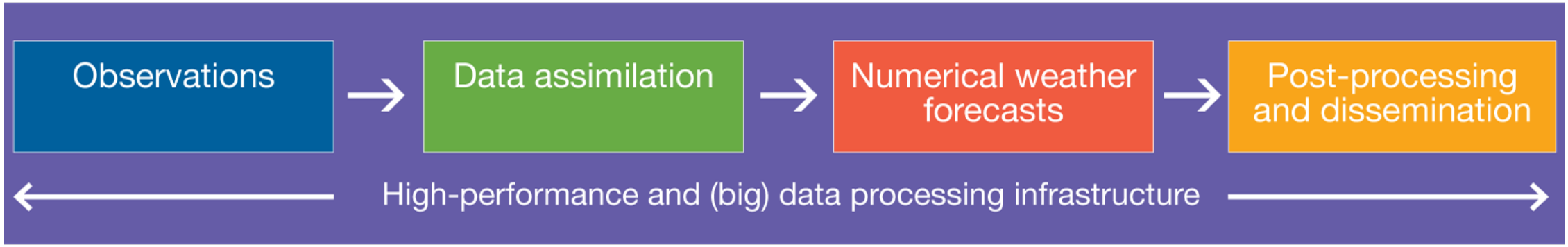
Data-driven weather forecasts and AIFS

Mariana Clare

ECMWF

© ECMWF November 16, 2026

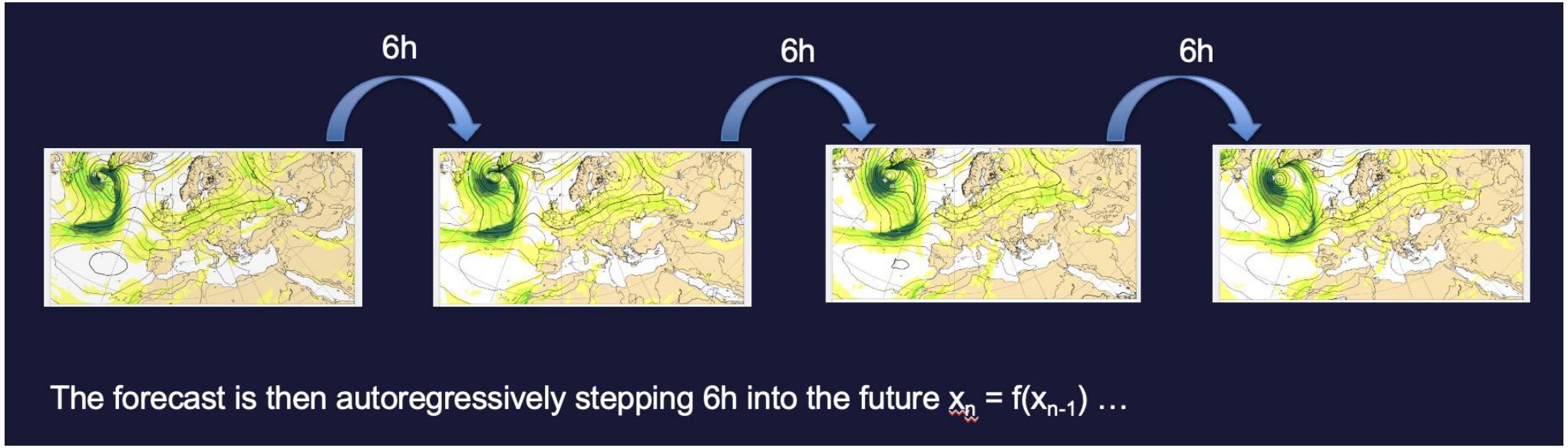
Classical NWP



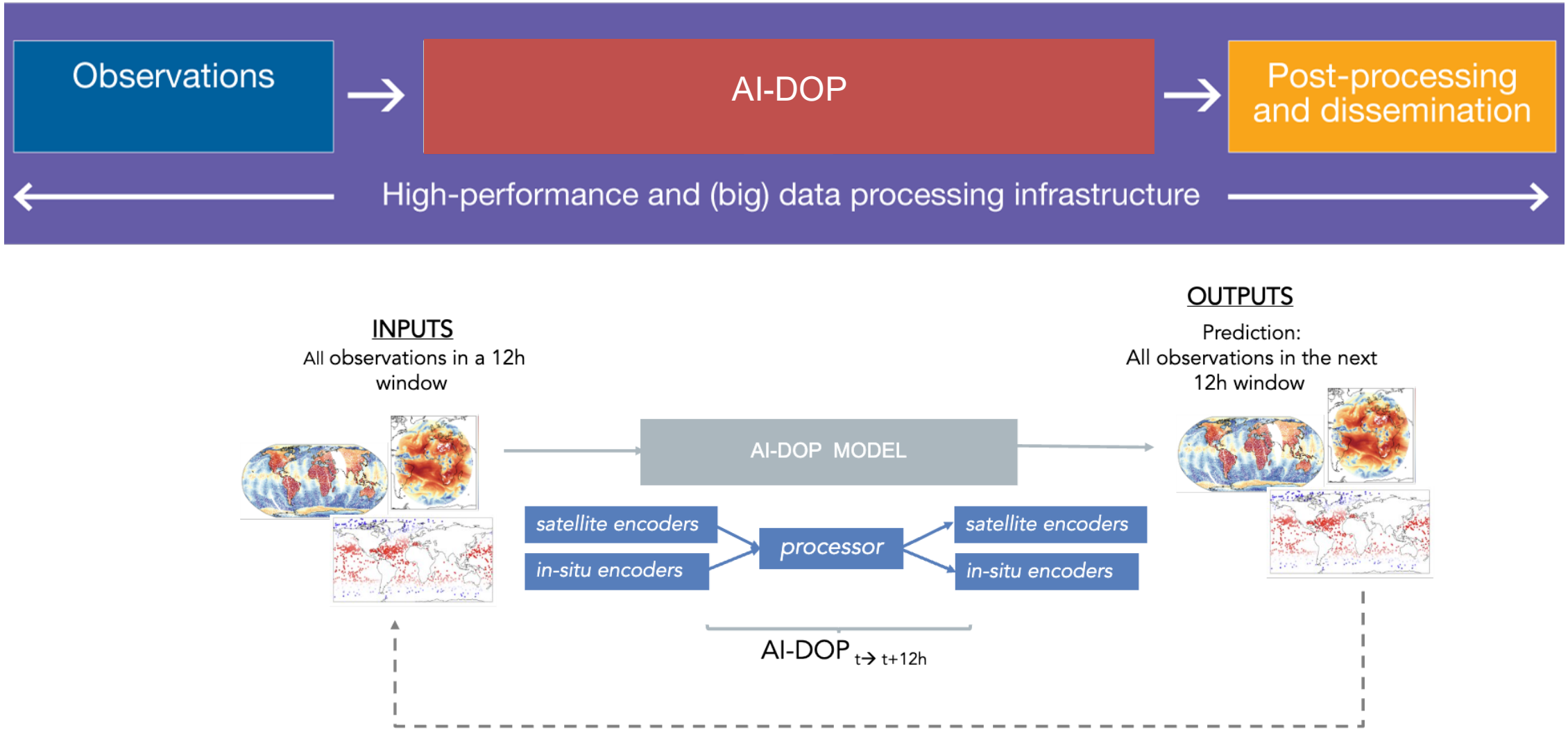
Data-driven weather forecasting system



Most data-driven weather forecasting models are trained on reanalysis datasets



Data-driven weather forecasting system



Training on observations (e.g. SYNOPs) at arbitrary locations allows for generalisation to standard gridded model output

A short history of data-driven weather forecasting

February 2022 – First competitive medium-range systems

- **Keisler –** GraphNN, competitive with GFS (USA)
- **NVIDIA** – **FourCastNet** Fourier+ , 0.25°, **O(10⁴) faster & more energy efficient than IFS**

December 2022

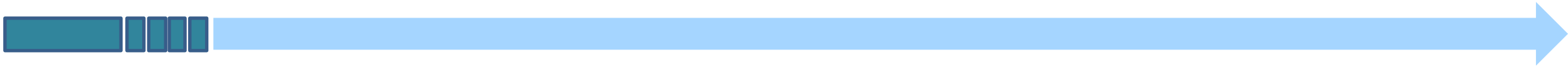
- **Deepmind – GraphCast**

GraphNN
0.25° **Many parameters with comparable skill to IFS.**

November 2022

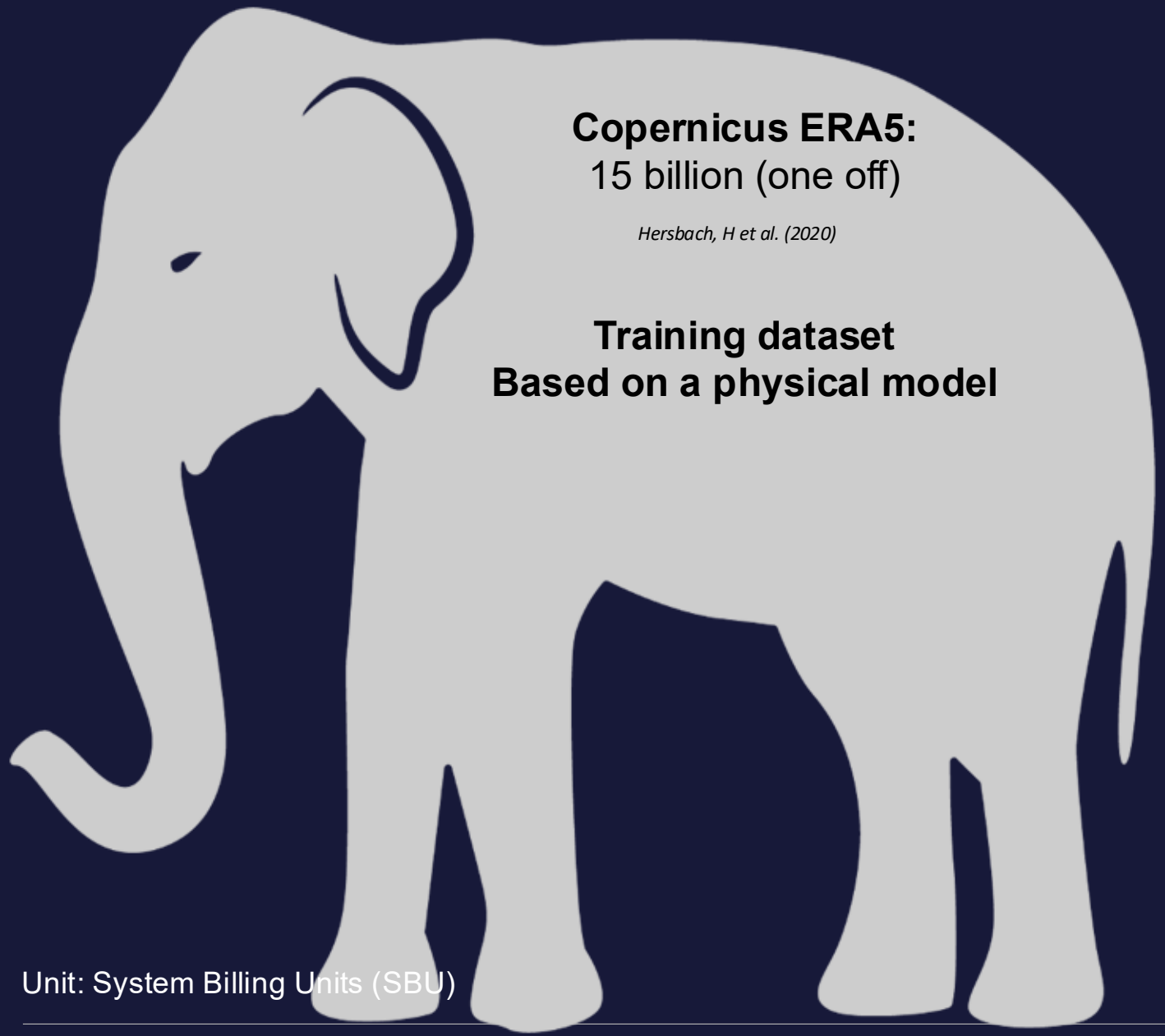
- **Huawei – PanguWeather**

Vision Transformer
0.25° **"More accurate tropical cyclone tracks" than the IFS.**



2018 – Concept explored (ECMWF and others)…

ECMWF



Copernicus ERA5:
15 billion (one off)

Hersbach, H et al. (2020)

Training dataset
Based on a physical model

Physics-based Forecast:
180 000
per forecast

AI Model:
0.3
per forecast

~1000x
reduction in
energy
Forecast time
reduced from
~30minutes to
~3minutes



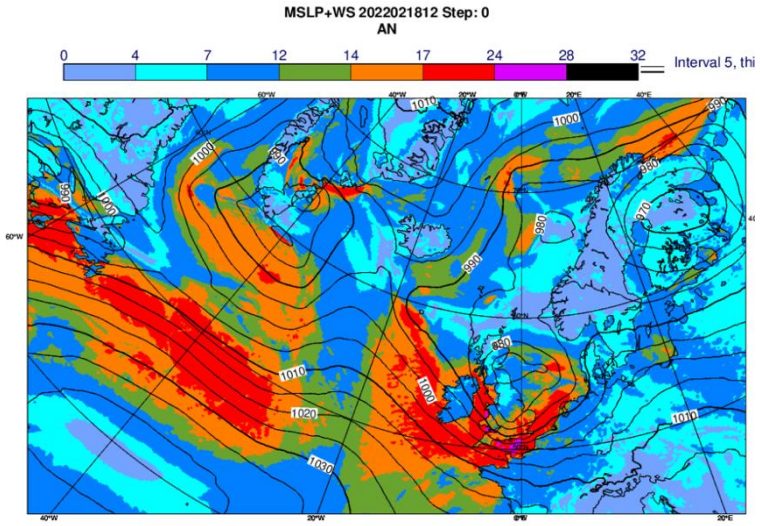


Unit: System Billing Units (SBU)

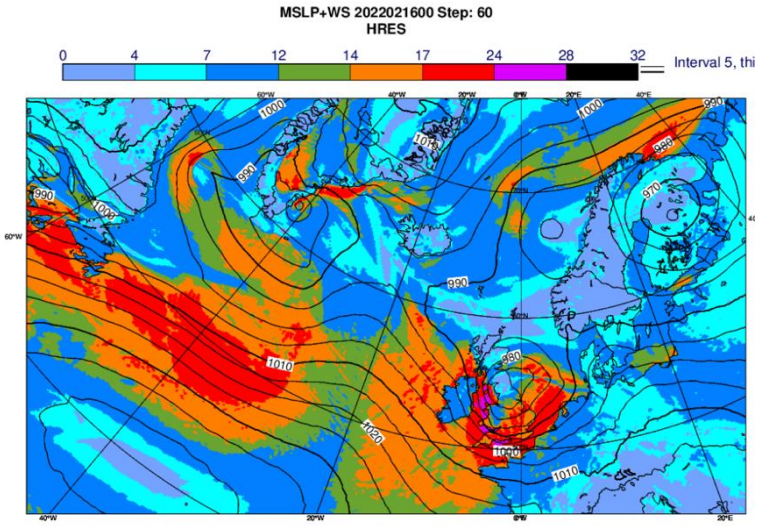
2022-02-16 00z + 60h (Storm Eunice over UK)

See ECMWF Newsletter 176

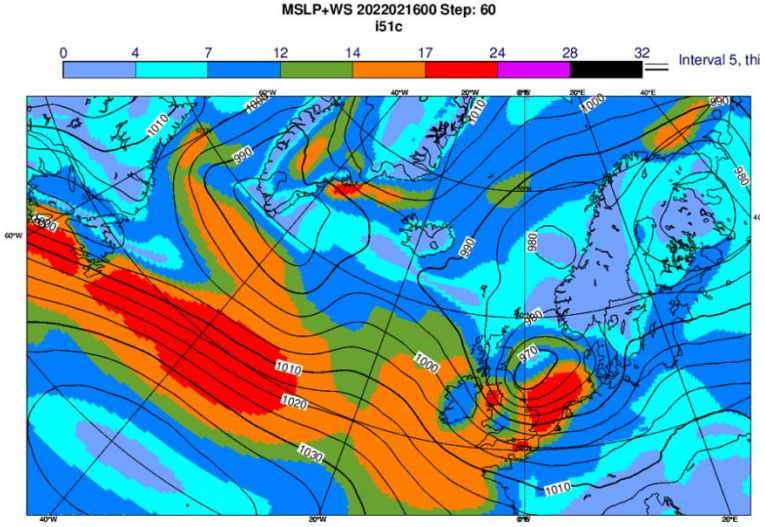
Analysis



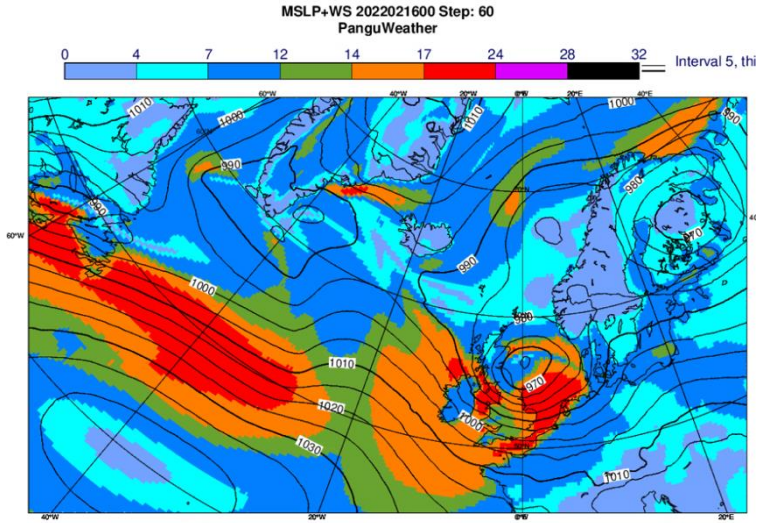
HRES



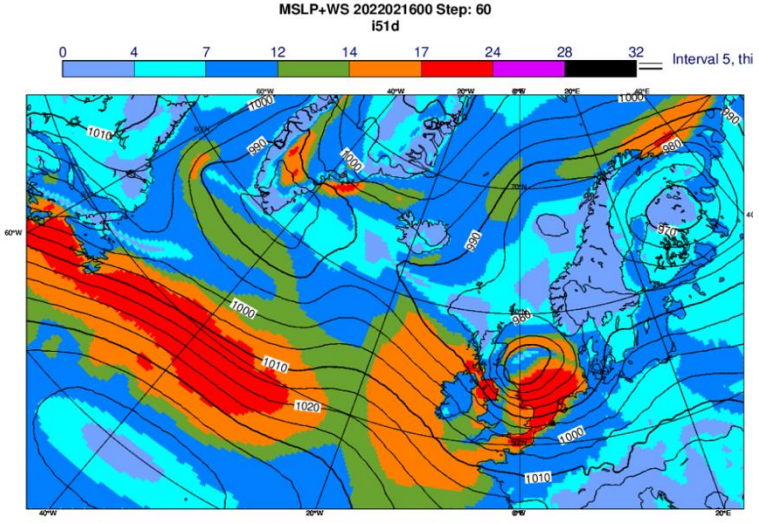
FourCastnet



PanguWeather



GraphCast



A short history of data-driven weather forecasting

2018 – Concept explored (ECMWF and others)…

February 2022 – First competitive medium-range systems

- **Keisler –** GraphNN, competitive with GFS (USA)
- **NVIDIA – FourCastNet** Fourier+ , 0.25°, **O(10⁴) faster & more energy efficient than IFS**

