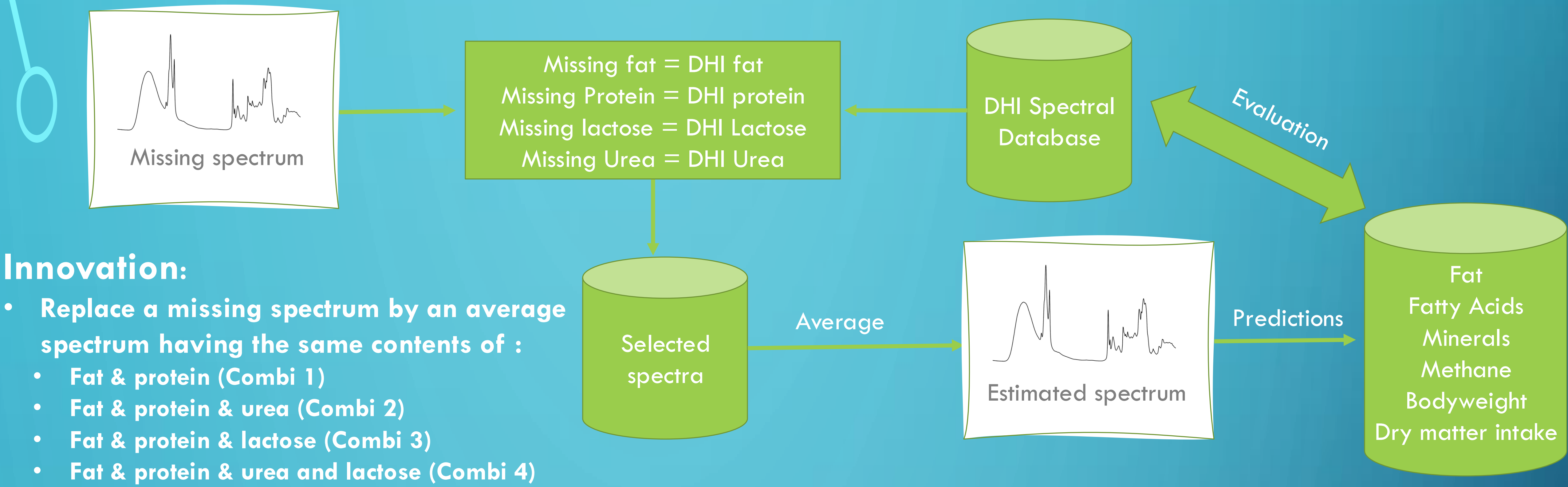


# Leveraging the DHI Databases to Estimate Missing Milk FT-MIR Spectra: A Strategy Toward Improving the Reliability of Breeding Values and Predictive Models

H. Soyeurt<sup>1</sup>, X.-L. Wu<sup>4,5</sup>, C. Grelet<sup>2</sup>, M.L. van Pelt<sup>3</sup>, N. Gengler<sup>1</sup>, F. Dehareng<sup>2</sup>, C. Bertozzi<sup>6</sup>, and J. Burchard<sup>4</sup>

<sup>1</sup> Research and Teaching Centre (TERRA), Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium; <sup>2</sup> Walloon Agricultural Research Center, 5030 Gembloux, Belgium; <sup>3</sup> Cooperation CRV, Animal Evaluation Unit, PO Box 454, 6800 AL Arnhem, the Netherlands; <sup>4</sup> Council of Dairy Cattle Breeding, Bowie, MD 20716, USA; <sup>5</sup> Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, USA; <sup>6</sup> Walloon Breeders Association, 5590 Ciney, Belgium.



## Discussion :

- Combi 1 and 3 gave the best performances
- Urea did not bring information
- If the number of merging criteria increases → lower number of selected spectra
- Better performances were reached if we used a tolerance of 0.1 around the lactose content (combi 3b). This combination applied to the full dataset gave the best predictions for the content of saturated fatty acids (Mean absolute error (MAE) = 0.15 g/dL of milk) and body weight (MAE = 12.80 kg).

Data	Variable	N	Mean	SD	Minimum	Maximum
Full dataset	Fat (g/dL milk)	6,625,570	4.10	0.76	0.10	18.46
	Protein (g/dL milk)	6,625,570	3.45	0.42	0.21	15.50
	Lactose (g/dL milk)	6,358,998	4.71	0.24	0.01	8.94
	Urea (mg/L)	6,621,432	248.50	86.92	1.00	2000.00
Subset	Fat (g/dL milk)	2,000,000	4.14	0.76	0.13	18.46
	Protein (g/dL milk)	2,000,000	3.48	0.41	0.50	13.05
	Lactose (g/dL milk)	1,985,770	4.75	0.22	0.69	5.67
	Urea (mg/L)	1,998,759	257.98	88.07	1.00	1870.00
Samples to predict	Fat (g/dL milk)	1,700	4.08	0.64	1.94	6.64
	Protein (g/dL milk)	1,700	3.46	0.41	2.41	7.14
	Lactose (g/dL milk)	1,700	4.72	0.22	2.90	5.25
	Urea (mg/L)	1,699	206.46	71.34	20.00	570.00

However, the predictions for the unsaturated fatty acids were less satisfactory (MAE = 0.13 and 0.018 g/dL of milk for monounsaturated and polyunsaturated fatty acids) mainly because of the poorer predictions of spectral regions related to the unsaturation of carbon chain. The same reason also accounted for the poorer methane predictions (MAE = 47.02 g/day).

Combination	1	2	3	4	1	2	3	4
	Number of matched samples				%Loss of samples			
	1,692	946	972	29	0.47	44.35	42.82	98.29
	Correlation				MAE			
%Fat	0.95	0.89	0.93	0.73	0.097	0.081	0.109	0.135
Methane	0.66	0.31	0.45	0.20	61.98	61.83	71.45	84.12
Saturated fatty acids (FA)	0.89	0.77	0.81	0.76	0.20	0.19	0.21	0.20
Monounsaturated FA	0.63	0.30	0.35	0.21	0.14	0.16	0.15	0.15
Polyunsaturated FA	0.68	0.40	0.39	0.18	0.024	0.019	0.026	0.023
Body weight	0.94	0.91	0.92	0.84	13.50	16.37	15.17	20.60
Dry matter intake	0.63	0.42	0.41	0.37	1.84	2.24	2.21	2.92
Calcium	0.66	0.36	0.40	0.02	68.57	81.95	79.98	86.94
Sodium	0.31	0.07	0.60	0.17	39.26	44.59	25.49	22.89
Phosphorus	0.51	0.27	0.34	0.10	75.75	84.64	82.49	88.60
Lactoferrin	0.37	0.11	0.23	0.20	80.41	99.06	81.67	110.51

## Conclusions :

- The matching criteria must have a high quality
- Finding new matching criteria could improve the performances
- Use only the most predictable regions