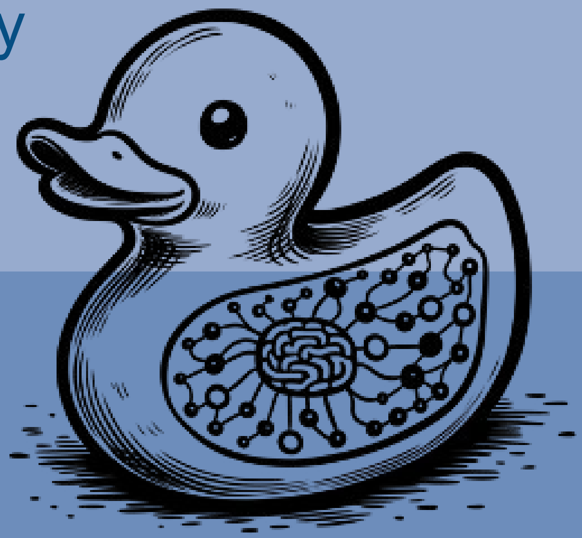
Funded by the European Union
Destination Earth implemented by ECMWF esa EUMETSAT

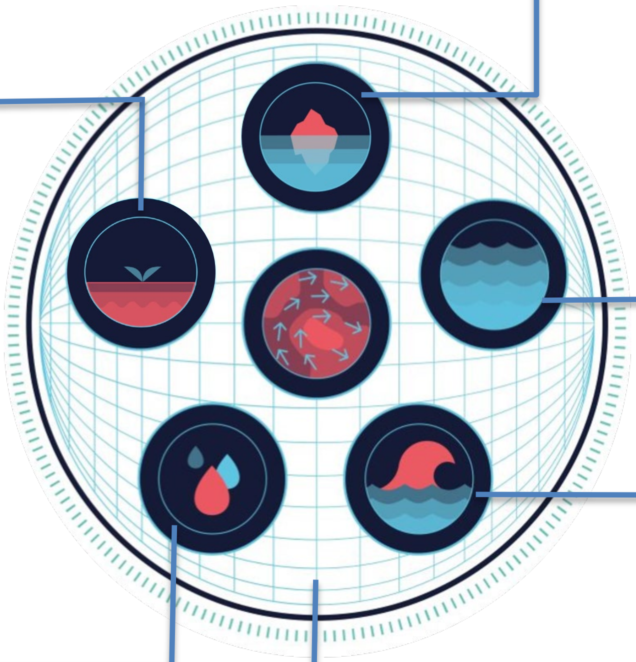
Destination Earth: Earth System Modelling

Ocean, Sea Ice, Land, Waves, and Hydrology



Nina Raoult nina.raoult@ecmwf.int
on behalf of the core ML Earth System Components group:
Rachel Furner, Sara Hahner, Ewan Pinnington, Nina Raoult, Mario Santa Cruz, Maria Luisa Taccari, Kenza Tazi, Lorenzo Zampieri
and many other contributors

Advancing towards an Earth system ML Model with DestinE



Land

Ewan

Nina

Ocean, ice and waves

Lorenzo

Rachel

Sara

Hydrology

Maria Luisa

Kenza

Technical support

Mario

Why has ML worked so well for NWP?

- **Data abundance**:
  Decades of high-resolution reanalyses and observational products.

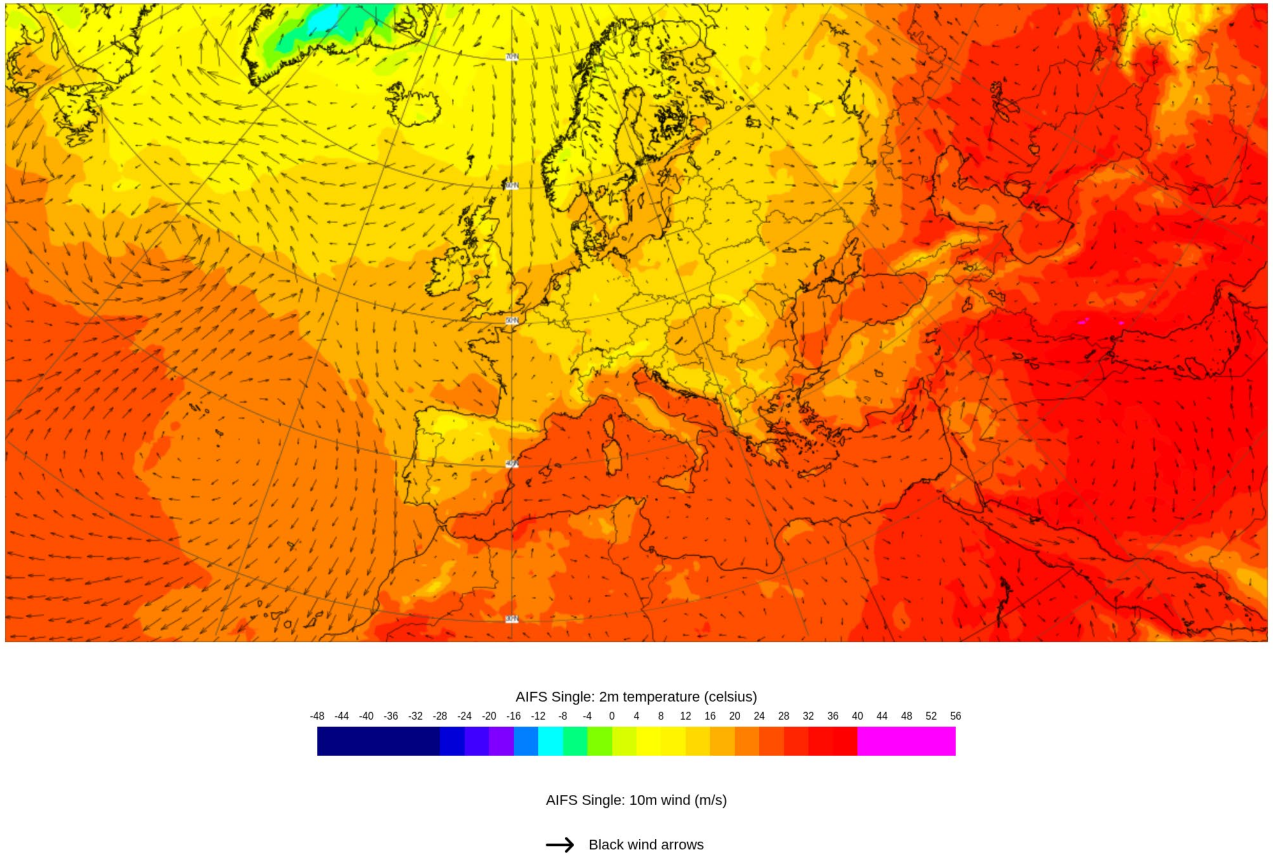
- **Predictability & formulation**:
  NWP to be posed as an initial-value, sequence-to-sequence learning problem.

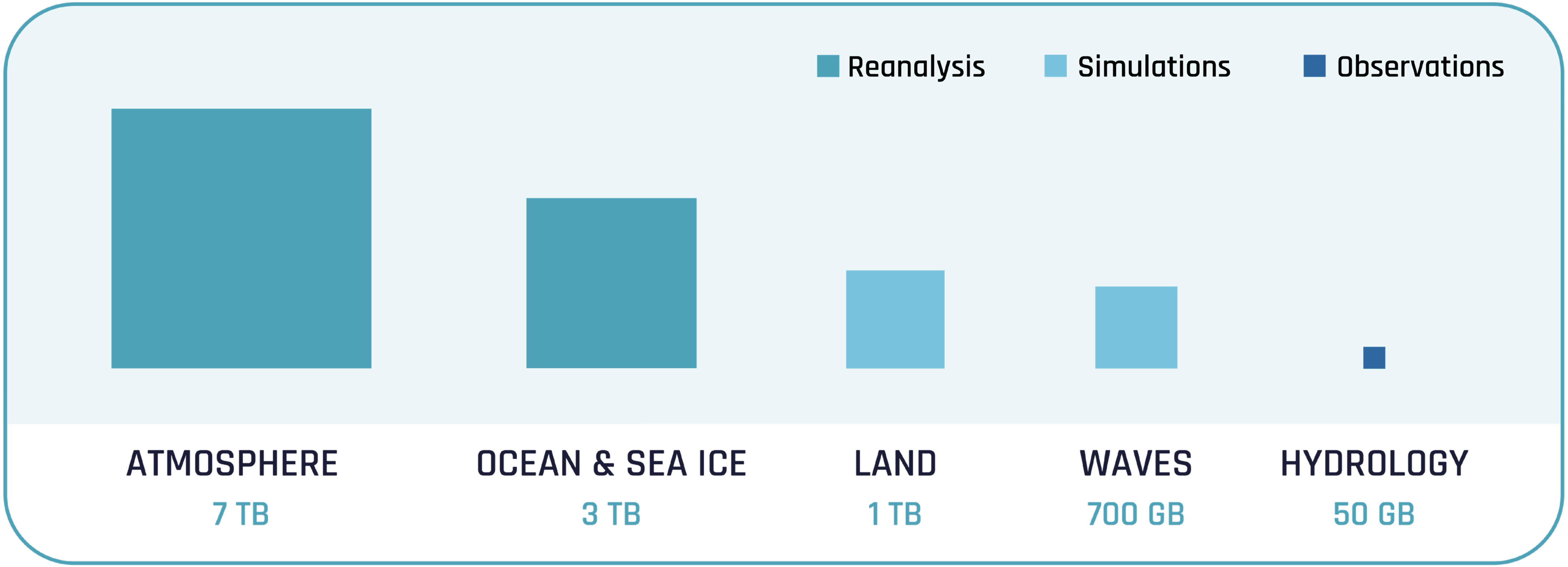
- **Bounded horizon**:
  Forecasts usually cover days to ~2 weeks, limiting error accumulation.

