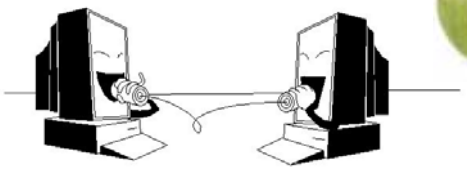
Action 3.5. : Comparison of methods for spectral standardization

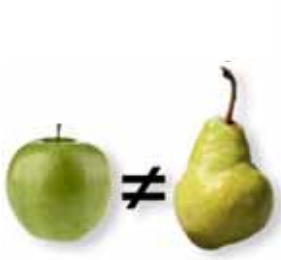
Clément Grelet, Mazen Bahadi, Pierre Broutin, Malia Caputo, Octave Christophe, Laura Dale, Maria Frizzarin, Julie Leblois, André Mensching, Juliana Petrini, Silvia Orlandini, Kevin Vancleef, Per Waaben Hansen, Valérie Wolf, Hélène Soyeurt & Frédéric Dehareng

Objective and importance of MIR spectral standardization

Harmonization of spectra across instruments, labs, countries, continents

- Sharing of model
- Sharing of data (costly reference)





Harmonization of spectra across time

- Useability of models in time
- Stability of predictions

Harmonization of historical database

- Use of new models on historical spectra
- Power of big data
- Genetic evaluation

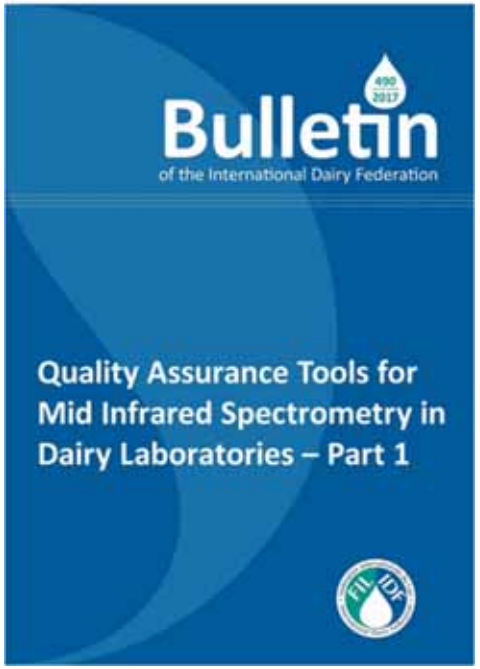
Different existing methods

Manufacturers “baseline methods”

- **Virtual Master Spectrum Approach (Foss)**
- **Universal Spectra Standardization (Bentley)**

Additional spectral standardization methods

- **Grelet et al., 2015**
- **Bonfatti et al., 2017**
- **Mensching et al., 2025**



→ Main goal of ExtraMIR 3.5 task is to evaluate/compare the standardization procedure

Methods tested in the comparison

Manufacturer Baseline
- "FTIR Equalizer" method
- Standardizing the intensity based on the intensity of the same chemical standard
- The method ensure transferability within one instrument platform

Grelet et al., 2015
- Based on monthly exchange of common raw milk samples (specifically designed for this)
- Across manufacturer brands
- > 100 models already developed

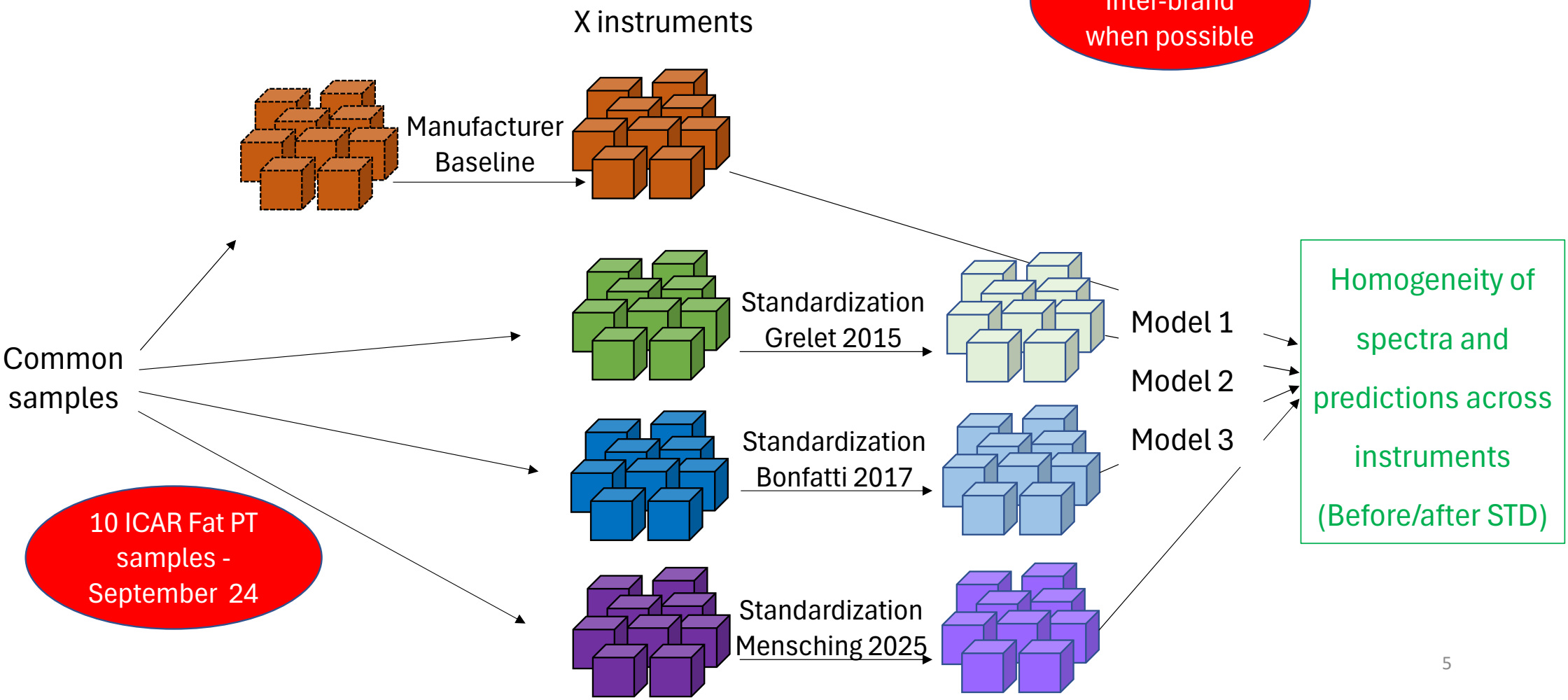
Bonfatti et al., 2017
- Based on historical spectral records
- No need for common samples exchange
- Requires the accumulation of considerable amount of milk spectra

Mensching et al., 2025
- Based on a statistical standardization framework incorporating additional data sources
- No common (additional) samples are required
- Applicable to both historical and daily incoming data
- Currently implemented for FOSS instruments

Methodology of comparison

10 instruments per organisation, within at least 2 labs

Intra-brand
Inter-brand
when possible



Data treatment

Collect and anonymization of data

Interpolation of all spectra into a common range

Switch of all spectra in absorbance

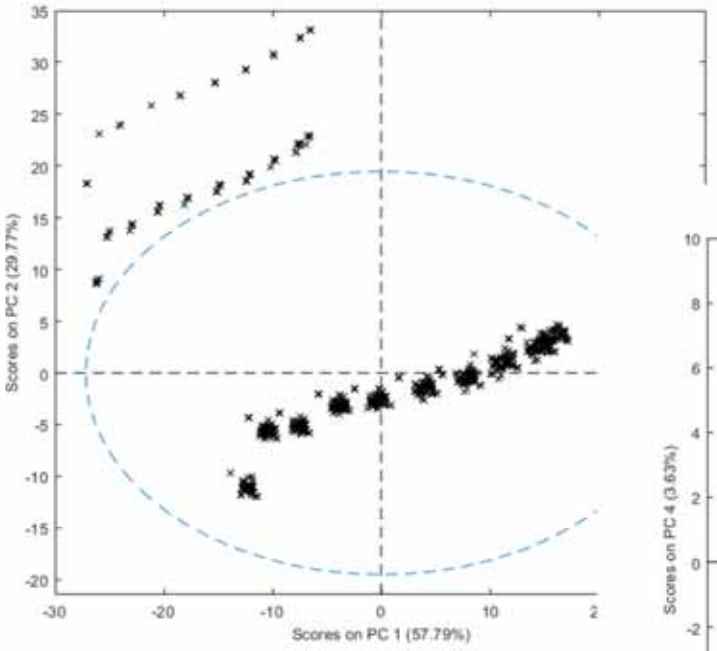
Application of 10 models from CRA-W/Ulg with different phenotypes and accuracies

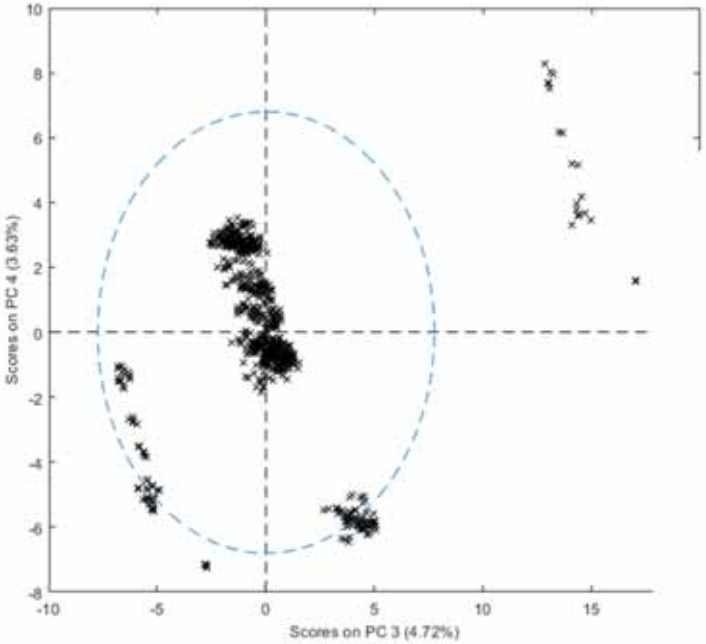
	Models	R^2cv	RMSEcv
Fatty acids	Fat (g/100ml)	1.00	0.040
	SAT (g/100ml)	0.99	0.072
	C18_1cis9 (g/100ml)	0.95	0.062
	C8 (g/100ml)	0.91	0.0040
	PUFA (g/100ml)	0.77	0.0210
Cow status	N efficiency (%)	0.74	5%
	Blood BHB (mmol/L)	0.70	0.270
	CH4 (g/day)	0.68	57.0
	Milk BHB (µmol/L)	0.61	64.6
	Blood Glucose (mmol/L)	0.44	0.360

