Machine Learning with satellite data for NWP

Peter Lean

With contributions from:

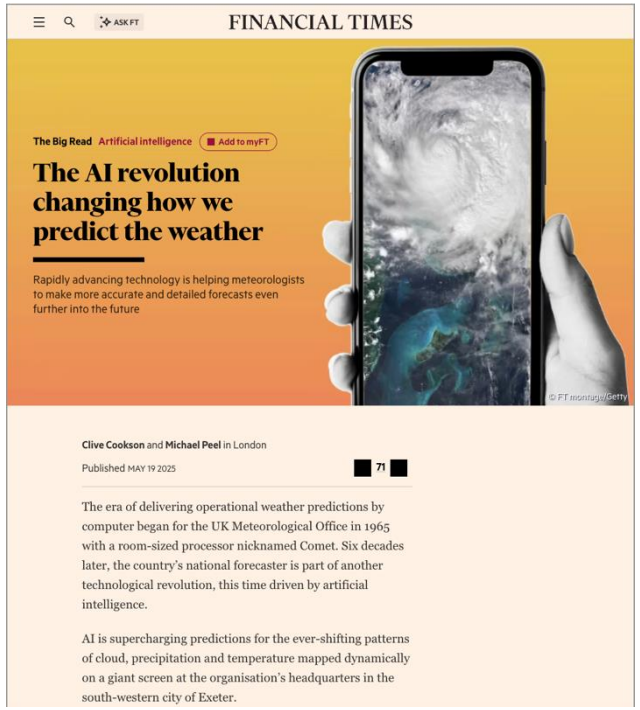
Sebastien Garrigues, Giovanna De Chiara, Alan Geer, Alban Farci, Marcin Chrust,

Mihai Alexe, Eulalie Boucher, Ewan Pinnington, Simon Lang, Patrick Laloyaux, Mat Chantry, Tony McNally

NWP SAF Training course

ECMWF

AI revolution in weather prediction



FINANCIAL TIMES

The Big Read Artificial intelligence

The AI revolution changing how we predict the weather

Rapidly advancing technology is helping meteorologists to make more accurate and detailed forecasts even further into the future

Clive Cookson and Michael Peel in London

Published MAY 19 2025

The era of delivering operational weather predictions by computer began for the UK Meteorological Office in 1965 with a room-sized processor nicknamed Comet. Six decades later, the country's national forecaster is part of another technological revolution, this time driven by artificial intelligence.

AI is supercharging predictions for the ever-shifting patterns of cloud, precipitation and temperature mapped dynamically on a giant screen at the organisation's headquarters in the south-western city of Exeter.

The Washington Post
Democracy Dies in Darkness

Weather Extreme Weather Climate Capital Weather Gang Environment Climate Lab

AI is transforming weather forecasting. Is the U.S. falling behind?

AI weather models have burst onto the scene during this 2024 hurricane season, with accurate forecasts days in advance.

October 25, 2024 More than 1 year ago

7 min 114 Make us preferred on Google

A weather simulation shown
Germany on Sept. 25. (Marcu

The Guardian

News Opinion Sport Culture Lifestyle

UK US politics World Climate crisis Middle East Ukraine Football Newsletters Business Environment UK politics Science Tech Global development Obituaries

Science

This article is more than 1 year old

Google DeepMind predicts weather more accurately than leading system

Ian Sample Science editor

Wed 4 Dec 2024 16.00 GMT

Share

Prefer the Guardian on Google

BAMS Article

The Rise of Data-Driven Weather Forecasting

A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context

Zied Ben Bouallègue,[a] Mariana C. A. Clare,[b] Linus Magnusson,[a] Estibaliz Gascón,[b] Michael Maier-Gerber,[b] Martin Janoušek,[b] Mark Rodwell,[a] Florian Pinault,[b] Jesper S. Dramsch,[b] Simon T. K. Lang,[a] Baudouin Raoult,[a] Florence Rabier,[a] Matthieu Chevallier,[a] Irina Sandu,[b] Peter Dueben,[b] Matthew Chantry,[a] and Florian Pappenberger[a]

KEYWORDS: Forecast verification/skill; Operational forecasting; Machine learning

ABSTRACT: Data-driven modeling based on machine learning (ML) is showing enormous potential for weather forecasting. Rapid progress has been made with impressive results for some applications. The uptake of ML methods could be a game changer for the incremental progress in traditional numerical weather prediction (NWP) known as the "quiet revolution" of weather forecasting. The computational cost of running a forecast with standard NWP systems greatly hinders the improvements that can be made by increasing model resolution and ensemble sizes. An emerging new generation of ML models, developed using high-quality reanalysis datasets like ERA5 for training, allows forecasts that require much lower computational costs and that are highly competitive in terms of accuracy. Here, we compare for the first time ML-generated forecasts with standard NWP-based forecasts in an operational-like context, initialized from the same initial conditions. Focusing on deterministic forecasts, we apply common forecast verification tools to assess to what extent a data-driven forecast produced with one of the recently developed ML models (PanguWeather) matches the quality and attributes of a forecast from one of the leading global NWP systems (the ECMWF IFS). The results are very promising, with comparable accuracy for both global metrics and extreme events, when verified against both the operational IFS analysis and synoptic observations. Overly smooth forecasts, increasing bias with forecast lead time, and poor performance in predicting tropical cyclone intensity are identified as current drawbacks of ML-based forecasts. A new NWP paradigm is emerging relying on inference from ML models and state-of-the-art analysis and reanalysis datasets for forecast initialization and model training.

The Washington Post
Democracy Dies in Darkness

Weather Extreme Weather Climate Capital Weather Gang Environment Climate Lab

How Big Tech AI models nailed forecast for Hurricane Lee a week in advance

U.S. and European weather agencies are escalating their engagement with artificial intelligence as the technology rapidly advances

September 21, 2023 More than 2 years ago

7 min 26 Make us preferred on Google

A surfer stands in the surf after the remnants of Hurricane Lee passed by Nahant, Mass., on Saturday. (C) Gunther/EPA-EFE/Shutterstock)

An exciting time to be working in meteorology!

ECMWF

Strategy to embed Machine Learning deeply into the ECMWF operational chain

Jan 2021
Machine Learning Roadmap

Deepmind – GraphCast
0.25° 6-hour

Many variables and pressure levels with comparable skill to IFS.

Dec 2022
Extensive predictions

ECMWF's
Peter Dueben and Peter Bauer publish a paper on using ERA5 at ~500km_resolution to predict future z500.

Feb 2022
Full medium-range NWP

Keisler - GraphNN 1°, competitive with GFS

NVIDIA – FourCastNet
Fourier+ , 0.25°

O(10^4) faster & more energy efficient than IFS

Huawei – PanguWeather
0.25° hourly product

"More accurate tracks" than the IFS.

ECMWF

Strategy to embed Machine Learning deeply into the ECMWF operational chain

Jan 2021
Machine Learning Roadmap

Deepmind – GraphCast
0.25° 6-hour

Many variables and pressure levels with comparable skill to IFS.

Dec 2022
Extensive predictions

FengWu – China academia + Shanghai Met Bureau
0.25° 6-hour product

Improves on GraphCast for longer leadtimes (still deterministic)

Apr 2023
7-d + scores improve

AIFS – ECMWF
1° 6-hour**

Deterministic GNN

Oct 2023
Experimental Operational DDWF

ECMWF's
Peter Dueben and Peter Bauer publish a paper on using ERA5 at ~500km_resolution to predict future z500.

Feb 2022
Full medium-range NWP

Keisler - GraphNN 1°, competitive with GFS

NVIDIA – FourCastNet
Fourier+ , 0.25°

O(10⁴) faster & more energy efficient than IFS

Nov 2022
Tropical cyclones

Huawei – PanguWeather
0.25° hourly product

"More accurate tracks" than the IFS.

Jan 2023
Global & Limited Area

Microsoft – ClimaX

Forecasting various lead-times at various resolutions, both globally and regionally

Jun 2023
Diffusion modelling & Spherical harmonics

Alibaba – SwinRDM
0.25° 6-hour product
Sharp spatial features

NVIDIA – SFNO
0.25° 6-hour product
Extension of FourCastNet to Spherical harmonics, improved stability

Dec 2023
Probabilistic DDWF

Deepmind – GenGast
1° 6-hour

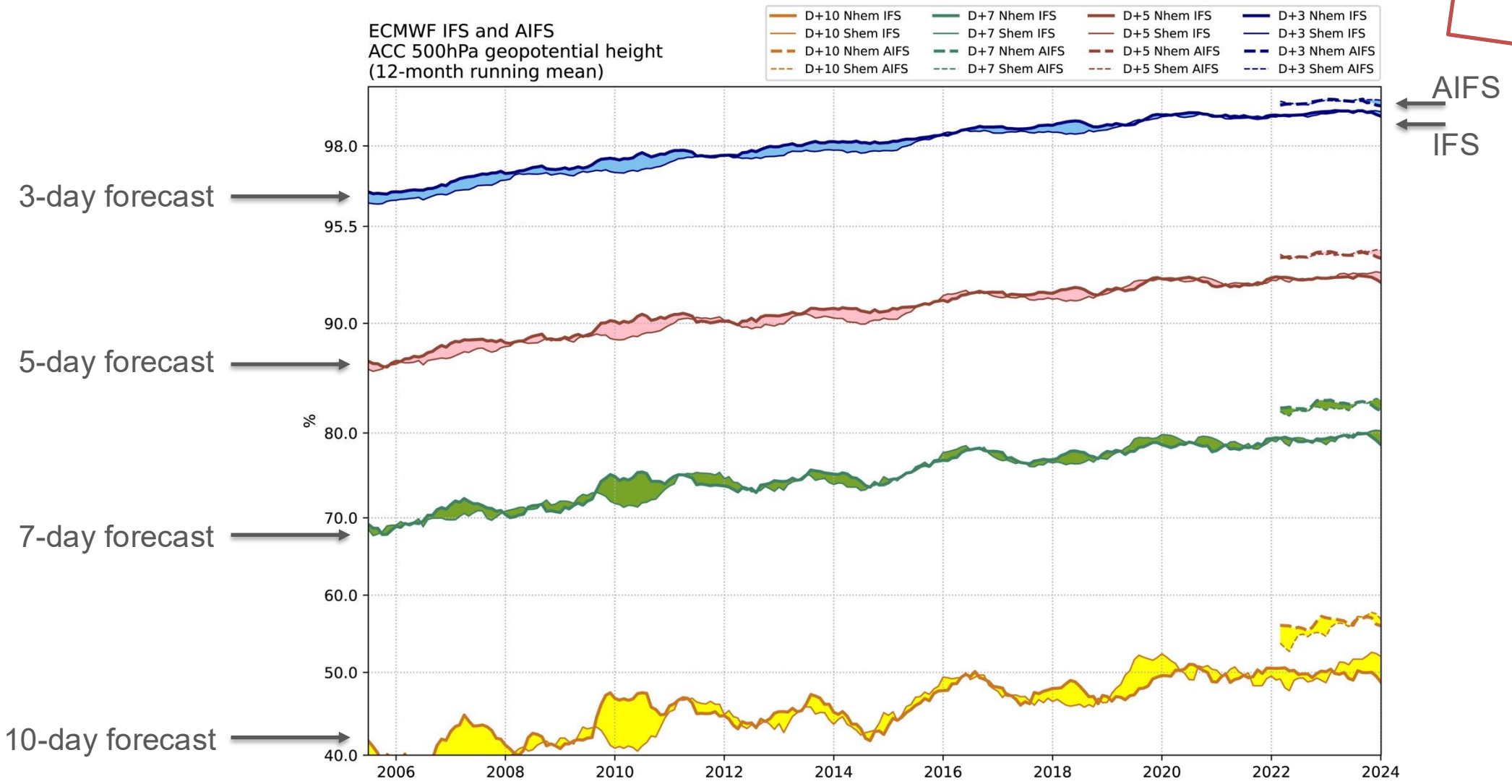
ML-based generative model for ensemble weather forecasting

Ja
AIFS

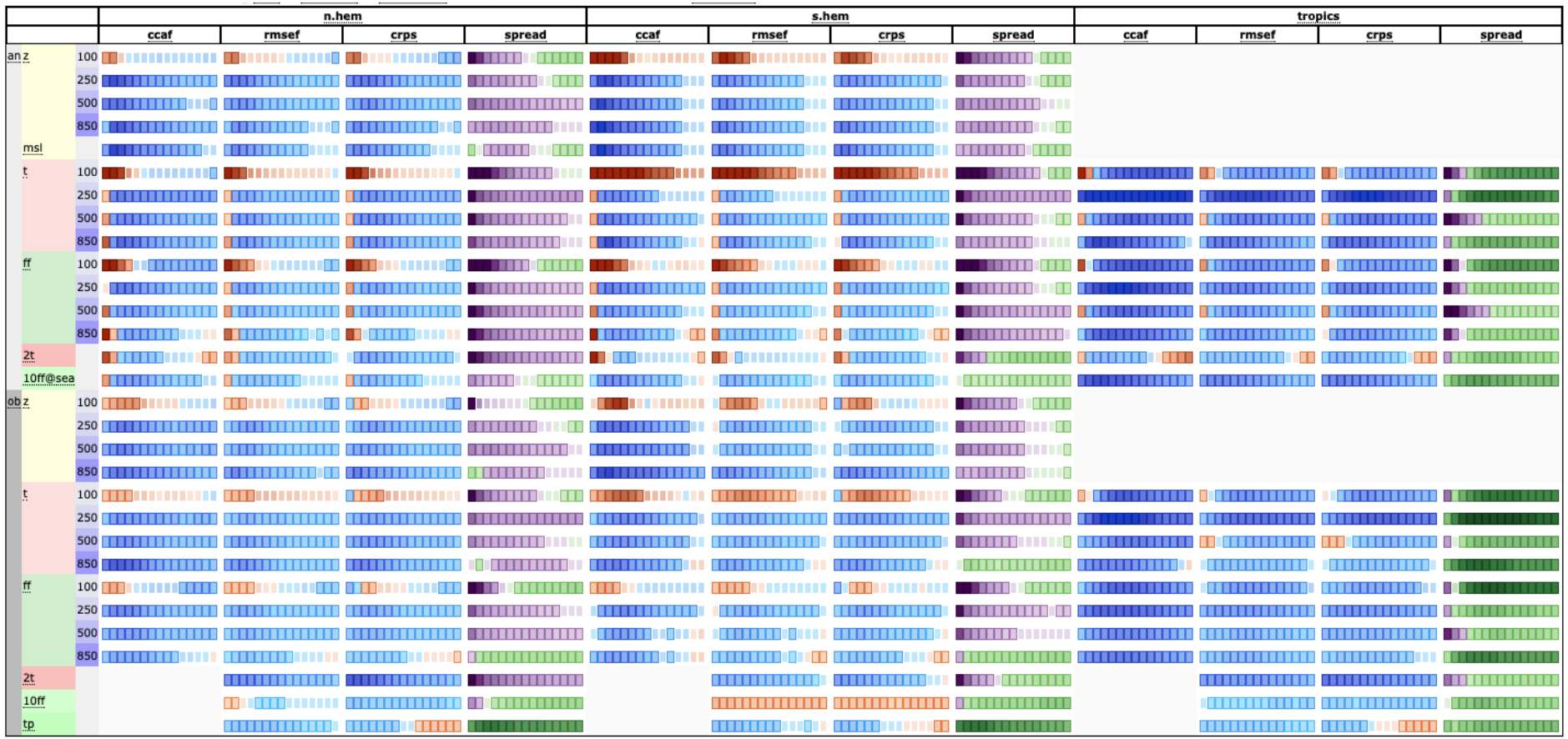
AIFS –
Push to

Verification of AIFS: Accuracy of Z500 forecasts compared to IFS

OPERATIONAL AIFS SINGLE Feb 2025



AIFS-CRPS Medium-Range evaluation, 50 member vs O1280 IFS

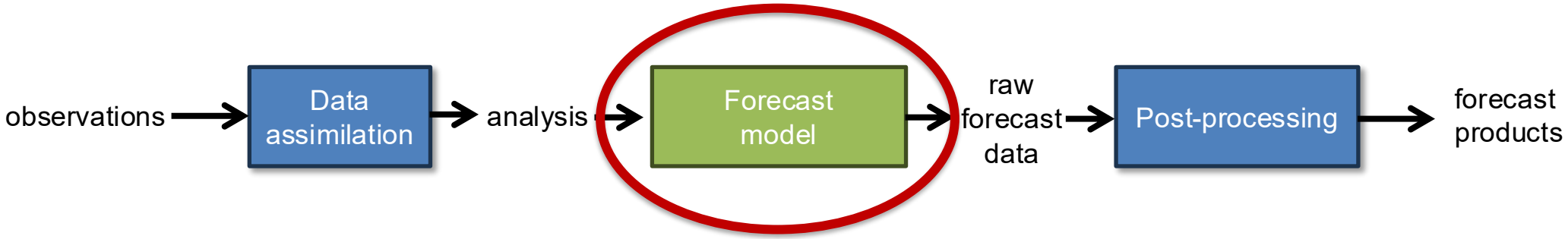


Verification: Oper An, Obs

OPERATIONAL
AIFS-CRPS
July 2025

The weather forecasting process

The weather forecasting process



The weather forecasting process

How will machine learning impact the area of observations and data assimilation?

A spectrum of possibilities

Use physical models where they are good

Use ML to learn components without a strong physical model

Use ML almost everywhere, but add physical constraints where data is lacking

Pure physics

(+ numerics, statistics + heuristics)

IFS

Pure data-driven

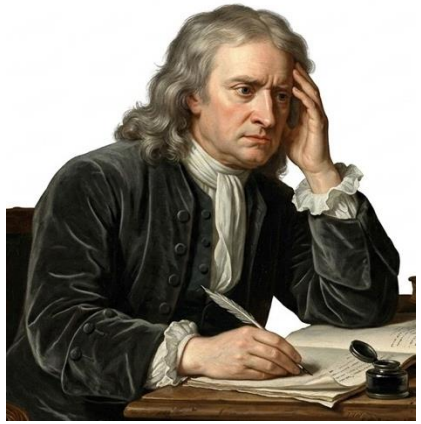
(+ inductive biases)

AI-DOP

IFS with learned observation operators

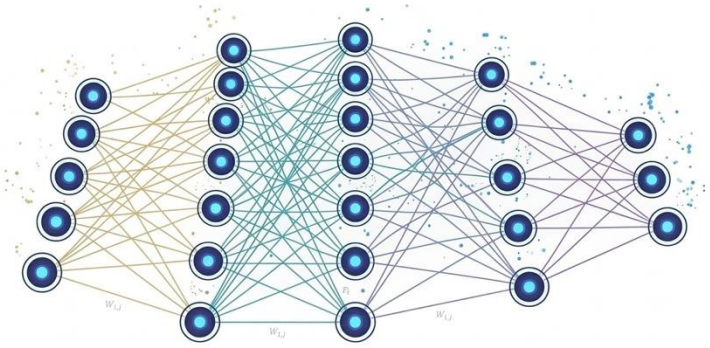
IFS with learned observation operators +learned model error correction

