AI models at ECMWF

Matthew Chantry – ML Coordinator & Head of Innovation Platform

Thanks to:
Zied Ben Bouallegue, Linus Magnusson, Simon Lang, Mark Rodwell, Mariana Clare, Mihai Alexe, Jesper Dramsch, Baudouin Raoult, Florian Pinault, Christian Lessig, Michael Maier-Gerber, Massimo Bonavita, Florian Pappenberger and many many more

Is reanalysis sufficient to learn a global forecasting system?

Simple problem framing.

- Given state of ERA5 at a random point in time, x(t). Typically u, v, t, z, q on ~10 pressure levels and 2t, 10u/v, sp.
- Construct a model F, a neural network parametrised by weights. Big models, $O(10^7)$ parameters.
- Predict a future state of ERA5, $x(t+dt) \simeq F(x)$. Typically 6-hour timestep!
- Seek to minimise $[x(t+dt) - F(x(t))]^2$ using gradient descent.
  - i.e. change the weights in such a way to decrease the MSE.
- Randomly draw a new x and repeat. Many many times, passing through ERA5 O(100) times.

A brief history of data-driven models

Defining the dataset, split, headline fields and metrics

2020 WeatherBench

2018
Exploring the idea

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

Nov 2022
Tropical cyclones

Huawei – PanguWeather
0.25° hourly product

"More accurate tracks" than the IFS.

Dec 2022
Extensive predictions

Deepmind – GraphCast
0.25° 6-hour

Many variables and pressure levels with comparable skill to IFS.

Jan 2023
Global & Limited Area

Microsoft – ClimaX

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

Apr 2023
7-day+ scores improve

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

Improves on GraphCast for longer leadtimes (still deterministic)

Jun 2023
Spherical harmonics

NVIDIA – SFNO
0.25° 6-hour product

Extension of FourCastNet to Spherical harmonics, improved stability

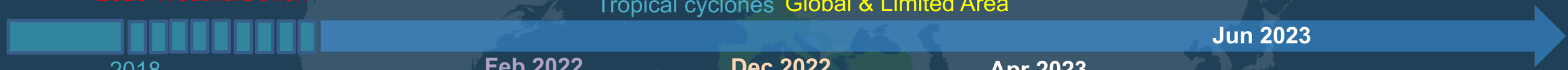
Diffusion modelling

Alibaba – SwinRDM
0.25° 6-hour product

Sharp spatial features

Last months
FuXi
AtmoRep
FuXi-extreme
NeuralGCM

A brief history of data-driven models



2018
Exploring the idea

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

2020 WeatherBench

Defining the dataset, split, headline fields and metrics

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

Nov 2022
Tropical cyclones

Huawei – PanguWeather
0.25° hourly product

"More accurate tracks" than the IFS.

Dec 2022
Extensive predictions

Deepmind – GraphCast
0.25° 6-hour

Many variables and pressure levels with comparable skill to IFS.

Jan 2023
Global & Limited Area

Microsoft – ClimaX

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

Apr 2023
7-day+ scores improve

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

Improves on GraphCast for longer leadtimes (still deterministic)

Jun 2023
Spherical harmonics

NVIDIA – SFNO
0.25° 6-hour product

Extension of FourCastNet to Spherical harmonics, improved stability

Diffusion modelling

Alibaba – SwinRDM
0.25° 6-hour product

Sharp spatial features

Last months
FuXi
AtmoRep
FuXi-extreme
NeuralGCM

Some first questions

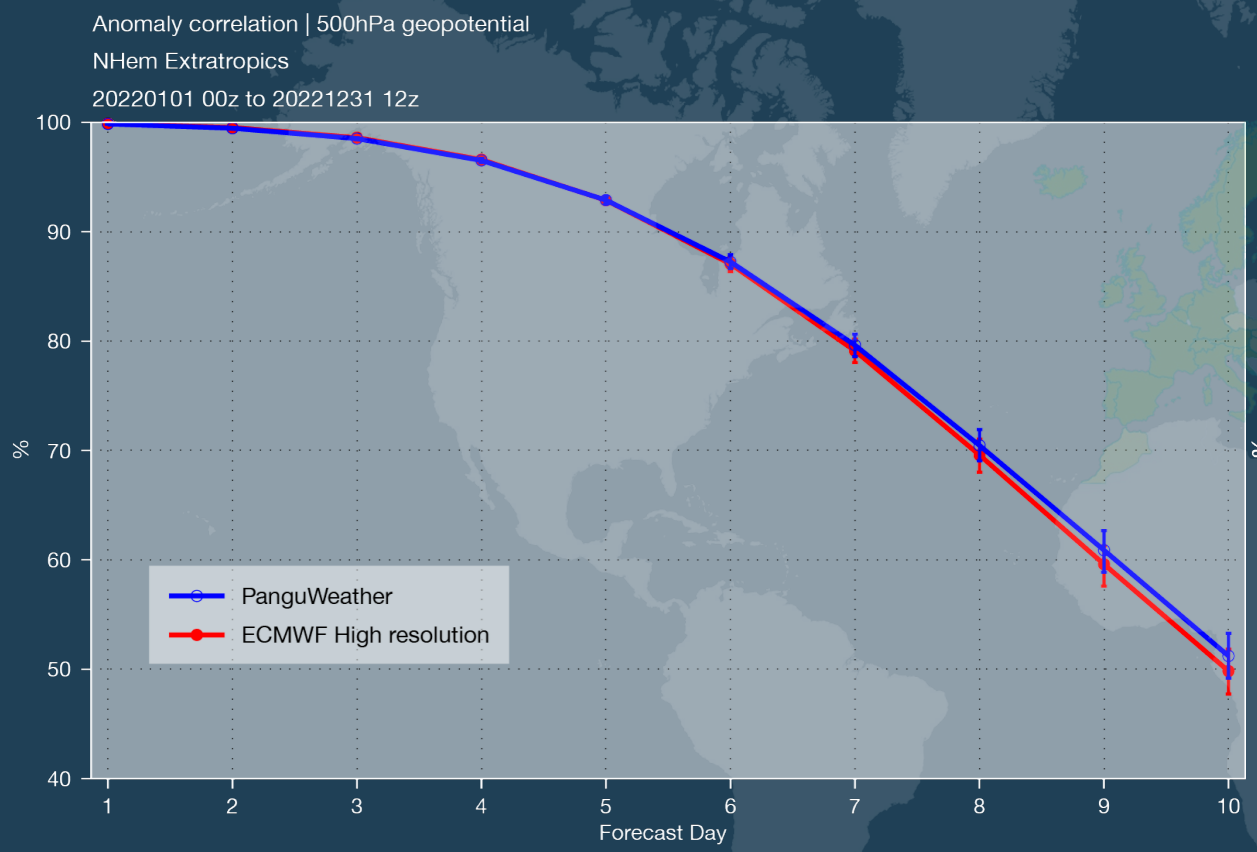
- Can we duplicate scores?
  - How big of an advantage is long-window DA?
  - How good is performance from operational analysis?
  - Are tropical cyclone results reproducible?

