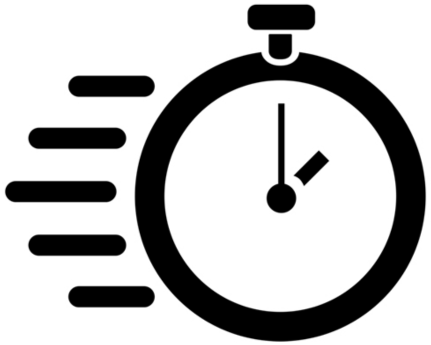
New phenotypes from milk MIR spectra:

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]

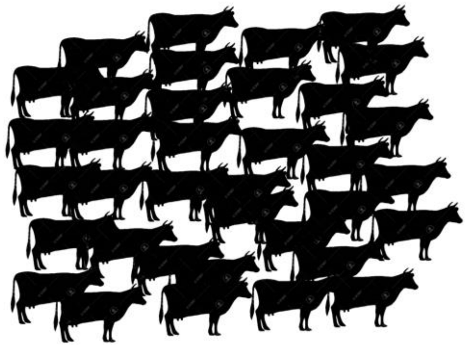
[1] Walloon Agricultural Research Centre, B-5030 Gembloux, Belgium
[2] Gembloux Agro-Bio Tech, ULiège, B-5030 Gembloux, Belgium

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[5], 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,*§[3] F. Dehareng,† and N. Gengler*[4]

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

PDO PGI

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 [d,*]

However...

Huge difference between

Developing a model in a research context	Using a model to generate predictions at a large scale

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 instruments**
- Evaluation
  - **Performances** (highest R^2, RMSE)

Using a model to generate predictions at a large scale

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**

However...

Huge difference between

Developing a model in a research context

- Optimized performances (highest R^2, RMSE)
- Limited variability

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

Objective…

Evaluate the impact of different factors on Robustness :

- Inclusion of variability in the model (breeds, days in milk…)
- Sampling scheme (oriented vs. random)
- Model development (spectral areas, PLS factors)
- Spectral standardization

Evaluated by :

- Error in external validation (RMSEP)

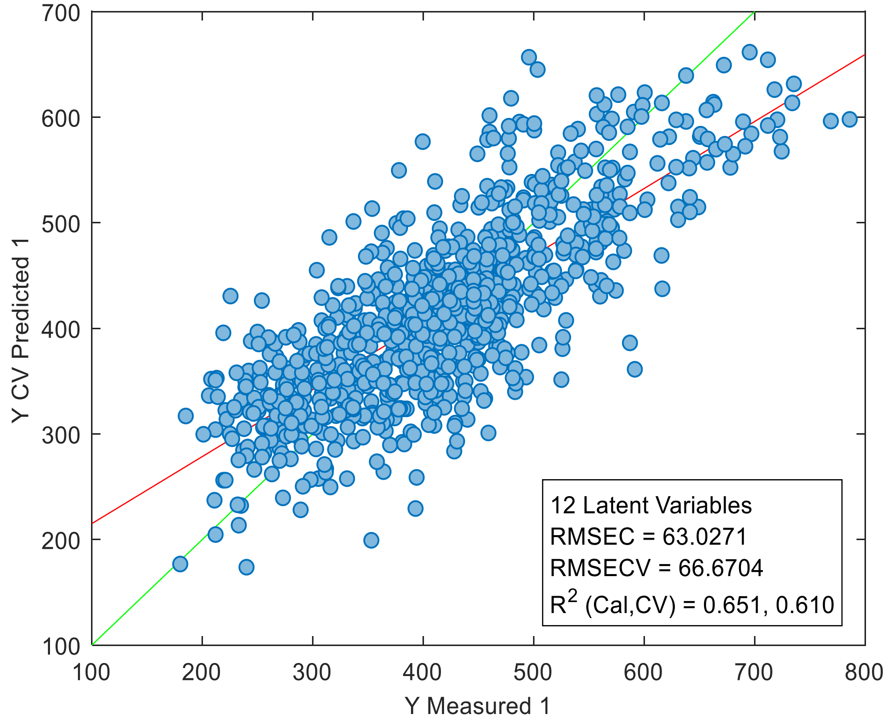
Inclusion of Variability

Effect of breeds in the model

Dataset used: CH4 by dairy cows

- 225 Holsteins

Step 1 : calibration with 225



RMSEcv = 67 g/d

External validation with 20

RMSEP = 85 g/d

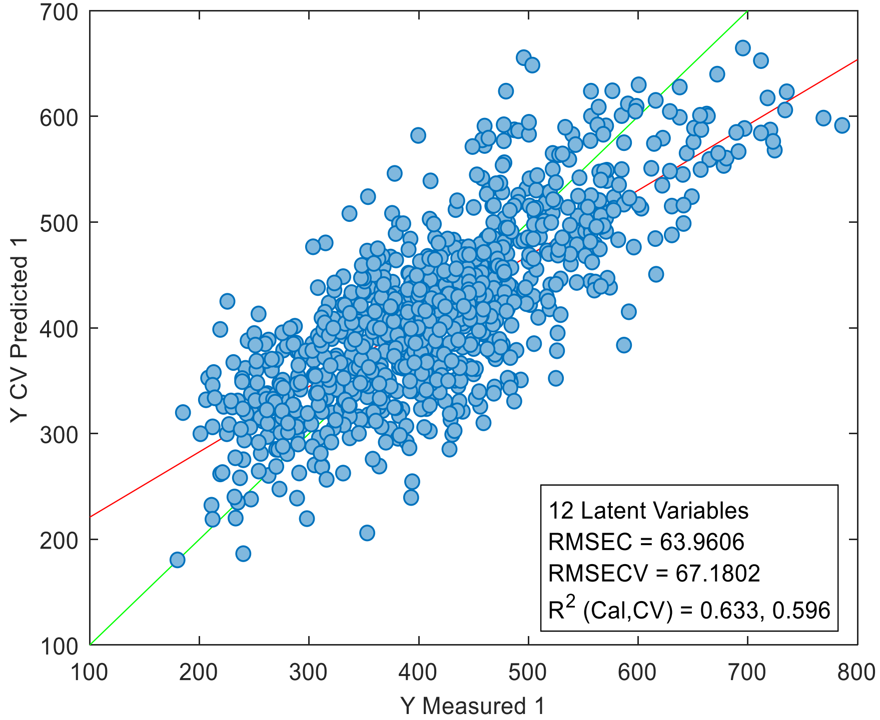
Effect of breeds in the model

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

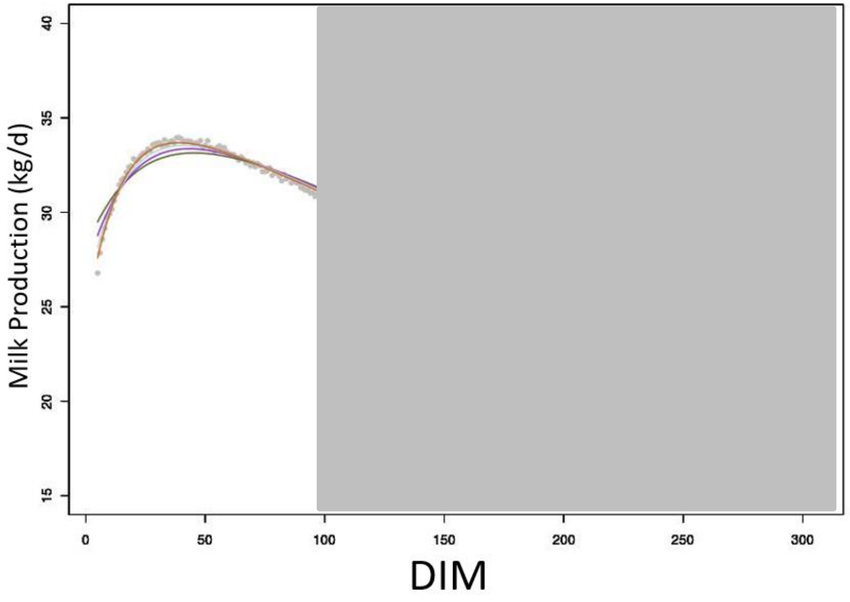
-19%

Effect of DIM in the model

Dataset used: <u>CH_4</u> by dairy cows

- 350 records from DIM 0 to DIM 100

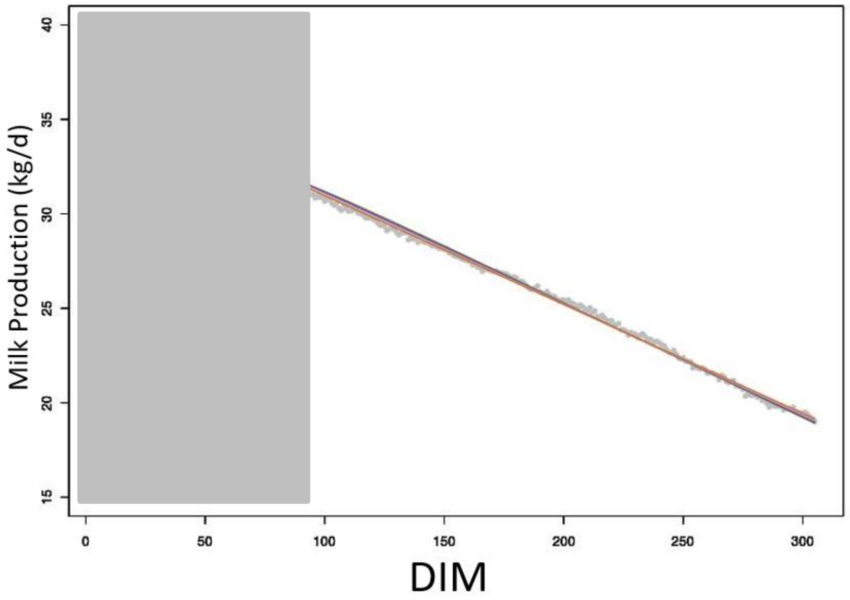
Step 1 : calibration with beginning of lactation



