Evaluation of ML emulators for covariance propagation in realistic DA workflows

Simon Toedtli, David Ahijevych, Wei Wang, Chris Snyder

NSF National Center for Atmospheric Research, Boulder, CO, USA

Research question and configuration

Emulators produce skillful forecasts, but can they also predict differences between forecasts and evolve ensemble perturbations[a,b,c]?

Forecast models

Physical model: MPAS[d]
- 120km quasi-uniform horizontal mesh
- 55 vertical levels, 30 km model top

Emulator: Graphcast[e]
- 1° horizontal resolution
- 13 pressure levels, 50 hPa (approx. 20 km) model top
- Interpolated to MPAS mesh, and extrapolated above 50 hPa, for DA step
- Missing soil state supplied by GFS

DA

1 month of cycling: 2018-04-15 – 2018-05-14

Leverages JEDI framework[f]

GETKF:
- Follows clear-sky configuration of Sun et al.[g]
- 20 member ensemble

3d EnVar:
- Similar to clear-sky configuration of Liu et al.[h]
- Includes variational bias correction

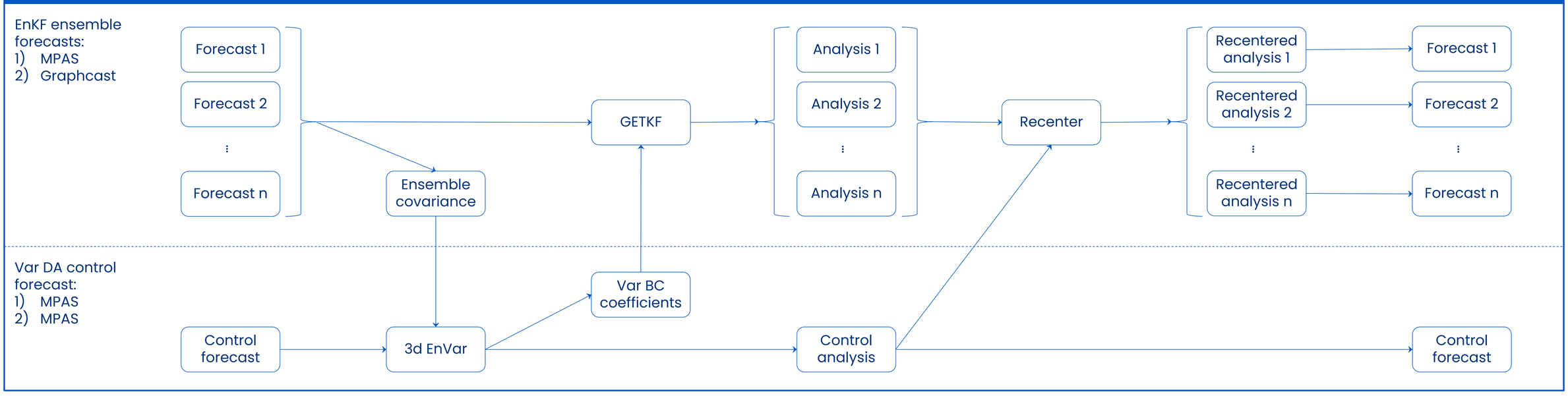
Assimilated observations

Conventional observations:
- Aircraft
- Radiosondes
- Atmospheric motion vectors
- Surface pressure
- GNSS radio occultation

