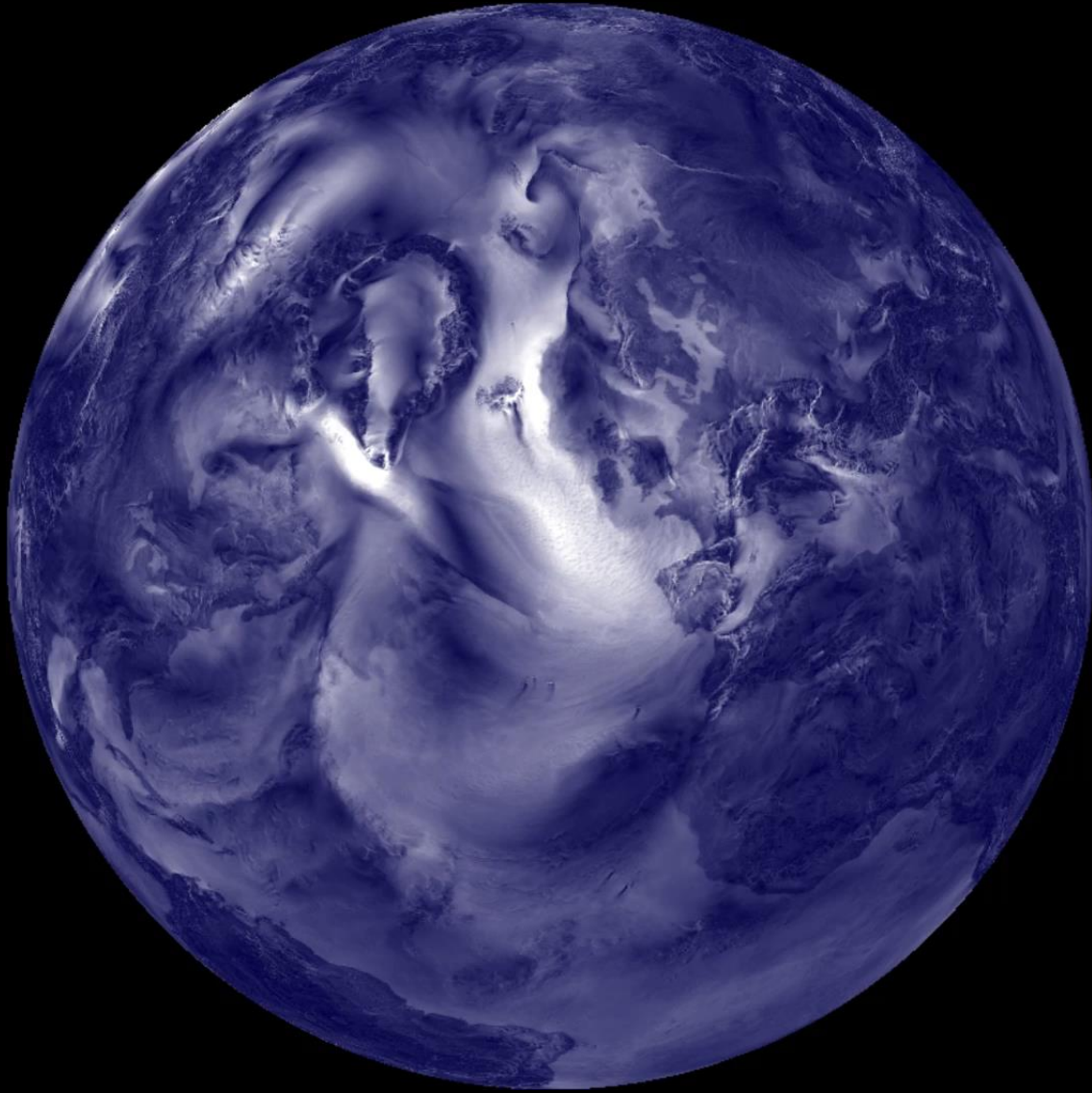


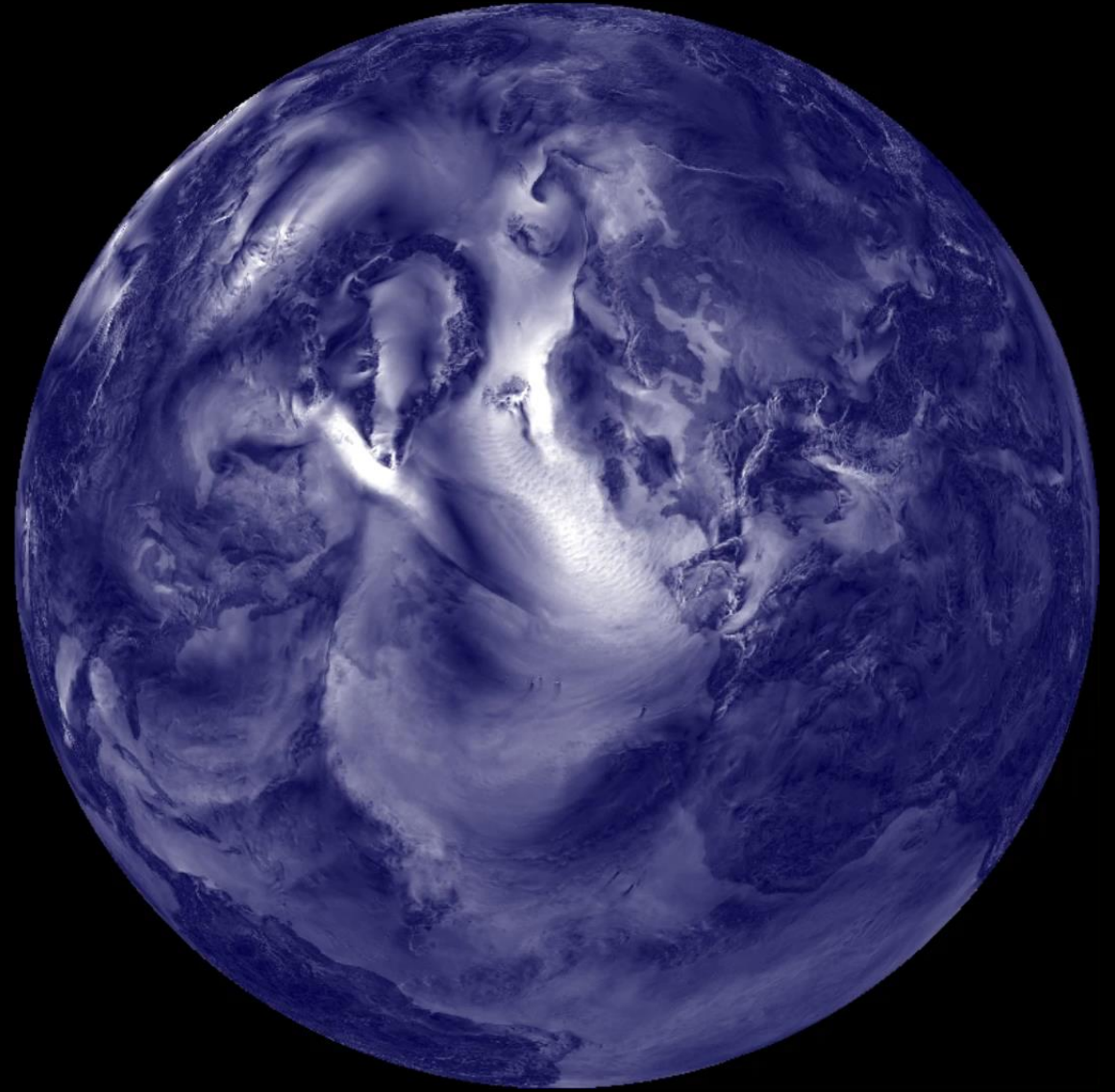
Ensemble forecasting with AIFS

Simon Lang

IFS 10m wind gusts, 2020-12-04 00 UTC 720h forecasts, 9 km spatial resolution



Control Member

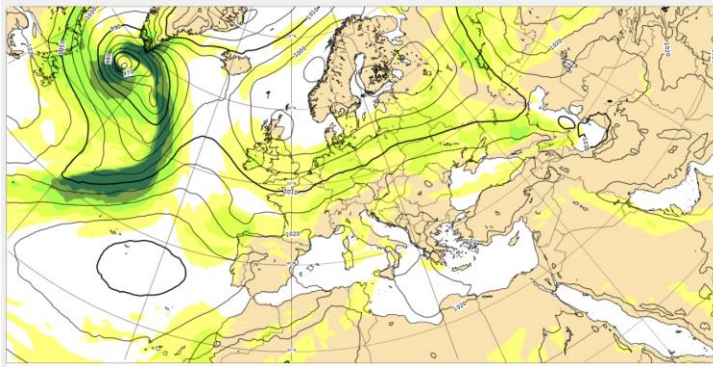


Perturbed Member 1

Weather Forecasts – NWP? Data Driven?

Traditionally weather forecasts are generated by running a NWP model – computer code that has been designed to represent the physical processes governing the evolution of the atmosphere. But can you produce a forecast without a NWP model?

Analysis

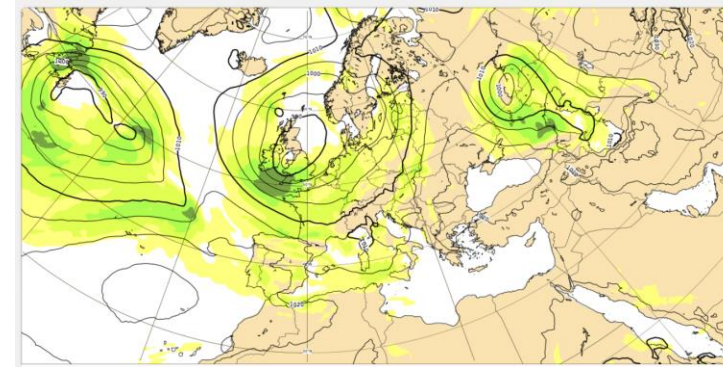


Fusion of a short-range forecast with the latest observations

NWP Model



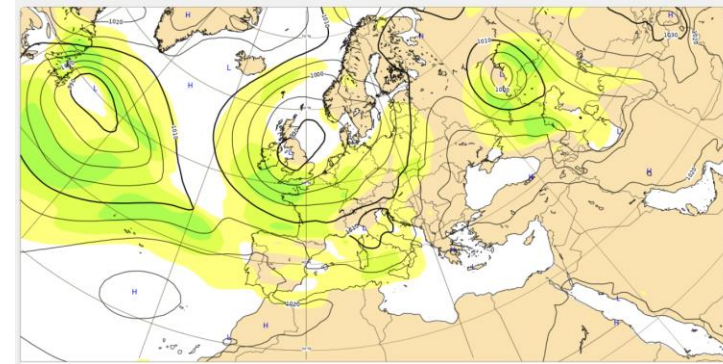
Forecast



Data-driven Model



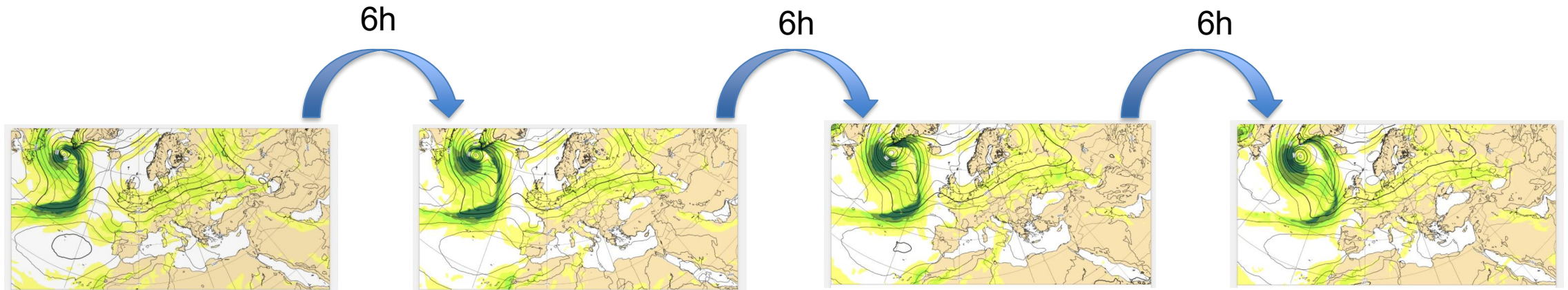
Learned from 40 years of analyses



Weather Forecasts – NWP? Data Driven?

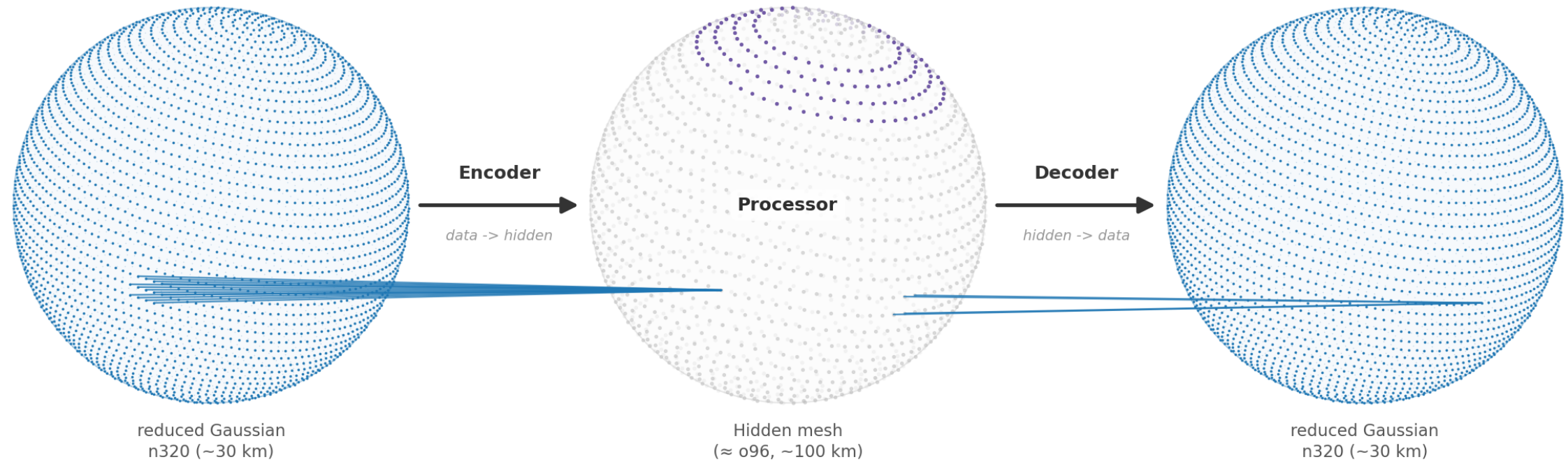
Recent advances show that this is possible ...

Here, the model learns from ca. 40 years of ERA5 reanalysis data, stepping, e.g., 6h from analysis to analysis

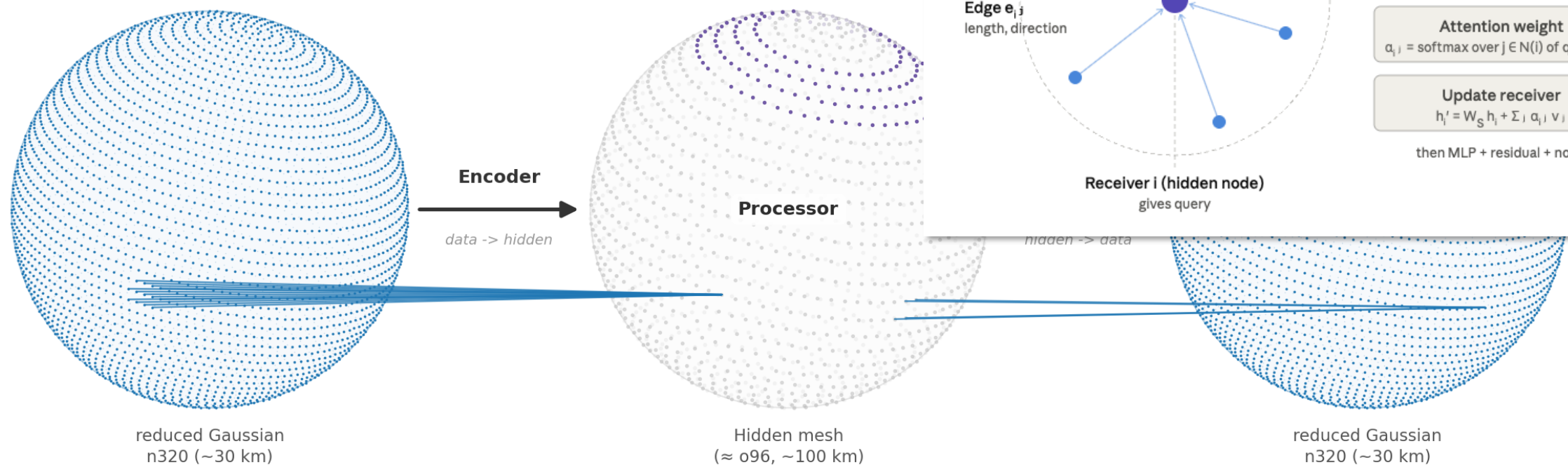


The forecast is then autoregressively stepping 6 h into the future $x_n = f(x_{n-1})$...

