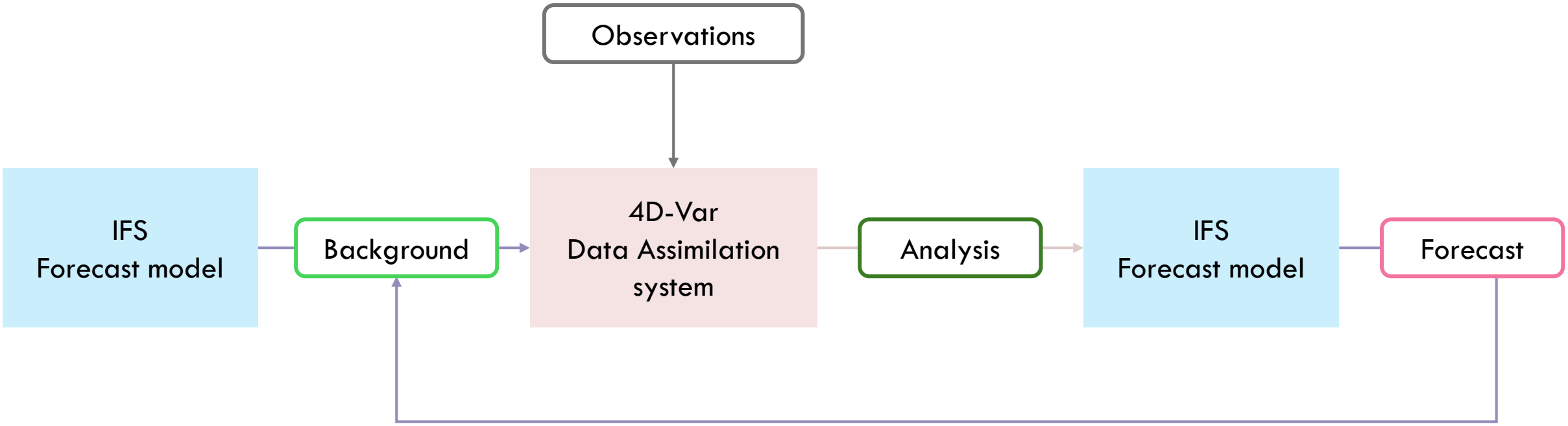
End-to-end real time weather forecasting from observations

Mihai Alexe and Eulalie Boucher

DOP team and many ECMWF colleagues

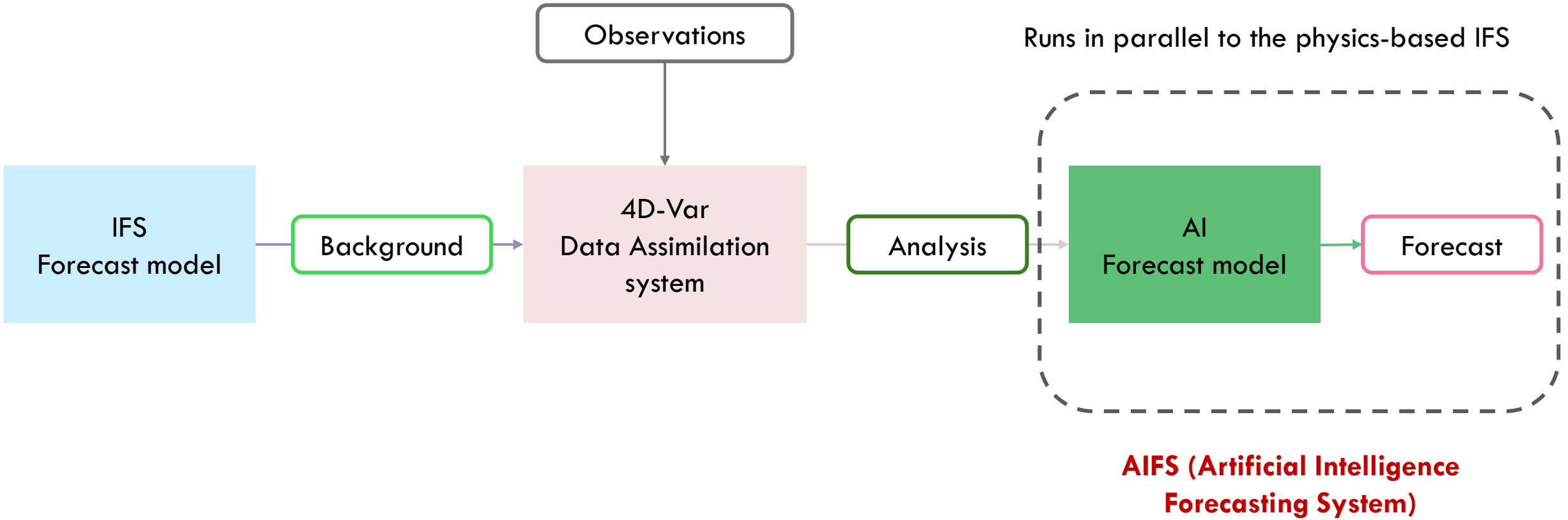


A (very) simplified view of the DA / forecast chain



We could **emulate** one or more parts of this processing chain using AI

Emulating the forecast model with AI



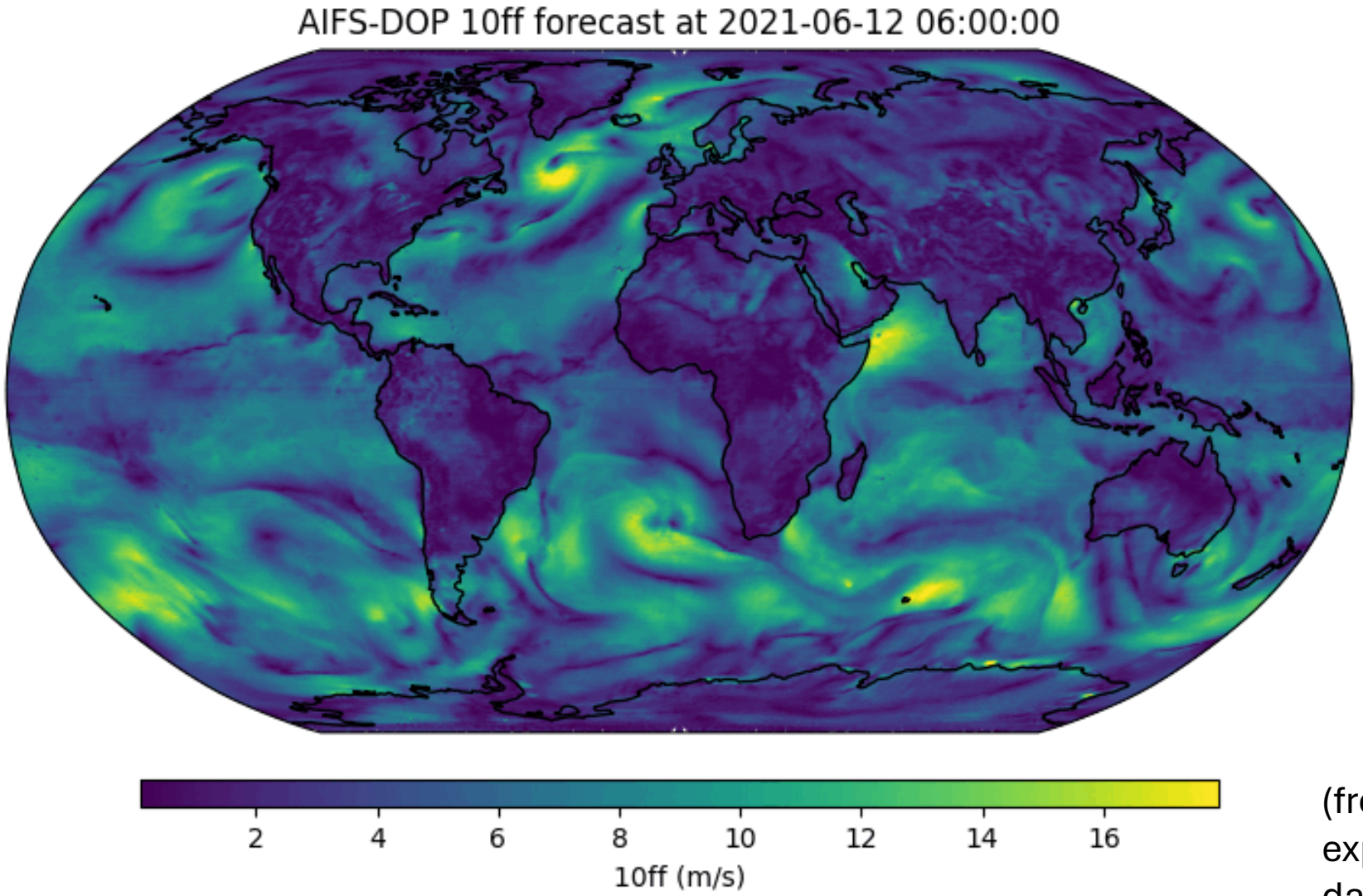
https://www.ecmwf.int/en/about/media-centre/aifs-blog

AI-DOP: Emulating the entire chain using only observations



Artificial Intelligence Direct Observation Prediction (AI-DOP)
goes **straight from Earth System observations to a medium-range forecast**

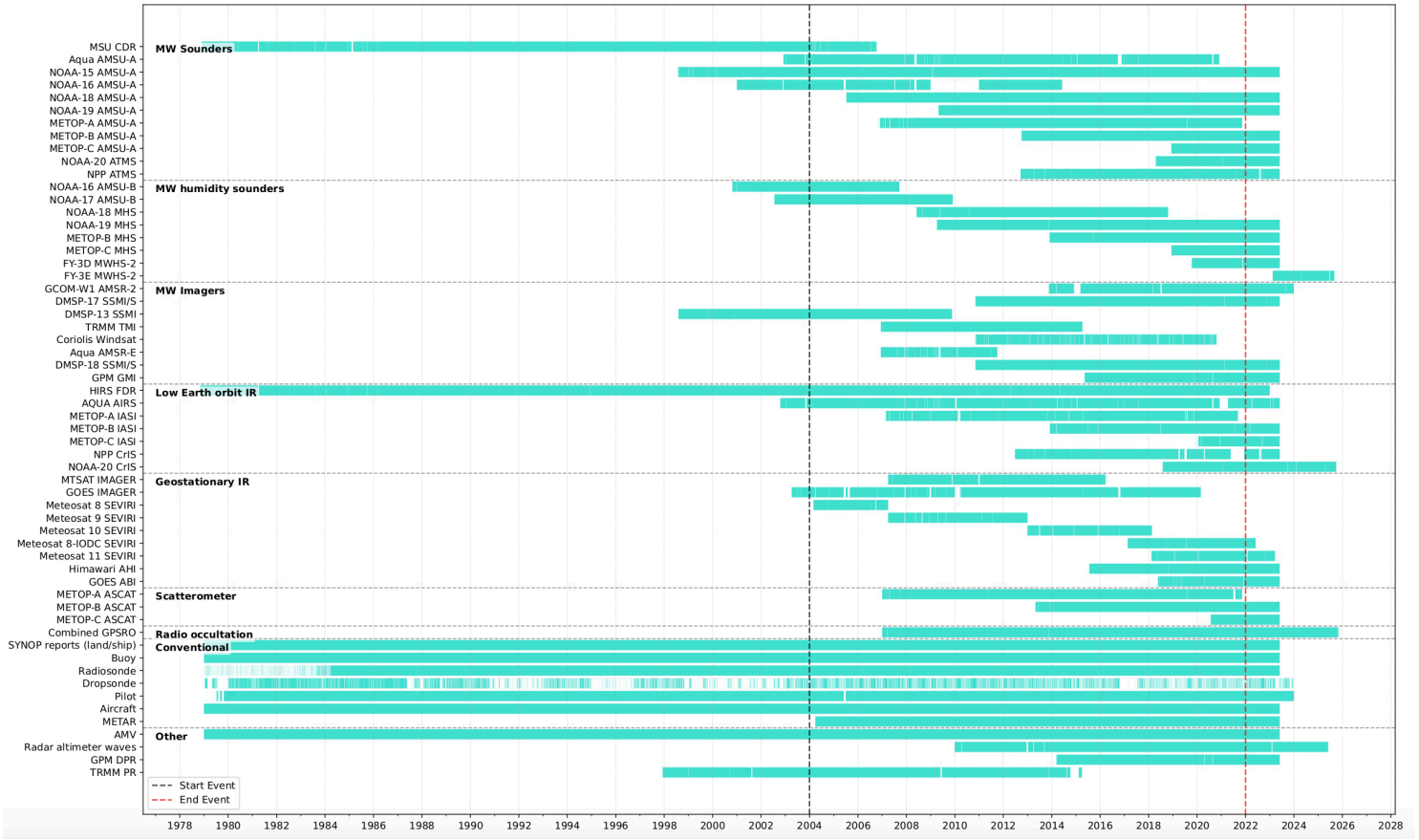
Can DOP produce a useful medium-range forecast?

