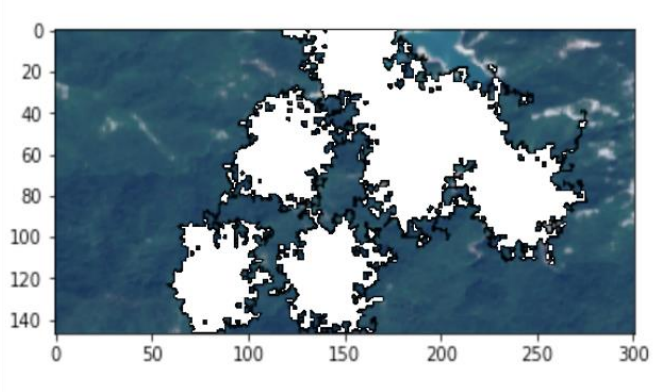
AutoML to Advance Earth Observation

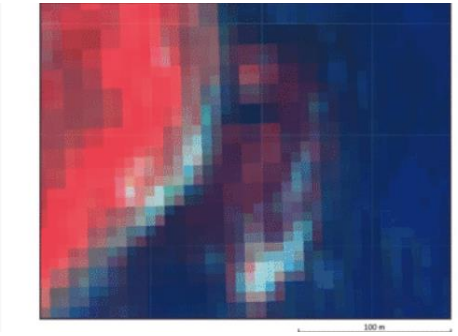
Mitra Baratchi

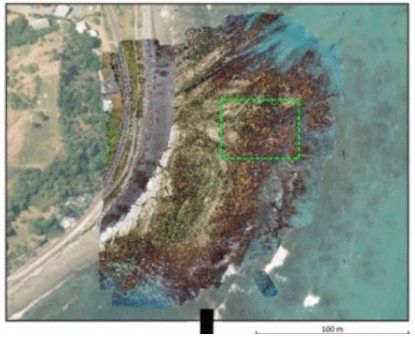
Leiden University – Leiden Institute of Advanced Computer Science (LIACS)
Spatio-temporal data Analysis and Reasoning (STAR)

The imperfect non-homogeneous spatio-temporal data



Missing data

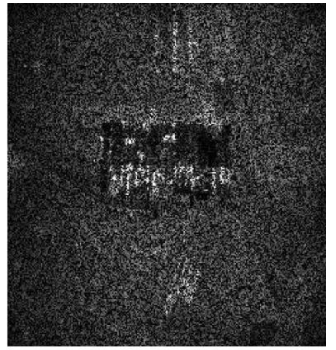




Different resolutions

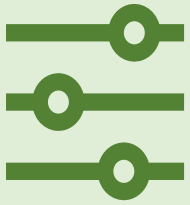
Ground truth

Observation



Noise

The AutoML building blocks



Search space

Possible algorithms and hyperparameter settings

Search strategy

How we explore them

Performance evaluation

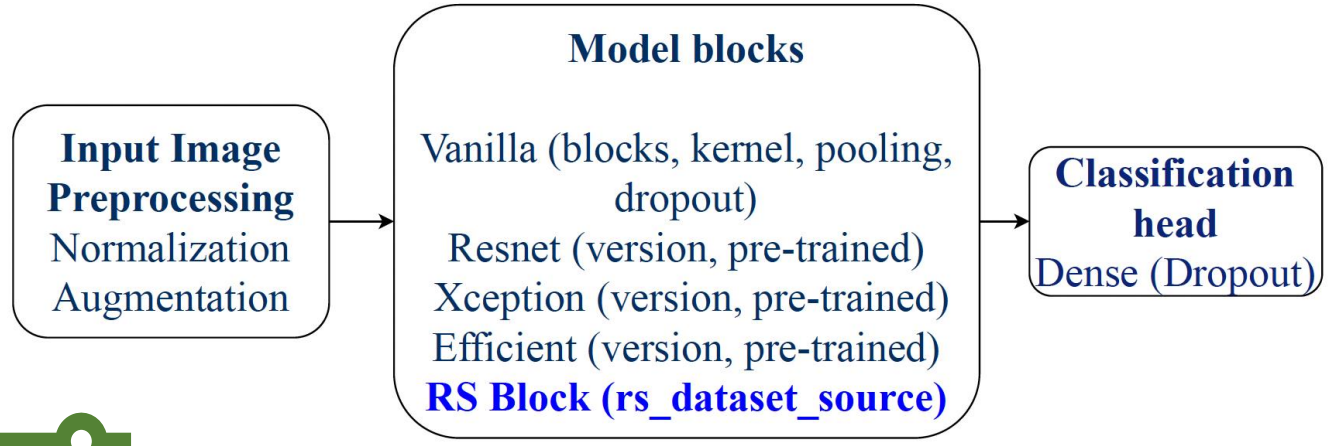
How we judge them

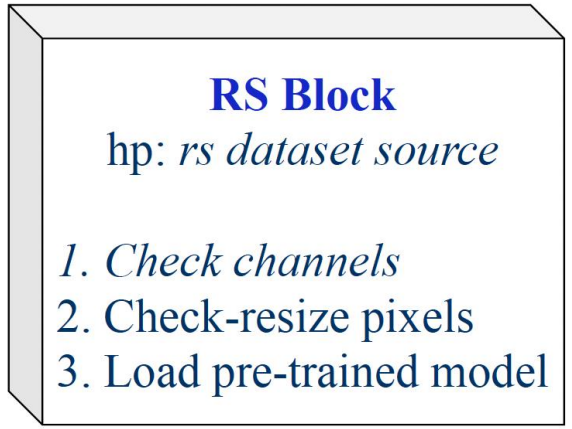
Central question in AutoML

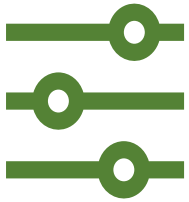
How to configure the **best model** for each **dataset** automatically?

Let's make a general system that works on all problems of the same type

Classification







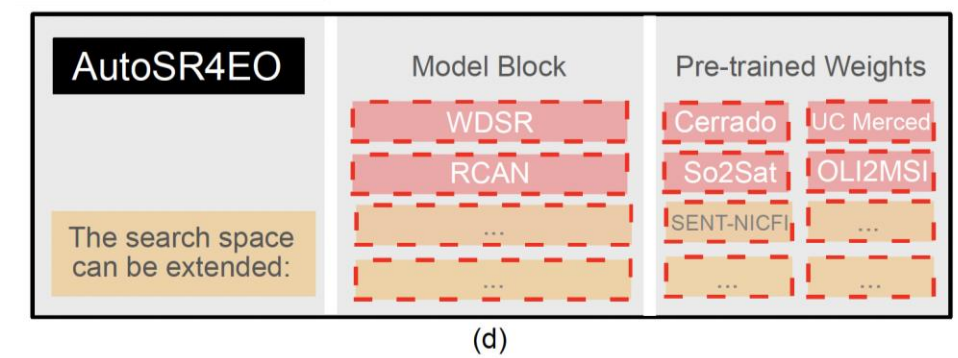
Search space

Palacios, N., Baratchi, M., van Rijn, J. N., & Vollrath, A., Automated machine learning for integrating remote sensing pre-trained models into AutoML systems. In Joint European conference on machine learning and knowledge discovery in databases, 2021.

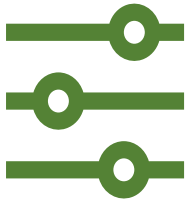
Super-resolution

Wąsala, J., Marselis, S., Arp, L., Hoos, H., Longepe, N., & Baratchi, M., AutoSR4EO: An AutoML approach to super-resolution for Earth observation images. *Remote Sensing*, *16*(3), 2024.

Super-resolution



(d)



Search space

Wąsala, J., Marselis, S., Arp, L., Hoos, H., Longepe, N., & Baratchi, M. AutoSR4EO: An AutoML approach to super-resolution for earth observation images. *Remote Sensing*, *16*(3), 2024.

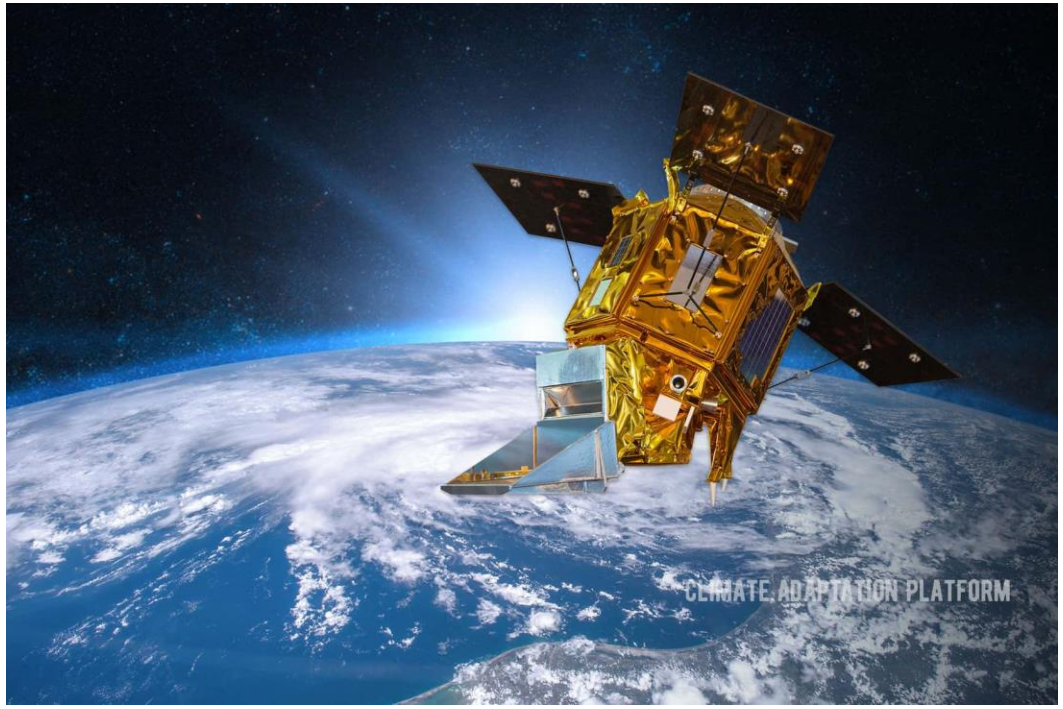
Few lessons learnt after working with downstream application users of EO data

AutoML for data fusion

Wąsala, J, Maasakkers, B., Schuit, B., Leguijt, G., Aben, I., Schneider, R., Hoos, H., and Baratchi, M., AutoMergeNet: AutoML-based m-source satellite data fusion evaluated with atmospheric case studies, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025.

