Data-driven weather forecasts

The future of weather forecasting?

Gabriel Moldovan

Thanks to the whole AIFS team

gabriel.moldovan@ecmwf.int

ECMWF

What are we going to explore?

What is data-driven forecasting?

What datasets are used?

How are data-driven models trained?

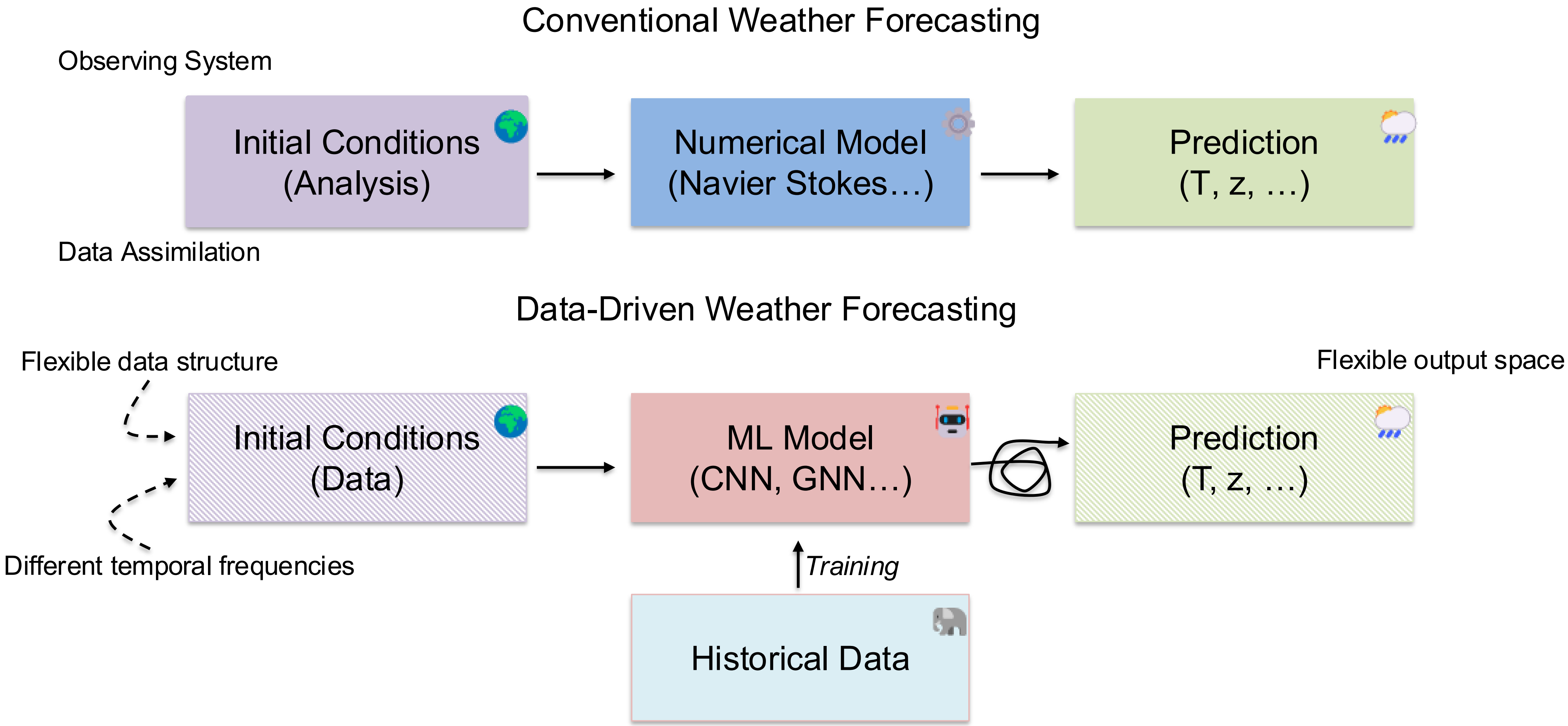
What architectures are used, and why?

What are the current strengths/weaknesses?

Are data-driven models physical?

Other approaches for ML forecasting

What is data-driven weather forecasting?



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(10^4) 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)…

Early 2023 Prototype AIFS developments begin

June 2023 ECMWF – ML project begins

October 2023 ECMWF – AIFS experimental forecasts live

Jan/Feb 2024 ECMWF – AIFS first updates

July 2024… ECMWF – First AIFS ENS experimental

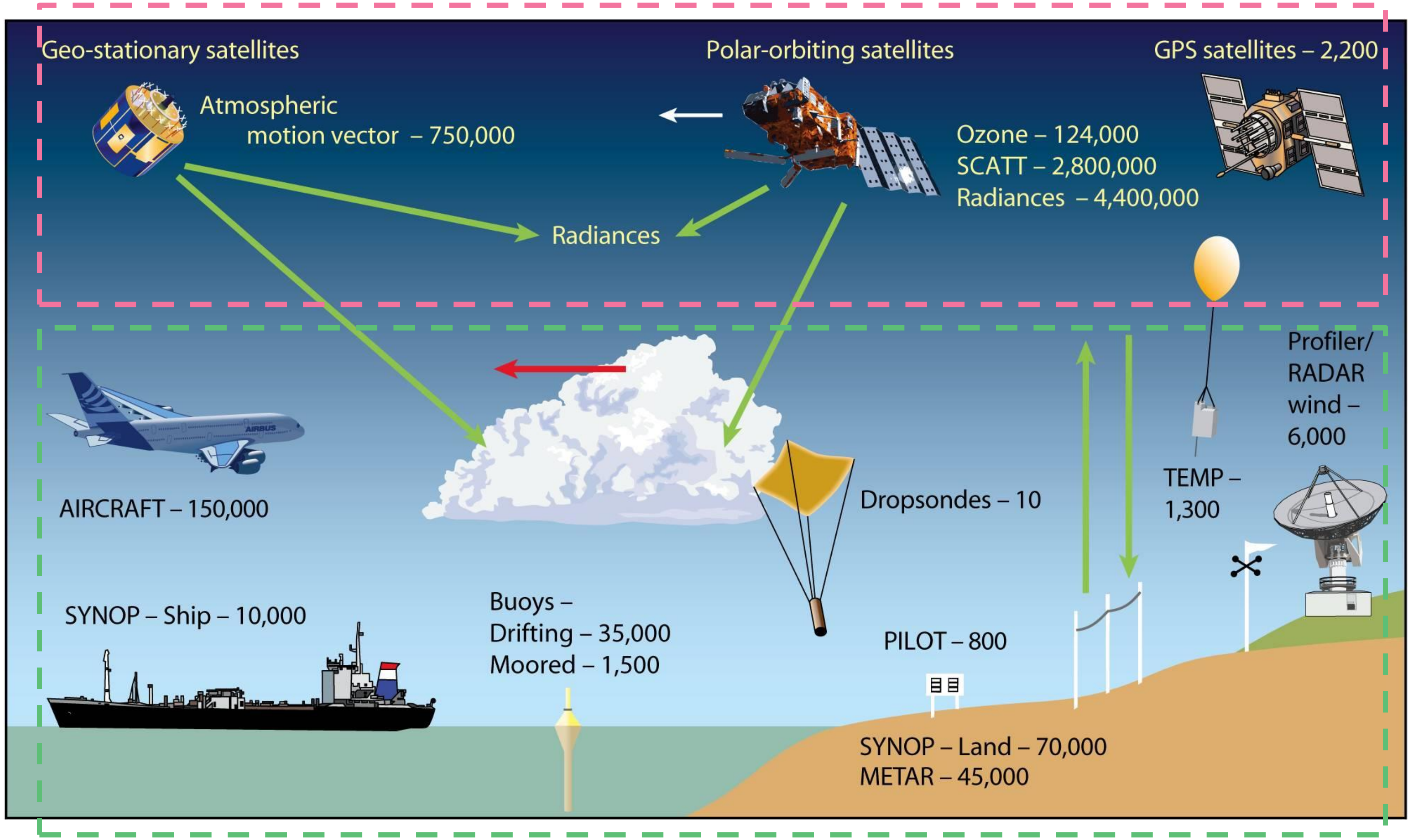
Feb 2025: ECMWF – AIFS Single 1 operational

July 2025: ECMWF – AIFS ENS operational

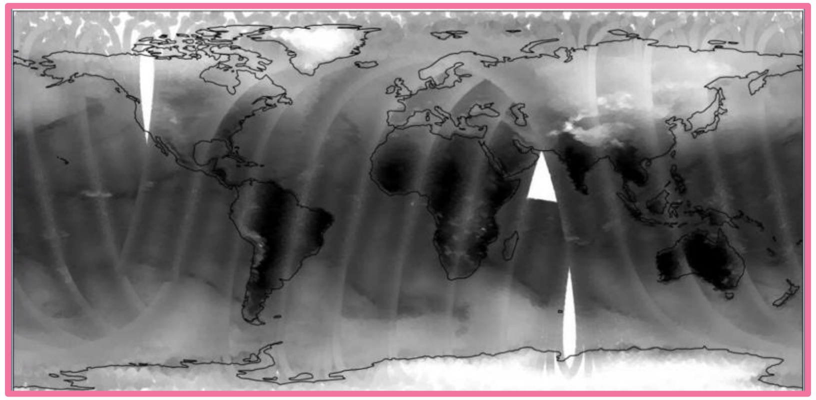
ECMWF

What data?

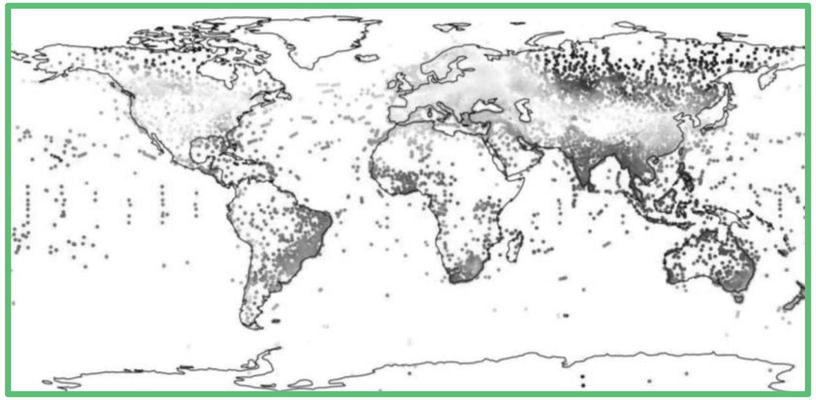
The observing system



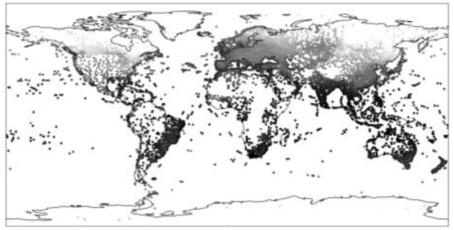
Satellite observations

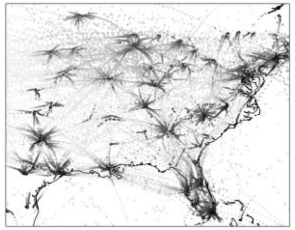


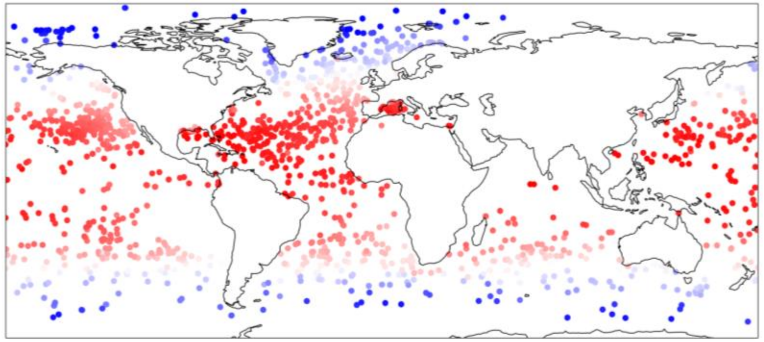
In-situ observations

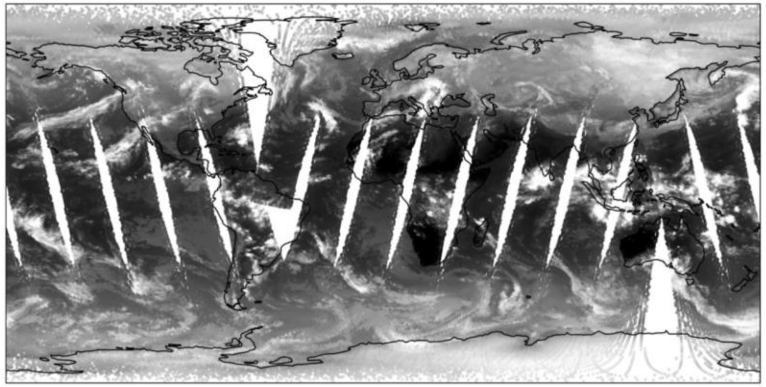


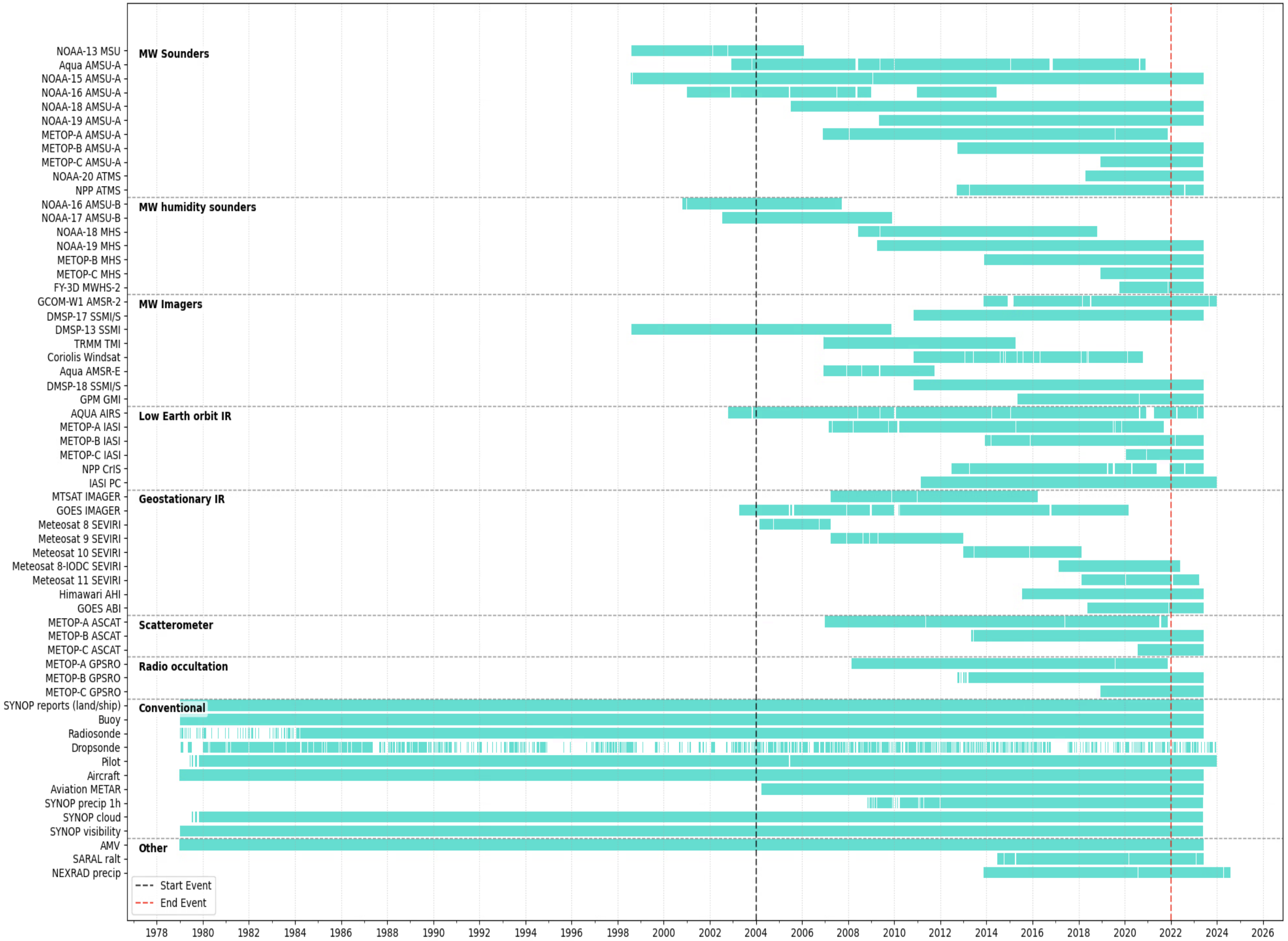
A complex system…





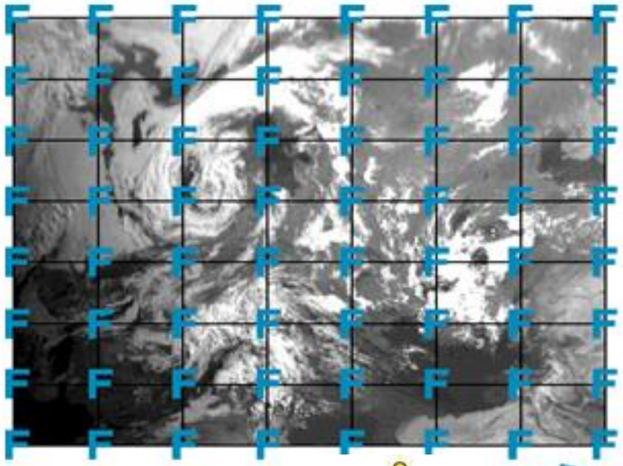




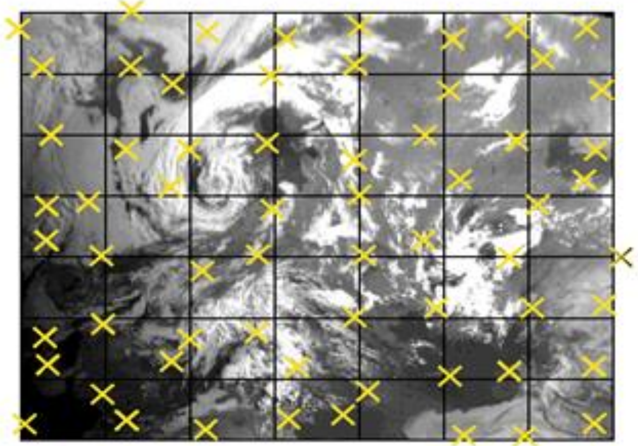


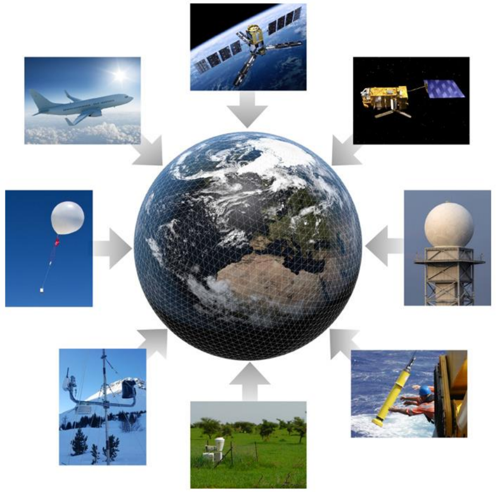
Data assimilation

Model Forecast (with errors)