November 2022

- **Huawei – PanguWeather**

Vision Transformer
0.25° **"More accurate tropical cyclone tracks" than the IFS.**

December 2022

- **Deepmind – GraphCast**

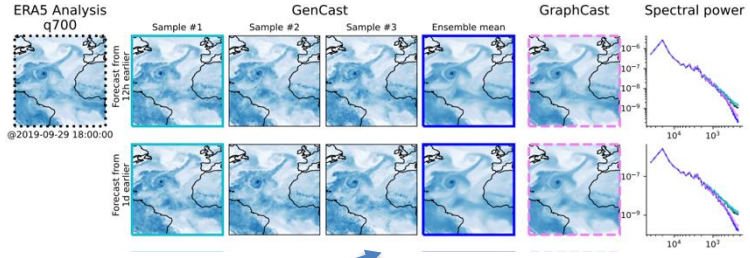
GraphNN
0.25° **Many parameters with comparable skill to IFS.**

January-June 2023

- **Microsoft – ClimaX**
- **China academia/Shangai Met – FengWu**
- **Alibaba – SwinRDM**
- **NVIDIA – SFNO**
- **…**

Early 2023
Prototype AIFS developments begin

June 2023
ECMWF – ML project begins



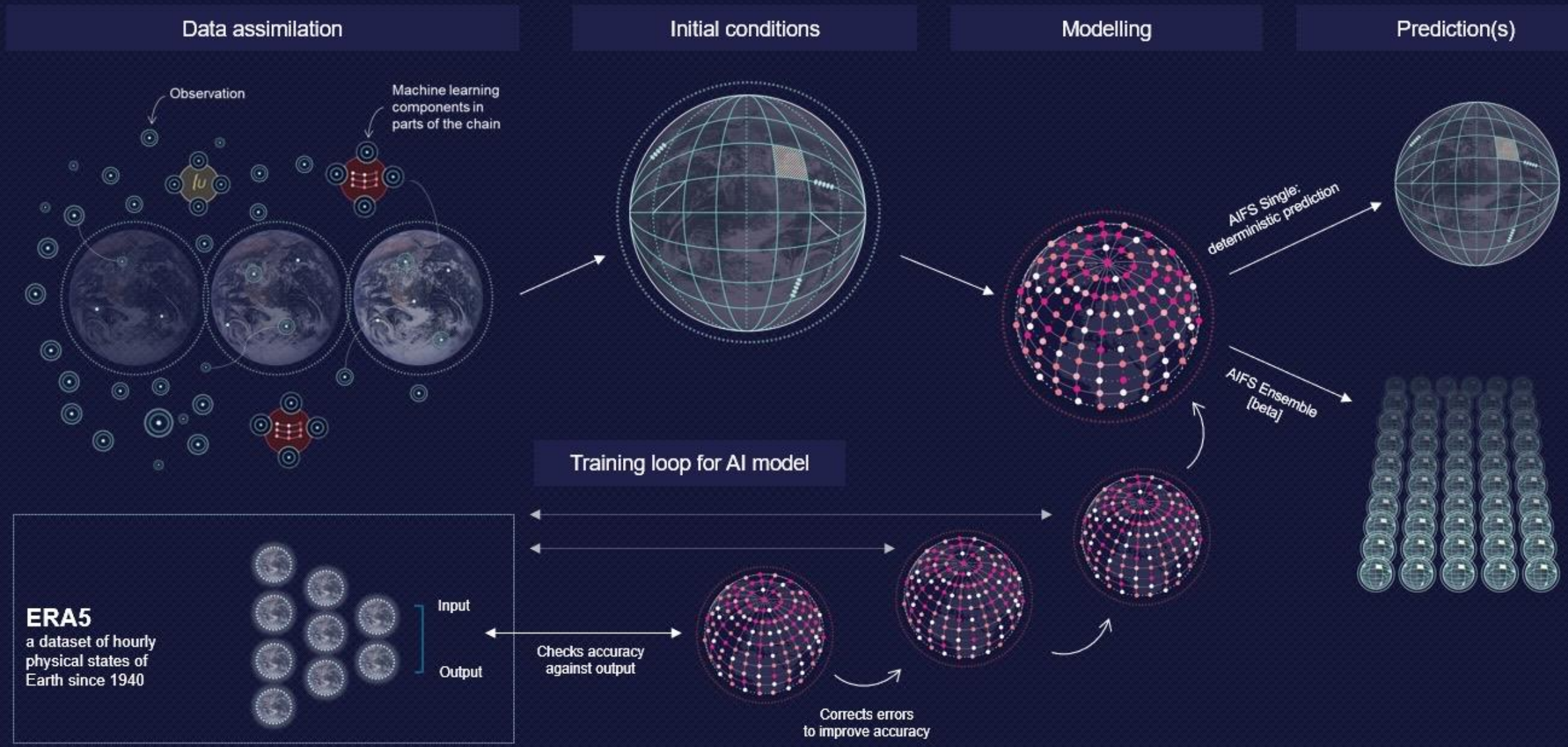
December 2023

- **Deepmind – GenCast**

Probabilistic forecast (ensemble) – 1°
"Outperforming the leading operational ensemble forecast" (aka ECMWF)

ECMWF

AIFS: Artificial Intelligence Forecasting System



A short history of data-driven weather forecasting

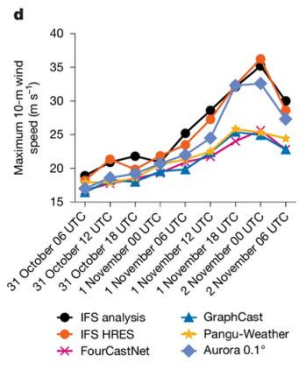
February 2022 – First competitive medium-range systems

- **Keisler –** GraphNN, competitive with GFS (USA)
- **NVIDIA – FourCastNet** Fourier+ , 0.25°, **O(10⁴) faster & more energy efficient than IFS**

December 2022

- **Deepmind – GraphCast**

GraphNN
0.25° **Many parameters with comparable skill to IFS.**



November 2022

- **Huawei – PanguWeather**

Vision Transformer
0.25° **"More accurate tropical cyclone tracks" than the IFS.**

January-June 2023

- **Microsoft – ClimaX**
- **China academia/Shangai Met – FengWu**
- **Alibaba – SwinRDM**
- **NVIDIA – SFNO**
- **…**

December 2023

- **Deepmind – GenCast**

Probabilistic forecast (ensemble) – 0.25°
"Outperforming the leading operational ensemble forecast" (aka ECMWF)

June 2024

- **Microsoft – Aurora**

Higher resolution – 0.1°
Atmospheric composition

2018 – Concept explored (ECMWF and others)…

June 2023
ML project begins

Jan/Feb 2024
AIFS first updates

July 2025:
AIFS ENS CRPS 1 operational

Early 2023
Prototype AIFS developments begin

October 2023
AIFS experimental forecasts live

Feb 2025:
AIFS Single 1 operational

ECMWF

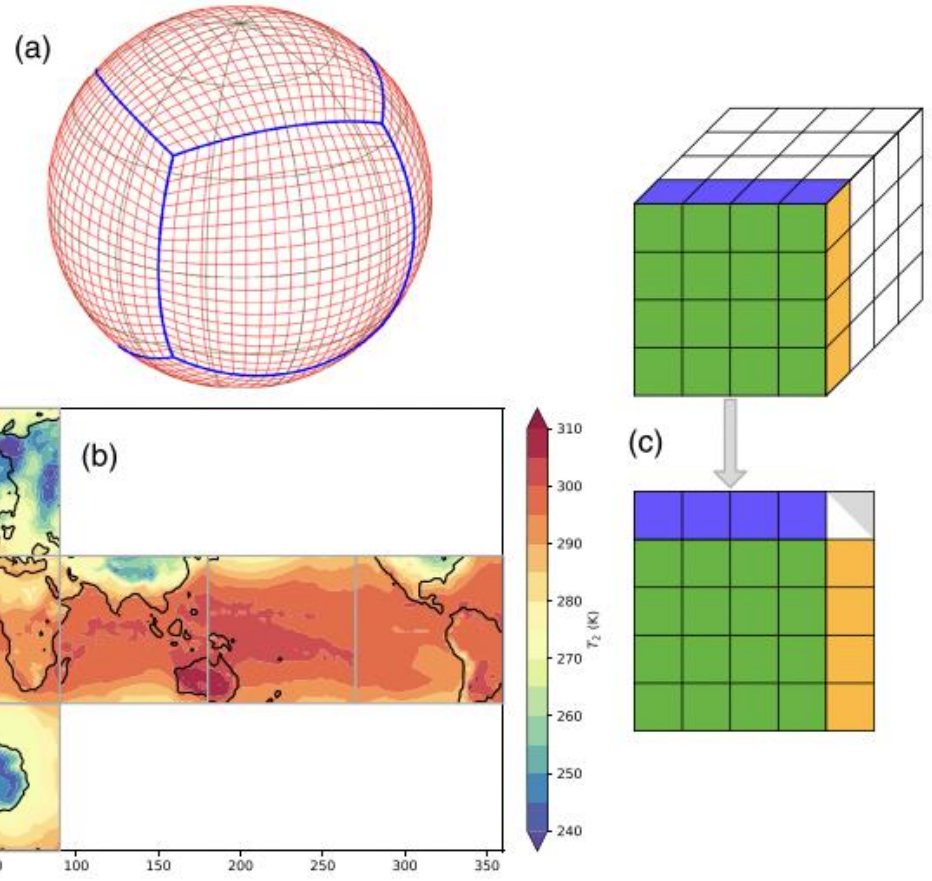
Different model architectures

ECMWF

Convolutions

- Simplest design, treat the Earth as a cylinder, use convolutions with periodicity in longitude.
  - No treatment of the pole. How can flow easily pass over the arctic?

- Weyn et al 2020 proposed a clever cubed-sphere approach.
  - One set of CNN for the side faces of the cube.
  - Another for the polar faces.



- Karlbauer et al (2023) do convolutions on the HealPix grid (see below)

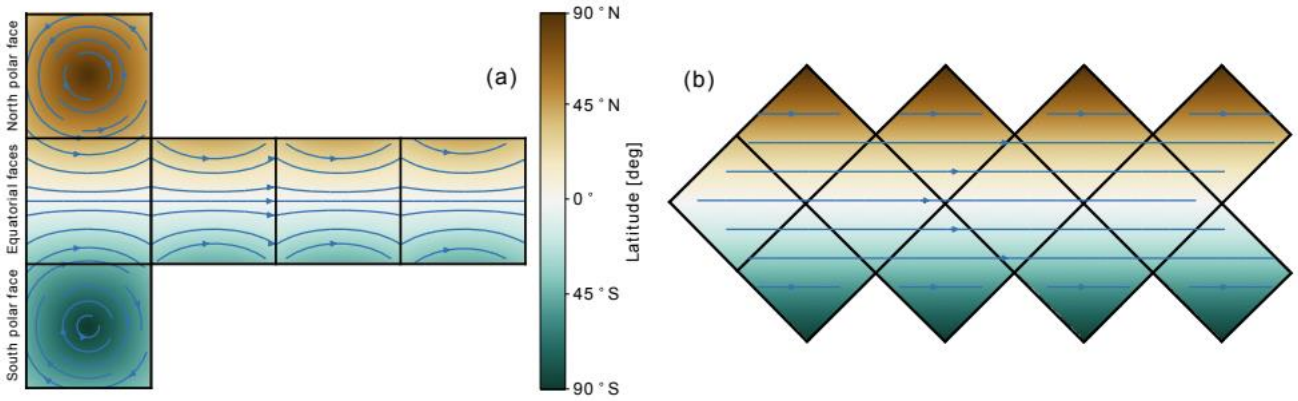
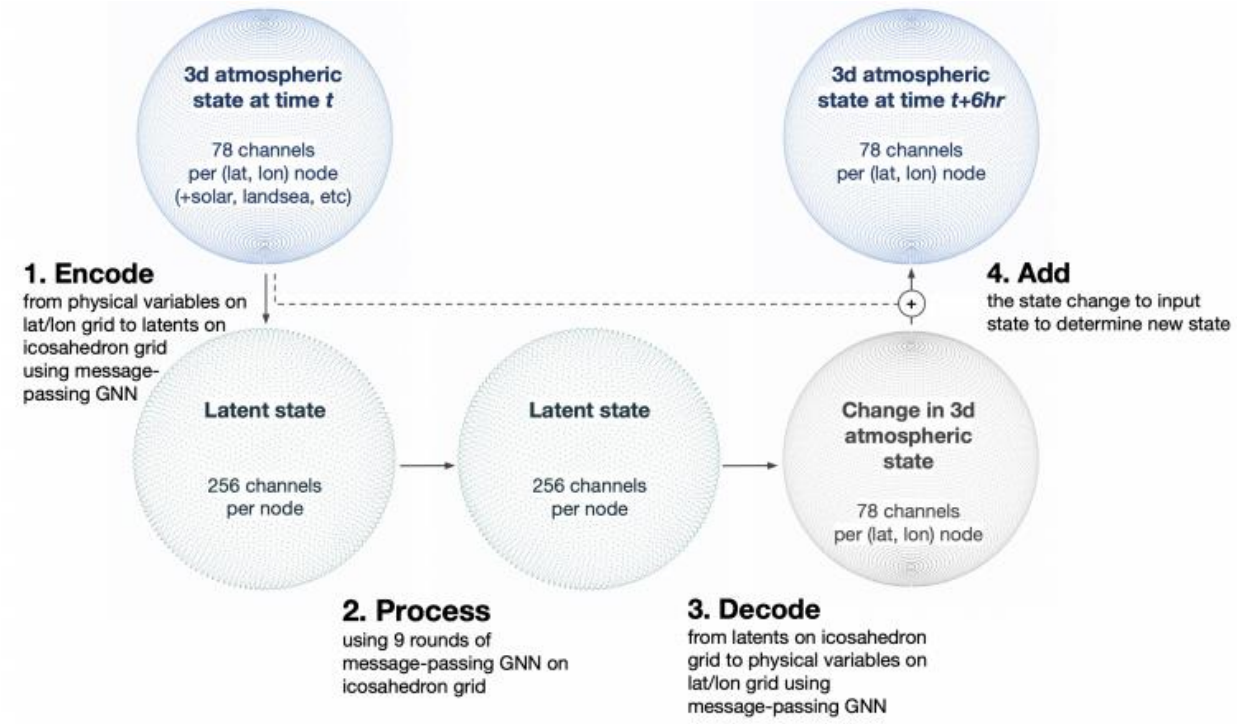
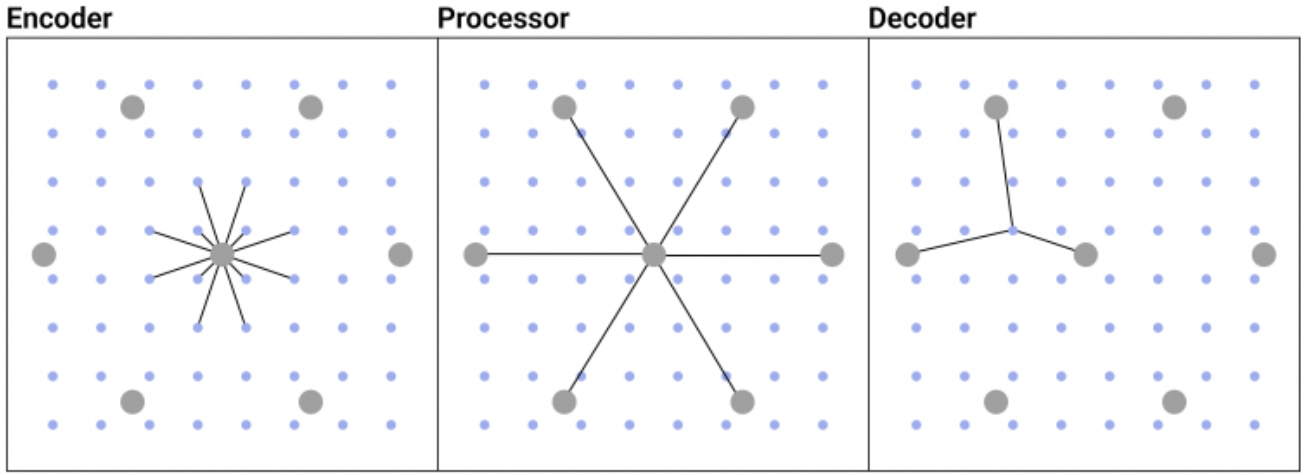


Figure 10: Lines of latitudes depicted as blue streamline arrows on the cubed sphere (a) and on the HEALPix (b). While the lines corresponding to constant eastward motion describe arcs of different radii on the cubed sphere mesh, the same motion translates to straight lines on the HEALPix mesh.

