cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

EU-LI-PHE
EUROPEAN NETWORK ON LIVESTOCK PHENOMICS

LIÈGE université
Gembloux
Agro-Bio Tech

Funded by
the European Union

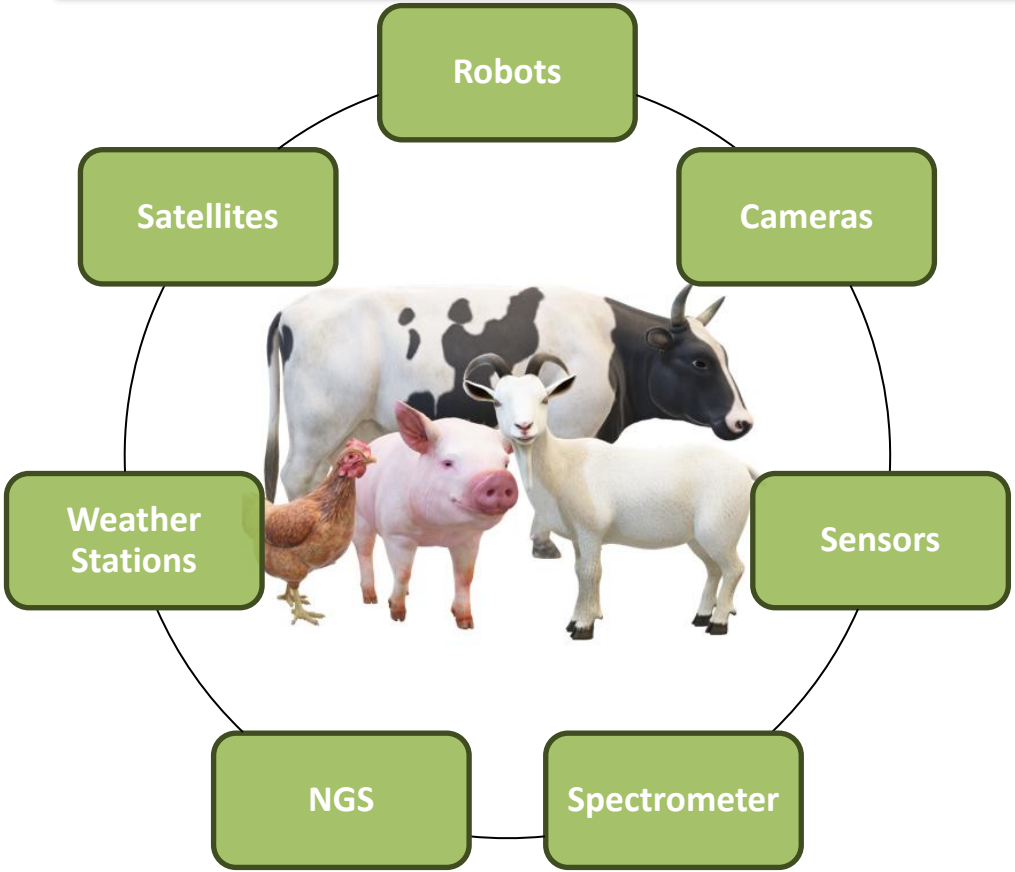
Combining high throughput phenotypes to study complex traits :

A case-study of negative energy balance using milk mid-infrared based predictions

Sébastien Franceschini

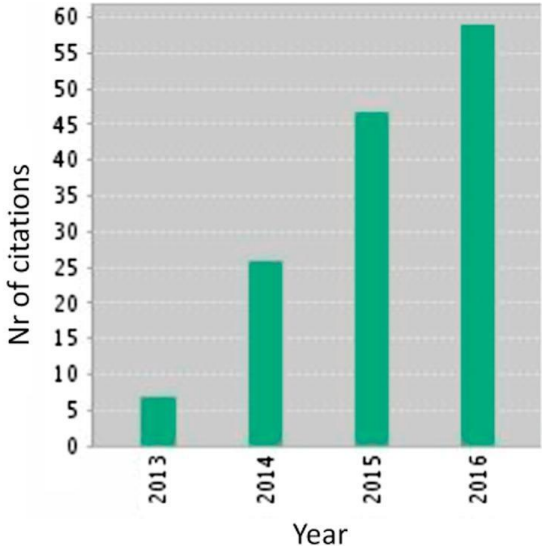
Supervised by N. Gengler and H. Soyeurt

New technologies = New phenotypes

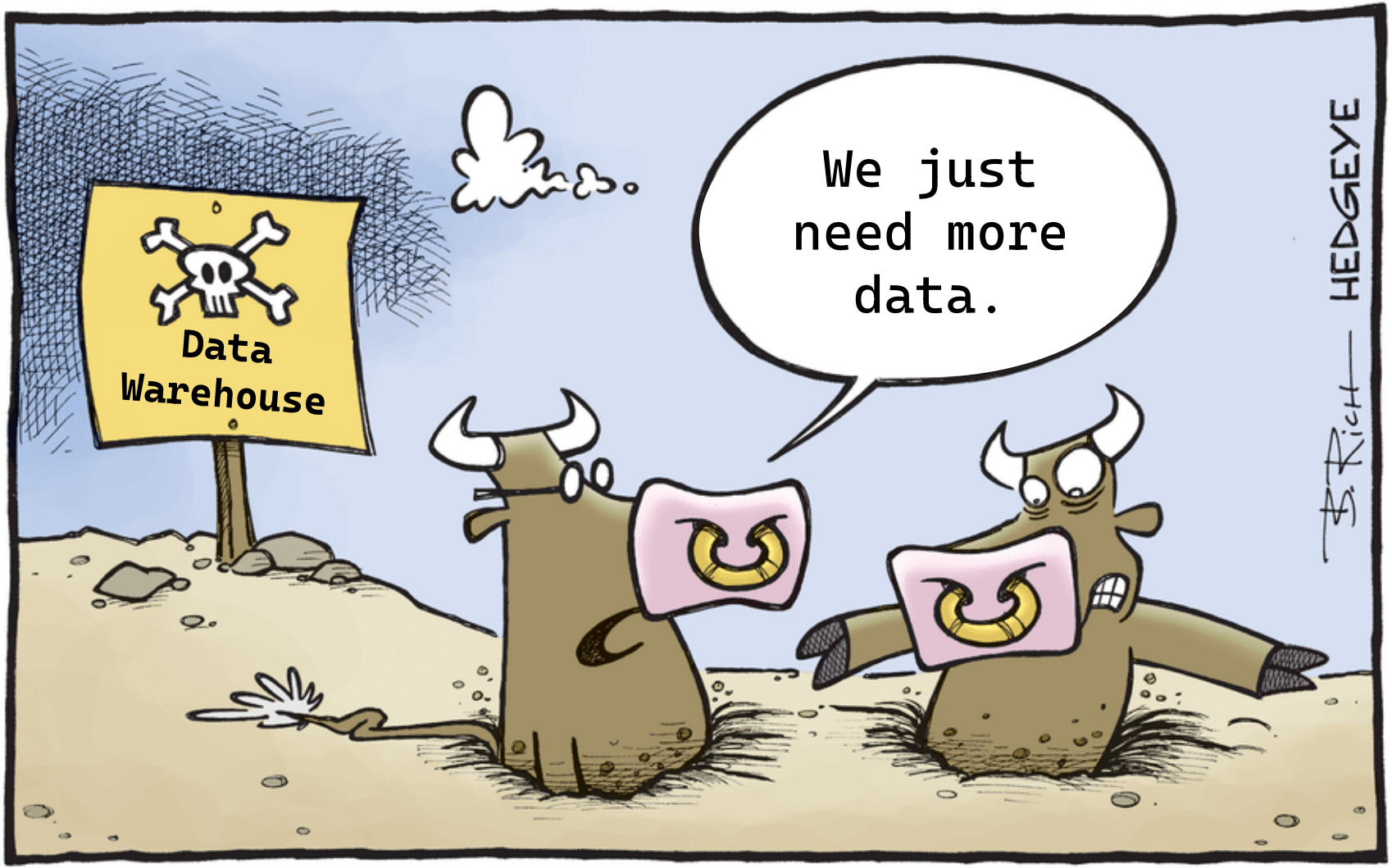


A huge opportunity !

« Big data in agriculture »



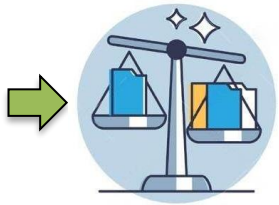
(A. Kamilaris et al., 2017)



(Modified from https://app.hedgeye.com/)

5 V's of big data

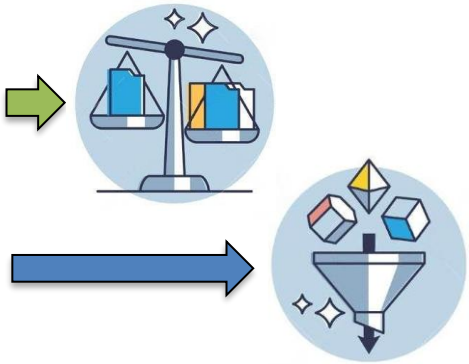
1. Volume



2. Variety
3. Velocity
4. Veracity
5. Value

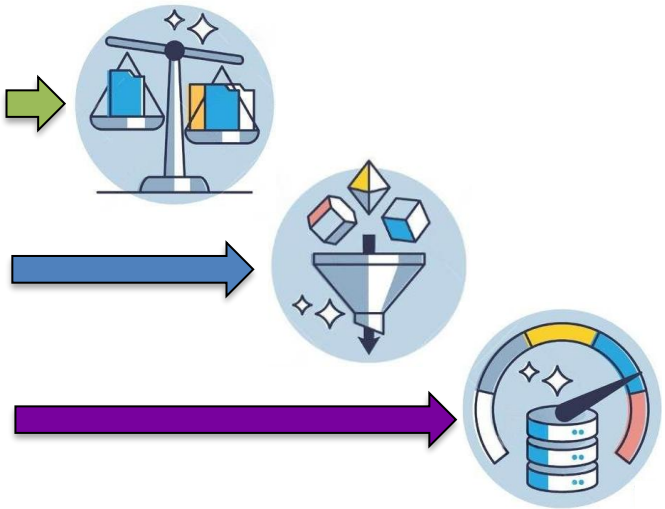
5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value



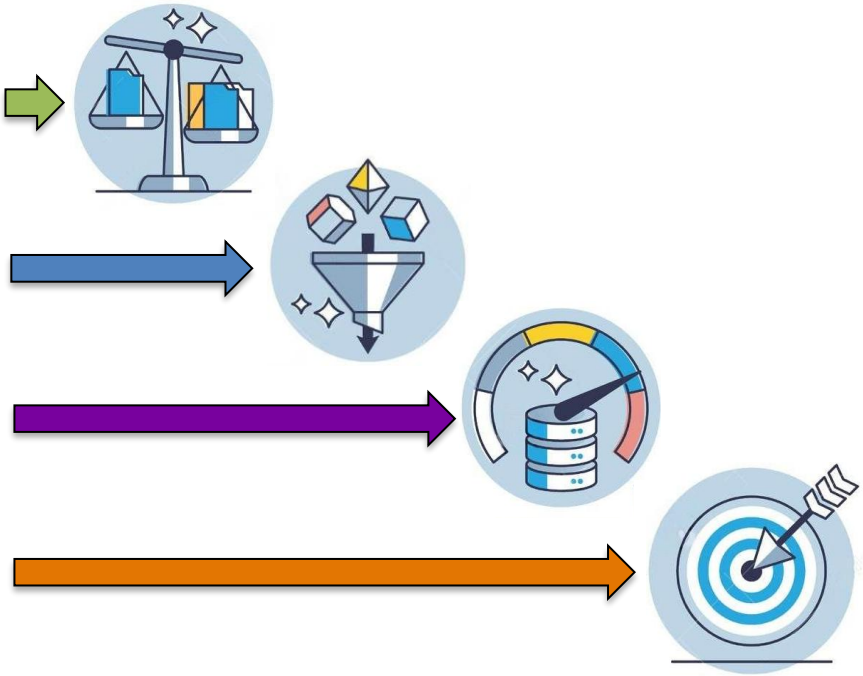
5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value



5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value



5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

Those are the challenges to overcome !