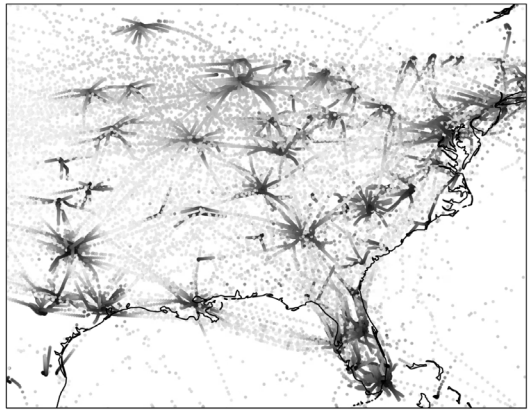


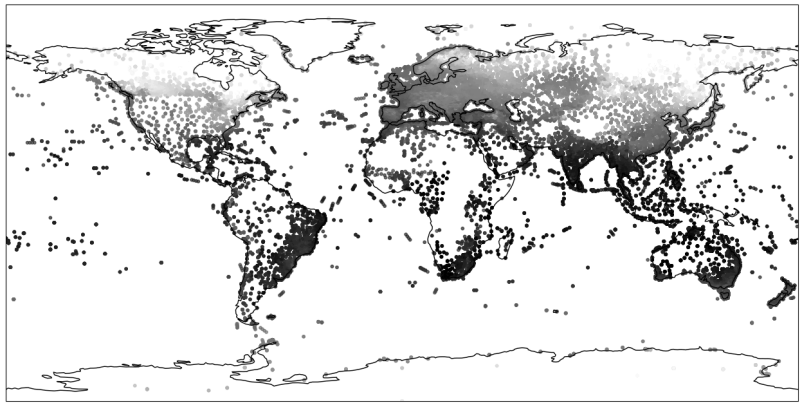
(from a preliminary n320 experiment, no scatt data)

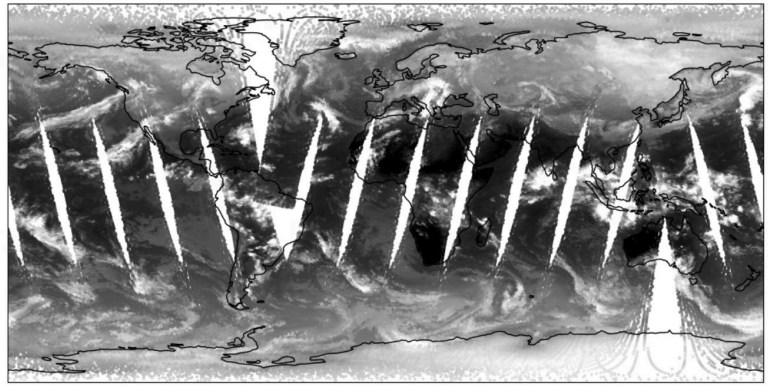
Observation data

- Curated zarr datasets for all major observation categories
- Includes observations/instruments not currently assimilated in IFS
- QC remains crucial (noisy observations can have a negative impact on model training)







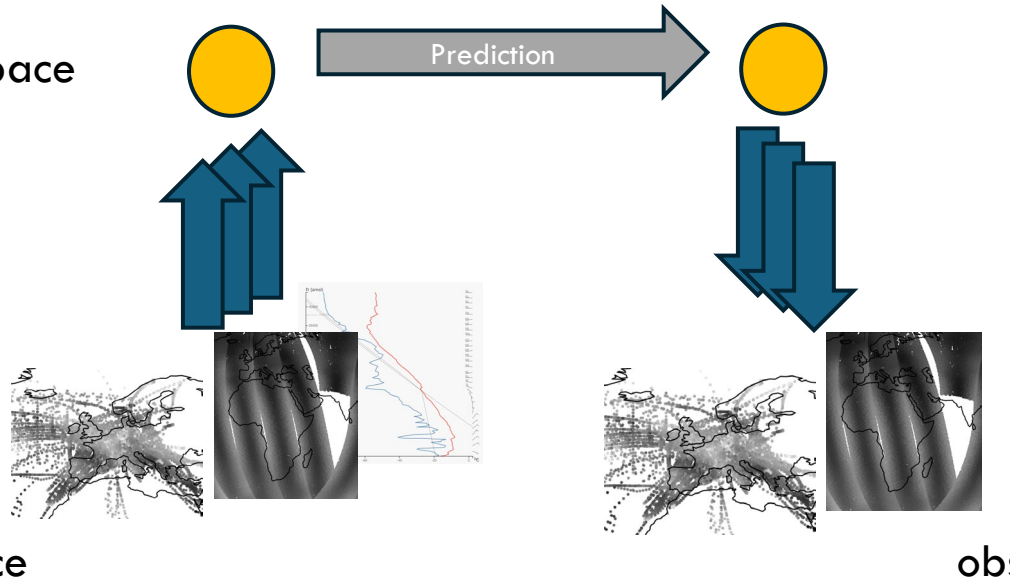


AI-DOP: Given observations in the past, predict observations in the future

Input: satellite and in-situ (conventional) observations in one or more input windows

Target: forecast all observations in the next window(s), with no NWP (re)analysis data used

A learned forecast model (**PROCESSOR**) that evolves the latent state forward in time



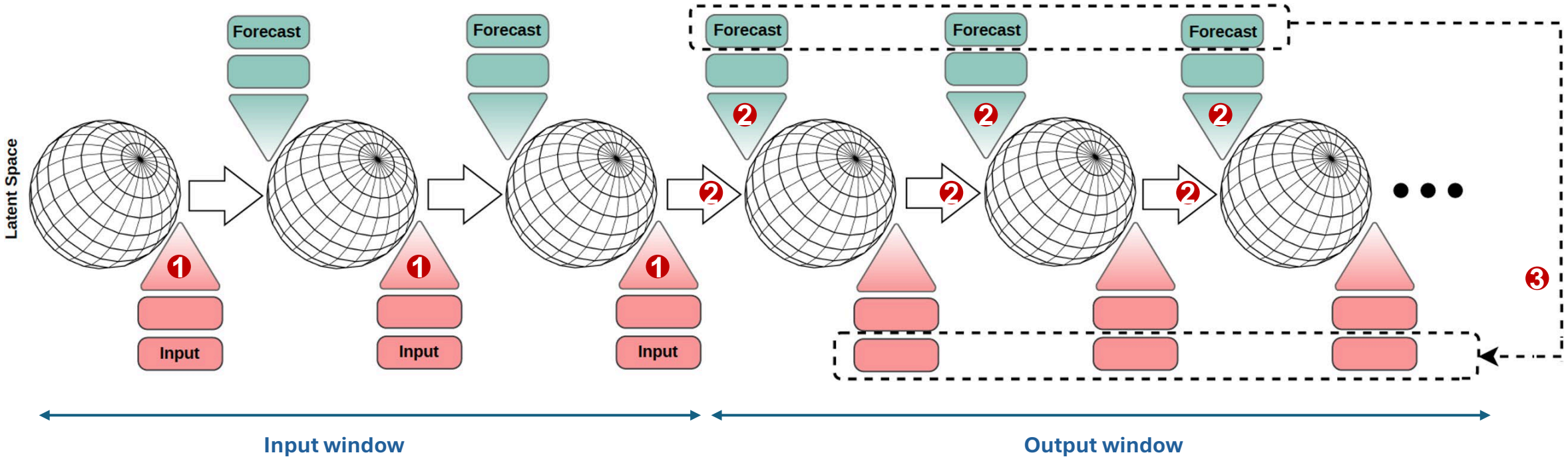
ENCODERs

Learned "DA" functions that combine information from different sources into a (learned) set of unified latent state variables

DECODERs

Learned observation operators which map from the network's internal model latent state to observation space

How DOP operates



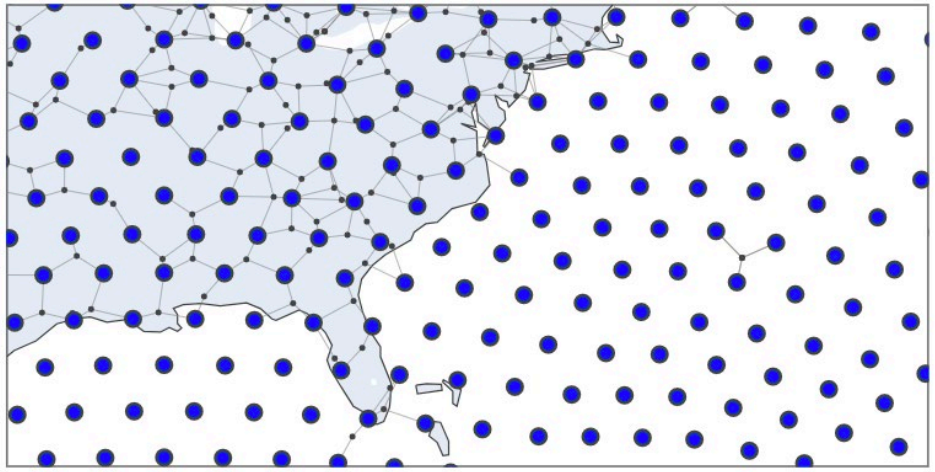
(Top figure from Patrick Laloyaux's talk on Wed)

