AI-Direct Observation Prediction

GraphDOP

End-to-end Weather Prediction from Observations

Eulalie Boucher

Mihai Alexe, Peter Lean, Ewan Pinnington, Simon Lang, Patrick Laloyaux, Tony McNally, Mat Chantry, Tomas Kral, Patricia de Rosnay, Niels Bormann, Chris Burrows, Sean Healy, and others across ECMWF

eulalie.boucher@ecmwf.int

Context: Two parallel but complementary forecast streams

1) **Physics/Hybrid**
(analysis-based)

Observations

4D-Var
LDAS
Nemovar

IFS / AIFS
forecast models

Forecasts

→ Operational AIFS still has a hard dependency on 4D-Var (NEMOVAR / LDAS)

2) **Pure ML**
(observation-based)

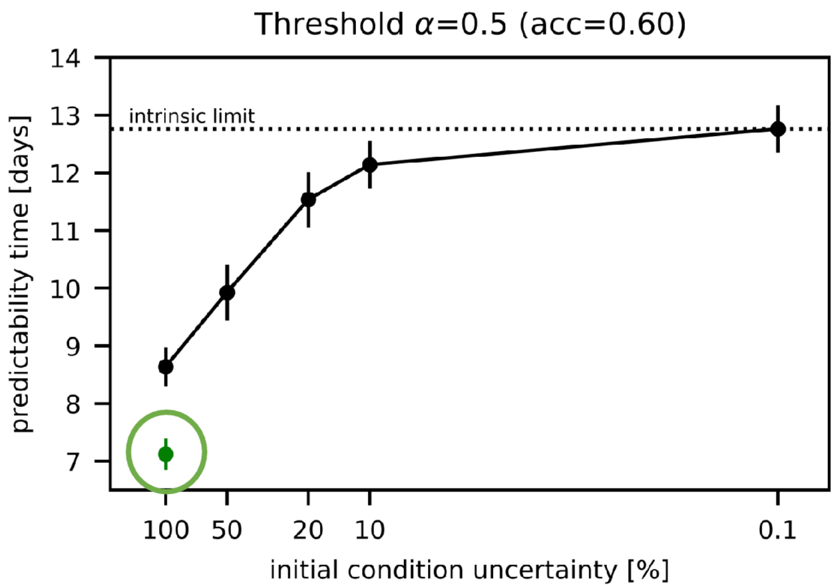
Observations

End-to-end data-driven ML
system

Forecasts
(+ reanalysis?)

→ Ambition is to have **a pure ML observation-based end-to-end system**, what that system will look like is still to be determined – AI-DOP is one possibility, but research incoming from AIFS and WeatherGenerator as well

Motivation:
End-to-end learning from observations has potential to reduce initial condition errors



George Craig : ECMWF DA workshop 2025

- 1 to 2 days improvement possible from improved model
- 4 to 5 days improvement possible from improved initial conditions

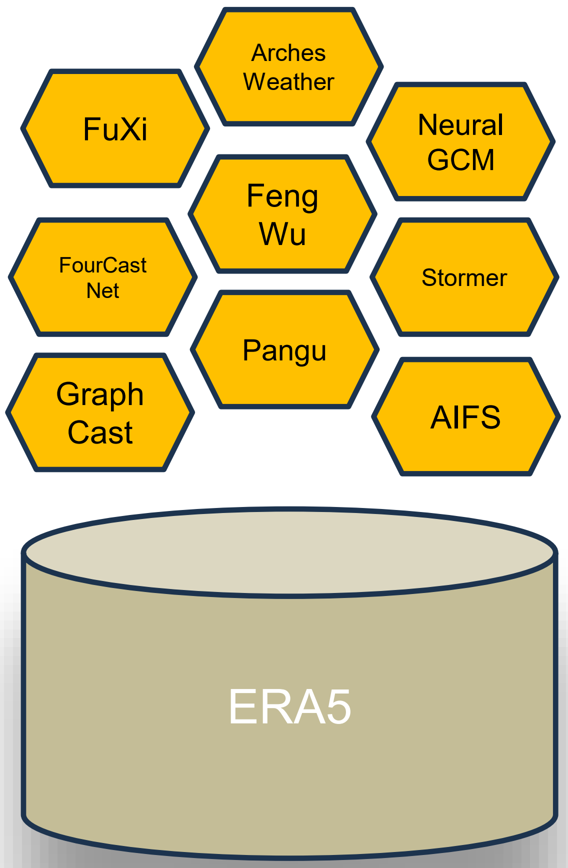
Reanalysis-driven AI models are trained and initialized from the IFS initial conditions
→ learn to **reduce model errors**

Observation-driven AI models
→ can potentially **reduce initial condition *and* model errors**

Reanalysis-driven AI models are trained and initialized from the IFS initial conditions

→ learn to **reduce model errors**

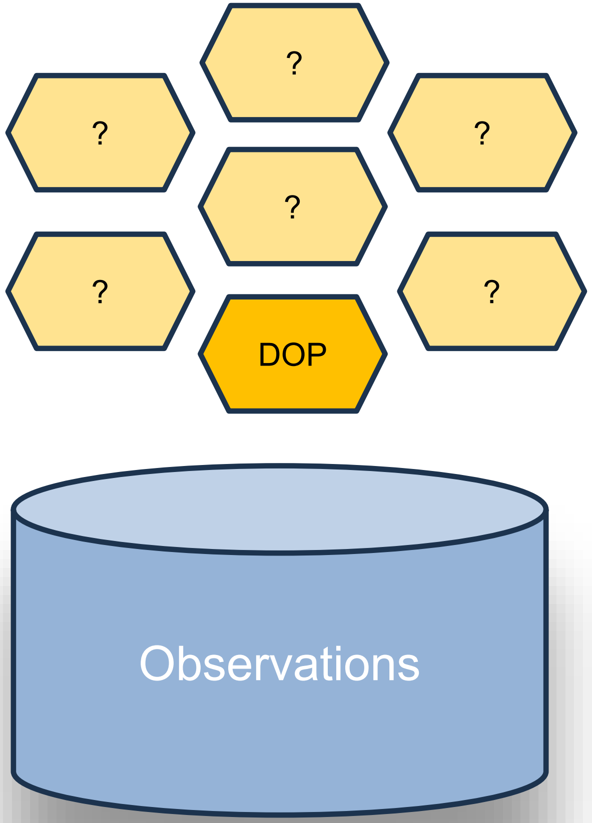
- ERA5 has provided the foundation for the rise of ML in weather forecasting
- Lots of models with different architectures – **all around the same skill**



Observation-driven AI models
→ can potentially **reduce initial condition *and* model errors**

- Ultimately the information content in ERA5 comes from observations (+ model to fill gaps)

- Yet, ERA5 4D-Var assimilates less than 5% of available observations
- To go beyond ERA5
  - Wait for ERA6
  - Use the observations directly



The AI-DOP concept

End-to-end observations to forecast

Training objective:

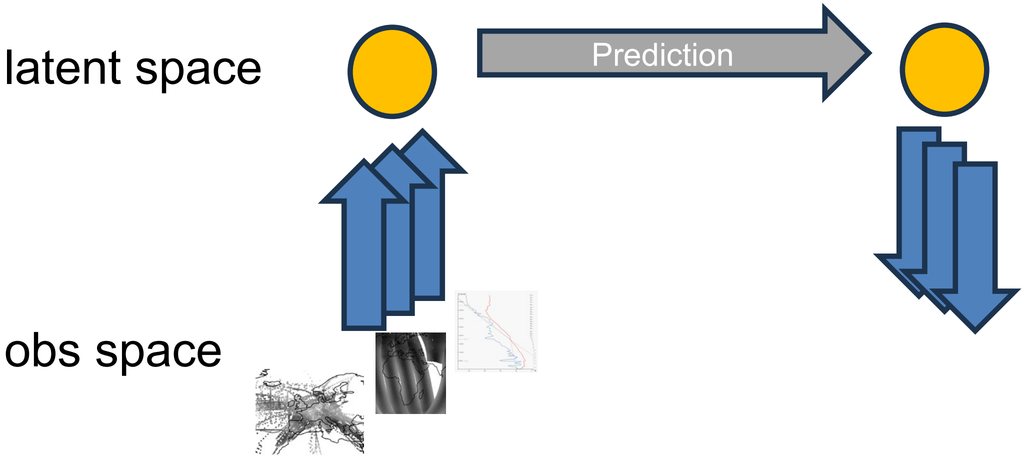
Using all satellite and in situ (conventional) observations in an input window

Forecast all observations in the next window **(with no reanalysis data used)**

To accurately predict future observations, the network must learn a representation of the state and processes that those observations are sensitive to

A **learned forecast model** which evolves the latent state forward in time

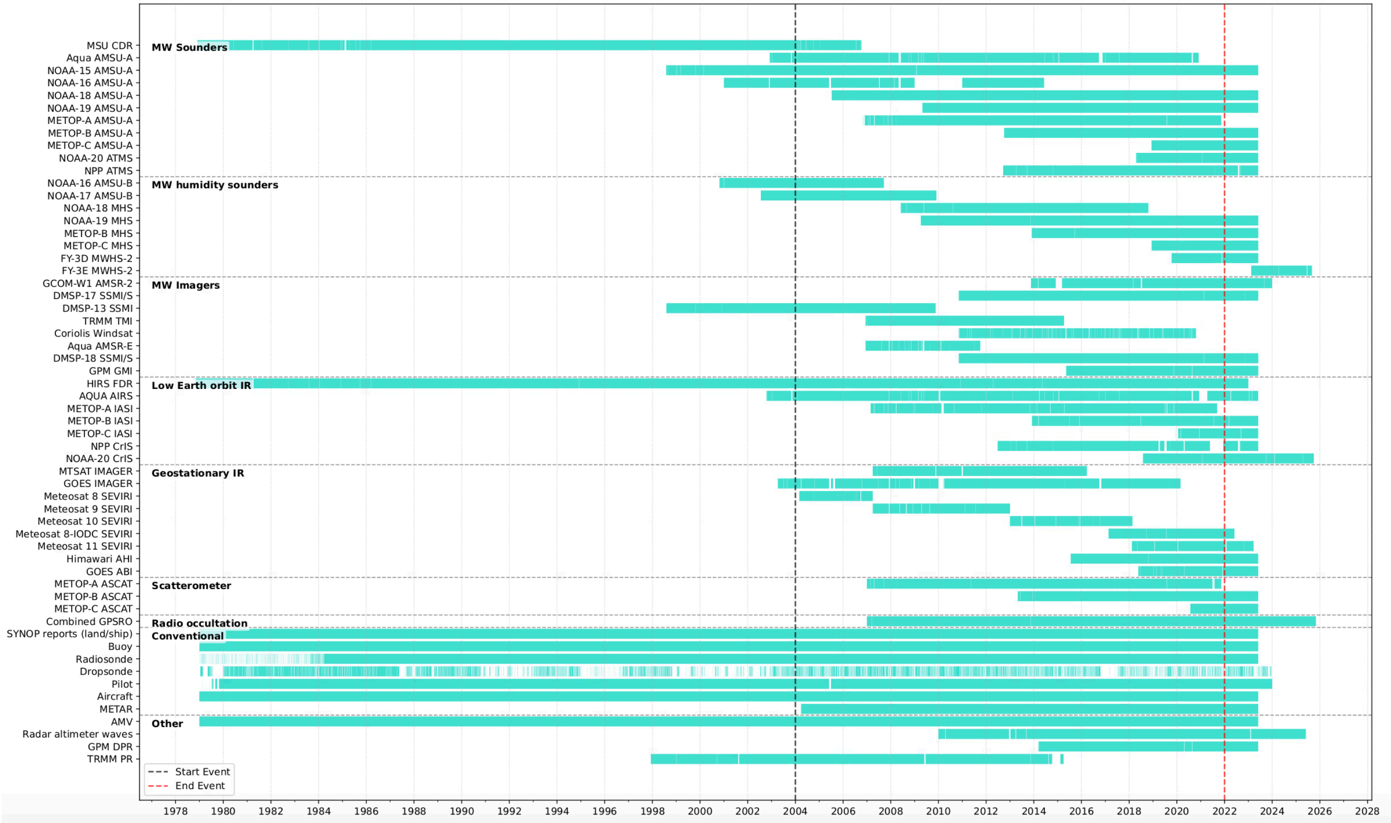
A **learned "DA" function** which combines information from different sources into a (learned) set of unified latent state variables

