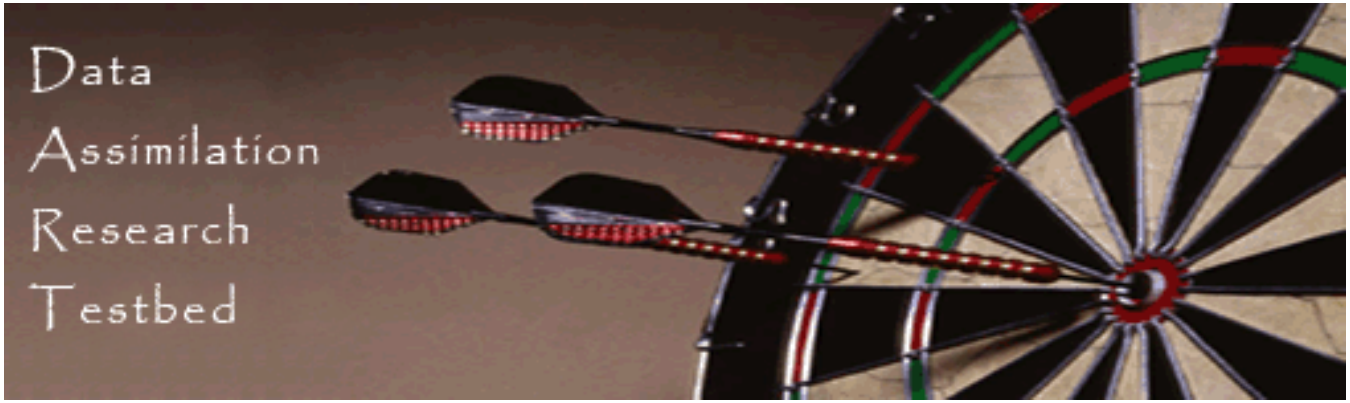
An Introduction to Ensemble Data Assimilation and the Data Assimilation Research Testbed (DART)



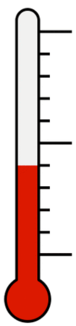
Jeffrey Anderson, Nancy Collins, Tim Hoar,

Hui Liu, Glen Romine, Kevin Raeder

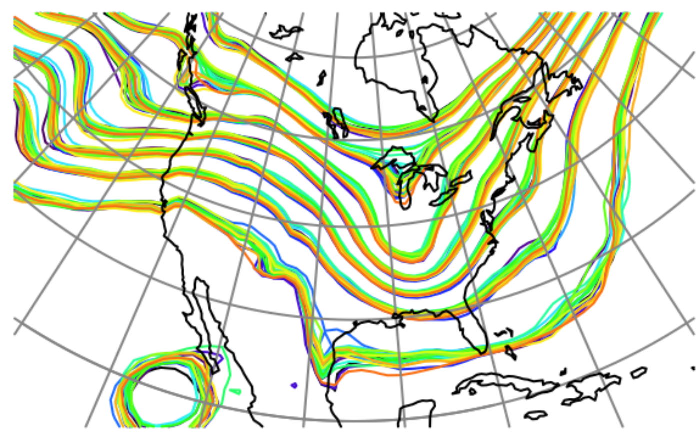
NCAR Institute for Math Applied to Geophysics

What is Data Assimilation?

Observations combined with a Model forecast…



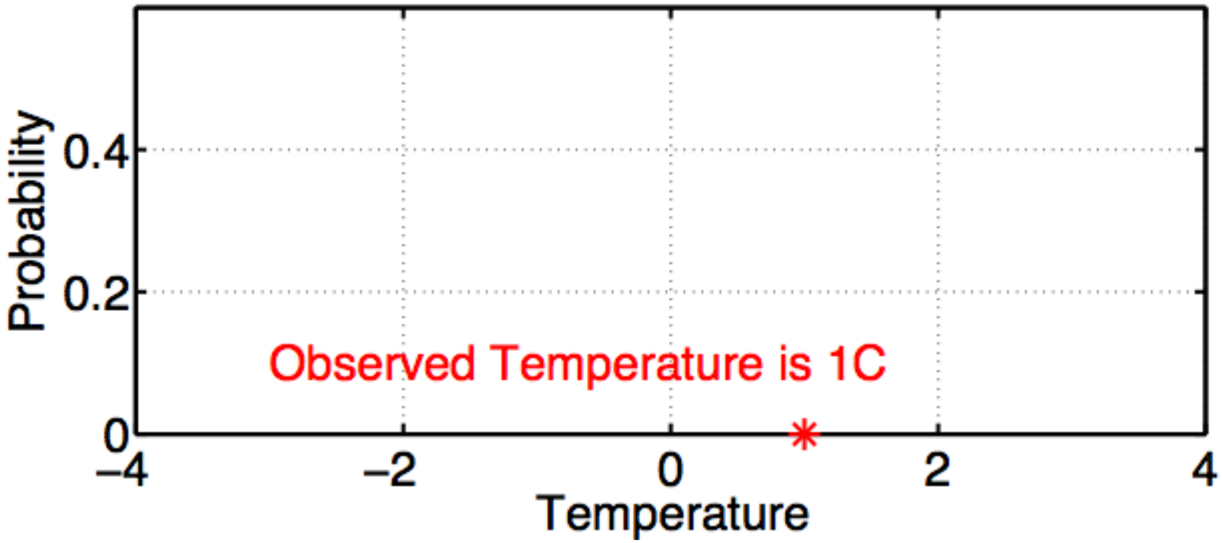
+



…to produce an analysis
(best possible estimate).

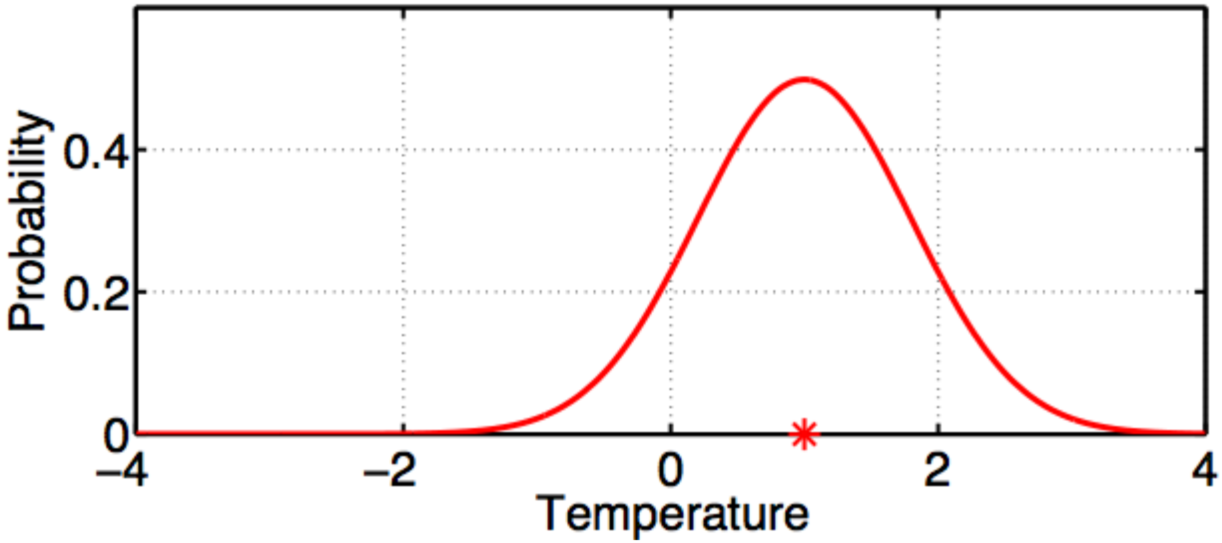
Example: Estimating the Temperature Outside

An observation has a value (∗),



Example: Estimating the Temperature Outside

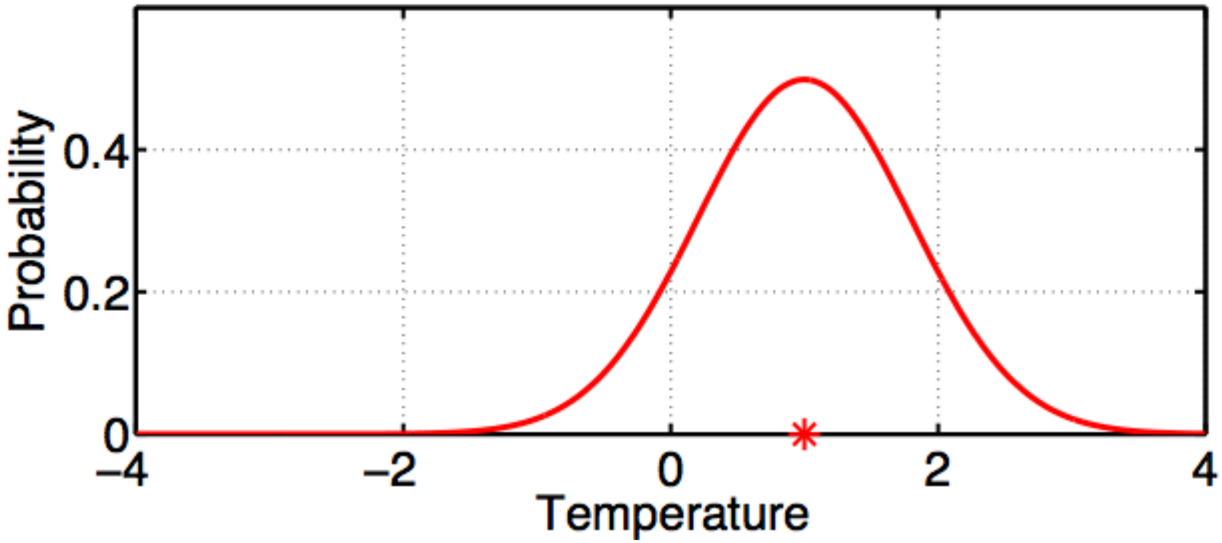
An observation has a value (*),

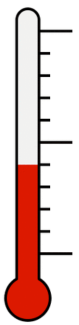


and an error distribution (red curve) that is associated with the instrument.

Example: Estimating the Temperature Outside

Thermometer outside measures 1C.

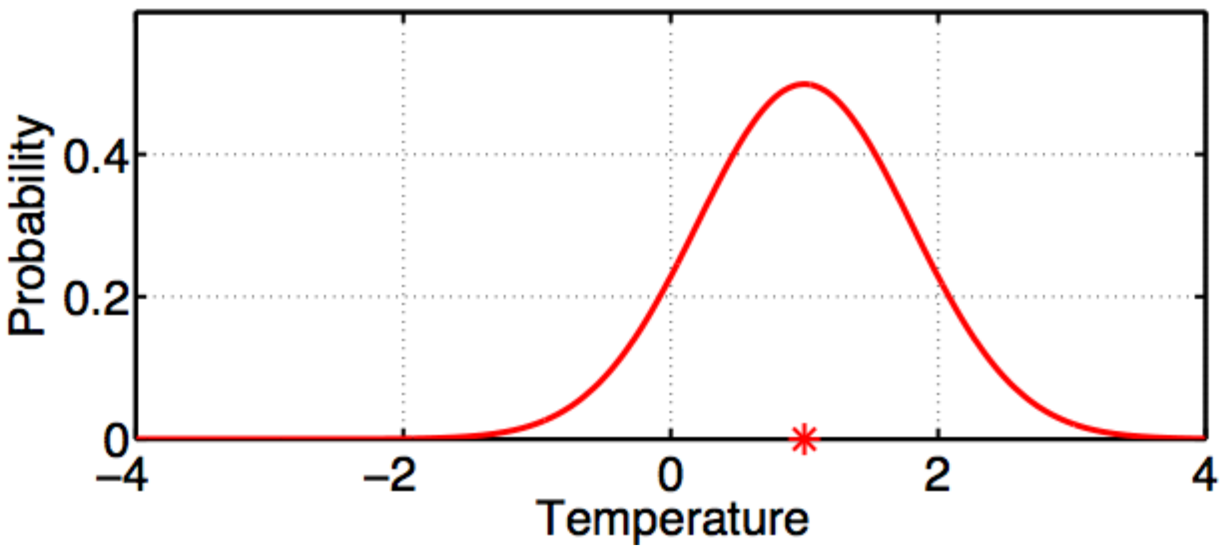


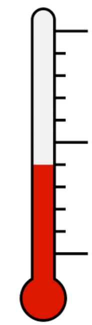


Instrument builder says thermometer is unbiased with +/- 0.8C gaussian error.

Example: Estimating the Temperature Outside

Thermometer outside measures 1C.

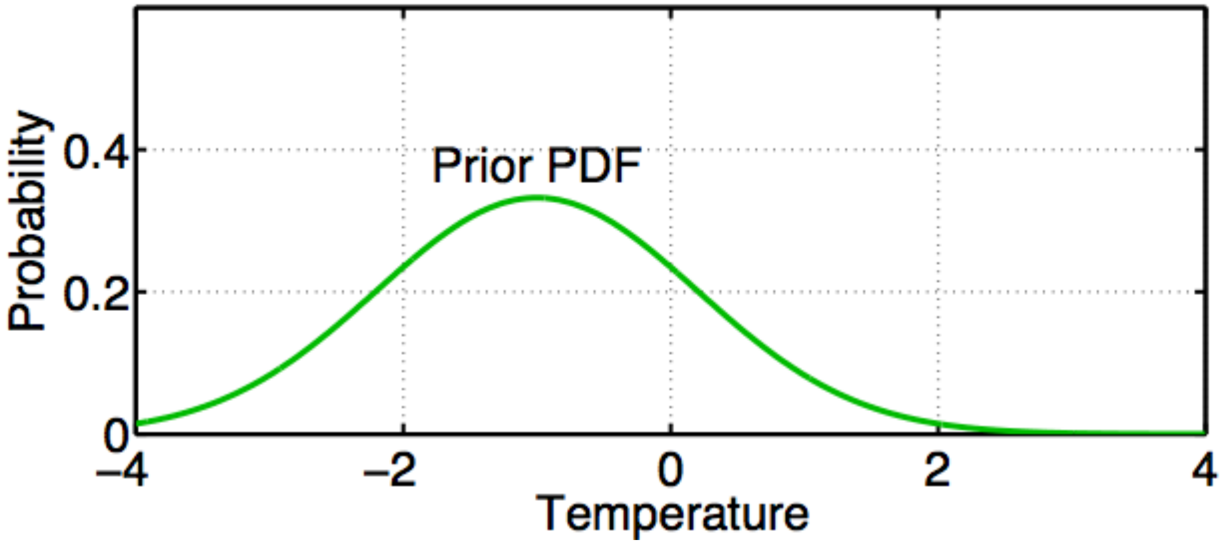




The red plot is $P(T \mid T_o)$, probability of temperature given that T_o was observed.

Example: Estimating the Temperature Outside

We also have a prior estimate of temperature.



The green curve is $P(T \mid C)$; probability of temperature given all available prior information C.

Example: Estimating the Temperature Outside

Prior information C can include:

1. Observations of things besides T;
2. Model forecast made using observations at earlier times;
3. *A priori* physical constraints (T > -273.15C);
4. Climatological constraints (-30C < T < 40C).

Combining the Prior Estimate and Observation

Bayes Theorem:

$$P\left(T \mid T_o, C\right) = \frac{P\left(T_o \mid T, C\right) P\left(T \mid C\right)}{Normalization}$$

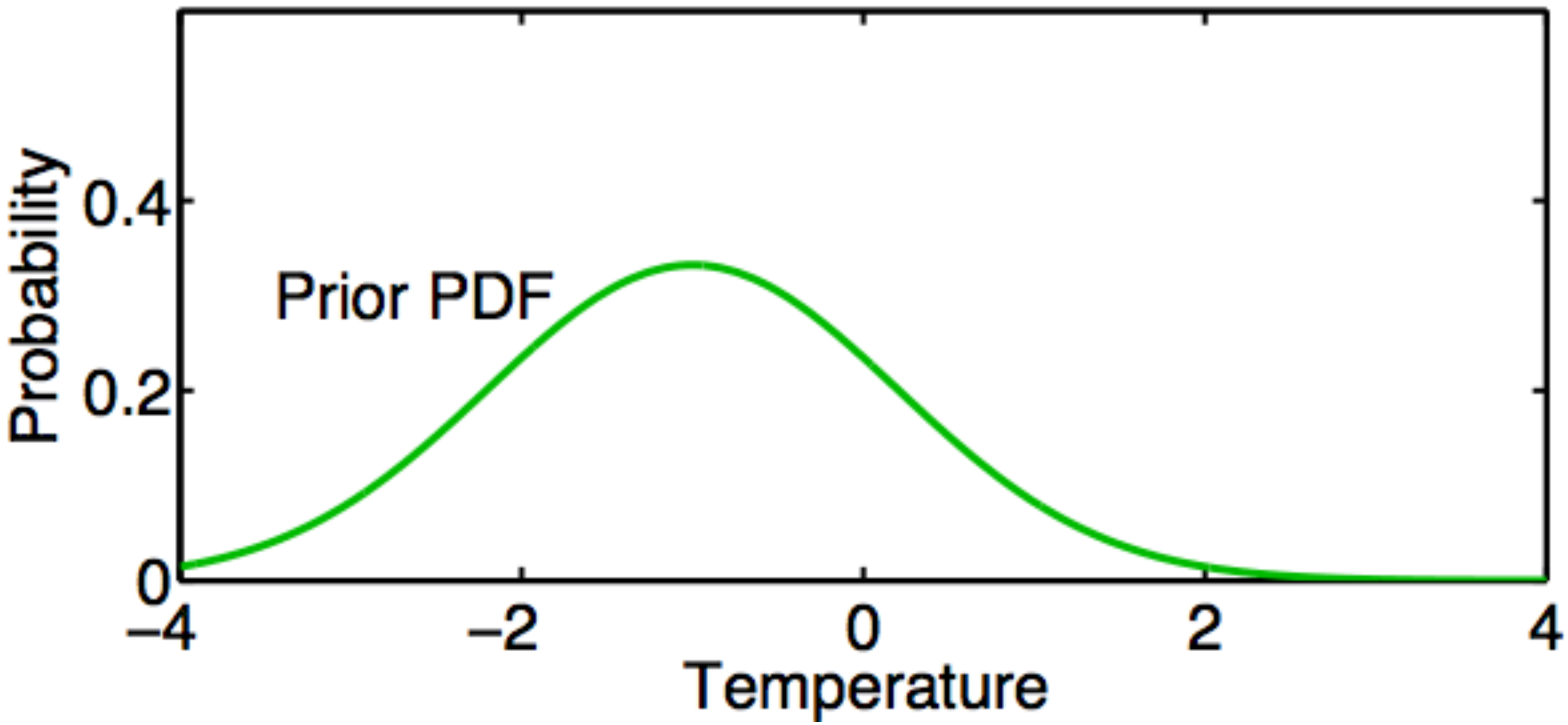
Prior

Posterior: Probability of T given observations and Prior. Also called update or analysis.

Likelihood: Probability that T_o is observed if T is true value and given prior information C.

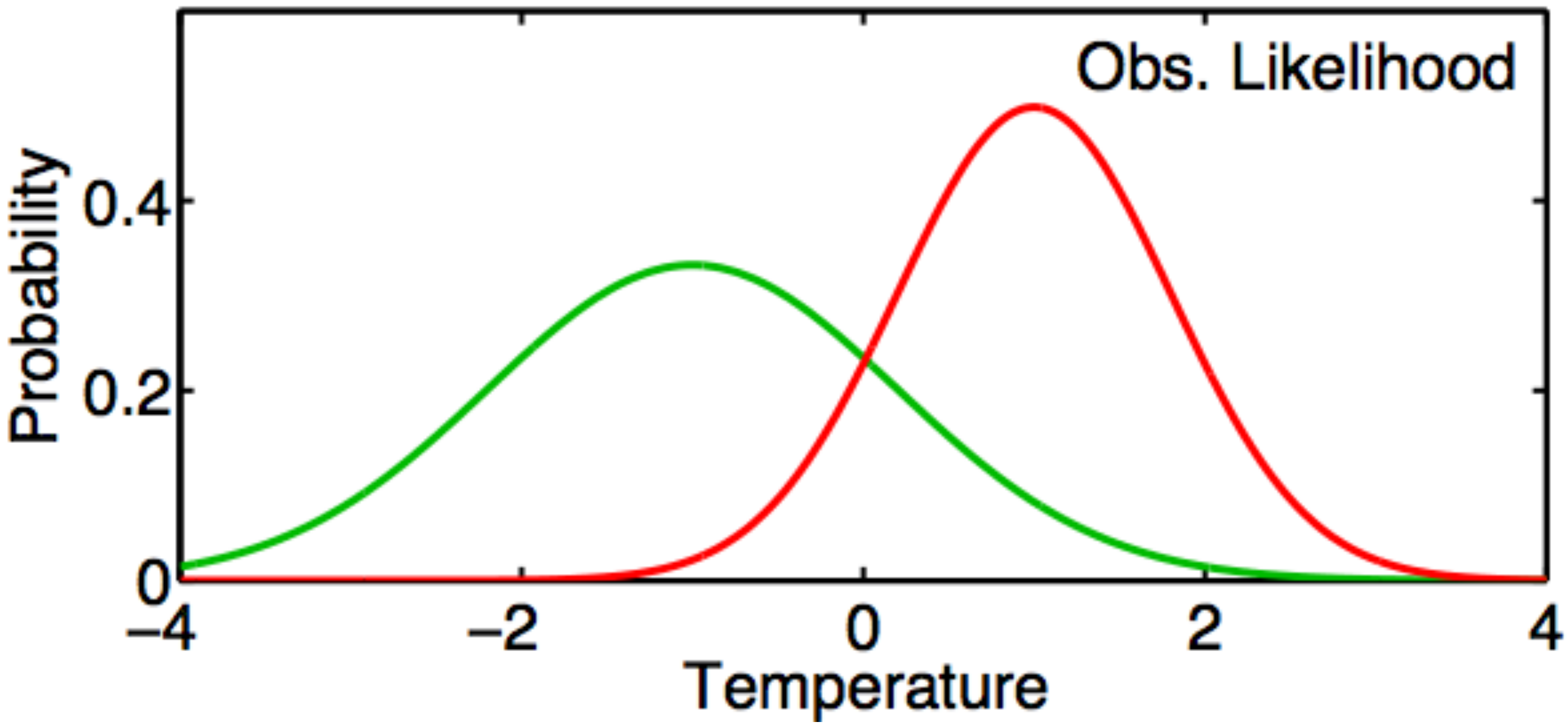
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



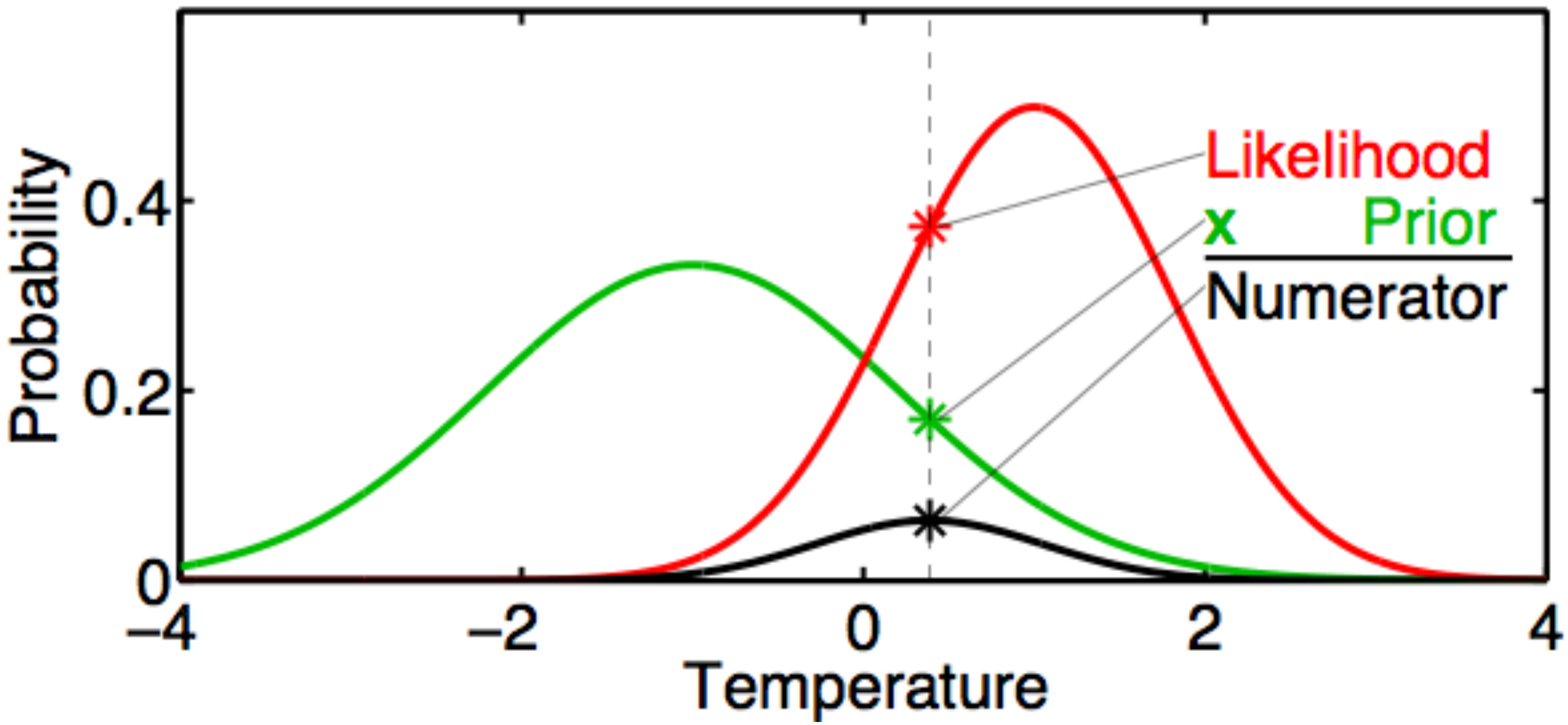
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



