Potentials of milk Mid-infrared Spectrometry:

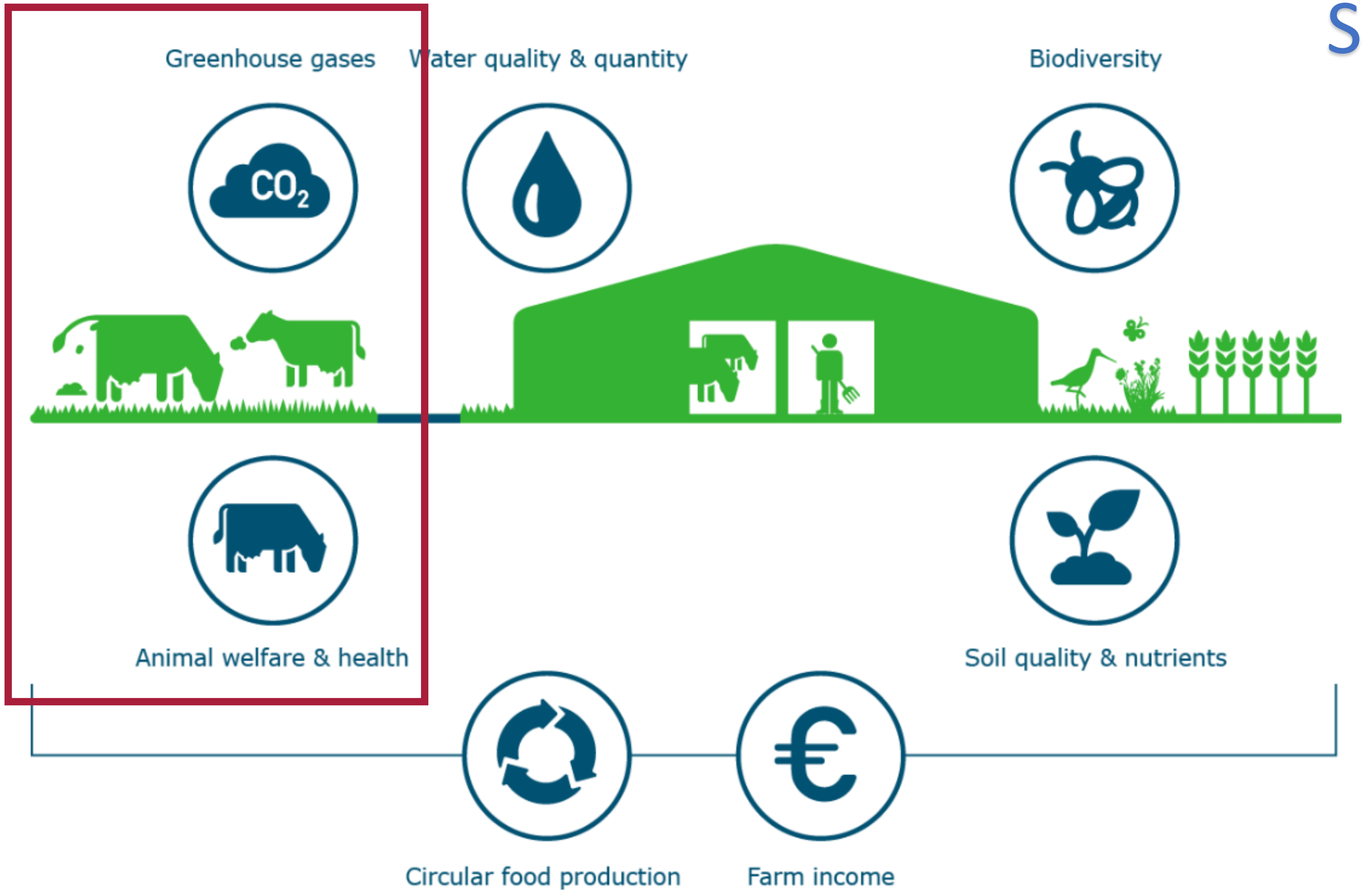
Where are we ? What are the next steps?

Prof. Hélène Soyeurt

hsoyeurt@uliege.be

Rinderzucht Austria – Wien – Austria – June 2025

Sustainable dairy production



Need data …

https://www.wur.nl/en/show/working-towards-sustainable-dairy-production.htm

vet

genetics

Need Phenotypes

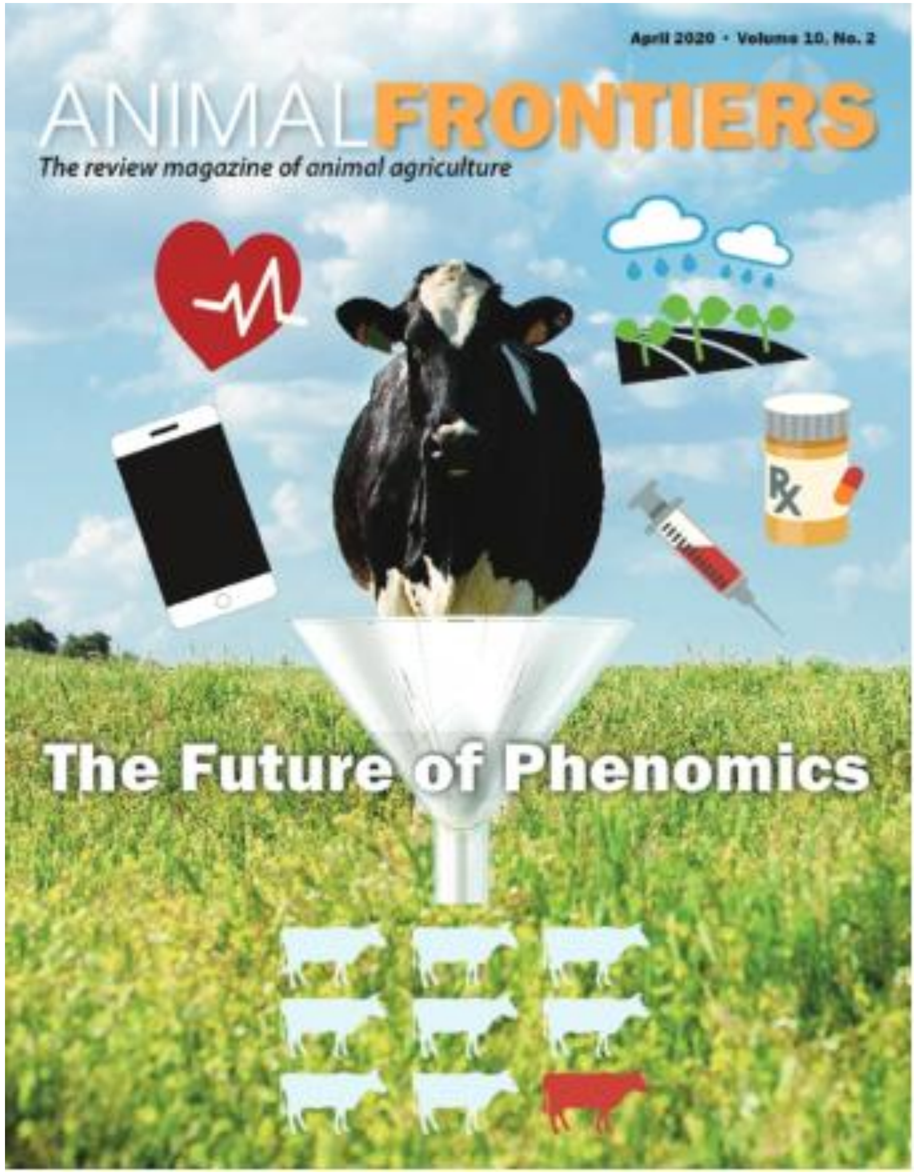
advicers

Decision tools

Everything is connected - Access to data is easy

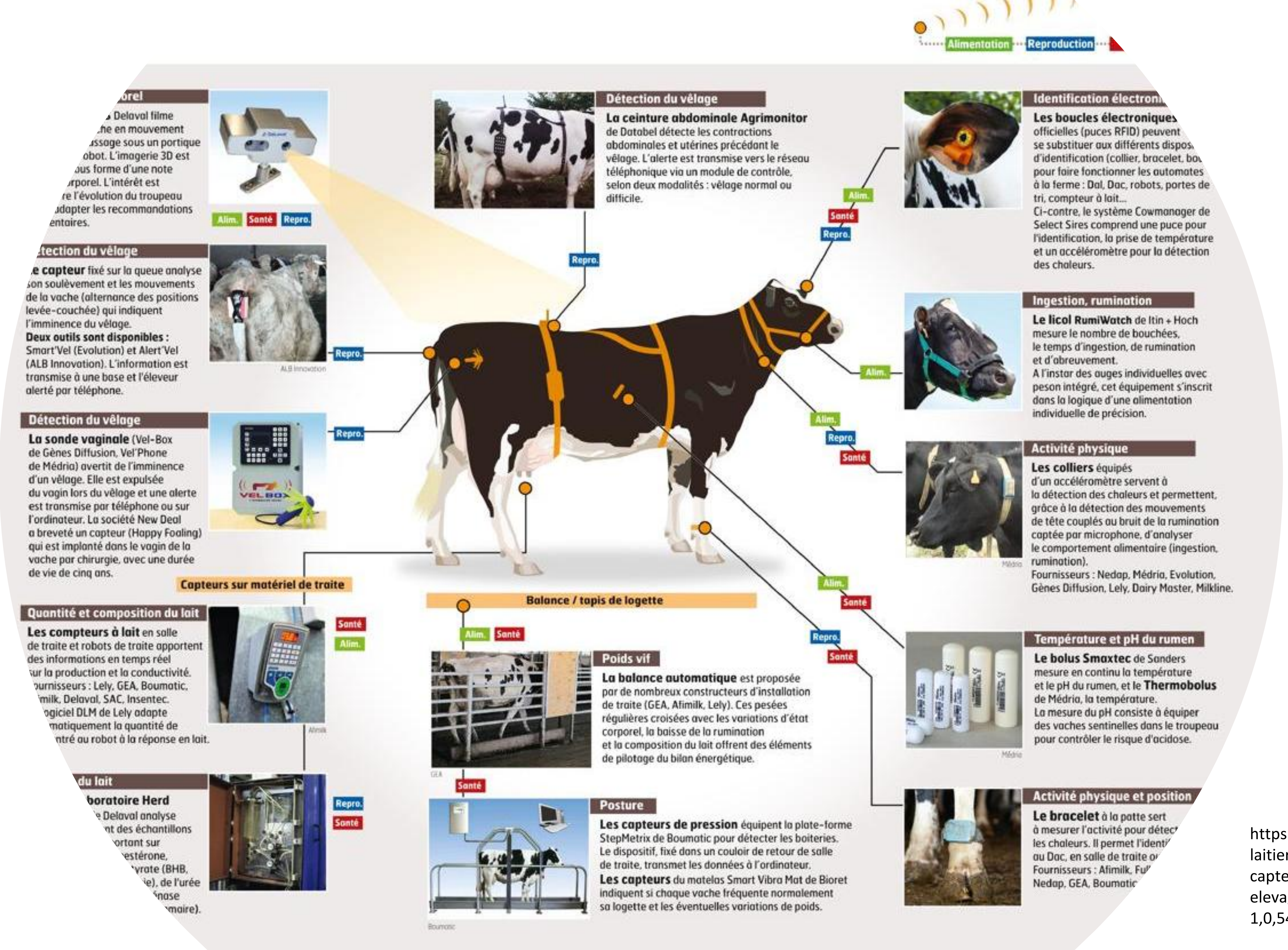
In the air …

On the ground …



April 2020 • Volume 10, No. 2
ANIMAL FRONTIERS
The review magazine of animal agriculture
RX
The Future of Phenomics

Directly on the cows ...



