Funded by the European Union

Coordinated by ECMWF

CopERnIcus climate change Service Evolution - CERISE

ML-BASED OBSERVATION OPERATORS FOR LAND-SURFACE DATA ASSIMILATION

5th ECMWF-ESA Machine Learning Workshop, Bologna, 15th April 2026

Pete Weston (ECMWF) & Patricia de Rosnay (ECMWF)

Contributors: Åsmund Bakketun, Jostein Blyverket, Cyril Palerme (Met Norway), Filipe Aires, Iris de Gelis, Catherine Prigent, Carlos Jimenez (Estellus)

CERISE project 🤝 ECMWF strategy

Funded by the European Union

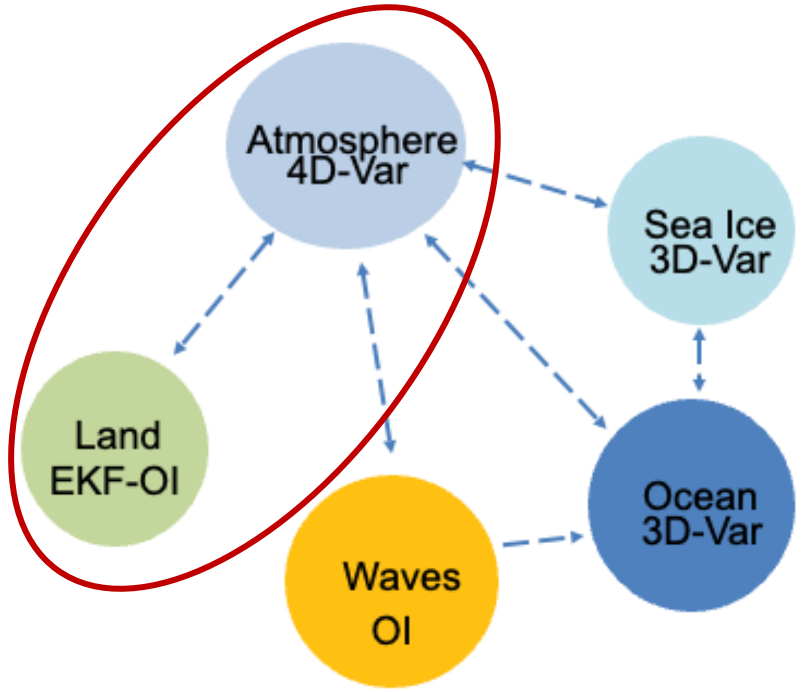
Support the long-term evolution of C3S for:

- **regional and global climate reanalysis and**
- **multi-system seasonal prediction,**

Towards an Earth system approach,
with a focus on land-atmosphere coupling.

- **Coupled data assimilation**
  - **Using physics-based and machine learning-based approaches**
- **Exploitation of interface observations**
  - **All-surface data assimilation**

