Forecasting Global Weather with Graph Neural Networks

Ryan Keisler

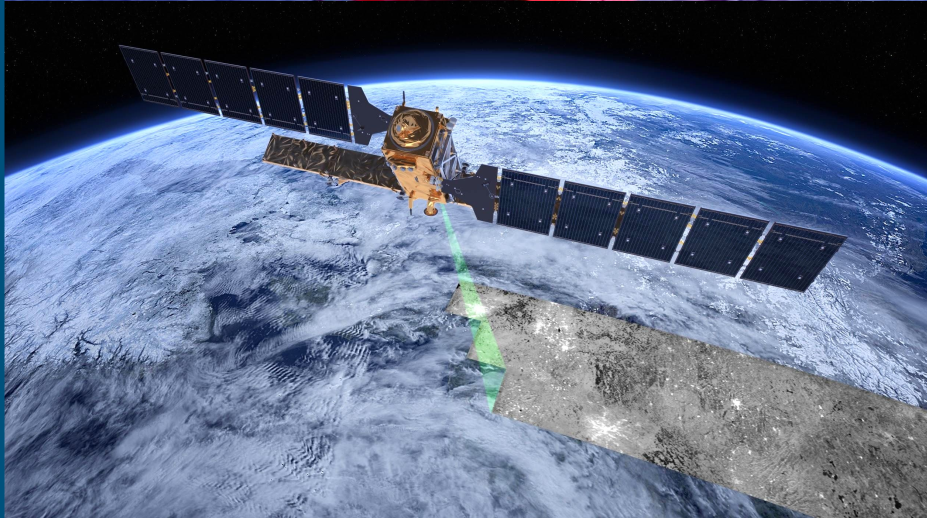
My background

10 years in physics + cosmology

(building things, data analysis, lots of stats)

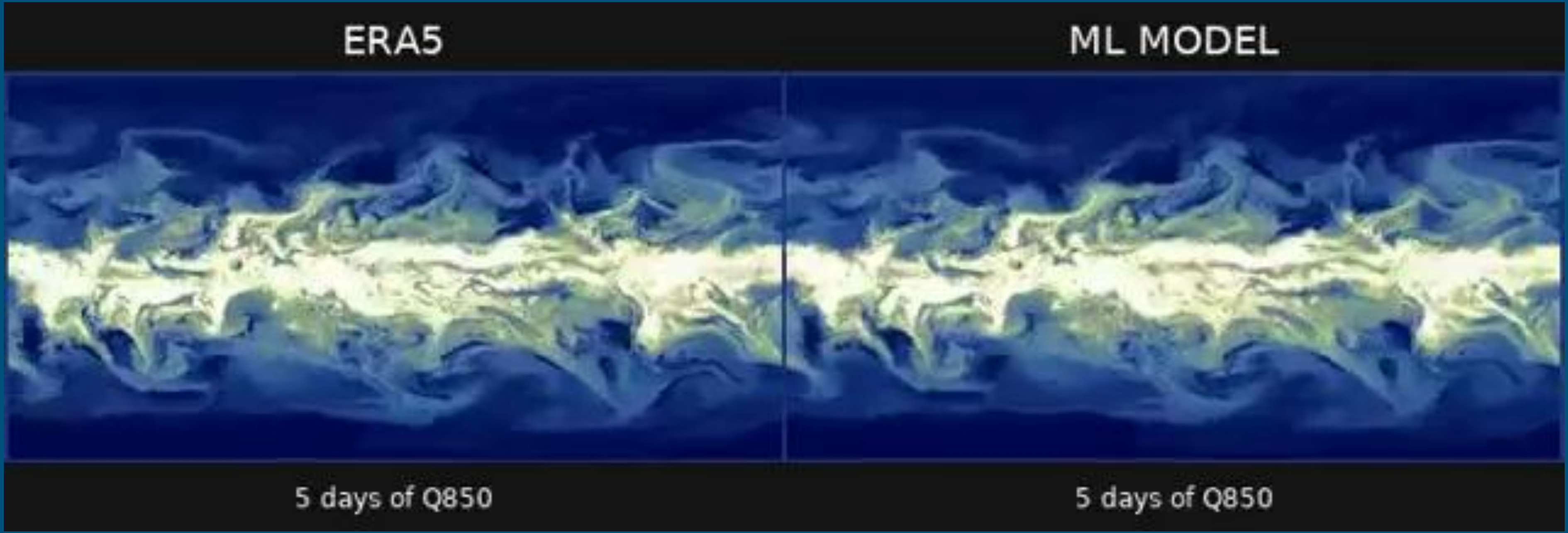
7 years in industry

(satellite imagery, weather, physical & statistical models)



Quick look at main result

Main result:
a new data-driven system for forecasting global weather



See bit.ly/graph_weather for more.

Motivation

Motivation

- Intellectual
- New opportunities
- Wouldn’t it be cool if…

Motivation

- Intellectual

- New opportunities

- Wouldn’t it be cool if…

  for <$1, I could run a high-quality, global weather forecast with an object that is
  - easy to share (some code + 20 MB weights)
  - easy to modify (fine-tune for specific applications)
  - easy to inspect (autodiff)
  - easy to glue (python)

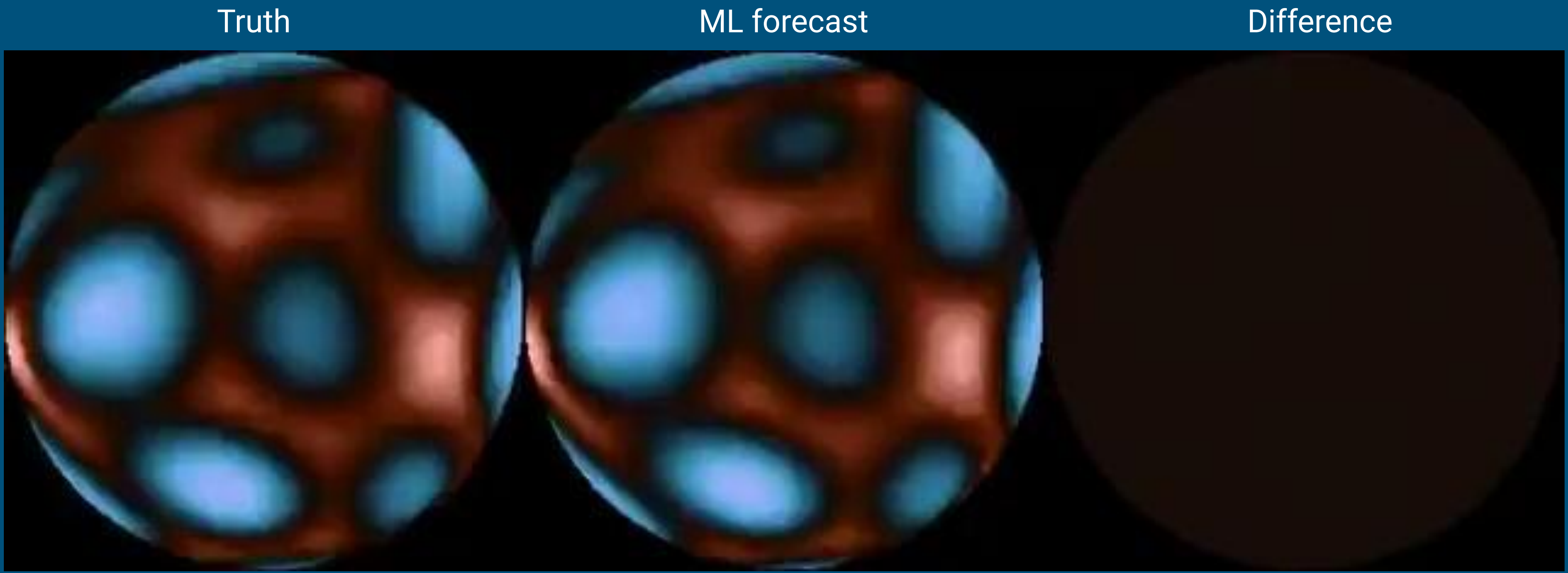
Warming up

Before tackling weather,
let’s try a toy problem.

Before tackling weather,
let's try a toy problem.

Let's learn chaotic dynamics
on the sphere.

Learning chaotic dynamics on the sphere



Learning Kuramoto-Sivashinsky with `jraph + haiku`

Learning Global Weather

Design Philosophy

Traditional NWP works really well so let's:

- model the variables that drive traditional NWP (z, t, q, u, v, w)
- model on a dense physical grid (whatever $ and GPU allow)
- pick an architecture that enables this (~MeshGraphNet)
- pick a dataset that enables this (ERA5)

ERA5

In my opinion, an incredible, under-appreciated scientific achievement!

In this work, I used a 2 TB subset of ERA5:

- Horizontal resolution: 1.0 degrees in lat/lon
- Vertical resolution: 13 pressure levels
- Time: every 3 hours, from 1979 through 2020
- Fields: 6 fields (z, q, t, u, v, w)

Data stored as a single `zarr` array.