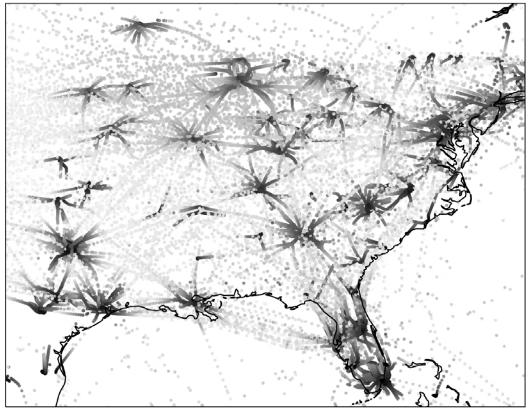


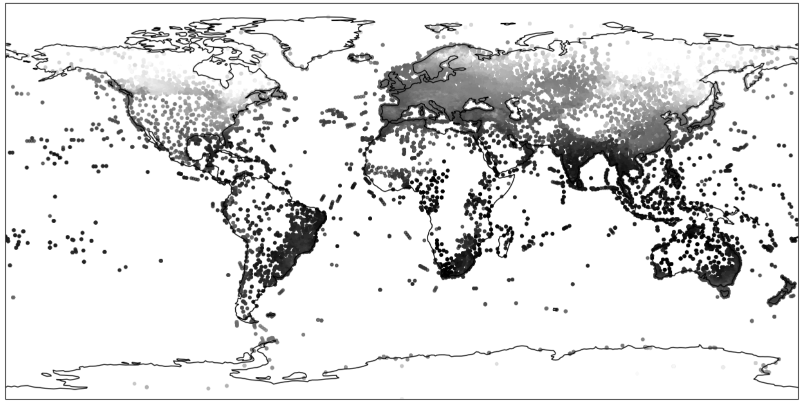
Learned observation operators which map from the network's internal model latent state to observation space

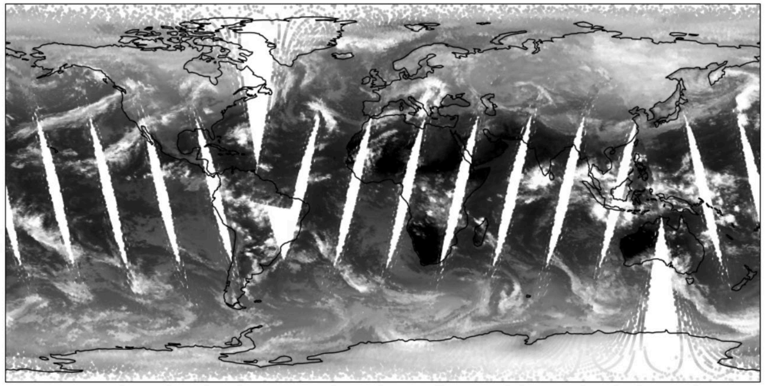
The training observations

- Zarr datasets of all important observation categories
- Includes observations/instruments not currently assimilated in IFS
- Quality control is still important



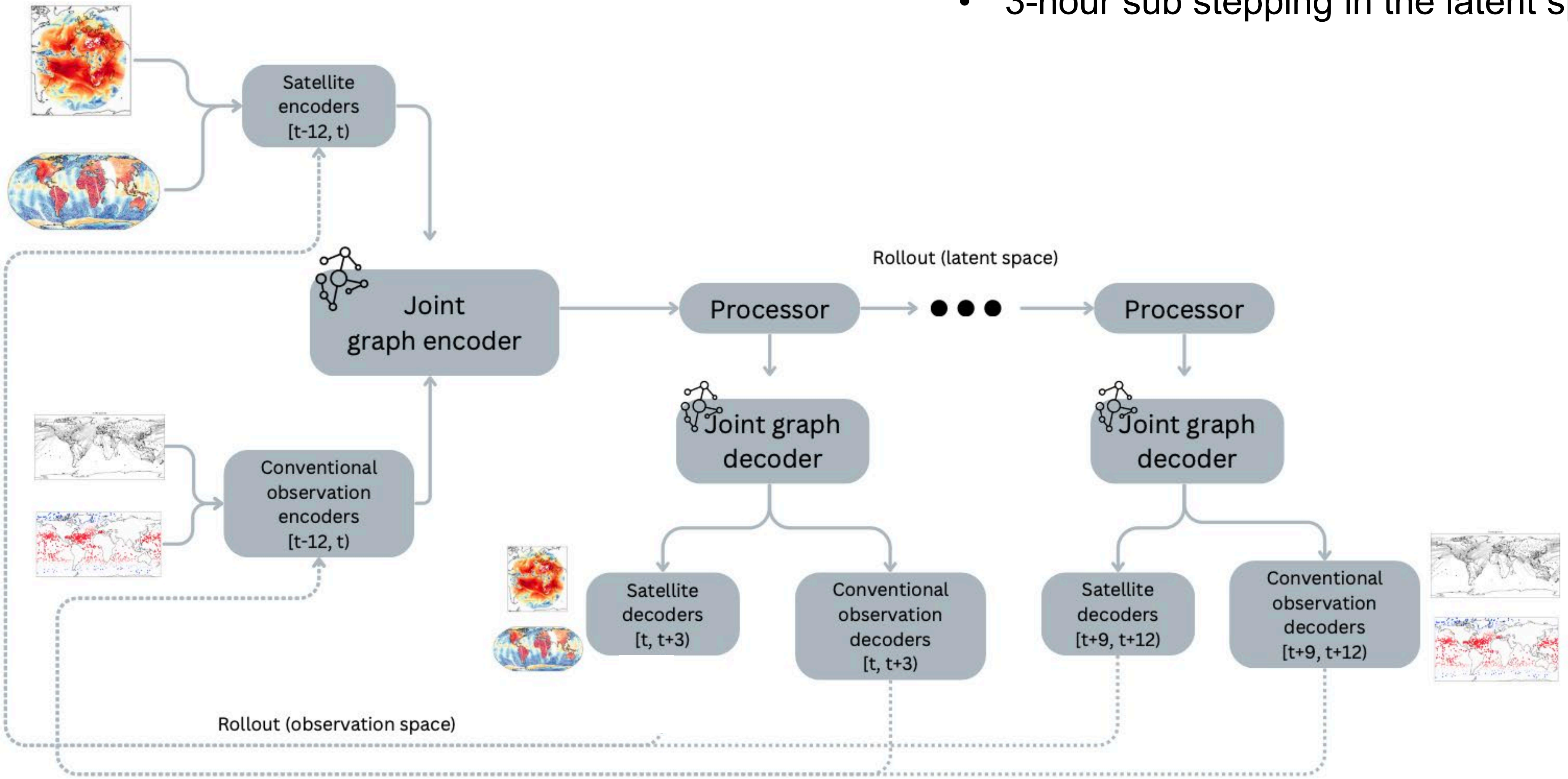






GraphDOP model architecture

- 10-20 year period for training
- 12-hour windows of observations
- 3-hour sub stepping in the latent space

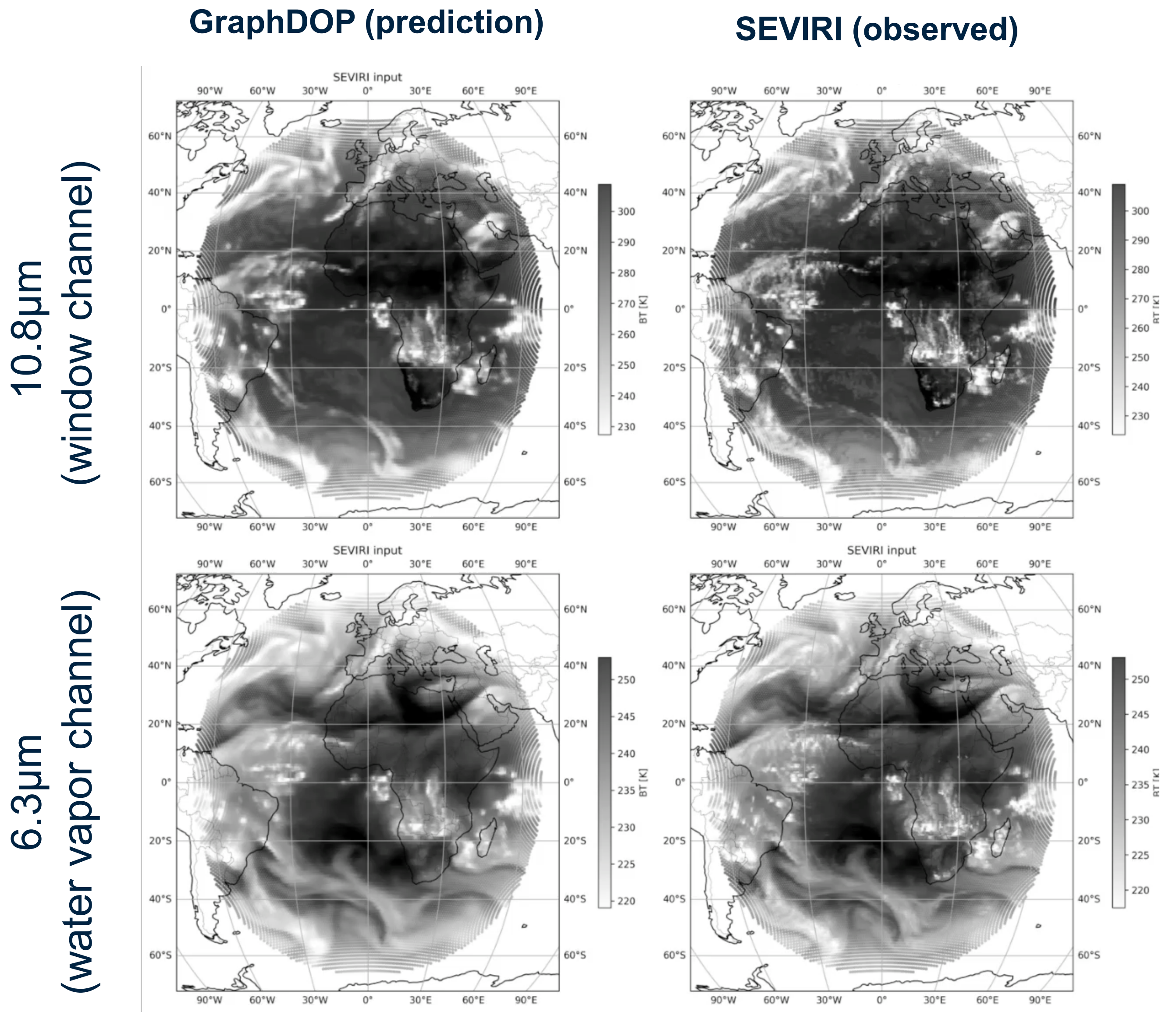


GraphDOP is trained to forecast all observations at their true location

- During training, GraphDOP is trained to forecast all observations at their true location

  → This includes satellite brightness temperatures

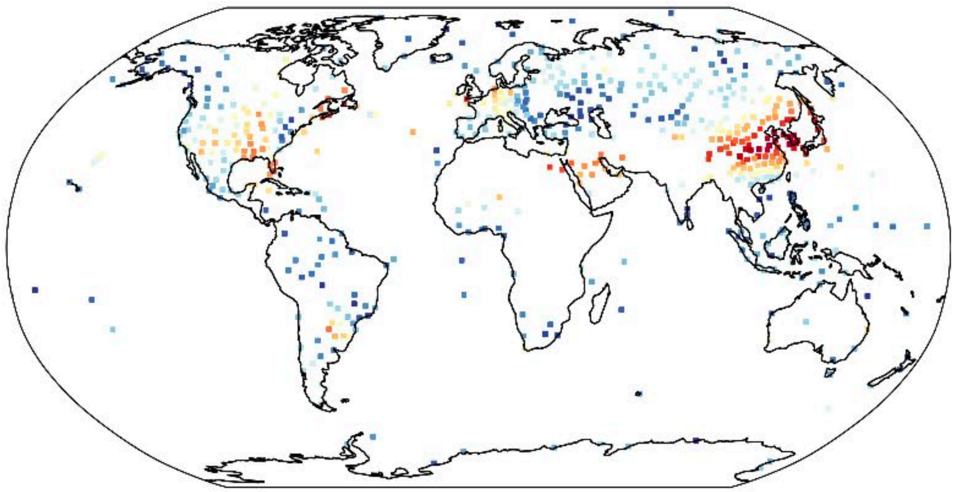
- **Example: 5-day forecast of SEVIRI**
- **Representations of convection and extra-tropical dynamics**