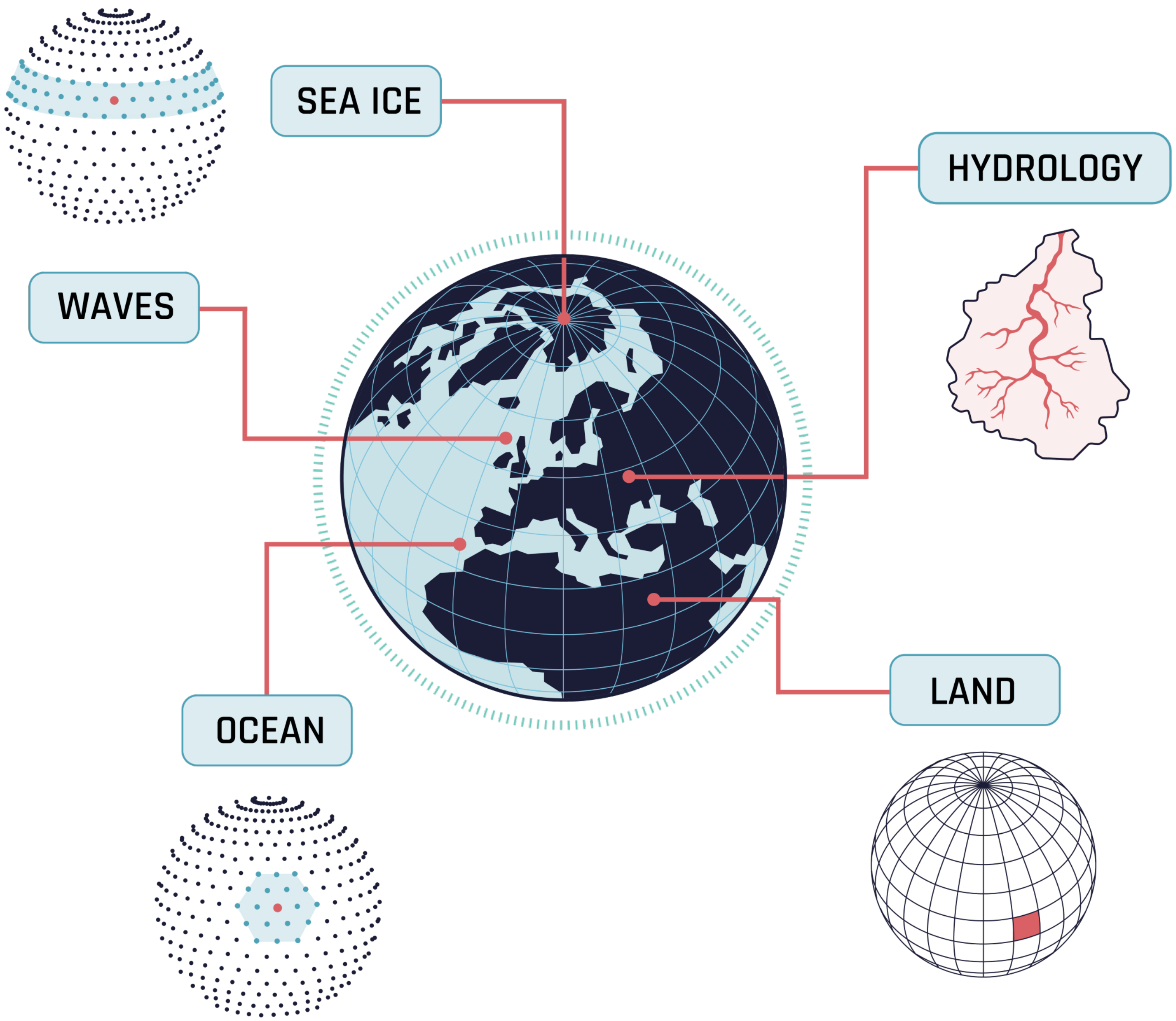
- **Community Momentum:**
  Broad engagement and open-sourced toolkits

AIFS Single: 2 m temperature and 10 m wind

Base time: Tue 19 Aug 2025 06 UTC Valid time: Tue 19 Aug 2025 06 UTC (+0h) Area : Europe




Reanalysis
Simulations
Observations
ATMOSPHERE
7 TB
OCEAN & SEA ICE
3 TB
LAND
1 TB
WAVES
700 GB
HYDROLOGY
50 GB

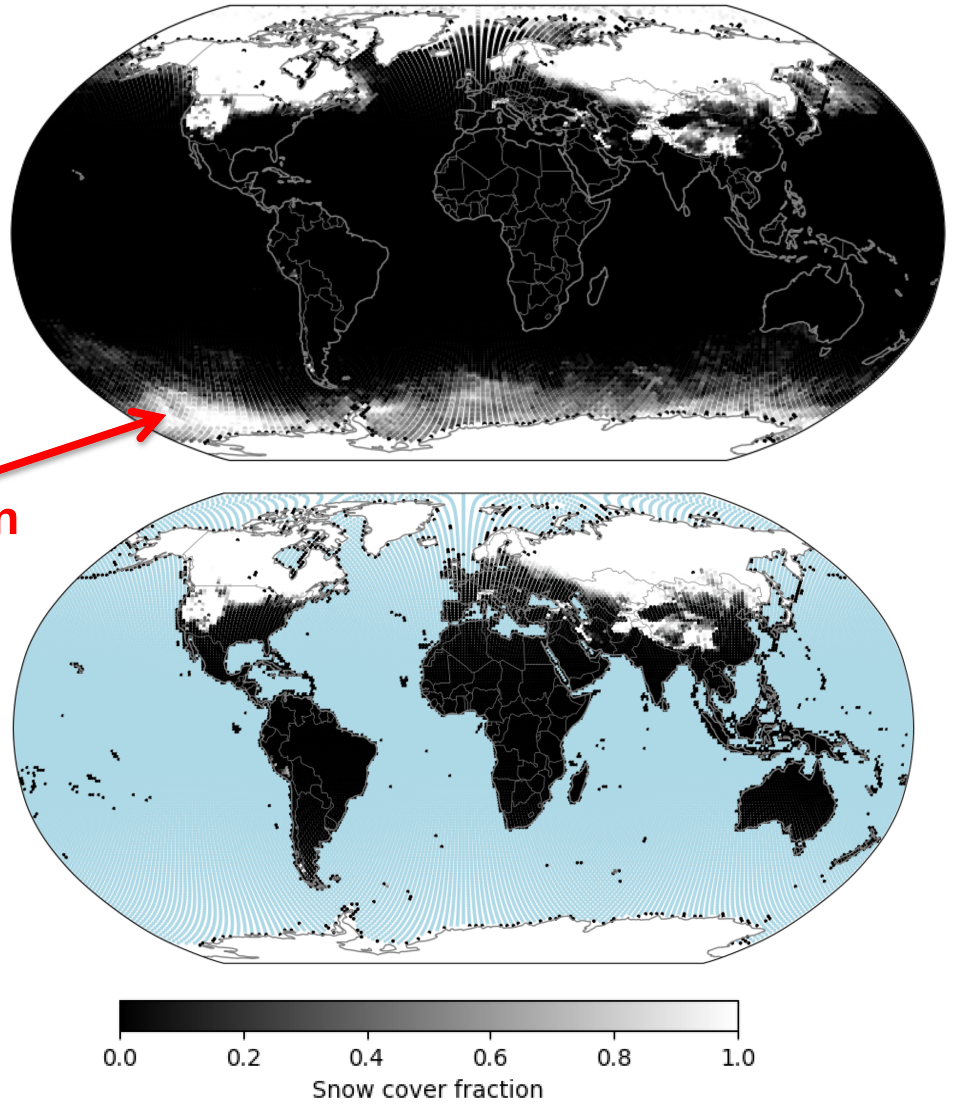

SEA ICE
HYDROLOGY
WAVES
OCEAN
LAND

Dealing with missing values

Without masking NaNs

Snow cover over the ocean

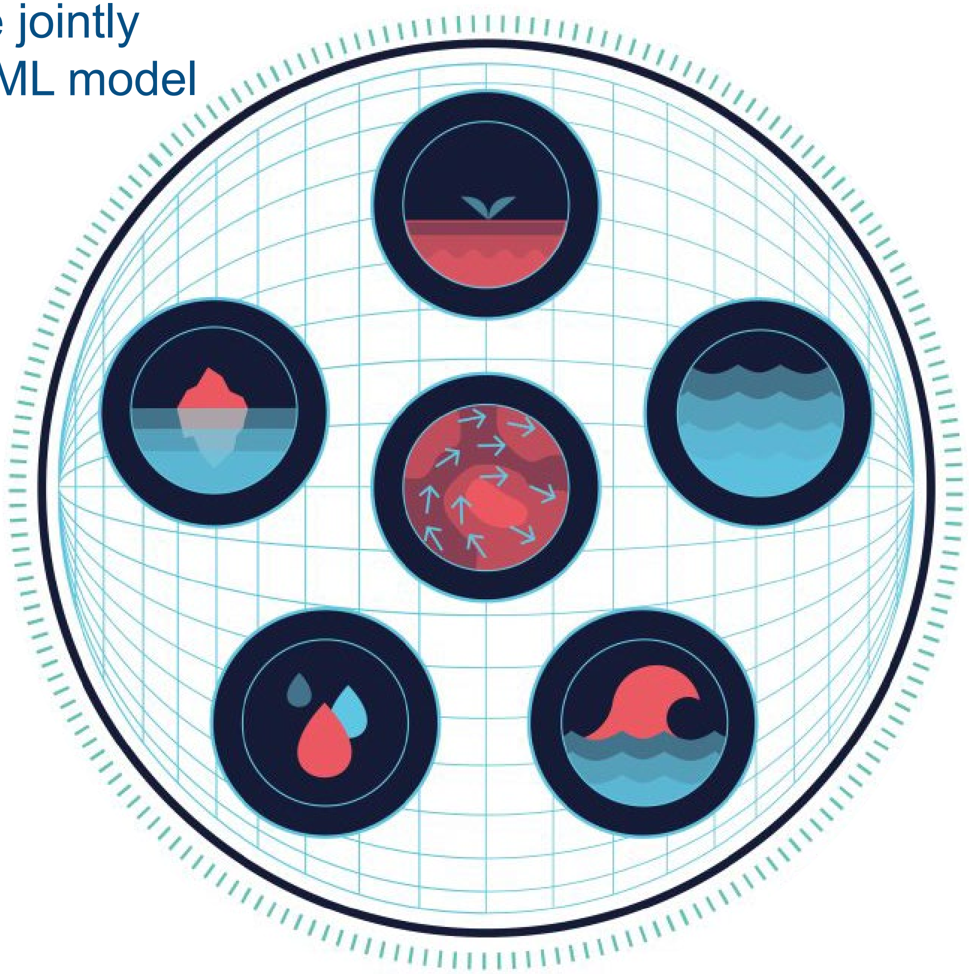
Masking NaNs (lightblue)