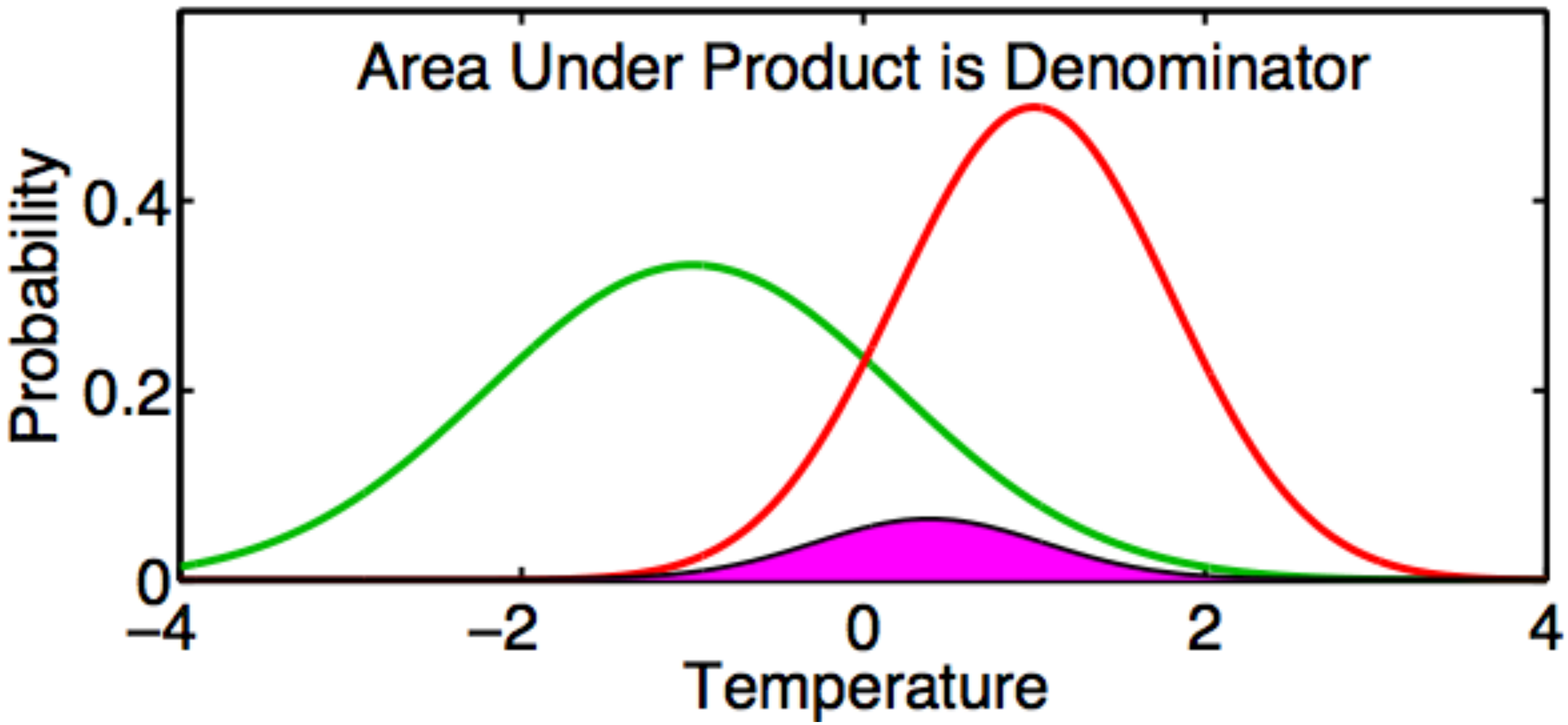
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



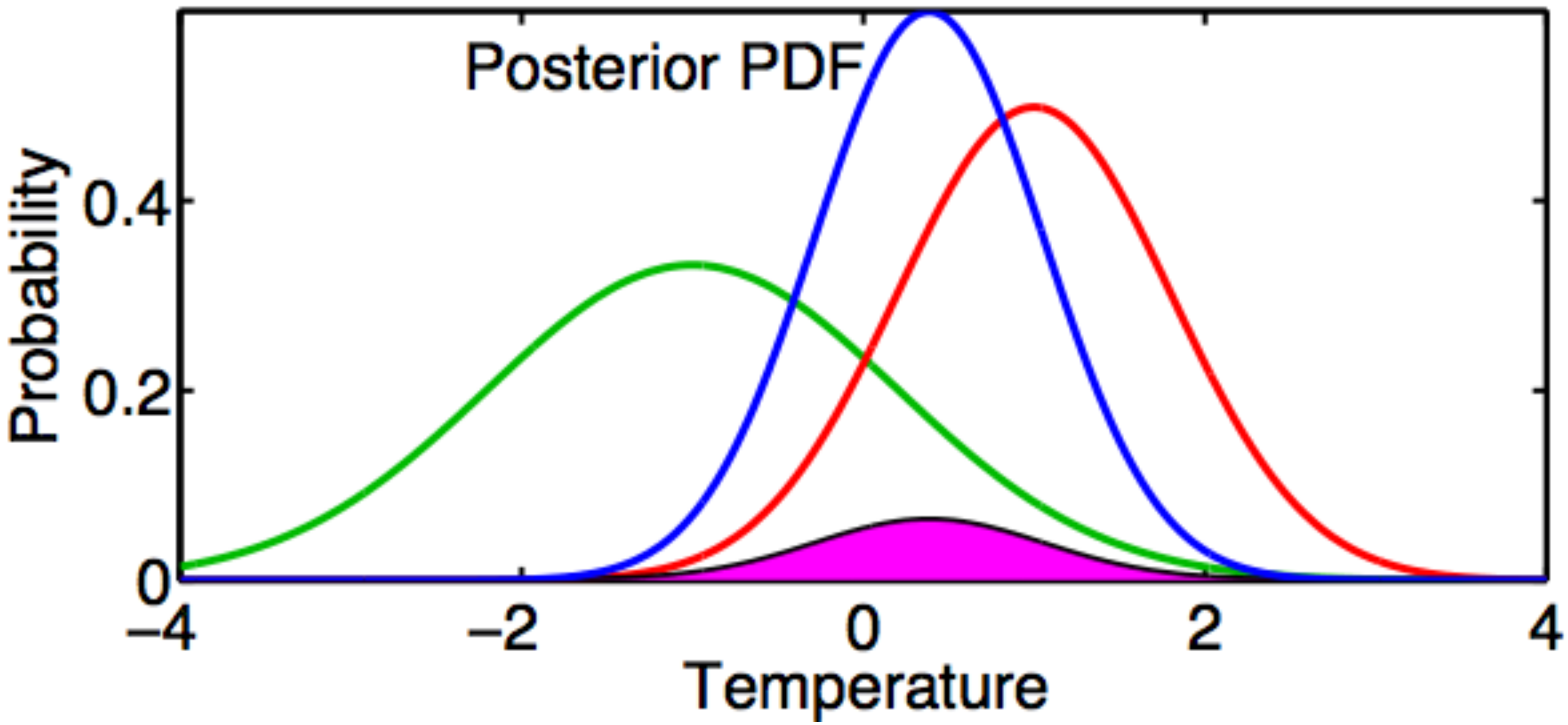
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



Consistent Color Scheme Throughout Tutorial

Green = Prior

Red = Observation

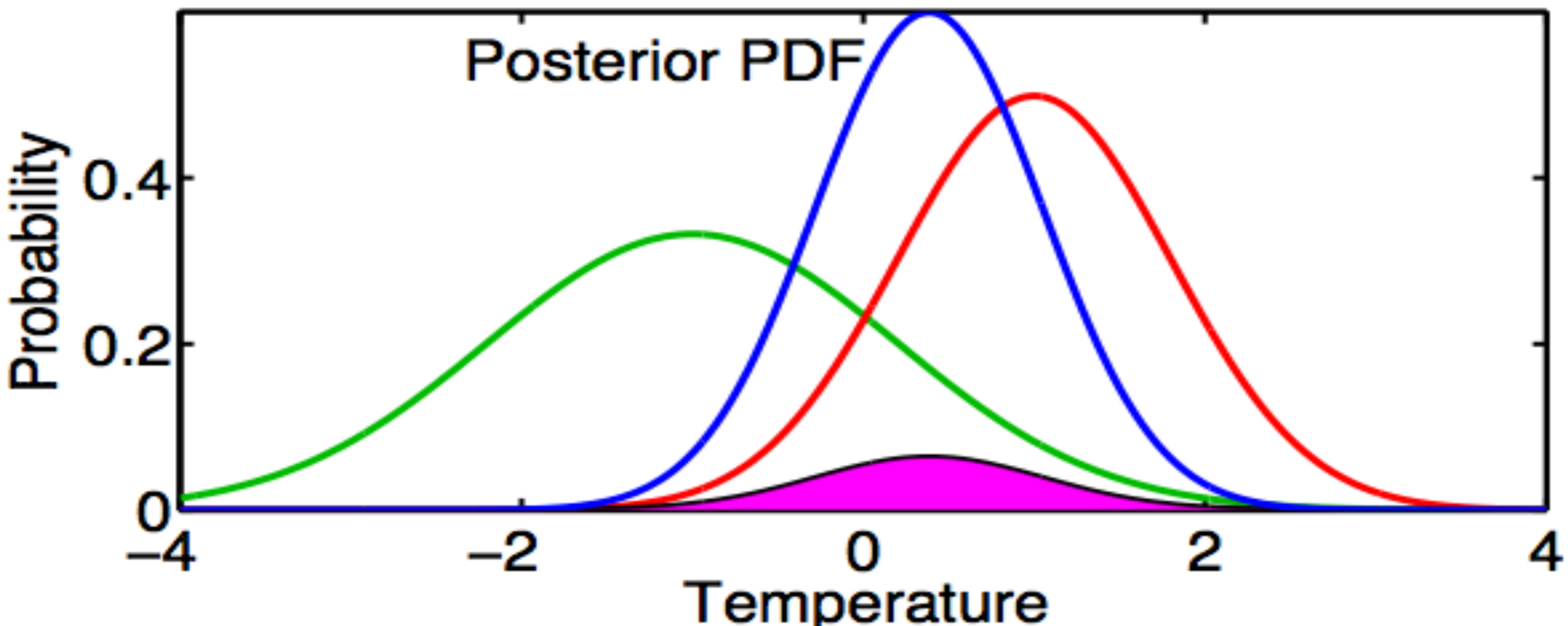
Blue = Posterior

Black = Truth

Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$

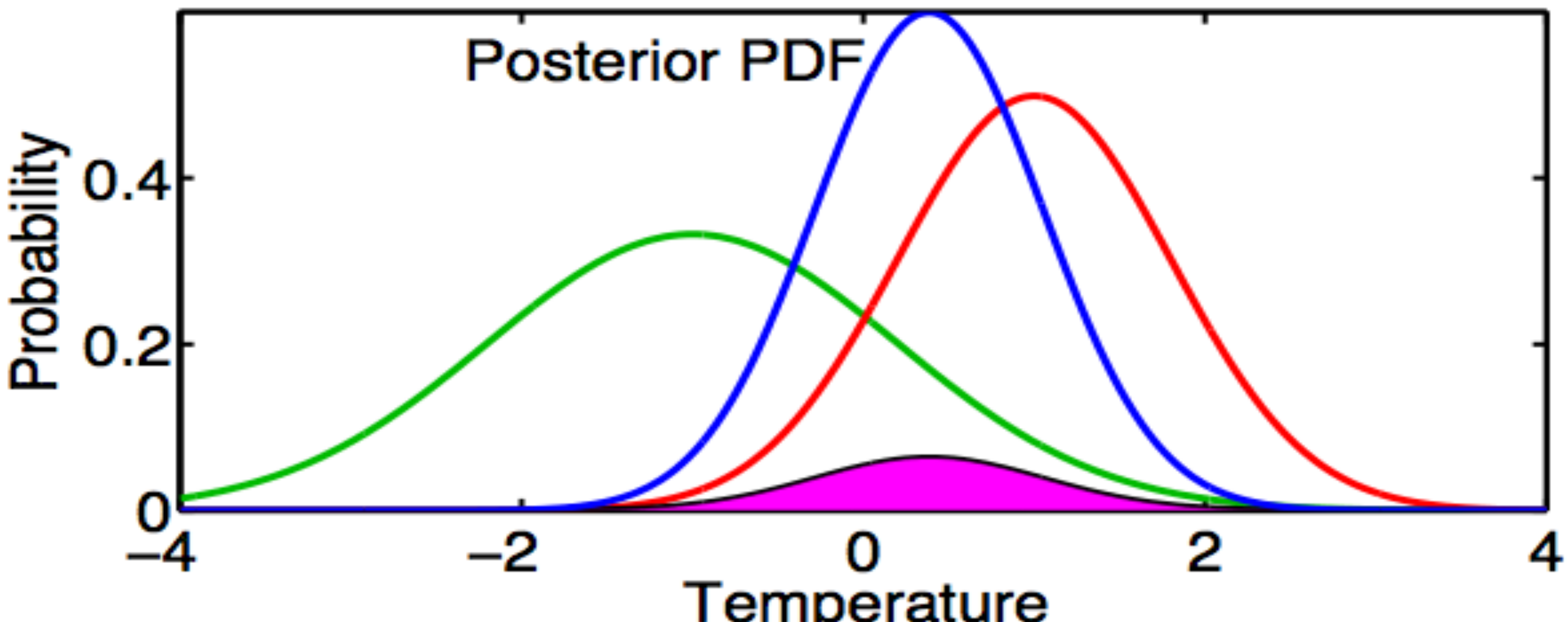
Generally no analytic solution for Posterior.



Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$

Gaussian Prior and Likelihood -> Gaussian Posterior



Combining the Prior Estimate and Observation

For Gaussian prior and likelihood…

Prior $$P(T \mid C) = Normal(T_p, \sigma_p)$$

Likelihood $$P(T_o \mid T, C) = Normal(T_o, \sigma_o)$$

Then, Posterior $$P(T \mid T_o, C) = Normal(T_u, \sigma_u)$$

With

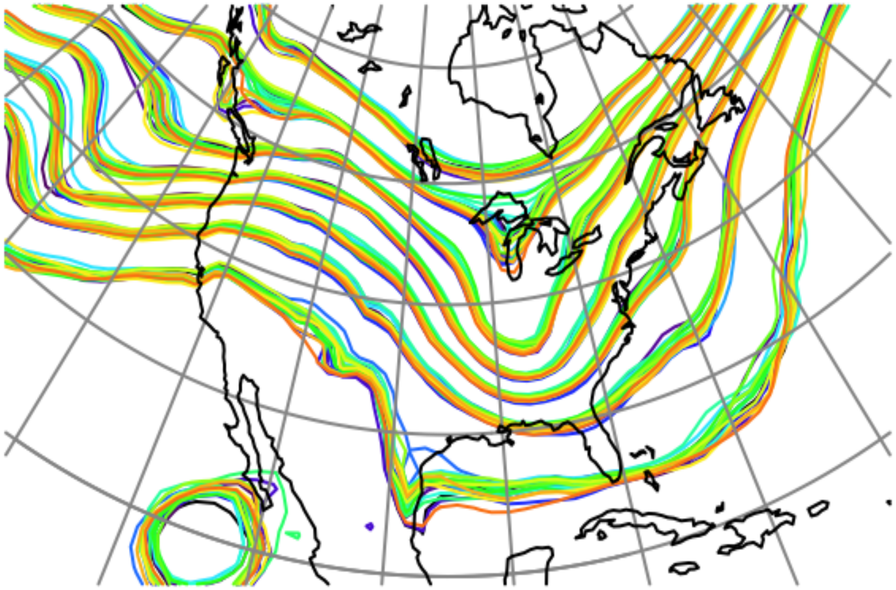
$$\sigma_u = \sqrt{\left(\sigma_p^{-2} + \sigma_o^{-2}\right)^{-1}}$$

$$T_u = \sigma_u^2\left[\sigma_p^{-2} T_p + \sigma_o^{-2} T_o\right]$$

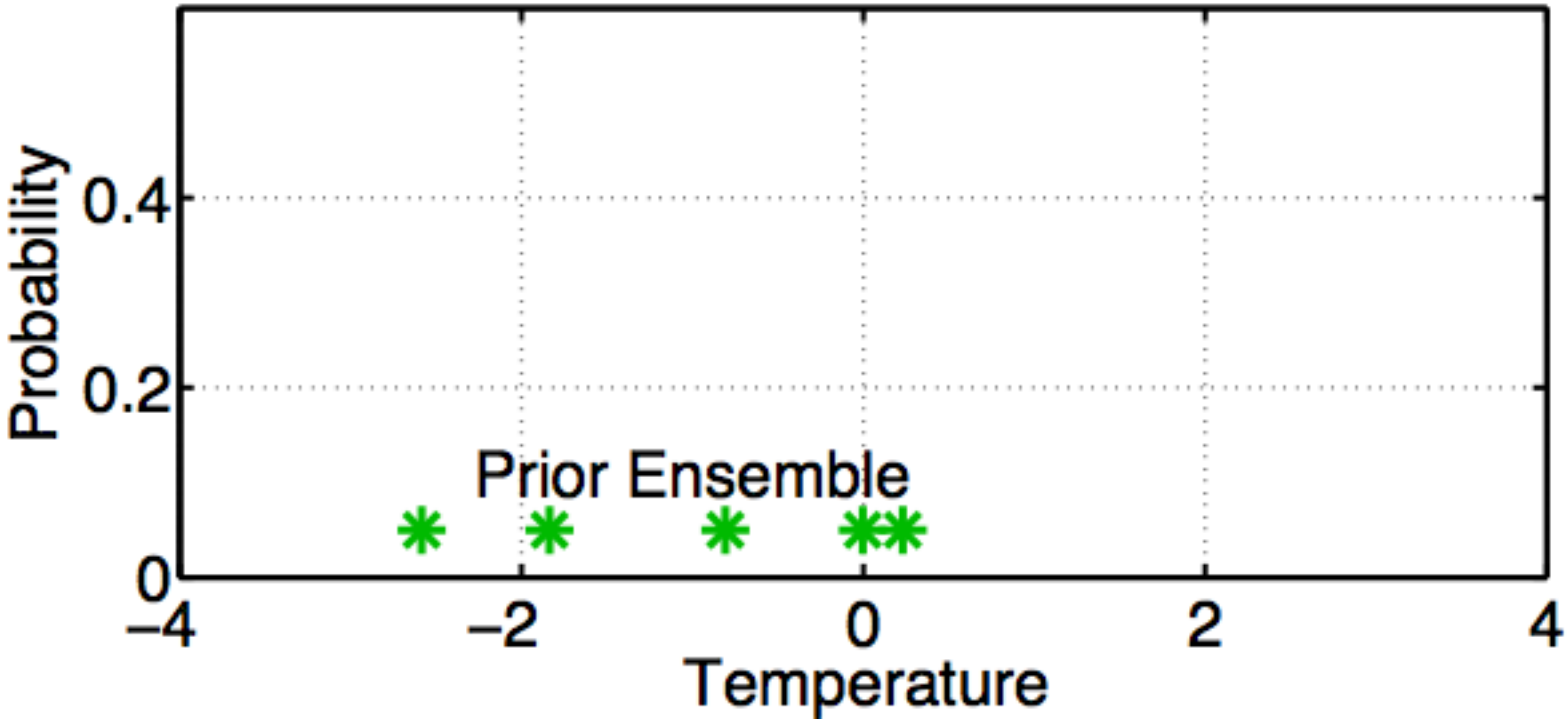
What is Ensemble Data Assimilation?

Use an ensemble (set) of model forecasts.

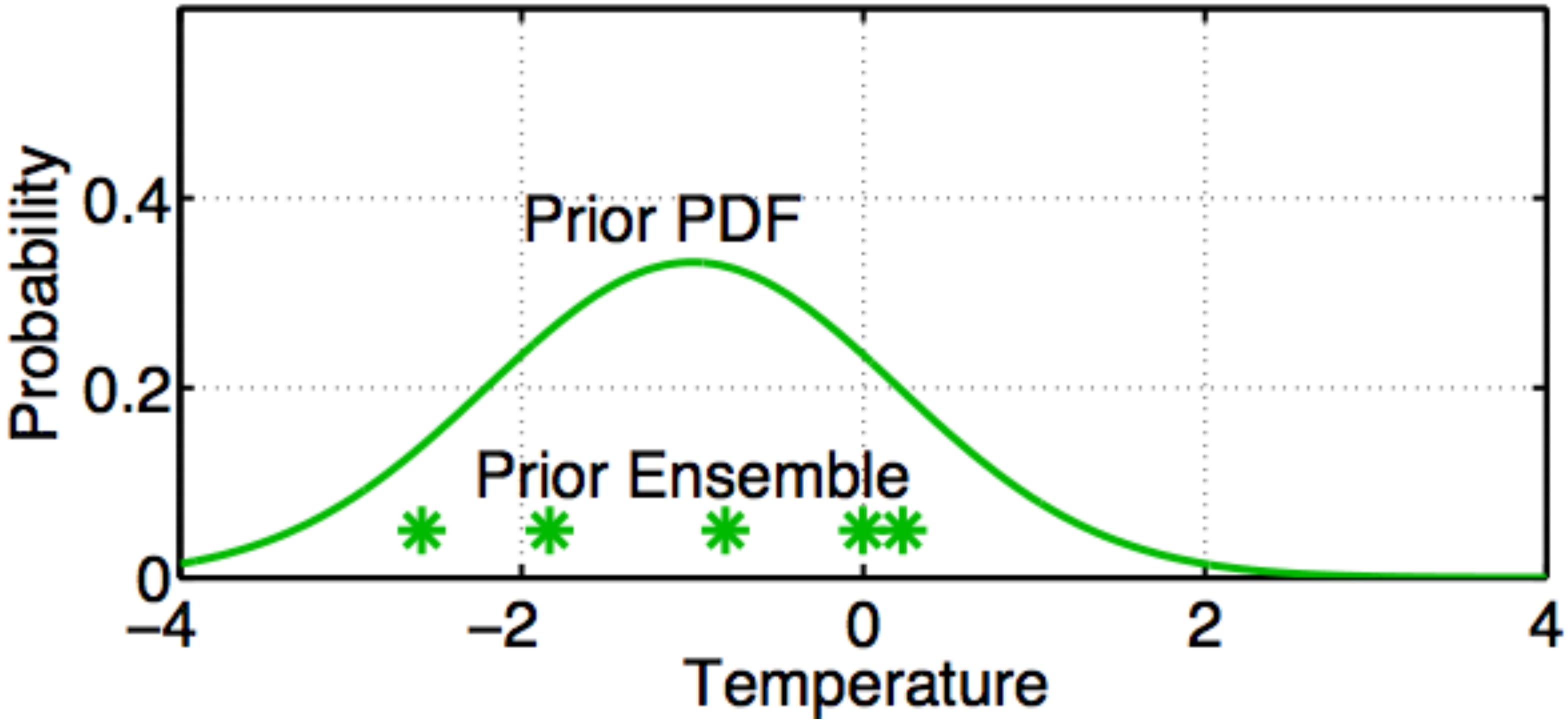
Use sample statistics to get covariance between state and observations.



A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation

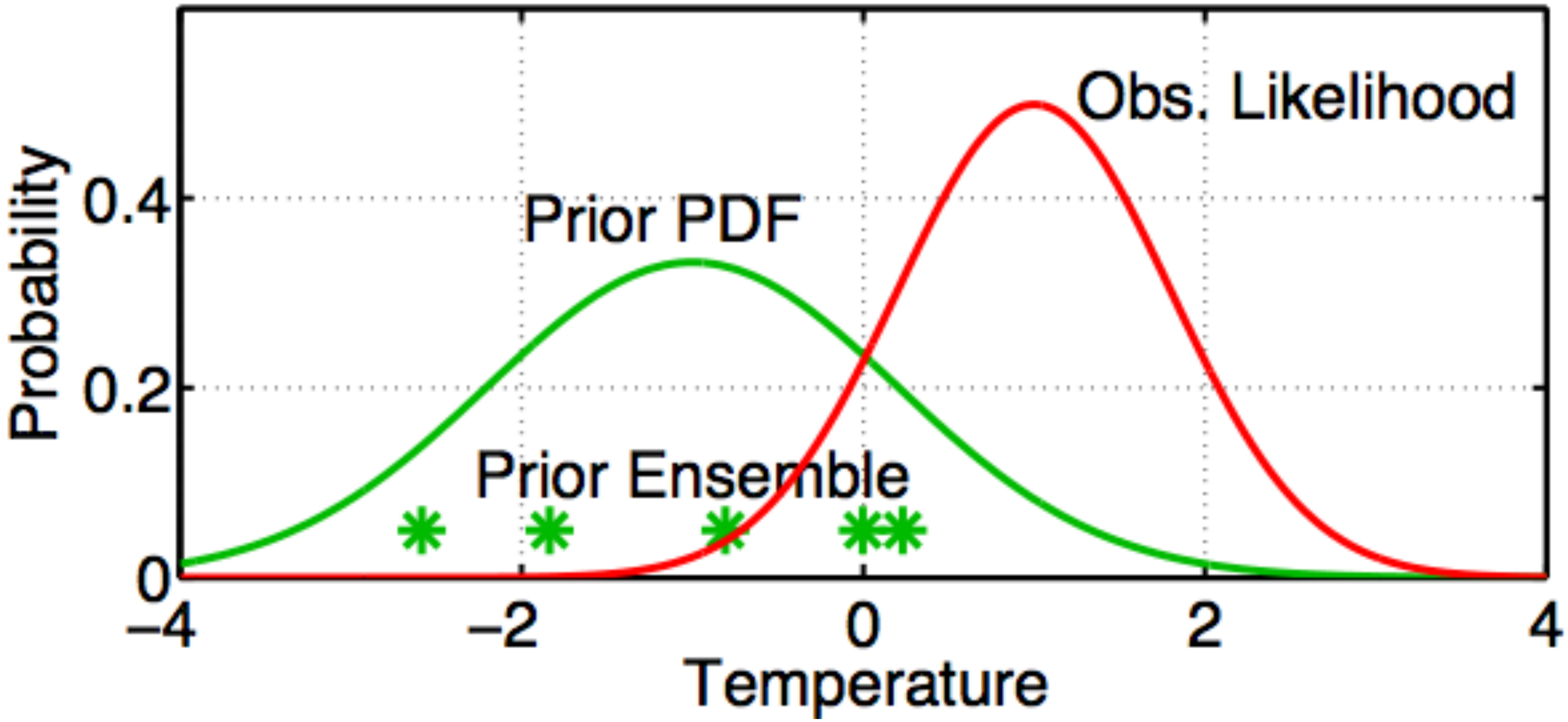


A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



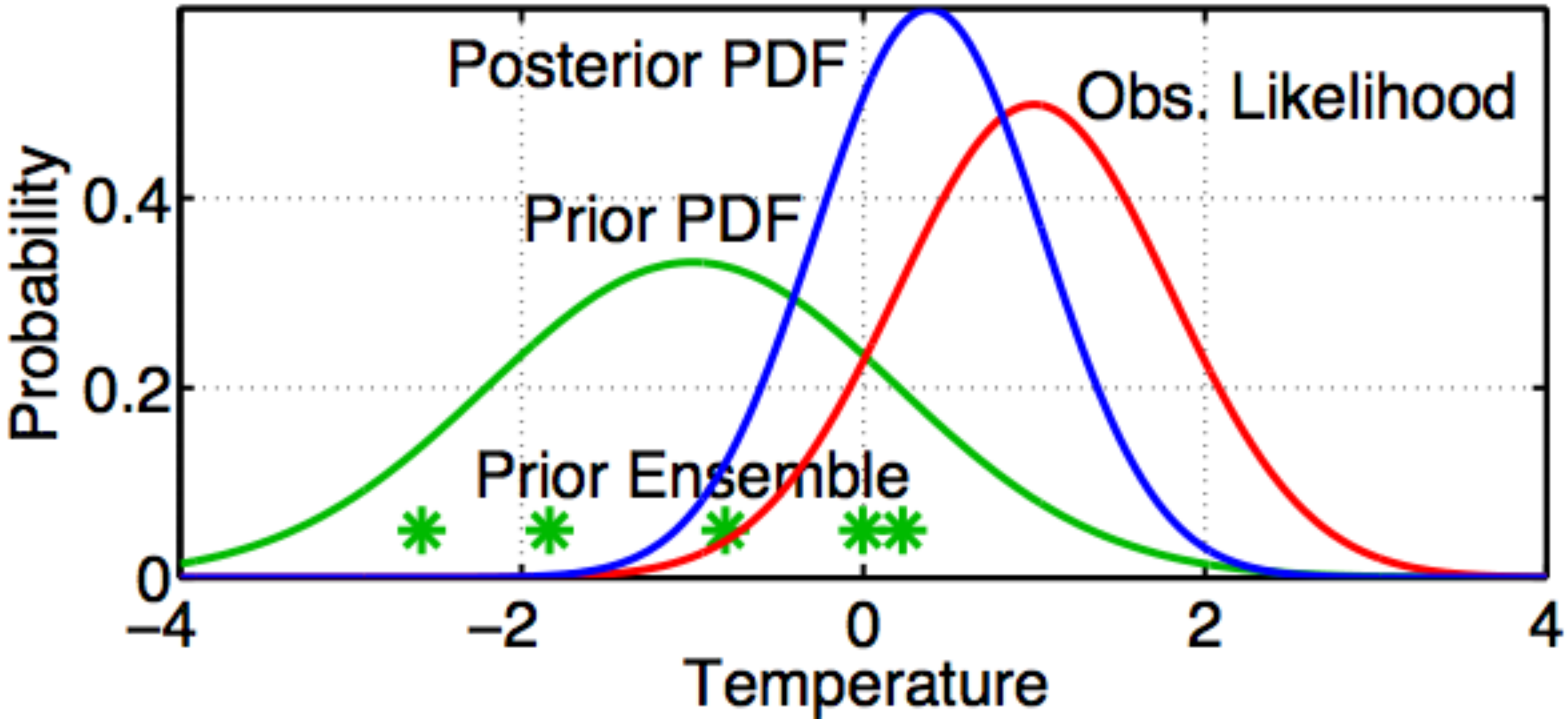
Fit a Gaussian to the sample.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



