Towards end-to-end machine learning models that combine forecast with data assimilation

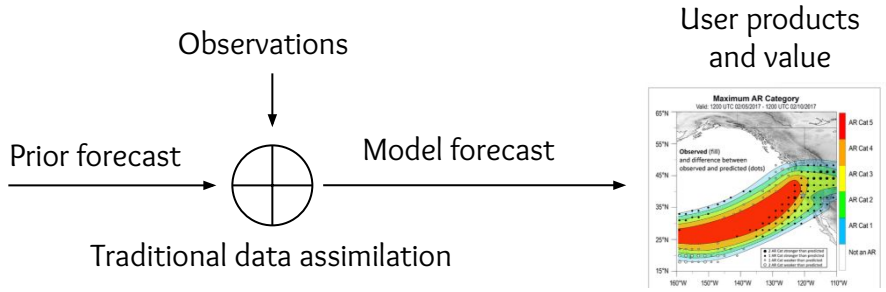
NOAA PSL: Sergey Frolov; Laura Slivinski; Tim Smith; Chong Chi Tong

MITRE: Matt Bender, Kelsey Lieberman, Joshua DaRosa, Nick Silverman, Chris Miller, Nick Krall, Mohammad Alam, Alex Philp

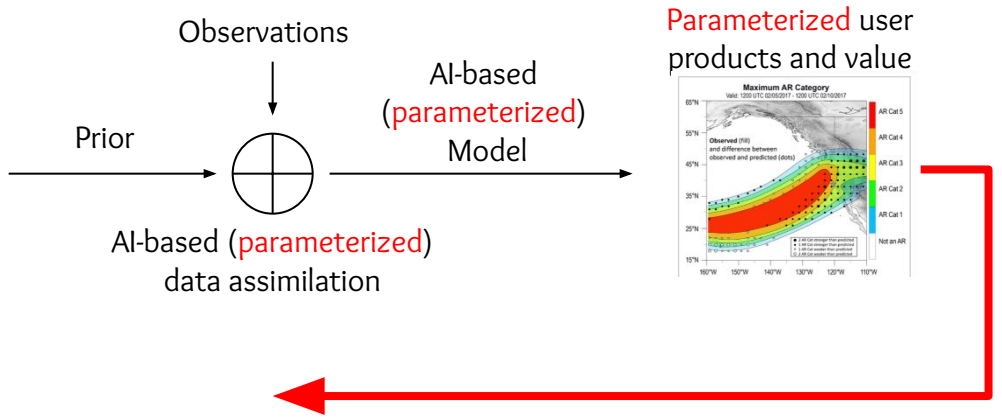
NVIDIA: Noah Brenowitz

Towards end-to-end forecast systems

Traditional weather services pipeline



End-to-end AI-based weather services pipeline



Optimization that maximizes value by adjusting parameters of the end-to-end system

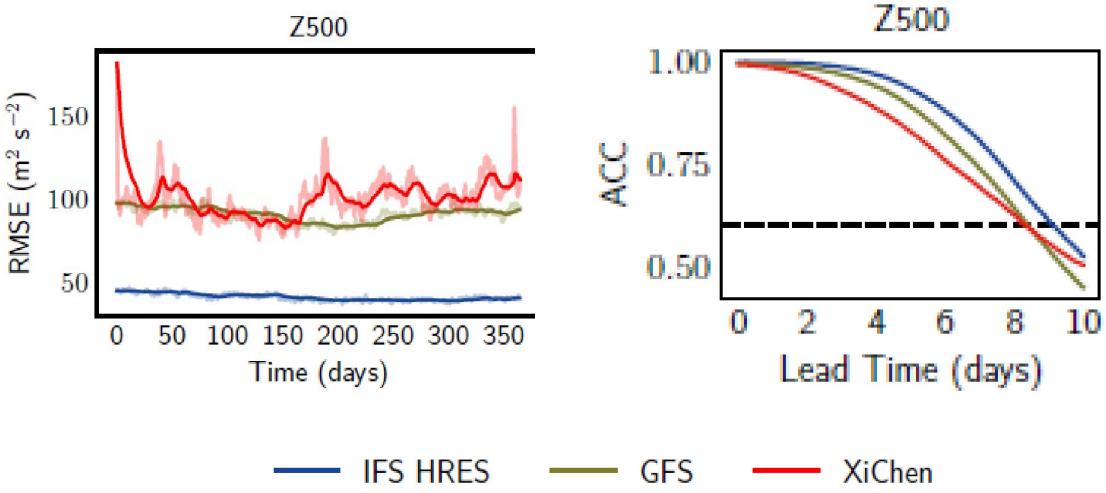
Can we develop an end-to-end DA + forecast system that improves use of observations and improve forecast products?

