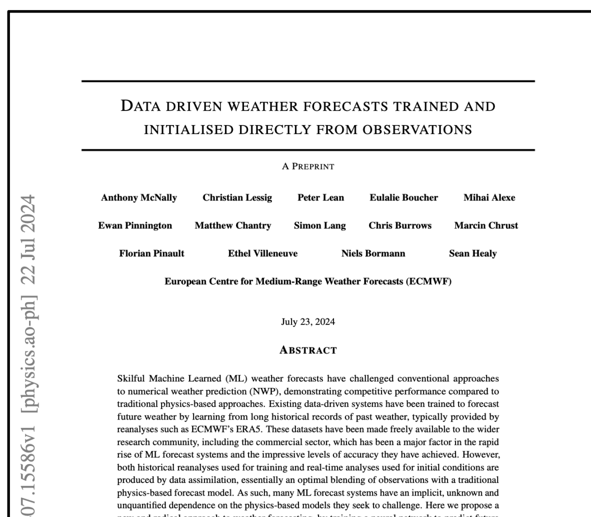
Skillful weather predictions from observations alone

AI-DOP

The AI-DOP team, notably Tony McNally, Mihai Alexe, Christian Lessig, Peter Lean, Ewan Pinnington, Eulalie Boucher and Simon Lang.

Presented by Chris Burrows



07.15586v1 [physics.ao-ph] 22 Jul 2024

DATA DRIVEN WEATHER FORECASTS TRAINED AND INITIALISED DIRECTLY FROM OBSERVATIONS

A PREPRINT

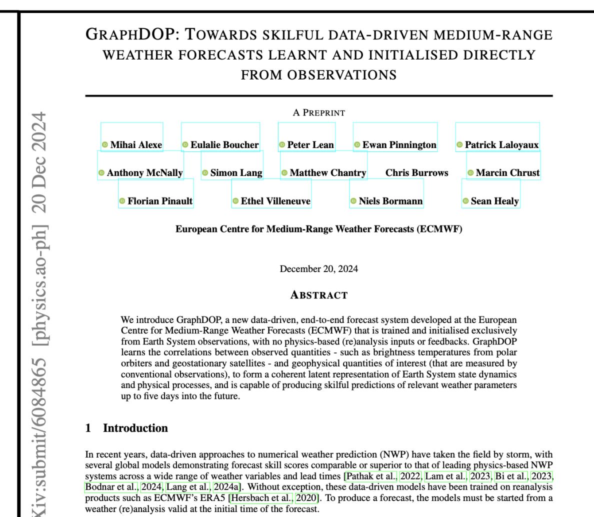
Anthony McNally Christian Lessig Peter Lean Eulalie Boucher Mihai Alexe
Ewan Pinnington Matthew Chantry Simon Lang Chris Burrows Marcin Chrust
Florian Pinault Ethel Villeneuve Niels Bormann Sean Healy

European Centre for Medium-Range Weather Forecasts (ECMWF)

July 23, 2024

ABSTRACT

Skilful Machine Learned (ML) weather forecasts have challenged conventional approaches to numerical weather prediction (NWP), demonstrating competitive performance compared to traditional physics-based approaches. Existing data-driven systems have been trained to forecast future weather by learning from long historical records of past weather, typically provided by reanalyses such as ECMWF's ERA5. These datasets have been made freely available to the wider research community, including the commercial sector, which has been a major factor in the rapid rise of ML forecast systems and the impressive levels of accuracy they have achieved. However, both historical reanalyses used for training and real-time analyses used for initial conditions are produced by data assimilation, essentially an optimal blending of observations with a traditional physics-based forecast model. As such, many ML forecast systems have an implicit, unknown and unquantified dependence on the physics-based models they seek to challenge. Here we propose a



Xiv:submit/6084865 [physics.ao-ph] 20 Dec 2024

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)

December 20, 2024

ABSTRACT

We introduce GraphDOP, a new data-driven, end-to-end forecast system developed at the European Centre for Medium-Range Weather Forecasts (ECMWF) that is trained and initialised exclusively from Earth System observations, with no physics-based (re)analysis inputs or feedbacks. GraphDOP learns the correlations between observed quantities - such as brightness temperatures from polar orbiters and geostationary satellites - and geophysical quantities of interest (that are measured by conventional observations), to form a coherent latent representation of Earth System state dynamics and physical processes, and is capable of producing skilful predictions of relevant weather parameters up to five days into the future.

