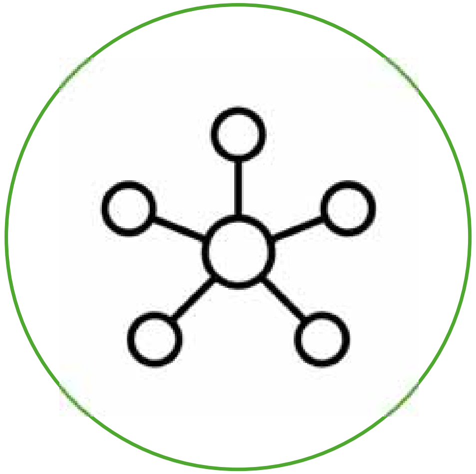
Modular Open-Source Framework for Spectral Phenotyping Using Neighbourhood-Adaptive PLSR: Application to Nitrogen Use Efficiency in a Dairy Herd

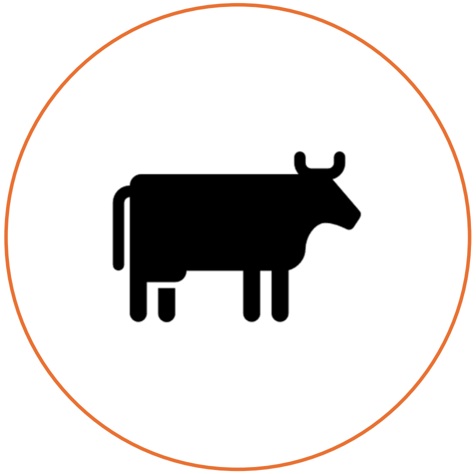
Ross Muers, PhD

SRUC, UK

TS5: From Milk Analysis to Decision Support Unlocking Insights for Sustainable Dairy management

ICAR 2026, Verona, Italy

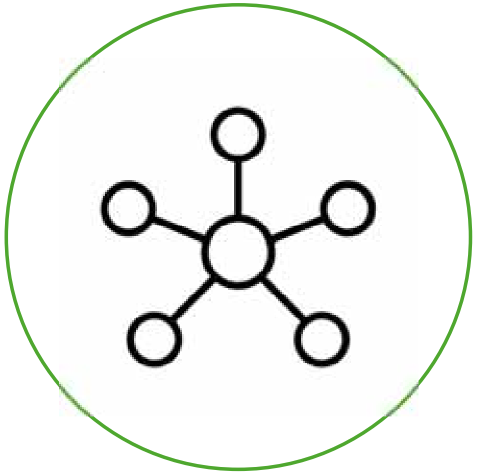




Prediction Framework (SPPF)

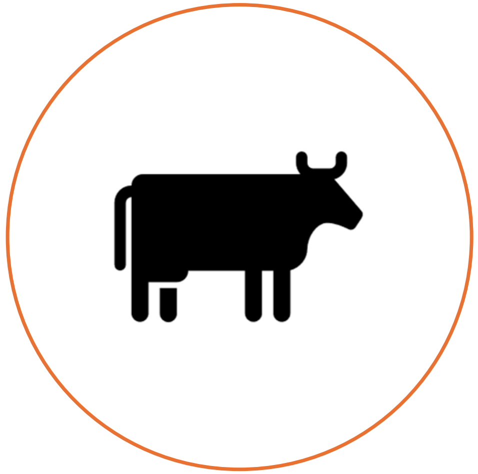
Novel local PLS model (NA-PLSR)

Case study: NUE



Prediction Framework (SPPF)

Novel local PLS model (NA-PLSR)



Case study: NUE

Why we need a standard framework (not just another model)

CONSTRAINTS

Technical barriers

Siloed data

Off-context training

Non-FAIR workflows

FAIR (Findable, Accessible, Interoperable and Reusable); Barker et al. (2022).

Why we need a standard framework (not just another model)

CONSTRAINTS

- Technical barriers
- Siloed data
- Off-context training
- Non-FAIR workflows

CONSEQUENCES

- Slower progress
- Wasted potential
- Poorer translation
- Restricted access

FAIR (Findable, Accessible, Interoperable and Reusable); Barker et al. (2022).

Why we need a standard framework (not just another model)

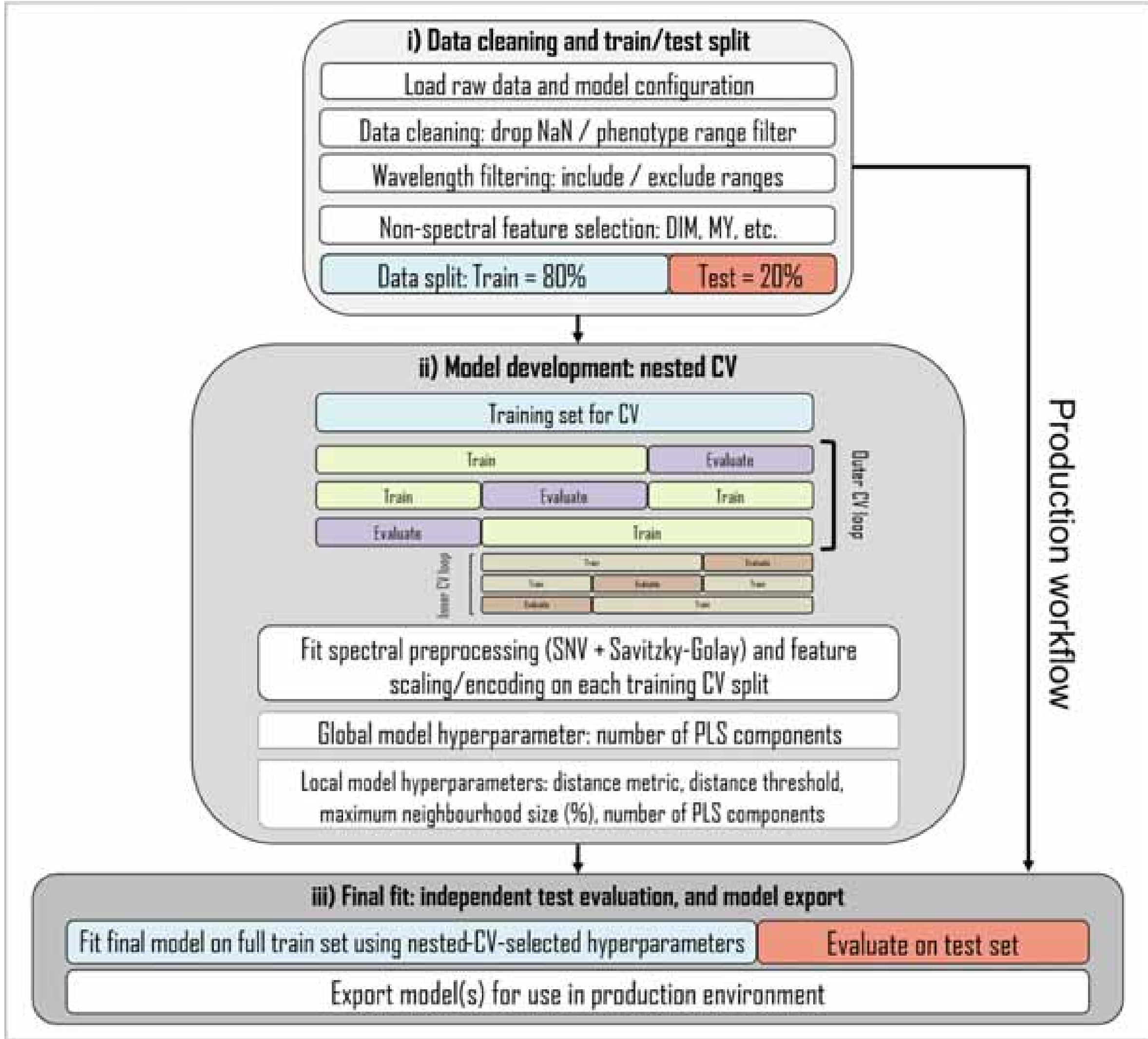
CONSTRAINTS	CONSEQUENCES	REQUIREMENTS
Technical barriers	Slower progress	Broad accessibility
Siloed data	Wasted potential	User-led development
Off-context training	Poorer translation	Decentralised modelling
Non-FAIR workflows	Restricted access	Open & FAIR

FAIR (Findable, Accessible, Interoperable and Reusable); Barker et al. (2022).