AI-DOP with loss terms to penalize unphysical predictions

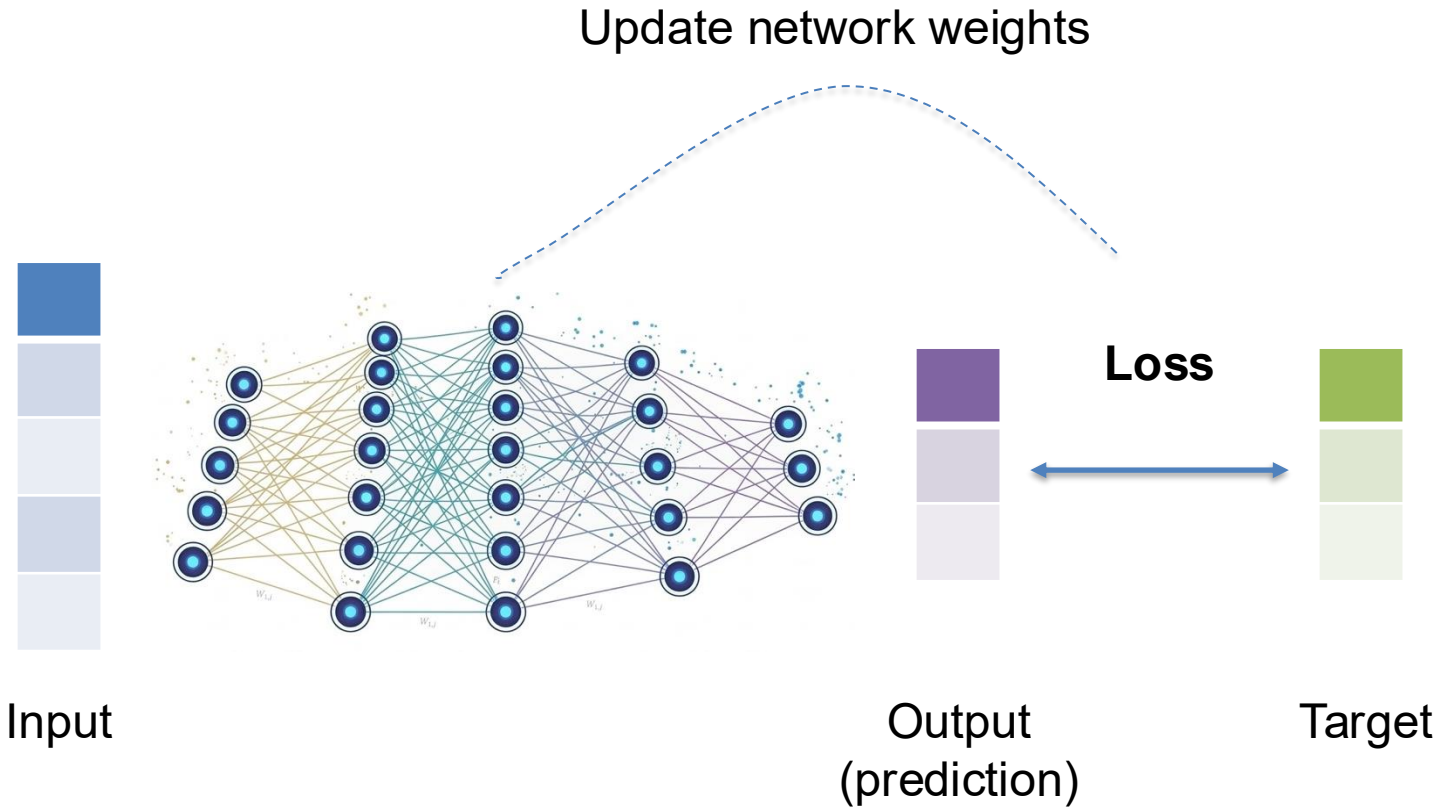


+ many possibilities

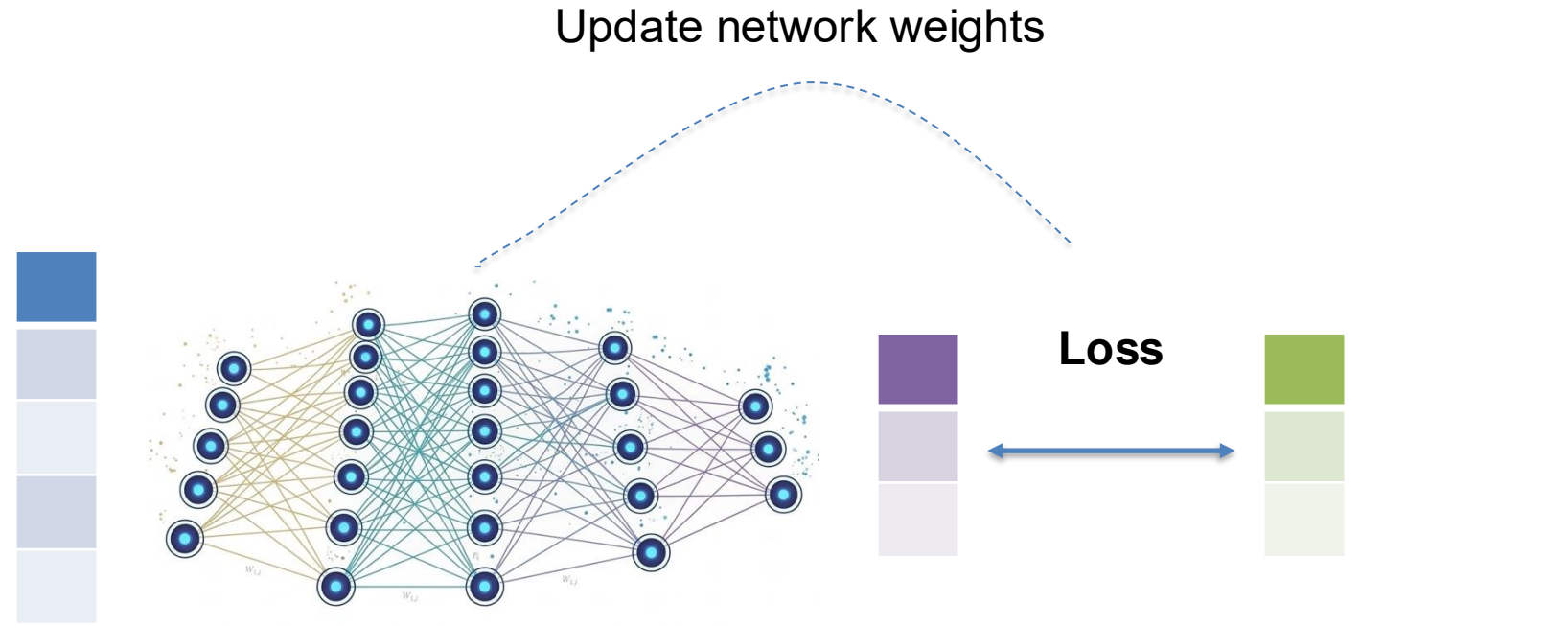
Cloud detection
Bias correction
QC / Instrument anomaly detection
…



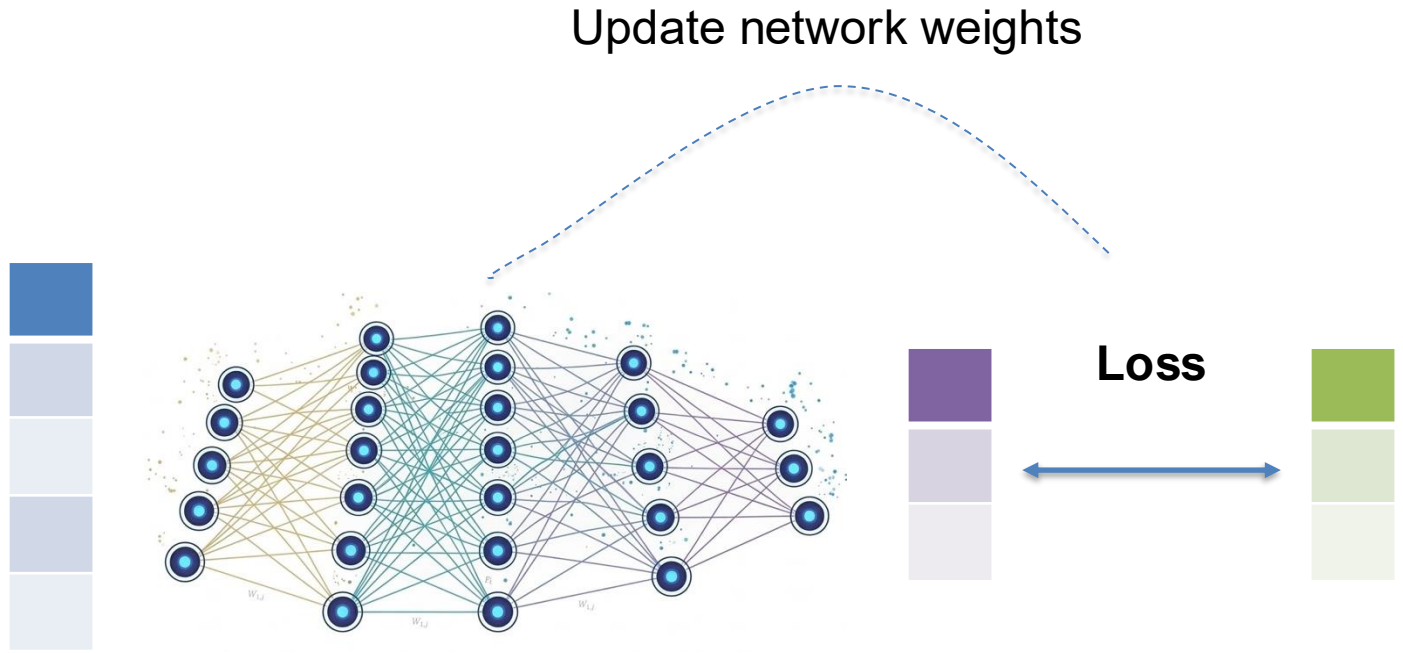
30-second primer on supervised machine learning



e.g. learn an observation operator

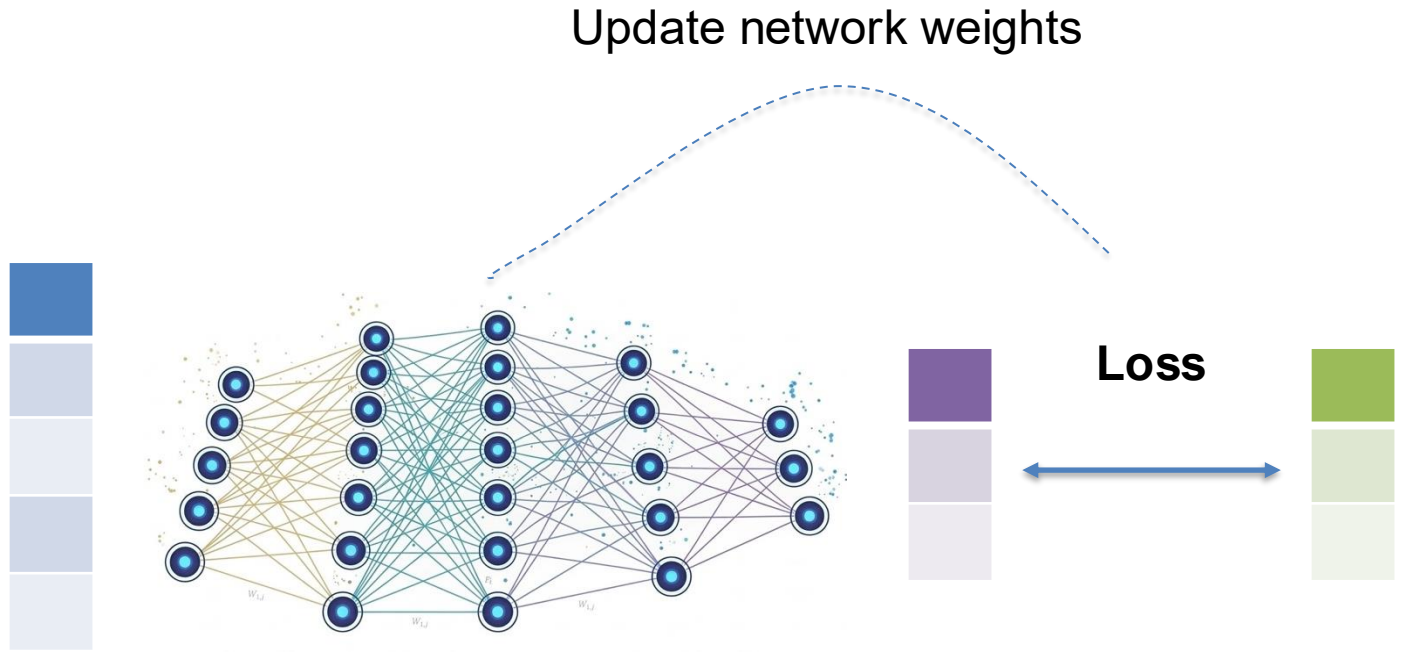


e.g. learn a forecast model that makes predictions in model space



ECMWF

e.g. learn a forecast model that makes predictions in observation space



Input

Observations in last 12h

Output (prediction)

Target

Observations in next 12h

ECMWF

Example:

Learning an observation operator for scatterometer backscatter

ECMWF

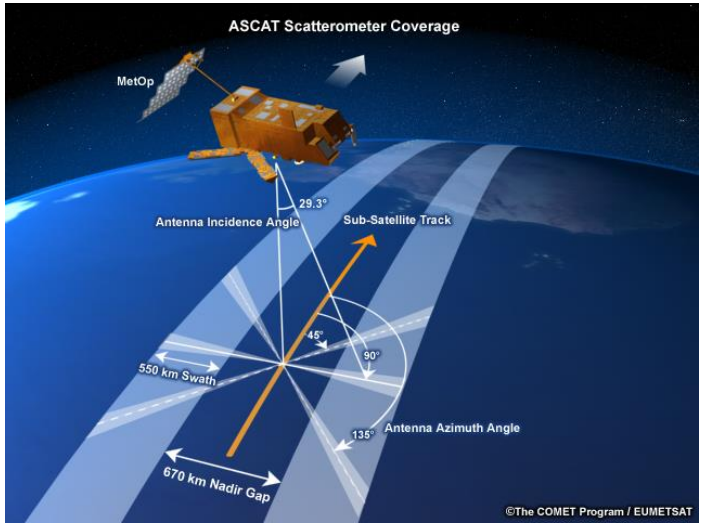
Learning an observation operator: example scatterometer sigma0

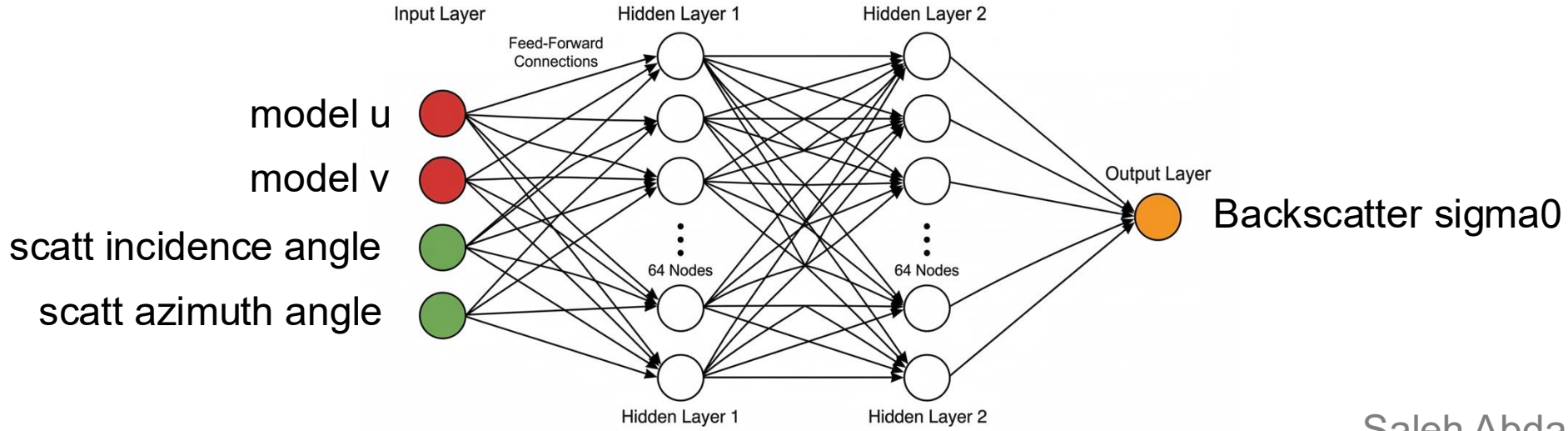
Over the ocean, scatterometer backscatter (sigma0) is sensitive to the small-scale surface roughness generated by near-surface winds.

Currently, we assimilate Level 2 scatterometer ambiguous wind u/v retrievals (for most other observation types we assimilated Level 1 data)

We don't have an observation operator for scatterometer backscatter in IFS

Can we train a neural network observation operator for scatterometer backscatter over the ocean?



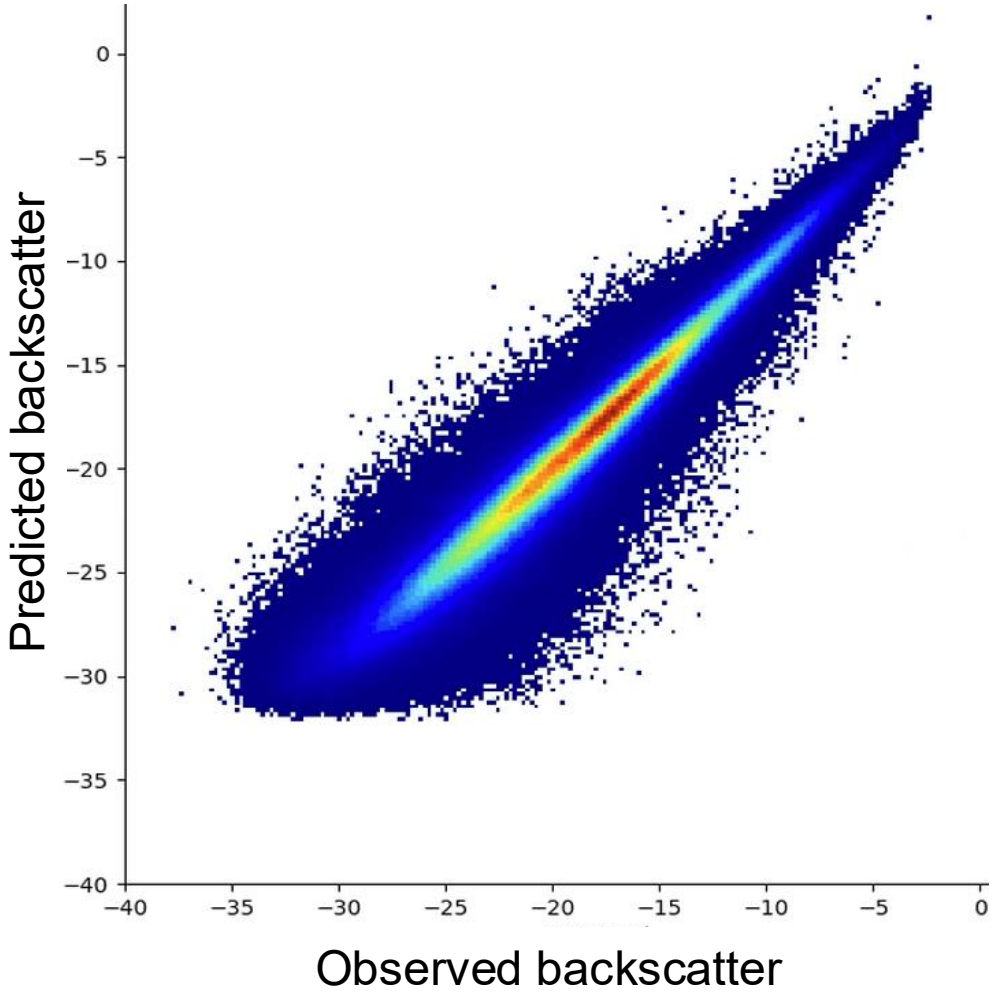


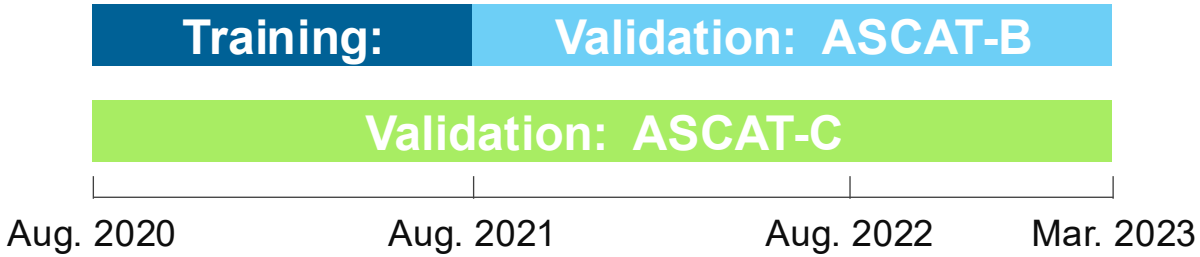
2 hidden layer
64 nodes per layer

ECMWF

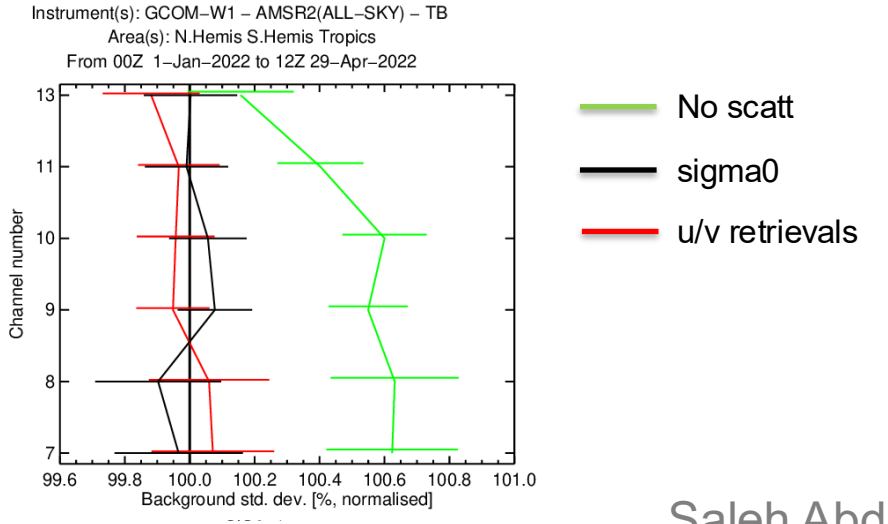
Saleh Abdalla,
Sean Healy,
Giovanna De Chiara

Learning an observation operator: example scatterometer sigma0





Results assimilating backscatter show comparable performance to assimilation of ambiguous u/v retrievals



Saleh Abdalla,
Sean Healy,
Giovanna De Chiara

Sebastien Garrigues

Learning a forward operator for backscatter over land

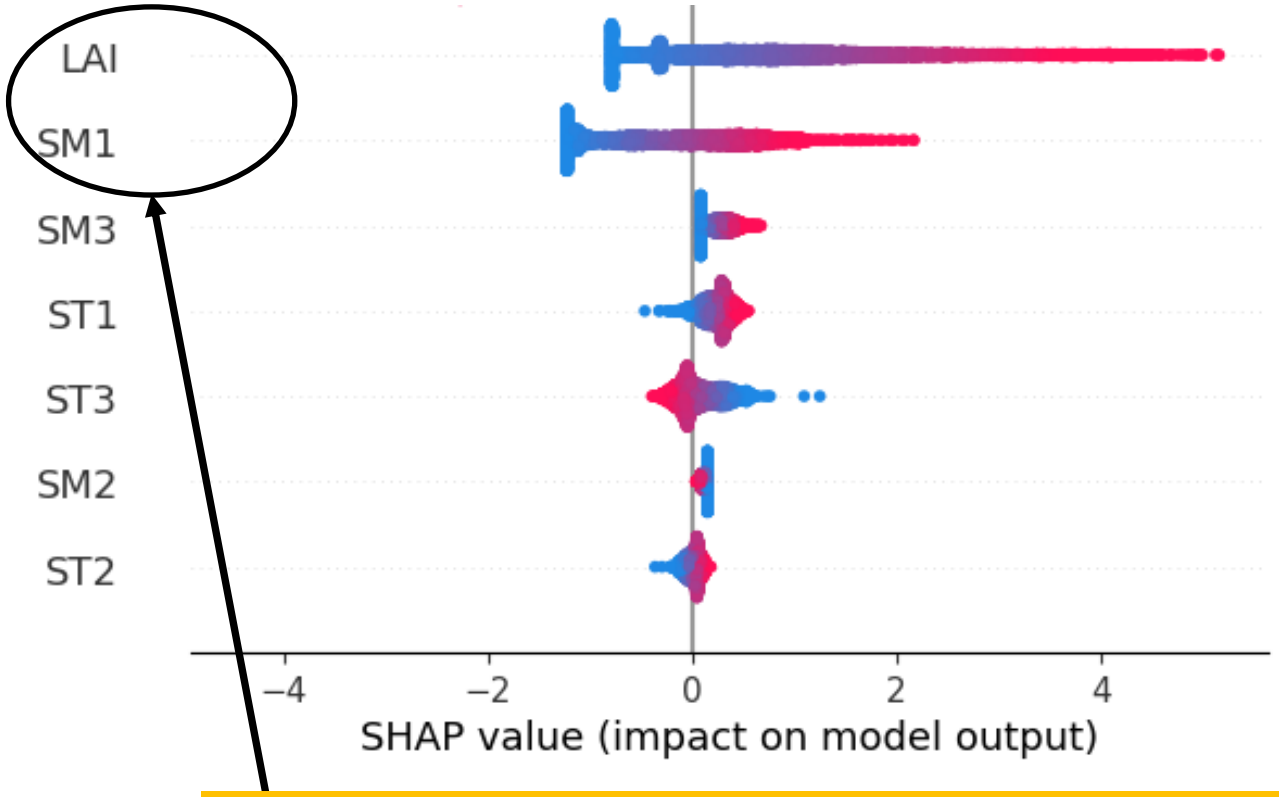
The geophysical sensitivity of scatterometer backscatter is very different over land compared to that over the ocean:

Training database (Aires, et al., QJRS 2021)

- Inputs ERA-5: Leaf Area Index (LAI), soil moisture (SM) (3 layers), soil temperature (ST) (3 layers)
- target: ASCAT backscatter normalized at 40°
- localization : latitude, longitude (sin/cos transform)
- period: 2016-2018 (training and validation), 2019 (testing)
- resolution: 0.25° grid.

