The prospects for future Machine Learning climate reanalysis directly from observations

DOP Team & Others

9093v1 [physics.ao-ph] 17 Jun 2026

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)

June 18, 2026

Abstract

We introduce the Artificial Intelligence Forecasting System for Direct Observation Prediction (AIFS-DOP). AIFS-DOP is trained on a 40-year harmonized dataset of gridded observations, without using numerical weather prediction (NWP) reanalysis or model data. The resulting model is competitive with ECMWF's Integrated Forecasting System (IFS) when scored on a one year period of forecasts across 2021/2022. This progress on Direct Observation Prediction represents the first time that a data-driven model, *trained solely on observations*, is competitive with the IFS at medium ranges for several key upper-air and surface headline scores, when verified against observation data.

[physics.ao-ph] 8 Jul 2026

Global reanalysis from observations alone with machine learning

Peter Lean[1,*], Ewan Pinnington[1], Patrick Laloyaux[1], Mihai Alexe[2], Eulalie Boucher[1], Simon Lang[1], Tomas Kral[1], Paul Poli[2], Hans Hersbach[1], Niels Bormann[1], Matthew Chantry[1], Anthony McNally[1]

[1]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
[2]ECMWF, Bonn, Germany
[*]Correspondence: peter.lean@ecmwf.int

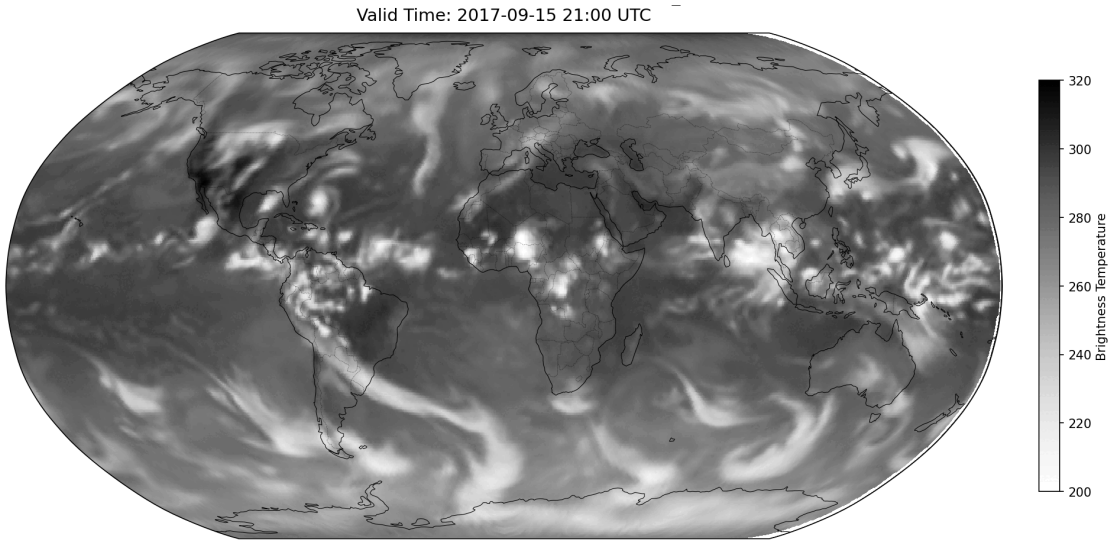
Abstract

Earth system reanalysis datasets are foundational for weather and climate research and provide the gridded training data used by most machine learning weather prediction systems. Here we show results from a prototype system that suggest that machine learning models trained only on Earth system observations can potentially be used to generate multi-decade global reanalyses without using physics-based numerical models. The resulting gridded fields capture large-scale atmospheric structure and variability across multiple timescales, while exhibiting signs of physical coherence in several key dynamical diagnostics. Evaluations of the prototype against held-out independent atmospheric observations indicate that the root mean square vector error of upper-level winds is close to that of ERA5 when compared at a consistent resolution, and that the standard deviation of the error at the surface is between that of 4th- and 5th-generation ECMWF reanalyses (ERA-Interim and ERA5). Furthermore, while traditional reanalysis production is computationally expensive, typically taking several years to produce, the reanalysis presented here was generated during the course of a single working day. These results suggest that observation-trained machine learning models offer a promising new approach for reanalysis production from observations alone.

Talk Outline

- Part 1: DOP in Anemoi & Medium-Range Scores
  - Training dataset & Model
  - Case Studies and Medium-Range Scores
  - What's new?
  - Summary

- Part 2: DOP Reanalysis Prototype
  - Running 40-years of Reanalysis
  - Case Studies
  - Validation
  - Summary



HeRRi
Climate Hazards Enhanced Reanalyses & Reasoning with AI

eRise

Funded by the European Union

Coordinated by ECMWF

[1] The CERISE project (grant agreement No101082139) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

ML and Observations

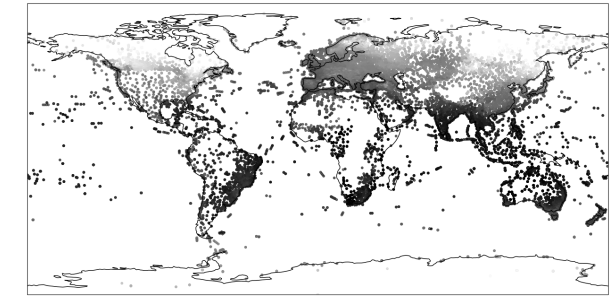
- Training ML models directly from observations
- Using Open-Source Anemoi packages: https://github.com/ecmwf/anemoi-core
- Architecture for this example very similar to the AIFS with some tweaks
- Some simplifications from GraphDOP to allow training under Anemoi packages

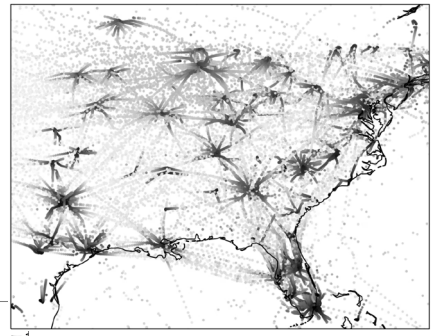
anemoi training

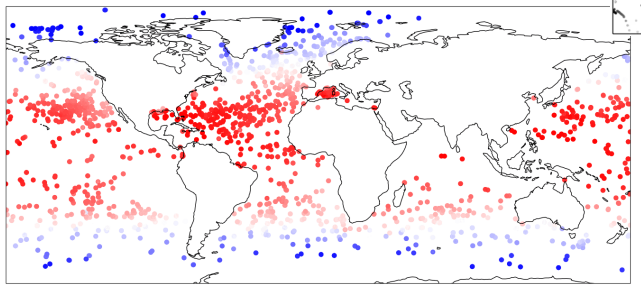
Model architecture the same as AIFS

First Anemoi Examples of Training with Observations

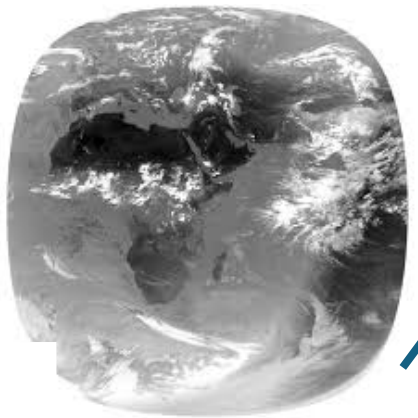
- Putting obs on the same resolution grid in 6-hour slices using DOP-dataset code
- Treat them as additional variables under dataloader
- Slight updates to NaN-handling to allow training/generalising in space (and new “InputOnlyImputer”)



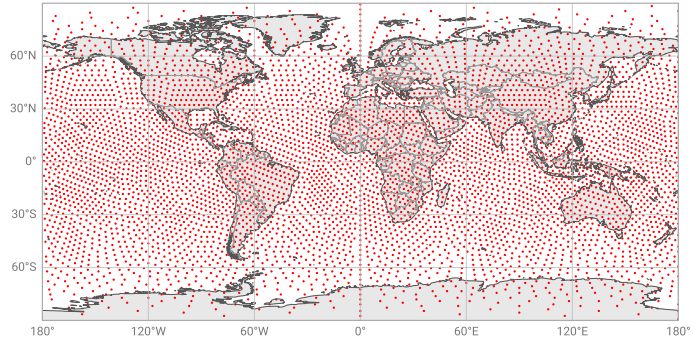








Put obs onto o96 grid
Add NaNs for missing data

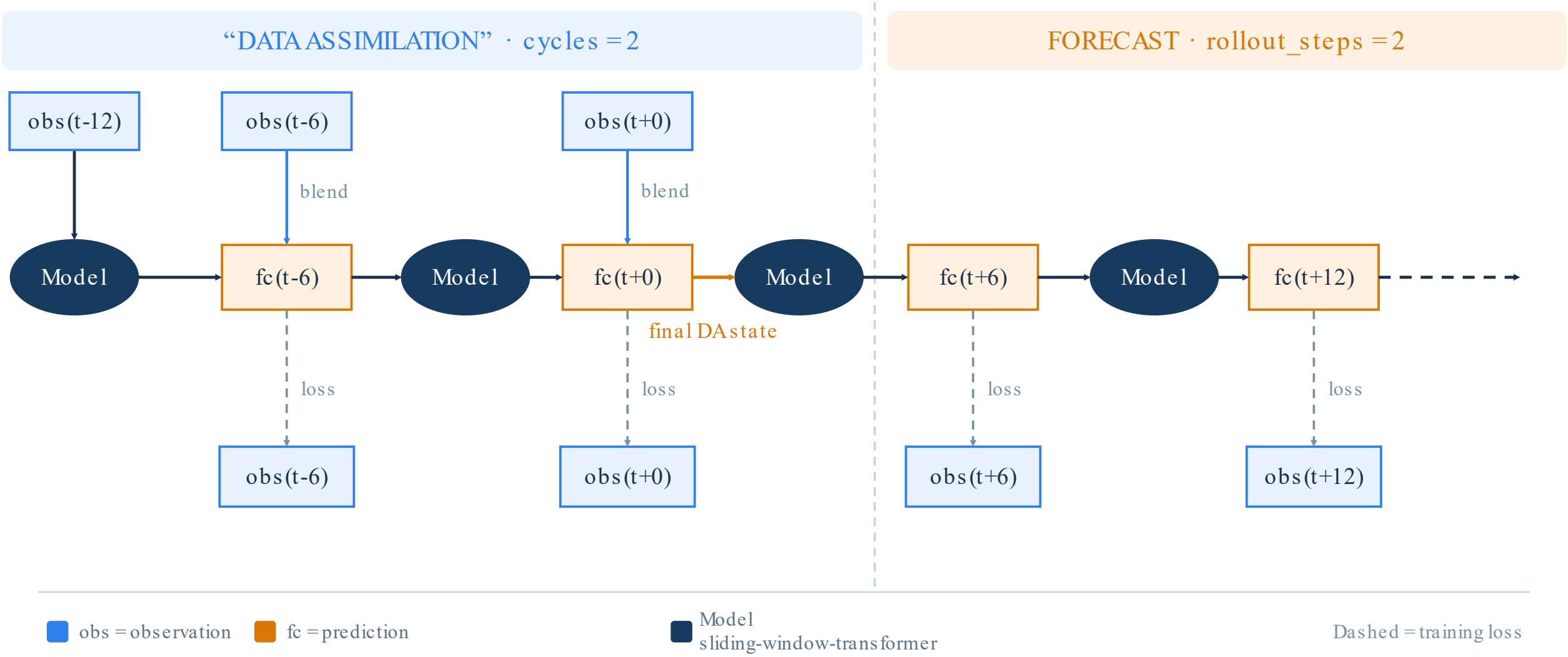


Make anemoi zarr from
Gridded obs nc files

anemoi datasets

From observation-guided assimilation to free-running forecast

Observation and context tensors enter during DA; forecast context contributes only to the training loss.



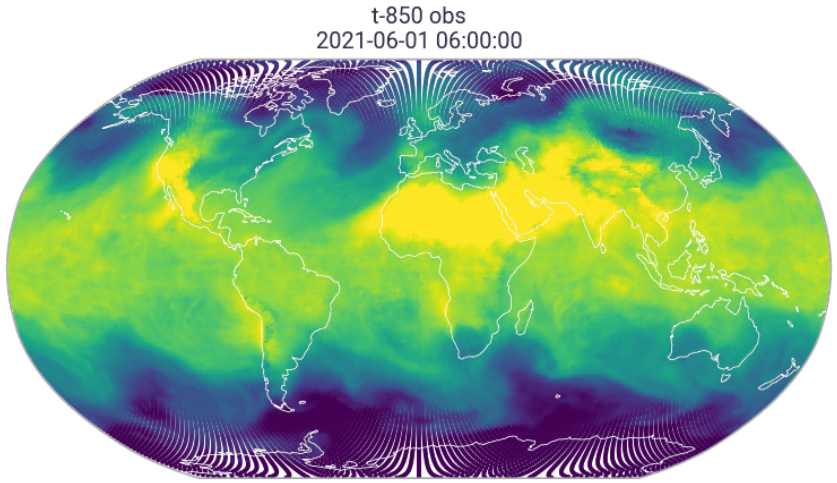


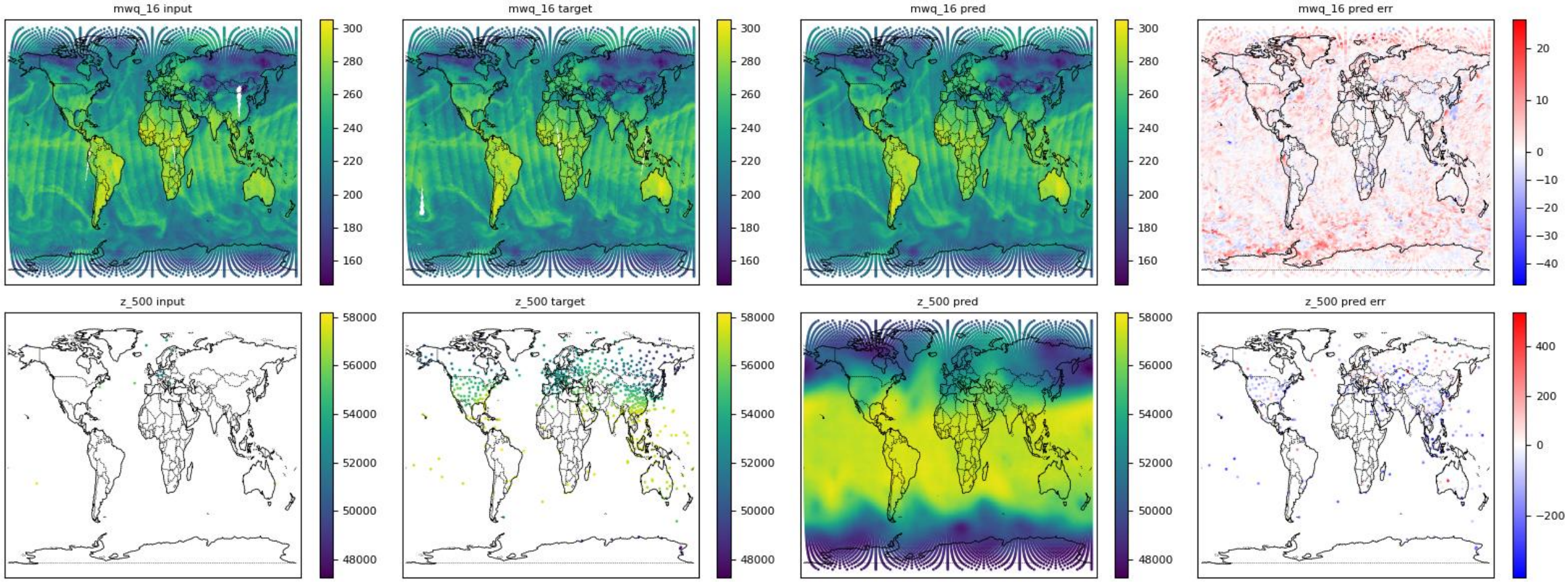
Table 1: Description of curated observation dataset

Category	Instruments	Period	Variables	Details
Infrared Sounder	HIRS	1980–2022	Brightness Temperatures	EUMETSAT Fundamental Data Record [EUMETSAT, 2024]
Microwave Sounders	MSU SSM/T-2 AMSU-A AMSU-B MHS ATMS	1980–2005 1994–2005 1998–2021 1998–2014 2005–2022 2012–2022	Brightness Temperatures	MSU taken from NOAA Climate Data Record [Zou et al., 2013], SSM/T2 taken from EUMETSAT Fundamental Data Record [EUMETSAT, 2020]
Surface Observations	SYNOP, Buoys, Ships	1980–2022	2t, 2d, msl, 10u, 10v	Existing ECMWF data archive
Upper-air Observations	Radiosonde, Aircraft, AMV	1980–2022	t, u, v, z on pressure levels	Existing ECMWF data archive
Geostationary Satellite	GridSat	1980–2022	Brightness Temperatures	NOAA Climate Data Record [Knapp et al., 2011]

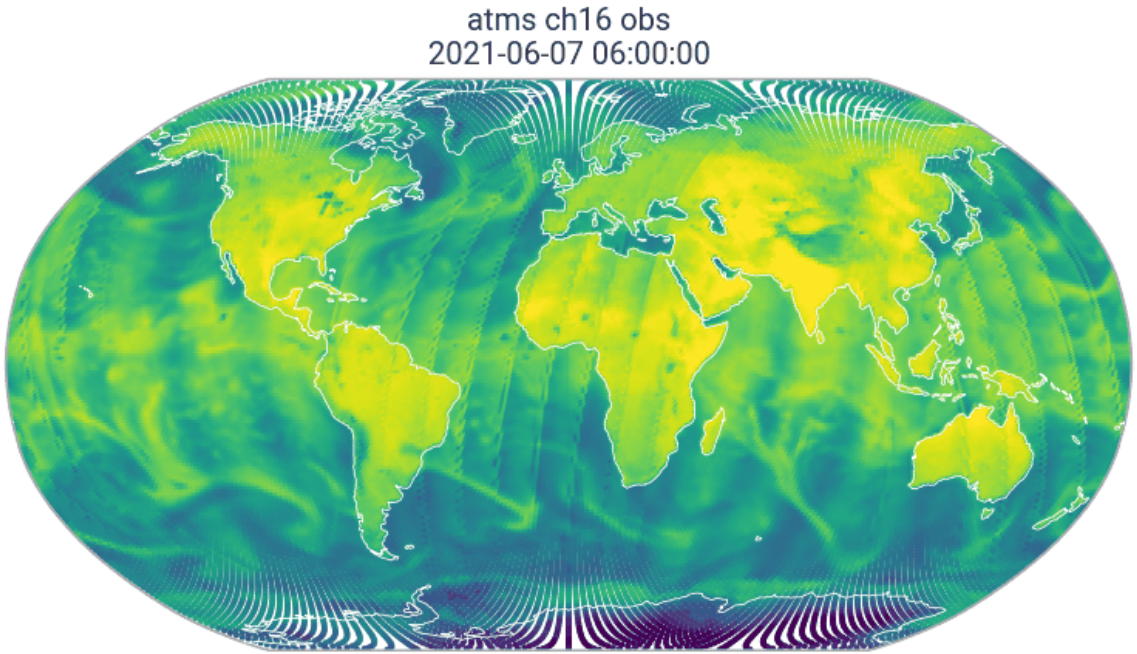
ECMWF

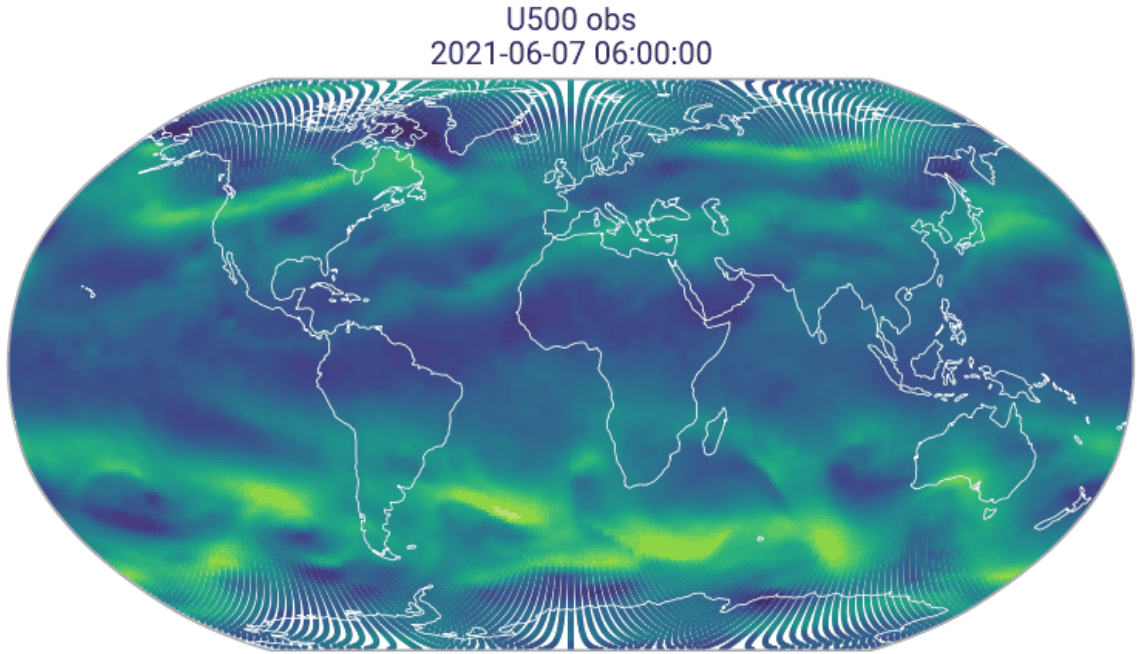
Training Example

Input | Target | Prediction | Error



Training Example





Training Setup

- Training periods
  - Training period: 1981 to 2020
  - Validation period: 2021-01 to 2021-05
  - Test period: 2021-06 to 2022-06
  - 3-day rollout finetuning: 2002 to 2020
- Model resolution o96 (~100 km resolution)

