Reconstructing representative daily Fourier transform mid-infrared milk spectra from partial milking robot samples using machine learning

C. Sneessens[1], S. Franceschini[1], P.-E. Jacoby[2], E. Reding[2], J. Leblois[2], H. Soyeurt[1]

[1] TERRA Research and Teaching Centre, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium
[2] AWé Groupe, Walloon Breeders Association, Ciney, Belgium

Automated milking systems (AMS) are increasingly adopted on European dairy farms to improve herd monitoring through automatically collected measurements. However, obtaining a representative daily milk sample for milk recording from those systems remains challenging because milk composition varies throughout the day and cows are milked voluntarily. In 2002, Peeters and Galesloot developed an ICAR-certified method to estimate daily fat content (%FAT24h), but not to reconstruct a representative Fourier-transform mid-infrared (FT-MIR) spectrum (SPECT24h) suitable for predicting multiple milk components. So, this study aimed to develop models allowing the estimation of %FAT24h and SPECT24h from partial milkings.

Two datasets were included: one with 49,631 historical records from five farms, three being independent and two being included in the experiment, and a validation dataset with 2,230 records from 746 cows across seven Belgian farms where every milking was sampled. Two sampling protocols were evaluated: RZ (one milking) and RM (two milkings), using three estimation methods: M1 (based on Peeters and Galesloot), M2 (weighted average), and M3 (linear regression). For the RZ protocol, M1 provided the most accurate %FAT24h estimates (RMSE = 0.275%). For RM, accuracy improved (RMSE = 0.193%). Optimal performance was observed when sampled milkings represented 40–60% of daily milk volume. M2 became the most accurate when this proportion exceeded 60% (RMSE = 0.190%), while M3 did not outperform the other methods.

Instead of predicting %FAT24h, the second step focuses on reconstructing a FT-MIR spectrum that should be representative of the day. From this spectrum, any prediction equation is applicable, but this study presents the performances for the prediction of fat content. Daily spectra reconstructed from partial milking samples enabled %FAT24h prediction with accuracy equal to or greater than traditional methods. For M1, RMSE was 0.272% (RZ) and 0.193% (RM). For M2, reconstructed spectra slightly improved accuracy (RMSE = 0.182%).

These results demonstrate the feasibility of reconstructing SPECT24h from partial milkings and using it to predict milk fat content, potentially enabling the full usage of milk recording FT-MIR spectra for management and monitoring in farms using AMS.