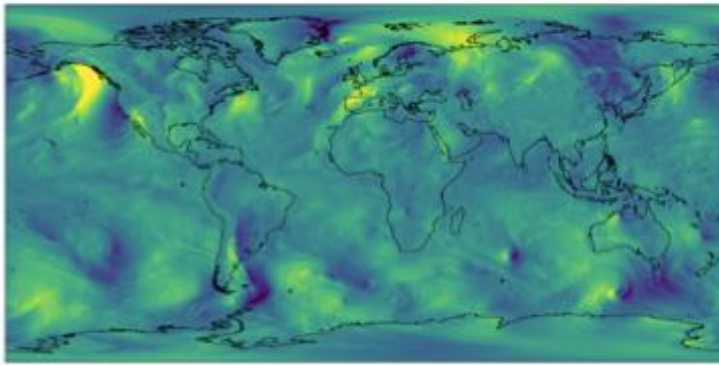
Encoder-Processor-Decoder Architecture



Encoder-Processor-Decoder Architecture

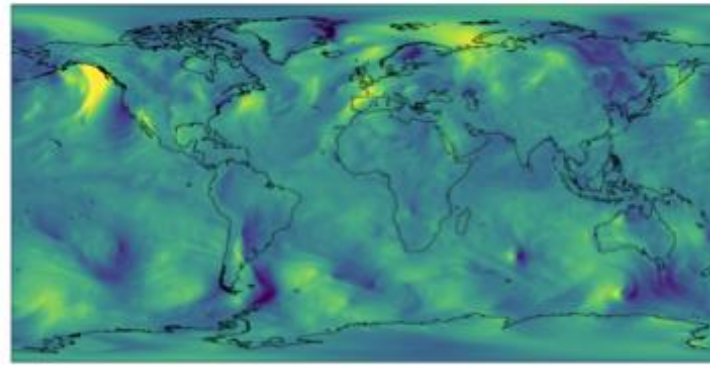
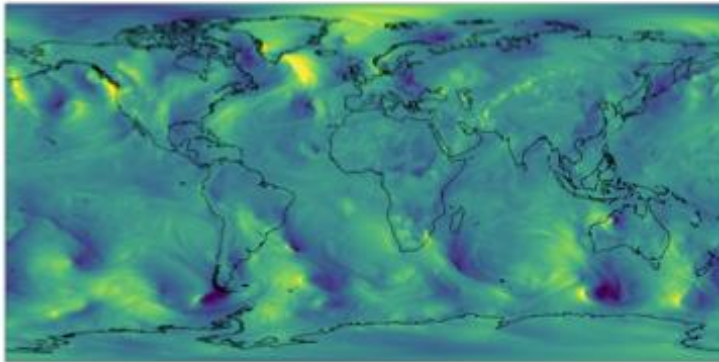


IFS

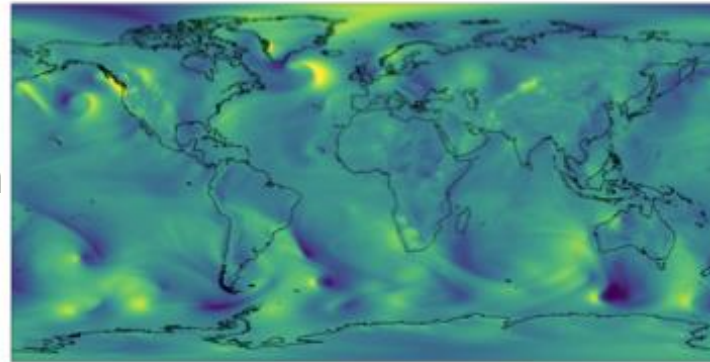
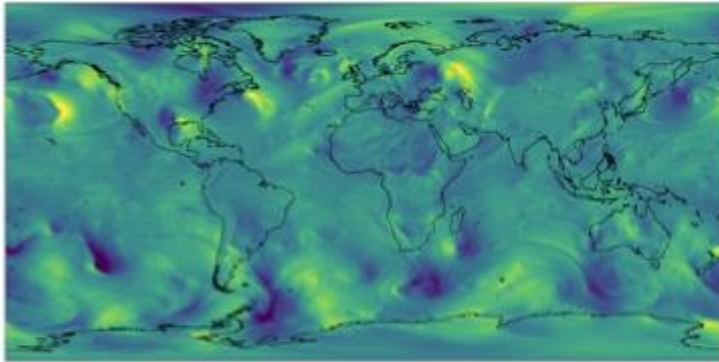


12 h

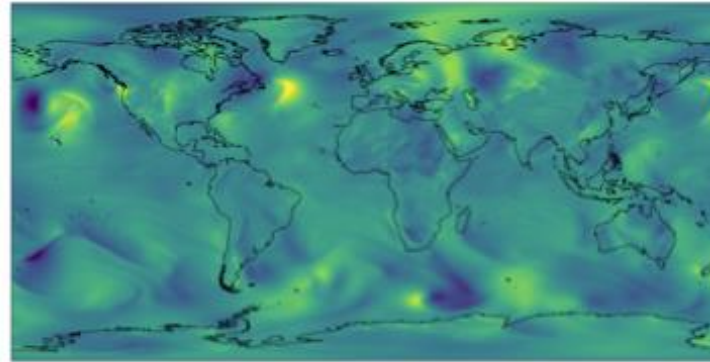
AIFS



120 h



240 h



Smoothing due to MSE
optimization –
deterministic training



Meridional wind at 850 hPa: IFS (left) and AIFS (right) for 1 January 2023 00 UTC date, 12 h (top), 120 h (middle) and 240 h (bottom) forecasts. For plotting, IFS and AIFS forecast fields were interpolated to a 0.25° regular latitude-longitude grid. The MSE objective used in AIFS training leads to more smoothing at longer lead times.

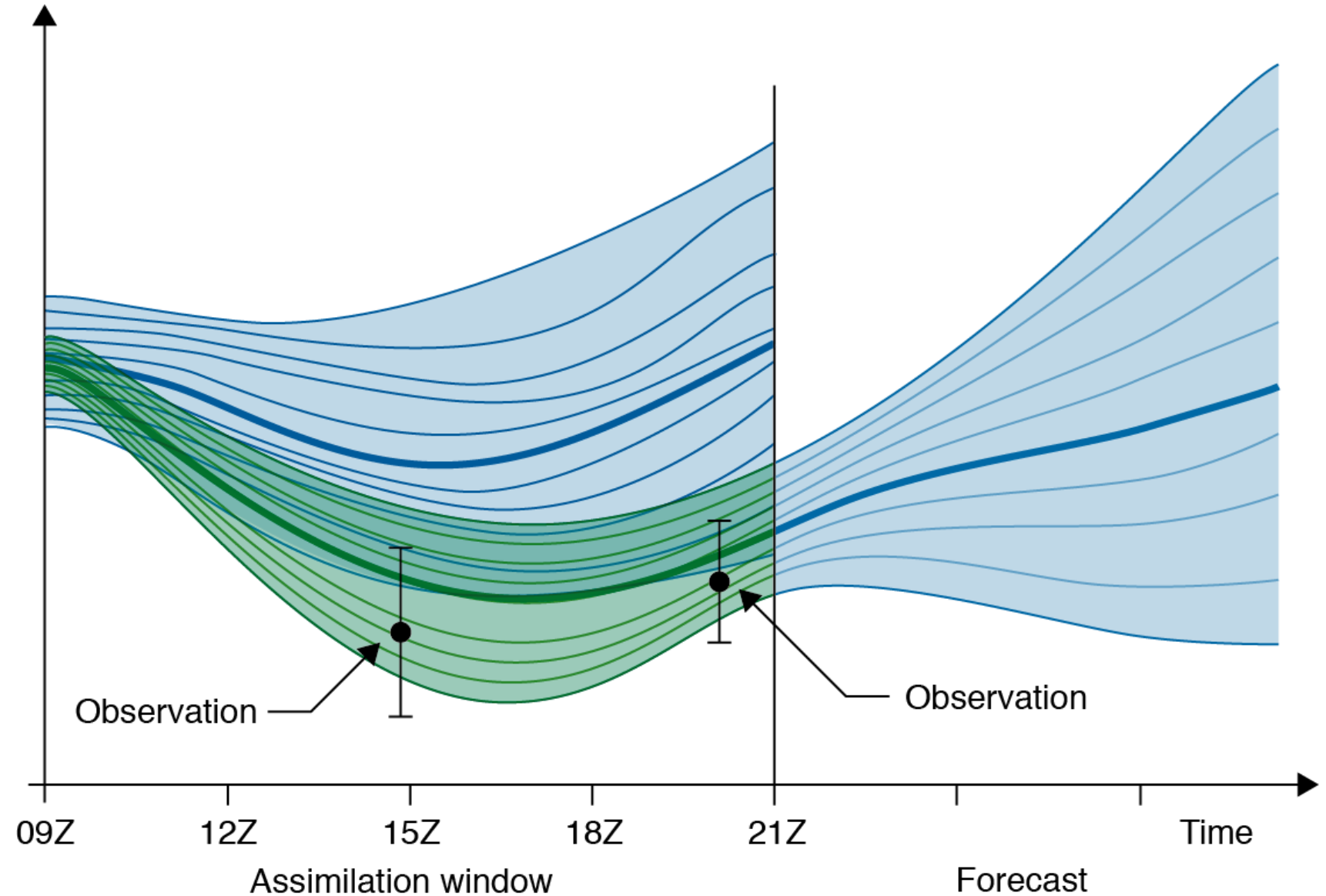
Ensemble Data Assimilation

Ensemble Forecast

NWP

Add perturbations to
physics-based forecast
model

Build an ensemble of
analyses states and
forecast states

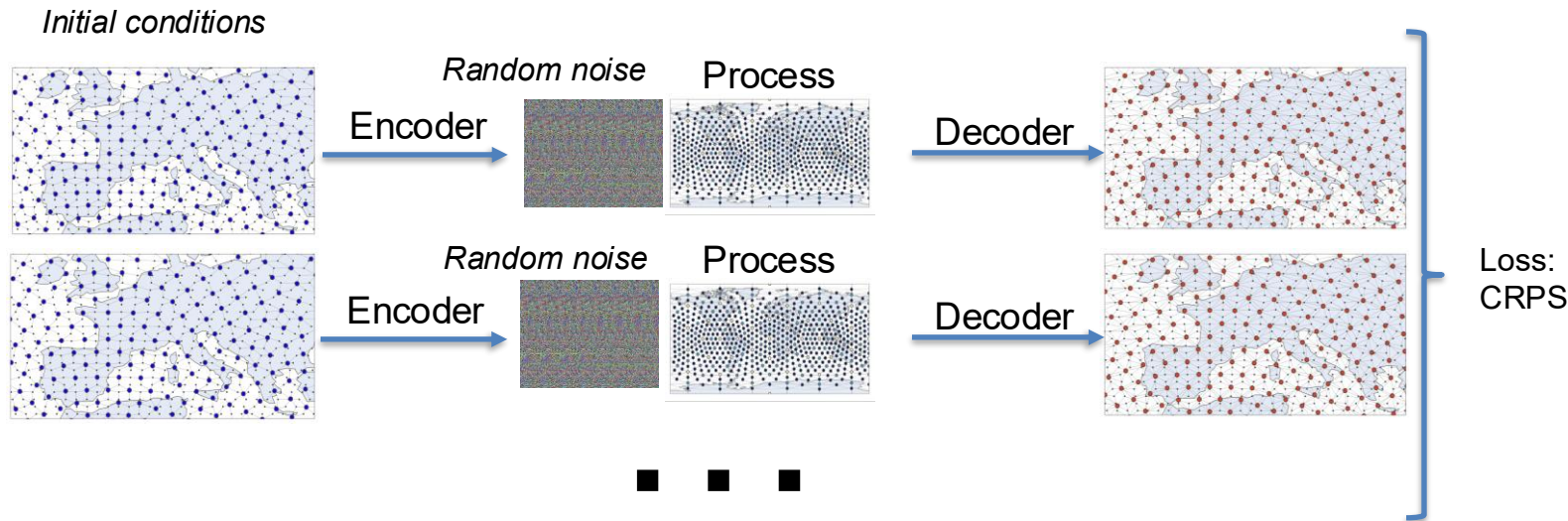
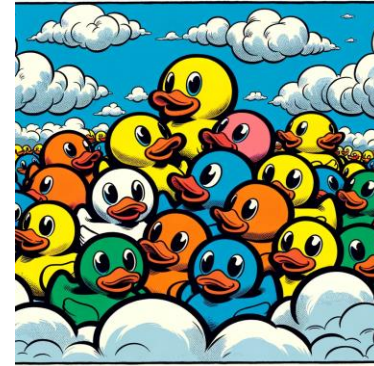


Ensemble forecasts ...

a) Instead of an MSE loss, learn an ensemble via optimizing probabilistic scores

for example:

AIFS-CRPS

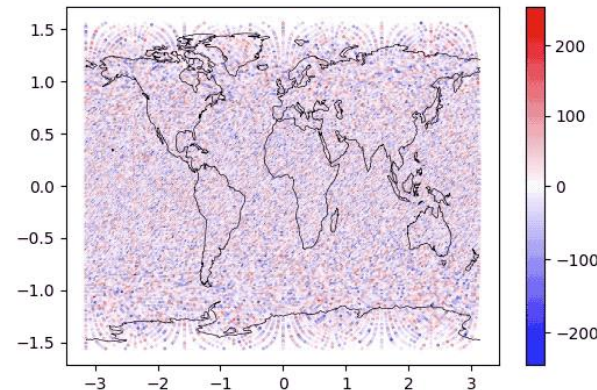


<https://www.nature.com/articles/s44387-026-00073-7>

b) Score matching via denoising (diffusion training)

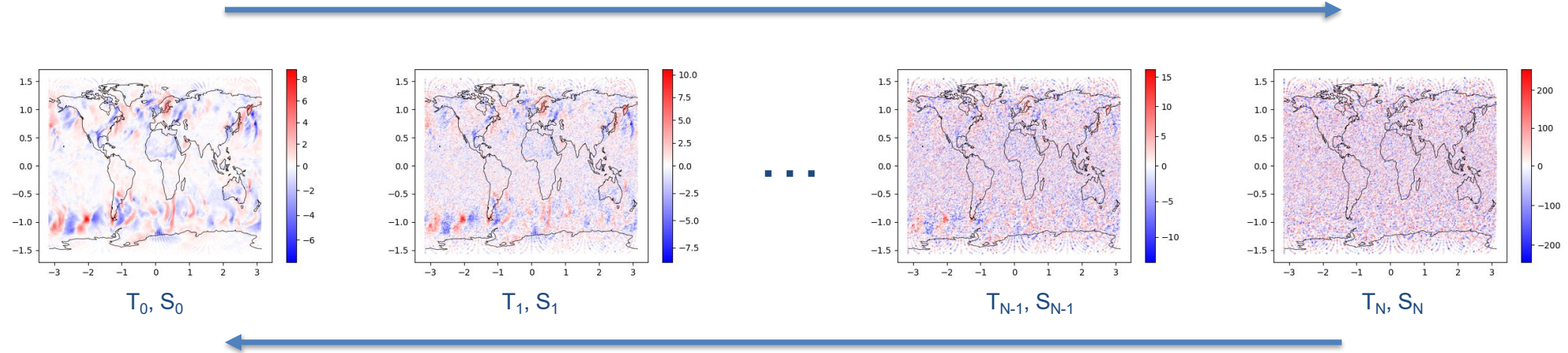
for example:

- Stable Diffusion -> Images
- Sora -> Video
- GenCast, AIFS-Diffusion -> Weather



<https://www.nature.com/articles/s41586-024-08252-9>

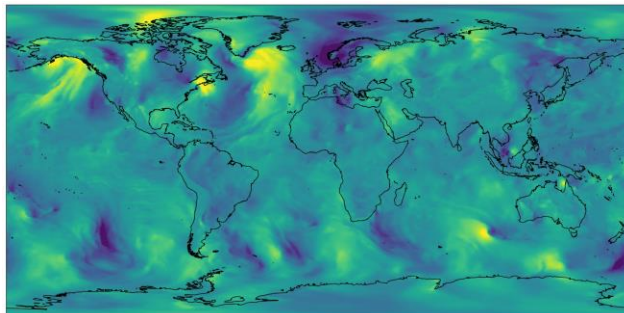
Forward process: add noise



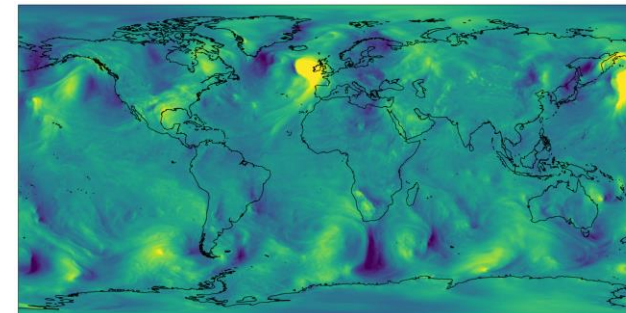
Reverse process (learned): remove noise

To generate a forecast (single member): start from pure noise with initial conditions, run the model, e.g. 40 times, to slowly create a forecast tendency step by step ; repeat for the next forecast step etc. -> 12 h step, 360 h lead time, call the model 1,200 times

AIFS Diffusion



AIFS MSE



CRPS-based training -> AIFS-ENS

Probabilistic training of AIFS:

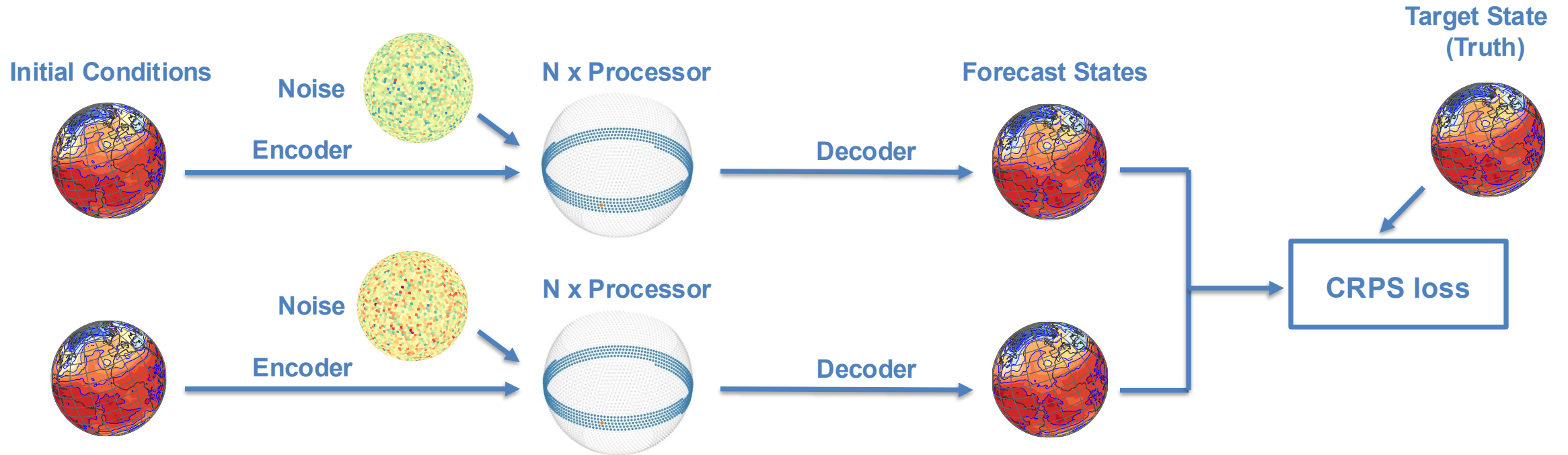
$$\text{afCRPS}_\alpha := \alpha \text{fCRPS} + (1 - \alpha) \text{CRPS}$$

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

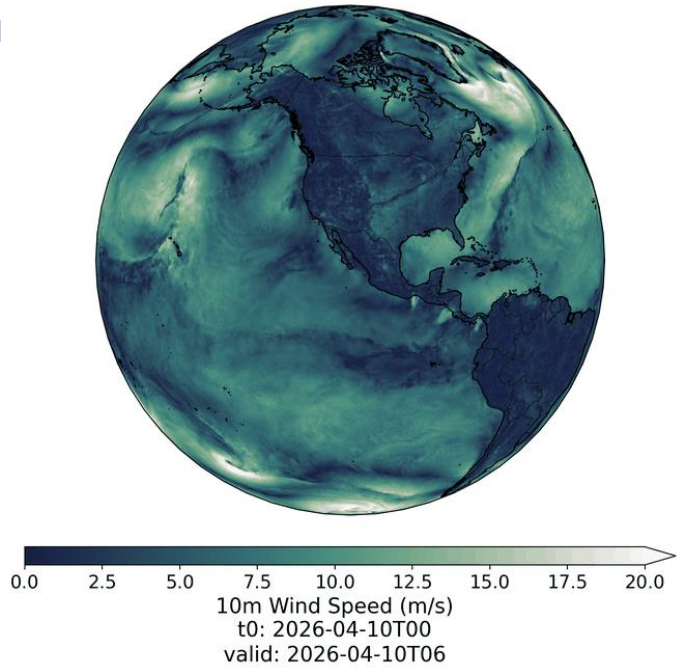
$$= \frac{1}{M} \sum_{j=1}^M |x_j - y| - \frac{1-\epsilon}{2M(M-1)} \sum_{j=1}^M \sum_{k=1}^M |x_j - x_k|$$

Members should be close to the truth

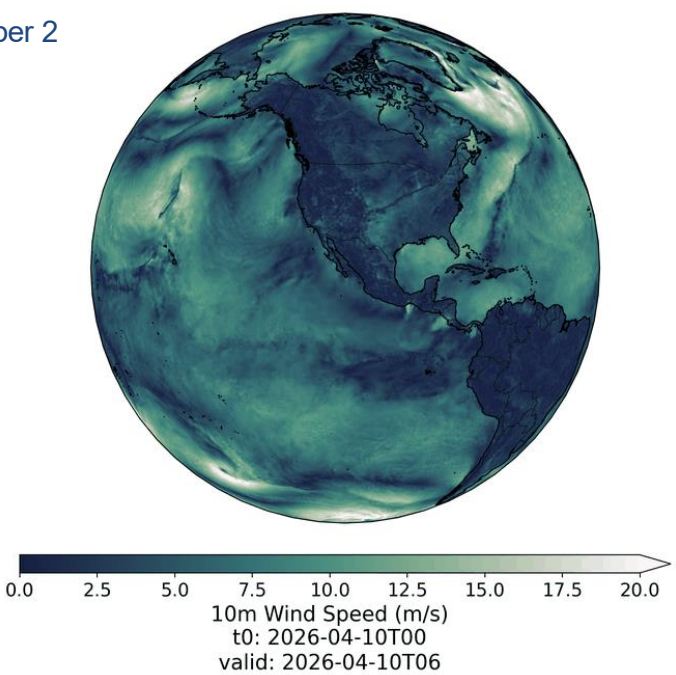
Differences between ensemble members should represent the forecast uncertainty



