Large-scale phenotyping from milk MIR spectra:

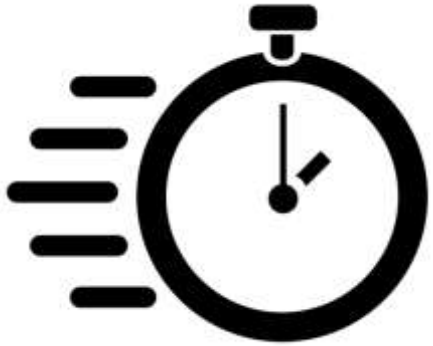
challenges to obtain reliable predictions

Grelet C. [1], Dardenne P. [1], Soyeurt H. [2], Fernandez J.A. [1], Gengler N. [2], Vanlierde A. [1], Dehareng F. [1]

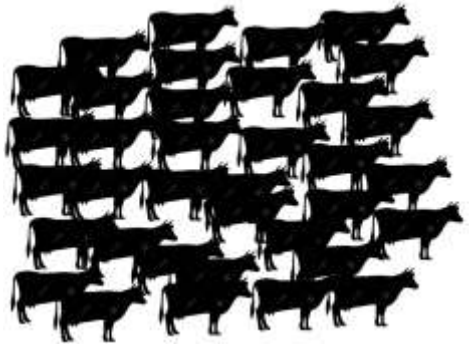
[1] *Walloon Agricultural Research Centre, B-5030 Gembloux, Belgium*
[2] *Gembloux Agro-Bio Tech, ULiège, B-5030 Gembloux, Belgium*

Context

Prediction of phenotypes by MIR
- -Fast
- -Cost effective
- -Easy to use in routine



Potentially usable for large scale applications
- -Management of cows
- -Genetic studies



→ Exponential researches to create MIR models

Context

Milk quality

Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry

H. Soyeurt,*†[1,2] P. Dardenne,‡ F. Deh
P. Mayeres,*[2] and N. Gengler*†[2]

Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry

H. Soyeurt,*[1] D. Bruwier,* J.-

Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics

A. McDermott,*† G. Visentin,*† M. De Marchi,† D. P. Berry,* M. A. Fenelon,‡ P. M. O'Connor,‡ O. A. Kenny,‡ and S. McParland*

Technological properties

Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy

M. De Marchi,*[1] C. C. Fagan,† C. P. O'Donnell,† A. Cecchinato,* R. Dal Zotto,* M. Cassandro,* M. Penasa,* and G. Bittante*

Cow phenotype

Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows

F. Dehareng[1*]
A. Vanlierde[1]

Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis

H. Soyeurt[1,2†]
F. Dehareng[6],
M. Coffey[7], L.

The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows[1]

S. McParland

Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network

C. Grelet,*[1] C.
F. G. Colinet,

Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk

N. Shetty,[1] P. L

Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra

A. Lainé,* C. Bastin,*[1] C. Grelet,† H. Hammami,* F. G. Colinet,* L. M. Dale,*[2] A. Gillon,‡ J. Vandenplas,*[§1] F. Dehareng,† and N. Gengler*[#]

Outliers, detection of contaminants

Use of a multivariate moving window PCA for the untargeted detection of contaminants in agro-food products, as exemplified by the detection of melamine levels in milk using vibrational spectroscopy

J.A. Fernández Pierna, D. Vincke, V. Baeten, C. Grelet, F. Dehareng, P. Dardenne *

Milk origin determination

Building of prediction models by using Mid-Infrared spectroscopy and fatty acid profile to discriminate the geographical origin of sheep milk

Marco Caredda [a], Margherita Addis [a], Ignazio Ibba [b], Riccardo Leardi [c], Maria Francesca Scintu [a], Giovanni Piredda [a], Gavino Sanna [a,*]

However…

Huge difference between

Developing a model in a research context

- Objective:
  - Evaluate a potential
  - Publication
- Development
  - Research herds
  - **With one or few herds, diets, breeds, countries, MIR instrument**
- Evaluation
  - **Performances** (highest R^2, l

≠

Using a model to generate predictions at a large scale

- Objective:
  - Generate correct predictions in all cases
- Evaluation:
  - **<u>Robustness:</u> capacity to be "all terrain" and provide good results in various conditions**

Potential issue when using research models

However…

Huge difference between

Using a model to generate predictions at a large scale

- Objective:
  - Generate correct predictions in all cases
- Evaluation:
  - **<u>Robustness:</u> capacity to be "all terrain" and provide good results in various conditions**

Objective…

Evaluate the impact of different factors on Robustness :

- Inclusion of variability in the model (breeds, days in milk…)
- Extrapolation (& sampling scheme)
- Model development (spectral areas)
- Spectral standardization

Evaluated by :

- Error in external validation (RMSEP)

