The keys of ensemble data assimilation for soil-vegetation-atmosphere systems.

Tim Hoar: *National Center for Atmospheric Research*

with a lot of help from:

Jeff Anderson, Nancy Collins, Kevin Raeder: *NCAR*

Yongfei Zhang: *University of Texas Austin*

Andrew Fox: *National Ecological Observatory Network (NEON)*

Rafael Rosolem: *University of Arizona*

NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

NSF

Motivation

1. The ecological state of the planet is the result of an unknowable disturbance history.
2. Model spinup cannot be counted on to accurately re-create that disturbance history.
3. Data assimilation can put the model state more in line with the current state. With that, we can:

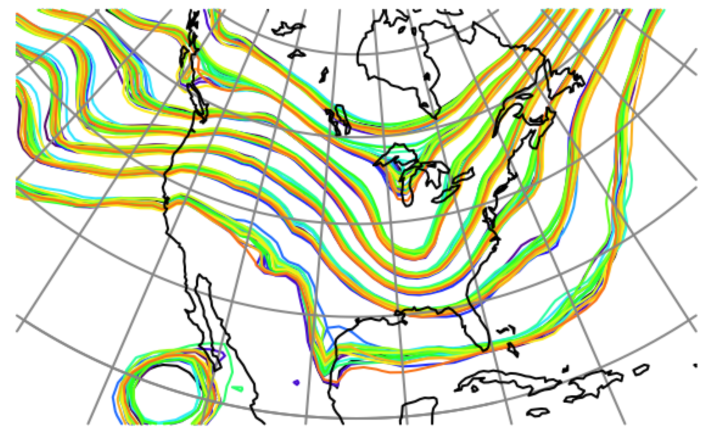
- Quantify ecological states
  - to establish a baseline
  - as a preface for ecological forecasting
- Better understand our models
- Improve our understanding of the underlying processes

What is Data Assimilation?

Observations combined with a Model forecast…

+

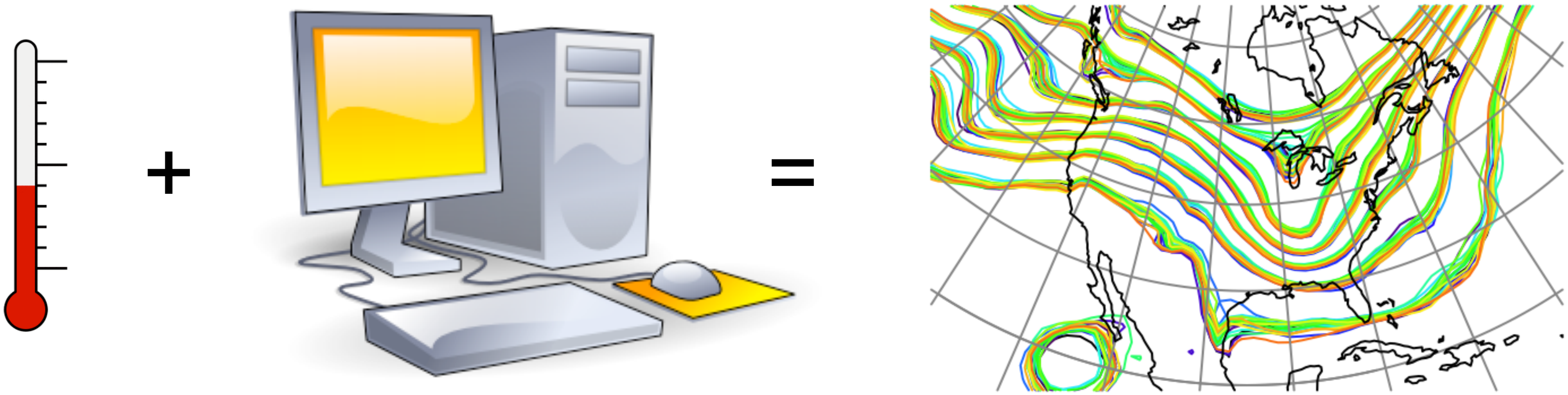
=



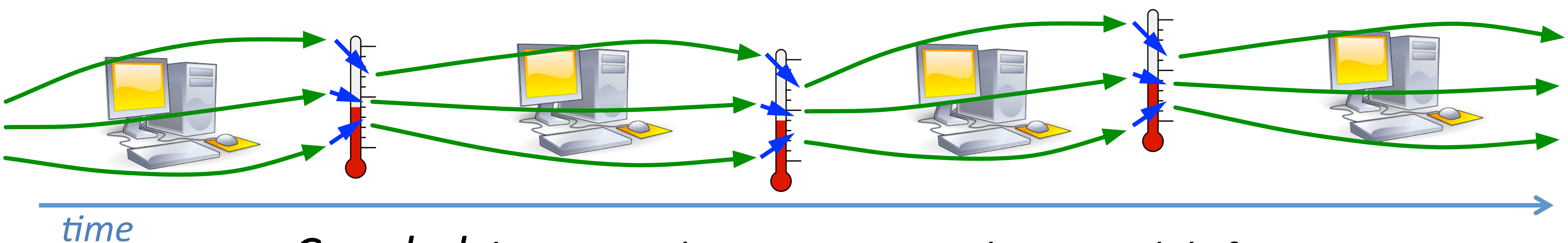
… to produce an analysis.

Overview article of the Data Assimilation Research Testbed (DART):

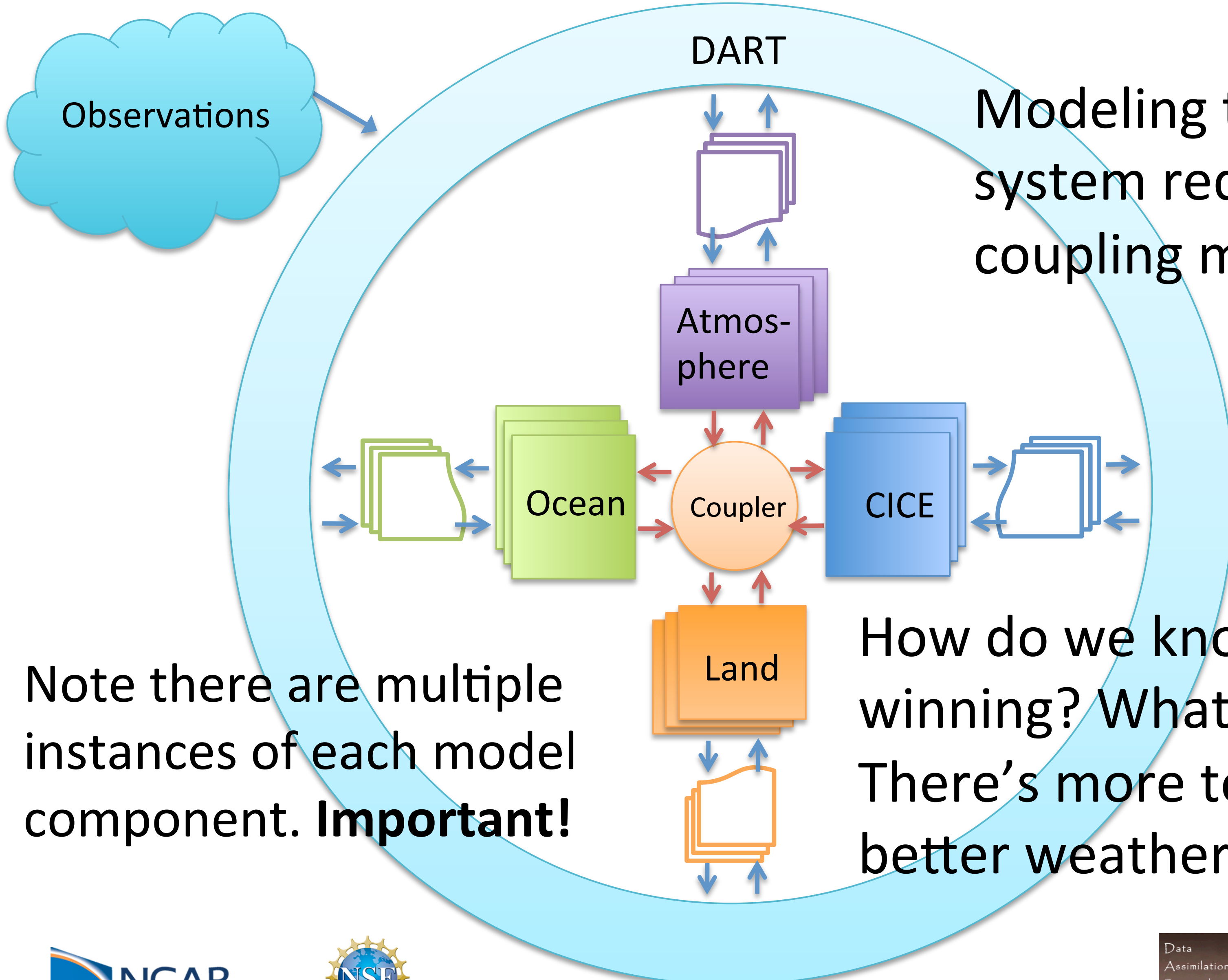
Anderson, Jeffrey, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, A. Arellano, 2009: The Data Assimilation Research Testbed: A Community Facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296. doi:10.1175/2009BAMS2618.1



We want to assimilate over and over to steadily make the model states more consistent with the observations.



Coupled data assimilation means we have models for atmosphere and ocean, or atmosphere and land, or all three, or Soil-Vegetation-Atmosphere, …

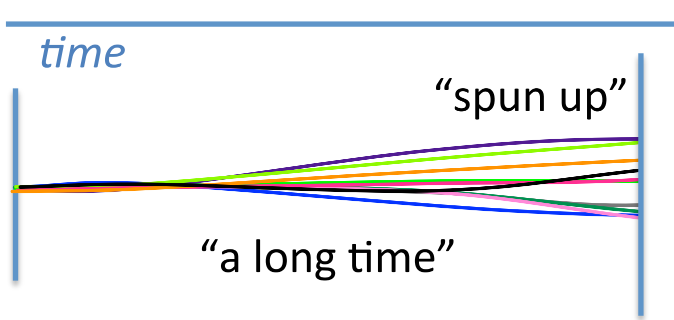


Modeling the Earth system requires coupling models.

Note there are multiple instances of each model component. **Important!**

How do we know if we are winning? What is success? There's more to it than a better weather forecast!

Creating the initial ensemble of ...



Getting a proper initial ensemble is an area of active research.

1. Replicate an equilibrated state N times.
2. Use a unique (and different!) *realistic* forcing for each to induce separate model trajectories.
3. Run them forward for “a long time”.

DART has tools we are using to explore how much spread we NEED to capture the uncertainty in the system.

Creating the initial ensemble of ...

Getting a proper initial ensemble is an area of active research.

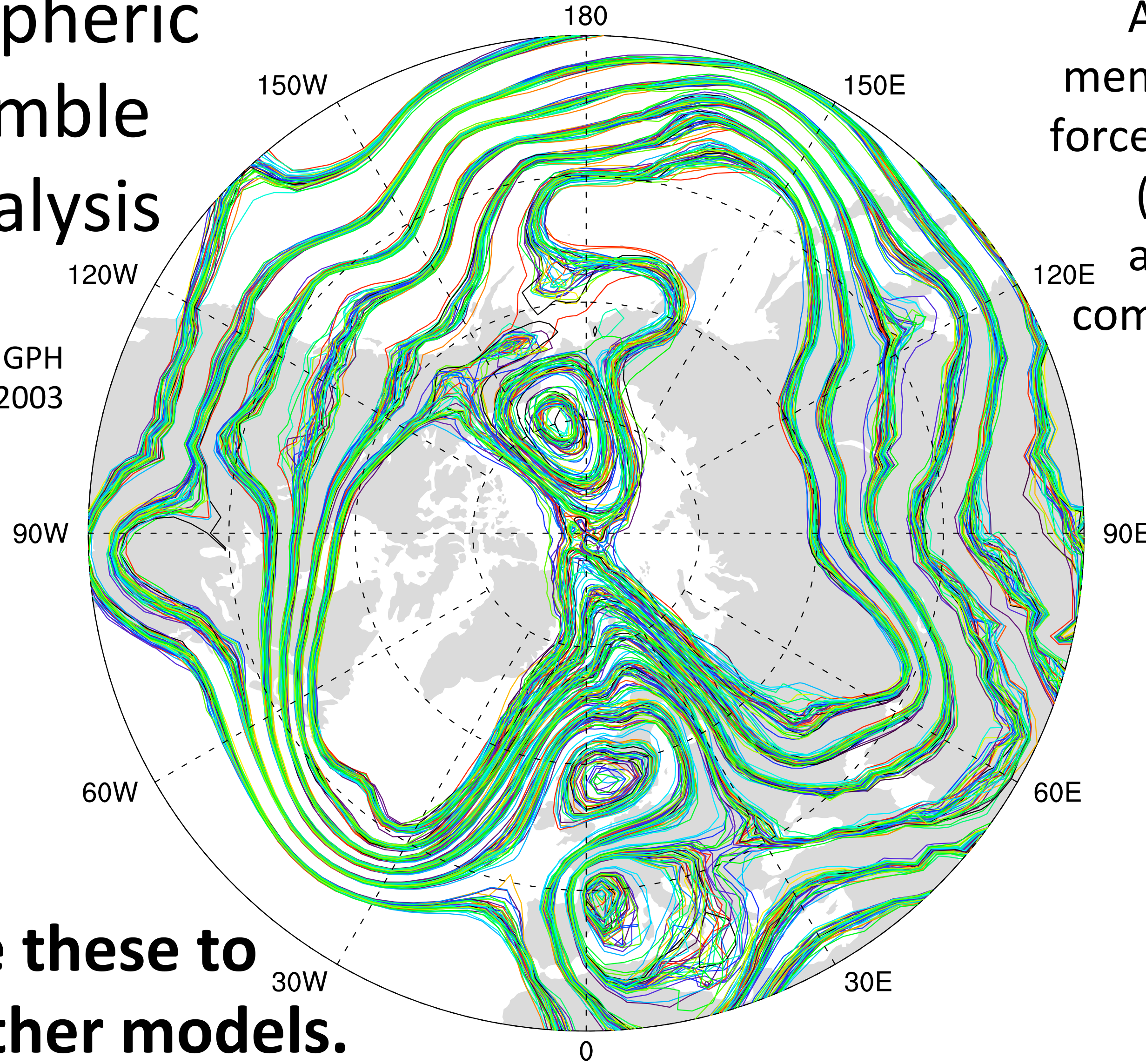
Does This Work for our Land Models?

