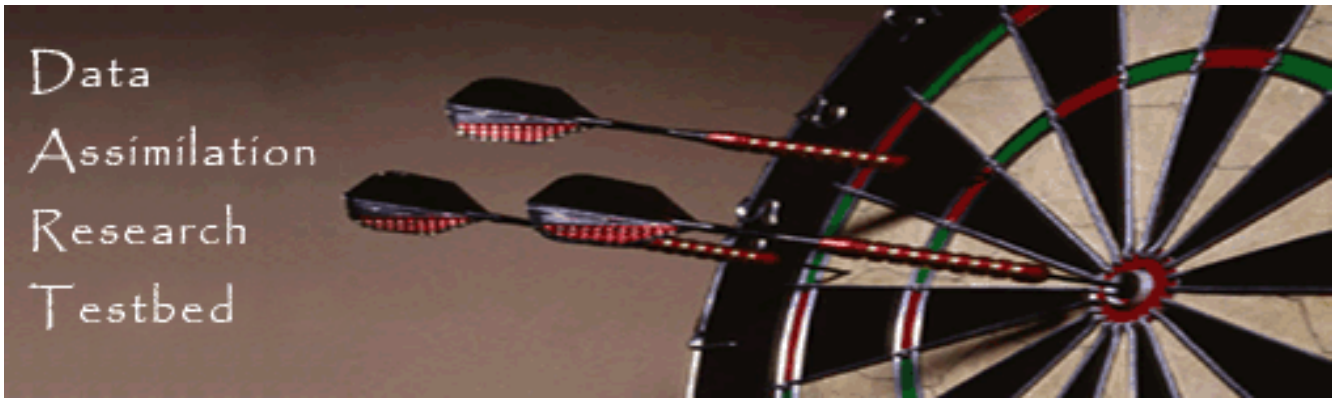
The Data Assimilation Research Testbed: A Community Ensemble DA Facility



Jeffrey Anderson, Nancy Collins, Tim Hoar,
Hui Liu, Glen Romine, Kevin Raeder
NCAR Institute for Math Applied to Geophysics

Data
Assimilation
Research
Testbed

DART is used at:

43 UCAR member universities
More than 100 other sites

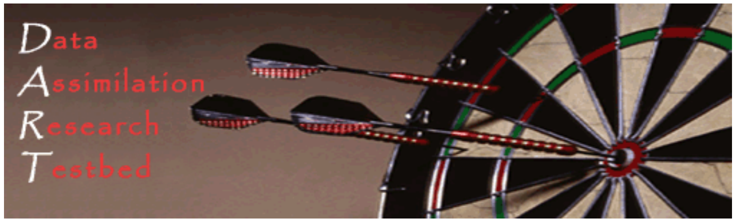
Public domain software for ensemble Data Assimilation
- Well-tested, portable, extensible, free!

Models
- Toy to HUGE

Observations
- Real, synthetic, novel

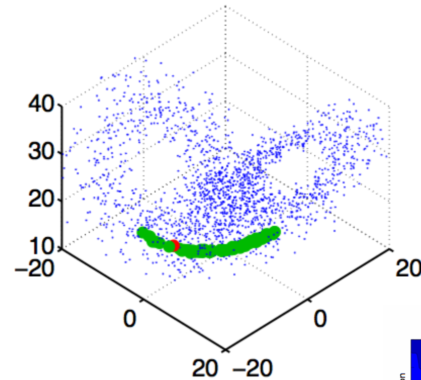
An extensive Tutorial
- With examples, exercises, explanations

People: The DAReS Team

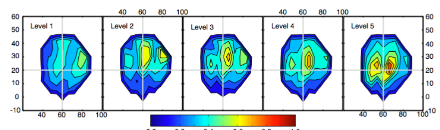


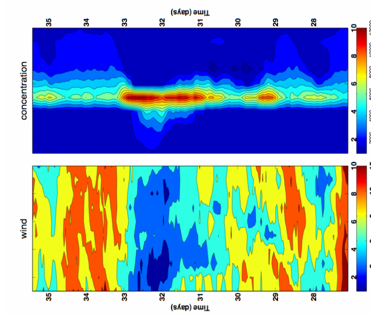
DART is:

Education

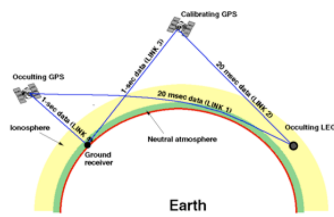


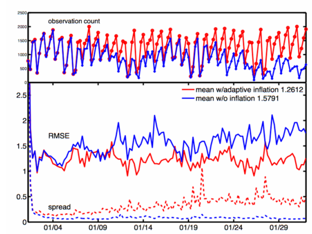
Exploration



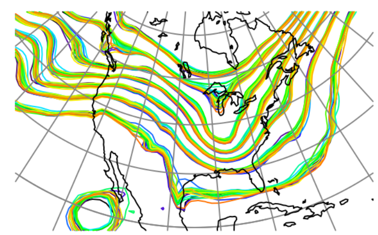


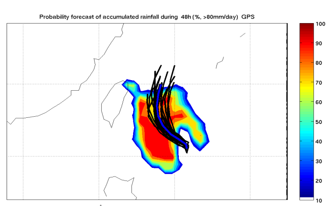
Research





Operations





DART works with many geophysical models

Global Atmosphere models:

CAM	Community Atmosphere Model	NCAR
CAM/CHEM	CAM with Chemistry	NCAR
WACCM	Whole Atmosphere Community Climate Model	NCAR
AM2	Atmosphere Model 2	NOAA/GFDL
NOGAPS	Navy Operational Global Atmospheric Prediction System	US Navy
ECHAM	European Centre Hamburg Model	Hamburg
Planet WRF	Global version of WRF	JPL
MPAS	Model for Prediction Across Scales (under development)	NCAR/DOE

DART works with many geophysical models

Regional Atmosphere models:

WRF/ARW	Weather Research and Forecast Model	NCAR
WRF/CHEM	WRF with Chemistry	NCAR
NCOMMAS	Collaborative Model for Multiscale Atmospheric Simulation	NOAA/NSSL
COAMPS	Coupled Ocean/Atmosphere Mesoscale Prediction System	US Navy
CMAQ	Community Multi-scale Air Quality	EPA

DART works with many geophysical models

Ocean models:

POP	Parallel Ocean Program	DOE/NCAR
MIT OGCM	Ocean General Circulation Model	MIT
ROMS	Regional Ocean Modeling System (under development)	Rutgers
MPAS	Model for Prediction Across Scales (Under development)	DOE/LANL

DART works with many geophysical models

Upper Atmosphere/Space Weather models:

ROSE		NCAR
TieGCM	Thermosphere Ionosphere Electrodynamic GCM	NCAR/HAO
GITM	Global Ionosphere Thermosphere Model	Michigan

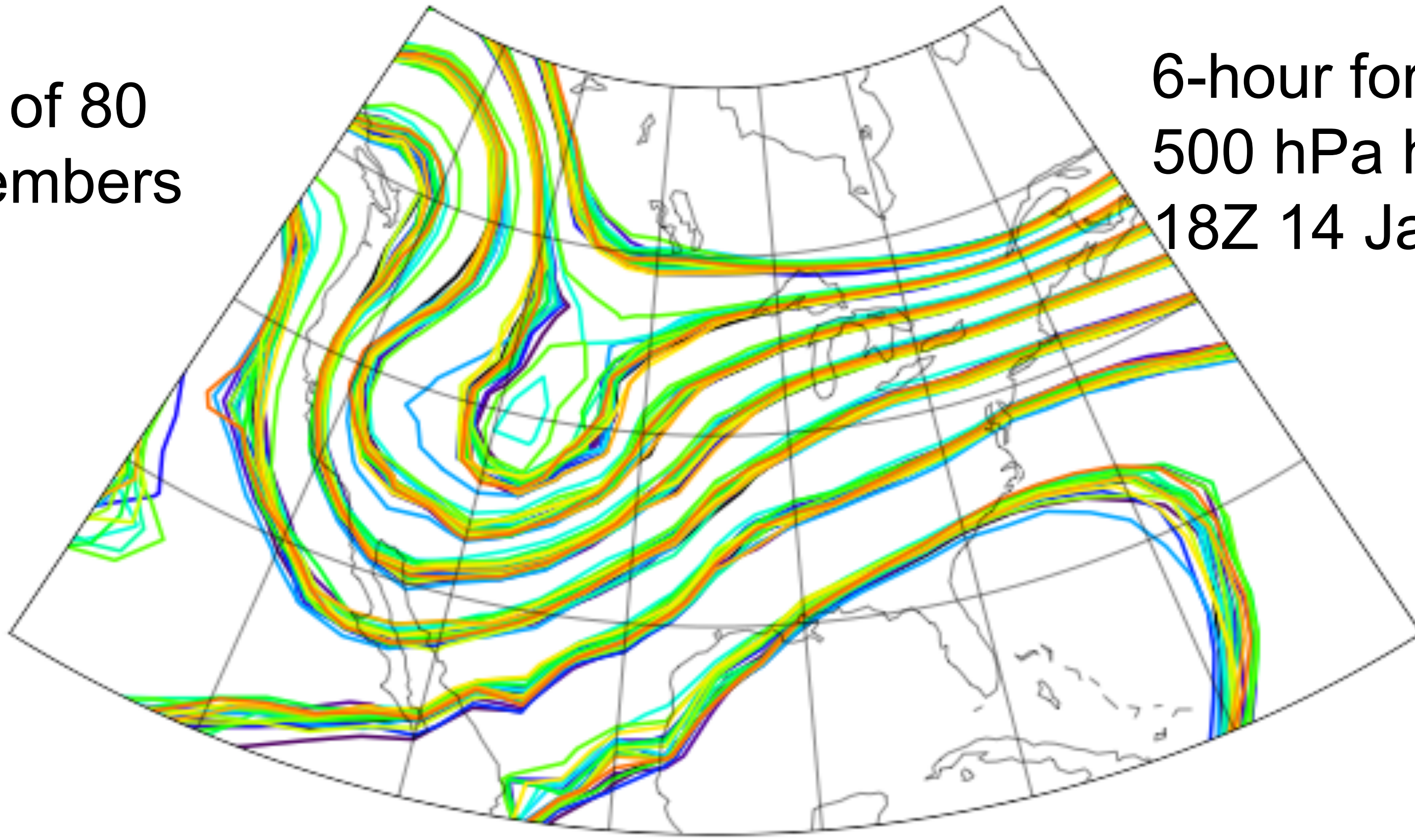
DART works with many geophysical models

Land Surface models:

CLM	Community Land Model	NCAR

Basic Capability: Ensemble Analyses and Forecasts in Large Geophysical Models

20 of 80 members



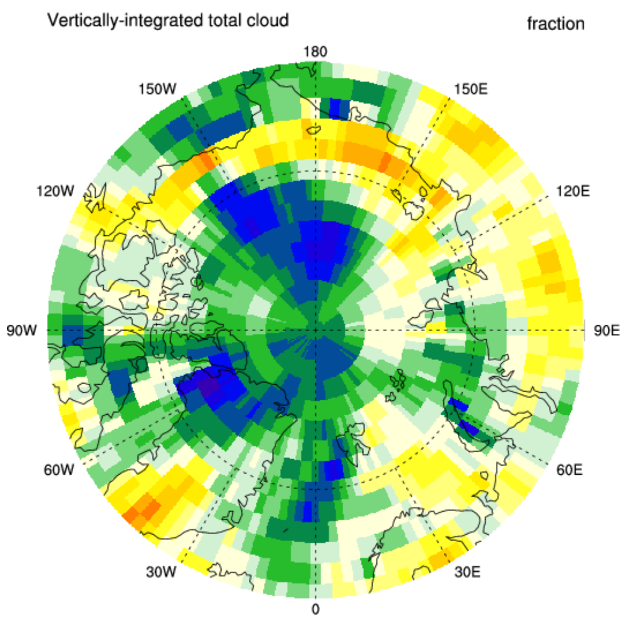
6-hour forecast
500 hPa height
18Z 14 Jan 2007

contours from 5400 to 5880 by 80

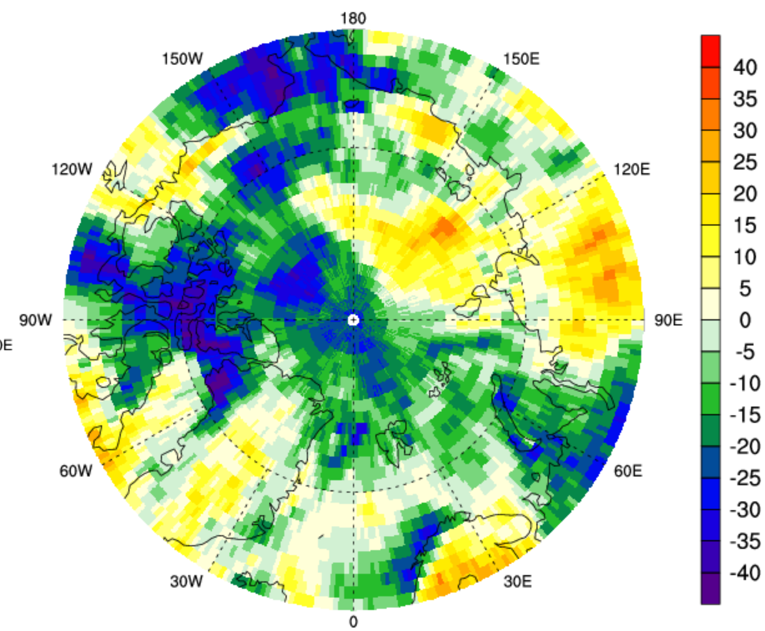
Forecast from CAM (Community Atmosphere Model)

Model improvement by confronting with observations.
(work by Jen Kay, CSU/NCAR)
Modeled vs. observed cloud changes
July 2007 minus July 2006