Features selection

SHAP values



Vegetation (LAI) and surface soil moisture (SM1) are the most influential variables

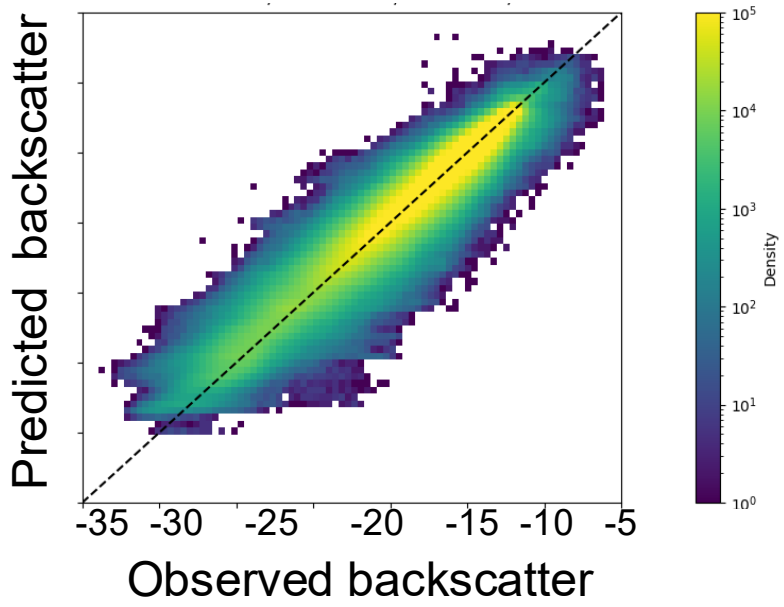
ECMWF

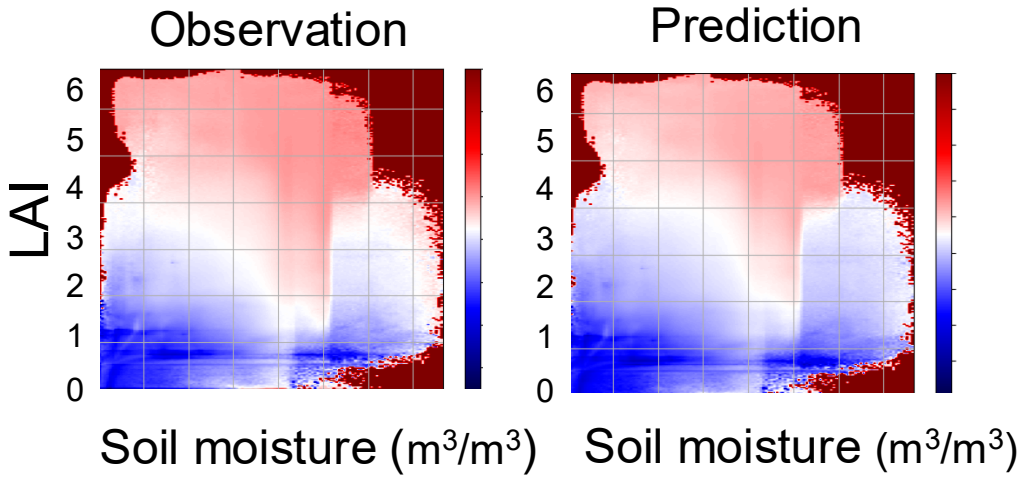
Garrigues et al., 2024, IGARSS

Sebastien Garrigues

Learning a forward operator for backscatter over land

One year evaluation (2019)





The learned observation operator shows accurate reconstruction of the relationships between backscatter and the geophysical variables

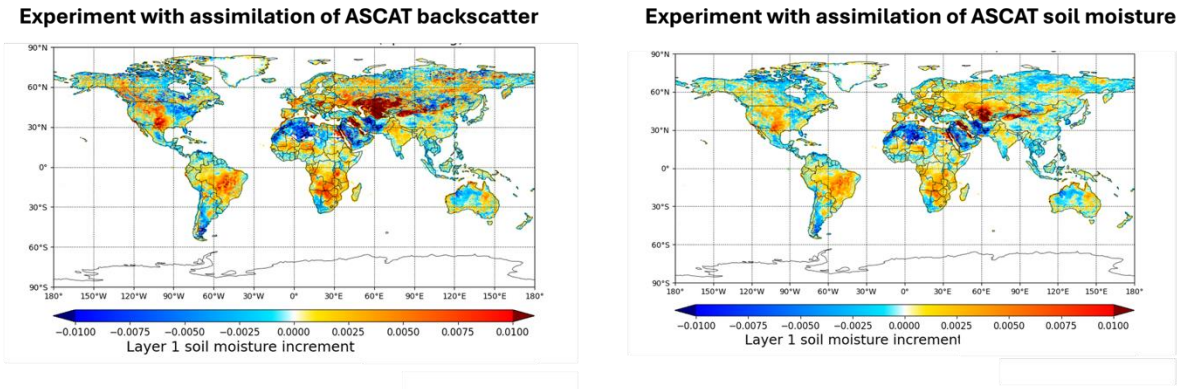
ECMWF

Garrigues et al., 2024, IGARSS

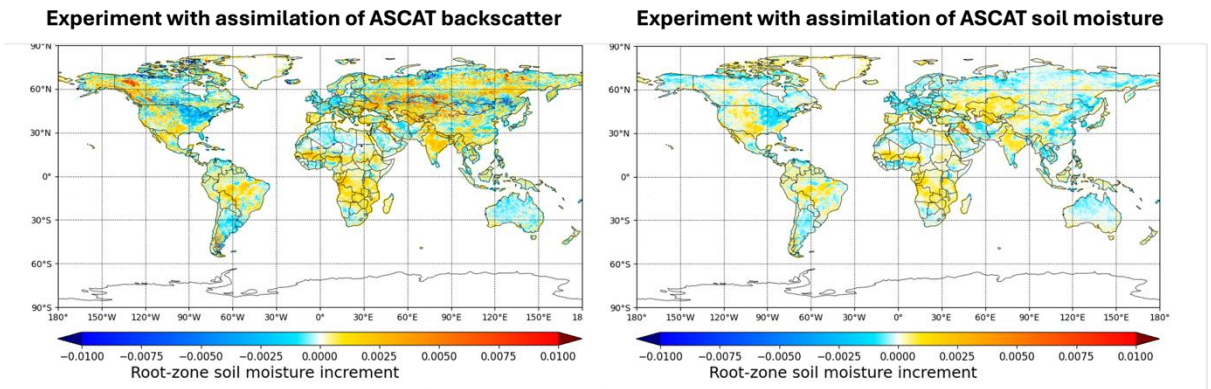
Sebastien Garrigues

Learning a forward operator for backscatter over land

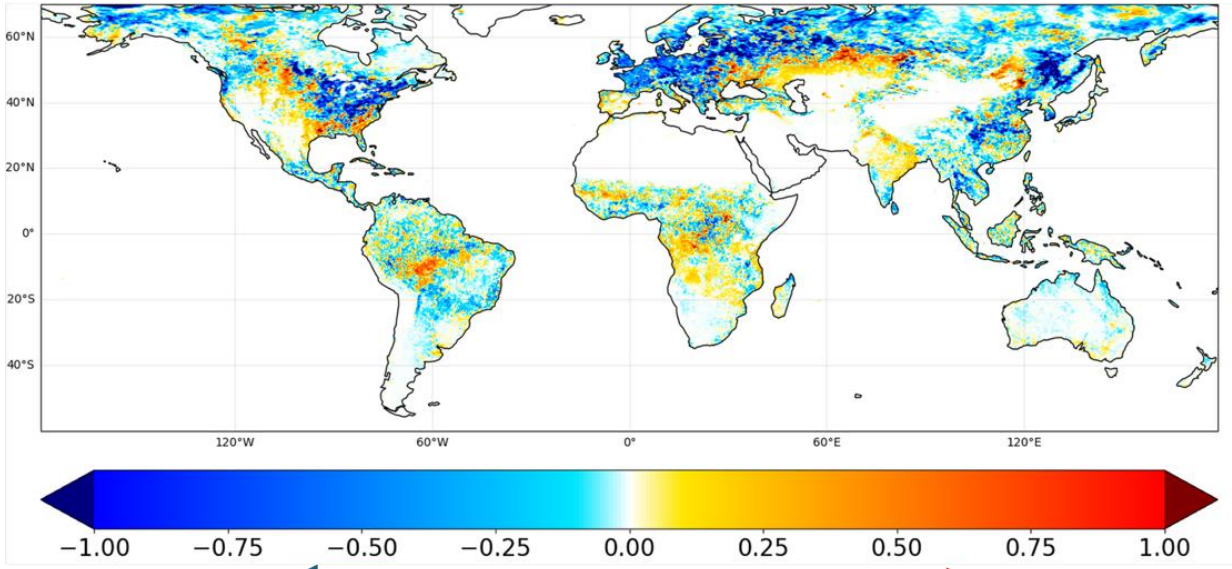
Surface soil moisture increments



Root-zone soil moisture increments



Impact on GPP: evaluation against satellite-based GPP (Gross Primary Production) - FLUXSAT



ECMWF

Other examples of learned observation operators:

MFASIS (visible reflectances)

Combining machine learning and data assimilation to retrieve sea ice concentration in 4D-Var

+ allow radiance assimilation over sea ice

ECMWF

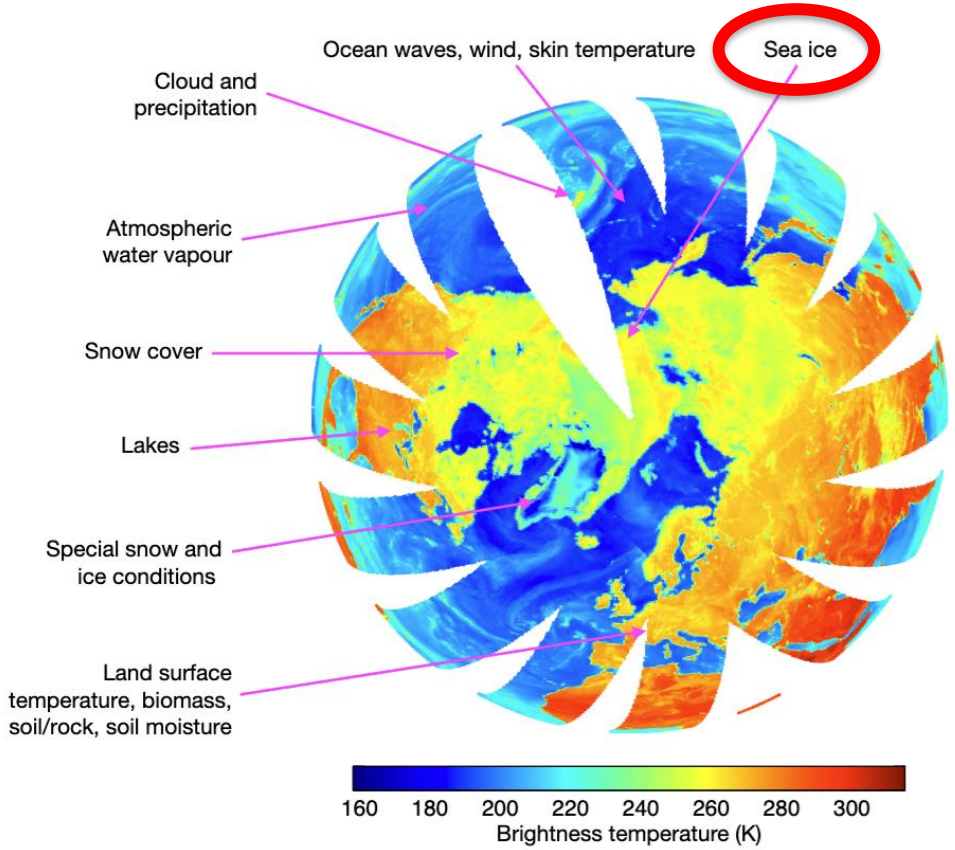
Combining machine learning and data assimilation to retrieve sea ice concentration in 4D-Var + allow radiance assimilation over sea ice

Alan Geer

Microwave brightness temperatures from channels with sensitivity to sea ice can provide valuable information on sea ice concentration and sea ice properties

However, we don't have a good model of sea ice emissivity and the surface properties that influence it…

Here an approach combining offline machine learning and data assimilation is used to solve this problem



Sea ice concentration

Alan Geer

Stage 1: Offline training

NN to estimate sea ice surface emissivity – used by RTTOV to provide output brightness temperatures

Training jointly updates:

i) 4 state variables: 3 sea ice empirical sea ice state variables + sea ice concentration (at obs locations)

iii) weights of a NN to estimate sea ice surface emissivity

Sea ice surface emissivity is used in combination with the sea ice concentration + ocean surface emissivity to provide a whole surface emissivity for RTTOV

Training targets:

AMSR-2 brightness temperatures

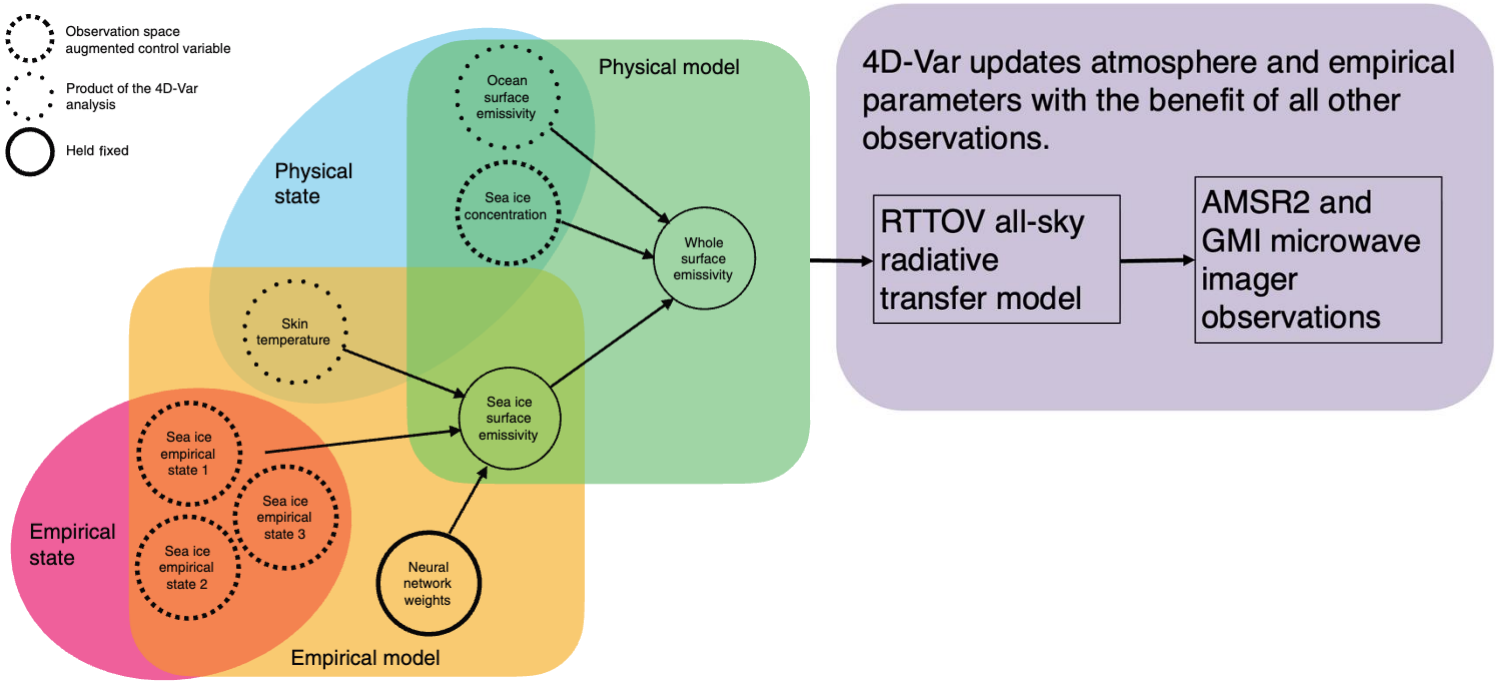
Stage 2: Online inside 4D-Var

Fixed NN weights

4D-Var updates:

3 empirical state variables

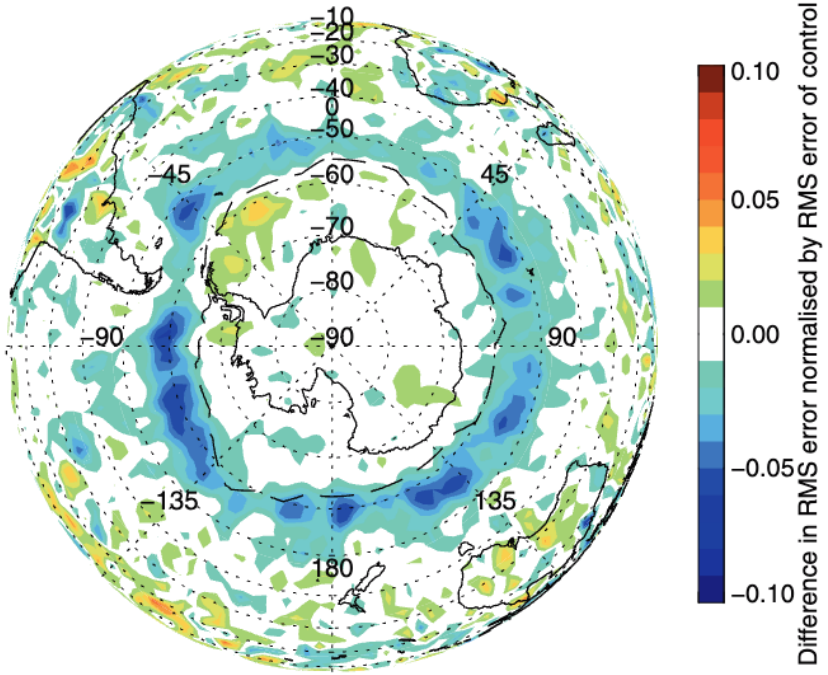
sea ice concentration



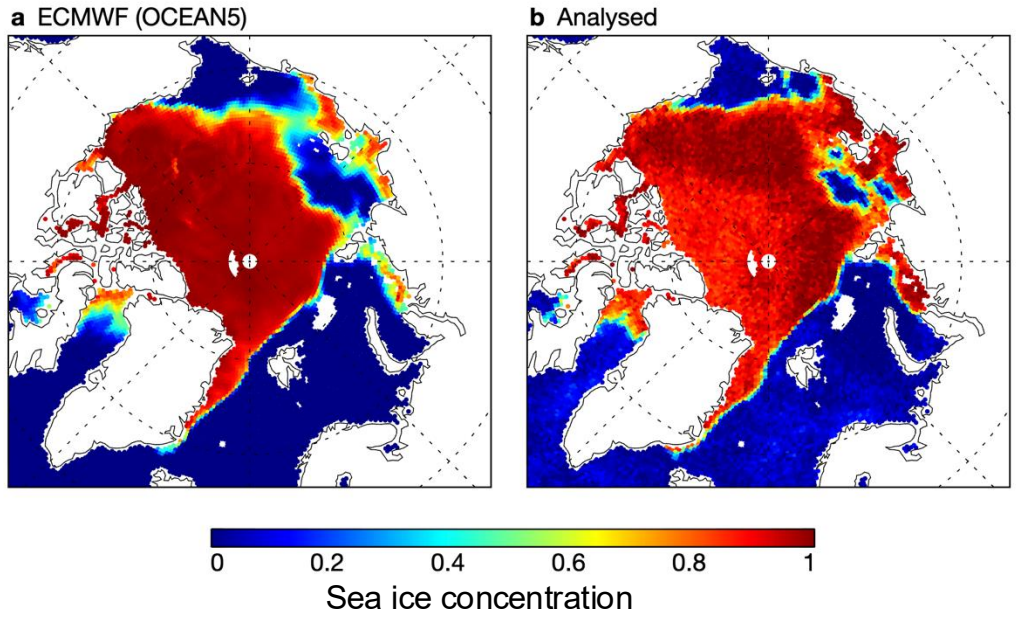
ECMWF

Alan Geer

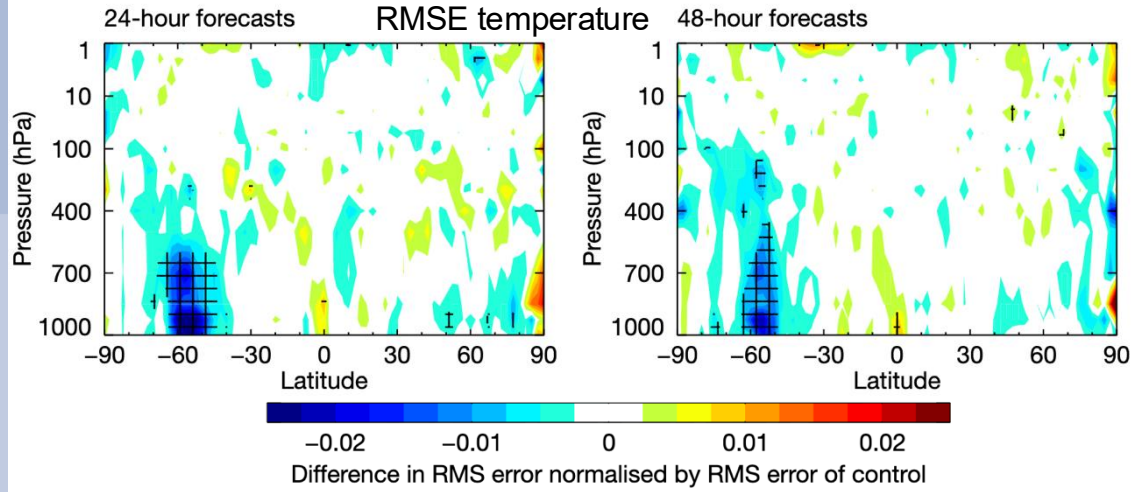
Sea ice concentration



Provides a new estimate of sea ice concentration



Evidence of improved temperature forecasts along the edge of the sea ice



Operational in CY49R1 (2024)

