

Constitution of an international dataset on blood biomarkers in dairy cows: A preliminary study to develop milk MIR models

C. Grelet¹, L. Millot², O. Christophe¹, V. Wolf², J. Leblois³, M. A. Crowe⁴, D.C. Wathes⁵, C.P. Ferris⁶, K.L. Ingvarlsen⁷, C. Marchitelli⁸, F. Becker⁹, B. Bapst¹⁰, A. Köck¹¹, J. Hummel¹², M. Gelé¹³, M.H. Lidauer¹⁴, D.E. Santschi¹⁵, F. Huot¹⁶, H. Naya¹⁷, A. Meikle¹⁷, B. Martin¹⁸, M. Coppa¹⁸, N. Gengler¹⁹, GplusE consortium⁴, and F. Dehareng¹

CRA-W, BEL, 2. CEL25-90, FRA, 3. Awé groupe, BEL, 4. UCD, IRE, 5. RVC, UK, 6. AFBI, UK, 7. Aarhus U, DEN, 8. CREA, IT, 9. FBN DEU, 10. Qualitas, CHE, 11. ZuchtData, AUT, 12. U.Göttingen, DEU, 13. Idele, FRA, 14. LUKE FIN, 15. Lactanet, CAN, 16. U.Laval, CAN, 17. U. República, URU, 18. INRAE, FRA, ULg BEL



Blood reference

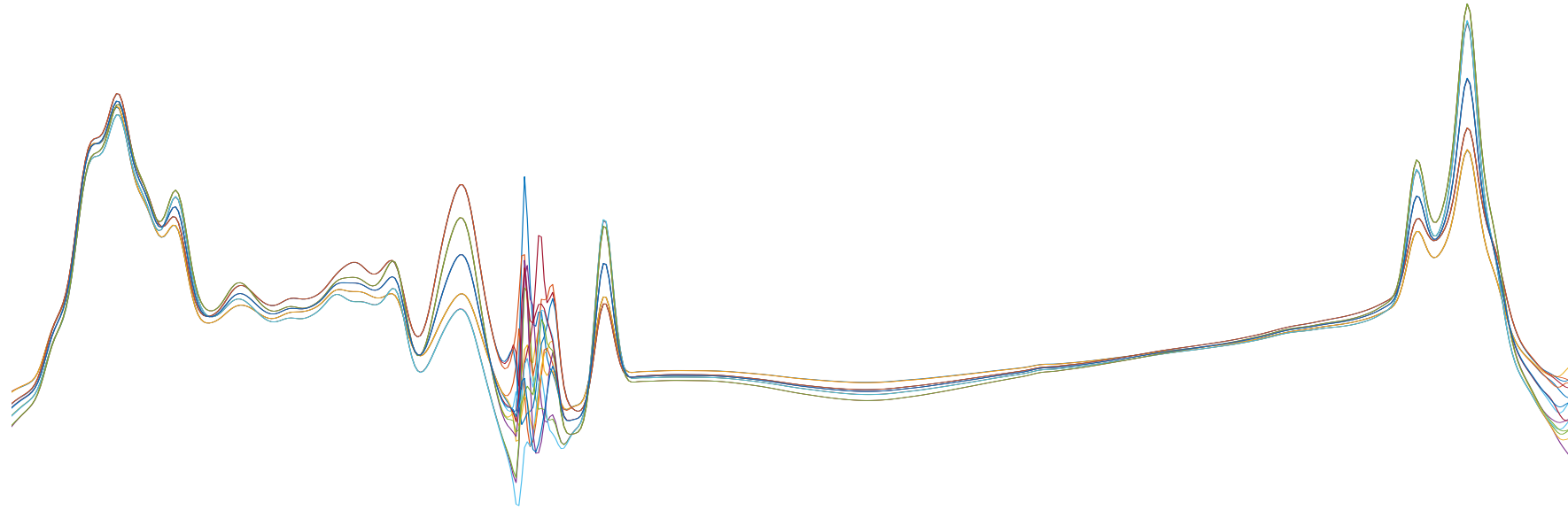
- Usual methodology for health diagnoses
- Invasive
- Expensive
- Not in routine sampling

Molecule	Use
BHB	Ketosis
NEFA	Body fat mobilization
Glucose	Glycemia
IGF-I	Liver status
Urea	Protein status
Cholesterol	Lipid metabolism
Fructosamin	Energy deficit, stress
Progesteron	Cyclicity



To predict by MIR?

- Cheap
- Accessible
- Large scale
- Management of cows
- Genetic studies



But to be usable in routine large-scale conditions

→ Need for Robustness: capacity to be “all terrain” and provide good results in various conditions

