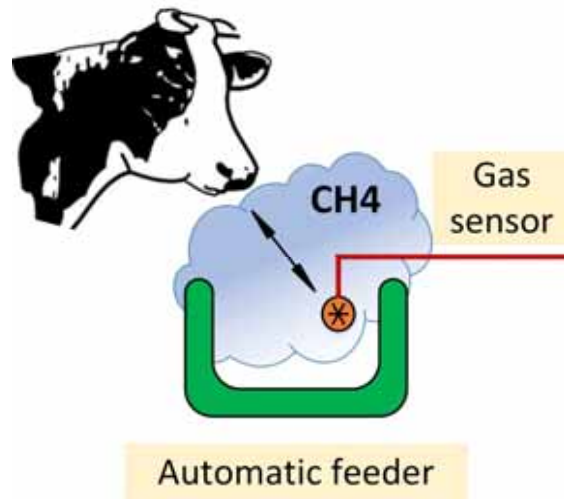


The unified framework for processing and storing large-scale sniffers-based emission data

ICAR 2026, Verona, June 4, 2026

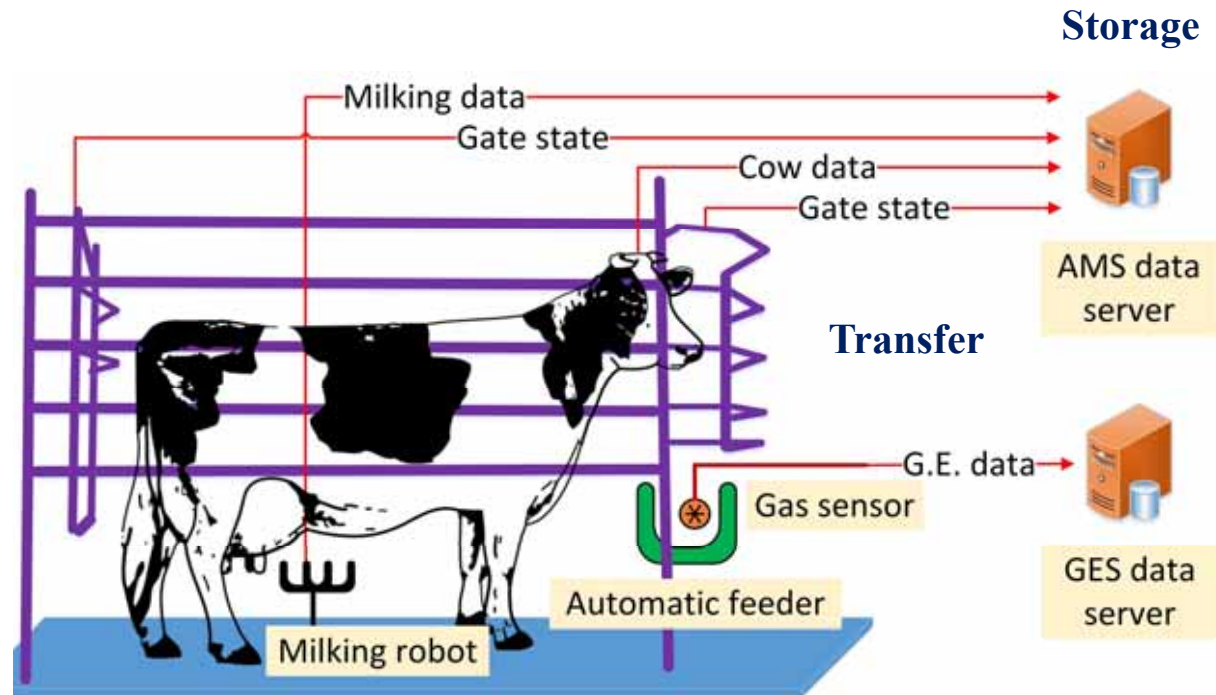
Measurement conditions



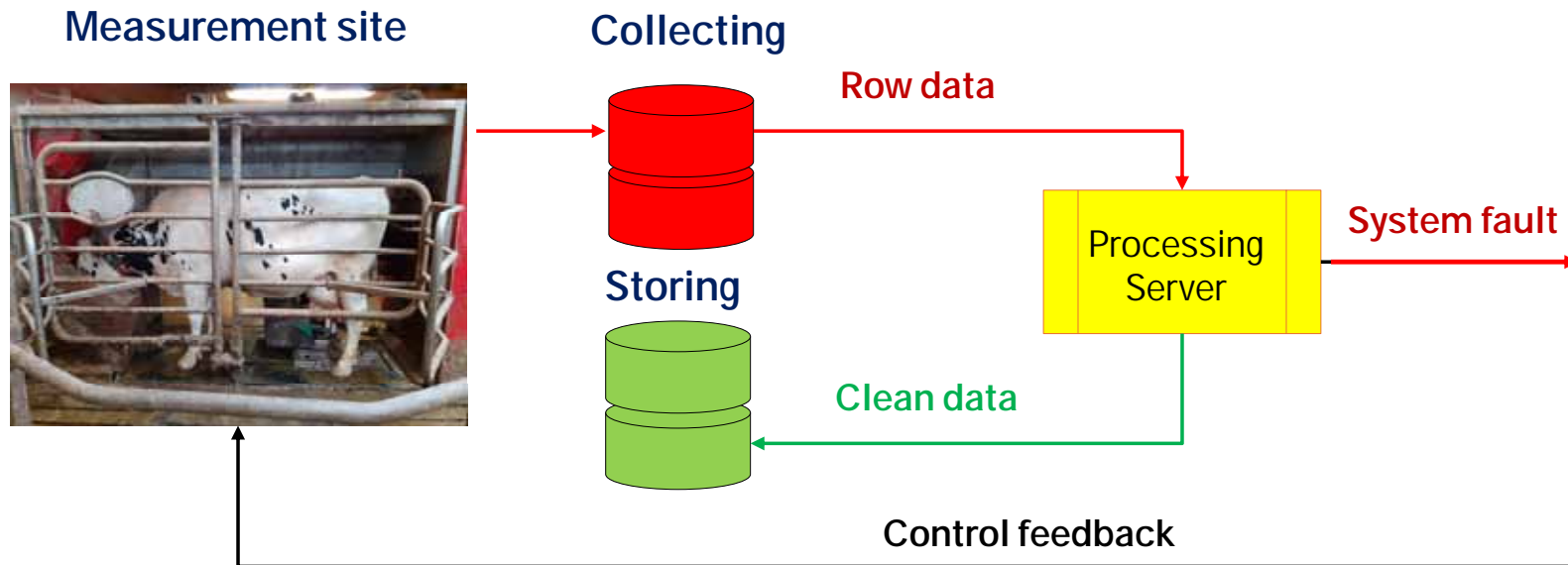
Collected CH₄ & CO₂:

- corrupted by noise,
- unsuitable for genetic studies as it is.