5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

Can we use the same algorithm ?

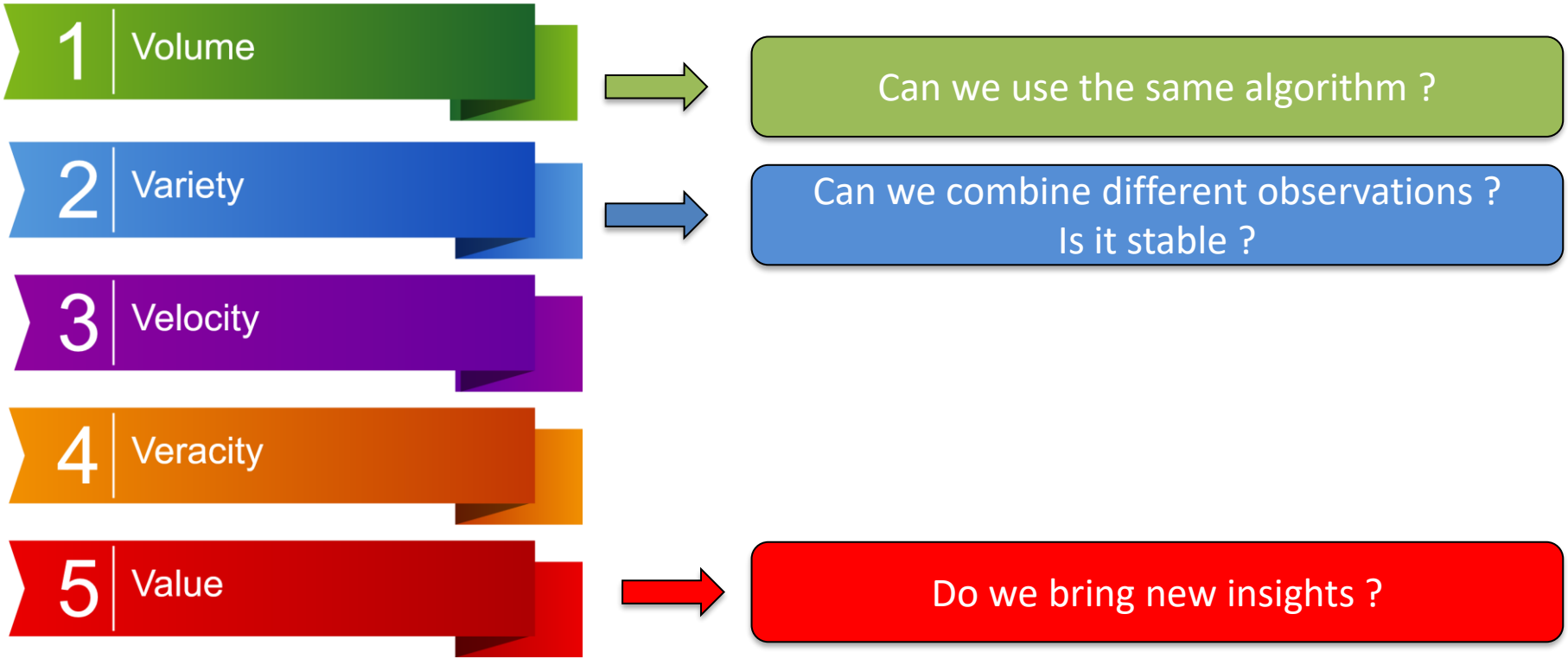
5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

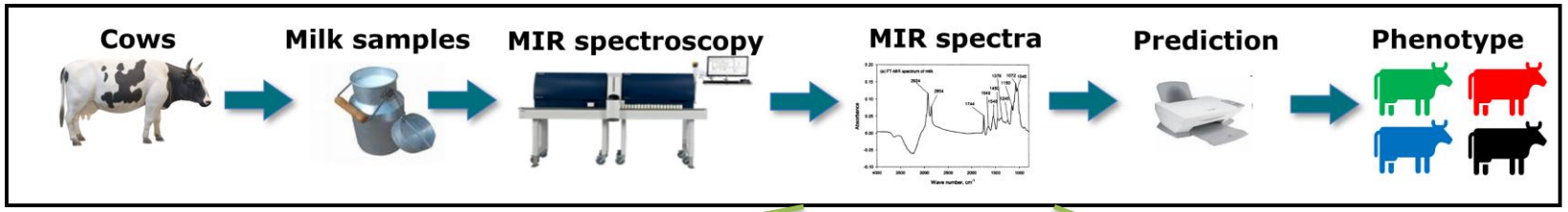
→ Can we use the same algorithm ?

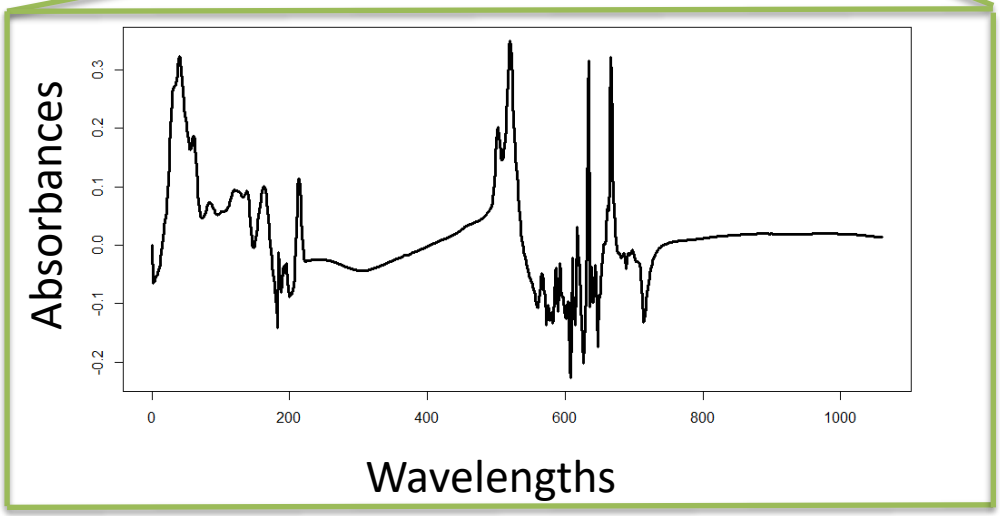
→ Can we combine different observations ?
Is it stable ?

5 V's of big data

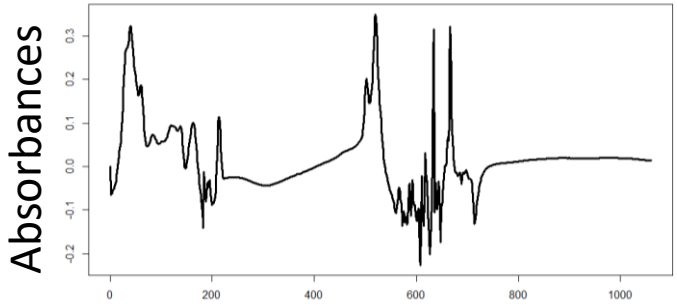


What is milk Mid-Infrared spectrometry ?





Predictions? Phenotypes ?



Fat (>0.99)

Protein (>0.80)

Fatty Acids (0.65-0.95)

Lactose (0.52)

Minerals (0.43-0.81)

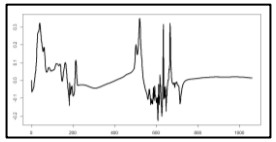
Methane (0.64)

Energy Balance (0.75)

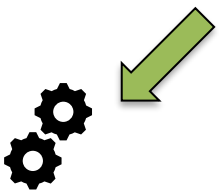
(Performance of the equation = R^2cv)

What can we do with our data ?

Supervised method



Collect Data

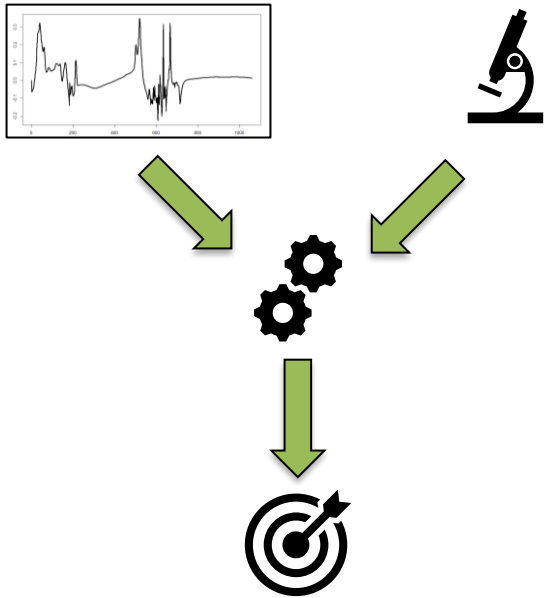


Create a model

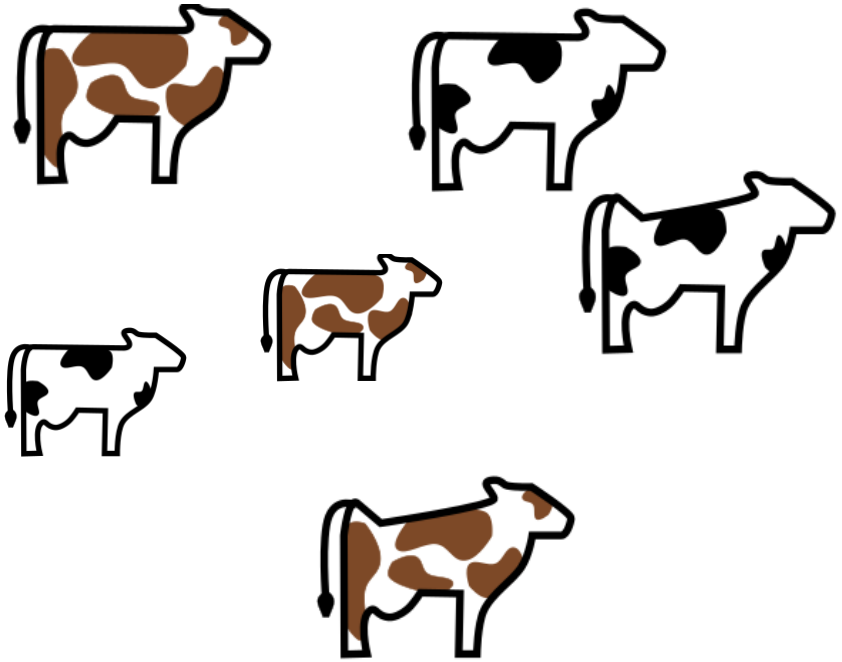
Predict new values

What can we do with our data ?

