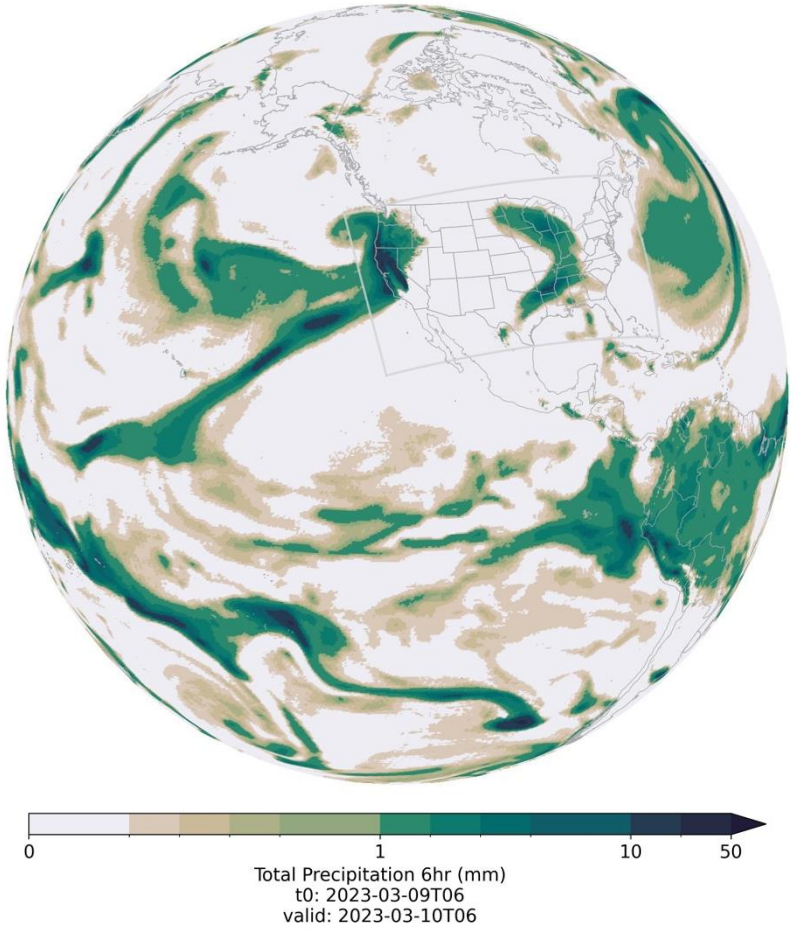
Nested-EAGLE: A Data Driven, Global Weather Model with High Resolution over the Contiguous US

Timothy Smith
previously NOAA PSL, now at the Nansen Center

Mariah Pope, EPIC
Sergey Frolov, NOAA PSL
Brett Basarab, CIRES/NOAA PSL
Paul Madden, CIRES/NOAA GSL
Bo Huang, CIRES/NOAA PSL
Daniel Abdi, CIRES/NOAA GSL
Isidora Jankov, NOAA GSL



Motivation: Seamless Prediction Across Scales

- Weather and climate phenomena can be broadly categorized by their spatiotemporal scales
- NOAA's Numerical Weather Prediction suite consists of individual forecast systems, each of which is uniquely tuned to its target application and scales of motion

Can we unify some of these applications into a single model?

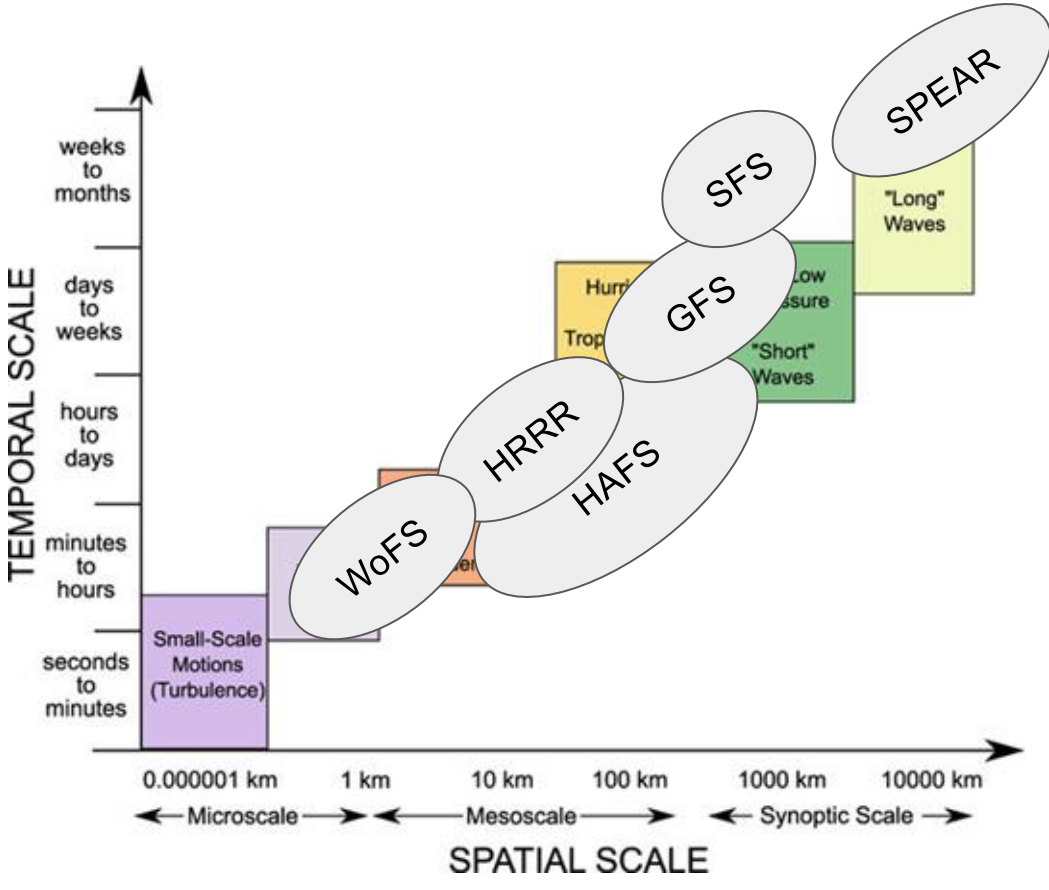
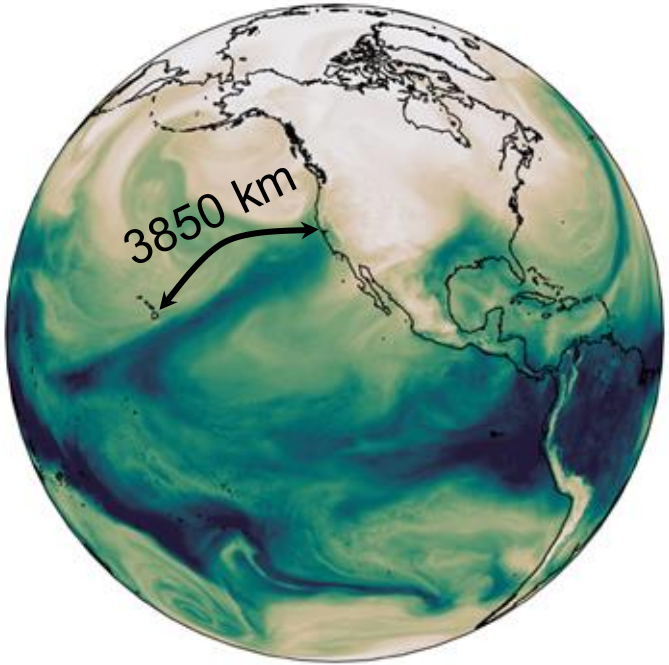
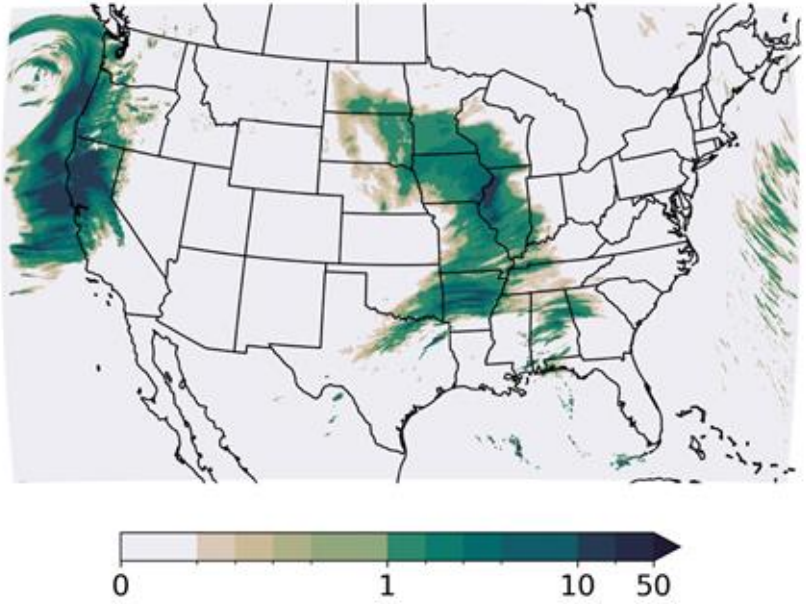