Spectrum of weather training data

Observations

(e.g. weather sat data)

"Pure" data, no physical modeling imposed

Reanalysis data

(e.g. ERA5)

A blend of observational data & physical modeling

Forecast data

(e.g. GFS or ECMWF IFS)

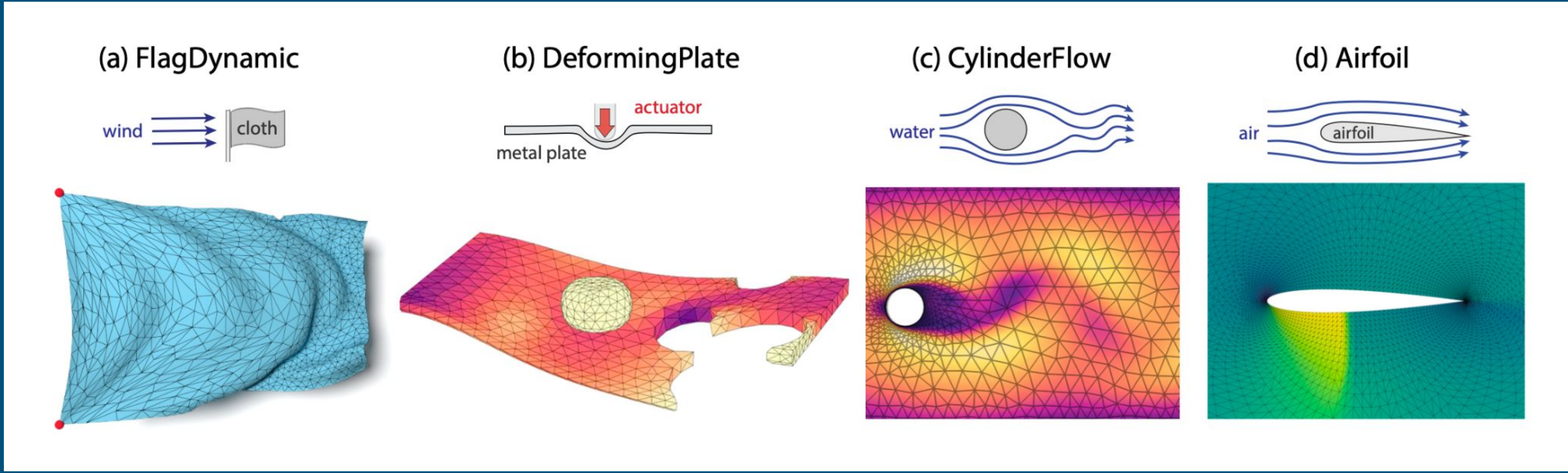
"Pure" physical model

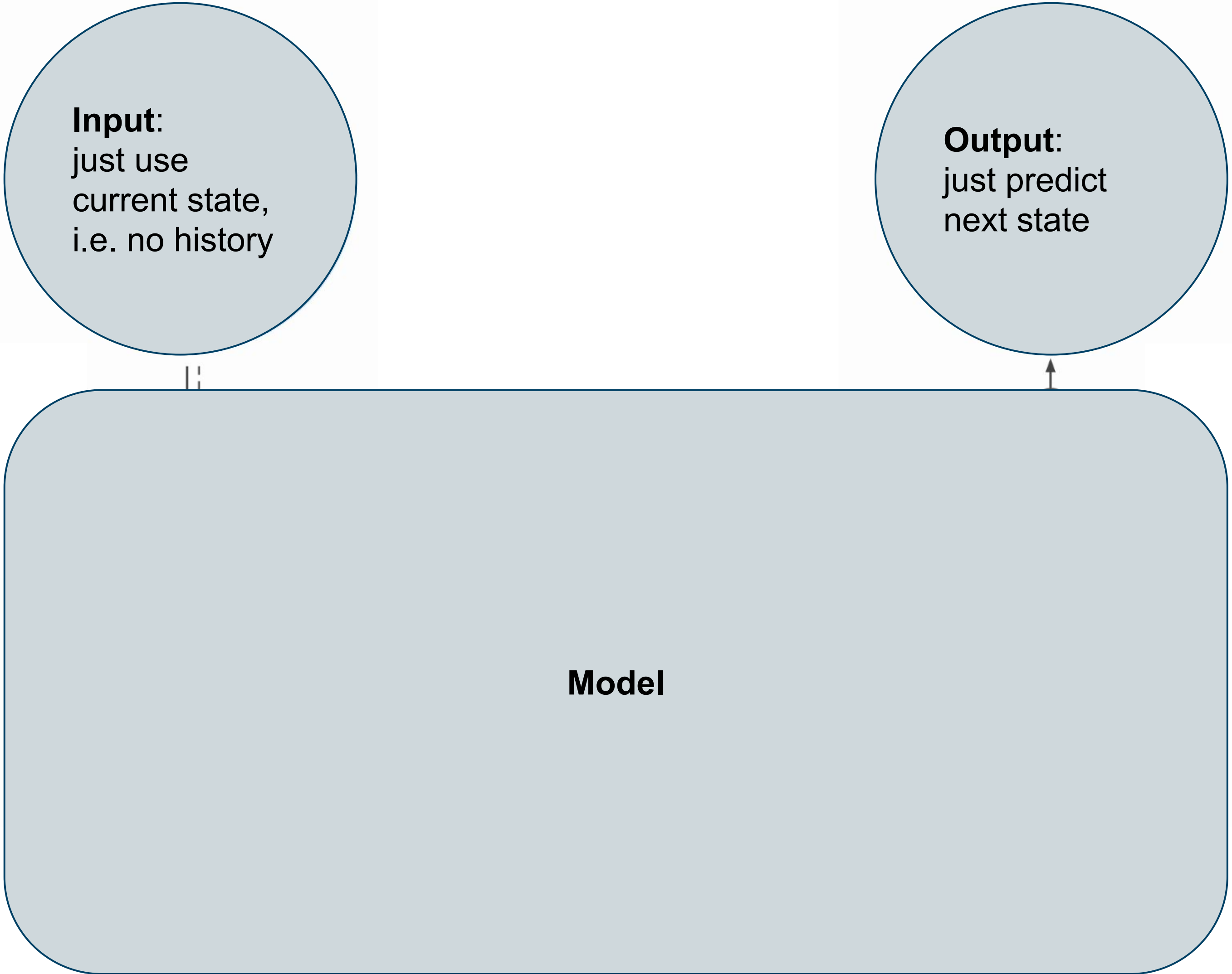
Spectrum of weather training data

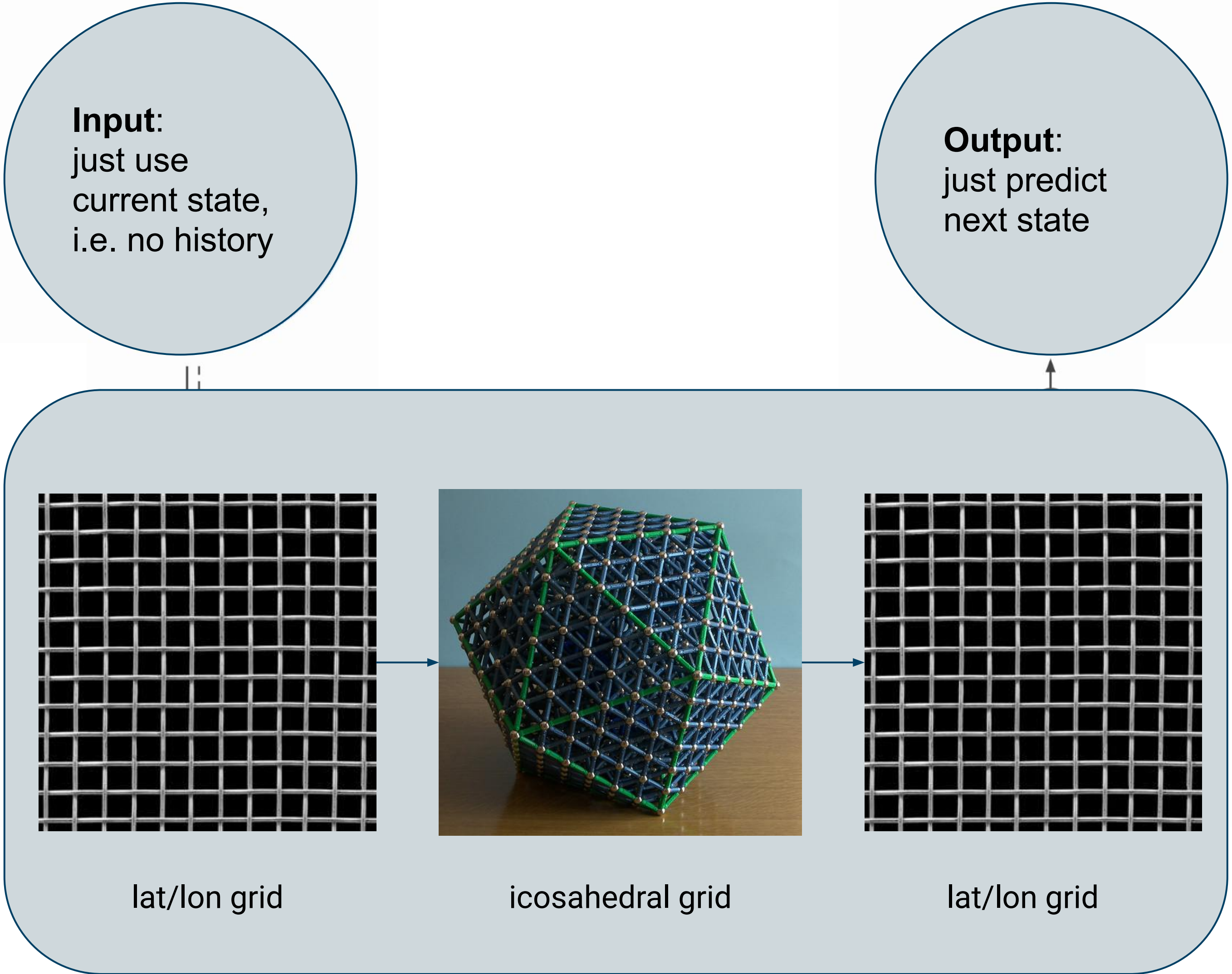
Observations	Reanalysis data	Forecast data
(e.g. weather sat data)	(e.g. ERA5)	(e.g. GFS or ECMWF IFS)
“Pure” data, no physical modeling imposed	A blend of observational data & physical modeling	“Pure” physical model
Best-case ML scenario You outperform NWP!	**Best-case ML scenario** ?	**Best-case ML scenario** You perfectly emulate (but don’t outperform) the NWP engine