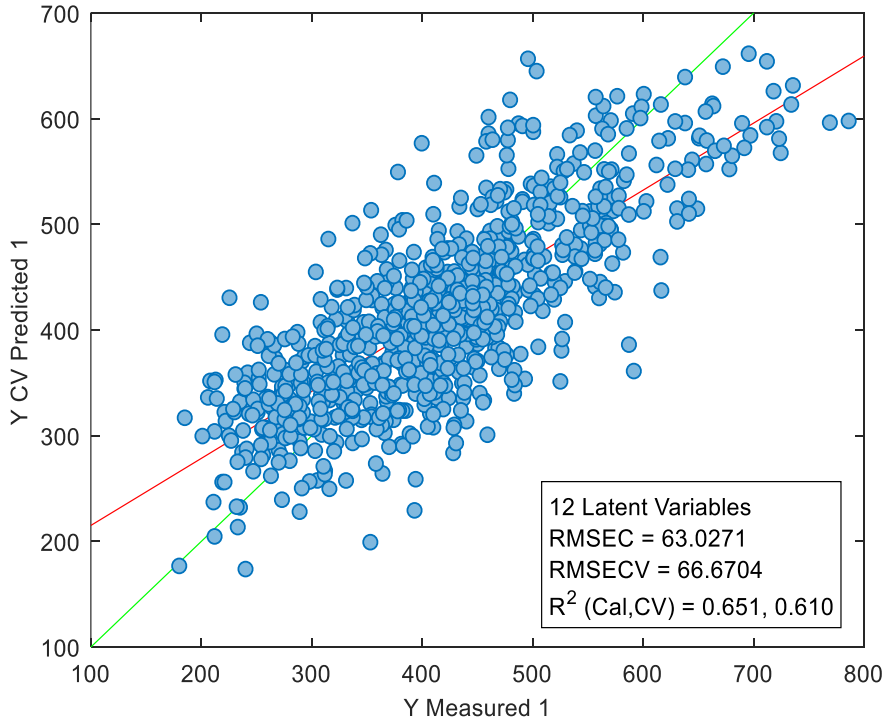
Inclusion of Variability

Effect of breeds in the model

Dataset used: CH_4 by dairy cows

- 225 Holsteins

Step 1 : calibration with 225



RMSEcv = 67 g/d

External validation with 20

RMSEP = 85 g/d

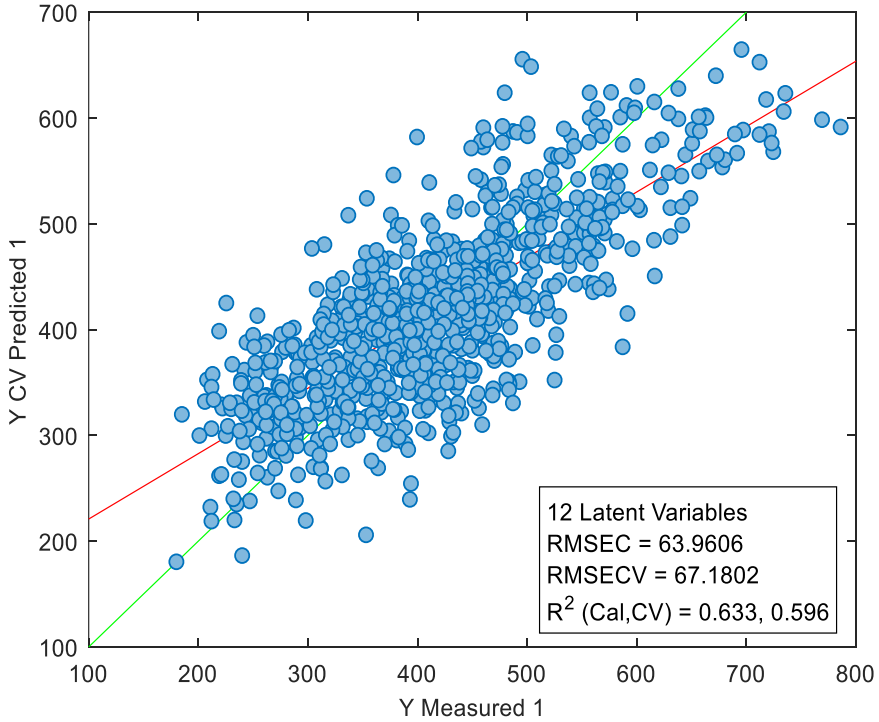
Effect of breeds in the model

Dataset used: CH_4 by dairy cows

- 225 Holsteins

Step 2 : calibration with 225

+ 19



RMSEcv = 67 g/d

External validation with 20

RMSEP = 69 g/d

-19%

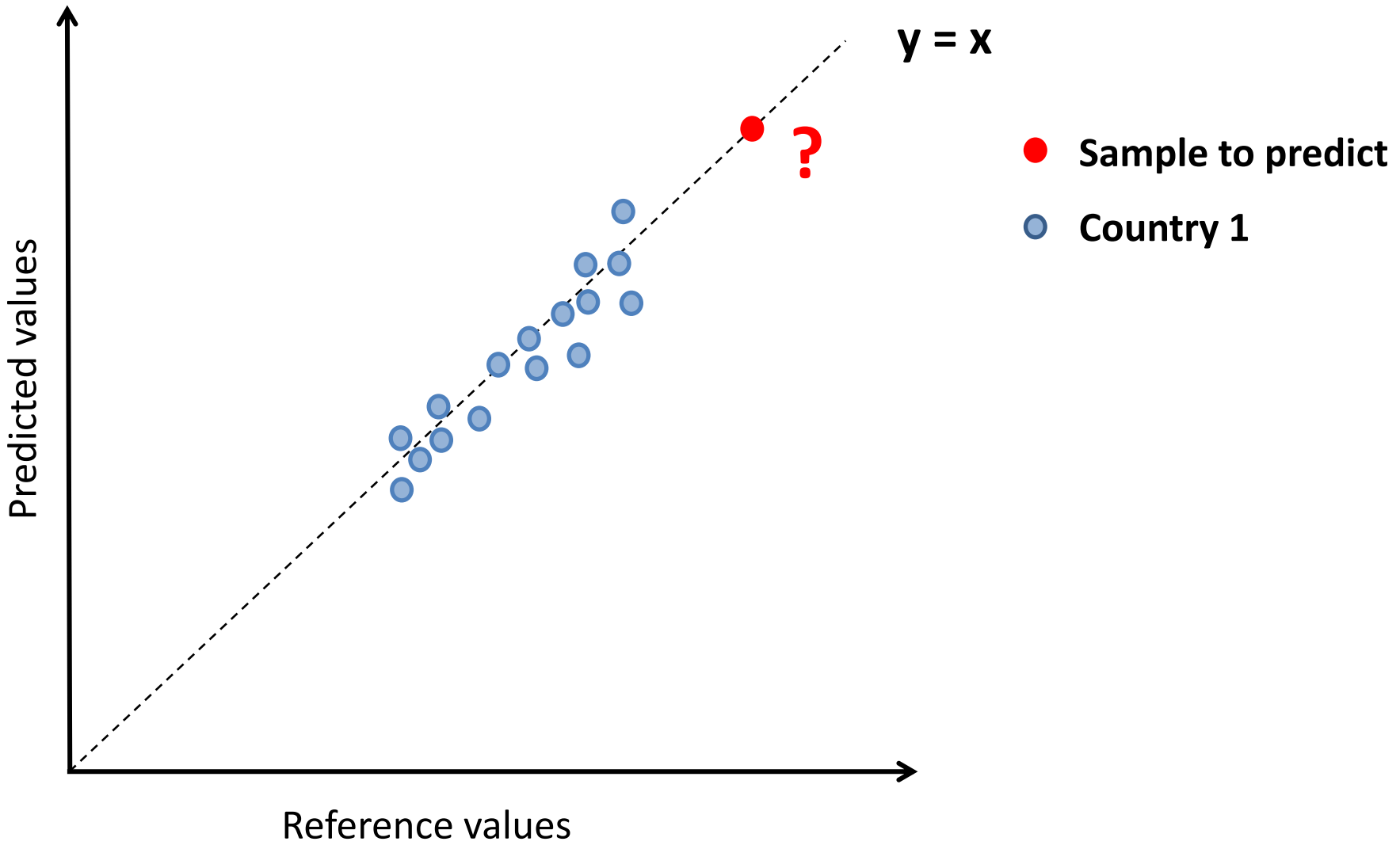
“IR models can only predict what they know”

Pierre Dardenne

“Extrapolation is dangerous!”