RMSEcv = 58 g/d

External validation with late lactation

- 689 samples from DIM 100 to 320



RMSEP = 90 g/d

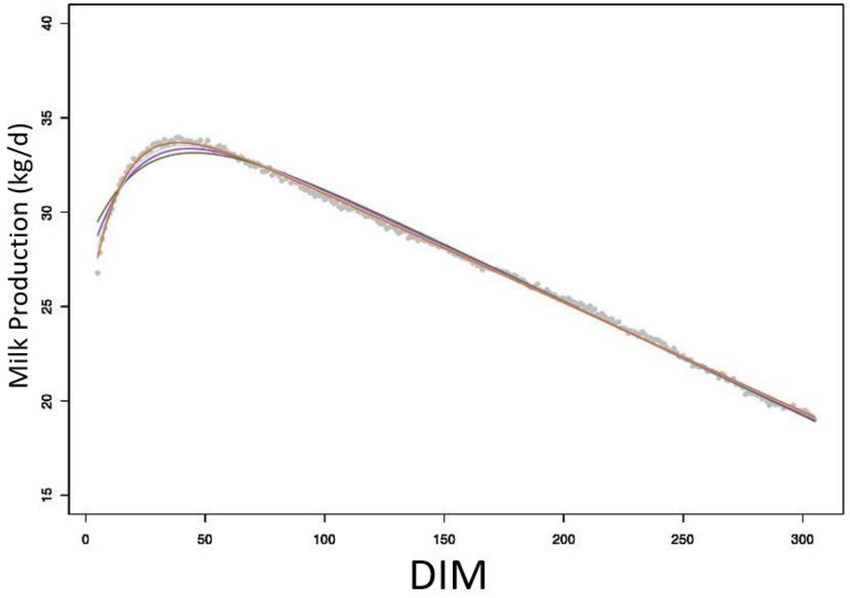
Effect of DIM in the model

Dataset used: <u>CH_4</u> by dairy cows

- 350 records from DIM 0 to DIM 100

Step 2 : calibration with beginning of lactation

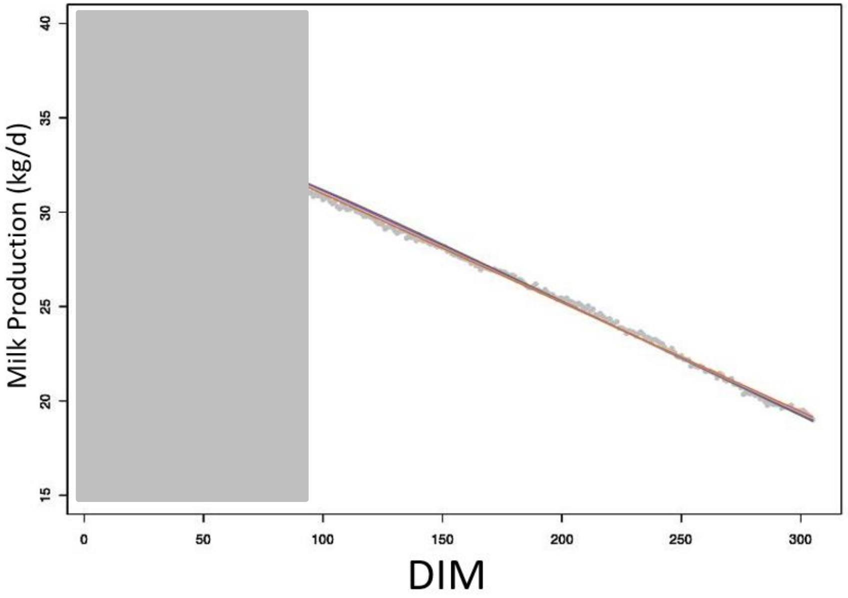
+ 50 records randomly selected between DIM 100 to 320



RMSEcv = 62 g/d

External validation with late lactation

- 689 samples from DIM 100 to 320



RMSEP = 78 g/d

-13%

“IR models can only predict what they know”

Pierre Dardenne

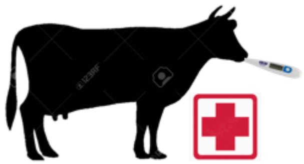
“Extrapolation is dangerous!”

IR maxim

Sampling method

Effect of sampling method

Dataset used: <u>Lactoferrin</u>



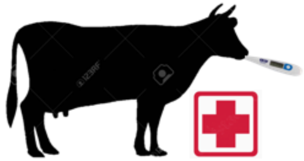
- 3506 as a global calibration population

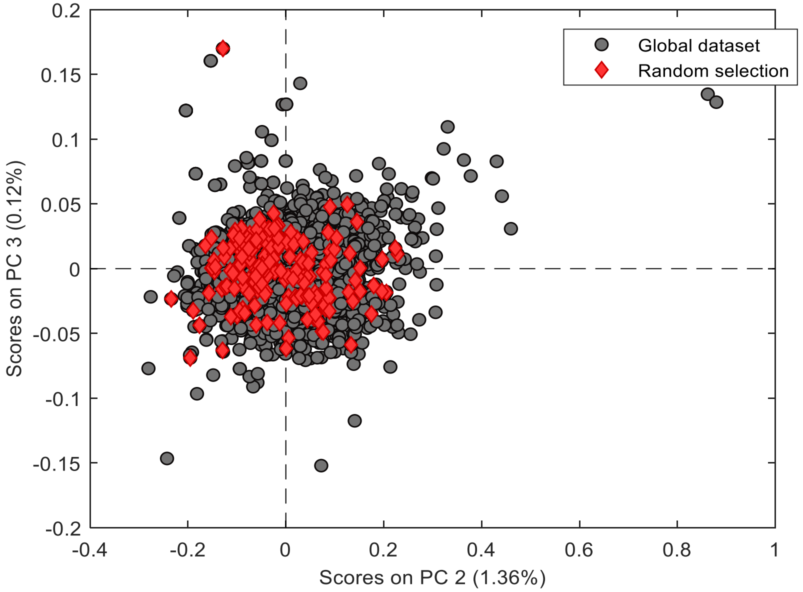
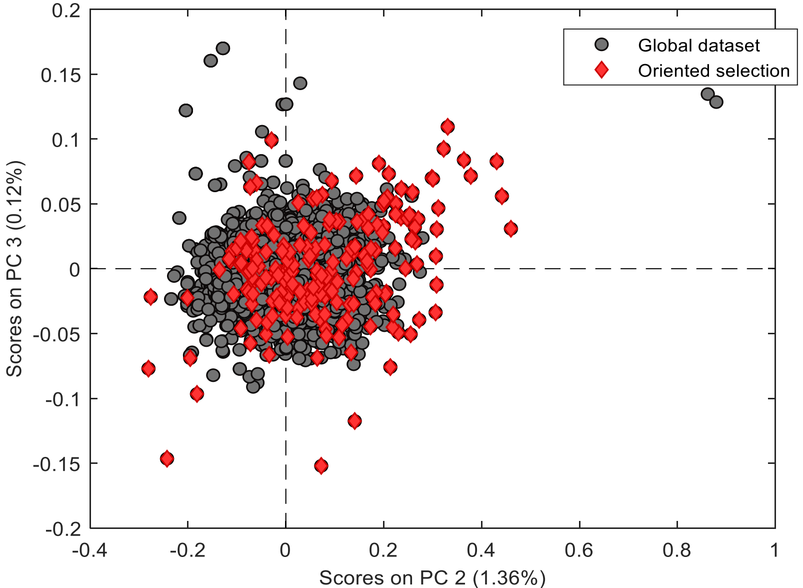
  →Selection of 200 samples to develop a model

  → Random selection

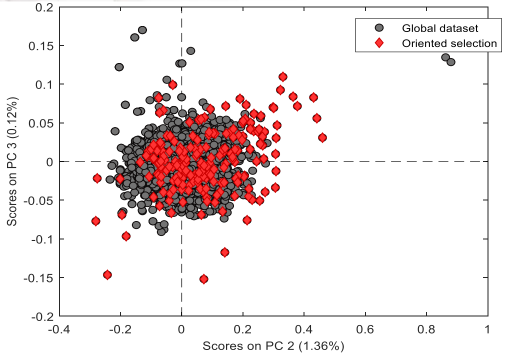
  → Oriented selection, to cover the spectral variability

External validation with 400 samples

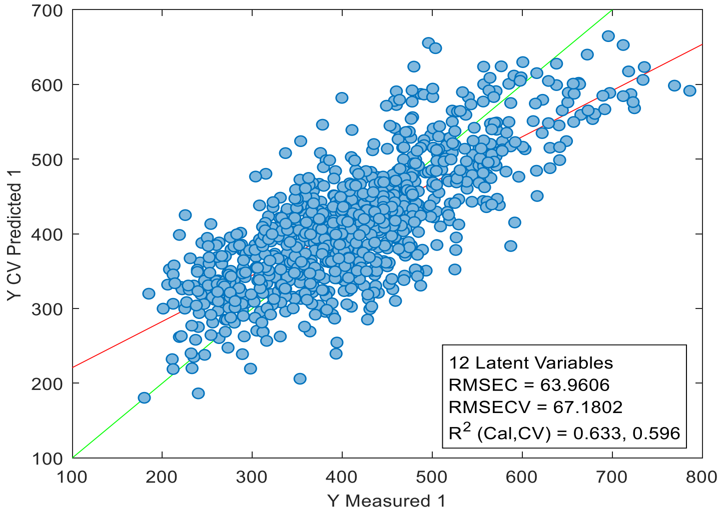


	Selection	Cross-validation	External-validation (400 external samples)
Random selection		RMSEcv = 126 g/L	• **RMSEP = 170 g/L** • 94.4% samples with GH<3
Oriented selection		RMSEcv = 176 g/L	• **RMSEP = 146 g/L** • 98.4% samples with GH<3 **-14%**

