An Operational-Quality Ensemble Assimilation System for NCAR's CAM Climate Model

Tim Hoar, Jeff Anderson, Kevin Raeder, Hui Liu
U. S. National Center for Atmospheric Research
18 October 2004

A group filter approach.

Supported by the U.S. National Science Foundation DMS

Who We Are / What We Do

The U.S. National Center for Atmospheric Research (NCAR) Data Assimilation Initiative (DAI) employs:

- Jeffrey Anderson (PI - formerly of GFDL),
- Kevin Raeder (integration of CAM),
- Hui Liu (BUFR obs & GPS obs),
- Alain Caya (WRF & Radar Obs),
- and me, Tim Hoar (none of above)

with the goal of providing a flexible, modular, environment for exploring Data Assimilation methodologies rather than differences of software implementation.

With this many people, we are focusing on Ensemble Methods; EnKF, EaKF...

Data Assimilation Research Testbed → DART

- Many low-order models: Lorenz63, L84, L96*,L2004, . . . ,
- Global 2-level PE model (from NOAA/CDC),
- NCAR's CAM 2.0 & 3.0 (global spectral model),
- NCAR's WRF (regional),
- GFDL FMS B-grid GCM (global grid point model),
- MIT GCM (from Jim Hansen),
- NCEP GFS (assisted by NOAA/CDC),
- GFDL MOM3/4 ocean model,
- NCAR's ROSE model (upper atmosphere with chemistry)

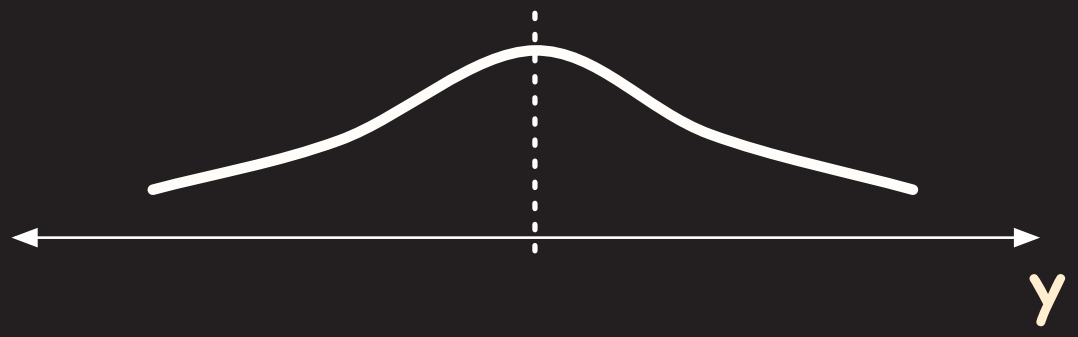
Forward Operators and Datasets

- Many linear, non-linear forward operators for low-order models
- U, V, T, Ps, Q, for realistic models
- Radar reflectivity, GPS refractivity for realistic models
- Observations from BUFR files (NCEP reanalysis flavor)
- Can create synthetic (i.e. perfect model) observations for any of these.

Roadmap to the rest of the talk

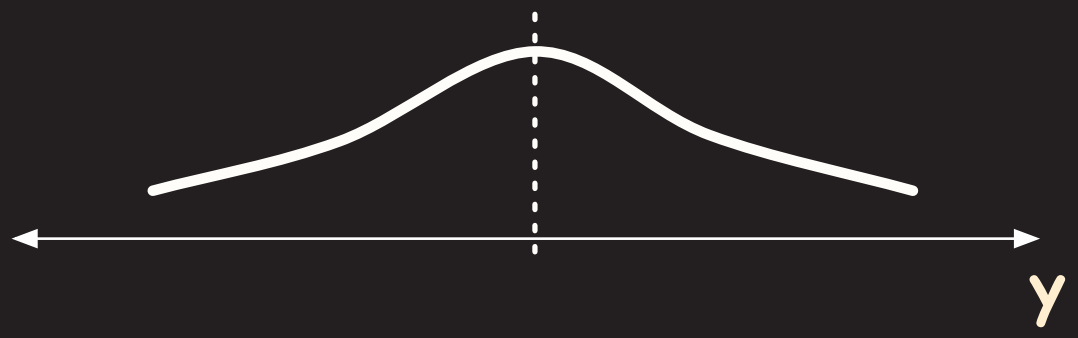
Ensemble Assimilation Outline	$\sim$ 1 minute
Identify the problem areas	$\sim$ 2 minutes
Describe Solution	$\sim$ 3 minutes
Describe Model & Obs	$\sim$ 1 minute
Show results	$\sim$ 5 minutes
Questions	$\sim$ closing

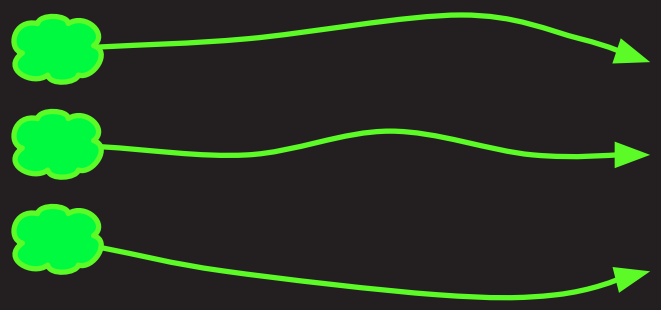
Ensemble Data Assimilation – Overview



An observation and its probability density function.

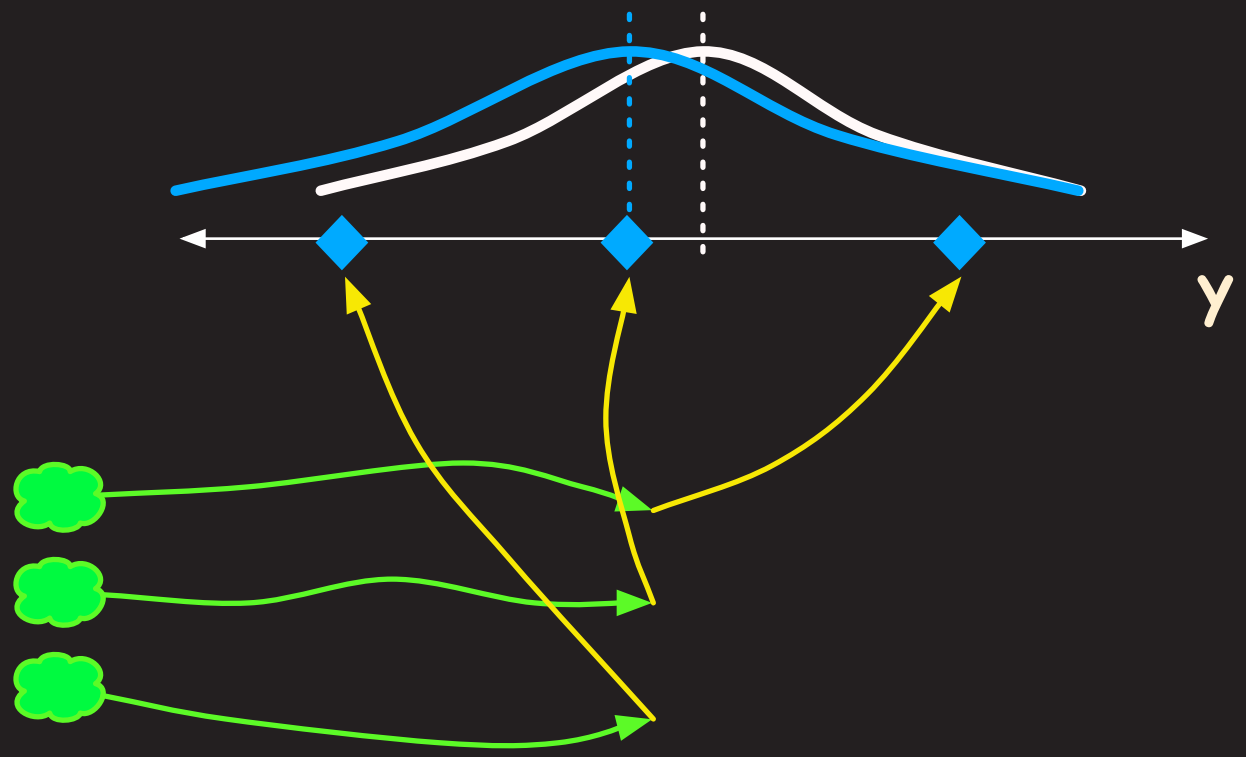
Ensemble Data Assimilation – Integrate





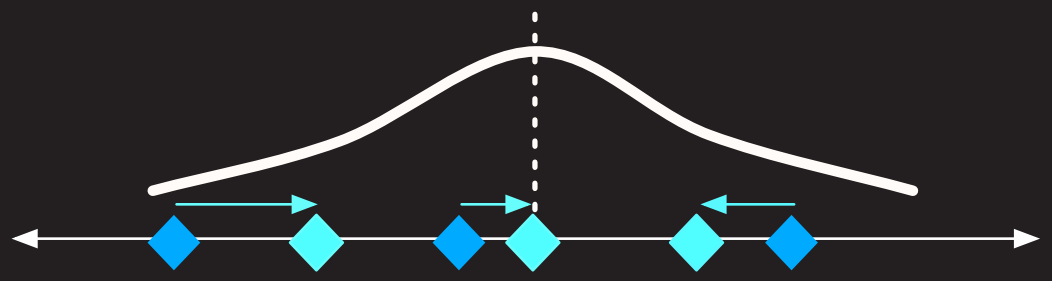
Integrate an ensemble of models to the time of the observation. This is the prior.

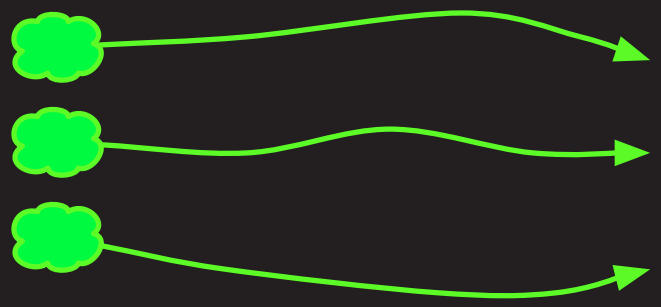
Ensemble Data Assimilation – $\hat{y} = \mathcal{H}(x)$



Apply a forward operator $\mathcal{H}$ to determine the *pdf* of the state variables in observation-space.

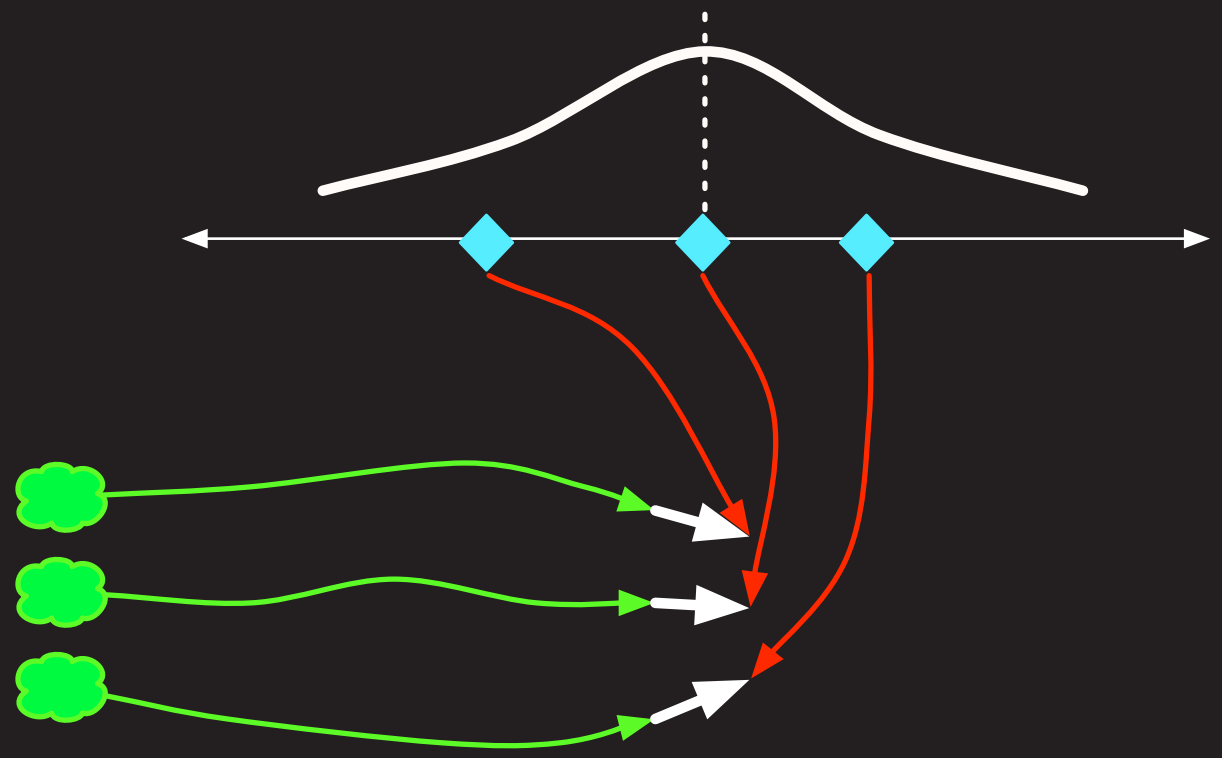
Ensemble Data Assimilation – Increments





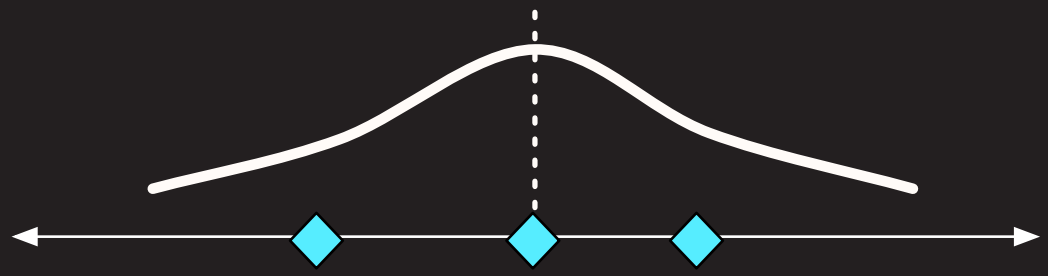
Determine and apply 'increments'.