1 Introduction

In recent years, data-driven approaches to numerical weather prediction (NWP) have taken the field by storm, with several global models demonstrating forecast skill scores comparable or superior to that of leading physics-based NWP systems across a wide range of weather variables and lead times [Pathak et al., 2022, Lam et al., 2023, Bi et al., 2023, Bodnar et al., 2024, Lang et al., 2024a]. Without exception, these data-driven models have been trained on reanalysis products such as ECMWF's ERA5 [Hersbach et al., 2020]. To produce a forecast, the models must be started from a weather (re)analysis valid at the initial time of the forecast.

AI is creating a big buzz in numerical weather prediction – some recent history...

- In 2022, Ryan Keisler showed that it was possible to train a data-driven weather prediction model on ECMWF **reanalysis** fields. https://doi.org/10.48550/arXiv.2202.07575

- Later that year, Google DeepMind unveiled **GraphCast** (https://doi.org/10.48550/arXiv.2212.12794), which appeared to have comparable accuracy with cutting edge NWP models.
- Many others soon appeared! (FuXi, Arches Weather, FourCastNet, Feng Wu, Pangu, Neural GCM, Stormer....). These are all **trained on reanalyses**.
- In 2023, ECMWF started working on **AIFS** (following the GraphCast concept). This became a **fully operational** forecast model in February 2025.

  https://www.ecmwf.int/en/about/media-centre/news/2025/ecmwfs-ai-forecasts-become-operational
- For some metrics, AIFS is better than IFS, but for others it is worse.

- What are the limitations of this approach?

These methods rely on physically-based **reanalyses**

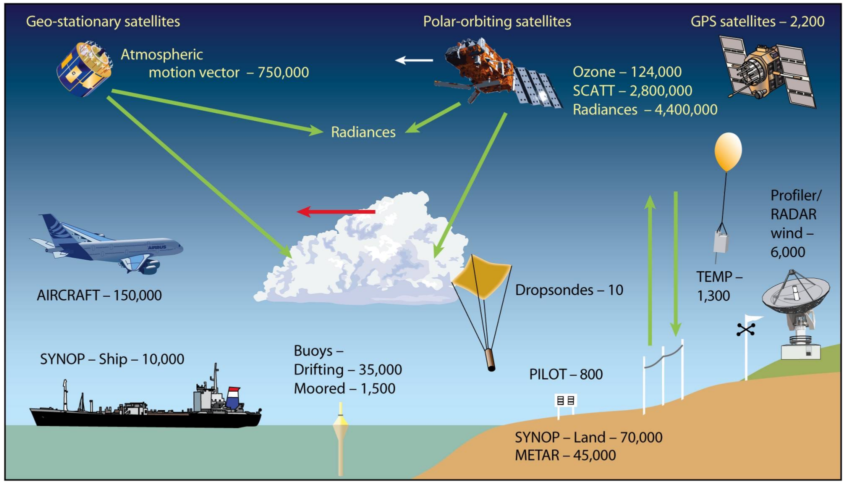
- Is it necessary to rely on these products?
- Are their errors capable of limiting our accuracy long-term?

Suggestion: train and run weather forecasts based on **observations *alone***

This means....

