Enhanced land surface data exploitation using machine learning in weather prediction systems

Patricia de Rosnay

Thanks to: Eulalie Boucher, Mariana Clare, David Fairbairn, Sébastien Garrigues, Christoph Herbert, Zdenko Heyvaert, Eleni Kalogeraki, Jana Kolassa, Tony McNally, Ewan Pinnington, Nina Raoult, Kirsti Salonen, Pete Weston, and many others

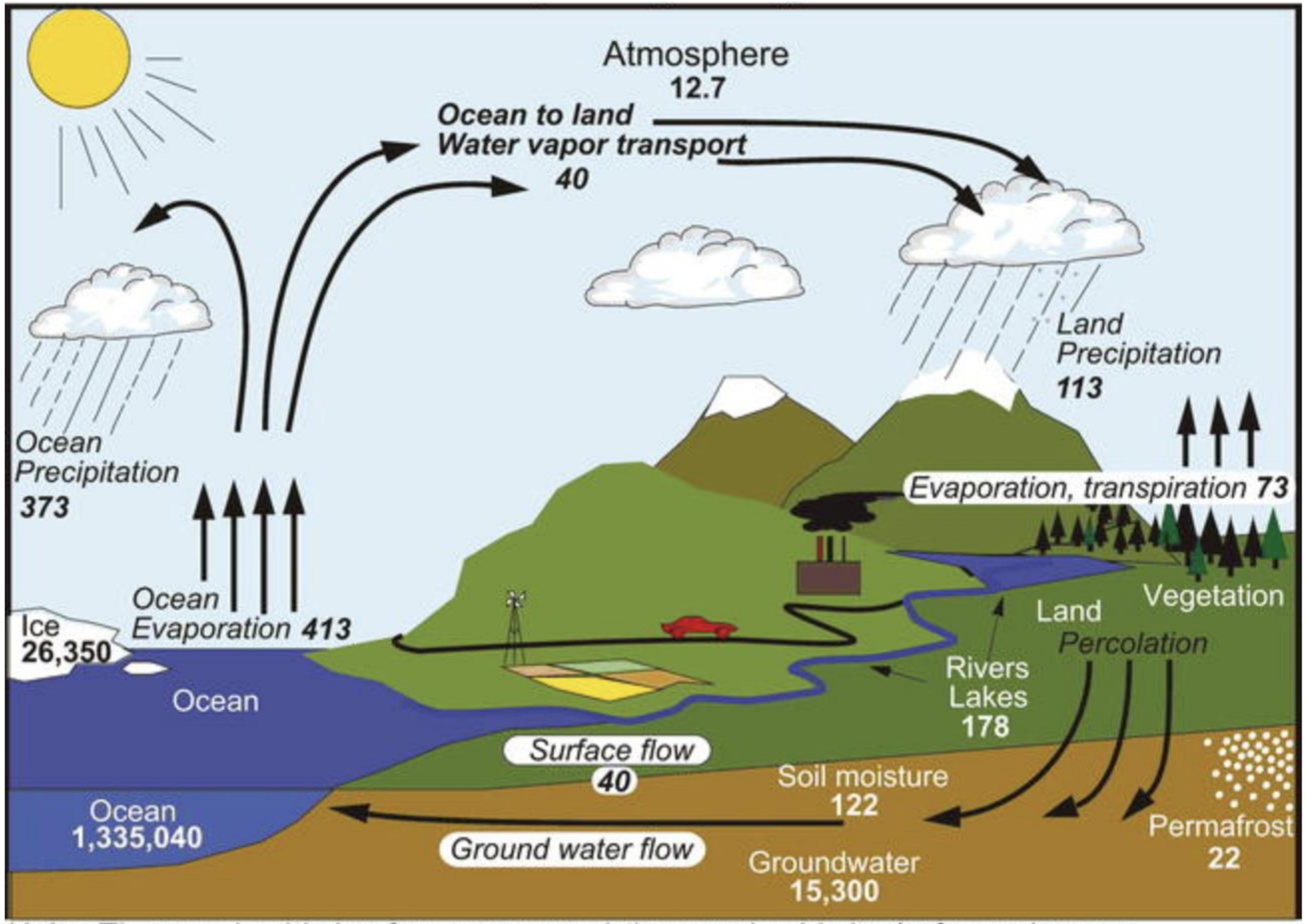
ML4LM webinar, 17 June 2026

Patricia.Rosnay@ecmwf.int

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

Earth system model

Image: Trenberth et al
J. Hydrometeorol. 2007



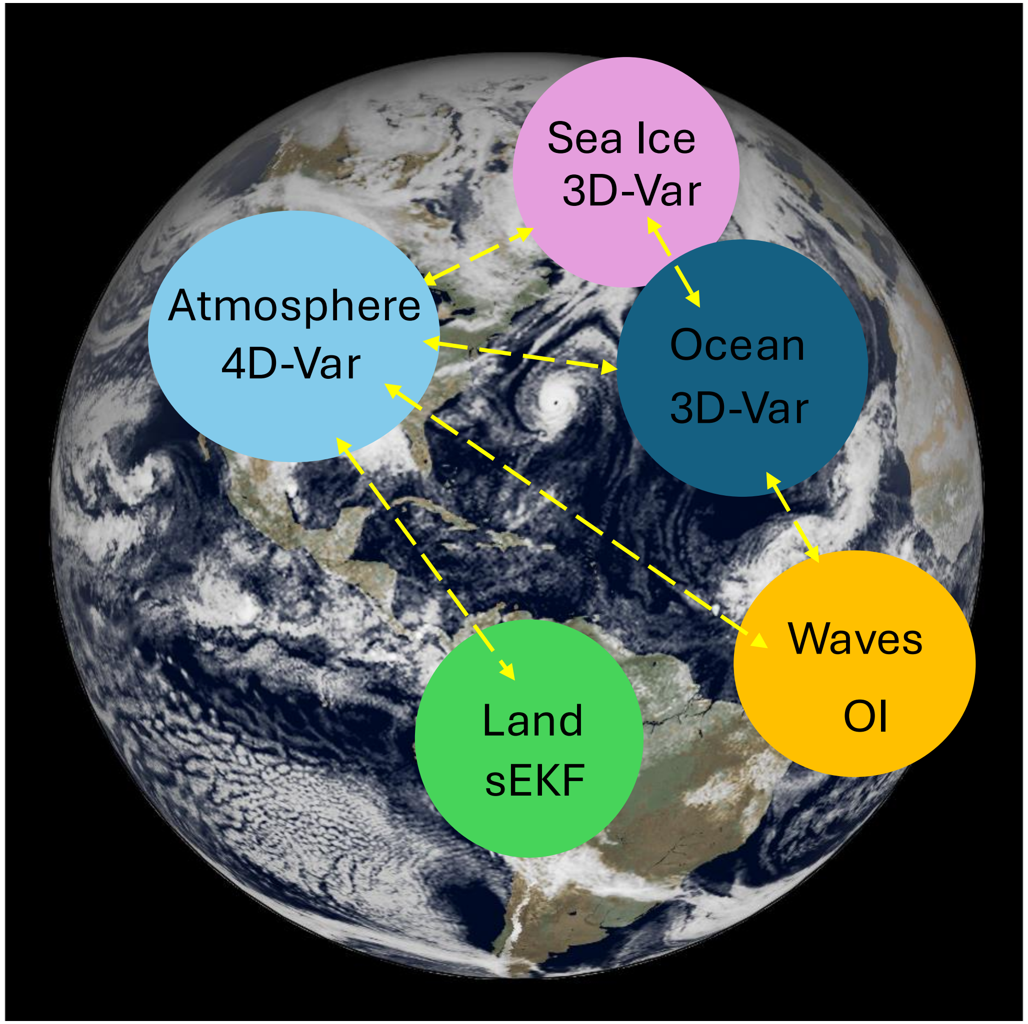
Units: Thousand cubic km for storage, and *thousand cubic km/yr* for exchanges

Earth system data assimilation

Integrated Forecasting system (IFS)

Over land: sEKF (simplified Extended Kalman Filter)

- de Rosnay P. et al QJRMS 2022 - coupled data assimilation (DA) strategy
- Browne et al. ECMWF Spring Newsletter 2026 - ocean-atmosphere outer loop coupling
- Herbert et al. ECMWF Spring Newsletter 2026 - land-atmosphere outer loop coupling



Coupled data assimilation

- Balanced initial conditions across the coupled forecast model components
- Opportunities to enhance the exploitation of surface sensitive satellite data
- Challenges related to complexity of land surfaces and related radiative transfer modelling

→ increasingly rely on ML approaches to link Level-1 satellite data to surface variables

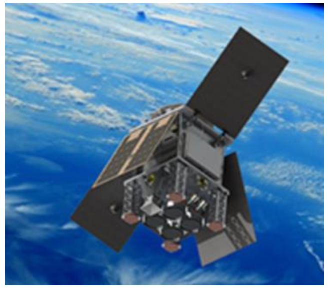
Existing missions | Future missions

ASCAT

SMOS

Sentinel-5p

HydroGNSS



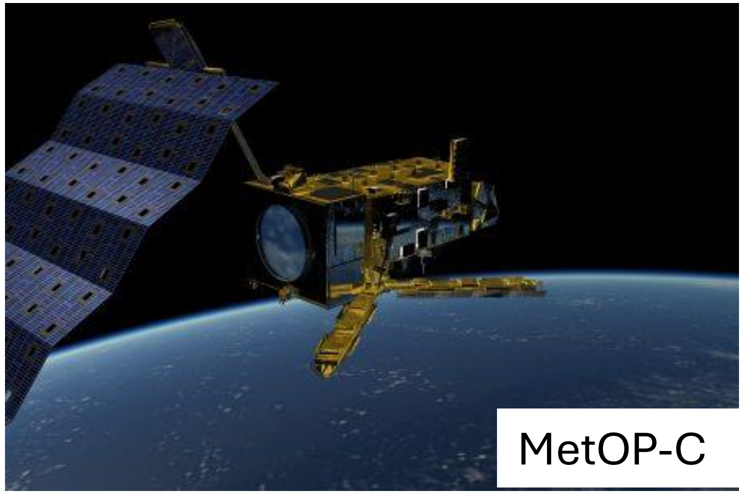
CRISTAL

LSTM

CIMR

Soil moisture satellite observations used for NWP (along with in situ T2m, RH2m screen level observations)

Active microwave data:
ASCAT: Advanced Scatterometer
MetOP-B (2012-), MetOP-C (2018-)
C-band (5.6GHz) **backscattering coefficient**
EUMETSAT Operational mission

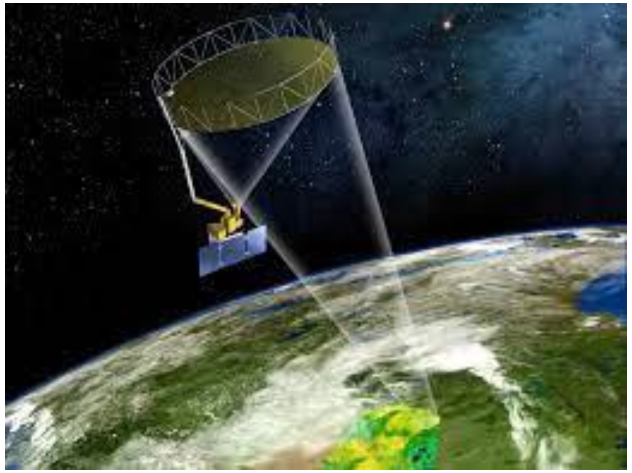


Scatterometer soil moisture → NWP, ERA5 & ERA6
(ERS-SCAT, Metop/ASCAT)

Passive microwave data:
SMOS: Soil Moisture & Ocean Salinity (2009-)
L-band (1.4 GHz) **Brightness Temperature (TB)**
ESA Earth Explorer, dedicated soil moisture mission
(Kerr et al., 2016)

SMOS → NWP & ERA6

SMAP (monitoring)
L-band TB 2015-
NASA Dedicated
soil moisture mission

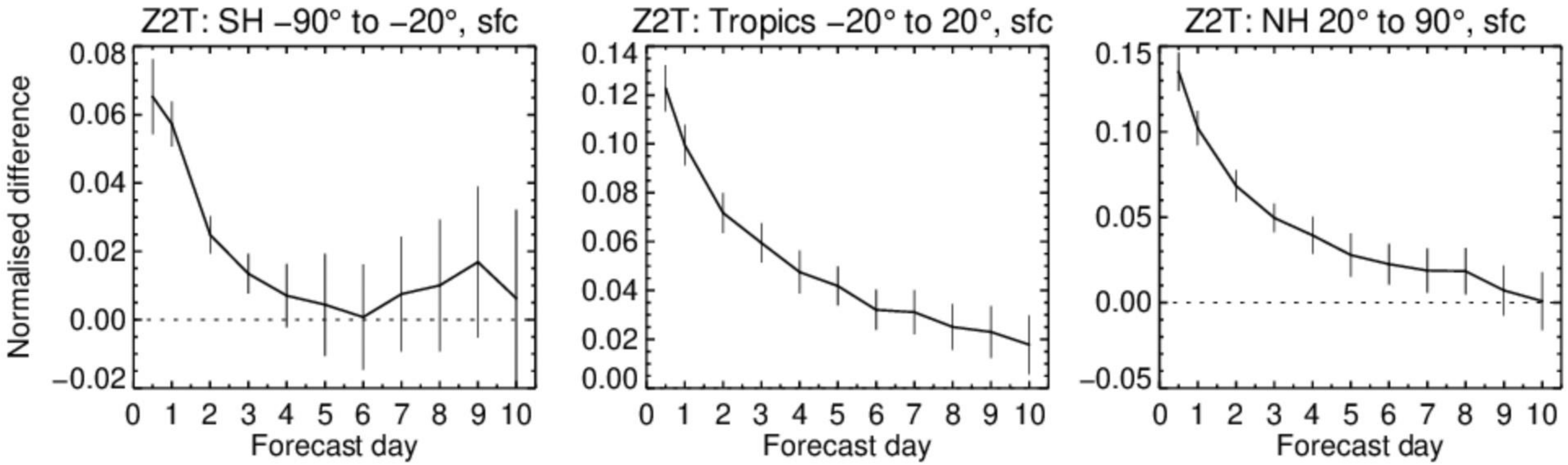


SMAP → monitoring

Unified soil analysis impact on the atmospheric forecast

June-August 2024
IFS cycle 50r1

Impact of soil analysis on T2m forecasts
T2m normalized RMSE difference without minus with soil analysis

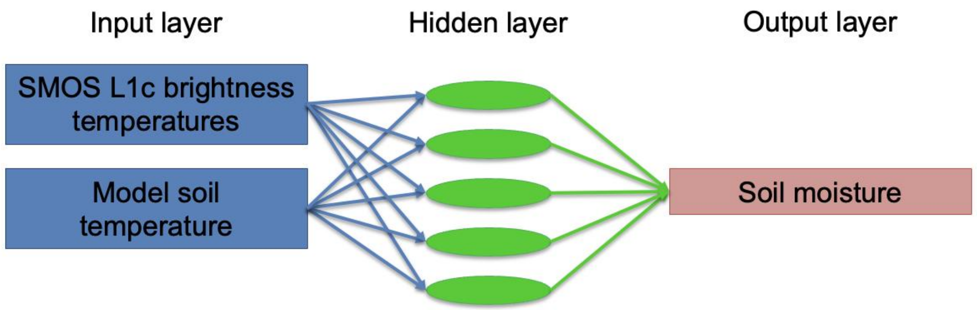


Degradation when soil analysis is off

→ **Significant positive impact of the soil analysis on low level atmospheric temperature forecasts**

The CERISE project (grant agreement No101082139) 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.