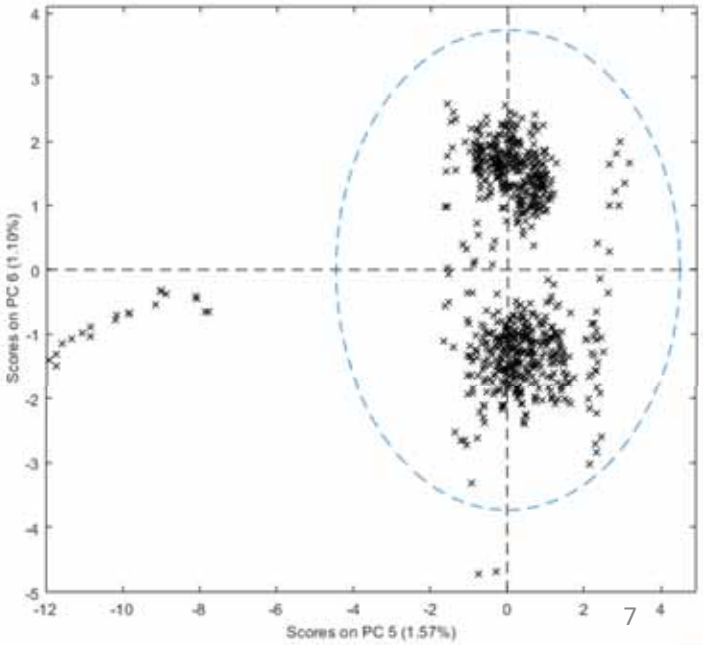
Editing of data (deletion of inconsistent results, mixing of samples…)

Evaluation on spectral homogeneity across instruments before standardization (within each methodology)

- BEFORE standardization
- PCA on 212 repeatable wavenumbers and 1rst derivative
- All brands

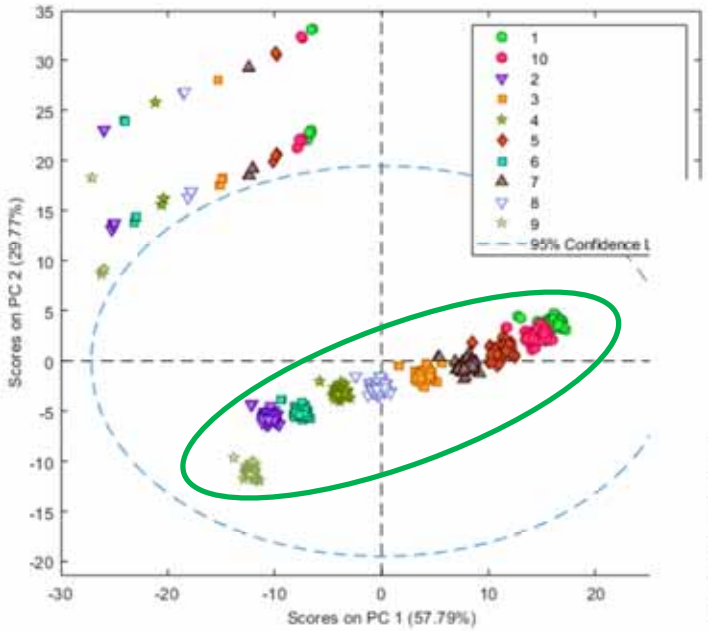




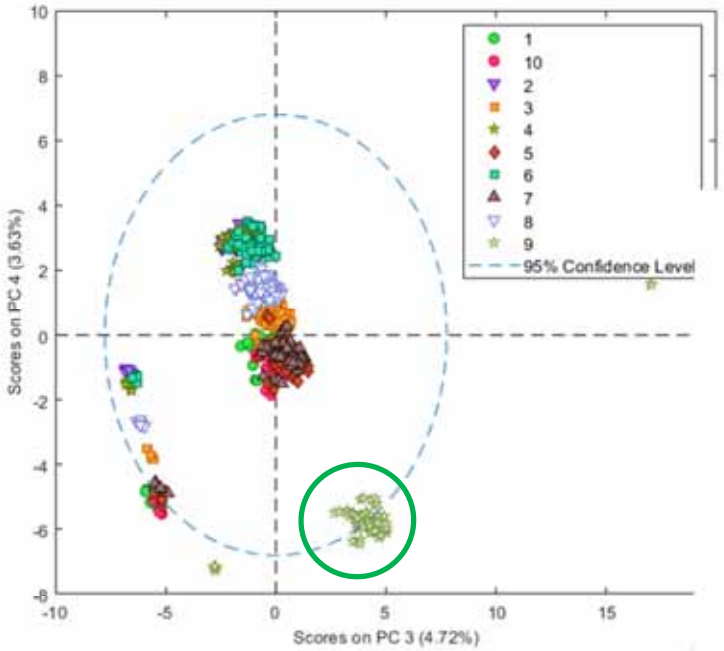


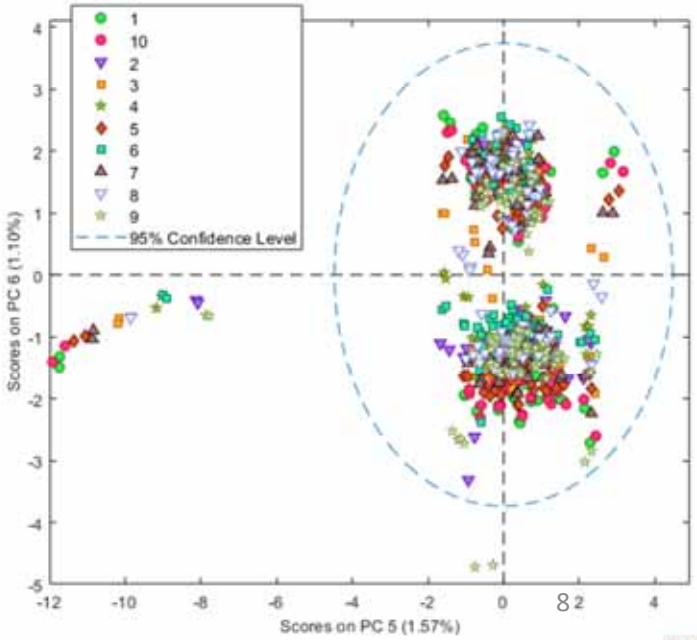
Class Set: #1

Evaluation on spectral homogeneity across instruments before standardization (within each methodology)

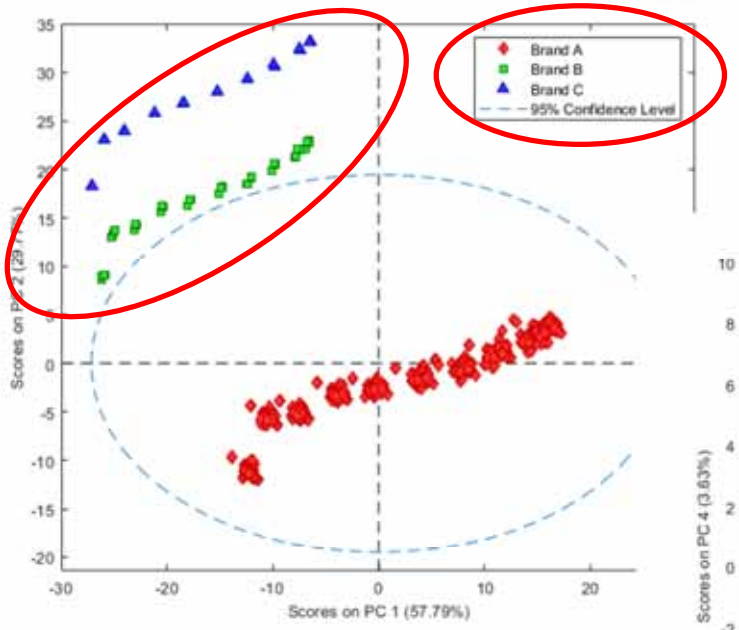


- BEFORE standardization
- PCA on 212 repeatable wavenumbers and 1rst derivative
- All brands
- **Sample effect**

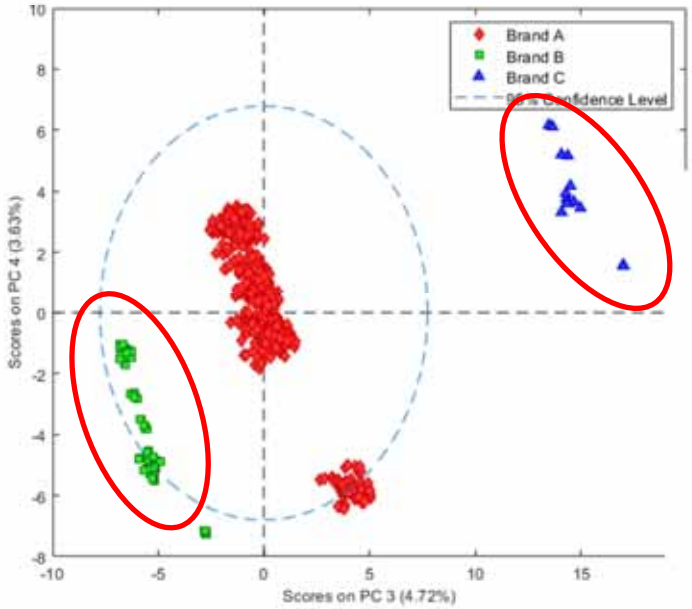


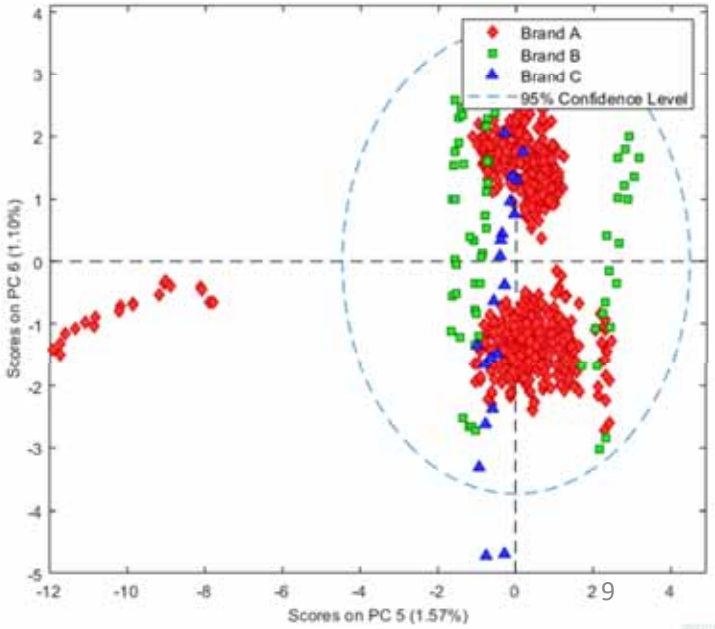


Evaluation on **spectral homogeneity across instruments** before standardization (within each methodology)