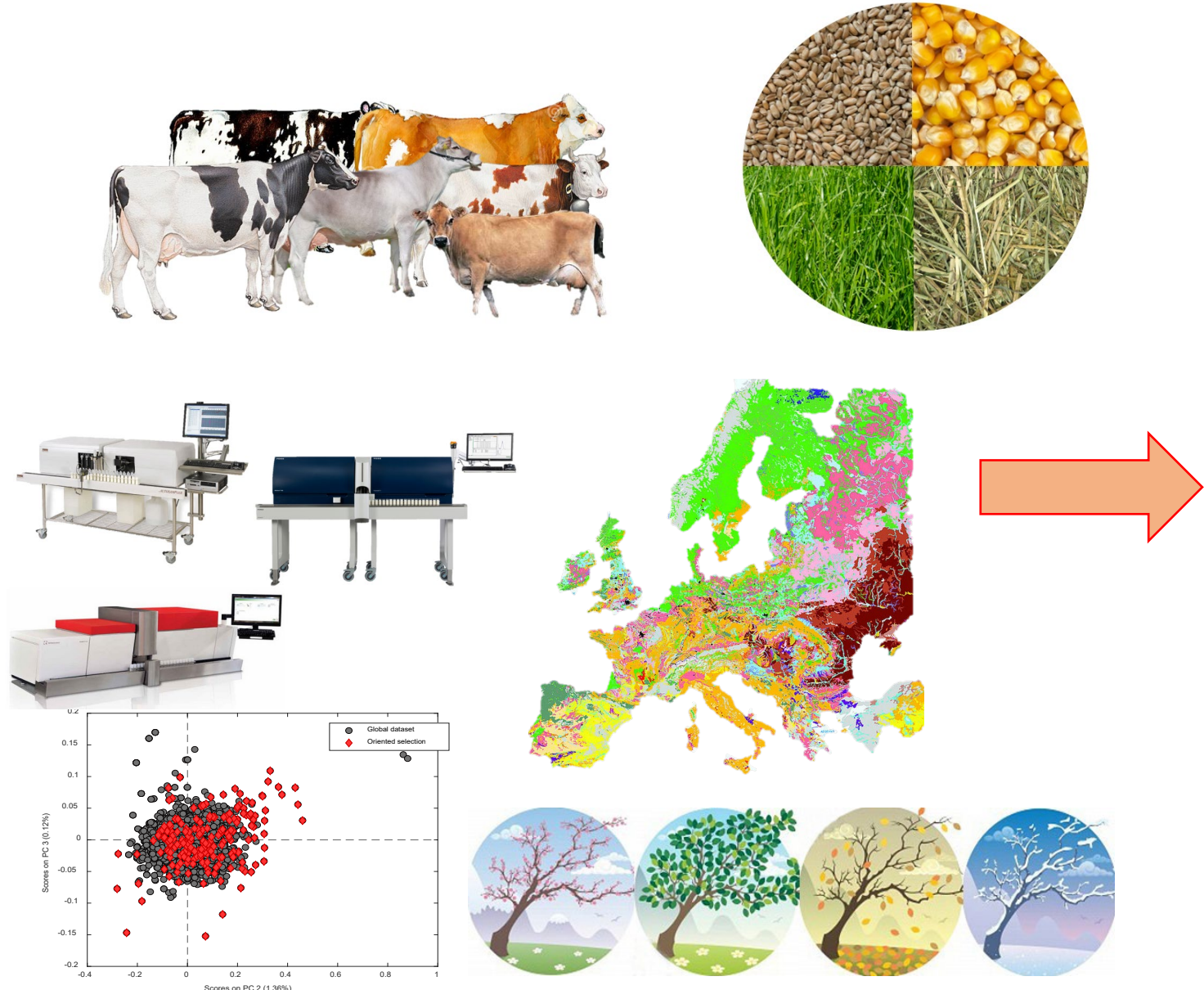
Research models



Robust models



IR models can only predict what they know



Robust models

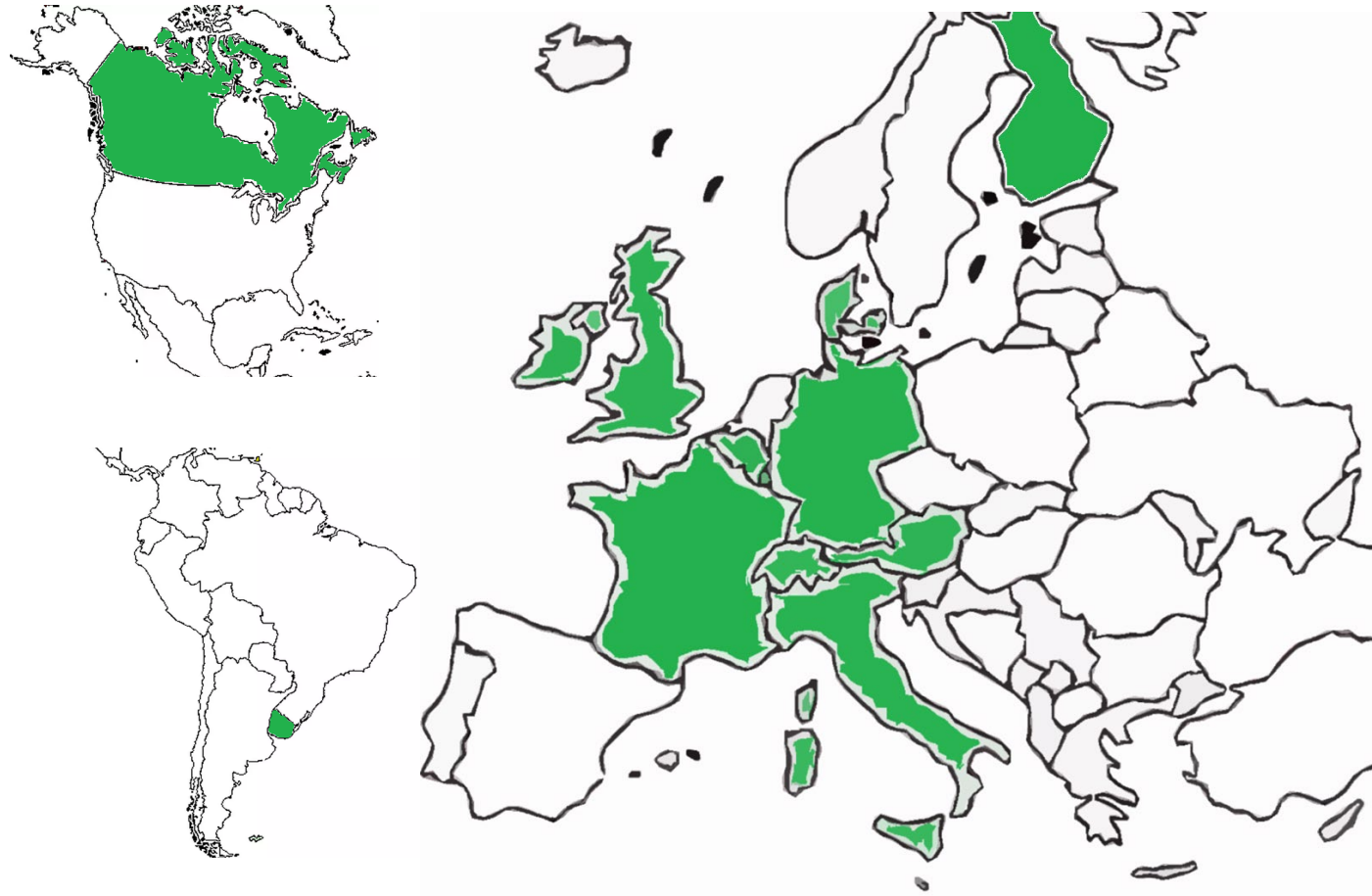
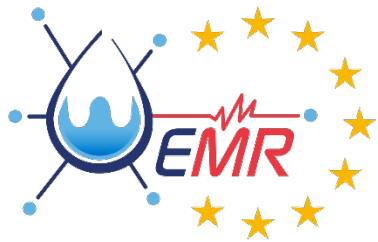


Creation of a consortium

Gathering of existing data (on a voluntary basis)

Simultaneous sampling of milk for MIR & blood for biomarker analysis

- Win/win
- 13 countries
- 14 projects
- 2013 to 2024
- Majority of MIR spectra standardized



Creation of a dataset

- Ongoing process
- 6059 records up to now:



- To be included:
- For a total final ≈ 10.000 records



+ other ?

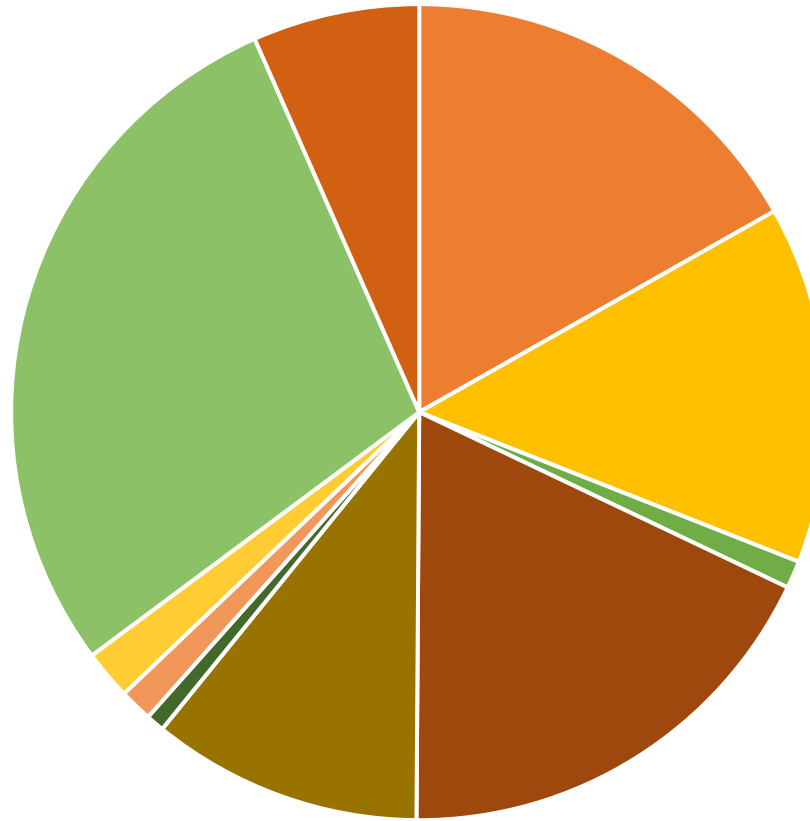
Maybe you? → c.grelet@cra.wallonie.be

Current dataset

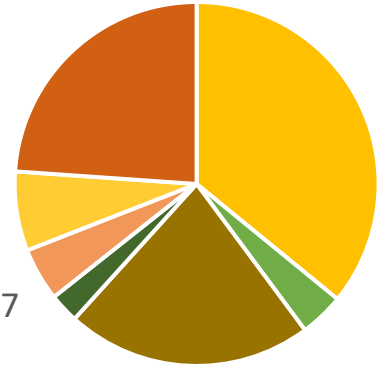
- 17 experiments
- 75 herds
- 2075 cows

■ Austria
■ Belgium
■ Denmark
■ France
■ Germany
■ Ireland
■ Italy
■ North_Ireland
■ Switzerland
■ UK

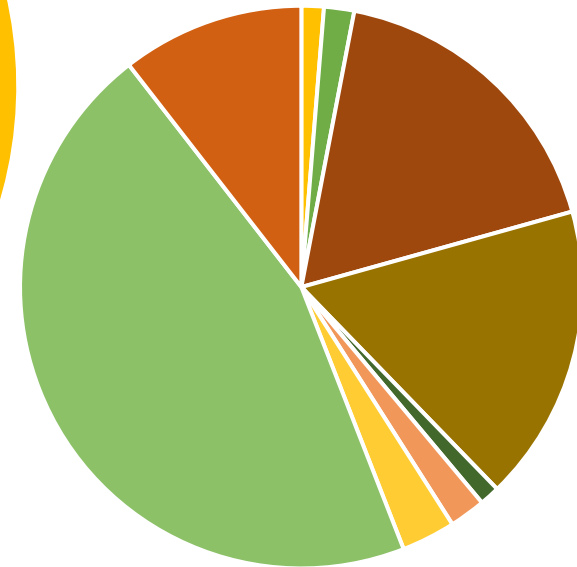
BHB (mmol/L) n = 6053



Glucose (mmol/L) n= 1674



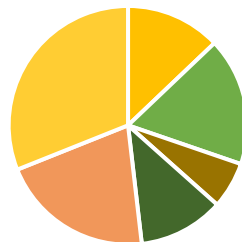
NEFA (mmol/L) n= 3807



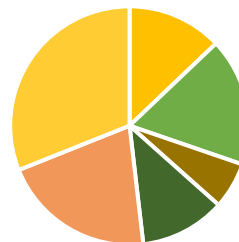
IGF-I (ng/ml) n= 534



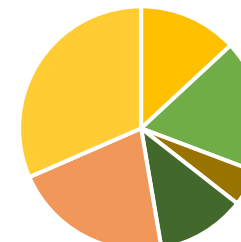
Urea (mmol/L) n= 380



Cholesterol (mmol/L) n= 380



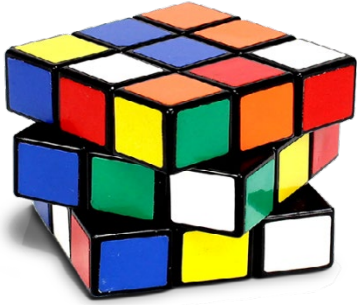
Fructosamine (μmol/L) n= 374



Progesterone (ng/ml) n= 374



Cleaning, treatment, modelling



Harmonization of datasets



6059 records

merging +/- 2 days between milk (MIR) and blood sampling → 5366

removing abnormal (zero...) values → 4907

fat difference between lab and spectra < 0.3 g/100ml → 4860 records

First derivative

212 wavenumbers (repeatable ones)