- In training: NaNs are replaced by values and **masked in the loss function with zeros**.
- Nans can be replaced with zeroes in the normalised space
- At inference: the model **'predicts'** values **in the masked areas** – these are **removed** in a **post-processing** step.
- Ability to handle **moving NaNs** during training and inference – relevant for ML wave modelling.

Investigating approaches for multi-component ML models

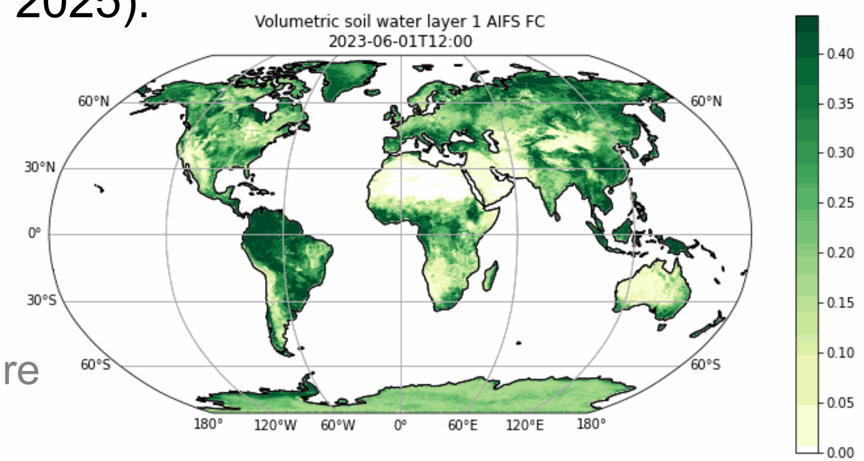
One jointly
learnt ML model



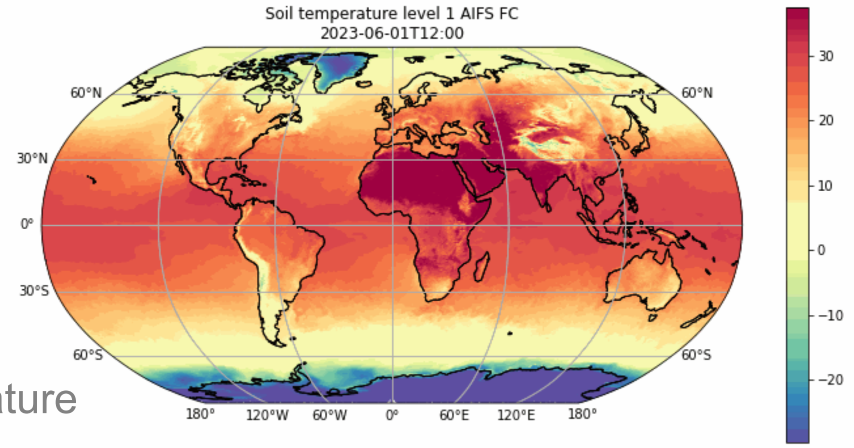
"joint model"

The AIFS is already a joint model

AIFS v1 (Feb 2025):



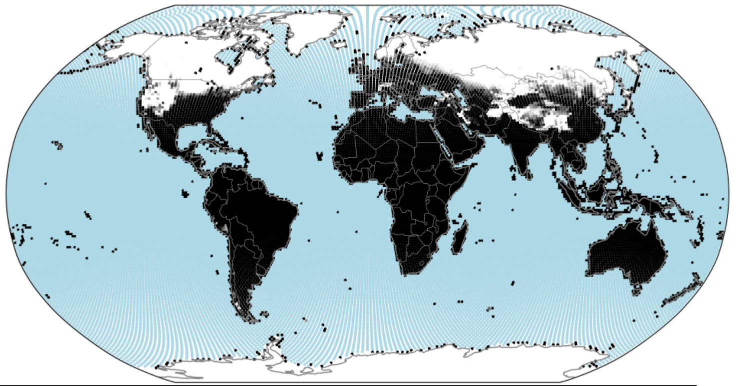
Soil moisture



Soil temperature

+ runoff

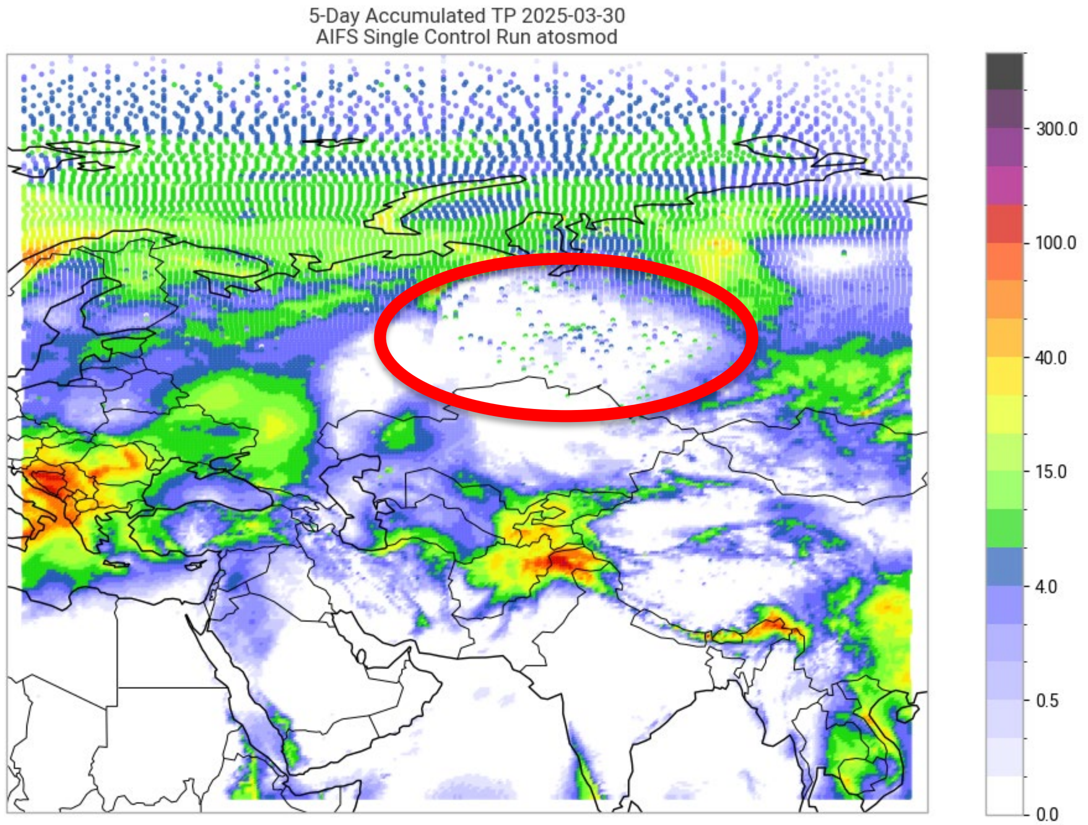
AIFS v2 (May 2026):