The sputnik moment

XiChen: An observation-scalable fully AI-driven global weather forecasting system with 4D variational knowledge

Wuxin Wang[1,2], Weicheng Ni[2], Lilan Huang[1,2], Tao Han[3], Ben Fei[3], Shuo Ma[2], Taikang Yuan[2], Yanlai Zhao[2], Kefeng Deng[2], Xiaoyong Li[2], Boheng Duan[2*], Lei Bai[3*], Kaijun Ren[1,2*]

[1*]College of Computer Science and Technology, National Univiersity of Defense Technology, Deya Street, Changsha, 410073, Hunan, China.
[2*]College of Meteorology and Oceanography, National Univiersity of Defense Technology, Deya Street, Changsha, 410073, Hunan, China.
[3*]Shanghai Artificial Intelligence Laboratory, 200030, Shanghai, China.



- Defense universities in China demonstrated (July 2025) that it is possible to stably cycle full AI models (AIDA + AIFC + AIObs) with just a small subset of real observations.
- And the results look encouraging and credible.

What is the simplest problem one could try?

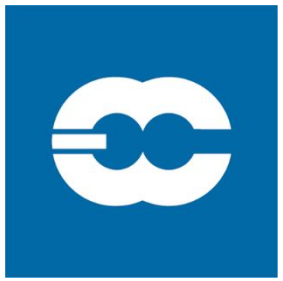
- Figuring out the complete end-to-end training and model pipeline is daunting.

  - What is the simple problem we can try and solve?

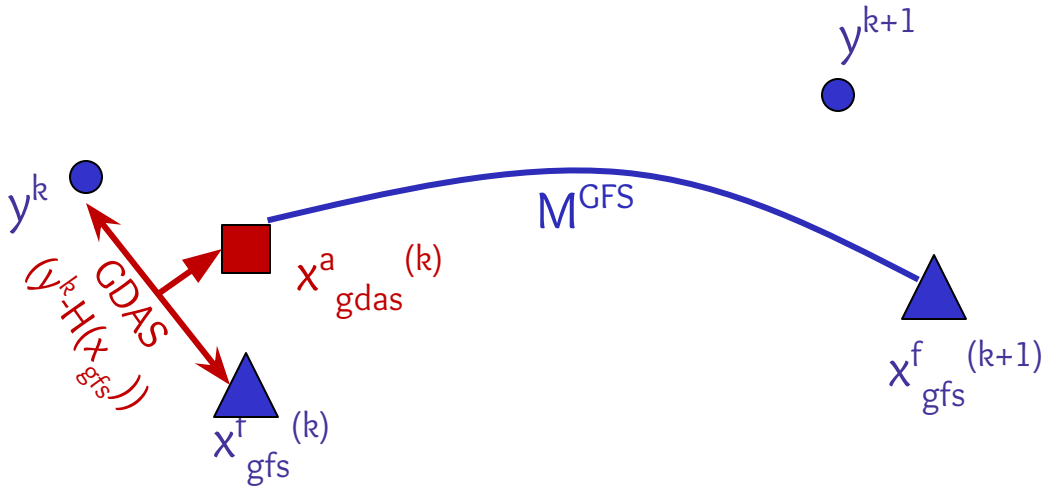
A collaboration between NOAA, MITRE, and NVIDIA
Using technology from ECMWF