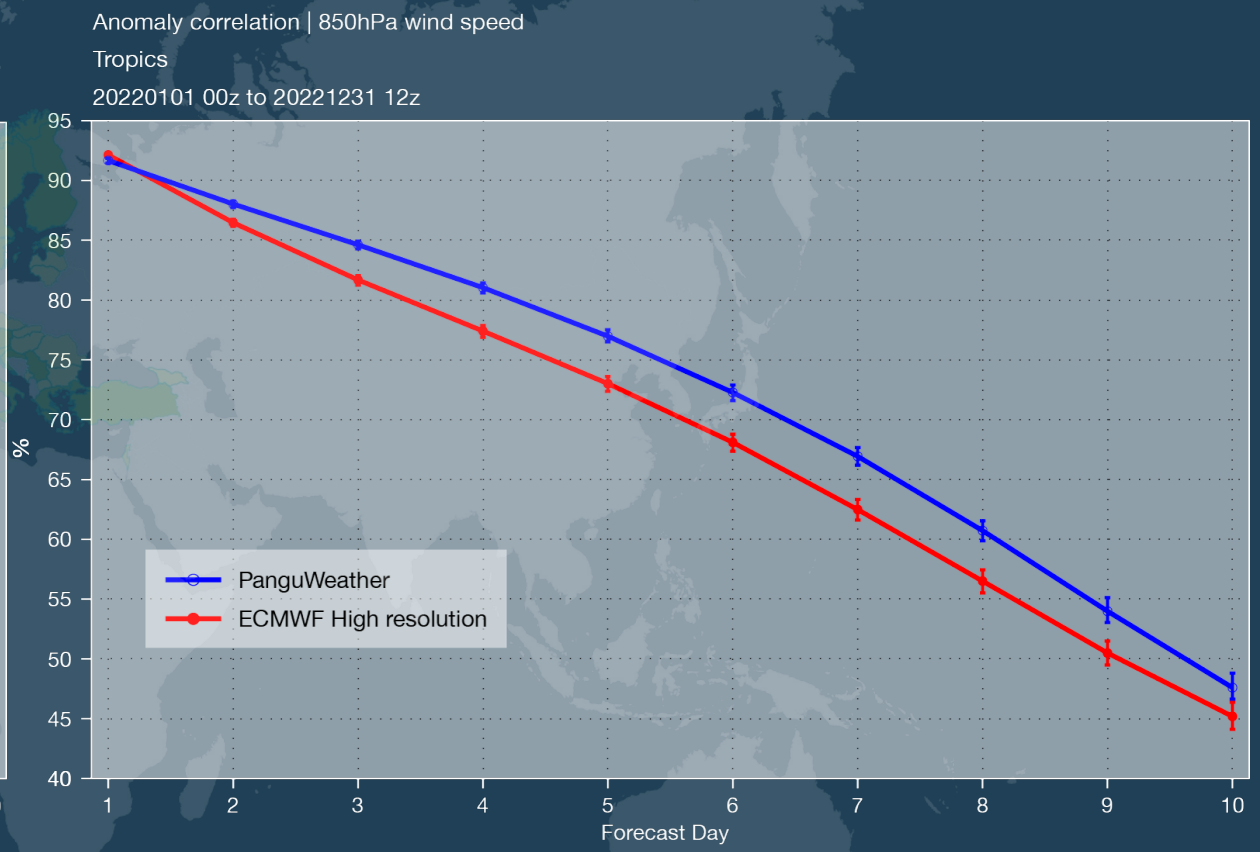
- Do data-driven models behave like physical NWP systems?
  - What do the spectra look like?
  - Do these models behave like an ensemble mean?
  - How physically consistent are they?

- What about precipitation?

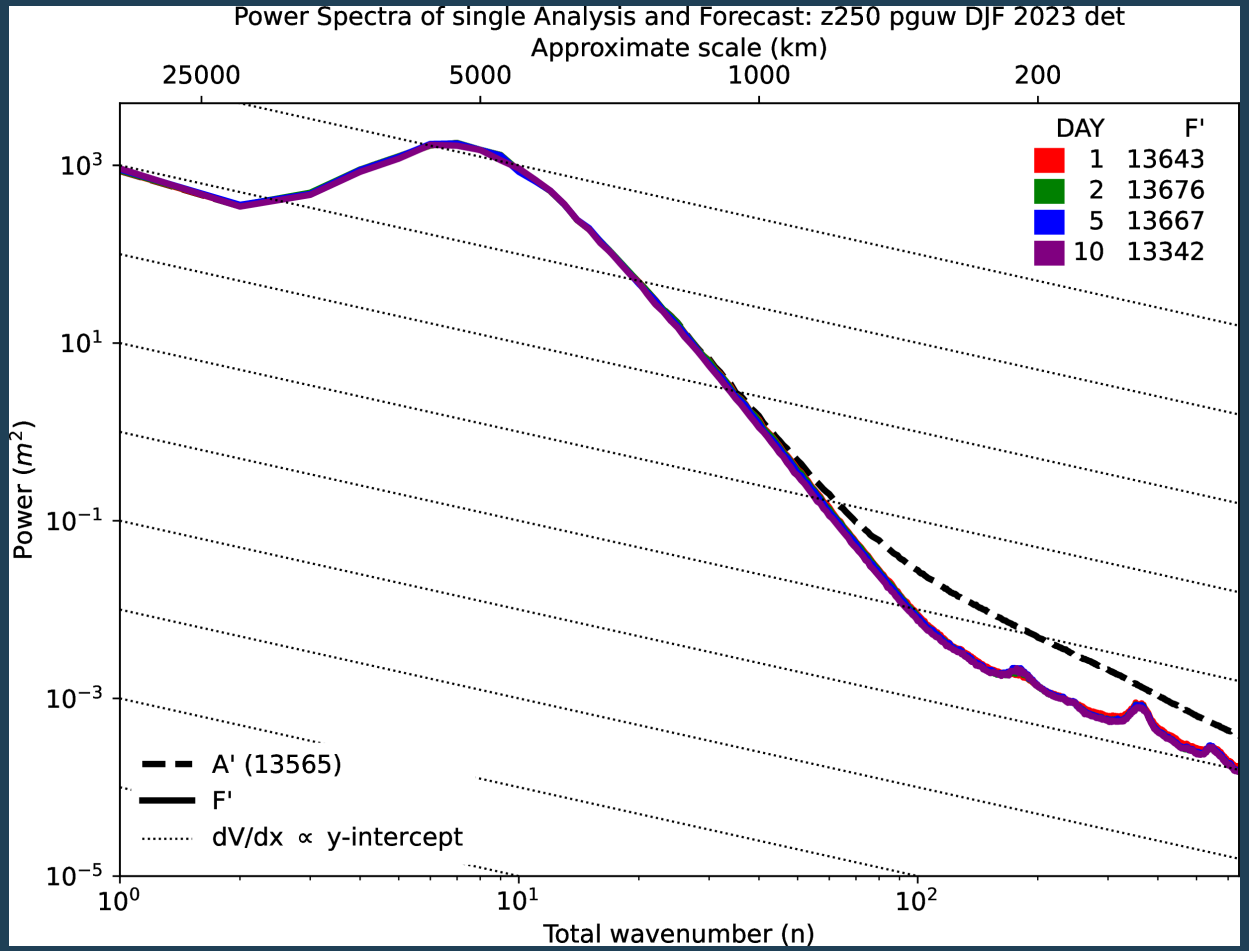
- What about extreme events?

