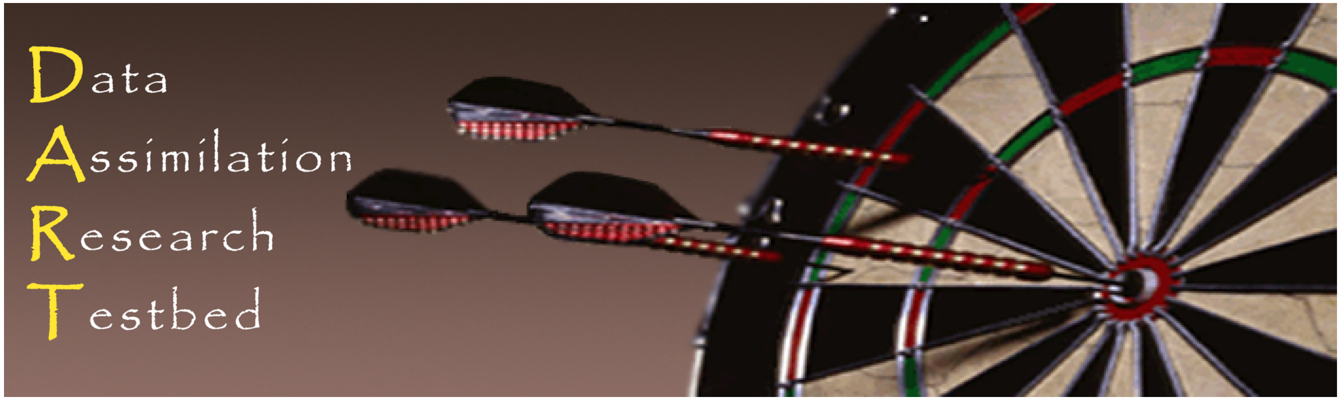
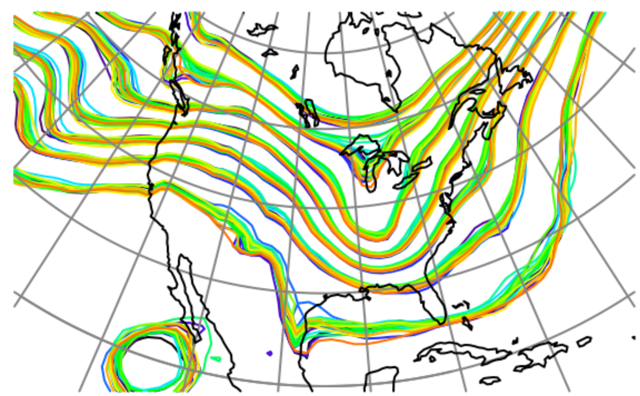


Using the Data Assimilation Research Testbed for Climate System Applications

Jeff Anderson (and many collaborators), NCAR/CISL

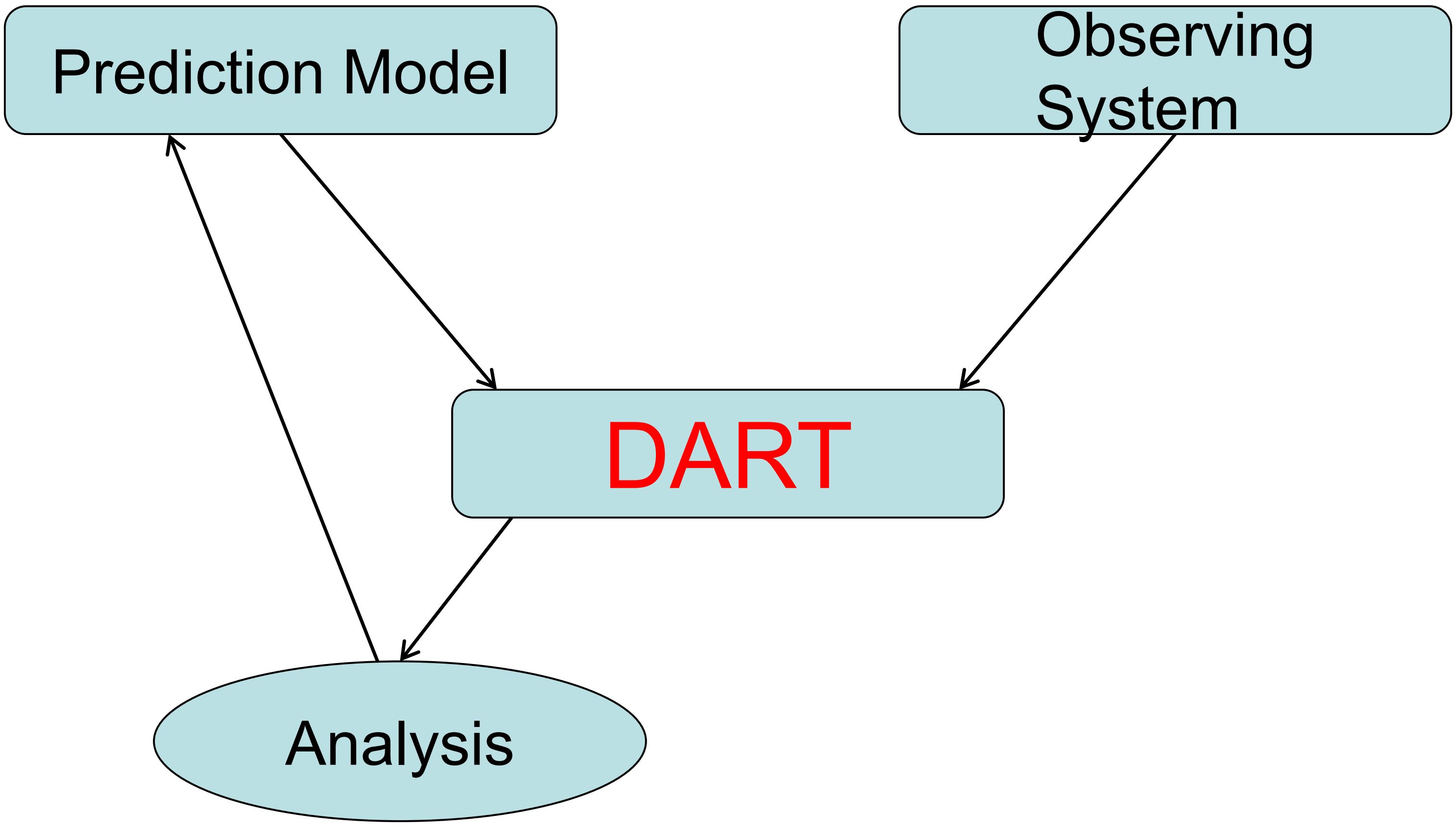


The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR | National Center for
UCAR | Atmospheric Research

The Data Assimilation Research Testbed (DART)

DART provides data assimilation ‘glue’ to build ensemble forecast systems for the atmosphere, ocean, land, …



Data Assimilation Research Testbed (DART)

- A state-of-the-art Data Assimilation System for Geoscience
  - Flexible, portable, well-tested, extensible, free!
  - Works with many models.
  - Works with any observations: Real, synthetic, novel.
- A Data Assimilation Research System
  - Theory based, widely applicable general techniques.
  - Localization, Sampling Error Correction, Adaptive Inflation, …
- Professionally software engineering
  - Carefully constructed and verified
  - Excellent performance
  - Comprehensive documentation
- People: The DAReS Team

DART is used at:

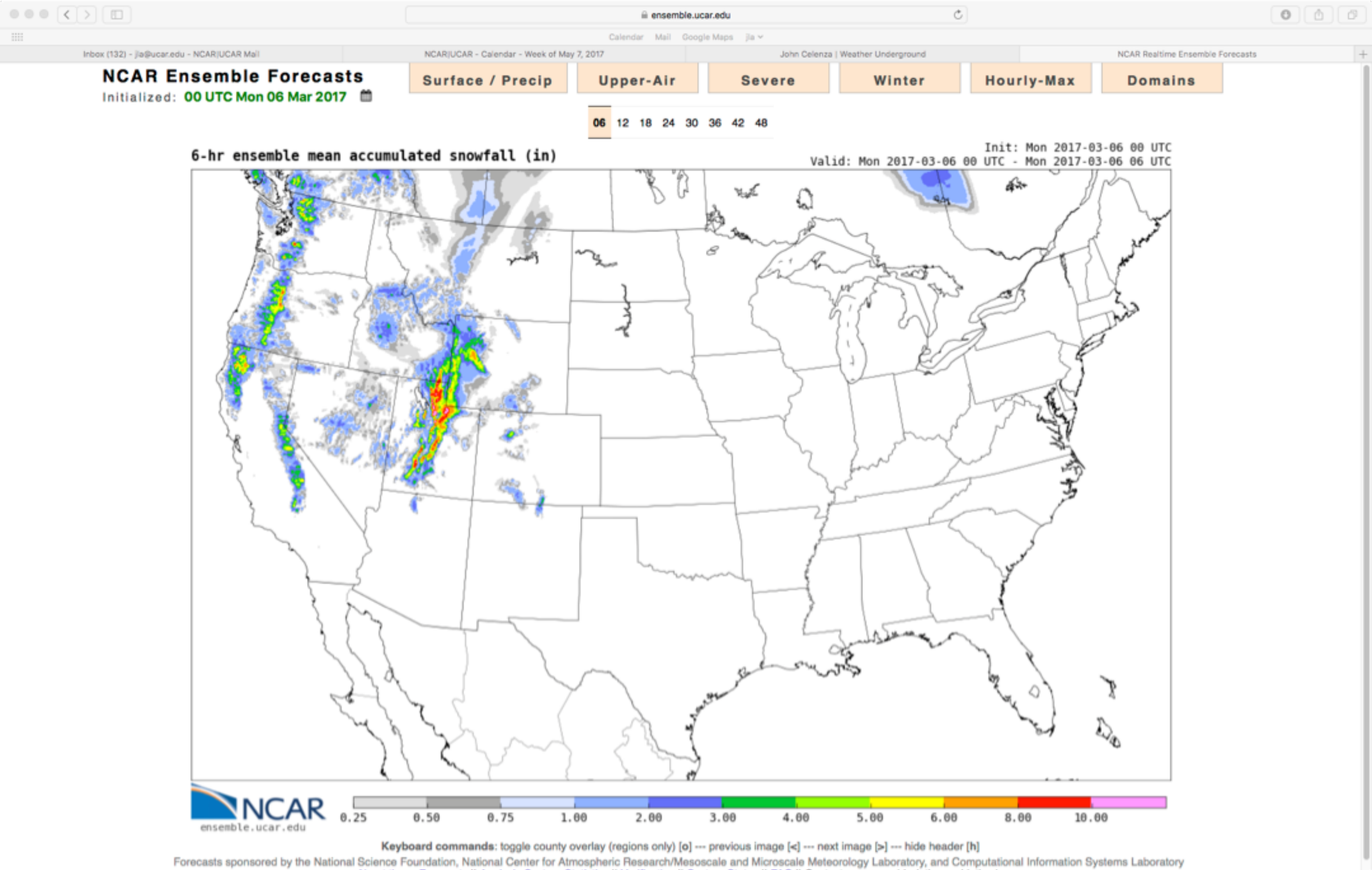
48 UCAR member universities,
More than 100 other sites,
(More than 1500 registered users).

AMS, 8 January 2018

DART Accelerates Forecast System Development

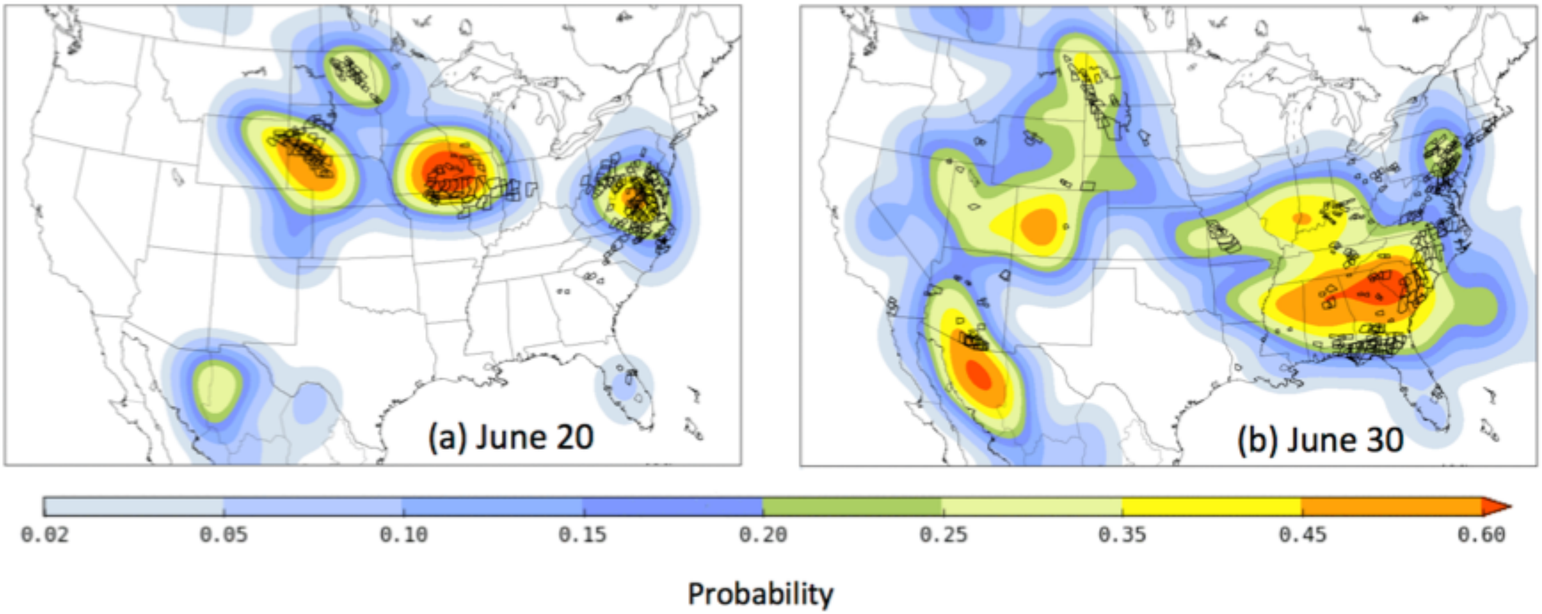
- Works with nearly all NCAR community models (dozens of other models, too).
- New models can be added in weeks.
- Adding new observations is even easier.
- Modular: models, observations and assimilation tools easily combined.
- Enables DA use by prediction scientists.
  Doesn't require assimilation expertise.
- Fast & efficient software: laptops to supers.