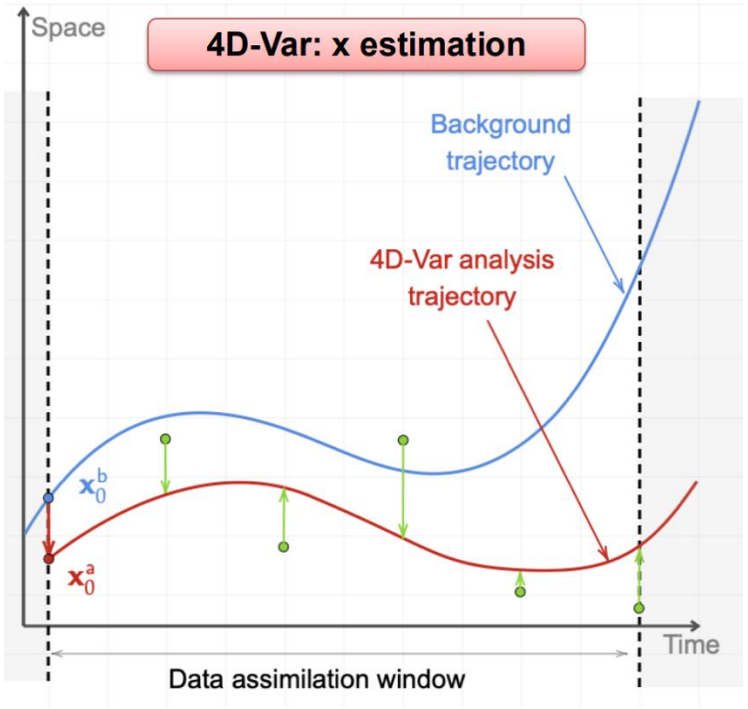
These techniques can also be applied
to many other uncertain components across IFS

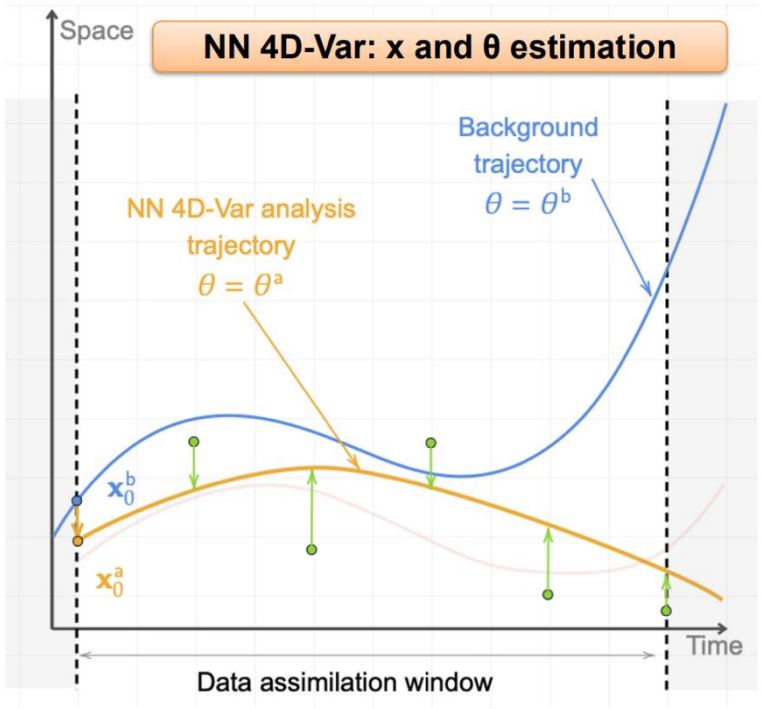
Example:

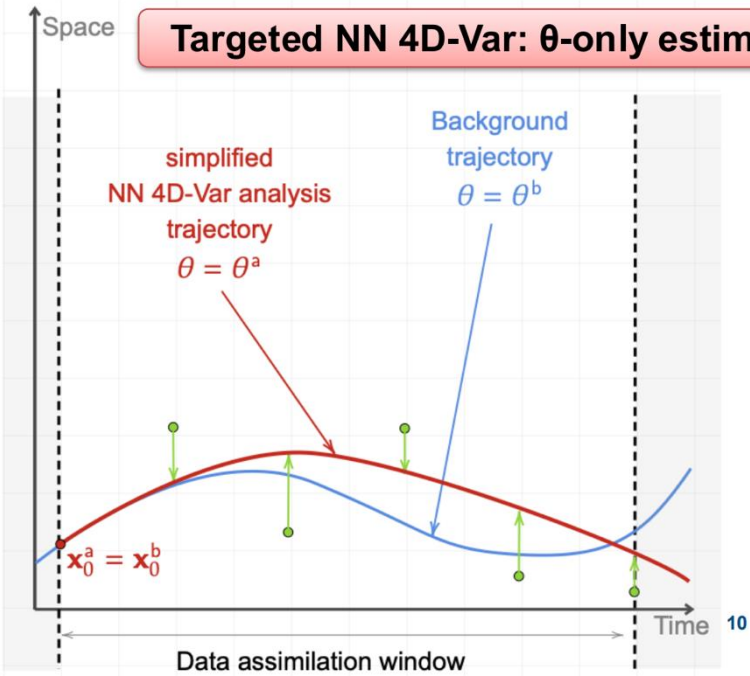
Learning (and correcting) model error

Learning a model error correction from observations

Use 4D-Var minimization to optimize the weights of a neural network embedded in the IFS which adjusts/corrects the model tendencies so as to fit the observations in the trajectory more closely







Results shown in this lecture are from a configuration with a separate 4D-Var step to updates NN weights only, fitting obs in a 48h window

ECMWF

Alban Farchi, Marcin Chrust

Learning a model error correction from observations

Online batched training:

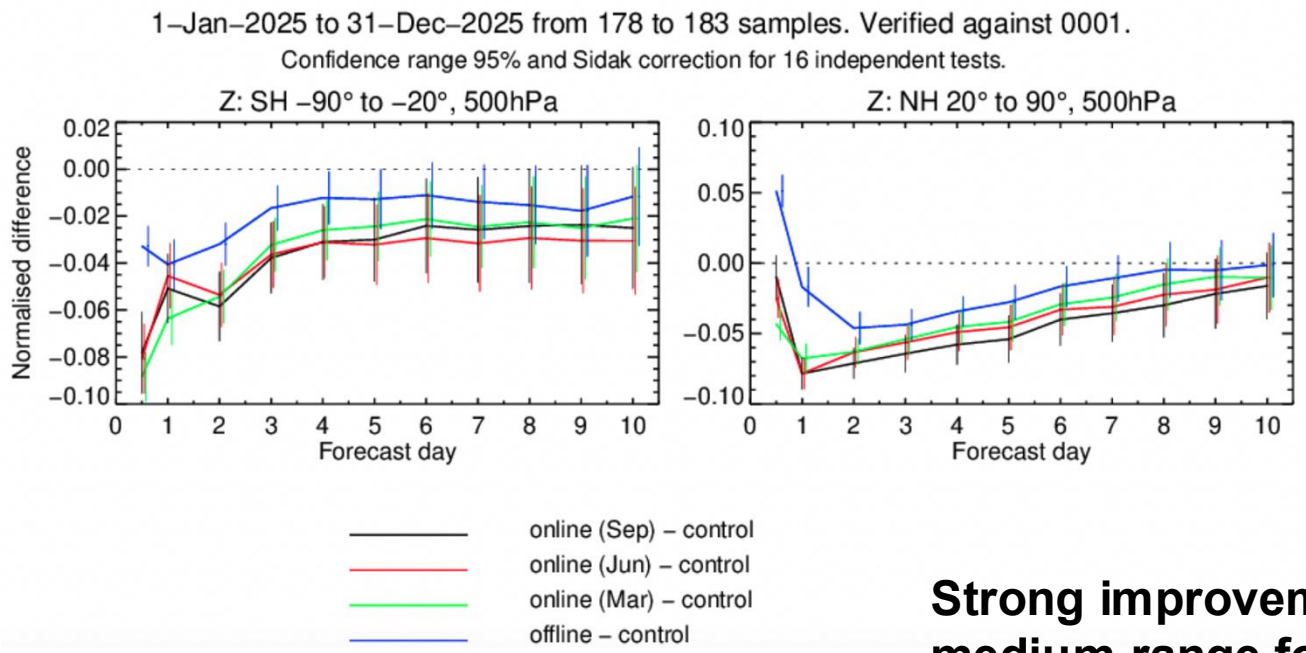
TCo399 resolution

8 members per batch (randomly sampling dates to improve generalization)

Training dataset: 182 48h-windows over 2024

Evaluation: forecast-only experiments in 2025

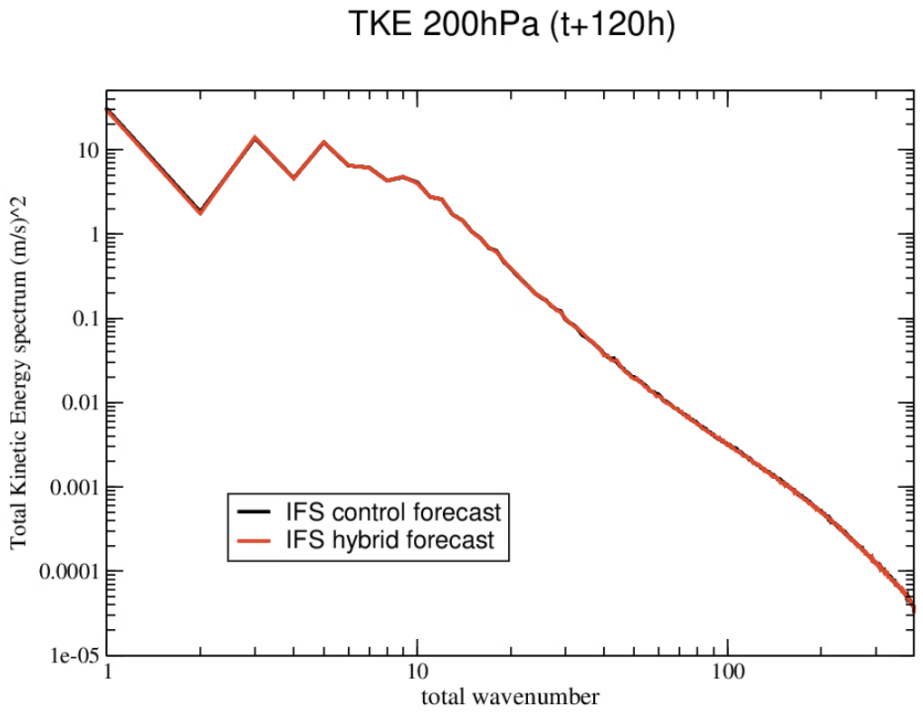
NN tendency corrections applied throughout the forecast

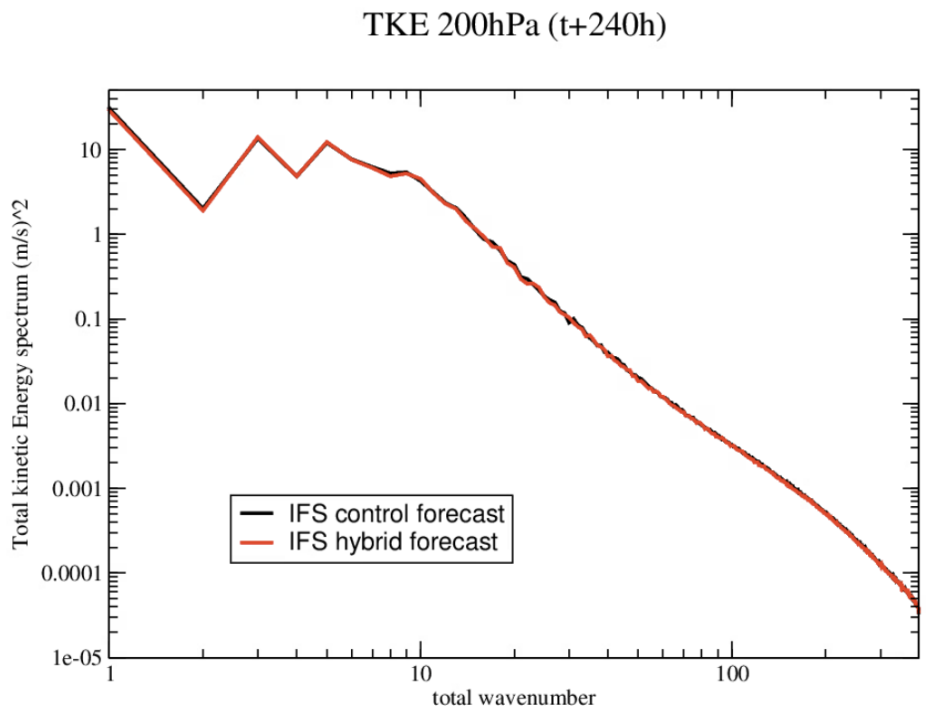


Strong improvements in medium-range forecast scores

Alban Farchi, Marcin Chrust

Learning a model error correction from observations





Dynamically indistinguishable from IFS forecasts

→ Paper coming soon!

ECMWF

Alban Farchi, Marcin Chrust

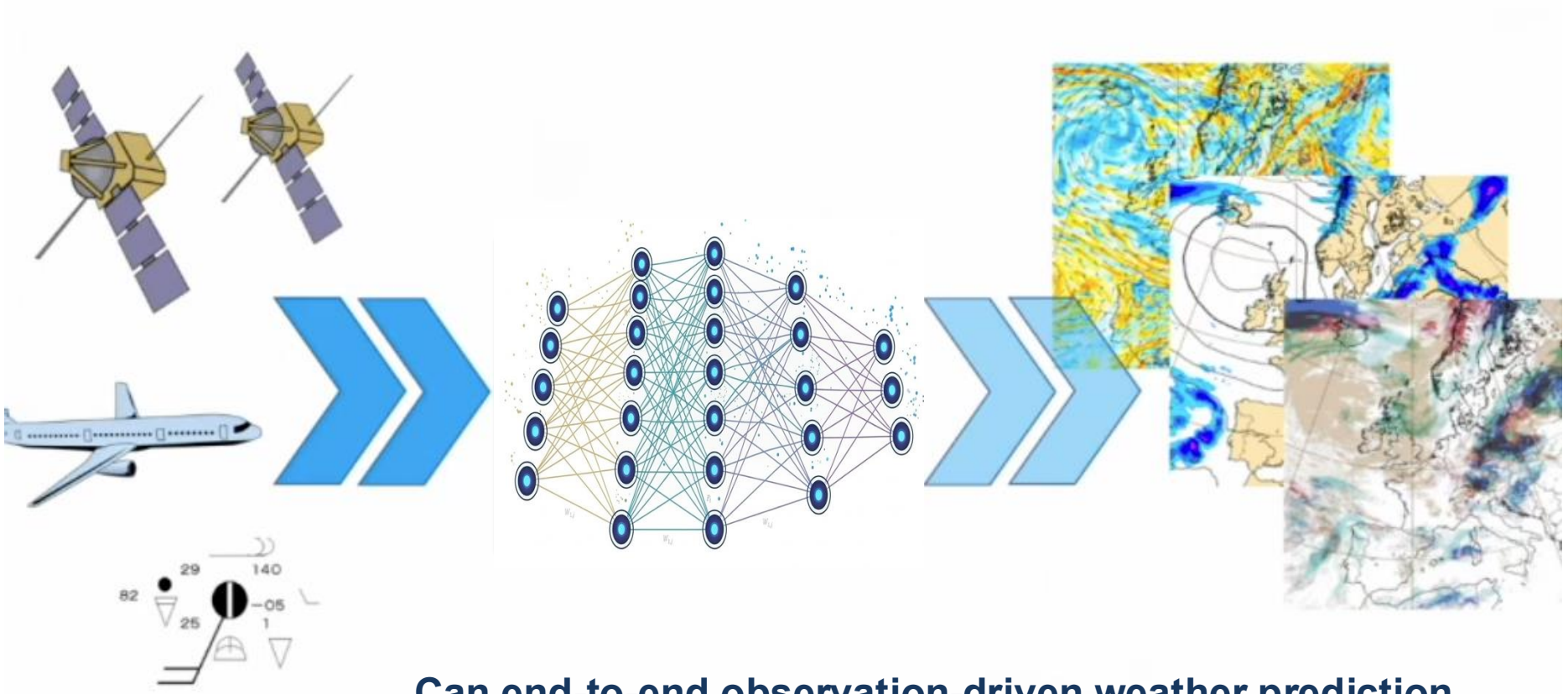
Example:

Learning everything - AI-Direct Observation Prediction

Thanks: Mihai Alexe, Eulalie Boucher, Ewan Pinnington, Patrick Laloyaux, Simon Lang, Tony McNally

ECMWF

Learning it all: AI-Direct Observation Prediction (AI-DOP)



Can end-to-end observation-driven weather prediction challenge the state of the art?

ECMWF

Learning it all: AI-Direct Observation Prediction (AI-DOP)

Why would we want to do this?

- because it gives the training process complete freedom to jointly optimize all parts of the system **end-to-end** so as to improve the thing we care about – the medium-range forecast scores
- speed – you can complete a full DA / forecast in ~ 2 minutes rather than 2 hours
- simplicity and sustainability of the system

Many open questions:

Are the observations complete enough to allow the network to learn sufficiently accurate representations of the physics that we normally encode into our NWP systems?

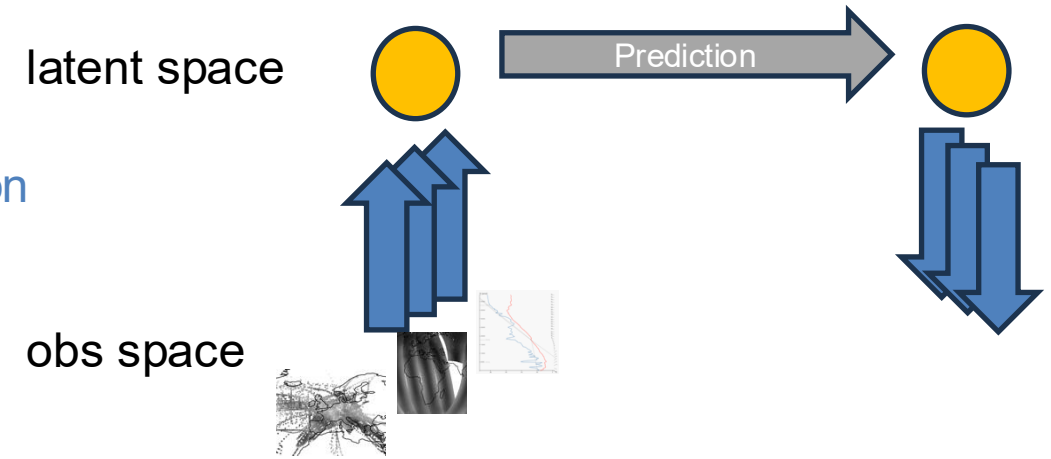
ECMWF

AI-Direct Observation Prediction concept

End-to-end observations → forecast (no explicit data assimilation step)