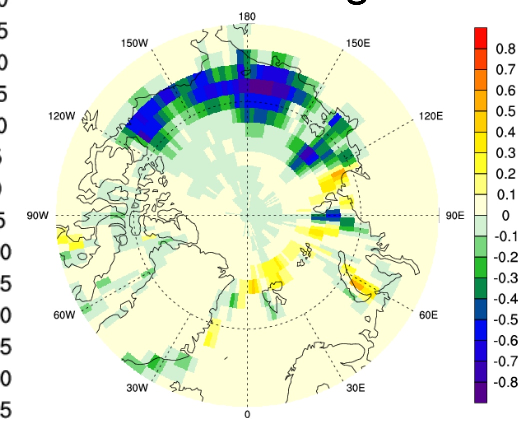
CAM Total Cloud Changes



MODIS Terra Cloud Changes



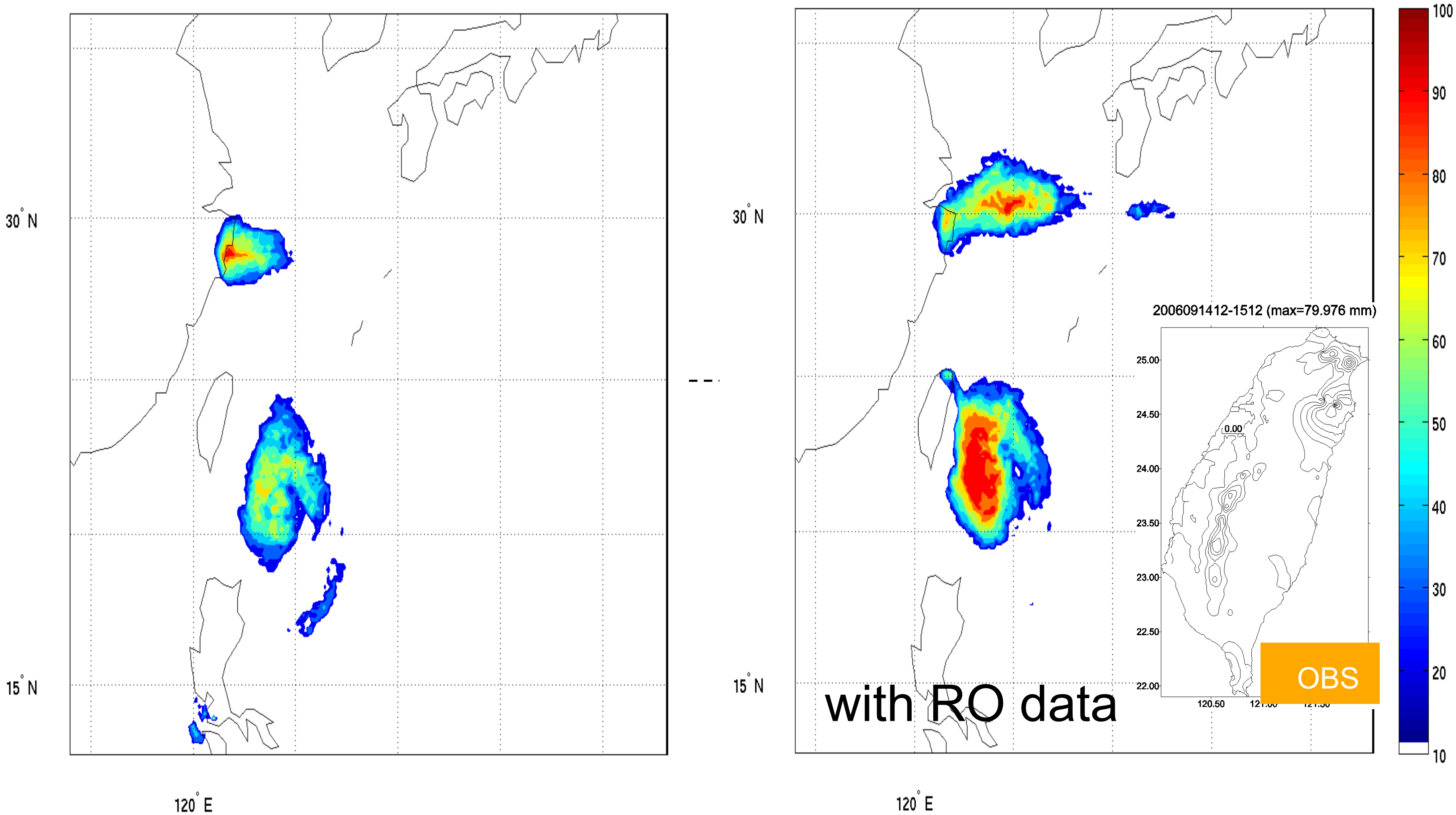
Sea Ice Area Fraction Changes



Unlike CAM, MODIS shows variability in the cloud response over open water.

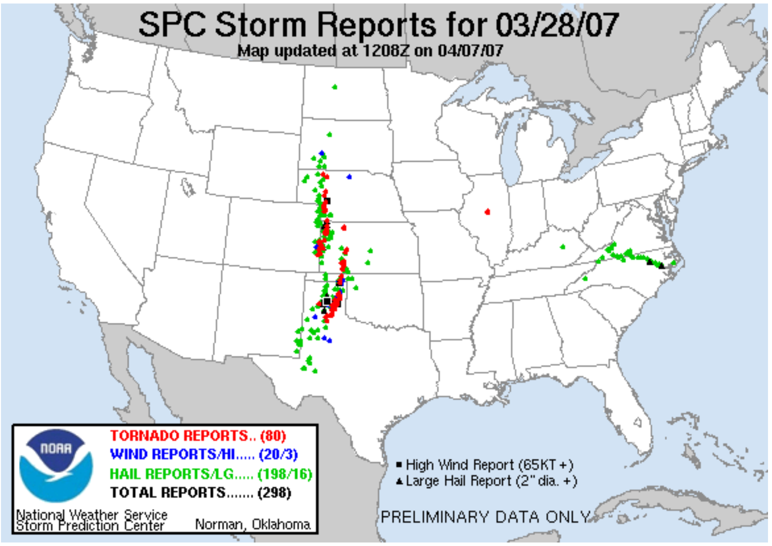
Probabilistic Prediction; Observing System Design

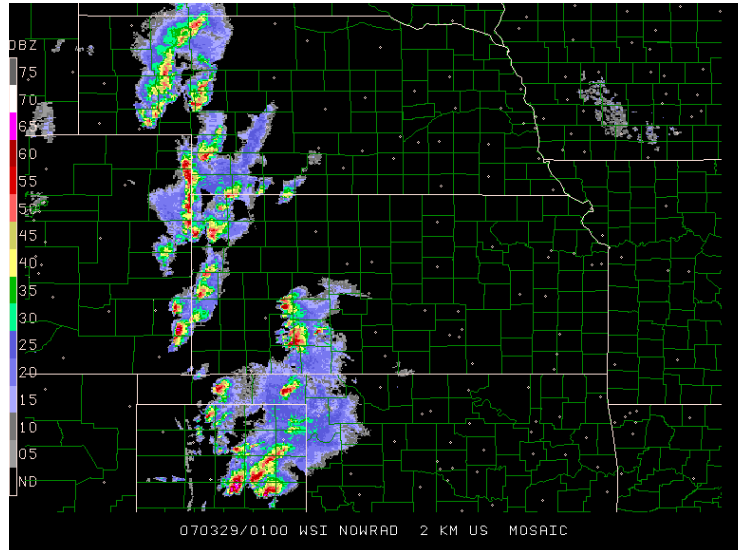
Forecast Probability of Rainfall >60mm/24h, 12Z 14-15 Sep



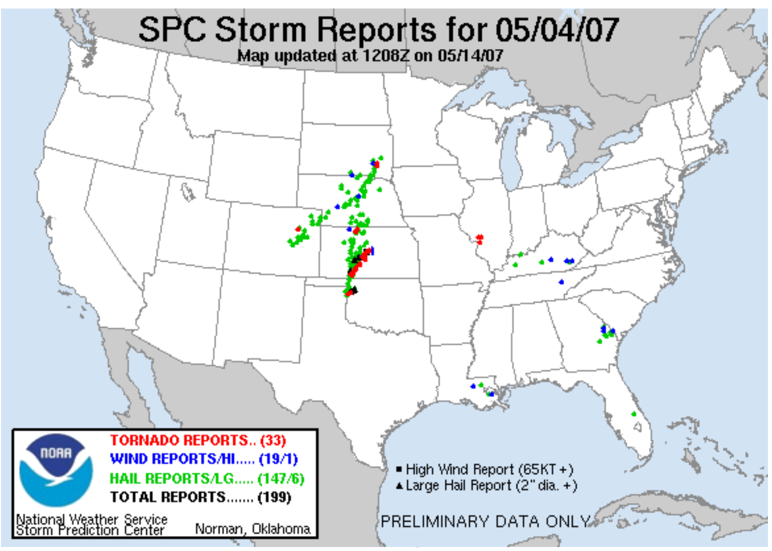
(David Dowell, NOAA)

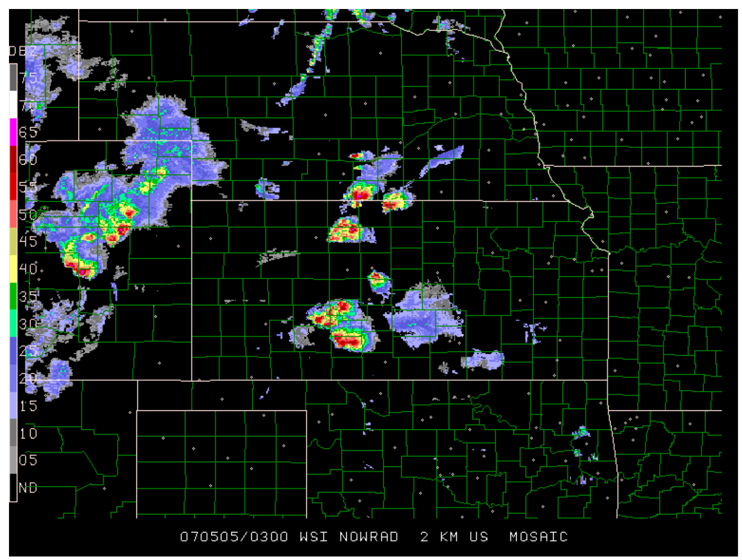
March 28 Tornado Outbreak



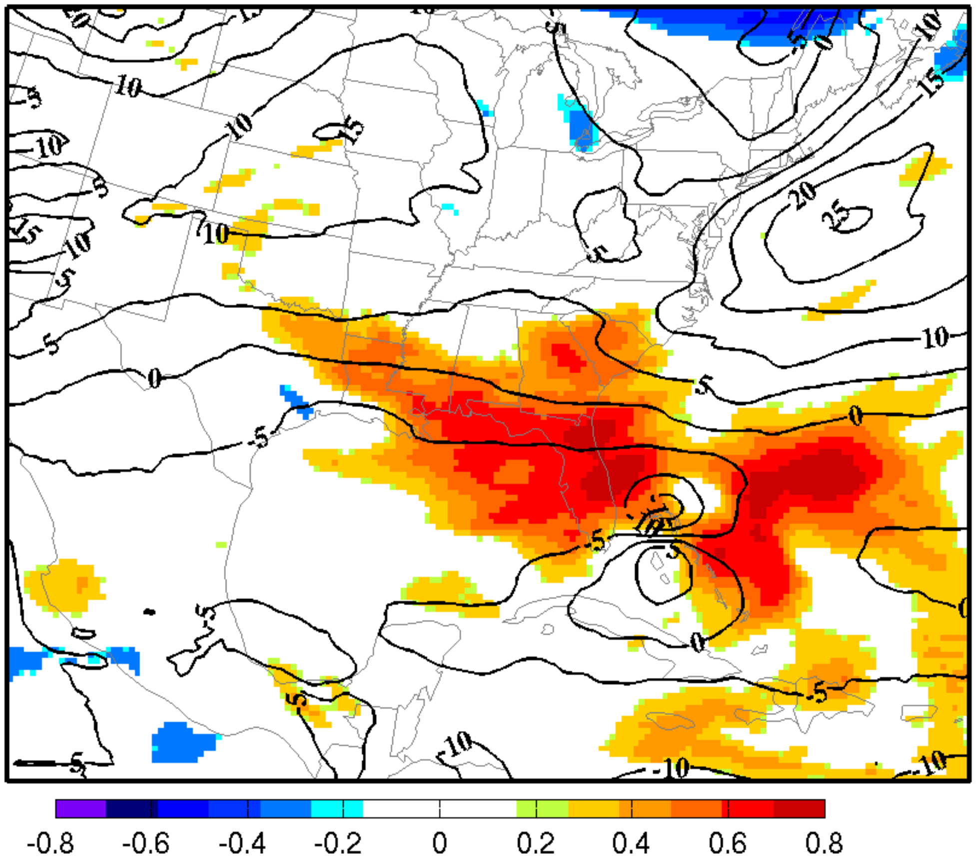


May 4 (Greensburg, KS) Tornado Case





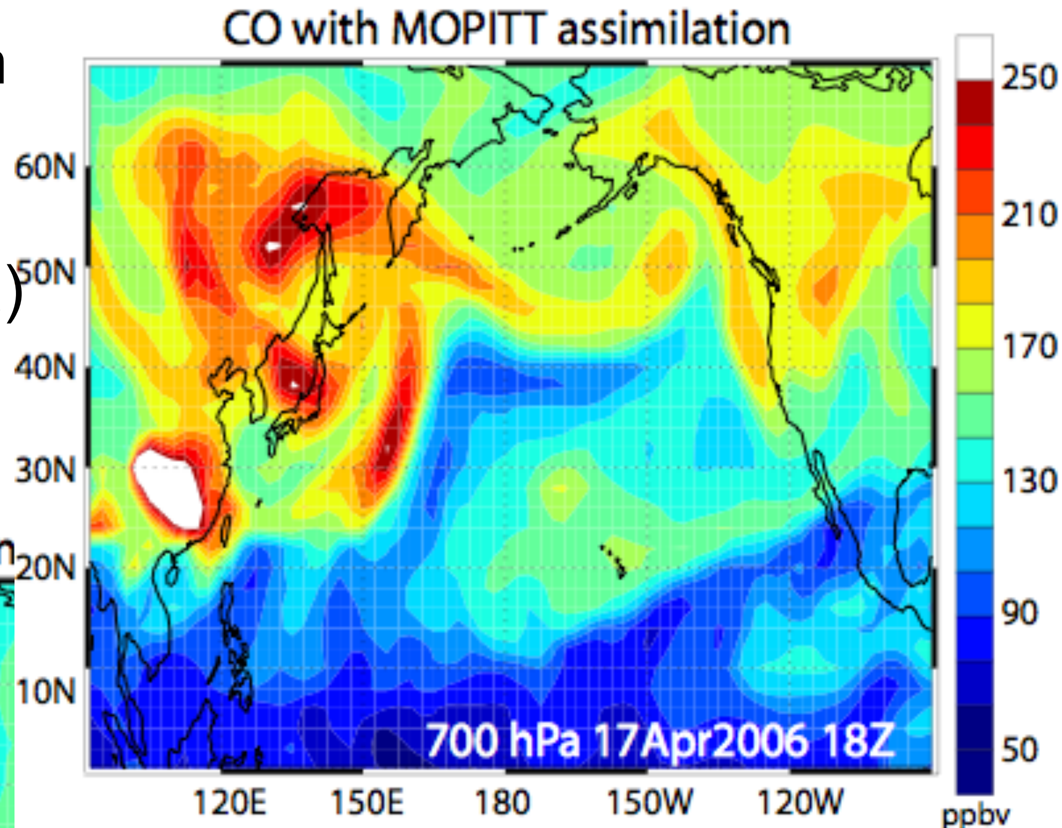
Hurricane Katrina Sensitivity Analysis (Ryan Torn, SUNY Albany)

