibration set was designed to reduce the correlation between fat and protein concentrations. There is a positive correlation between the fat and protein contents in producer milks (Schaefer, 2003). The

Table 8. Reference chemistry ranges (mean, largest, and smallest) for fat, true protein, lactose, and total solids within modified milk and producer milk calibration sample sets used in Experiments 1 and 2

Calibration set source	Range of			
	Fat	True protein	Lactose	Total Solids
	(%)			
Experiment 1				
Modified milk (n = 4 sets of 12 samples each)				
Mean	3.98	2.23	1.32[1]	4.83
Largest	4.04	2.30	1.34	4.92
Smallest	3.92	2.20	1.28	4.76
Producer milk (n = 7 sets of 12 samples each)				
Mean	2.29	1.01	0.47[2]	3.27
Largest	2.59	1.10	0.70	3.62
Smallest	1.83	0.91	0.29	2.78
Experiment 2				
Modified milk (n = 3 sets of 14 samples each)				
Mean	5.58	2.27	1.14[1]	6.81
Largest	5.73	2.30	1.16	6.96
Smallest	5.31	2.24	1.11	6.59
Producer milk (n = 3 sets of 12 samples each)				
Mean	3.00	0.94	0.36[1]	3.62
Largest	3.09	0.98	0.39	3.87
Smallest	2.87	0.88	0.34	3.47

[1]Lactose determined by enzyme method.

[2]Lactose determined by difference method.

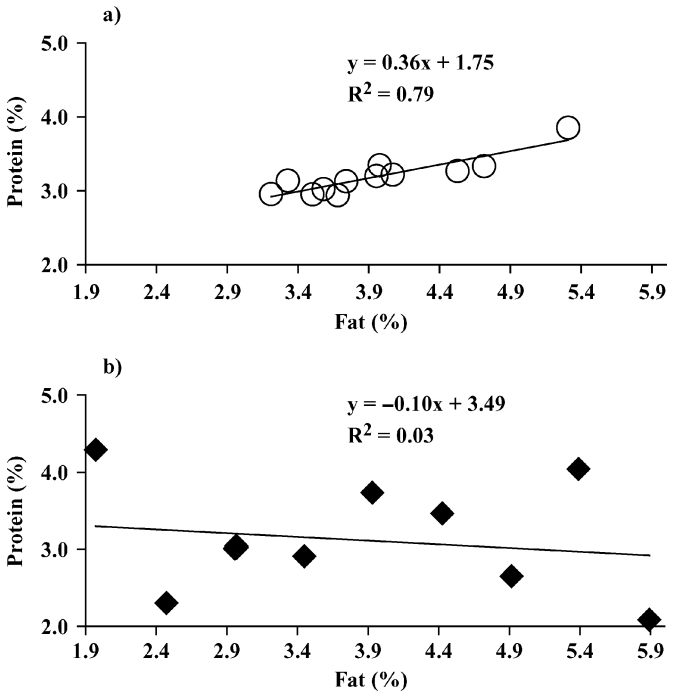

Figure 3. Correlation of percent fat and protein content of a) producer milk calibration samples, and b) modified milk calibration samples used in Experiment 1.

correlation is shown for one producer milk calibration set and one modified milk calibration set in Figures 3a and b, respectively. Producer calibration sets in Experiment 1 had a higher correlation between fat and protein than modified milk calibration sets (Table 9). The slope of the correlation between fat and protein in the producer milk calibration sets used in Experiment 1 ranged from 0.357 to 0.423 and the R^2 ranged from 0.61 to 0.83, compared with much lower values for the modified milks (Table 9). The values for producer samples are consistent with the correlation observed in producer milk samples in 2002 for the Upper Midwest USDA Federal Milk Market (Schaefer, 2003). Similar values were reported by the USDA for a 10-yr period (1992 to 2002) for that region in separate publications. A practical implication of the correlation of the fat and protein components in the producer milk calibration samples is that they are not well-suited for the use of a multiple linear regression approach to calibration (Barbano and Clark, 1989), which assumes independence of all terms in the regression equation.

