Assimilating real observations with ML-based models

Laura Slivinski, Jeff Whitaker, Niraj Agarwal, Sergey Frolov, & Tim Smith

Physical Sciences Laboratory & CIRES

Motivation

- Ensemble forecast is expensive, and accuracy in ensemble DA is limited by small ensemble size. ML models offer the potential to generate very large ensembles cheaply
- Many ML models have shown impressive performance in forecasts
- How do these models perform when confronted with real observations in cycled DA?

Article

Neural general circulation models for weather and climate

https://doi.org/10.1038/s41586-024-07744-y
Received: 13 November 2023
Accepted: 15 June 2024
Published online: 22 July 2024

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, Michael P. Brenner & Stephan Hoyer

Article

Accurate medium-range global weather forecasting with 3D neural networks

https://doi.org/10.1038/s41586-023-06185-3
Received: 5 January 2023

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu & Qi Tian

Science RESEARCH ARTICLE

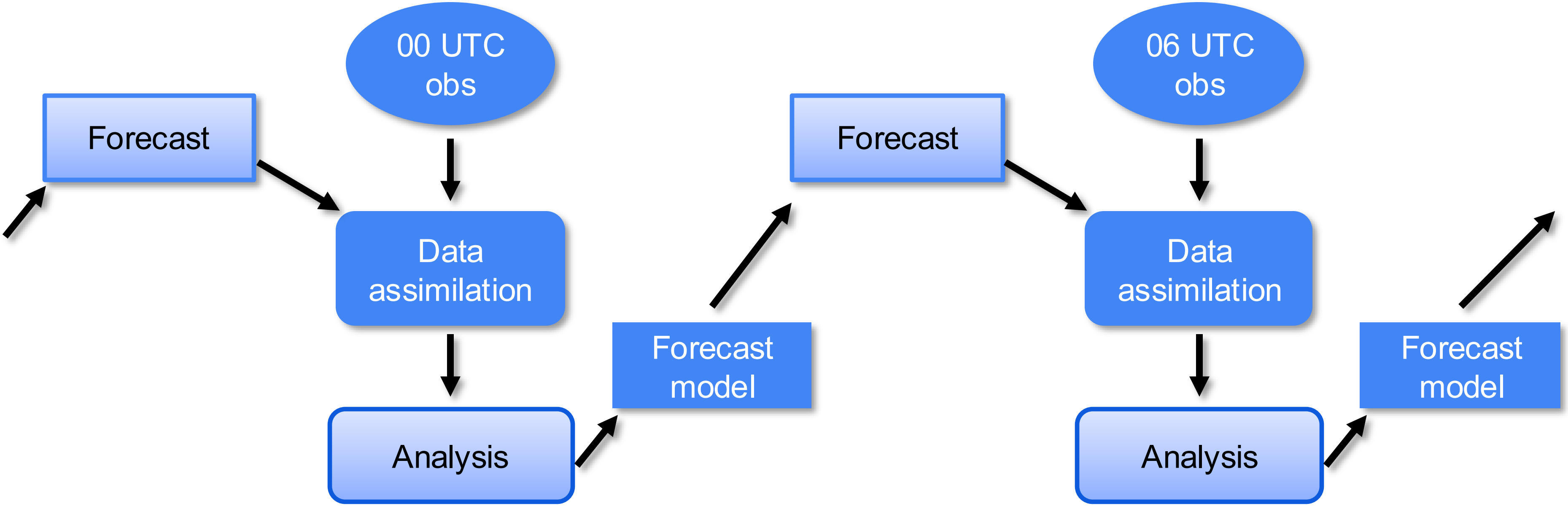
Cite as: R. Lam *et al.*, *Science* 10.1126/science.adi2336 (2023).