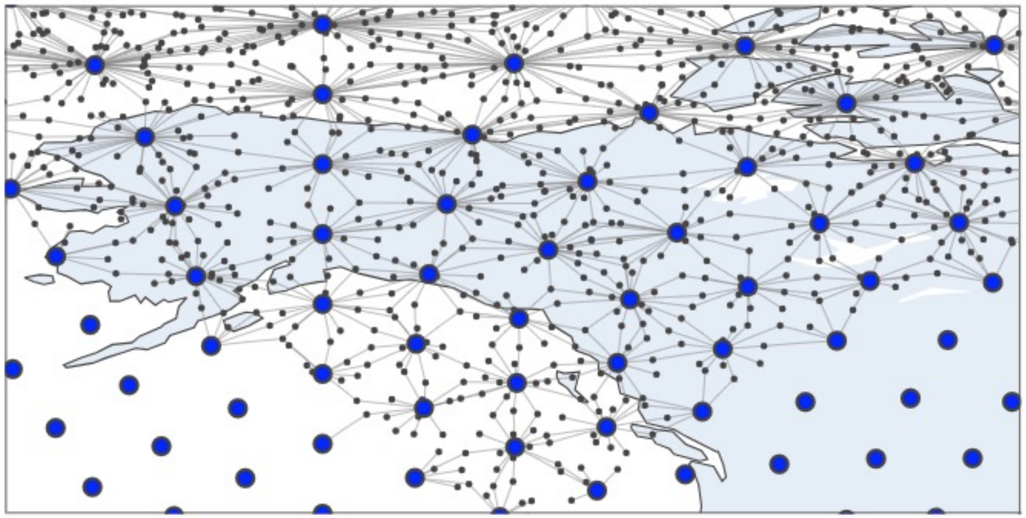
No NWP fields or physical equations are used in:

- observation pre-processing
- quality control
- training
- running the forecast.



18 months ago, DOP was born

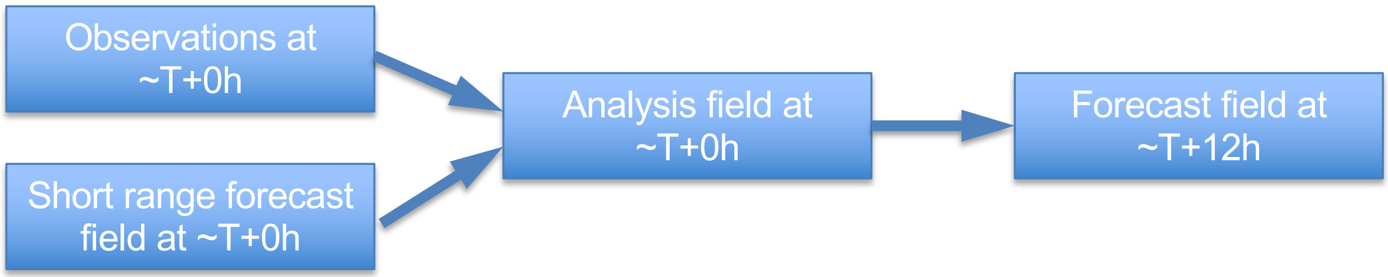
Direct Observation Prediction



≠

Disclaimer: developments are happening rapidly, so results shown here are from a mixed set of configurations.

Traditional NWP framework:



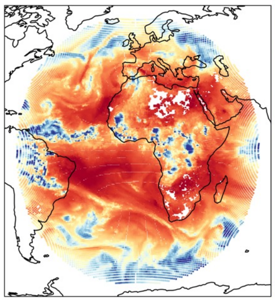
AI-DOP framework:

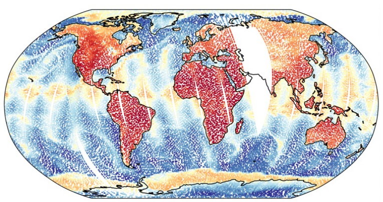
Cycling this with a purely obs-based “background” is planned.

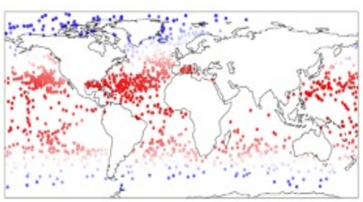
Train AI-DOP to predict future observations in their native units

Inputs:

All observations in a
12 hour window



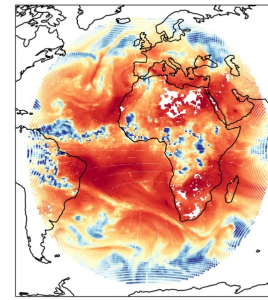


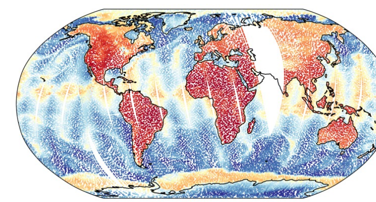


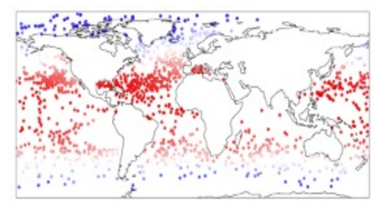
AI-DOP model

Outputs:

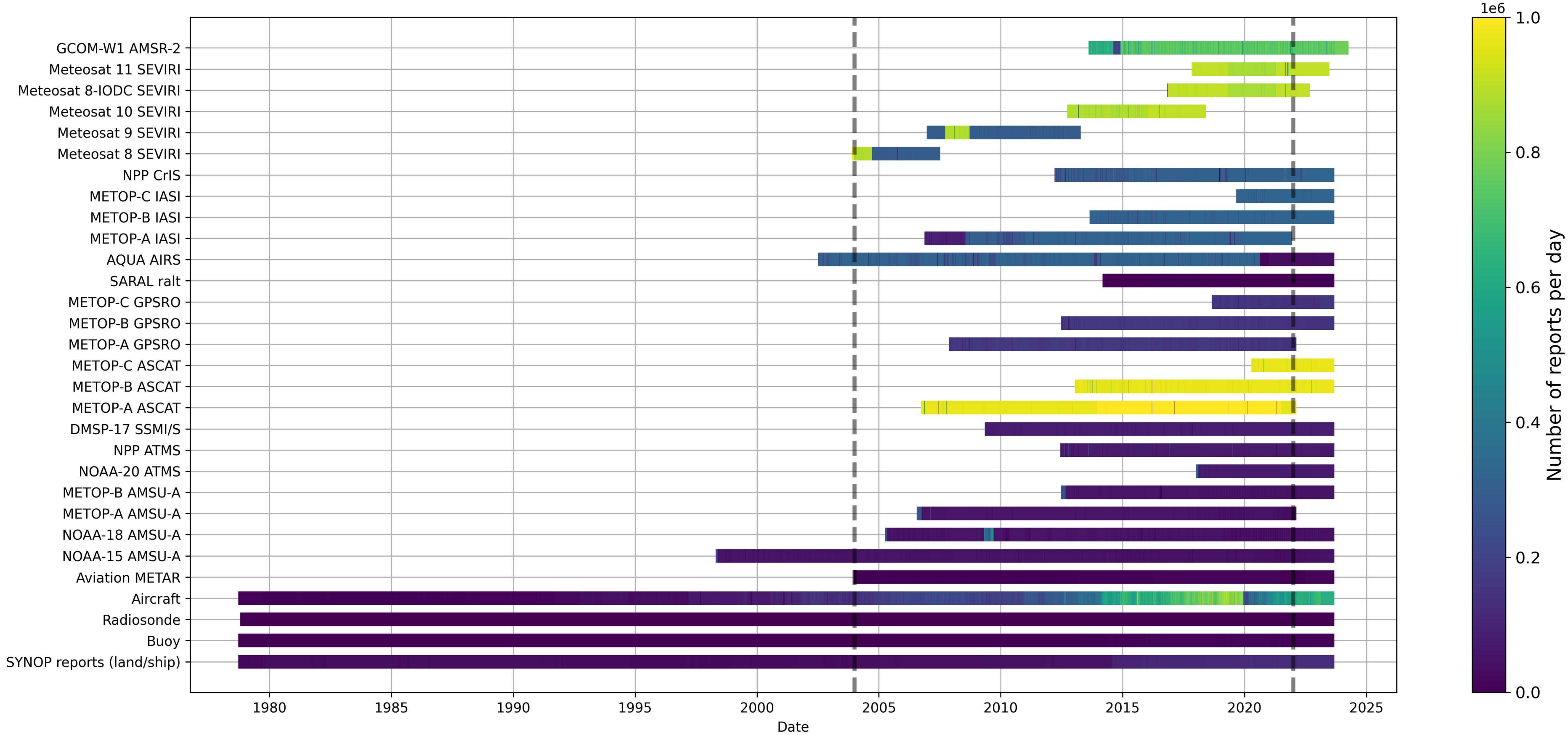
All observations in the **next**
12 hour window





