CESM Unified Postprocessing and Diagnostics (CUPiD)

Teagan King and Mike Levy

Mar 4th, 2024 - SEWG Meeting

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

GENERAL THOUGHTS FOR FLOW OF PRESENTATION - DELETE SLIDE LATER

- Mike [5 min]
  - ESDS Background
  - Vision of project / anticipated milestones
- Teagan [10 min]
  - Where we're at
  - A few things more relevant to SEWG:
    i. Specific framework-related things we're addressing
    ii. GeoCAT collaboration thoughts
    iii. Collaborative work time thoughts?
    iv. Any ways we may be able to improve CUPiD or strengthen collaborations with help from folks in this meeting?

About ESDS (Earth System Data Science)

Collaboration and grassroots effort started in Fall 2020 among NCAR/UCAR labs, centers, programs, and offices (LCPOs) focused on improving Earth system data science

- Led by a committee with representation from CGD, CISL, EOL, MMM, RAL, and Unidata
- Has been integral in the “push to python” (helping users transition from NCL to open source python packages)
- Community-building and creating a culture of open science and development

CESM Diagnostics and ESDS

- CESM users were hitting the limits of what NCL-based analysis could handle (and NCL is no longer being developed)
  - Advantage to making sure everyone replaced NCL with the same product
  - ESDS helped community find a solution
    - Pangeo stack: numpy, xarray, matplotlib, dask, etc
- Still had several independent efforts for analyzing CESM output
  - ADF, NBscuid, individuals writing one-off notebooks
- Next step: build a common framework
  - **CESM Unified Postprocessing and Diagnostics (CUPiD)**

ADF

CUPiD Collaborators

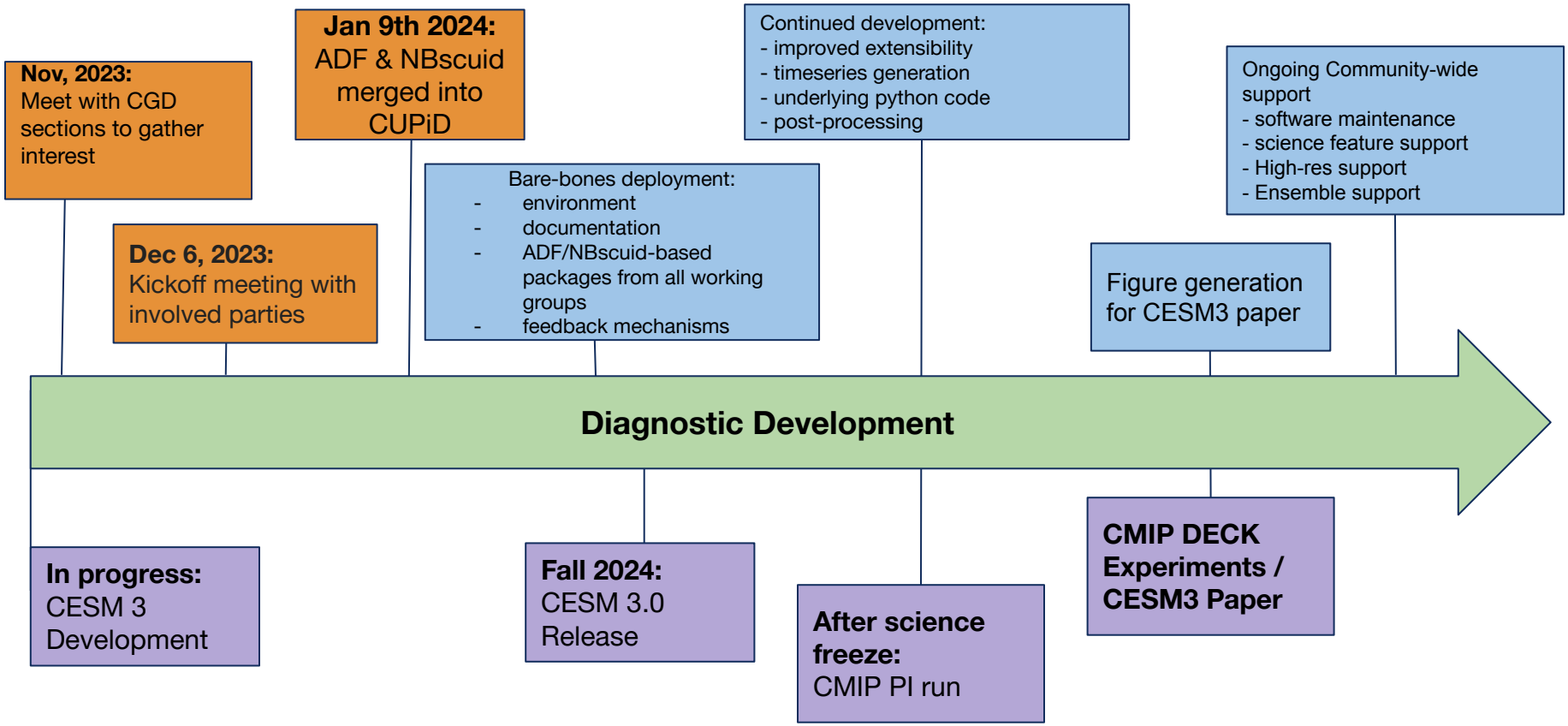
- **AMP:** Dani Coleman, Cecile Hannay, Brian Medeiros, Christina McCluskey, Jesse Nusbaumer, Justin Richling
- **CAS:** John Fasullo, Adam Phillips, Isla Simpson
- **CCR:** Gary Strand
- **CSEG:** Brian Dobbins
- **CESM:** Dave Lawrence
- **ESDS:** Katie Dagon, Teagan King, Mike Levy
- **ESMF:** Bill Sacks
- **GeoCAT** (CISL)**:** Orhan Eroglu, Katelyn FitzGerald, Anissa Zacharias
- **OS:** Anna Deppenmeier, Gustavo Marques, Lev Romashkov
- **PPC:** Dave Bailey, Kate Thayer-Calder
- **TSS:** Sam Levis, Will Wieder

