CSEG Update

Mariana Vertenstein
CESM Software Engineering Group

19 JUNE 2019
Outline

• Status of releases and CMIP6 Experiments
• Improvements in the CESM model development process
• Progress in PIO
• Infrastructure efforts
  – CMEPS: (inter-component coupling)
  – LILAC/CTSM (land modeling)
  – CPF (atmosphere physics)
• New workflow effort targeting Earth System prediction
Releases and CMIP6 Simulations
CESM2.1.z Releases and CMIP6 Experimental Status

• CESM2.1.0 and CESM2.1.1 Releases
  – Incremental releases of CESM2 that further expand the available set of out-of-the-box configurations of the model for readily performing CMIP6 simulations
    http://www.cesm.ucar.edu/models/cesm2/

• CMIP6 Experimental Status
  – Since last August, have run 900 different CESM cases for CMIP6. Have about 150 left to run.
  – Generated about 1 PB of compressed (lossless) time series files.
  – Published about 230 TB of compressed CMIP6 files to ESGF from over 500 CESM cases. Have about 143 TB waiting to be published
The creation script talks to the Experiment Database to get case details.

Case simulation’s progress and status is updated in the Experiment Database after each job finishes.

Manual publication is triggered through the Experiment Database.
Data factory

Data warehouse

Data delivery services

CESM data & metadata flow

Data headquarters

Dataset metadata

Datasets list

Other NCAR data sets

Dataset metadata

CESM2 Experiments Database

Scientifically Validated Configurations

Data Google (DASH Search)

Data consumers

archive_metadata

Other apps/websites?
Improvements in CESM Model Development
CESM2.1.z Model Development Improvements

- **Migration to GitHub**
  - All components in GitHub except for CAM (scheduled for this summer)
  - GitHub has dramatically improved interaction with the community
    - issue tracking and code reviews lead to a more robust code base and increased transparency

- **Testing**
  - New system and unit test capabilities
  - New use of continuous integration tools such as Travis CI

- **Handling complexity of repositories via manage_externals**
  - New manage_externals capabilities
  - Keep git submodule functionality if that is desired
  - NOAA UFS and possibly NASA is adopting manage_externals

- **Encourage community to contribute to the software development via GitHub!**
Progress in PIO
PIO Status and Upcoming Developments

- **PIO1**
  - is default in cesm2.1.z releases

- **PIO2**
  - will be default in cesm2.2 series
  - plan is for PIO2 to become part of NetCDF in fall 2019

- **PIO2 new capabilities**
  - New asynchronous write capability that will be leveraged in CMEPS
  - New data rearrangement strategies lead to improved scaling
Community Mediator for Earth Predictive Systems (CMEPS)

(Rocky Dunlap’s talk will give more details)
CMEPS enables new needed functionality and collaborations

**NUOPC** is a layer on top of ESMF that simplifies the creation of coupled modeling systems.
Status of CMEPS

• What can we now run in CMEPS:
  – **CESM:** prognostic components (including MOM6) - validation is underway
  – **UFS:** seasonal forecast system (FV3GFS/MOM6/CICE)
  – Swap data models for prognostic components in above

• Benefits to CESM:
  – On-line regridding
  – Run time sequencing via simply text file
  – New coupling strategies – use of exchange grid for atm/ocn flux calculation in mediator
  – Unification of component caps across models (MOM6 done – CICE next)
  – Easily bring in new versions of models that have NUOPC caps (e.g. WW3)
  – CMEPS targets CESM, NOAA/EMC, and GFDL needs – facilitates new collaborations
New Land Model Infrastructure Developments: CTSM and LILAC
LILAC coupling architecture
Light Weight Infrastructure for Land-Atm Coupling

alternative coupling of CTSM directly into host atmosphere

CESM hub and spoke architecture

Lilac architecture

LILAC uses ESMF as the underlying coupling infrastructure
LILAC coupler: initial target atmosphere models

- **WRF**: Weather Research and Forecasting model
- **SAM**: CSU's System for Atmospheric Modeling
- **COSMO**: ETH-Zurich’s coupled biosphere-atmosphere Regional Climate Model
- **MPAS-A**: Model for Prediction Across Scales – Atmosphere
- **RegCM**: Regional climate model originally developed at ICTP
- **CESM DATM**: Data atmosphere for testing and verification
The Community Terrestrial Systems Model

- A unified model for research and prediction in climate, weather, water, and ecosystems
- Will permit a hierarchy of complexity
- NWP, climate, water, and ecology applications
- Bringing in many Noah-MP features into CTSM
- Major refactoring underway to
  - modularize processes in order to enable and accelerate user-friendly science development
  - reduce accumulated technical debt
New Atmospheric Community Physics Framework (CPF) and New CAM Infrastructure

