Modelling the Earth system

Sara Hahner and Nina Raoult

Machine Learning Scientists, ECMWF

on behalf of many colleagues across ECMWF



Outline



WHY?



HOW



EXAMPLES

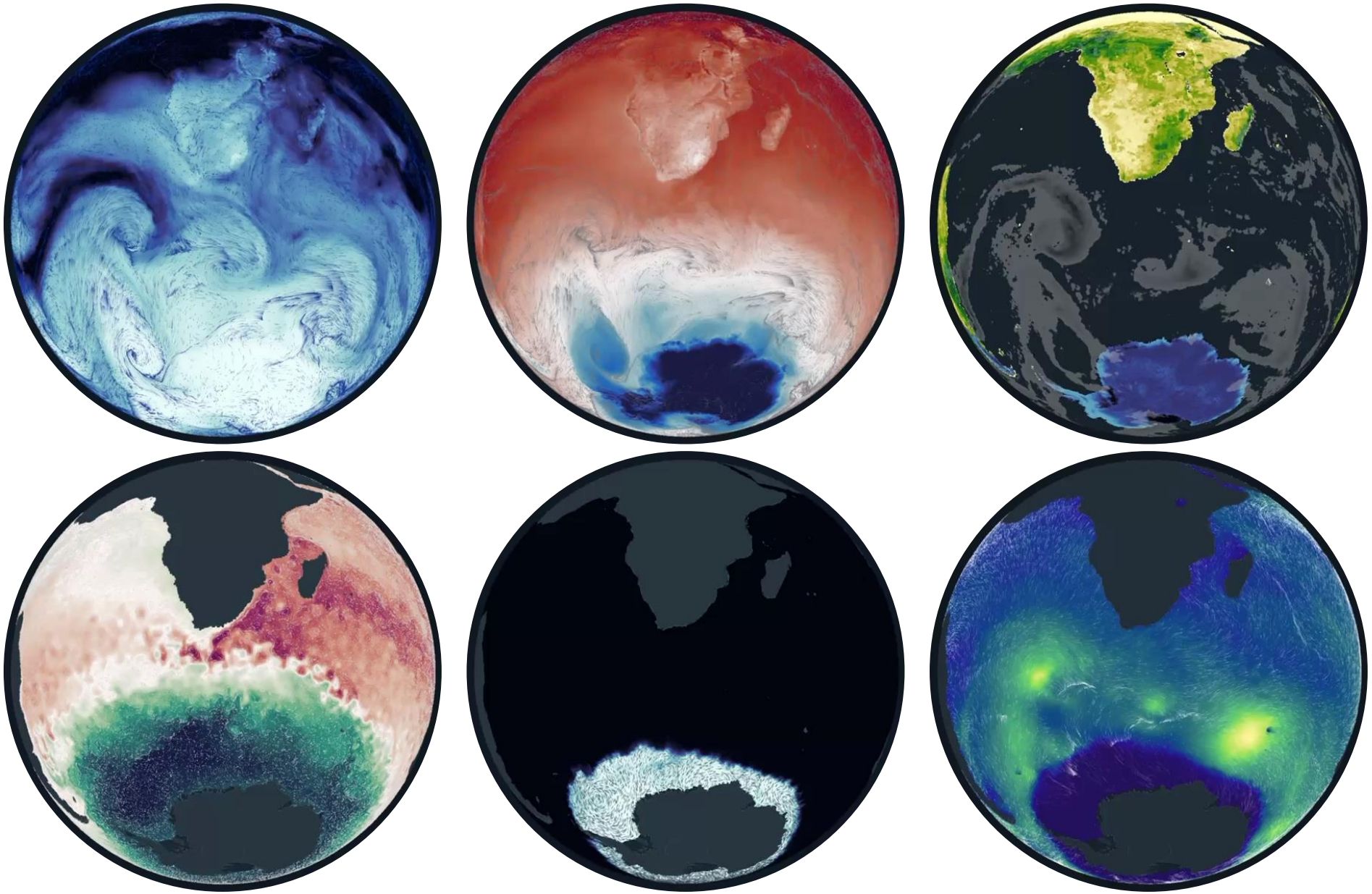


LESSONS LEARNED

Why an Earth System Model

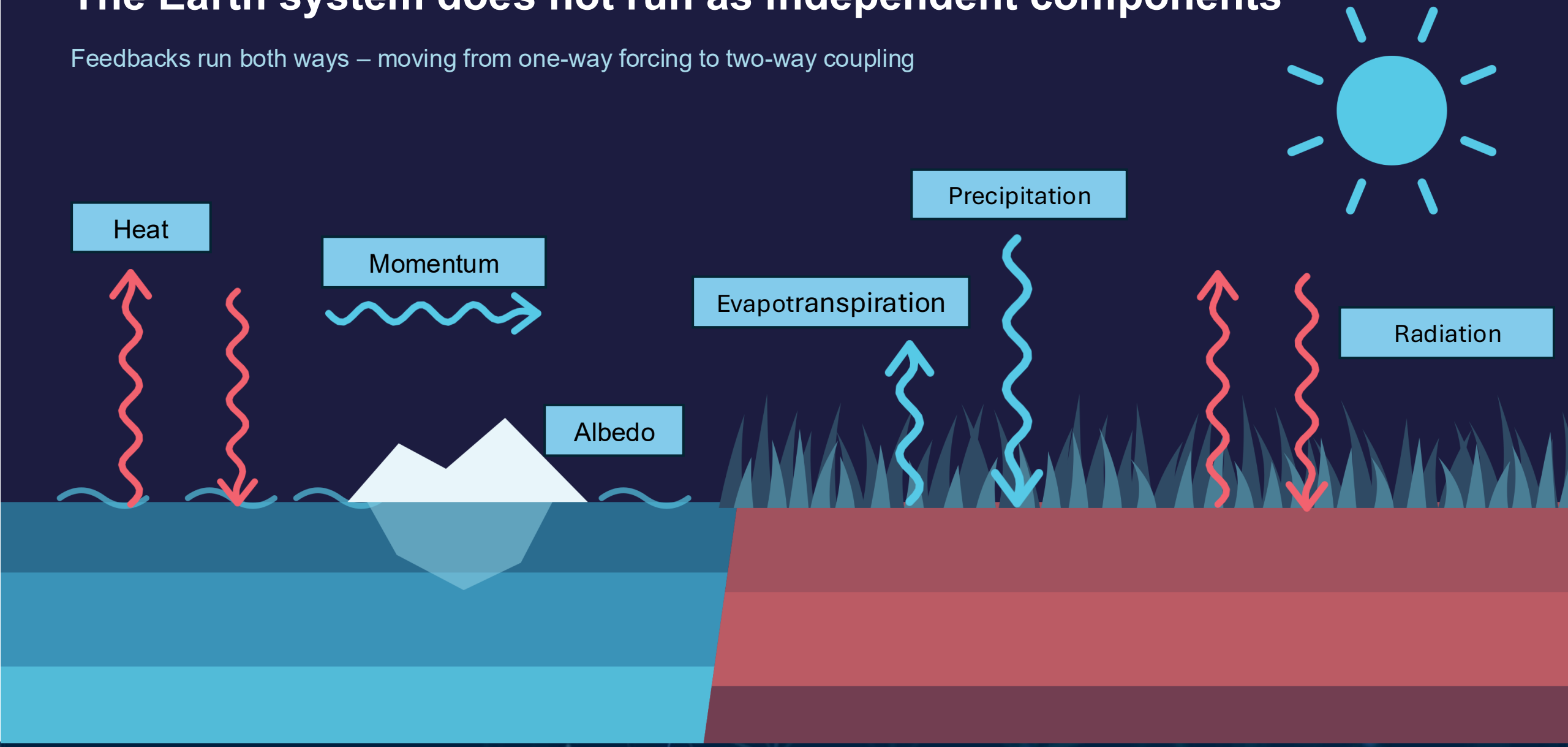
From atmosphere-only forecasts to coupled components



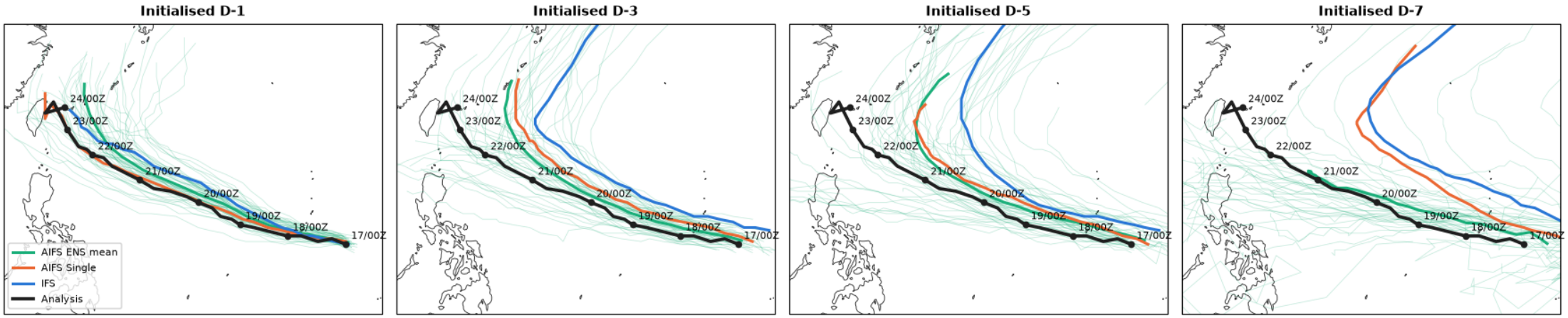


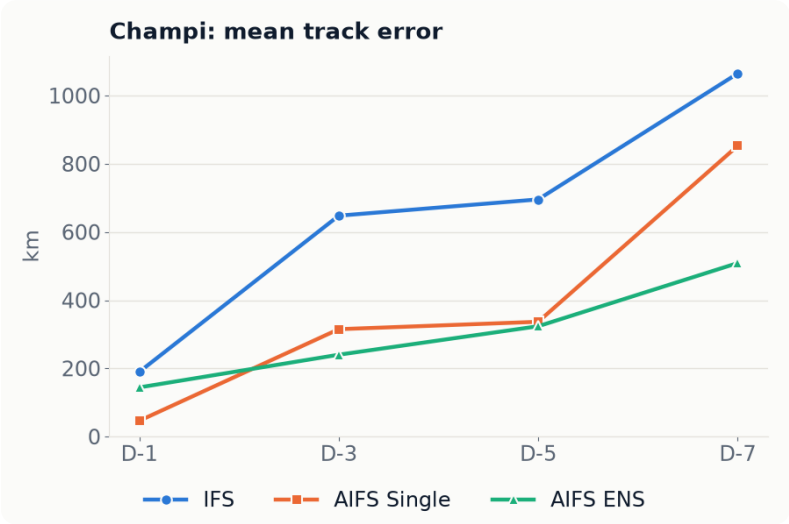
The Earth system does not run as independent components

Feedbacks run both ways – moving from one-way forcing to two-way coupling



Tropical Cyclone Champi: 17th-24th October 2026

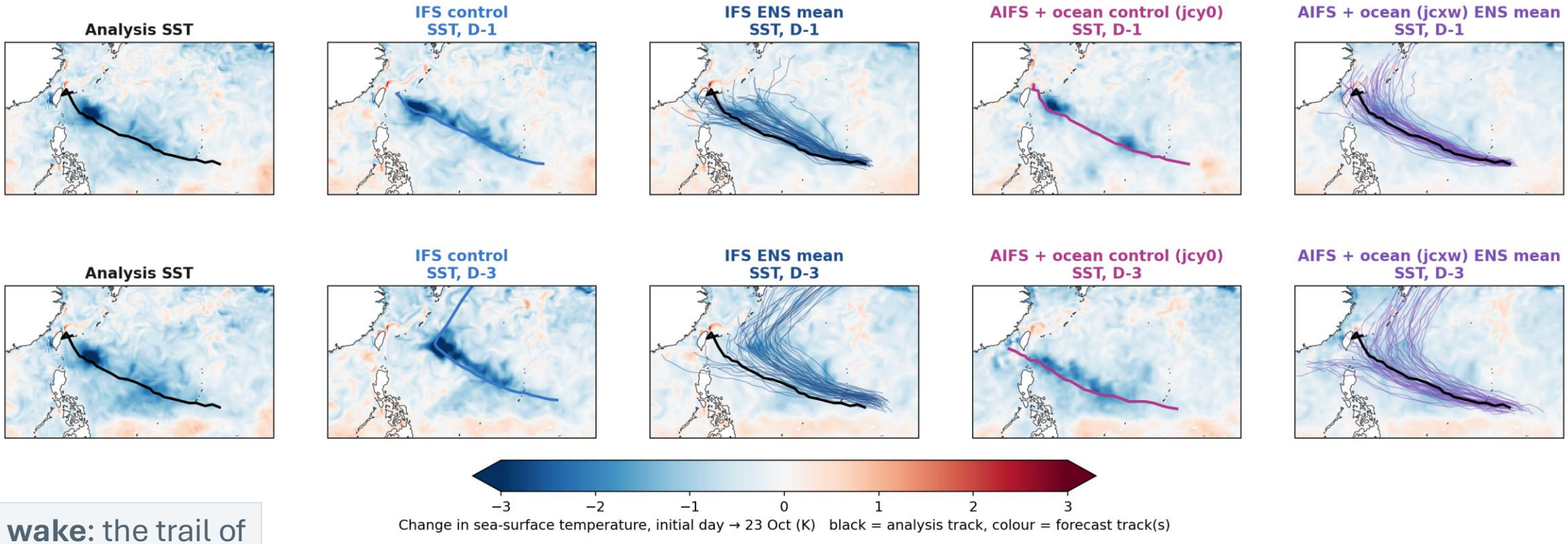




Model	Day 1	Day 7
Analysis	944 hPa · 45 m/s	944 hPa · 45 m/s
IFS	889 hPa · 64 m/s	895 hPa · 62 m/s
AIFS Single	947 hPa · 32 m/s	960 hPa · 29 m/s
AIFS ENS	957 hPa · 29 m/s	980 hPa · 22 m/s