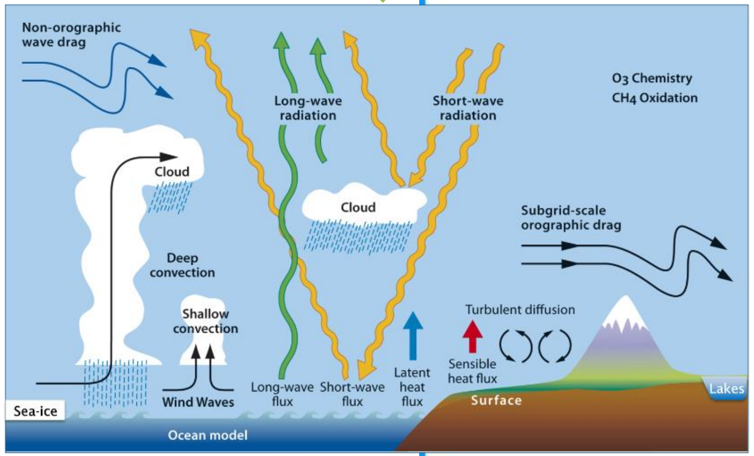
Observations (with errors)



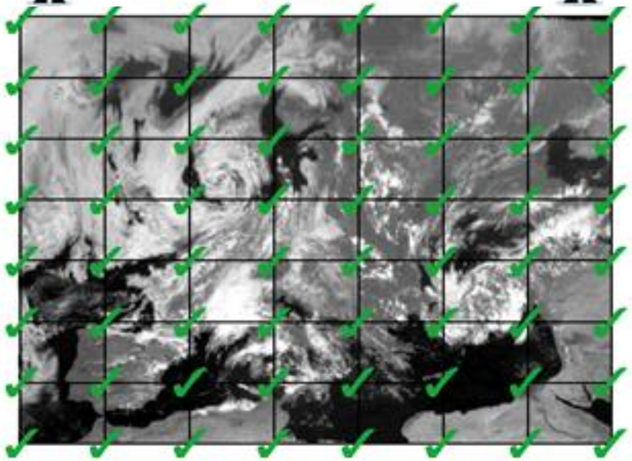


Earth system modelling needs accurate information from a wide range of observing systems, including satellites, radar, ocean floats and buoys, weather stations, weather balloons and aircraft. (Image credits below)

Physical Models



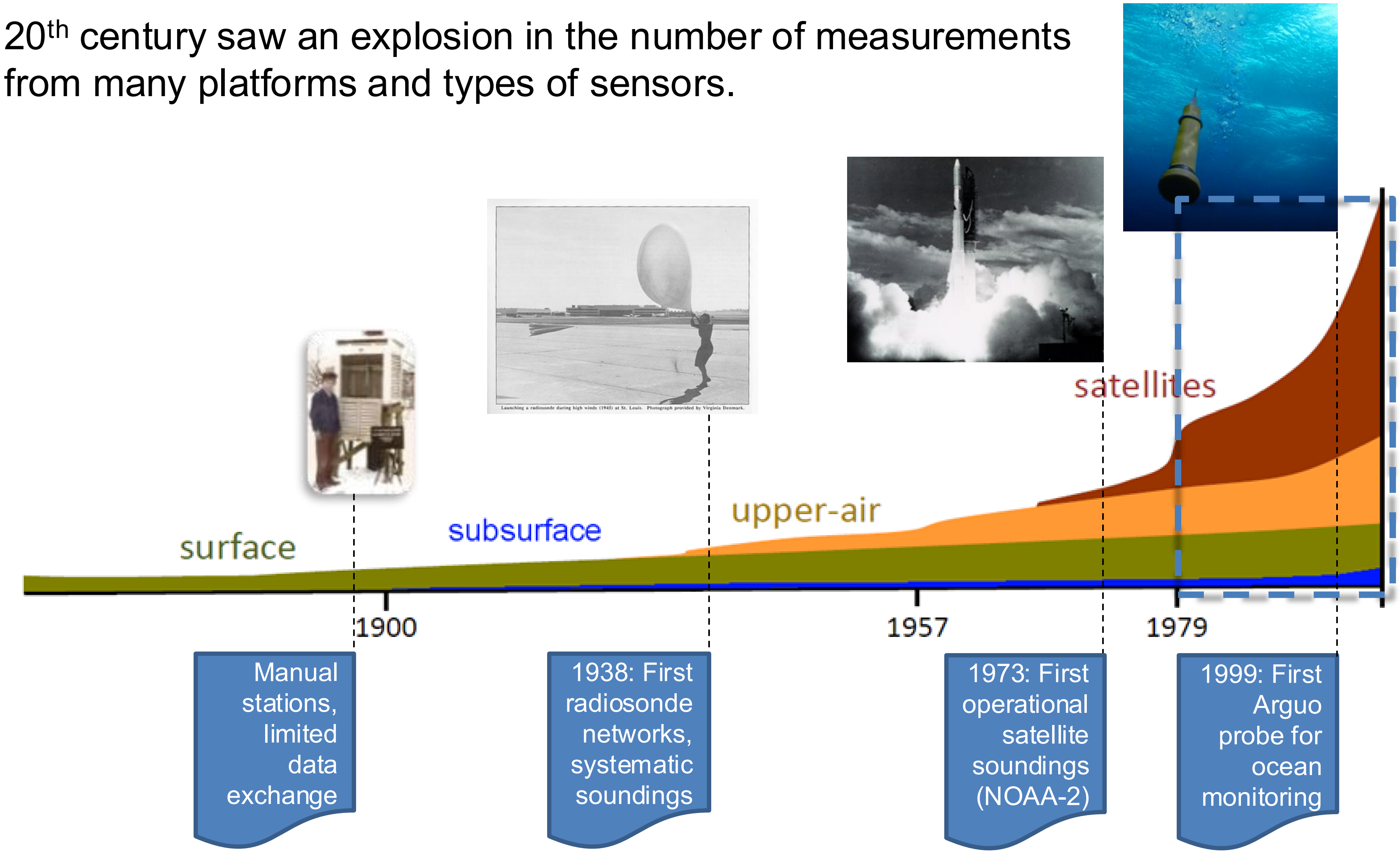
$$J(\mathbf{x}) = \underbrace{(\mathbf{x}^{\mathrm{b}} - \mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}^{\mathrm{b}} - \mathbf{x})}_{\text{prior state constraints}} + \underbrace{(\mathbf{y} - \mathbf{h}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}))}_{\text{observational constraints}}$$



Analysis (with smaller errors)

Choosing the Right Data: From Observations to Reanalysis: **ERA5**

20th century saw an explosion in the number of measurements from many platforms and types of sensors.

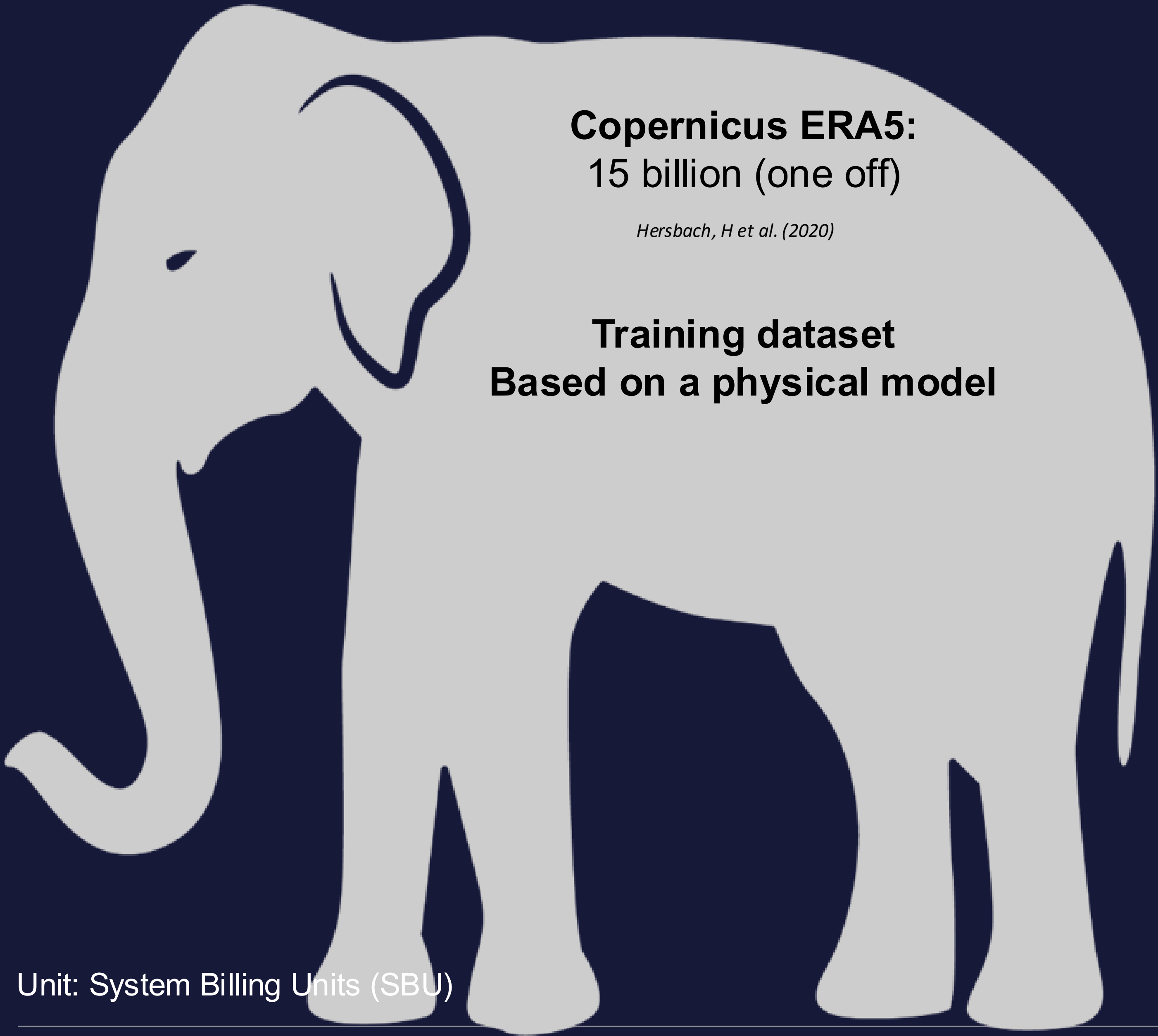


🌍 **Reanalysis (IFS Cycle 41r2)**
1940→, *consistent* ✅

⚙️ **Operational (IFS HRES)**
2016→, *inconsistent* ⚠️

ERA5 provides **accurate**, **reliable**, and **easy-to-access** data, the perfect foundation for machine learning exploration.

Operational analysis data are used for **fine-tuning** and **model operationalisation**, as AI models rely on them for initialization.



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)

Training setup

Variables?

Training task?

How to handle time?

Loss function?

What variables to use?

- Two driving motivations:
  - What helps me predict better.
  - What do users want from the system.
    - If high quality data exists, then it can be added directly to the training…

- Typical set used by many models:
  - ~13 pressure level (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, z
    - But no direct cloud information.
  - At the surface: 2t, msl/sp, 10u/v and precipitation for some.

Variables used in AIFS1.1.0

A mix of **prognostic**, **diagnostic**, and **forcing** variables:

Upper Atmosphere (13 pressure levels)

- Geopotential height
- Wind components (u, v)
- Specific humidity
- Temperature

Surface Variables

- 2 m temperature
- 10 m wind speed
- Precipitation

External Forcings

- Orography
- Insolation
- Latitude / Longitude
- Time of day / Time of year

Variable name	Short name	Level type Pressure level (50-1000 hPa) or Surface	Variable type: Prognostic, Diagnostic, Forcing	Normalization	Scaling
Geopotential	z	Pl	P	Z-score	12
Horizontal wind components	u, v	Pl	P	Z-score	0.8, 0.5
Specific humidity	q	Pl	P	Std	0.6
Temperature	t	Pl	P	Z-score	6
Surface pressure	sp	S	P	Z-score	10
Mean sea-level pressure	msl	S	P	Z-score	1
Skin temperature	skt	S	P	Z-score	1
2 m temperature	2t	S	P	Z-score	1
2 m dewpoint temperature	2d	S	P	Z-score	0.5
10 m horizontal wind components	10u, 10v	S	P	Z-score	0.5, 0.5
Total column water	tcw	S	P	Std	1
Volumetric soil water level 1 and 2*	swvl1, swvl2	S	P	None	1, 2
Soil temperature level 1 and 2*	stl1, stl2	S	P	None	1, 10
Total precipitation	tp	S	D	Std	0.025
Convective precipitation	cp	S	D	Std (tp)	0.0025
Snowfall*	sf	S	D	Std (tp)	0.025
Total cloud cover*	tcc	S	D	None	0.1
High cloud cover*	hcc	S	D	None	0.1
Medium cloud cover*	mcc	S	D	None	0.1
Low cloud cover*	lcc	S	D	None	0.1
Runoff*	ro	S	D	Std	0.005
Surface solar radiation downwards*	ssrd	S	D	Std	0.05
Surface thermal radiation downwards*	strd	S	D	Z-score	0.1
100 m horizontal wind components*	100u, 100v	S	D	Z-score	0.1, 0.1
Land-sea mask	lsm	S	F	None	
Orography	z	S	F	Max	
Standard deviation of sub-grid orography	sdor	S	F	Max	
Slope of sub-scale orography	slor	S	F	Max	
Insolation	insolation	S	F	None	
Latitude/longitude (cos/sin)	lat/lon	S	F	None	
Time of day/day of year	local time, julian day	S	F	None	

Table 1: Variables used in the training of AIFS, with their short names, level type, variable type, normalization method, and scaling factors. Variables marked with * were newly introduced compared to AIFS v0.2.1.

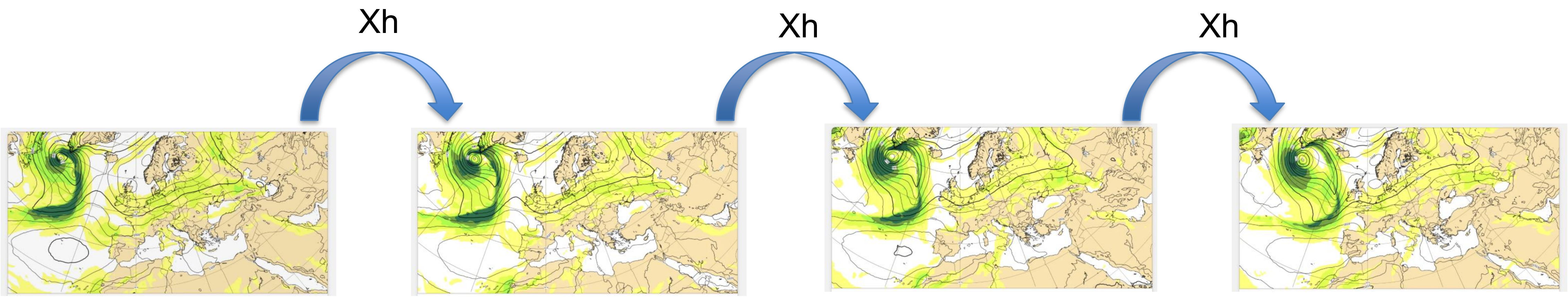
https://doi.org/10.5194/egusphere-2025-4716

Training Task?