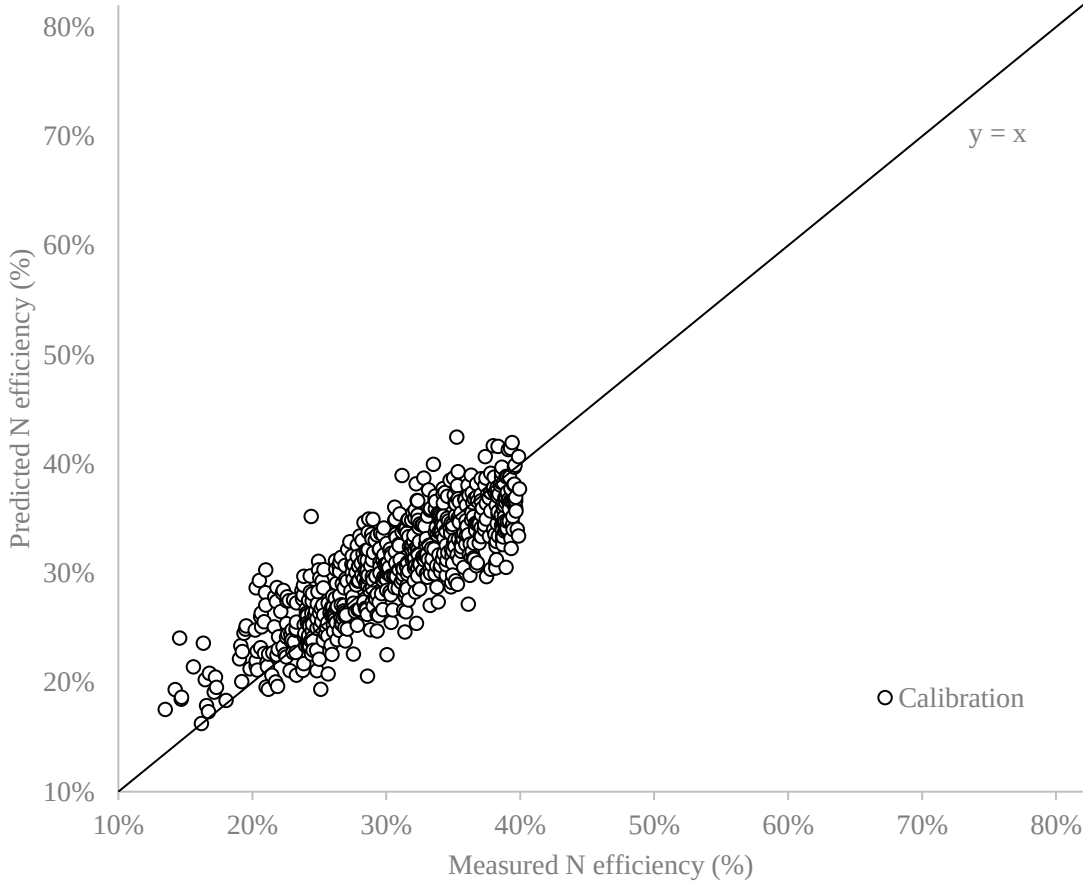
IR maxim

Cover the Y (reference data) range

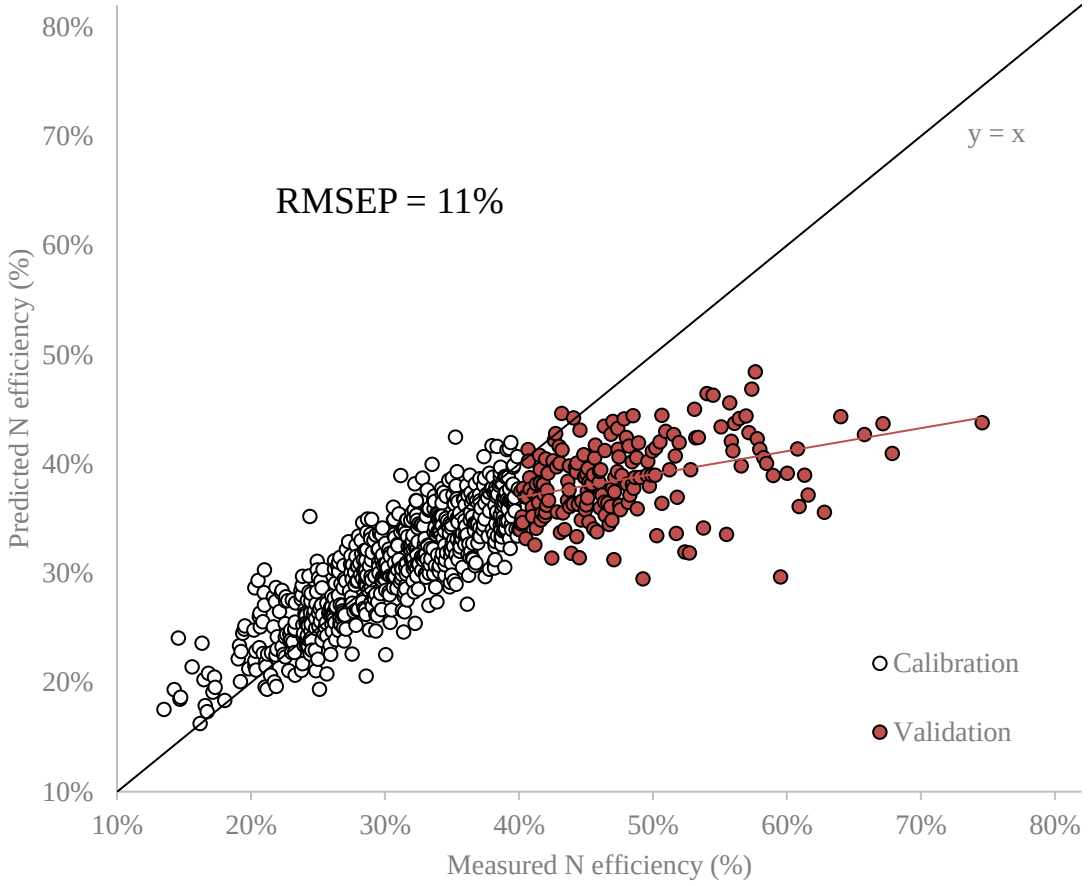


Cover the Y (reference data) range – *test with N efficiency model*

Calibration with expected values : 10 to 40%

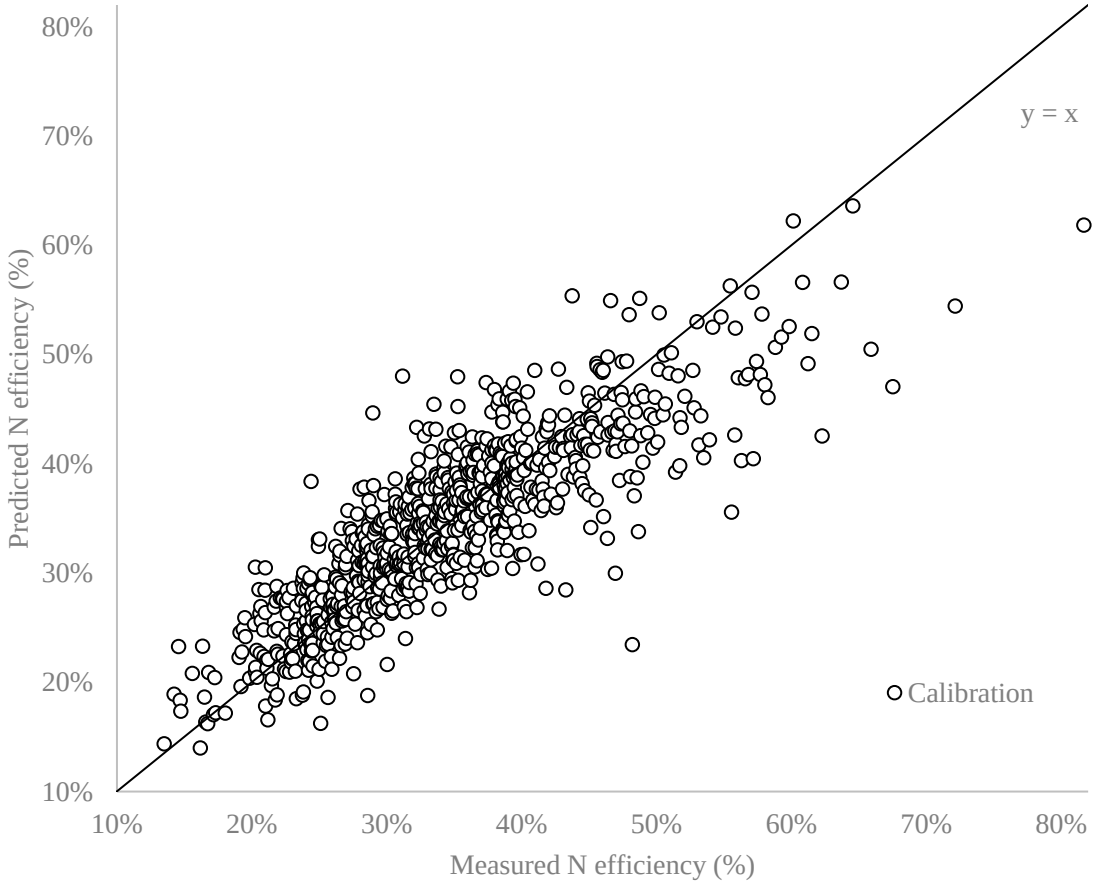


Test with extreme high values

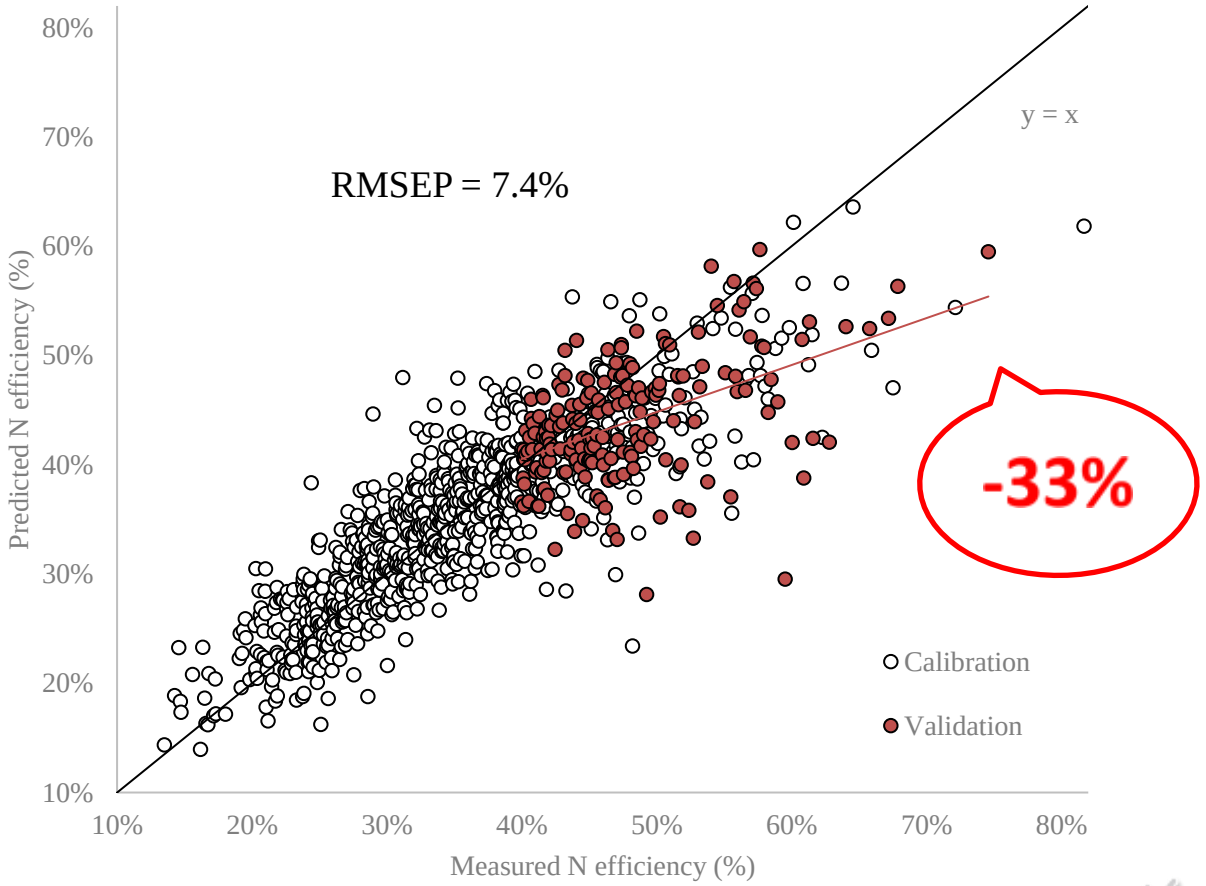


Cover the Y (reference data) range – *test with N efficiency model*

Calibration with expected and extreme high values : 10 to 80%



Test with extreme high values

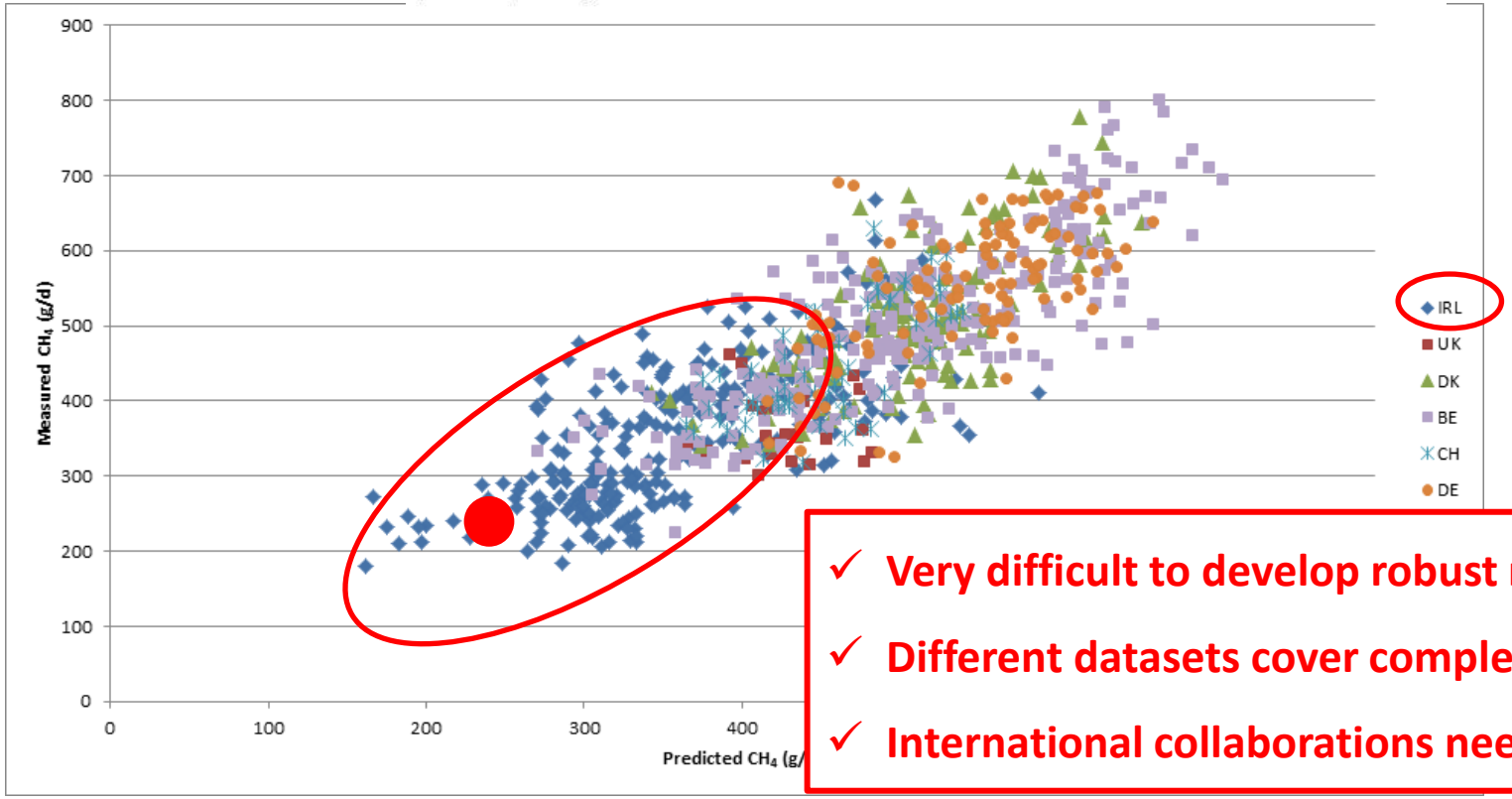


Methane model :