Supervised method

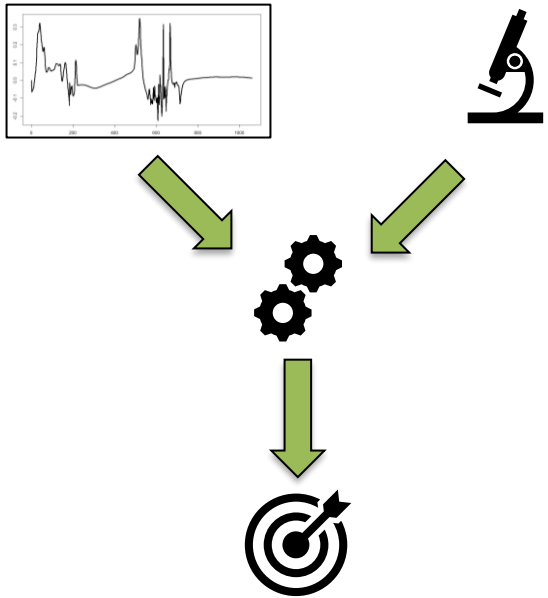


Unsupervised method

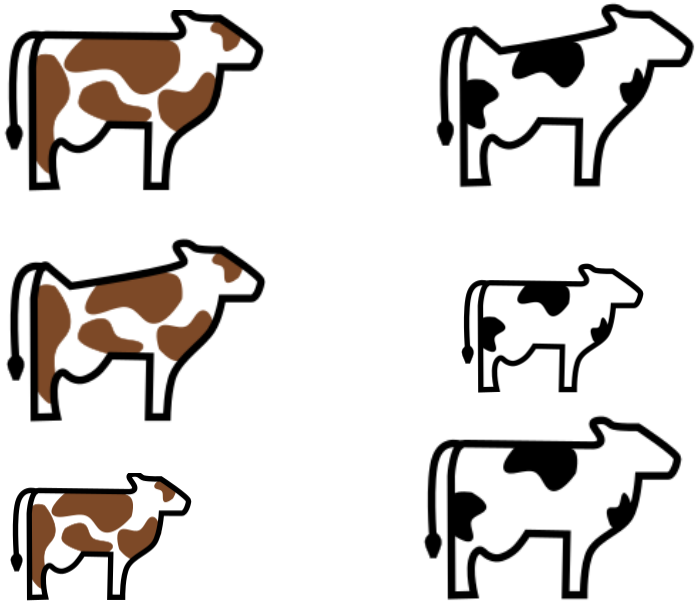


What can we do with our data ?

Supervised method

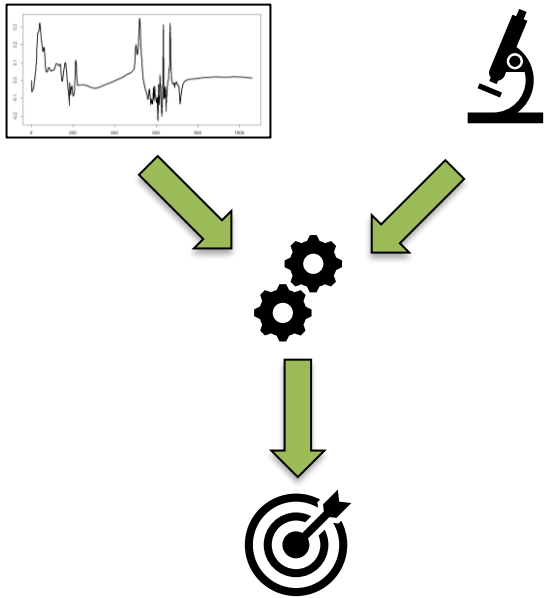


Unsupervised method

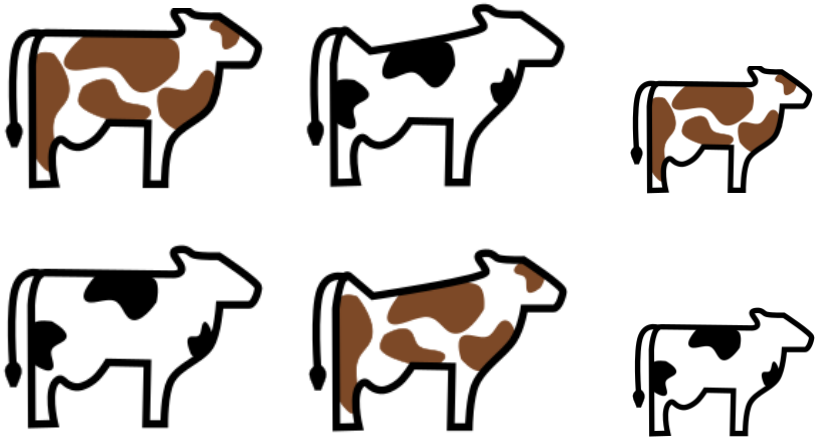


What can we do with our data ?

Supervised method

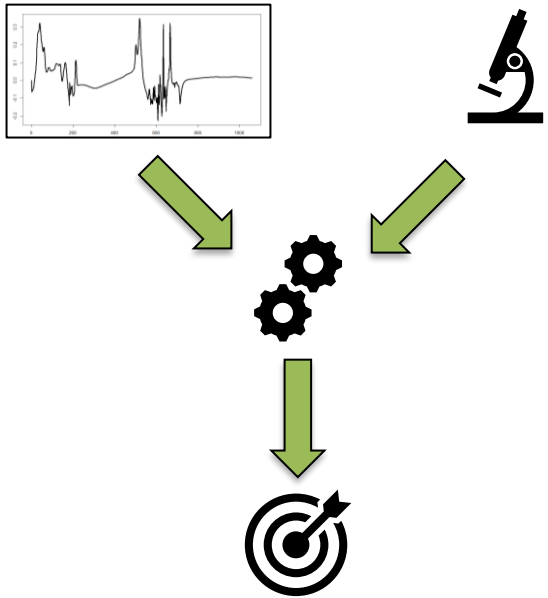


Unsupervised method

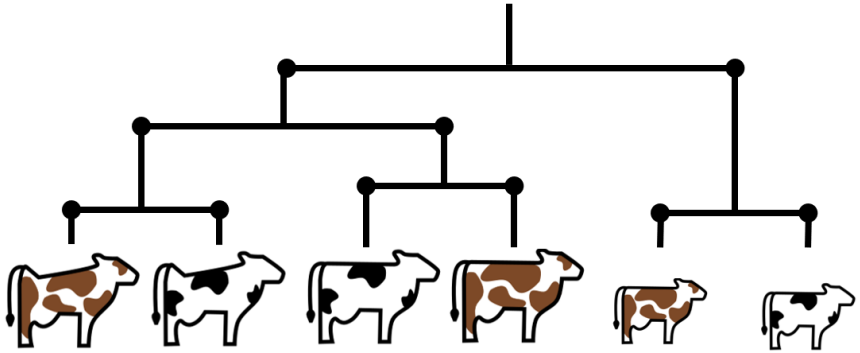


What can we do with our data ?

Supervised method

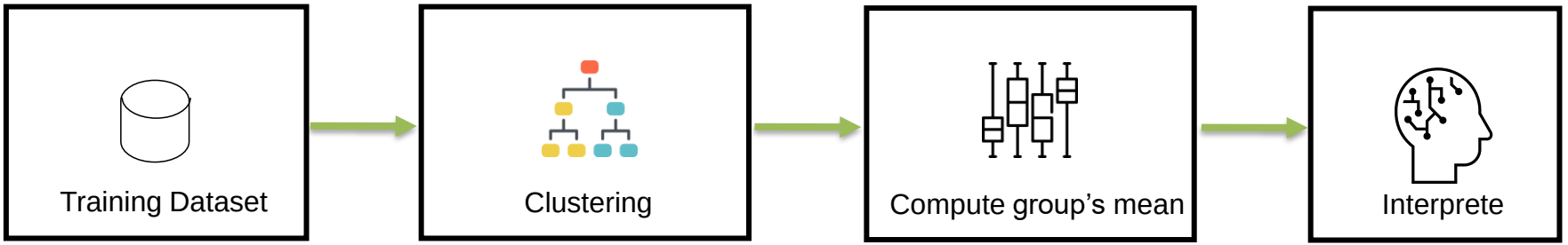


Unsupervised method



Hierarchical Clustering

A simple methodology



Dataset

1,712,417 DHI records

- **From 139,203 Holstein cows in 1082 herds**
- **With DIM between 5 and 365 days**
- **Between January 2012 and March 2020**
- **From first to third lactations**

Milk Component	Minerals	Fatty acids	Metabolism	Feeding	Other
•Milk Yield	•Natrium	•Saturate	•β-Hydroxybutyrate	•Weight	•Somatic Cells Count
•Fat	•Calcium	•C18:1cis9	•Protein Efficiency	•Dry Matter Intake	•Methane
•Protein	•Magnesium	•Monounsaturated	•Energy Balance	•Dry Matter Intake 2	•Lactoferrin
•Lactose	•Phosphorus	•Polyunsaturated	•Acetone	•Consumption Index	
•FPCM	•Potassium		•Citrate	•Residual Feed Intake	
			•Urea	•Residual Feed Intake 2	

