DART and Land Data Assimilation

Tim Hoar: National Center for Atmospheric Research
with a whole lot of help from:
Jeff Anderson, Nancy Collins, Kevin Raeder, Bill Sacks: NCAR
Yongfei Zhang: University of Texas Austin
Andrew Fox: National Ecological Observatory Network (NEON)
Rafael Rosolem: University of Arizona
What is Data Assimilation?

Observations combined with a Model forecast...

\[ \text{Observations} + \text{Model forecast} = \text{Analysis} \]

... to produce an analysis.

Overview article of DART:

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 ...
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
Fully coupled assimilation will need data from all models at the same time.
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.
• One unbreakable rule: “Do Not Invade the model code.”
• Unique routines communicate between each model and DART to provide this separation.
• I want to use COSMOS observations with “all” of our land models. This means our forward observation operator must be pretty generic.
• The land models are an area where observations are needed to help constrain the model states – so we can learn about and improve the models.
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.
Creating the initial ensemble of ...

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.

"spun up"
The ensemble advantage.

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

If we can reduce the ensemble spread and still be indistinguishable, all the better.

With a good assimilation the ensemble is indistinguishable from the true model state in any meaningful way.

You can represent uncertainty.
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.

Can use these to force other models.

1998-2010 4x daily is available.
**Pros and Cons**

- 80 realizations/members
- Model states are self-consistent
- Model states consistent with observations
- Available every 6 hours
- 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?

Since Rafael is going to be showing results of NOAH, I’ll explore some of these (and other) issues with results from CLM – one for a flux tower, one for global snow 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
- (40º 02' 09'' N; 105º 32' 09'' W; 3021 m)
- Single column of Community Land Model
- 64 ensemble members of CLM
- 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 Run of CLM

Leaf Area Index

Carbon

Nitrogen

Ecosystem N (gN m$^{-2}$)

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

FYI: We are assimilating flux observations...

Focus on the ensemble means (for clarity)
These are all unobserved variables.
Effect on short-term forecast on unobserved variables.

Leaf Area Index

Free Run
Assim
Forecast

Carbon

Nitrogen

Ecosystem N (gN m\(^{-2}\))
Effect on longer-term forecast

Again, these are model variables.

Leaf Area Index

Carbon

Nitrogen

Year

Ecosystem N (gN m$^{-2}$)
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 snowcover fraction on total snow water equivalent in CLM.

Prior for Nov 30, 2002

Focus on the non-zero increments

Increments (Prior – Posterior)

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

Thanks Yongfei!
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?
- Crusty Snow?
- Old Snow?
- Packed Snow?
- 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!
As I see it, problems to be solved:

- Proper initial ensemble
- Can models tolerate new assimilated states?
- Snow ... depths, layers, characteristics, content.
  - When all ensembles have identical values the observations cannot have any effect with the current algorithms.
  - COSMOS forward observation operator for NOAH-MP, CLM ...
- Forward observation operators
  - many flux observations are over timescales that are inconvenient
  - need soil moisture from last month and now ... GRACE
  - Multisensor soil moisture assimilation?
- Bounded quantities
  - Negative SW fluxes, for example.
  - Soils dry beyond their limits.
For more information:

- CAM
- CLM
- AM2
- COAMPS
- MITgcm_ocean
- SQG
- NAAPS
- MPAS_OCN
- NCOMMAS
- TIEGCM
- PE2LYR
- POP
- BGRID
- WRF
- NOAH
- MPAS_ATM
- COAMPS_nest
- PBL_1d

www.image.ucar.edu/DARes/DART
dart@ucar.edu