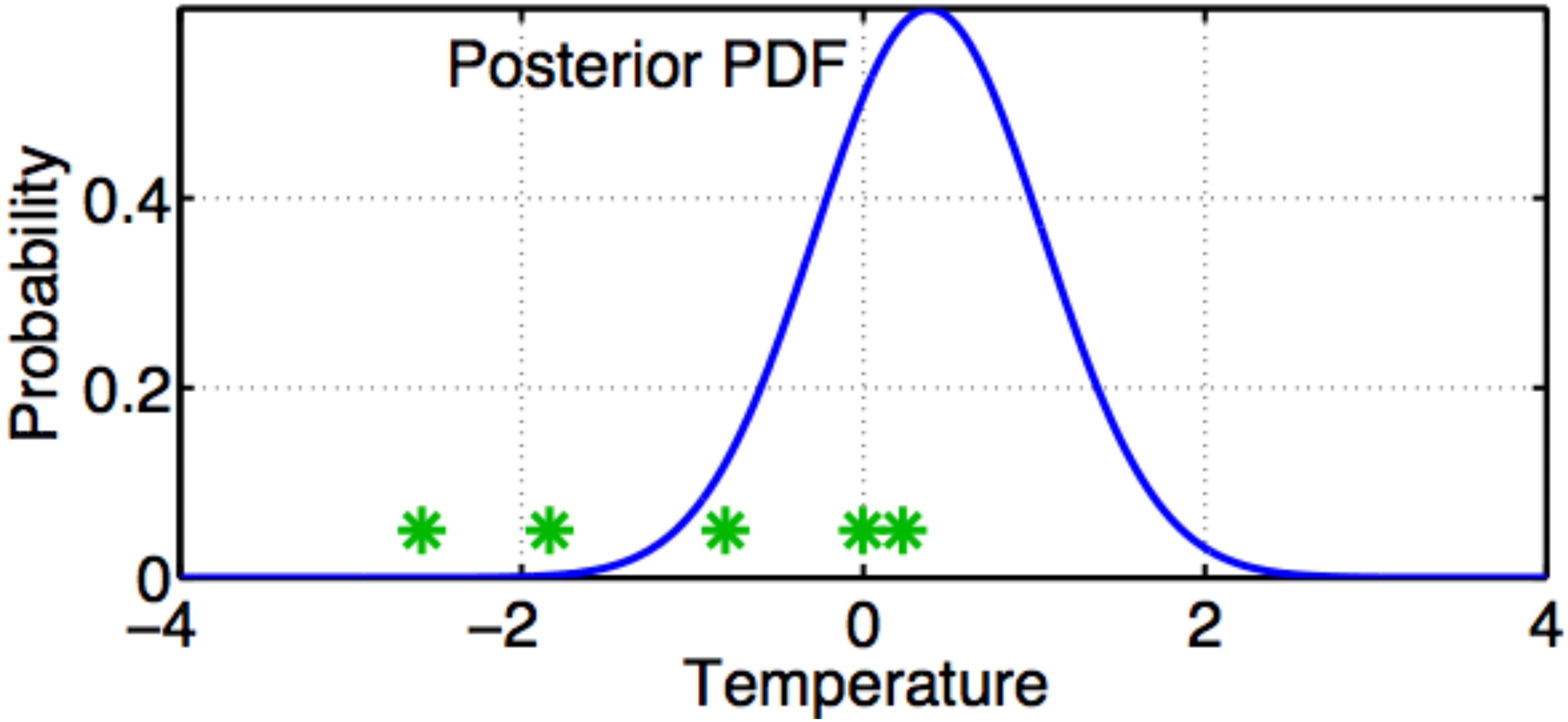
Get the observation likelihood.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



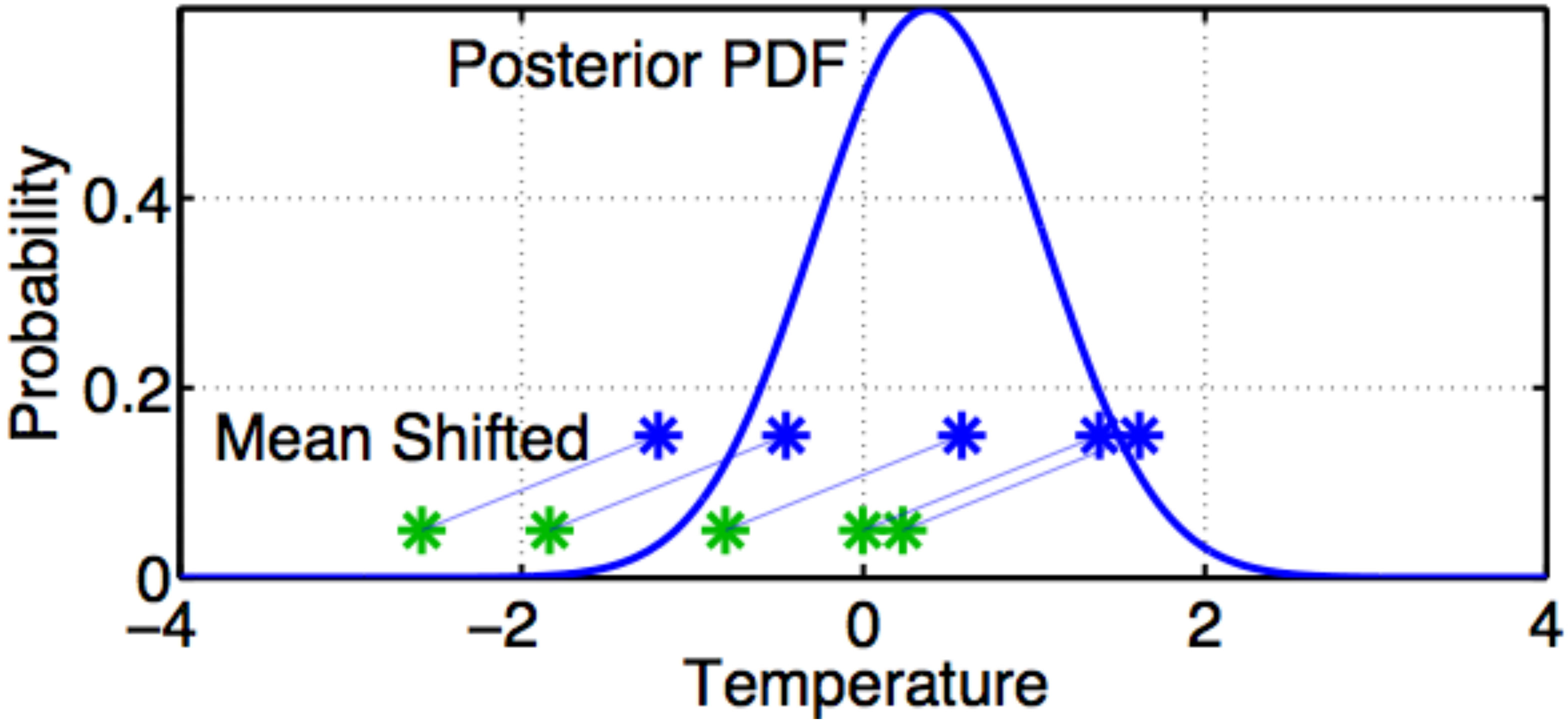
Compute the continuous posterior PDF.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



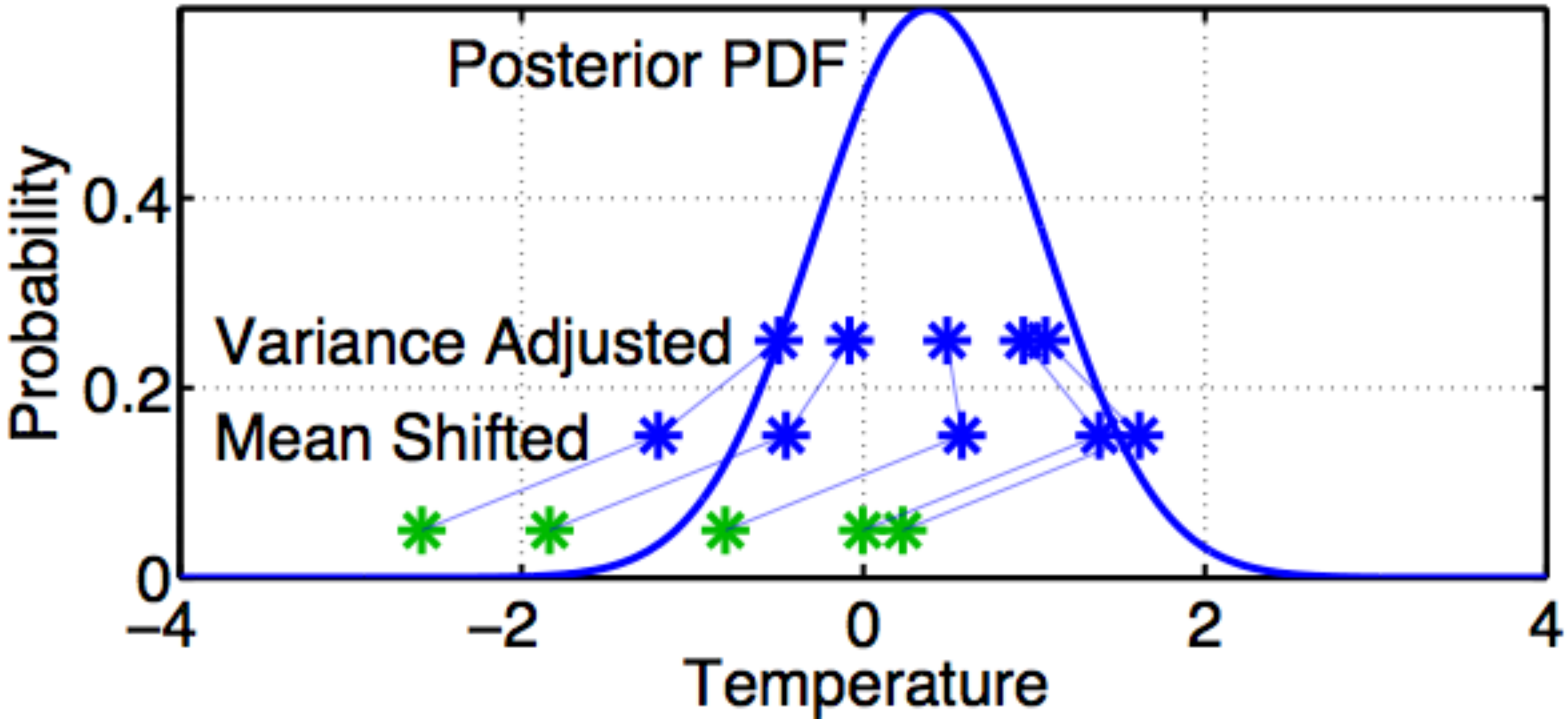
Use a deterministic algorithm to ‘adjust’ the ensemble.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



First, 'shift' the ensemble to have the exact mean of the posterior.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



First, ‘shift’ the ensemble to have the exact mean of the posterior.

Second, linearly contract to have the exact variance of the posterior.

Sample statistics are identical to Kalman filter.

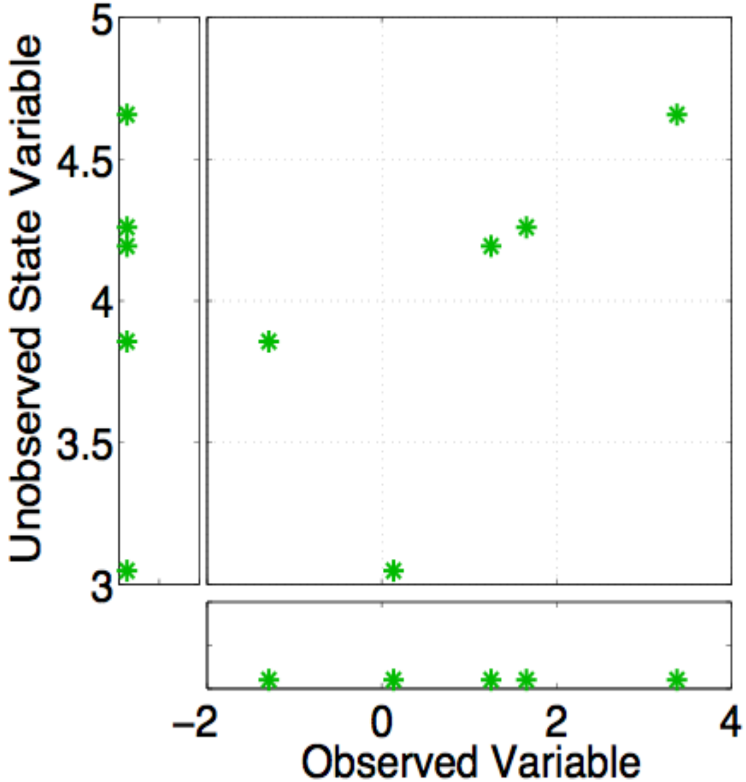
Multivariate Ensemble Kalman Filter

So far, we have an observation likelihood for single variable.

Suppose the model prior has additional variables.

Use linear regression to update additional variables.

Ensemble filters: Updating additional prior state variables

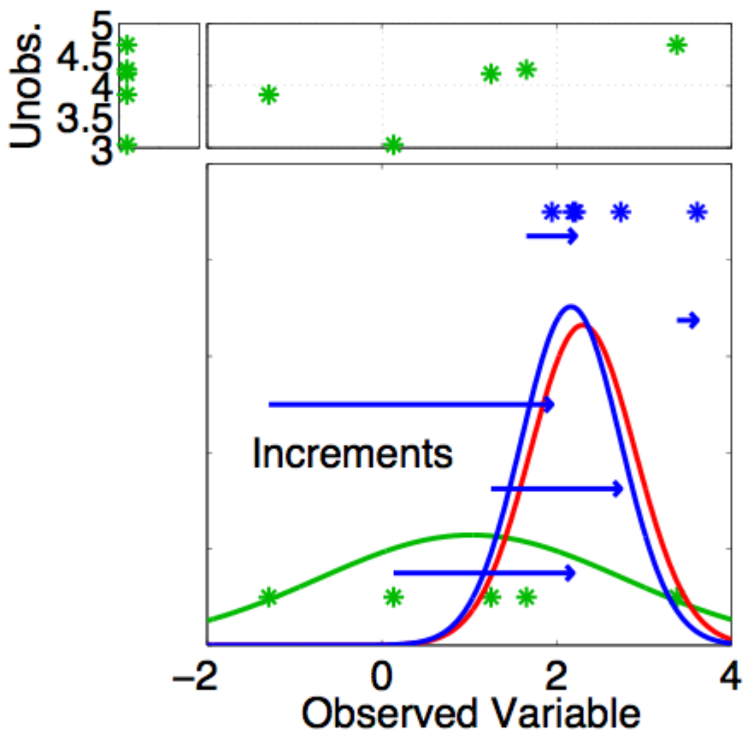


Assume that all we know is prior joint distribution.

One variable is observed.

What should happen to the unobserved variable?

Ensemble filters: Updating additional prior state variables

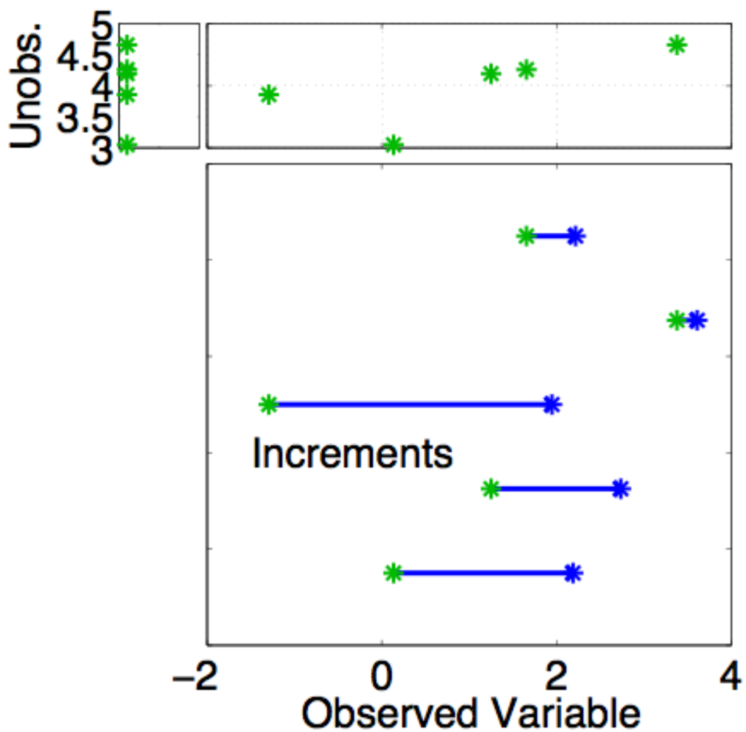


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

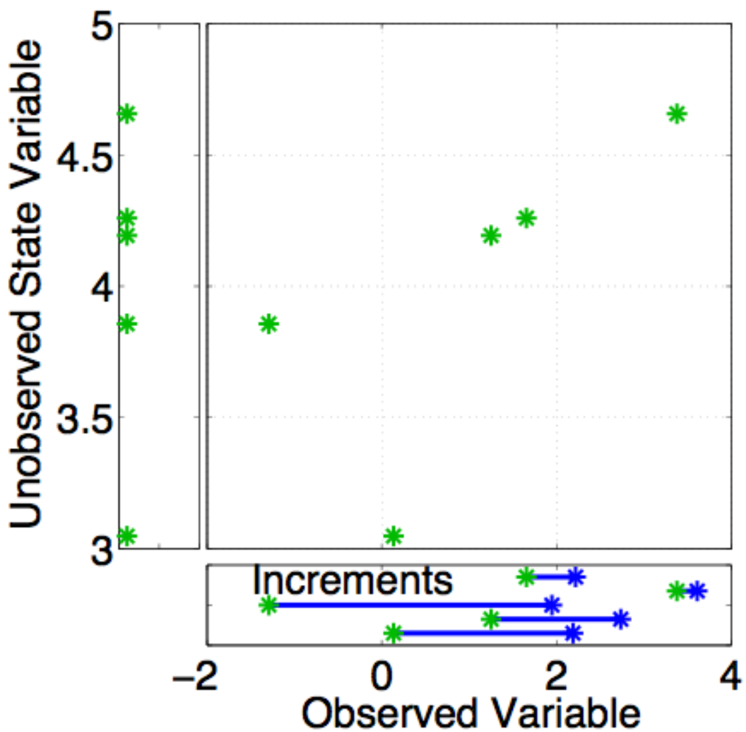


Assume that all we know is prior joint distribution.

One variable is observed.

Using only increments guarantees that if observation had no impact on observed variable, unobserved variable is unchanged (highly desirable).

Ensemble filters: Updating additional prior state variables



Assume that all we know is prior joint distribution.

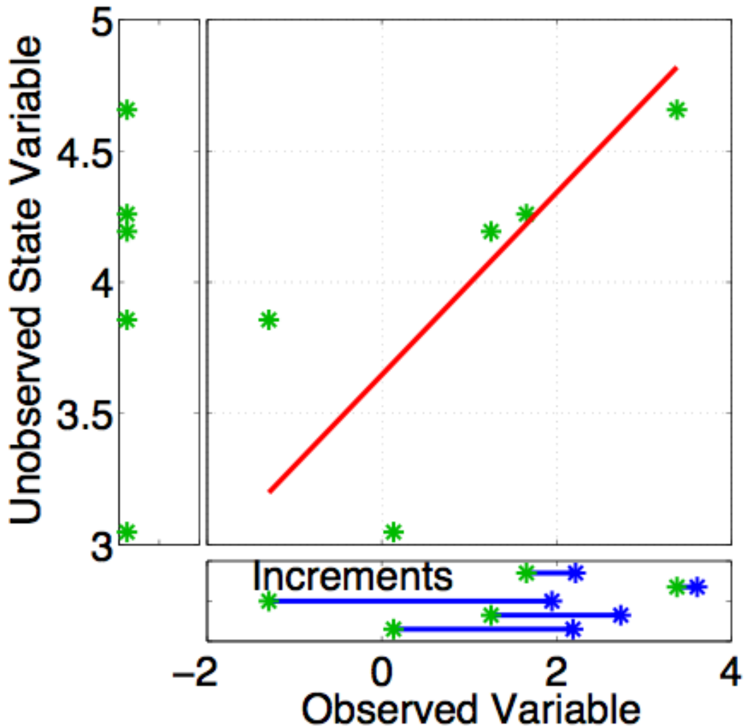
How should the unobserved variable be impacted?

First choice: least squares.

Equivalent to linear regression.

Same as assuming binormal prior.

Ensemble filters: Updating additional prior state variables



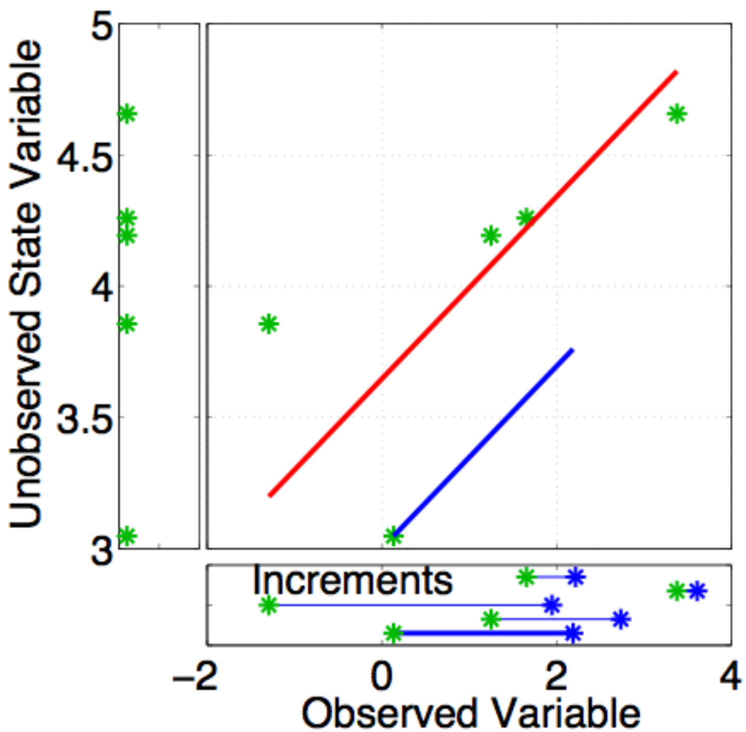
Have joint prior distribution of two variables.

How should the unobserved variable be impacted?

First choice: least squares.

Begin by finding least squares fit.

