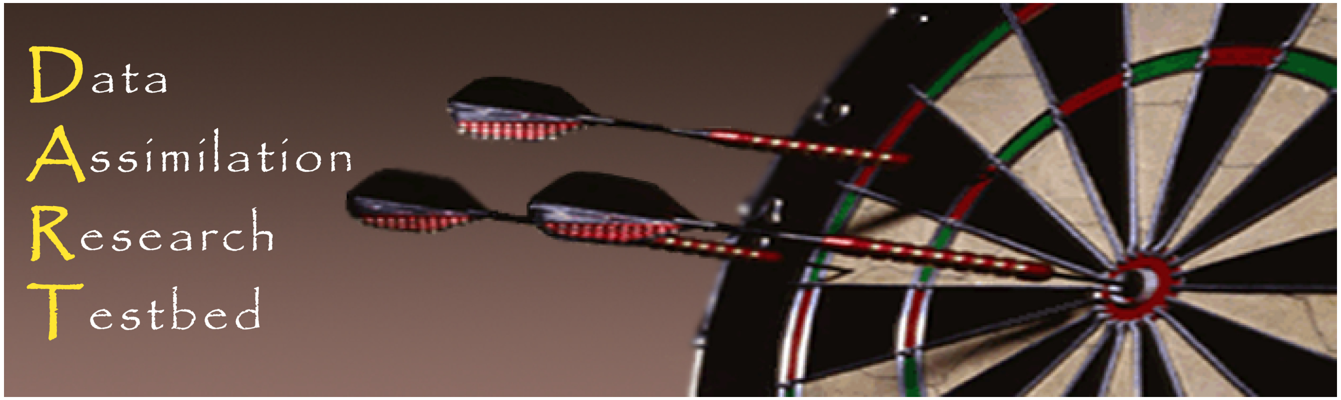
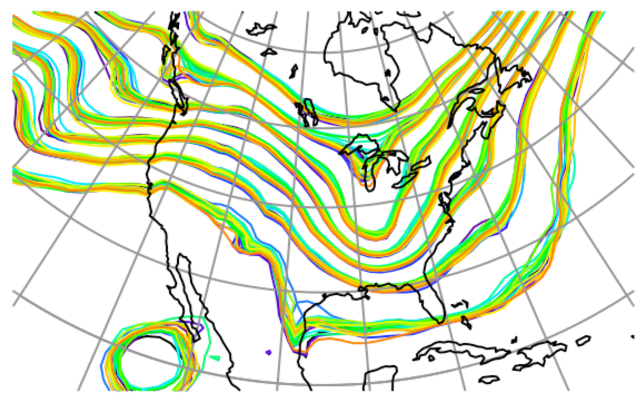


CISL's Data Assimilation Research Section:
Accelerating NCAR Science with Ensemble Data Assimilation



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.

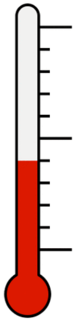
NCAR | National Center for
UCAR | Atmospheric Research

Outline

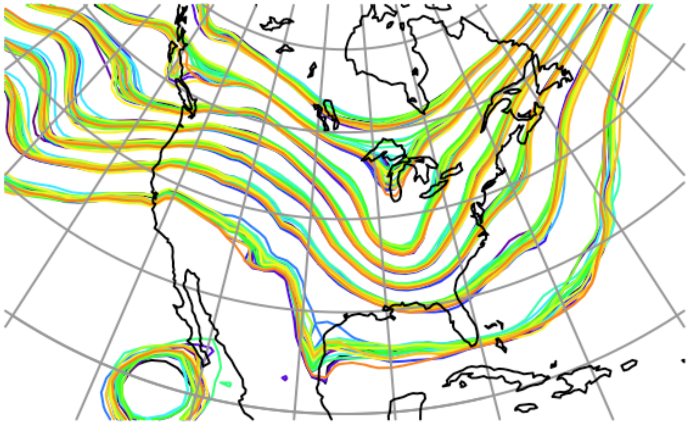
1. Intro to Ensemble Data Assimilation
2. Intro to DART
3. Example Collaborative Projects (with team introductions)
4. Data Assimilation Science in DAReS
5. Exciting new projects

What is Data Assimilation?

Observations combined with a Model forecast ...



+



...to produce an analysis
(best possible estimate).

Uncertainty and Ensemble Data Assimilation

Uncertainty is a key aspect of Earth System Data Assimilation (DA).

All observations have random errors (my thermometer is not exact).

Usually not as many observations as one would like (it's a big world).

Errors grow as forecasts get longer (models are 'chaotic').

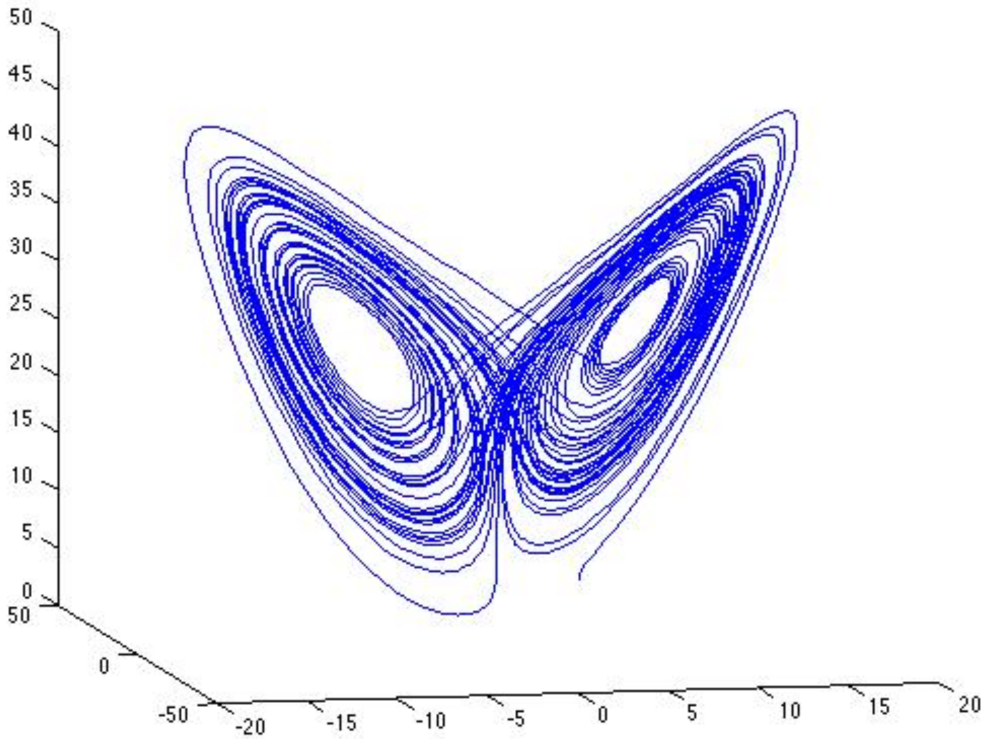
Use an ensemble (a set) of forecasts.

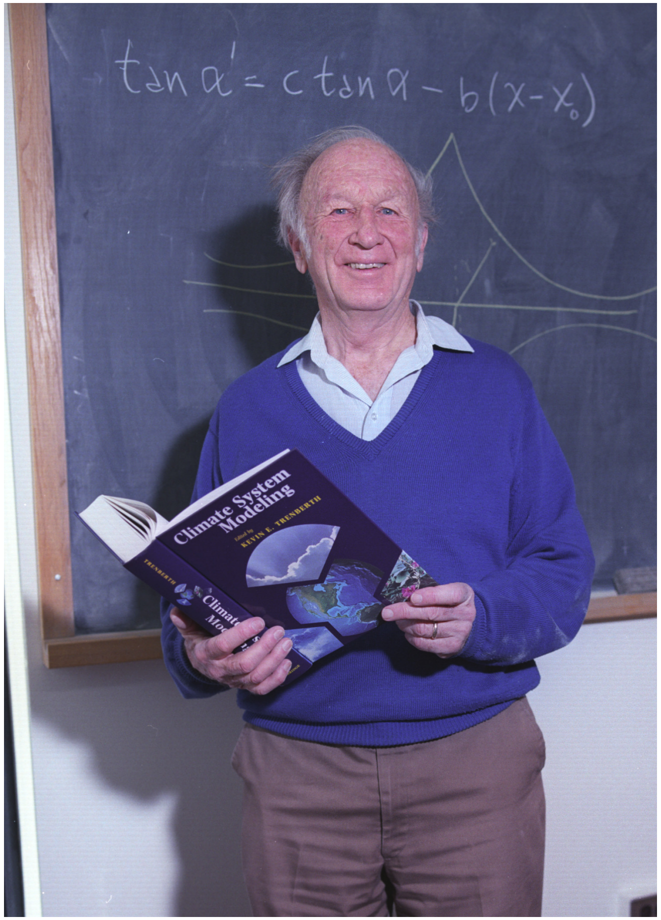
These can give an idea of the uncertainty.

Ensemble DA in the Lorenz Model

In 1963, Ed Lorenz made a very simple model of convection.
It only has 3 variables.
Surprise! Very small differences at the start become HUGE for long forecasts.

Model also describes how a ball moves in space.





Ensemble DA in the Lorenz Model

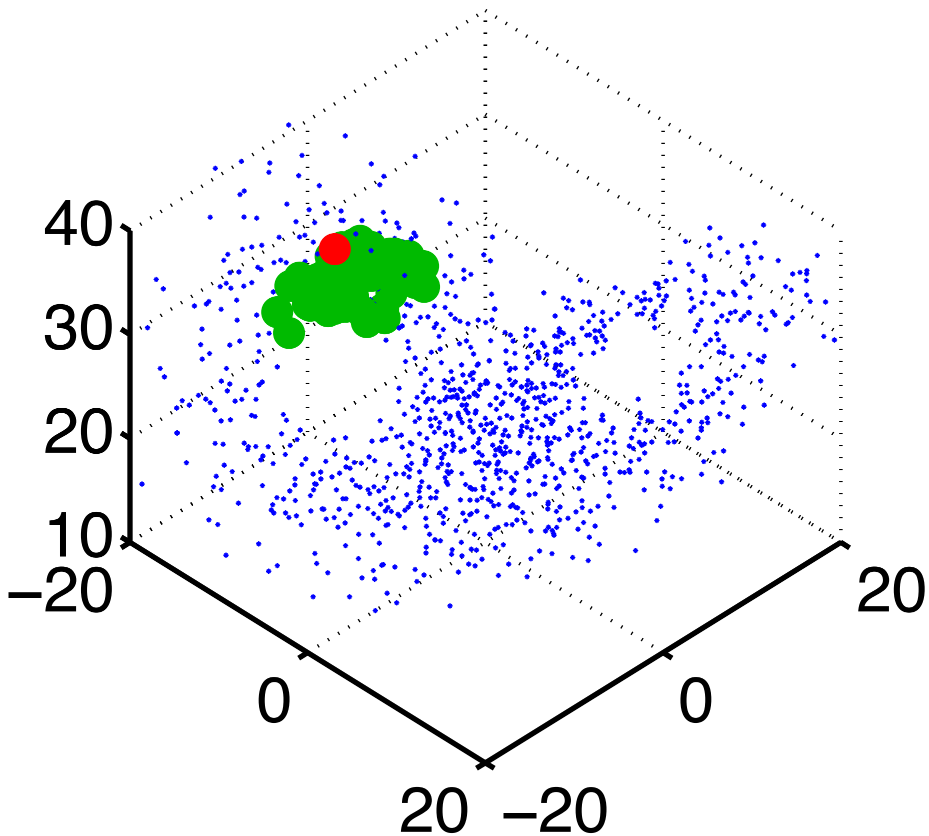
We’ll show an example with 20 ensemble members.

One solution of model is defined to be the ‘truth’.

‘Observations’ are created by adding random error to truth.

Observations only every 6 hours.

Simple Example: Lorenz 63
3-variable chaotic model



Observation in red.

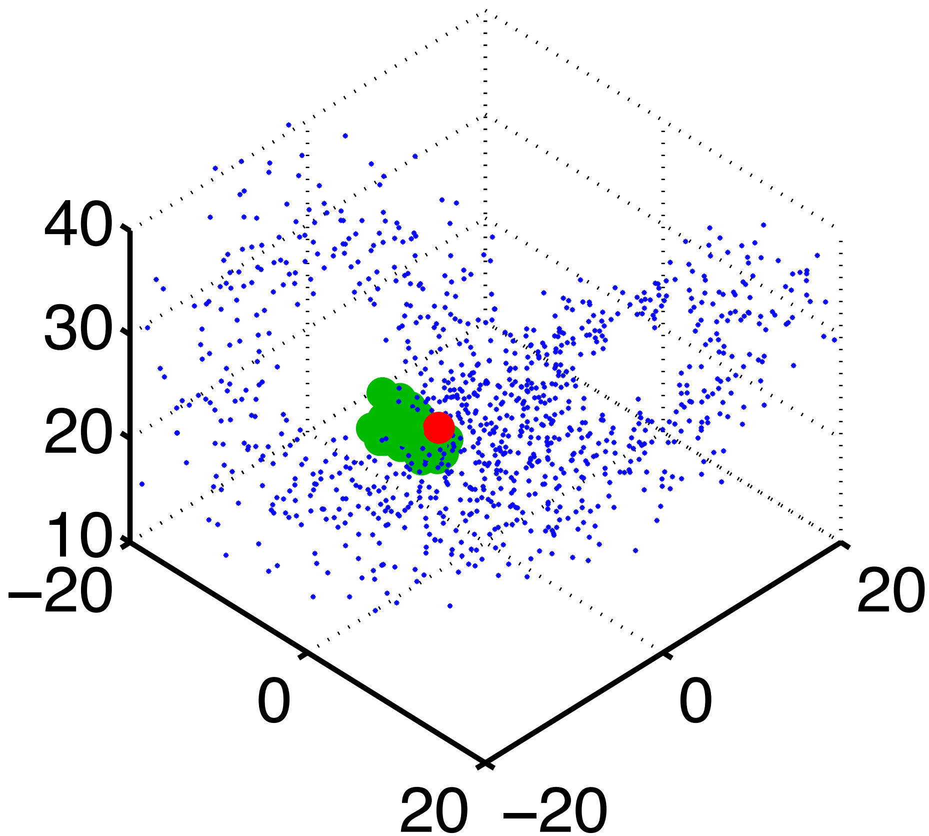
Prior ensemble in green.

Observing all three state variables.

Obs. Error variance = 4.0.

Four 20-member ensembles.

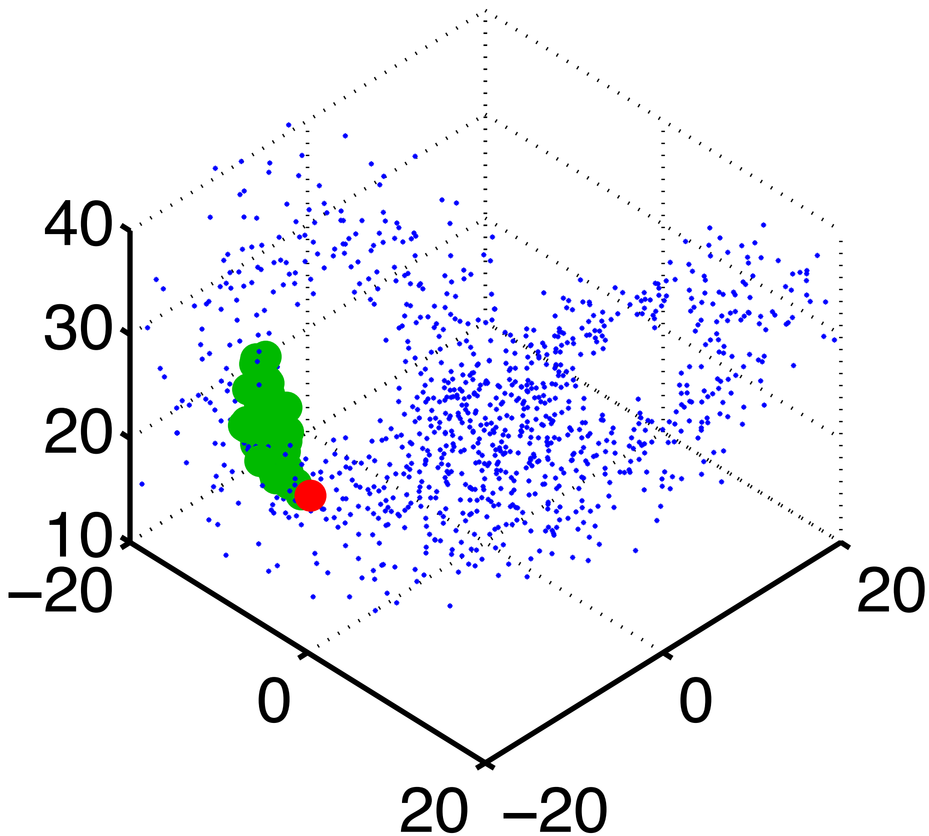
Simple Example: Lorenz 63
3-variable chaotic model



Observation in red.

Prior ensemble in green.