IFS forecasts a far stronger storm than AIFS whose winds stay near 30 m/s

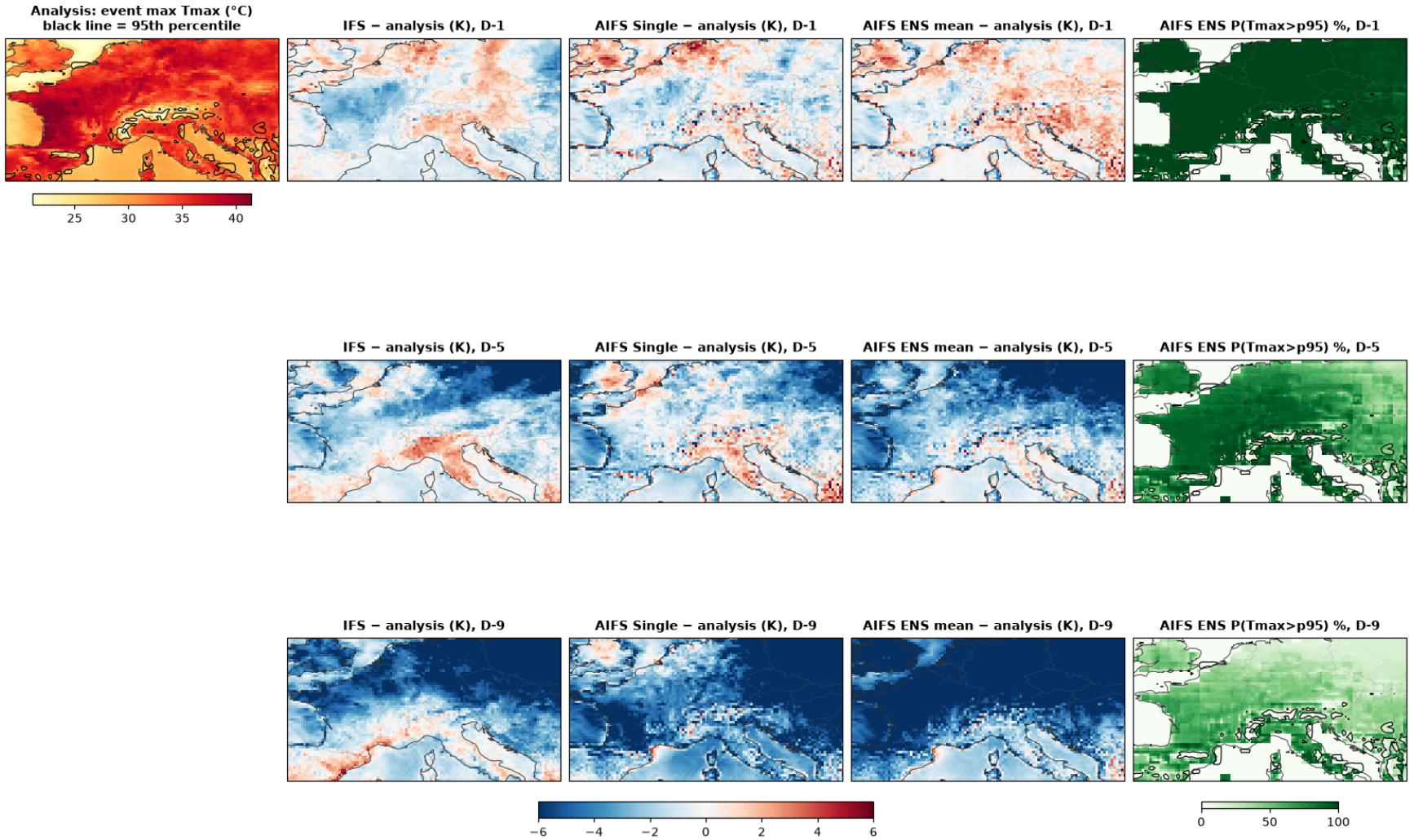
Tropical Cyclone Champi: 17th-24th October 2026 – cold wake



Cold wake: the trail of cooler sea a tropical cyclone leaves behind as its winds stir cold deep water to the surface.

With an ocean, AIFS cools the sea by up to 2 K behind Champi, close to the analysis and IFS.
Operational AIFS v2 has no ocean, so no cold wake.

Heatwave in Europe: 23rd-28th June 2026

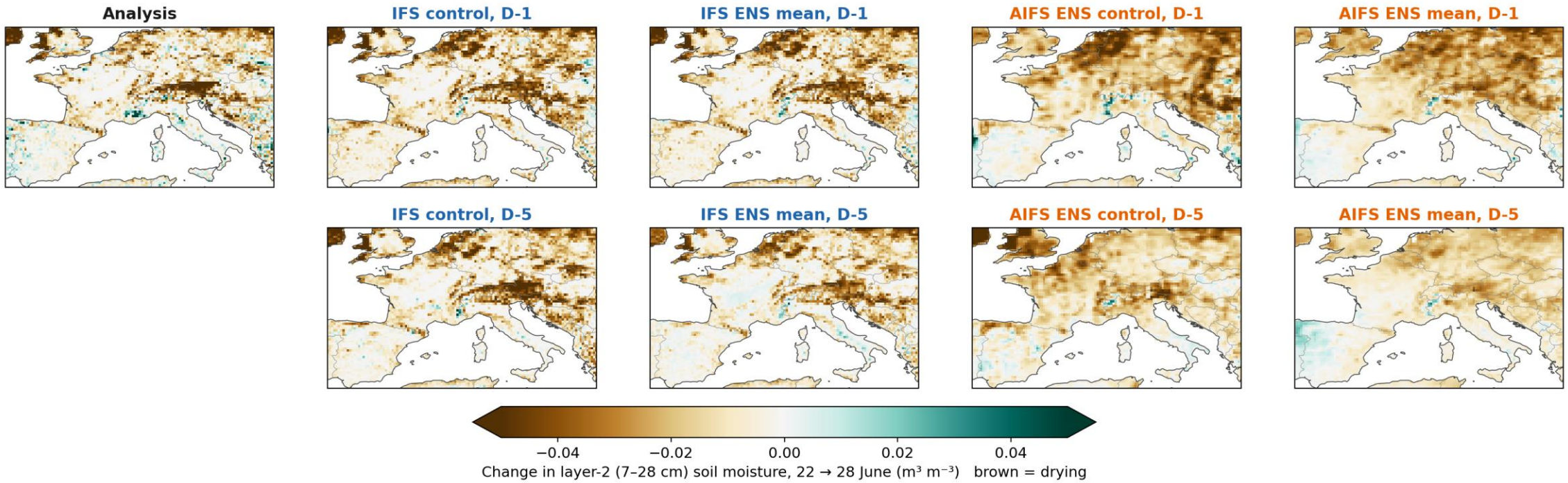


Day 1: all three are within about 2 K of the analysis, and AIFS ENS puts high probability above the 95th percentile almost everywhere.

Day 5: every model is cold over the north and east. RMSE is 5.3 K for IFS and 3.2–3.3 K for the AIFS models.

Day 9: errors of 4–6 K too cold over most of the area, and the AIFS ENS probabilities fade.

Heatwave in Europe: 23rd-28th June 2026



Representing the land lets us see the heatwave drying the soils.

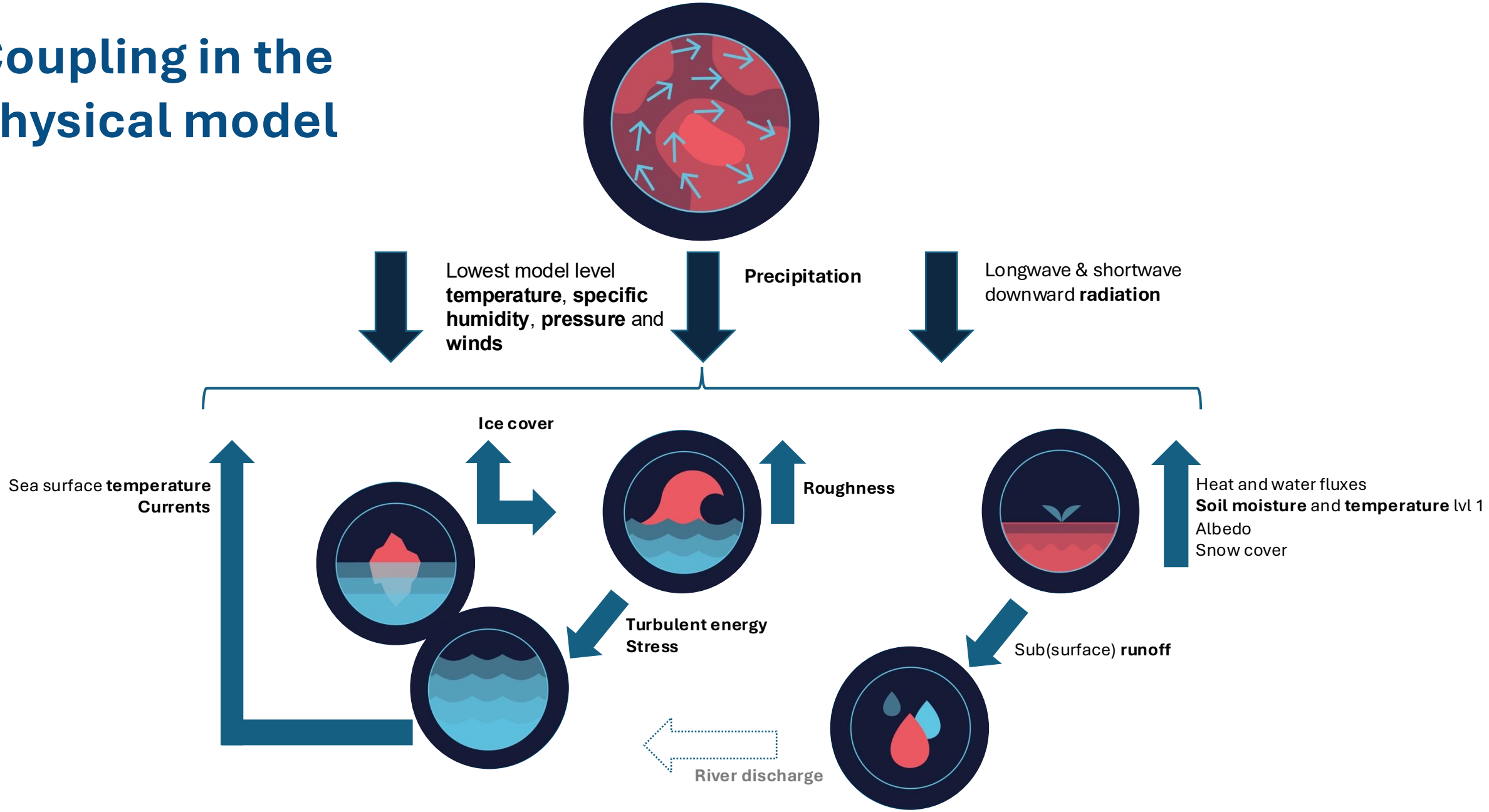
AIFS ENS captures the pattern, though it dries a little faster than IFS.