J. Dairy Sci. 98:5740–5747
http://dx.doi.org/10.3168/jds.2014-8436

Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra

A. Vanlierde,*[1] M.-L. Vanrobays,†[1] F. Dehareng,* E. Froidmont,‡ H. Soyeurt,† S. McParland,§ E. Lewis,§ M. H. Deighton,# F. Grandl,‖ M. Kreuzer,‖ B. Gredler,¶ P. Dardenne,* and N. Gengler†[2]
*Walloon Agricultural Research Centre, Valorization of Agricultural Products Department, 5030 Gembloux, Belgium
†Agriculture, Bio-engineering and Chemistry Department, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
‡Walloon Agricultural Research Centre, Production and Sectors Department, 5030 Gembloux, Belgium
§Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland
#Agriculture Research Division, Department of Economic Development, Jobs, Transport and Resources, Ellinbank Centre, Ellinbank, 3821 Victoria, Australia
‖ETH Zürich, Institute of Agricultural Sciences, 8092 Zürich, Switzerland
¶Qualitas AG, 6300 Zug, Switzerland

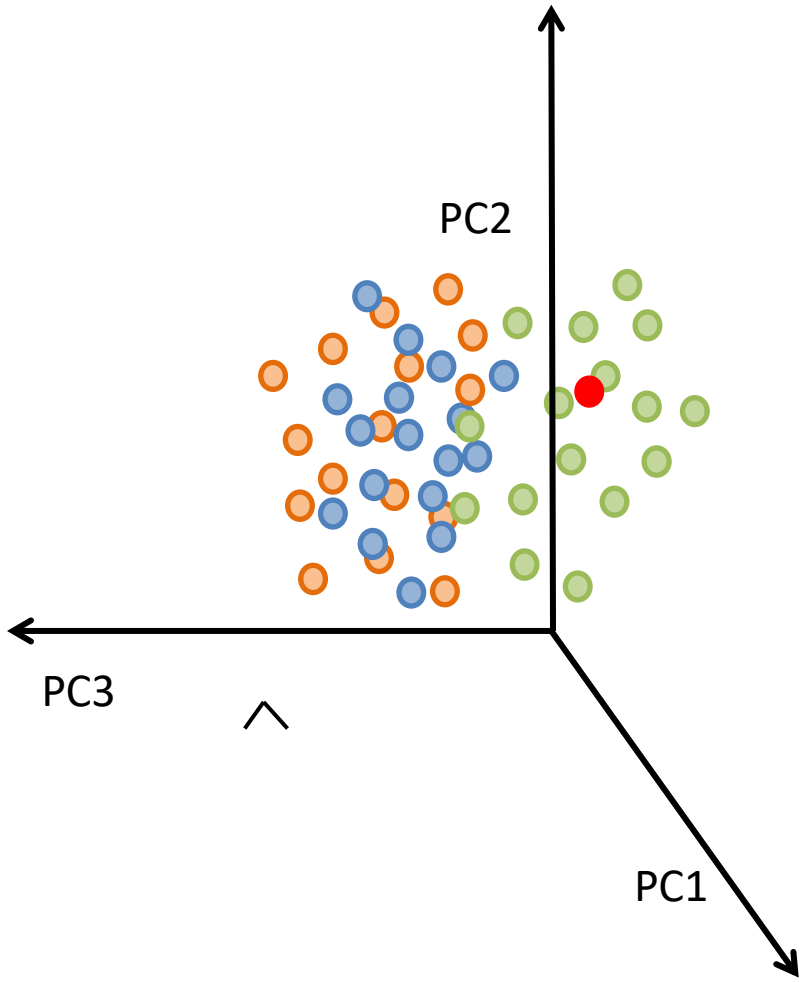


- ✓ **Very difficult to develop robust model in only one country!**
- ✓ **Different datasets cover complementary variability**
- ✓ **International collaborations needed!**