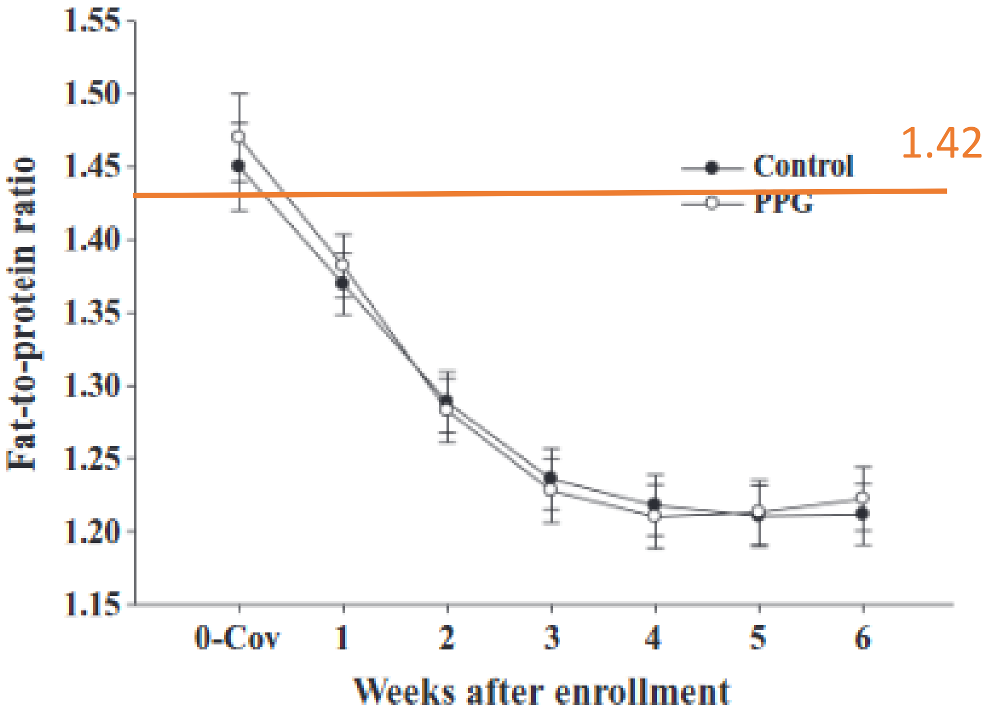
https://www.eleveur-laitier.fr/dossier/des-capteurs-au-service-de-l-elevage-de-precision-1,0,549358506.html

One of the most consumed food

An important source of nutritive elements for humans

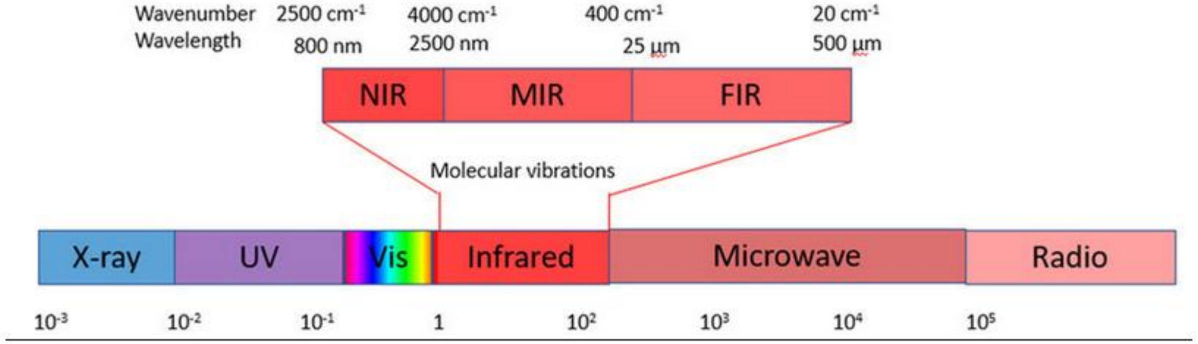
Its composition is the mirror of the animal and its changes reflect its health status

Knowing its composition is therefore of interest



Jenkins, 2015

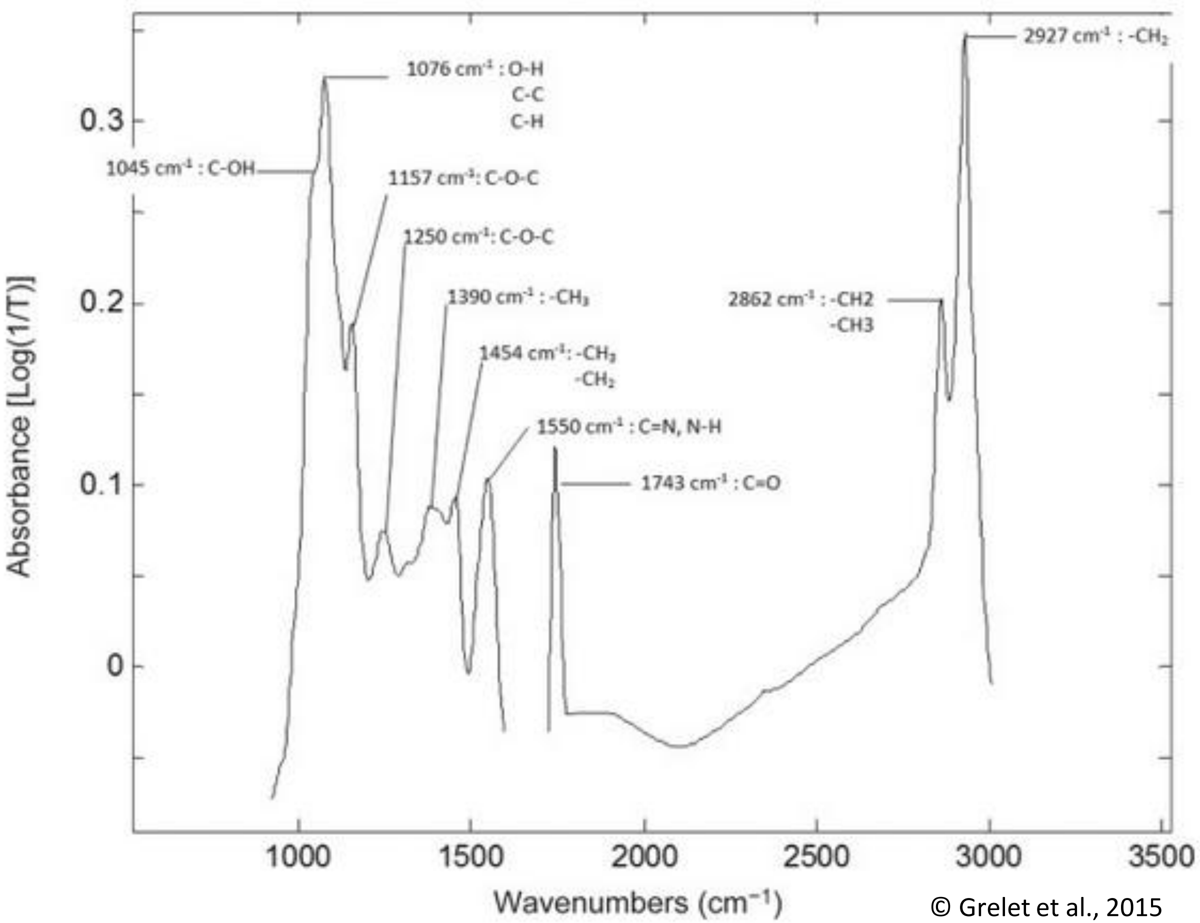
Milk Analysis



Milk MIR spectrum

=

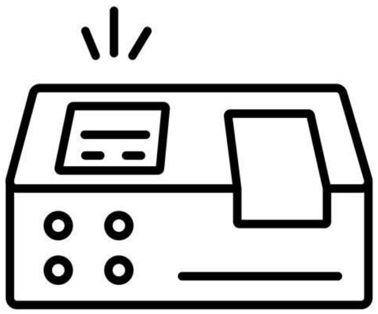
Absorption of infrared ray at frequencies related to the vibrations of specific chemical bounds in milk



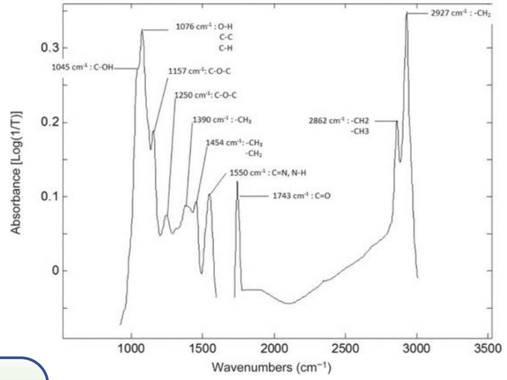
Milk Analysis

Milk payment : 1-3 days (Bulk tank milk)
Milk recording : 4-6 weeks (individual cow milk)

Milk samples



Milk FT-MIR



© Grelet et al., 2015

Usually,
Fat and Protein contents

Predictive model = Equation

But we used the same spectral information, could we go further ?

What exists ?

Nutritional quality

Fat, protein, lactose, fatty acids, Ca, lactoferrin,…

Sustainability

Consumption index, nitrogen efficiency …

Technological properties

Cheese yield, yoghurt yield, butter yield, spreadability …

Environmental fingerprint

Methane, P, urea …

Animal Health

Na, lactoferrin, Energy balance, body weight, dry matter intake, acetone, BHB, citrate …

Abnormal milk samples

Methods
Volume 186, February 2021, Pages 97-111

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. Stevens [a], N. Gengler [b], F. Dehareng [a]

Cluster	RPDcv	Relative RMSEcv	R^2cv	Interpretation for application
1	> 6	<5%	> 0.97	Any application
2	4.2–6	<10%	0.94–0.97	Quality control
3	3–4.2	<10%	0.89–0.94	Quantitative screening
4	2–3	<25%	0.74–0.89	Rough screening
5	1.5–2	<25%	0.55–0.74	Allows to compare groups, discriminate high or low values
6	1.5–2	>25%	0.55–0.74	Highly imprecise, can be used to detect extreme values
7	< 1.5	–	< 0.55	Not recommended

From our experience …

Table 1
Datasets used in the study.

Traits	N records	N cows	N countries	Sampling years	References
Milk Fatty acids, 30 models	1822	1822	7	2005–2015	Soyeurt et al. [39,40]
Milk Minerals, 5 models	1340	1340	4	2005–2015	Soyeurt et al. [41]
Milk Lactoferrin	3906	3906	3	2005–2009	Soyeurt et al. [42]
Methane emitted (CH4)	1089	299	7	2010–2019	Dehareng et al. [43] and Vanlierde et al. [18]
Milk Fresh Cheese Yield (FCY), Coagulation time r, Time when the curd is firm enough for cutting (k20)	283	283	1	2011–2014	Colinet et al. [44]
Milk Casein	996	*	1	2011–2014	Not published
Milk Acetone, β-hydroxybutyrate (**BHB**) and Citrate	566	346	3	2013–2014	Grelet et al. [45]
Blood BHB, Non-Esterified Fatty Acids (NEFA), Insulin Growth Factor I (IGF-I), Glucose	387	241	6	2014–2016	Grelet et al. [6]
Nitrogen efficiency (NUE), Nitrogen losses, Dry matter intake (DMI), Body weight	1034	129	3	2014–2015	Grelet et al. [14]
Milk Glucose free, Glucose-6-phosphate, Uric acid, Iso-citrate, Progesterone	2175	241	6	2014–2016	Not published

*The casein model was constituted by 790 samples from individual cows and 206 samples from bulk tank.

	Phenotype	Min	Max	Mean	SD	R^2cv	Relative RMSEcv	RPDcv	Cluster
Any application	Milk SAT FA(g/dL)	0.31	6.97	2.70	0.75	0.99	3%	10.22	1
Quality control	Milk C18_1cis9 (g/dL)	0.08	2.69	0.76	0.29	0.95	8%	4.35	2
	Milk Casein (g/100 g)	1.61	4.05	2.66	0.34	0.95	3%	4.46	
	Milk LCFA (g/dL)	0.19	4.79	1.59	0.52	0.95	7%	4.52	
	Milk MCFA (g/dL)	0.22	5.48	2.00	0.60	0.97	5%	5.53	
	Milk MONO FA (g/dL)	0.12	3.42	1.08	0.35	0.97	5%	5.83	
	Milk Tot18_1cis (g/dL)	0.09	2.77	0.82	0.31	0.95	8%	4.58	
	Milk Total_C18_1 (g/dL)	0.10	2.98	0.94	0.33	0.96	7%	5.18	
	Milk UNSAT (g/dL)	0.14	3.86	1.25	0.39	0.97	5%	5.75	
Quantitative screening	Milk C10 (g/dL)	0.02	0.32	0.11	0.04	0.91	9%	3.37	3
	Milk C12 (g/dL)	0.02	0.41	0.13	0.04	0.92	9%	3.62	
	Milk C14 (g/dL)	0.05	1.20	0.45	0.13	0.93	7%	3.88	
	Milk C16 (g/dL)	0.12	3.32	1.20	0.40	0.94	8%	4.18	
	Milk C4 (g/dL)	0.01	0.23	0.10	0.03	0.93	8%	3.67	
	Milk C6 (g/dL)	0.01	0.16	0.07	0.02	0.91	9%	3.32	
	Milk C8 (g/dL)	0.01	0.11	0.05	0.01	0.91	9%	3.29	
	Milk Citrates (mmol/L)	3.88	16.12	9.04	2.21	0.89	8%	3.04	
	Milk SCFA (g/dL)	0.05	0.80	0.35	0.10	0.93	7%	3.88	
Rough screening	Milk C17 (g/dL)	0.00	0.09	0.03	0.01	0.80	13%	2.24	4
	Milk C18 (g/dL)	0.05	1.32	0.40	0.15	0.84	14%	2.51	
	Milk Calcium (mg/kg)	593	1743	1149	135	0.82	5%	2.34	
	Milk Odd fatty acids (g/dL)	0.03	0.50	0.16	0.04	0.83	10%	2.41	
	Milk PUFA (g/dL)	0.02	0.53	0.16	0.05	0.77	13%	2.10	
	Milk Total_Trans (g/dL)	0.02	0.75	0.16	0.08	0.80	19%	2.26	
	Tot18_1trans (g/dL)	0.01	0.57	0.13	0.06	0.79	21%	2.17	