No outlier removal

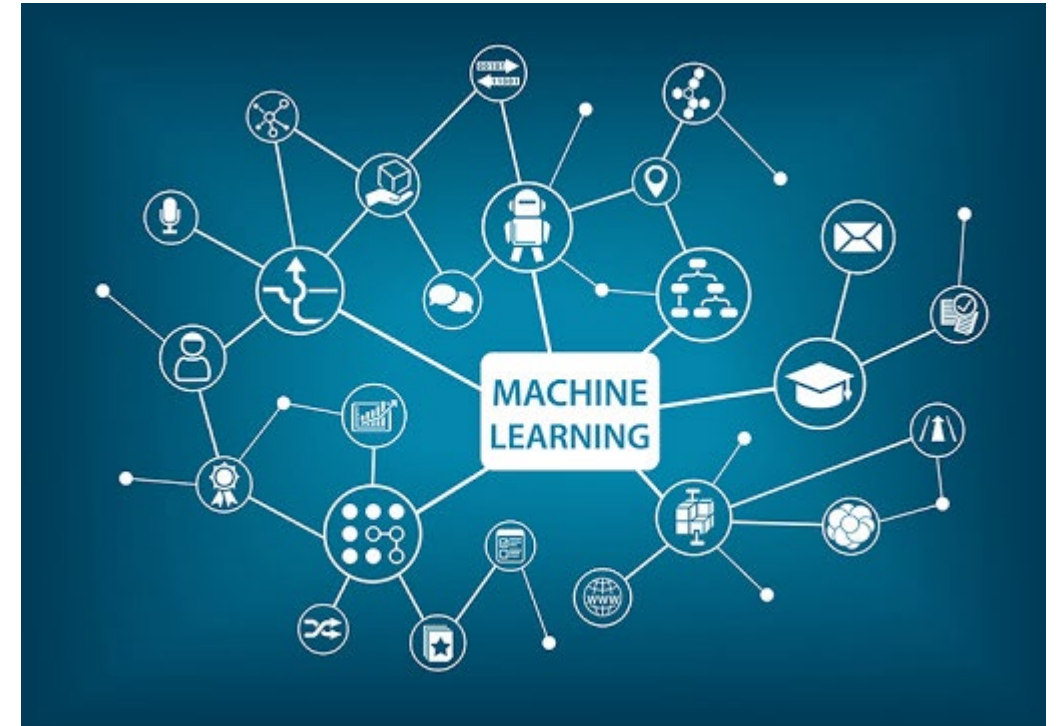
Cleaning, treatment, modelling

In a first time

- PLS
- PLS with Log
- SVM (support vector machine)

Other algorithms to be tested in future

- Kernel based
- Weighted local
- ...



External-Herd-Validation by excluding 1/4 of herds

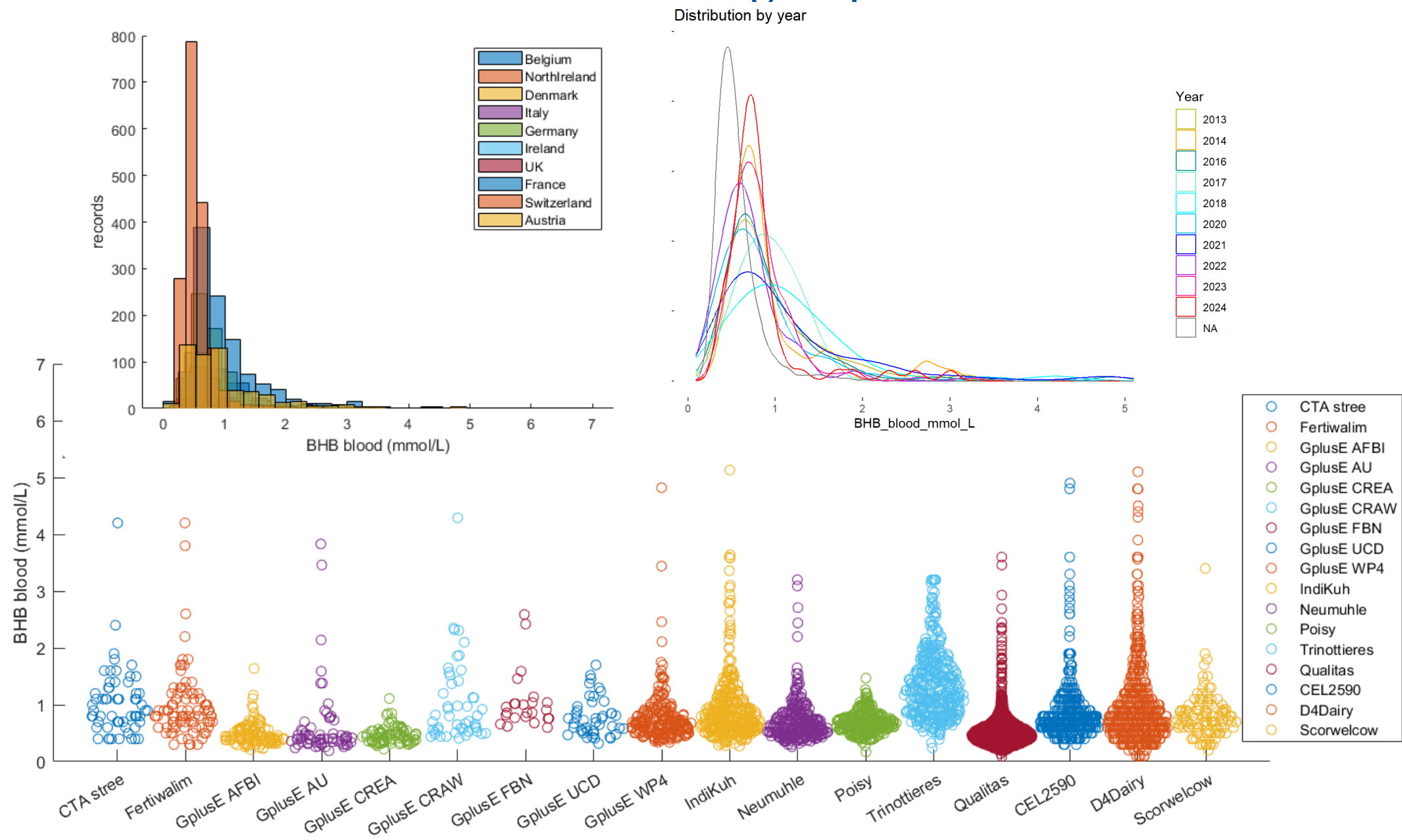
→ Repeated 4 times to turn all herds in the loop