Graph Neural Networks

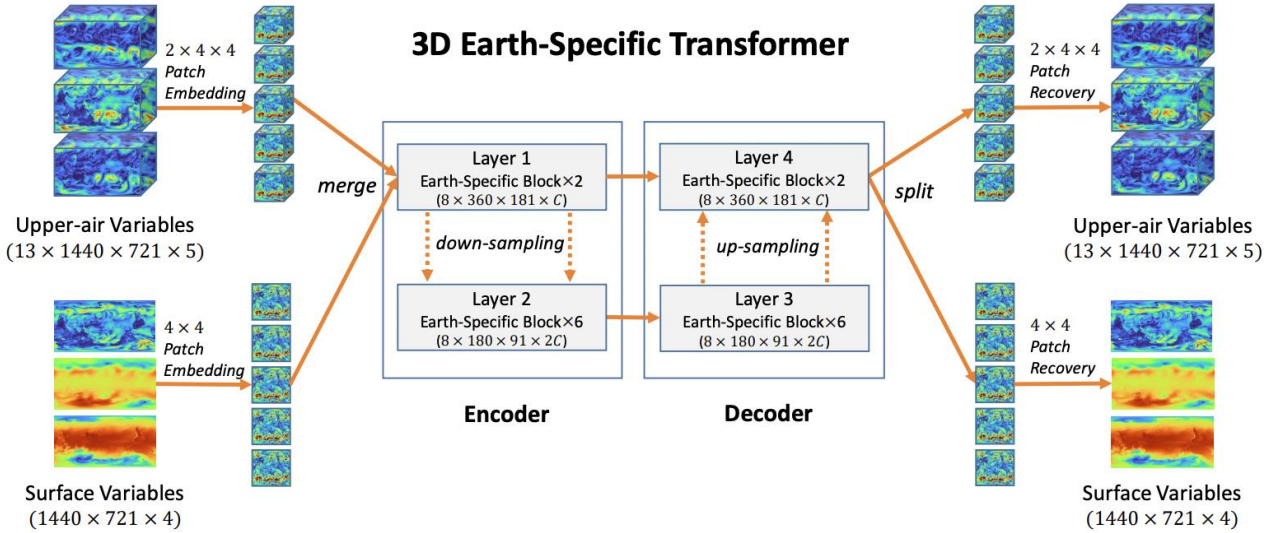
- First demonstrated by Keisler 2022.
- Data is structured in nodes and the connections between them (edges)
- Most popular, the message passing GNN.
  - Involves MLPs on the edges, and nodes.
  - Alternates are GraphConvolutions & Graph Attention
- Can handle unstructured data in space.



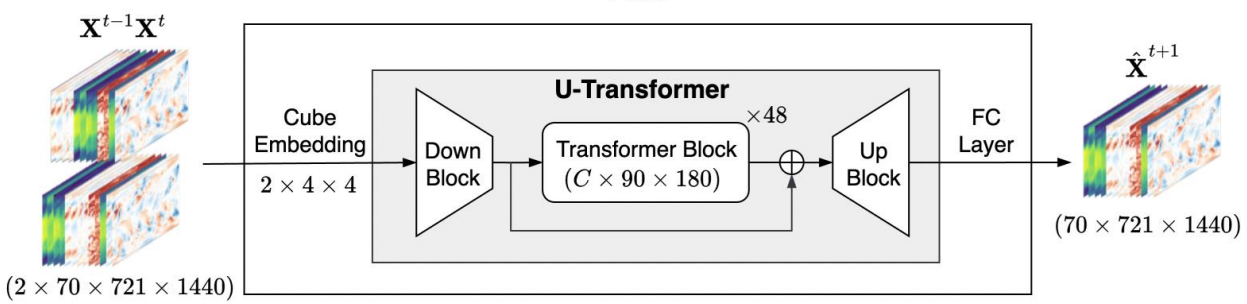


(Vision) Transformers

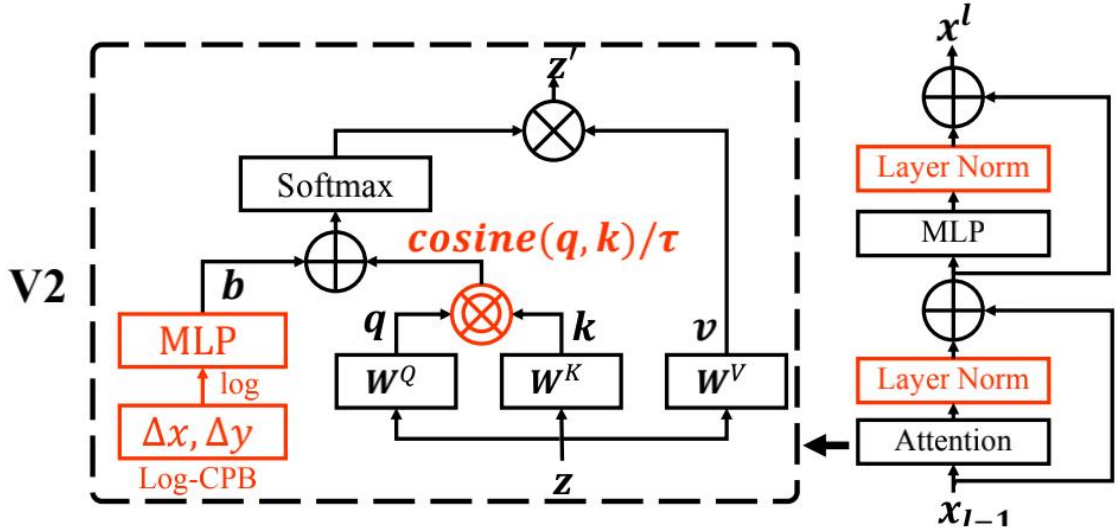
- Build heavily on advances in vision transformers
- Pangu, FuXi, FengWu.
- Embed to a coarser resolution.
- Shifted-window approach adapted to include longitudinal periodicity but poles are not explicitly handled.



a) The overall architecture of FuXi model

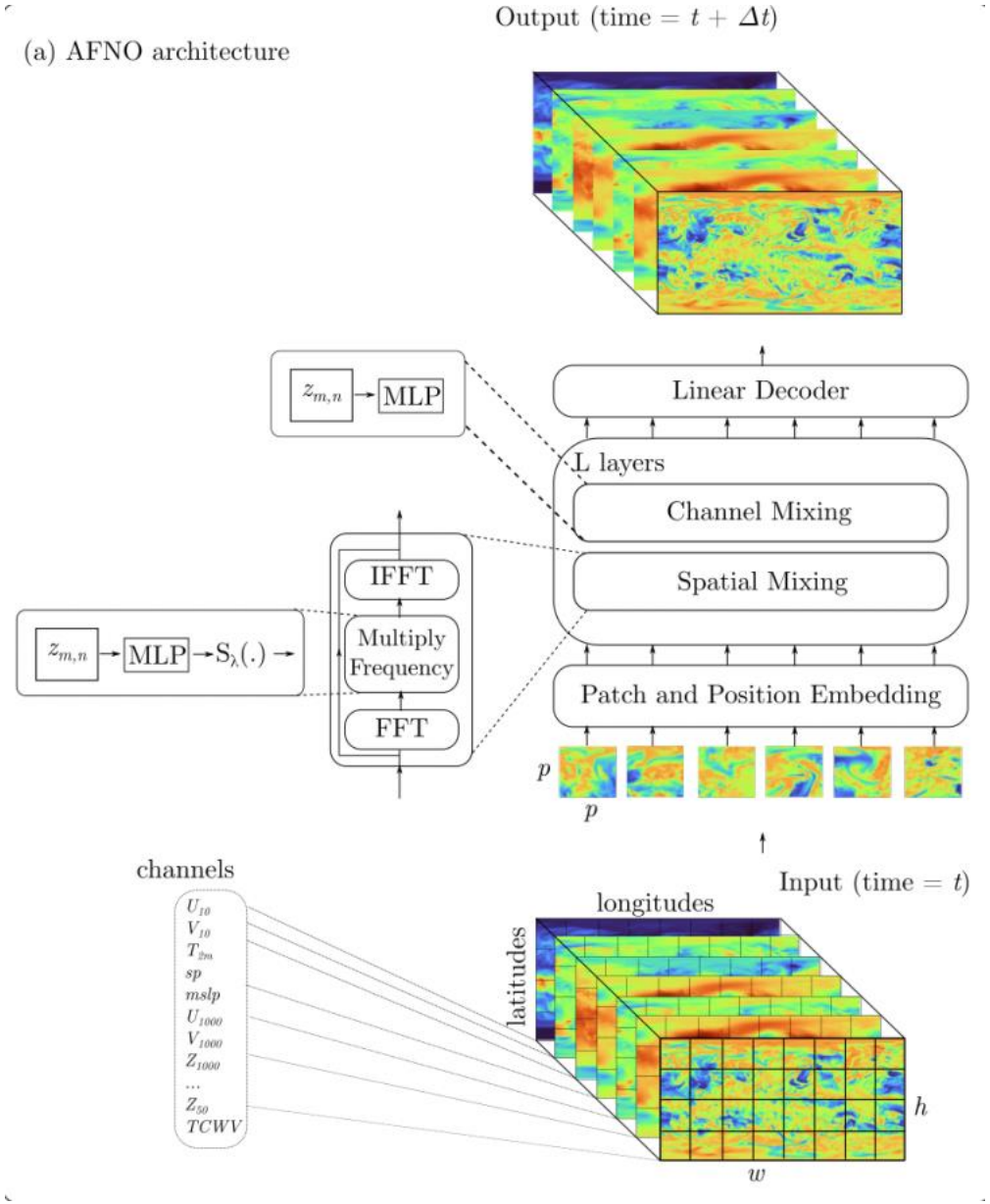


SwinTransformerV2 block



Fourier Neural operators

- e.g. FourCastNet, popularised by NVIDIA.
- Part of neural network carried out in frequency space and part in grid-point space.
- Grid-invariance built in.
- Spherical version encodes the symmetries of the sphere.
- Also used in “ACE”, the climate emulator.



So, which to choose?

- Vision transformer and GNN solutions both hit comparable levels of skill.
  - SFNO a little further behind in skill, unclear why.
  - CNN not been implemented at the same scale.

- GNN naturally encodes the sphere and allows use of unstructured grids.

- Vision transformers (and SFNO) appear to converge faster than GNN.

What variables to use?

Two driving motivations:

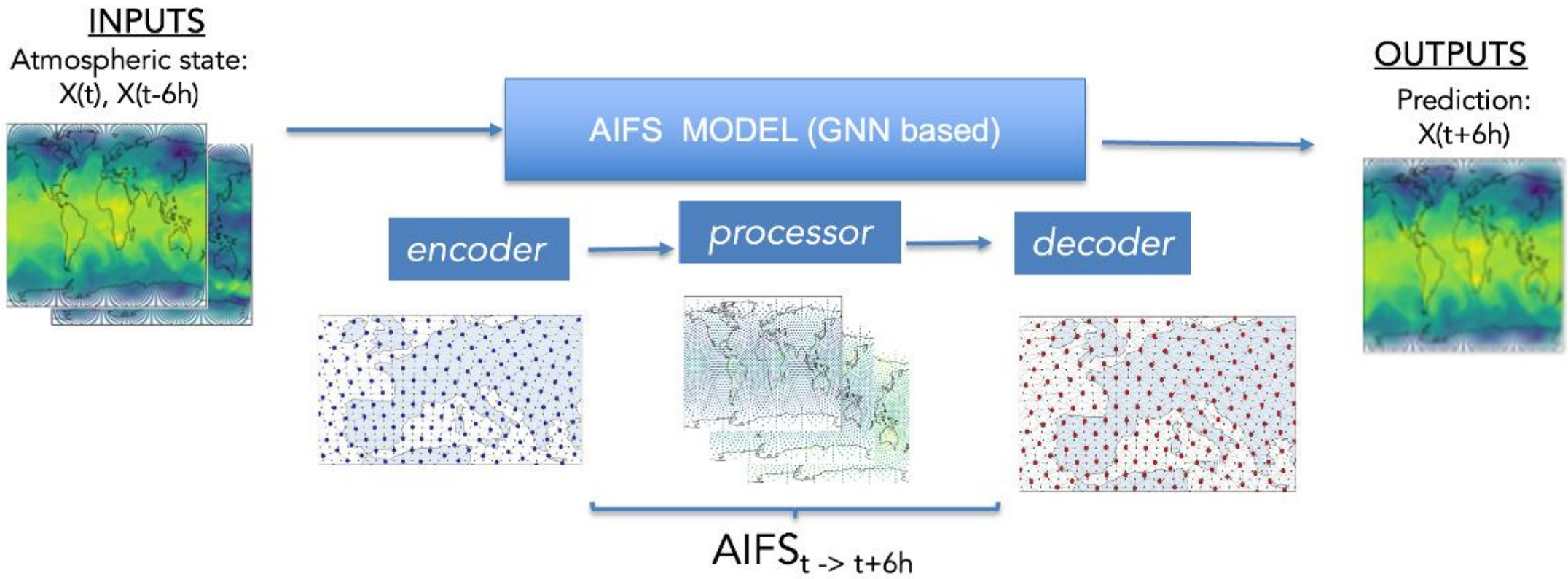
- What helps me predict better.
- What do users want from the system.

Typical set used by many models:

- ~13 pressure levels (with model top at 50hPa)
  - Contrast with 137 model levels in the IFS.
  - GraphCast version with 37 levels isn't more skilful than 13 level version.
- q, t, u, v, w, z
  - But no direct cloud information.
- At the surface: 2t, msl/sp, 10u/v and precipitation for some.
- Increasingly training on Earth System Components

AIFS Single v1

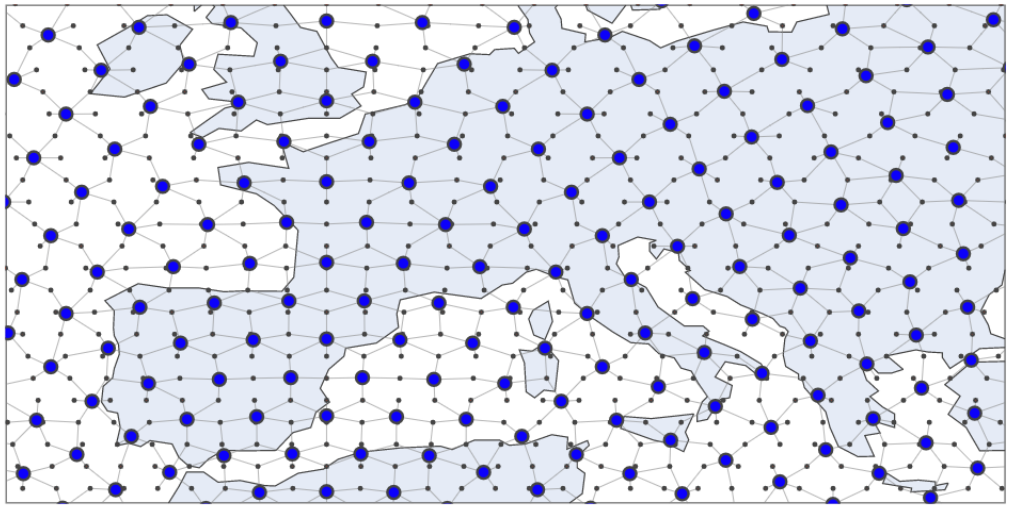
Lang et al 2024a
Operational system
from 25/2/25



AIFS – Encoder and Decoder

AIFS works with the native IFS reduced gaussian grids ;
possible to split model and input data across multiple GPUs to handle large memory requirements

Encoder, GNN



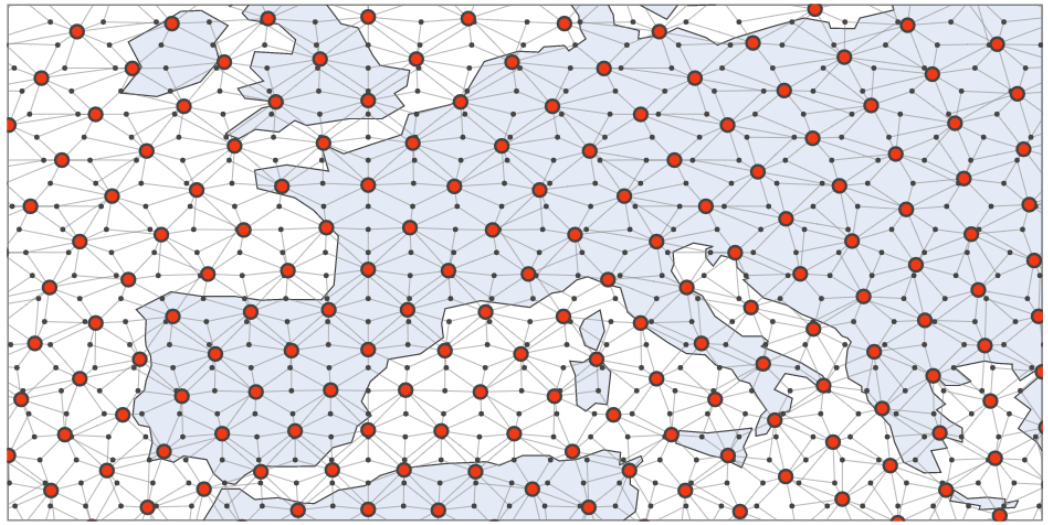
ERA5 – n320

~ 540 000 Nodes,
~ 1 Million Edges

16 Processor Layers

O96 ~ 40 000 Nodes

Decoder, GNN



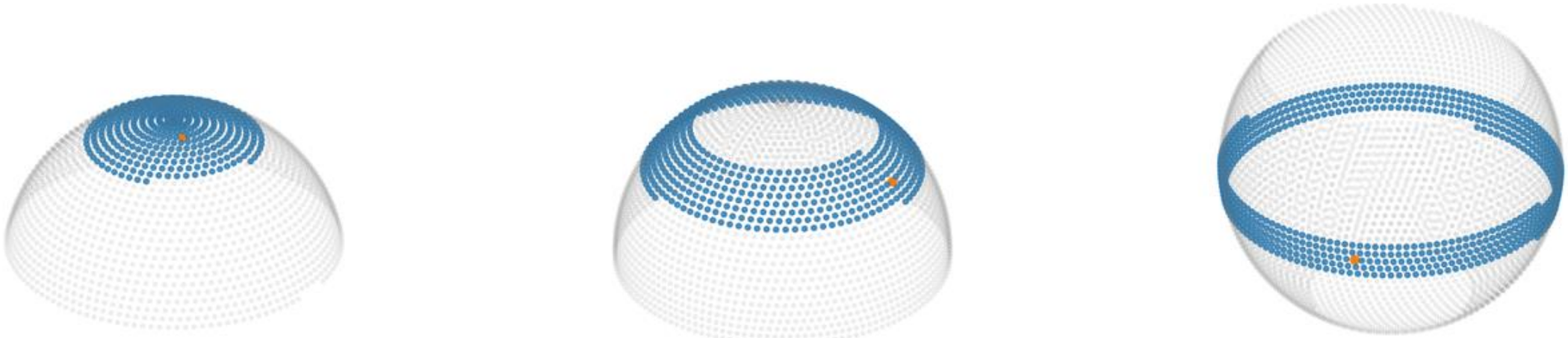
ERA5 – n320

AIFS – Processor

Transformer (like LLMs) that works with a sliding attention window -> attention bands around the globe

Red: target node
Blue: Nodes target node attends to in one processor layer
Grey: How far information can travel within e.g. 6 processor layers