Example: NCAR Real-time ensemble prediction system



Example: NCAR Real-time ensemble prediction system

Severe weather forecast for *two* days compared to NWS warnings



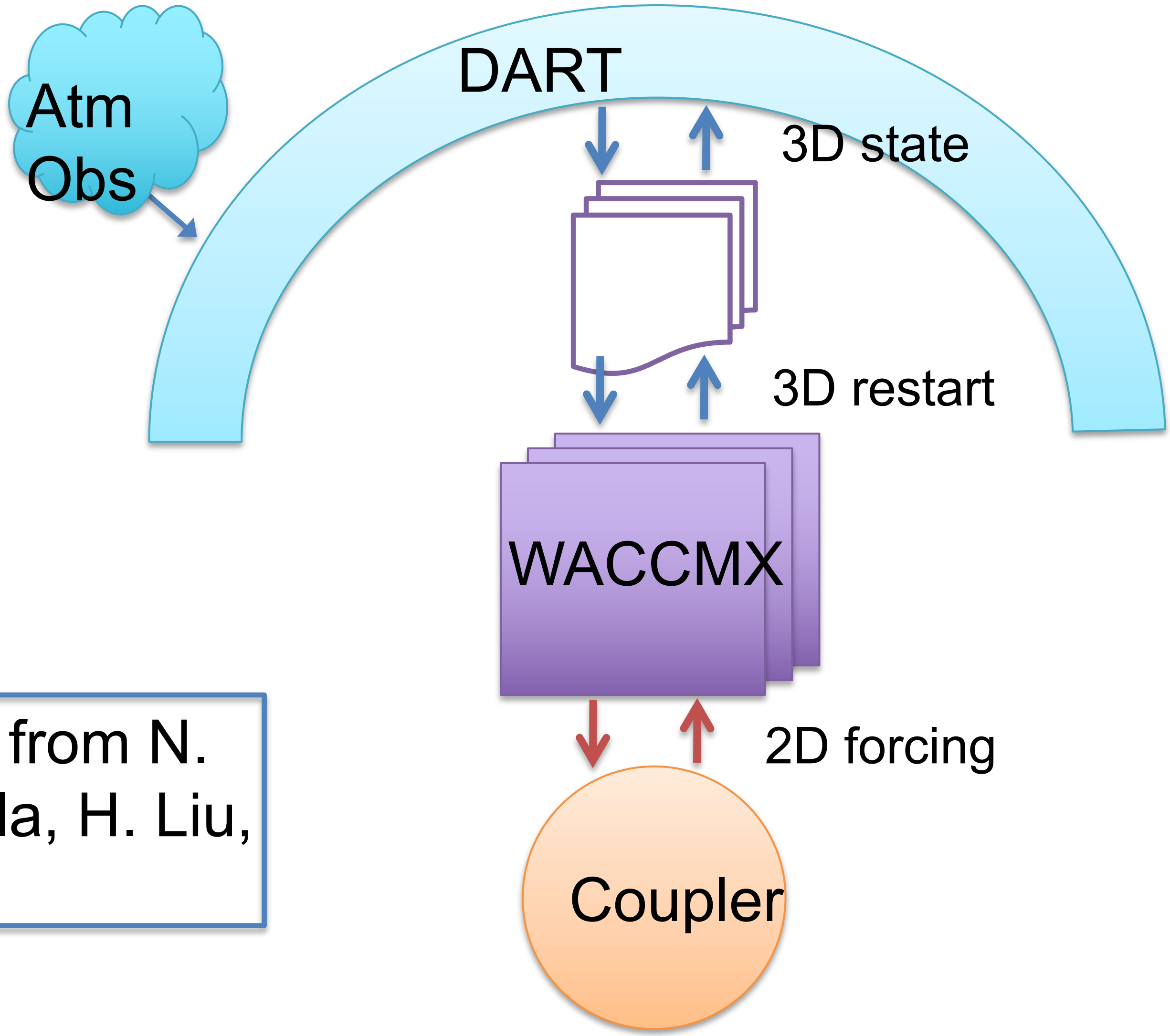
- WRF, 10 member ensemble, GFS for boundary conditions
- Continuous operation since April, 2015
- 48 hour forecasts at 3km resolution
- First continuously cycling ensemble system for CONUS

DART Applications with CESM Earth System Models

DART interfaces exist for many components of NCAR's Community Earth System Model:

- Lower atmosphere: CAM-FV, CAM-SE, MPAS
- Upper atmosphere, ionosphere: WACCM, WACCMX
- Atmospheric Chemistry: CAM/Chem
- Ocean: POP
- Land surface / biosphere: CLM
- Sea Ice: CICE
- Weakly coupled DA combinations of the above

Deep Atmospheric Component Coupled DA



Results from N. Pedatella, H. Liu, J. Liu

Deep Atmospheric Component Coupled DA

WACCMX:

- 2 degrees, 126 levels, top at 4.1×10^{-10} hPa
- High-top extension of CAM
- Includes ionospheric processes
- Persistence forecasts of solar and geomagnetic forcing

Observations:

- All in situ plus GPS refractivity in trop/lower strat.
- Temperature from AURA Microwave Limb Sounder (MLS)
- Temperature from TIMED/SABER
- Temperatures only up to 100km

DART:

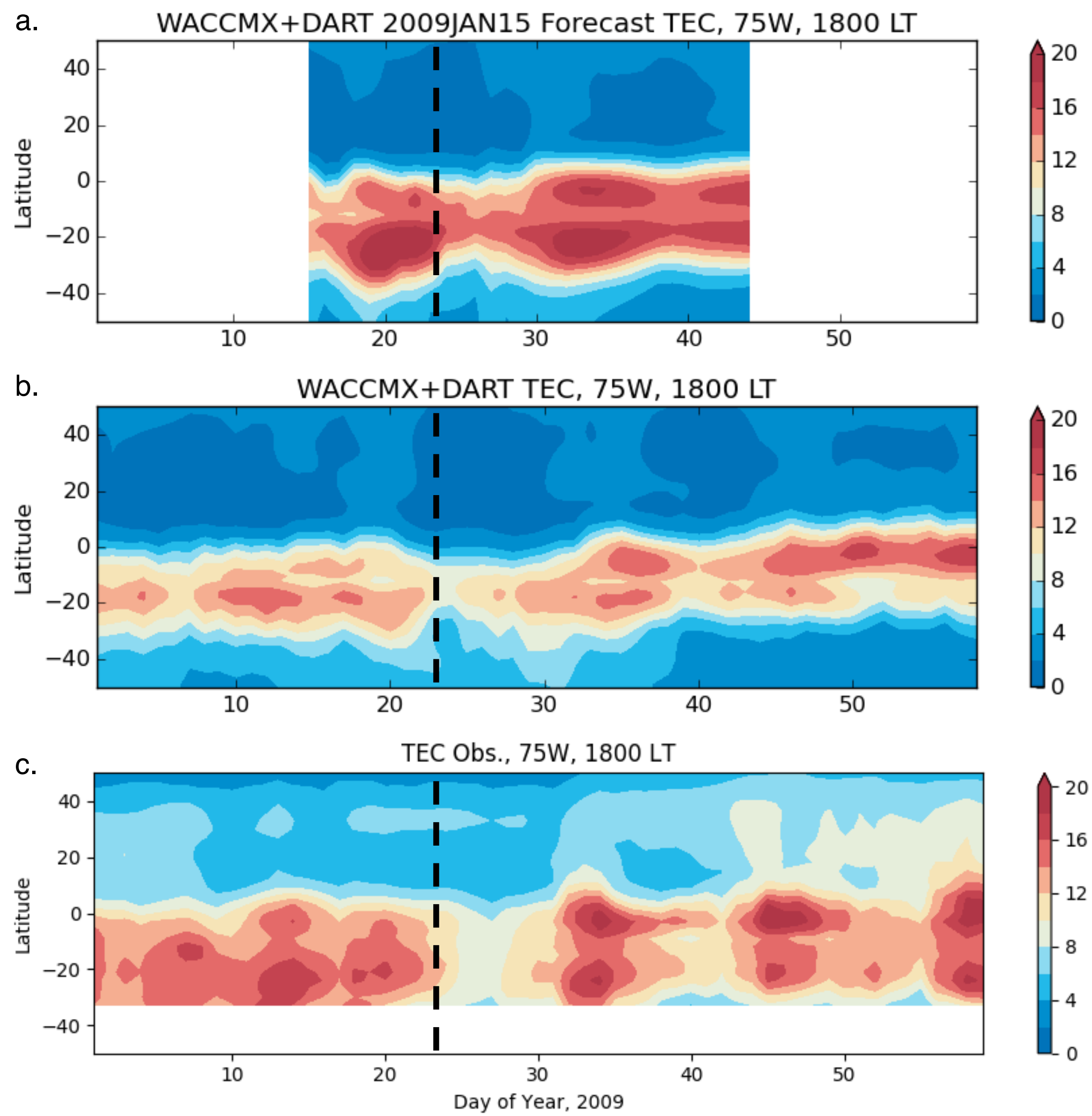
- 40 members
- Adaptive inflation, GC localization
- 6-hour window

Deep Atmospheric Component Coupled DA

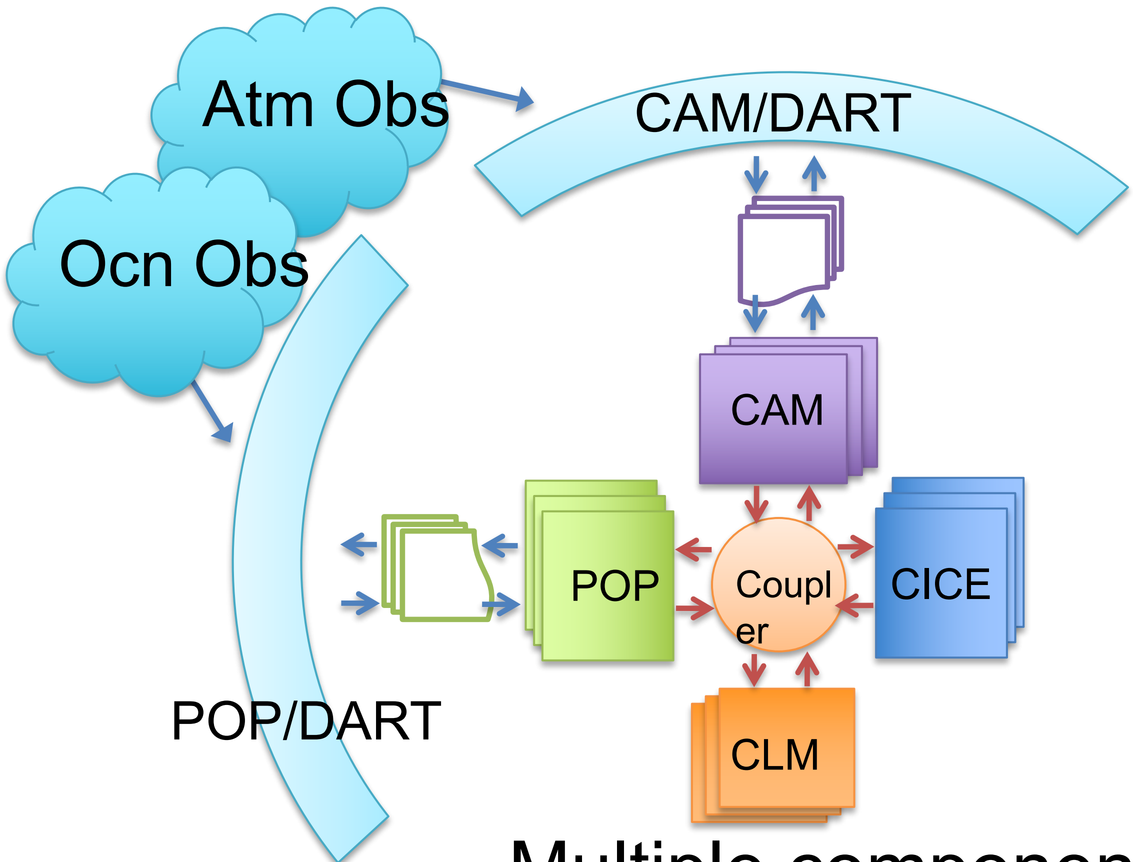
Impact of SSW on ionosphere

Forecast (top panel), reanalysis (middle), and independent obs of Total Electron Content.

Agreement of forecast with observations indicates significant prediction skill.



Multiple Component POP/CAM Coupled DA



Multiple components assimilating with different DART(s) in fully-coupled CESM.