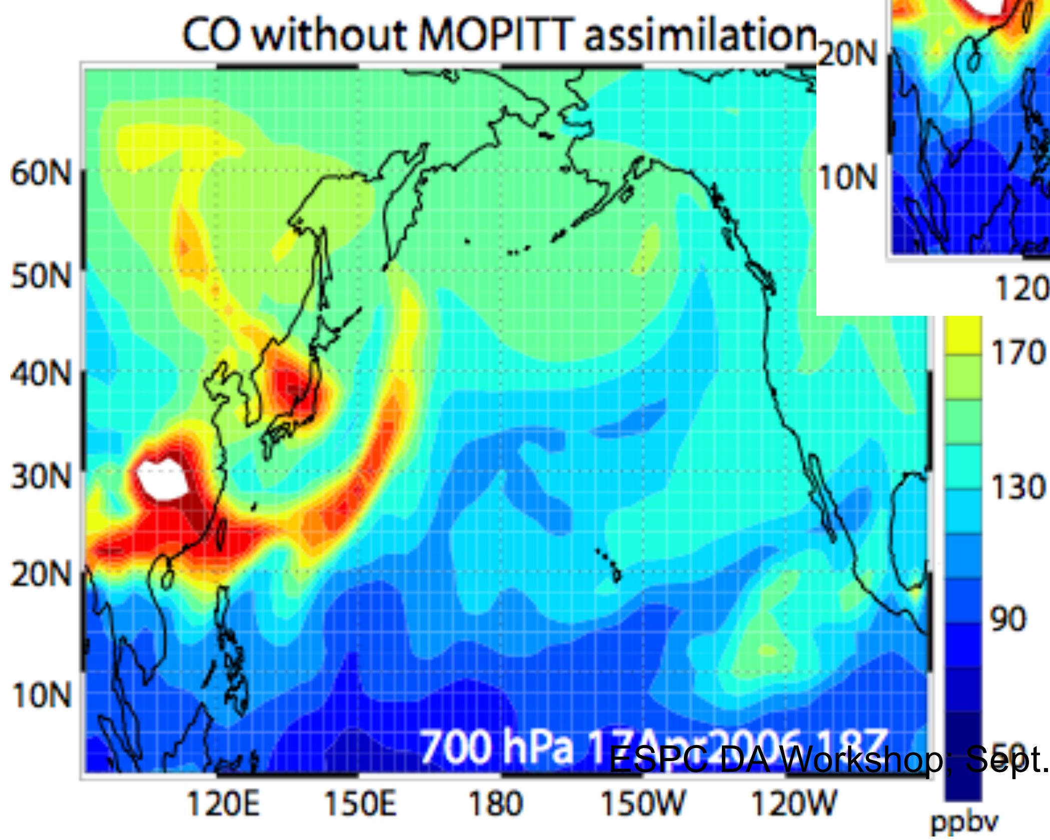


Contours are ensemble mean 48h forecast of deep-layer mean wind.

Color indicates change in the longitude of Katrina.

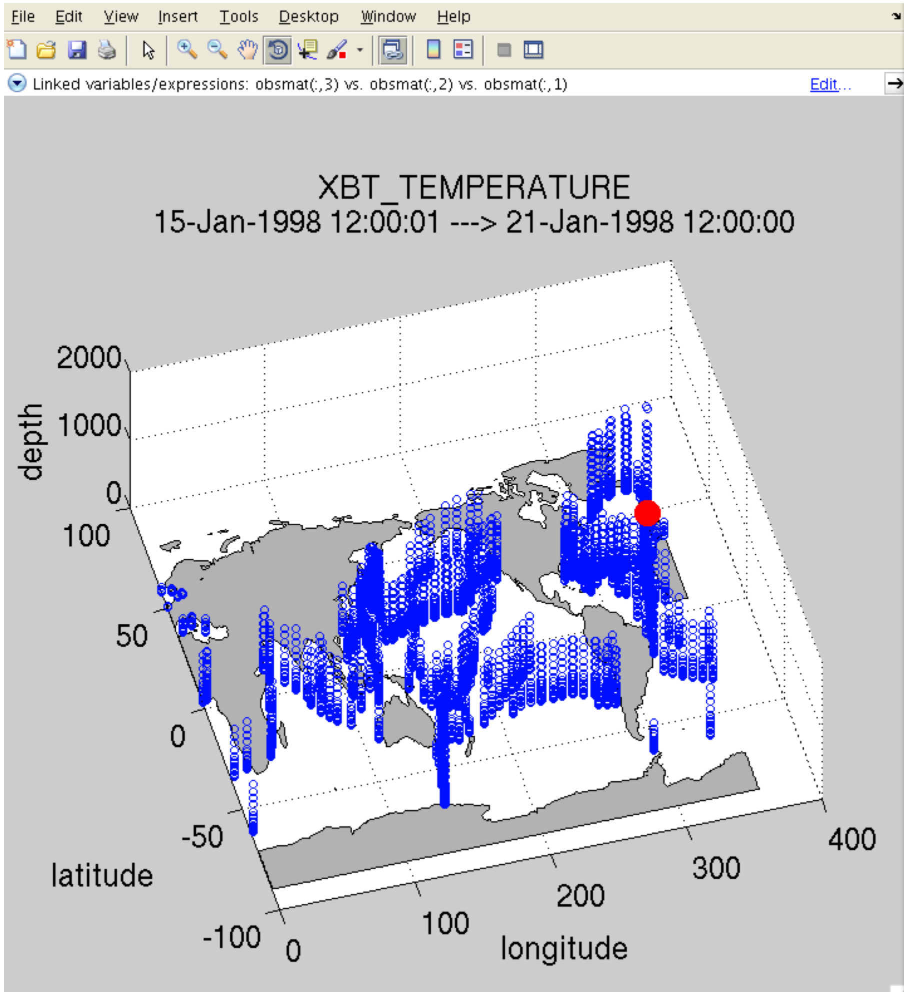
MOPITT CO assimilation prototype
(CAM/CHEM model)
(Ave Arellano, U. Arizona)

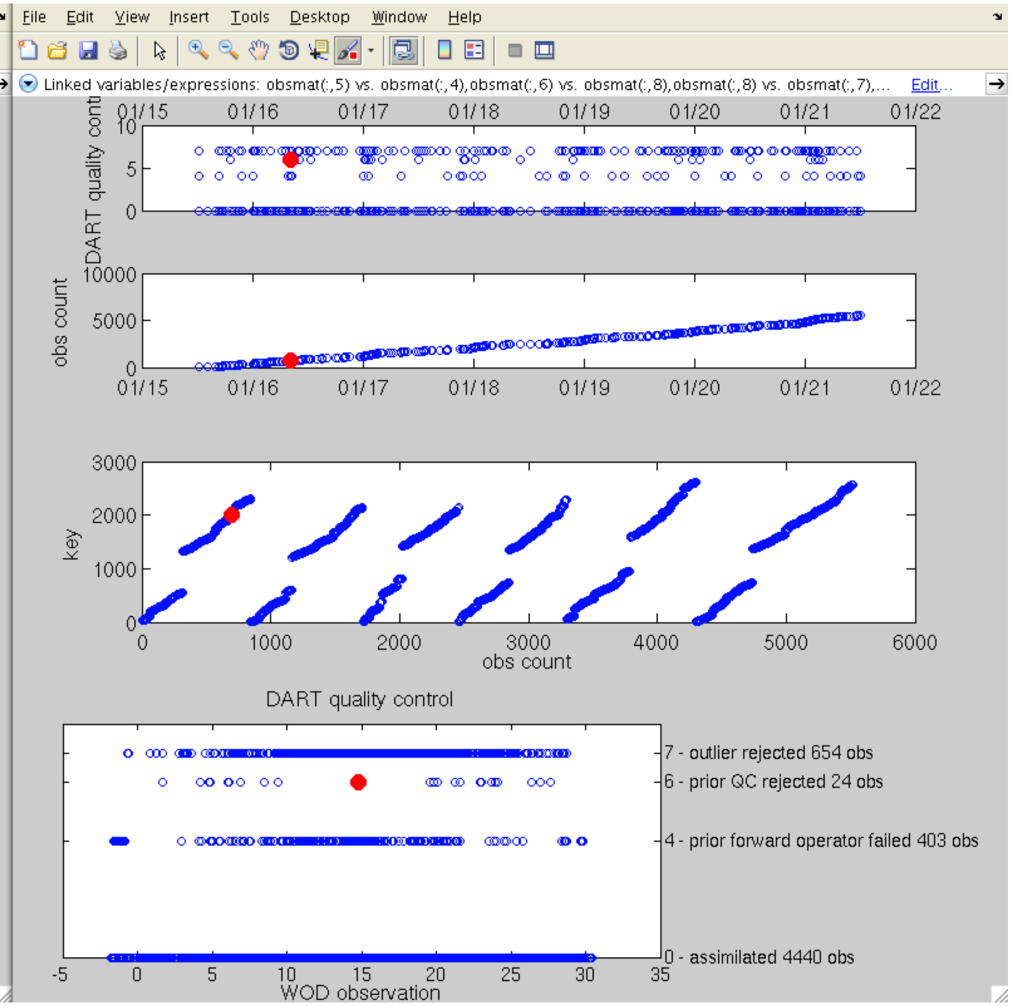




Support for ARCTAS field experiment.

DART Includes Many Diagnostic Tools
Observation Visualization Example





DART Strategy for Generic Ensemble DA

Challenges:

- Only have 4 FTEs plus additional fractional FTE.
- Need to maintain and support existing models and users.
- Add new models, currently about four per year.
- Add new observation types.
- Support users on many different supercomputers.
- Support an assortment of compilers.
- Support new users and students.

DART Strategy for Generic Ensemble DA

Strategies:

- <u>Strict</u> boundaries between DA and models / observation operators.
- Basic implementation leaves forecast model unchanged.
- Interface between DA and models has small set of interfaces.
- Careful coding of tasks that are common to most models.
- Extensive documentation, tutorials and examples.

DART Strategy for Generic Ensemble DA

Parallel performance is key issue:

•Need algorithm that is independent of model grid, other details.
•Scales well for small or large applications.
•Avoids load balancing problems.

Ensemble Filter for Large Geophysical Models

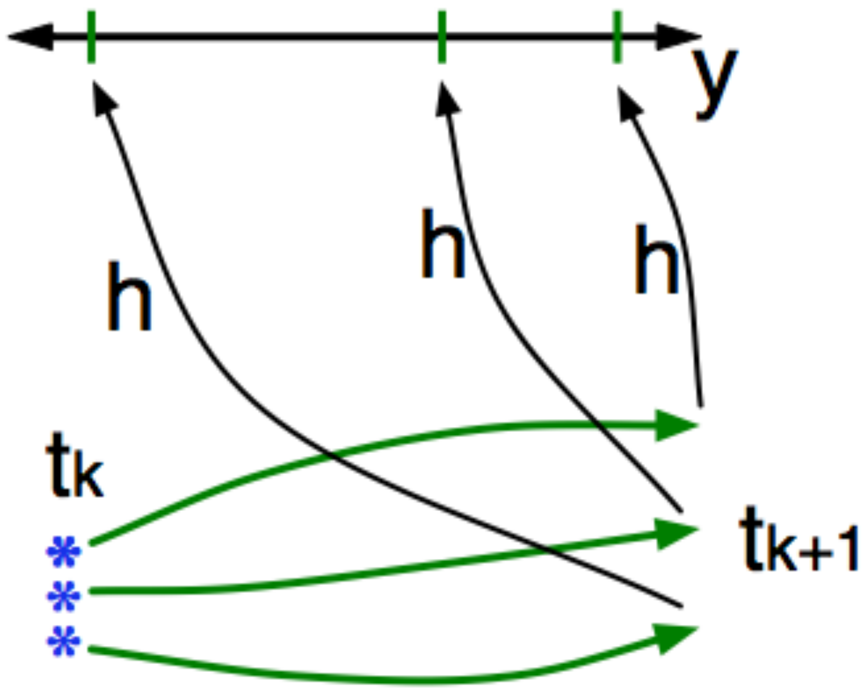
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Ensemble Filter for Large Geophysical Models

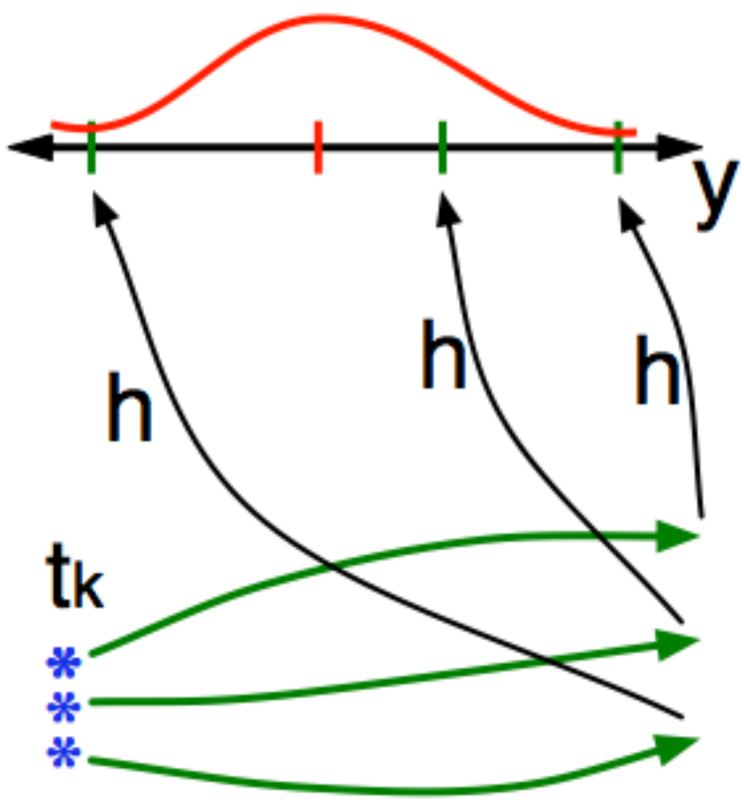
2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.



