A "replay" reanalysis dataset for initializing coupled reforecast ensembles with the UFS
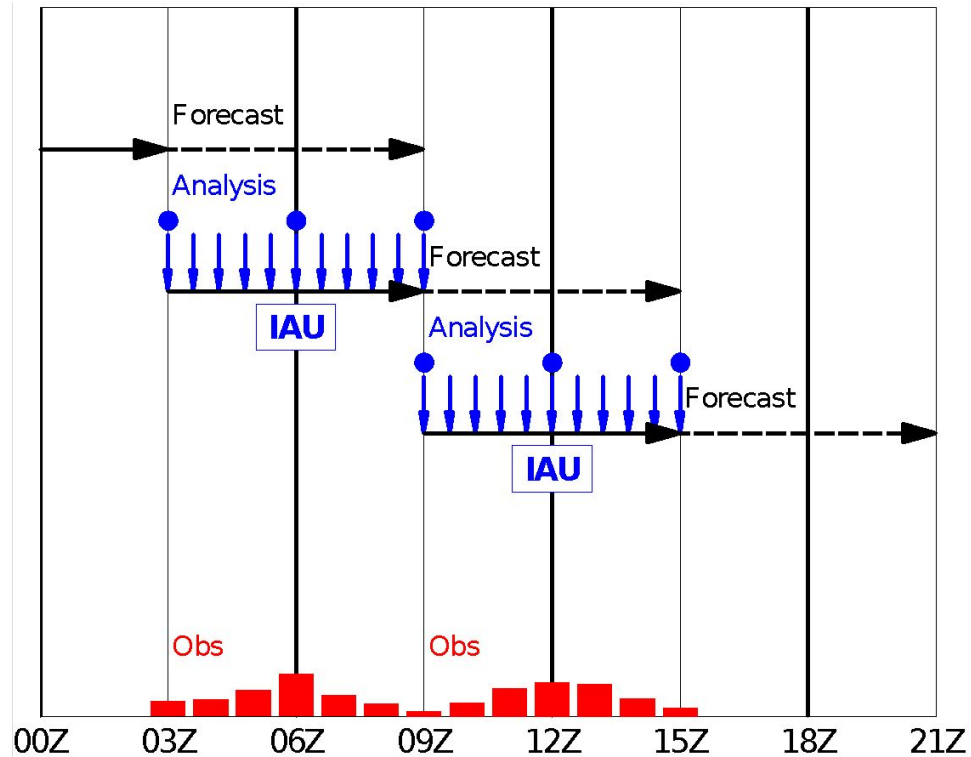
Jeff Whitaker and Phil Pegion
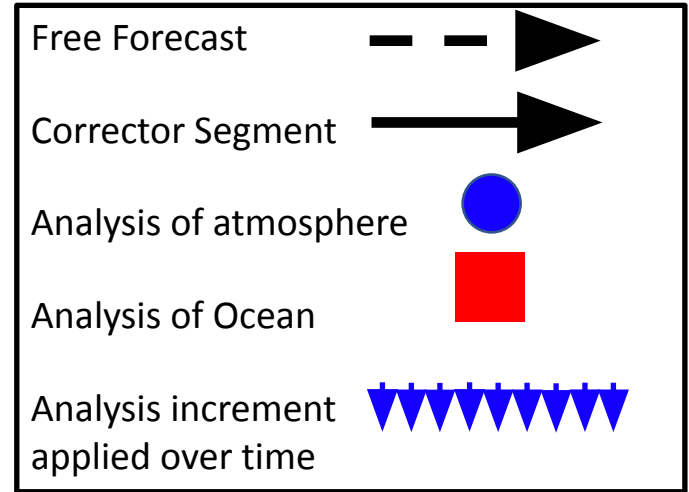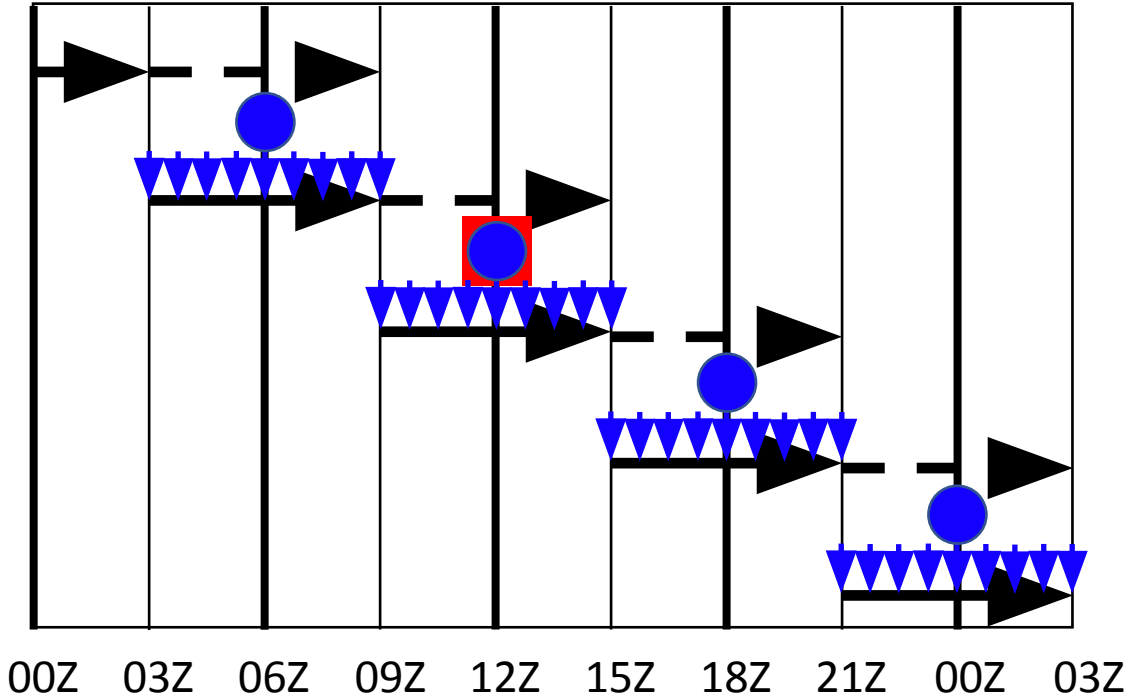*NOAA Physical Sciences Lab*