Figure from [Brisch & Kantz (2019)]

Our Focus:
Short & Medium Range Weather



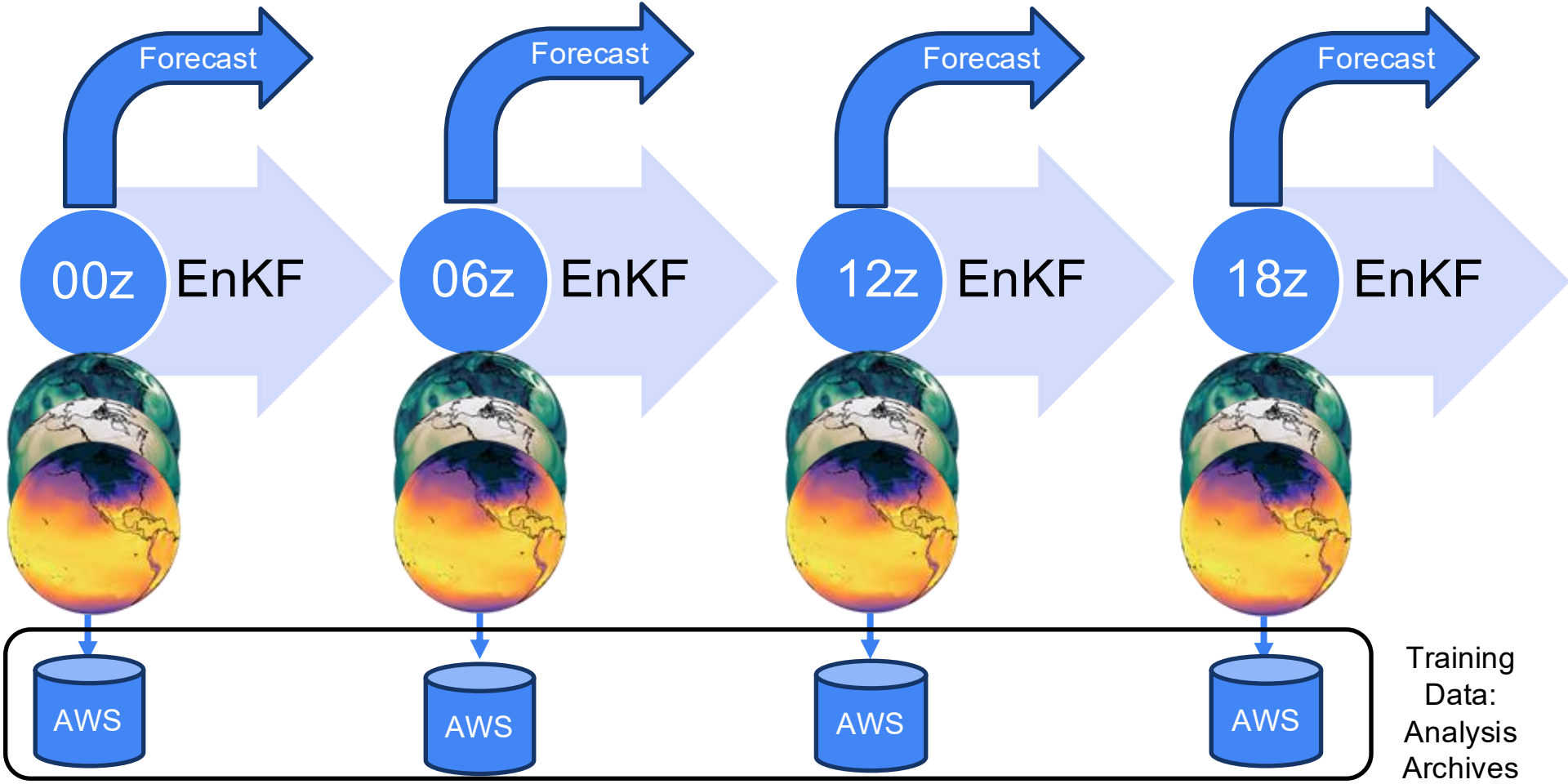
Global Forecast System (GFS)
Medium range: global forecasts to 15 days every 6h at 13km
(Precipitable Water Shown Here)



High Resolution Rapid Refresh (HRRR)
Short range: US forecasts to 48 hours (18 hours) every 6 hours (1 hour) at 3km
6h Accumulated Precipitation Shown Here

Data driven approach: use ML to compress ~1 decade of GFS & HRRR knowledge into a single prediction system