Calibration Regression Confidence Interval. Data on confidence intervals of calibration regression lines used for MIR analysis are not available in the literature. Calibration uncertainty was assessed by comparing the width of the 95% CI around the calibration regression lines at three concentrations for each component, representing the low, midpoint, and high values typical for raw milk (Table 10). The modified milk calibration sets in Experiment 1 consistently had smaller CI for all components with the CI at the midpoint concentration being smaller ($< \pm 0.013$) than the value at the midpoint for producer milk calibration sets ($< \pm 0.045$). The width of the CI for each component was more consistent among the modified milk calibration sets than among the producer milk calibration sets (Table 10). The larger CI at the midpoint of the range for the producer milks would be expected to contribute to more bias error from set to set on unknown validation samples when using a producer milk-based calibration than a modified milk calibration.

High Leverage Samples. Quantitative data on the occurrence of moderate and high leverage samples in calibration sets used for MIR analysis are not available in the literature. An example of moderate (yellow point) and high (red point) leverage samples is shown in Figure 2b for fat B for a producer milk calibration set. The

Table 9. Slope, intercept, and R^2 values for fat and protein correlation for modified milk and producer milk calibration sets used in Experiment 1 and Experiment 2

Calibration source and set	Slope		Intercept		R^2	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2	Experiment 1	Experiment 2
Modified milk						
1	–0.095	–0.089	3.461	3.578	0.034	0.050
2	–0.096	–0.089	3.509	3.558	0.033	0.044
3	–0.097	–0.089	3.491	3.603	0.032	0.050
4	–0.103	—	3.564	—	0.035	—
Producer milk						
1	0.419	0.238	1.552	2.308	0.789	0.499
2	0.357	0.308	1.840	2.005	0.723	0.768
3	0.376	0.287	1.696	2.057	0.776	0.581
4	0.401	—	1.591	—	0.614	—
5	0.364	—	1.751	—	0.790	—
6	0.423	—	1.493	—	0.836	—
7	0.404	—	1.563	—	0.658	—

Table 10. Width of 95% confidence interval for calibration linear regression for fat A, fat B, protein, and lactose at 3 different component concentrations for modified milk and producer milk calibration sets used in Experiments 1 and 2

	Width (±) of 95% CI for component slopes at different concentrations (%)											
	Fat A			Fat B			Protein			Lactose		
Calibration set	2.5	4.0	5.5	2.5	4.0	5.5	2.5	3.25	4.0	4.2	4.6	5.0
Experiment 1												
Modified milk												
1	0.016	0.011	0.018	0.015	0.010	0.017	0.012	0.009	0.015	0.017	0.009	0.013
2	0.018	0.013	0.022	0.017	0.012	0.021	0.011	0.007	0.013	0.010	0.005	0.006
3	0.012	0.009	0.016	0.013	0.009	0.016	0.013	0.009	0.016	0.016	0.009	0.012
4	0.011	0.008	0.014	0.011	0.008	0.014	0.009	0.006	0.010	0.017	0.009	0.014
Producer milk												
1	0.081	0.033	0.073	0.034	0.014	0.031	0.036	0.013	0.034	0.125	0.038	0.067
2	0.061	0.025	0.064	0.024	0.009	0.025	0.043	0.012	0.043	0.125	0.042	0.053
3	0.071	0.029	0.062	0.033	0.013	0.029	0.031	0.010	0.031	0.134	0.045	0.056
4	0.088	0.033	0.092	0.052	0.018	0.054	0.029	0.011	0.035	0.092	0.032	0.035
5	0.069	0.028	0.072	0.033	0.012	0.034	0.039	0.012	0.044	0.055	0.023	0.020
6	0.074	0.030	0.070	0.045	0.017	0.042	0.041	0.014	0.046	0.093	0.038	0.033
7	0.089	0.037	0.084	0.043	0.017	0.040	0.027	0.011	0.030	0.066	0.026	0.041
Experiment 2												
Modified milk												
1	0.011	0.012	0.018	0.012	0.013	0.019	0.006	0.004	0.006	0.010	0.008	0.014
2	0.010	0.012	0.018	0.010	0.012	0.019	0.008	0.005	0.007	0.006	0.004	0.007
3	0.010	0.011	0.017	0.009	0.010	0.015	0.005	0.003	0.004	0.009	0.007	0.012
Producer milk												
1	0.055	0.025	0.051	0.043	0.019	0.040	0.037	0.011	0.035	0.027	0.012	0.040
2	0.049	0.025	0.050	0.024	0.012	0.024	0.050	0.017	0.053	0.057	0.015	0.067
3	0.036	0.017	0.038	0.020	0.009	0.021	0.050	0.018	0.057	0.048	0.014	0.057