IR models can only predict what they know



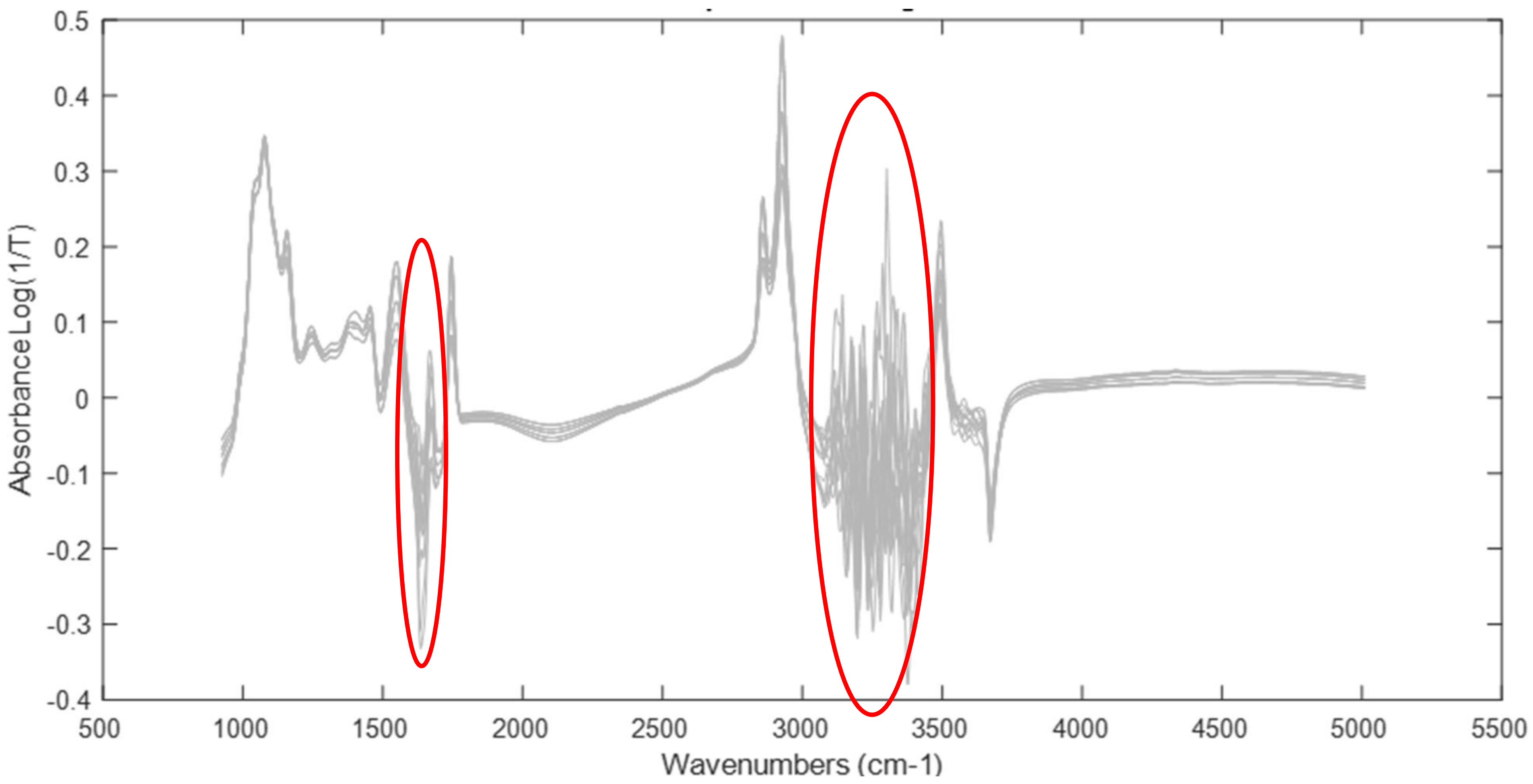
Robust models



Model development

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

ELSEVIER

Journal of Dairy Science
Volume 101, Issue 3, March 2018, Pages 2260-2272

Research

Genome-wide association study for milk infrared wavenumbers

Qiuyu Wang, Henk Bovenhuis

Journal of Dairy Science

Genetic and environmental variation in bovine milk infrared spectra

Qiuyu Wang, Alex Hulzebosch, Henk Bovenhuis

ELSEVIER

Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds

G. Bittante, A. Cecchinato

Journal of Dairy Science

Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows

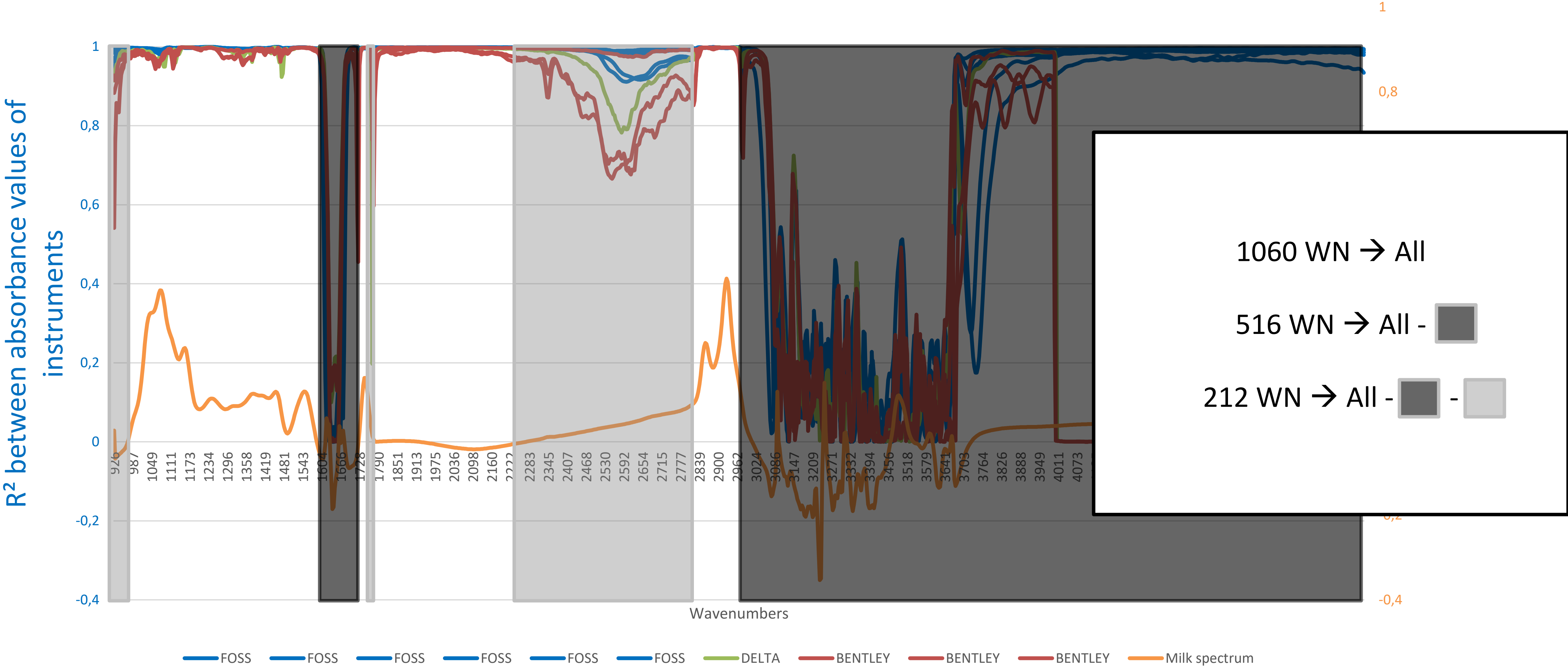
Hugo Toledo-Alvarado *, Ana I. Vazquez †, Gustavo de los Campos †, Robert J. Tempelman ‡, Giovanni Bittante *, Alessio Cecchinato *

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 7 Foss instruments + 3 Bentley + 1 Delta

For each wavenumber, correlation between the absorbance values of a reference and the others instruments

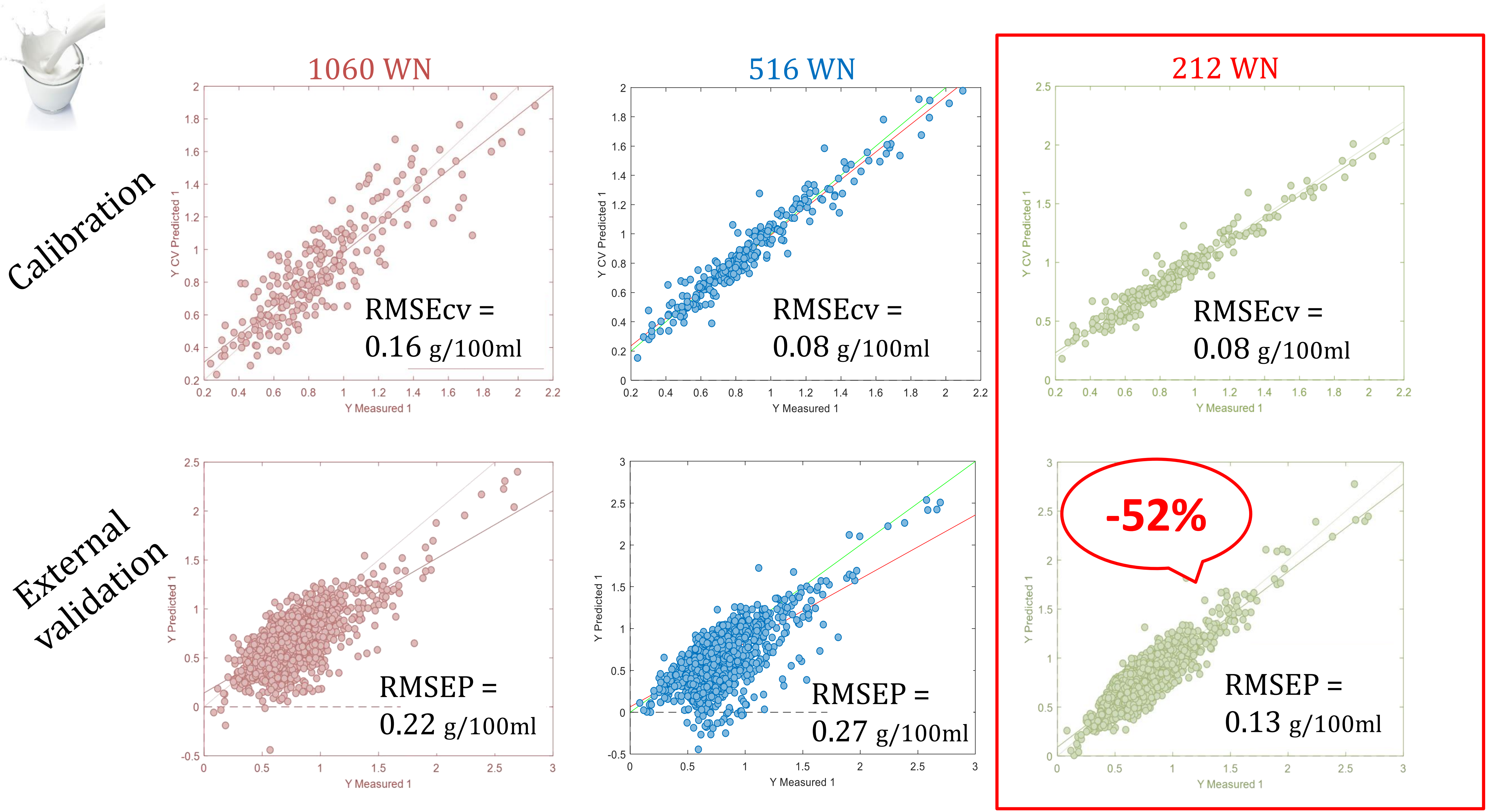


Effect of model development: Wavenumber selection

Dataset used: C18_1 cis9 fatty acid

- 250 samples in calibration
- 1572 samples in external validation

Same number of PLS factors



1060 WN
516 WN
212 WN
Calibration
External validation
RMSEcv = 0.16 g/100ml
RMSEcv = 0.08 g/100ml
RMSEcv = 0.08 g/100ml
RMSEP = 0.22 g/100ml
RMSEP = 0.27 g/100ml
RMSEP = 0.13 g/100ml
-52%
Y CV Predicted 1
Y Predicted 1
Y Measured 1