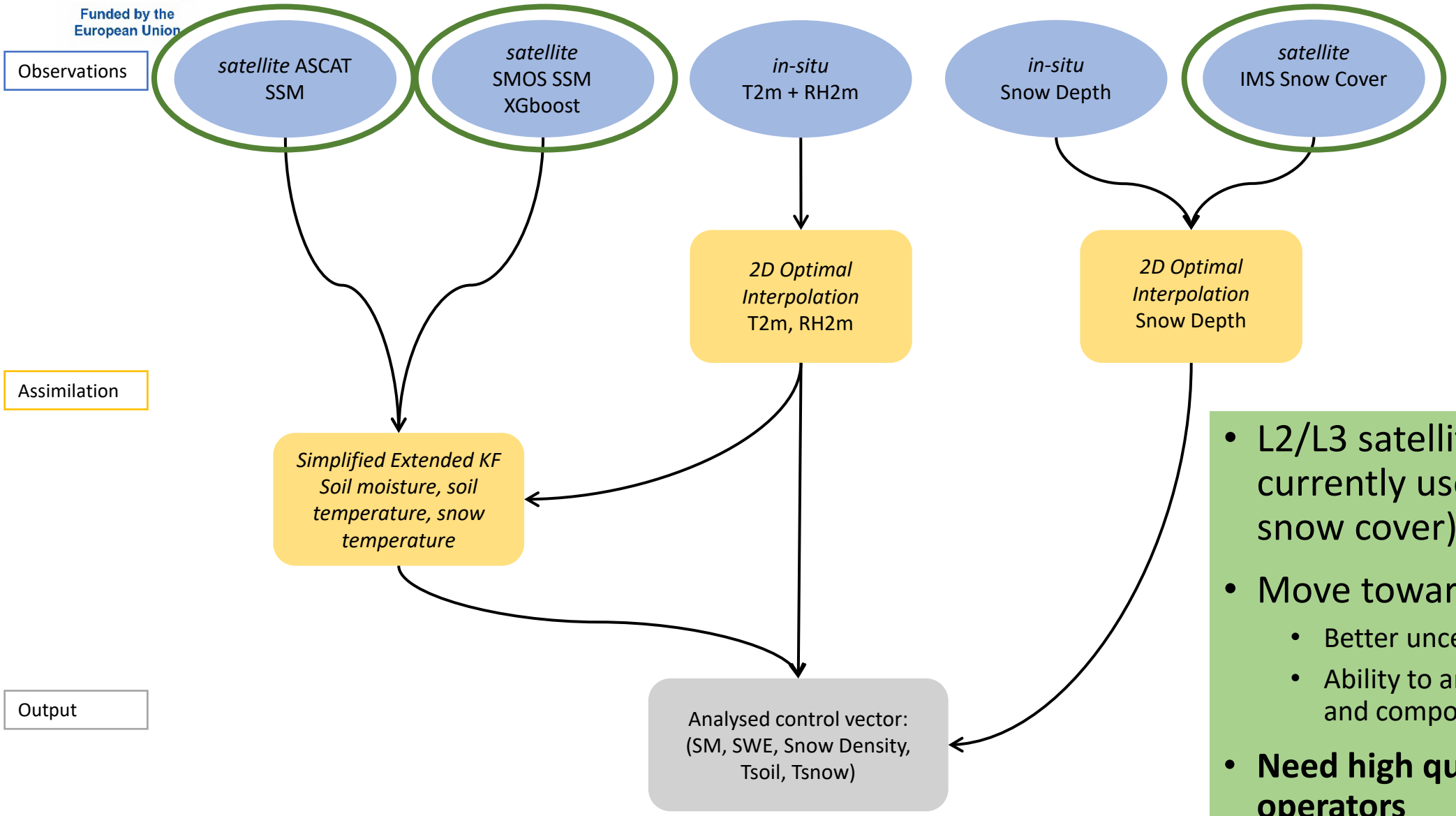
Earth system approach



de Rosnay et al., 2022

ECMWF's LDAS in 2026

Funded by the European Union

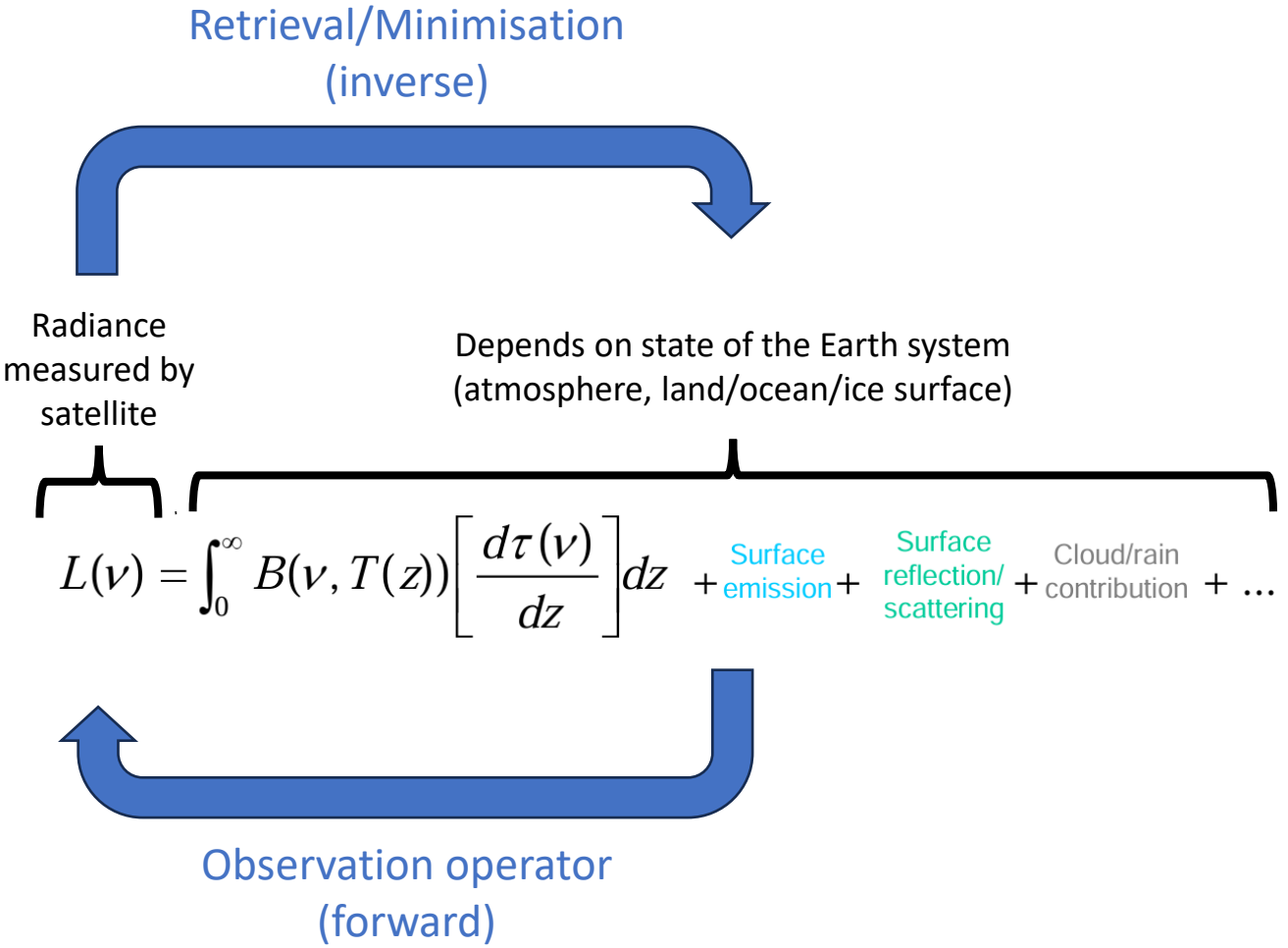


- L2/L3 satellite products currently used (soil moisture, snow cover)
- Move towards L1:
  - Better uncertainty characterisation
  - Ability to analyse multiple variables and components simultaneously
- **Need high quality observation operators**

Funded by the European Union

Forward vs Inverse problem

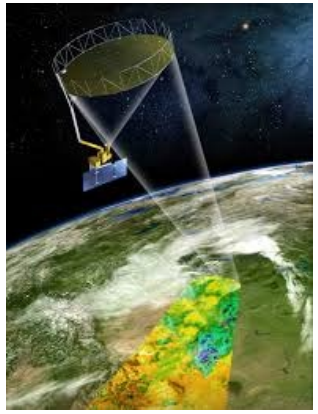
- Traditionally physics-based methods have been used for both:
  - Forward: RTTOV, CMEM etc.
  - Inverse: 1D-Var, 4D-Var, EnKF etc.
- ML-based methods increasingly used:
  - Forward: ML-based observation operators (**this talk**)
  - Inverse: SMOS SM XGboost, AI-DOP etc.