Results from A. Karspeck

Multiple Component POP/CAM Coupled DA

Weakly coupled reanalysis from 1970-1981

Model:

- POP, 1 degree, standard CESM configuration
- CAM-FV, 1 degree, standard CESM configuration

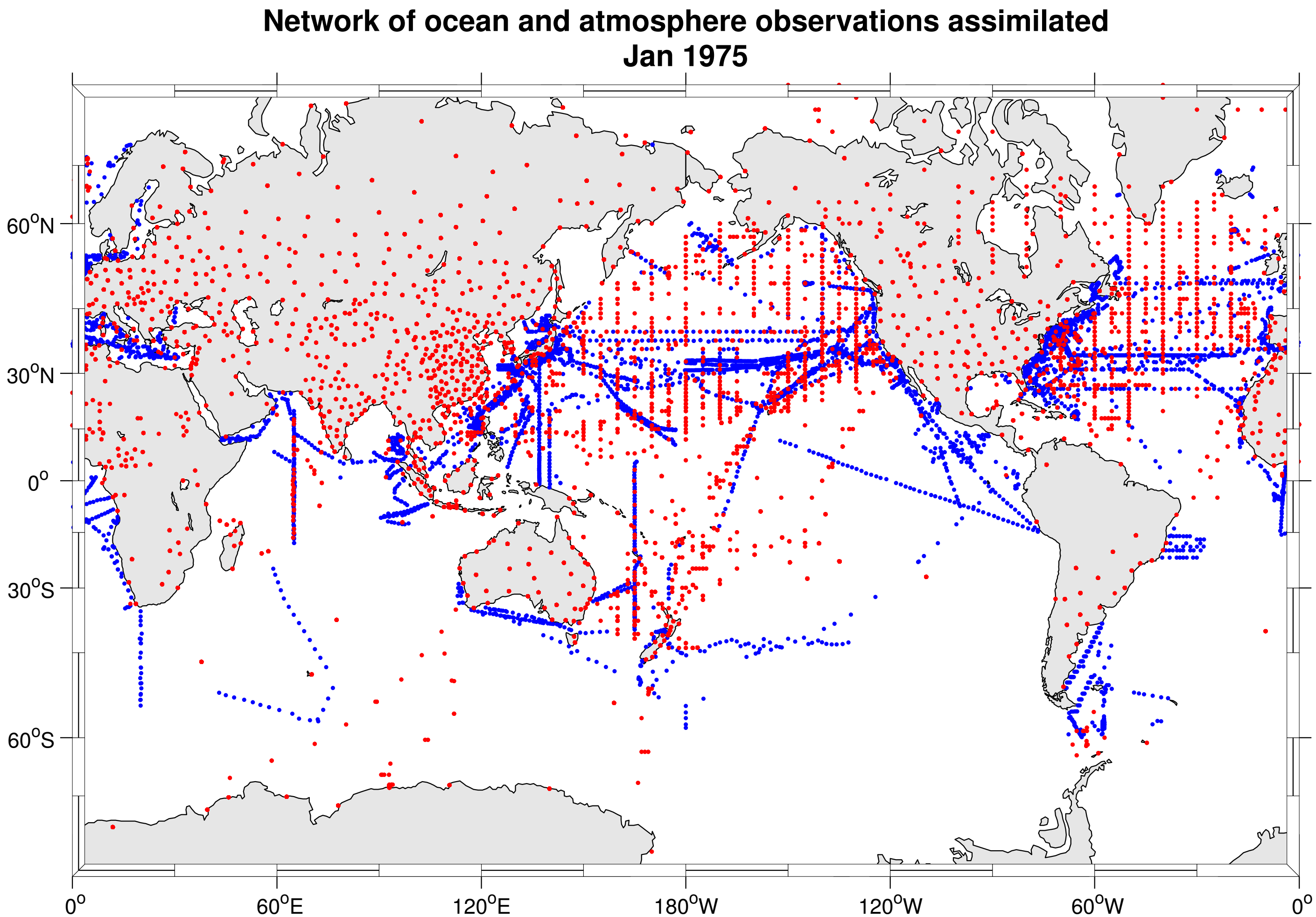
Observations:

- In-situ atmosphere observations from NCEP reanalysis
- Ocean temperature and salinity, World Ocean Database

DART:

- 30 members
- Limited adaptive inflation in ocean
- Fully adaptive inflation in atmosphere
- GC localization

Multiple Component POP/CAM Coupled DA



Observations are sparse for this period.

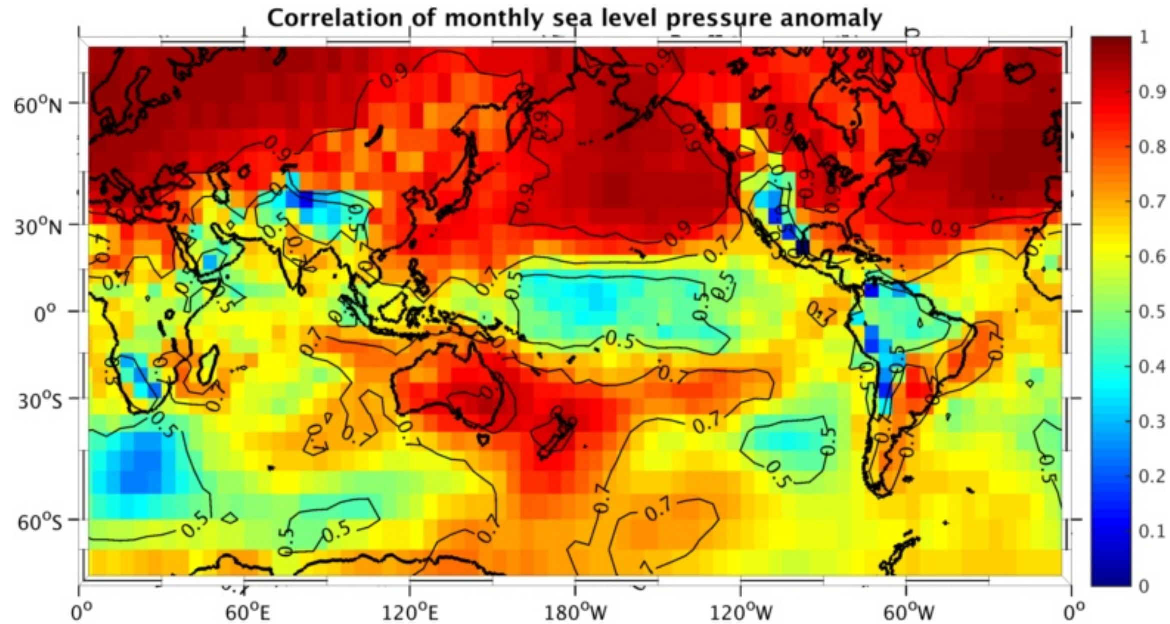
Multiple Component POP/CAM Coupled DA

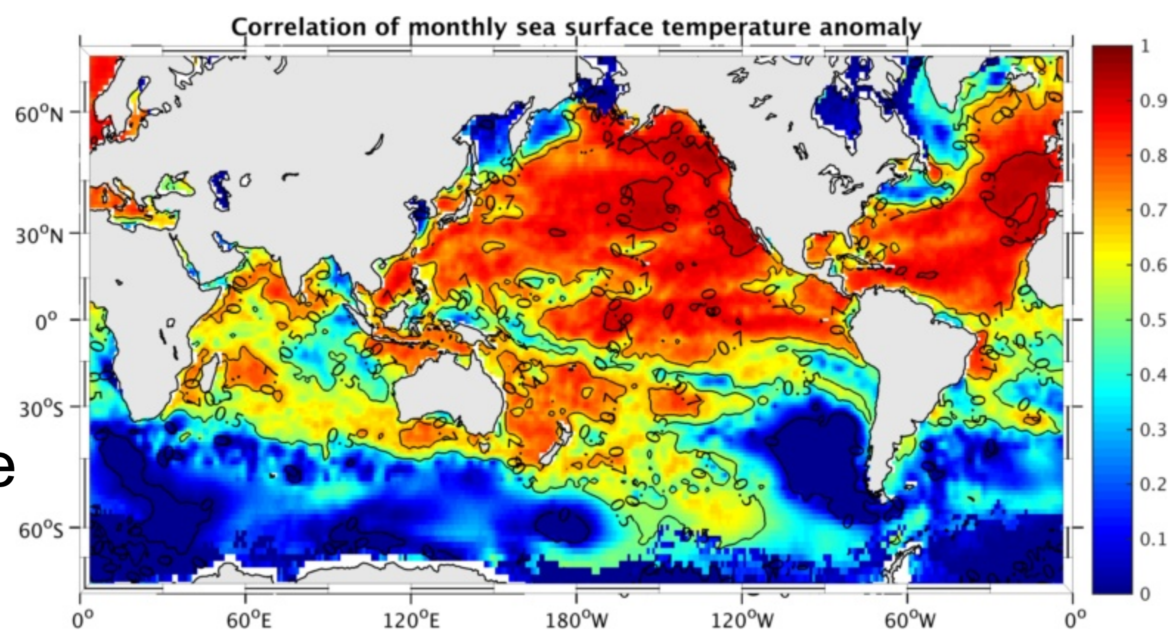
Comparisons to HADISST and HADSLP.

Correlation high where observations existed.

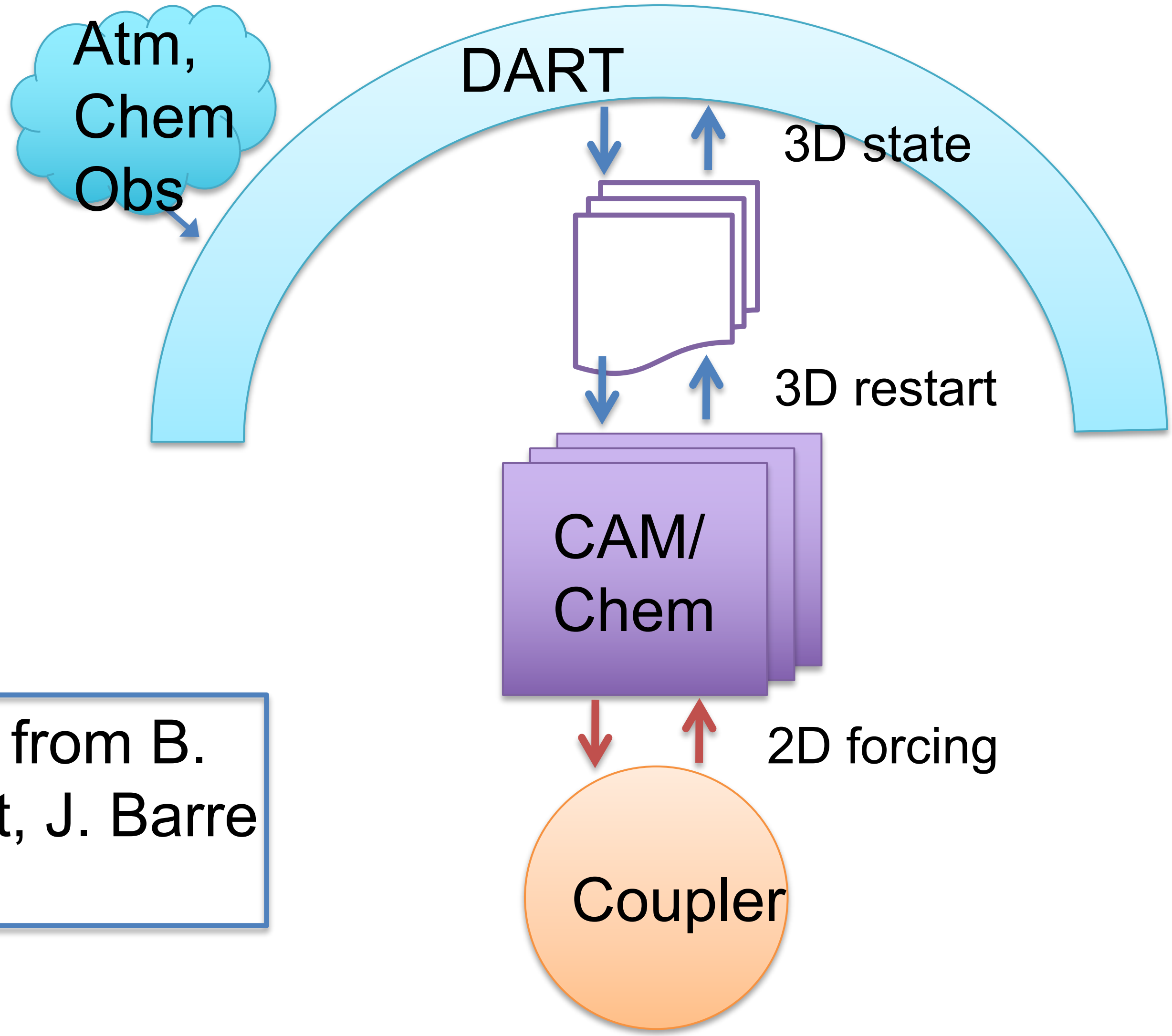
DART did not assimilate SST products or observations.

Produces competitive reanalysis.





Tropospheric Chemical Weather DA



Results from B. Gaubert, J. Barre

Tropospheric Chemical Weather DA

Chemical Weather Reanalysis for Summer 2008

Model:

- CAM-FV 2 degree 30 levels
- Mozart-4 tropospheric chemistry

Observations:

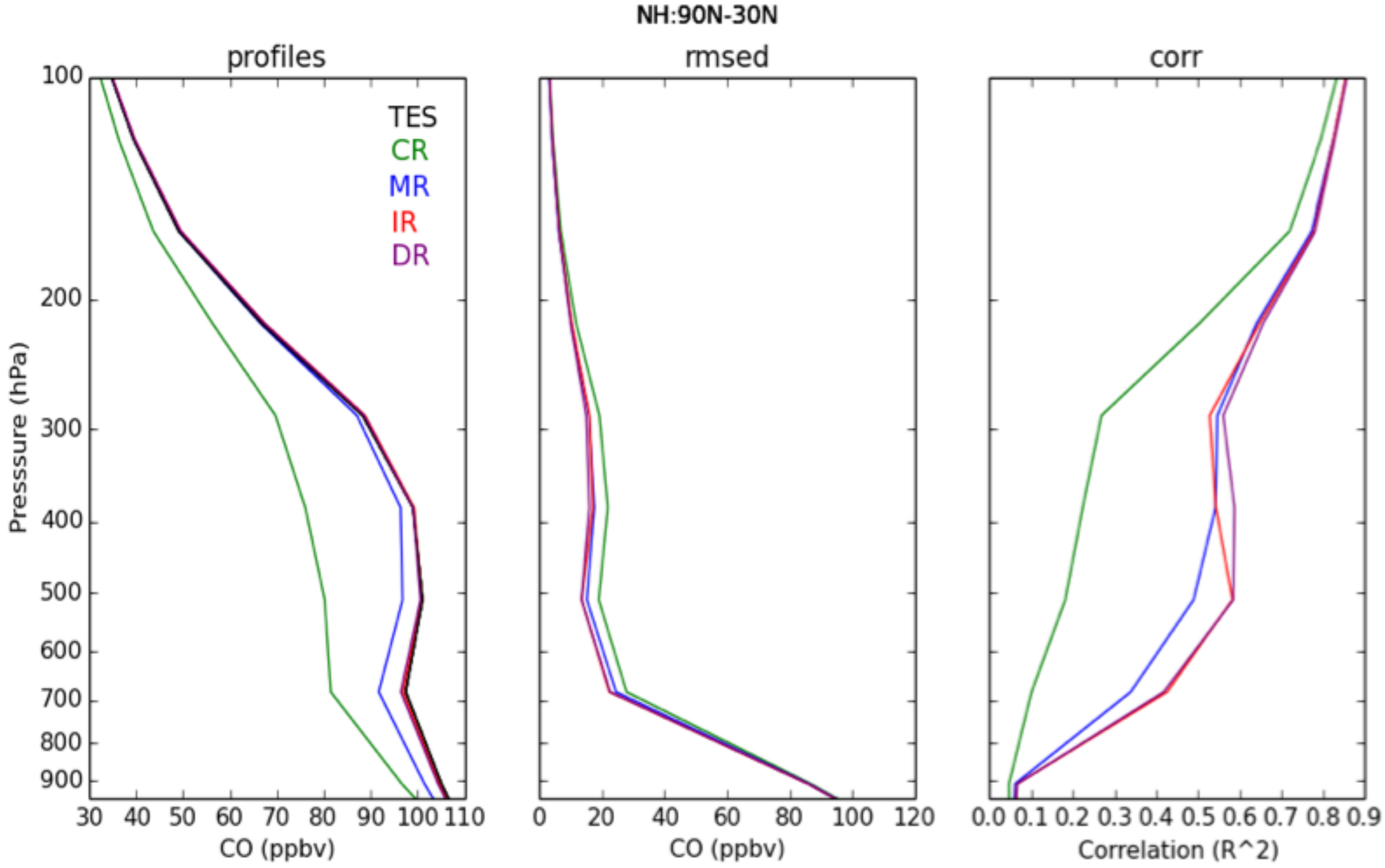
- In-situ atmosphere observations from NCEP reanalysis
- MOPITT and IASI CO retrieved profiles

