ESGF
Earth System Grid Federation

Update on ESGF and Model Benchmarking

10th United States Climate Modeling Summit (2–3 May 2024)

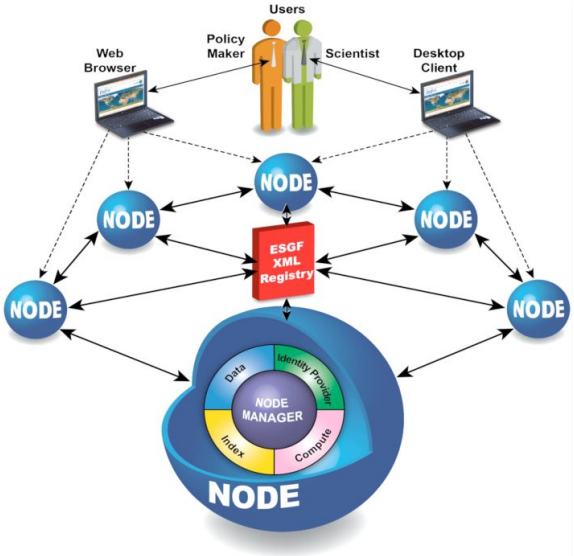
Forrest Hoffman, Oak Ridge National Laboratory

3 May 2024

What is the Earth System Grid Federation?

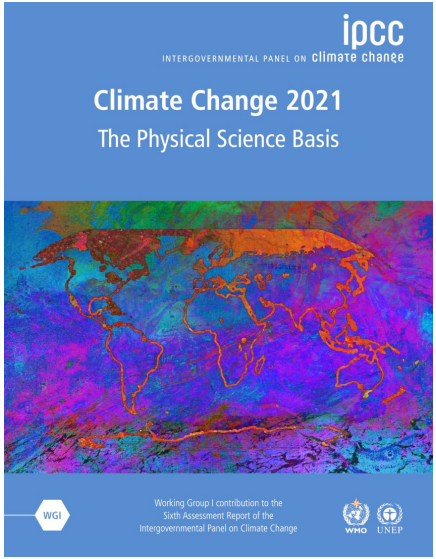
- **Earth System Grid Federation (ESGF)** is an *international consortium* and a *globally distributed peer-to-peer network of data servers* using a common set of protocols & interfaces to archive and distribute climate & Earth system model output and related input, observational, and reanalysis data
- **Open Science data** are used by scientists all over the world to investigate consequences of possible climate change scenarios and Earth system feedbacks

ESGF Conceptual Diagram



Model output data from ESGF are used for research that underpins IPCC Assessment Reports, like AR6

ESGF
Earth System Grid Federation



Logos represent primary international contributors: US Department of Energy, NASA, NOAA, NSF, European IS-ENES Project, and Australian NCI

ESGF Organization

ESGF Steering Committee

Justin Hnilo (Chair), DOE
Ben Evans, NCI
Adrienne Simonson, NOAA
Sylvie Joussaume, IPSL (ENES)
Tsengdar Lee, NASA
Forrest Hoffman (Ex-Officio), DOE/ORNL
Philip Kershaw (Ex-Officio), CEDA

ESGF Executive Committee

Forrest Hoffman (Co-Chair), DOE/ORNL
Philip Kershaw (Co-Chair), CEDA
Sasha Ames, DOE/LLNL
Rachana Ananthakrishnan, DOE/ANL
Laura Carriere, NASA
Ben Evans, NCI
Stephan Kindermann, DKRZ
Christian Pagé, CERFACS
Aparna Radhakrishnan, NOAA/GFDL

Tenth ESGF Conference

- Held last week in Rockville, Maryland
- John Dunne joined the meeting to share CMIP priorities and current timeline
- ~50 in-person attendees from 8 countries (Australia, France, Germany, Italy, Japan, Sweden, United Kingdom, USA)
- ~69 virtual registrants from 18 countries
- **Primary objectives** of conference were to
  - **Share all current development activities** across the Federation
  - **Develop a roadmap** for collaborative activities necessary to deploy operational ESGF infrastructure to support CMIP AR7 Fast Track

Conference Agenda

- Accomplished through a minimal set of plenary presentations and a **large amount of time addressing specific topics in breakout groups**
- Message from Jay Hnilo (DOE) and the ESGF Steering Committee
  - US-DOE is committed to continue helping deliver a solution for CMIP
  - The SC is committed to collaborating across institutions for building out innovative solutions
  - We have confidence in the assembled groups to build a modernized framework that is more available and resilient and that enables more community access and innovation
- Agenda is available at https://bit.ly/3vZlTsj

Tenth Earth System Grid Federation (ESGF) Hybrid Conference

April 23–26, 2024

Bethesda North Marriott Hotel & Conference Center
5701 Marinelli Road, Rockville, Maryland 20852, USA
Phone +1 301-822-9200

Agenda

This agenda is designed to facilitate collaborative interactions and creation of a roadmap for ESGF development activities aimed at modernizing the system for CMIP AR7 Fast Track simulation data. While important categories of development, deployment, and operations are included, the agenda may be changed when necessary to maintain productive use of face-to-face meeting time. Some open meeting times are available in breakout rooms. Plenary Sessions will be broadcast by video conference. Separate meeting rooms will be available for breakout sessions.

