J. Dairy Sci. 98:4174–4181
http://dx.doi.org/10.3168/jds.2014-8776

A novel statistical approach to detect differences in fat and protein test values among mid-infrared spectrophotometers[1]

Michael C. Adams and David M. Barbano[2]
Northeast Dairy Foods Research Center, Department of Food Science, Cornell University, Ithaca, NY 14853

ABSTRACT

Our objective was to develop a statistical approach that could be used to determine whether a handler's fat, protein, or other solids mid-infrared (MIR) spectrophotometer test values were different, on average, from a milk regulatory laboratory's MIR test values when split-sampling test values are not available. To accomplish this objective, the Proc GLM procedure of SAS (SAS Institute Inc., Cary, NC) was used to develop a multiple linear regression model to evaluate 4 mo of MIR producer payment testing data (112 to 167 producers per month) from 2 different MIR instruments. For each of the 4 mo and each of the 2 components (fat or protein), the GLM model was *Response* = *Instrument* + *Producer* + *Date* + *2-Way Interactions* + *3-Way Interaction*. *Instrument* was significant in determining fat and protein tests for 3 of the 4 mo, and *Producer* was significant in determining fat and protein tests for all 4 mo. This model was also used to establish fat and protein least significant differences (LSD) between instruments. Fat LSD between instruments ranged from 0.0108 to 0.0144% ($\alpha = 0.05$) for the 4 mo studied, whereas protein LSD between instruments ranged from 0.0046 to 0.0085% ($\alpha = 0.05$). In addition, regression analysis was used to determine the effects of component concentration and date of sampling on fat and protein differences between 2 MIR instruments. This statistical approach could be performed monthly to document a regulatory laboratory's verification that a given handler's instrument has obtained a different test result, on average, from that of the regulatory laboratory's and that an adjustment to producer payment may be required.

Key words: statistics, mid-infrared, least significant difference, milk payment testing

INTRODUCTION

Typically, in the United States, thousands of raw milk samples are tested each day using mid-infrared (**MIR**) spectrophotometers to determine fat, protein, and other solids. These component measures are used in conjunction with delivery weights to determine payment to milk producers (i.e., dairy farmers) on a component weight basis (Barbano and Lynch, 2006). Because the trend is for fewer producers to provide larger volumes of milk per day (von Keyserlingk et al., 2013) to handlers (i.e., dairy plants), small errors in payment testing could amount to large amounts of money. The USDA Federal Milk Marketing Orders ensure accuracy of milk component testing by comparing MIR results from handlers' labs to the corresponding regulatory laboratory. Currently, this comparison can be made in 2 ways: through (1) split-sampling, or (2) statistical analysis of routine testing data for a group of producers (e.g., 100). Split-sampling is ideal because the same milk can be tested in both labs, thereby removing the uncertainty contributed by day-to-day variation and differences among samples. However, split-sampling can be difficult to carry out consistently from a logistical standpoint. It also requires twice as much milk and twice as many sample containers, making it a less sustainable practice. Split-sampling also has the disadvantage, from a verification perspective, that the handler knows these samples are being used to verify accuracy of testing. Statistical analysis of existing routine data could be used as an alternative to split-sampling to accomplish verification of handler testing accuracy when both labs are testing milks from the same group of producers during a month but on milks from different days if appropriate statistical procedures are used. In this approach, the normal day-to-day variation in milk composition for the same producer will present a challenge for accuracy, but this approach has the advantage that handlers can be checked at any time without prior knowledge that their performance is being evaluated.

Received August 23, 2014.
Accepted February 2, 2015.
[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

By building multiple linear regression models, least significant differences (**LSD**) can be determined at various confidence levels that take into account instrument, producer, and day-to-day variation. The statistical models can be constructed using only a few tests per producer per month (e.g., 3) and are not constrained to using identical duplicate milk samples for both MIR analyses on the same day. This method offers flexibility in sample collection and would be robust to the logistical problems associated with split-sampling of large groups of individual producers. In addition, random selection of milk samples within a given period (e.g., a 10-d period) for statistical analysis would prevent handlers from knowing when their results were being compared with the regulatory laboratory's results.

Our objective was to develop a statistical approach that could be used to determine whether a handler's fat, protein, or other solids MIR test values were different, on average, from a regulatory laboratory's MIR test values when split-sampling is not practical. To accomplish this objective, we developed a multiple linear regression model to evaluate MIR producer payment testing data; then, fat and protein LSD between instruments using this model were calculated. In addition, regression analysis was used to determine the effects of component concentration and date of sampling on fat and protein differences between 2 MIR instruments. This statistical approach could be performed monthly to document a regulatory laboratory's verification that a given handler's instrument has obtained a different test result, on average, from that of the regulatory laboratory's and that adjustments to producer payment and the handler's MIR instrument may be required.

MATERIALS AND METHODS

Data Acquisition and Organization

Fat and protein testing data from 2 anonymous MIR instruments (arbitrarily designated "Instrument 1" and "Instrument 2") used to test producer raw milks collected from a common set of northeast UF milk producers were supplied by USDA Federal Milk Marketing Order 1 (Albany, NY) for the months of September 2010, December 2010, March 2011, and July 2011. Within these months, payment testing data from 112, 149, 150, and 167 producers were analyzed, respectively.

Eight data sets (2 components: fat and protein × 4 mo: September, December, March, and July) were created and independently analyzed using the methods described below. To populate each of these data sets, 3 test values were randomly chosen from each instrument for each producer within each month. These 3 test values represented milk samples collected during each of 3 periods within a month: the beginning of the month, the middle of the month, and the end of the month. The beginning, middle, and end of the month periods encompassed d 1 to 10 of the month, d 11 to 20 of the month, and d 21 to the end of the month, respectively. The dates on which milk samples were collected were not the same for the 2 instruments for a given producer during a given period. This choice was made to permit flexibility when selecting samples for comparison of results from 2 laboratories. The average difference in sample date between 2 instruments' test values within a common 10-d period for a common producer was 2.2 d for the 8 data sets studied. The median difference was 2 d for all 8 data sets.

Statistical Model Development and Analyses

Multiple Linear Regression Model. A multiple linear regression model was constructed using SAS software (version 9.2, SAS Institute Inc., Cary, NC). The following general linear model (PROC GLM) was used:

```
proc glm data= work.month;
class Instrument Producer;
model Fat Protein = Instrument Producer Date
                    Instrument*Producer
                    Instrument*Date
                    Producer*Date
                    Instrument*Producer*Date;
means Instrument / lsd alpha=0.0001;
means Instrument / lsd alpha=0.001;
means Instrument / lsd alpha=0.005;
means Instrument / lsd alpha=0.01;
means Instrument / lsd alpha=0.05;
means Instrument / lsd alpha=0.10;
run;
```

The ANOVA model included factors to account for variation contributed by the 2 MIR instruments used (*Instrument*), the producers from which raw milk samples were collected (*Producer*), the date of milk sampling (*Date*), and the interactions among these factors. *Instrument* and *Producer* were treated as categorical factors and *Date* was treated as a continuous factor. Furthermore, *Date* was mean-centered before model development to reduce the likelihood of multicollinearity and to improve the precision of its parameter estimates (Glantz and Slinker, 2001). Mean centering was performed by subtracting the average of all sample dates within a month from each sample date. For this application, the only factor levels of interest for a given analysis would be those instruments and producers be-