Constituent	N	Mean	SD	R^2c	R^2cv	SEC	SECV
CH_4	863	459	123	0.71	0.67	66	71

METHAGENE

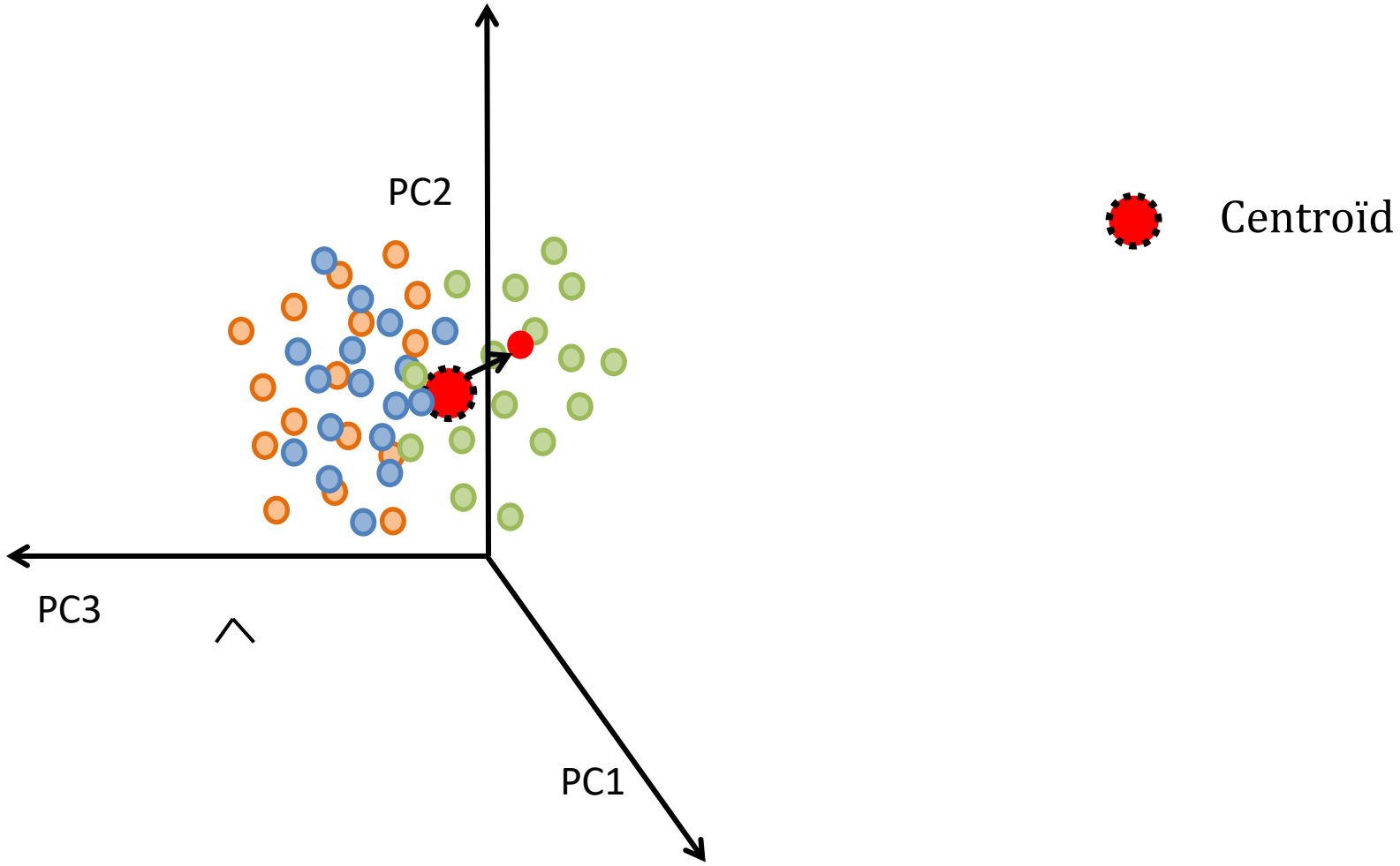
Cover the X (spectral data) range



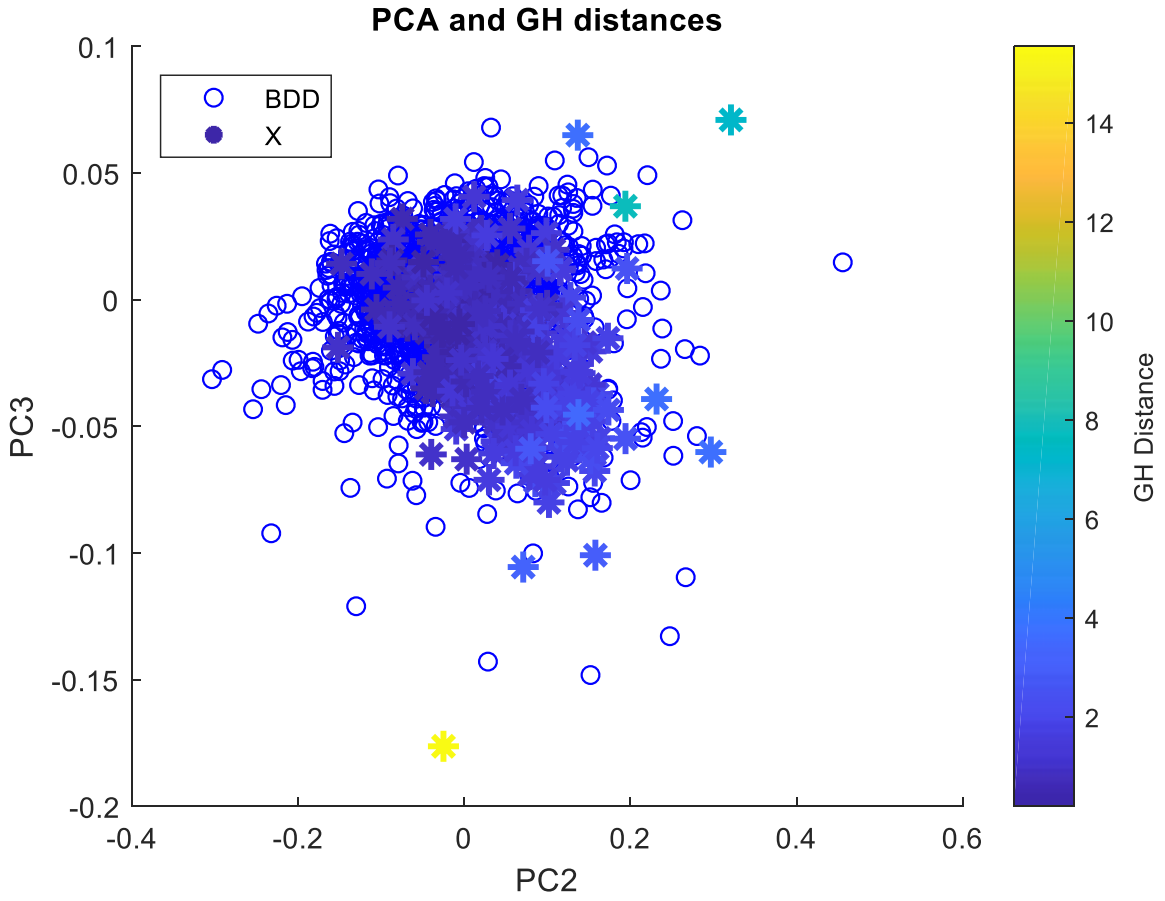
- **Sample to predict**
- Country 1
- Country 2
- Country 3

Mahalanobis distances (GH):

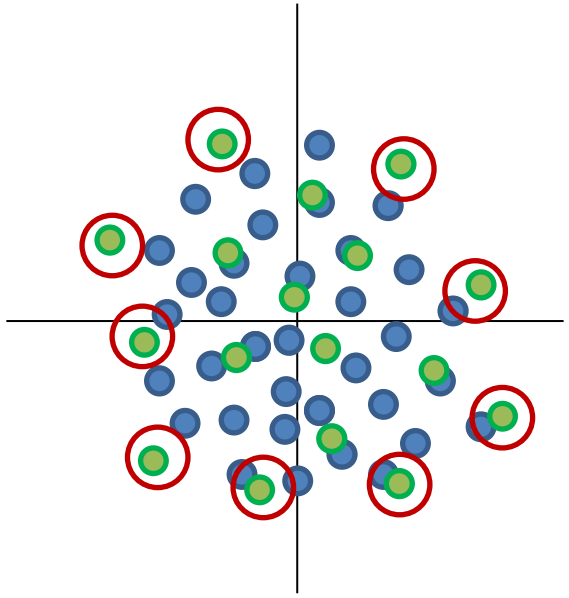
Distance of a sample to the centroid of the dataset



Do we want extreme samples (high GH) in the Calibration datasets??



What samples to select/ to keep?



Effect of sampling method

Dataset used: Lactoferrin

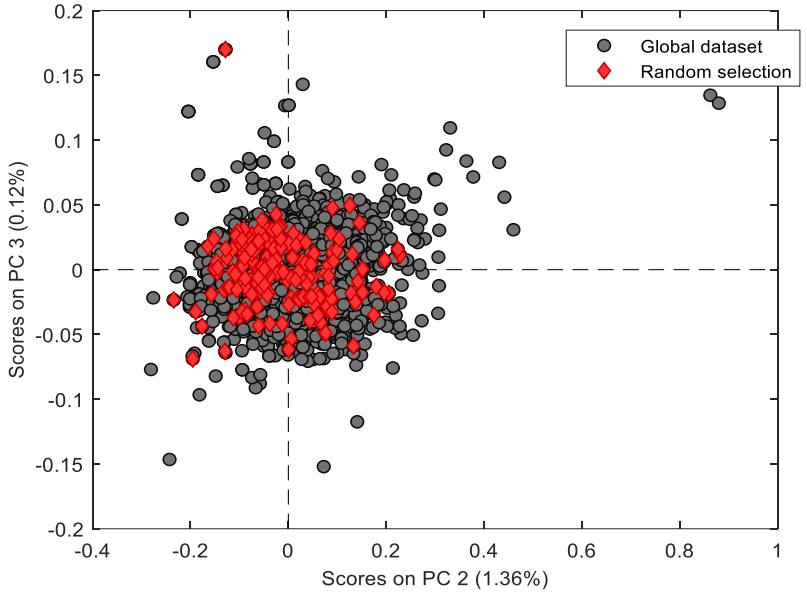
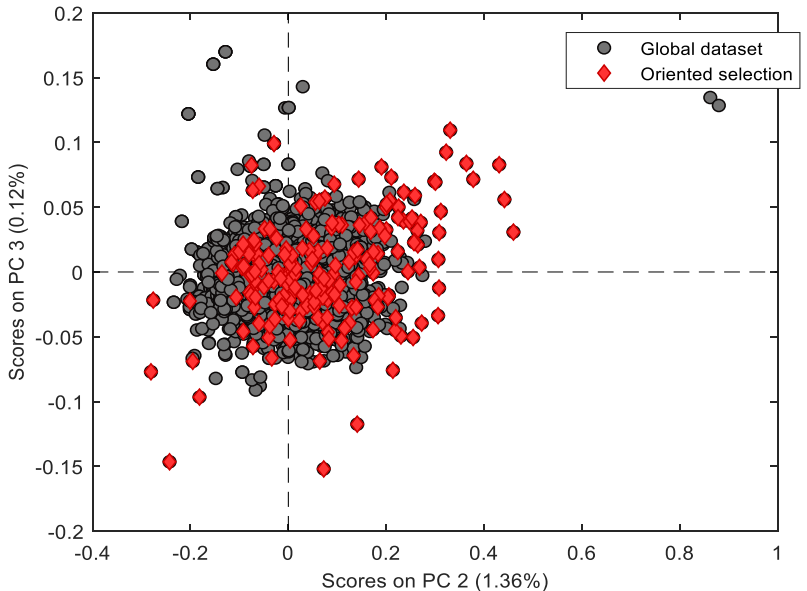
- 3506 as a global calibration population

  →Selection of 200 samples to develop a model

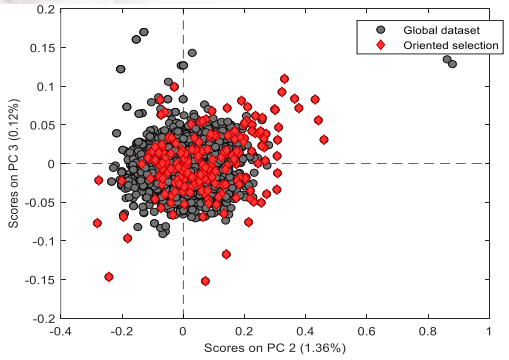
→ Random selection

→ Oriented selection, to cover the spectral variability (extreme GH)

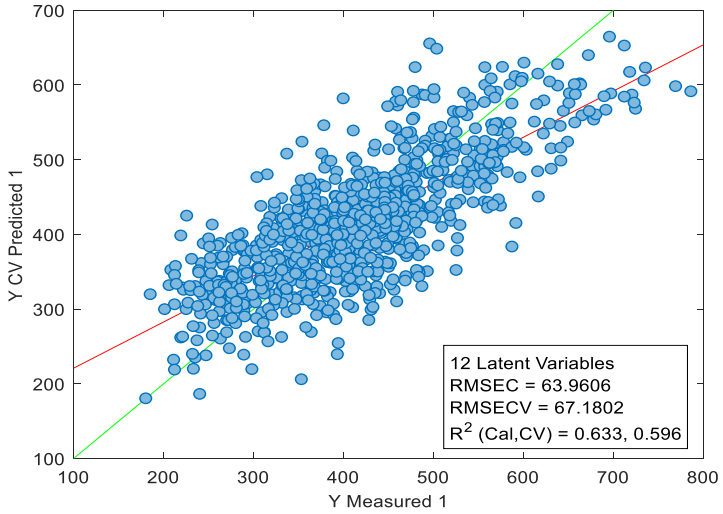
External validation with 400 samples

	Selection	Cross-validation	External-validation (400 external samples)
Random selection		RMSEcv = 126 g/L	• **RMSEP = 170 g/L** • 6.6% samples out of the range (GH>3)
Oriented selection		RMSEcv = 176 g/L	• **RMSEP = 146 g/L** • 1.6% samples out of the range (GH<3)