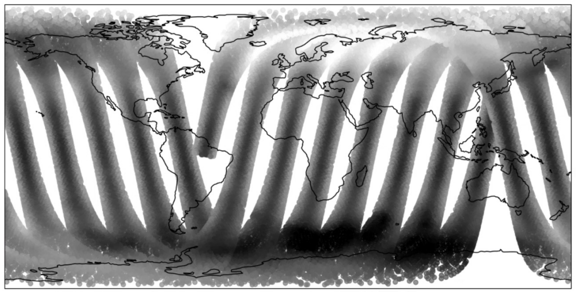


Observations used to train DOP

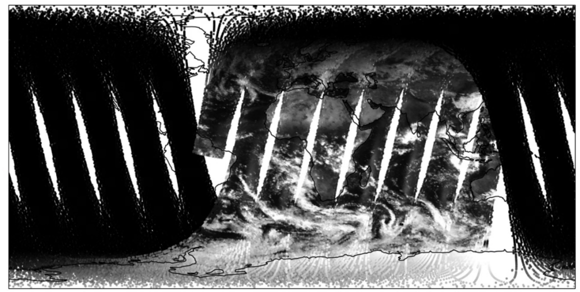


As of April 2025, ca. **81 billion reports** available for us to train on – ca. **7TB of zarr data**

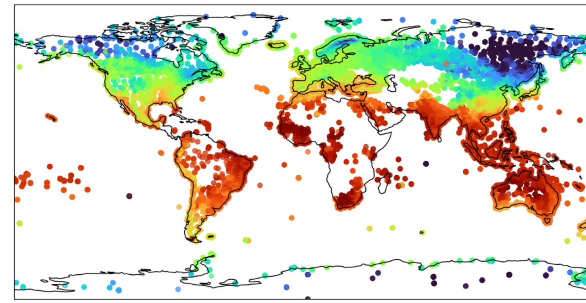
More instruments are being added …



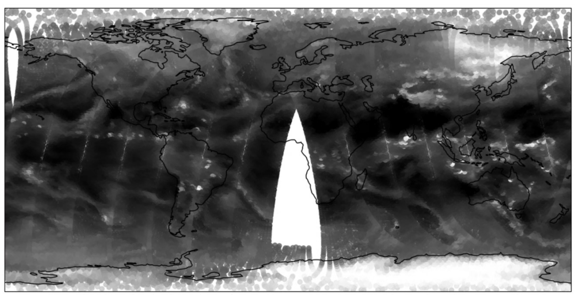
METOP-B AMSU-A ch7



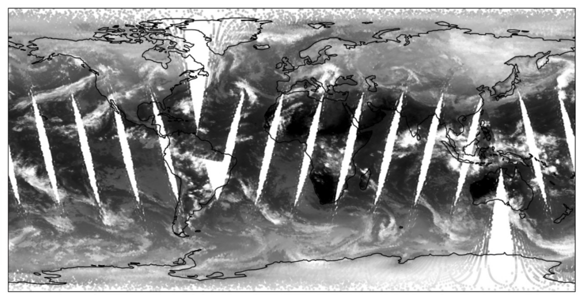
AVHRR visible reflectances



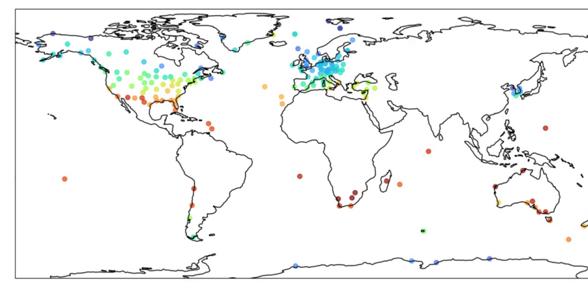
BUFR Land SYNOP 2m Temperature



NPP ATMS ch18



METOP-B IASI ch756



BUFR Land TEMP 850hPa Temperature

Clearing up some potential confusion...