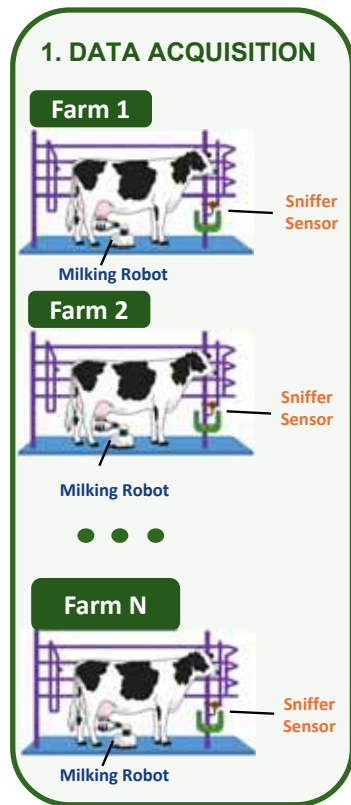
Automation of data transfer & storage



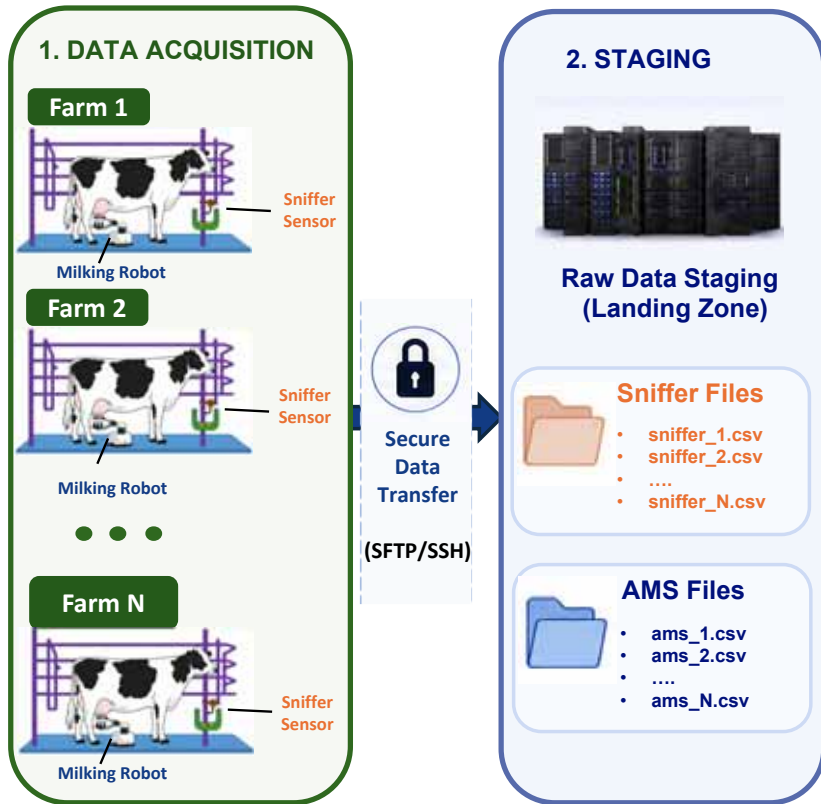
Online monitoring & faults detection



Continuous data integration: Architecture



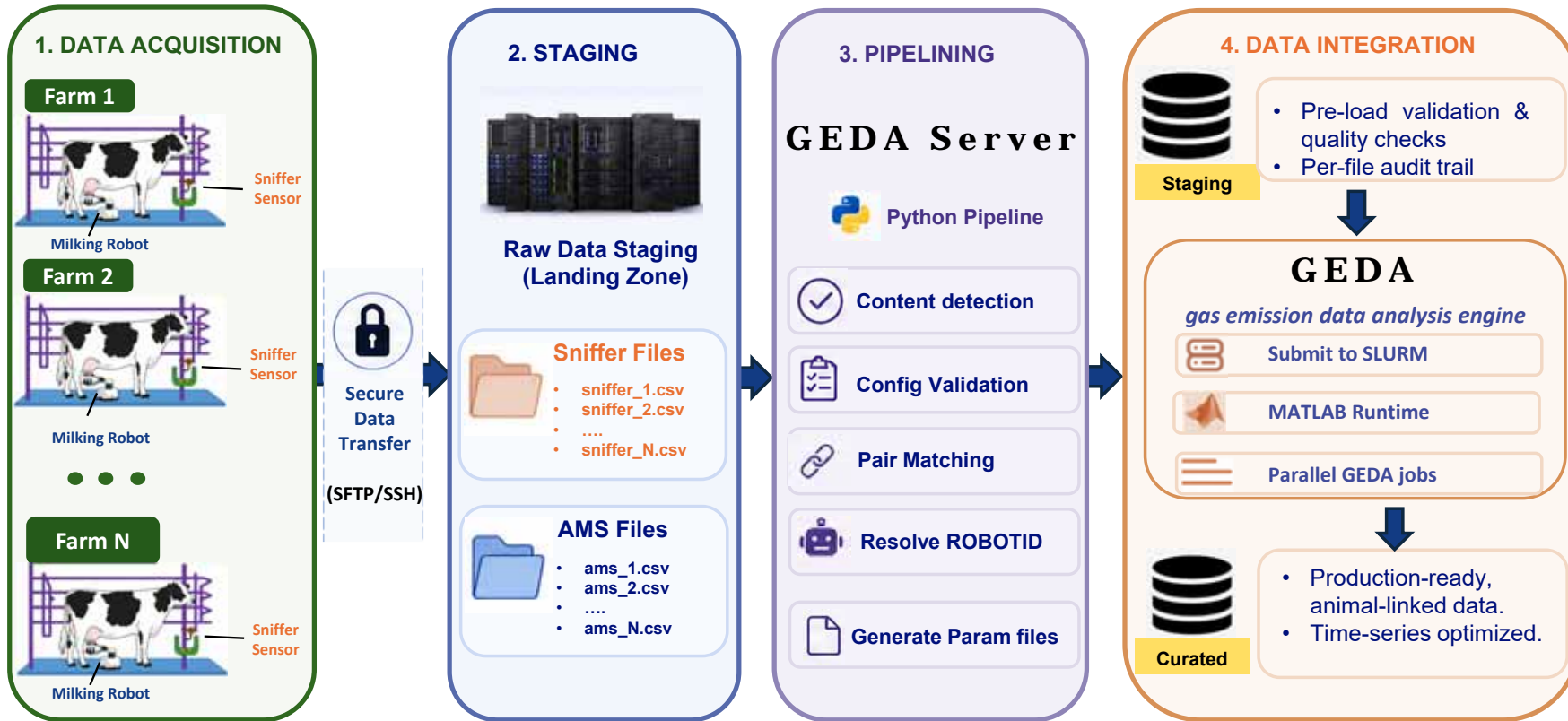
Architecture



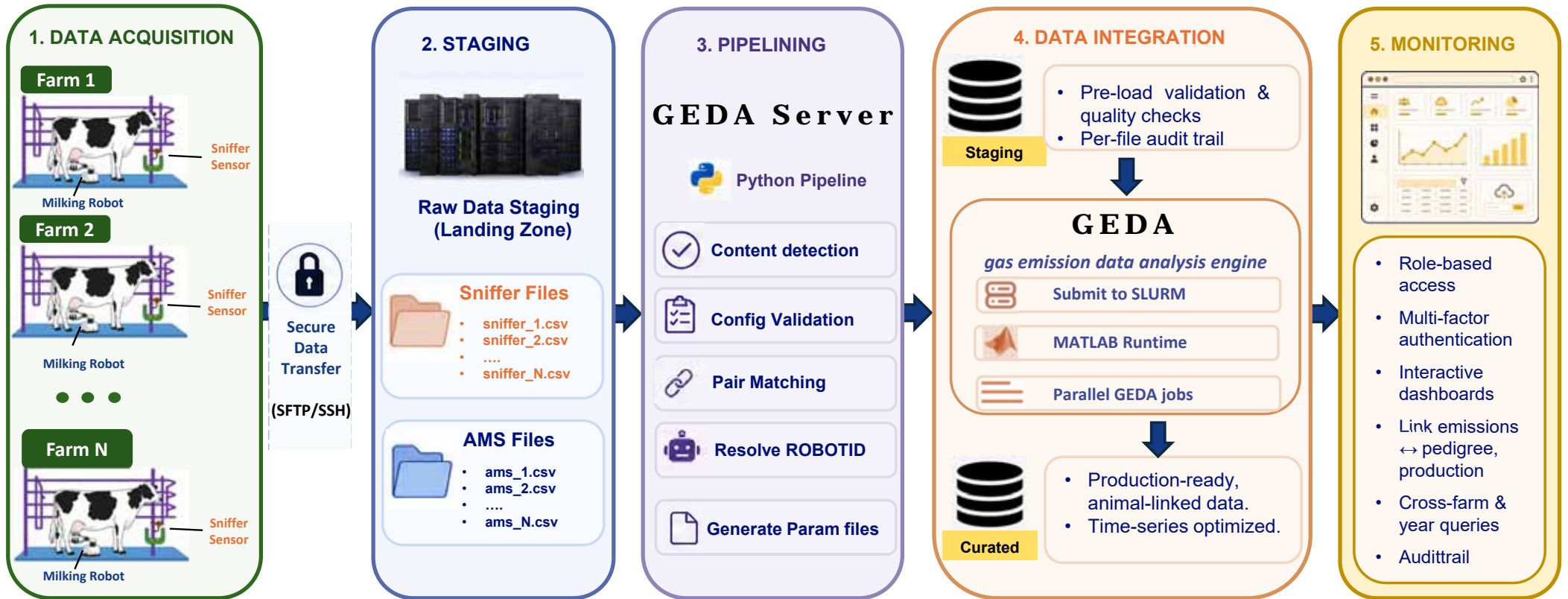
Architecture



Architecture



Architecture



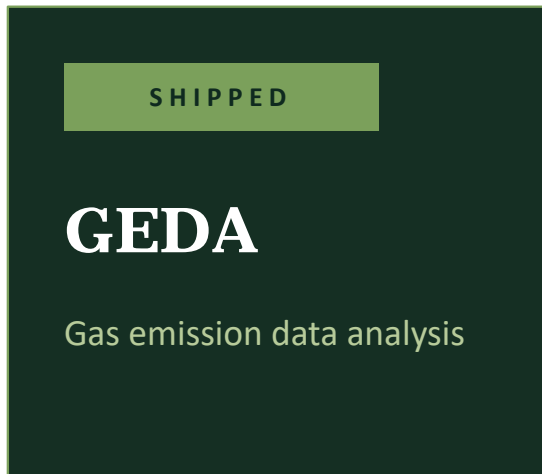
