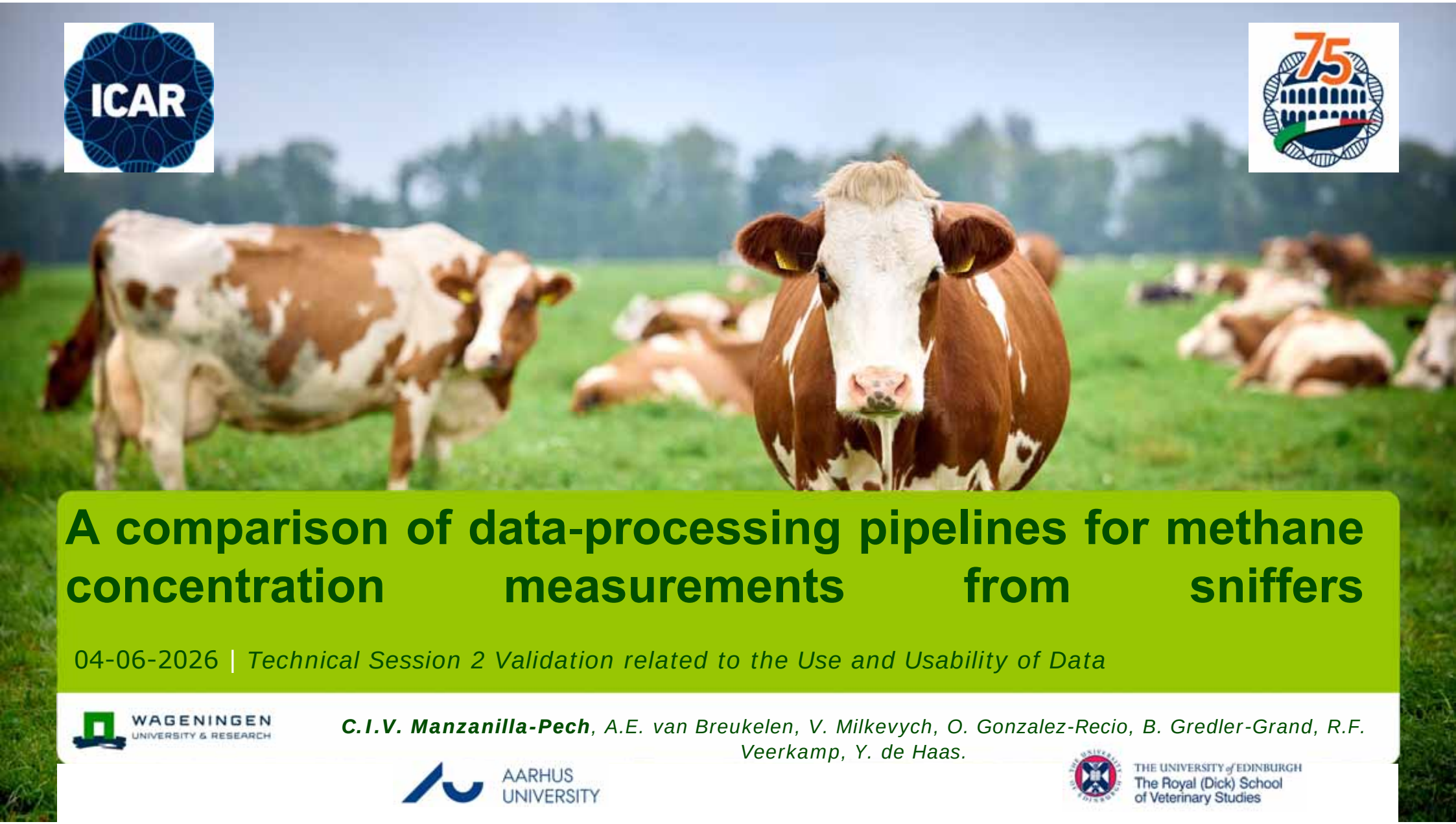


A comparison of data-processing pipelines for methane concentration measurements from sniffers

04-06-2026 | *Technical Session 2 Validation related to the Use and Usability of Data*

***C.I.V. Manzanilla-Pech**, A.E. van Breukelen, V. Milkevych, O. Gonzalez-Recio, B. Gredler-Grand, R.F. Veerkamp, Y. de Haas.*

WAGENINGEN UNIVERSITY & RESEARCH

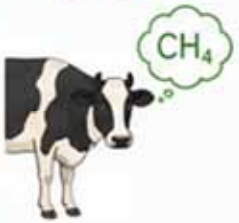
AARHUS UNIVERSITY

THE UNIVERSITY of EDINBURGH
The Royal (Dick) School of Veterinary Studies

Introduction

Why methane phenotyping?