Task:
Given ERA5 initial conditions $\mathbf{x}_t$,learn a mapping

$$f_\theta : \mathbf{x}_t \rightarrow \mathbf{x}_{t+\Delta t}$$

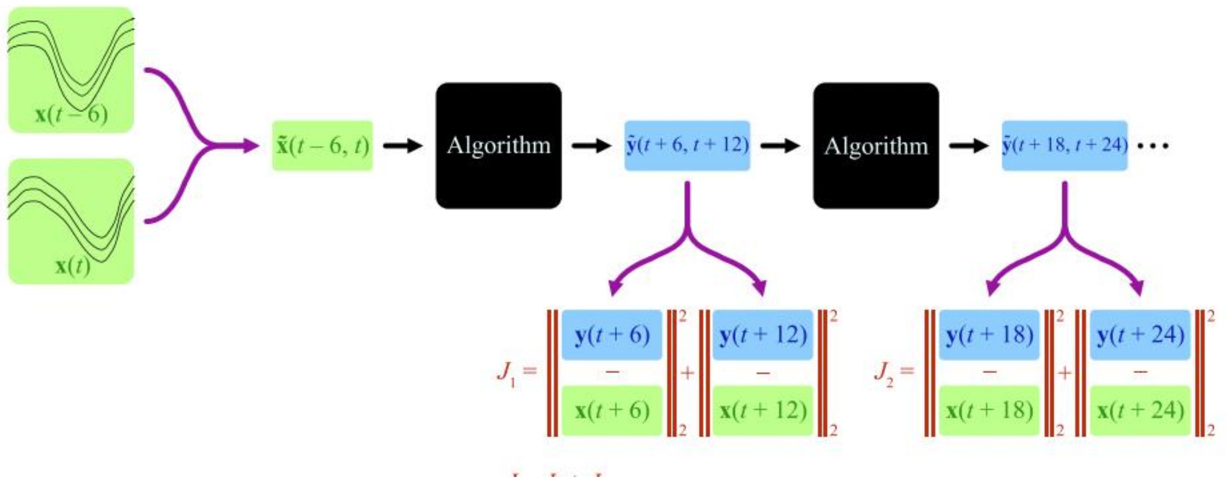
such that the predicted state $\hat{\mathbf{x}}_{t+\Delta t}$approximates the true ERA5 atmospheric state $\mathbf{x}_{t+\Delta t}$.
This corresponds to **learning a data-driven approximation of the atmospheric time-evolution operator**, analogous to solving the prognostic equations of NWP, but directly from reanalysis data.



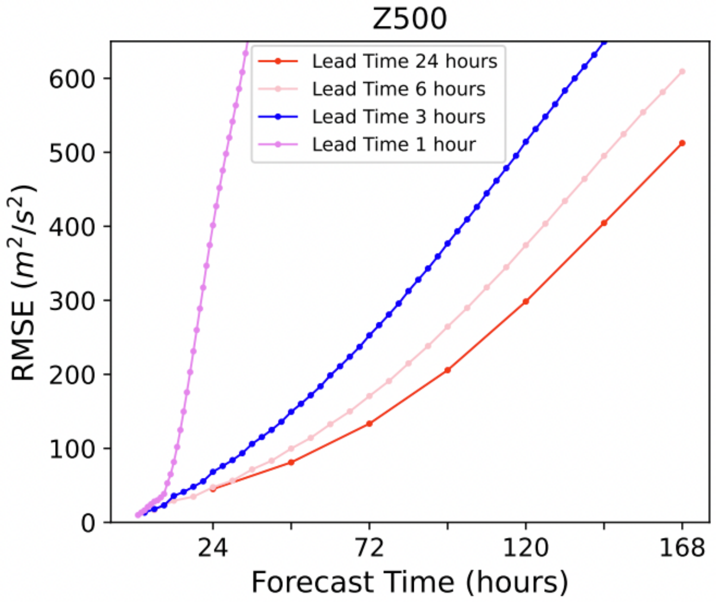
The forecast is then **autoregressively** stepping Xh into the future $x_n = f(x_{n-1})$ …

How to handle time?

- How many time slices to provide as input?
  - Weyn et al (2020) provide 2 time slices and get out the next two.
  - This is now used by many others.

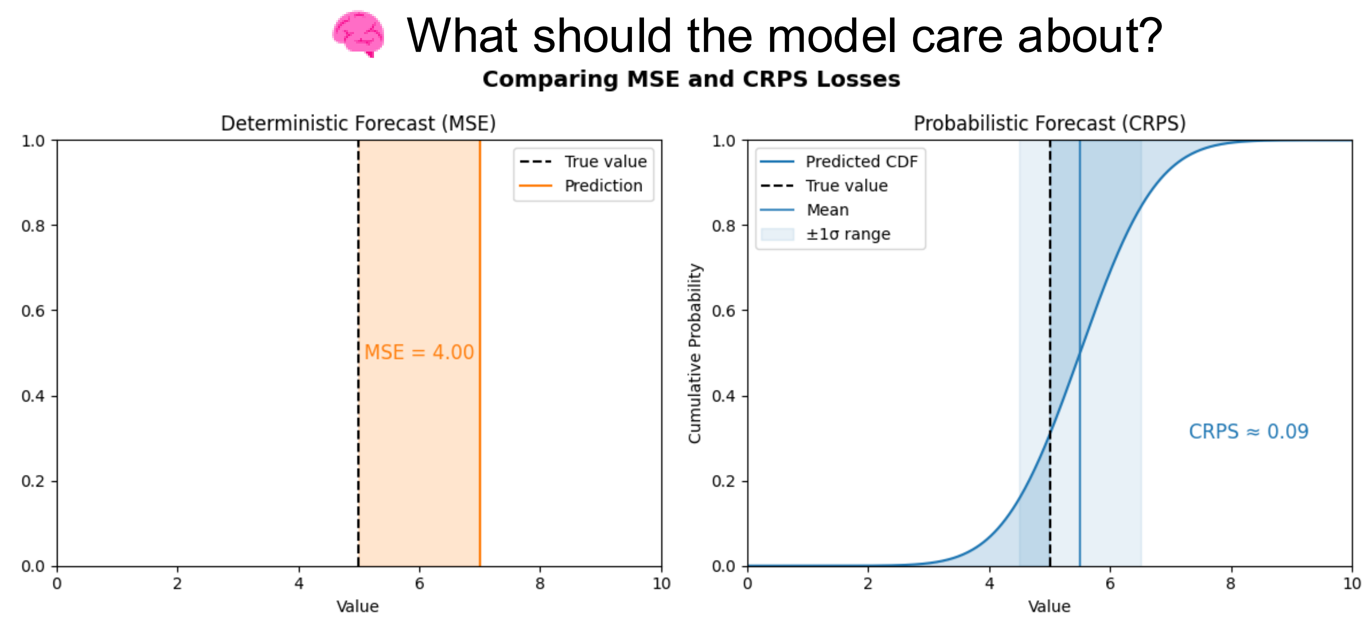


- How big of a timestep to make?
  - Early work tried 3-day steps but failed to compete with physics models.
  - Many choose 6-hours, but this limits the granularity of the output.
  - Pangu Weather created multiple models for different timesteps…
  - but this leads to inconsistencies in time.

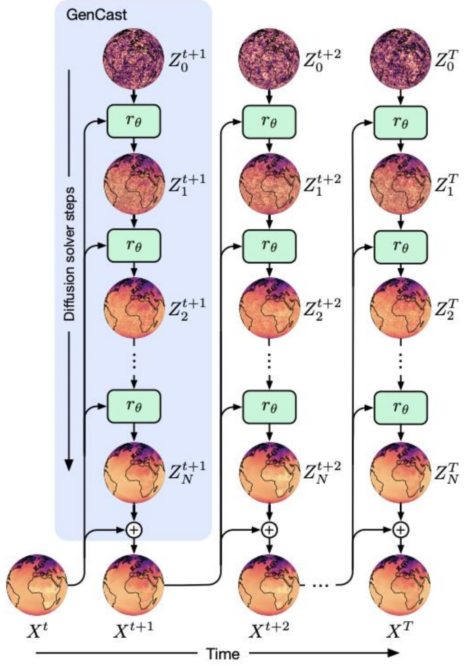


What loss to optimise?

- Regression losses (MSE, MAE):
  - Robust and easy to implement
  - Penalises large errors strongly
  - Will produce smooth fields.
- Probabilistic (CRPS, Energy Score):
  - Harder to implement (ensemble needed)
  - Evaluates full predictive distribution
- Generative (Diffusion models):
  - Slightly modified training task
    - From input noise, produce forecasts conditioned on initial conditions
  - Inherently stochastic
- For all losses, you need to decide how to aggregate over variables & heights.



CRPS: CDF should match the observation step CDF function
MSE: minimise forecast error



Probabilistic weather forecasting with machine learning, Price et al 2025

Which model?

Architecture choices

Convolutions

Graphs

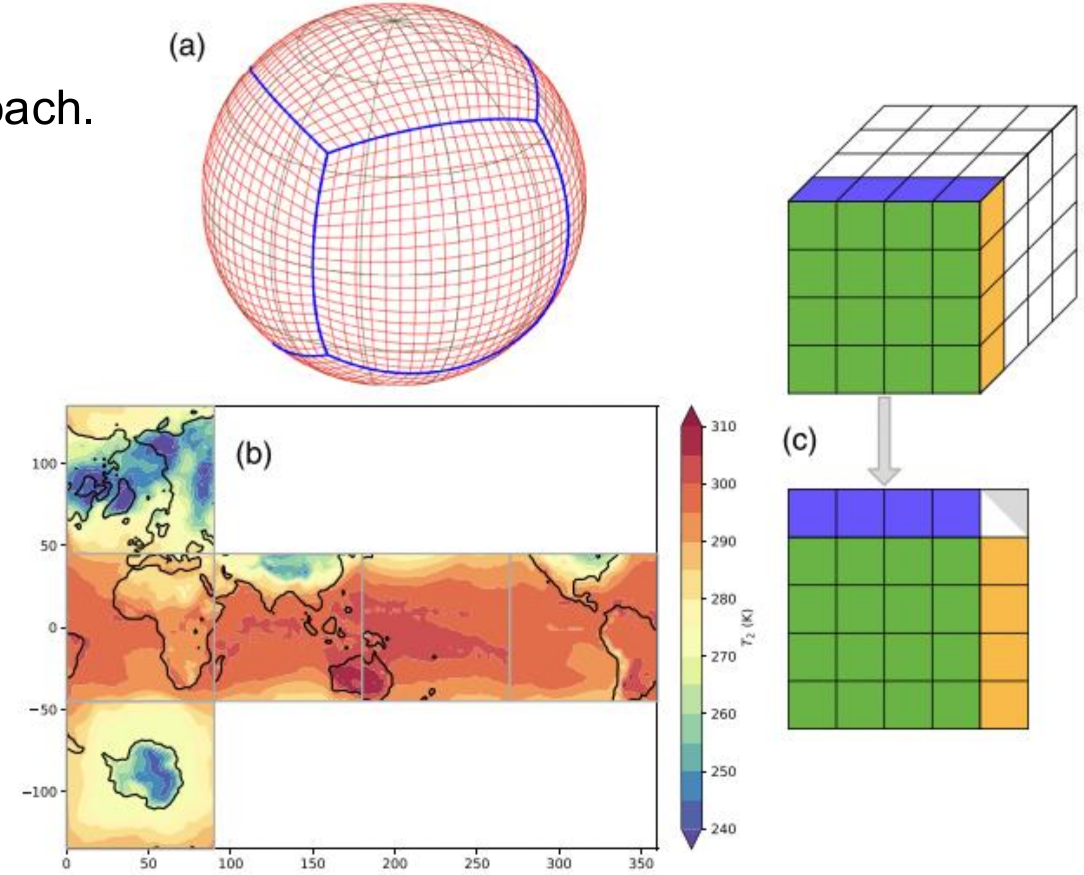
Transformers

Fourier Neural Operators

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 artic?

- 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)

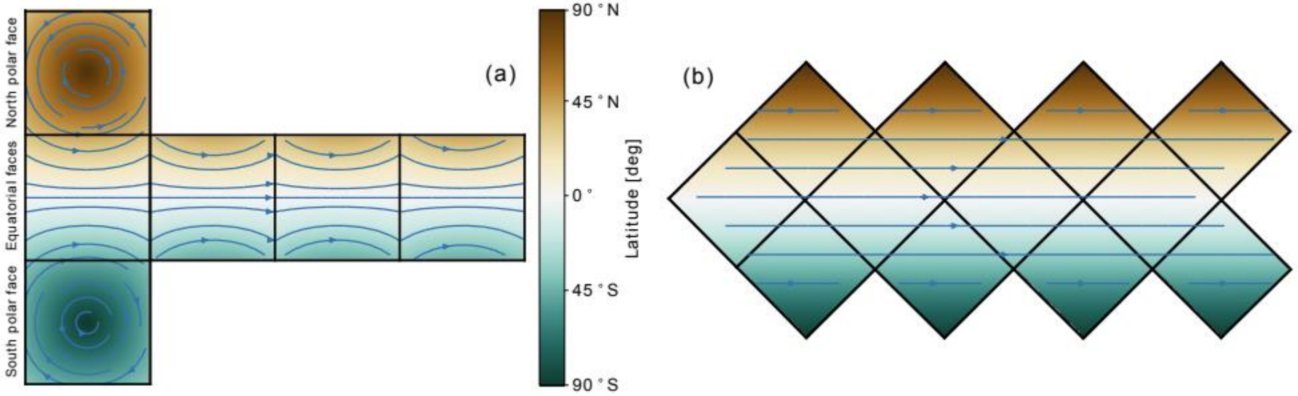
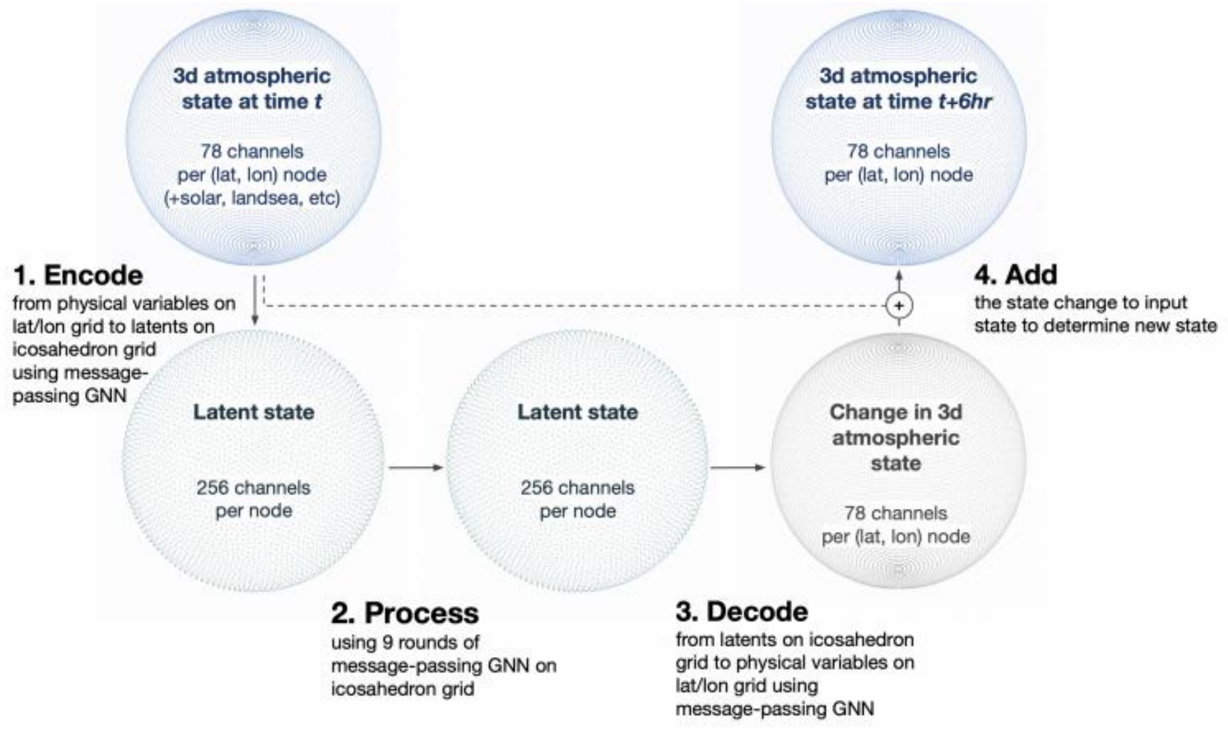
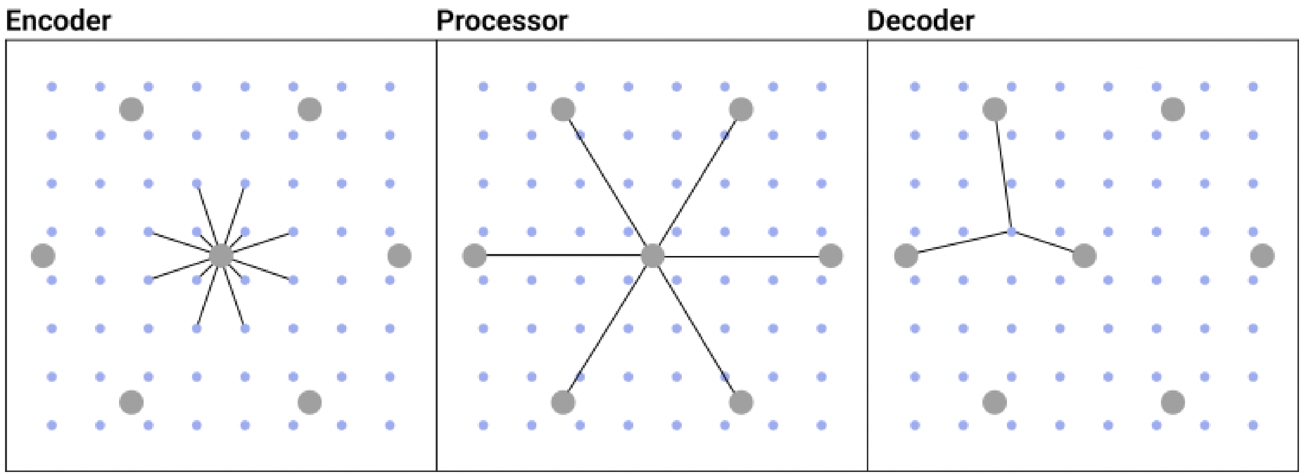


Figure 10: Lines of latitudes depicted as blue streamline arrows on the cubed sphere (a) and on the HEALPix (b). While the lines corresponding to constant eastward motion describe arcs of different radii on the cubed sphere mesh, the same motion translates to straight lines on the HEALPix mesh.