What the analysis is showing: **an undeniable skill**

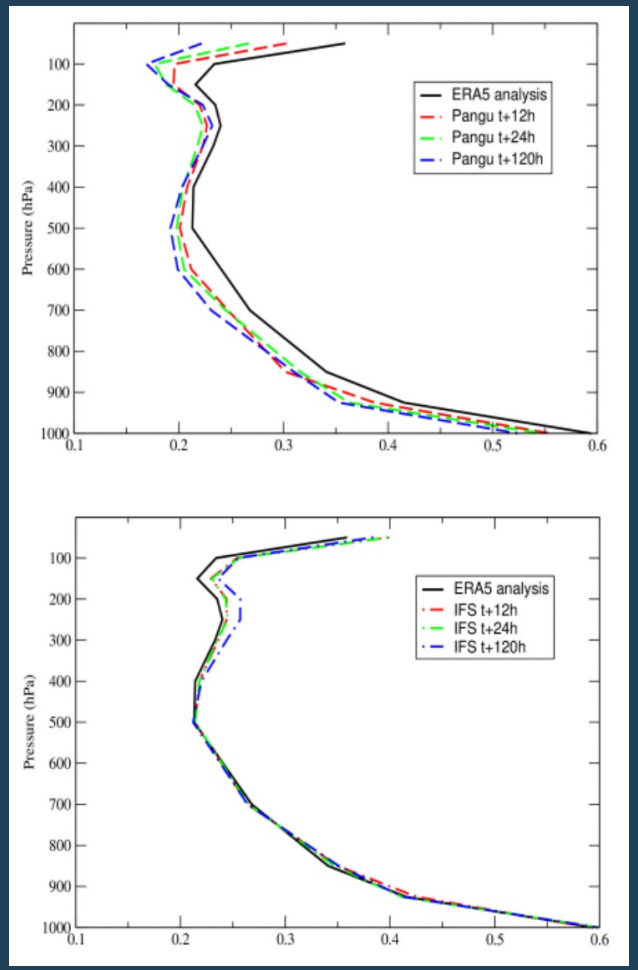




Physical consistency



Ratio of ageostrophic and geostrophic winds

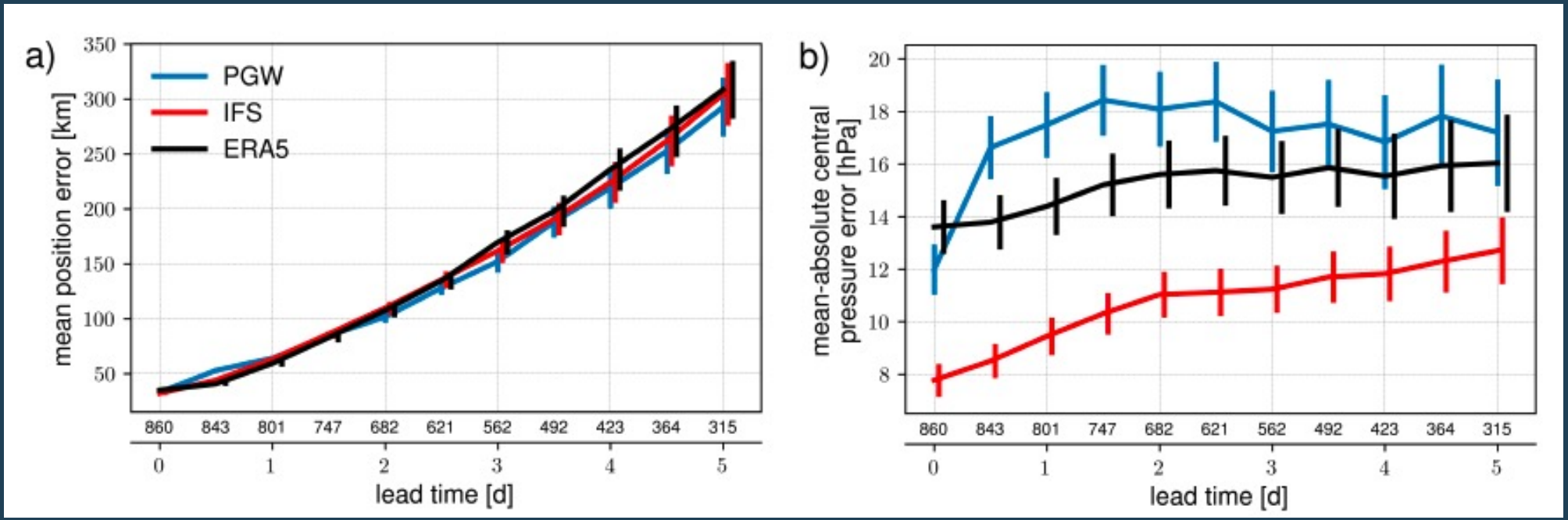


See Bonavita 2023

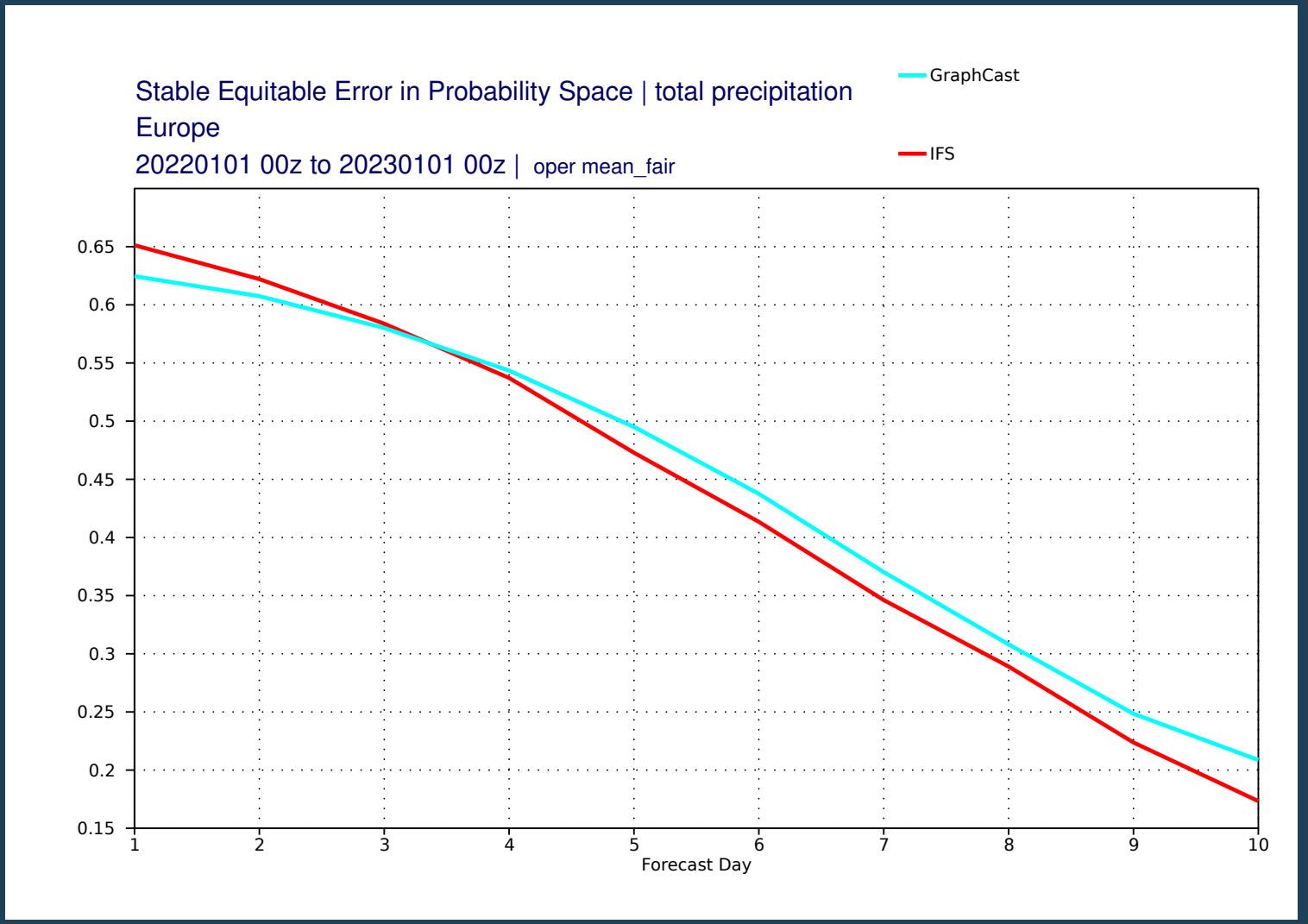
Tropical cyclone verification