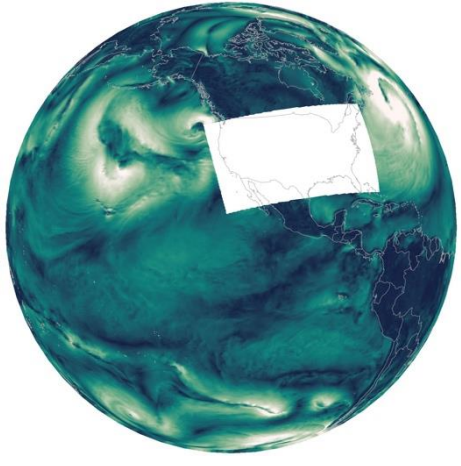
Training Data from Archived Analysis of NOAA's Operational DA and Forecast Systems



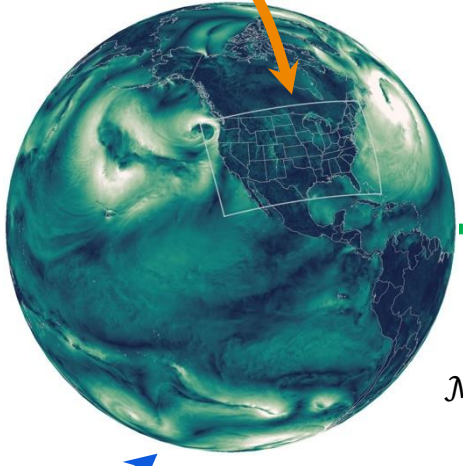
GFS + HRRR = Nested-EAGLE

Train the model so that forecast looks like the next analysis, i.e.
$\mathcal{L} = MSE(\boldsymbol{x}(t+6h) - \boldsymbol{y}(t+6h))$

HRRR fhr = 0

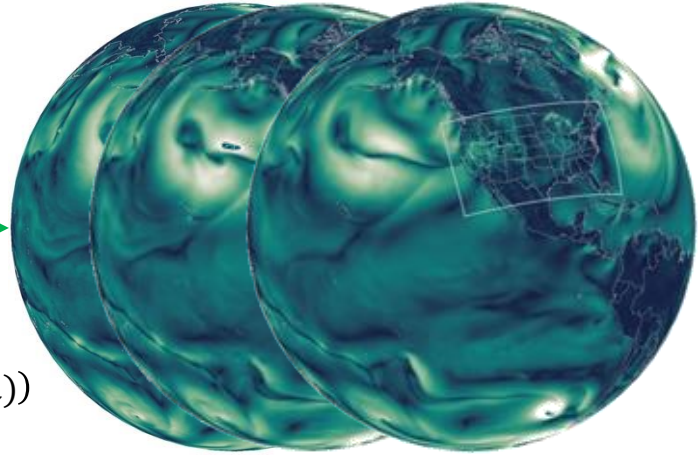


GFS fhr = 0
(HRRR Cutout)



Nested-EAGLE
Initial Conditions
$\boldsymbol{x}(t)$

Nested-EAGLE
$\mathcal{M}(\boldsymbol{x}(t), \boldsymbol{x}(t-6h))$



$$\boldsymbol{x}(t+6h) = \mathcal{M}(\boldsymbol{x}(t), \boldsymbol{x}(t-6h))$$
$$\boldsymbol{x}(t+12h) = \mathcal{M}(\boldsymbol{x}(t)+6h, \boldsymbol{x}(t))$$
$$\dots$$
$$\boldsymbol{x}(t+24h) = \mathcal{M}(\boldsymbol{x}(t+18h), \boldsymbol{x}(t+12h))$$

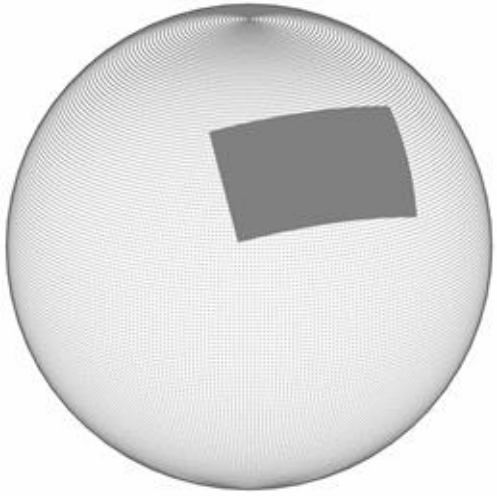
Follows approach pioneered by Nipen et al., (2026) at Met Norway, and directly benefits from open-source contributions by their team.

State Space Definition

Prognostic Fields (Inputs & Outputs)

3D Atmospheric Fields	2D Fields
Geopotential Height Zonal Velocity Meridional Velocity Vertical Velocity Temperature Specific Humidity	Surface Pressure 10m Zonal Velocity 10m Meridional Velocity Skin Temperature 2m Temperature 2m Specific Humidity

12 Vertical Levels:
100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, 1000 hPa



Current model resolution:
¼ degree globally (GFS)
6km over CONUS (HRRR)

Diagnostic Fields (Outputs Only)

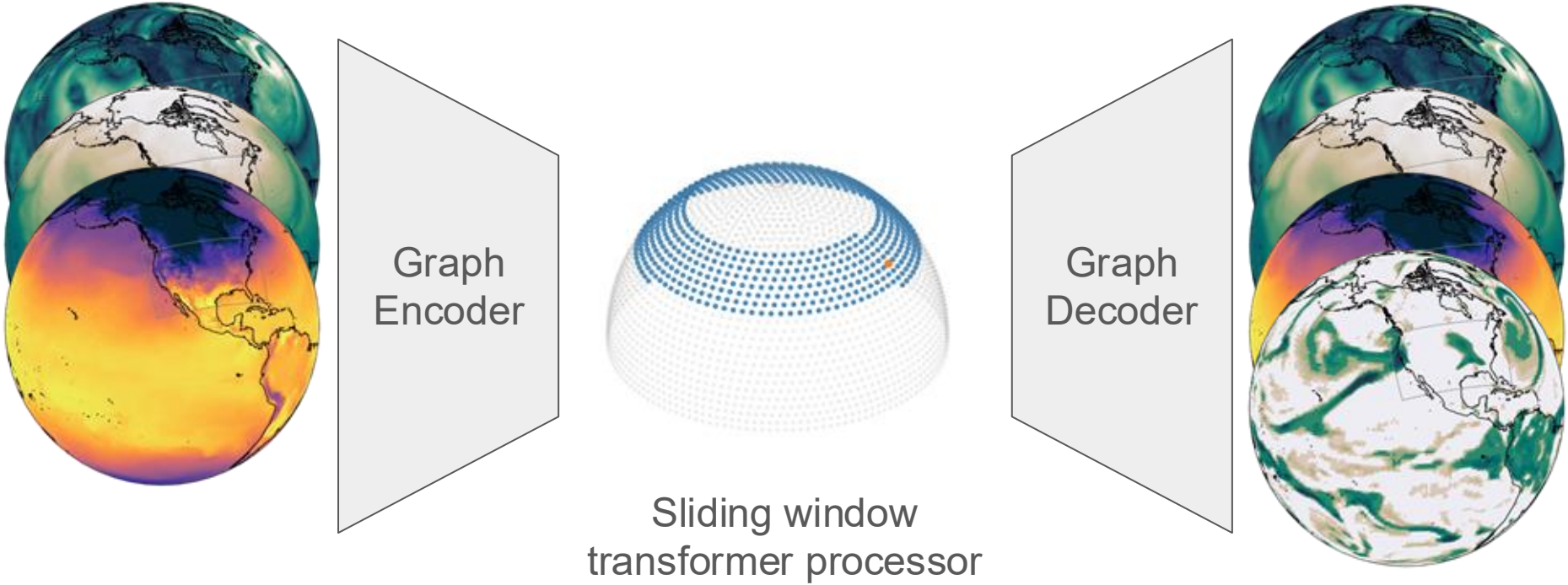
2D Fields
6-hour Accumulated Precipitation 80m Zonal Velocity 80m Meridional Velocity