→ Representative of real routine performances

I – Understanding the dataset



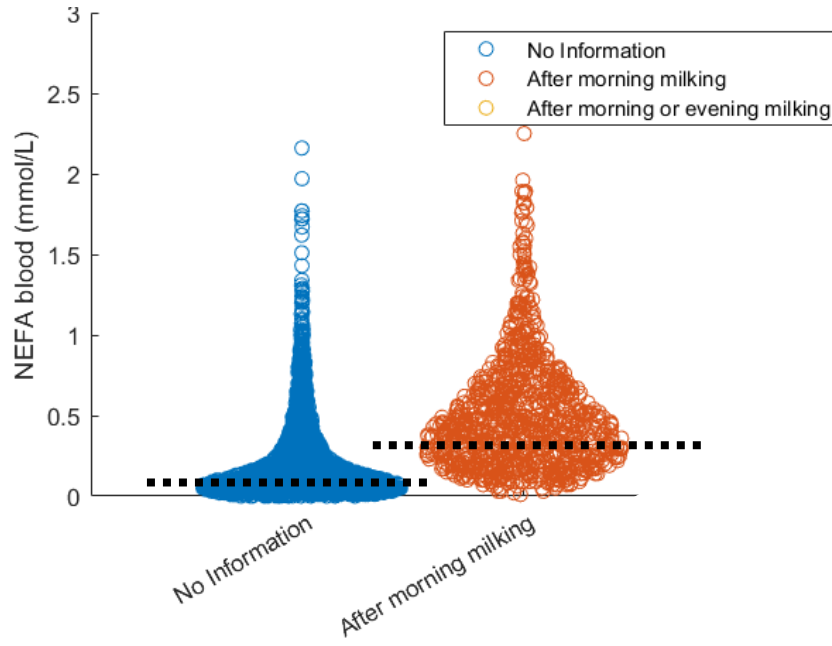
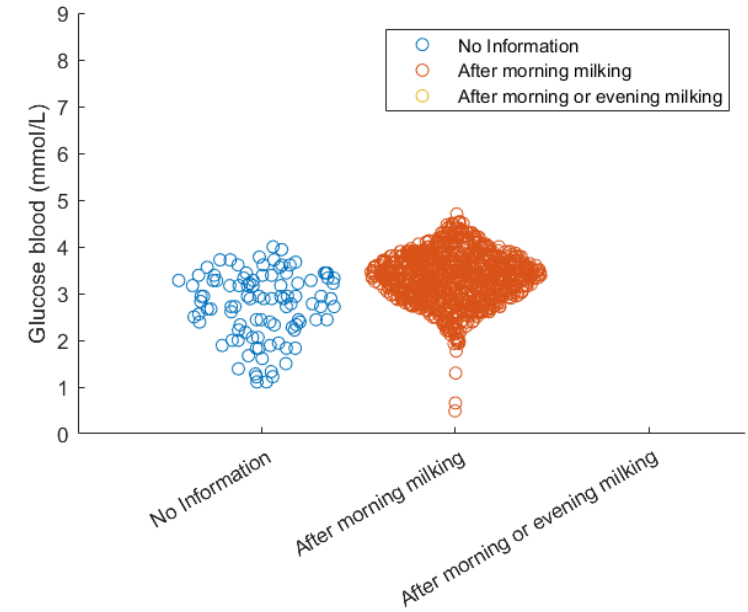
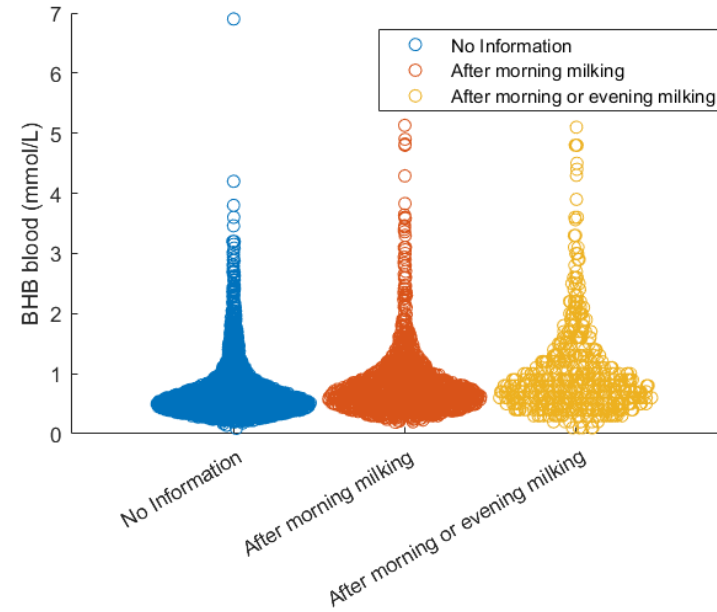
Are the data harmonized ? Can we group them?



Methodological aspects

Hour of sampling:

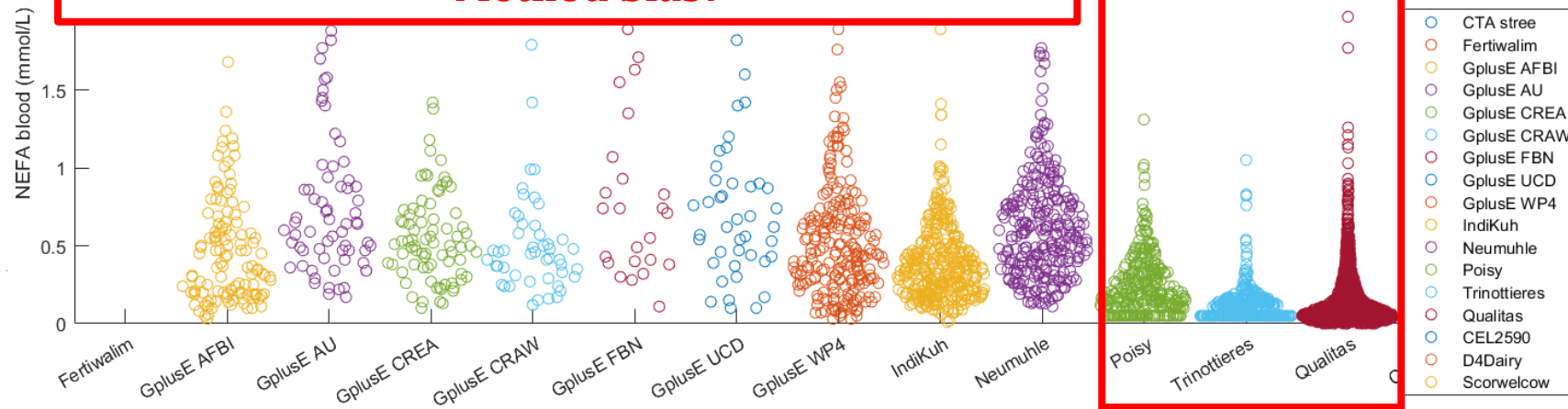
- 1470 after AM milking
- 558 after AM or PM milking
- 3385 with no information



Real differences (breed, diet, management...)

Or

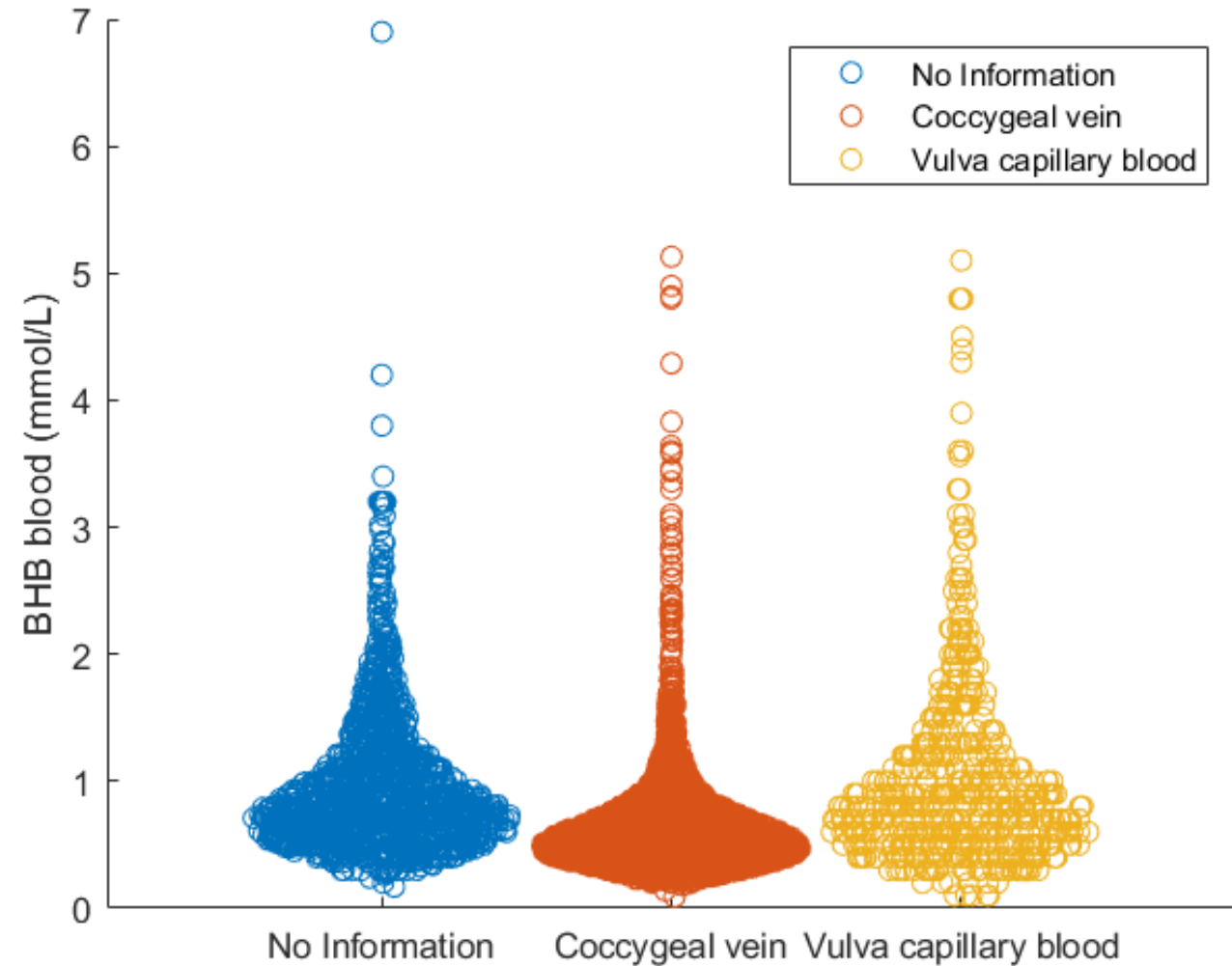
Method bias?