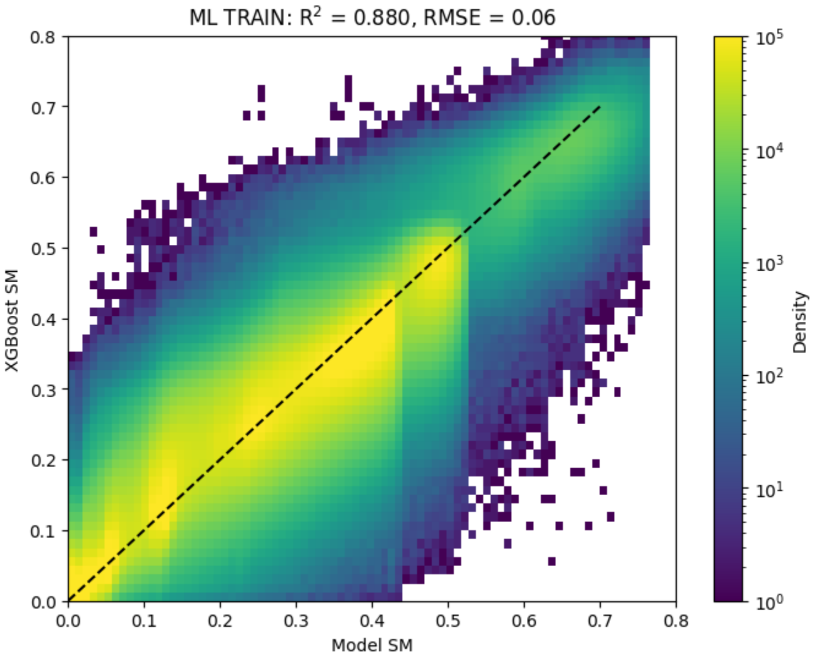
SMOS soil moisture retrieval for data assimilation

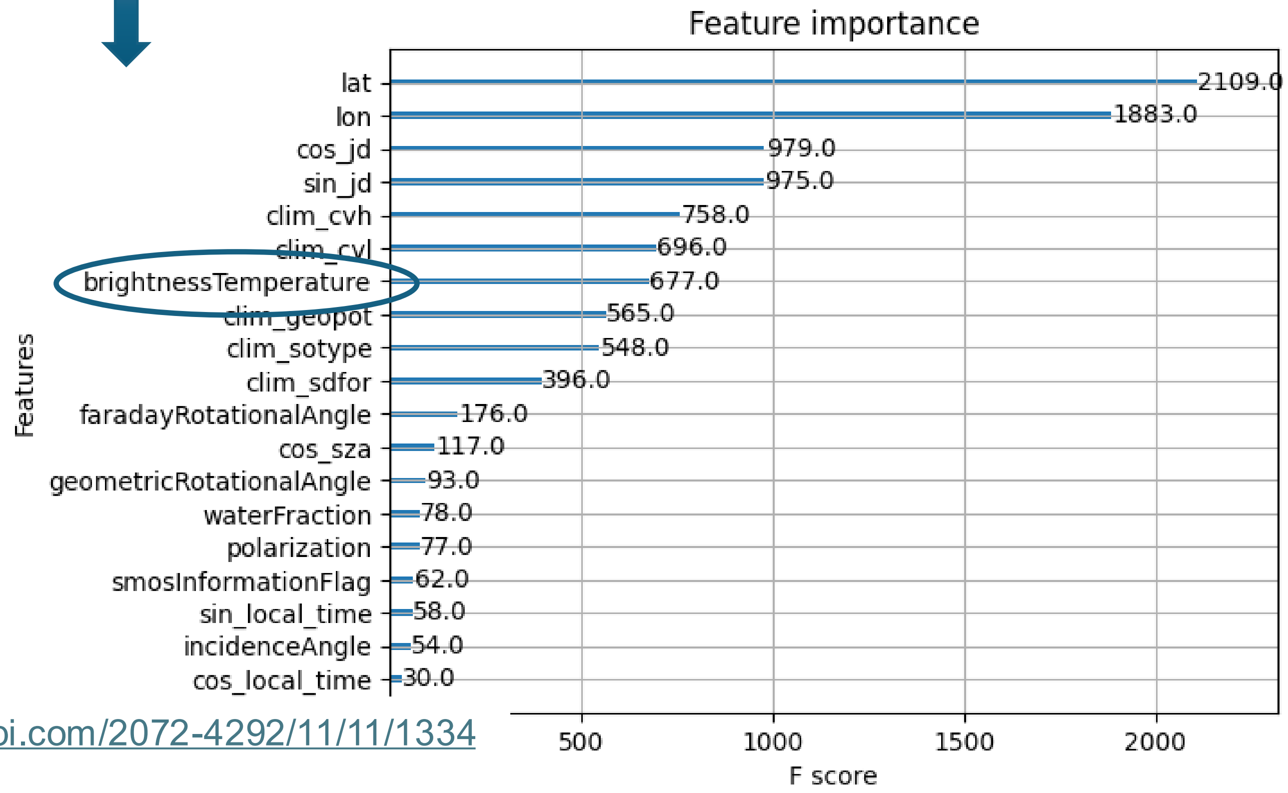


Used for operational NWP since 2019

using a Neural Network (2019-2025)

And XGBoost (since Feb 2025)





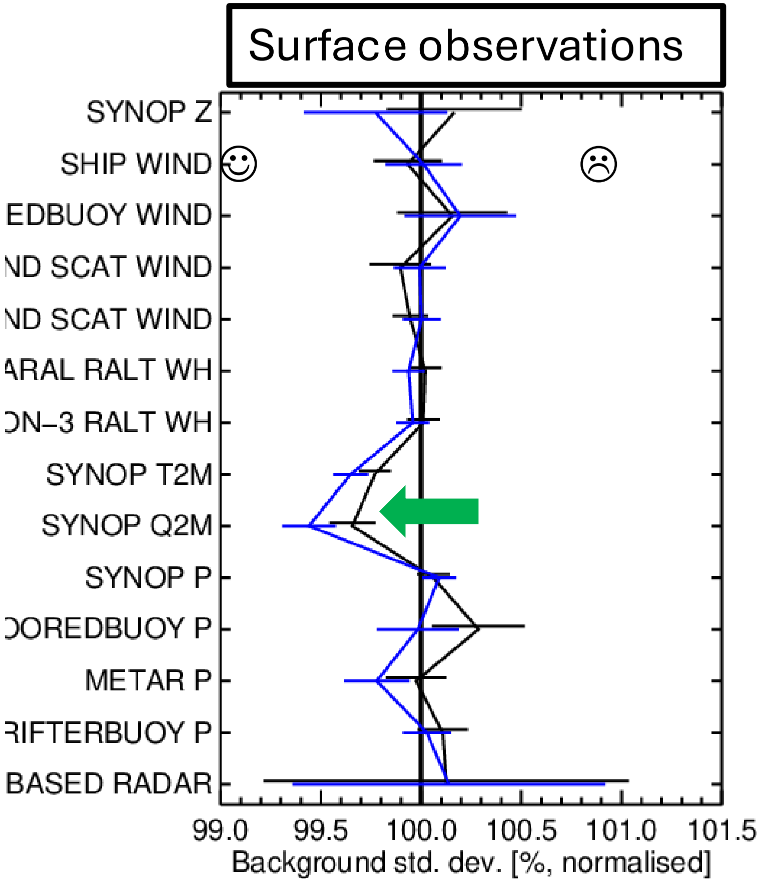
Rodriguez-Fernandez et al, Remote sensing, 2019 https://www.mdpi.com/2072-4292/11/11/1334

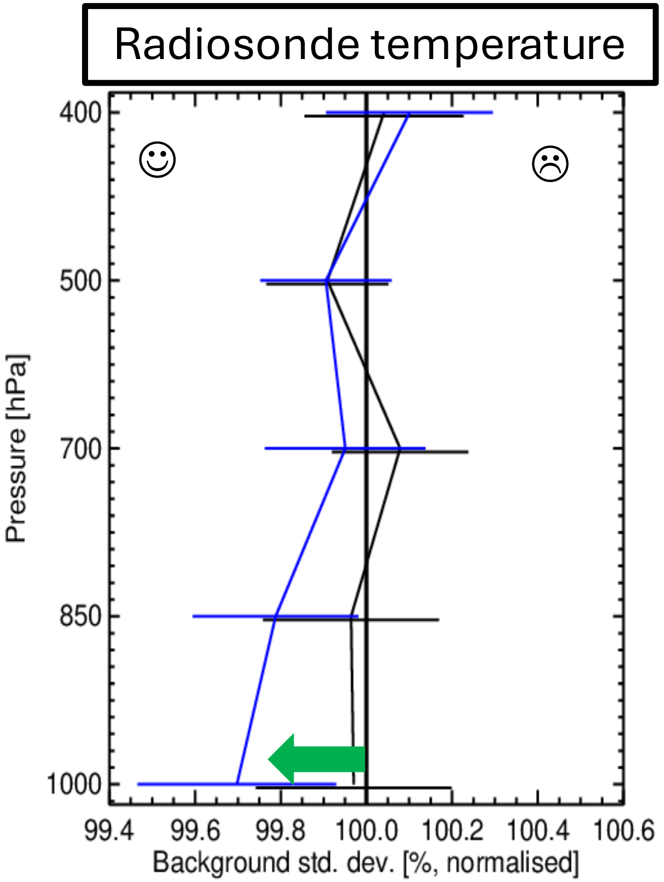
Salonen et al., in prep 2026

SMOS data assimilation impact

- Comparing tree-based method (XGBoost) and Neural Network (Rodriguez-Fernandez, RS 2019) SMOS DA
- **SMOS soil moisture DA has positive near surface impact when compared to no SMOS DA (100% line).**
- The positive impact of SMOS DA is stronger with **XGBoost** than **NN** SMOS soil moisture in IFS cycle 49r1.

SMOS XGBoost DA
SMOS NN DA



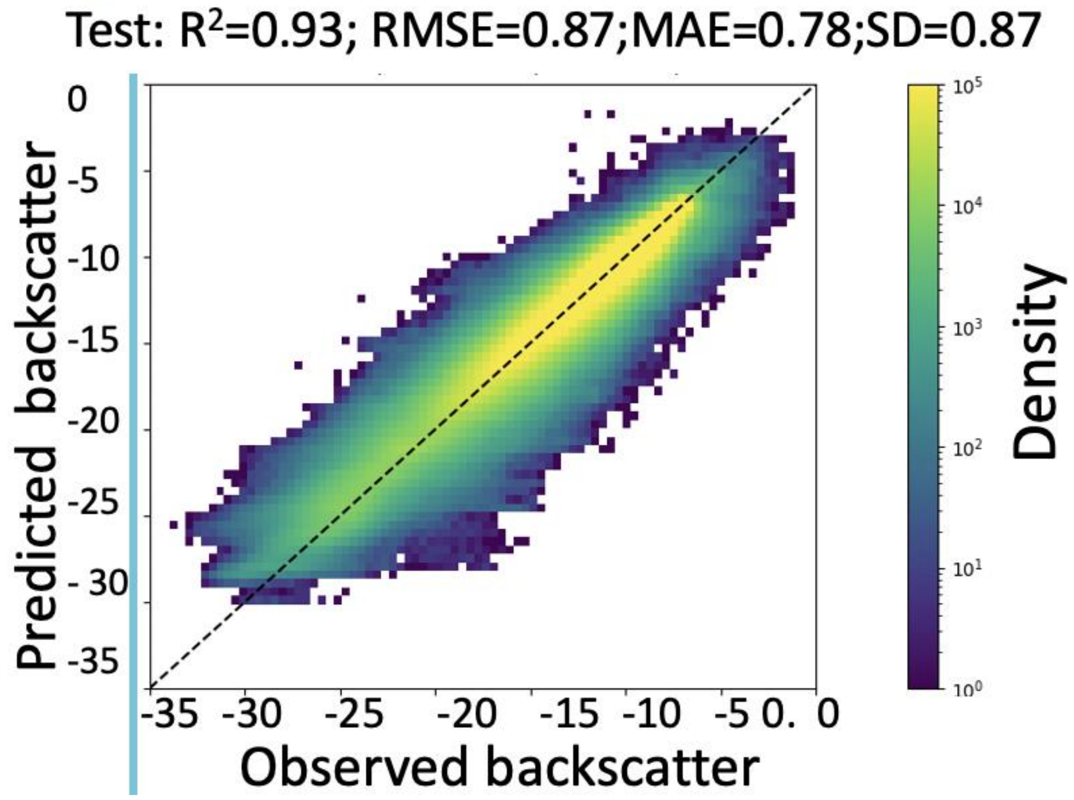


Kirsti Salonen

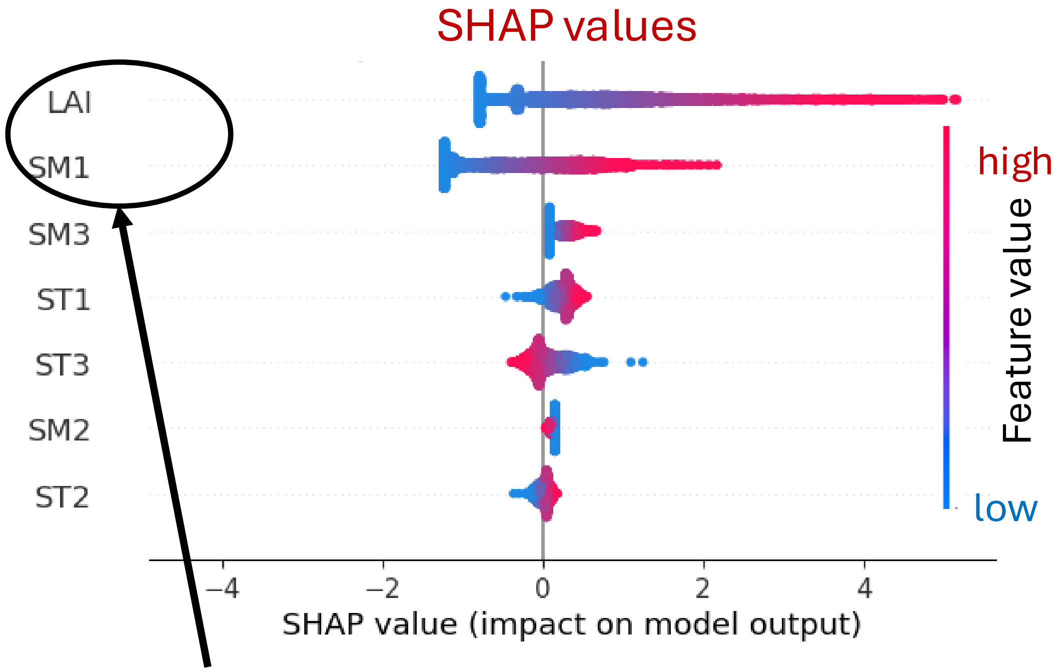
Microwave observation operators

Enhance the exploitation of satellite observations in coupled land-atmosphere assimilation to constrain vegetation water and carbon cycle variables.

