S05(T)-OP-4

Creation of a functional spectral database – Can the calculation of a spectral Global h distance ensure the quality of international based MIR predictions?

Lei Zhang[1,2], YaBin Ma[3], Frédéric Dehareng[4], Clément Grelet[4], Frédéric Colinet[2], Nicolas Gengler[2], Hélène Soyeurt[2]

[1]AgroBioChem Department, University of Liège – Gembloux Agro-Bio Tech, Gembloux, Belgium

[2]TERRA Research Centre, University of Liège – Gembloux Agro-Bio Tech, Gembloux, Belgium

[3]Hebei Livestock Breeding Station, Shijiazhuang, China

[4]Walloon Agricultural Research Centre, Gembloux, Belgium

Milk mid-infrared (MIR) spectrometry provides new traits related to milk nutritional quality, cow well-being and environmental footprint of dairy production. The standardization of MIR data allows sharing international based prediction equations between different spectrometers. When used in routine, the accuracy of prediction depends partly on the capacity of the calibration dataset to cover the variability of the new spectra to predict. Unfortunately, the current way of predicting rarely take into account those specificities. To cope with this issue, the calculation of standardized Mahalanobis distance (GH) between the average calibration spectrum and the spectrum to be predicted is of interest. In this work, the effect of GH limitation on the MIR predicted traits were estimated using 198,394 milk samples from Chinese Holstein cows and analyzed by 3 Bentley FTS spectrometers. The content of fat, protein, monounsaturated fatty acid (MFA), unsaturated FA (UFA), saturated FA (SFA), polyunsaturated FA (PFA) predicted by the models of the Manufacturer were assumed to be the reference values. Averages of fat, protein, lactose, MFA, PFA, SFA and UFA were 3.97, 3.43, 5.09, 0.86, 0.07, 2.62 and 0.93 g/dL of milk. By analyzing common milk samples with the European Milk Recording network, the recorded Bentley MIR spectra were standardized to master MIR spectra. Then the studied traits were predicted on the standardized spectra using international based equations published in the literatures. The averaged predicted content of fat, protein, MFA, PFA, SFA, and UFA were 3.99, 3.53, 1.15, 0.15, 2.64, 1.29 g/dL of milk. GH ranged from 0 to 475. The correlation coefficients estimated between predicted and reference contents ranged from 0.92 to 0.98, except for PFA (0.59). Root mean squared errors (RMSE) were 0.19, 0.18, 0.32, 0.09, 0.21, and 0.39 g/dL for fat, protein, MFA, PFA, SFA, and UFA. The corresponding correlation values between the squared residuals and GH were positive (0.17-0.42) suggesting the need to sort the predictions based on the GH. When a threshold of GH≤3 was applied, the percentage of loss data was 13.03%, 12.26%, 19.53%, 11.88%, and 19.62% for fat, protein, MFA, PFA, SFA and UFA. The correlation coefficients increased with the GH limitation (0.94-0.98, 0.64 for PFA). Globally, the correlation value increased with the quality of the prediction equations, represented by cross-validation coefficient of determination. RMSE decreased by 11.75%, 4.25%, 11.82%, 3.62%, 8.23%, and 8.62% for fat, protein, MFA, PFA, SFA, and UFA. The increase of correlation values and the decrease of RMSE confirm the interest of using GH limitation to ensure the quality of the MIR predictions. However, this limitation allows avoiding spectral extrapolation but this does not mean that the predictions are accurate. To improve the robustness, samples with GH higher than 3 must be added in the calibration set in order to cover the spectral variability. Moreover, to make predictions in an optimal situation, the spectral standardization must be performed on a regular basis in order to check the potential spectral deviation of an instrument. In this study, the

standardization was only performed once a time.

Keywords: milk components, MIR, prediction accuracy, GH