Motivation

- GEFSv13 will be coupled, need a coupled (atm/ocn/land/sea-ice) reanalysis to initial reforecasts.
- Don't have time or resources to run a coupled reanalysis (and then reforecasts) in time for implementation.
- As a stop-gap measure, we are using a technique called 'replay' to generate balanced coupled initial conditions for the UFS from existing reanalyses.
    - We do plan to run a coupled reanalysis to initialize SFSv1 reforecasts.
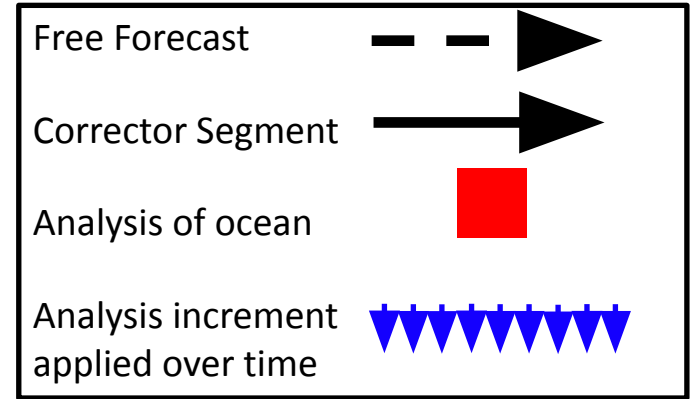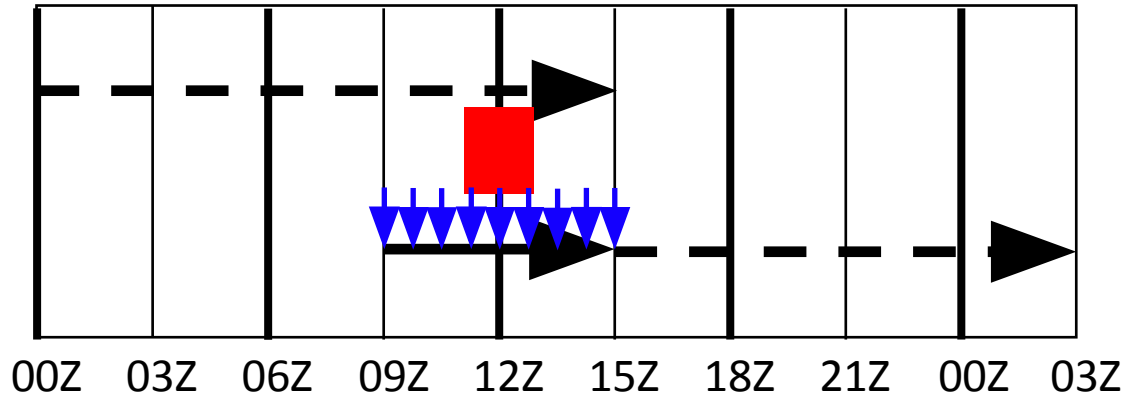
The data assimilation cycle