→ Machine Learning (ML)-based observation operators. Example for ASCAT:



Features selection



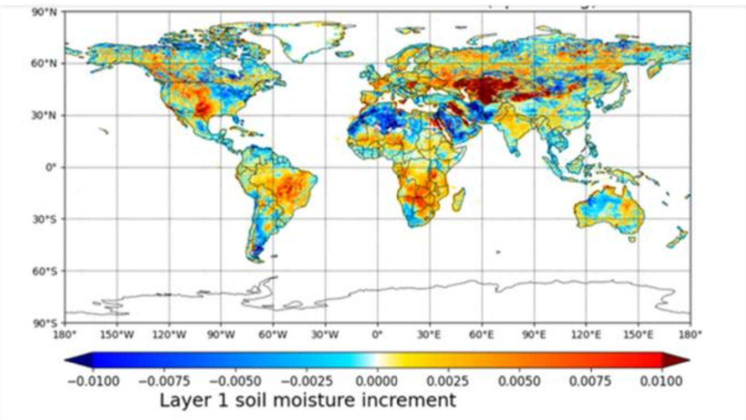
Vegetation (LAI) and surface soil moisture (SM1) are the most influential variables

Sébastien Garrigues

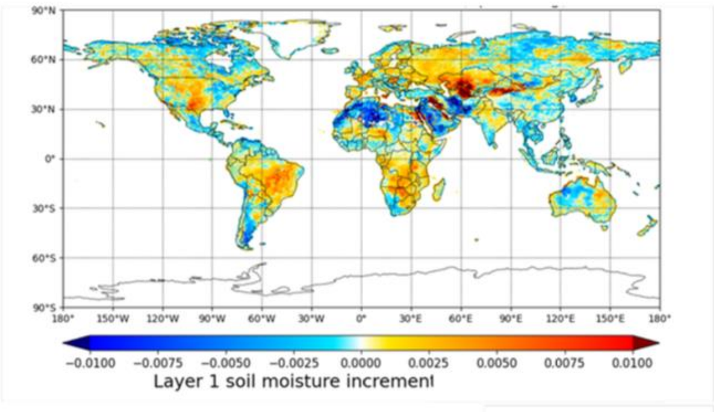
ASCAT backscatter 40 degrees DA in the IFS

Surface

Experiment with assimilation of ASCAT backscatter



Experiment with assimilation of ASCAT soil moisture

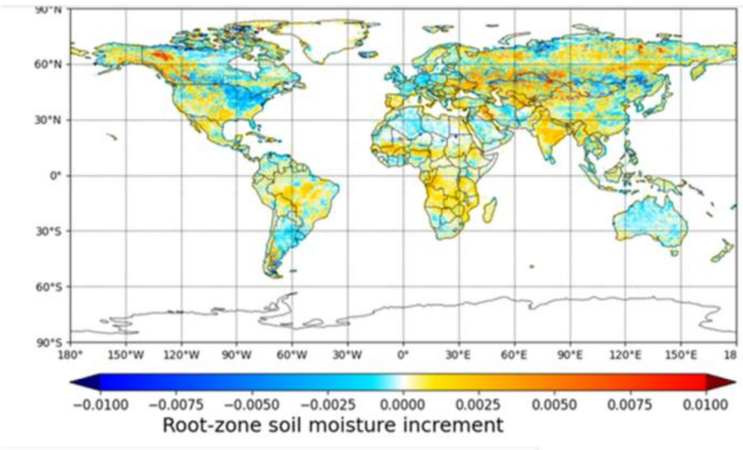


Impact of assimilating ASCAT backscatter 40 degrees instead of Level-2 soil moisture on soil moisture increments ($m^3.m^{-3}$)

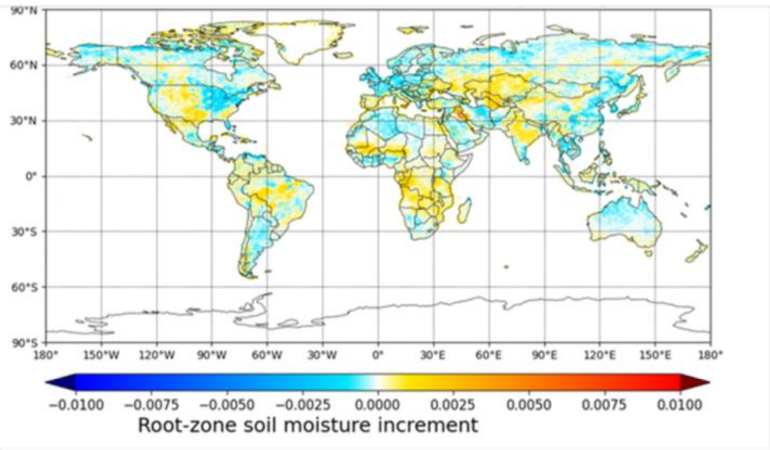
Tested period Summer 2022

Root zone

Experiment with assimilation of ASCAT backscatter

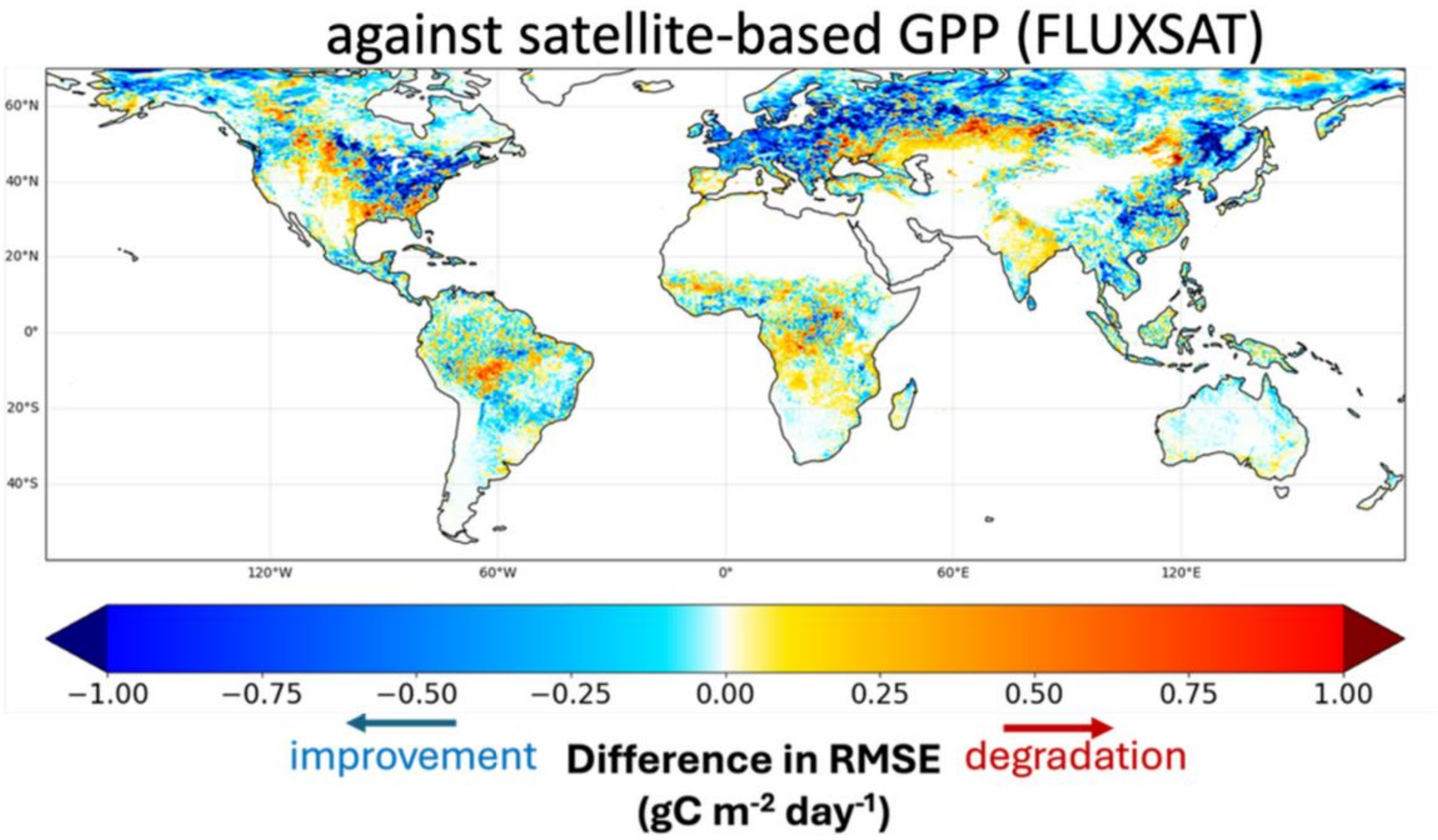


Experiment with assimilation of ASCAT soil moisture



Larger increments in deeper layers reflecting sensitivity of ASCAT backscatter to deep soil moisture in the the observation operator

ASCAT backscatter 40 degrees DA in the IFS



(diff between ASCAT_40 and ASCAT_SM data assimilation)

- Feedforward neural network observation operator provides accurate prediction of ASCAT backscatter from the IFS fields.
- The assimilation of ASCAT backscatter at 40° improves the simulation of the root-zone soil moisture, leading to improved GPP over Northern latitudes.

LAI analysis using Solar Induced Fluorescence (SIF)

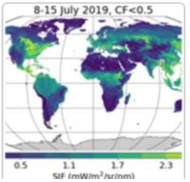
https://doi.org/10.5194/essd-13-5423-2021

Article | Assets | Peer review | Metrics | Related articles

Data description paper | 19 Nov 2021

The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission

Luis Guanter, Cédric Bacour, Andreas Schneider, Ilse Aben, Tim A. van Kempen, Fabienne Maignan, Christian Retscher, Philipp Köhler, Christian Frankenberg, Joanna Joiner, and Yongguang Zhang



SIF: electromagnetic signal emitted by the chlorophyll of assimilating plants

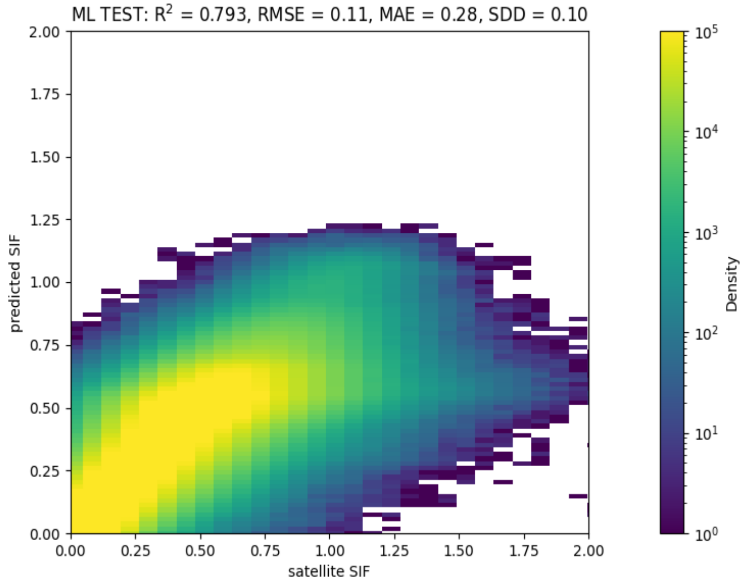
→ part of the energy absorbed by chlorophyll is not used for photosynthesis but emitted at longer wavelengths in the near infrared

→ Relevant to analyse vegetation LAI and Gross Primary Production (GPP)

TROPOMI SIF 743-758 nm

Exploratory work to use SIF at ECMWF.
Observation operator development

Predicted SIF



Satellite SIF

Sébastien Garrigues et al., QJRMS 2026
http://dx.doi.org/10.1002/qj.70111

Funded by the European Union

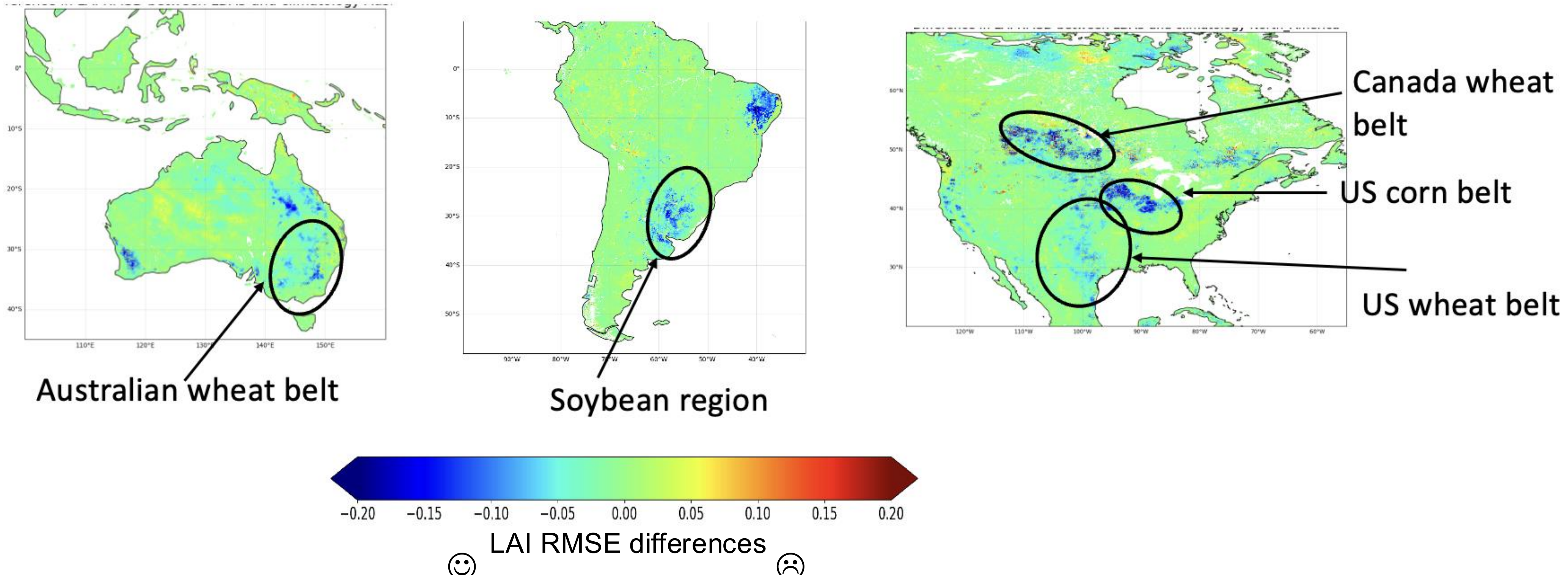
LAI analysis using Solar Induced Fluorescence (SIF)

CORSO
CO2MVS Research on Supplementary Observations

Funded by the European Union

Impact of SIF DA shown as LAI RMSE differences with vs without SIF data assimilation against satellite (Copernicus Land) LAI for 2022

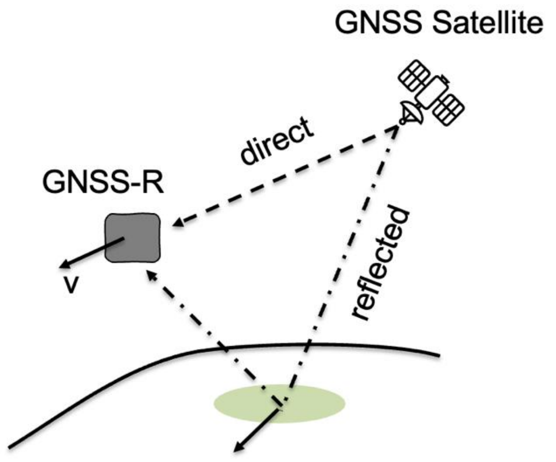
Improvement for cropland



The CORSO project (grant agreement No101082194) is funded by the European Union. Views and opinions expressed are however thoseof 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.