Learning skillful medium-range global weather forecasting

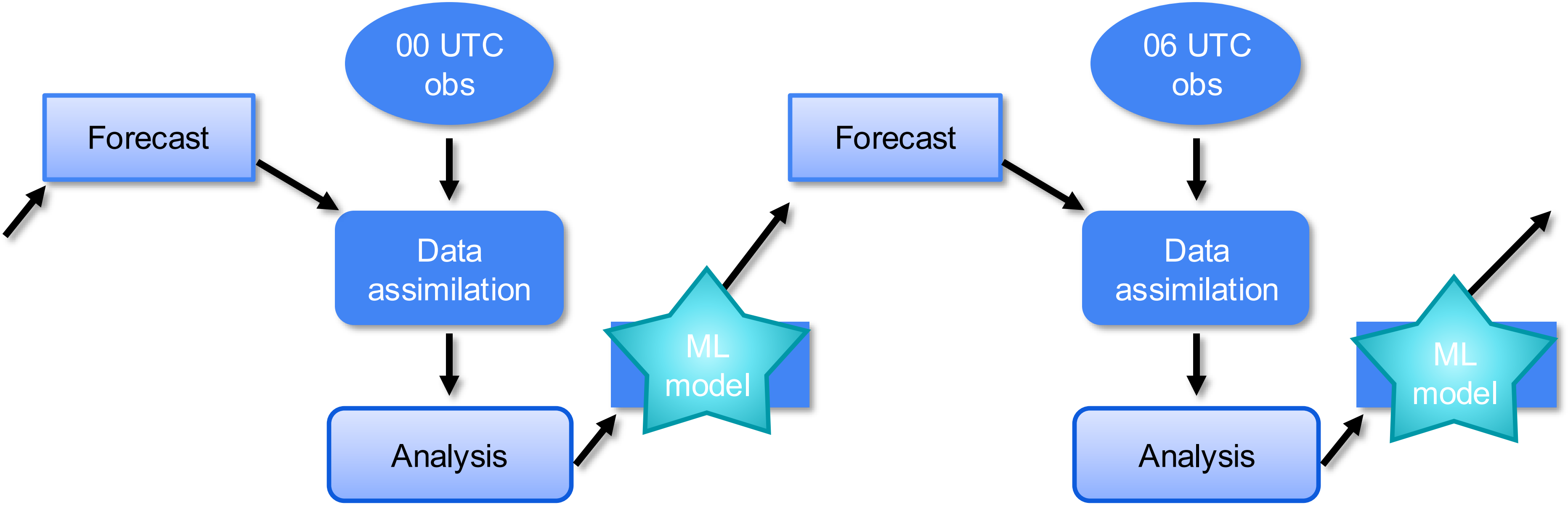
Remi Lam†, Alvaro Sanchez-Gonzalez†, Matthew Willson†, Peter Wirnsberger†, Meire Fortunato†, Ferran Alet†, Suman Ravuri†, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed*, Peter Battaglia*

[1]Google DeepMind, London, UK. [2]Google Research, Mountain View, CA, USA.

Can physical models be replaced by ML models in a cycling DA system for NWP?

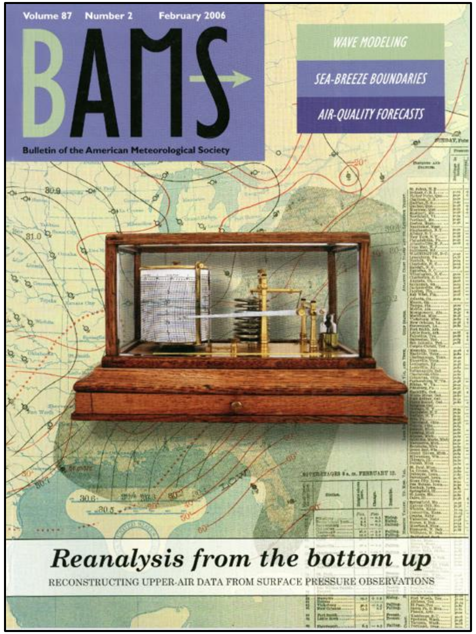


Can physical models be replaced by ML models in a cycling DA system for NWP?



Test assimilating *only surface pressure obs* with EnKF

- **Motivation**
  - Sfc pres obs can get "most" of the answer *(20th Century Reanalysis; Compo et. al. 2008)*
  - Sfcp-only is a stringent test of ensemble covariances: good estimates of background-error covariances are essential
  - Surface pressure forecast skill often not optimized in training



Test assimilating *only surface pressure obs* with EnKF

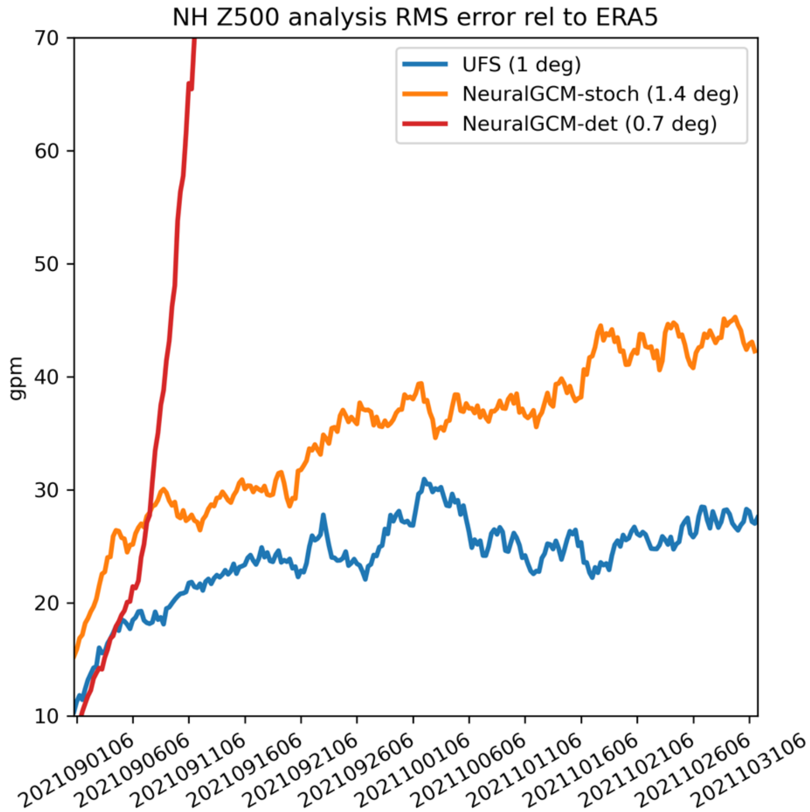
- **Models tested**
  - **Unified Forecast System (UFS)** (roughly 1 deg, 127 vertical levels)
    - Physical atmospheric model with stochastic parameterizations
  - **NeuralGCM-stoch** (1.4 deg, 32 levels)
    - Stochastic hybrid model, stable on climate timescales
  - **NeuralGCM-det** (0.7 deg, 32 levels)
    - Deterministic hybrid model, unstable on long timescales
  - **PanguWeather** (0.25 deg, 13 levels)
    - Deterministic pure emulator, medium-range weather forecasting, unstable on long timescales
  - **GraphCast** (0.25 deg, 13 levels)
    - Deterministic pure emulator, medium-range weather forecasting, unstable on long timescales
  - **Spherical Fourier Neural Operators (SFNO)** (0.25 deg, 13 levels)
    - Deterministic pure emulator, medium-range weather forecasting, stable up to 1 year