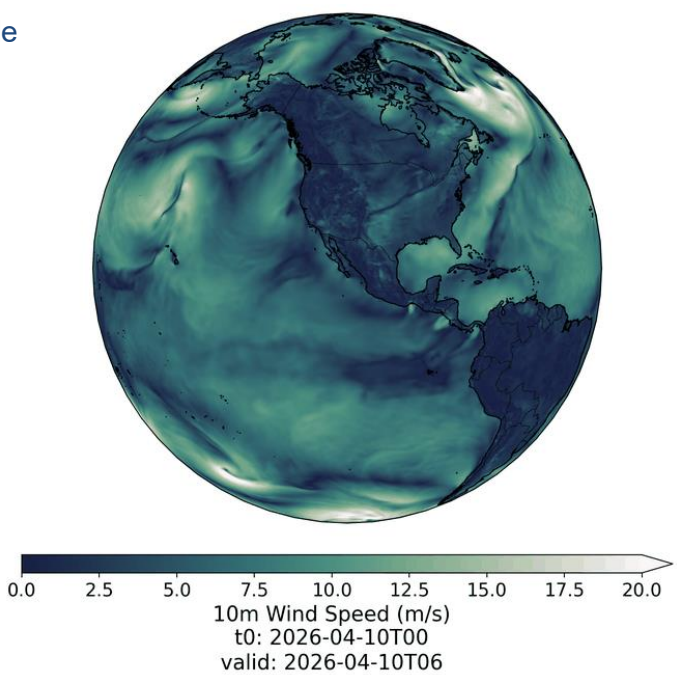
Member 1



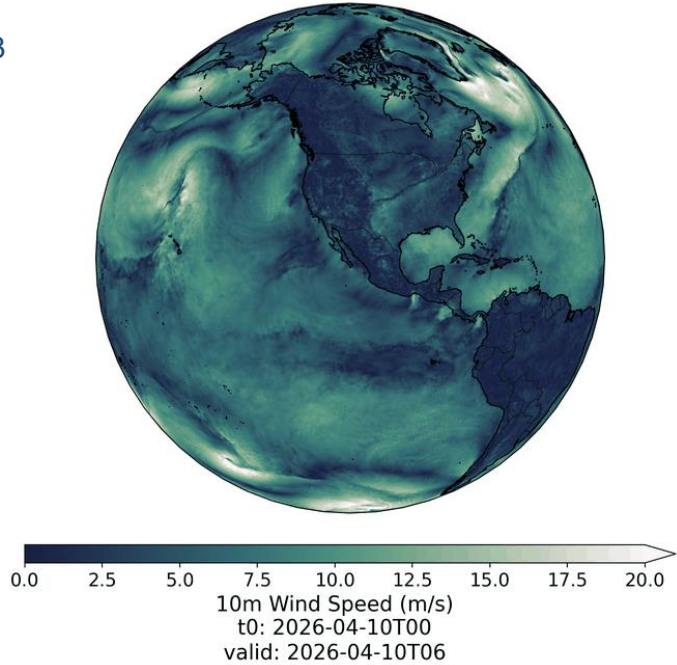
Member 2



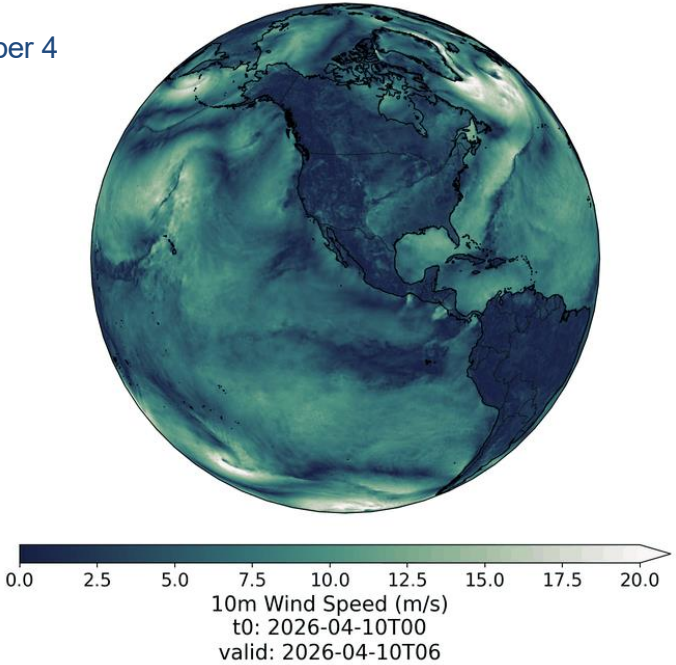
Single



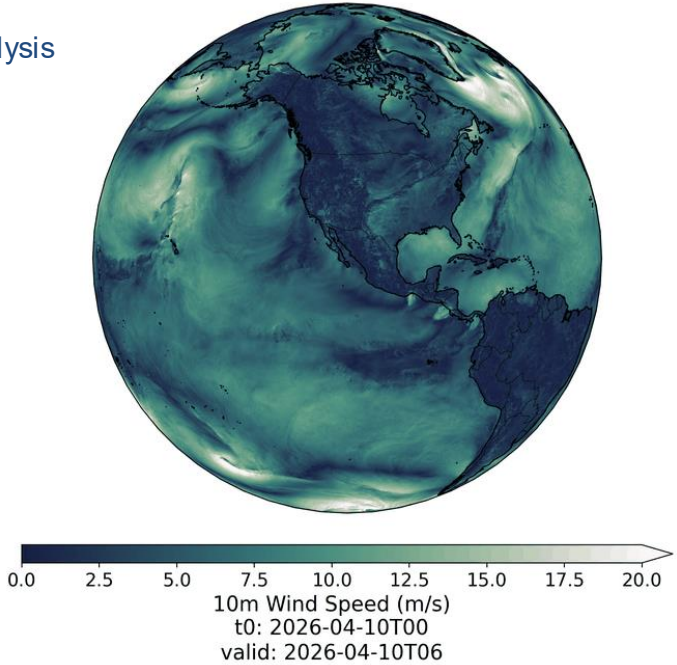
Member 3



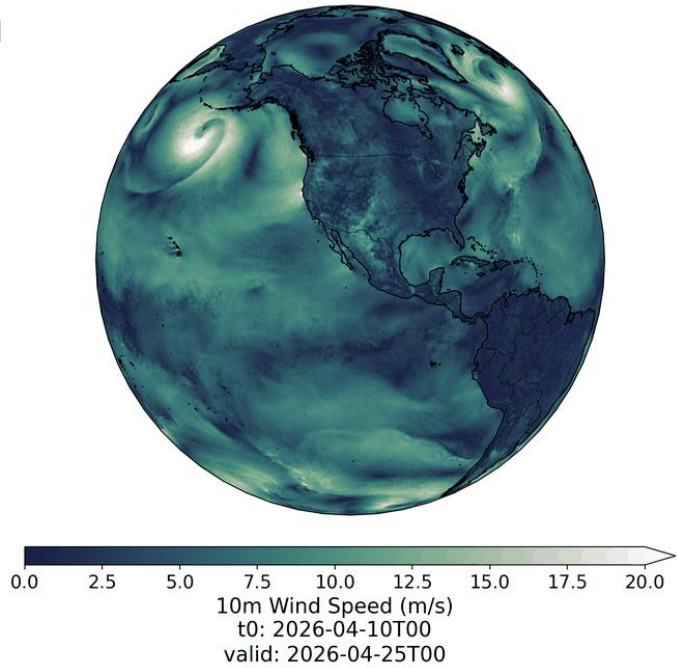
Member 4



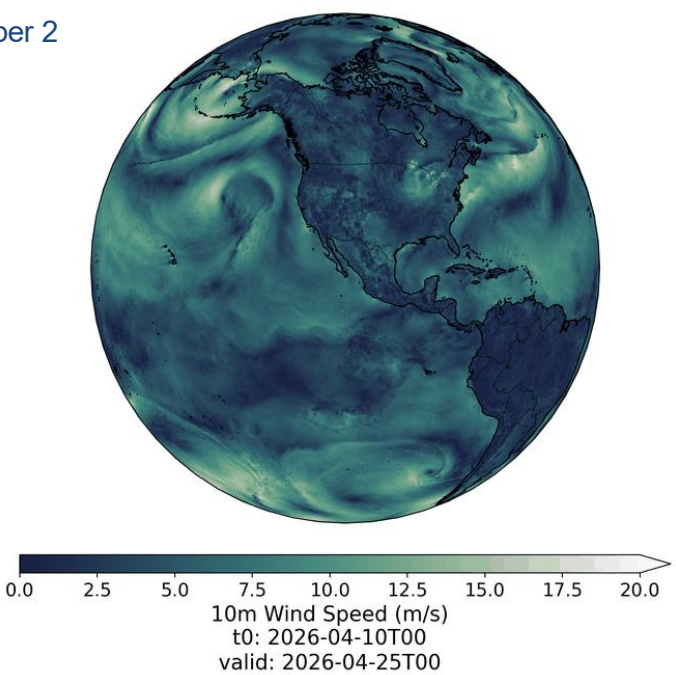
Analysis



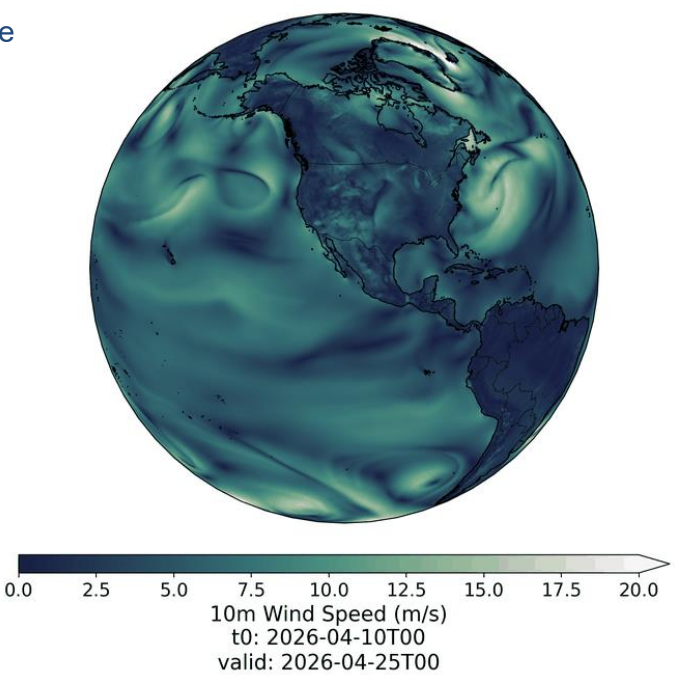
Member 1



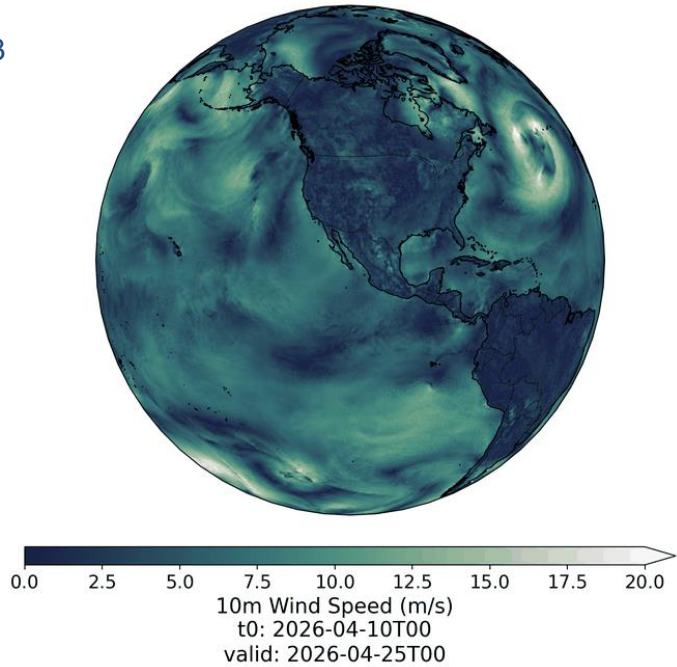
Member 2



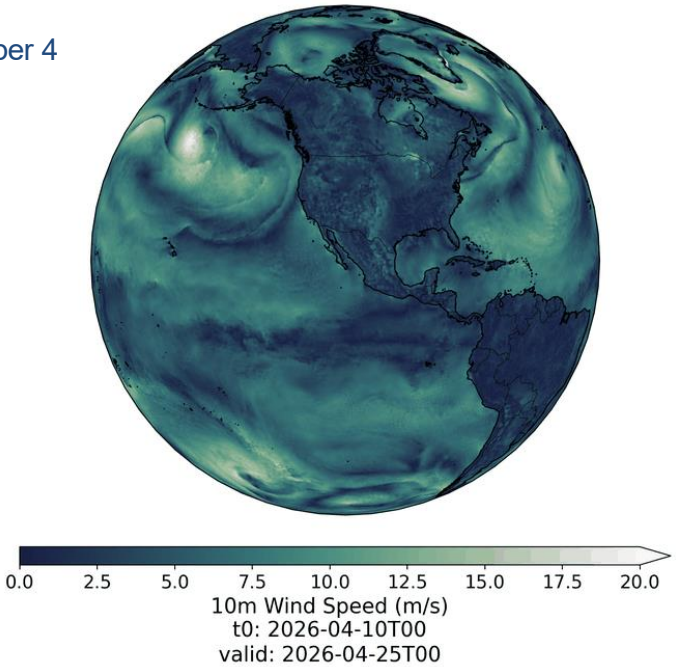
Single



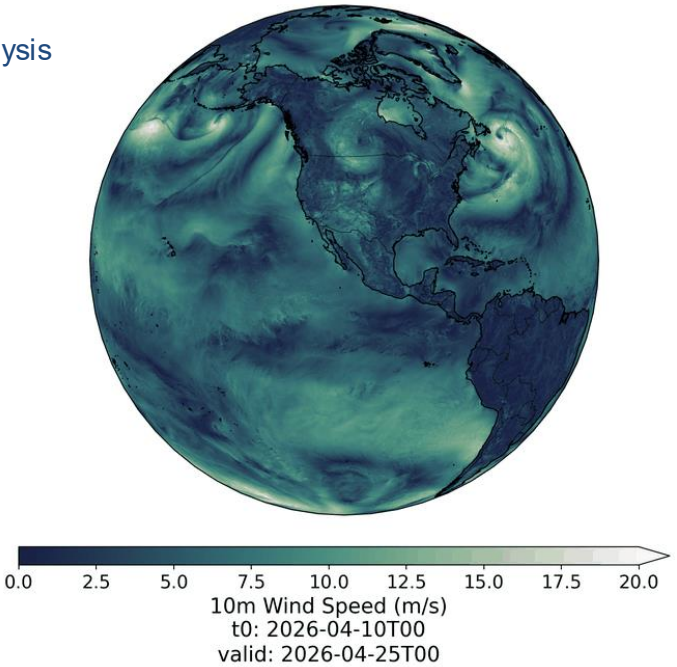
Member 3



Member 4

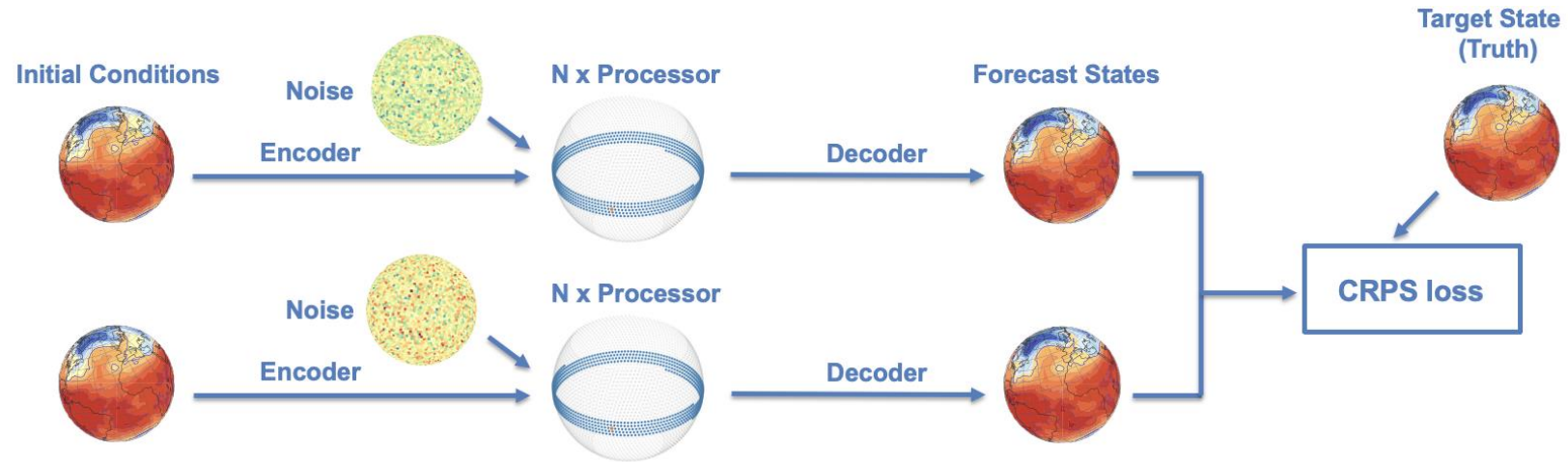
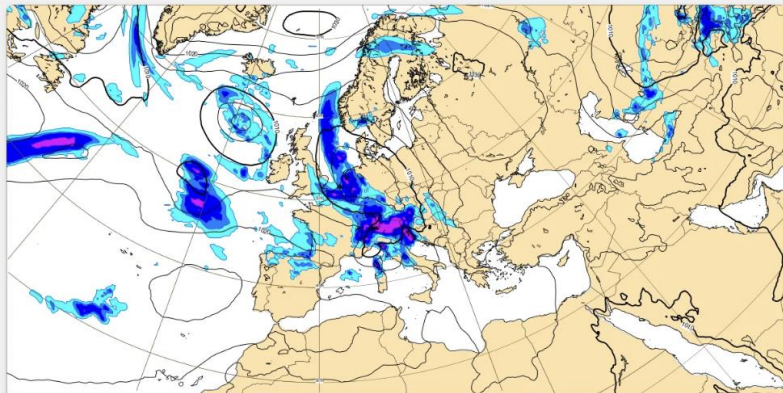


Analysis

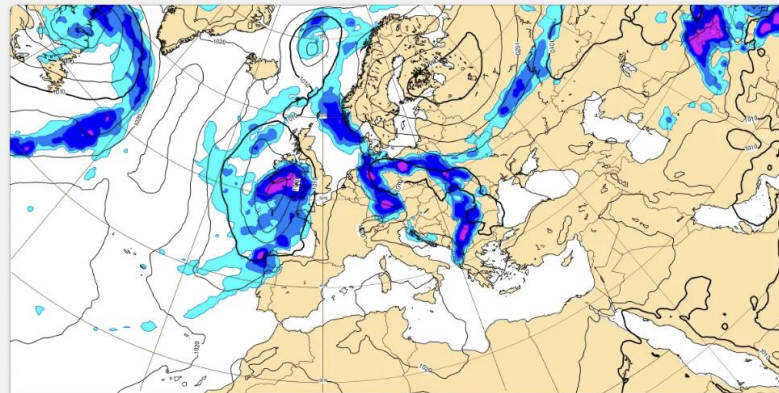


AIFS-ENS:

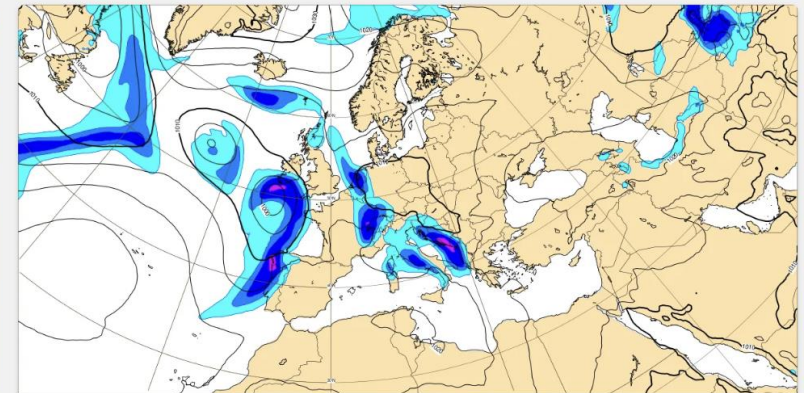
Probabilistic training of AIFS:



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

Make the loss scale-aware:

CRPS is computed point-wise on the full output field. However, atmospheric processes are inherently multi-scale, and different scales contribute to a different degree to the loss function -> introduce a scale-aware loss

Scale-unaware: $\mathcal{L} = c \int_{\mathcal{M}} \mathcal{S}([x_j \mid j = 1, \dots, M], y) d\mu$ scoring rule $\mathcal{S}$ for scalars

Scale-aware: $\mathcal{L}_{n\text{-scale}} = \sum_{i=1}^n \zeta_i c \int_{\mathcal{M}} \mathcal{S}([x_{j,\text{scale } i} \mid j = 1, \dots, M], y_{\text{scale } i}) d\mu$

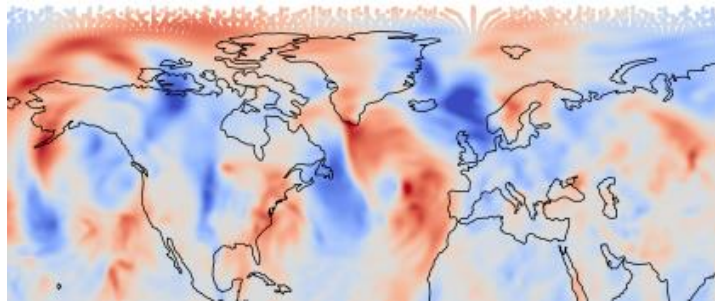
$$\phi_{\text{scale } 1} = D_1(\phi)$$

$$\phi_{\text{scale } 2} = D_2(\phi) - D_1(\phi)$$

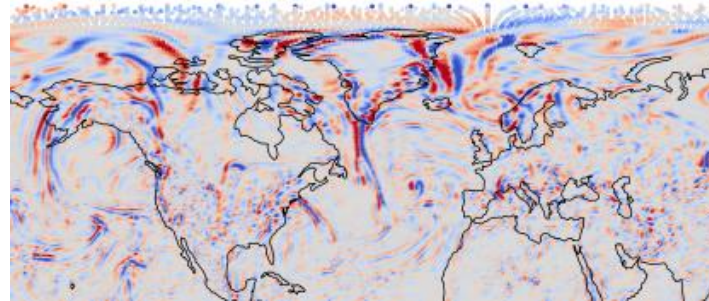
$$\vdots$$

$$\phi_{\text{scale } n} = \phi - D_{n-1}(\phi)$$

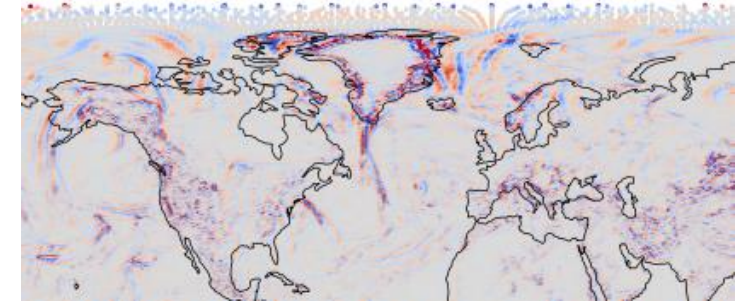
With ordered smoothing operators D_i that partition of a function ϕ on the manifold M into n scales:



Scale 1



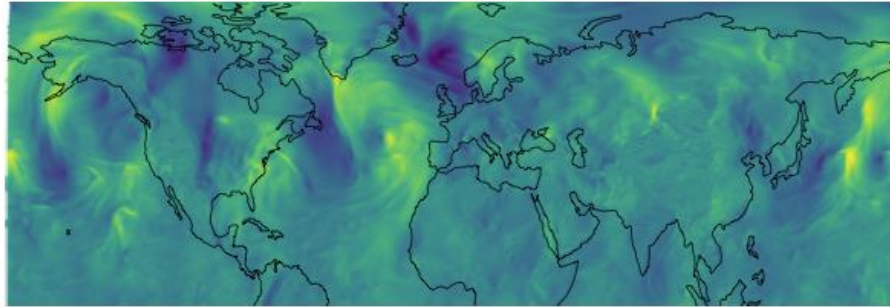
Scale 2



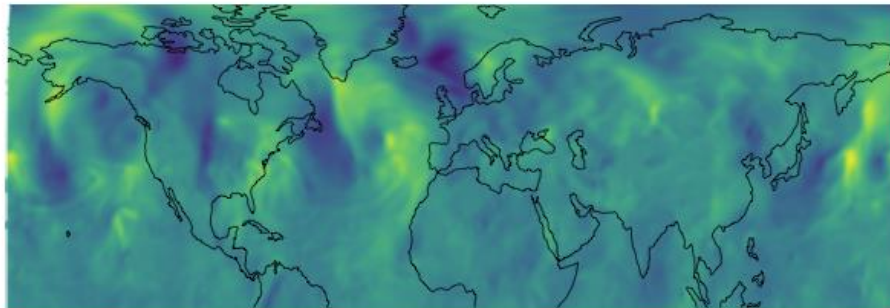
Scale 3

Multi-Scale Loss removes noise from predictions (here two scales):

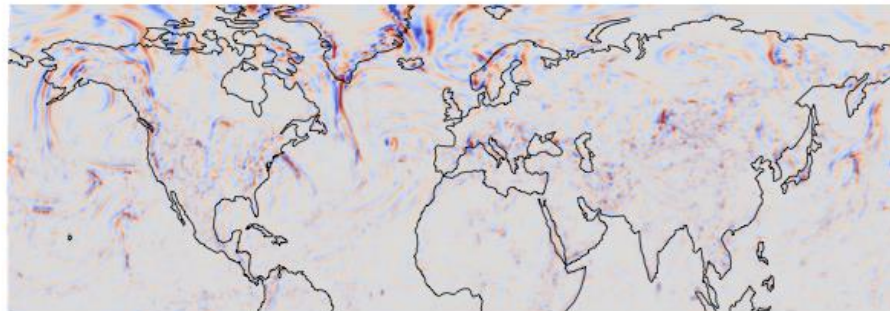
(a) target



(b) target, scale 1

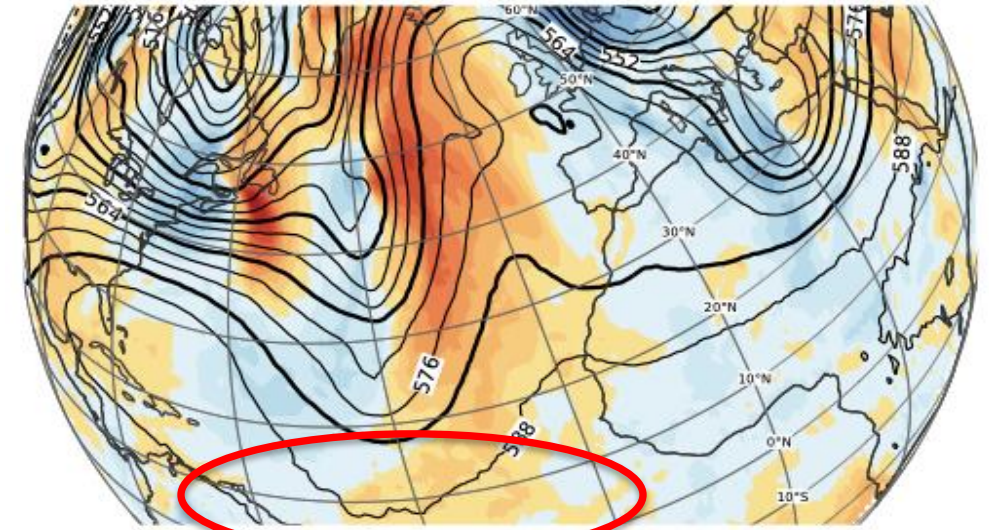


(c) target, scale 2

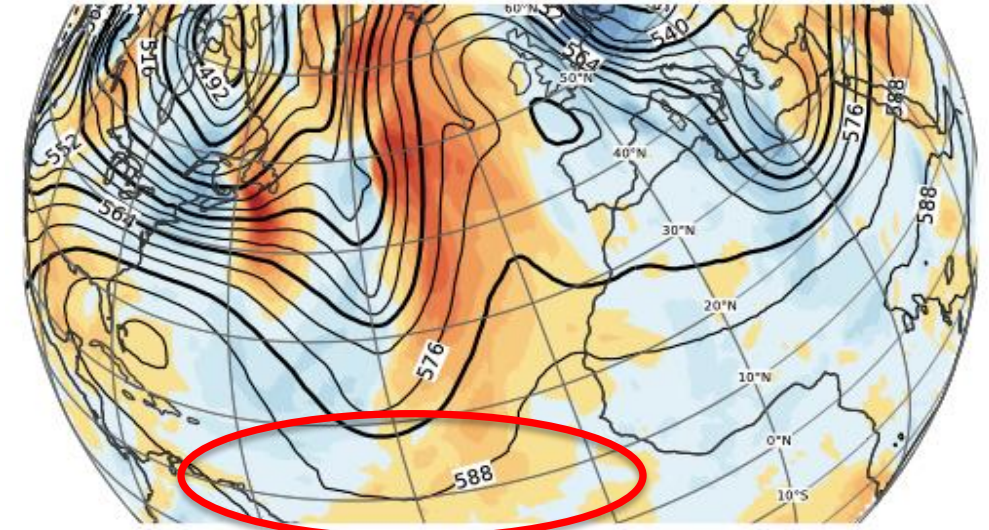


ERA5 v-component of wind (in m s^{-1}) at 850 hPa, full field (a), field after filtering (b) and difference between filtered and full field (c).