relative impact of a high leverage sample on the linear regression slope is shown by the red line in Figure 2b. If the high leverage sample (large red dot) was removed from the calibration set, then the calibration slope, shown as the horizontal black line at 0.00 of the 95% CI funnel curve, would shift by the relative amount represented by the red line. Thus, high leverage samples can cause slope variation in a linear regression.

The number of moderate and high leverage samples in all calibration sets used in Experiment 1 are presented in Table 11. The modified milk calibration sets had no high leverage samples but had 4 to 8 moderate leverage samples depending on the component. The producer milk calibration sets had between 3 to 5 high and 2 to 5 moderate leverage samples for each component. The absence of high leverage samples in the modified milk calibration sets was a result of the wider component concentration range and a more uniform distribution of concentrations within the range compared with the producer milk calibration sets. As the number of moderate and high leverage samples in any calibration set increases, more variation in calibration slope and intercept can be expected on a day-to-day and set-to-set basis.

Calibration Slope and Intercept Consistency. The calibration slopes for the modified milk calibration sets were more consistent for all components both between calibration sets and within calibration sets than for the producer milk calibration sets. An example of slope consistency is shown for the protein component for modified milk samples in Figure 4a and for producer milk samples in Figure 4b. The protein slope for the modified milk calibration samples across 4 calibration sets ranged from 1.035 to 1.048, whereas the protein slope for the producer milk calibration samples across 7 calibration sets ranged from 1.043 to 1.105 over the course of Experiment 1. The modified milk calibration sample sets had a change in protein slope that ranged from 0.005 to 0.009 from the beginning to the end of the 4-wk set life, compared with the producer milk calibration sets, which had a larger change in slope ranging from 0.006 to 0.017 over the 2-wk set life.

In general, the protein slope decreased within the useful life of each individual producer calibration set (Figure 4b). The decrease in slope could have been caused by systematic quality deterioration of one or more high or moderate leverage samples. Therefore, on this instrument, tests of unknown milk samples that had a protein concentration higher than the mean of the producer calibration set would be lower relative to reference chemistry, and those with low protein would be higher. These trends for slope change within the life of the producer milk calibration set were not seen in the modified milk calibration sets (Figure 4a) that were run at the same time on the same instrument.

Table 11. Number of high and moderate leverage samples in modified and producer calibration sets used in Experiments 1 and 2

	Number of high and moderate leverage samples[1]			
	Component filter			
Calibration set source	Fat A	Fat B	Protein	Lactose
Experiment 1				
Modified milk	H = 0	H = 0	H = 0	H = 0[2]
(4 sets of 12 samples each)	M = 4	M = 4	M = 4	M = 8
Producer milk	H = 3	H = 3	H = 5	H = 4[3]
(7 sets of 12 samples each)	M = 3	M = 3	M = 2	M = 5
Experiment 2				
Modified milk	H = 0	H = 0	H = 0	H = 0[2]
(3 sets of 14 samples each)	M = 0	M = 0	M = 0	M = 0
Producer milk	H = 1	H = 1	H = 3	H = 1[2]
(3 sets of 12 samples each)	M = 5	M = 5	M = 0	M = 3

[1]H = High leverage; M = moderate leverage.