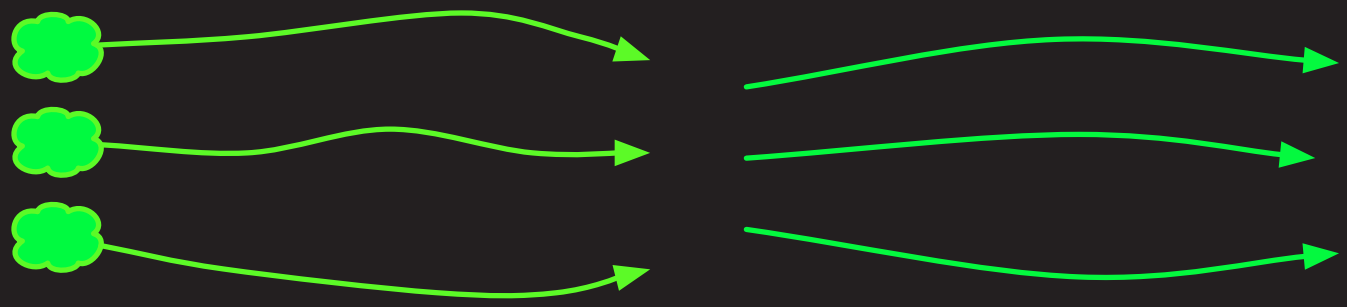
Ensemble Data Assimilation – Regress



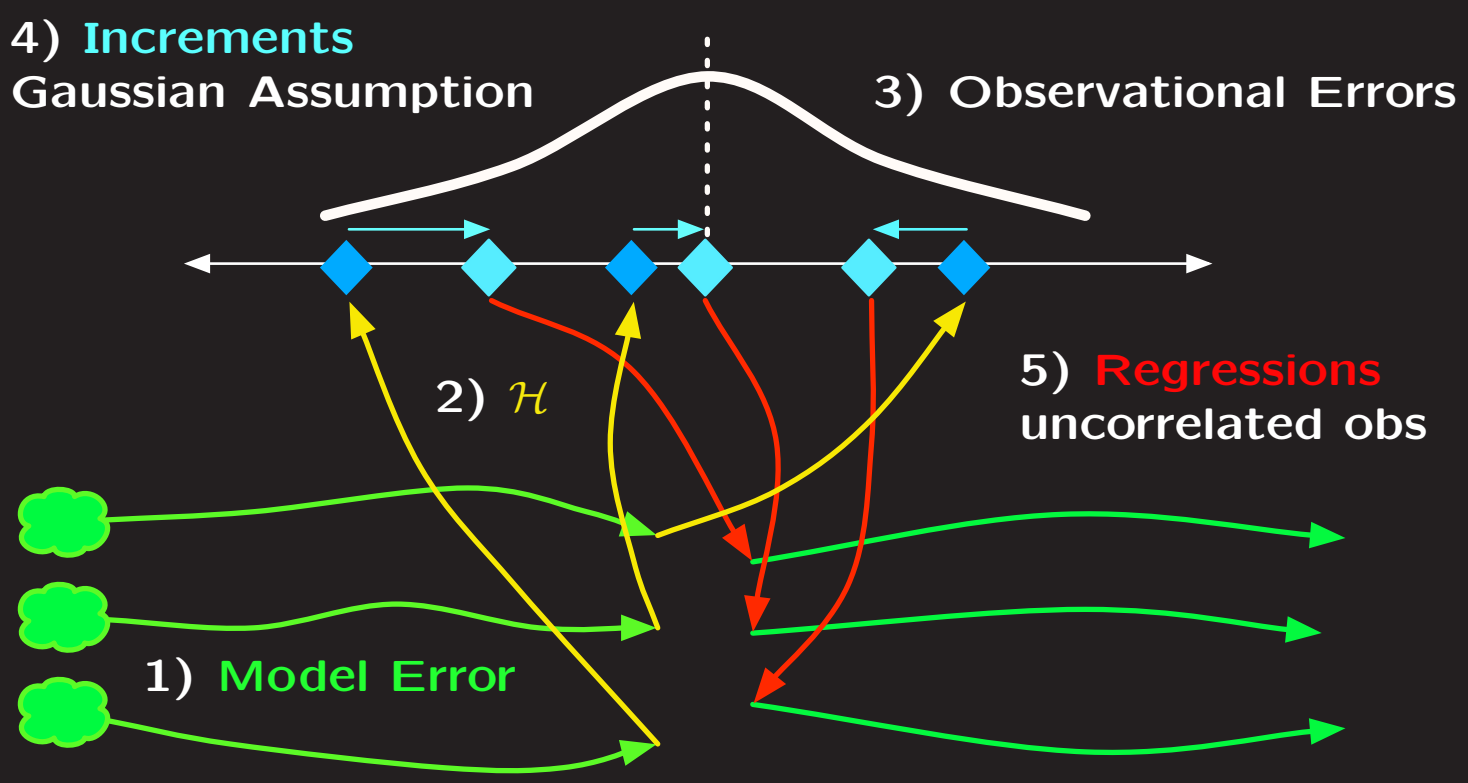
Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments. Can be done sequentially if observations are conditionally uncorrelated.

Ensemble Data Assimilation – Proceed





Error Sources and Treatments



- 1,2,3 → Independent, Hard, ongoing . . .
- 1-4 → Adaptive Observation-Space Algorithm
- 5 → Localization, Group Filter

Group Filter – augments *Localization*

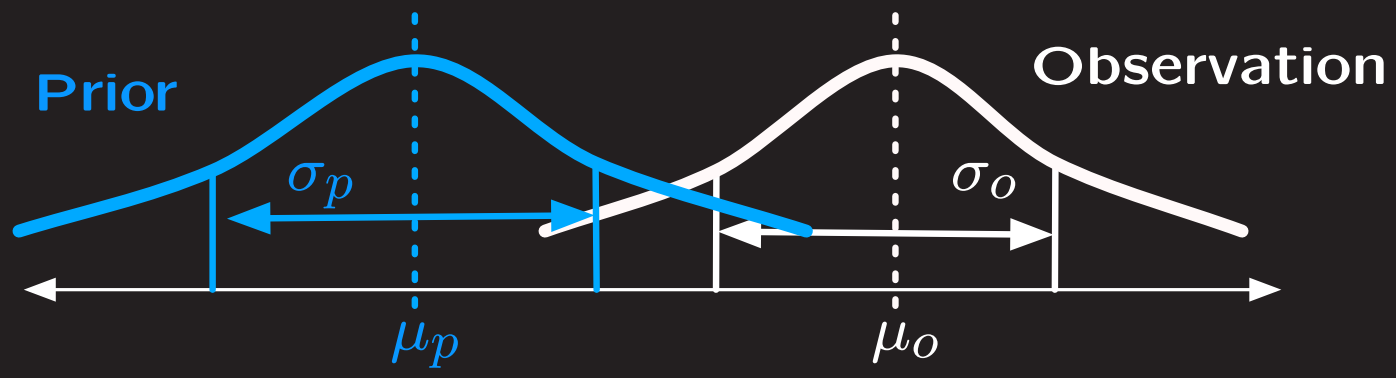
A Monte Carlo algorithm to address regression sampling errors.

- **Create M groups of N ensembles.**
- **Compute the observation increments as usual.**
- **For a given observation/state variable pair:**

  1. **M estimates of the regression coefficient r**
  2. **Uncertainty in r implies state variable increments $\Downarrow$**
  3. **Can compute a** *regression confidence factor* $\rightarrow \alpha$

http://www.cgd.ucar.edu/DAI/publications.html

Adaptive Observation-Space Treatments



Construct a ratio $R = \frac{|\mu_p - \mu_o|}{\sqrt{\sigma_p^2 + \sigma_o^2}}$

Best Possible Scenario: $R \equiv 1 \rightarrow$ no covariance inflation needed.

Reality: As R differs from 1, adaptively inflate the variance of both the Prior and the Observation.

Why assimilate with a climate model?

- Systematic error: Can one detect long-term biases quickly?
- Model quality evaluation: Being a good forecast model increases climate model credibility.
- Parameter estimation: Can model be 'tuned' with assimilation?
- Observation system evaluation experiments: determine value of different observations, OSSEs

Experiment Description

CAM 2.0.1 T42L26

- U, V, T, Q, and PS impacted by observations
- Land Model (CLM 2.0) not impacted by observations
- Climatological SSTs

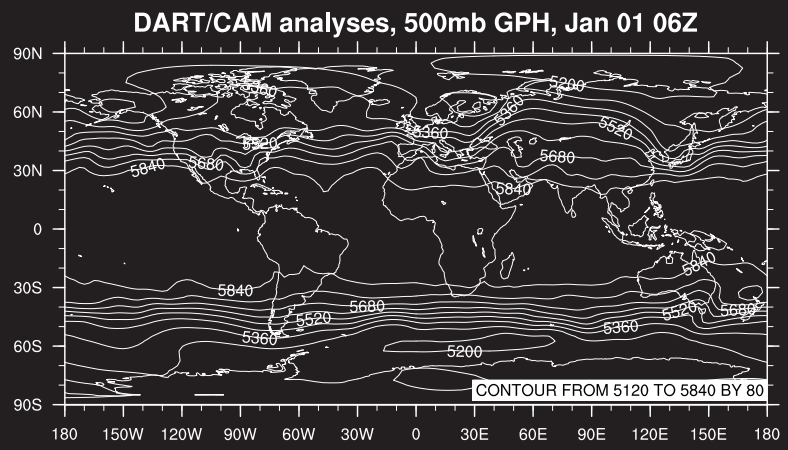
Assimilation/Prediction Experiment

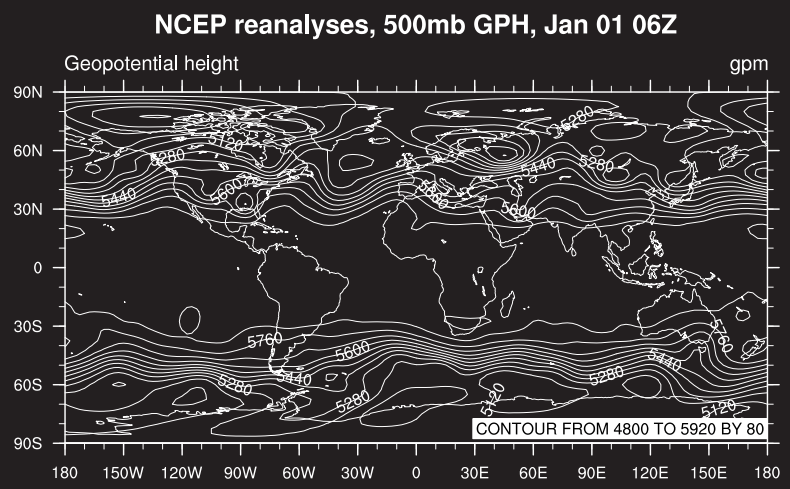
- Initialized from a climatological distribution (huge spread)
- 80-member ensemble divided into 4 equal groups
- Run on a 'university-sized' computer – 16 compute nodes.
- Assimilation cycling every 6 hours; $\pm$ 1.5 hour window for obs.
- Uses most observations used in NCEP reanalysis (ACARS, Radiosondes (no Q), Satellite Winds ..., no surface obs.)

NCEP Reanalysis – used for comparison only

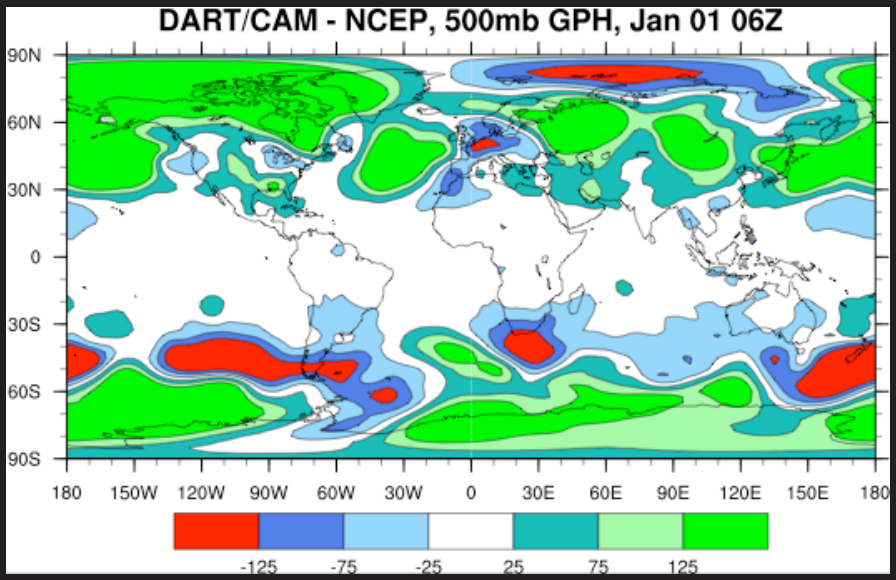
- T62L28
- uses radiances (DART does not)

06Z 1 Jan 2003 → after first cycle.

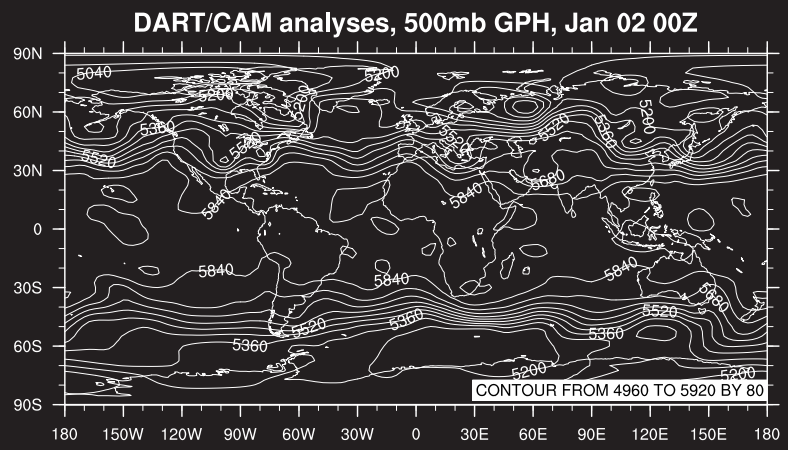


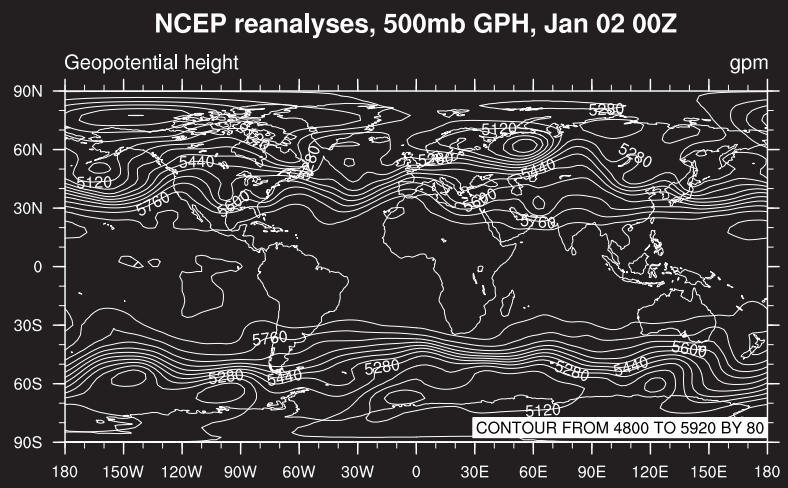


Ensemble Mean; large differences $\approx \pm$ 125 m

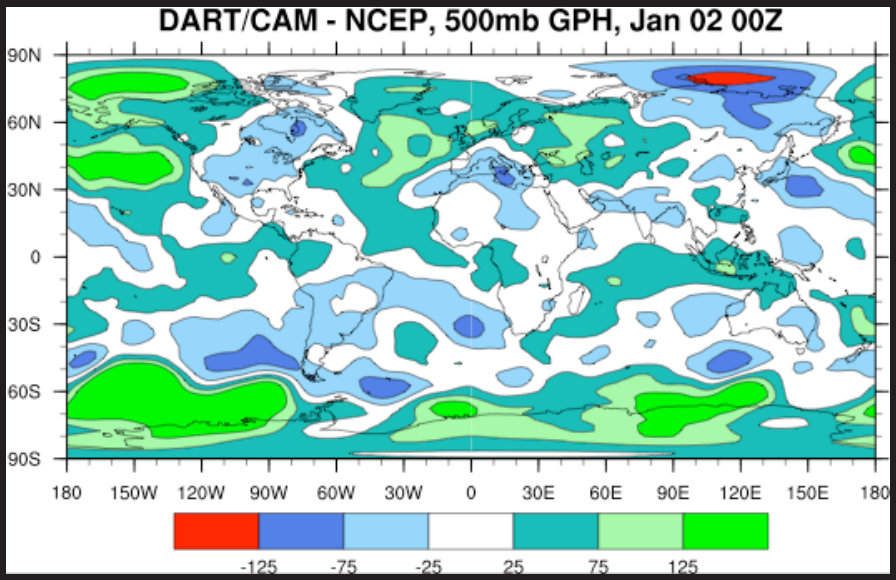


00Z 2 Jan 2003 → 4 Assimilation cycles.