1. Replicate an equilibrated state N times.
2. Use a unique (and different) *realistic* forcing for each to induce separate model trajectories.
3. Run them forward for "a long time".

DART has tools we are using to explore how much spread we NEED to capture the uncertainty in the system.

Atmospheric Ensemble Reanalysis

500 hPa GPH
Feb 17 2003



Assimilation uses 80 members of 2° FV CAM forced by a single ocean (Hadley+ NCEP-OI2) and produces a very competitive reanalysis.

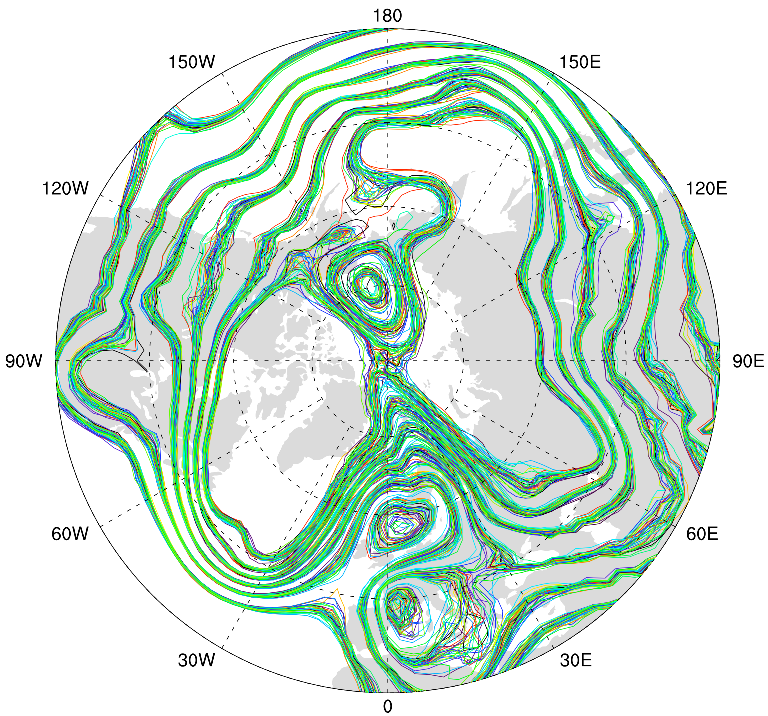
O(1 million) atmospheric obs are assimilated every day.

1998-2010+
4x daily is available.

Can use these to force other models.

NCAR

Data Assimilation Research Testbed



Pros and Cons

- **80 realizations/members**
- **Model states are self-consistent**
- **Model states consistent with obs**
- **Available every 6 hours for 12+ years**

- Relatively low spatial resolution has implications for regional applications.
- Suboptimal precipitation characteristics.
- Available every 6 hours
  - higher frequency available if needed.
- Only have 12 years … enough?

I'm not going to prove it here, but I believe having an **ensemble** of forcing data is **crucial** to land data assimilation.

Creating the initial ensemble of ...

Getting a proper initial ensemble is an area of active research.

1. Replicate what we have N times.
2. Use a unique (and different!) *realistic* forcing for each.
3. Run them forward for “a long time”.

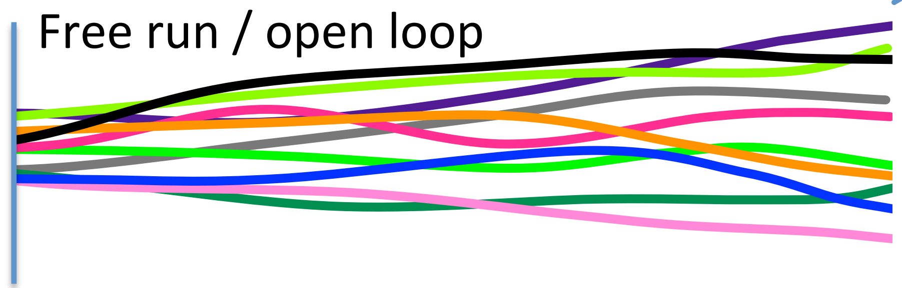
We have tools we are using to explore how much spread we NEED to capture the uncertainty in the system.

The ensemble advantage.

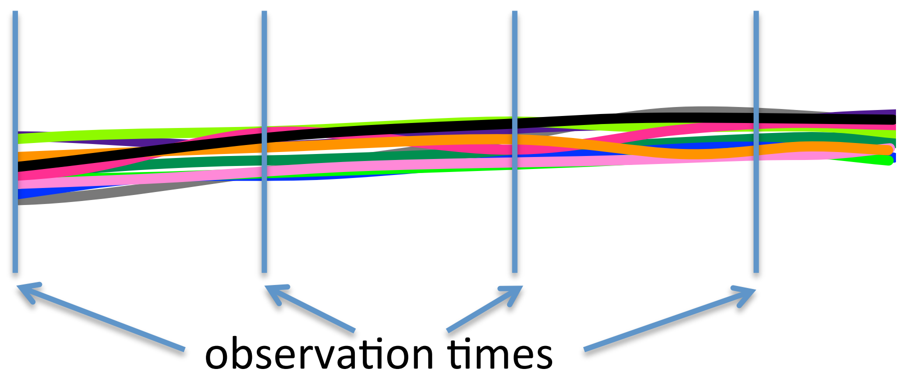
You can represent uncertainty.

time

The ensemble spread frequently grows in a free run of a dispersive model.

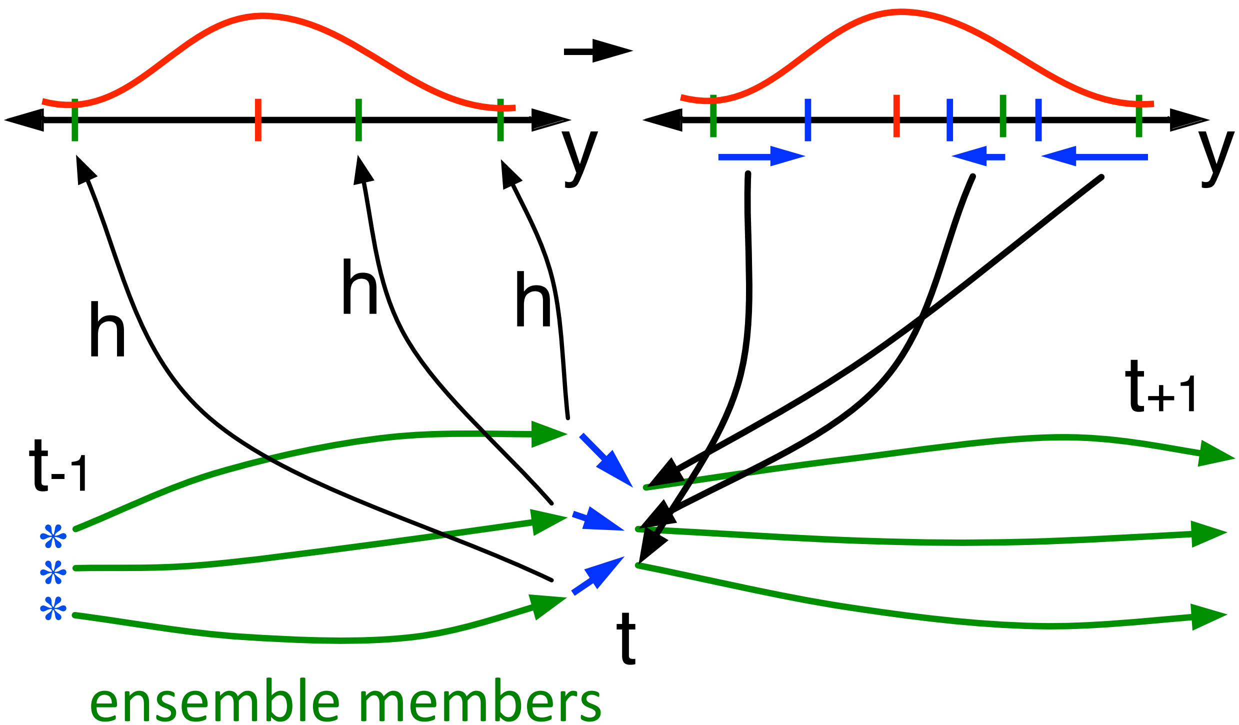


A good assimilation reduces the ensemble spread and is still representative and informative.



A generic ensemble filter system like DART needs:

1. A way to make model forecasts.
2. A way to estimate what the observation would be – given the model state. This is the forward observation operator – h.

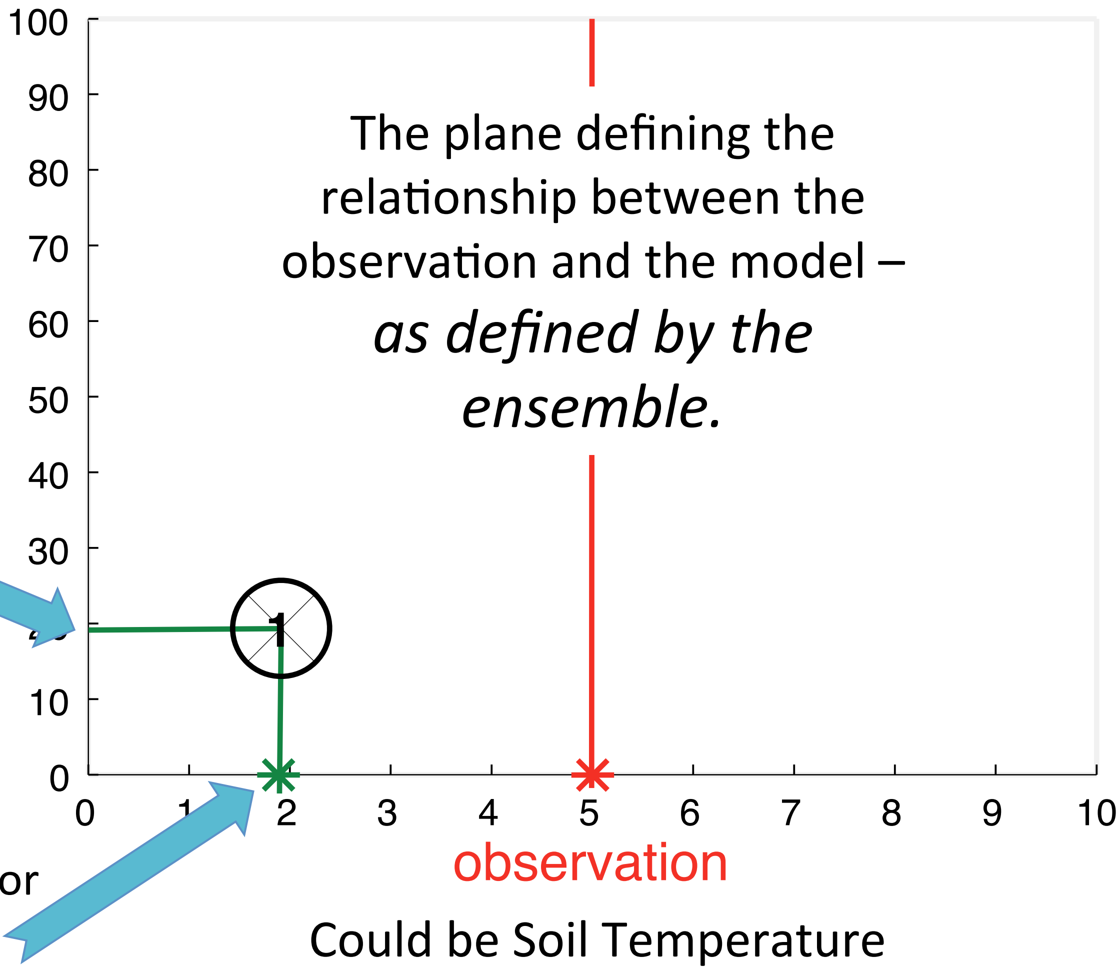


The **increments** are regressed onto as many **state variables** as you like. If there is a correlation, the state gets adjusted in the restart file.

Looking at it another way:

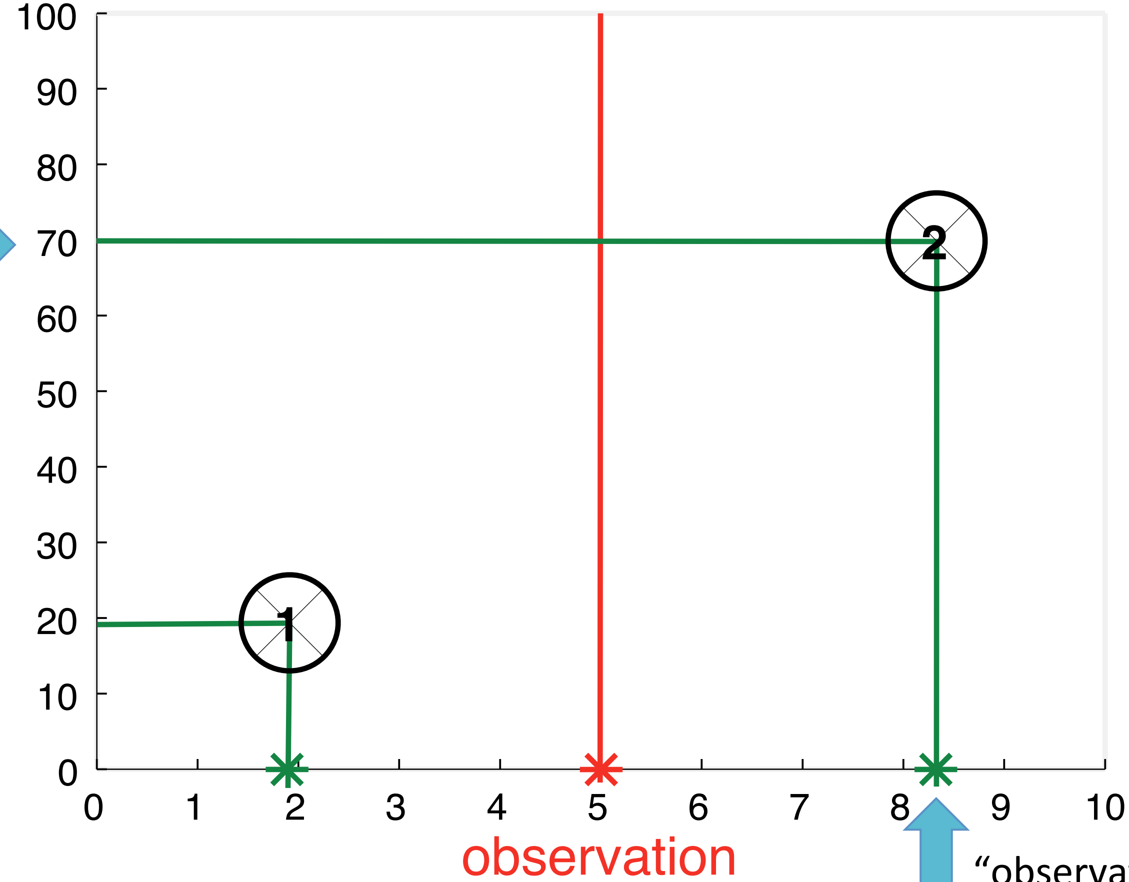
Some unobserved state variable. e.g. live root carbon, dead root carbon, canopy water …