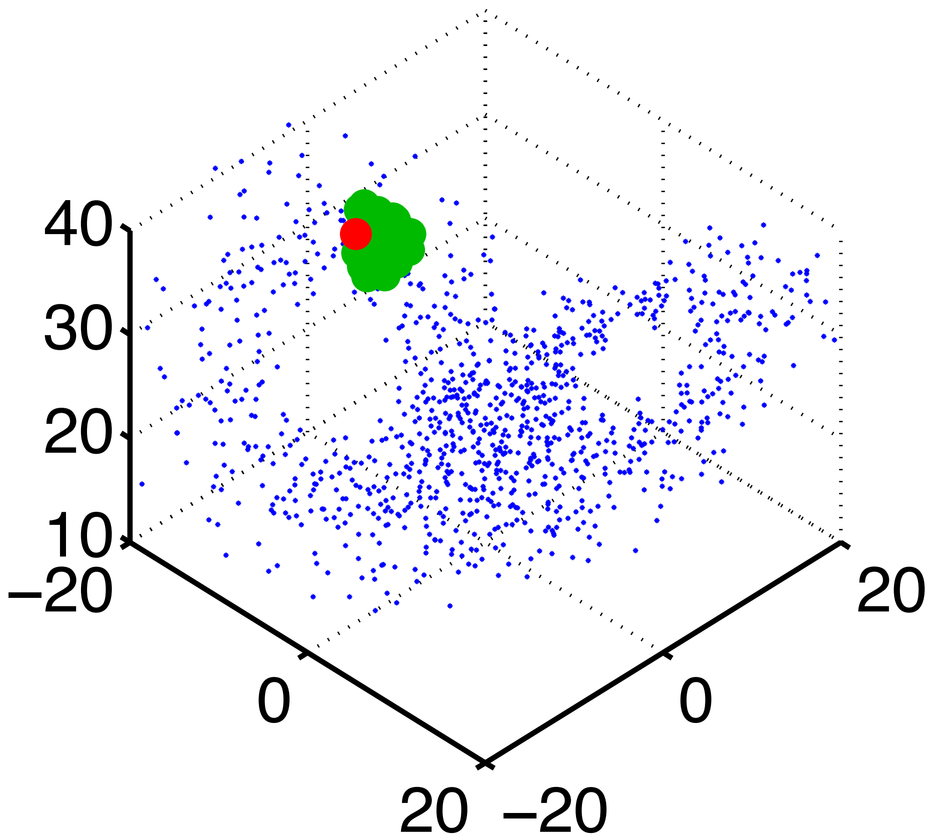
Simple Example: Lorenz 63
3-variable chaotic model



Observation in red.

Prior ensemble in green.

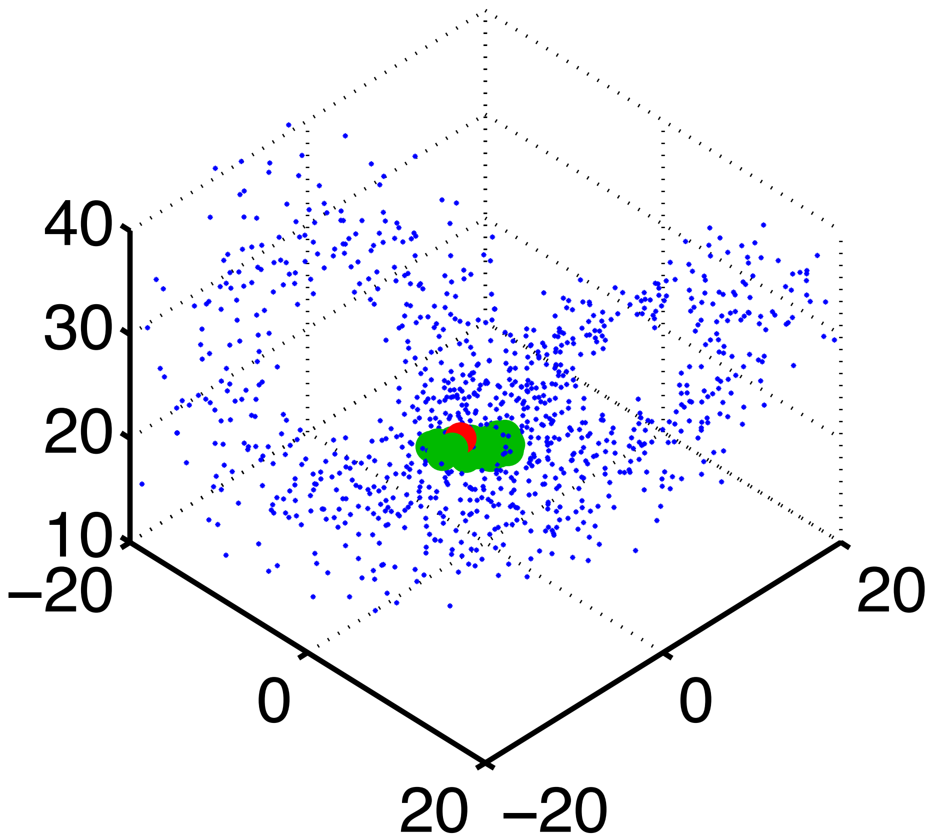
Simple Example: Lorenz 63
3-variable chaotic model



Observation in red.

Prior ensemble in green.

Simple Example: Lorenz 63
3-variable chaotic model

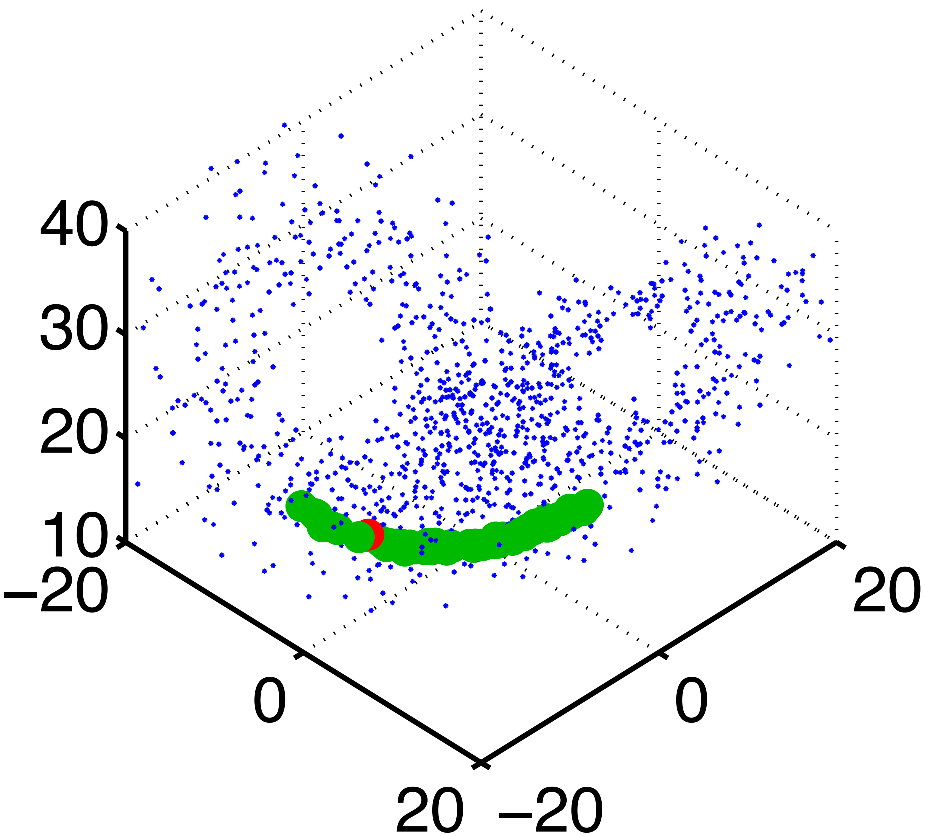


Observation in red.

Prior ensemble in green.

Ensemble is passing through an unpredictable region.

Simple Example: Lorenz 63
3-variable chaotic model

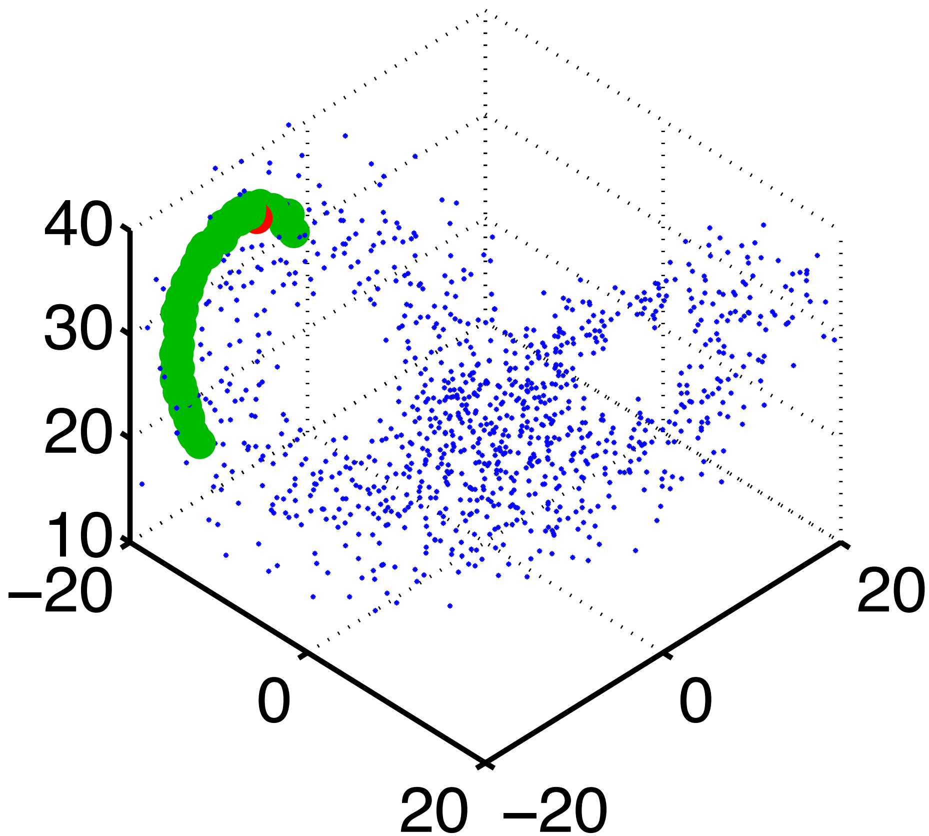


Observation in red.

Prior ensemble in green.

Part of the ensemble heads for one lobe, the rest for the other..

Simple Example: Lorenz 63
3-variable model



Observation in red.

Prior ensemble in green.

Ensemble Kalman Filter: The Details

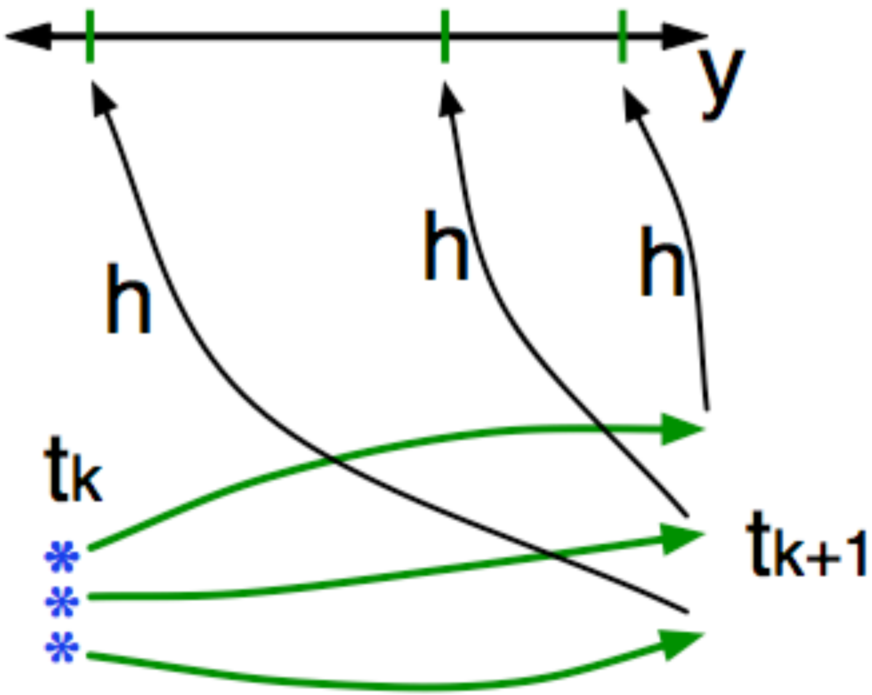
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Ensemble Kalman Filter: The Details

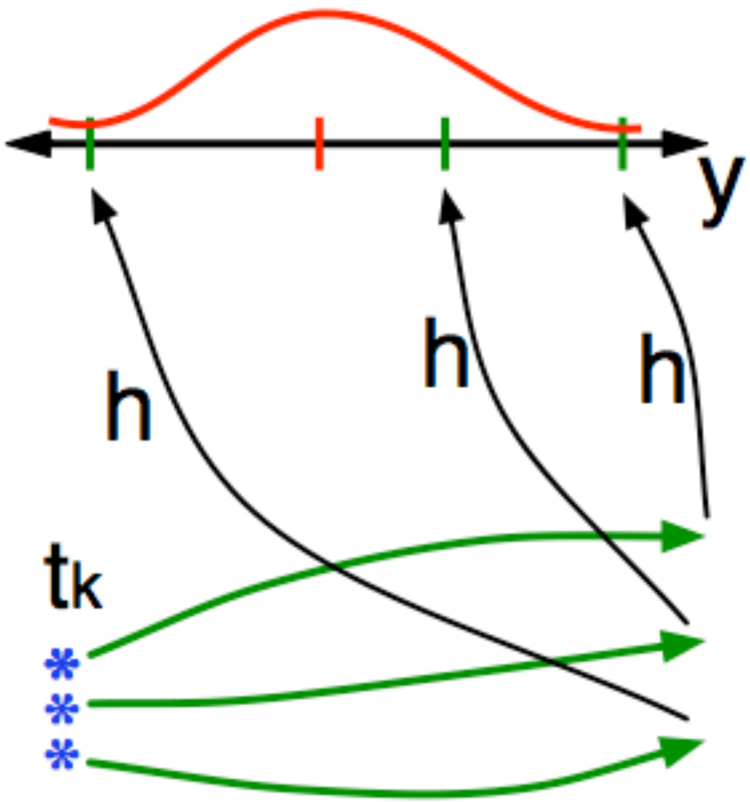
2. Get prior ensemble sample of observation, $y = h(x)$, by applying forward operator **h** to each ensemble member.



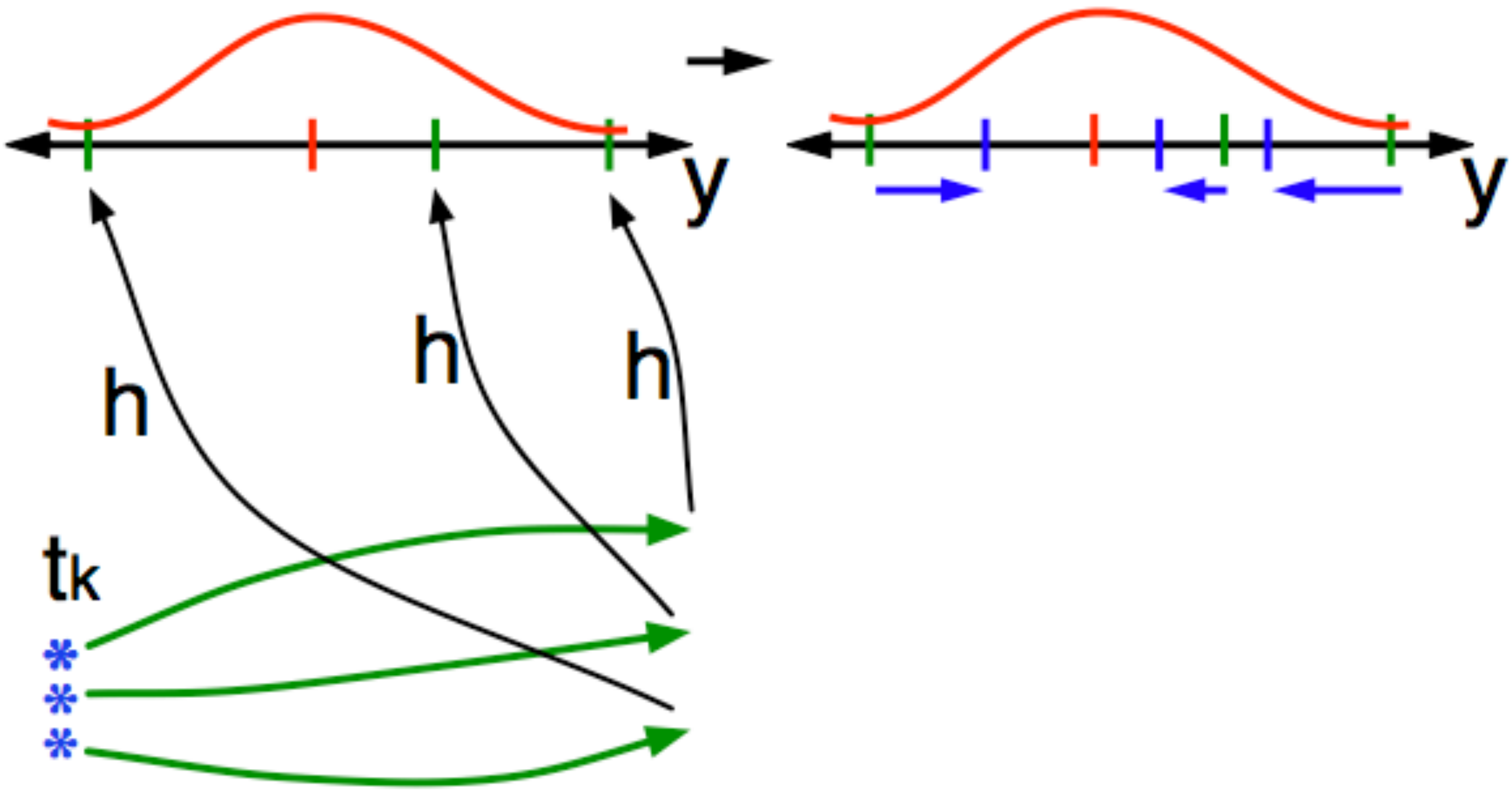
Theory: observations from instruments with uncorrelated errors can be done sequentially.