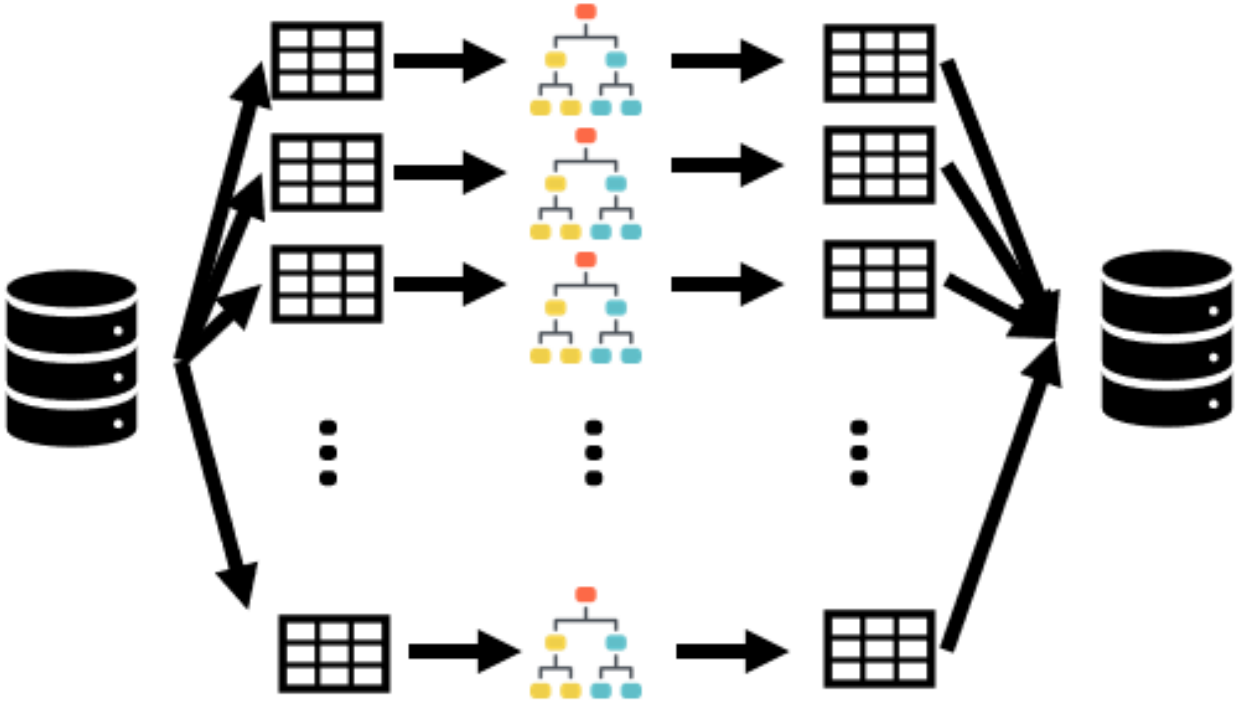
Dataset

1,712,417 DHI records => Requires adaptation for computation

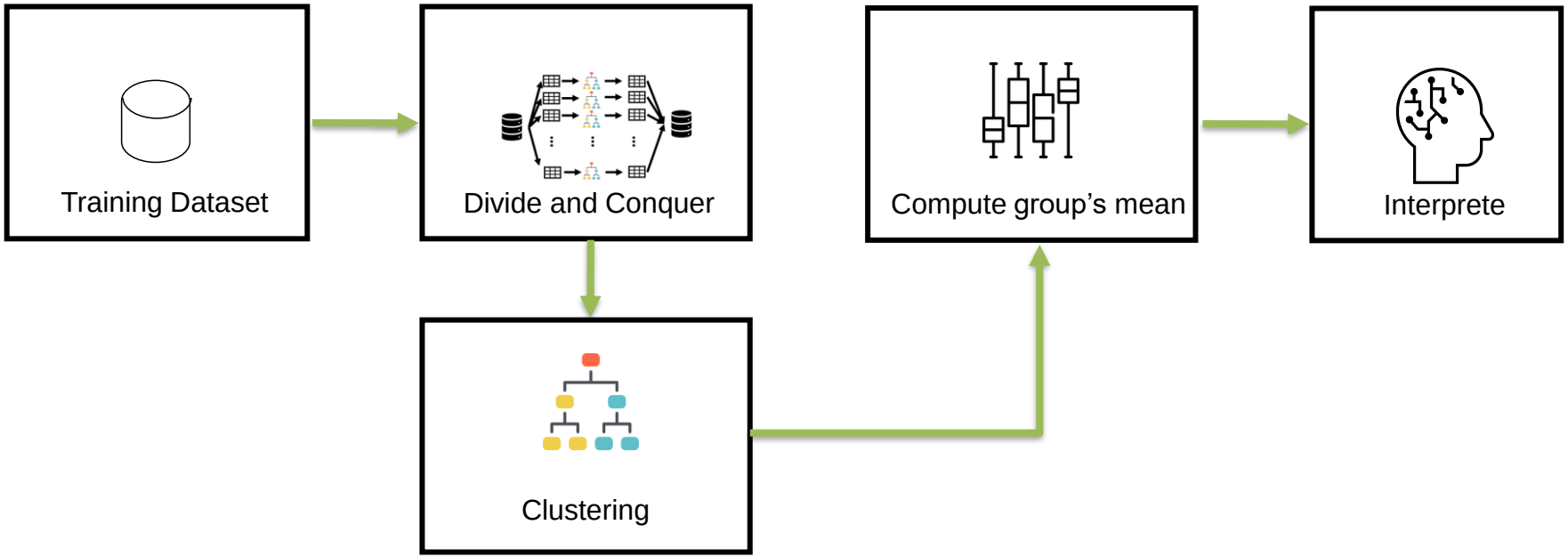
- From 139,203 Holstein cows in 1082 herds
- With DIM between 5 and 365 days
- Between January 2012 and March 2020
- From first to third lactations

Milk Component	Minerals	Fatty acids	Metabolism	Feeding	Other
•Milk Yield	•Natrium	•Saturate	•β-Hydroxybutyrate	•Weight	•Somatic Cells Count
•Fat	•Calcium	•C18:1cis9	•Protein Efficiency	•Dry Matter Intake	•Methane
•Protein	•Magnesium	•Monounsaturated	•Energy Balance	•Dry Matter Intake 2	•Lactoferrin
•Lactose	•Phosphorus	•Polyunsaturated	•Acetone	•Consumption Index	
•FPCM	•Potassium		•Citrate	•Residual Feed Intake	
			•Urea	•Residual Feed Intake 2	

Divide and Conquer Clustering Approach



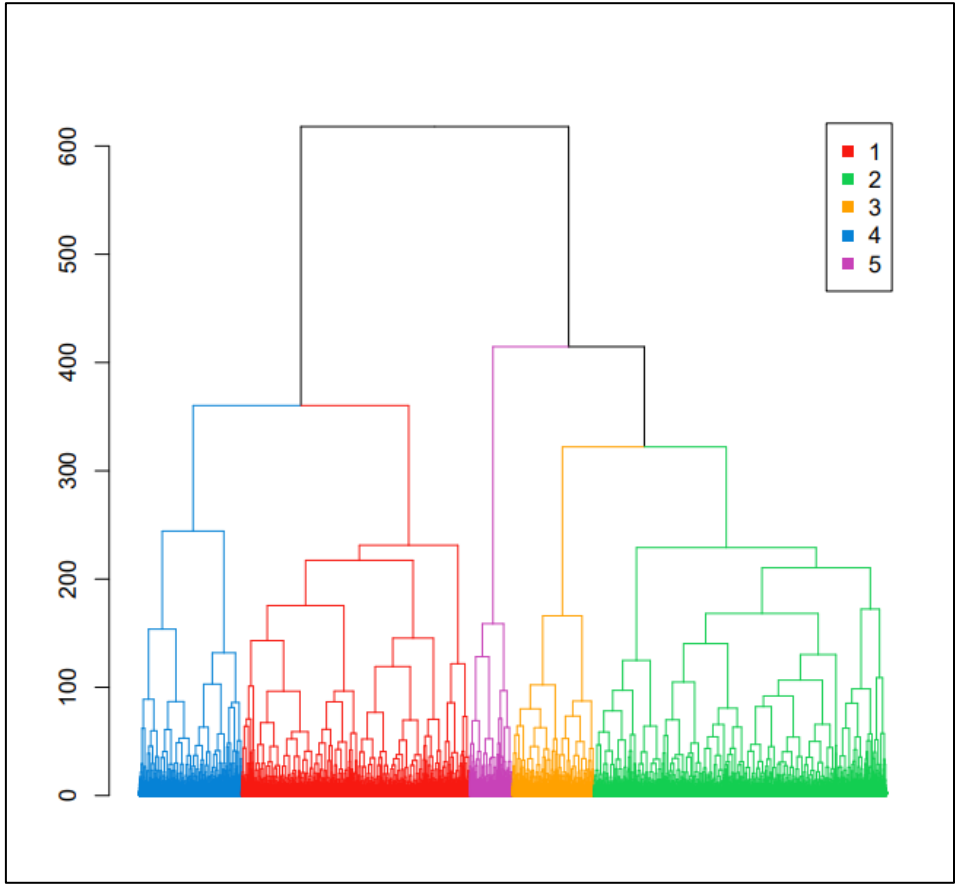
Adapting the methodology I



J. Dairy Sci. 105:6760–6772
https://doi.org/10.3168/jds.2022-21975

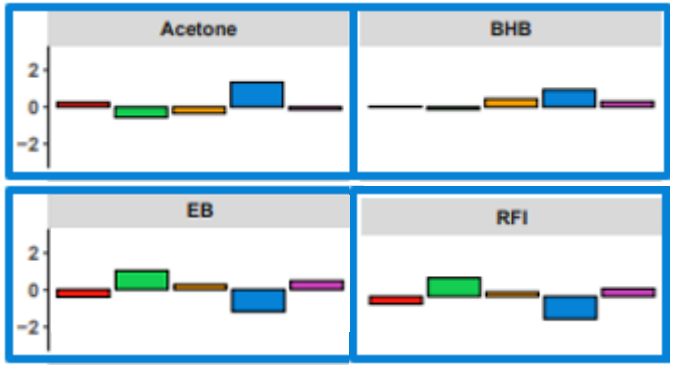
Can unsupervised learning methods applied to milk recording big data provide new insights into dairy cow health?