Software roadmap

Three steps, the same problem solved more completely each time.

Each builds on the one before. Effort → Access → Collaboration.

Software roadmap

Three steps, the same problem solved more completely each time.



SHIPPED

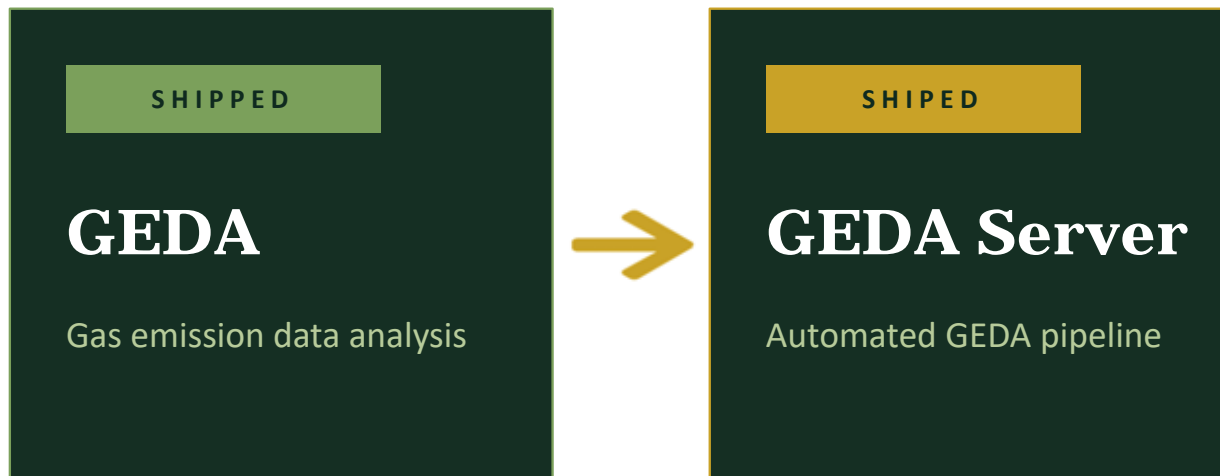
GEDA

Gas emission data analysis

Each builds on the one before. Effort → Access → Collaboration.