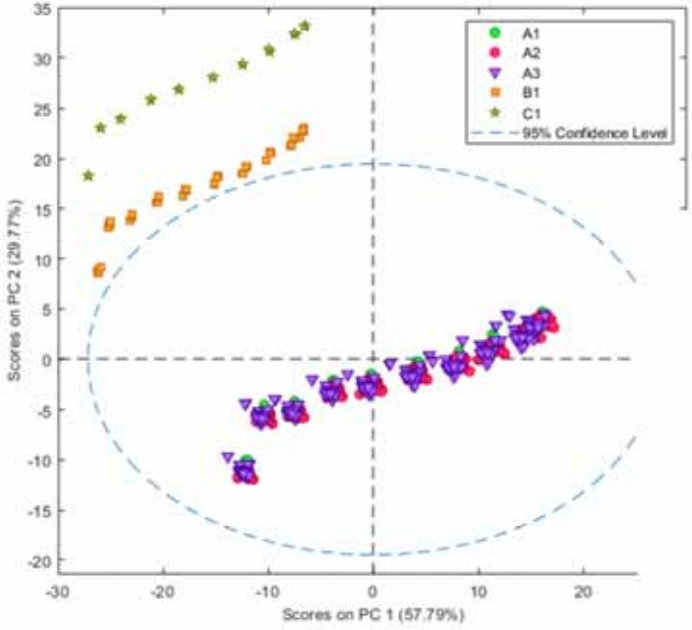


- BEFORE standardization
- PCA on 212 repeatable wavenumbers and 1rst derivative
- All brands
- **Brand effect**



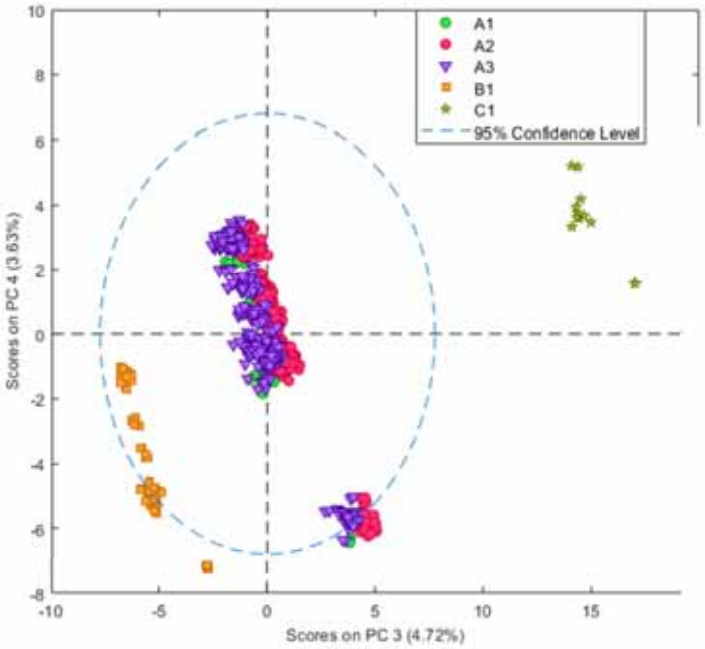


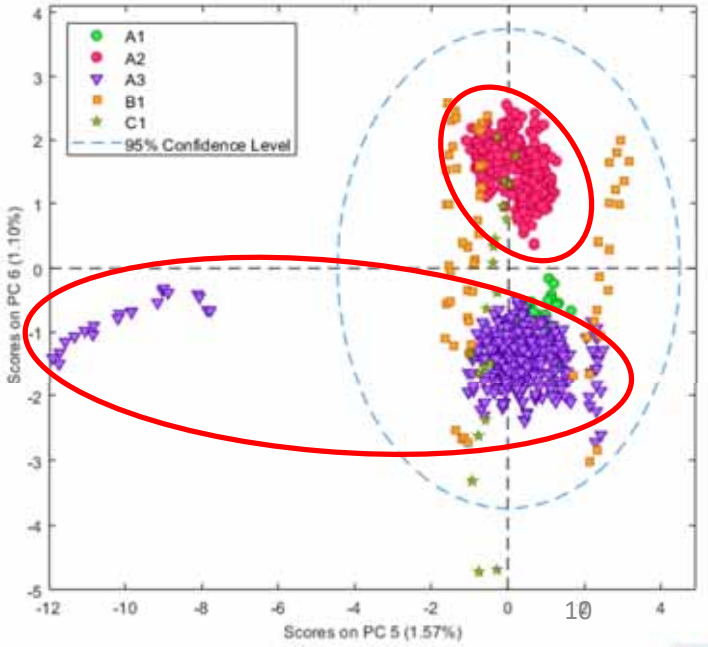
Evaluation on spectral homogeneity across instruments before standardization (within each methodology)



- BEFORE standardization
- PCA on 212 repeatable wavenumbers and 1rst derivative
- All brands
- **Instrument model effect**

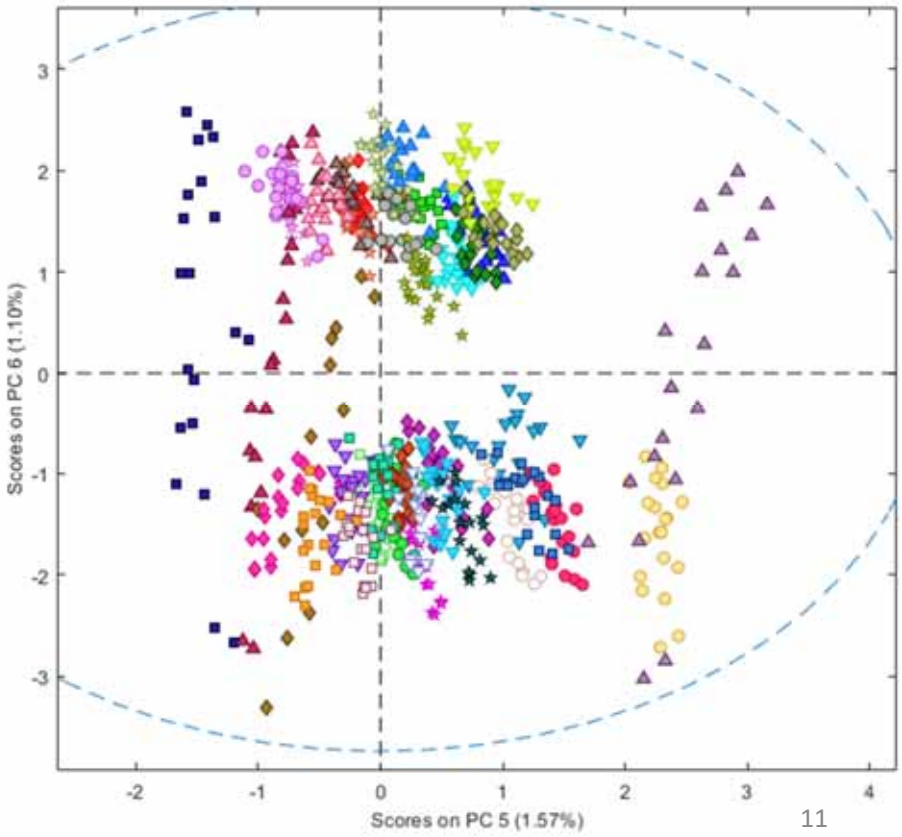
Brand	Model
Brand A	A1
Brand A	A2
Brand A	A3
Brand B	B1
Brand C	C1





Evaluation on spectral homogeneity across instruments before standardization (within each methodology)

- BEFORE standardization
- PCA on 212 repeatable wavenumbers and 1rst derivative
- All brands
- **Machine effect**



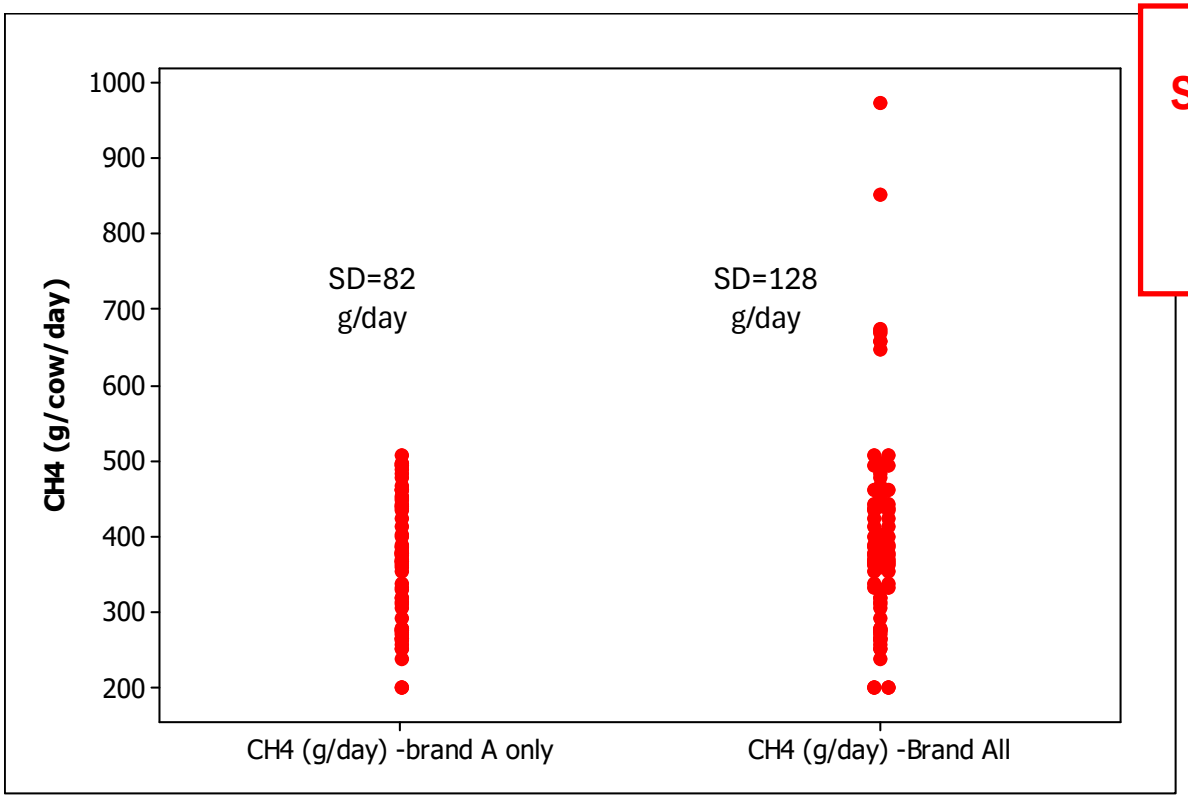
Conclusion 1:
Spectra vary according to
- **samples**
- **instrument brands**
- **instrument models**
- **individual machines**

Evaluation on spectral homogeneity across instruments before standardization (within each methodology)

Analysis of one common sample by all machines (sample 5 ~ 4.13 g/100g)