Table 1: Description of curated observation dataset

Category	Instruments	Period	Variables	Details
Infrared Sounder	HIRS	1980–2022	Brightness Temperatures	EUMETSAT Fundamental Data Record [EUMETSAT, 2024]
Microwave Sounders	MSU SSM/T-2 AMSU-A AMSU-B MHS ATMS	1980–2005 1994–2005 1998–2021 1998–2014 2005–2022 2012–2022	Brightness Temperatures	MSU taken from NOAA Climate Data Record [Zou et al., 2013], SSM/T2 taken from EUMETSAT Fundamental Data Record [EUMETSAT, 2020]
Surface Observations	SYNOP, Buoys, Ships	1980–2022	2t, 2d, msl, 10u, 10v	Existing ECMWF data archive
Upper-air Observations	Radiosonde, Aircraft, AMV	1980–2022	t, u, v, z on pressure levels	Existing ECMWF data archive
Geostationary Satellite	GridSat	1980–2022	Brightness Temperatures	NOAA Climate Data Record [Knapp et al., 2011]

Not including!: GPSRO, hyperspectral IR, MW imagers, scatterometer + others

Storm Eunice, February 2022

Results

NH

SH

T850 ACC

Tropics

ECMWF

NH

SH

Z500 ACC

Tropics

ECMWF

NH

SH

Wind 250 ACC

Tropics

ECMWF

NH

SH

T2m RMSE

Tropics

ECMWF

NH

SH

10m Wind RMSE

Tropics

ECMWF

Moving to real time

- Eulalie's talk yesterday!
- Retraining model up to end of 2024
- Transitioning from older sensors to newer ones
  - *e.g.* HIRS missing – pickup CrIS
- Initial results look promising

Table 1: Description of curated observation dataset

Category	Instruments	Period	Variables	Details
~~Infrared Sounder~~	~~HIRS~~	~~1980–2022~~	~~Brightness~~ Temperatures	EUMETSAT Fundamental Data Record [EUMETSAT, 2024]
Microwave Sounders	MSU SSM/T-2 AMSU-A AMSU-B MHS ATMS	1980–2005 1994–2005 1998–2021 1998–2014 2005–2022 2012–2022	Brightness Temperatures	MSU taken from NOAA Climate Data Record [Zou et al., 2013], SSM/T2 taken from EUMETSAT Fundamental Data Record [EUMETSAT, 2020]
Surface Observations	SYNOP, Buoys, Ships	1980–2022	2t, 2d, msl, 10u, 10v	Existing ECMWF data archive
Upper-air Observations	~~Radiosonde~~, Aircraft, ~~AMV~~	1980–2022	t, u, v, z on pressure levels	Existing ECMWF data archive
~~Geostationary~~ Satellite	~~GridSat~~	~~1980–2022~~	~~Brightness~~ Temperatures	NOAA Climate Data Record [Knapp et al., 2011]

Moving to real time

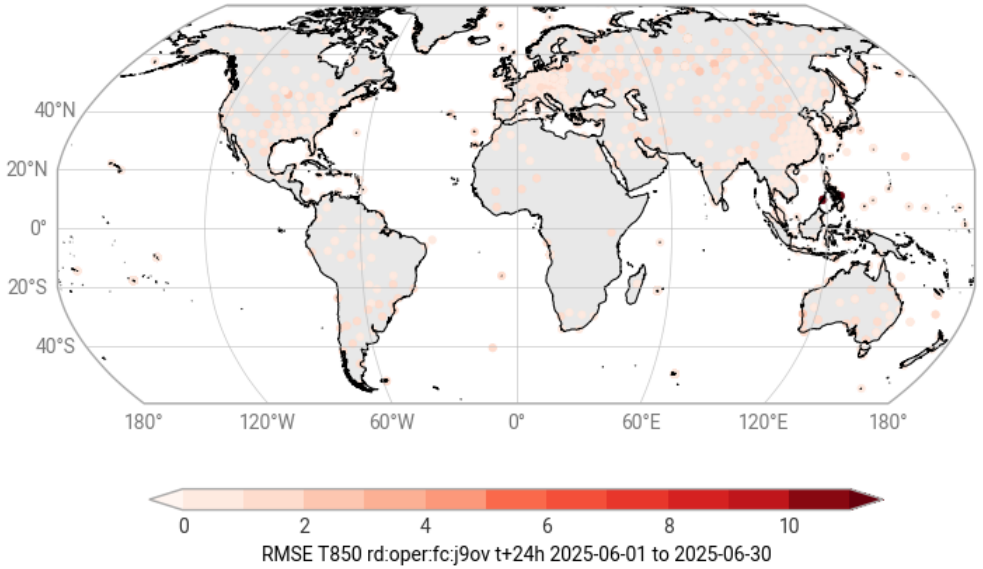
- Eulalie's talk yesterday!
- Retraining model up to end of 2024
- Transitioning from older sensors to newer ones
  - *e.g.* HIRS missing – pickup CrIS
- Initial results look promising

NH T850
RMSE

SH T850
RMSE

Moving to real time

- Eulalie's talk yesterday!
- Retraining model up to end of 2024
- Transitioning from older sensors to newer ones
  - *e.g.* HIRS missing – pickup CrIS
- Initial results look promising



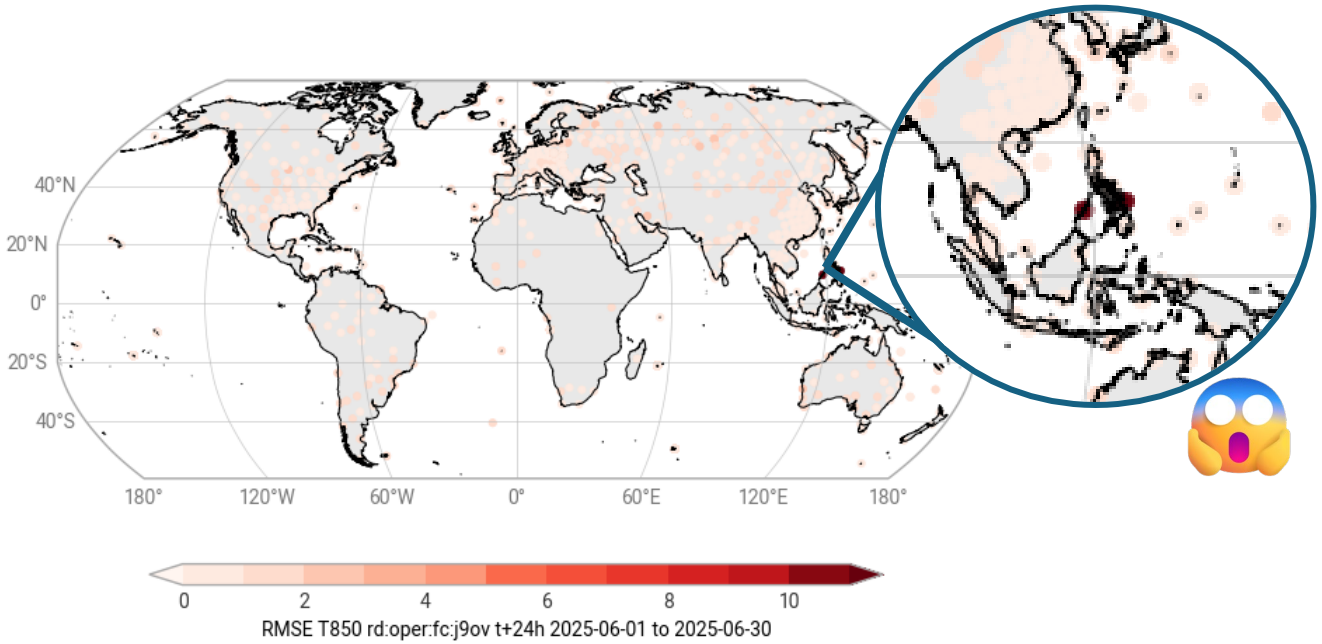
NH T850
RMSE

SH T850
RMSE

Tropics T850
RMSE 😱

Moving to real time

- Eulalie's talk yesterday!
- Retraining model up to end of 2024
- Transitioning from older sensors to newer ones
  - *e.g.* HIRS missing – pickup CrIS
- Initial results look promising



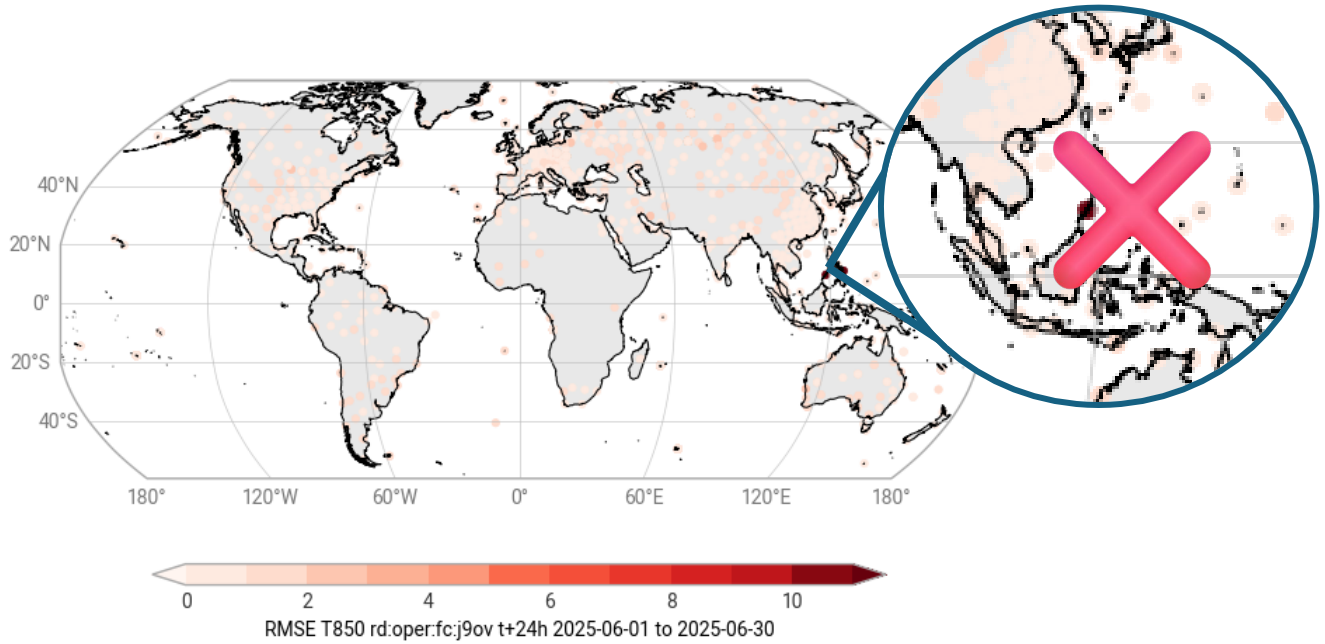
NH T850
RMSE

SH T850
RMSE

Tropics T850
RMSE

Moving to real time

- Eulalie's talk yesterday!
- Retraining model up to end of 2024
- Transitioning from older sensors to newer ones
  - *e.g.* HIRS missing – pickup CrIS
- Initial results look promising



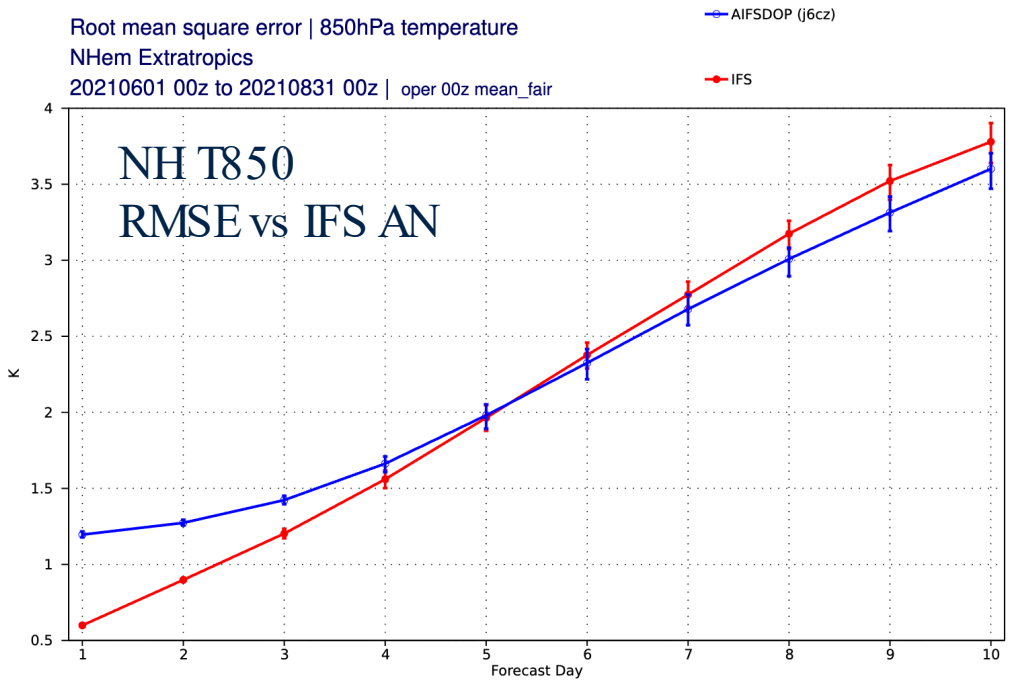
NH T850
RMSE

SH T850
RMSE

Tropics T850
RMSE

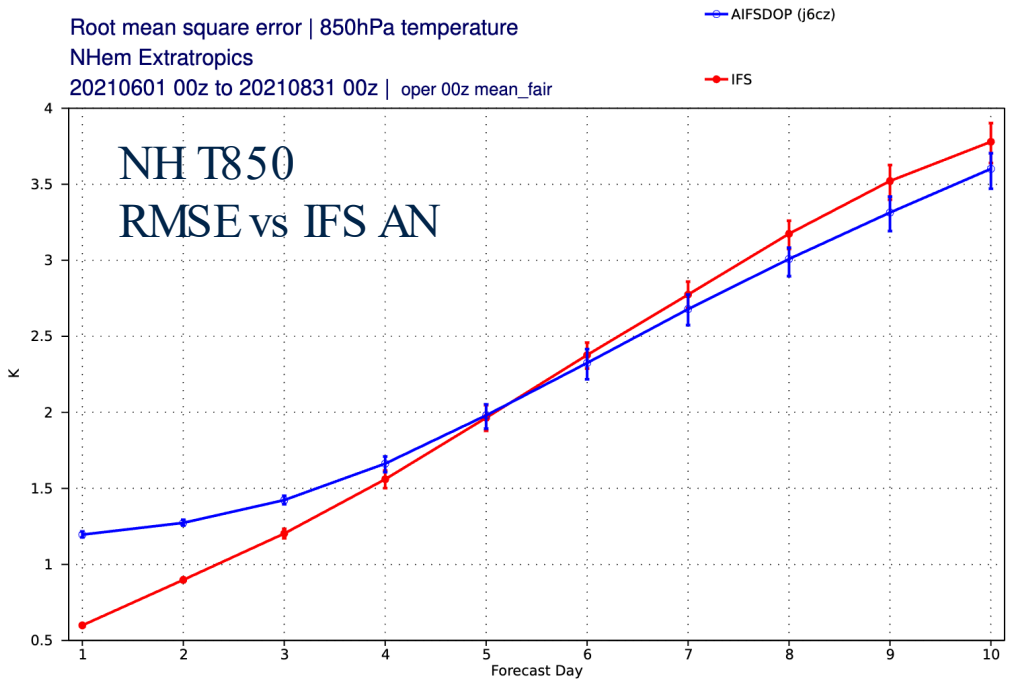
Lies, damned lies and Analysis verification…

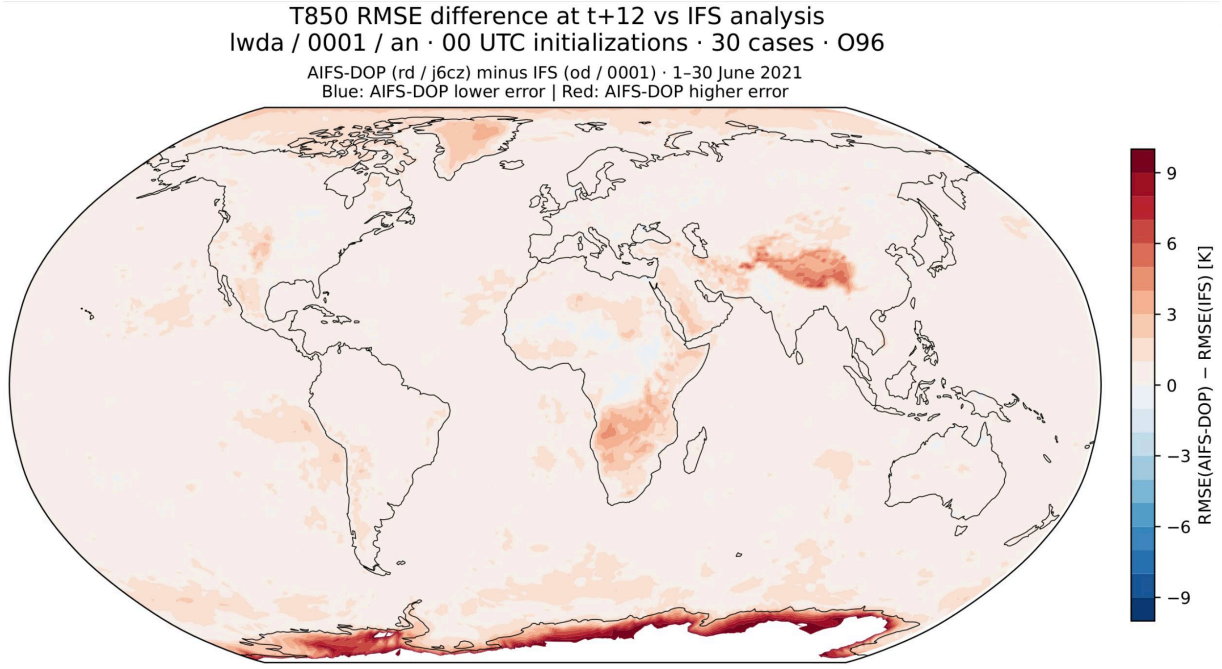
- Work started to understand analysis scores
  - Care needed when making comparisons!



Lies, damned lies and Analysis verification…

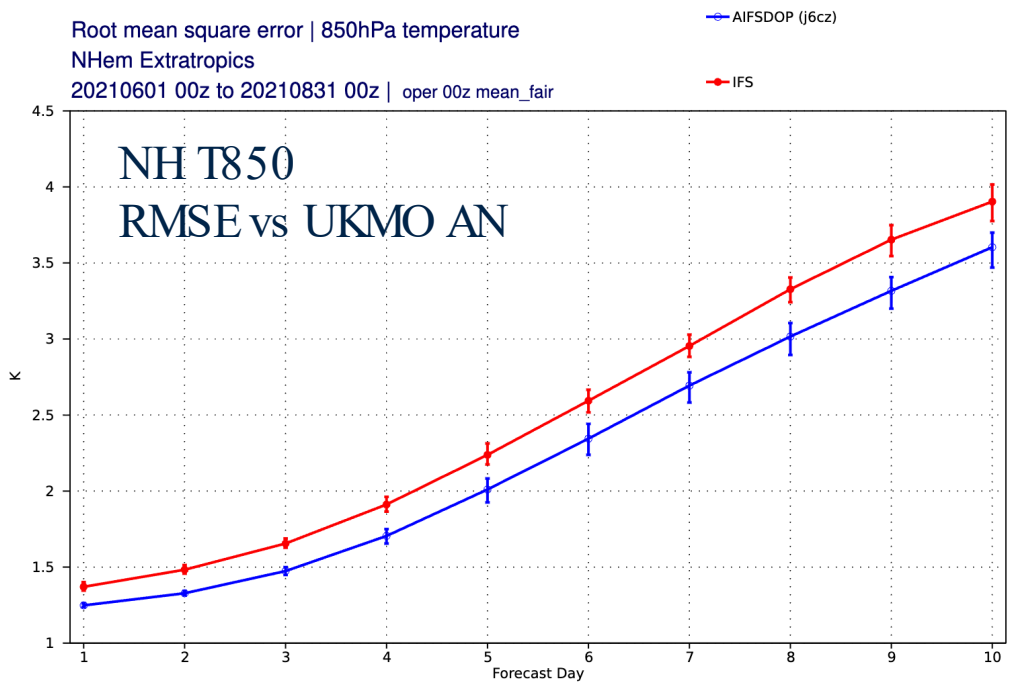
- Work started to understand analysis scores
  - Care needed when making comparisons!