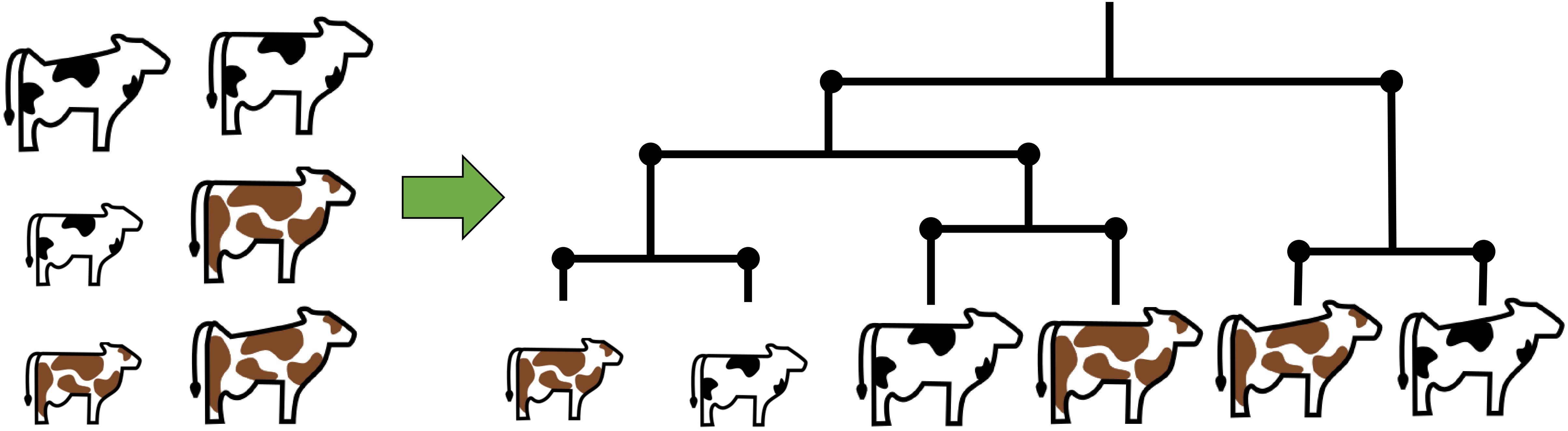
Group	Phenotype	Min	Max	Mean	SD	R^2cv	Relative RMSEcv	RPDcv	Cluster
High and Low values	Cheese process r (s)	319	1653	906	231	0.58	16%	1.54	
	Dry matter intake (kg/d)	8.8	36.2	19.9	4.5	0.71	12%	1.83	
	Fresh cheese yield (g curd/100 g milk)	7.40	47.93	26.76	6.45	0.73	12%	1.91	
	Methane emitted (g/d)	180	786	413	102	0.68	14%	1.79	
	Milk C14_1 (g/dL)	0.00	0.15	0.04	0.02	0.68	21%	1.78	
	Milk C16_1c (g/dL)	0.01	0.24	0.07	0.03	0.73	20%	1.91	
	Milk C18_2c9c12 (g/dL)	0.00	0.17	0.06	0.02	0.72	19%	1.91	
	Milk C18_3c9c12c15 (g/dL)	0.00	0.09	0.02	0.01	0.68	22%	1.77	
	Milk isoanteiso FA (g/dL)	0.02	0.28	0.09	0.03	0.75	14%	2.00	
	Milk Magnesium (mg/kg)	61	157	100	13	0.72	7%	1.88	5
	Milk omega3 (g/dL)	0.00	0.11	0.03	0.01	0.66	22%	1.73	
	Milk omega6 (g/dL)	0.01	0.33	0.10	0.03	0.72	14%	1.89	
	Milk Phosphorus (mg/kg)	509	1447	999	124	0.75	6%	1.99	
	Milk Potassium (mg/kg)	819	1985	1524	147	0.55	6%	1.48	
	Milk Tot18_2 (g/dL)	0.01	0.32	0.10	0.03	0.69	15%	1.79	
	N efficiency (%)	9.8	81.7	36.9	10.3	0.71	15%	1.87	
	N losses (kg/d)	0.04	0.81	0.31	0.11	0.65	20%	1.69	
	Weight of cows(kg)	448	832	617	73	0.70	6%	1.83	
Extreme values	Blood BHB (mmol/L)	0.19	3.46	0.77	0.48	0.70	35%	1.81	
	Blood IGF-I (mg/L)	13	436	107	71	0.61	42%	1.59	
	Lactoferrin (mg/L)	7	1248	299	222	0.66	44%	1.71	6
	Milk BHB (mmol/L)	0.05	1.60	0.22	0.17	0.75	46%	1.97	
	Milk C18_2c9t11 (g/dL)	0.00	0.14	0.03	0.02	0.74	37%	1.95	

Too bad

Phenotype	Min	Max	Mean	SD	R^2cv	Relative RMSEcv	RPDcv	Cluster
Blood Glucose (mmol/L)	1.93	4.51	3.47	0.47	0.44	10%	1.33	
Blood NEFA (µekv/L)	26	1956	672	440	0.39	51%	1.28	
Cheese process k20 (s)	160	386	225	39	0.34	13%	1.24	
Milk Glucose Free (mmol/L)	0.00	0.69	0.24	0.11	0.50	32%	1.41	
Milk Glucose6Phosphate (mmol/L)	0.00	0.93	0.16	0.08	0.49	36%	1.40	7
Milk IsoCitrate (mmol/L)	0.02	2.90	0.17	0.10	0.11	55%	1.06	
Milk Natrium (mg/kg)	234	1273	356	91	0.44	15%	1.34	
Milk Progesterone (ng/ml)	0.50	22.44	5.22	2.74	0.08	50%	1.05	
Milk Uric Acid (µmol/L)	2.4	348.5	158.8	54.6	0.32	28%	1.22	

Bring together predictors even if there are less informative …

Well-being: The agglomerative hierarchical clustering



Milk
- Milk Yield
- Fat
- Protein
- Lactose
- FPCM

Minerals
- Sodium
- Calcium
- Magnesium
- Phosphote
- Potassium

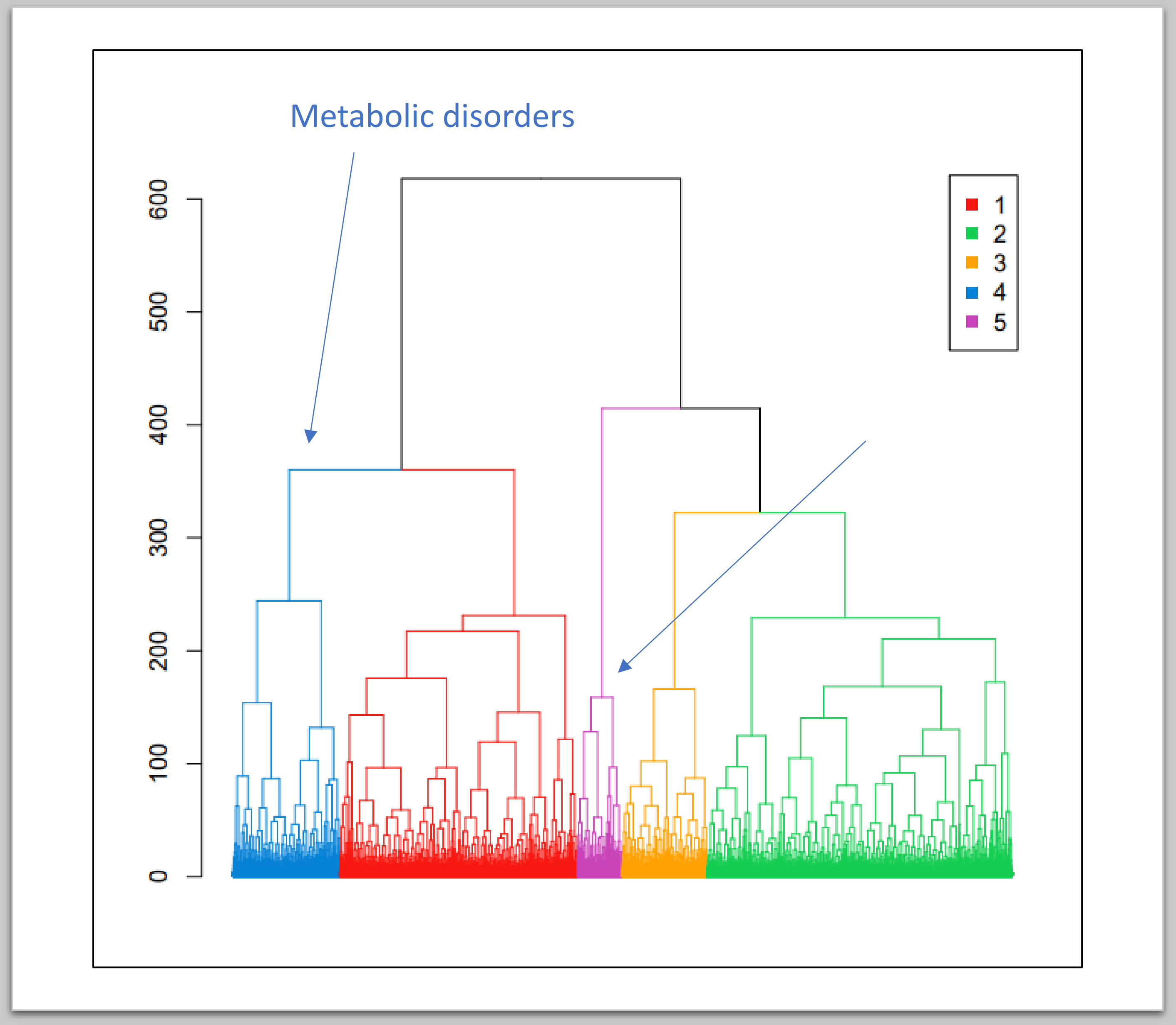
Fat
- Saturated FA
- C181 FA
- Monounsaturated FA
- Polyunsaturated FA