Spectral Phenotype Prediction Framework

python

GitHub

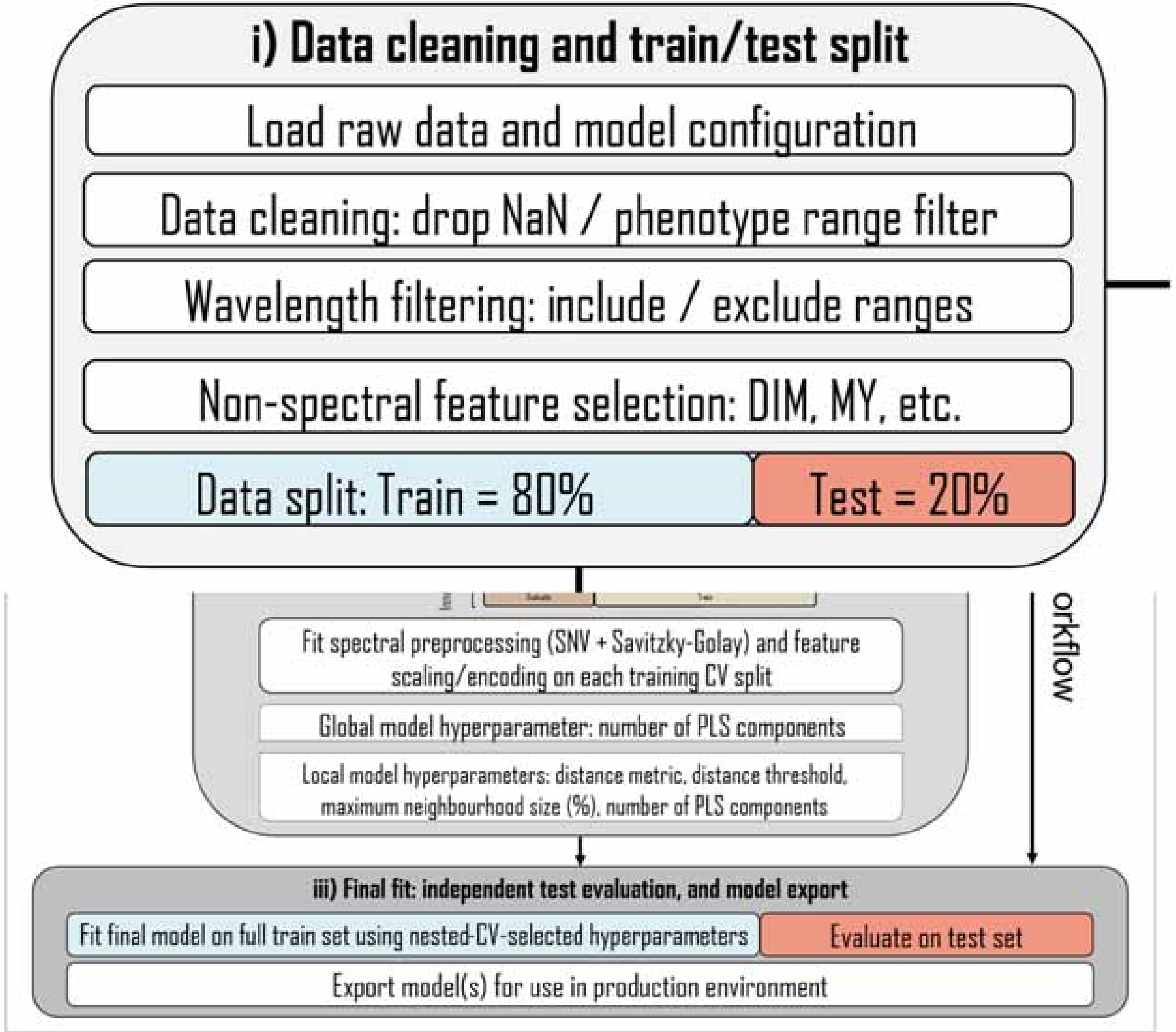


SRUC

Design principles

- Open-source
- Out-of-the-box accessibility
- Modular and extensible
- Trained in context