DART:

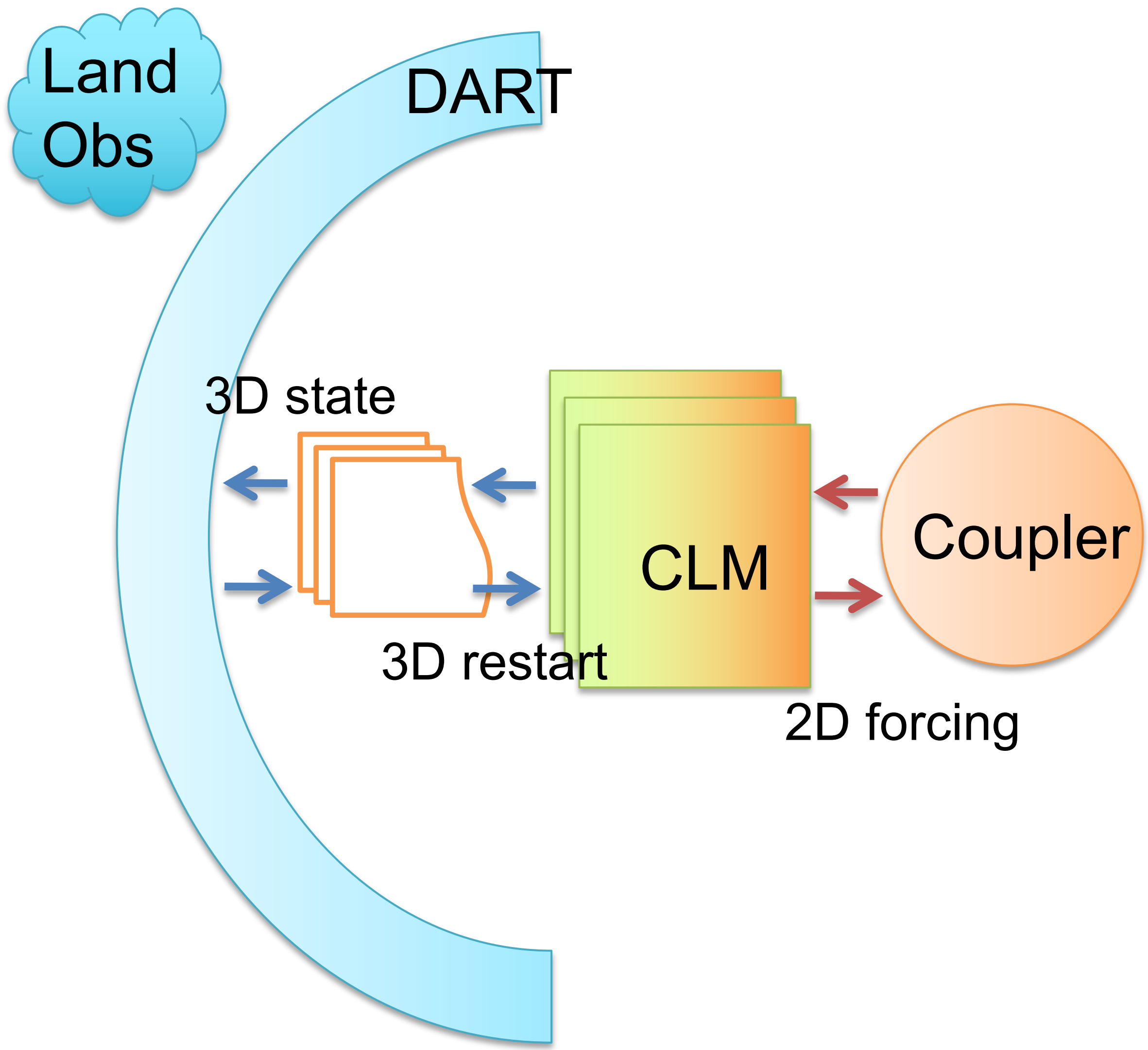
- 30 members
- Adaptive inflation
- GC localization, more localized for CO obs

Tropospheric Chemical Weather DA

CO forecast fits to observations improved with DA.
Comparison to independent TES CO obs greatly improved.



CLM Land Component Coupled DA

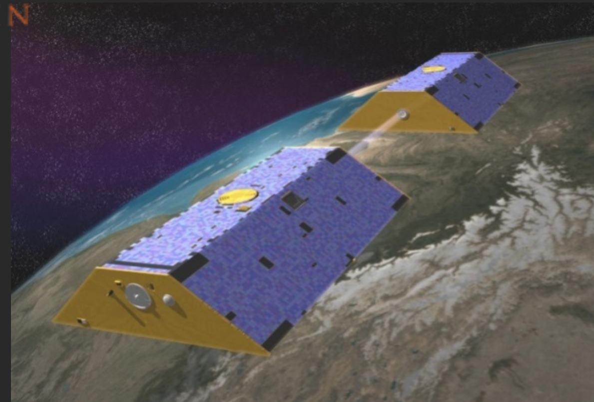


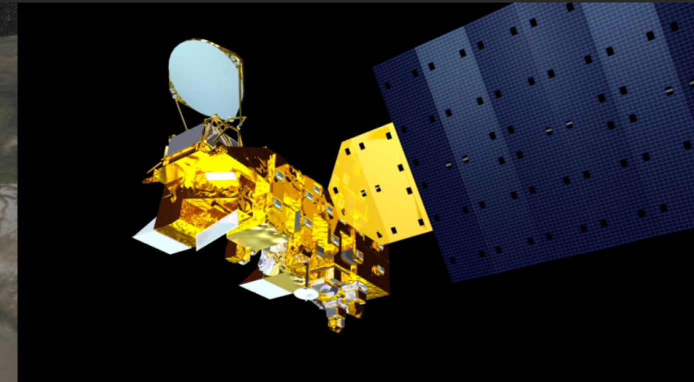
Some of the researchers using CLM/DART

- **Yong-Fei Zhang** (UT Austin)
  - multisensor snow data assimilation
- **Andy Fox** (NEON)
  - flux observations/state estimation
- **Hanna Post** (Jülich)
  - assimilation & parameter estimation
- **Raj Shekhar Singh** (UC Berkeley)
  - groundwater
- **Long Zhao** (UT Austin)
  - AMSR-E radiances, empirical vegetated surface RTM, soil moisture (SMAP)
- **Ally Toure** (NASA-Goddard USRA)
  - brightness temperatures
- **Yonghwan Kwon** (UT Austin)
  - sensitivity of assimilation of brightness temperatures from multiple radiative transfer models on estimates of snow water equivalent.

THE UNIVERSITY OF TEXAS AT AUSTIN
JACKSON
SCHOOL OF GEOSCIENCES

THE UNIVERSITY OF TEXAS AT AUSTIN
CIESS
CENTER FOR INTEGRATED EARTH SYSTEM SCIENCE





Improving Estimates of Snowpack Water Storage in the Northern Hemisphere Through a Newly Developed Land Data Assimilation System

Yong-Fei Zhang[1], Zong-Liang Yang[1,2], Yonghwan Kwon[1], Tim J. Hoar[3], Hua Su[1], Jeffrey L. Anderson [3], Ally M. Toure [4,5], and Matthew Rodell [5]

[1]Jackson School of Geosciences, University of Texas at Austin, Austin, TX, United States.
[2]Key Lab of Regional Climate-Environment for Temperate East Asia (RCE-TEA), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.
[3]The National Center for Atmospheric Research, Boulder, CO, United States.
[4]Universities Space Research Association (USRA), Columbia, MD, United States.
[5]NASA Goddard Space Flight Center, Greenbelt, MD, United States.

NASA

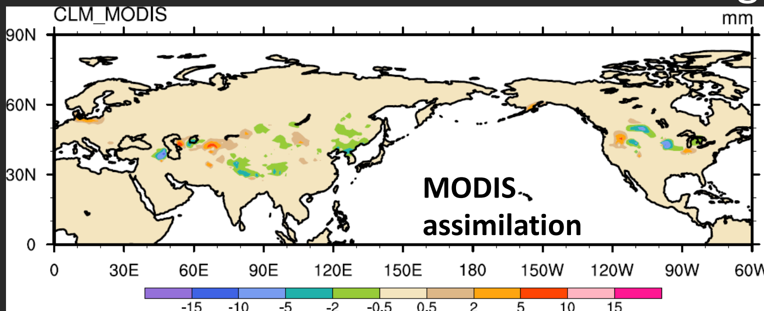
NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