Graph Neural Networks

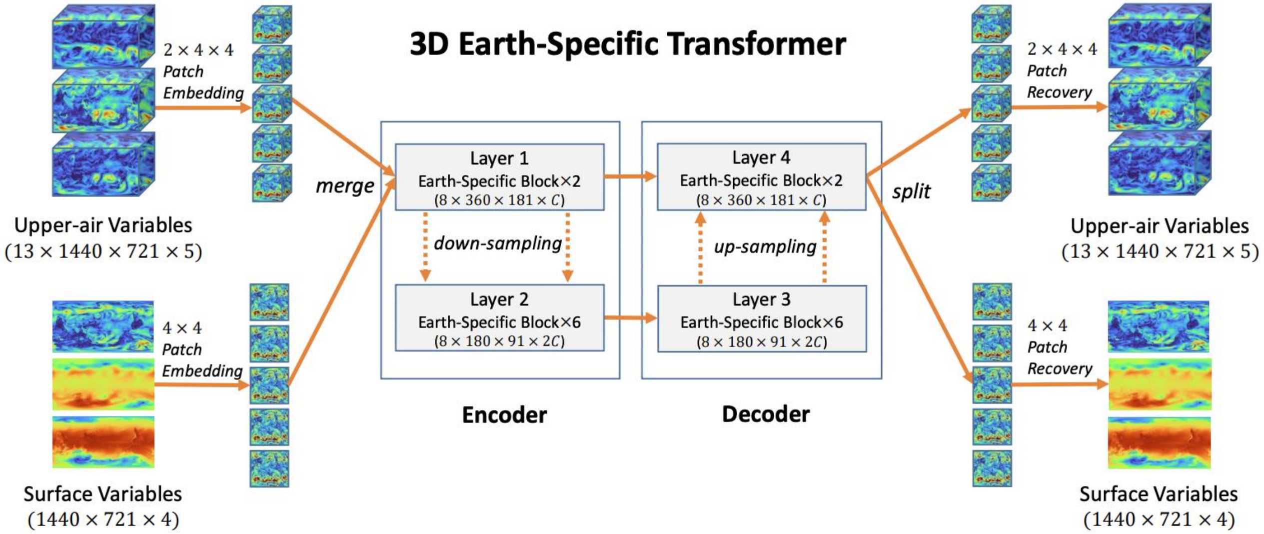
- First demonstrated by Keisler 2022.
- Data is structured in nodes and the connections between them (edges)
- Most popular, the message passing GNN.
  - Involves MLPs on the edges, and nodes.
  - Alternates are GraphConvolutions & Graph Attention
- Further developed in GraphCast
  - And used in early versions of the AIFS.
- No issues at the poles.
- Can handle irregular data in space.



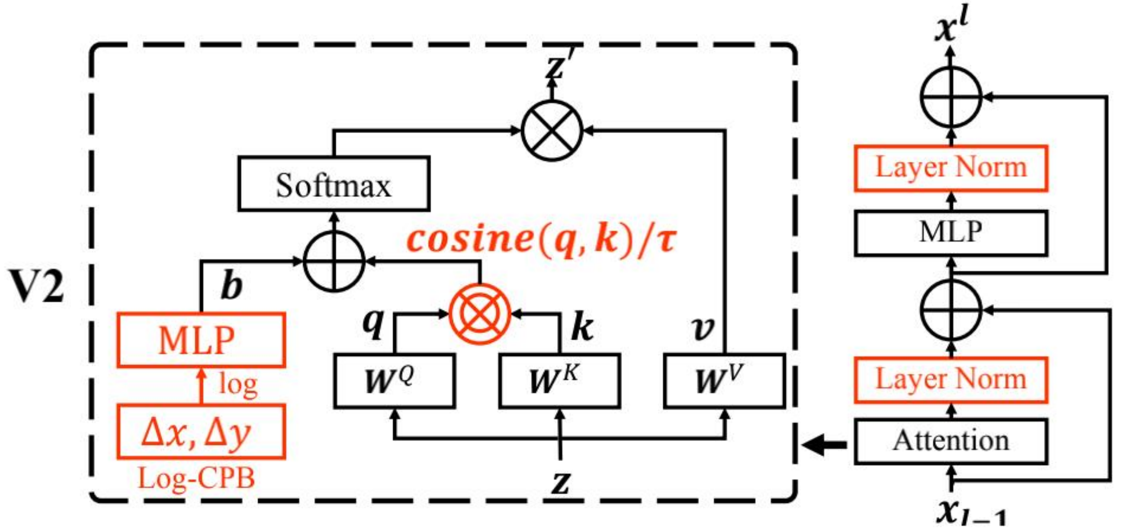


(Vision) Transformers

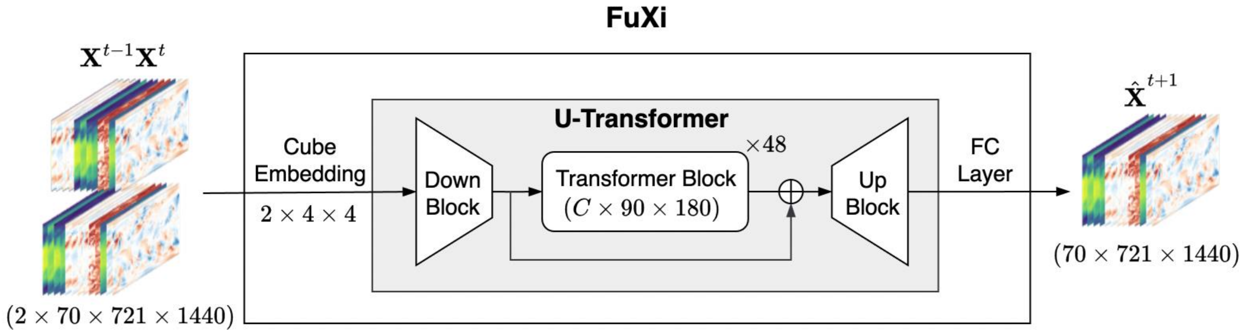
- Build heavily on advances in vision transformers
  - Specifically Shifted Window (SWIN) approaches.
- Pangu, FuXi, FengWu.
- Embed to a coarser resolution.
- Shifted-window approach adapted to include longitudinal periodicity.
  - But poles are not explicitly handled.



SwinTransformerV2 block

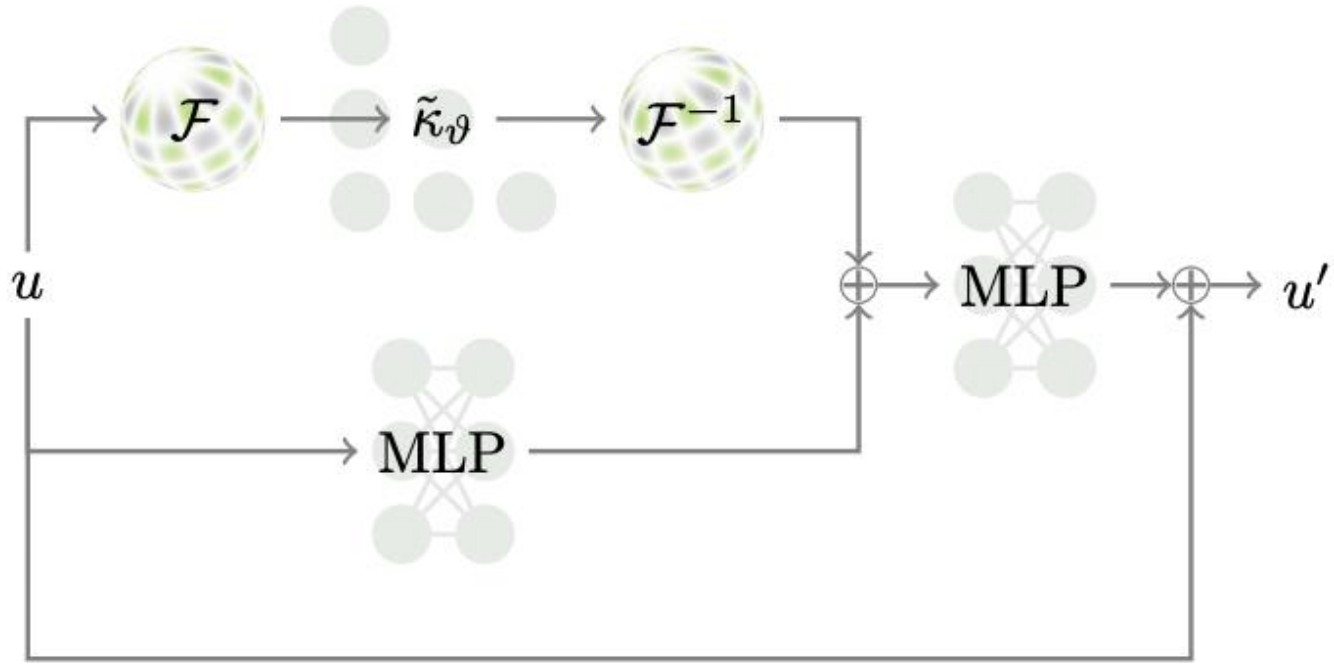


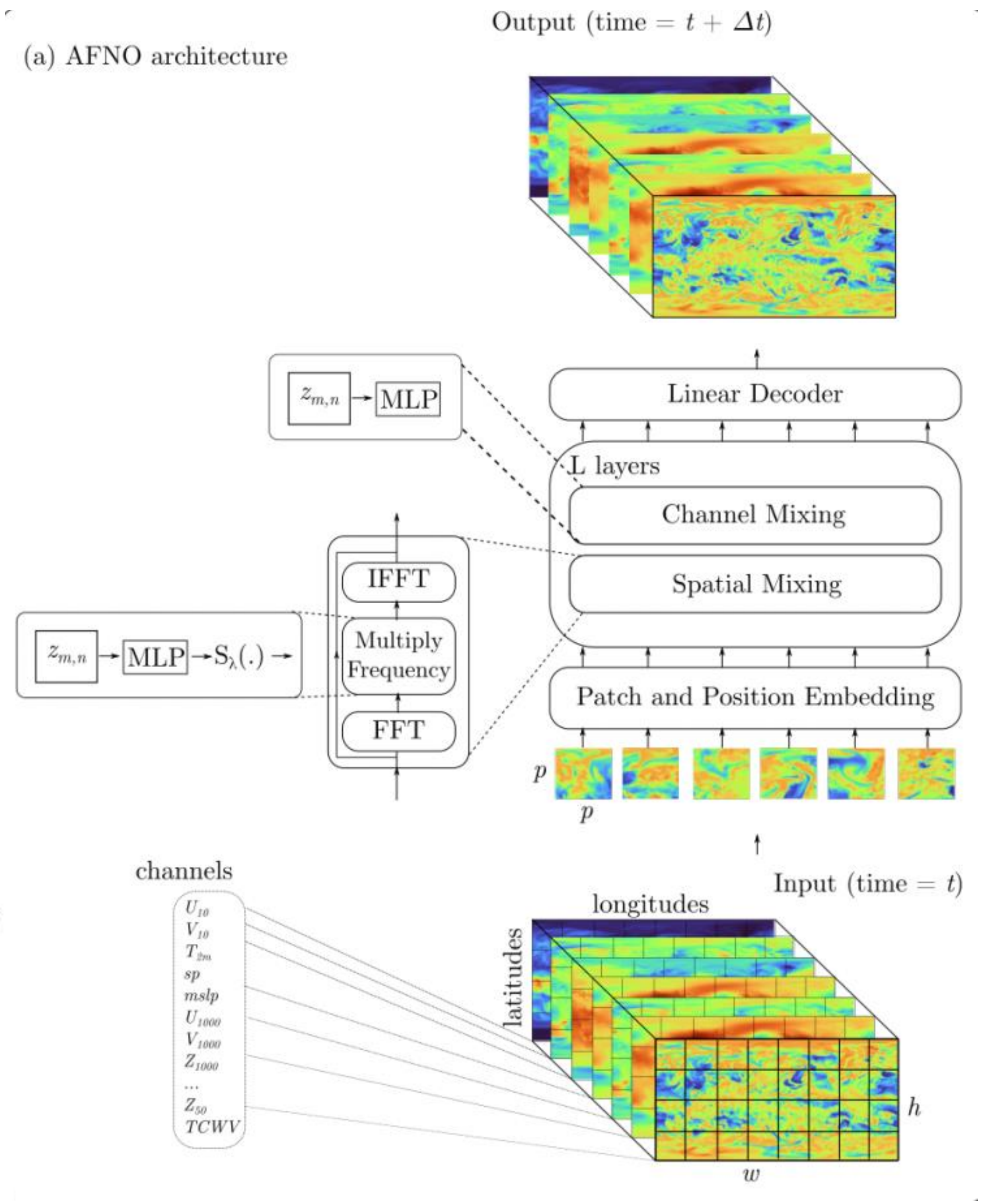
a) The overall architecture of FuXi model



Fourier Neural operators

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





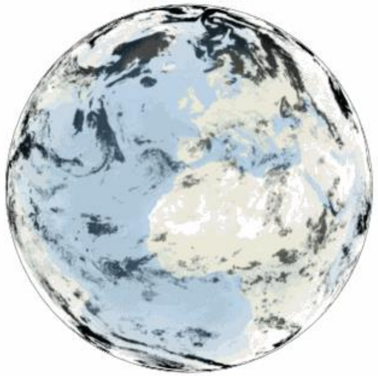
So, which to choose?

- 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 equi-spaced grids.

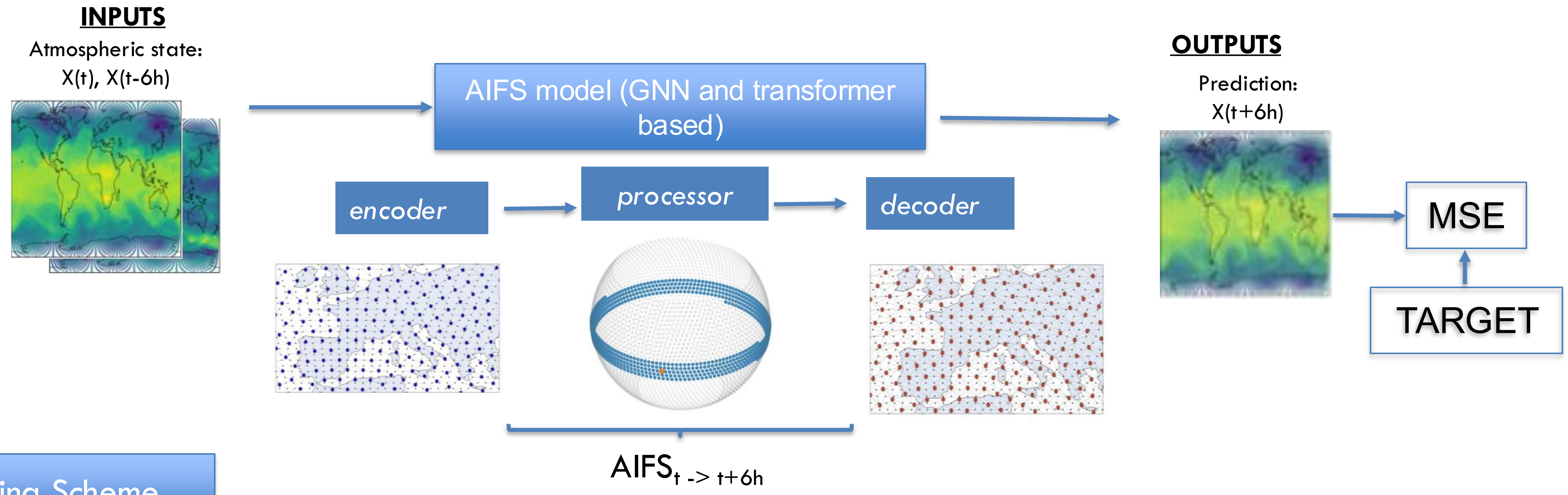
- Vision transformers (and SFNO) appear to converge faster than GNN.