Spectral Phenotype Prediction Framework



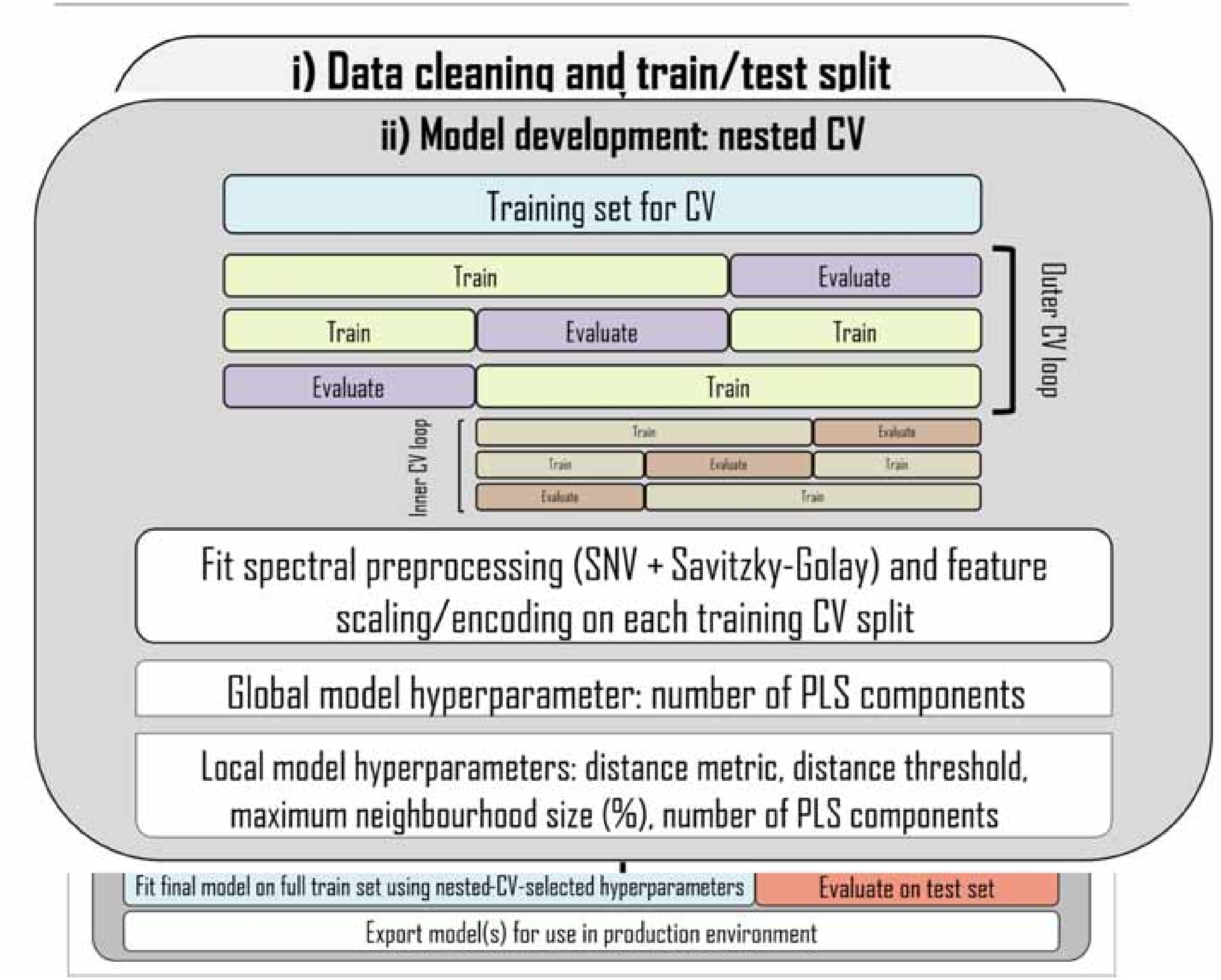
python
GitHub

SRUC

Design principles

- Open-source
- Out-of-the-box accessibility
- Modular and extensible
- Trained in context

Spectral Phenotype Prediction Framework

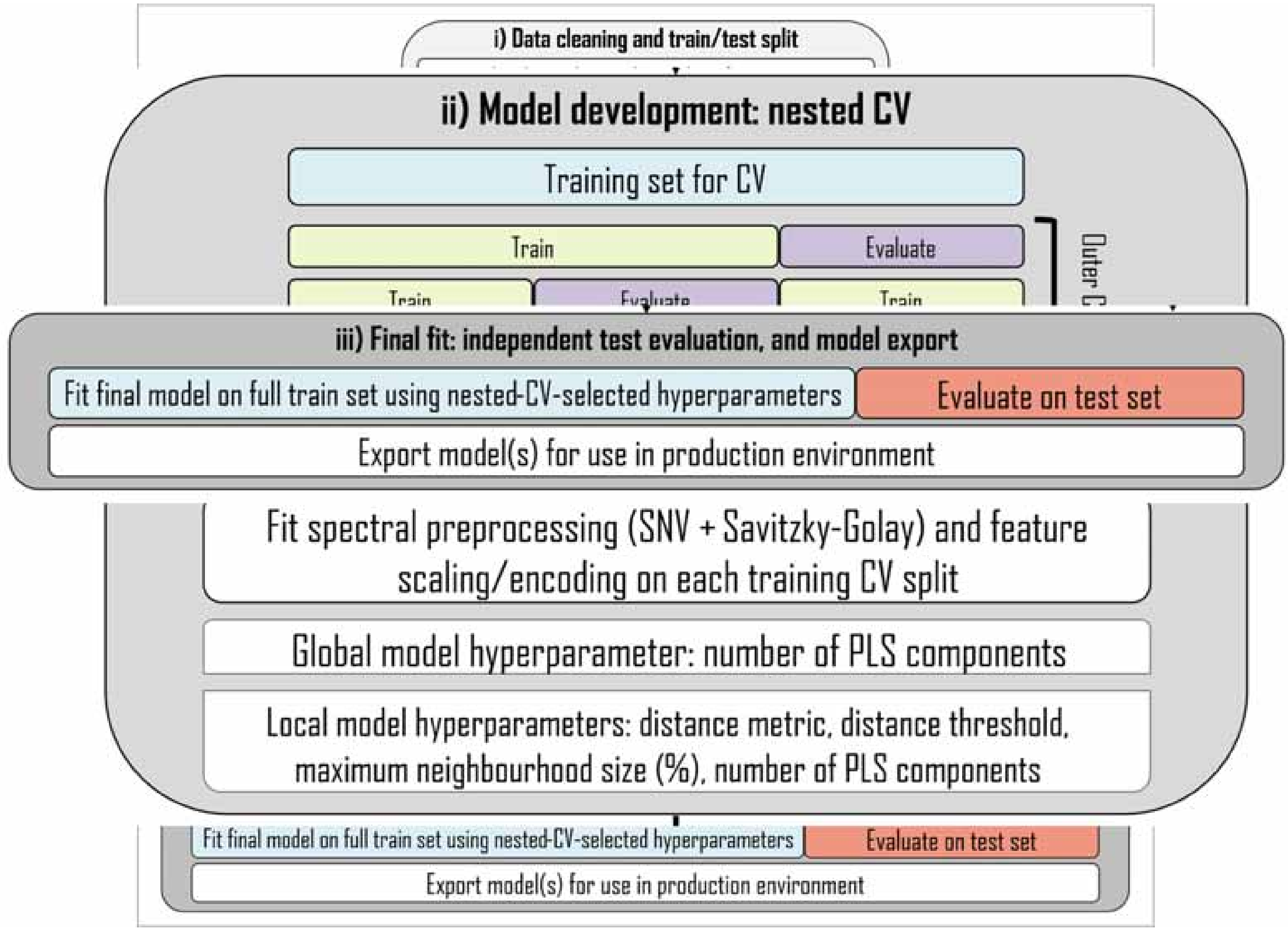


SRUC

Design principles

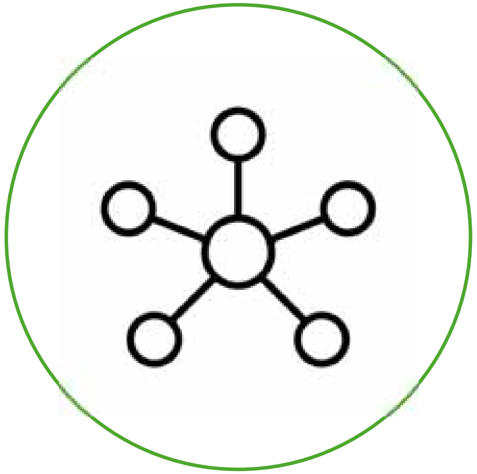
- Open-source
- Out-of-the-box accessibility
- Modular and extensible
- Trained in context

Spectral Phenotype Prediction Framework



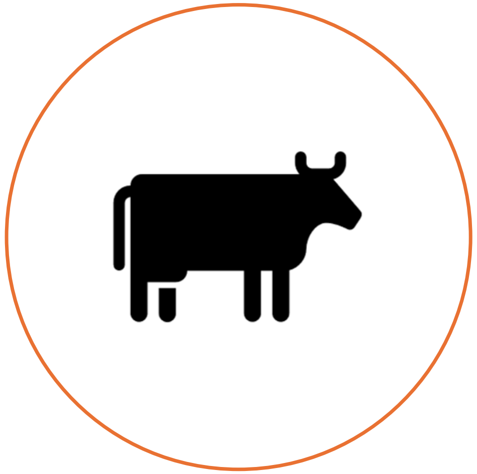
Design principles

- Open-source
- Out-of-the-box accessibility
- Modular and extensible
- Trained in context



Prediction Framework (SPPF)

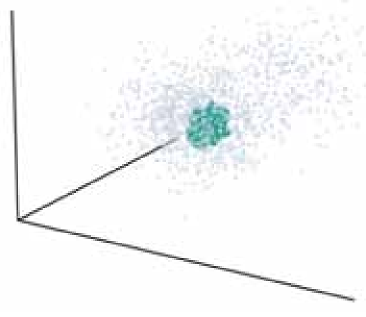
Novel local PLS model (NA-PLSR)