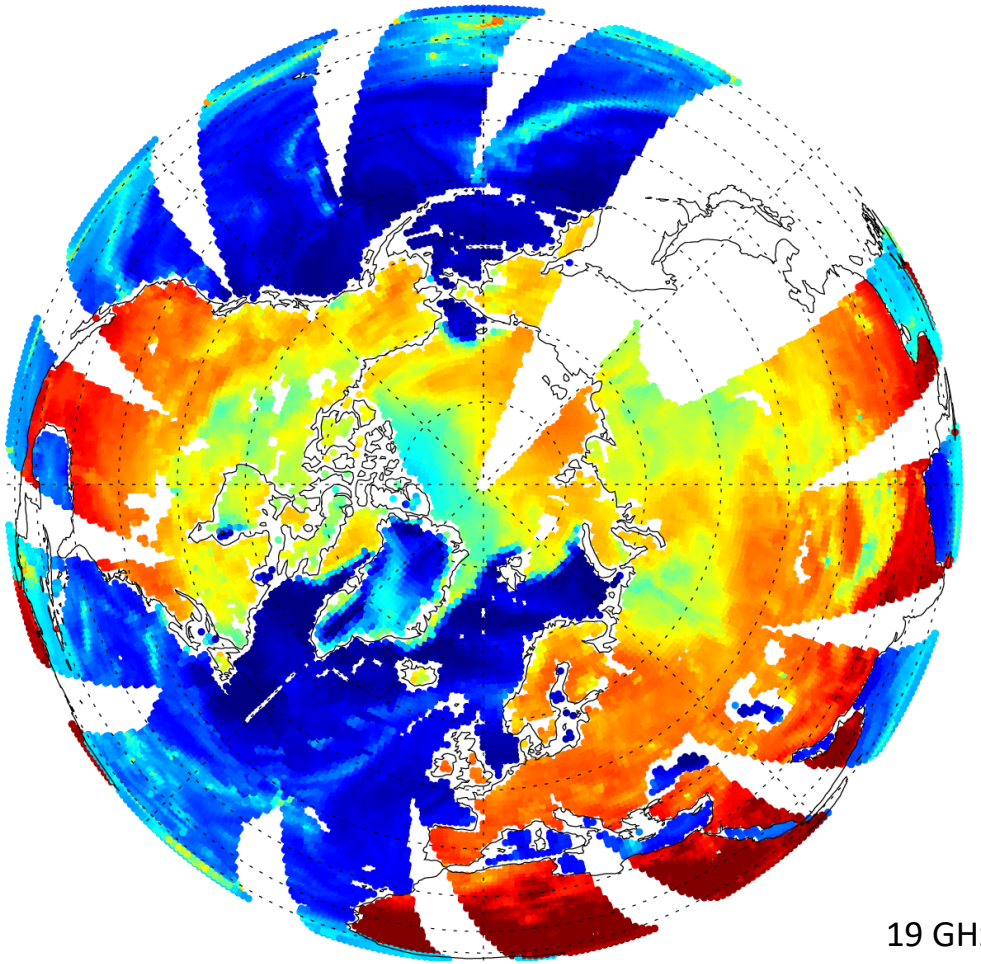


Funded by the European Union

Target frequencies and instruments

- Targeting low frequency passive microwave with sensitivity to:
  - Soil moisture
  - Vegetation
  - Snow properties
  - (Sea-ice and SST)
- 1.4-36GHz channels on SMOS, SMAP, AMSR2, GMI, MWI, CIMR etc.





19 GHz V,
brightness temperatures

A ML approach to model large scale surface contributions - Methodology

Same method for snow-free and snow-covered areas, but with different predictors, resulting in two operators.

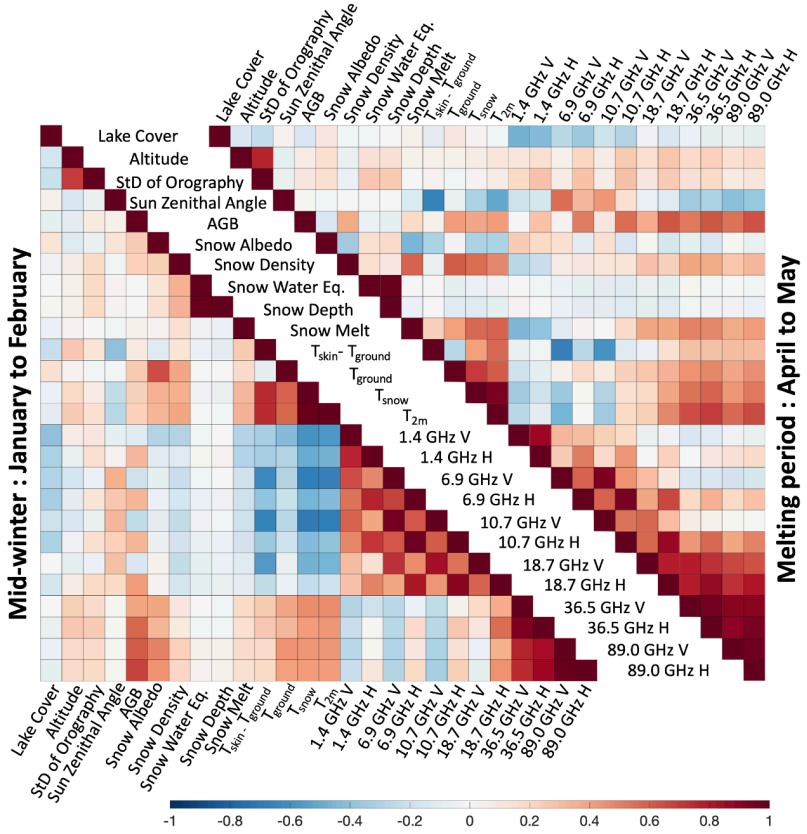
- **Methodology :**
  - ➜ Satellite-derived microwave emissivity database : SMAP, SMOS and AMSR2
  - ➜ Select potential model predictors to parameterise the emissivities
  - ➜ Statistically relate the satellite-derived emissivities to relevant geophysical parameters at global scale
    - ➜Use of machine-learning methods to account for complex relationships between geophysical parameters and emissivity