TROPOMI

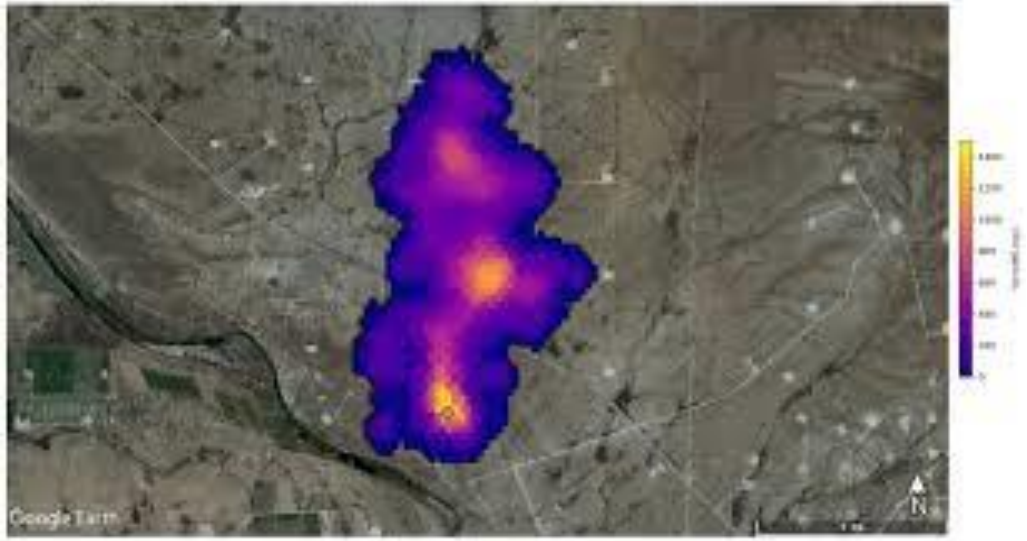
- Launched in 2017
- On board of Sentinel-5P
- Maps global air pollution (e.g., nitrogen dioxide, methane, carbon monoxide)



SRON

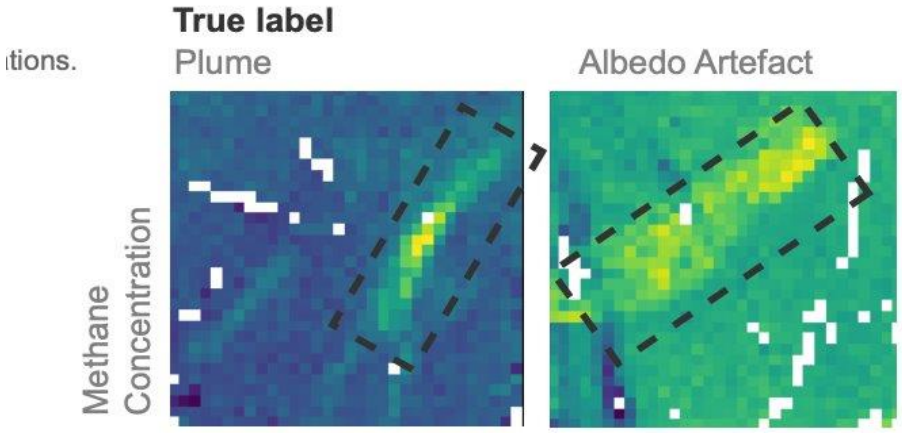
Methane plume

- Invisible clouds of methane gas
- An important cause of global warming



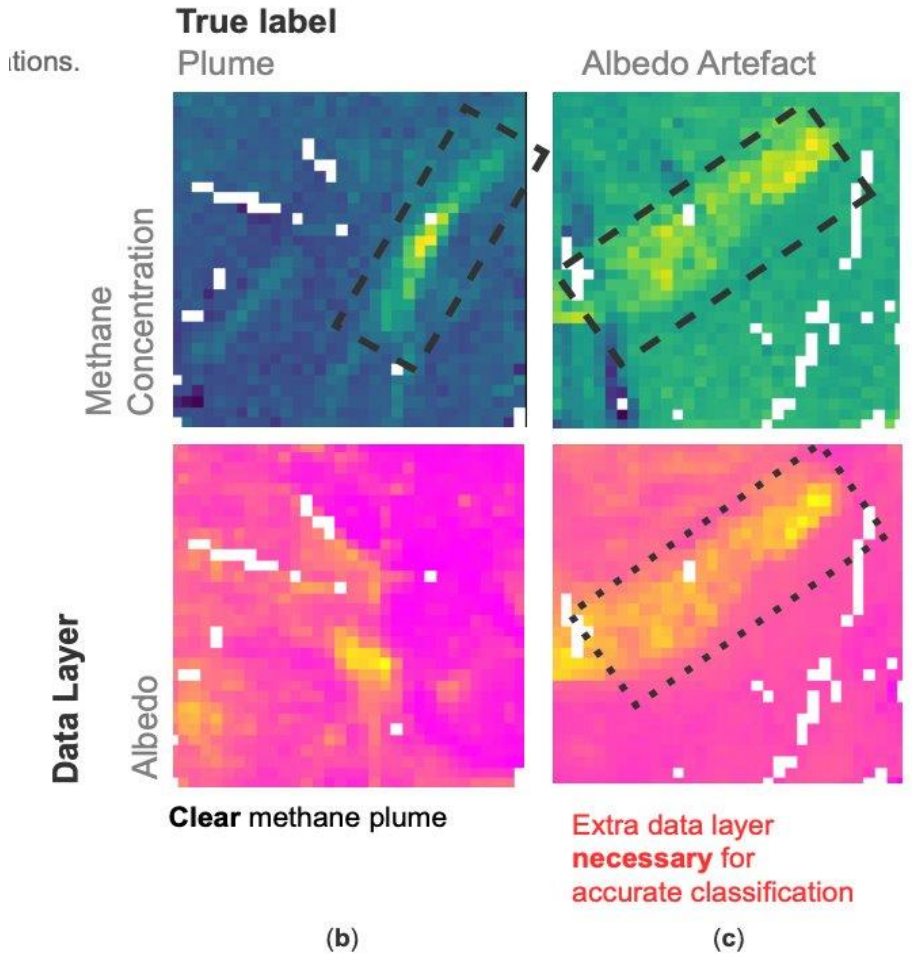
Binary classification problem

- Many Earth observation problems are detecting one concept against everything else



Concept detection problem

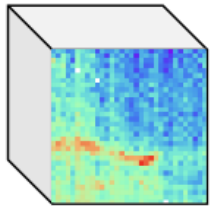
- Many Earth observation problems are detecting one concept against everything else
- Iterative improvement of the model and feature engineering



The task

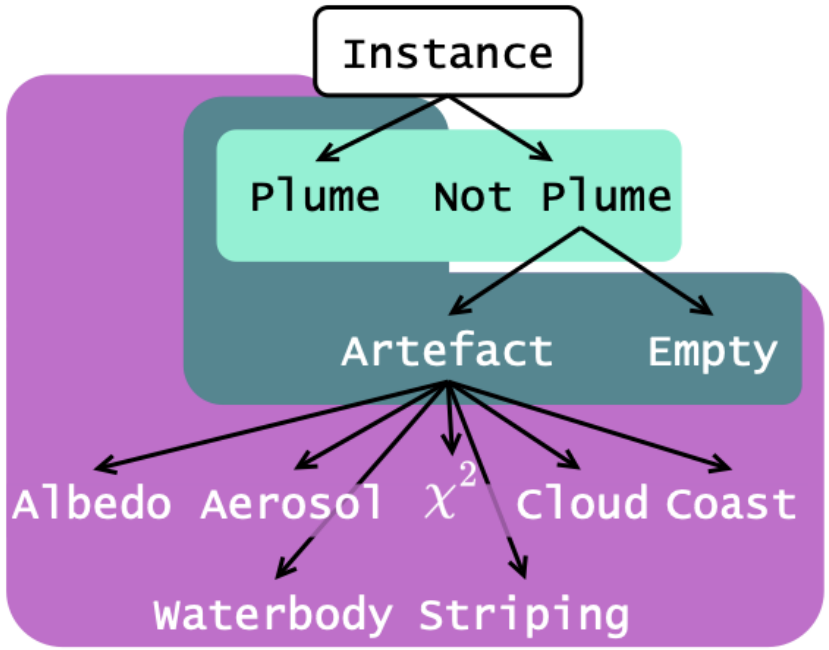
Instance

Data cube with 7 channels

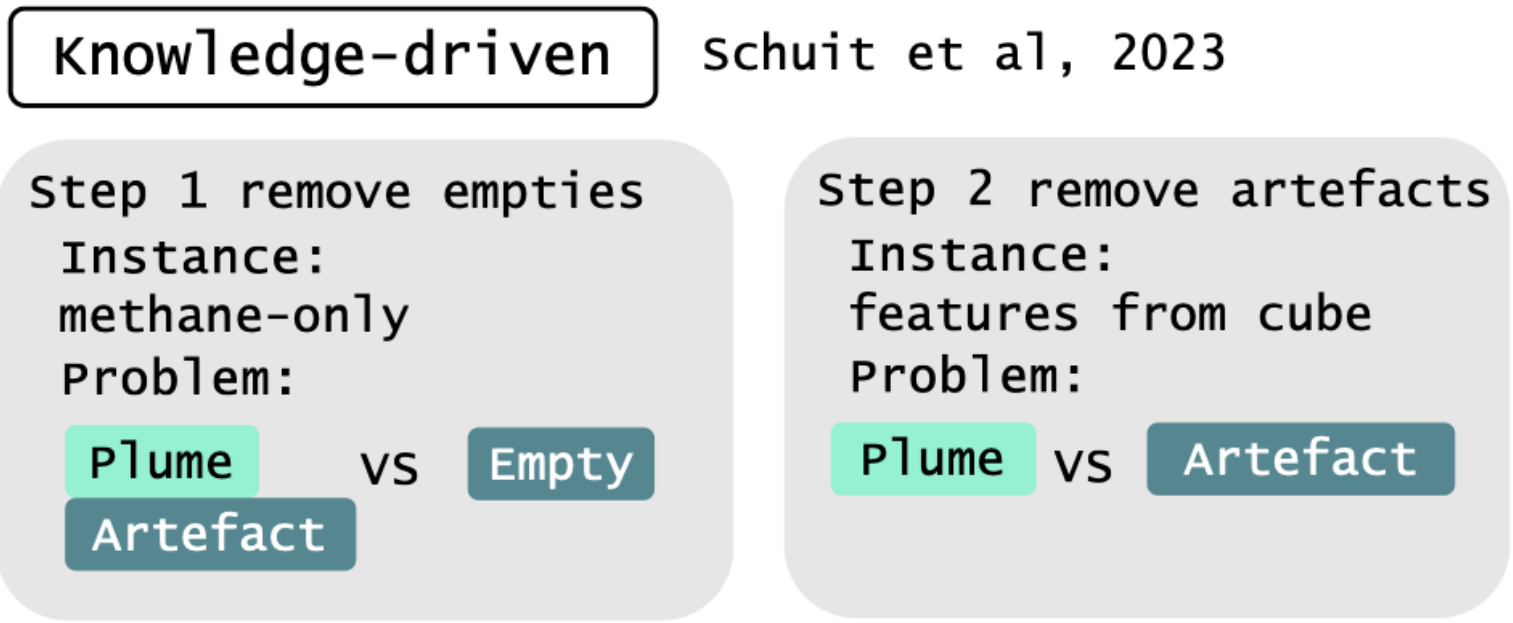


0: Methane
1: Surface pressure
2: Albedo
3: Aerosol optical thickness
4: Quality assurance
5: Cloud fraction
6: Land cover type