Theory: observations from instruments with uncorrelated errors can be done sequentially.

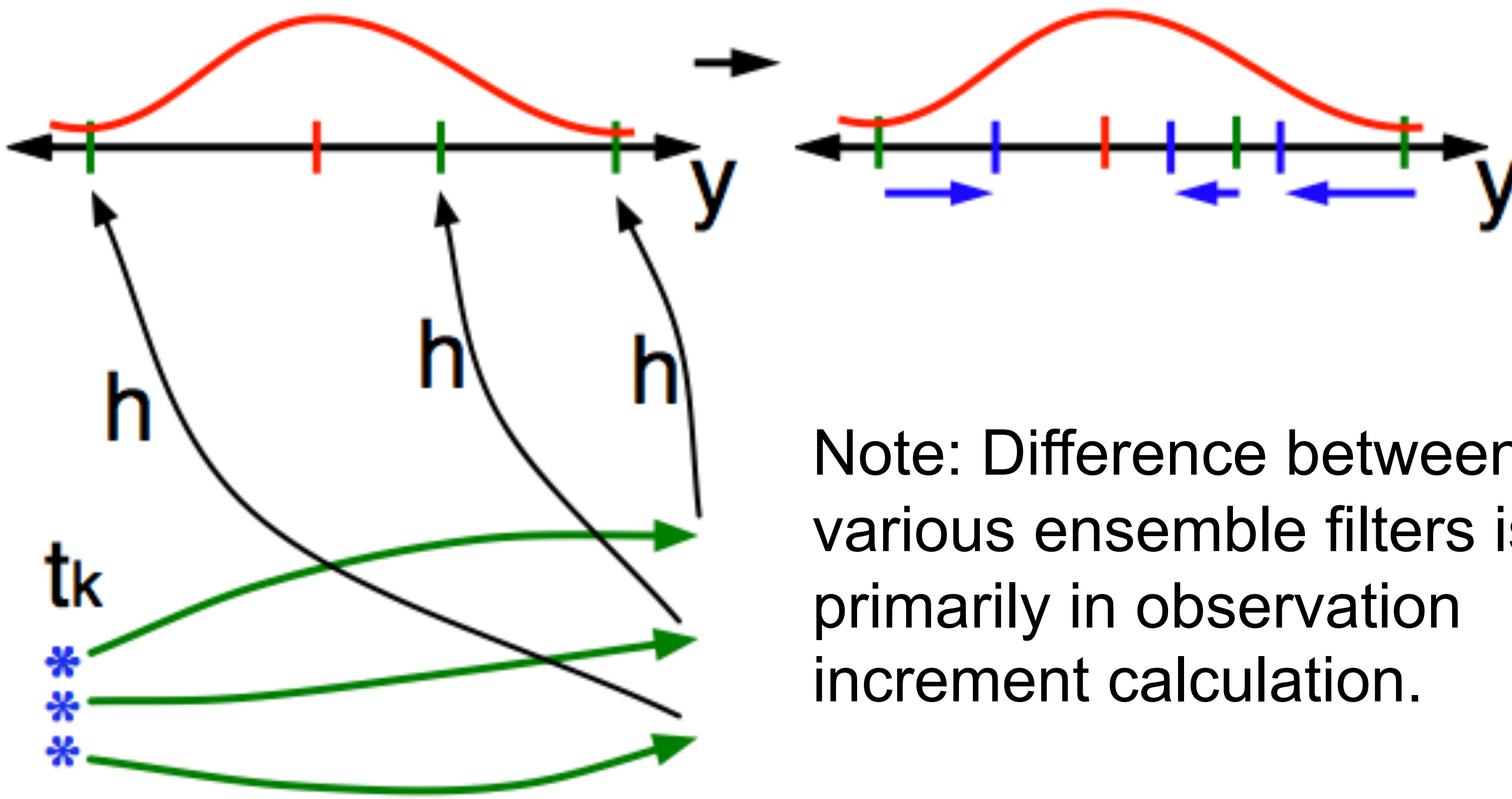
Ensemble Filter for Large Geophysical Models

3. Get observed value and observational error distribution from observing system.



Ensemble Filter for Large Geophysical Models

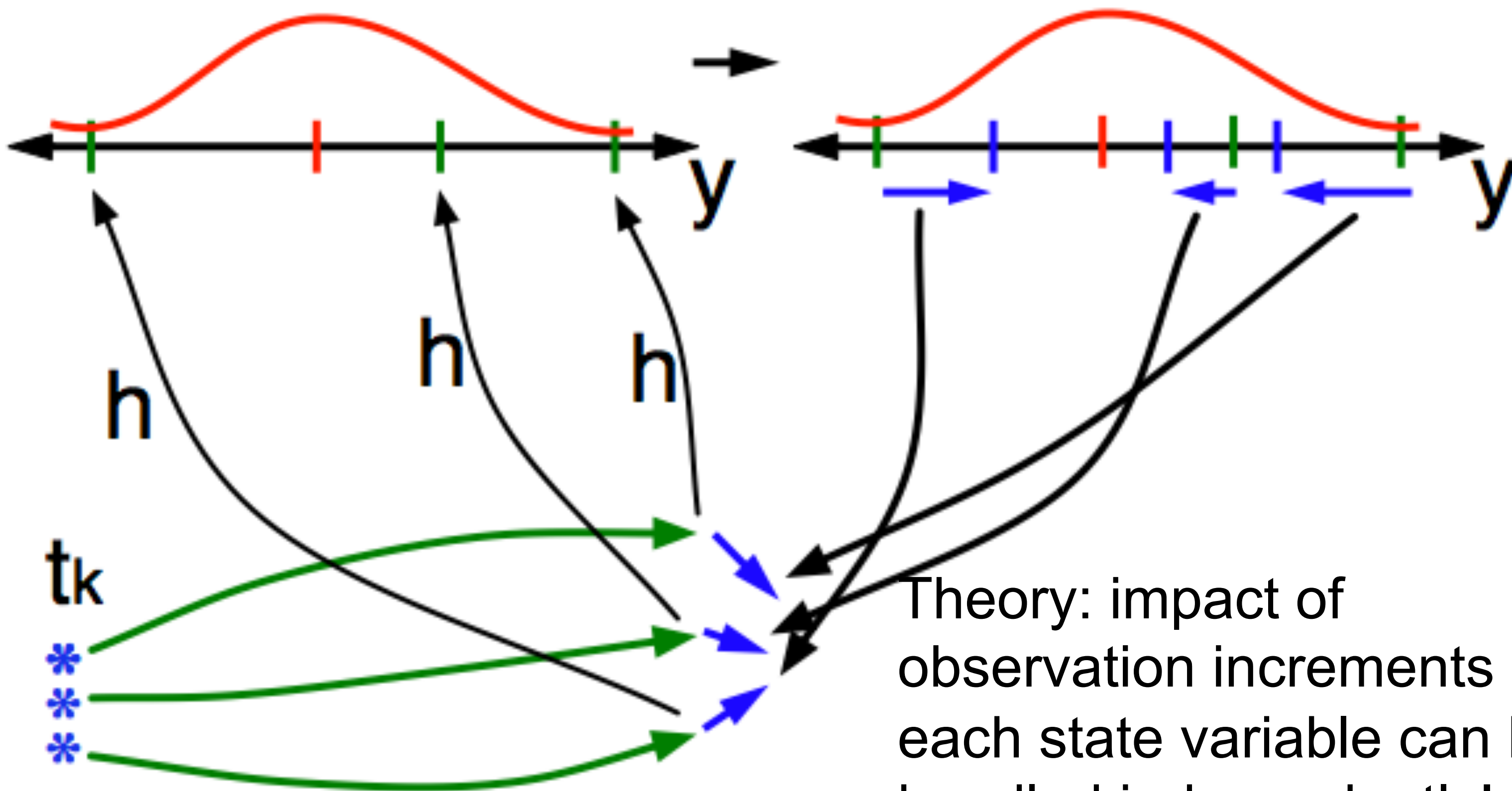
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



Note: Difference between various ensemble filters is primarily in observation increment calculation.

Ensemble Filter for Large Geophysical Models

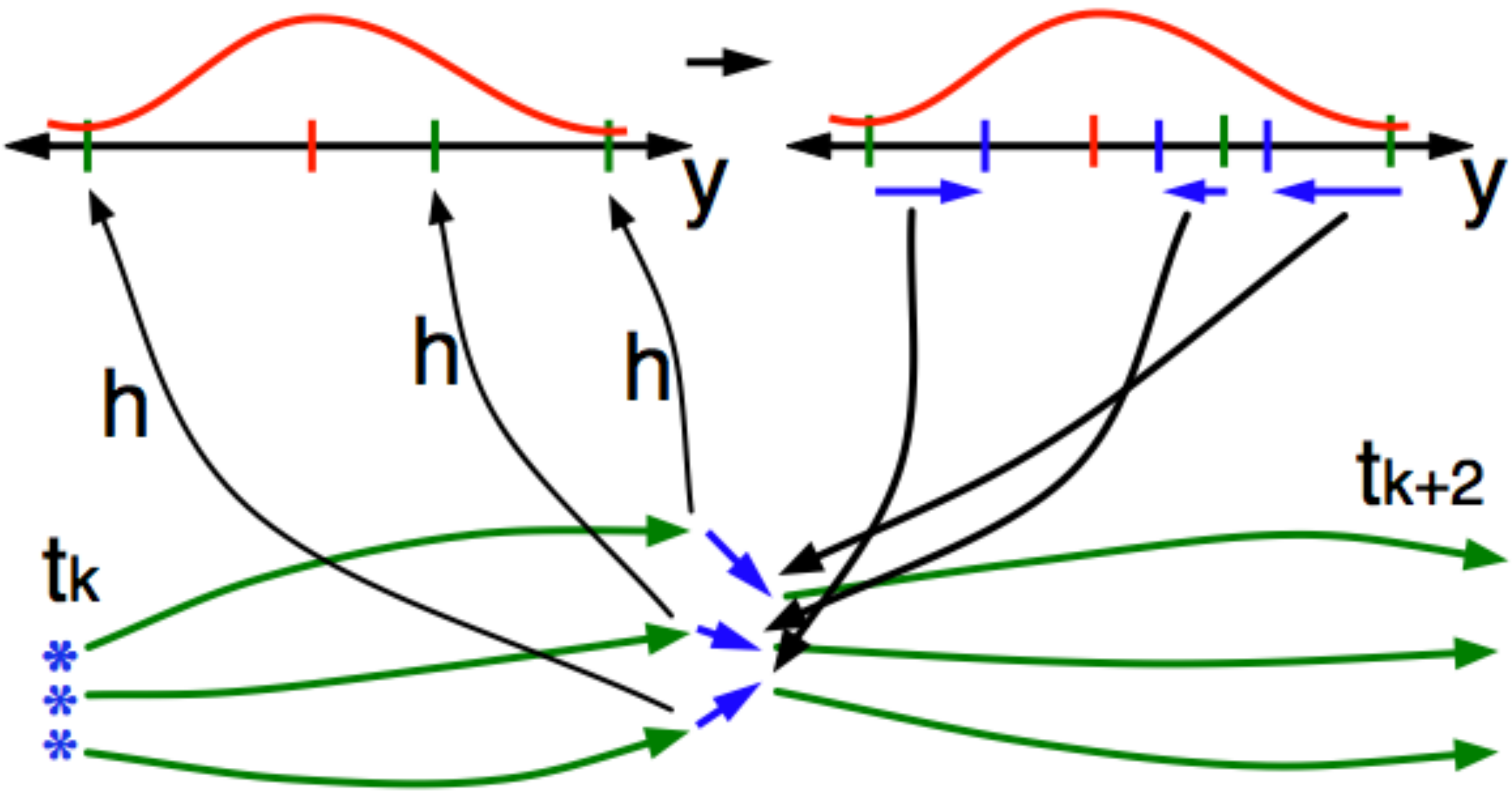
5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



Theory: impact of observation increments on each state variable can be handled independently!

