Data Assimilation for CLM:
a comprehensive overview

in 12 minutes!

Tim Hoar: NCAR 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)
Data assimilation can put the model state more in line with the current state. This allows us to:

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

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.
What is Data Assimilation?

Observations combined with a Model forecast...

... to produce an analysis.

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

A generic ensemble DA 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.
Keys to ensemble land DA:

1. What parts of the model ‘state’ do we update?
   1. The stock CLM restart files have *hundreds* of variables in them. Knowing which ones to update is up to the researcher!

2. What is a “proper” initial ensemble?
   1. How many model instances do we need?
   2. How do we get them?
   3. Does it maintain realistic uncertainty? Is it still informative?

3. We have imperfect knowledge of the “forcing” fields.
   1. Will the inference change with slightly different forcing?
   2. Does the forcing overwhelm the sparse observations?

4. Can models tolerate new assimilated states?
   1. Model variables not necessarily ‘in balance’ or consistent anymore. *What happens in a coupled framework?*
   2. Silently fail?
Keys to ensemble land DA (cont’d):

5. What happens when CLM and the observations are in violent disagreement? *Can only be answered by the researcher!*
   1. Snow vs. bare ground
   2. Senescence, etc.

6. Assimilation affects bounded quantities.
   1. Soils dry beyond their physical limits, for example.

7. Need forward observation operators.
   1. How do we estimate the observation value given the CLM state? Ally Touré [NASA] here now to do this for AMSR-E brightness temperatures.

8. Observation metadata is very important for accurate forward observation operators. *This is the next thing on my to-do list.*
   1. Location information alone is insufficient. Land cover type needed.
CLM abstracts the gridcell into a “nested gridcell hierarchy of multiple landunits, snow/soil columns, and Plant Function Types”. This is particularly troublesome when trying to convert the model state to the expected observation value because:
Location information is contained at this level ONLY!

Observations occur here!

CLM abstracts the gridcell into a “nested gridcell hierarchy of of multiple landunits, snow/soil columns, and Plant Function Types”. This is particularly troublesome when trying to convert the model state to the expected observation value because: 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!
DART Multiple Component Data Assimilation

Important! There are multiple instances of each model component.

DART assimilates the observations into the components separately.

Started with CCSM4 20th Century 30-member ensemble for all model components.

COUPLER

CAM

CLM

DART

POP

ROF

CICE

B compset CESM1_1_1

TJH CESM 2013 pg 9
Assimilation of the MODIS Snow Cover Fraction data through DART/CLM4

Y. Zhang\textsuperscript{1}, T. J. Hoar\textsuperscript{2}, Z-L. Yang\textsuperscript{1}, J. L. Anderson\textsuperscript{1}, A. Tourp\textsuperscript{1}, M. Rodell\textsuperscript{1}

1. Jackson School of Geosciences, University of Texas at Austin, Austin, TX, United States.
2. The National Center for Atmospheric Research, Boulder, CO, United States.
4. NASA Goddard Space Flight Center, Greenbelt, MD, United States.

(yongfei@utexas.edu)

Check out Yongfei's poster!

The DART/CLM4 Data Assimilation System

- Assimilation of Snow
- As an extension to previous discussions
- Experiments with SAM

Meteorological Forcings from DART/CLM4

- A freely available ensemble reanalysis data created by DART and the Community Atmospheric Model (CAM4) is used to drive the CLM ensemble members.
- The CAM4 reanalysis is similar to the NCEP reanalysis, except for the ensemble and a product of the coupled DART and CAM4.
- The CAM4-produced ensemble reanalysis forcing fields are physically and mutually consistent for a given member and exhibit ensemble spread sparsely temporally.

Results

- A localization distance of 0.05 degrees stands out among a series of localization distances, producing the smallest RMSE.
- In winter, SFC ensemble spread is mainly located in lower-middle latitude regions. In spring, the spatial pattern of SFC ensemble spread extends northward, indicating that the uncertainty of modeled snow in high latitude regions increases as snow start melting.
- Snow data assimilation shows little change on SFC at higher-middle and high latitudes in winter due to the fact that SFC in CLM4 reaches the unity too fast compared to MODIS data.
- The effectiveness of data assimilation on model state varies with vegetation types, with mixed-pixel performance over forest regions and consistently good performance over grassland areas.

Conclusions

Acknowledgements

This work is supported by NASA Grant NNX09AF82G and the NCAR Advanced Study Program.
Slides held in reserve
Looking at it another way:

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

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

Result of the forward observation operator for ensemble member 1

Directly from ensemble member 1

Could be Soil Temperature
Directly from ensemble member 2

“observation” from ensemble member 2
In our global atmospheric assimilations, we use 80.

3 IS NOT ENOUGH!
Regression Error!

Least-squares fit

Now, we can calculate out observation increments any way we want.
a) The “observation” Posterior for member 1

b) which projects to here:

c) Which means the unobserved Posterior should be:
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.*

Could be Soil Temperature
I got these from Dave Lawrence. I don’t know if he made them or not – but Thanks to whomever did!
I got these from Dave Lawrence. I don’t know if he made them or not – but Thanks to whomever did!
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
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”?*
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.
The ensemble advantage.

You can represent uncertainty.

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.
Atmospheric Ensemble Reanalysis

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.
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.
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
- (40° 02' 09" N; 105° 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: LEAFCC, LIVEROTTC, LIVESTEMC, DEADSTEMC, LITR1C, LITR2C, SOIL1C, SOIL2C, SOILLIQ ... all of these are unobserved.
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
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”? Slushy Snow? Dirty Snow?

Early Season Snow? Snow Density?

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!
NOAH-DART: Integrated Soil Moisture

Posterior Mean of 40 members

Santa Rita site

Raphael at Tonzi Ranch