Land Surface Data Assimilation: DART and CLM

Yongfei Zhang: *University of Texas Austin*

<u>Tim Hoar</u>, Bill Sacks, Tony Craig, Jeff Anderson: *National Center for Atmospheric Research*

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

Zong-Liang Yang: *University of Texas Austin*

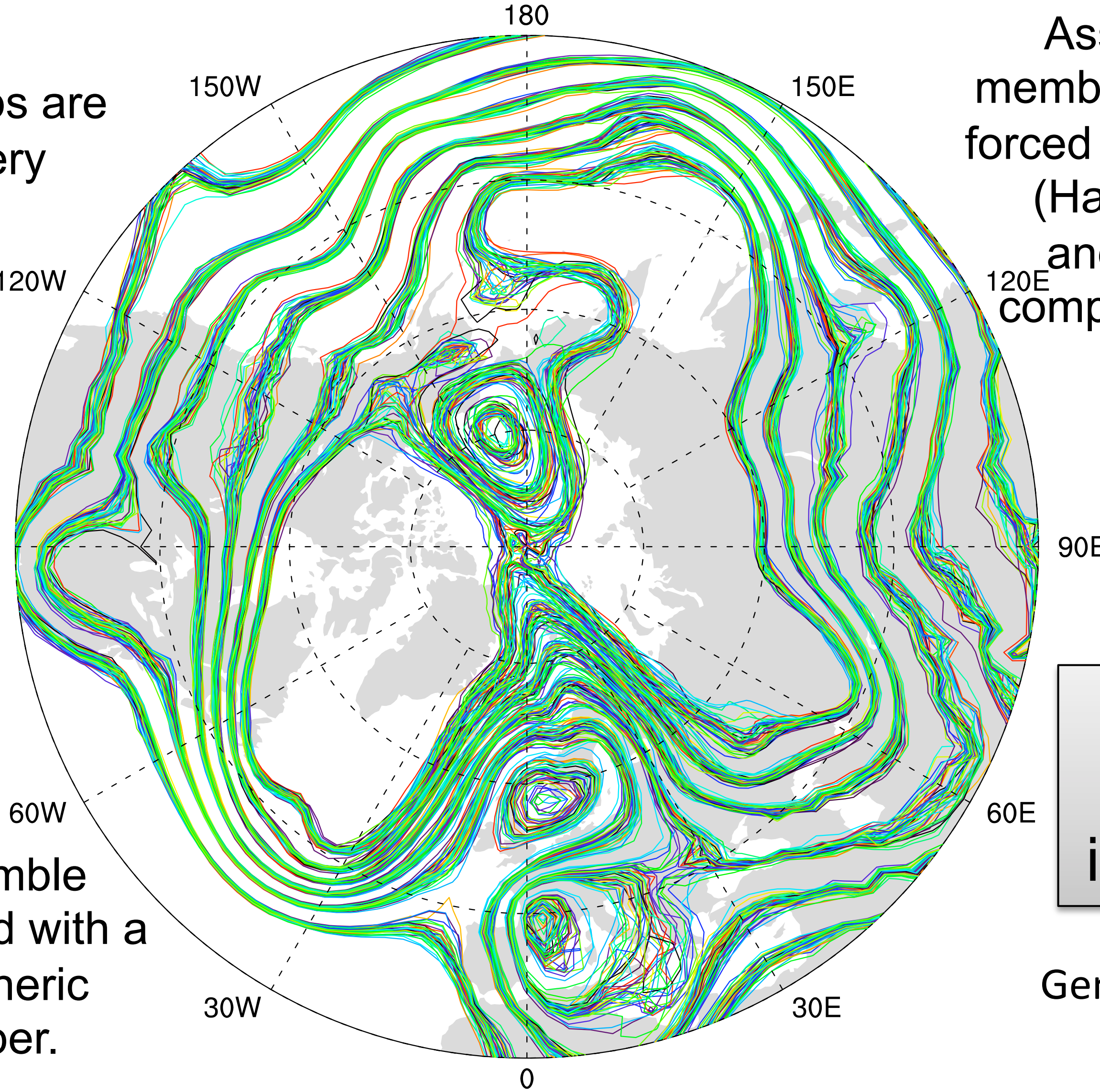
Talk Outline

1. Overview of DART and CLM and how they relate to one another (1 slide).
2. Proof of concept with synthetic leaf carbon observations in a ‘perfect model’ scenario.
3. Assimilation with MODIS snow cover fraction, restricted to updating just the CLM snow cover fraction. *Snow is tricky!*
4. Discussion of problems, potential solutions, and what’s next.