Multiple ways to build a coupled ML Earth system

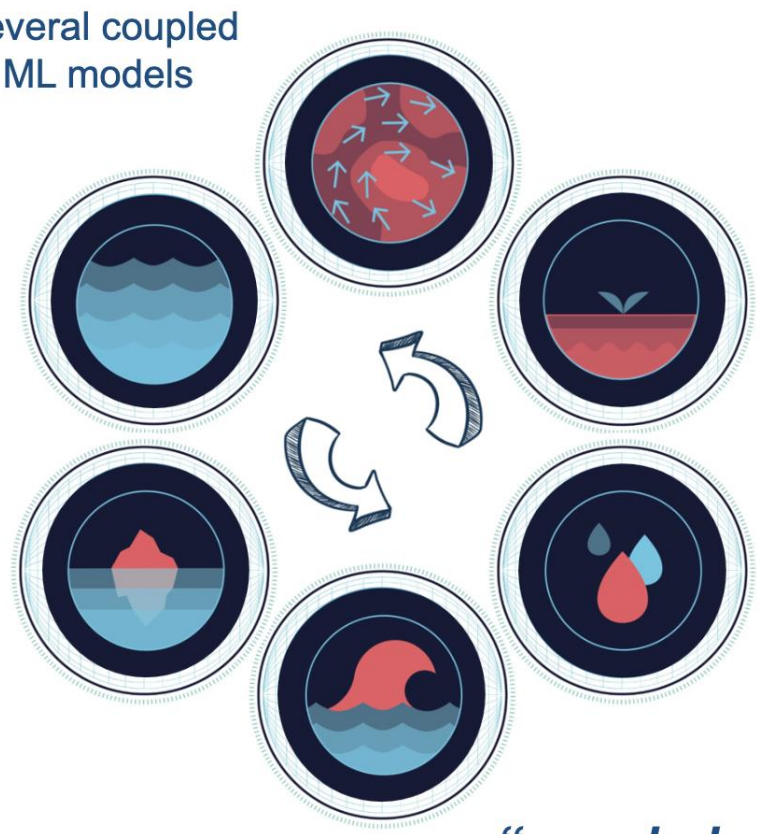
Examples from AIFS and Destination Earth



Coupling in the physical model



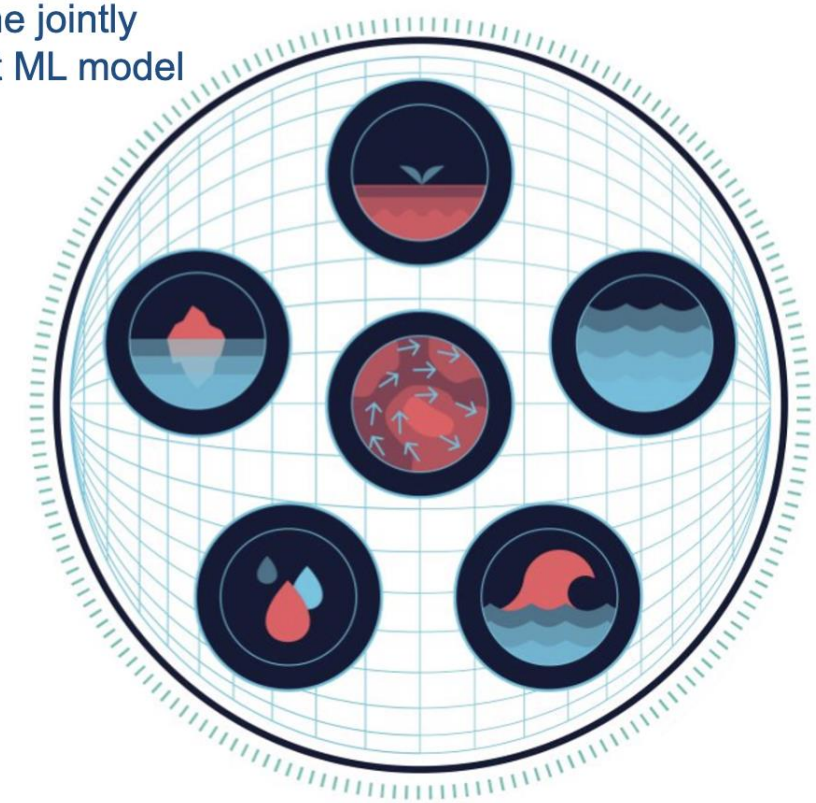
Two ways to build a coupled ML Earth System



“coupled model”

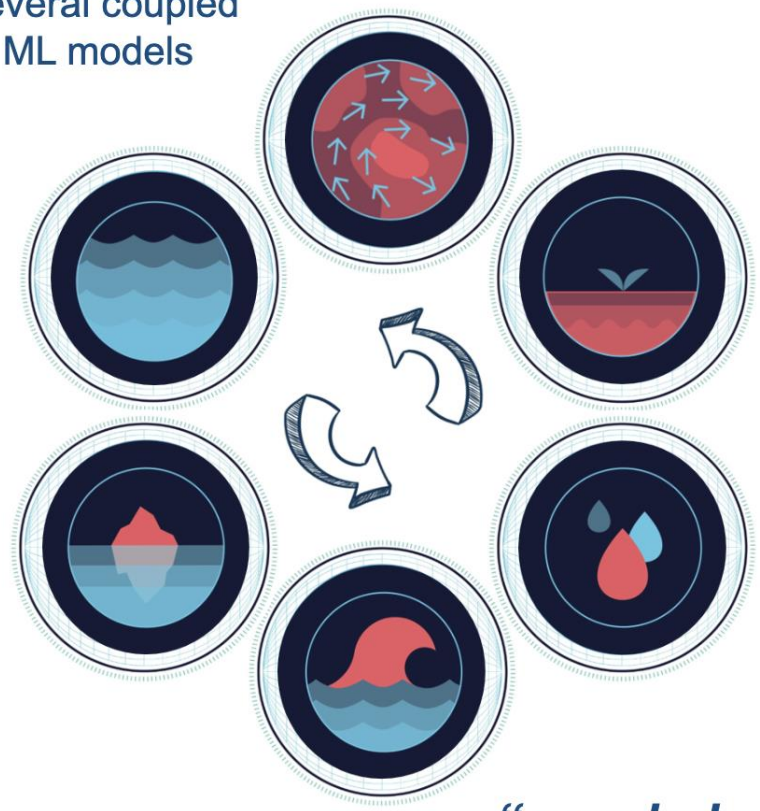
Two ways to build a coupled ML Earth System

One jointly learnt ML model



"joint model"

Several coupled ML models



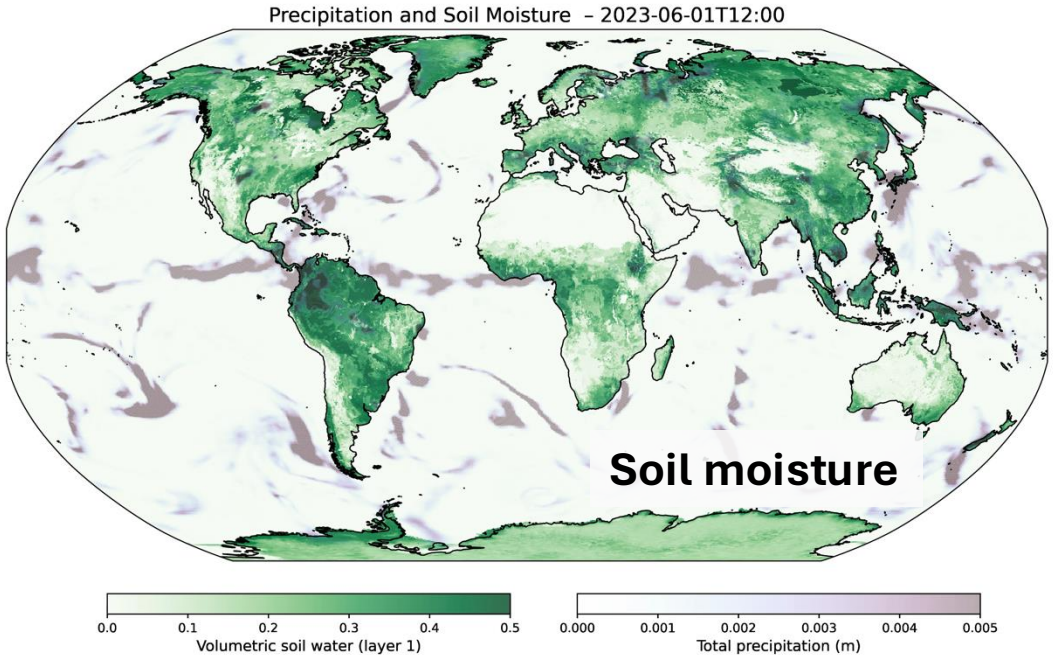
"coupled model"

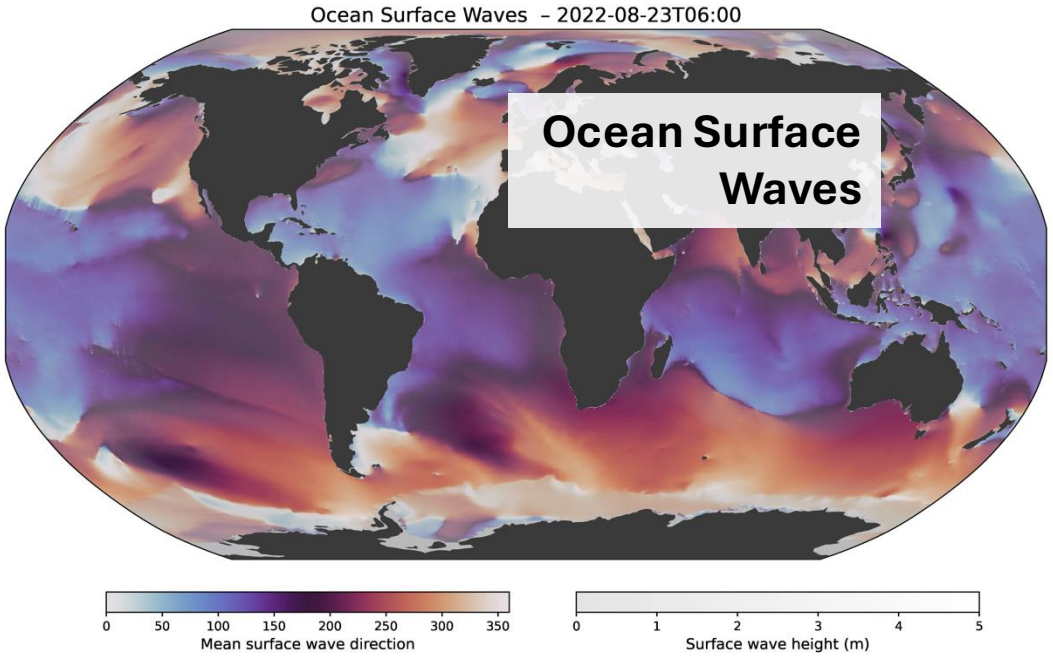
AIFS v2
ECMWF

Atmosphere, Land & Waves

Coming soon: surface ocean and sea ice

AIFS – an example of a joint system





AIFSv1:

- soil moisture (lvls 1-2)
- soil temperature (lvls 1-2)
- runoff

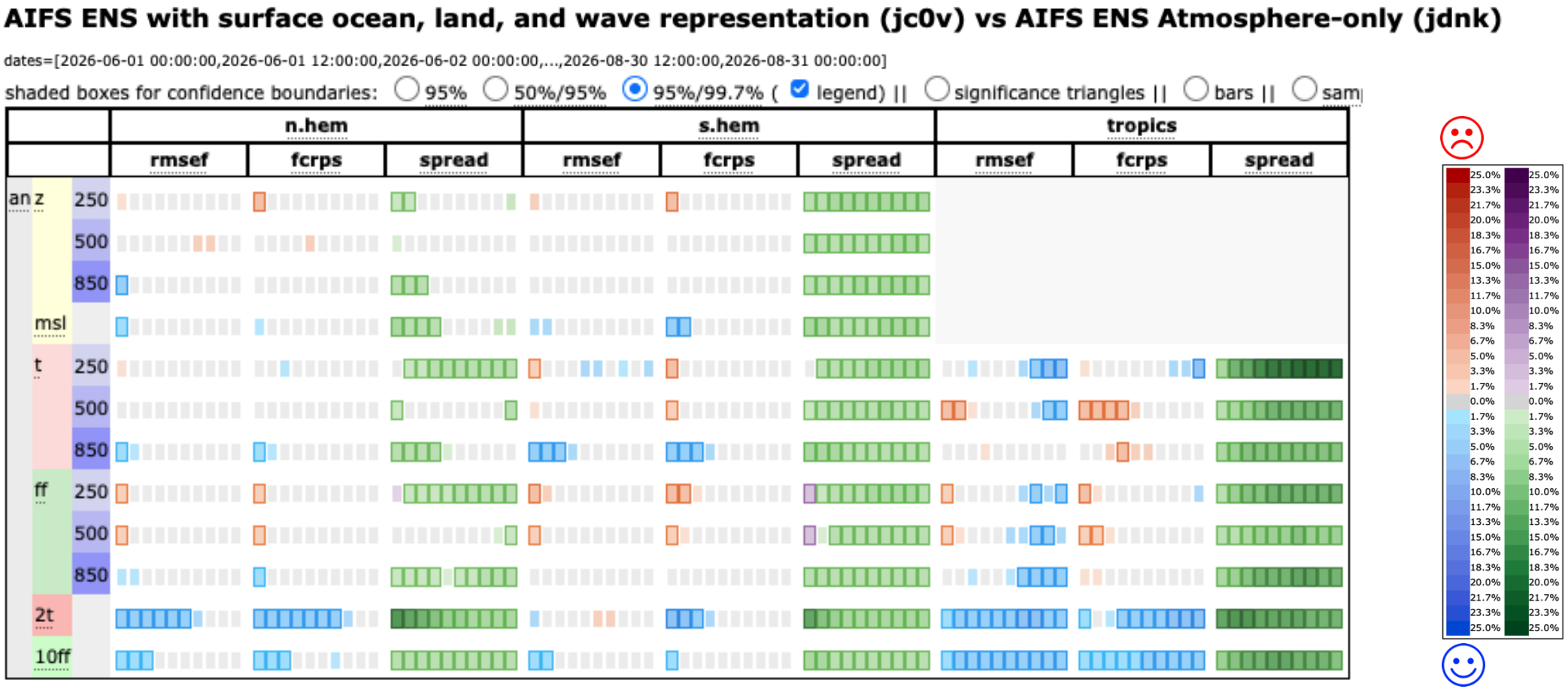
AIFSv2:

- ocean surface waves
- snow cover

Under consideration for AIFSv3:

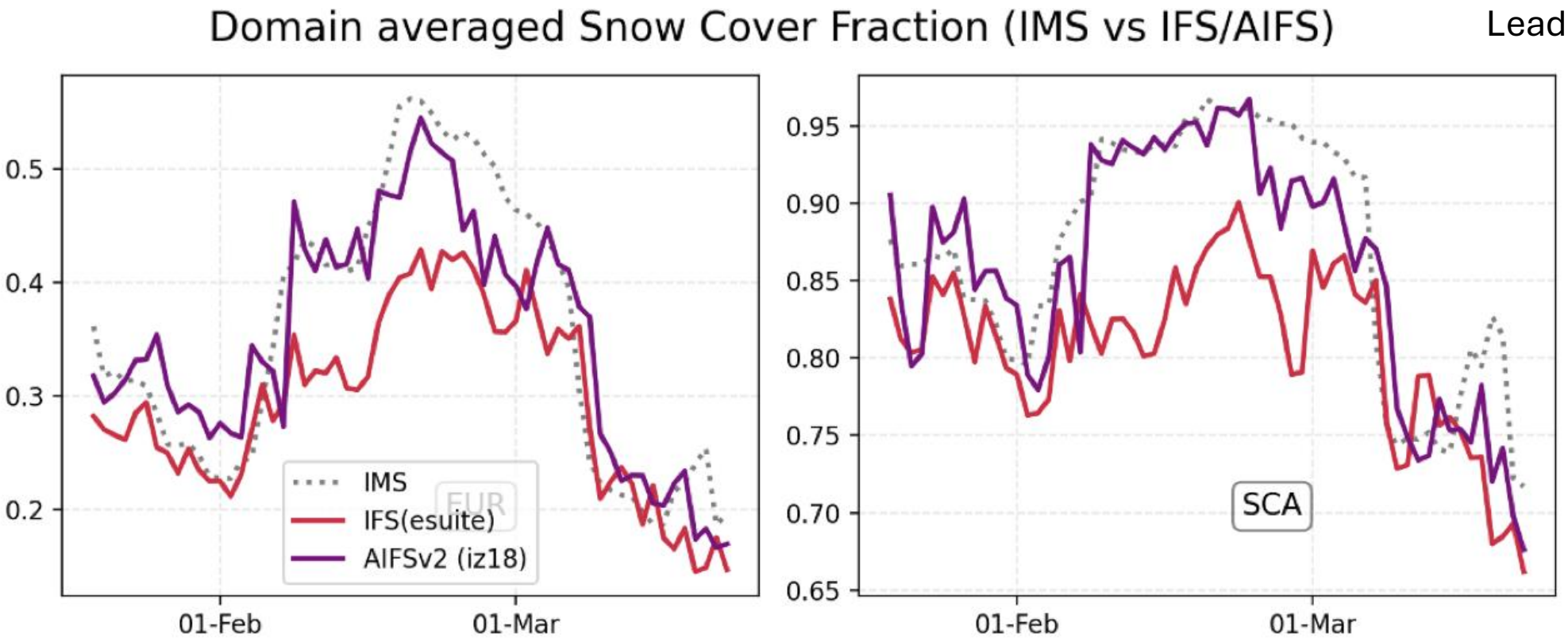
- urban & lake cover forcing
- runoff partitioning
- surface ocean & sea ice

Representing the earth system components improves the surface forecast scores in the AIFS



Land representation in AIFS-ENS

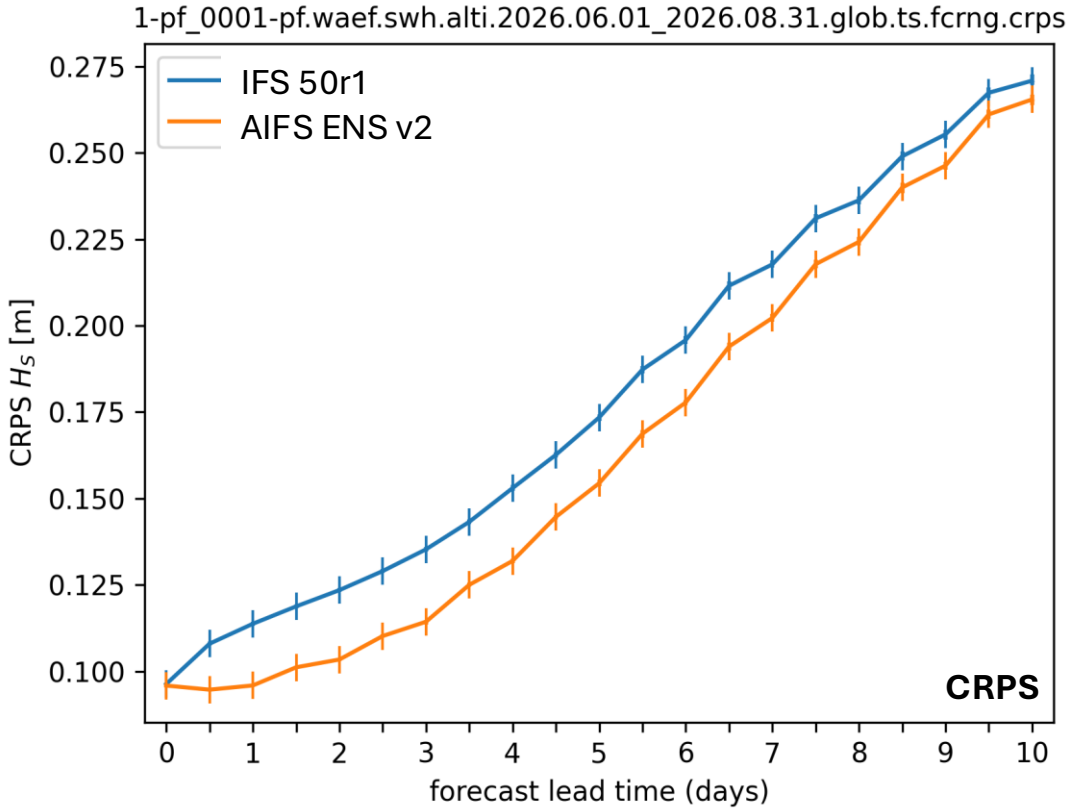
The joint model has learnt the correct land processes

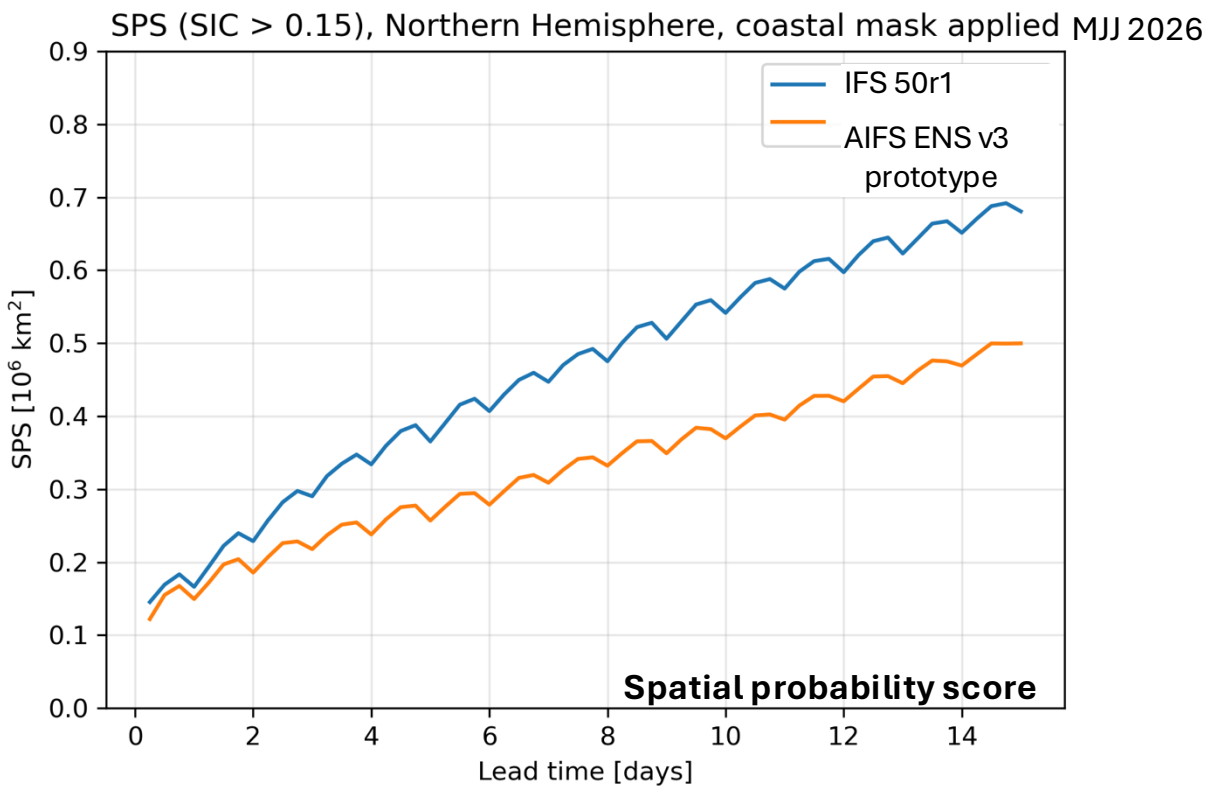


Verified against satellites observation: the AIFS tracks observation more closely that the IFS

Marine forecasting in AIFS-ENS

Improving forecasts of significant wave height and sea ice edge position against IFS

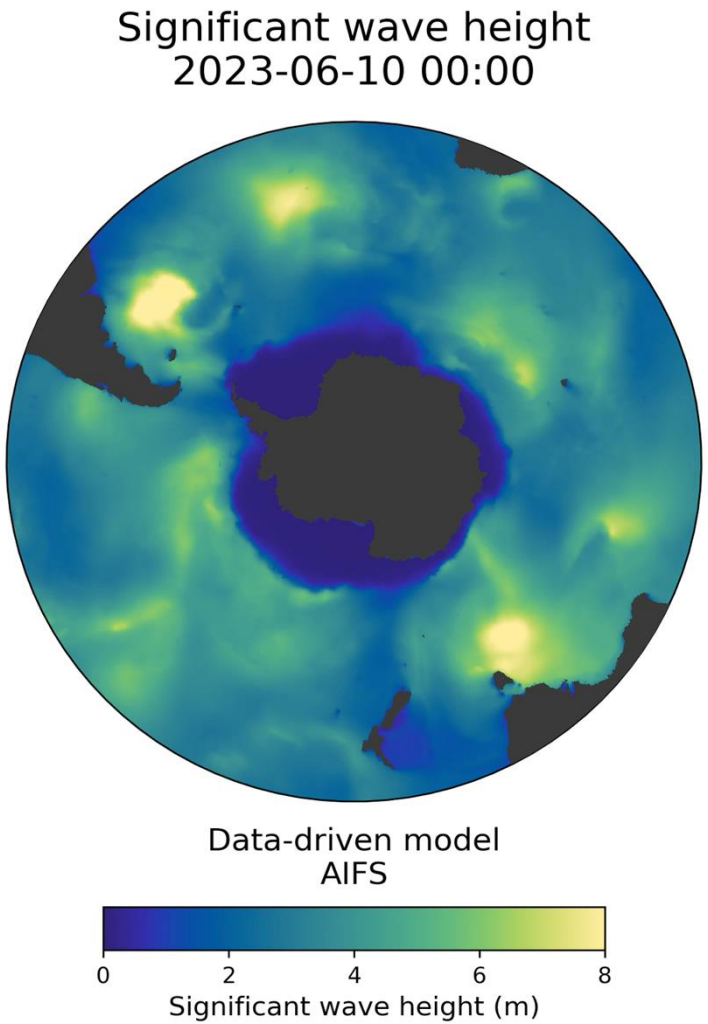




AIFS has about 10-20% smaller errors for SWH forecasts

AIFS has about 25% smaller errors for sea ice edge position forecasts

AIFS v2: implicit information about sea ice

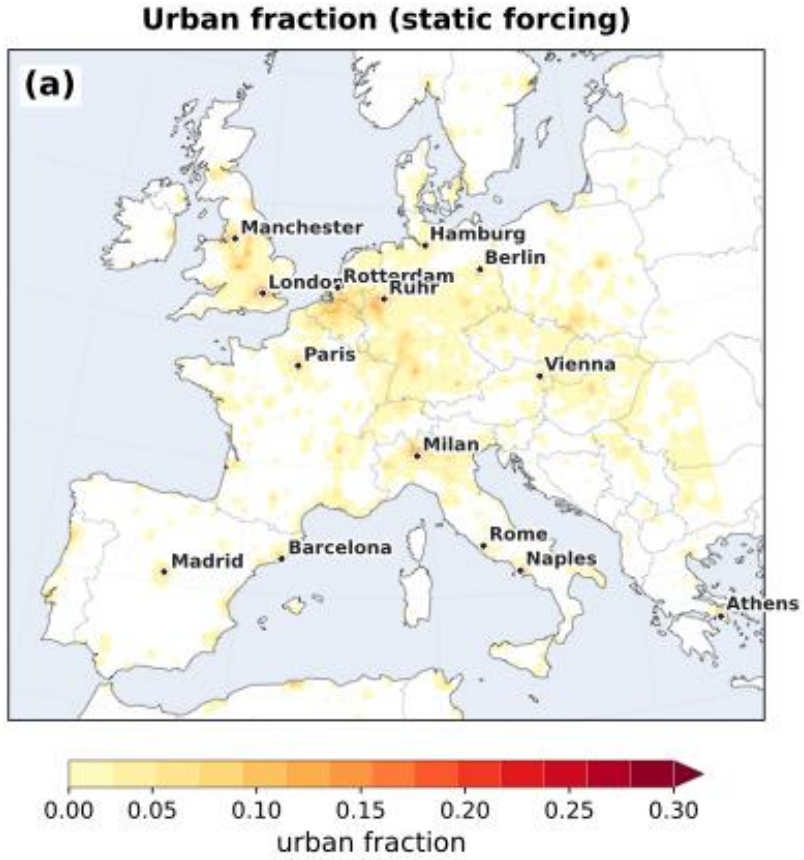


Implicit sea ice representation through waves:

- Under ice, the significant wave height is dramatically reduced
- The ML model learns an implicit representation of sea-ice – the region of reduced SWH realistically varies over time with the *implicit* ice-cover
- With an explicit sea ice representation (AIFS v3) the wave forecasts around sea ice improve

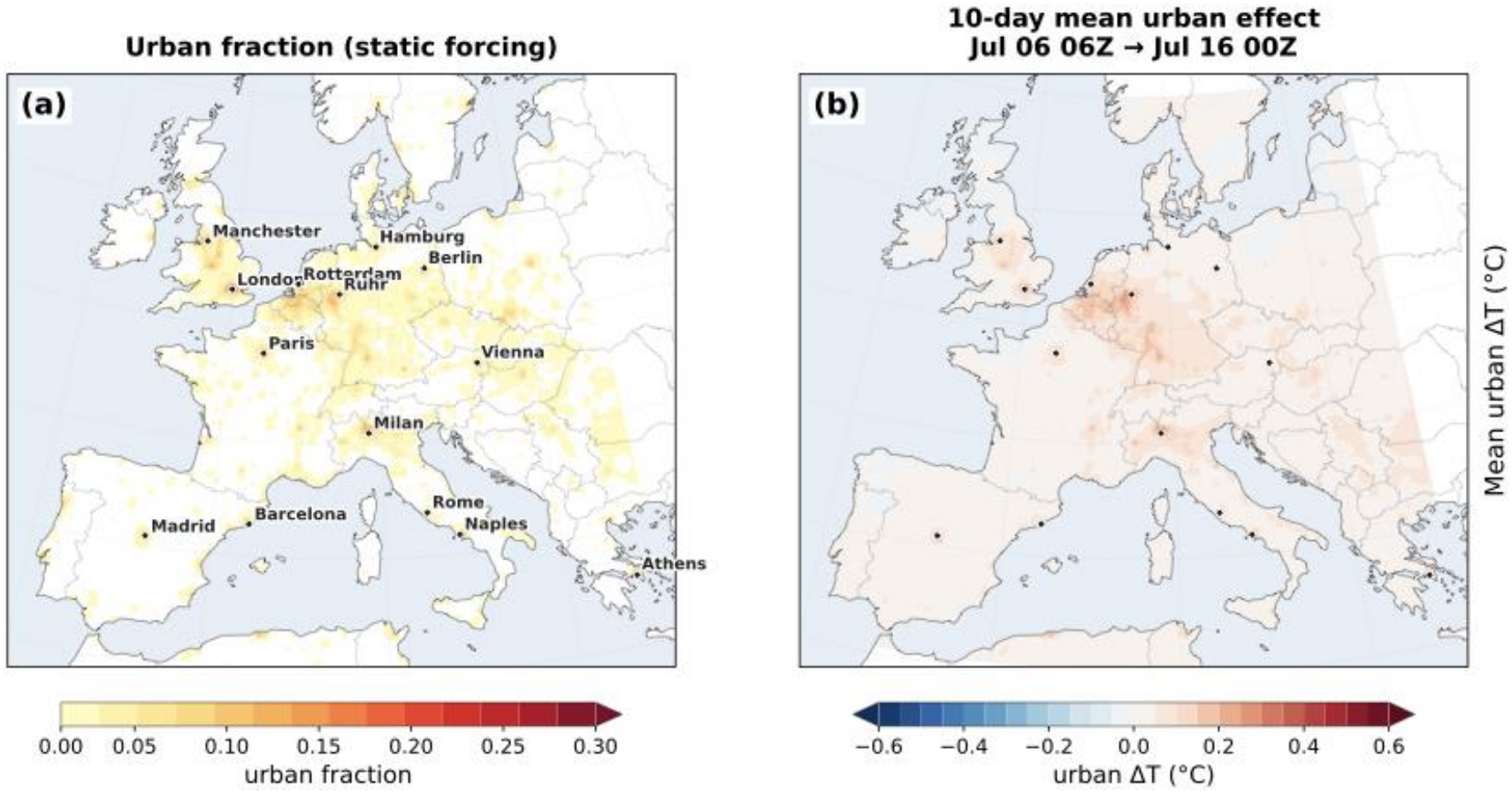
Probing physical consistency in land representation in AIFS-ENS

The joint model has learnt the correct land processes



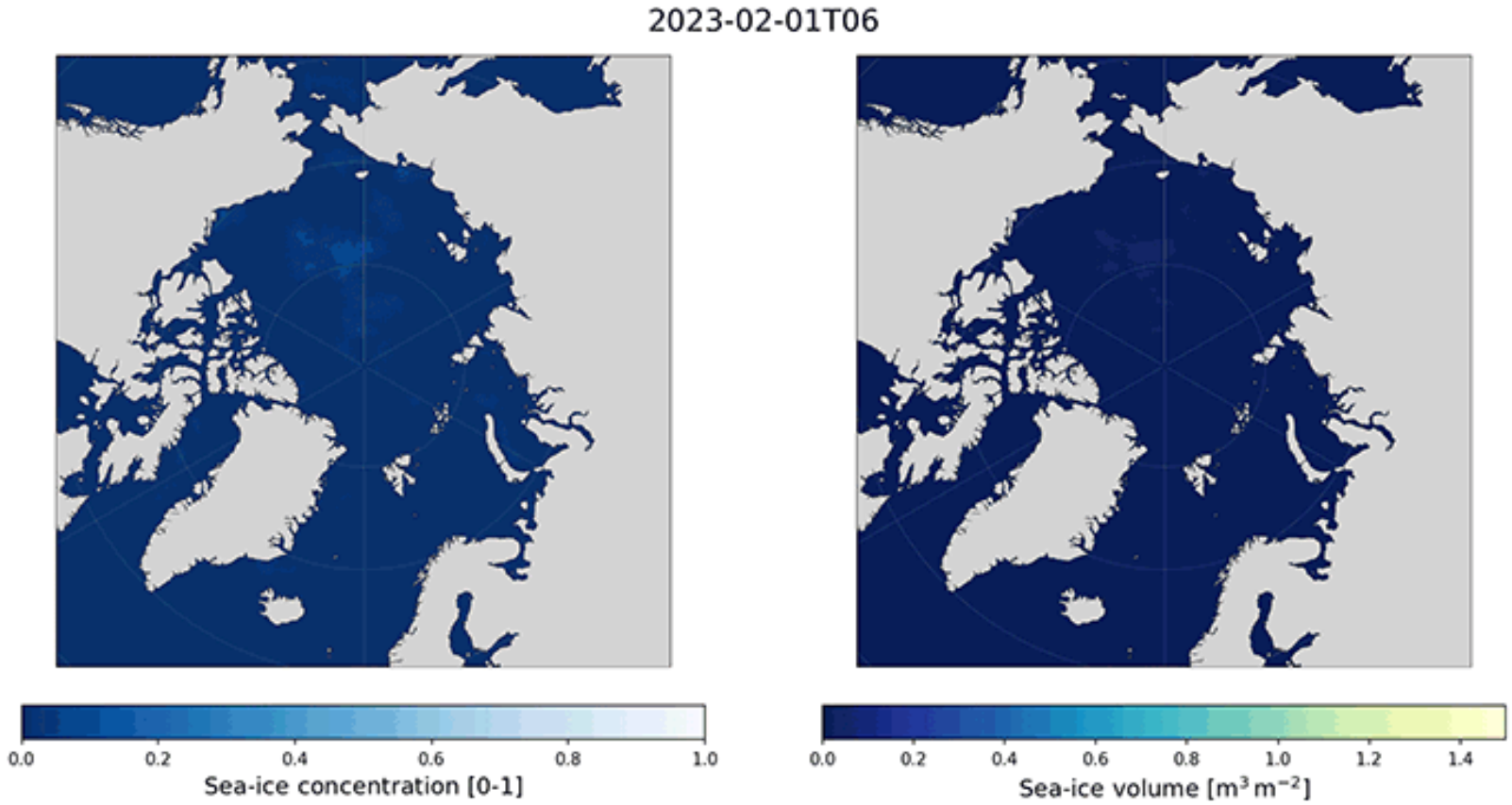
Probing physical consistency in land representation in AIFS-ENS

The joint model has learnt the correct land processes



Emergent from static inputs: the AIFS produces an urban heat island

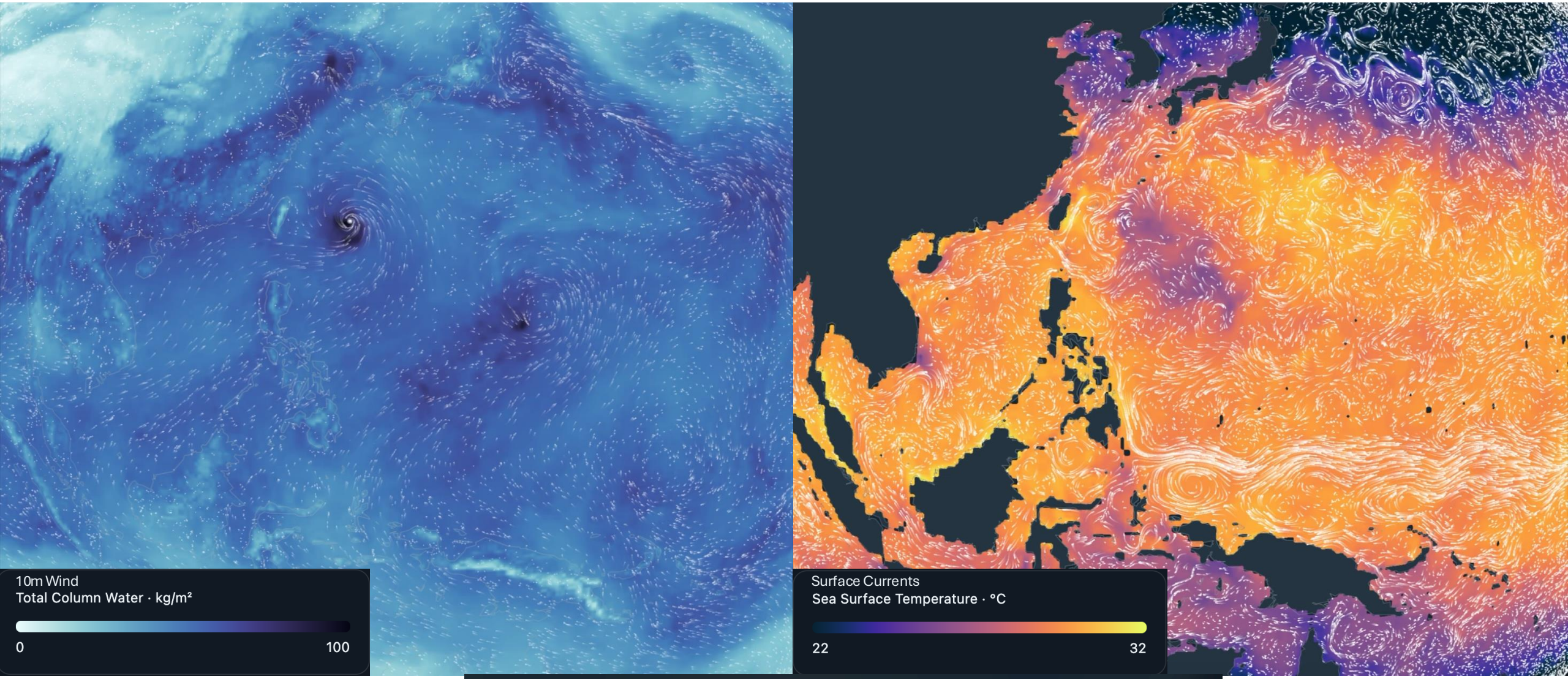
Probing physically consistent coupled behaviour



Arctic sea-ice recovery from 1 February 2023 after removing ice from the initial conditions.