S. Franceschini,[1]* C. Grelet,[2] J. Leblois,[3] N. Gengler,[1] GplusE consortium,† and H. Soyeurt[1]
[1]University of Liège, Gembloux Agro-Bio Tech (ULiège-GxABT), 5030 Gembloux, Belgium
[2]Walloon Agricultural Research Center (CRA-W), 5030 Gembloux, Belgium
[3]Walloon Breeders Association Group (Elevéo by Awé groupe), 5590 Ciney, Belgium



Best cut is 5 five groups:

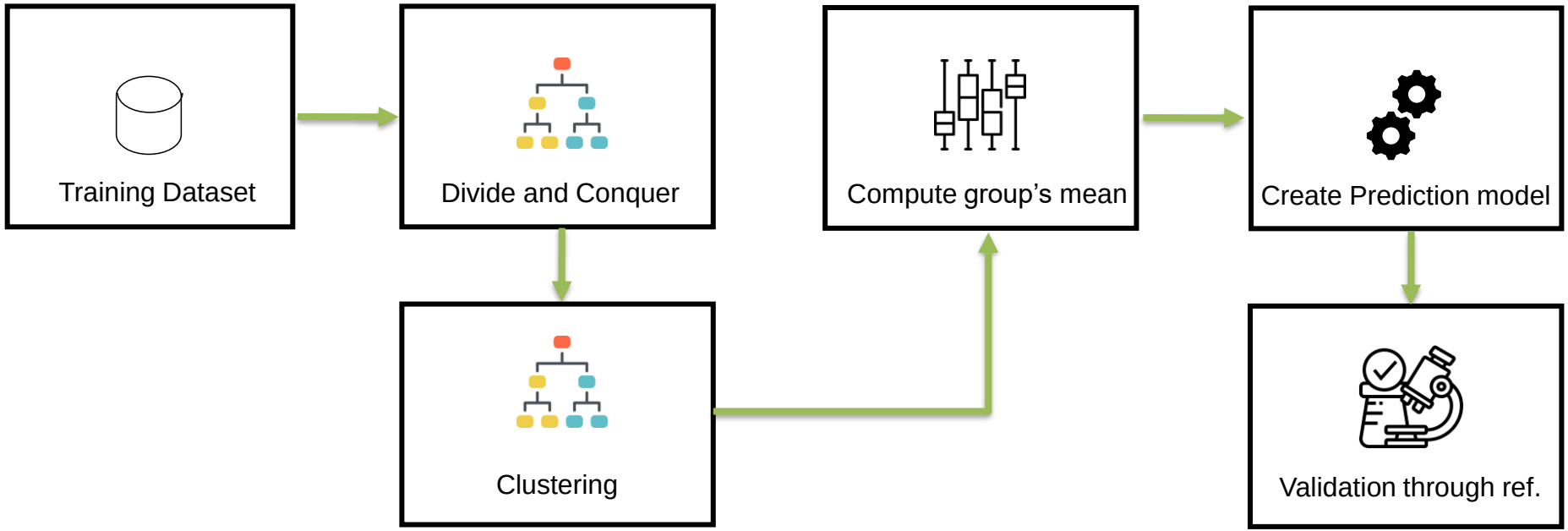
- **Blue is related to Negative Energy Balance**



- **Also validated through reference values**

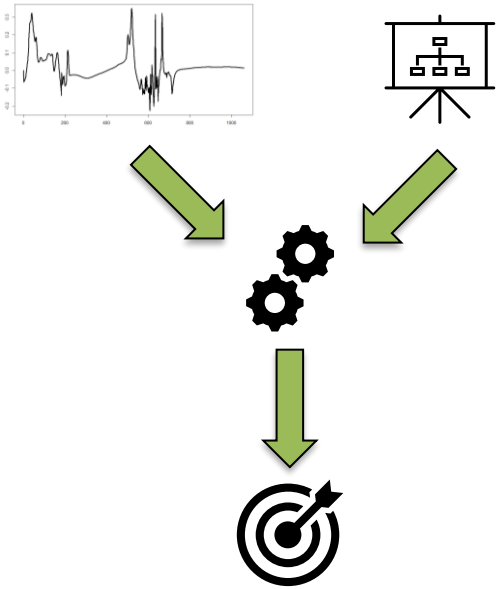
(Franceschini et al., 2022)

Adapting the methodology II



Prediction of the NEB group

Supervised method

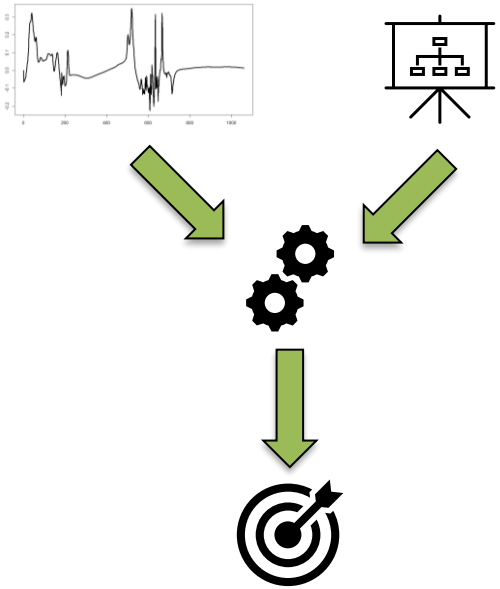


From the MIR spectra, we use

- Partial Least Square Discriminant Analysis (PLSDA)
- On 1,712,417 spectra
- With 10 latent variables

Prediction of the NEB group

Supervised method



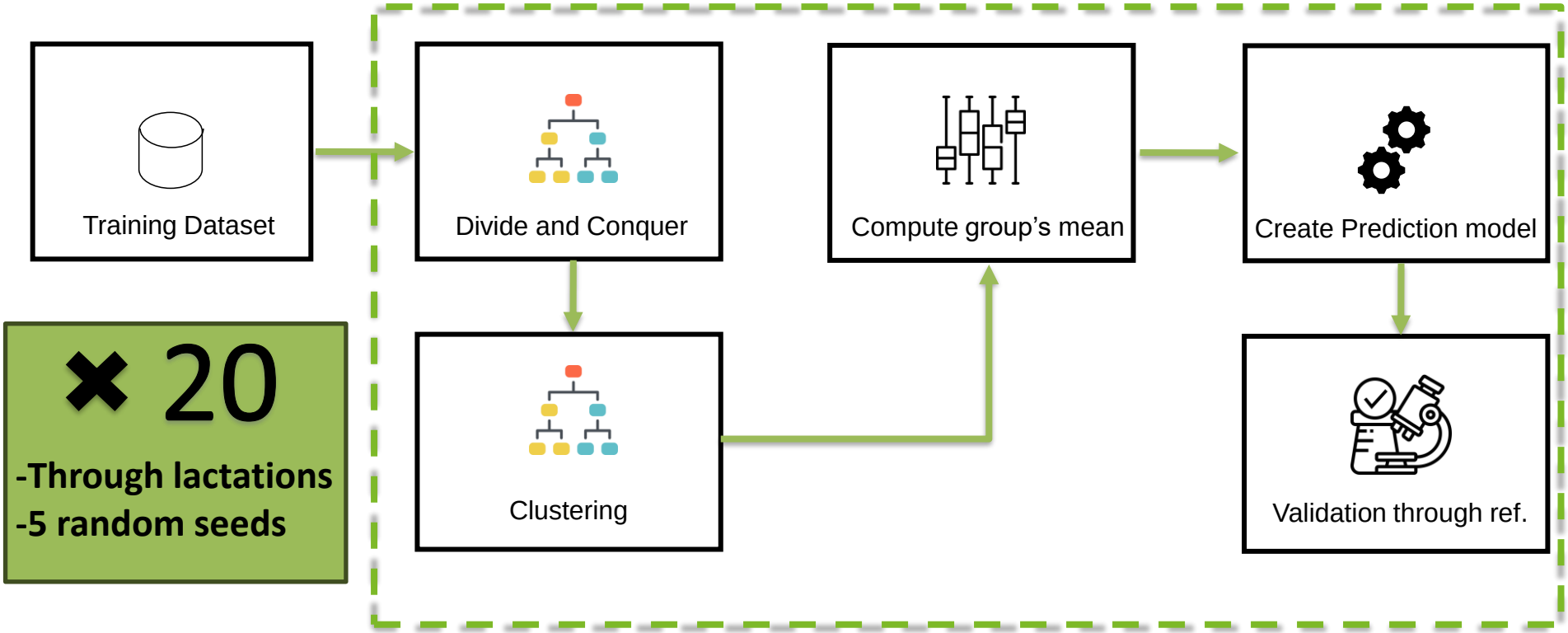
From the MIR spectra, we use

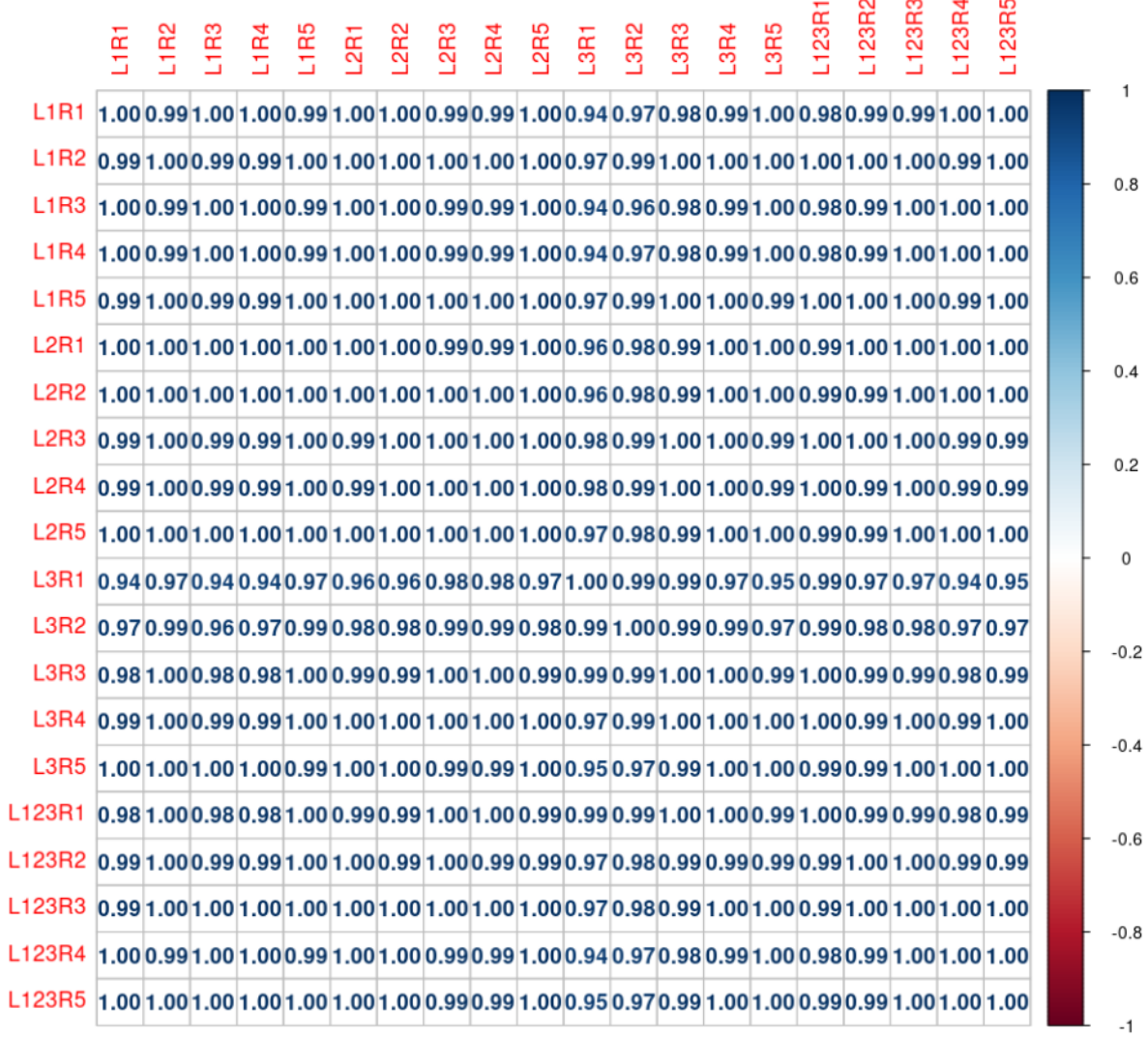
- Partial Least Square Discriminant Analysis (PLSDA)
- On 1,712,417 spectra
- With 10 latent variables