Atmospheric Reanalysis

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

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



500 hPa GPH
Feb 17 2003

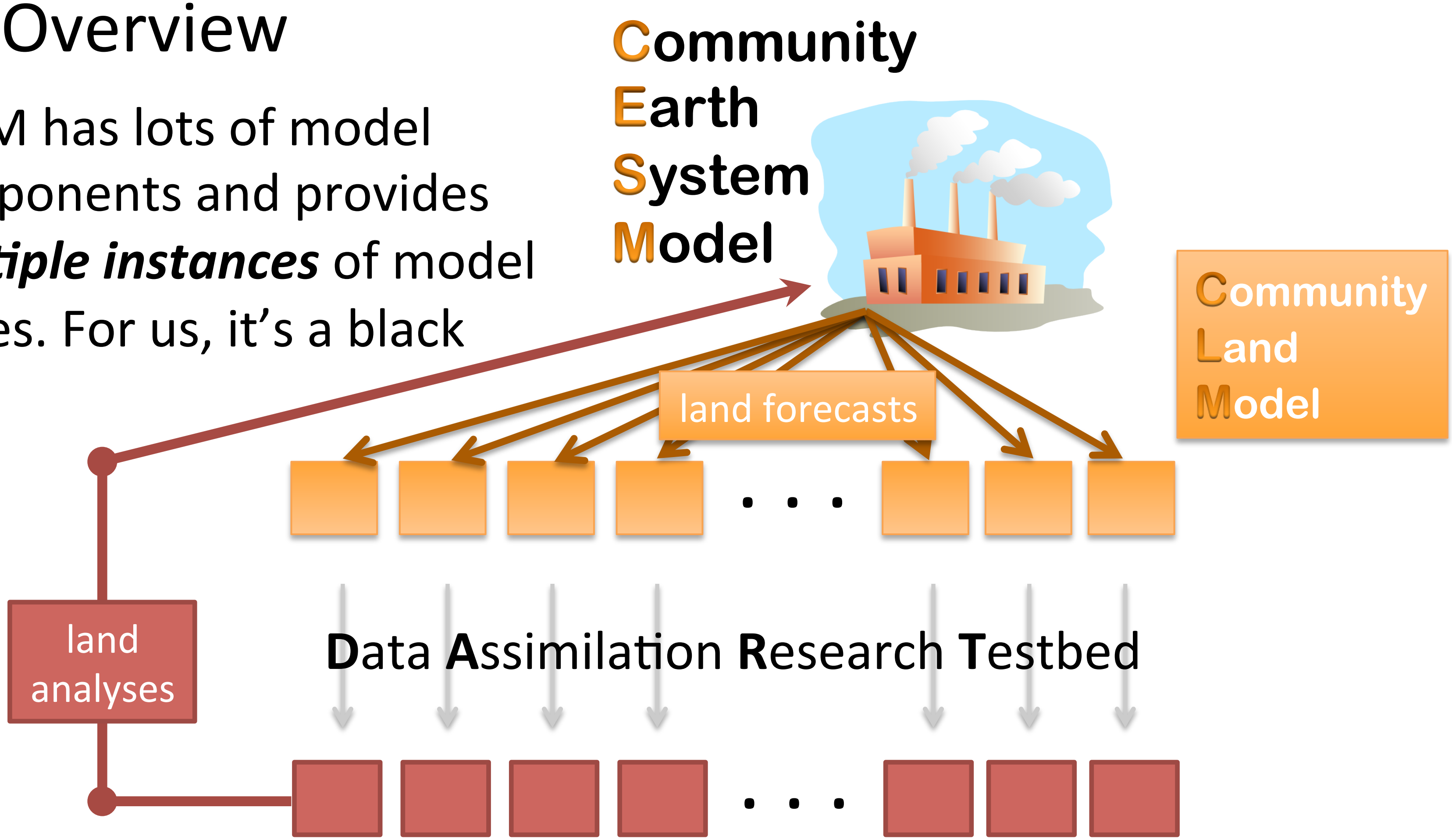
1998-2010
4x daily
is available.