Software roadmap

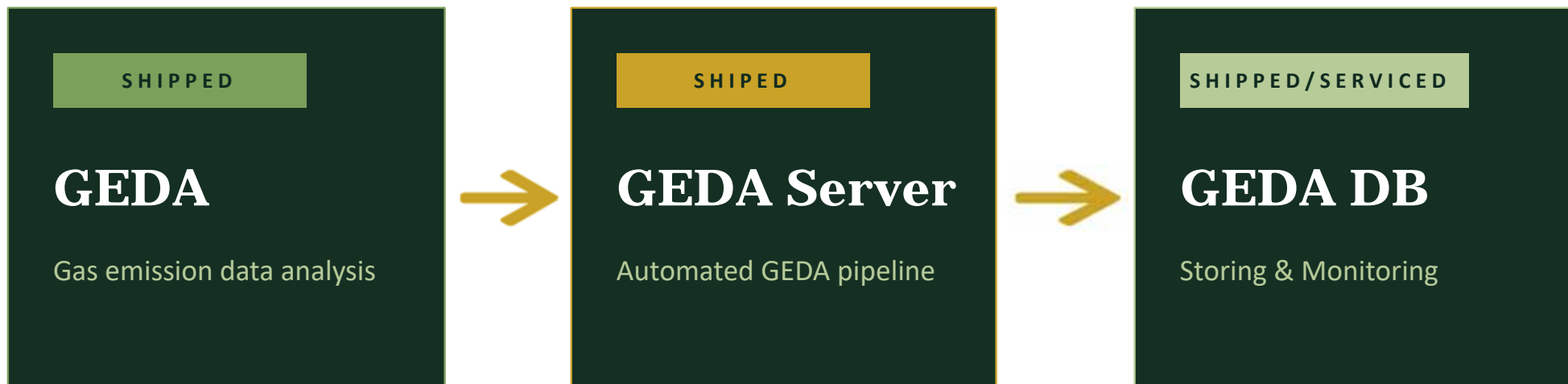
Three steps, the same problem solved more completely each time.



Each builds on the one before. Effort → Access → Collaboration.

Software roadmap

Three steps, the same problem solved more completely each time.



Each builds on the one before. Effort → Access → Collaboration.

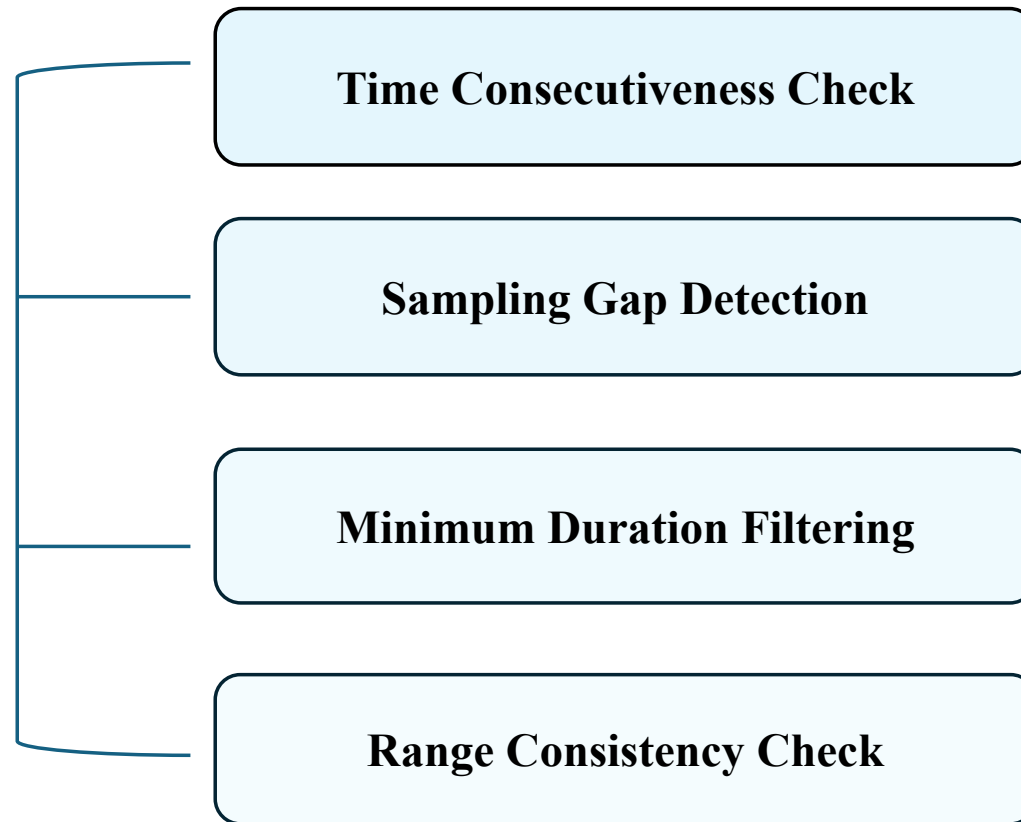
GEDA: Gas Emission Data Analysis

Allows an automated processing a sniffers data.

The program provides:

- Initial quality control and filtering.
- Data alignment.
- Detection and filtering off an unreliable data.
- Phenotypes estimation.
- Quality assessment of estimated phenotypes.
- Comprehensive reporting.