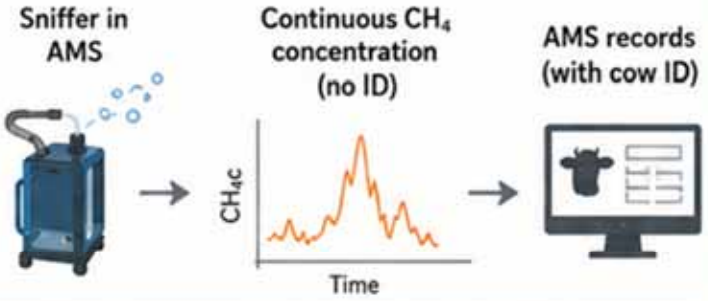
Reducing greenhouse gas emissions is a **global challenge.**



Genetic improvement in dairy cattle can **mitigate methane emissions.**

This requires **large-scale, reliable methane phenotypes.**

How are methane data obtained?



Sniffer data must be synchronized with AMS records to link measurements to individual cows.

Challenges: different time systems (CET vs UTC), clock drift, continuous measurements, and idle periods require precise alignment.

Development of data processing pipelines

Methane phenotyping with sniffers is an emerging field.

Pipelines follow similar steps: **synchronization, cleaning, background correction,** and **phenotype derivation.**

Recent advances include matched filter alignment, fault detection, improved peak detection, and local background estimation.

Comparing and harmonizing processing approaches is essential.

Introduction

Genetic selection requires reliable methane phenotypes

Sniffers are largely used to sample CH_4

- Sniffers are set on AMS
  - Do not have an ID tag reader
    - Rely on the AMS ID reader
    - Need to be align with AMS data to obtain cow ID
    - Both devices have a clock, different time

Sniffer output

Date	Time	Ch4	Co2
01-09-2022	00:00:02	236.0	5293.0
01-09-2022	00:00:07	277.0	5676.0
01-09-2022	00:00:13	269.0	4833.0
01-09-2022	00:00:19	183.0	3305.0
01-09-2022	00:00:25	64.0	2144.0
01-09-2022	00:00:30	64.0	1982.0
01-09-2022	00:00:36	7.0	2728.0
01-09-2022	00:00:42	52.0	3198.0
01-09-2022	00:00:48	52.0	3477.0
01-09-2022	00:00:53	97.0	3331.0
01-09-2022	00:00:59	89.0	2325.0
01-09-2022	00:01:05	89.0	1084.0
01-09-2022	00:01:11	0.0	899.0