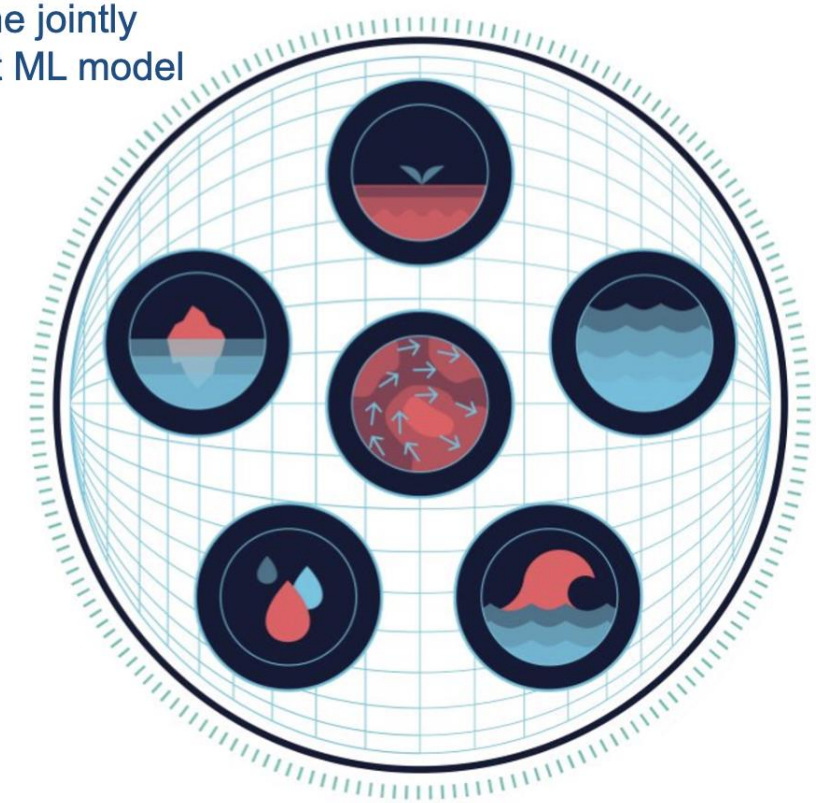
Regrowth follows the seasonal cycle, and ice volume (right) recovers more slowly than extent.

Cold Wakes under Tropical Cyclones – 05.07.2026 – Typhoon Bavi



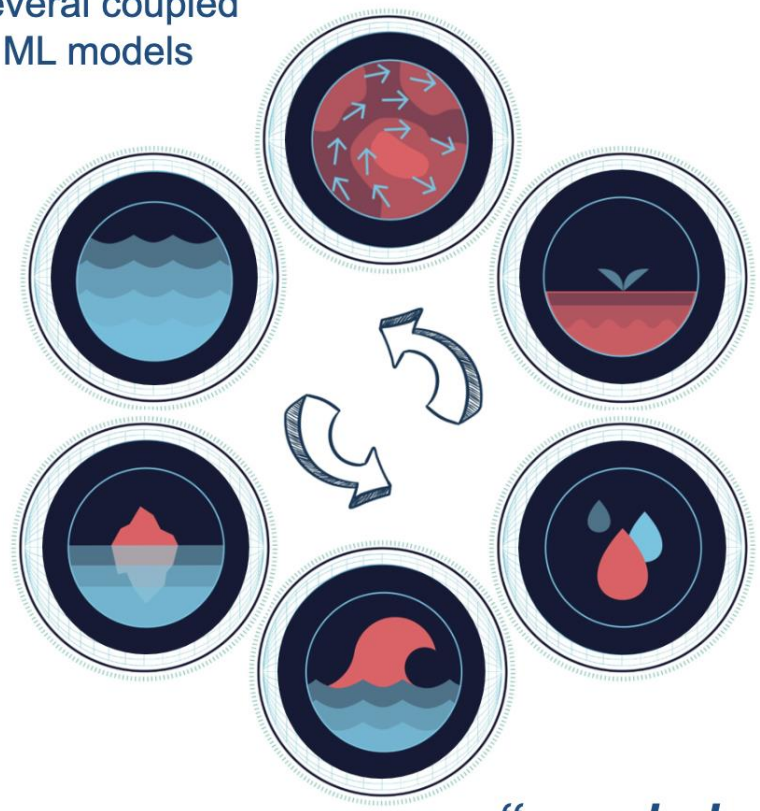
Two ways to build a coupled ML Earth System

One jointly learnt ML model



"joint model"

Several coupled ML models



"coupled model"

AIFS v2
ECMWF

Atmosphere, Land & Waves

Coming soon: surface ocean and sea ice

Two ways to build a coupled ML Earth System

Single *jointly*-trained component-agnostic network

- **No coupling mechanism to design:** no exchanged fields, coupling frequency, time-stepping or grid matching
- **No inherited assumptions:** interactions are learned in a shared latent space, not limited to variables the modeller chose
- **No train–inference gap:** the model sees the same fields in training as at run time
- **One model** to train, run and deploy

But: *one architecture, resolution, and loss must serve every domain, and any change means retraining the whole system.*

Separate components, *coupled* at inference

- **Domain-specific design:** each domain gets the architecture, grid and timestep its dynamics need
- **Focused capacity and loss:** small components are not overshadowed by larger ones
- **Domain expertise:** decades of understanding built into the components and the coupling, keeping models within known physics rather than spurious correlations
- **Best available components:** each trained on the best data for its domain, and swapped or upgraded independently

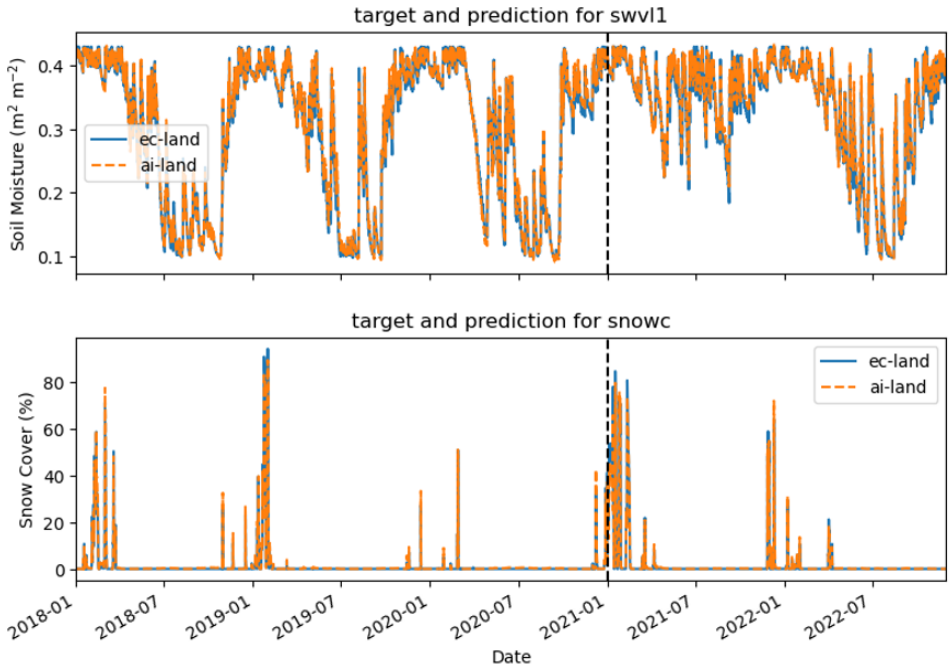
But: *each component is trained under near-perfect forcing and then run on its partner's errors*

These two approaches live on a spectrum – *the answer is likely different for each use case, and sometimes a combination*

The stand-alone land component

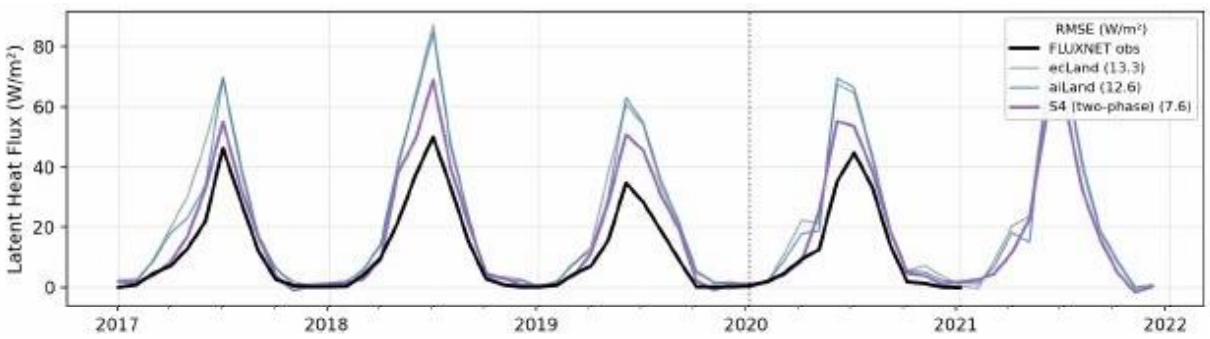
Emulates the physical model

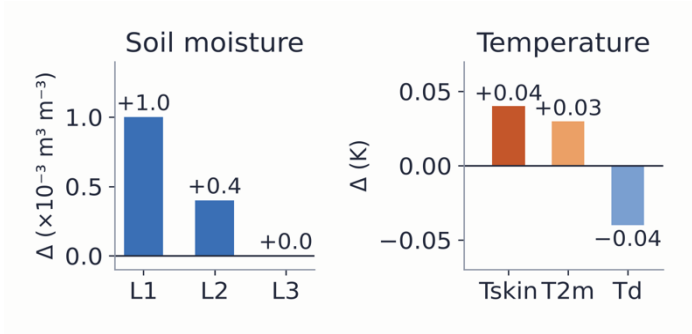
A simple MLP (ai-land) reproduces ecLand's outputs, including beyond the training period.



Improved by fine-tuning on observations

By fine-tuning on observations, we are able to correct known biases in the physical model



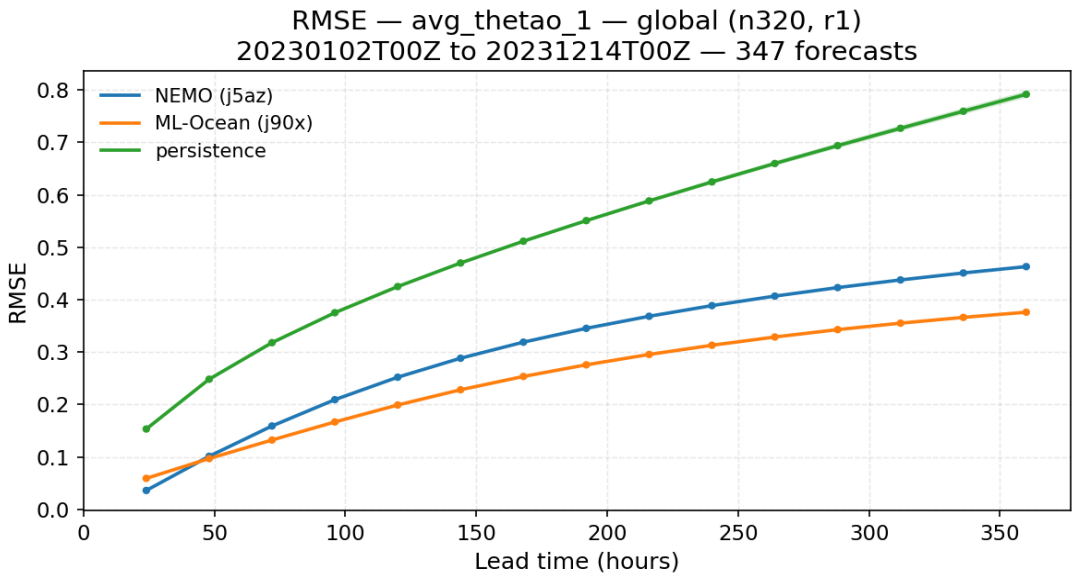


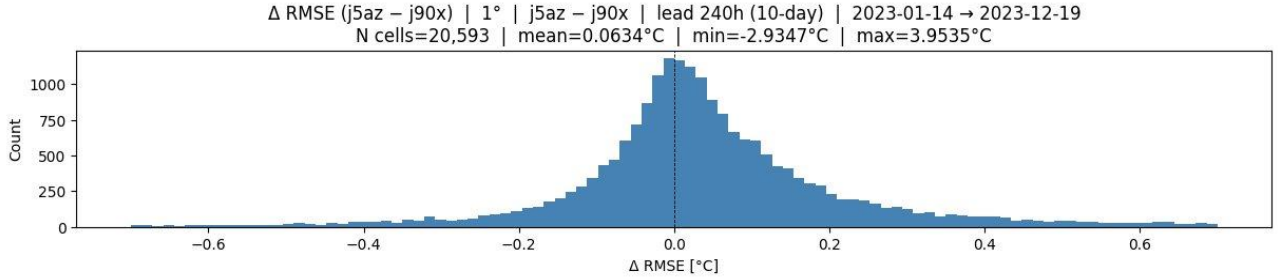
Knock-on effects are physical: lower latent heat → wetter soil → less evaporative cooling.