Forcing Fields (Inputs Only)

Time Varying	Static 2D Fields
Solar Insolation Cosine & Sine of Latitude Cosine & Sine of Longitude Cosine & Sine of Julian Day Cosine & Sine of Local Time	Land Sea Mask Orographic Height

For now, only use variables we have in both GFS & HRRR

Model Design

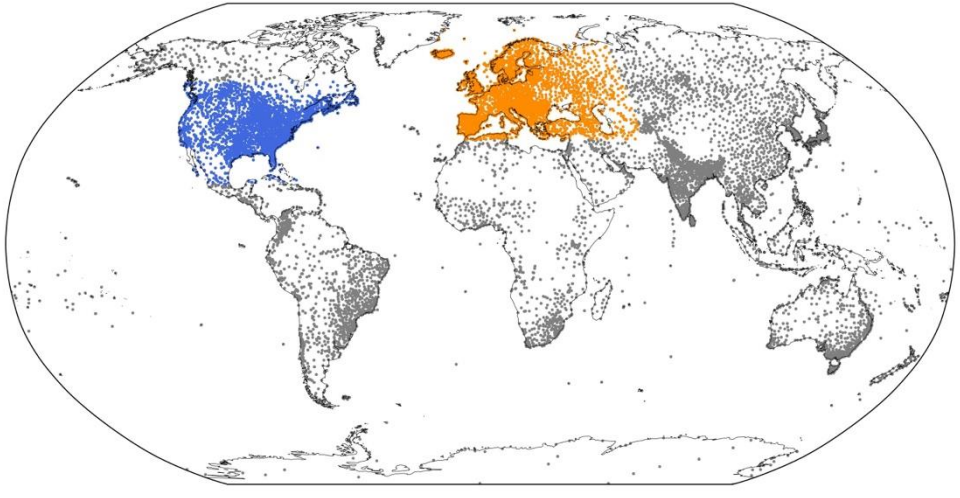


- Latent state lives on mesh that is ~16x (4x lat, 4x lon) coarser than data space grids
- Sliding window image from AIFS paper (Lang et al., 2024)

Evaluation

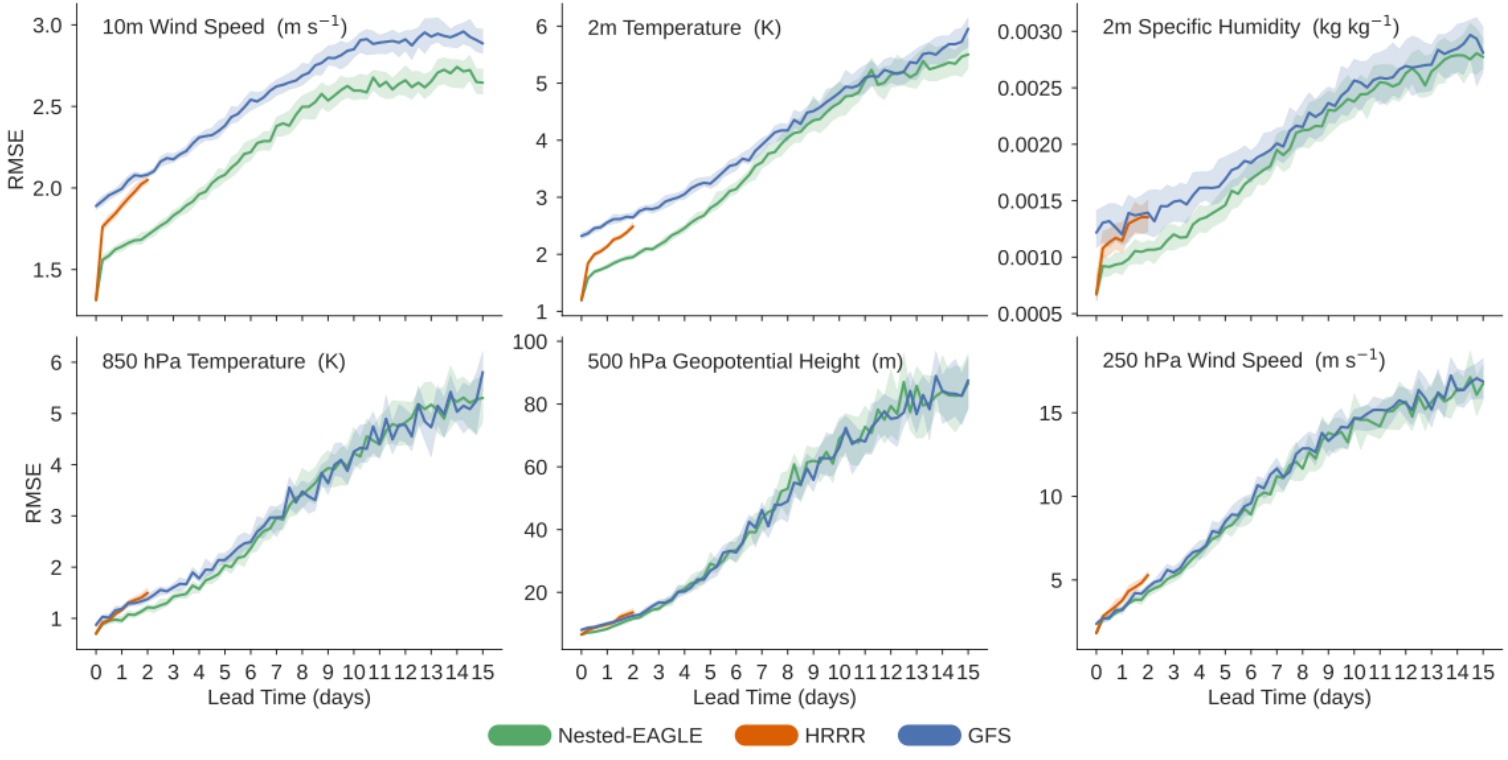
Prognostic Skill Over US & Europe

- 293 forecasts initialized during Feb 2024 - Jan 2025
- Conservatively regridded to 6km (US) and 0.25° grid (Europe/Globe)
  - Processing with ufs2arco
- Bilinearly interpolate to conventional observations from NOAA/Brightband NNJA-AI dataset
  - Includes e.g. dropsondes, rawinsondes, aircraft, SYNOP/METAR station data
  - Data ingest via nnja-ai and comparison via eagle-tools