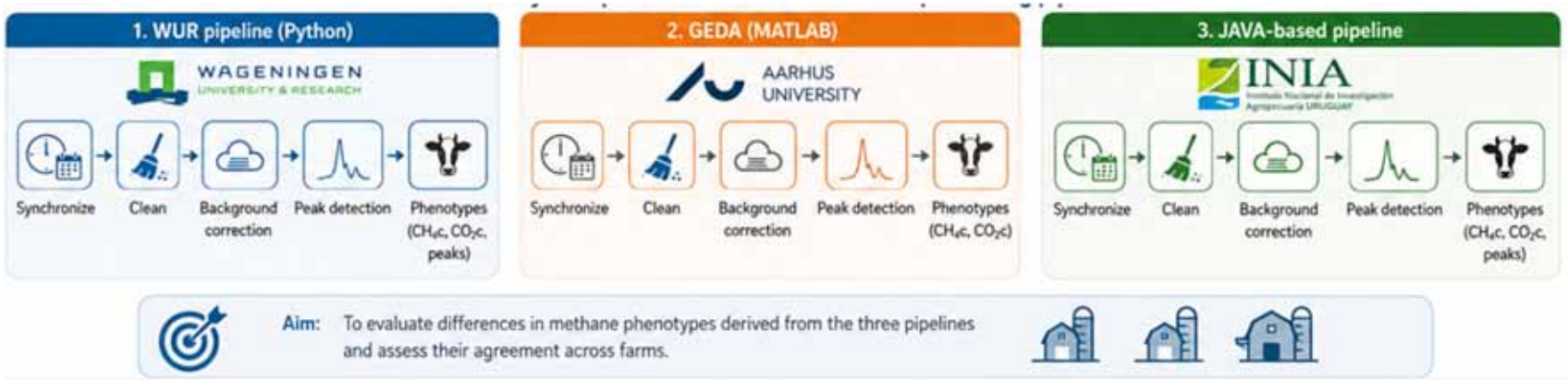
AMS file

ID	Start	End	Robot	UBN
1045	2022-09-01 00:01:00	2022-09-01 00:07:00	1	40
14112	2022-09-01 00:08:00	2022-09-01 00:12:00	1	4
6972	2022-09-01 00:08:00	2022-09-01 00:14:00	2	40
8340	2022-09-01 00:15:00	2022-09-01 00:20:00	2	40
6202	2022-09-01 00:16:00	2022-09-01 00:21:00	1	40
14408	2022-09-01 00:22:00	2022-09-01 00:26:00	1	4
6745	2022-09-01 00:22:00	2022-09-01 00:28:00	2	40
14982	2022-09-01 00:28:00	2022-09-01 00:32:00	1	4
14096	2022-09-01 00:29:00	2022-09-01 00:34:00	2	4
14992	2022-09-01 00:35:00	2022-09-01 00:40:00	2	4
9019	2022-09-01 00:37:00	2022-09-01 00:42:00	1	40

Objectives

a) Compare three different pipelines for methane data processing

b) Evaluate differences in methane phenotype derived from these pipelines and assess their agreement across farms.



Three pipelines

WUR

- In-house self written script in Phyton (van Breukelen, ABG-WUR)