Position error

Intensity bias

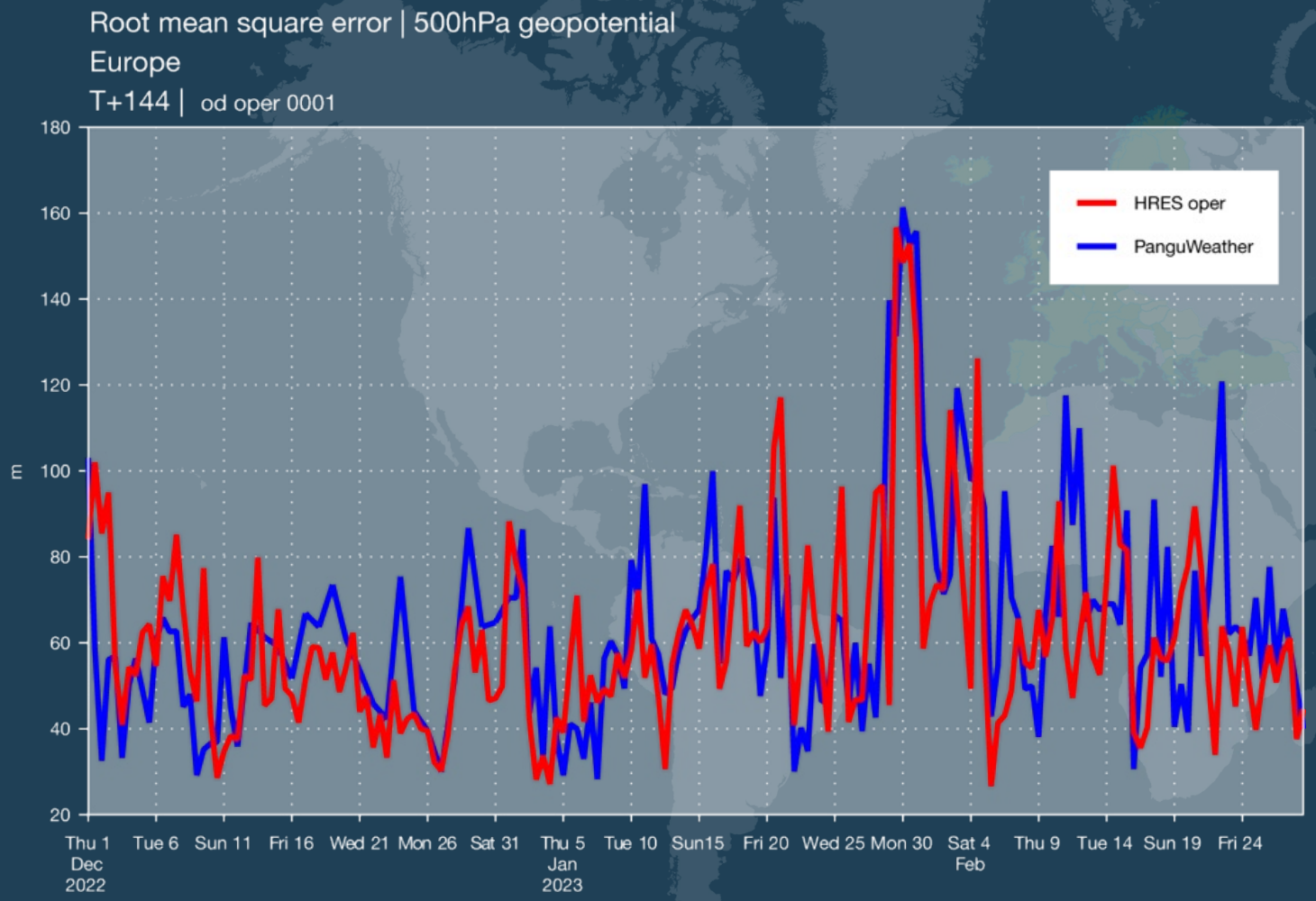


Precipitation

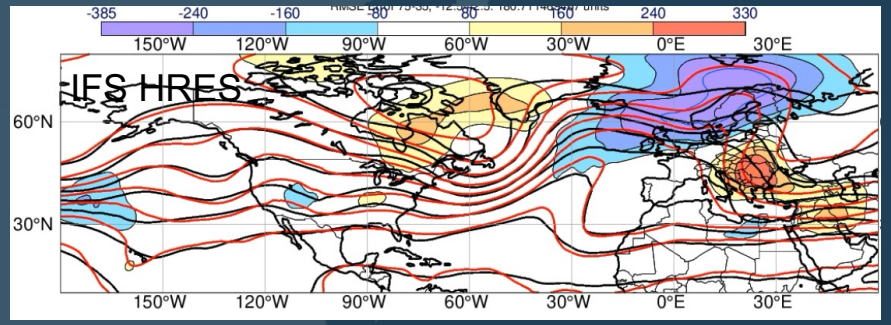


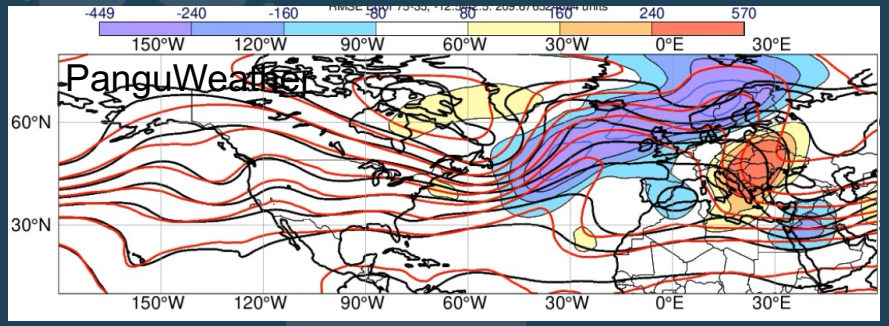
Time-series of day 6, RMSE over Europe

Same starting point....similar results



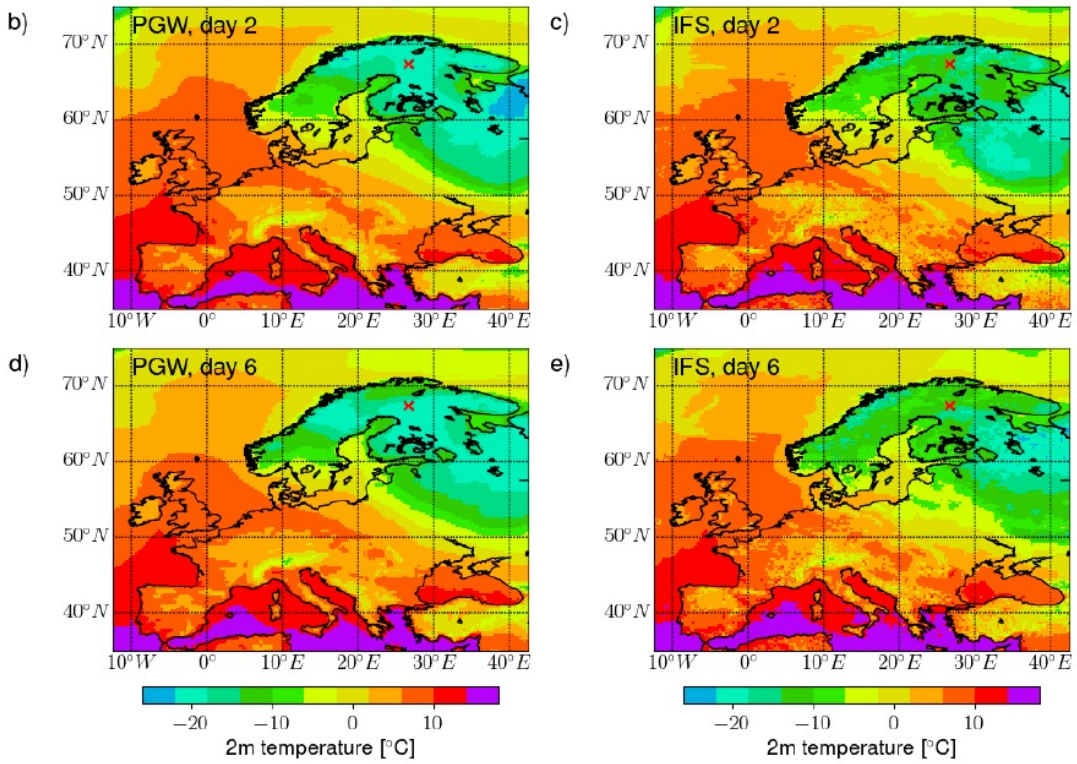
