

**The RGMA CMIP6 Analysis Activity
and the
Coupled Earth System Model
Analytics Consortium (CESMAC)**

2020 RGMA Fest

Overview

- The RGMA CMIP6 analysis activity
- CESMAC: Coupled Earth System Model Analytics Consortium
- The status of the repository
- HighResMIP data
- Discussion: Future needs

The RGMA CMIP6 Analysis Activity

- In early 2019 we initiated the RGMA CMIP6 analysis activity
- Main goal: ***Facilitate CMIP6 analysis for the RGMA community***
- Here we will review the activity, and provide updates on the next phase

Organizing Committee:



Renu Joseph
(DOE)



Forrest Hoffman
(ORNL)



Paul Ullrich
(UC Davis)



Michael Wehner
(LBNL)




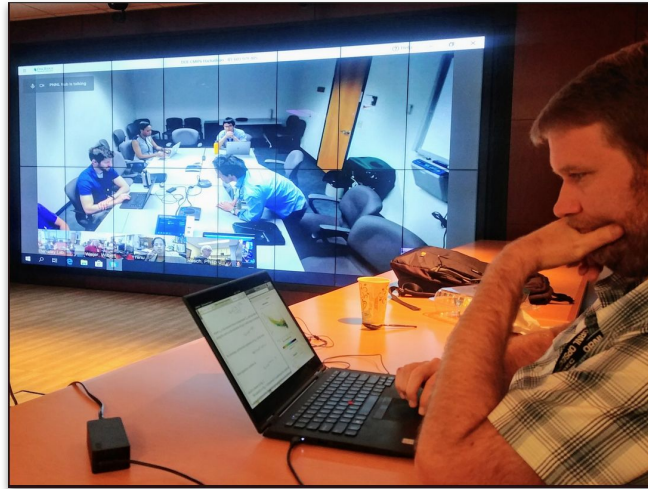
Wilbert Weijer
(LANL)

Goals

- **To build a common data repository**, accessible from a powerful analysis platform by a large number of scientists.
 - *We downloaded > 4 PB of CMIP6 data from ESGF nodes to NERSC storage.*
 - *We added ancillary data collections (including ana4mip, obs4mip, reanalyses).*
 - *We organized a CMIP6 tutorial to familiarize the community with CMIP6*
- **To build a common analysis environment**, capable of dealing with large data volumes.
 - *We organized a tutorial on (V)CDAT (led by Charles Doutriaux)*
- **To build a community of scientists**, collaborating towards the common goal of producing policy-relevant science.
 - *We organized the CMIP6 Hackathon*

CMIP6 Hackathon

- The activity culminated in the CMIP6 Hackathon
 - July 31 through August 6, 2019.
 - Participants at six hubs around the country.
 - Communication through videoconferencing, Slack.
 - Exchange analysis scripts through github.



Hackathon Speeds Progress Toward Climate Model Collaboration

Climate scientists collaborated in a nationwide event to analyze and compare archived Earth system model simulations and to generate input for the IPCC's upcoming climate change report.



By William Hladik, Forrest M. Hoffman, Paul A. Ullrich, Michael Watson, and Justin Liu

In summer 2019, scientists from the U.S. Department of Energy (DOE) gathered at six hubs across the United States to participate in a 20-hour model comparison “hackathon.” They profiled computing resources and capacities, and they collaborated to parse and fix interoperability. By joining forces, these scientists got results much faster, reduced duplication of efforts, and spent less time solving in-house problems than they would have had they worked on their own.

These findings will contribute to a reporting requirement set by the Environmental Protection Agency (EPA). This report reviews the state of climate change science, the impacts of socioeconomic implications, and identifies viable response strategies. The IPCC has published five assessment reports so far,

and the Sixth Assessment Report (AR6) is currently in preparation.

Analyses of the Earth system based on observational data from sensors on the ground, in the ocean, and in space form an important basis for these reports. Not only do they complement each other, but they also provide important complementary information because they enable insights into Earth environmental conditions and help address observed changes to specific causes.

Each model used here are inputs to create a new body of water data, atmospheric, and agricultural. Thus, the base model system of Earth’s climate changes when results from several models are compared, using sets of the strengths and limitations of each. However, this type of comparison poses challenges to individual researchers.



Exit Poll

- 37 Respondents
- Overall satisfaction with hackathon: 4.3 out of 5
- Only 1 respondent would *not* participate in future hackathons

***“...I do not think we are what the name [hackathon] suggests. Our hackathon felt more like supporting people to do the work they were already going to do and less like a massive collaborative effort. This is a good thing, just different.*”**

***“The camaraderie was really great.”*”**

***“This activity has definitely saved a lot of my time that I would have spent otherwise learning by myself.”*”**

***“... the hackathon served the purpose of professional development for many. I think this is a strength and should be part of the advertising.”*”**

Impact

- So far RGMA scientists reported **14** publications with help from this activity
 - **6** currently in review
 - But only **2** of these papers are mentioned in the draft AR6
- More than 130 RGMA scientists have access to 4 PB of CMIP6 data from a powerful analysis platform

CESMAC: Coupled Earth System Model Analytics Consortium

- Next phase of the RGMA CMIP6 analysis activity:

CESMAC: Coupled Earth System Model Analytics Consortium

- New NERSC Project led by Forrest
 - Active AY 2020 allocation
 - AY 2021 request submitted
- All data and *cmip6* file group users are now part of **m3522**
- Our data repository migrated to NERSC's new **Community File System (CFS)** at `/global/cfs/projectdirs/m3522/cmip6`
- Small computational allocation available

The Status of the Repository

- Our 4 PB drive is full
 - Downloads are temporarily suspended
- Forrest is moving files to HPSS
 - He will make database of files that were moved
 - Still accessible by *cmip6* group members
- Download will be resumed shortly
 - Filling is missing files for existing experiments
 - More downloads for ScenarioMIP
 - Individual requests, if you see that files are missing (but check that they are available on ESGF)

HighResMIP Data

- The HighResMIP data is very useful for studies of weather and climate phenomena at the highest resolutions available from CMIP6.
- Experiments include CMIP and AMIP-style simulations:
 - Time periods: historical (1950 to 2015) and future (2015 to 2050)
 - Two resolutions: mainstream ($\sim 1^\circ$) and high resolution ($\geq 0.25^\circ$)
- To facilitate tracking these phenomena, CASCADE has downloaded all required 1-6 hourly fields identified by WACCEM and CASCADE
 - These fields enable running TEMPEST, TECA, etc to track ARs, TCs, ETCs, etc. and computing ETCCDI's.
 - The data include ancillary data on surface and TOA fluxes, rainfall, etc.
- The data is union of output from PRIMAVERA models available from CEDA (UK), [esp. HadGEM output not in public domain] and non-PRIMAVERA models on ESGF.
- All downloads are finished except for final HadGEM retrievals from tape.
- Total holdings: ~ 525 TB, from 10 modeling groups:
 - CMCC, CNRM-CERFACS, EC-Earth, ECMWF, INM, IPSL MIROC, MOHC/NERC, MPI-M, and MRI
- Please contact Bill Collins to arrange access.

Discussion: Future Needs

- Are there future needs for coordinated activities?
 - Another CMIP6 hackathon?
 - Maybe focused on HighResMIP?
 - Tutorials on analysis tools?
 - (V)CDAT?
 - Pangeo?
 - Hackathon and/or tutorials on RGMA metrics packages?
 - CMEC, ILAMB, etc.
 - Machine Learning tools?
 - Webinar series presenting RGMA CMIP6 studies?