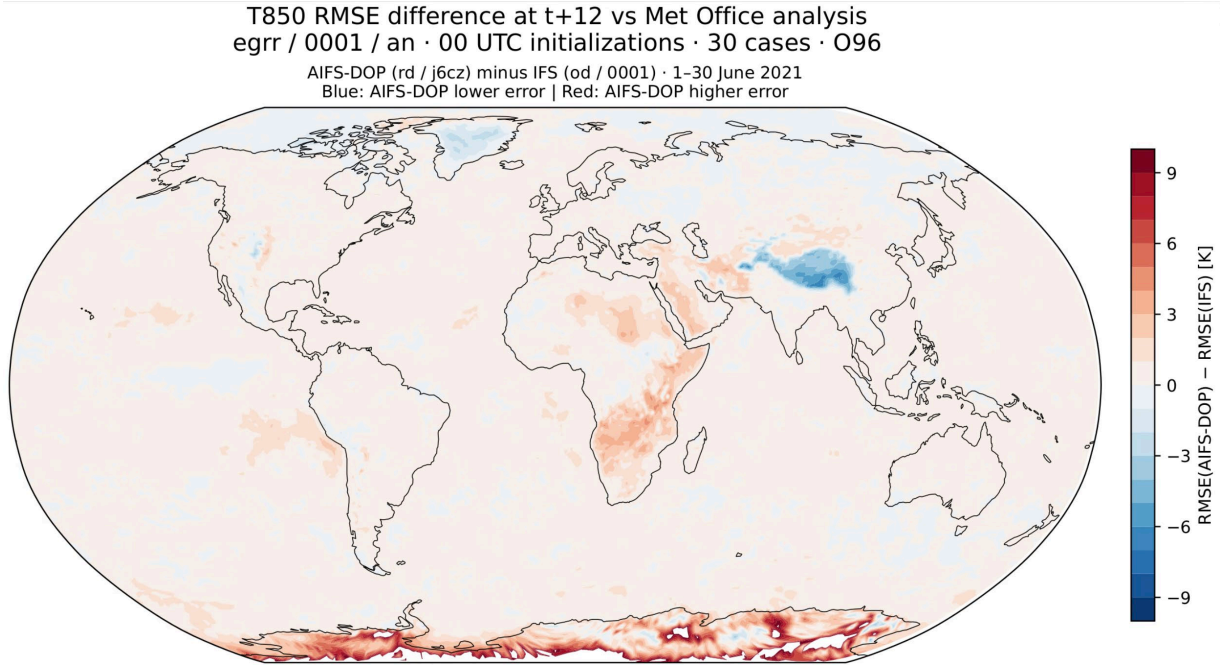




Lies, damned lies and Analysis verification…

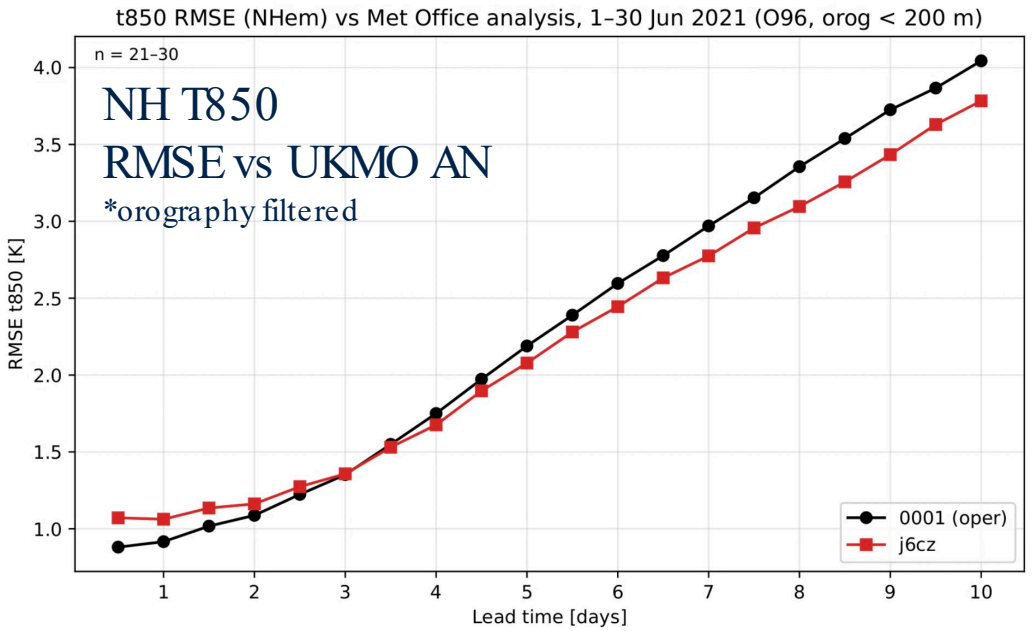
- Work started to understand analysis scores
  - Care needed when making comparisons!





Lies, damned lies and Analysis verification…

- Work started to understand analysis scores
  - Care needed when making comparisons!



Comparisons to Analysis?

- Work started to understand analysis scores
  - Very helpful to diagnose issues and improve results!

NH Wind 250
RMSE

NH T850
RMSE

SH Z500
RMSE

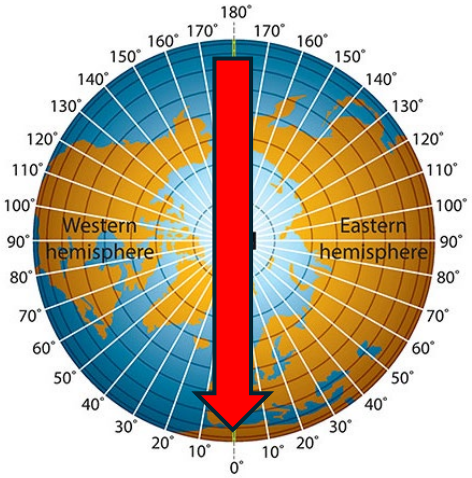
Tropics T850
RMSE

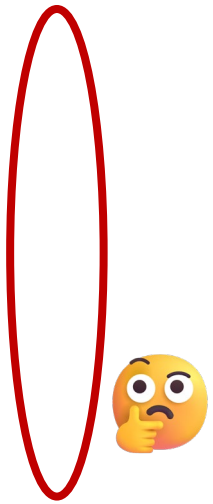
Comparisons to Analysis?

- Work started to understand analysis scores
  - Very helpful to diagnose issues and improve results!
- First paper results has some large errors
  - Strange artefacts??

Comparisons to Analysis?

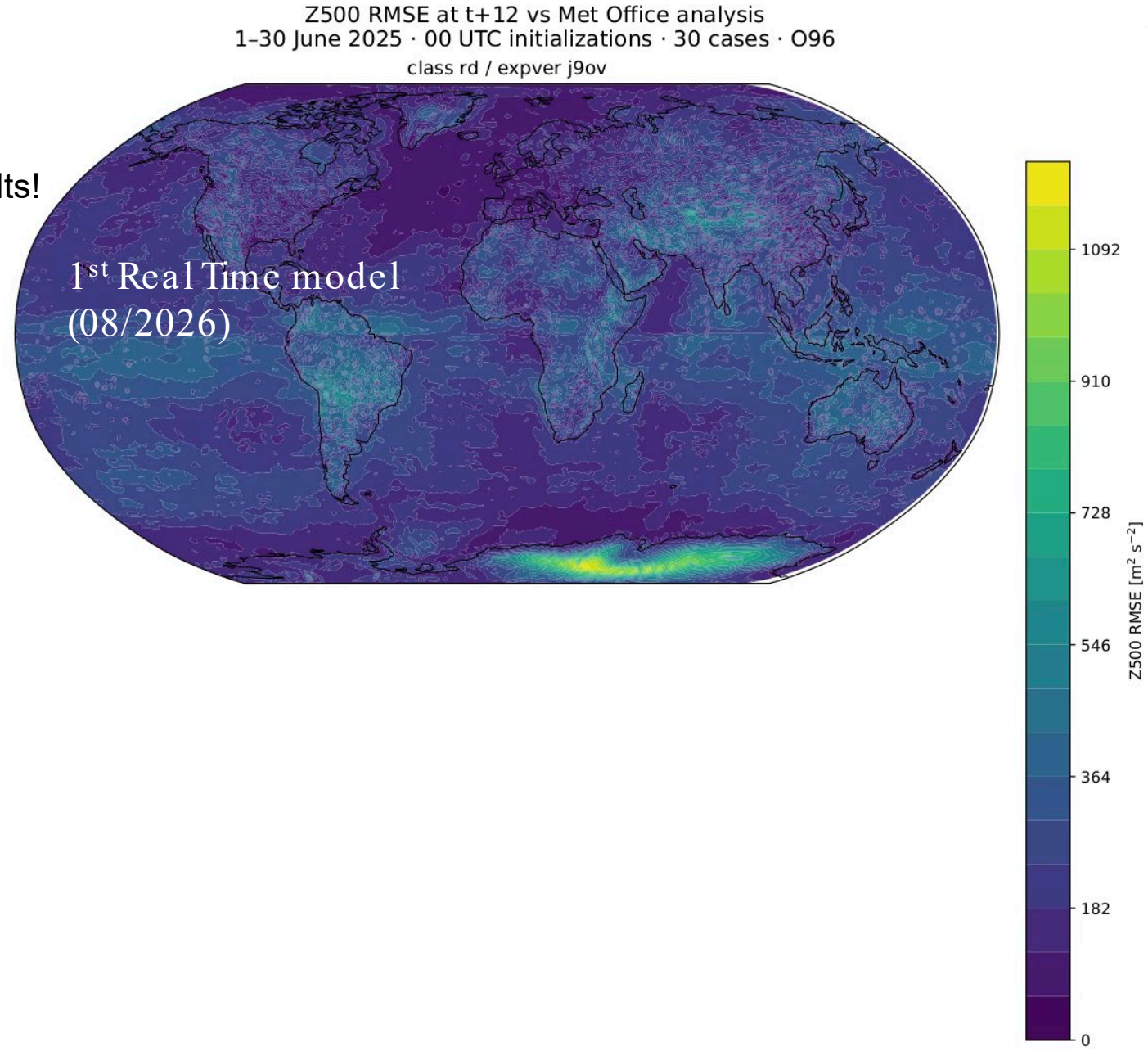
- Work started to understand analysis scores
  - Very helpful to diagnose issues and improve results!
- First paper results has some large errors
  - Strange artefacts??





Comparisons to Analysis?

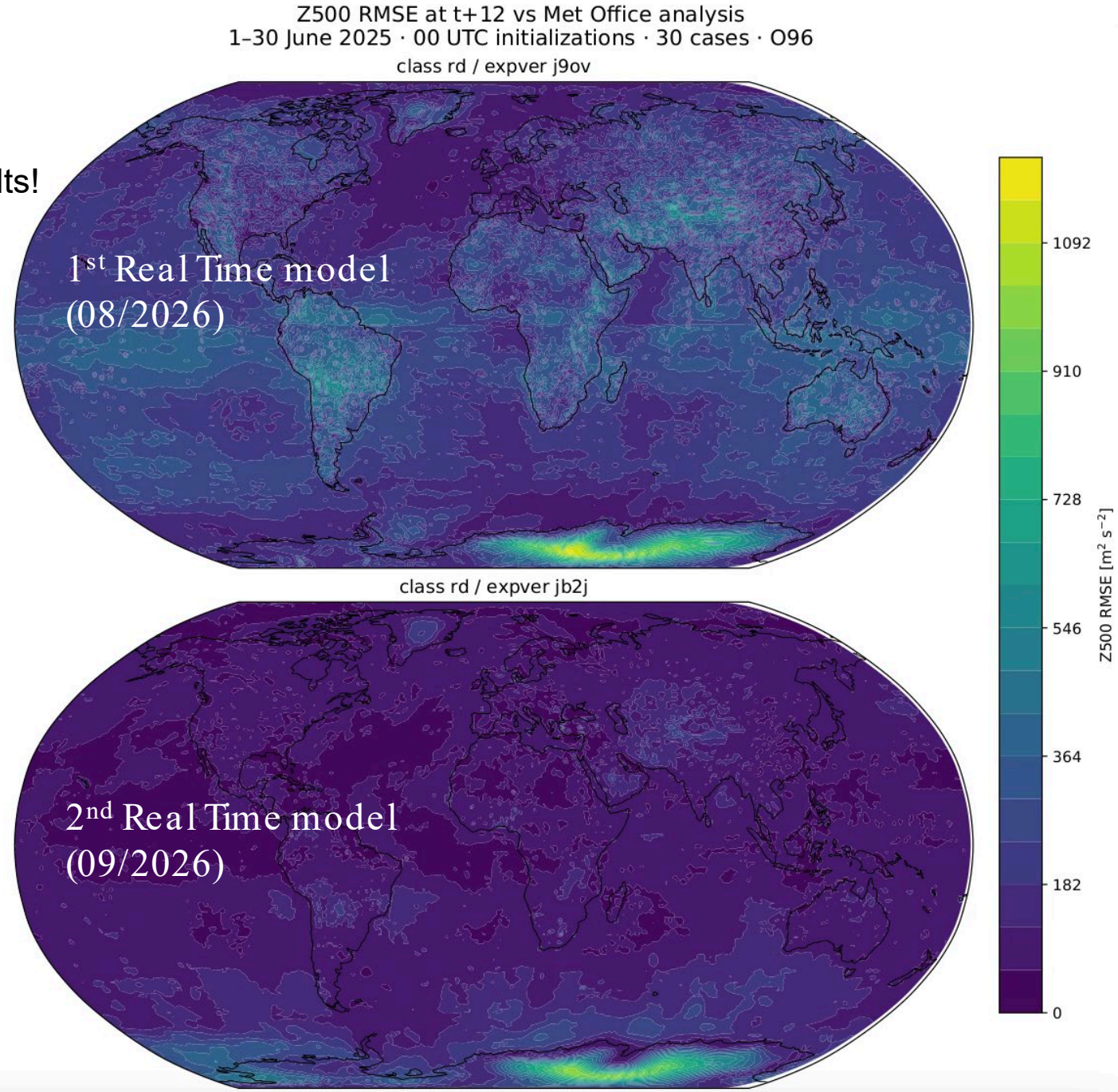
- Work started to understand analysis scores
  - Very helpful to diagnose issues and improve results!
- First paper results has some large errors
  - Strange artefacts??



Comparisons to Analysis?

- Work started to understand analysis scores
  - Very helpful to diagnose issues and improve results!
- First paper results has some large errors
  - Strange artefacts??

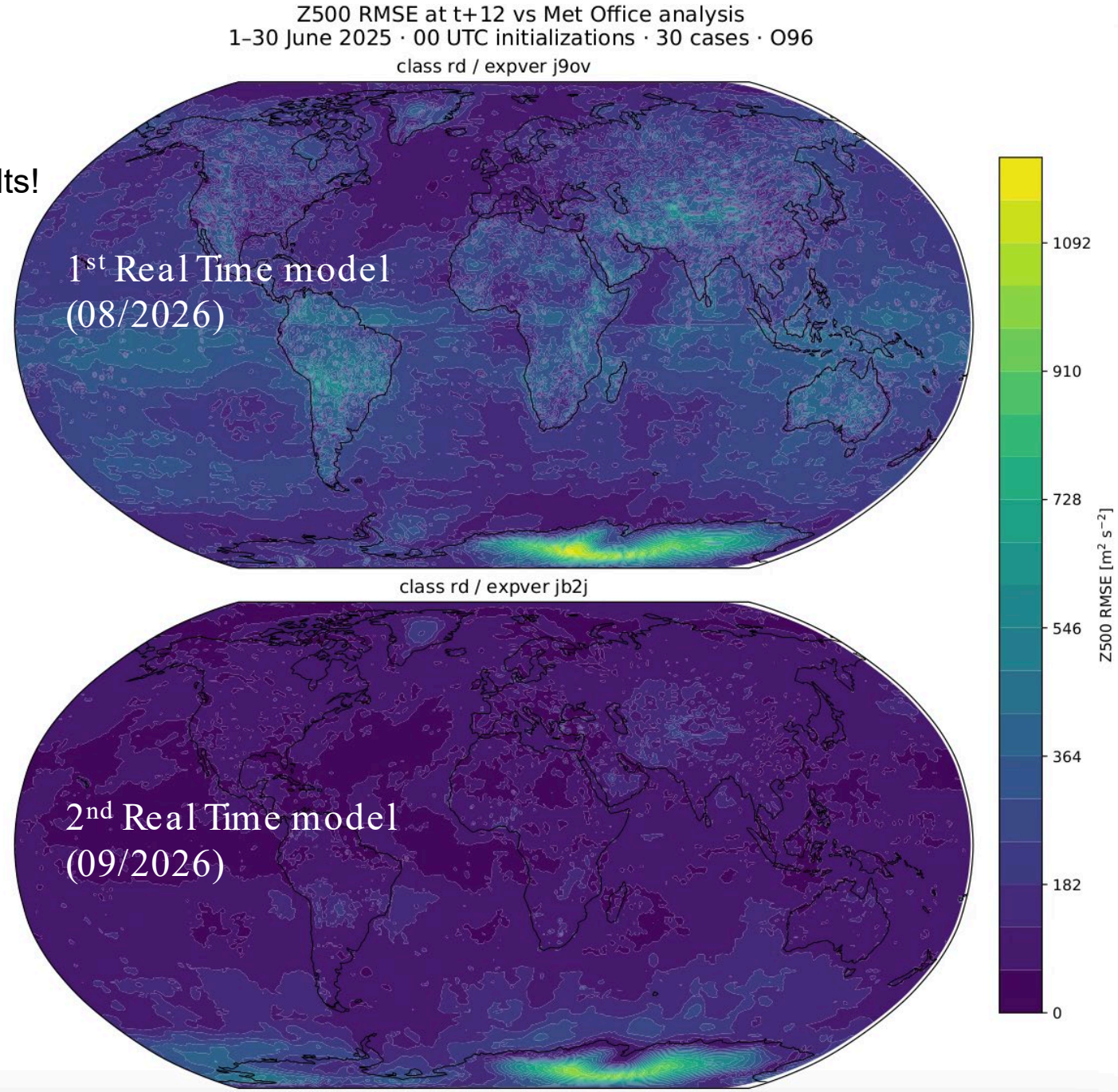
NH Z500
RMSE



Comparisons to Analysis?

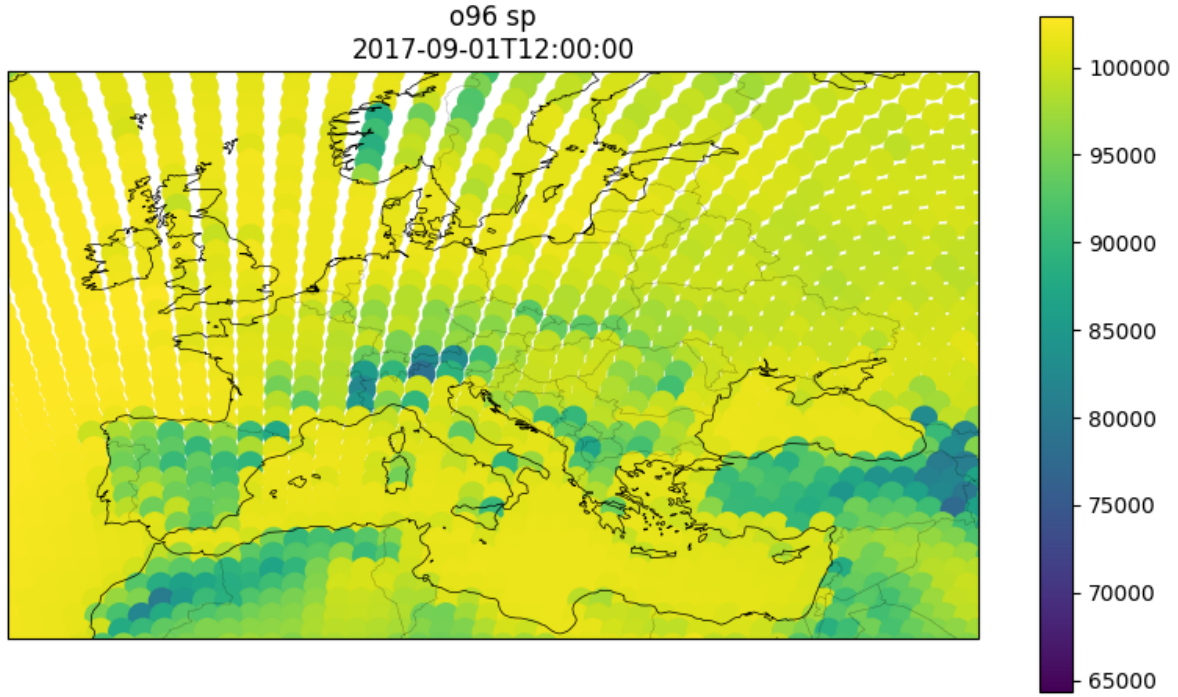
- Work started to understand analysis scores
  - Very helpful to diagnose issues and improve results!
- First paper results has some large errors
  - Strange artefacts??

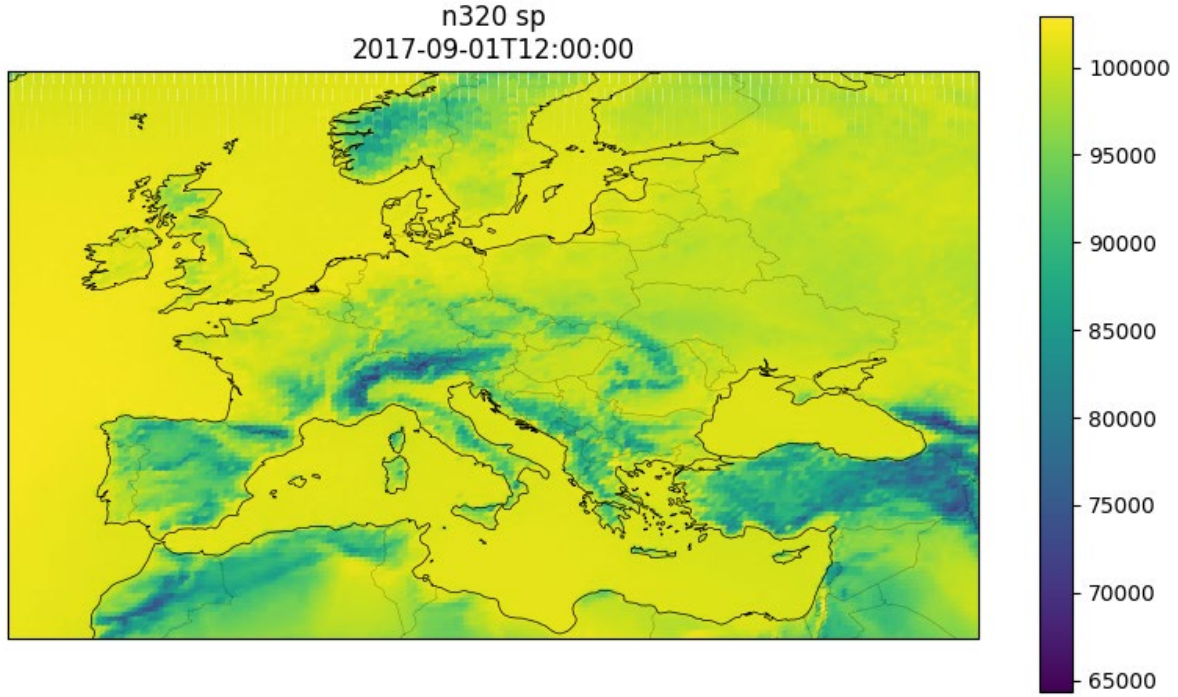
SH Z500
RMSE



Going Higher Resolution?