Global Navigation Satellite System - Reflectometry (GNSS-R)

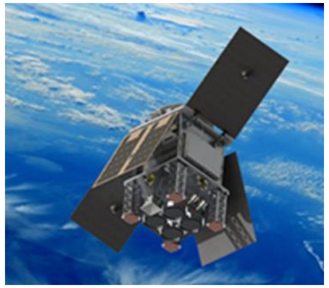
Concept:
GNSS-R gets information from the signal reflected at the surface.
Wet soils reflect more strongly than dry soils.



Strengths:

- Uses existing GNSS (e.g. GPS) signals → opportunity information
- Frequent observations (Daily and Sub-daily)
- Potentially high spatial resolution
- Low cost compared to other dedicated SM missions

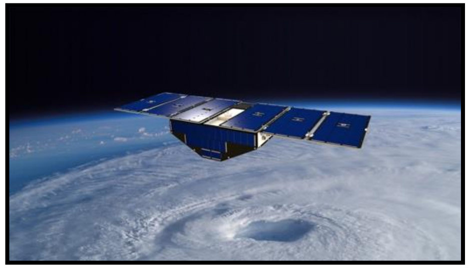
HydroGNSS
Launched Nov 2025



Using CYGNSS (Cyclone GNSS) in preparation of hydroGNSS

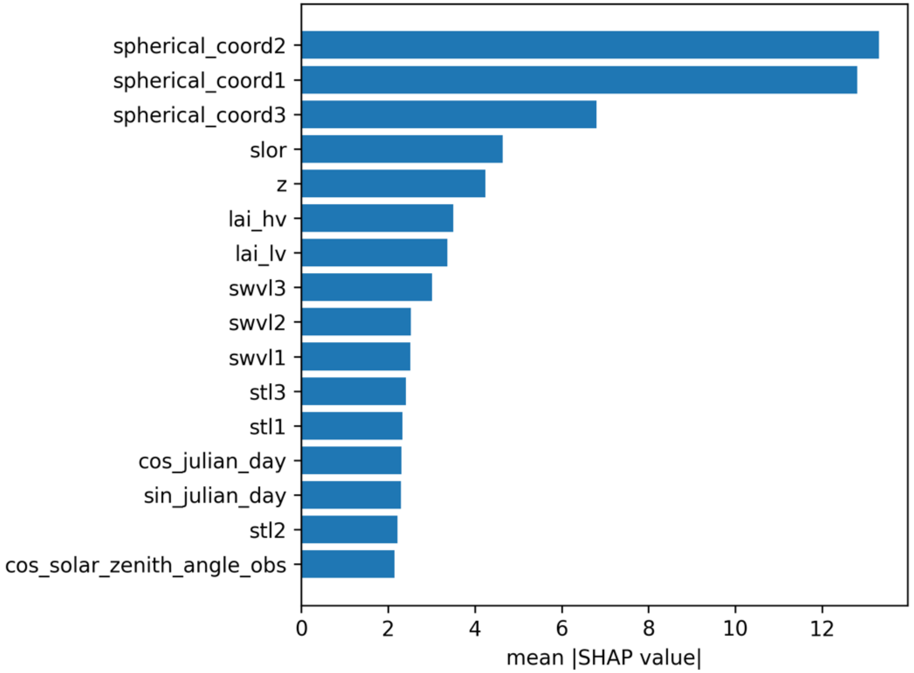
Eleni Kalogeraki
Jana Kolassa

GNSS-R ML-based observation operator

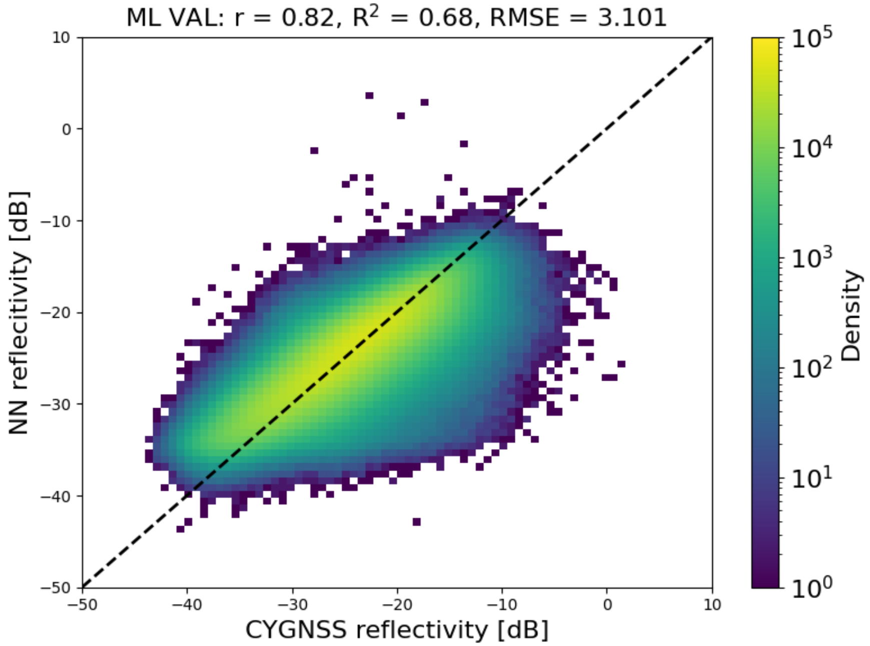


CYGNSS
Constellation of 8 satellites
Orbit between ±38° latitude

Training data ~36 million samples (across 2022/2023)
Validation data ~9 million samples
(20% retention, 80% rejection)

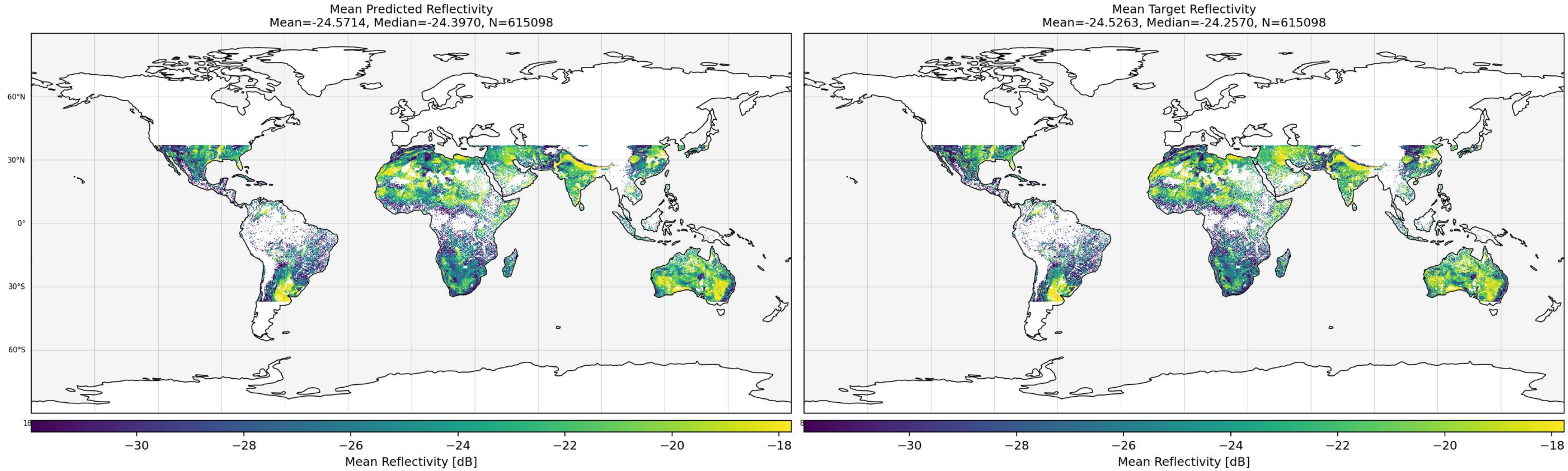


Importance of orography, geolocation, vegetation, soil moisture and temperature



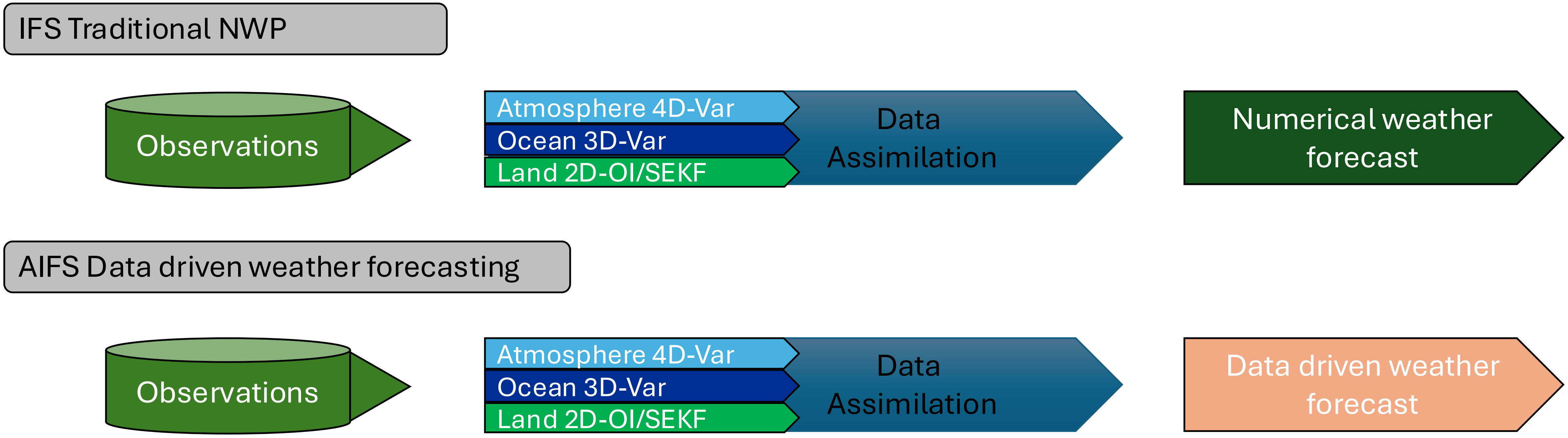
Jana Kolassa

GNSS-R ML-based observation operator



Next steps:
Implementation in the ECMWF Land Data Assimilation System (ongoing) and impact study to assess potential of GNSS-R L1 DA on NWP and reanalysis.

ECMWF weather prediction systems



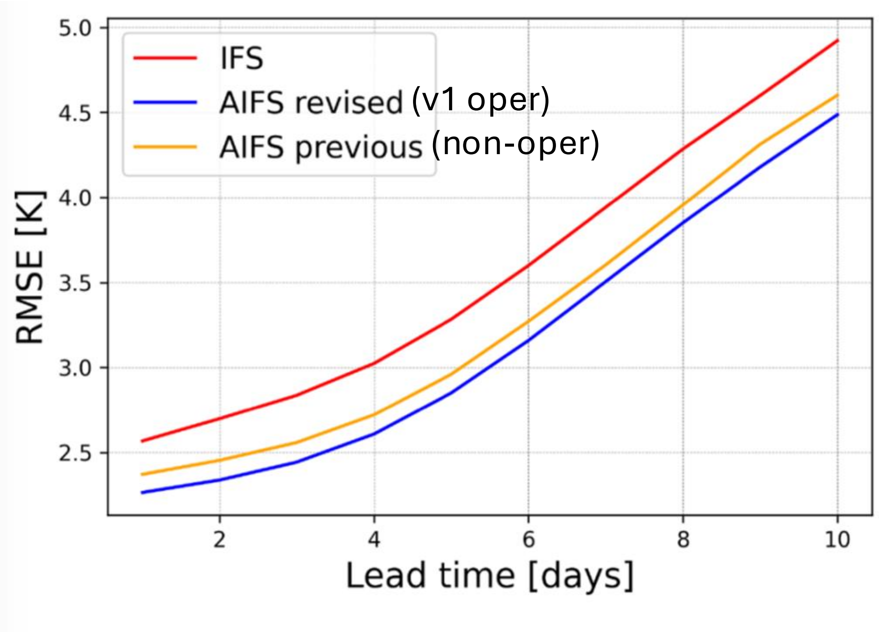
Most data driven forecasting models are trained on reanalysis datasets, specifically the ERA5 reanalysis produced as part of the Copernicus Climate Change Service at ECMWF

Land surfaces in machine learning forecasts models: AIFS

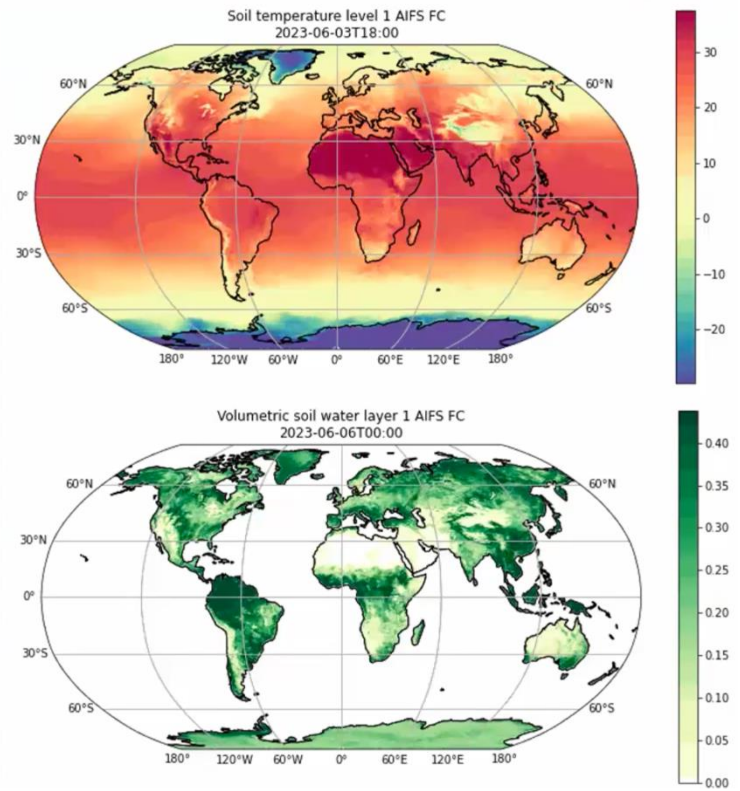
Adding land surface variables:

- v1 from Feb 2025 to May 2026 with soil moisture, temperature, runoff

AIFS v1
Evaluation of T2m



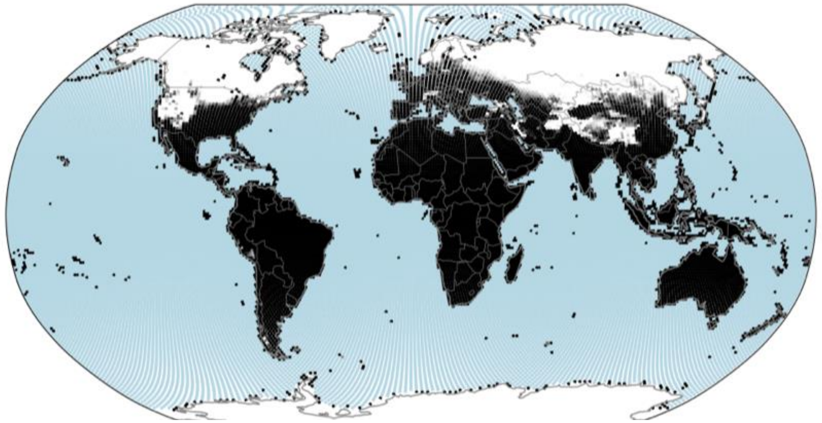
(a) 2-m temperature

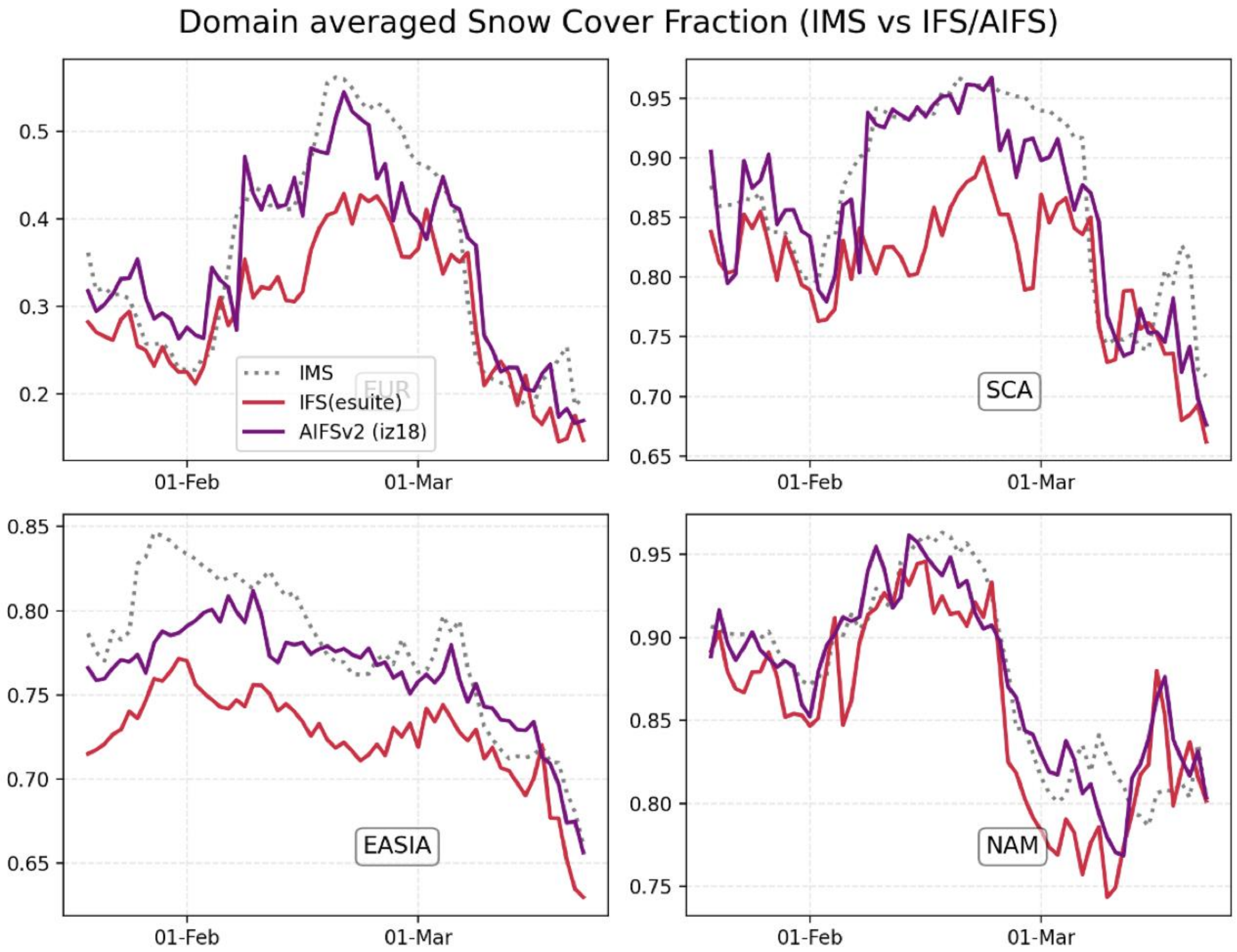


Moldovan et al., GMD 2026 https://doi.org/10.5194/gmd-19-4703-2026

Land surfaces in machine learning forecasts models: AIFS

AIFS v2 with snow cover, oper since May 2026





- By training on ERA5, the AIFS has learnt the snow DA

- Verified against satellite IMS data, snow cover in the AIFS outperforms that of the IFS

Nina Raoult et al. ECMWF NL winter 2026

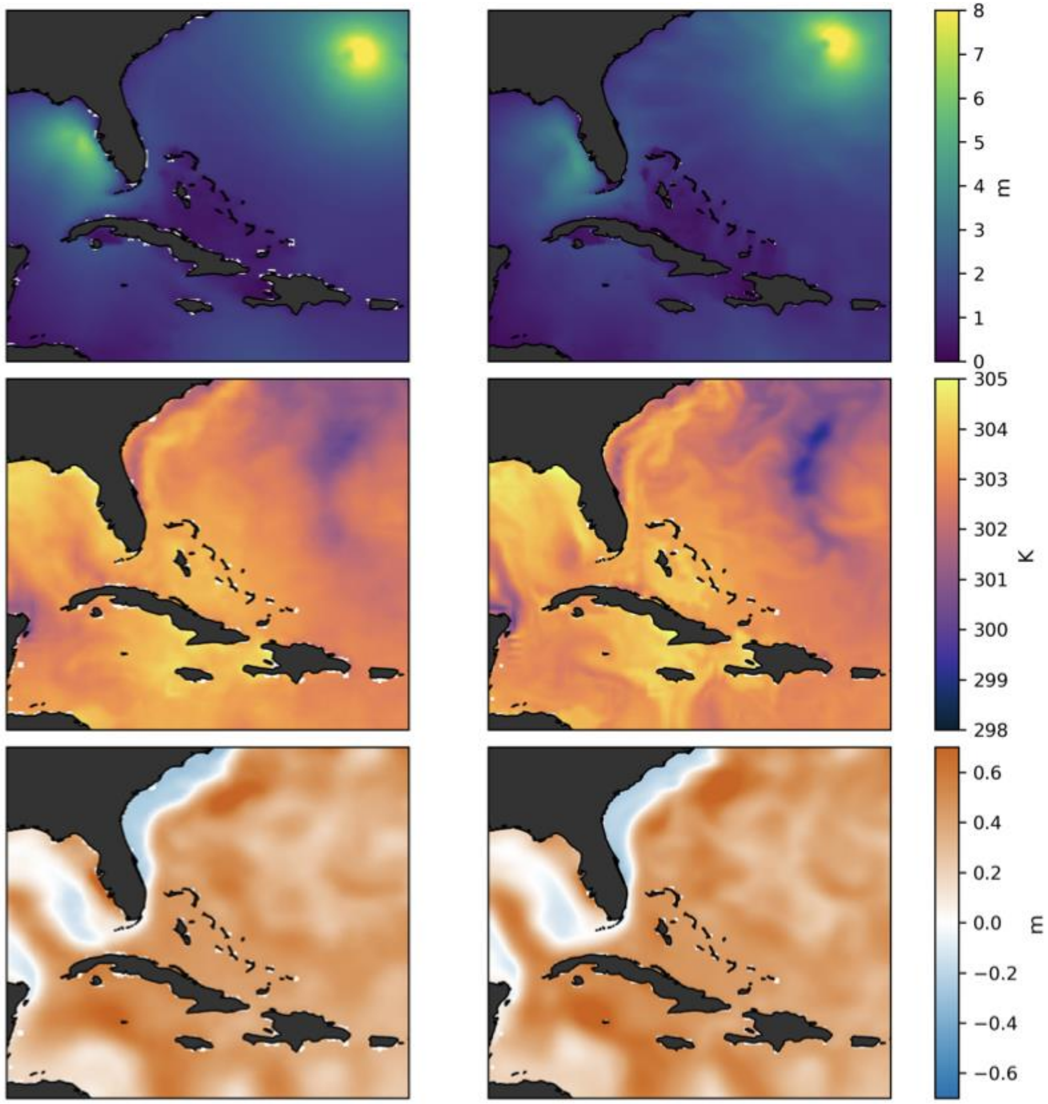
https://www.ecmwf.int/en/newsletter/186/news/integrating-snow-fields-aifs-v2

Ocean surfaces in machine learning forecasts models: AIFS

Significant wave height

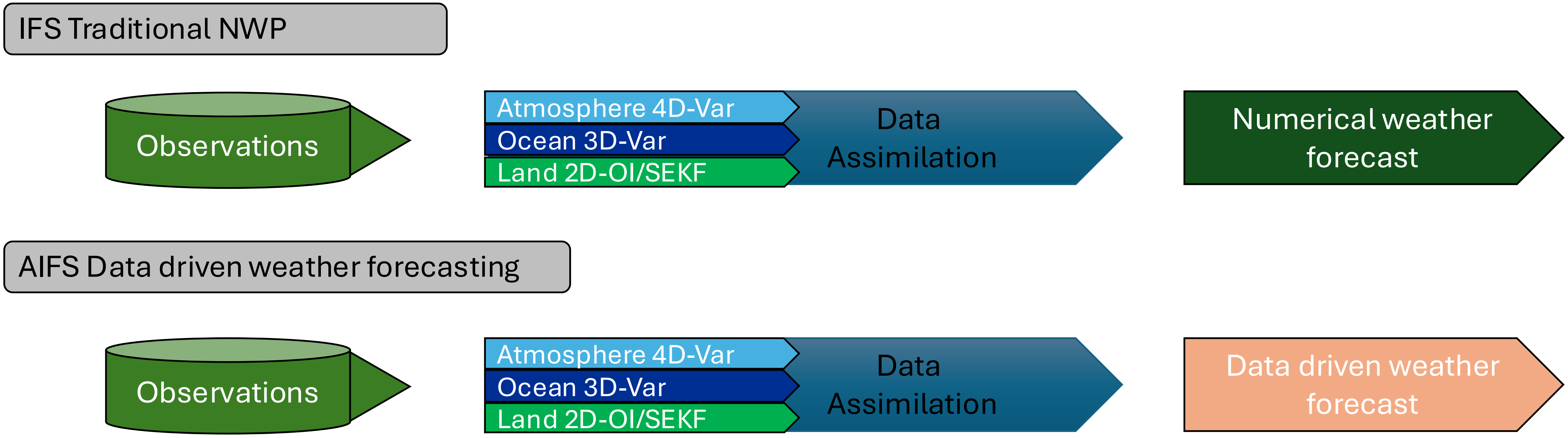
Sea surface temperature

Sea surface height



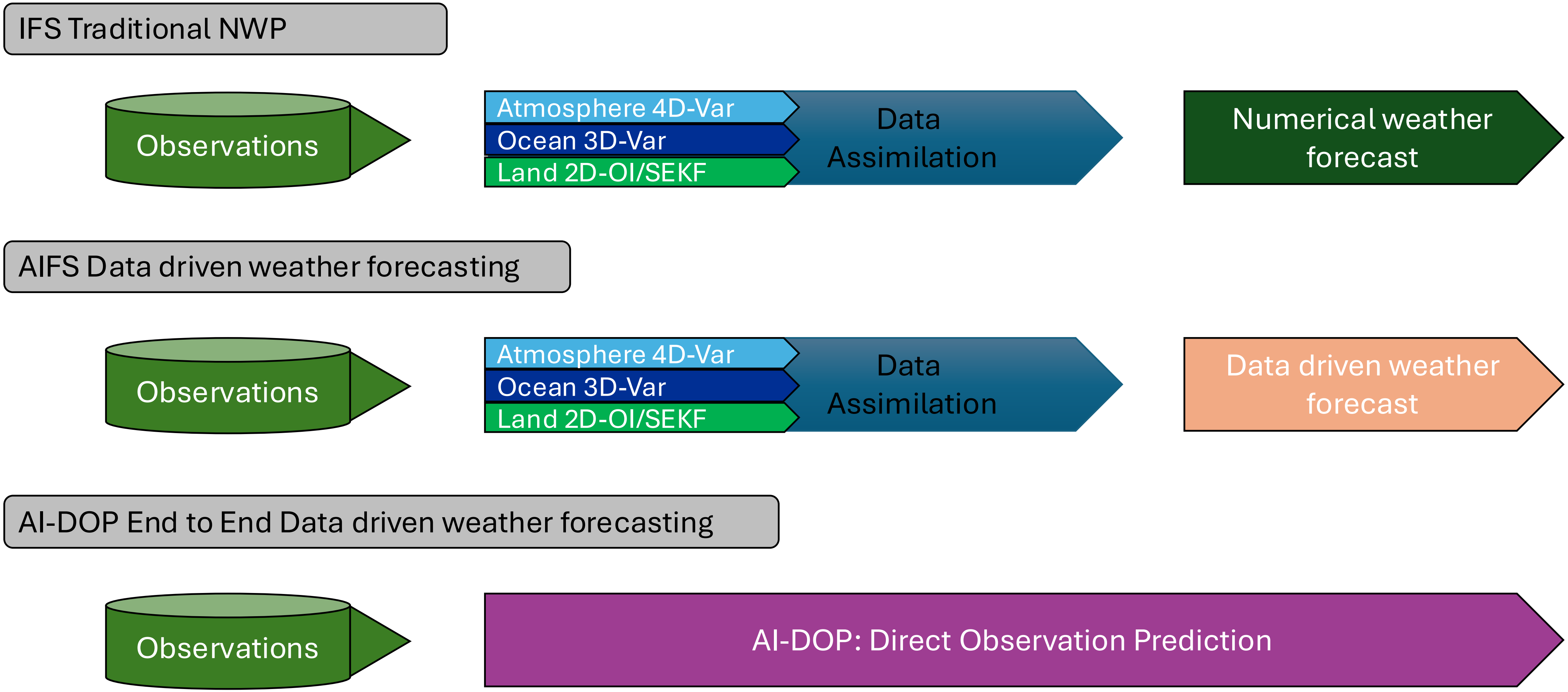
Hahner et al 2026
https://arxiv.org/abs/2604.25559