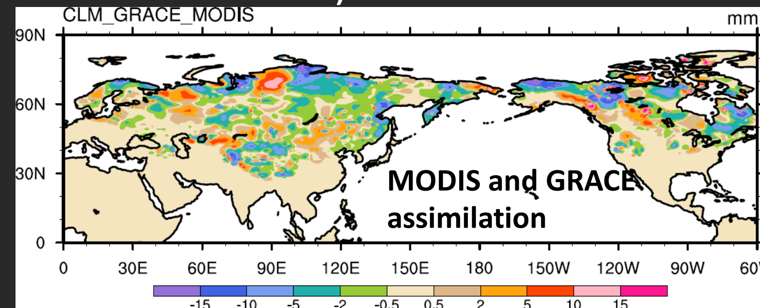
AMS, 8 January 2018

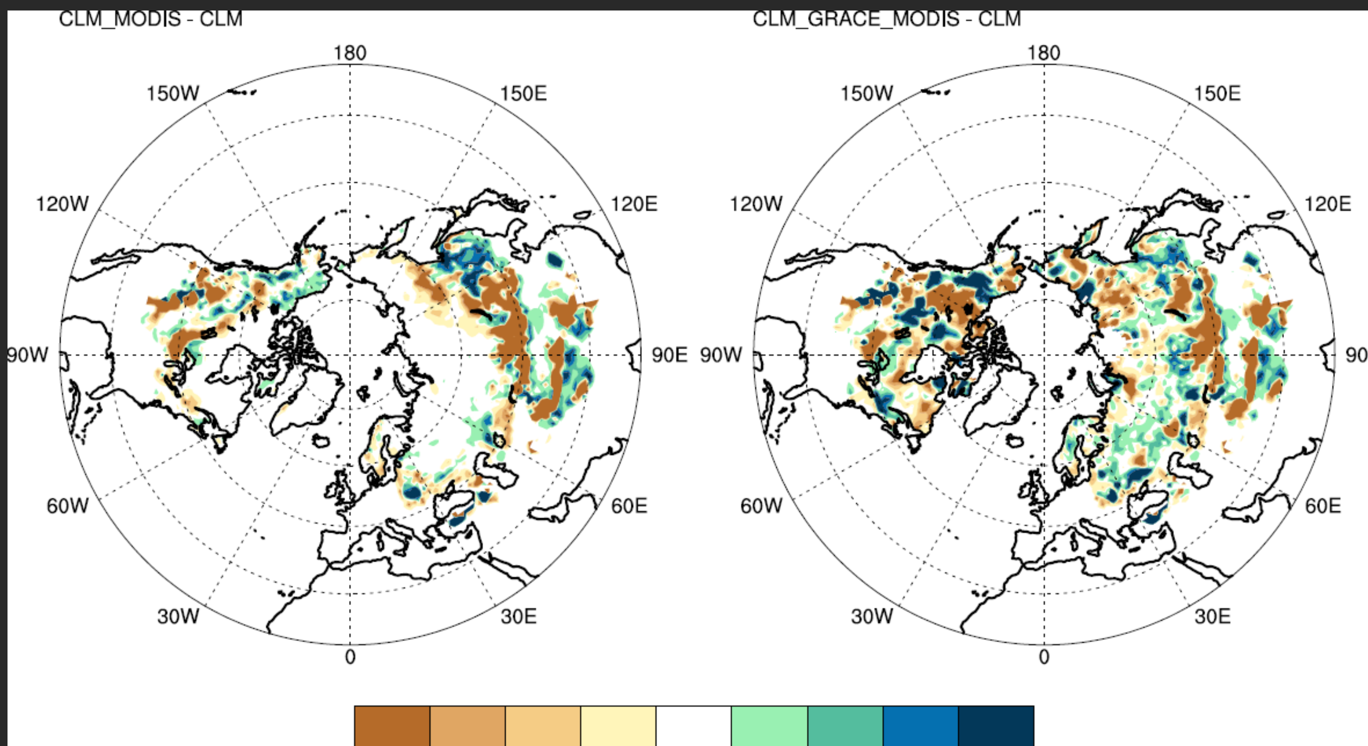
RCTEA

Assimilation Results

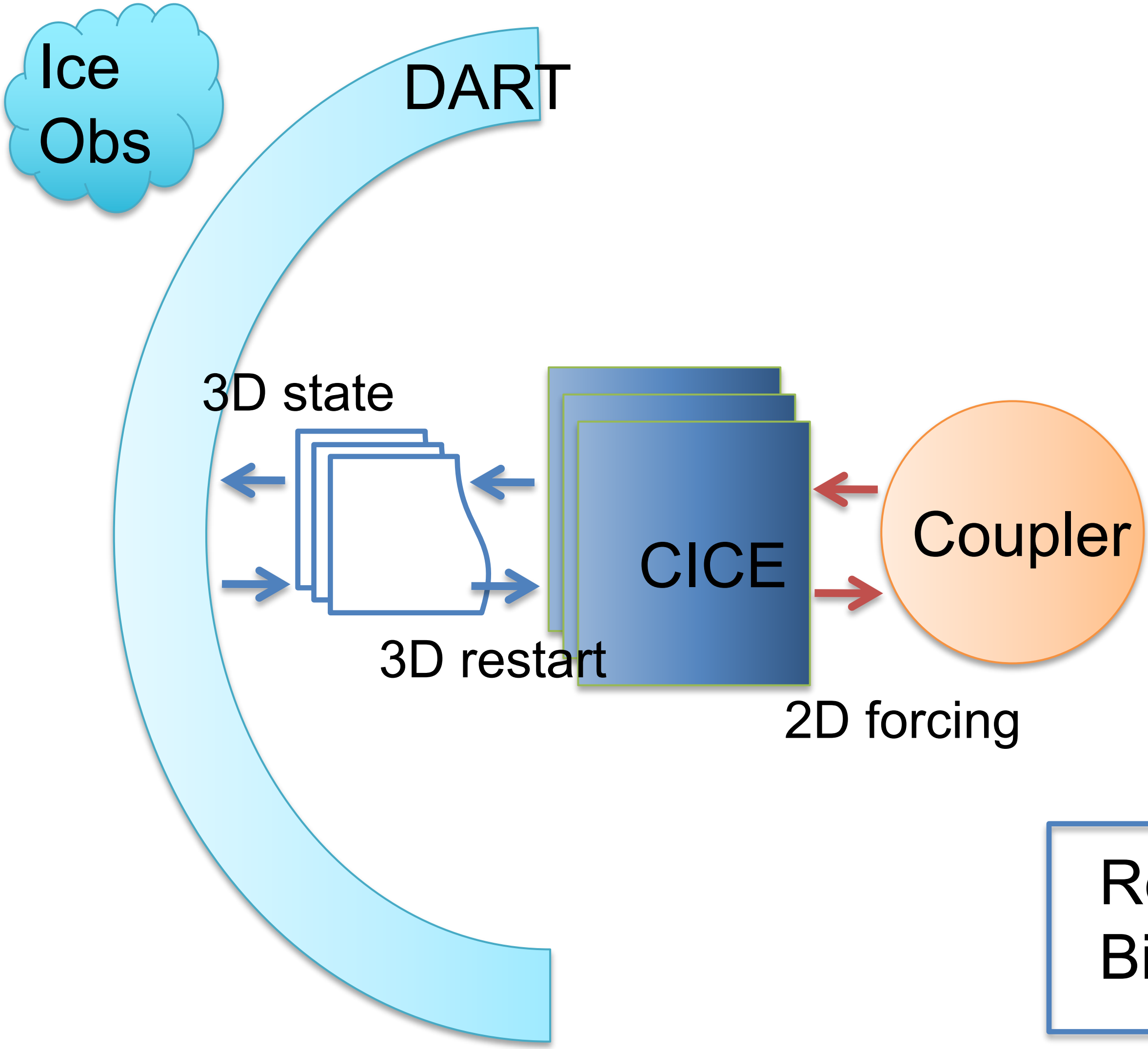
Snow Water Storage (Posterior minus Prior)







CICE Sea Ice Component Coupled DA



Results from C. Bitz, Y. Zhang

CICE Sea Ice Component Coupled DA

Sea Ice OSSE

Model:

- CICE-5 forced by slab ocean
- Atmospheric forcing from CAM ensemble reanalysis

Observations:

- Sea ice concentration, age, thickness

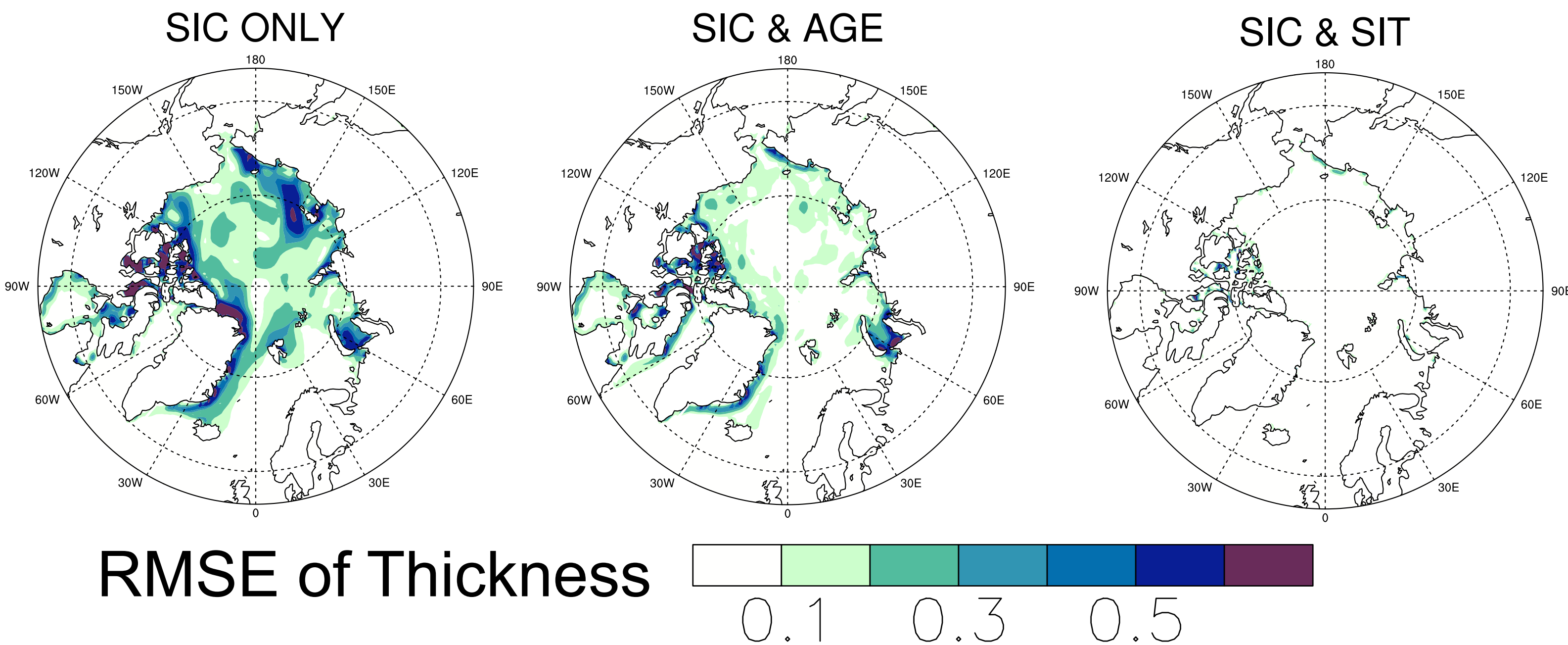
DART:

- 30 members
- Adaptive inflation
- GC localization

CICE Sea Ice Component Coupled DA

OSSE used to explore information content of different obs.

Sea ice concentration alone not as good as when combined with age or thickness observations.

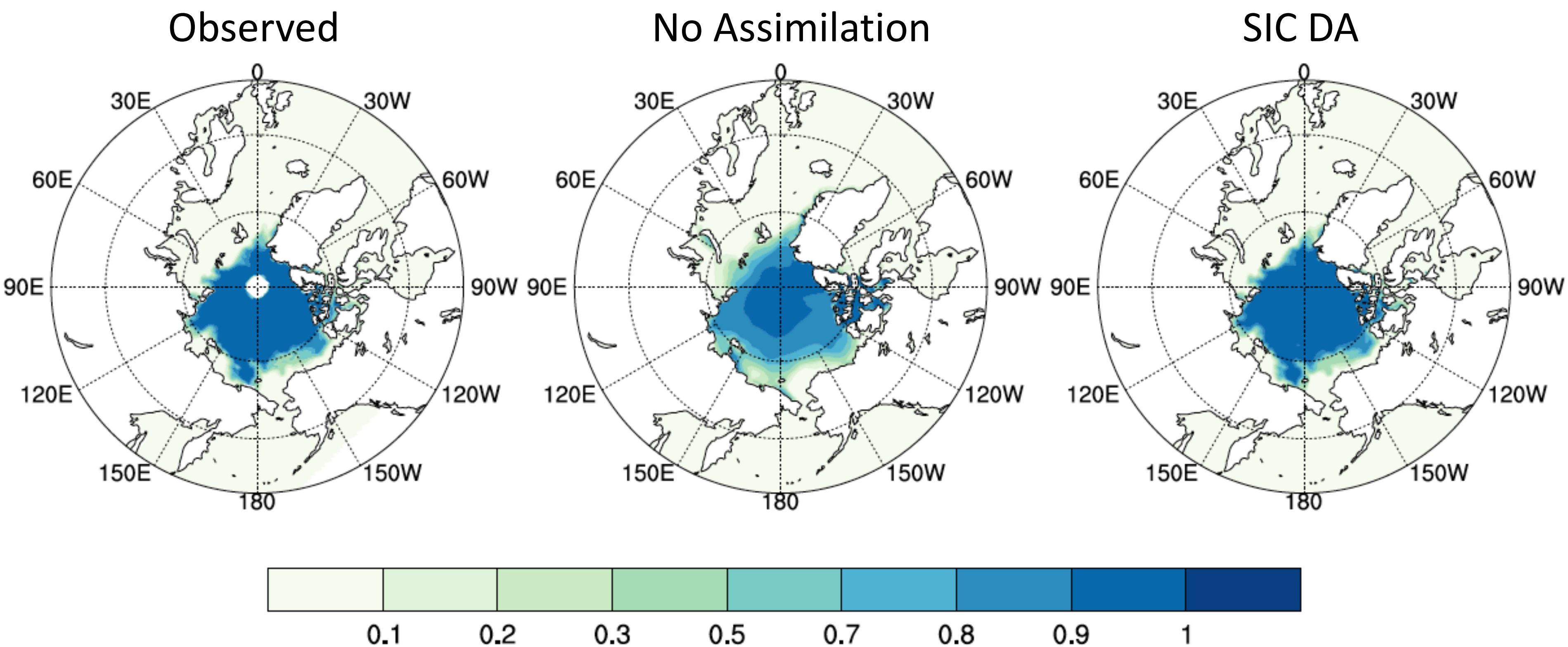


AMS, 8 January 2018

CICE Sea Ice Component Coupled DA

Sea ice concentration from SSM/I retrievals as next step. Reanalysis is moved much closer to observed concentration.

Sea ice concentration for September 2001



What might DA do for climate model applications?

- ICs for predictions,
- Produce reanalyses to help increase understanding,
- Confront models with observations, find inconsistencies (but almost never gives direct path to improving model),
- Parameter estimation for model ‘tuning’.

Ensemble methods can provide information about uncertainty (and other aspects of distributions) for all of these.

DA Challenges for All Applications

Many assumptions of Bayesian theory are violated:

- Unbiased model and observations,
- Uncorrelated observation errors,
- Exact estimates of observation error distribution,
- Exact estimates of representation error.

Prior (forecast)

Likelihood

Posterior (analysis)

$$P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k|\mathbf{x})P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$

DA Challenges for All Applications

For Kalman Filter class algorithms following are violated:

- Gaussian priors and observation errors,
- Linear relation between model state and observations.

For Ensemble Kalman filter algorithms add:

- Sufficiently large ensembles,
- Model provides accurate estimates of second moments.

DA Challenges for All Applications

Because of all these violations of assumptions, there is no way to assess the quality of DA results a priori.

It is **essential to calibrate and validate** DA results.

This is even true for very mature NWP systems.

For novel climate system applications it is even more vital.

Requires lots of observations spanning many decorrelation times for model dynamics.

DA Challenges for All Applications

Calibration and validation must include ensemble statistics.

This requires even more observations.

Ensemble statistics aside from 1st, 2nd moment are suspect.

- Kalman filter does not generate estimates of these.
- Not clear why ensemble Kalman filters should.

DA Challenges for All Applications

Non-equilibrium ”off attractor” model evolution.

E.g., spurious numerical gravity waves in NWP.

DA can cause state variables not found in free runs.

These can challenge the model numerics.

Example from WACCMX:

Model damping and diffusion had to be increased to reduce gravity wave amplitude with DA.

DA approximates solutions to this problem

$$P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k | \mathbf{x}) P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$

This may be inconsistent with what modelers expect.

Example: Parameter estimation for gravity wave drag in CAM.

- Estimate surface roughness at each horizontal gridpoint.
- Result was a very bumpy tropical Pacific, with improved forecasts.

Unless known exactly, ‘conserved’ quantities shouldn’t be conserved.

DA Challenges for Earth System Component Models

$$P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k|\mathbf{x})P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$

DA requires a (stochastic) forecast model:

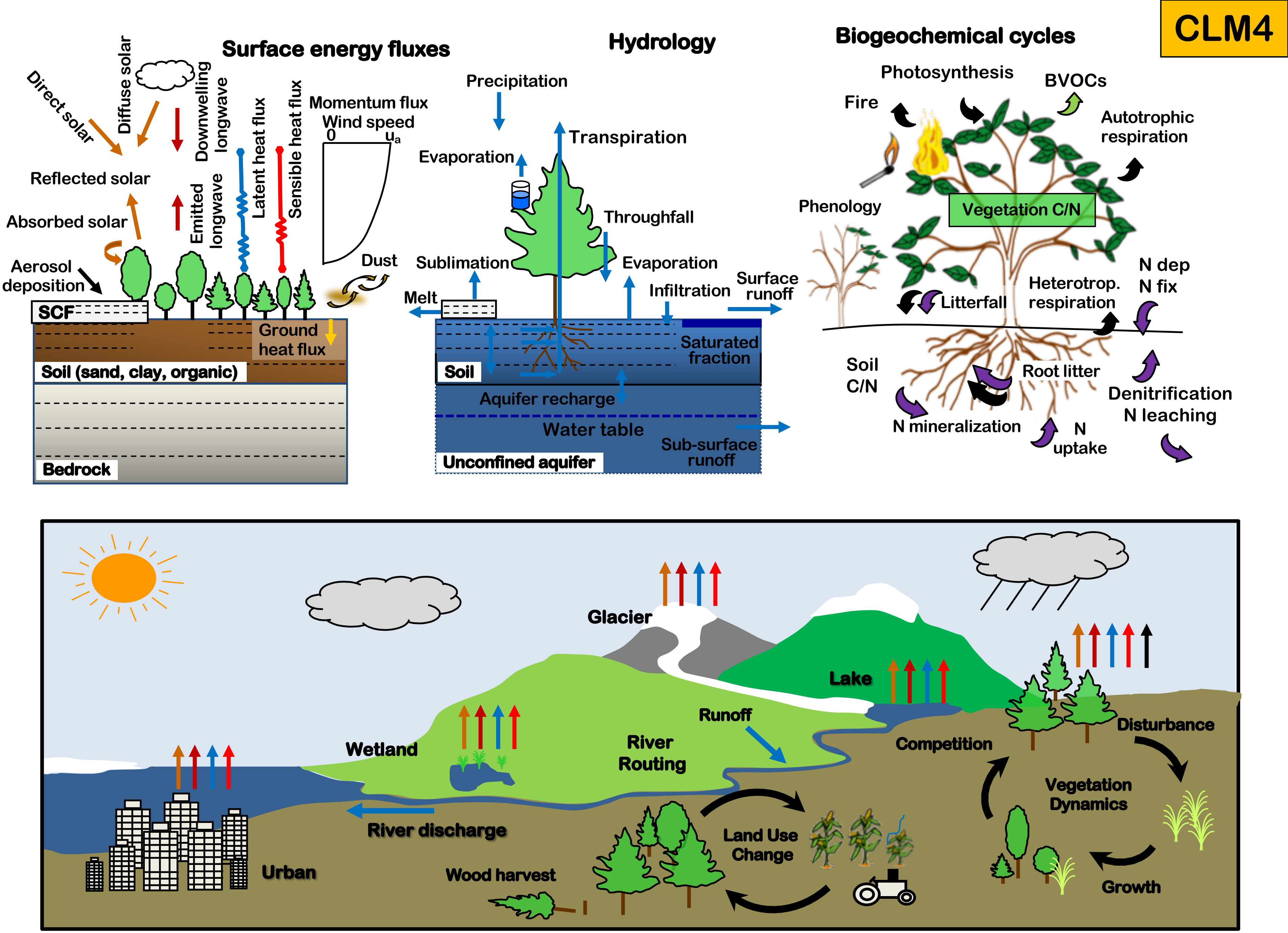
$$m_{k:k+1}\left(\mathbf{x}_{t_k}\right) = f_{k:k+1}\left(\mathbf{x}_{t_k}\right) + g_{k:k+1}\left(\mathbf{x}_{t_k}\right).$$

When applied to a correct analysis distribution ensemble at a previous time, model should produce a correct forecast ensemble distribution for subsequent observations.