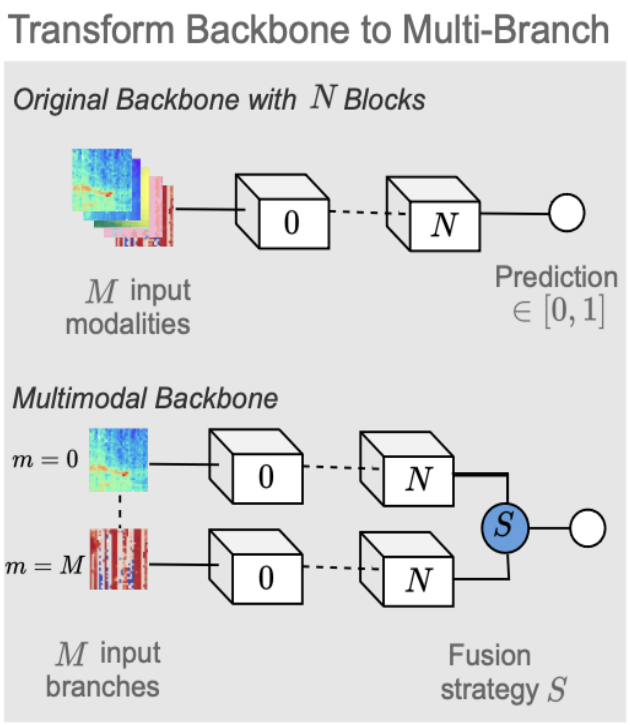
Label

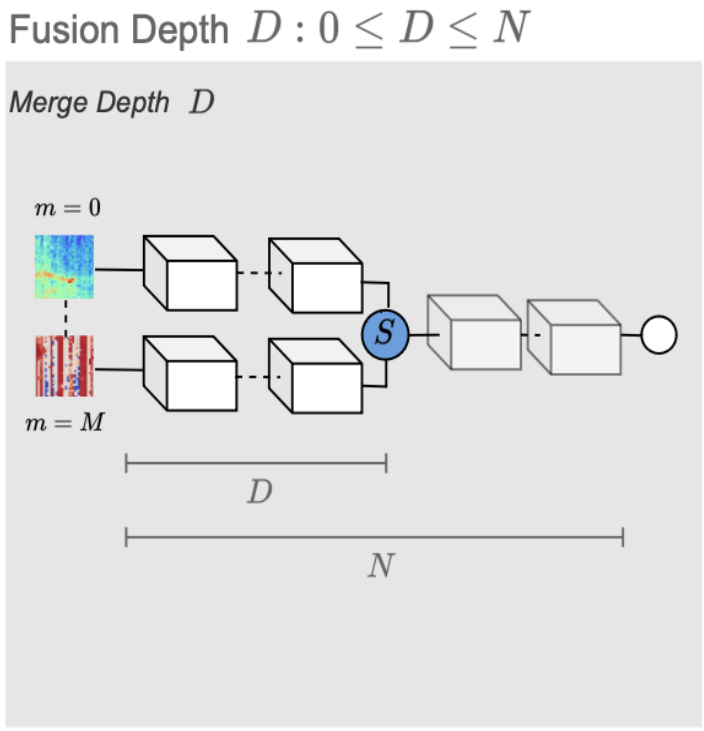


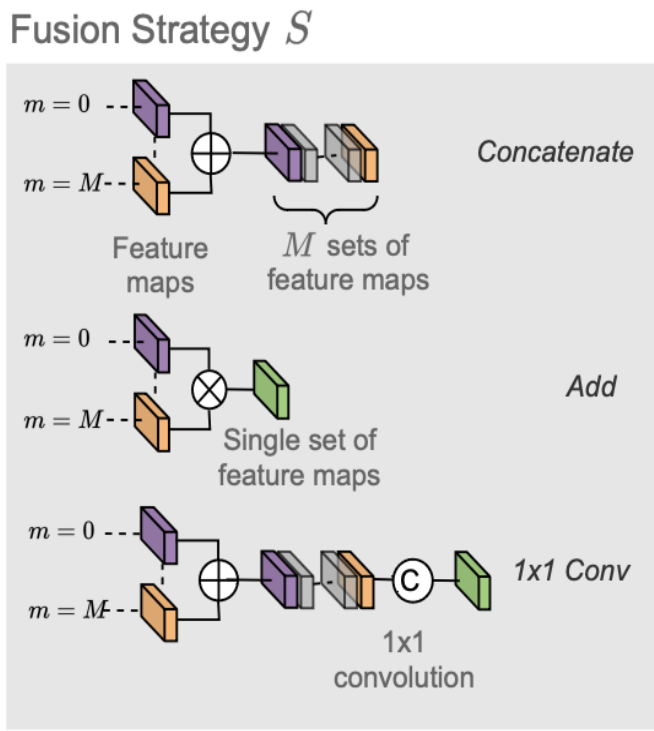
A knowledge-driven pipeline



AutoMergeNet







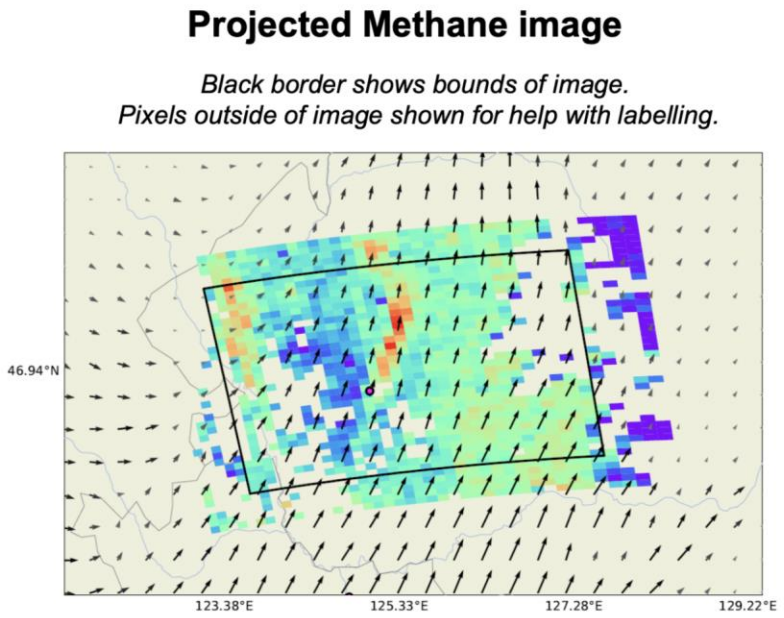
J. Wąsala, J. D. Maasakkers, B. J. Schuit, G. Leguijt, I. Aben, R. Schneider, H. Hoos, and M. Baratchi., AutoMergeNet: AutoML-based m-source satellite data fusion evaluated with atmospheric case studies, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025.

Results against baselines (Methane dataset)

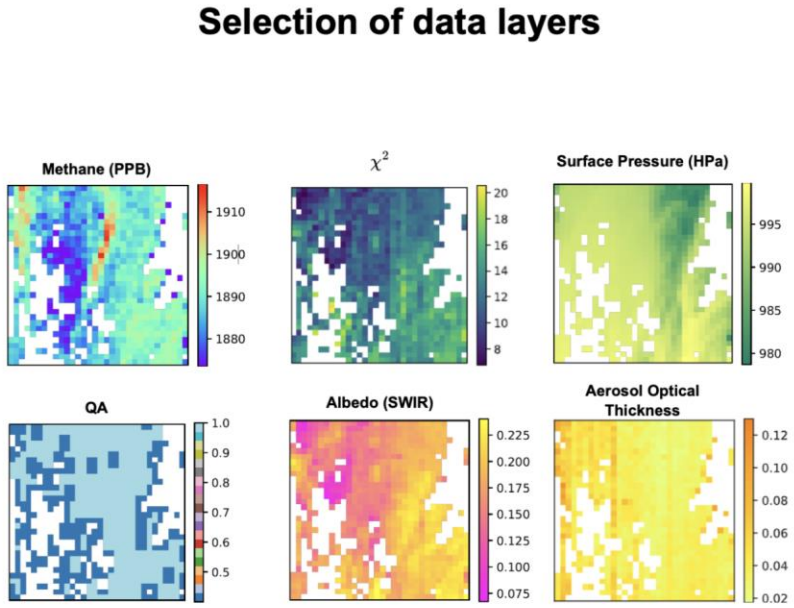
Model	Accuracy	Precision	Recall
BANet [7]	0.89 ± 0.01	0.84 ± 0.01	0.82 ± 0.05
CNN [15]	0.89 ± 0.01	0.85 ± 0.02	0.82 ± 0.03
CvT [10]	0.89 ± 0.02	0.84 ± 0.02	0.85 ± 0.08
EpsaNet [35]	0.88 ± 0.02	0.84 ± 0.02	0.80 ± 0.07
ResNet [6]	0.88 ± 0.02	0.84 ± 0.03	0.81 ± 0.10
MobileNetV2 [9]	0.87 ± 0.04	0.87 ± 0.03	0.89 ± 0.3
AutoMergeNet (ours)	$\mathbf{0.94} \pm 0.01$	$\mathbf{0.91} \pm 0.03$	$\mathbf{0.91} \pm 0.04$

Unseen test set

- 17760 images (one week)



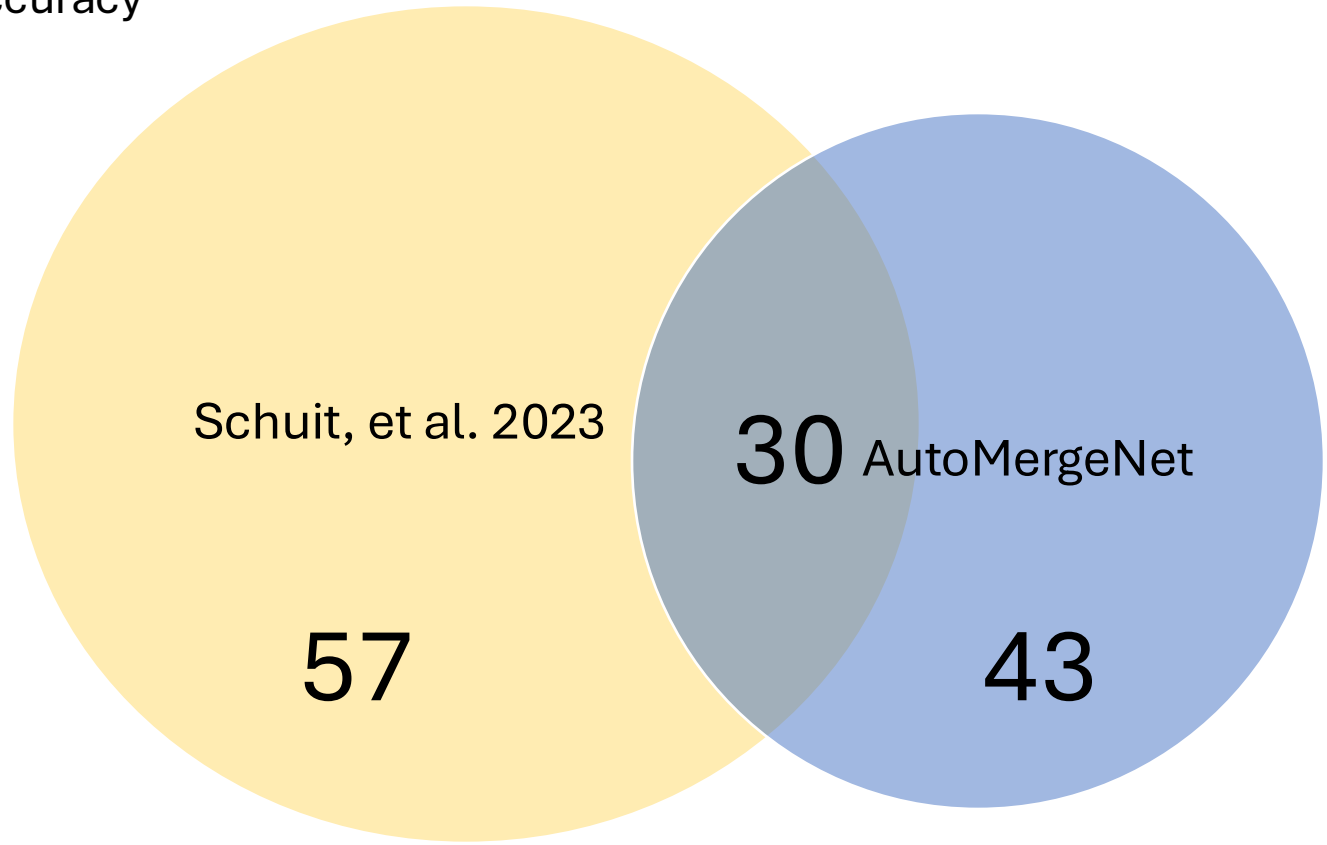
Clear plume shape aligned with wind direction. Plume starts at a known emission location.



No artefacts visible in supporting data layers.

