J. Dairy Sci. 96:5933–5942
http://dx.doi.org/10.3168/jds.2012-6068

Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats

B. S. Dagnachew,[1] T. H. E. Meuwissen, and T. Ådnøy
Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, PO Box 5003, N-1432 Ås, Norway

ABSTRACT

The usual practice today is that milk component phenotypes are predicted using Fourier-transform infrared (FTIR) spectra and they are then, together with pedigree information, used in BLUP for calculation of individual estimated breeding values. Here, this is referred to as the indirect prediction (IP) approach. An alternative approach—a direct prediction (DP) method—is proposed, where genetic analyses are directly conducted on the milk FTIR spectral variables. Breeding values of all derived milk traits (protein, fat, fatty acid composition, and coagulation properties, among others) can then be predicted as traits correlated only to the genetic information of the spectra. For the DP, no need exists to predict the phenotypes before calculating breeding values for each of the traits—the genetic analysis is done once for the spectra, and is applicable to all traits derived from the spectra. The aim of the study was to compare the effects of DP and IP of milk composition and quality traits on prediction error variance (PEV) and genetic gain. A data set containing 27,927 milk FTIR spectral observations and milk composition phenotypes (fat, lactose, and protein) belonging to 14,869 goats of 271 herds was used for training and evaluating models. Partial least squares regression was used for calibrating prediction models for fat, protein, and lactose percentages. Restricted maximum likelihood was used to estimate variance components of the spectral variables after principal components analysis was applied to reduce the spectral dimension. Estimated breeding values were predicted for fat, lactose, and protein percentages using DP and IP methods. The DP approach reduced the mean PEV by 3.73, 4.07, and 7.04% for fat, lactose, and protein percentages, respectively, compared with the IP method. Given the reduction in PEV, relative genetic gains were 2.99, 2.78, and 4.85% for fat, lactose, and protein percentages, respectively. We concluded that more accurate estimated breeding values could be found using genetic components of milk FTIR spectra compared with single-trait animal model analyses on phenotypes predicted from the spectra separately. The potential and application is not only limited to milk FTIR spectra, but could also be extended to any spectroscopy techniques implemented in other species and for other traits.

Key words: Fourier-transform infrared spectra, breeding value, prediction error variance, genetic gain

INTRODUCTION

In Europe, goat milk is mainly used for cheese production. Producing quality cheese requires quality raw milk—milk with good cheese-making ability. The quality of goat milk for cheese making is strongly related to the total amount and type of proteins, minerals, lipids, and coagulation properties (Grosclaude and Martin, 1997; Hallén, 2008). These properties of the milk are influenced by several genetic and environmental factors (Jaubert, 1996). Selection of dairy animals for improving the cheese-making ability of the milk is possible because genetic variation for most milk composition and quality traits exist (Ikonen, 2000; Manfredi, 2003; Cassandro et al., 2008; Schopen, et al., 2009; Cecchinato et al., 2011b).

Genetic evaluation of dairy animals requires rather large data sets, possibly larger than what is feasible to observe directly through chemical analyses for milk composition and quality traits because they are costly and time consuming. An alternative and more convenient method—use of infrared (**IR**) spectra—has been proposed (Goulden, 1964). Fourier-transform IR (**FTIR**) spectrometry is among several vibration spectroscopy techniques applied in food analysis, which use the mid-infrared region of the light spectrum (van de Voort, 1992). The FTIR spectra, exhibiting reduced cost of analysis, high throughput, and possibly large-scale application, are widely used today in milk-recording programs for milk payment and prediction of major milk components (e.g., protein, fat, and lactose percentages).

Received August 18, 2012.
Accepted May 13, 2013.

[1] Corresponding author: binyam.dagnachew@umb.no

The common practice today is that milk composition phenotypes are collected using FTIR spectra. These phenotypes are obtained through application of calibration equations on the spectral data. The predicted phenotypes are then, together with pedigree information and variance component estimates, used in BLUP for estimation of EBV, and selection candidates are ranked accordingly. In this paper, this approach is referred to as the indirect prediction (**IP**) method.