Ensemble filters: Updating additional prior state variables

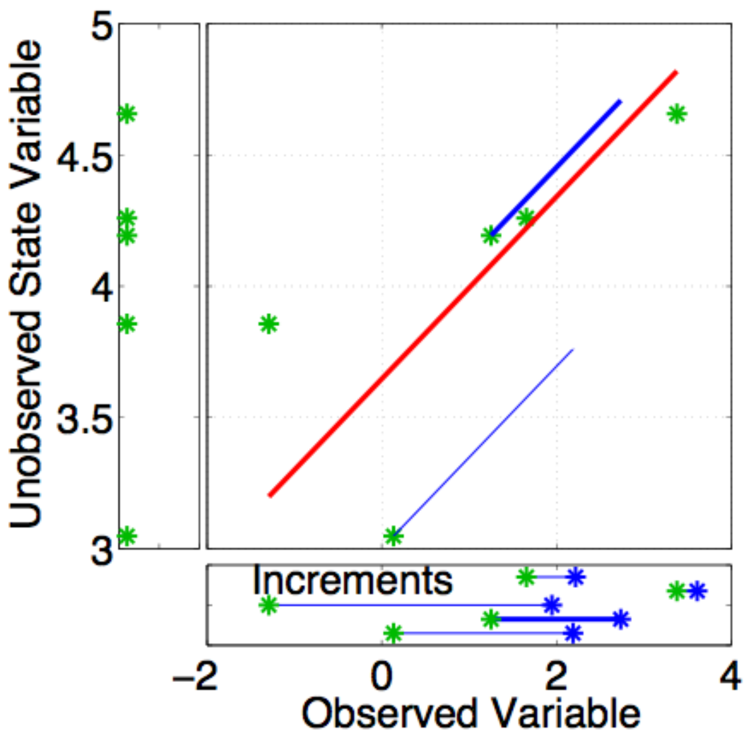


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

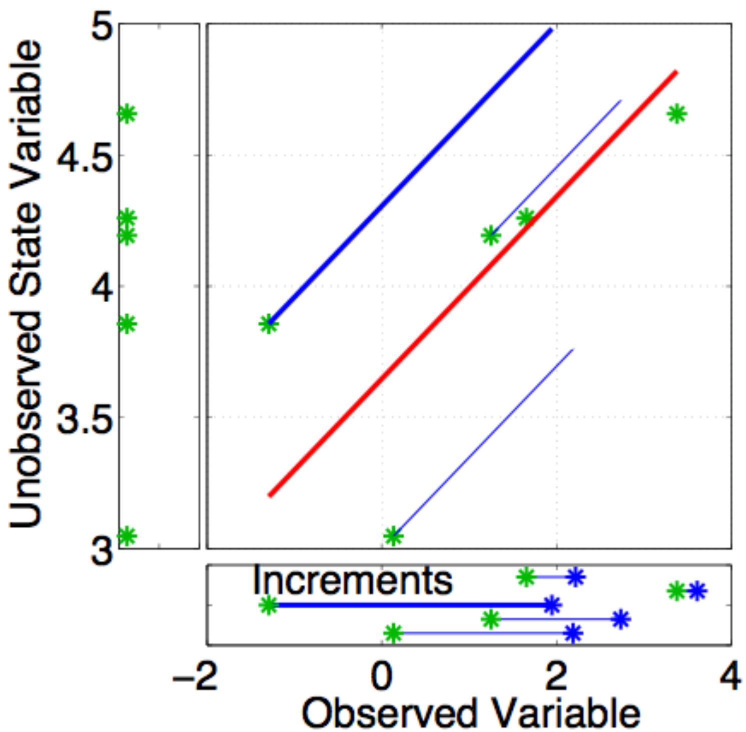


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

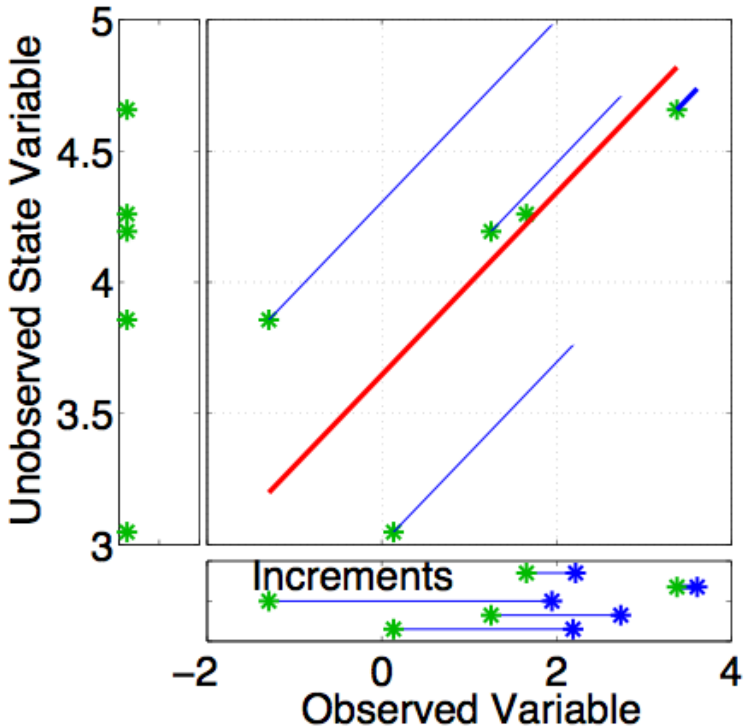


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

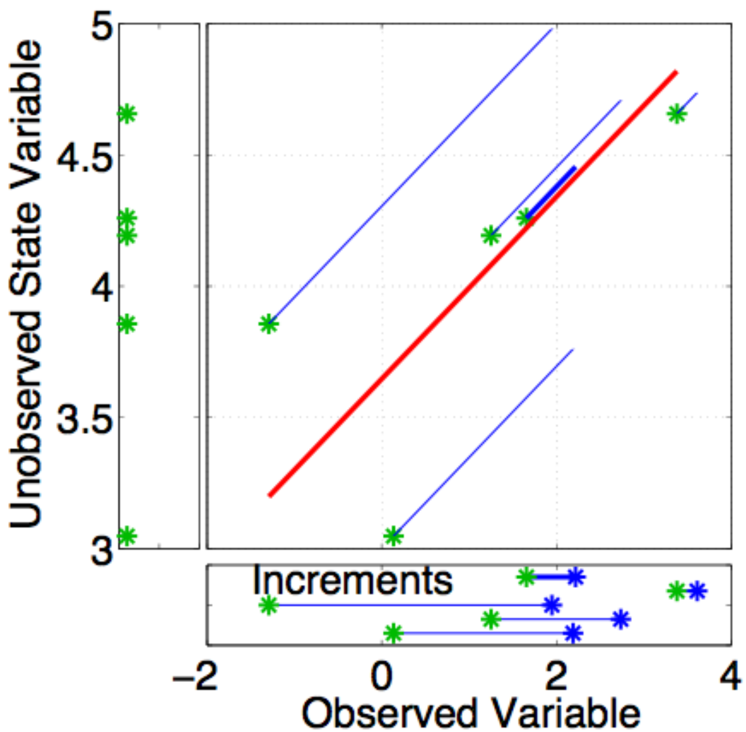


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

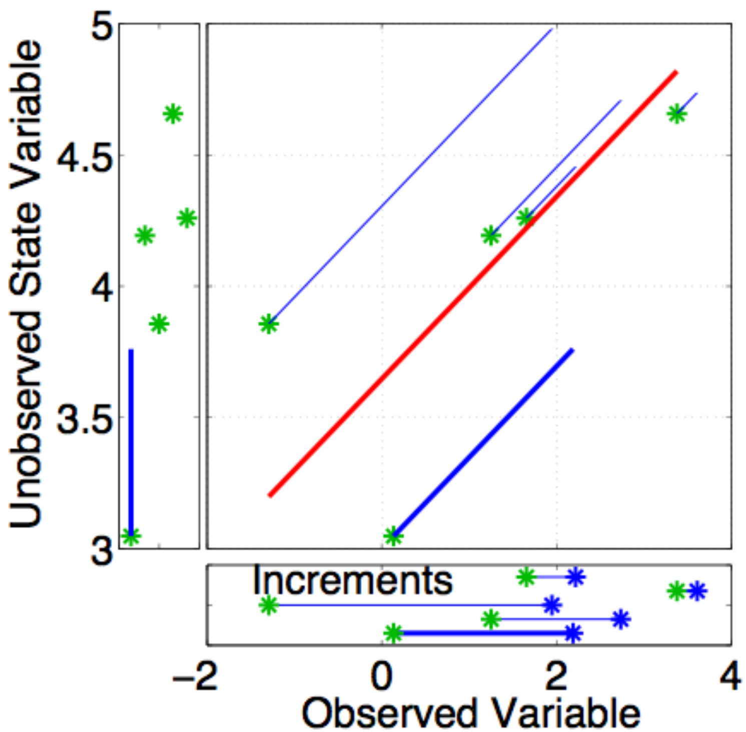


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

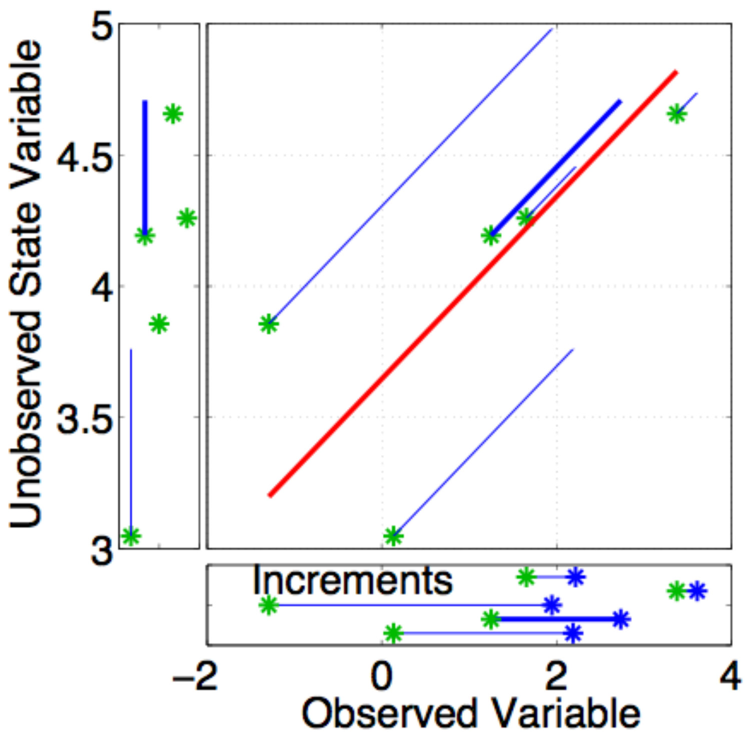


Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

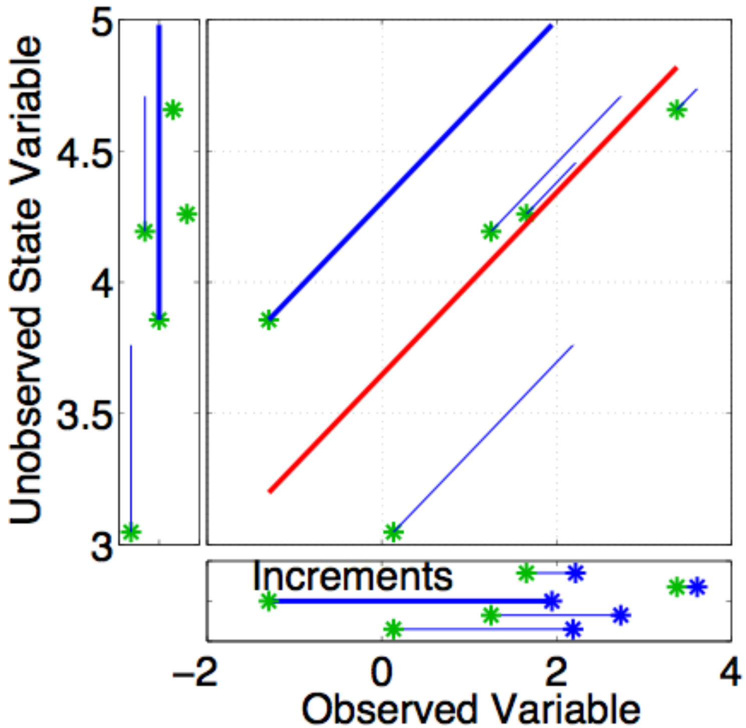


Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

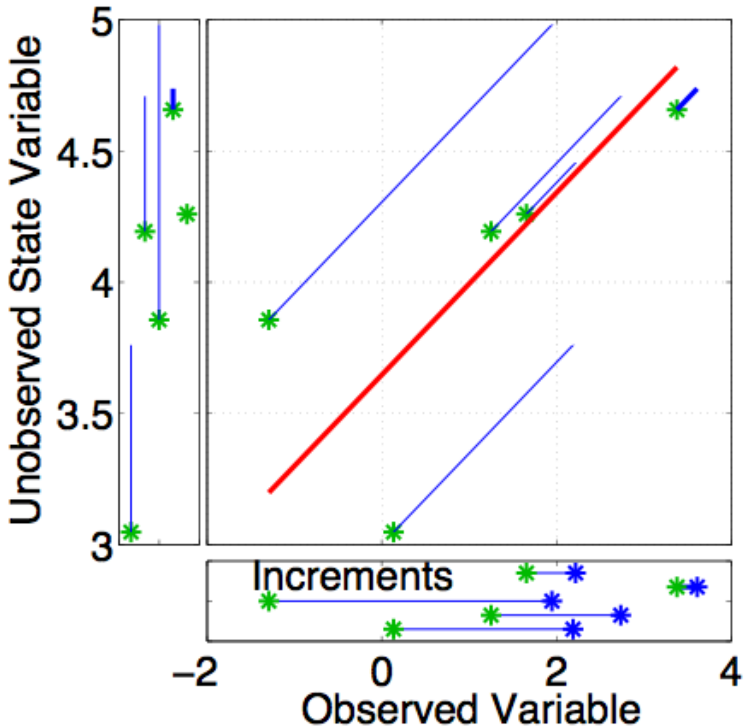


Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

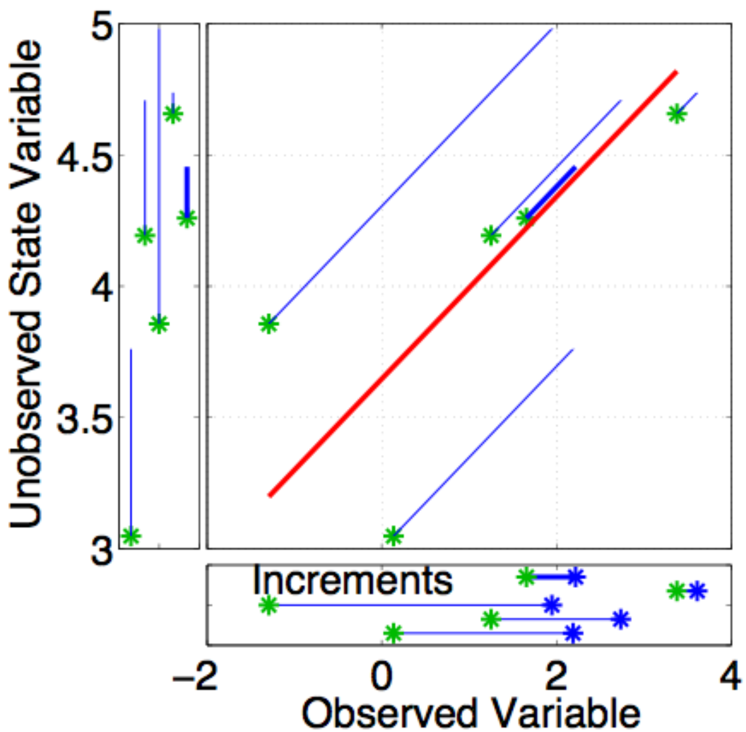


Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

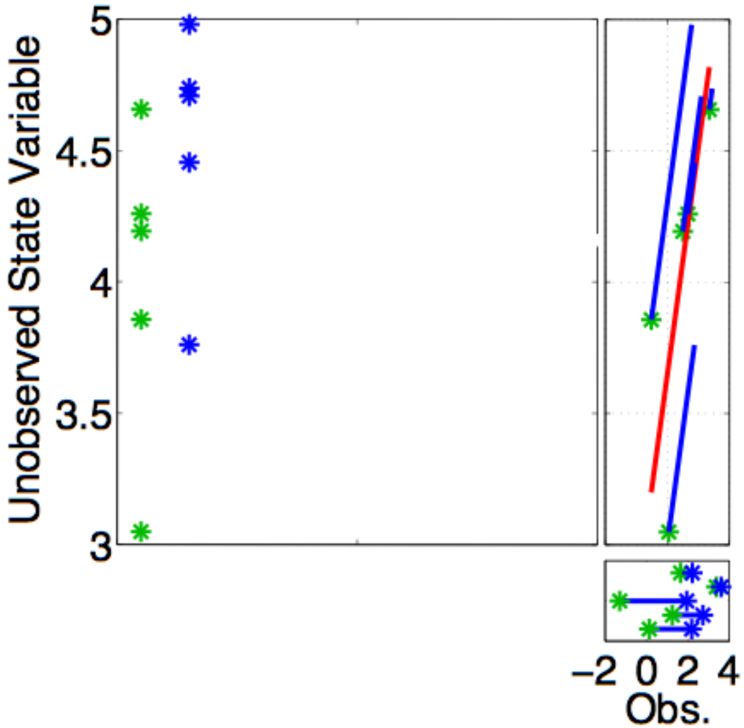


Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

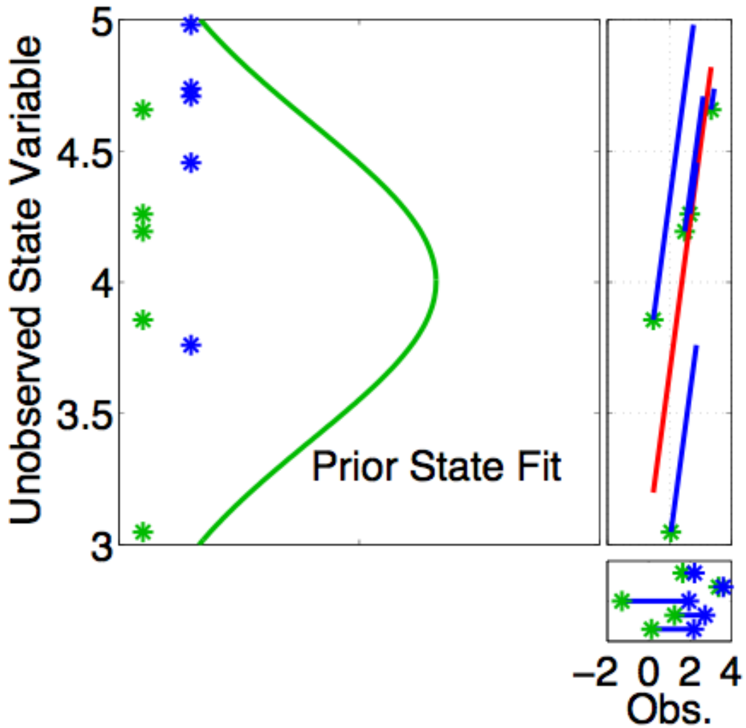
Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables



Now have an updated (posterior) ensemble for the unobserved variable.

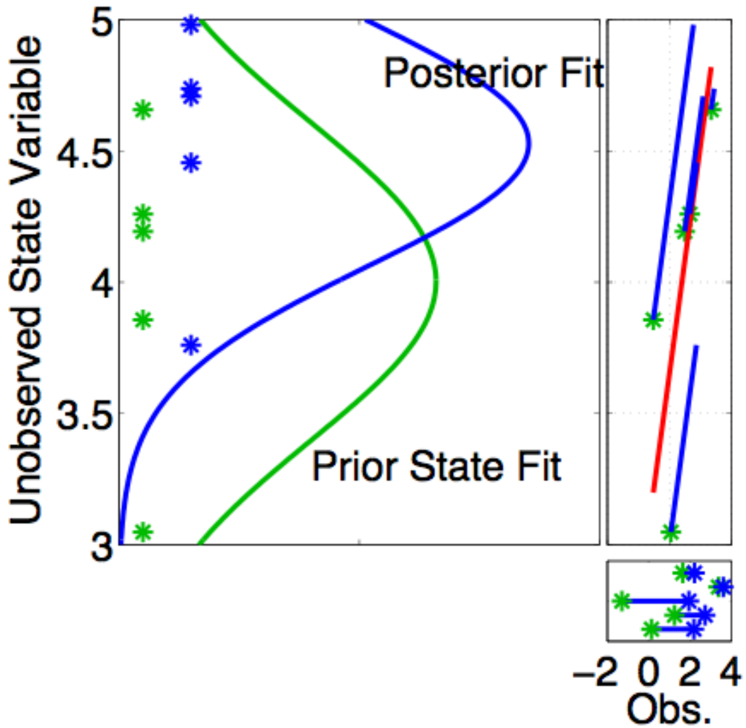
Ensemble filters: Updating additional prior state variables



Now have an updated (posterior) ensemble for the unobserved variable.

Fitting Gaussians shows that mean and variance have changed.

Ensemble filters: Updating additional prior state variables



Now have an updated (posterior) ensemble for the unobserved variable.

Fitting Gaussians shows that mean and variance have changed.

Other features of the prior distribution may also have changed.

Ensemble Filter for Large Geophysical Models

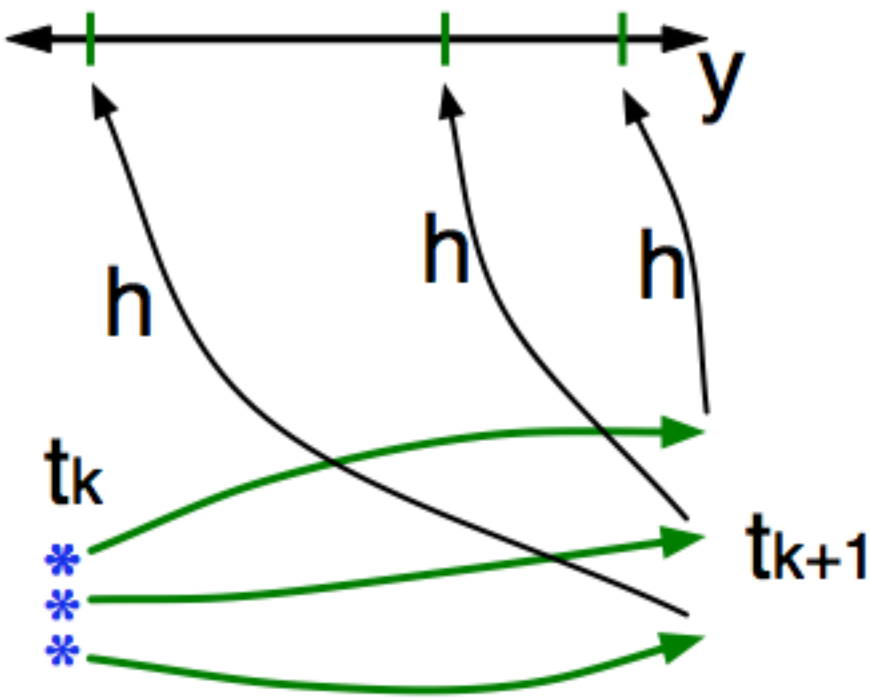
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Ensemble Filter for Large Geophysical Models

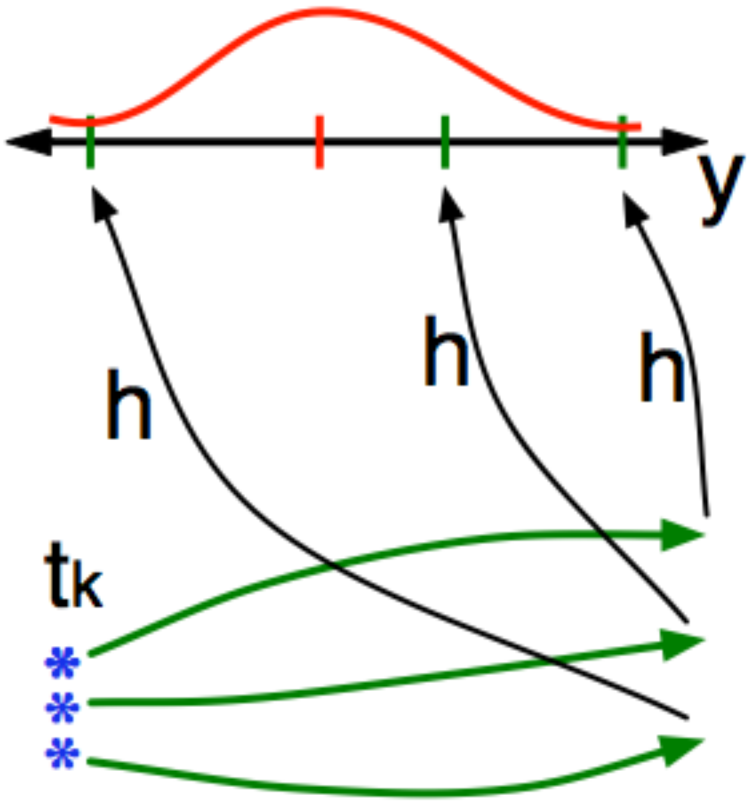
2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.



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

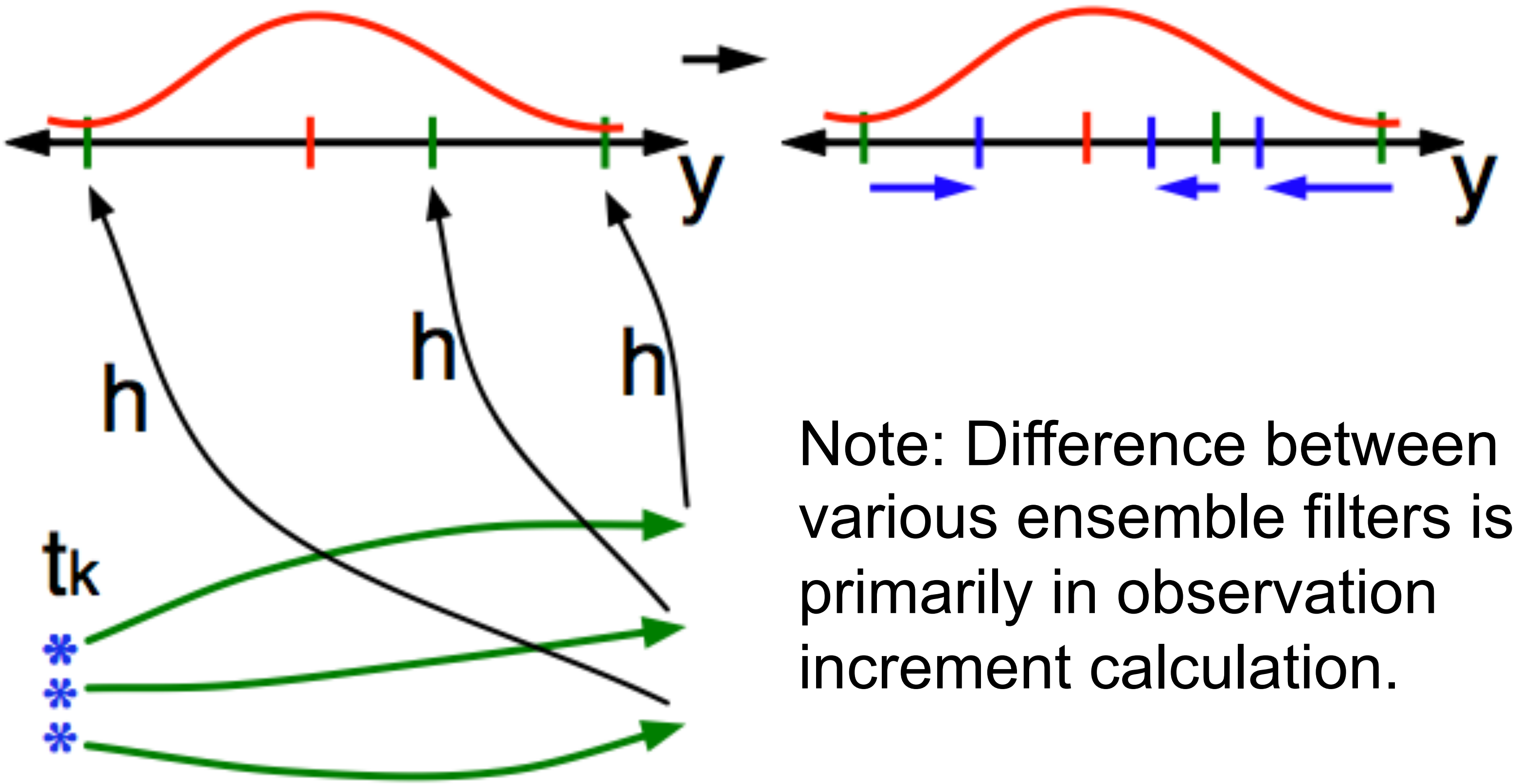
Ensemble Filter for Large Geophysical Models