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

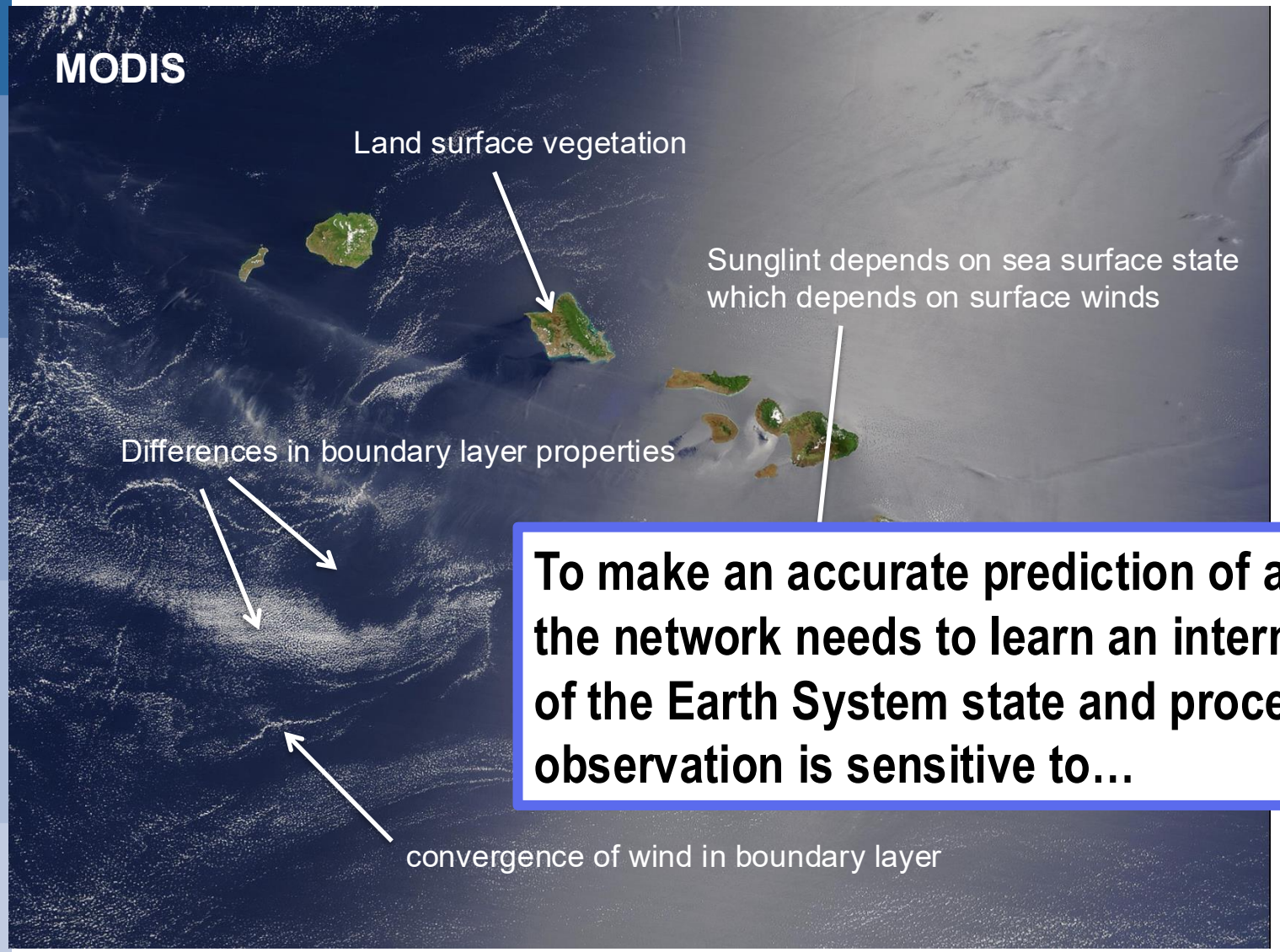
A **learned "DA" function** which combines information from different sources into a (learned) set of unified latent state variables

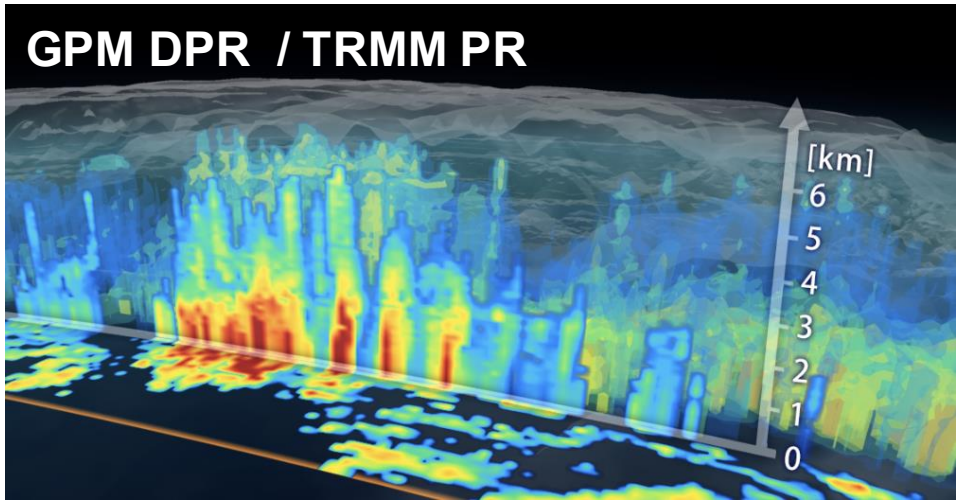


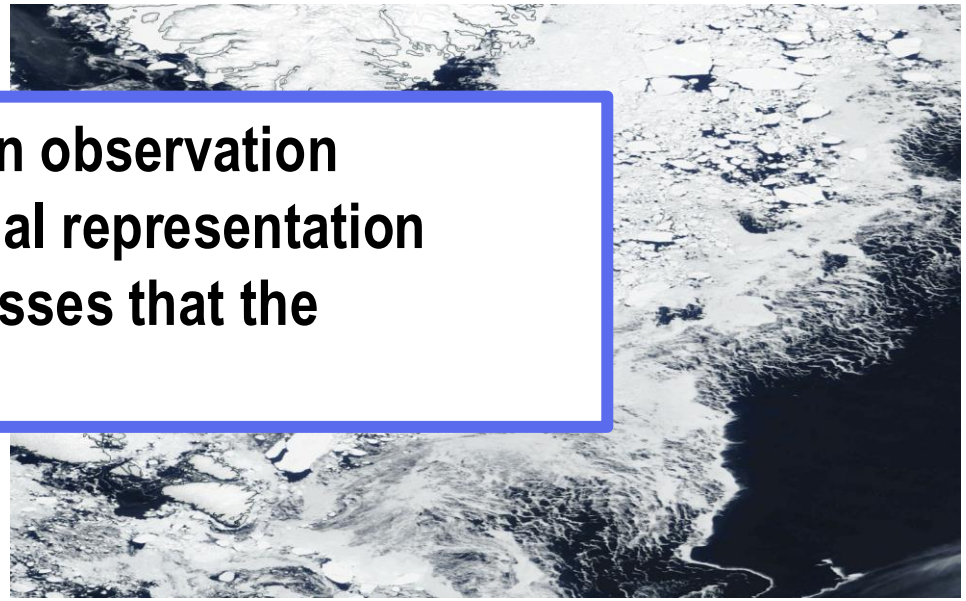
Learned observation operators which map from the network's internal model latent state to observation space

ECMWF

Learning from the rich information content of observational data





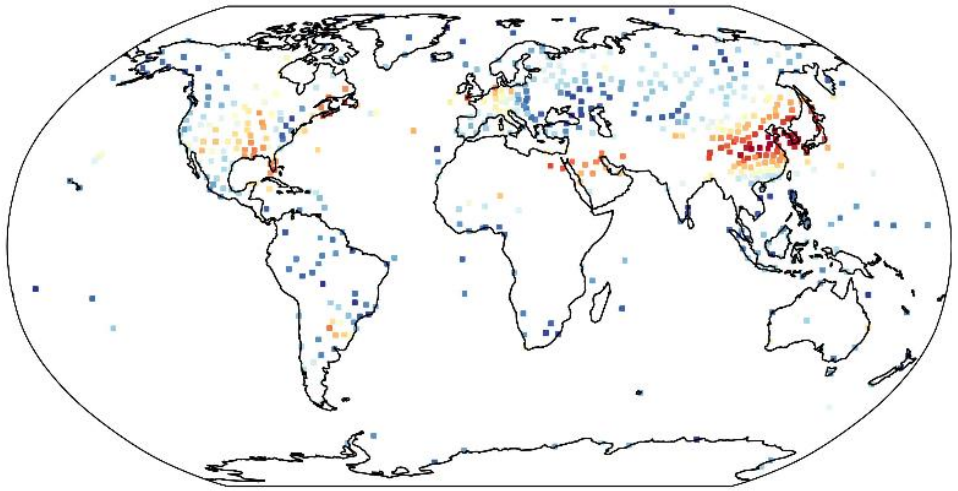


To make an accurate prediction of an observation the network needs to learn an internal representation of the Earth System state and processes that the observation is sensitive to…

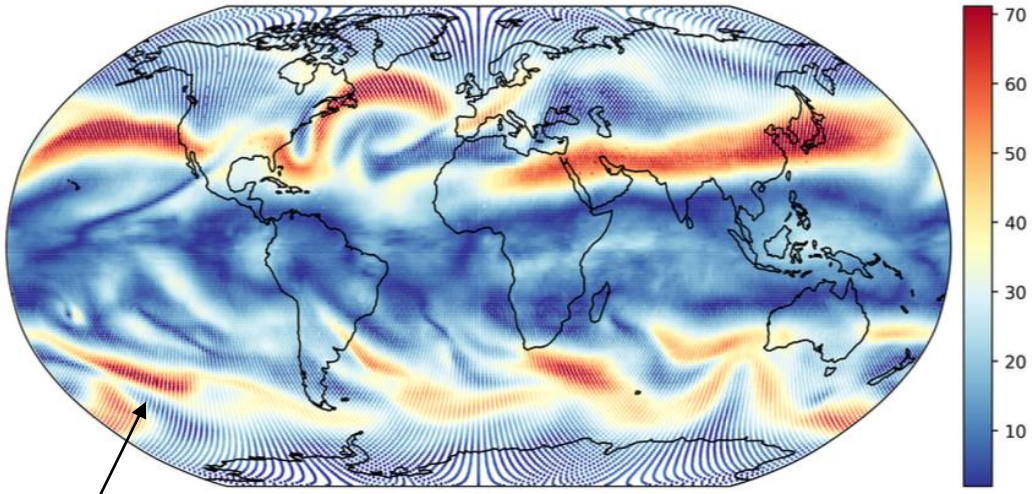
ECMWF

Learning a data assimilation function to combine information from multiple sources: GraphDOP can make predictions on a regular grid

Input 200hPa wind obs



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

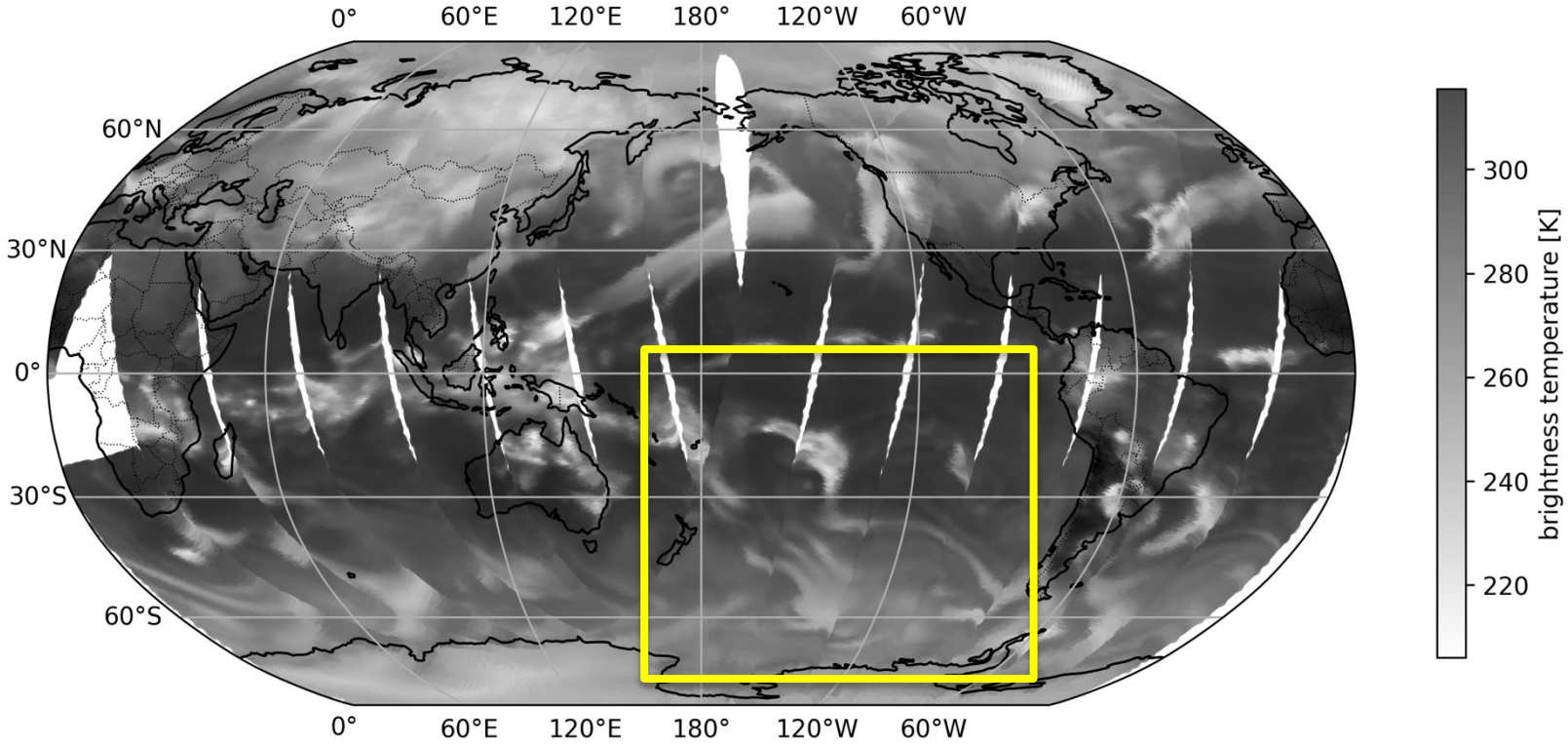


Even in areas where there are no direct observations for that variable

- evidence that it has developed a model of the relationship between different observed variables

ECMWF

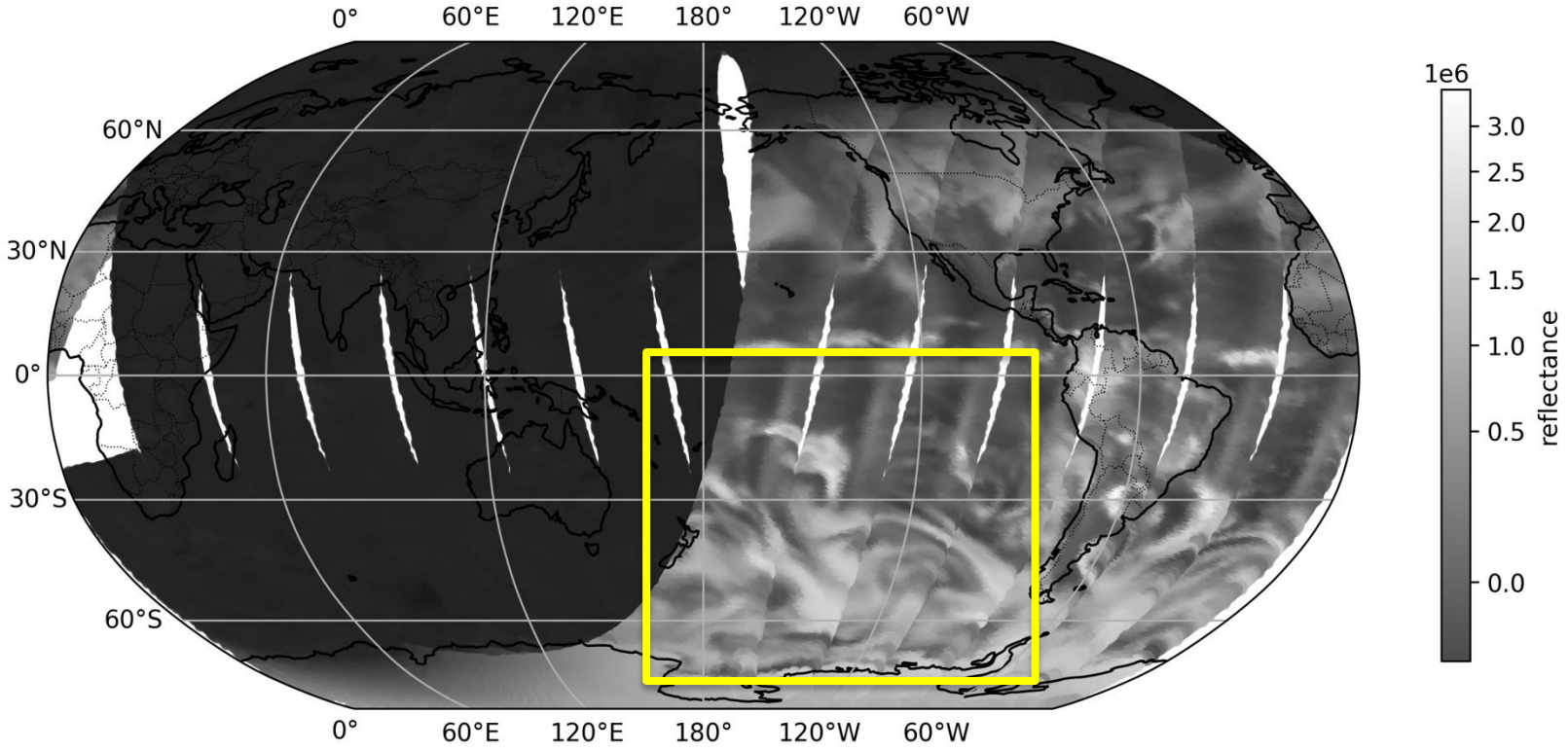
Learning a data assimilation function to combine information from multiple sources into a unified latent state: cloud example



Predicted IASI ch 756

ECMWF

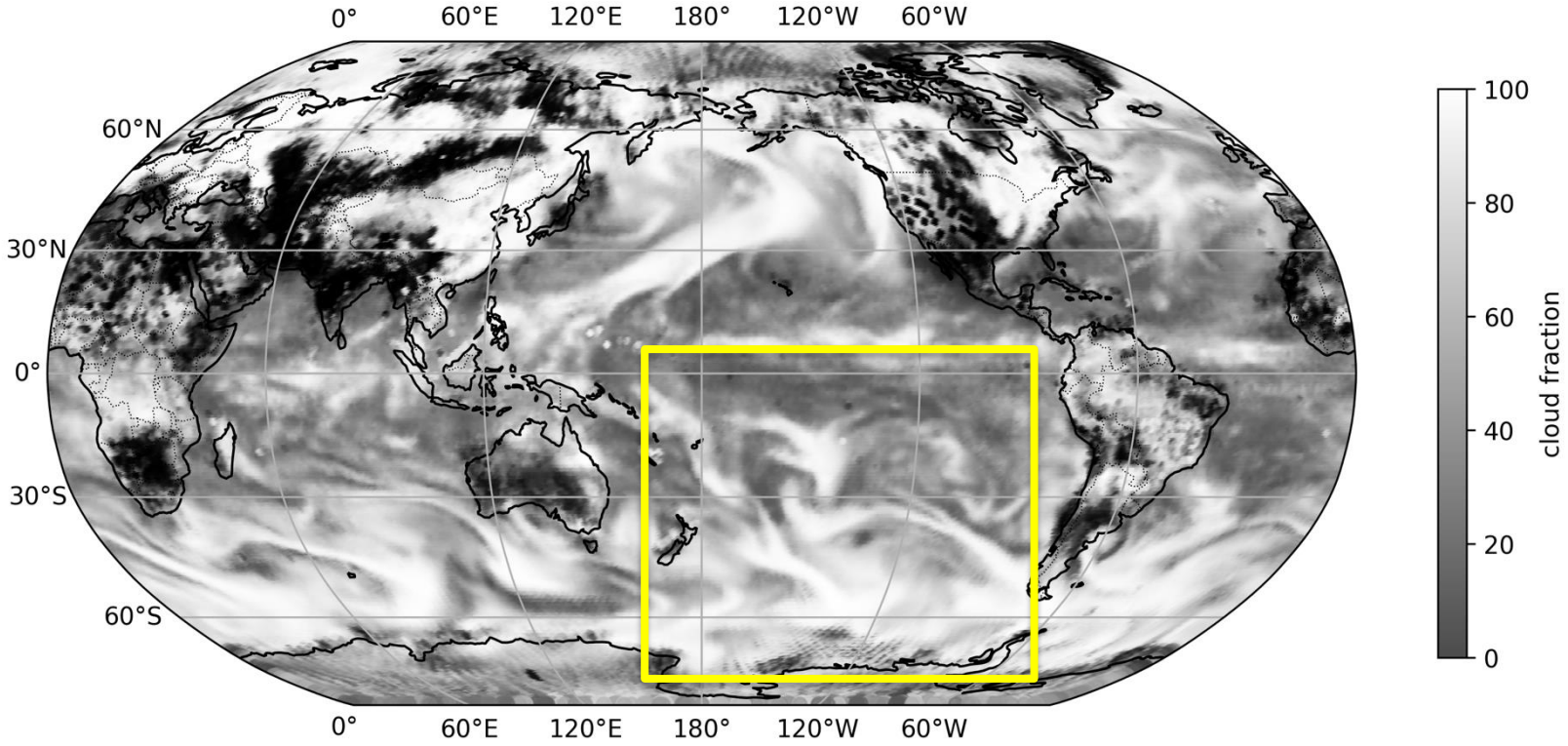
Learning a data assimilation function to combine information from multiple sources into a unified latent state: cloud example



Predicted AVHRR visible

ECMWF

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

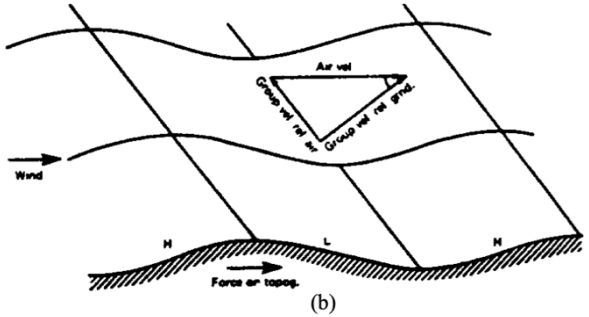


Predicted SYNOP cloud fraction
(on a regular grid)