Methodological aspects

Area of sampling

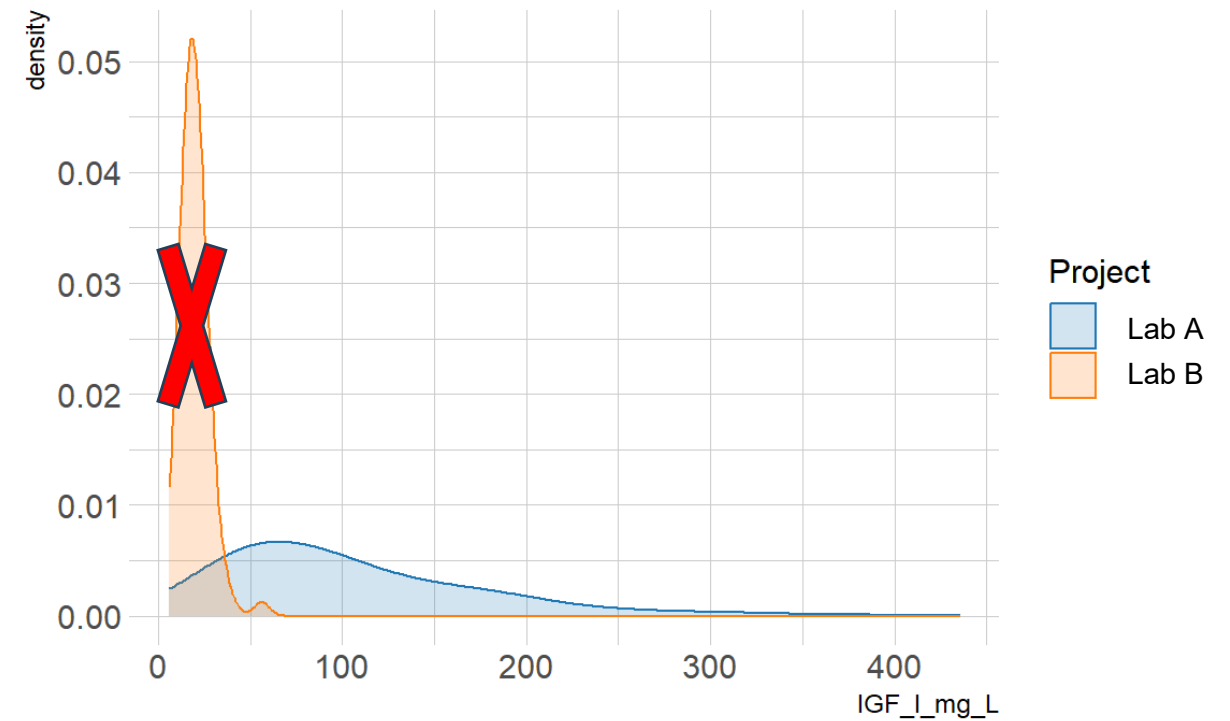
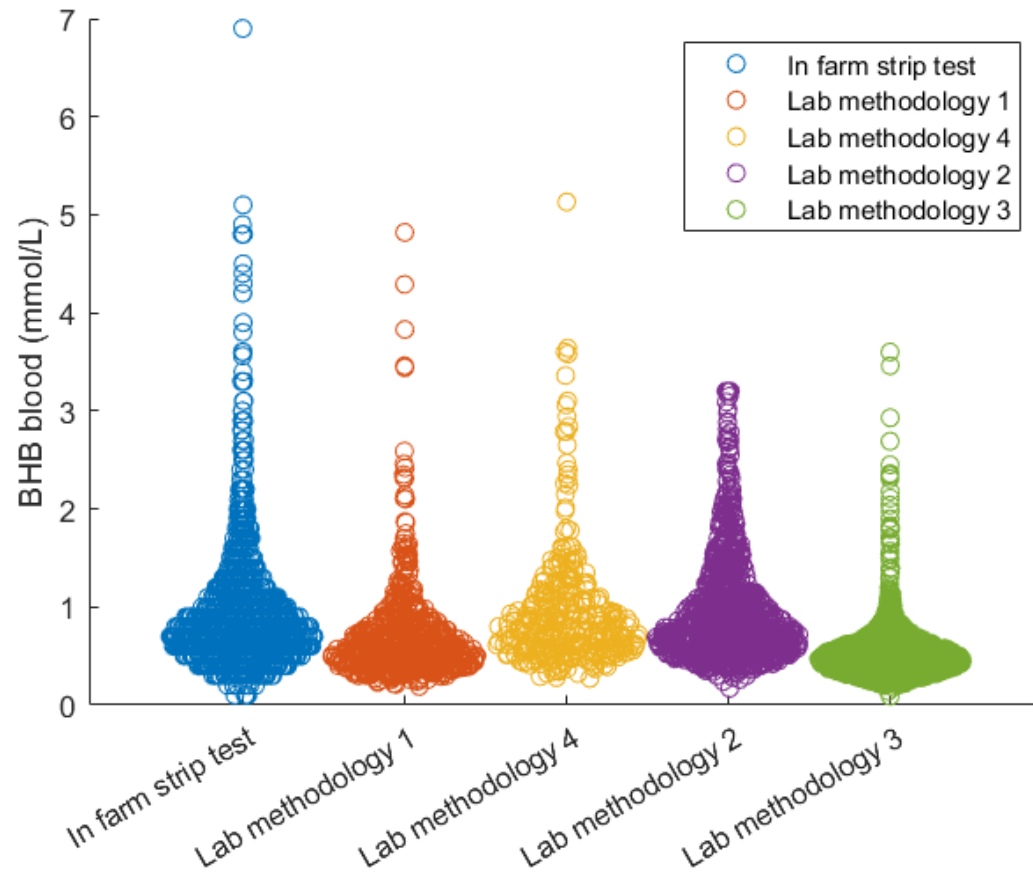
- Coccygeal vein : 3084
- Vulva capillary blood : 558
- No information : 1217