ECMWF weather prediction systems



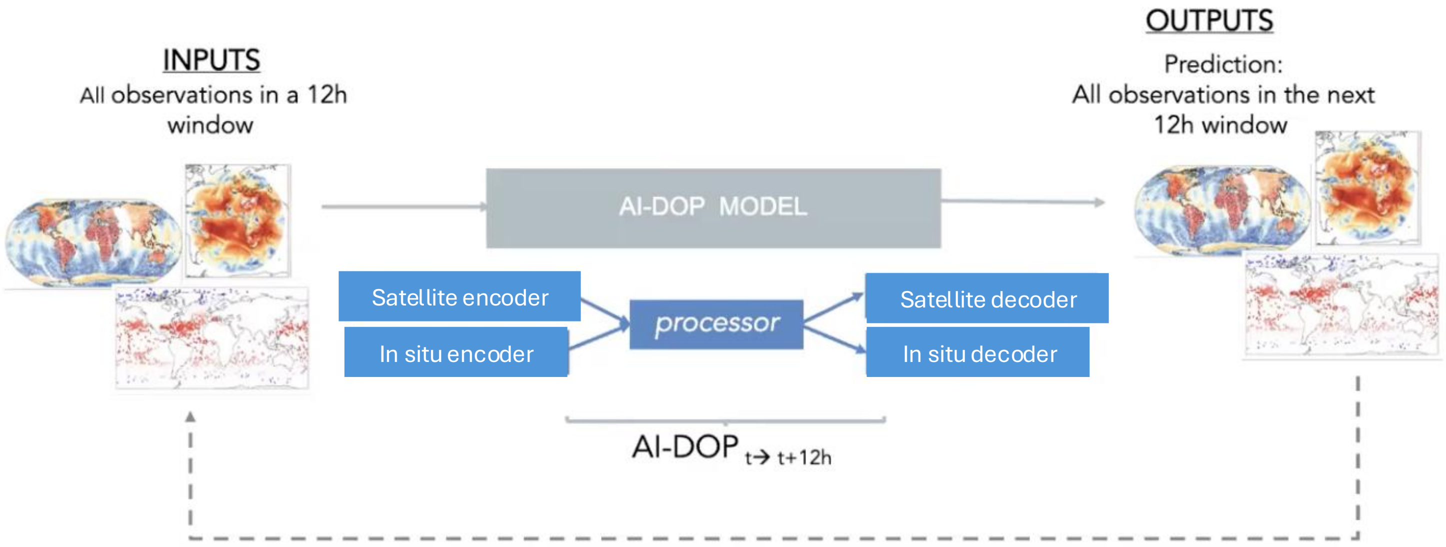
Most data driven forecasting models are trained on reanalysis datasets, specifically the ERA5 reanalysis produced as part of the Copernicus Climate Change Service at ECMWF

ECMWF weather prediction systems



Direct Observation Prediction (AI-DOP)

- Using historical observations to train a Neural Network to forecast future observations
- Can include observations of the full Earth system (atmosphere, ocean, land) simultaneously

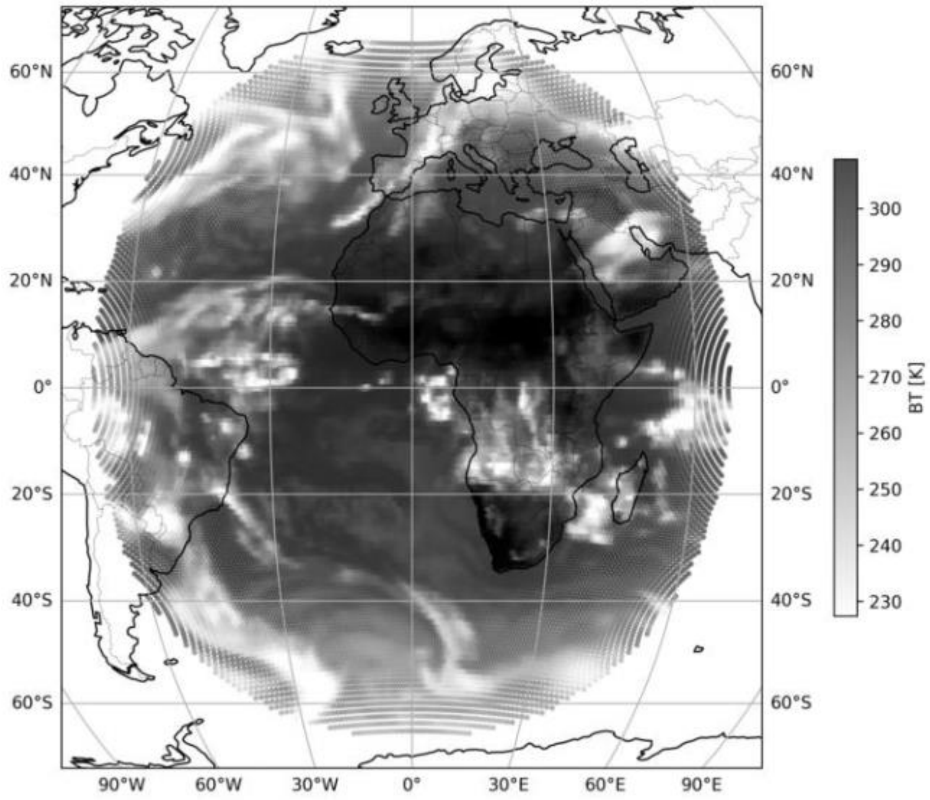


AI-DOP is learning a physical model of the Earth system directly from observations

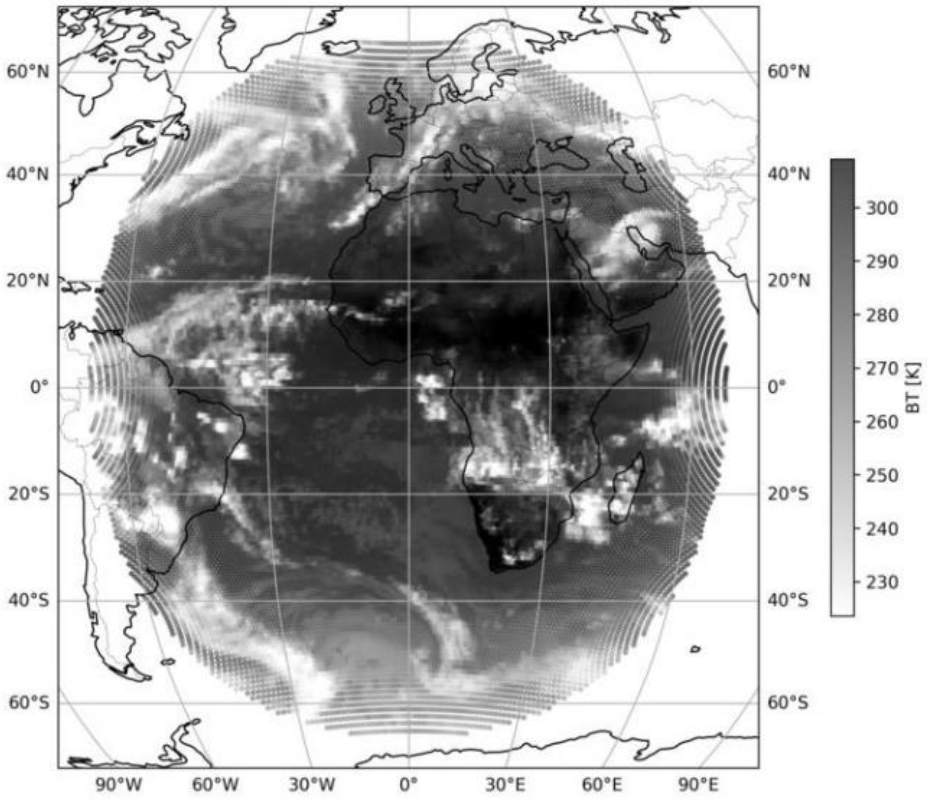
GraphDOP:
Alexe et al., 2024, https://arxiv.org/abs/2412.15687
Boucher et al., 2025 https://arxiv.org/abs/2510.20416

Direct Observation Prediction (AI-DOP)

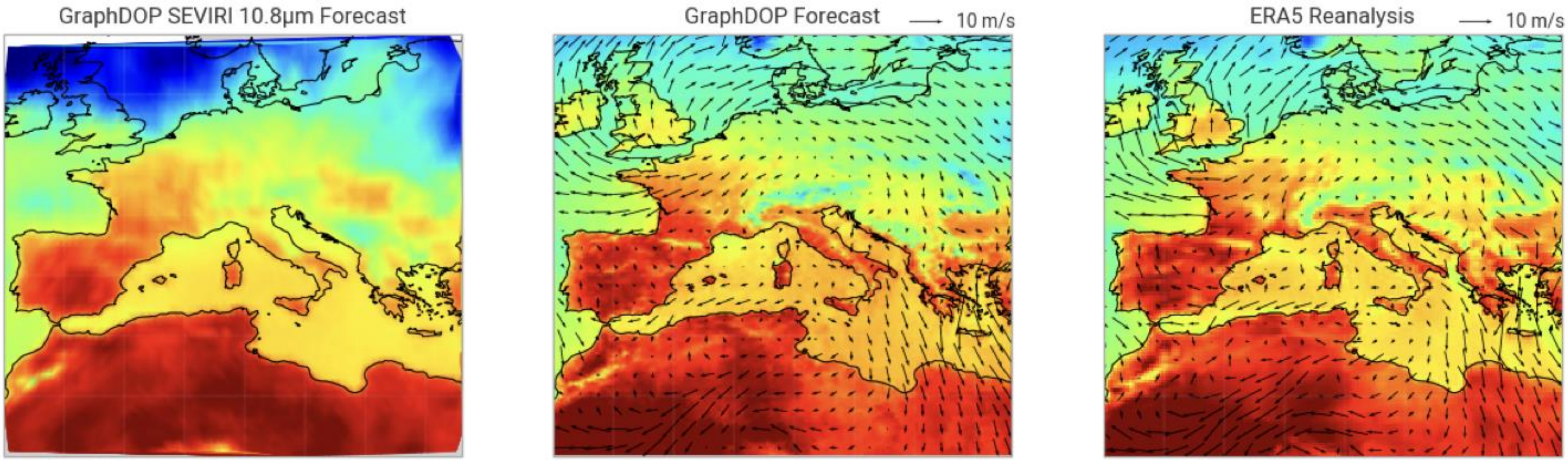
AI-DOP forecast



SEVIRI observation target



Direct Observation Prediction (AI-DOP)

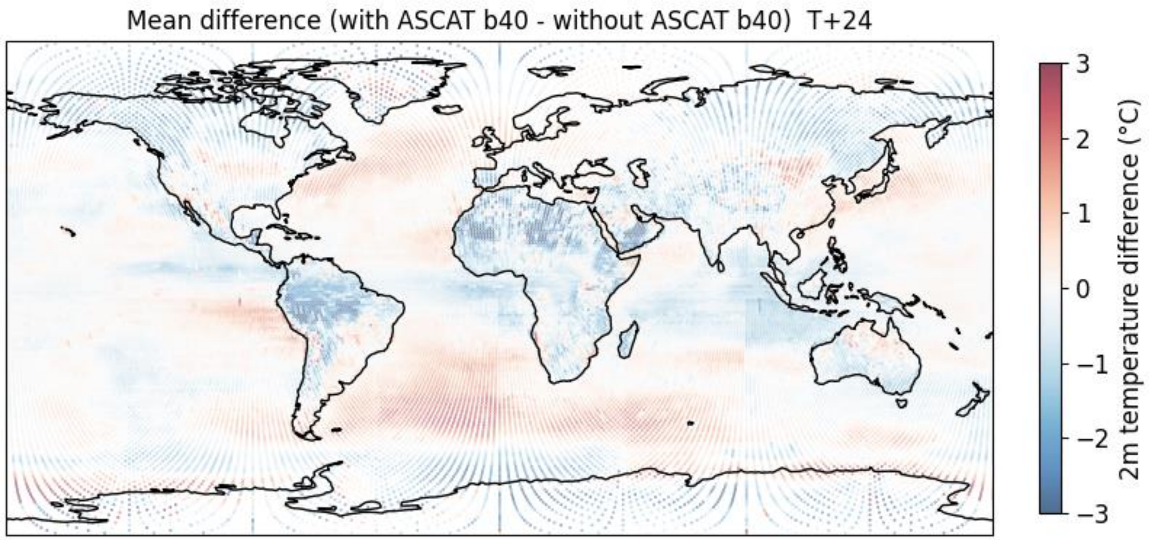


- AI-DOP captures heatwave evolution well over western Europe, but degrades further north
- Maintains physically consistent surface fields

Boucher et al., 2025 https://arxiv.org/abs/2510.20416

Land surfaces in data driven machine learning forecasts models: AI-DOP

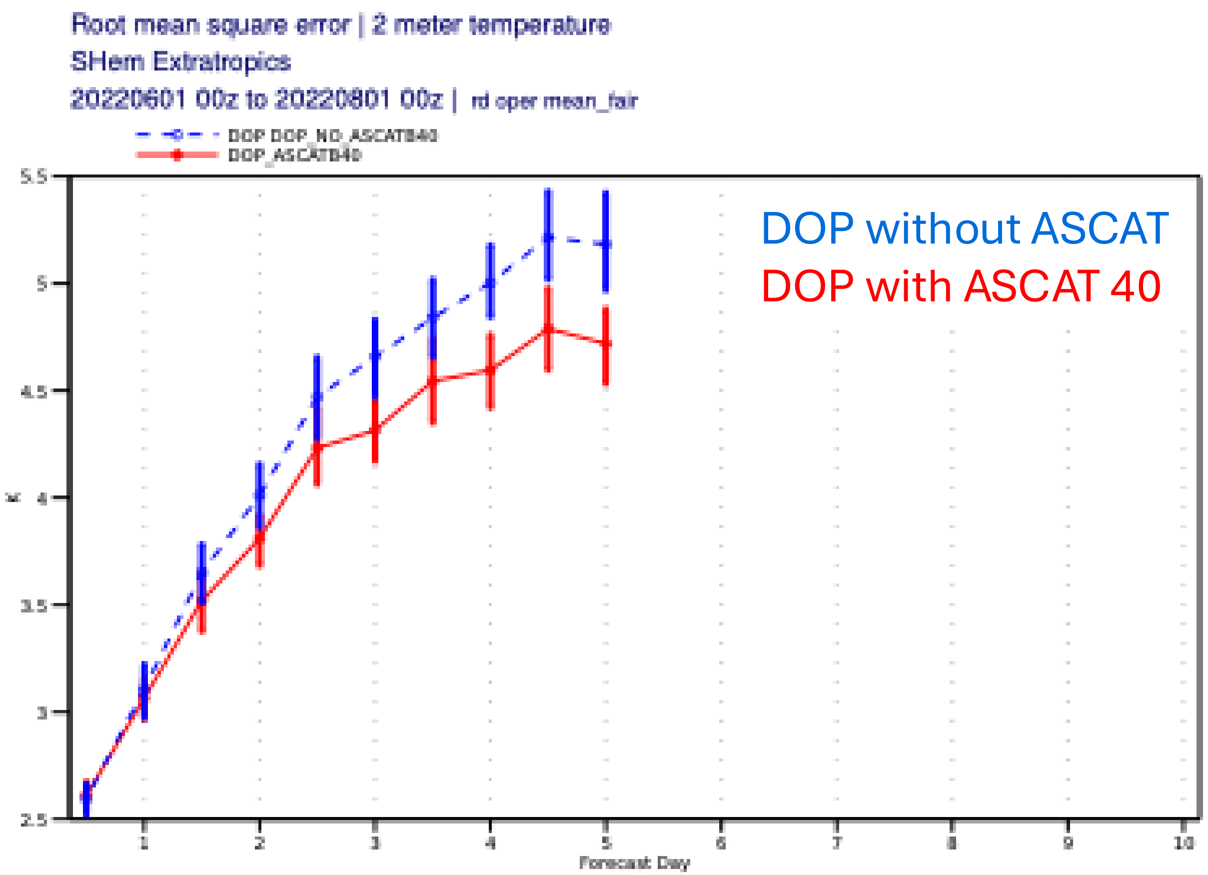
- Using observations to learn a physical model of the Earth system
- Progressive addition of observations sensitive to the surface