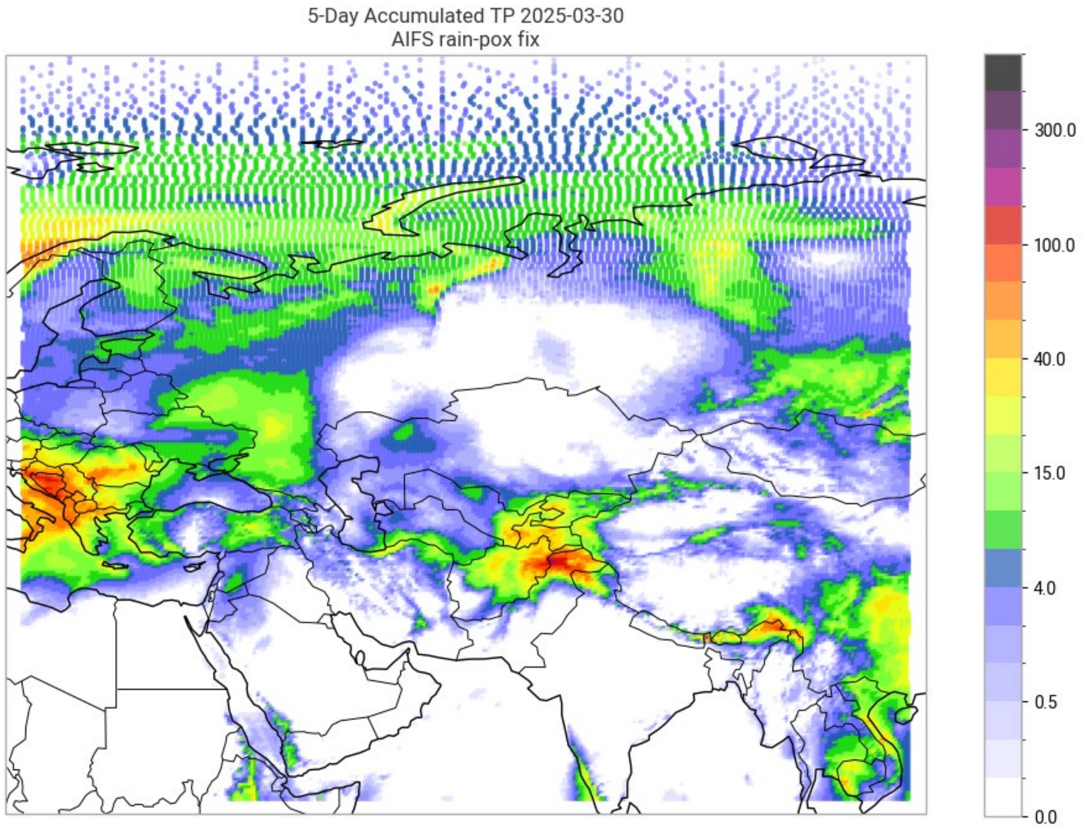


Snow

Wa

Land variables in the AIFS – soil moisture

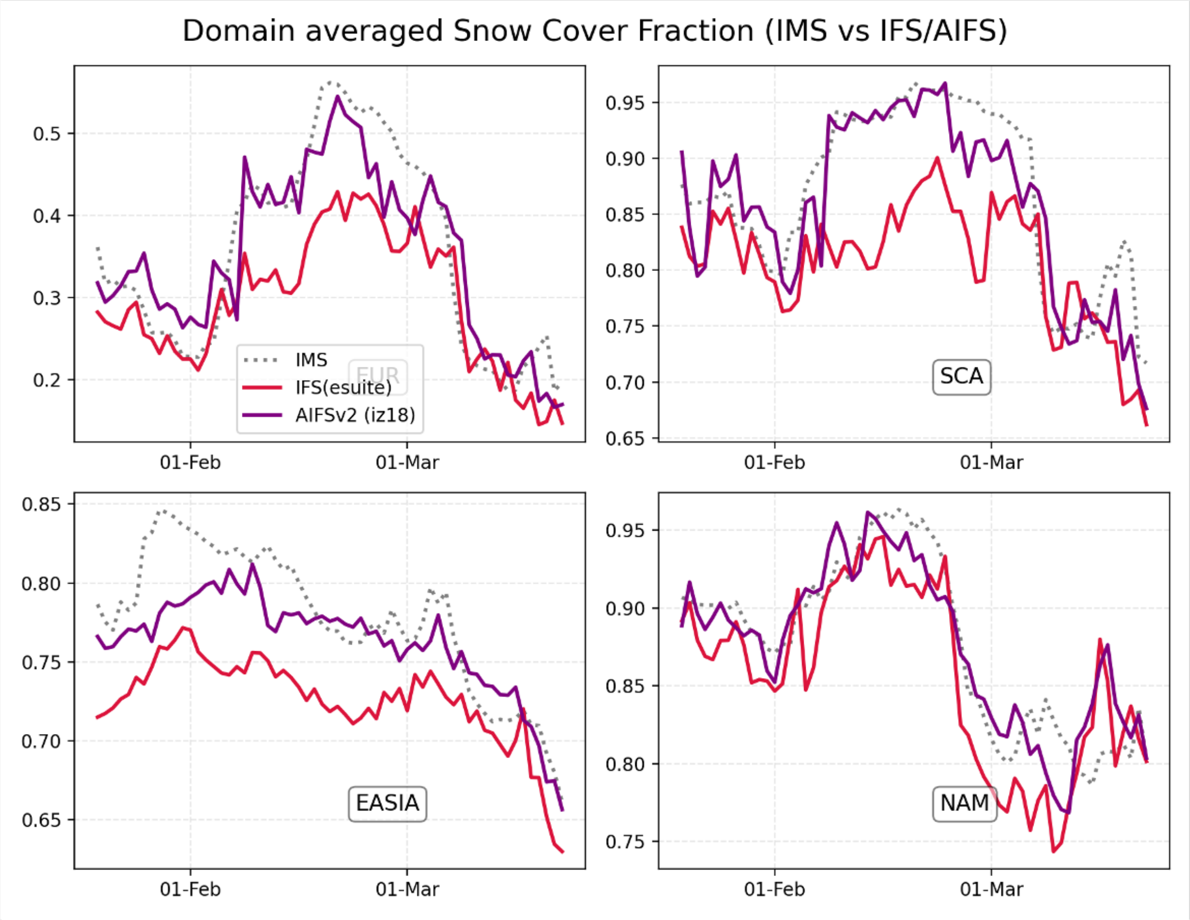




Land variables in the AIFS – snow cover

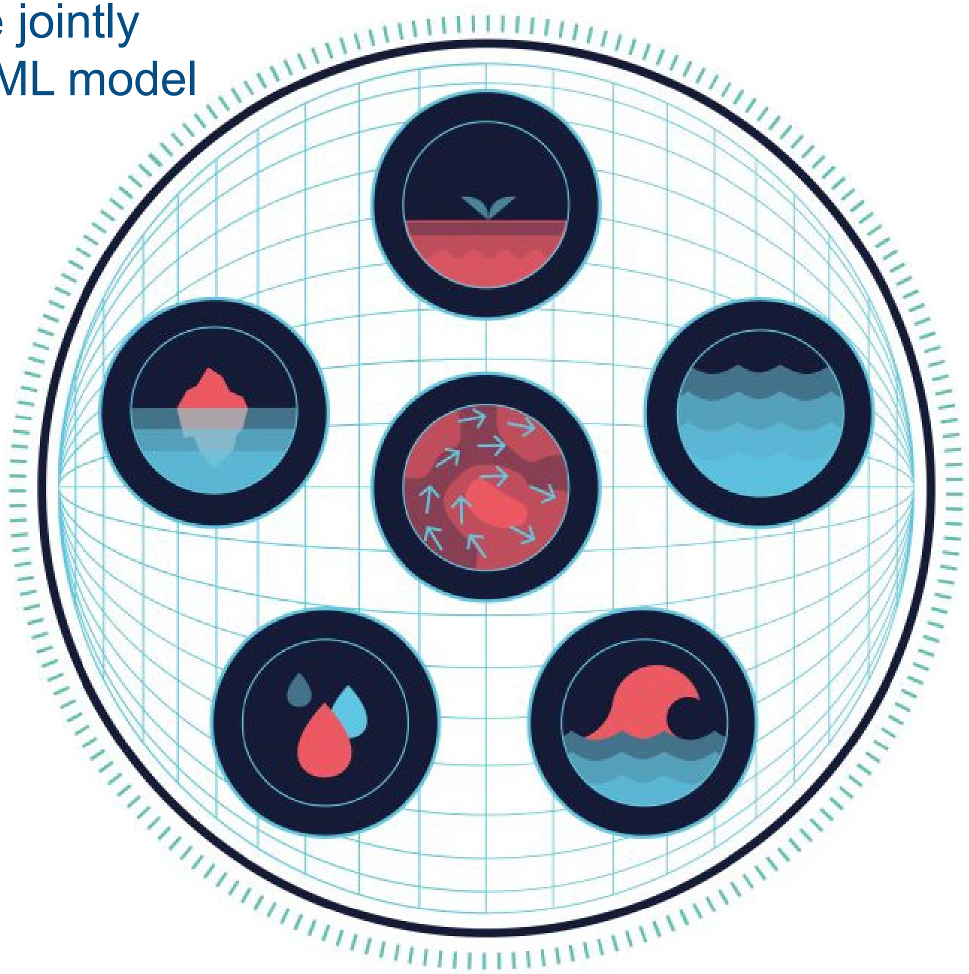
- By training on ERA5, the AIFS has learnt the snow DA

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



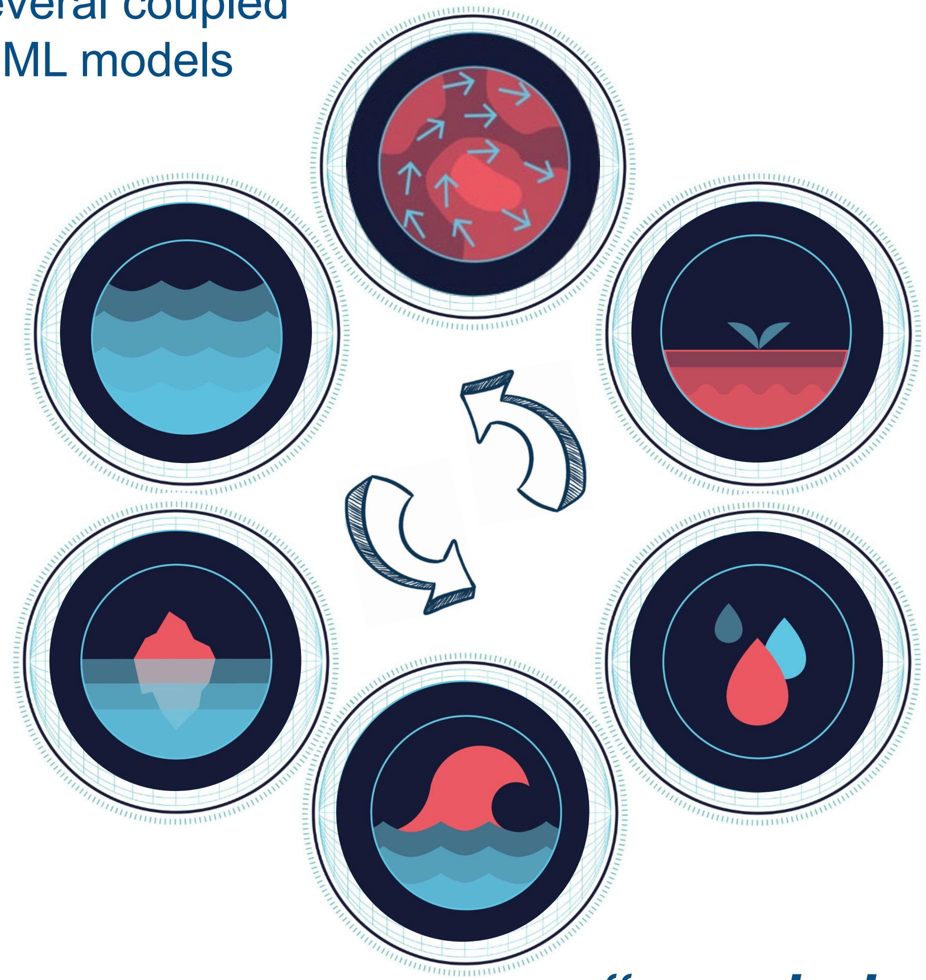
Investigating approaches for multi-component ML models