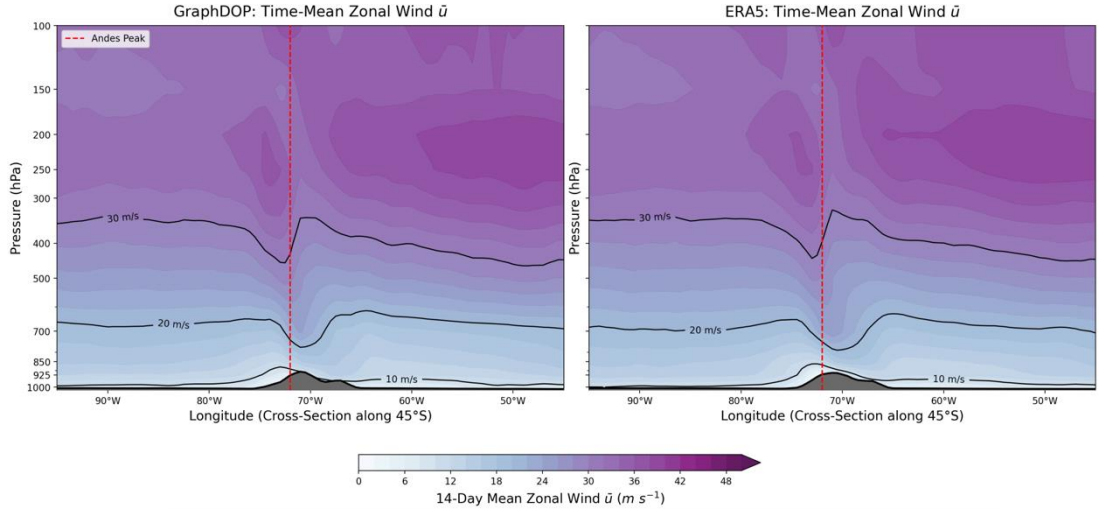
ECMWF

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

Flow over orography



GraphDOP



ERA5

Ageostrophic flow

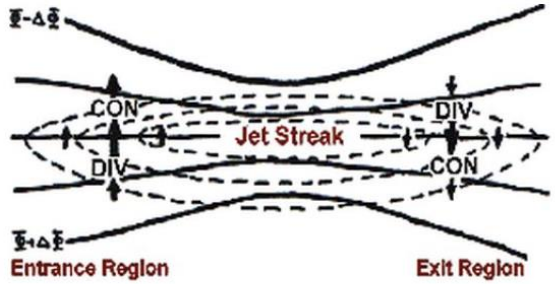
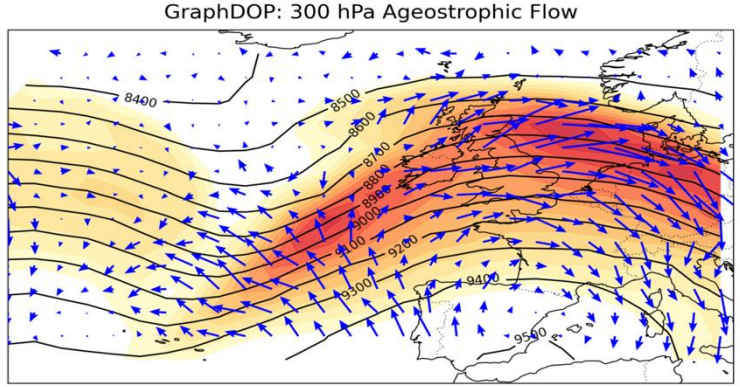
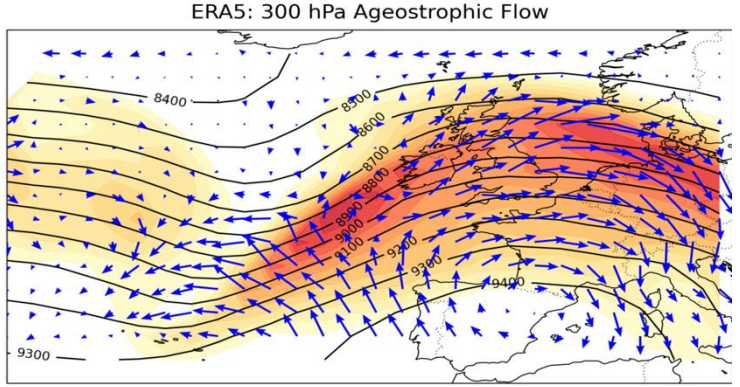


Figure 2: Location of convergence (CON) and divergence (DIV) areas in the entrance and exit region of a jet streak, © NOAA

GraphDOP



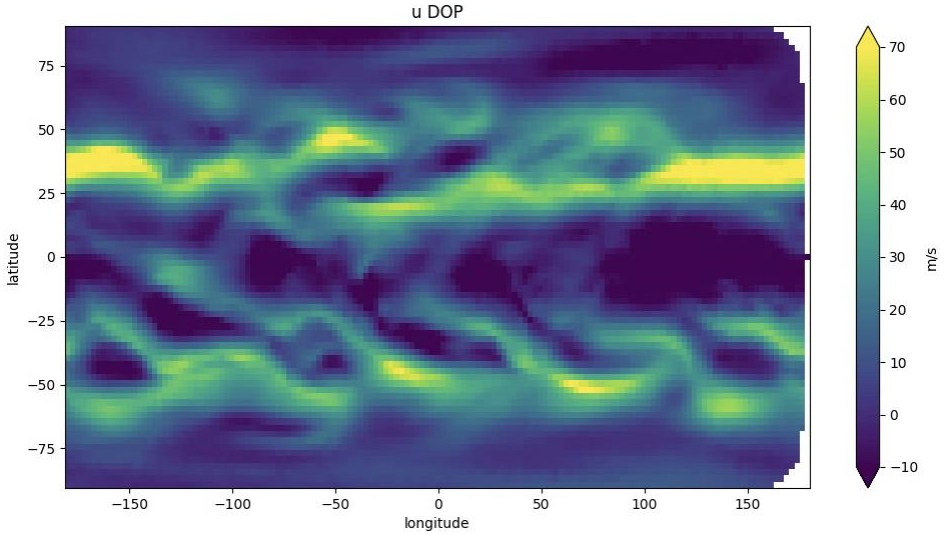
ERA5



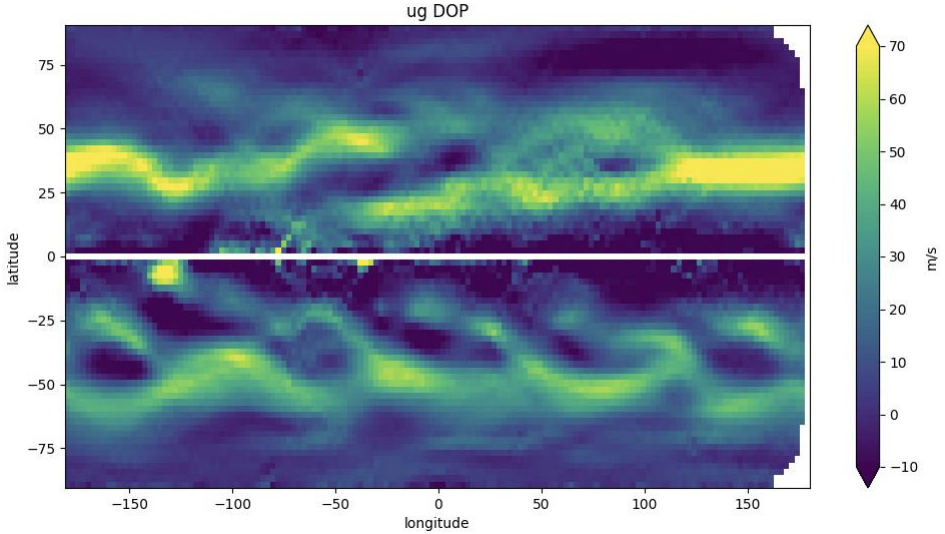
ECMWF

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

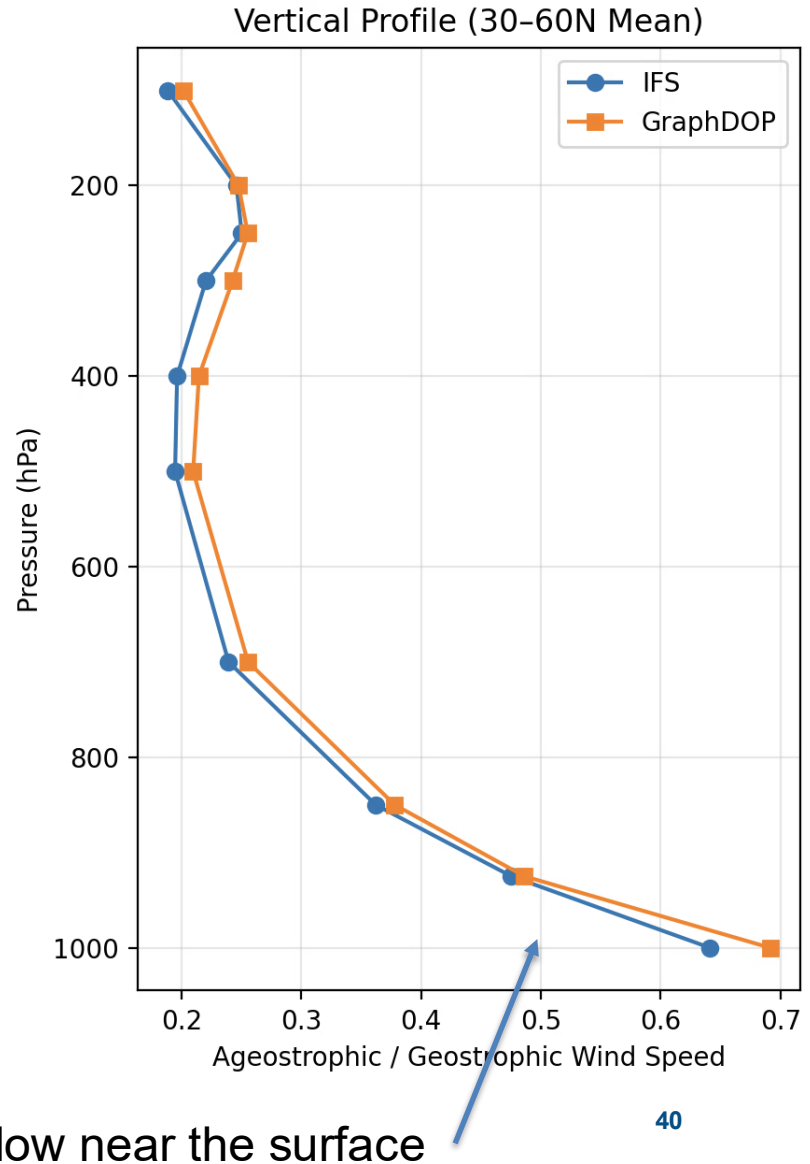
Balance relationships



u-wind
200hPa
DOP prediction



Geostrophic
u-wind
200hPa
derived from
geopotential
prediction



Increased ageostrophic flow near the surface

ECMWF

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

Semi-diurnal solar atmospheric tides

$$S_2^* = \frac{(p_0 + p_{12}) - (p_6 + p_{18})}{4}$$

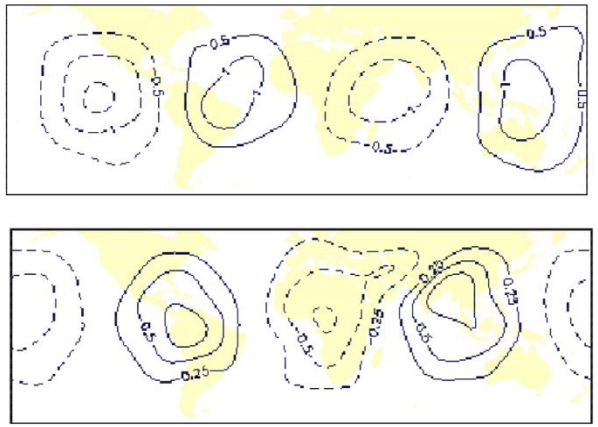
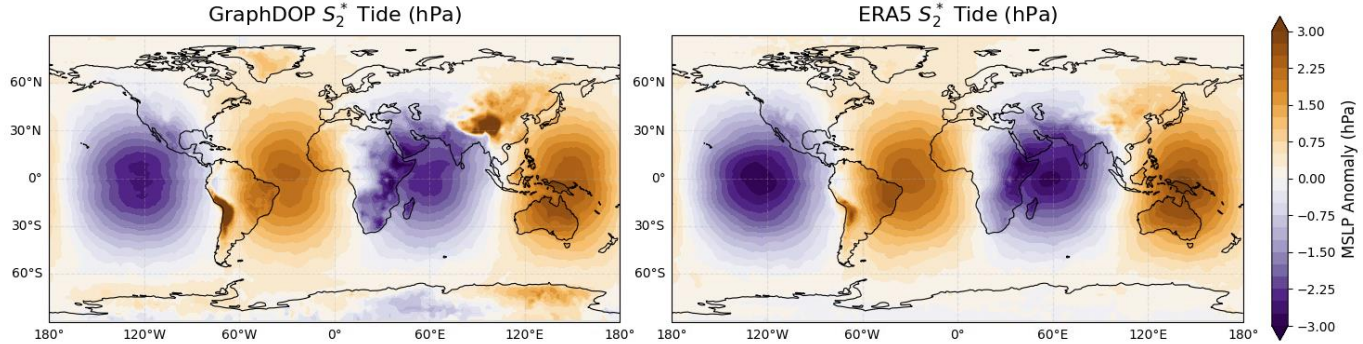


Figure 5. The semi-diurnal tide. The upper plot shows surface pressure (hPa) and the lower plot the tendency of pressure (hPa/hour) at the lowest model level. The quantity plotted is $(1/4)(A_{00} - A_{06} + A_{12} - A_{18})$ where A_{xx} denotes the mean ECMWF analysis at xx UTC for January 1997, truncated spectrally at T10 to remove local orographic and station-specific features. Solid contours denote positive values and dashed contours negative values. The nature of atmospheric tides is described in Chapman and Lindzen (1970).



GraphDOP ERA5

ECMWF

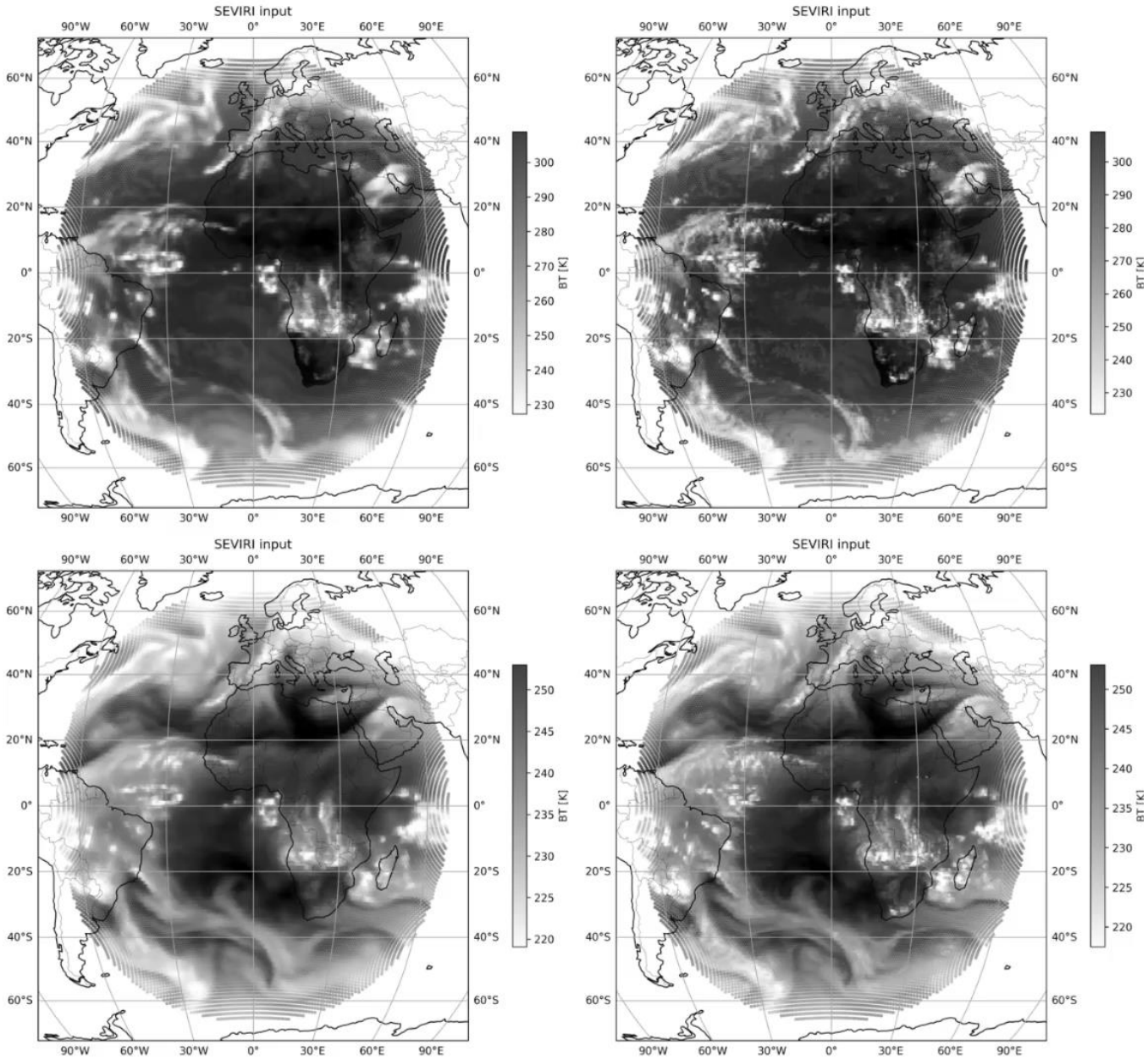
5-day forecast:

10.8µm (window channel)

Representations of **convection** and
extra-tropical dynamics

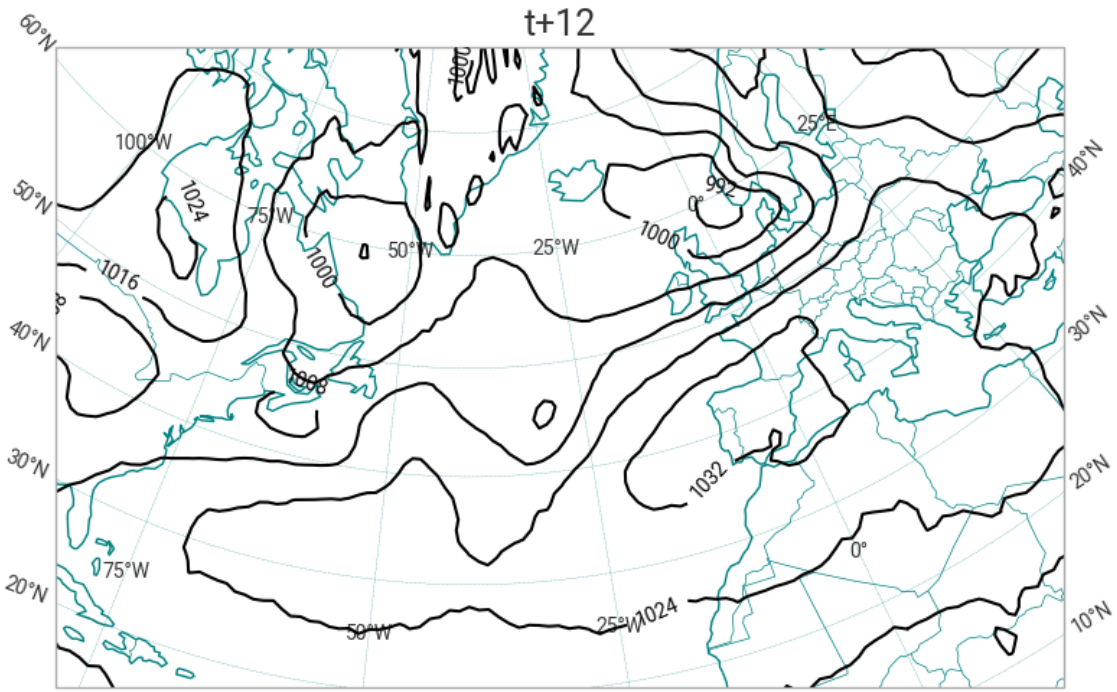
6.3µm (water vapor channel)

GraphDOP (prediction) **SEVIRI (observed)**

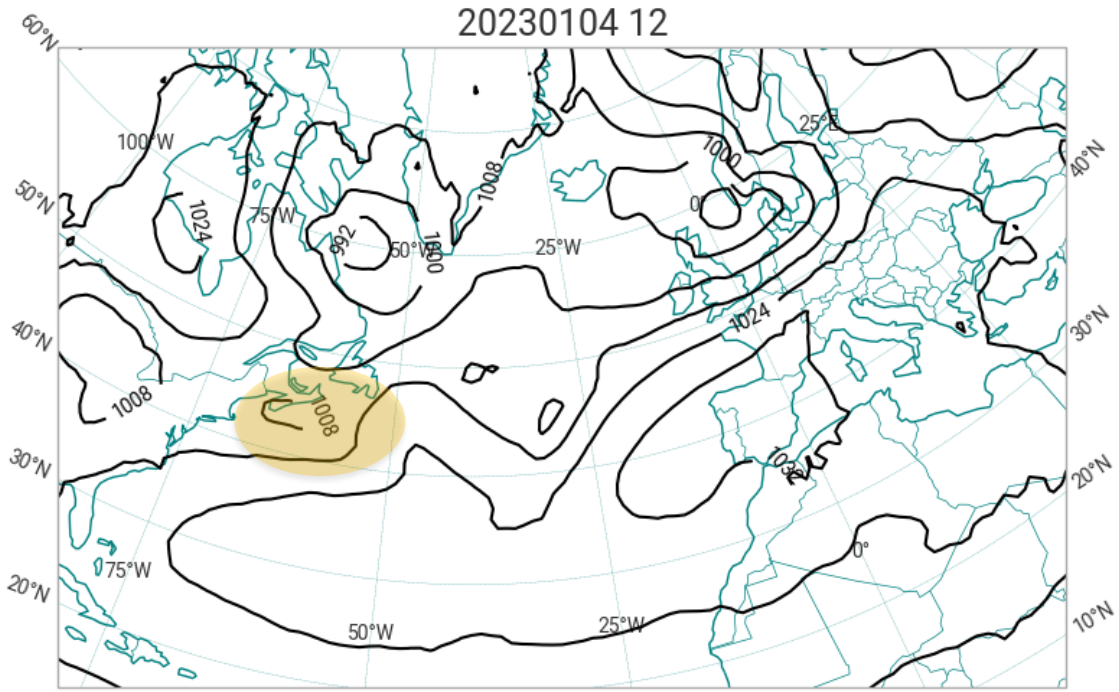


Extra-tropical dynamics: explosive cyclogenesis

Mean sea level pressure



GraphDOP prediction (gridded)