(a)

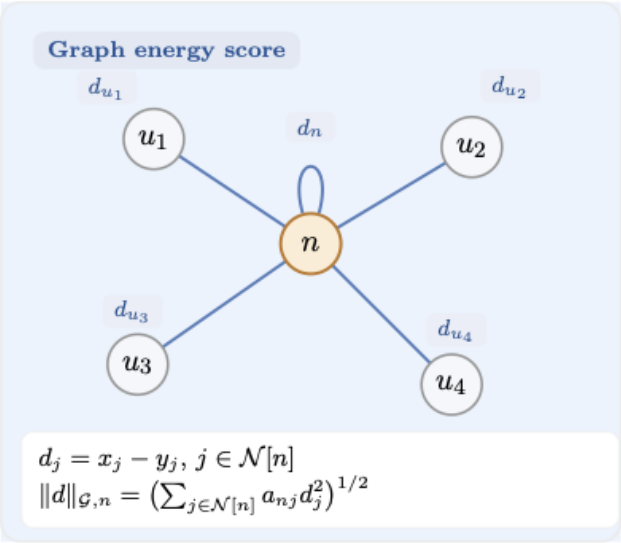
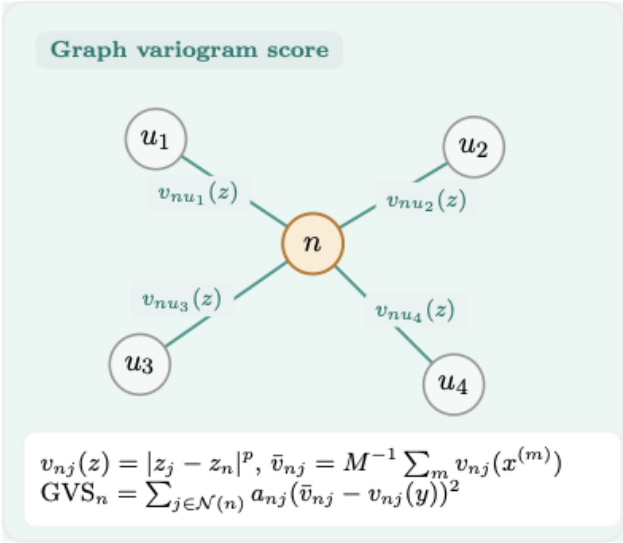
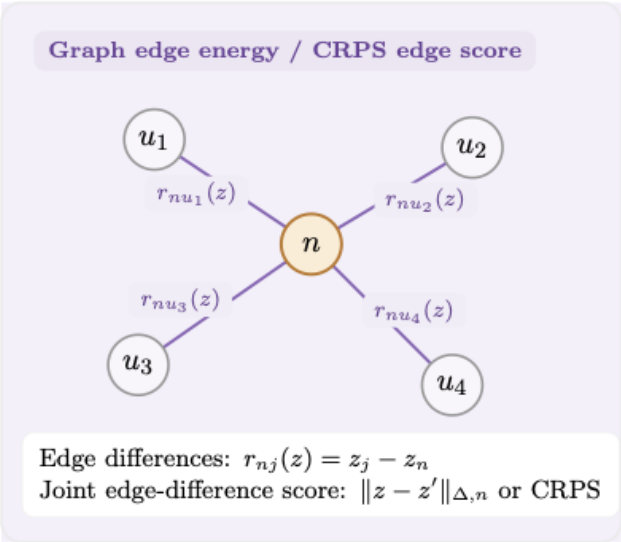


(b)



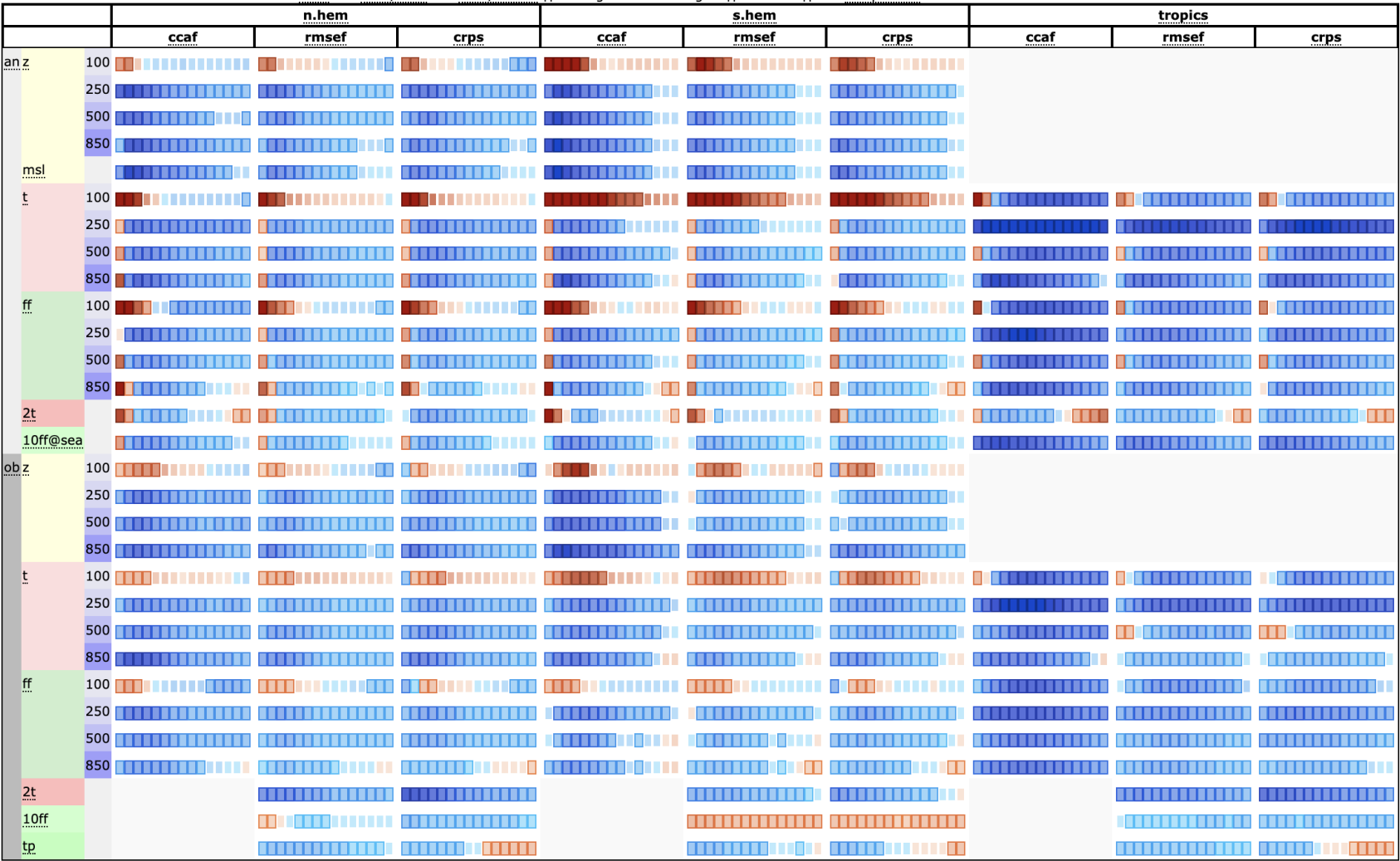
Geopotential at 500 hPa (dam, contours) and v-component of wind at 850 hPa (shaded) of (a) the scale-unaware loss experiment, (b) the multi-scale loss experiment
Shown are 24 h forecasts of member 1, initialised on 2019-01-01 00 UTC (a, b) and verifying analysis on 2019-01-02 00 UTC (c).

Proper scores defined on graphs:



AIFS-CRPS Medium-Range evaluation

AIFS-CRPS N320 vs IFS ENS, 20240201 – 20240930, 00, 12 UTC, 50 members



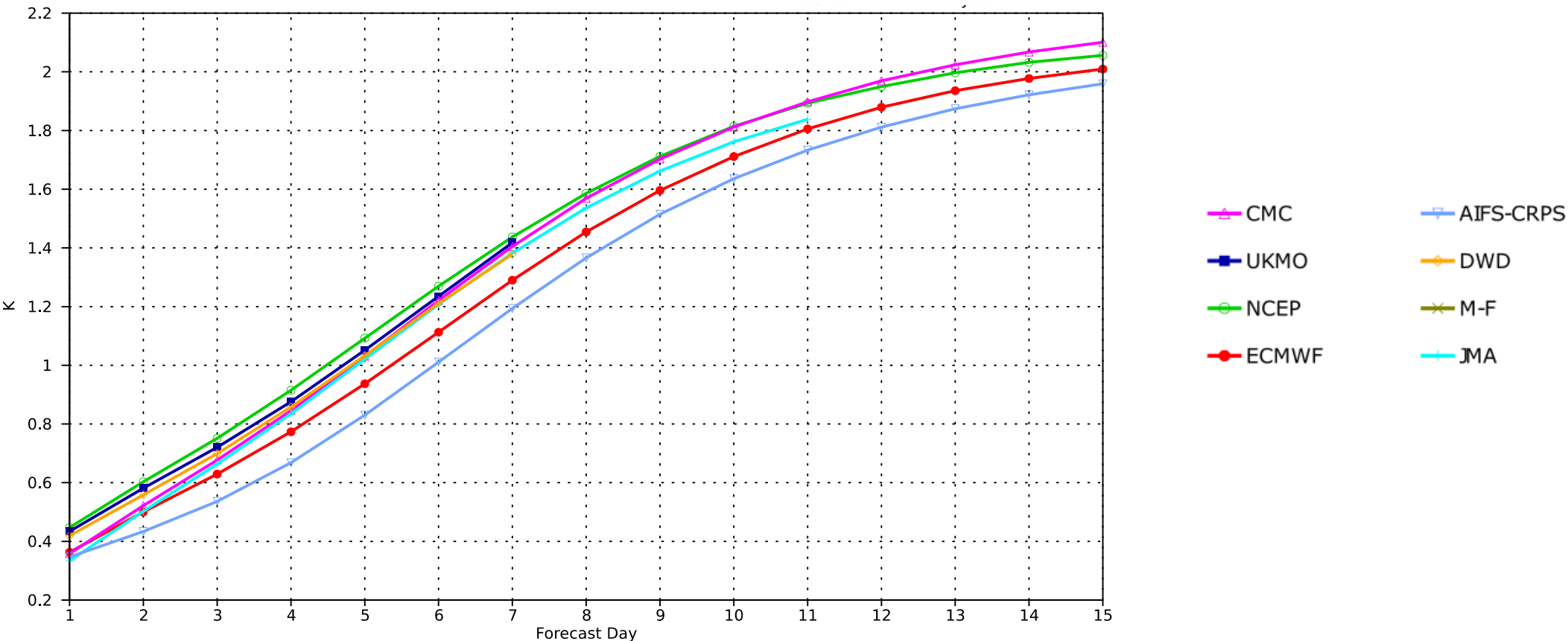
AIFS-ENS, comparison with TIGGE ensembles

Continuous ranked probability score | 850hPa temperature

NHem Extratropics

20240201 00z to 20240930 00z | enfo mean_standard |

Population: 484,3*483,2*482,481,480,479,478,477,474,473,472,471 (averaged)



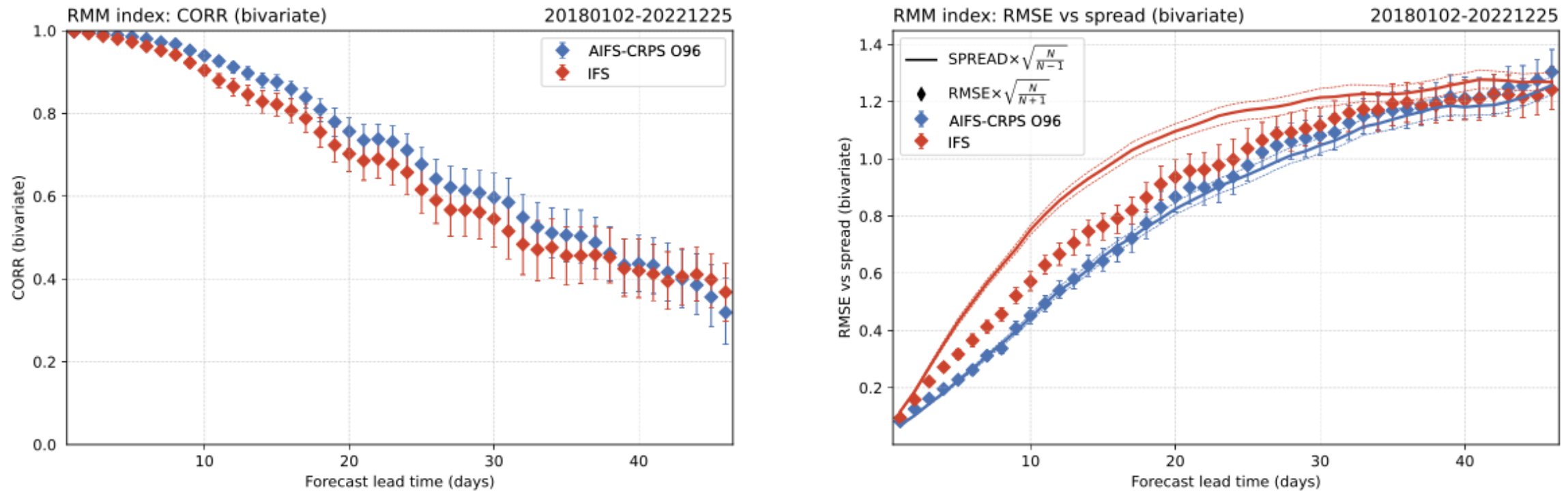


Figure 11: (a) Bivariate correlations for an MJO index calculated from 200 hPa and 850 hPa zonal wind anomalies for AIFS-CRPS (blue) and operational IFS reforecasts run in 2023 (red). The MJO index used here is an approximation for the full Wheeler and Hendon [2004] Real-time Multivariate MJO index as it excludes contributions from outgoing longwave radiation that are not available from AIFS-CRPS. For both systems, correlations are calculated with respect to the same indices calculated from ERA5. Error bars represent the 2.5th and 97.5th percentiles of the distribution created by block-bootstrap resampling of the available start dates. (b) Estimates of root mean square error (RMSE; diamonds) and average ensemble spread (solid lines) for the MJO index described in the text. Spread and RMSE are scaled by factors of $\sqrt{\frac{N}{N-1}}$ and $\sqrt{\frac{N}{N+1}}$, respectively, to ensure estimates are unbiased with sample size (N) as described in Leutbecher and Palmer [2008].

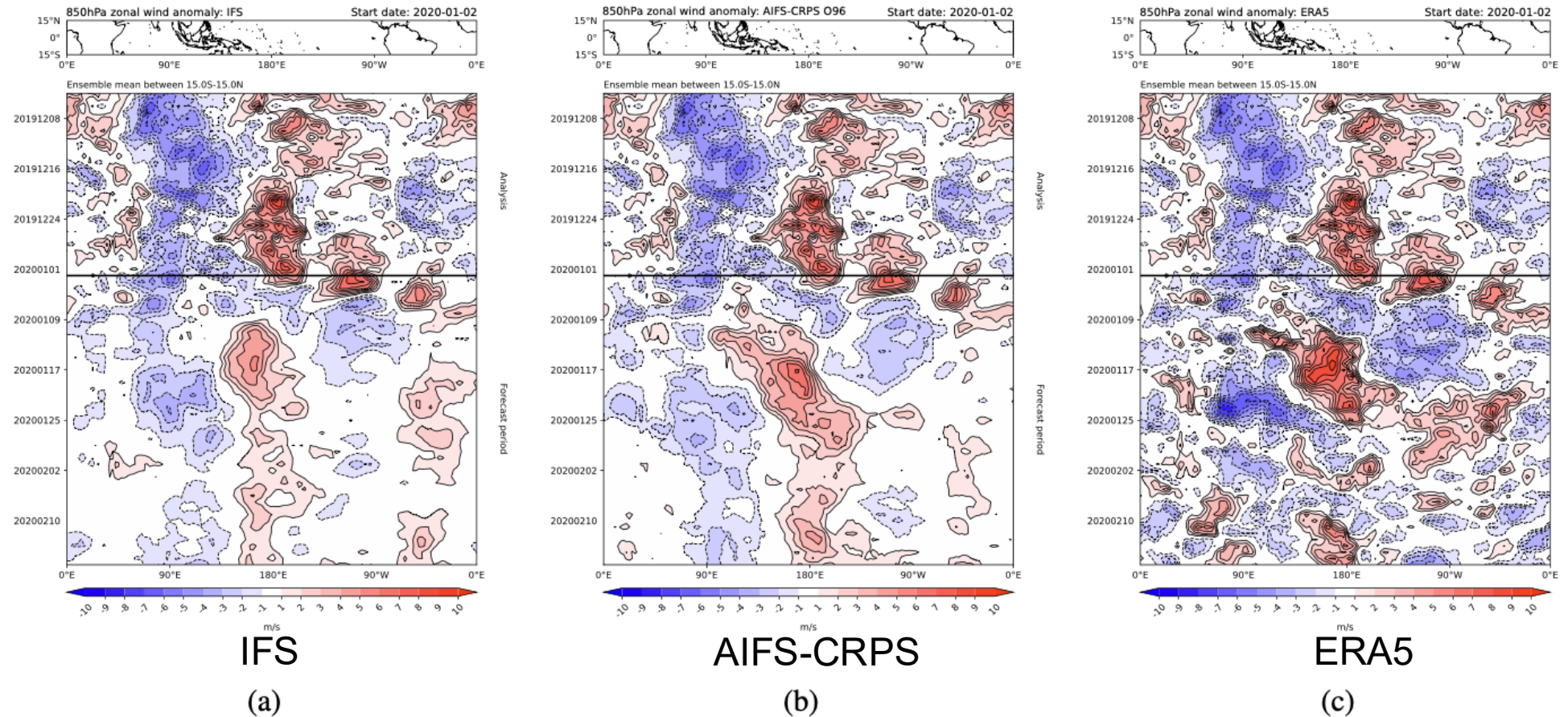
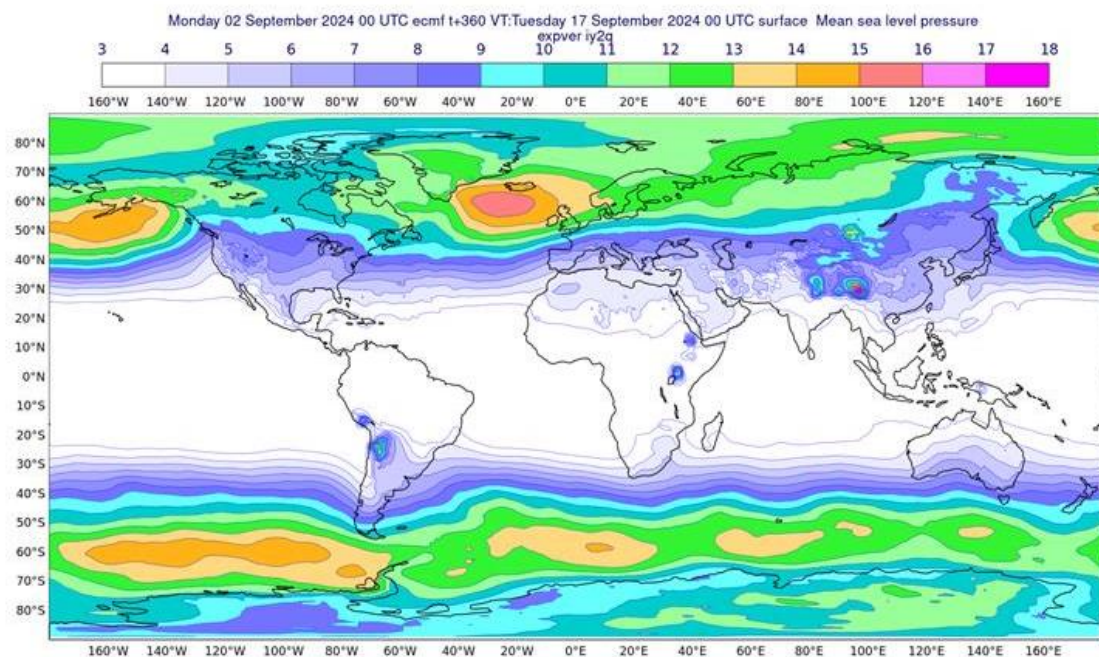