Metabolism
- Methane
- BHB
- Protein Efficiency
- Energy Balance
- acetone

Health
- Somatic Cells Count
- Citrate

Feeding
- Weight
- DMI
- DMI2
- CI
- RFI
- RFI2

- Detection of cows with abnormal behaviors

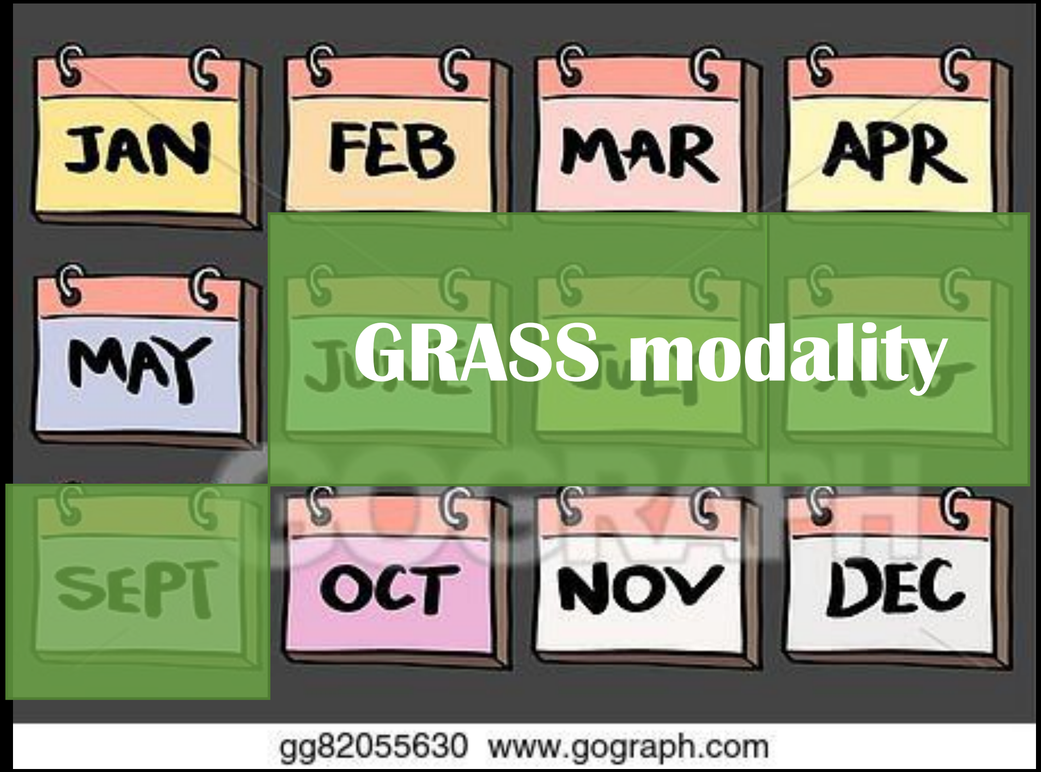


Franceschini et al., 2022

Grass indicators

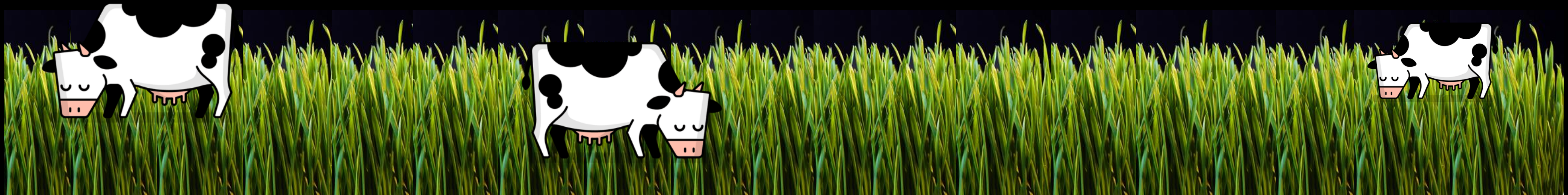
Soyeurt et al., 2022

Grass-based prediction

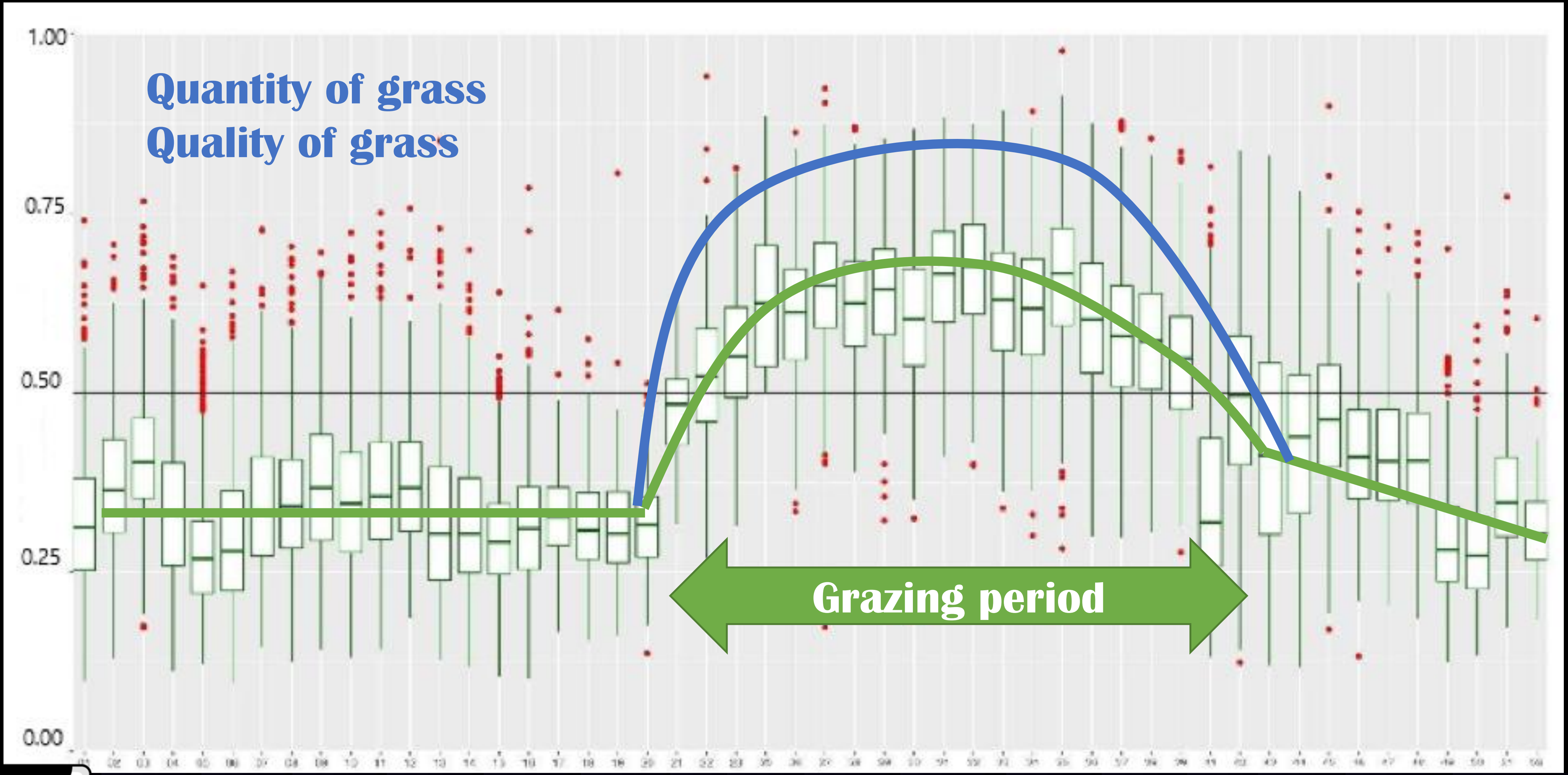


- **Calibration accuracy: 88 %**
- **Validation accuracy: 87%**

Coppa et al. (2021)
Frizarrin et al. (2021)



GRASS : Not a binary variable

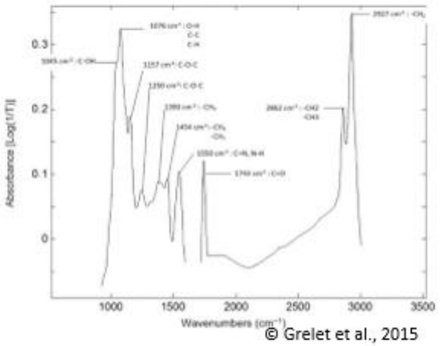


Many reasearch developments BUT …

Limited applications on field at herd and (a little bit less for) individuals …

Intrinsic constraints

Inspird from ©Pixabay



Predictive model = Equation

Methods 186 (2021) 97–111

Contents lists available at ScienceDirect

Methods

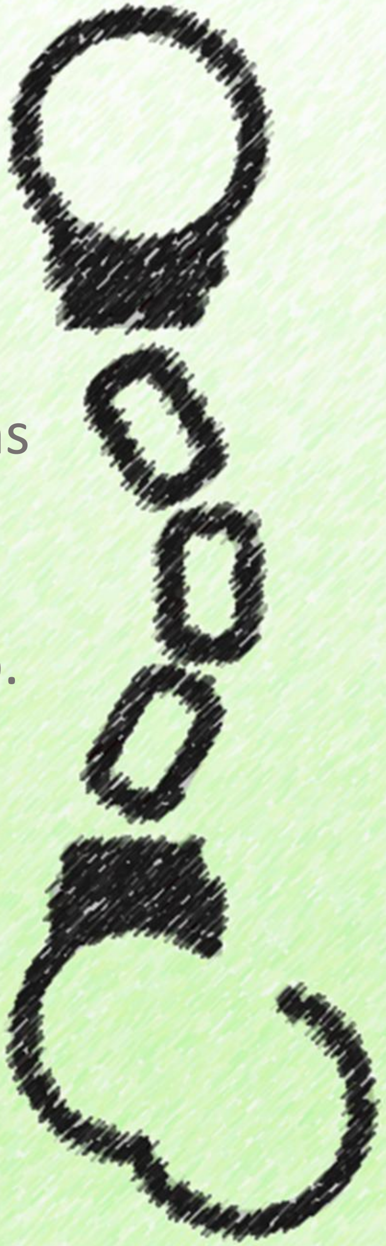
journal homepage: www.elsevier.com/locate/ymeth

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. Stevens [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*

Solutions
&
Future
Develop.



Inspird from ©Pixabay

Intrinsic Constraints

Providing a detailed information about the calibration dataset used (records + model)

➔ Discussion with spectrometer providers

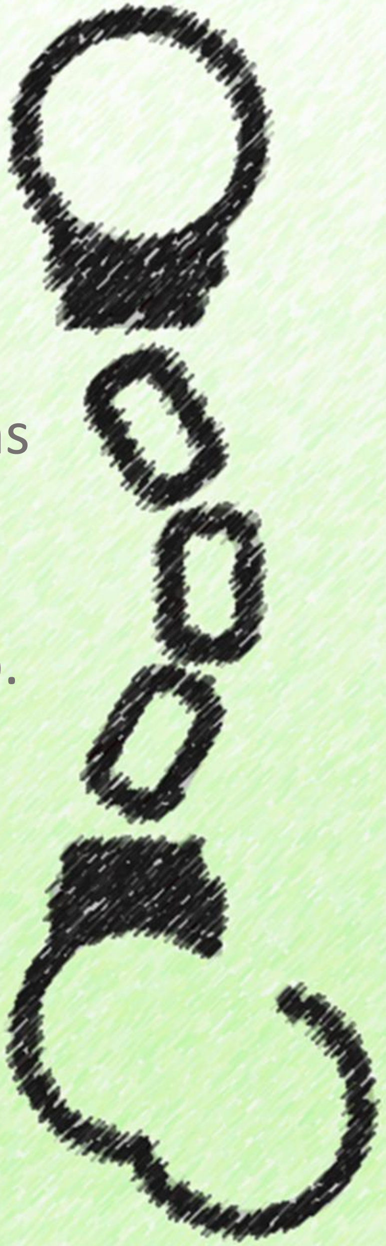
Promoting a large spectral and y variability in the calibration set

➔ International collaborations

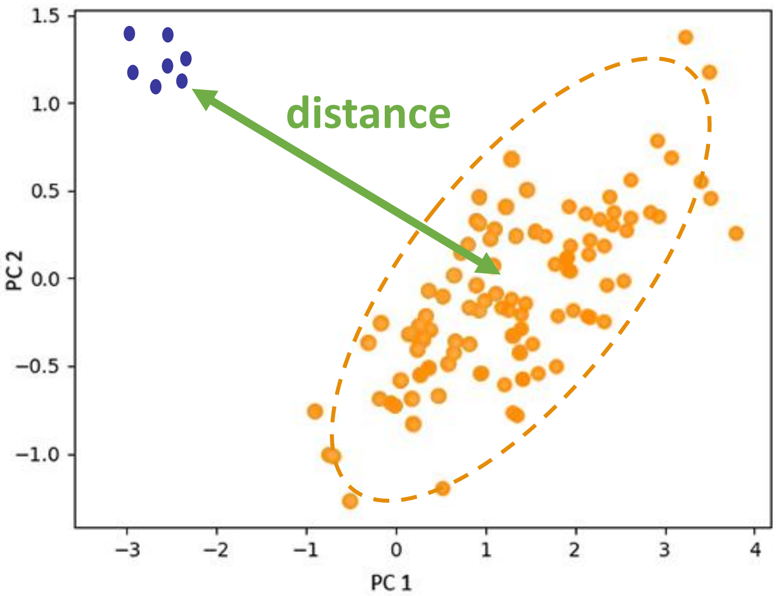
Limiting the extrapolation by sharing a « outlier » file

➔ Distance could be calculated to avoid spectral extrapolation

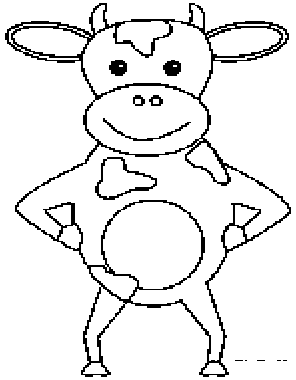
Solutions
&
Future
Develop.



Outlier file



© Modified from scikit-learn. 2019



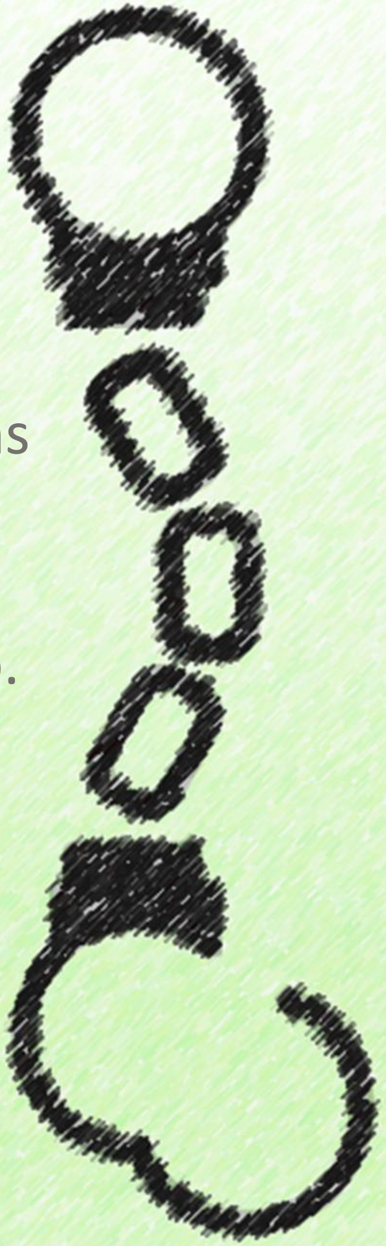
$$GH = \left((\overline{x} - \overline{\mu})^T S^{-1} (\overline{x} - \overline{\mu})\right) / nPC$$

GH > 3 ➔ no prediction

We don’t need the calibration set but just few statistical parameters:

- The **averaged spectrum** calculated from the calibration set
- The **matrix of eigenvectors** obtained after applying a Principal Component Analysis on the calibration spectra

Solutions
&
Future
Develop.