Architecture

"Learning Mesh-Based Simulation with Graph Networks"
by Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, Peter W. Battaglia
arXiv:2010.03409

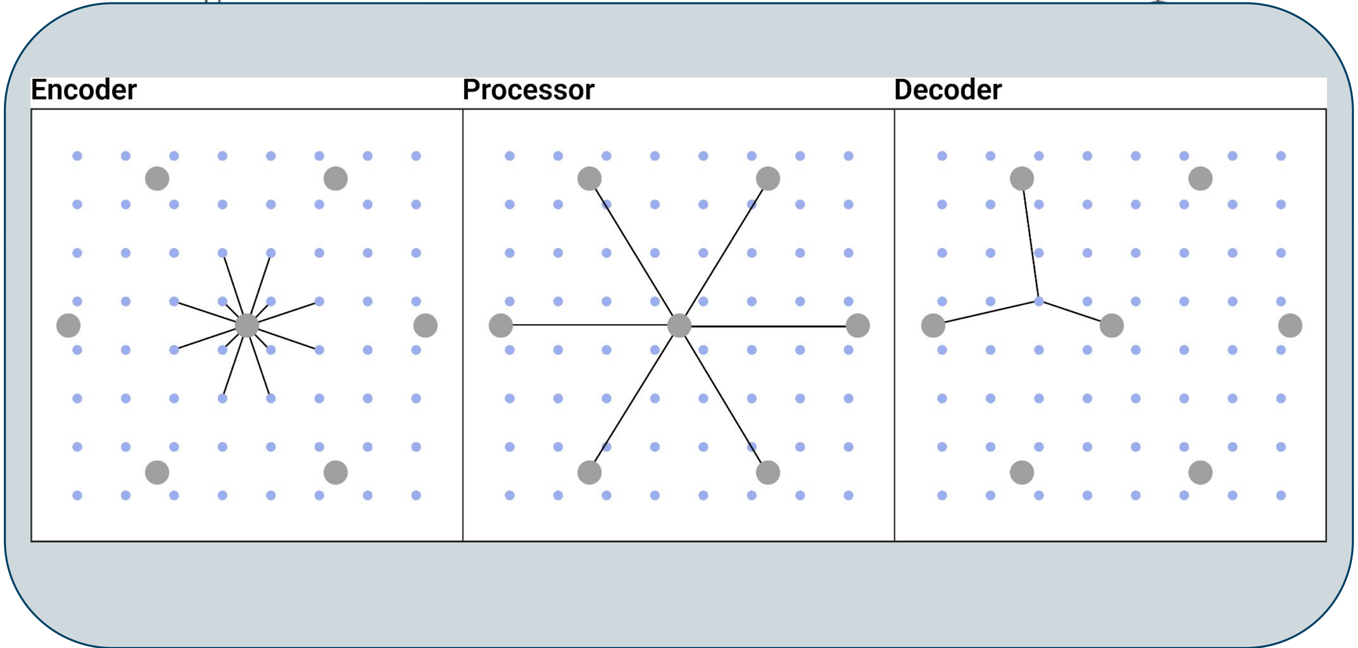


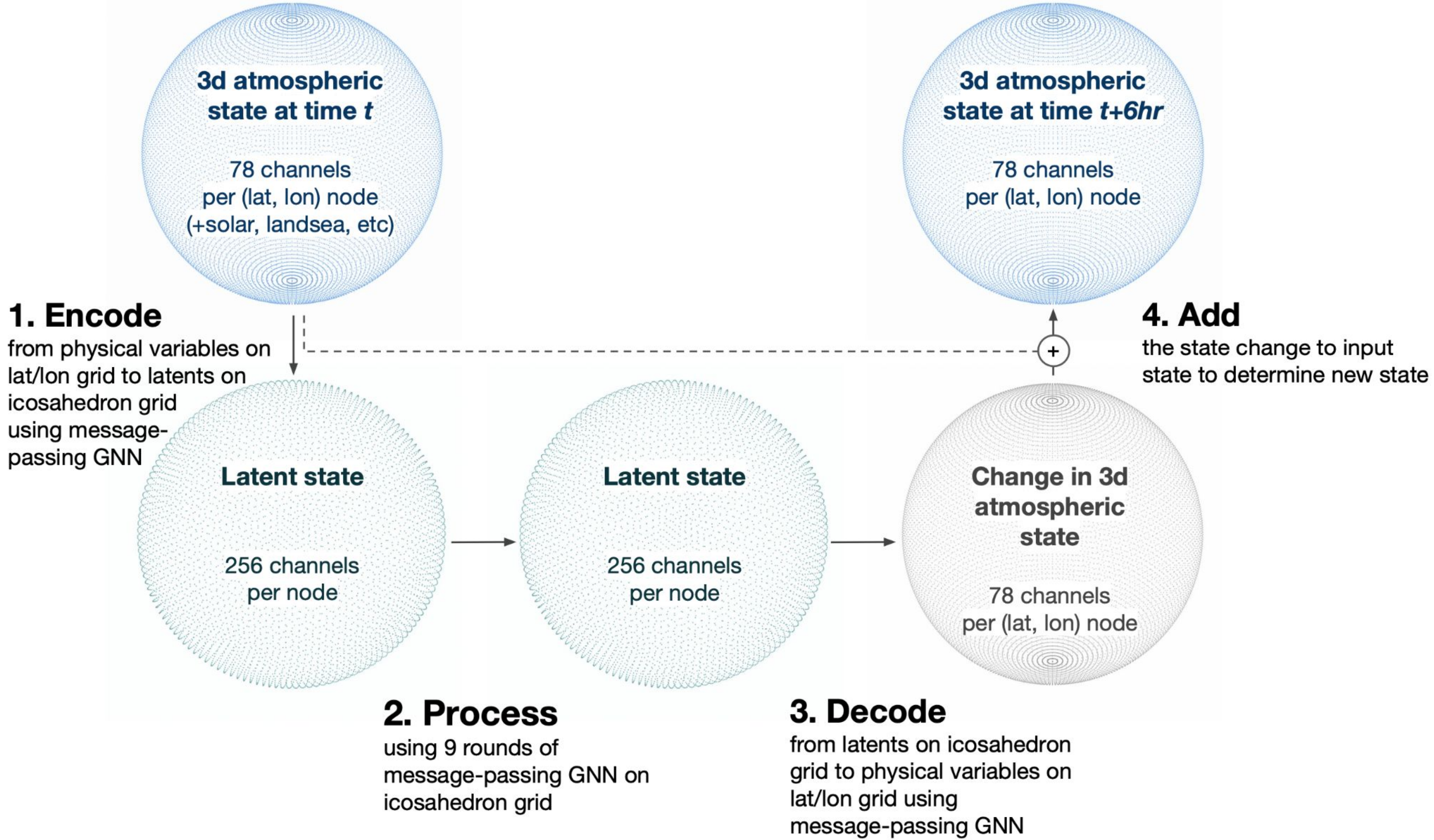

Input:
just use
current state,
i.e. no history
Output:
just predict
next state
Model


Input:
just use
current state,
i.e. no history
Output:
just predict
next state
lat/lon grid
icosahedral grid
lat/lon grid

Input:
just use
current state,
i.e. no history

Output:
just predict
next state





3d atmospheric state at time t
78 channels per (lat, lon) node (+solar, landsea, etc)
3d atmospheric state at time t+6hr
78 channels per (lat, lon) node
1. Encode
from physical variables on lat/lon grid to latents on icosahedron grid using message-passing GNN
4. Add
the state change to input state to determine new state
+
Latent state
256 channels per node
Latent state
256 channels per node
Change in 3d atmospheric state
78 channels per (lat, lon) node
2. Process
using 9 rounds of message-passing GNN on icosahedron grid
3. Decode
from latents on icosahedron grid to physical variables on lat/lon grid using message-passing GNN

GNNs are well suited to NWP

- Easy to handle the spherical geometry of earth
    - Just nodes in 3d space
- Potential for multi-resolution models
    - e.g. learn from GFS *and* HRRR?
- Potential for adaptive meshing
    - i.e. put the compute where it is needed

Counting bits

25 MB

Model
weights

Counting bits

5 PB	>>	2 TB	>>	20 GB	>>	25 MB	~	5 MB	>>	20 kB	>>	1 kB
ERA5 full		ERA5 used		ERA5 used, compressed		Model weights		GFS FV3+FMS source code		GFS atmos_model.F90		Primitive Equations

Counting bits

5 PB	>>	2 TB	>>	20 GB	>>	25 MB	~	5 MB	>>	20 kB	>>	1 kB
ERA5 full		ERA5 used		ERA5 used, compressed		Model weights		GFS FV3+FMS source code		GFS atmos_model.F90		Primitive Equations