Each CLM ensemble member is forced with a different atmospheric reanalysis member.

Generates spread in the land model.

Overview

CESM has lots of model components and provides ***multiple instances*** of model states. For us, it's a black box.



Given this ***ensemble*** and observations, DART determines increments for the model states, the model states get updated, and the ensemble is fed back to CESM to be advanced to the next desired time.

Details

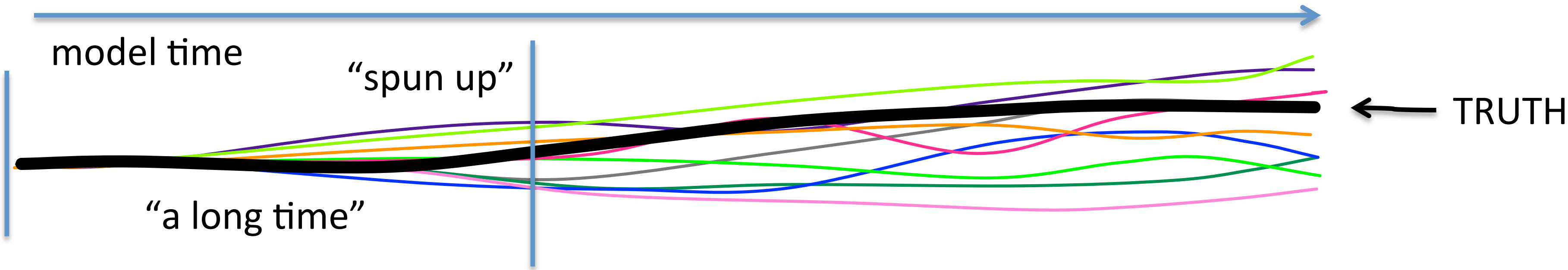
- DART allows you to choose what **CLM variables** get updated by the assimilation.

```
&clm_vars_nml
   clm_state_variables = 'frac_sno',     'KIND_SNOWCOVER_FRAC',
                         'DZSNO',        'KIND_SNOW_THICKNESS',
                         'H2OSNO',       'KIND_SNOW_WATER’,
                         'T_SOISNO',     'KIND_SOIL_TEMPERATURE',
                         ‘leafc’,        ‘KIND_LEAF_CARBON’ /
```

- These are read from a CLM restart file and reinserted after the assimilation.
- **Potential problem … balance/consistency?**

Proof-of-concept schematic.

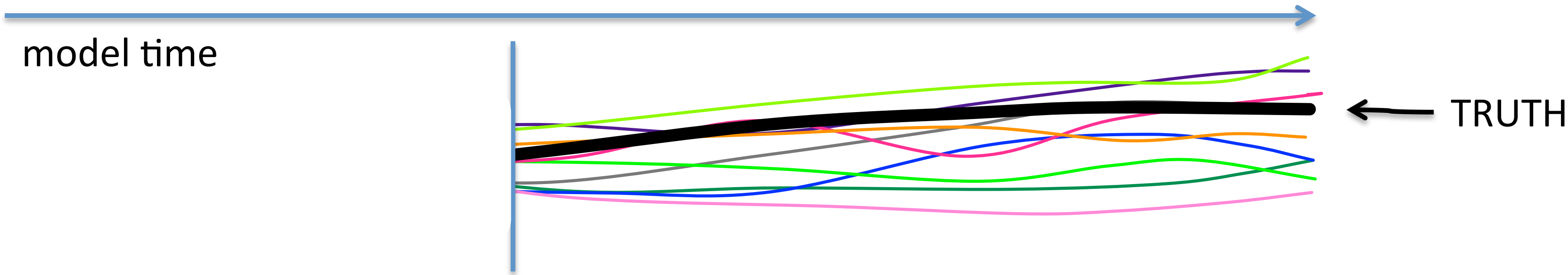
The lines represent the evolution of individual instances of CLM. Pick any one and declare it to be the TRUTH