(here lower resolution processor grid than operational AIFS for visualization)



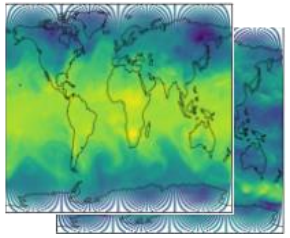
Training setup

Training

TRAINING

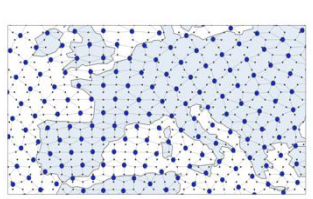
INPUTS

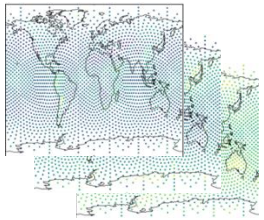
Atmospheric state:
X(t), X(t-6h)

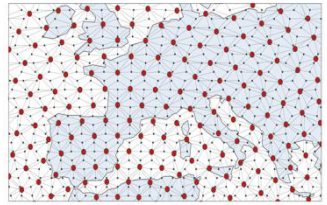


AIFS MODEL (GNN based)

encoder → *processor* → *decoder*



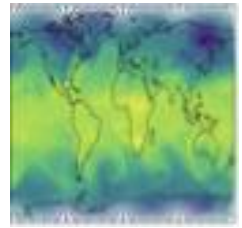




AIFS$_{t \rightarrow t+6h}$

OUTPUTS

Prediction:
X(t+6h)



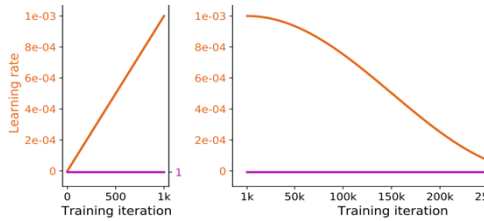
Lang et al. (2024)

Training Scheme

Step 1: pre-training phase, during which the model is given the task to forecast 6 h ahead

Step 2: model is trained to produce forecast up to 72 hr ahead

Step 3: model is fine-tuned on operational real-time analysis



Scheme – Step 1

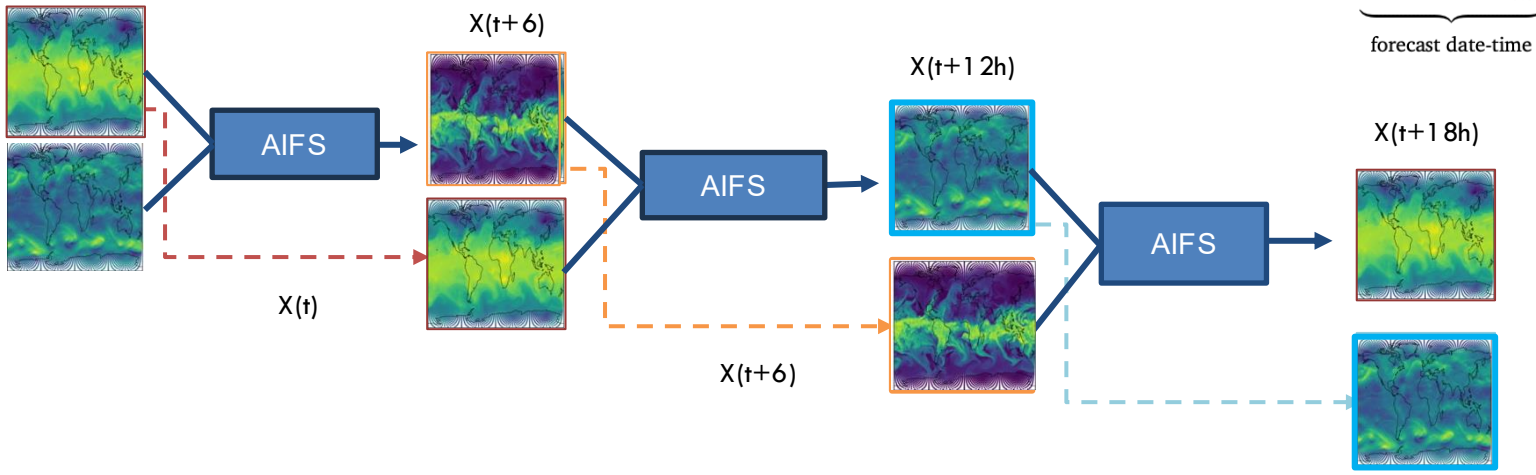
Scheme – Step 2 & 3

72 hr forecast - "AKA" Rollout

Training

TRAINING SCHEME – Step 2

Atmospheric state:
X(t), X(t-6h)



WMSE - area-weighted mean squared error

$$\mathcal{L}_{\text{MSE}} = \underbrace{\frac{1}{|D_{\text{batch}}|}\sum_{d_0 \in D_{\text{batch}}}}_{\text{forecast date-time}} \underbrace{\frac{1}{T_{\text{train}}}\sum_{\tau \in 1:T_{\text{train}}}}_{\text{lead time}} \underbrace{\frac{1}{|G_{0.25^\circ}|}\sum_{i \in G_{0.25^\circ}}}_{\text{spatial location}} \underbrace{\sum_{j \in J}}_{\text{variable-level}} s_j w_j a_i \underbrace{(\hat{x}_{i,j}^{d_0+\tau} - x_{i,j}^{d_0+\tau})^2}_{\text{squared error}}$$

AR predictions
Up to rollout 12 (72 hours)

WMSE t+6h + WMSE t+12h + WMSE t+18h + ... + WMSE t+72h

model is trained to produce forecast up to 72 hr ahead by aggregating the WMSE

ROLLOUT 1 - 6h

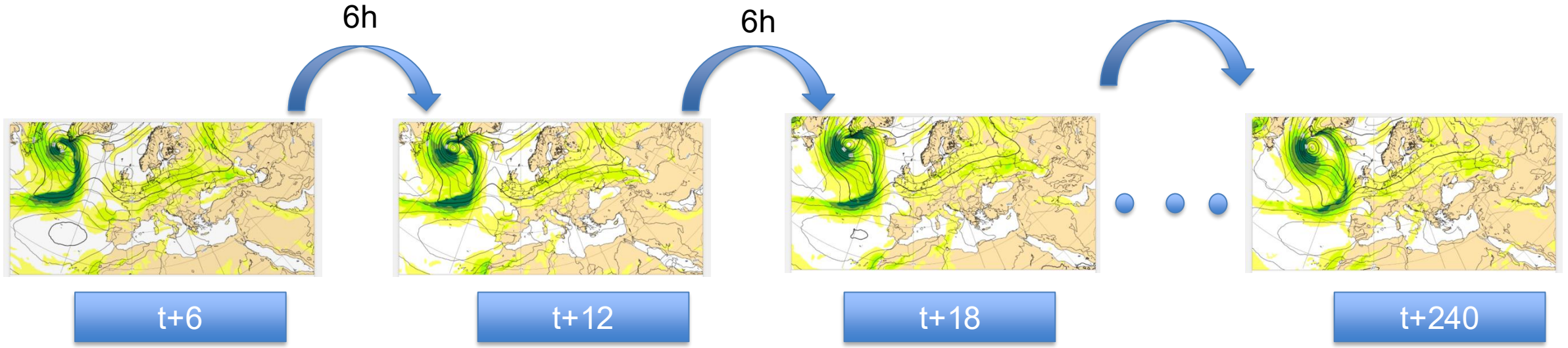
ROLLOUT 2 - 12hr

ROLLOUT 3 – 18hr

ROLLOUT 12- 72hr

Inference

Once the model is trained, we can generate predictions stepping e.g. 6h from analysis to analysis

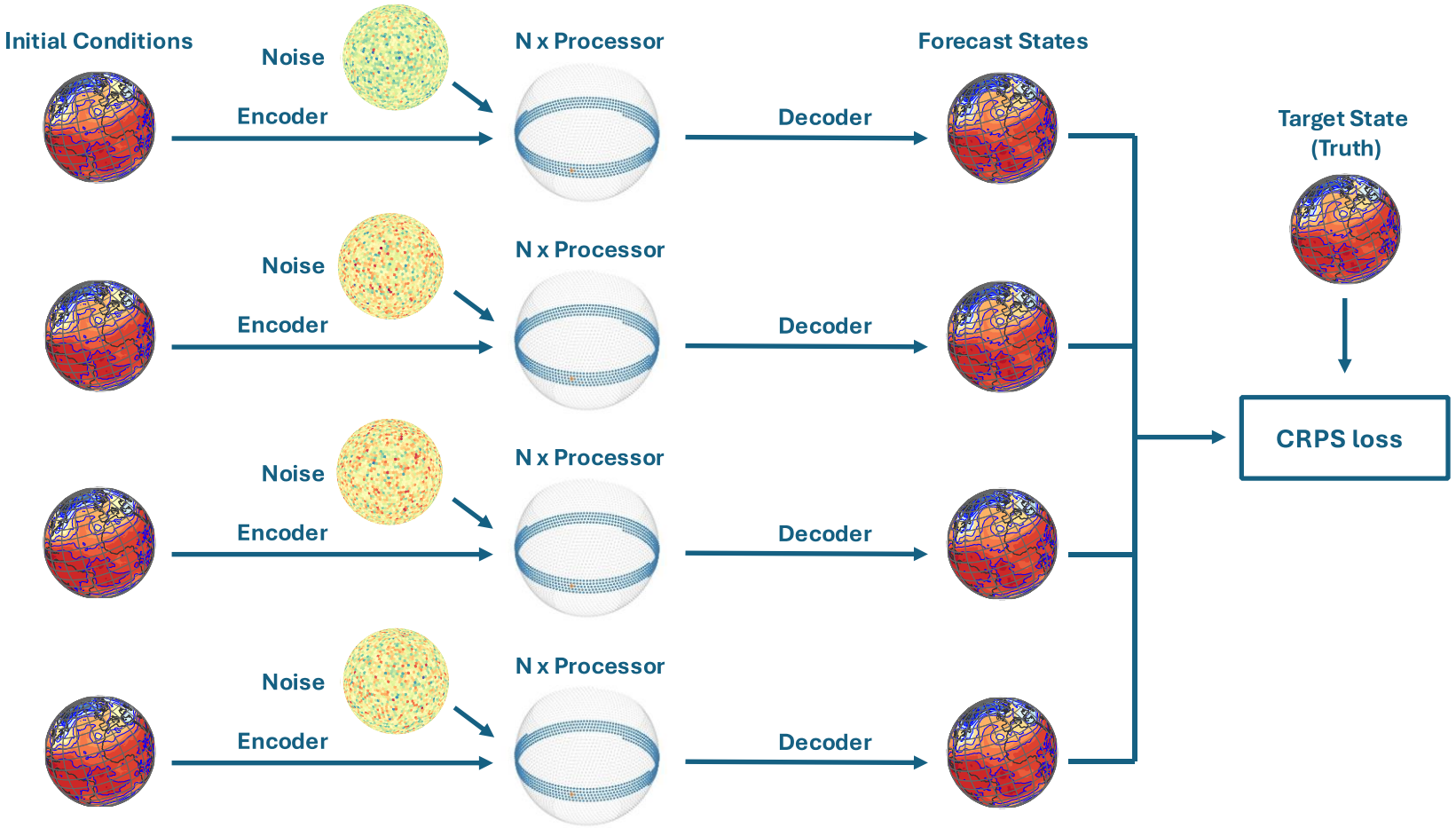


The forecast is then autoregressively stepping 6h into the future $x_n = f(x_{n-1})$ …

Proper score loss – AIFS-ENS (AIFS-CRPS):

Lang et al 2024b
Operational system from 1/7/25

In training: run (small) ensemble:



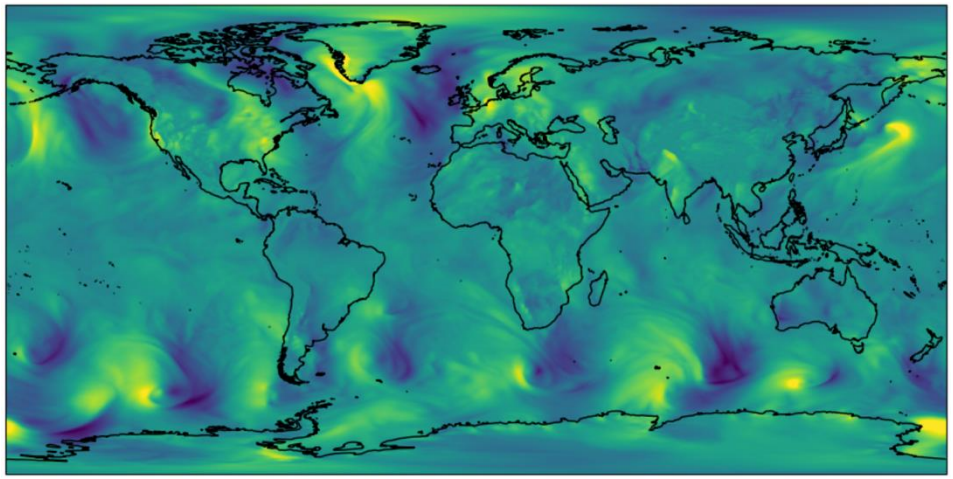
To generate a forecast (single member): run model with noise realization for each forecast step

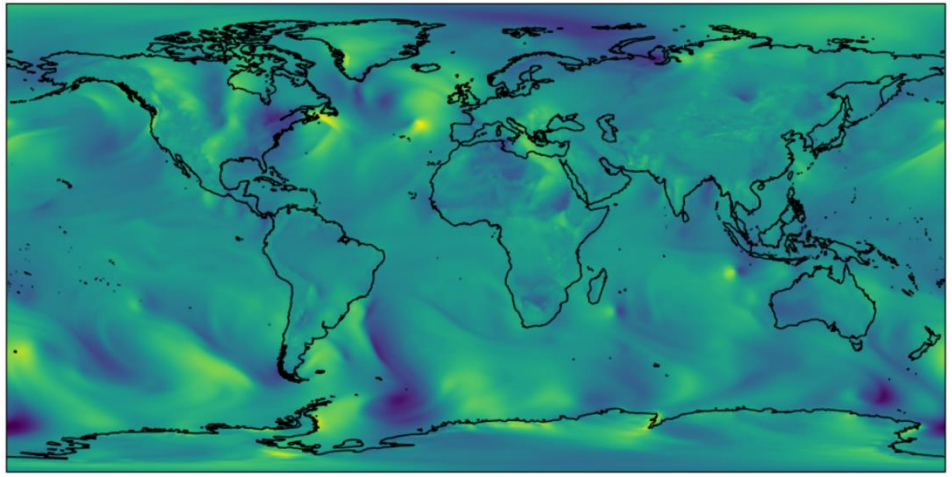
AIFS CRPS: CRPS training loss avoids smoothed forecasts

meridional wind at 850 hPa

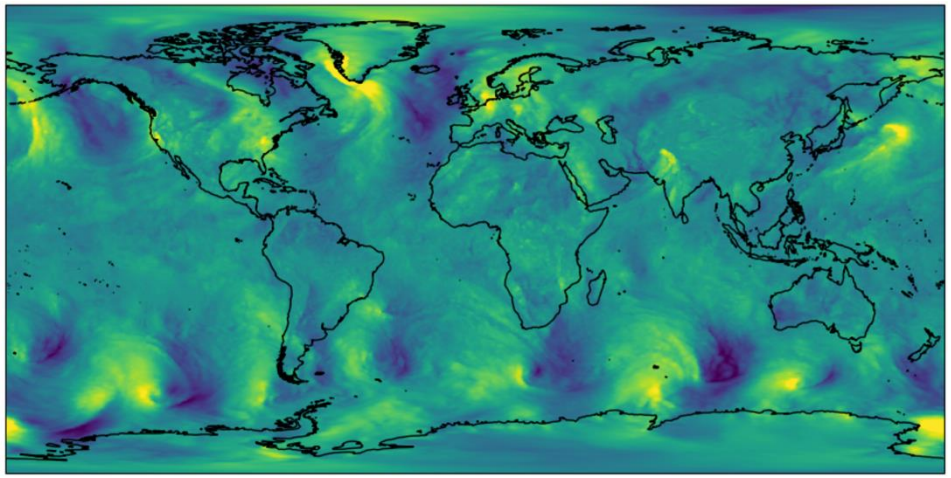
1-day forecast **10-day forecast**

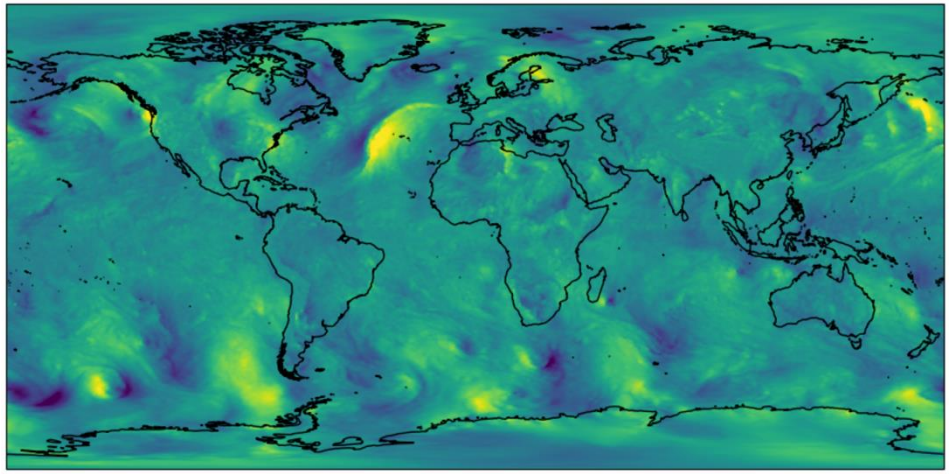
AIFS Single





AIFS CRPS





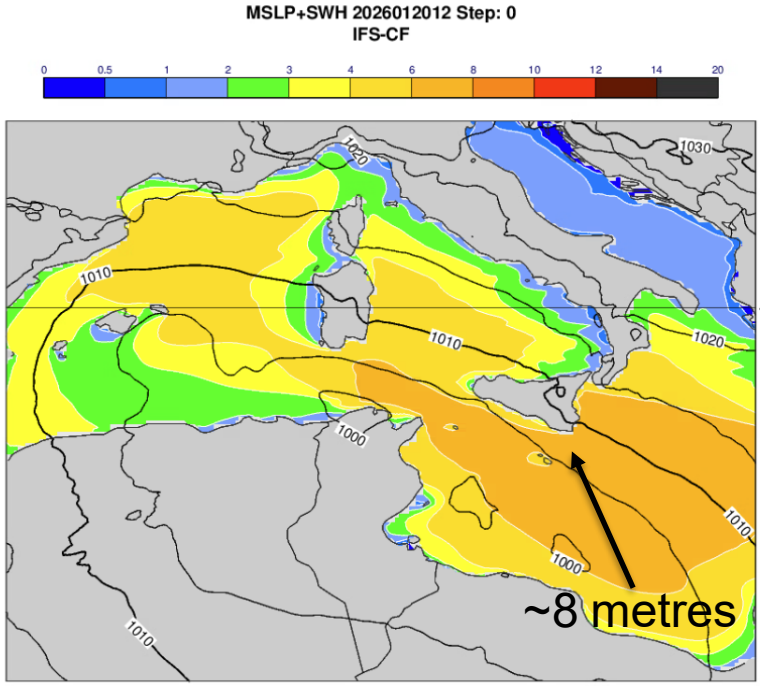
AIFS v2 – Operational on 12th May 2026

- General improvement in forecast accuracy and less smoothing in AIFS-single forecasts
- ECMWF will issue data-driven wave forecasts for the first time, leading to a substantial improvement in medium-range skill of wave forecasts
- New data-driven land surface variables
- Marked improvements in forecast accuracy at the top of the atmosphere