+144h forecast errors 30 January 00UTC



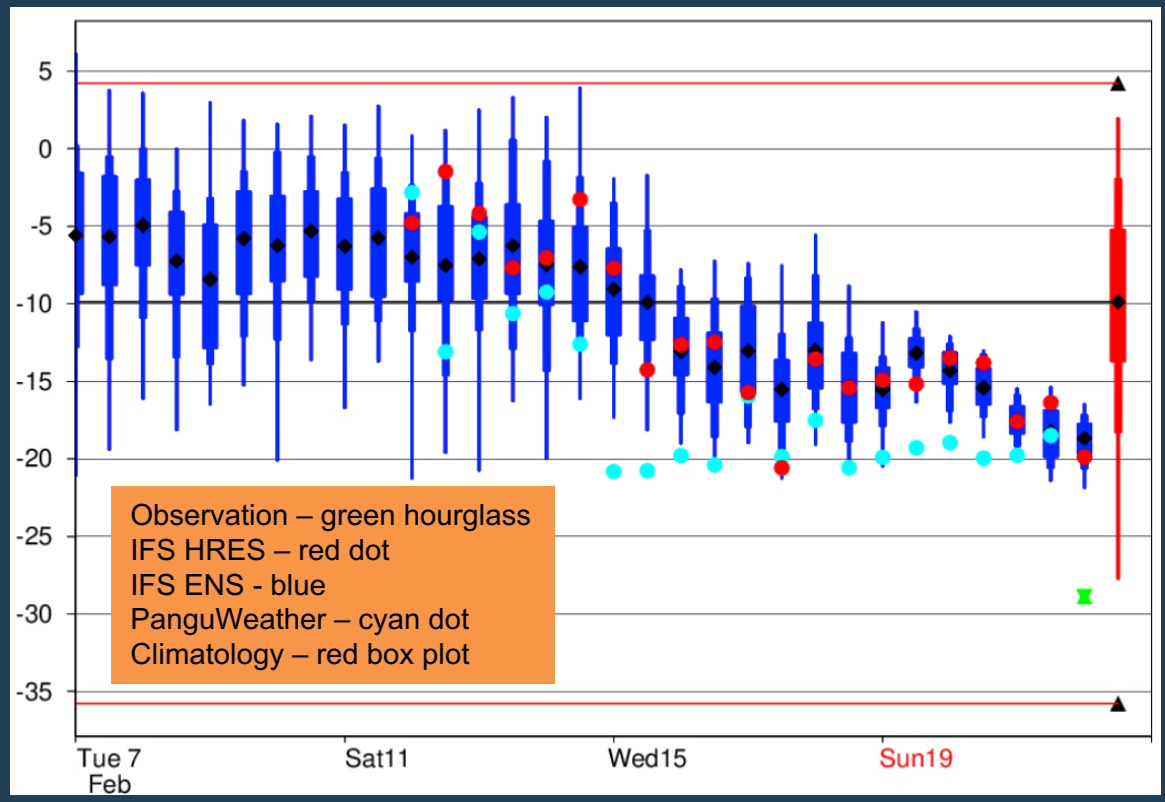


Cold snap over northern Europe Feb 2023

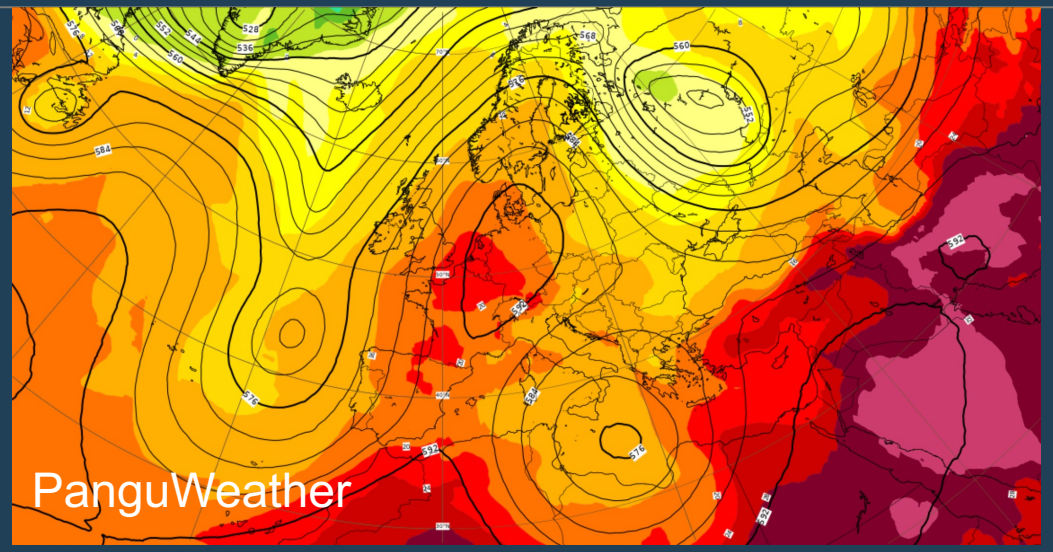
PanguWeather IFS HRES

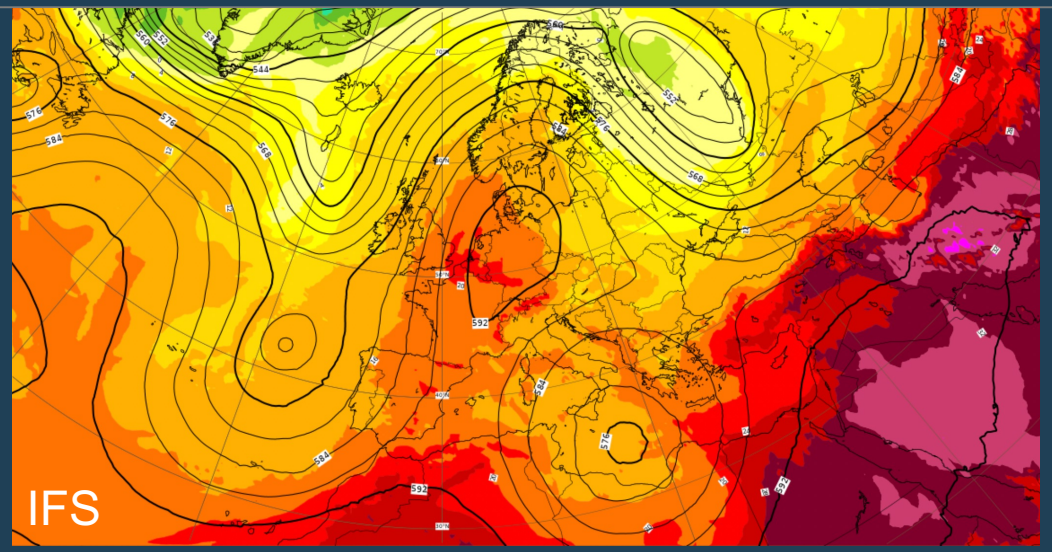


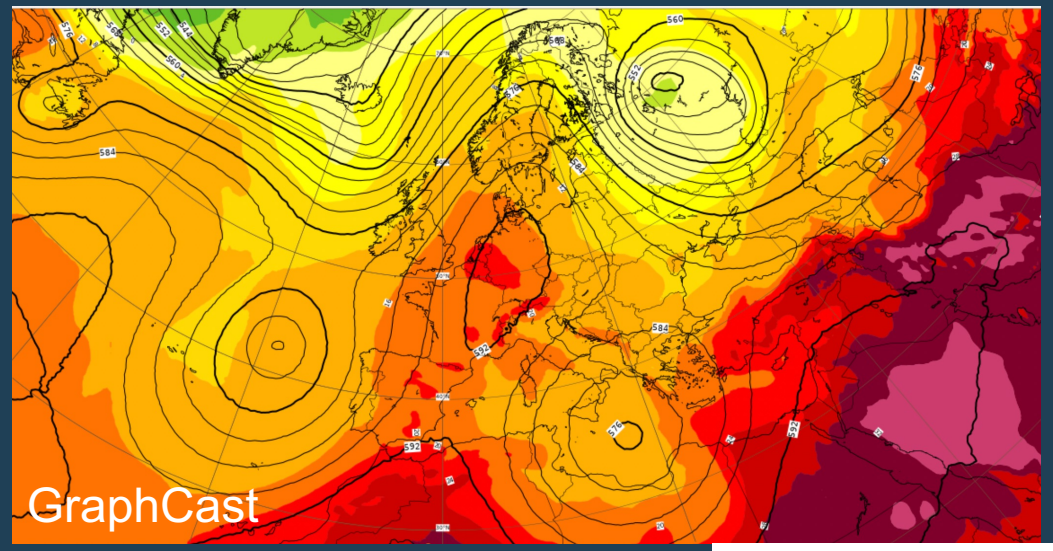
2m temperature Sodankyla 22 February 00UTC