Encoding and decoding the observations with graph neural networks

Unstructured observations and dynamic graphs

anemoi

Under development



GRAPHDOP: TOWARDS SKILFUL DATA-DRIVEN MEDIUM-RANGE WEATHER FORECASTS LEARNT AND INITIALISED DIRECTLY FROM OBSERVATIONS

A PREPRINT

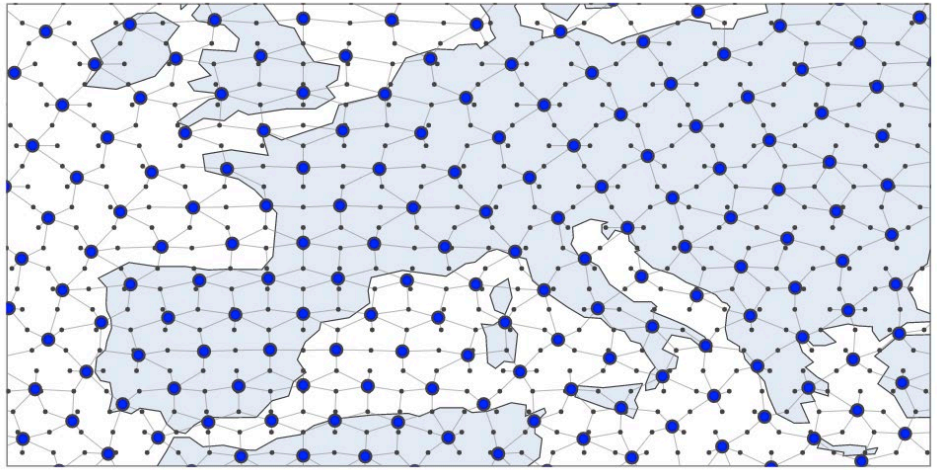
Mihai Alexe · Eulalie Boucher · Peter Lean · Ewan Pinnington · Patrick Laloyaux · Anthony McNally · Simon Lang · Matthew Chantry · Chris Burrows · Marcin Chrust · Florian Pinault · Ethel Villeneuve · Niels Bormann · Sean Healy

European Centre for Medium-Range Weather Forecasts (ECMWF)

https://arxiv.org/pdf/2412.15687

Gridded observations and static graphs

anemoi



AIFS-DOP: END-TO-END MEDIUM-RANGE WEATHER PREDICTION FROM OBSERVATIONS ALONE WITH MACHINE LEARNING

Ewan Pinnington · Peter Lean · Mihai Alexe · Eulalie Boucher · Simon Lang · Patrick Laloyaux · Gert Mertes · Tomas Kral · Patricia de Rosnay · Matthew Chantry · Anthony McNally

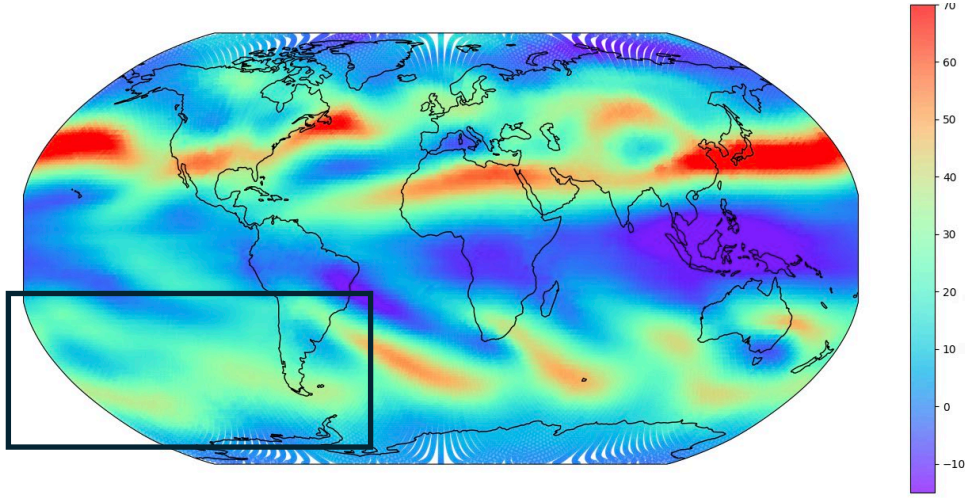
European Centre for Medium-Range Weather Forecasts (ECMWF)

https://arxiv.org/pdf/2606.19093

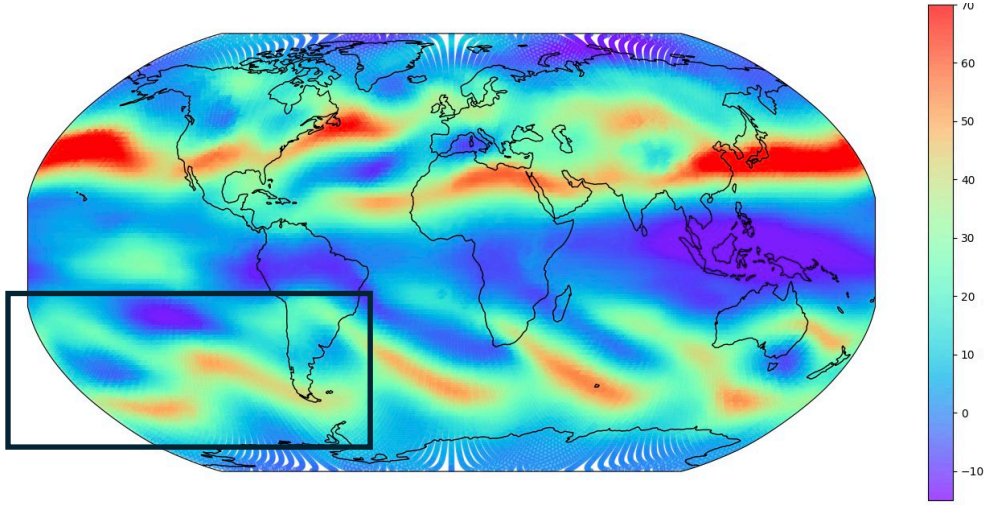
The “optimal” projection for each obs type is still an open research question

Impact of “DA-style” input cycling in DOP (low-res experiment)

Sonde only input – 12h input window

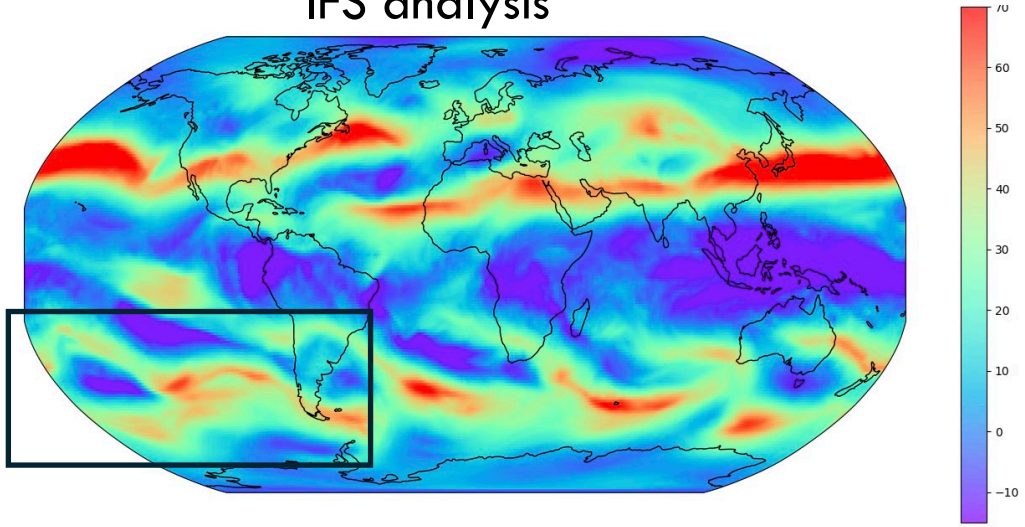


Sonde only input – 72h input window



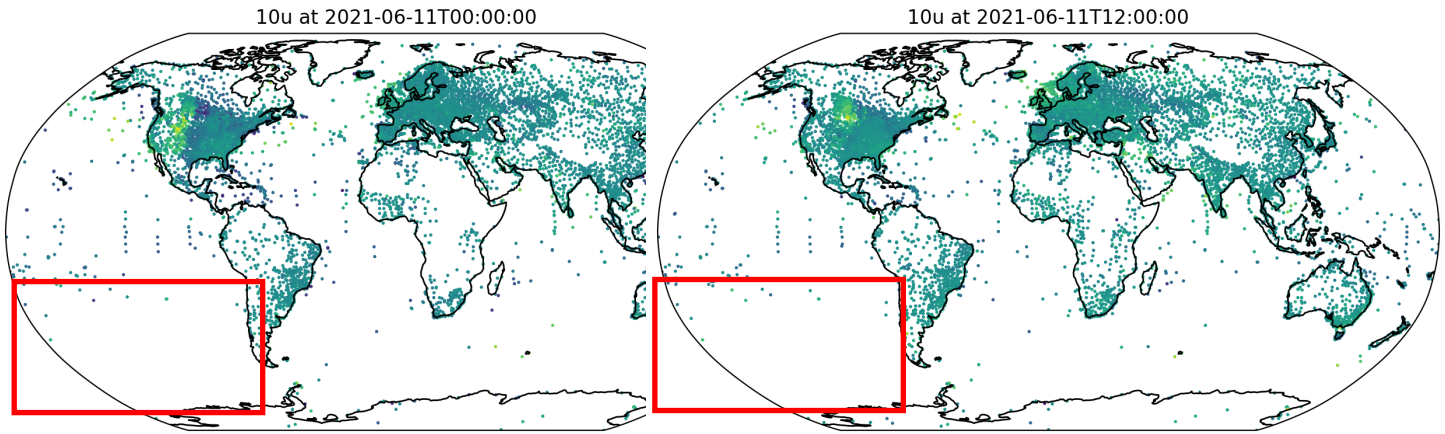
200 hPa zonal wind

IFS analysis

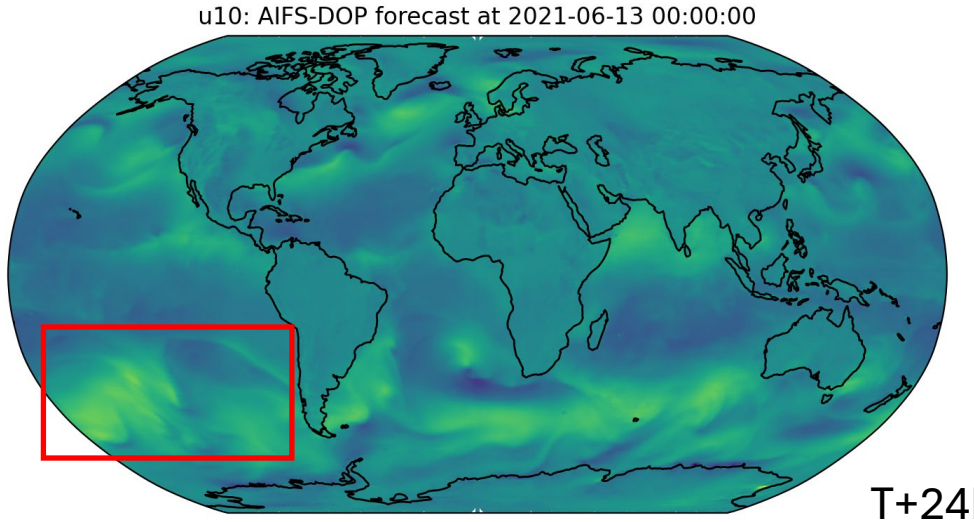


Targets: all conventional observations including AMVs, aircraft, sondes, surface

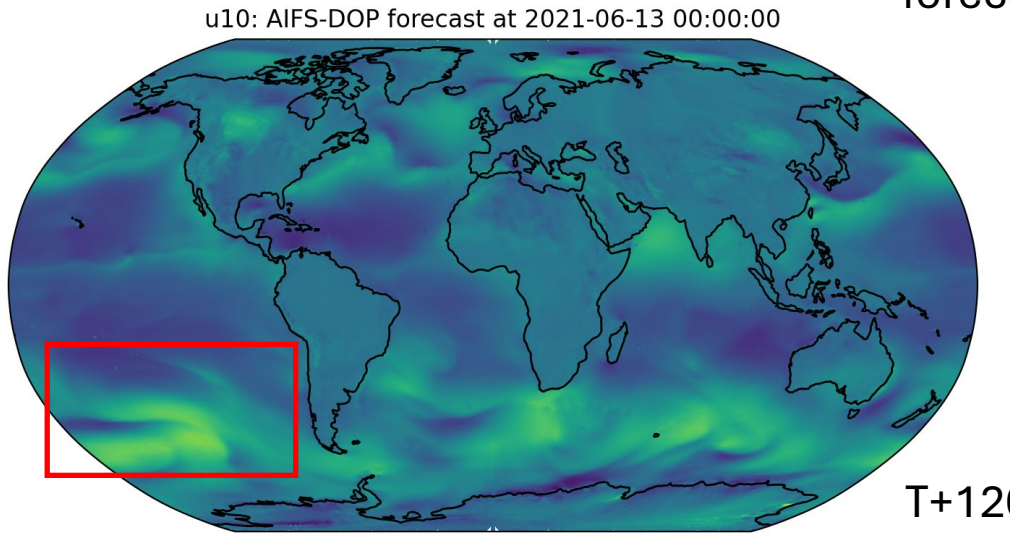
DOP forecasts are produced from very **sparse** conventional observations