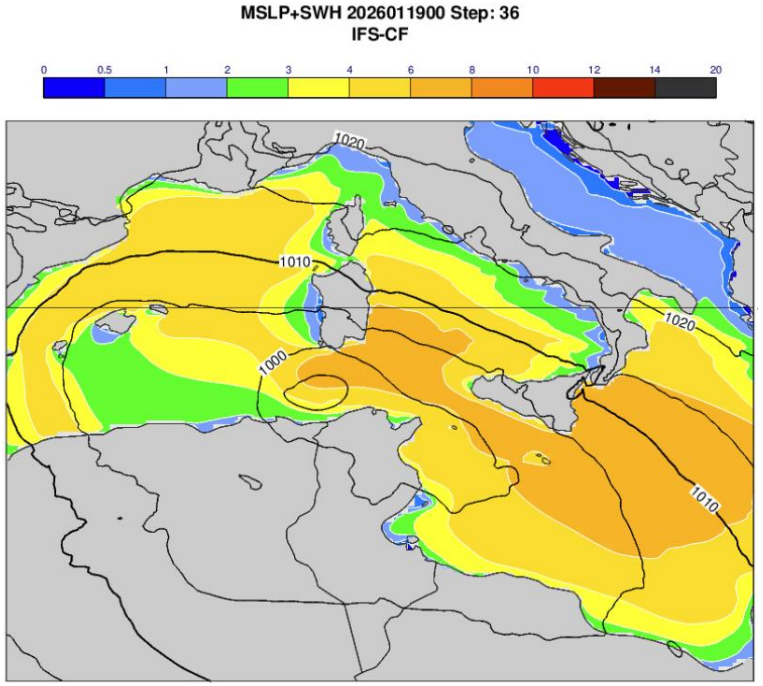
Wave forecasts for Storm Harry in January 2026

Significant wave height

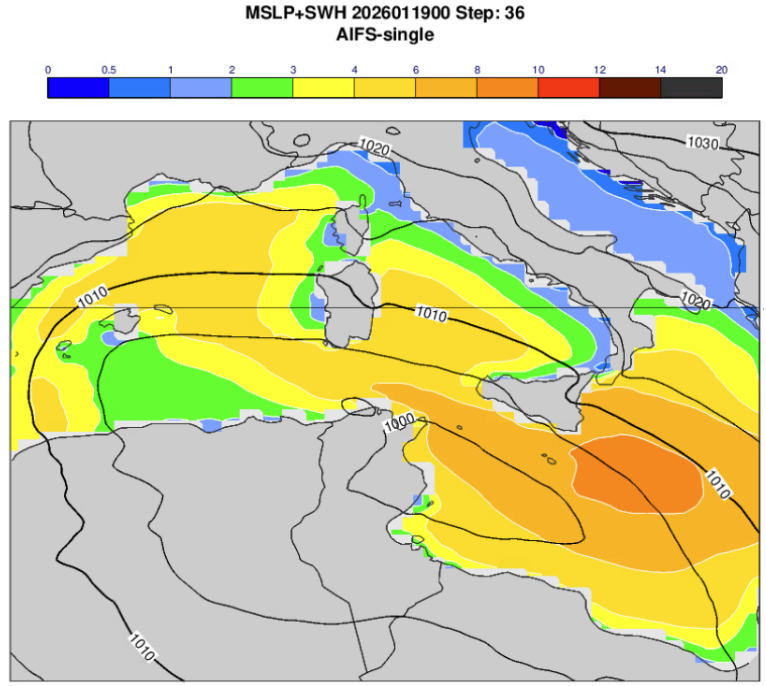
IFS Analysis 20 Jan 12UTC



IFS 19 Jan 00UTC + 36h



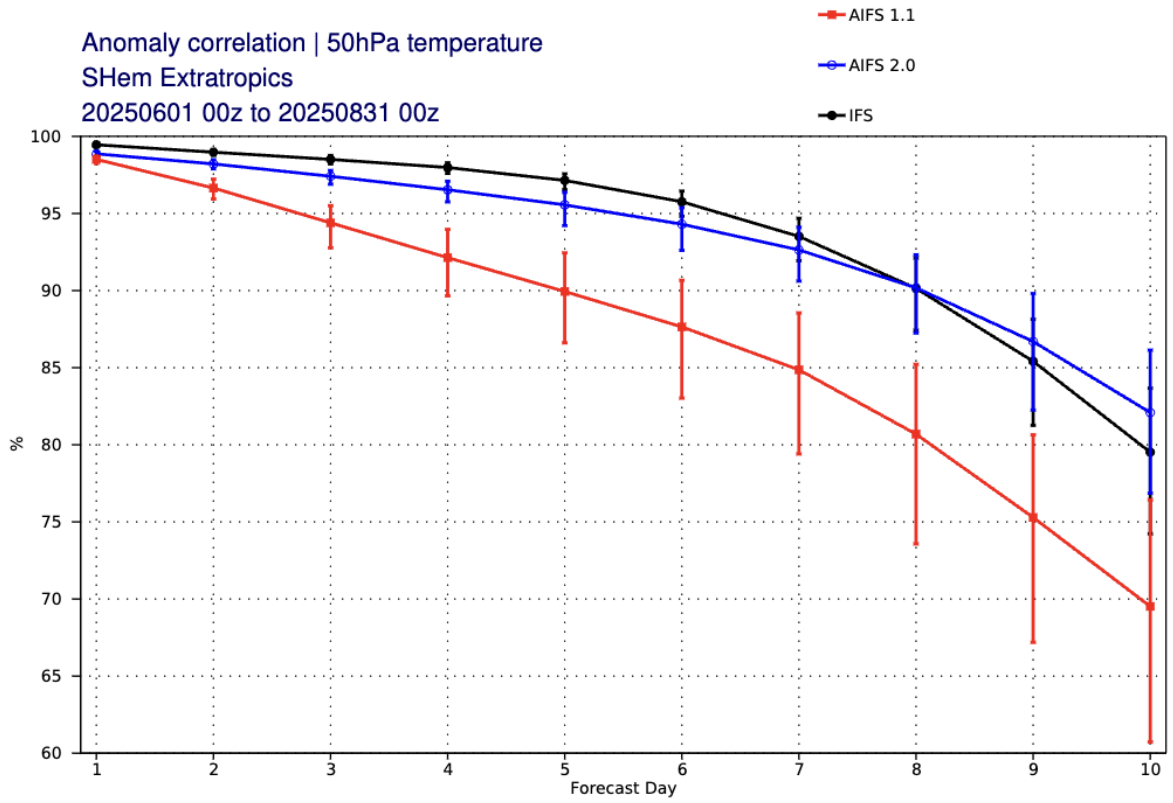
AIFS v2 19 Jan 00UTC + 36h



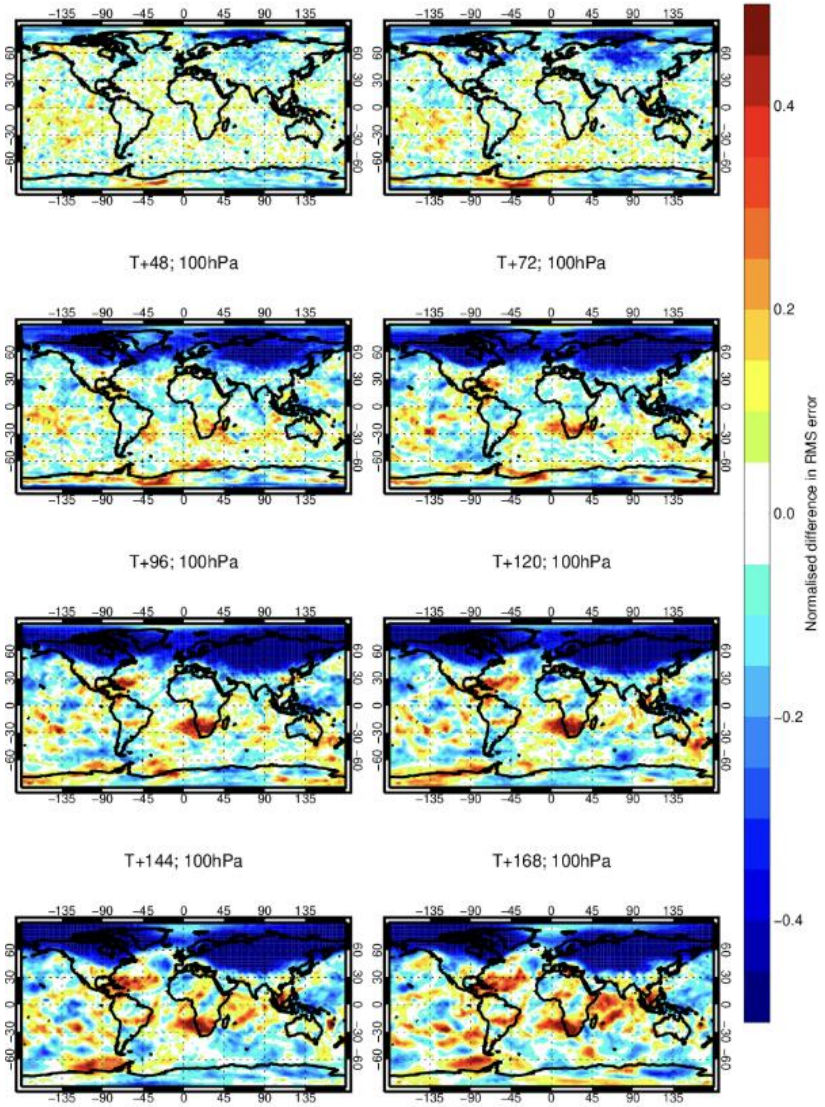
Case study of extreme wave height in the Mediterranean

Improved stratosphere

Anomaly correlation vs analysis (summer 25)



Improvements in temperature at 100hPa (2.0 from 50r1 vs 1.1 from 49r1) - winter 25



Impact of bounding on Precip:

Snowfall accumulation

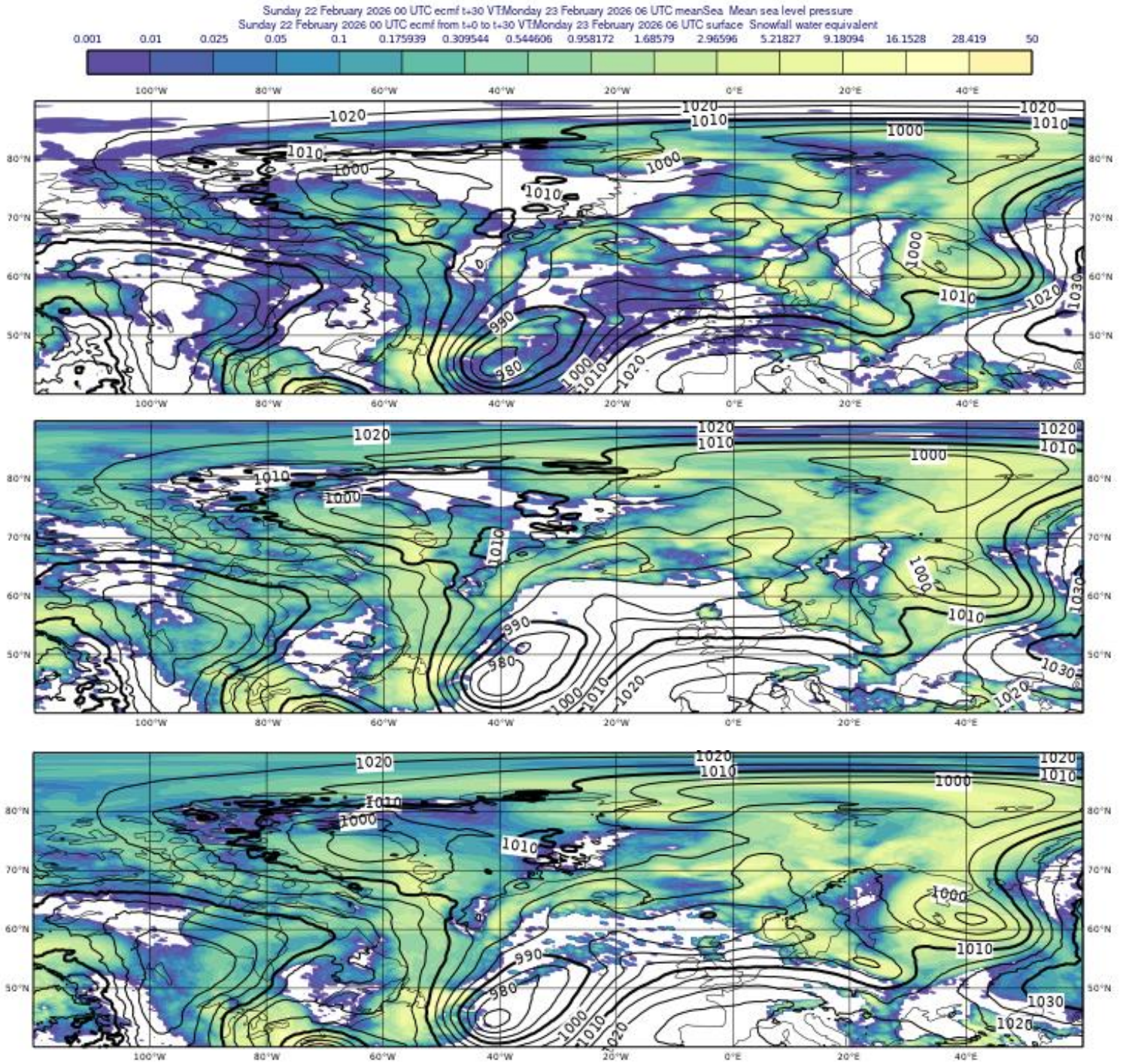
Perturbed member 2

Less spurious small values in many precip fields

v1

v2

IFS



AIFS v3 and beyond

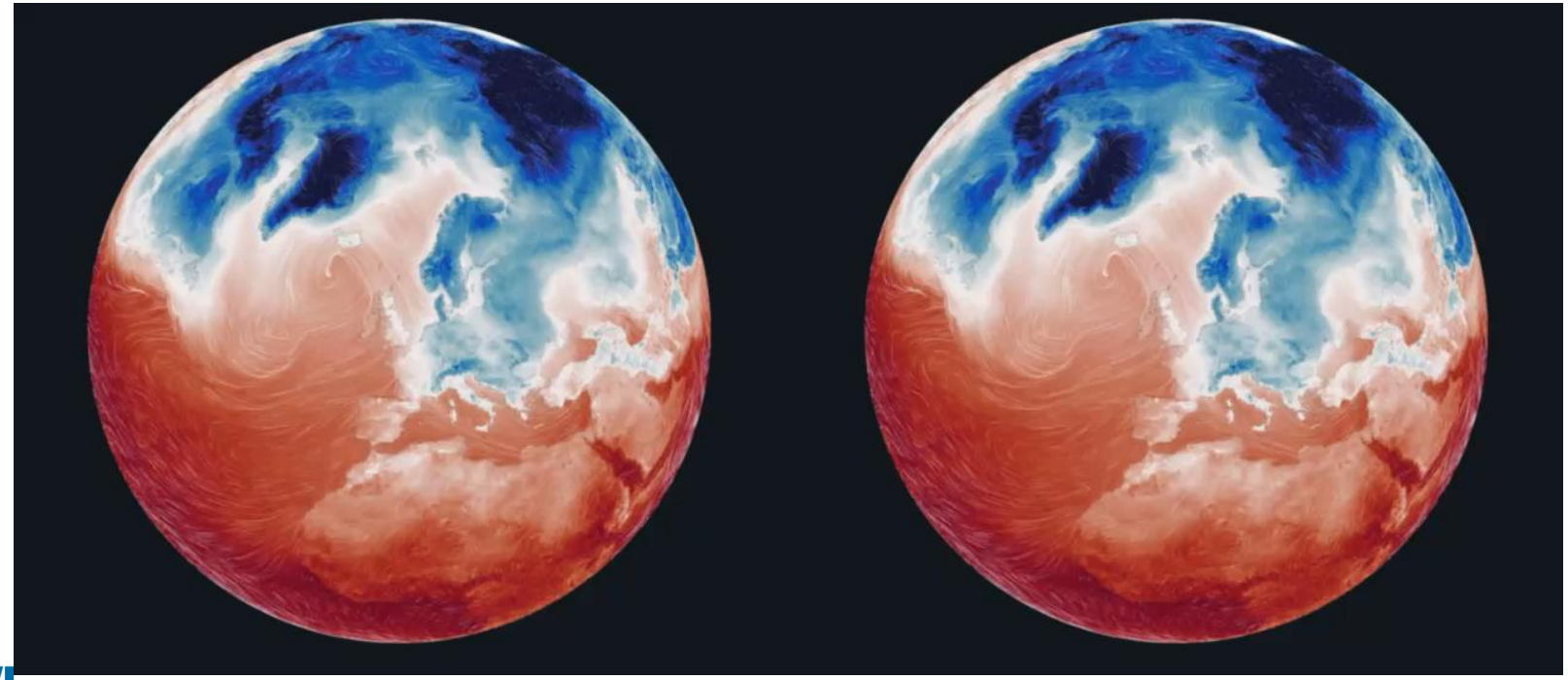
Prioritise Ensemble

More Earth System Components

Make use of observations

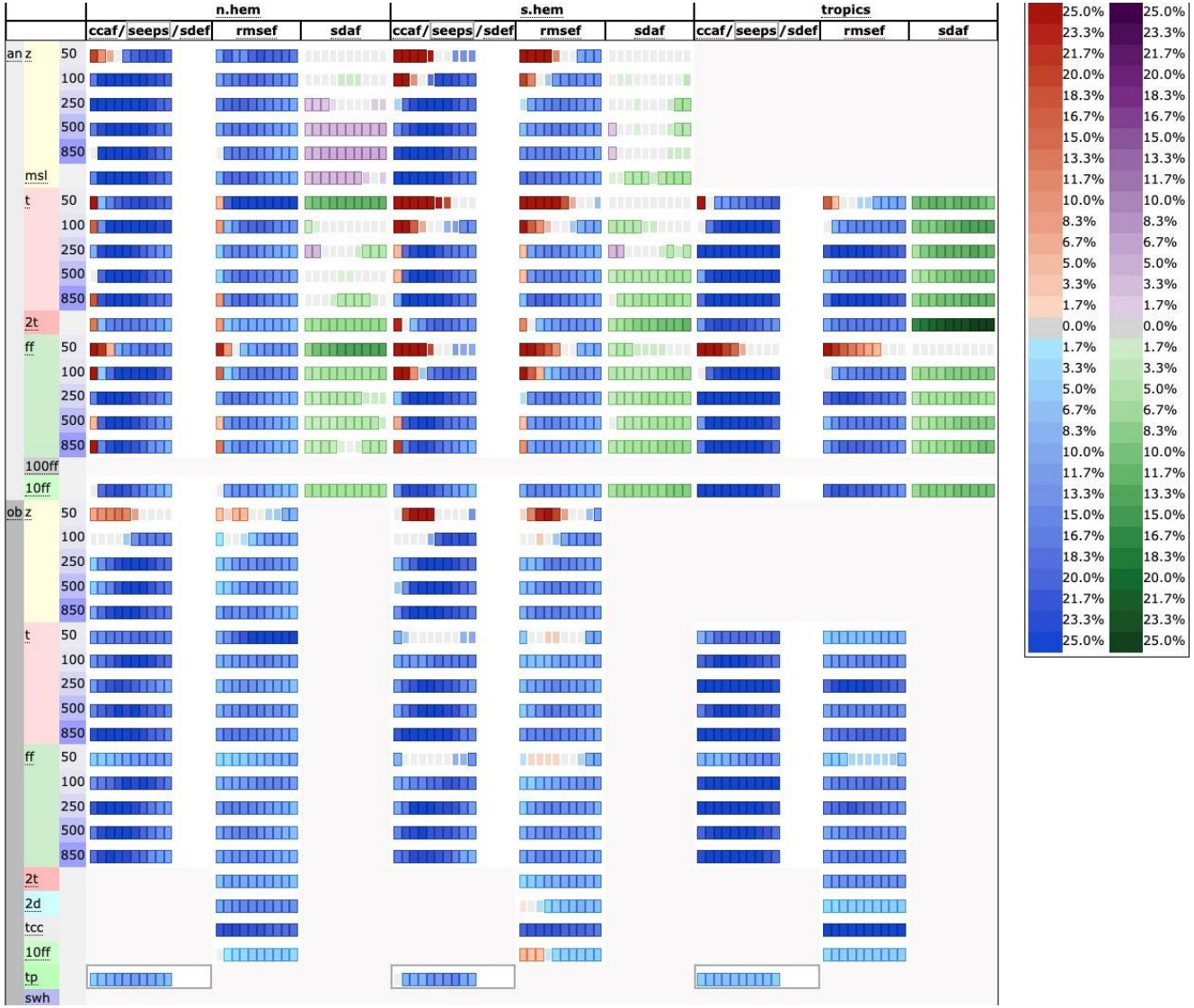
Higher time resolution

Higher spatial resolution

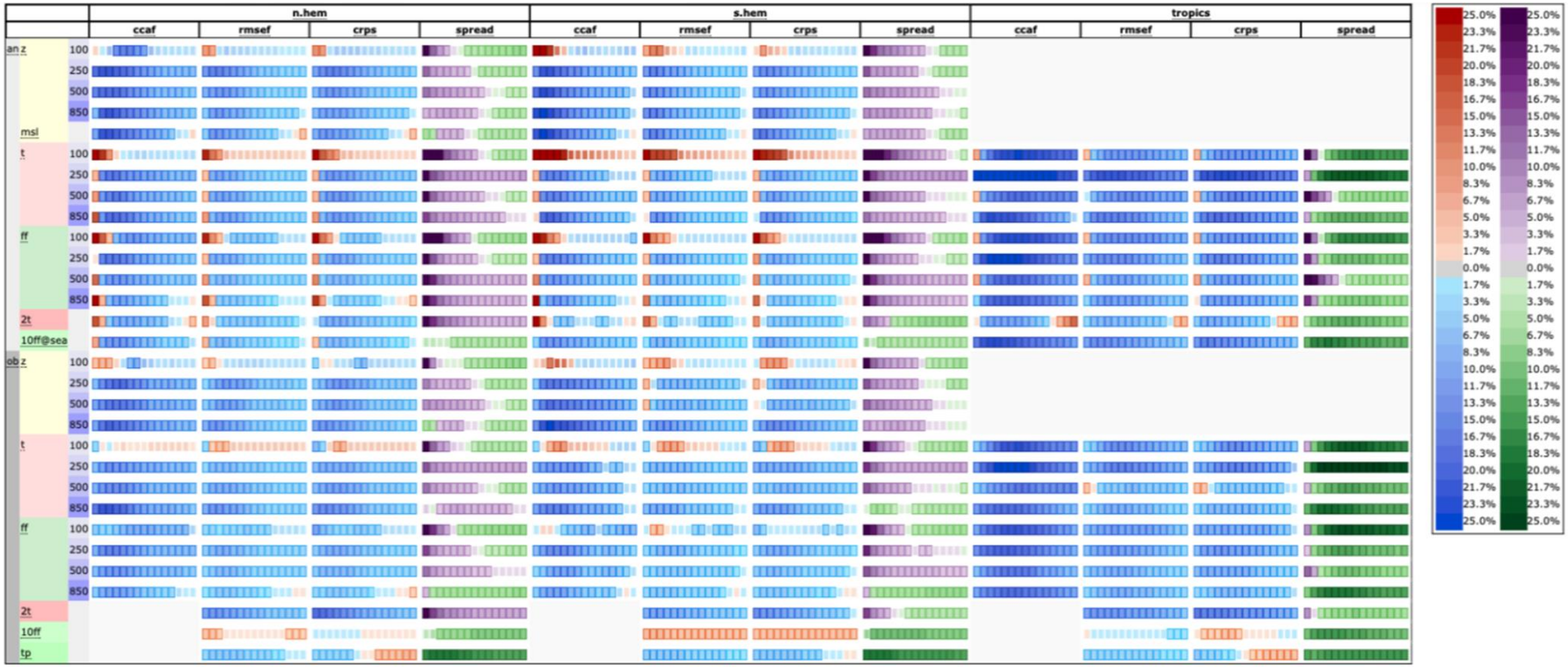


Are these models skillful?

AIFS Single vs IFS



AIFS ENS vs IFS



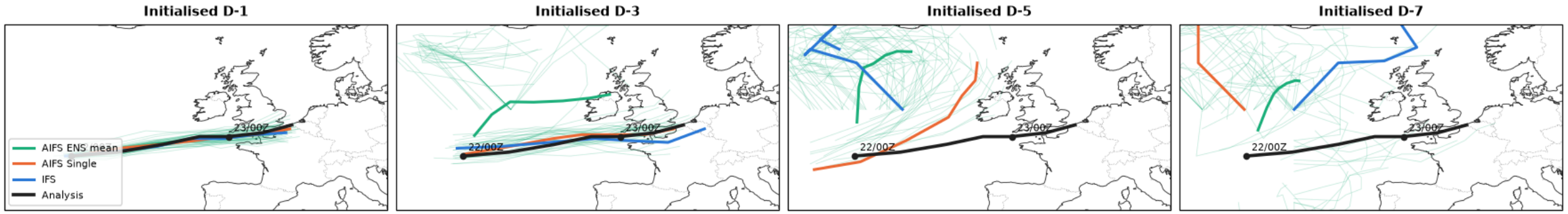
Are these models useful?

Some recent case studies

ECMWF

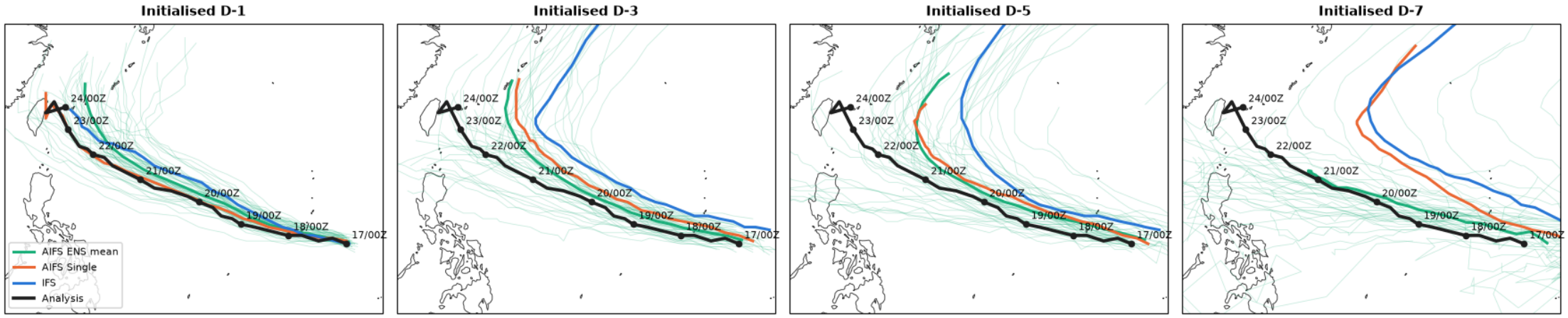
Storm Austen: 22-23rd October 2026

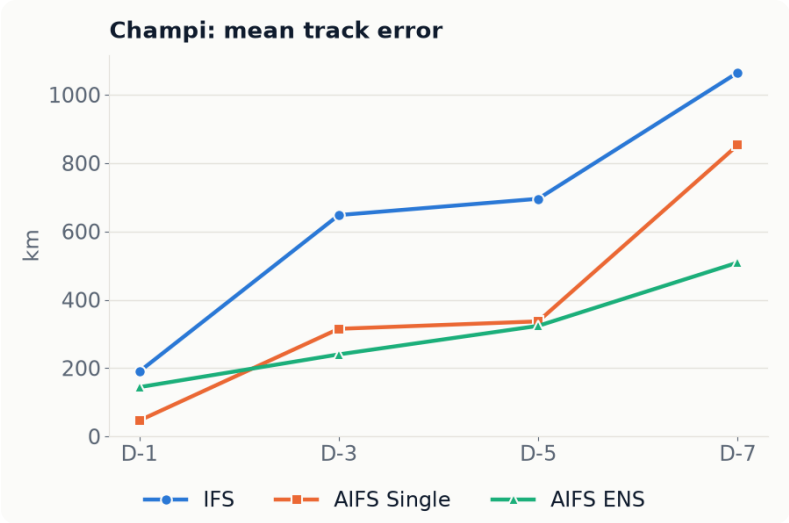
AIFS tracked the storm well, but its peak wind is 12–14 m/s too low



Model	Track, day 1	Track, day 3	Min. pressure	Peak wind
Analysis	–	–	974 hPa	39 m/s
IFS	76 km	114 km	978 hPa	40 m/s
AIFS Single	45 km	84 km	978 hPa	26 m/s
AIFS ENS	46 km	655 km*	979 hPa	25 m/s

Tropical Cyclone Champi: 17th-24th October 2026

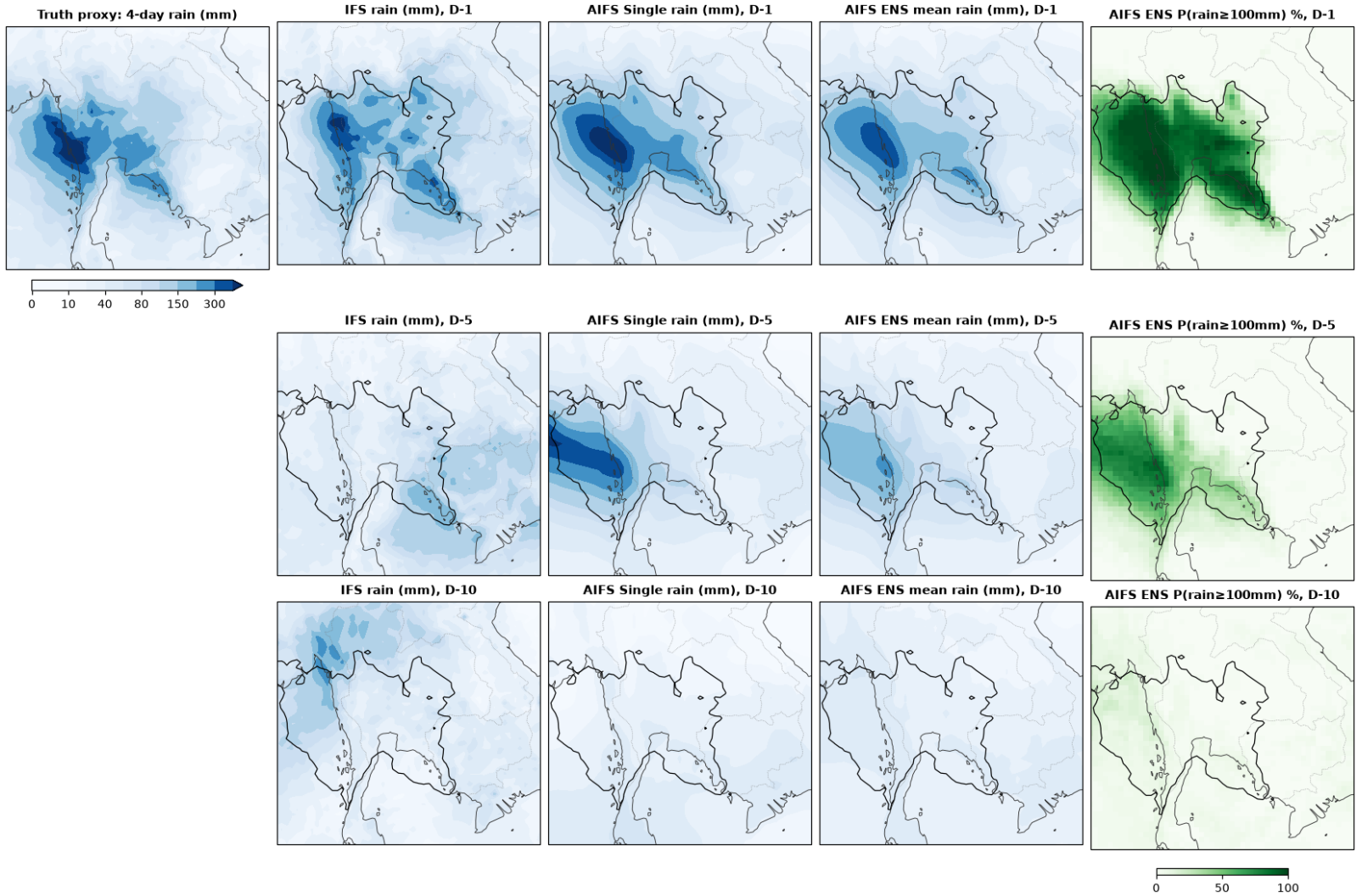




Model	Day 1	Day 7
Analysis	944 hPa · 45 m/s	944 hPa · 45 m/s
IFS	889 hPa · 64 m/s	895 hPa · 62 m/s
AIFS Single	947 hPa · 32 m/s	960 hPa · 29 m/s
AIFS ENS	957 hPa · 29 m/s	980 hPa · 22 m/s

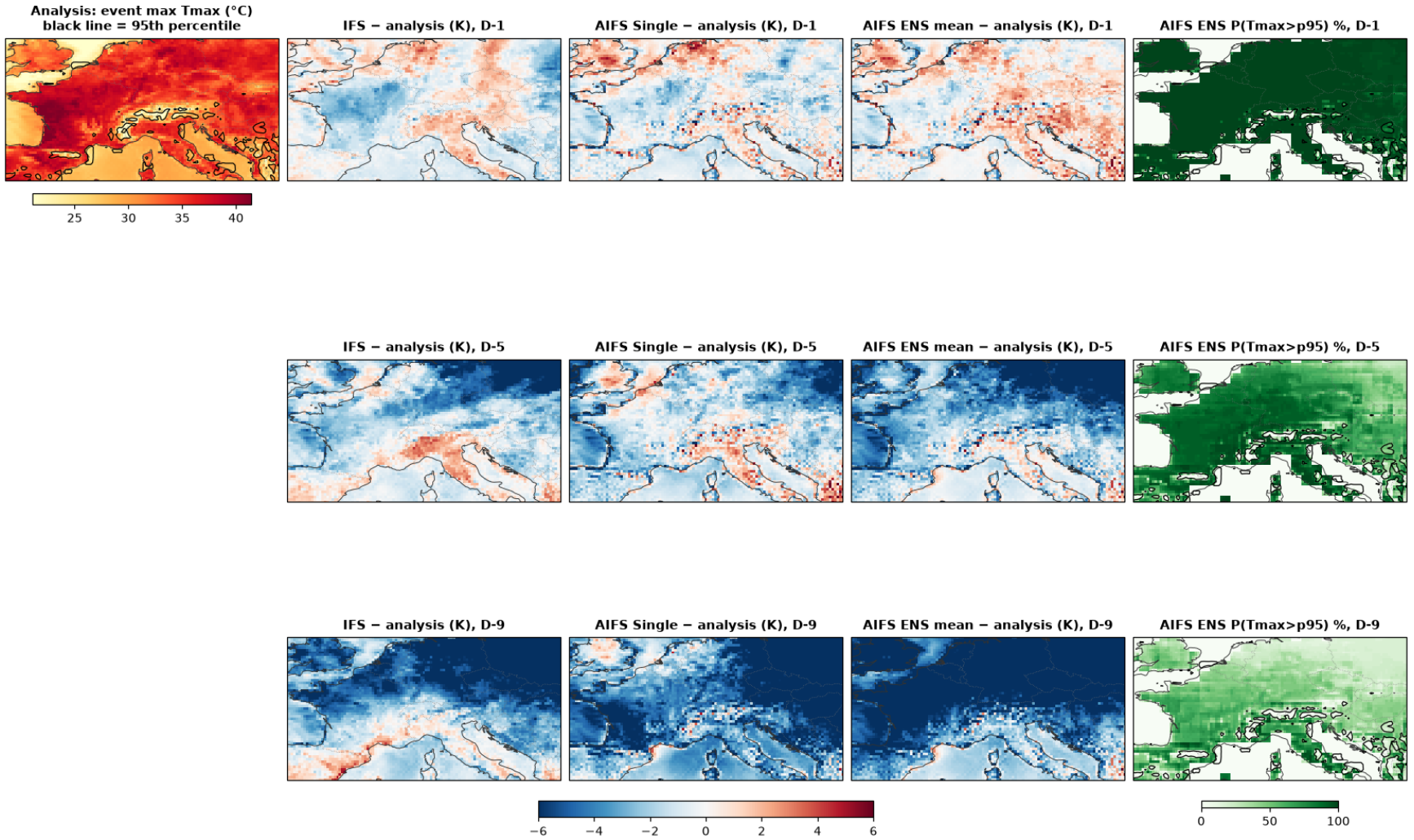
IFS forecasts a far stronger storm than AIFS whose winds stay near 30 m/s

Bangkok floods: 24th-28th October 2026



The flood is captured by all models at day 1, partly at day 3 in AIFS and mostly lost by day 5

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.