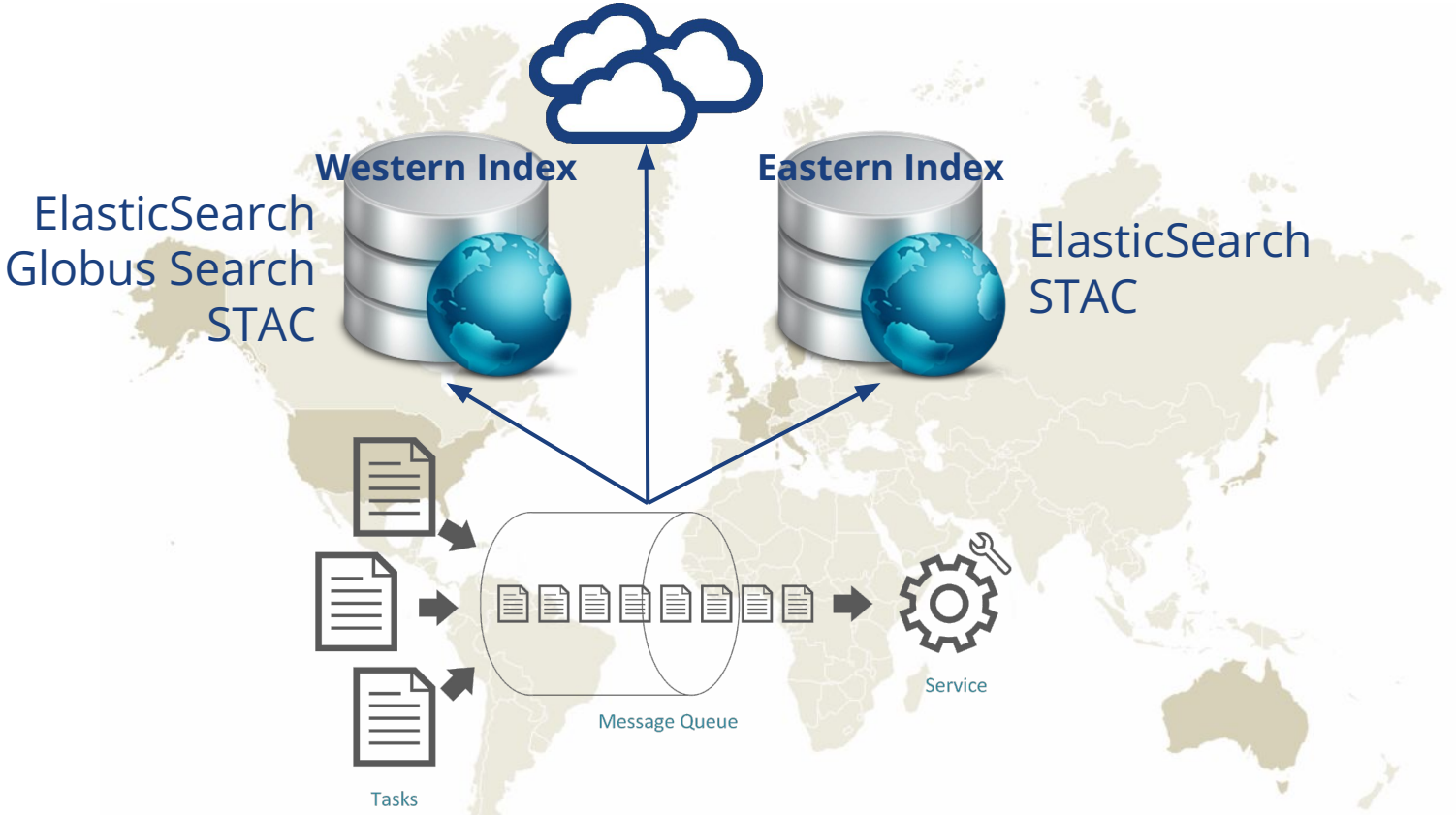
Plenary sessions will be held in (Grand Ballroom) Salon F&G on the Lobby Level of the hotel. Breakout Room 1 is in Salon H, and Breakout Room 2 is in Salon C, both also on the Lobby Level of the hotel. Side meetings can be held in the Seneca Boardroom, one level down from the meeting rooms.

Contacts: Taylor Clower-Mettica <clowertm@ornl.gov> and Forrest M. Hoffman <forrest@climatemodeling.org>

Tuesday, April 23, 2024

EDT	Description	Session Chair(s)	Location
7:30	ESGF Conference Registration Desk Opens		
8:00	Buffet Breakfast		
8:00	Networking Session		Plenary (F&G)
9:00	Session 1: Welcome and Introductions		Plenary (F&G)
9:00	Welcome to the ESGF Conference and Meeting Logistics	Gary Geernaert, Forrest Hoffman	
9:10	Introduction to the ESGF Steering Committee and Supporting Programs	Justin Hnilo, Tsengdar Lee, Sylvie Joussaume, Ben Evans, Adrienne Simonson	
9:20	Introduction to the ESGF Executive Committee	Forrest Hoffman, Philip Kershaw	
9:25	Introduction to the Coupled Model Intercomparison Project (CMIP), the CMIP International Program Office (CMIP-IPO), CMIP Task Teams & Fresh Eyes on CMIP, and the CMIP7 Timeline	John Dunne	
9:45	Introduction to the WGCM Infrastructure Panel (WIP)	Matthew Mizielinski, Paul Durack	
9:55	Overview of the CMIP Node Operations Team (CDNOT)	Sasha Ames	Plenary (F&G)
10:00	Charge for the meeting: Establish roadmap for CMIP7 readiness • ESGF Data Lifecycle: from Model to Analysis - https://docs.google.com/presentation/d/18kJXCShAl2RBrFeDBpzhF2Aiv08bXPbxyj-v1H6_pFE/edit	Phil Kershaw, Sasha Ames	Plenary (F&G)
10:30	Morning Break		
11:00	Charge for the meeting (Continued) • Present ESGF Roadmap for CMIP7 as it stands (25 min) • Outline session structure (5 min): - Charge - SMART objective (which must tie into the roadmap: what must be done, who will do it, what timeline, what are the dependencies) - Discussion time, work (up to three breakouts) - Report back, decision on next step i.e. how it fits into roadmap (including timings, e.g. task X will take 2 months starting in Oct 2024), what effort/resources needed and who will do it	Phil Kershaw, Forrest Hoffman	Plenary (F&G)
11:30	Session 2: Data Publication	Sasha Ames	Plenary (F&G)
11:40	Breakout Sessions		