Preprocessing



AMS & Sniffer data alignment + Faults detection

1: Split the gas emission signal to the smaller sub-signals

Each sub-signal will be aligned independently



2: Time skew estimation

$$v[l] = \sum_{i=1}^m c[i]g[i+l]$$

Cross Correlation



$$\tau_* = \arg \max_l v(l)$$



$$\tau = t_{g_l} - t_{c_1}$$

Time Skew



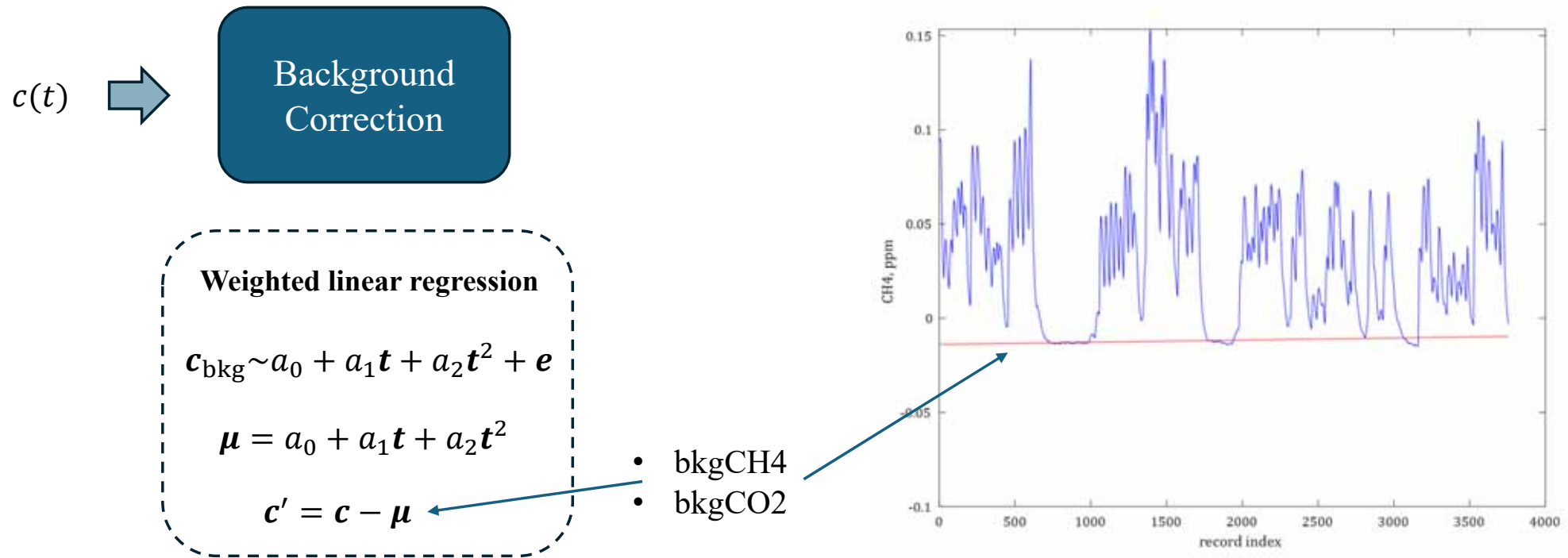
$$\Delta T = 1 \dots 24 h$$