Experimental Details

- GSI EnKF, updates model level (hybrid sigma/pressure) or pressure level data on rectangular lat/lon grid (Gaussian or regular) every 6-hours.
- Serial EnSRF, configured as in 20CRv3 *(Slivinski et. al. 2019)*, but without observation bias-correction and non-linear QC.
- 80 member ensemble, with horizontal and vertical localization & relaxation to prior spread inflation (and also relaxation to prior perturbation inflation for non-stochastic emulators like GraphCast).
- Forward observation operator MAPS pressure reduction *(Benjamin & Miller 1990)* from model terrain to station height (for Pangu and GraphCast MSLP is used, so model terrain = 0).
- **All experiments initialized from (interpolated) operational UFS ensemble.

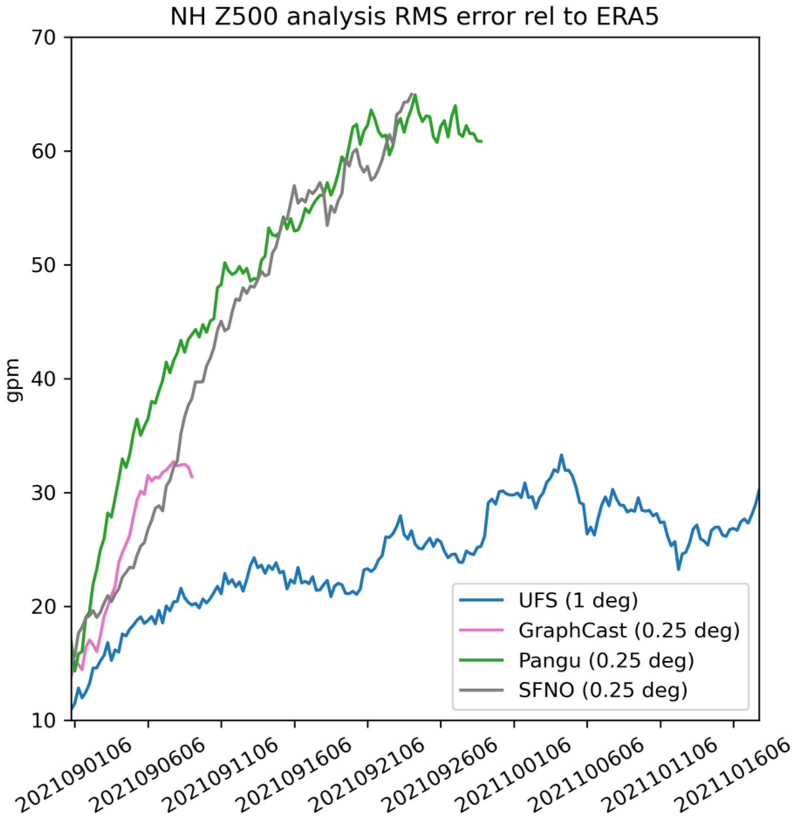
NeuralGCM results (spectral dycore, ML column physics)

- Stochastic version is stable but performs worse than UFS (but is also lower resolution)
- Deterministic version starts out performing well but becomes unstable *(consistent with developers' expectations)*



