CSEG Update

Mariana Vertenstein
CESM Software Engineering Group
• CIME Update
• Data Assimilation in CESM
• New Efforts in Modularity
• Future Concurrency
CIME is being used to facilitate external collaborations

Infrastructure
PUBLIC Open Source Github Repository

Paradigm for DOE, NOAA, NSF Infrastructure Collaborations

Driver-Coupler Data Models Scripts System/Unit testing Mapping Utilities

Science code Restricted or Public Repositories

CESM

DOE/ACME

ESPC and/or NOAA/NEMS

ACME/CSEG has established a very successful informal collaboration
• **Major pythonization refactor effort**
  – all CIME5 scripting/testing infrastructure is now in python
  – Joint ACME/CSEG collaboration

• **CIME5 is no longer CESM specific**
  – Share code where appropriate
  – Model specific configurability is provided
    • (e.g. $CIMEROOT/cime_config/cesm....)
  – Each prognostic component determines their own stand-alone configurations and pe-layouts

• **Both ACME and CESM will adopt CIME5**
  – Target is July
Modularity of coupling infrastructure permits users to easily switch prognostic/data components.

Inter-component feedbacks are easily activated and deactivated – critical for model development of target component.

Note that all boxes are in CIME.
Introduction of alternative NUOPC driver-mediator in CIME will permit coupling to new prognostic components.

The inter-agency National Earth System Prediction Capability project (ESPC) would like to couple to CESM components.
Data Assimilation in CESM
New External System Processing Component
Upcoming Pause/Resume Capability
CIME coupling infrastructure has unique multiple instance capability for Kalman-Filter DART data assimilation.
ESP component will allow data assimilation — DART — to interact directly with component data via I/O files.

More Details on how ESP works

External tools can easily interact with CESM during simulation.
New Efforts in Concurrency
Why prioritize modularity?

Modularity: the separation of each aspect of model functionality with self-contained code structures

• New science can be used in multiple models – removes duplication of effort
• Should not need to know about all the model to modify a small part of it
• Supports multiple alternatives for a parameterization
• Supports a hierarchy of model complexity - use certain processes, exclude others
• Supports testing individual pieces in isolation (unit testing)
• Enables addition of new concurrency
MARBL: Marine Biogeochemistry Library

\[ \frac{\partial x}{\partial t} + \nabla \cdot (u x) - \nabla \cdot (K \cdot \nabla x) = B_x(x) \]

- DOE funded effort
- Enables portability to alternative physical frameworks
  - Implemented in POP – part of CESM2
  - Being implemented in MPAS-O
- Enables question dependent configurations

Old Model

Base model (POP2)

BGC Driver

BGC Model

\[ \text{ecosys nutrient/carbon} \]
\[ \text{runtime param.} \]
\[ \text{initialization} \]
\[ \text{forcing} \]
\[ \text{output} \]
\[ \text{optional tracer} \]

\[ \text{share} \]

New Marbl-ized verison

CSEG Update 21 June 2016
CVMix: Community Vertical Mixing Library

- **Robust and flexible library for parameterizing ocean vertical mixing processes**
- **Developed within a community of scientists who make use of CVMix modules for a variety of research needs**
- **Ability to be used stand-alone or as part of an ocean model**
- **Reduces duplicate code – e.g. static mixing occurs as a step in many parameterizations**

CVMix used in:
- POP (NCAR)
- MPAS-O (LANL)
- MOM6 (GFDL)
**FATES: Functionally-Assembled Terrestrial Ecosystem Simulator**

- Core NGEE Tropics (DOE) demographic vegetation model
- Modular interface that can be run within multiple host land models (CLM, ALM)
- Built on the CLM(ED) representation that combines ED structured vegetation representation with CLM biophysics
- Future goals:
  - Embed fine-scale laterally-resolved hydrology model to determine roles of subsurface heterogeneity and connectivity on ecosystem structure and function
  - Plan to incorporate gross land-use transitions via CLM/ALM subgrid hierarchy as transfer between distinct land-use columns

<table>
<thead>
<tr>
<th>FATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>° Photosynthesis</td>
</tr>
<tr>
<td>° Growth &amp; Mortality of cohorts</td>
</tr>
<tr>
<td>° Disturbance and patch aging</td>
</tr>
<tr>
<td>° Water and Energy exchange with atmosphere</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physics interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>(moisture, energy exchange)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BGC interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>(litterfall, nutrient uptake)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Host Land Model (CLM, ALM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>° Soil Moisture and Temperature</td>
</tr>
<tr>
<td>° Decomposition and nutrient cycling</td>
</tr>
<tr>
<td>° Lakes, snow, ice, subgrid structure, etc</td>
</tr>
</tbody>
</table>
CLM Modularity Example: Simplifying Adding New Science

- Wood products are split 10 and 100-year lifetime pools
- Old code: 12 modules had information about these two pools
  - Adding another pool, or changing the lifetime of a pool required changes to all 12 modules
- New code: Only 2 modules have information about pool
  - main science module and a module that reads parameters from a file
- Result: easy to now add a pool or change the definition of the pools - all changes will be in the single science module
New Efforts in Concurrency
Inter-component concurrency

Concurrent Layout

- Driver
- CPL (regridding, merging)
- CAM
- CLM
- POP
- CICE
- MOSART
- WW3
- CPL

Sequential Layout

- Driver (controls time evolution)
- CPL (regridding, merging)
- CAM
- CLM
- CICE
- POP

Hybrid Sequential/Concurrent Layout

- Driver
- CAM
- POP
- CPL
- CLM
- CICE

CSEG Update 21 June 2016
For intra component concurrency need:

- **Driver**
  - control time evolution of components, calls coupling infrastructure
- **Coupling infrastructure**
  - carries out redistribution and regridding
- **Modularity**
  - Components must be self contained modular units with APIs