IR models can only predict what they know

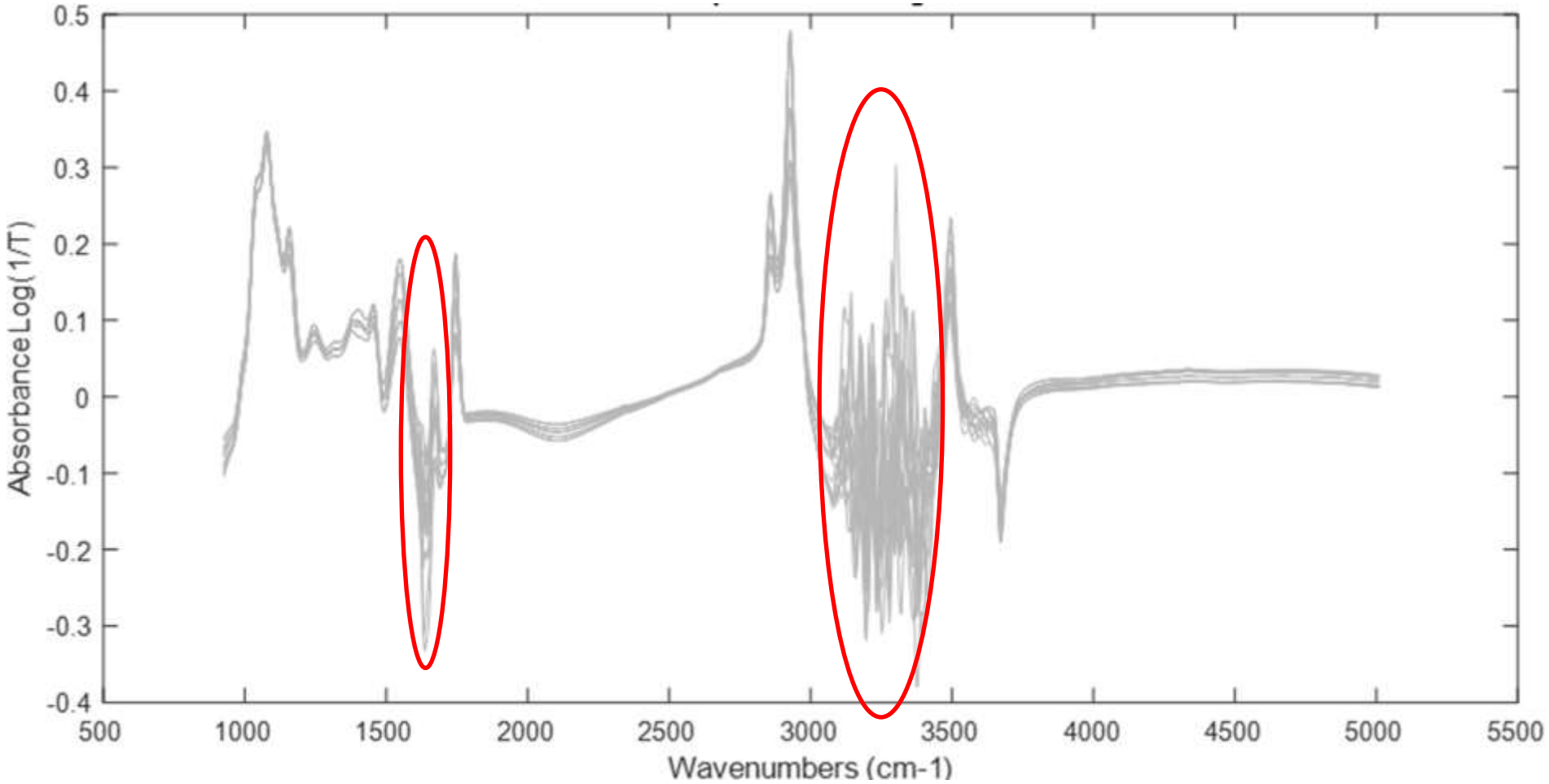


Robust models

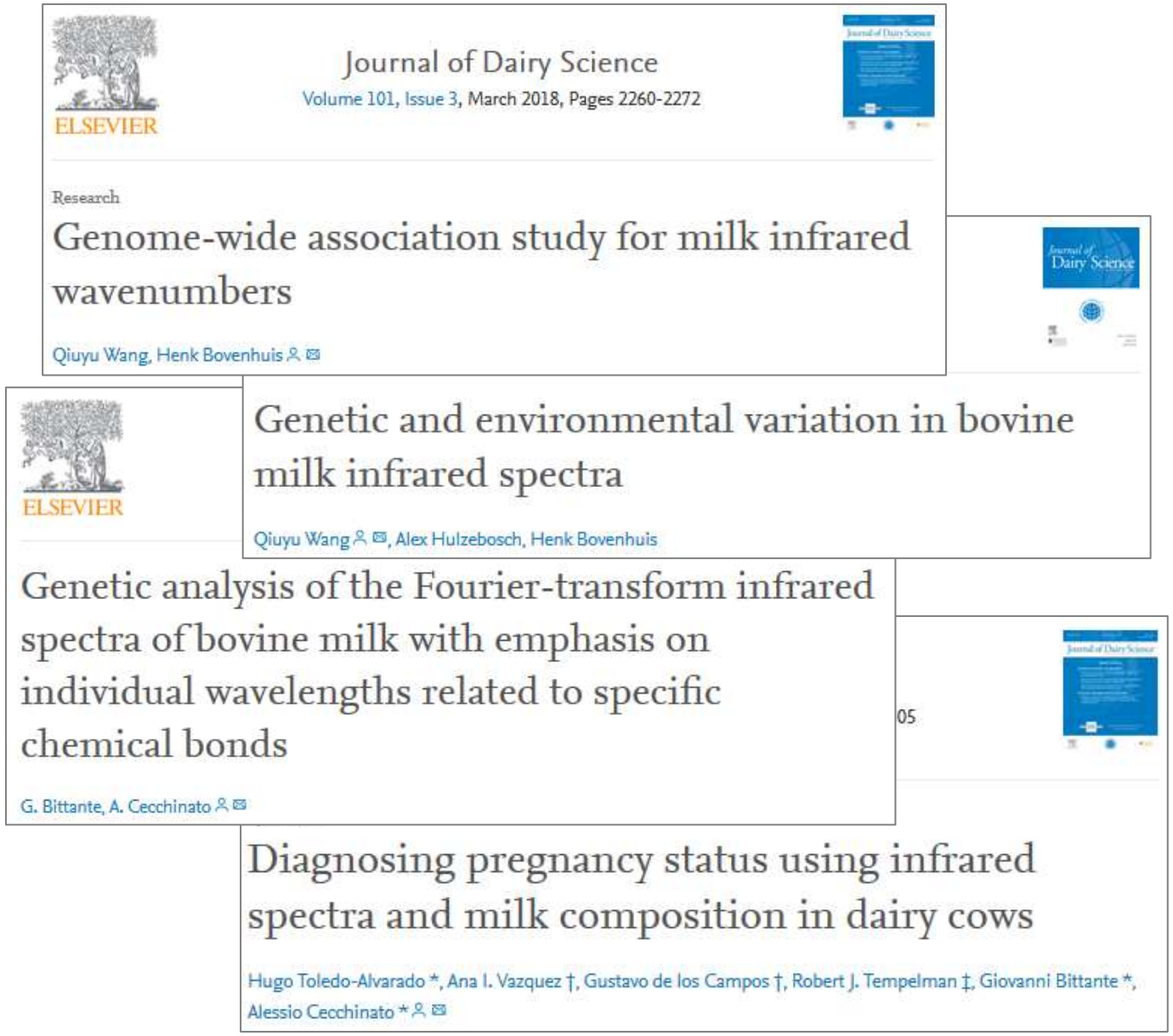


Effect of model development: Wavenumber selection

Noisy areas induced by water absorption
→ usually considered without valuable information and not used



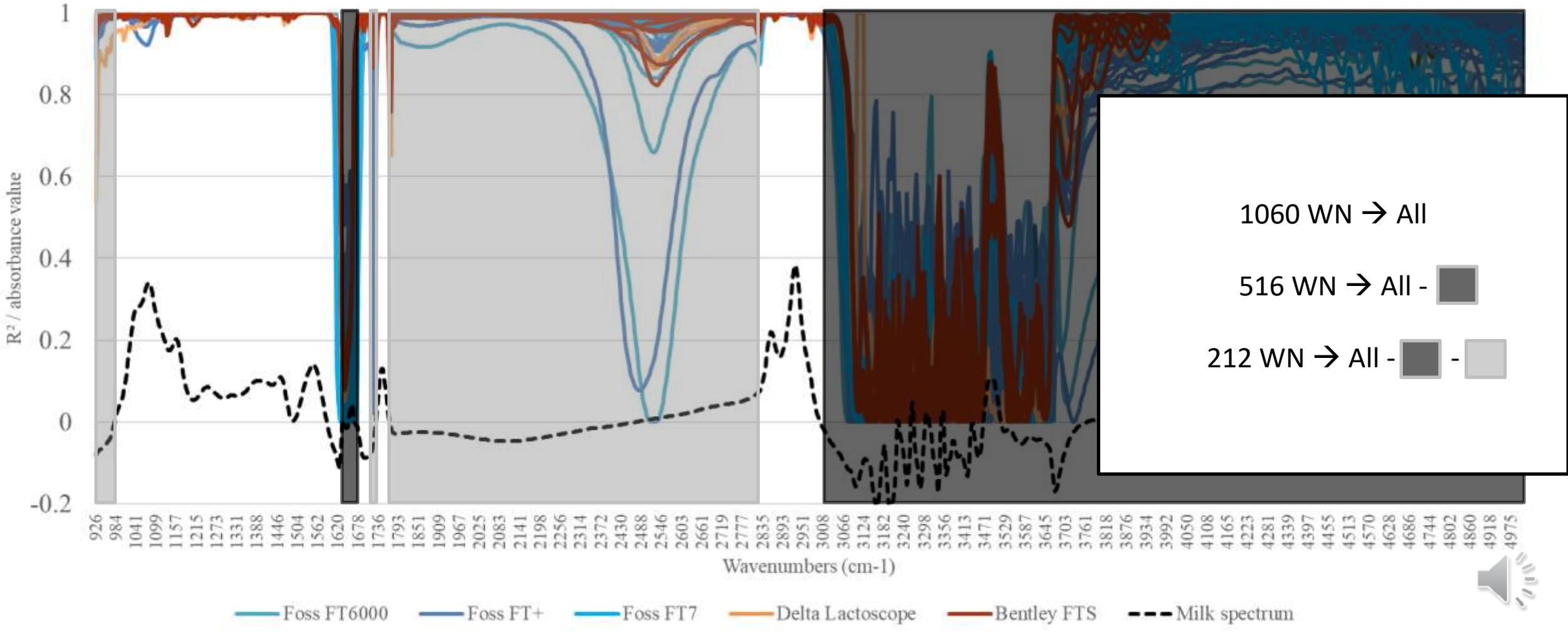
Effect of model development: Wavenumber selection