Ensemble Filter for Large Geophysical Models

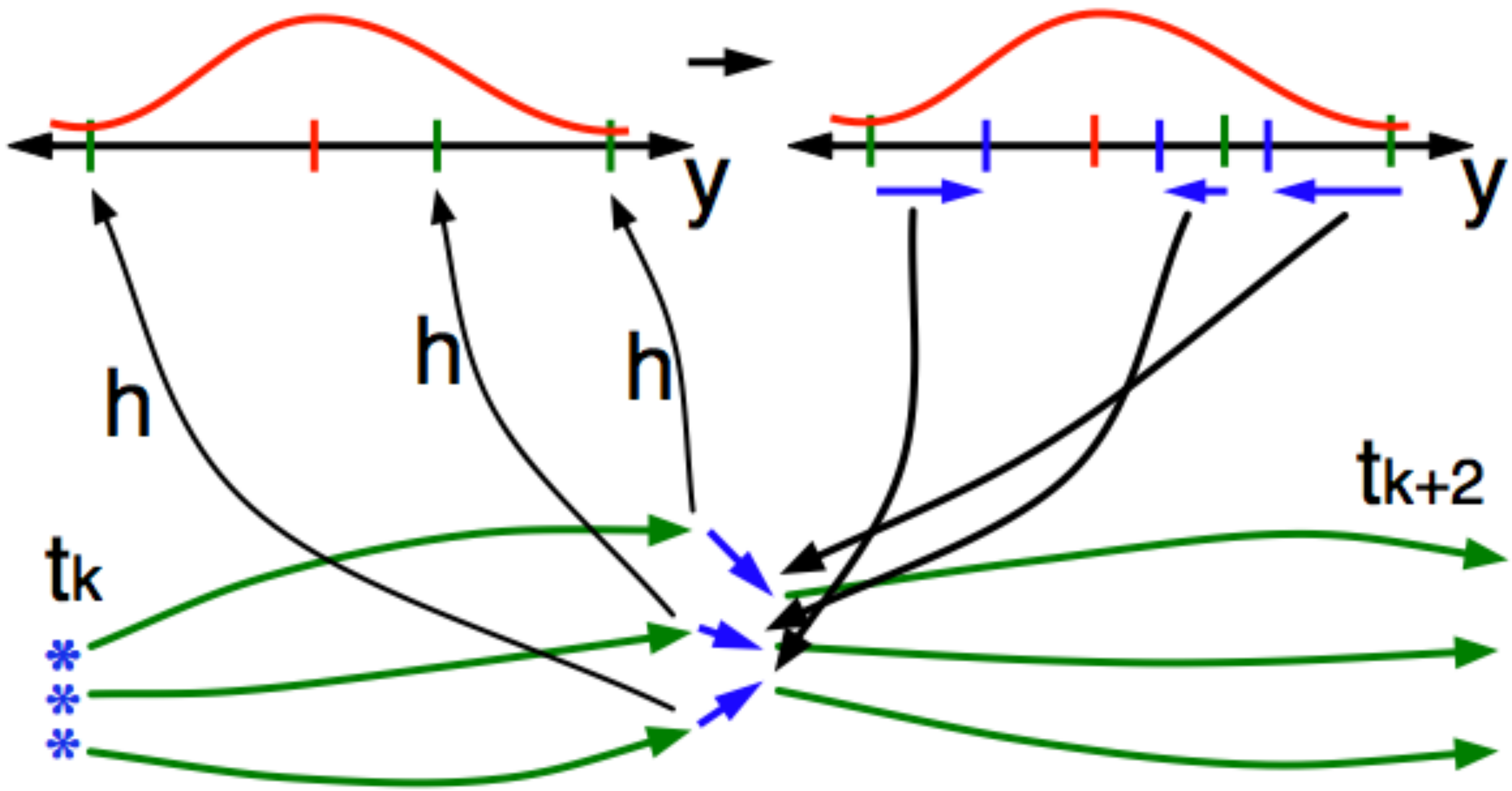
6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation …



Ensemble Filter for Large Geophysical Models

A generic ensemble filter system like DART just needs:

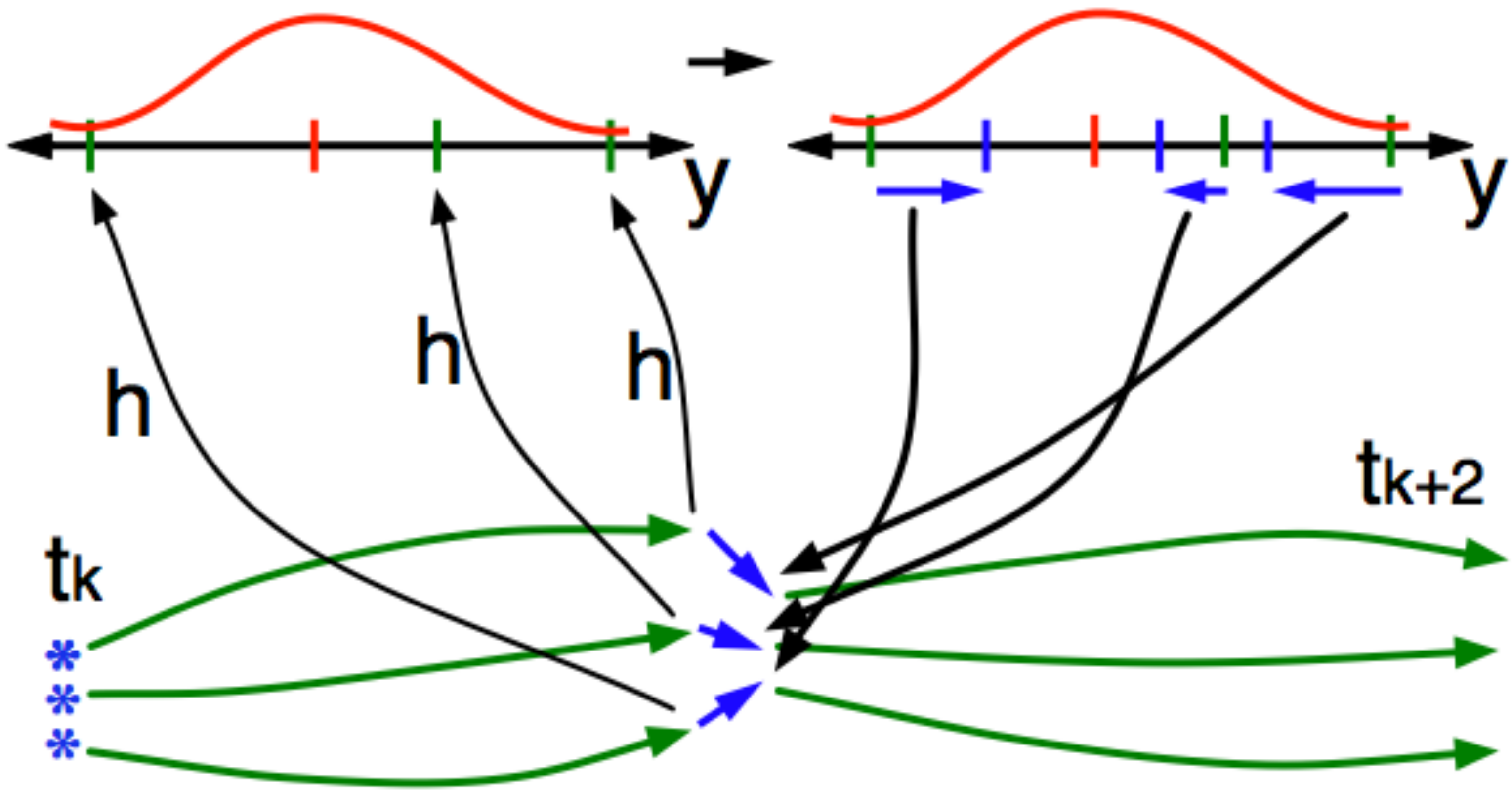
1. A way to make model forecasts.



Ensemble Filter for Large Geophysical Models

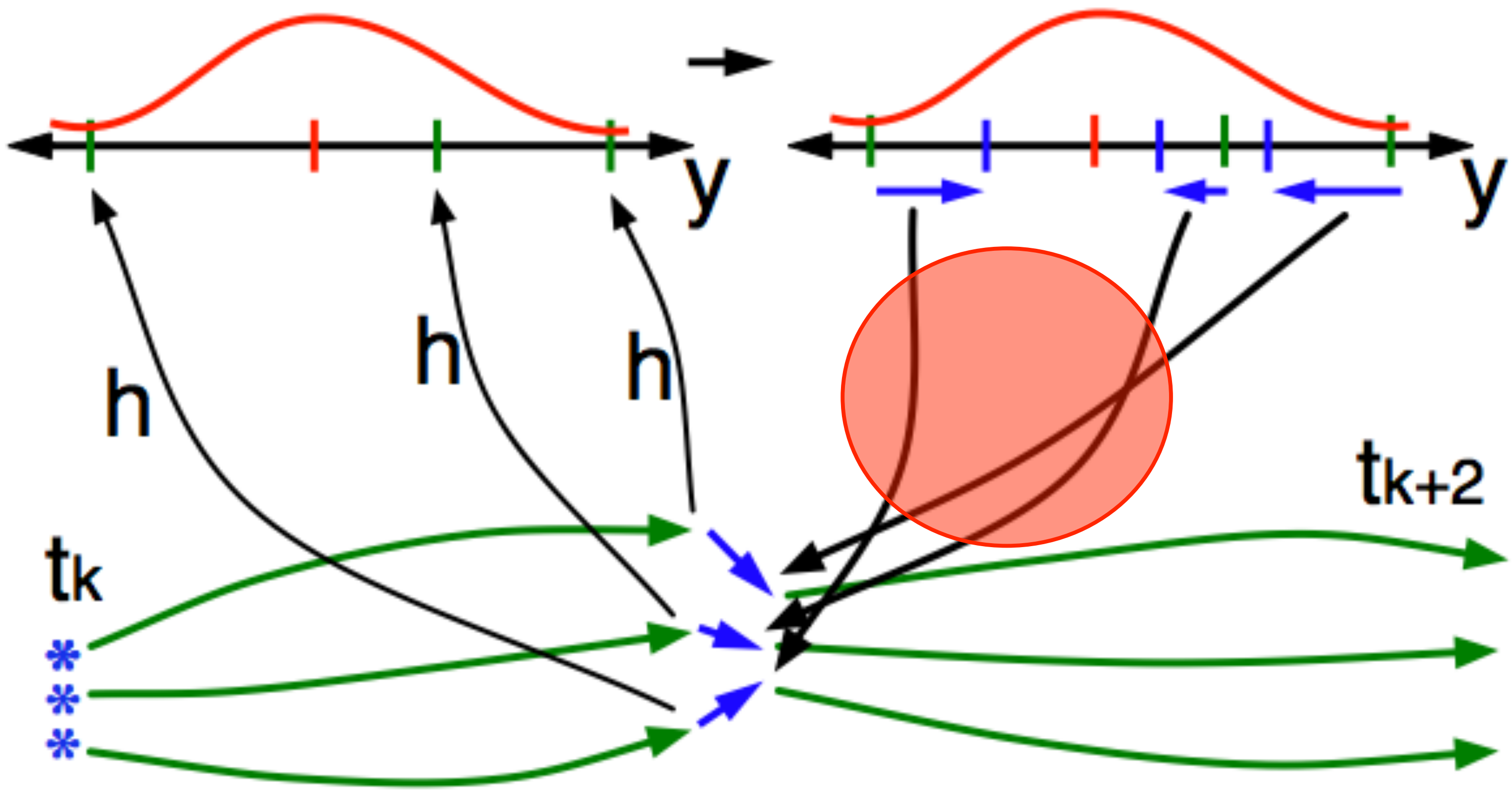
A generic ensemble filter system like DART just needs:

1. A way to make model forecasts.
2. A way to compute forward operators, h.



Parallel Implementation of Sequential Filter

For large models, regression of increments onto each state variable dominates time.



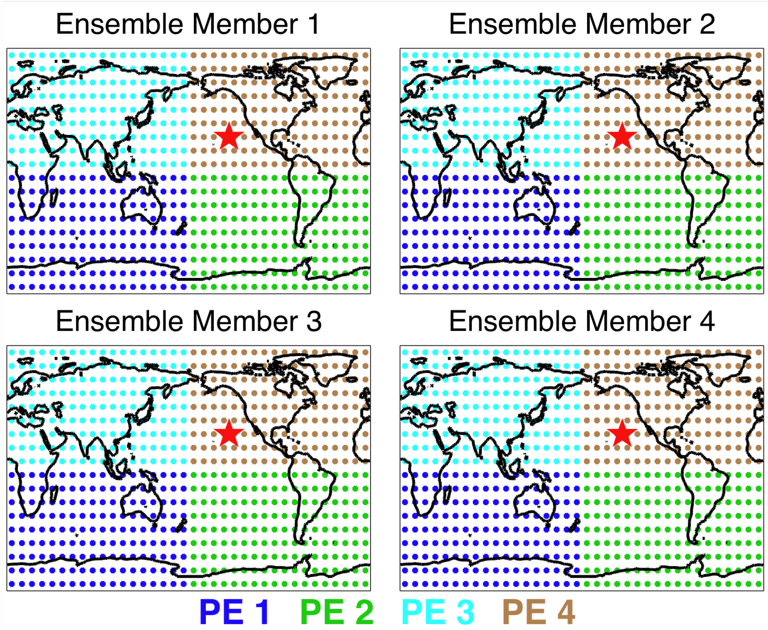
Parallel Implementation of Sequential Filter

For large models, regression of increments onto each state variable dominates time.

Simple example:

4 Ensemble members;
4 PEs (colors).

Observation shown by red star.



Parallel Implementation of Sequential Filter

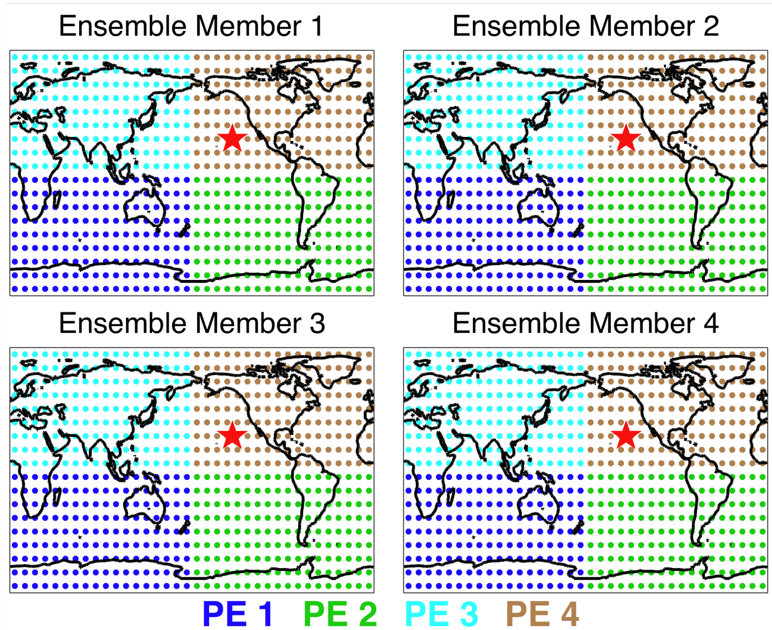
For large models, regression of increments onto each state variable dominates time.

One PE broadcasts obs. increments.

All ensemble members for each state variable are on one PE.

Can compute mean, variance without communication.

All state increments computed in parallel.