Input u10 observations
(this model doesn't use any scatterometer data)



T+24h forecast

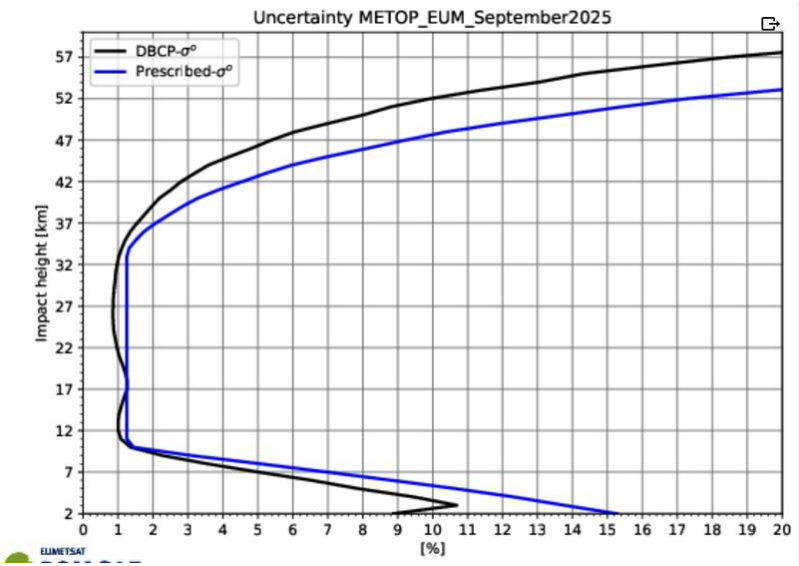


T+120h forecast

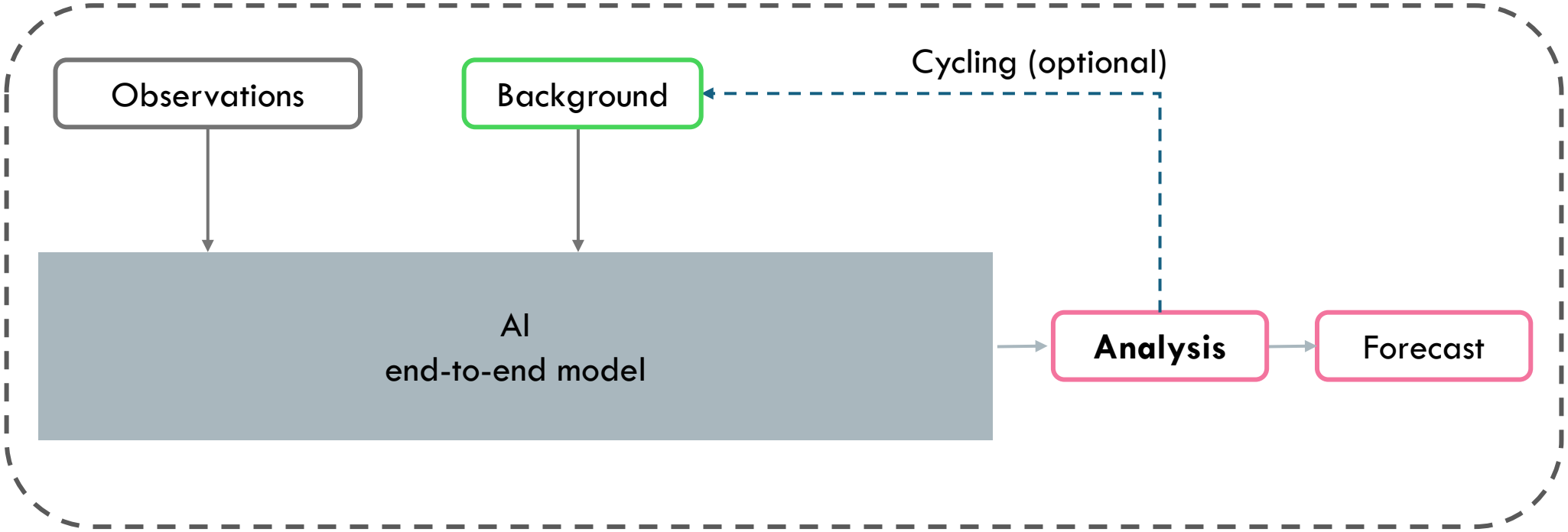
Is there a role for **R** in DOP?

t+12h forecast results from a small experiment that incorporates the GNSS-RO uncertainties into the training procedure (MSE loss term scaled by the prescribed $1/\sigma^2$)

Input: GNSS-RO
Output: radiosondes, AMVs, PMSL, GNSS-RO



A “hybrid” approach: introduce a background

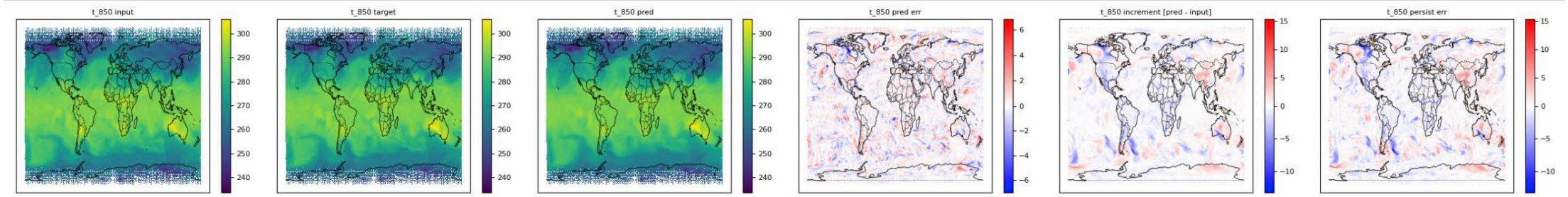


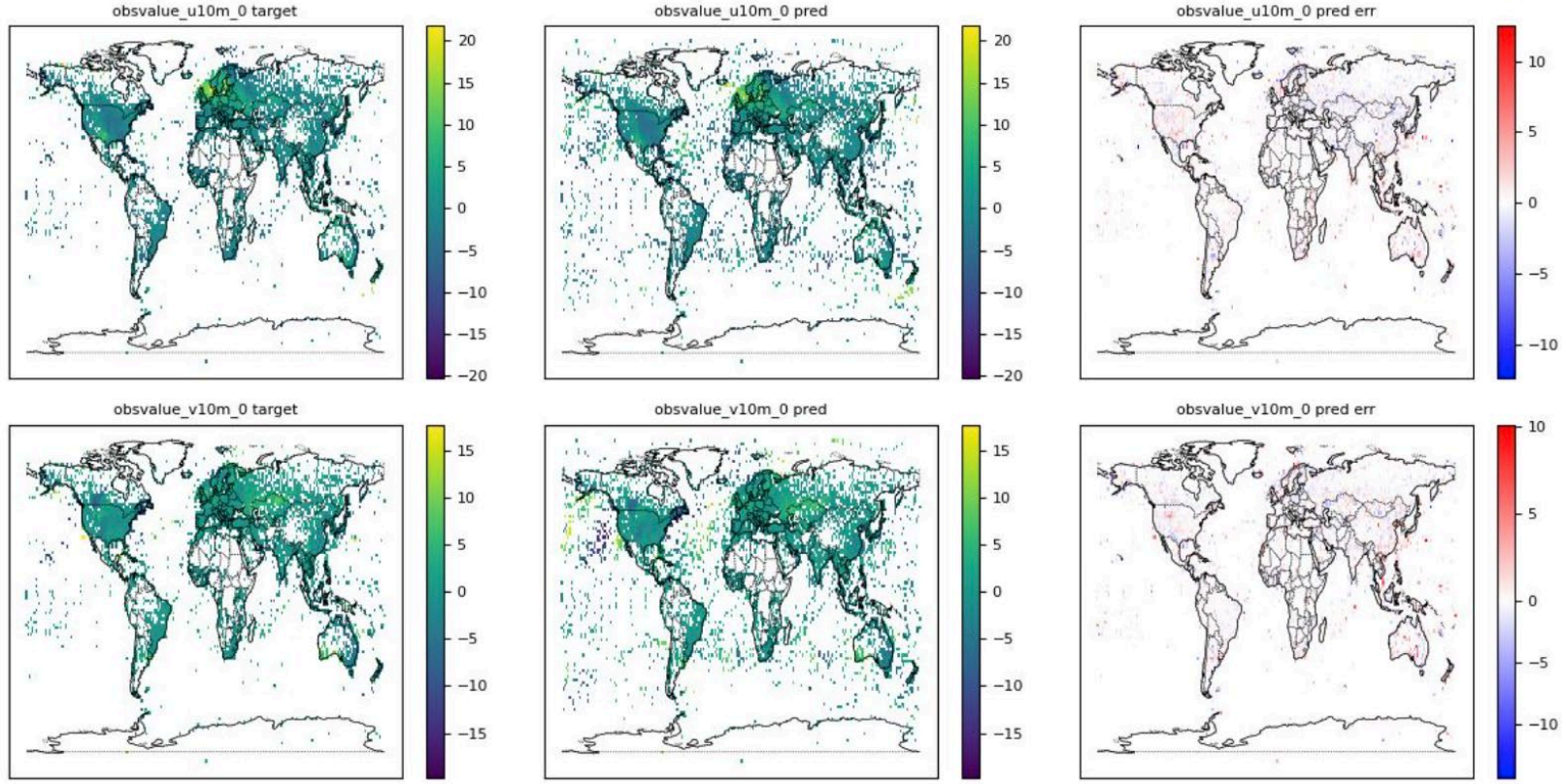
A “hybrid” system, trained end-to-end

driven by recent Earth System observations and a background atmospheric state

The background can originate from a physics-based or ML model trajectory

Joint learning from (re)analysis and observations with anemoi





Support for unstructured observational data is currently being added to the Anemoi framework – major refactoring across datasets, models, training and inference