de Gelis et al., 2025

Funded by the European Union

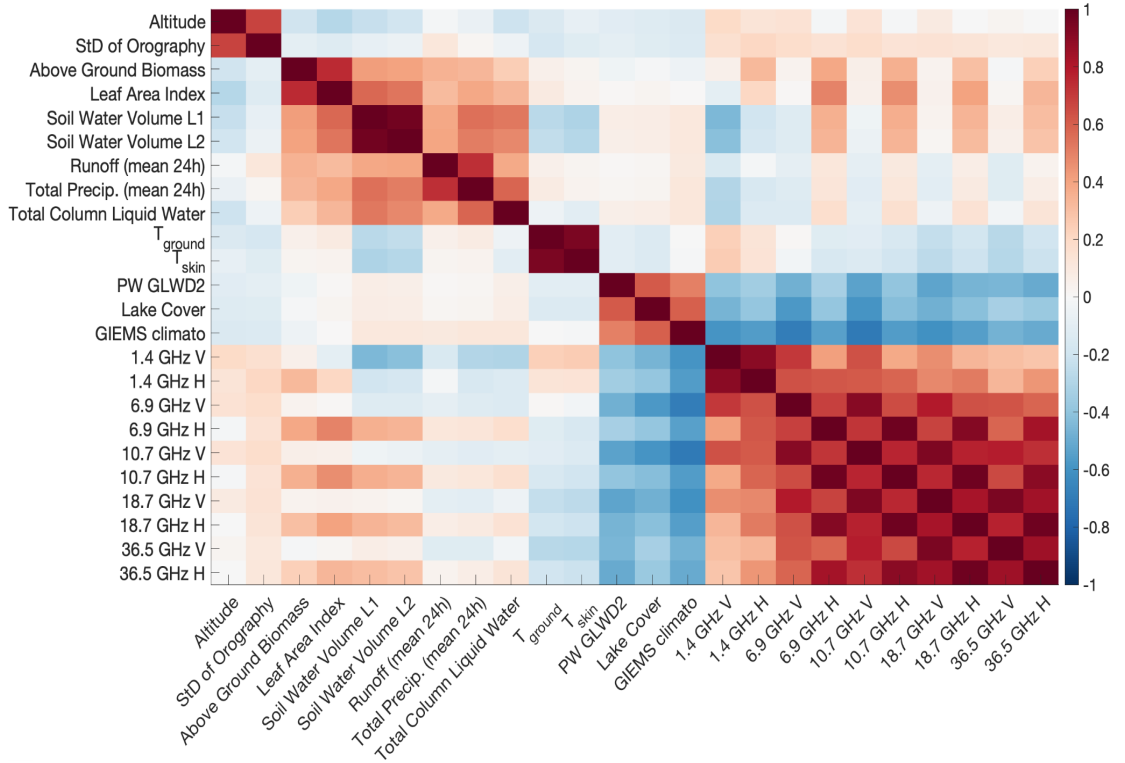
Feature importance and selection

Snow-covered land



Geophysical parameters		
Fixed	Snow	Temperature
Lake Cover StD of Orography AGB	Albedo Density Depth Melt	Snow (T_{snow}) Ground (T_{ground}) Gradient ($T_{skin} - T_{ground}$)