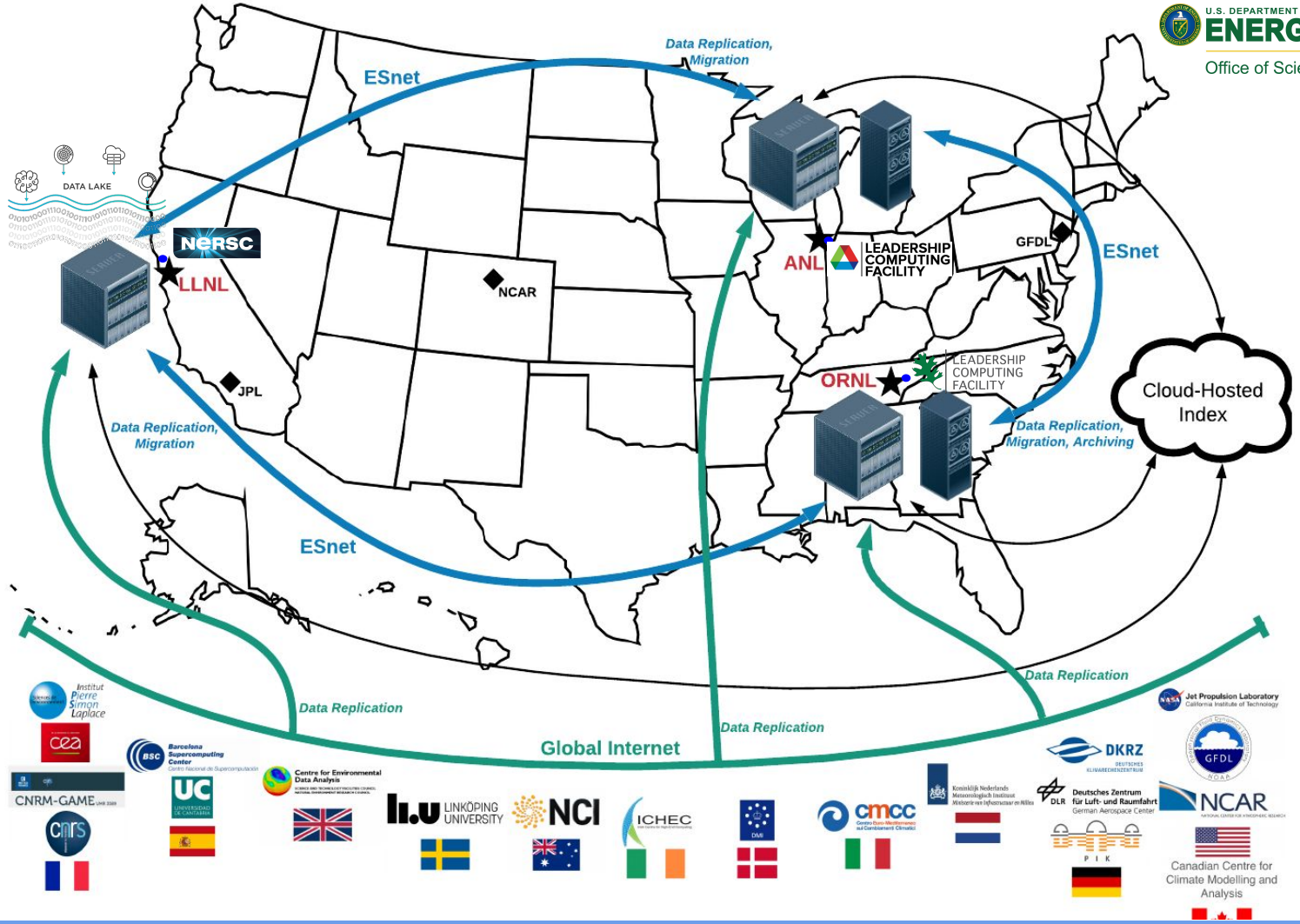
Major Accomplishment: New Index Strategy



Publication and Replication

DOE's Next Generation Earth System Grid Federation

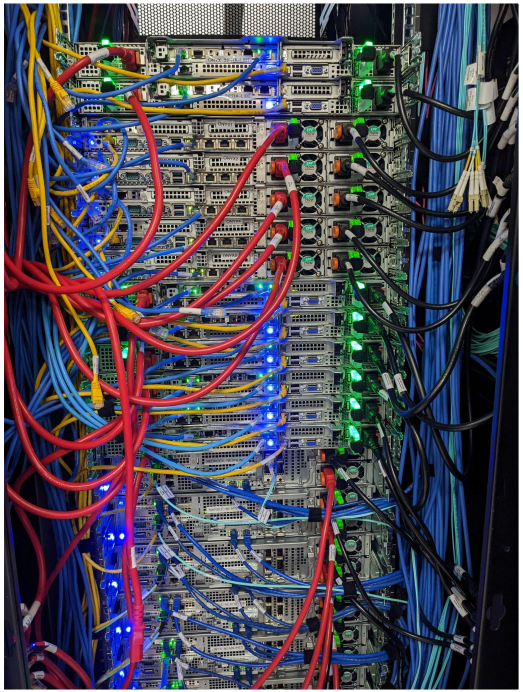
- As many as three nodes co-located at DOE's major computing facilities
- Replicating data from the global Federation
- Providing cloud indexing and tape archiving



ESGF2 US Data and Index Nodes Deployed at ORNL

- Containerized server software deployed on the shared Onyx cluster is serving 8 PB of Coupled Model Intercomparison Project (CMIP5 and CMIP6) data at ORNL
- Data are stored on the new Themis hierarchical storage platform, providing on-disk copy for fast access to frequently used data and backup copies on two tapes for all data
- Hardware investment at ORNL has been in storage capacity (fully operational)
  - 15 PB of disk
  - 30 PB of tape (for redundant backup)

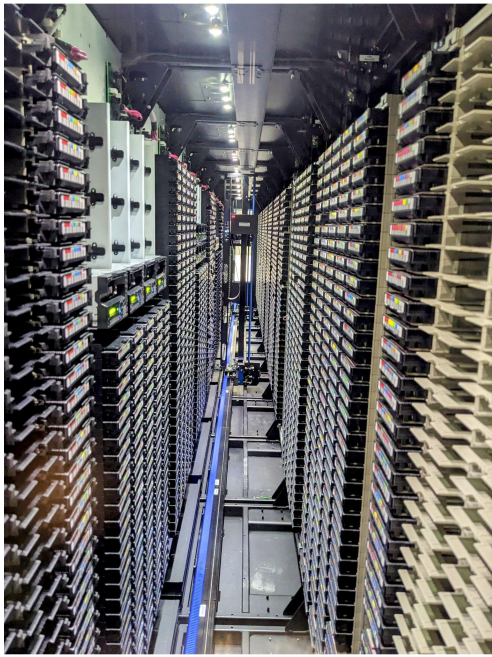
Delivered ahead of schedule and under budget!