Coupled Replay Cycle

Free Forecast

Corrector Segment

Analysis of atmosphere

Analysis of Ocean

Analysis increment applied over time

00Z 03Z 06Z 09Z 12Z 15Z 18Z 21Z 00Z 03Z

Ocean Replay Cycle

We could use a full 24-hour forecast and corrector segment, but this way is consistent with the coupled replay.

Advantages of replay

- Can easily re-run the reanalysis without needing the data assimilation if new output needed (different frequency, missing variables) or if model is upgraded but there is not yet a consistent reanalysis.
  - Envision reanalysis production on a 7 year cycle, model upgrades more like every 2 years.
- Can add additional components to modeling system and re-run (such as aerosols: Orbe et al. 2017).
- Use mean replay increments to assess model errors, train ML-based bias correction schemes: (Chen et al 2022).
- Constrain certain portions of the system, i.e. only replay tropics: Dias et al. 2021
- Avoids 'spin-up/spin-down' problems seen in 'cold start' forecasts by gradually adding in increments. Unconstrained components (e.g. land) can adjust consistently.

Implementation details

- 1994-2023, full coupled restarts once per day (at 03Z, informed by analyses up to including 00Z).
- Publicly available on AWS (no egress charges)
  - https://noaa-ufs-gefsv13replay-pds.s3.amazonaws.com/index.html
- Land unconstrained
  - Except for snow, which is updated using JEDI OI-based DA.
- ORAS5 sea-ice inserted.
- Ensemble perturbations for reforecasts from ERA5 analysis perturbations in atmosphere, and lagged differences between four reanalyses in ocean (no initial perturbations for land and sea-ice).
- Being run in 6 overlapping streams, with first year discarded from each stream.
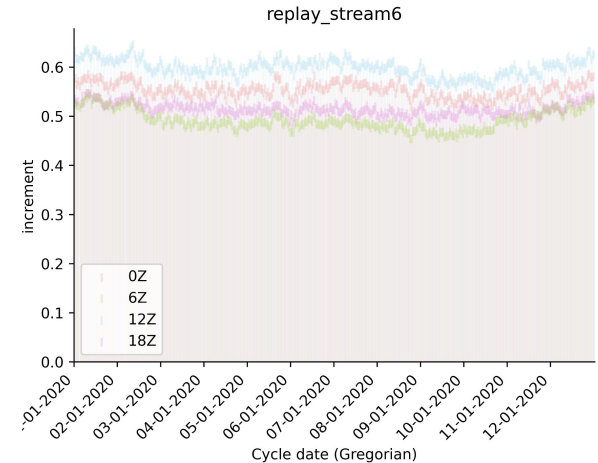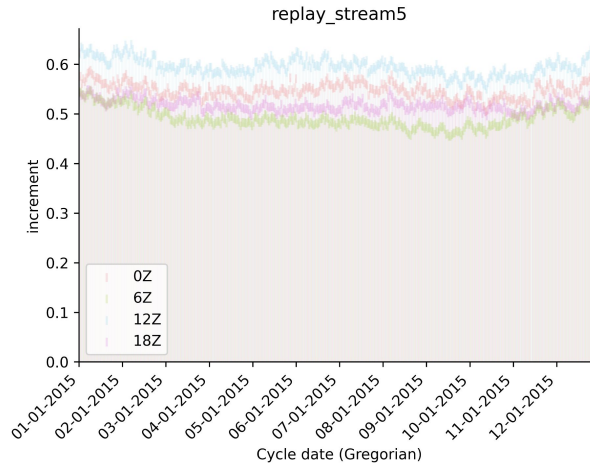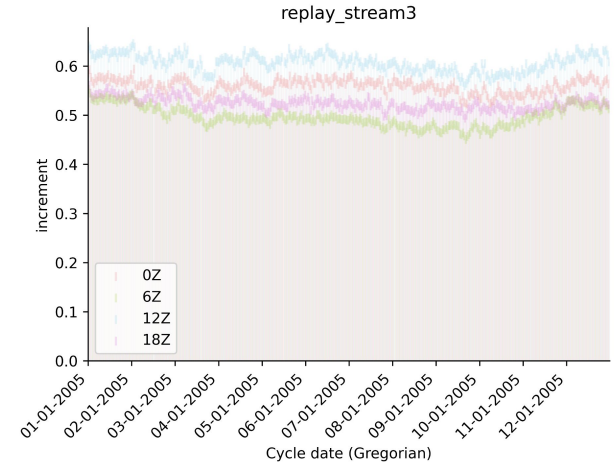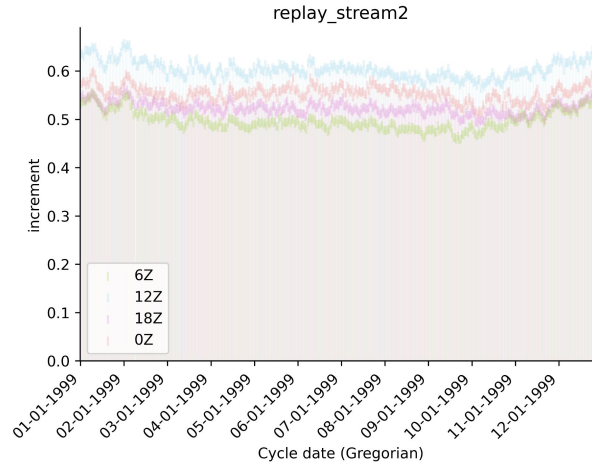 - Streams initialized from ORAS5/ERA5 and NOAH-MP forced by ERA5.