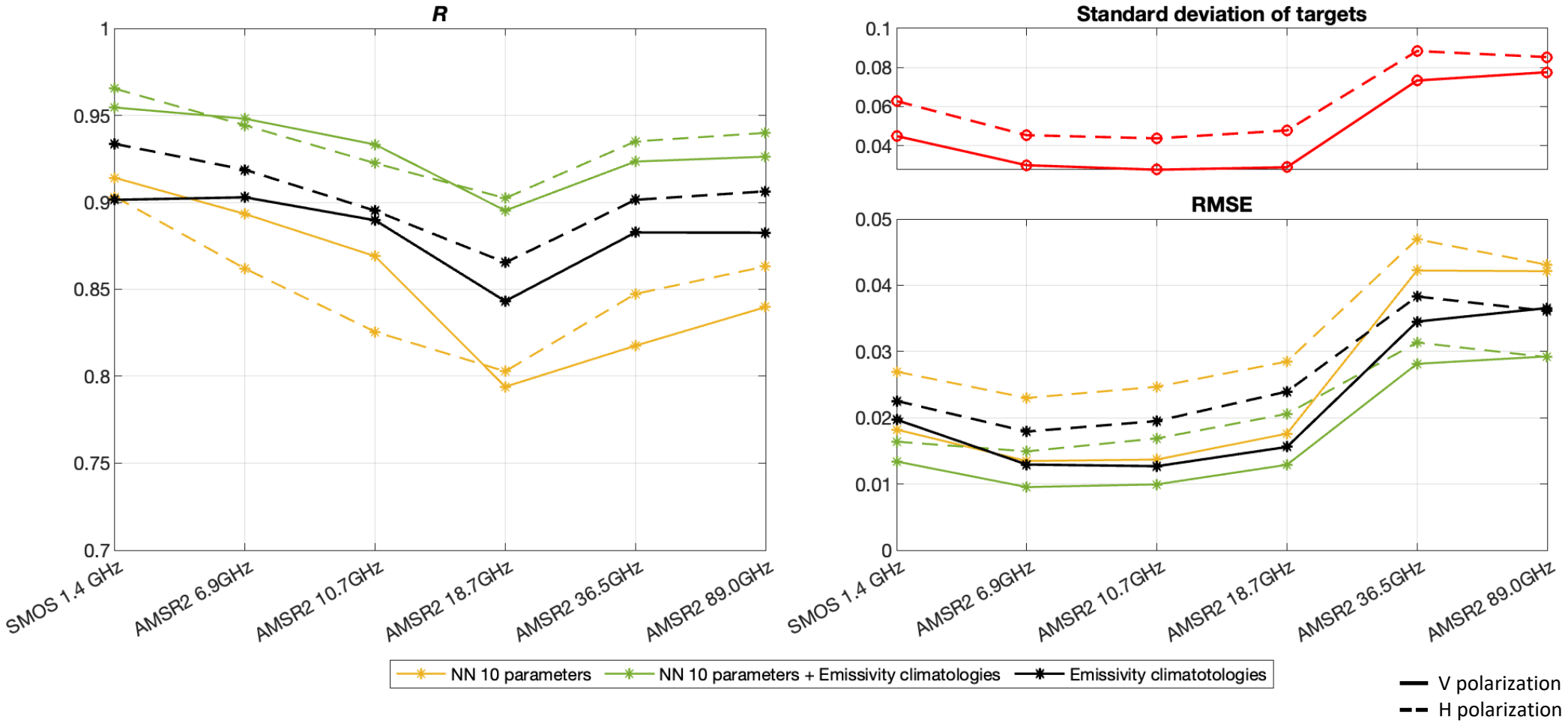
Snow-free land



Geophysical parameters		
Vegetation	Water and moisture presence	Other
AGB LAI	**Soil Water Volume L1** GIEMS climatology Permanent Water GLWD2	StD of Orography **Soil Temperature L1**

Funded by the European Union

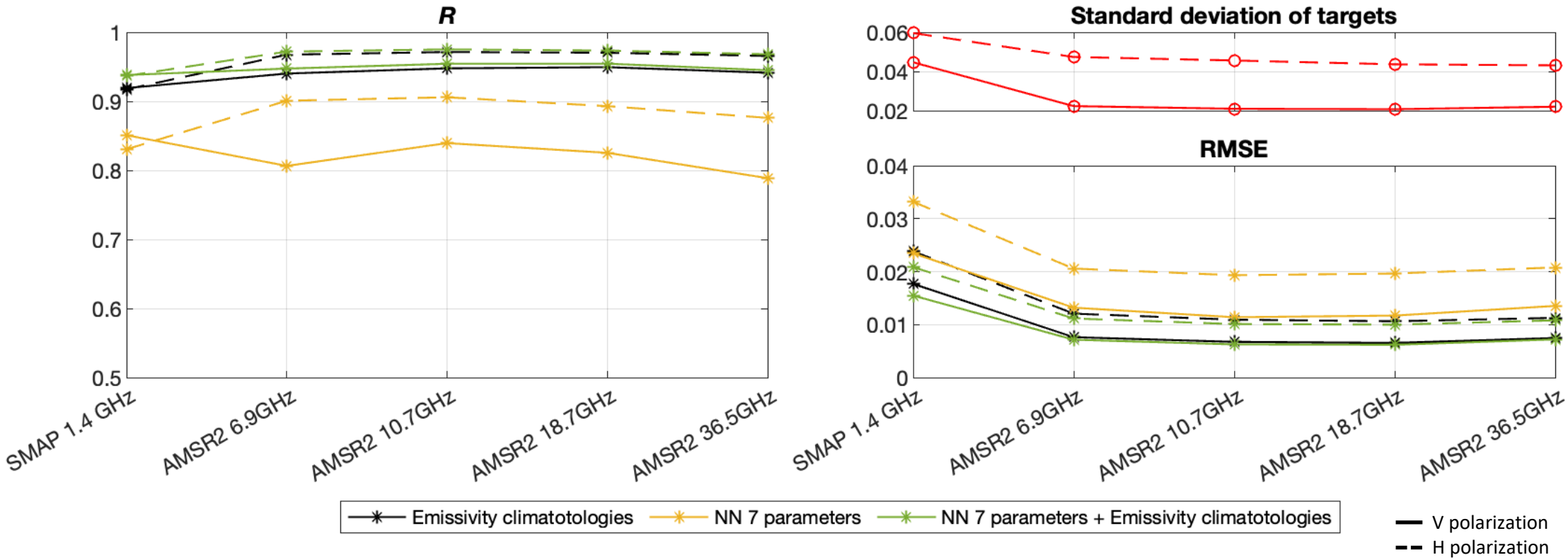
Snow-covered land results



- Combining the neural network and the emissivity climatologies gives best performance

Funded by the European Union

Snow-free land results

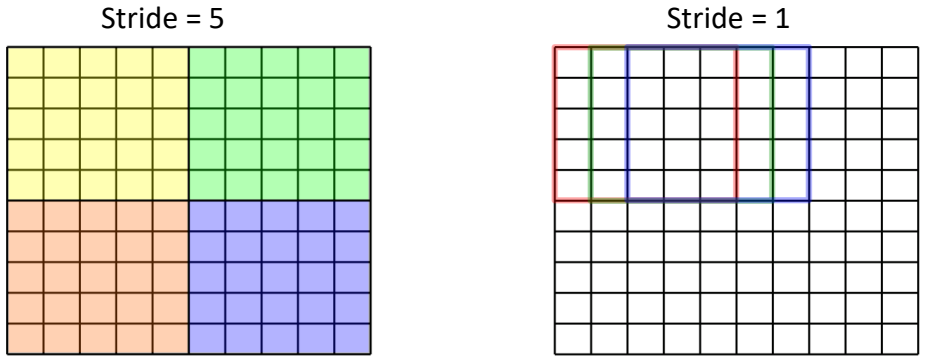


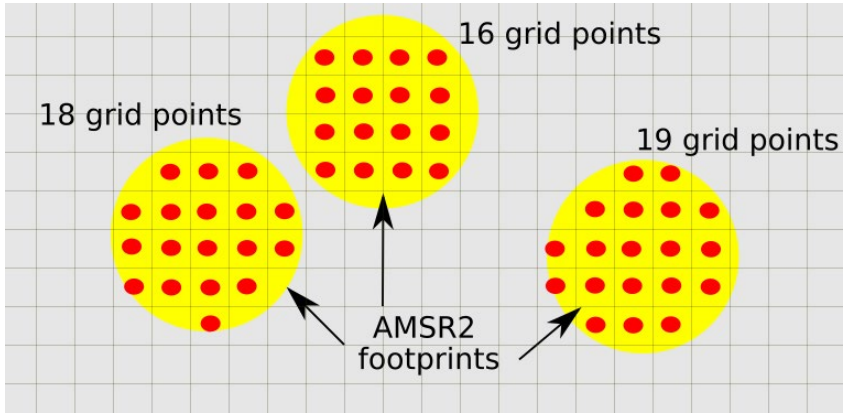
- Not much improvement over emissivity climatologies
- Possibility to improve results by using more dynamical vegetation information