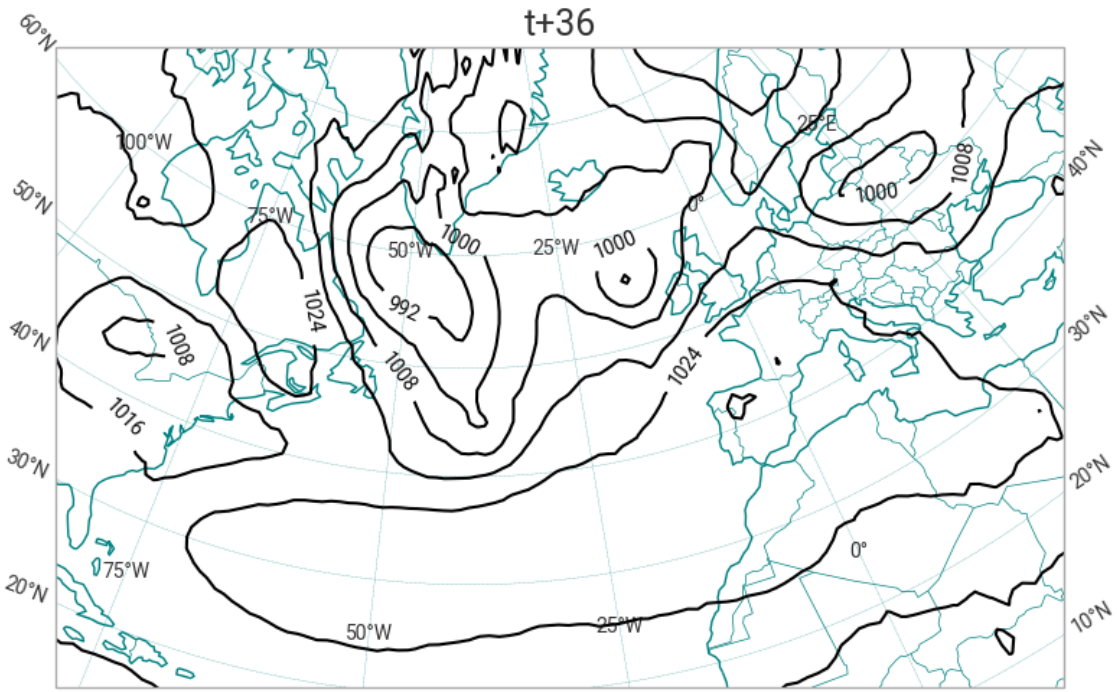


ECMWF operational analysis
(for reference only)

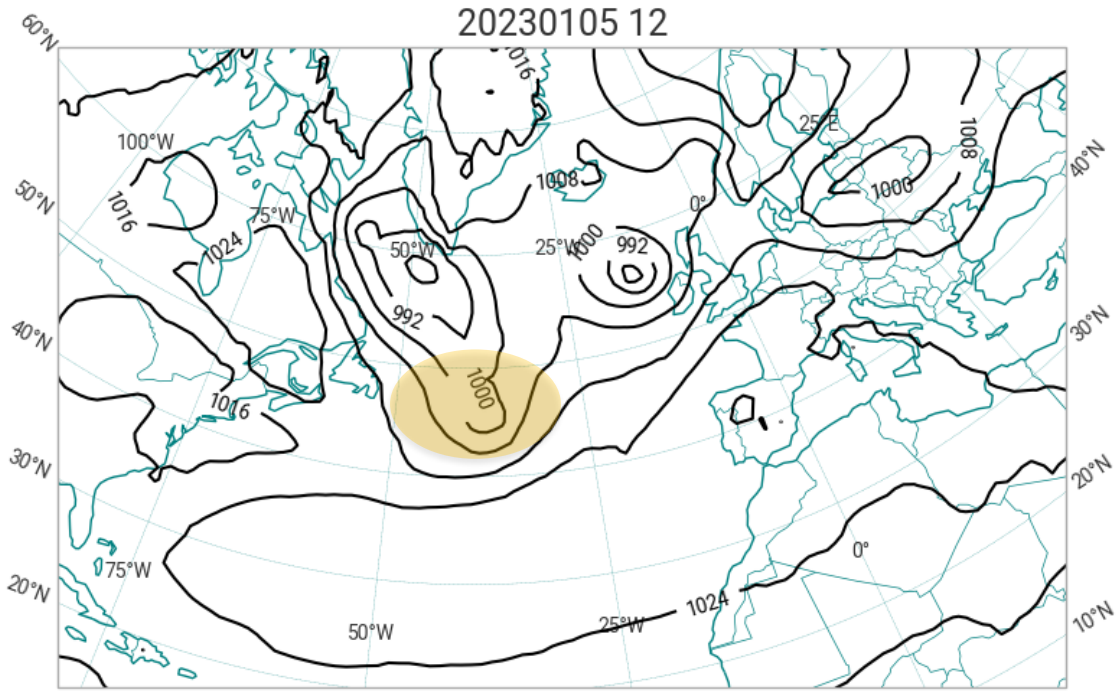
ECMWF

Extra-tropical dynamics: explosive cyclogenesis

Mean sea level pressure



GraphDOP prediction (gridded)

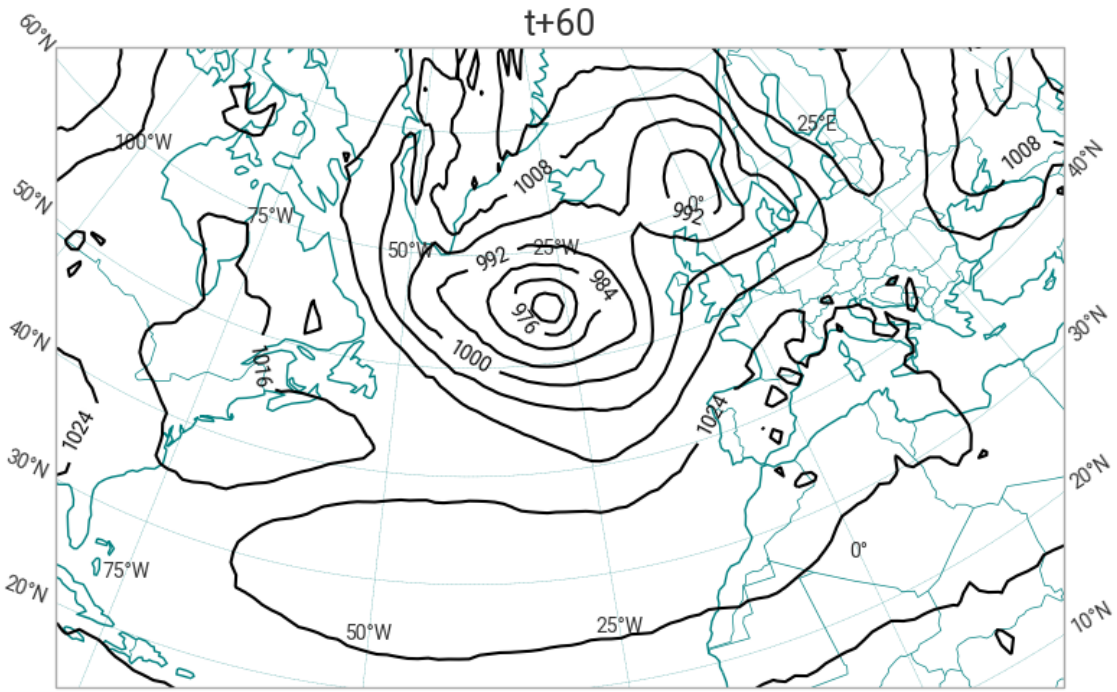


ECMWF operational analysis
(for reference only)

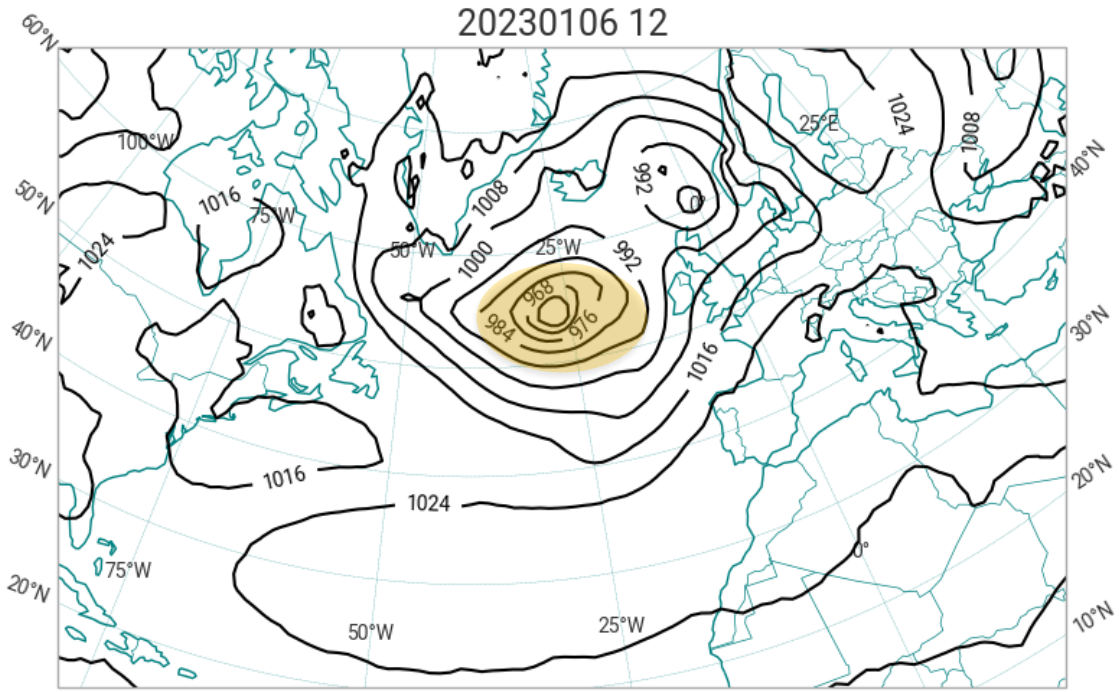
ECMWF

Extra-tropical dynamics: explosive cyclogenesis

Mean sea level pressure



GraphDOP prediction (gridded)

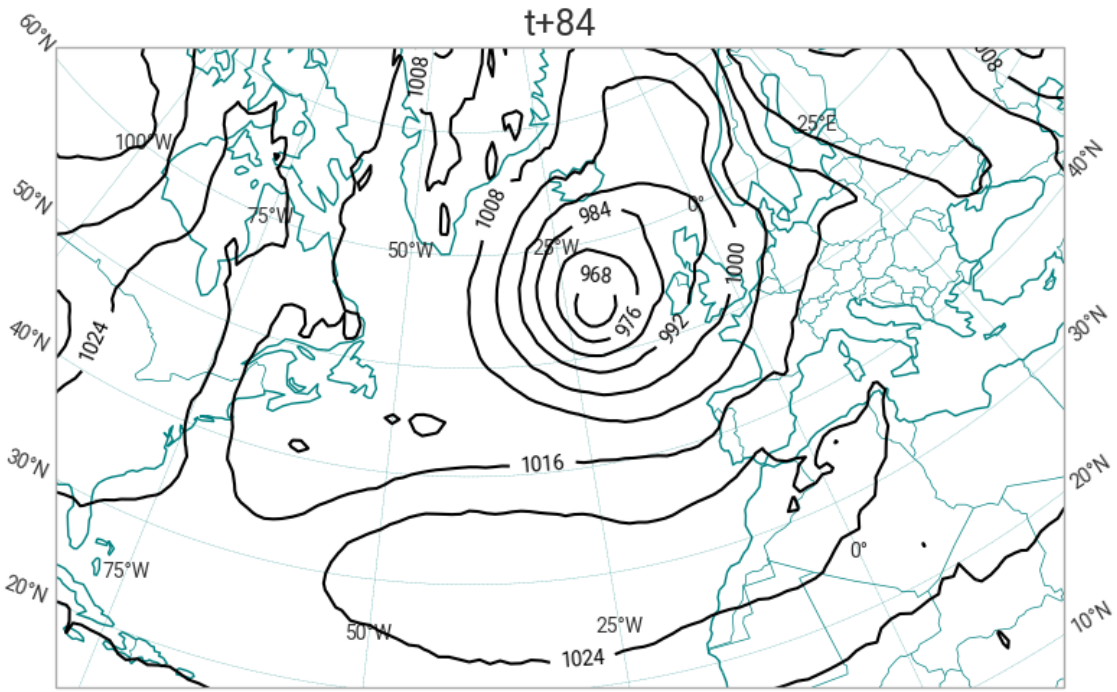


ECMWF operational analysis
(for reference only)

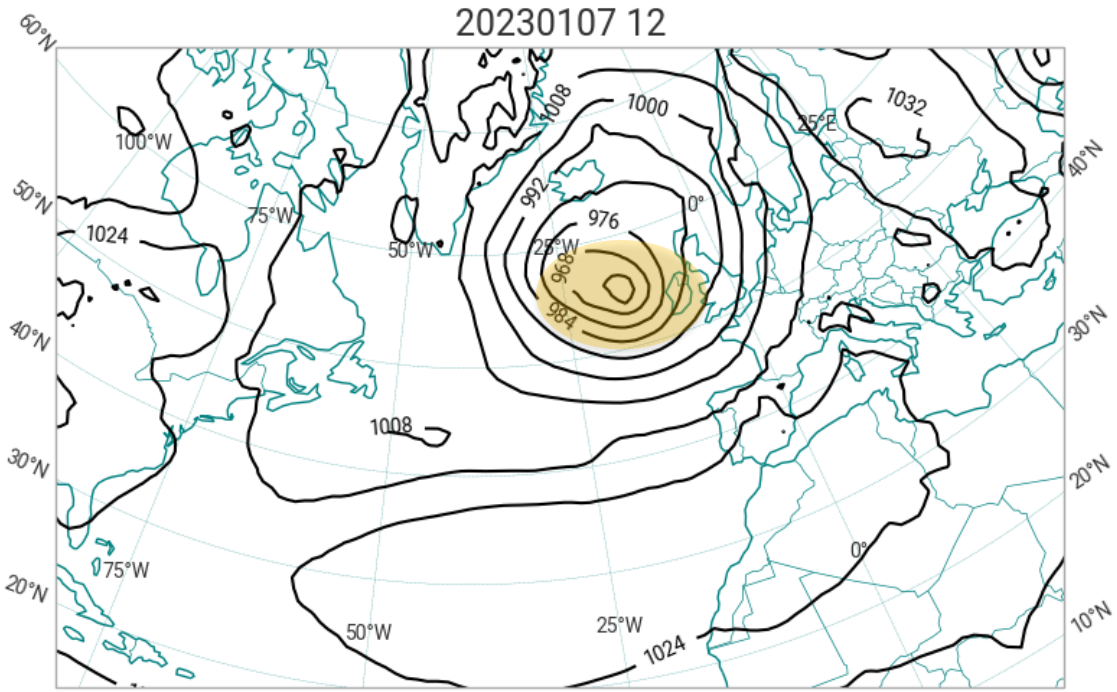
ECMWF

Extra-tropical dynamics: explosive cyclogenesis

Mean sea level pressure



GraphDOP prediction (gridded)

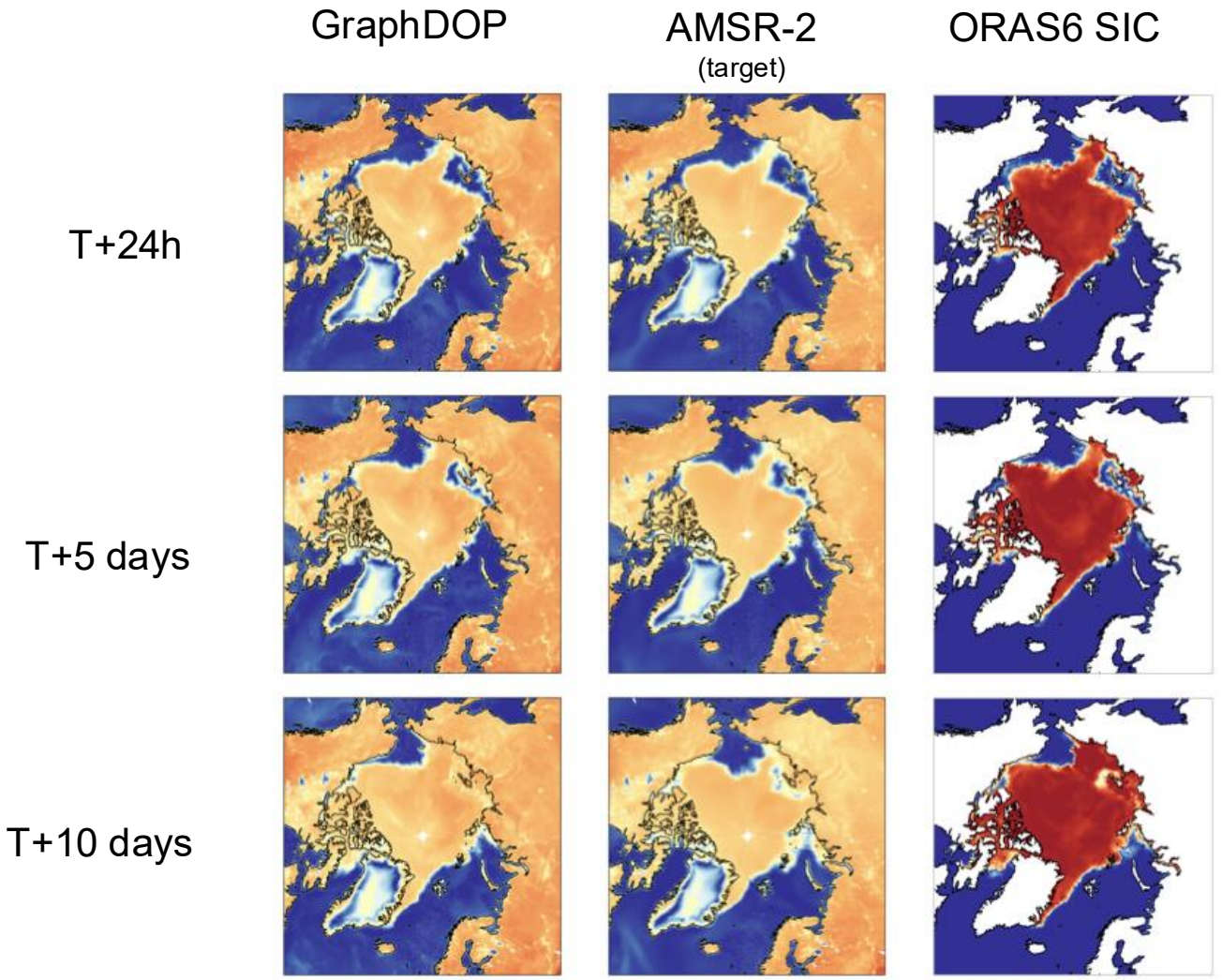


ECMWF operational analysis
(for reference only)

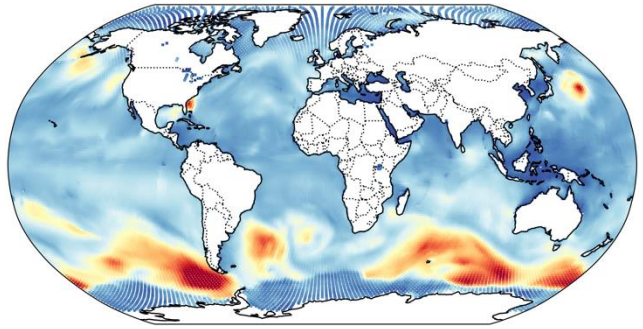
ECMWF

Case study: Arctic rapid freezing event

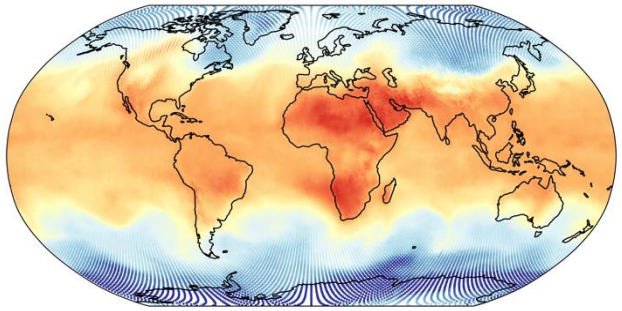
- AMSR-2 (5v) MW brightness temperature signal of the sea ice extent
  - **Captures a rapid freezing event**
  - Reproduces the signature of sea ice in the brightness temperatures
  - Underscores the capacity to learn representations of sea ice dynamics from indirect observations without access to explicit sea ice labels



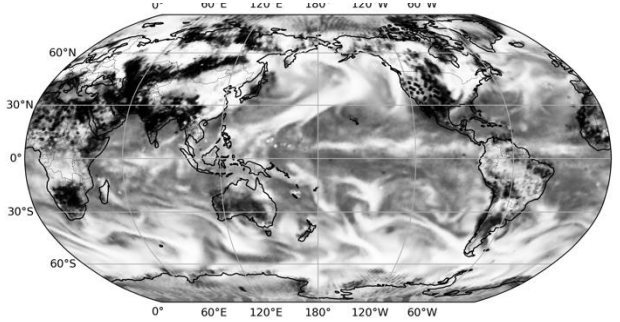
Learning observation operators that map back to observed variables