CUPiD Project Vision

A collaborative effort that unifies all CESM component diagnostics and provides

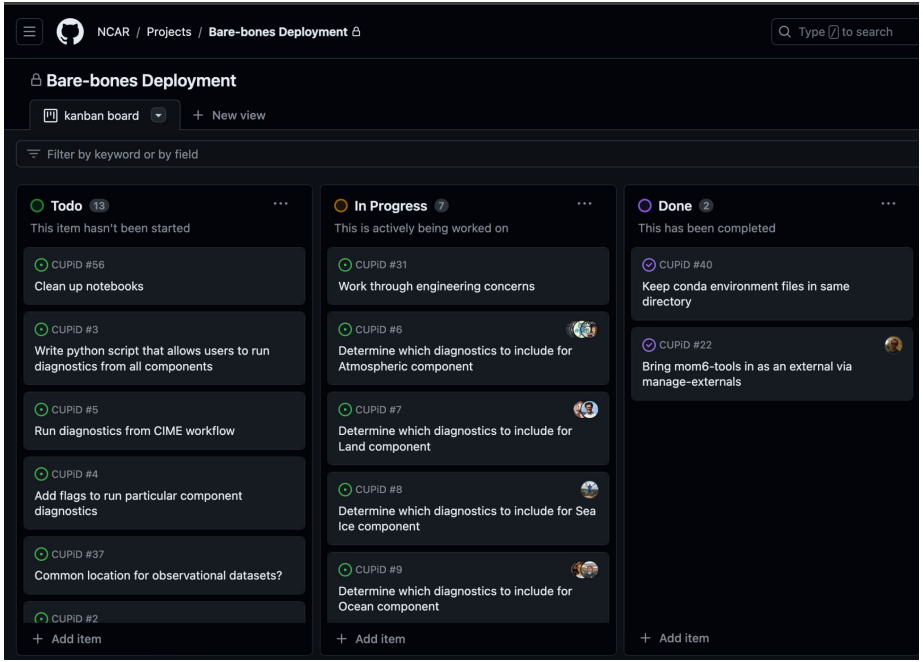
- Python code that
  i. runs in an easy-to-generate conda environment, and
  ii. can be launched via CIME workflow or independently
- Diagnostics for single/multiple runs and single/multiple components
- Ability to call post-processing tools that other groups are working on
- An API that makes it easy to include outside code
- Ongoing support and software maintenance