[2]Lactose determined by enzymatic method.

[3]Lactose determined by difference method.

The regression intercepts for the modified milk calibration sets were more consistent for all components both between and within calibration sets than for the producer milk calibration sets. The regression intercept values for protein (Figures 5a, b) showed the same consistency trends as the protein slope; namely, the intercept for the modified milk samples was more consistent from set to set and within a set than the producer milk samples. The change in intercept values was inversely

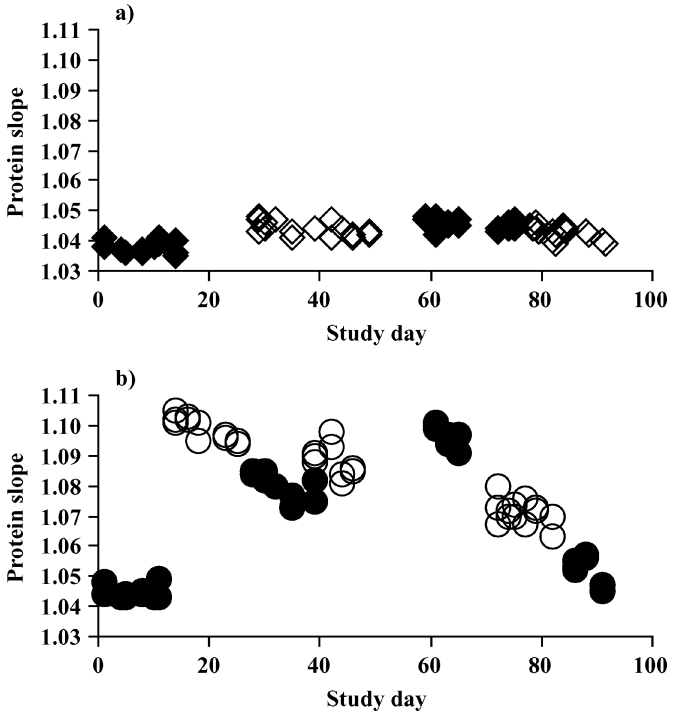

Figure 4. Protein slope for an infrared analyzer calibrated with a) modified milk samples, and b) producer milk samples in Experiment 1 plotted as a function of study day. Filled (sets 1 and 3 for modified and sets 1, 3, 5, and 7 for producer) and empty symbols (sets 2 and 4 for modified and sets 2, 4, and 6 for producer) represent alternating calibration sets in a series.

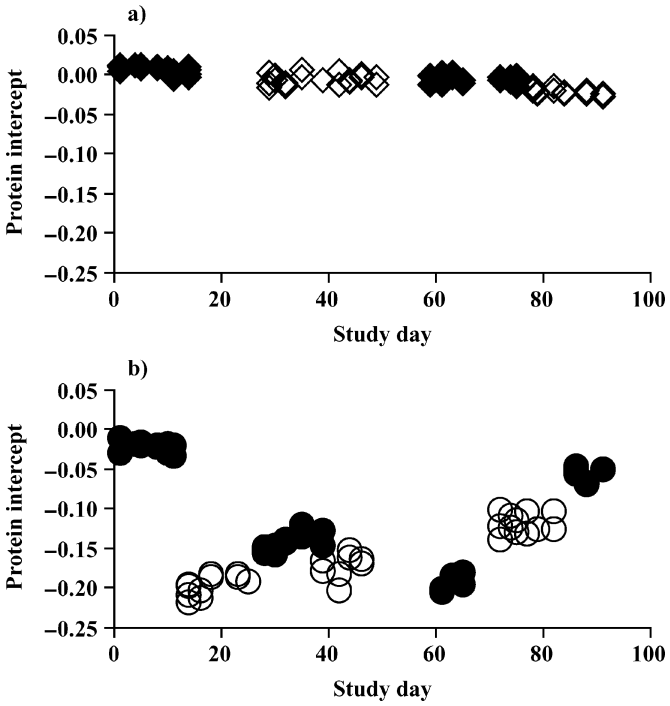

Figure 5. Protein intercept for an infrared analyzer calibrated with a) modified milk samples, and b) producer milk samples in Experiment 1 plotted as a function of study day. Filled (sets 1 and 3 for modified and sets 1, 3, 5, and 7 for producer) and empty (sets 2 and 4 for modified and sets 2, 4, and 6 for producer) symbols represent alternating calibration sets in a series.

related to the change in slope values. The same trends observed in the protein slope and intercept were observed in the fat and lactose components (data not shown).