Application of 1 common CH4 model

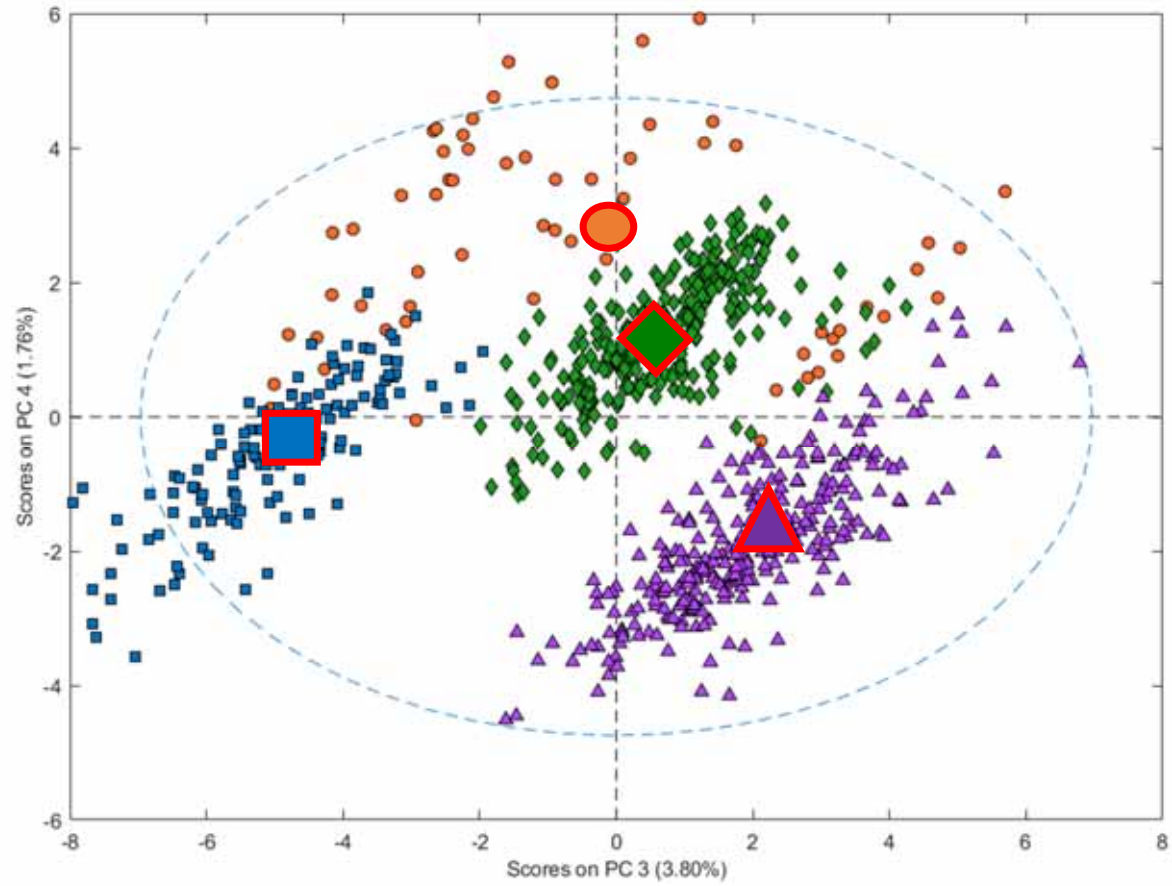
→ **Impact on predictions**



Conclusion 2:
Spectral heterogeneity across instruments leads to prediction heterogeneity across instruments

Evaluation on spectral homogeneity across instruments after standardization (within each methodology)

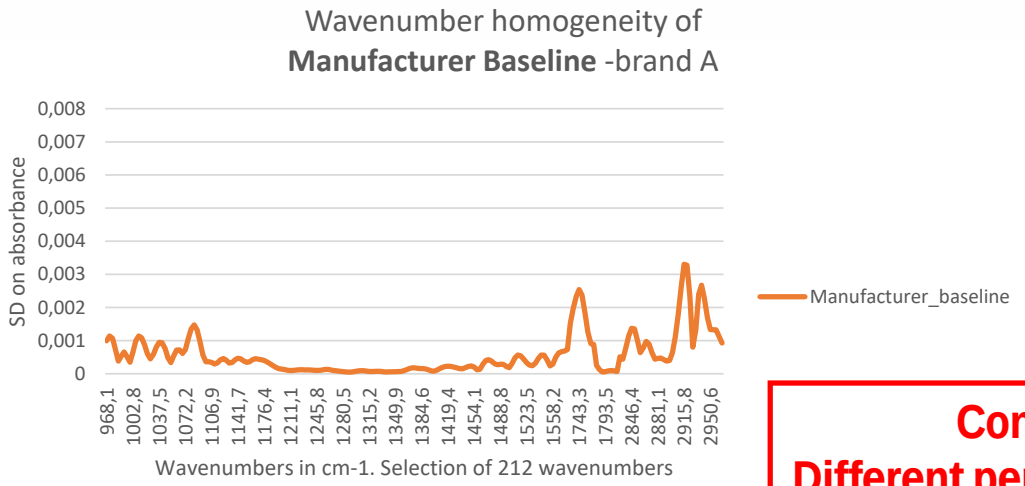
- **AFTER standardization within each methodology**
- PCA on 212 repeatable wavenumbers and 1rst derivative
- Brand A only
- **Method effect**

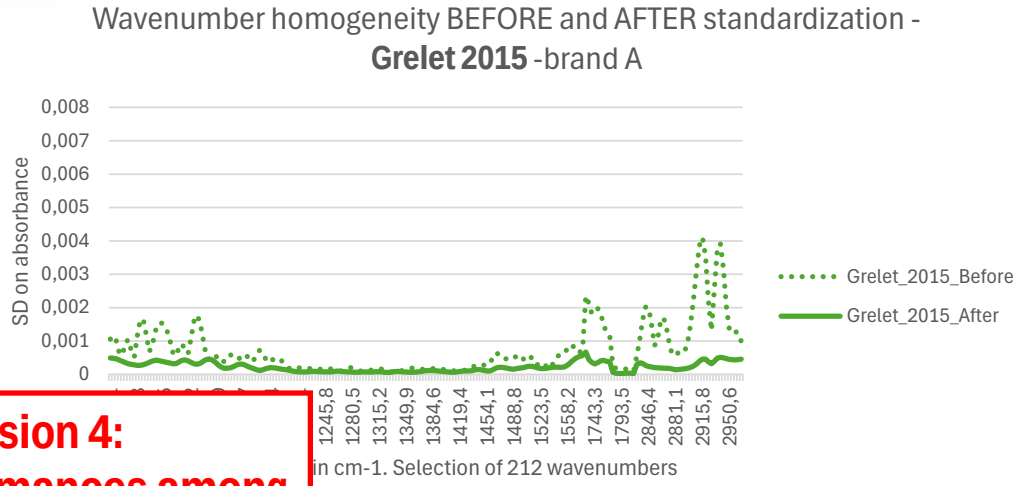


Manufacturer Baseline
Grelet et al., 2015
Bonfatti et al., 2017
Mensching et al., 2025

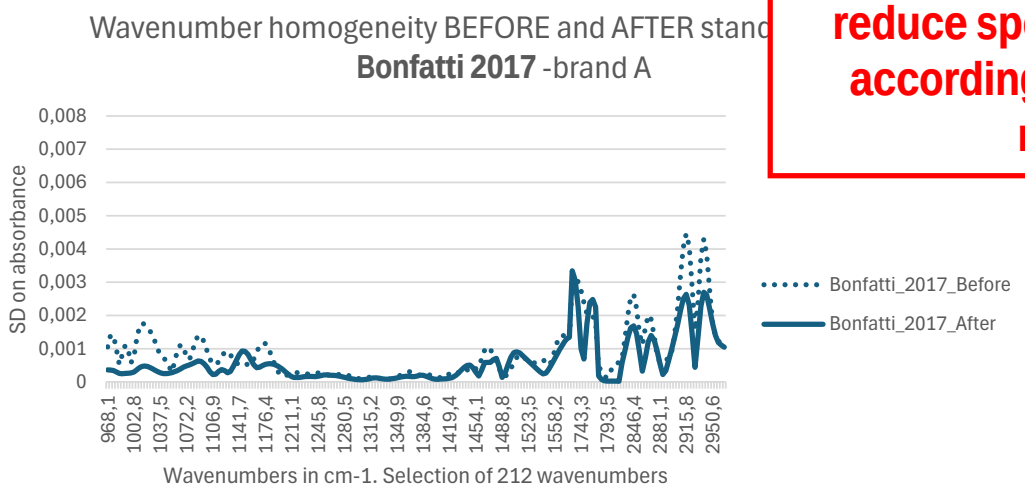
Conclusion 3:
After standardization, the spectra of each network has a different centroid, a different format

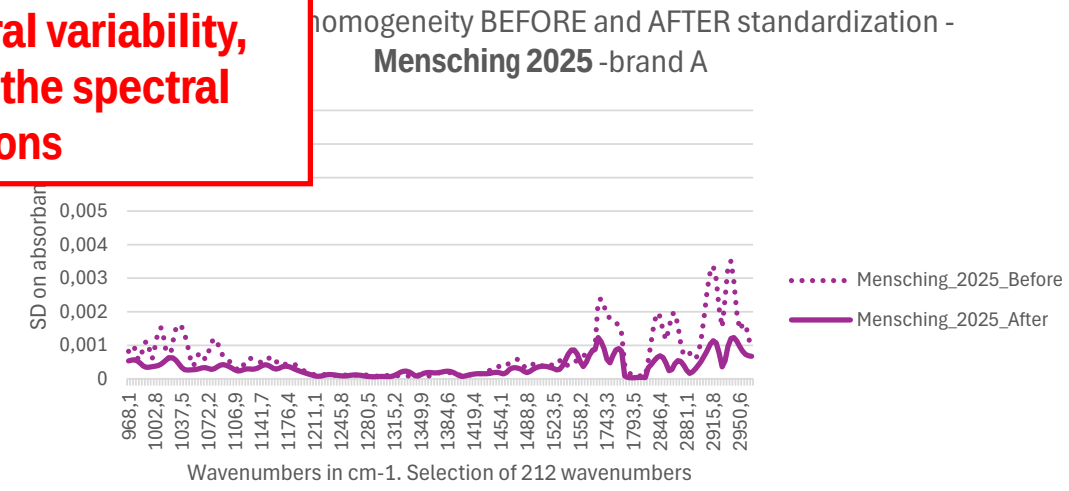
Evaluation on spectral homogeneity across instruments before and after standardization (within each methodology)