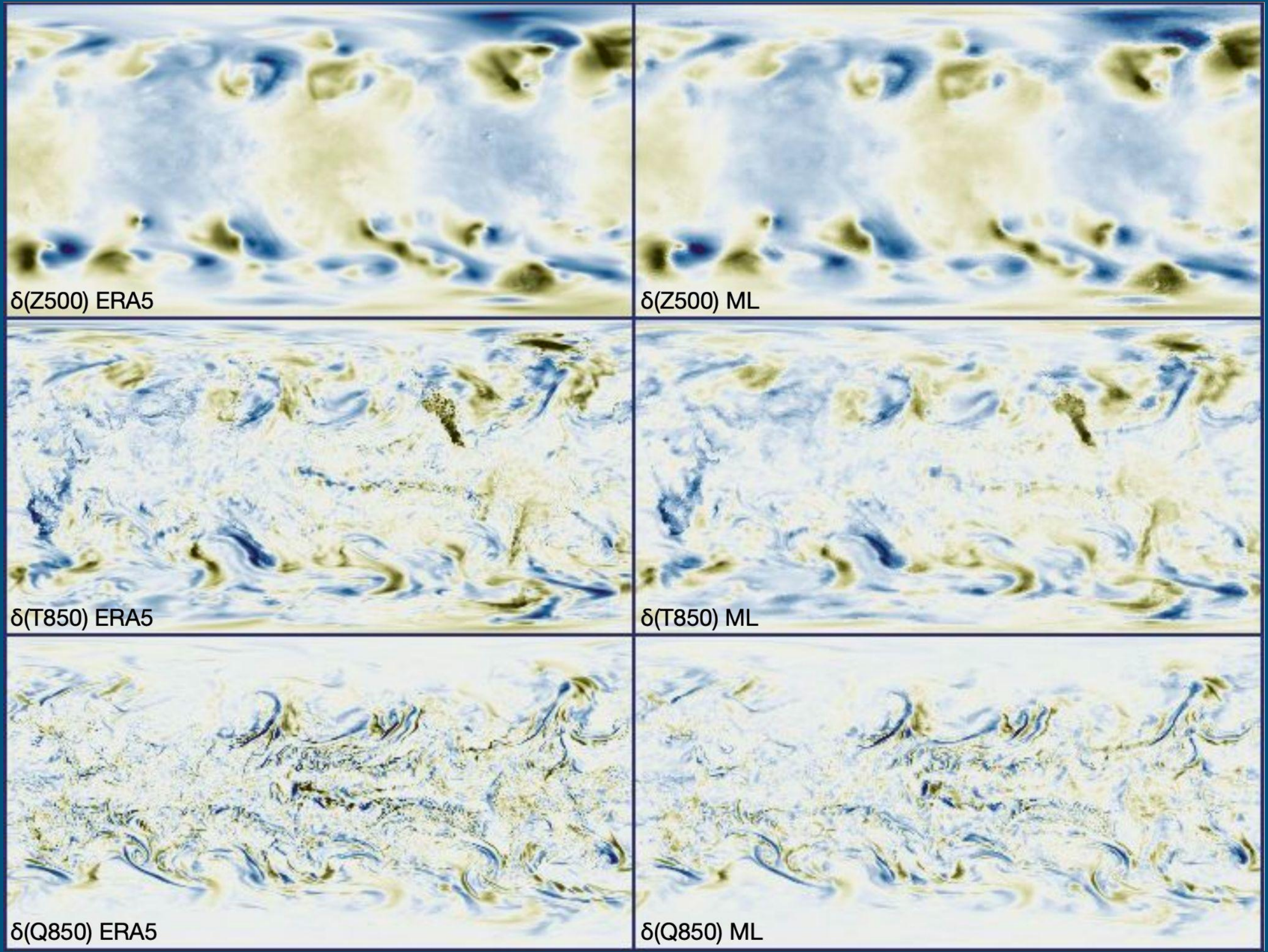
No overfitting

There is a lot more data available.

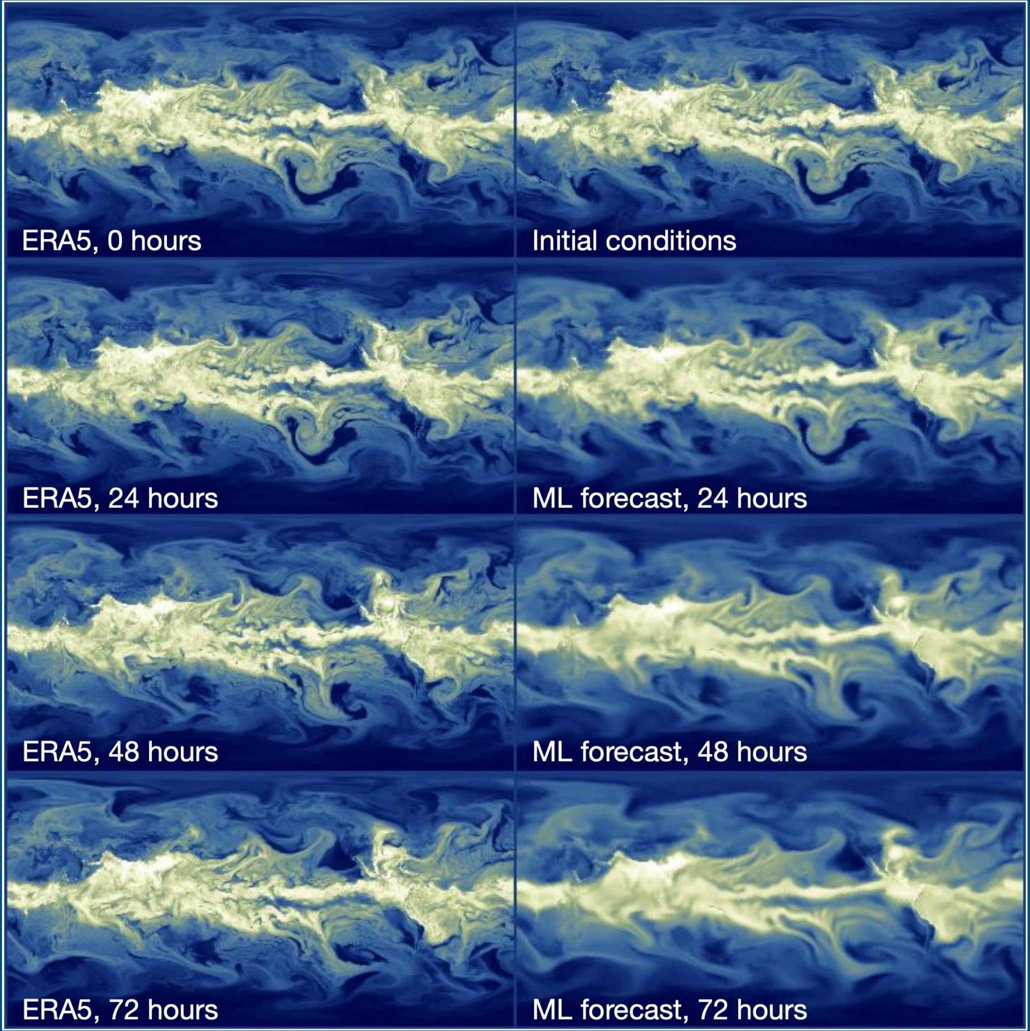
But do you need it?