(**A**)RTIFICIAL (**I**)NTELLIGENCE (**F**)ORECASTING (**S**)YSTEM



AIFS - Artificial Intelligence Forecasting System

TRAINING



Training Scheme

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

Step 2: model is autoregressively trained to optimise forecasts between 6h and 72h ahead

ECMWF

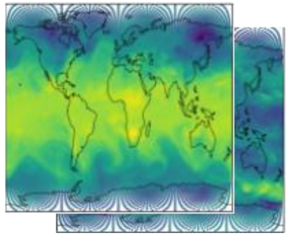
AIFS - Artificial Intelligence Forecasting System

https://arxiv.org/abs/2406.01465
https://doi.org/10.5194/egusphere-2025-4716

TRAINING

INPUTS

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

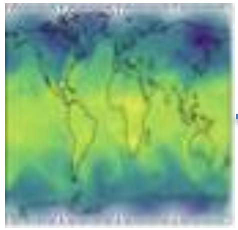


AIFS model (GNN and transformer based)

encoder → processor → decoder

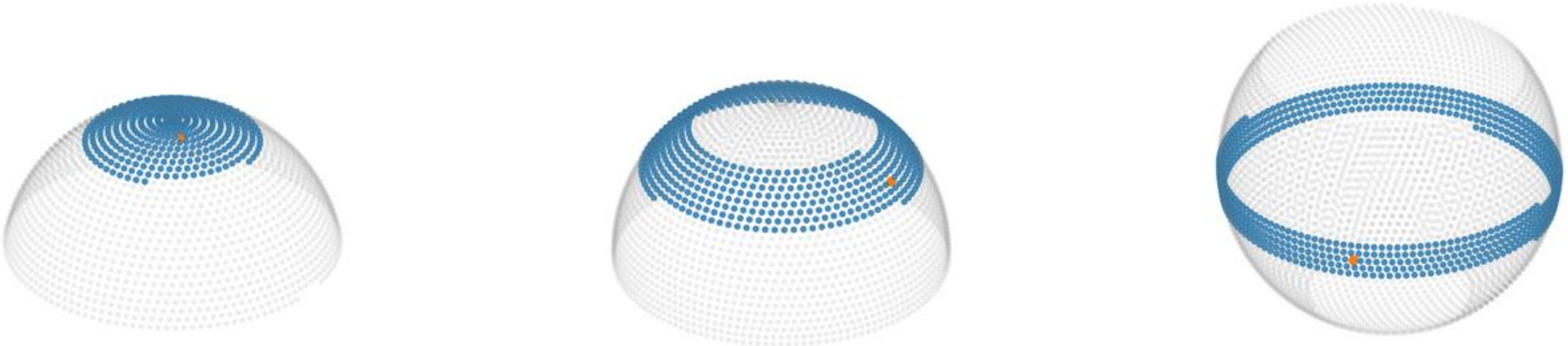
OUTPUTS

Prediction:
X(t+6h)



MSE

TARGET

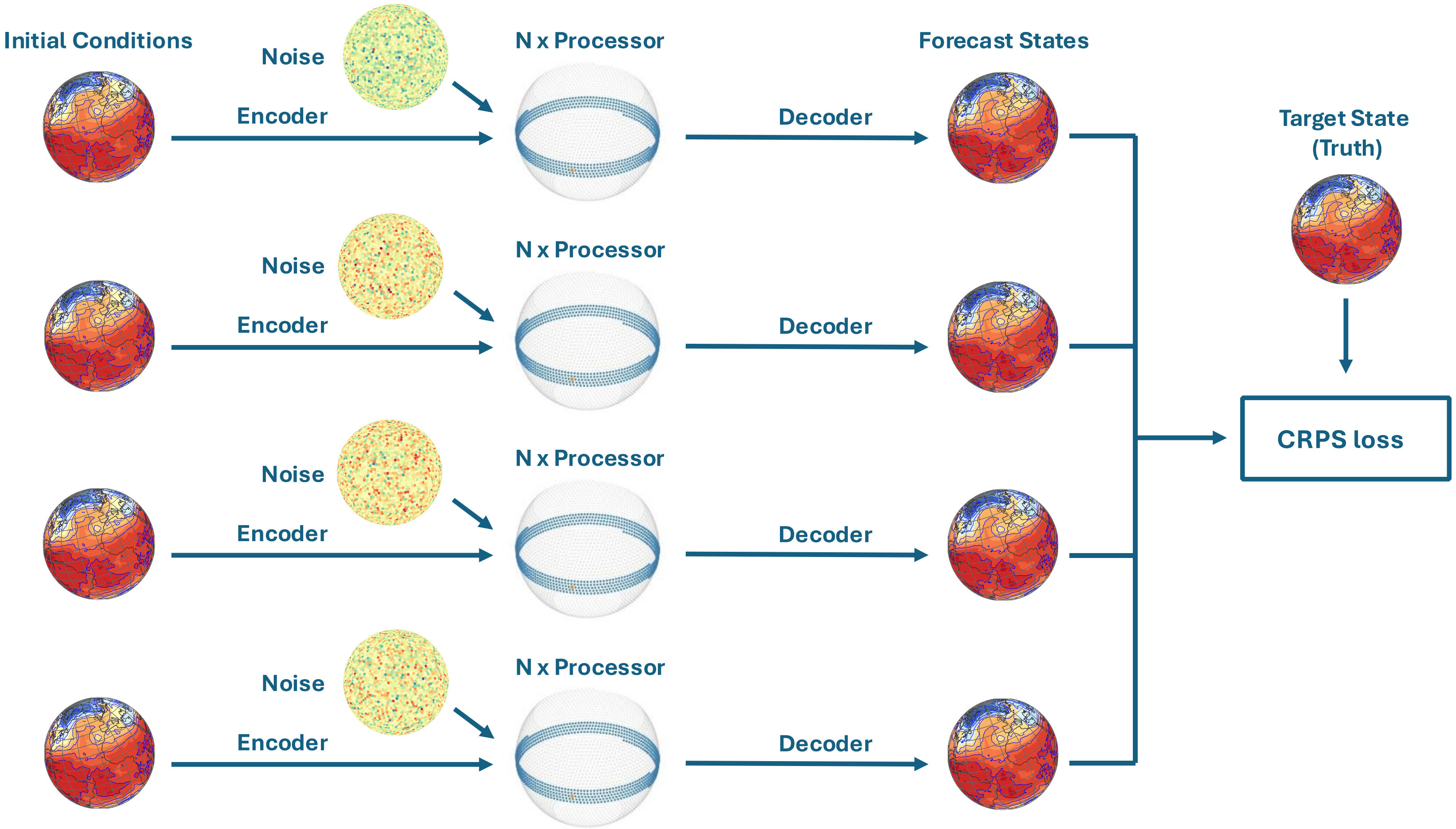


Transformer (e.g., like LLMs) that works with a sliding attention window → attention bands around the globe.

https://arxiv.org/abs/2412.15832

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

In training: run (small) ensemble:



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

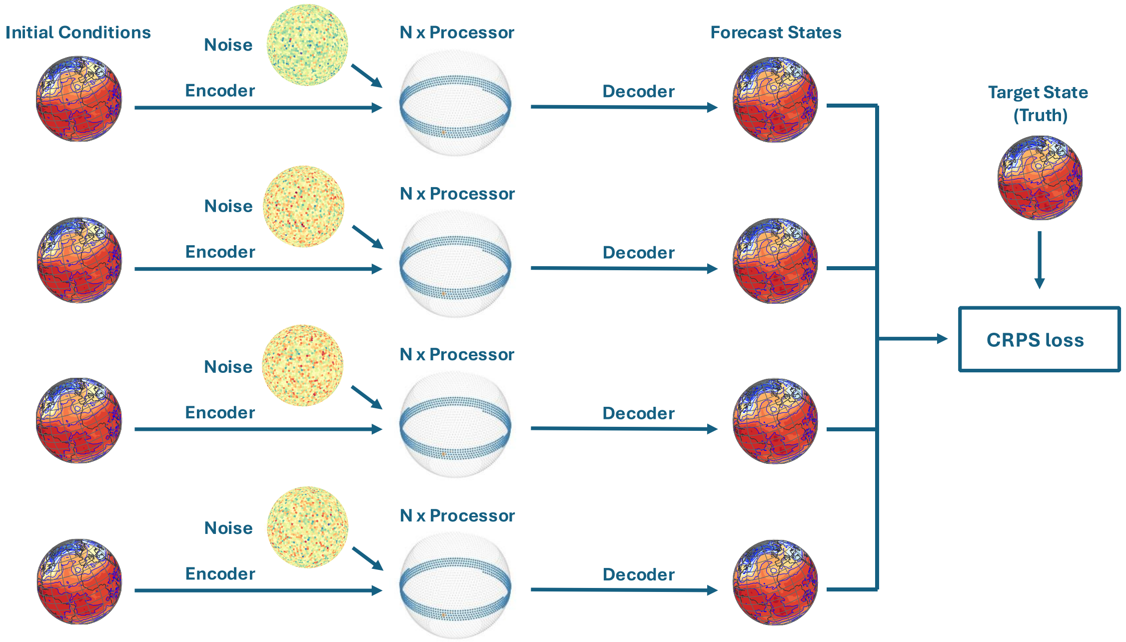
AIFS-ENS:

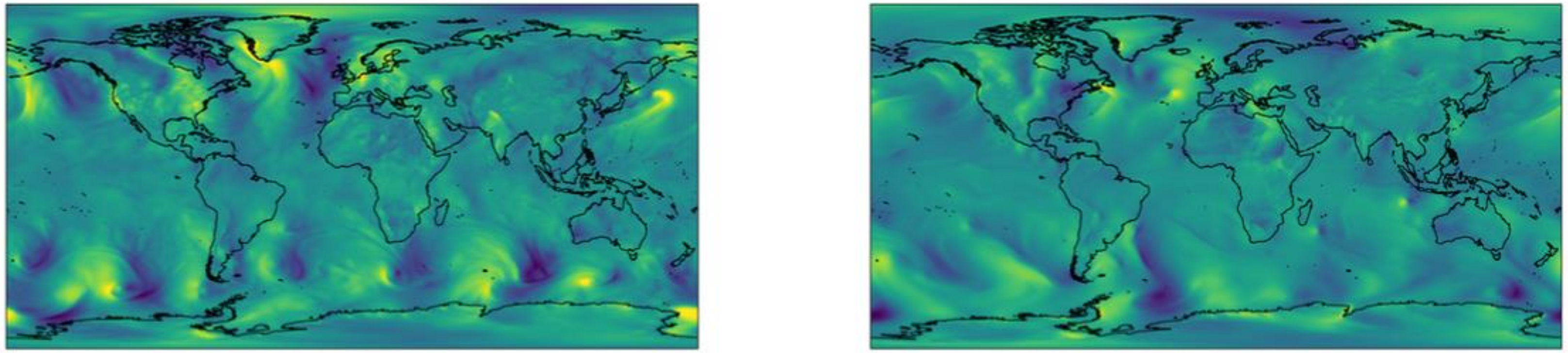
Probabilistic training of AIFS:

$$\mathrm{afCRPS}_\alpha := \alpha\,\mathrm{fCRPS} + (1-\alpha)\mathrm{CRPS}$$

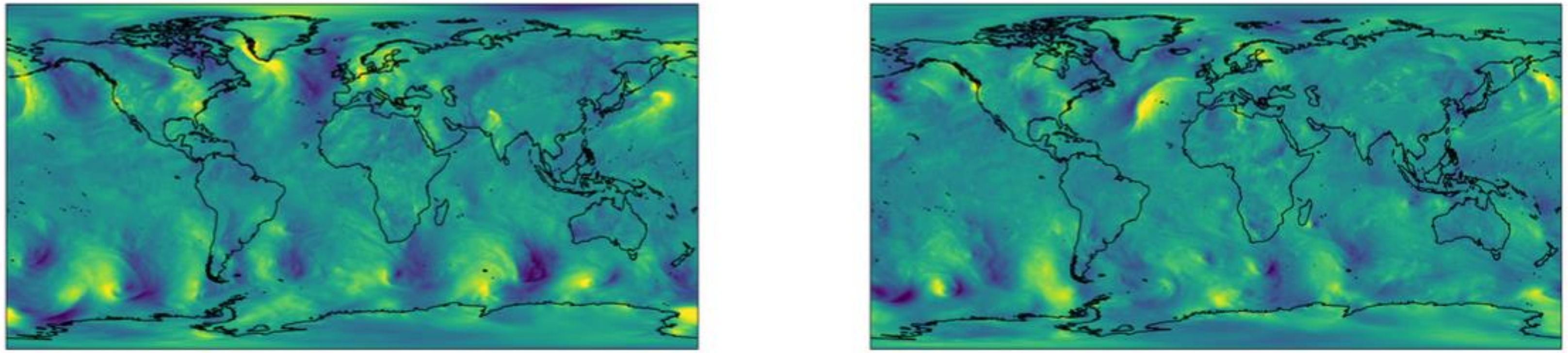
$$= \frac{1}{M}\sum_{j=1}^{M}|x_j - y| - \frac{M-1+\alpha}{2M^2(M-1)}\sum_{j=1}^{M}\sum_{k=1}^{M}|x_j - x_k|$$

$$= \frac{1}{M}\sum_{j=1}^{M}|x_j - y| - \frac{1-\epsilon}{2M(M-1)}\sum_{j=1}^{M}\sum_{k=1}^{M}|x_j - x_k|$$





AIFS-single, day 1 and 10



AIFS-ENS, day 1 and 10

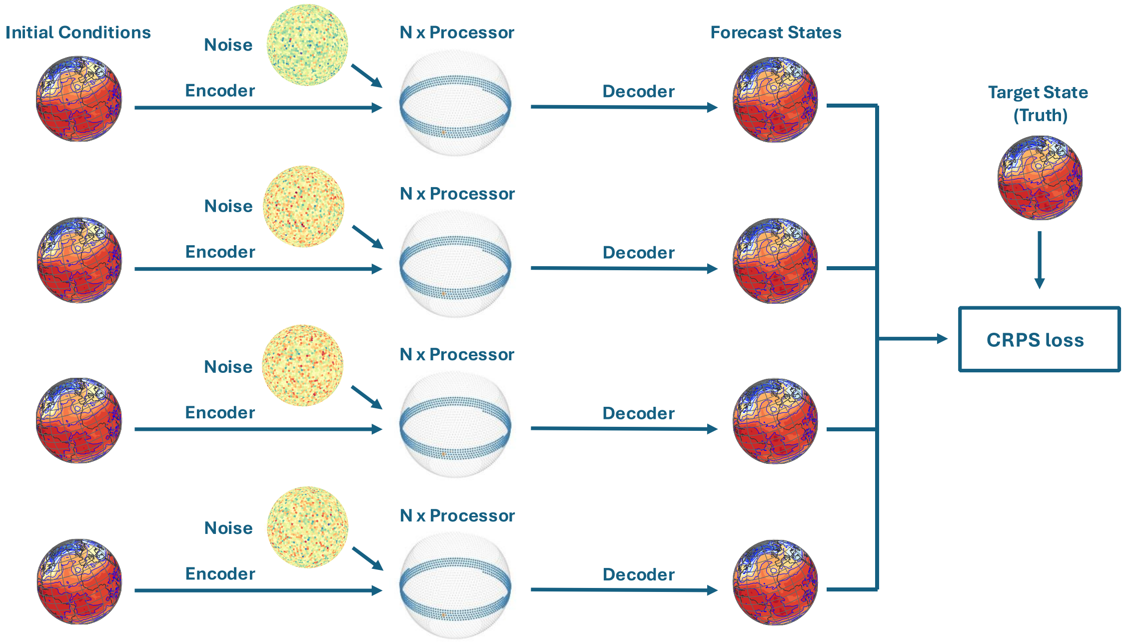
AIFS-ENS:

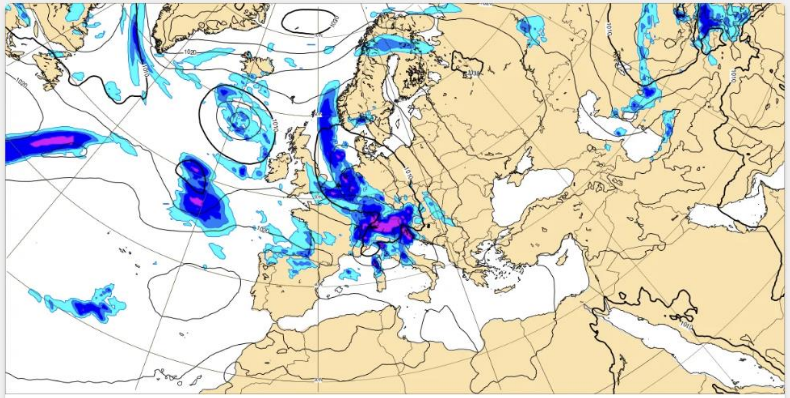
Probabilistic training of AIFS:

$$\mathrm{afCRPS}_\alpha := \alpha\,\mathrm{fCRPS} + (1-\alpha)\mathrm{CRPS}$$

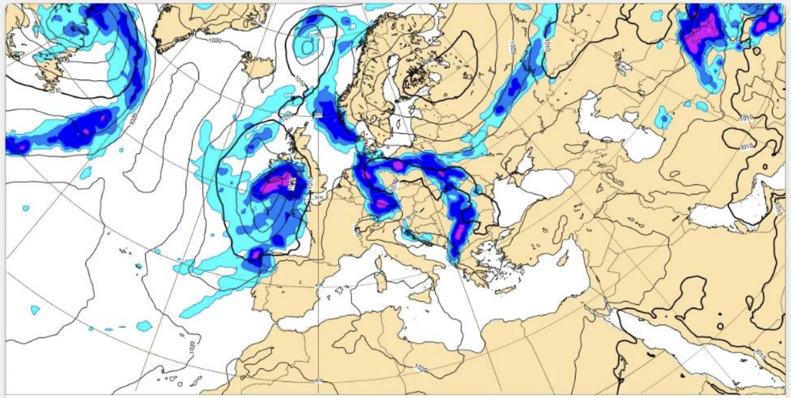
$$= \frac{1}{M}\sum_{j=1}^{M}|x_j - y| - \frac{M-1+\alpha}{2M^2(M-1)}\sum_{j=1}^{M}\sum_{k=1}^{M}|x_j - x_k|$$

$$= \frac{1}{M}\sum_{j=1}^{M}|x_j - y| - \frac{1-\epsilon}{2M(M-1)}\sum_{j=1}^{M}\sum_{k=1}^{M}|x_j - x_k|$$

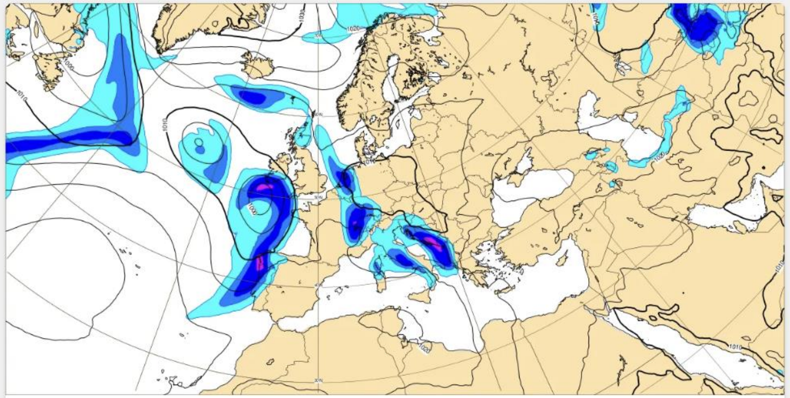




Rain and mean sea level pressure



AIFS CRPS : Rain and mean sea level pressure



AIFS Single: Rain and mean sea level pressure

10-day forecasts

AIFS-ENS and AIFS-Single:

Operational since February 2025 AIFS-Single and July 2025 AIFS-CRPS

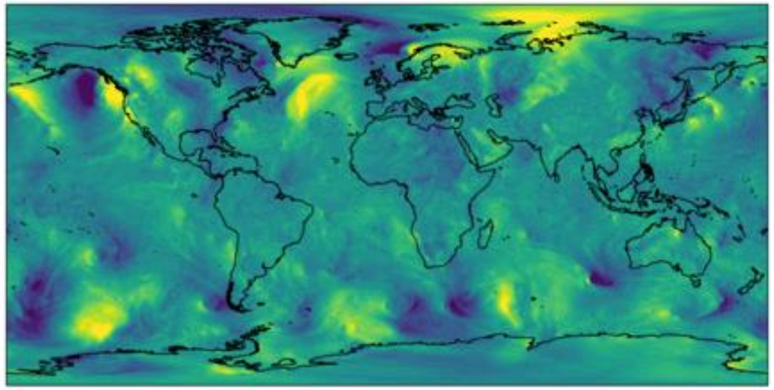
AIFS-ENS Forecast configuration:

50 perturbed member, starting from the perturbed initial conditions of the IFS-ENS

1 control member, starting from the unperturbed initial conditions of the IFS-ENS control

15-day forecasts, N320 (~ 0.25) resolution, 6 hourly output, like AIFS-Single

AIFS-ENS training is inherently probabilistic, the control member of AIFS-ENS still has representation of uncertainty (similar to stochastic representation of model uncertainty in IFS-ENS)! This means, it is somewhat less skillful on average than the IFS-ENS control



Are these models skilled?