Directly from ensemble member 1



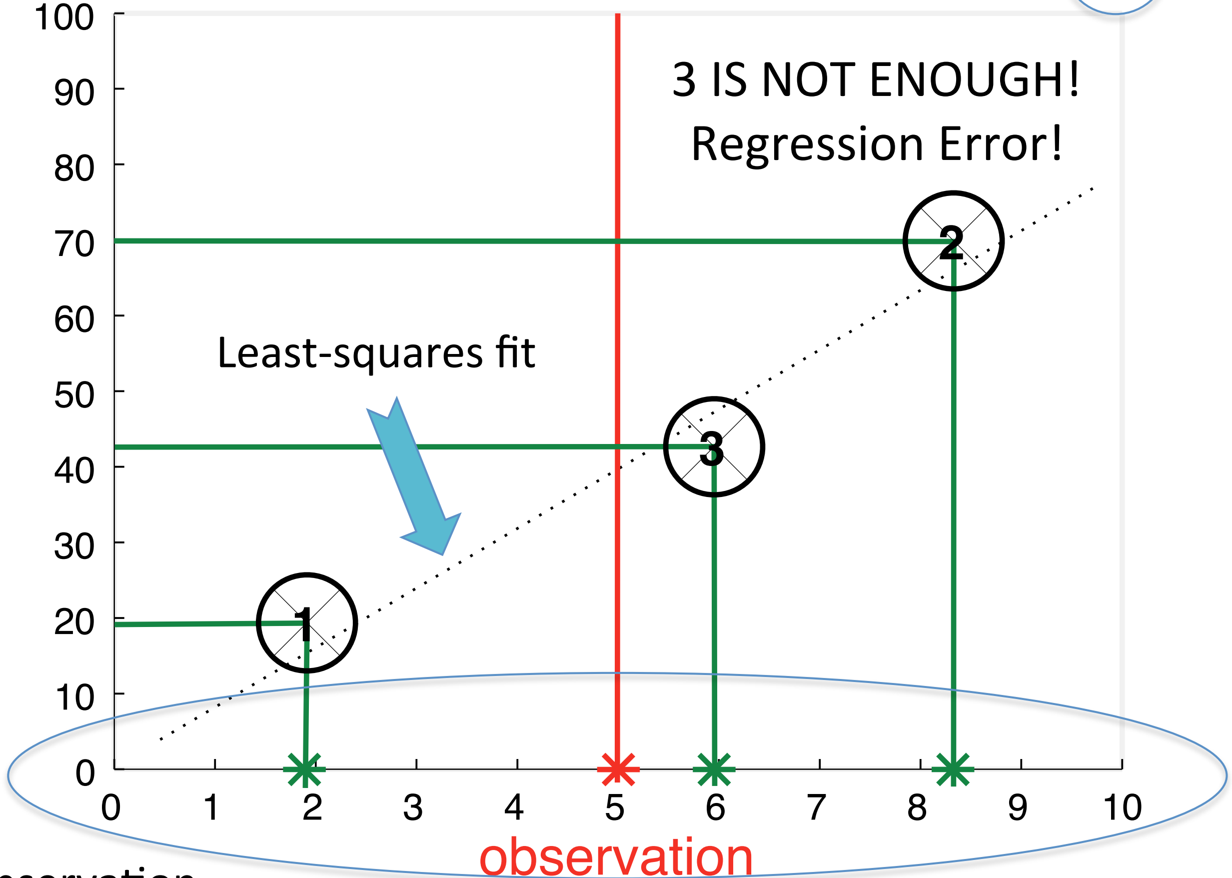
Result of the forward observation operator for ensemble member 1

Directly from ensemble member 2

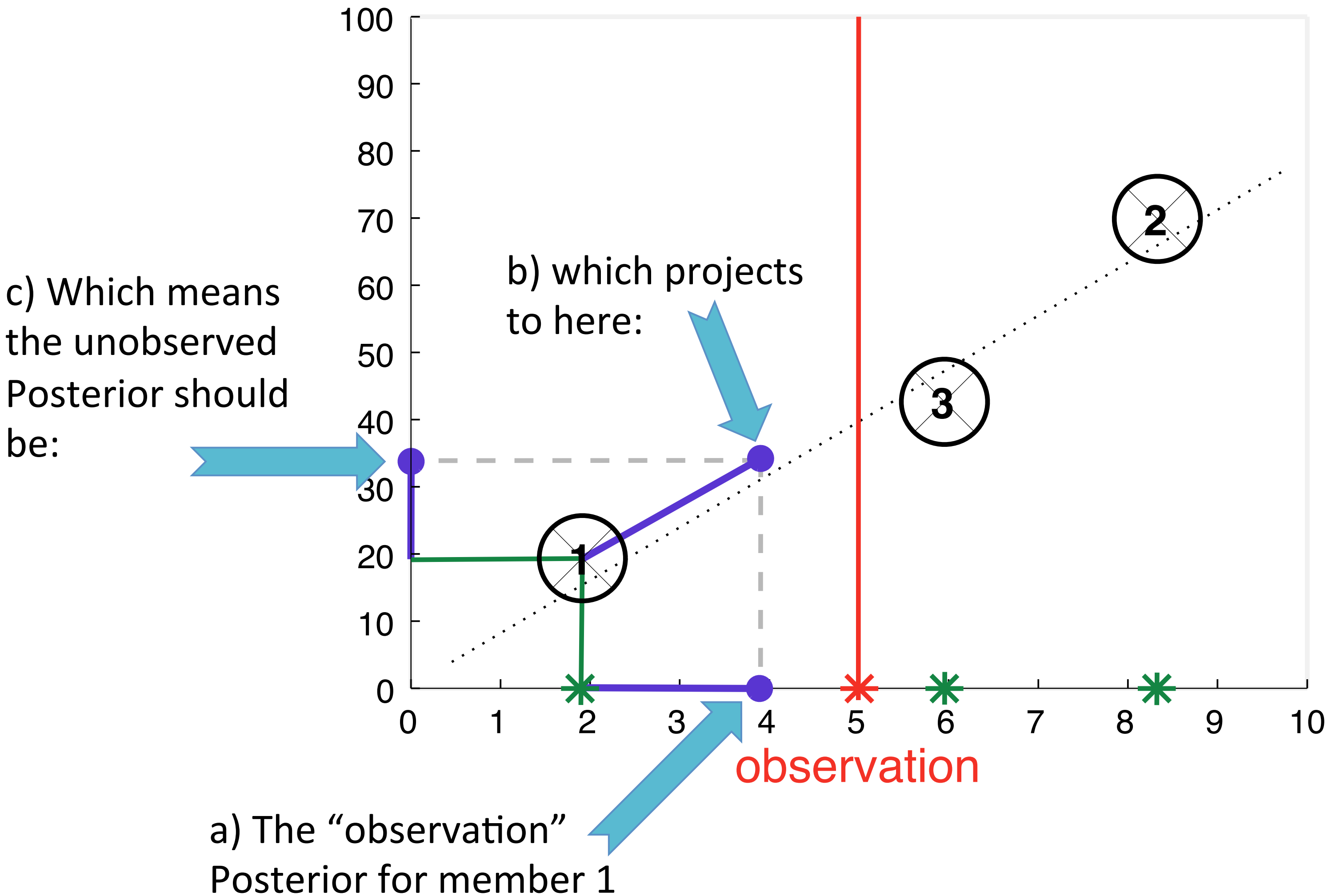


“observation” from ensemble member 2

In our global atmospheric assimilations, we use 80.

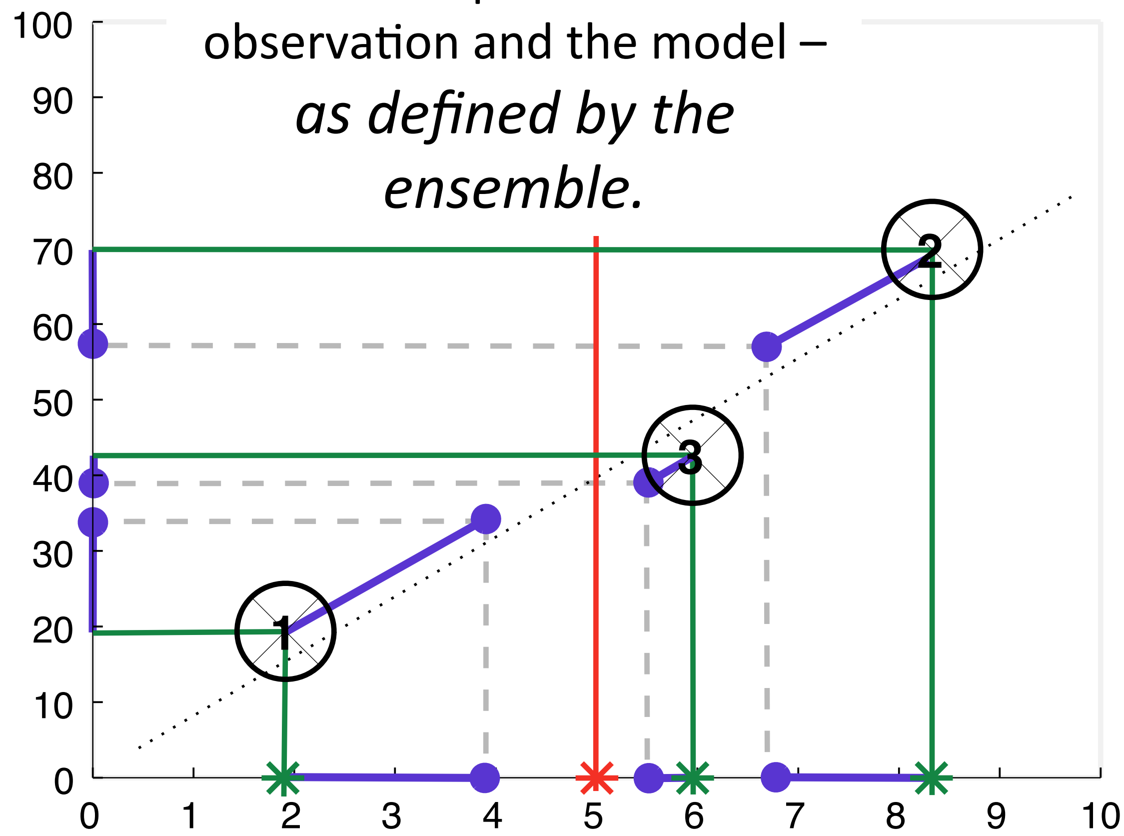


Now, we can calculate out observation increments any way we want.

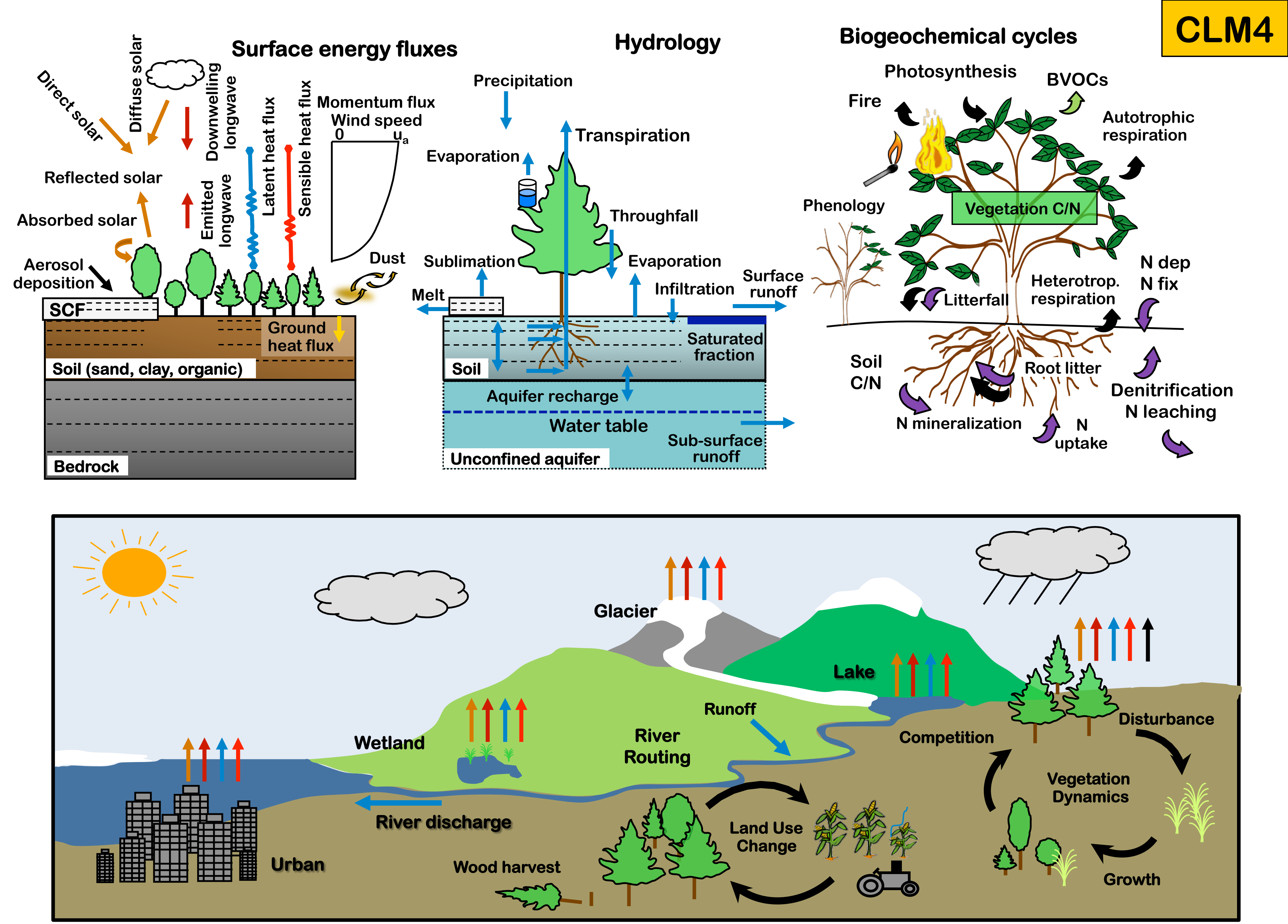

100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
2
3
1
observation
b) which projects
to here:
c) Which means
the unobserved
Posterior should
be:
a) The “observation”
Posterior for member 1