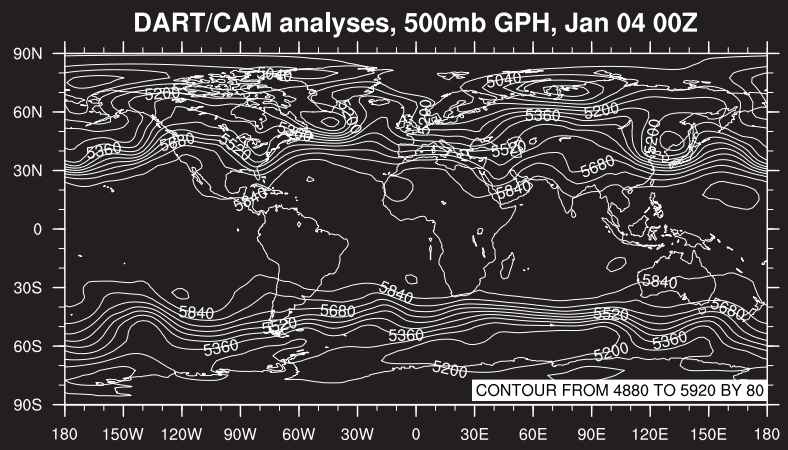


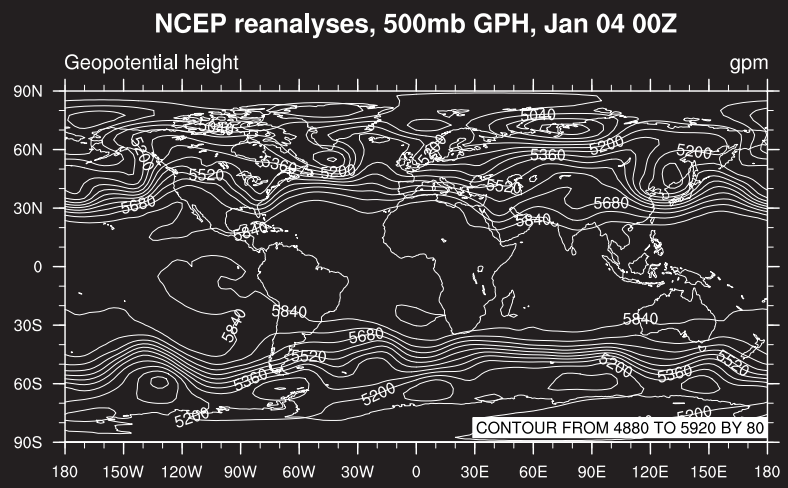


Adapting . . .

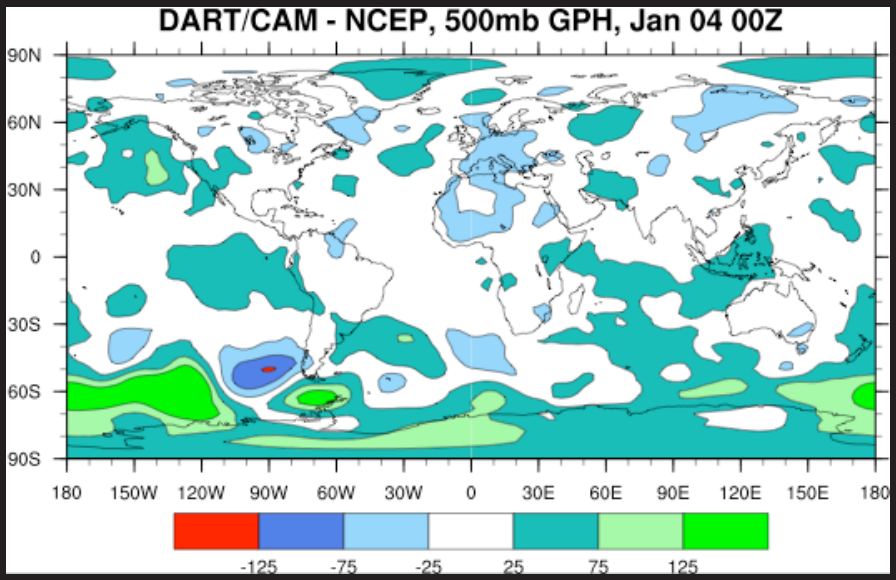


00Z 4 Jan 2003 . . .

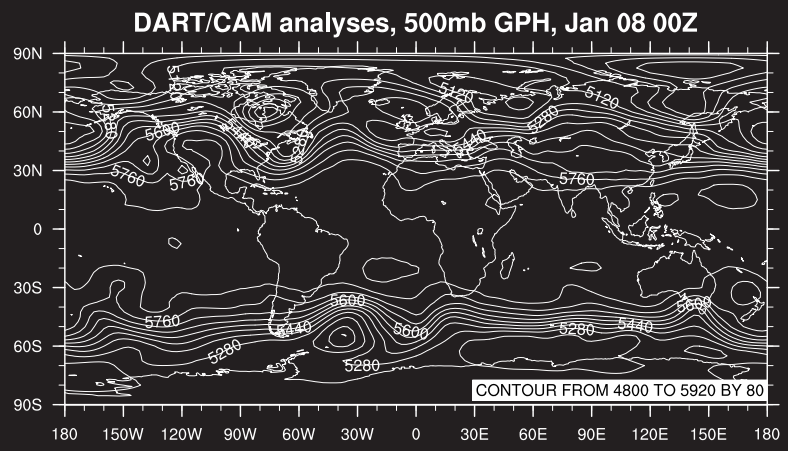


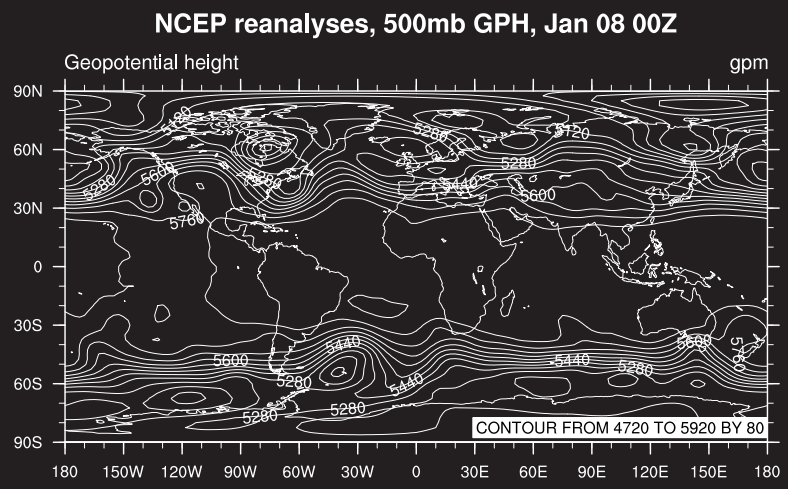


Still Improving . . .



00Z 8 Jan 2003 . . .

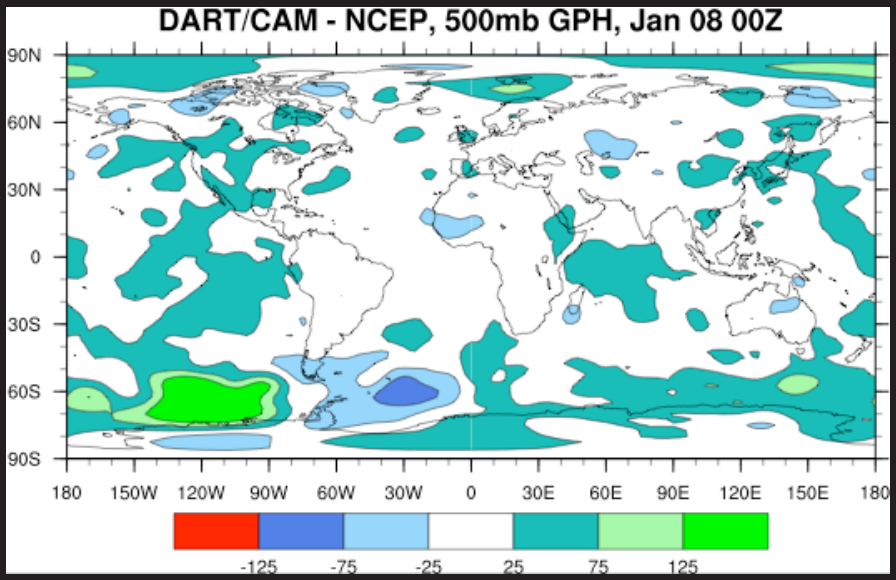




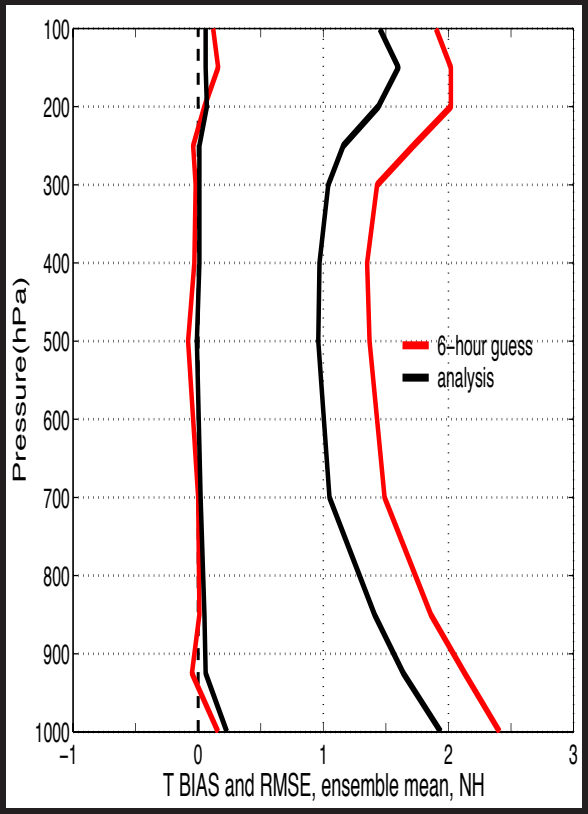
Small differences.

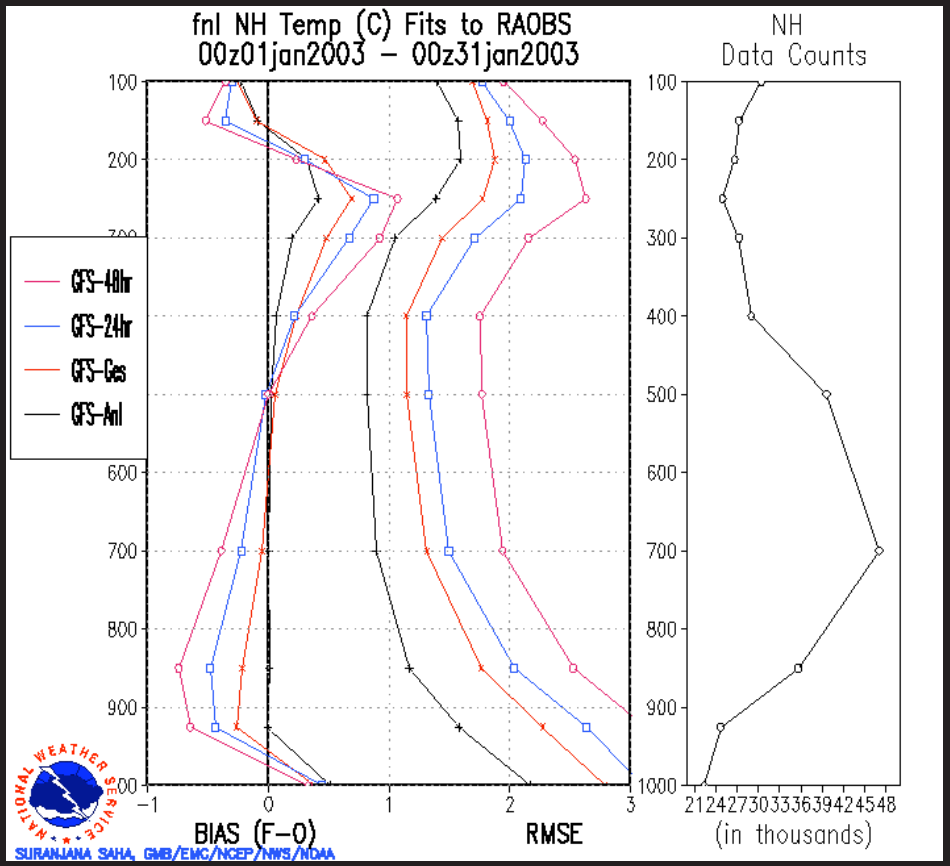
Proof we are assimilating.

Excellent amount of structure from a T42 model!