Funded by the European Union

ML observation operator in regional HARMONIE-AROME

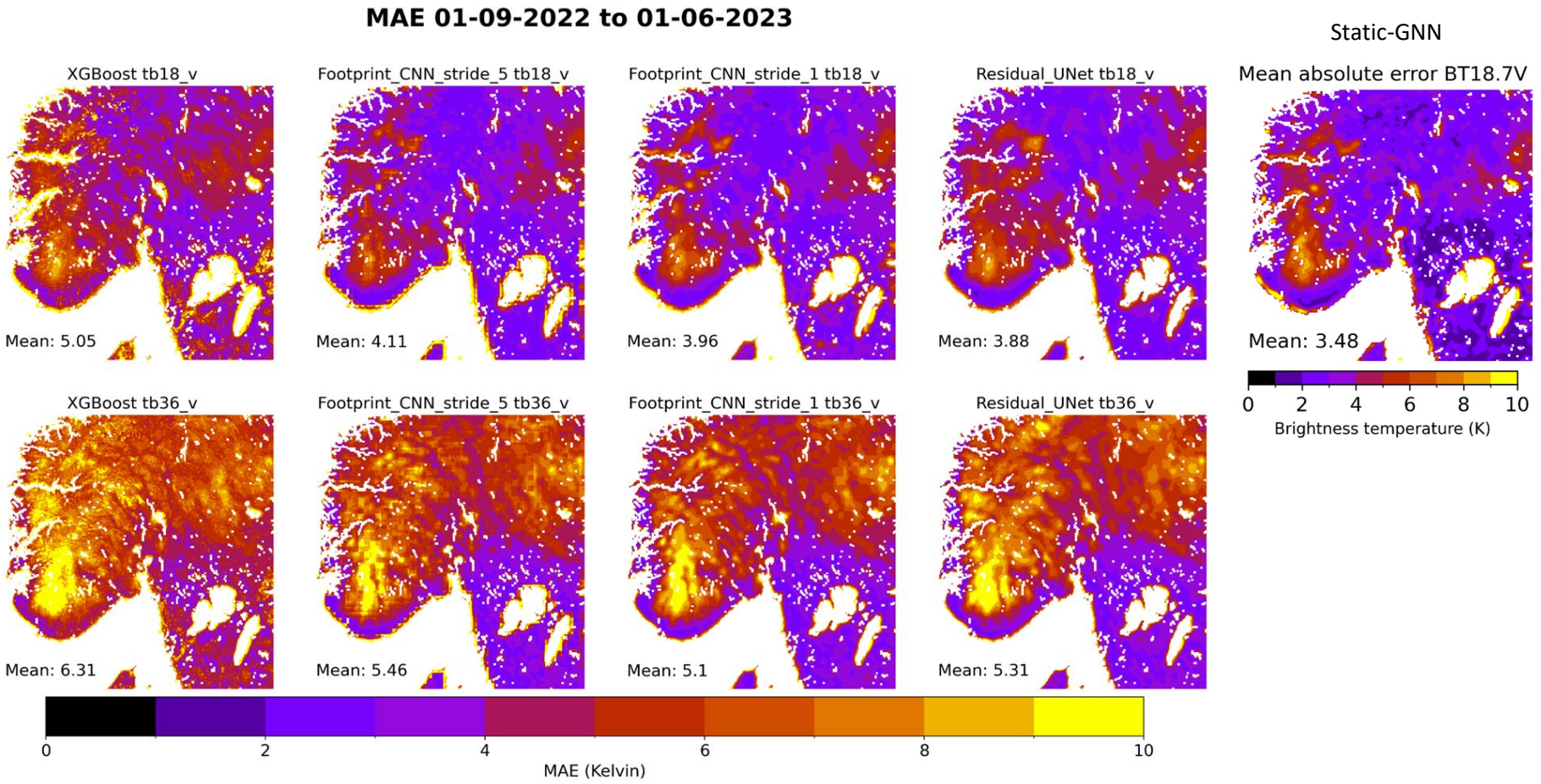
- Several machine learning methods were considered and tested
  - XGboost
    - No spatial information
  - Footprint Convolutional Neural Network (CNN)
    - Learn spatial features within the patches (figure to the top right)
  - Residual U-Net
    - Learn spatial features at different scales
    - Domain dependent
  - Static and dynamic graph neural networks (GNN)
    - Where each observation point has model information from the grid-cells within the satellite footprint (varies within the scan, dynamic) or keep static.
    - Add variables such as distance to footprint centre, allows the model to learn how to weight the different grid-cells within a footprint





Funded by the European Union

Comparison of different machine learning algorithms



Mean absolute errors for increasing machine learning model complexity from left to right. XGboost, footprint Convolutional Neural Network stride 5, footprint CNN stride 1, residual U-Net and static-GNN. The static-GNN has the lowest MAE for 18GHz V-pol for this time-period.