1. What is the use of predicting observations at **observation locations only**?
   – For training, the observations are all we have. But when we run the forecast, we can predict the observations anywhere! We can even predict them at gridded locations 😃.

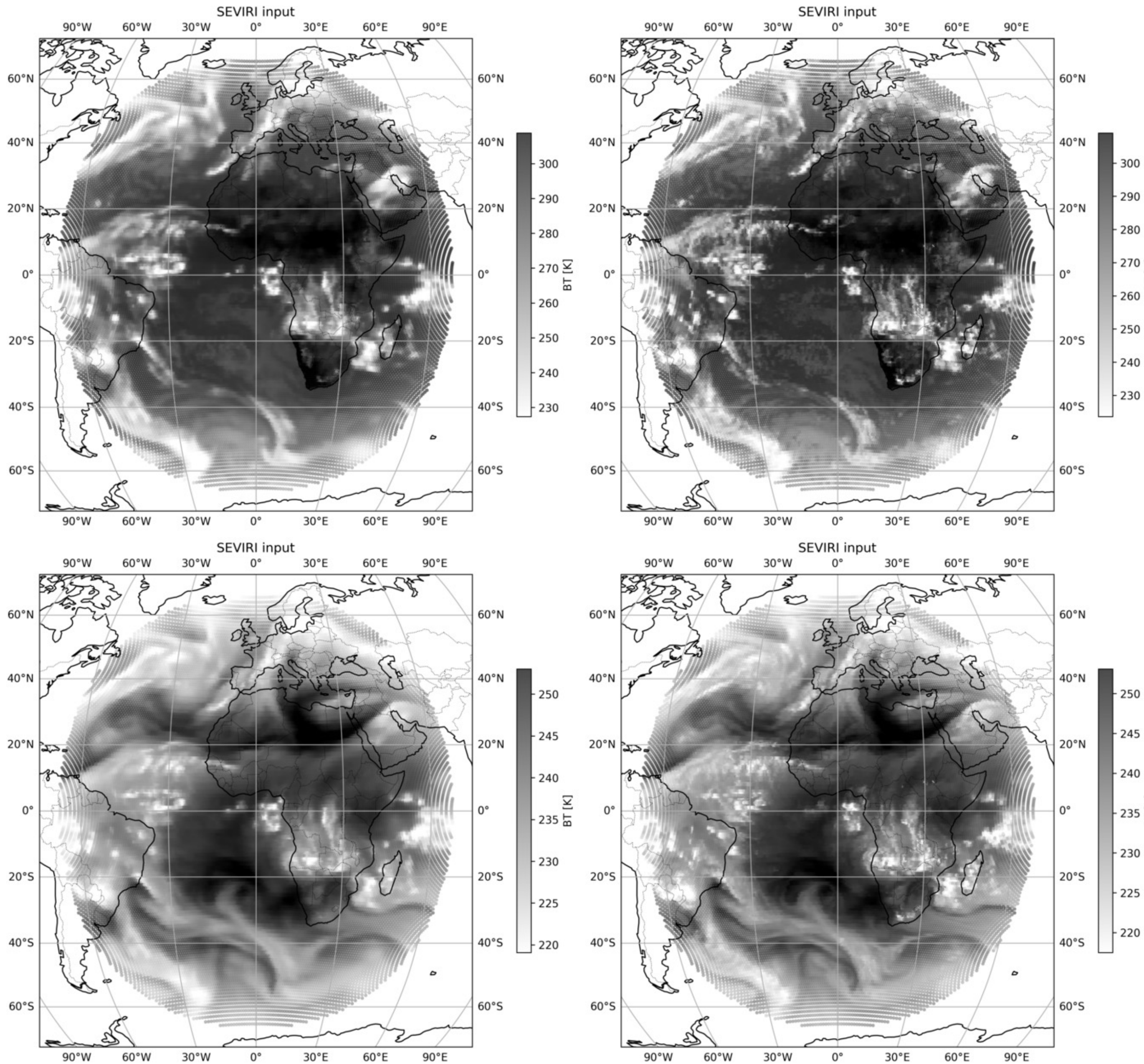
2. What is the value in predicting brightness temperatures and bending angles etc? **We want to know T, P, Q!**
   – This is a good point. In practice, users would mostly be interested in simulated synop/radiosonde observations as these measure geophysical variables.
   – However, including the satellite observations in the training allows the network to learn correlations between the various observation types, so in regions with few in-situ observations, the satellite data can allow us to predict "pseudo-radiosondes" etc.