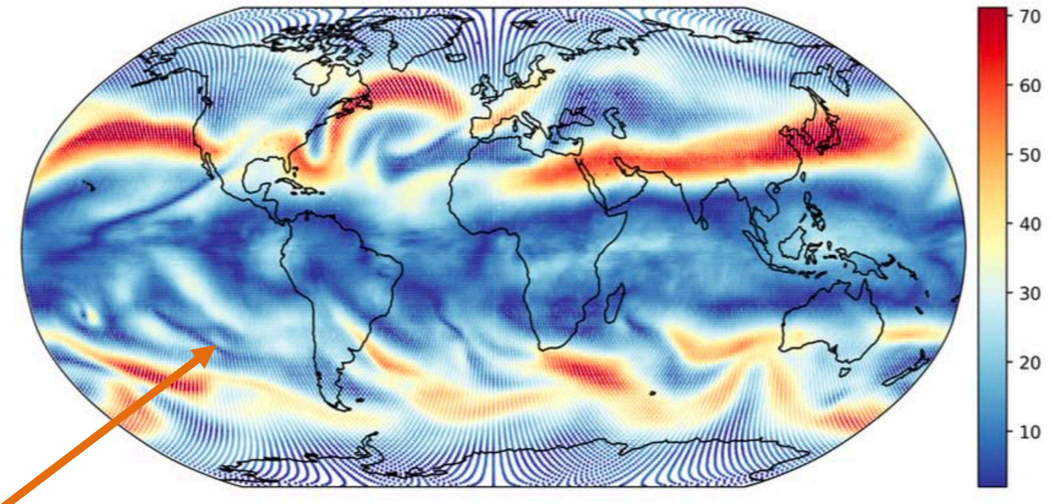
GraphDOP can also forecast on a global grid

- At inference time, the graph decoders can be used to produce a forecast at any arbitrary point

Input 200hPa wind obs



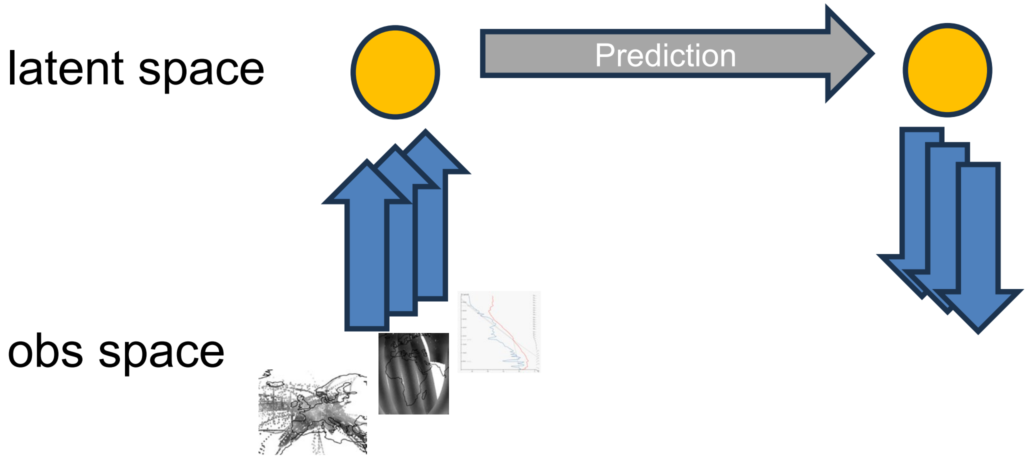
Predicted 200hPa wind at t+24h on O96 grid



Even in areas where there are no direct observations for that variable
→ evidence that it **has developed a model** of the relationship between different observed variables

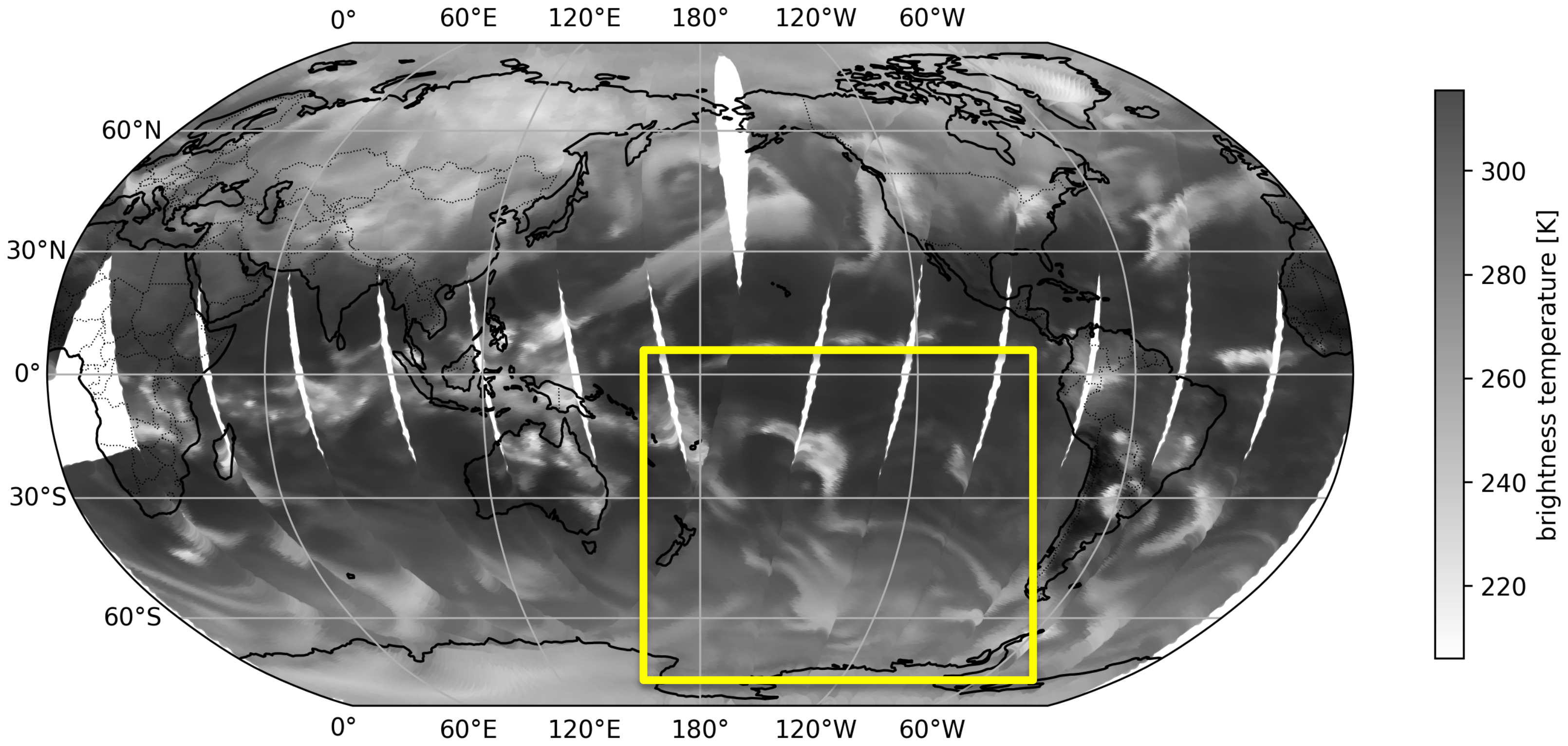
A **learned forecast model** which evolves the latent state forward in time

A **learned "DA" function** which combines information from different sources into a (learned) set of unified latent state variables



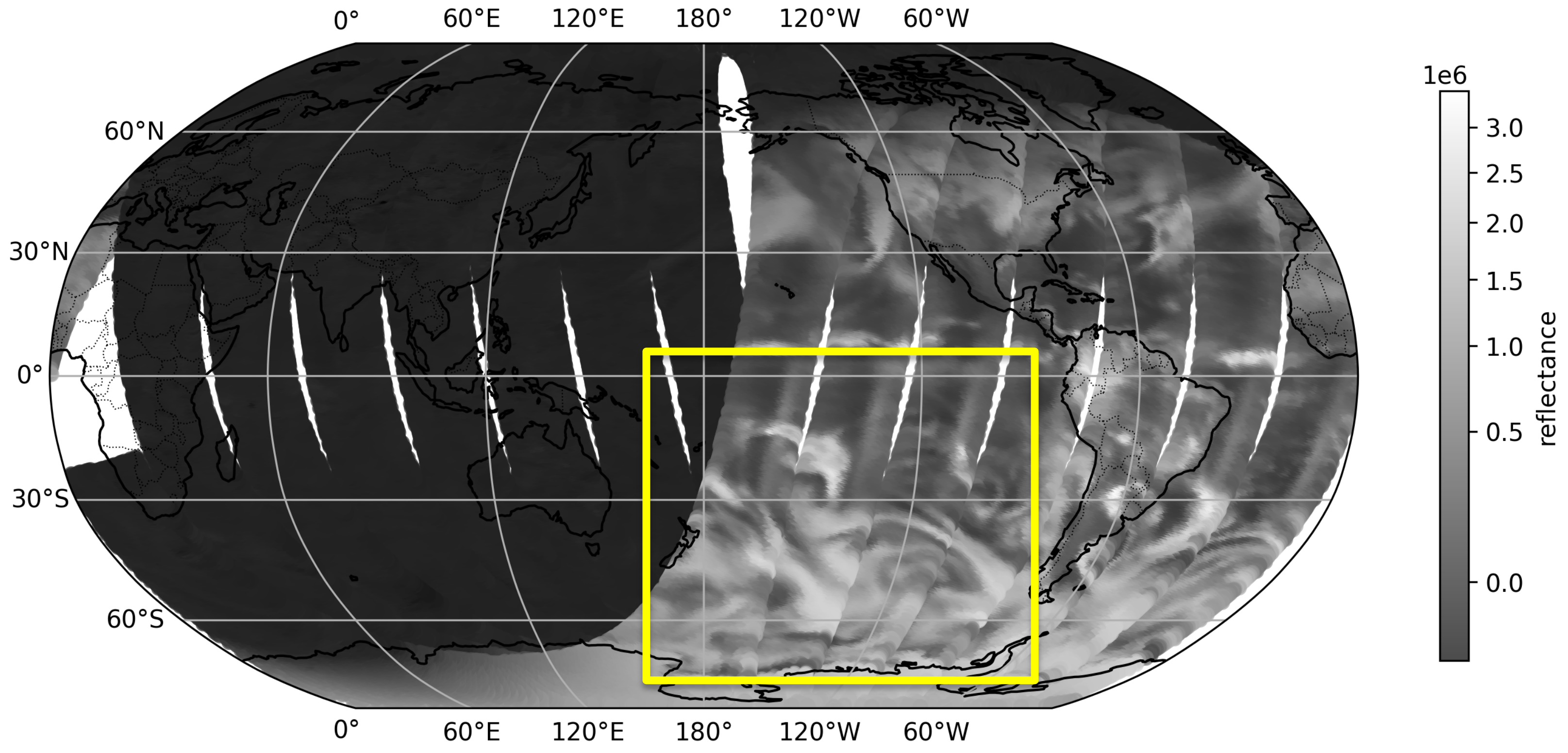
Learned observation operators which map from the network's internal model latent state to observation space

Learning a data assimilation function to combine information from multiple sources into a unified latent state: cloud example