GEDA

- Software developed (Milkevych, QGG-AU) and written in MATLAB

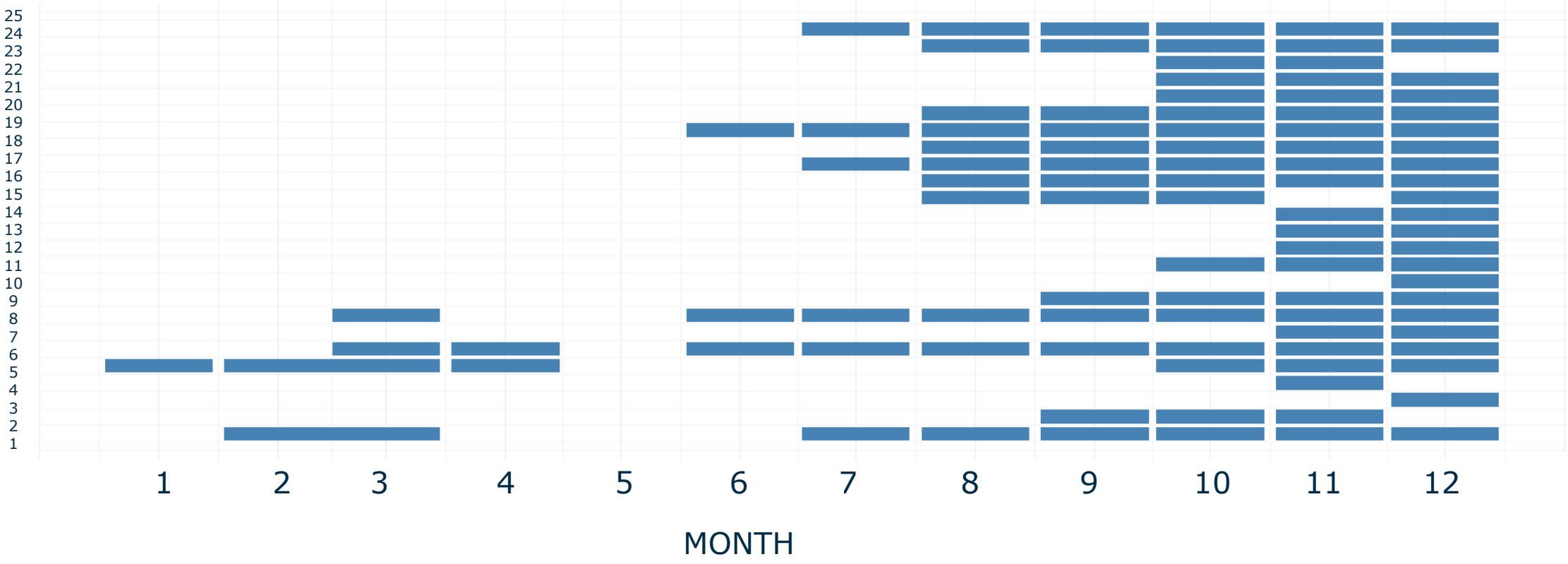
JAVA

- In-house self written script in Java (Gonzalez-Recio, INIA)

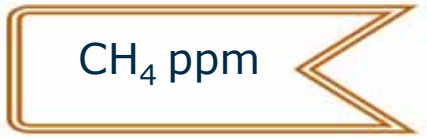
Main Specifications per Pipeline

Variable	WUR	GEDA	JAVA
Requirements	Duration ≥ 150 s	Set in program (default = 2 min)	Duration > 1 min
Average	60–300 s	Entire visit	≥ 5 min
Lag time	Assumed within first 60 s	0.1 min	Defined per farm during program execution
Background	Default: no background correction[1]	Estimated using idle period	Based on 5 lowest records
Concentration in input	ppm	ppm	As decimal fraction (internally multiply by 10,000)
Concentration in output	ppm	ppm	ppm
Time in input	AMS and Sniffer may have different time zones, which is specified manually	AMS and Sniffer may have different time zones and automatically detected	Requires AMS and Sniffer timestamps aligned
Time in output	Sniffer time format; start time of visit reported	AMS time format	Time format as provided in both input files minus the lag time.
Main phenotypes	CH4c, CO2c, eructation peaks phenotypes	CH4c, CO2c	CH4c, CO2c, eructation peaks phenotypes

Data Available: 25 farms, year 2022



Outputs per pipeline per visit



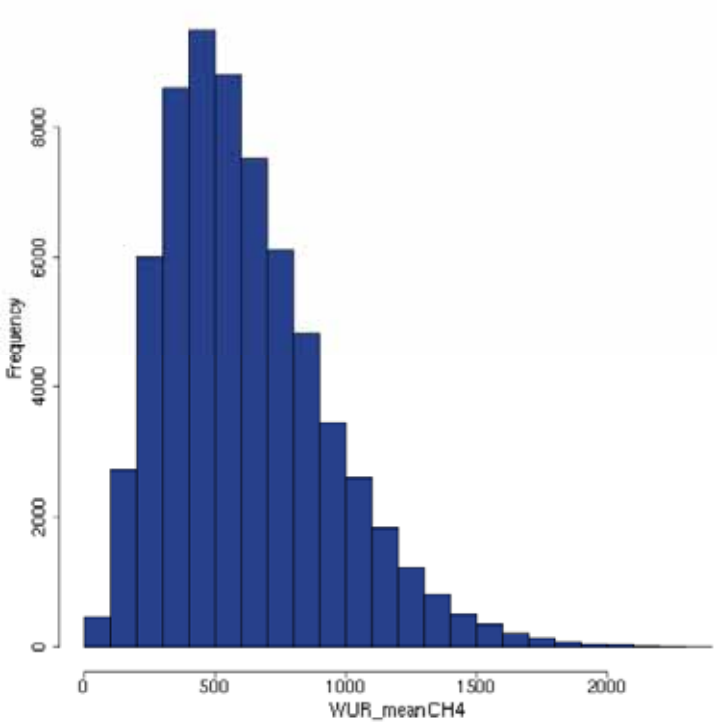
WUR	GEDA	JAVA
25	25	25
2,745	2,636	2,778
262,111	145,760	385,901