So, is it any good???

SEVIRI prediction SEVIRI target

10.8µm

6.2µm



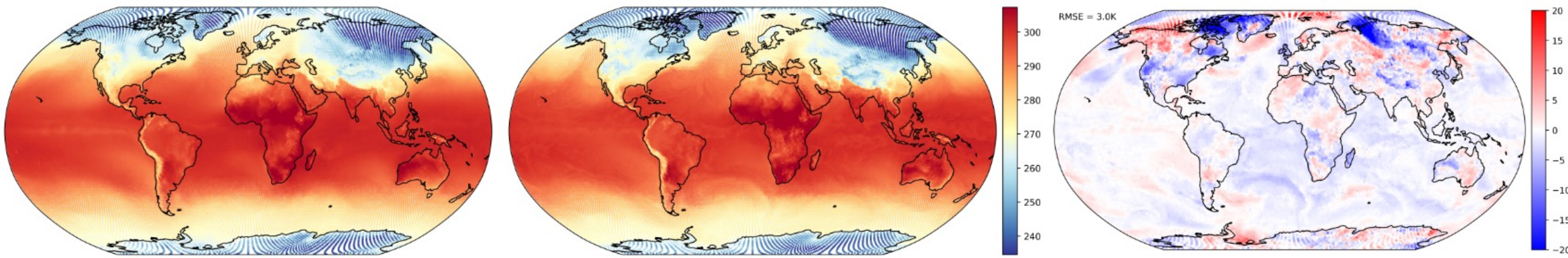
Reminder:

There is no explicit physics in these predictions.

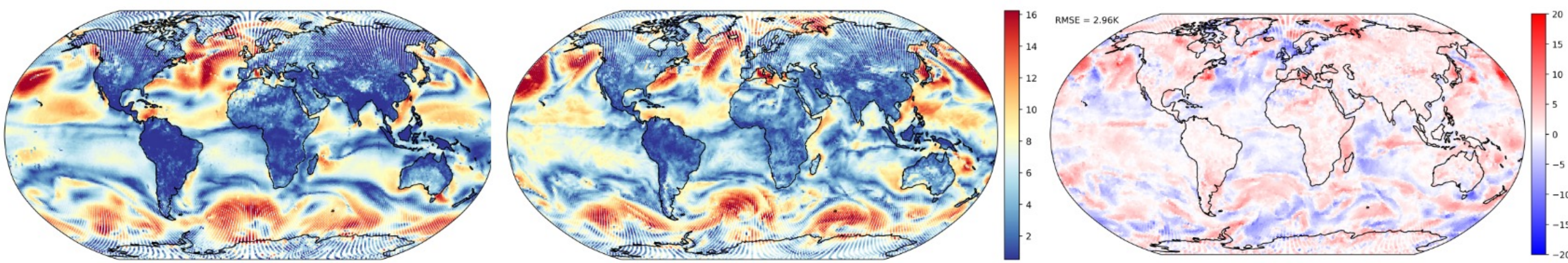
All the atmospheric dynamics have been learned from observations!