Example of evaluation coverage on 7 Dec 2024

Prognostic Skill Over CONUS vs Conventional Observations

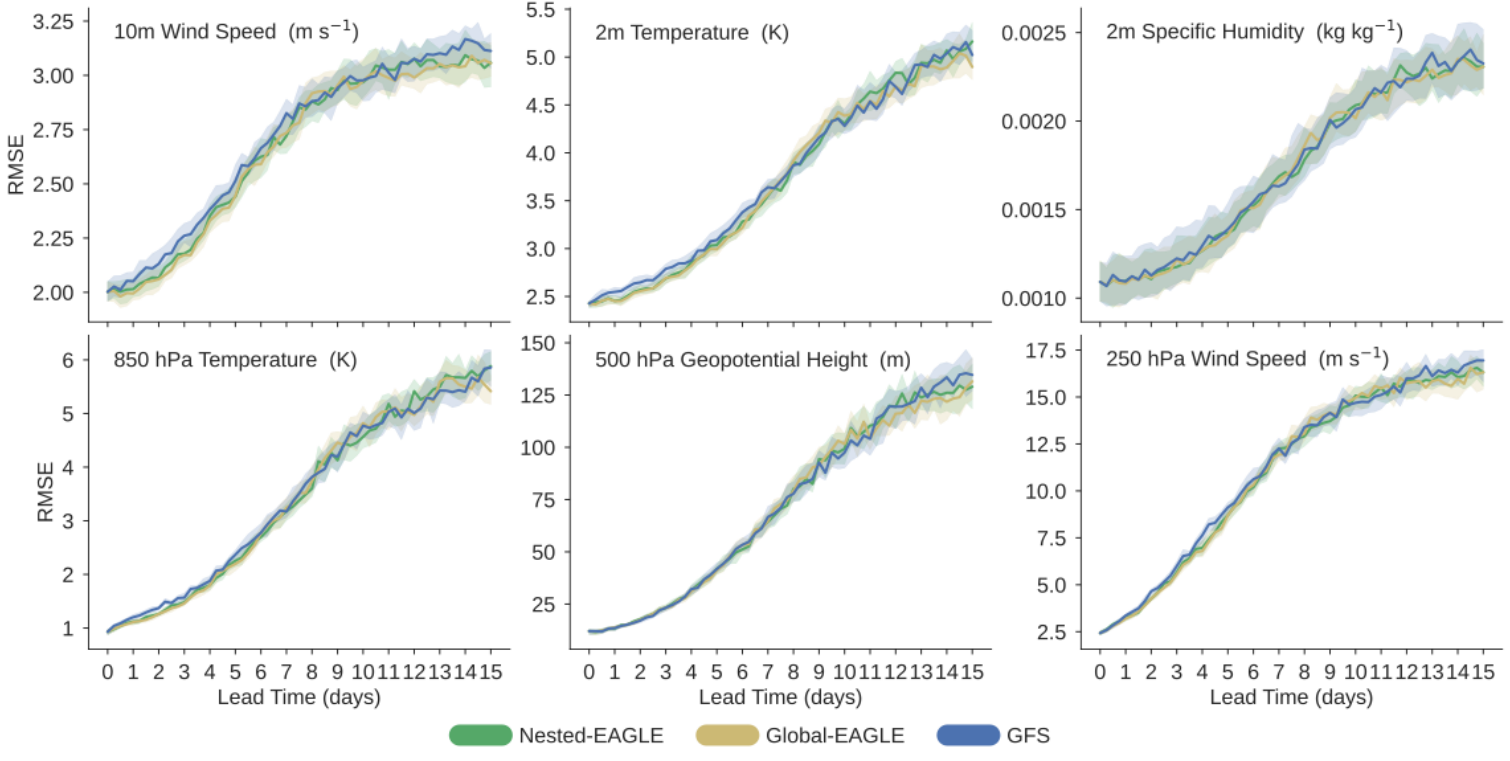


~2 day skill gain to 5 days in 2m Temperature, 10m Winds

Slight advantage to ~5 day 850 hPa Temperature

Competitive Z500 and upper air winds

Prognostic Skill Over Europe vs Conventional Observations

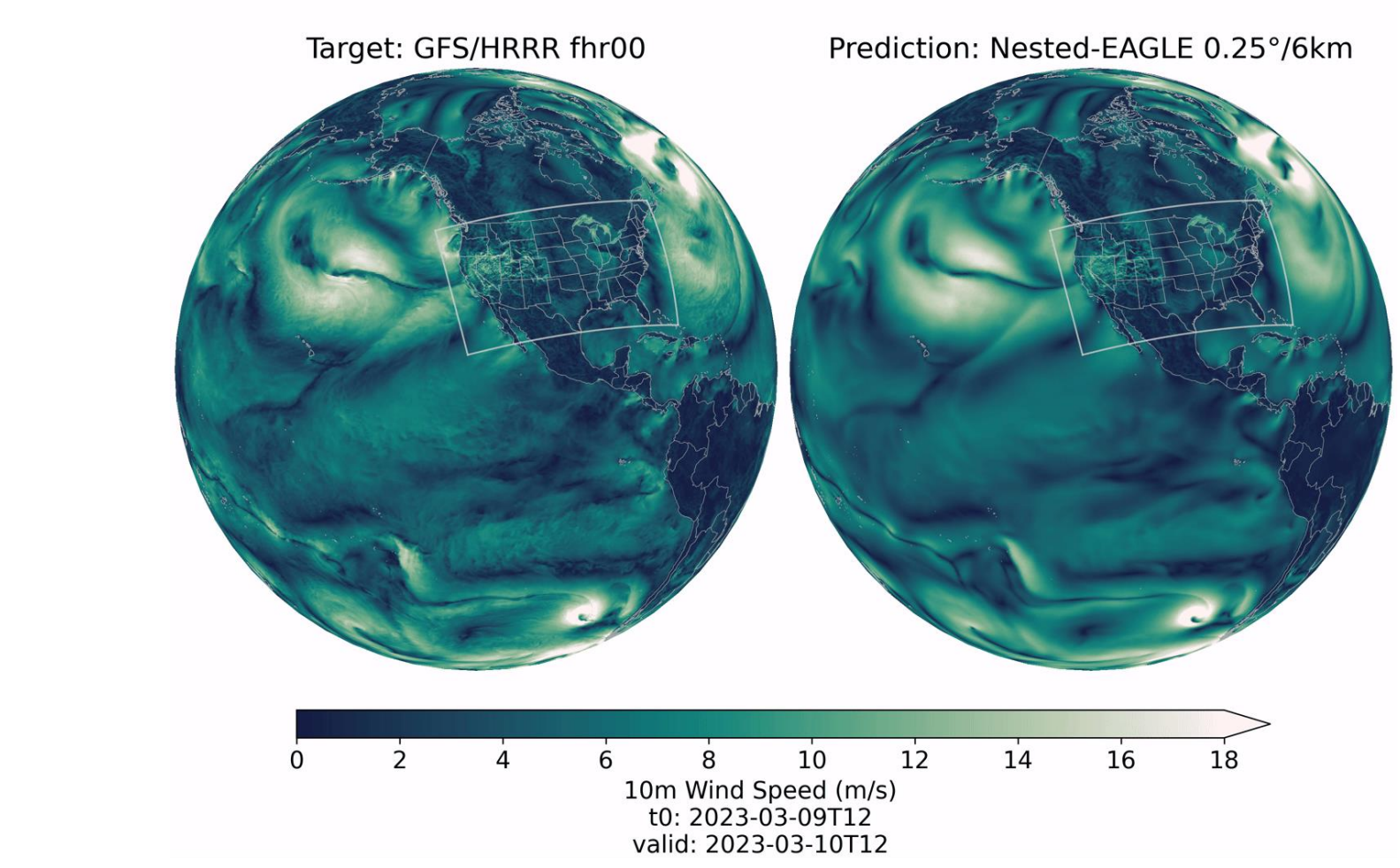


Outside of the US, Nested-EAGLE tracks a model trained only on GFS analysis

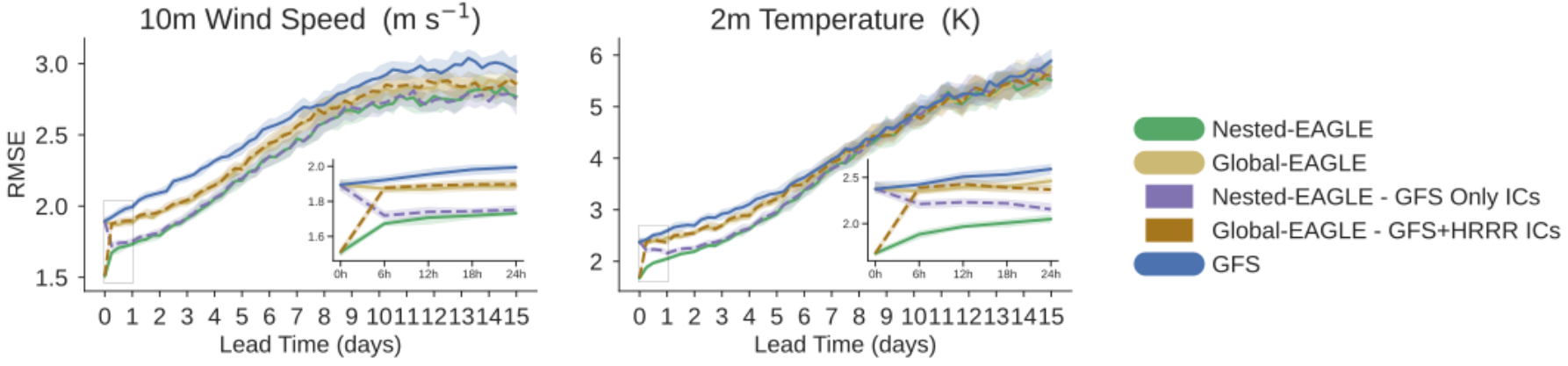
Unfortunate US benefits don't carry over due to supervised learning task

At least not overtrained to US - still a good global MRW model!