Effect of model development: PLS factor selection

Dataset used: <u>Weight of cows</u>

1033 records from 241 cows

- 75% cows in calibration (781 records)
- 25% cows in validation (252 records)

Different ways to choose the PLS factor number

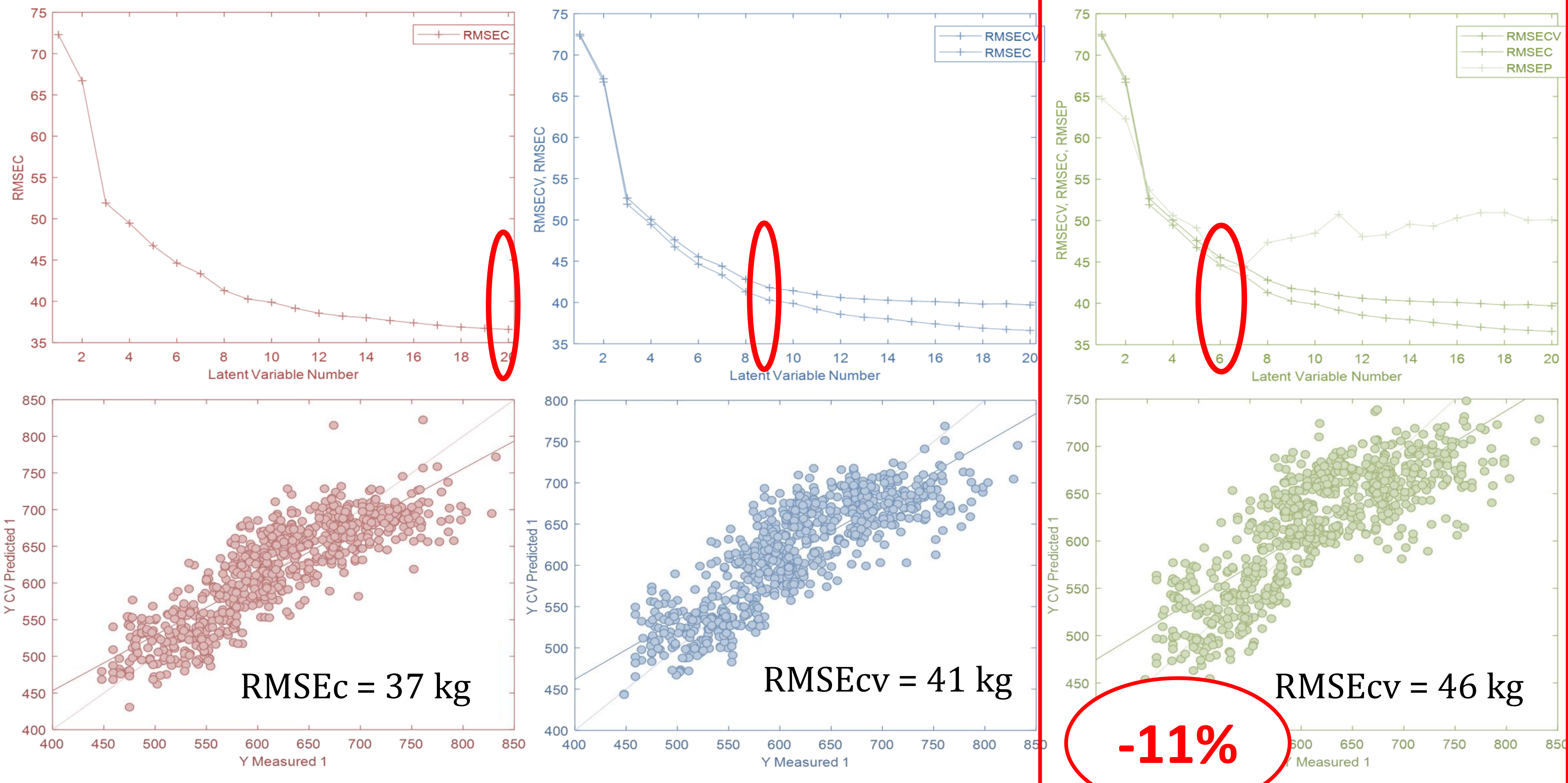
Calibration step: 20 LV

Cross-validation step: 9 LV

External validation : 6 LV

PLS factors selection

CV performances



Validation performances

RMSEP = 50 kg

RMSEP = 48 kg

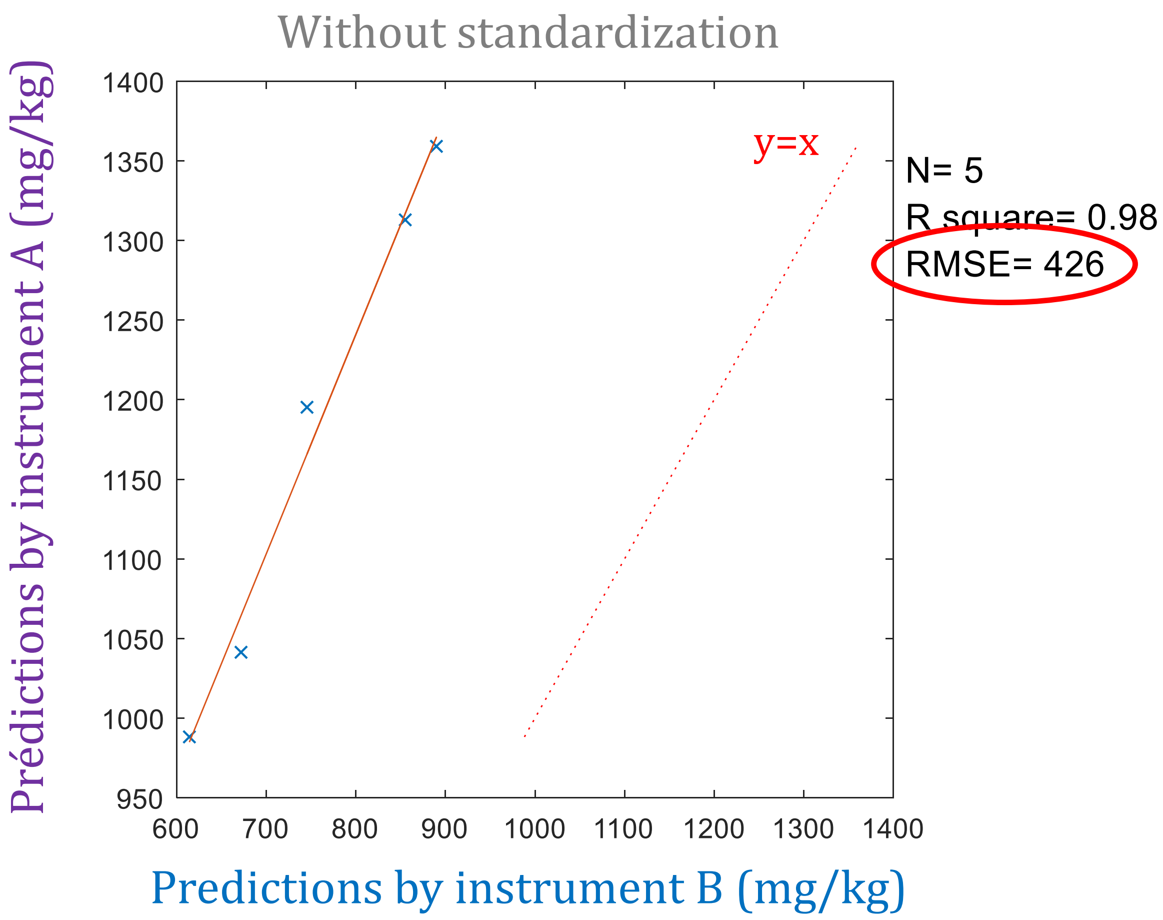
RMSEP = 44 kg

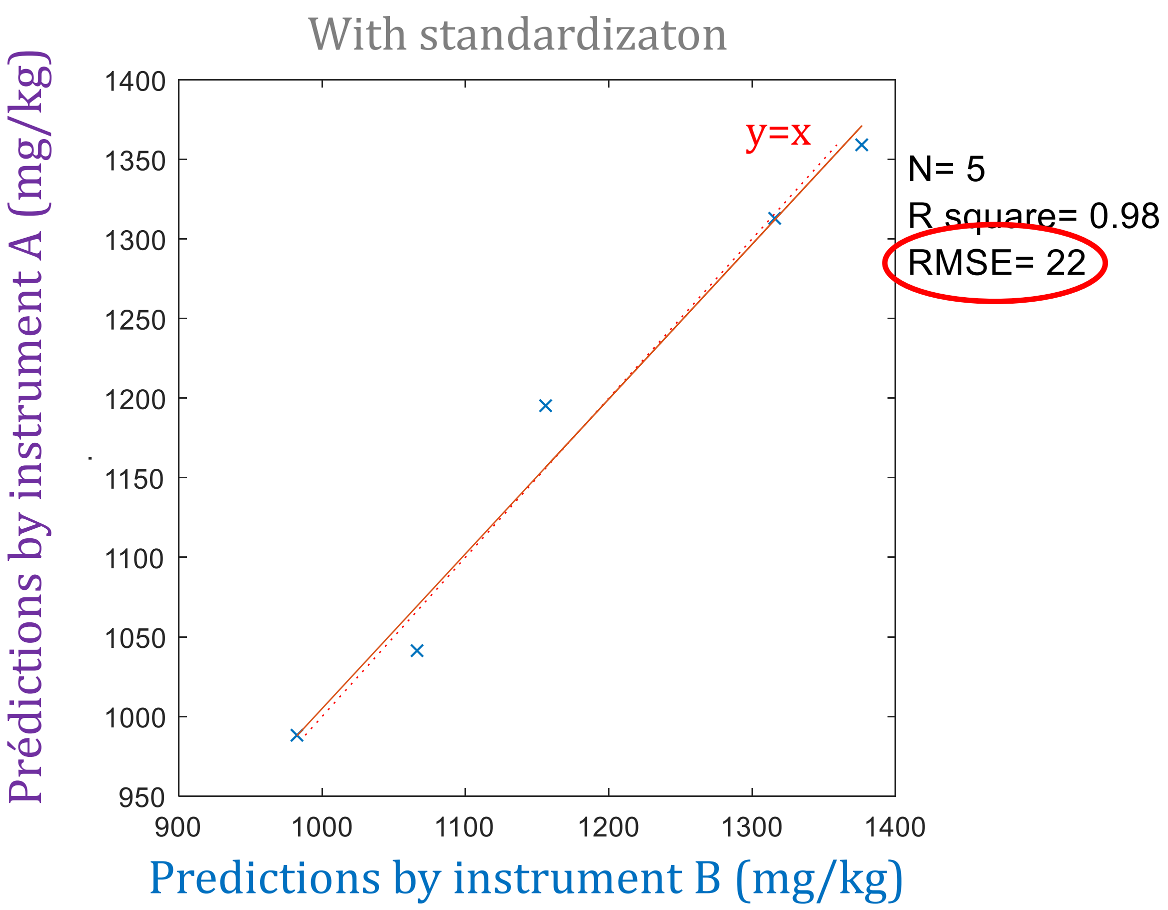
Spectral standardization

Spectral standardization

- <u>Calcium</u> model developed on instrument A

- Model applied on instrument B after analysis of common samples





Spectral standardization

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

Bias when using a model into another instrument!!