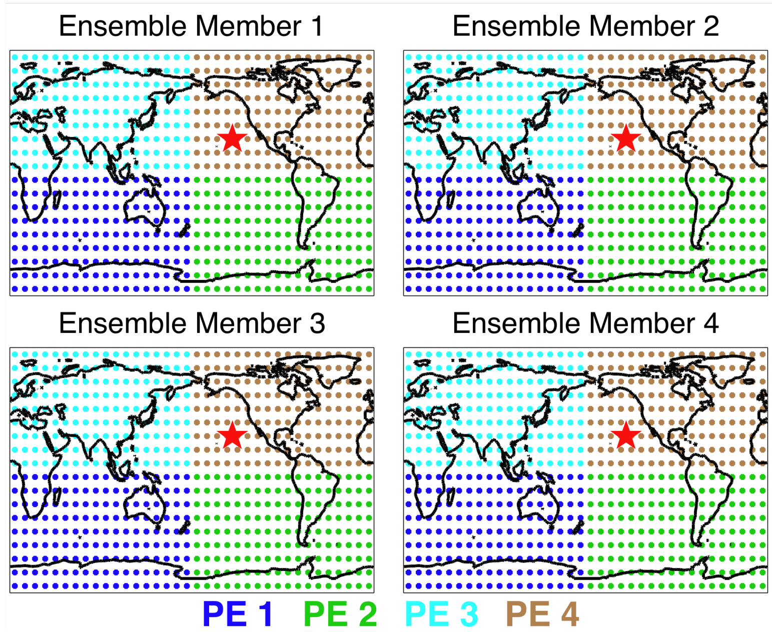


Parallel Implementation of Sequential Filter

For large models, regression of increments onto each state variable dominates time.

Computing forward operator, h, is usually local interpolation.

Most obs. require no communication.

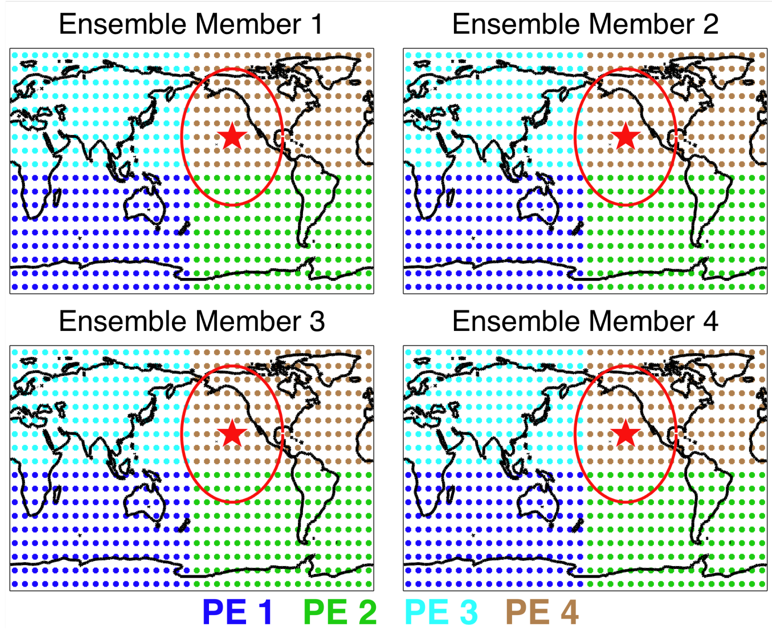


Parallel Implementation of Sequential Filter

For large models, regression of increments onto each state variable dominates time.

Observation impact usually localized.
- Reduces sampling error.
- Observation in N. Pacific not expected to change Antarctic state.

Now have a load balancing problem.

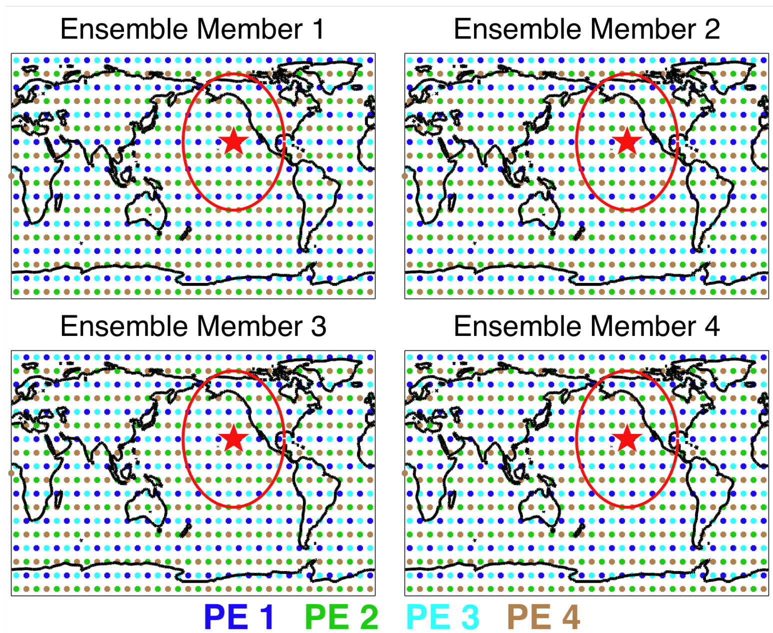


Parallel Implementation of Sequential Filter

For large models, regression of increments onto each state variable dominates time.

Can balance load by ‘randomly’ assigning state ensembles to PEs.

Now computing forward operators, h, requires communication.

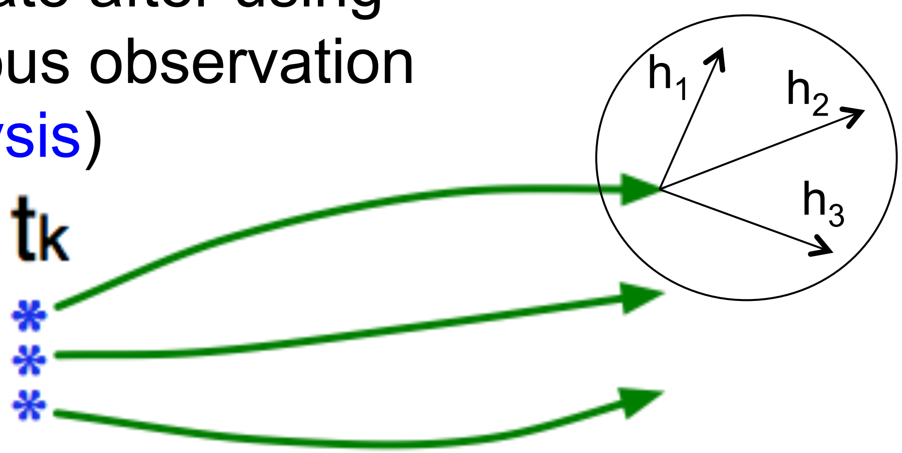


Ensemble Filter for Large Geophysical Models

1a. Compute ALL forward operators in a time window.

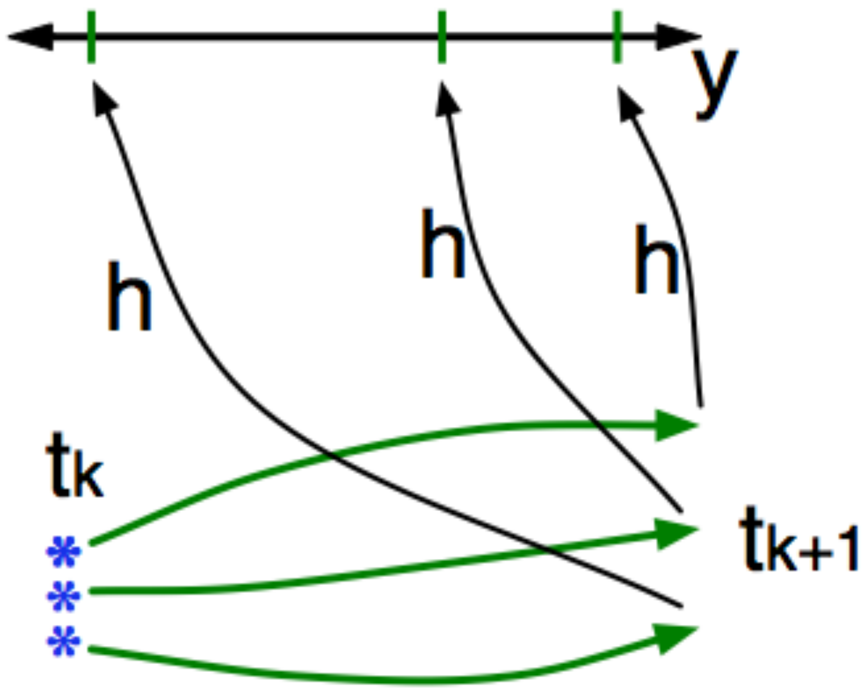
Define extended 'joint' state: $x_j=\{x, H(x)\}$ for each ensemble member

Ensemble state estimate after using previous observation (analysis)



Ensemble Filter for Large Geophysical Models

2. Get prior ensemble sample of observation, $y = h(x_j)$, by applying forward operator **h** to each ensemble member.



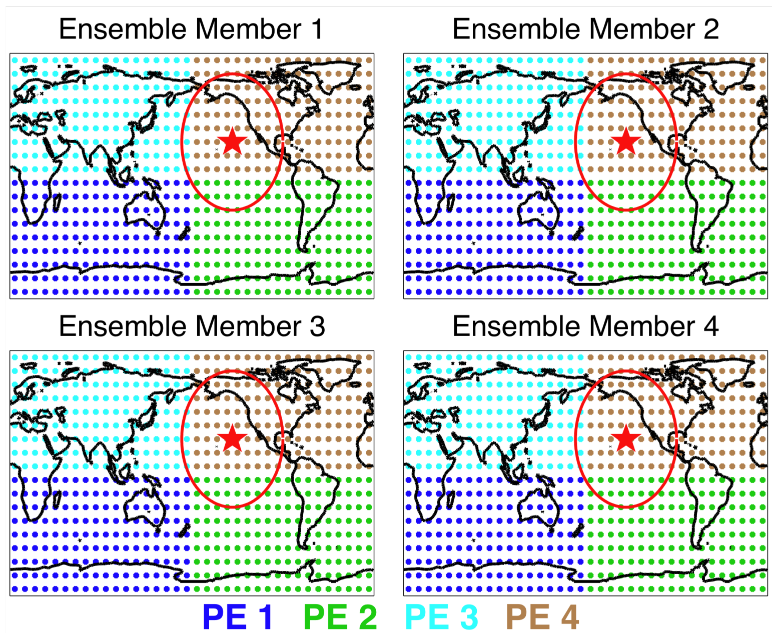
With joint state, forward operator is identity, no communication required. However, more regressions to do.

Parallel Implementation of Sequential Filter

Compute forward operators to get joint state before starting assimilation.

If each PE has a complete ensemble, forward operators require no communication.

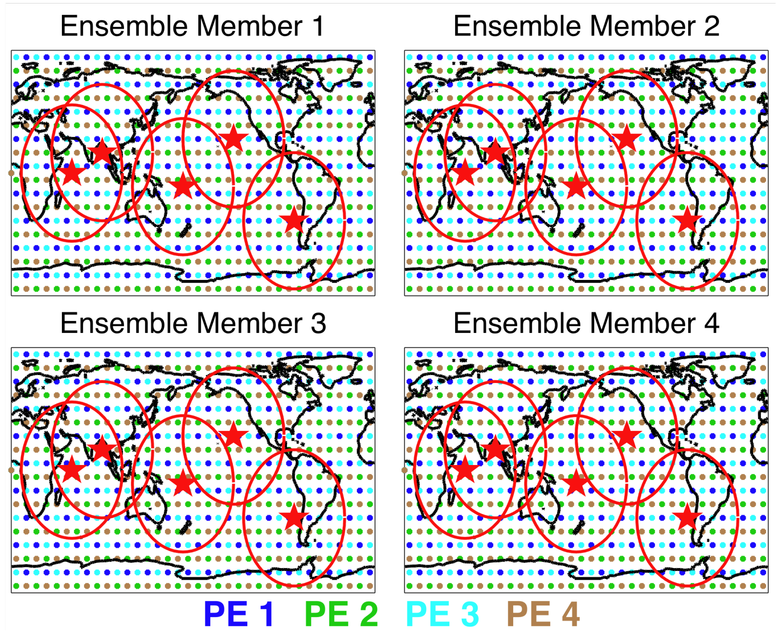
Can do many forward operators in parallel.



Parallel Implementation of Sequential Filter

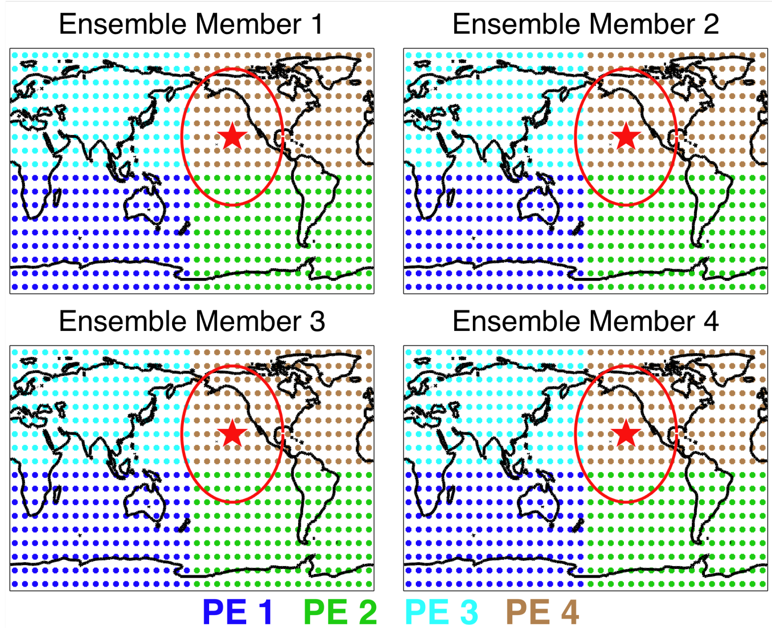
Do a data transpose, using all to all communication to get random layout.

Can do state increments for many obs in parallel for extra cost $O(n^2)$
(n is number of obs)



Parallel Implementation of Sequential Filter

Then transpose back to do more forward operators or advance model.