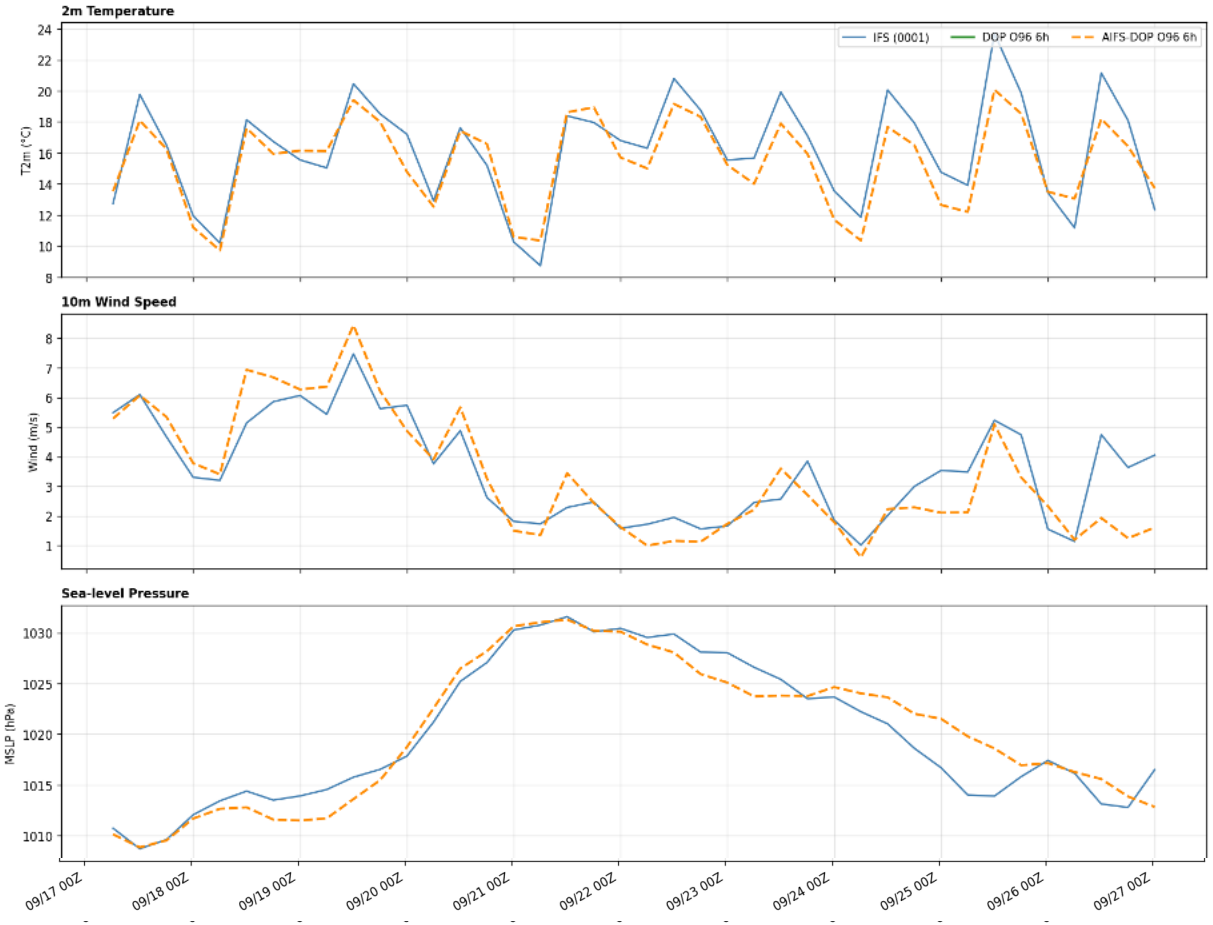
You can now leverage the full Anemoi toolkit to build DOP-like models (including SOTA probabilistic approaches)