Alignment

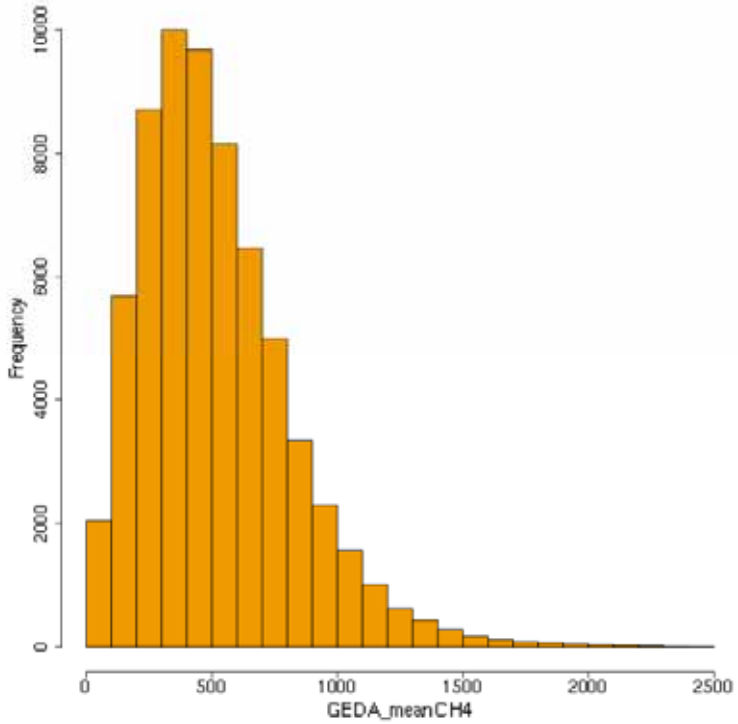
CH4 concentration (ppm) visit level

- Extract hour from the date-time of the output
  - Merge by windows 5, 10, 15 min
  - Merge by CowID, Date, Hour
  - Keep only the complete cases, phenotype for all 3 pipelines