But recent studies concluding with the presence of valuable information within those noisy regions

Effect of model development: Wavenumber selection

5 identical samples analyzed on 83 instruments (72 Foss + 10 Bentley + 1 Delta)
For each wavenumber, correlation between the absorbance values of a reference and the others instruments



Effect of model development: Wavenumber selection

Dataset used: C18_1 cis9 fatty acid

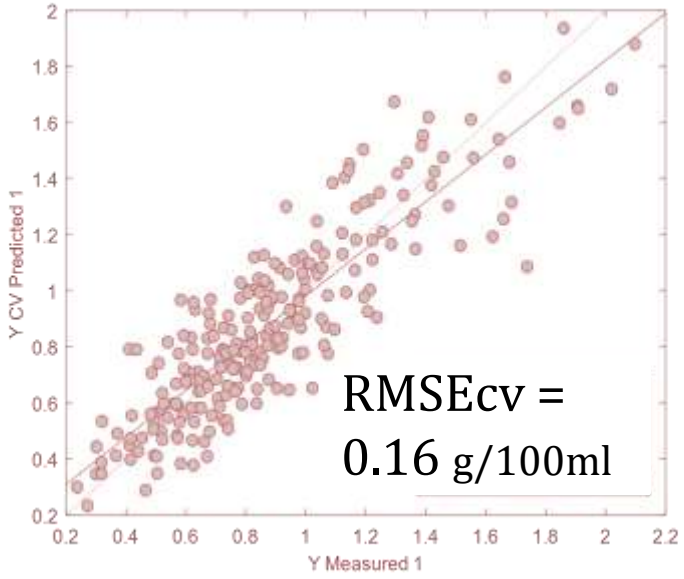
- 250 samples in calibration
- 1572 samples in external validation