Static-GNN vs CMEM

- Comparison of the static-GNN vs the physical Community Microwave Emission Model (CMEM, de Rosnay et al., 2020)
- CMEM run on graph level, i.e., for each node we compute the footprint average value of the ISBA LSV inputs for fair comparison
- Bias:
  - CMEM large difference in H vs V-pol, smaller difference for the static-GNN
  - Comparable bias for CMEM vs static-GNN V-pol
  - Increasing bias amplitude for the static-GNN from August
- Mean absolute error (MAE):
  - Large errors in CMEM H-pol, considerably lower errors for static-GNN H-pol
  - static-GNN V-pol errors ~2 K lower than for CMEM V-pol

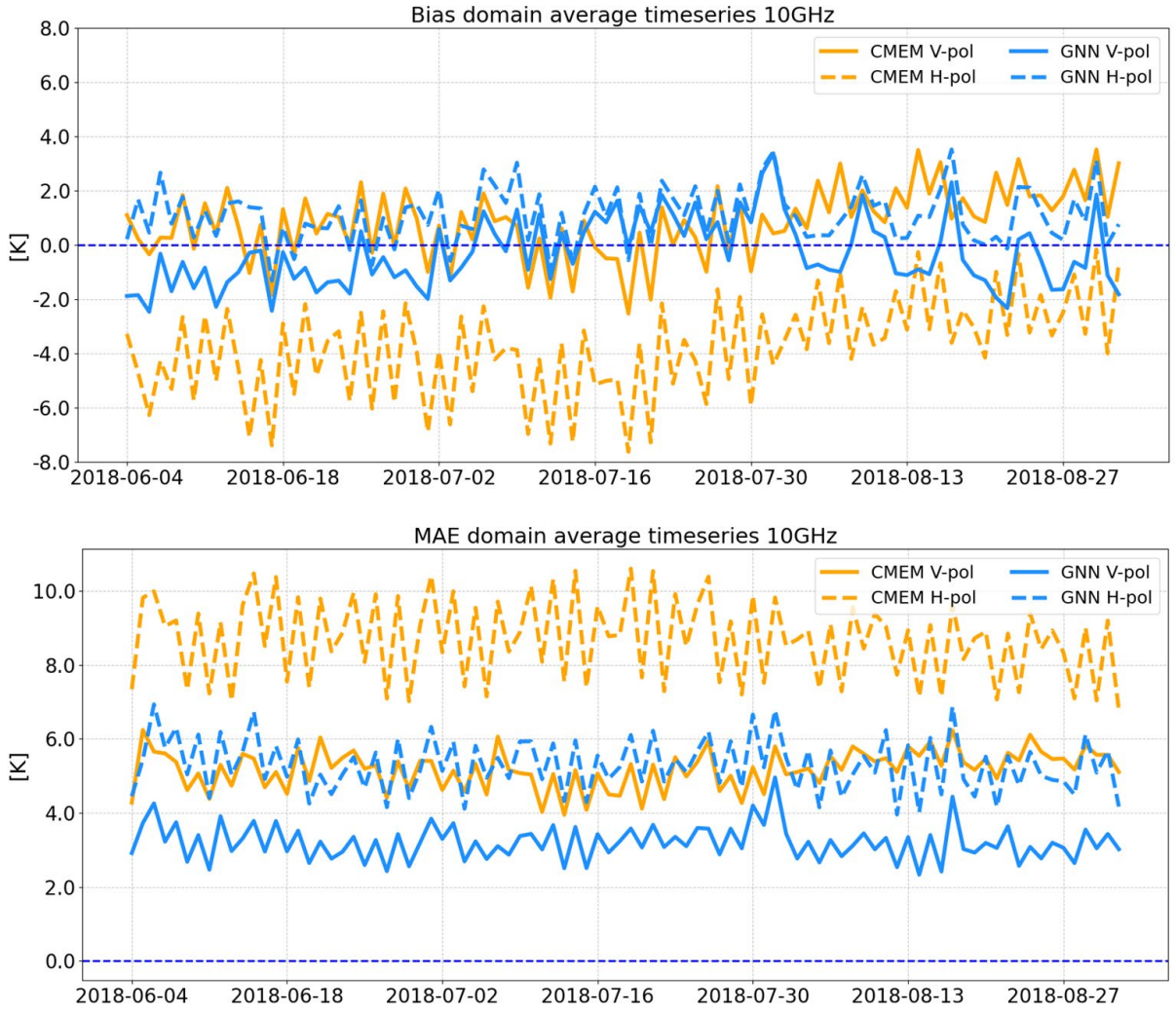


Figure: Domain average bias (top) and MAE (bottom) for CMEM (orange) and static-GNN (blue) V-pol (solid) and H-pol (dotted) covering the test period June, July and August 2018.

Funded by the European Union

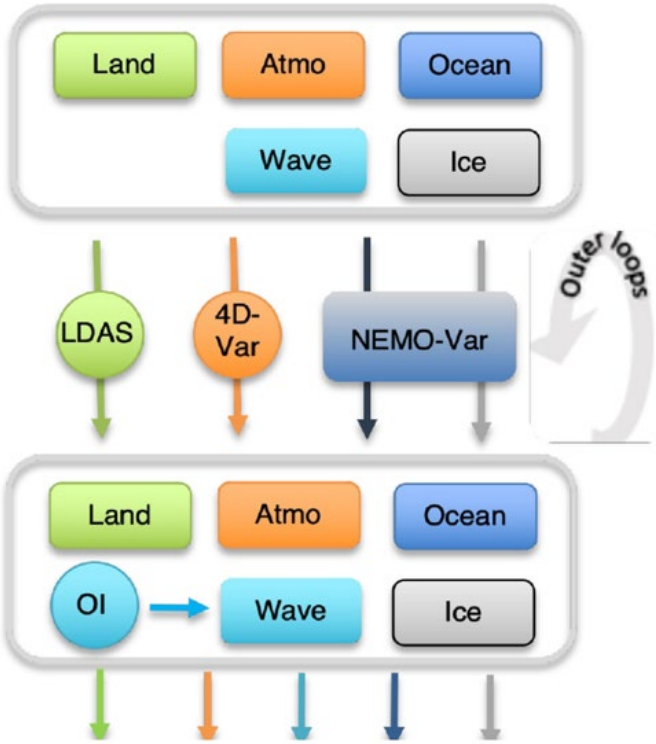
Conclusions

- ML-based observation operators have been developed for low frequency passive MW over global and regional domains
- Global model results:
  - Best results by combining with climatologies, especially over snow-covered land
  - Snow-free land results impacted by lack of dynamic vegetation information in training
- Regional model results:
  - Different ML methods have been compared:
    - Static graph neural network performs best
  - ML-based model outperforms state of the art physical model