Monthly Mean NH Temperature Jan 2003

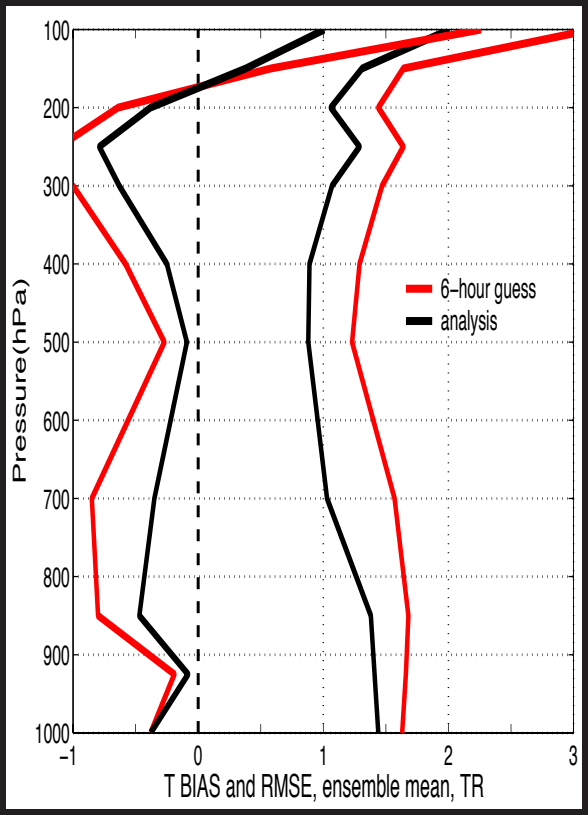


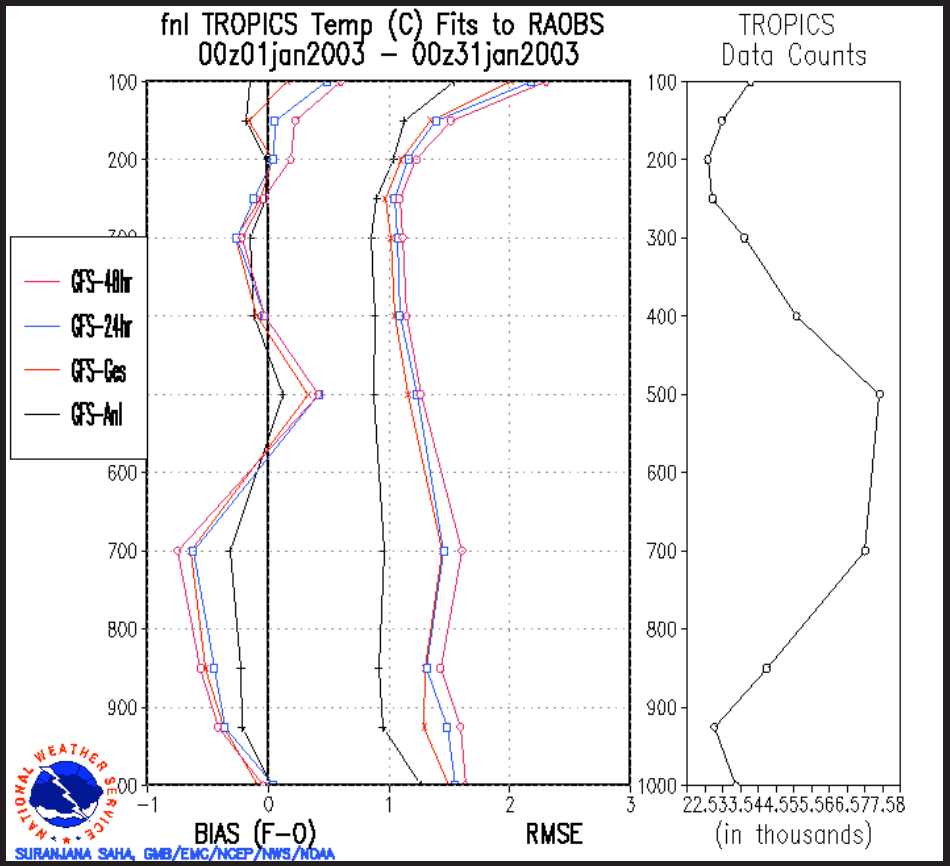


CAM/DART
6-hour guess is pretty comparable

NCEP GFS
Operational Forecast System
T15x

Monthly Mean TR Temperature Jan 2003

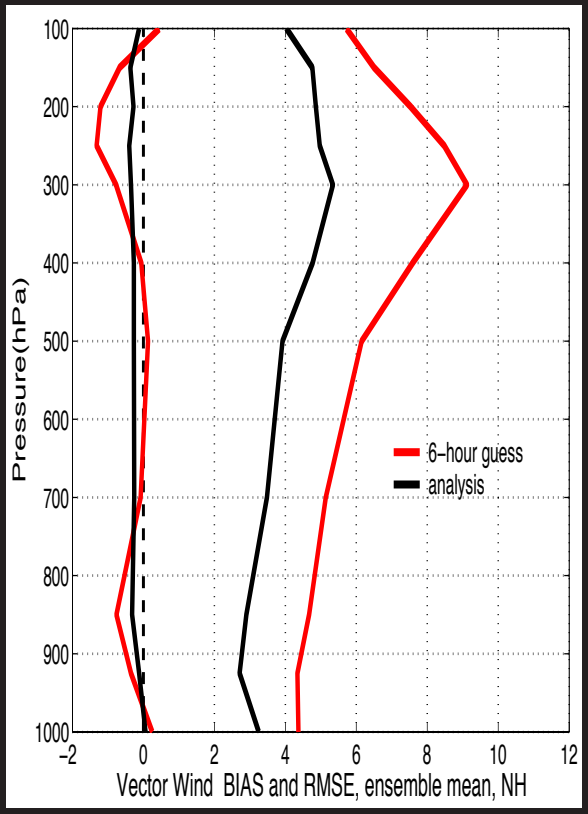


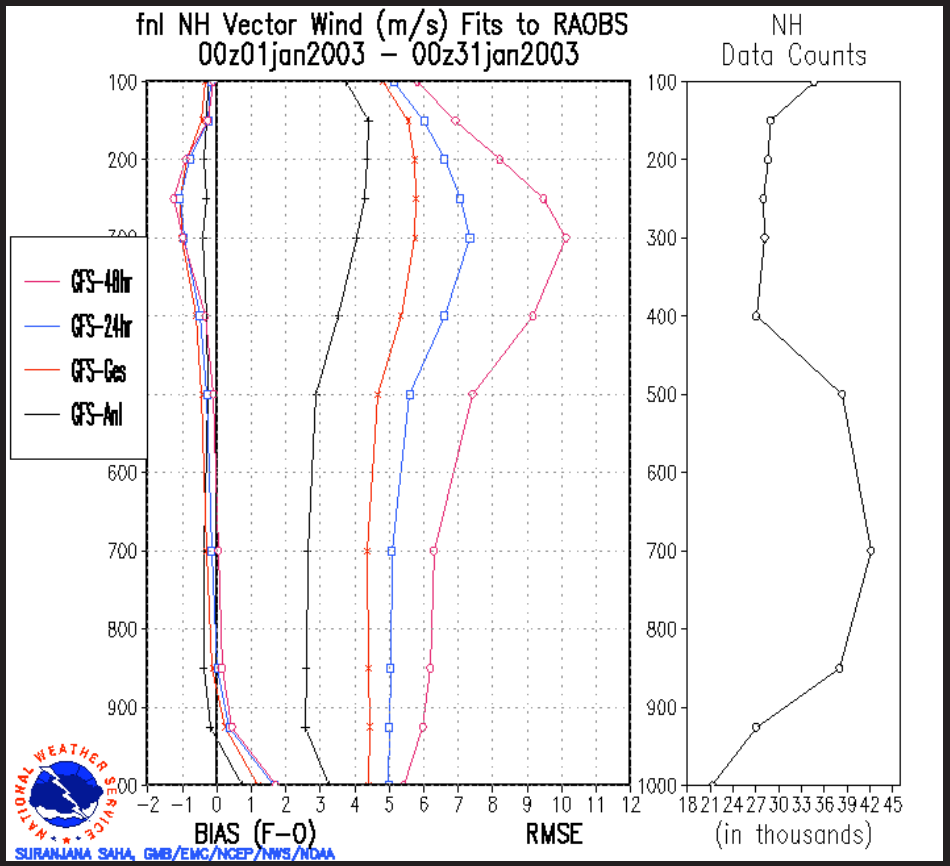


CAM/DART
bias large
trouble in upper tropo

NCEP GFS

Monthly Mean NH Wind Jan 2003



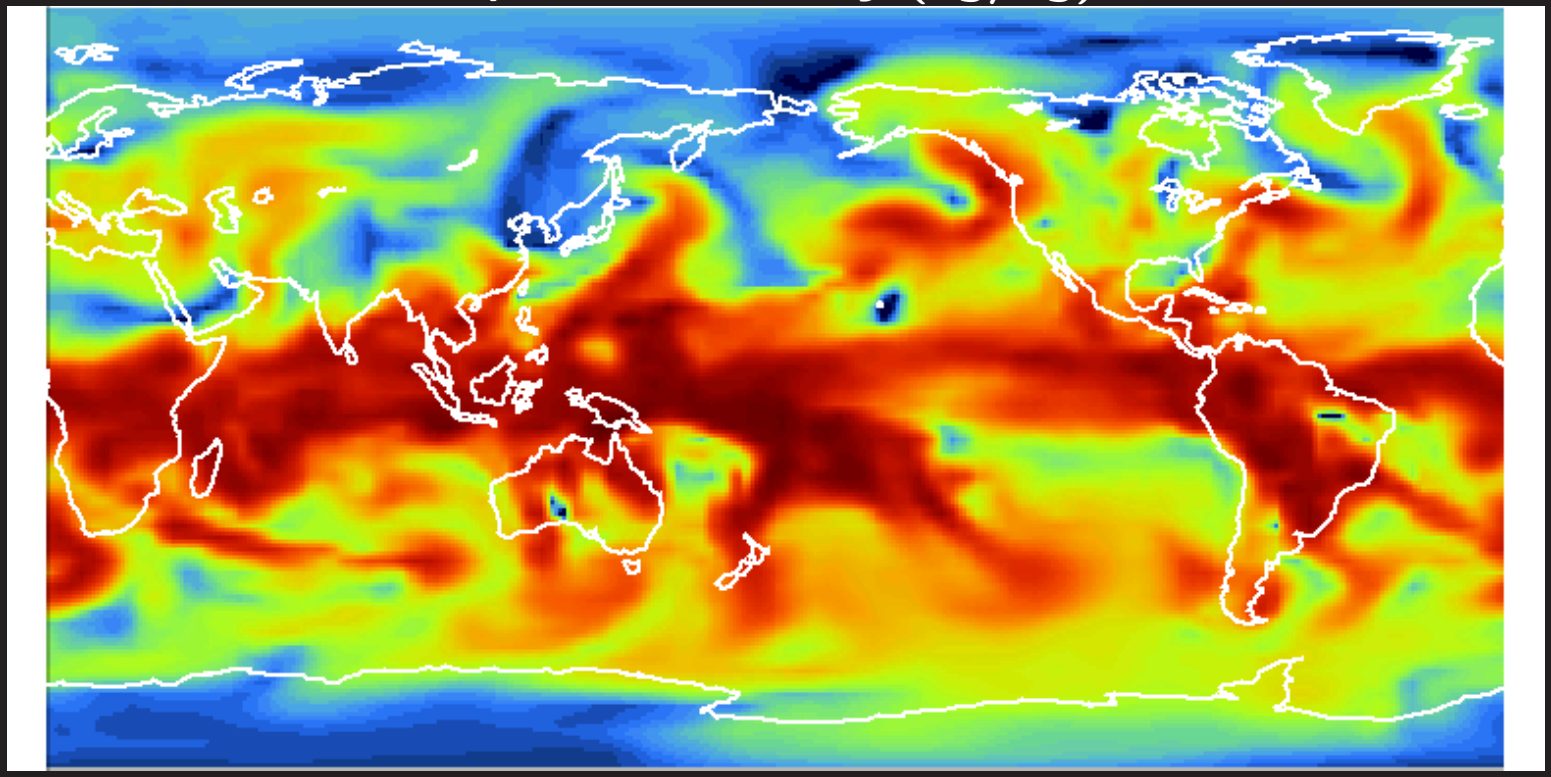


CAM/DART
worse in upper tropo
but similar shapes

NCEP GFS

12 GMT 04 Jan 2003 CAM Analysis

Specific Humidity (kg/kg)

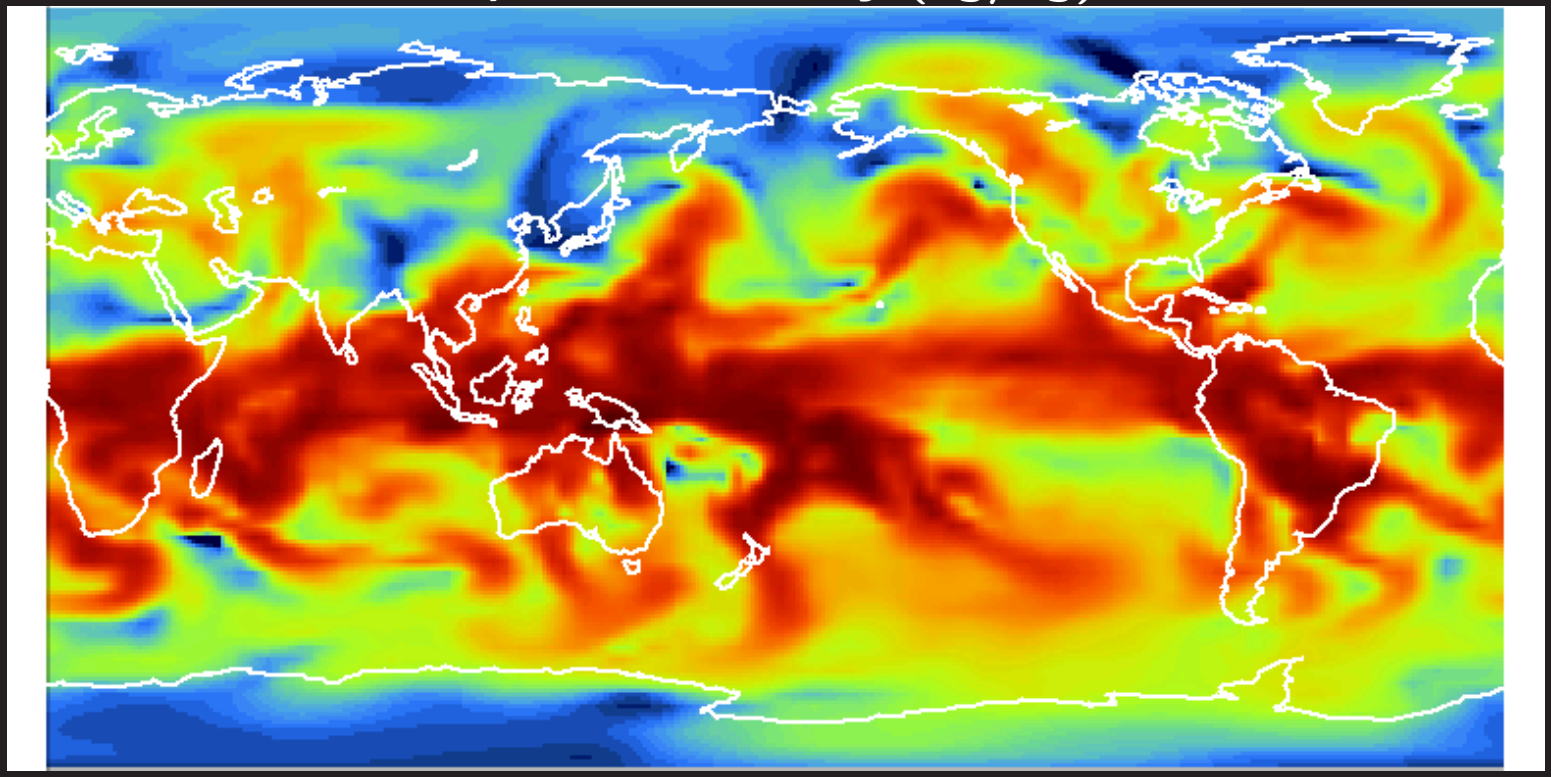


(WET, DRY)