Results

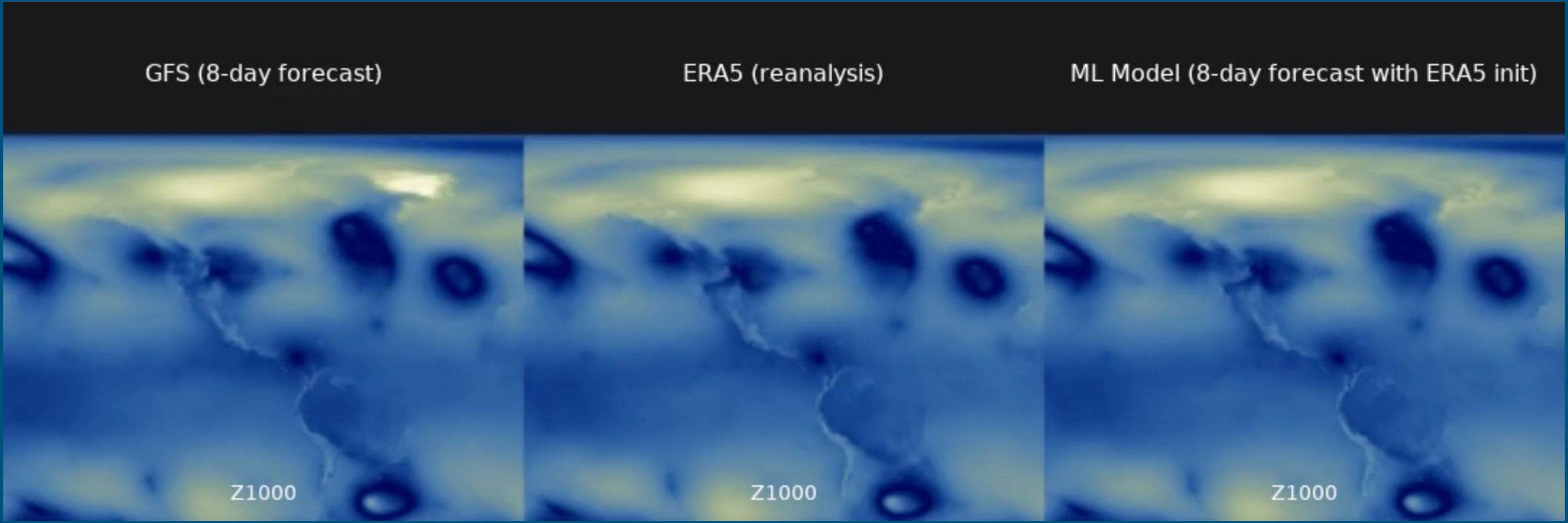
6-hour Differences



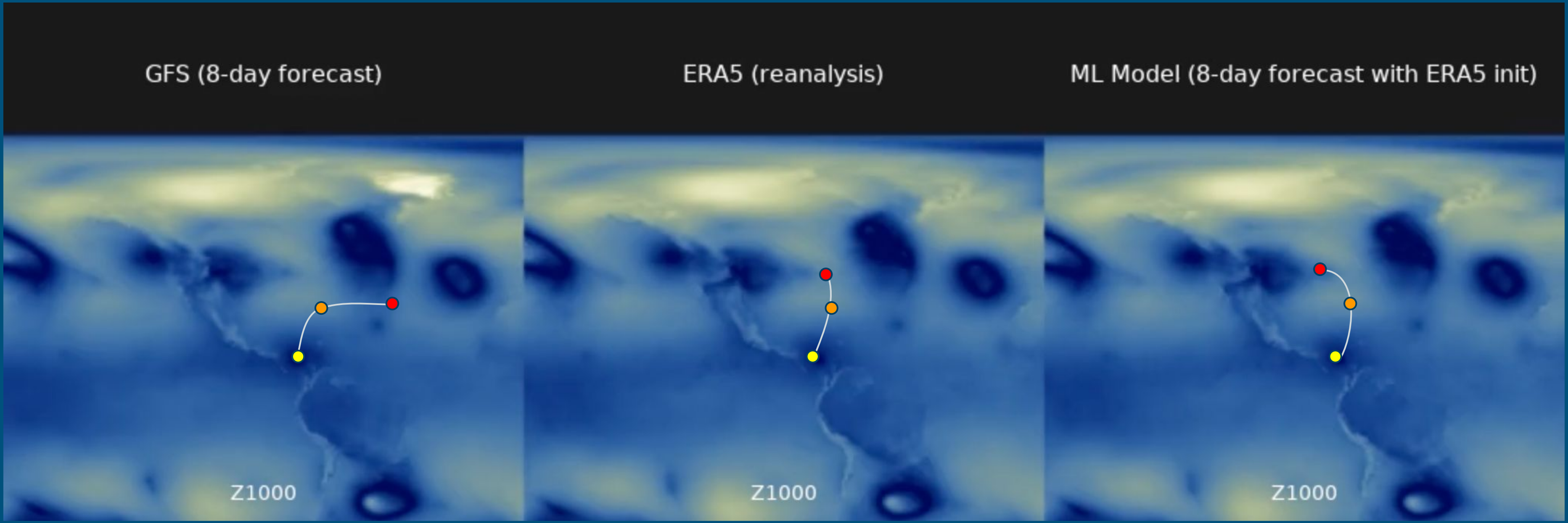
3-day Rollout



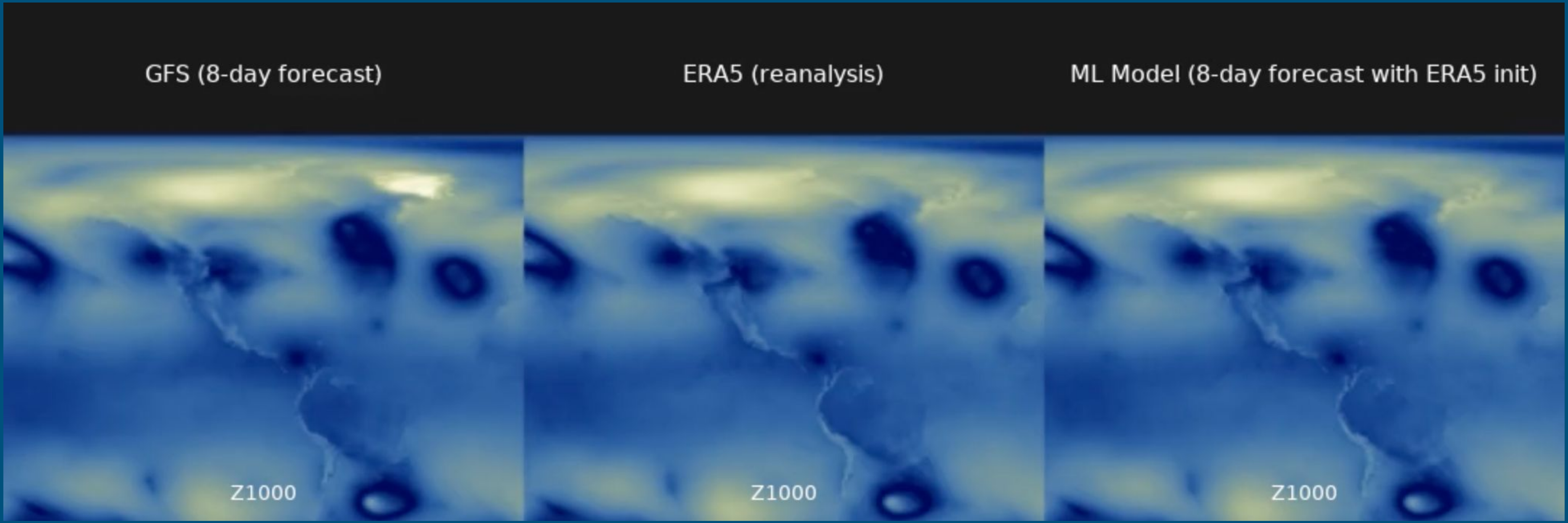
Hurricane Sandy



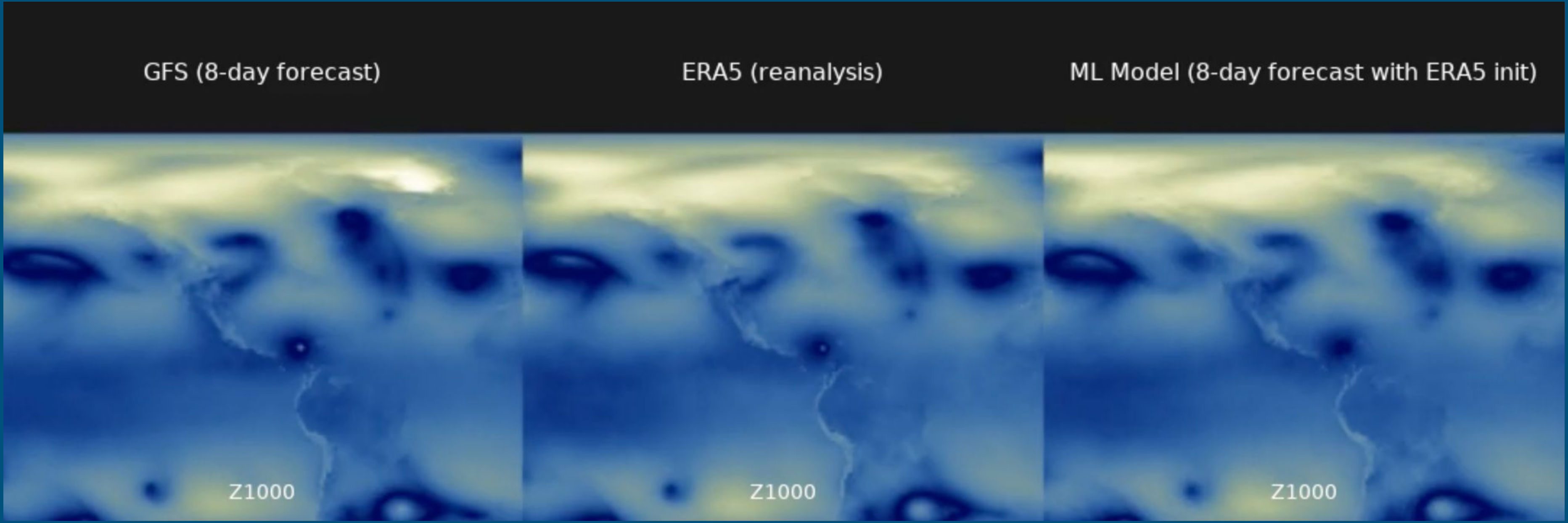