MITRE

NVIDIA

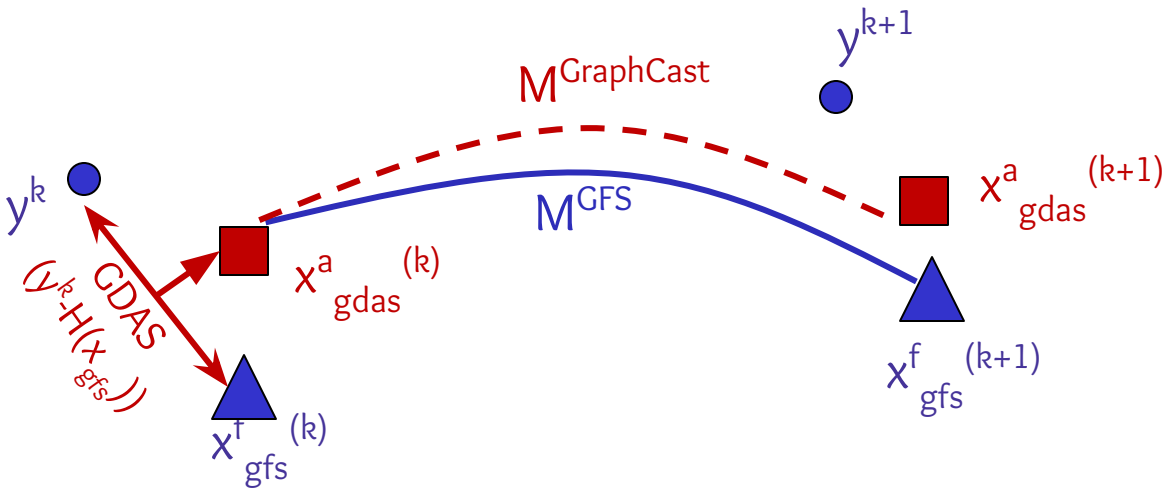


ADD-DA: step towards end-to-end training



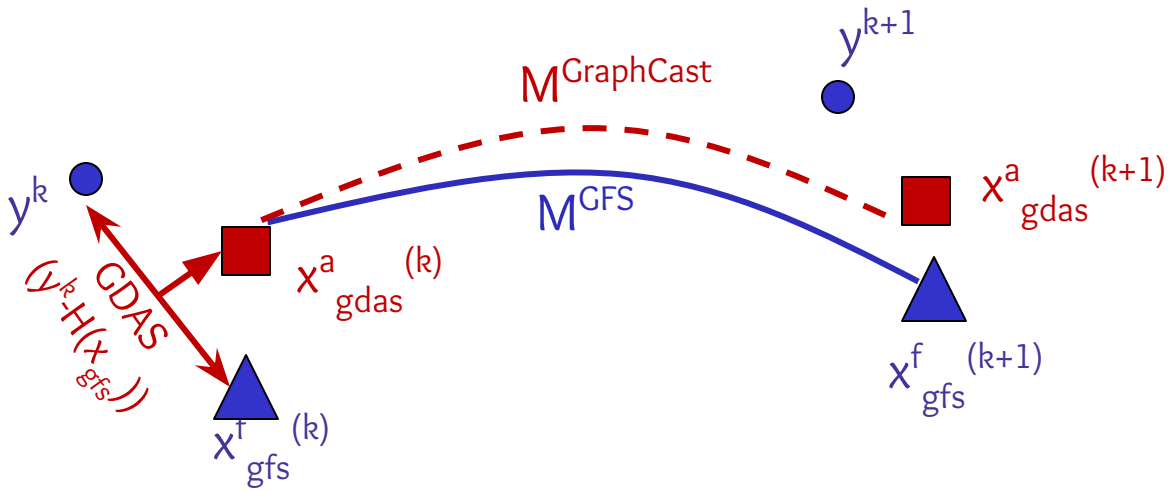
Single forecast step of the traditional NWP

ADD-DA: step towards end-to-end training



In GraphCast-like algorithm, we assume that "GDAS" or ERA5 analysis is perfect.