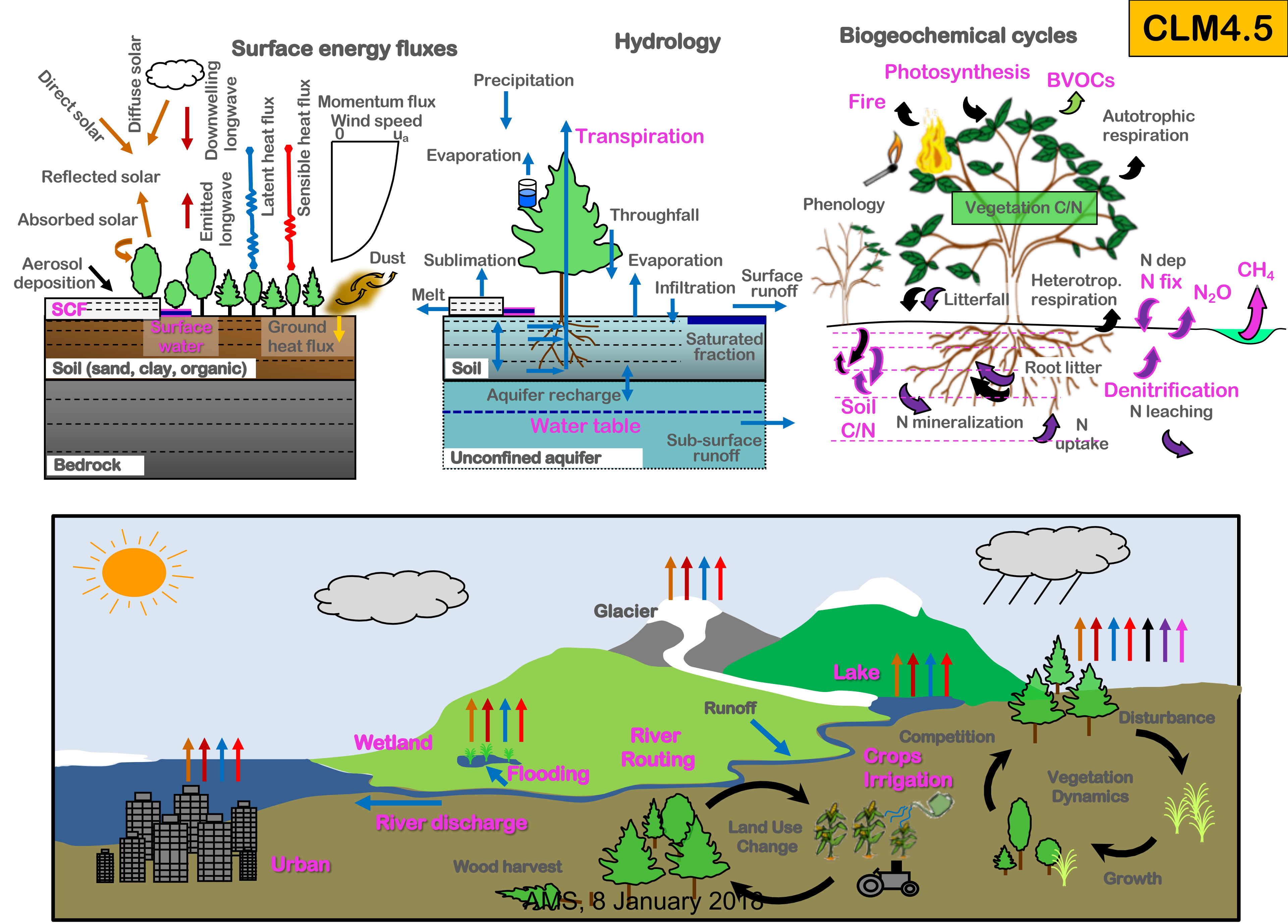
Challenges for Earth System Models (CLM examples)

Earth system component models may not make good forecasts:

- Not as mature as NWP models, especially for forecasts,
- No set of nice PDEs like Navier Stokes,
- Extreme complexity of modeled system,
- Developed as ‘process’ model, not prediction model,
- Lack of model error growth, especially if strongly forced,
- Not developed with DA/prediction as primary objective.



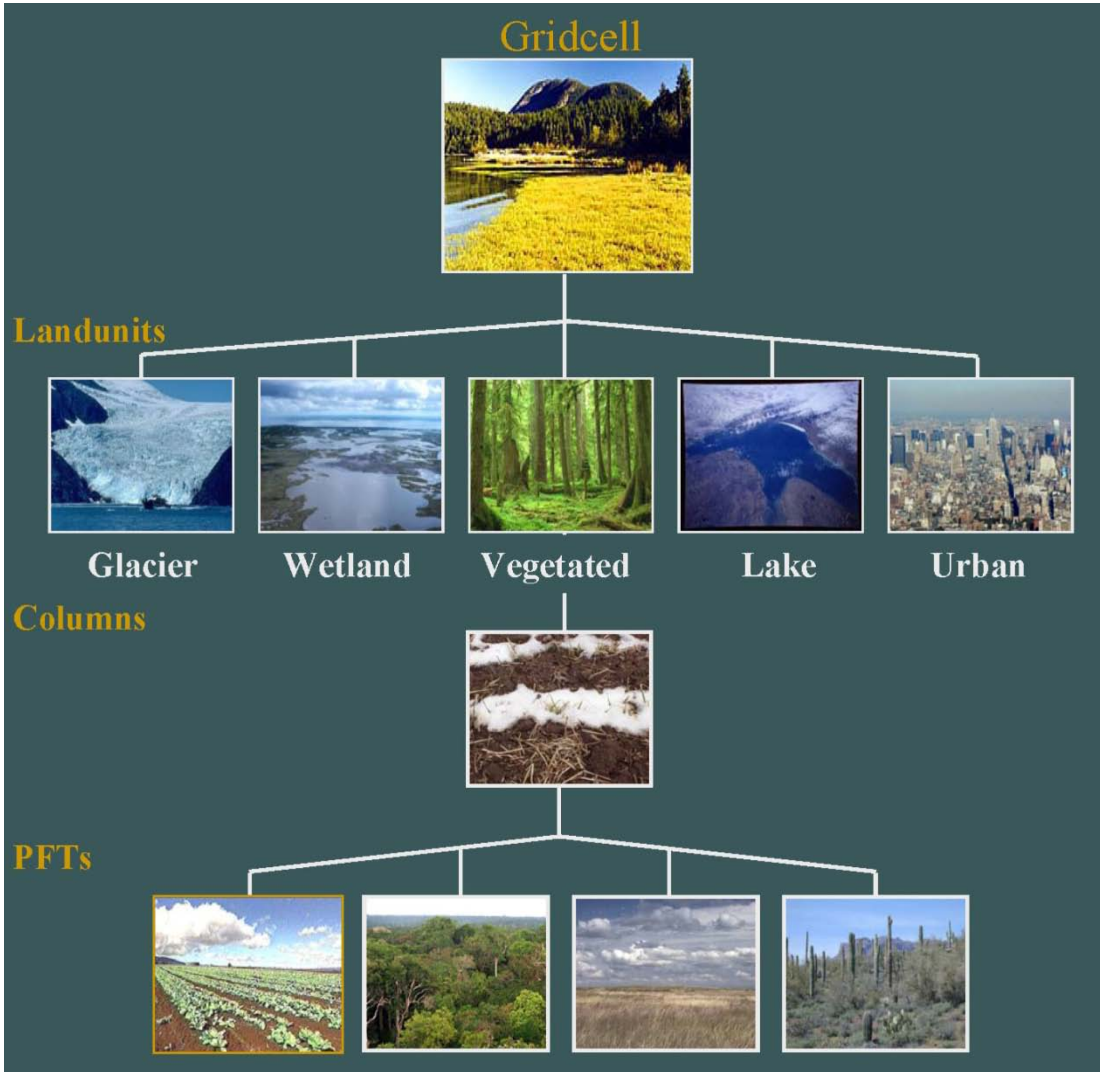
I got these from Dave Lawrence. I don't know if he made them or not – but Thanks to whomever did!



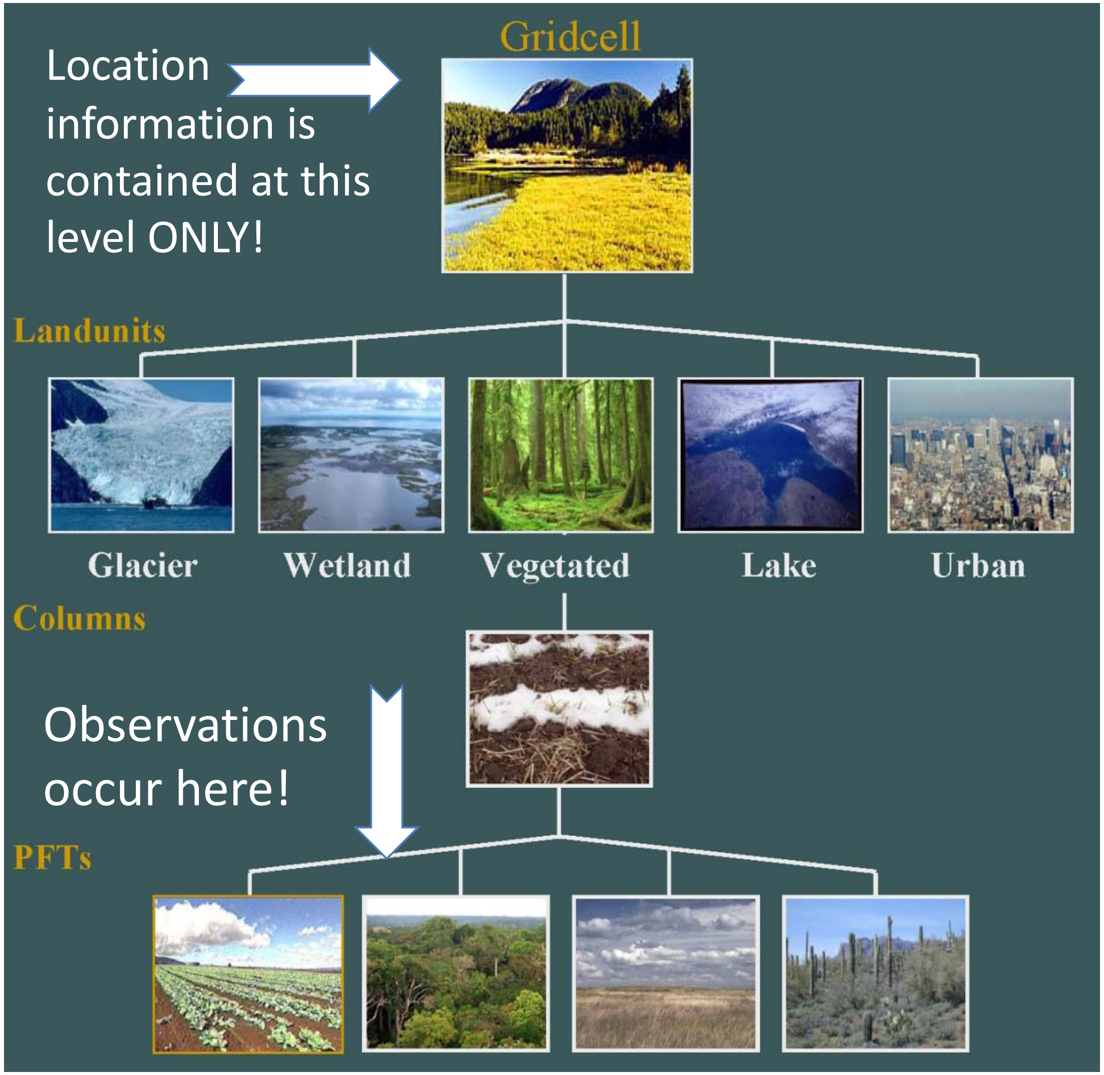
I got these from Dave Lawrence. I don't know if he made them or not – but Thanks to whomever did!

Challenges for Earth System Models (CLM examples)

- Relation between state variables and observations unclear.