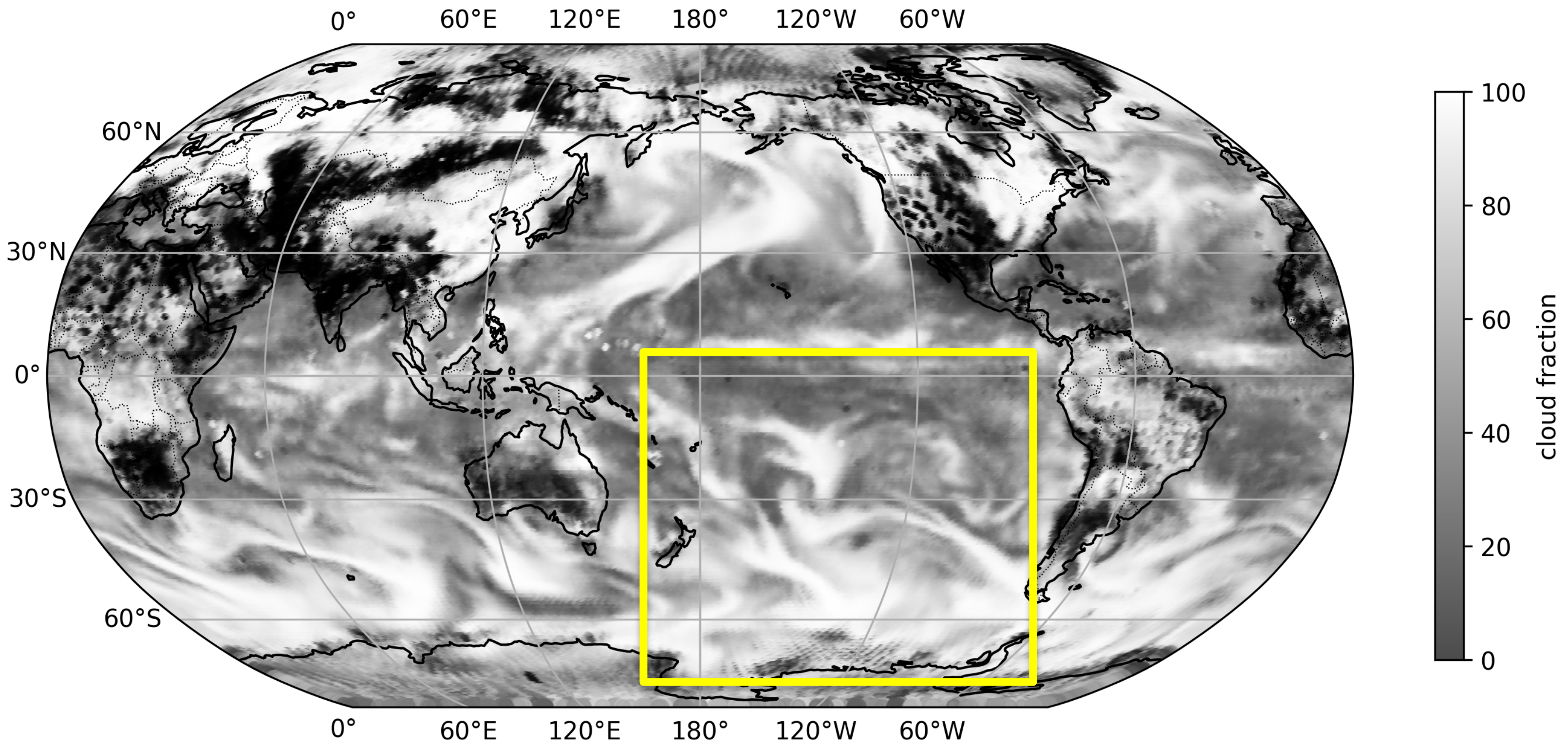
Predicted IASI ch 756

Learning a data assimilation function to combine information from multiple sources into a unified latent state: cloud example



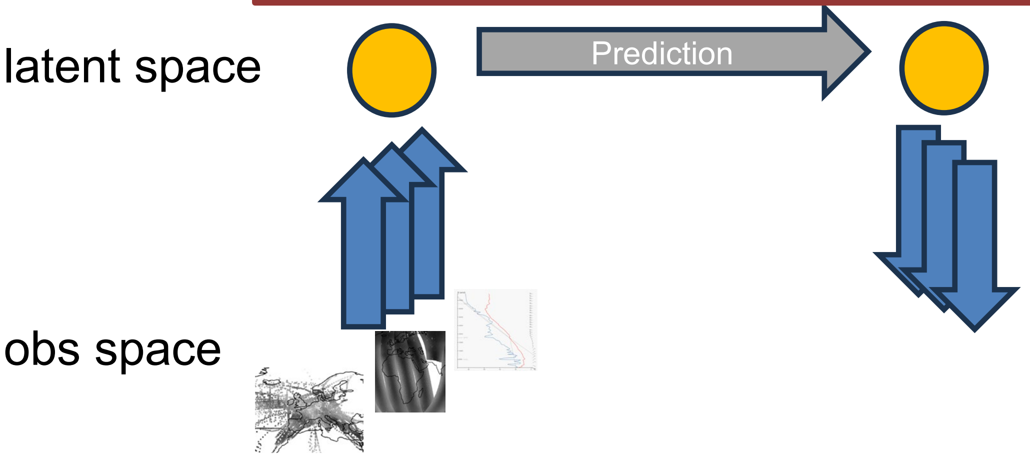
Predicted AVHRR visible

Learning a data assimilation function to combine information from multiple sources into a unified latent state: cloud example



Predicted SYNOP cloud fraction
(on a regular grid)

A **learned forecast model** which evolves the latent state forward in time

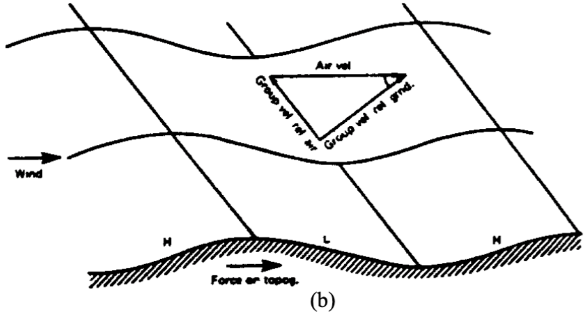


A **learned "DA" function** which combines information from different sources into a (learned) set of unified latent state variables

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

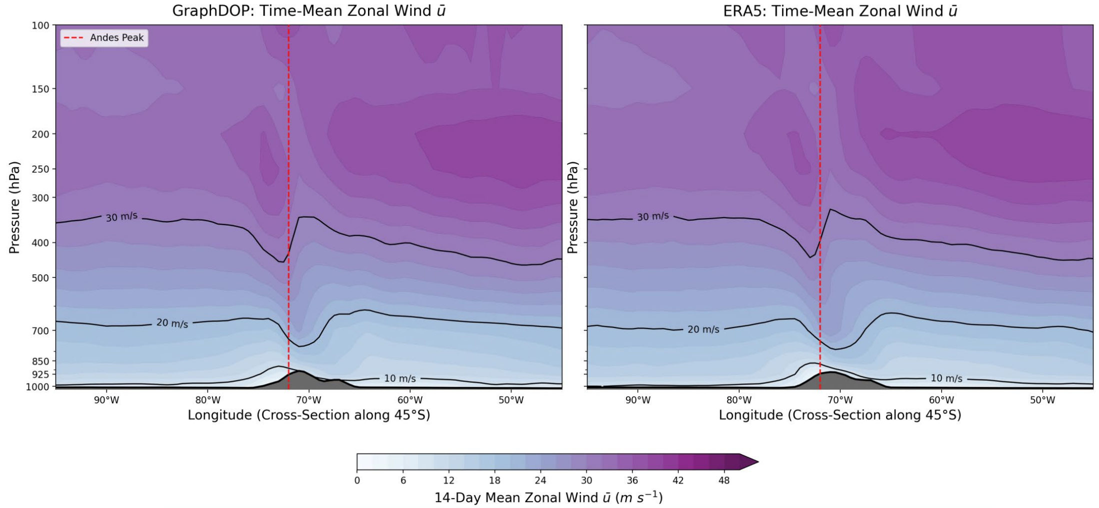
Learning a model of the Earth System from sparse, indirect observations

Flow over orography



GraphDOP

ERA5



Ageostrophic flow

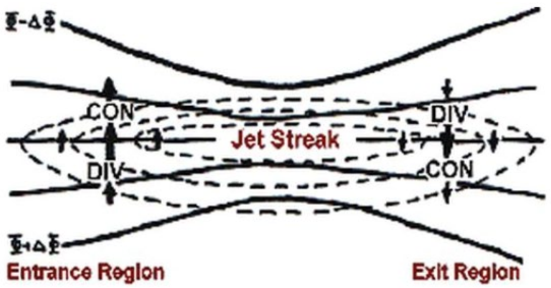
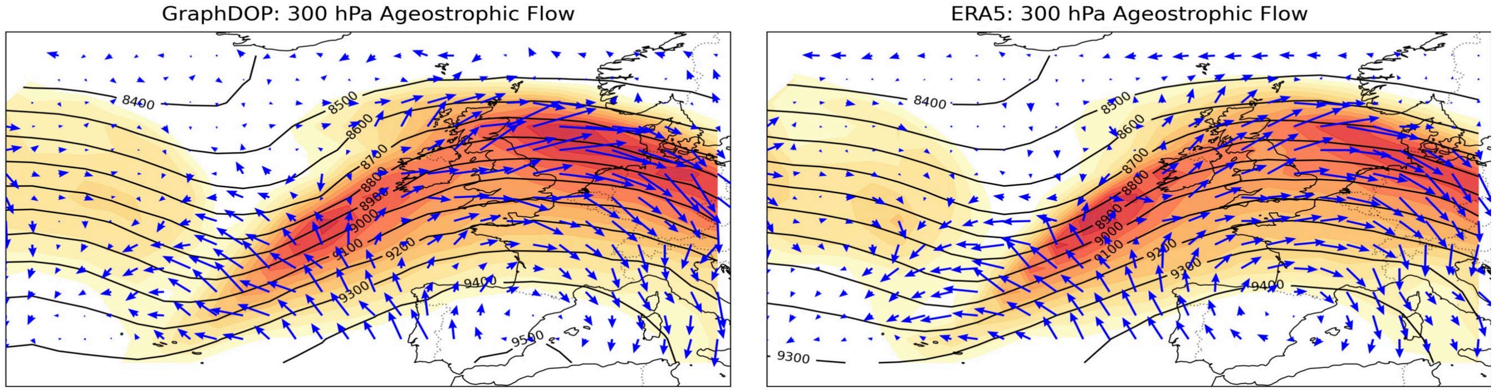


Figure 2: Location of convergence (CON) and divergence (DIV) areas in the entrance and exit region of a jet streak, © NOAA