ADD-DA: step towards end-to-end training

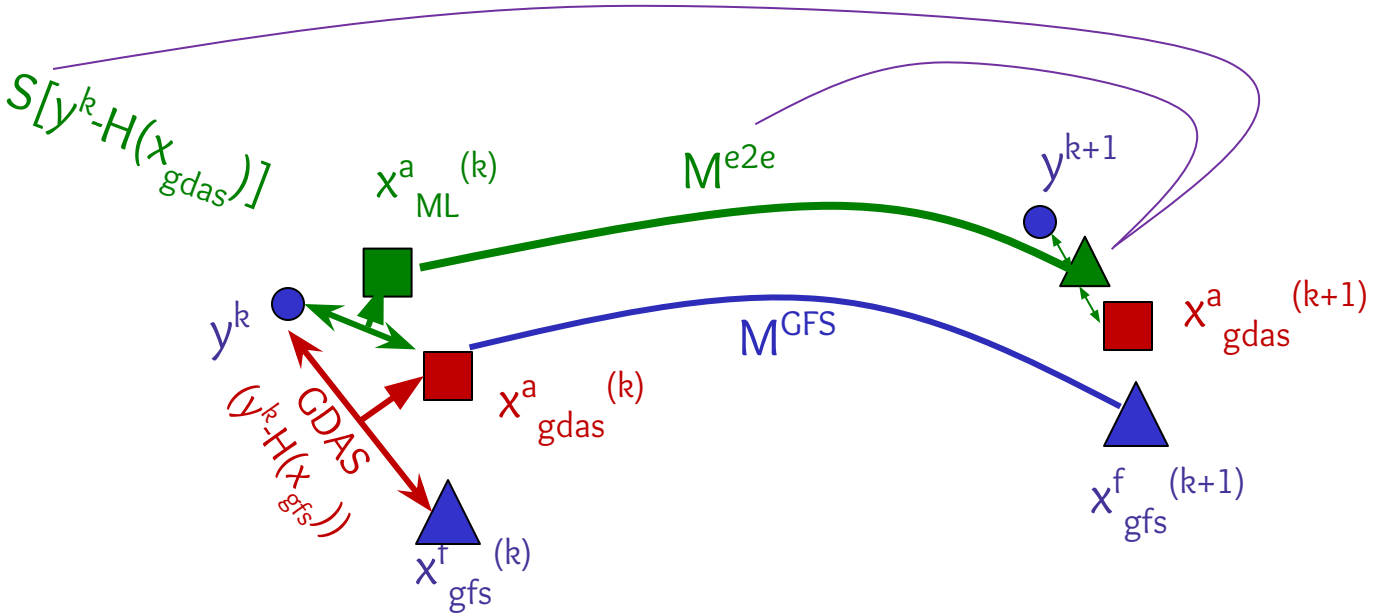


In GraphCast-like algorithm, we assume that "GDAS" or ERA5 analysis is perfect.

Assumptions that we want to relax:

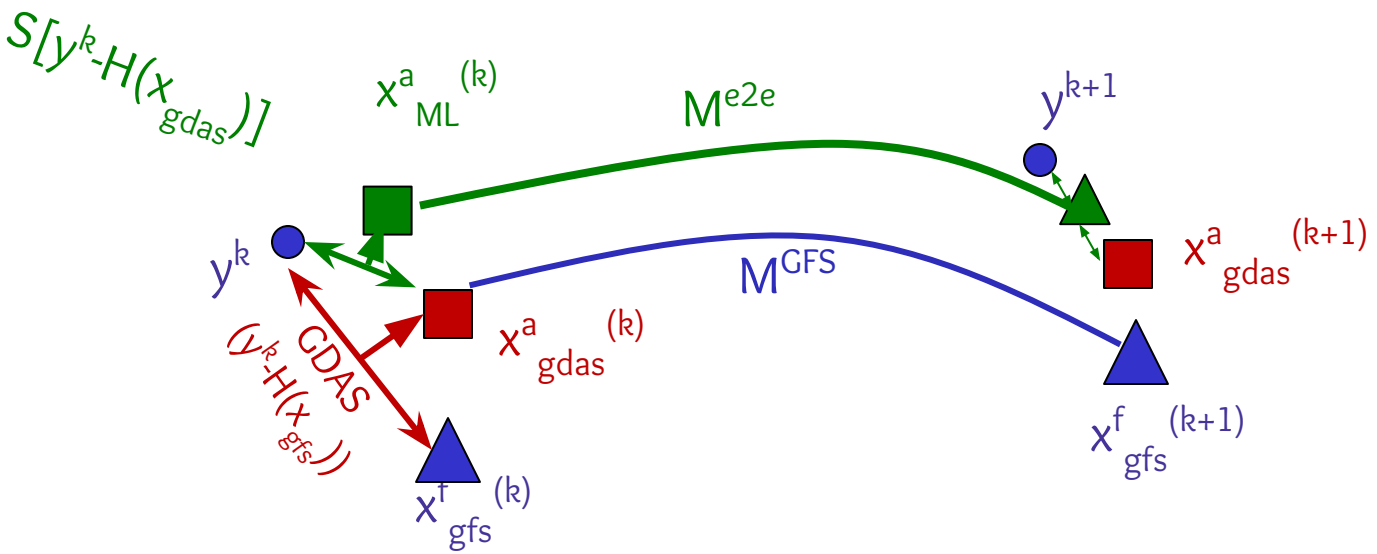
- Quality of the analysis is constrained by the quality of the IFS/UFS.
  - Could ML model improve our representation of the boundary layer, precipitation, and clouds over traditional models?
  - Could we increase effective resolution of ML models?
- Not all observations are used to their full extent to generate GDAS analysis.
  - Could we start using direct measurements of precipitation, surface obs, surface-sensitive radiances in the end-to-end ML framework?

ADD-DA: step towards end-to-end training



- Suggestion:
  - Start with the GDAS analysis x^a_{gdas}.
  - Then generate an improved initial condition x^a_{ML} that uses:
    - Innovations between GDAS analysis and obs. $[y^k\text{-}H(x_{gdas})]$
    - Learns from backpropagation of errors in the end-to-end forecast system against future analysis and observation.
  - Simultaneously train initialization operator S and model forecast M^{e2e}.