Case study: NUE

Neighbourhood-Adaptive PLSR (NA-PLSR)

- Local instance-wise PLSR model
- Data driven sub-population selected per predictand
  - Distance metric, threshold, ceiling sample size
- Disposible PLSR model fit using this sub-sample
- In use: ~10ms per prediction



Neighbourhood-Adaptive PLSR (NA-PLSR)

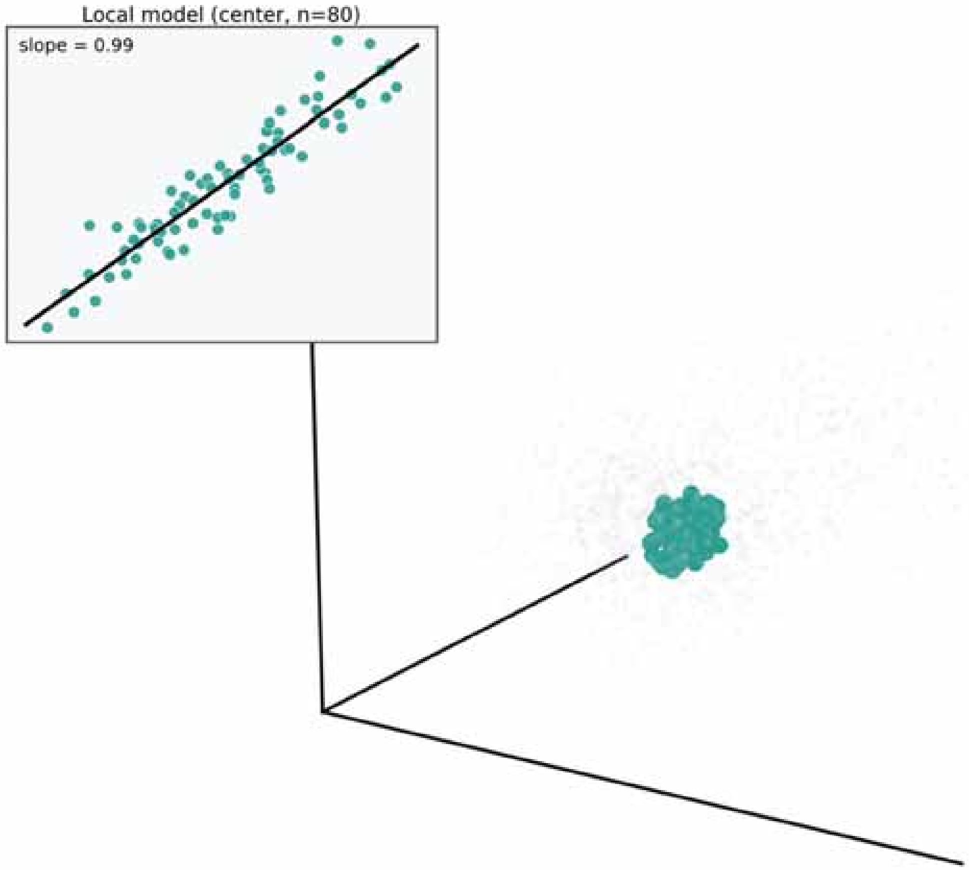
Tuneable parameters

Number of components	`# Separate PLS LV ranges.`
	`global_n_components_range: [20, 30, 40, 50, 60, 80, 100, 120, 150, 175, 200]`
	`comparison_n_components_range: [10, 20, 30, 40, 45]`

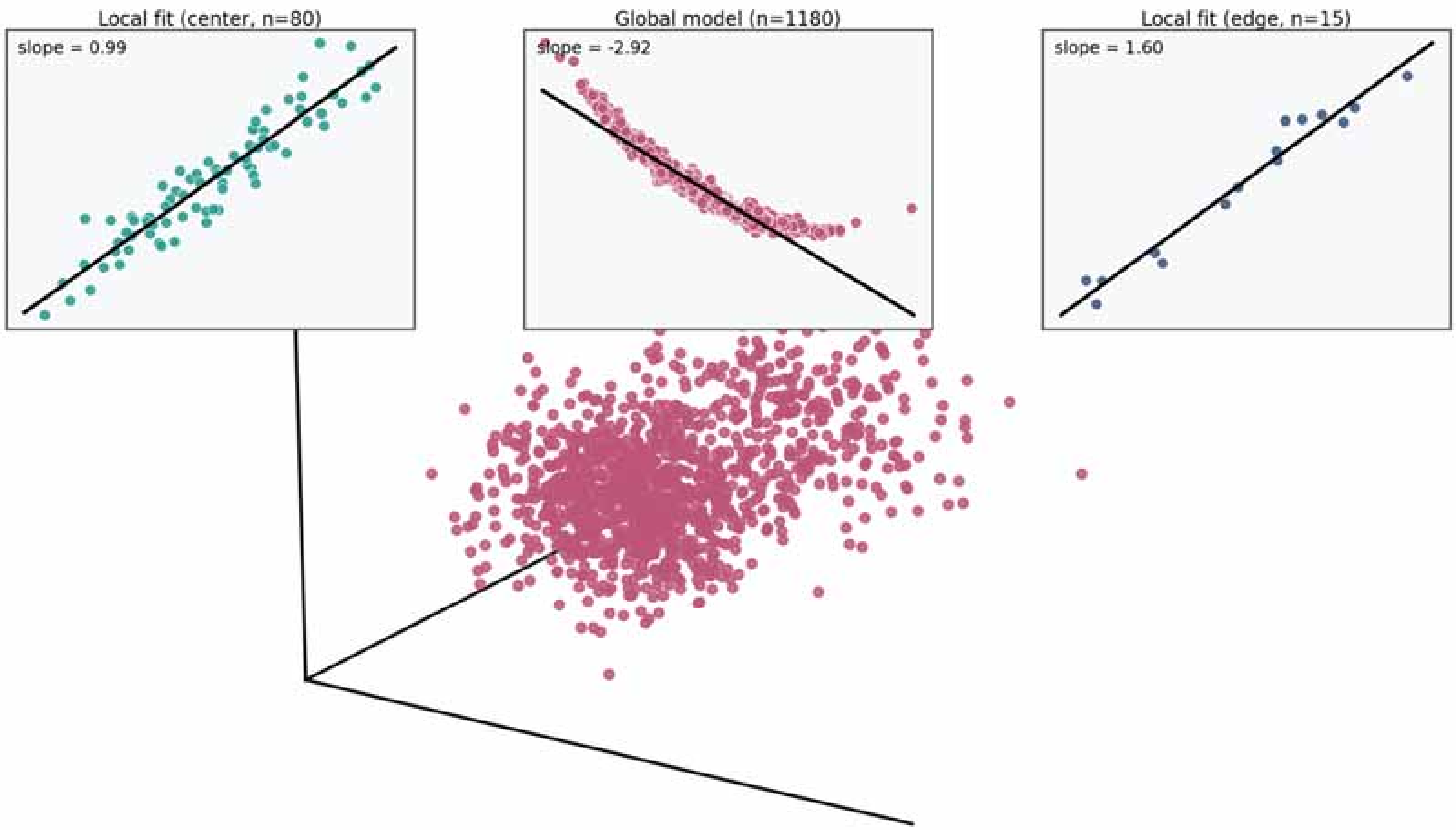
Max. neighbours	`# Maximum number of neighbours` `max_neighbors_grid: [0.05, 0.1, 0.2]  # 0.1 => 10% of samples`
Min. neighbours	`# Minimum neighbourhood size` `min_samples_per_comparison_env: 50`