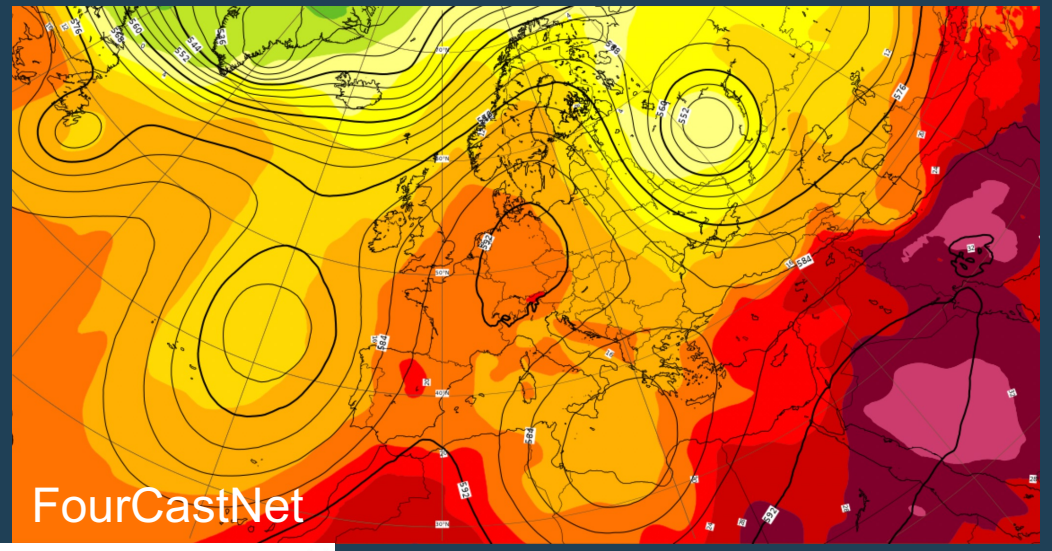


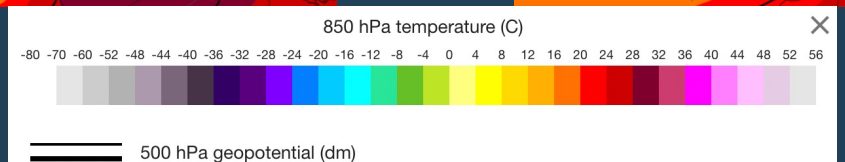
Now available on charts.ecmwf.int



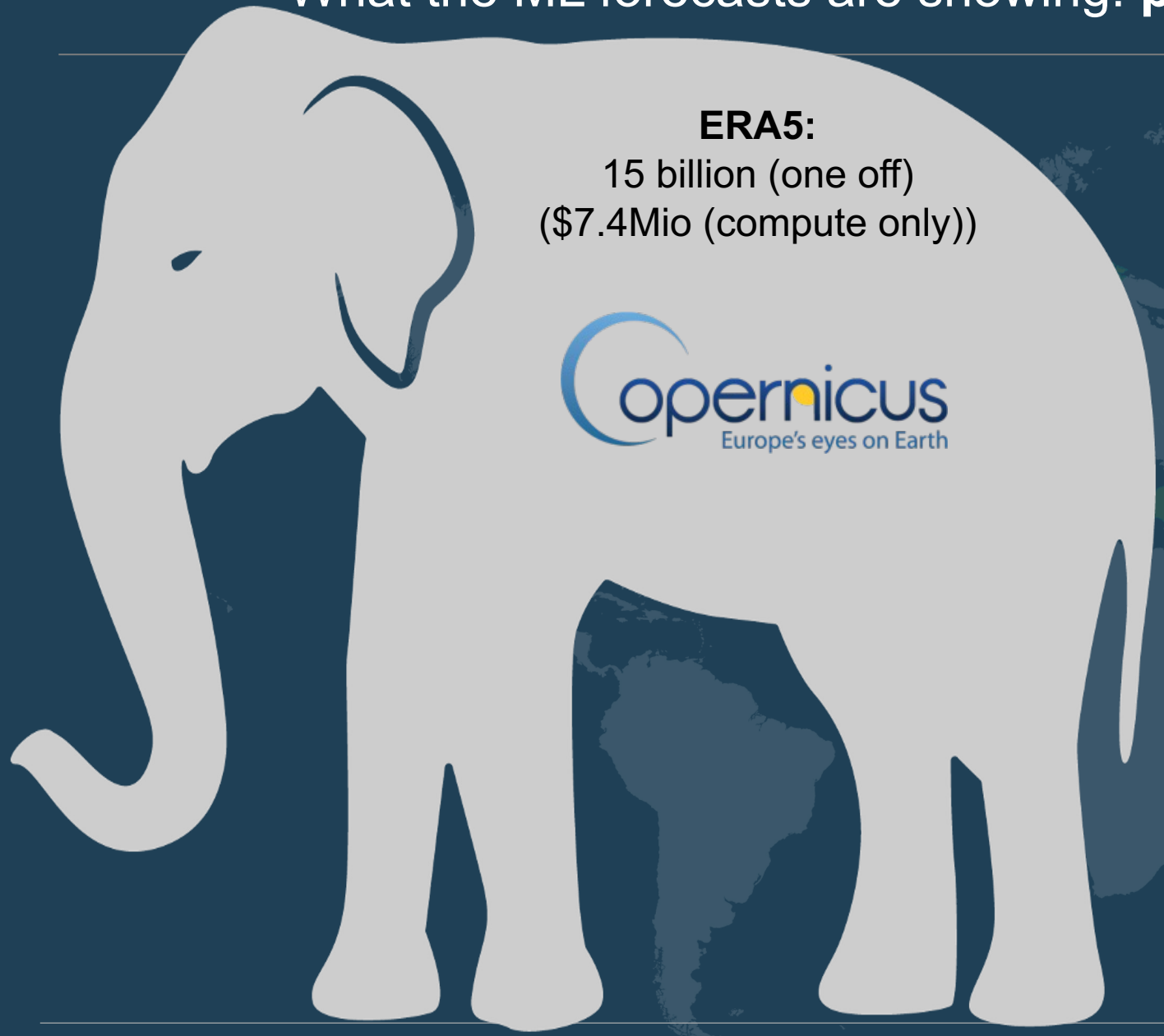








What the ML forecasts are showing: **potential gain in time and energy**



ERA5:
15 billion (one off)
($7.4Mio (compute only))

Copernicus
Europe's eyes on Earth

ECMWF HRES:
180 000 ($90)
per forecast

Pangu:
0.3 (<¢1)
per forecast

Embracing the technology… building the AIFS

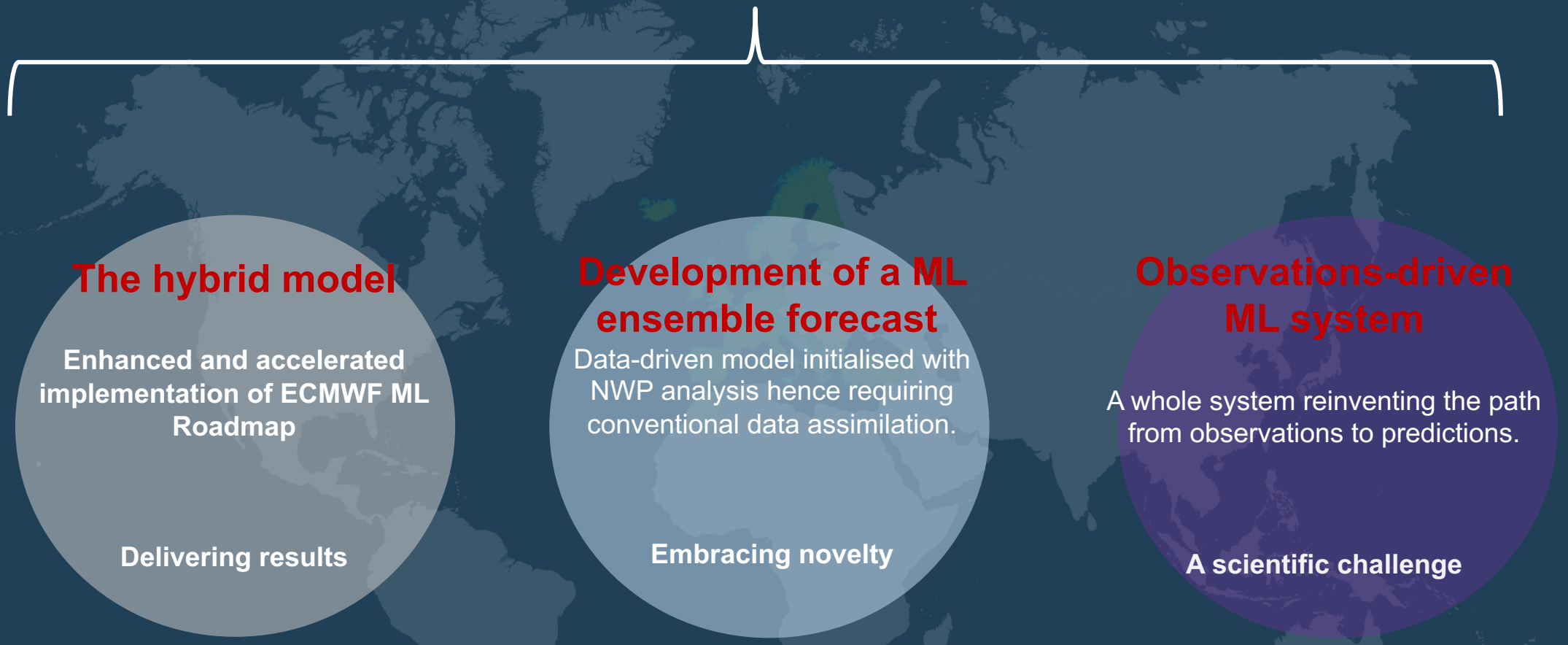
Why?

- We view that this will be a component of how medium-range forecasts are made.
  - Still unclear the balance between physics and data-driven forecasts.

- Further work is still required.
  - Building reliable ensembles.
  - Improving representation of extreme winds/precip.
  - Utilise knowledge of data strengths.
  - Oceans and extended range predictions

- Need to learn to develop these systems ourselves.
  - Like conventional modelling, there will be a regular cycle of regular model upgrades.

Project overview: different paths **towards a ML ensemble prediction at ECMWF**



Resourcing

- All three project aspects are important.

- Utilise existing ML talent.
  - A lot of enthusiasm to work in this area.

- Using skills from across the centre.
  - Verification.
  - Building ensembles.
  - Data pipelines.
  - Production.

- ~15FTEs.
  - Based on available resources, hard decisions on slowing development in some areas.

Computing is key
To train, and particularly develop data-driven models will require significant numbers of GPUs.