The performances in validation were:

- Accuracy of 0.99
- Sensitivity of 0.95
- Specificity of 0.92

Inproving the methodology III

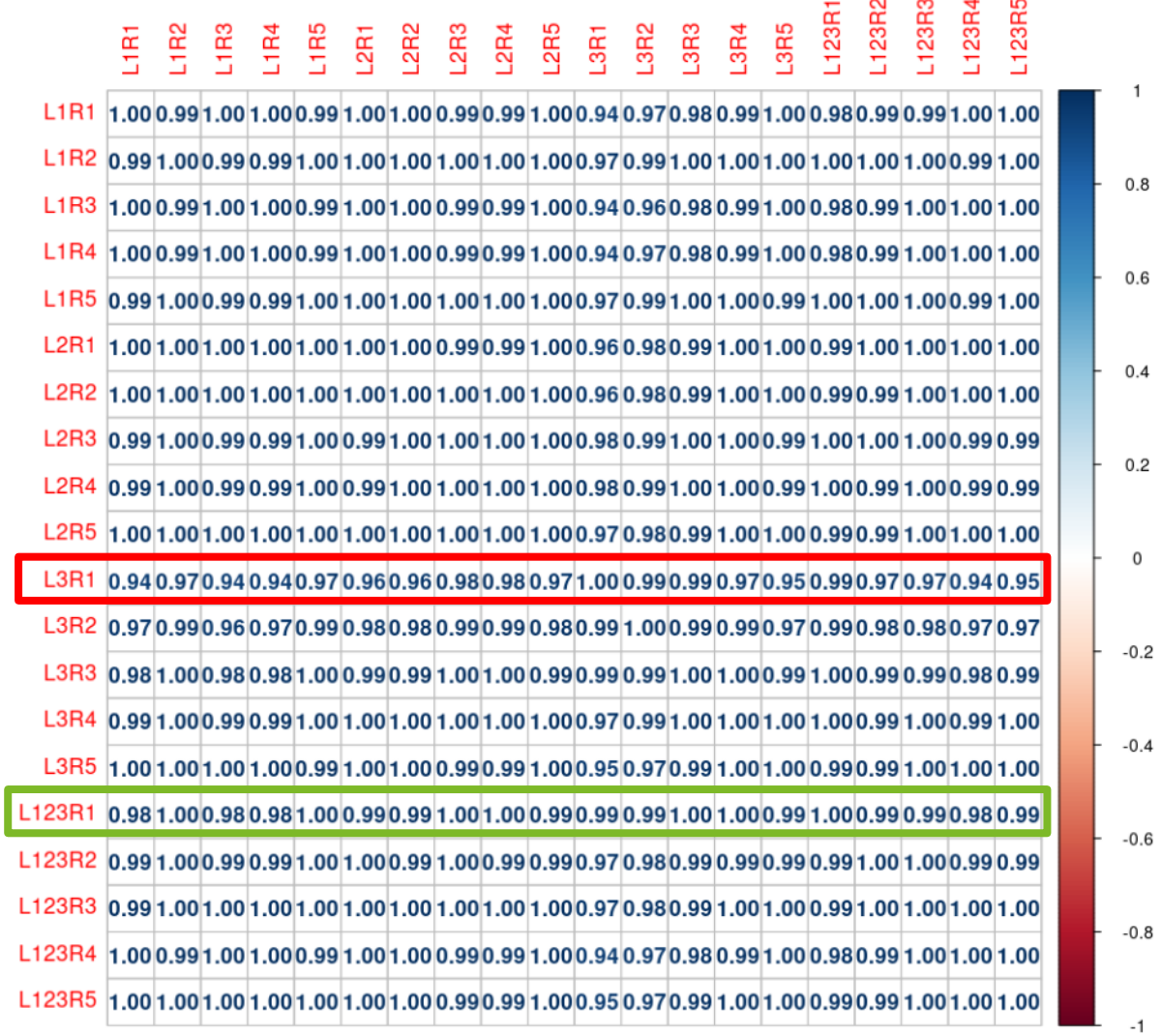




Correlations Matrix

-Between 20 prediction models

-On external spectral data



Correlations Matrix

-Between 20 prediction models

-On external spectral data

The correlations between the predictions are between 0.98 and 1.

Take Home message

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

Can we combine different observations ?
Is it stable ?

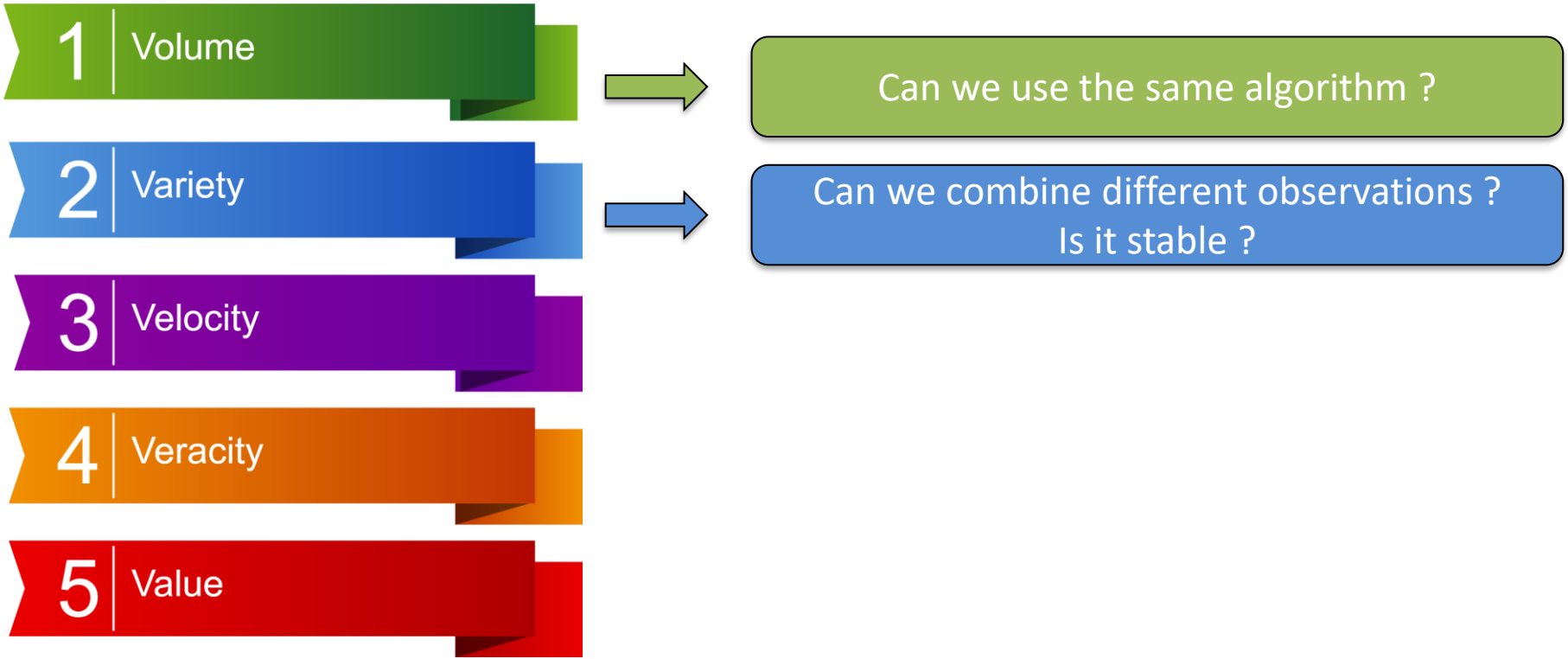
Take Home message

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

→ Can we combine different observations ?
Is it stable ?

YES

Take Home message



Take Home message

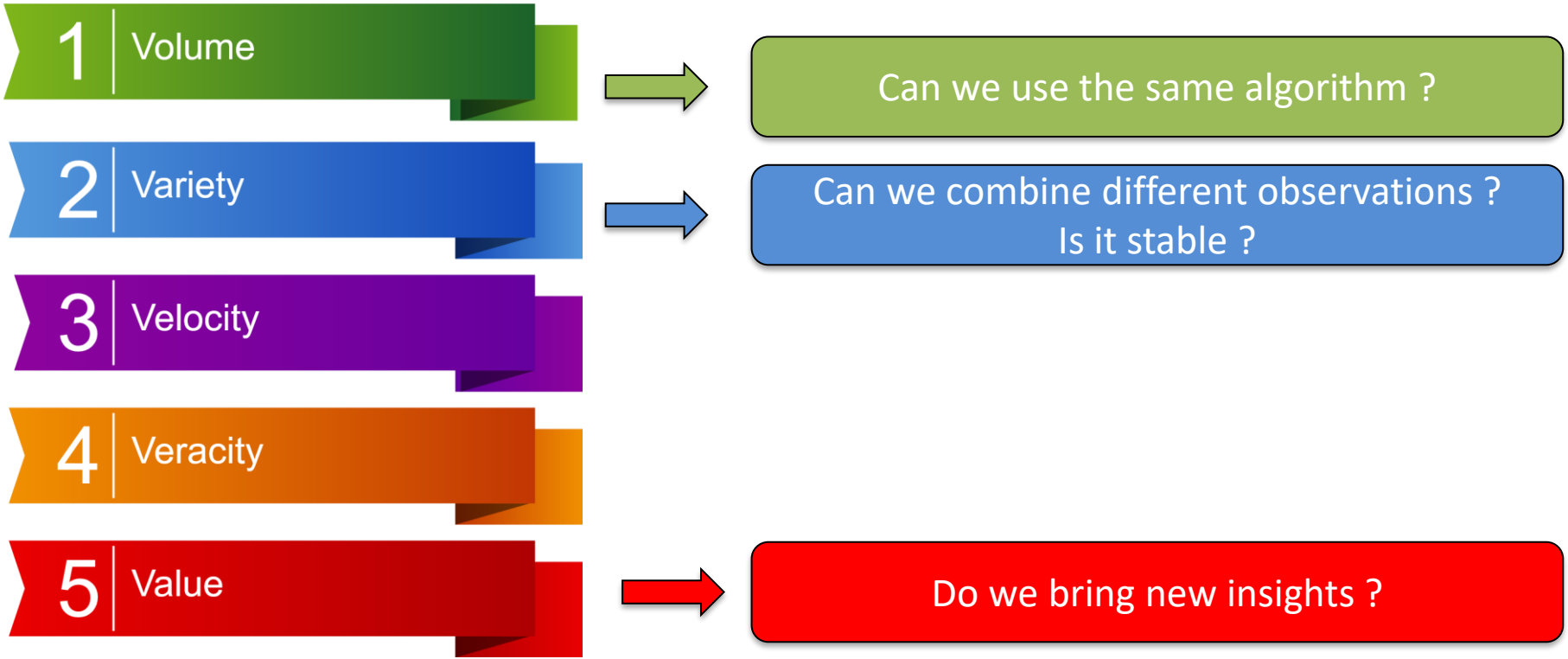
1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

→ Can we use the same algorithm ?