→ Standardization needed to use models at a large scale

ELSEVIER
Journal of Dairy Science
Volume 98, Issue 4, April 2015, Pages 2150-2160

Standardization of milk mid-infrared spectra from a European dairy network

C. Grelet [1], J.A. Fernández Pierna [1], P. Dardenne, V. Baeten, F. Dehareng

Research

Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models

C. Grelet [*], J. A. Fernández Pierna [*], P. Dardenne [*], H. Soyeurt [†], A. Vanlierde [*], F. Colinet [†], C. Bastin [‡], N. Gengler [†], V. Baeten [*], F. Dehareng [*]

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!!!**
- ✓ Keep models as simple as possible (Wavenumbers and PLS factors)
- ✓ Think spectroscopy (not only mathematics)
- ✓ Standardization

Thank you for your attention!

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*

Experience from 30 years NIR in feed…

Extensive databases :
thousands of feed NIR spectra from years.

Global models ?

- very robust to sample composition variation
  - but prediction accuracy decreased

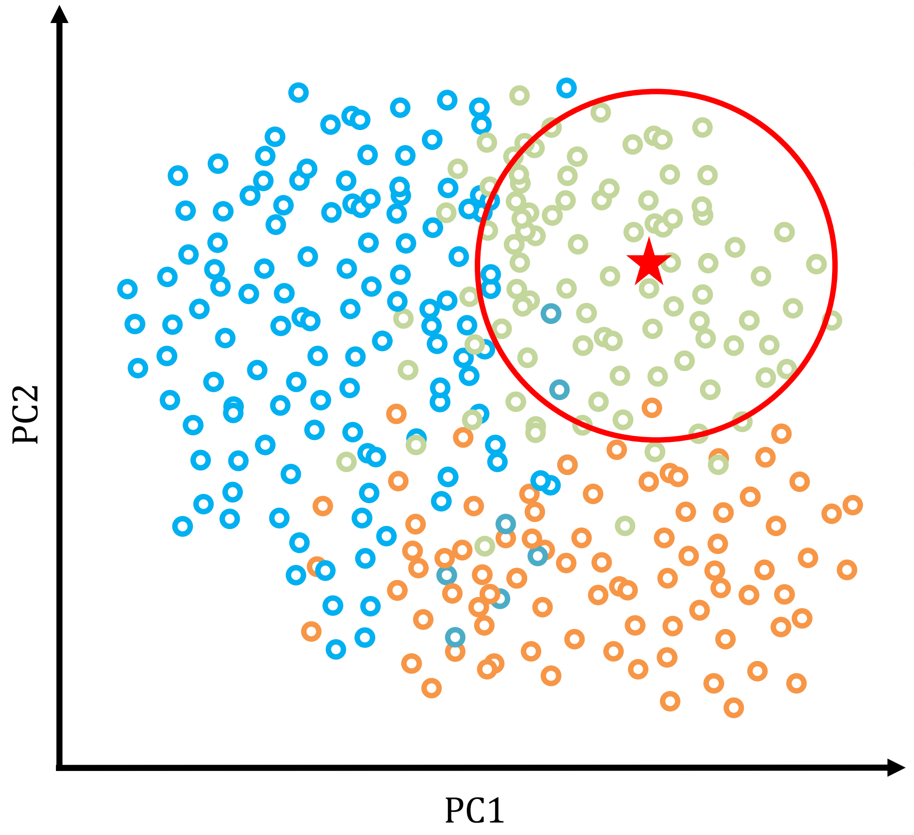
Specific models for small groups of similar samples ?

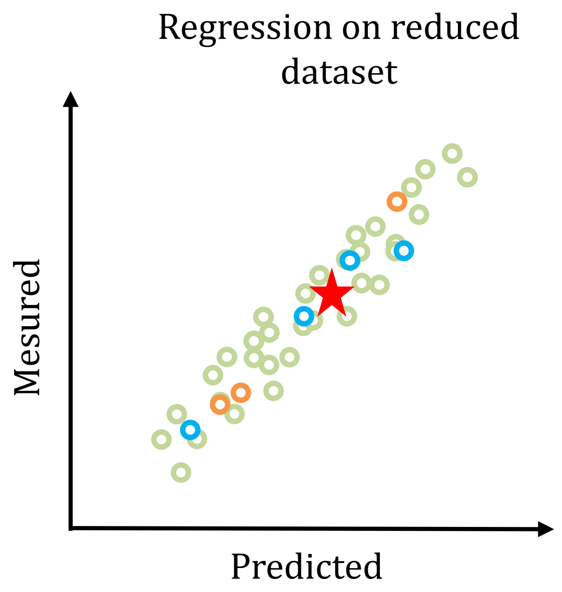
- difficult, time consuming, tedious in practice
  - increased complexity

Local regression

for each sample, compute specific model using a reduced calibration data extracted from a large library

Local methods





Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Regression models based on new local strategies for near infrared spectroscopic data

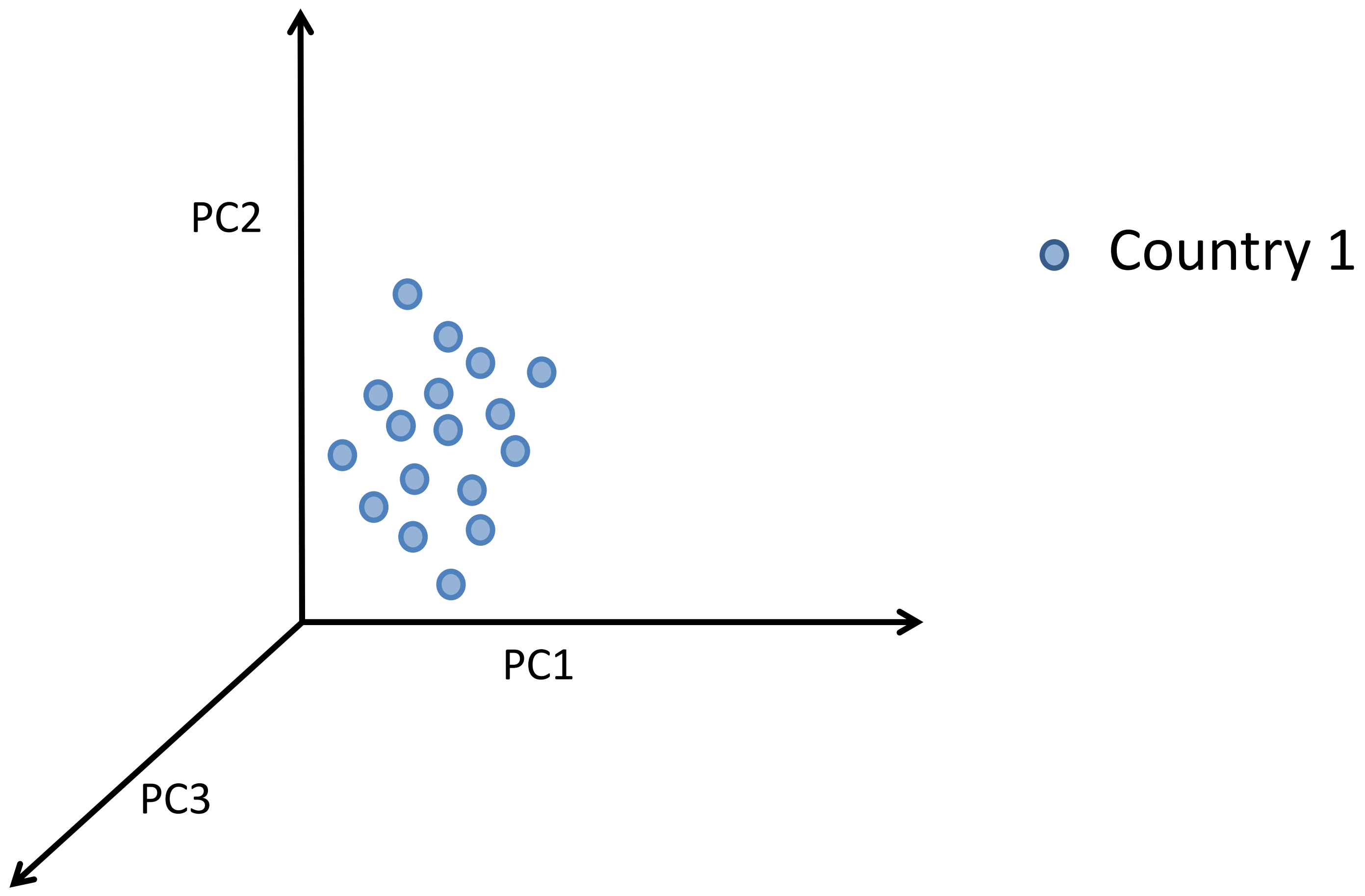
F. Allegrini [a,*], J.A. Fernández Pierna [b], W.D. Fragoso [c], A.C. Olivieri [a], V. Baeten [b], P. Dardenne [b]

[a] Univ. Nacional de Rosario, Facultad de Ciencias Bioquímicas y Farmacéuticas, IQUIR, CONICET, Argentina
[b] Valorisation of Agricultural Products Dpt, Walloon Agricultural Research Centre, Gembloux, Belgium
[c] Departamento de Química, Universidade Federal da Paraíba, Campus I, João Pessoa, Brazil