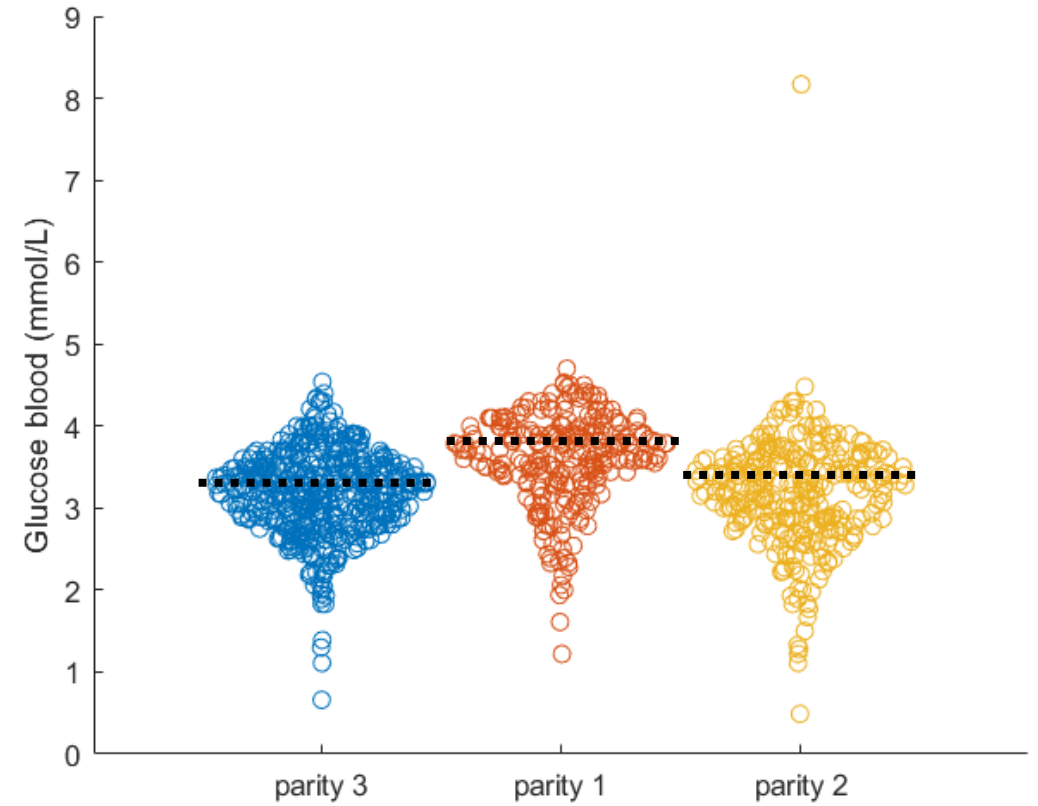
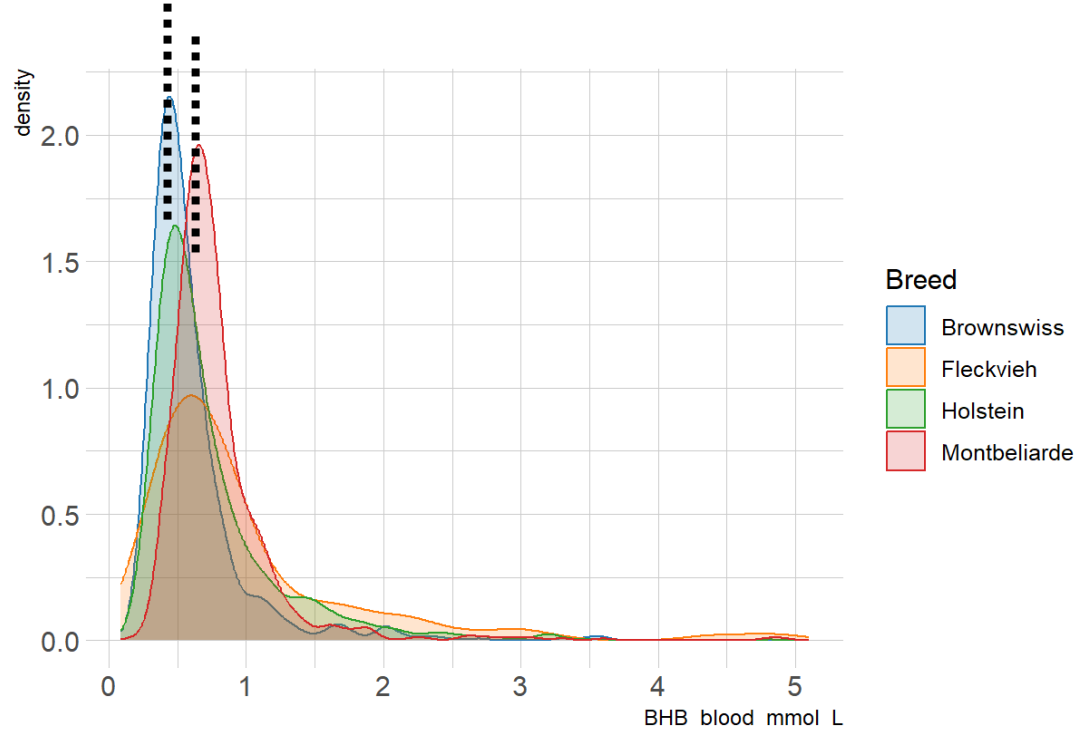
Methodological aspects