Plumes detected in 17760 images

Far from 94% accuracy



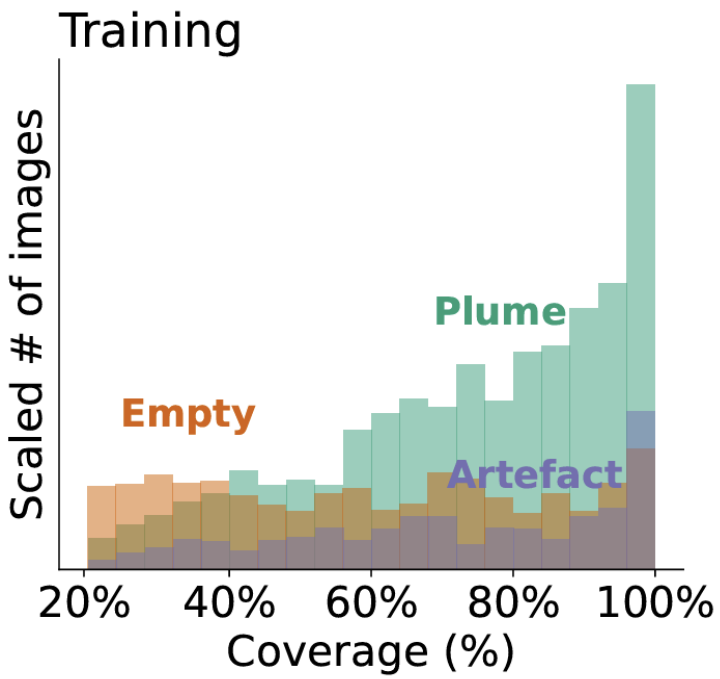
What is the underlying cause of the generalisation gap?

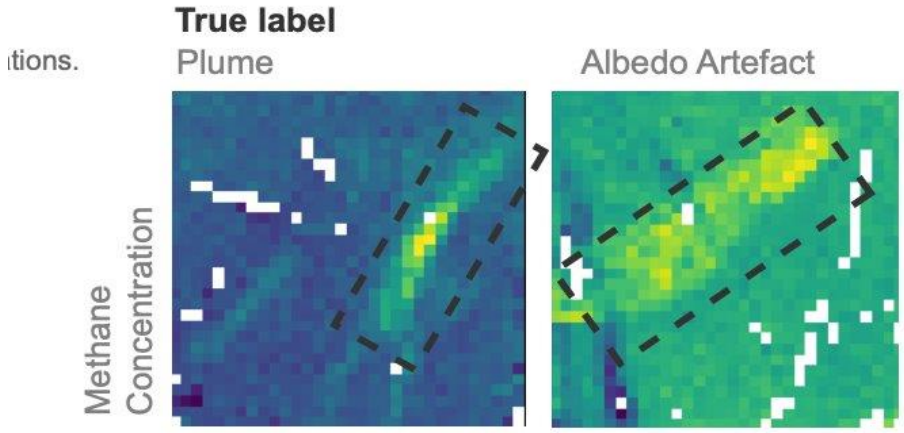
Mitigating representation bias caused by missing pixels

Best paper

Wąsala, J., Maasakkers, J., D., Schuit, B., Leguijt, G., Aben, I., Schneider, R., Hoos, H., Baratchi, M., Mitigating representation bias caused by missing pixels in methane plume detection, In proceedings of the MACLEAN workshop @ ECML-PKDD 2025.

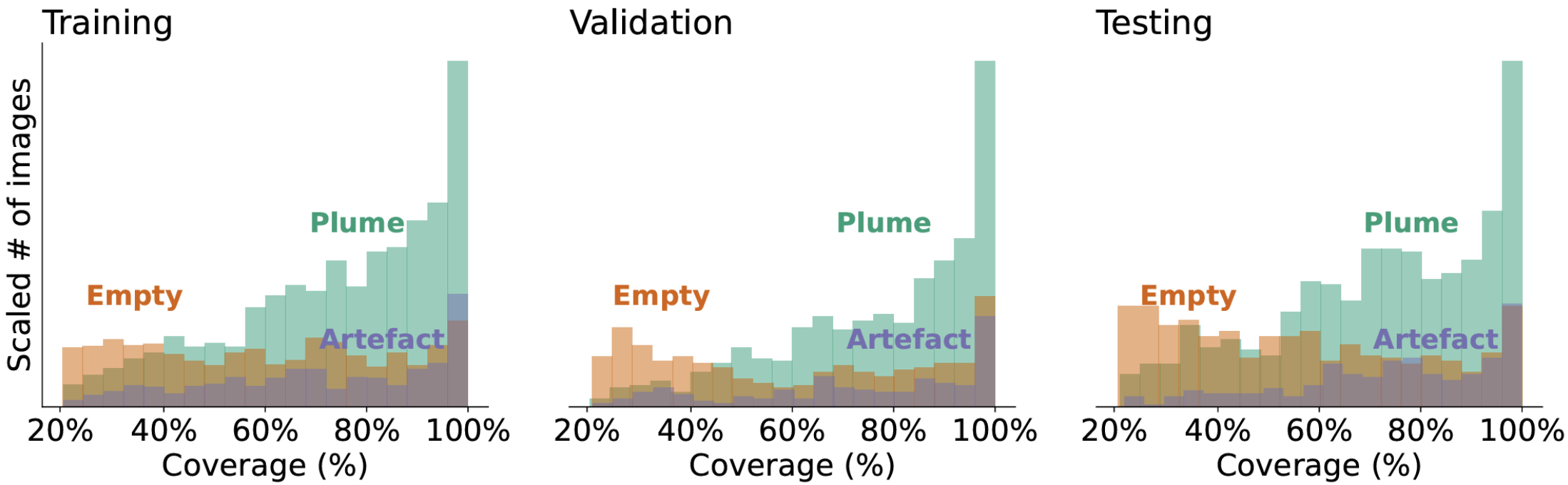
Distribution of images





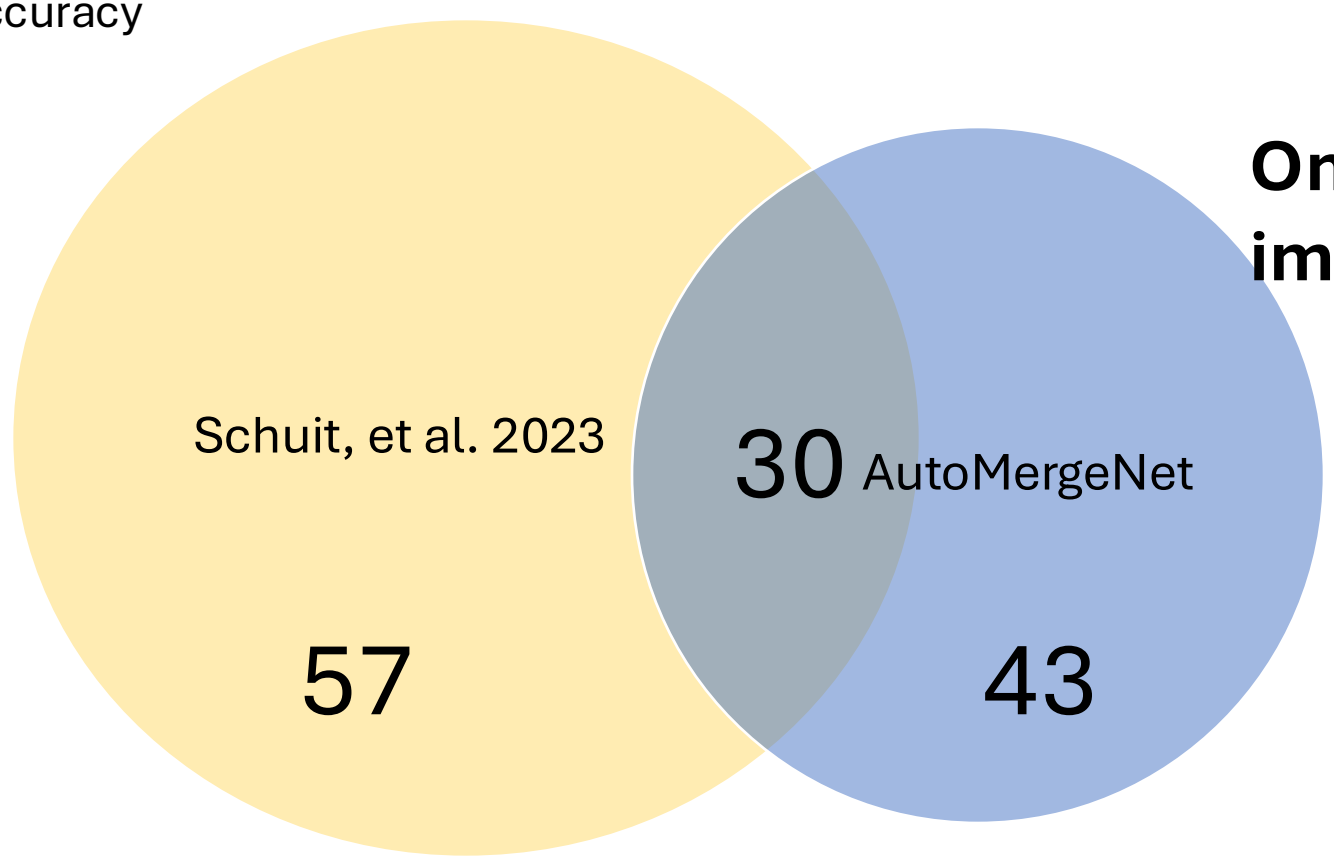
Distribution of images

Plumes have higher coverage than empty images -> the model learns to associate coverage to the label



Plumes detected in 17760 images

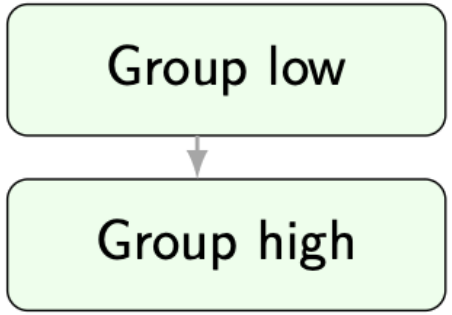
Far from 94% accuracy



Only high coverage images

Fairness evaluation metrics

Compare groups



We ask whether the model behaves differently across groups we care about.