Hurricane Sandy



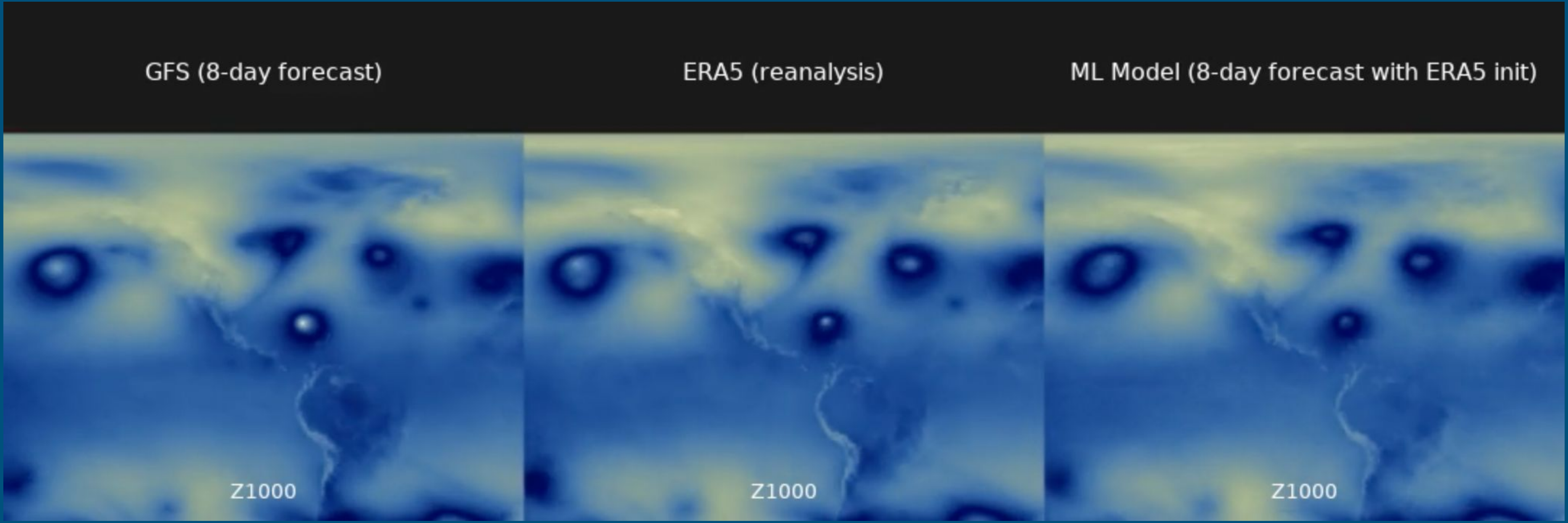
Hurricane Sandy



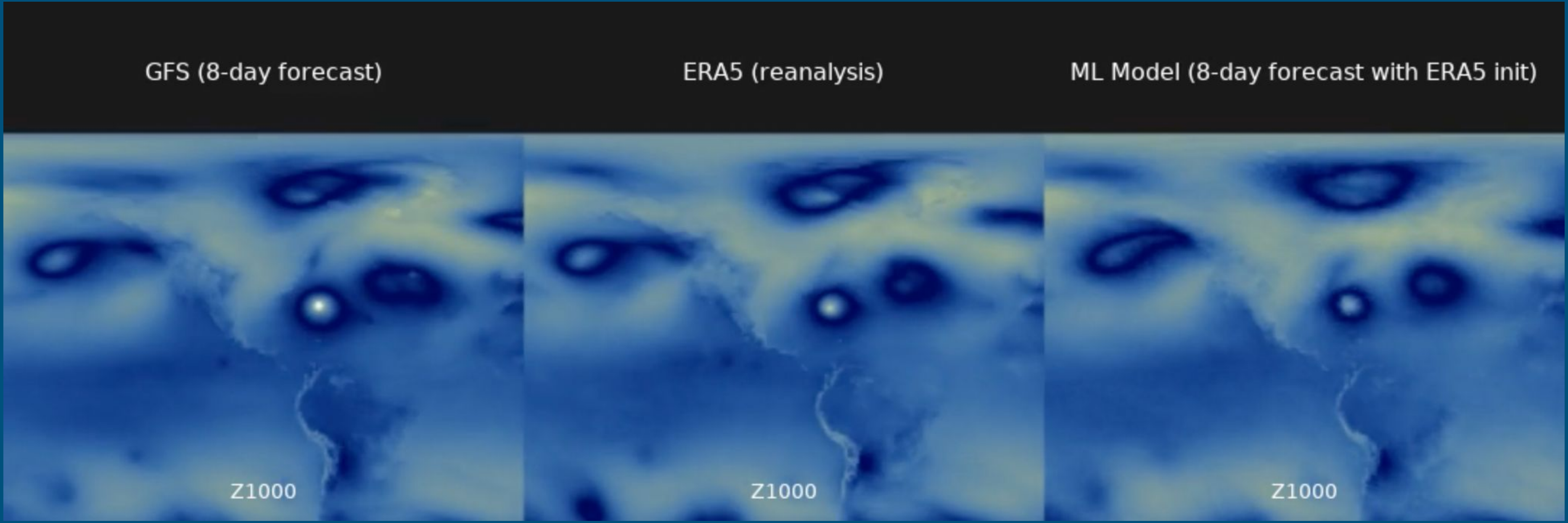
Hurricane Sandy



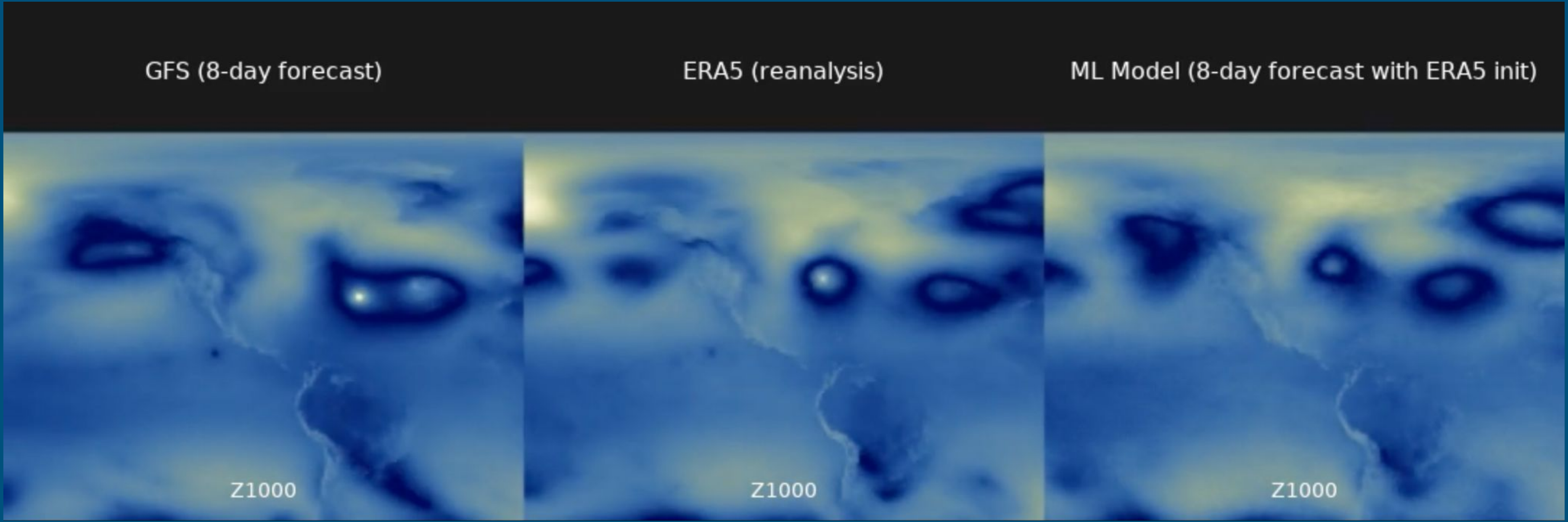
Hurricane Sandy



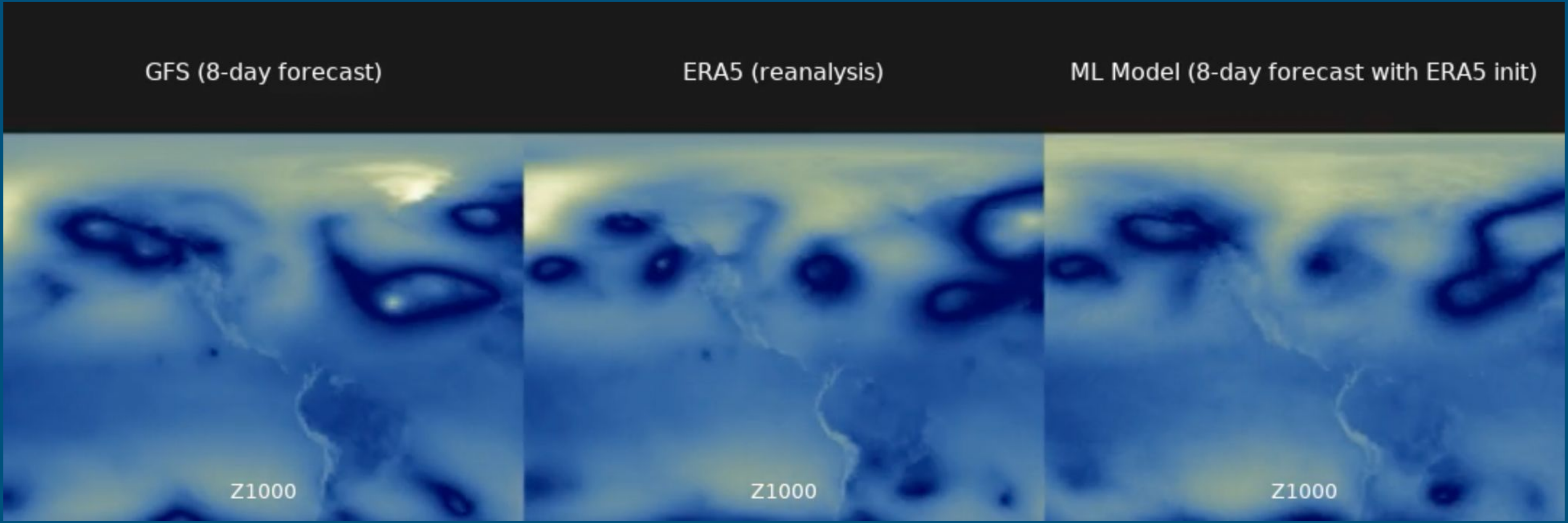
