Data Assimilation in CESM without breaking the bank: Pause-Resume to the rescue?

Steve Goldhaber

National Center for Atmospheric Research
CGD: Atmospheric Modeling & Predictability

March 3, 2016
Outline

- Data assimilation and CESM
- What’s the problem?
- Can we make it better?
- Plans
Acknowledgments

Software Engineering

- Nancy Collins (DAReS ← IMAGE ← CISL)
- Hellen Kershaw (Resides In Providence) ← No longer at NCAR
- Brian Eaton
- Sheri Mickelson (ASAP ← TDD ← CISL)

Science Support

- Alicia Karspeck
- Gokhan Danabasoglu
- Joe Tribbia
- Mariana Vertenstein
- Jeff Anderson (DAReS ← IMAGE ← CISL)
Data assimilation and CESM

- The Data Assimilation Research Section (DAReS) provides the Data Assimilation Research Testbed (DART).
- CSEG and DAReS have provided a platform for data assimilation studies (e.g., decadal prediction) with CESM and DART.
CESM run in a multi-instance configuration with the atmosphere and/or ocean nudged with observations processed by the DART filter.
Data assimilation and CESM

- The Data Assimilation Research Section (DARes) provides the Data Assimilation Research Testbed (DART).
- CSEG and DARes have provided a platform for data assimilation studies (e.g., decadal prediction) with CESM and DART.
- So, what is the problem?
- It is really slow (e.g., 2 sim-years per wallclock month)
- So where does the time go?
Where does the time go?

- CESM initializing (35%)
- CESM running (20%)
- Assimilation (30%)
- Other (15%)

Figure courtesy of Alicia Karspeck / NCAR
Current CESM-DART Sequence

1. CESM Initialization
2. Components run (atm, Ind, ocn, rof, ice, glc, wav, cpl)
3. CESM writes restart files and stops
4. Script runs to marshal files for DART input
5. DART runs
6. Script runs to marshal files for CESM restart
7. Repeat steps 1 - 6
Current CESM-DART Sequence

- CESM Initialization
- CESM Run (driver, coupler, components)
- Other (e.g., file manipulation)
- Data Assimilation
Current CESM-DART Sequence

- CESM Initialization
- CESM Run (driver, coupler, components)
- Other (e.g., file manipulation)
- Data Assimilation
Can we make it better?

• Currently, CESM and DART run as separate executables.

• In order to shrink the time taken up by CESM initialization, we plan to incorporate data assimilation as part of CESM.

• CESM currently recognizes seven component types (atm, lnd, ocn, ice, glc, rof, and wav).

• External processing, such as data assimilation will be an eighth component type for CESM, dubbed External System Processing (ESP).

• However, CESM still has to communicate with DART through restart files.
Can we make it better?

- Currently, CESM and DART run as separate executables
- In order to shrink the time taken up by CESM initialization, we plan to incorporate data assimilation as part of CESM.
- CESM currently recognizes seven component types (atm, lnd, ocn, ice, glc, rof, and wav).
- External processing, such as data assimilation will be an eighth component type for CESM, dubbed External System Processing (ESP).
- However, CESM still has to communicate with DART through restart files.
• We call this new workflow pause-resume.
• However, a model can only take advantage of data assimilation if it implements pause-resume.
We call this new workflow pause-resume.

However, a model can only take advantage of data assimilation if it implements pause-resume.
Implementing pause-resume in a model

Model without pause-resume

<table>
<thead>
<tr>
<th>Model Initialization (namelist, data structures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Initial Data  ----- or -----  Read Restart Data</td>
</tr>
<tr>
<td>Model Run</td>
</tr>
<tr>
<td>Return to Driver</td>
</tr>
</tbody>
</table>

CAM initialization and restart does *not* follow this pattern. Brian Eaton is modifying CAM to meet these requirements.
Implementing pause-resume in a model

Model with pause-resume

Model Initialization (namelist, data structures)

Read Initial Data ----- or ----- Read Restart Data

Model Run

Write Restart Files

Return to Driver

Pause

Resume

Read Restart Data (possibly modified by DART)

Model Run

Write Restart Files

Return to Driver

Pause

CAM initialization and restart does not follow this pattern.
Brian Eaton is modifying CAM to meet these requirements.
Current targets (this year)

- Implement CESM with stub ESP model
- Implement data ESP model to test pause-resume functionality
- Implement DART ESP component and perform data assimilation with an active ocean component.
- Implement pause-resume in CAM for data assimilation experiments.
Future plans

- Implement pause-resume capability in more models (e.g., CLM)
- Find more efficient way to communicate restart data to/from DART
- Use pause-resume to implement fault resilience
Towards fault resilience in CESM

- If a fault is detected (e.g., a filesystem glitch) and the job PEs are okay to keep running then CESM could 'resume' from a previous known good state.
- We call this feature 'rewind'
- Sheri Mickelson is conducting research on the first two items
- Note that for this to work, all active components will need to support pause-resume
Thanks!

Questions?

Contact: goldy@ucar.edu