Selection parity

PR_{low}
PR_{high}

$$\text{parity} = \frac{PR_{high}}{PR_{low}}$$

False positive rate gap

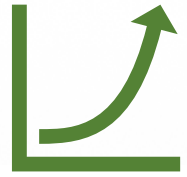
FPR_{low}
FPR_{high}

$$\Delta FPR = FPR_{low} - FPR_{high}$$

True positive rate gap

TPR_{low}
TPR_{high}

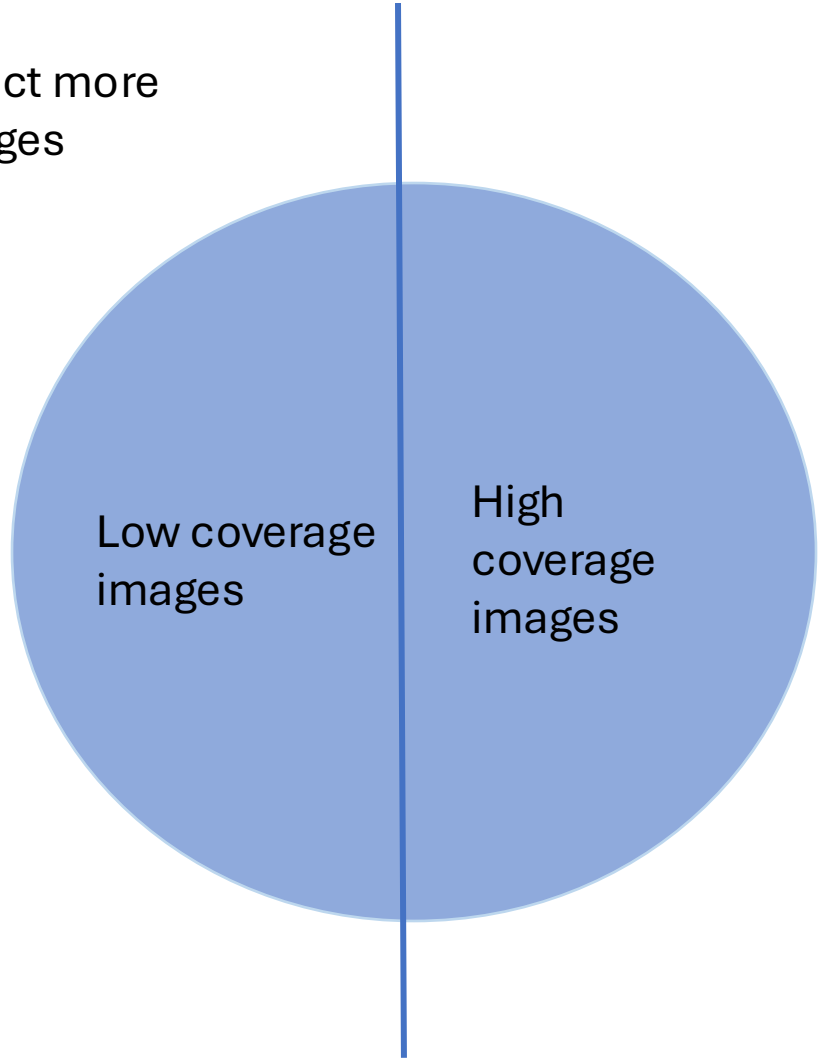
$$\Delta TPR = TPR_{low} - TPR_{high}$$



Performance evaluation

Plumes detected in 17760 images

More balanced models detect more plumes in low coverage images

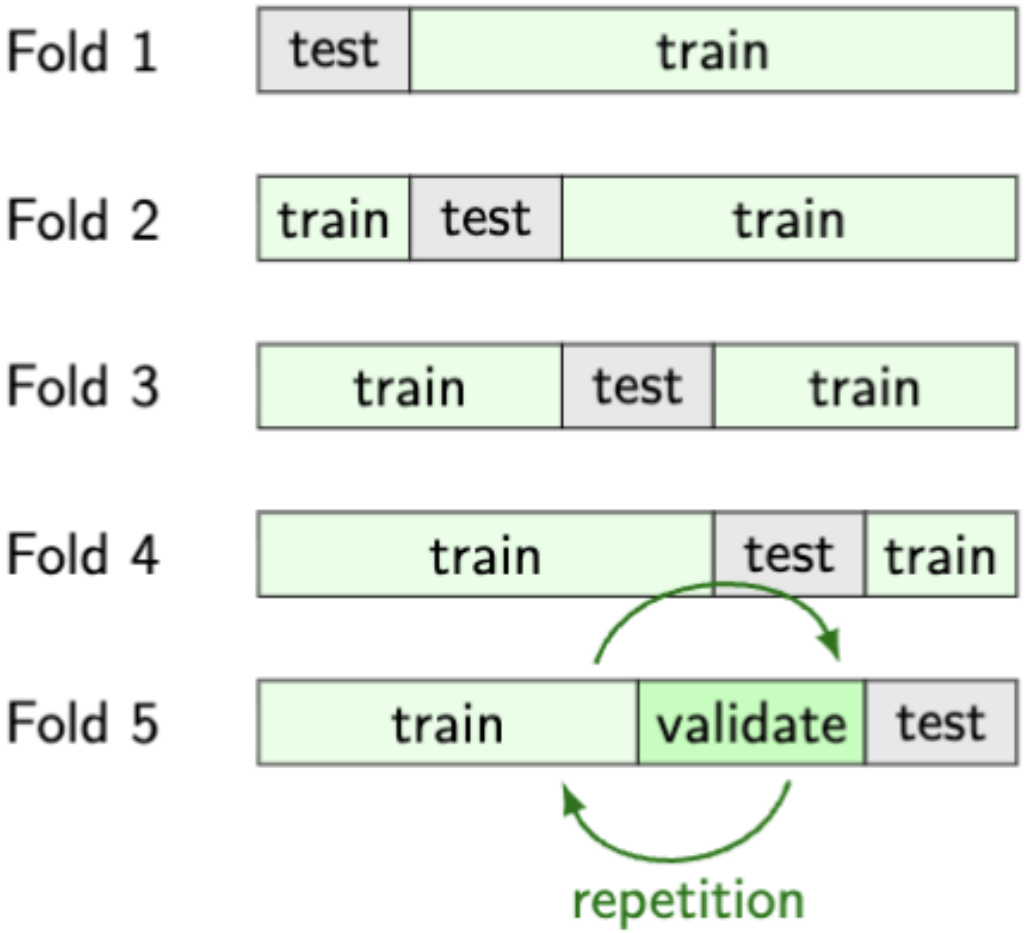


Overfitting in combined algorithm selection and hyperparameter optimization

Best paper
Runner up

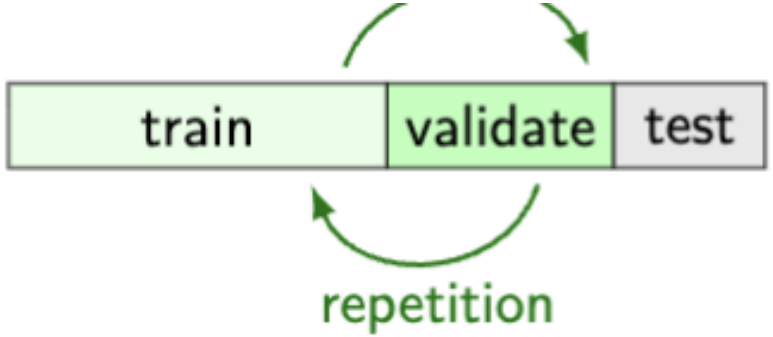
Schröder, S., Baratchi, M., and van Rijn, J., "Overfitting in combined algorithm selection and hyperparameter optimization." International Symposium on Intelligent Data Analysis, 2025.

Nested k-fold cross-validation

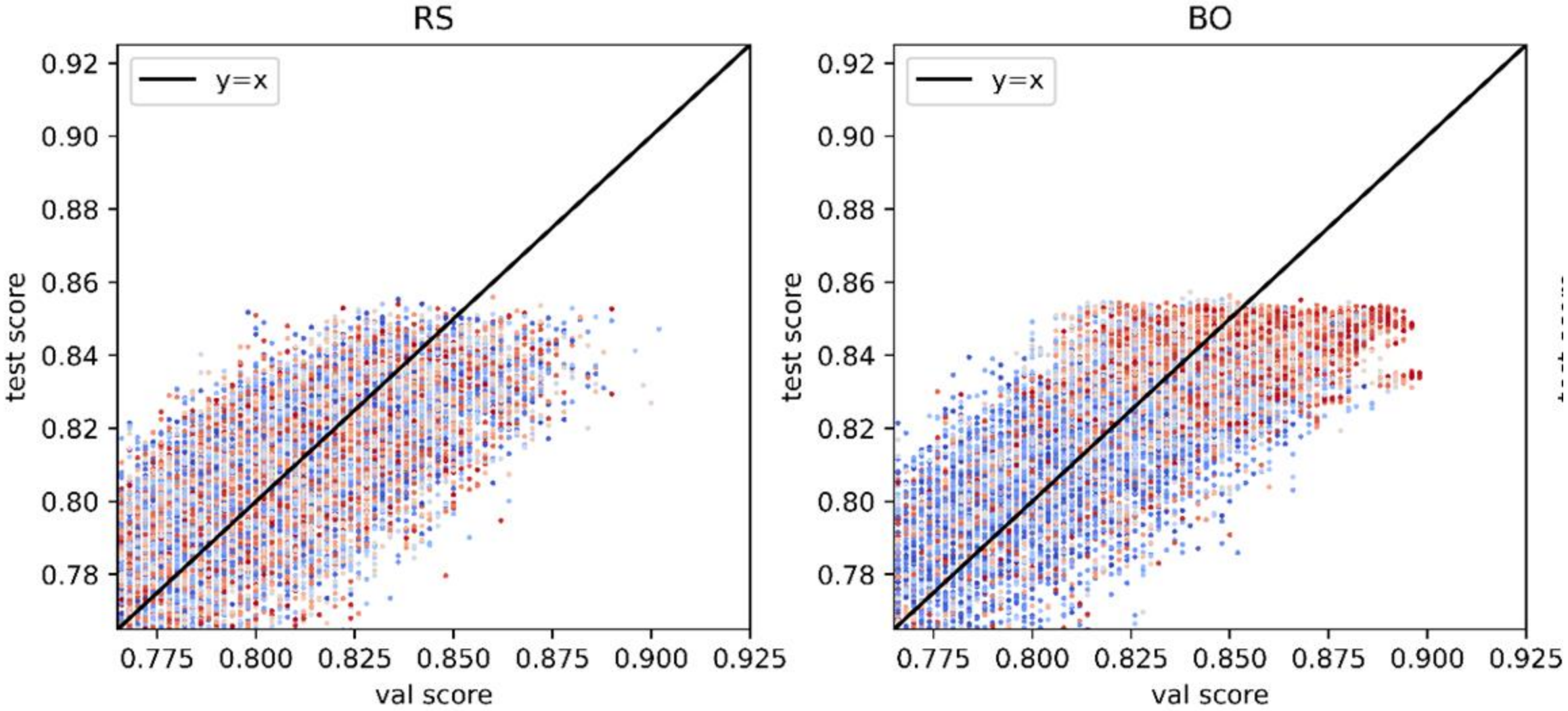


Understanding overfitting

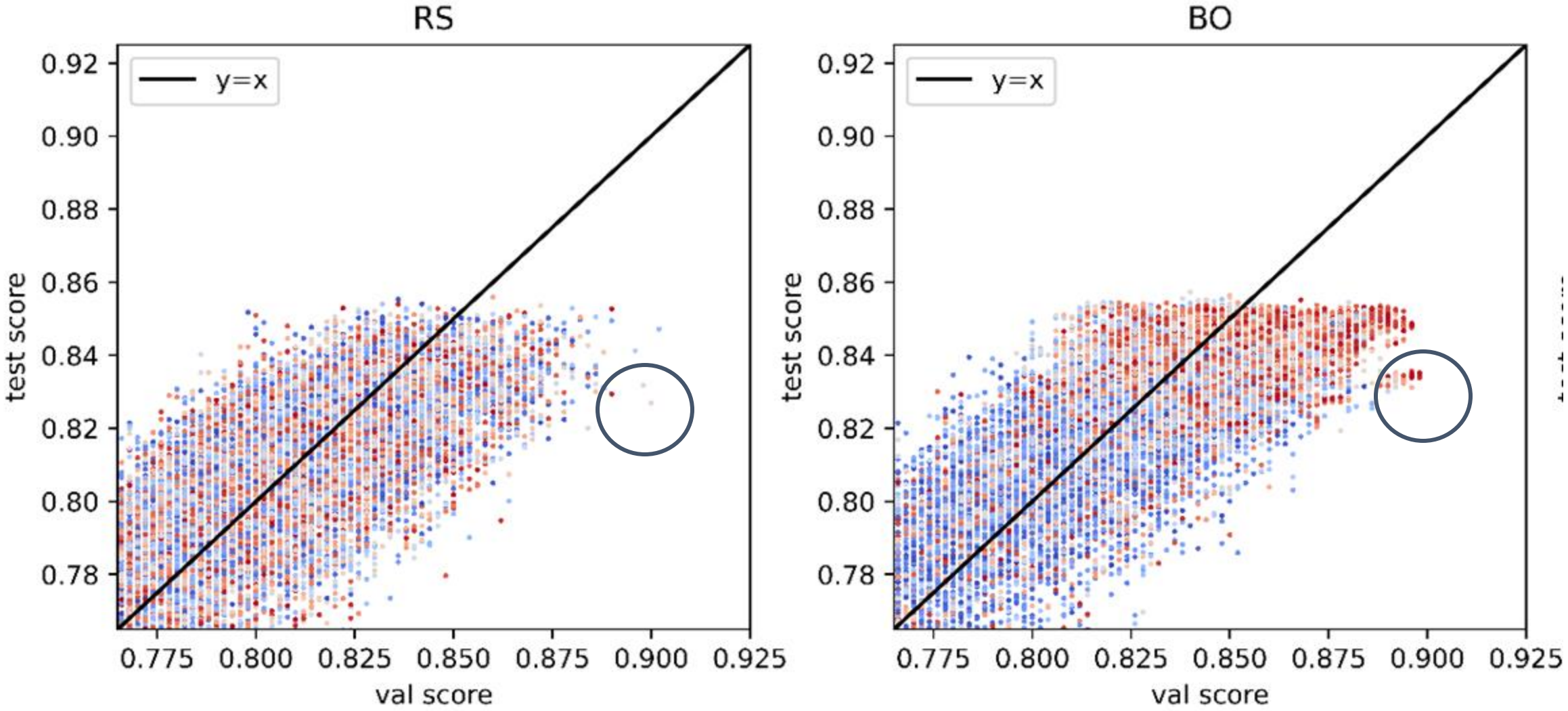
Meta-overfitting error (MOE) = validation performance - test performance