(see Steve Goldhaber’s talk for more details)
CPF Framework Update

• CPF based on new model metadata standards that
  – auto-generates host model interface to physics
  – auto-generates the code to run a suite of physics schemes
  – facilitates creation of flexible host model interface

• NCAR implementations underway
  – MICM (chemistry)
  – WRF physics in MPAS
  – CAM physics (soon)

• Formal governance agreement between NCAR and NOAA
  – CPF (also known as CCPP Framework at NOAA) repository is jointly managed by NOAA and NCAR
  – NOAA has accepted additional requirements from NCAR
New Workflow Effort Targeting Earth System Prediction
Next Steps in CIME-Compatible Workflows

• Need new end-to-end capabilities for earth system prediction
• Improved flexibility in defining workflows will be in cesm2.1.2
  – create_newcase has new optional workflow option
     
     ./create_newcase .... –workflow-case default
  – new CIME xml file, config_workflow.xml, extends the original implementation of workflow job requirements in config_batch.xml
  – key requirement will be to flexibly interface with external workflow engines such as Cylc
• Will handle requirements for CESM2 large ensemble simulations starting in early Fall
• Will handle new seasonal forecast end-to-end requirements for CESM/WACCM
Acknowledgements

CMIP6

Software Developers
Sheri Mickelson, Kevin Paul, Alice Bertini

Data Management and Advice
Gary Strand

Advice and Guidance
John Dennis, Mariana Vertenstein

Running of the Experiments
Cecile Hannay, Peter Lawrence, Michael Mills, Simone Tilmes, Robert Tomas, Brian Medeiros, Lantao Sun, Keith Lindsay, Kate Thayer-Calder, John Fasullo

Publication and Documentation
Eric Nienhouse, Adam Phillips

CSEG Members and Contributors

Atmosphere
Brian Eaton, Cheryl Craig

Ocean
Alper Altuntas

Sea Ice
David Bailey

Land
Erik Kluzek, Negin Sobhani, Bill Sacks

Land-Ice
Kate Thayer-Calder, Bill Sacks

Paleo
Mathew Rothstein

Workflow/Releases/Testing
Alice Bertini, Chris Fischer, Sheri Mickelson

CIME-CCS/Performance/Porting/ESMF
Rocky Dunlap, Jim Edwards, Robert Oehmke, Bill Sacks, Ufuk Turungoglu
Thank you!
EXTRA SLIDES
CESM2 Experiments Database
https://csegweb.cgd.ucar.edu/expdb2.0

- **Mysql database stores:**
  - Case metadata including XML, namelists and source mods
  - Process workflow status via cylc and archive_metadata
  - Diagnostics links
  - Dataset publication links (ESGF, CDG, glade, HPSS, DASH)
  - CMIP6 data request fields

- **Scientists and developers interfaces - login required**
  - Web enable dashboard at https://csegweb.cgd.ucar.edu/expdb2.0
  - Automated email parser for cylc workflow updates
  - Caseroot script archive_metadata for “push” and “pull” access via HTTP

- **Public interface**
  - Scientifically Validated Configurations at https://csegweb.cgd.ucar.edu/exp2-public
What default options do we want for the initial NWP configuration? #456

@barlage @dlawrenncar @martynpclark - can you please use this issue to list the options we want for the initial NWP configuration? My impression is that we're starting with clm5_0 but then changing some options. What I know so far is that we want SP with a few things changed:

- Surface dataset with only one PFT per grid cell
- 4 soil layers
- turn off PHS
- turn off BVOCs (if they're on by default with SP)

What else do we want to change?
Transparency via projects
Collaboration via Pull Requests

I have spotted this piece of code that seems wasteful. We could remove `dewmx` and change

```plaintext
fwet(p) = ((dewmx/veg)+h2ocan)*0.66666666666666666666_r8
and
fcansno(p) = ((dewmx / (veg * params_inst%sno_stor_max * 10.0_r8)) * snocan(p))*0.15_r8
```
to

```plaintext
fwet(p) = (h2ocan / (params_inst%dewmx * veg)))*0.66666666666666666666_r8
and
fcansno(p) = (snocan(p) / (params_inst%dewmx * veg * 
params_inst%sno_stor_max * 10.0_r8))*0.15_r8
```

These changes would not be BFB. Should I add comments in the code for a future clean-up?

I'm confused: your new versions don't seem to have `dewmx` in them at all. Is that a