The Onyx cluster hosts the ESGF containerized data & index nodes

Data and services reside in the Open Network Enclave of NCCS to provide fast and open access to data

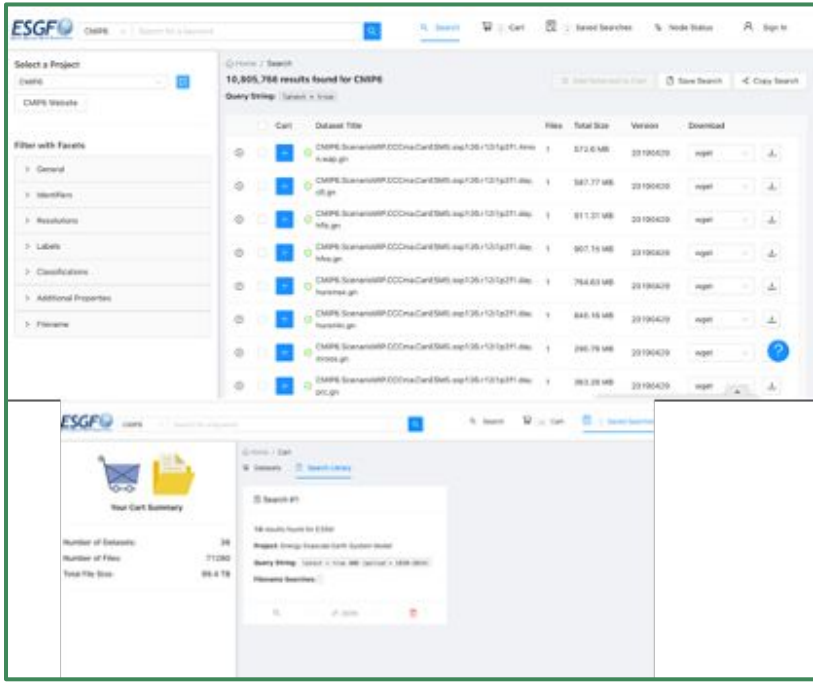
In partnership with the ORNL National Center for Computational Sciences (NCCS)



Expandable tape subsystem of the Themis storage system

ESGF2 US Metagrid Enhances ESGF Search

- New **Metagrid faceted search user interface**, developed at LLNL on popular React Javascript framework, deployed at ORNL, LLNL and ANL
- Offers new features, including a **shopping cart**, ability to **save and share searches**, integration with **Globus authentication & transfer** and a search page **tour & support dialog**
- User experience enhancements make it **faster and easier** to discover published data
- **Globus integration** offers faster and more reliable data access
- Will be deployed internationally across the Federation by mid-2024



The Metagrid Web Interface for ESGF search is a completely redesigned interface from CoG. It features a familiar faceted search and a new capability to save searches.

ESGF2 US Outreach Activities

- Organize **Webinars, Tutorials, and Bootcamps**
  - Data management lessons learned, ingest best practices
  - Data discovery and access, analysis frameworks and tools

➜ **ESGF Webinar series playlist at** https://www.youtube.com/@esgf2432

- **Hackathons and Workshops**
  - Data standards, data node deployment and user compute resources
  - Hold at large relevant conferences, e.g., AGU, EGU, AMS

➜ **Open ESGF Workshop at AGU 2022 (Chicago)**

➜ **Open ESGF Workshop & Tutorial at AGU 2023 (San Francisco)**

- Organize / host annual **ESGF Developer and User Conferences**

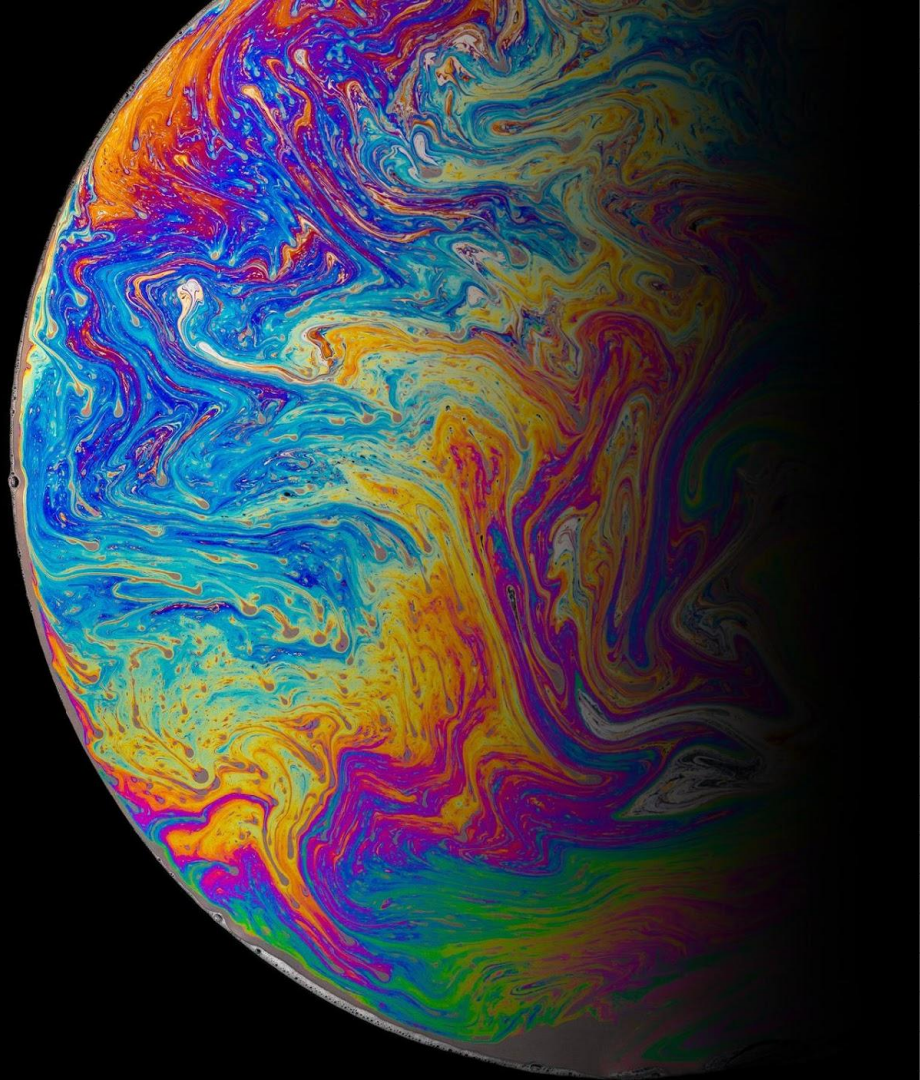
➜ **Ninth ESGF Developer and User Dual-Hybrid Conference was held January 18–20, 2023 at ORNL and Toulouse**