The use of preserved pasteurized milk contributed to the improved consistency of slope and intercept within the useful life of the modified milk calibration sets compared with the use of preserved raw milk for the producer calibration sets. It is very important to note that all of these calibration samples were preserved, stored at 4°C, and run on the same instrument on the same day. Variations in slope and intercept within the useful life of one calibration set, or from set to set, were not due to differences in the instrument but to the differences in the characteristics of the calibration sets. This would cause differences in results when testing unknown samples on the same instrument that was calibrated using the 2 different types of calibration sample sets.

Experiment 2

Component Range of the Calibration Sample Sets. Compared with Experiment 1, the modified milk calibration sets for Experiment 2 were increased from 12 to 14 samples to increase the fat concentration range, from 2.00 to 6.00 in Experiment 1 (Table 3) to 0.20 to 5.70 in Experiment 2 (Table 4). The change in the lowest value from 2.00% fat in Experiment 1 to 0.20% fat in Experiment 2 extended the utility of the calibration set from the analysis of raw milk to include finished fluid milks (e.g., skim, lowfat). The mean, largest, and smallest concentration range for each component of the modified milk and producer milk calibration sets used in Experiment 2 is summarized in Table 8. The mean concentration range for all components was greater in the modified milk than in the producer milk calibration sets. The component ranges for both the modified milk and producer milk calibration sets were more consistent from set to set in Experiment 2 than in Experiment 1.

Correlation of Fat and Protein Components Within Calibration Sets. Compared with Experiment 1, the modified milk calibration sets were redesigned to further reduce the component correlation (i.e., increase the orthogonality of the calibration set). The slope and R^2 for the correlation of fat and protein concentrations for the modified milk were low and consistent in Experiments 1 and 2, whereas producer milk calibration sets for Experiment 2 had a smaller slope and about the same R^2 as producer milk calibration sets in Experiment 1 (Table 9).

Calibration Regression Confidence Interval. The width of the regression CI for modified milk calibration sets was smaller for all components than the producer milk calibration sets used in Experiment 2 (Table 10), indicating more certain component slope values for the modified milk calibration sets. Both the modified milk and producer milk calibration sets used in Experiment 2 had smaller values for the width of the CI for each component compared with the calibration sets used in Experiment 1 (Table 10). Factors that contributed to the smaller CI in Experiment 2 for modified milk calibration sets included increased component concentration range and better distribution within the range in the modified milk calibration sets. The use of all lab mean chemistry for producer calibration sets contributed to a smaller CI for producer milk calibration sets in Experiment 2.

High Leverage Samples. The modified milk calibration sets used in Experiment 2 had no high or moderate leverage samples, and were an improvement over Experiment 1 (Table 11). Reformulation of the modified milk calibration samples and increasing from 12 to 14 samples improved orthogonality and uniform distribution of component concentrations used in Experiment 2 compared with those used in Experiment 1, and this eliminated all high and medium leverage samples (Table 11).

The producer milk calibration sets used in Experiment 2 had 13 moderate and 5 high leverage samples (Table 11). This is probably a best-case scenario for producer milk calibration sets because the laboratory assembling these sets consistently had the best producer calibration sets based on the preliminary survey. The presence of moderate and high leverage samples in natural producer milk calibration sets is unavoidable because the laboratory assembling the calibration sets has little control over the range, and distribution of concentrations within the range, for these samples.