Ensemble Kalman Filter: The Details

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

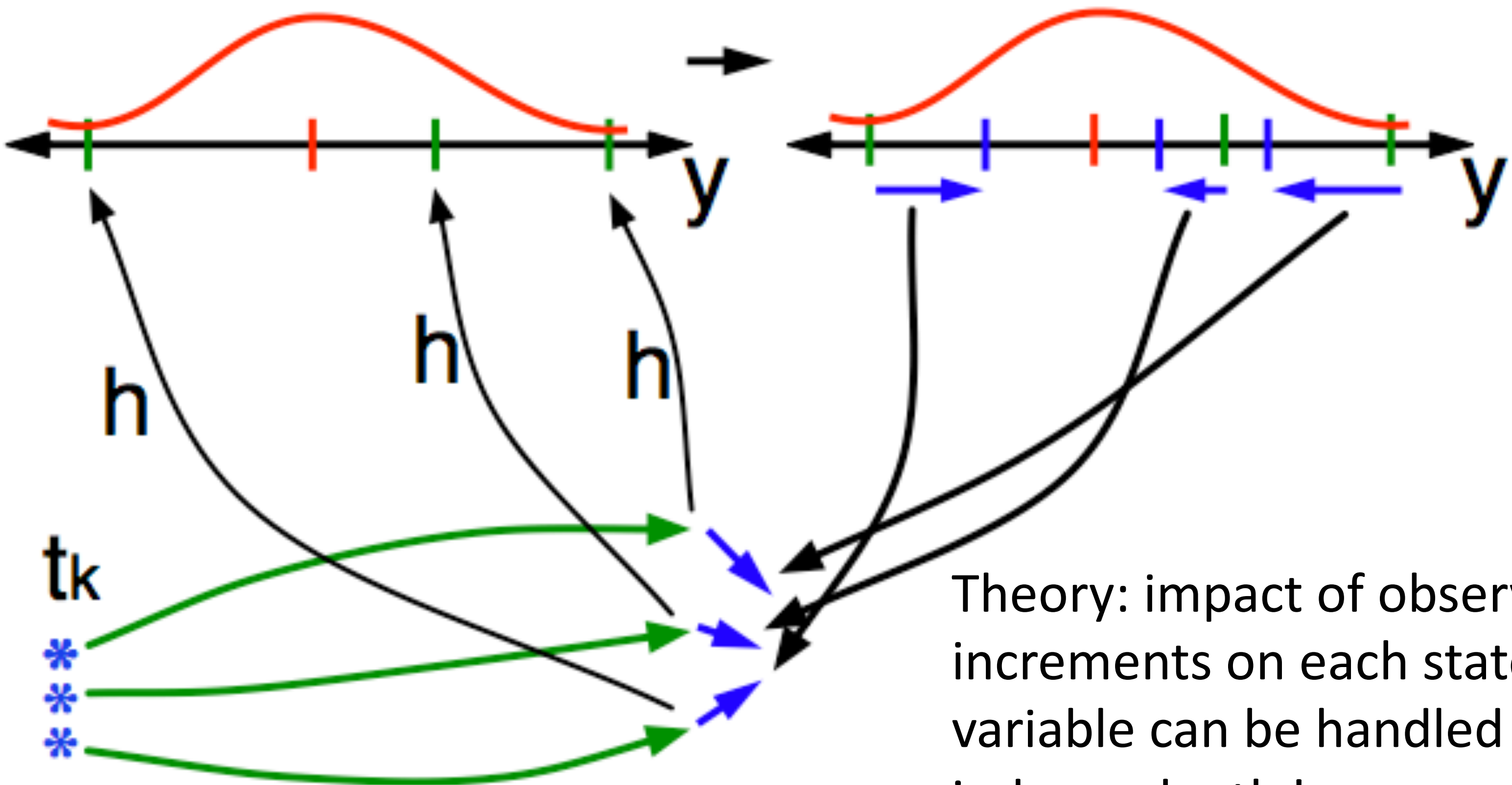


4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



Ensemble Kalman Filter: The Details

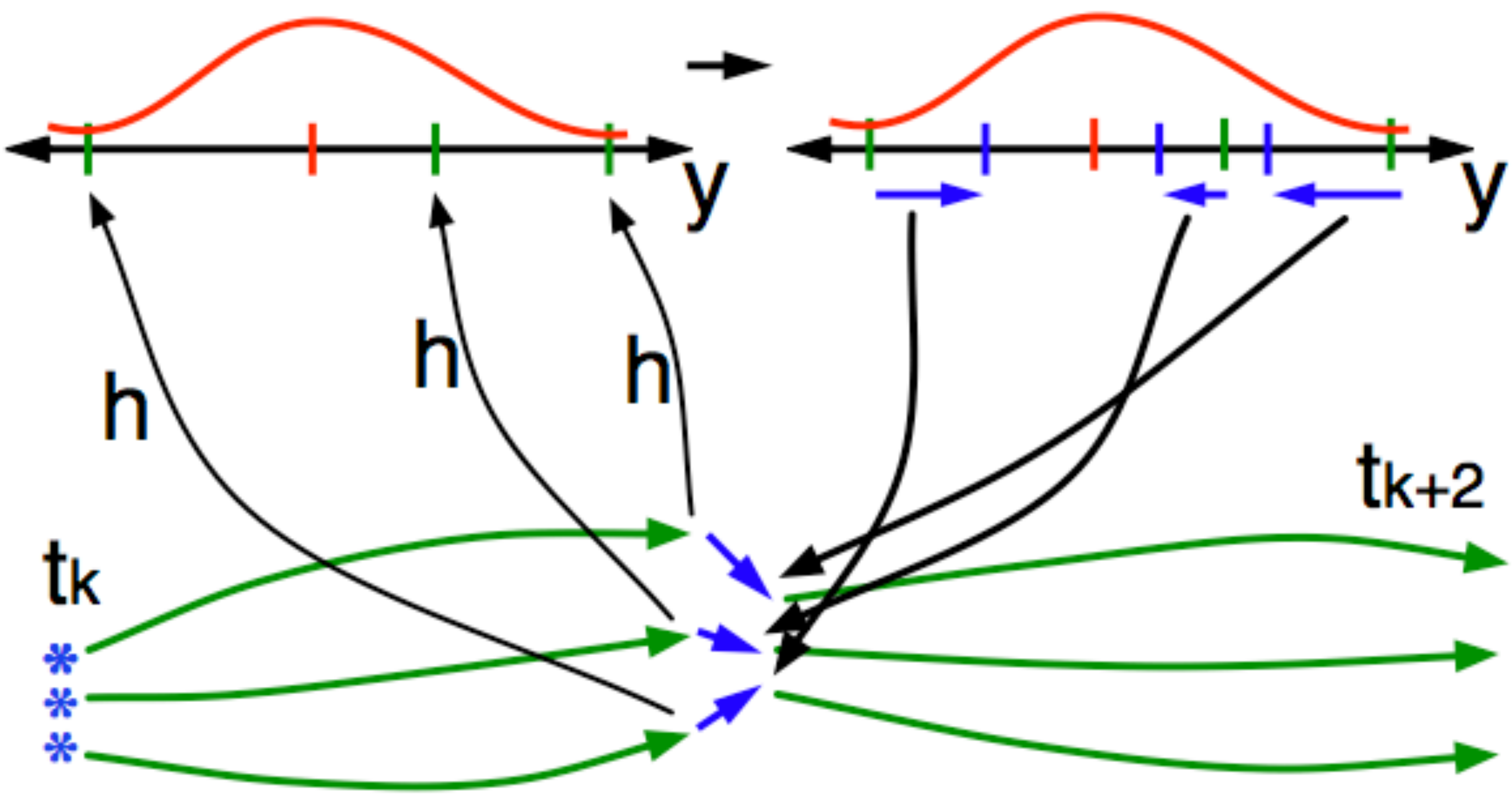
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.



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

Ensemble Kalman Filter: The Details

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



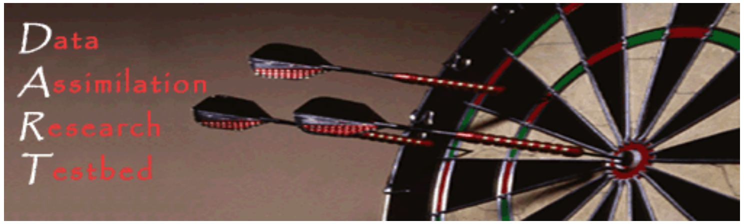
The Data Assimilation Research Section (DAReS)

Mission: To accelerate progress in Earth System Science at NCAR, UCAR Universities, and in the broader science community by providing state-of-the-art ensemble DA capabilities.

Method: DAReS develops and maintains the Data Assimilation Research Testbed, a community facility for ensemble data assimilation.

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, …
- Professional software engineering
  - Carefully constructed and verified.
  - Excellent performance.
  - Comprehensive documentation, examples, tutorials.
- People: The DAReS Team



DART is used at:

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

中央氣象局 Central Weather Bureau

UNIVERSITY OF WASHINGTON

CIRES

GFDL Geophysical Fluid Dynamics Laboratory Princeton, NJ

THE UNIVERSITY OF UTAH

NOAA

PENNSTATE

JPL
Jet Propulsion Laboratory
California Institute of Technology

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.

Some DART Capabilities

- Ensemble forecasts,
- Ensemble reanalysis,
- Explore predictability,
- Sensitivity analysis,
- Model improvement,
- Observing system evaluation,
- Observing system design,
- DA algorithm improvement.

DART is still Unique after nearly 2 Decades

Two Critical Science/Engineering Design Choices

1. We keep our fingers out of your model.

   No changes required to forecast model.

2. Single observation changing single model variable.

   Without loss of generality,

   Simplifies algorithms, parallelism.

Cool Science Example 1

DAReS Lead
Glen Romine (50% MMM)

Glen is at the Mesa Lab on Tuesday each week.

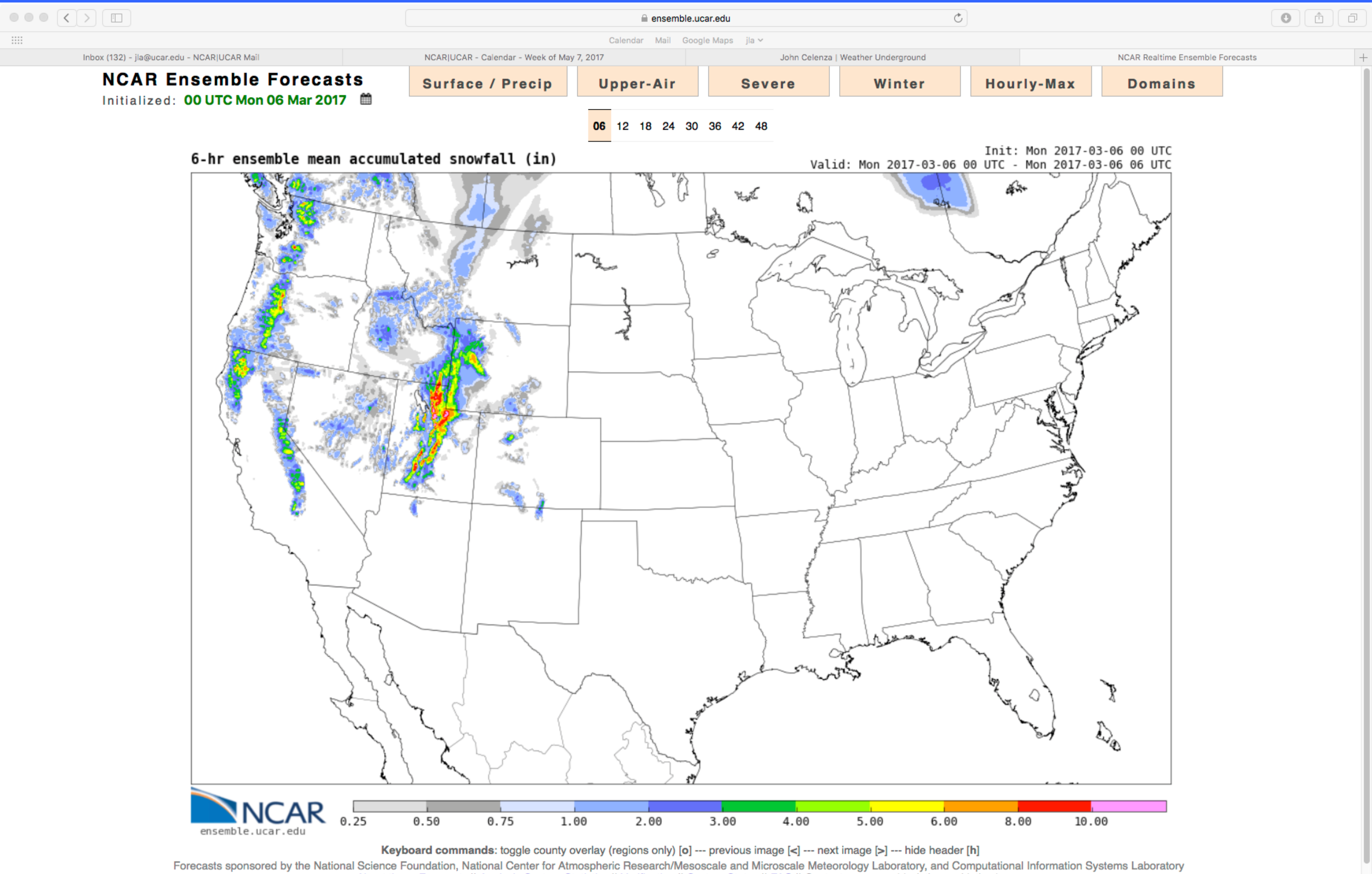
With DAReS since 2009.

Forecast Model
WRF, Weather Research and Forecasting Model

Science Collaborator
MMM, Oklahoma

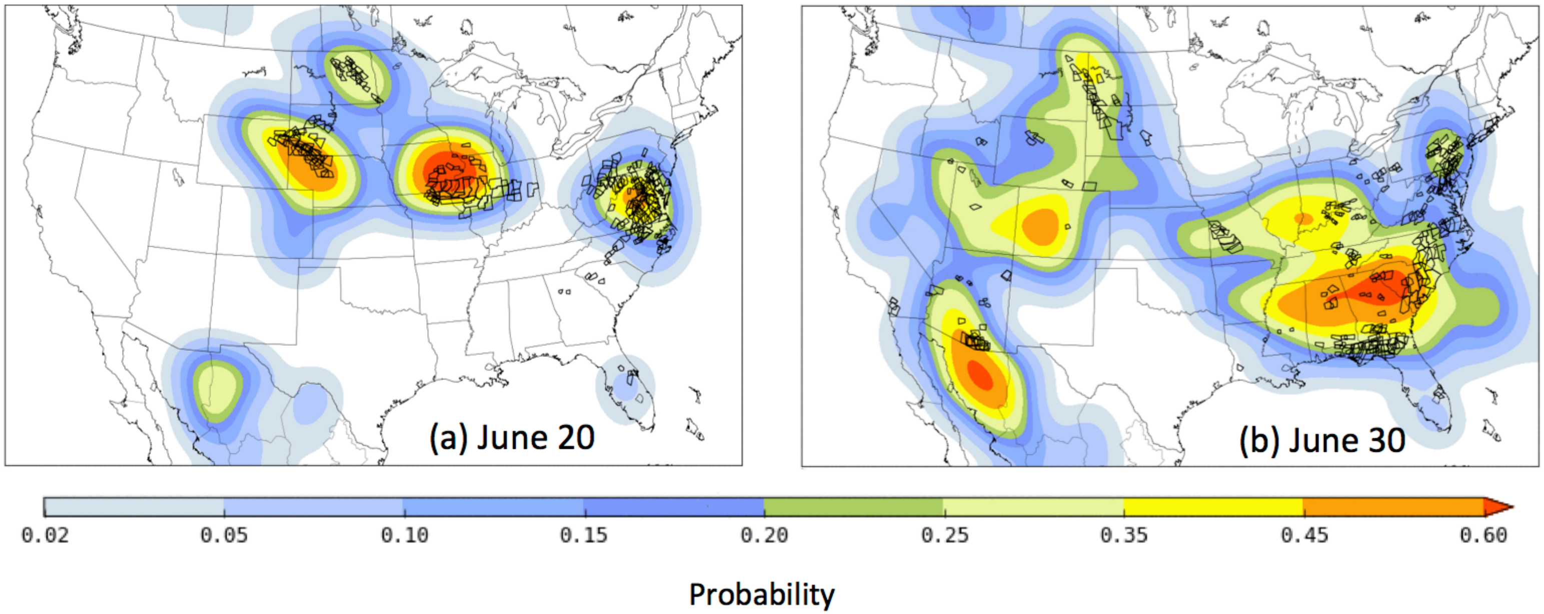
DA Capability
Ensemble Prediction

NCAR Real-time ensemble prediction system



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 from April 2015 to December 2017
- 48 hour forecasts at 3km resolution
- First continuously cycling ensemble system for CONUS
- CISL Dedicated Queues and Computing Support were Vital

Cool Science Example 2

DAReS Lead
Kevin Raeder

With DAReS since 2003.

Previously with CGD.

Forecast Model
CAM6 (Community Atmosphere Model)
CESM (Community Earth System Model)

Science Collaborator
CGD, Washington,
Arizona, Utah

DA Capability
Ensemble Reanalysis

An Ensemble Reanalysis with CAM in CESM: Motivation

1. Evaluate weather prediction capabilities of CAM
    - Confront climate model with observations
    - Identify systematic short-term forecast errors
    - Compare to earlier CAM reanalysis

2. Provide forcing for CESM component model simulations
    - POP ocean model
    - CLM land surface
    - CICE sea ice model
    - Offline chemistry transport models

An Ensemble Reanalysis with CAM in CESM: Logistics

Target period from 1999-present.

Two overlapping streams starting in 1999, 2010.

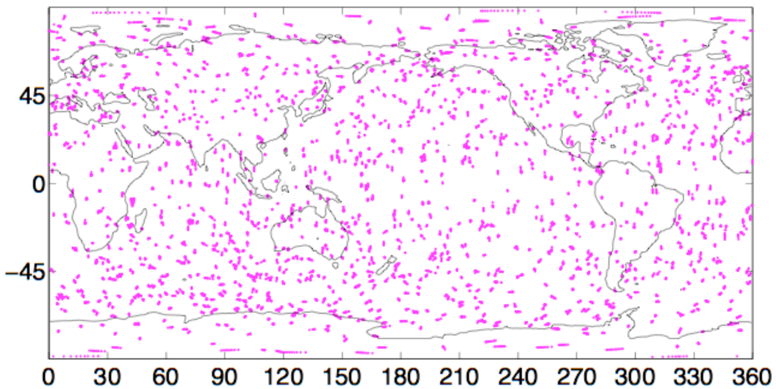
Lots of computing (50 million Cheyenne core hours).

Thanks to NSC allocation and help from many CISL staff.