Inspird from ©Pixabay

Intrinsic Constraints

Providing a detailed information about the calibration dataset used (records + model)

➔ Discussion with spectrometer providers

Promoting a large spectral and y variability in the calibration set

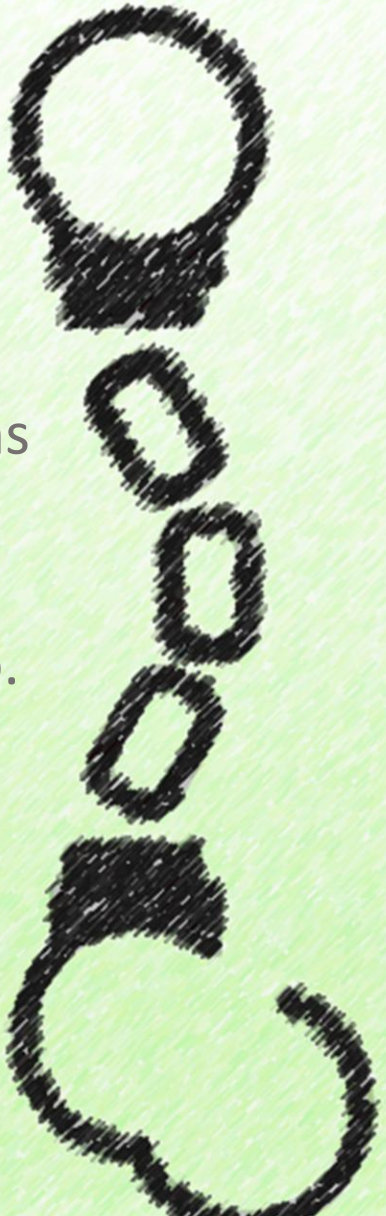
➔ International collaborations

Limiting the extrapolation by sharing a « outlier » file

➔ Distance could be calculated to avoid spectral extrapolation

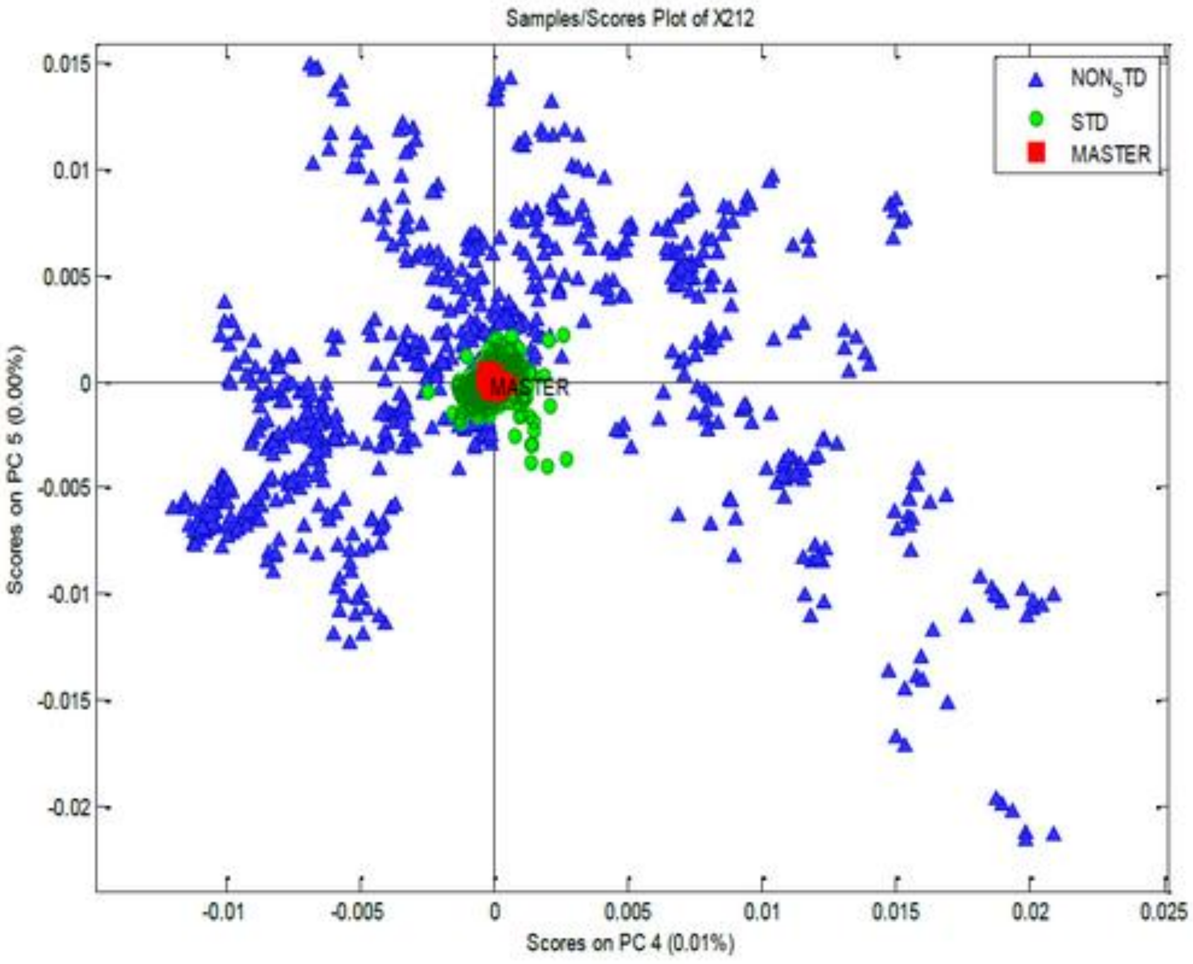
Use comparable spectral data

➔ Spectral standardization

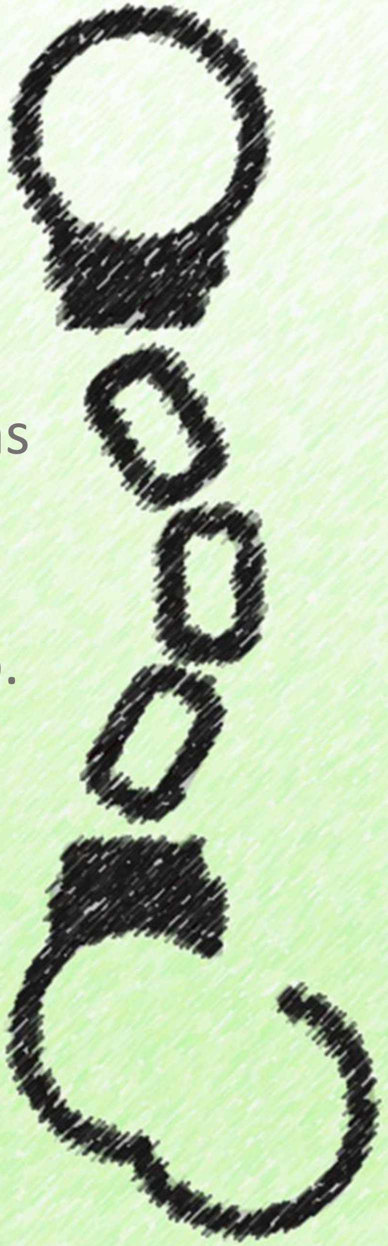


Solutions
&
Future
Develop.

Spectral standardization



Solutions
&
Future
Develop.



Inspird from ©Pixabay

Intrinsic Constraints

Providing a detailed information about the calibration dataset used (records + model)

➔ Discussion with spectrometer providers

Promoting a large spectral and y variability in the calibration set

➔ International collaborations

Limiting the extrapolation by sharing a « outlier » file

➔ Distance could be calculated to avoid spectral extrapolation

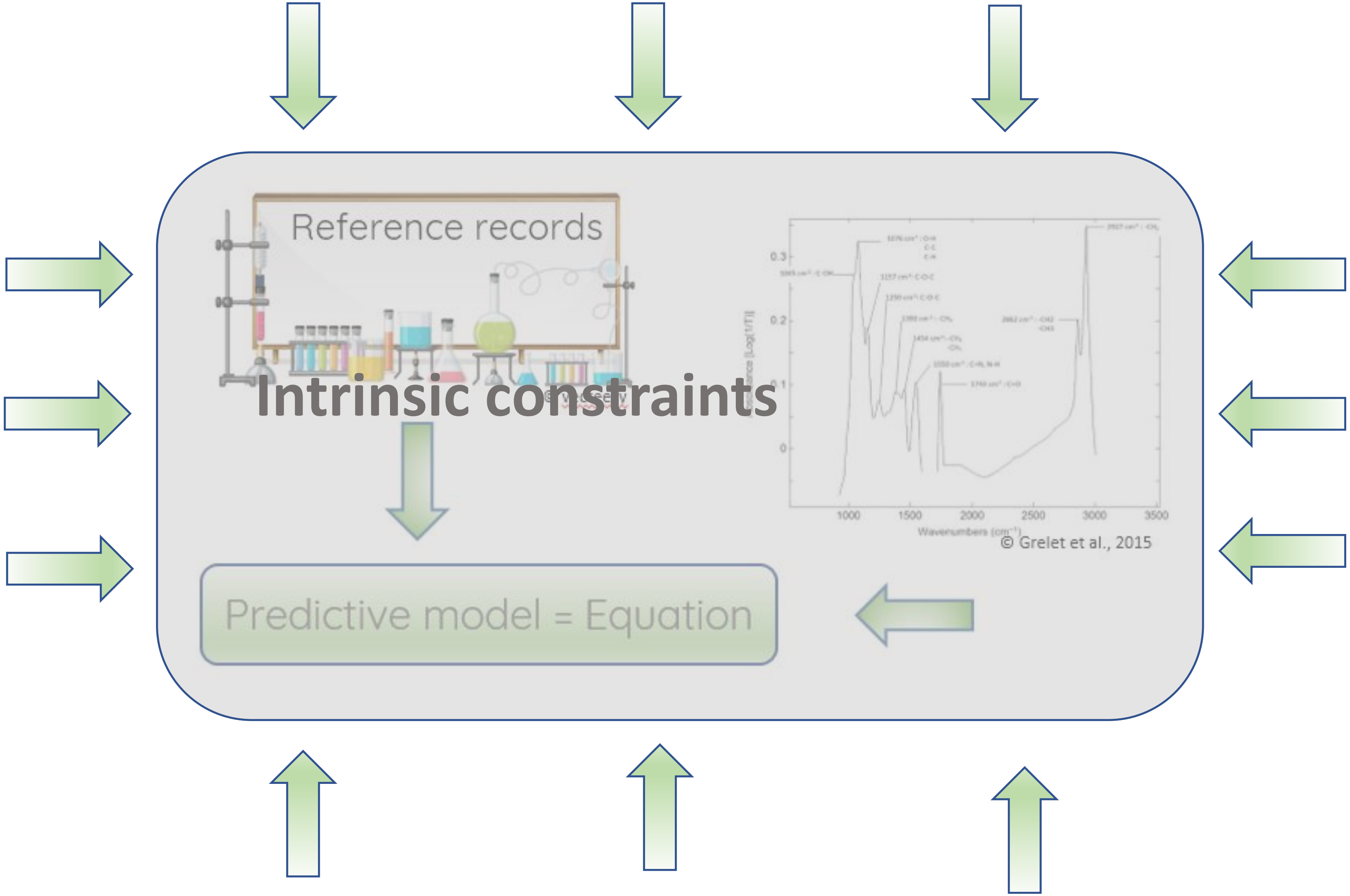
Use comparable spectral data

➔ Spectral standardization

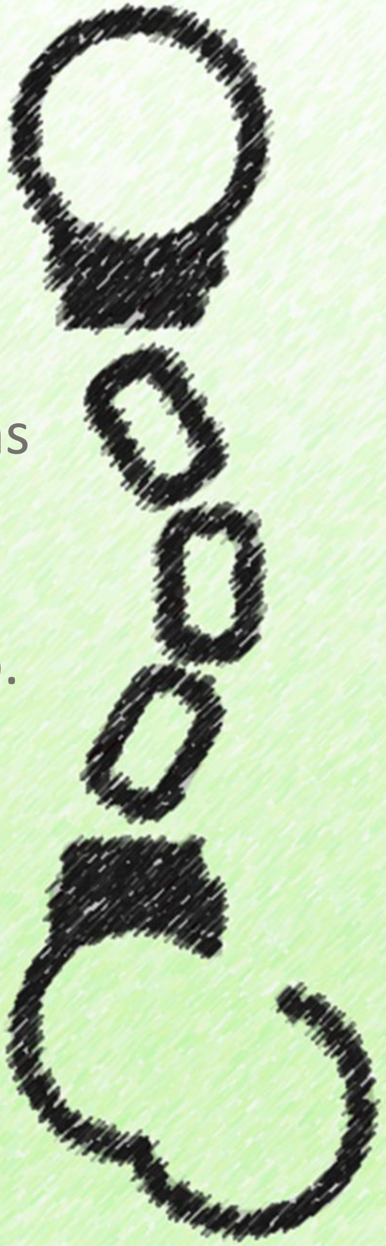
As multiple equations exist for the same trait, realizing a common validation managed by an independent institution.

➔ All metrics of validation will be directly comparable

Extrinsic constraints



Solutions
&
Future
Develop.



Inspird from ©Pixabay

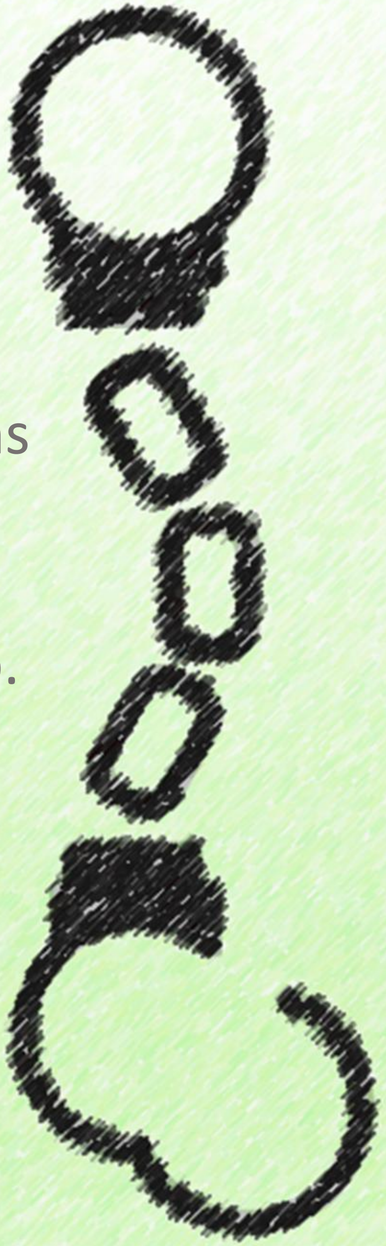
Extrinsic Constraints

Problem of equation access or equation representativeness

➔ Promoting a **common effort** in the development of future equations (cost decrease, better representativity)

➔ Creating an **appropriate environment in independent structure** (IP, data confidentiality, model dissemination and validation …)

Solutions
&
Future
Develop.