Now, harvest some synthetic observations from the instance we declared to be the TRUTH.

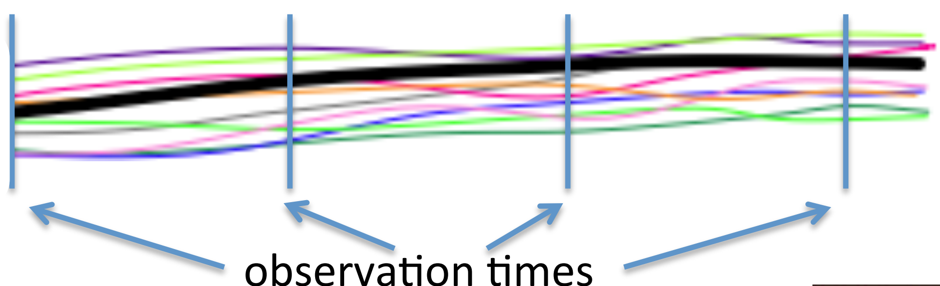
Proof-of-concept schematic.

Without Assimilation:
Frequently, the ensemble spread simply grows.



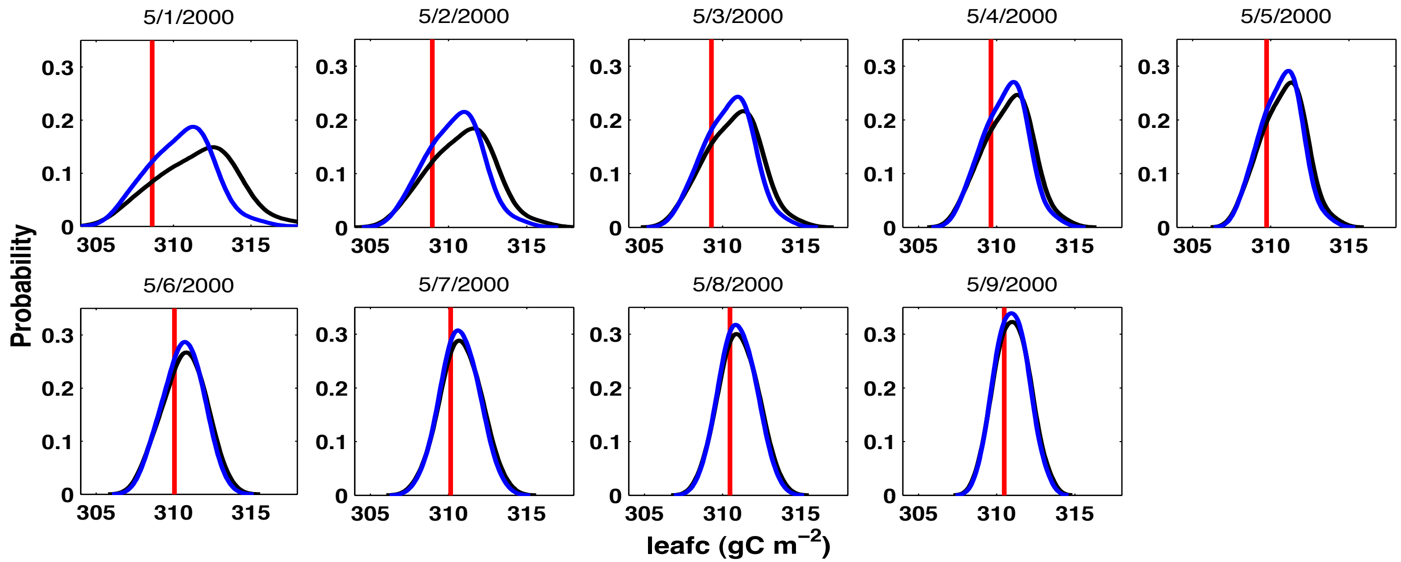
With Assimilation: ensemble spread ultimately remains stable and small enough to be informative, but not so small that it collapses away from the Truth.