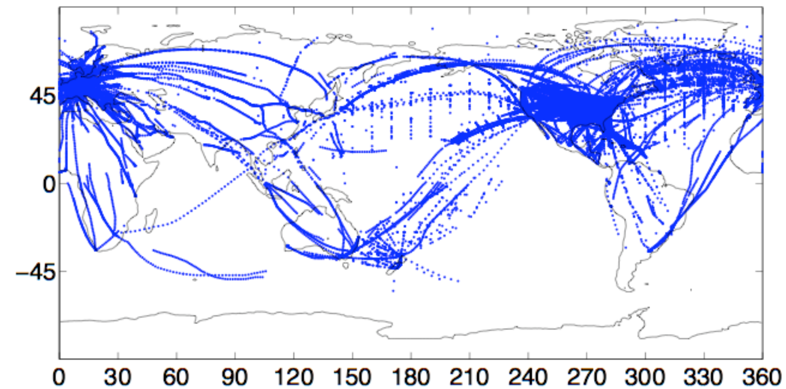
An Ensemble Reanalysis with CAM in CESM: Observations

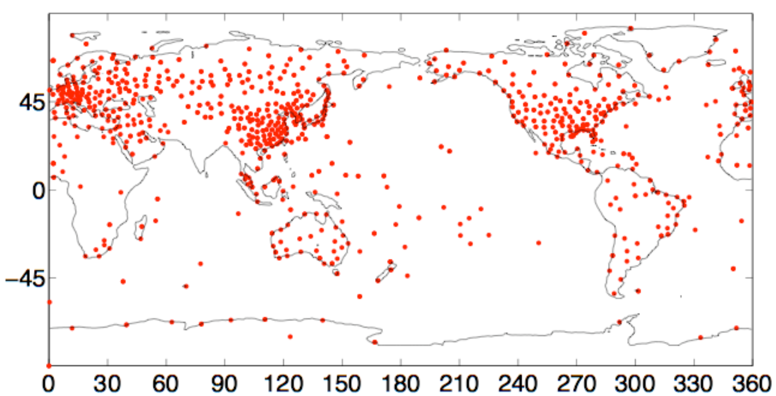
Examples for single assimilation window, also assimilating AIRS satellite temperatures.

GPS

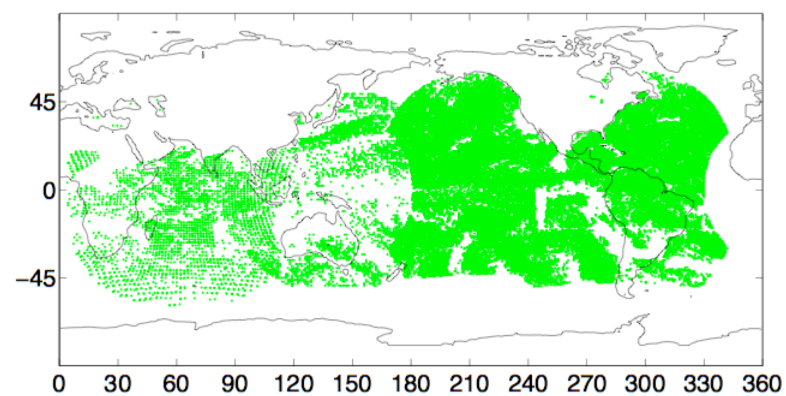


ACARS and Aircraft



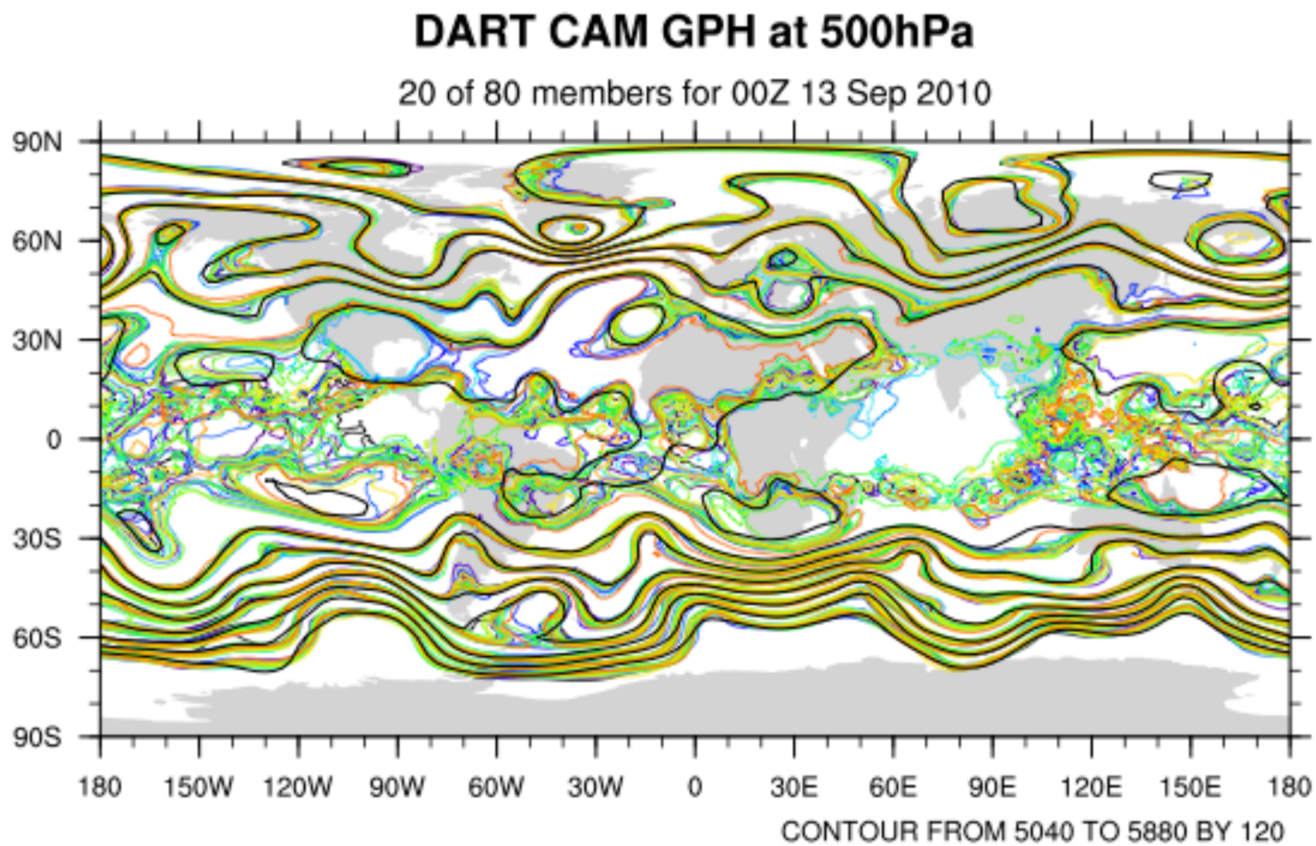


Radiosondes



Sat Winds

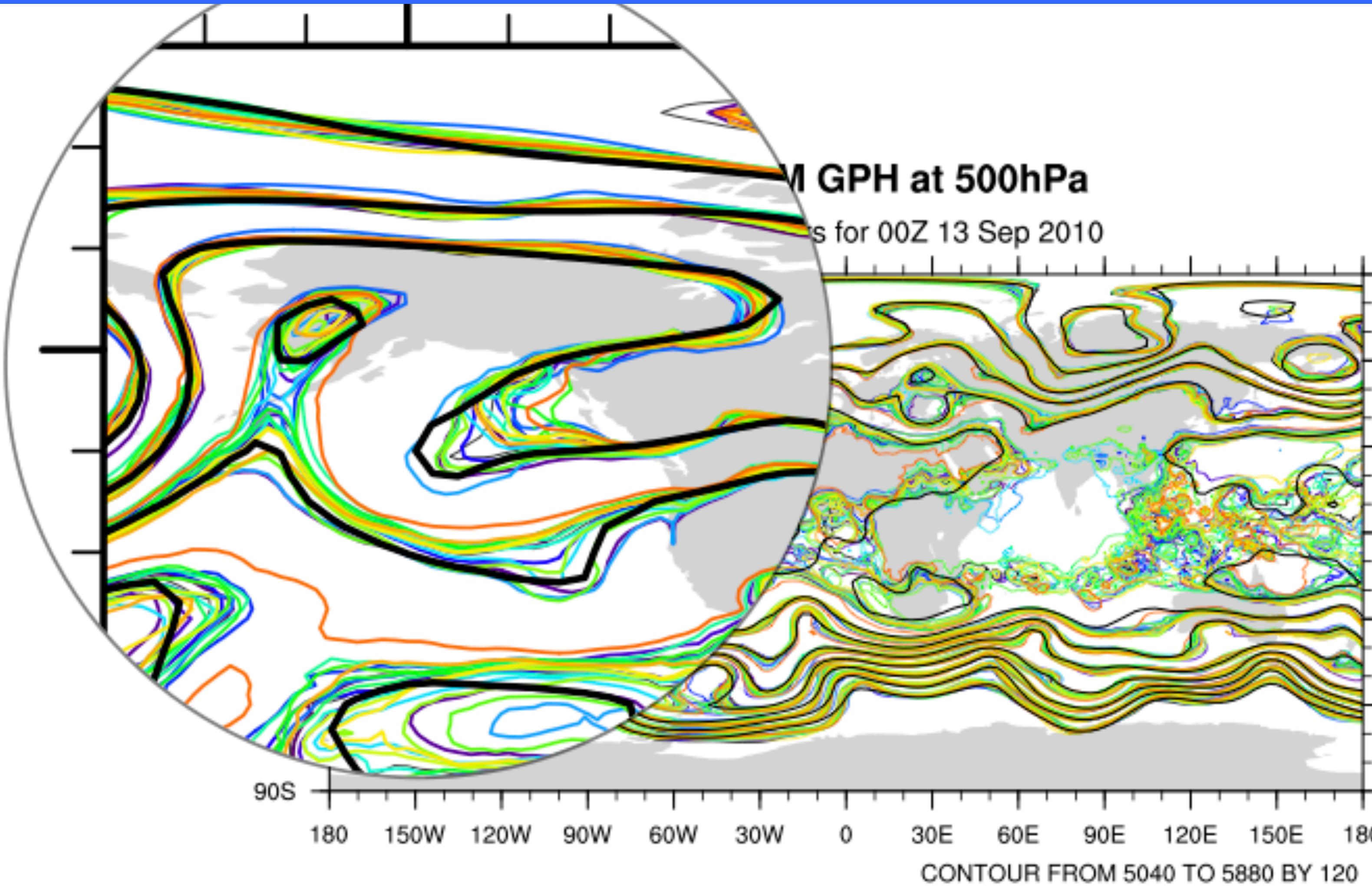
An Ensemble Reanalysis with CAM in CESM: Results



Color contours from DART (20 of 80 ensemble members). Show Uncertainty.

Black from operational NCEP analysis.

An Ensemble Reanalysis with CAM in CESM: Results



Color contours from DART (20 of 80 ensemble members). Show Uncertainty.

Black from operational NCEP analysis.

Cool Science Example 3

DAReS Leads

Moha El Gharamti

With DAReS since 2016.

Ben Johnson

With DAReS since September.

Forecast Model

WRF-Hydro®

Nearly the same as the National Water Model

Science Collaborator

RAL (James McCreight), Texas Arlington

DA Capability

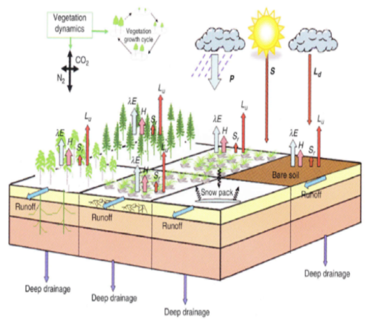
Model Improvement

WRF-Hydro/DART: The Model

Weather Forcing Engine

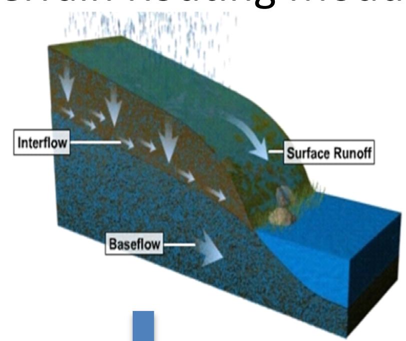
WRF-Hydro: https://www.ral.ucar.edu/projects/wrf_hydro