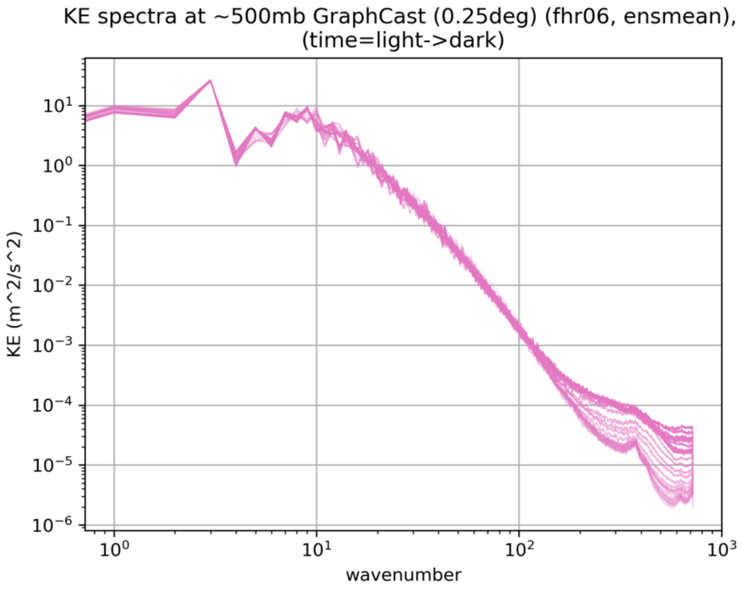
GraphCast, PanguWx, SFNO results (pure emulators)

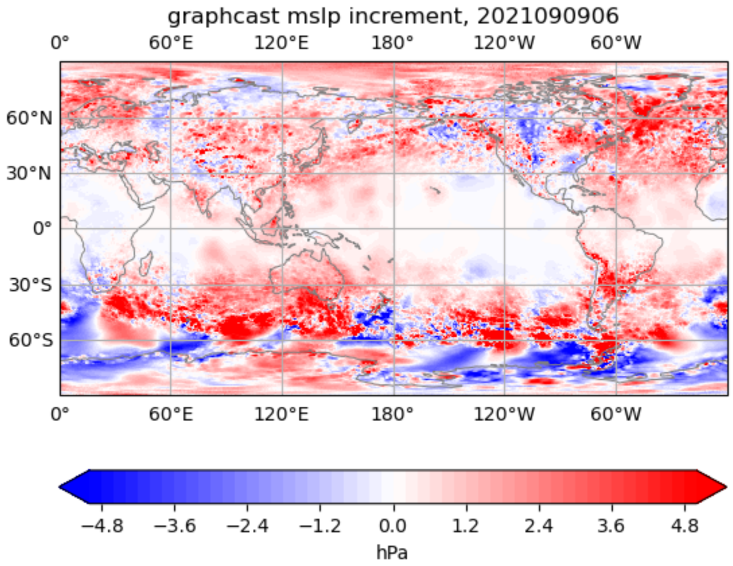
- Pure emulators are not stable (curves end when NaNs appear in output)



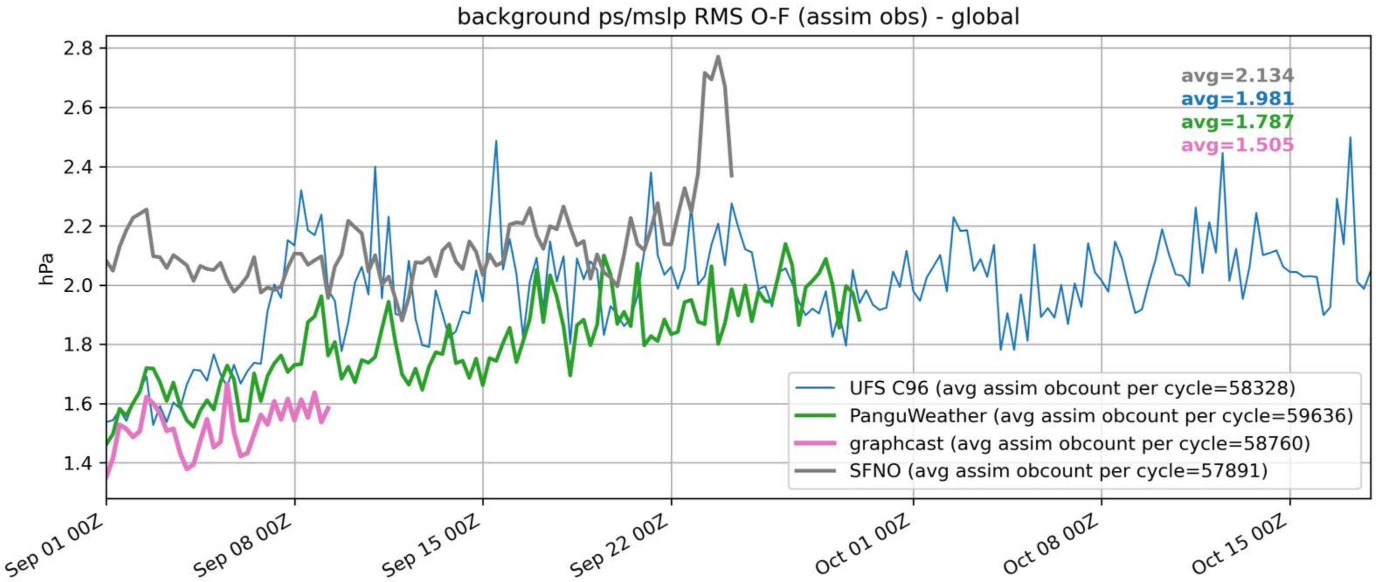
Pure emulators quickly develop instability