➜ **Tenth ESGF Developer and User Conference scheduled for Rockville, MD, on April 23–26, 2024**

Ninth ESGF Developer and User Conference, held jointly between Oak Ridge National Laboratory (USA) and Toulouse (France), January 18–20, 2023

CMIP Coupled Model Intercomparison Project
Model Benchmarking

Climate Model Benchmarking for CMIP7



Forrest M. Hoffman and Birgit Hassler
CMIP Climate Model Benchmarking Task Team Co-Leads

The Model Benchmarking TT

- Rebecca Beadling, *USA*
- Ed Blockley, *UK*
- Jiwoo Lee, *USA*
- Valerio Lembo, *Italy*
- Jared Lewis, *Australia*
- Jianhua Lu, *China*
- Luke Madaus, *USA*
- Elizaveta Malinina, *Canada*
- Brian Medeiros, *USA*
- Wilfried Pokam Mba, *Cameroon*
- Enrico Scoccimarro, *Italy*
- Ranjini Swaminathan, *UK*

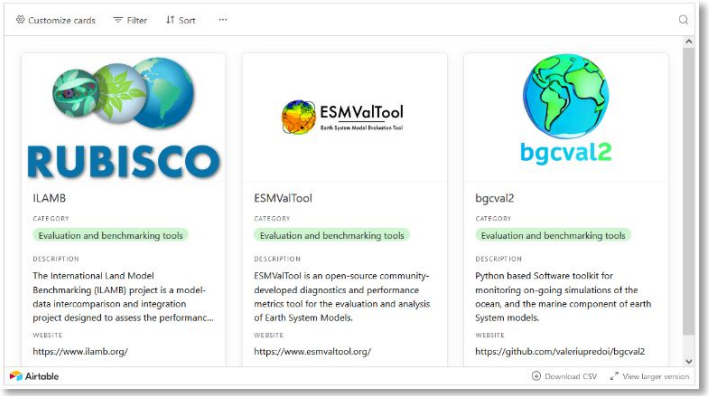
- Diversity in expertise (realms and methods), user group representation, gender, location, career stage
- Overarching goals:
  - Systematic and rapid performance assessment of the expected models participating in CMIP7 (including the model output and documentation)
  - Enhancing existing community evaluation tools that facilitate performance assessment of models
  - Integration of evaluation tools into CMIP publication workflows and fostering publication of their diagnostic outputs alongside the model output on the ESGF
- Collaboration with two Fresh Eyes on CMIP Subgroups
  - Model Evaluation
  - Data Analysis

#CMIP

Model Benchmarking Tools – Info “Cards” & Videos

- Main characteristics of (open source) benchmarking and evaluation tools available for analyses of CMIP-style data summarized in an overview “card” or an information video
- Collected information presented centrally on the CMIP website for easy access
- Cards can be filled out for all available open source benchmarking and evaluation tools if they can be used for CMIP data analysis; pre-defined questionnaire available on the CMIP website

https://wcrp-cmip.org/tools/model-benchmarking-and-evaluation-tools/



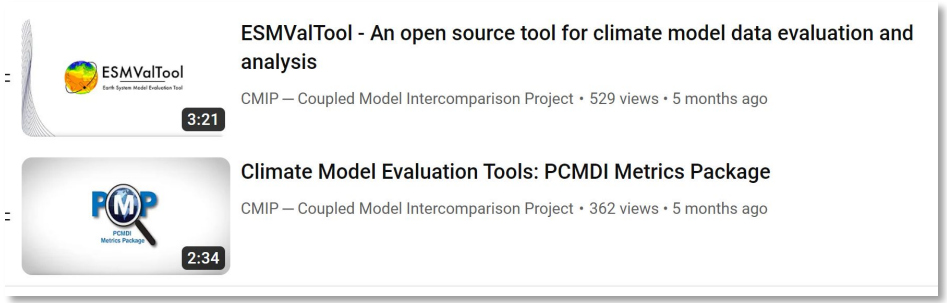
Status: first cards available

Started: October 2023

#CMIP

Model Benchmarking Tools – Information Videos

- Videos with descriptions of different benchmarking and evaluation tools
- Contain also the main characteristics of the different tools, just presented in a different way than the "cards"
- Videos can also be of different style
- All videos are presented in one central location linked to CMIP



- More videos of tools welcome!
- More info cards about evaluation/benchmarking tools welcome!

https://wcrp-cmip.org/tools/model-benchmarking-and-evaluation-tools/

#CMIP

CMIP Model Benchmarking | WCRP World Climate Research Programme

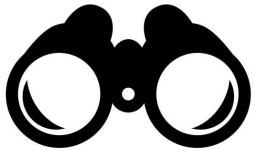
Retrospective paper

- Definitions of “evaluation”, “validation”, and “benchmarking”
- Retrospective look at evolution of evaluation & benchmarking metrics
- What tools were available for CMIP6 (methods, philosophies, tools)?
- What approaches were used for CMIP6?
- Which of them worked well for CMIP6 and what did not work for CMIP6?
- Extensive information about different benchmarking and evaluation tools

Status: Currently being finalized

Planned submission: May 2024

CMIP Model Benchmarking | WCRP World Climate Research Programme



What is the way forward?