Calibration Slope and Intercept Consistency. The calibration slope for the modified milk calibration sets were more consistent both between calibration sets and within a calibration set than for the producer milk calibration sets in Experiment 2, as in Experiment 1. An example of slope consistency is shown for the protein component for the modified milk samples (Figure 6a) and for the producer milk calibration samples (Figure 6b). There was more day-to-day variation in slope within the set life of the producer milk calibration sets (Figure 6b) than for the modified milk calibration sets (Figure 6a). Although the protein slopes for the producer calibration sets for Experiment 2 (Figure 6b) were more consistent from set to set than the producer calibration sets used in Experiment 1 (Figure 4b), the producer milk calibration sets were always less consistent than the modified milk calibration sets in both experiments.

Within both experiments in this study, the modified milk sets and the producer milk sets were stored in the

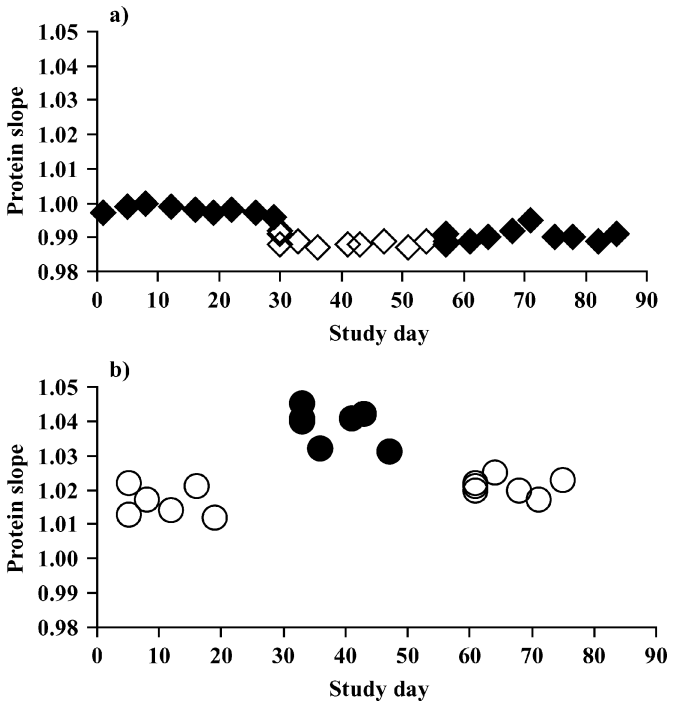

Figure 6. Protein slope for an infrared analyzer calibrated with a) modified milk samples, and b) producer milk samples in Experiment 2 plotted as a function of study day. Filled (sets 1 and 3 for modified and set 2 for producer) and empty (set 2 for modified and sets 1 and 3 for producer) symbols represent alternating calibration sets in a series.

same 4°C cooler and calibrations were run on the same instrument on the same day. Any day-to-day variation in instrument conditions would be reflected in day-to-day variation in the slopes for both calibration sets and this was not evident in the data for either experiment (Figures 4 and 6). Thus, the larger variation observed in calibration slopes and intercepts within and between producer calibration sets compared with the modified milk calibration sets was caused by differences in the characteristics of the producer calibration sets, rather than variations in the response characteristics of the MIR milk analyzer. The improved consistency during the useful life of the modified milk calibration sets was also due to the use of preserved pasteurized milk, compared with the producer milk sets that use preserved raw milks. Calibration of MIR milk analyzers with modified milk calibration sets has the potential to produce improved validation accuracy (i.e., smaller mean difference and standard deviation of the difference from reference chemistry) in MIR milk analysis than when the same analyzer is calibrated with producer milk sets.