The plane defining the relationship between the observation and the model – *as defined by the ensemble.*

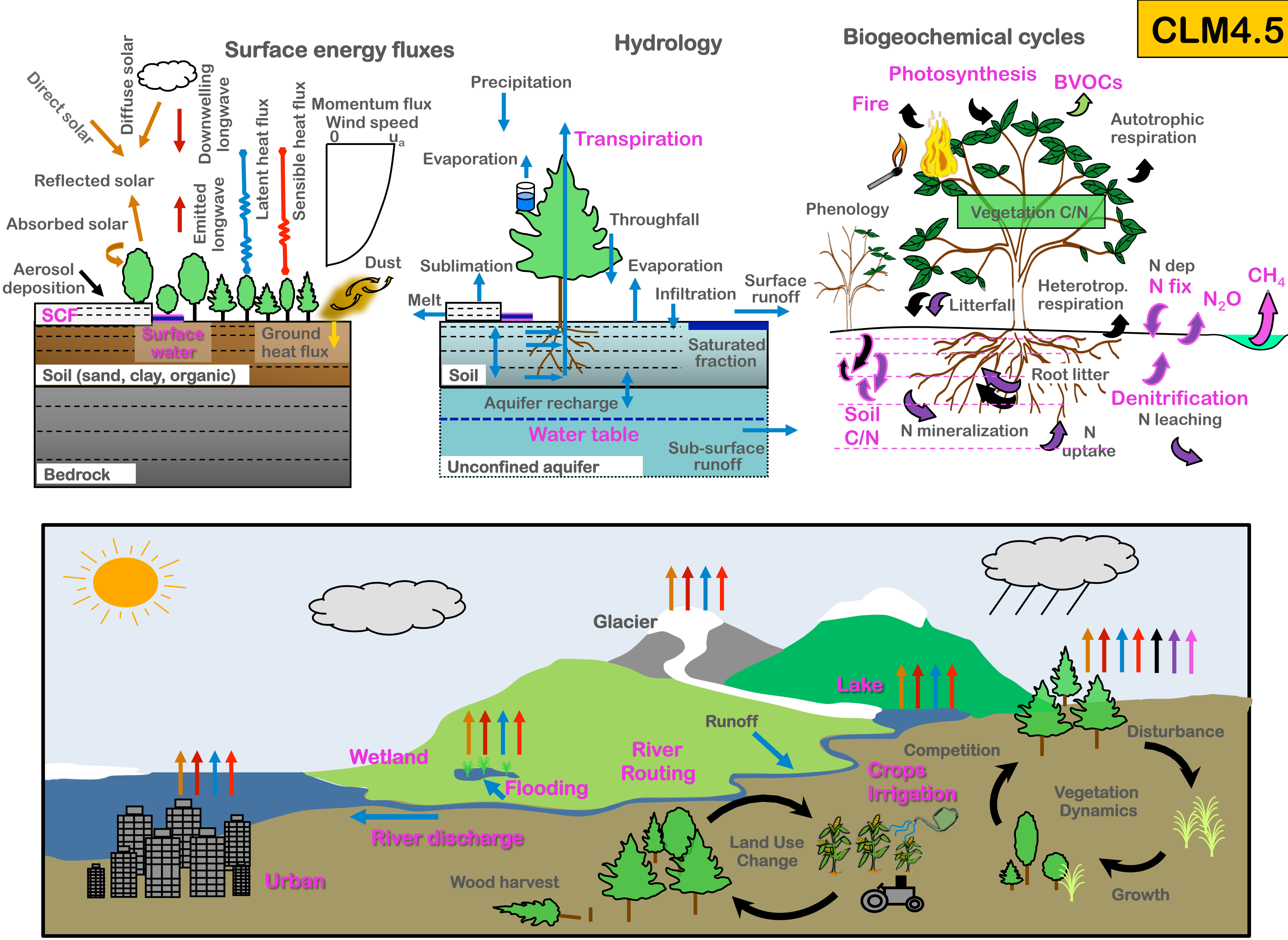
Some unobserved state variable like: live root carbon, dead root carbon, canopy water …



Could be Soil Temperature



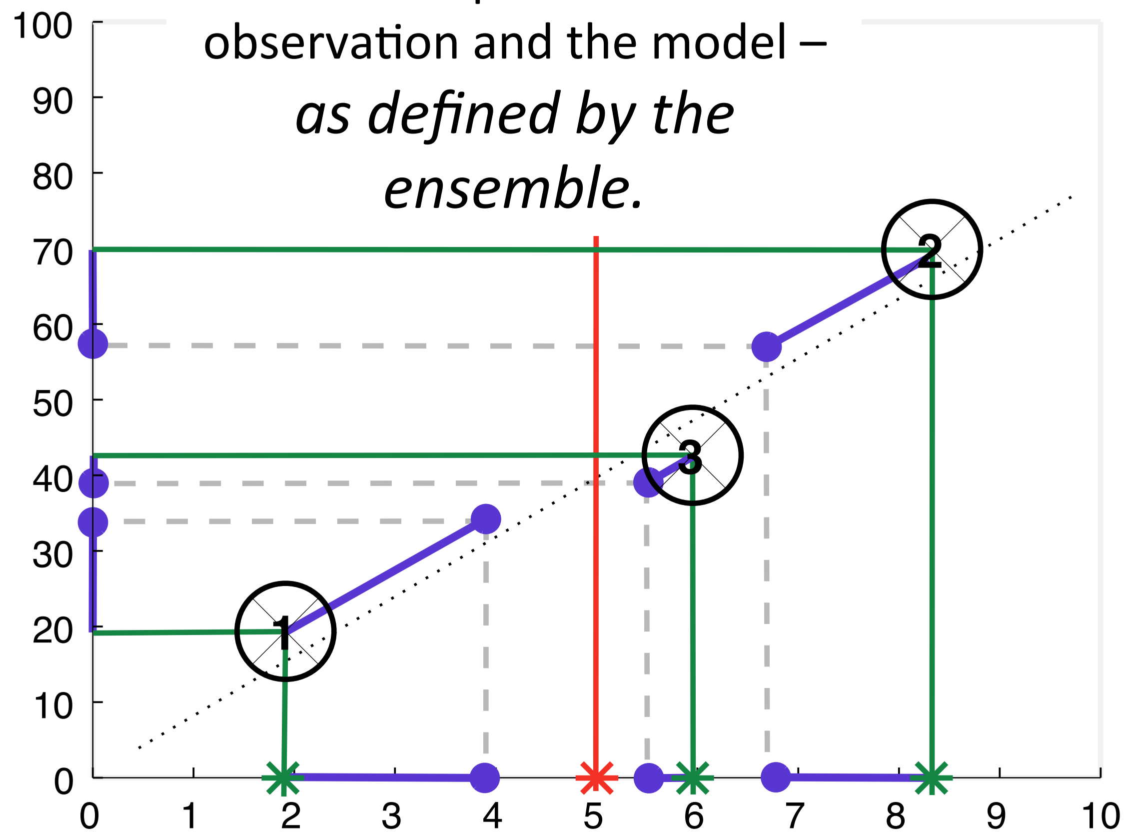
CLM4
Surface energy fluxes
Direct solar
Diffuse solar
Downwelling longwave
Latent heat flux
Sensible heat flux
Momentum flux
Wind speed
0
u_a
Reflected solar
Absorbed solar
Emitted longwave
Aerosol deposition
Dust
SCF
Ground heat flux
Soil (sand, clay, organic)
Bedrock
Hydrology
Precipitation
Transpiration
Evaporation
Throughfall
Sublimation
Evaporation
Melt
Infiltration
Surface runoff
Saturated fraction
Soil
Aquifer recharge
Water table
Sub-surface runoff
Unconfined aquifer
Biogeochemical cycles
Photosynthesis
BVOCs
Fire
Autotrophic respiration
Phenology
Vegetation C/N
N dep
N fix
Litterfall
Heterotrop. respiration
Soil C/N
Root litter
Denitrification
N leaching
N mineralization
N uptake
Glacier
Lake
Wetland
Runoff
River Routing
Competition
Disturbance
River discharge
Vegetation Dynamics
Land Use Change
Urban
Wood harvest
Growth



CLM4.5
Surface energy fluxes
Direct solar
Diffuse solar
Downwelling longwave
Latent heat flux
Sensible heat flux
Momentum flux
Wind speed
0
u_a
Reflected solar
Absorbed solar
Emitted longwave
Aerosol deposition
Dust
SCF
Surface water
Ground heat flux
Soil (sand, clay, organic)
Bedrock
Hydrology
Precipitation
Transpiration
Evaporation
Throughfall
Sublimation
Evaporation
Melt
Infiltration
Surface runoff
Saturated fraction
Soil
Aquifer recharge
Water table
Sub-surface runoff
Unconfined aquifer
Biogeochemical cycles
Photosynthesis
BVOCs
Fire
Autotrophic respiration
Phenology
Vegetation C/N
Litterfall
Heterotrop. respiration
N dep
N fix
N_2O
CH_4
Root litter
Soil C/N
N mineralization
N uptake
Denitrification
N leaching
Glacier
Lake
Wetland
Runoff
River Routing
Flooding
River discharge
Competition
Disturbance
Crops
Irrigation
Vegetation Dynamics
Land Use Change
Urban
Wood harvest
Growth

Some unobserved state variable like: live root carbon, dead root carbon, canopy water …

The plane defining the relationship between the observation and the model – *as defined by the ensemble.*