Meta-overfitting



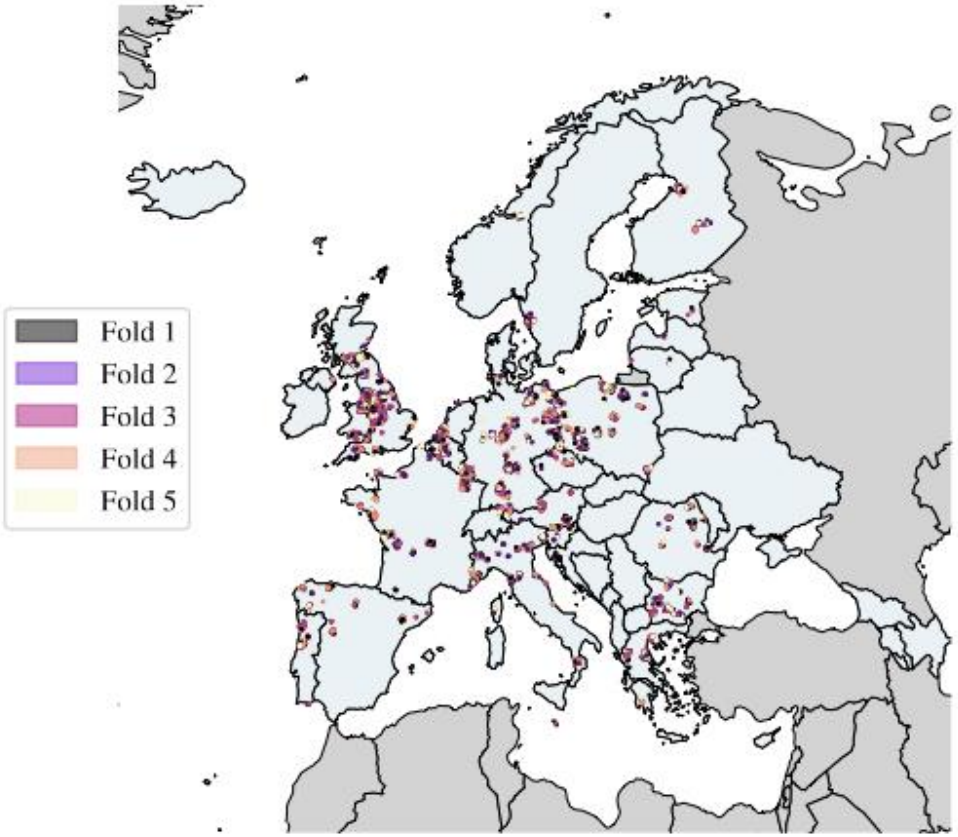
Selecting overfitting configurations



Model selection for spatial datasets

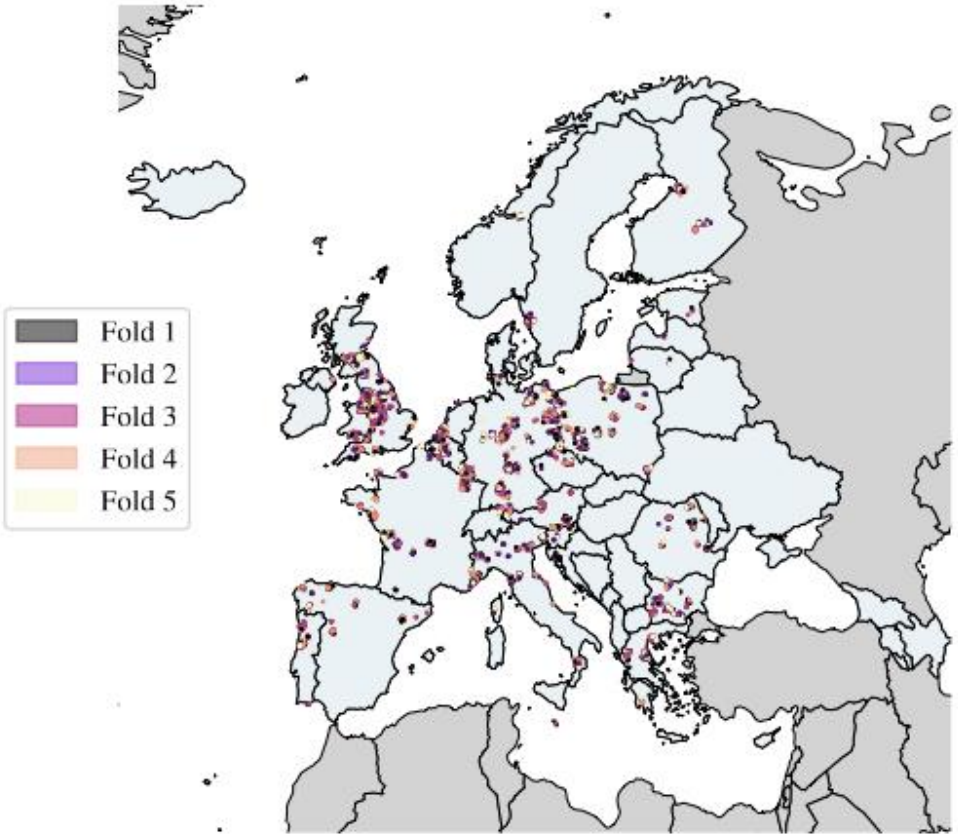
Stegena, A., Wasala, J., van Rijn, J., Baratchi, M. "Spatial Cross-Validation for Machine Learning Systems: towards comprehensive geospatial dataset evaluation", in preparation, 2026.