Hurricane Sandy



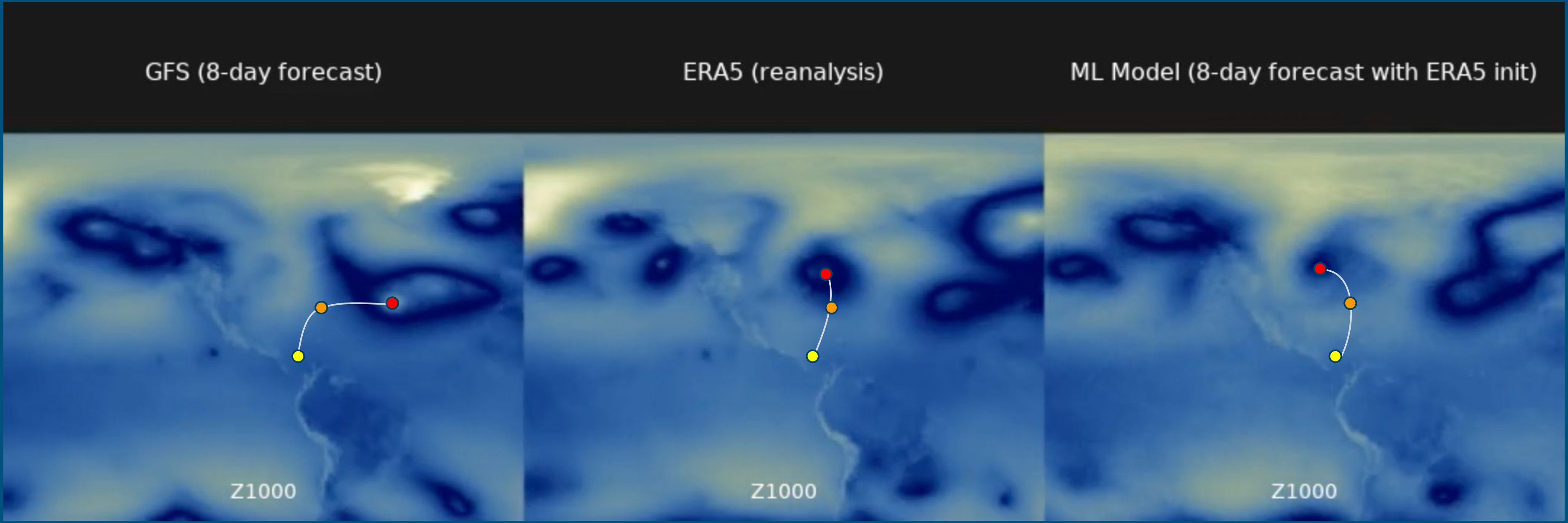
Hurricane Sandy



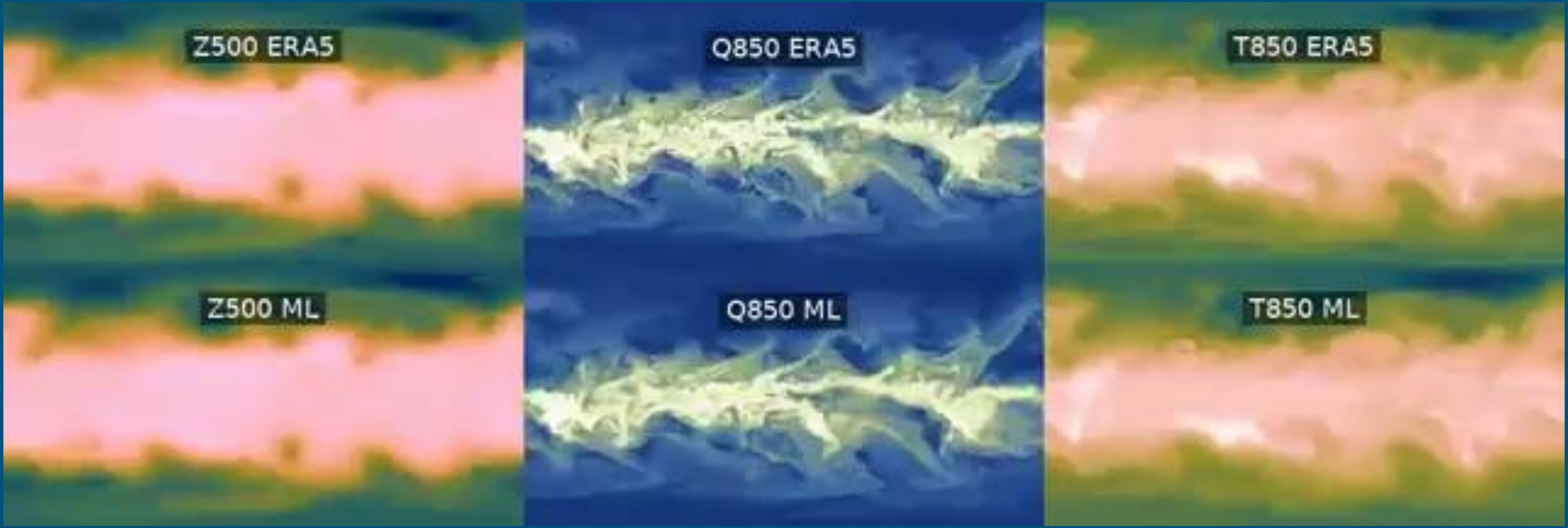
Hurricane Sandy



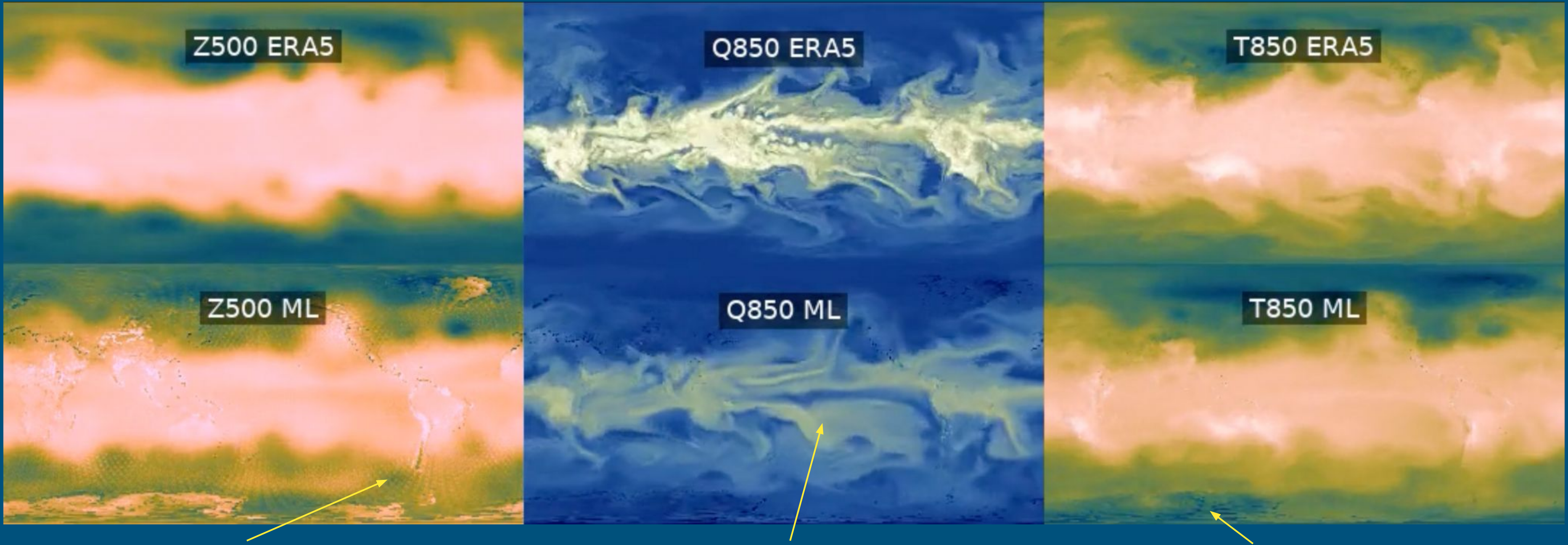
Hurricane Sandy



1-year rollout



1-year rollout (final frame)