Parallel Implementation of Sequential Filter

Algorithm can be tuned for problem size, # of PEs;

Number of observations per transpose;
Selection of subsets of obs. to do in parallel;

How to assign state variables to PEs to:

1). Minimize transpose cost;
2). Minimize forward operator cost;
3). Minimize communication for updates.

Really fun for heterogeneous communication paths!

Parallel Implementation of Sequential Filter

Scaling for large atmospheric models:

Naïve random algorithm scales to O(100) PEs for mid-size climate / regional prediction models.

Expect modern NWP model to scale to O(1000).

O(10,000) seems viable with custom algorithm design.

Ensemble DA for Coupled Models

Straightforward from DA engineering perspective.
View coupled model as a single model.
Doesn't care which component state variable is from.
Doesn't matter what model observations are from.
Parallel implementations work unchanged.

In Process: Coupled DA for CESM Models

CESM is Community Earth System Model,
NCAR's coupled model for climate change.

Have ensemble DA for component models:

- CAM: atmosphere,
- POP: ocean,
- CLM: land,
- CICE: future development.

Coupled DA for CESM: What we are doing now.

CAM

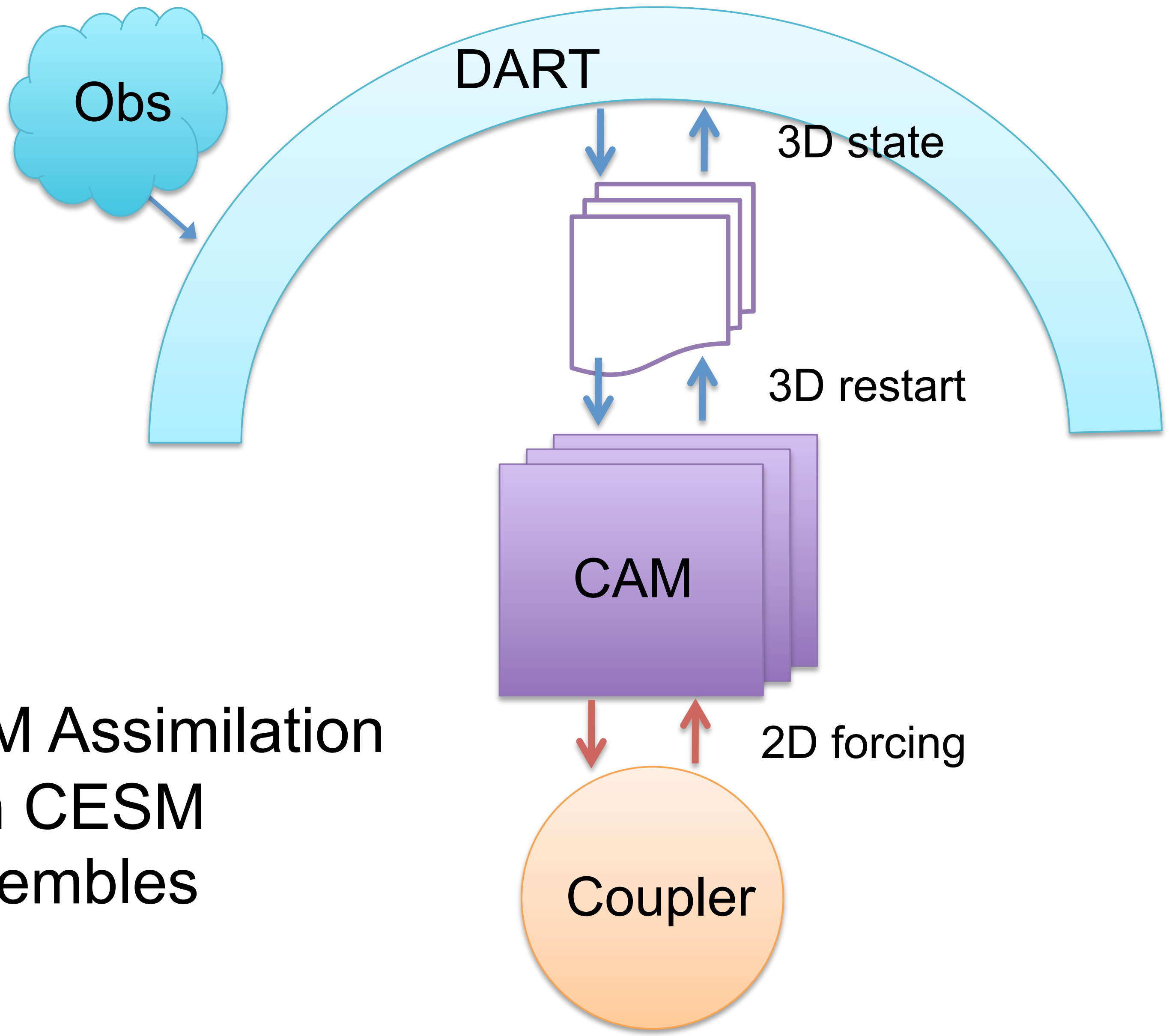
- Assimilating ATM obs with multiple executables of CAM.
- Could now also use CESM coupler w/ ensembles of CAM.

POP

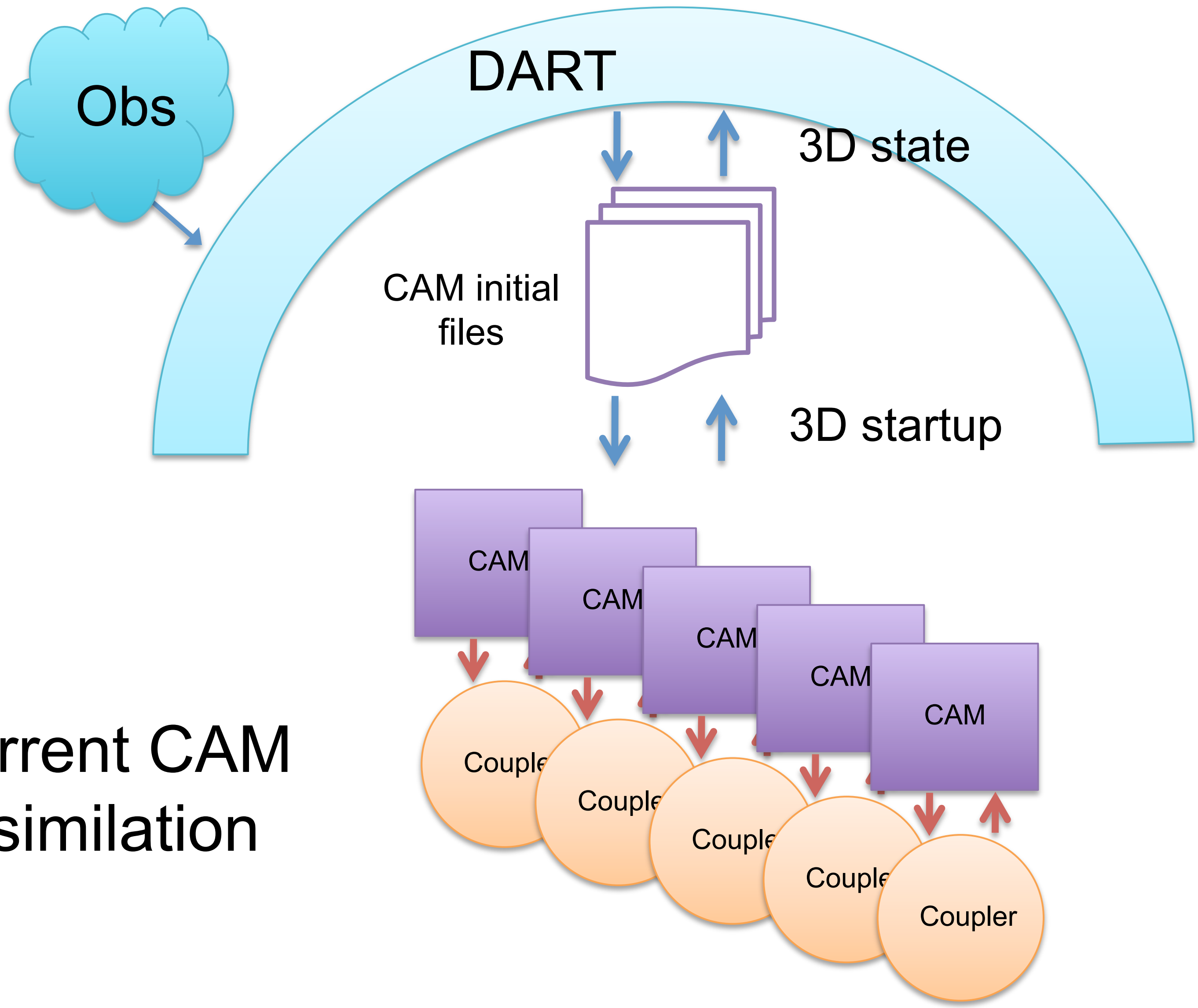
- Use new CESM ensemble capability.
- Assimilating OCN obs with CESM POP.
- Start and stop CESM each day.
- CESM job script calls DART assimilation script.
- Transfer state by reading/writing restart files.

CLM

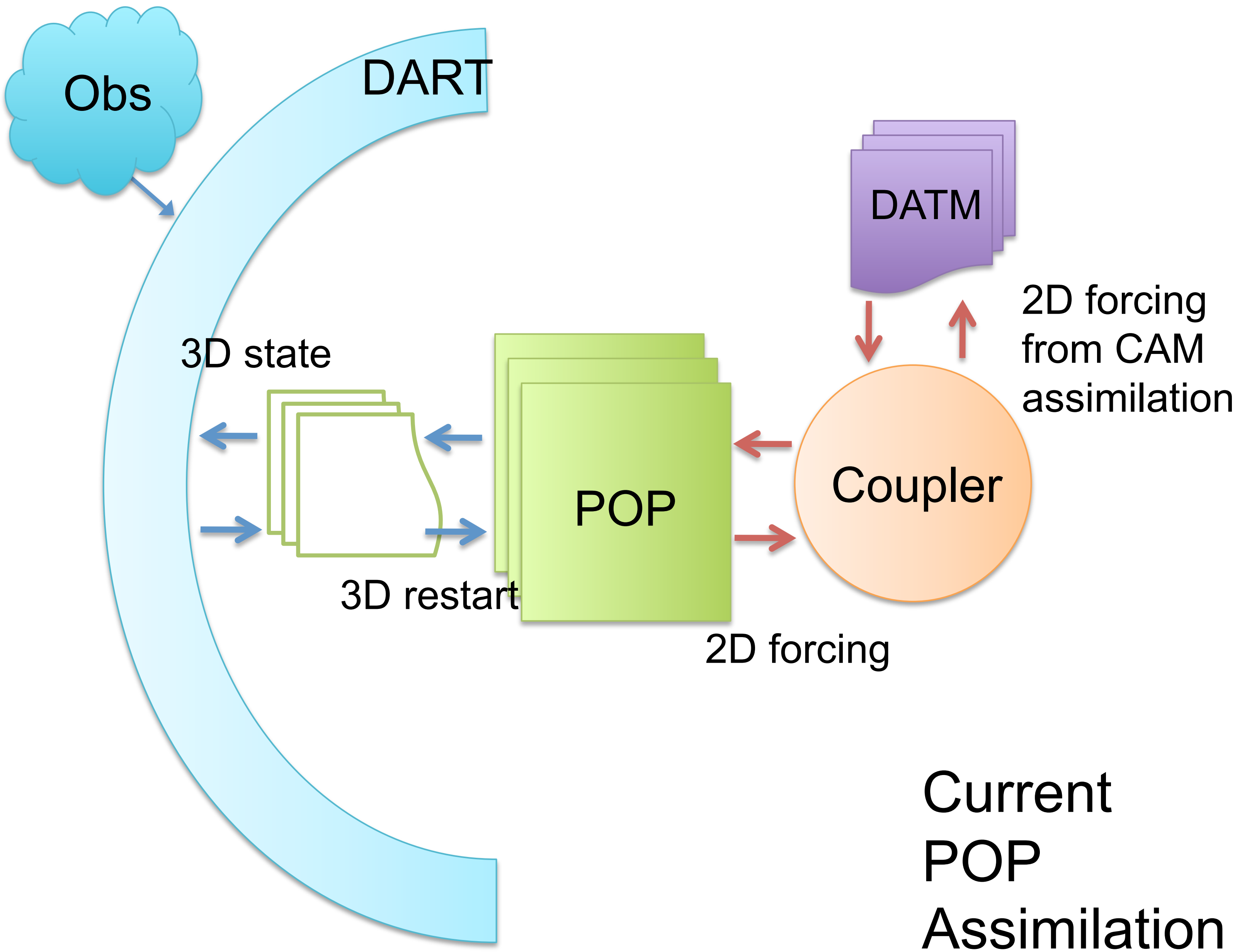
- DA implemented, challenges remain.