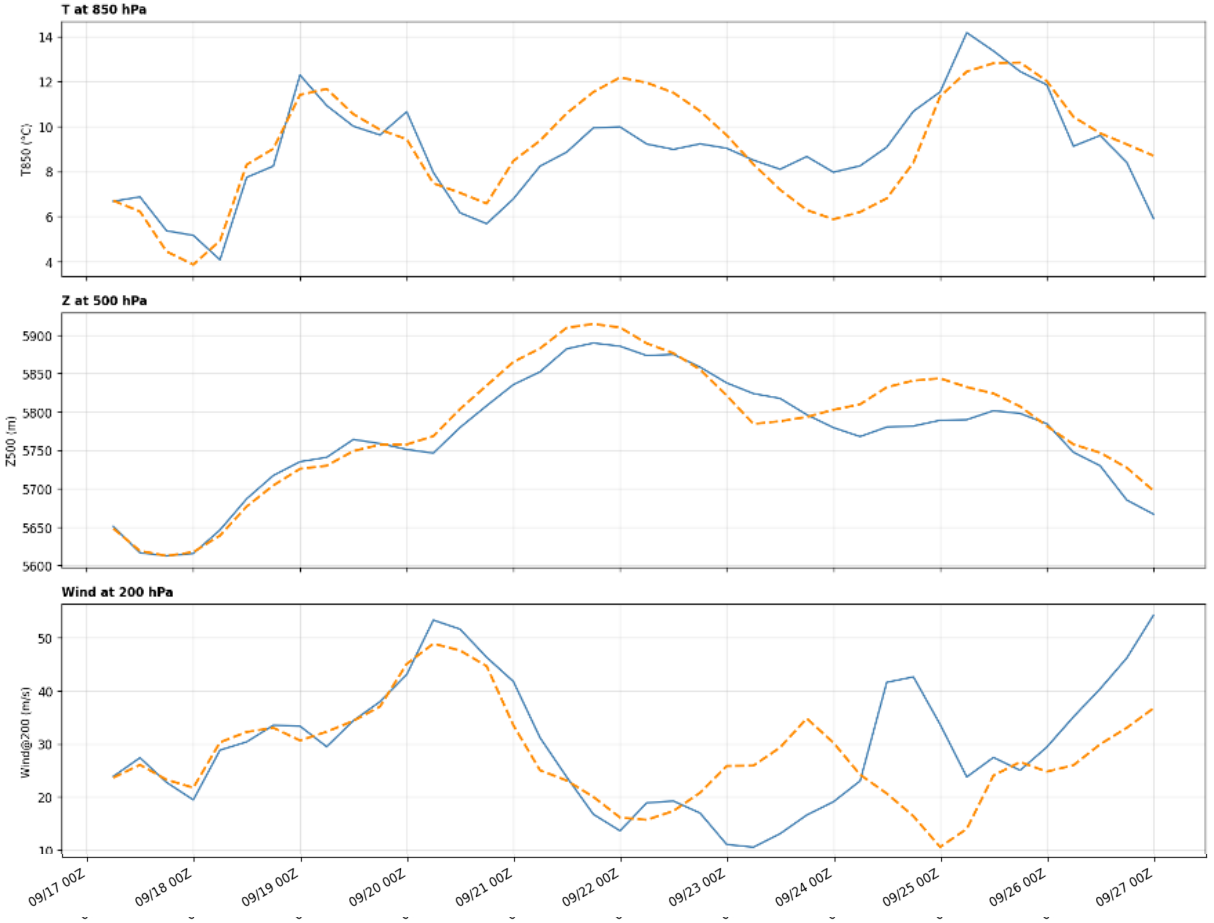
Producing real-time forecasts

Weather for next week in Reading

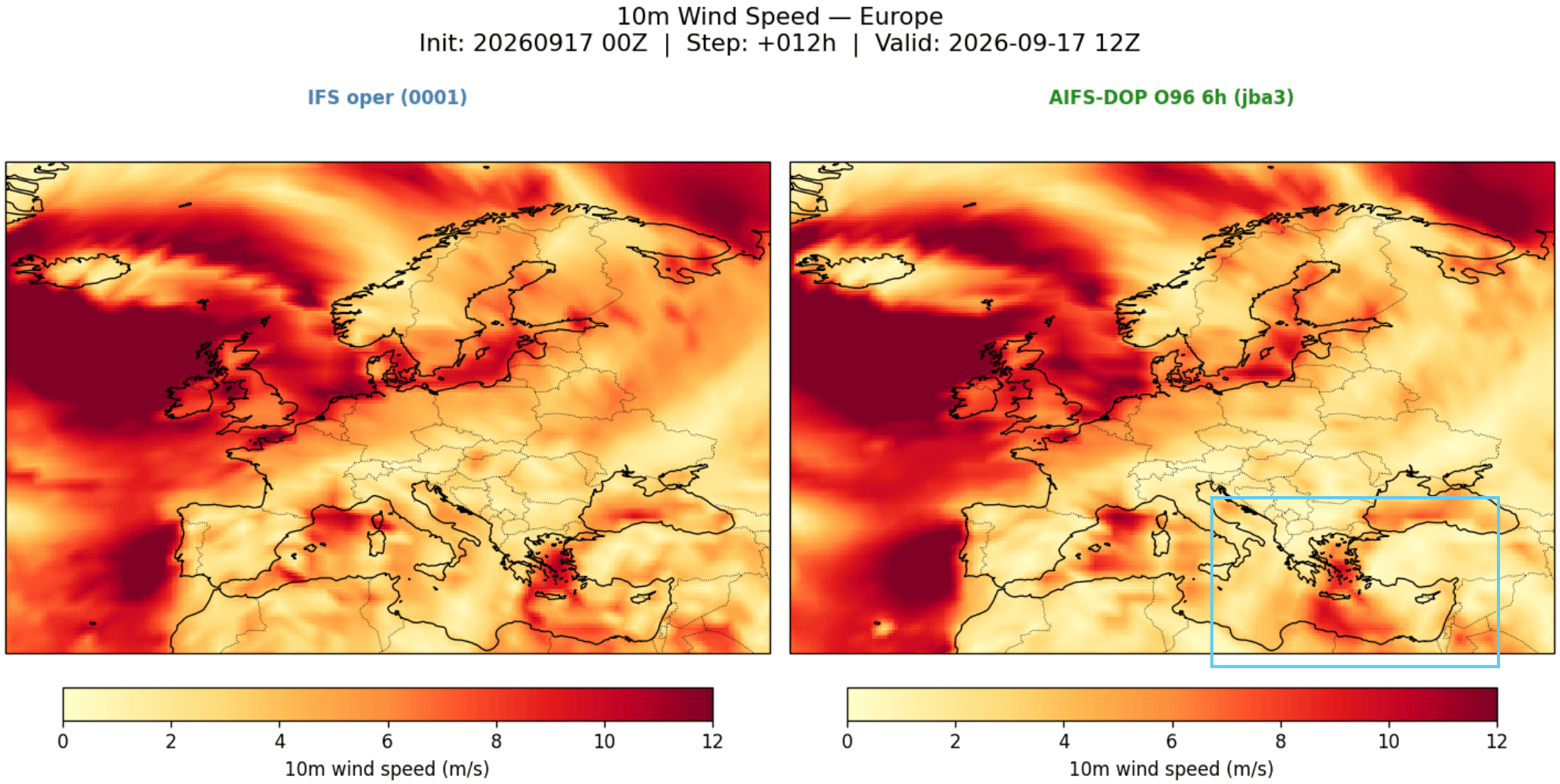


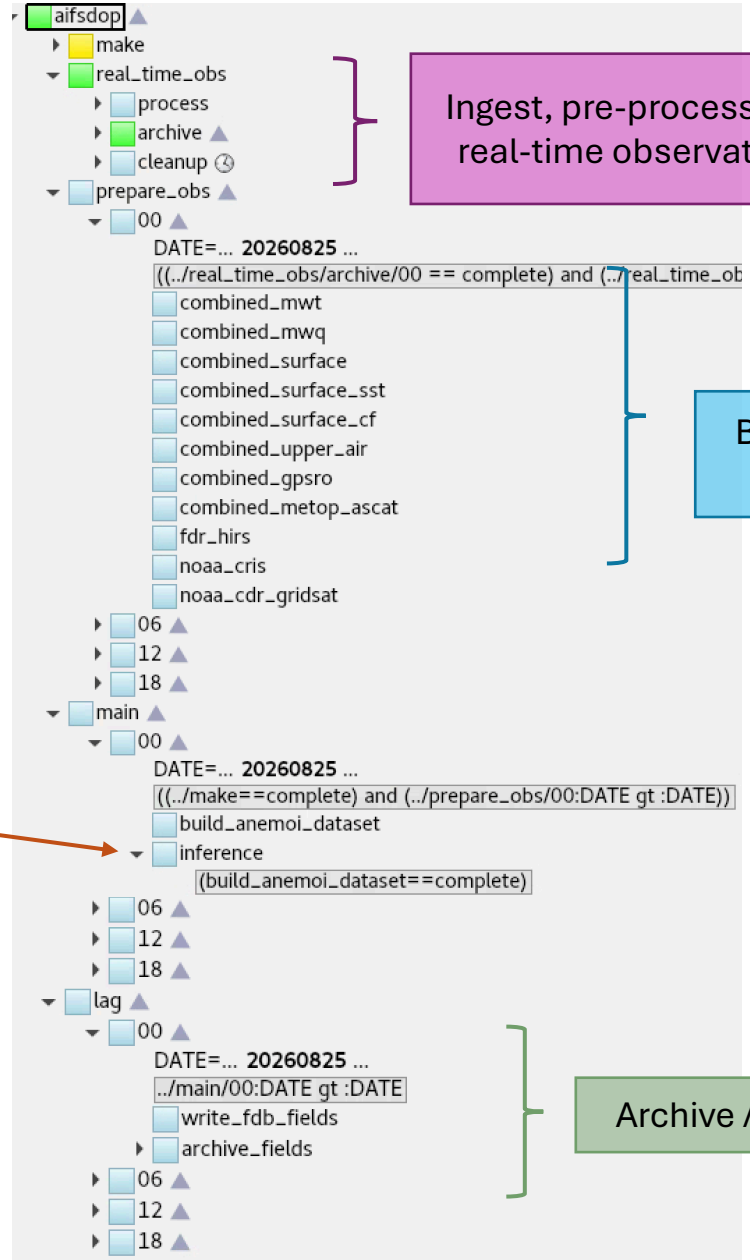
Reading (51.46°N, -0.97°E) — Forecast Timeseries
Init: 20260917 00Z | steps 6–240h / 6h
IFS | DOP O96 6h (j1fw) | AIFS-DOP O96 6h (jba3)





And in Greece for Tony





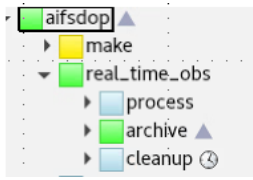
Ingest, pre-process and archive real-time observations (SAPP)

Build input ZARR required by inference model (e.g. grid, tabular, 6hr, 36hr...)

Option to plug in multiple inference models with different architecture and training (but **only** supplied with OBS for real-time input)

Archive / verify

Observation acquisition



- Work is done in close collaboration with production section
- Embedded in the existing real-time operational observation handling workflow
  - SAPP (Scalable Acquisition and Pre-Processing System) is designed to support real-time observation delivery to 4D-Var
  - Its role is to consolidate millions of observations arriving in a wide variety of dissemination formats into a single standardised representation. SAPP performs periodic extractions of newly processed observations
  - This has been adapted such that the same repository now serves as common source of real-time observations for both 4D-Var and DOP
- For DOP, all steps are performed quasi-continuously, in lockstep with the SAPP 10-minute cycle - in principle DOP forecasts could be run hourly…
  - Equivalent 4D-Var screening is run only after the analysis cut-off time

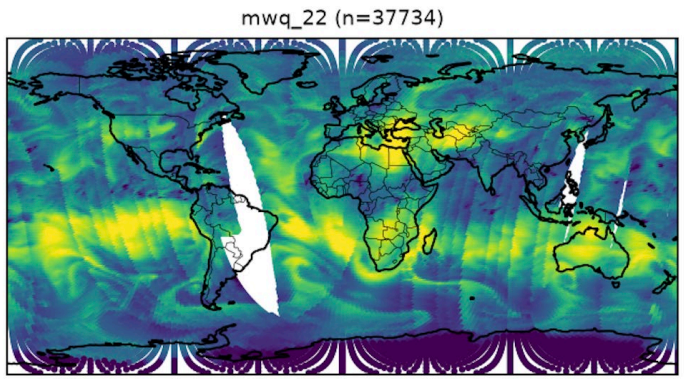
Forecast model AIFS-DOP

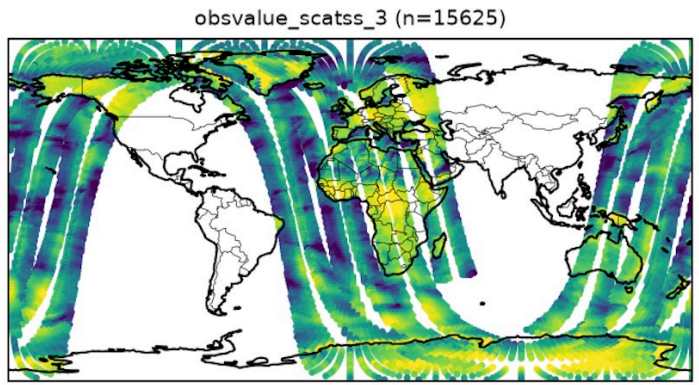
More information in Ewan and Peter's talk tomorrow!

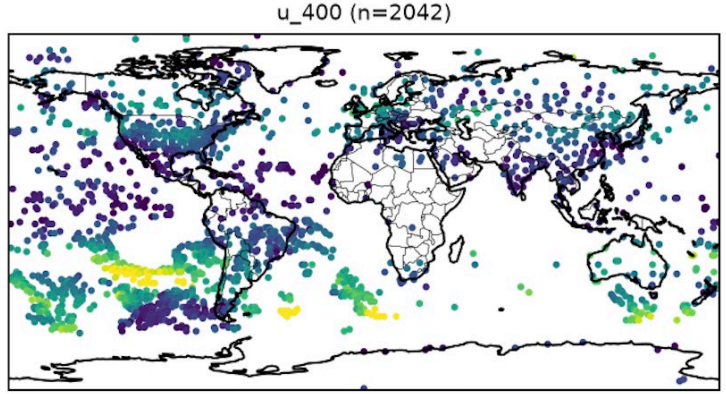
anemoi

- 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.
- Currently O96 (~ 100km) resolution

*** Used to initialise model in real-time**







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

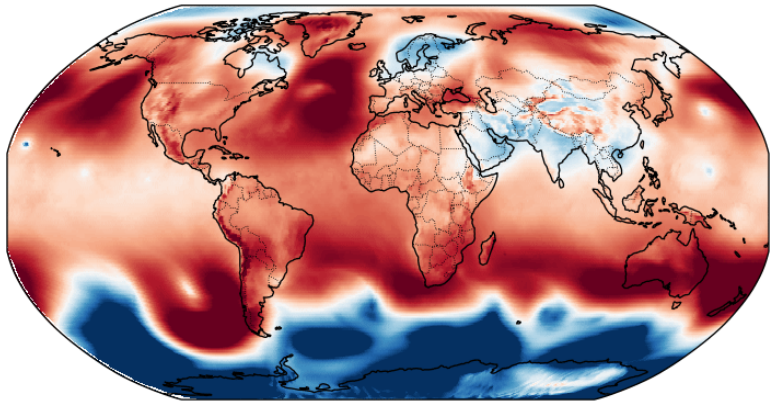
Init: 20260820 00Z | Step: +012h | Valid: 2026-08-20 12Z