This study proposed an alternative approach—direct prediction (**DP**) method—where genetic analyses are directly conducted on the milk FTIR spectral variables. The BLUP breeding values for milk composition traits are then predicted as correlated traits to the genetic component of the spectra. The proposed method involves genetic analysis of several hundreds of spectral variables and is expected to benefit from use of information-rich spectral data. Studies have showed that milk FTIR spectral variables have strong correlations among each other in cow milk spectra (Soyeurt et al., 2010) and in goat milk spectra (Dagnachew and Ådnøy, 2011). Direct genetic analysis on these correlated spectral variables may provide a better estimation of genetic parameters and, hence, improve accuracy of animal evaluation methods.

Genetic evaluation of Norwegian dairy goats is based on major milk components predicted from FTIR spectra. These milk component measurements, together with pedigree information, are used for selection of superior individuals following the IP approach as explained above. The main aim of this study was to investigate the effect of using the DP approach for prediction of breeding values instead of the IP approach on accuracy and prediction error variance (**PEV**) for major milk components (i.e., fat, lactose, and protein percentages). We did not have independent chemical analyses of the milk components, but relied on predictions of milk content based on the same FTIR spectra as found in the Norwegian Dairy Goat Control.

MATERIALS AND METHODS

FTIR Spectra and Phenotype Data

In the Norwegian Dairy Goat Control, daily milk yield is to be recorded at least 5 times per farm per year. Of the 5 recording days, at least 3 d of individual goat samples should be analyzed for major milk components (fat percentage, protein percentage, lactose percentage, SCC, FFA, and urea content). The milk samples are analyzed by the MilkoScan CombiFoss 6500 instrument (Foss Electric A/S, Hillerød, Denmark) and the components are predicted using equations applied to the spectral data. The predictions are transferred to the Norwegian dairy goat control. The raw spectral data are also stored in a database for research and development purposes. For this study, 2 yr (2007 and 2008) of raw goat milk FTIR spectral data were used. Phenotype measurements (i.e., fat, protein, and lactose) were obtained from the Norwegian Dairy Goat Control recording system. These phenotypes were predictions based on the spectra. We use these predicted phenotypes because no phenotype was available from chemical analysis.

Table 1. Data structure: number of herds, goats, and individual test Fourier-transform infrared (FTIR) spectra

Item	Herds (no.)	Goats (no.)	Tests (no.)
Data set[1]			
Calibration			
Lactation 1	252	5,575	9,825
Lactation 2	233	5,162	9,074
Lactation 3	121	709	1,028
Total	271	11,446	19,927
Test			
Lactation 1	245	1,623	3,748
Lactation 2	232	1,428	3,485
Lactation 3	91	372	767
Total	271	3,423	8,000
Total		14,869	27,927

[1] Lactation group 3 includes lactation number 3 and above.

Two data sets were created randomly, one for calibrating a model (calibration data set) and other for testing model performance (test data set). The sizes of the data sets were 19,927 and 8,000 for calibration and test data sets, respectively. The structure of the data sets used in this study is presented in Table 1.

A pedigree file containing all animals with records was available. Their ancestors were traced back as far as possible and the final pedigree file contained 42,770 animals. The number of sires was 4,102 and number of dams was 20,644. Among the sires, 1,890 of them had progeny with records in the data set. Among the dams, 2,632 of them had both their own record and progeny with records, and 9,971 of them had progeny with records in the data set.

Calibration of Prediction Model

Due to the large amount of water in milk, both the O–H bending region (between 1,600 and 1,700 cm^{-1}) and the O–H stretching region (above 3,020 cm^{-1}) are more or less opaque to IR light in milk samples (Afseth et al., 2010). Another region, between 1,801 and 2,600 cm^{-1}, does not contain useful chemical information (Andersen et al., 2002). These spectral regions were removed and the remaining frequencies (i.e., 941 to 1,586 cm^{-1}, 1,705 to 1,801 cm^{-1}, and 2,693 to 3,067 cm^{-1}), giving a total of 321 spectral data points, were selected