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



Ensemble Filter for Large Geophysical Models

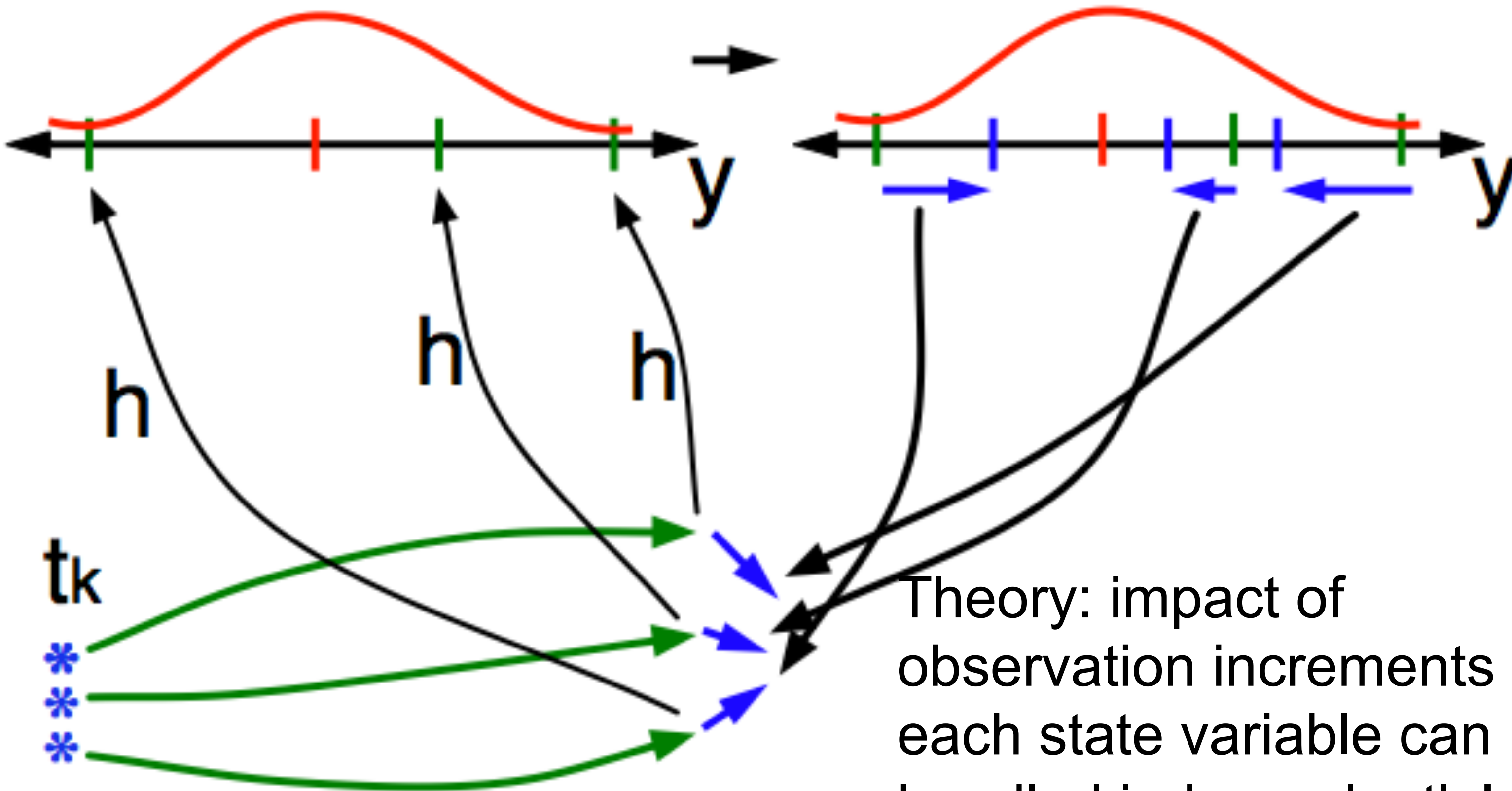
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



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

Ensemble Filter for Large Geophysical Models

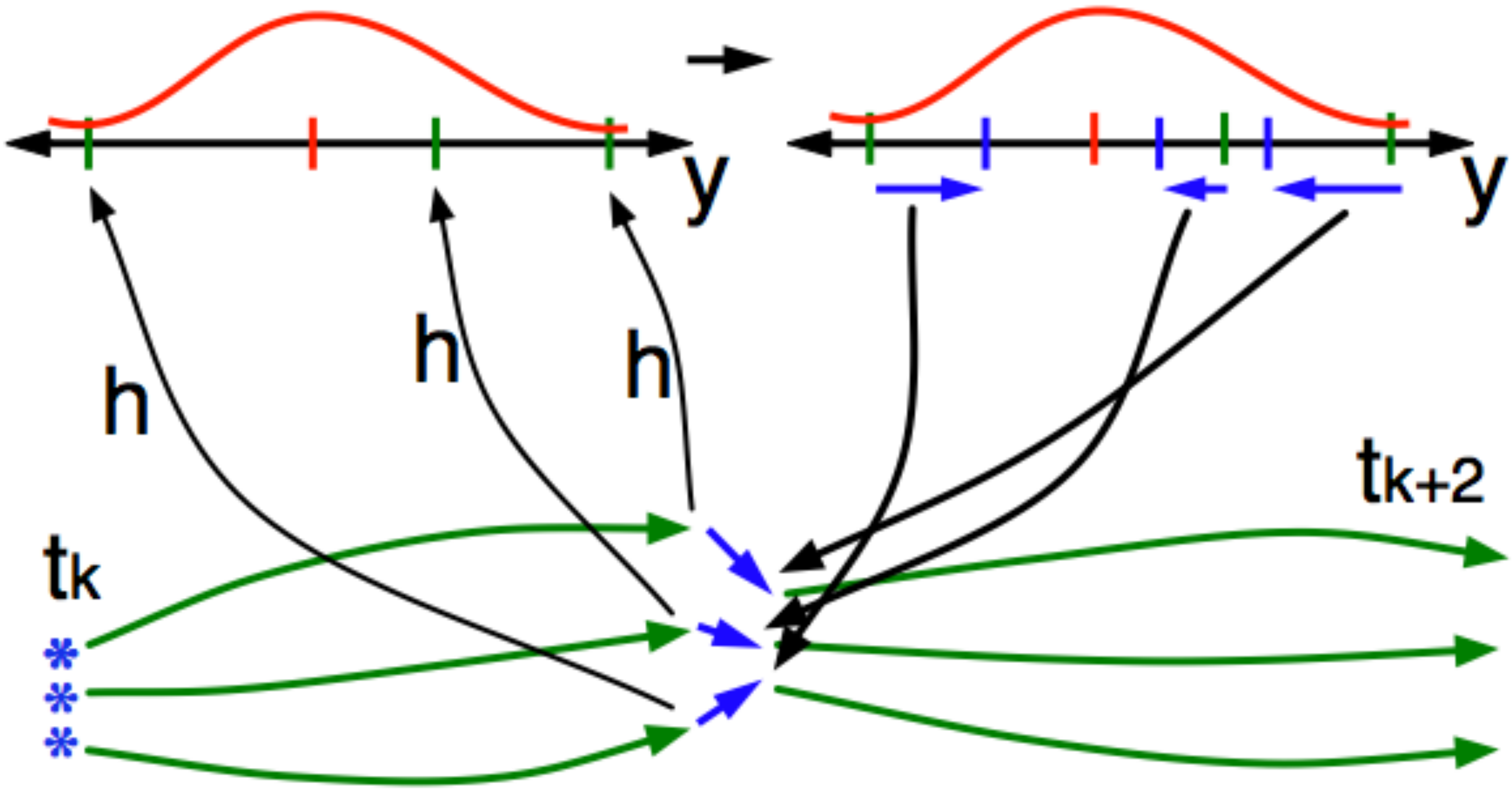
5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



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

Ensemble Filter for Large Geophysical Models

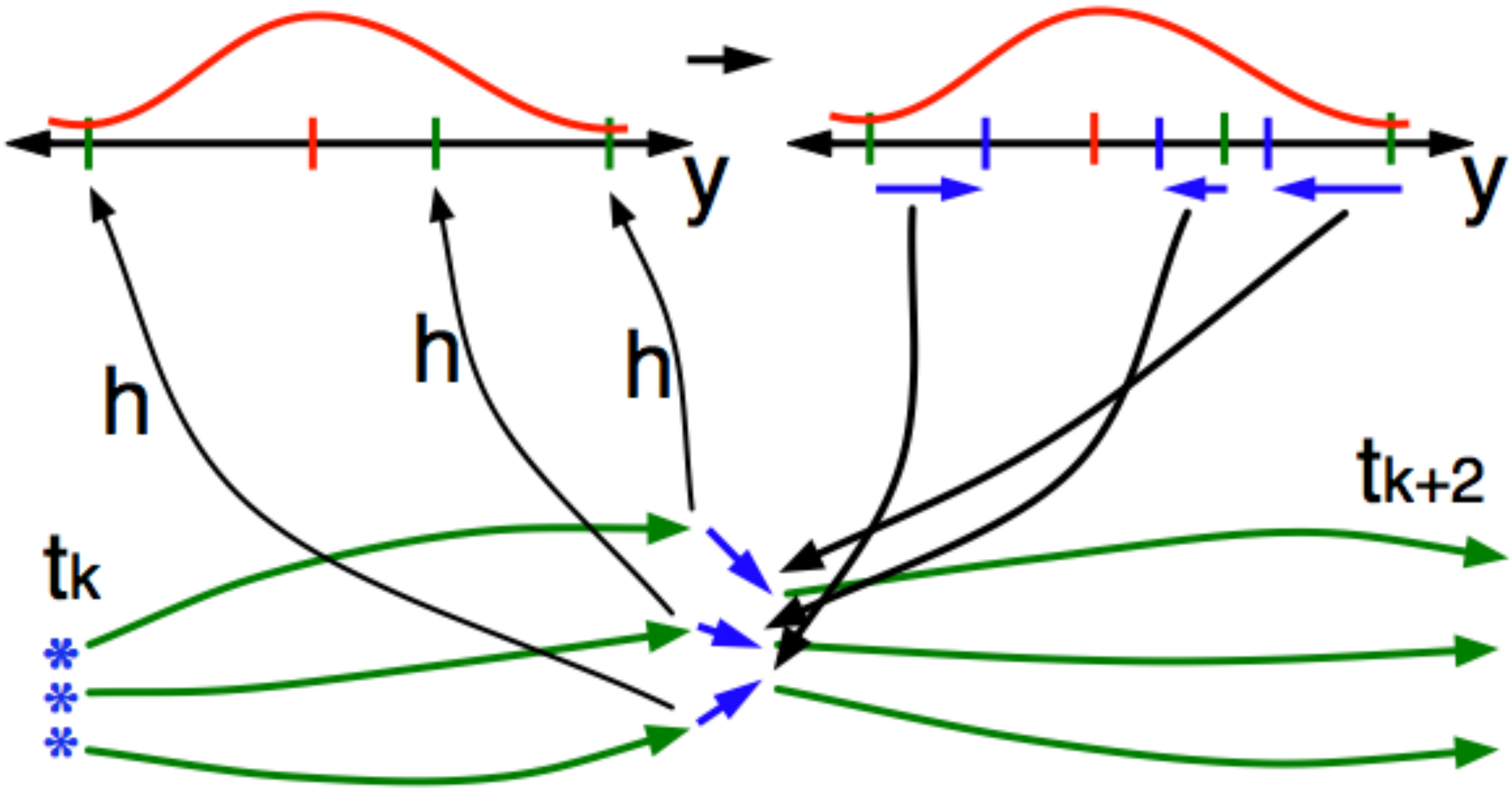
6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation …



Ensemble Filter for Large Geophysical Models

A generic ensemble filter system like DART just needs:

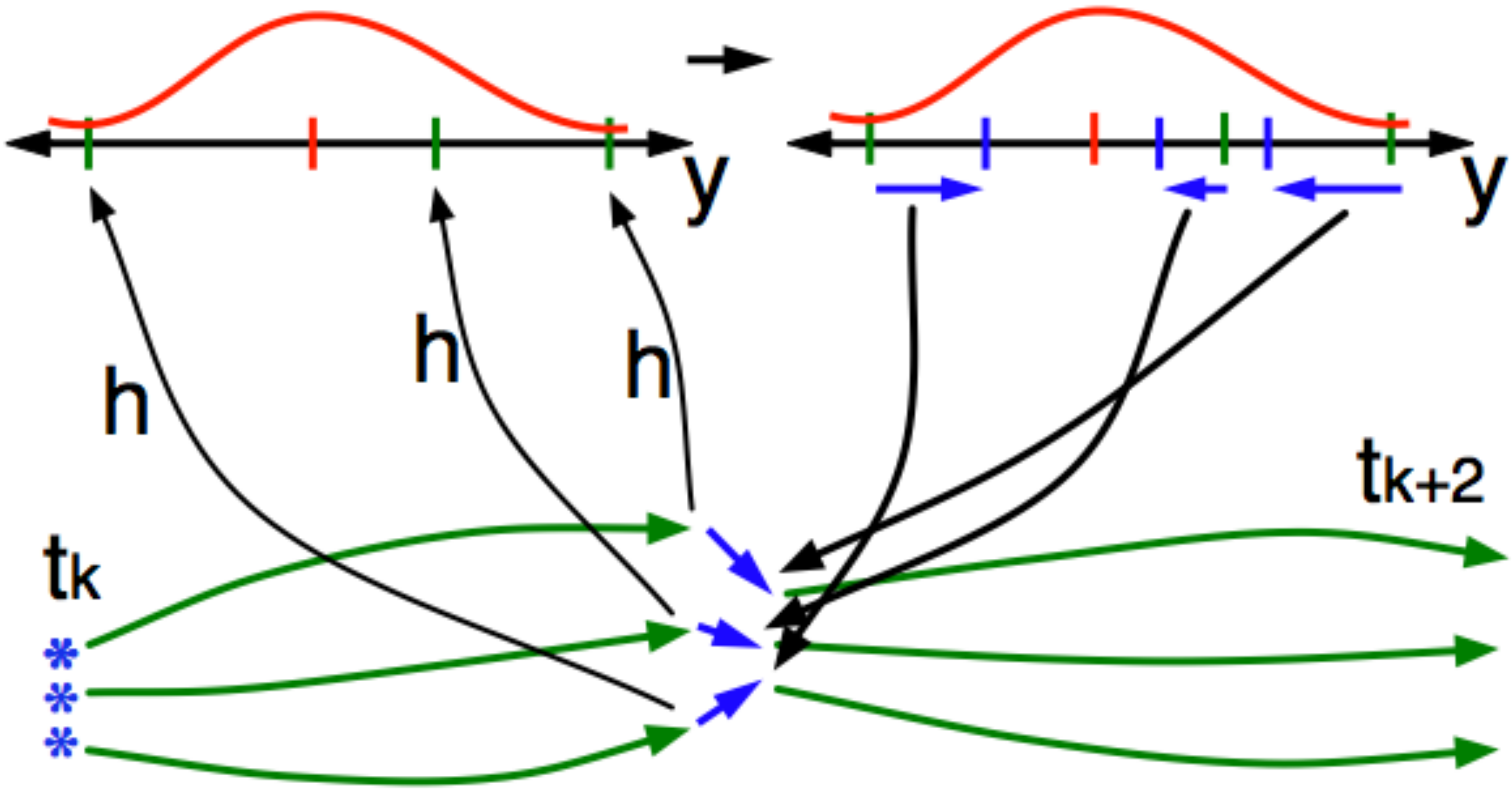
1. A way to make model forecasts;



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.



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

UNIVERSITY OF WASHINGTON

CIRES

THE UNIVERSITY OF UTAH

中央氣象局 Central Weather Bureau

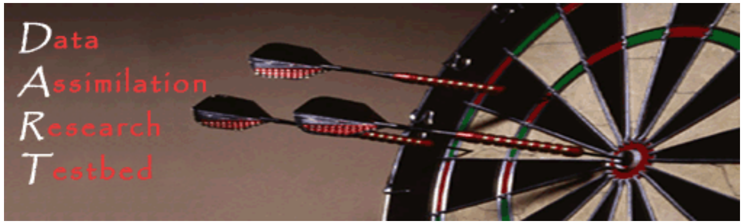
GFDL Geophysical Fluid Dynamics Laboratory Princeton, NJ

NOAA

PENNSTATE

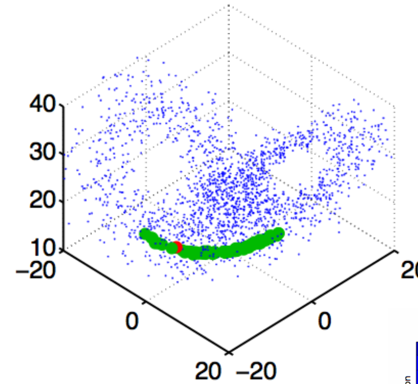
CALIFORNIA INSTITUTE OF TECHNOLOGY

JPL Jet Propulsion Laboratory California Institute of Technology

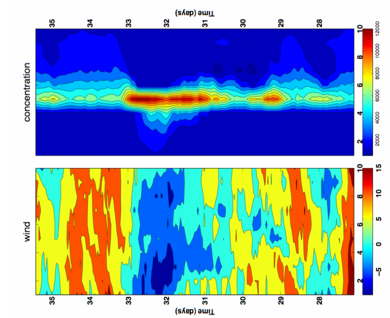


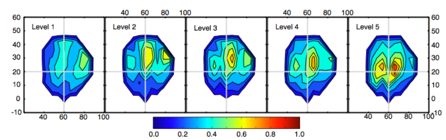
DART is:

Education

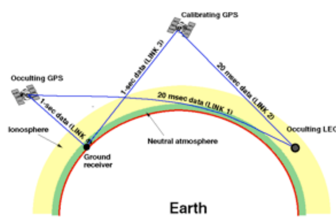


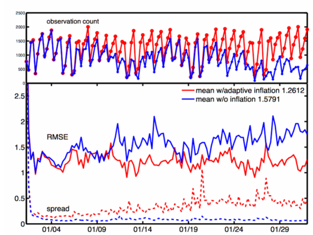
Exploration



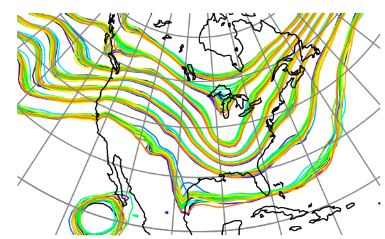


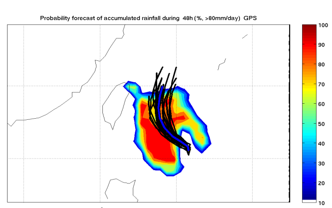
Research





Operations





DART works with many geophysical models

<u>Global Atmosphere models</u>:

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
COSMO	Consortium for Small-Scale Modeling	DWD

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

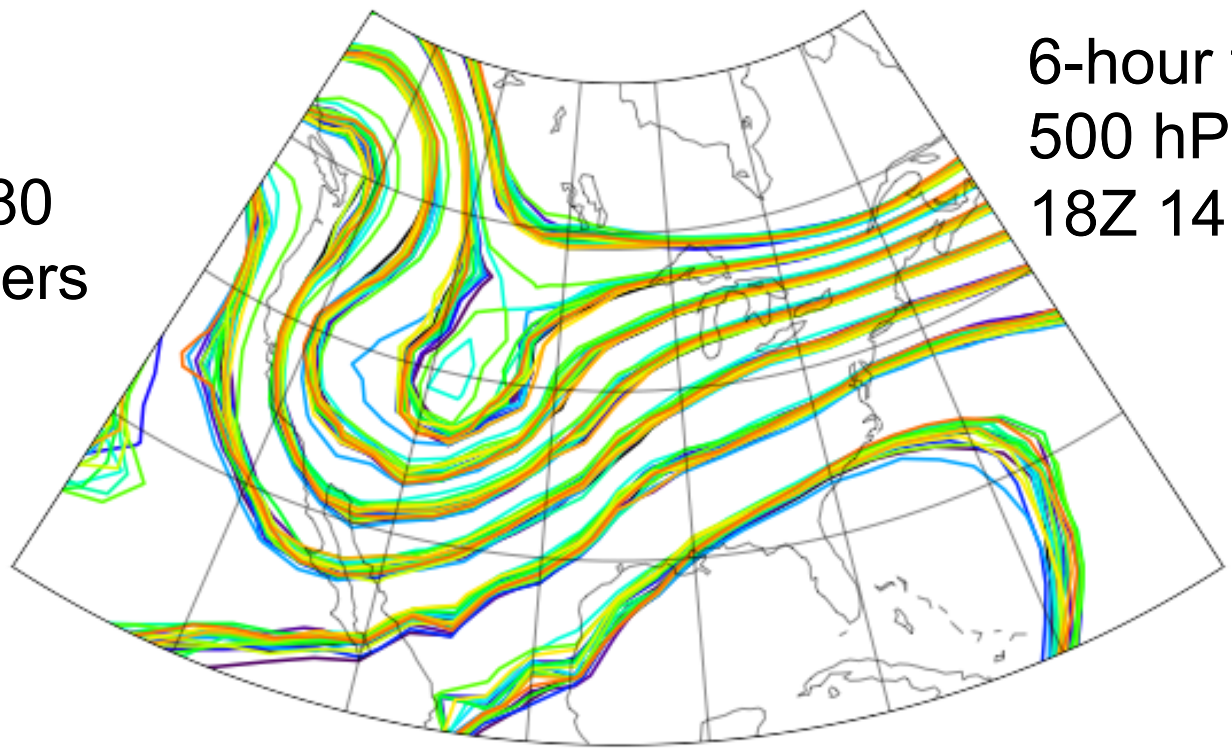
Land Surface models:

CLM	Community Land Model	NCAR

Focus on DART Science with CAM

Basic Capability: Ensemble Analyses and Forecasts
Works for all CAM versions since 2002

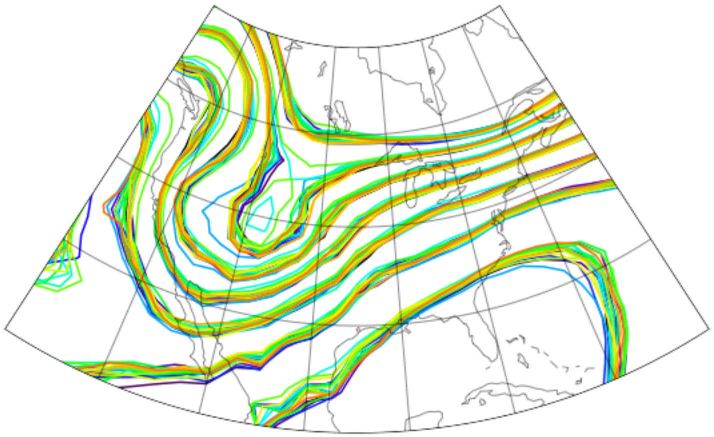
20 of 80 members



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

contours from 5400 to 5880 by 80

Ensemble Analyses and Forecasts



contours from 5400 to 5880 by 80

Sample collaborations:

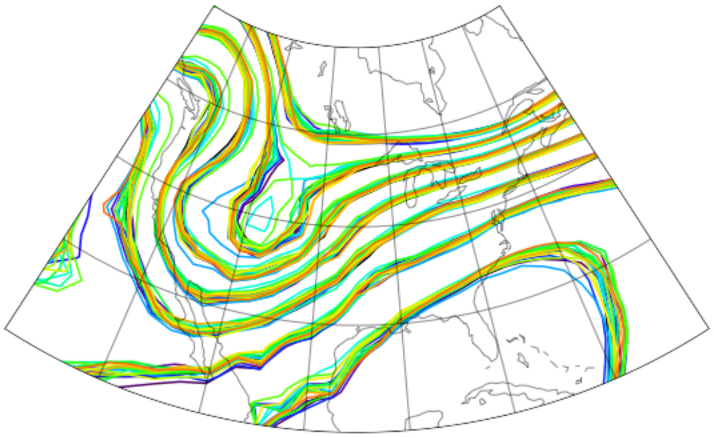
Edmund Chang, Stony Brook
Pacific storm track/cyclogenesis

Nedjeljka Zagar, Ljubljana University
Normal mode analysis of general circulation

Maria Tsukernik, Monash/Brown
Antarctic cyclones

Ensemble Analyses and Forecasts

Sample collaborations:

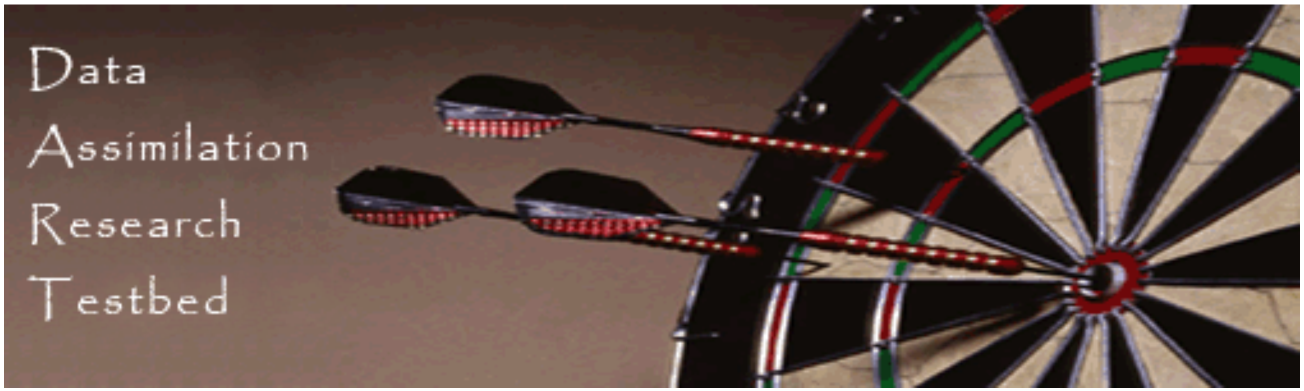

contours from 5400 to 5880 by 80

Rahul Mahajan, U. Washington
Real-time ensemble forecasts
For Pacific Northwest.

Ibrahim Hoteit, KAUST Saudi Arabia
Gulf of Mexico Ocean Prediction.

Ryan Torn, SUNY Albany
Real-time Atlantic hurricane forecasts.