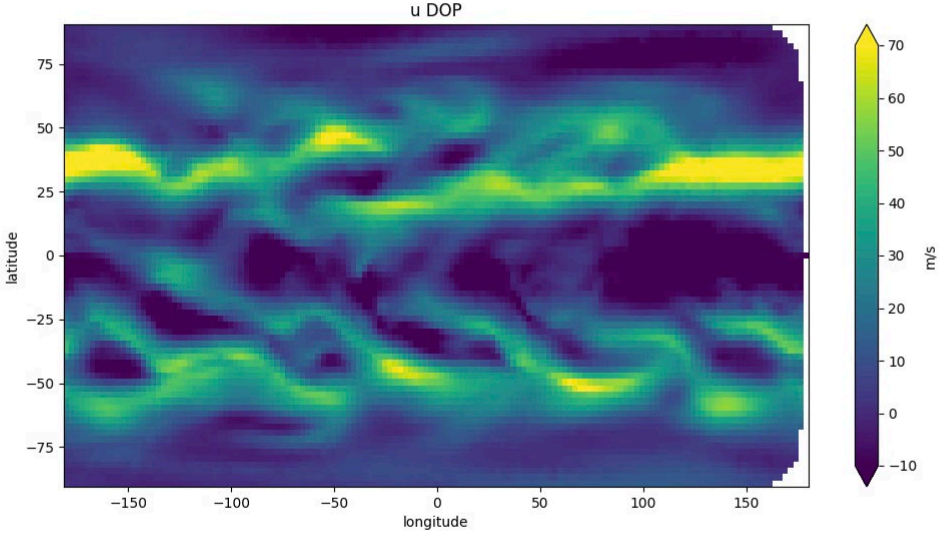
GraphDOP

ERA5

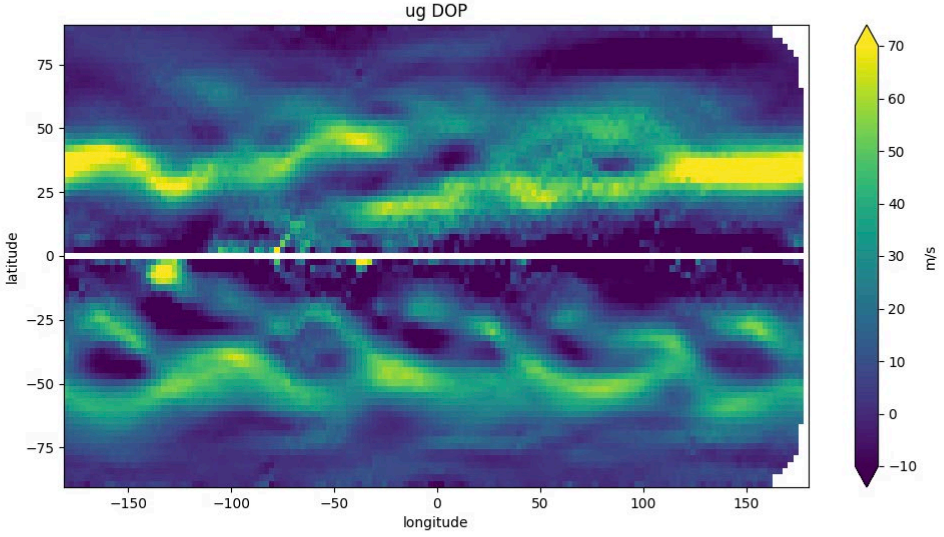


Learning a model of the Earth System from sparse, indirect observations

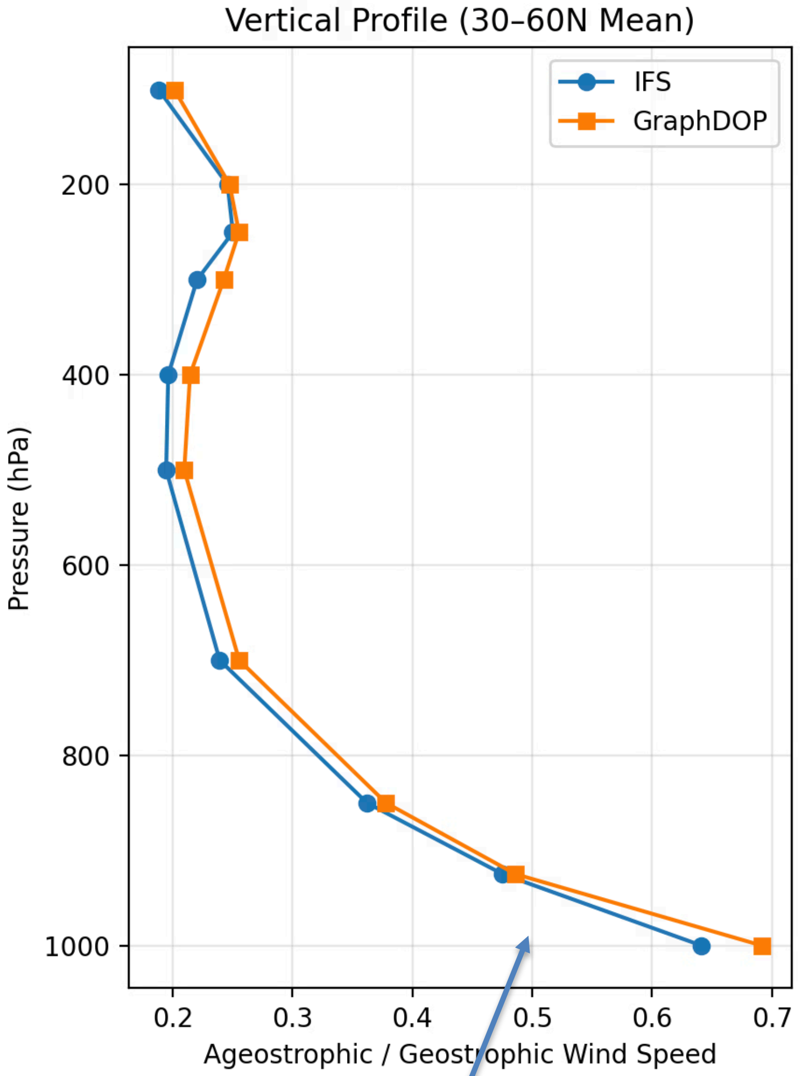
Balance relationships



u-wind
200hPa
DOP prediction

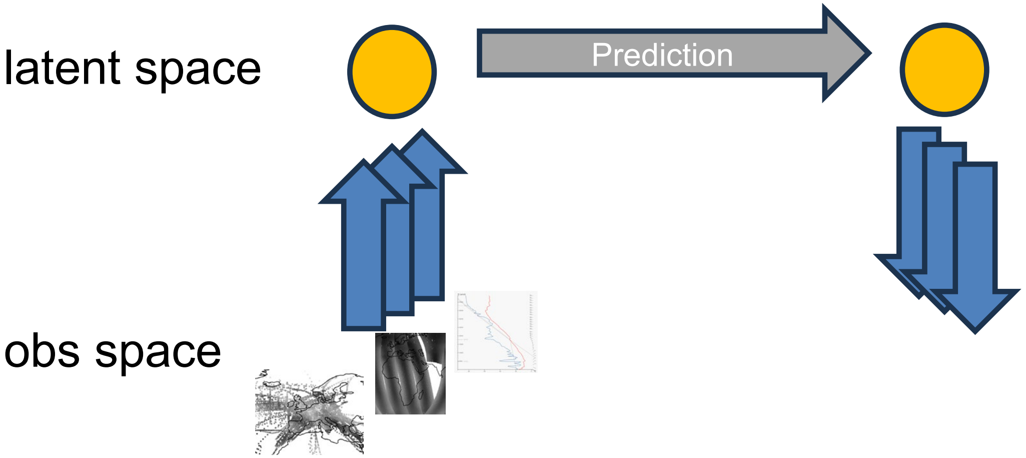


Geostrophic
u-wind
200hPa
derived from
geopotential
prediction



Increased ageostrophic flow near the surface

A **learned forecast model** which evolves
the latent state forward in time

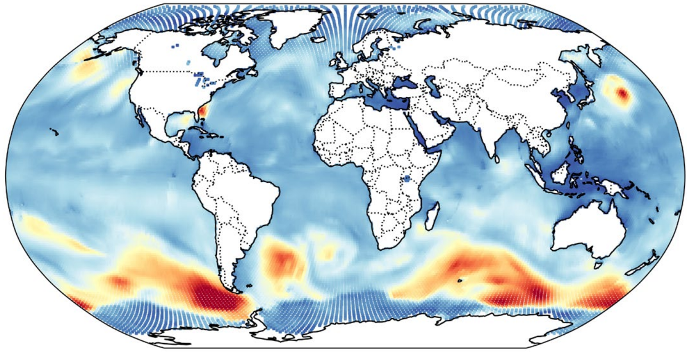


A **learned "DA" function**
which combines information
from different sources
into a (learned) set of
unified latent state
variables

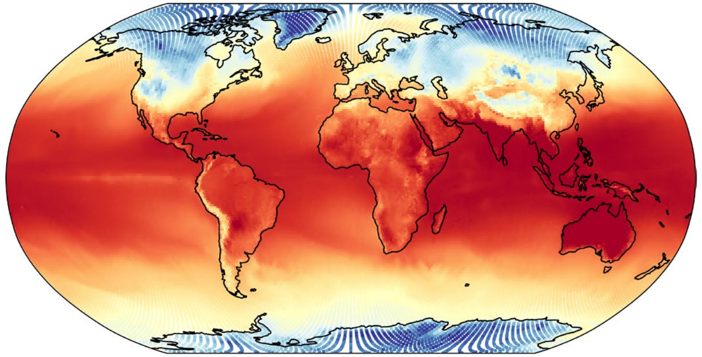
Learned observation operators
which map from the network's internal
model latent state
to observation space