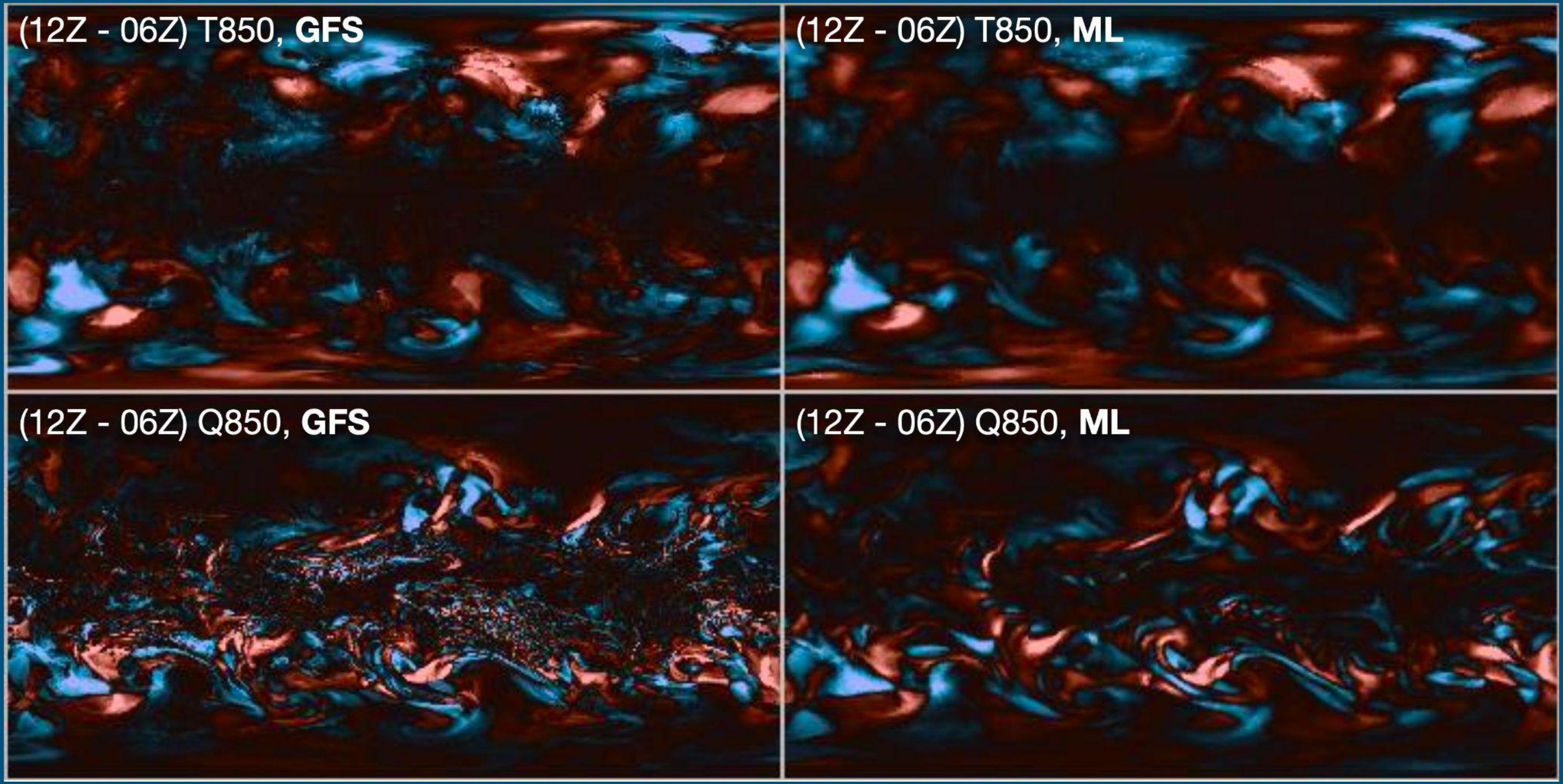


grid pattern

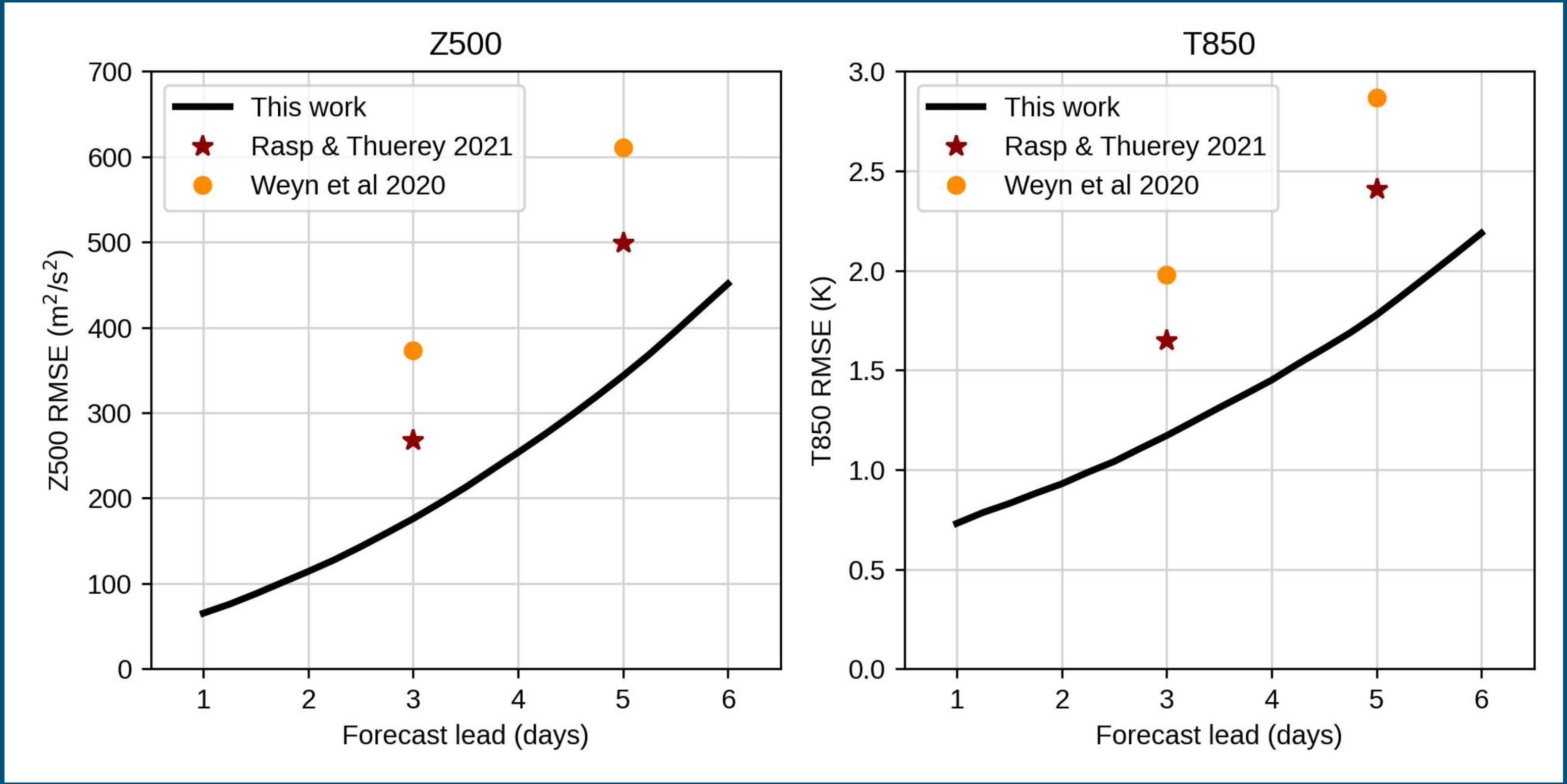
overly smooth

location-specific problems
(topography, land-sea, etc)

Anticipating GFS



Improves upon previous data-driven approaches

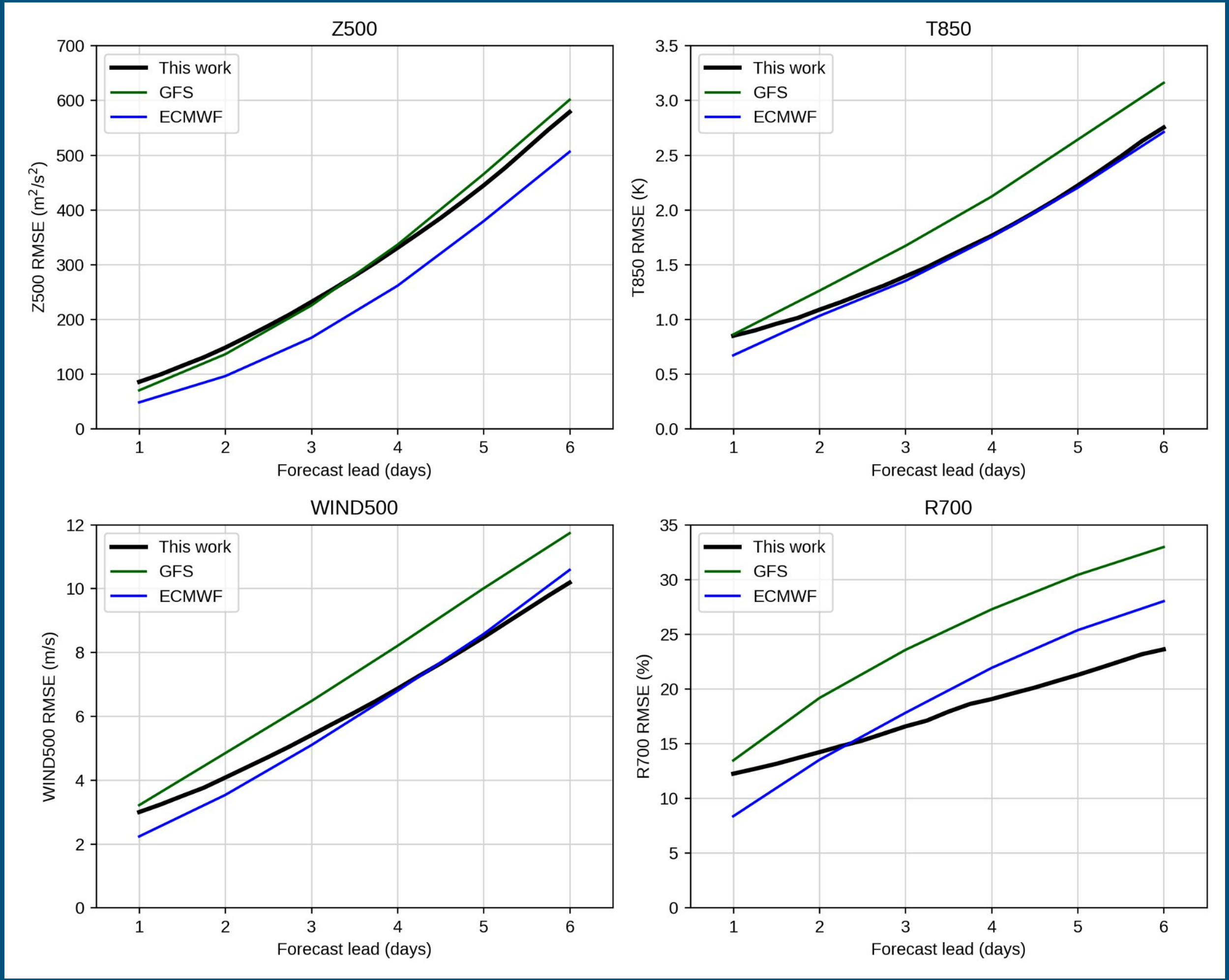


Comparable to Operational NWP

…when high-res op models are evaluated at ~1-deg scale

…when using reanalysis initial conditions

…but still, it works surprisingly well!



Thank you

Please see bit.ly/graph_weather for more.