Diagnosing and Correcting Errors in the CAM Finite Volume core with DART



Kevin Raeder*

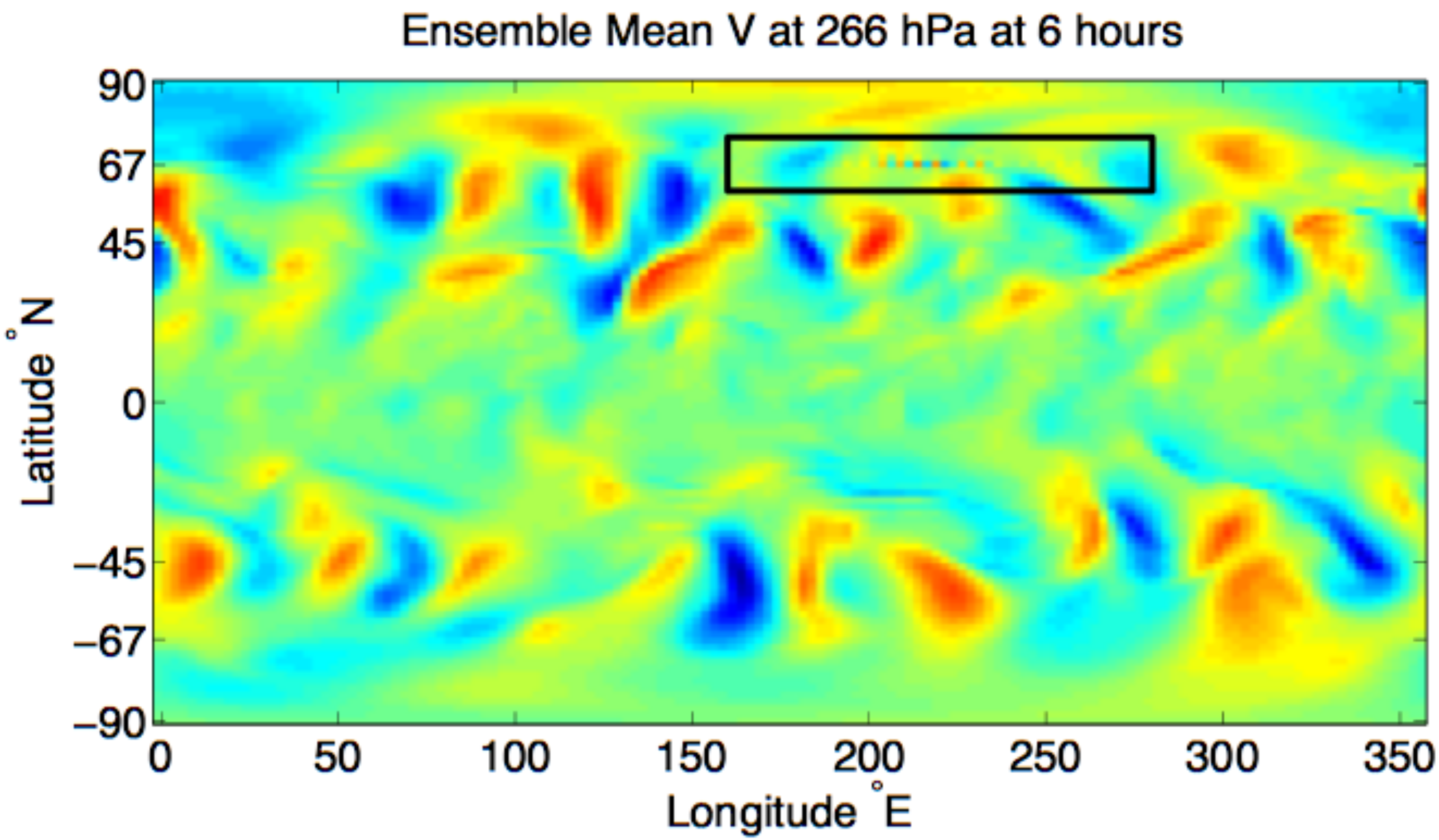
Jeff Anderson*

Peter Lauritzen+

Tim Hoar*

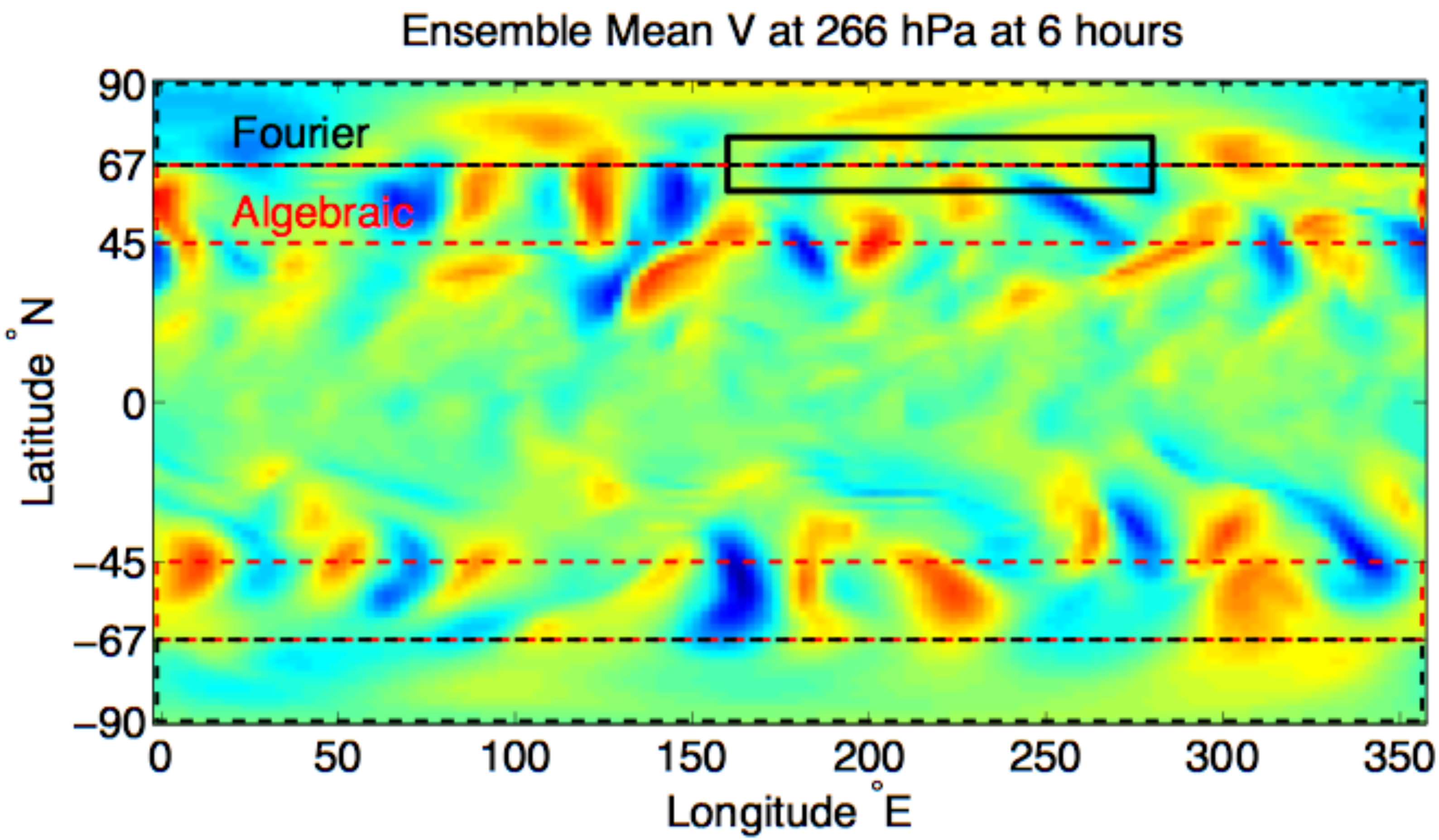
*NCAR/CISL/IMAGe/DAReS
+NCAR/ESSL/CGD/AMPS

Gridpoint noise detected in CAM/DART analysis



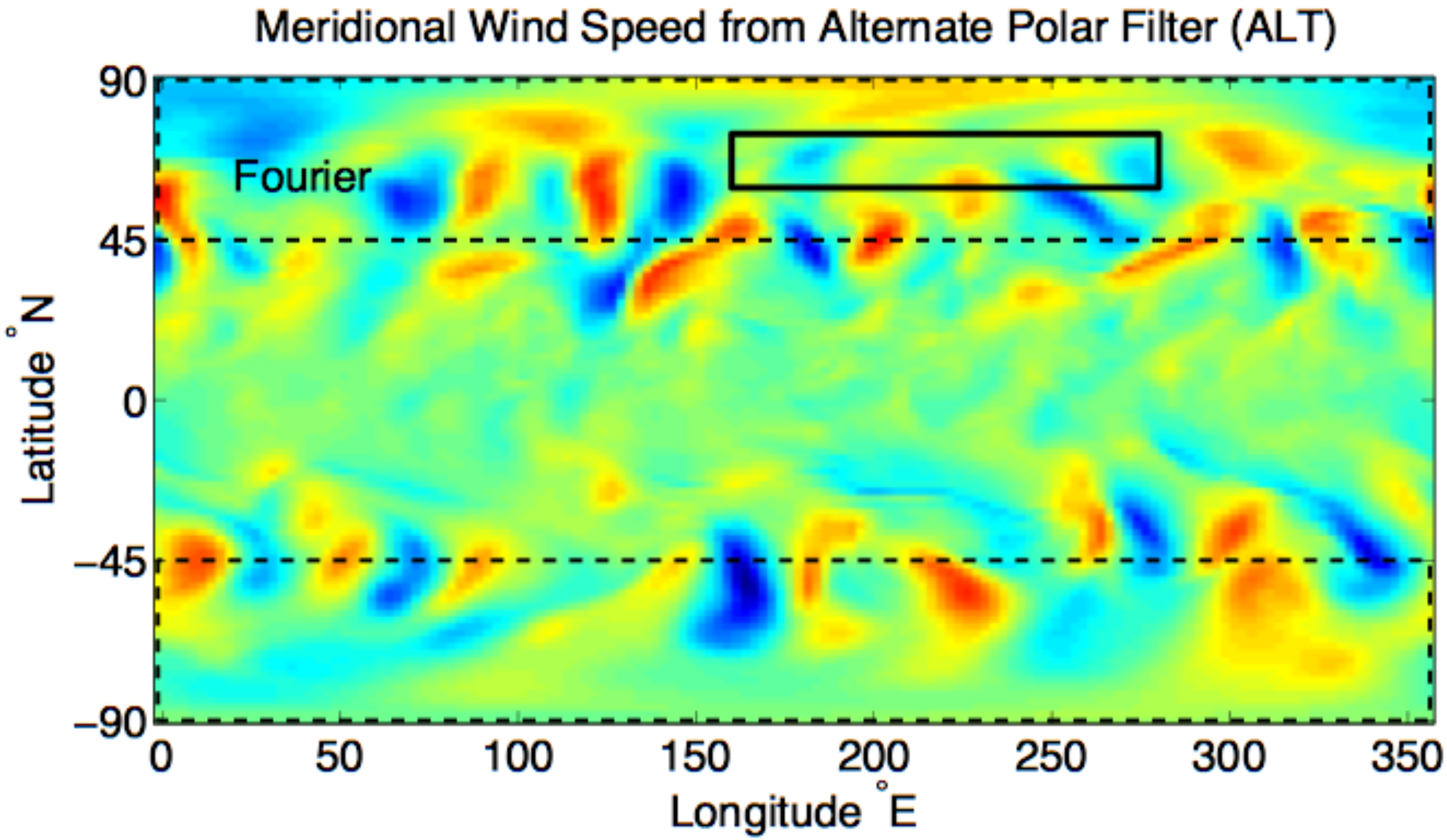
CAM FV core - 80 member mean - 00Z 25 September 2006

Suspicions turned to the polar filter (DPF)

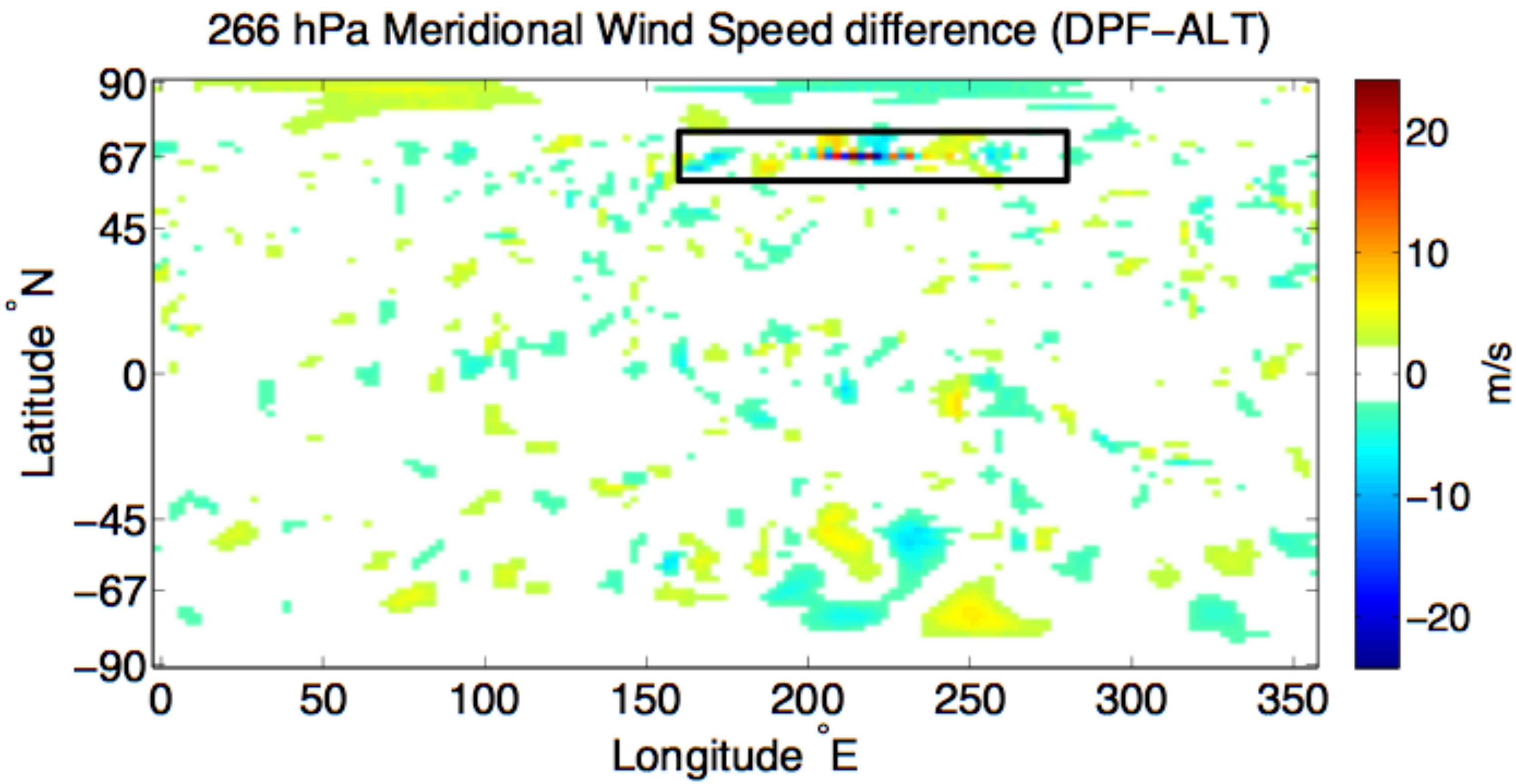


CAM FV core - 80 member mean - 00Z 25 September 2006

Continuous polar filter (alt-pft) eliminated noise.



Differences mostly in transition region of default filter.



Diagnosing and Correcting Errors in the CAM Finite Volume core with DART

- 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.

Cloud response to the 2007 Arctic sea ice loss in CAM3.5 and CAM4

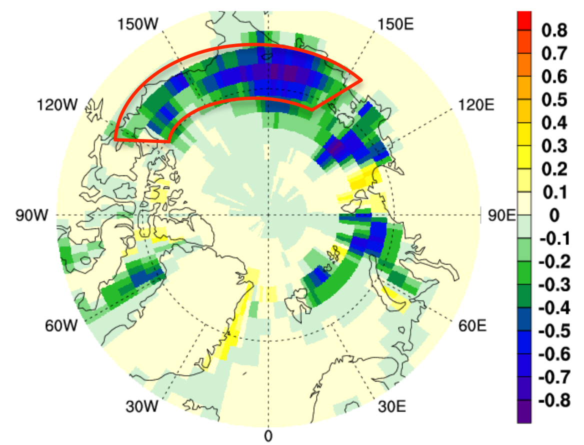
Jennifer E. Kay
National Center for Atmospheric Research (NCAR)
Colorado State University (CSU)
Collaborators: Julienne Stroeve (NSIDC),
Andrew Gettelman, Kevin Raeder, Jeff Anderson (NCAR),
Graeme Stephens, Tristan L'Ecuyer, Chris O'Dell (CSU)

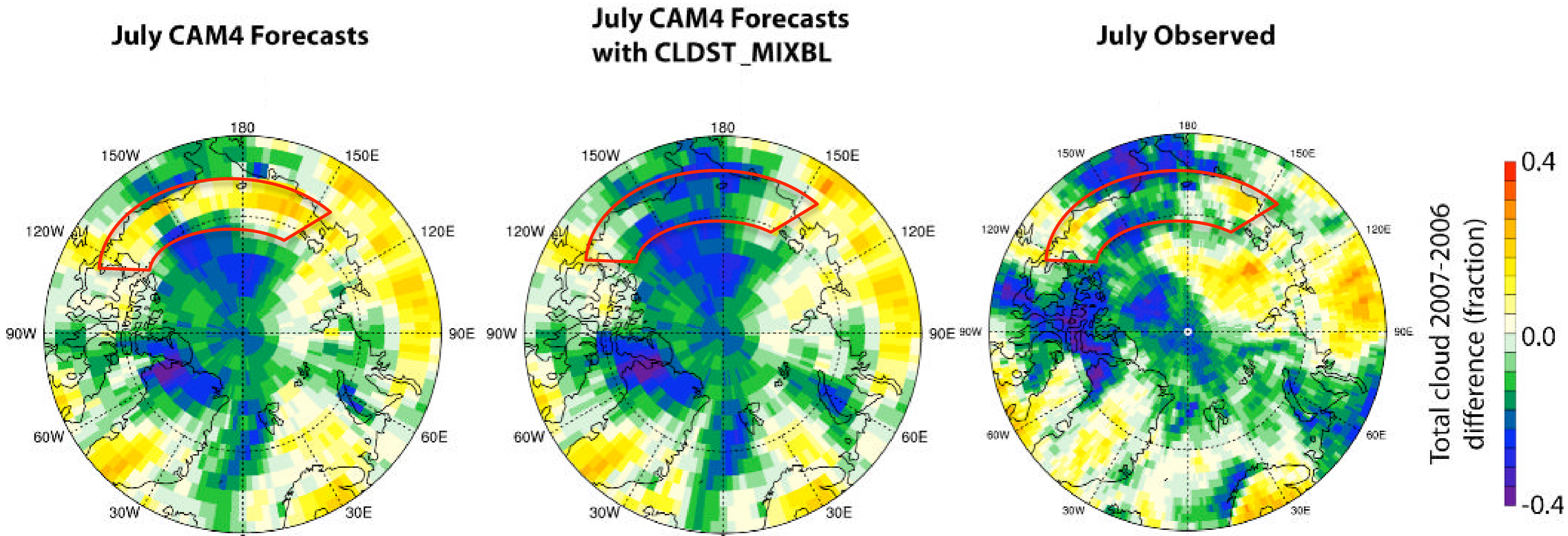
CAM4's cloud response to sea ice loss; July 2006 to 2007

24-hour forecasts started from DART/CAM analyses identified erroneous cloud response to disappearing sea ice.
Jen Kay found that low clouds were only diagnosed over open water, not ice, and the low cloud scheme should have required a well mixed boundary layer.

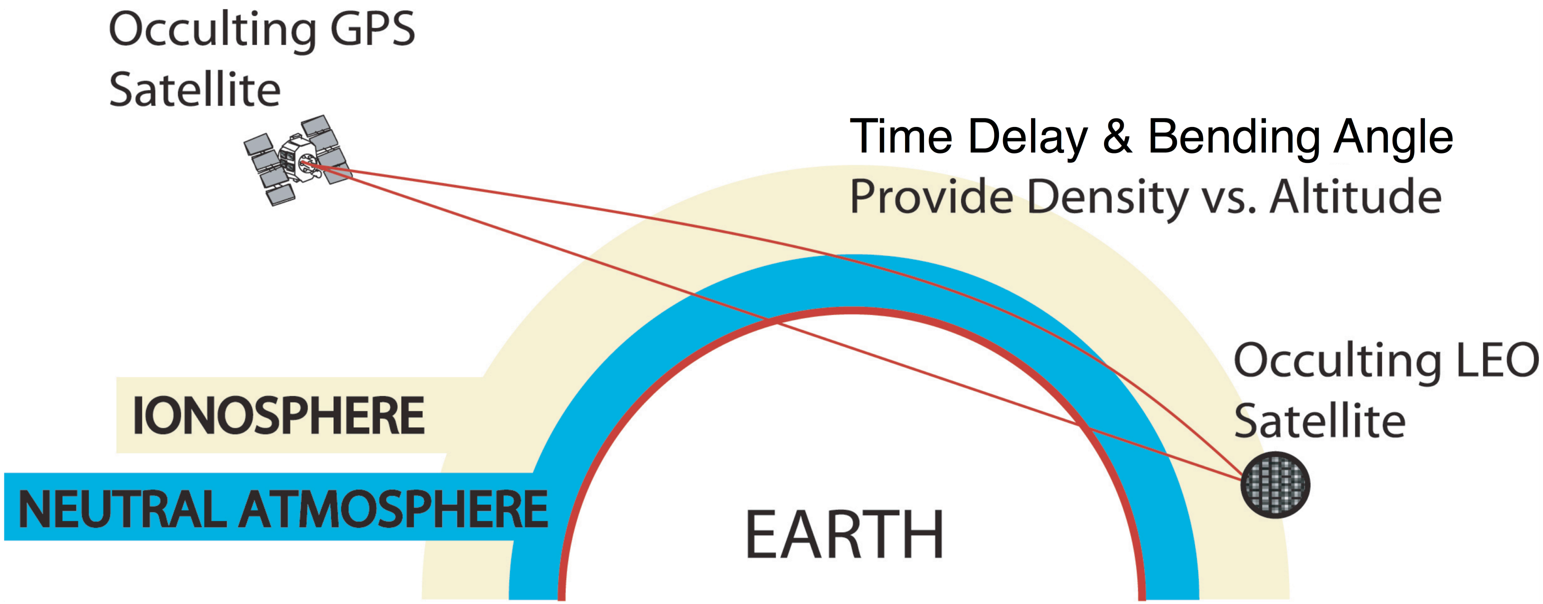
Short forecasts with a climate model from analyses, compared against observations, point to model improvements.

Observed ice fraction loss





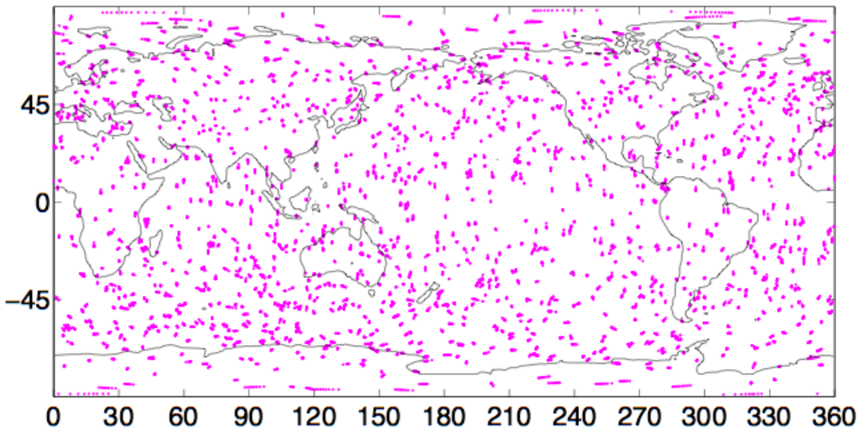
Exploring the Impact of Novel Observations: Impact of COSMIC GPS Observations in Cam Analyses



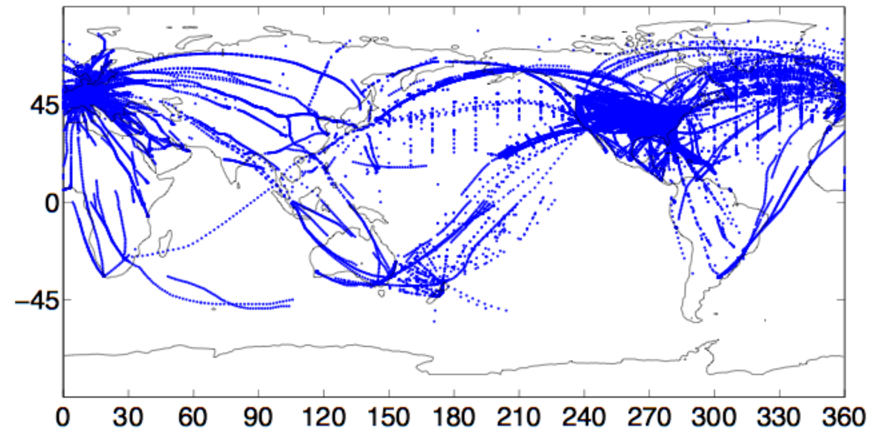
Impact of COSMIC GPS Observations in Cam Analyses