- Everything so far has been at o96 (~100 km)
  - Quick to train and run! Learning lots!
- Preliminary runs at n320 (~30 km)

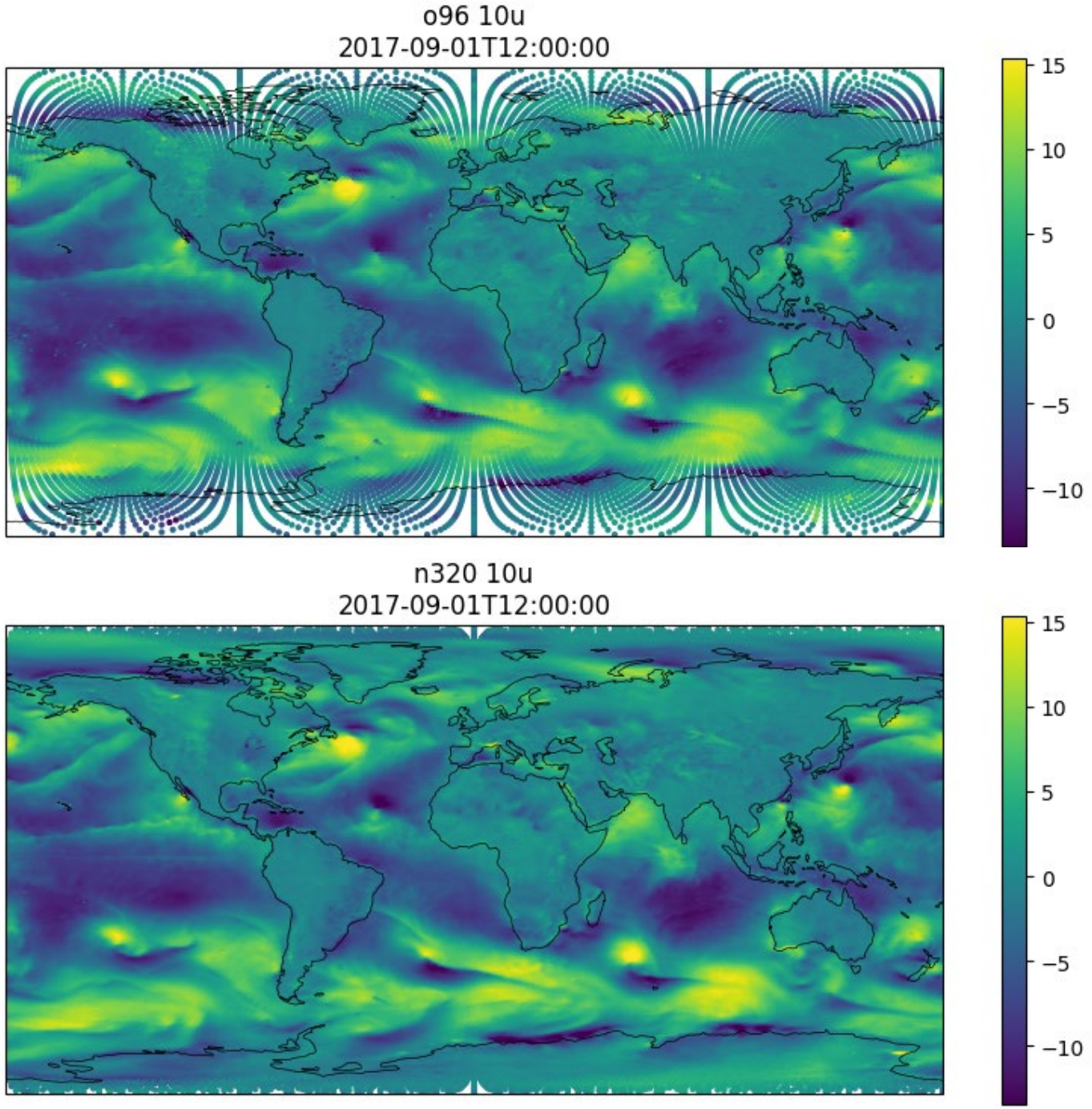




ECMWF

Going Higher Resolution?

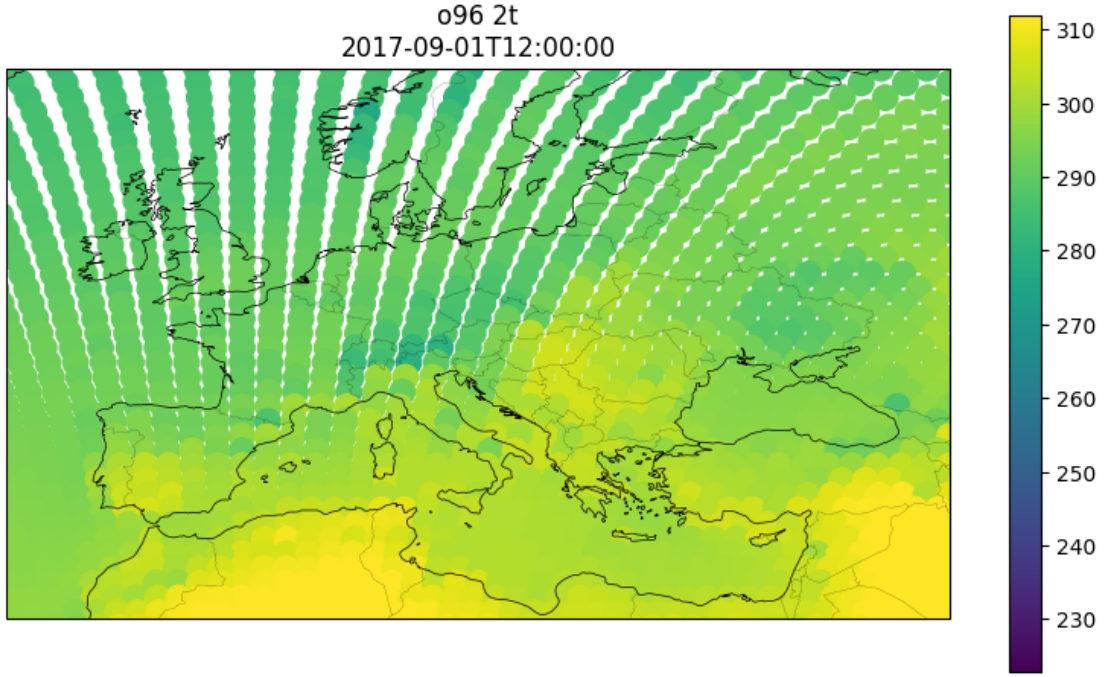
- Everything so far has been at o96 (~100 km)
  - Quick to train and run! Learning lots!
- Preliminary runs at n320 (~30 km)

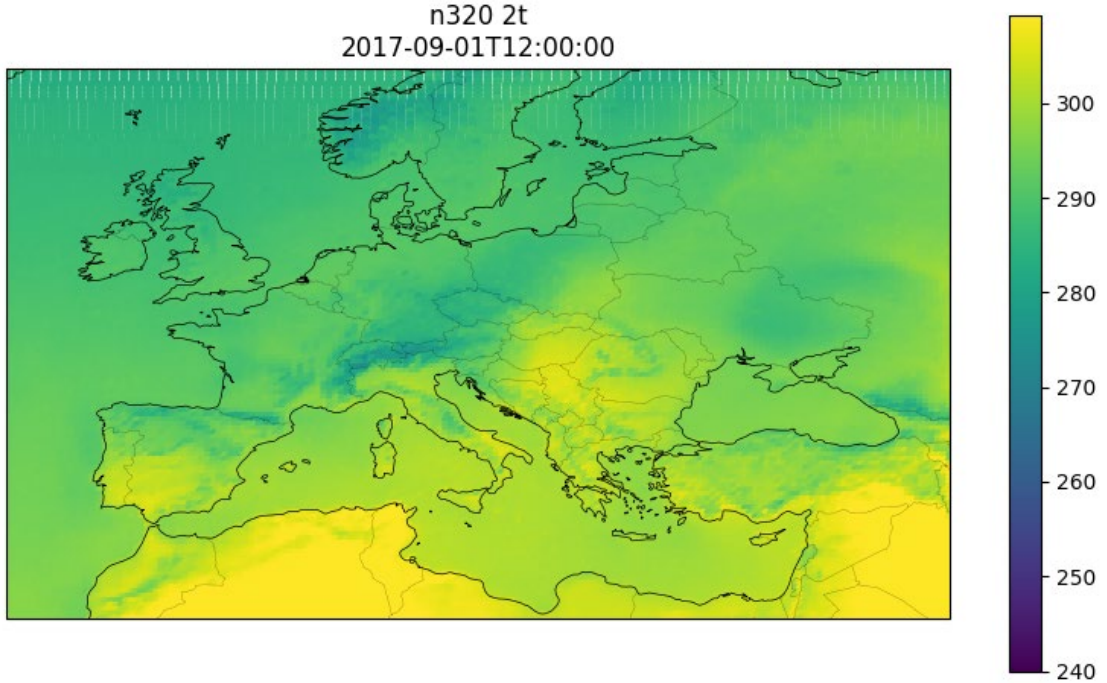


Going Higher Resolution?

- Everything so far has been at o96 (~100 km)
  - Quick to train and run! Learning lots!
- Preliminary runs at n320 (~30 km)

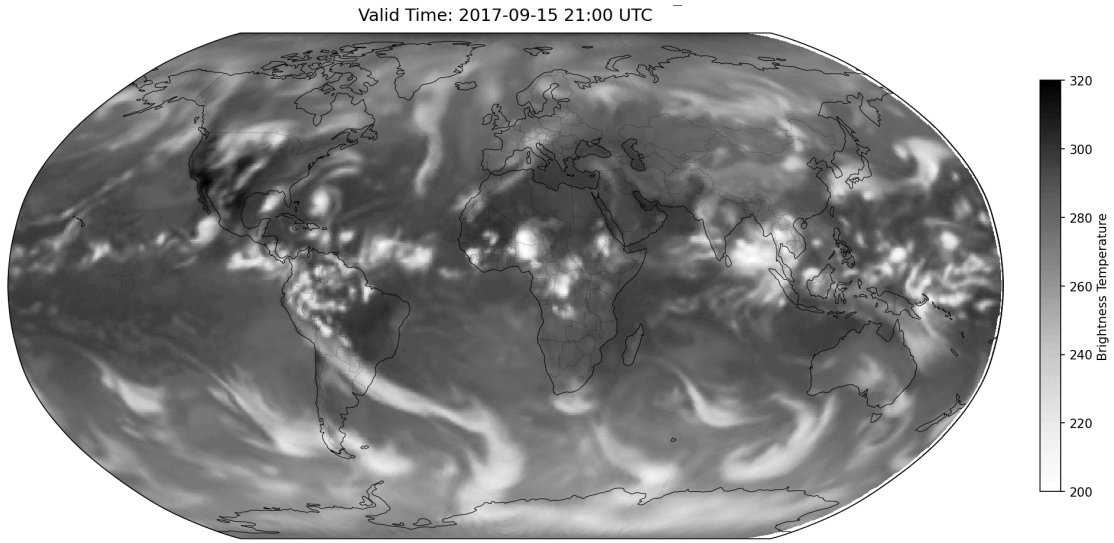
NH T2m
RMSE





Summary

- Pulling DOP work under Anemoi, enable more collaboration with Member States and beyond
    - Initial results there look promising!
- Extending satellite records back in time using harmonized EUMETSAT & NOAA records
- Extended observation dataset has allowed production of prototype “reanalysis”
    - Initial results there look promising! See Peter’s slides next 👀
- Pushing on to *real time* system and bridging gap with previous data records
- Lots more development need!



Outlook

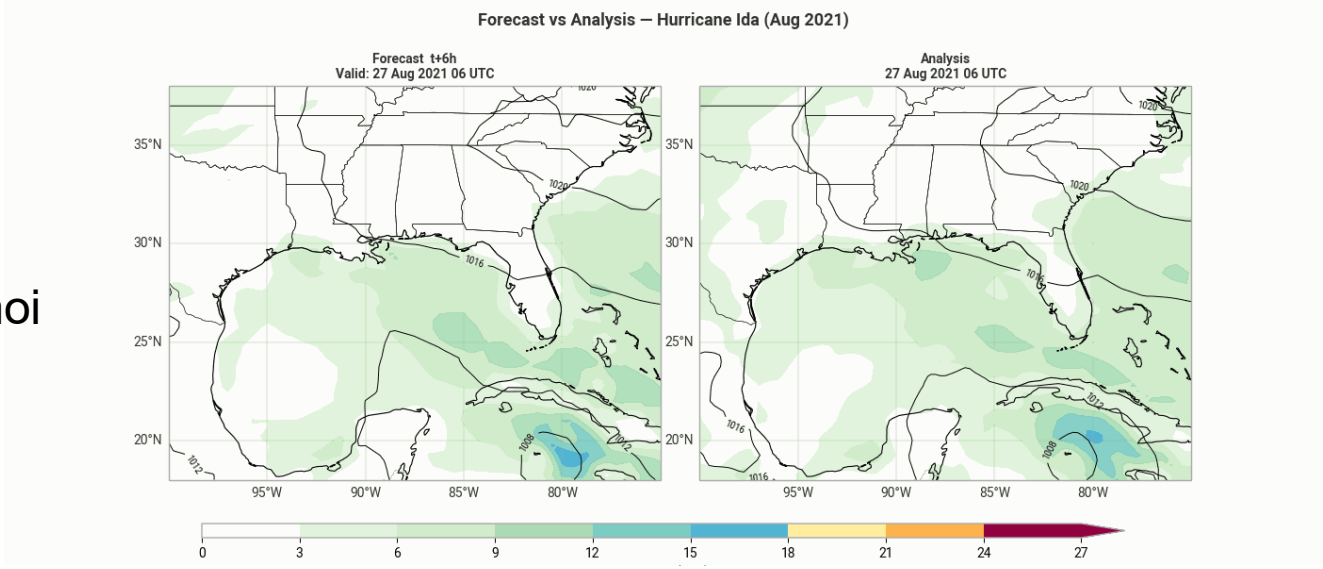
- Increasing resolution (o96 → n320)
  - Usage of dynamic-graphs under Anemoi
- Training with probabilistic loss
  - Current results using MSE-loss, will be subject to smoothing!
- Adding more observations
  - In presented work reduced set of obs was considered
  - Improved Quality Control of current obs too!
- More model development!
- More analysis of results!
- ML Observation machinery developed under Anemoi applicable to many different problems!

License: CC BY-NC-ND 4.0
arXiv:2606.19093v1 [physics.ao-ph] 17 Jun 2026

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)



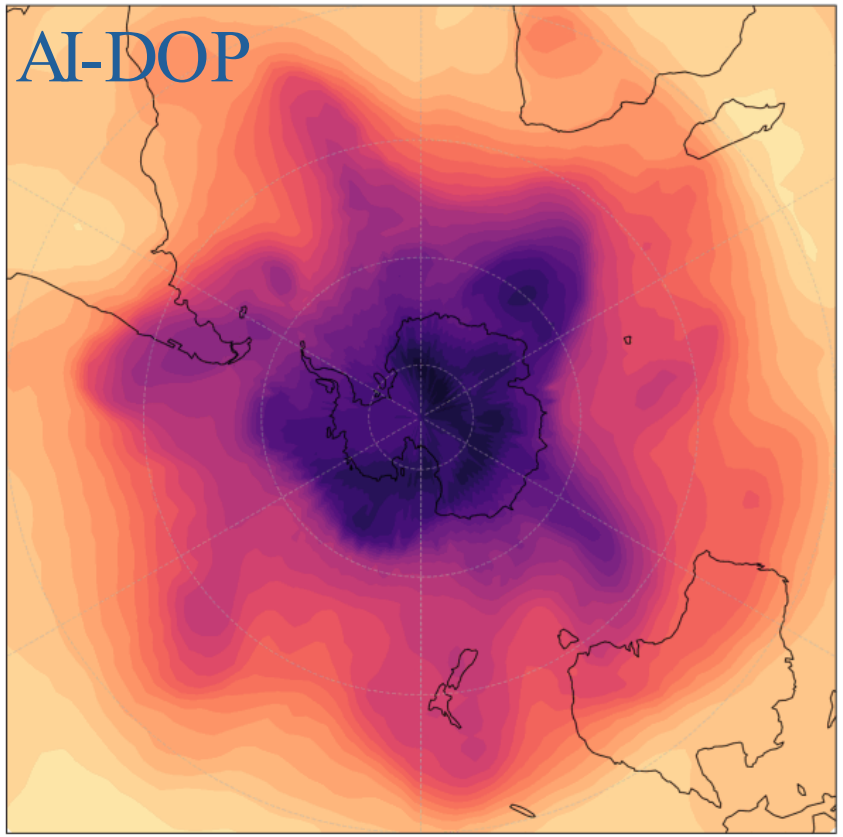
Prospects for using DOP for reanalysis



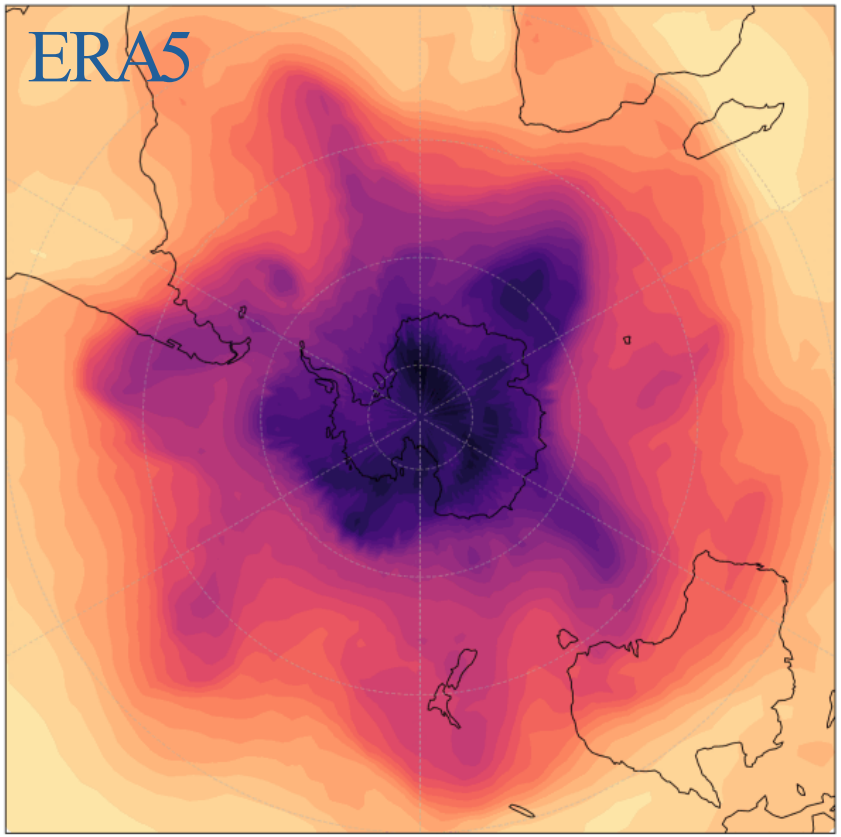
500hPa Temperature

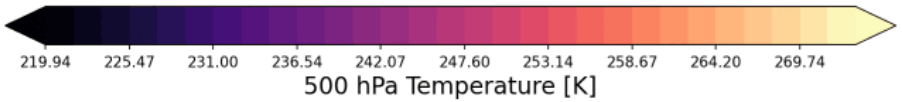
SH Extratropical T500 Evolution
Valid Time: 2017-09-03 00:00 UTC

AIFS-DOP



ERA5





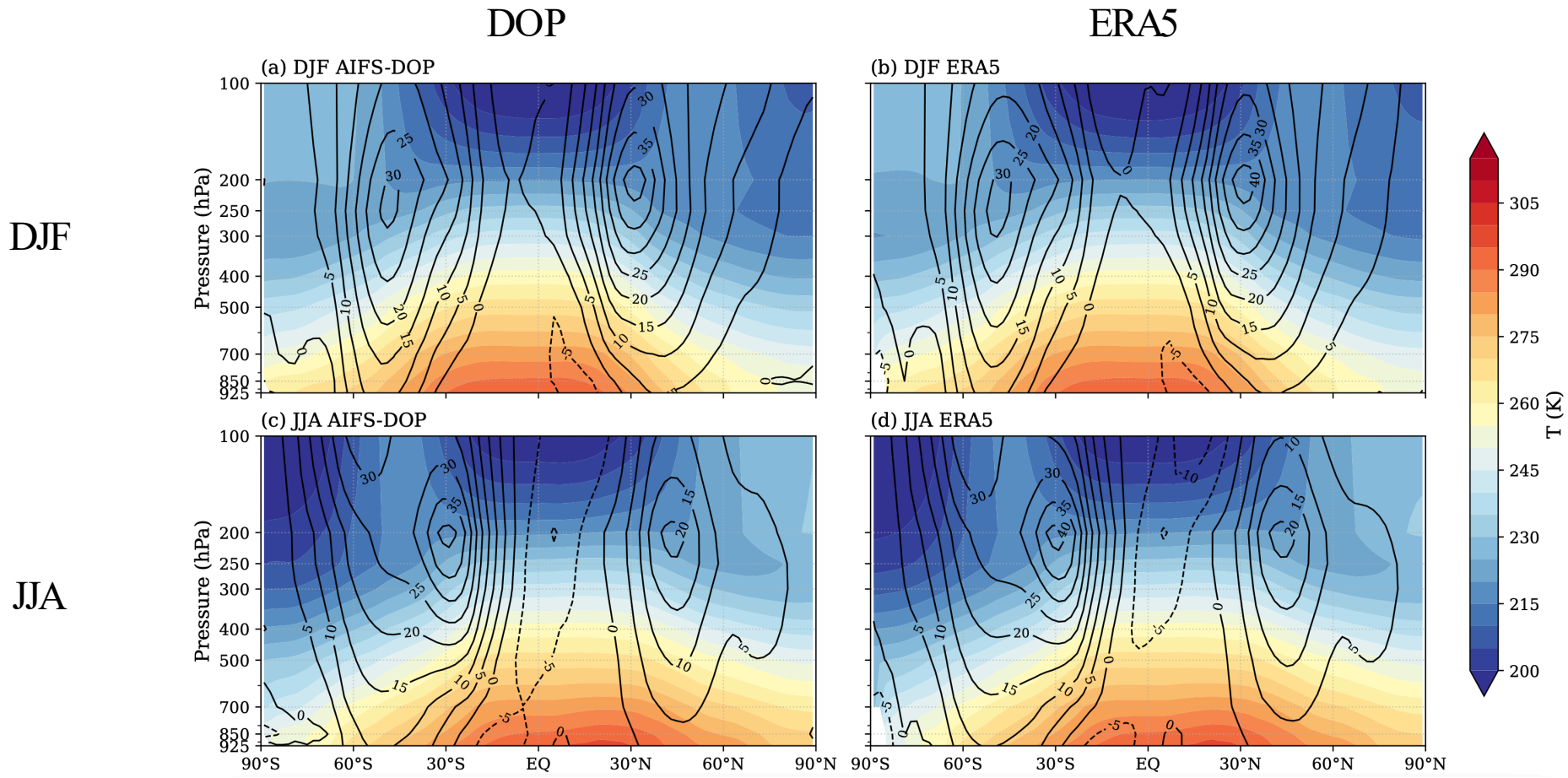
* Sequence of analyses

Could Direct Observation Prediction be used to generate a reanalysis?