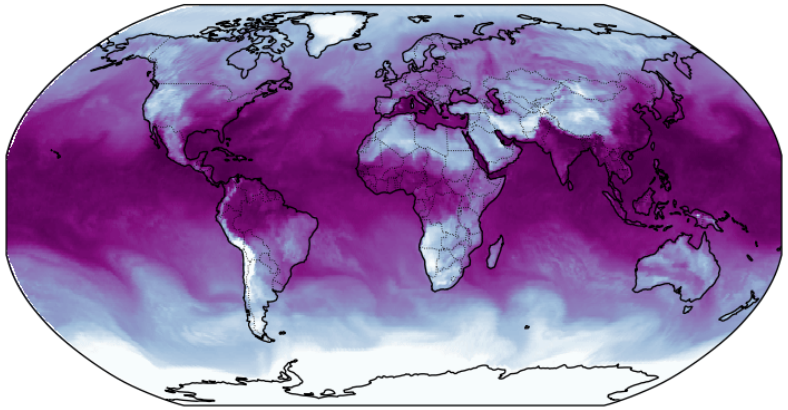
MSLP

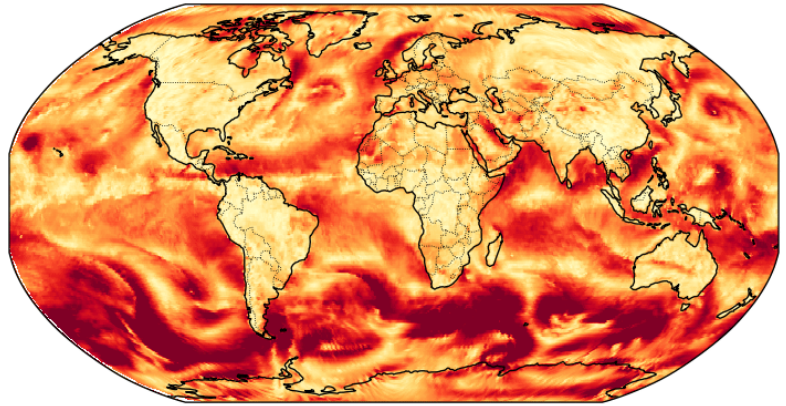
2m Dewpoint temperature

10-metre wind speed

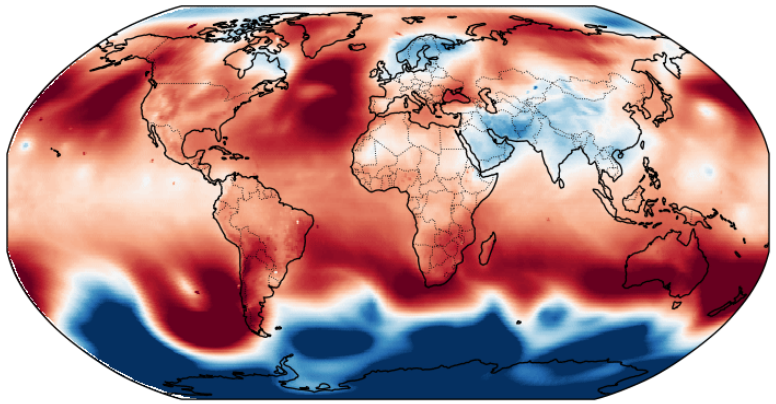
IFS

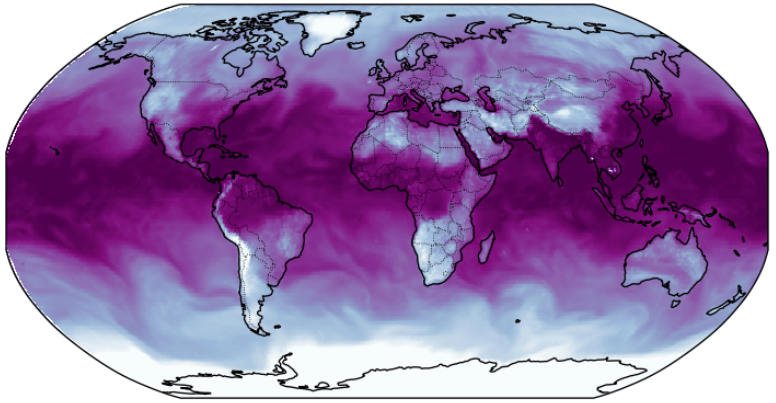


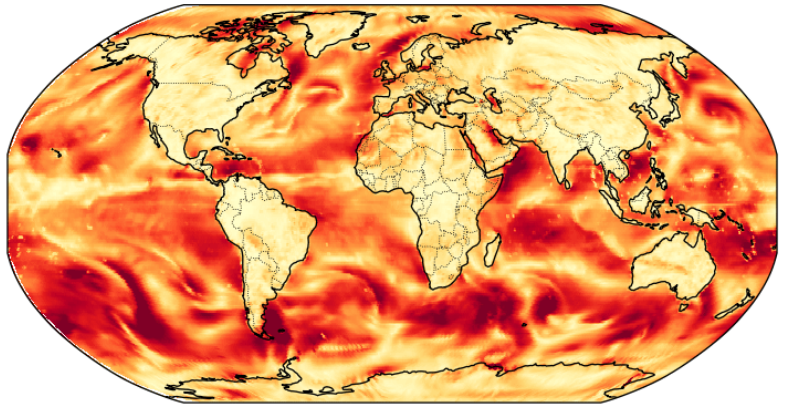


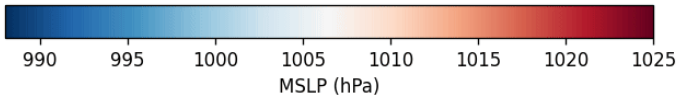


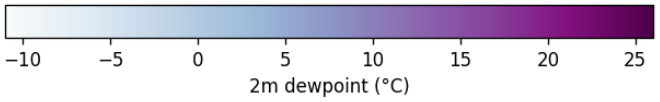
AIFS-DOP

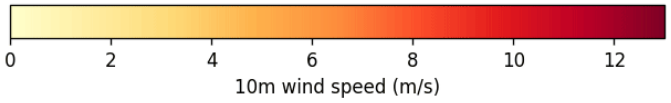










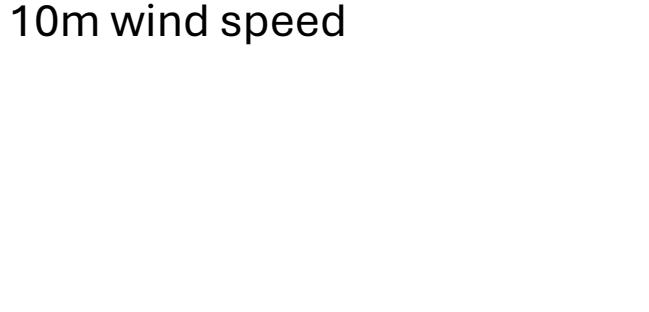


Observation-based model evaluation: surface

AIFS-DOP O96 IFS O96

N Hem

MSLP



10m wind speed

T2m

S Hem

Observation-based model evaluation: upper air

AIFS-DOP O96 IFS O96

N Hem

T850

200hPa wind speed

Z500

S Hem

Analysis-based model evaluation

AIFS-DOP O96 IFS O96

T850

NHem Tropics SHem

Z500

DANA

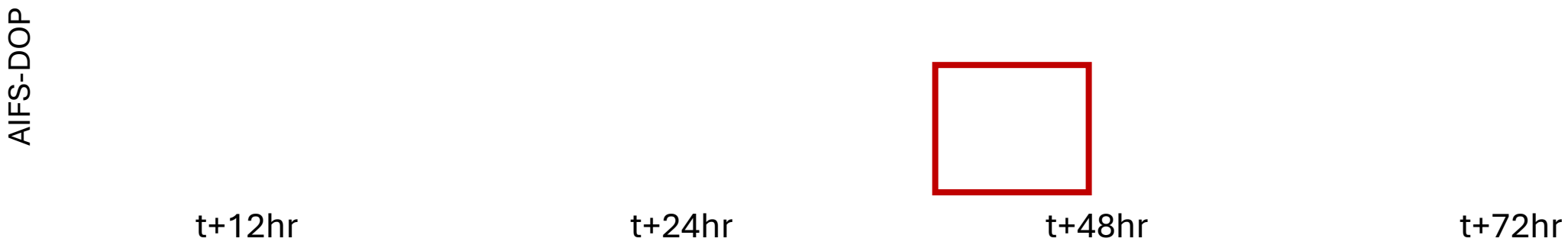
- Deep low pressure suddenly forms off the coast of Spain

  →This is a DANA ("Depresión Aislada en Niveles Altos"), a pocket of cold air that becomes cut off from the main upper-level circulation.

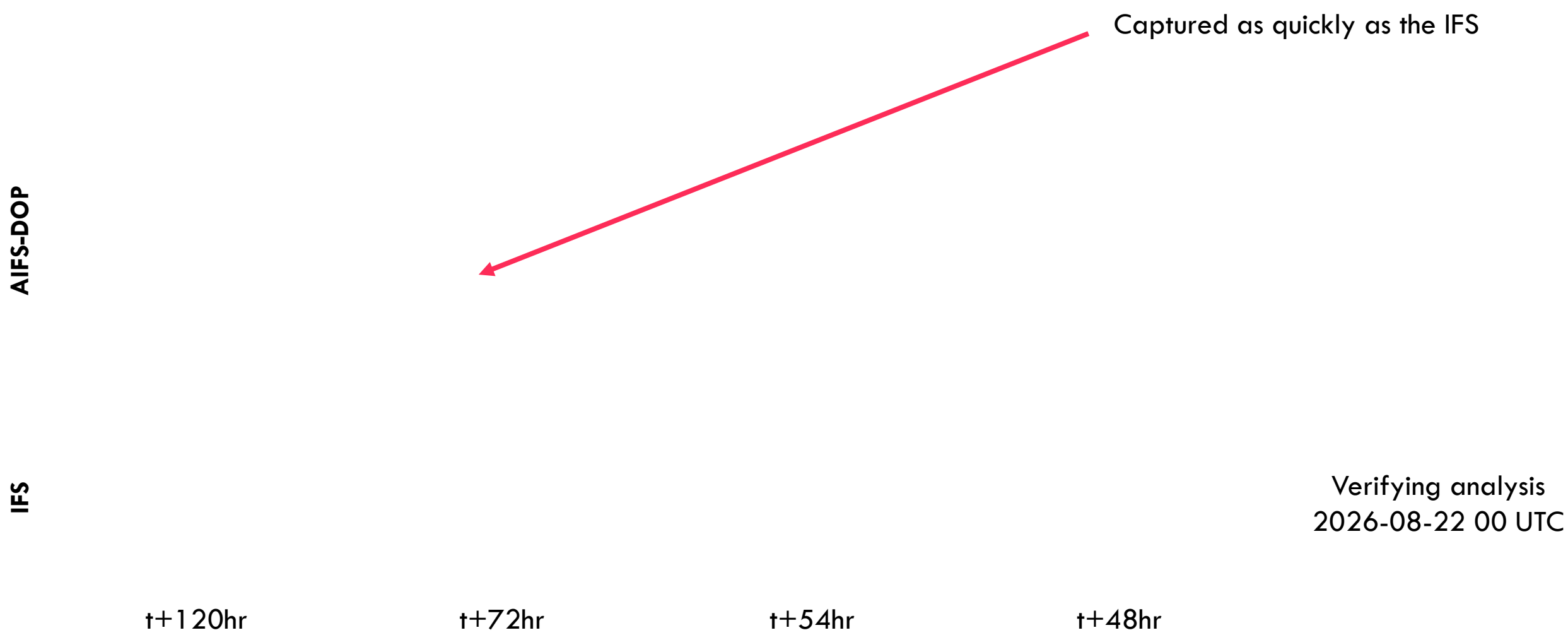
  → The cold air aloft creates strong atmospheric instability when it sits above warm, moist air near the surface.

  → This can trigger intense, slow-moving thunderstorms and extreme rainfall, particularly around the Mediterranean.

- Unusual for mid-August: DANAs are not exceptionally rare over Spain/Mediterranean, but a deep, isolated low this early in summer is unusual and can produce exceptional rainfall

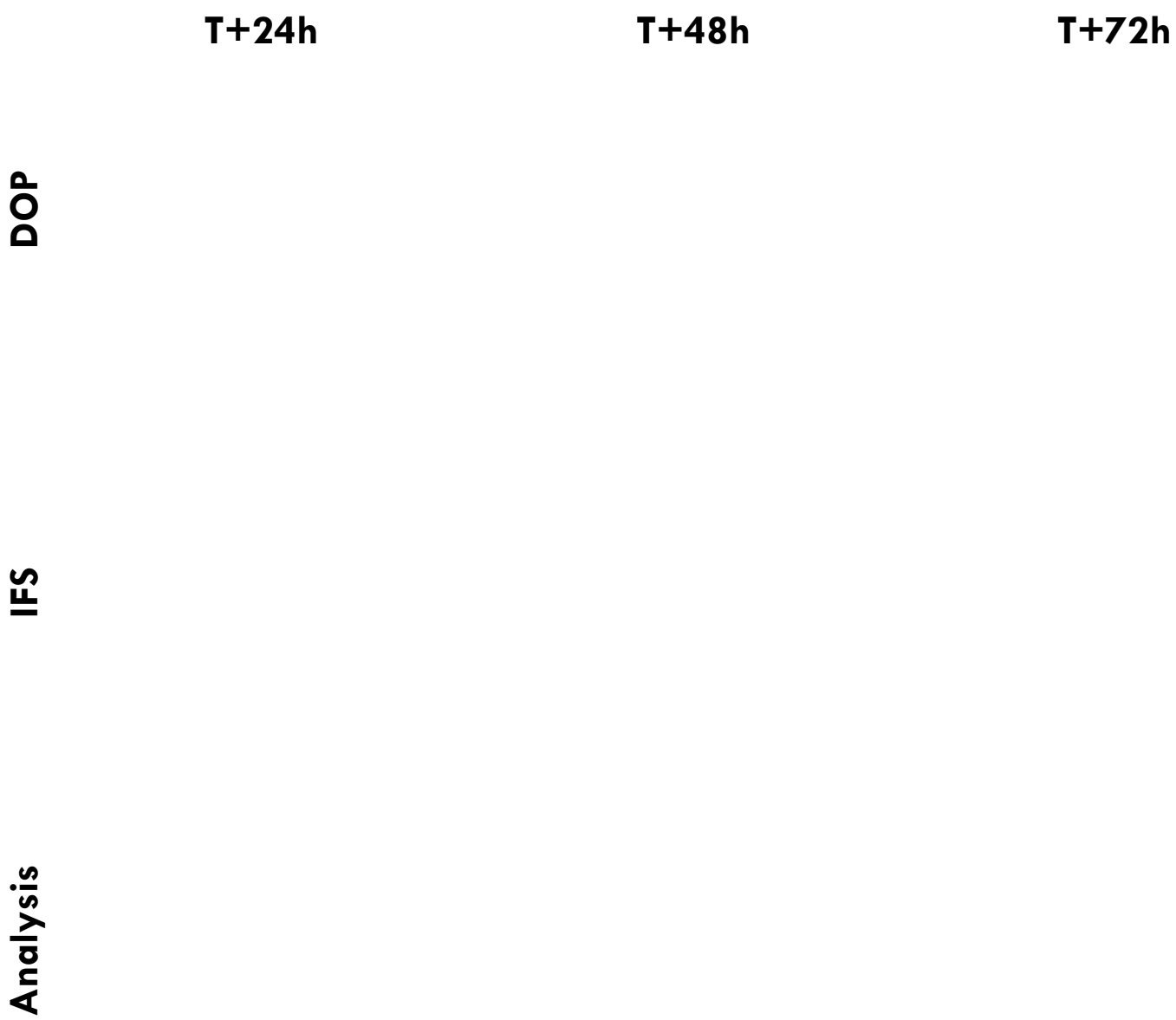


DANA


Captured as quickly as the IFS
AIFS-DOP
IFS
Verifying analysis
2026-08-22 00 UTC
t+120hr
t+72hr
t+54hr
t+48hr

Three hurricanes

- Warming in the Eastern Pacific Ocean caused by El Nino has created conditions for more storms to develop in this region in 2026
- Hurricanes Lowell, Karina, and Marie formed in late August and early September 2026
- The last time there were three hurricanes simultaneously in the Eastern or Central Pacific was in 2015
- AIFS-DOP anticipates the formation of all three hurricanes up to 72 hours ahead



Challenges of running in real-time

Changes in the observing system

- Global observing system is extremely dynamic : scheduled transitions and unplanned events (degraded performance, failure, receiving problem etc.)
- In the IFS, an unexpected observation change will trigger immediate action to manually exclude instruments to prevent poor-quality observations entering the 4D-Var → Minor detrimental impact upon the IFS forecast quality

What about the impact in a DOP-like system?