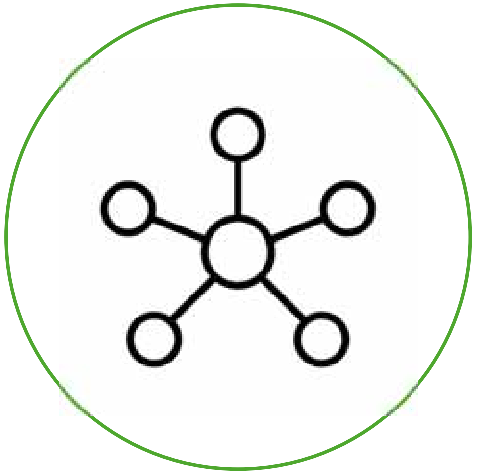
Distance metric / threshold	`# Distance thresholds by metric (keys define the metrics searched in CV)`
	`distance_thresholds:`
	`  cosine: [0.2, 0.3]`
	`  euclidean: [4.0, 6.5]`
	`  mahalanobis: [0.1, 3.0]`
	`  manhattan: [30, 45]`

Neighbourhood-Adaptive PLSR (NA-PLSR)



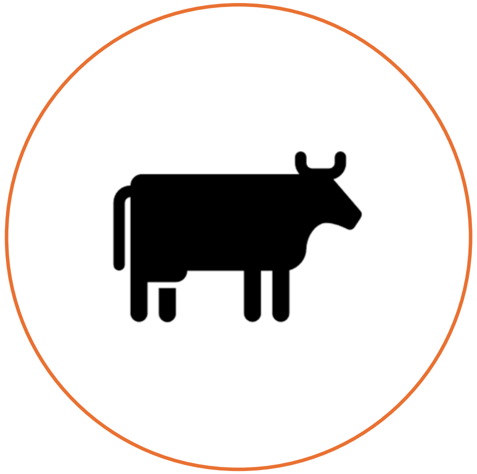
Neighbourhood-Adaptive PLSR (NA-PLSR)





Prediction Framework (SPPF)

Novel local PLS model (NA-PLSR)



Case study: NUE

Case study: Nitrogen-use efficiency

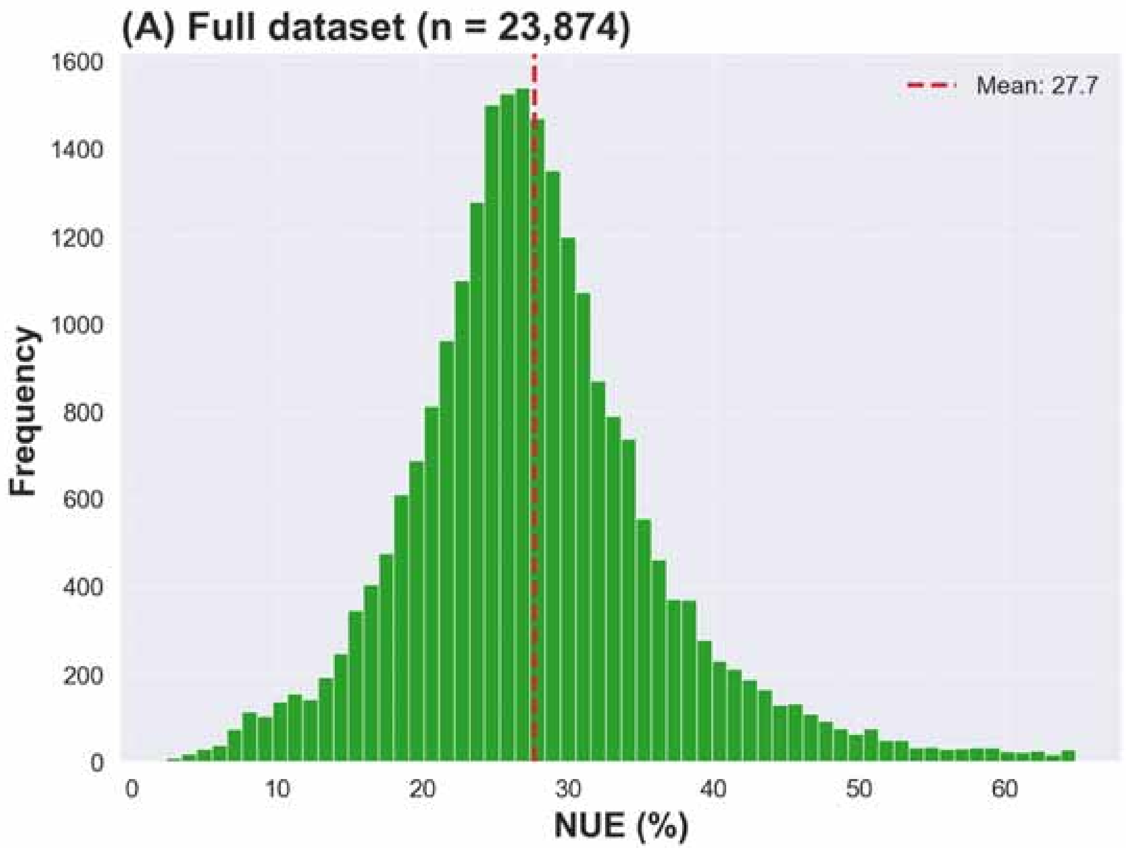
SRUC

$$\text{NUE} = \frac{N_{out}}{N_{in}}$$

$$N_{out} = MY * \frac{CP_{milk}}{6.38}$$

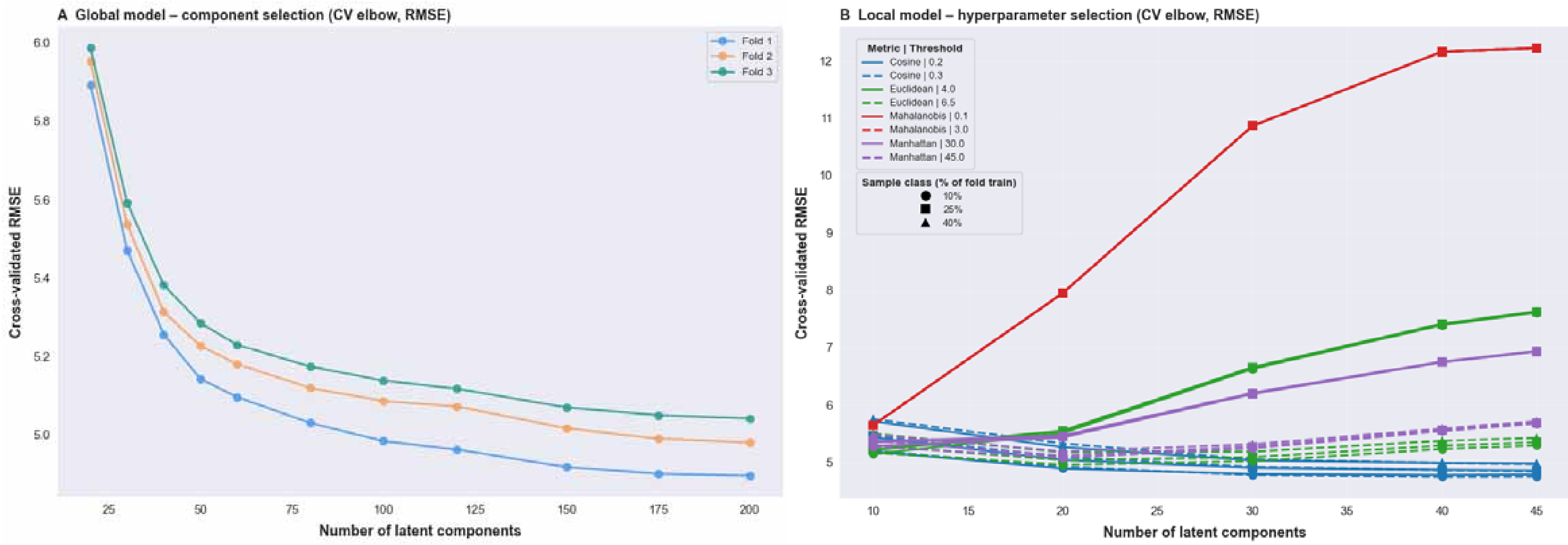
$$N_{in} = TMRin * DM * \frac{CP}{6.25}$$