One jointly
learnt ML model



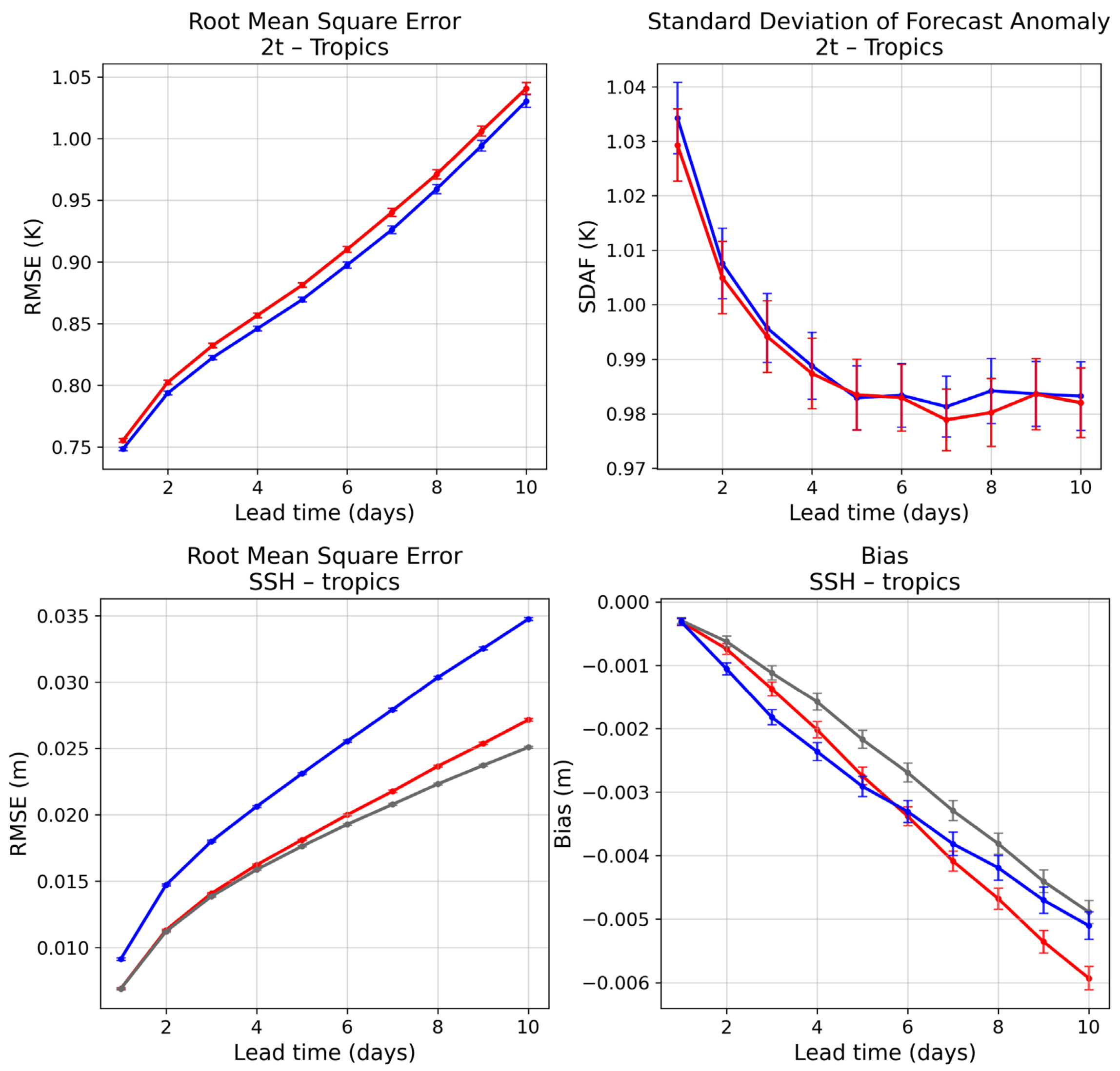
"joint model"

Several coupled
ML models



"coupled model"

Exploring coupling strategies: an example from the ocean



ATMOSPHERIC Perspective

--- Atmosphere + ocean trained **separately** and then **coupled**
--- Atmosphere + ocean trained **together**
--- Atmosphere **forced** by the ocean

- Both approaches perform similarly
- Slight better skill & more active forecasts when components trained together
- Atmospheric forced by the ocean is similar to coupling them, suggesting a reliance on intrinsic ocean representation over forcing.

OCEAN Perspective

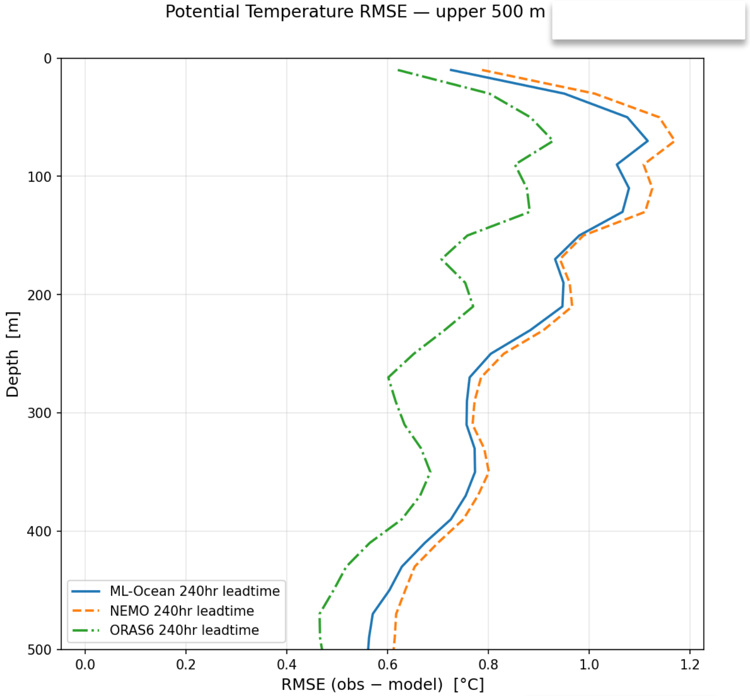
--- Atmosphere + ocean trained **separately** and the **coupled**
--- Atmosphere + ocean trained **together**
--- Ocean **forced** by the atmosphere

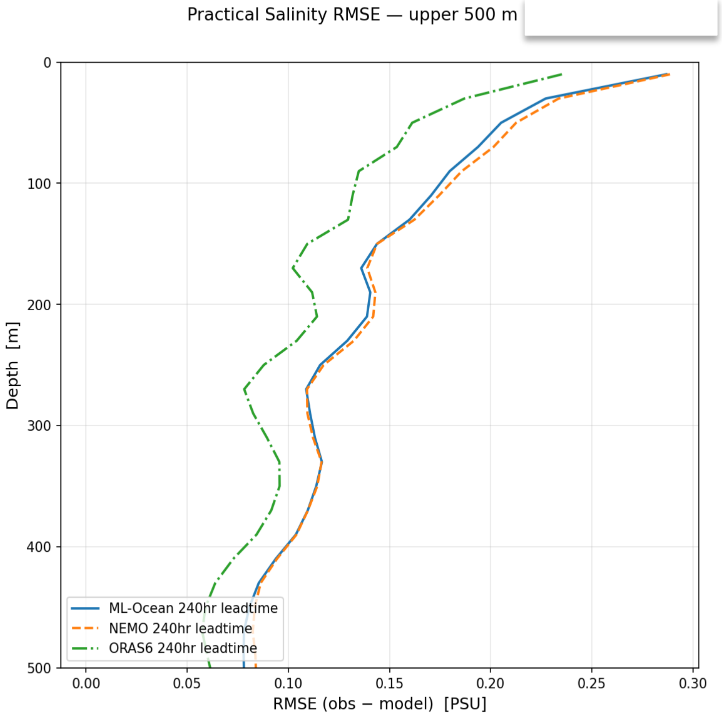
- Models trained separately substantially better for ocean fields
- Forced model perfect best, indicating the atmosphere has a driving role