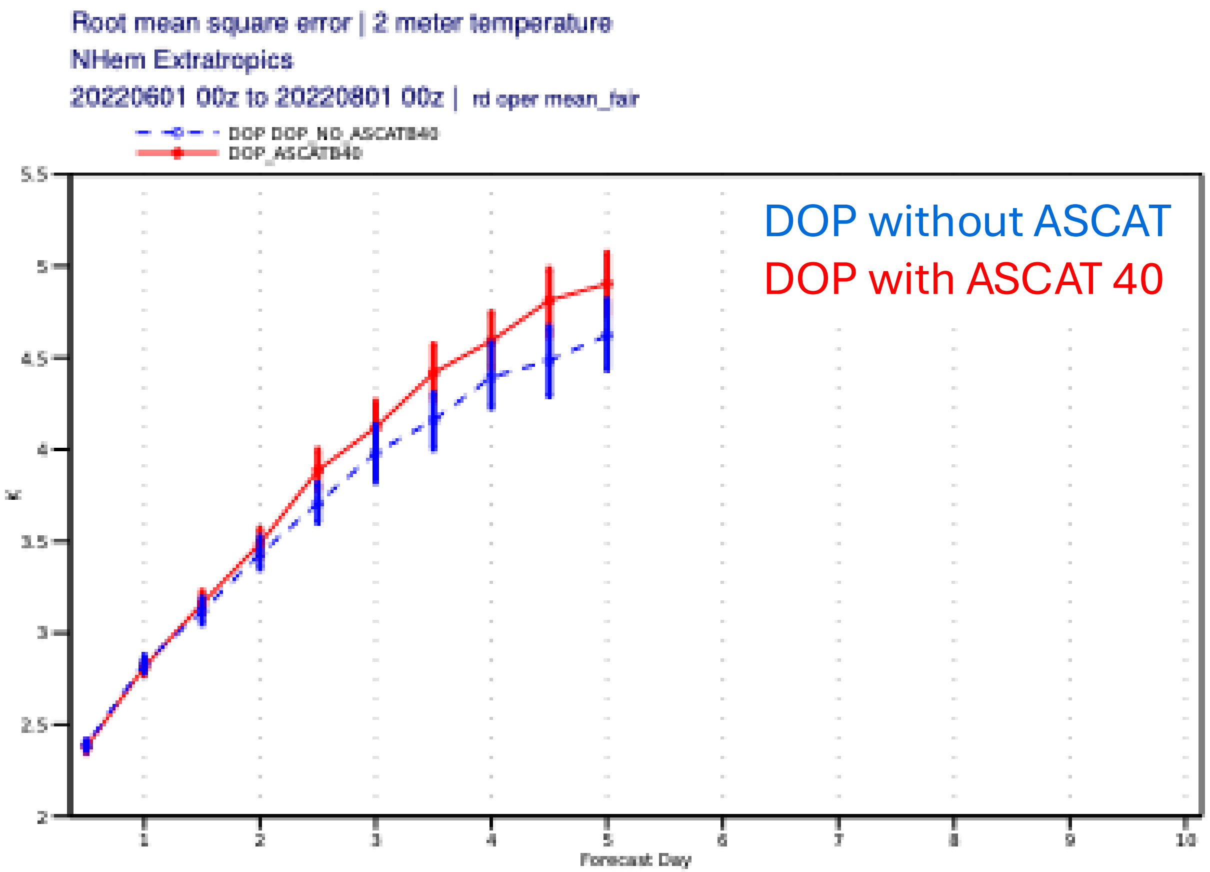


- Predictors: radiosondes, satellites (ATMS,AMSUA, IASI, GPSRO, AMV), surface observations (buoy, ship, automatic weather station.) **+ASCAT backscatter at 40° over land**
- Training : 2012 to 2020; Validation: 2021; Evaluation: 2022

Sébastien Garrigues et al.

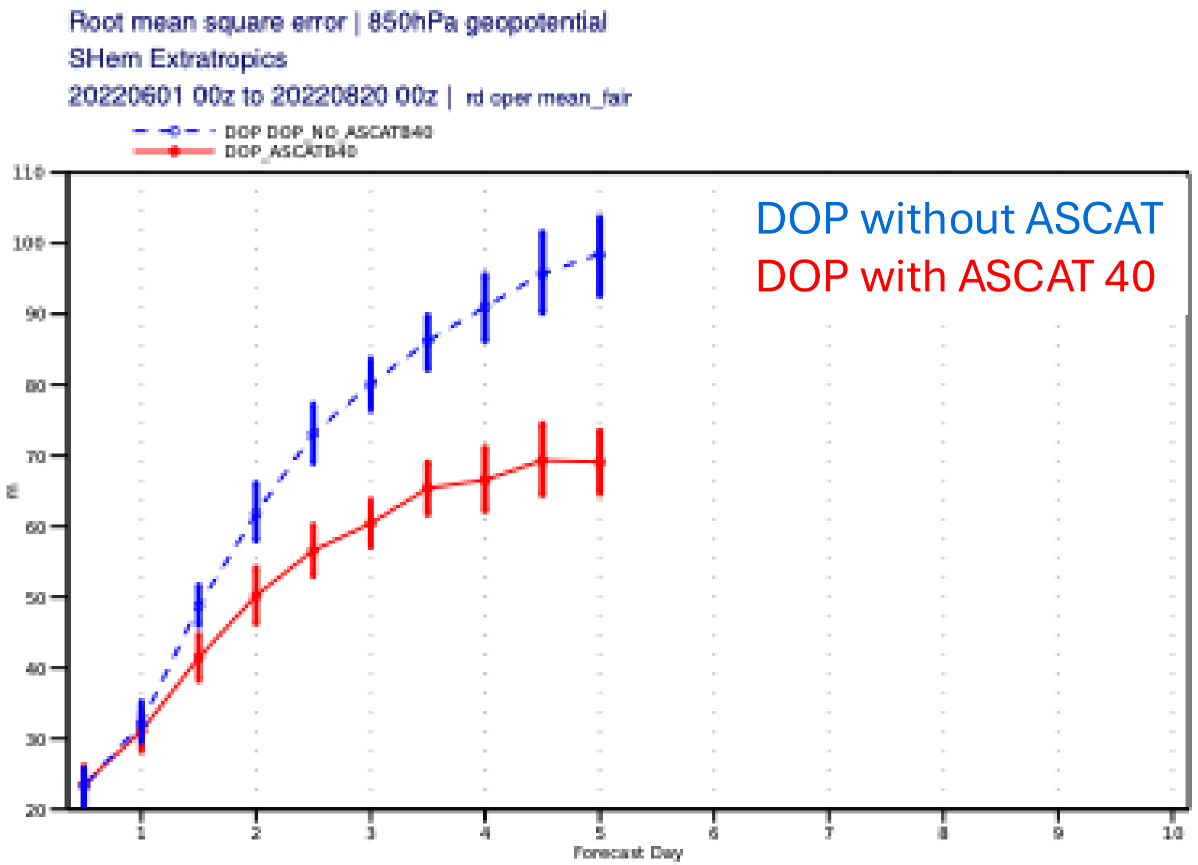
Impact of ASCAT backscatter 40 degrees in GRAPH-DOP on T2m

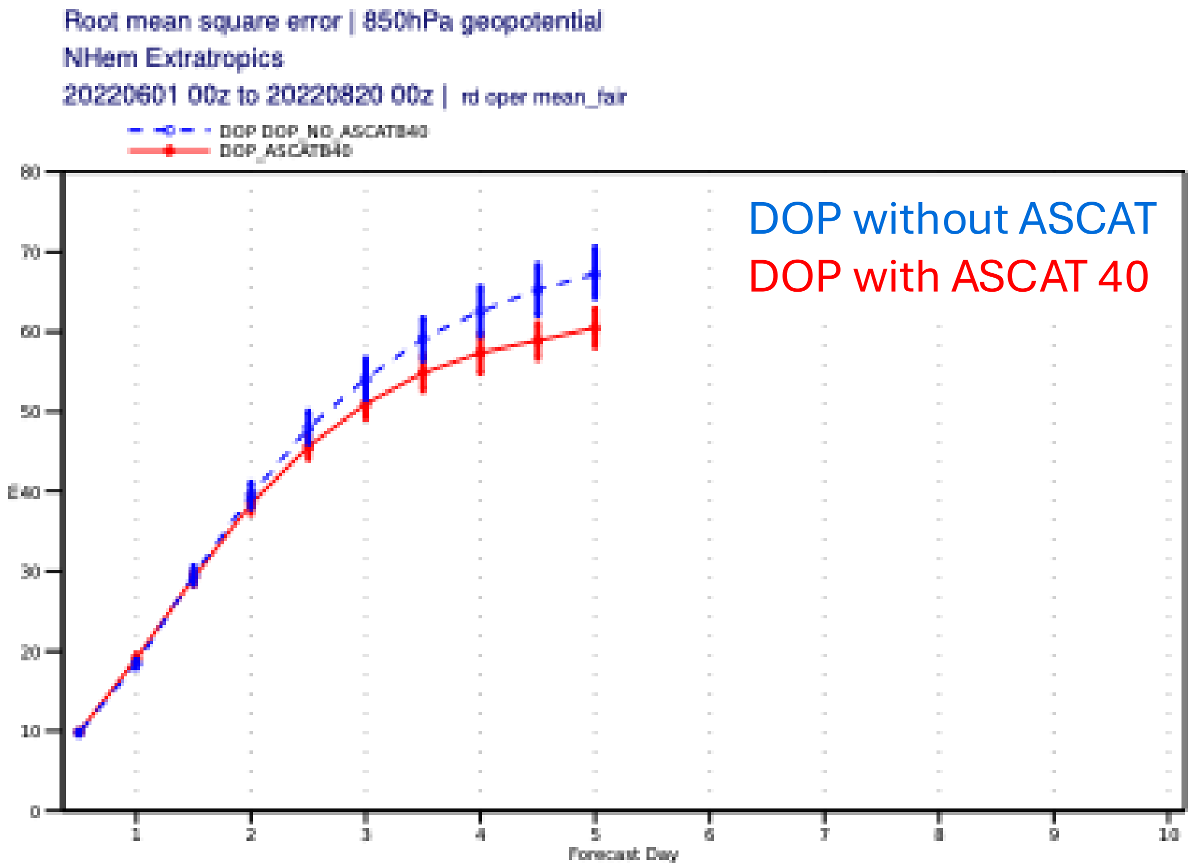




ASCAT backscatter 40 degrees improves GRAPH-DOP T2m forecast in the southern hemisphere, but slightly degrades in the northern hemisphere

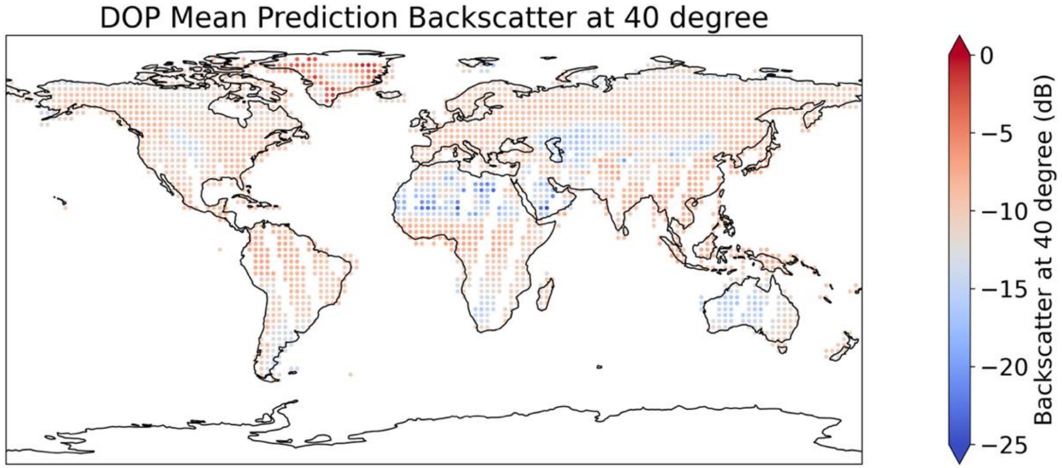
Impact of ASCAT backscatter 40 degrees in GRAPH-DOP on Z850 hPa

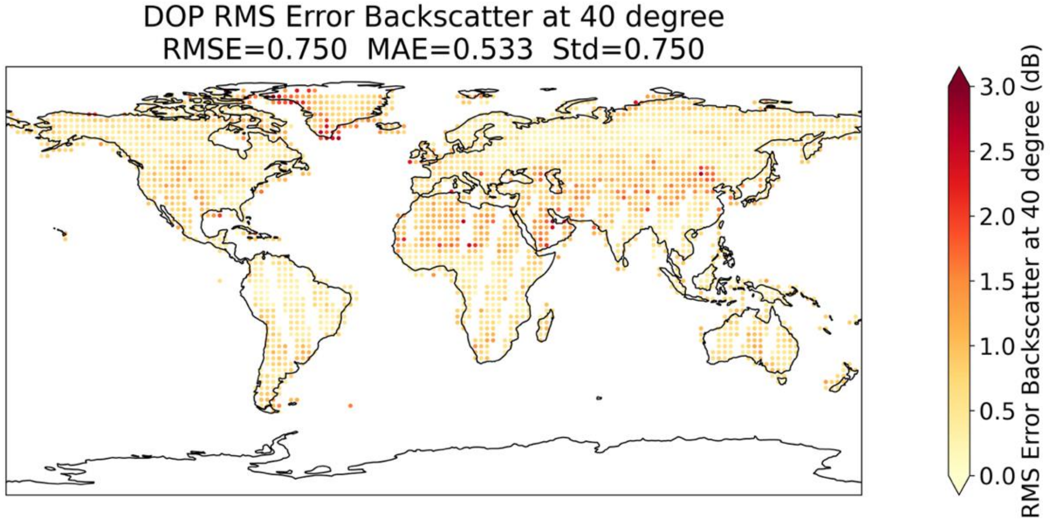




ASCAT backscatter 40 degrees globally improves GRAPH-DOP geopotential height forecast