ADD-DA: step towards end-to-end training

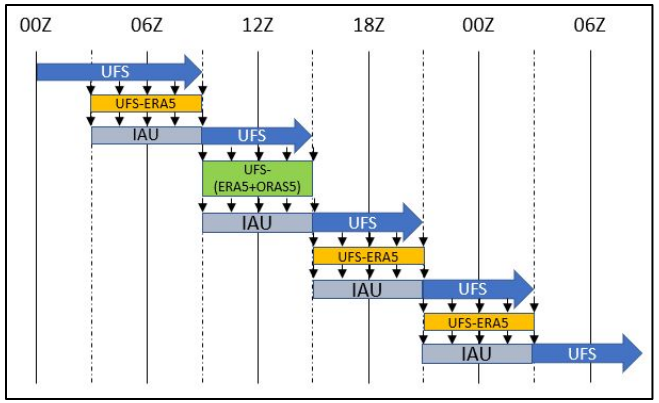


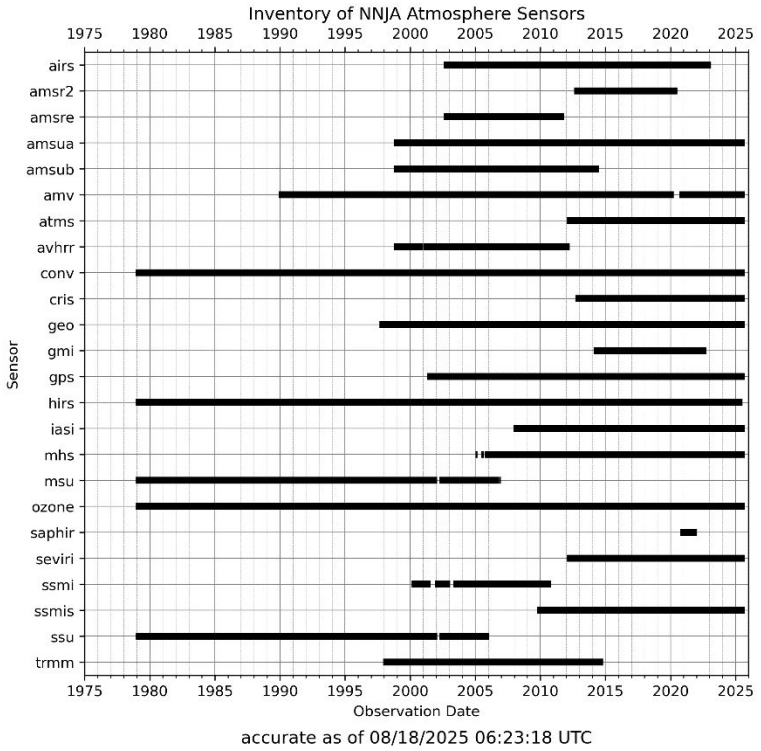
- Suggestion:
  - Start with the GDAS analysis x^a_{gdas}.
  - Then generate an improved initial condition x^a_{ML} that uses:
    - Innovations between GDAS analysis and obs. $[y^k\text{-}H(x_{gdas})]$
    - Learns from backpropagation of errors in the end-to-end forecast system against future analysis and observation.
  - Simultaneously train initialization operator S and model forecast M^{e2e}.

Example: To improve representation of the water vapor burden in the AR, we need to simultaneously:

- Improve the I.C.,
- Improve the forecast model, and
- Draw to surface-sensitive observations more closely than to the surface analysis
- Still draw closely to the GDAS analysis for the free atmosphere.