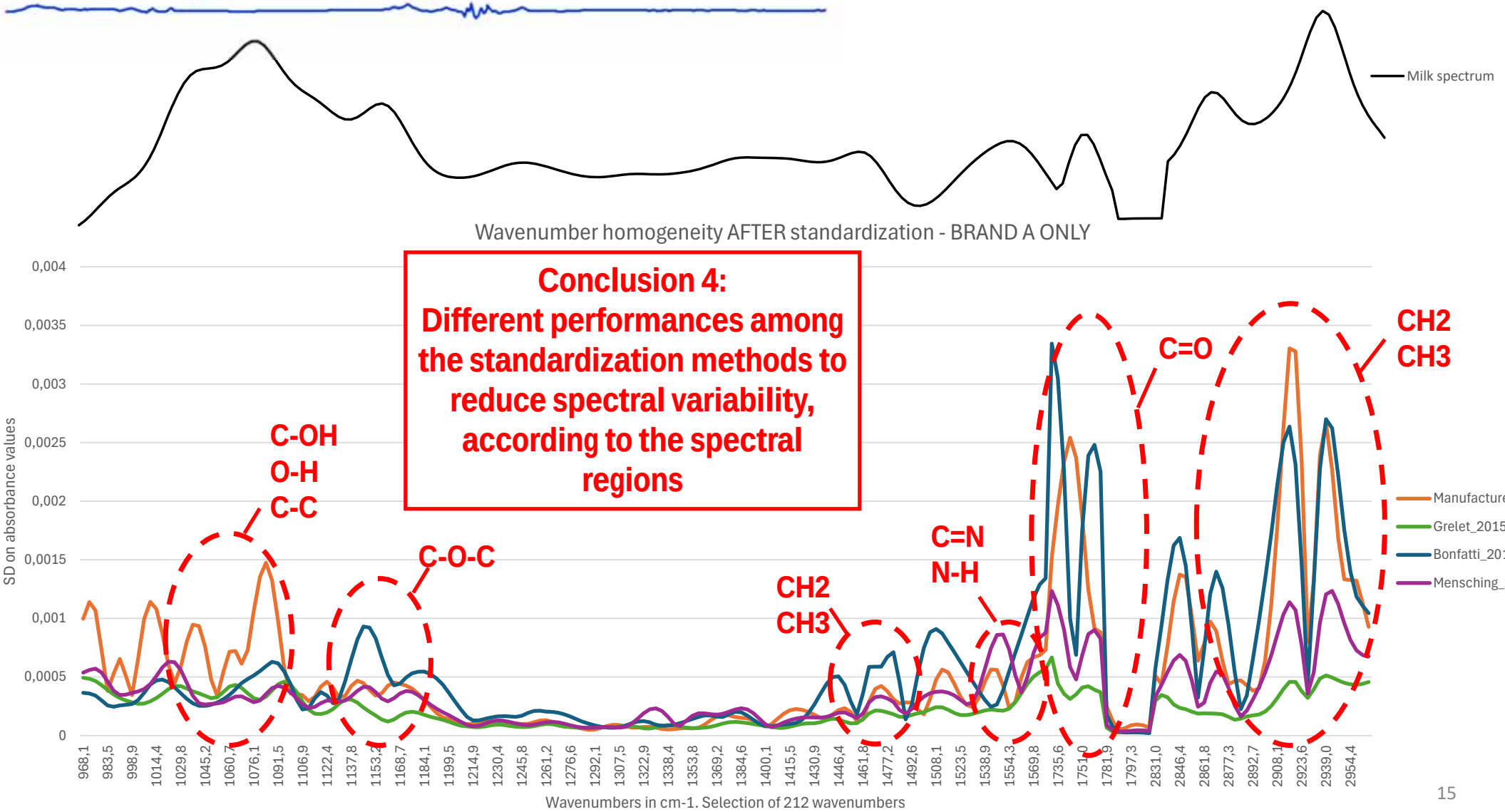


Conclusion 4:
Different performances among the standardization methods to reduce spectral variability, according to the spectral regions





Evaluation on spectral homogeneity across instruments after standardization (within each methodology)

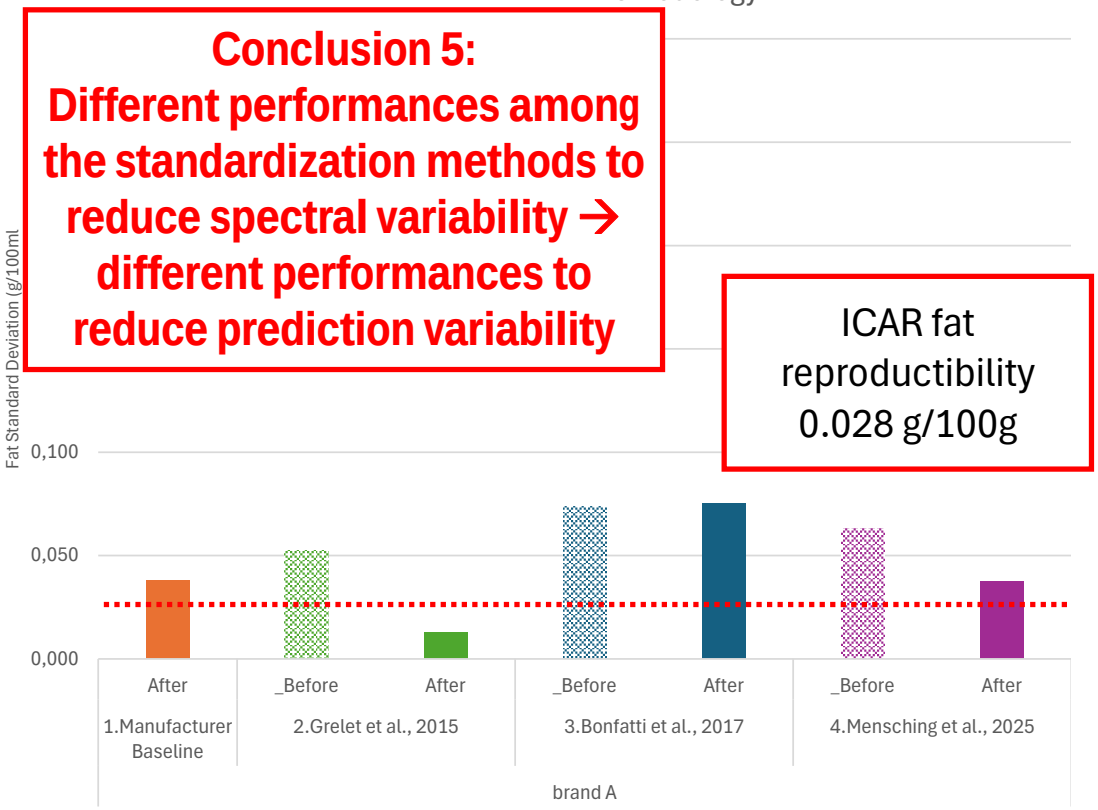


Evaluation on Predictions homogeneity across instruments before and after standardization (within each methodology)

Standard deviation calculated per samples and equation and averaged per methodology (for the 10 samples)

	Method	STD	n machines	Fat (g/100ml)
	Manufacturer Baseline		5	0.038
brand A	Grelet et al., 2015	Before	11	0.053
brand A	Grelet et al., 2015	After	11	0.013
	Bonfatti et al., 2017	Before	7	0.074
	Bonfatti et al., 2017	After	7	0.075
	Mensching et al., 2025	Before	14	0.063
	Mensching et al., 2025	After	14	0.038

Standard deviation of fat predictions, for the 10 samples individually and per methodology



Conclusion 5:
Different performances among the standardization methods to reduce spectral variability → different performances to reduce prediction variability

ICAR fat reproductibility 0.028 g/100g

Lab results (all labs): Slope and bias corrected fat SD=0.042 g/100g

Evaluation on Predictions homogeneity across instruments before and after standardization (within each methodology)

Standard deviation calculated per samples and equation and averaged per methodology (for the 10 samples)

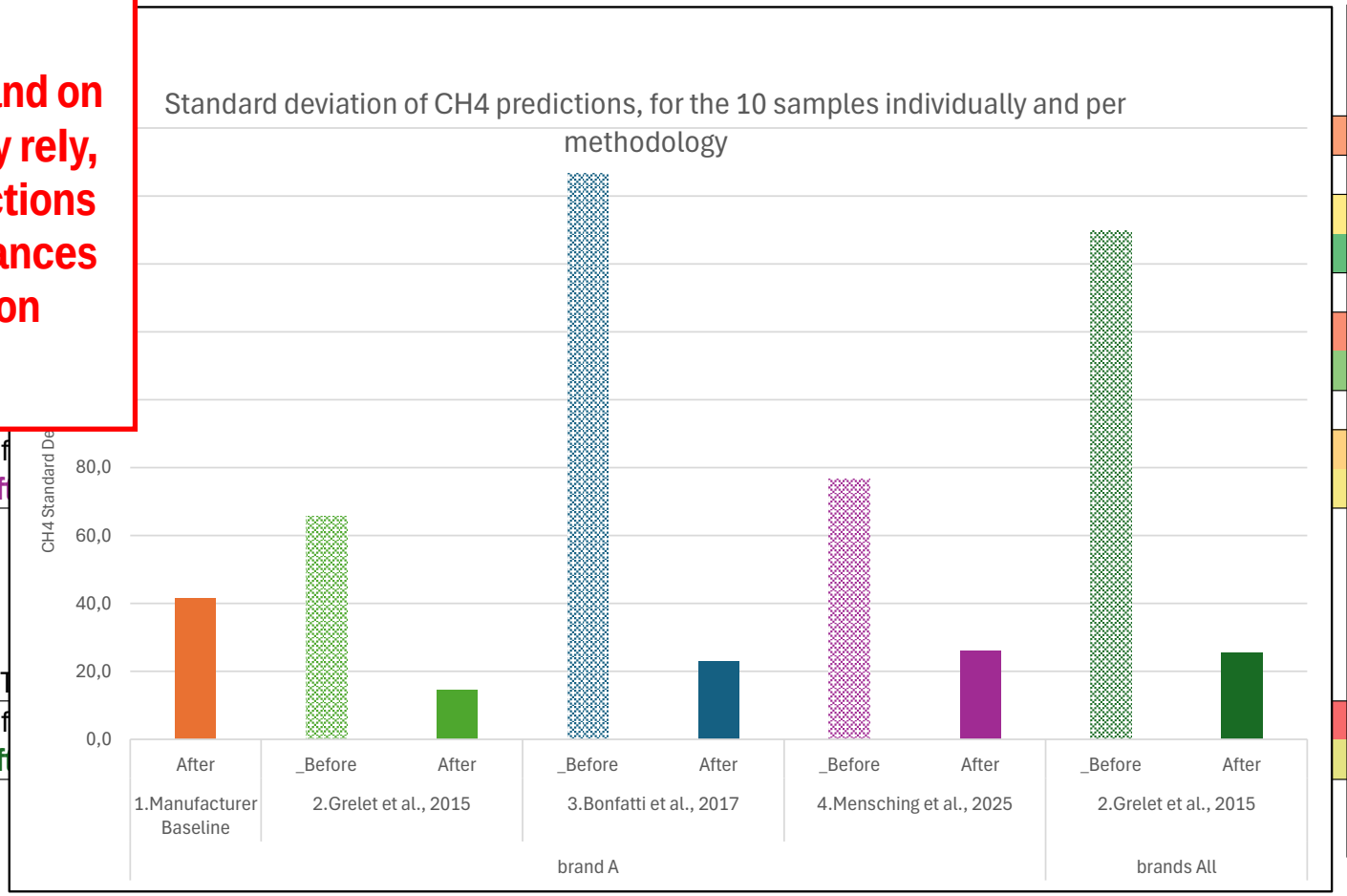
	Method	STD	n machines	Fat (g/100ml)	SAT (g/100ml)	C18_1cis9 (g/100ml)	C8 (g/100ml)	PUFA (g/100ml)	Blood BHB (mmol/L)	Blood Glucose (mmol/L)	Milk BHB (µmol/L)	CH4 (g/day)	N efficiency (%)
	Manufacturer Baseline		5	0.038	0.045	0.033	0.0035	0.0261	0.031	0.084	31.6	41.6	5%
brand A	Grelet et al., 2015	Before	11	0.053	0.079	0.085	0.0030	0.0348	0.091	0.101	17.8	65.7	6%
	Grelet et al., 2015	After	11	0.013	0.027	0.024	0.0014	0.0060	0.029	0.039	8.4	14.4	1%
	Bonfatti et al., 2017	Before	7	0.074	0.087	0.076	0.0028	0.0431	0.057	0.119	34.5	166.7	6%
	Bonfatti et al., 2017	After	7	0.075	0.070	0.056	0.0022	0.0099	0.064	0.048	11.1	21.7	2%
	Mensching et al., 2025	Before	14	0.063	0.069	0.085	0.0029	0.0374	0.062	0.110	22.6	76.8	5%
	Mensching et al., 2025	After	14	0.038	0.047	0.040	0.0019	0.0093	0.072	0.083	17.3	26.1	3%