- Can DOP produce dense, physically coherent fields from sparse observations?
  - Data assimilation systems use a physical model and covariance statistics to produce dense, physically coherent fields from sparse observations
  - Can the representations of the physical Earth System learned during training perform a similar role?
- Overlapping training and test period, is this a problem?
- Prototype presented here: state estimates from t+6h DOP forecasts, every 6 hours at O96 resolution
- Very fast to produce: inference is cheap + embarrassingly parallel

Takes < 1 day to produce a 40-year reanalysis!

Zonal mean state

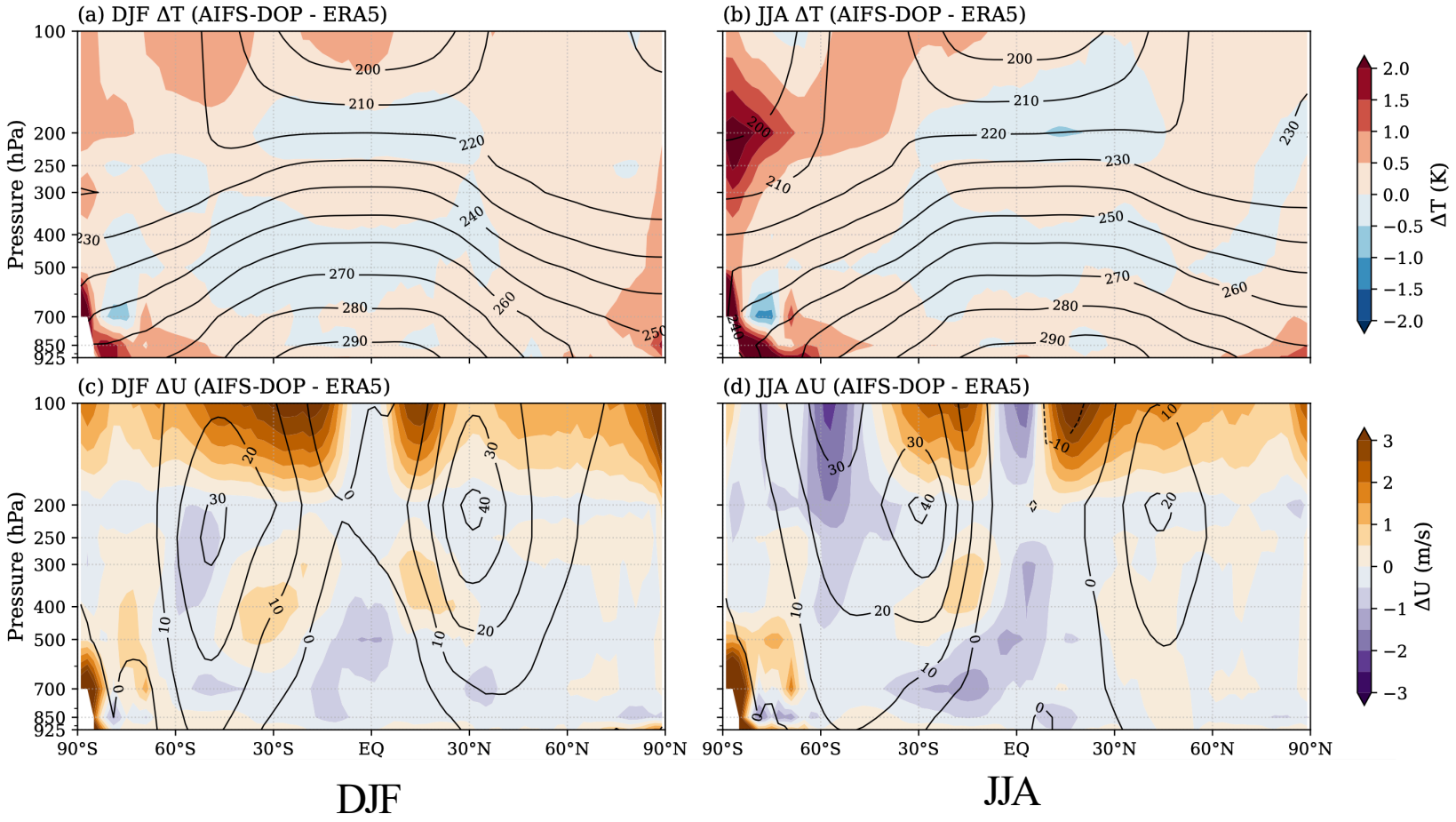


Zonal mean u-wind contours

Mean difference v ERA5

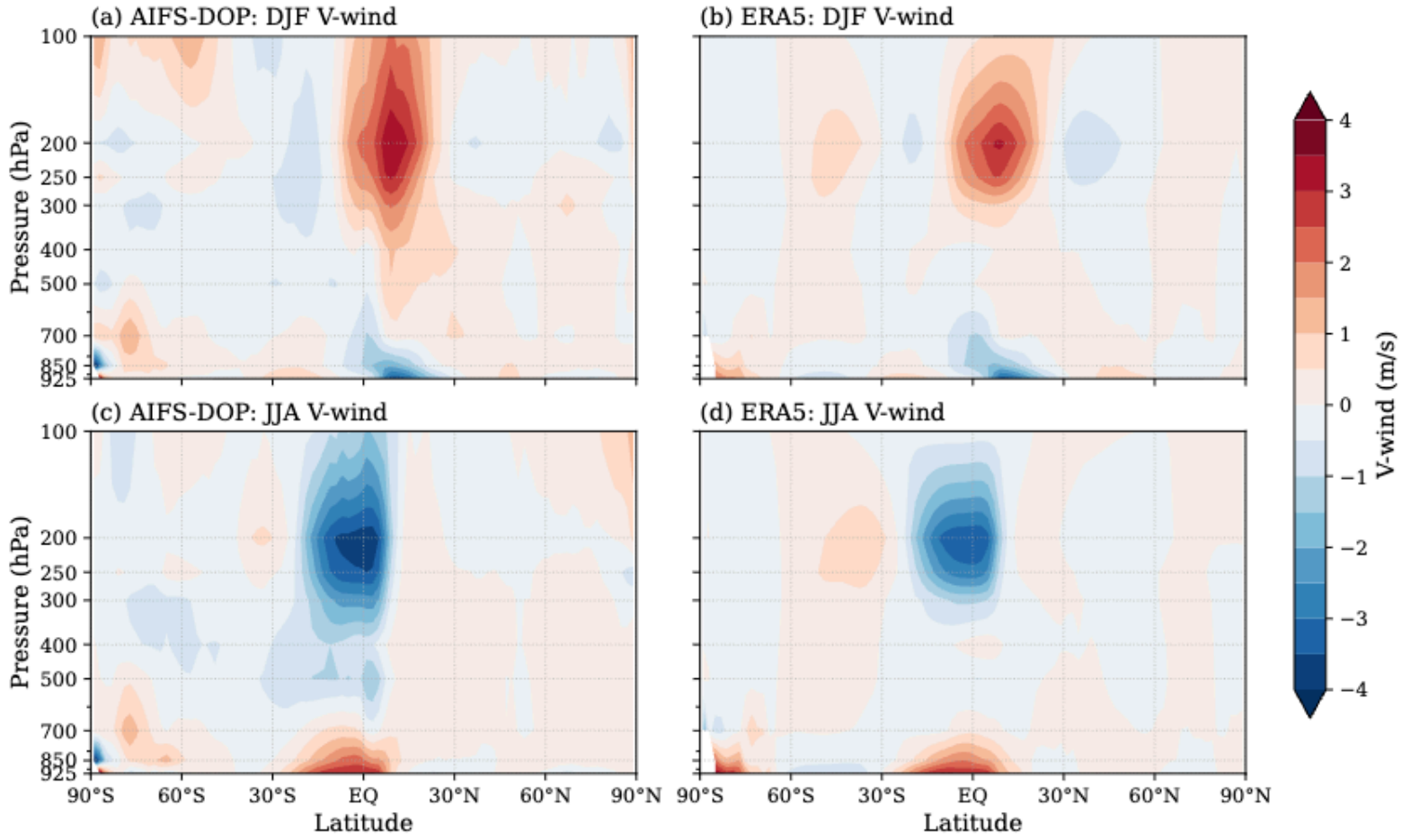
Temp diff

Wind diff

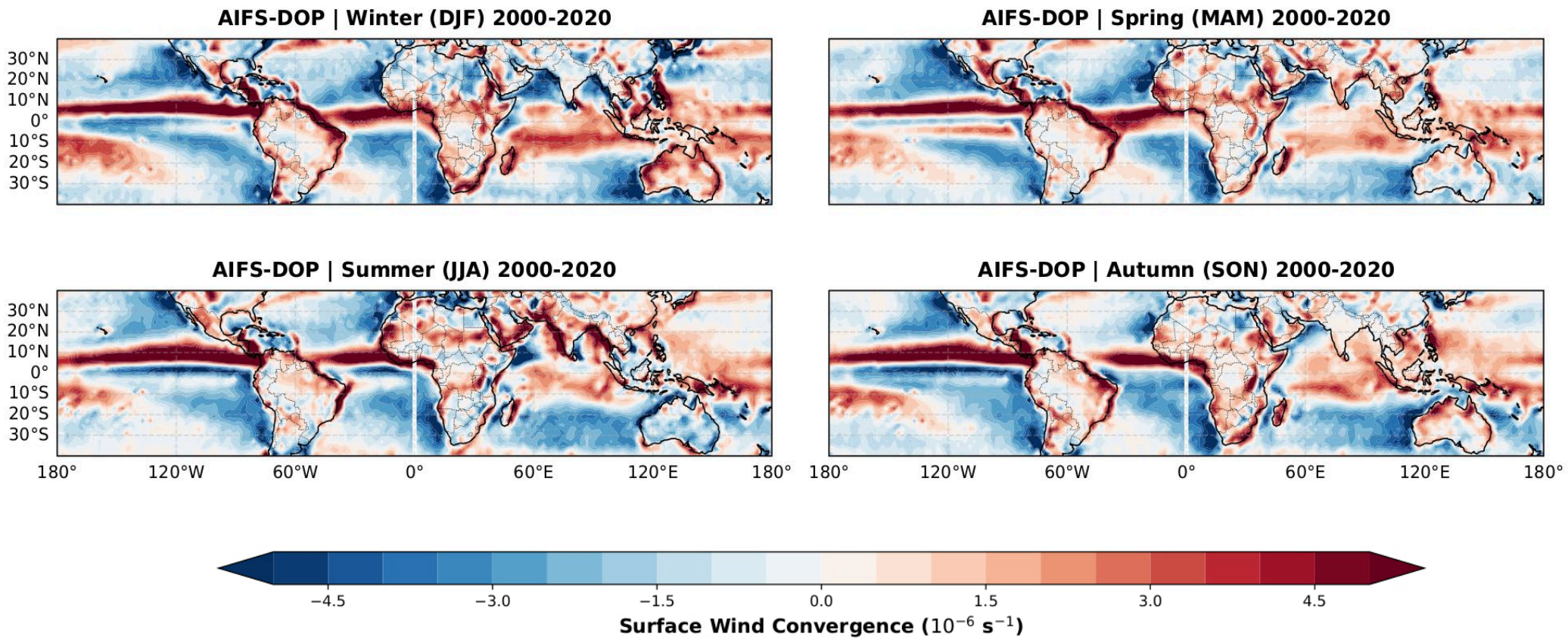


Zonal mean u-wind contours

Issues with the meridional wind

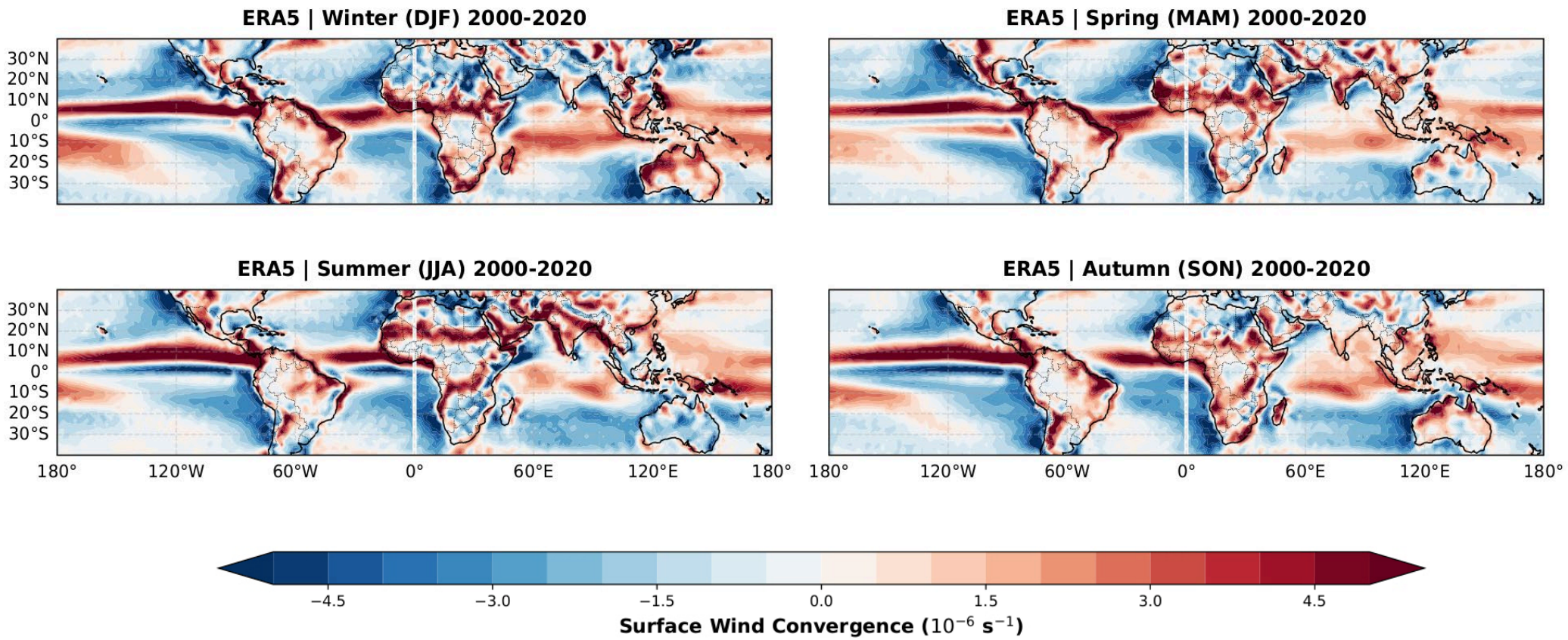


ITCZ: DOP



* No scatterometer data used in DOP yet

ITCZ: ERA5



* No scatterometer data used in DOP yet

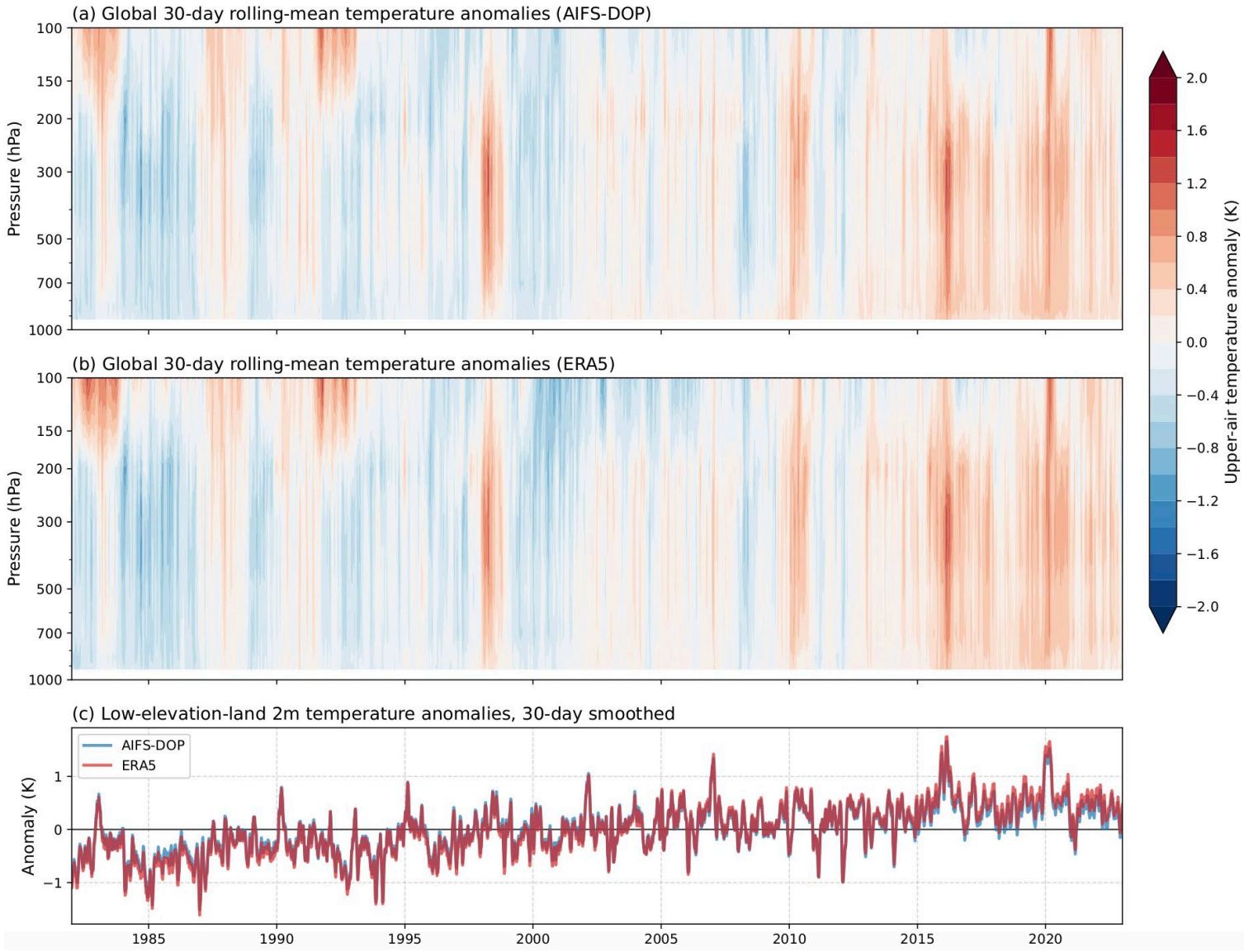
Temperature anomalies

DOP

Upper air

ERA5

2m temp
low-elevation land



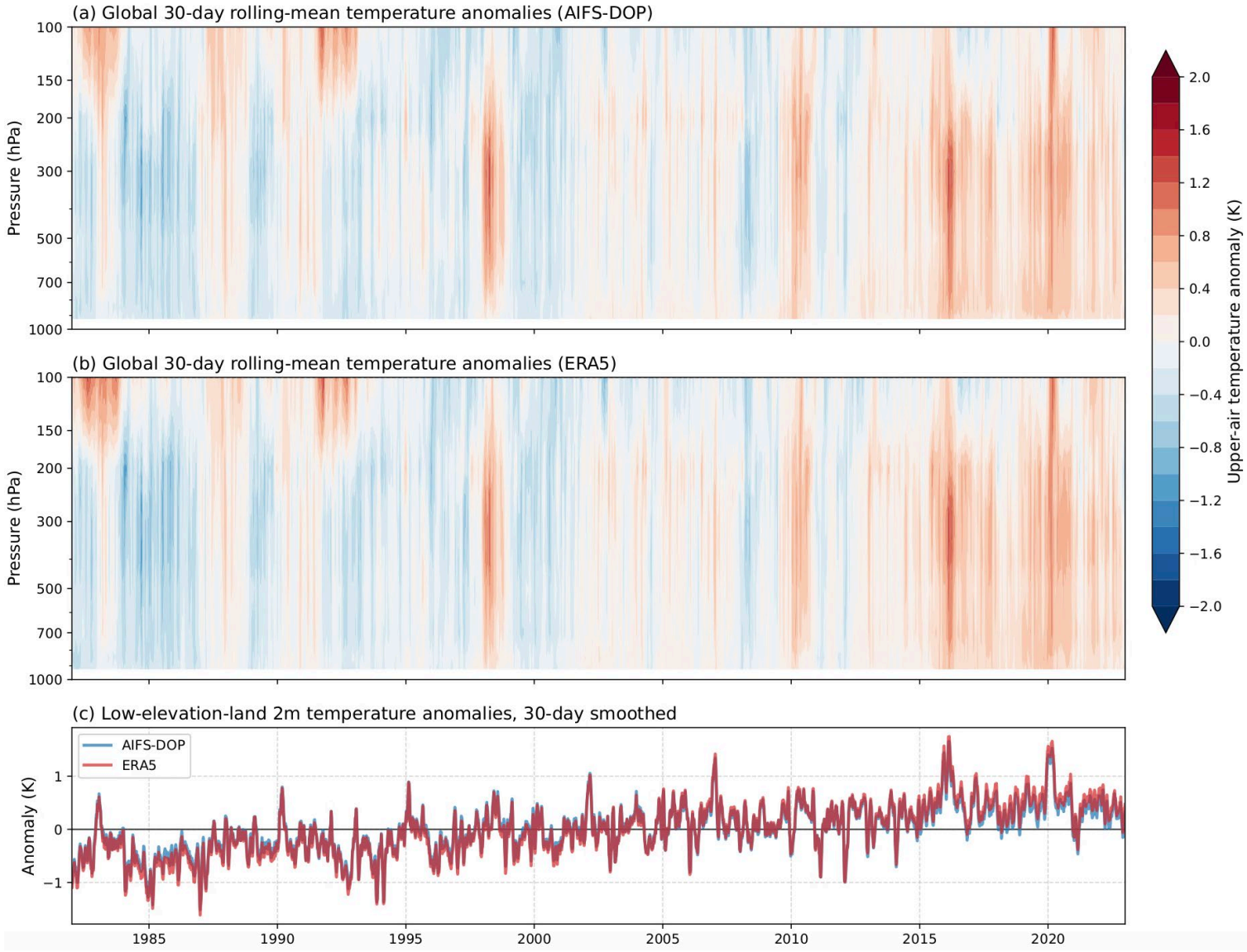
Temperature anomalies

DOP

Upper air

ERA5.1

2m temp
low-elevation land

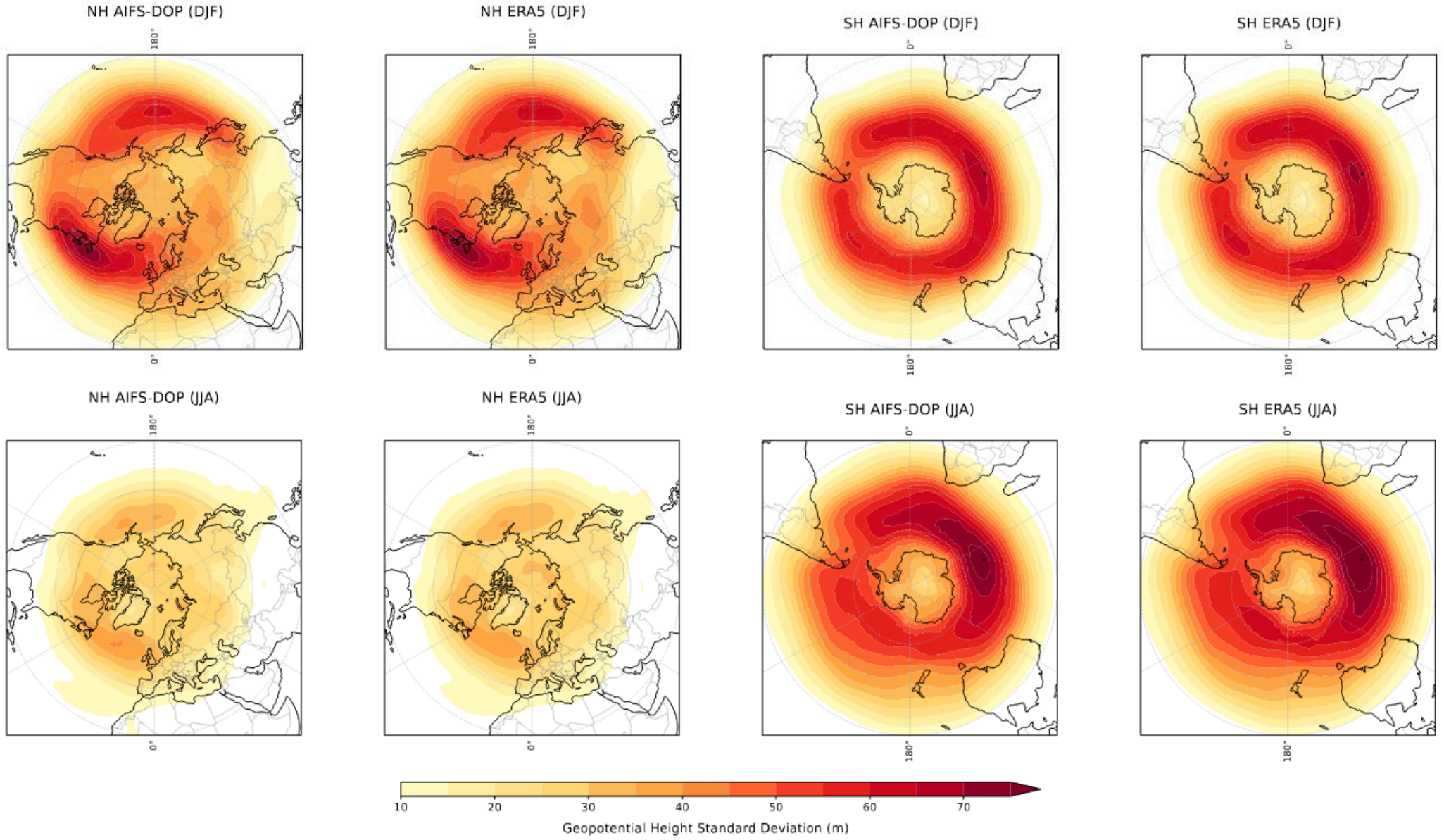


Variability - storm tracks

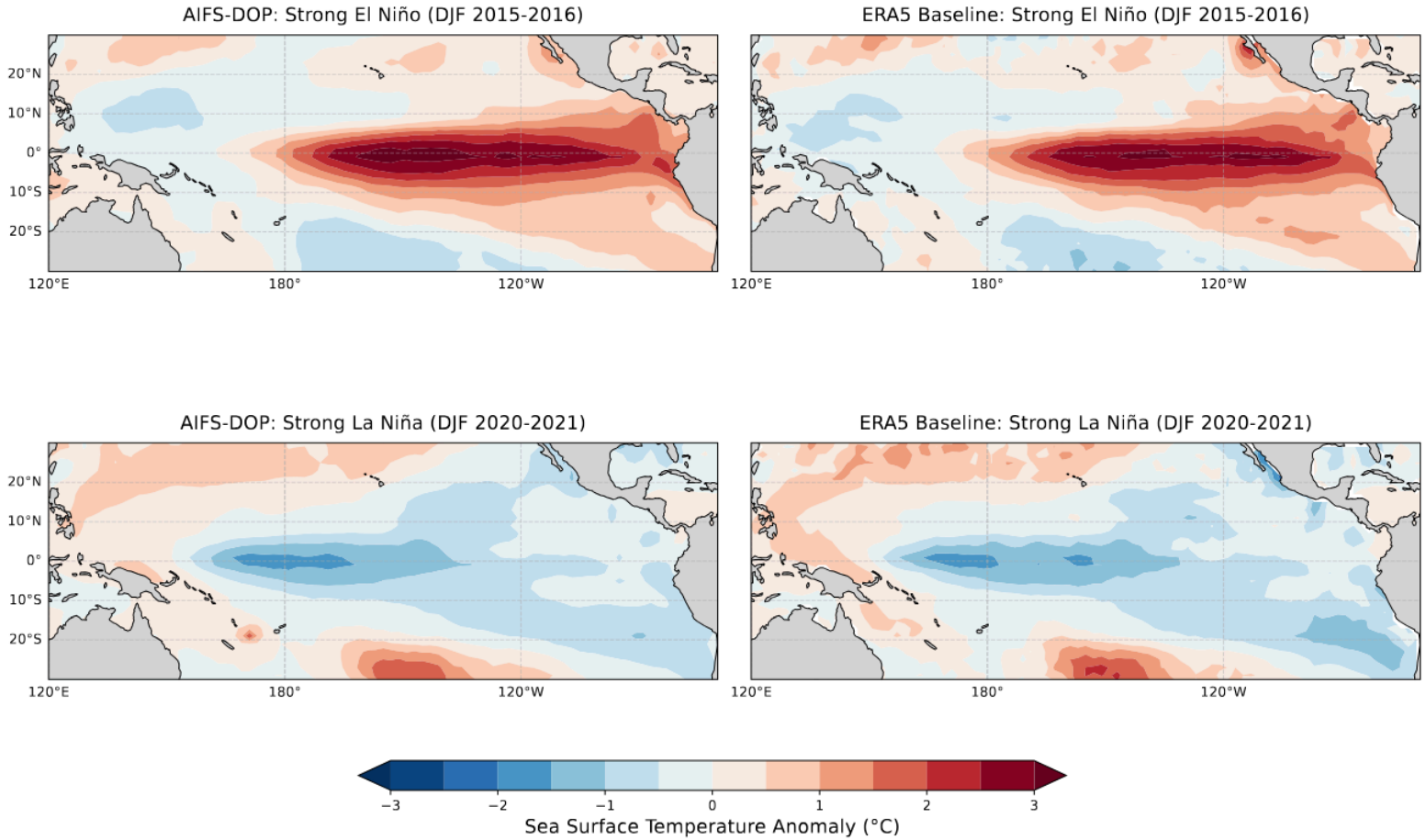
Standard deviation of band-pass filtered z500