→ Can we combine different observations ?
Is it stable ?

Sometimes

Take Home message



Take Home message

1 Volume → Can we use the same algorithm ?

2 Variety → Can we combine different observations ?
Is it stable ?

Work in progress!

5 Value → Do we bring new insights ?

LIÈGE université
Gembloux
Agro-Bio Tech

Thank you for attention

Sfranceschini@uliege.be

Let’s keep it touch !

LinkedIn

5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

How 

vations, variables ?

5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

How many sources, structures, distribution ?

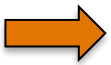
5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

What are the frequence of data acquisition ?

5 V's of big data

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value



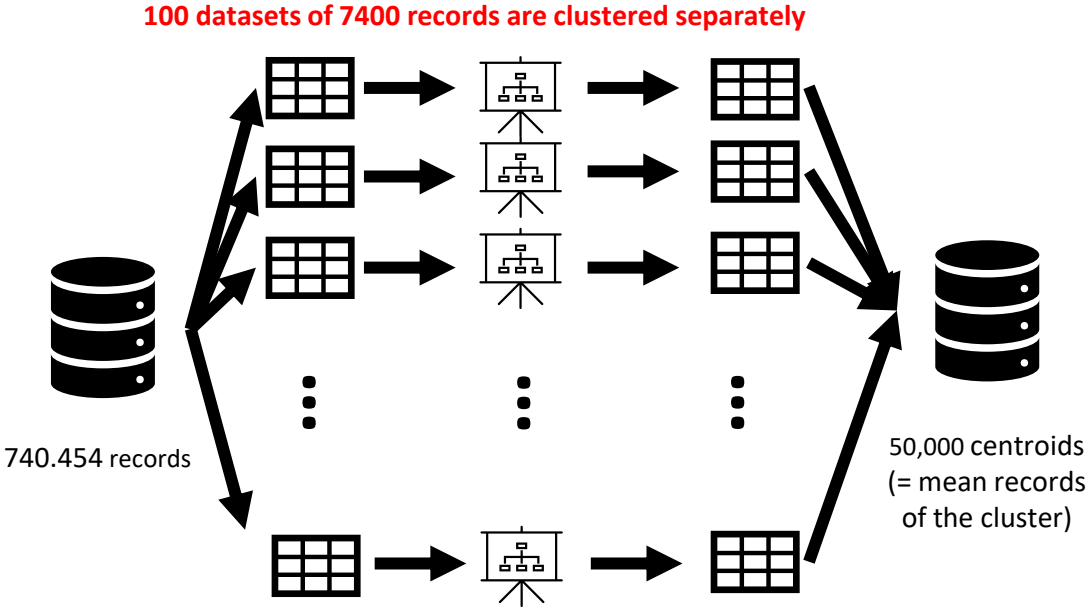
What is the data quality ?

5 V's of big data

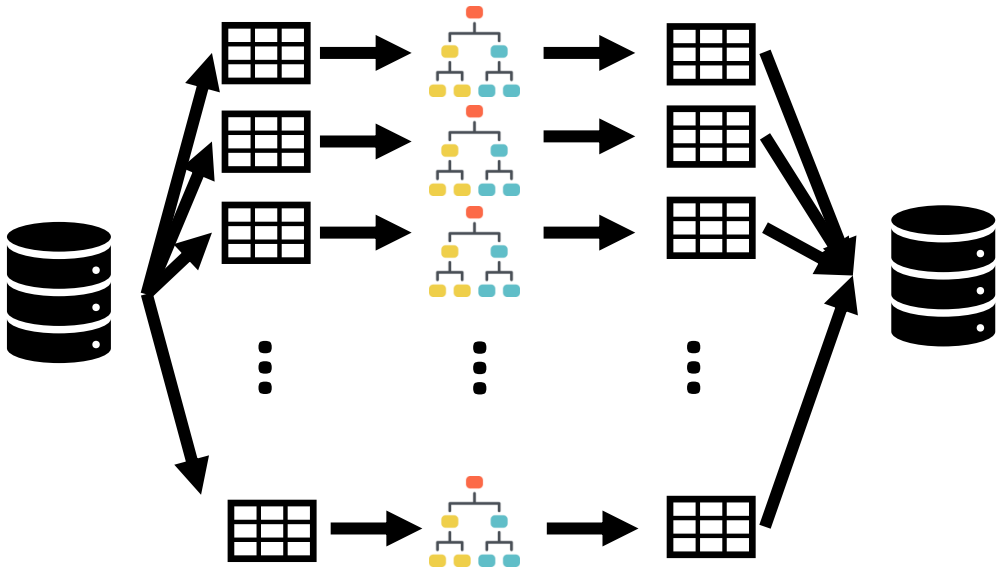
1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

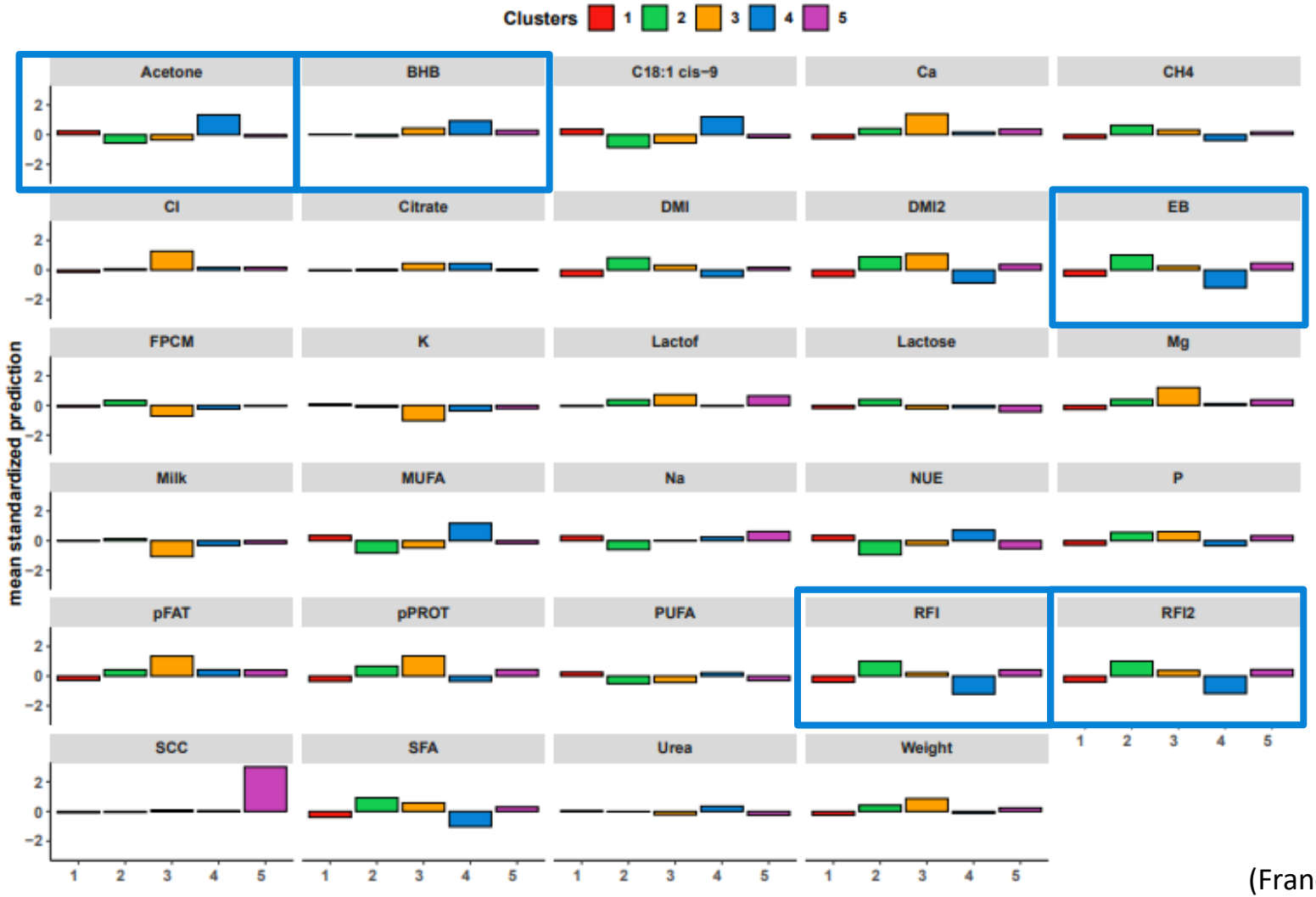
What can we do with this data ?