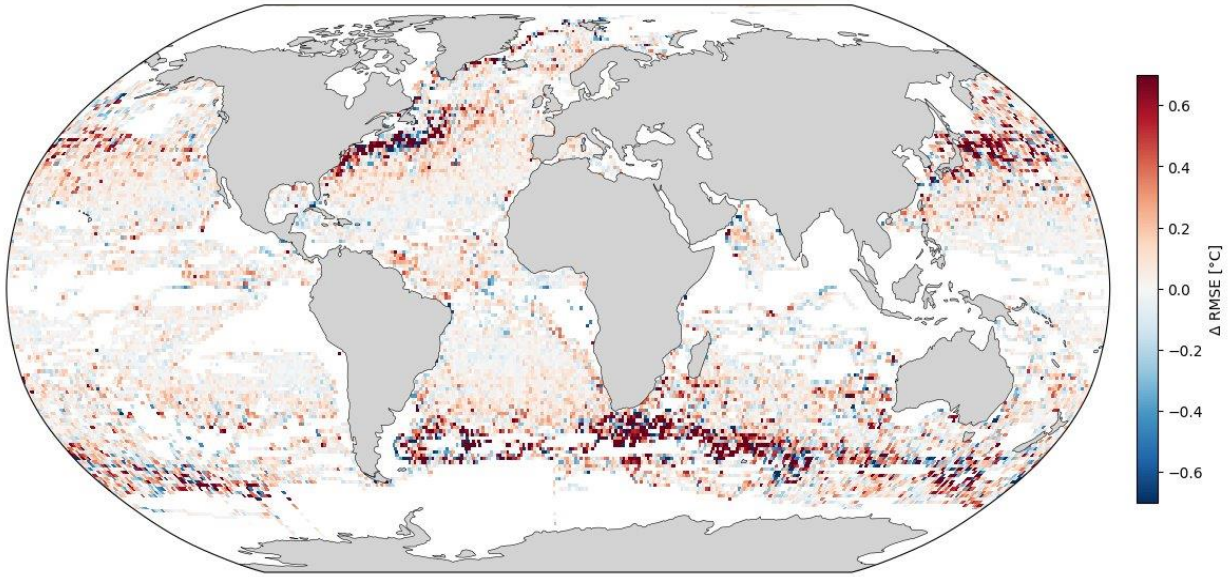
Pretaining on physics & fine-tuning on observations is a promising route when reanalysis products are not available and observations are sparse

The stand-alone ocean component

Learns from the ORAS 6 ocean reanalysis to model the global 3d ocean.







Assessment against both the reanalysis and observation data, show the ML gives improved RMSE compared to phyics based models, with the skill gain increasing with lead time.

Building a coupled system under DestinE: the stand-alone components



Sea ice: concentration, volume, albedo and drift in both hemispheres, with about 1.5 days more ice-edge skill than the numerical system



Hydrology: an LSTM predicting daily river streamflow at some 18,000 gauges worldwide, including catchments it has never seen



Ocean: temperature, salinity and currents through the full depth, along with sea surface height. Shows improvements over physics-based model predictions for almost all variables.



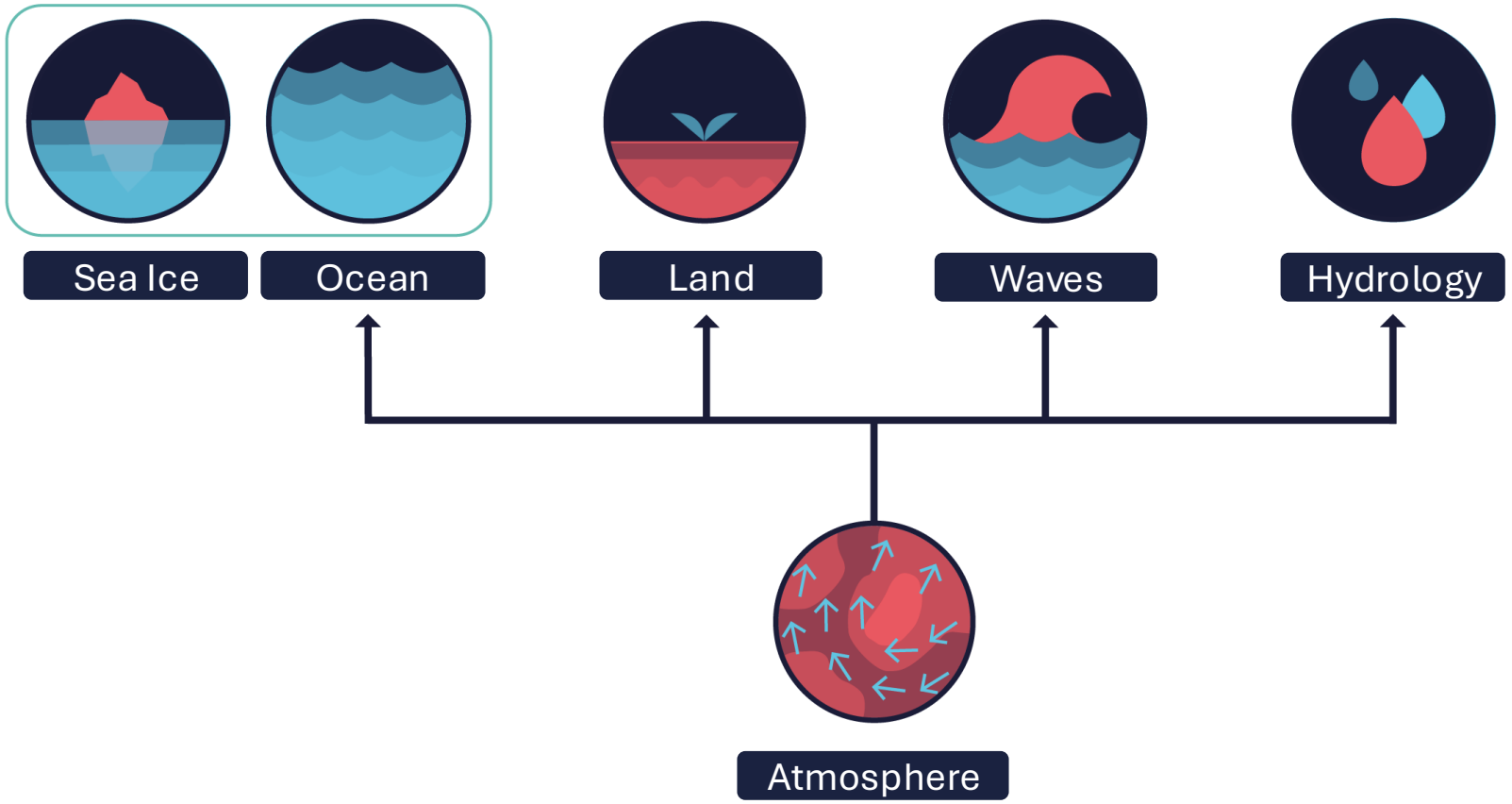
Land: a single MLP mimicking the column structure of land-surface models gives a stable multi-year emulator we can fine-tune on observations



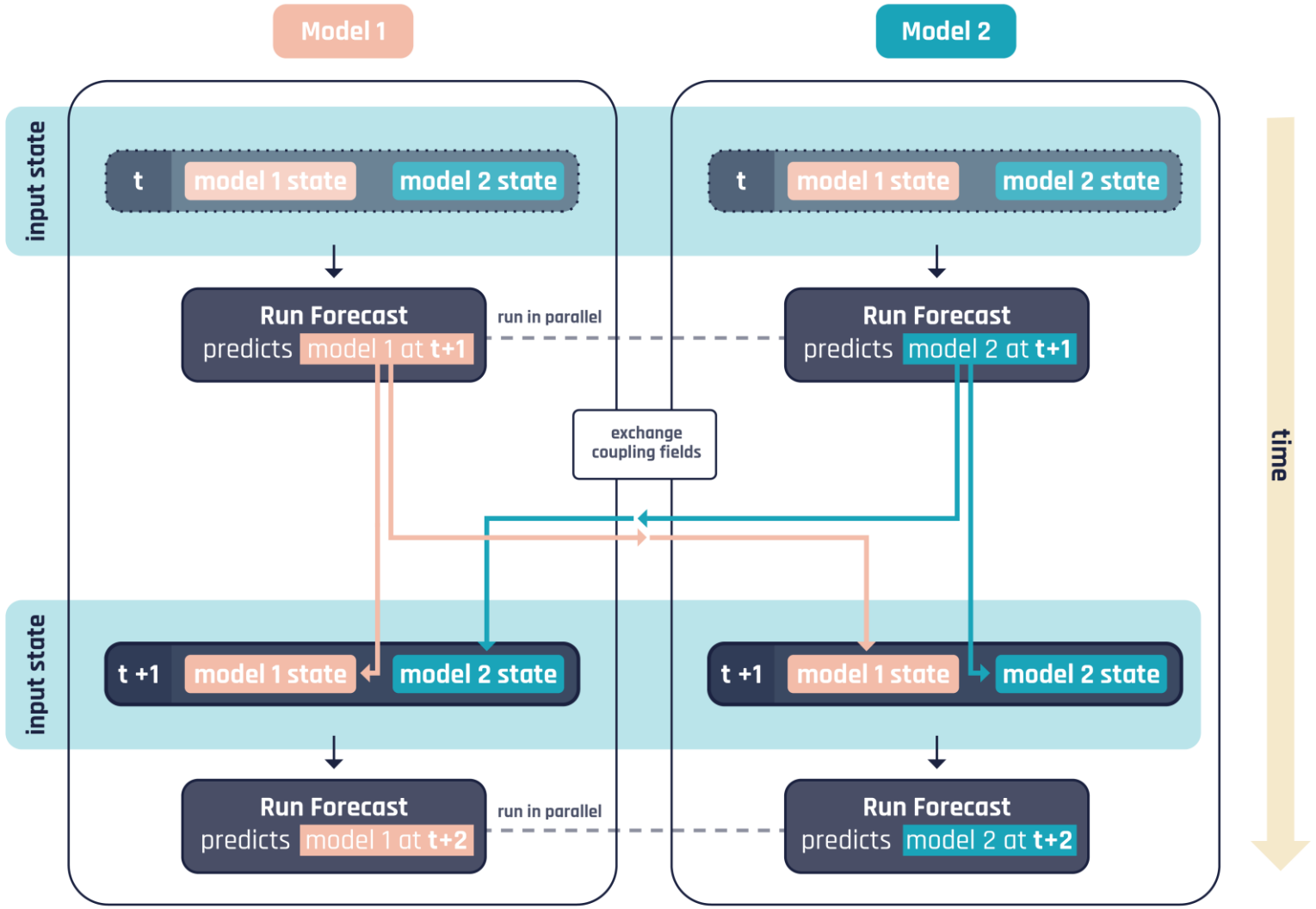
Waves: wave height, period and direction, separating wind-sea from swell; roughly one forecast day of skill over the physical model

More on the components:
See Destination Earth blogs on data-driven Earth system modelling

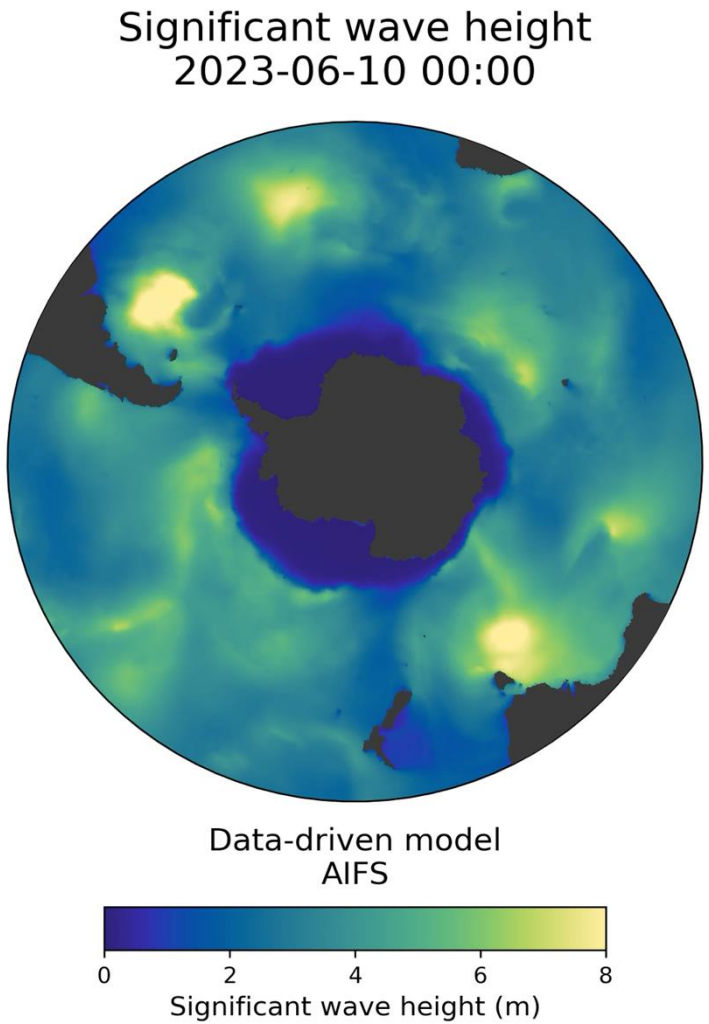
Building AI coupled Earth System Components



Passing information between components in coupled approaches



New mindset: implicit coupling coming from the data



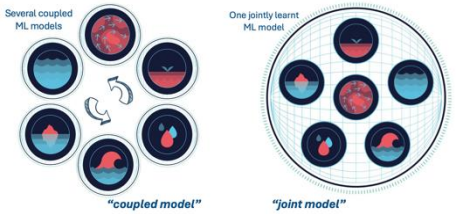
Even if models may not be coupled, often the datasets are.