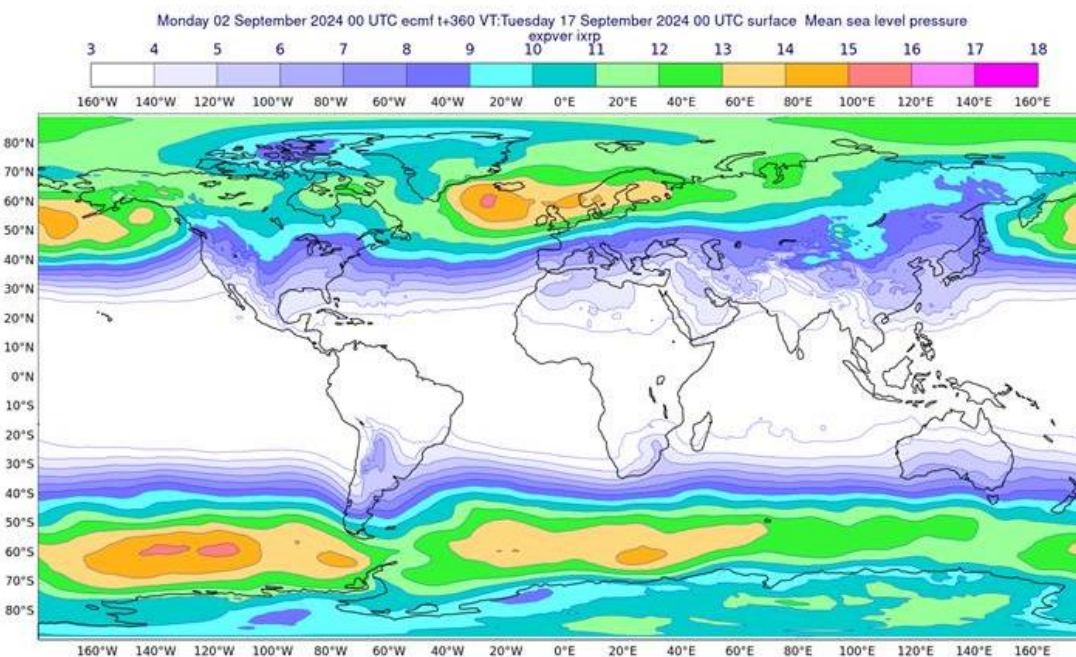
Figure 12: Hovmöller diagrams showing the evolution of zonal wind anomalies at 850 hPa meridionally averaged from 15°S-15°N. All panels show the evolution of zonal wind anomalies in ERA5 for the 30 days prior to the forecast start date (i.e. data above the grey line). Anomalies below the grey line are from (a) IFS ensemble mean forecast initialized on 2020-01-02, (b) AIFS-CRPS ensemble mean forecast initialized on 2020-01-02, and (c) ERA5.

Impact of artefacts in forecast, 2t and MSL – orography

v1



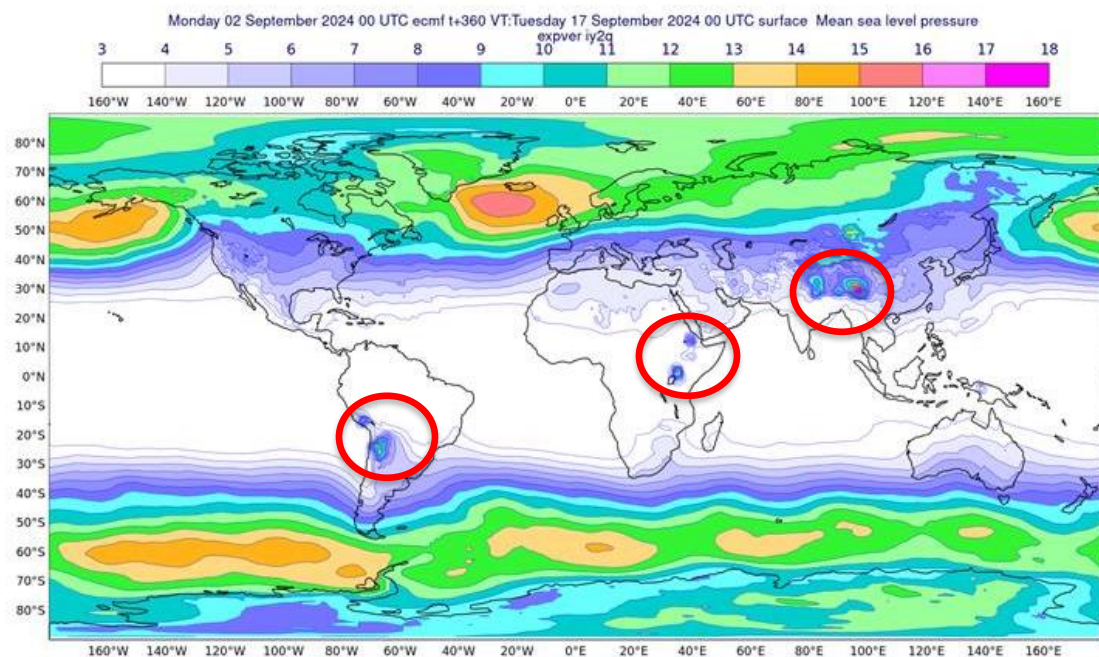
v2



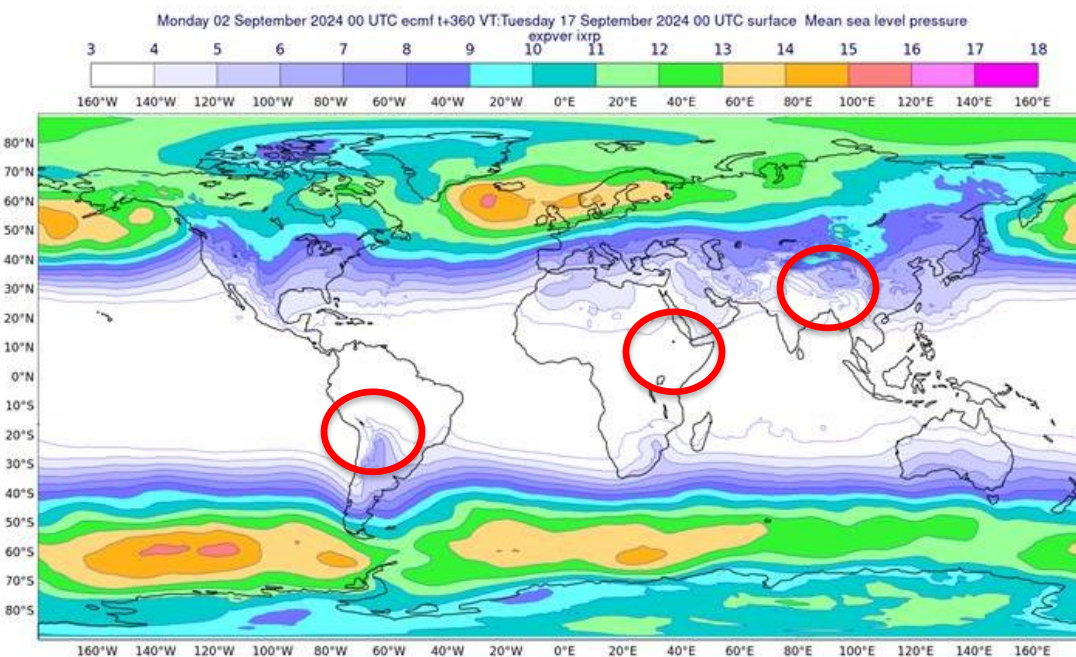
Standard deviation of all perturbed forecasts at day 15 for mean sea level pressure

Impact on artefacts in forecast, 2t and MSL

v1



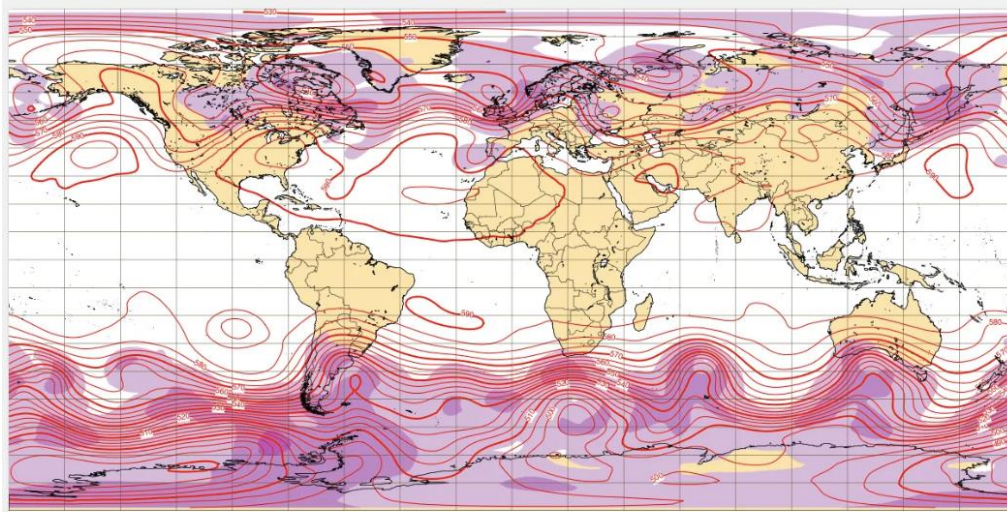
v2



standard deviation of all perturbed forecasts at day 15 for mean sea level pressure

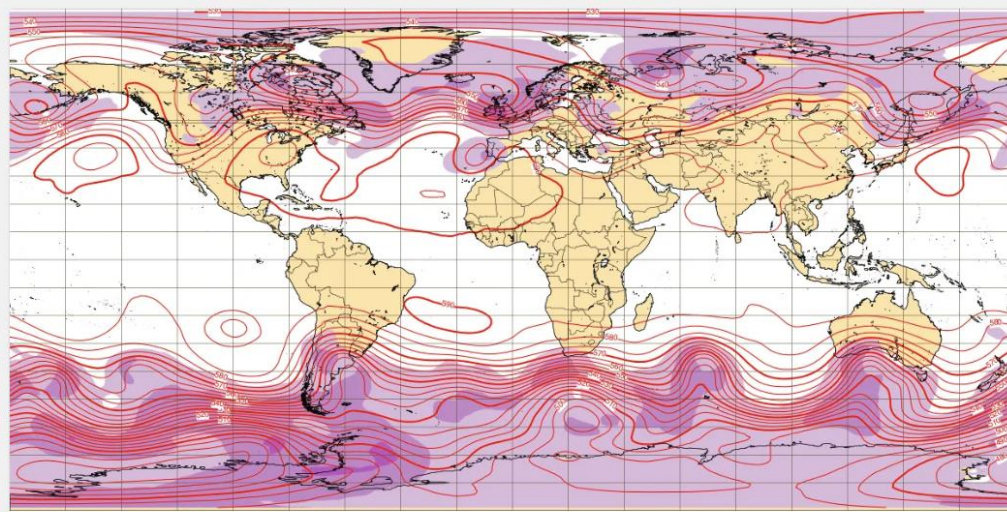
Charts

IFS



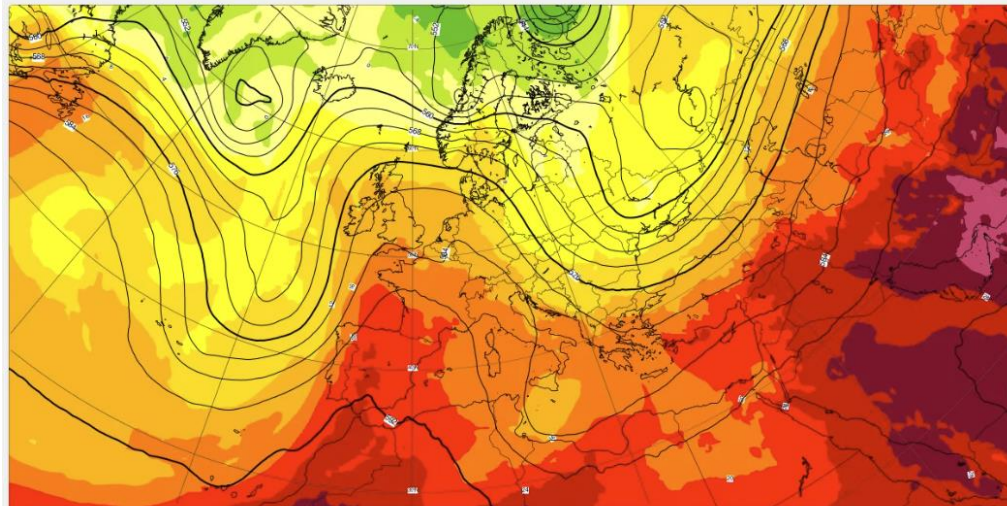
Ensemble mean and spread: 500 hPa geopotential height

AIFS-ENS

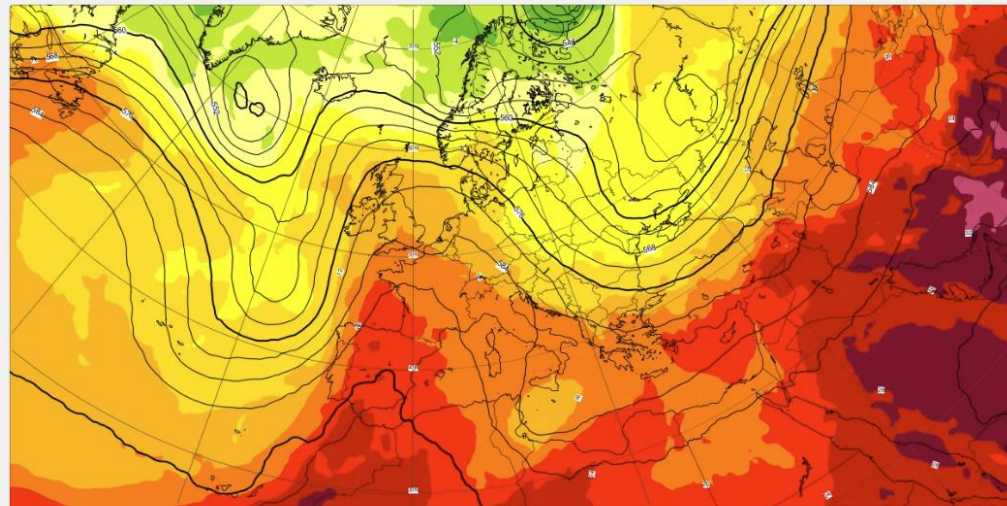


Experimental: AIFS CRPS ENS (ECMWF): Ensemble mean and spread: 500 hPa geopotential height

Z500 hPa,
mean and
spread



500 hPa geopotential height and 850 hPa temperature



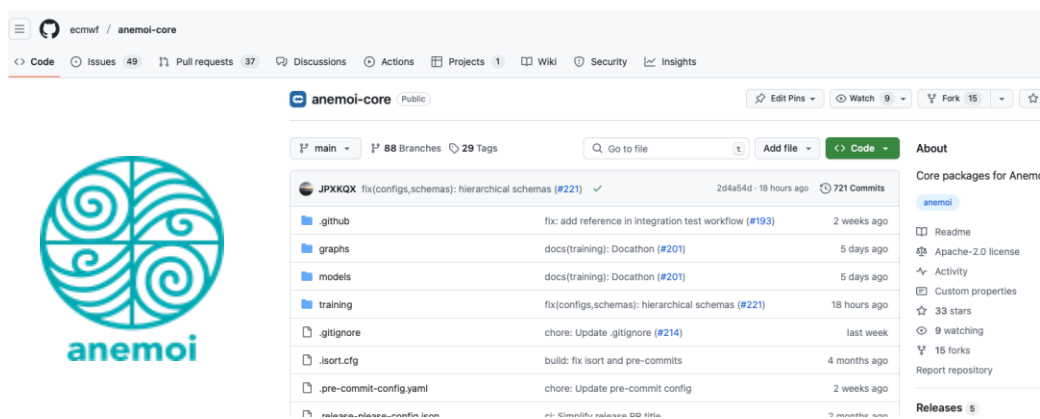
AIFS CRPS Control: 500 hPa geopotential height and 850 hPa temperature