Clear-sky radiances:
- AMSU-A
- MHS

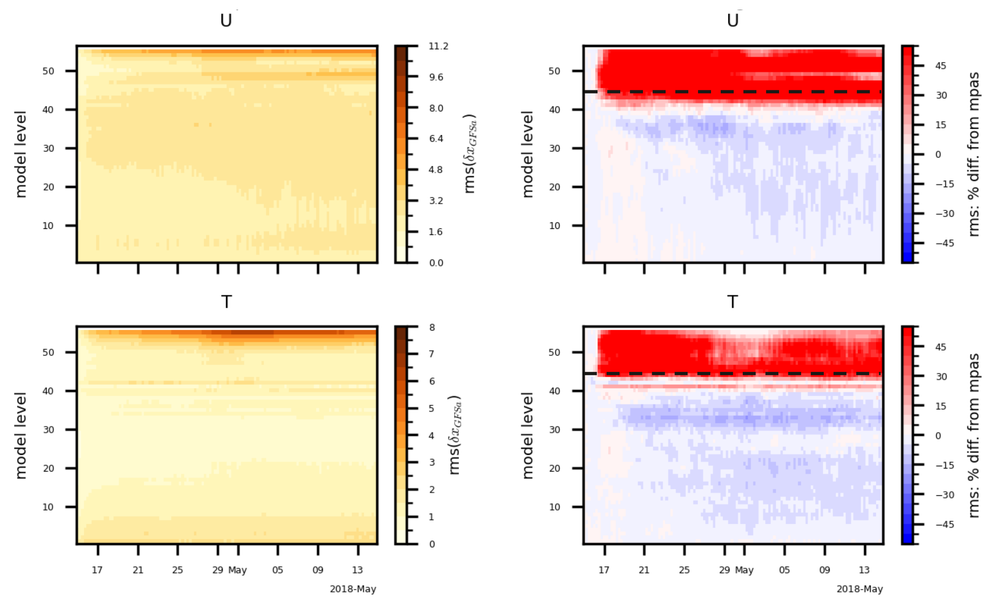
[a] Penny et al. (2022), JAMES, DOI: 10.1029/2021MS002843
[b] Tian et al. (2024), arXiv:2411.14677
[c] Slivinski et al. (2025), Geophys. Res. Lett., DOI: 10.1029/2024GL114396
[d] Skamarock et al. (2012), Mon. Weather Rev., DOI: 10.1175/MWR-D-11-00215.1
[e] Lam et al. (2023), Science, DOI: 10.1126/science.adi2336
[f] Trémolet & Auligné (2020), JCSDA Q. Newsletter, DOI: 10.25923/RB19-0Q26
[g] Sun et al. (2025), Geosci. Model Dev., DOI: 10.5194/gmd-18-8569-2025
[h] Liu et al. (2022), Geosci. Model Dev., DOI: 10.5194/gmd-15-7859-2022

DA workflow



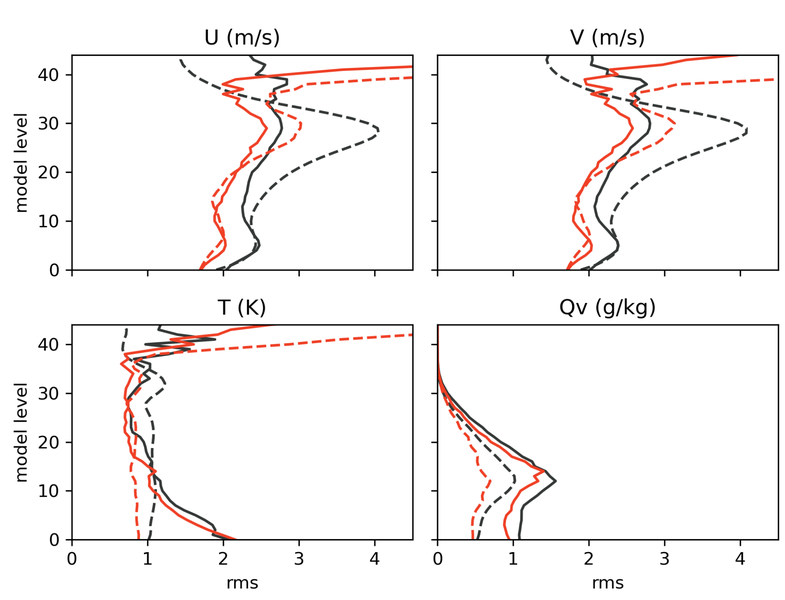
Results

Despite limitations as full forecast replacement in DA workflows[c], Graphcast can accurately evolve ensemble perturbations in DA cycling experiments



Verification of the 6-hourly control background vs. GFS. Left: rms difference (GFS vs. config 1: MPAS ensemble). Right: relative rms change (config 1 vs config 2: Graphcast ensemble).

The Graphcast ensemble is stable and improves relative to MPAS over time. The degradation in the stratosphere is likely due to the lower Graphcast model top (dashed line)



6-hourly background ensemble spread (final week average, levels below the Graphcast model top). Dashed lines: ensemble spread; solid: RMS vs. GFS analysis (optimal spread). Black: MPAS ensemble, orange: Graphcast ensemble.

The Graphcast ensemble maintains a realistic spread over time.

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of NSF.