Observations 1 December 2006

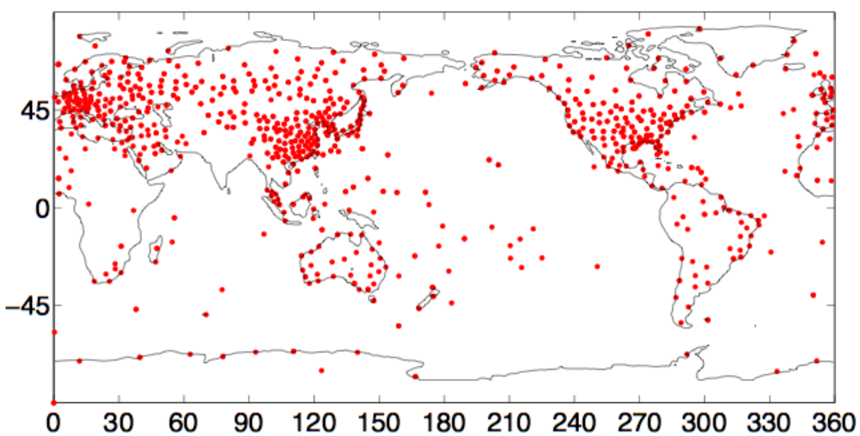
GPS



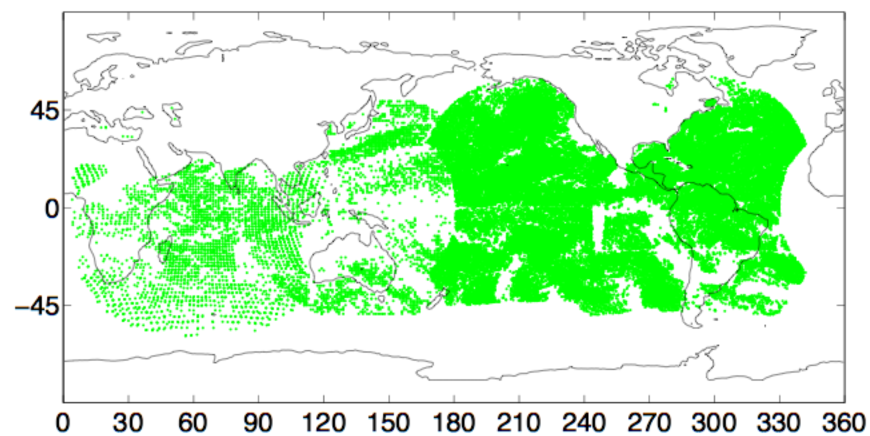
ACARS and Aircraft



Radiosondes

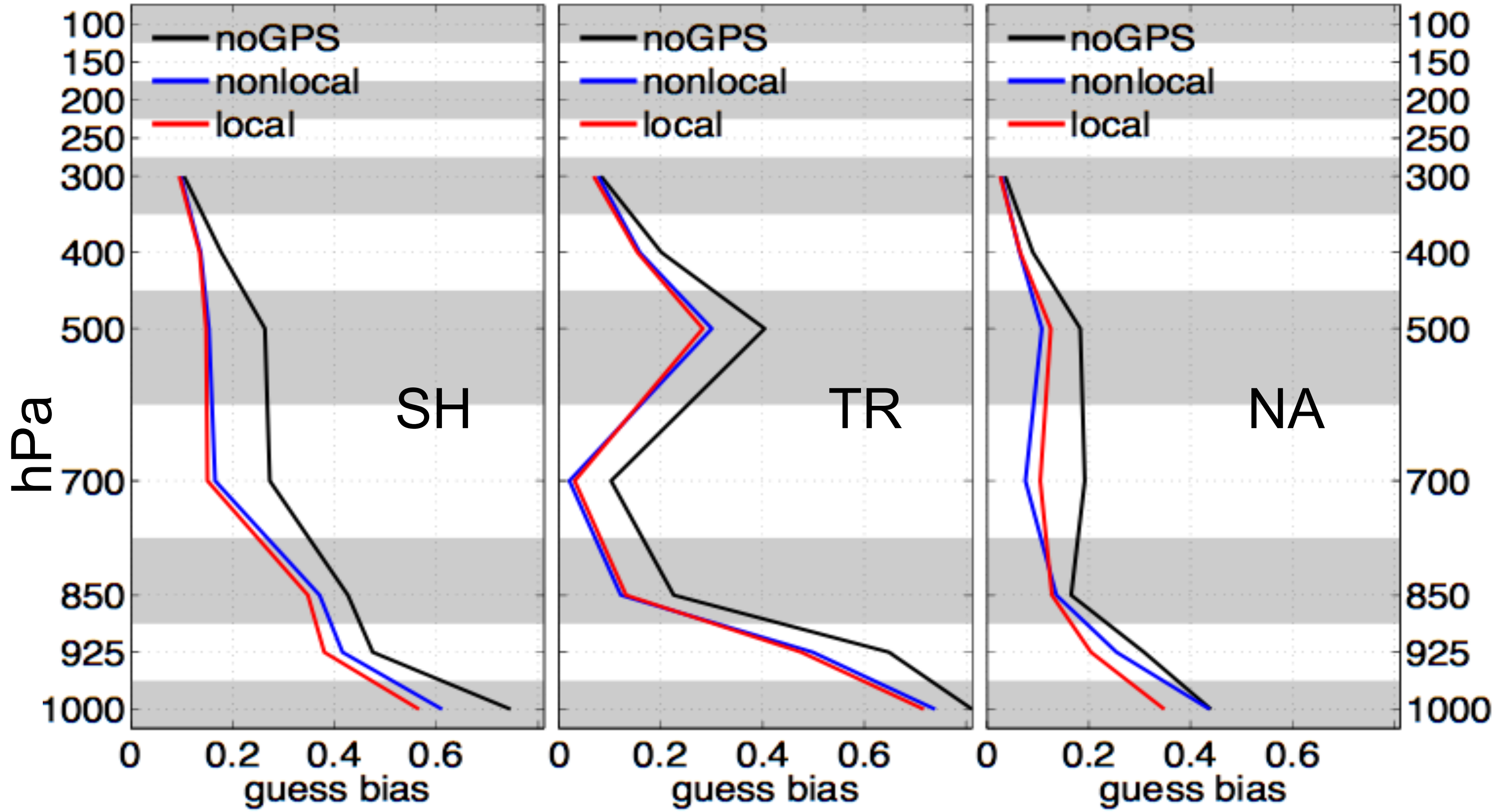


Sat Winds



CAM 6-hour forecast Bias from Radiosonde Specific Humidity (Q)

December 2006

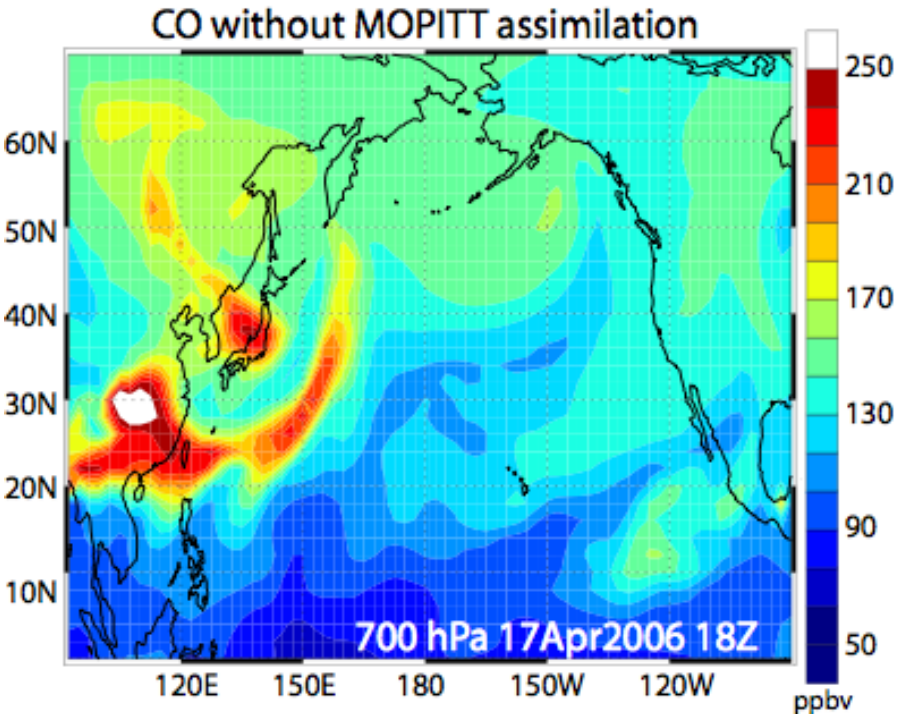


Conclusions

- GPS has significant information, especially about moisture;
- Most important where other observations are sparse;
- Ensemble assimilation can do full multivariate improvement;
- Must carefully consider planning of future obs systems.

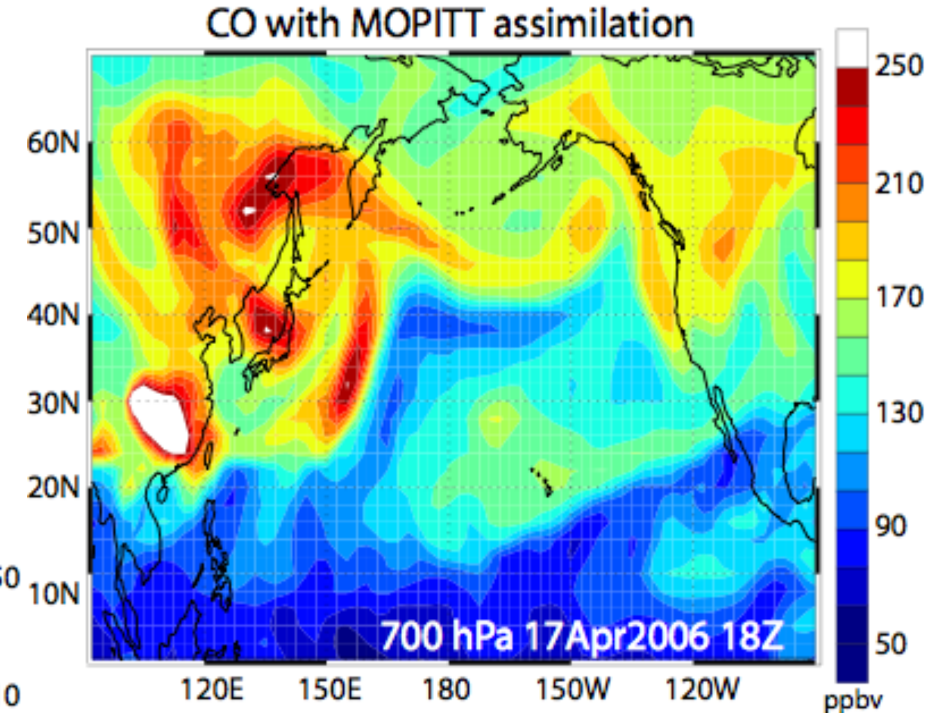
- CAM biases can be reduced with GPS observations.

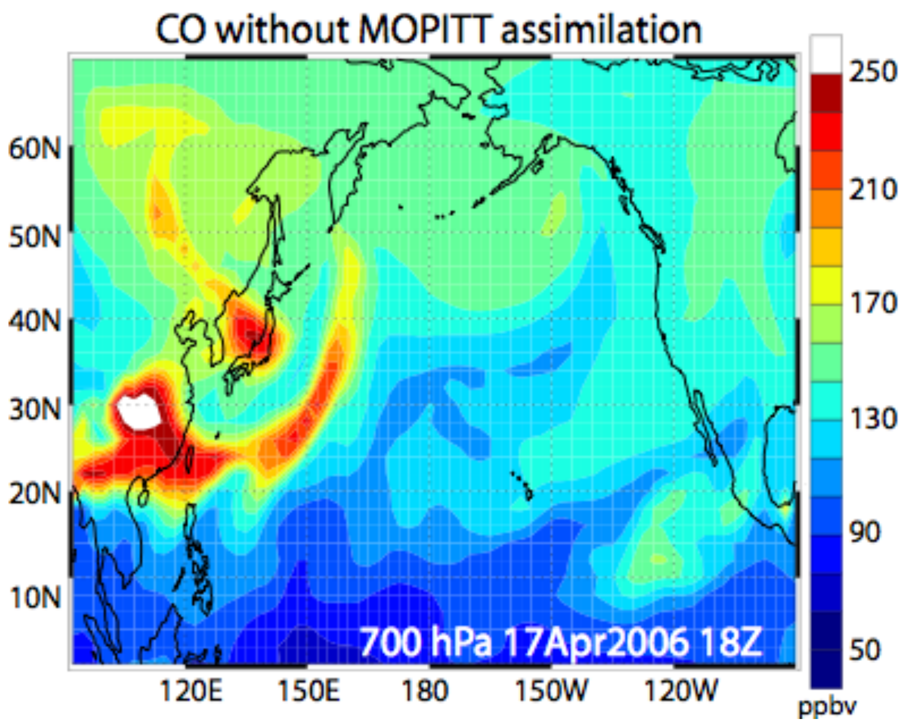
Estimating CO with MOPITT remote sensing observations in CAM/Chem
Ave Arellano, ACD
(now U. Arizona)



Ave extended DART/CAM for CAM/Chem.

Estimating CO with MOPITT remote sensing observations in CAM/Chem
Ave Arellano, ACD
(now U. Arizona)

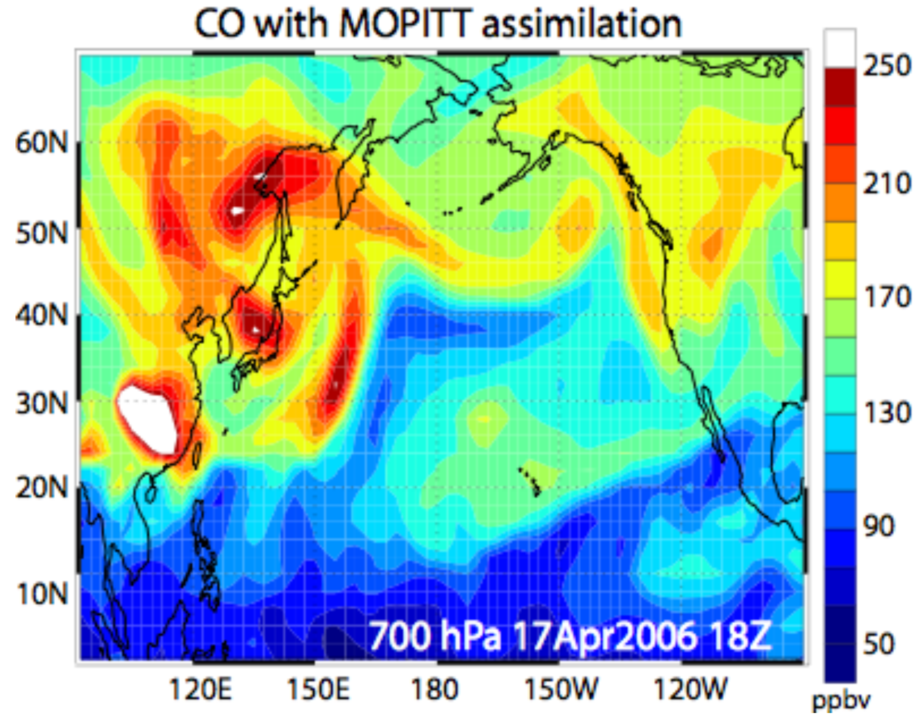




Then he added MOPITT observations. Improved fit to aircraft CO obs.

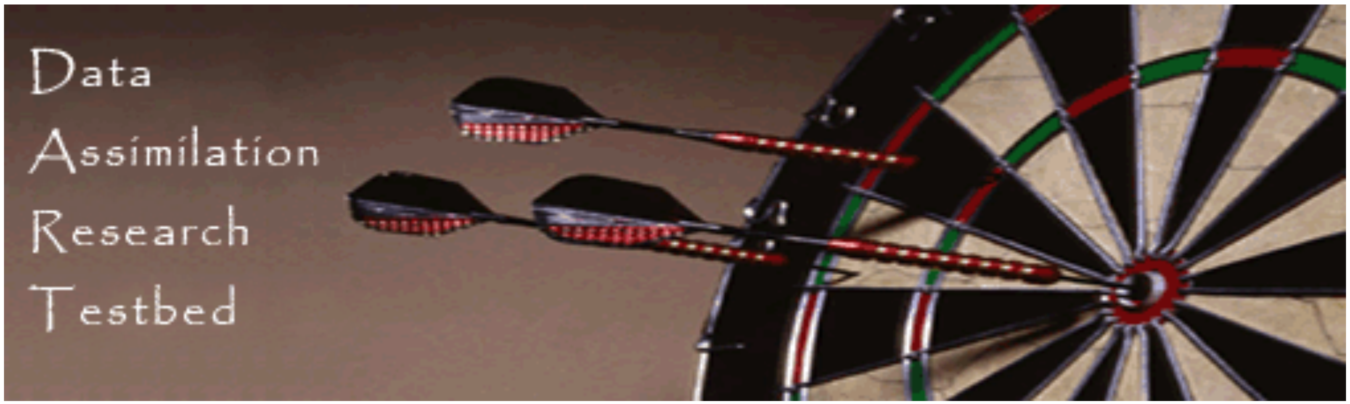
Estimating CO with MOPITT remote sensing observations in CAM/Chem

Ave Arellano, ACD
(now U. Arizona)



This system was used for real-time support for ARCTAS field campaign.

Moving towards coupled assimilation for earth system models.



Tim Hoar, Nancy Collins, Kevin Raeder, Jeffrey Anderson,
NCAR Institute for Math Applied to Geophysics
Data Assimilation Research Section

Steve Yeager, Mariana Vertenstein, Gokhan Danabasoglu, Alicia Karspeck, and Joe Tribbia
NCAR/NESL/CGD/Oceanography

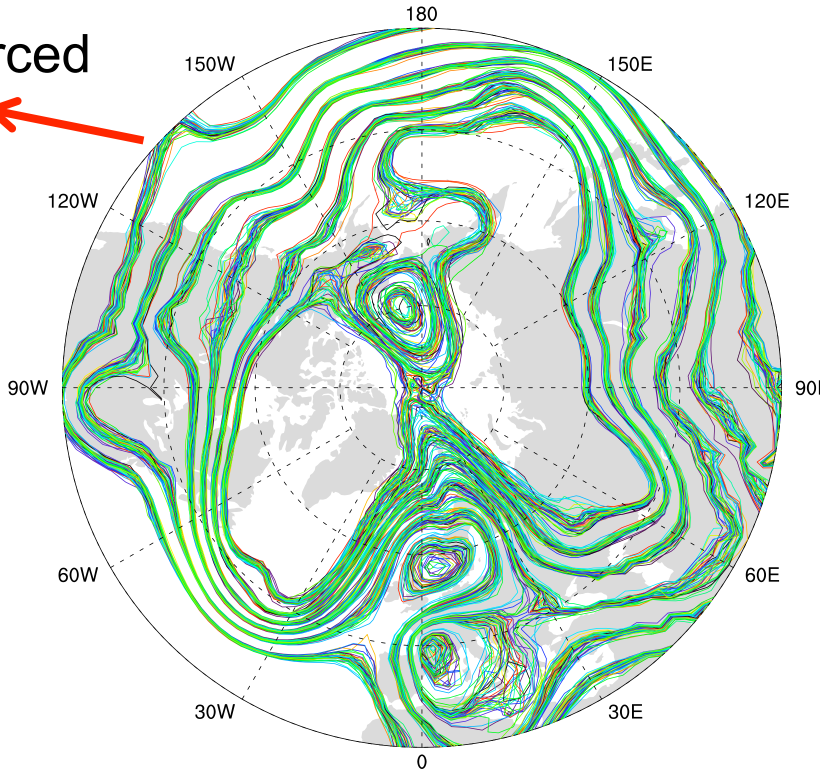
Ocean Data Assimilation Motivation

- Climate change over time scales of 1 to several decades has been identified as very important for mitigation and infrastructure planning.
- CGD needs ocean initial conditions for the IPCC decadal prediction program.

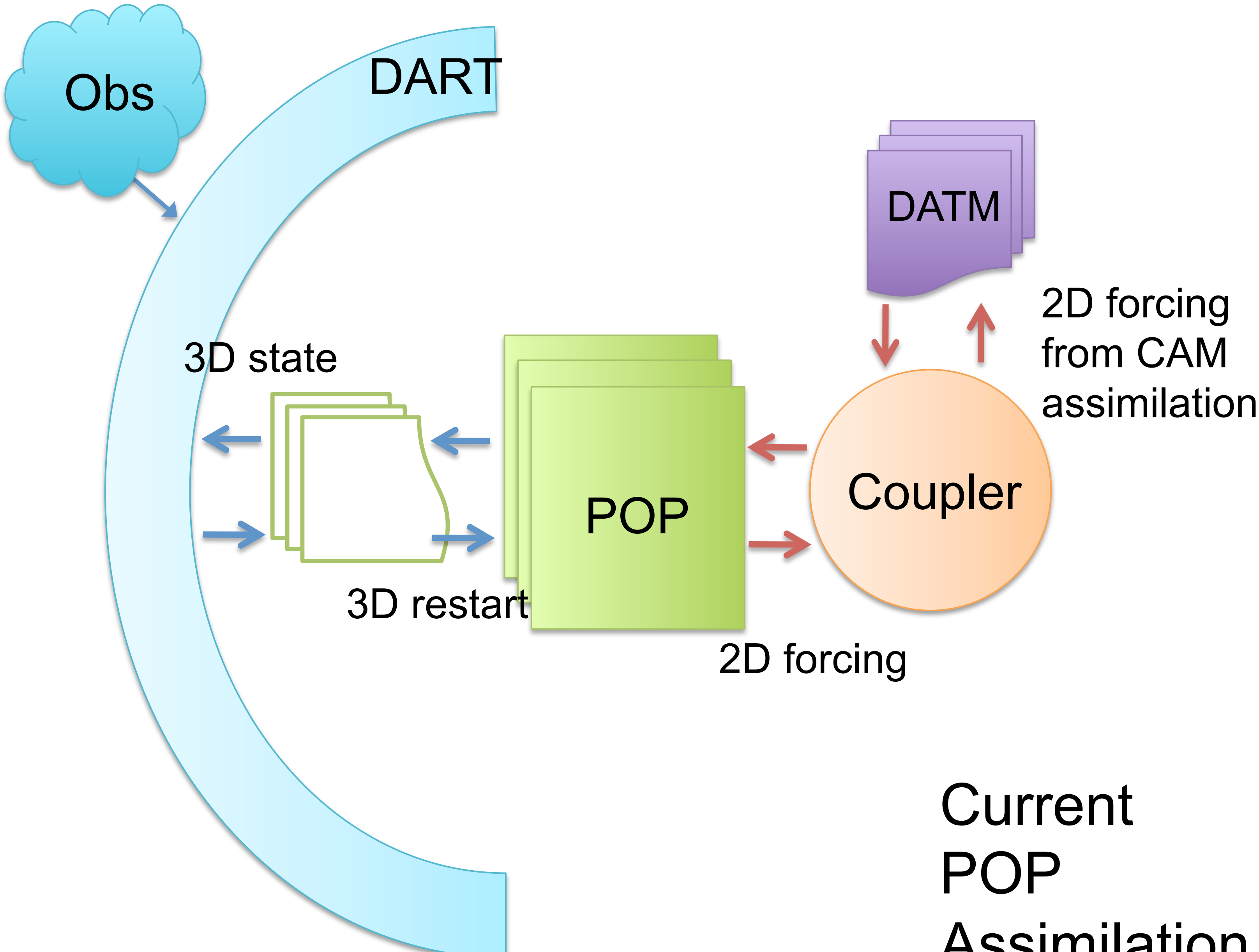
Hypothesis: Need Ensemble of Atmospheres to Force Ensemble Assimilation for Ocean

- Case 1: 23 POP members forced by a single atmosphere.
- Case 2: 48 POP members forced by 48 CAM/DART analyses.
- Generates additional ocean spread, improved analyses.

500 hPa GPH
Feb 17 2003



CONTOUR FROM 5200 TO 5700 BY 100

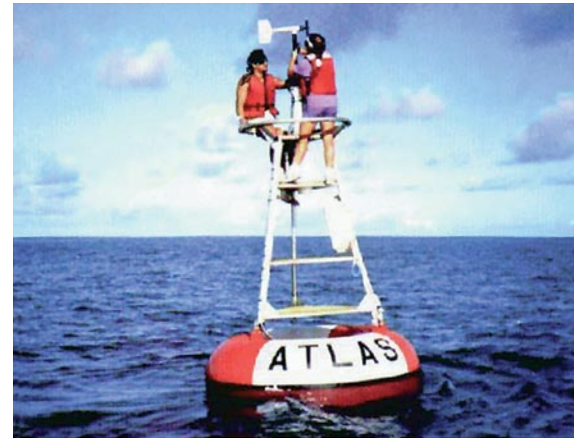


Current POP Assimilation

World Ocean Database T,S observation counts

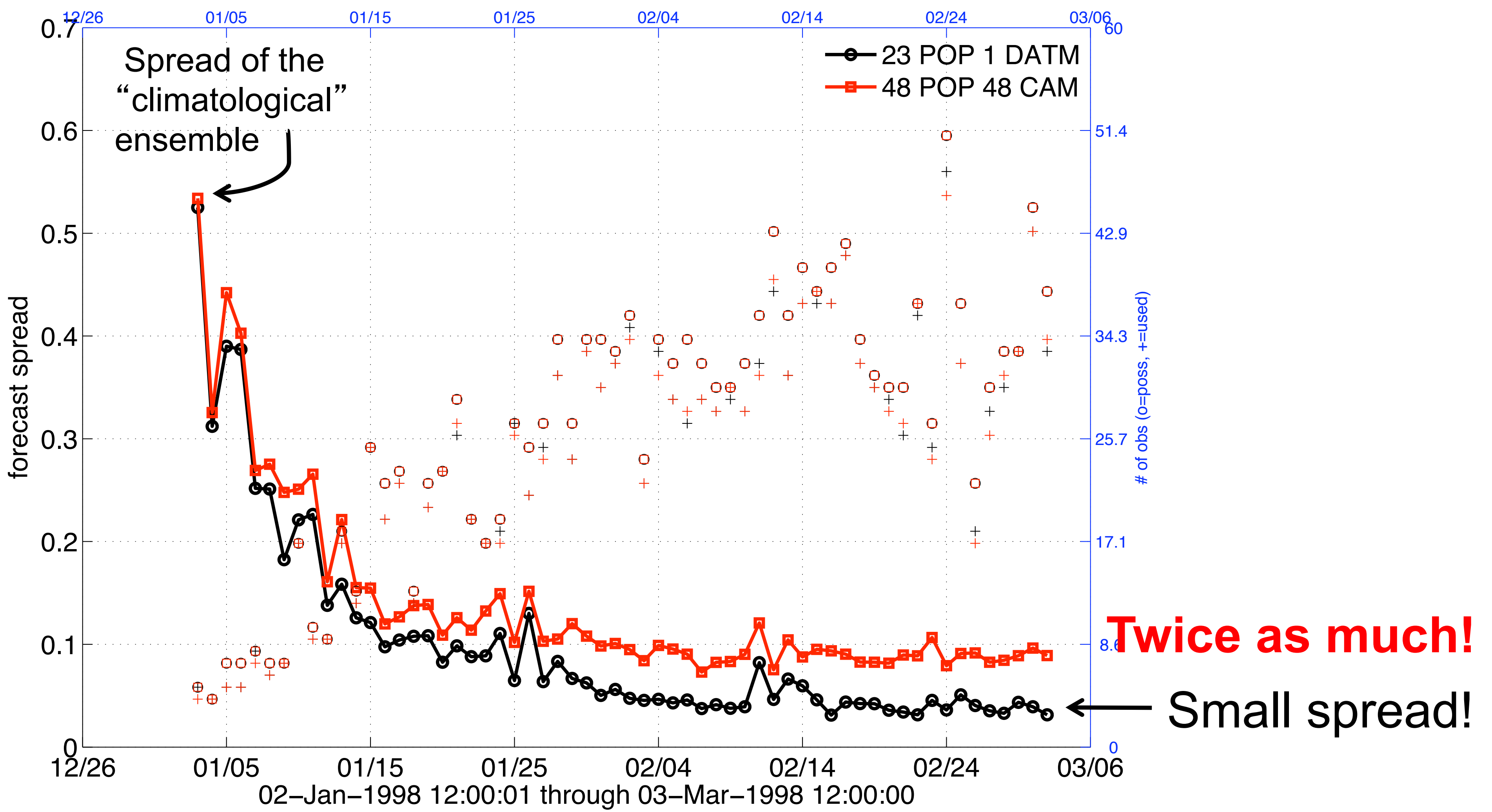
These counts are for 1998 & 1999 and are representative.