observation

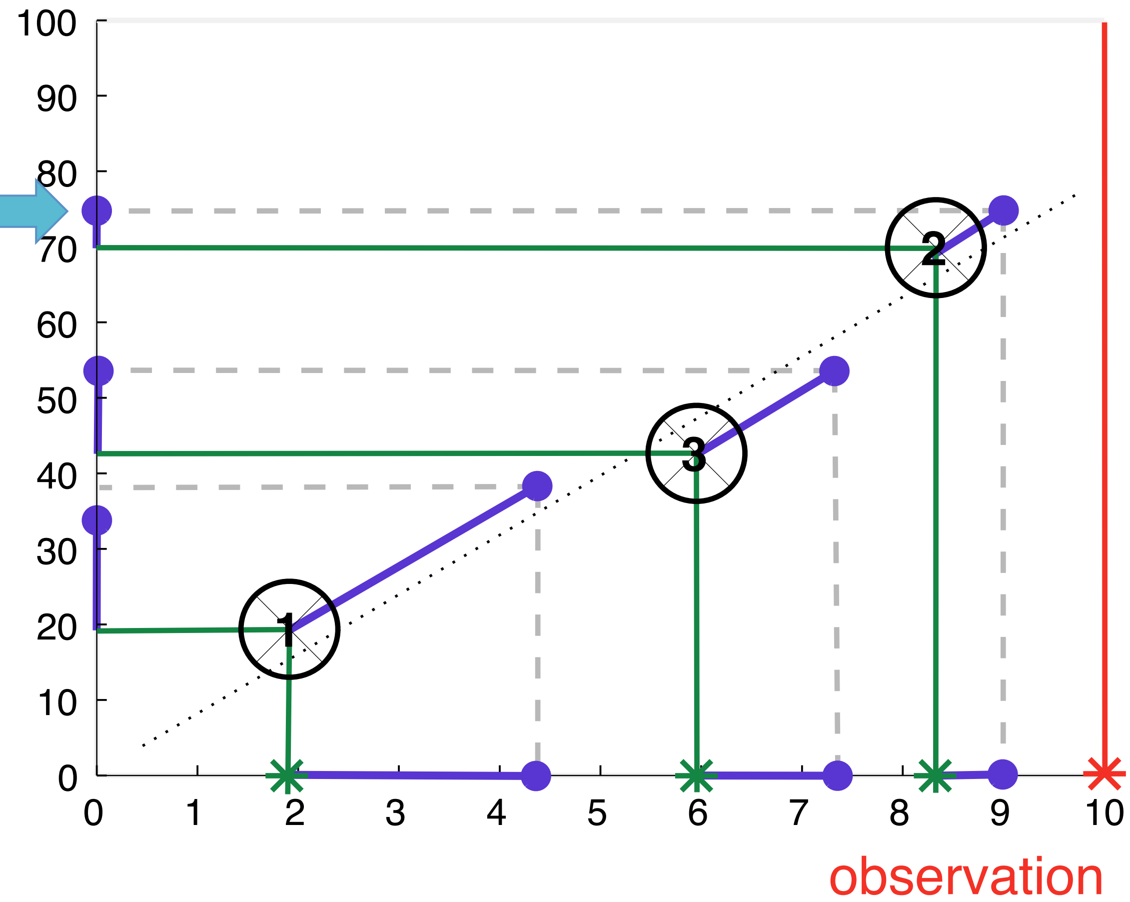
Could be Soil Temperature

REPEATED FOR REFERENCE

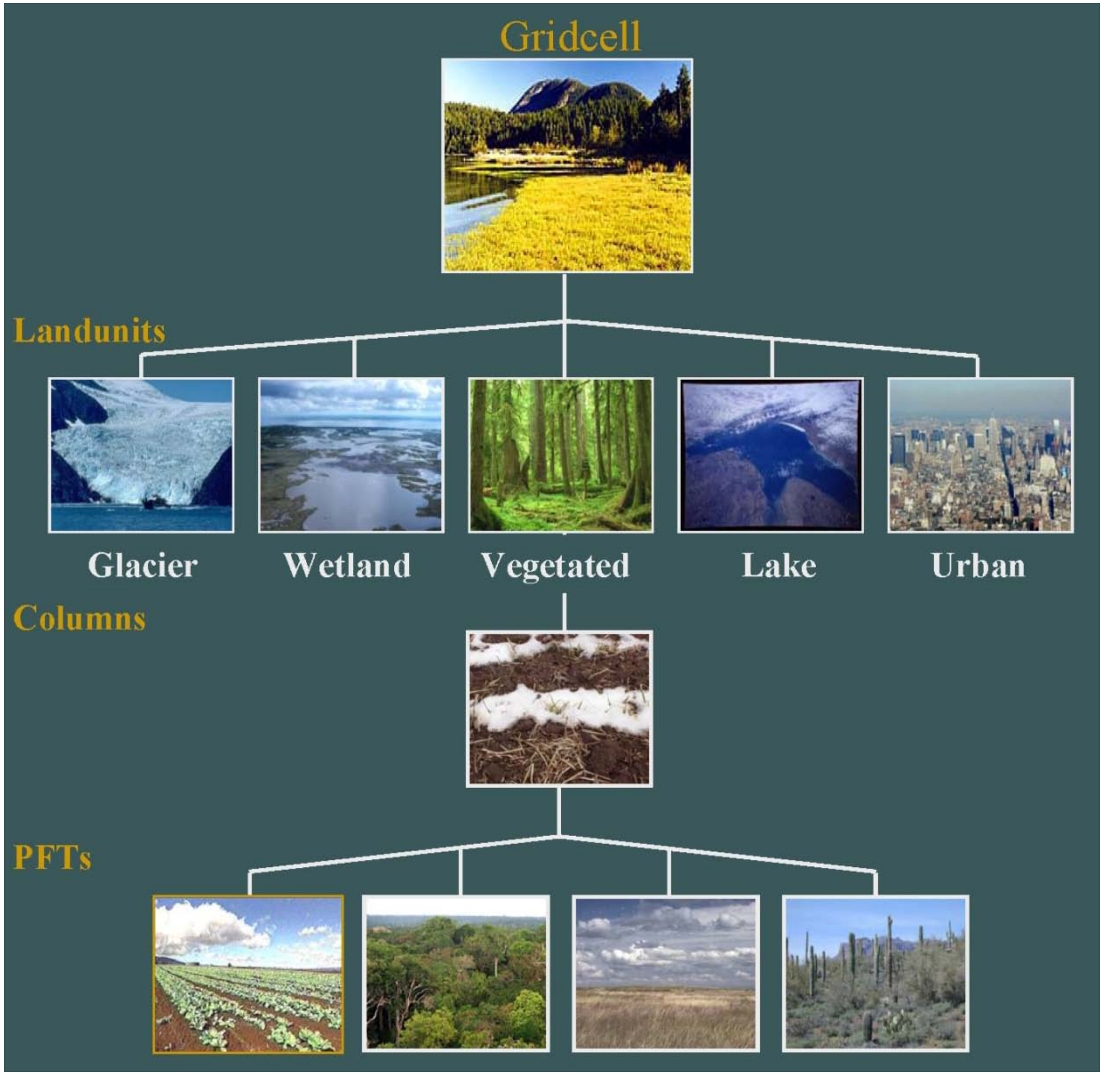
Potential Problem

This posterior MAY or MAY NOT be realistic!

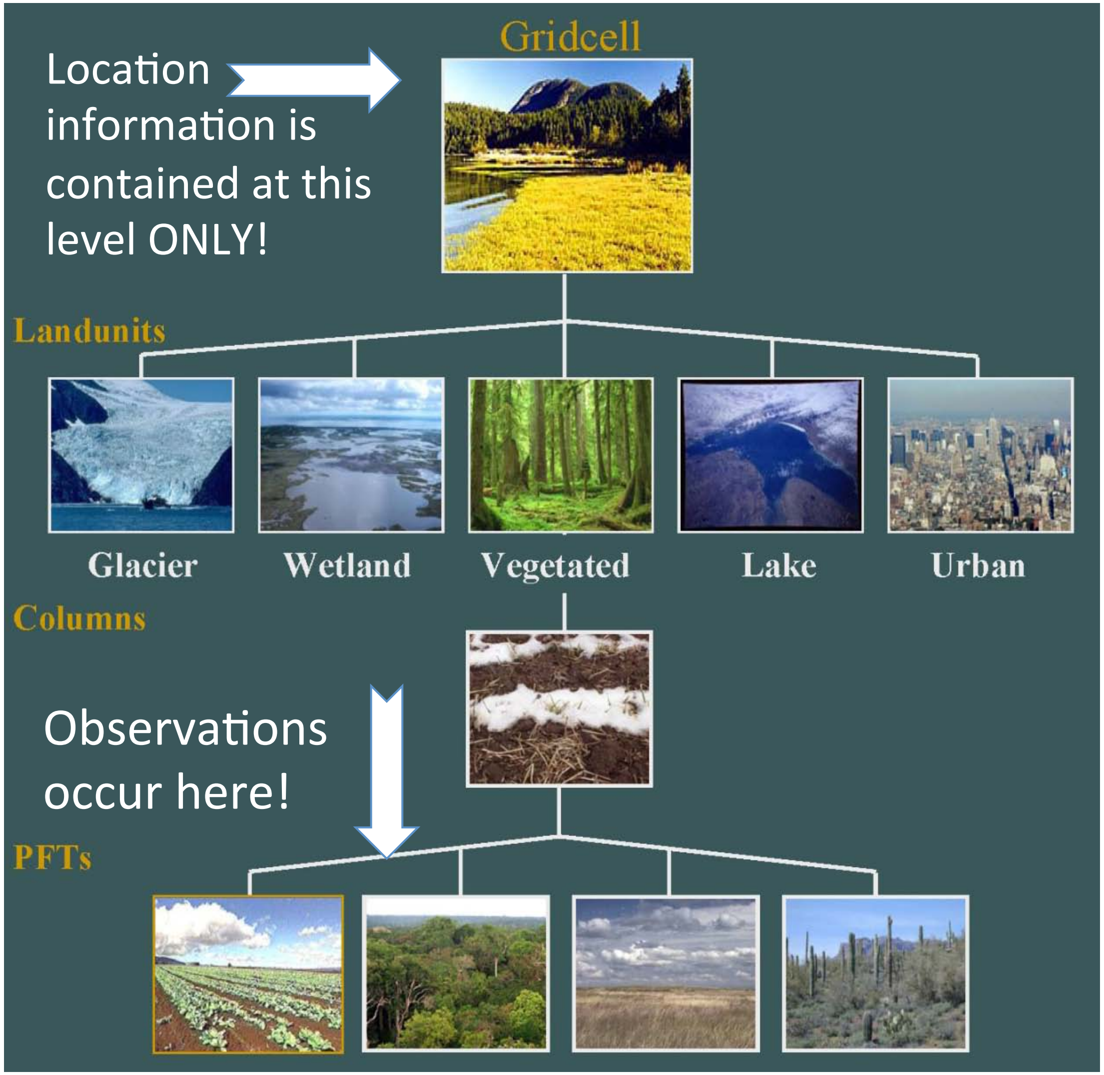
Can the model tolerate this new state?



If the observation is “too far” away, it is rejected.
What is “too far”?



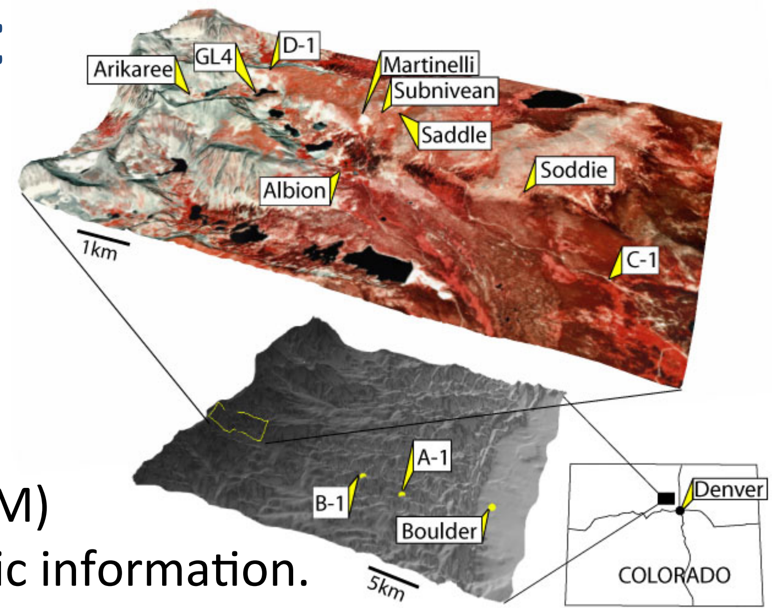
Models that abstract the gridcell into a "nested gridcell hiearchy of of multiple landunits, snow/soil columns, and Plant Function Types" are particularly troublesome when trying to convert the model state to the expected observation value.



Models that abstract the gridcell into a “nested gridcell hiearchy of of multiple landunits, snow/soil columns, and Plant Function Types” are particularly troublesome when trying to convert the model state to the expected observation value.

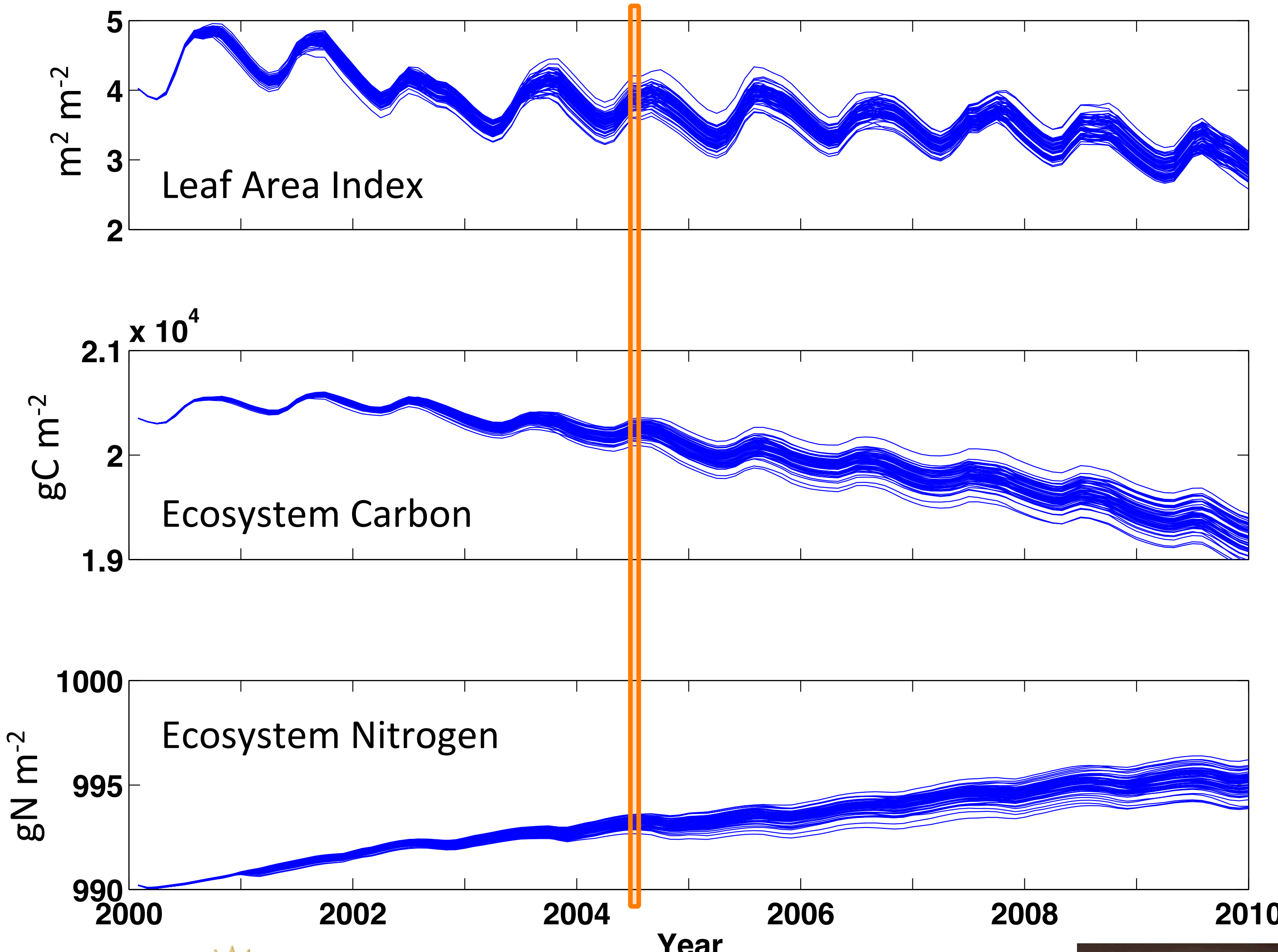
Given a soil temperature observation at a specific lat/lon, which PFT did it come from? **No way to know!** *Unless obs have more metadata!*

In collaboration with Andy Fox (NEON): An experiment at Niwot Ridge

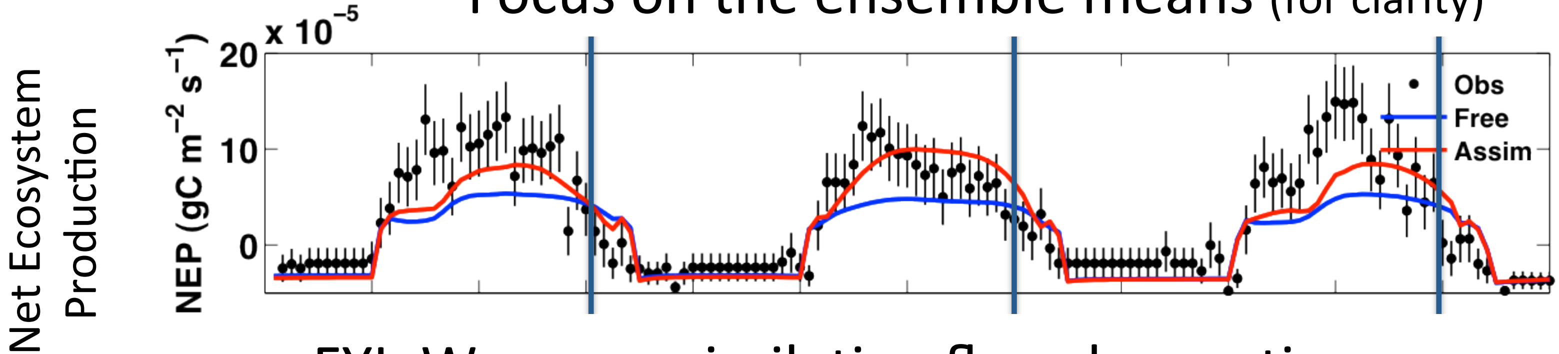


- 9.7 km east of the Continental Divide
- C-1 is located in a Subalpine Forest
- (40º 02' 09'' N; 105º 32' 09'' W; 3021 m)
- One column of Community Land Model (CLM)
  - Spun up for 1500 years with site-specific information.
- 64 ensemble members
- Forcing from the DART/CAM reanalysis,
- Assimilating 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*.