- Increments develop small-scale noise, and fields become unphysical





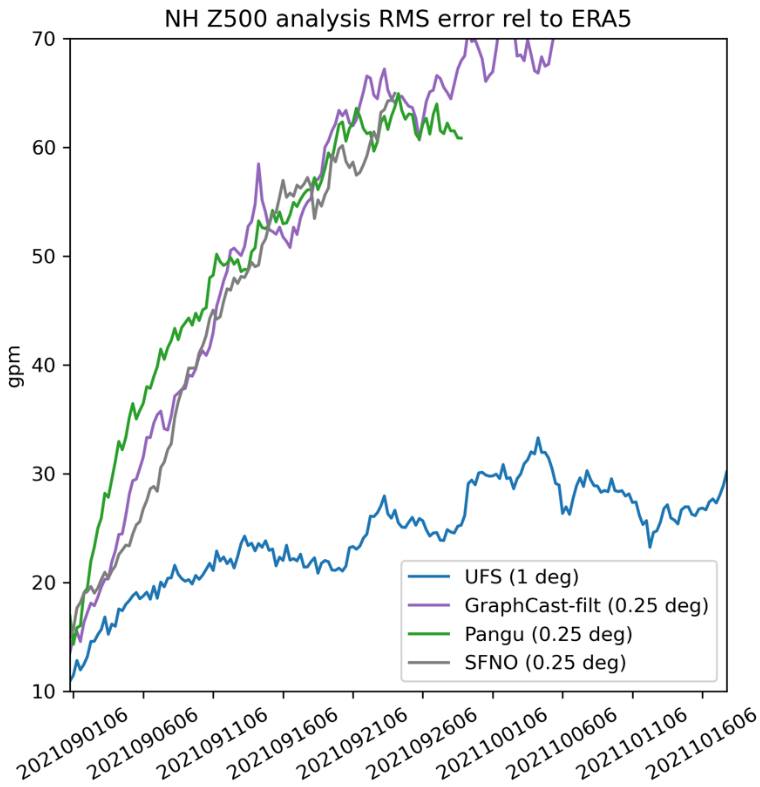
Pure emulators quickly develop instability



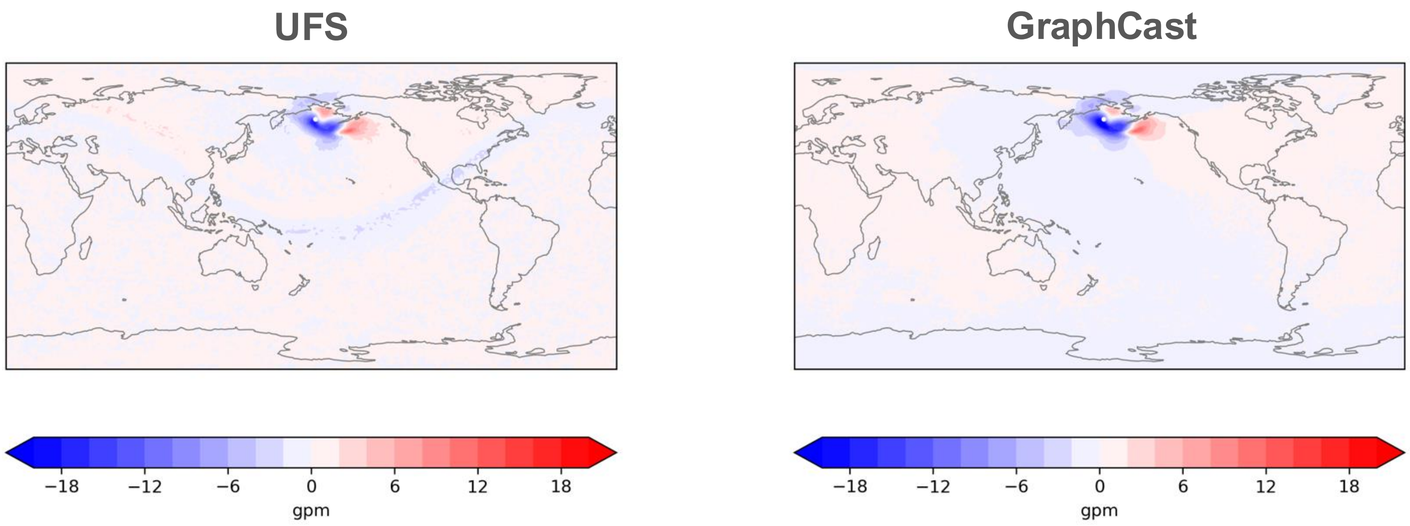
- O-F statistics perform well right up until emulators diverge (MSLP fields remain realistic at observation locations, and unobserved fields become unphysical.)
- 0.25deg emulators (Pangu & GraphCast) actually fit obs a little better than 1deg UFS