NoahMP
Land Surface Model

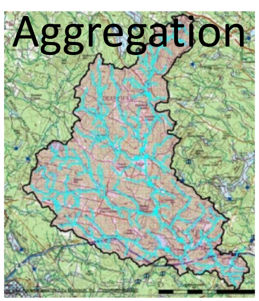


2-way coupling

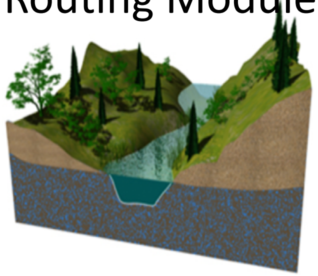
Terrain Routing Module

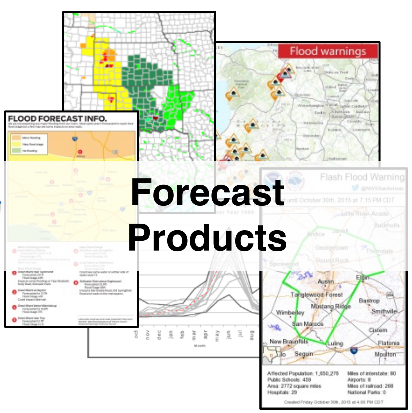


NHDPlus Catchment
Aggregation

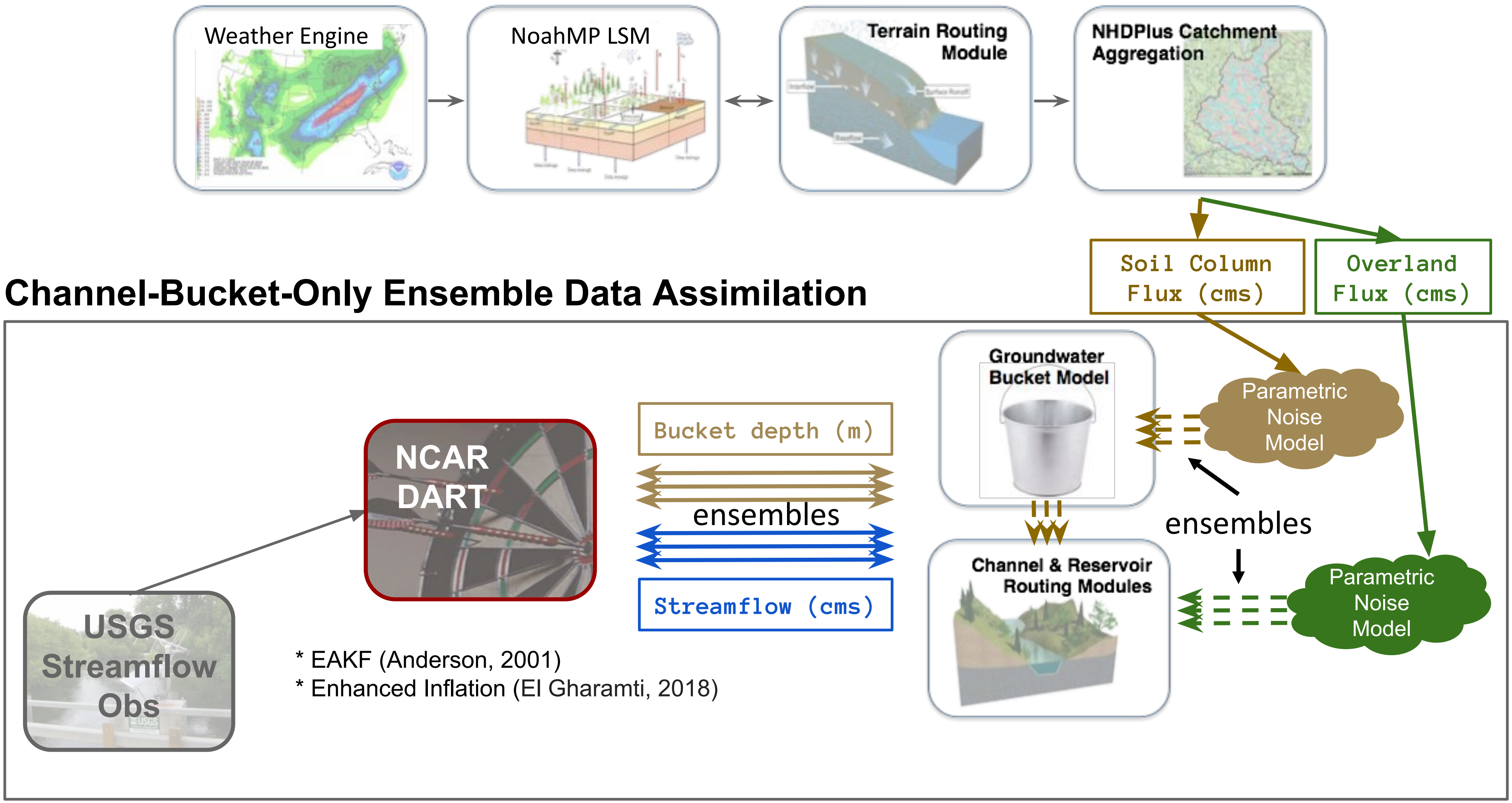


Channel & Reservoir
Routing Module



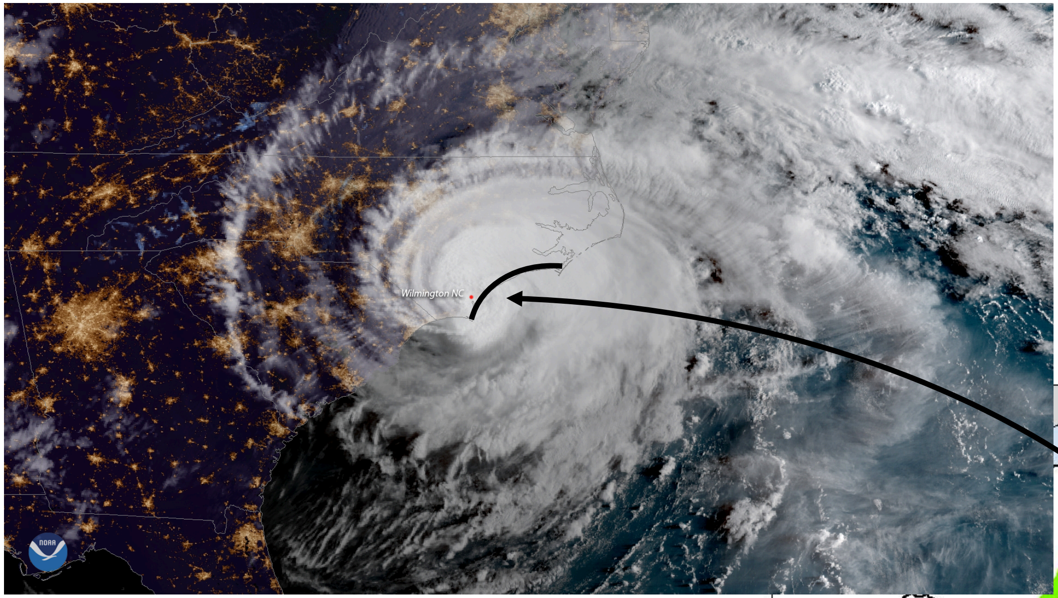


WRF-Hydro/DART: DA System

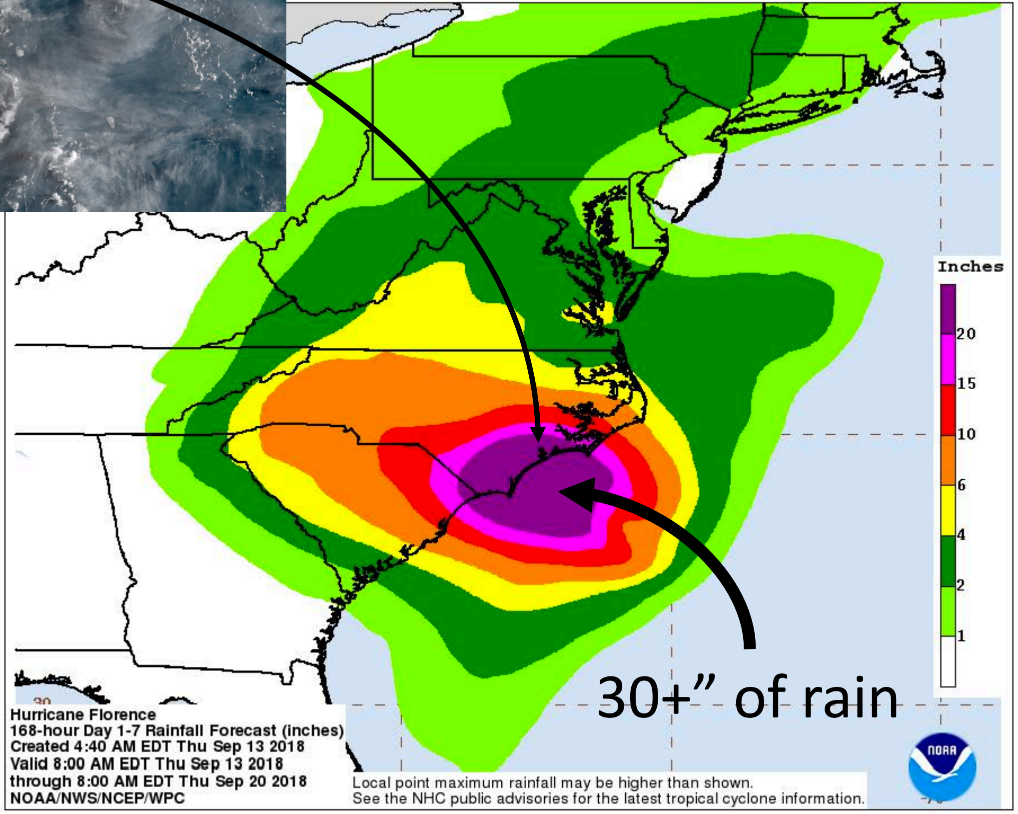


Python environment github.com/NCAR/wrf_hydro_py.git

WRF-Hydro/DART: Florence 2018



Hurricane Florence made landfall near Wrightsville Beach, North Carolina at **7:15 a.m. ET September 14**. The GOES East satellite captured this geocolor image at 7:45 a.m. ET



Winds up to 150 mph (240 km/hr)
Damage: $24.23 billion
NOAA/NWS/NCEP/WPC

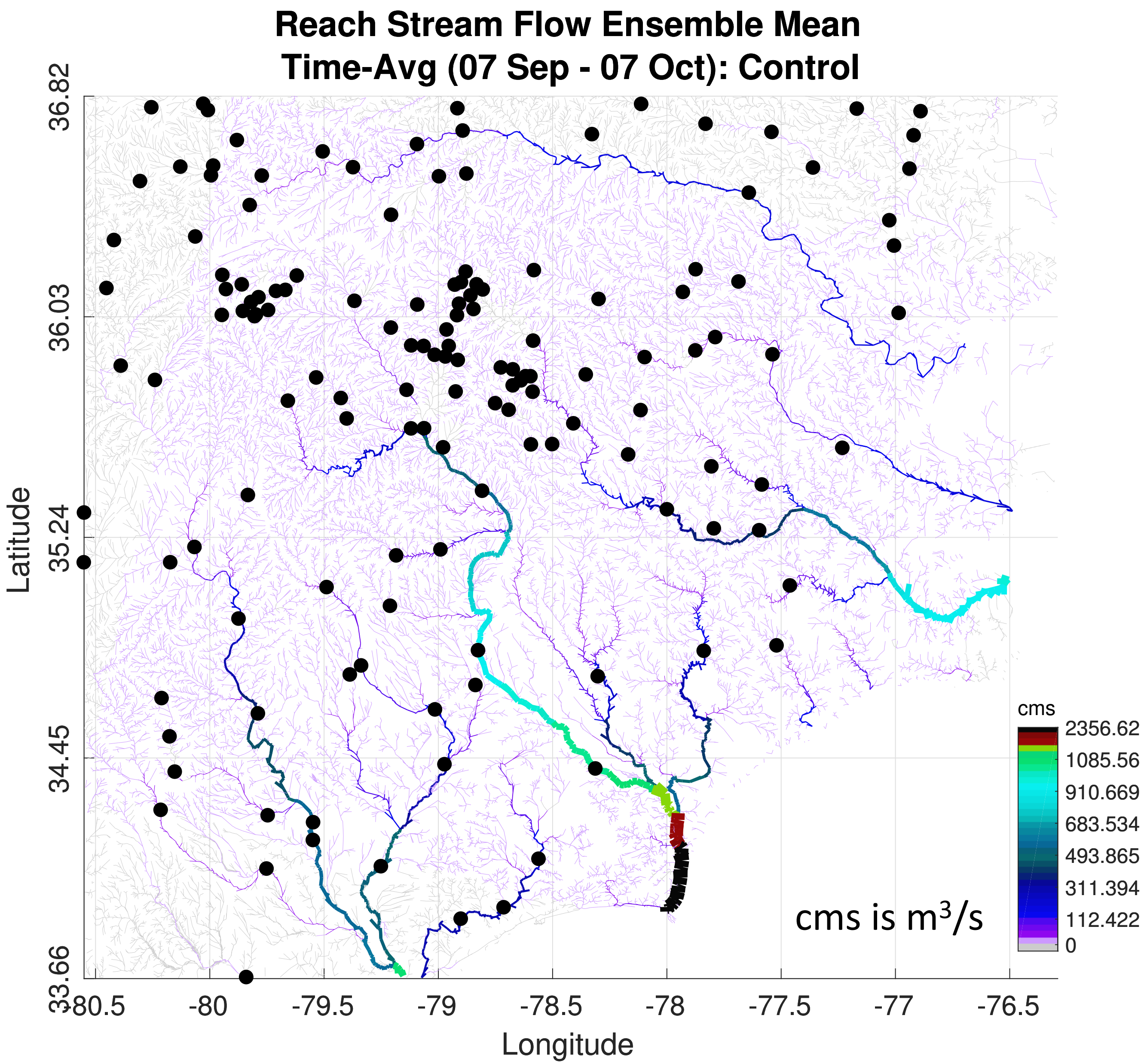
WRF-Hydro/DART: The Florence Region

WRF-Hydro/DART: No DA Control

Monthly mean of the model. The streamflow is driven by the precipitation.

More than 100 gauges, reporting every 15 mins.

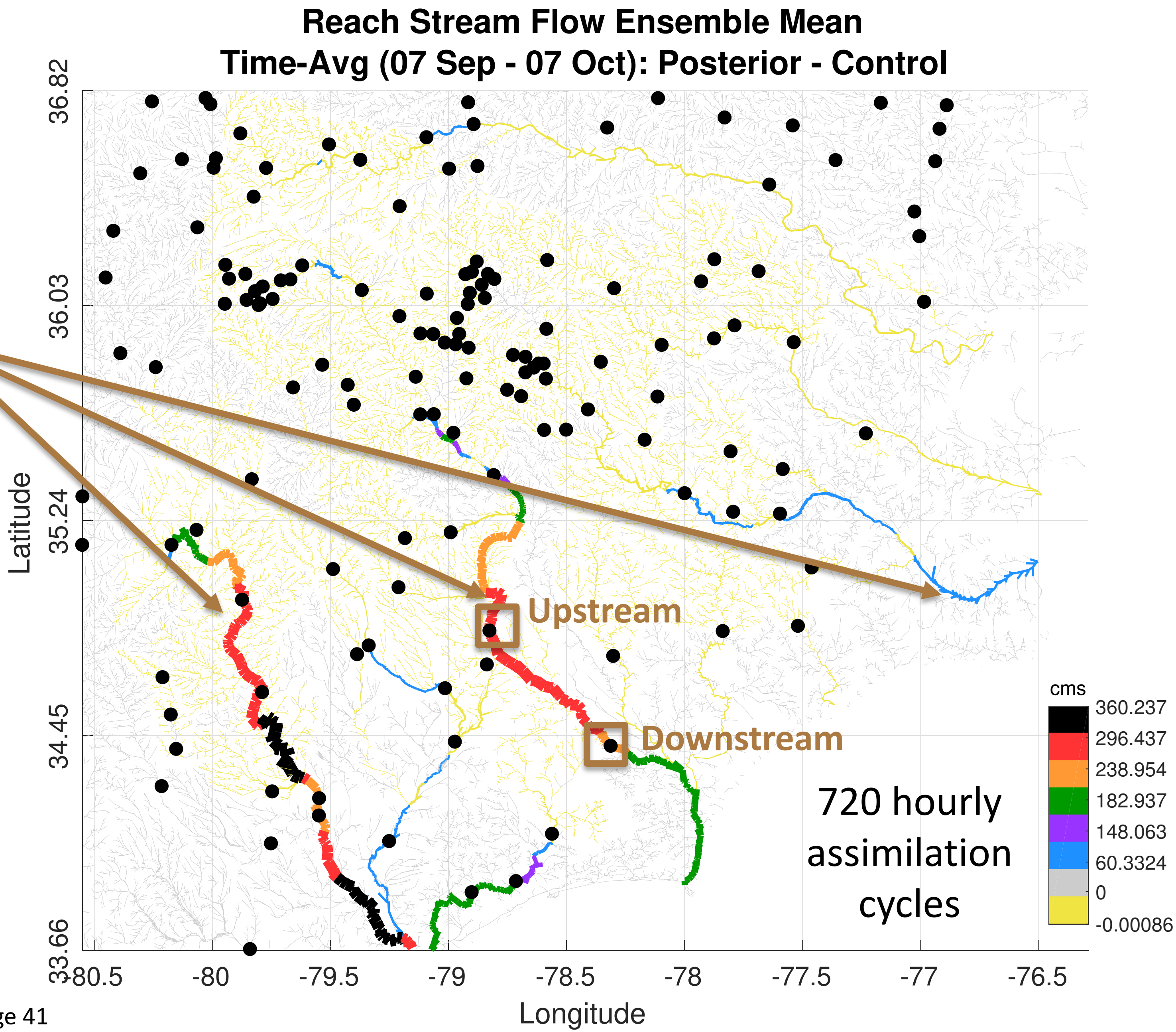
Now, what happens when streamflow gauge data is incorporated through DA?



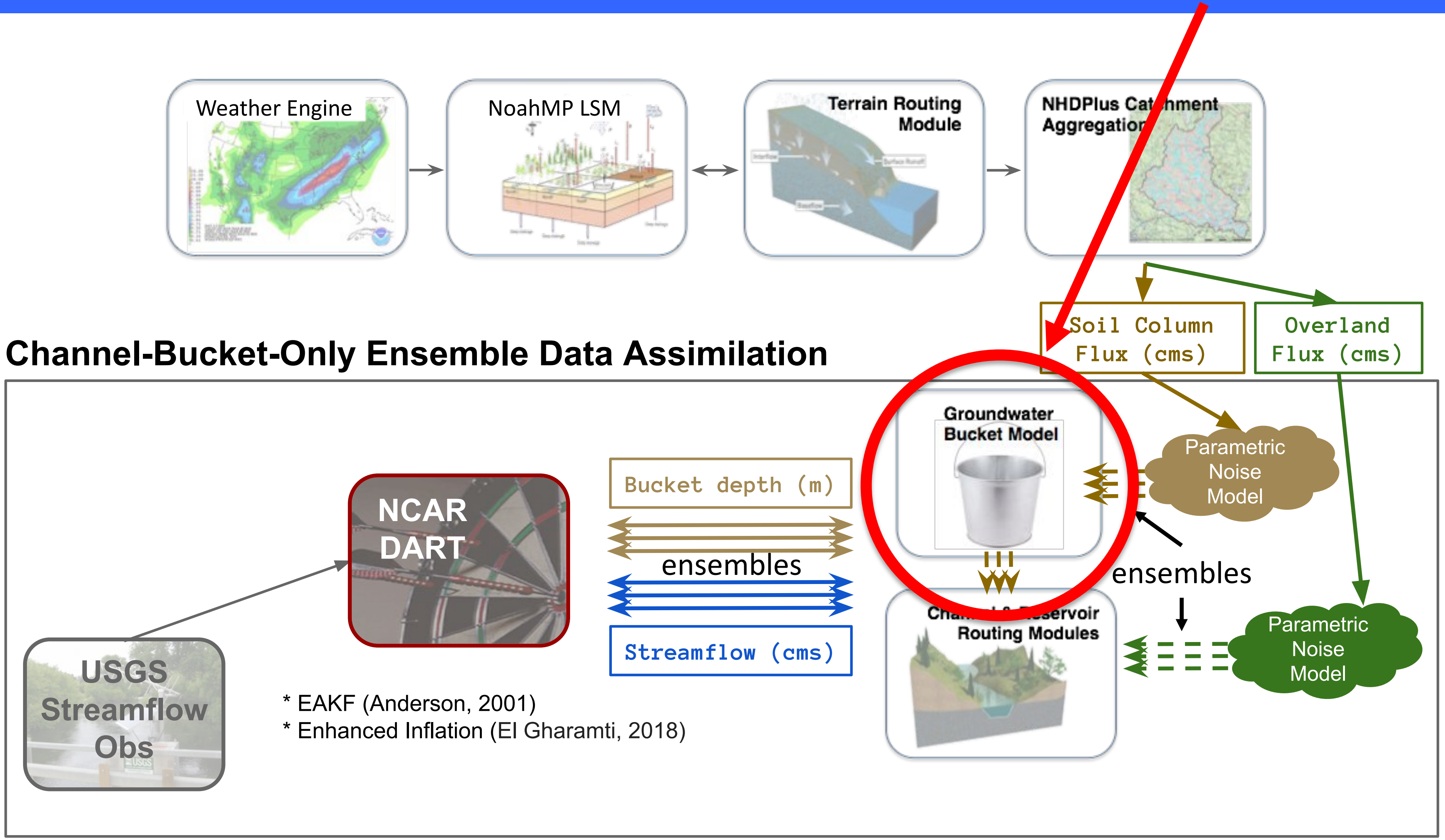
WRF-Hydro/DART: DA Impact

Assimilation happens every hour

Correction along major reaches. DA is adding water to the stream channels.



WRF-Hydro/DART: Bucket Problems



* EAKF (Anderson, 2001)
* Enhanced Inflation (El Gharamti, 2018)

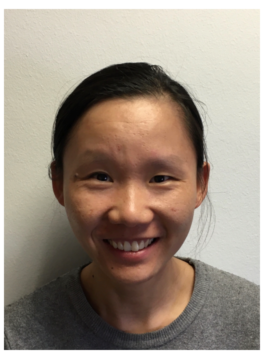
Python environment github.com/NCAR/wrf_hydro_py.git

Cool Science Example 4

<u>DAReS Leads</u>
Tim Hoar

With DAReS since 2003.