AIFS ENS 1 vs IFS scorecard

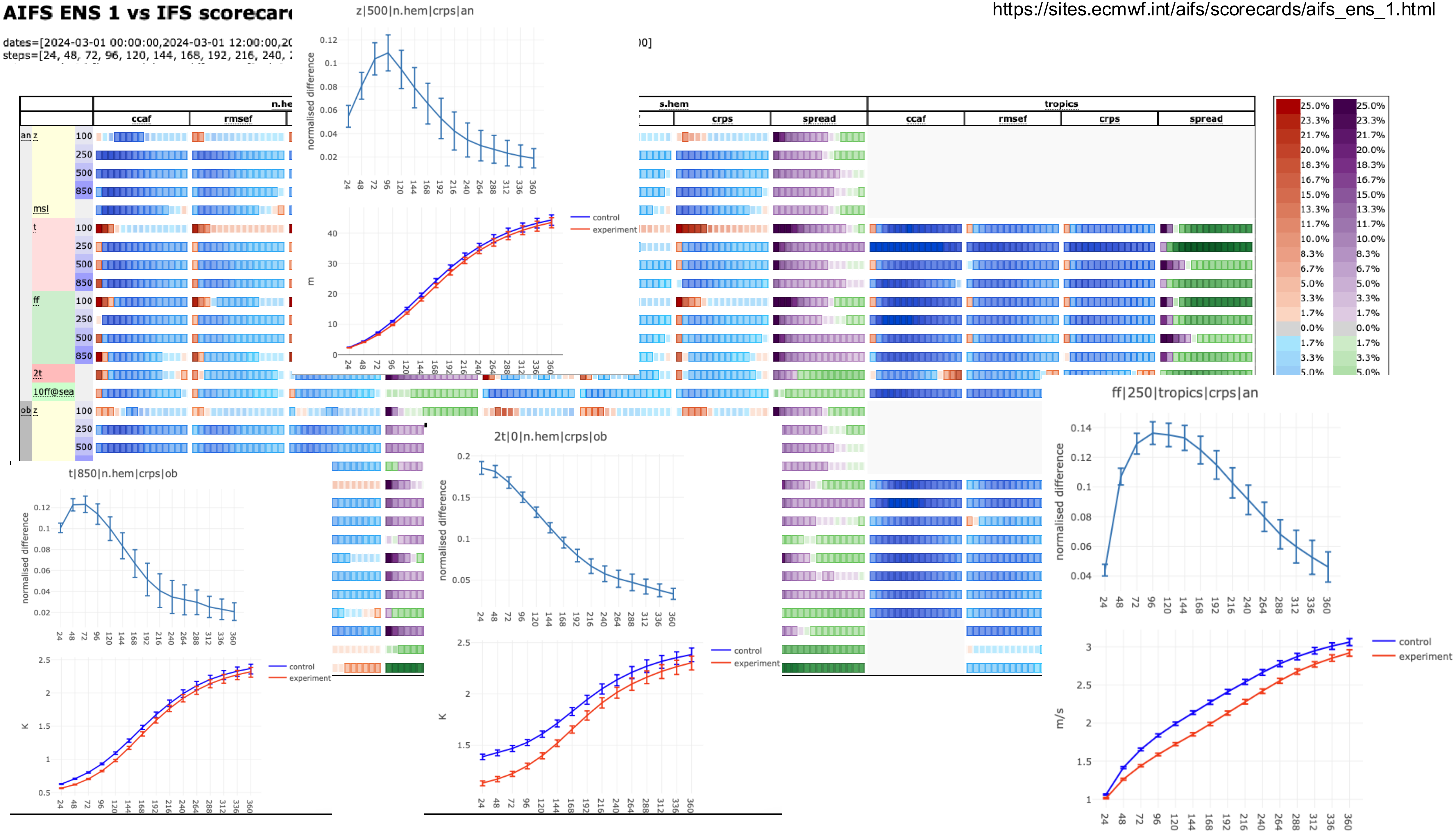
https://sites.ecmwf.int/aifs/scorecards/aifs_ens_1.html

dates=[2024-03-01 00:00:00,2024-03-01 12:00:00,2024-03-02 00:00:00,...,2025-02-28 12:00:00,2025-03-01 00:00:00]
steps=[24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 264, 288, 312, 336, 360]

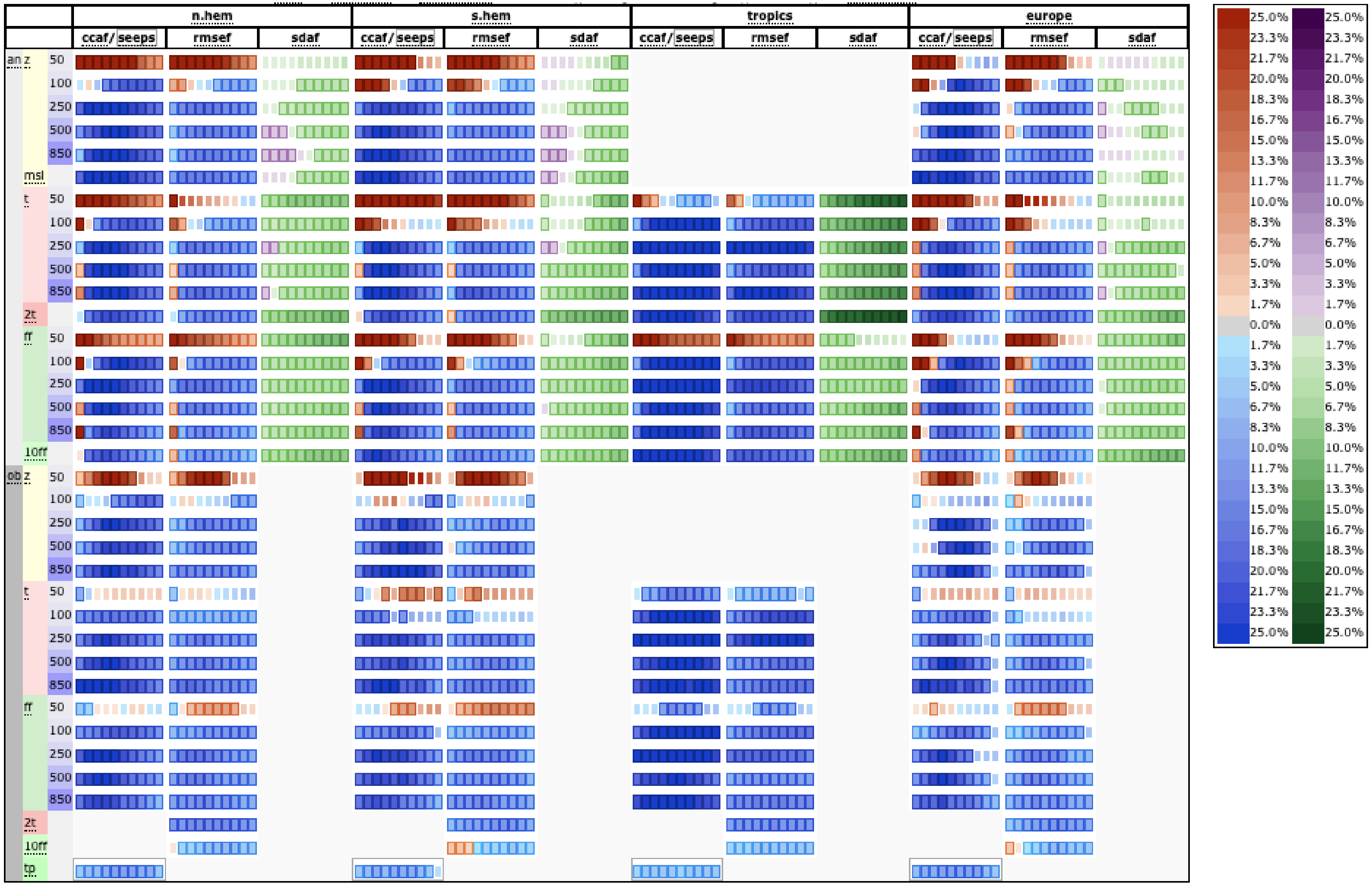
AIFS ENS 1 vs IFS scorecar

https://sites.ecmwf.int/aifs/scorecards/aifs_ens_1.html

dates=[2024-03-01 00:00:00,2024-03-01 12:00:00,20 ... 0]
steps=[24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 2



AIFS vs IFS 2024

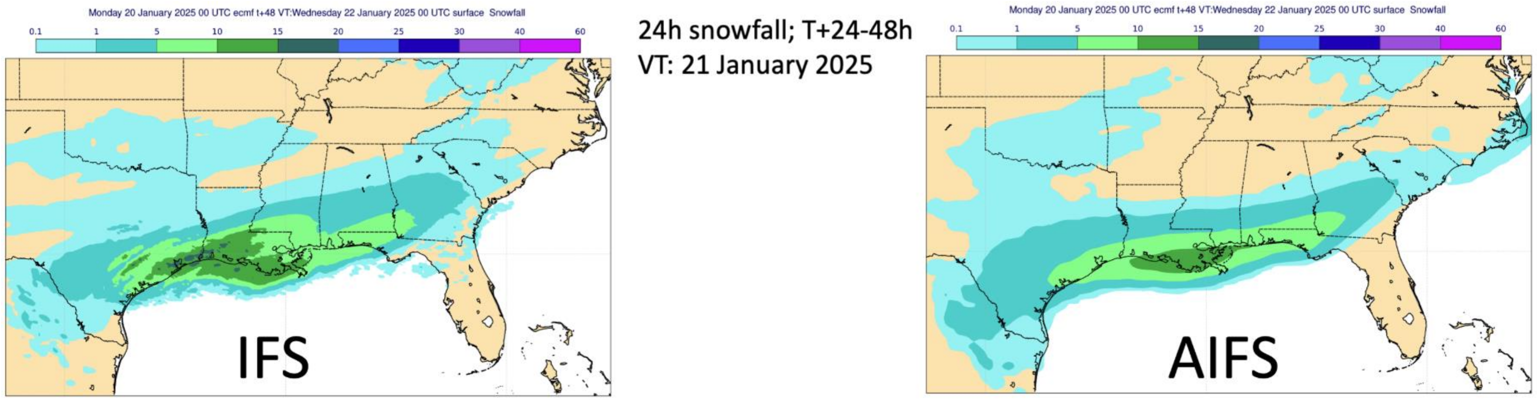


- Better performance overall for AIFS
- Less forecast activity (smoothing)
- Issues in the stratosphere
- Scores tend to be worse for short lead times 1D

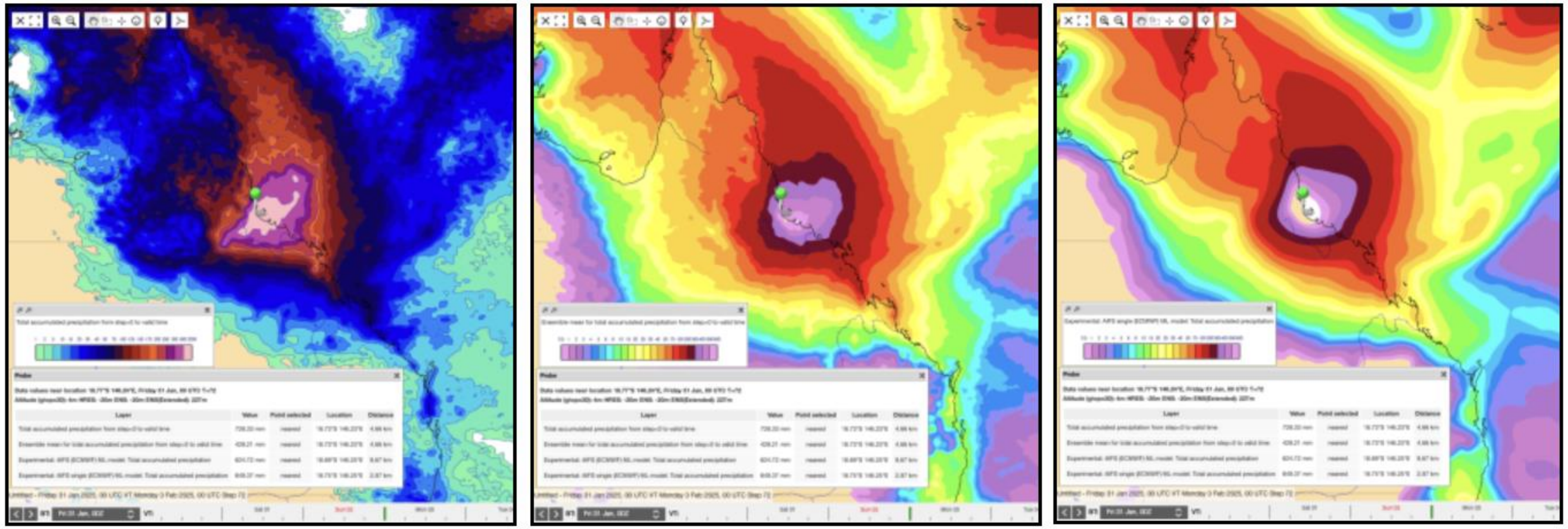
ECMWF

Are these models useful?

Case Studies: AIFS Single v1



Rare snow along the Gulf Coast
Structure well-predicted but underestimated intensity.



Heavy precipitation event in Queensland
AIFS predicts more extreme precipitation than the IFS

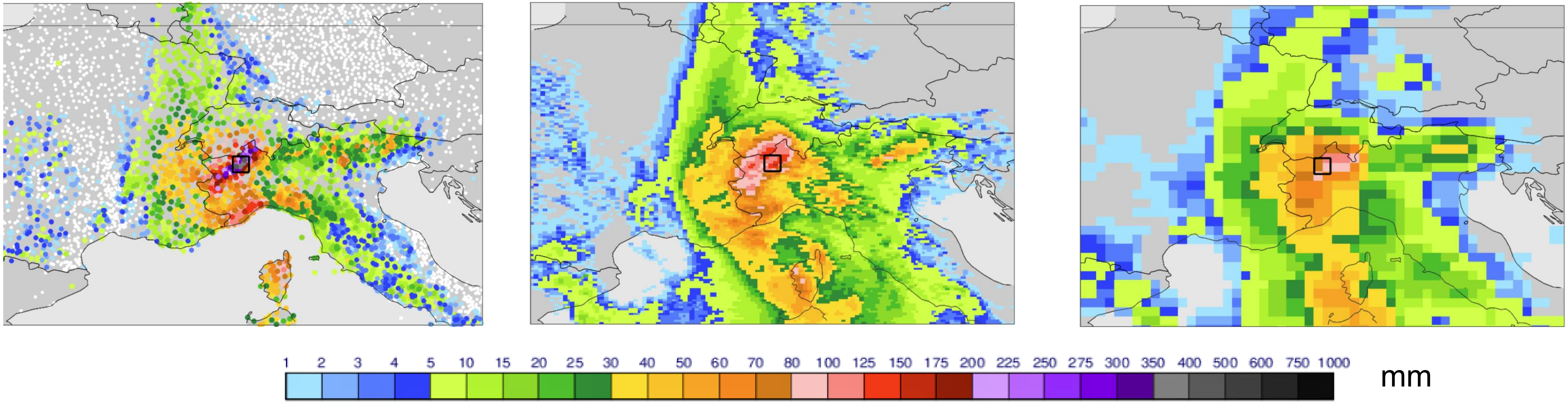
Case Studies

24-hour precipitation 16 April 06UTC – 17 April 06UTC in a 0.5 degree box in the Italian Alps