Analytical method

- Lab analysis (4 types) : 3618
- In farm strip tests : 1241



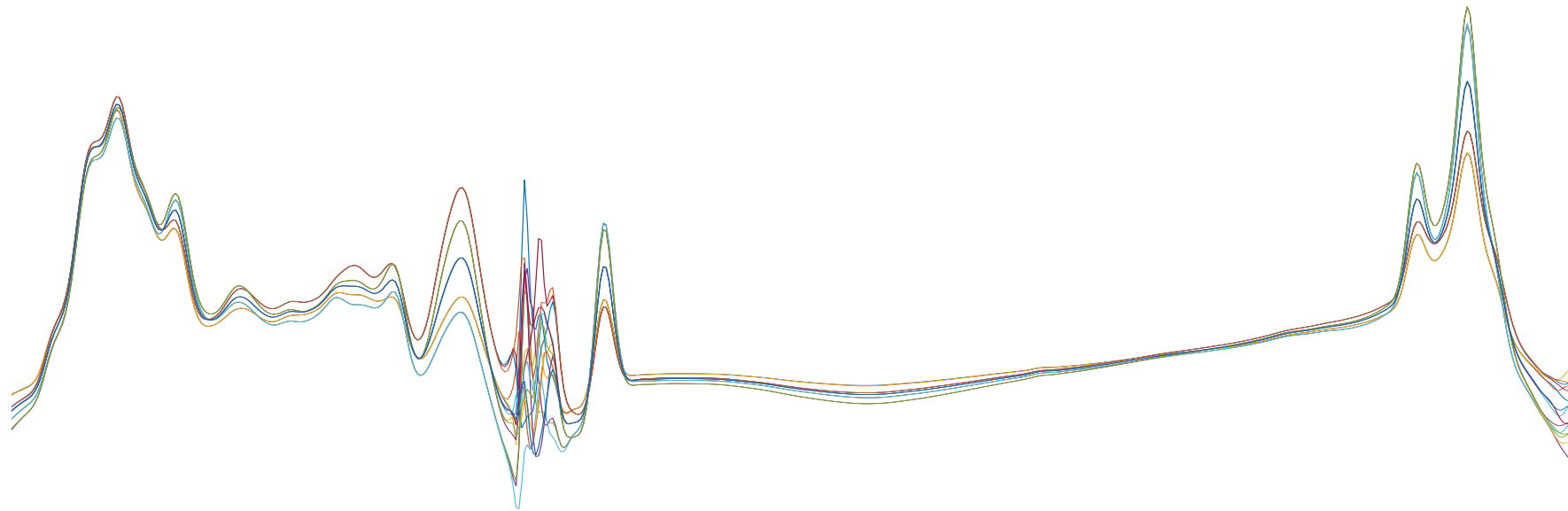
Animal effects

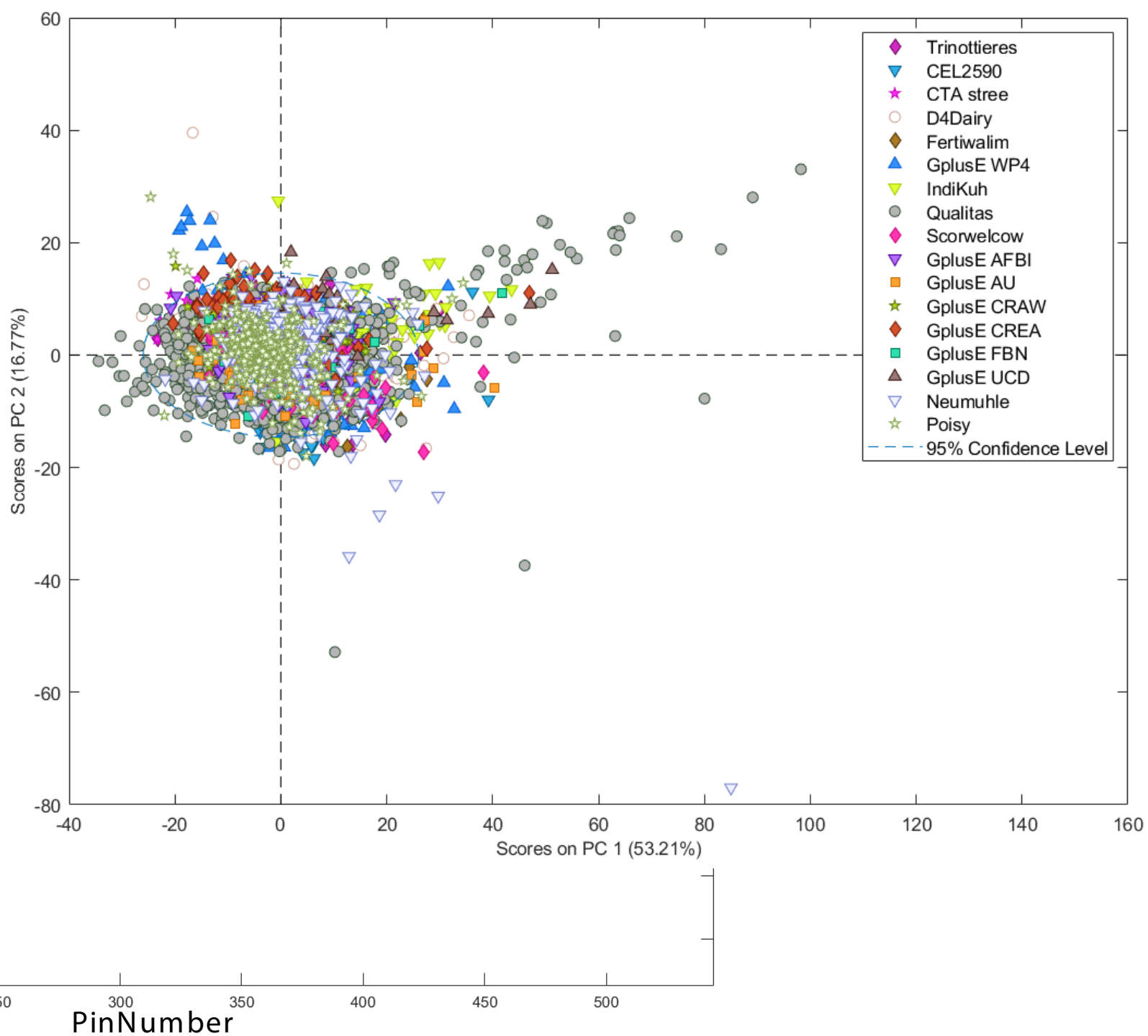
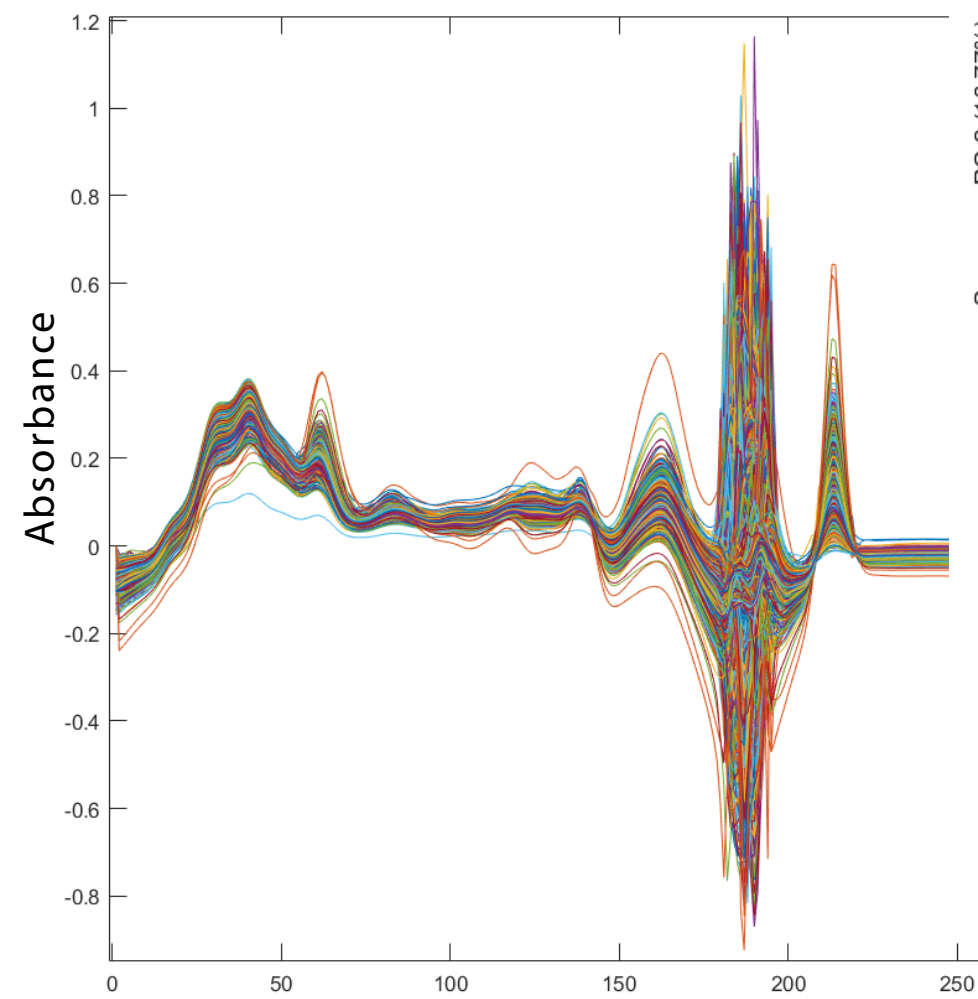


	BHB (mmol/L)	NEFA (mmol/L)	Glucose (mmol/L)	IGF-I (ng/ml)	Urea (mmol/L)	Cholesterol (mmol/L)	Fructosamin (μmol/L)	Progesteron (ng/ml)
Parity	0.08	0.06	-0.19	-0.29	-0.12	0.02	-0.05	0.10
Parity²	0.03	0.01	-0.14	-0.2	-0.15	-0.01	-0.03	0.07
DIM	-0.17	-0.46	-0.24	0.26	0.16	0.67	0.07	0.45
DIM²	-0.11	-0.34	-0.23	0.2	0.17	0.67	0.07	0.45
MY24h	0.08	0.17	-0.19	-0.07	0.09	0.26	0.03	0.20

Breed (Holstein vs others)
Parity
MY
DIM
→ To include in models

II – First draft of MIR models

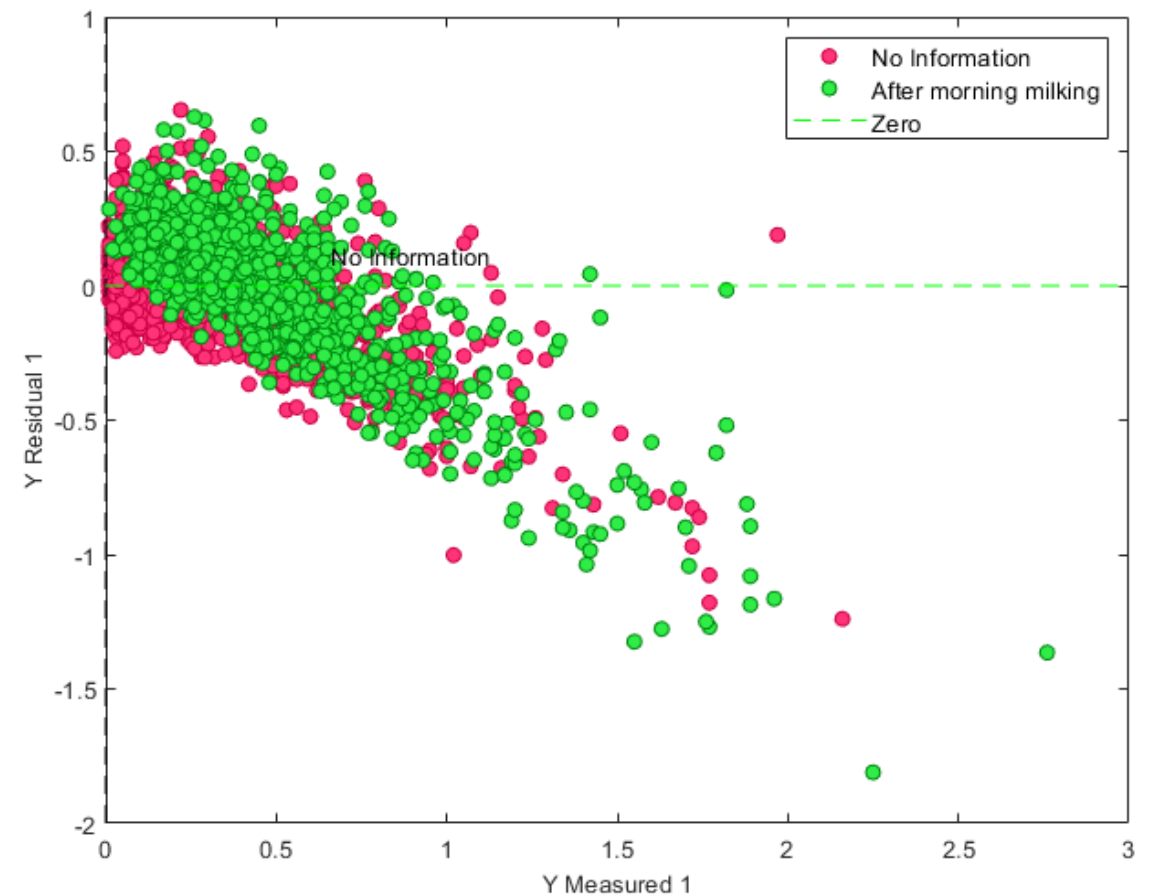
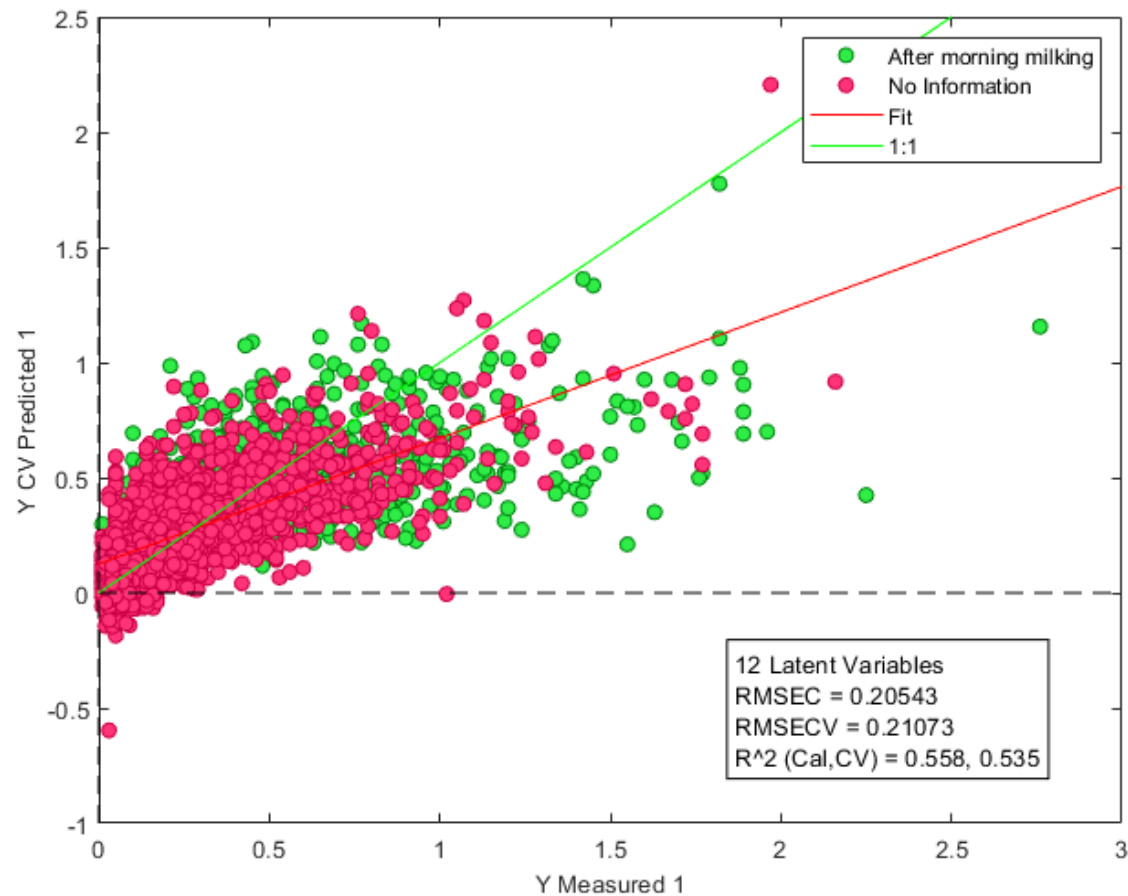