- Based on the findings of the extensive information collected about different tools, and the retrospective paper – What do we think should be the benchmarking/evaluation focus for CMIP7?
- What framework would ideally be available for instantaneous benchmarking and evaluation at the time of data submission? Is such a framework even possible?
- How to avoid the bottlenecks encountered in CMIP6 benchmarking/ evaluation?
- Comprehensive community evaluation in near-real time possible (i.e., on ESGF)?

Status: Under development

Planned submission: Summer 2024

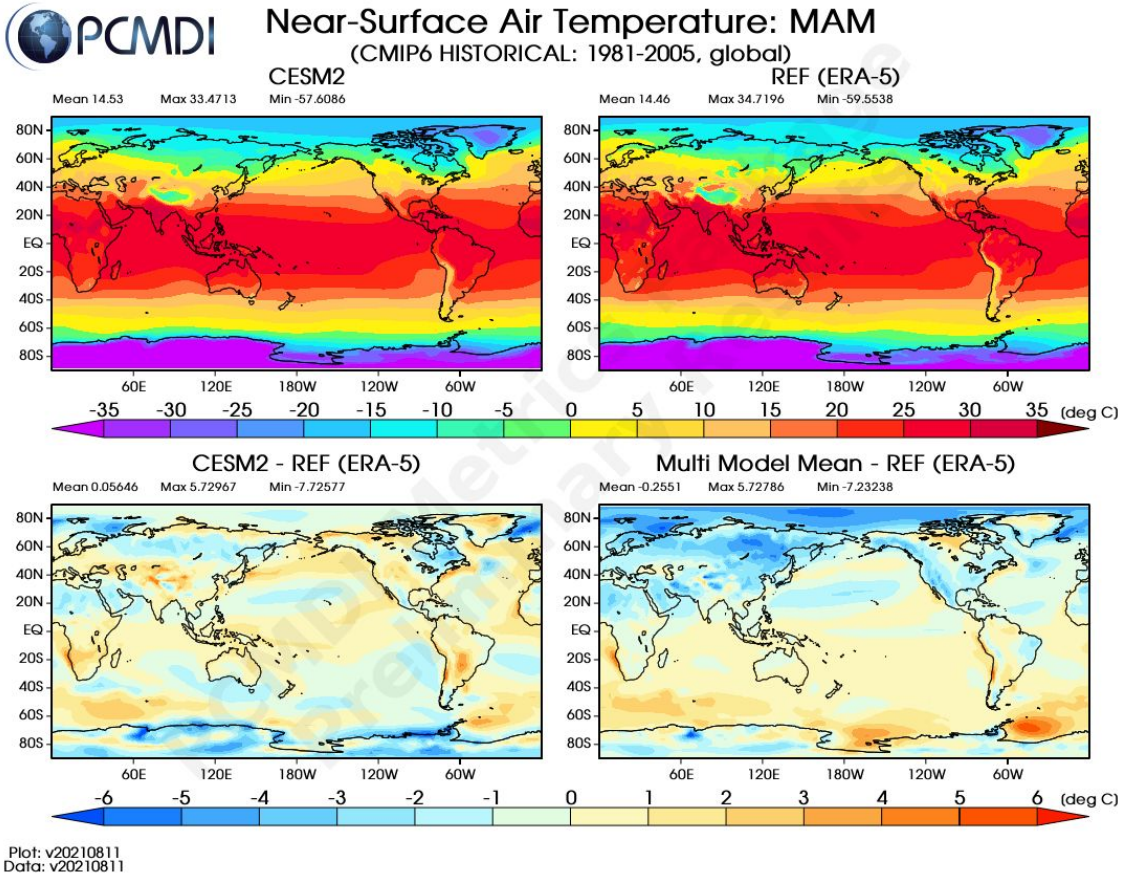
#CMIP

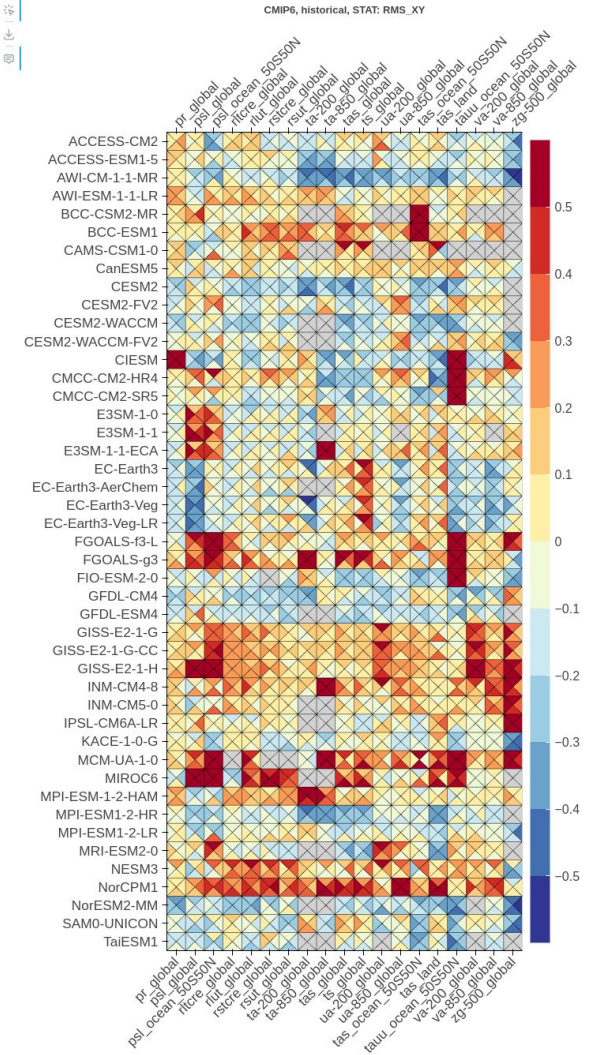
Other Planned TT Activities

- CMIP Model Benchmarking TT face-to-face meeting in May at the German Aerospace Center (DLR) to finish off the retrospective paper and establish vision for the perspectives paper
- In collaboration with Fresh Eyes on CMIP groups
  - Develop scope for better quality assurance / quality control (QA/QC) for CMIP model output
  - Develop a white paper on data needs for model benchmarking, including uncertainties
  - Develop a final report from the Model Benchmarking TT that proposes an architecture for automated benchmarking capabilities at the time of CMIP data publication