Observations

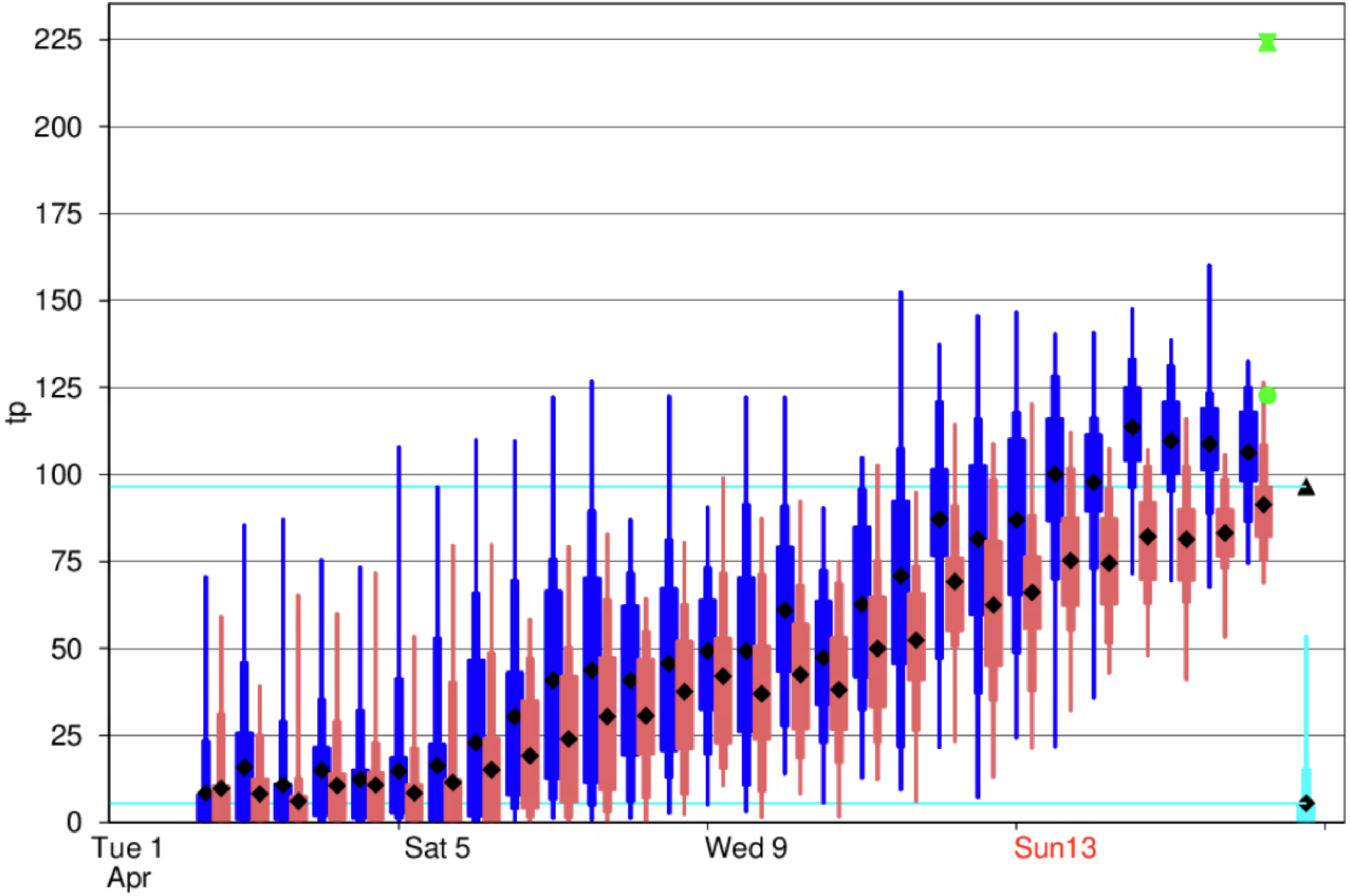
30-54h forecast from IFS-ENS Mem 1

30h-54h forecast from AIFS-ENS Mem 1



Case Studies

24-hour precipitation 16 April 06UTC – 17 April 06UTC in a 0.5 degree box in the Italian Alps



Observation – green hourglass
Analysis – green dot

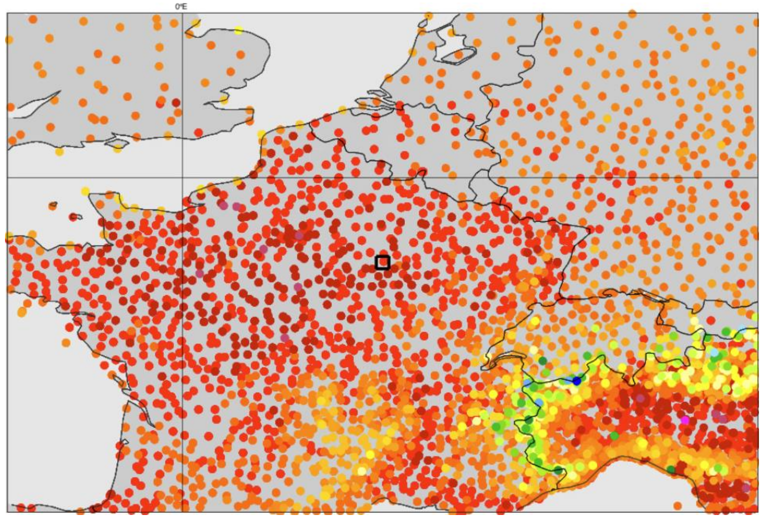
IFS-ENS – blue
AIFS-ENS – pink

M-climate – cyan

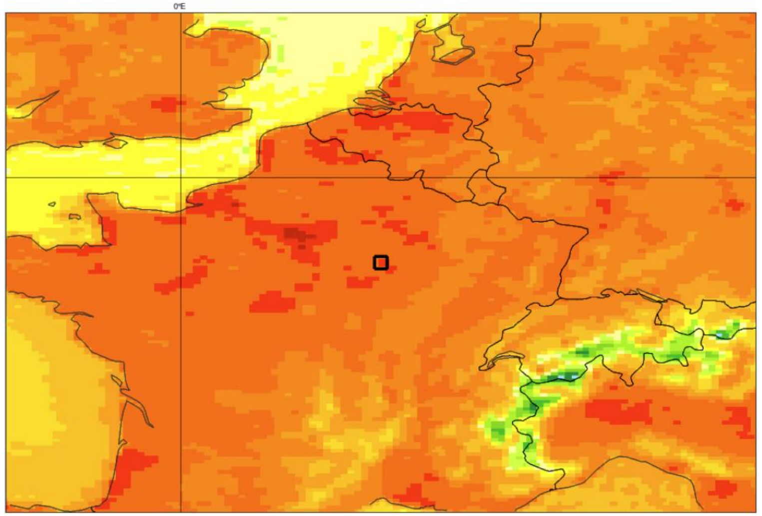
Case Studies

2-metre temperature 30 April 12UTC

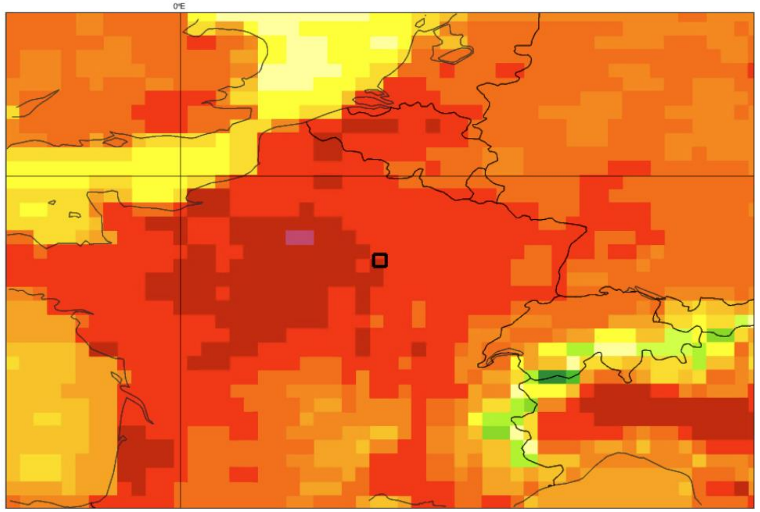
Observations

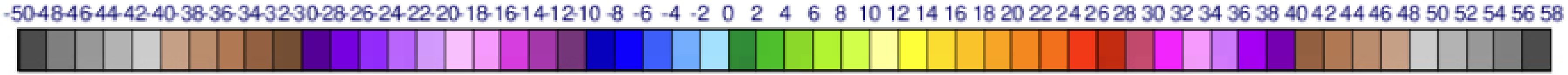


24h forecast from IFS-ENS Mem 1



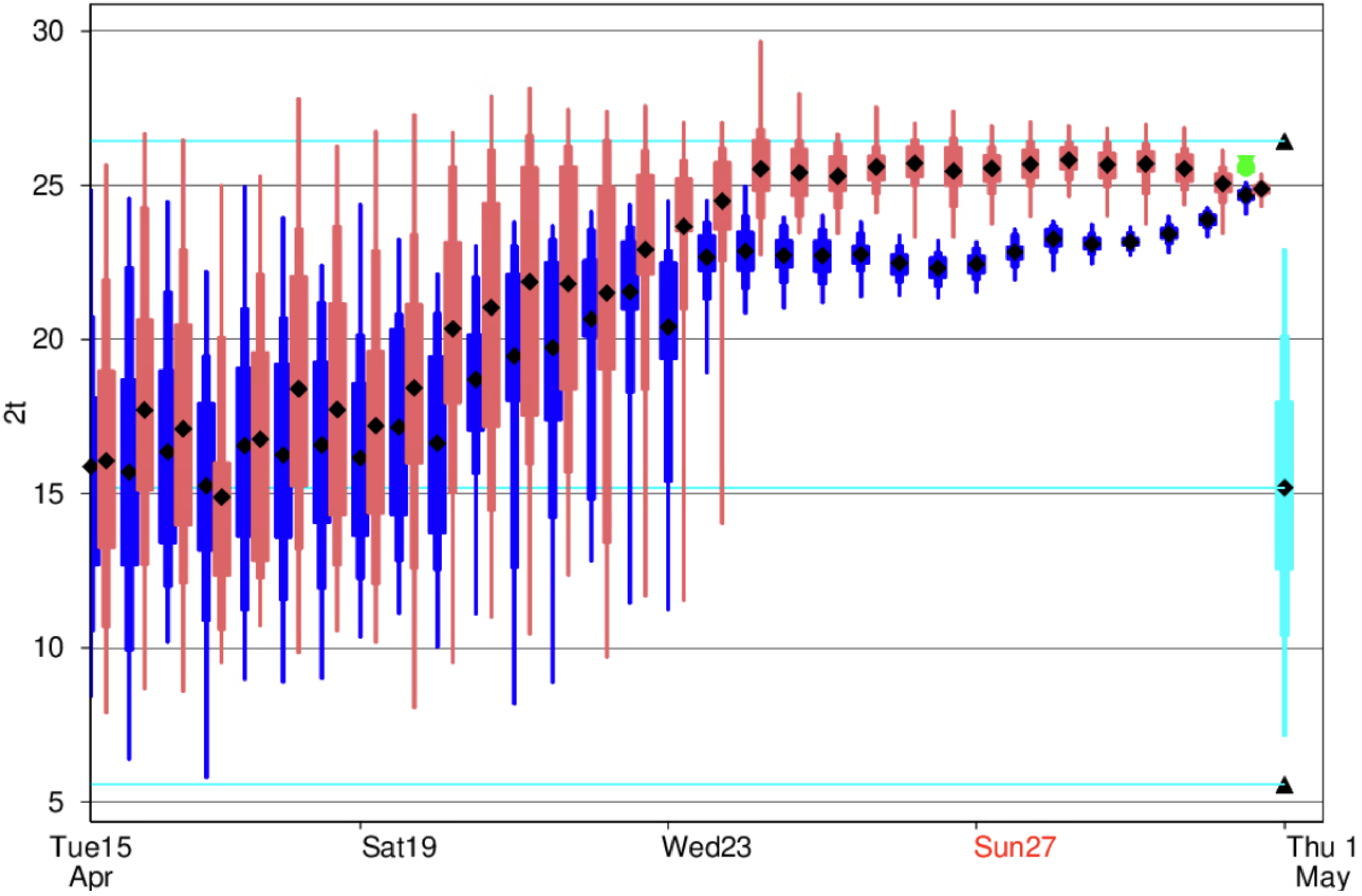
24h forecast from AIFS-ENS Mem 1





Case Studies

Evolution of forecasts for 2-metre temperature 30 April 12UTC in Troyes, France



Observation – green hourglass
Analysis – green dot

IFS-ENS – blue
AIFS-ENS – pink

M-climate – cyan

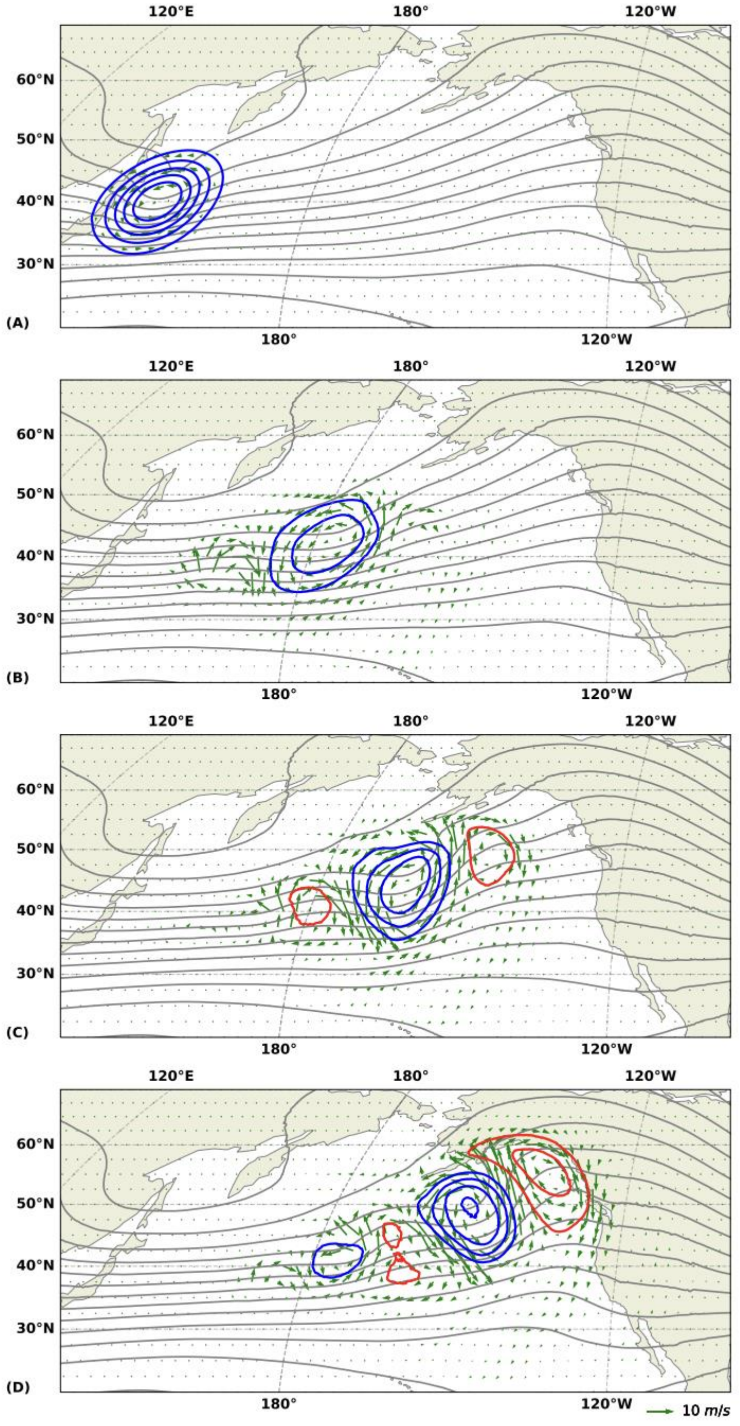
Are data-driven weather forecasts physical?

- 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.

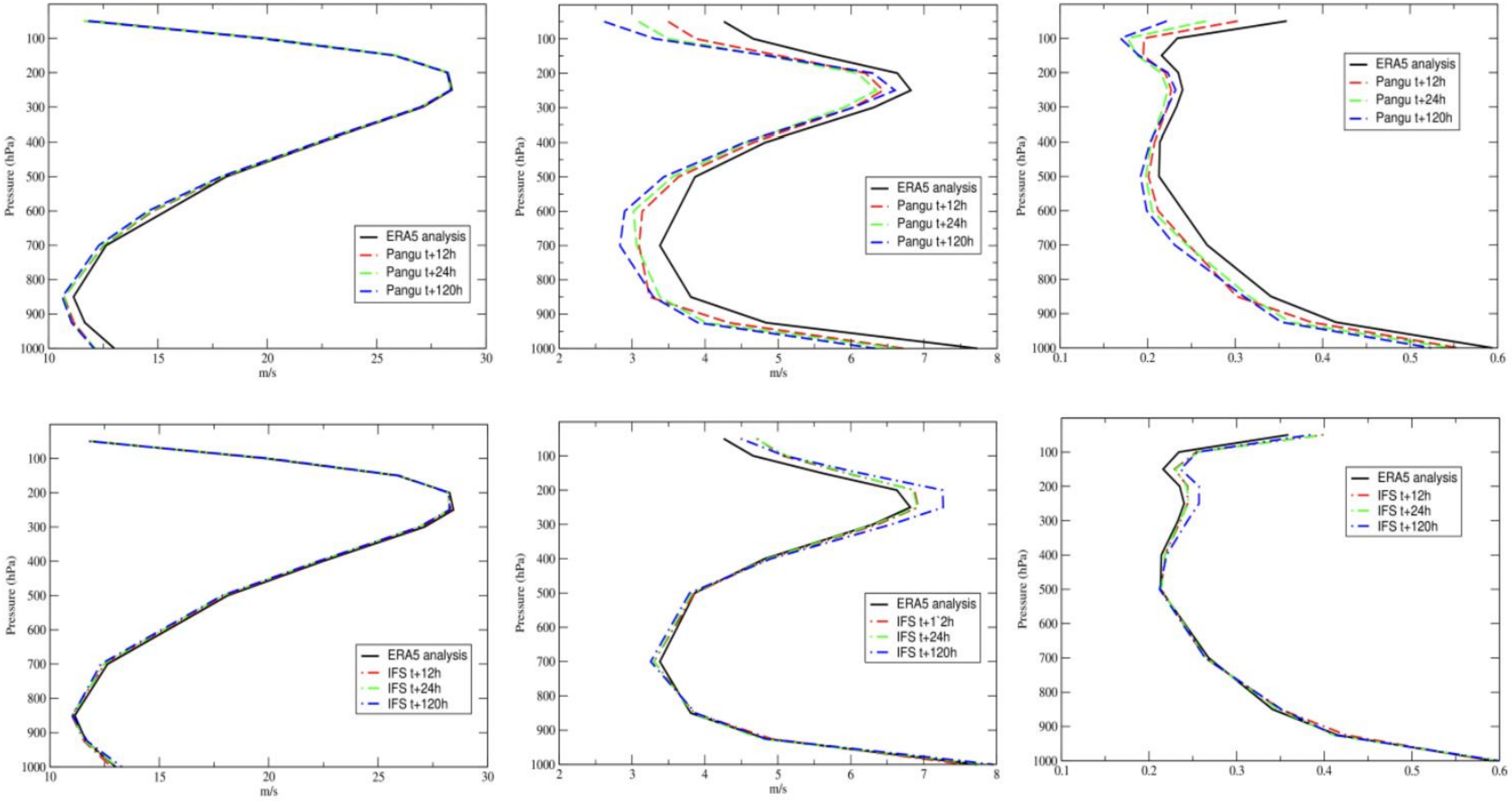
- Overall, Pangu behaves as expected, compared with theory.
  - The 1h model is best for the faster evolving processes, which aren’t well captured by the longer timestep model.