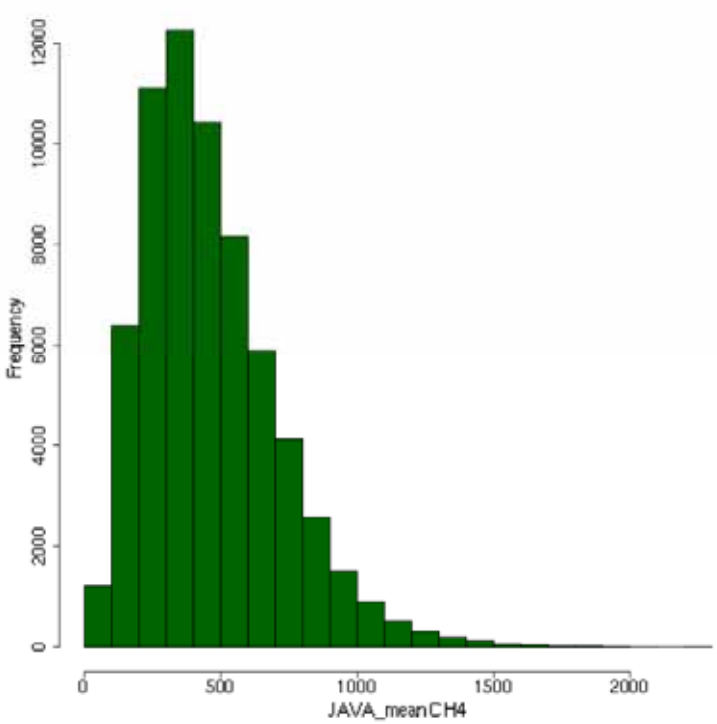
Histograms per phenotype



M= 618
SD=317

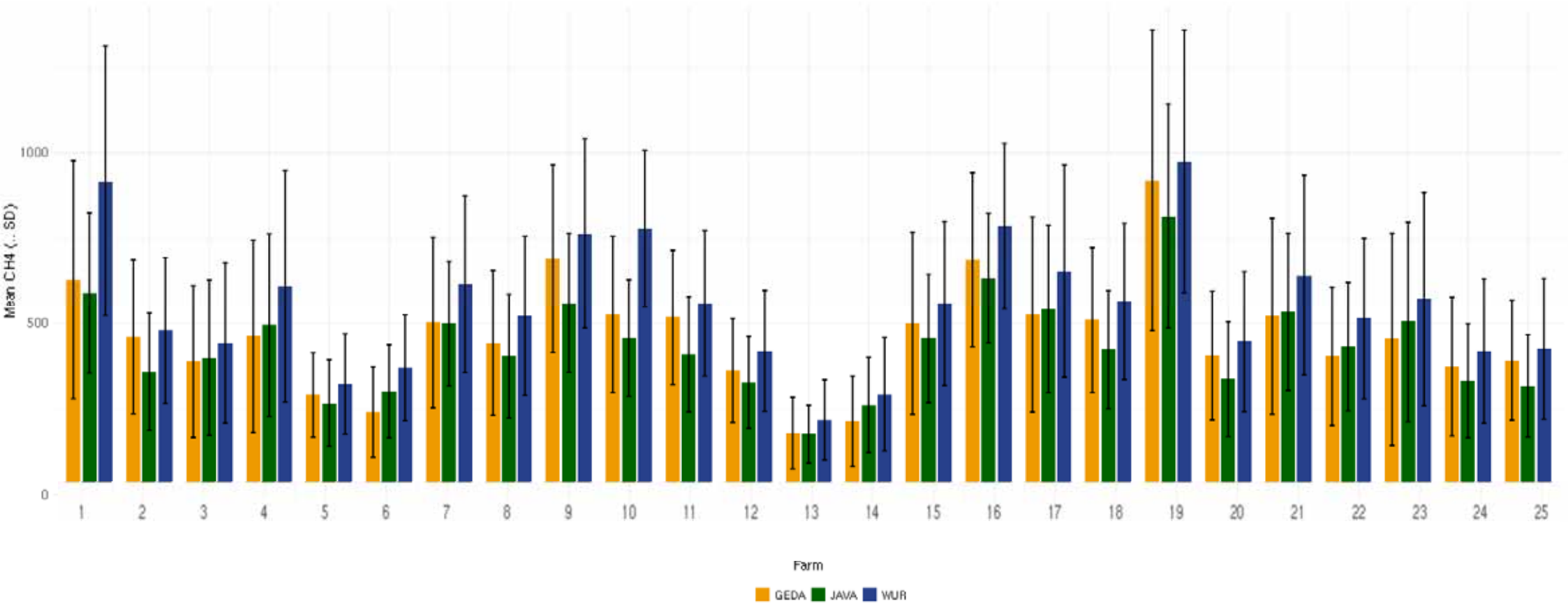


M= 521
SD=304



M= 463
SD=250

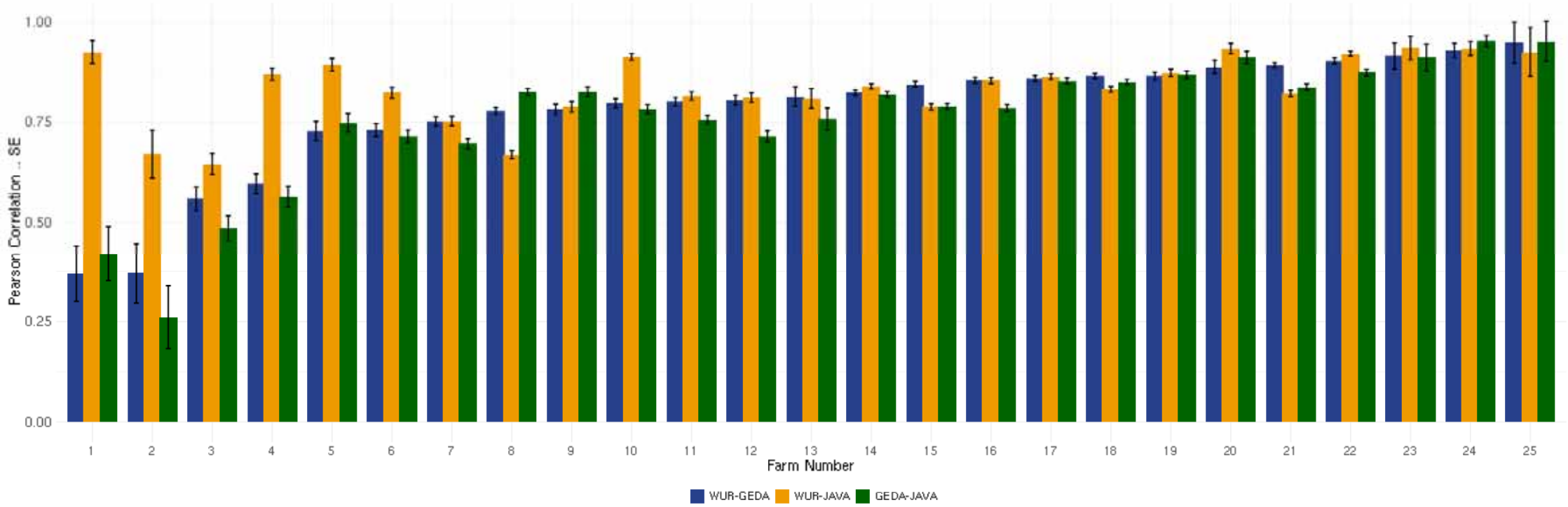
Average per Farm