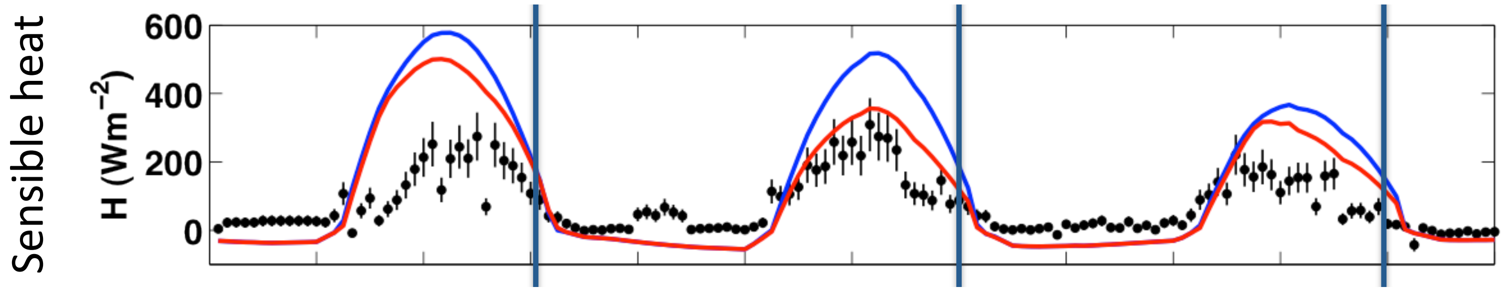
Free Runs of CLM driven by 64 CAM reanalyses



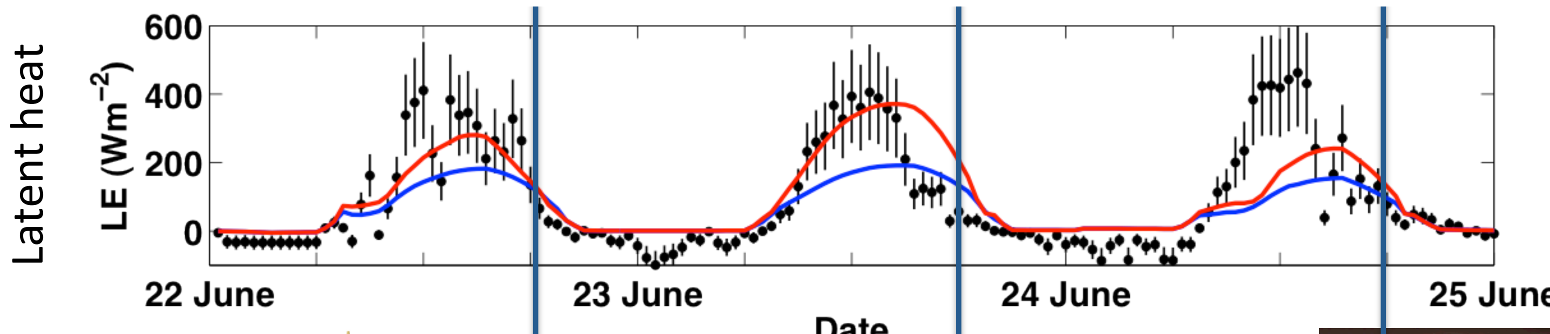
Focus on the ensemble means (for clarity)

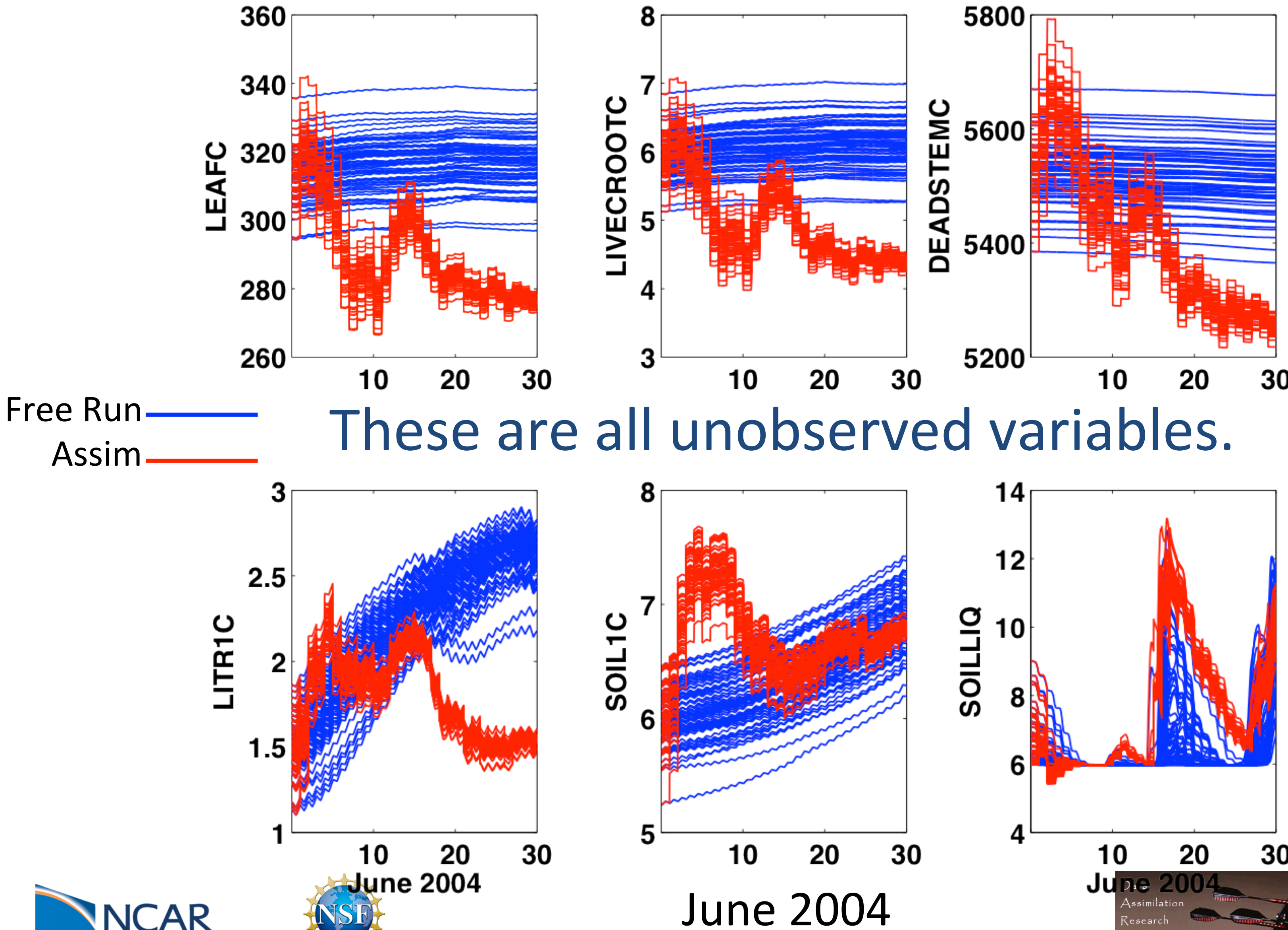


FYI: We are assimilating flux observations …



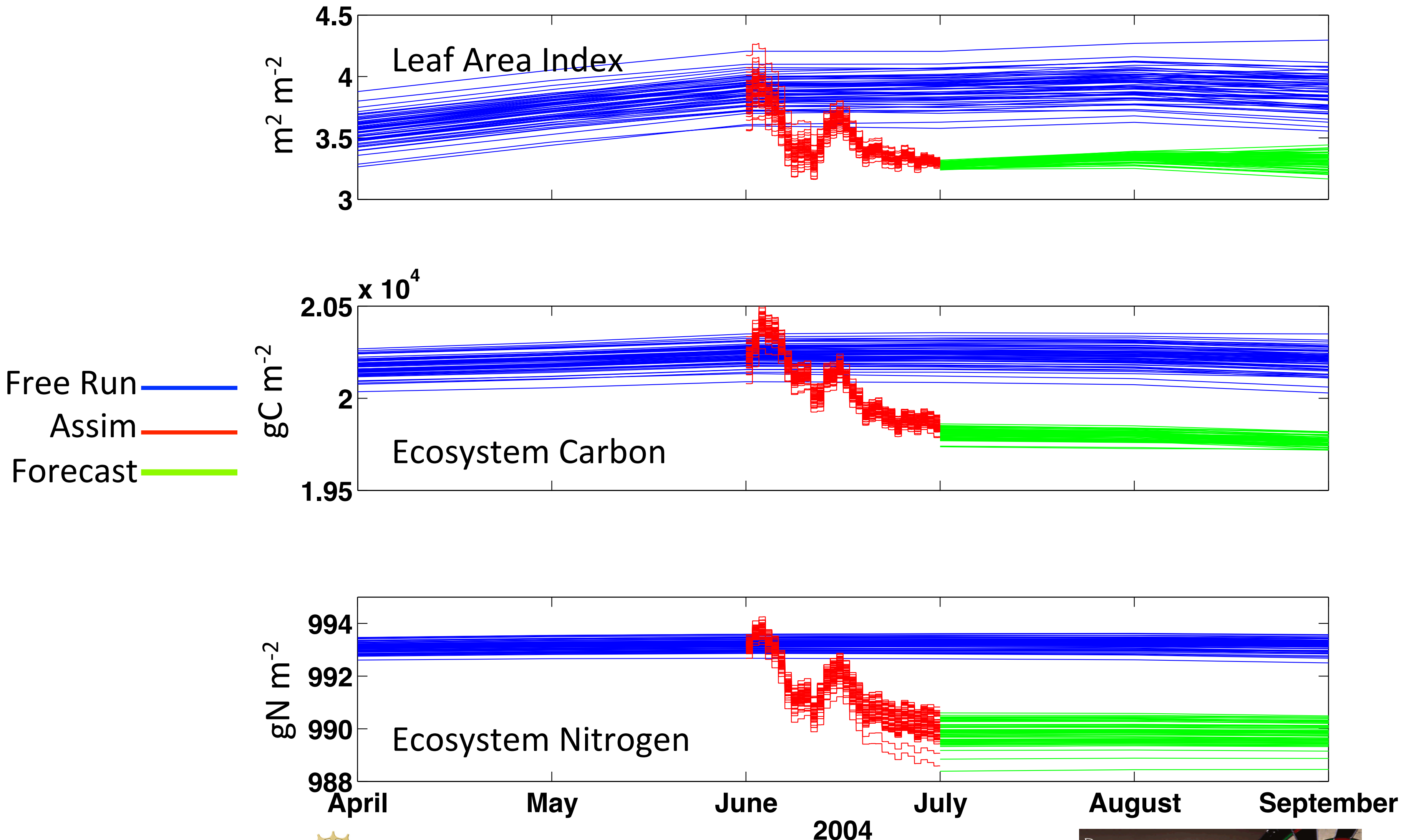
The model states are being updated at about 8PM local time.



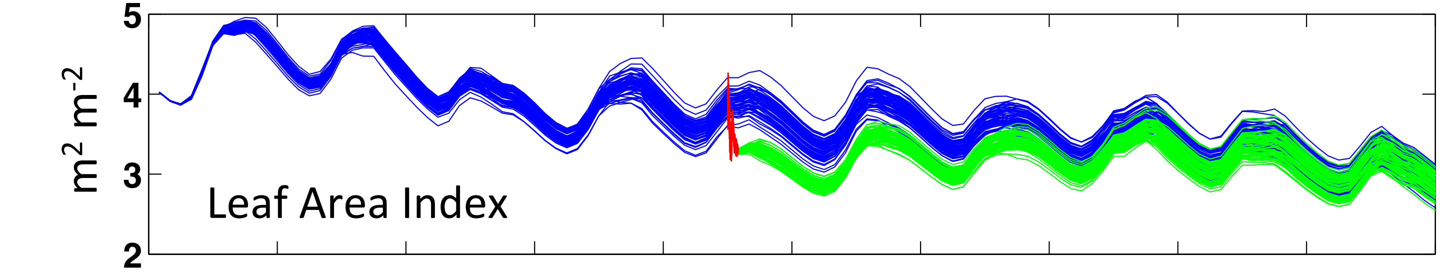


These are all unobserved variables.

