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

Timothy A. Smith, NOAA PSL
Mariah Pope, EPIC
Sergey Frolov, NOAA PSL
Brett Basarab, CIRES / NOAA PSL
Daniel Abdi, CIRES / NOAA GSL
Isidora Jankov, NOAA GSL

Goal

Develop a global medium range weather prediction model that:

- captures synoptic scale dynamics
- represents precipitation at scale useful for decision makers
- produces forecasts at a low computational expense

Data

- Train on GFS & HRRR "Analysis" (fhr=0) for all variables, except precipitation, which uses 0-6h forecast accumulations
- Implement nested or "stretched" grid approach, following Met Norway [1]: cut out CONUS portion from GFS, stick in HRRR
- Conservatively regrid archived GFS: 1/4°→1° and HRRR: 3km→15km
- Use full archives available on NCAR RDA and AWS:
  - Training: Feb 2015-Jan 2023
  - Validation: Feb 2023-Jan 2024
  - Testing Feb 2024-Jan 2025
- Use `ufs2arco` [2] for all data processing

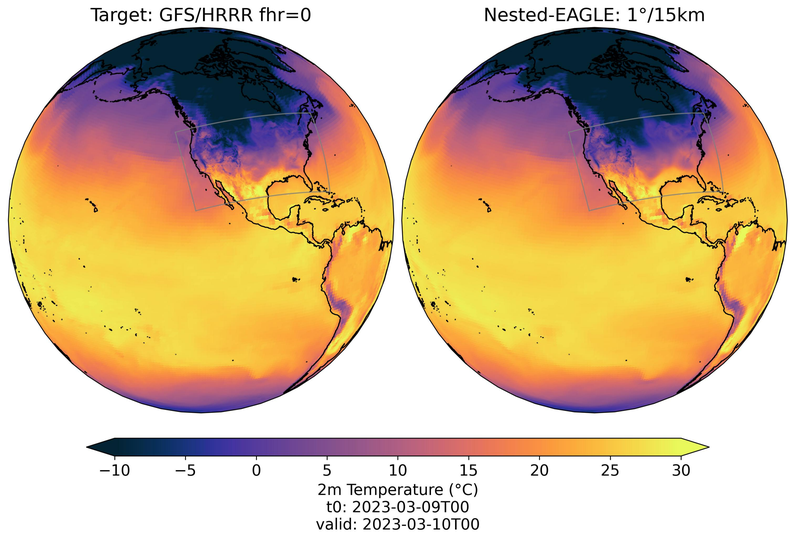
Design Choices that Mattered

- Moving from "GraphCast Style" multimesh to single, high resolution mesh with shifted window processor removed GFS/HRRR boundary artifacts
- Reduce CONUS loss weight 50%→10% improved skill significantly

References

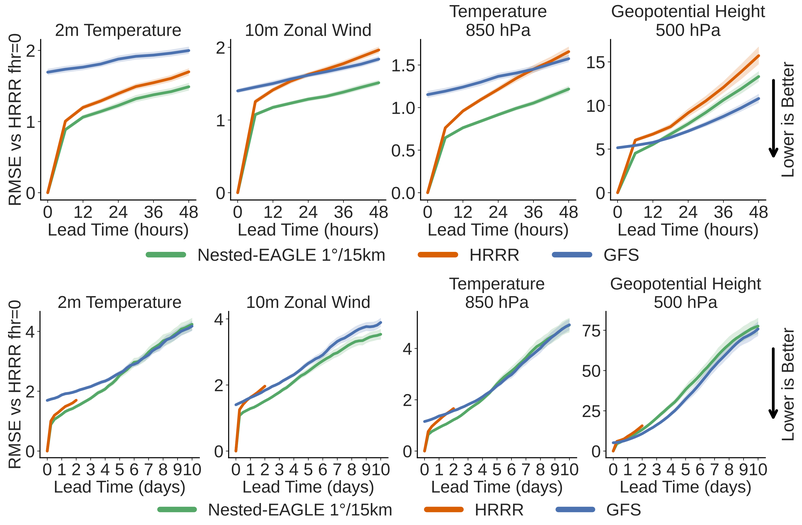
[1] Nipen et al., (2024). doi: 10.48550/arXiv.2409.02891
[2] `ufs2arco` Documentation. `ufs2arco.readthedocs.io`
[3] AORC Dataset. `registry.opendata.aws/noaa-nws-aorc`