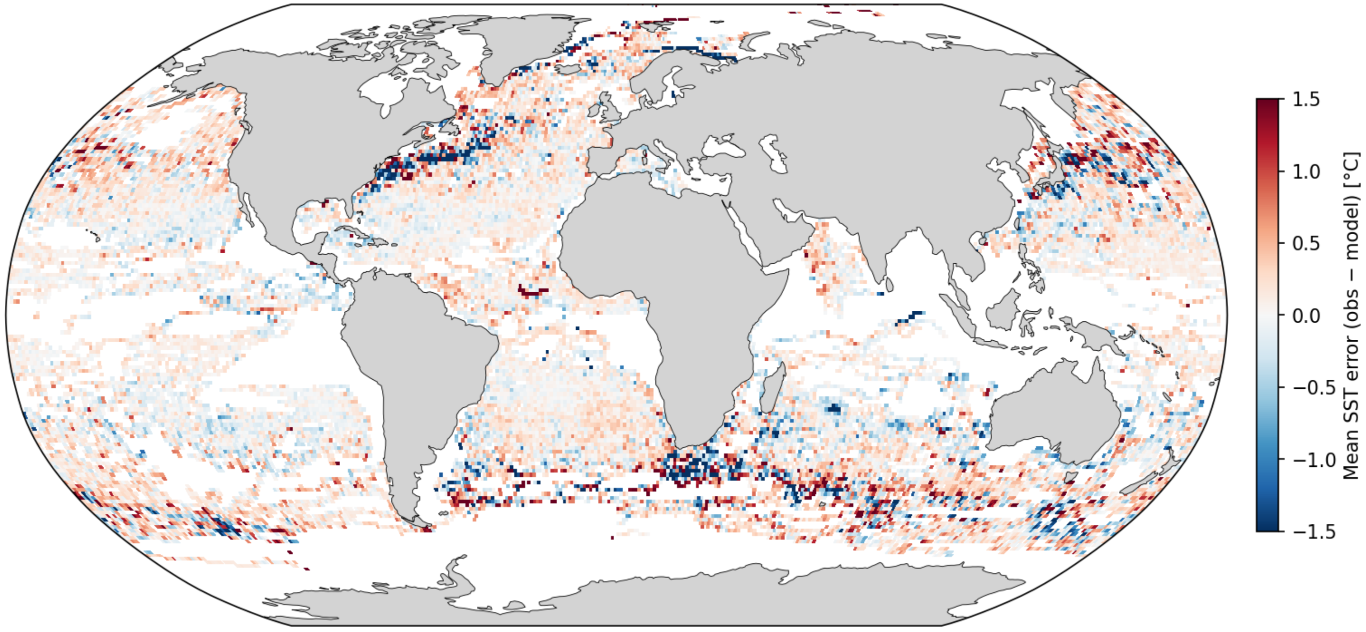


Highlights from ML component models

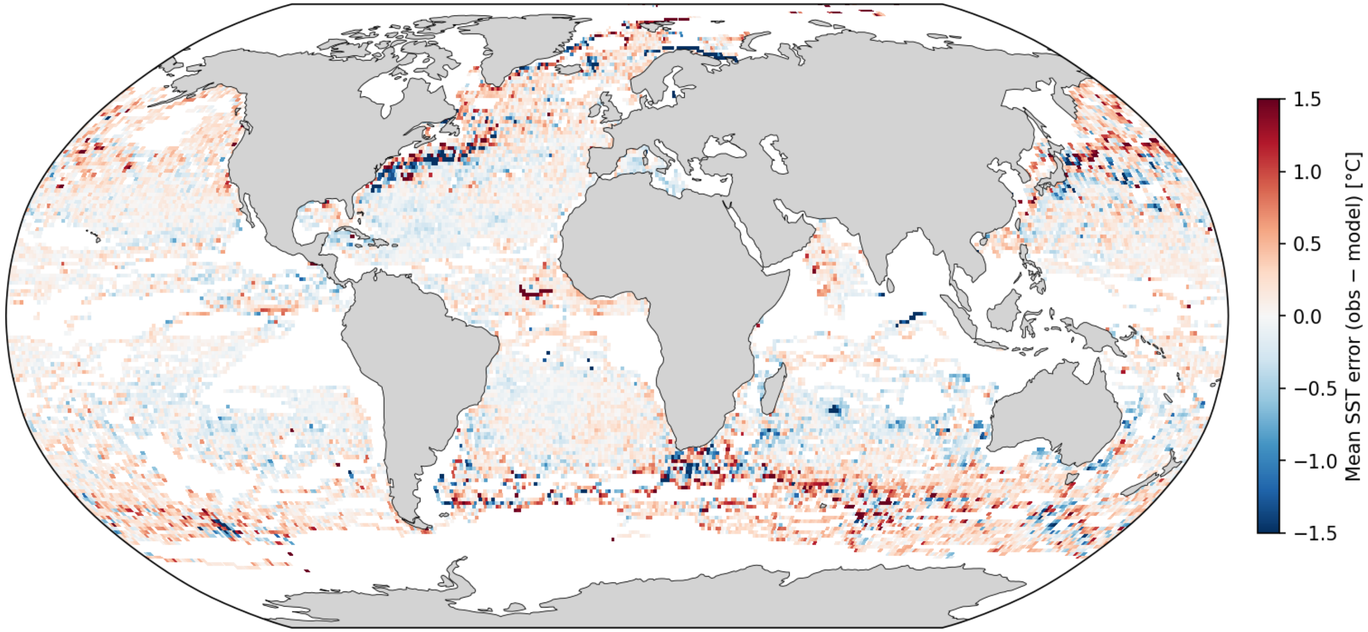

ML-Ocean model comparable to physics based NEMO





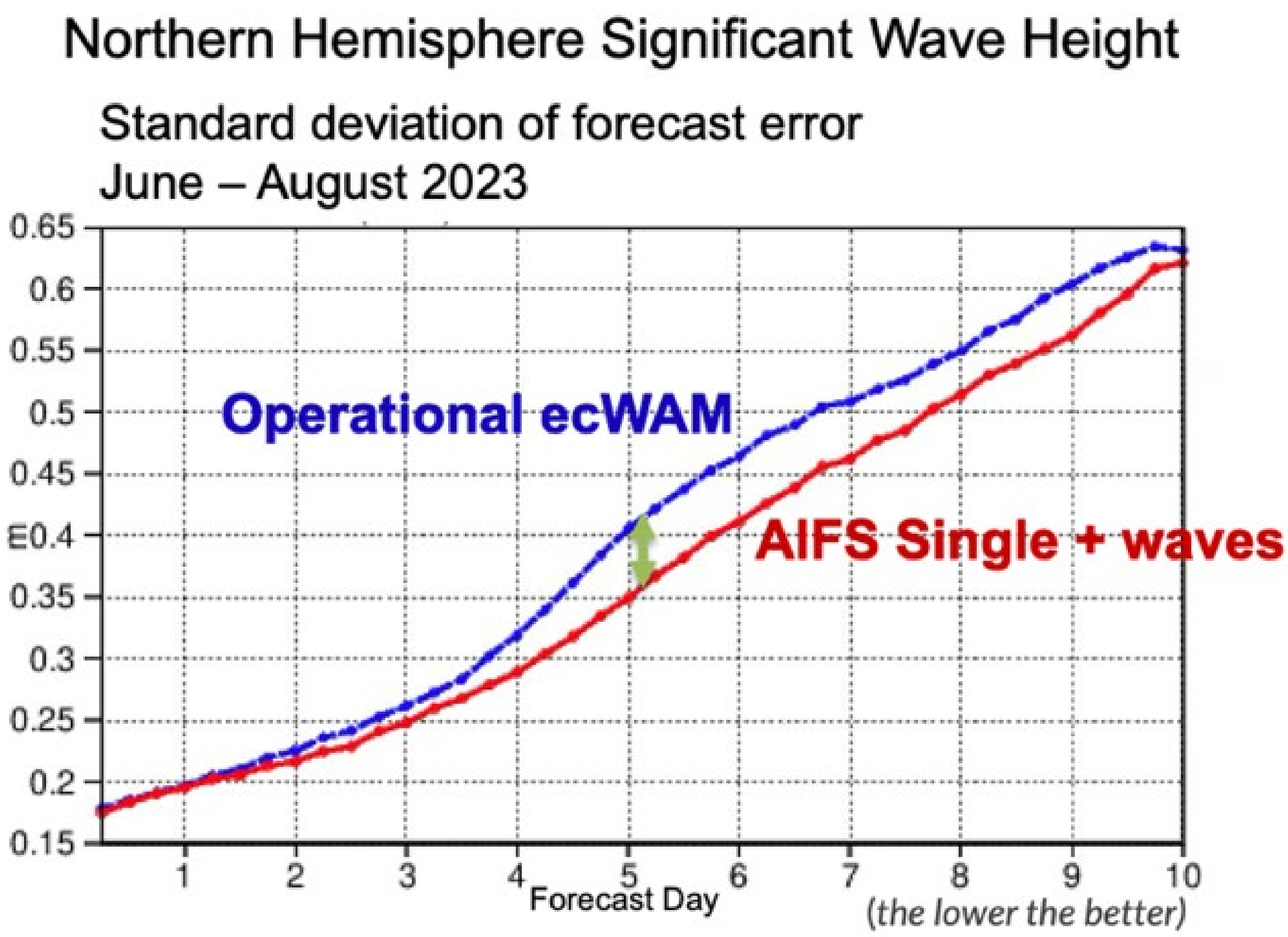


NEMO



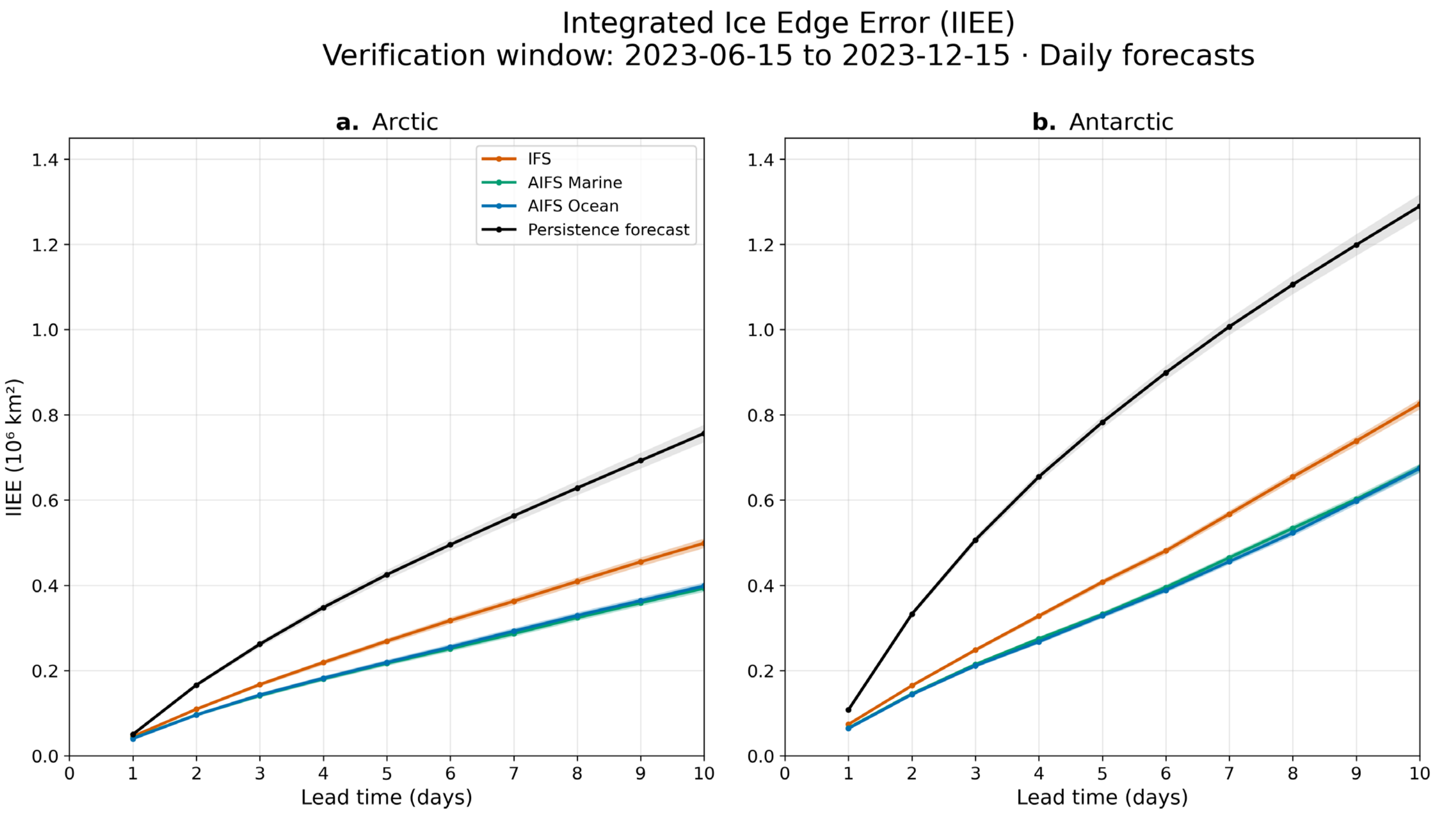
ML-
ocean

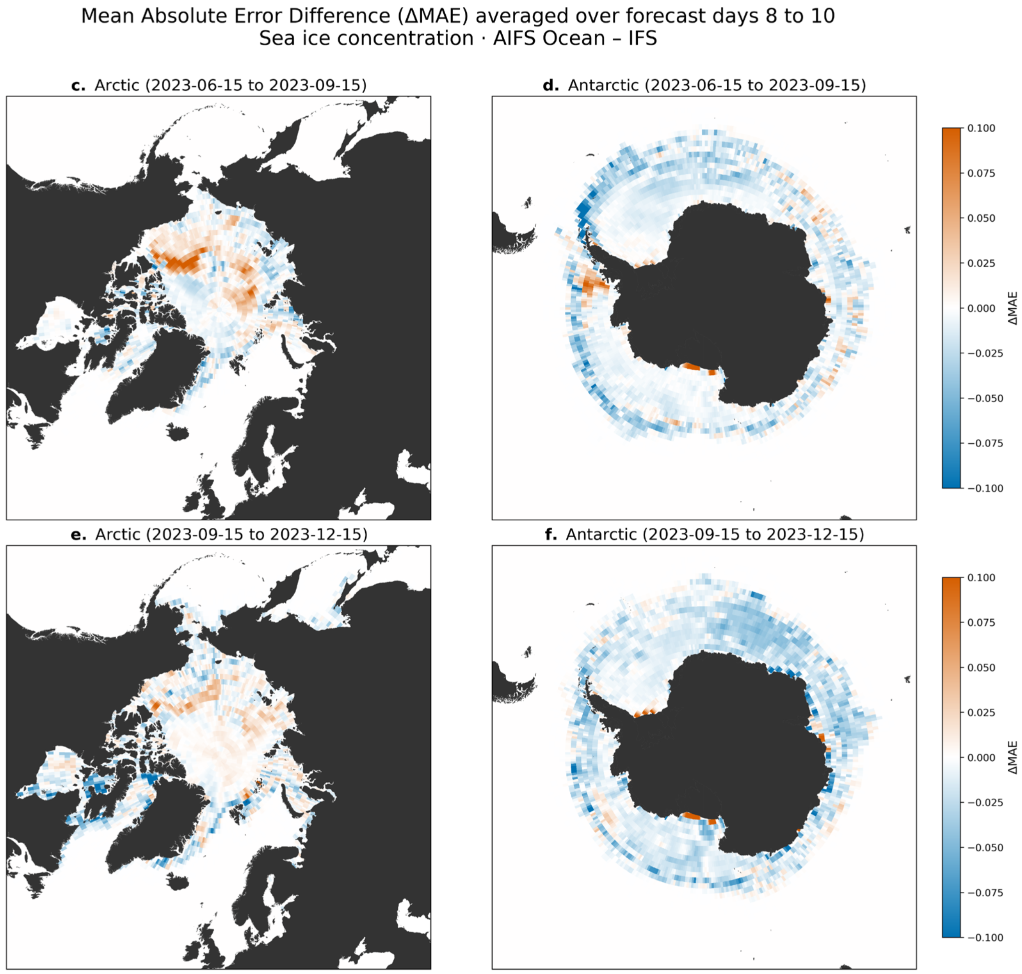