Xueli Huo

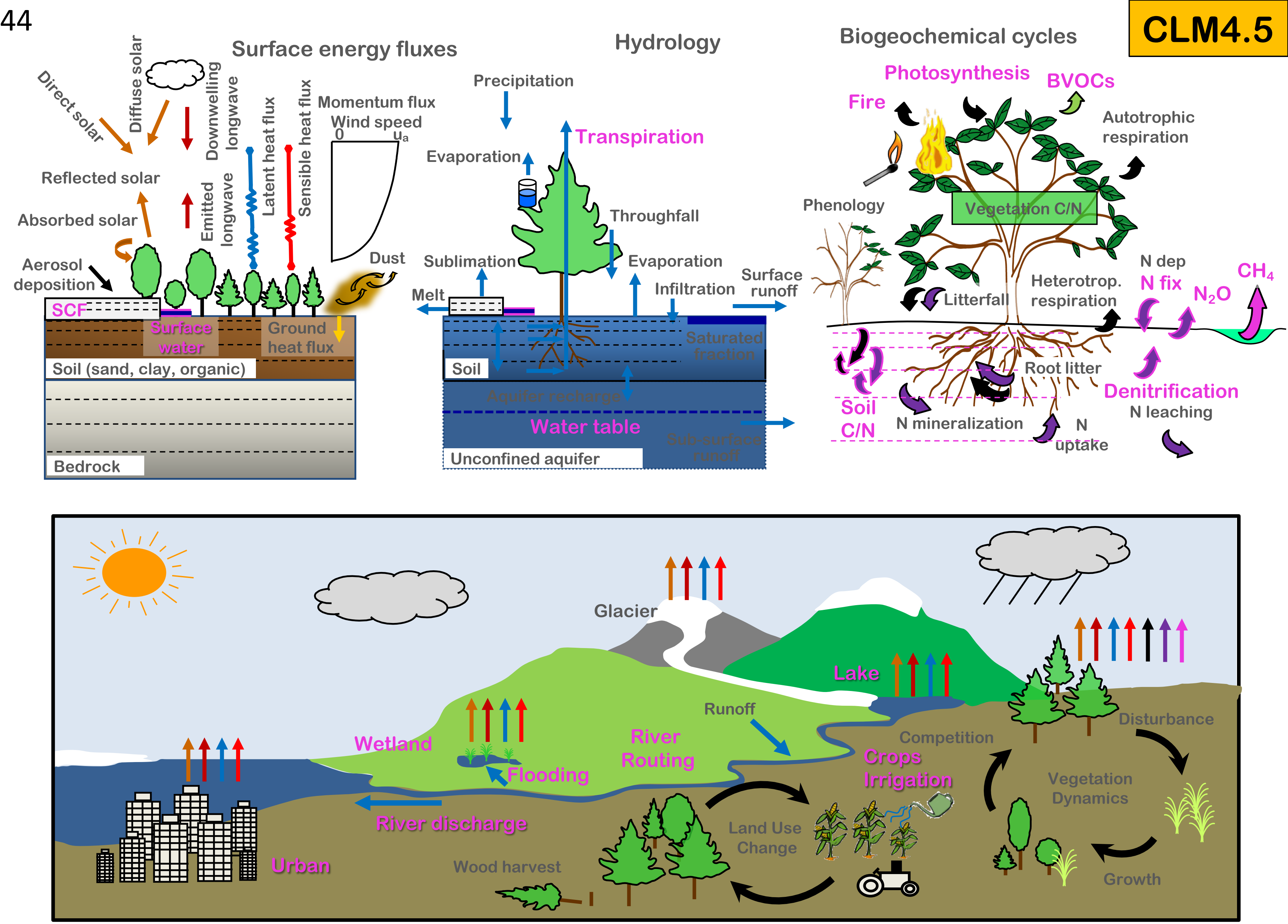


Long-term visitor from Arizona.

<u>Forecast Model</u>
CLM (Community Land Model)

<u>Science Collaborator</u>
UCAR (Andy Fox)
CGD, Arizona, Utah

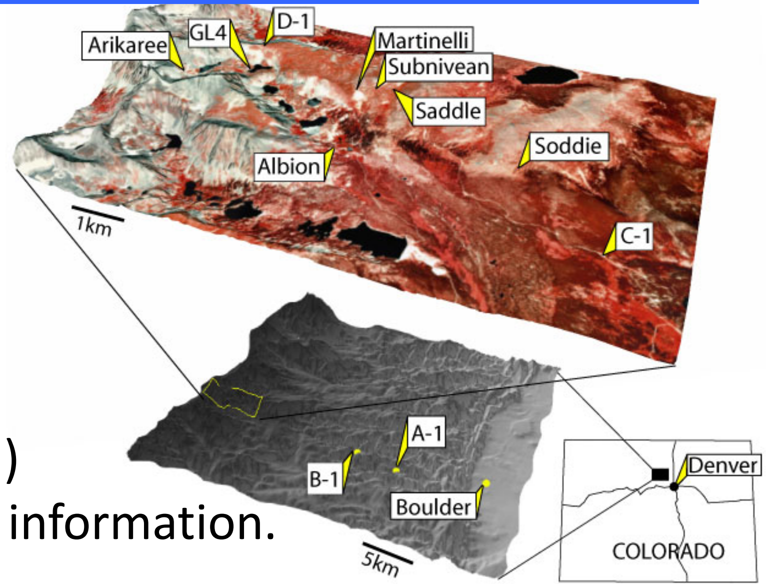
<u>DA Capability</u>
Predictability



Slides from Dave Lawrence, CGD.

Estimating Ecosystem Variables in CLM

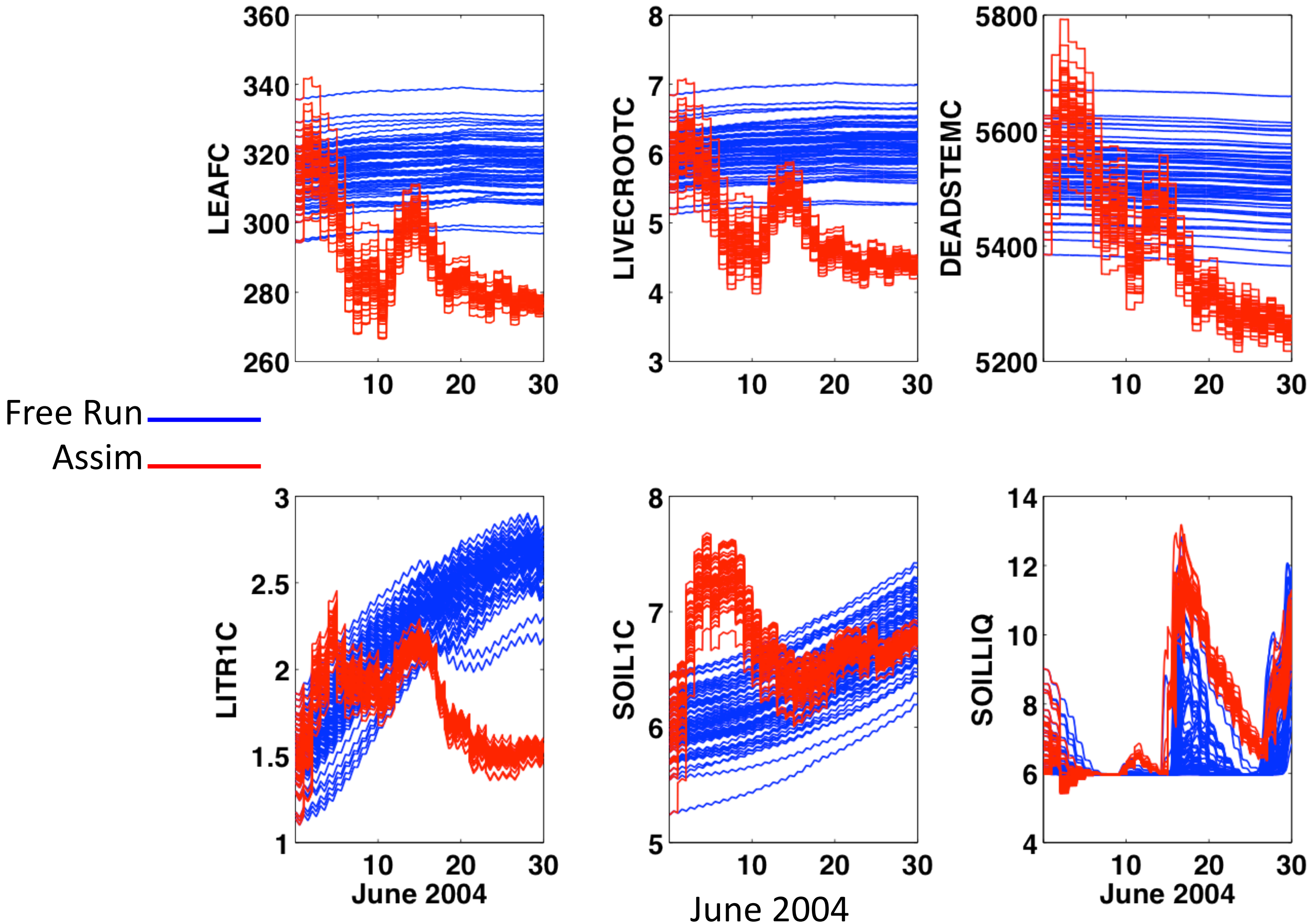
In collaboration with Andy Fox (U. Arizona)
An experiment at Niwot Ridge



- 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*.

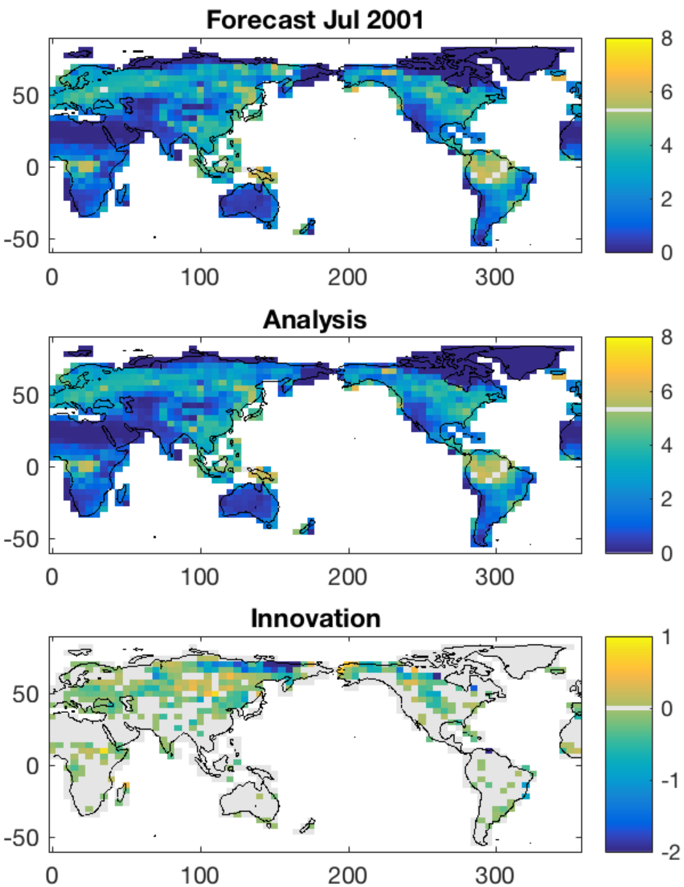
Estimating Ecosystem Variables in CLM

Unobserved variables are updated.



Estimating Ecosystem Variables in CLM

Global case. Remote sensing DA changes to Leaf Area Index estimates.



Cool Science Example 5

<u>DAReS Lead</u>
Jeff Anderson

With DAReS since before it existed.

<u>Forecast Model</u>
WACCM-X
Whole Atmosphere Community Climate Model, Extended Top

<u>Science Collaborator</u>
HAO (Nick Pedatella)
Colorado

<u>DA Capability</u>
Observing System Evaluation

Deep Atmospheric Component Coupled DA

WACCMX:
- 2 degrees, 126 levels, top at 4.1×10^{-10} hPa (more than 500 km)
- High-top extension of CAM
- Includes ionospheric processes
- Persistence forecasts of solar and geomagnetic forcing

Observations:
- All in situ plus GPS refractivity in troposphere/lower stratosphere
- Temperature from AURA Microwave Limb Sounder (MLS)
- Temperature from TIMED/SABER
- Temperatures only up to 100km

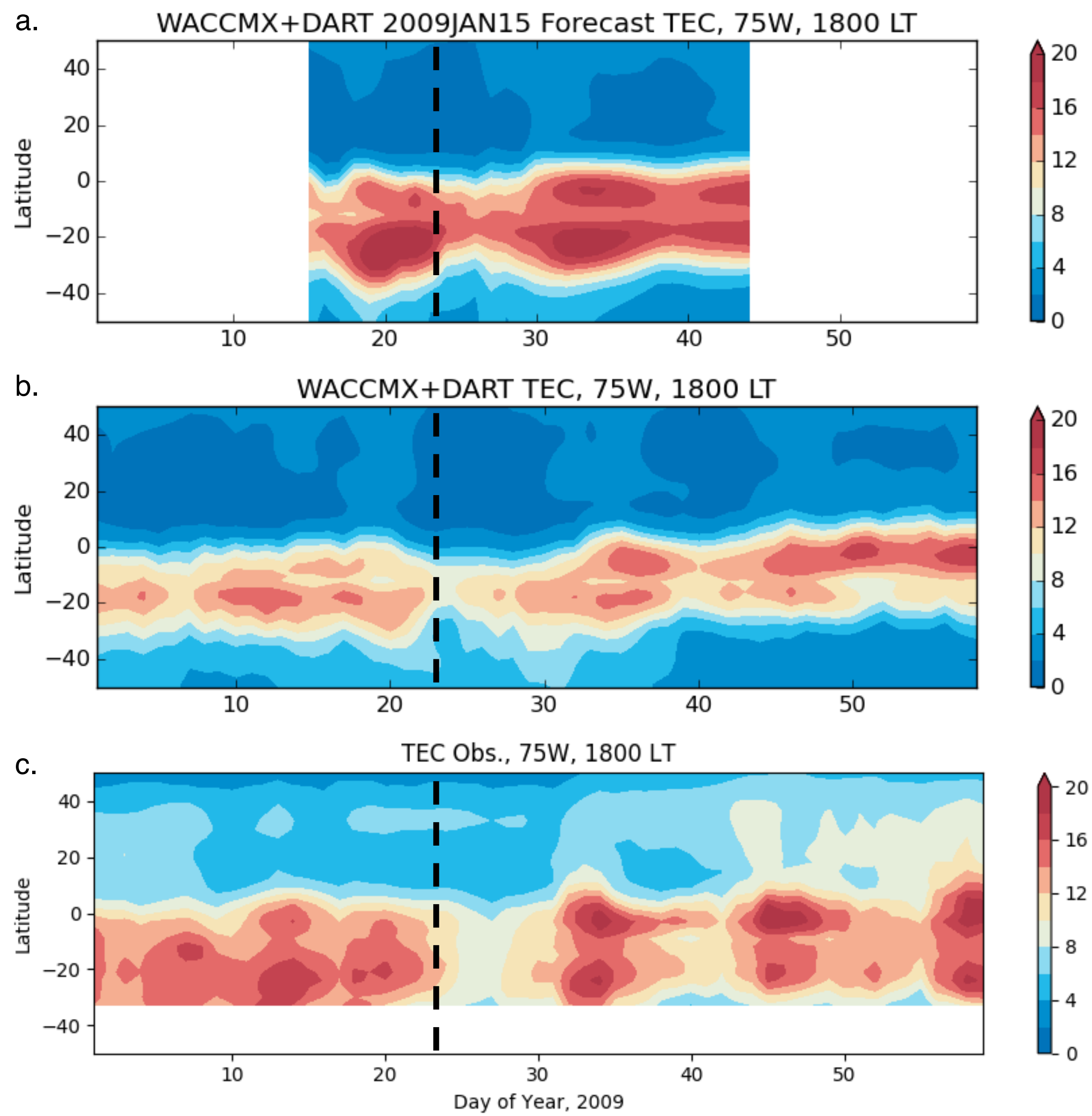
DART:
- 40 members
- Adaptive inflation, GC localization
- 6-hour window

Deep Atmospheric Component Coupled DA

Impact of Stratospheric Sudden Warming on ionosphere

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

Agreement of forecast with observations indicates significant prediction skill.



Cool Science Example 6

<u>DAReS (Honorary) Lead</u> Arthur Mizzi (Had ACOM/CISL joint)  Now with State of Colorado. Often at Mesa Lab on Friday.	<u>Forecast Model</u> WRF/Chem (WRF with Chemistry) <u>Science Collaborator</u> State of Colorado ACOM UC Berkeley <u>DA Capability</u> Observing System Design

Air Quality Prediction Example

Xueling Liu, Ron Cohen, Inez Fung UC Berkeley.

Build NO_2 prediction system for Denver Metro.

Model is WRF/CHEM.

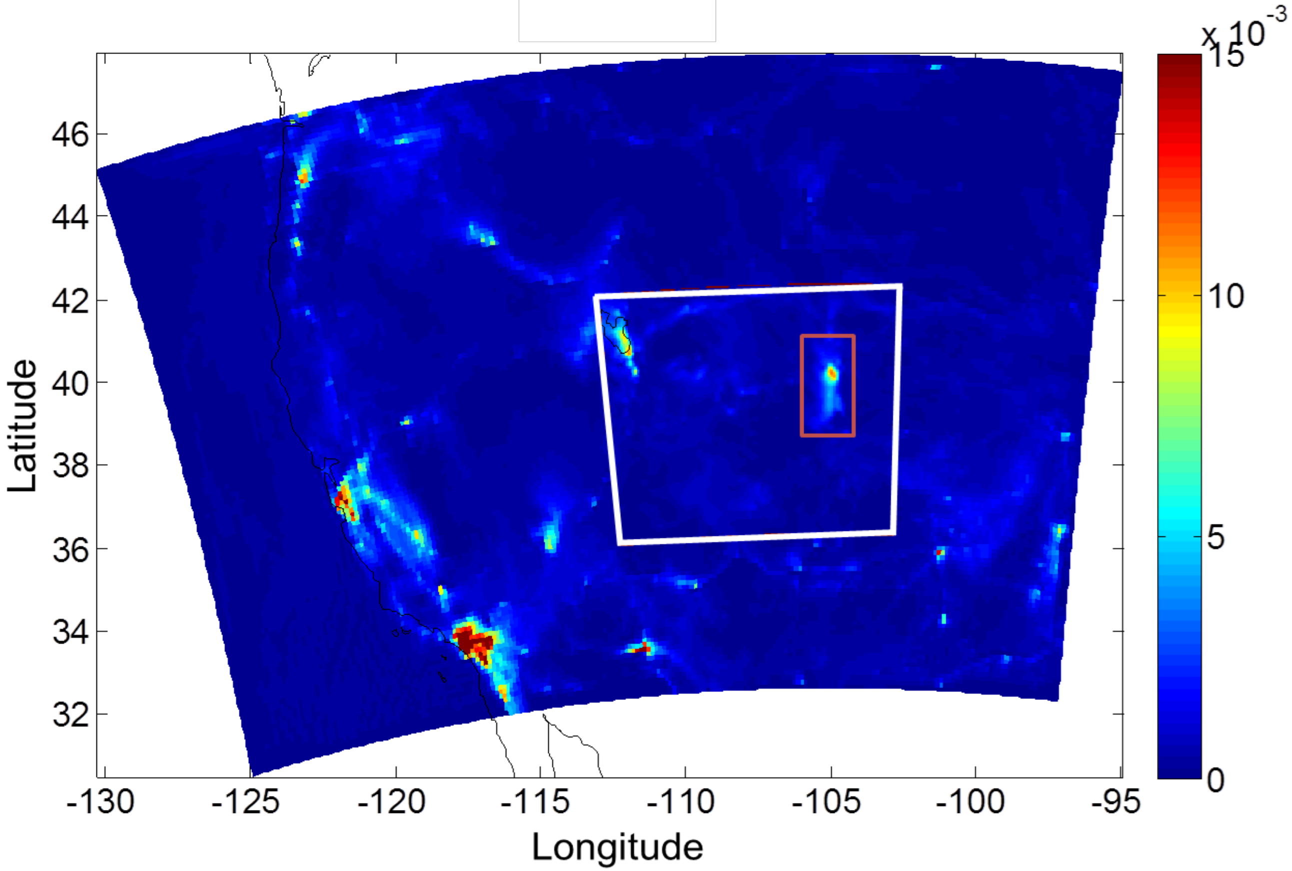
Observations are in situ plus satellite NO_2 (and NWP obs).