DJF

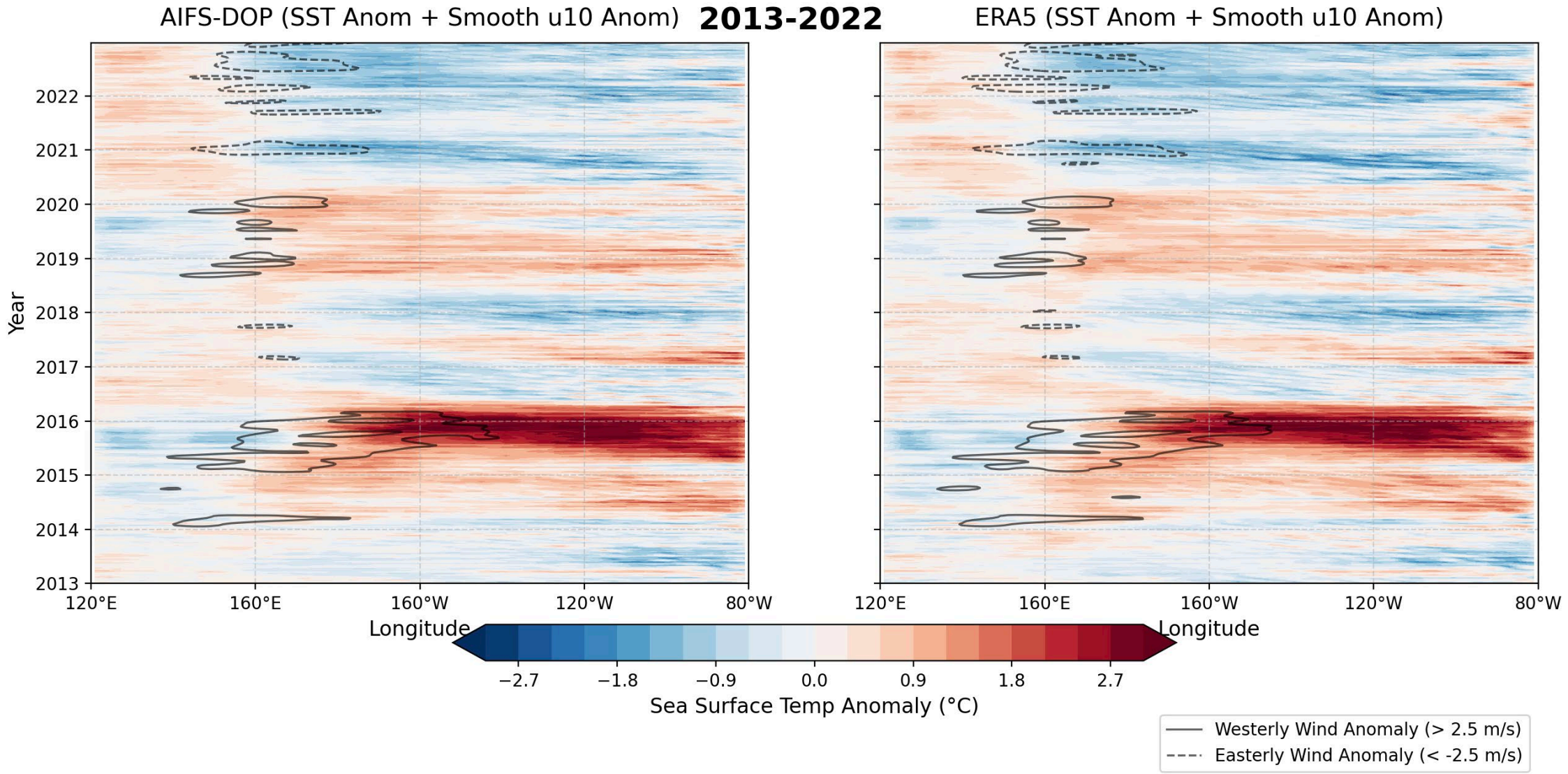
JJA



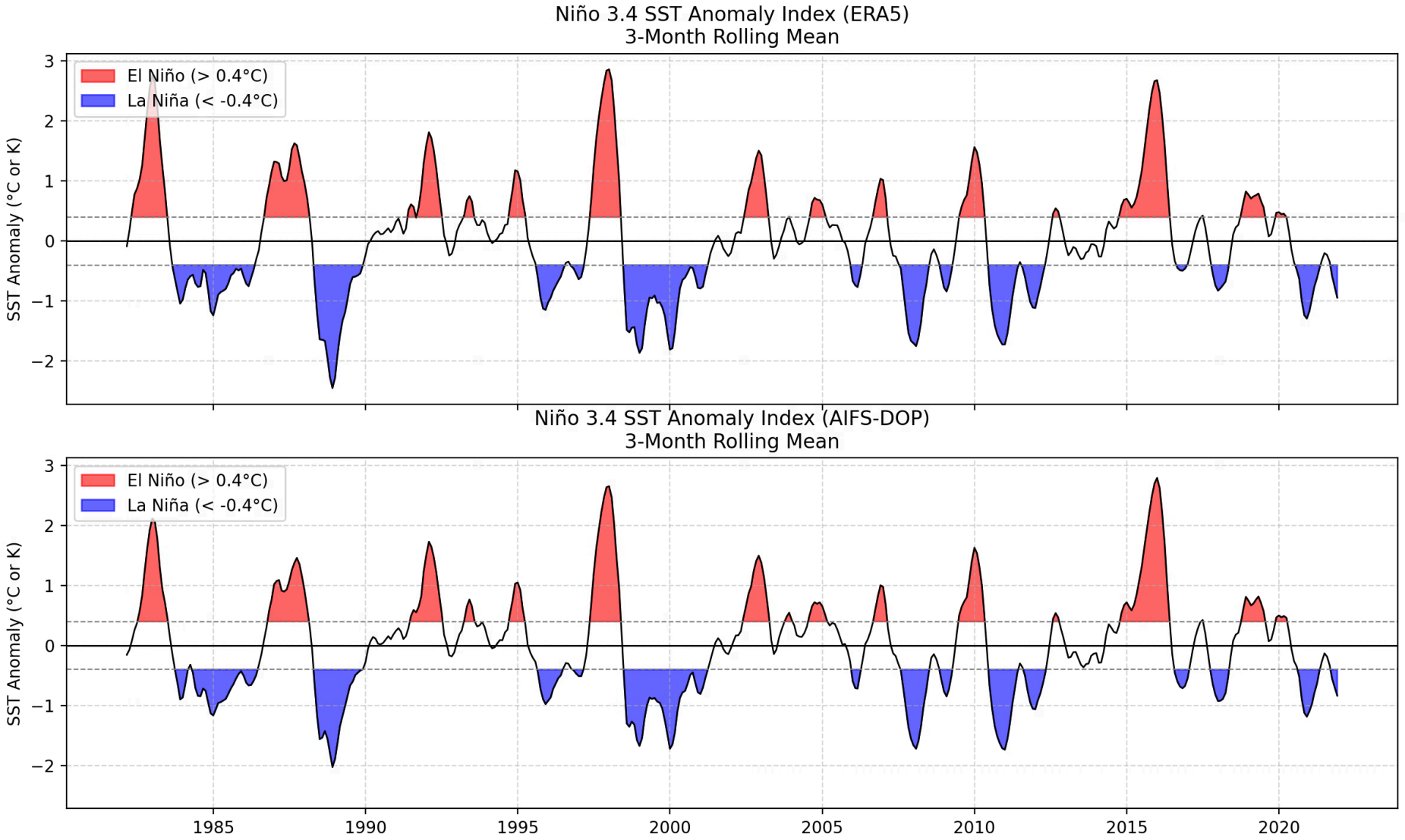
ENSO anomalies



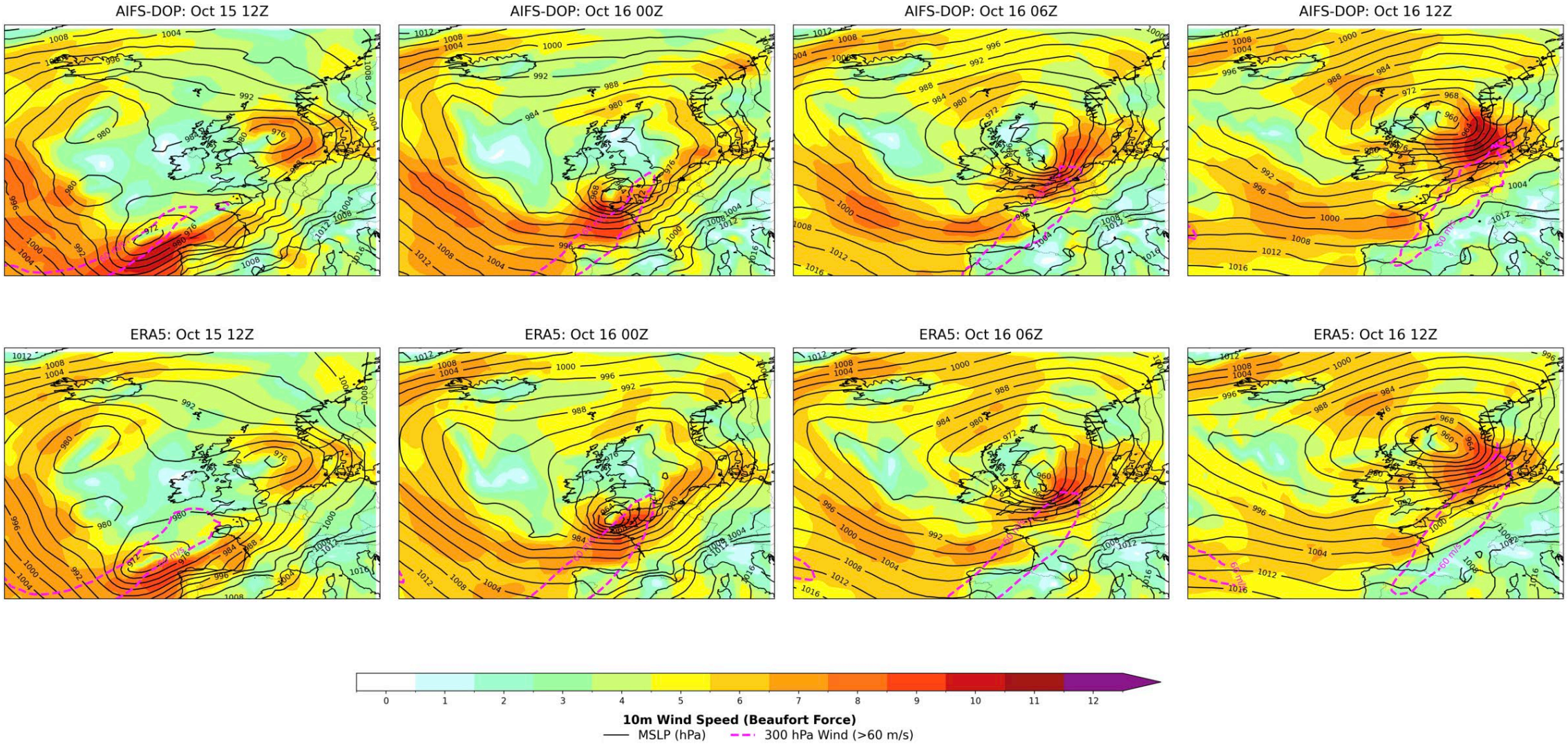
Nino 3.4 SST anomalies



Nino 3.4 SST anomalies



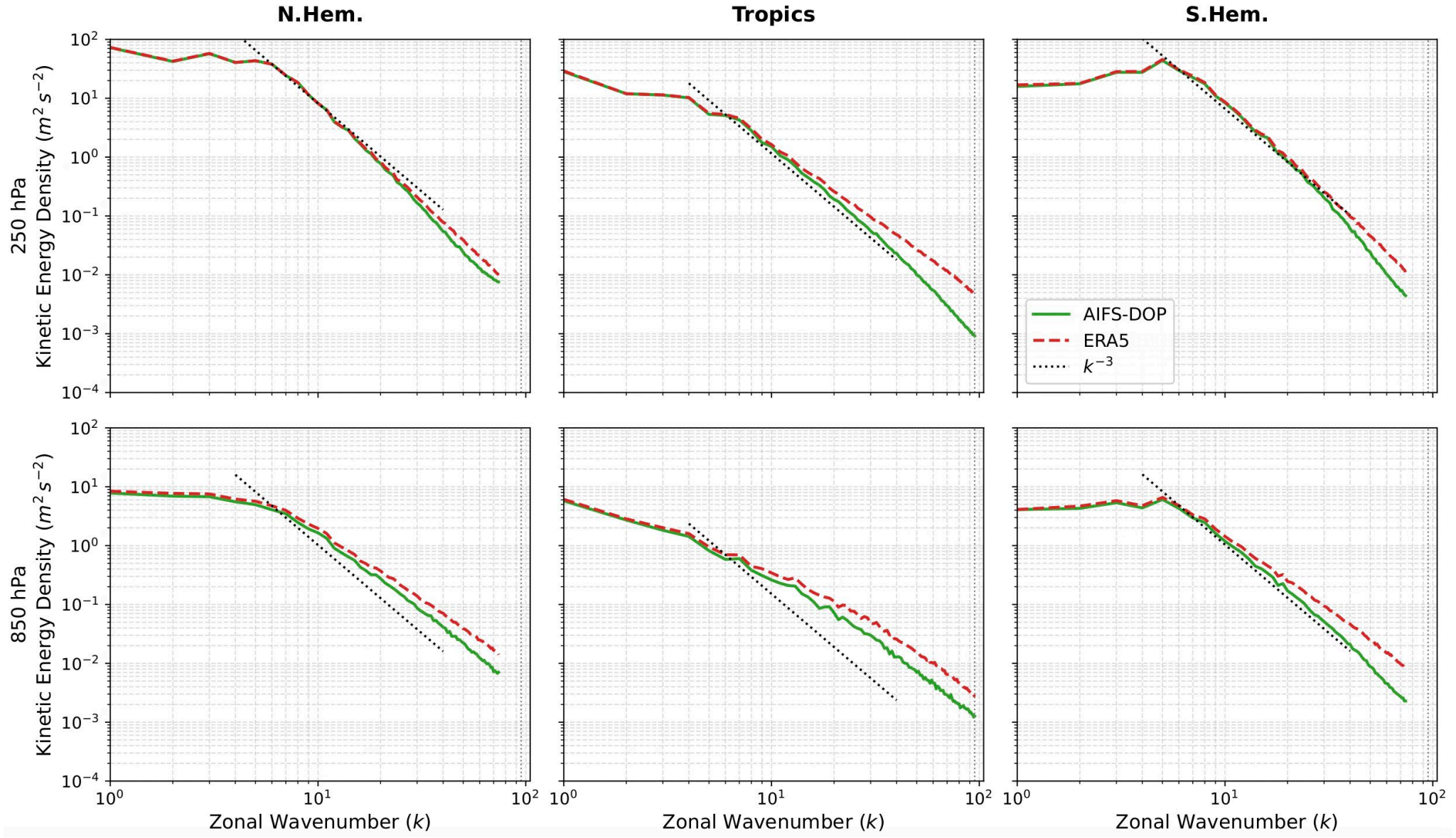
Case study: 1987 extreme wind storm



Kinetic Energy Spectra

KE 250hPa

KE 500hPa

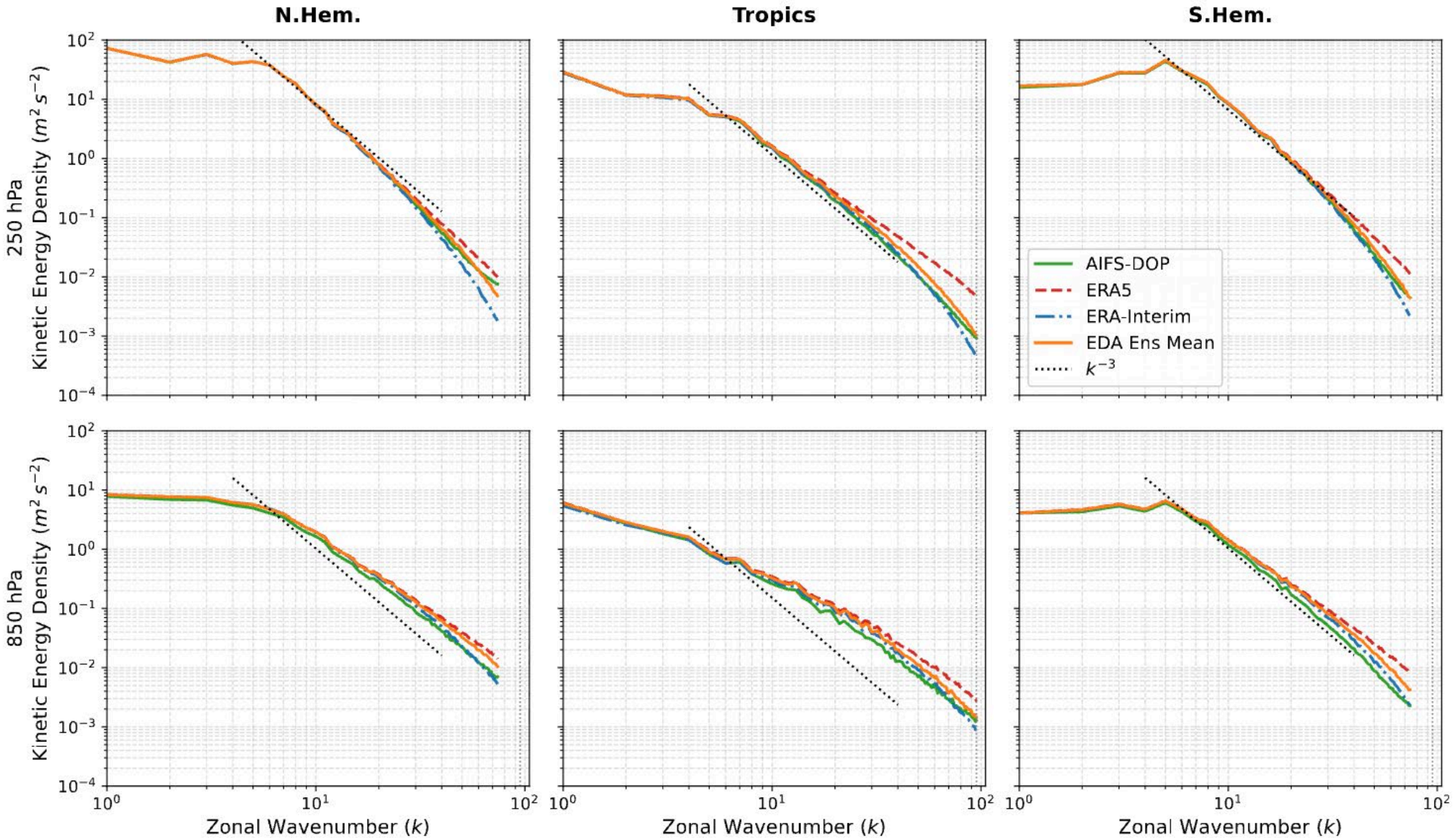


More smoothing in tropics

Kinetic Energy Spectra

KE 250hPa

KE 500hPa

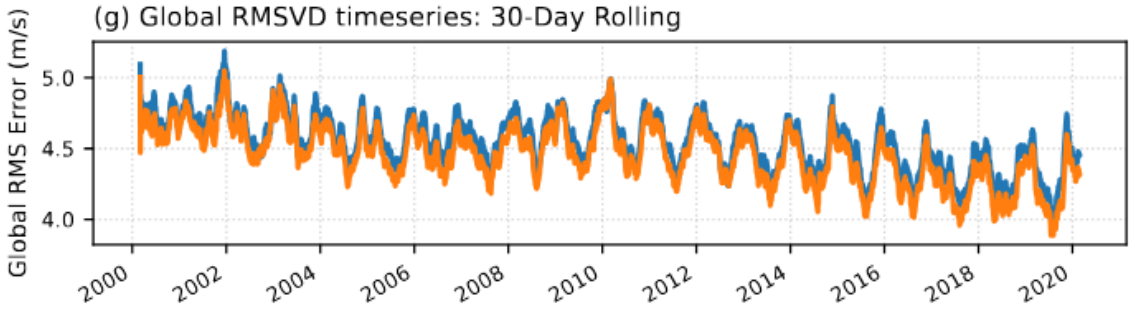


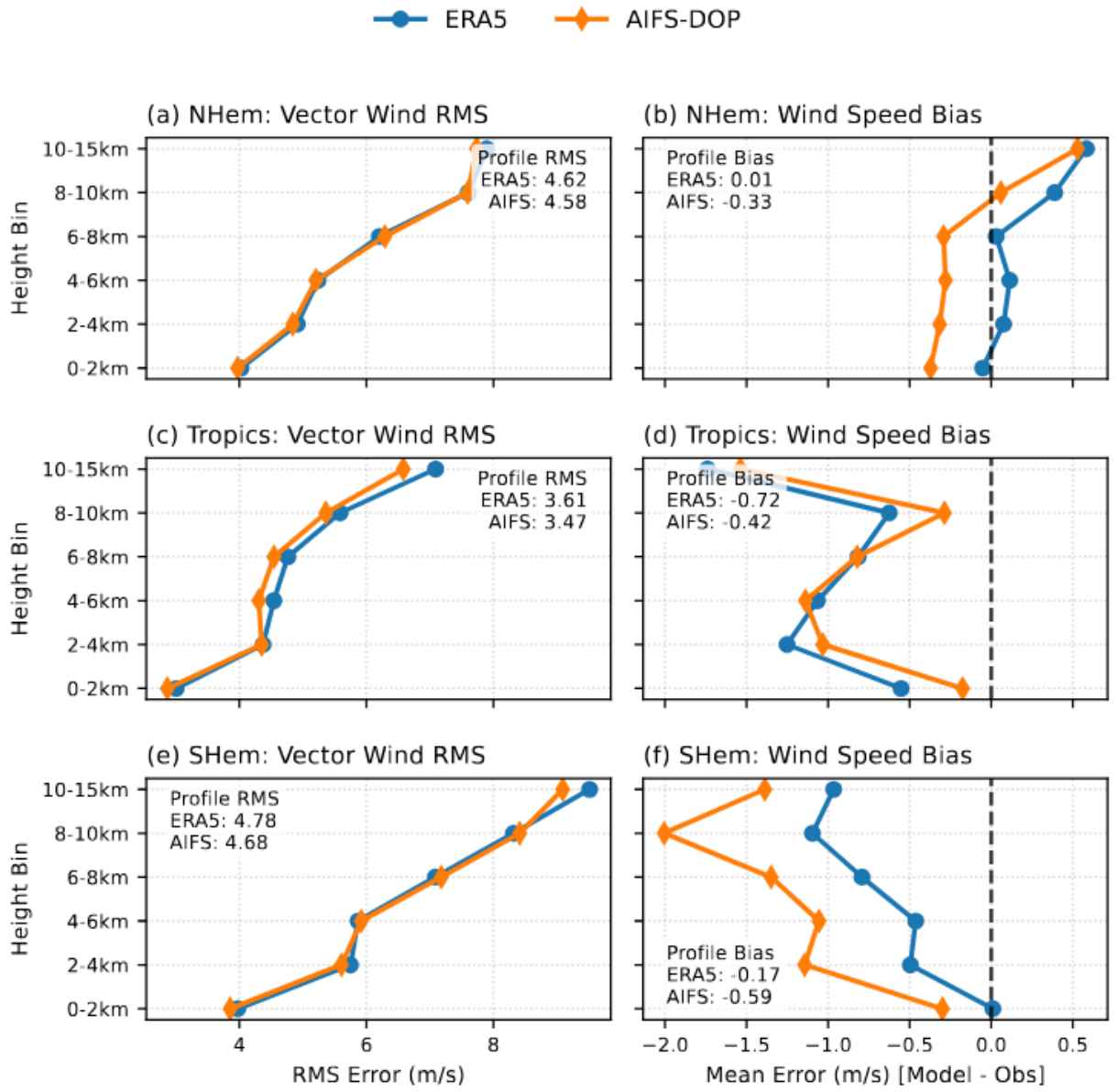
More smoothing in tropics

Evaluation against held-out observations: MISR AMVs

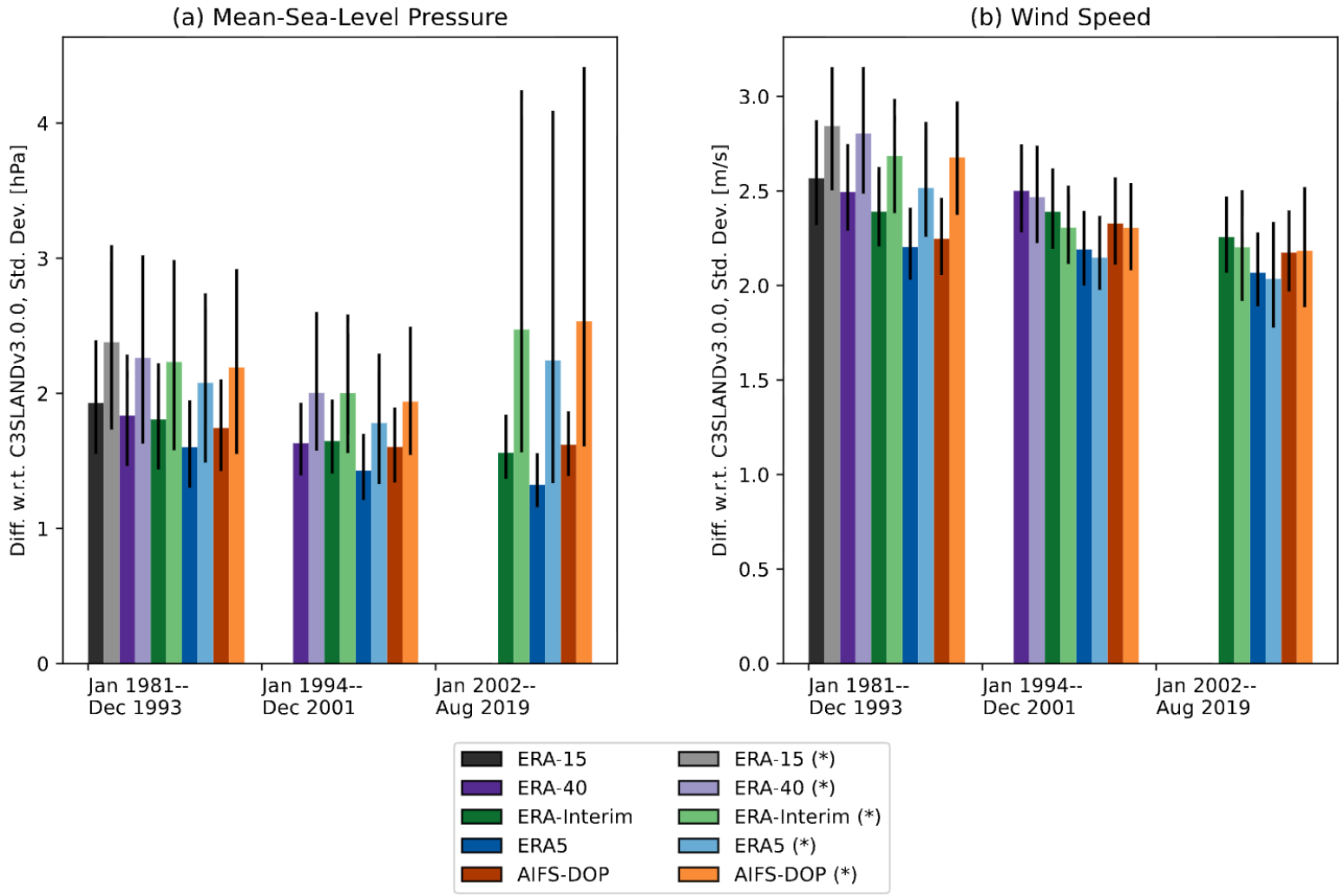