Data-driven wave forecasts reduce error by 10%



Significant wave height forecast errors are reduced by approximately 10% for medium-range forecasts against operational wave model.

Advancing sea ice edge position predictions at the medium range





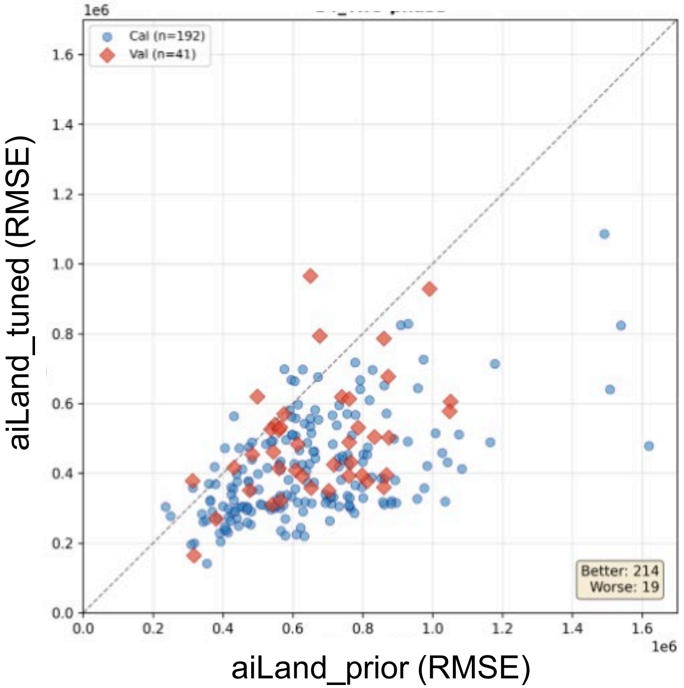
Implicit modelling:
Sea ice representation through waves