An observing system simulation experiment.

(No real observations yet).

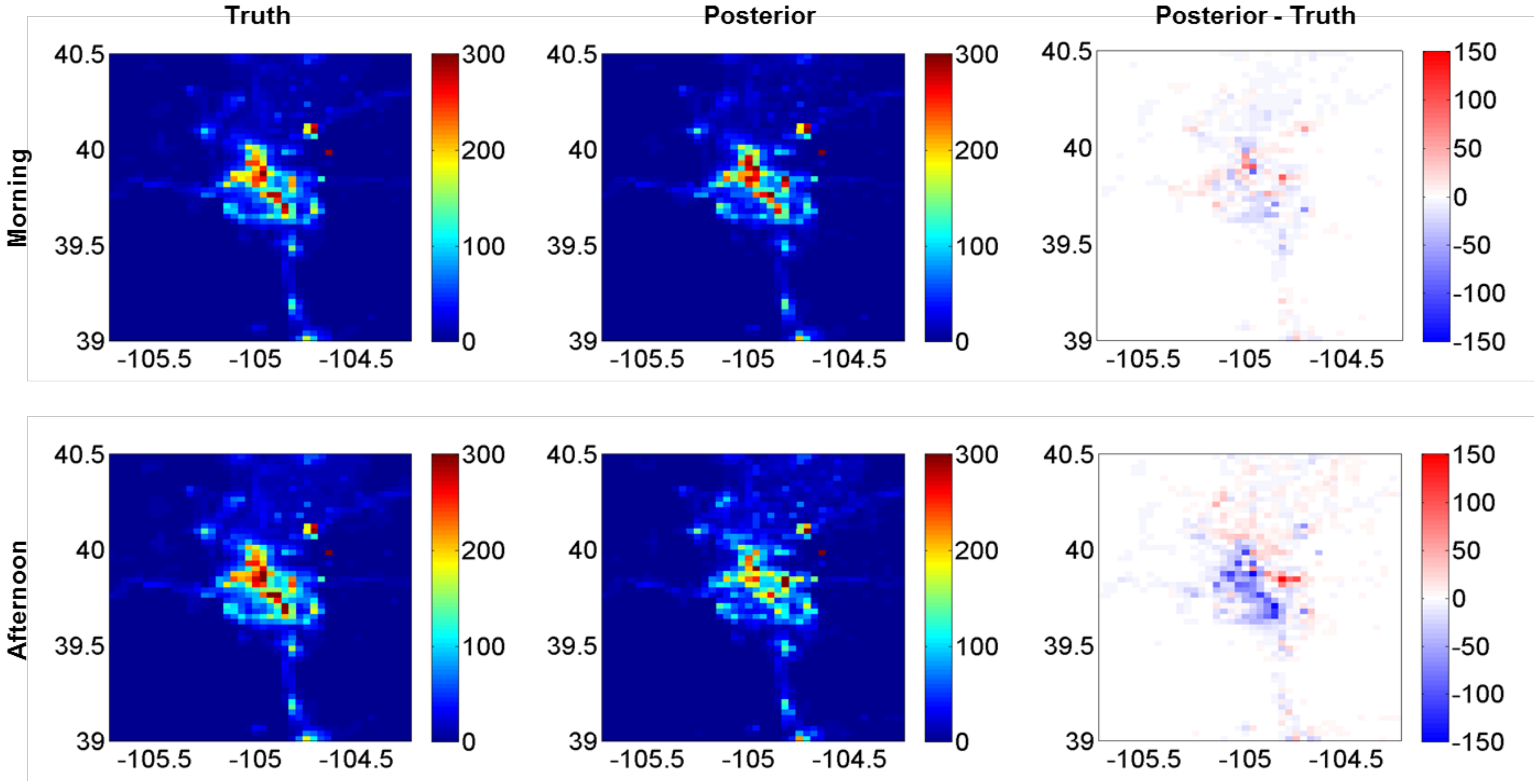
Source estimation at km scale is eventual goal.

Air Quality Prediction Example



12 km outer domain and 3 km inner domain.
Weather observations assimilated on inner domain.
TEMPO NO_2 observations assimilated in red rectangle.

Air Quality Prediction Example



System also estimates emissions.

9:00 am (top) 4:00 pm (bottom) on July 3rd.

Comparison of analysis to specified truth.

Cool Science Example 7

DAReS Leads

Jeff Steward

With DAReS since September.

Nancy Collins

With DAReS since 2006.

Forecast Model

WP-GITM (Ocean Wave Propagation-Global Ionosphere Thermosphere Model)

Science Collaborator

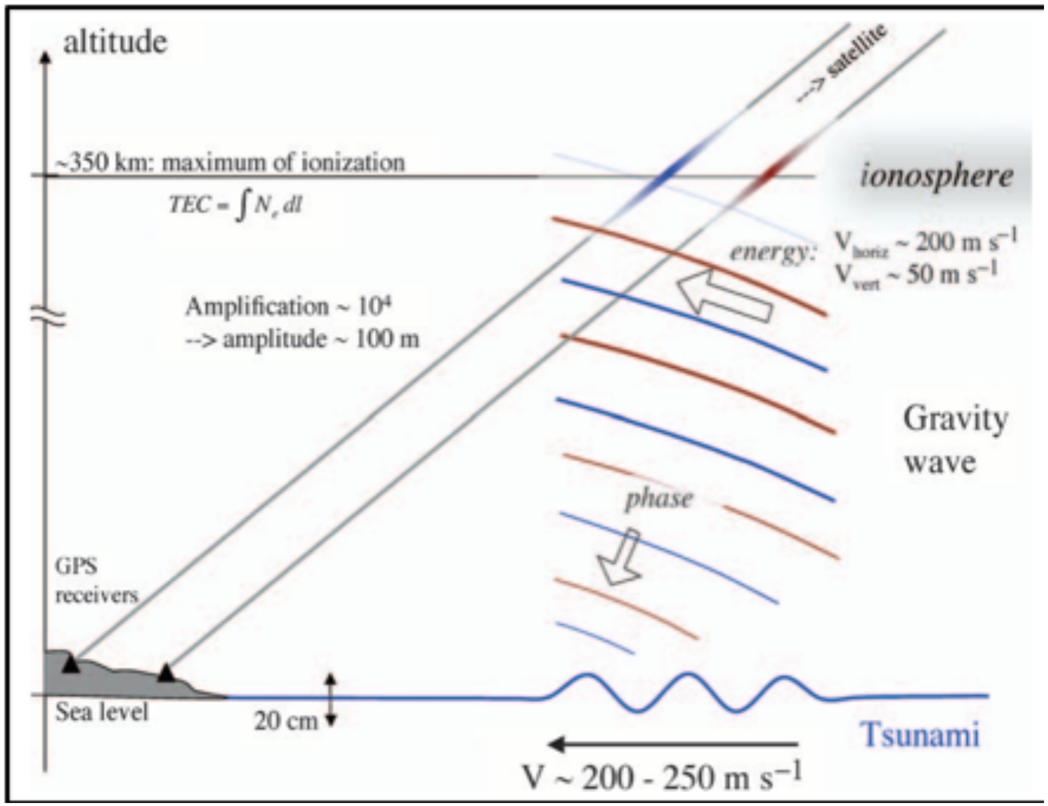
JPL (Panagiotis)

DA Capability

Sensitivity

Observing Tsunamis via the Ionopshere

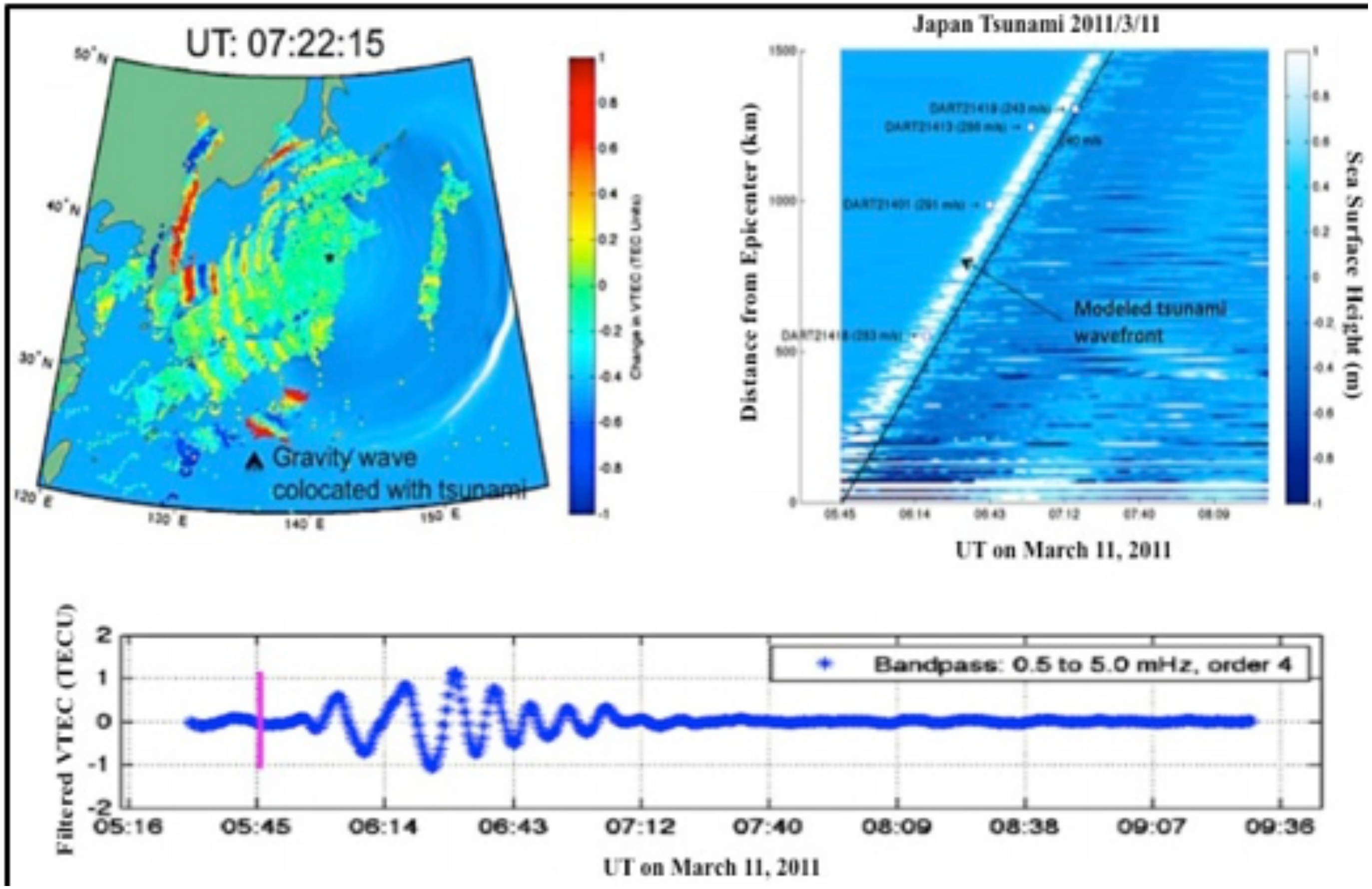
Tsunamis make very small changes to sea surface height in open ocean.
But, waves amplify in the atmosphere, 100m plus amplitude in ionosphere.
Changes Total Electron Content (TEC) of Ionosphere.
GPS signals are slowed by electrons.
Delays at ground stations can detect tsunami impacts in ionosphere (no way)!



Observing Tsunamis via the Ionopshere

Tohuku example: Gravity waves in ionosphere over tsunami waves in ocean.

Use DART to find tsunami by assimilating GPS observations of TEC.



Cool Science Example 8

<u>DAReS Lead</u>
Dan Amrhein (80% CGD)

Dan sits in A-tower but joins DAReS software standups.

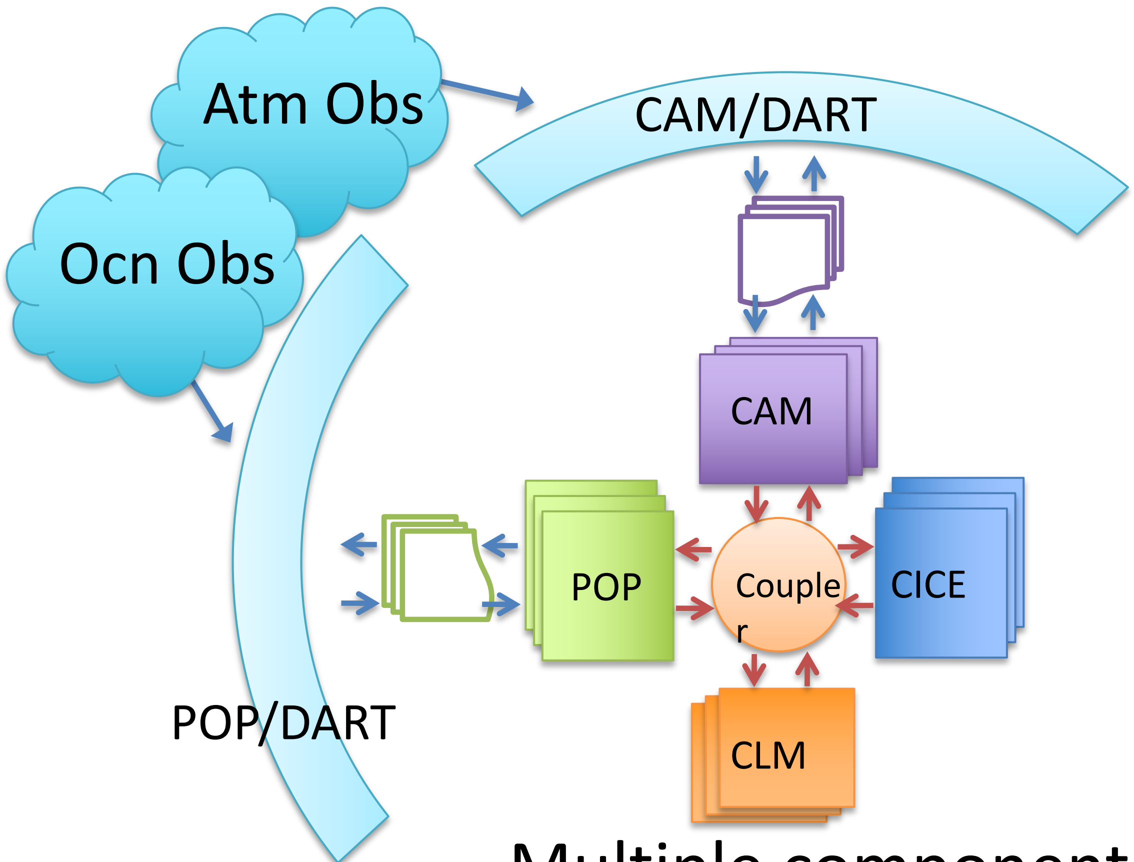
With DAReS since September.

<u>Forecast Model</u>
POP/CAM/CLM/CICE in CESM

<u>Science Collaborator</u>
CGD (Alicia Karspeck), Washington, Oklahoma

<u>DA Capability</u>
Assimilation methodology.

Multiple Component POP/CAM Coupled DA



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

Results from
Alicia 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

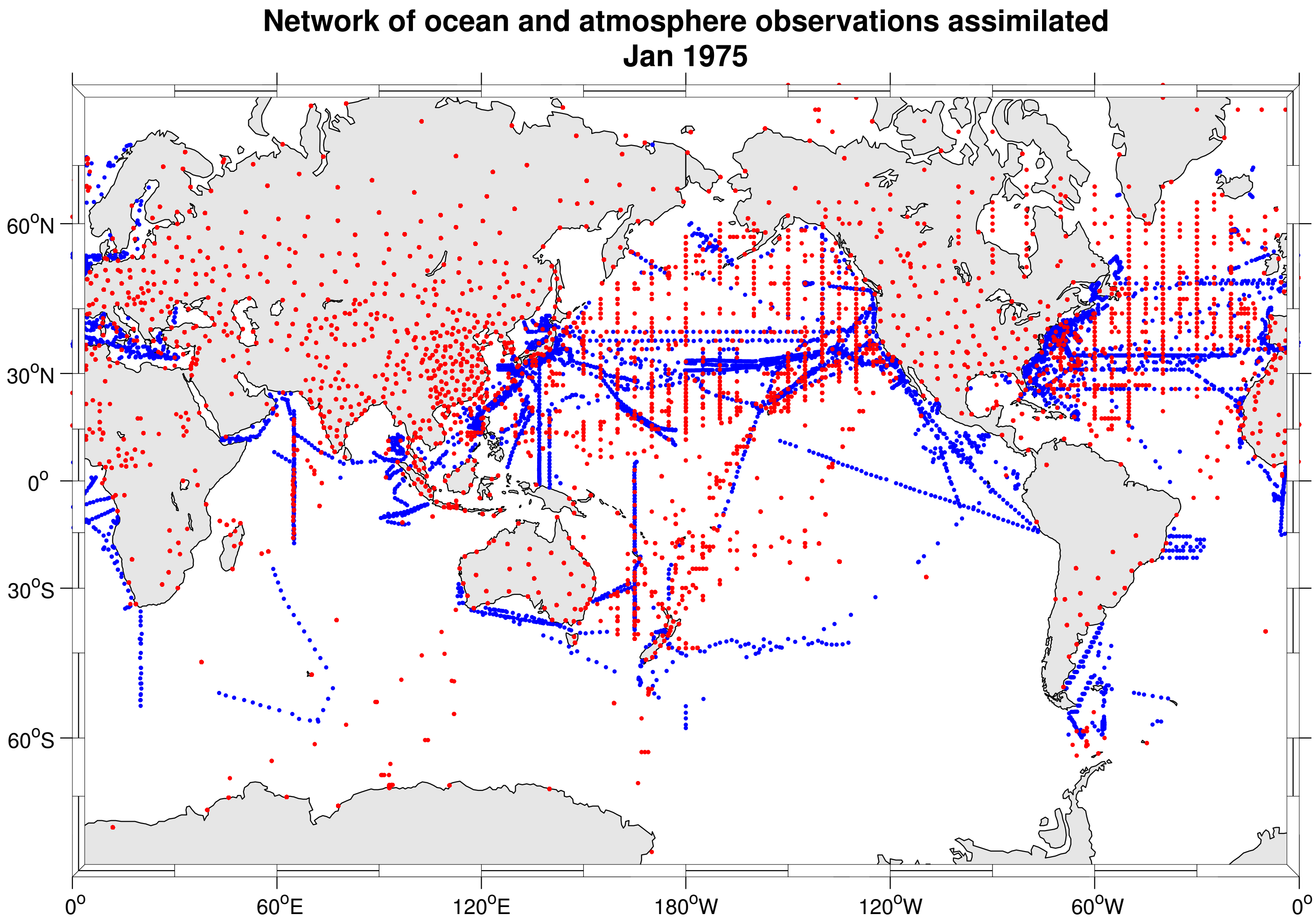
Observations:

- 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.