Observations are impacting all state variables – including Q.
Level 21 (of 26) $\approx$ 850 hPa, WET $\simeq$ 0.01

00 GMT 05 Jan 2003 CAM Analysis

Specific Humidity (kg/kg)

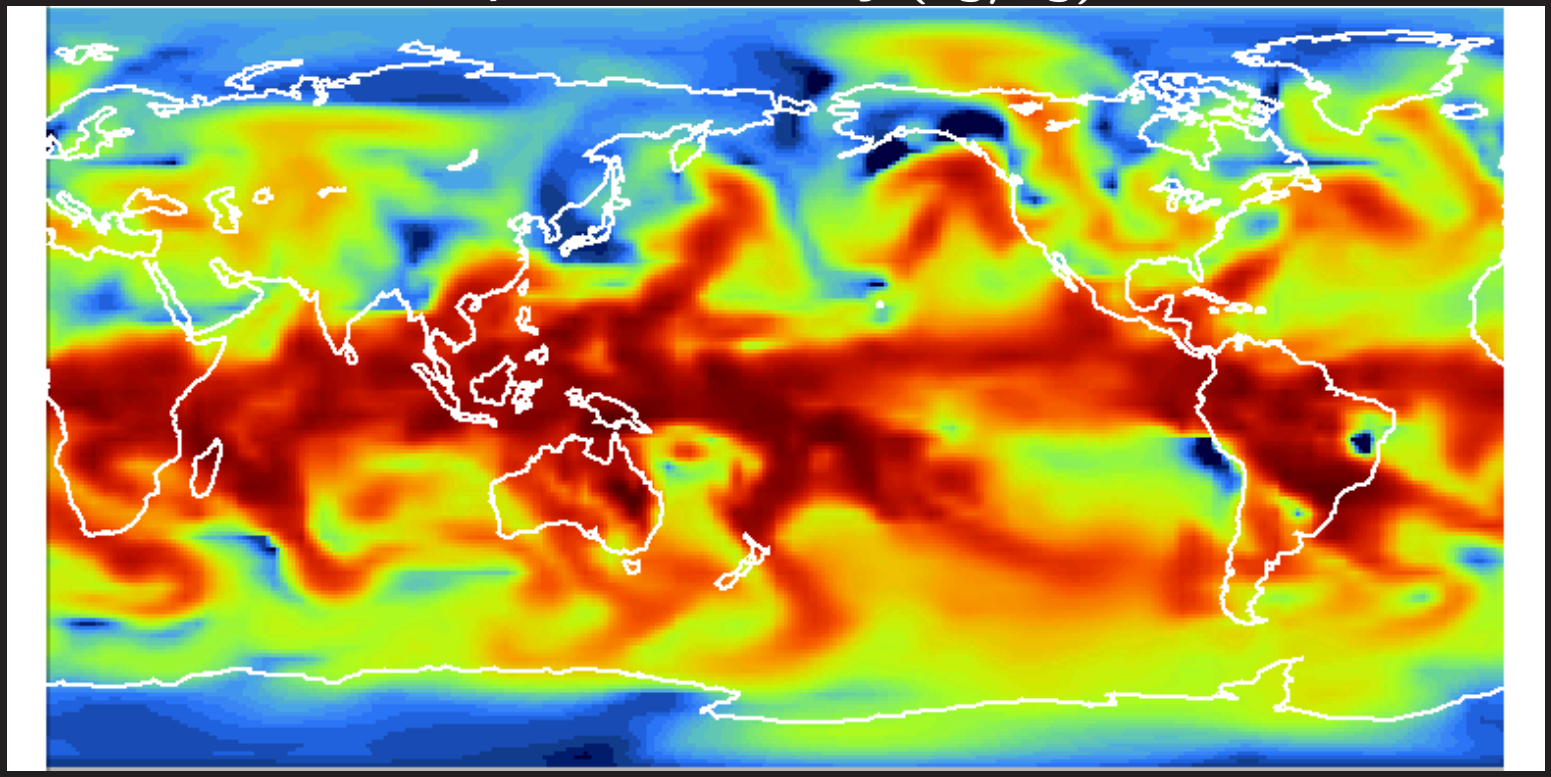


(WET, DRY)

T42 gets realistic structure consistent with obs, much more so than climate model

12 GMT 05 Jan 2003 CAM Analysis

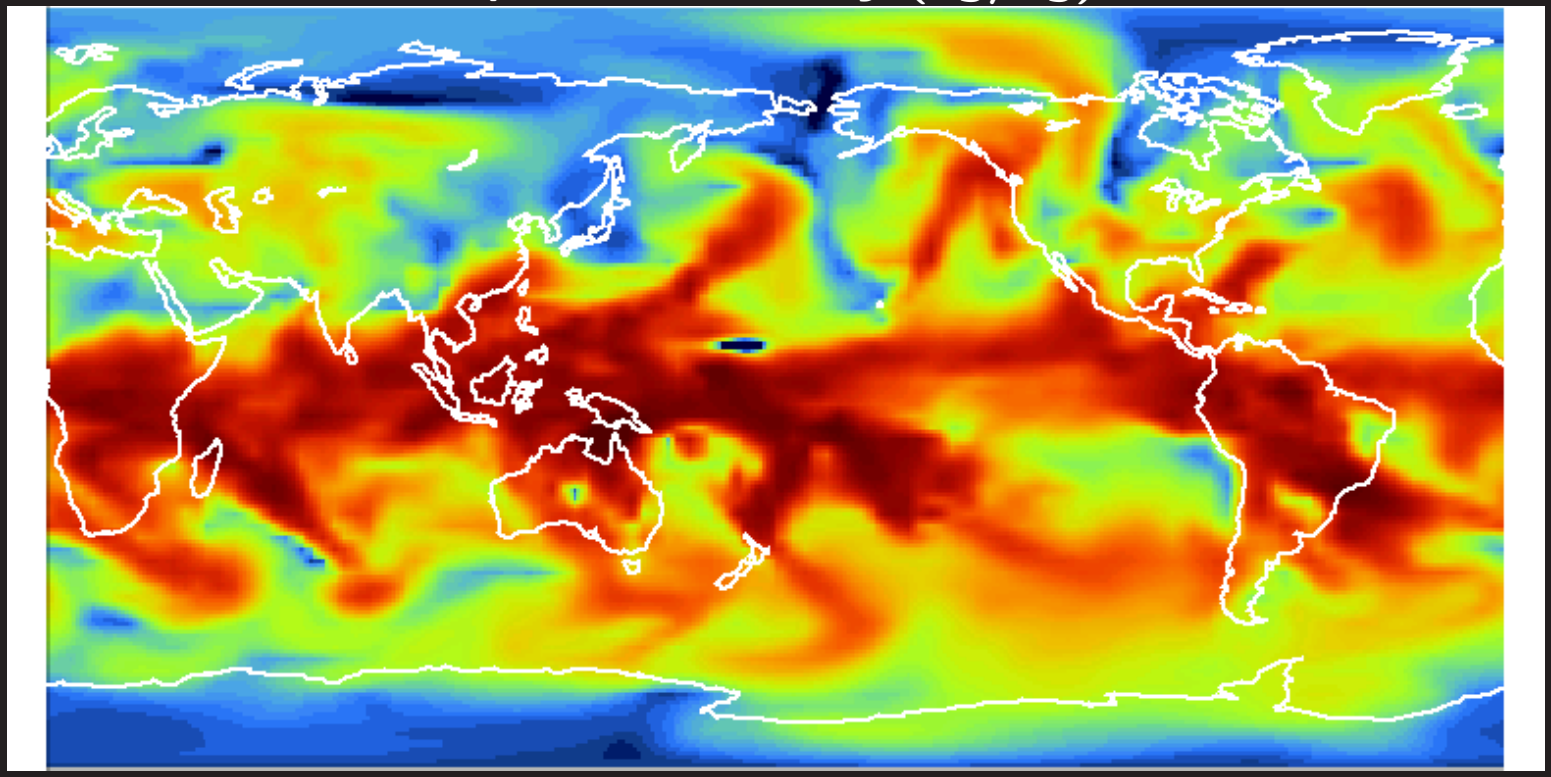
Specific Humidity (kg/kg)



(WET, DRY)

00 GMT 06 Jan 2003 CAM Analysis

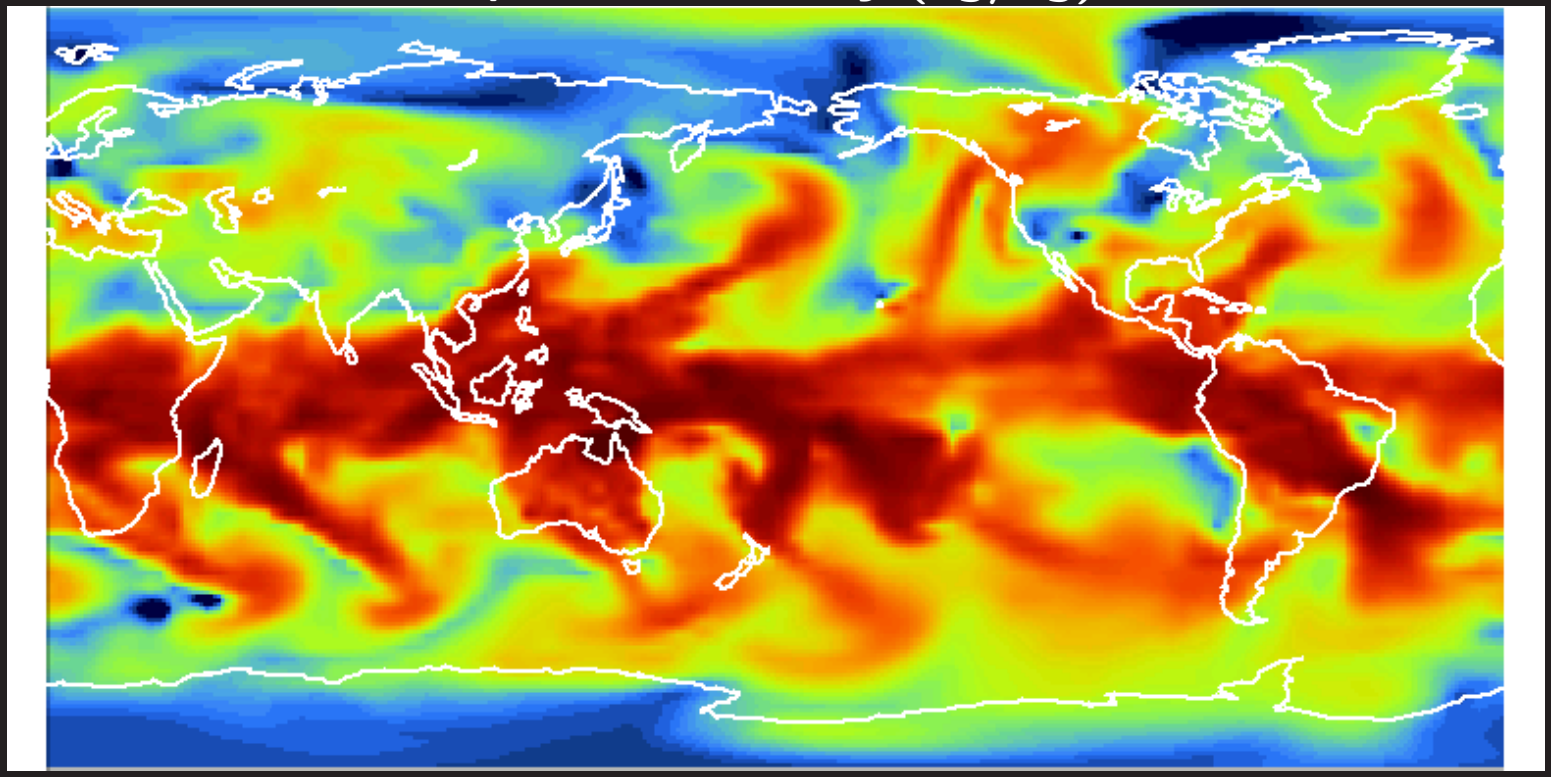
Specific Humidity (kg/kg)



(WET, DRY)