CLM abstracts the gridcell into a “nested gridcell hiearchy of of multiple landunits, snow/soil columns, and Plant Function Types”. This is particularly troublesome when trying to convert the model state to the expected observation value ***because***:



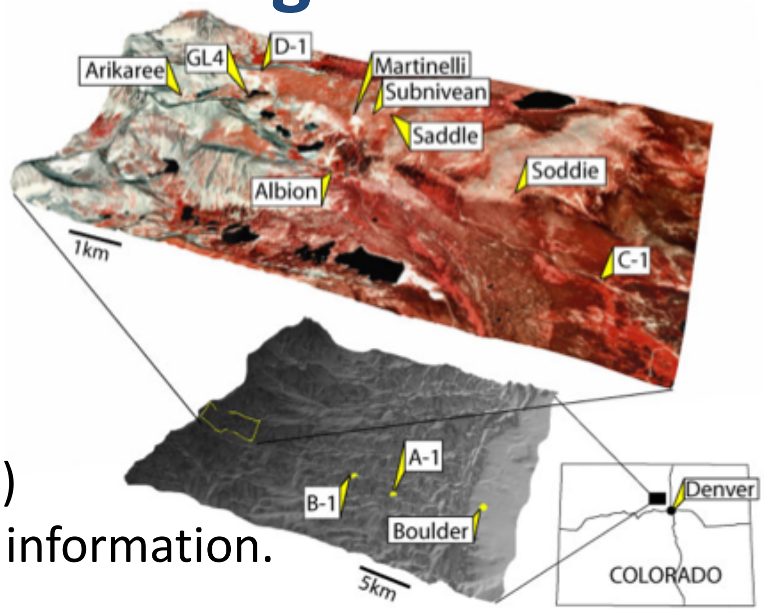
CLM abstracts the gridcell into a “nested gridcell hiearchy of of multiple landunits, snow/soil columns, and Plant Function Types”. This is particularly troublesome when trying to convert the model state to the expected observation value ***because***:

Given a soil temperature observation at a specific lat/lon, which PFT did it come from? **No way to know!** *Unless obs have more metadata!*

Challenges for Earth System Models (CLM examples)

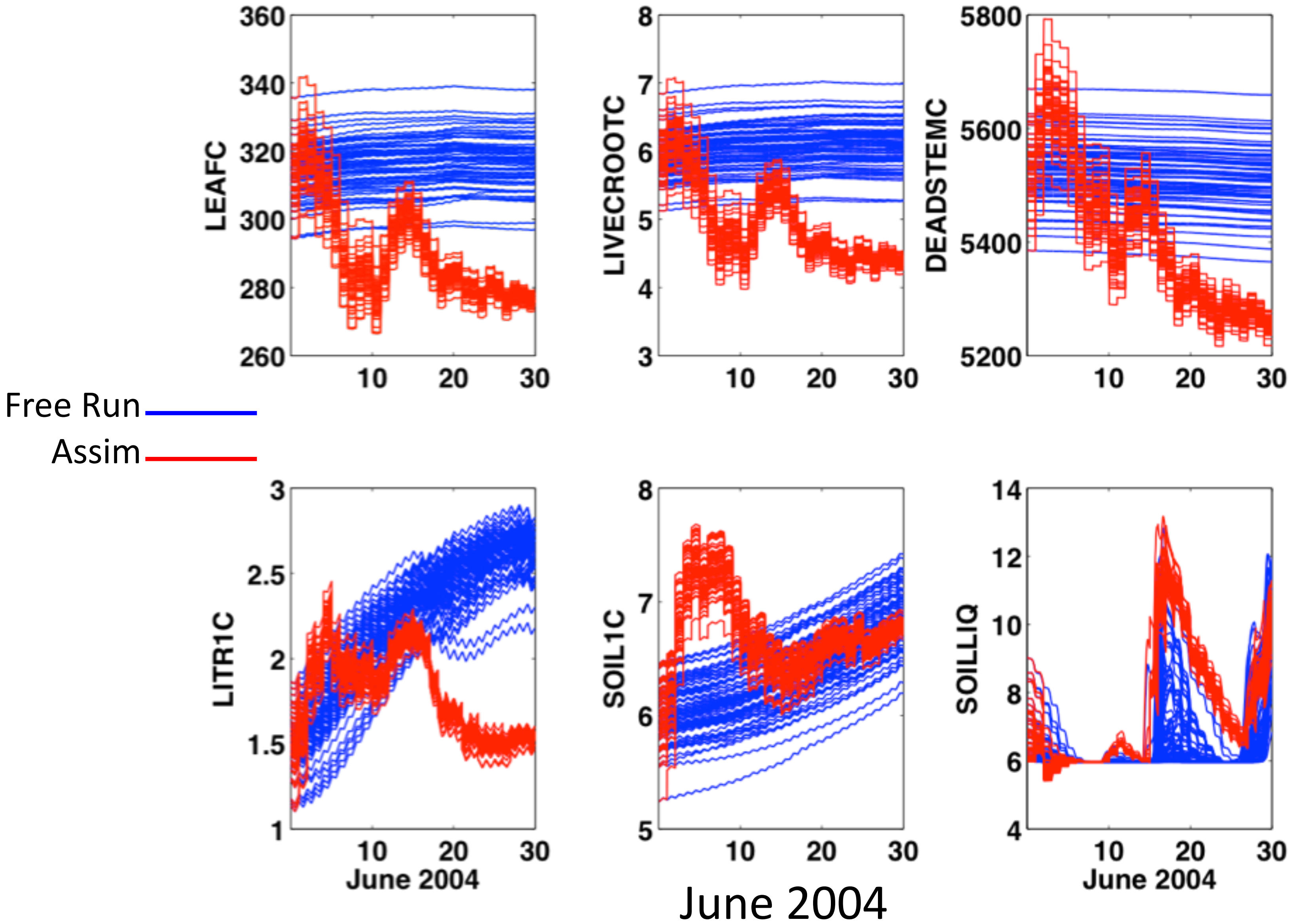
- State variables that are nearly unobserved.

In collaboration with Andy Fox (U. Arizona)
An experiment at Niwot Ridge



- 9.7 km east of the Continental Divide
- C-1 is located in a Subalpine Forest
- (40º 02' 09'' N; 105º 32' 09'' W; 3021 m)
- One column of Community Land Model (CLM)
  - Spun up for 1500 years with site-specific information.
- 64 ensemble members
- Forcing from the DART/CAM reanalysis,
- Assimilating tower fluxes of latent heat (LE), sensible heat (H), and net ecosystem production (NEP).
- Impacts CLM variables: LEAFC, LIVEROOTC, LIVESTEMC, DEADSTEMC, LITR1C, LITR2C, SOIL1C, SOIL2C, SOILLIQ … all of these are *unobserved*.

These are all unobserved variables.



Challenges for Earth System Models (CLM examples)

Additional challenges:
- Model variable definitions with non-Gaussian distributions,
- Creating unobserved things.

Snow example, If prior has no snow, but observations do:
- Must partition snow amongst five layers, normally depends on history of snowfall. Highly non-Gaussian,
- Must assign age, dust content, ice content, … to each layer.

Challenges for Earth System Models (CLM examples)

Additional challenges:

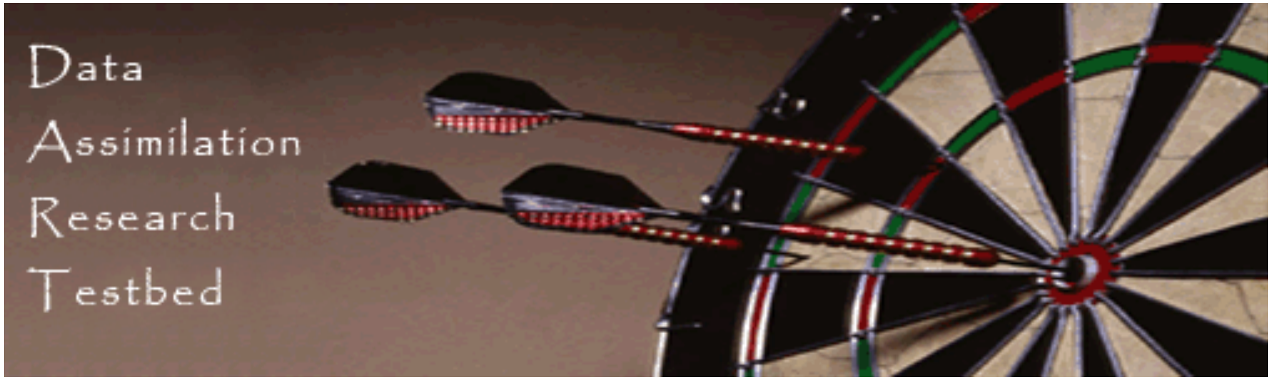
- Observations with poor error characterization,
- Short periods of observations compared to system timescales.

Promising Research Directions

- Just do it. Can get useful results by ignoring the problems.
  Try new models and observations.

- Develop more appropriate models for prediction applications.
  (with modelers)
- Explore value of existing or proposed observations.
  (with observation folks)
- Apply novel techniques for parameter estimation.
  (with statisticians)
- Identify new important quantities that might be predicted.
  (with impacts folks)

Learn more about DART at POSTER 171 TODAY

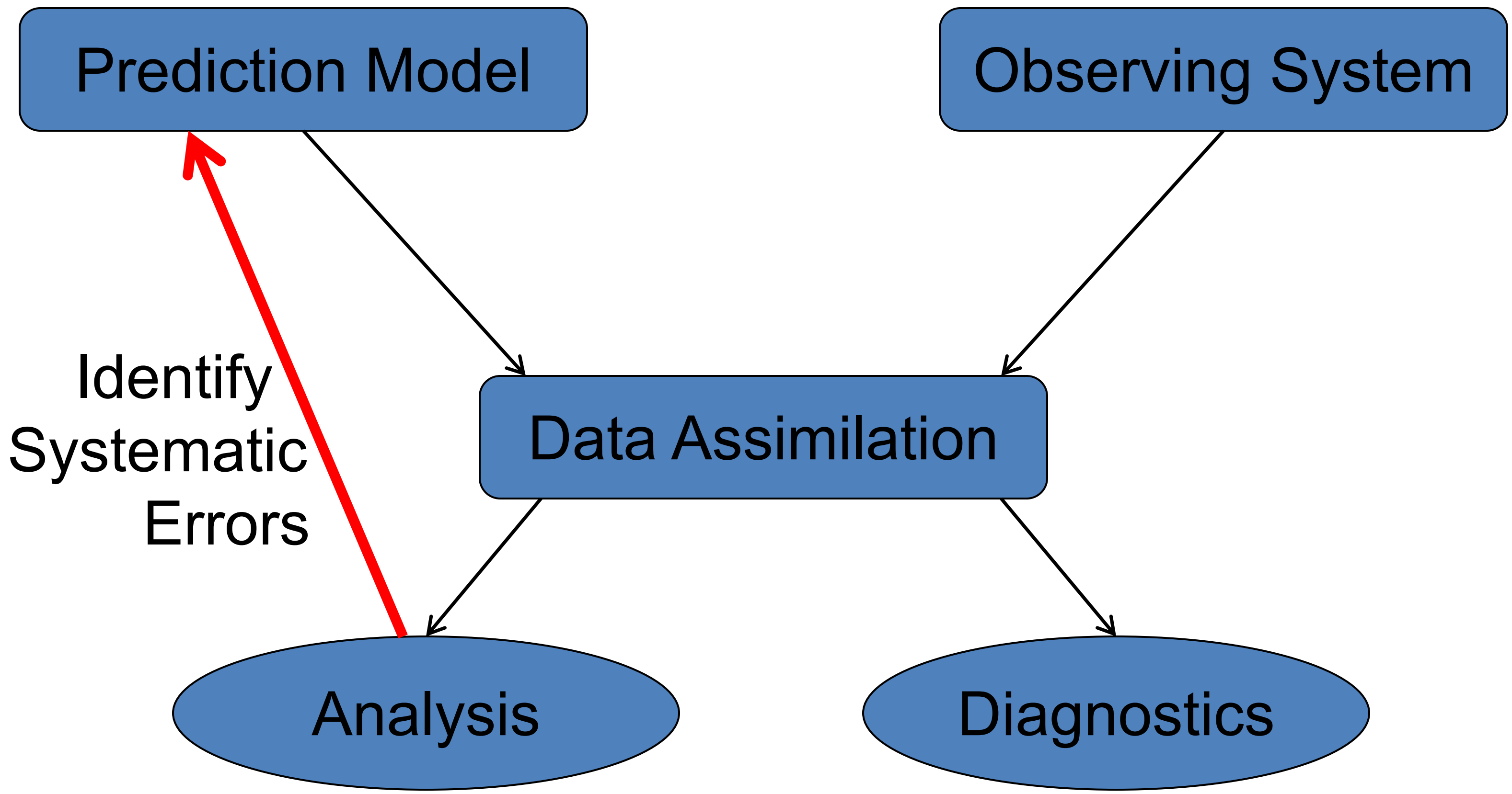


www.image.ucar.edu/DAReS/DART

dart@ucar.edu

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Arellano, A., 2009: *The Data Assimilation Research Testbed: A community facility.* BAMS, **90**, 1283—1296, doi: 10.1175/2009BAMS2618.1