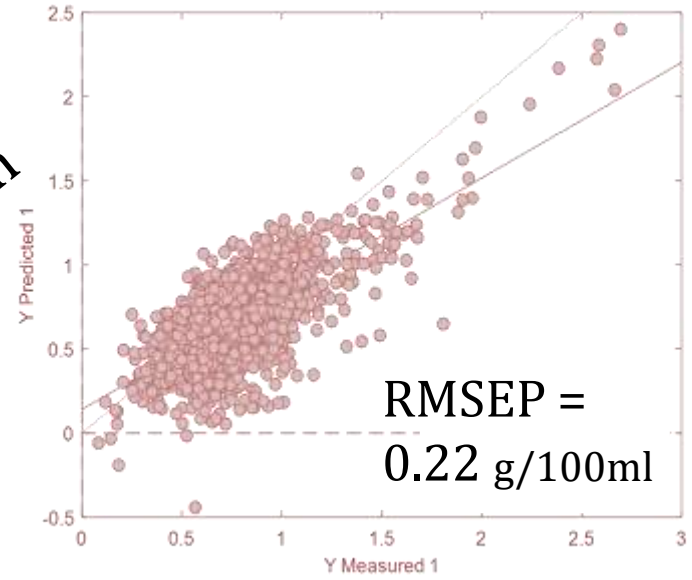
Same number of PLS factors

Calibration

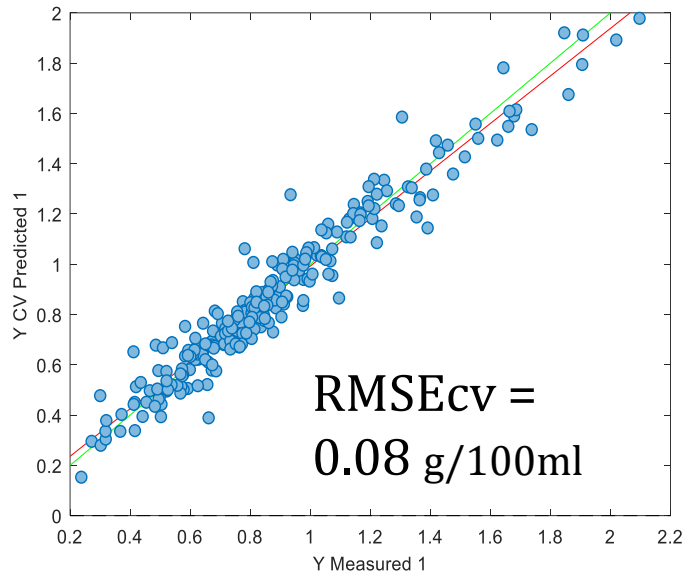
External validation

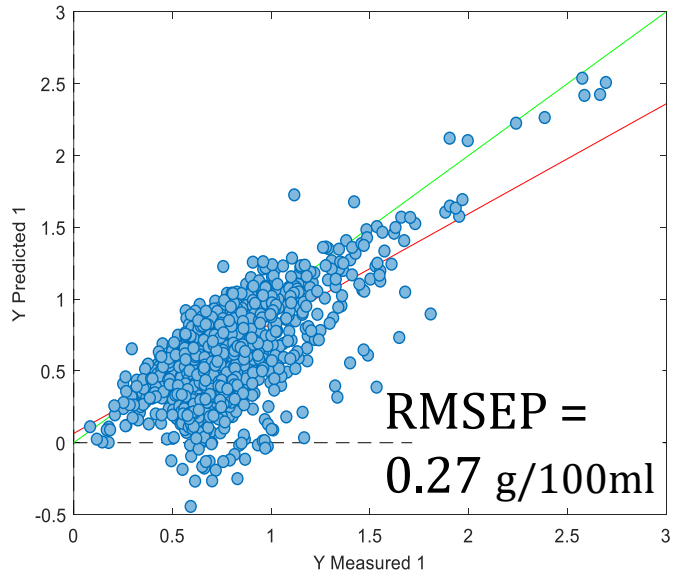
1060 WN



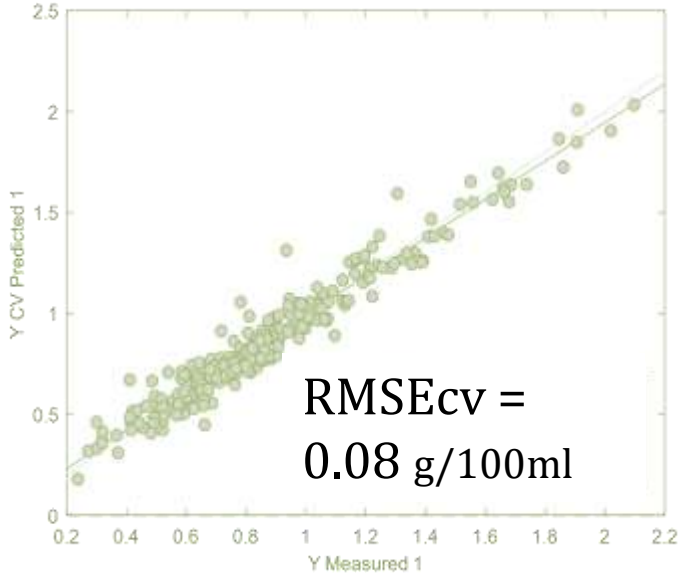


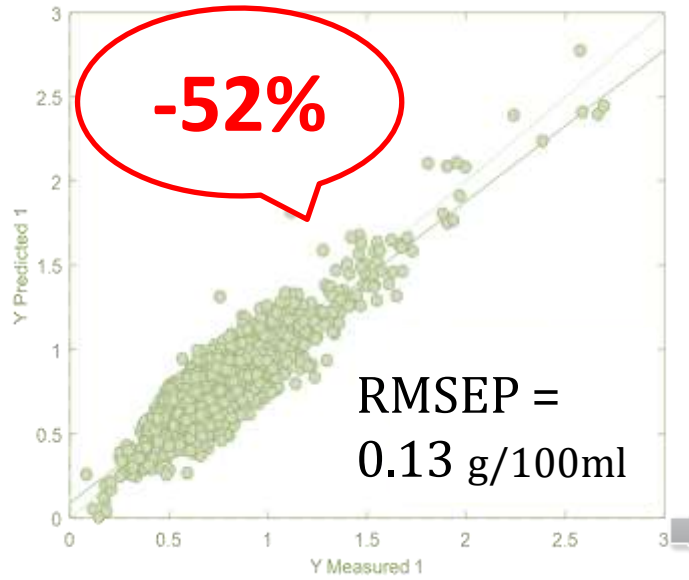
516 WN





212 WN

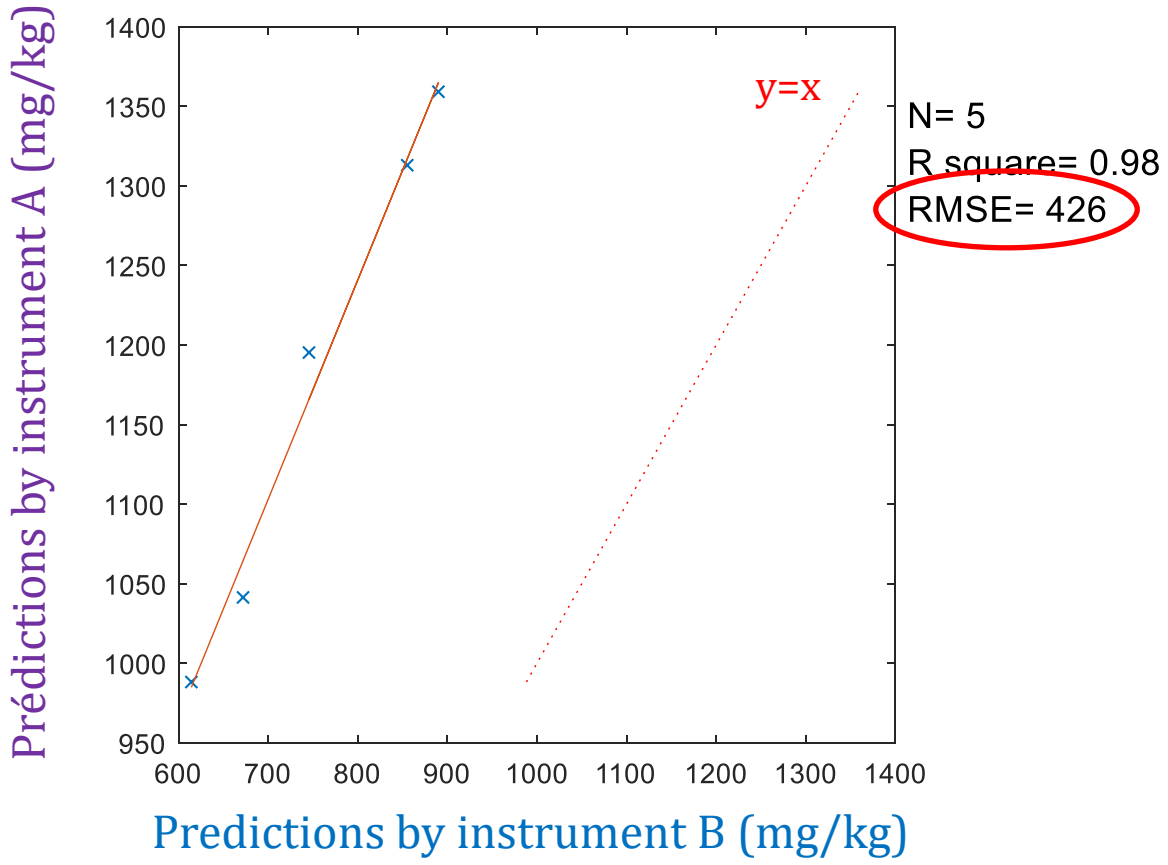




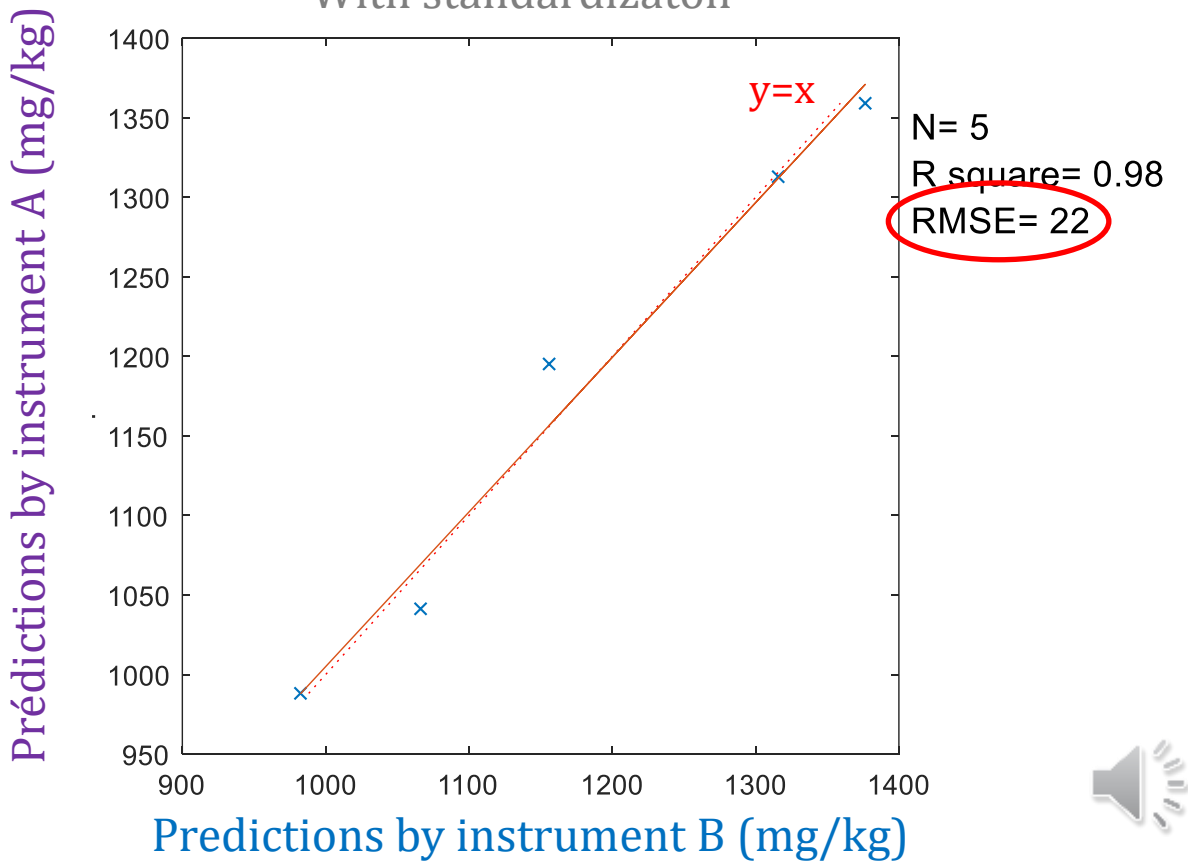
Spectral standardization

- Calcium model developed on instrument A
- Model applied on instrument B after analysis of common samples

Without standardization

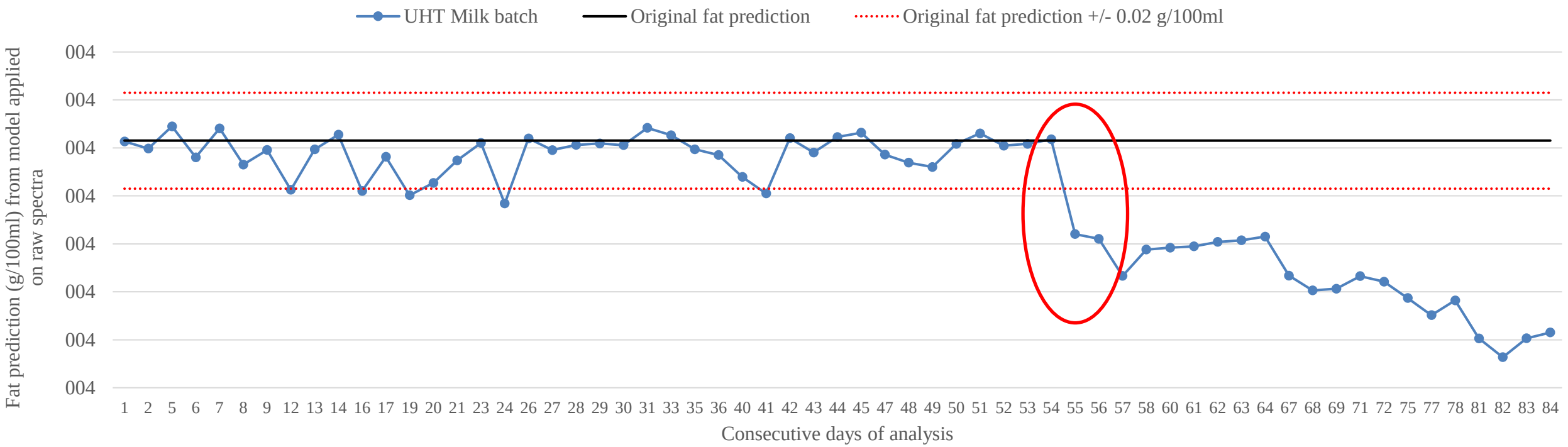


With standardizaton



Spectral Daily monitoring

- Monitoring of the <u>daily spectral stability</u> of each individual instrument in between two standardization



Take home message

This is only examples, with specific datasets…

…but highlight some elements to take into account

- ✓ Look for variability (reference and spectral data)
- ✓ **Collaboration to merge datasets!!!**
- ✓ Reliability of spectral areas
- ✓ Standardization & spectral monitoring

Complementary information

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions

C. Grelet[a], P. Dardenne[a], H. Soyeurt[b], J.A. Fernandez[a], A. Vanlierde[a], F. Steevens[a], N. Gengler[b], F. Dehareng[a,*]

[a] Walloon Agricultural Research Center (CRA-W), 24 Chaussée de Namur, 5030 Gembloux, Belgium
[b] TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium


Thank you for your attention!
SAVE THE DATE
27 > 29/04/2022
in Namur, Belgium
DAIR'INNOV
congress
Innovations to benefit cow welfare
and dairy farming sustainability
www.dairinnov.eu
Interreg
North-West Europe
HappyMoo
European Regional Development Fund
D4Dairy