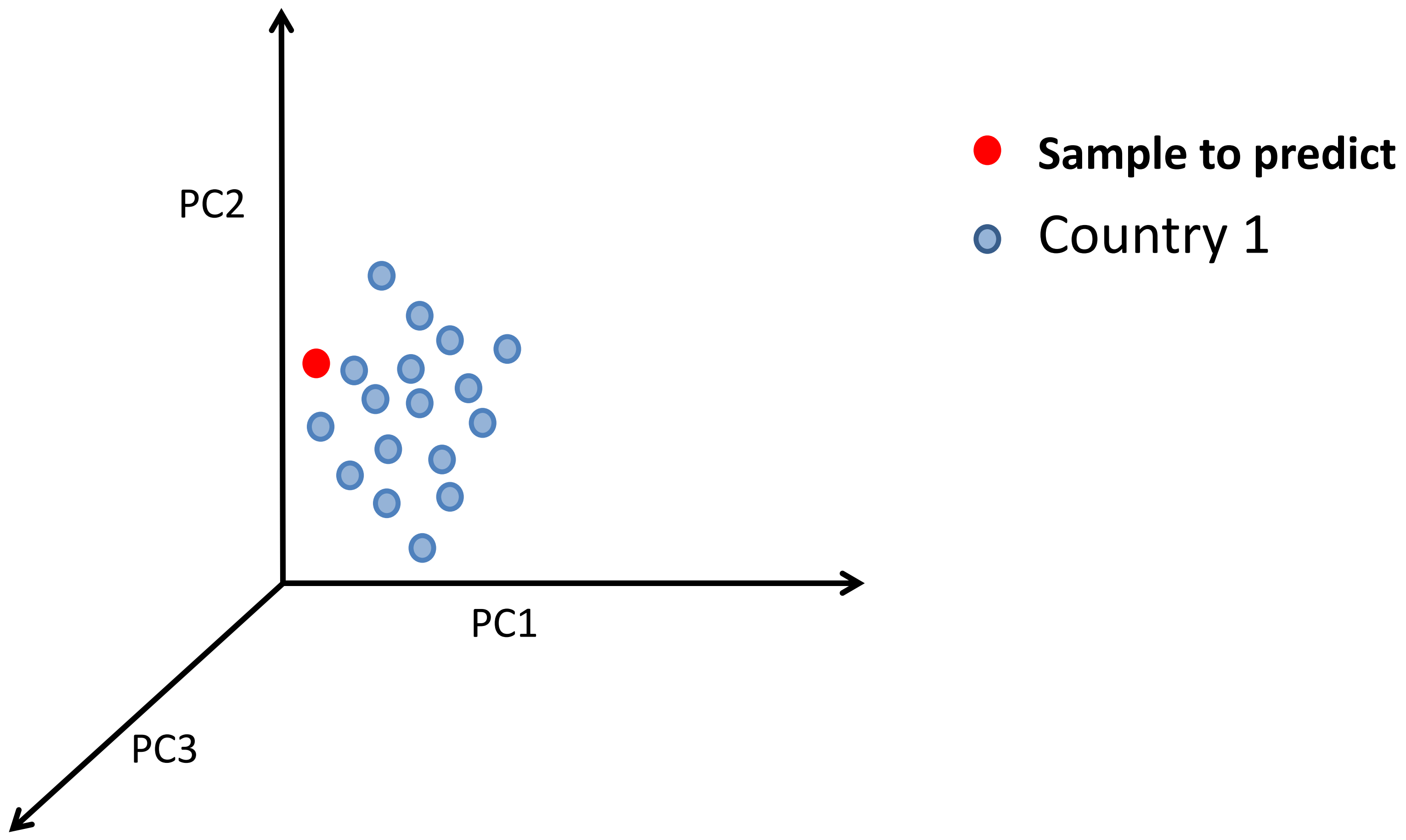
-12 to -15% RMSEP

Robustness & Specific

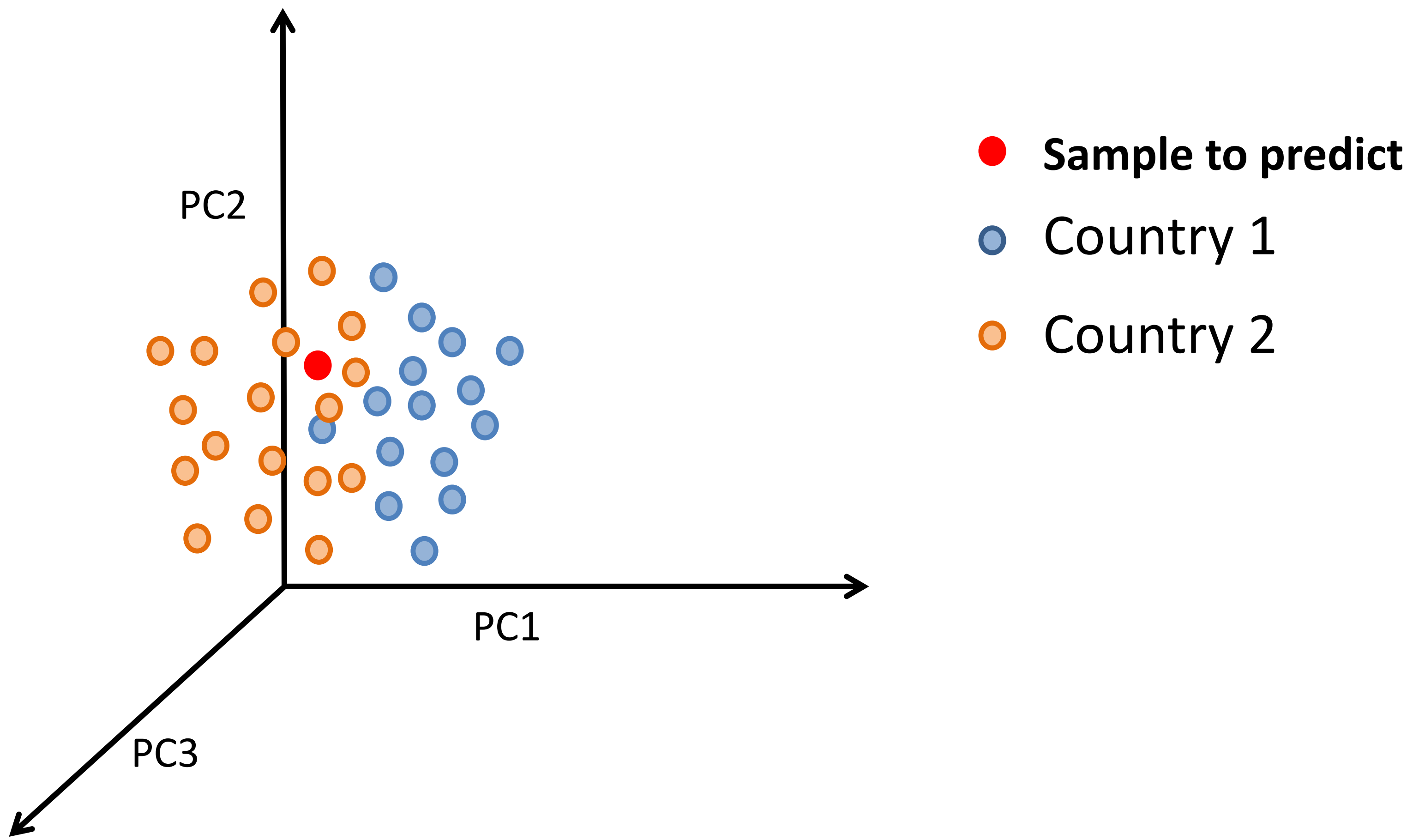
However...