Learning an observation operator; latent state → observation space

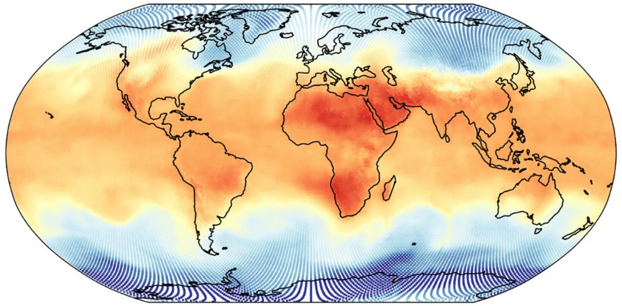
- For any observed variable/channel



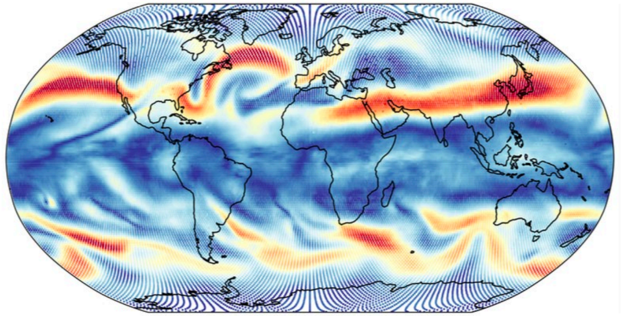
Significant wave height



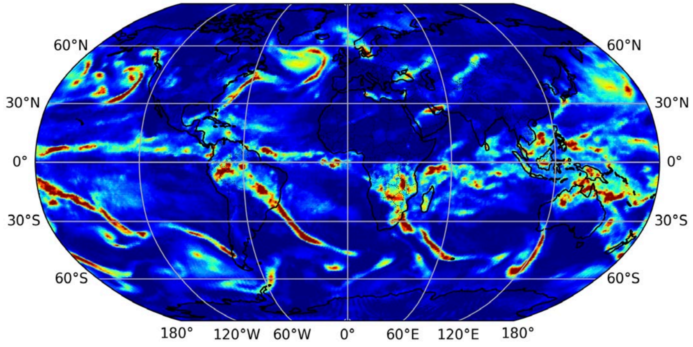
2-meter temperature



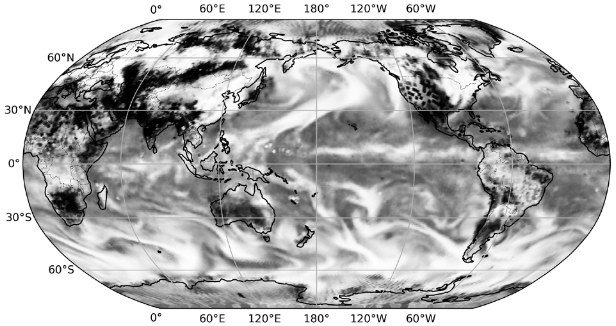
850hPa temperature



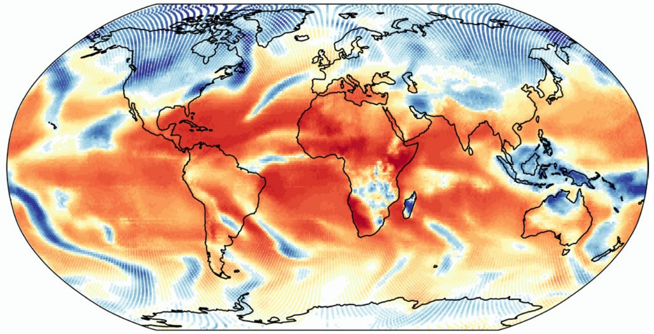
200hPa winds



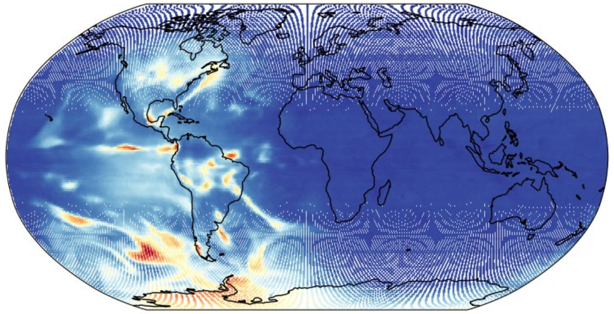
1 hour precipitation accumulation



Cloud fraction

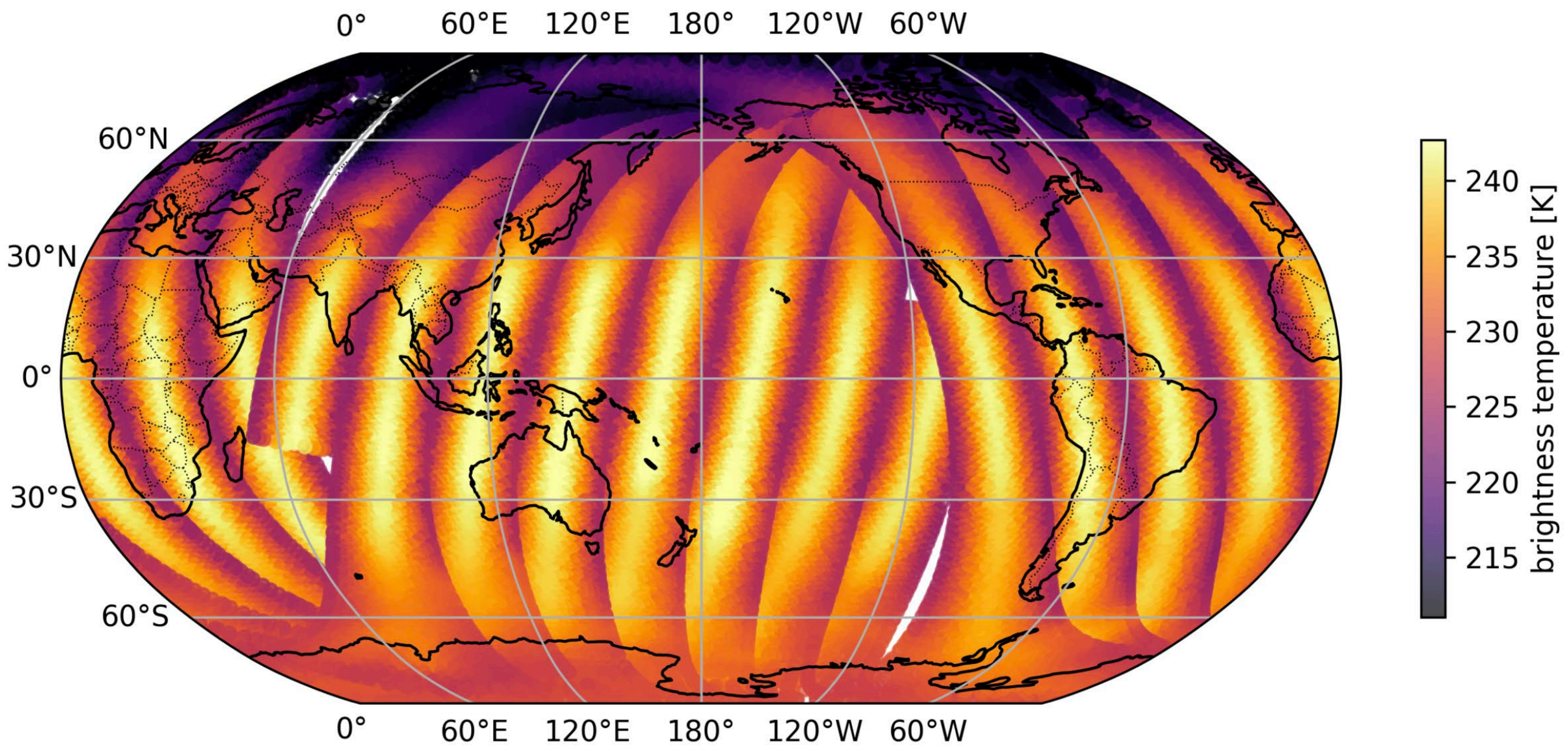


SEVIRI infrared window channel



AVHRR visible channel

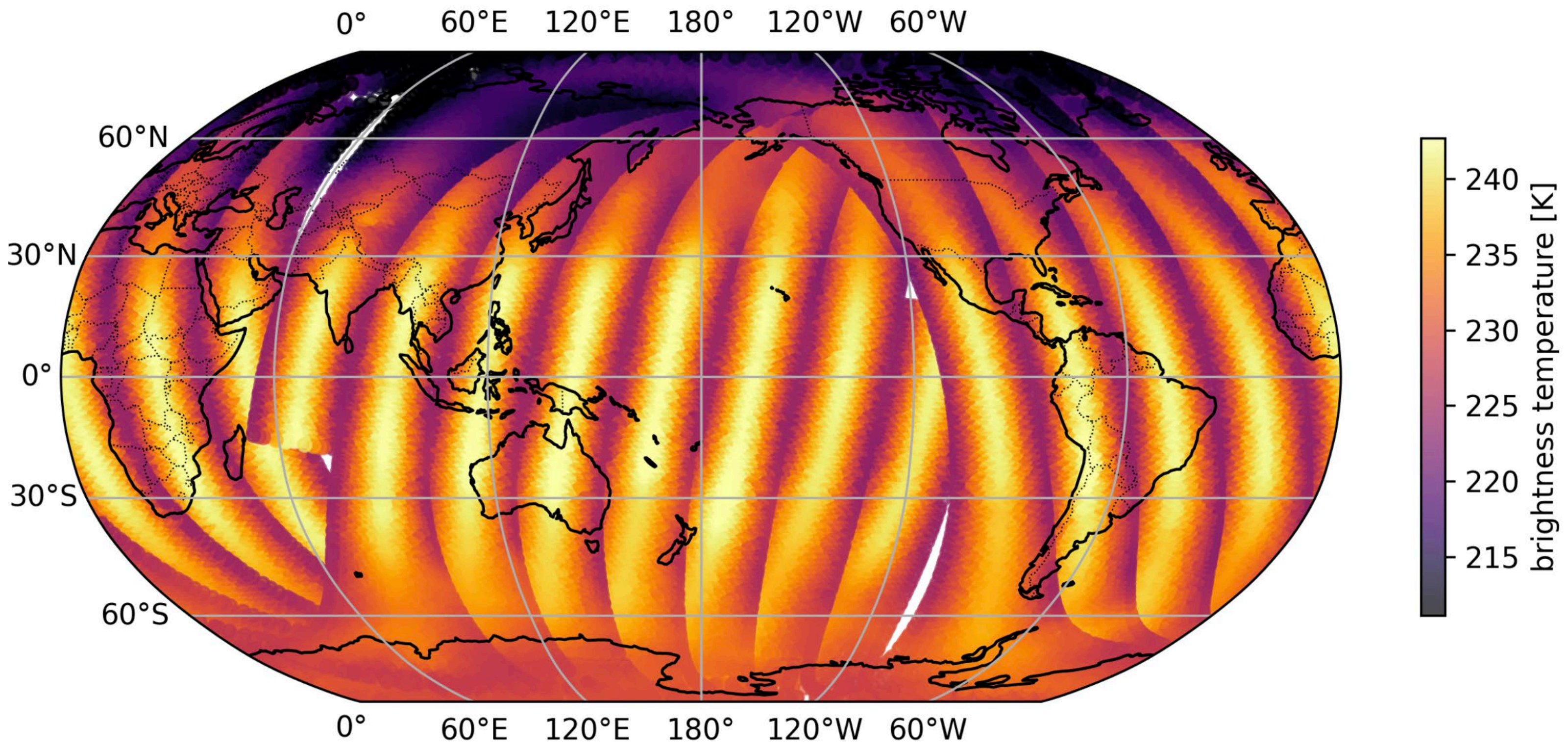
Learning an observation operator; latent state → observation space: Representation of limb effects



Target

NPP ATMS ch7

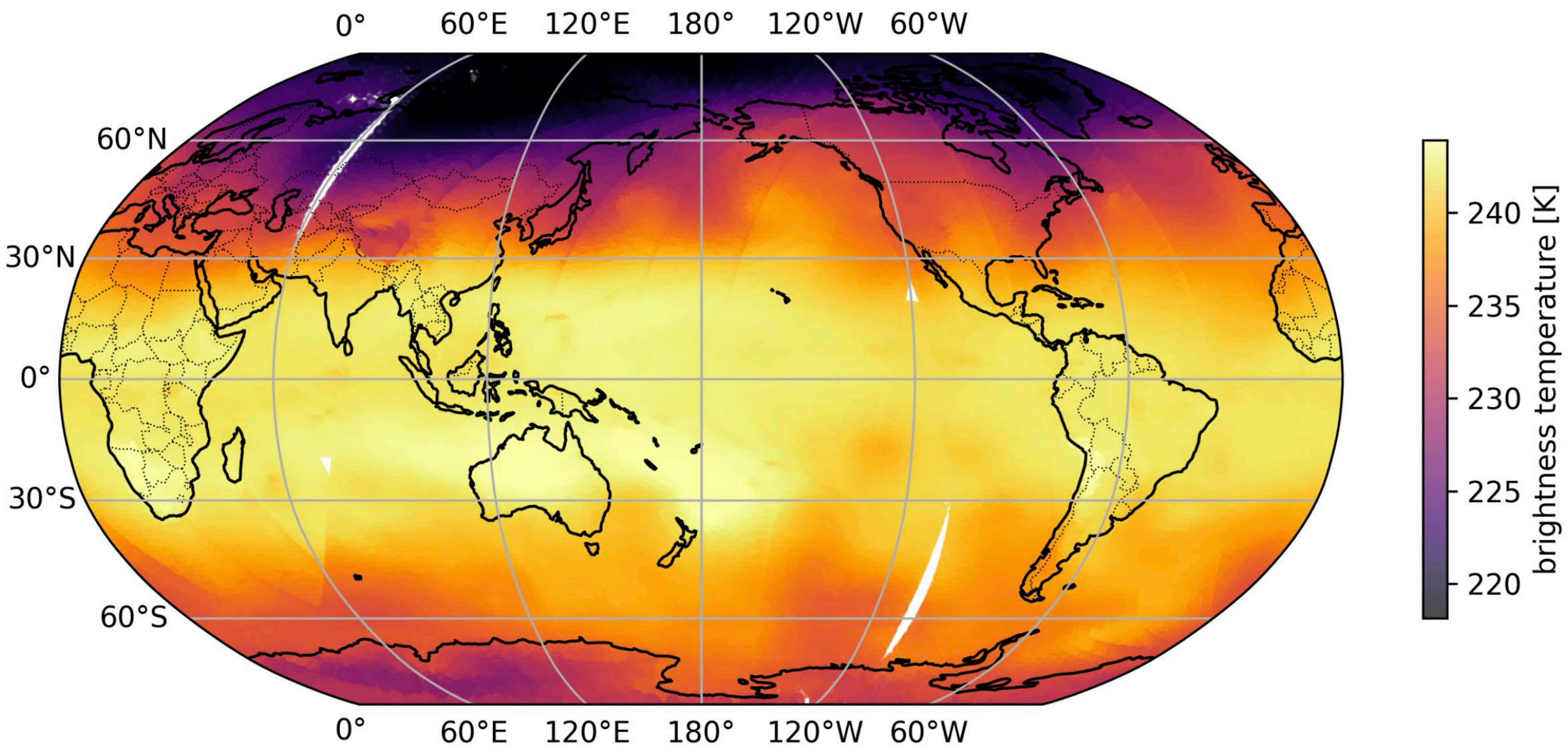
Learning an observation operator; latent state → observation space: Representation of limb effects



Prediction

NPP ATMS ch7

Learning an observation operator; latent state → observation space: Representation of limb effects



NPP ATMS ch7

Prediction **with zenith angle set to zero**

→ **Evidence that it has a learned a representation (in some form) of limb effects**