Gridded 5-day forecasts

AI-DOP ERA5 RMSE



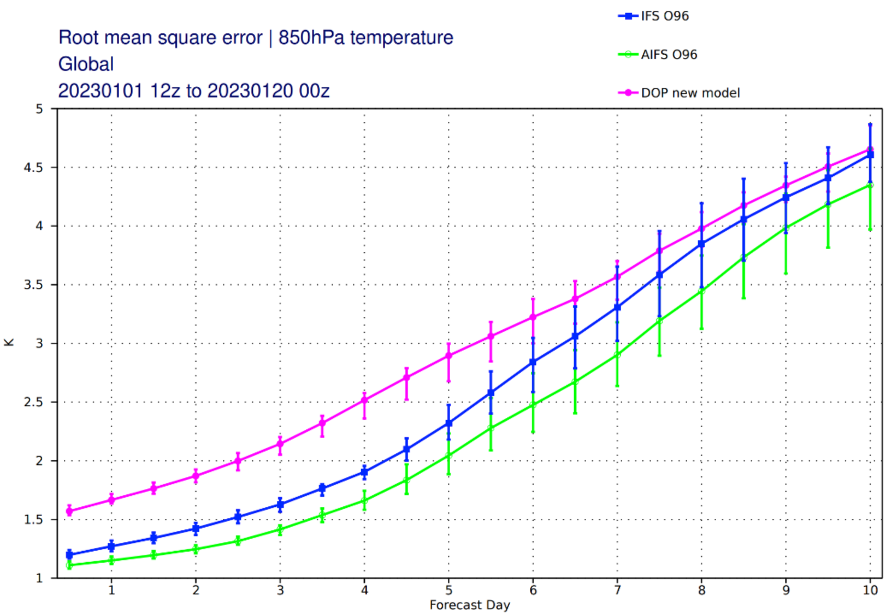
2m temperature



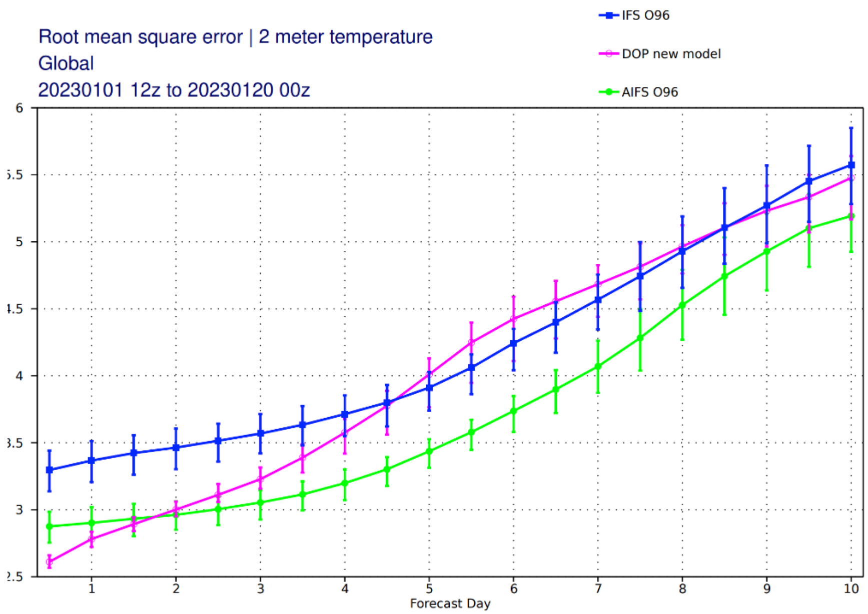
10m wind speed

AI-DOP is progressing rapidly, but there are still many deficiencies compared to IFS/AIFS. However, in some surface metrics, AI-DOP is performing comparably.

Global 850hPa temperature RMSE



Global 2m temperature RMSE



Summary/outlook

- Accurate weather forecasts can be made from observations alone, with no model fields, observation operators or dynamical equations used in the training or inference. **The concept works**.
- The training dataset is being augmented all the time. For example, we are testing with IASI PC scores now.
- Upper air scores are still 1-2 days worse than IFS, but are improving rapidly.
- Surface results are competitive with IFS/AIFS.

- The value of good observations to produce a forecast is clear!
- If we relax the no-model rule or implement a hybrid approach, could we do better? “Background cycling” of the latent space could help.

- Niels Bormann’s presentation will cover more detail about the impact of the observations in AI-DOP.