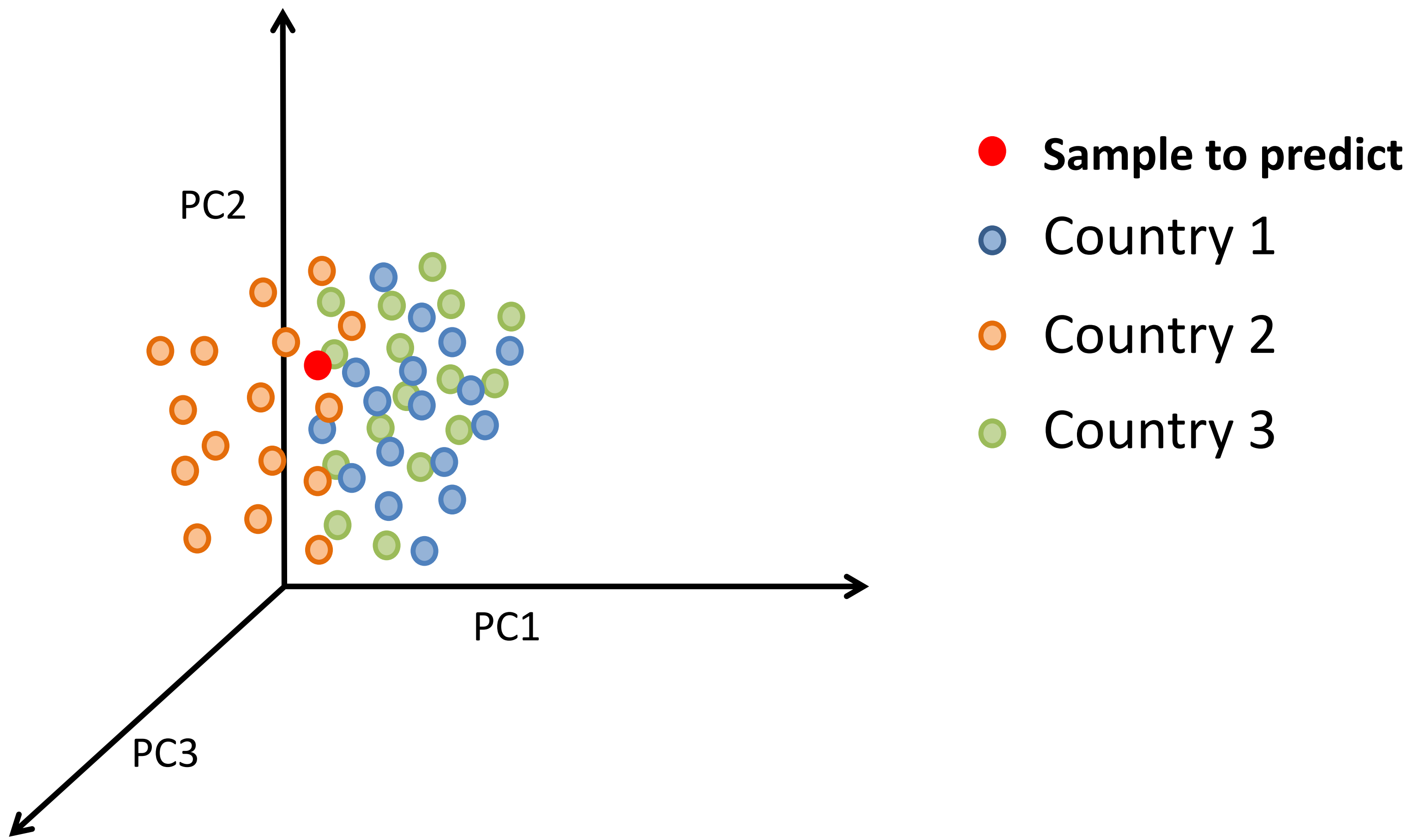
Cover the X (spectral data) range



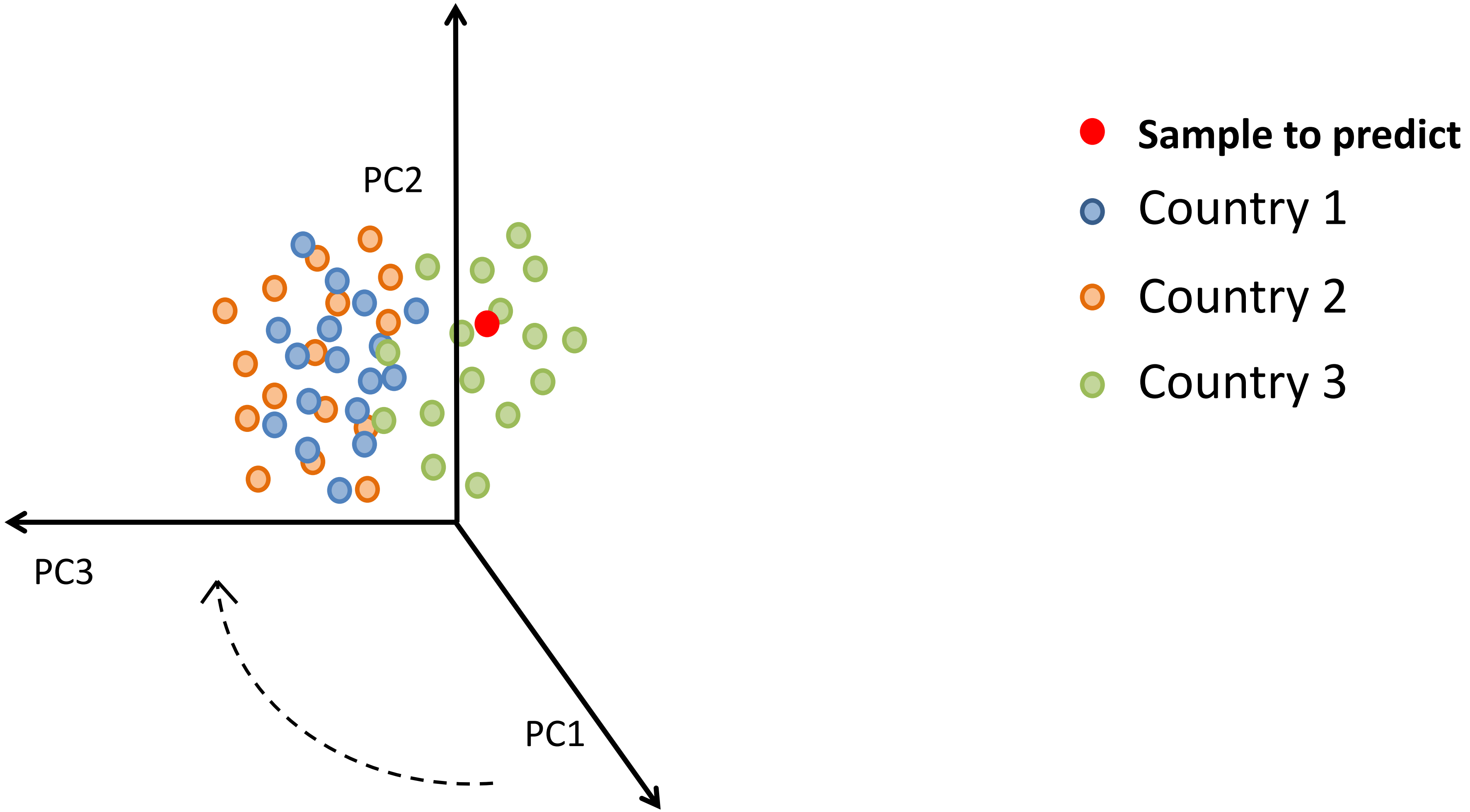
Cover the X (spectral data) range



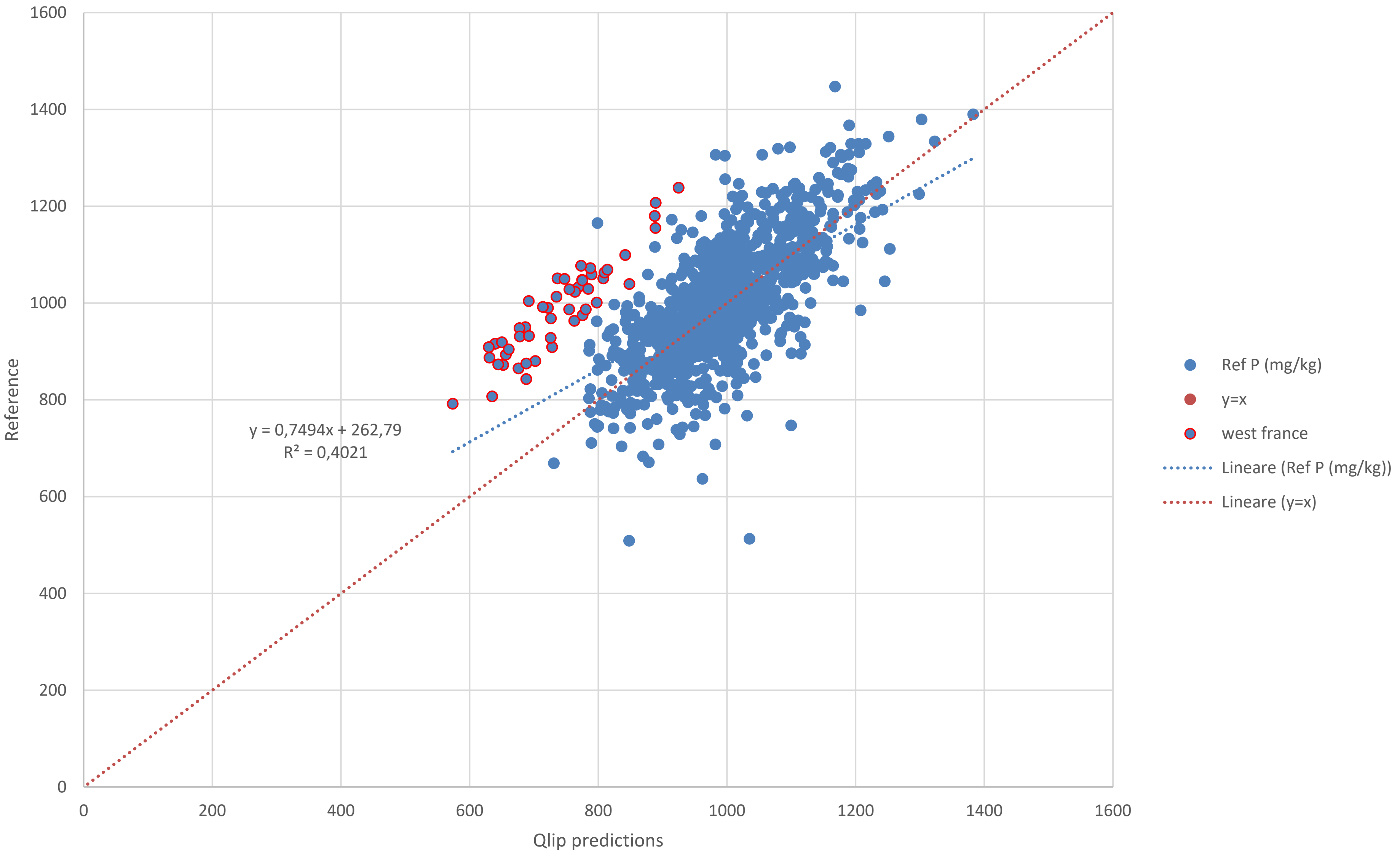
Cover the X (spectral data) range



Cover the X (spectral data) range



Real test in external validation

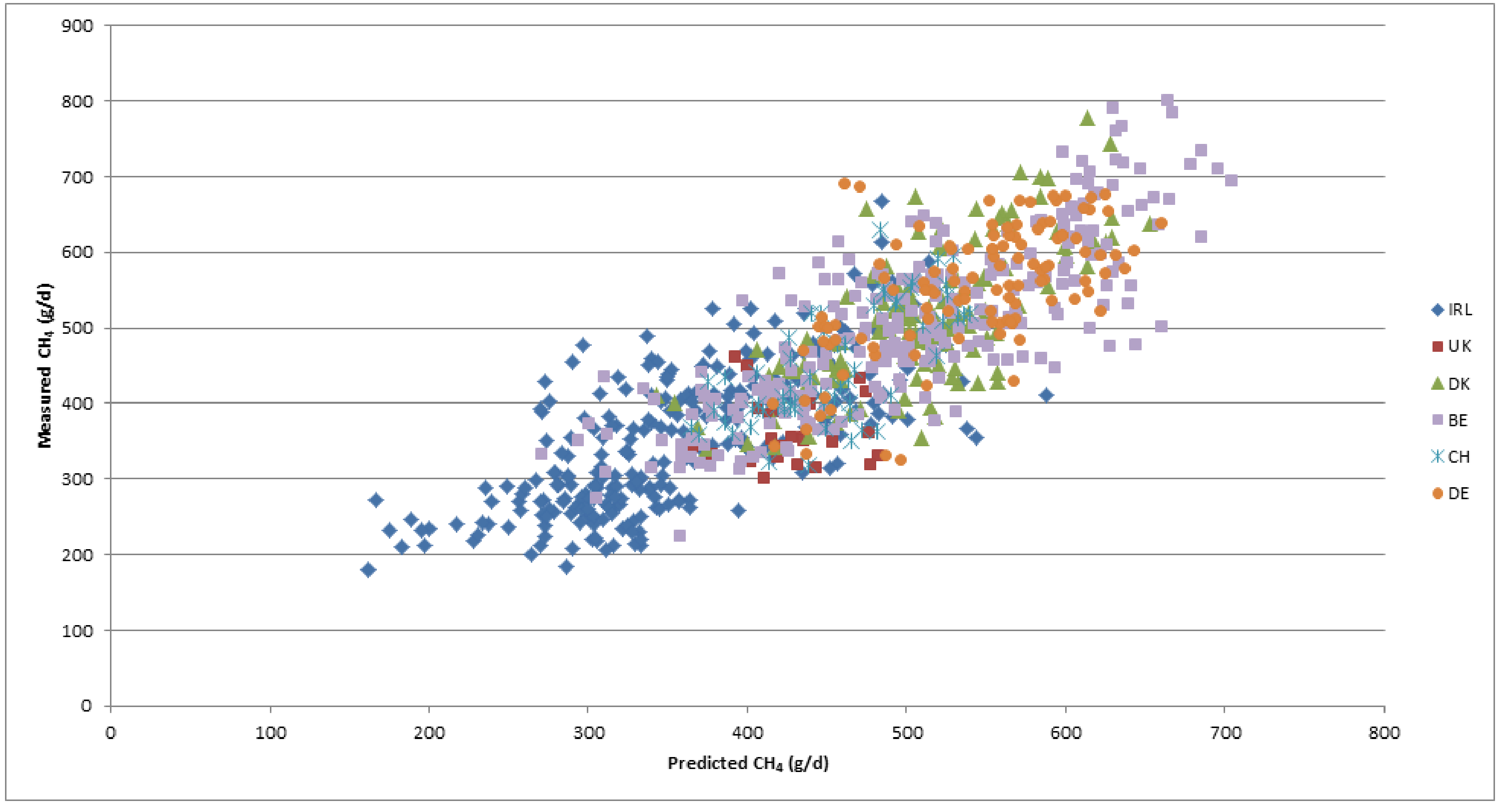


Methane model :

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



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

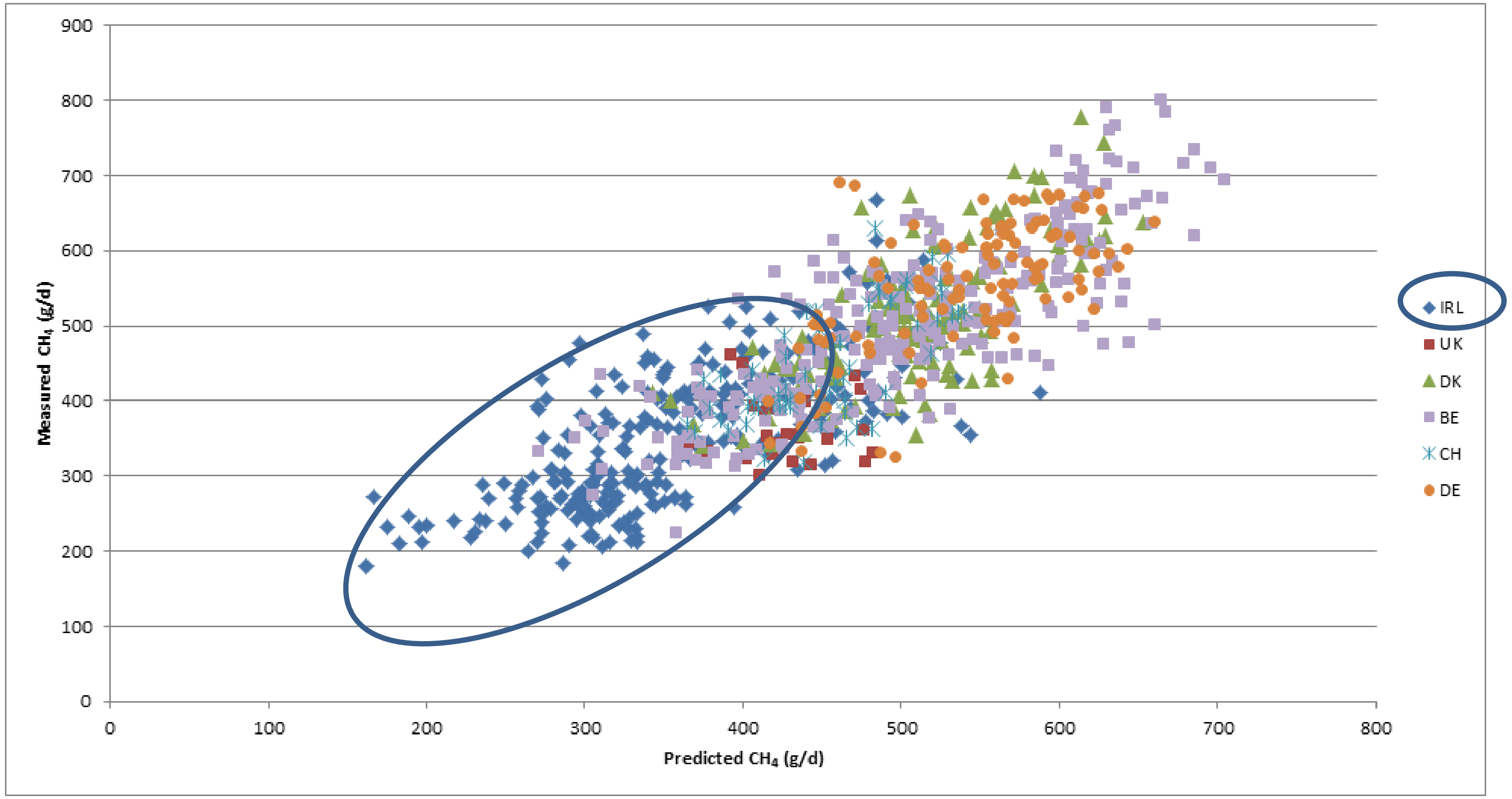
cra-w
Centre wallon de Recherches agronomiques
METHAGENE
QUALITAS
afbi Agri-Food and Biosciences Institute
FBN
AARHUS UNIVERSITY
Universität Zürich UZH
teagasc Agriculture and Food Development Authority