Inspird from ©Pixabay

Extrinsic Constraints

Problem of equation access or equation representativeness

➔ Promoting a **common effort** in the development of future equations (cost decrease, better representativity)

➔ Creating an **appropriate environment in independent structure** (IP, data confidentiality, model dissemination and validation …)

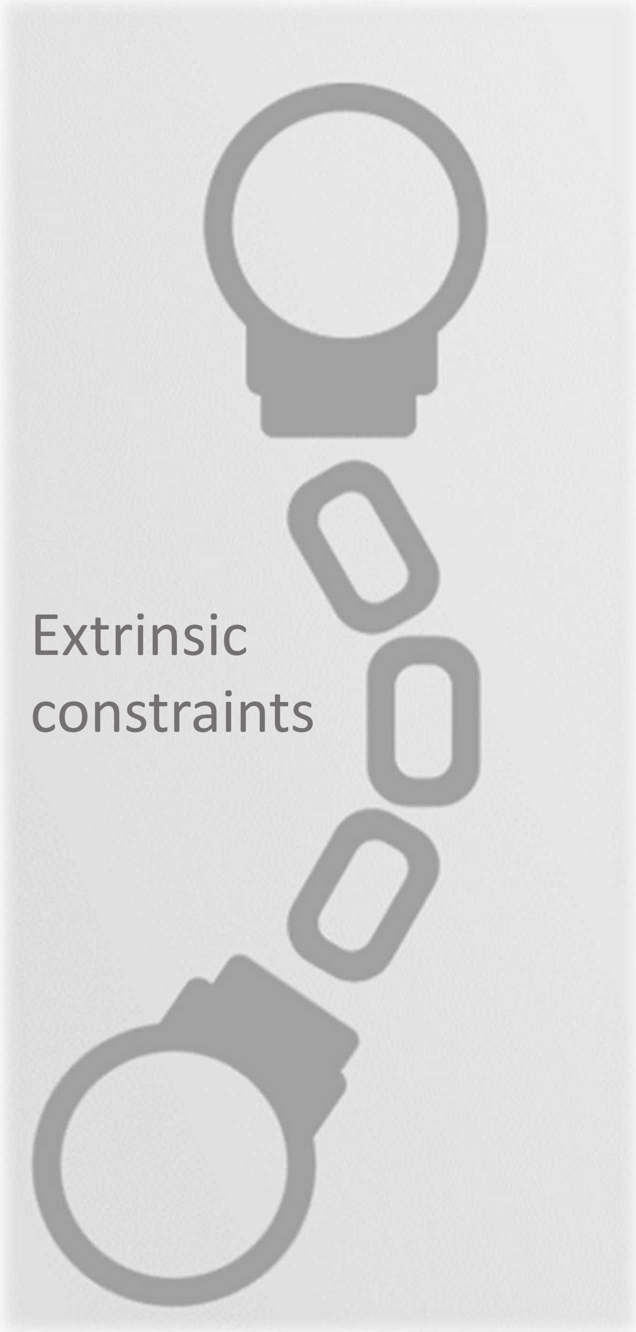
No economic or administrative incentive

➔ Don’t be afraid to change the rule

➔ Improving the **communication** between all stakeholders

➔ Sharing the data collected by different stakeholders (herd vs individuals): common data platform ?

➔ What is possible and what is not practically and scientifically?



Inspird from ©Pixabay

Extrinsic constraints

Problem of equation access or equation representativeness

No economic or administrative incentive

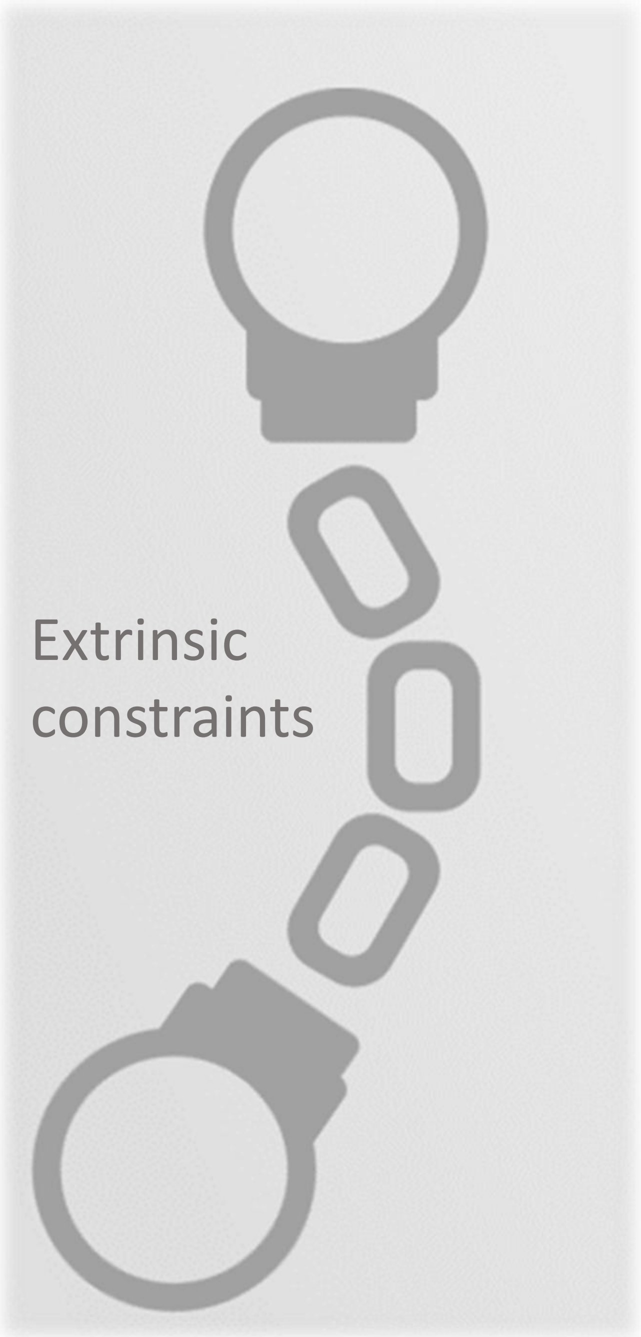
➔ Fatty acids : Milk price still defines based on milk fat and protein contents

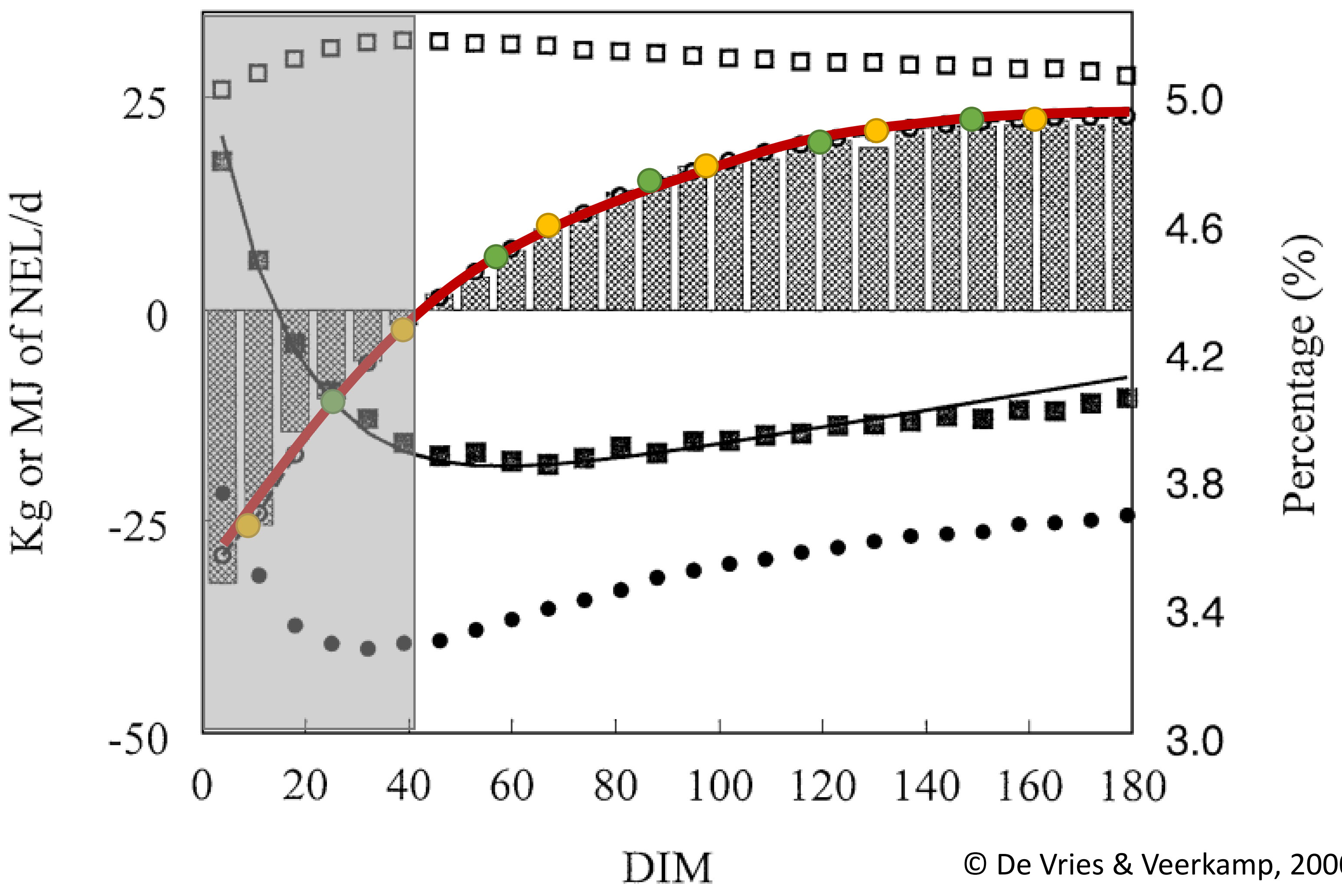
➔ Technological properties of milk

➔ Methane or nitrogen efficiency : No real economic interest to take into account the environmental fingerprint

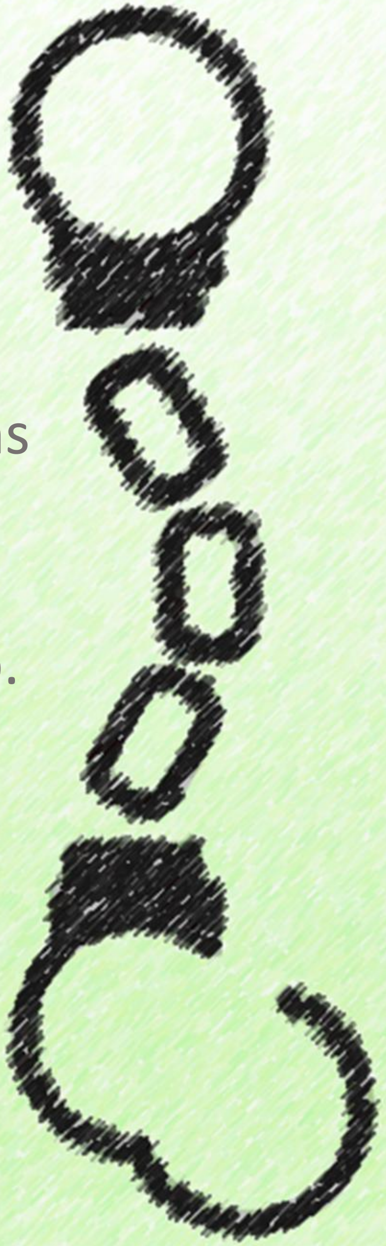
Too low frequency data acquisition

➔ Energy balance, body weight, mastitis, ….


Extrinsic
constraints


Kg or MJ of NEL/d
25
0
-25
-50
0
20
40
60
80
100
120
140
160
180
DIM
5.0
4.6
4.2
3.8
3.4
3.0
Percentage (%)

Solutions
&
Future
Develop.



Inspird from ©Pixabay

Extrinsic Constraints

Problem of equation access or equation representativeness

➔ Promoting a **common effort** in the development of future equations (cost decrease, better representativity)

➔ Creating an **appropriate environment in independent structure** (IP, data confidentiality, model dissemination and validation …)

No economic or administrative incentive

➔ Don't be afraid to change the rule

➔ Improving the **communication** between all stakeholders

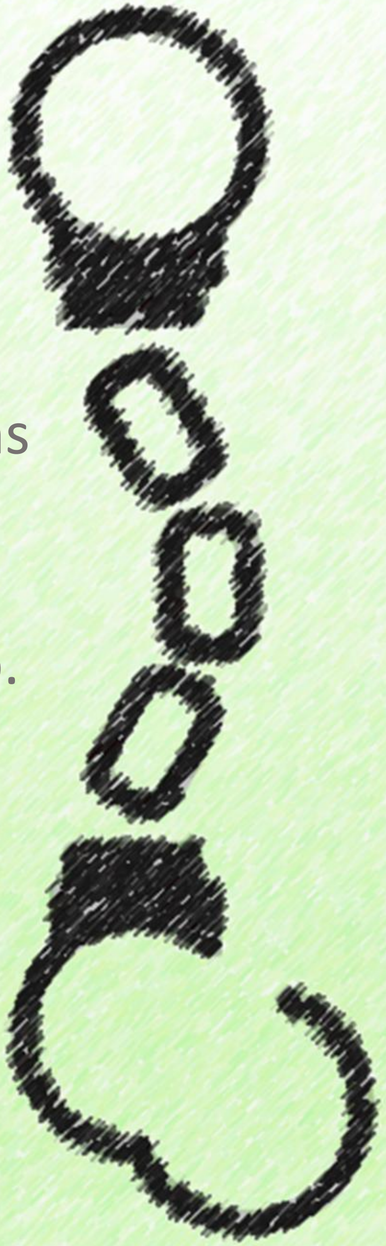
➔ Sharing the data collected by different stakeholders (herd vs individuals): common data platform ?