12 GMT 06 Jan 2003 CAM Analysis

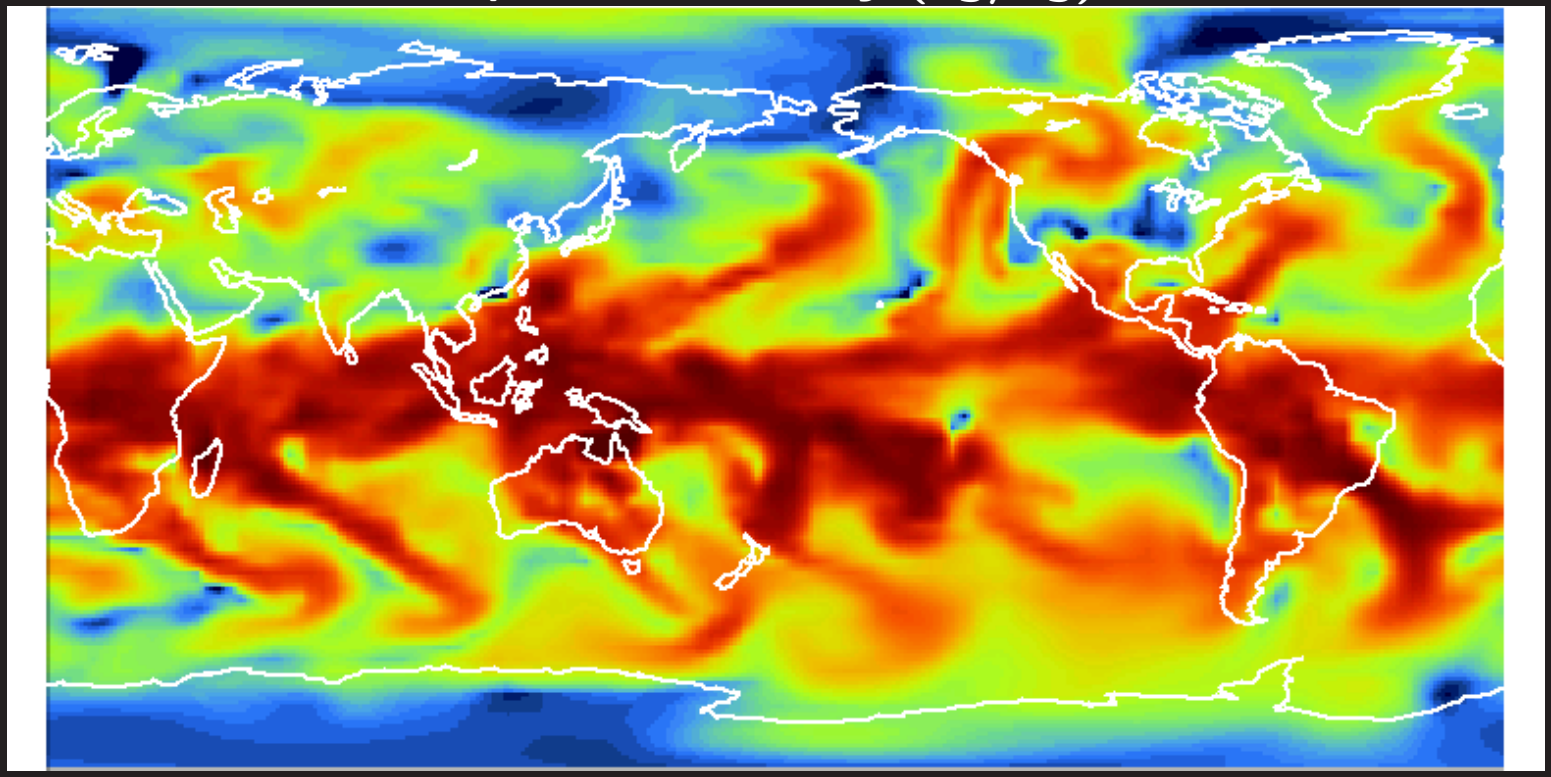
Specific Humidity (kg/kg)



(WET, DRY)

00 GMT 07 Jan 2003 CAM Analysis

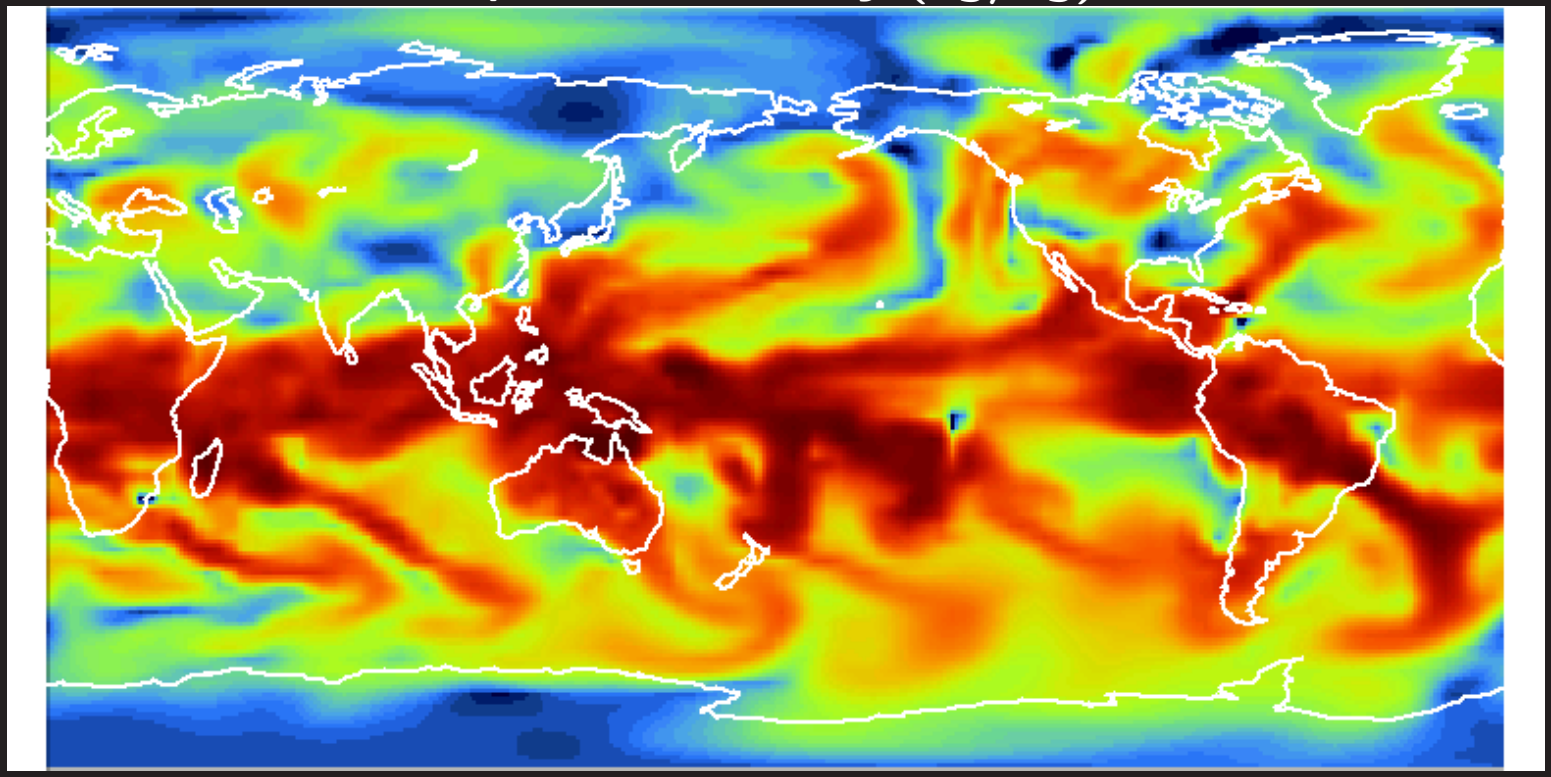
Specific Humidity (kg/kg)



(WET, DRY)

12 GMT 07 Jan 2003 CAM Analysis

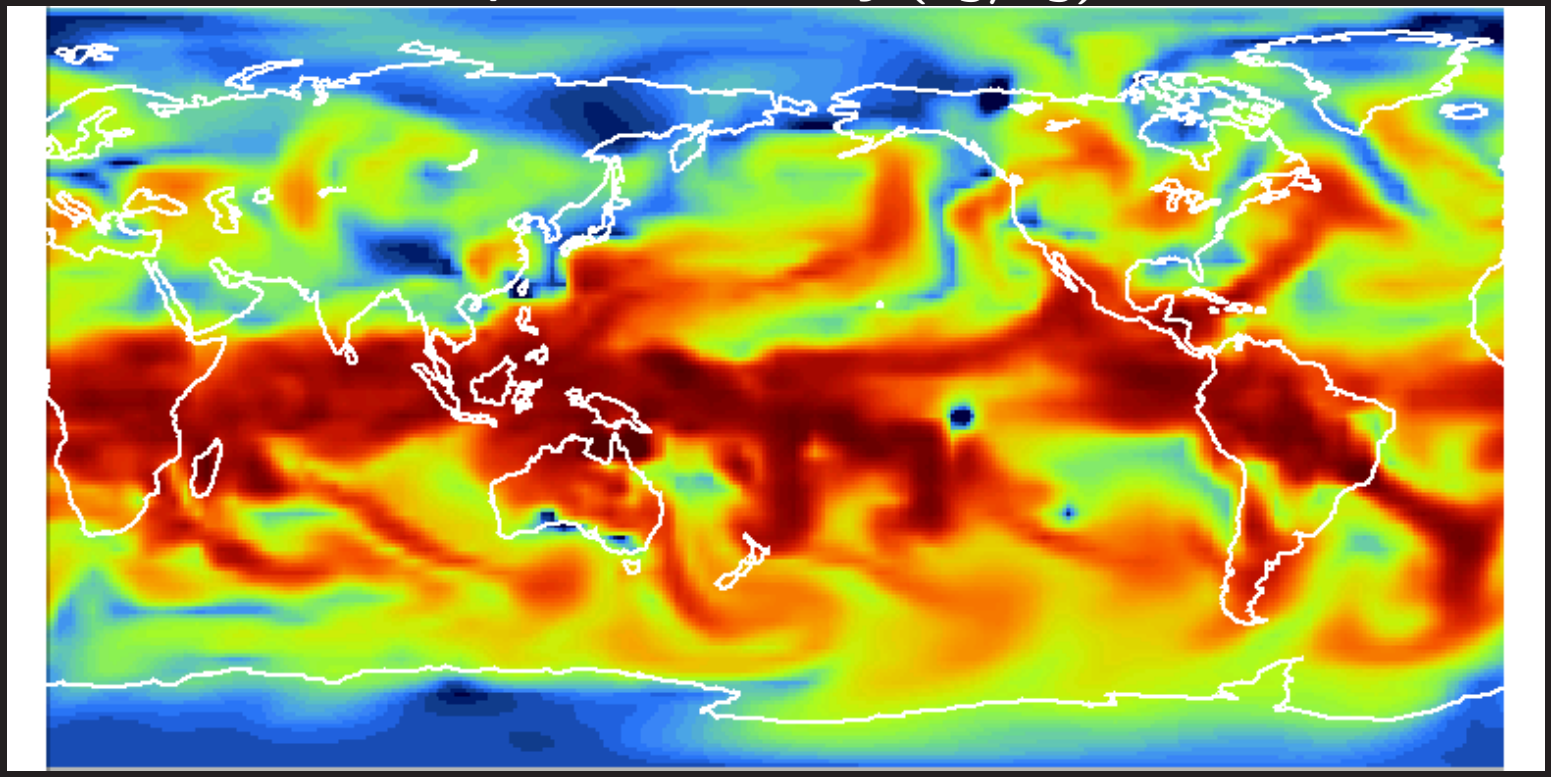
Specific Humidity (kg/kg)



(WET, DRY)

00 GMT 08 Jan 2003 CAM Analysis

Specific Humidity (kg/kg)



(WET, DRY)

Cyclogenesis off Newfoundland; dry air in front, dry air in occlusion.

Conclusions

- Assimilation of real observations in a climate model is a reality.
- The observations can impact all the state variables if desired.
- The diagnostics can focus research efforts in model development.
- A static covariance inflation is not needed to address filter divergence.
- Augmenting localization with a small 'ensemble of ensembles' approach proves useful.

My thanks go to: Jeff Anderson, Kevin Raeder, Hui Liu, Alain Caya

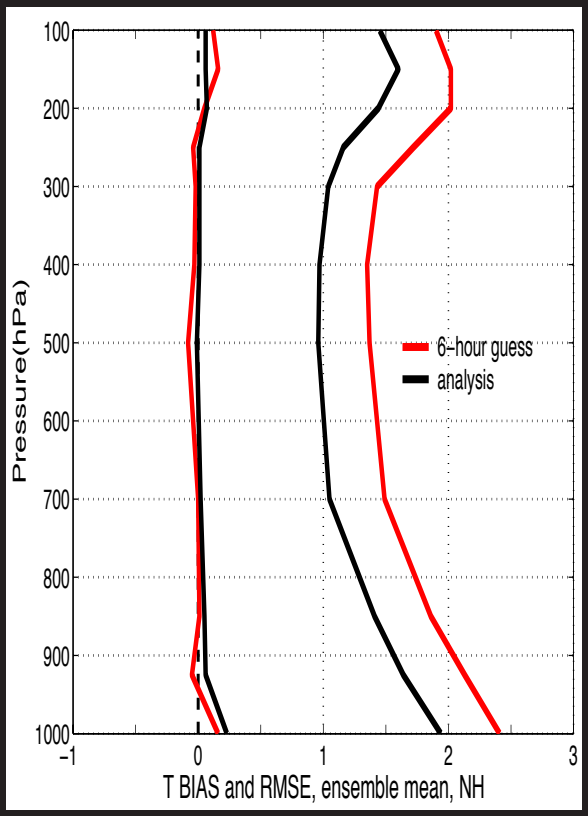
http://www.cgd.ucar.edu/DAI

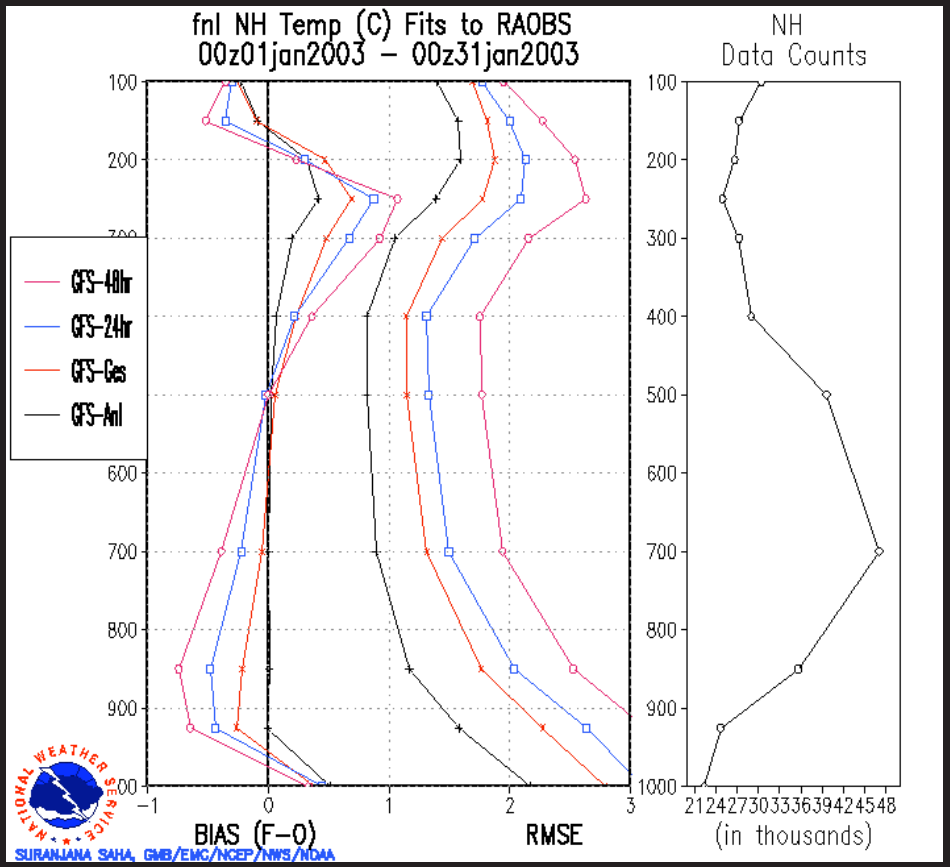
The Geophysical Statistics Project is Supported by the U.S. National Science Foundation DMS