Prognostic Skill Over CONUS



Evaluation against HRRR Forecast Hour 0 (fhr=0):

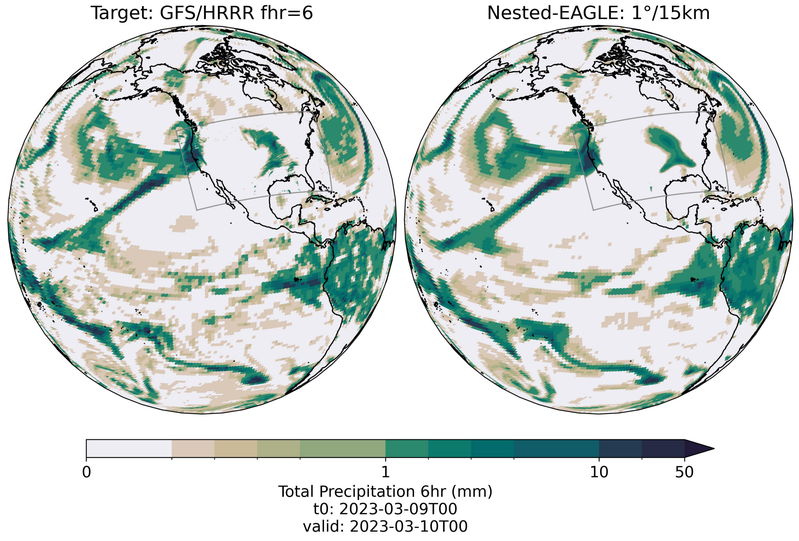
- 158 forecasts initialized throughout validation period
- all datasets conservatively regridded to 15km
- plots show median Root Mean Squared Error (RMSE), shading indicates 95% confidence interval



Main Results

- Lower RMSE than HRRR vs HRRR fhr=0 for all prognostic variables analyzed
- At 10 days, RMSE is competitive with GFS
- At 6h lead, monthly mean precipitation matches HRRR, but longer lead times have larger bias
- Fractional Skill Score highlights blurring in the model, owing to MSE loss in deterministic model training

Precipitation Skill Over CONUS

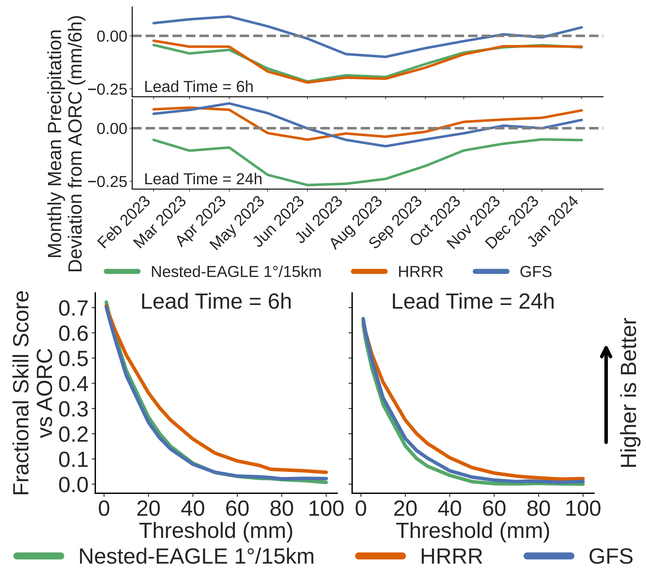


Evaluation against NOAA's Analysis of Record for Calibration (AORC) Dataset [3]:

- ∼1450 forecasts initialized during validation period
- compare 6h accumulations from each model at lead times: 6h, 12h, 24h, and 48h
- all datasets conservatively regridded to 15km

Here we show:

- monthly mean precipitation, relative to AORC
- Fractional Skill Score (FSS), using a ∼33km radius



Next Steps

- Scale to 0.25° global and 6km CONUS resolution
- Train with CRPS loss for ensemble uncertainty estimation and better feature representation
- Incorporate observations into evaluation and training