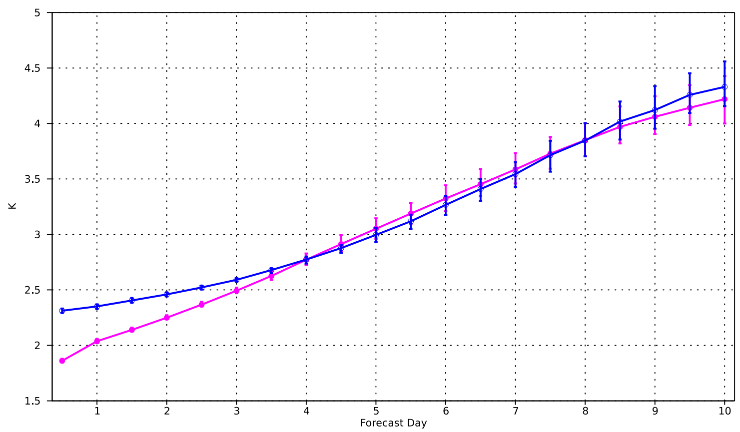
GraphDOP: most recent scores

June 2022, NHem, RMSE against observations

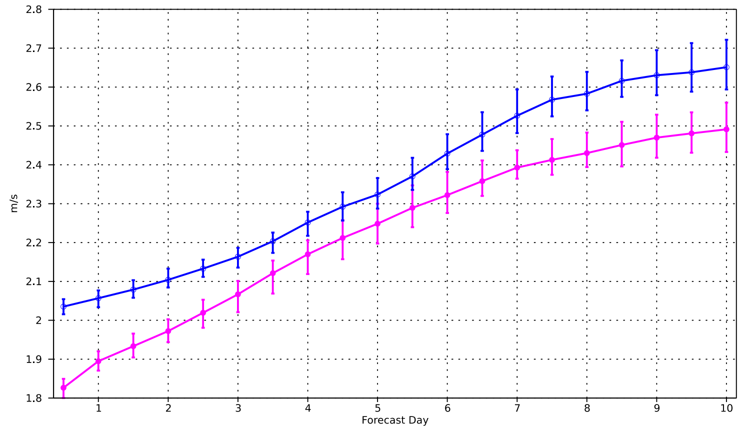
IFS

GraphDOP

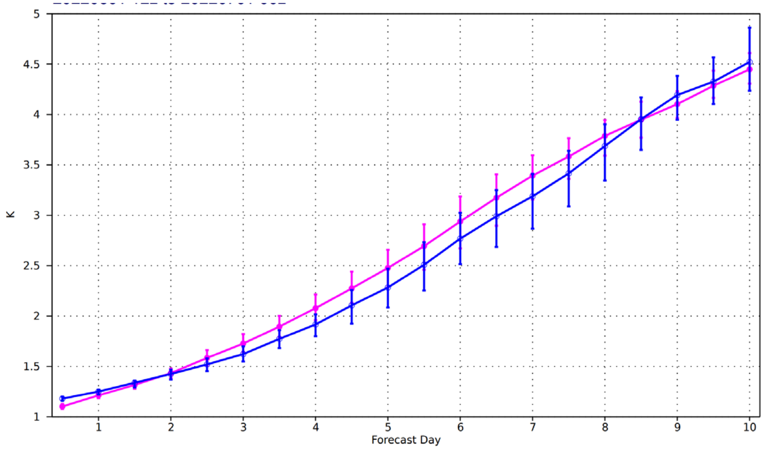
2m temperature



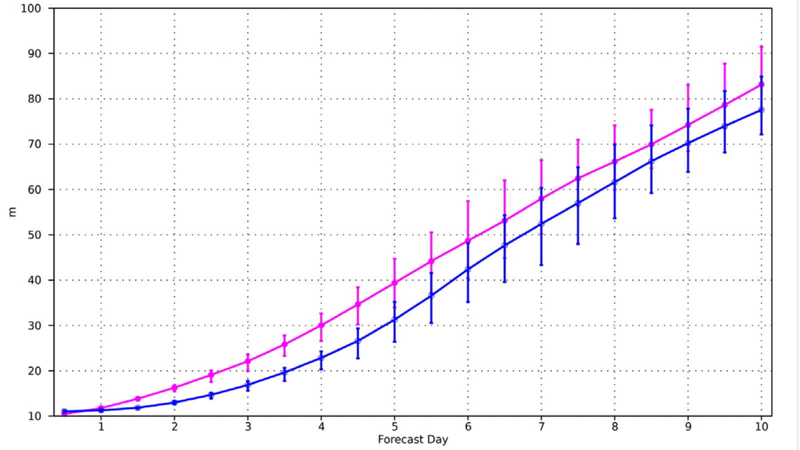
10m wind speed



Temperature at 850hPa



Geopotential height at 500hPa



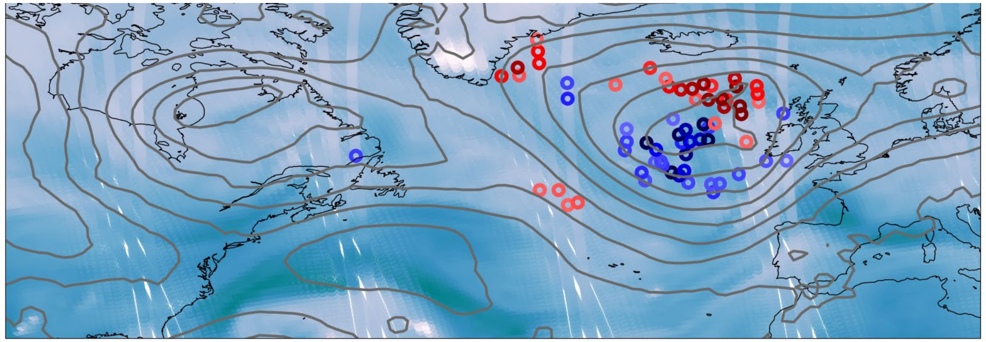
- Especially good for surface parameters (where near-surface conditions are constrained directly by the observations)
- Upper air is still lagging but now GraphDOP beating IFS at t+12h
- Errors still grow more rapidly than in traditional approaches → work in progress on the rollout strategy

Using DA tools to understand GraphDOP

Backpropagation used in training can be extended to compute sensitivities of inputs to the forecast

➔ adjoint of GraphDOP

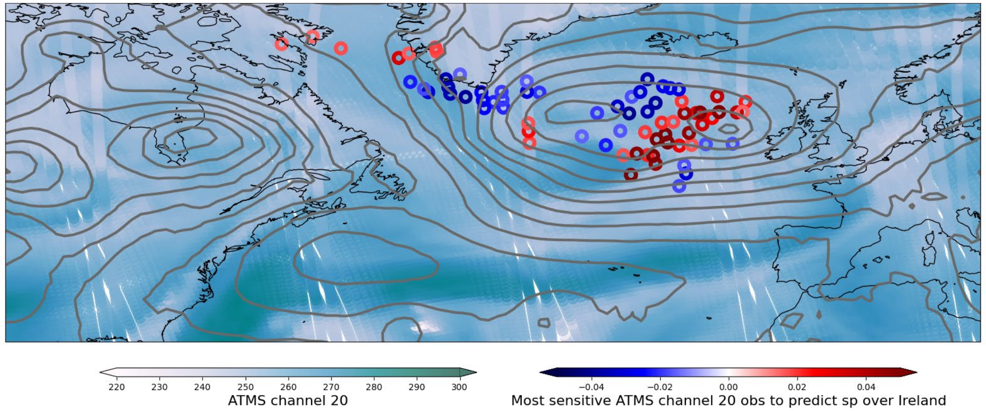
ATMS channel 20 DOP input (24h before landfall)



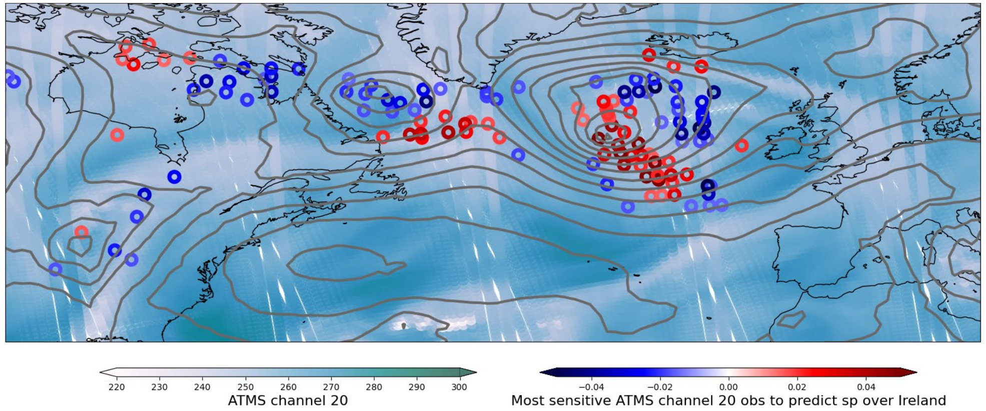
Which observations were the most critical to predict a storm hitting Ireland on the 8th Nov 2022?

Let's run the adjoint backwards in time

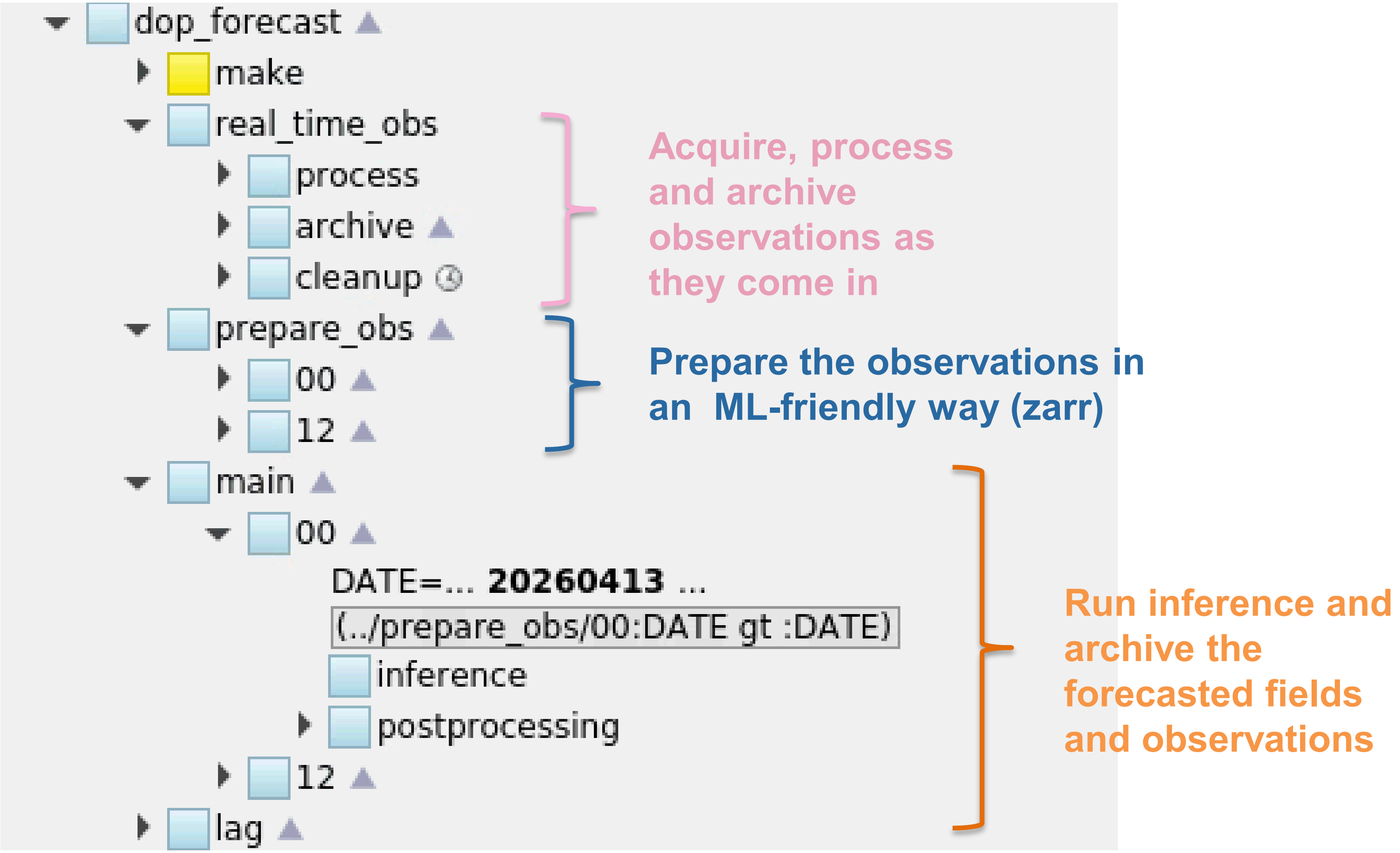
ATMS channel 20 DOP input (48h before landfall)



ATMS channel 20 DOP input (72h before landfall)



Prototype near-real-time system

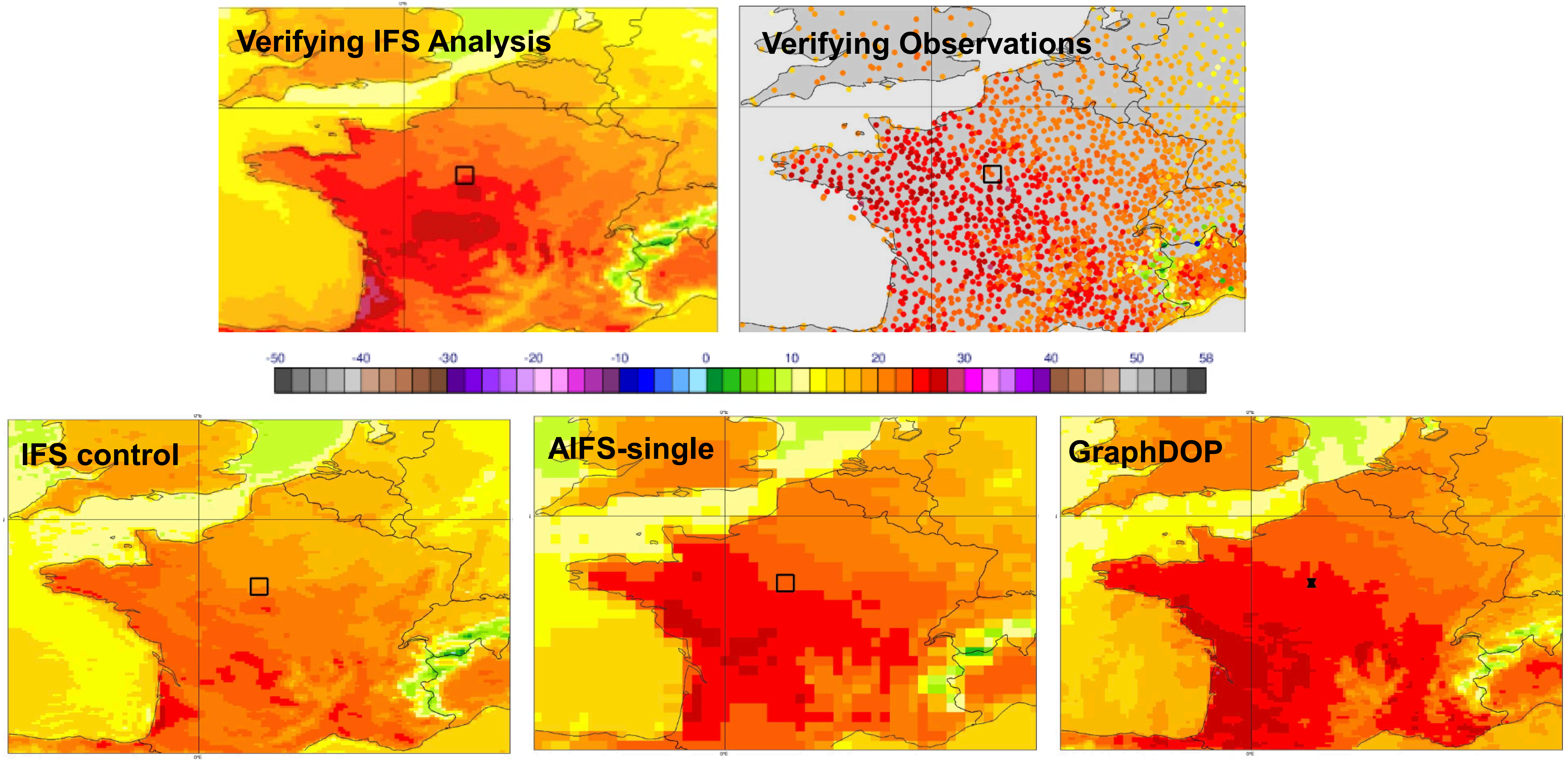


- Starting to tackle a real-time end-to-end system and all the challenges that come alongside this
  - Quality control of the observations
  - Deciding where to forecast the future observations as we don't know their locations in advance
  - Handling the loss of an instrument/channel
  - Integrating new satellites
  - Etc…
- Forecast delivered **10 minutes after observation cut-off** (still room for improvement)

