J. Dairy Sci. 100:2057–2067
https://doi.org/10.3168/jds.2016-11951

# Comparison between direct and indirect methods for exploiting Fourier transform spectral information in estimation of breeding values for fine composition and technological properties of milk

**V. Bonfatti,*[1] D. Vicario,† L. Degano,† A. Lugo,‡ and P. Carnier***
*Department of Comparative Biomedicine and Food Science (BCA), University of Padova, 35020, Legnaro, Italy
†Italian Simmental Cattle Breeders Association (ANAPRI), 33100, Udine, Italy
‡Friuli Venezia Giulia Milk Recording Agency (AAFVG), 33033, Codroipo, Italy

## ABSTRACT

The aim of this study was to compare the common method of exploiting infrared spectral data in animal breeding; that is, estimating the breeding values for the traits predicted by infrared spectroscopy, and an alternative approach based on the direct use of spectral information (direct prediction, DP) to predict the estimated breeding values (EBV). Traits were pH, milk coagulation properties, contents of the main casein and whey protein fractions, cheese yield measured by micro-cheese making, lactoferrin, Ca, and fat composition. For the DP method, the number of spectral variables was reduced by principal components analysis to 8 latent traits that explained 99% of the original spectral variation. Restricted maximum likelihood was used to estimate variance components of the latent traits. (Co)variance components of the original spectral traits were obtained by back-transformation and EBV of all derived milk traits were then predicted as traits correlated with the genetic information of the spectra. The rank correlation between the EBV obtained for the infrared-predicted traits and those obtained from the DP method was variable across traits. Rank correlations ranged from 0.07 (for the content of saturated fatty acids expressed as g/100 g of fat) to 0.96 (for dry matter cheese yield, %) and, for most traits, was <0.5. This result can be explained by the nature of the principal components analysis: it does not take into account the covariance between the spectral variables and the reference traits but produces latent traits that maximize the spectral variance explained. Thus, the direct approach is more likely to be effective for traits more related to the main sources of spectral variation (i.e., protein and fat). More research is required to study spectral genetic variation and to determine the best way to choose spectral regions and the type and number of considered latent traits for potential applications.

**Key words:** infrared spectroscopy, fatty acid, protein fraction, breeding value

## INTRODUCTION

Fourier-transform infrared spectroscopy (**FTIR**) is a useful tool to predict individual phenotypes for traditional and innovative milk traits and a candidate method to replace gold standard methodologies, which are often not applicable for population-wide phenotyping due to high cost or other practical limitations. Infrared prediction of individual phenotypes relies on the availability of calibration equations developed using gold standard measures of traits of concern and FTIR spectra for a limited number of reference samples. Together with pedigree information and variance component estimates, predicted phenotypes can be used in BLUP to obtain EBV. This approach is referred to as the indirect prediction method (**IP**; Dagnachew et al., 2013b) because the spectral information is not directly used in EBV prediction procedures, although the spectra provide insights into the genetic variation in milk components (Soyeurt et al., 2010).

Starting from evidence that milk FTIR spectral variables exhibit tight correlations among each other (Soyeurt et al., 2010; Dagnachew et al., 2013a) and considering that direct genetic analysis on such correlated spectral variables may enhance the accuracy of genetic evaluation methods, a direct prediction (**DP**) approach has been proposed (Dagnachew et al., 2013b). In the DP approach, EBV prediction is performed using the milk FTIR spectral variables directly, and EBV for traits of concern are derived from the predicted EBV for the spectral variables (i.e., EBV for the traits of interest are predicted as EBV of traits correlated with the genetic component of the spectra).

---

Received September 1, 2016.
Accepted December 5, 2016.
[1] Corresponding author: valentina.bonfatti@unipd.it

The DP method has some benefits over the IP method: there is no need to predict phenotypes from the spectra to estimate the EBV for the traits, and EBV are predicted once (only for the spectra) and then used to derive the EBV of traits. This is particularly relevant when considering the high number of traits for which FTIR calibration equation are being developed (Bonfatti et al., 2016; Gengler et al., 2016). In addition, direct analysis of the spectral variables may increase the precision of the estimated genetic parameters and the accuracy of EBV predictions and genetic gains, particularly for low-heritability traits, as a consequence of exploiting the genetic relationships among many spectral variables (Dagnachew et al., 2013b).