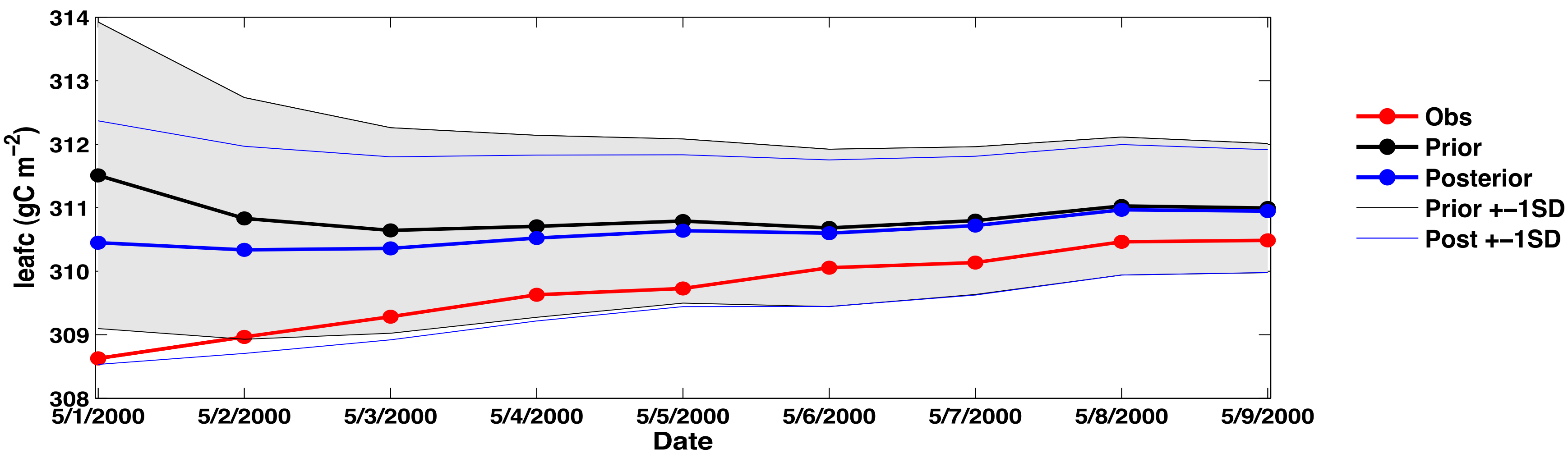
Problem: Getting a proper initial ensemble is an area of active research.



Proof-of-concept with leaf carbon

Prior and posterior probability distributions of leaf carbon in a single grid cell at 60°W, 4°S for nine days of assimilation