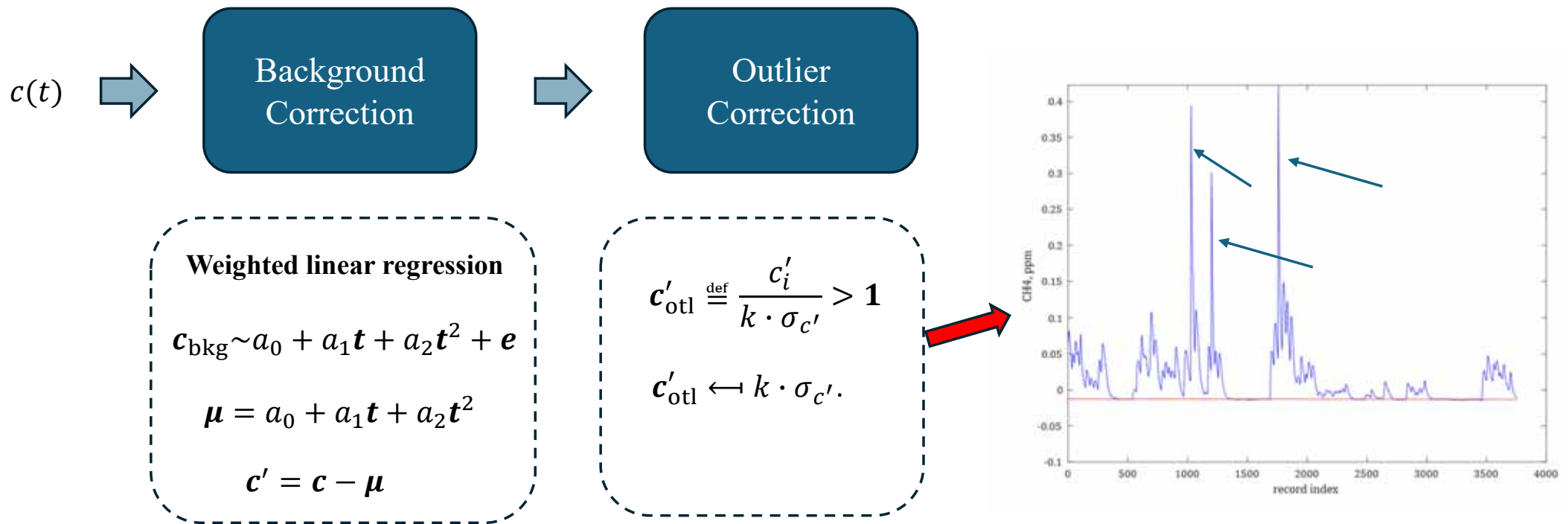
\$SIGLENGTH
6

\$SIGLENGTH
-6

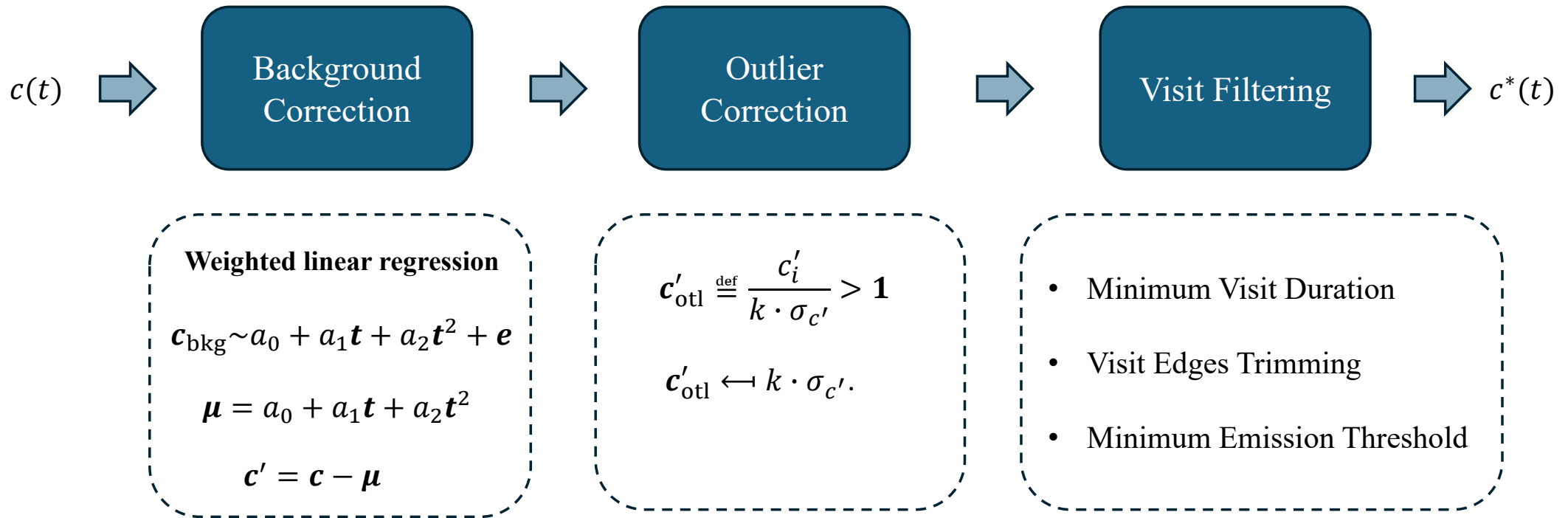
Filtering Aligned Signals



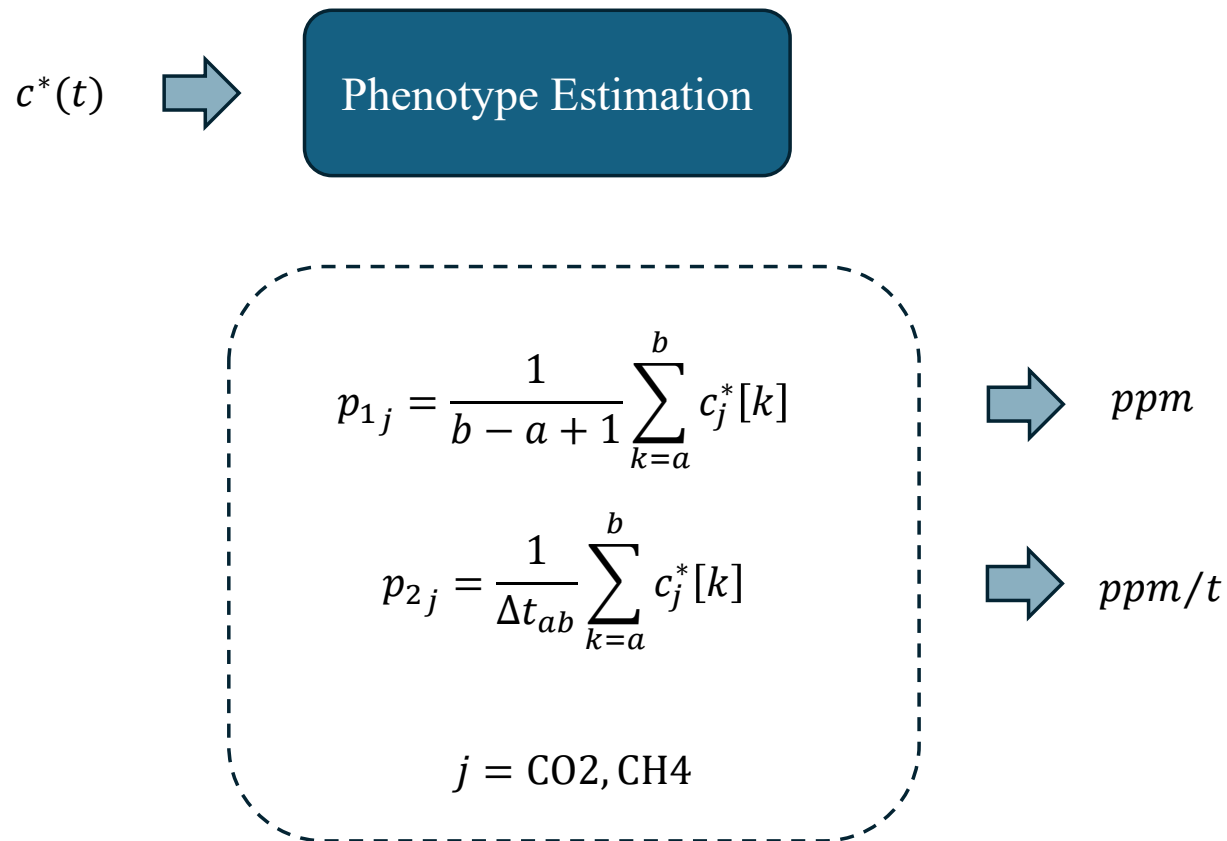
Filtering Aligned Signals



Filtering Aligned Signals



Phenotypes Estimation



Phenotypes Estimation

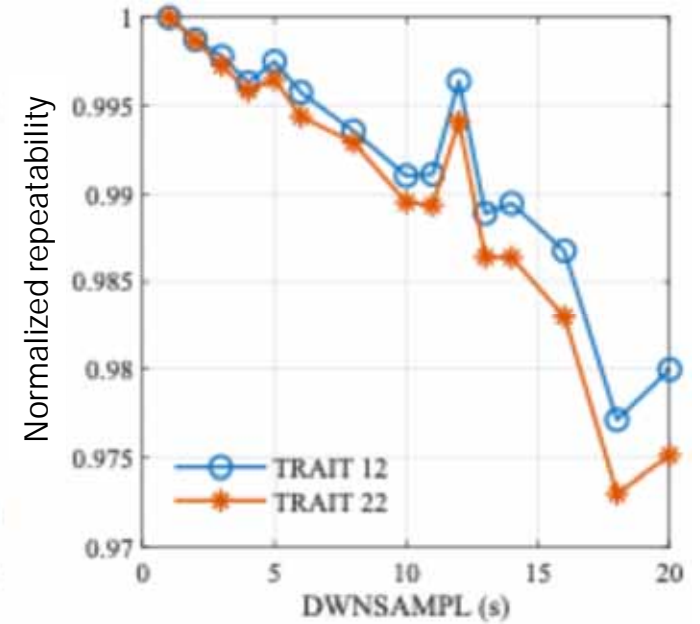
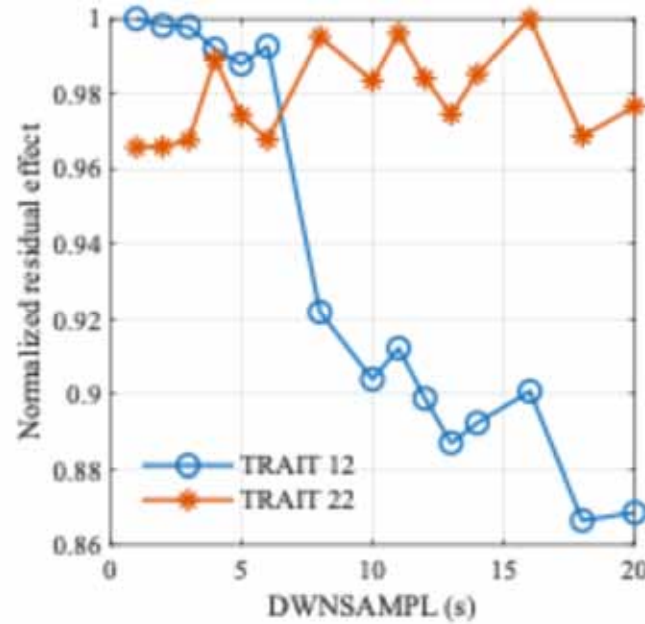
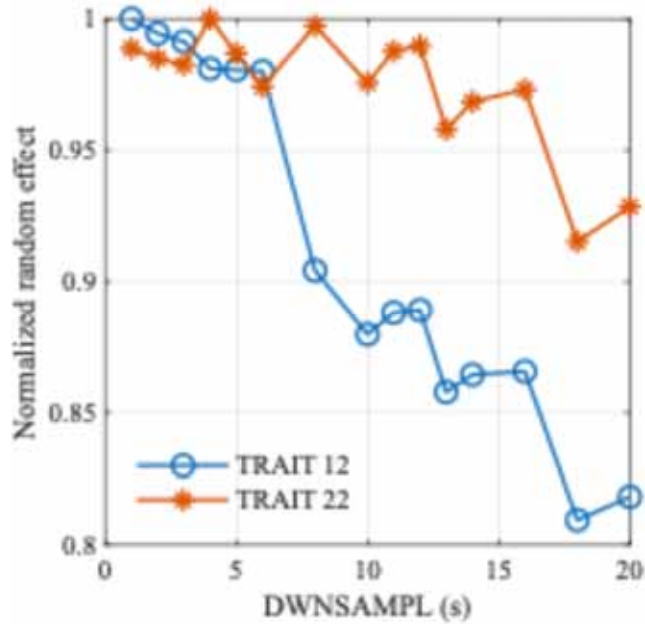
$c^*(t)$ →