Case study: Nitrogen-use efficiency



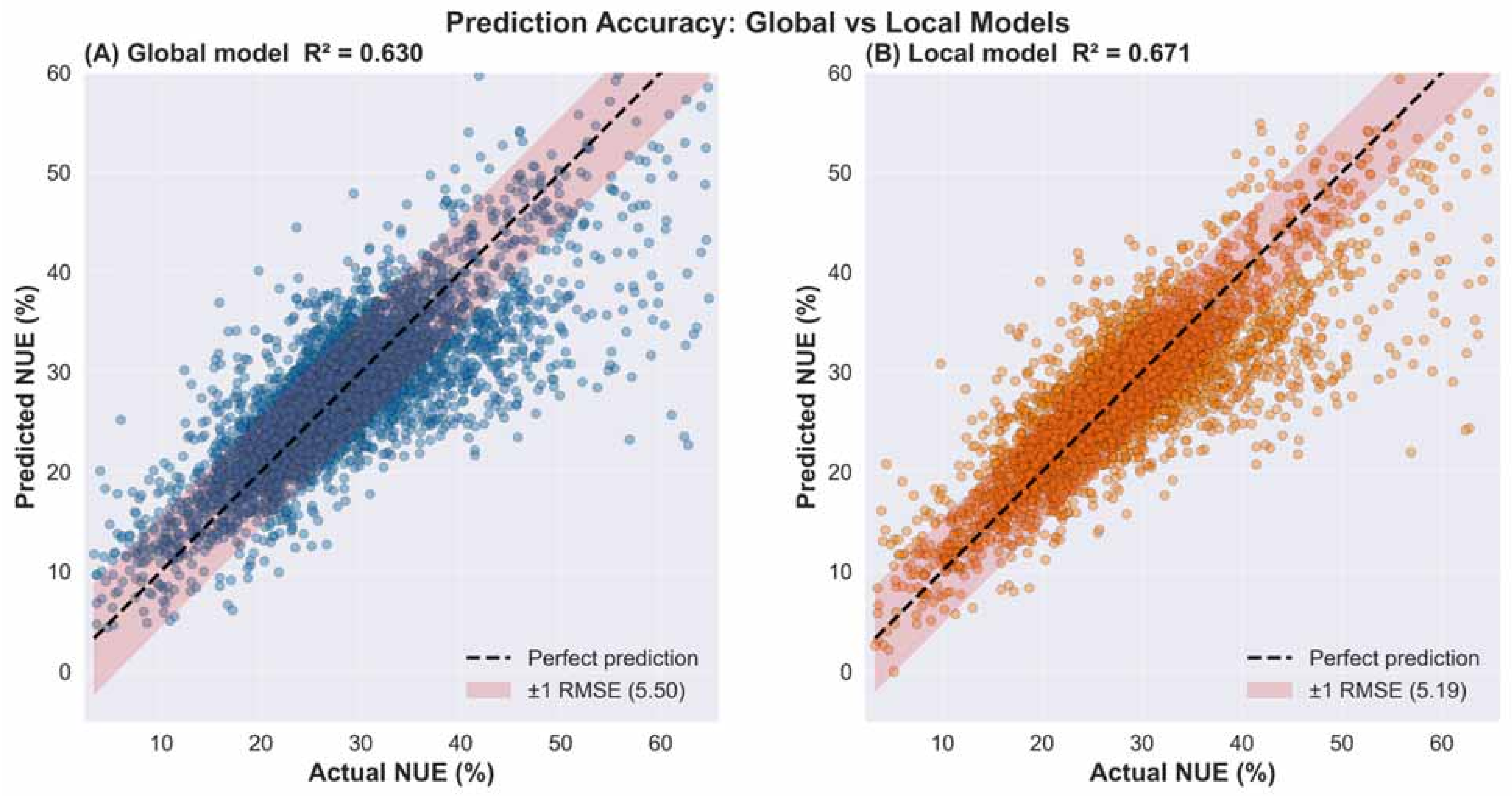
- Matched to spectra date
- 715 Holstein Friesian cows
- 2012 to 2019
- Additional co-variates: MY + DIM + lactation number + month
- FOSS MilkoScan FT6000