Control
forecasts,
z500 hPa and
t850 hPa

AIFS Single and AIFS-ENS are now operational

- ➔ Forecasts available via charts.ecmwf.int and data via ECMWF's open data policy
- ➔ Code, training recipes etc. now available via Anemol
- ➔ Model weights open

Code (including training etc.)



ecmwf / anemol-core

Code Issues 49 Pull requests 37 Discussions Actions Projects 1 Wiki Security Insights

anemol-core Public

main 88 Branches 29 Tags

Go to file Add file Code

JPXKQX fix(configs,schemas): hierarchical schemas (#221) 2d4a54d · 18 hours ago 721 Commits

.github	fix: add reference in integration test workflow (#193)	2 weeks ago
graphs	docs(training): Docathon (#201)	5 days ago
models	docs(training): Docathon (#201)	5 days ago
training	fix(configs,schemas): hierarchical schemas (#221)	18 hours ago
.gitignore	chore: Update .gitignore (#214)	last week
.isort.cfg	build: fix isort and pre-commits	4 months ago
.pre-commit-config.yaml	chore: Update pre-commit config	2 weeks ago
release_manifest.toml	ci: Simplify release SD kit	9 months ago

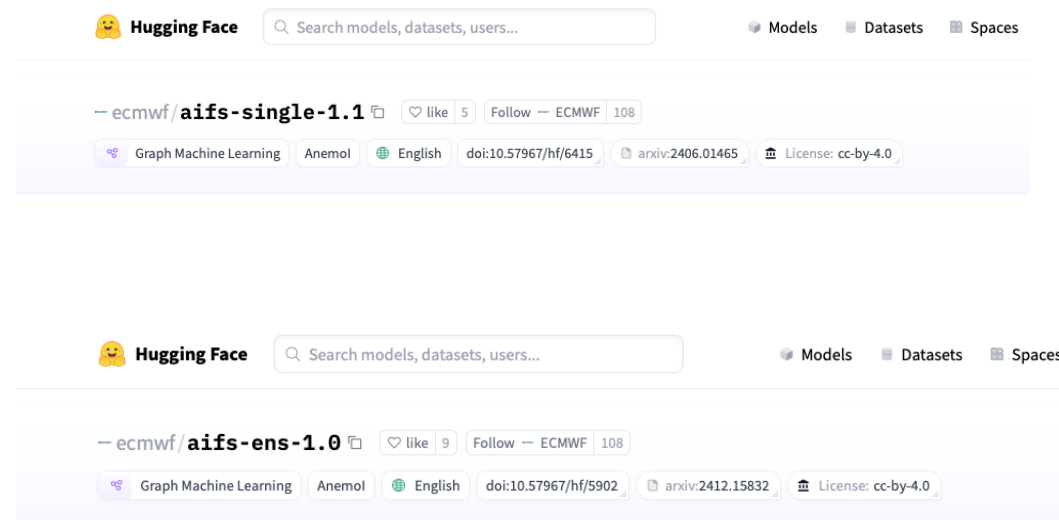
About

Core packages for Anemol

- anemol
- Readme
- Apache-2.0 license
- Activity
- Custom properties
- 33 stars
- 9 watching
- 15 forks
- Report repository

Releases 5

Model weights:



Hugging Face

Search models, datasets, users...

Models Datasets Spaces

ecmwf/aifs-single-1.1 like 5 Follow ECMWF 108

Graph Machine Learning Anemol English doi:10.57967/hf/6415 arxiv:2406.01465 License: cc-by-4.0

Hugging Face

Search models, datasets, users...

Models Datasets Spaces

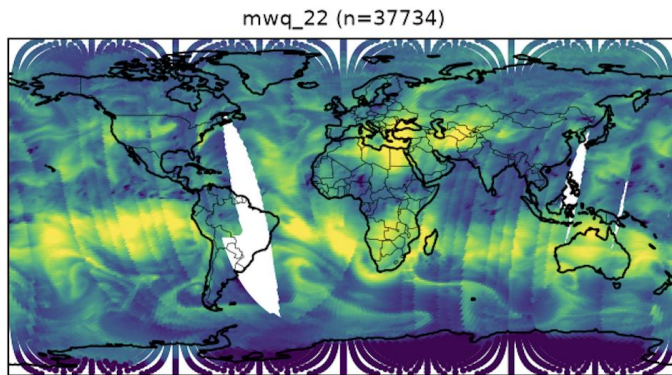
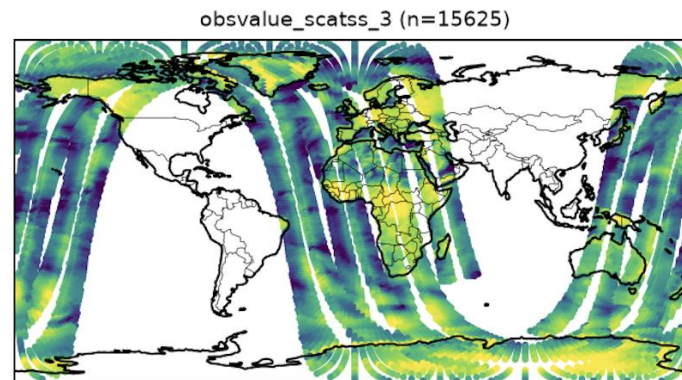
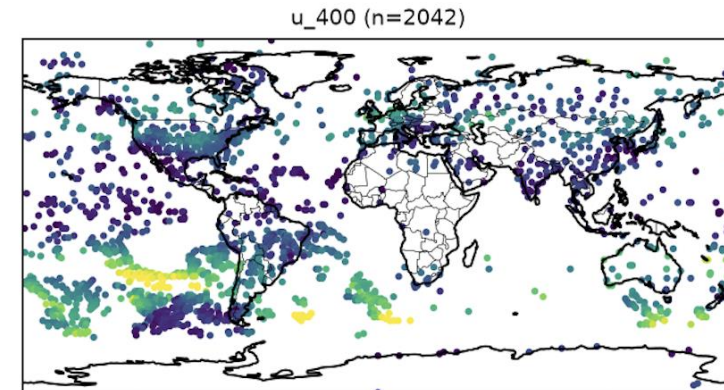
ecmwf/aifs-ens-1.0 like 9 Follow ECMWF 108

Graph Machine Learning Anemol English doi:10.57967/hf/5902 arxiv:2412.15832 License: cc-by-4.0



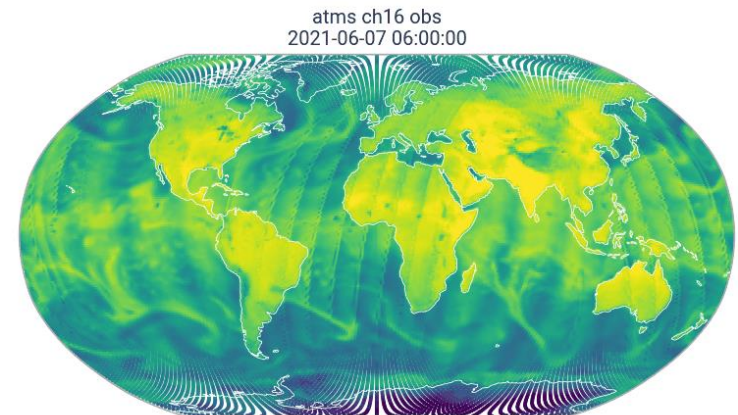
Exploratory work: AIFS-DOP (direct observation prediction)

- Trained on 1981–2024 observations, combining long, harmonised records with recent observations representative of real-time acquisition.
→ HIRS, MSU, SSM/T-2, **AMSU-A**, AMSU-B, **MHS**, **ATMS**, **SYNOPS**, **Radiosondes**, **CrIS**, **AMV**, **ASCAT**, **GPSRO**, **IASI**, GridSAT
- Uses 36 hours of input observations, provided on a 6-hourly grid.
- Here, O96 (~ 100km) resolution



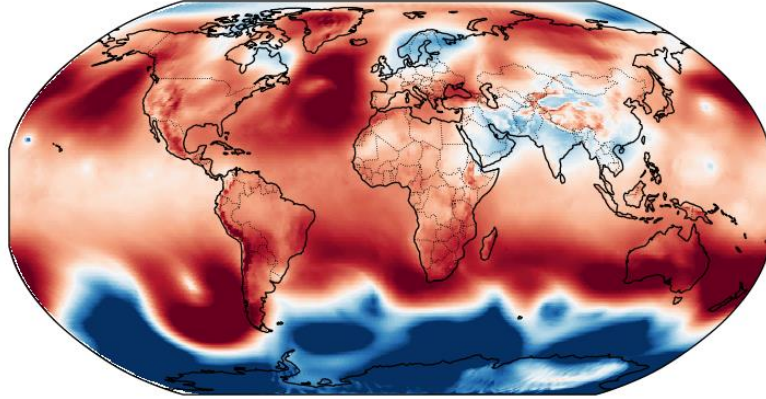
- Runs every 6 hours
- Produces a 10-day forecast in ~ **0.2 seconds**

<https://arxiv.org/abs/2606.19093>

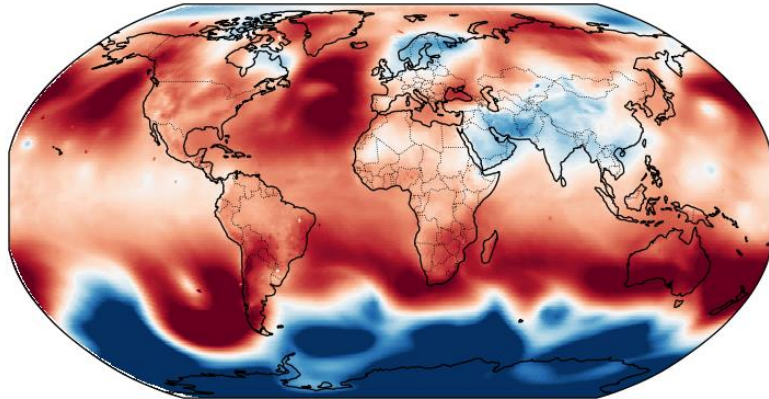


MSLP

IFS oper (0001)

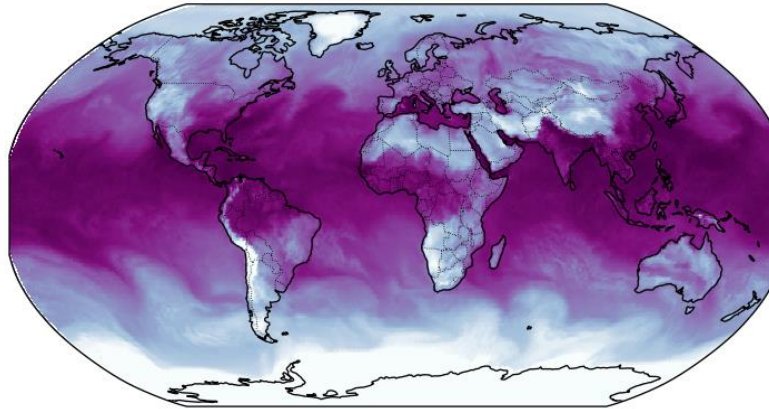


AIFS-DOP O96 6h (j84i)

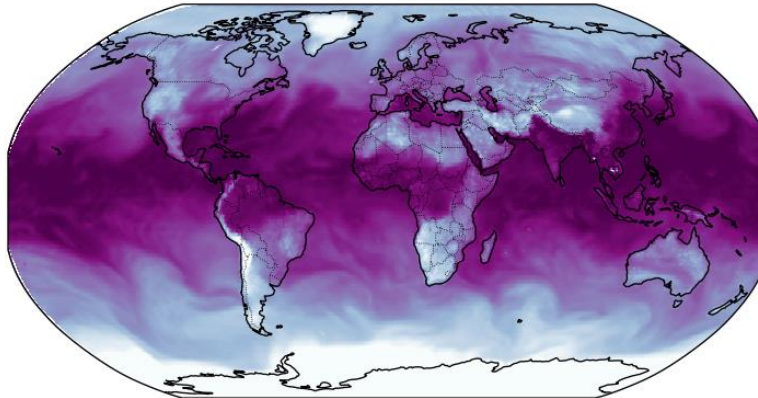


2m Dewpoint temperature

IFS oper (0001)

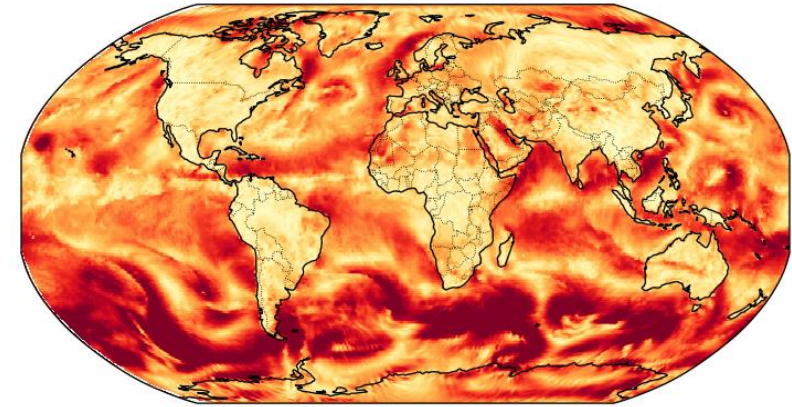


AIFS-DOP O96 6h (j84i)

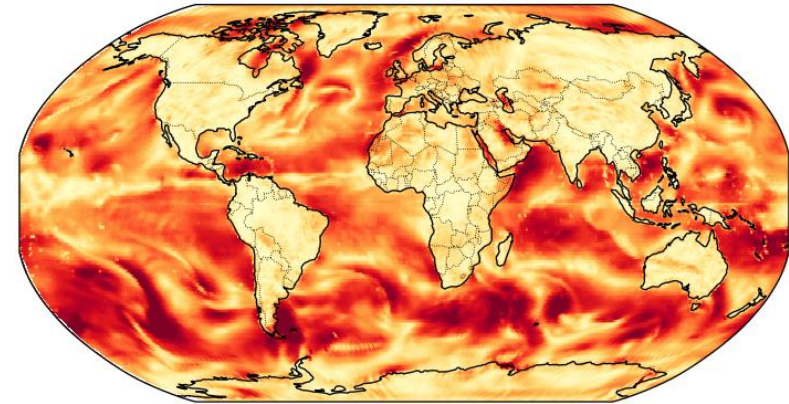
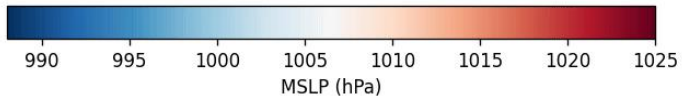
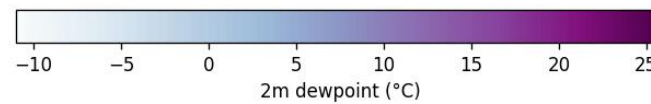
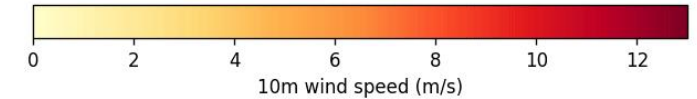


10-metre wind speed

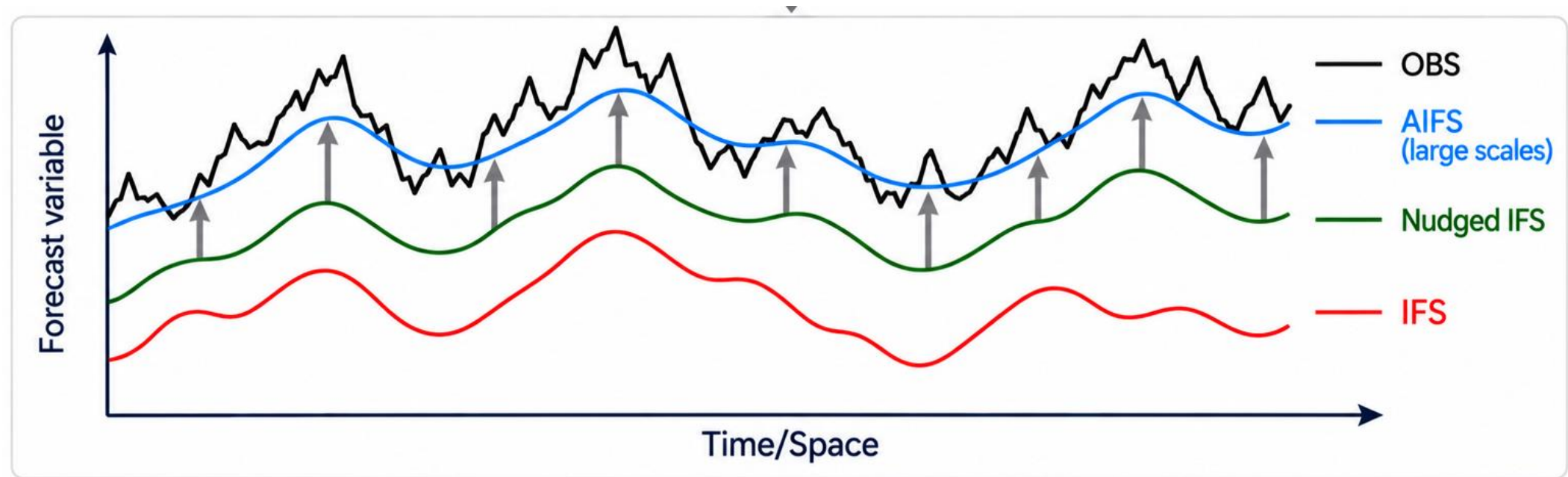
IFS oper (0001)



AIFS-DOP O96 6h (j84i)