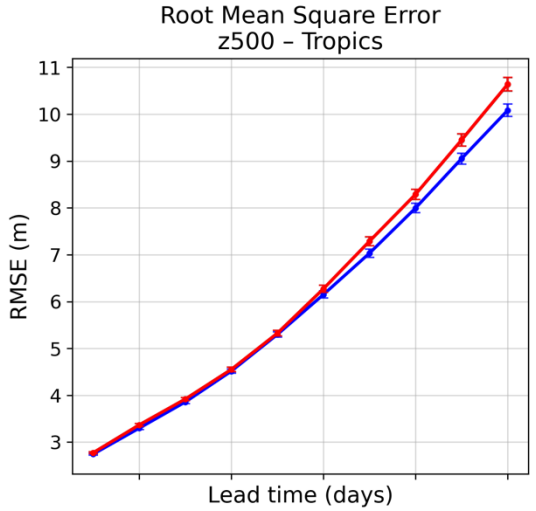
Example of sea ice representation through waves:

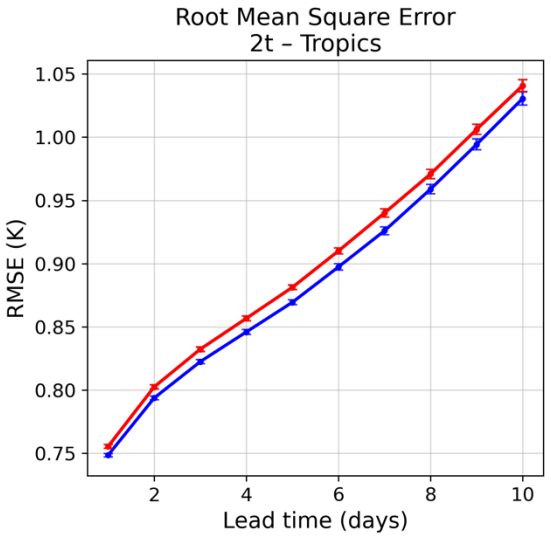
- Under ice, the significant wave height is dramatically reduced
- The ML model learns an implicit representation of sea-ice – the region of reduced SWH realistically varies over time with the *implicit* ice-cover

We see this type *implicit representation* throughout Earth system models

Comparing **coupled** and **joint** model configurations







ATMOSPHERIC Perspective

Both approaches are viable and present overall similar behaviors.

2t representation benefits from a joint training, particularly in the tropics, where knowledge about the ocean circulation is more relevant.

Joint model shows slightly better scores and more active forecasts.

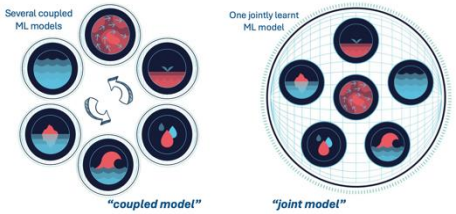
Similarity between coupled and forced (AMIP) atmospheric model suggests that the atmosphere uses its intrinsic ocean representation rather than the forcing information.

Atmosphere (786) + ocean (512), **coupled model** — 174 forecasts
Atmosphere (786) **forced model** — 174 forecasts
Atmosphere + ocean (1024), **joint model** — 174 forecasts

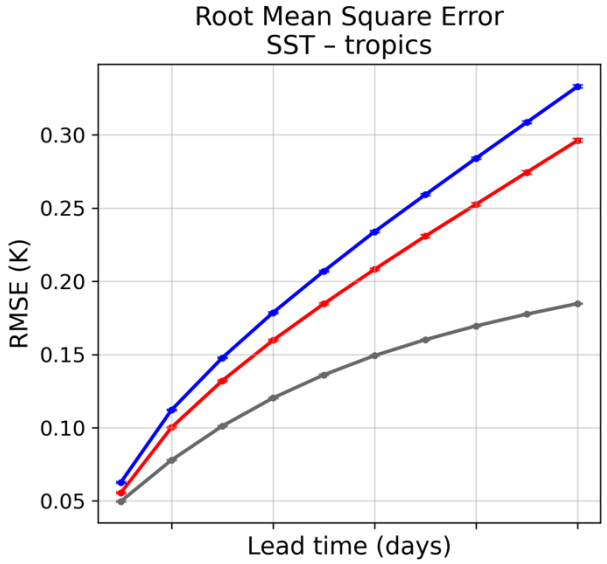
Coupling timestep : 6h (also the forecast timestep)

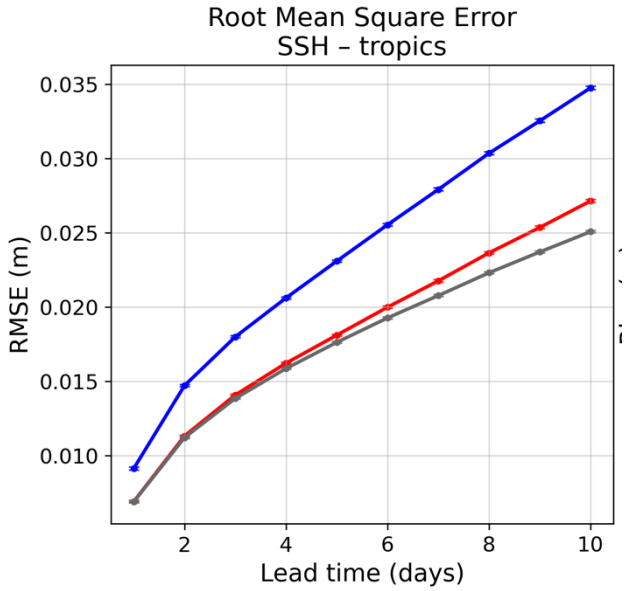
Verification year : 2023

Comparing **coupled** and **joint** model configurations



OCEAN Perspective





Coupled model substantially better approach for ocean fields. Joint training leads to degradation for all variables.

Forced ocean model (OMIP) shows best performances, indicating that at these timescales the atmosphere has a driving role on the surface ocean.

SSH bias develops similarly for all configurations.

Sea ice scores are less affected by differences in coupling strategy.

Experiment Configuration :

- Atmosphere (786) + ocean (512), **coupled model** — 696 forecasts
- Ocean (512) **forced model** — 696 forecasts
- Atmosphere + ocean (1024), **joint model** — 696 forecasts

Coupling timestep : 6h (also the forecast timestep)

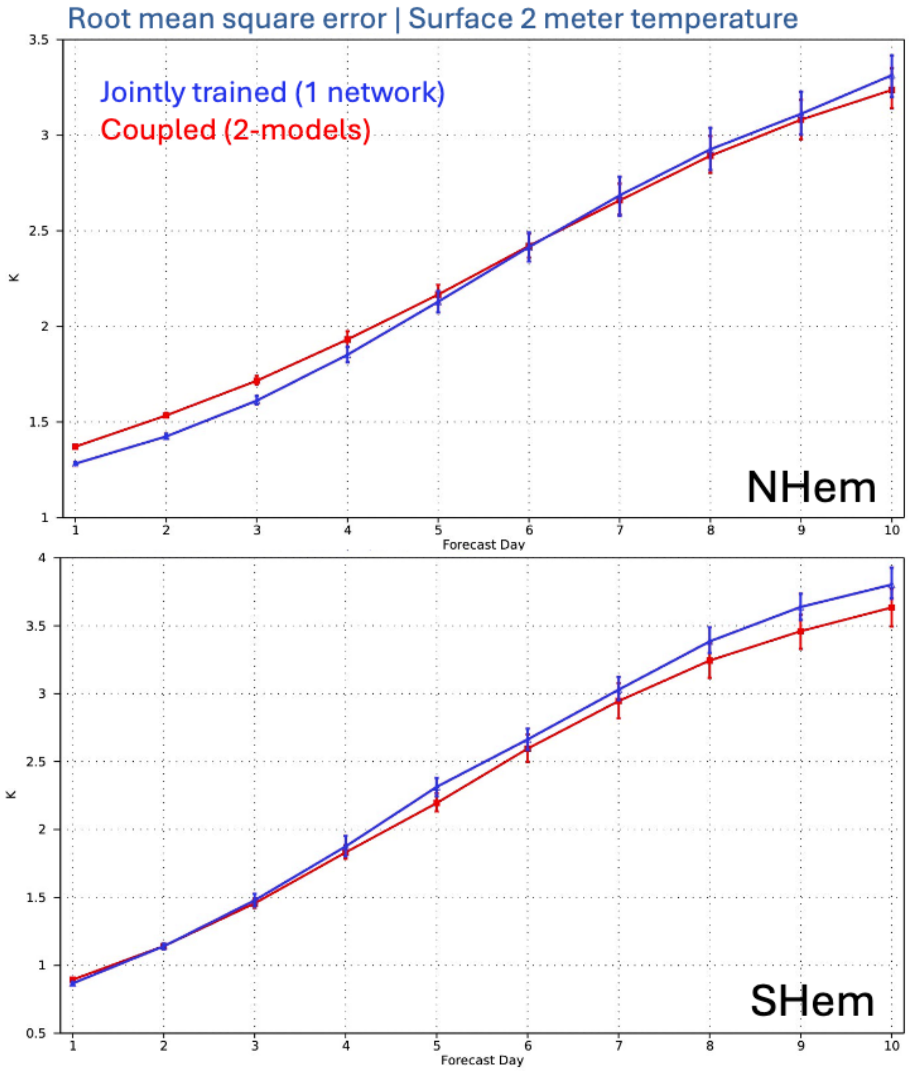
Verification year : 2023

Comparing **coupled** and **joint** model configurations

Example from the land

Using a coupled setup means we keep the land as its own component; its own architecture, grid and timestep, its own capacity and loss, trained on the best data for land and upgradeable on its own.

Coupled model wins at longer lead times; the memory in the soil moisture as a positive impact on 2m temperature

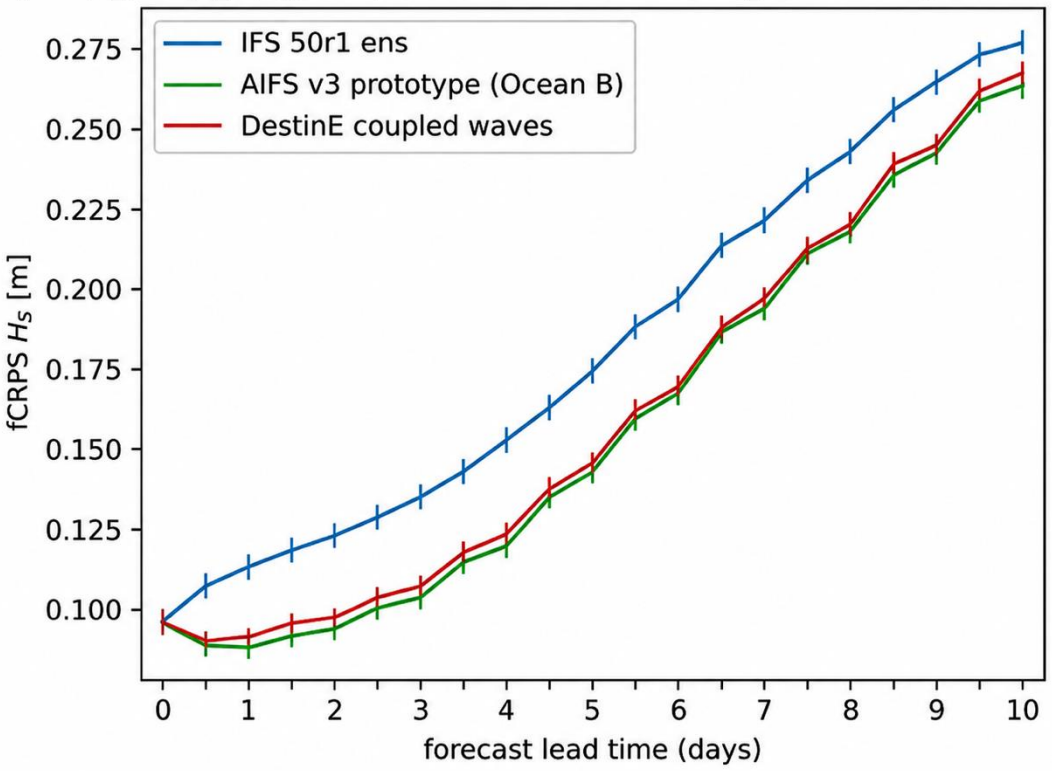


Comparing coupled and joint model configurations

Example from the wave component

For waves, the joint ensemble model and the coupled wave ensemble forced by an ensemble atmosphere model show very similar performance for significant wave height forecasts.

Fair CRPS score for significant wave height forecasts against altimeter observations, June 26 – August 26



Lessons Learned



Lessons learned

Data comes first

Model quality is limited by the training data. Most effort goes into datasets, normalisation and checking for gaps.

The data is already coupled

Reanalysis fields carry the interactions between components. Models learn them implicitly, such as waves vanishing under sea ice.

Scores are not enough

Check case studies, long rollouts, spectra and extremes, not only RMSE at day 1–5.

Engineering matters

Software packages like anemoi make configs reproducible, datasets versioned and checkpoints self-describing, which saves a lot of time.

Involve domain experts

Meteorologists and component experts spot unphysical behaviour that scores miss.

Purpose of model defines training target and architecture

Forecast performance, water conservation, stability, etc. Choose the loss for how the model will be used.

Building an Earth system model in anemoi

Sara Hahner and Nina Raoult

Machine Learning Scientists, ECMWF