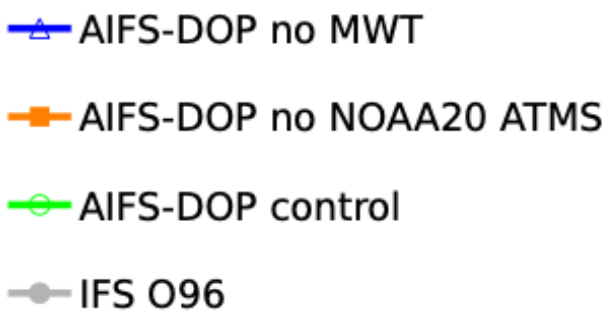
- Started to run **Observing System Experiments (OSE)** to estimate the impact of observation types in DOP and analyse the impact of losing a system unexpectedly.

Challenges of running in real-time

Changes in the observing system

RMSE against observations for T850 forecasts in the Nhem / July-August 2026

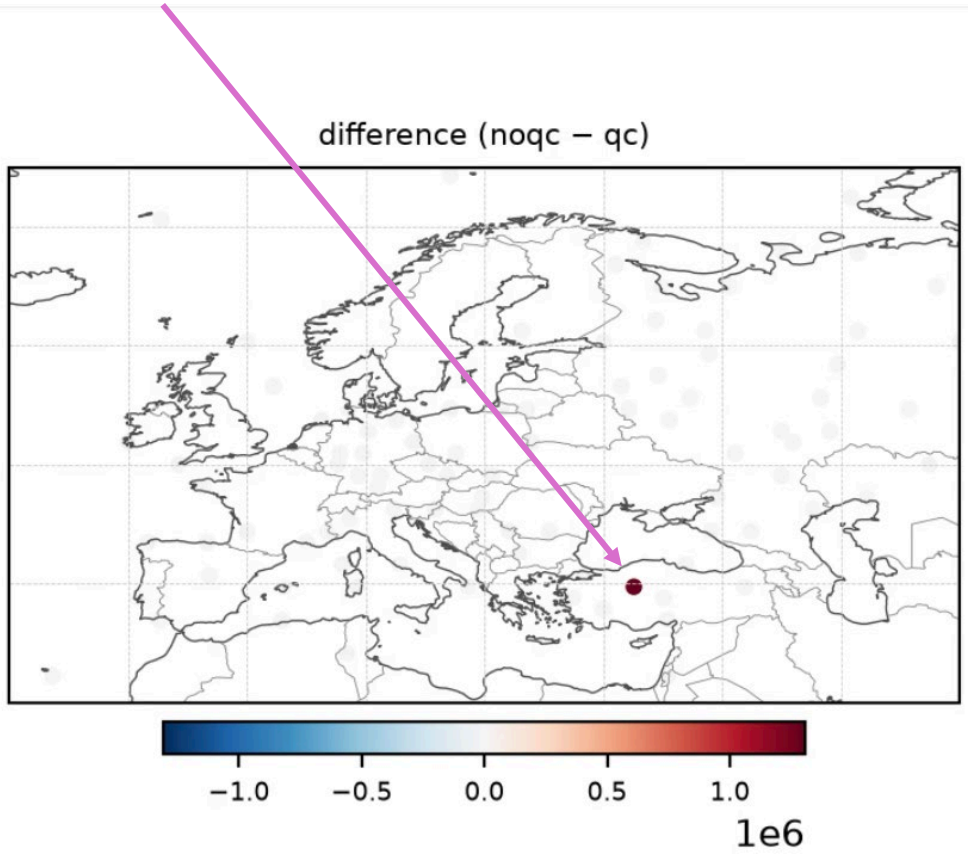
- Very little degradation if we lose just one MW instrument (in orange)
- However quite a considerable degradation if we lose all MW temperature sounding channels (in blue), especially at longer lead times.
- Still more work needed to try and control this degradation

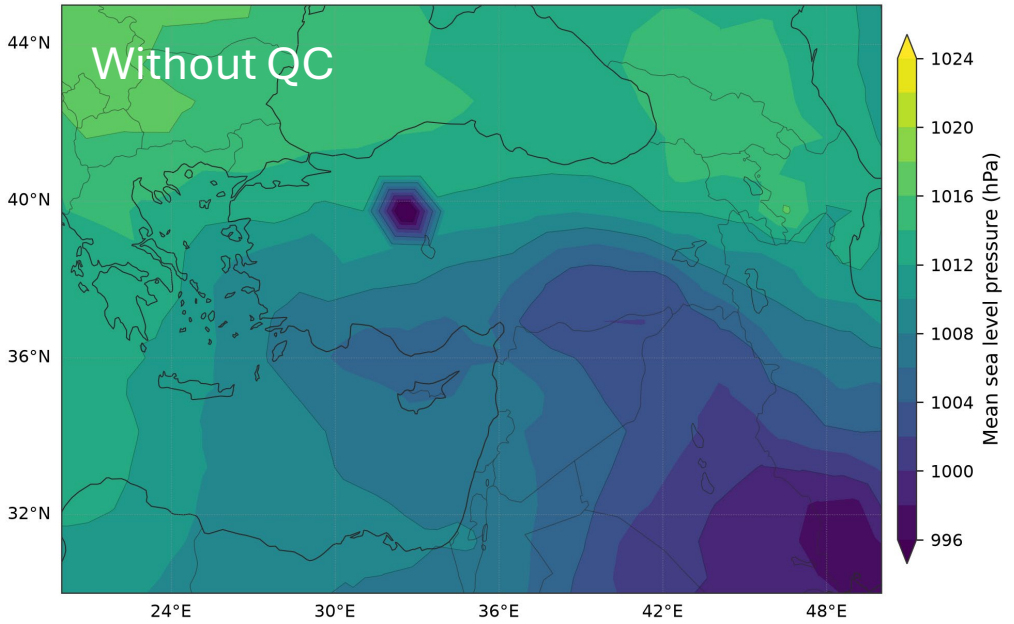


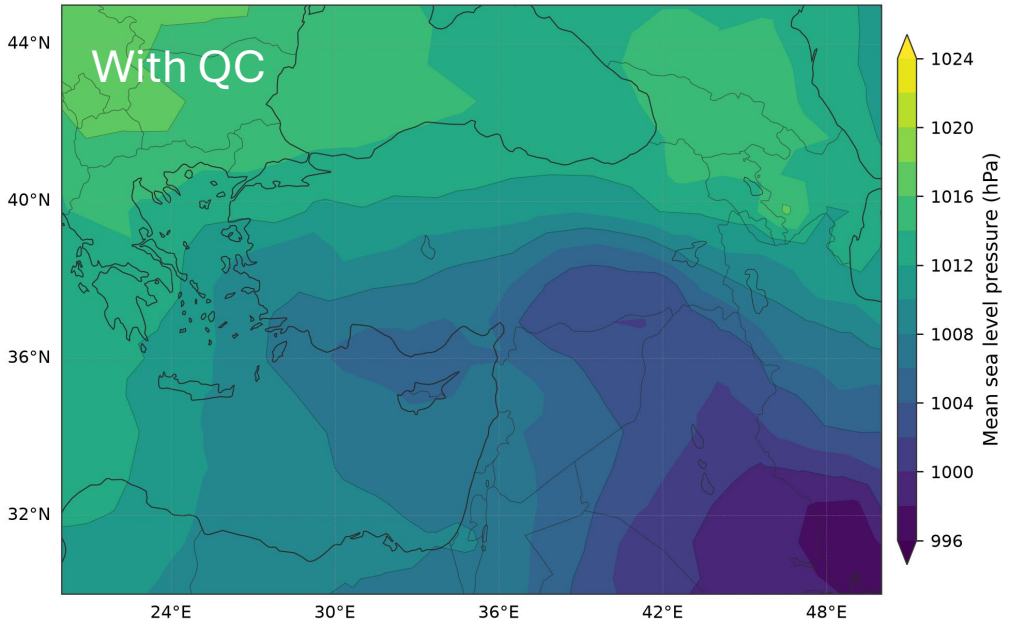
Challenges of running in real-time

Managing quality control

- A single erroneous radiosonde measurement over Türkiye generated a large pressure error in the forecast.







Challenges of running in real-time

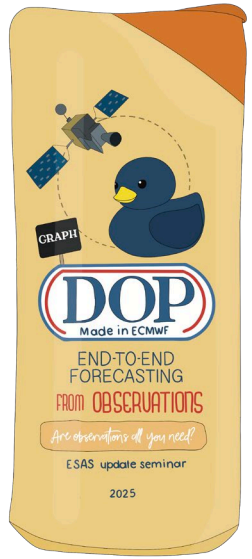
Managing quality control

- Observations from active data sources can show intermittent or random anomalies.
- 4D-Var protects against poor-quality observations through real-time QC, rejecting or downweighting suspect data (VarQC / Huber Norm).

- **QC is also important for observation-driven ML models, both in training and inference:**
  - *Training*: manual QC, IFS blacklist, reprocessed/homogenised datasets (if ignored, can lead to instabilities in training, encourages built-in cross-attention QC if training data is clean)
  - *Real time inference*: range checks and masking of known bad channels/stations/radiosondes
  - Next step: background checks for AIFS-DOP - compare incoming observations with values predicted from the previous observation window and reject observations with large discrepancies.

Conclusions and future plans

- **We implemented a prototype for producing real-time medium-range forecasts from observations alone**, with observations ingested, processed and delivered directly to the AIFS-DOP inference model.
- AIFS-DOP can operate **quasi-continuously alongside the real-time observing system** and can capture rapidly developing weather systems.
- Running an observation-driven model in real time introduces a number of challenges beyond the model itself:
  - **Observation quality control:** addressing erroneous or anomalous observations in both training and real-time inference.
  - **Changes in the observing system:** understanding and mitigating the impact of instrument failures and changes in observation availability.
- **Several challenges remain to be investigated**, e.g. how to incorporate new instruments and manage model retraining as the observing system evolves.
- **Next steps include training/inference at higher resolution, accounting for observation errors, onboarding new observation types, probabilistic training, improvements in analysis-based verification …**



Thank you!
Questions?



www.ecmwf.int