Anticipated Timeline



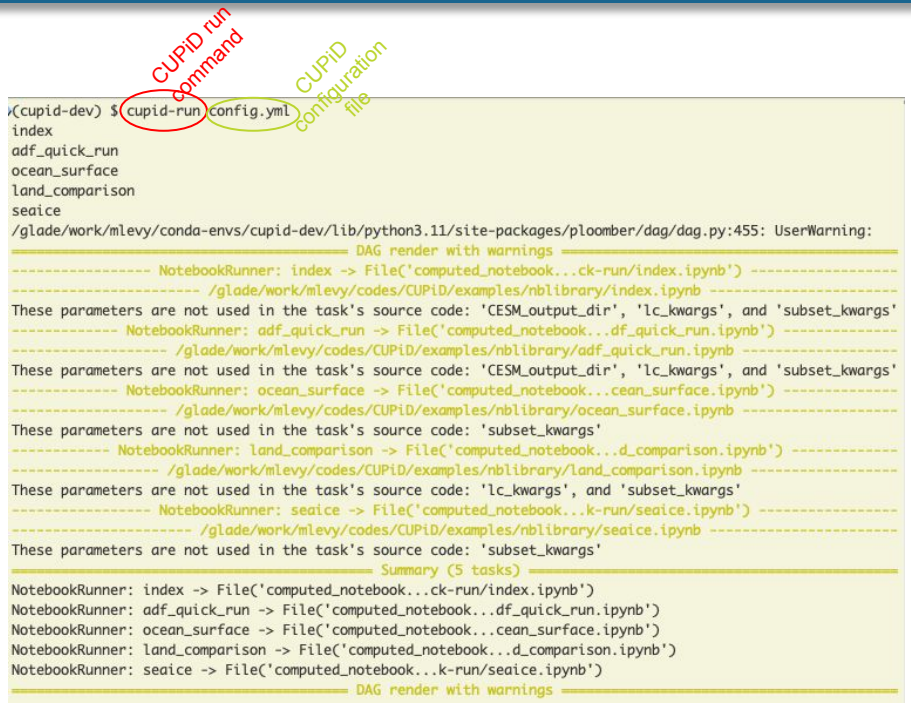
CUPiD Project Organization

- Open development on github (under NCAR organization)
  - Documentation on github.io
  - Fork / branch / PR strategy
- Issues and Project Board



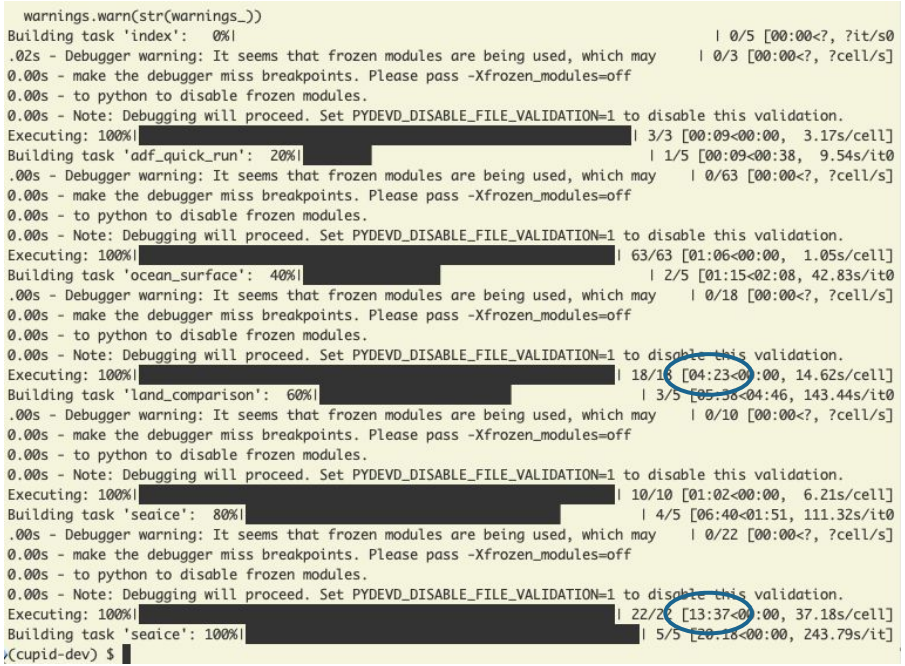
Current State of CUPiD

- Single command line call to produce basic diagnostics for atmosphere, ocean, land, and sea-ice
  - User-editable YAML file controls what is run
  - Uses ploomber (which wraps papermill) to run notebooks and pass parameters



Current State of CUPiD

- Each notebook is run sequentially
- Runtime is fairly variable
- Notebooks can use dask LocalCluster for parallelization
  - Ocean notebook ~10 minutes without dask
    - With dask: 4:23 here, 3 minute best
  - Seaice notebook ~25 minutes without dask
    - With dask: 13:37 here, 9 minute best