Has the model learned to generalize spatially or is it just memorizing the training dataset?

MISR AMVs are independent observations that aren't used by either ERA5 or DOP





Evaluation against held-out observations: surface stations



Paul Poli

Can a **data-driven** model handle **data-sparse** settings?

- Pre-satellite era

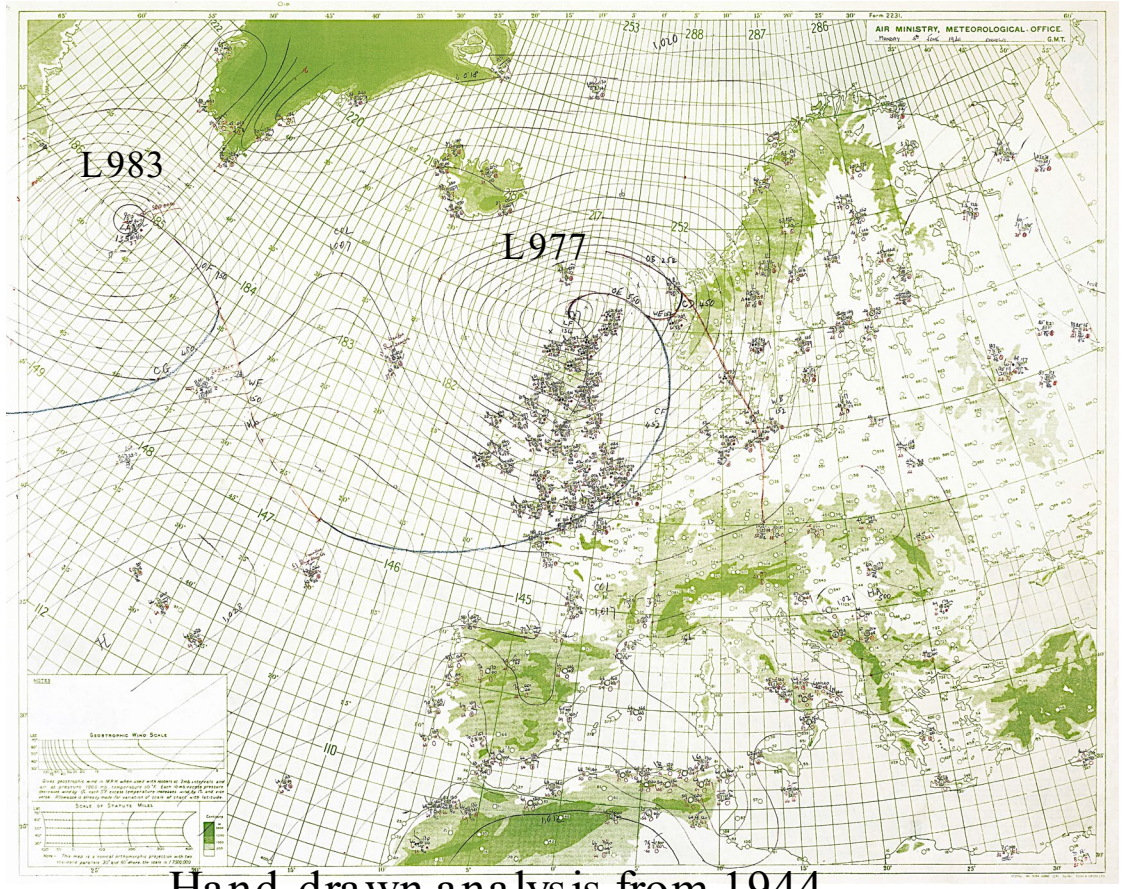
- Ocean

Paul Poli, Patrick Laloyaux

D-Day case study

Fine-tune model to only use conventional observations

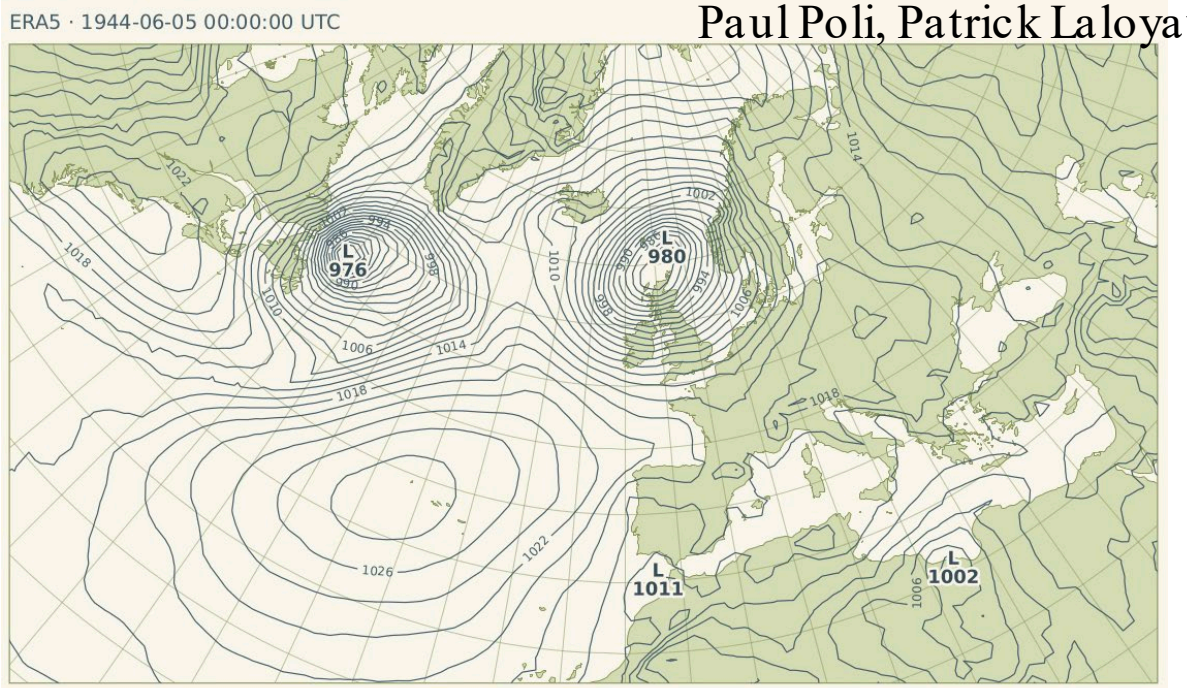
June 5th, 1944 00 UTC



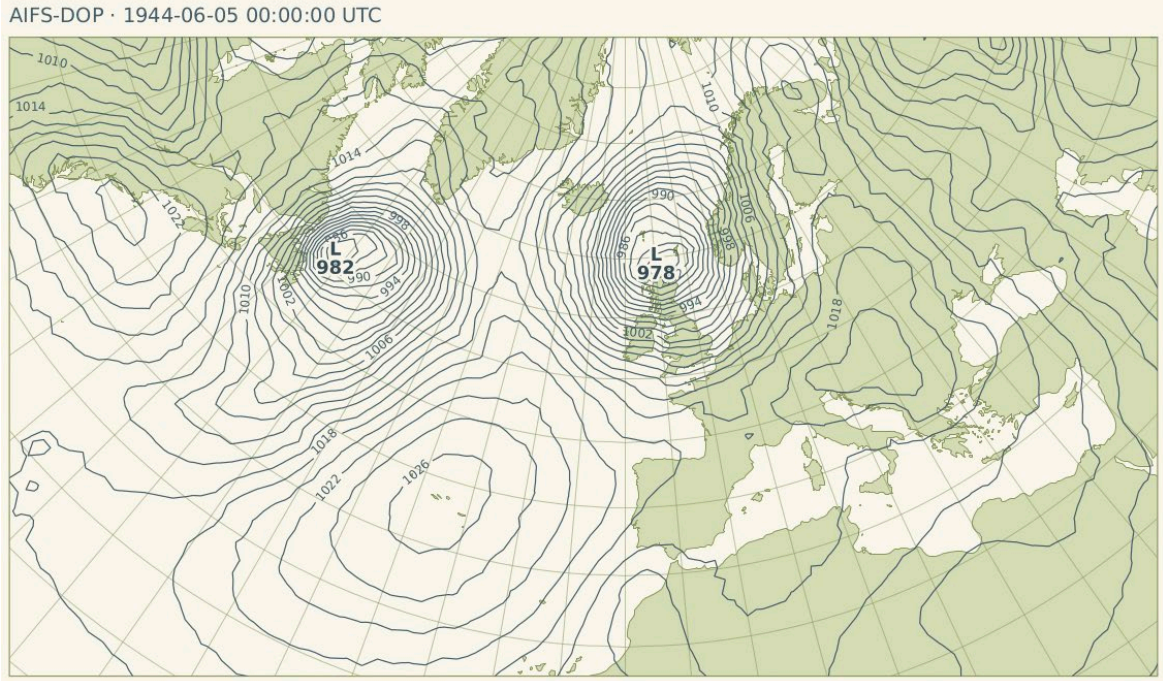
Hand-drawn analysis from 1944

Information provided by the National Meteorological Library and Archive – Met Office, UK
© Crown Copyright 1944.