➔ What is possible and what is not practically and scientifically

Too low frequency data acquisition

➔ Allowing to the farmers to make additional analysis especially for cows in early lactation

Solutions
&
Future
Develop.



Inspird from ©Pixabay

Extrinsic Constraints

Problem of equation access or equation representativeness

➔ Promoting a **common effort** in the development of future equations (cost decrease, better representativity)

➔ Creating an **appropriate environment in independent structure** (IP, data confidentiality, model dissemination and validation …)

No economic or administrative incentive

➔ Don’t be afraid to change the rule

➔ Improving the **communication** between all stakeholders

➔ Sharing the data collected by different stakeholders (herd vs individuals): common data platform ?

➔ What is possible and what is not practically and scientifically

Too low frequency data acquisition

➔ Allowing to the farmers to make additional analysis especially for cows in early lactation

Problem of data visualization and communication

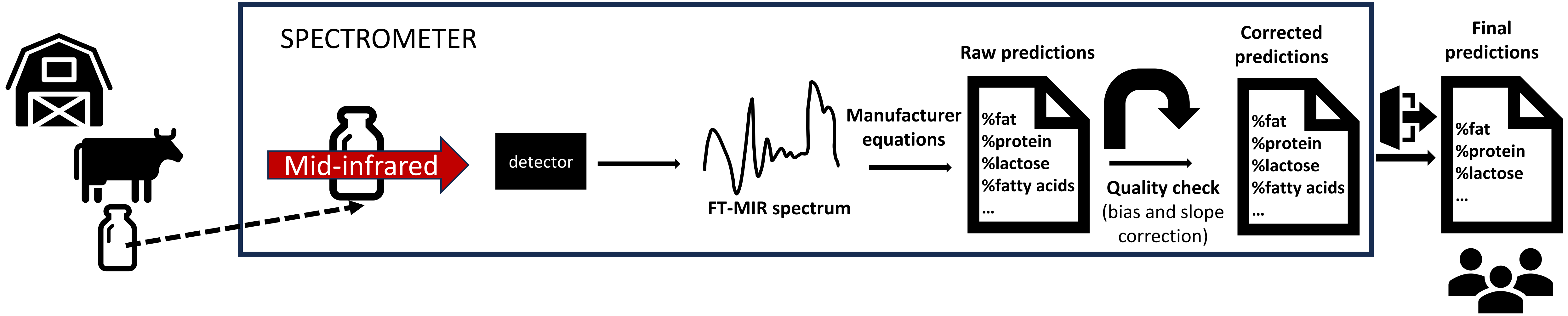
➔ Better communication with specialists of data visualization, scientists and farmers

Extra value

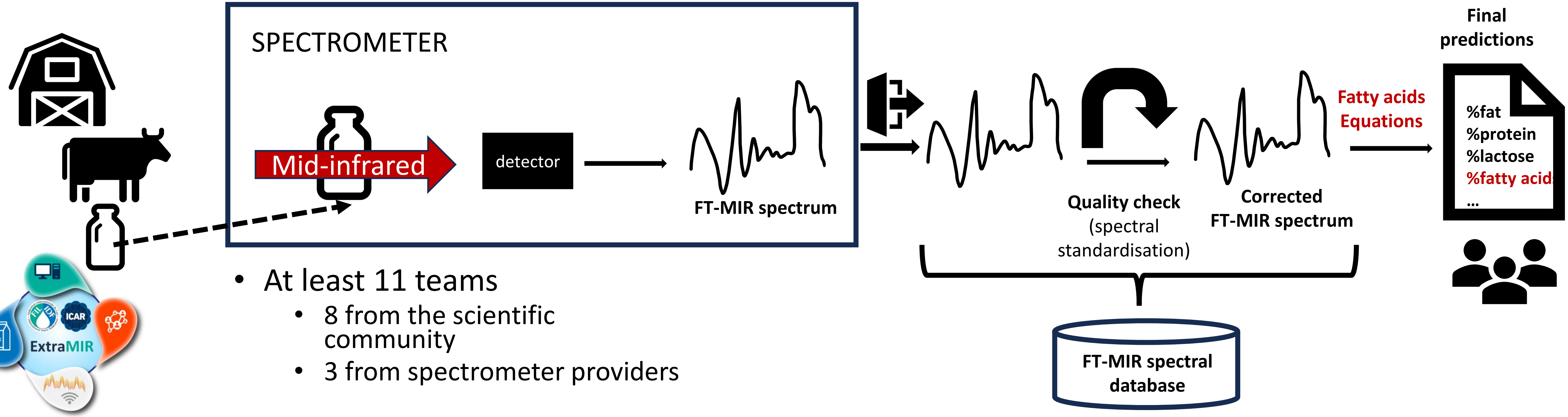
from- smart use of

-MIR spectra

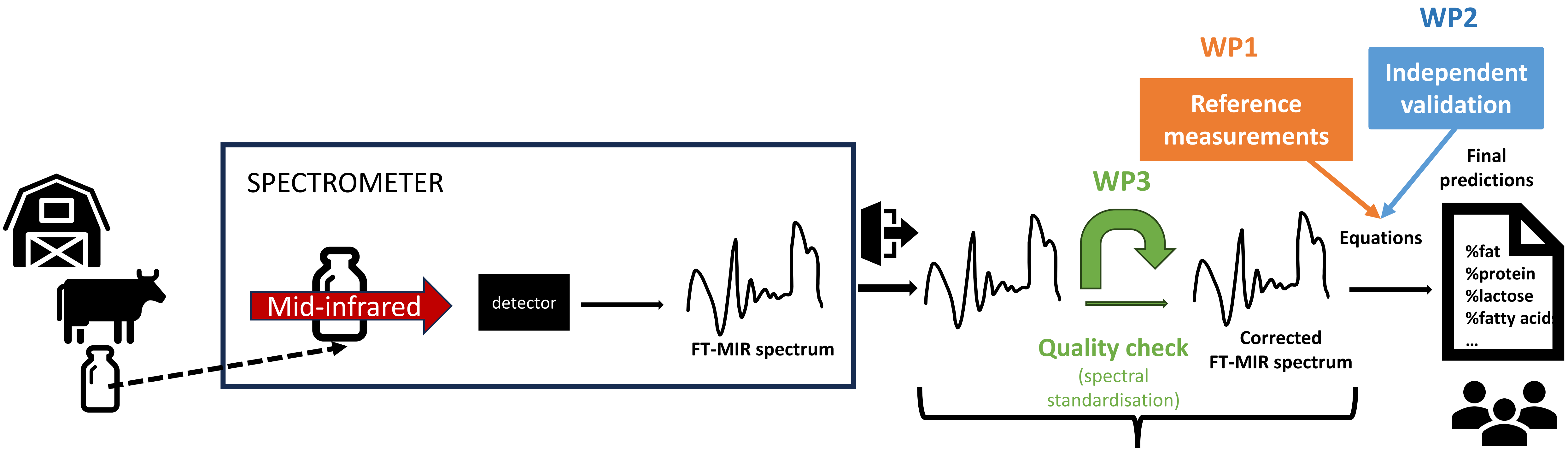
Existing Milk analysis process

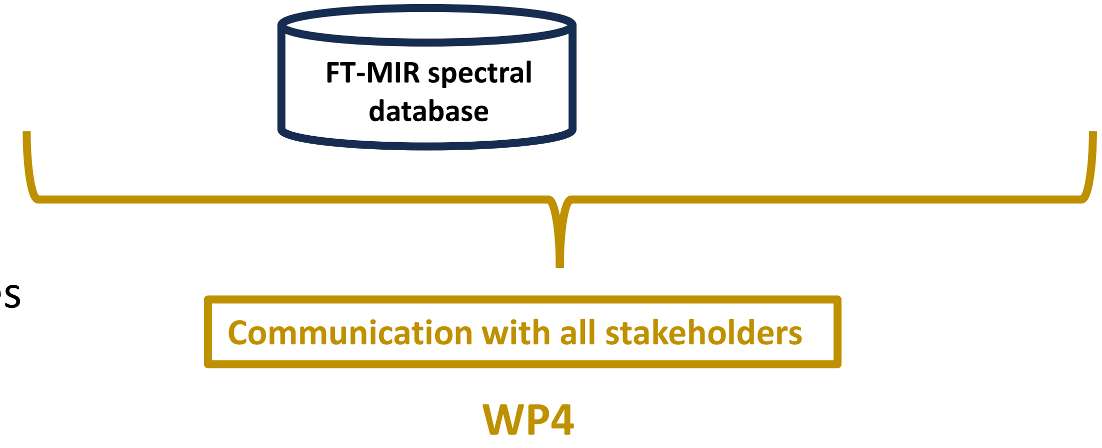


Proposal



- At least 11 teams
  - 8 from the scientific community
  - 3 from spectrometer providers