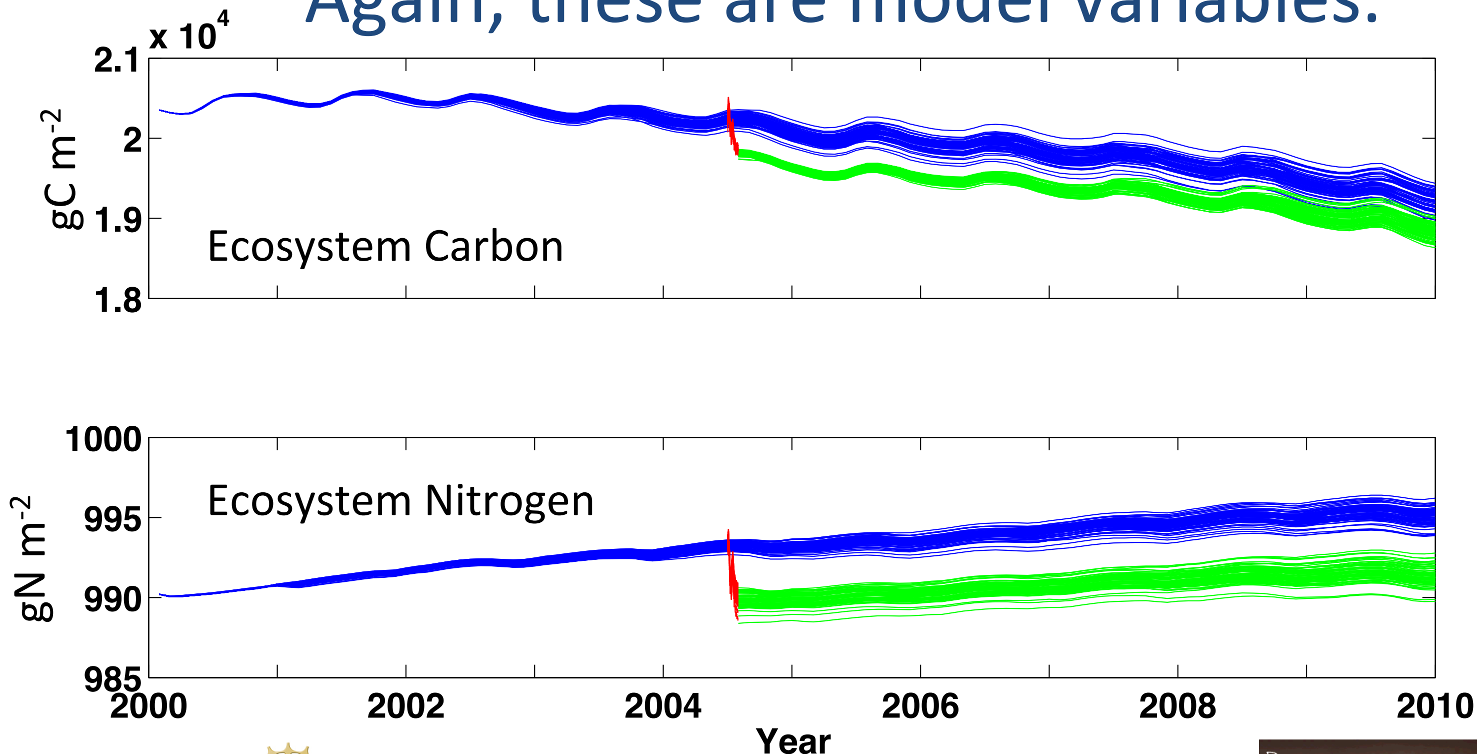
Effect on short-term forecast on unobserved variables.



Effect on longer-term forecast



Again, these are model variables.



Assimilation of the MODIS Snow Cover Fraction Dataset through the Coupled Data Assimilation Research Testbed (DART) and the Community Land Model (CLM4)

Yongfei Zhang, Zong-Liang Yang

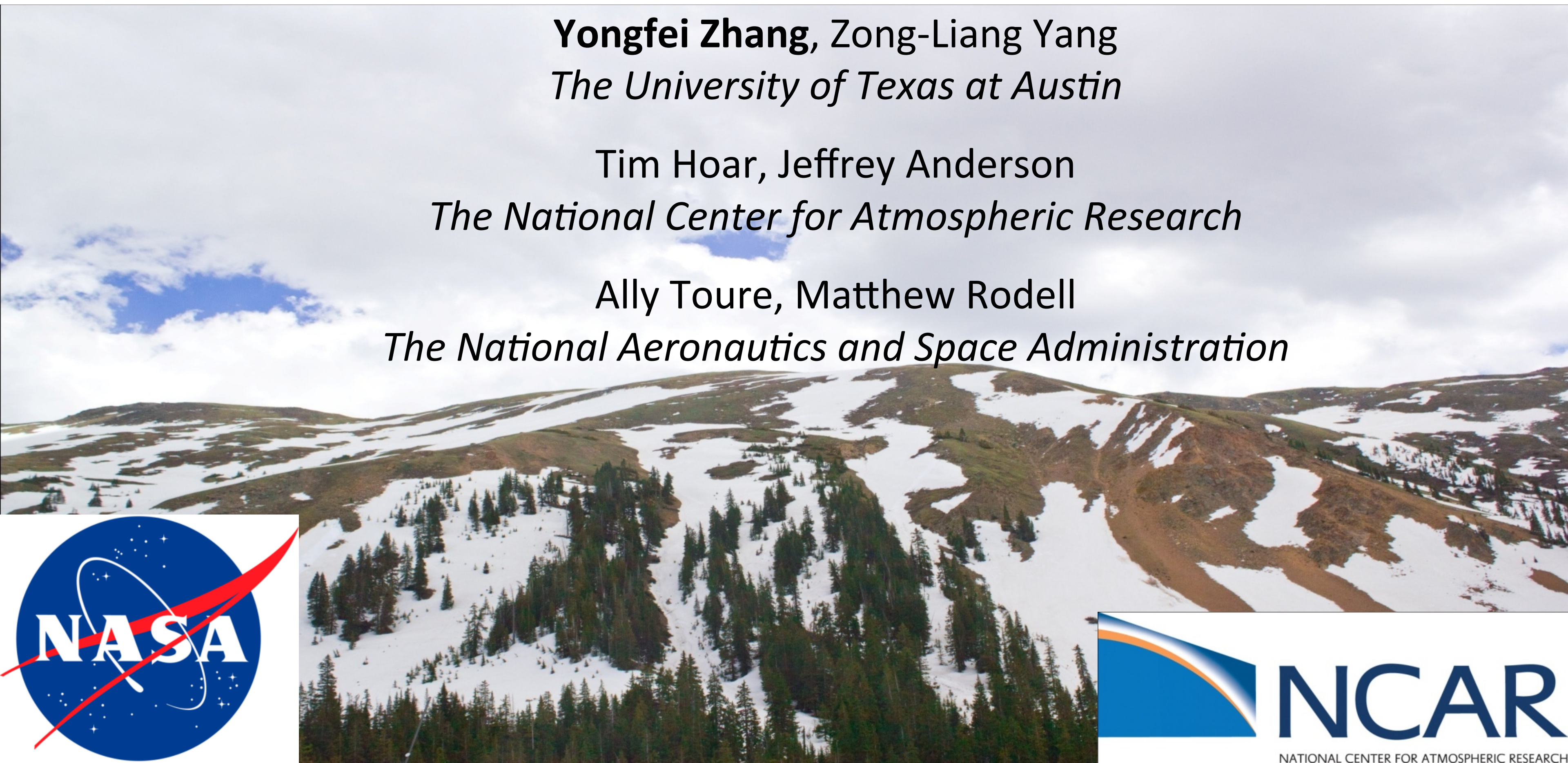
The University of Texas at Austin

Tim Hoar, Jeffrey Anderson

The National Center for Atmospheric Research

Ally Toure, Matthew Rodell

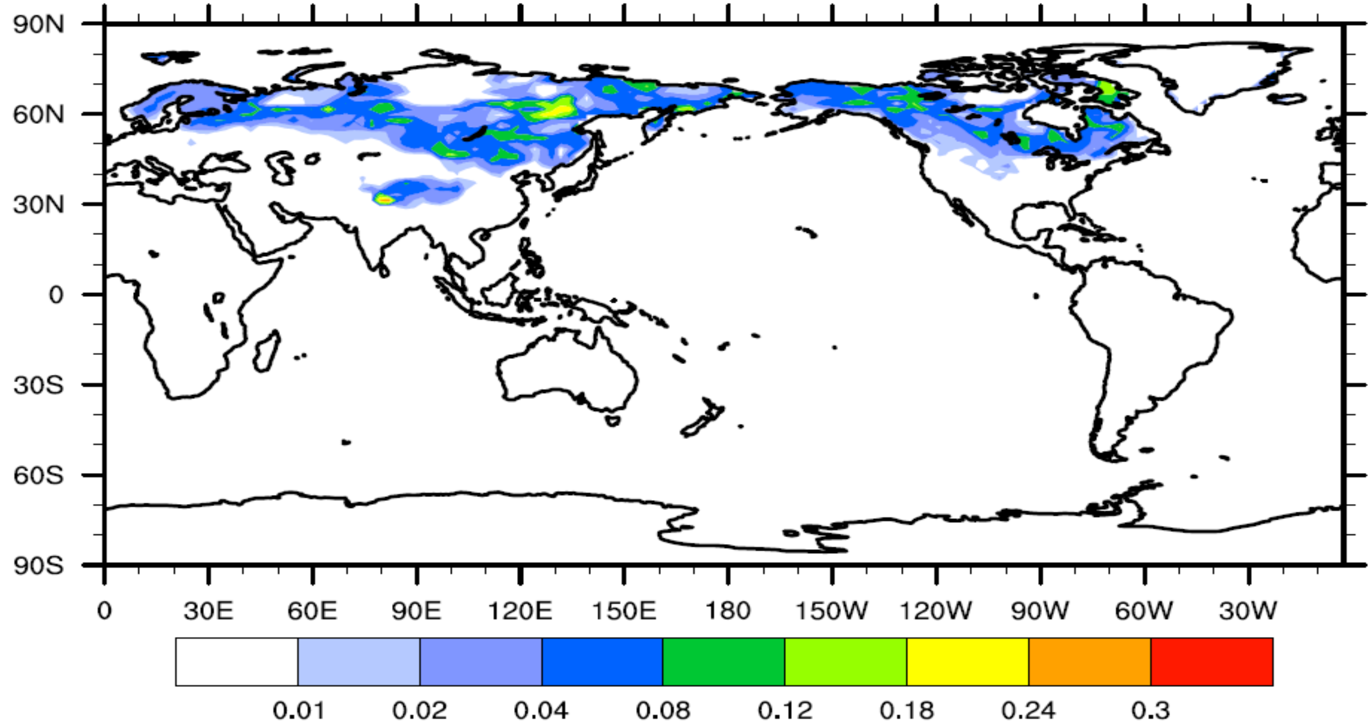
The National Aeronautics and Space Administration



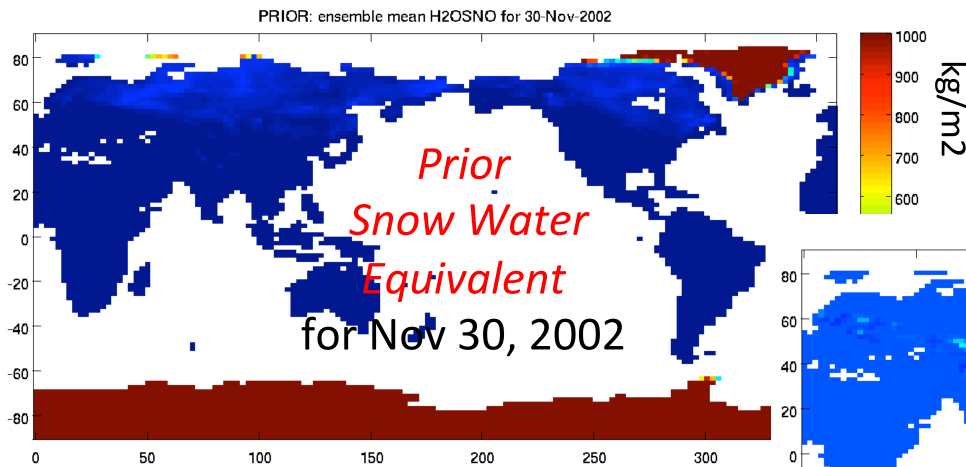
Assimilation of MODIS snow cover fraction

- 80 member ensemble for onset of NH winter, assimilate once per day
- Level 3 MODIS product – regridded to a daily 1 degree grid
- Observations can impact state variables within 200km
- CLM variable to be updated is the snow water equivalent "H2OSNO"
- **Analogous to precipitation** …

Standard deviation of the CLM snow cover fraction initial conditions for Oct. 2002

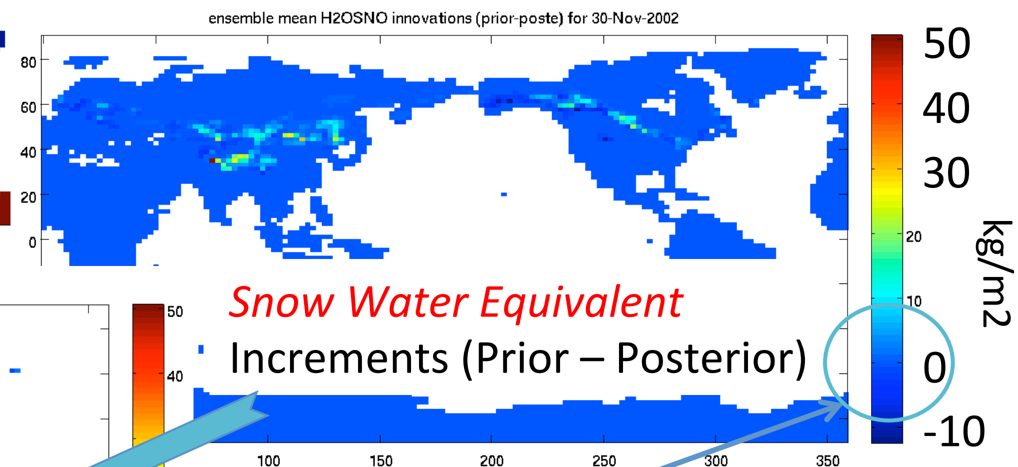


An early result: assimilation of MODIS *snow cover fraction* on total *snow water equivalent* in CLM.

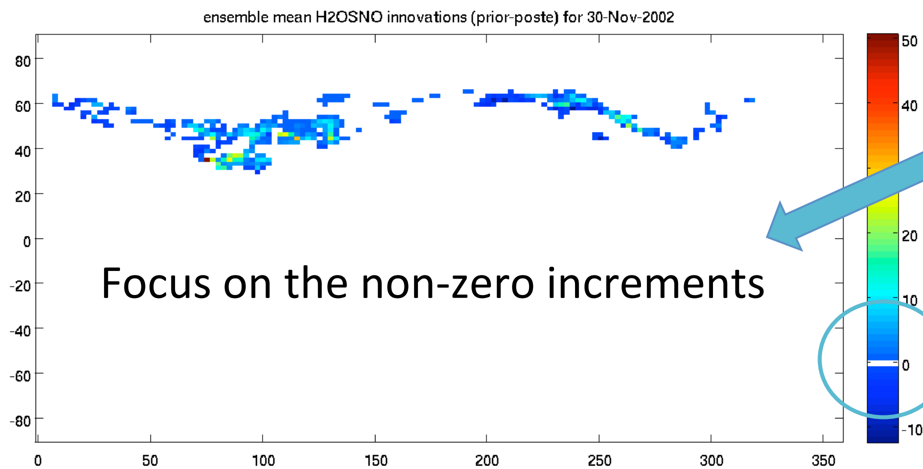


Prior Snow Water Equivalent
for Nov 30, 2002

Thanks Yongfei!