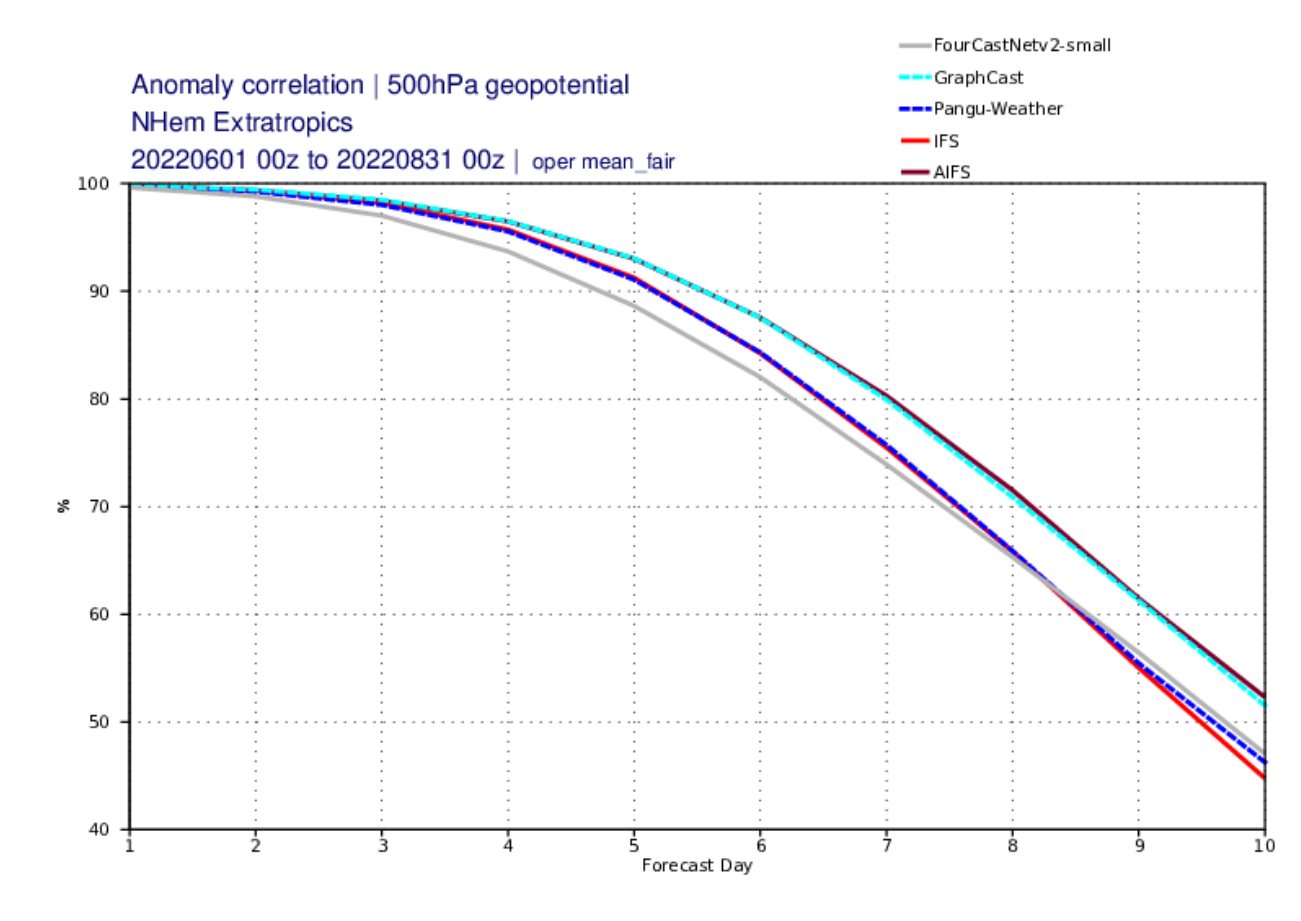
The AIFS

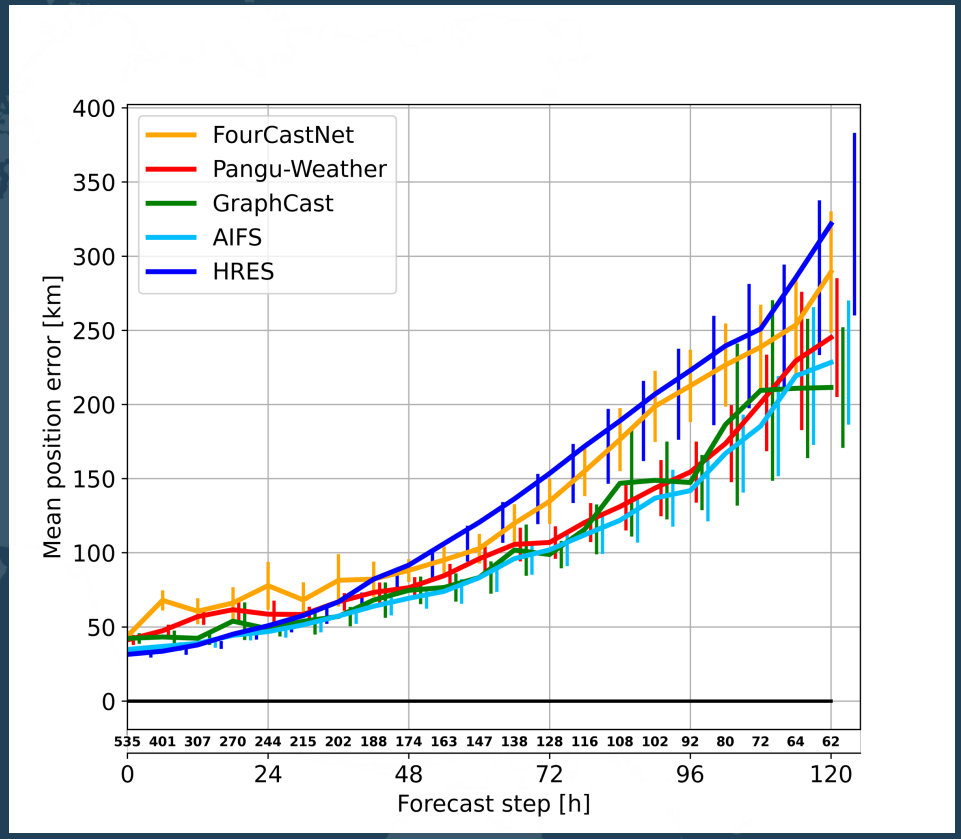
- Particularly inspired by Keisler & GraphCast.
  - GNNs naturally encode the sphere and allow use of optimal grids.
  - Several innovations, to be submitted for publication soon.

- Only 1° but rivals atmospheric scores of others.
  - Significantly cheaper to train, useful for exploring ensemble approaches.

- Already running daily and producing live and open forecasts.
  - As with other ML models, we want as many eyes on forecasts as possible.