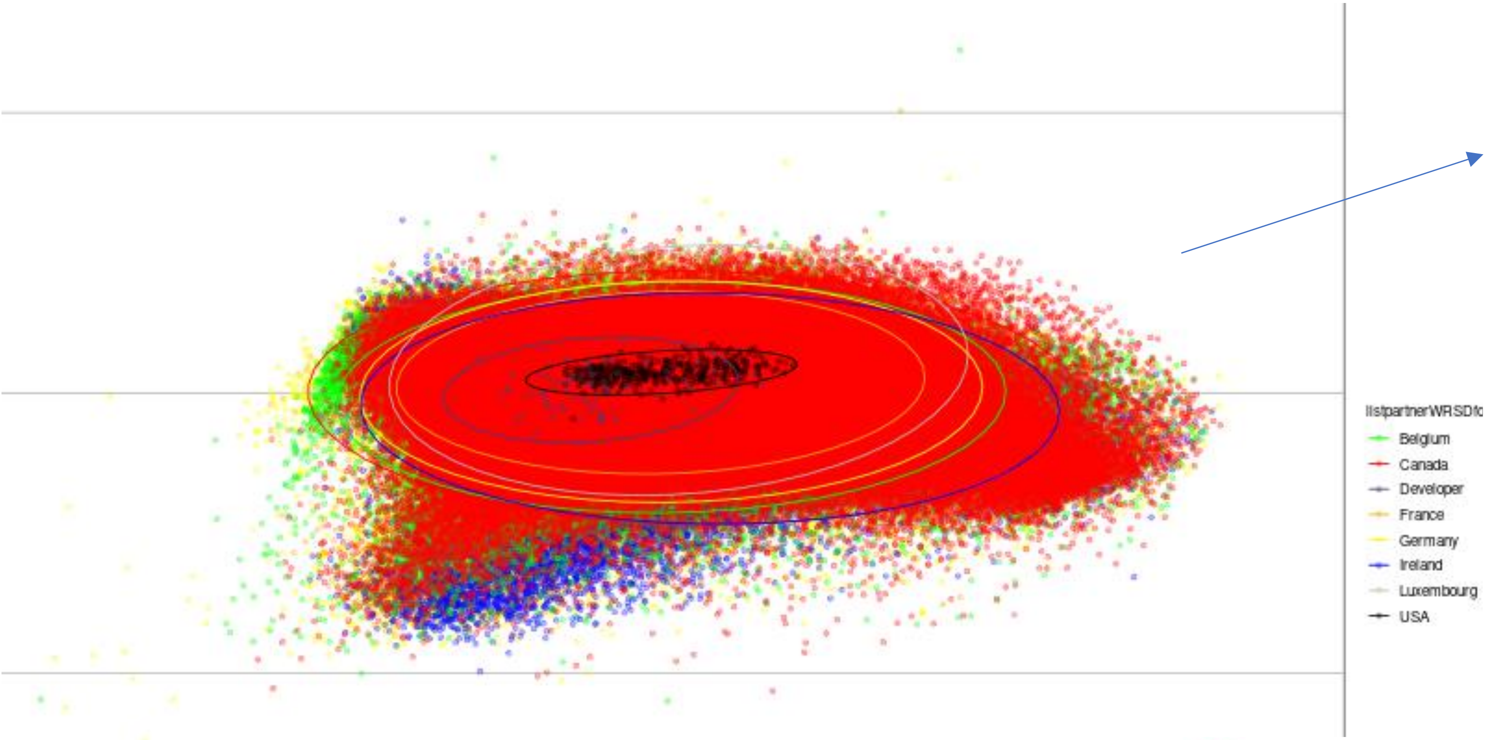


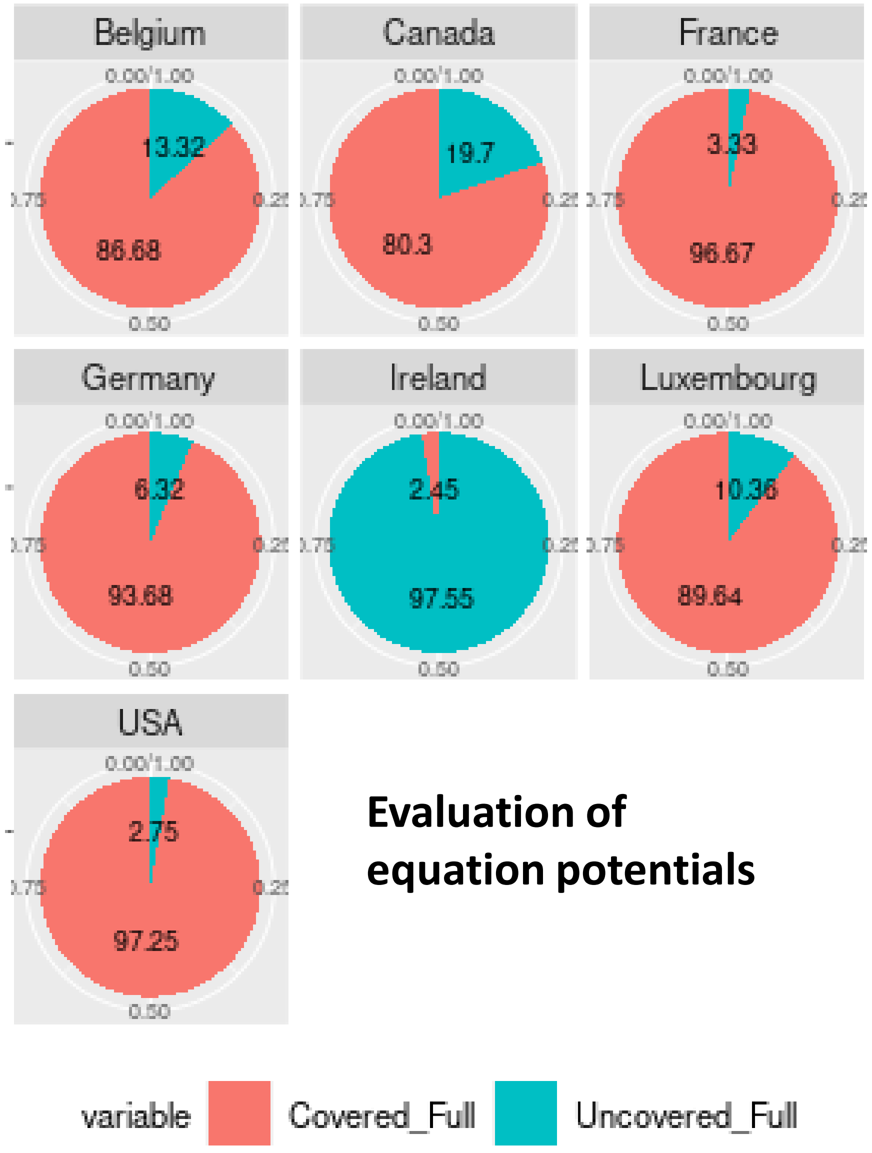
ExtraMIR will contribute to develop guidelines and tools to limit the prediction bias.

Can we collaborate to create equations ?

But …

ExtraMIR is developping a world representative spectral database …





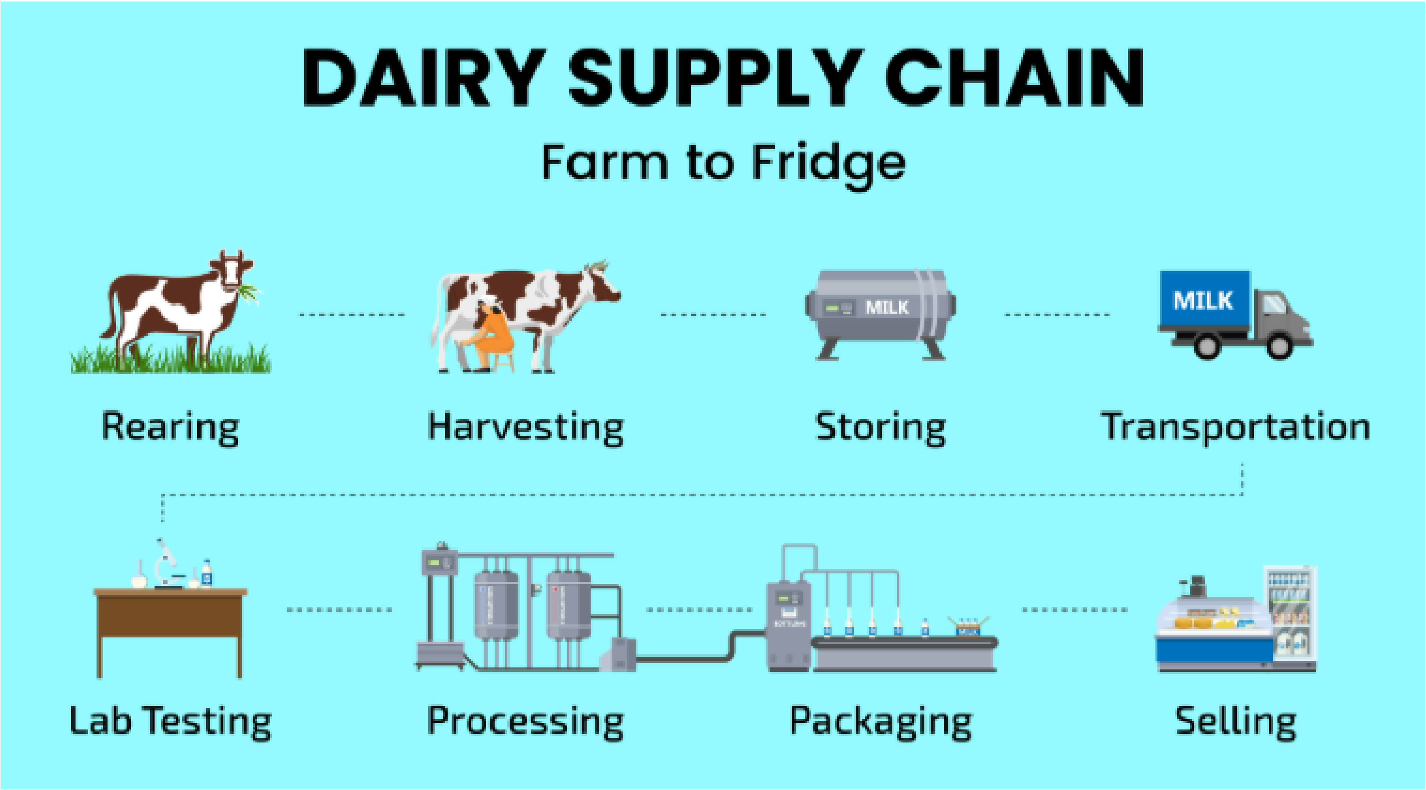
2 WRSD : one for Foss and one for Bentley

Canada (red) and Belgium (green) are ok.

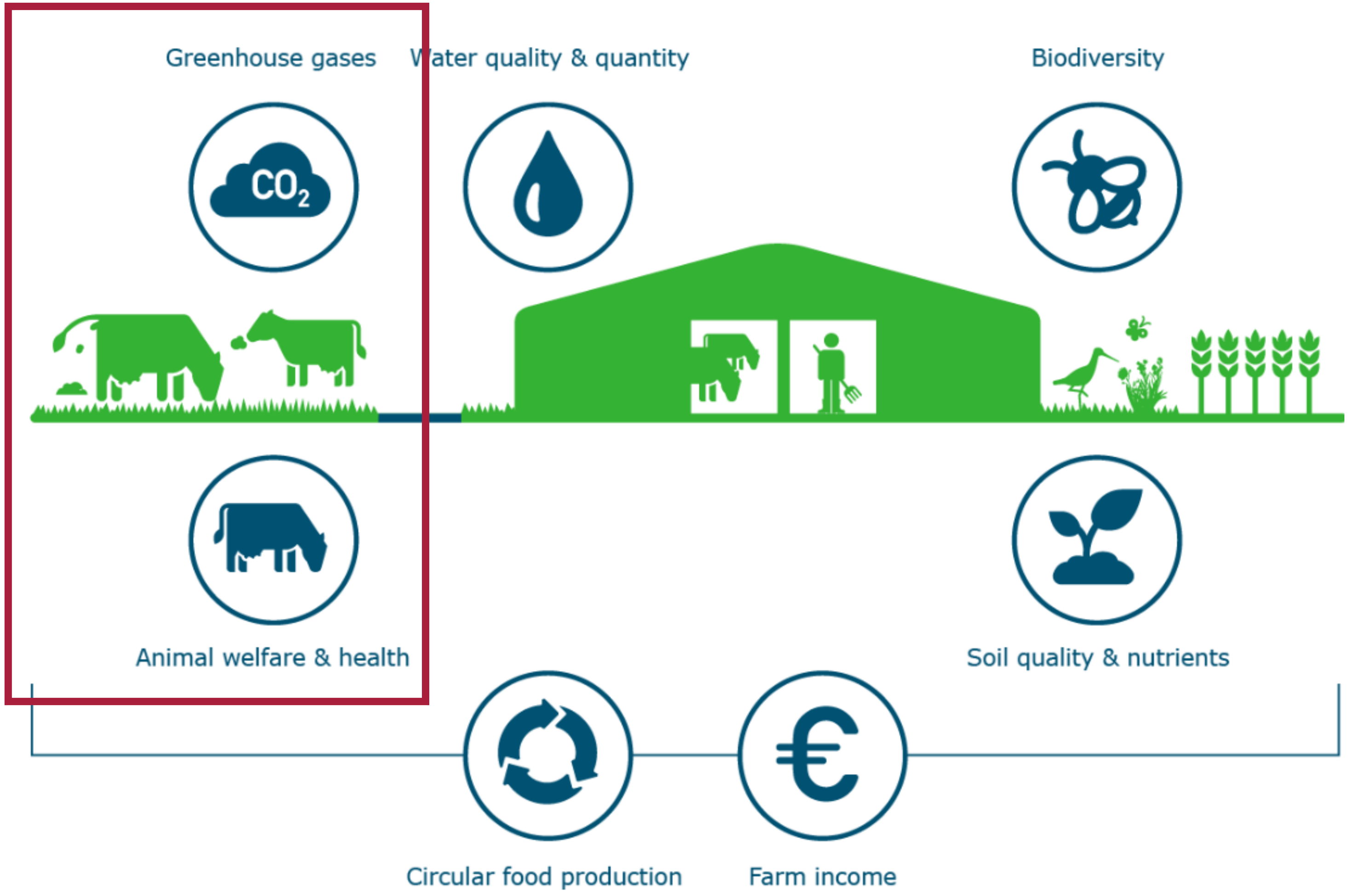
No data for The Netherlands

Evaluation of equation potentials

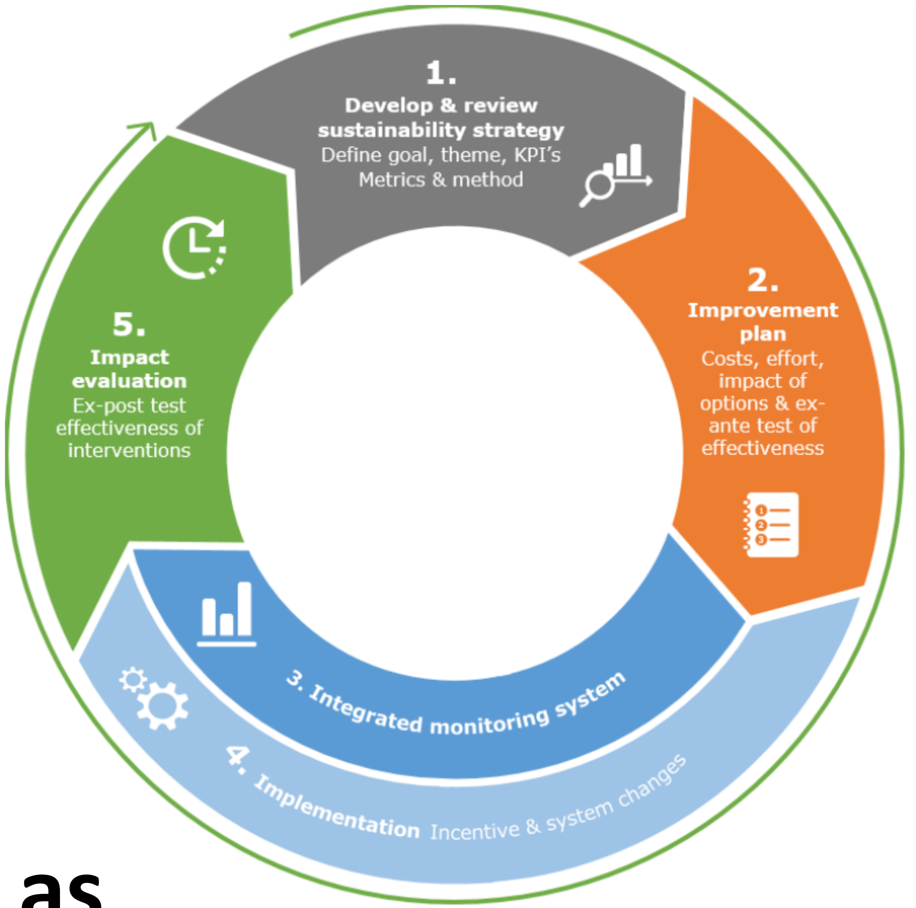
Consider the entire dairy chain ...



Having data at each step ...



Sustainable dairy production



New management and genetic tools as well as new products

https://www.wur.nl/en/show/working-towards-sustainable-dairy-production.htm

CONCLUSIONS

- MIR spectrometry
  - Largely used on practice
  - Many equations developed by researchers and companies to increase the number of phenotypes available
- Lack of guidelines to use efficiently the spectral data to take the maximum profit of this technology ➔ ExtraMIR + spectral standardisation
- Increase the data acquisition at the beginning of the lactation
- International collaborations to obtain cheaper equations with a larger dissemination
- Communication the potentiality to all stakeholders : bulk vs. Individual milk samples : combining the sources of information
- Take time to develop the most relevant decision tool for dairy farmers and industry in agreement with their expectations/needs

LIÈGE université
Gembloux
Agro-Bio Tech

Potentials of milk Mid-infrared Spectrometry:

Where are we ? What are the next steps?

Prof. Hélène Soyeurt

hsoyeurt@uliege.be

Rinderzucht Austria – Wien – Austria – June 2025