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

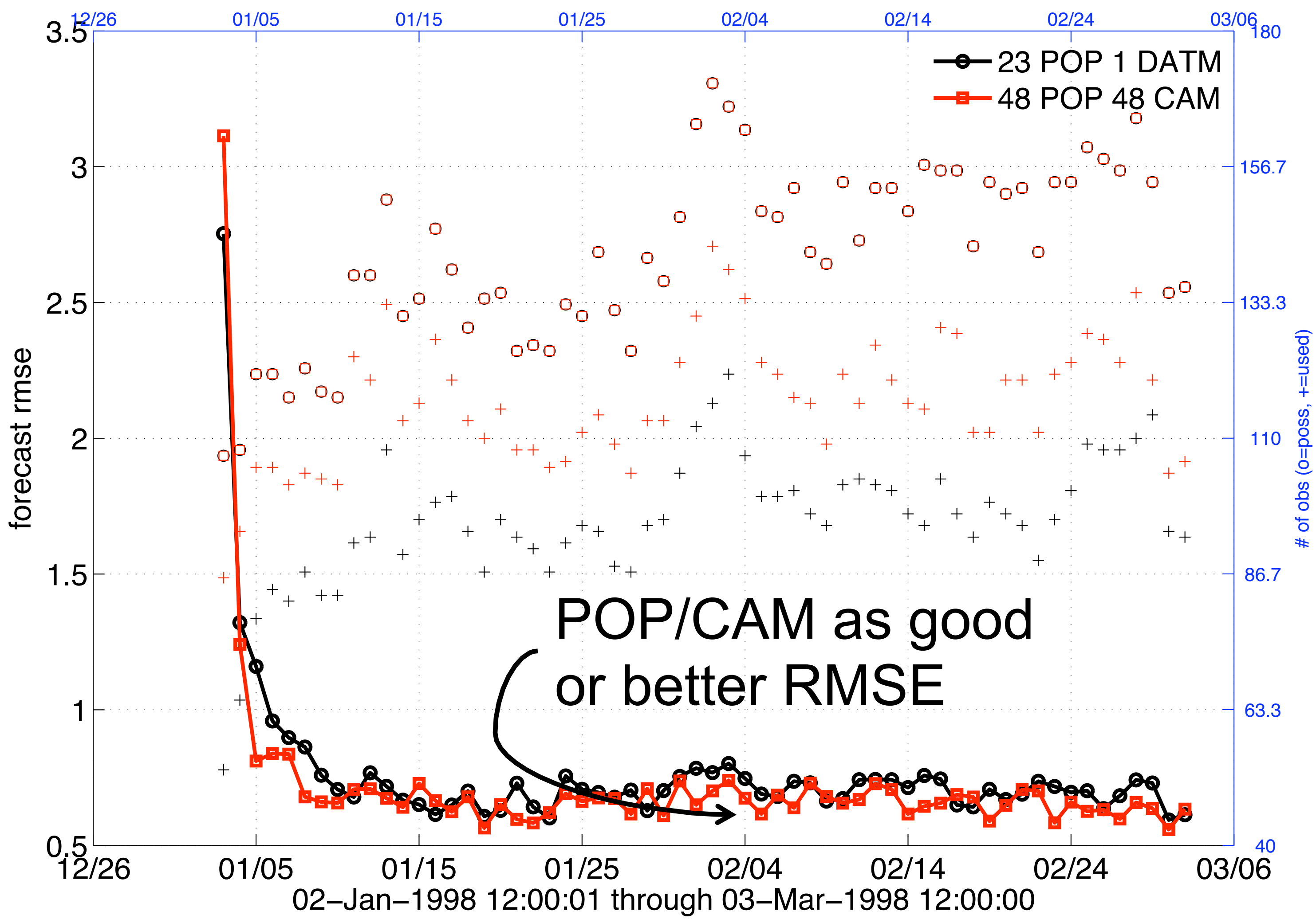


- temperature observation error standard deviation == 0.5 K.
- salinity observation error standard deviation == 0.5 msu.

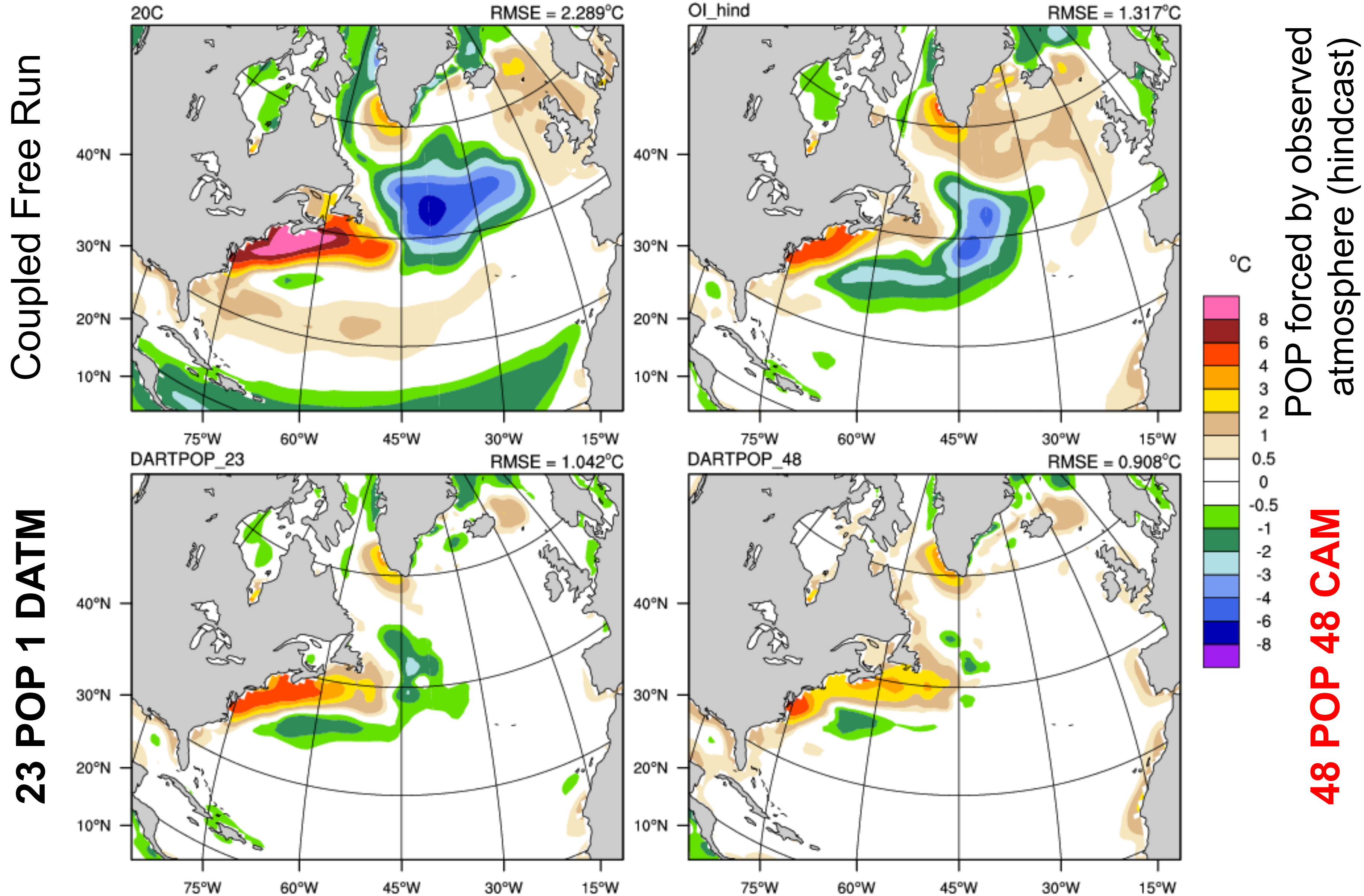
Ensemble *Spread* for Pacific 100m XBT

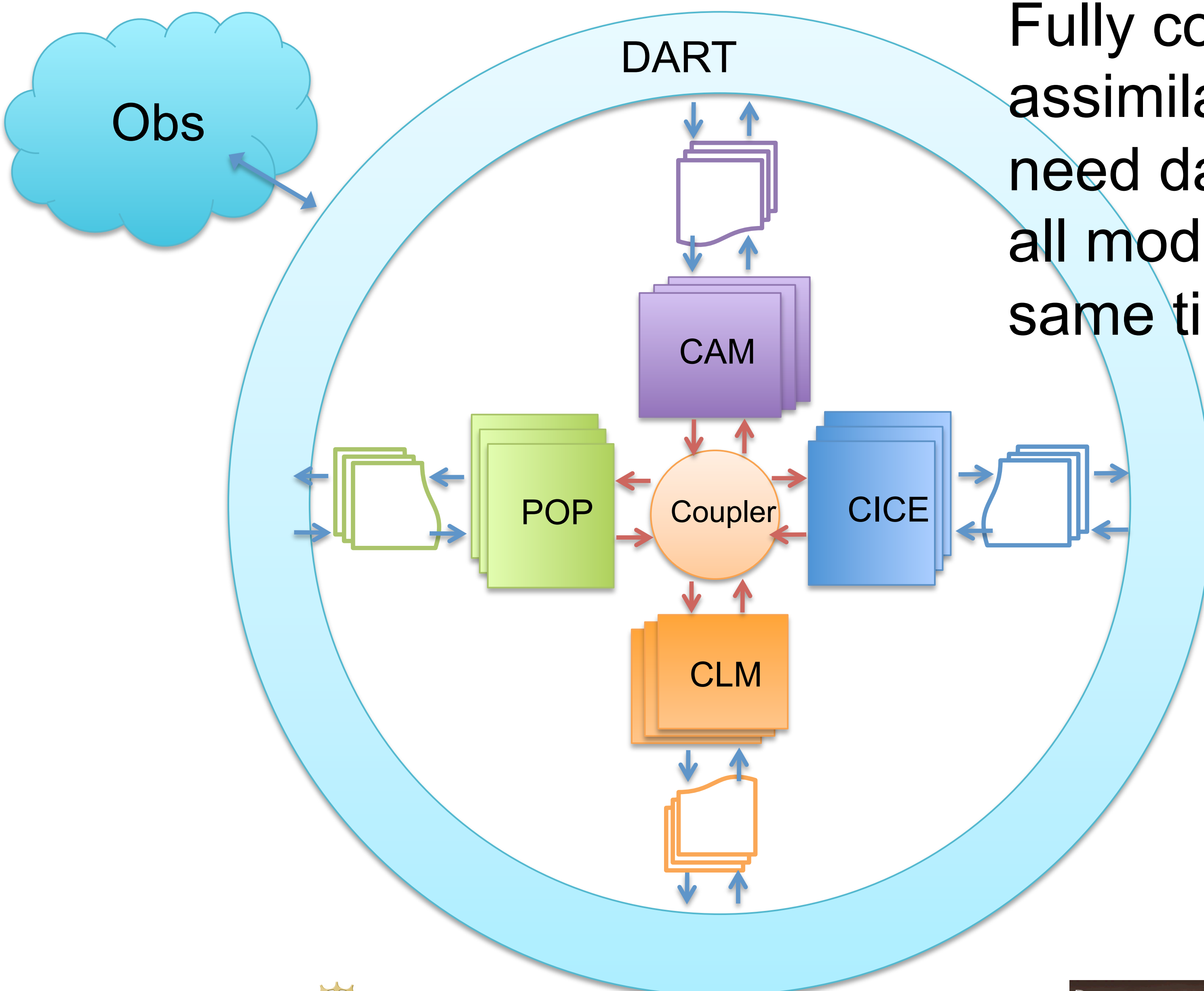


100m Mooring Temperature RMSE – Pacific



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

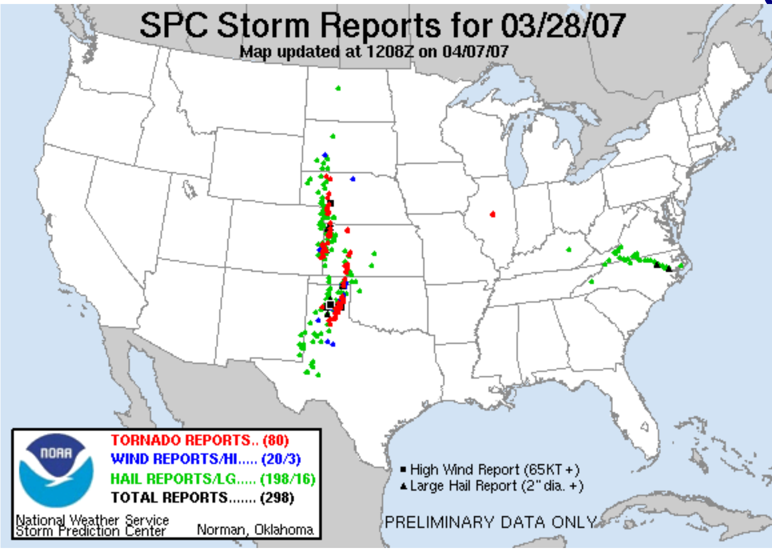


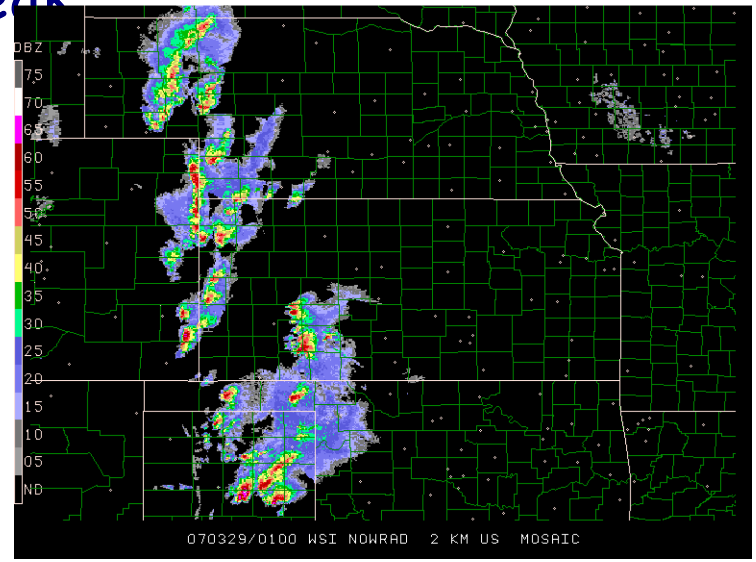


Fully coupled assimilation will need data from all models at the same time

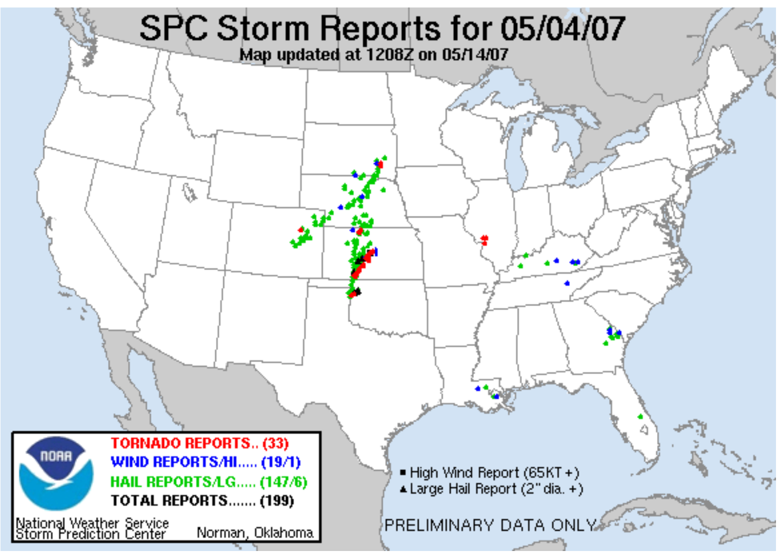
(David Dowell, NOAA)

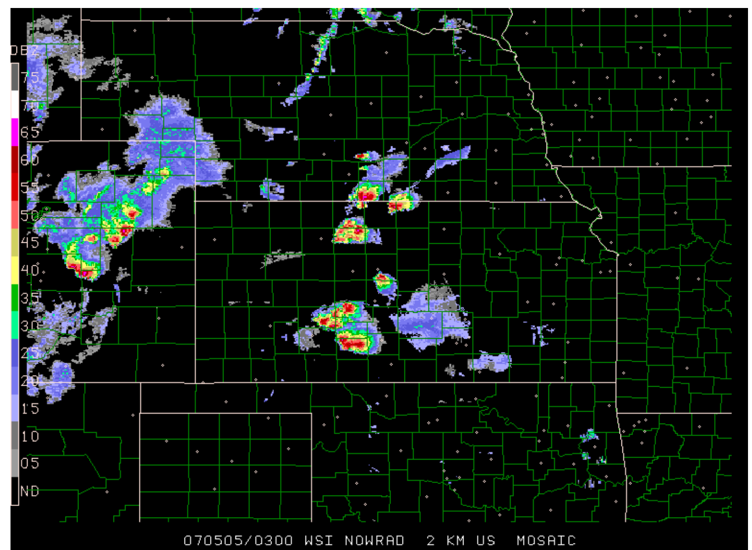
March 28 Tornado Outbreak



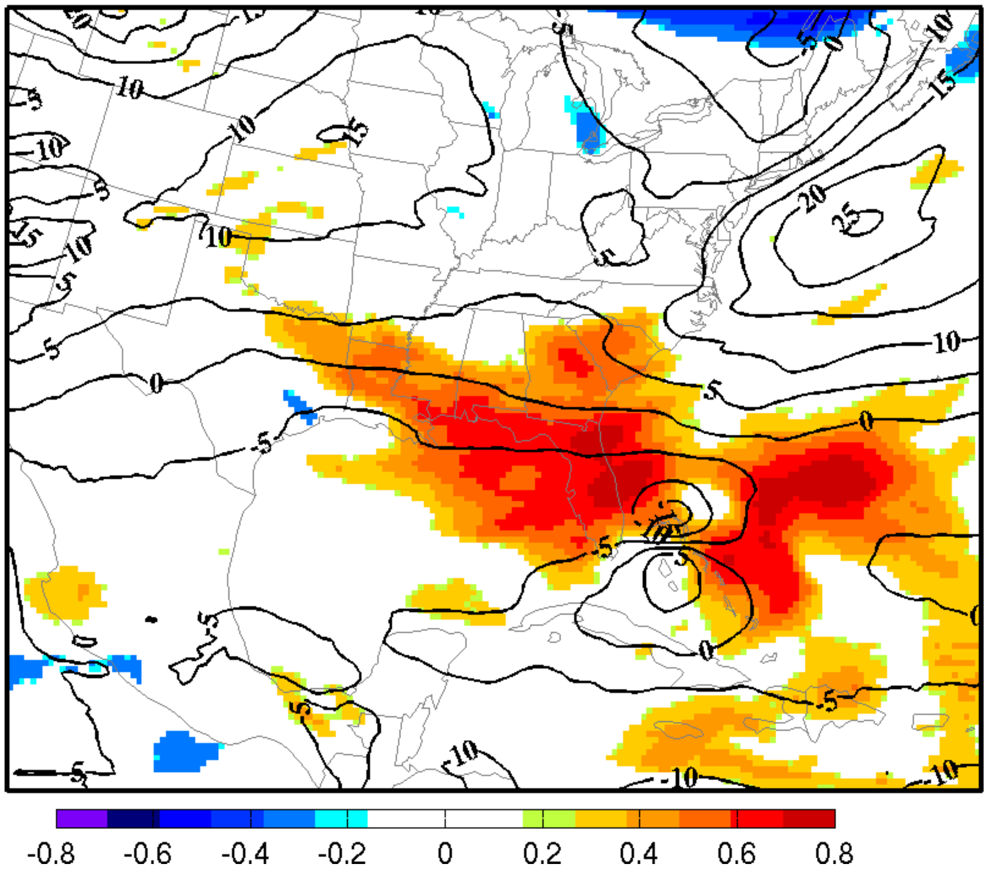


May 4 (Greensburg, KS) Tornado Case





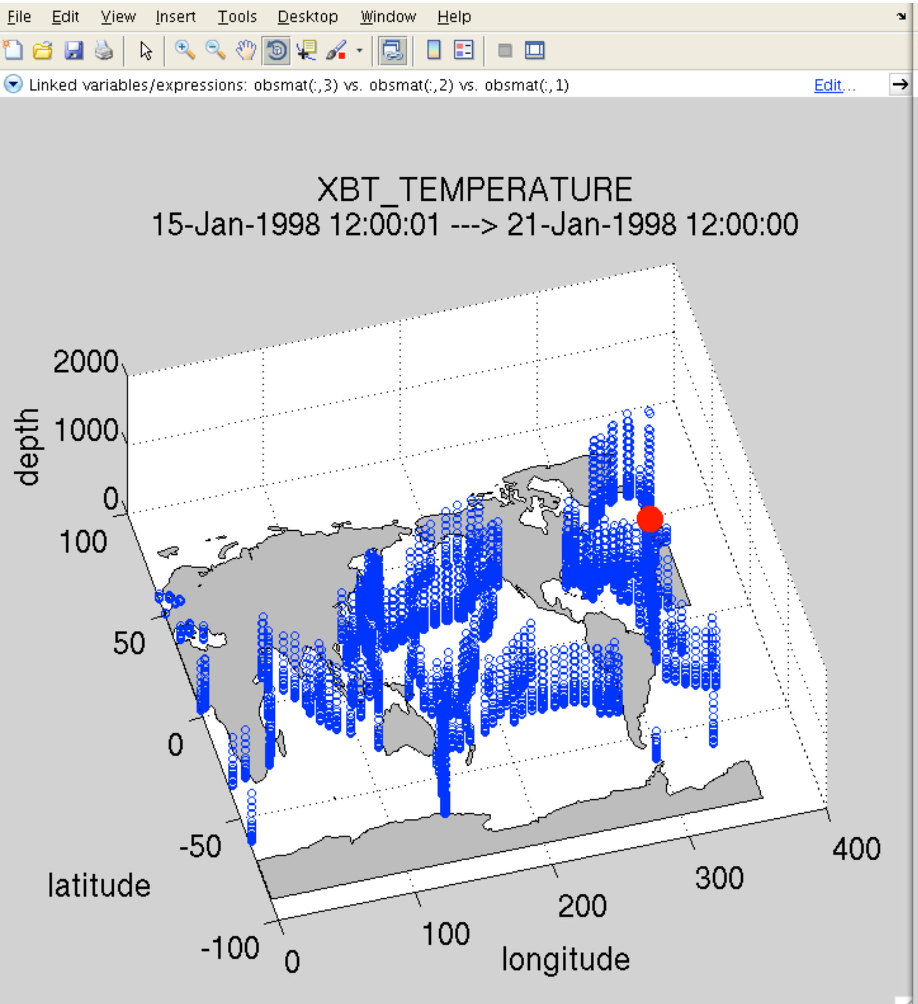
Hurricane Katrina Sensitivity Analysis (Ryan Torn, SUNY Albany)

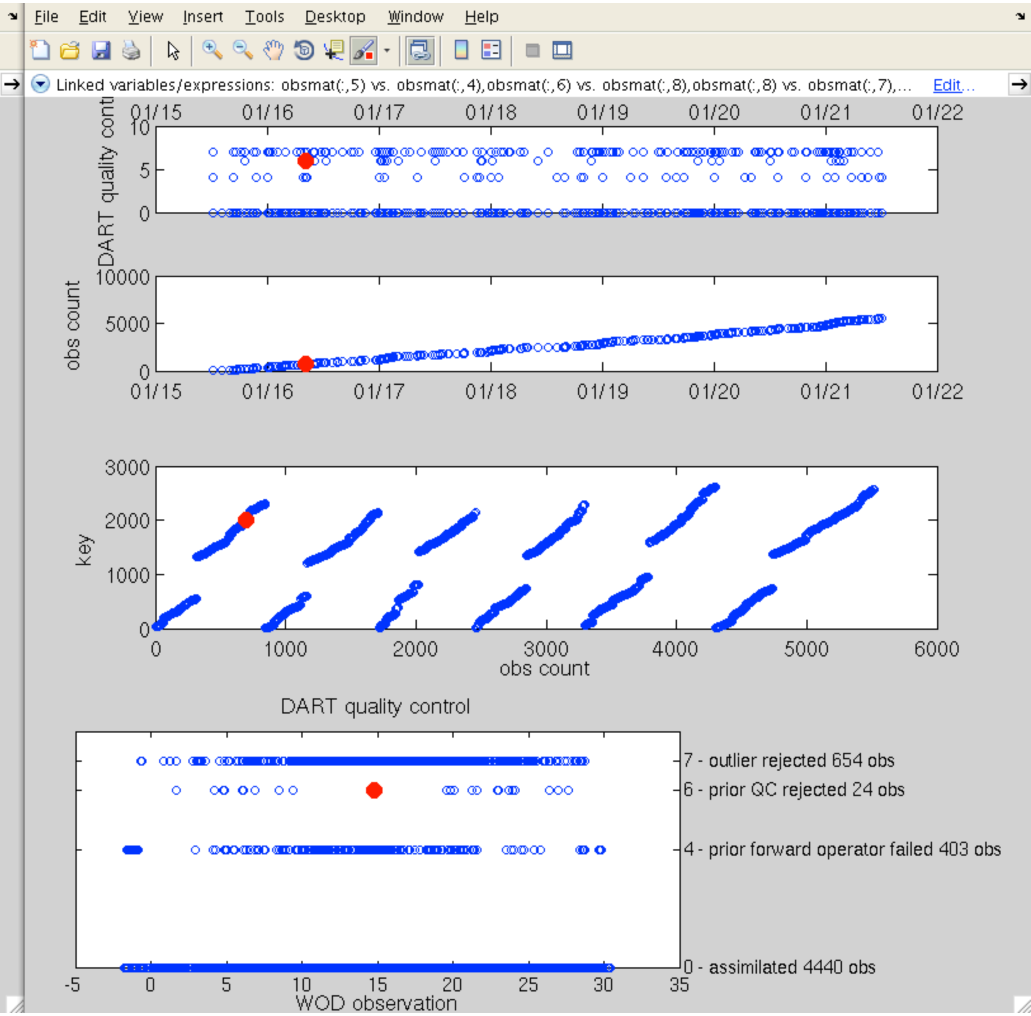


Contours are ensemble mean 48h forecast of deep-layer mean wind.

Color indicates change in the longitude of Katrina.

DART Includes Many Diagnostic Tools
Observation Visualization Example





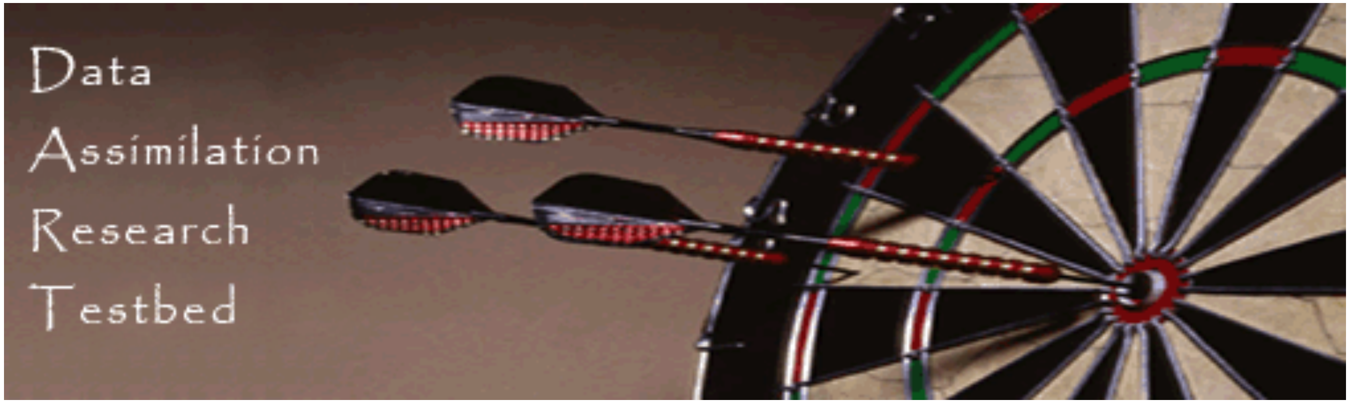
Data
Assimilation
Research
Testbed

DART Includes Many Algorithms to Improve Performance

- Adaptive inflation to maintain spread
- Adaptive localization to reduce computation
- Group filter to design localization
- Sampling error correction to reduce errors

- General parallel implementation

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



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