Community Atmosphere Model → CAM

- Atmospheric component of NCAR's Community Climate System Model
- Designed for climate scenario integrations (IPCC) and research
- Open source; estimated to be $\approx$ 1000+ users
- This experiment uses spectral Eulerian T42 resolution. (Semi-lagrangian and finite-volume cores, T5 and T85 also available.)
- Horizontal localization: 0.2 radians half-width
- Multiplicative vertical localization: 200hPa half-width
- Observational errors are directly from BUFR files

Monthly Mean NH Temperature Jan 2003

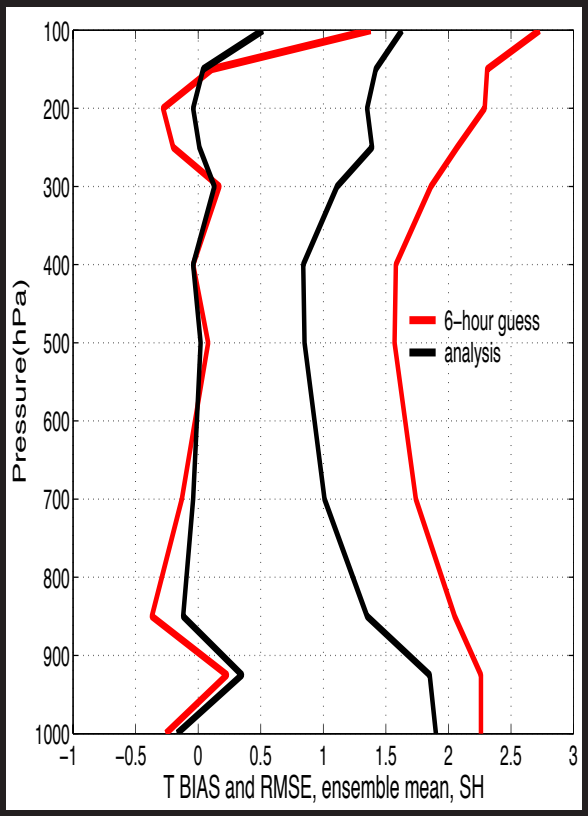


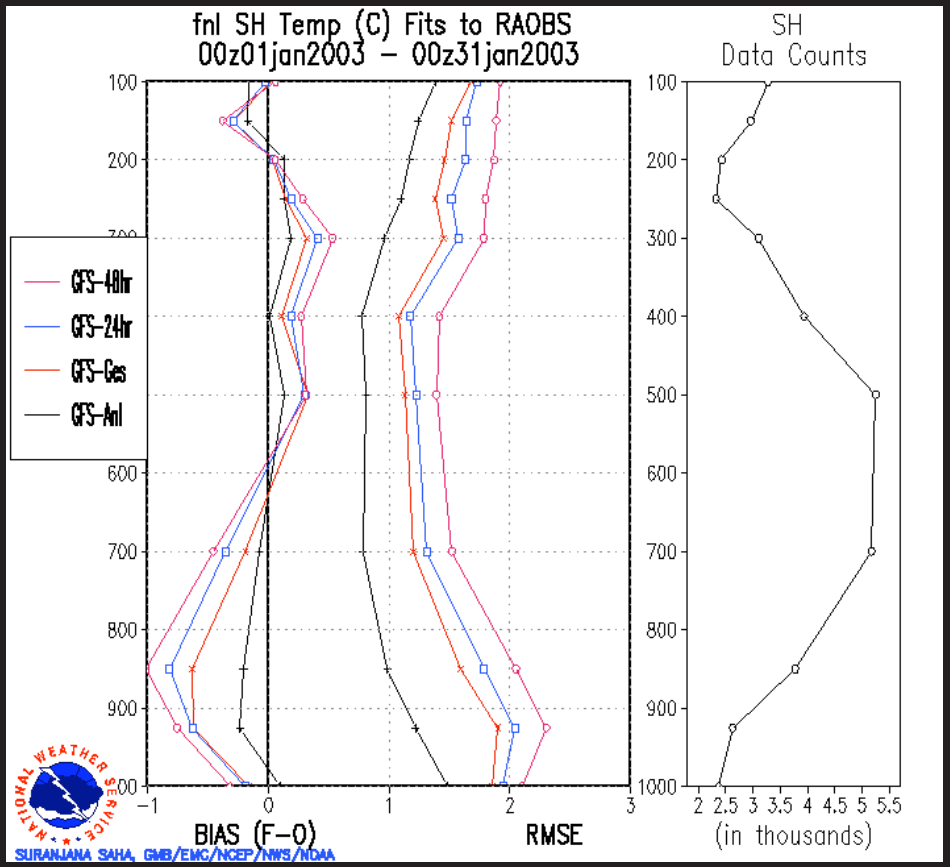


CAM/DART

NCEP GFS

Monthly Mean SH Temperature Jan 2003

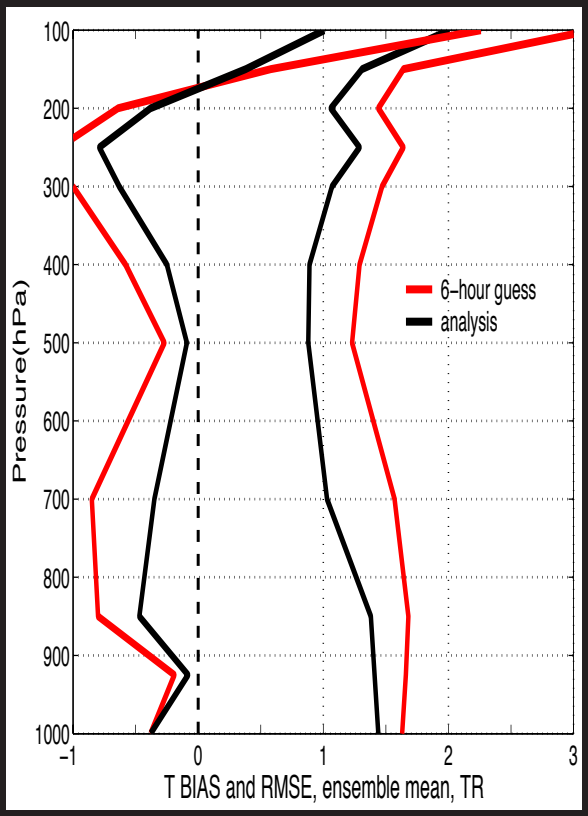


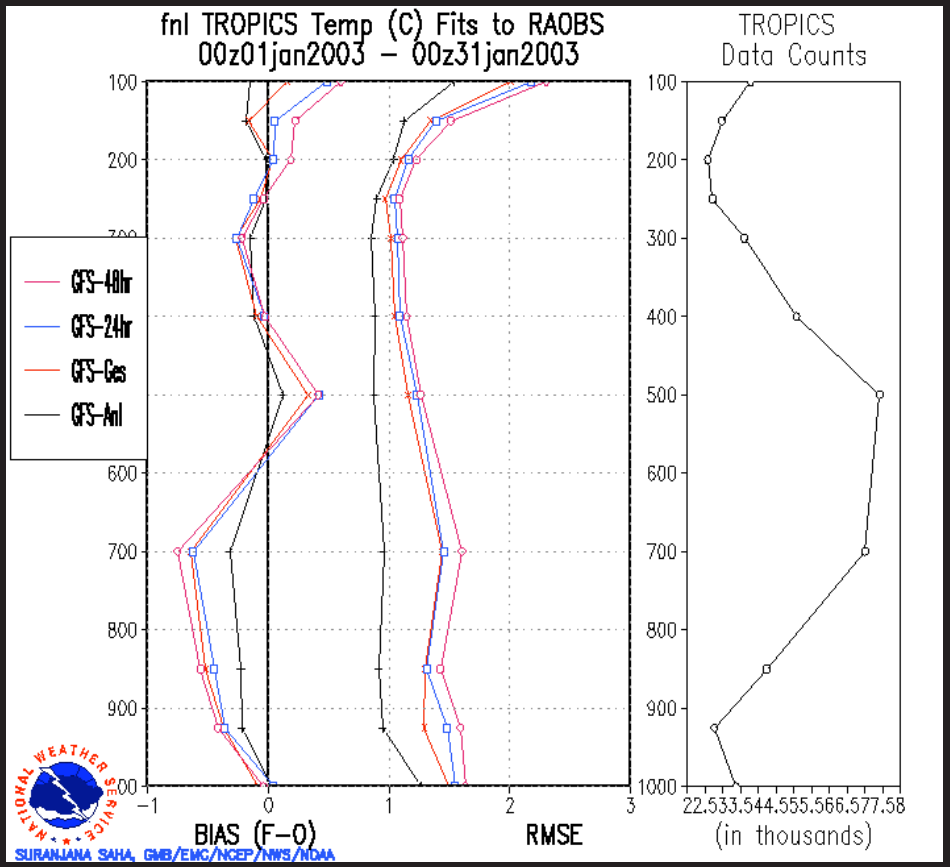