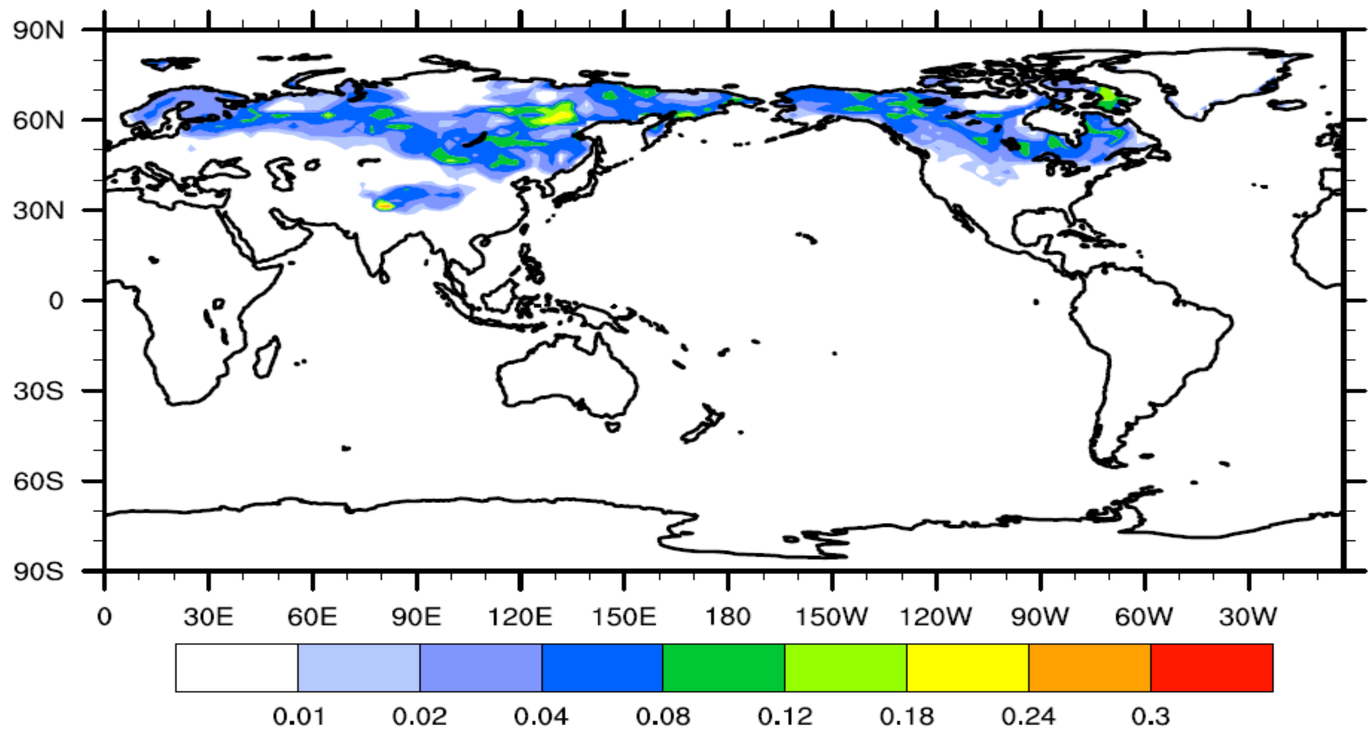




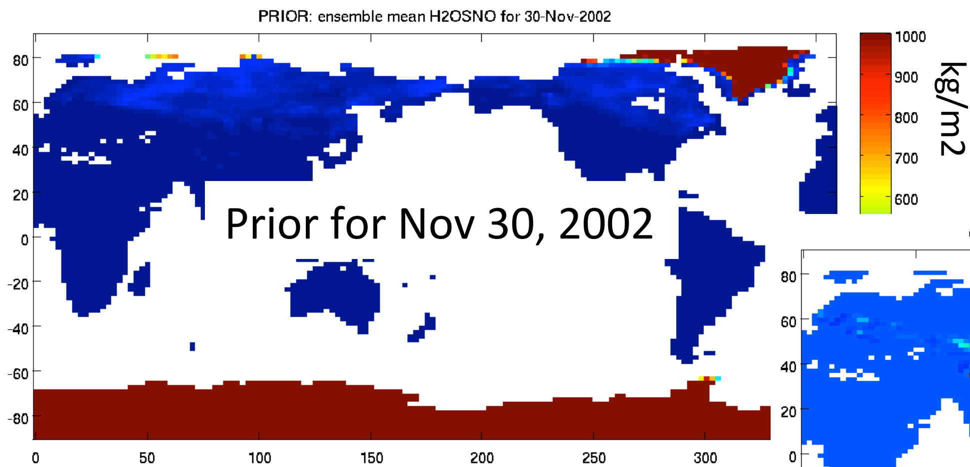
MODIS assimilation experiment

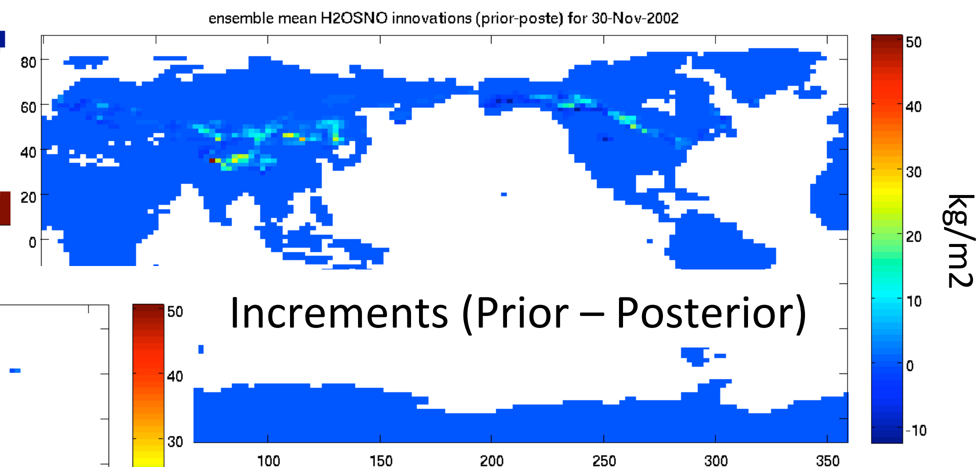
- 80 member ensemble for onset of NH winter
- Assimilate once per day
- Regridded MODIS product, not raw observations (suboptimal)
- Observation error variance is 0.1 (for lack of a better value)
- Localization set to 0.03 radians ~ 200km half-width