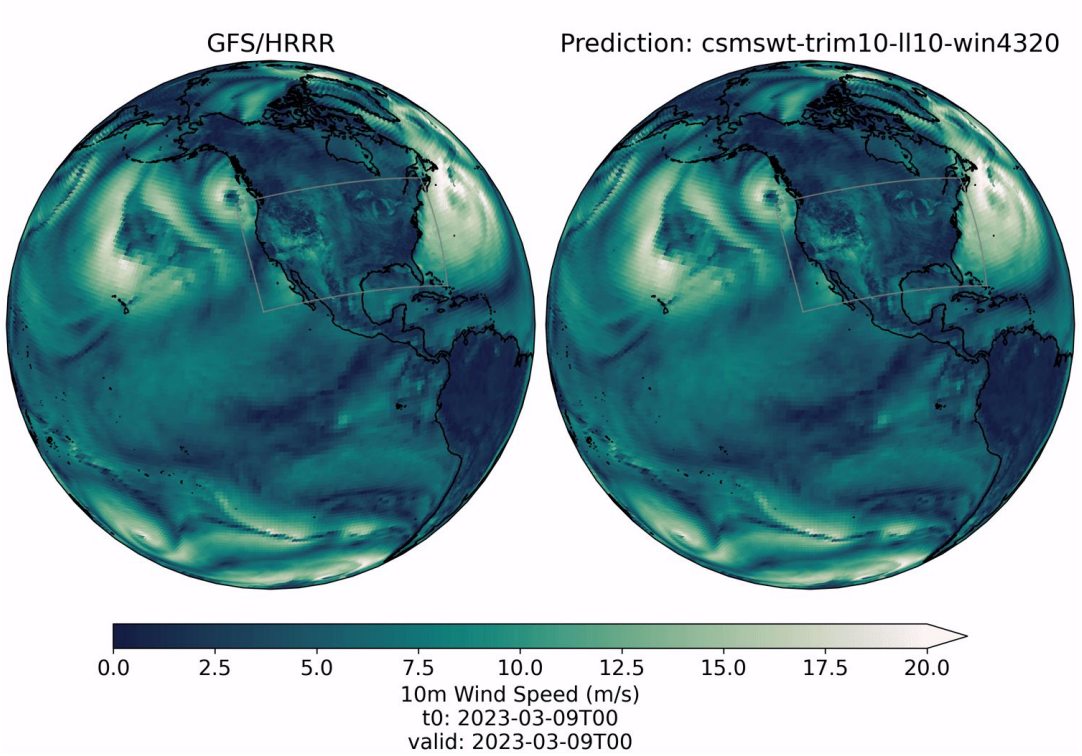
Training datasets





- UFS-Replay:
  - UFS coupled model "replayed" to ERA5 and ORAS5 reanalysis.
  - All of the gridded output available in zarr on GCS.
- Replay observer:
  - GSI observer ran on 30 years of UFS-Replay backgrounds to generate CRTM targets and observation bias corrections.
  - Available on S3
- NNJA: NOAA NASA Joint archive of observations for reanalysis:
  - Available in legacy, AI-ready parquet, and traditional netcdf formats on S3 and GCS.

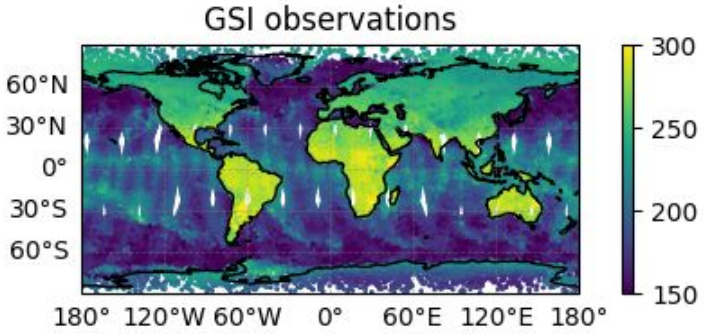
Current state of work: Forecast operator M

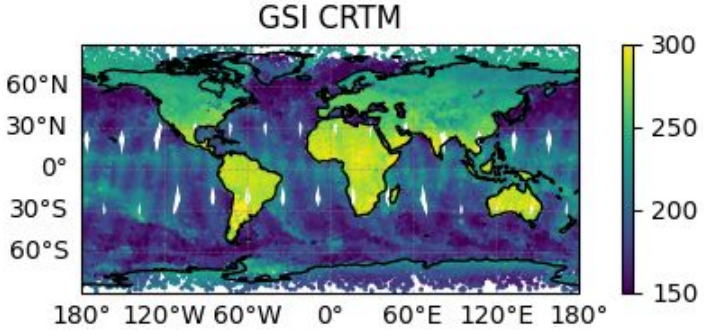


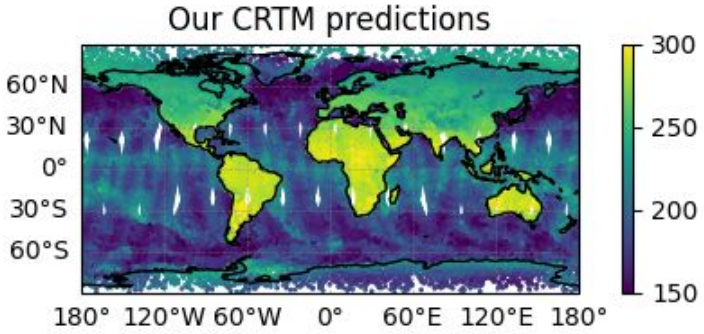
- Anemoi:
  - Developed by ECMWF and European partners (20-50 FTEs)
  - Basis for AIFS operational model
  - Adopted by NWP centers in Europe, UKMO, Canada, and now NOAA
- Adopted European Anemoi framework for AI forecast modeling to NOAA data:
  - GFS, GDAS global analysis;
  - HRRR analysis and forecast;
  - UFS-Replay;
  - Future: NNJA observations.
- Global, Nested, and LAM models are all under development.