CAM/DART

NCEP GFS

Monthly Mean TR Temperature Jan 2003

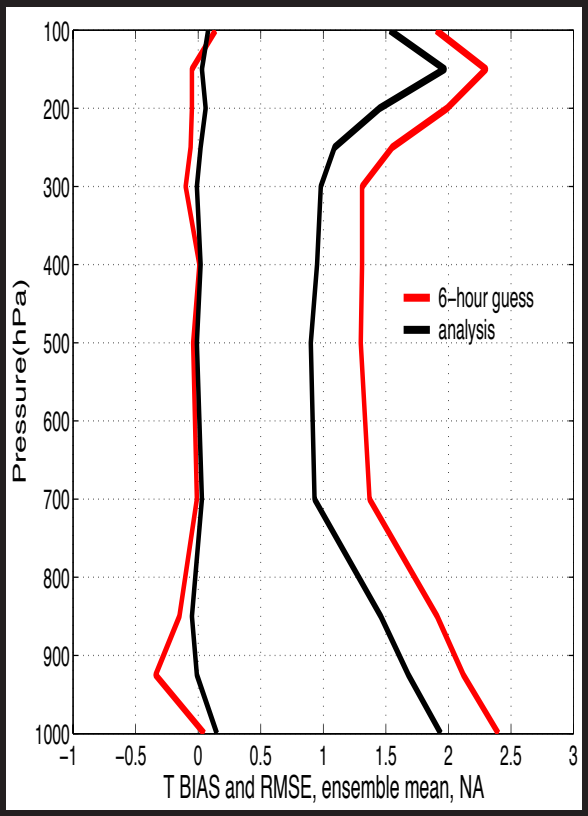


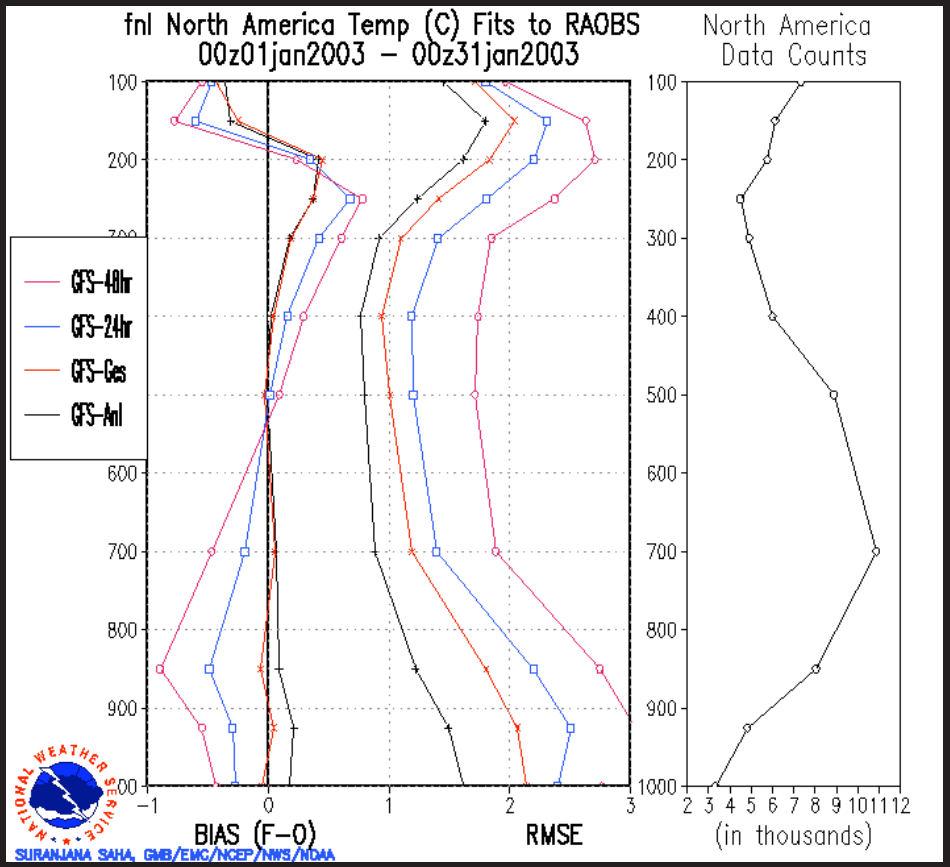


CAM/DART

NCEP GFS

Monthly Mean NA Temperature Jan 2003

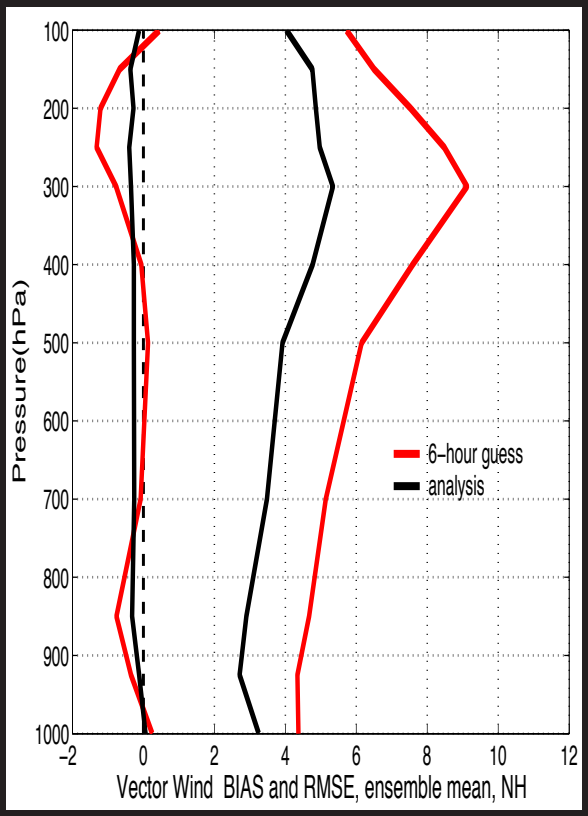


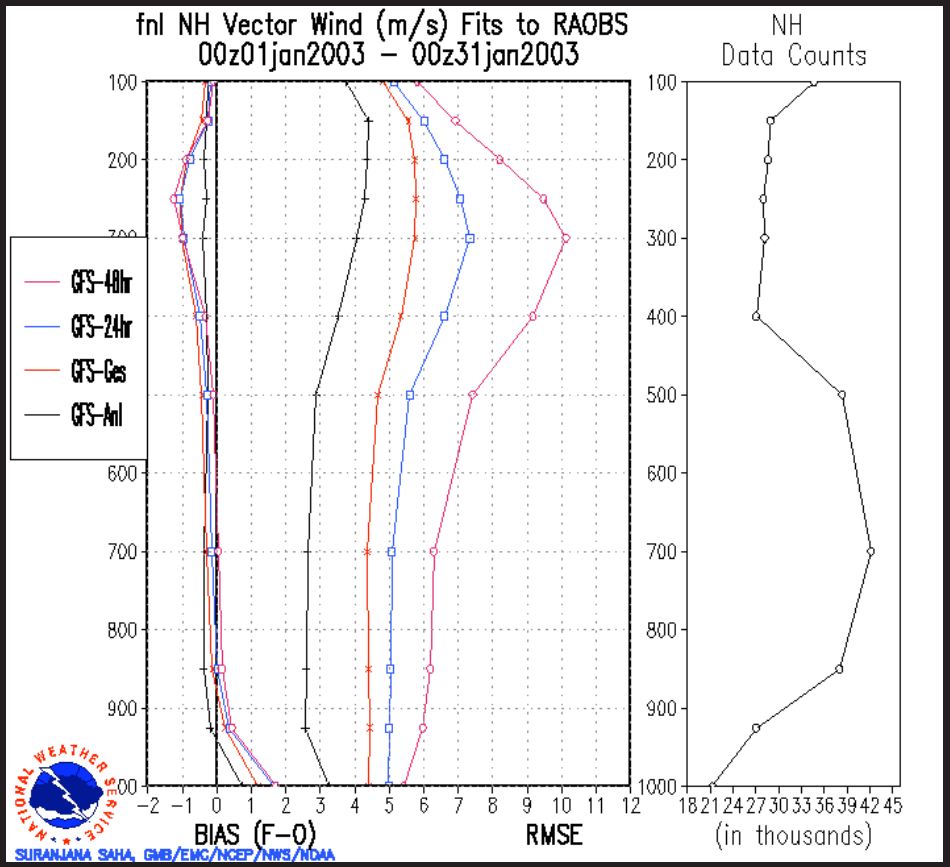


CAM/DART

NCEP GFS

Monthly Mean NH Wind Jan 2003

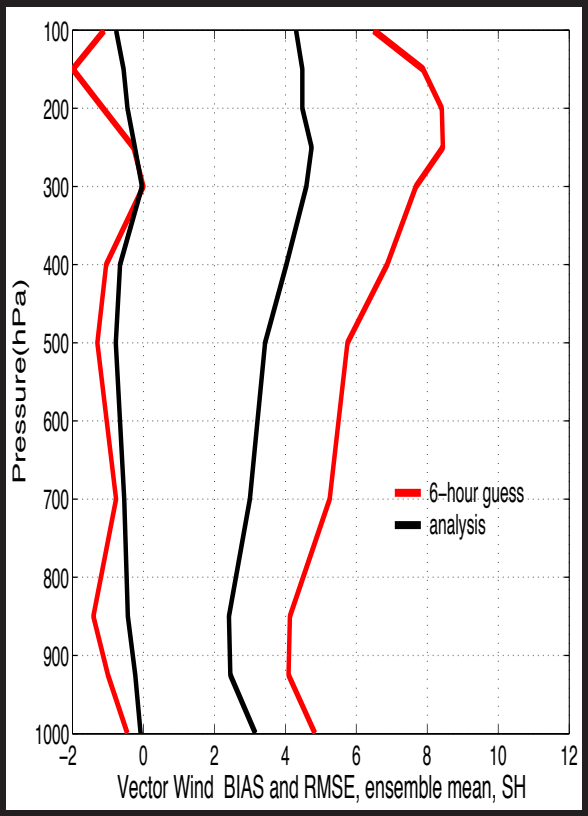


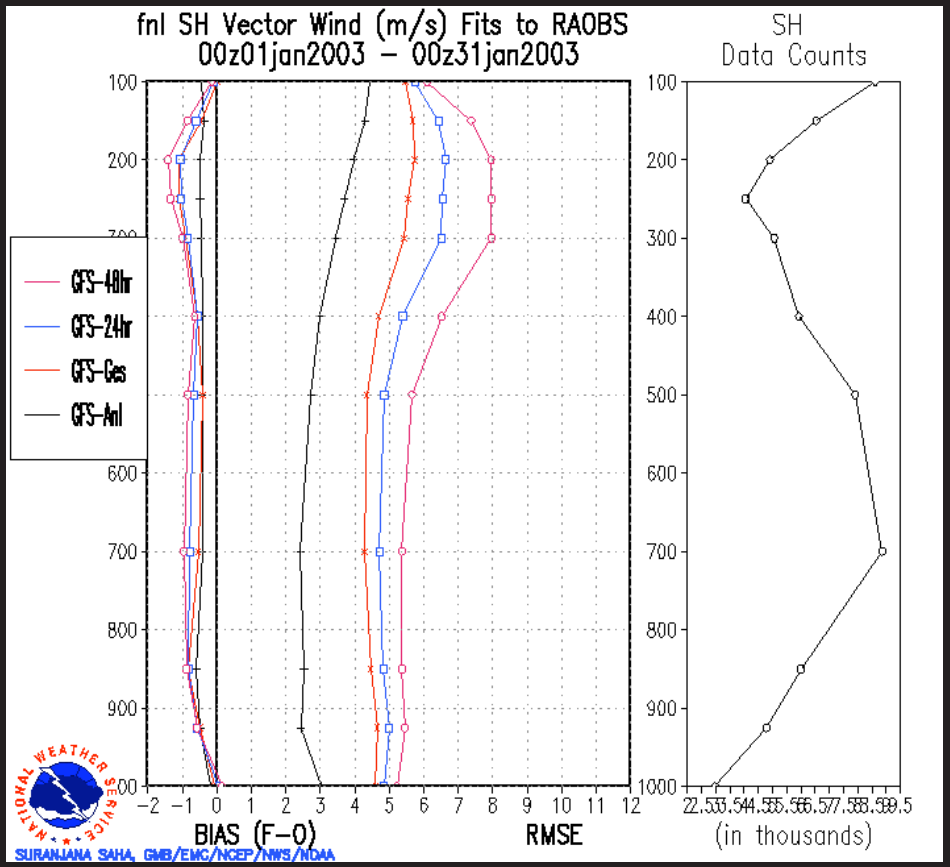


CAM/DART

NCEP GFS

Monthly Mean SH Wind Jan 2003

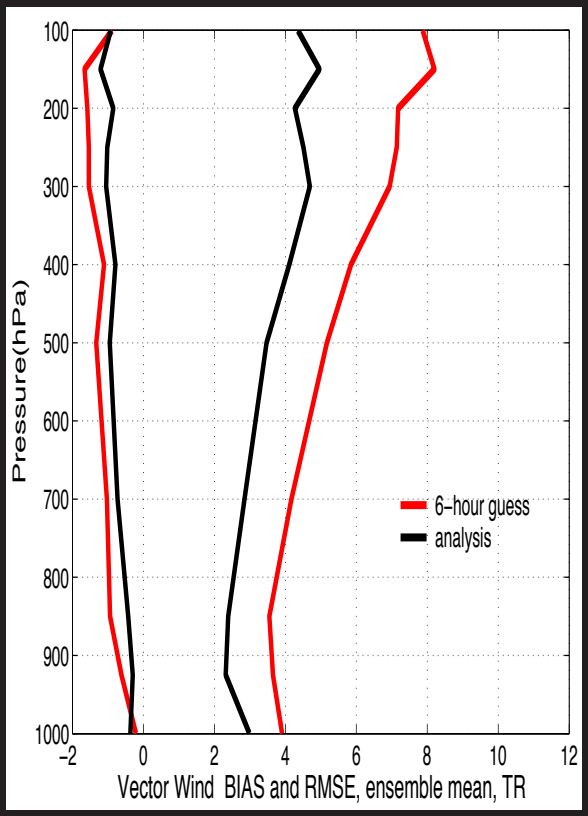


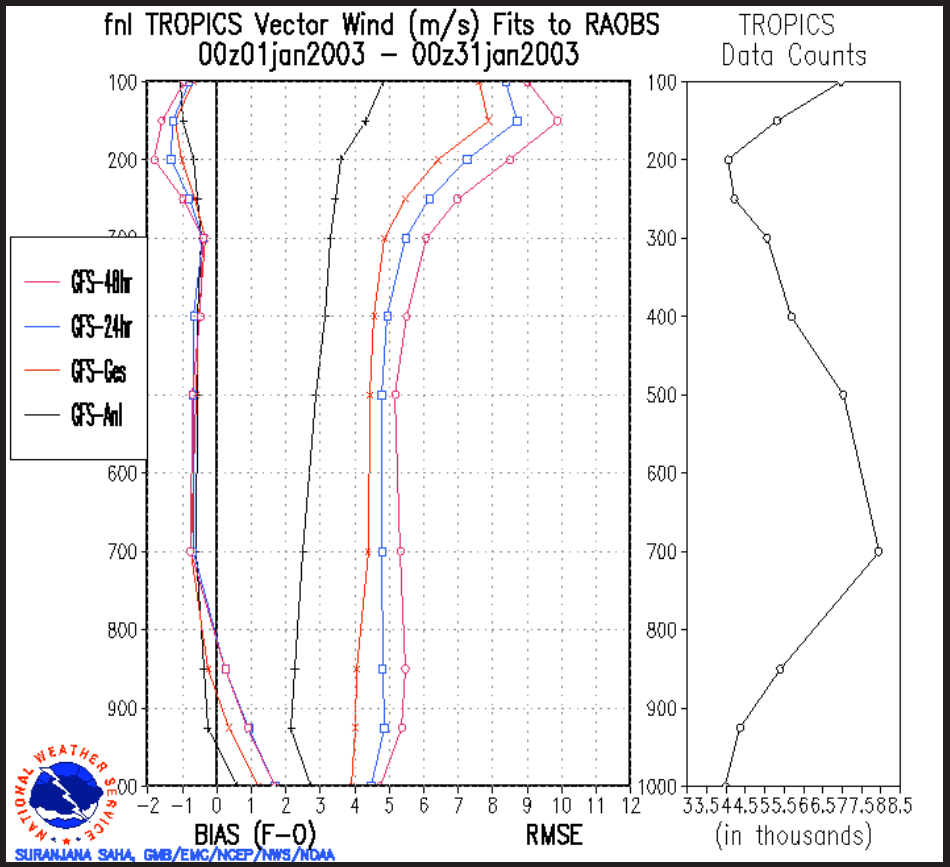


CAM/DART

NCEP GFS

Monthly Mean TR Wind Jan 2003

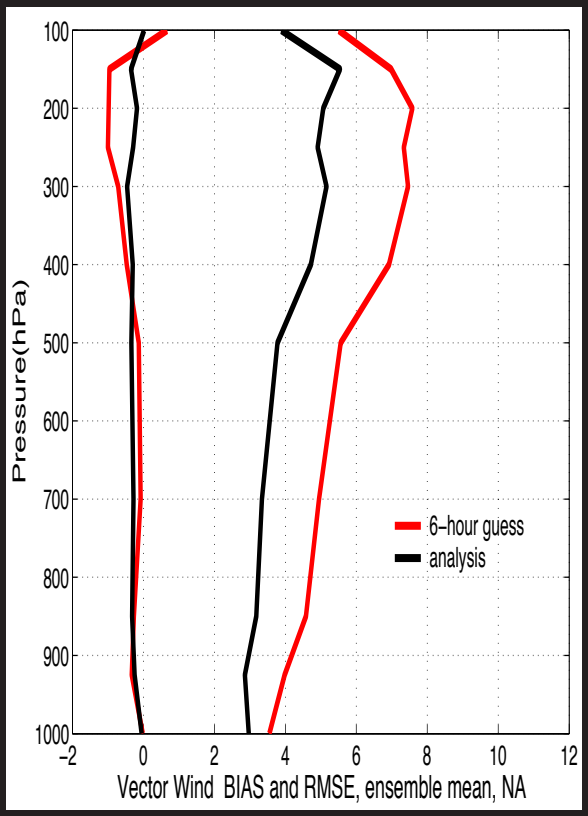


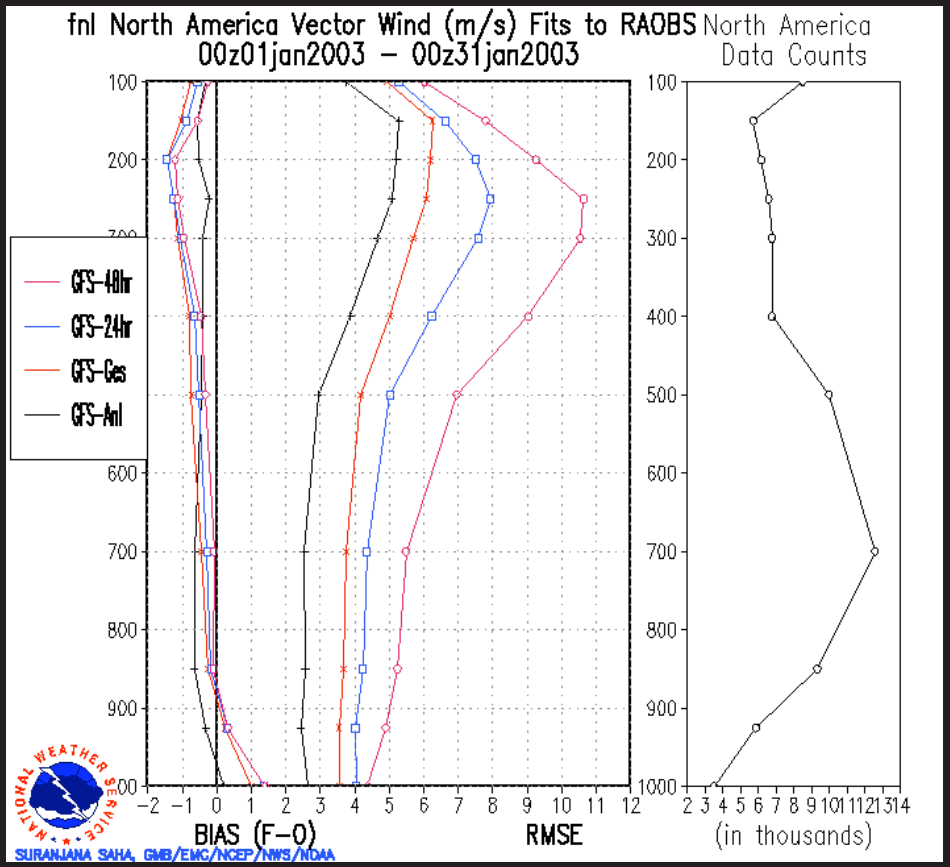


CAM/DART

NCEP GFS

Monthly Mean NA Wind Jan 2003





CAM/DART

NCEP GFS