

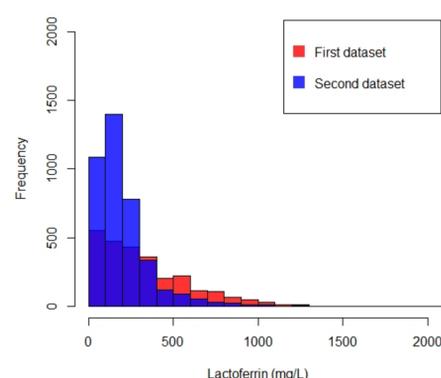
Large scale dataset to improve and validate the prediction of lactoferrin content using milk mid-infrared spectrometry

Soyeurt H.¹, Dehareng F.², McParland S.³, Camels M.⁴, Gengler N.¹, Grelet C.²

¹TERRA research and teaching centre, Gembloux Agro-Bio Tech / University of Liège, Gembloux, Belgium, ²Valorisation of agricultural products, Walloon Research Centre, Gembloux, Belgium, ³Animal & Grassland, Teagasc, Moorepark, Ireland, ⁴Research and development, Seenovia, Saint-Berthevin, France

Context

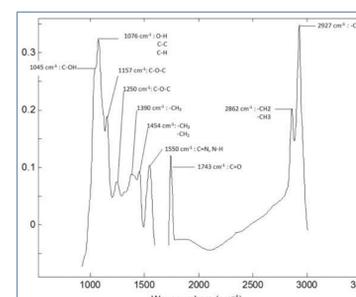
Lactoferrin (LF), a glycoprotein with interesting immunological properties, is measured in milk using ELISA kit which limits its use for management and breeding purposes. Past investigations highlighted the feasibility of predicting an indicator of LF using milk mid-infrared (MIR) spectra.



2 datasets

- 2,654 milk samples [data1]
- 3,965 milk samples [data2]

LF was measured using 2 different ELISA kits. MIR milk spectra obtained from Foss MilkoScan FT6000.



	data1	data2	All
N	2,587	3,689	6,059
R ² cv	0.61 ± 0.03	0.47 ± 0.03	0.63 ± 0.04
RMSEpcv	160 ± 7.8 mg/L	80 ± 2.6 mg/L	99 ± 2.8 mg/L
N validation	3,965	2,602	
R ² validation	0.20	0.46	
RMSEpv	164 mg/L	198 mg/L	

Methods

Partial Least Squares (PLS) regressions were performed to predict the content of LF in milk from milk mid-infrared spectrum.

10-fold cross-validation to assess the robustness.

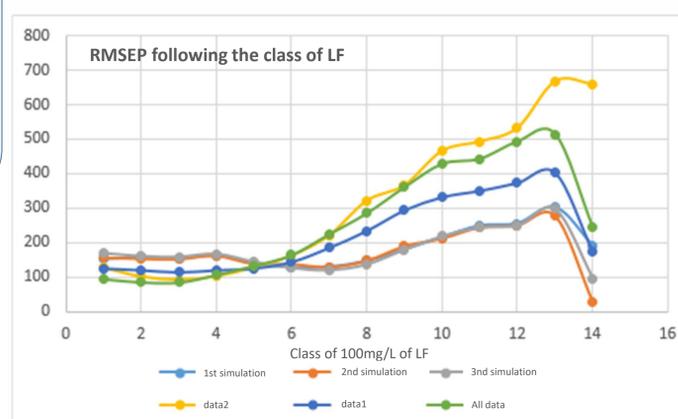
Data cleaning based on the residual values.

	N	R ²	RMSEP
Calibration	989	0.72	81 mg/L
Cross-validation	989	0.68	169 mg/L
Validation	5,293	0.41	157 mg/L



Balanced dataset per classes of LF

The best equation was the one using 100 samples for the 5 first classes of 100 mg of LF per L of milk, all samples in the following classes were kept.



Conclusion

- Even if highest LF contents were better predicted using balanced datasets, RMSE stayed largely higher than the ones observed for the smaller contents. A hypothesis explaining those observations could be that the milk spectra can be different for the samples having effectively LF content higher than 600 mg/L. Therefore the use of LF class dependent models could be a potential interesting improvement.
- The obtained RMSE was largely different between datasets suggesting a reference method error requiring the definition of standard procedure for LF measurement.

Acknowledgements

One dataset was provided by EMR (European Milk Recording EEIG) and included standardisation of MIR spectral data. The other dataset was built thanks to a financial support from the European Commission, Directorate-General for Agriculture and Rural Development, under Grant Agreement 211708; from the Commission of the European Communities, FP7, KBBE-2007-1; and also from the Ministry of Agriculture of the Walloon Region of Belgium (Service public de Wallonie, Direction générale de l'Agriculture, des Ressources naturelles et de l'Environnement, Direction de la Recherche).