Divide and Conquer Clustering Approach

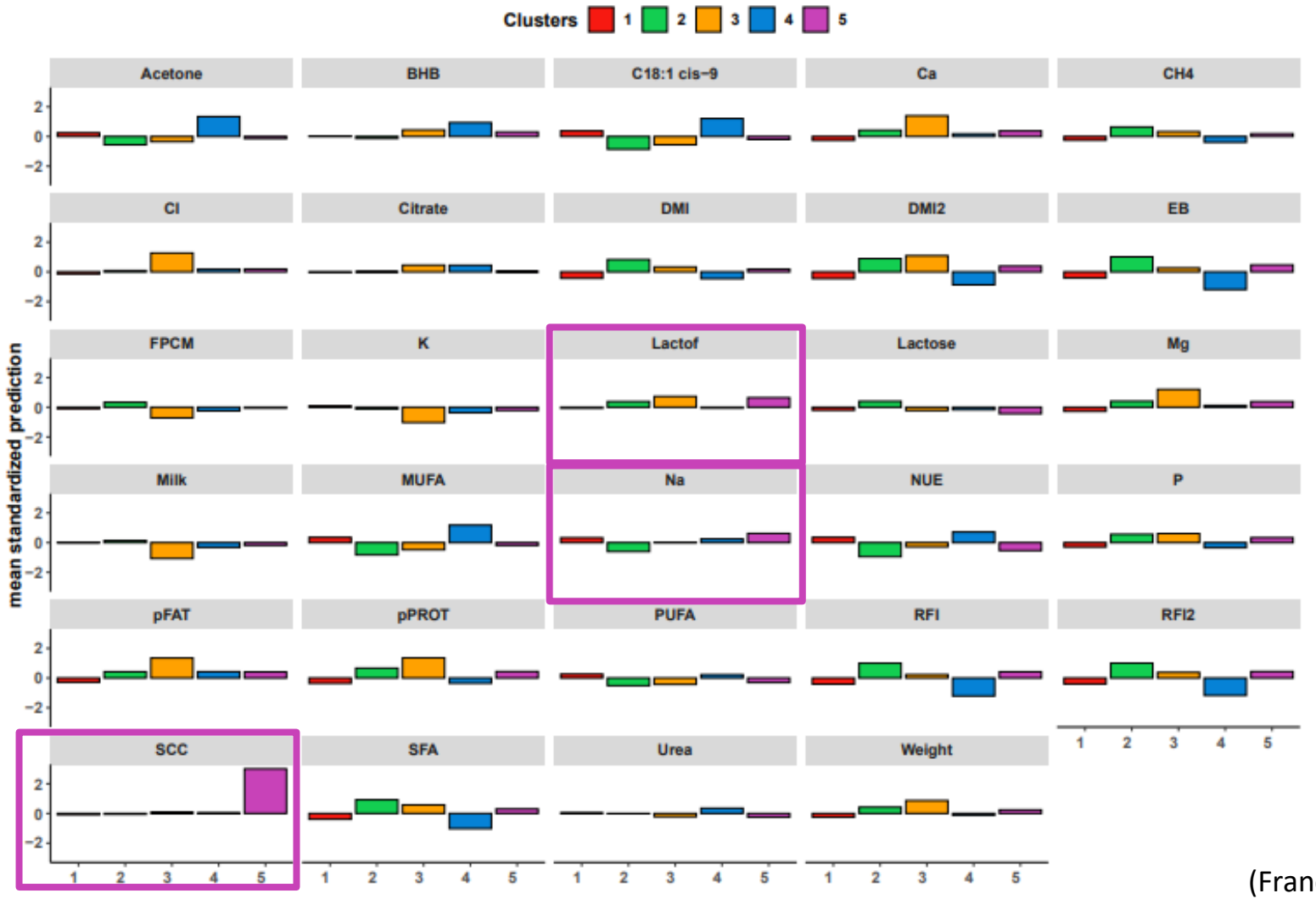


Divide and Conquer Clustering Approach





(Franceschini et al., 2022)

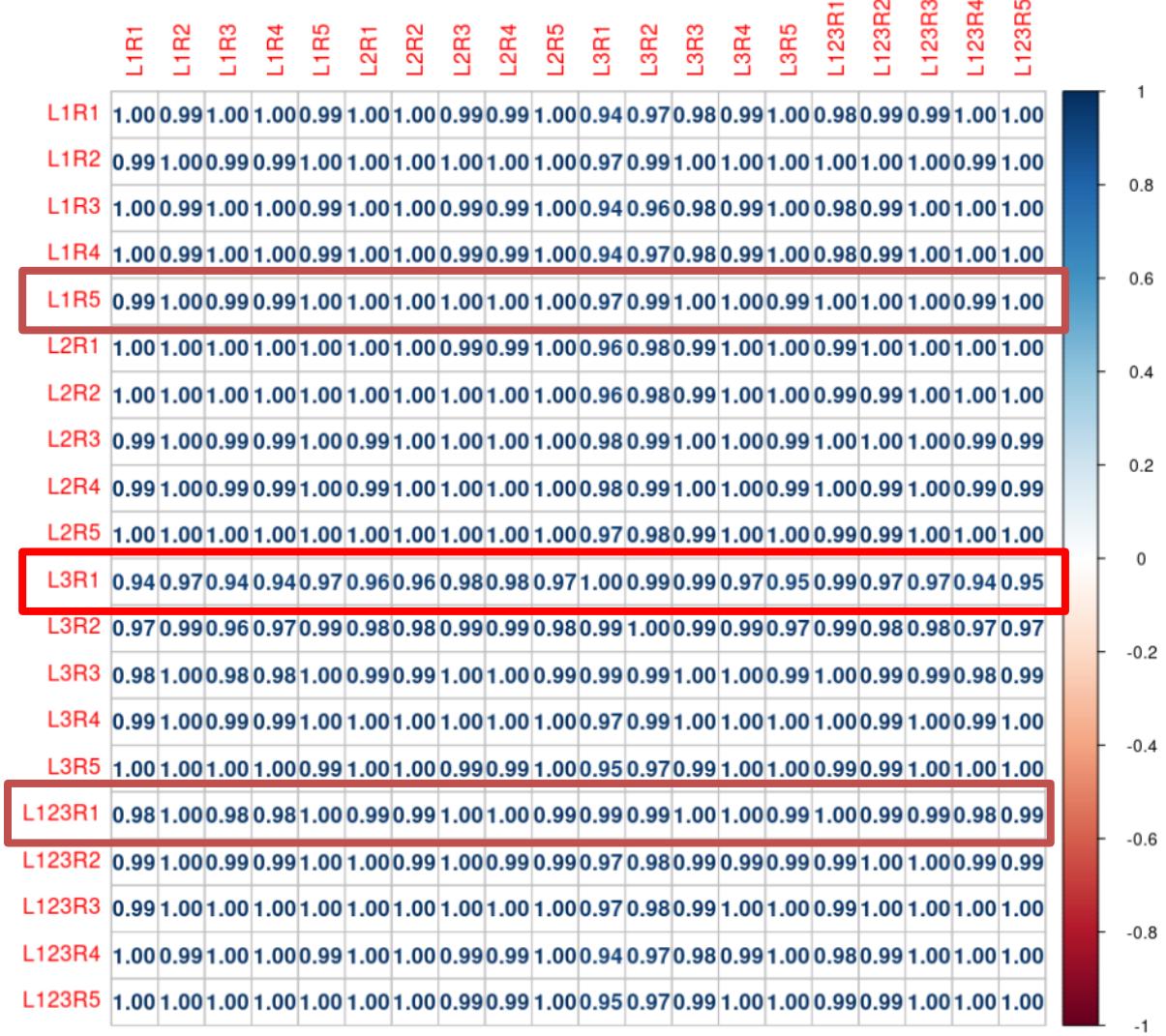


(Franceschini et al., 2022)

Correlation with laboratory measures

	Lactation 1 (N = 50)		All Lactation (N = 204)	
Reference Analysis	Cluster 4	Cluster 5	Cluster 4	Cluster 5
Blood cholesterol	-0.09	0.02	-0.06	-0.07
Blood Fructosamine	-0.15	0.19	-0.19**	0.03
Blood BHB	0.70***	-0.4**	0.54***	-0.06
Blood NEFA	0.67***	-0.36**	0.44***	-0.01
Blood P4	-0.09	-0.11	0.05	-0.06
Blood Glucose	-0.60***	0.42**	-0.32***	-0.06
Milk Glu6P	-0.32*	0.29*	-0.23***	0.16*
Milk GluFree	0.16	-0.26	-0.15*	-0.18**
Milk BHB	0.53***	-0.21	0.35***	0.04
Milk Isocitrate	0.5***	-0.16	0.38***	-0.1
Milk Urea	0.32*	-0.15	0.03	-0.03
Milk NAGase	0.1	0.02	0.06	0.46***
Milk LDH	-0.07	0.43**	-0.13	0.69***

Stability


L1R1 L1R2 L1R3 L1R4 L1R5 L2R1 L2R2 L2R3 L2R4 L2R5 L3R1 L3R2 L3R3 L3R4 L3R5 L123R1 L123R2 L123R3 L123R4 L123R5
L1R1 1.00 0.99 1.00 1.00 0.99 1.00 1.00 0.99 0.99 1.00 0.94 0.97 0.98 0.99 1.00 0.98 0.99 0.99 1.00 1.00
L1R2 0.99 1.00 0.99 0.99 1.00 1.00 1.00 1.00 1.00 1.00 0.97 0.99 1.00 1.00 1.00 1.00 1.00 1.00 0.99 1.00
L1R3 1.00 0.99 1.00 1.00 0.99 1.00 1.00 0.99 0.99 1.00 0.94 0.96 0.98 0.99 1.00 0.98 0.99 1.00 1.00 1.00
L1R4 1.00 0.99 1.00 1.00 0.99 1.00 1.00 0.99 0.99 1.00 0.94 0.97 0.98 0.99 1.00 0.98 0.99 1.00 1.00 1.00
L1R5 0.99 1.00 0.99 0.99 1.00 1.00 1.00 1.00 1.00 1.00 0.97 0.99 1.00 1.00 0.99 1.00 1.00 1.00 0.99 1.00
L2R1 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 0.96 0.98 0.99 1.00 1.00 0.99 1.00 1.00 1.00 1.00
L2R2 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.96 0.98 0.99 1.00 1.00 0.99 0.99 1.00 1.00 1.00
L2R3 0.99 1.00 0.99 0.99 1.00 0.99 1.00 1.00 1.00 1.00 0.98 0.99 1.00 1.00 0.99 1.00 1.00 1.00 0.99 0.99
L2R4 0.99 1.00 0.99 0.99 1.00 0.99 1.00 1.00 1.00 1.00 0.98 0.99 1.00 1.00 0.99 1.00 0.99 1.00 0.99 0.99
L2R5 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.97 0.98 0.99 1.00 1.00 0.99 0.99 1.00 1.00 1.00
L3R1 0.94 0.97 0.94 0.94 0.97 0.96 0.96 0.98 0.98 0.97 1.00 0.99 0.99 0.97 0.95 0.99 0.97 0.97 0.94 0.95
L3R2 0.97 0.99 0.96 0.97 0.99 0.98 0.98 0.99 0.99 0.98 0.99 1.00 0.99 0.99 0.97 0.99 0.98 0.98 0.97 0.97
L3R3 0.98 1.00 0.98 0.98 1.00 0.99 0.99 1.00 1.00 0.99 0.99 0.99 1.00 1.00 0.99 1.00 0.99 0.99 0.98 0.99
L3R4 0.99 1.00 0.99 0.99 1.00 1.00 1.00 1.00 1.00 1.00 0.97 0.99 1.00 1.00 1.00 1.00 0.99 1.00 0.99 1.00
L3R5 1.00 1.00 1.00 1.00 0.99 1.00 1.00 0.99 0.99 1.00 0.95 0.97 0.99 1.00 1.00 0.99 0.99 1.00 1.00 1.00
L123R1 0.98 1.00 0.98 0.98 1.00 0.99 0.99 1.00 1.00 0.99 0.99 0.99 1.00 1.00 0.99 1.00 0.99 0.99 0.98 0.99
L123R2 0.99 1.00 0.99 0.99 1.00 1.00 0.99 1.00 0.99 0.99 0.97 0.98 0.99 0.99 0.99 0.99 1.00 1.00 0.99 0.99
L123R3 0.99 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.97 0.98 0.99 1.00 1.00 0.99 1.00 1.00 1.00 1.00
L123R4 1.00 0.99 1.00 1.00 0.99 1.00 1.00 0.99 0.99 1.00 0.94 0.97 0.98 0.99 1.00 0.98 0.99 1.00 1.00 1.00
L123R5 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 1.00 0.95 0.97 0.99 1.00 1.00 0.99 0.99 1.00 1.00 1.00
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

LIÈGE université
Gembloux Agro-Bio Tech

EU-L-PHE
EUROPEAN NETWORK ON LIVESTOCK PHENOMICS

Combining high throughput phenotypes to study complex traits : a case-study of negative energy balance using milk mid-infrared based predictions

Sébastien Franceschini

Supervised by N. Gengler and H. Soyeurt

Funded by the European Union