Model selection approaches

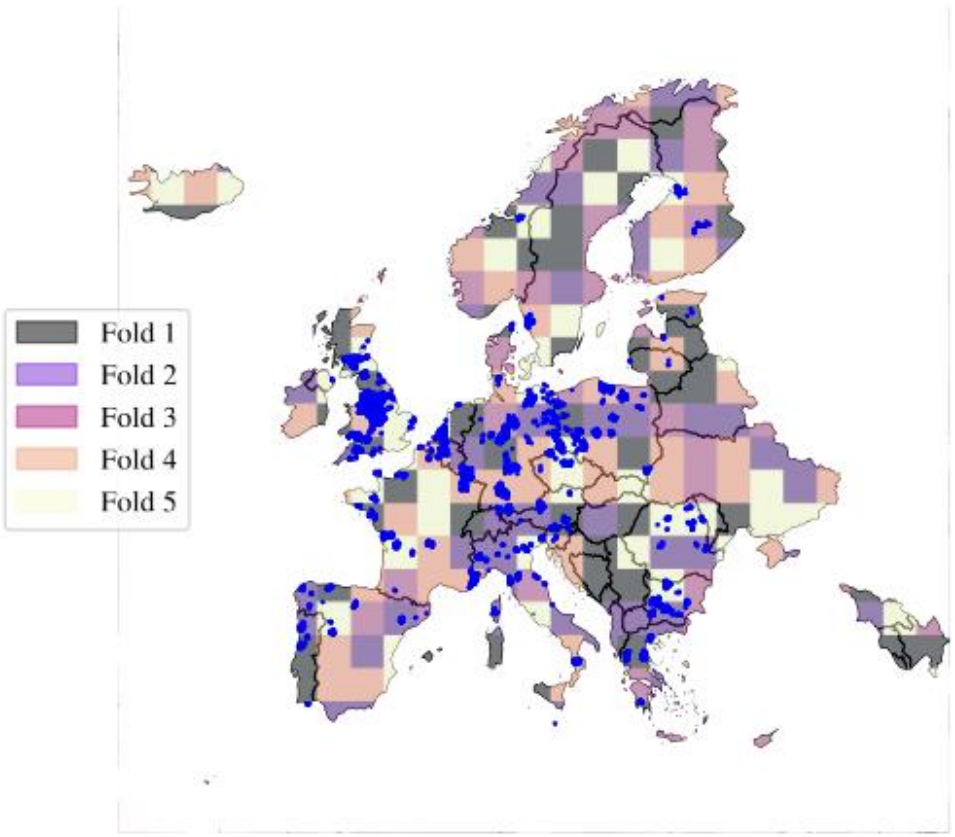


Random cross-validation

Model selection approaches

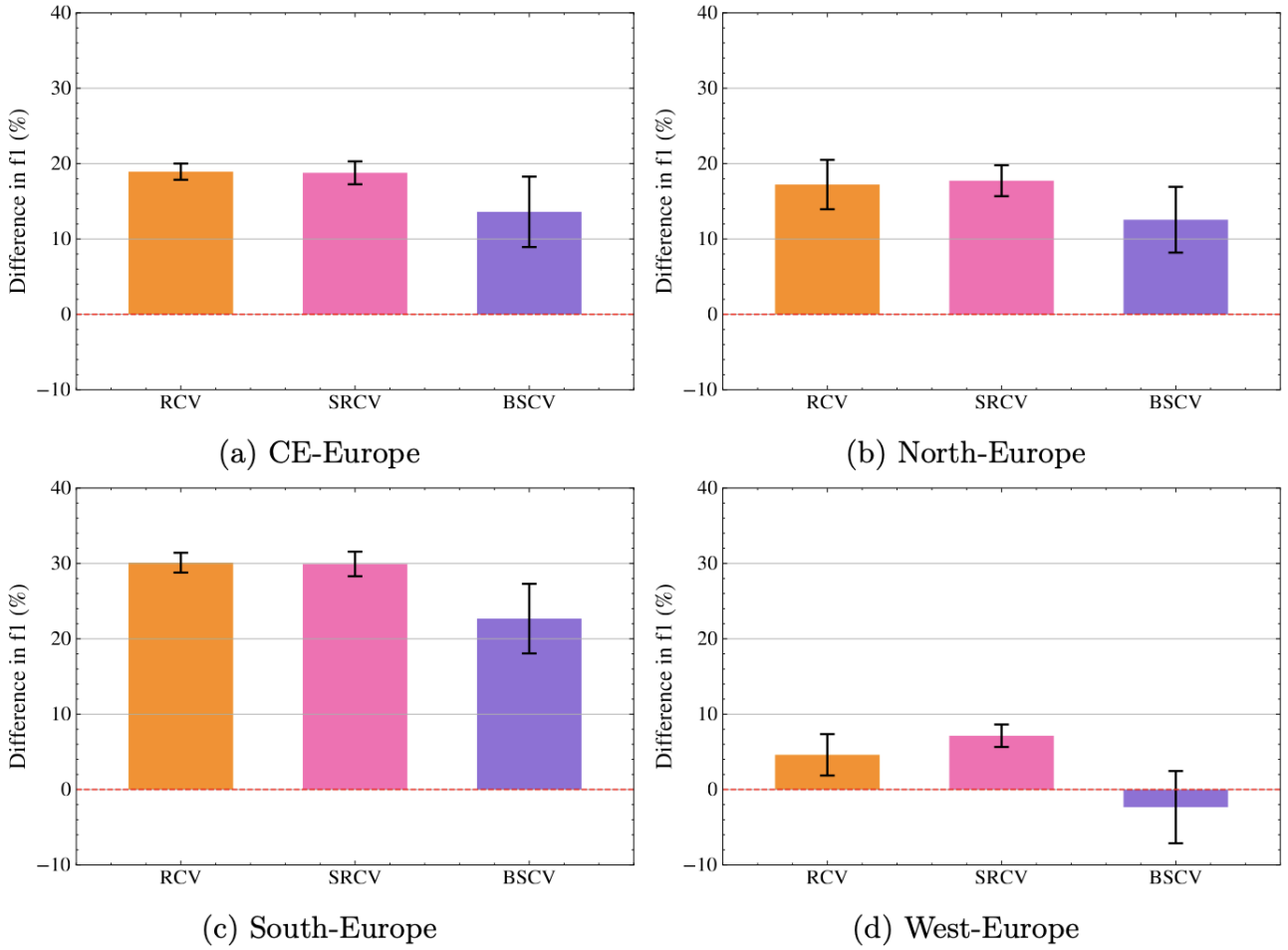


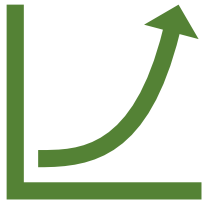
Random cross-validation



Spatial cross-validation

Performance overestimation





Performance evaluation

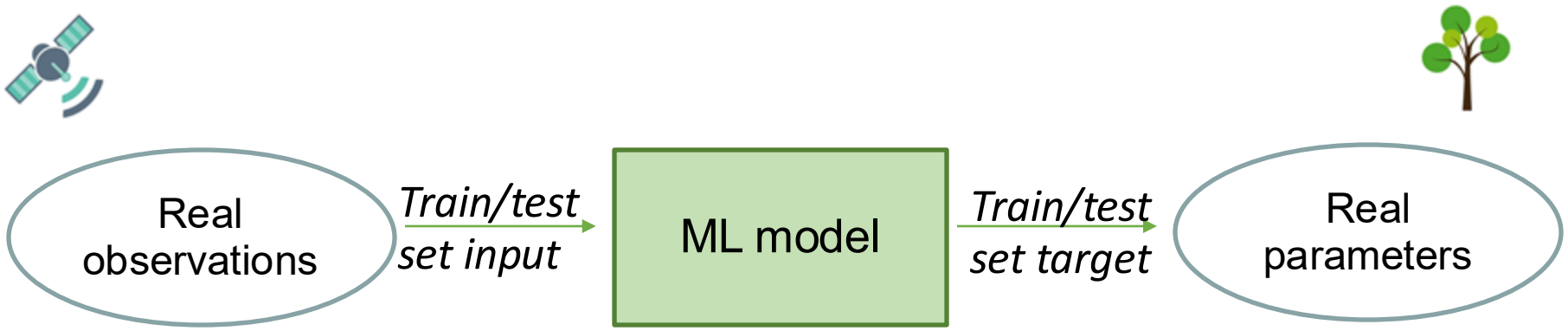
Incorporating domain knowledge

Using domain knowledge encoded in physical models

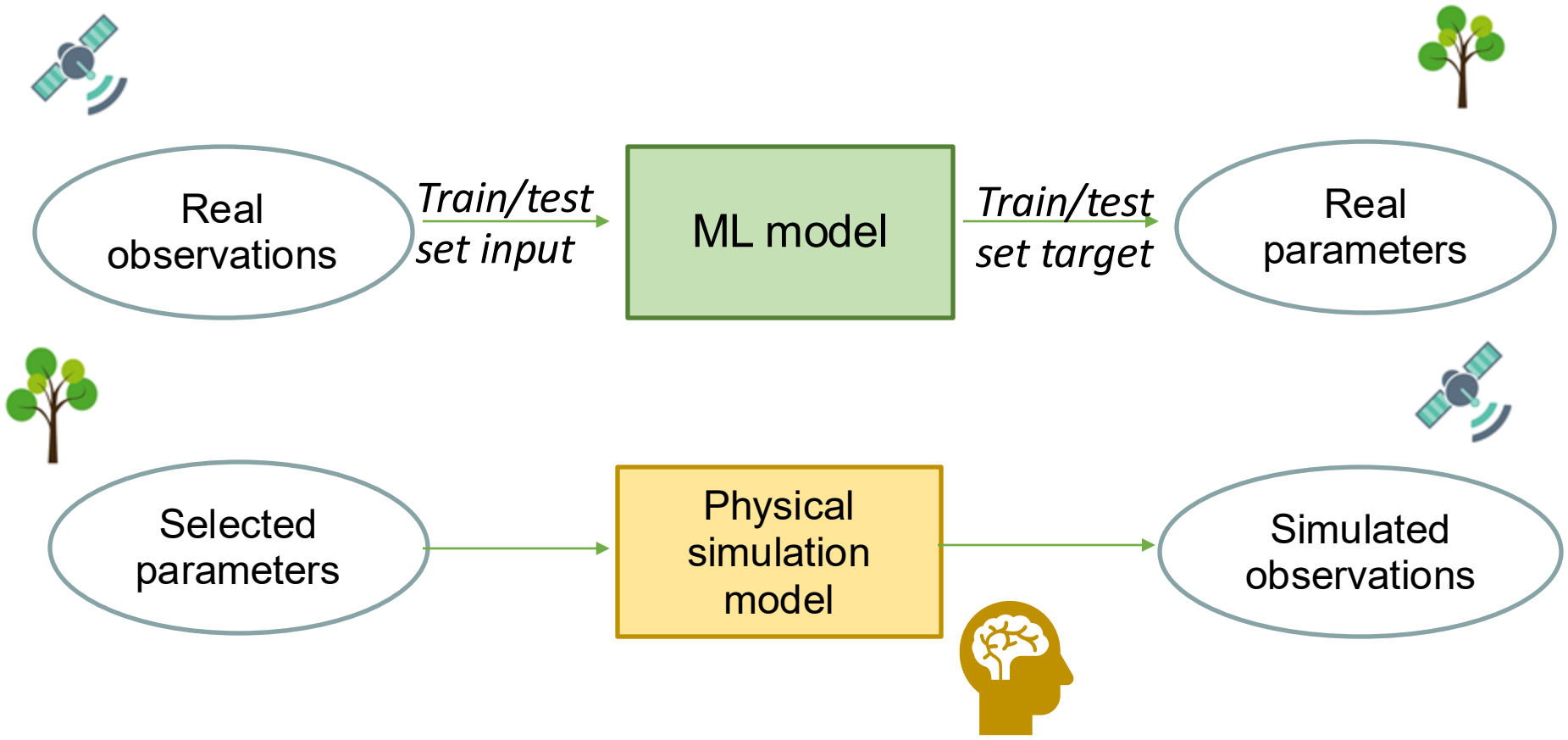
Arp, L., van Bodegom, P. M., Hoos, H. H., & Baratchi, M., Characterising the ill-posedness of PROSAIL inversion for biophysical parameter retrieval. European Journal of Remote Sensing, 59(1), 2026.

Arp L., van Bodegom P., Dang N., Francis A., Hoos, H., & Baratchi M., Inference from Noisy Observations through Model Inversion: Constructing ε-Manifolds of Potentially Valid Solutions". Under review by the Journal of Artificial Intelligence, 2026.

Parameter estimation in scientific problems



Parameter estimation in scientific problems



The problem

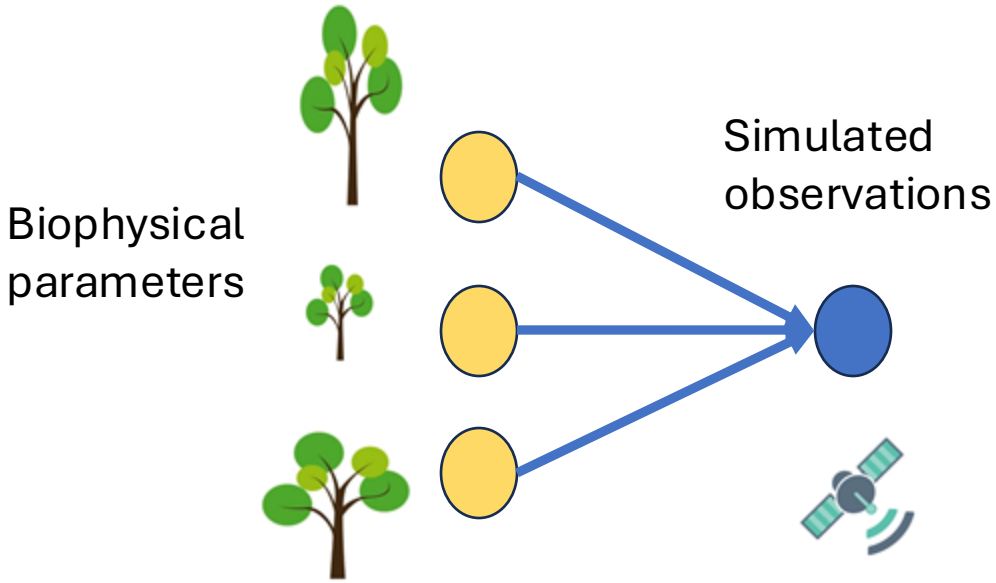
Physical model (e.g., PROSAIL model)

Vactor of ecological parameters

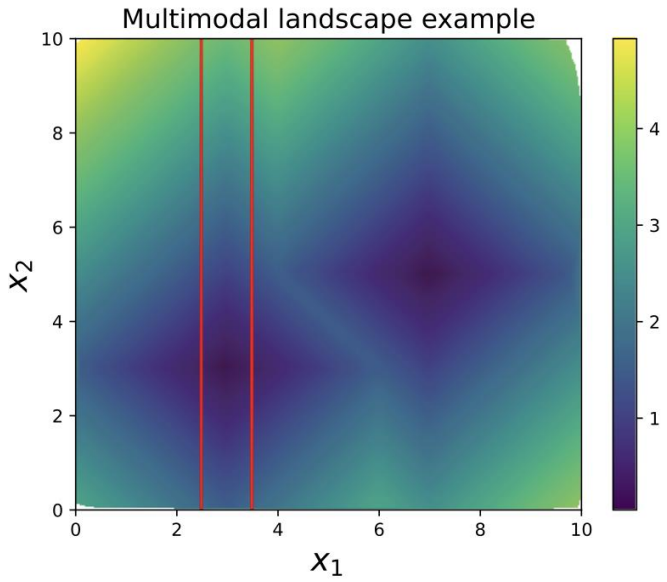
$$\boldsymbol{\theta} \in Arg \min_{\boldsymbol{\theta}} L\left(M(\boldsymbol{\theta}), x\right)$$