GraphCast with spectrally-filtered increments

- To prevent build-up of noise at small scales, a Gaussian spectral filter is applied to increments with e-fold scale n=80
- This stabilizes the cycling DA, but errors are more than twice as large as UFS – indicating that covariances are not effectively ‘spreading out’ the information from obs



Single ob experiments: Z500 forecast difference

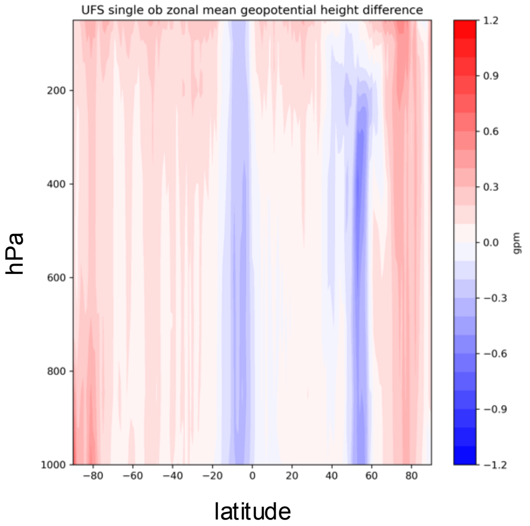


- Initialize two 6-hour forecasts with/without assimilating a single sfcp ob
- Map of Z500 differences are surprisingly similar between UFS and GraphCast (though GraphCast does not have gravity waves)

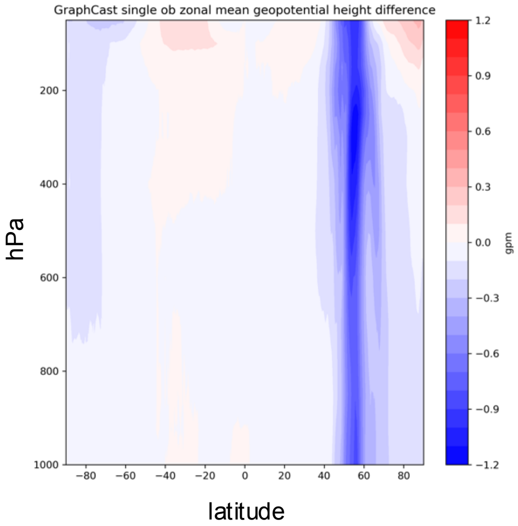
Single ob experiments: zonal mean geopotential height forecast difference

- Magnitude of difference in GraphCast is too strong, extends too far in the vertical

UFS



GraphCast



Summary and future work

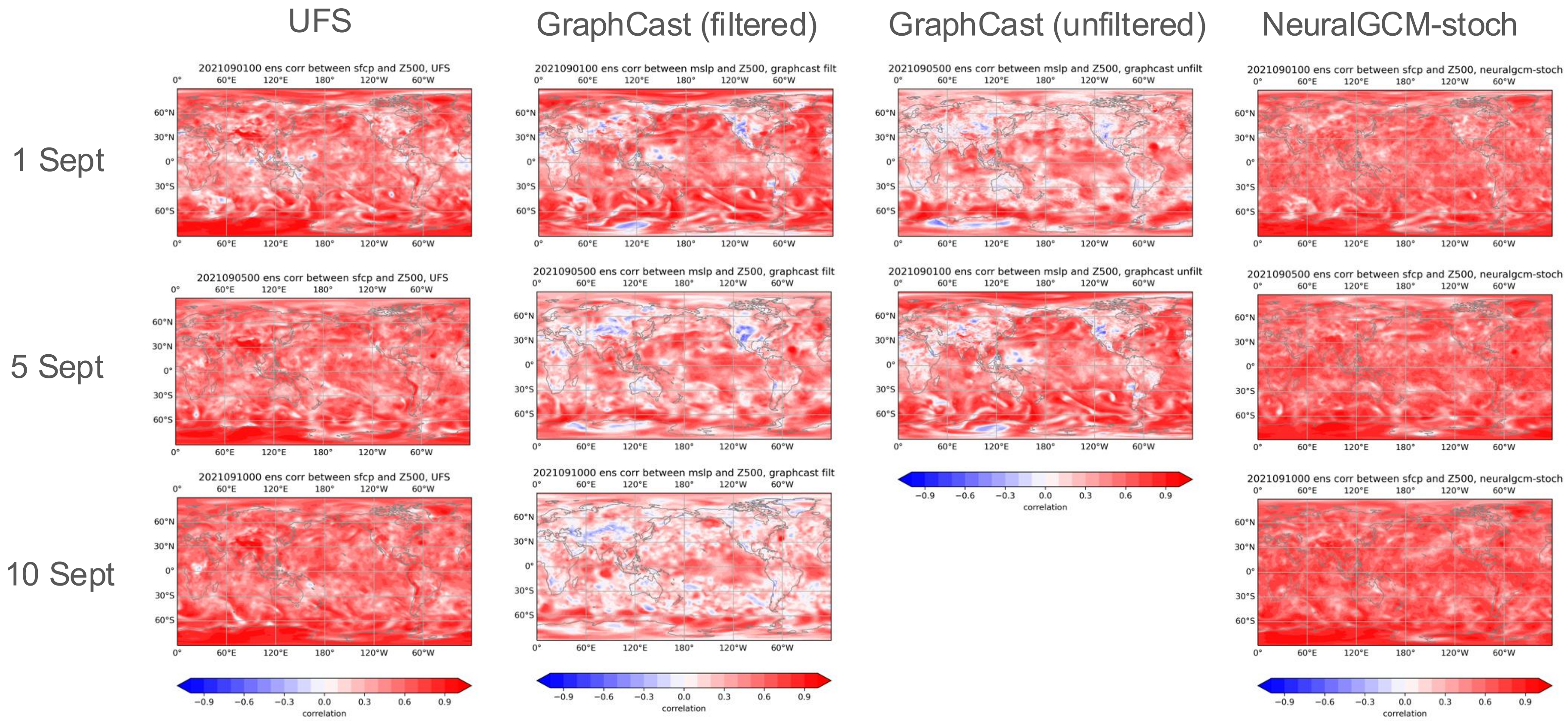
- Confronting models with observations via DA can effectively expose issues in the models
- Hybrid stochastic model (traditional dycore but emulated physics) can perform stably, but is less accurate than UFS
- Off-the-shelf pure emulators develop instability within 1-2 weeks of cycling
    - Build up of energy at small scales, development of unphysical fields
    - Filtering out small scales in increments stabilizes the system, but errors are large compared to UFS (indicating poor covariance estimate – supported by single ob experiments)
- Other work that has diagnosed problems with ML models:
    - Selz & Craig (2023): ML models cannot represent “butterfly effect”
    - Bonavita (2024): ML models lack physical consistency and cannot reproduce small-scale features
- Next: Develop our own emulators to investigate how training and architecture need to be modified to ensure stability and optimize performance in cycling DA