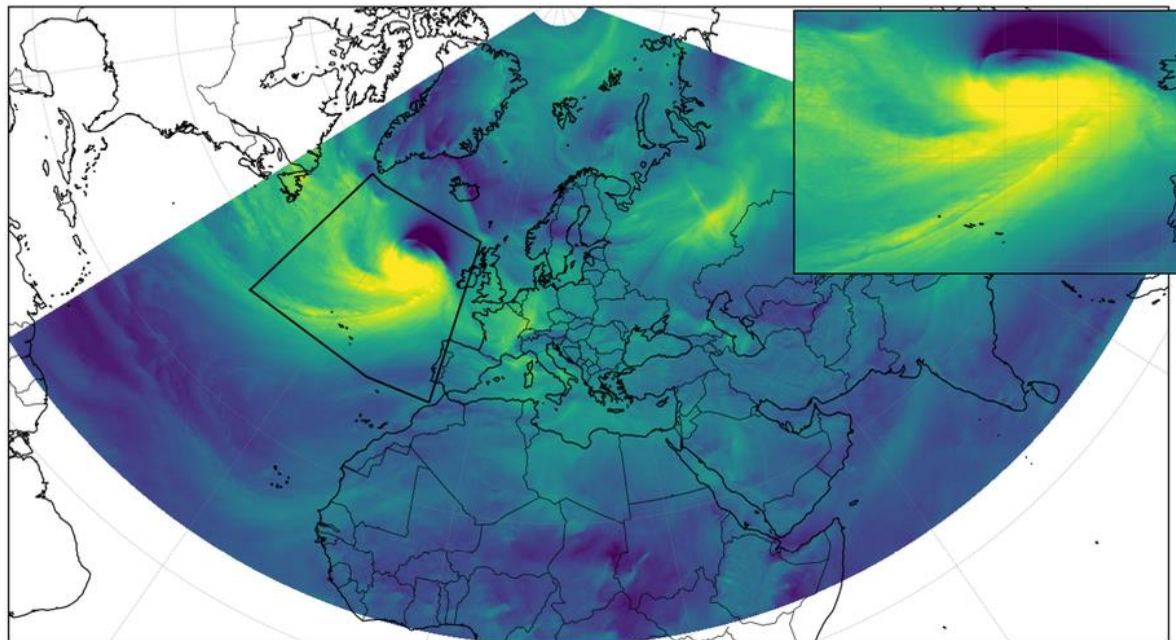


Nudging the IFS to AIFS for both the control and ensemble forecasts

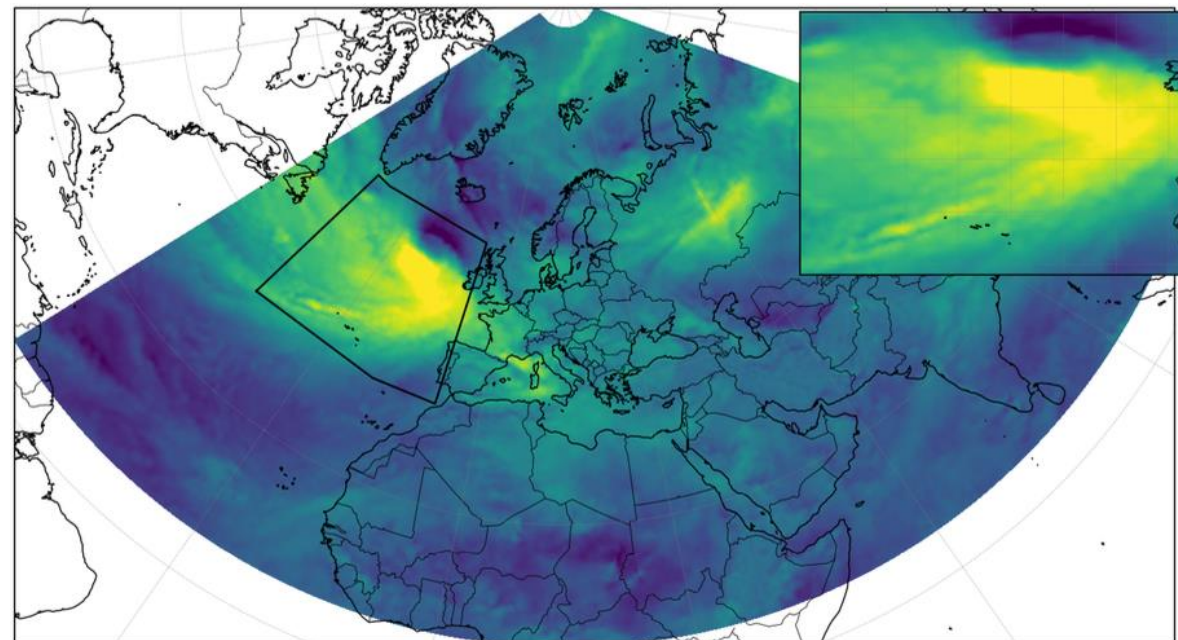


Why: Translate large-scale skill improvement from AIFS to IFS while benefiting from high-resolution and physical consistency.

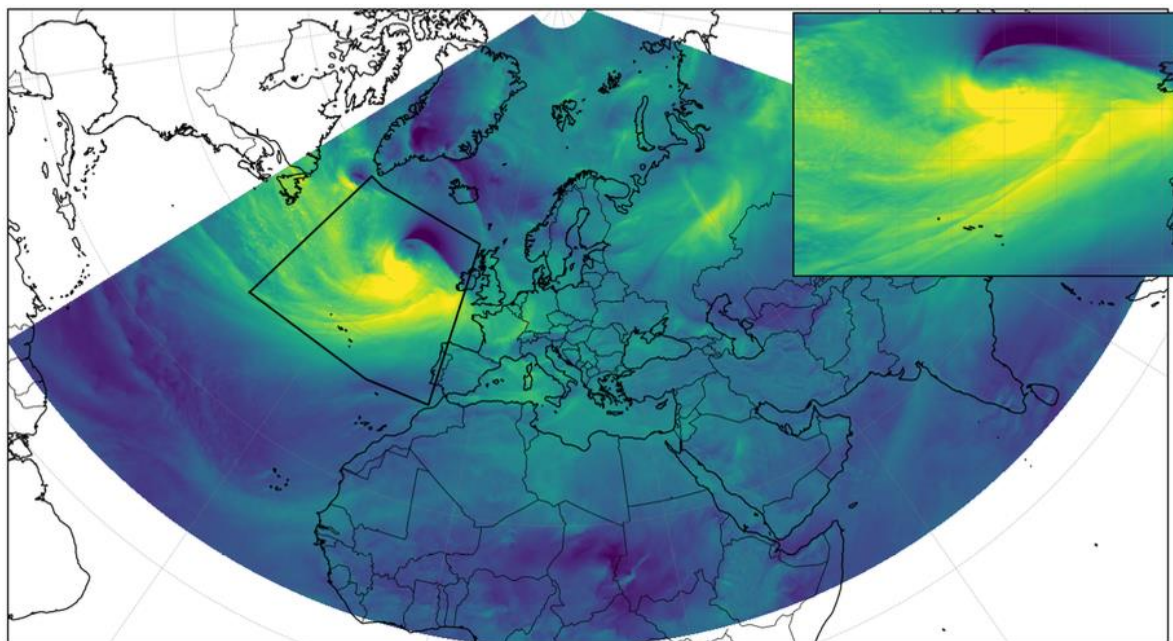
IFS-ENS



AIFS-ENS



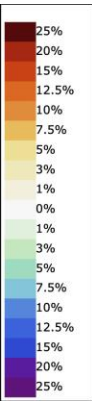
Hy-IFS-ENS (nudged)



Nudging the IFS to AIFS for both the control and ensemble forecasts

Control

Ensemble

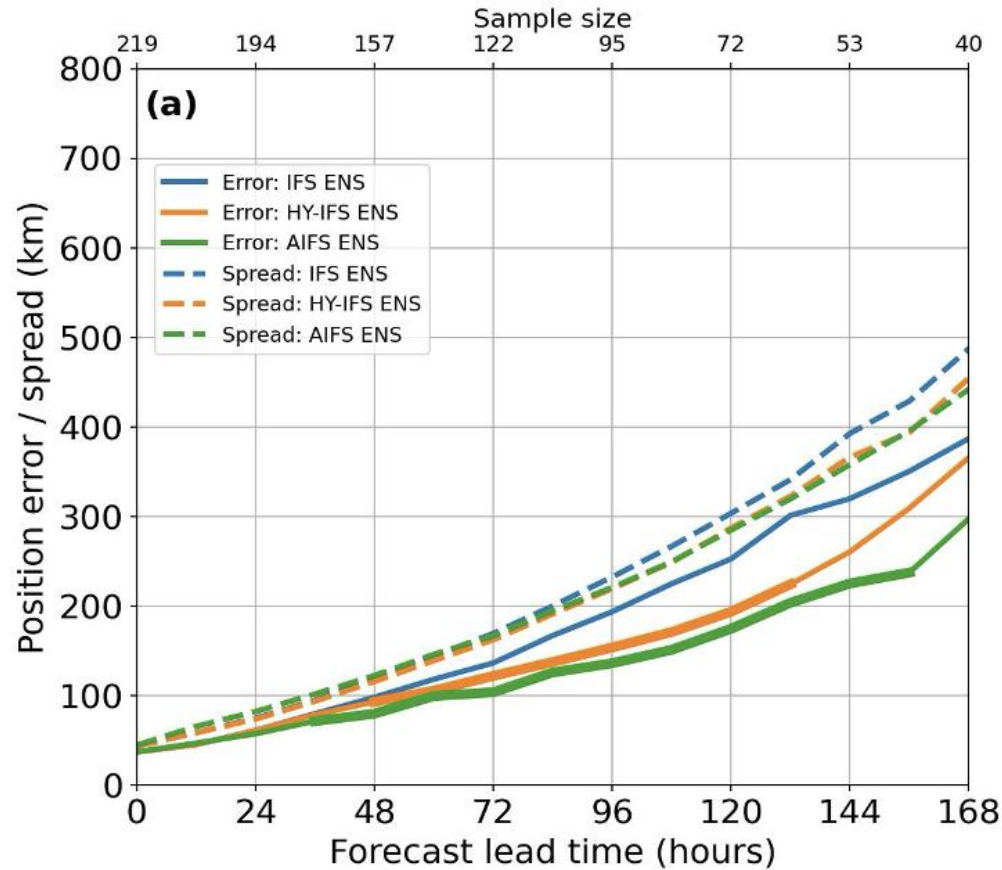
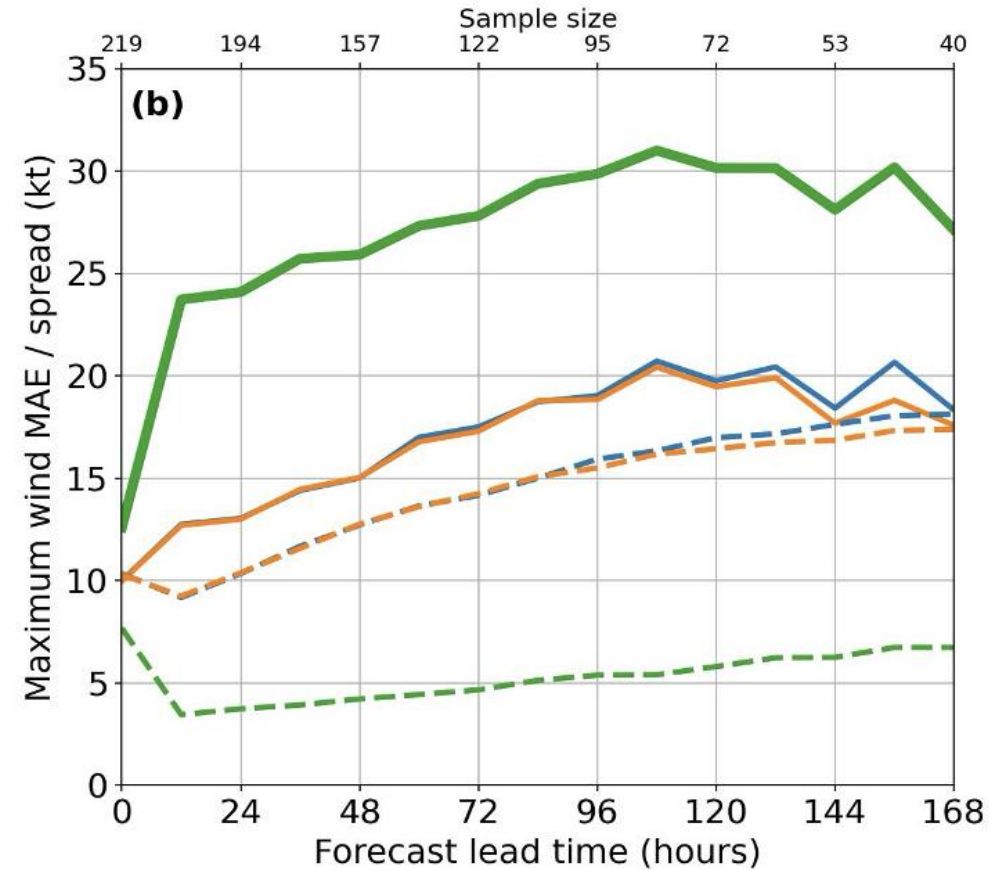


Forecast skill improvement by up to 18h in extra-tropics and up to 1.5days in the tropics.

Extra-tropics skill similar to operational AIFS.

Nudging the IFS to AIFS: Tropical cyclones

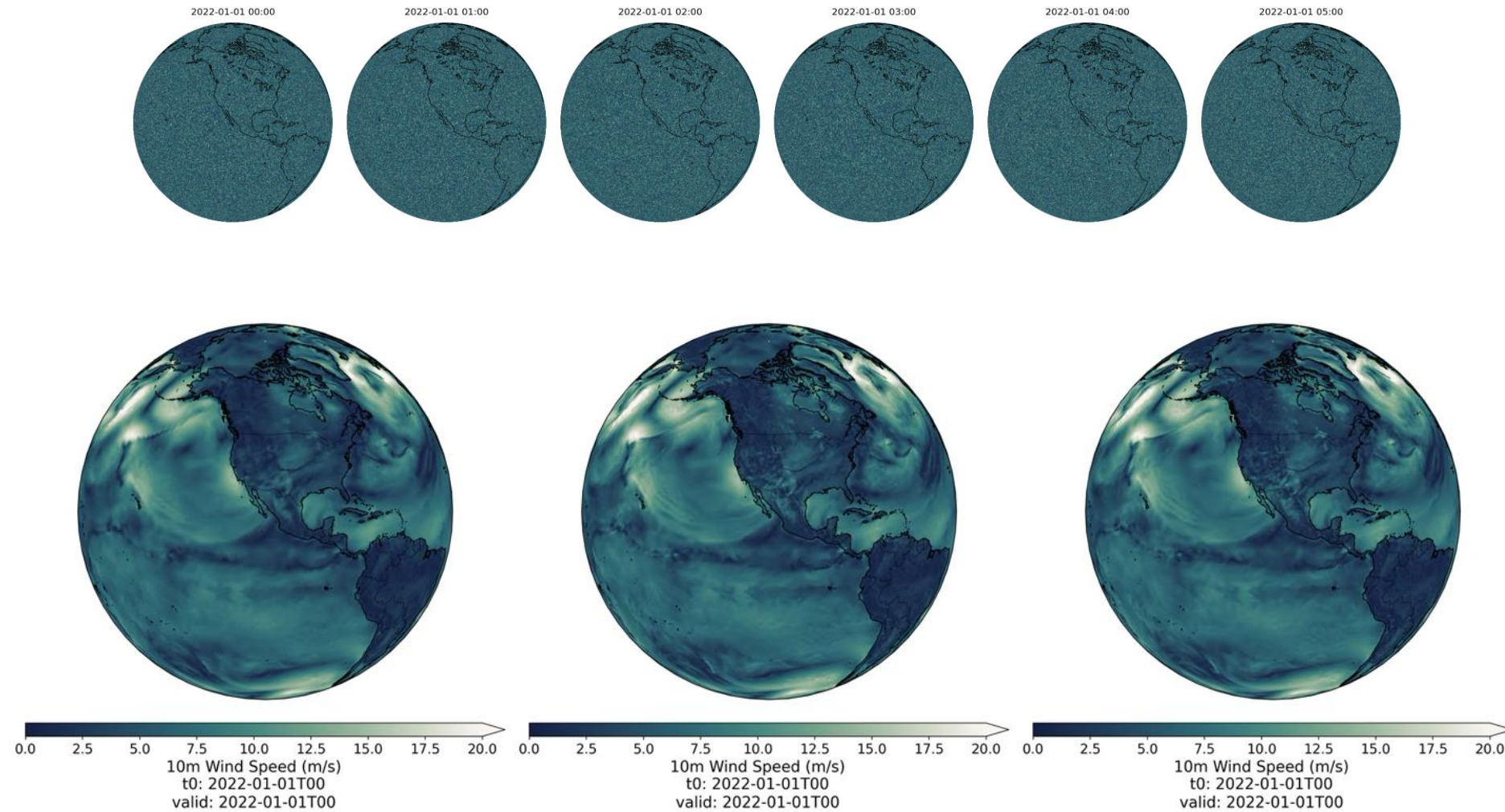
Tropical cyclone track improved, while intensity stays the same as in operational IFS.



What else can we do with ML?

Example: Spatio-Temporal Downscaling

<https://www.ecmwf.int/en/about/media-centre/aifs-blog/2026/flow-matching-spatio-temporal-downscaling>



What is next?

- Higher spatial resolution
- Higher temporal resolution
- Further improve scalability (GPU kernels etc.)
- Longer time ranges
- Integrate / develop more Earth system components
atmospheric composition, ocean, etc.
- Further improve physical consistency
- More and newer datasets
- Exploit observations
- Products
- ...