#CMIP

PCMDI Metrics Package





ESMValTool

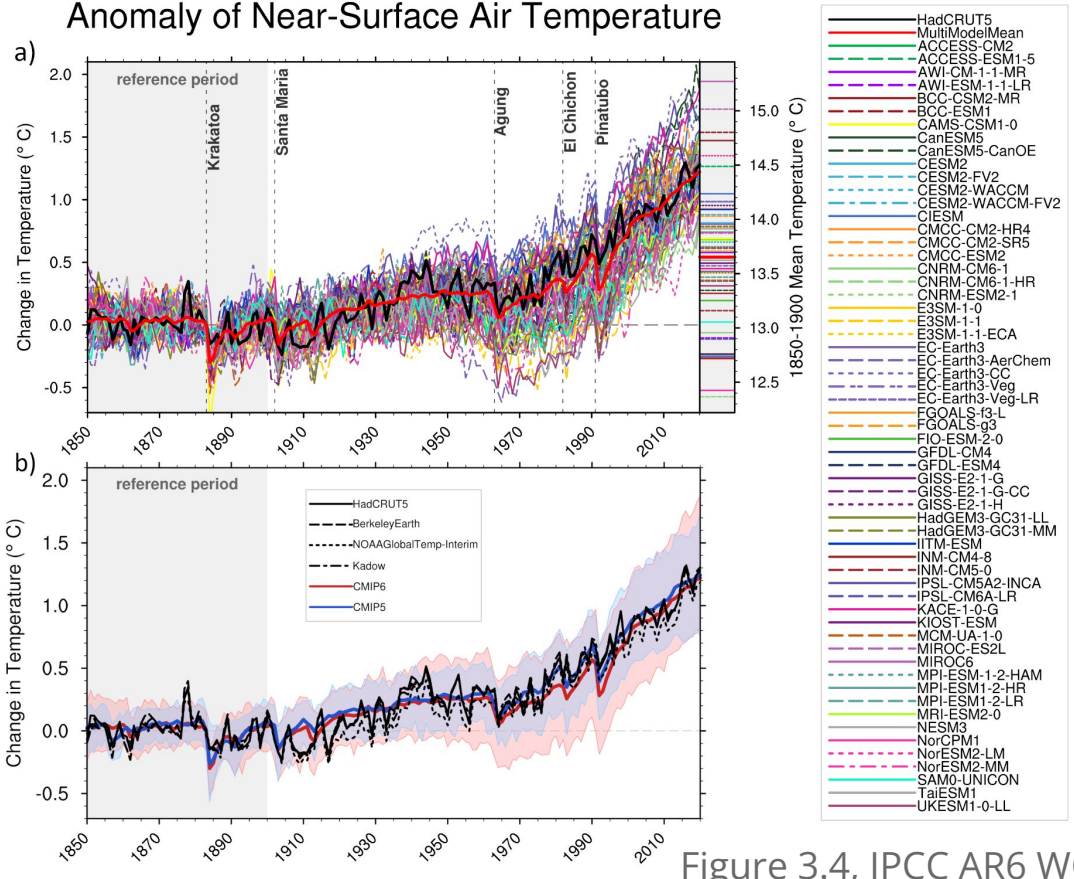
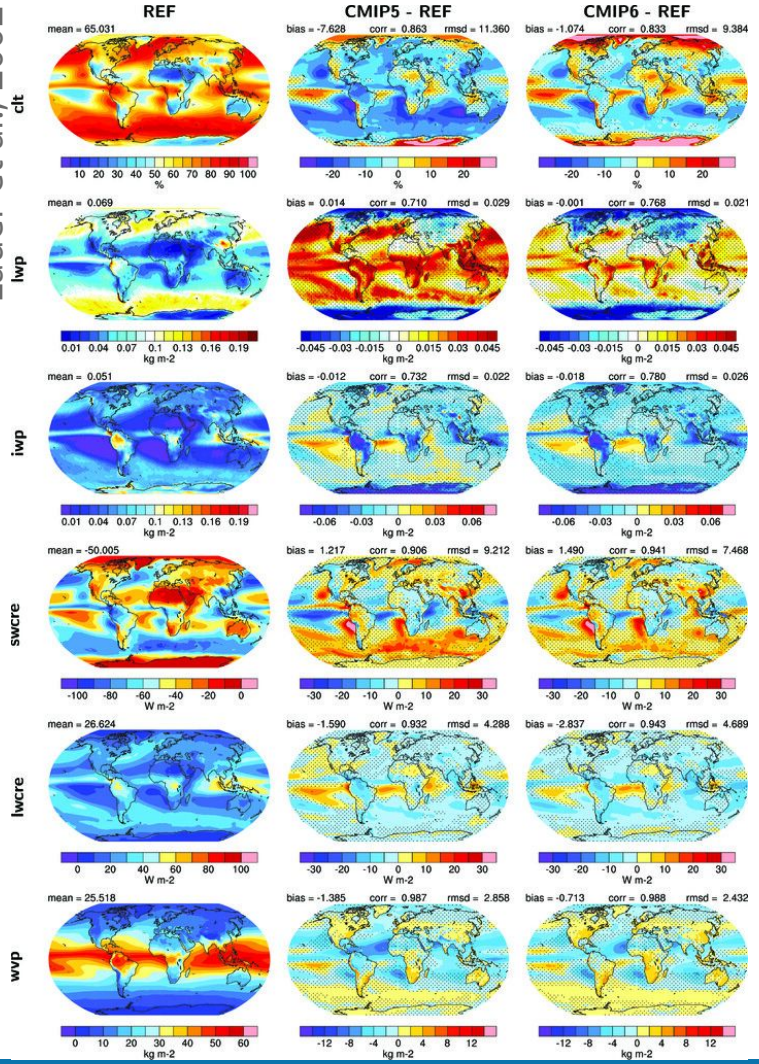