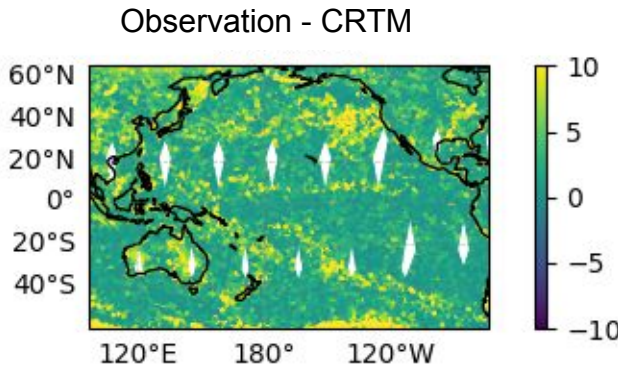
Figures courtesy of Tim Smith from NOAA PSL

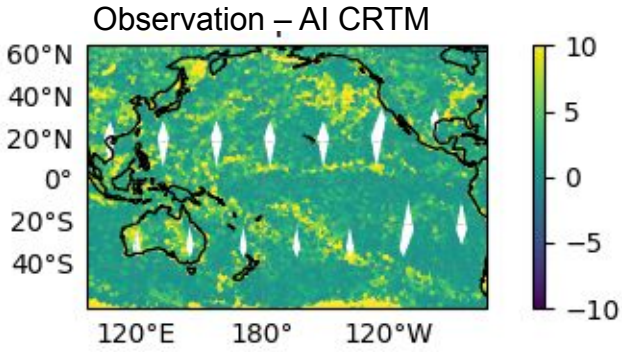
Current state of work: observation operator H





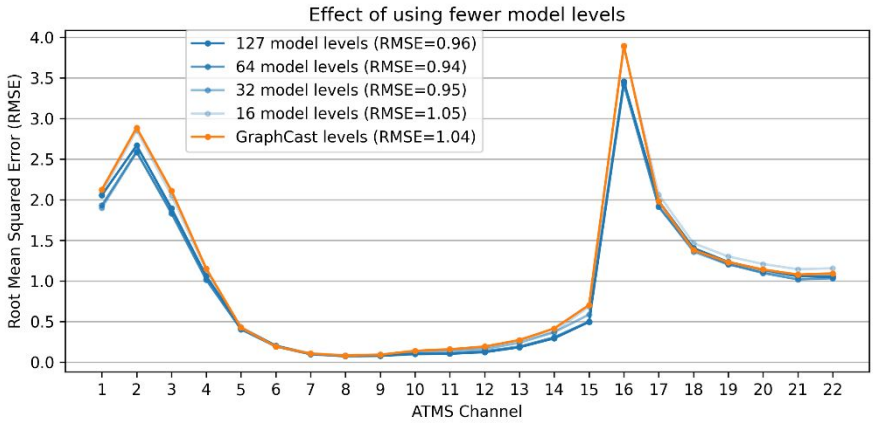






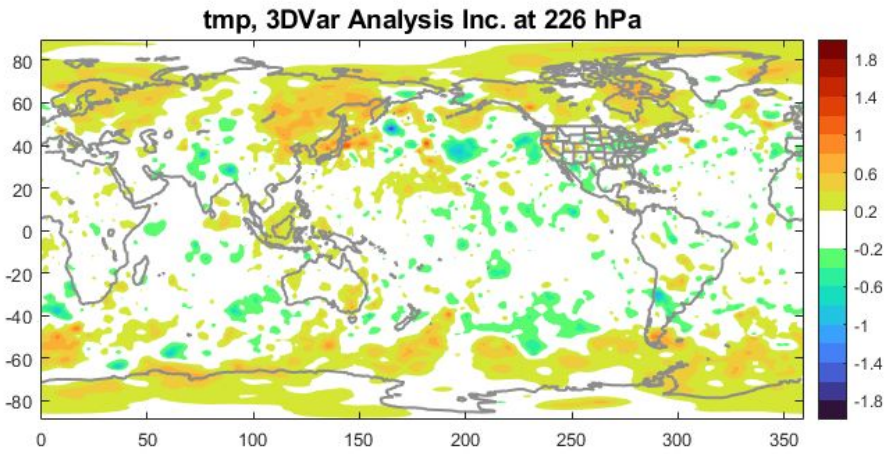
- Developing AI observational emulator appropriate for AI forecast models:
  - Focus on accurate simulation of OMB.
  - With fewer input layers from AI model.
  - Following AI-CRTM work from NESDIS.