Prototype near-real-time system

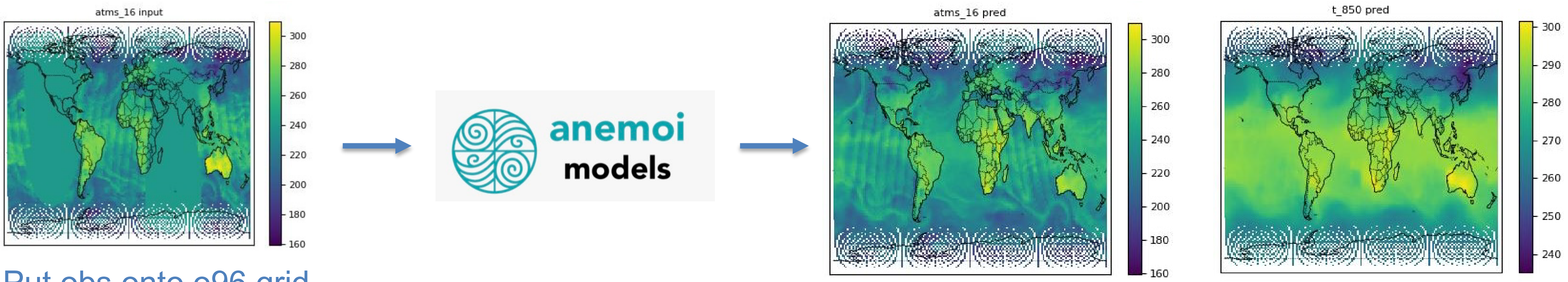
2-metre temperature in 84h forecasts valid 8 April 12UTC

Linus Magnusson



Integration into Anemoi

- Work underway to integrate observations into Anemoi!
  - Anemoi Datasets
    - Creation and loading of tabular observation zarrs from GraphDOP is being implemented
  - Anemoi Models
    - “multiple datasets” branch now merged into main, allowing for training a model with datasets at different time and spatial resolutions
    - Dynamic graphs from GraphDOP is being prototyped
- A first example of training from observations with Anemoi using gridded observations (Input = two 6-hour grids, Output = the next 6-hour grid)



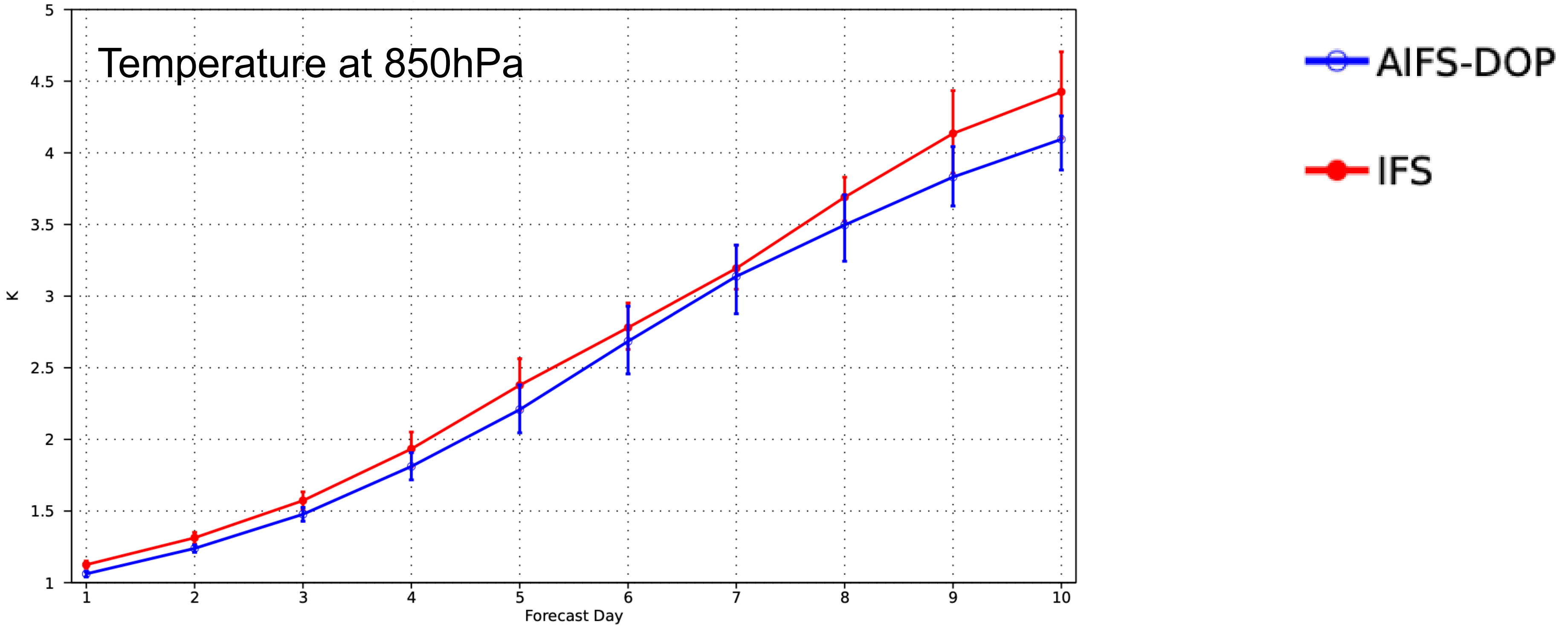
Put obs onto o96 grid
Add nans for missing data

Integration into Anemoi

anemoi

- Promising results **especially to constrain the error growth** in the rollout despite a very depleted set of o96 observations

June 2021, NHem, RMSE against observations



Summary and next steps

- **The AI-DOP project at ECMWF is exploring the extent to which fully observation-driven end-to-end weather forecasting can challenge the state of the art**
  - Initial results indicate that the concept - *learning a model from observations alone* - is working
  - First NRT system up and running daily
- **Verification scores continue to improve:**
  - Surface verification scores are already competitive (especially in the short range)
  - Upper air scores are more challenging (mainly in N.Hem winter)
  - Promising preliminary results from AIFS-DOP
- **Next steps:**
  - Continued Anemoi integration
  - Using a longer input window through background cycling
  - CRPS training
  - Strategies to handle changing observing system; systematic errors etc

Further details in our Arxiv preprints:

arXiv:2407.15586v1 [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 new and radical approach to weather forecasting, by training a neural network to predict future weather purely from historical observations with no dependence on a physics-based model or reanalysis datasets. We use raw observations (level-1) to initialise a model of the atmosphere (in observation space) learned directly from the observations themselves. Forecasts of crucial weather-related parameters (such as surface temperature and wind) are obtained by predicting weather parameter observations (such as SYNOP surface data) at future times and arbitrary locations. We present preliminary results on forecasting observations 12-hours into the future. These already demonstrate successful learning of the time evolution of the physical processes captured in real observations. We argue that this new approach, by staying purely in observation

arXiv:2412.15687v1 [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 23, 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.

A (re)analysis is the product of data assimilation, a family of algorithms that aim to optimally combine the best available estimate of the current global atmospheric state - e.g., a previous short-range forecast from a physics-based weather model - with information obtained from Earth System observations. For example, the ECMWF runs four-dimensional variational data assimilation (4D-Var; see, e.g., Rabier et al. [2000]) to produce high-resolution gridded analyses (currently ca. 9 km along the horizontal and 137 vertical levels) that are then used to initialise the medium-range

18018v1 [physics.ao-ph] 25 Aug 2025

Learning from nature: insights into GraphDOP's representations of the Earth System

A Preprint

Peter Lean Mihai Alexe Eulalie Boucher Ewan Pinnington Simon Lang
Patrick Laloyaux Niels Bormann Anthony McNally
European Centre for Medium-Range Weather Forecasts (ECMWF)

August 26, 2025

Abstract

Through a series of experiments, we provide evidence that the GraphDOP model - trained solely on meteorological observations, using no prior knowledge - develops internal representations of the Earth System state, structure and dynamics as well as the characteristics of different observing systems. Firstly, we demonstrate that the network constructs a unified latent representation of the Earth System state which is common across different observation types. For example, cloud structures maintain physical consistency whether viewed in predictions for satellite radiances from different sensors, or for direct in-situ measurements of the cloud fraction. Secondly, we show examples that suggest that the network learns to emulate viewing effects - learned observation operators that map from the unified state representation to observed properties. Microwave sounder limb effects and geometric viewing effects, such as sunglint in visible imagery, are both well captured. Finally, we demonstrate that the model develops rich internal representations of the structure of meteorological systems and their dynamics. For instance, when the network is only provided with observations from a single infrared instrument, it is able to infer unobserved, non-local structures such as jet streams, surface pressure patterns and warm and cold air masses associated with synoptic systems. This work provides insights into how neural networks trained solely on observations of the Earth System spontaneously develop coherent internal representations of the physical world in order to meet the training objective

PEAN C

20416v1 [physics.ao-ph] 23 Oct 2025

Learning Coupled Earth System Dynamics with GraphDOP

Eulalie Boucher Mihai Alexe Peter Lean
Ewan Pinnington Simon Lang Patrick Laloyaux Lorenzo Zampieri
Patricia de Rosnay Niels Bormann Anthony McNally
European Centre for Medium-Range Weather Forecasts (ECMWF)

October 24, 2025

Abstract

Interactions between different components of the Earth System (e.g. ocean, atmosphere, land and cryosphere) are a crucial driver of global weather patterns. Modern Numerical Weather Prediction (NWP) systems typically run separate models of the different components, explicitly coupled across their interfaces to additionally model exchanges between the different components. Accurately representing these coupled interactions remains a major scientific and technical challenge of weather forecasting. GraphDOP is a graph-based machine learning model that learns to forecast weather directly from raw satellite and in-situ observations, without reliance on reanalysis products or traditional physics-based NWP models. GraphDOP simultaneously embeds information from diverse observation sources spanning the full Earth system into a shared latent space. This enables predictions that implicitly capture cross-domain interactions in a single model without the need for any explicit coupling. Here we present a selection of case studies which illustrate the capability of GraphDOP to forecast events where coupled processes play a particularly key role. These include rapid sea-ice freezing in the Arctic, mixing-induced ocean surface cooling during Hurricane Ian and the

Using data assimilation tools to dissect GraphDOP

A Preprint

Patrick Laloyaux Mihai Alexe Eulalie Boucher Peter Lean Ewan Pinnington
Simon Lang Tobias Necker Anthony McNally
European Centre for Medium-Range Weather Forecasts (ECMWF)

October 28, 2025

Abstract

The Data Assimilation (DA) community has been developing various diagnostics to understand the importance of the observing system in accurately forecasting the weather. They usually rely on the ability to compute the derivatives of the physical model output with respect to its initial condition. For example, the Forecast Sensitivity-based Observation Impact (FSOI) estimates the impact on the forecast error of each observation processed in the DA system. This paper presents how these DA diagnostic tools are transferred to Machine Learning (ML) models, as their derivatives are readily available through automatic differentiation. We specifically explore the interpretability and explainability of the observation-driven GraphDOP model developed at the European Centre for Medium-Range Weather Forecasts (ECMWF). The interpretability study demonstrates the ef-