Multi-instance CESM for Fully Coupled Data Assimilation Capability using DART

Kevin Raeder, Tim Hoar, Nancy Collins, Jeffrey Anderson: DARES
Mariana Vertenstein, Tony Craig, Jim Edwards: SEWG
Alicia Karspeck: CGD/OS
Yongfei Zhang: U. of Texas Austin
We are building an ensemble data assimilation system for NCAR’s Community Earth System Model using DART (Data Assimilation Research Testbed)

+ Strong support from SEWG: Vertenstein, Craig, Edwards
+ Use of new multi-instance capability: CESM advances an ensemble of CAMs and/or POPs ... simultaneously.
+ CESM/CAM+DART is nearly as easy as CESM
+ Fully coupled data assimilation with any/all CESM components is within view.
Past CAM Assimilation

DART runs the show

Obs
Past CAM Assimilation

DART runs the show

Obs

CAM initial files

3D state
Past CAM Assimilation

DART runs the show

Obs

CAM initial files

3D state

Limited parallelization required a new algorithm

AMWG Winter 2012
Current CAM Assimilation with CESM ensemble.

In a mind-boggling example of anticipating users’ needs, SEWG was already working on the multi-instance capability and was looking for something to exercise it.
Current CAM Assimilation with CESM ensemble.

CAM Coupler runs the show

2D forcing

3D restart
Current CAM Assimilation with CESM ensemble.

Obs

DART

3D state

3D restart

CAM

2D forcing

Coupler runs the show

Current CAM Assimilation with CESM ensemble.

AMWG Winter 2012
Current CAM Assimilation with CESM ensemble.

- Obs
- DART
- 3D state
- 3D restart
- CAM
- Output @ each observation time
- 2D forcing
- Coupler runs the show

Current CAM Assimilation with CESM ensemble.

AMWG Winter 2012
Advantages & Opportunities

- Any atmospheric model;
  - FV and Eulerian dy-cores, any resolution
  - CAM-MPAS interface; alpha testing nearing completion
  - CAM-SE interface is not done yet
  - Physics packages (CAM4, CAM5, ...)
  - WACCM; available, not tested, beta testers wanted
  - CAM+SKEBS (Berner), and other CAM variants
- Consistent with how the POP and CLM assimilations are being done
- Facilitates assimilation into a fully coupled model
- Advantages/applications we haven't thought of . . .
How hard is this really?

"I'm impressed by how easy you've made this to run" — Robert Pincus, beta tester
How hard is this really?

"I'm impressed by how easy you've made this to run ” Robert Pincus, beta tester

The CESM+DART setup script:

- Defines and builds a standard case (F_2000 for now),
- but uses the multi-instance capability for the atmospheric component.
- Modifies CESM scripts and namelists to
  - use namelist files appropriate for actual dates,
  - define the ensemble size,
  - manage the startup files,
  - run DART between the forecast and the archiving,
  - archive the new DART output.
How hard is this really?

"I'm impressed by how easy you've made this to run " Robert Pincus, beta tester

The CESM+DART setup script:

✓ Defines and builds a standard case (F2000 for now),
✓ but uses the multi-instance capability for the atmospheric component.
✓ Modifies CESM scripts and namelists to

• use files appropriate for actual dates,
• define the ensemble size,
• manage the startup files,
• run DART between the forecast and the archiving,
• archive the new DART output,

= A few hundred lines of code

= A few dozen lines which a user might want to change

Validated on bluefire and hopper (NERSC Cray XT5).
Should run anywhere that CESM does.
1-deg CAM4+DART 6 hour Assimilation

- 20 bluefire nodes × 2 hours/day
- in /ptmp (/glade is 10-100% slower)
- CAM5 has more levels, chemistry, larger state vector, ...
- Significant efficiencies are in the works
- and we get useful results from days, not decades,
- but yellowstone will be welcome! (we hope)
What do we get out of this?
Atmospheric Ensemble Reanalysis, 1998-2010

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.

500 hPa GPH
Feb 17 2003
Ensemble Mean Increment Due to Assimilation

These are some of the corrections to CAM’s moisture resulting from assimilation of T, U, and V (no Q) observations.

Specific humidity

- ~200 hPa
- -.06 < ΔQ < .05 g/kg
  
  (.05 < Q < .10 g/kg)
6-hour forecast Bias of Radiosonde Specific Humidity (Q)

December 2006

AMWG Winter 2012
Coupled Free Run

Physical Space: 1998/1999 SST Anomaly from HadOI-SST

48 POP 48 CAM

POP forced by observed atmosphere (hindcast)
Current Land (CLM) Assimilation.

Implementing **THIS IS EASY!!!**
If we restrict ourselves to the simple cases ... here is the early result of an assimilation of MODIS snowcover fraction on total snow water equivalent in CLM.

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.
Coupler
CAM
DART
Ocn Obs
POP
Atm Obs
CLM
CICE
Lnd Obs
AMWG Winter 2012
Fully coupled assimilation will need data from all components at the same time.

Each component corrected by all kinds of observations.
Challenge for Earth System Model

DA

- Atmospheric components of earth system models may not be as mature as NWP models.
- Model systematic or algorithmic errors may be large.
- Can lead to reduced quality analyses.
- But, DA can help to detect and correct errors!
Summary

CESM components+DART ensemble DA exists for:

- CAM: Multiyear ensemble reanalysis available,
- POP: Ensemble analyses used for decadal prediction initial conditions,
- CLM: Ensemble snow cover analyses and leaf area index research.
Plans and Challenges

- Build fully coupled earth system model ensemble assimilation system.
- Methods for dealing with land surface variables.
- Methods for dealing with strongly biased models.
Extras
Ensemble DA Sensitivity of Cyclone Central Pressure to Mean Sea Level Pressure (left) and 300 hPa geopotential height (right)

slide held in reserve
These counts are for 1998 & 1999 and are representative.

- temperature observation error standard deviation == 0.5 K.
- salinity observation error standard deviation == 0.5 msu.

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLOAT_SALINITY</td>
<td>68200</td>
</tr>
<tr>
<td>FLOAT_TEMPERATURE</td>
<td>395032</td>
</tr>
<tr>
<td>DRIFTER_TEMPERATURE</td>
<td>33963</td>
</tr>
<tr>
<td>MOORING_SALINITY</td>
<td>27476</td>
</tr>
<tr>
<td>MOORING_TEMPERATURE</td>
<td>623967</td>
</tr>
<tr>
<td>BOTTLE_SALINITY</td>
<td>79855</td>
</tr>
<tr>
<td>BOTTLE_TEMPERATURE</td>
<td>81488</td>
</tr>
<tr>
<td>CTD_SALINITY</td>
<td>328812</td>
</tr>
<tr>
<td>CTD_TEMPERATURE</td>
<td>368715</td>
</tr>
<tr>
<td>STD_SALINITY</td>
<td>674</td>
</tr>
<tr>
<td>STD_TEMPERATURE</td>
<td>677</td>
</tr>
<tr>
<td>XCTD_SALINITY</td>
<td>3328</td>
</tr>
<tr>
<td>XCTD_TEMPERATURE</td>
<td>5790</td>
</tr>
<tr>
<td>MBT_TEMPERATURE</td>
<td>58206</td>
</tr>
<tr>
<td>XBT_TEMPERATURE</td>
<td>1093330</td>
</tr>
<tr>
<td>APB_TEMPERATURE</td>
<td>580111</td>
</tr>
</tbody>
</table>
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”?
- Slushy Snow?
- Dirty Snow?
- Old Snow?
- Early Season Snow?
- Packed Snow?
- Snow Density?
- Snow Albedo?
- 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!