Qualitative 10 Day Forecasts



Skill Gain from Errors of the Day, or Learned Constraints?

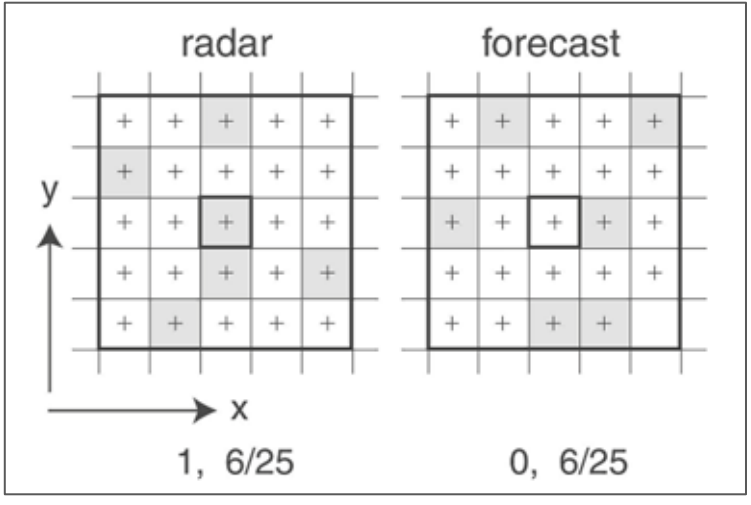


The models relax to the errors corresponding to their training

In this framework, improved initial conditions must be incorporated into training

Precipitation Skill Over CONUS

- Evaluate against NOAA's Analysis of Records for Calibration (AORC) dataset
- ~1450 forecasts Feb 2023 - Jan 2024
- Conservatively regridded to 6km, 6 hourly
- Fractions Skill Score:
  - Radius: ~2 grid cells (~33km)
  - Above a given threshold did forecast meet or exceed the threshold?

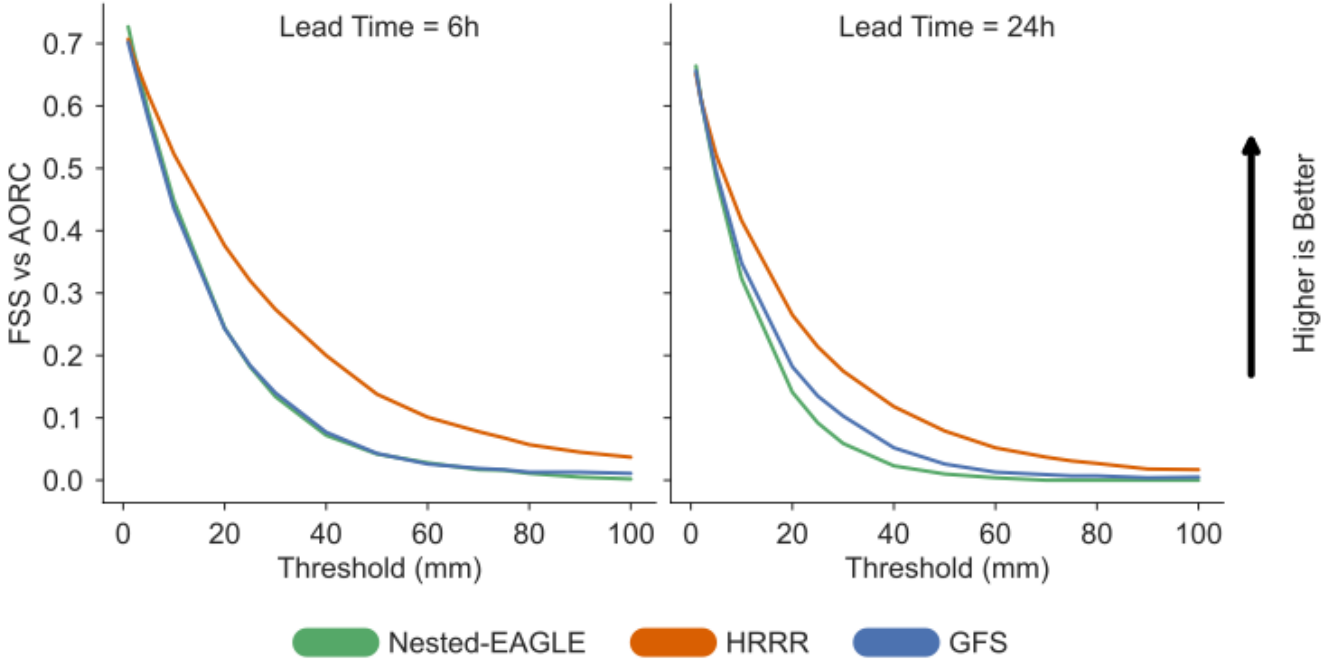


Roberts and Lean (2008) Figure 2:

- Forecast is wrong at central grid cell
- However, forecast & observations have equal fractions of events in the defined 5x5 grid cell neighborhood
- High FSS correspond to forecasts that "look a lot" like observations

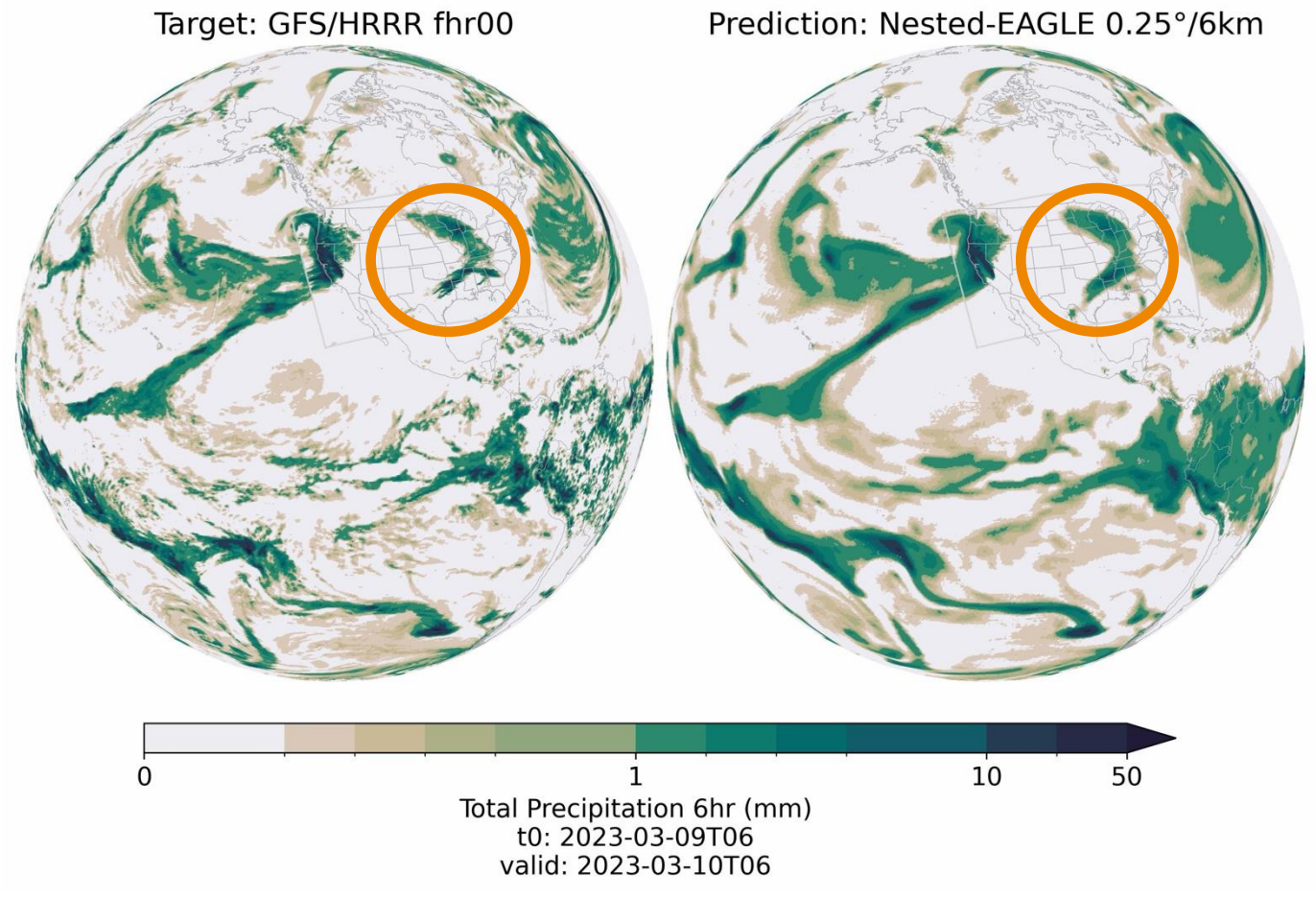
Thanks to Brett Basarab for this slide, and for my ongoing precipitation education

Fractions Skill Score Over CONUS

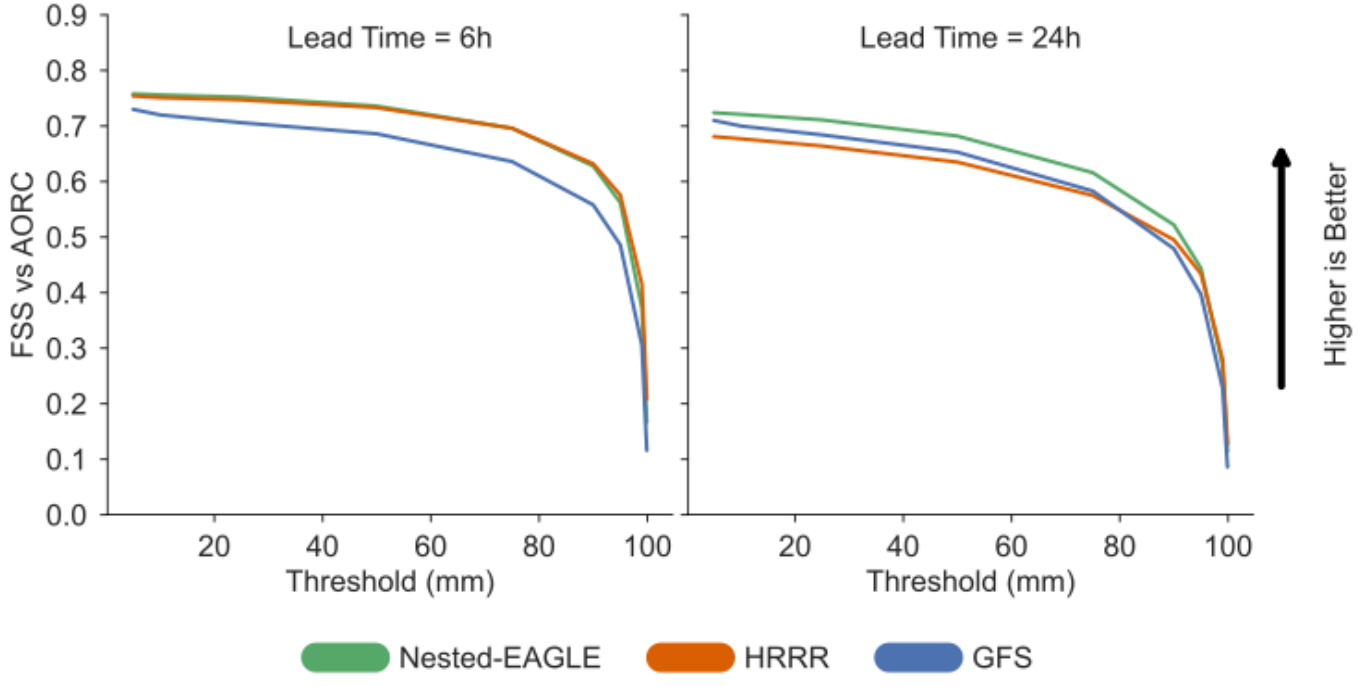


FSS quantitatively shows negative impacts of blurring from MSE-based loss function