Phenotype Estimation

CH4

TRAIT 12 => ppm

TRAIT 22 => ppm/min



Processed Data Quality Estimation



Linear mixed model:

$$\mathbf{p} = \boldsymbol{\theta} + \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

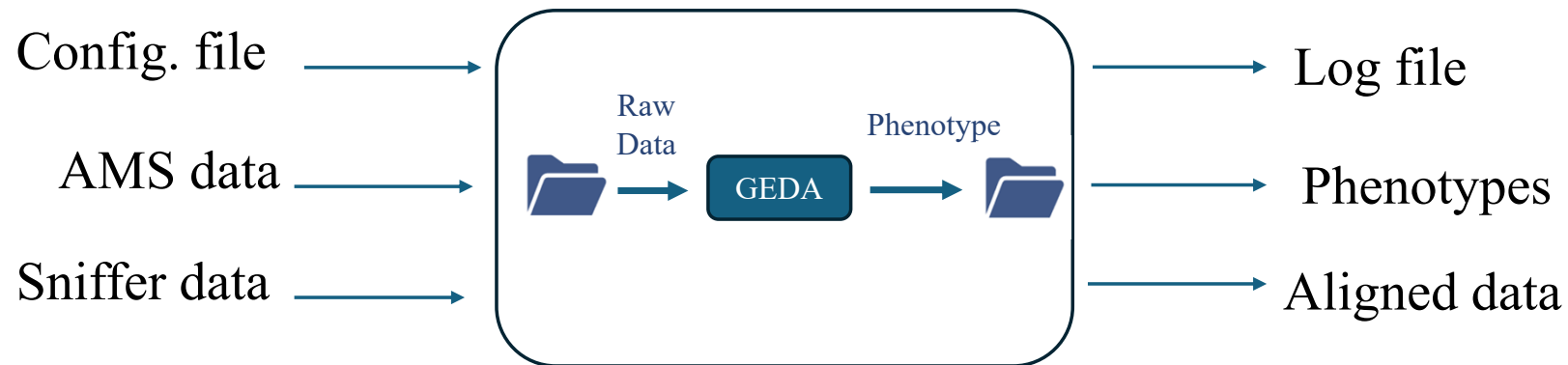
Intraclass correlation coefficient
(local quality metric)

$$r = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

From the GEDA log file:

```
CH4 model: CH4 ~ 1 + Date + bkgCH4 + hour + (1|CowID)
num. observations per ID:      3.018
fit quality, adjusted r^2:    0.57438
random effect variance (var_id): 0.00019982
residual variance (var_err):  0.00018711
var_id/(var_id+var_err):     0.51642
```

What is needed to run GEDA?



The manual GEDA workflow — per farm, per sniffer:

- 1 Find the right robot and sniffer files
- 2 Write a configuration file
- 3 Submit the job and wait
- 4 Collect the log & results, read metrics, analyze
- 5 Repeat — every pair, every farm, every month

GEDA Server - a service pipelining GEDA

Launches GEDA on a schedule

Edit one file. The server does the rest.

- ✓ Scans farm data folders
- ✓ Pairs each robot file with its sniffer file
- ✓ Resolves robot and sniffer hardware IDs
- ✓ Writes a GEDA parameter file for every pair
- ✓ Submits jobs and collects & analyzes every log

Only edit one file: config.py

```
gedaservice-edit-config # opens config.py in your editor
```

FIVE REQUIRED PATHS

BASE_DIR

RAW_DATA_DIR

PATH_SAVE

GEDA_BINARY_PATH

MATLAB_RUNTIME_PATH

ONE REQUIRED NAME

TRAIT_NAME

Short identifier with no spaces (e.g. "trait_11").

Names your tracking files and result folders.

Everything else has a sensible default. Full reference in the researcher guide.

Run modes

"normal "

DEFAULT

Skip processed pairs

The default. Safe to repeat — fast when nothing new.

"update"

Also, re-check n-times

Reprocess pairs whose source CSV was modified.

"full "

CAUTION

Ignore tracking CSV

Reprocess everything found. Pair with TARGET_FARMS.







Combine with TARGET_FARMS to scope a run

```
TARGET_FARMS = ["17895"] # one farm only  
MODE = "full" # reprocess from scratch
```



Bare MODE = "full" reprocesses everything. Pair it with TARGET_FARMS.

All results under a PATH_SAVE

 log_summary.txt	Cumulative scientific results	SCIENCE
 run_report_<ts>.txt	Per-run audit report — config, decisions, failures	AUDIT
 combined_geda/* .txt	Per-bucket pooled phenotype files	
 pipeline_run.log	Execution log — first stop for debugging	DEBUG
 pair_warnings/* .txt	Unmatched pairs, ID fallbacks	
 <herd>/sniffer_R*/	Per-farm folders: param files + raw GEDA+ output	

Deployment environment

GEDA Server deploys on whatever infrastructure a user have.



HPC cluster

SLURM auto-detected ·
jobs via sbatch



Standalone Linux

Parallel local runner · no
cluster



Cloud

AWS · Azure · GCP · private
cloud



Windows / macOS

Docker image · no MATLAB
Runtime setup



Drive from any OS

Windows · macOS · Linux



Your data stays on your infrastructure in every case.

GEDA Database Integration

Results stop being scattered files. They live in one shared, queryable place.