Wallonie

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



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

cra-w

Centre wallon de Recherches agronomiques

METHAGENE

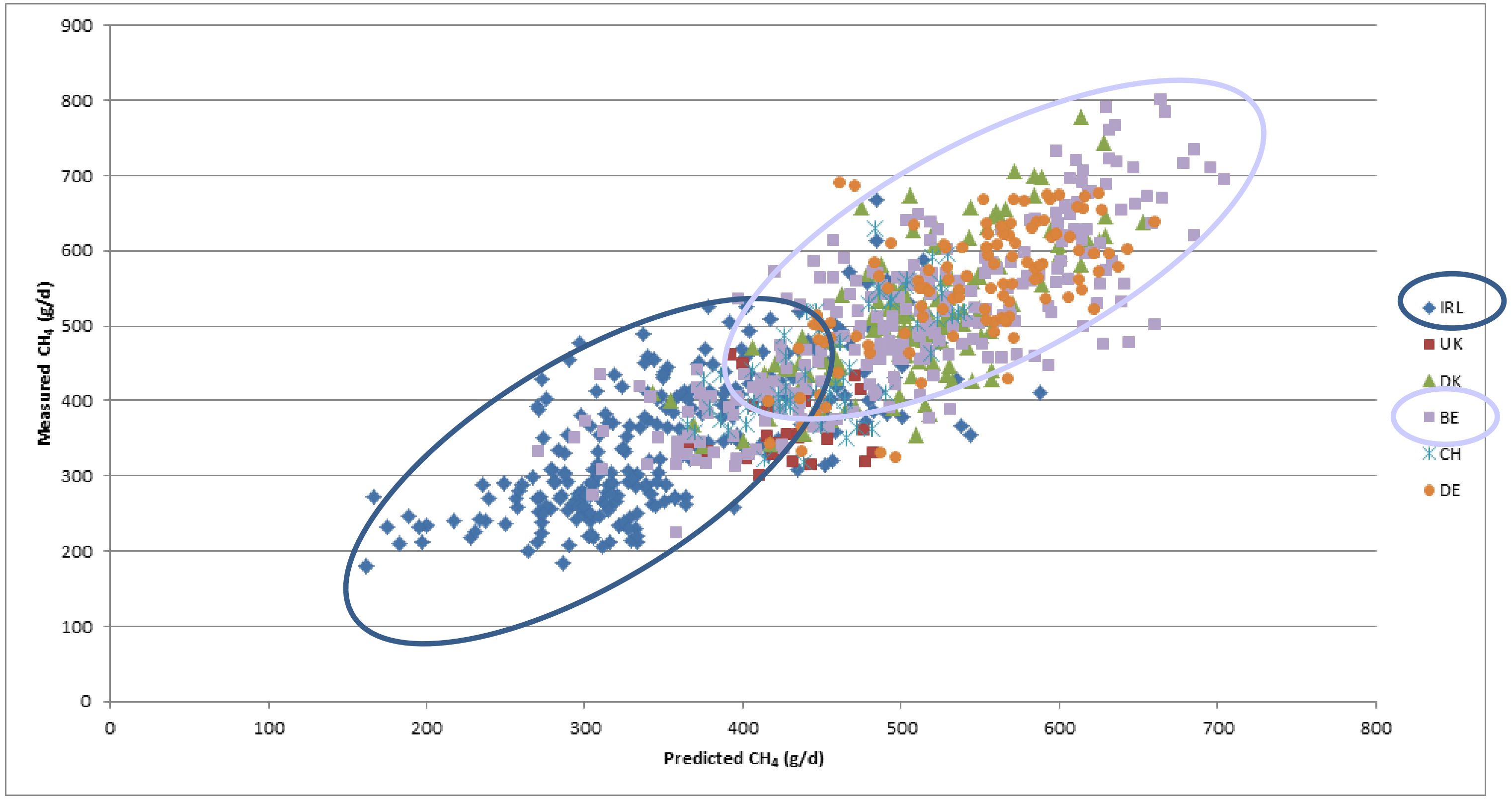
QUALITAS afbi Agri-Food and Biosciences Institute FBN AARHUS UNIVERSITY Universität Zürich teagasc Wallonie

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



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

Centre wallon de Recherches agronomiques

Effect of model development: Wavenumber selection

5 identical samples analyzed on 7 Foss instruments

For each wavenumber, correlation between the absorbance values of a reference and the others instruments