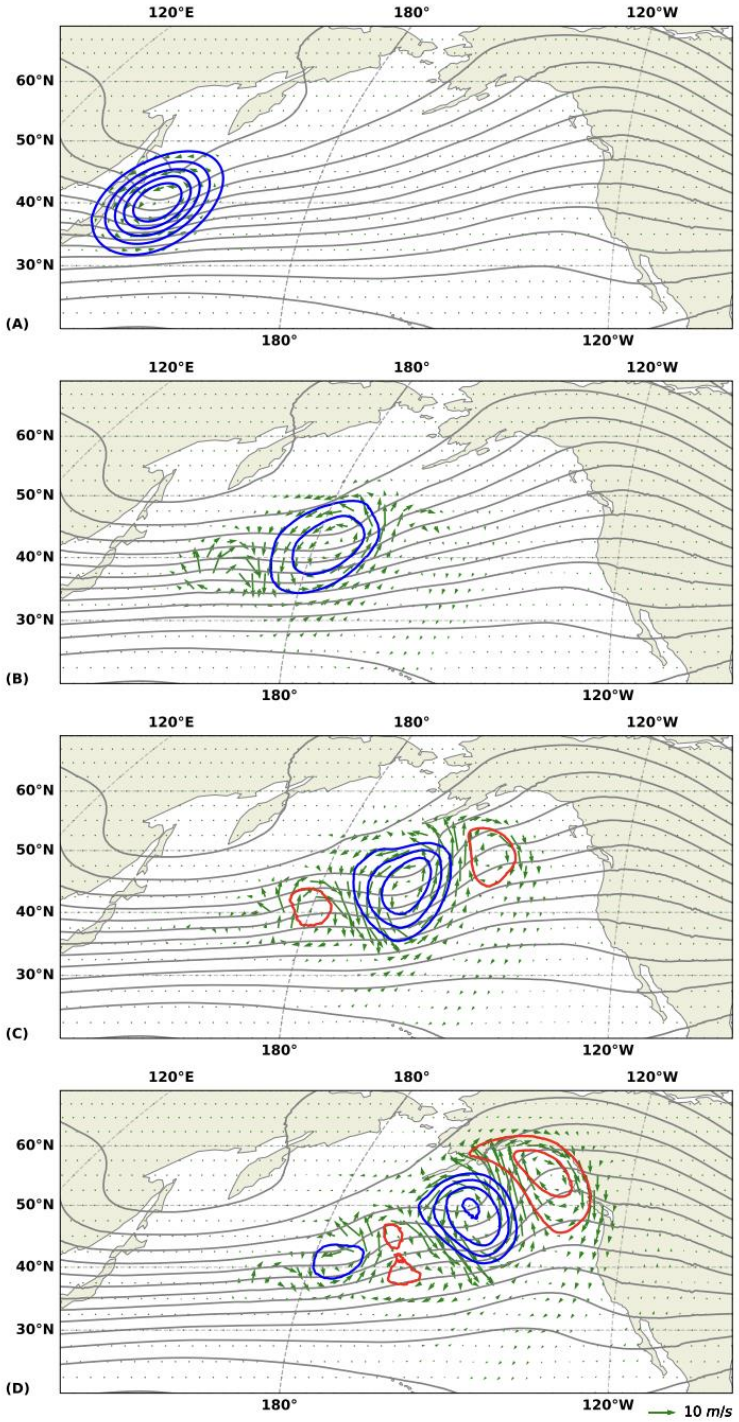
Are data-driven weather forecasts physical?

- Highly recommend reading Hakim & Masanam 2023:
  Dynamical Tests of a Deep-Learning Weather Prediction Model

- Take Pangu weather, and test it on a series of classical dynamical core test cases.
  - Cases need to be applied as deviations from climatology.
  - Apply localised disturbances and study the reaction of the system.

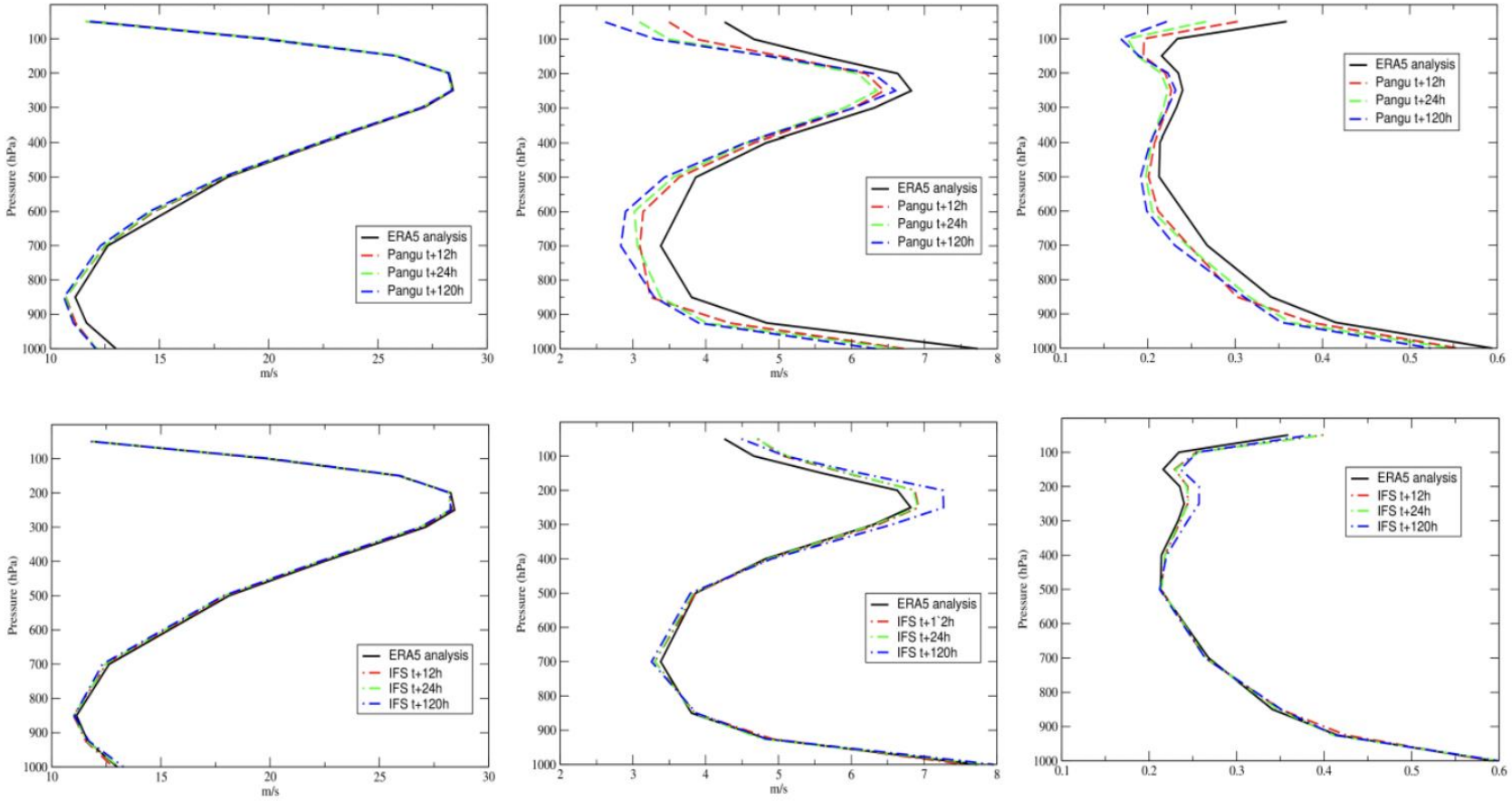
- Overall, Pangu behaves as expected, compared with theory.
  - The 1h model is best for the faster evolving processes, which aren't well captured by the longer timestep model.

- Hopefully we see lots more studies of this type.



Are data-driven weather forecasts physical?

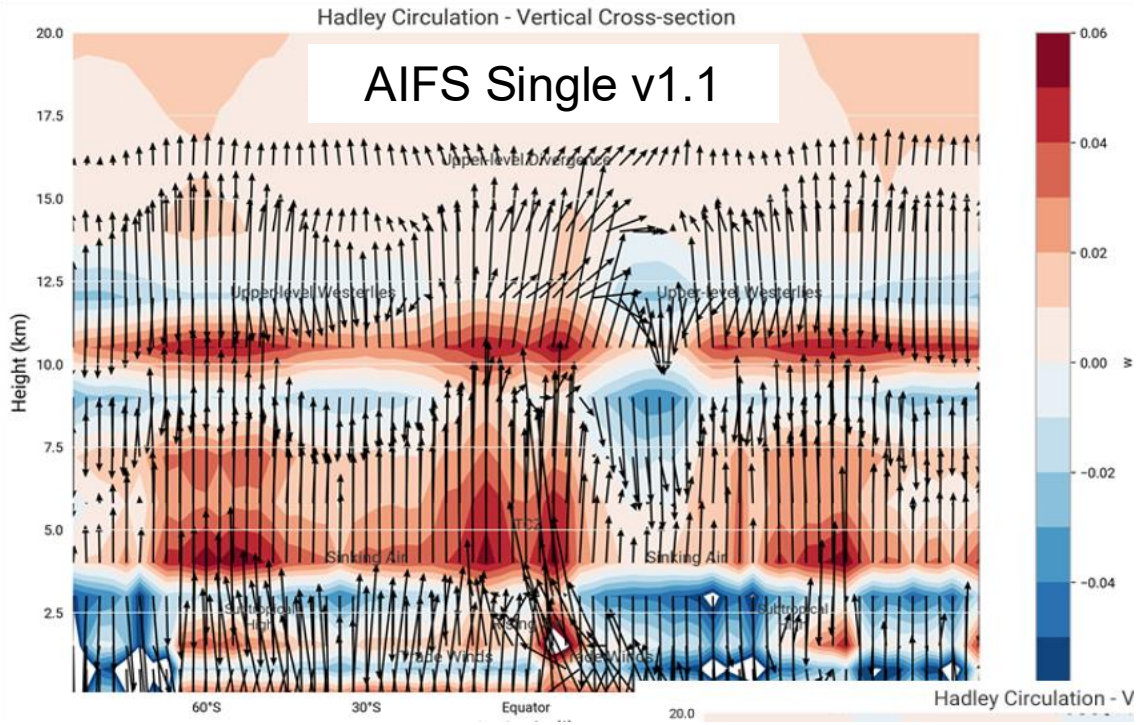
- Bonavita 2023 explore other tests with Pangu Weather.
- Geostrophic balance fairly well represented, but not as well as the IFS.

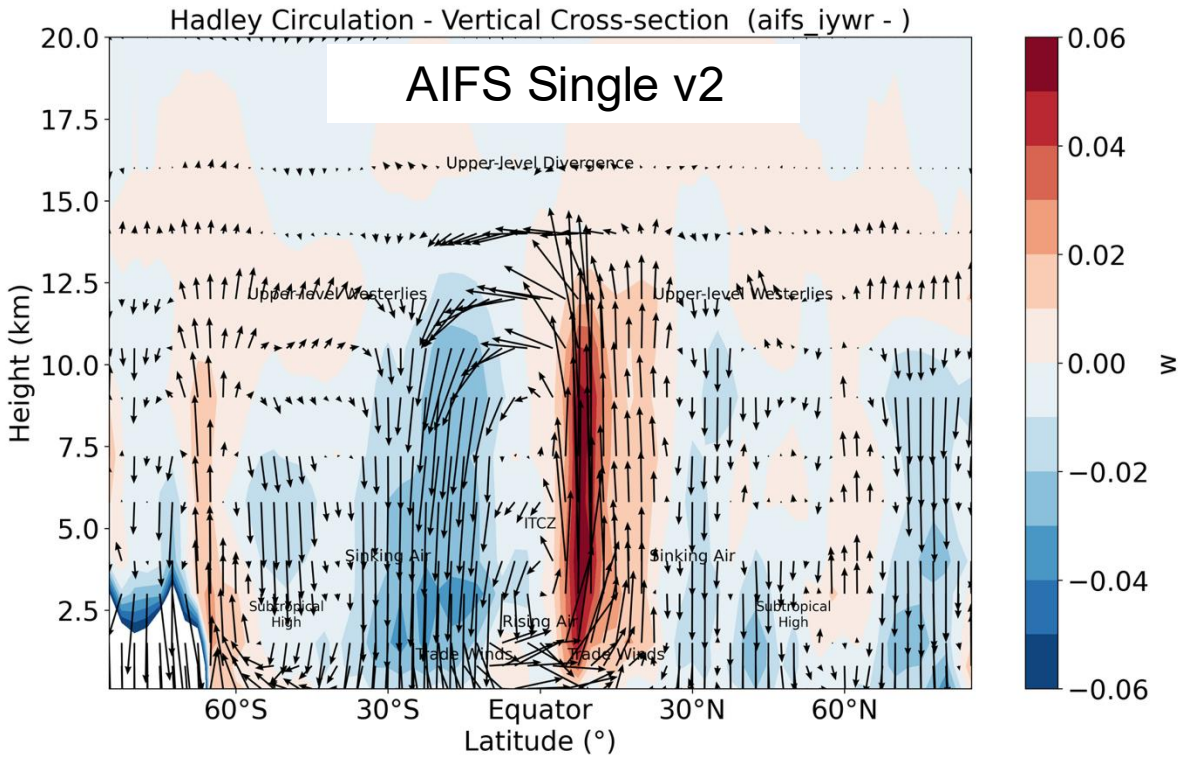


Geostrophic winds | Ageostrophic winds | Ratio

But we can also work on improving their physical realism

Hadley circulation now has correct strength and the convection cells are now properly formed

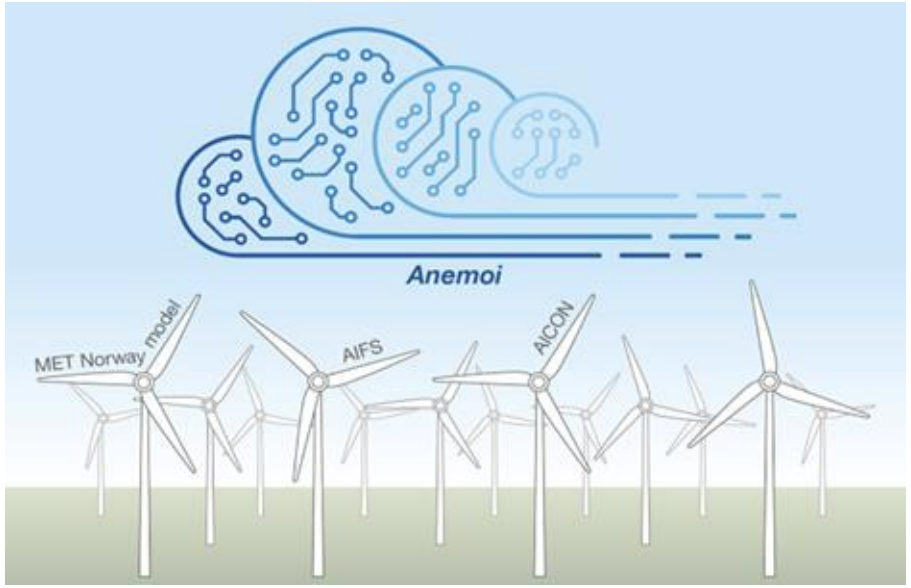




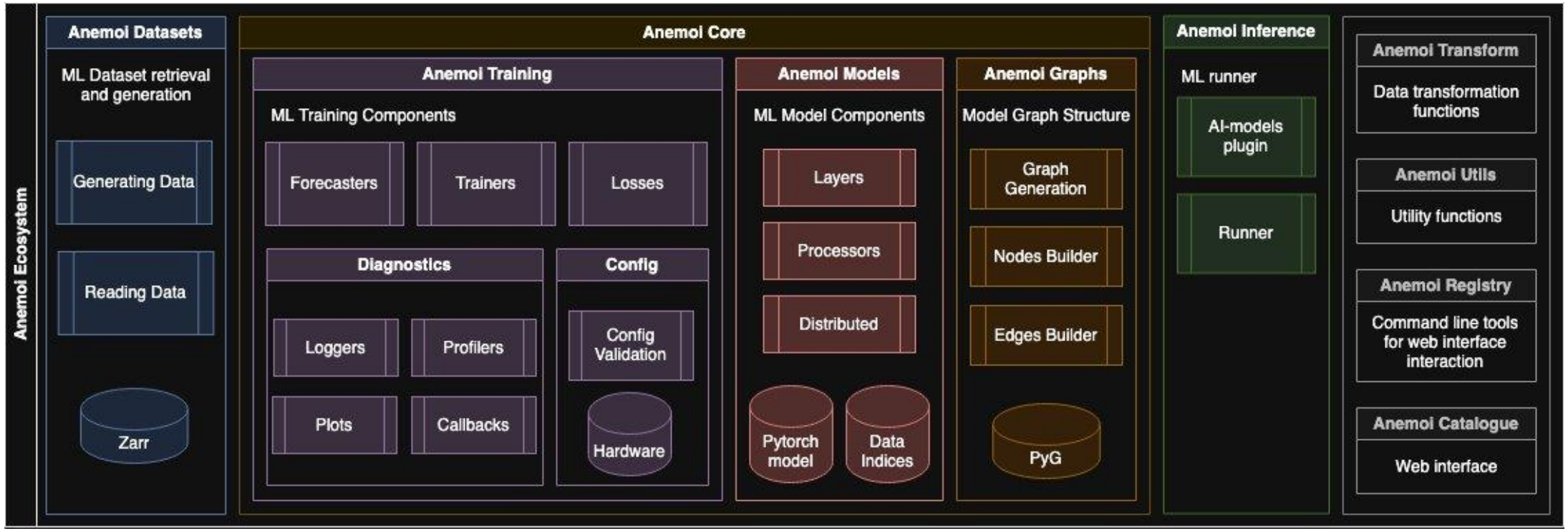
Anemoi: the tool behind it all

What is Anemoi?