Real observation

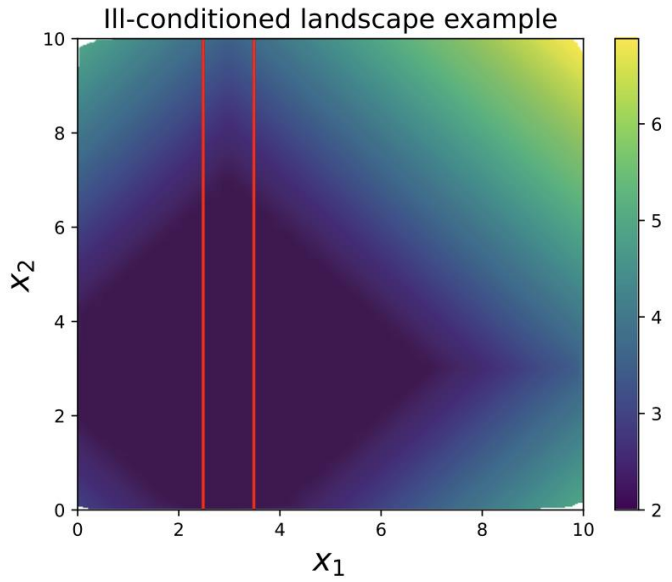
The ill-posedness problem



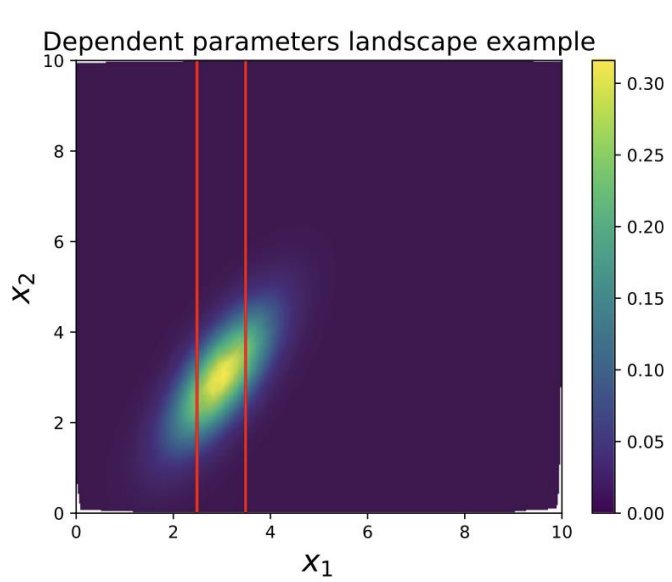
Understanding ill-posedness



(a) Multimodality (mechanism i))

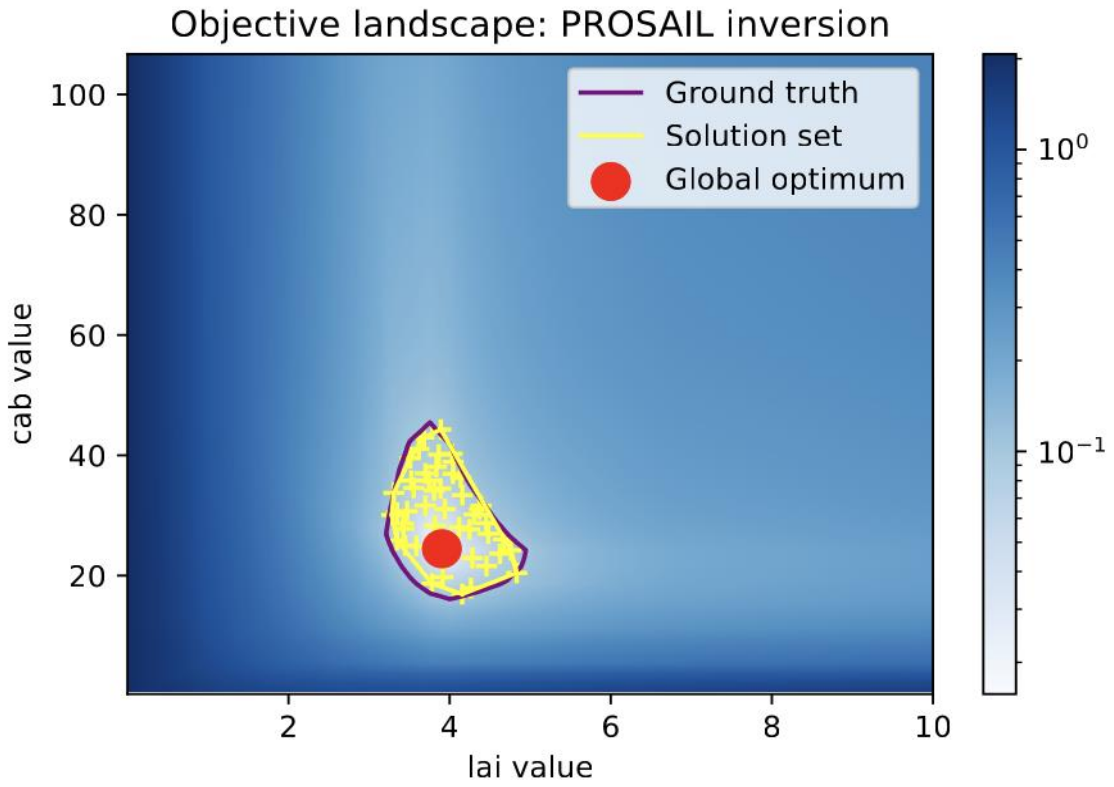


(b) Ill-conditioning (mechanism ii))



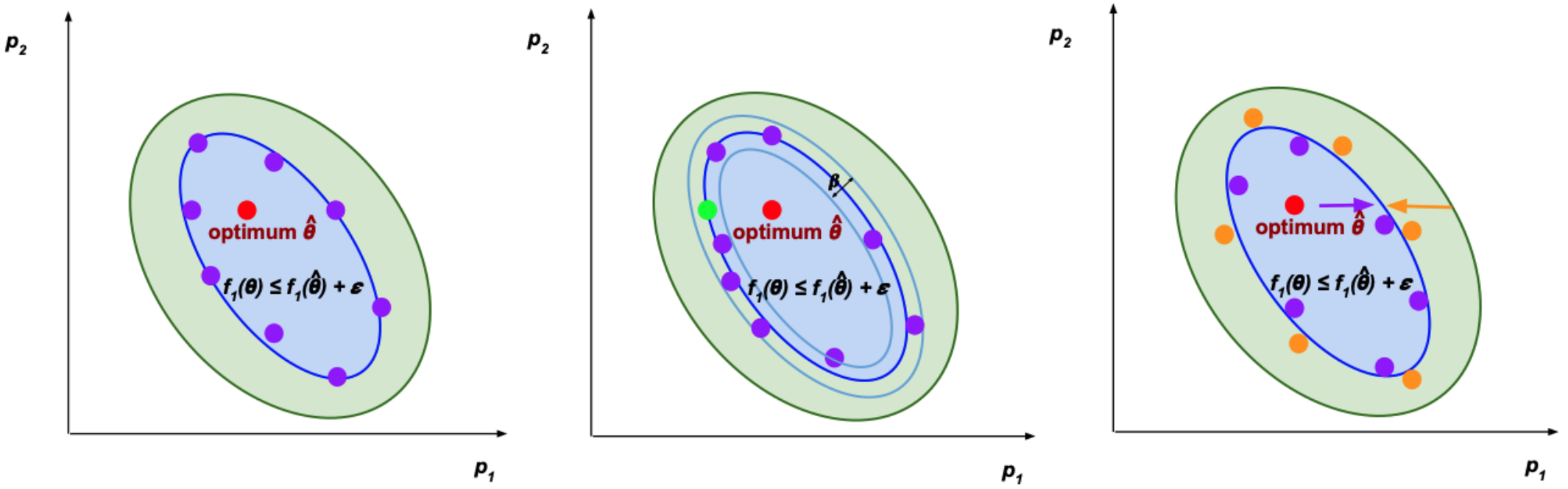
(c) Dependency (mechanism iii))

Automatic ill-posedness quantification



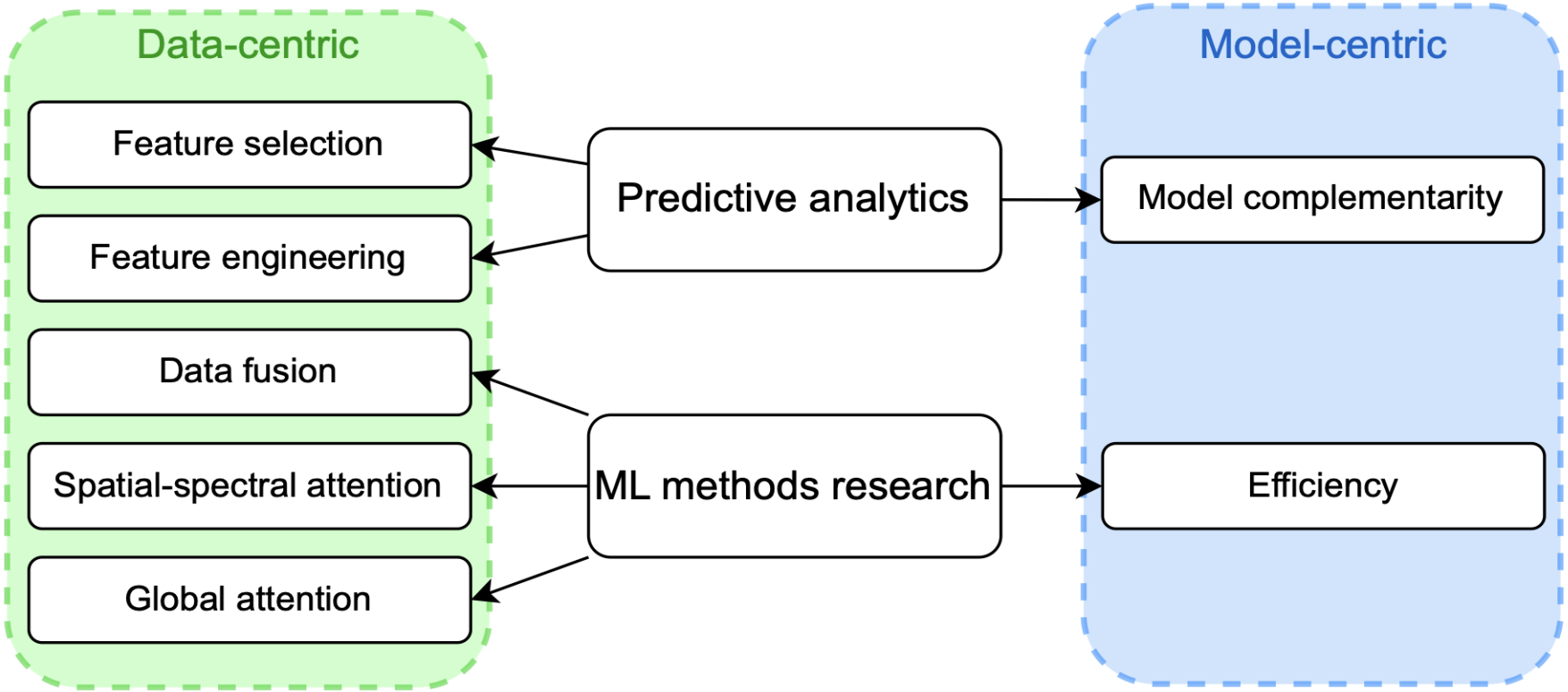
Search strategy

EMMI



(a) Conv-eMMI (b) Seq-eMMI (c) Dual-eMMI

AutoML research



Wąsala, J., Maasakkers, J. D., Schneider, R., Hoos, H. H., & Baratchi, M. (2026). AutoML to advance Earth Observation. *Journal of Data-centric Machine Learning Research*.

Outlook

- Improving evaluation
  - Spatial and spatio-temporal cross-validation
  - Out of distribution robustness
- Mitigating biases
  - Multi-objective optimisation (subgroup fairness)
- Transparency/expressiveness (domain user in the loop)
  - Improving interpretability
  - Are trained networks physically consistent

Lessons learnt

- AutoML can make advanced machine learning more accessible to domain users

But …

- There is a severe generalisation gap in automatically configured models
- Both data-centric and model-centric approaches are needed to close this gap

https://star.liacs.nl

https://autoai4eo.liacs.nl