Replay progress

- All streams in at least 2nd year of production.
- Estimated completion Oct 1.

replay stream progress as of August 16 2023 - 16:15 UTC
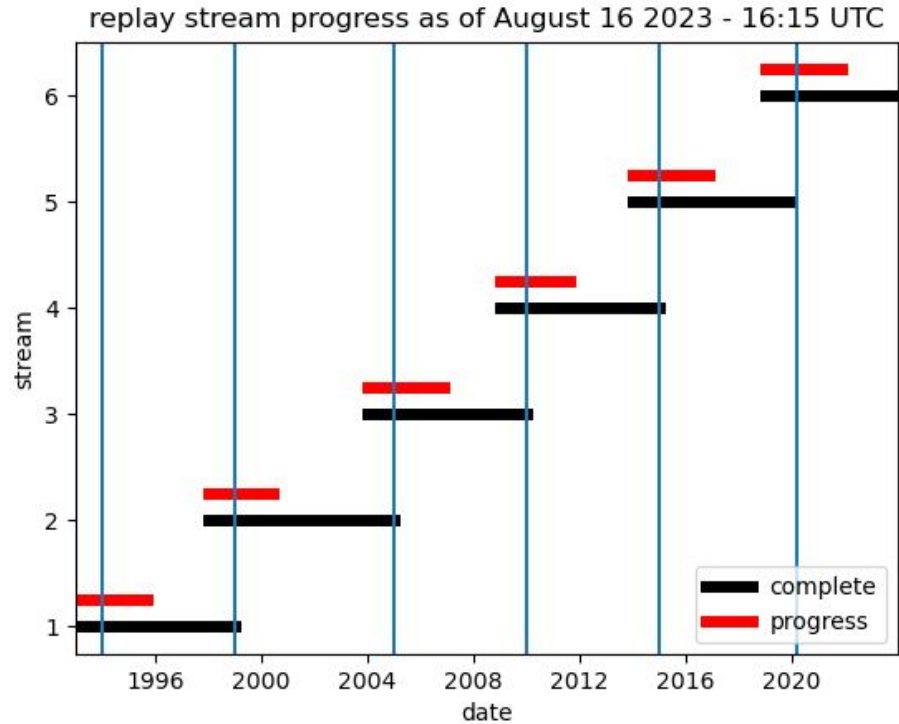
QC/Monitoring

- Plots routinely generated to monitor quality, detect problems.
- Global mean atm T increment RMS show here.