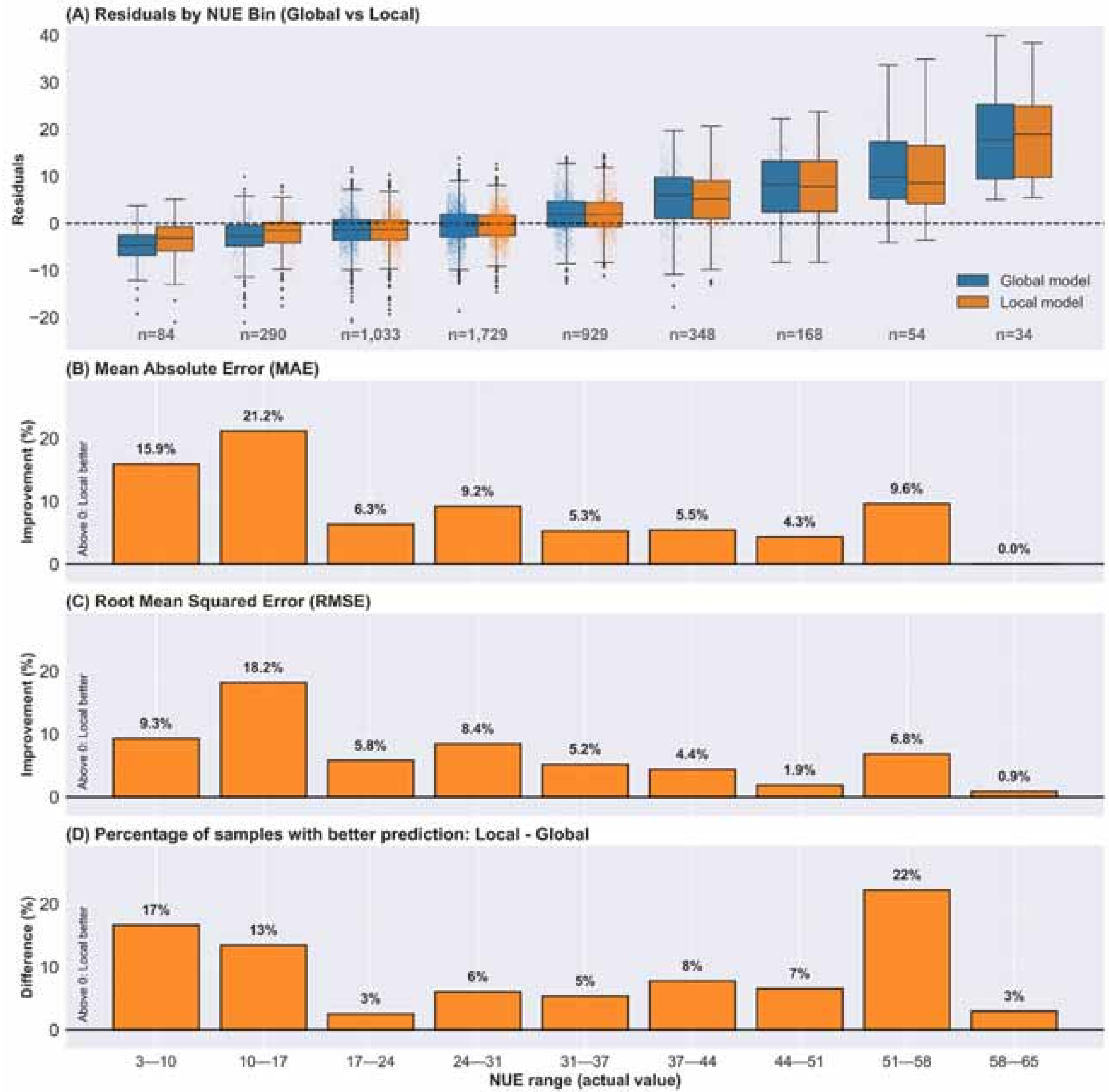
Case study: Nitrogen-use efficiency



200 components

Case study: Nitrogen-use efficiency



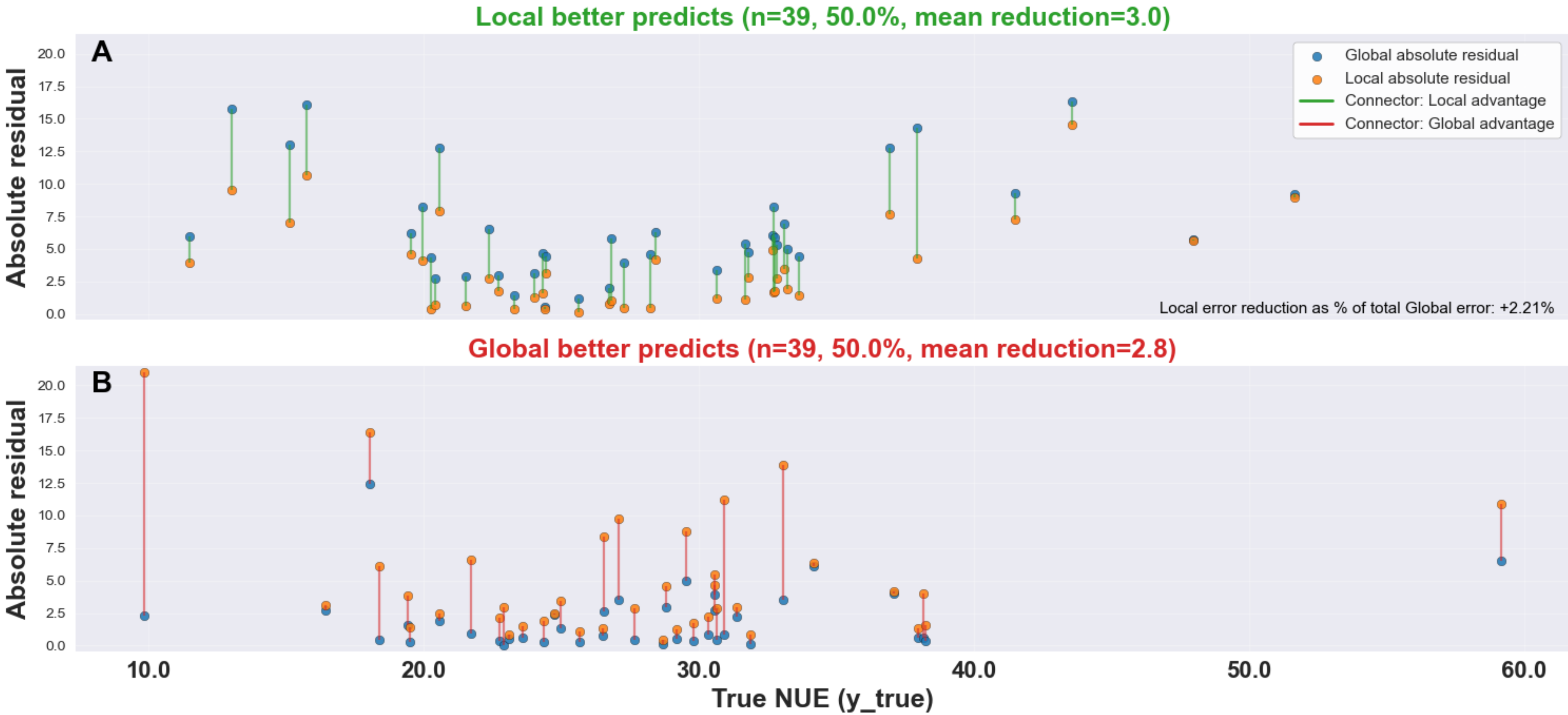


SRUC

<u>Local advantage not linear across NUE range</u>

- Local advantage highest in non-modal ranges
- Important gains for samples that are likely the most biologically important
- Improved management for animals exhibiting non-typical performance

Improvements not solely from edge-cases



Implications and limits

Current contribution:

- SPPF lowers technical barrier for spectroscopy-based predictions
- Open-source approach to promote collaboration and faster development
- Novel local PLSR implementation provided gains for a complex trait

Continued limitations:

- NA-PLSR needs validation on further phenotypes and datasets
- Practical limits of NA-PLSR needs investigated (i.e., min. dataset size)
- Framework limited to Python, needs conversion to other languages (i.e. R)

Future hopes:

- Complements large harmonisation efforts (e.g., CRA-W)
- Gives siloed datasets a FAIR-aligned context-specific evaluation route
- Supports practical testbed for novel methods while preserving comparability to established baselines

Conclusions

- Open-source SPPF lowers barriers for robust spectral phenotyping predictions
- NA-PLSR improved NUE prediction vs global PLSR in this case study
- Benefit is practical: better prediction of non-modal animals with management relevance
- Next: further validation and broader trait testing, implement further ML models, explore federated learning opportunities