AI-DOP forecasts of ASCAT backscatter 40 degrees

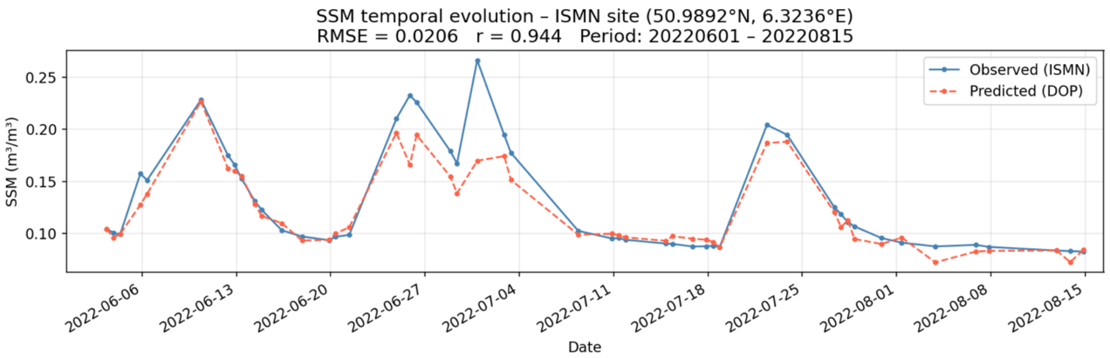


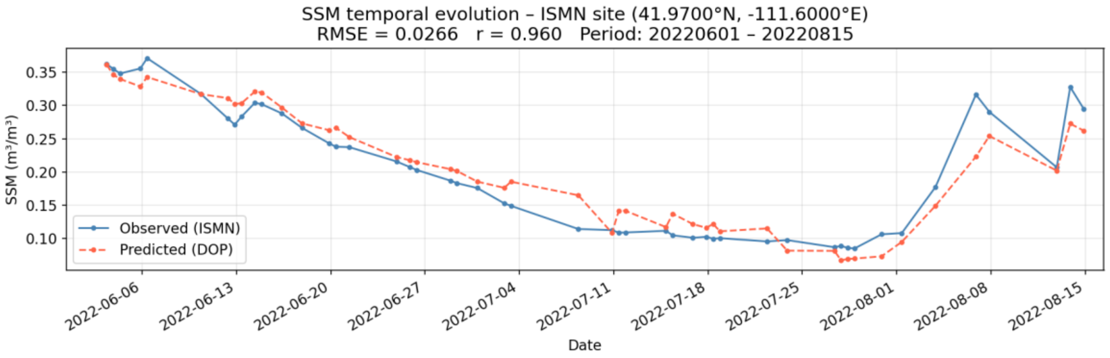


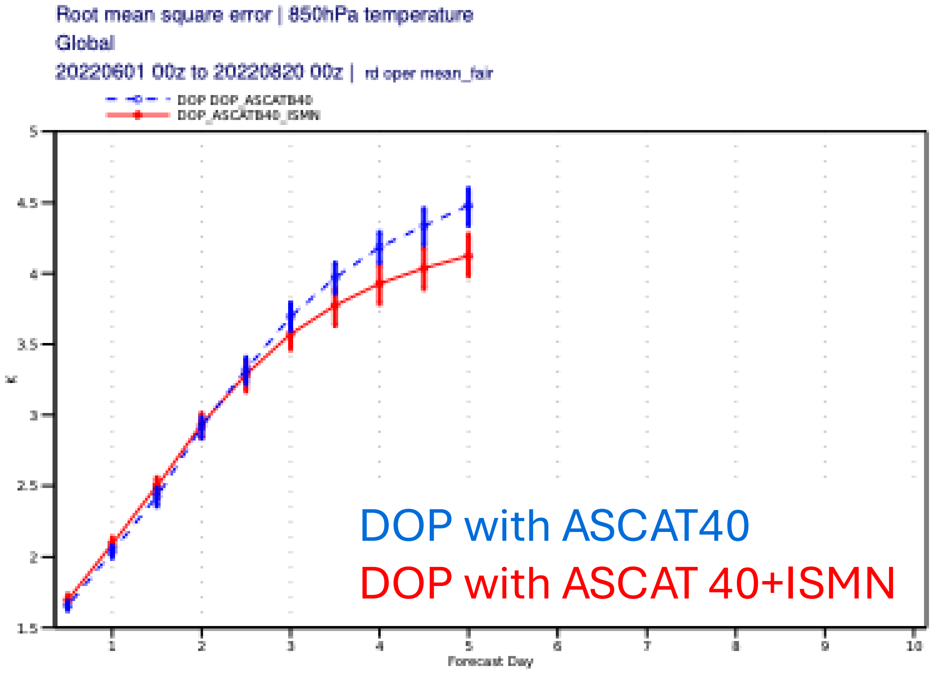
- AI-DOP learning to forecasts ASCAT 40 degrees over land surfaces
- Influences upper air forecasts

Land surface in data driven machine learning forecasts models: AI-DOP

And adding in situ soil moisture from the ISMN (International Soil moisture network)
→ Learning the physical relation between soil moisture and multiple satellite information
→ Extending the AI-DOP model to soil moisture







Sébastien Garrigues et al.

- Promising preliminary results with soil moisture
- Next steps: consolidation of the AI-DOP configuration and extention to deeper soil level and more variables (soil temperature, snow depth, ...)

Exploring AI-DOP for reanalysis

anemoi datasets

AIFS-DOP version using gridded observations in the Anemoi framework

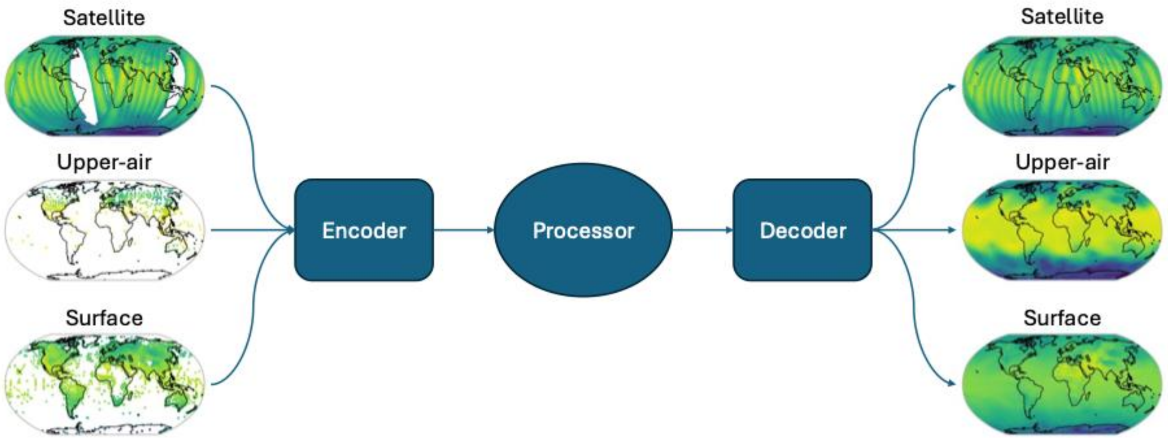


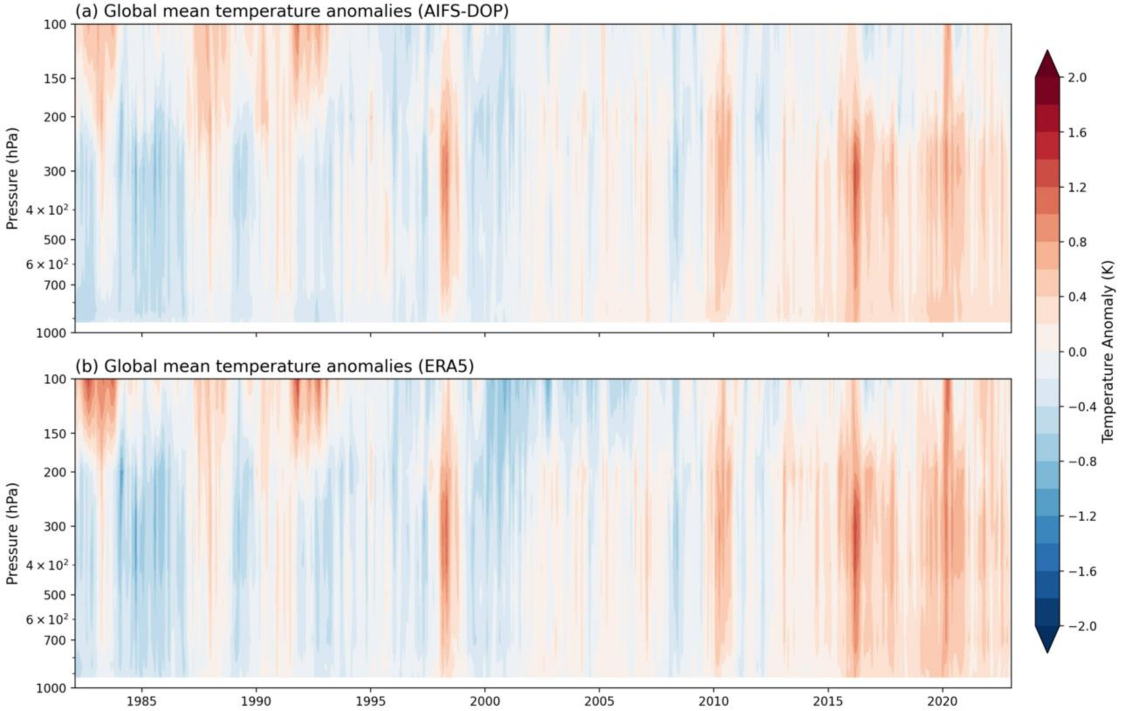
Table 1: Description of curated observation dataset

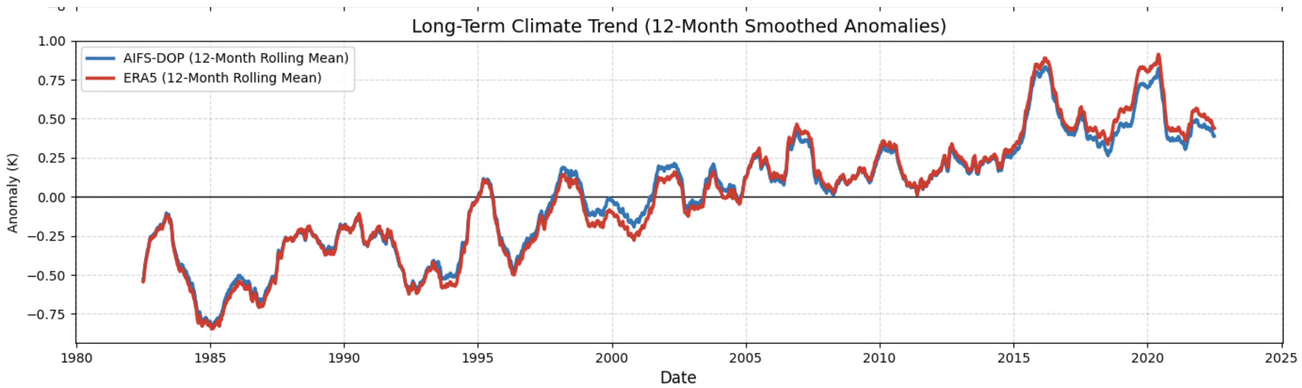
Category	Instruments	Period	Variables	Details
Infrared Sounder	HIRS	1990–2021	Brightness Temperatures	EUMETSAT Fundamental Data Record [EUMETSAT, 2024]
Microwave Sounders	MSU SSM/T-2 AMSU-A AMSU-B MHS ATMS	1990–2005 1994–2005 1998–2021 1998–2014 2005–2021 2012–2021	Brightness Temperatures	MSU taken from NOAA Climate Data Record [Zou et al., 2013], SSM/T2 taken from EUMETSAT Fundamental Data Record [EUMETSAT, 2020]
Surface Observations	SYNOP, Buoys, Ships	1990–2021	2t, 2d, msl, 10u, 10v	Existing ECMWF data archive
Upper-air Observations	Radiosonde, Aircraft, AMV	1990–2021	t, u, v, z on pressure levels	Existing ECMWF data archive
Geostationary Satellite	GridSat	1990–2021	Brightness Temperatures	NOAA Climate Data Record [Knapp et al., 2011]

Ewan Pinnington, Peter Lean, Eulalie Boucher et al.

Exploring AI-DOP for reanalysis

Ewan Pinnington et al., in prep 2026
Peter Lean et al., in prep 2026





Running 40 years global reanalysis took less than a day with AI-DOP
Very promising exploratory results with more surface observations to be included

Summary

- Strong and significant impact of land data assimilation analyses on NWP
- Coupled data assimilation → possibilities to enhance the exploitation of surface observations
- Transition towards using raw observations at level 1 to simultaneously analyse soil, vegetation and atmospheric variables
- Machine-learning forward models allow to enhance exploitation of current satellites (e.g. ASCAT), as well as passive microwave data, and to explore new land surface observations (e.g. SIF, GNSS-R),
- Integration of land surface variables in AIFS (soil moisture, temperature, runoff, snow cover, etc)
- Potential of surface observations into end-to-end Machine learning systems such as Graph-DOP
  → unprecedented capability to exploit surface sensitive observations for weather prediction & climate reanalyses

2026 Workshop on Machine Learning for Land and Hydrology

3-5 November | Reading, UK

Prof Pierre Gentine
Columbia University

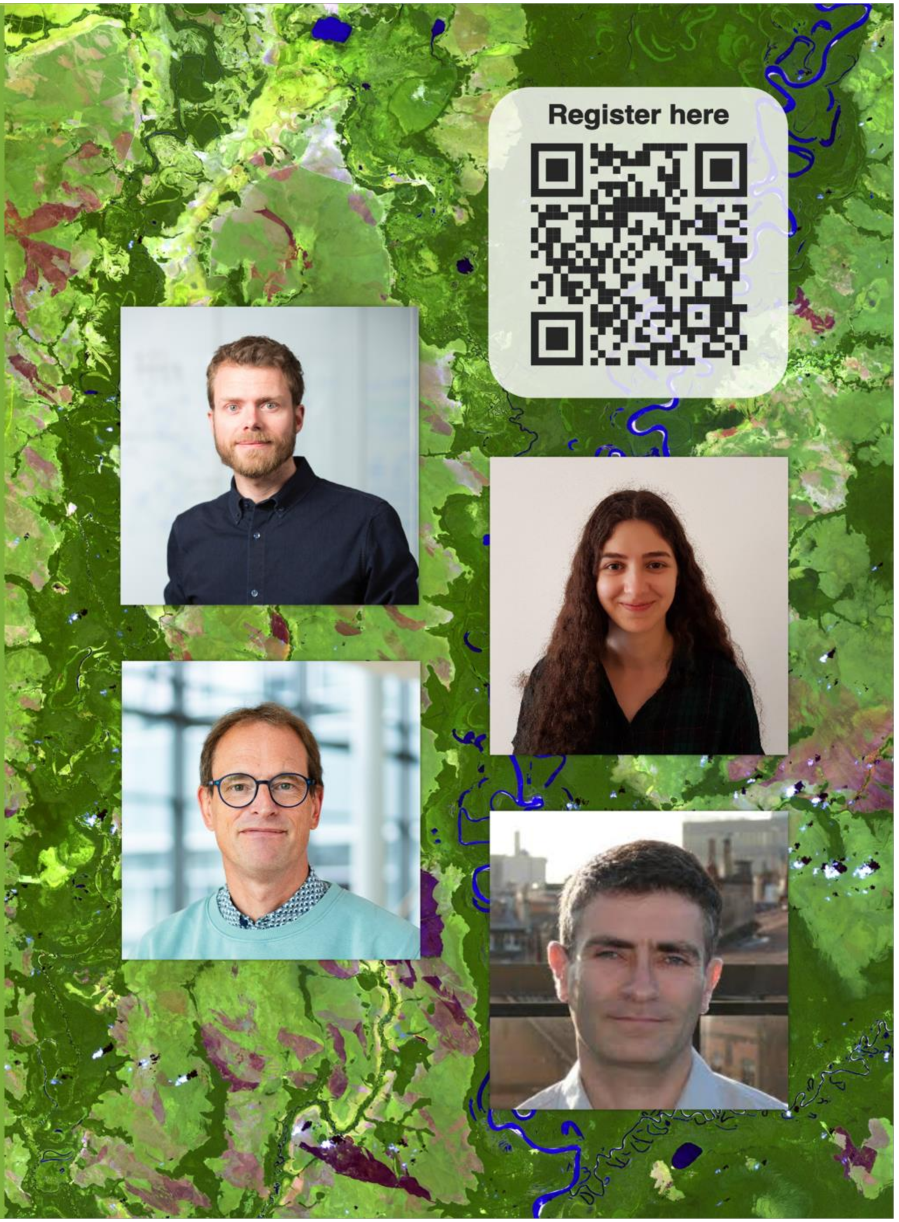
Rotem Mayo
Google

Prof Markus Reichstein
Max-Planck Institute for Biogeochemistry

Dr Jean-Christophe Calvet
CNRM | Météo-France

More speakers to be confirmed!

ECMWF



https://events.ecmwf.int/event/530/

Patricia de Rosnay

Coupled Assimilation Team Leader

Patricia.Rosnay@ecmwf.int

www.ecmwf.int