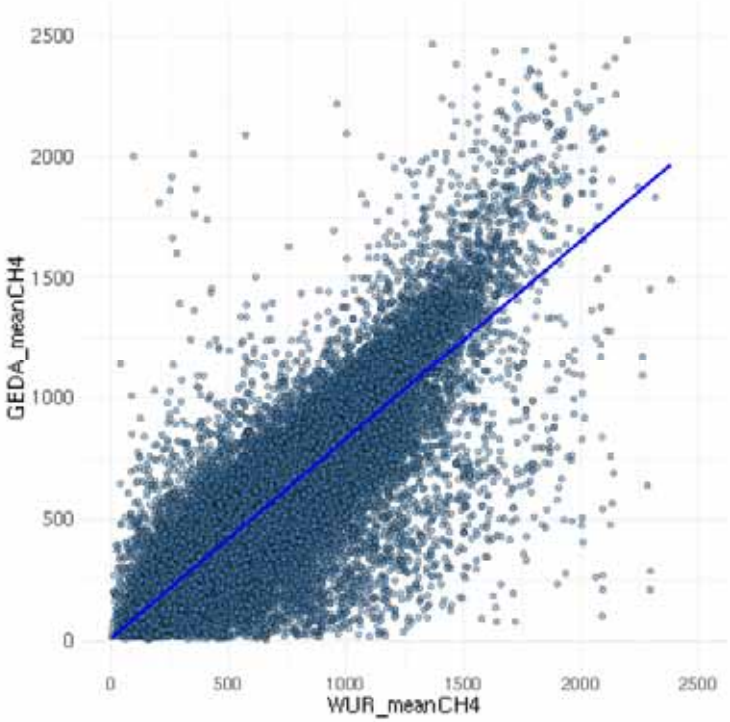
Overall Pearson and Spearman Correlations

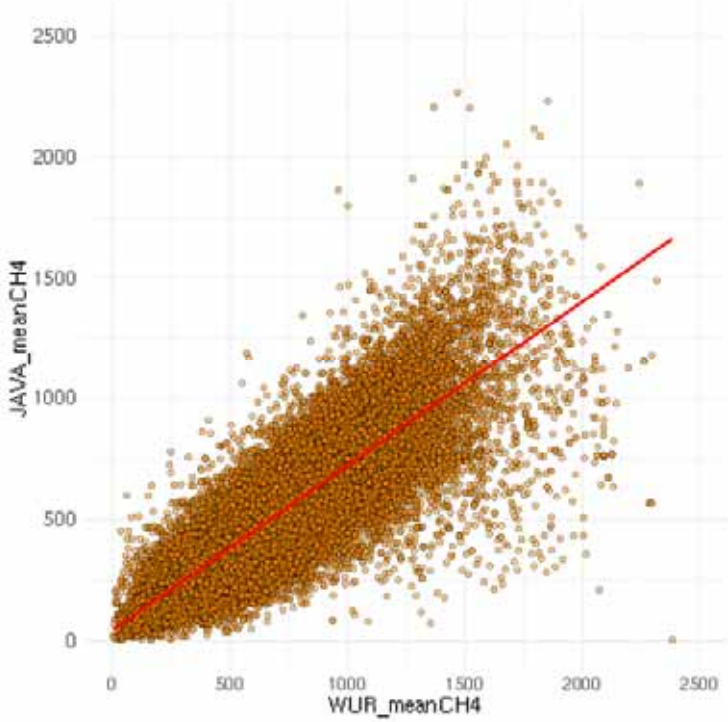
WUR - GEDA	WUR - JAVA	GEDA - JAVA
0.86 ***P***	0.86 ***P***	0.85 ***P***
0.86 ***S***	0.87 ***S***	0.85 ***S***
	SE = 0.001	