Acknowledgements

Team

Dr Mazdak Salavati

Prof. Richard Dewhurst

Prof. John Newbold

Funding

DIGITALDAIRYCHAIN – UKRI Strength in Places Fund

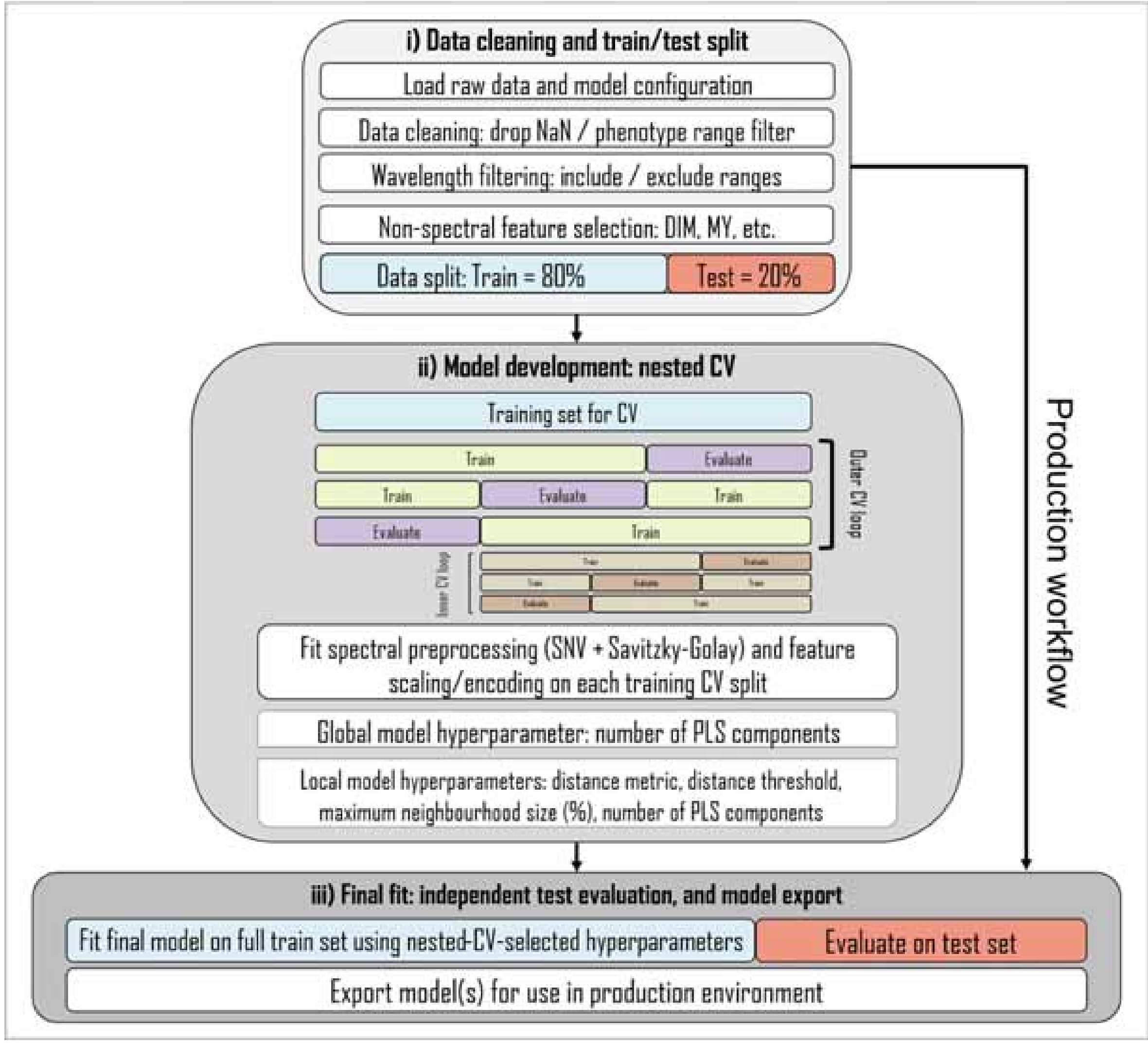
Why we need a framework (not just another model)

CONSTRAINTS	CONSEQUENCES	REQUIREMENTS
Technical barriers	Slower progress	Broad accessibility
Siloed data	Wasted potential	Decentralised modelling
Off-context training	Poorer translation	User-led development
Non-FAIR workflows	Restricted access	Open & FAIR principles

FAIR (Findable, Accessible, Interoperable and Reusable); Barker et al. (2022).

Spectral Phenotype Prediction Framework

python
GitHub

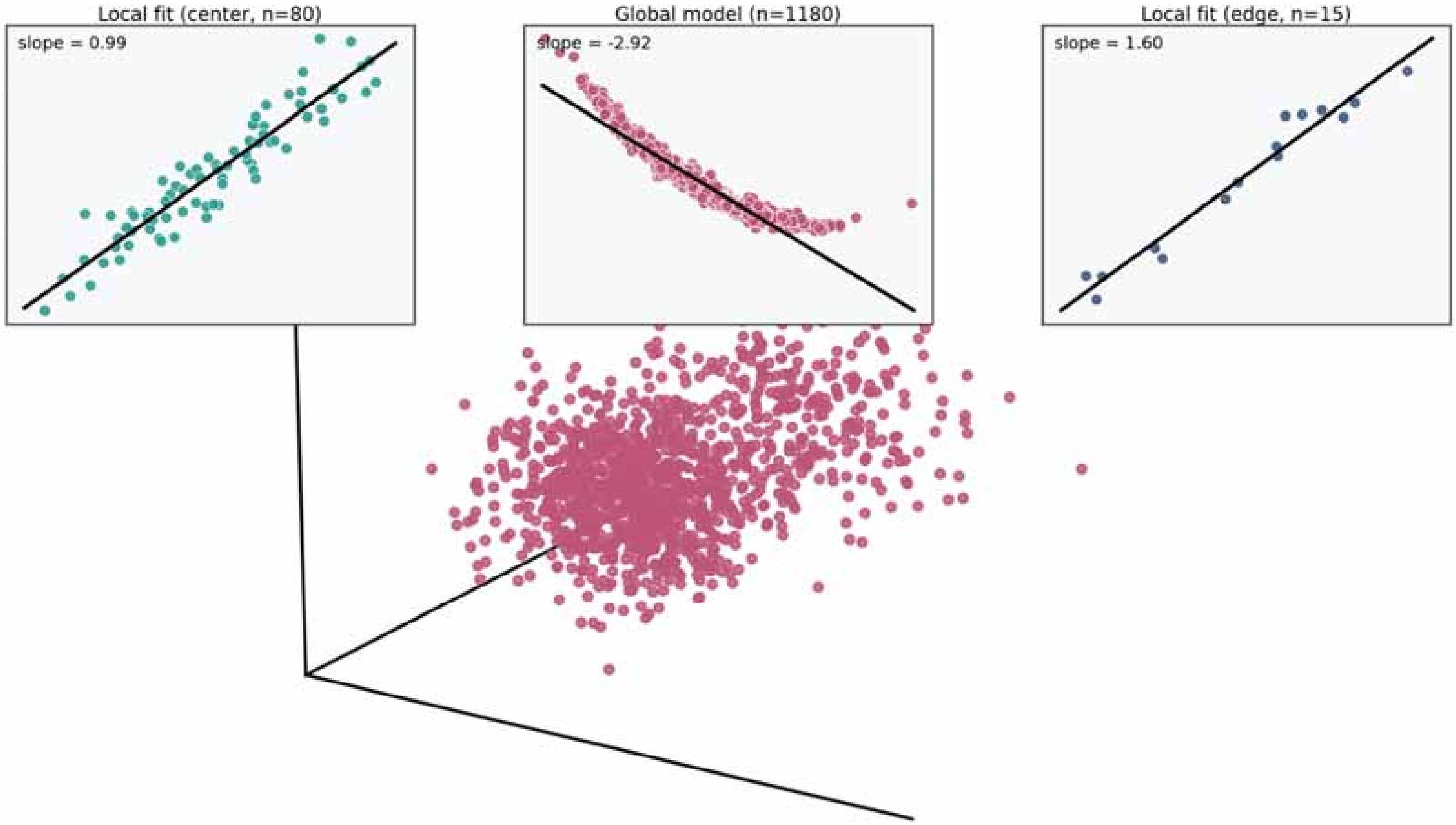


SRUC

Design principles

- Open-source
- Out-of-the-box accessibility
- Modular and extensible
- Trained in context

Neighbourhood-Adaptive PLSR (NA-PLSR)



Why we need a standard framework (not just another model)

CONSTRAINTS

- Technical barriers
- Siloed data
- Off-context training
- Non-FAIR workflows

CONSEQUENCES

- Slower progress
- Wasted potential
- Poorer translation
- Restricted access

REQUIREMENTS

- Broad accessibility
- User-led development
- Decentralised modelling
- Open & FAIR

FAIR (Findable, Accessible, Interoperable and Reusable); Barker et al. (2022).