CAM Assimilation with CESM ensembles

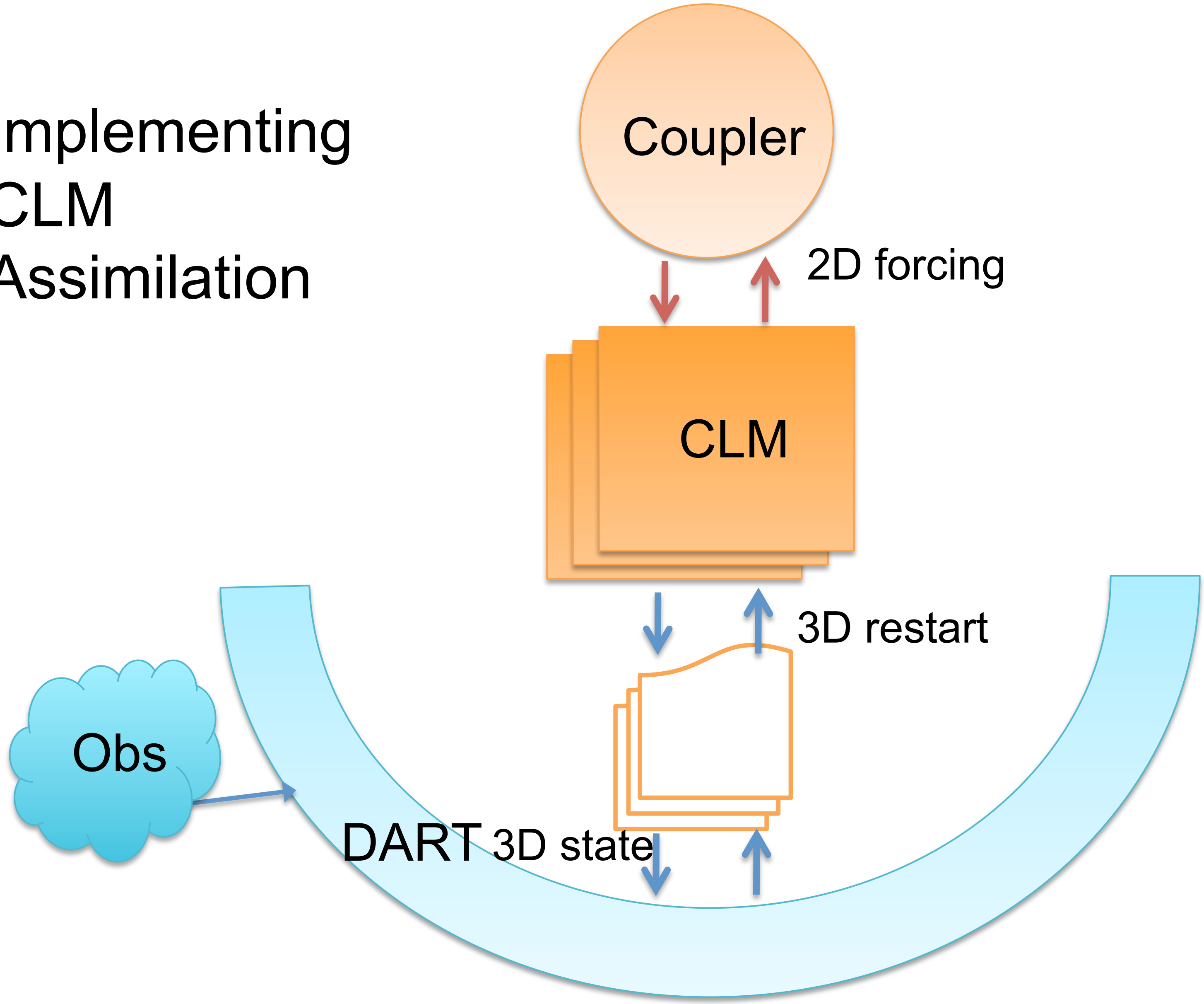


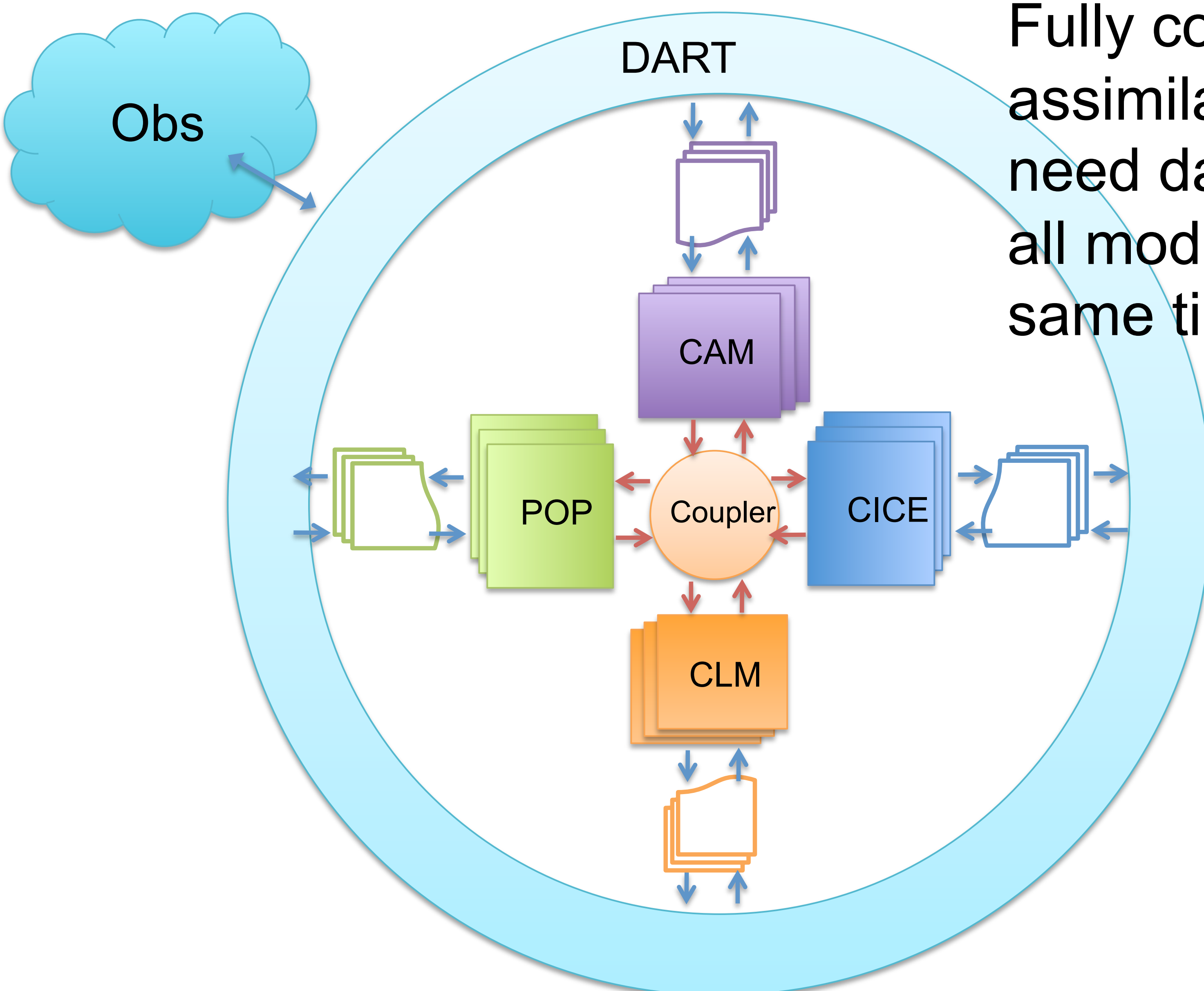
Current CAM Assimilation



Current POP Assimilation

Implementing CLM Assimilation



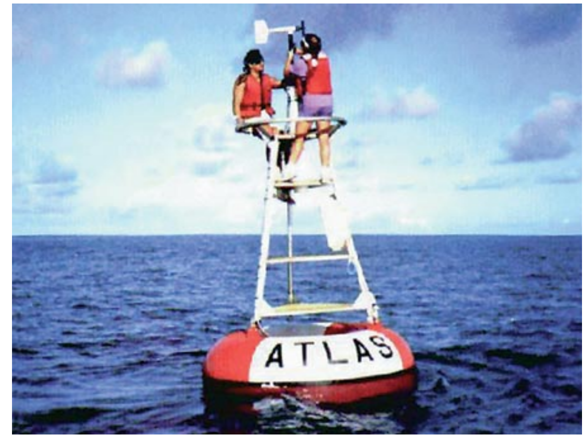


Fully coupled assimilation will need data from all models at the same time

World Ocean Database T,S observation counts

These counts are for 1998 & 1999 and are representative.

FLOAT_SALINITY	68200
FLOAT_TEMPERATURE	395032
DRIFTER_TEMPERATURE	33963
MOORING_SALINITY	27476
MOORING_TEMPERATURE	623967
BOTTLE_SALINITY	79855
BOTTLE_TEMPERATURE	81488
CTD_SALINITY	328812
CTD_TEMPERATURE	368715
STD_SALINITY	674
STD_TEMPERATURE	677
XCTD_SALINITY	3328
XCTD_TEMPERATURE	5790
MBT_TEMPERATURE	58206
XBT_TEMPERATURE	1093330
APB_TEMPERATURE	580111



- temperature observation error standard deviation == 0.5 K.
- salinity observation error standard deviation == 0.5 msu.

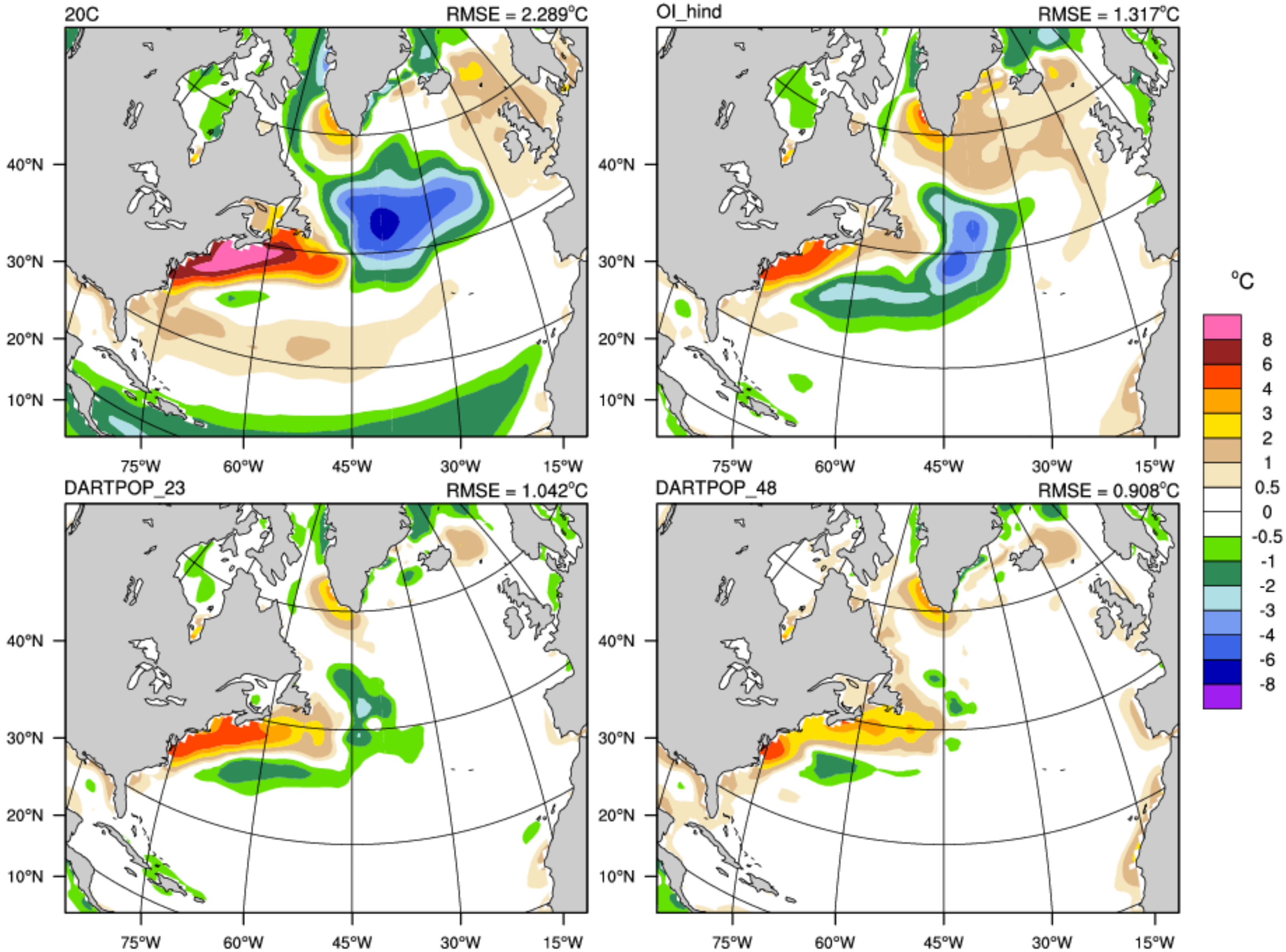
Physical Space: 1998/1999 SST Anomaly from HadOI-SST

Coupled Free Run

POP forced by observed atmosphere (hindcast)

23 POP 1 DATM

48 POP 48 CAM



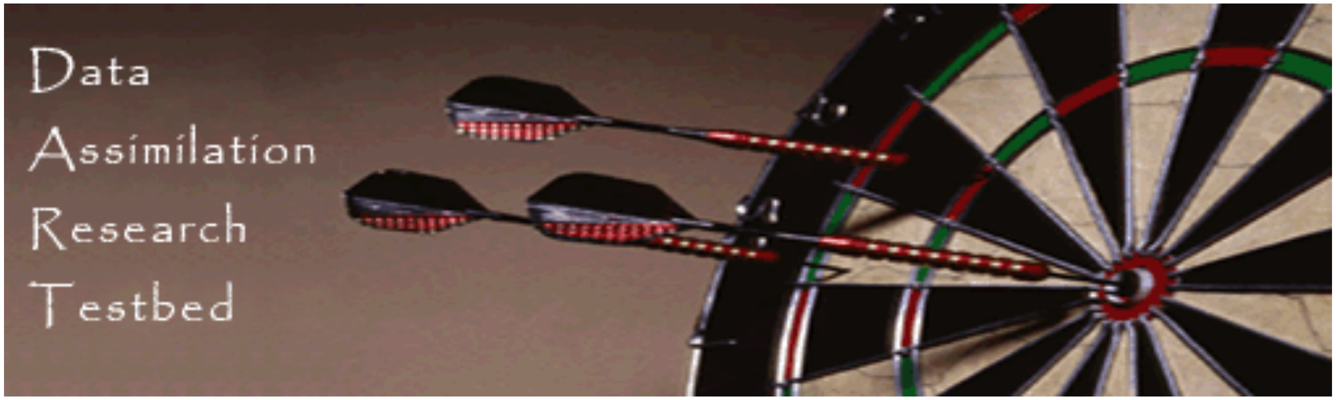
Challenges for Coupled Ensemble DA

Engineering ensemble DA system is not hard but…

- Frequent restarting of coupled model.
- State variables that don't have well-defined priors.
  - Snow temperature example.
- Interaction of different time/space scales.
- Localization of observations across boundaries.
  - I think we know how to get guidance for this.

- <u>Models that don't make accurate predictions.</u>

Code to implement all of the algorithms discussed are freely available from:



http://www.image.ucar.edu/DAReS/DART/