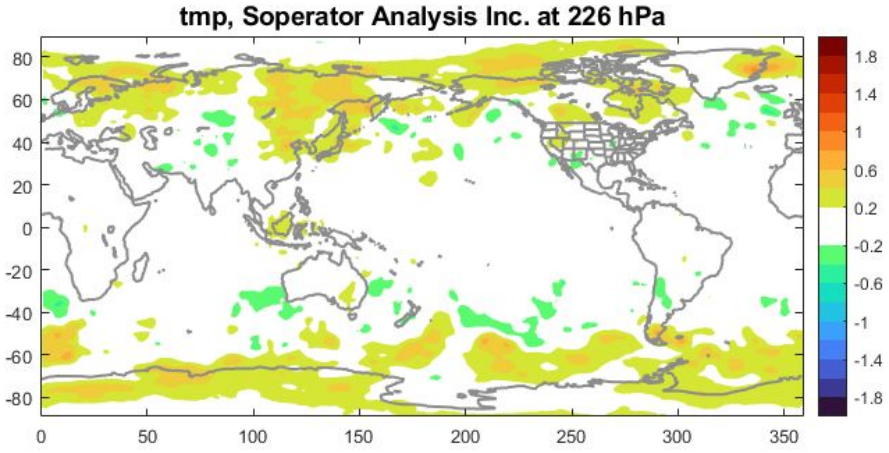
Ablation study with number of input levels



Figures courtesy of Kelsey Liberman from MITRE

Current state of work: DA operator S





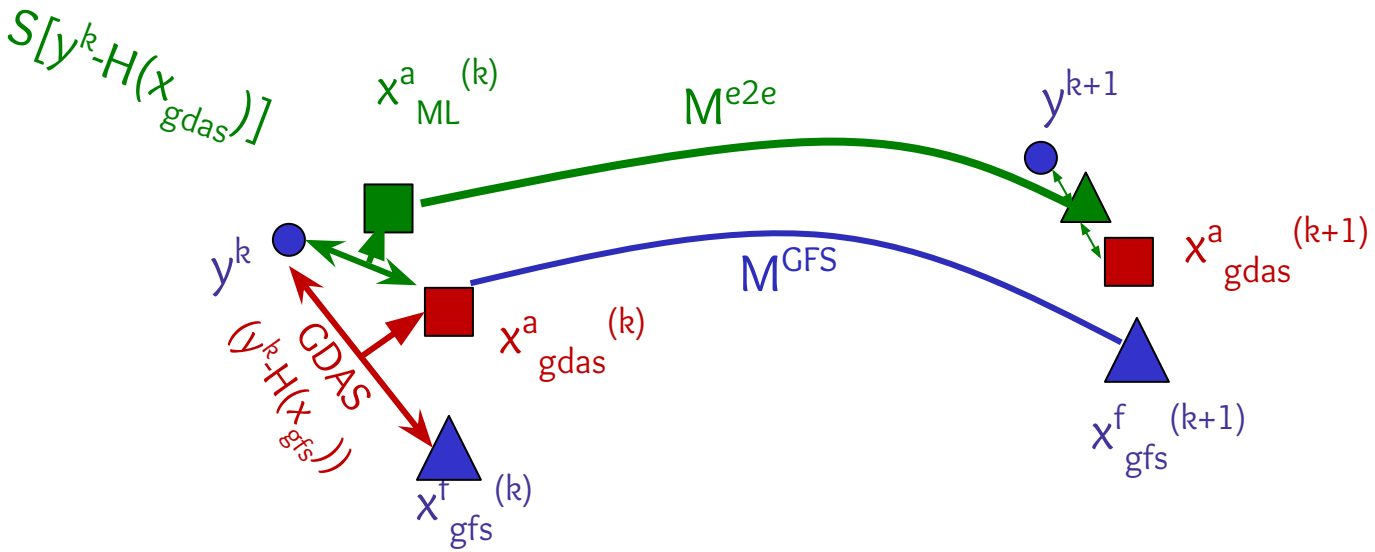
Pre-trained DA operator:

- Trained on a history of observations (conventional and ATMS) and 3DVAR increments from the UFS replay observer.
- Resulting increments are credible but are too smooth compared to the 3DVAR increment.

Next steps:

- Fine-tune the operator to produce additional increment to the GDAS initial condition that will produce better 3-day forecast compared to the forecast from GDAS.

Current state of work: putting initial conditions and forecast together



- MITRE is combining all parts of the pipeline together so we can start fine tuning S operator against future analysis and observations:
  - Discussing strategies for incremental development rather than throwing all obs at the same time.
- Very encouraged by the XiChen and FuXi results that are taking similar approach.

Next steps

- Fine tune data assimilation model to increment GDAS analysis to optimize day forecast.

- Fine tune data assimilation and forecast model against observations.

- Focus on improved utilization of the underexploited observations like the surface sensitive channels.

- Extend the problem to the ocean.

End