Impact of protocols on models (hour of sampling)

NEFA example (mmol/L)

Differences of sampling time on NEFA are probably real differences, not methodological artefact



Best results up to now

	n records	n herds	Modelling Predictors		RMSEcal	RMSEcv	RMSEval	R ² cal	R ² cv	R ² val
BHB (mmol/L)	4221	74	SVM	MIR	0.30	0.36	0.40	0.65	0.47	0.39

BHB model

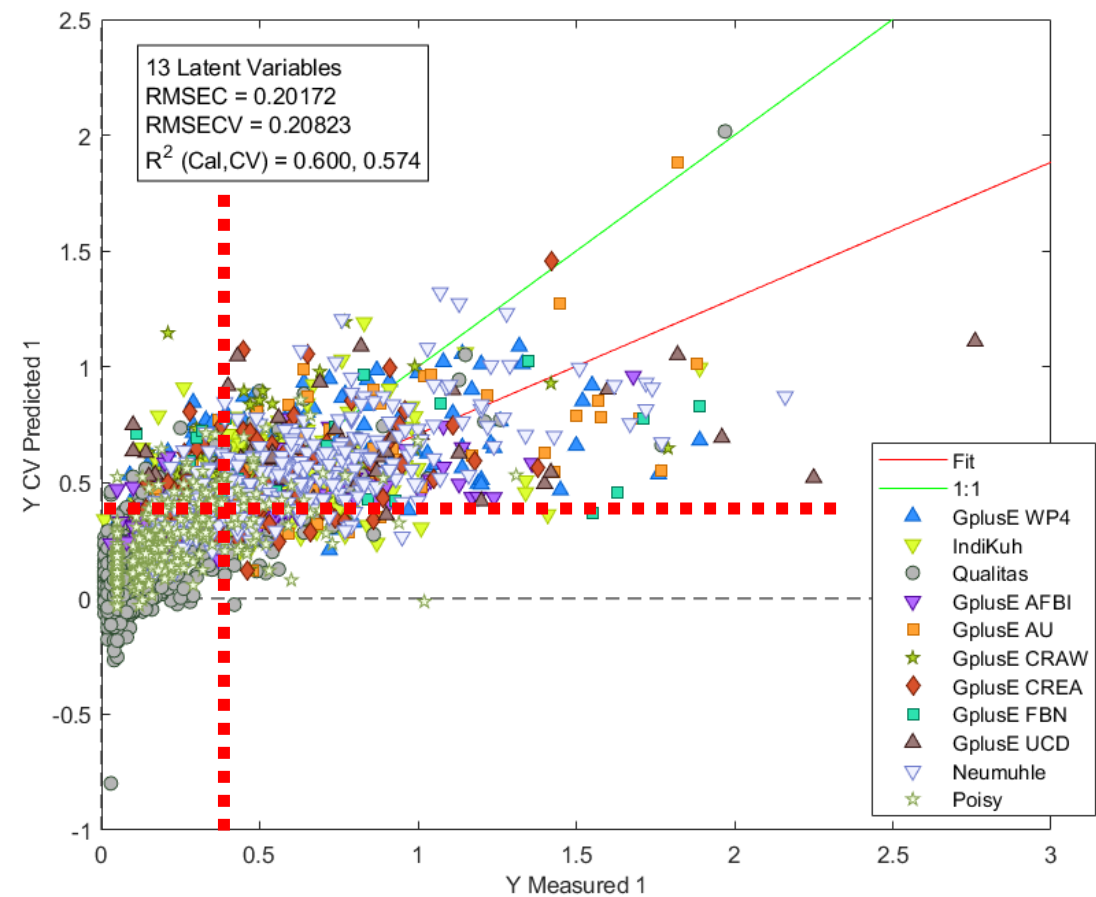
4 subsets excluded on 17 (13% excluded for data quality: sampling, storage, analysis...?)

Best results up to now

BHB (mmol/L)

NEFA (mmol/L)

n records	n herds	Modelling Predictors	RMSEcal	RMSEcv	RMSEval	R ² cal	R ² cv	R ² val
4221	74	SVM MIR	0.30	0.36	0.40	0.65	0.47	0.39
3180	27	PLS MIR+breed+MY+parity+DIM	0.20	0.21	0.23	0.61	0.57	0.48



Best results up to now

	n records	n herds	Modelling	Predictors	RMSEcal	RMSEcv	RMSEval	R ² cal	R ² cv	R ² val
BHB (mmol/L)	4221	74	SVM	MIR	0.30	0.36	0.40	0.65	0.47	0.39
NEFA (mmol/L)	3180	27	PLS	MIR+breed+MY+parity+DIM	0.20	0.21	0.23	0.61	0.57	0.48
Glucose (mmol/L)	1038	22	PLS+Log	MIR+breed+MY+parity	0.44	0.49	0.58	0.44	0.31	0.17
IGF-I (ng/mL)	380	6	PLS+Log	MIR+breed+MY+parity	41	52	56	0.67	0.50	0.44
Urea (mmol/L)	380	6	PLS	MIR	0.66	0.85	1.07	0.67	0.48	0.29
Cholesterol (mmol/L)	380	6	PLS+Log	MIR+breed+MY+parity+DIM	0.64	0.79	1.12	0.69	0.53	0.25
Fructosamin (μmol/L)	374	6	PLS+Log	MIR	13.7	15.9	22.1	0.38	0.20	0.03
Progesteron (ng/ml)	373	6	PLS	MIR+breed+MY+parity+DIM	2.3	2.6	2.9	0.37	0.22	0.18

6 herds
 → not robust enough
 → 1 or 2 herds in
 external validation

Conclusion

- Win/win international collaboration (still open)
- Ongoing process, to reach >10.000 records
- Data quality !!
- Importance of harmonization of protocols, sampling, analytical methods
 - ICAR, IDF, ISO ?
- First draft of models are promising for some molecules



Next steps

- To include the other datasets (Canada, France, Uruguay, Finland...others?)
- Qualitative models (high vs low)?
- Test new methodologies
- To validate on field
- To be used in real life for farm management (on standardized instrument)



Thank you for your attention!

C.grelet@cra.wallonie.be