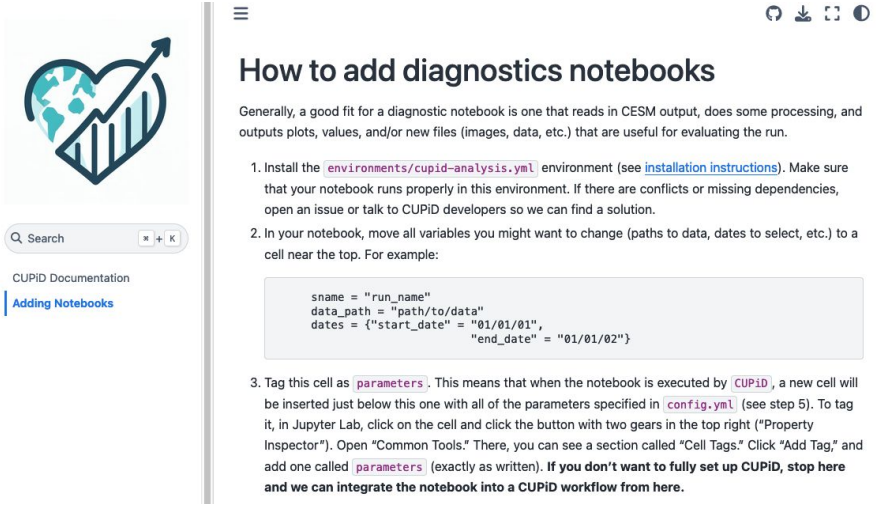


Current State of CUPiD

- CUPiD provides two environments
  - cupid-run executable in cupid-dev (plan: have CISL add to NPL)
  - All notebooks run in cupid-analysis
    - Build on pangeo stack
    - Uses geoCAT packages
      - Active migration of functions to GeoCAT

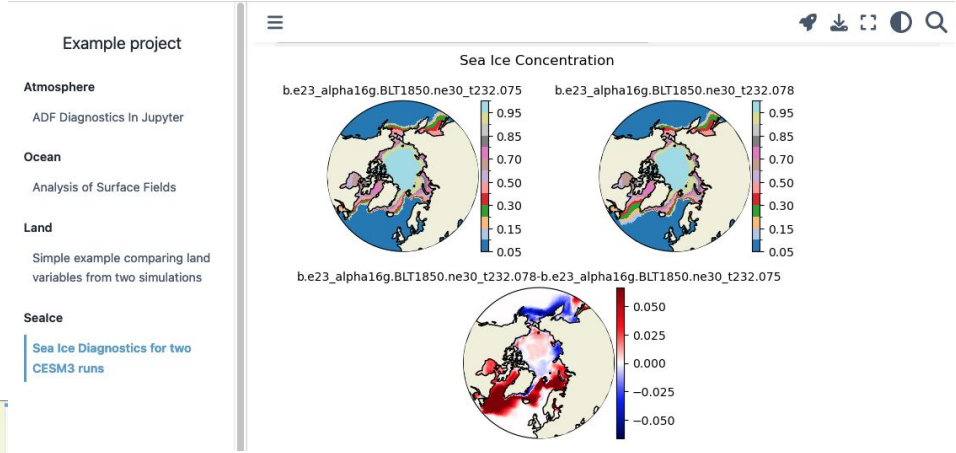
Current State of CUPiD

- Step-by-step instructions for adding new diagnostics in this framework are available online



CUPiD Output

- CUPiD can use JupyterBook to turn notebooks into web pages
  - cupid-build function name likely to change




```
(cupid-dev) $ cupid-build config.yml
Running Jupyter-Book v0.15.1
Source Folder: /glade/work/mlevy/codes/CUPiD/examples/coupled_model/computed_notebooks/quick-run
Config Path: /glade/work/mlevy/codes/CUPiD/examples/coupled_model/computed_notebooks/quick-run/_config.yml
Output Path: /glade/work/mlevy/codes/CUPiD/examples/coupled_model/computed_notebooks/quick-run/_build/html
Running Sphinx v5.0.2
making output directory... done
```

Where we're headed

- Conda kernel available on Casper & Derecho
- Launch diagnostics from CIME workflow
- Port NCL-based code to python

Summary

- ESDS is a cross-laboratory initiative at NCAR
- CUPiD is an ESDS project focused on CESM3 (and beyond) diagnostics
- Still in early stages of development, but showing lots of promise