Laura.Slivinski@noaa.gov

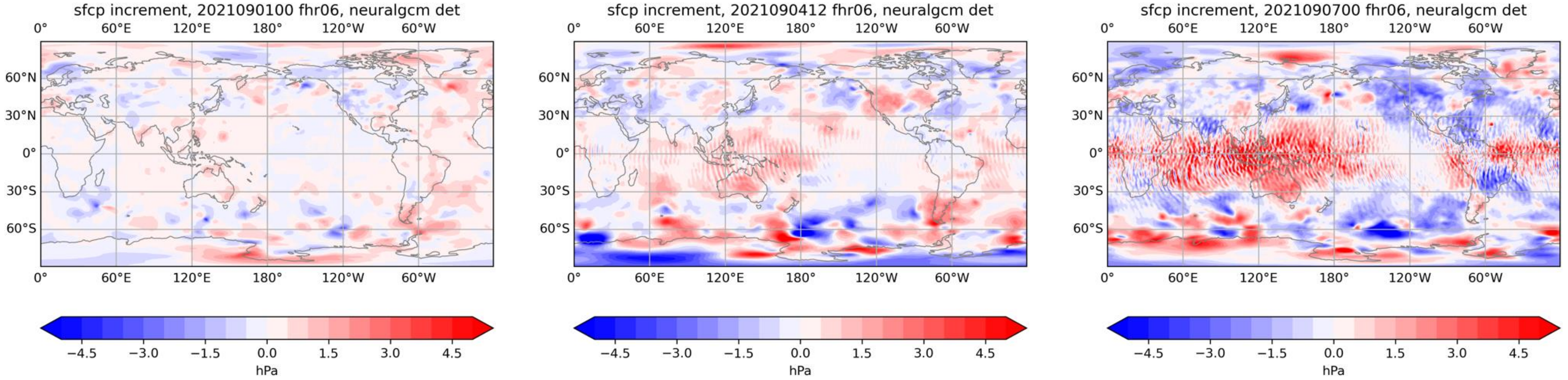
Thank you!