CONCLUSIONS

Pasteurized modified milk calibration samples achieved smaller regression CI (i.e., calibration uncertainty) and improved slope and intercept consistency compared with producer milk calibration samples. The larger component concentration range, more even concentration distribution within the range, and the lower correlation of fat and protein concentrations for the modified milk calibration sets resulted in a smaller 95% CI around the regression line and eliminated moderate and high leverage samples from the modified milk calibration sets compared with the producer milk calibration sets. The CI for the producer calibration sets were about 2 to 12 times larger than the CI for the modified milk calibration sets, depending on the component. The preserved pasteurized modified milk sets exhibited more consistent day-to-day and set-to-set calibration slope and intercept values for all components, than the preserved raw producer milk calibration sets. Modified milk calibration samples have the potential to produce calibrations for MIR milk analyzers that will perform better in validation checks than producer milk-based calibrations by reducing the mean difference and standard deviation of the difference between instrument values and reference chemistry.

ACKNOWLEDGMENTS

The authors would like to thank all the USDA Federal Milk Market laboratories and affiliated laboratories for their collaboration and sample analysis in this work. The technical assistance in sample preparation by Maureen Chapman, Laura Landolf, Bob Kaltaler, Mark Schweisthal, and Pat Wood was important for the success of this project. The authors thank the Test Procedures Committee of the USDA, Dairy Programs, Federal Milk Markets for their financial support of this research.

REFERENCES

AOAC. 2000. Official Methods for Analysis. 17th ed. AOAC International, Gaithersburg, MD.

Barbano, D. M., and J. L. Clark. 1989. Infrared milk analysis – Challenges for the future. J. Dairy Sci. 72:1627–1636.

Barbano, D. M., and J. M. Lynch. 1992. Crude and protein nitrogen bases for protein measurement and their impact on current testing accuracy. J. Dairy Sci. 75:3210–3217.

Biggs, D. A., G. Johnsson, and L. O. Sjaunja. 1987. Analysis of fat, protein, lactose, & total solids by infra-red absorption. Pages 21–30 in Monograph on Rapid Indirect Methods for Measurement of the Major Components of Milk. Bull. Int. Dairy Fed. No. 208. Int. Dairy Fed., Brussels, Belgium.

Biggs, D. A., and D. McKenna. 1989. Alternative methods for infrared analysis of fat in milk: Interlaboratory study. J. AOAC 72:724–734.

Cook, R. D. 1977. Detection of influential observation in linear regression. Technometrics 19:15–18.

Cook, R. D., and S. Weisberg. 1980. Characterizations of an empirical influence function for detecting influential cases in regression. Technometrics 22:495–508.

IDF. 2000. Whole milk: Determination of milkfat, protein and lactose content. Guidance on the operation of mid-infrared instruments. IDF Standard 141C:2000. Int. Dairy Fed., Brussels, Belgium.
Lynch, J. M. 1998. Use of AOAC International method performance statistics in the laboratory. J. AOAC Int. 81:679–684.
Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1990. Variation in the ash and nonprotein nitrogen content of milk, and use of milk protein content to predict ash content. J. Dairy Sci. 73(Suppl. 1):92. (Abstr.)
Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1995. Evaluation of commercially available milk powders for calibration of mid-infrared analyzers. J. AOAC Int. 78:1219–1224.
Schaefer, H. H. 2003. Upper Midwest Marketing Area analysis of component levels and somatic cell count in individual herd milk at the farm level 2002. Staff paper 03-01. Fed. Milk Market Admin. Office, Minneapolis, MN.
Smith, E. B., D. M. Barbano, J. L. Lynch, and J. R. Fleming. 1993a. Performance of homogenizers in infrared milk analyzers: A survey. J. AOAC Int. 76:1033–1041.
Smith, E. B., D. M. Barbano, J. L. Lynch, and J. R. Fleming. 1993b. A quantitative linearity evaluation method for infrared milk analyzers. J. AOAC Int. 76:1300–1308.
Smith, E. B., D. M. Barbano, J. L. Lynch, and J. R. Fleming. 1995. Infrared analysis of milk: Effect of homogenizer and optical filter selection on apparent homogenization efficiency and repeatability. J. AOAC Int. 78:1225–1233.
van de Voort, F. R., A. A. Elkashef, and B. L. Mills. 1990. Dry calibration milks for infrared milk analyzers. J. AOAC 73:688–692.