Figure 3.4, IPCC AR6 WG1

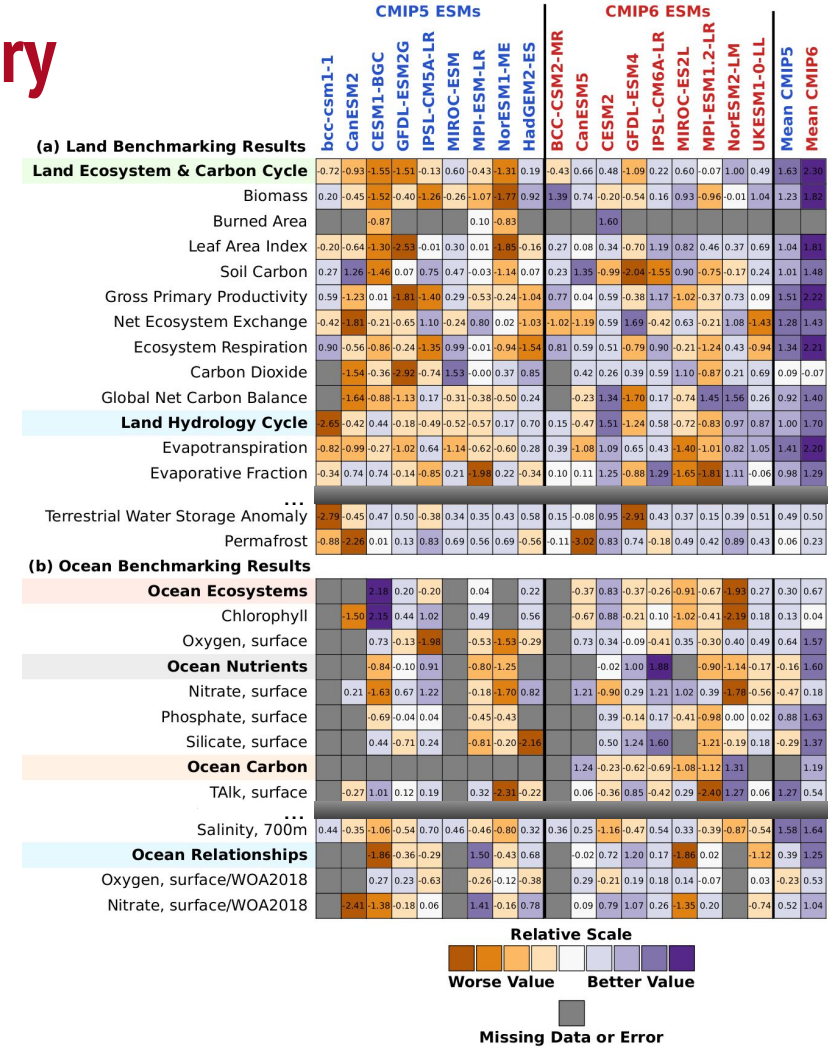
Lauer et al., 2002



Land & Ocean Biogeochemistry

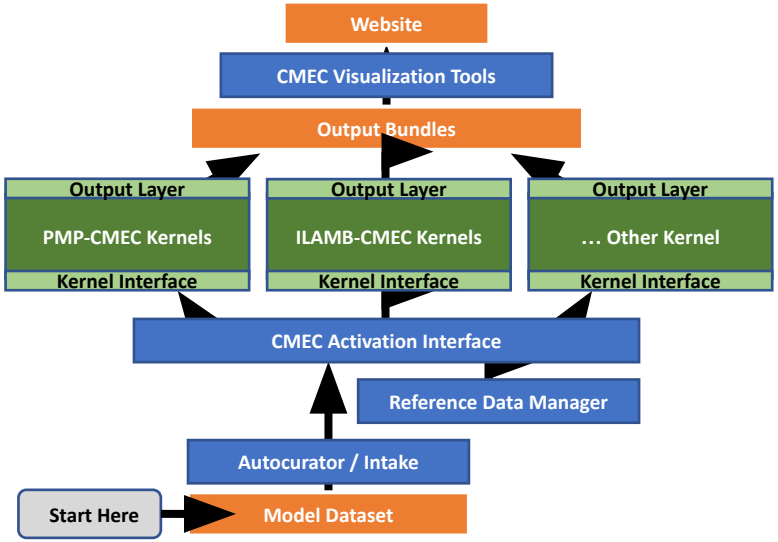
Evaluation of CMIP5 vs CMIP6 with ILAMB and IOMB

- (a) ILAMB and (b) IOMB have been used to evaluate how land and ocean model performance has changed from CMIP5 to CMIP6
- Model fidelity is assessed through comparison of historical simulations with a wide variety of contemporary observational datasets
- The UN's Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) from Working Group 1 (WG1) Chapter 5 contains the full ILAMB/IOMB evaluation as Figure 5.22



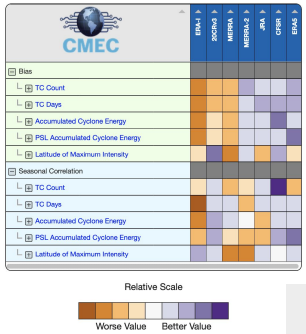
Coordinated Model Evaluation Capabilities

The **CMEC Project** is developing uniform standards and software tools that allow multiple metrics and diagnostics packages to be executed from a unified interface, and results explored in a unified manner.



Project Goals

1. Develop robust **standards** for the development of metrics and diagnostics packages.
2. Develop accompanying tools for **coordinated execution** of metric packages and **interactive analysis of** metrics and diagnostics package output.
3. Build **connections and standards across projects and agencies** related to model evaluation (e.g., MDTF).



CMEC supports a comprehensive software suite for analysis of model evaluation output, including both climate data metrics and diagnostics.

PCMDI Earth System Model Evaluation Project

Contact Paul Ullrich
ullrich4@llnl.gov