Qualitative Precipitation Analysis



Fractions Skill Score, Per Represented Threshold

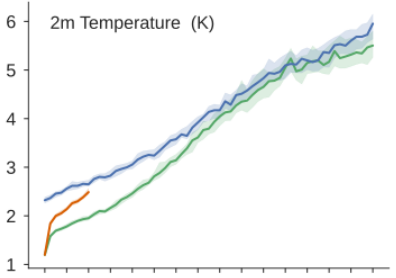


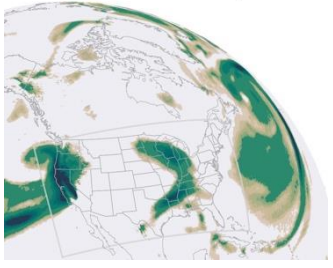
Relative to the values each model does represent, Nested-EAGLE looks good

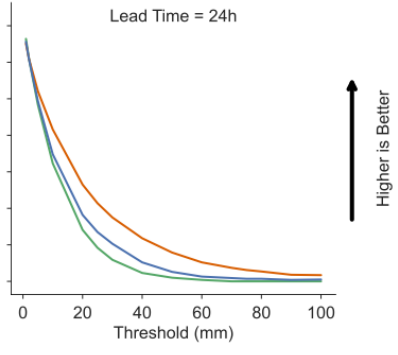
Promising, if extrema blurring can be resolved

Summary

- Nested-EAGLE outperforms NOAA's current NWP suite for near surface quantities, providing short to medium forecasts all-in-one
- These improvements are learned in the model
  - Future analysis improvements need to be learned too
- Atmosphere variables are comparable to GFS
- Precipitation shows blurred features
  - Per-threshold comparison is promising! Good location, bad amplitudes
- Open, reproducible, and accessible. The pipeline consists of pip & conda installable packages and yaml files





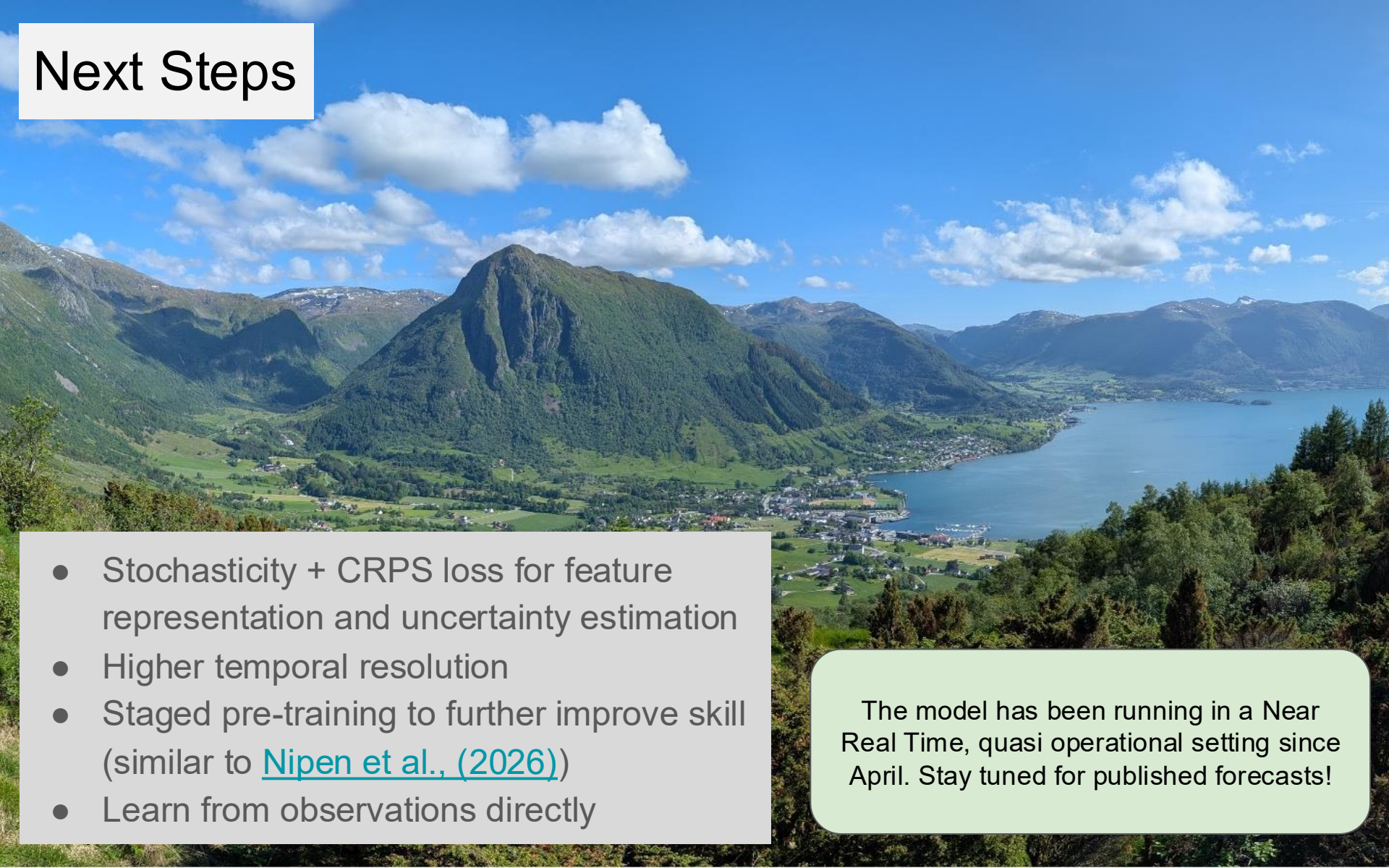


ufs2arco

anemoi

wxvx

Next Steps



- Stochasticity + CRPS loss for feature representation and uncertainty estimation
- Higher temporal resolution
- Staged pre-training to further improve skill (similar to Nipen et al., (2026))
- Learn from observations directly

The model has been running in a Near Real Time, quasi operational setting since April. Stay tuned for published forecasts!