- CLM variable to be updated is the snow water equivalent "H2OSNO"

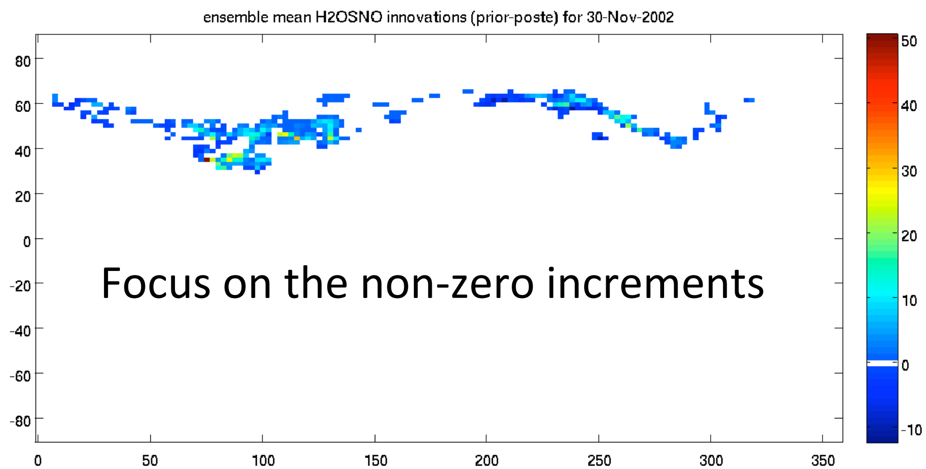
Standard deviation of the snow cover fraction initial conditions for Oct. 2002



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







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

The HARD part is: ***What do we do when only SOME (or none!) of the ensembles have [snow,leaves,…] 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!