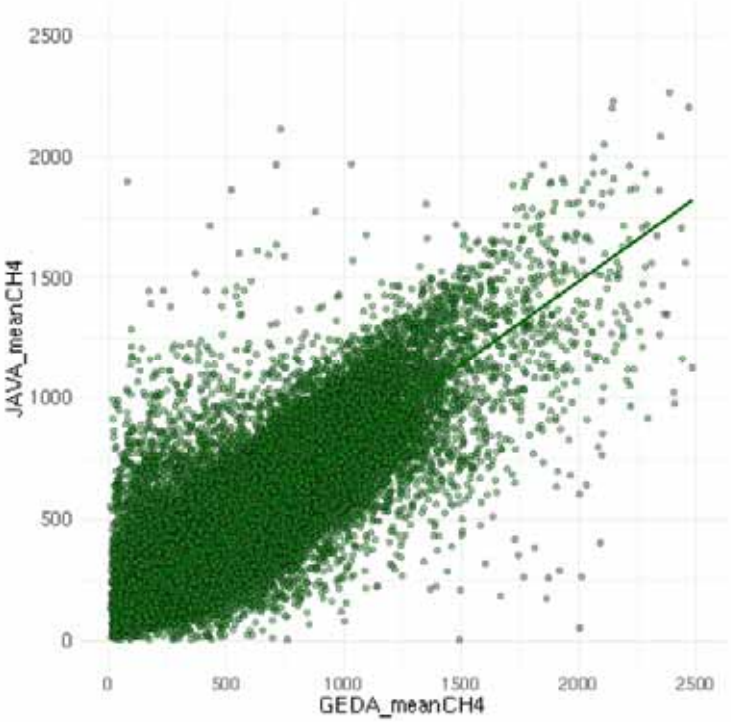
Pearson Correlations per Farm



Pairwise Scatter Plots and Correlations of CH_4 Measurements

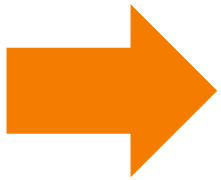






Extra test: Mimicking the background of JAVA

WUR: Uncorrected phenotype



WUR background:5 lowest measurements

WUR - JAVA

WURbc - JAVA

0.86

M= 618
SD=317

0.94

M= 501
SD=272

Extra test: Mimicking the background of JAVA

WUR: Uncorrected phenotype

WUR background:5 lowest measurements

WUR - JAVA

WUR - JAVA

0.86

0.94

Background Comparison
Poster 2286327
Thursday 5:30-6:15 PM

Preliminary Results

Genetic correlations

WUR - GEDA

0.99

Further genetic comparison results

Session 1 "Breeding for reduced methane, feed efficiency, resiliency and its potential consequences"

Take home messages

The three pipelines showed strong agreement in methane phenotyping, with high correlations indicating that relative ranking of cows and farms is similar despite methodological differences.

Differences in final phenotype can be due to background correction, lag-time handling, and averaging procedures.

Acknowledgements

Global Methane Genetics initiative
Led by WAGENINGEN UNIVERSITY & RESEARCH
In Partnership with
Ministerie van Landbouw, Natuur en Voedselkwaliteit

coralia.manzanillapech@wur.nl