ERA5



AIFS-DOP



ECMWF

Paul Poli, Patrick Laloyaux

D-Day case study

June 6th, 1944 12 UTC

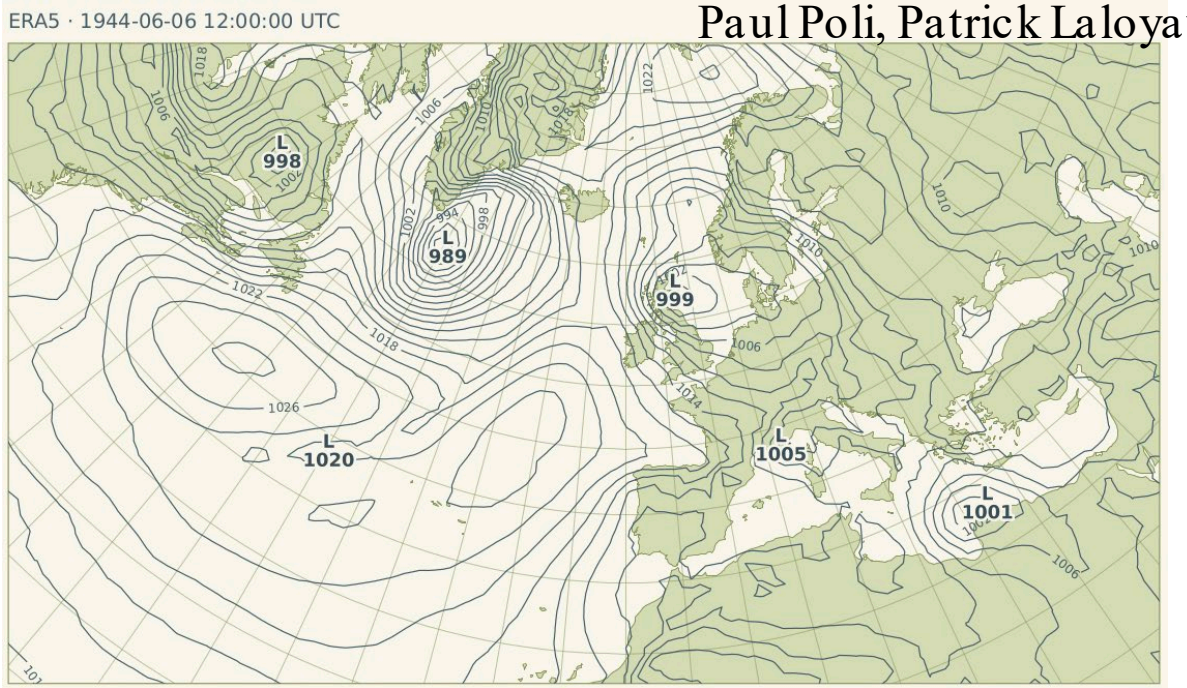
L979

L995

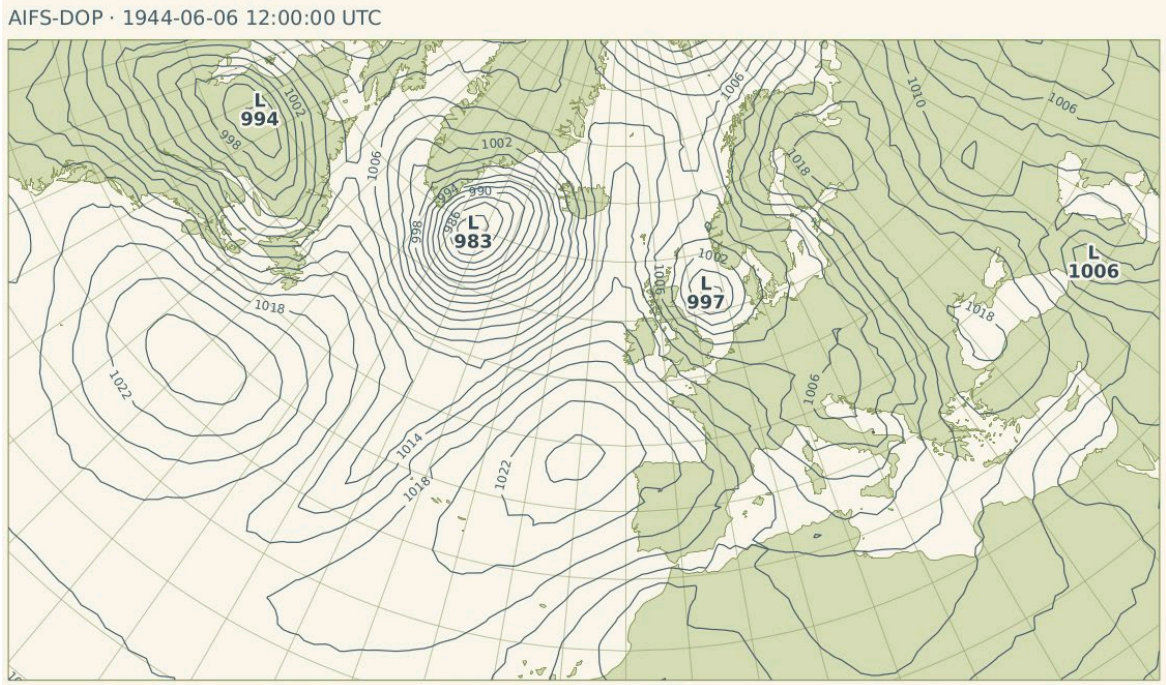
Hand-drawn analysis from 1944

Information provided by the National Meteorological Library and Archive – Met Office, UK
© Crown Copyright 1944.

ERA5



AIFS-DOP



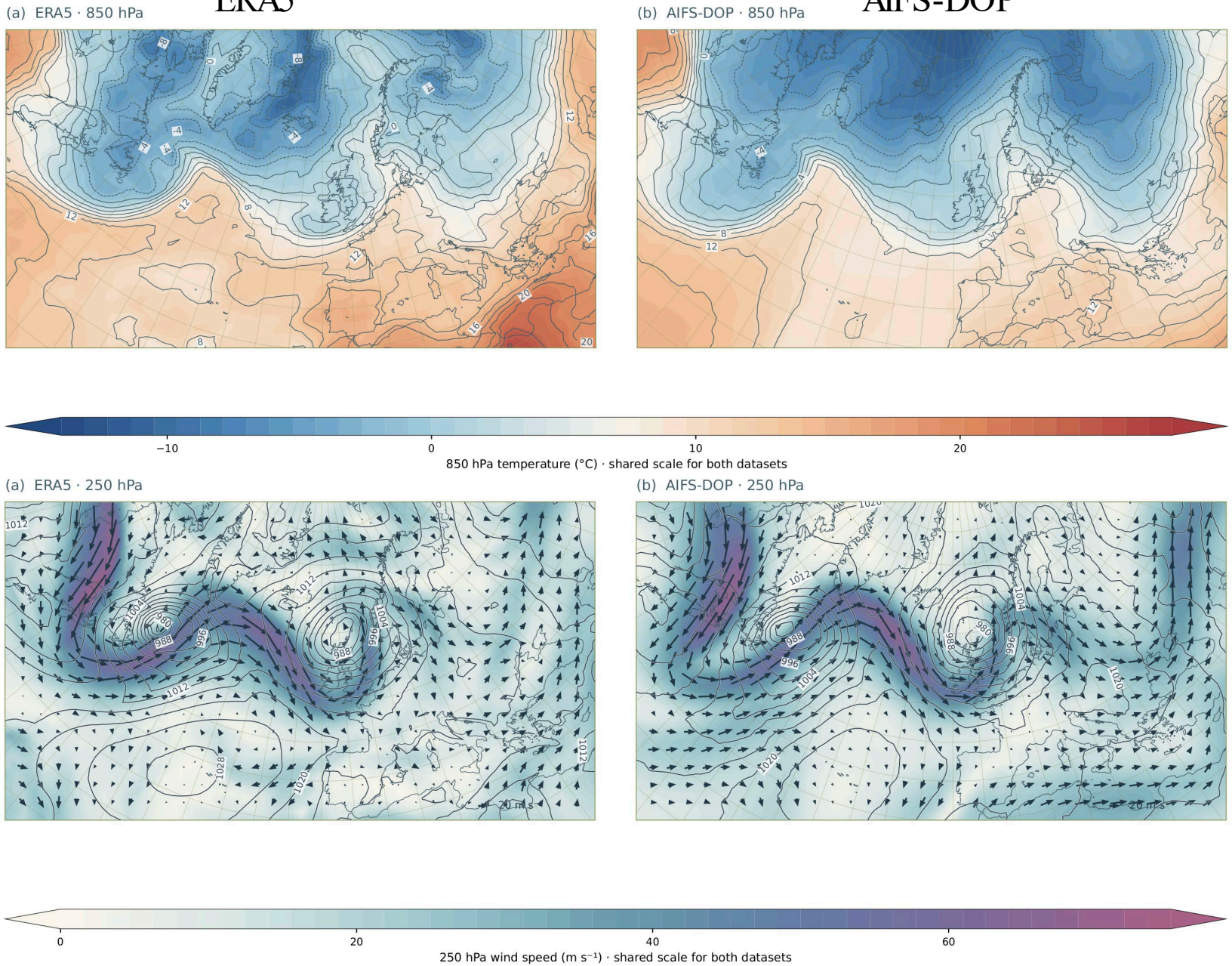
Case study: D-Day

June 5th, 1944 00 UTC

- No upper air temperature observations were input to AIFS-DOP
- No wind observations above 500hPa

ERA5

AIFS-DOP



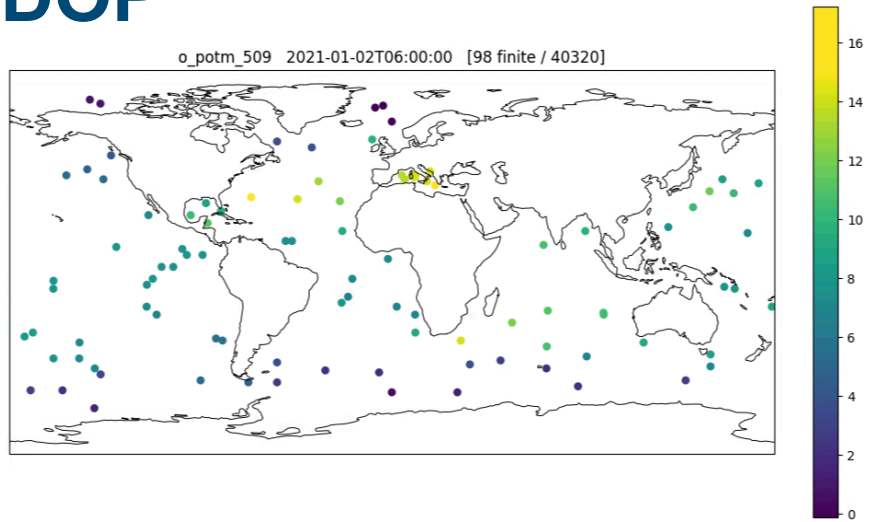
850hPa temp

250hPa wind

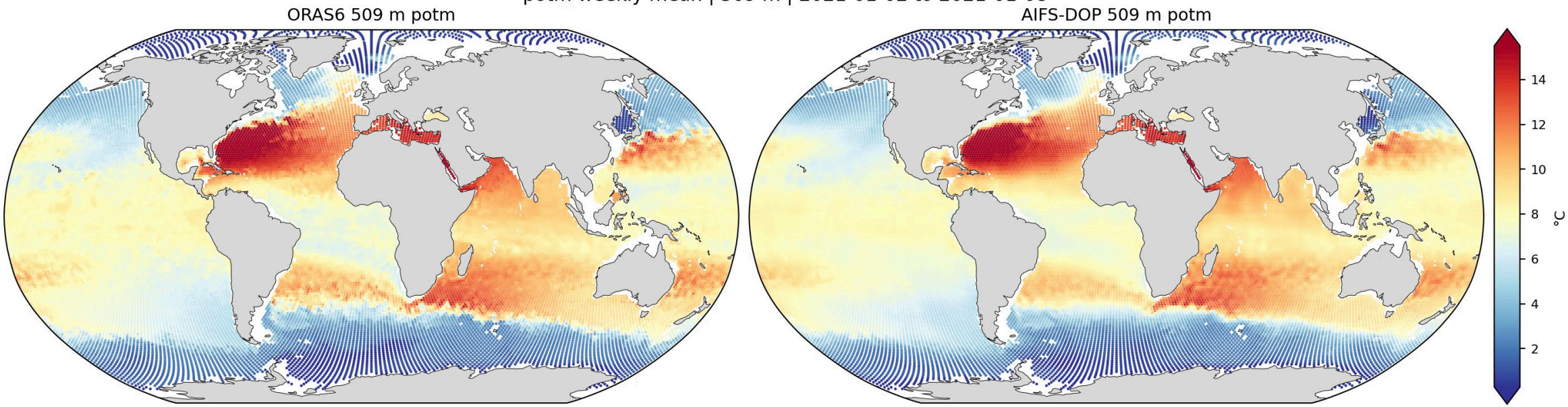
Towards fully-coupled DOP

Phil Browne

Ocean



Potential temperature at 509m depth

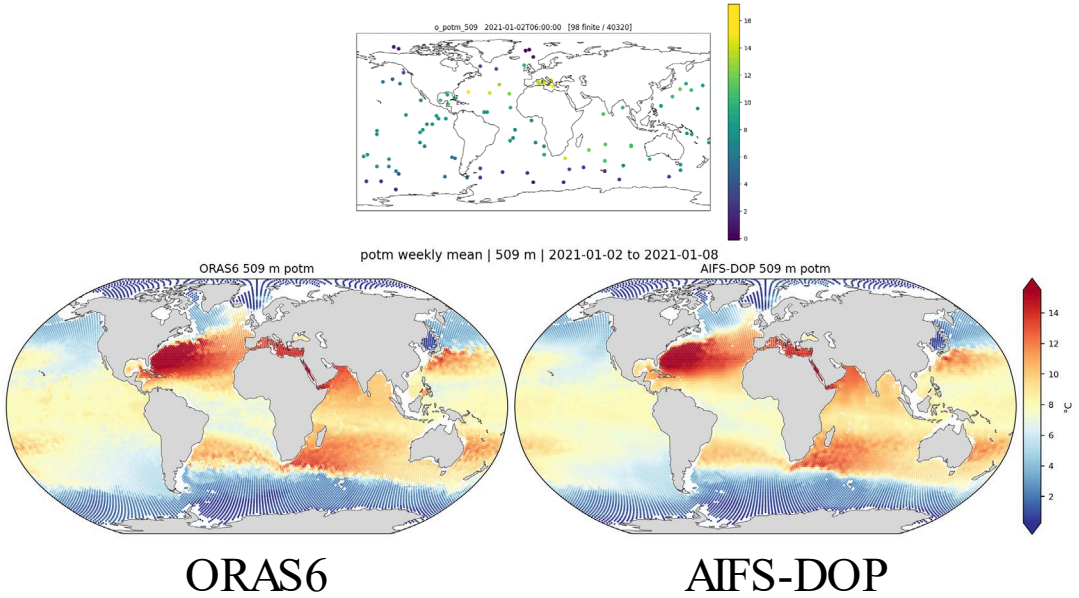


ORAS6

AIFS-DOP

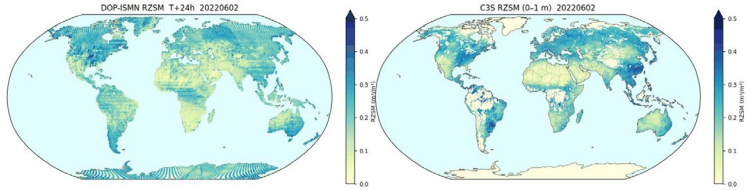
Towards fully-coupled DOP

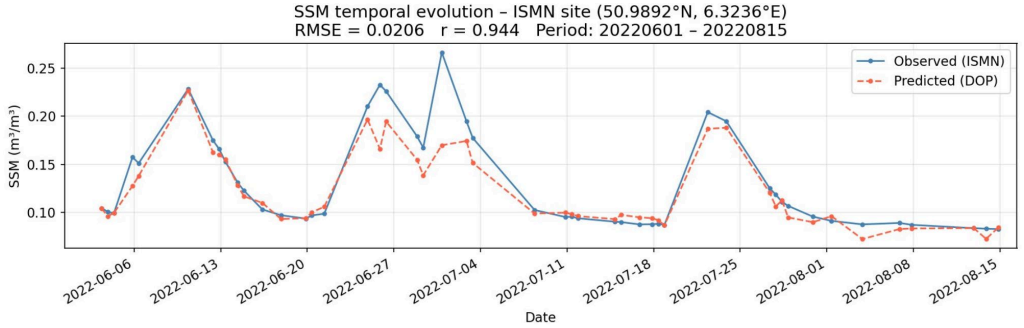
Ocean



ORAS6 AIFS-DOP

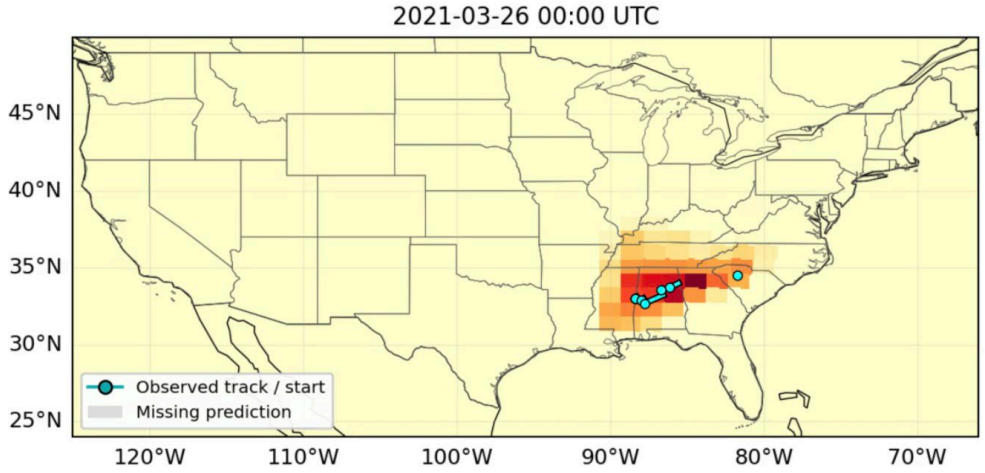
Land surface



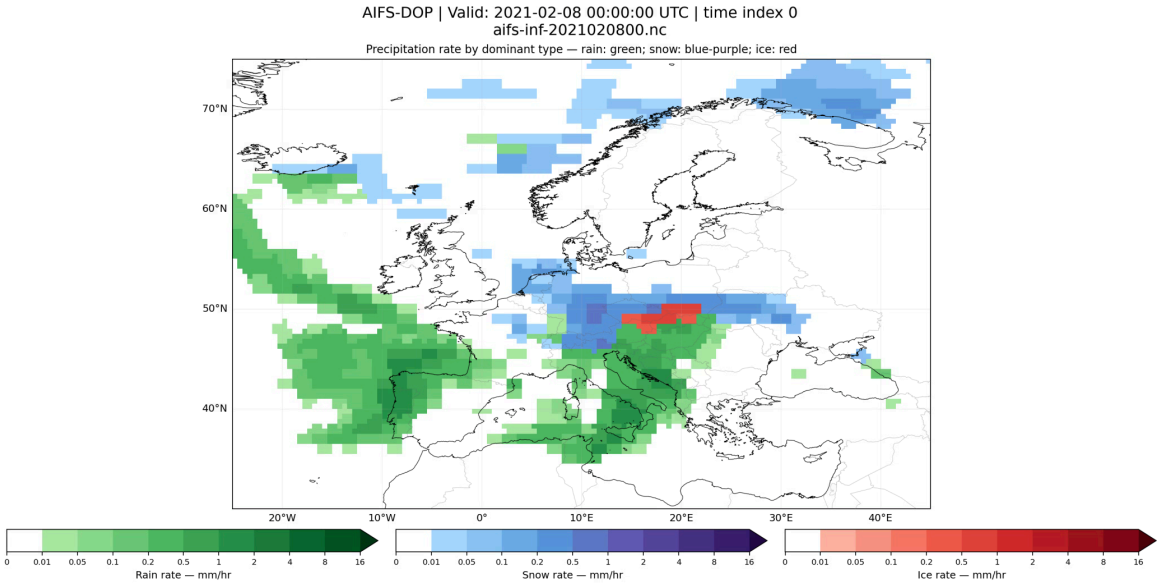


* Sebastien Garrigues

Making use of more exotic observation types



Tornado tracks



SYNOP present weather codes : precipitation type

Summary of Direct Observation Prediction reanalysis prototype

- Data assimilation combines observations with a model prior to provide a spatially complete state estimate
- DOP learns a model of the Earth System during training which performs a similar role at inference time;
  - allowing gaps between sparse observations to be filled in a coherent way
- 40-year DOP reanalysis can be completed in **O(1-day)** opening-up many new possibilities

- **Encouraging results from this first prototype, but plenty of work still to do**

  to better understand, evaluate and improve the performance of data-driven models for reanalysis