Dagnachew et al. (2013b) compared DP and IP using goat milk spectra and reported very high rank correlations between the EBV provided by the 2 methods. In that study, infrared predictions of fat, protein, and lactose contents were used as phenotypes because no data from chemical analysis were available. The investigated traits were directly linked to spectral information, and the calibration equations developed by Dagnachew et al. (2013b) had very high predictive ability, with $R^2$ values ranging from 0.95 to 0.98. The DP and IP methods have not been compared for calibration equations developed using independent reference data obtained by chemical analysis and for traits predicted with intermediate to low accuracy.

The aim of this study was to compare DP and IP as methods for routine prediction of EBV in a dairy cattle population for a group of traits that describe the fine composition and technological properties of milk and that are predicted with variable accuracy using FTIR spectra.

## MATERIALS AND METHODS

### *FTIR Spectra and Calibration Models*

A total of 100,272 milk samples were collected (from February 2013 to June 2014) from 11,216 Italian Simmental cows (92 herds) during the routine milk recording operated in Italy in the Friuli Venezia Giulia region by the regional milk recording agency (AAFVG, Codroipo, Italy). On average, each cow provided 6.9 milk samples, with a minimum of 1 and a maximum of 12 samples. Samples were analyzed using a MilkoScan FT6000 (Foss Electric A/S, Hillerød, Denmark), and the generated FTIR absorbance spectral data (1,060 variables per spectrum) were recorded.

Calibration equations used in this study were the outcome of a research project (MilCo project, CPDA122982; University of Padova, Padova, Italy), which started in 2013 with the aim of developing procedures to estimate EBV for the Italian Simmental cattle population for FTIR predictions of detailed protein and FA composition, minerals, lactoferrin, coagulation properties, cheese yield, and curd composition. Reference data for the development of calibration equations were obtained for milk protein composition by reversed phase HPLC (Bonfatti et al., 2008), for FA composition by accelerated extraction (Dionex application note 345; Thermo Scientific Dionex, 2016) followed by 2-dimensional gas chromatography separation (Pellattiero et al., 2015), for minerals by inductively coupled plasma atomic emission spectroscopy (Soyeurt et al., 2009), for lactoferrin by ELISA (Soyeurt et al., 2007), for milk coagulation properties by lactodynamography (Bonfatti et al., 2016), and for cheese yield and curd composition by micro-cheese making (Bonfatti et al., 2016). In total, 92 traits were measured and calibration equations were developed using more than 1,000 samples for each of the investigated traits, with the exception of minerals (n = 689) and lactoferrin (n = 558). Details on procedures providing reference data for the traits investigated in this study can be found in Bonfatti et al. (2016).

Due to the interference of water absorption, the O–H bending and stretching regions of the spectra (between 1,628 and 1,658 $\text{cm}^{-1}$ and between 3,105 $\text{cm}^{-1}$ and 3,444 $\text{cm}^{-1}$, respectively) were removed from each spectrum, as suggested by Soyeurt et al. (2010). Spectral outliers were identified based on the standardized Mahalanobis distance (Burns and Ciurczak, 2007).

Calibration equations to be used in this study were developed using the remaining 872 spectral variables and modified partial least square (**MPLS**; Shenk and Westerhaus, 1991) regression procedures, as implemented in the software WinISI II (Infrasoft International Inc., State College, PA), and were cross-validated using a 10-random-segments procedure.

### *Estimates of EBV Under the DP and IP Methods*

Sixteen traits were used to compare the EBV obtained from application of IP and DP. The traits were selected from the 92 traits investigated in the MilCo project to compare the 2 methods under scenarios in which FTIR predictions had variable accuracy. Performance of the calibration equations used in the prediction of the 16 selected traits is reported in Table 1. Values of the coefficient of determination of cross-validation ($R^2_{CV}$) and the ratio of performance to deviation (i.e., the ratio of the SD of a trait to the standard error in cross-validation) ranged from 0.35 to 0.86 and from 1.24 to 2.85, respectively.

A schematic representation of IP and DP methods is depicted in Figure 1. The estimation of genetic and