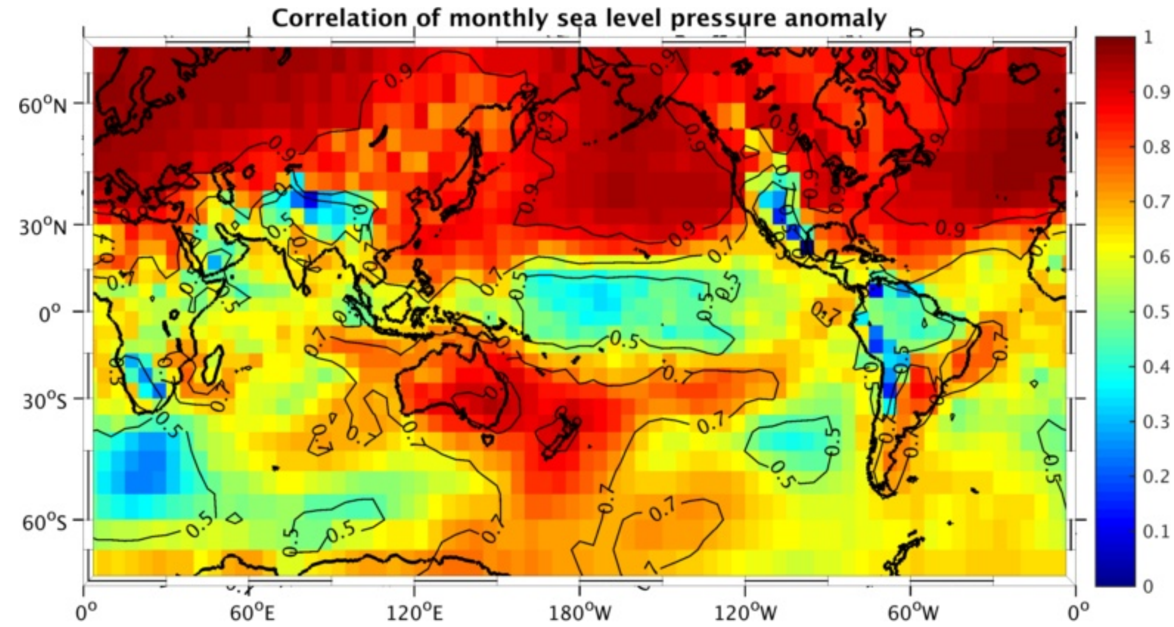
Multiple Component POP/CAM Coupled DA

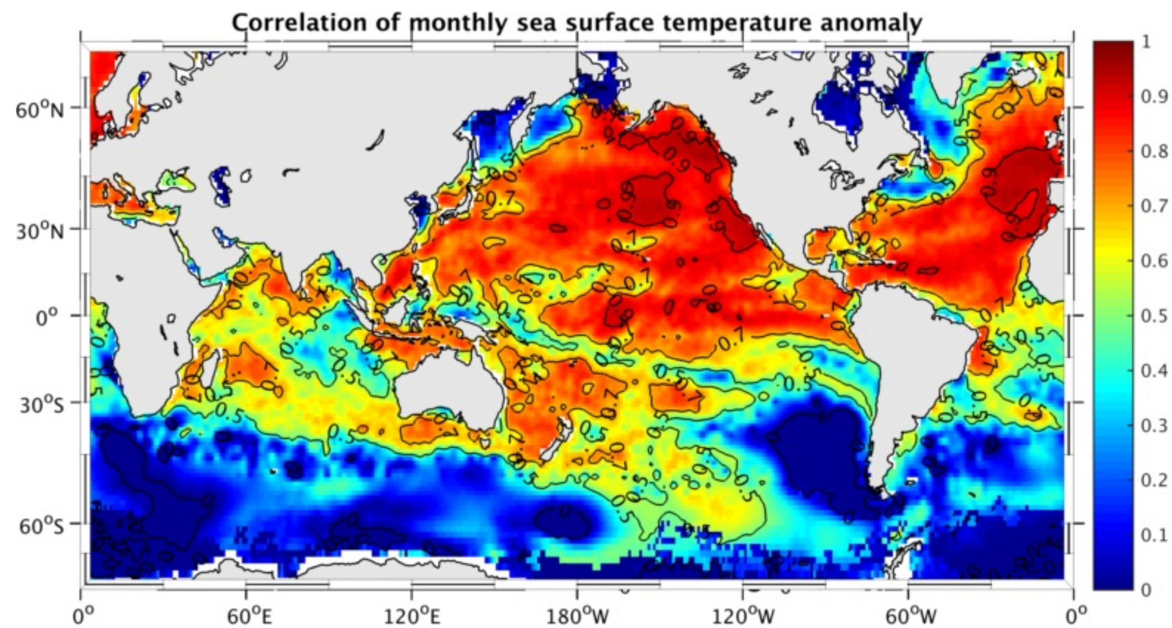
Comparisons to HADISST and HADSLP.

Correlation high where observations existed.

DART did not assimilate SST products or observations.

Produces competitive reanalysis.





DAReS DA Science Research

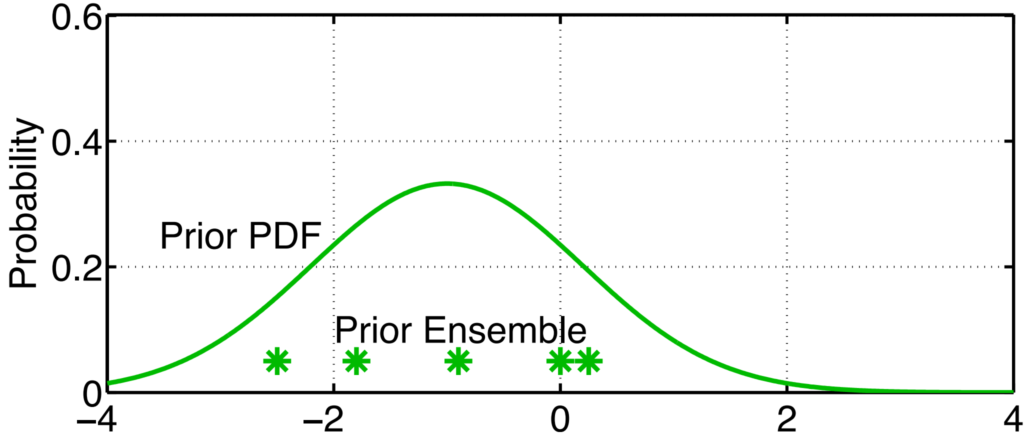
DAReS staff develop improved algorithms that can be added to DART.

DAReS world leader in complex algorithms for:

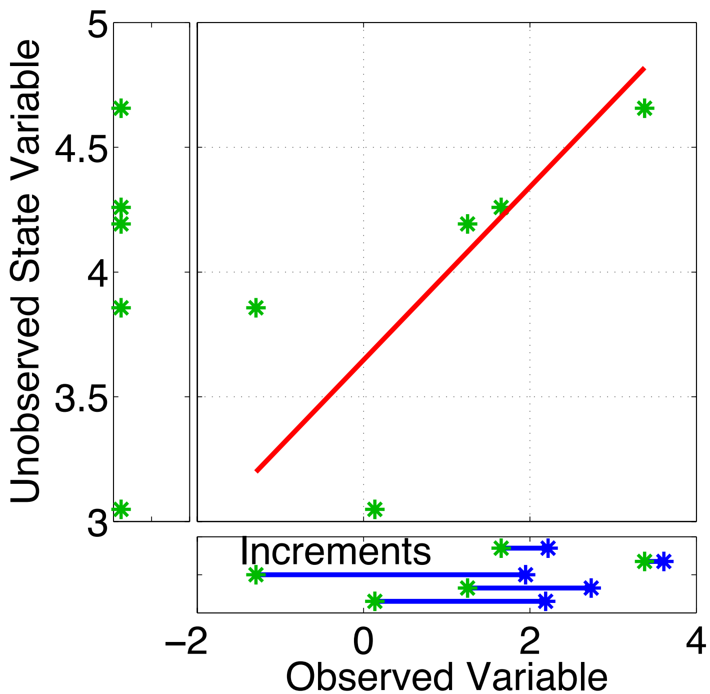
- Adaptive inflation,
- Localization,
- Sampling error reduction.

All of these are available and nearly universally used in DART.

Most DA algorithms are Gaussian and linear.



Fit a Gaussian to ensemble members.



Do a least squares fit to ensemble.

DAReS DA Science Research

All earth system applications are Non-Gaussian and Nonlinear

Especially true for bounded quantities:

Streamflow
Relative humidity
Emission of carbon dioxide

Two Non-Gaussian, Nonlinear algorithms being tested in DART

1. Anamorphosis, Moha

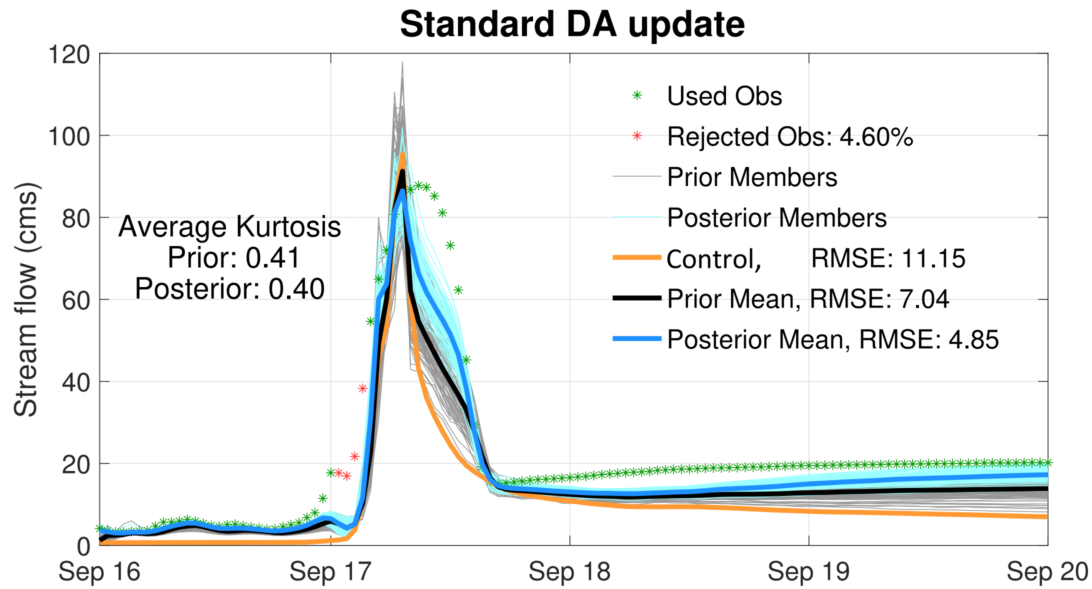
2. Marginal Correction Rank Histogram Filter, Jeff A.

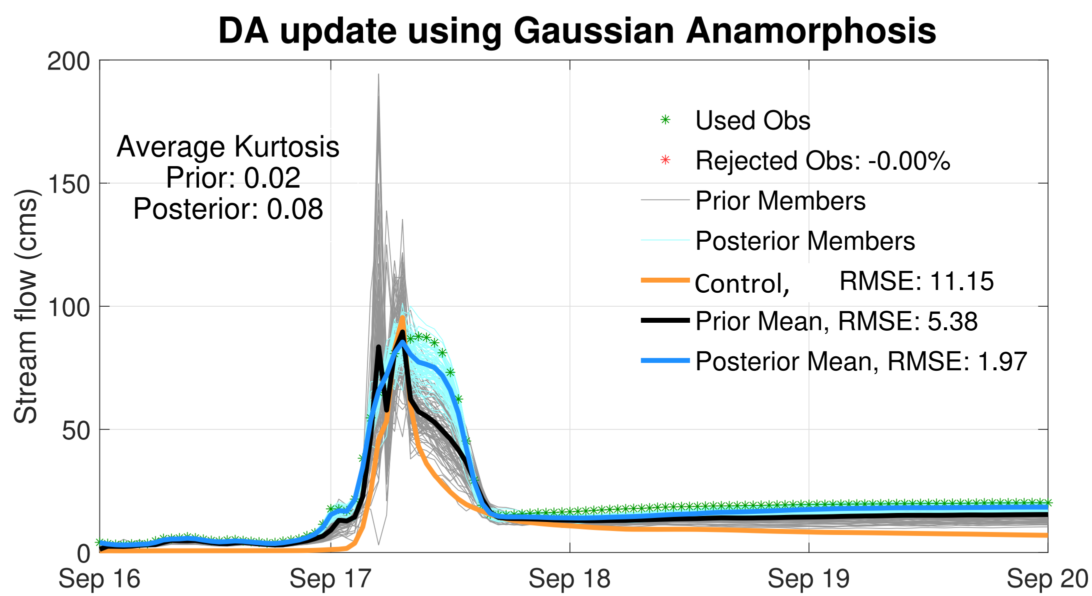
Gaussian Anamorphosis

Observation rejection is improved with GA.

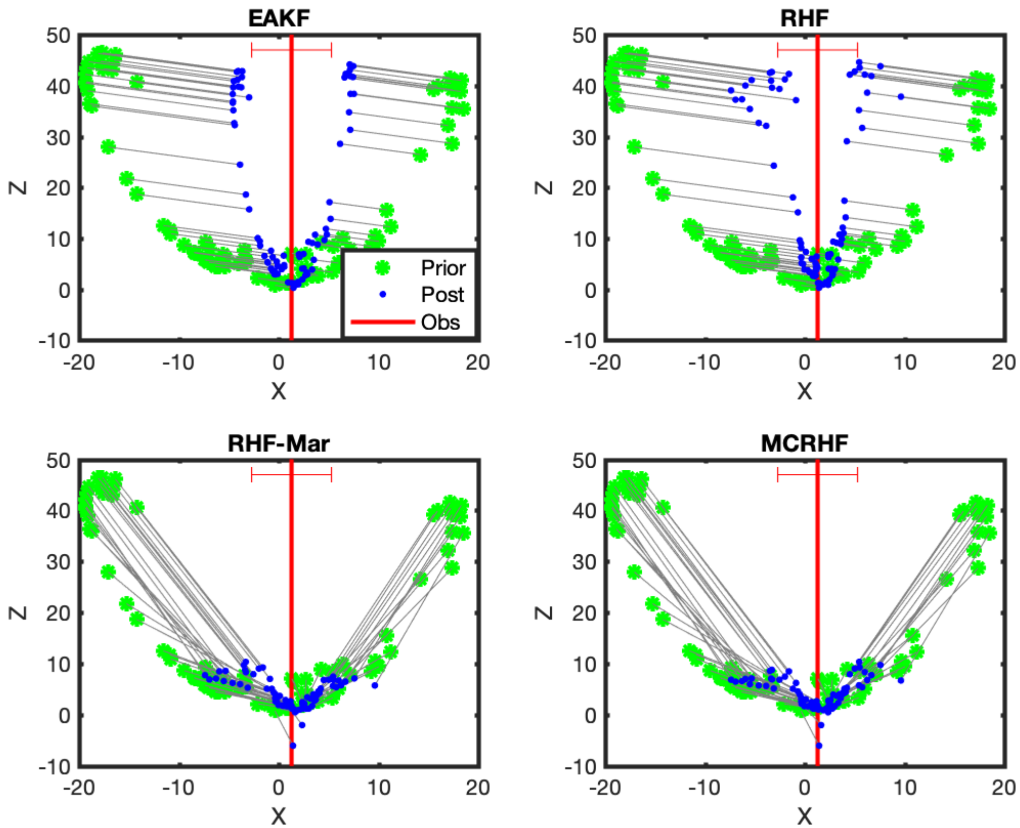
Better fit to the observations on Sep. 17^{th}.

Higher order moments are almost completely eliminated using GA.





Marginal Correction Rank Histogram



Green prior.
Blue Posterior.

Lorenz-63 example: x is observed, z is not observed.
Red indicates observation and error s.d.

Interactions with CISL Colleagues

Brian Dobbins, John Dennis, Mick Coady, D. J. Gagne, many others.

Observations.

Really, really big data.

Large user of super-computing resources, need optimization.

Machine learning.

SIParCS, we love working with these students.

Why are we in CISL? Because this is the best place for us.

Some Exciting Next Steps

- Fully-coupled atmosphere, ocean (land, sea-ice) DA with CGD.
- DA for satellite radiance observations with CMCC/Italy.
- New software for handling observations:
  Better IO,
  Handle correlated observation errors,
  Better parallelism.
- DART for SIMA, System for Integrated Modeling of the Atmosphere.
- DART for MUSICA, new chemistry modeling.
- Prediction with MPAS and regional MPAS (Soyoung Ha, MMM).

Priority: Stay nimble so we can work with more novel science applications.