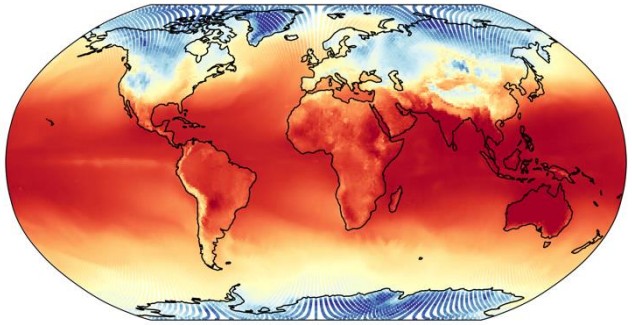
Significant wave height



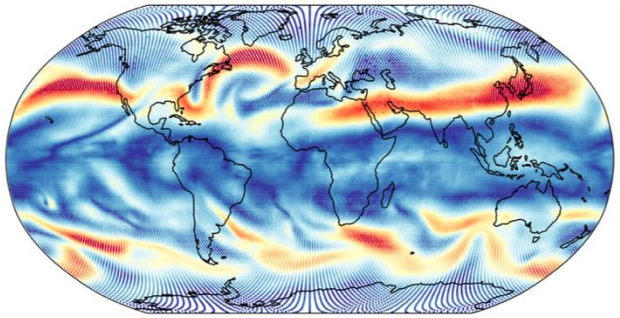
850hPa temperature



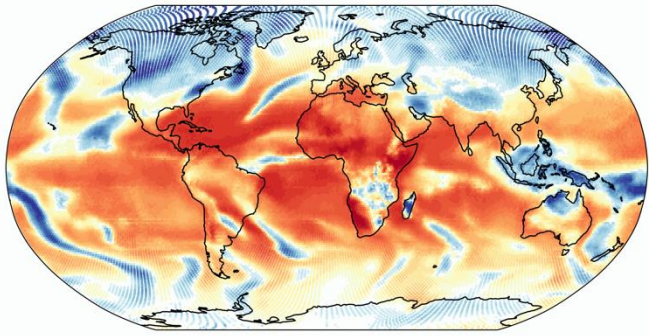
Cloud fraction



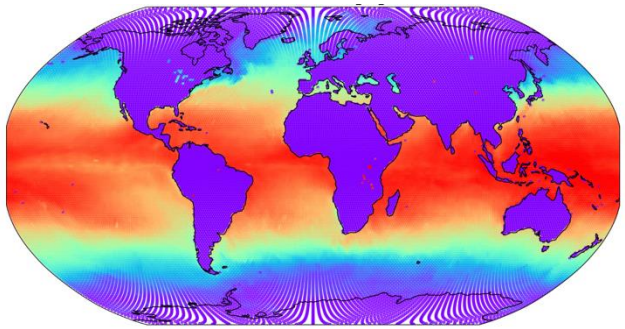
2-meter temperature



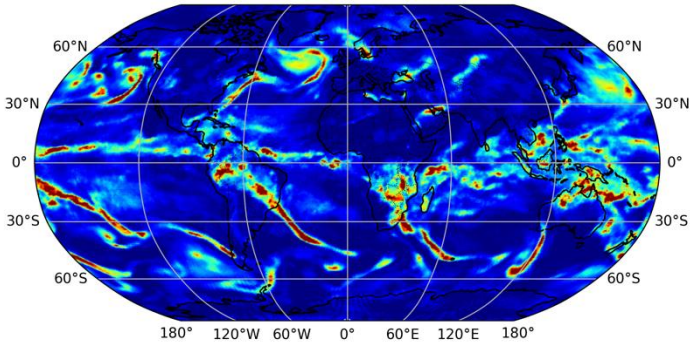
200hPa winds



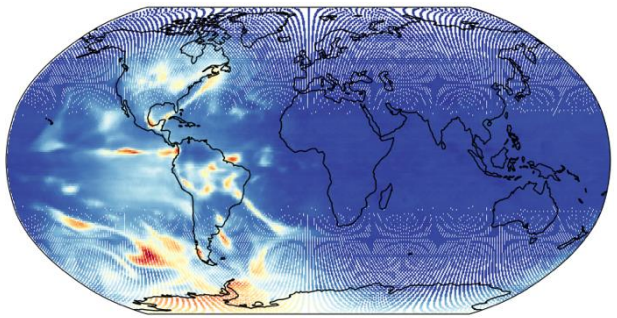
SEVIRI infrared window channel



SST

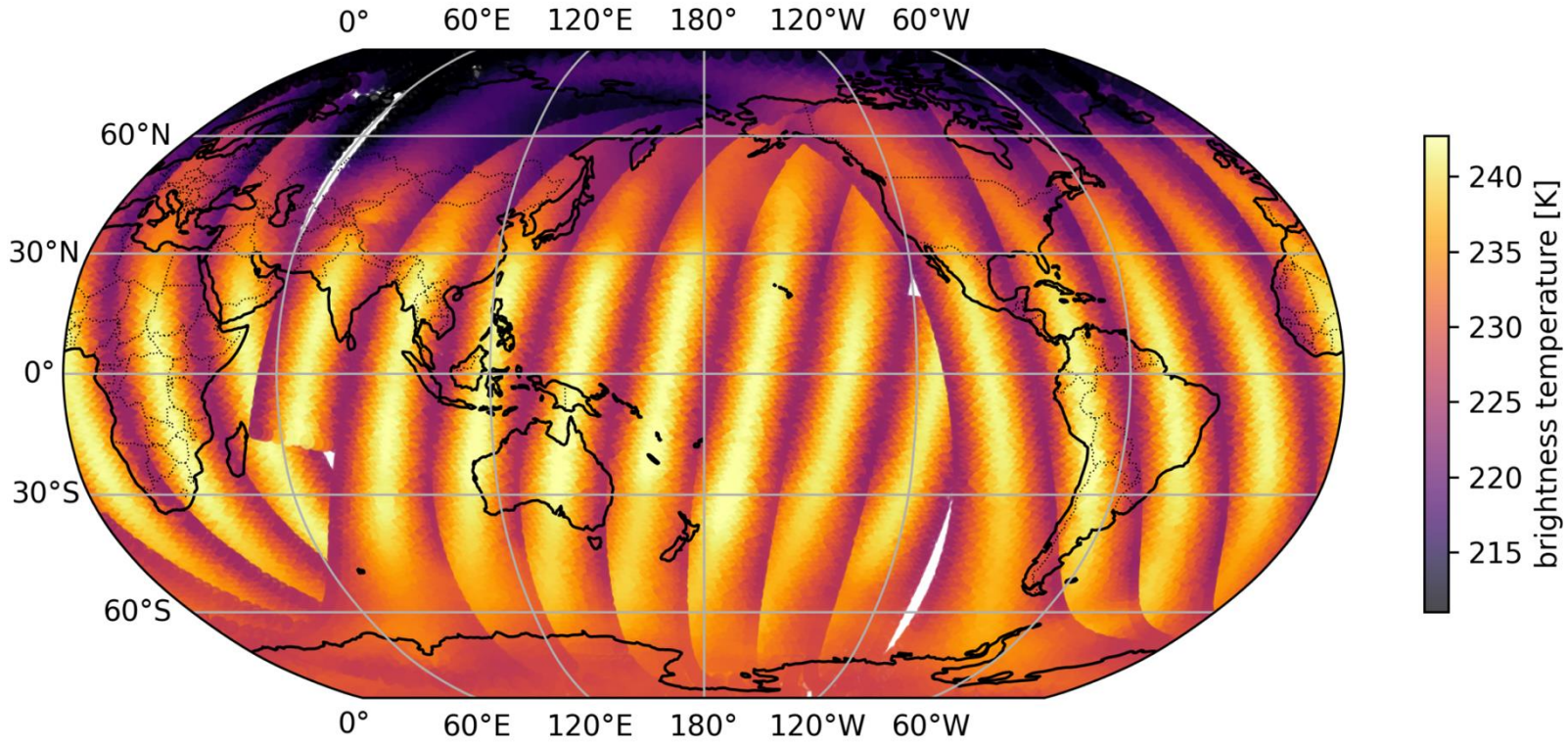


1 hour precipitation accumulation



AVHRR visible channel

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

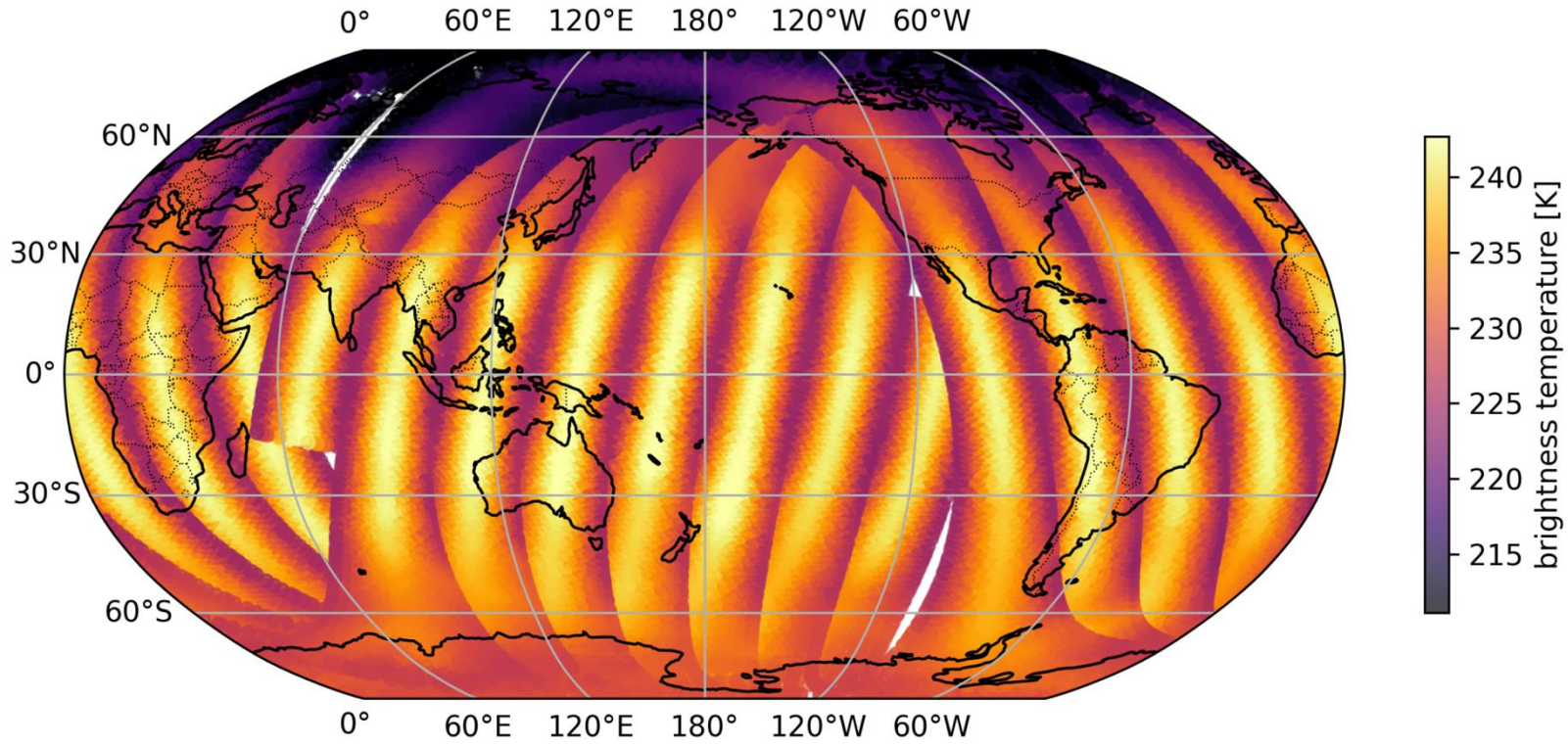


Target

NPP ATMS ch7

ECMWF

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

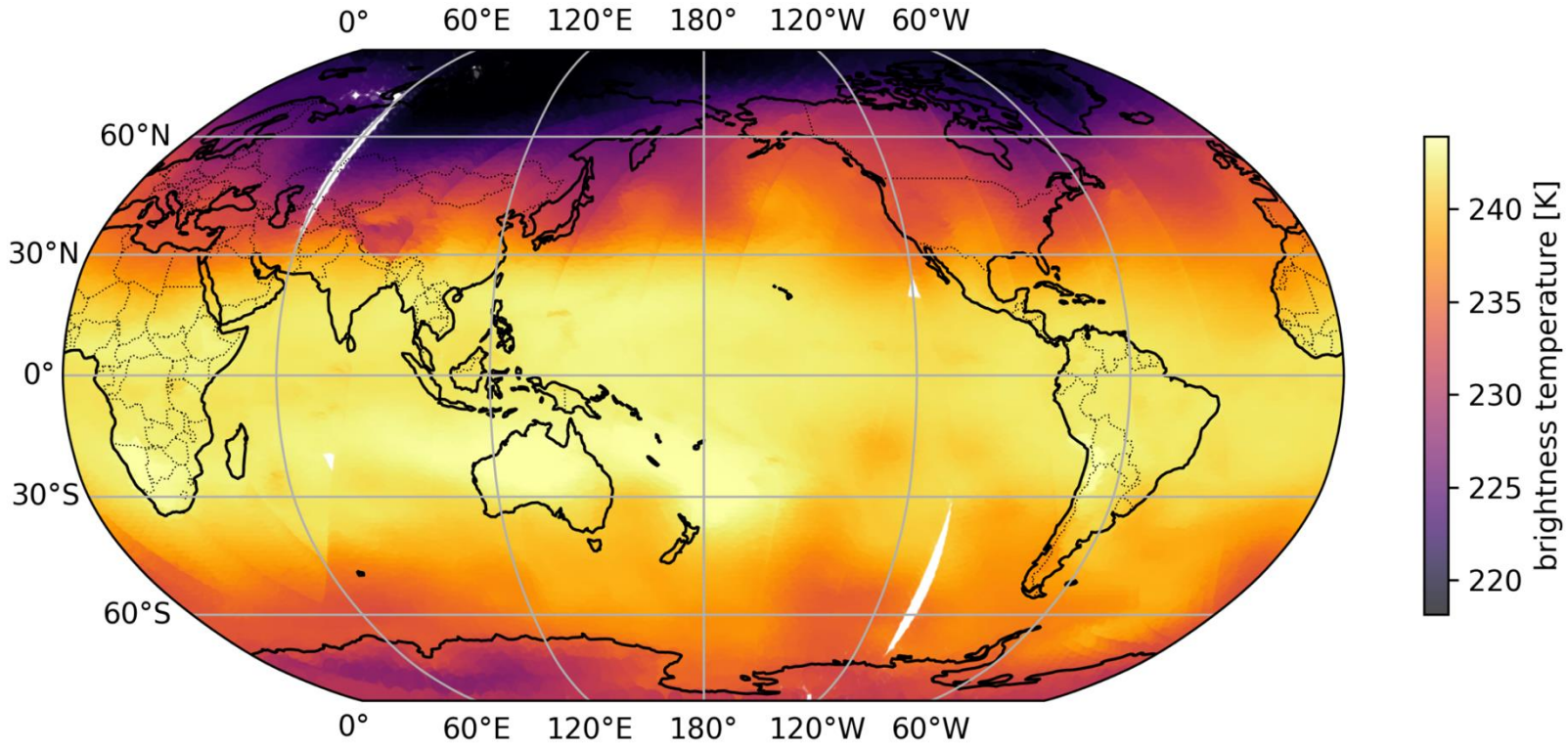


Prediction

NPP ATMS ch7

ECMWF

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



Prediction **with zenith angle set to zero**

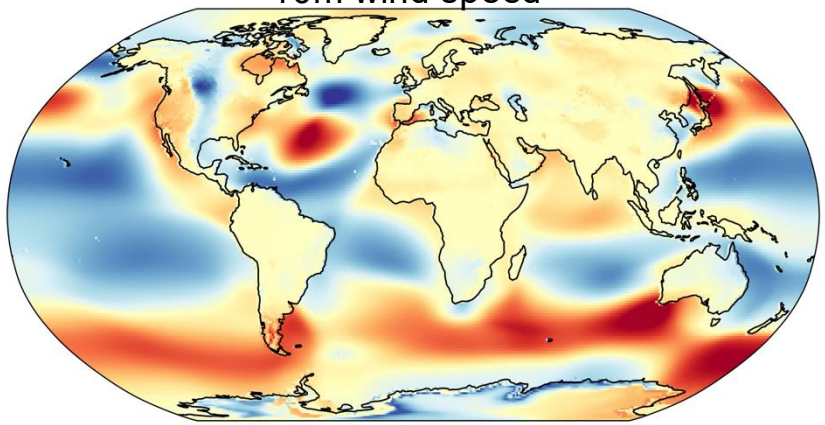
NPP ATMS ch7

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

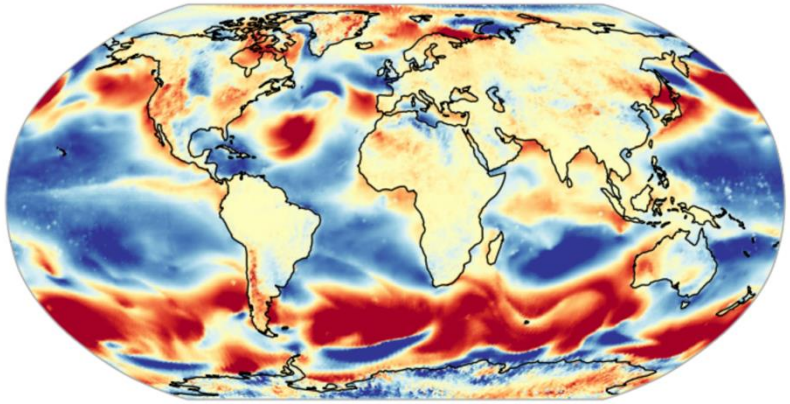
ECMWF

AI-DOP: still progressing rapidly, but can it challenge the state of the art?

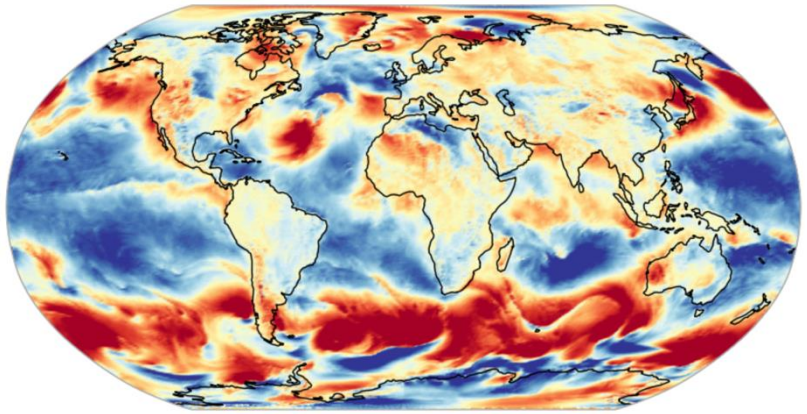
10m wind speed



Results from Sept 2024

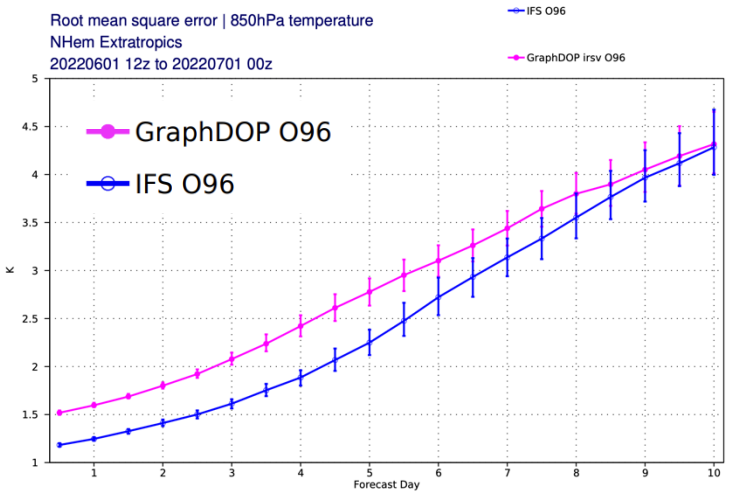


Sept 2025

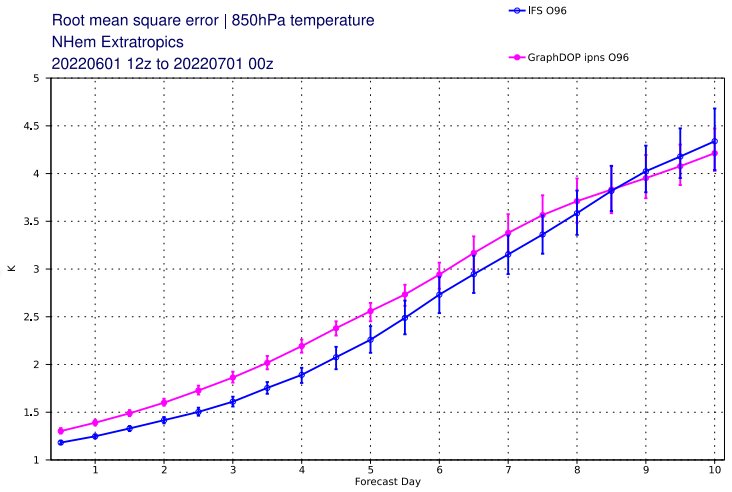


IFS analysis

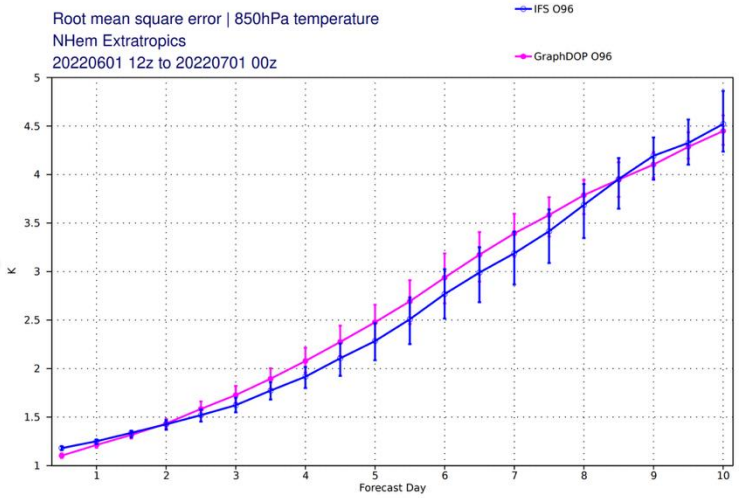
T850 N.Hem RMSE:



May 2025



June 2025



August 2025

ECMWF

Encouraging results, many open research questions to be addressed

Summary

- The “AI revolution” in weather prediction has mainly affected the model component so far, but attention is increasingly turning towards data assimilation and observations.

- Machine learning offers **many opportunities** to improve our use of satellite data for weather prediction, from individual learned components inside traditional NWP systems to more radical end-to-end approaches.

- Several of these ML developments are already operational with more in the pipeline.

- **It’s an exciting time to be a researcher in weather prediction**
  - lots of low hanging fruit!