	Method	STD	n machines	Fat (g/100ml)	SAT (g/100ml)	C18_1cis9 (g/100ml)	C8 (g/100ml)	PUFA (g/100ml)	Blood BHB (mmol/L)	Blood Glucose (mmol/L)	Milk BHB (µmol/L)	CH4 (g/day)	N efficiency (%)
all brands	Grelet et al., 2015	Before	15	0.271	0.764	0.350	0.0330	0.0618	0.964	0.539	41.0	149.7	35%
	Grelet et al., 2015	After	15	0.017	0.058	0.045	0.0024	0.0138	0.037	0.102	16.2	25.5	4%

Evaluation on Predictions homogeneity across instruments before and after standardization (within each methodology)

Standard deviation calculated per samples and equation and averaged per methodology (for the 10 samples)

Conclusion 6:
Depending on models, and on which spectral area they rely, harmonization of predictions show different performances across standardization methodologies



	CH4 (g/day)	N efficiency (%)
	41.6	5%
	65.7	6%
	14.4	1%
	166.7	6%
	21.7	2%
	76.8	5%
	26.1	3%

Mensching et al., 2025 Bef
Mensching et al., 2025 Af

	Method	ST
all brands	Grelet et al., 2015	Bef
	Grelet et al., 2015	Af

	CH4 (g/day)	N efficiency (%)
	149.7	35%
	25.5	4%

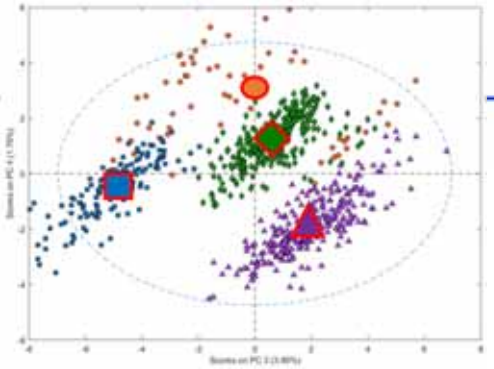
Conclusions

- MIR instruments show a very large spectral variability without standardization (brands, models, machines…)
- Spectral heterogeneity impacts **prediction reproducibility across spectrometers**

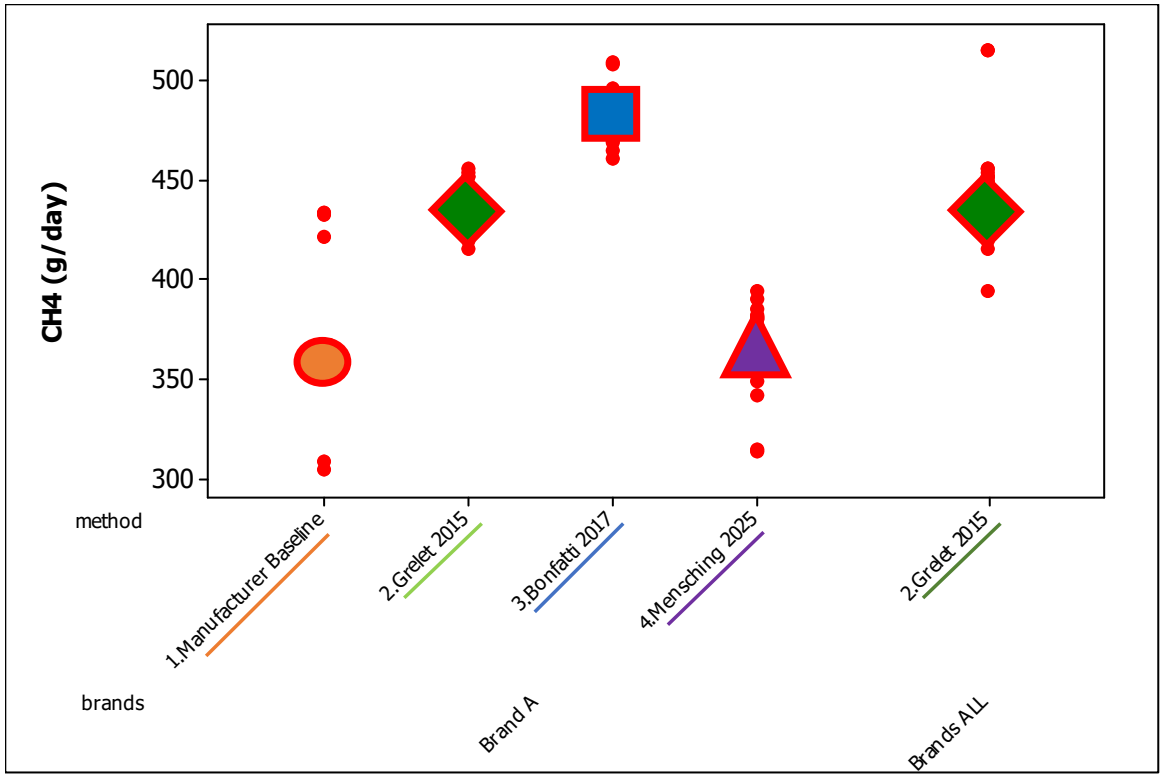
- 4 methodologies of standardization compared
  - 1 Baseline method & 3 *A posteriori* methods
  - Different concepts (historical dataset, sharing of common samples…)
  - Different limitations (sufficient routine amount of samples, cost, transport of samples…)
  - 3 methods intra-brand & 1 inter-brand

- Standardization methodologies enable to reduce spectral variability
- **Very different performances among the standardization methods** to reduce spectral variability, **according to the spectral regions**
- Depending of models (quality), and on which spectral area they rely, **harmonization of predictions show different performances across standardization methodologies**

Conclusions



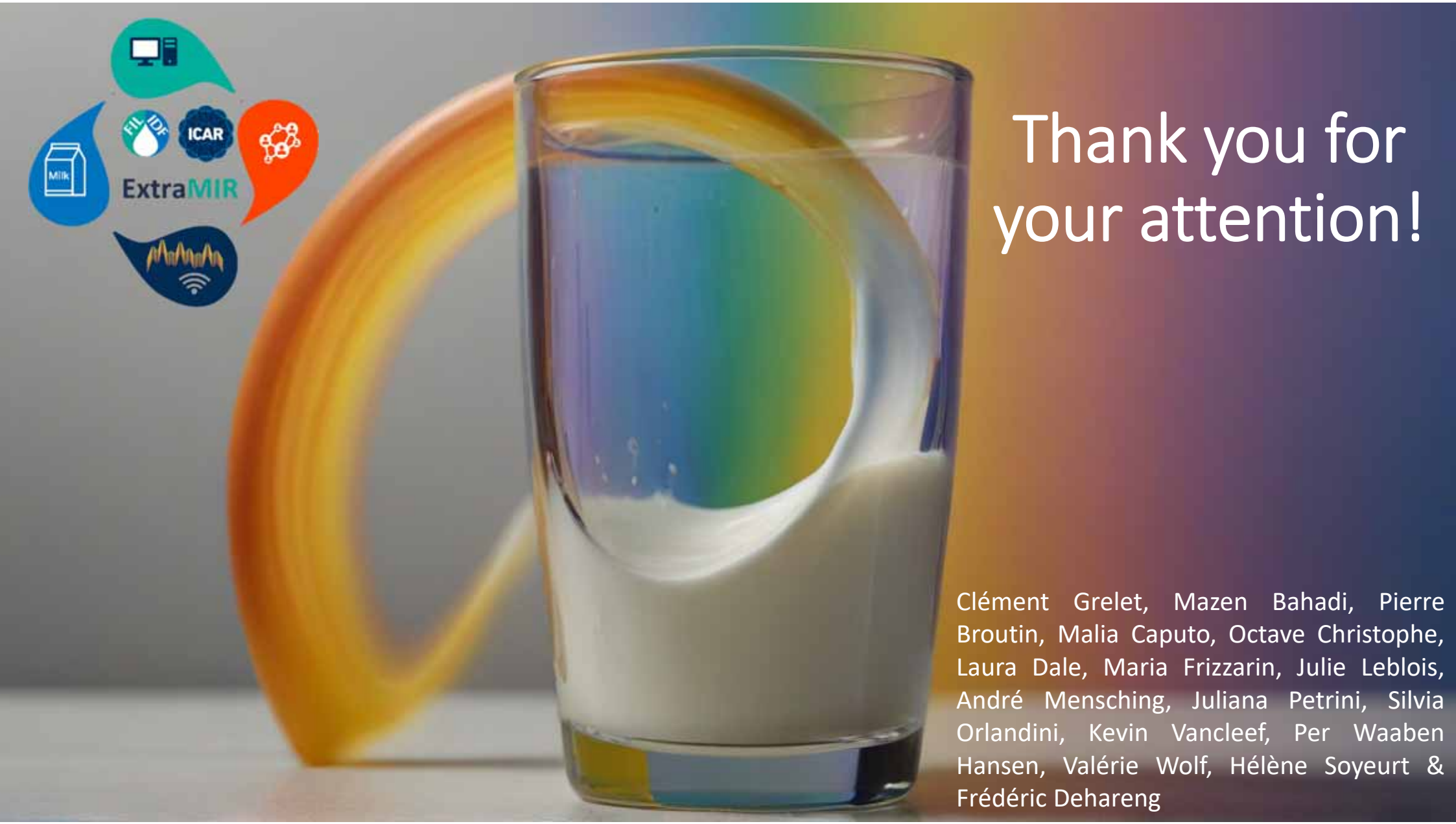
→ *application of 1 common CH4 model on standardized spectra of one milk sample across the different methodologies*



Differences in spectral standardization (centroids) across methodologies

→ **may constrain model transferability between networks** (i.e. sharing models may require standardising spectra to the same standardization process to build the model and for predictions).

Need for further research to better understand and potentially overcome these limitations



Thank you for your attention!