Extra slides

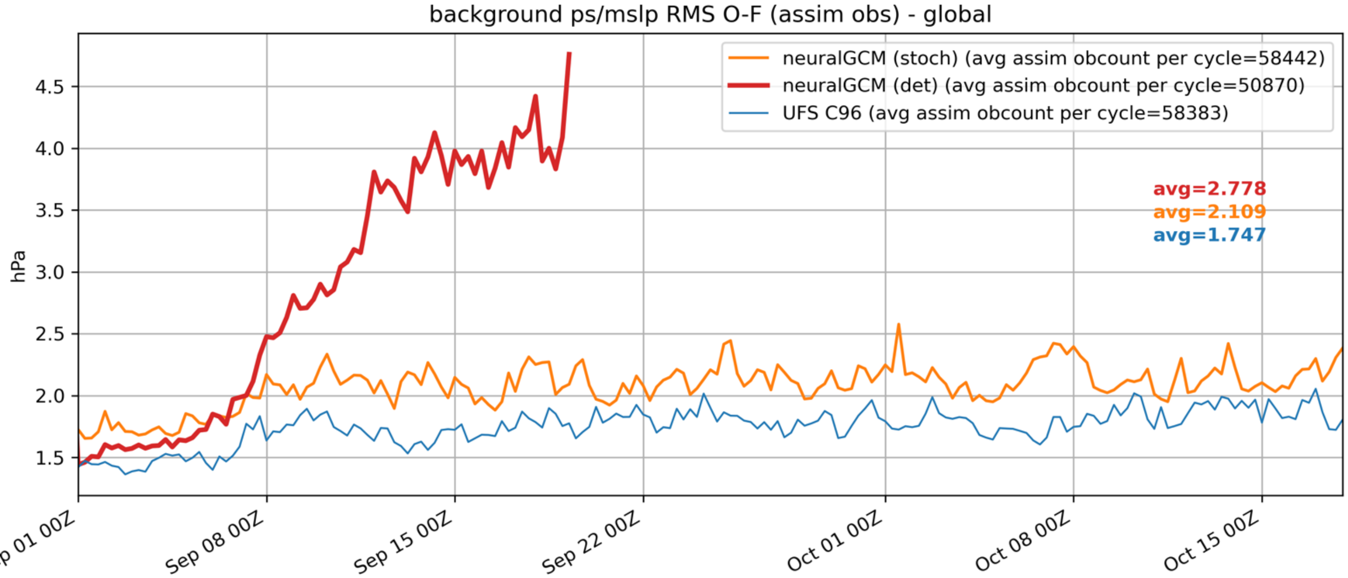
Ensemble correlations (sfcp/mslp and Z500)



Surface pressure increments (NeuralGCM-det)



NeuralGCM OmFs



Serial ensemble square-root filter update (1/2)

Given a single ob y^o with expected error variance R, an ensemble of model forecasts $\mathbf{x}^b$ (model priors – MxN matrix, where M is number of model state variables, N is ensemble size), and an ensemble of predicted observations $\mathbf{y}^b = \mathbf{H}\mathbf{x}^b$ (observation priors – vector length N):

Step 1: Update observation priors

(1a) $\overline{\mathbf{y}}_{\mathrm{a}} = (1-\mathrm{K})\overline{\mathbf{y}}_{\mathrm{b}} + \mathrm{K}\mathrm{y}^{\mathrm{o}}$ *update for obs prior means*

(1b) $\mathbf{y}_{\mathrm{a}}^{'} = \sqrt{(1-\mathrm{K})}\mathbf{y}_{\mathrm{b}}^{'}$ *rescaling of obs prior perturbations*

where the scalar $K = var(\mathbf{y}^b)/(var(\mathbf{y}^b)+ R)$, overbar denotes means, prime denotes perturbations, superscript b denotes prior, a denotes analysis.

Linear interpolation between observation and observation prior mean with weight K (0<=K<=1), rescaling of observation prior ensemble so posterior variance is consistent with Kalman filter covariance update, i.e. $var(\mathbf{y}^a)=(1\text{-}K)\ var(\mathbf{y}^b)= var(\mathbf{y}^b)R/(var(\mathbf{y}^b)+R)$.

when $var(\mathbf{y}^b) \ll R$, all weight given to prior.
when $var(\mathbf{y}^b) \gg R$, all weight given to observation.

Serial ensemble square-root filter update (2/2)

Step 2: Update model priors.
Let $\Delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b$ be analysis increment (change) for model priors, $\Delta y = y^a - y^b$ is analysis increment for obs priors.

(2) $\Delta\mathbf{x} = \mathbf{G}\Delta\mathbf{y}$ *computation of increments to model priors*

where $\mathbf{G} = \mathrm{cov}(\mathbf{x}^b, \mathbf{y}^{bT})/\mathrm{var}(\mathbf{y}^b)$ (an MxN matrix)

Linear regression of model priors on observation priors

Model priors only updated when $\mathbf{x}^b$ and $\mathbf{y}^b$ are correlated within the ensemble.
Localization: $\mathrm{cov}(\mathbf{x}^b, \mathbf{y}^{bT})$ *tapered to zero with distance from observation (mitigates sampling error)*

If there is more than one ob to be assimilated, the observation priors for other (not yet assimilated) obs ($\mathbf{Y}^b$, a KxN matrix, where K is the number of obs not yet assimilated) should also be updated using (2) with $\Delta\mathbf{x}$ replaced by $\Delta\mathbf{Y}$. Next iteration, replace $\mathbf{y}^b$ with next column of $\mathbf{Y}^b$, removing that column from $\mathbf{Y}^b$. After each iteration the model priors and observation priors are set to the latest analysis values ($\mathbf{x}^a$ replaces $\mathbf{x}^b$, $\mathbf{Y}^a$ replaces $\mathbf{Y}^b$). Continue iterating until $\mathbf{Y}^b$ is empty.