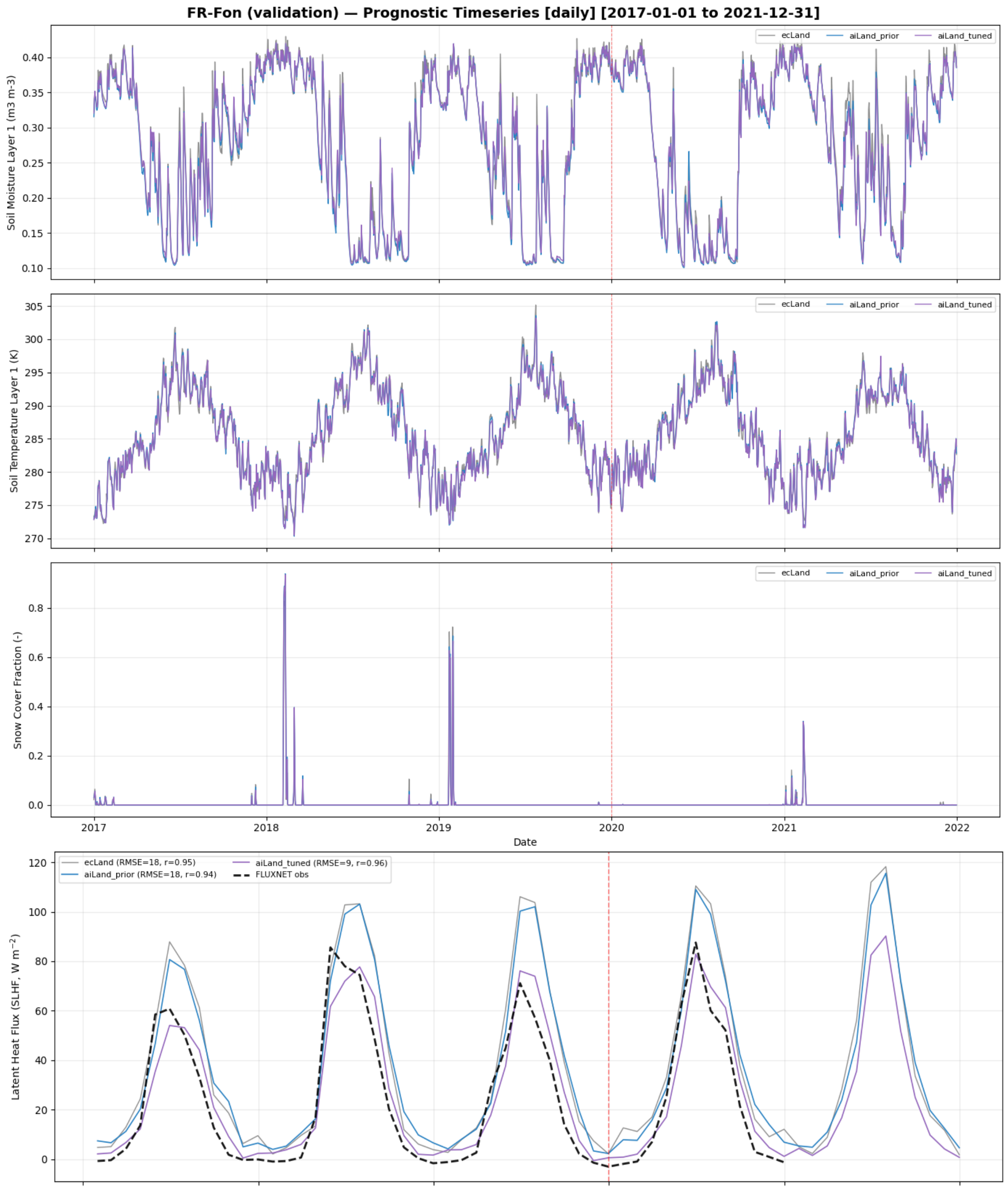
- Under ice, the significant wave height (SWH) is dramatically reduced
- The ML model learns an implicit representation of sea-ice – the region of reduced SWH realistically varies over time
- We see this *implicit representation* throughout Earth system models
- The models may not be coupled, but the datasets are

Ensuring physical consistency

Land emulation

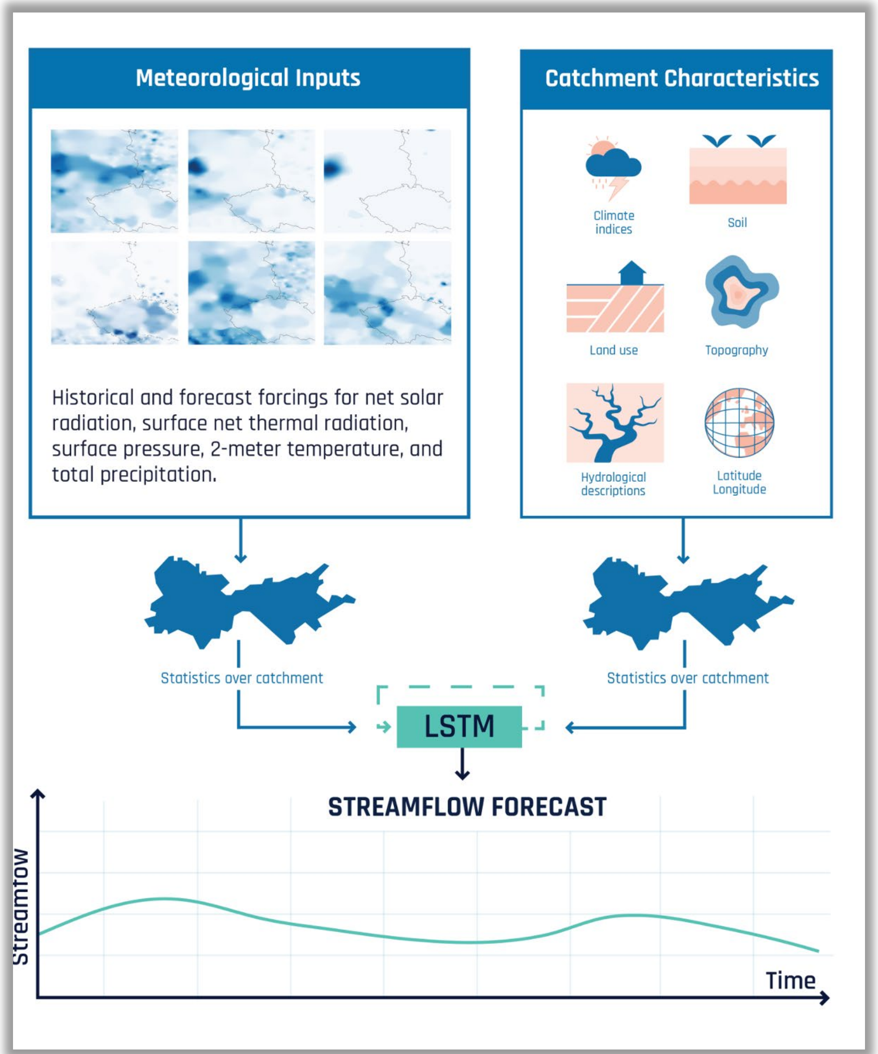
- Developing an emulator of ECMWF's land surface model:
  - Trained on physical model output
  - Fine tuned on observations (in situ FLUXNET)





Hydrology

AIFL: Towards ECMWF's data-driven flood forecasting system



- Flooding is widespread and the most destructive natural disasters worldwide
- River discharge is a key independent diagnostic of hydrological systems

AIFL is a Global LSTM-based flood forecasting trained with up to 20,000 basins

Conclusions

- **Successful prototypes** have been developed across Earth system components.
- **Extending ML success** beyond the atmosphere requires tailored approaches to data, scales, and physical processes.
- **Progress relies on unifying expertise** across domains.
- **Technical advances** are essential to overcome challenges in adapting NWP methods to other components.

> ➢ *many of these development have already fed into **operational** models at ECMWF!!!*

2026 Workshop on Machine Learning for Land and Hydrology

3-5 November

Reading, UK

Prof Pierre Gentine

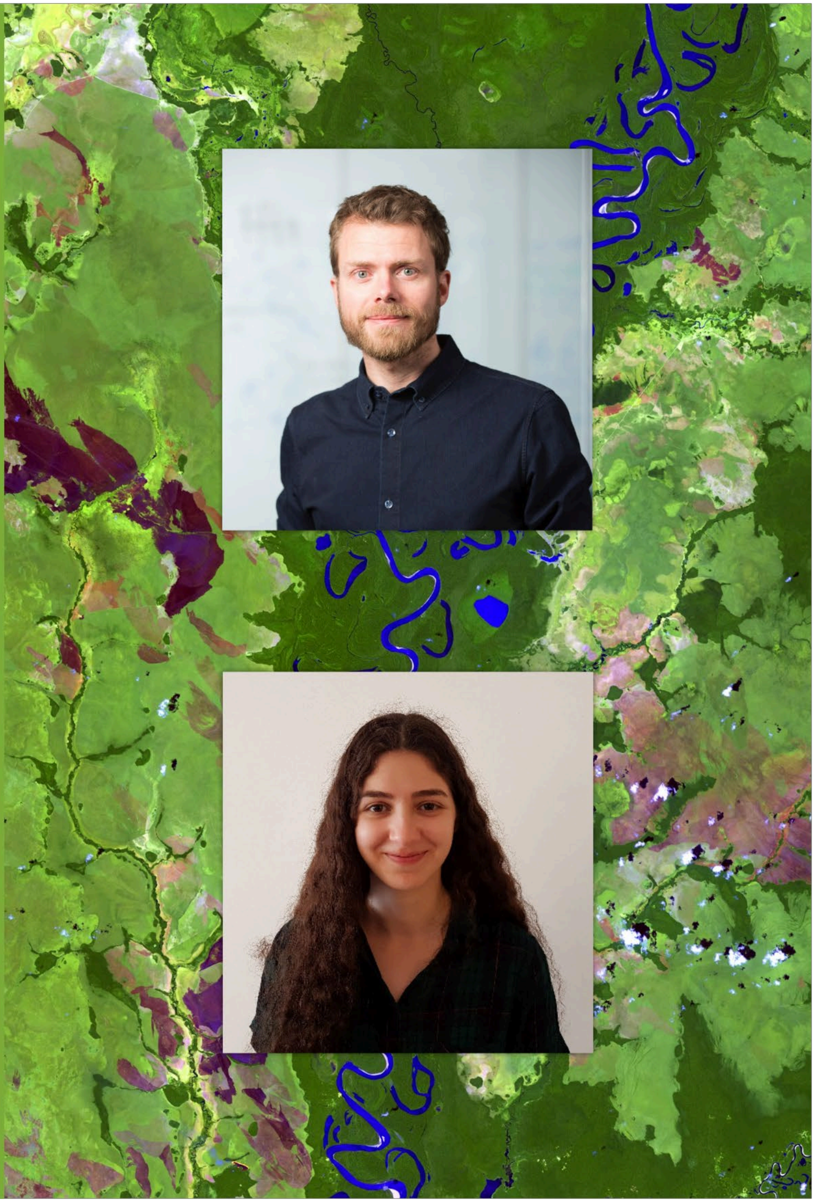
Columbia University

Rotem Mayo

Google

More speakers to be confirmed!

ECMWF



DestinE Earth System Components

Abstract submission deadline 26th June