Clément Grelet, Mazen Bahadi, Pierre Broutin, Malia Caputo, Octave Christophe, Laura Dale, Maria Frizzarin, Julie Leblois, André Mensching, Juliana Petrini, Silvia Orlandini, Kevin Vancleef, Per Waaben Hansen, Valérie Wolf, Hélène Soyeurt & Frédéric Dehareng

Evaluation of data

- Evaluation of spectral homogeneity across all instruments before standardization
  - Spectral homogeneity through PCA and predictions homogeneity

- Within each methodology, evaluation on spectral homogeneity across instruments before and after standardization.
  - To calculate the Standard Deviation of absorbance intensities per wavenumber, across 10 instruments, before and after standardization. To be applied only on stable wavenumbers only.

- Within each methodology, evaluation on prediction homogeneity across instruments before and after standardization.
  - To calculate the Standard deviation of predictions across 10 instruments before and after standardization for each methodology and each model.

Dataset description

4 methodologies (44 machines in total)

Manufacturer Baseline – network 1

- **5** instruments brand A (7 in total but 2 with missing spectra)
- After standardization only (5)

Grelet et al., 2015– network 2

- **15** instruments
- 11 instruments brand A
- 3 instruments brand B
- 1 instrument brand C
- All instruments Before and After standardization

Bonfatti et al., 2017 – network 3

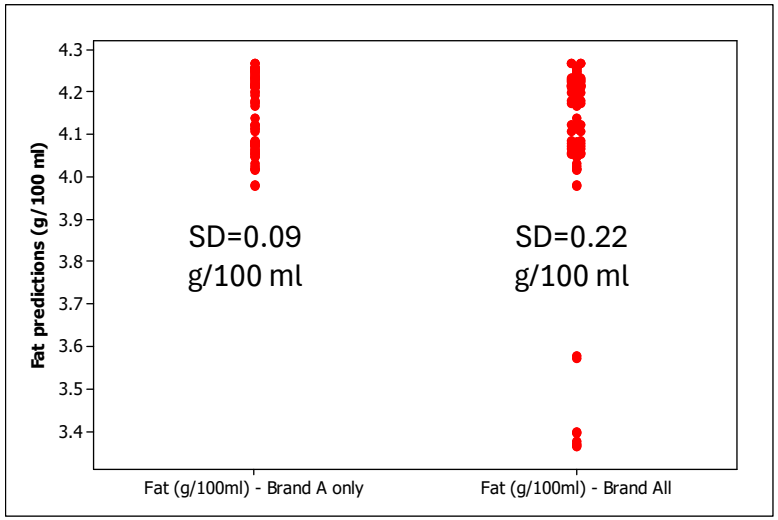
- **10** instruments brand A
- Before (8 instrument) and After (7 instruments) standardization
- One machine excluded because of technical limitation (not enough samples analyzed on the instrument)

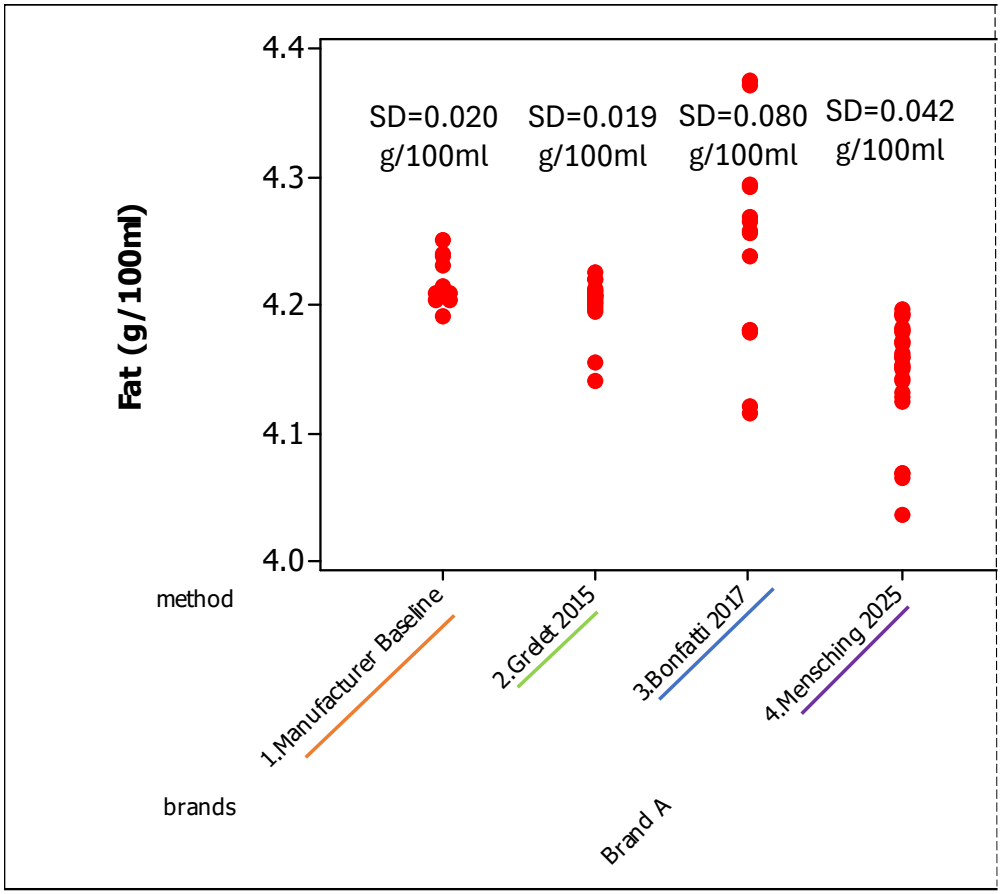
Mensching et al., 2025– network 4

- **14** instruments brand A
- All instruments Before and After standardization

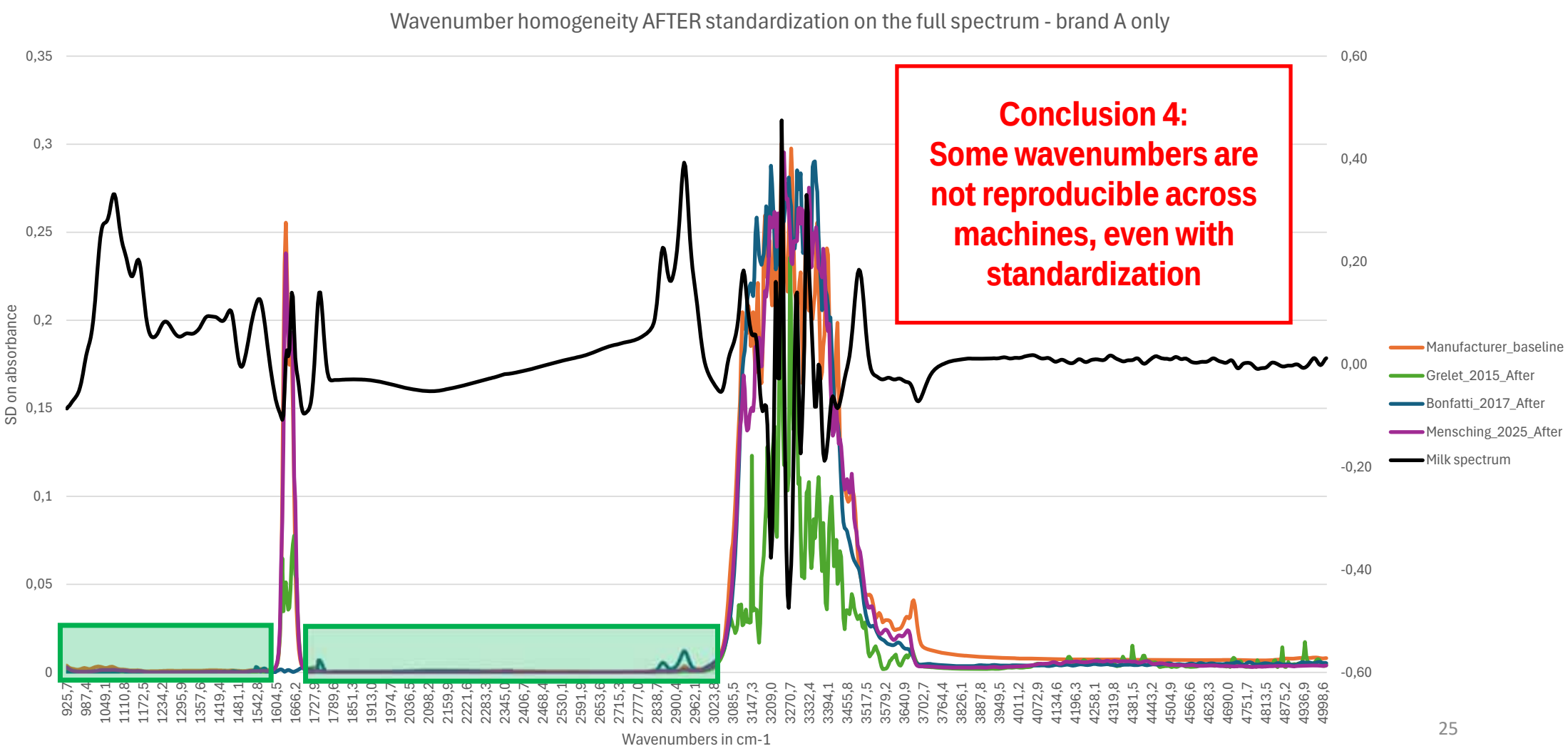
Evaluation on Predictions homogeneity across instruments before and after standardization (within each methodology)

→ Impact on predictions homogeneity across instruments, application of 1 common fat model
(one sample only, sample 5 ~ 4.13 g/100g)





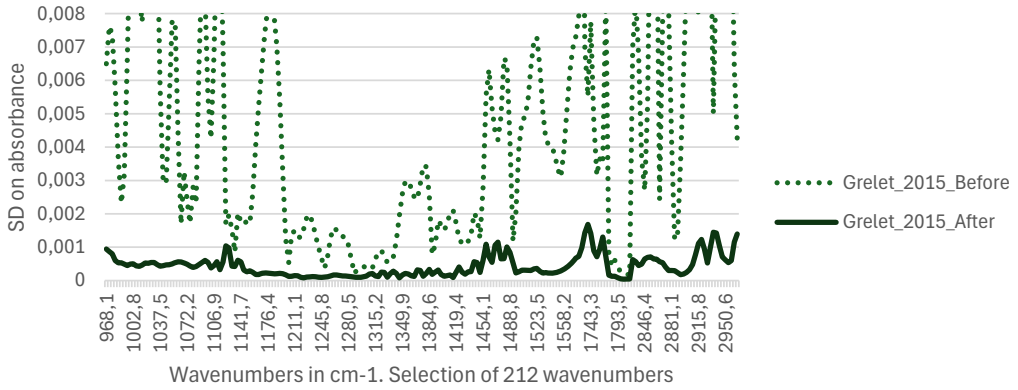
Evaluation on spectral homogeneity across instruments **AFTER** standardization (within each methodology)