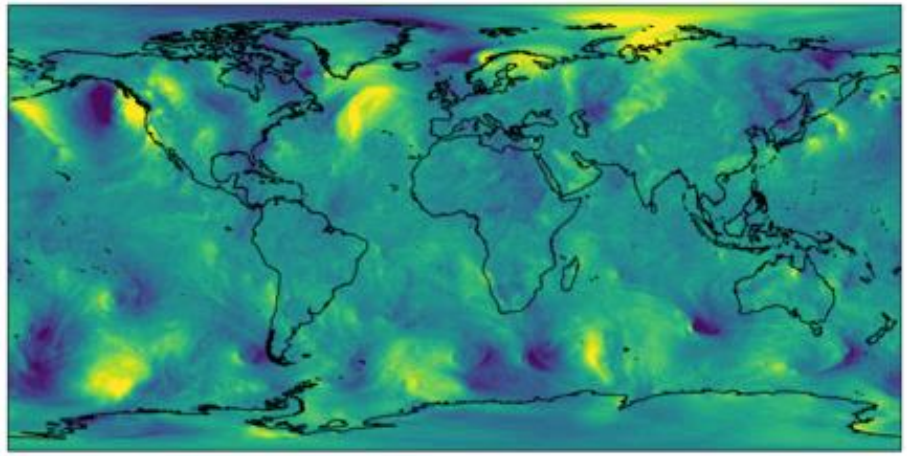
- An *open-source framework* that provides a complete toolkit to develop data-driven weather models – from data preparation through to inference
- Empowering *real-time running* and operationalisation of data-driven models at meteorological centres
- A *highly modular framework* organised into different Python packages



What is Anemoi?



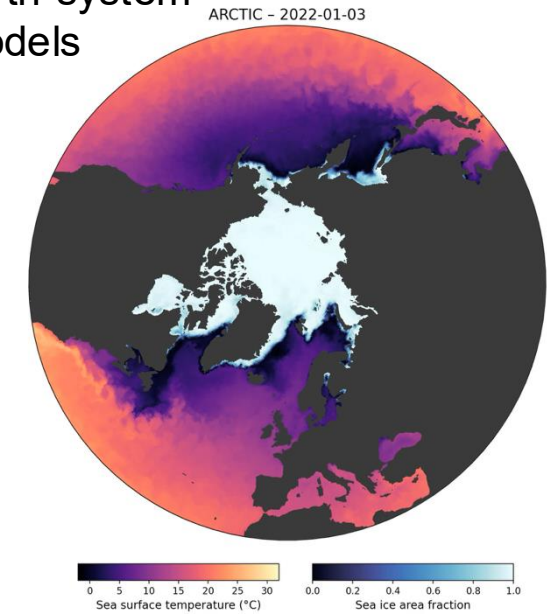
What Anemoi enables

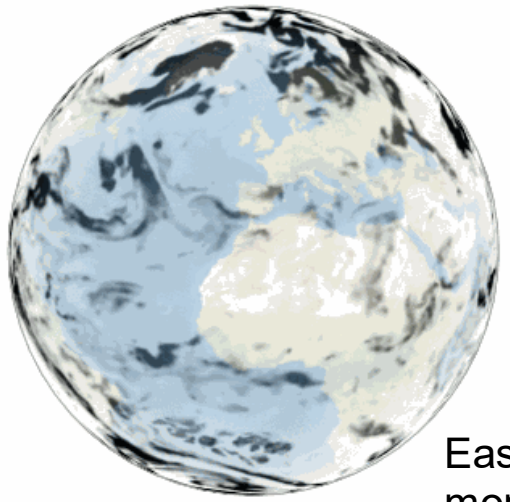


Training deterministic & ensemble models

Lang et al. 2024a & b

Building coupled earth-system models

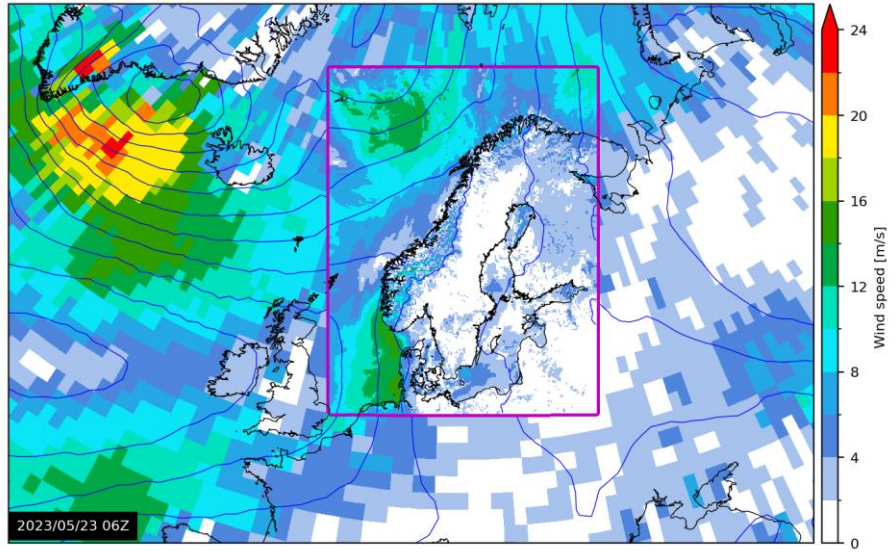




Easy exploration of more atmospheric fields

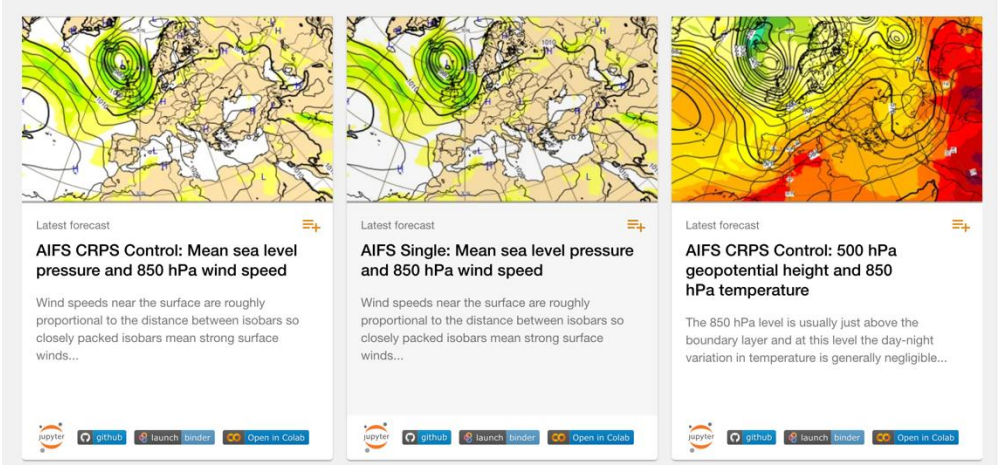
Regional high-resolution modelling:

Nipen et al. 2024



Real-time running at operation centres

Anemoi is powering ECMWF's operationalisation of AIFS in 2025.

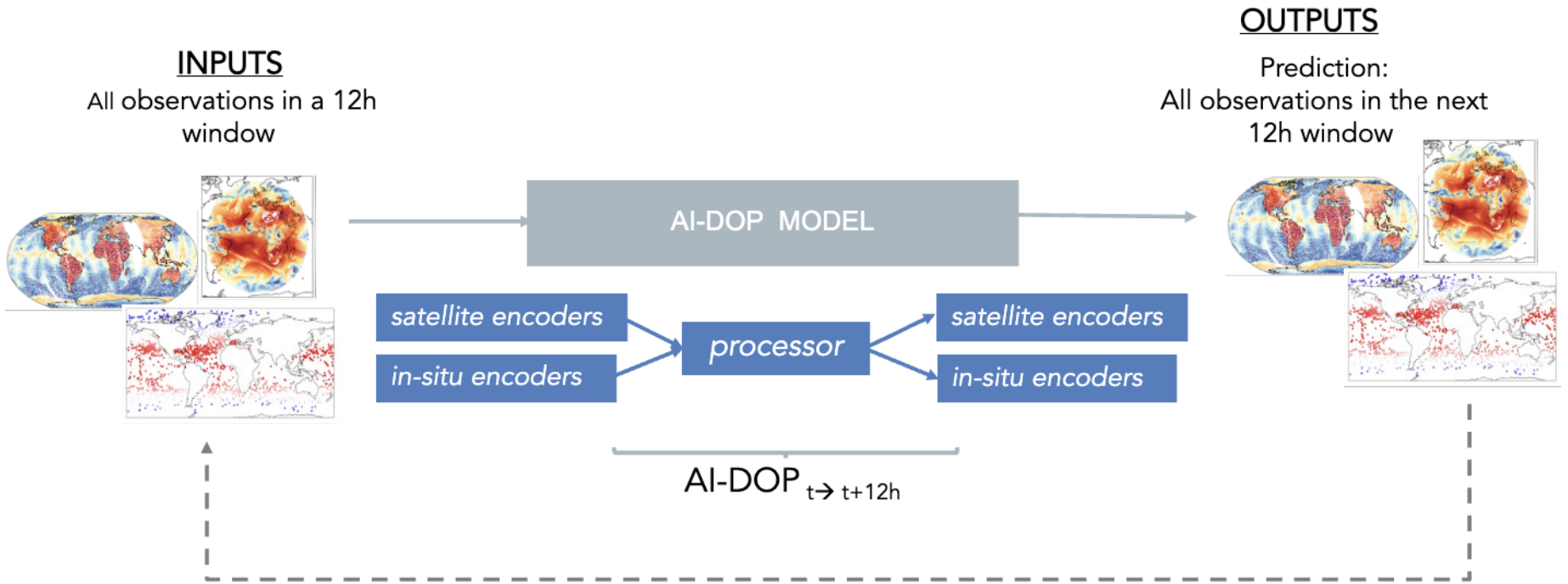


Learning directly from observations

AI-DOP: Emulating the entire chain using only observations

Alexe et al. 2024

- Use historical observations to train a Neural Network to forecast future observations (don't need analysis)
- Include all available observations of the full Earth system (atmosphere, ocean, land) simultaneously

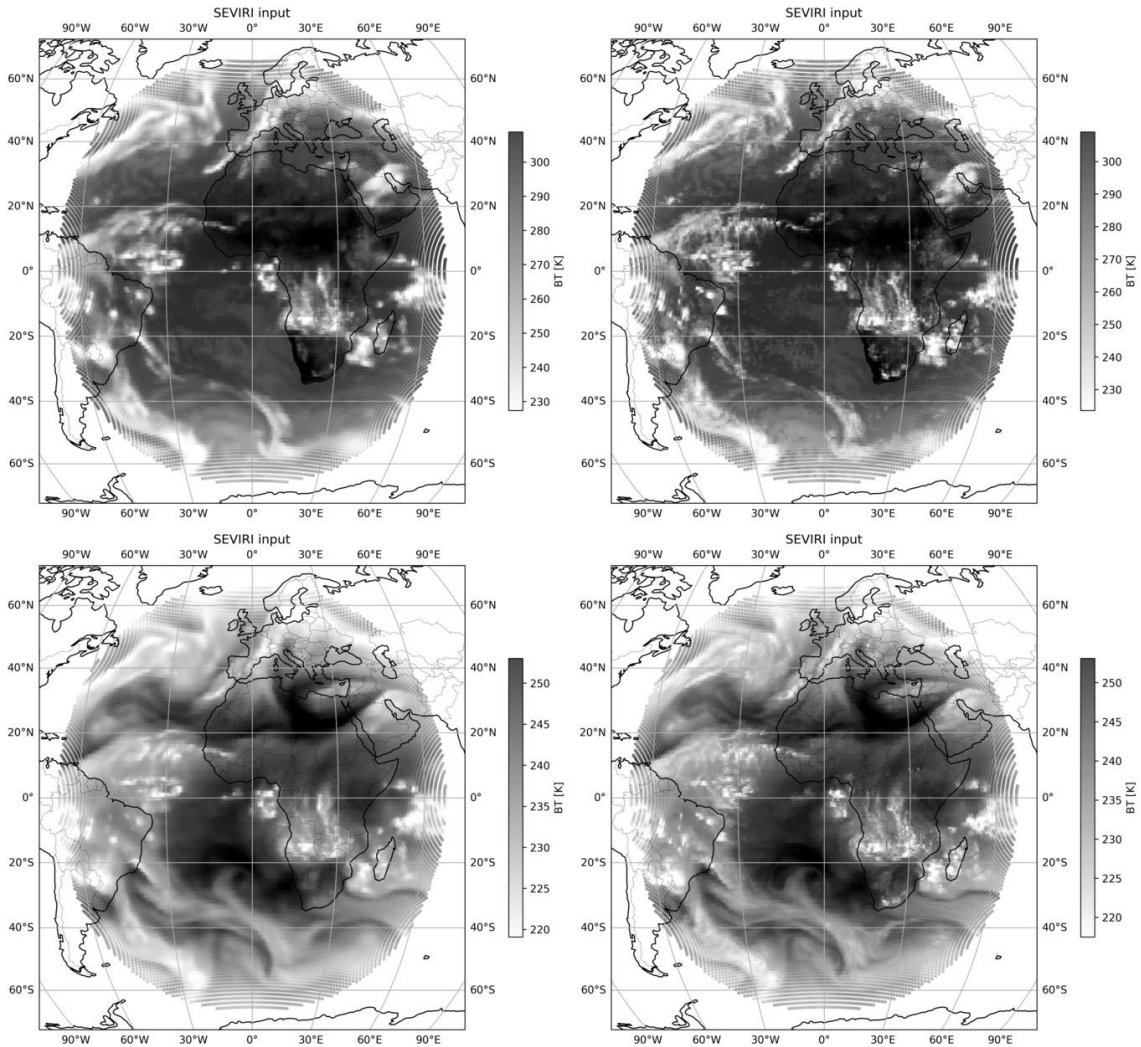


AI-DOP: Emulating the entire chain using only observations

Alexe et al. 2024

SEVIRI prediction

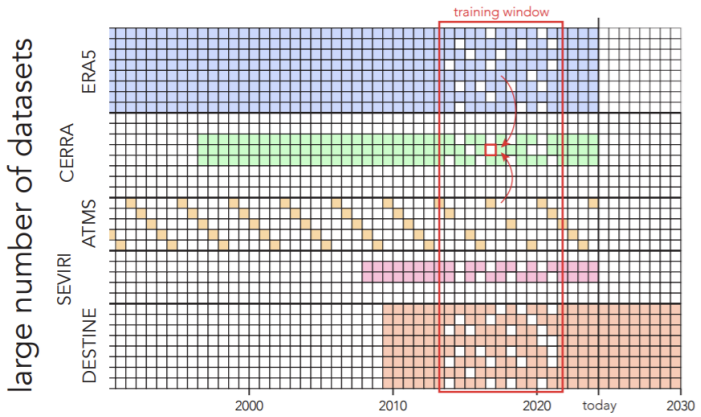
SEVIRI target



WeatherGenerator

A pan-European initiative to build a single open-source foundational model for weather and climate

pre-train



self-supervised training on petabytes of reanalysis, forecasts, observations

core model

$p_\theta(x, y)$

capture Earth system dynamics across scales in space and time

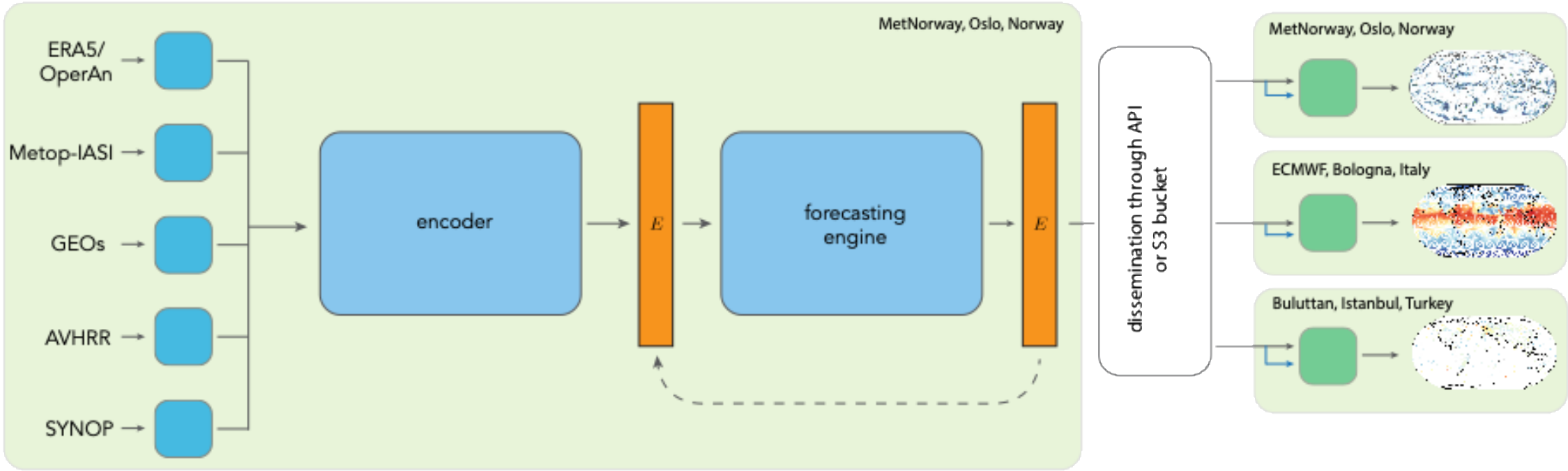
fine-tuning

- Forecasting across scales in space and time
- Downscaling
- Reanalysis
- Renewable energy output
- Extreme event risk
- ...

fine-tune to calibrate/adapt model to applications

Real-time WeatherGenerator

Met Norway is running WeatherGenerator in real-time since September



Summary

- Data-driven weather forecasting is a fast-evolving and exciting landscape
- With more advances to come:
  - Full Earth System models
  - Direct Observation Prediction models
  - Foundational models
  - ….