Snow Water Equivalent
Increments (Prior – Posterior)



Focus on the non-zero increments

The model state is changing in reasonable places, by reasonable amounts. At this point, that's all we're looking for.

Data Assimilation Research Testbed

The HARD part is: ***What do we do when SOME (or none!) of the ensembles have [snow,leaves,precipitation, …] and the observations indicate otherwise?***

Corn Snow?

New Snow?

Sugar Snow?

Dry Snow?

Wet Snow?

"Champagne Powder"?

Crusty Snow?

Slushy Snow?

Old Snow?

Dirty Snow?

Packed Snow?

Early Season Snow?

Snow Density?

Snow Albedo?

The ensemble ***must*** have some uncertainty, it cannot use the same value for all. The model expert must provide guidance. It's even worse for the hundreds of carbon-based quantities!

Data Assimilation of Cosmic-ray Derived Soil Moisture

Rafael Rosolem

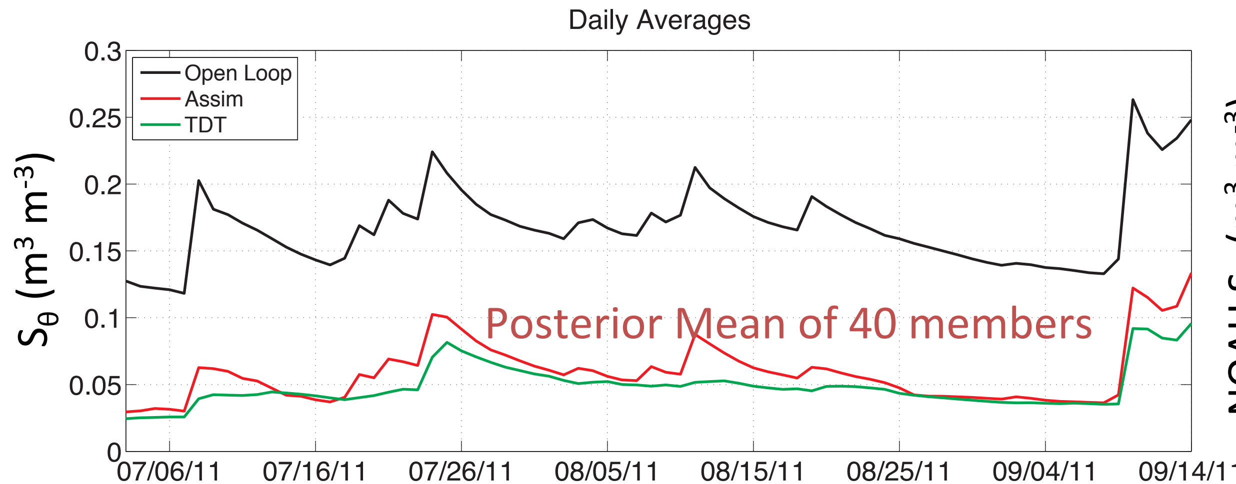
With acknowledgments to W. J. Shuttleworth, M. Zreda, X. Zeng, A. Arellano, T. Hoar, J. Anderson, T. Franz, S. A. K. Papuga, M. Barlage, C. Zweck, D. Desilets, G. Womack, J. Broermann, R. Chrisman, A. M. Karczynski

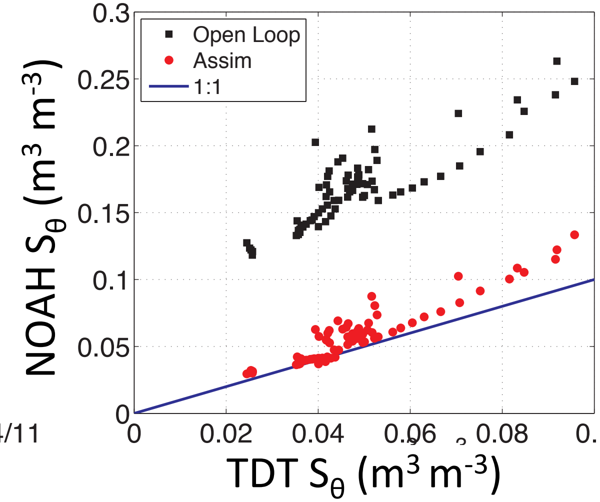
THE UNIVERSITY OF ARIZONA®

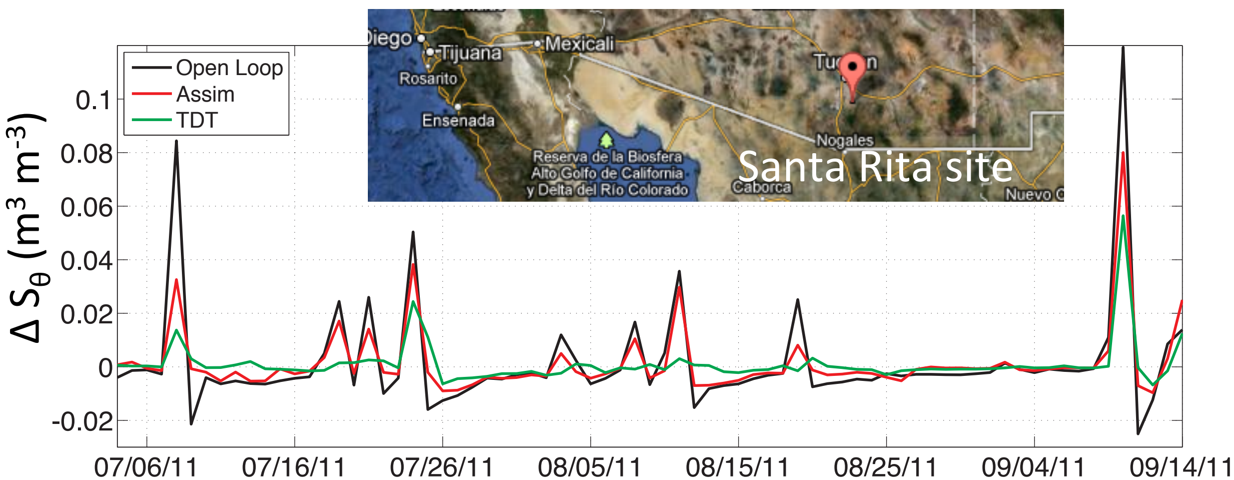
AGU Fall Meeting
December 04, 2012

NSF

NOAH-DART: Integrated Soil Moisture







Keys to ensemble land DA:

- What parts of the model 'state' do we update?
- Proper initial ensemble
- Can models tolerate new assimilated states? Silently fail?
- Observation location insufficient – ?standard? land cover type needed.
- Snow (vegetation) … depths, layers, characteristics, content.
  - When all ensembles have identical values the observations cannot have any effect with the current algorithms.
- Forward observation operators
  - many flux observations are over timescales that are inconvenient
  - need soil moisture from last month and now … GRACE
- Bounded quantities
  - Soils dry beyond their physical limits, for example.

For more information:

CAM

GITM

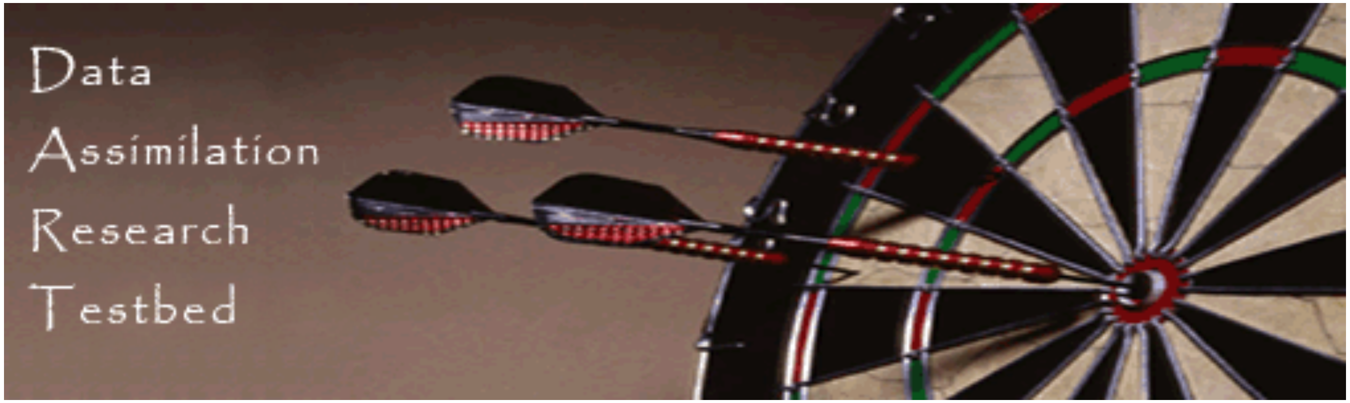
WRF

CLM

POP

AM2

BGRID



COAMPS

www.image.ucar.edu/DAReS/DART

NOAH

dart@ucar.edu

MPAS_ATM

MITgcm_ocean

COAMPS_nest

SQG NAAPS MPAS_OCN TIEGCM

PBL_1d

PE2LYR

NCOMMAS

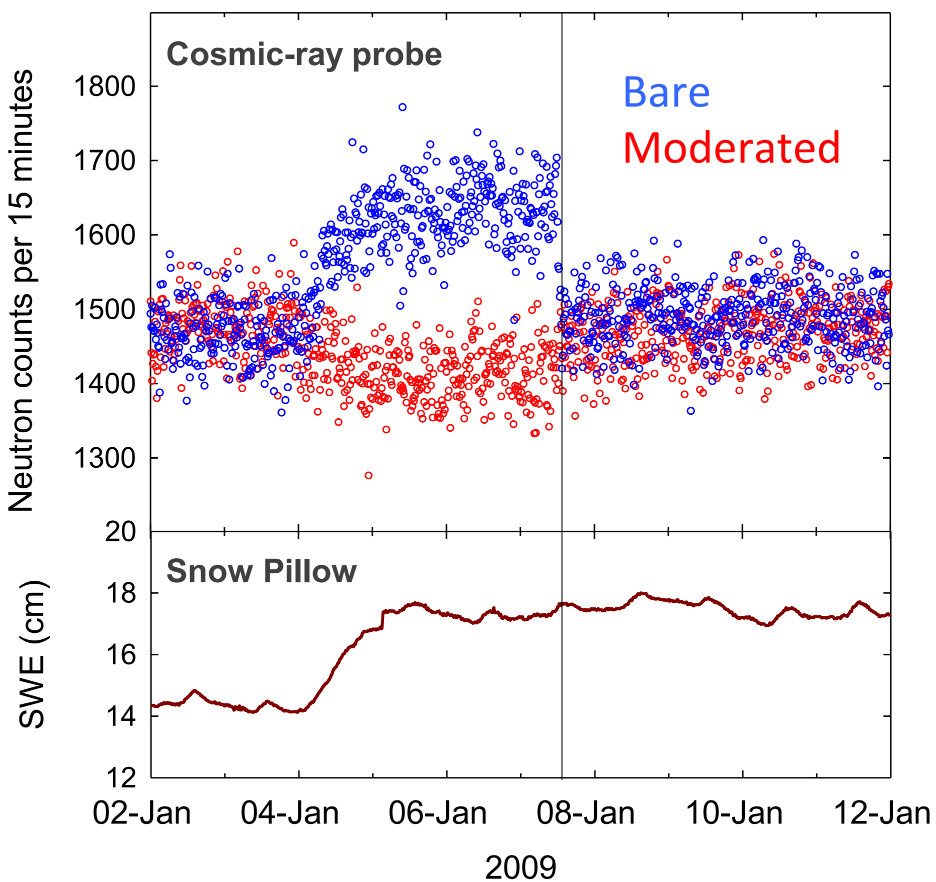
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

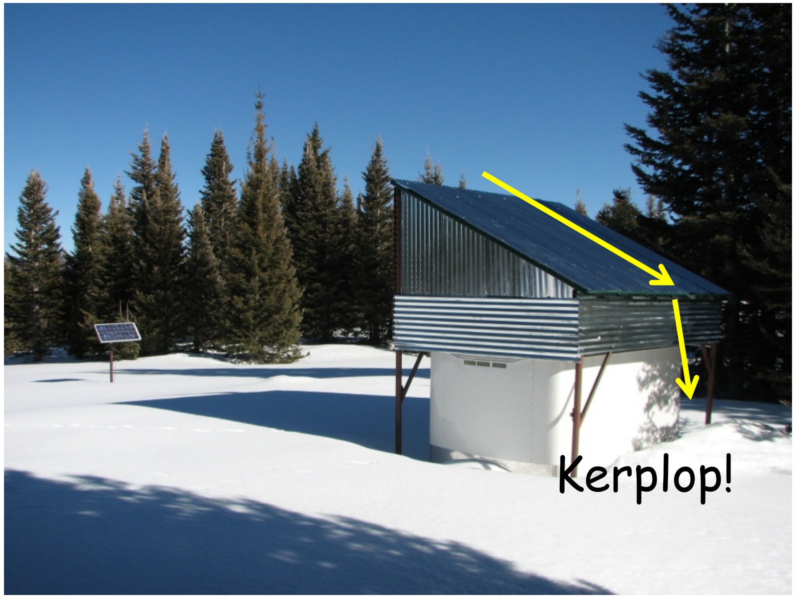
NSF

Data Assimilation Research Testbed


My favorite land use …

Bare counter enhancements?





- **DARIN DESILETS**
  Hydroinnova

Data Assimilation in CLM

- Main Objective
  - Process understanding
  - Regional quantification
  - Forecasting
- Dynamics
  - Physical, biological, chemical –
  - Only partially known, empirical relationships
- Observations
  - Very different spatial and temporal characteristics
- Mathematical problem
  - Initial value problem (e.g. pools)
  - Boundary conditions (e.g. fluxes)
  - Parameter optimization

A short list of models that can assimilate with DART:

- CAM: Community Atmosphere Model
- POP: Parallel Ocean Program
- WRF: Weather Research and Forecasting Model
- AM2: GFDL Atmospheric Model
- COAMPS: Coupled Atmosphere/ Ocean Mesoscale Prediction System
- CLM: Community Land Model
- NOAH: Land Surface Model
- … many more