Replay-initialized ensemble reforecast evaluation (courtesy Bing Fu/EMC)

- Test period that spans May 2019 through November 2020
- Reference forecasts are our best proxy for operational initial conditions for GEFSv13 which consists of experimental JEDI/SOCA ocean and ice analyses, GFSv16 atmospheric analysis and spun-up Noah-MP land initial conditions. Labeled ctrl
- Replay ensemble forecasts are initialized with restarts at the end of the IAU window (FHR03) and perturbations are applied to the atmosphere and ocean initial conditions. Labeled rply
- 40 cases of 6-member ensemble forecasts out 16-days

500 hPa spread/error

Northern Hemisphere 500hPa Height
Ensemble Mean RMSE and Ensemble SPREAD
Average For 20190509 — 20201105
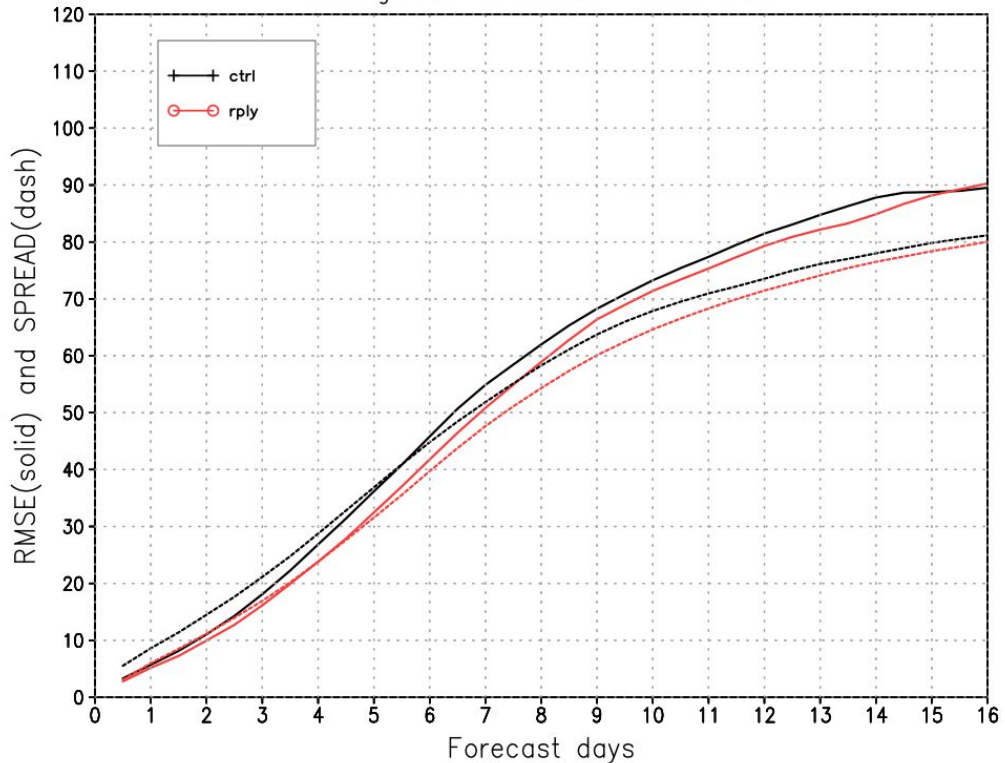
Ensemble forecasts with replay initial conditions are substantially better than the control experiments with about 12-hours of increased skill at day-5.

Ensemble spread is also lower in replay and matches the RMS error quite well for the first 6 days.

Summary

- A 30 year daily dataset (1994-2023) of coupled UFS restarts (for ATM, OCN, land, sea-ice and waves) at ¼ degree will be publicly available on AWS this fall.
    - Informed by ECMWF reanalysis of atmosphere, ocean and sea-ice, plus snow observations.
    - Includes perturbations for atmosphere and ocean for generating ensembles.
- Will be used to initialize reforecasts for GEFSv13.
- Also useful for research, model development.
- Increments should be useful for identifying regime-dependent biases
    - Ongoing work at PSL to use these increments to train an in-line ML-based bias correction.