Problems to be solved:

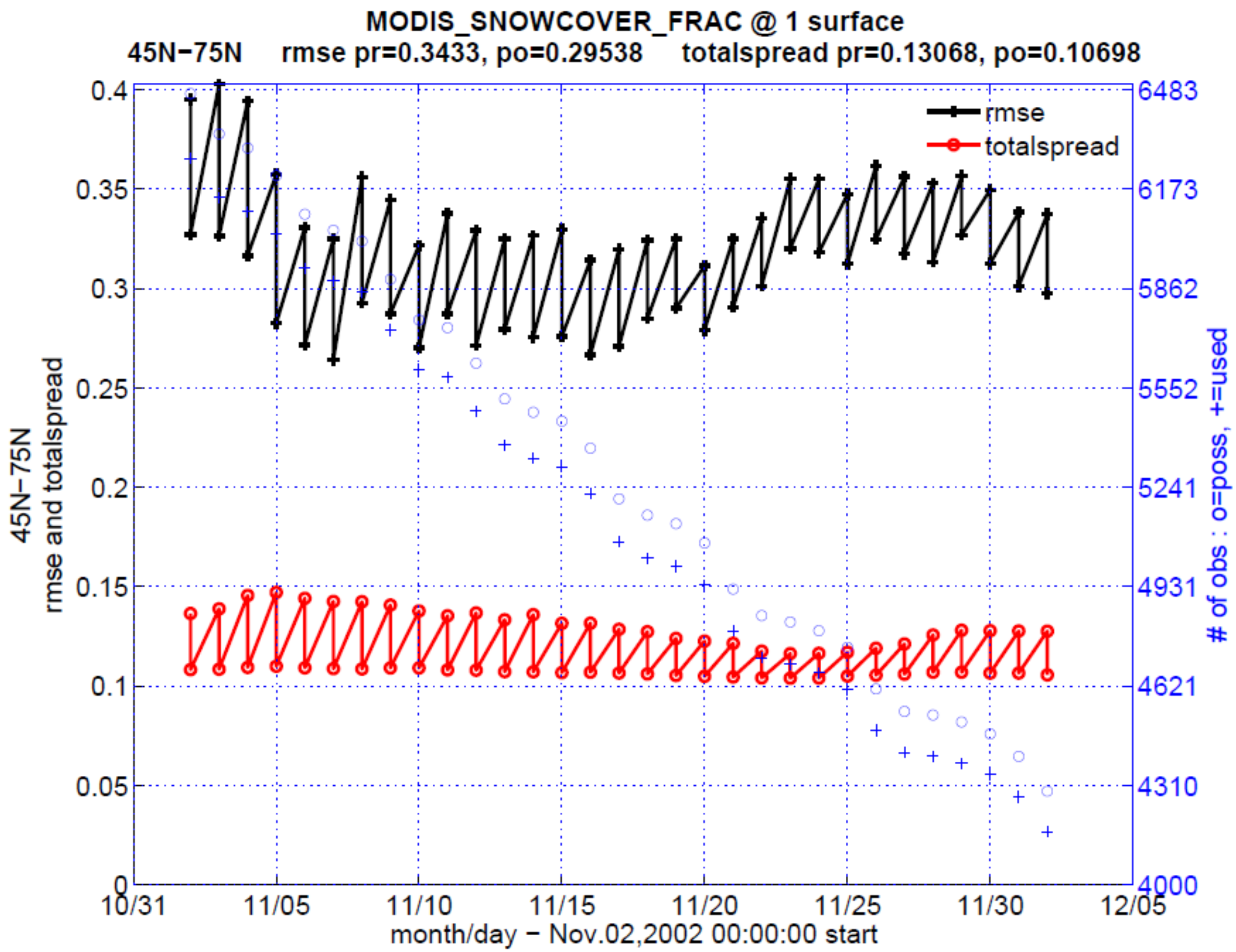
- Proper initial ensemble
- Creating snow with the right characteristics
- Bounded quantities – when all ensembles have identical values the observations cannot have any effect with the current algorithms
- Forward operators – many flux observations are over timescales that are inconvenient – need soil moisture from last month and now…
- CLM has **a lot** of carbon species, hard to support all the forward operators required
- CLM's abstraction of grid cells, land units, etc., make the treatment of observations very peculiar. All land units in a grid cell share a location. Easy to have 'contradictory' observations.

For more information:

www.image.ucar.edu/DAReS/DART

dart@ucar.edu

slide held in reserve



slide held in reserve