Identifying Model Systematic Errors



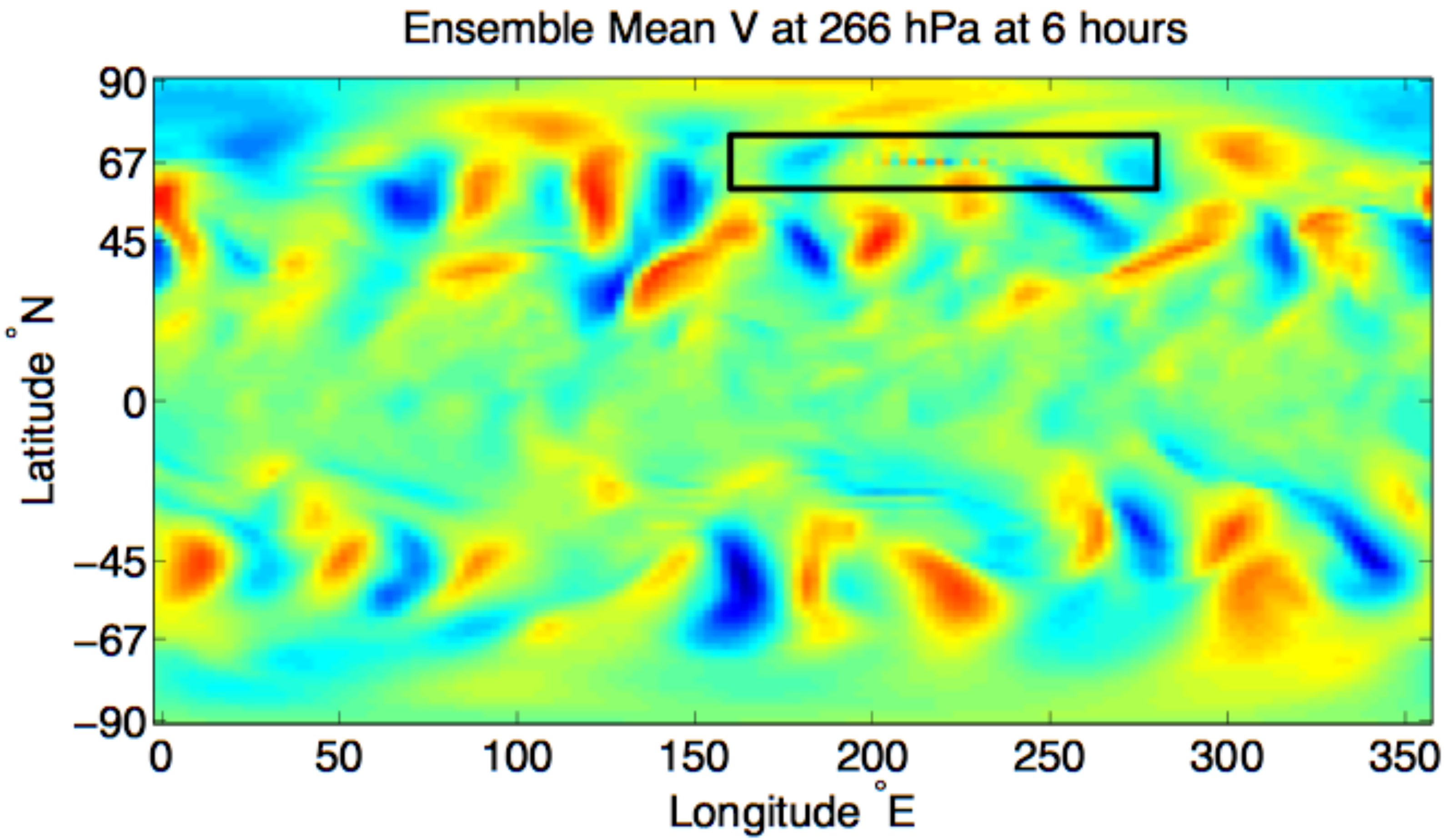
DART Science and Collaborators (4)

Science: Diagnosing and correcting errors in the CAM FV core.

Collaborator: Peter Lauritzen, CGD.

DART Science and Collaborators (4)

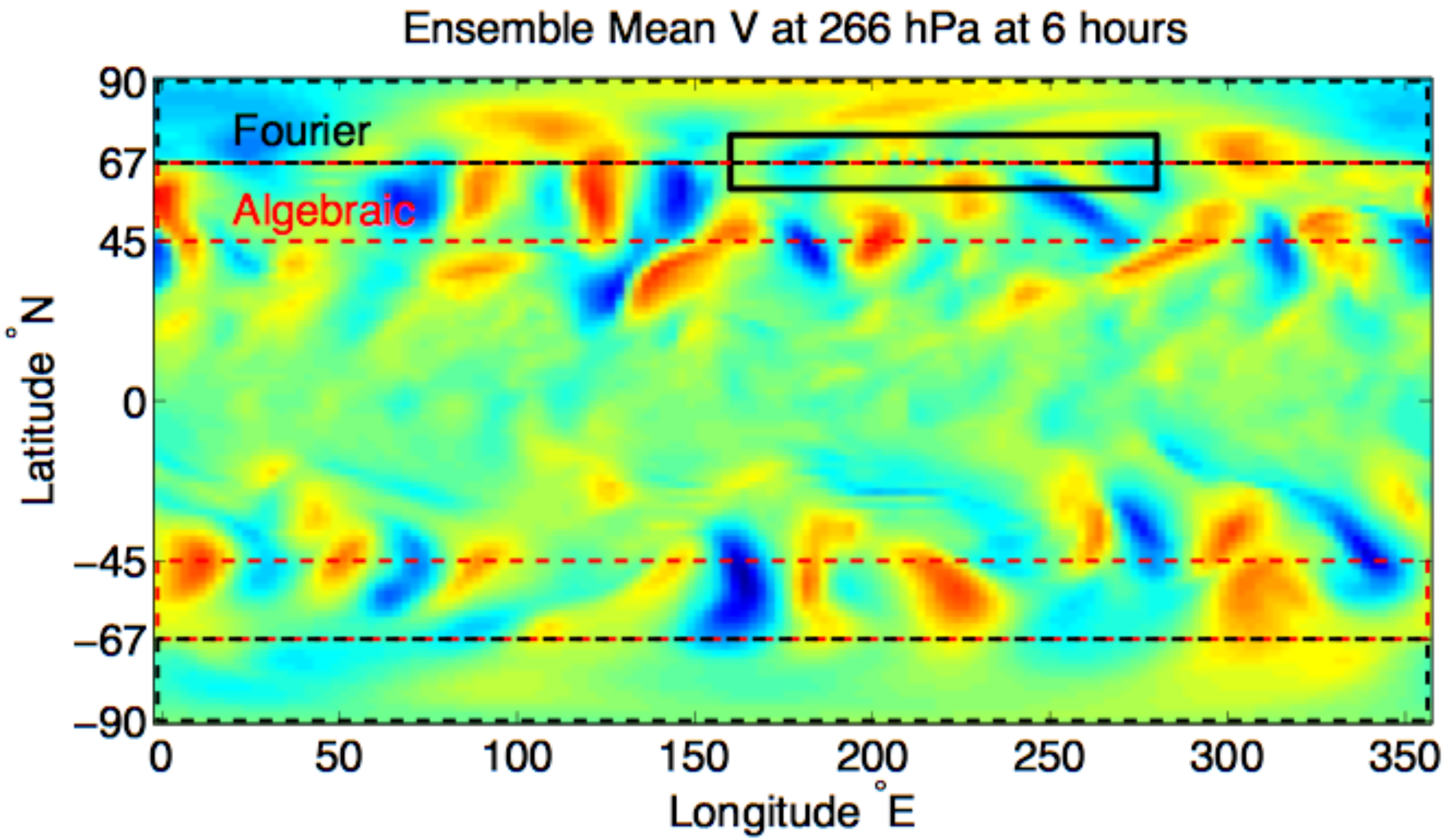
Gridpoint noise detected in CAM/DART analysis



CAM FV core - 80 member mean - 00Z 25 September 2006

DART Science and Collaborators (4)

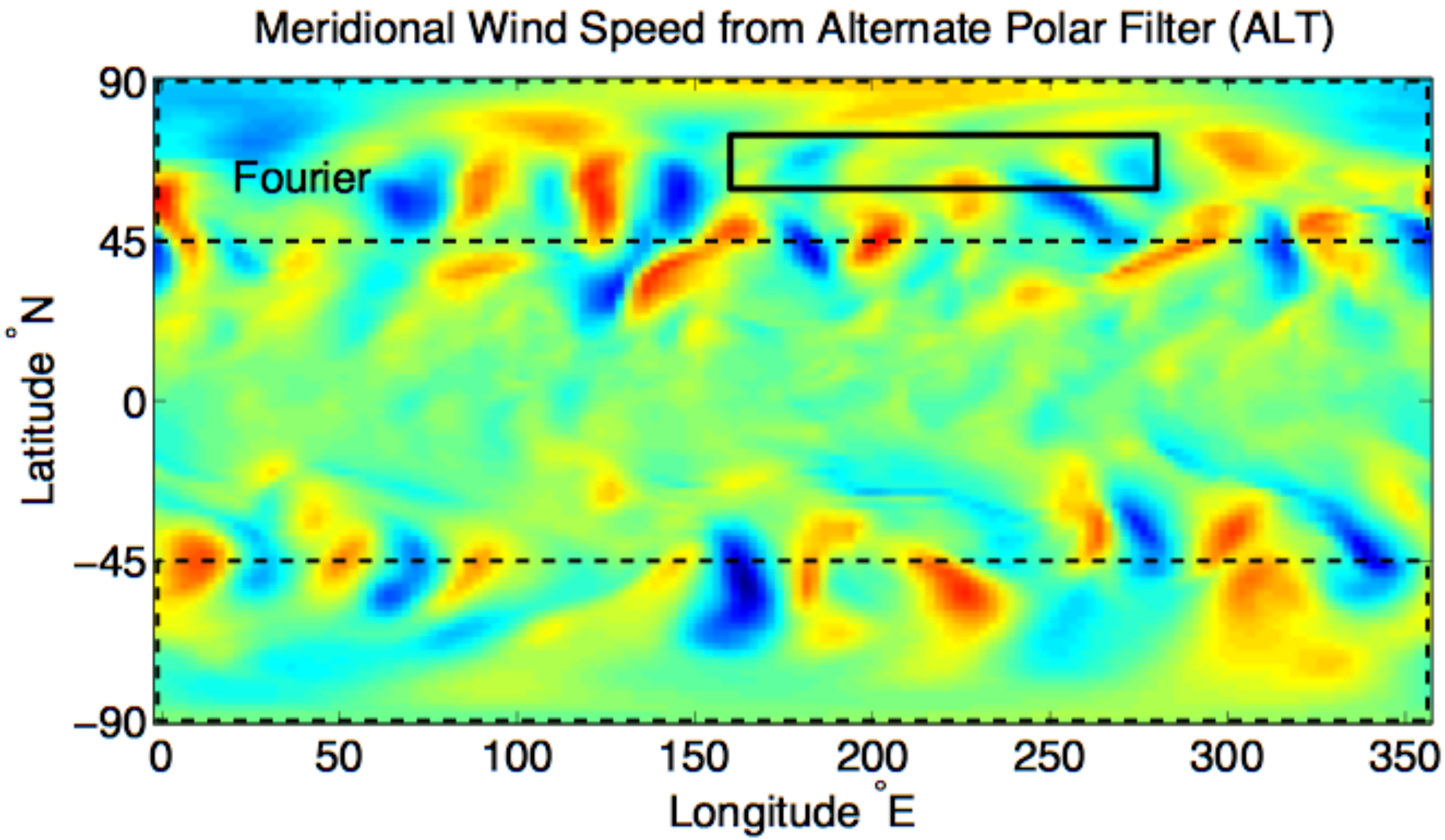
Suspicions turned to the polar filter (DPF)



CAM FV core - 80 member mean - 00Z 25 September 2006

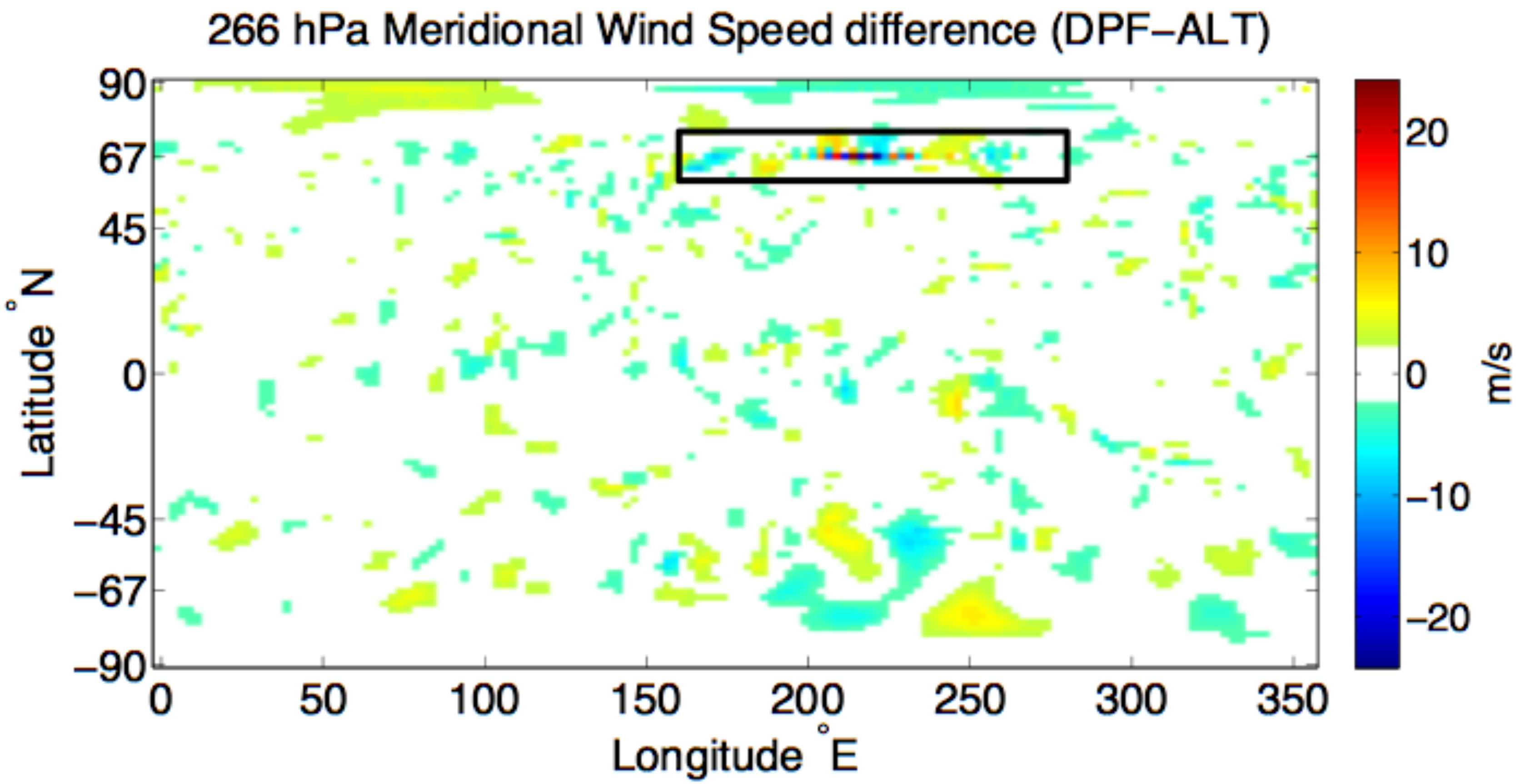
DART Science and Collaborators (4)

Continuous polar filter (alt-pft) eliminated noise.



DART Science and Collaborators (4)

Differences mostly in transition region of default filter.



DART Science and Collaborators (4)

- The use of DART diagnosed a problem that had been unrecognized (or at least undocumented).
- Could have an important effect on any physics in which meridional mixing is important.
- The problem can be seen in ‘free runs’ - it is not a data assimilation artifact.
- Without assimilation, can’t get reproducing occurrences to diagnose.