- 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.

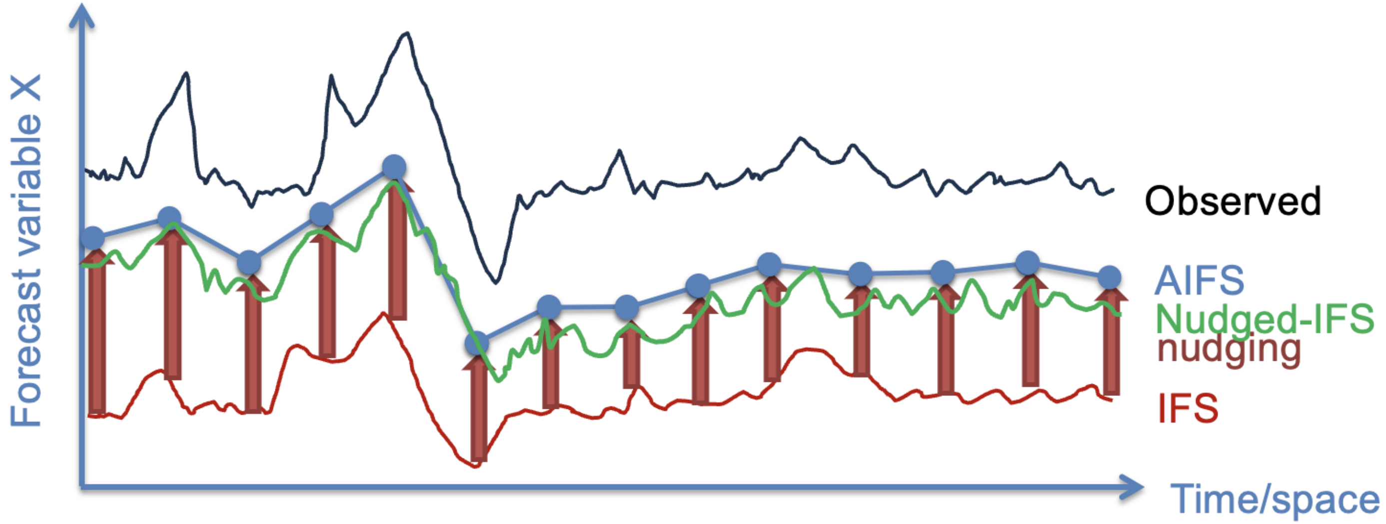


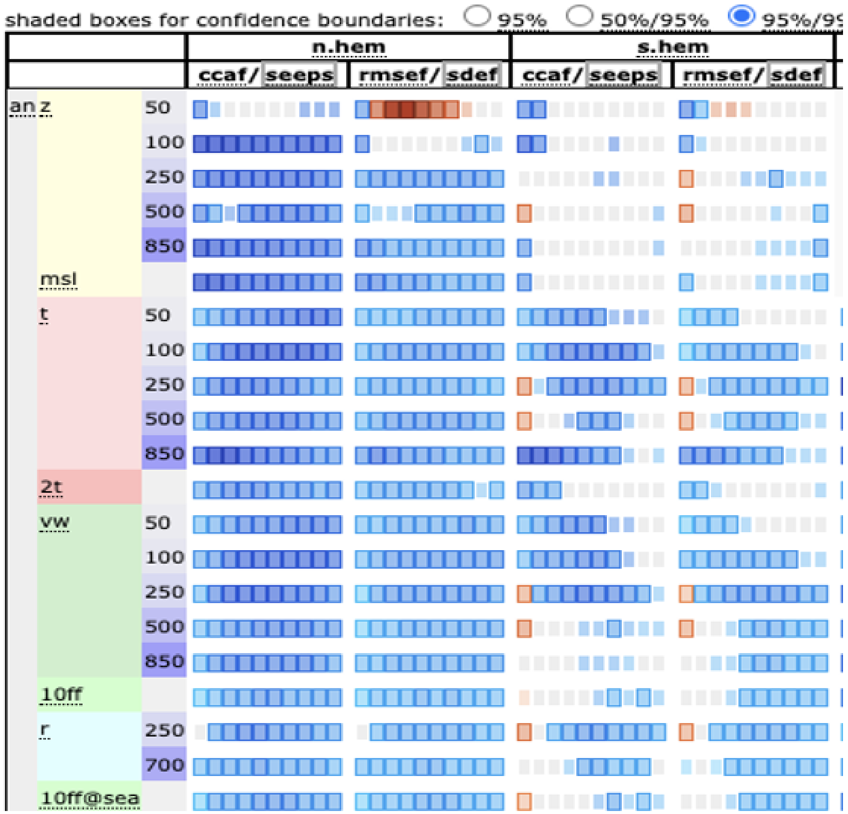
Geostrophic winds Ageostrophic winds Ratio

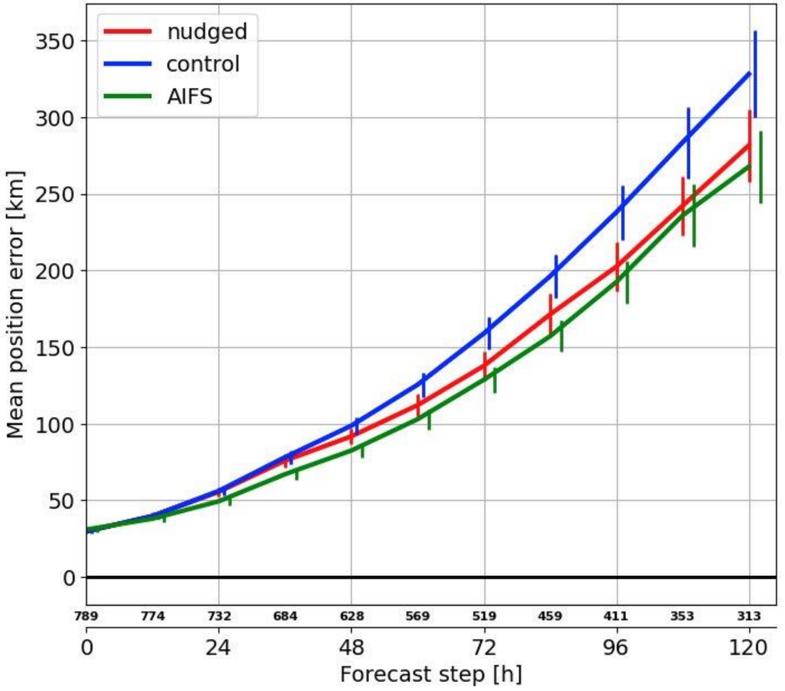
Other ML approaches

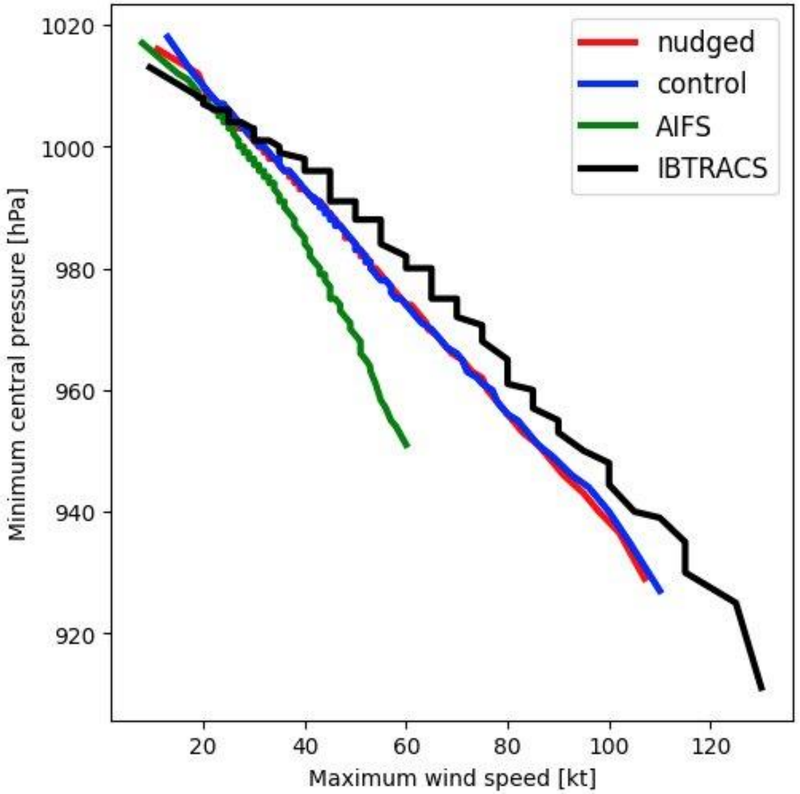
Driving the IFS with the AIFS

- Following the work by Hussain et al (2024)
- Develop custom AIFS version that operates on 137 model levels.
- Up to 15% improvements in ACC/RMSE



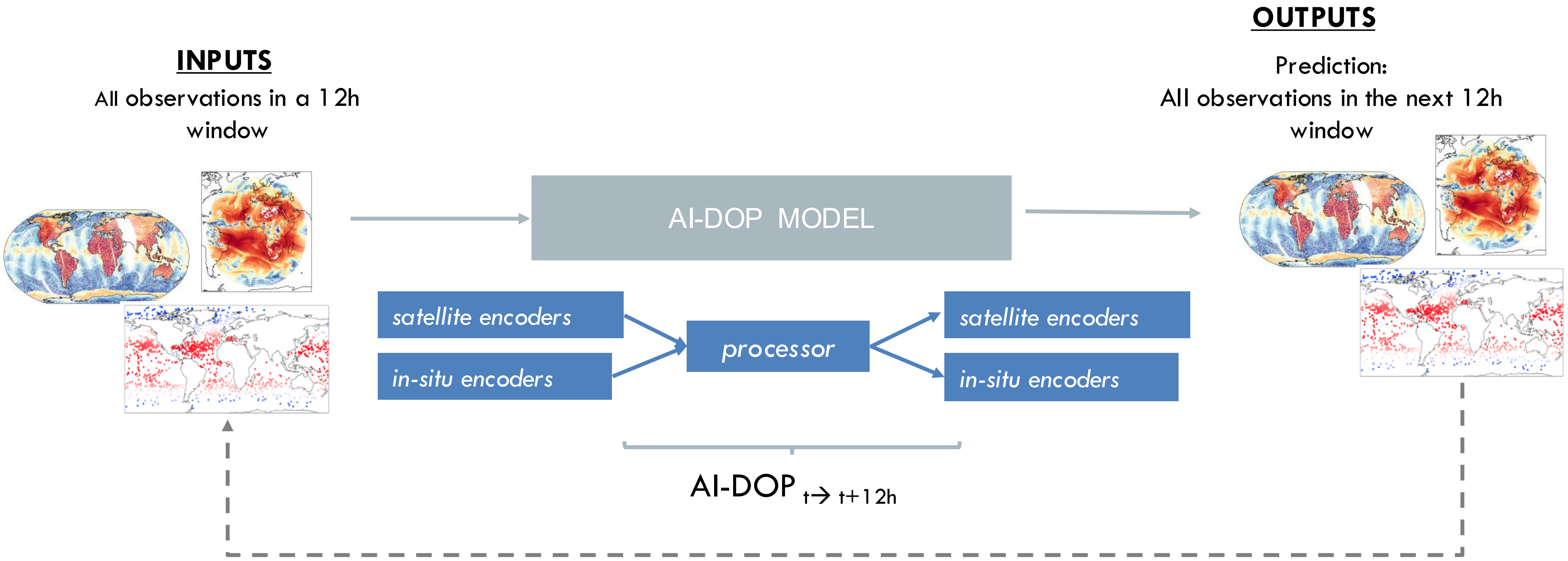






Artificial Intelligence Direct Observation Prediction

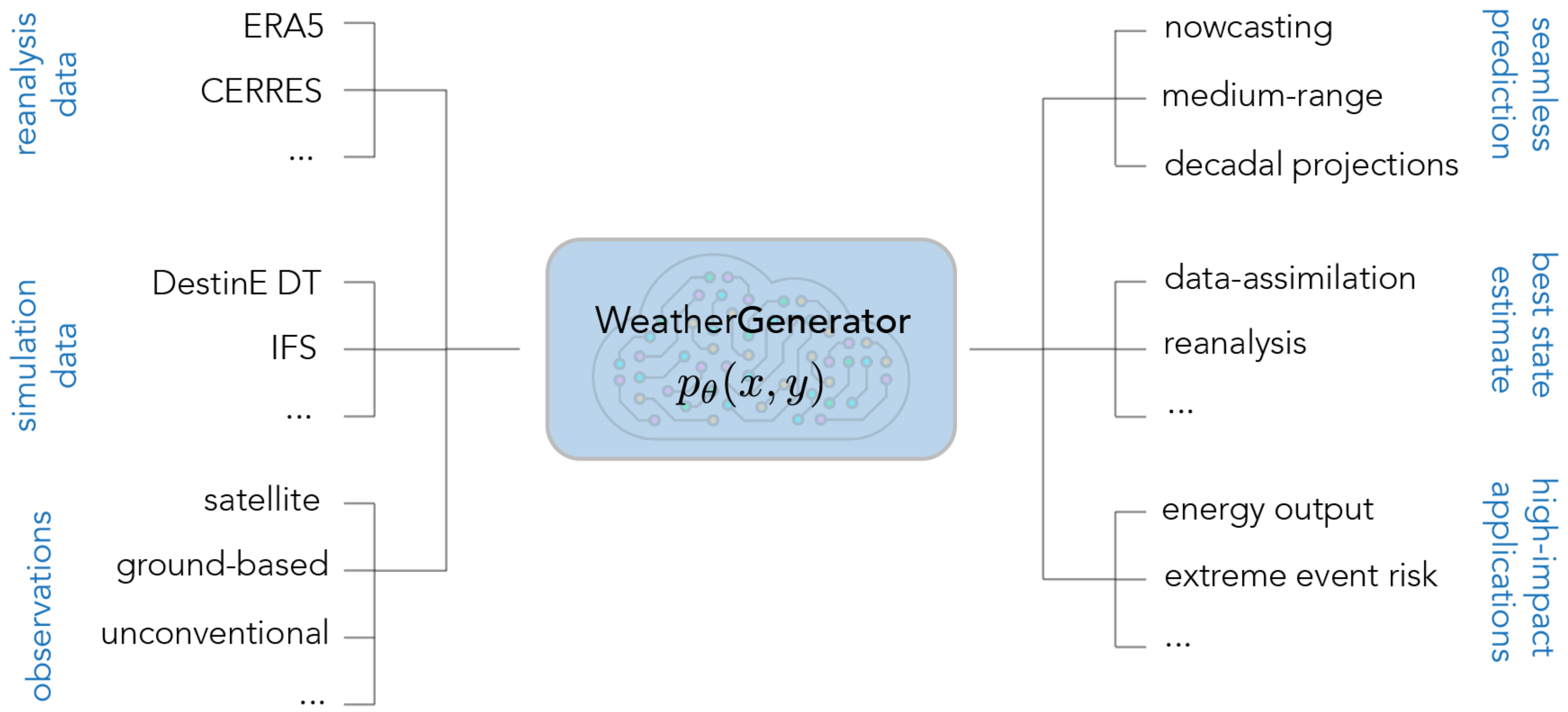
TRAINING



https://arxiv.org/abs/2412.15832

A. McNally, M. Alexe et al. (2024)

Weather Generator



Exploring the next generation of ML models.
Single foundation models across time, space and task

https://www.ecmwf.int/en/about/media-centre/news/2024/weathergenerator-project-aims-recast-machine-learning-earth-system

Where are we?

- For headline scores, data-driven models are best.
- Don’t represent all the spatial scales correctly when trained deterministically.
  - Still useful, despite this, and probabilistic framing solves this.
- Extreme events
  - AIFS Single underpredicts extreme events but still useful. AIFS CRPS and IFS-ENS similar scores

Where is the field going?

- Earth system data-driven models
  - Capture land, ocean and more processes.

- Extended range predictions, pushing beyond 2 weeks.
- Higher resolution
- Use of observations to predict the future state.
  - Incorporate data-assimilation into the training.

- Collaboration between ECMWF and MS on data-driven models.
  - Opportunities to share code/infrastructure whilst still having bespoke models.

What do you think the future will hold?

Anemoi

Open source ML software framework for earth system modelling. Underpins AIFS, DestinE AI activities and more activities across Europe.

Open recipes for training the AIFS and open models.

<u>Developed</u> and <u>used</u> by meteorological centers across Europe.
AEMET, DWD, FMI, GeoSphere, KNMI, MET Norway, Meteo Swiss, Meteo France, RMI, & ECMWF

Pooling of resources without resulting in a single forecasting model.