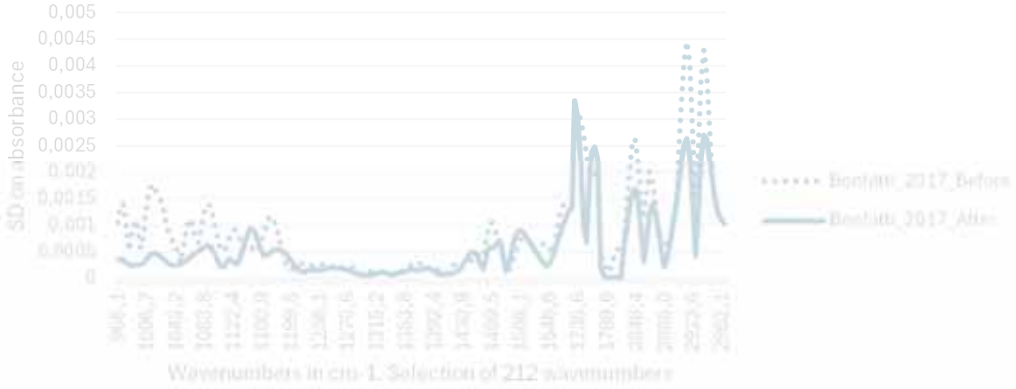
Evaluation on spectral homogeneity across instruments before and after standardization (within each methodology)

Wavenumber homogeneity of
Manufacturer Baseline -brand A

Wavenumber homogeneity BEFORE and AFTER standardization–
Grelet 2015 - all brands



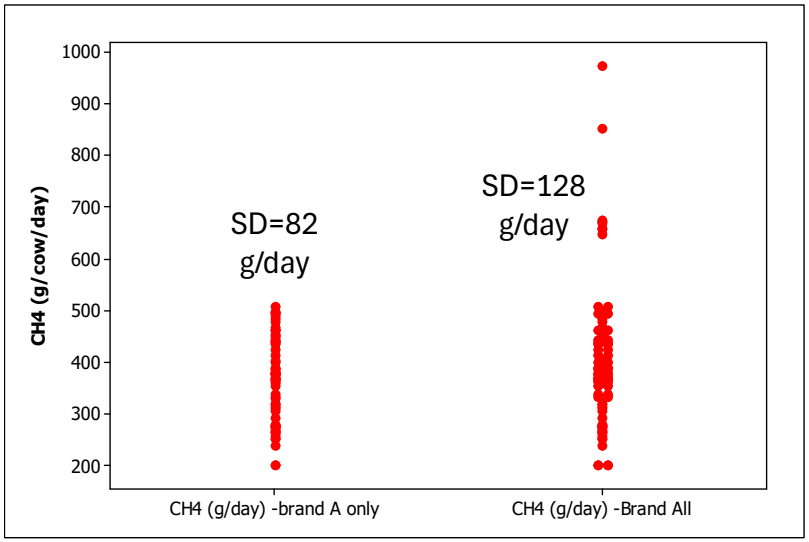
Wavenumber homogeneity BEFORE and AFTER standardization -
Bonfatti 2017 -brand A

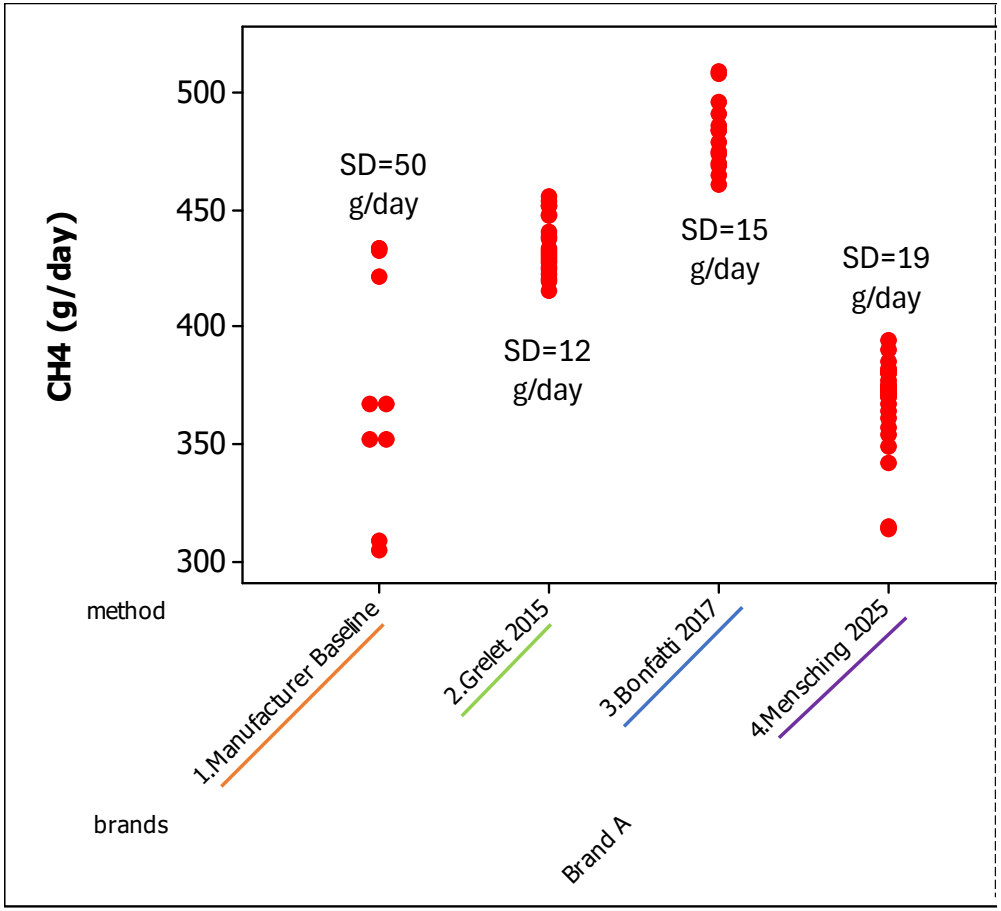


Wavenumber homogeneity BEFORE and AFTER standardization -
Mensching 2025 -brand A

Evaluation on **Predictions homogeneity across instruments** before and after standardization (within each methodology)

→ Impact on predictions homogeneity across instruments, application of 1 common CH4 model
(one sample only, sample 5 ~ 4.13 g/100g)





Evaluation on **Predictions homogeneity across instruments** before and after standardization (within each methodology)

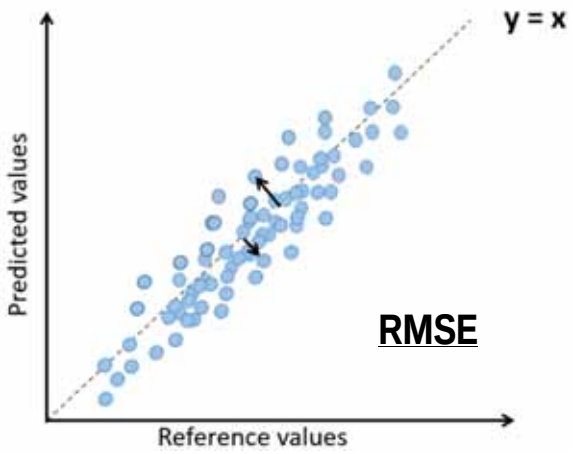
What does it mean in lab real life ?

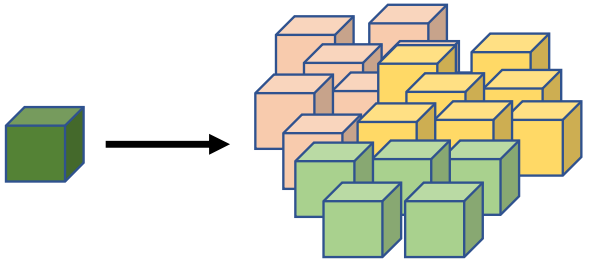
Real total error of prediction when using a MIR model on a spectrometer

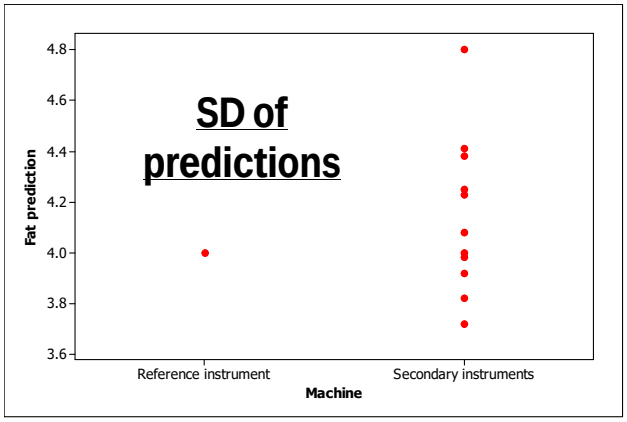
Error of the model itself

+ / &

Error due to the transfer of model on this instrument







Evaluation on Predictions homogeneity across instruments before and after standardization (within each methodology)

Error of the model itself

	Fat (g/100ml)	SAT (g/100ml)	C18_1cis9 (g/100ml)	C8 (g/100ml)	PUFA (g/100ml)	Blood BHB (mmol/L)	Blood Glucose (mmol/L)	Milk BHB (µmol/L)	CH4 (g/day)	N efficiency (%)
R^2cv	1.00	0.99	0.95	0.91	0.77	0.70	0.44	0.61	0.68	0.74
RMSEcv	**0.040**	**0.072**	**0.062**	**0.0040**	**0.0210**	**0.270**	**0.360**	**64.6**	**57.0**	**5%**

Error due to the transfer of model on this instrument

Relative homogeneity, SD compared to model error

	Method	STD	n machines	Fat (g/100ml)	SAT (g/100ml)	C18_1cis9 (g/100ml)	C8 (g/100ml)	PUFA (g/100ml)	Blood BHB (mmol/L)	Blood Glucose (mmol/L)	Milk BHB (µmol/L)	CH4 (g/day)	N efficiency (%)
	Manufacturer Baseline		5	95%	63%	53%	87%	124%	11%	23%	49%	73%	88%
brand A	Grelet et al., 2015	Before	11	132%	109%	138%	76%	166%	34%	28%	28%	115%	104%
	Grelet et al., 2015	After	11	31%	37%	39%	35%	29%	11%	11%	13%	25%	22%
	Bonfatti et al., 2017	Before	7	185%	121%	123%	69%	205%	21%	33%	53%	292%	110%
	Bonfatti et al., 2017	After	7	186%	98%	91%	54%	47%	24%	13%	17%	38%	45%
	Mensching et al., 2025	Before	14	157%	96%	137%	73%	178%	23%	31%	35%	135%	85%
	Mensching et al., 2025	After	14	94%	65%	64%	47%	44%	27%	23%	27%	46%	52%
all brands	Grelet et al., 2015	Before							7%	150%	63%	263%	670%
	Grelet et al., 2015	After							4%	28%	25%	45%	71%

Conclusion 7:
Interaction with models. Standardization is adding less (relative) error to predictions from models with large errors