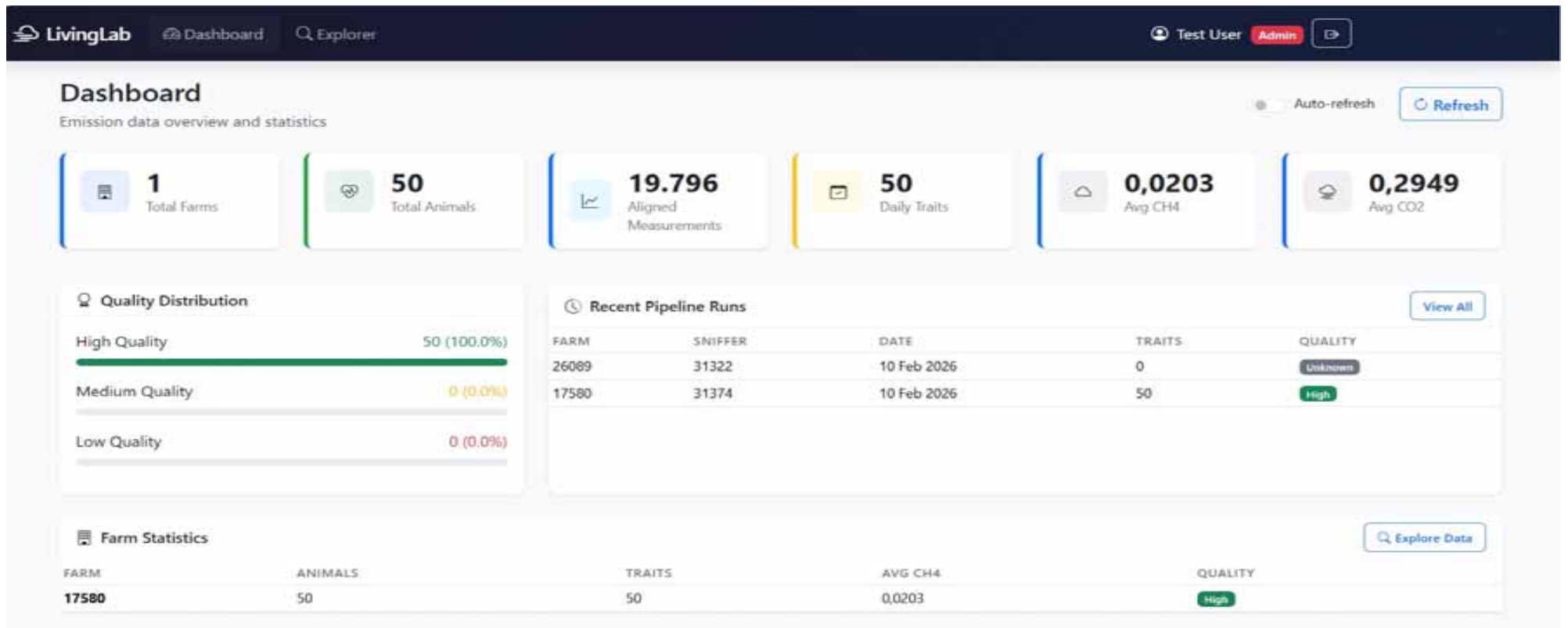
GEDA DB Integration

Emissions + pedigree + production + genotypes + MIR

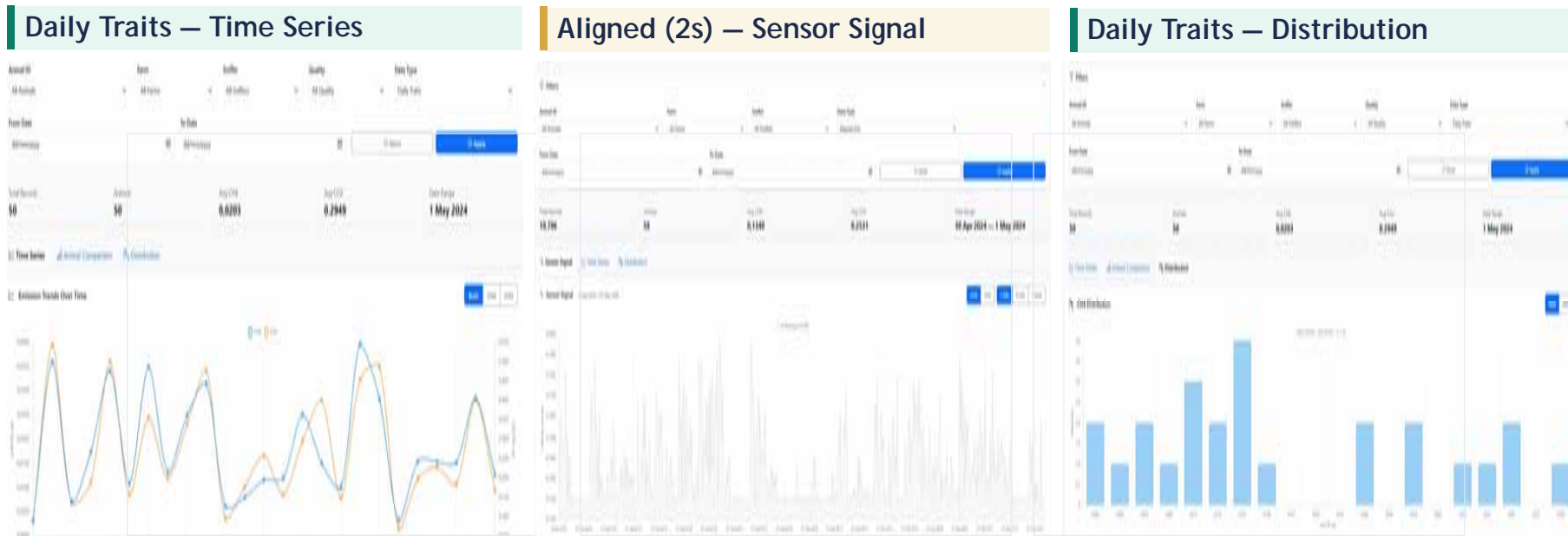
- ✓ **Multi-user, multi-institution** · the right people see the right data
- ✓ **Cross-herd, country, year** · comparisons in one query
- ✓ **Linked to the animal** · pedigree, production, genotypes, MIR
- ✓ **Programmatic access** · straight into R / Python pipelines
- ✓ **Full history and traceability** · every change recorded, every run reproducible

A platform for collaborative dairy-science research, not a folder of files.

Database dashboard



Database data explorer



- ✓ Switch between 3 data types with a single dropdown
- ✓ Dual-axis CH4/CO2 charts with min-max range bands
- ✓ Multiselect filters: animal, farm, sniffer, date, quality
- ✓ CSV export up to 100,000 records

Summary

The GEDA-based framework

- ❑ Provides unique automatization solution of entire sniffer's system.
- ❑ Allows continuous large-scale data integration.

Authors

Leila Ashayeri



Rasmus B. Stephansen



Trine M. Villumsen



Martin Bjerring



Muhammad Javed



Viktor Milkevych



Contact:

vimi@qgg.au.dk