Funded by the European Union

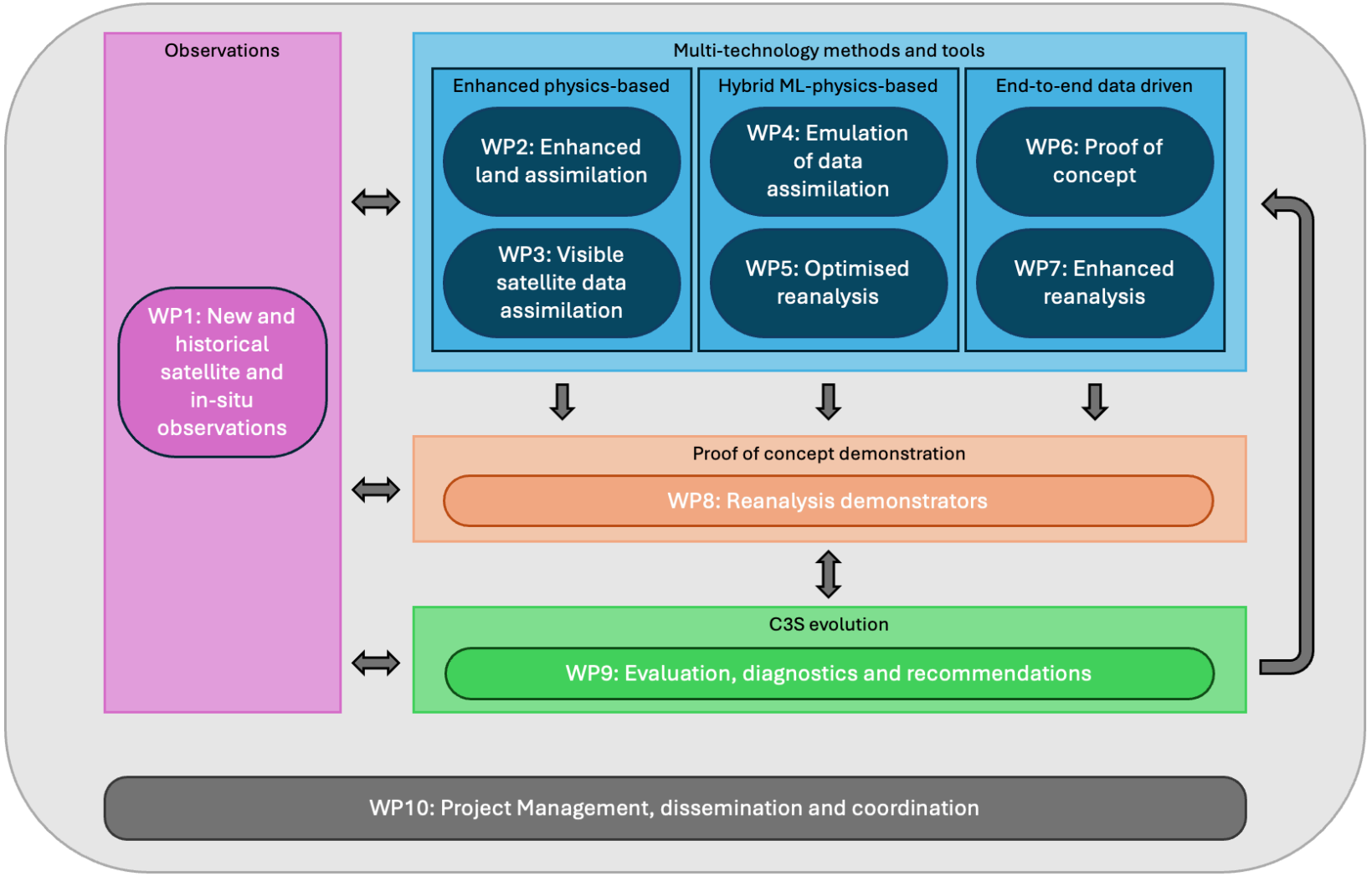
Perspectives

- Plan to implement global ML observation operator in IFS
- Run monitoring experiments to compare against CMEM
- Run assimilation experiments:
  - SMOS/SMAP L1 Tb assimilation for soil moisture
  - All-sky AMSR2 L1 Tb assimilation over snow and snow-free land
- Future work may involve extending Geer (2024) sea-ice approach over snow-covered and snow-free land
- **Long-term aim is to exploit full information content of satellite observations across all Earth system components via coupled assimilation framework**



What's next?

- **C**limate **H**azards **E**nhanced **R**eanalyses and **R**easoning with a**I**
- Exploiting new and historic satellite and in-situ data
- Multi-technology approach to reanalysis and DA
- Proof of concept demonstration
- Evaluation to inform design of future reanalyses



Funded by the European Union

References

- CERISE deliverable D1.4 https://www.cerise-project.eu/deliverables
- de Gelis et al, 2025 https://doi.org/10.1016/j.rse.2025.114821
- de Rosnay et al, 2020 https://doi.org/10.1016/j.rse.2019.111424
- de Rosnay et al, 2022 https://doi.org/10.1002/qj.4330
- Geer, 2024 https://doi.org/10.1002/qj.4797

Funded by the European Union

Coordinated by ECMWF

CopERnIcus climate change Service Evolution - CERISE

Thank you!

ECMWF · Norwegian Meteorological Institute · SMHI · Deutscher Wetterdienst (DWD) Wetter und Klima aus einer Hand · cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici · BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

METEO FRANCE · The Danish Meteorological Institute · Estellus · IPMA · NILU · Met Office

The CERISE project (grant agreement No 101082139) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.