The AIFS

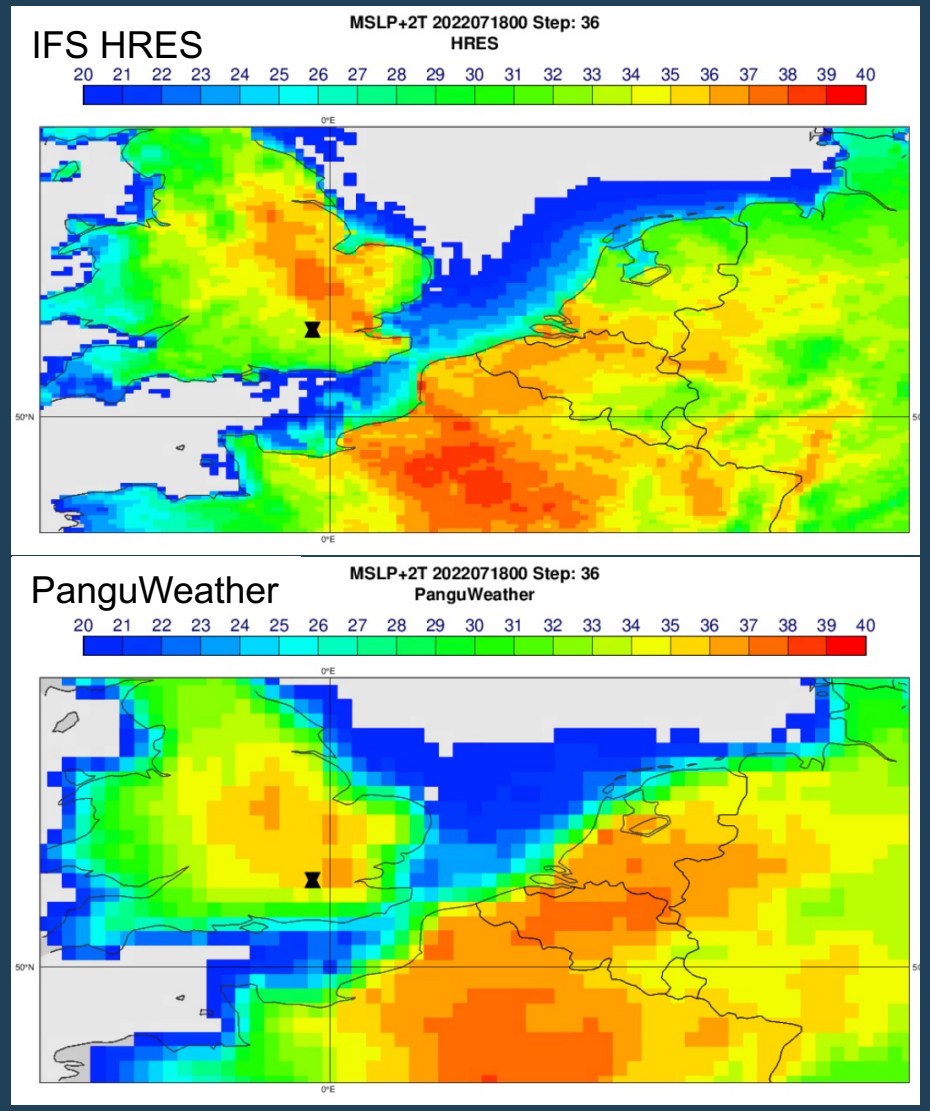




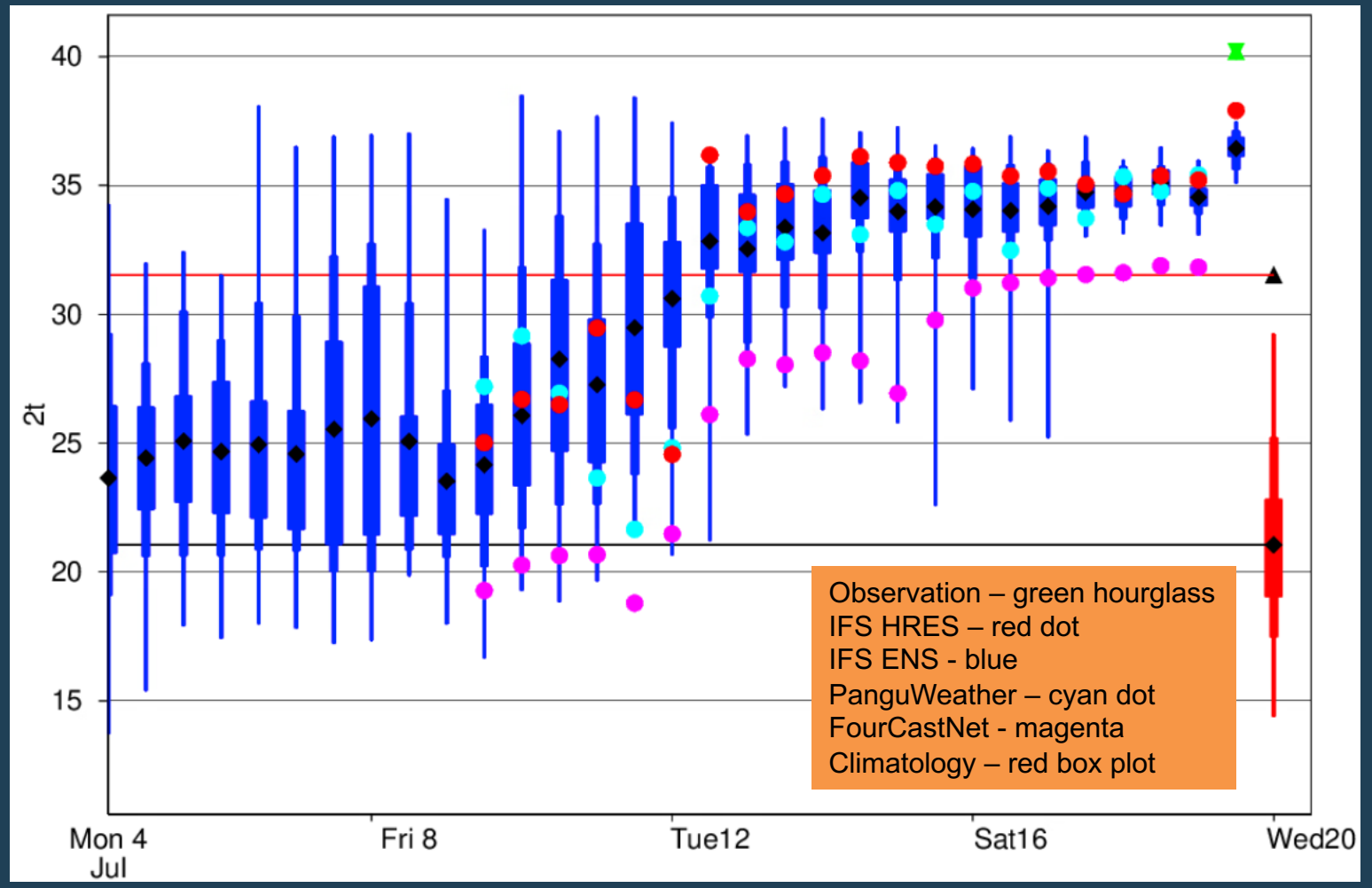
Caveats: reduced number of TC (resolution induced) and underestimation of intensity.


What’s next?

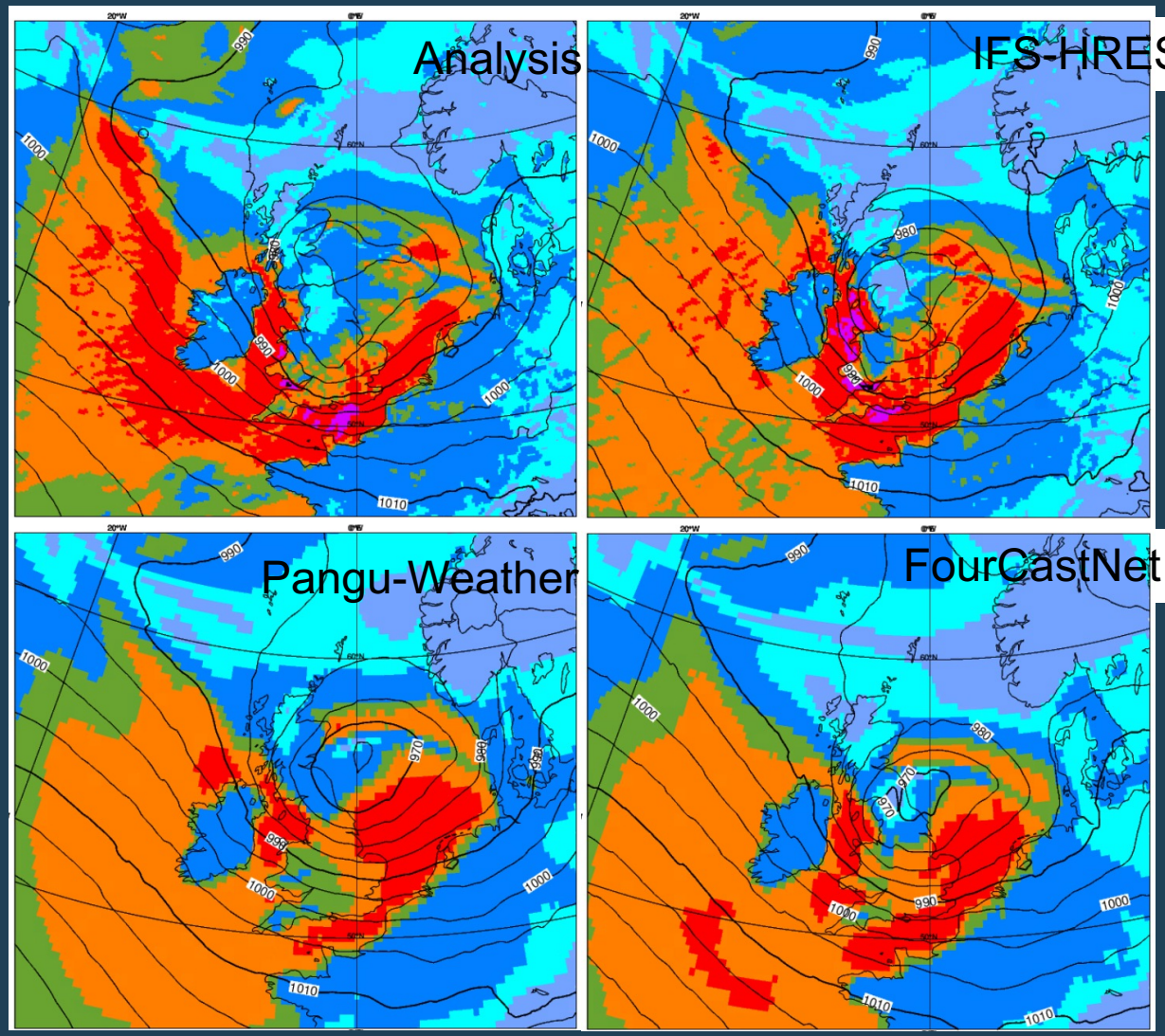
UK heatwave 2022



2m temperature Heathrow 19 July 12UTC



Storm Eunice (2.5-day forecasts valid18th Feb 2022 12UTC)



Maximum mean wind Heathrow 18 Feb (00, 06, 12 and 18