GC51X-0294: Climate Model Benchmarking for CMIP7 – A CMIP Task Team

Forrest M. Hoffman[1], Birgit Hassler[2], Rebecca Beadling[3], Ed Blockley[4], Jiwoo Lee[5], Valerio Lembo[6], Jared Lewis[7], Jianhua Lu[8], Luke Madaus[9], Elizaveta Malinina[10], Brian Medeiros[11], Wilfried Pokam Mba[12], Enrico Scoccimarro[13], and Ranjini Swaminathan[14]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America; [2]Germany Aerospace Center (DLR), Oberpfaffenhofen, Germany; [3]Temple University, Philadelphia, Pennsylvania, United States of America; [4]UK Met Office, Exeter, United Kingdom; [5]Lawrence Livermore National Laboratory, Livermore, California, USA; [6]Institute of Atmospheric Sciences and Climate, Bologna, Italy; [7]Climate Resources Pty. Ltd., Melbourne, Australia; [8]Sun Yat-Sen University, Guangzhou, Guangdong, China; [9]Jupiter Intelligence, San Mateo, California, United States of America; [10]Environment Canada, Victoria, British Columbia, Canada; [11]National Center for Atmospheric Research, Boulder, Colorado, United States of America; [12]University of Yaoundé I, Yaoundé, Cameroon; [13]CCMC Foundation, Lecce, Italy; and [14]University of Reading, Reading, United Kingdom;

Coupled Model Intercomparison Project (CMIP) and Model Benchmarking Task Team

- The **Coupled Model Intercomparison Project (CMIP)** is an initiative of the World Climate Research Programme (WCRP) that provides climate projections to support essential WCRP activities and climate science worldwide, including the decision and policy-maker communities.
- The objective of the CMIP is to better understand past, present and future climate changes arising from natural, unforced variability or in response to changes in radiative forcing in a multi-model context.
- CMIP and its associated data infrastructure have become essential to the **Intergovernmental Panel on Climate Change (IPCC)** and other international and national climate assessments.
- The **CMIP International Project Office (CMIP-IPO)** coordinates the project under the governance of the WCRP **Working Group on Coupled Modelling (WGCM)** and is part of the developing **Earth System Modelling and Observations (ESMO)** Core Project, which coordinates all modelling, data and observations activities across WCRP and their key partners.
- ESMO, through its **Working Group on Coupled Models' (WGCM) CMIP Panel** and **WGCM Infrastructure Panel (WIP)**, has established a number of **Task Teams** to support the design, scope, and the evolution of CMIP7.

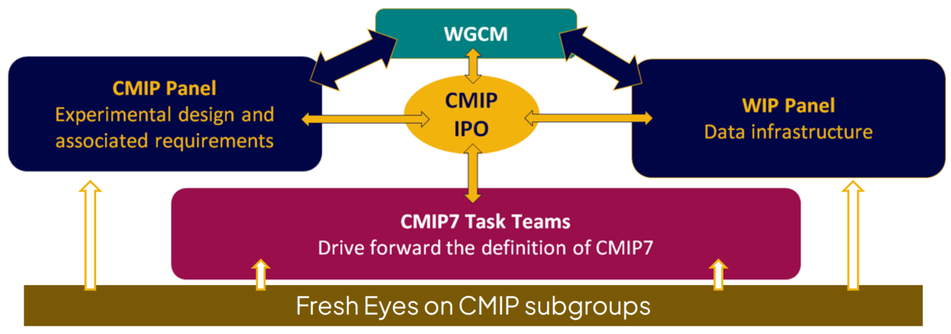


Figure 1: The CMIP-IPO coordinates the CMIP project under the governance of the WCRP Working Group on Coupled Modelling (WGCM).

CMIP Climate Model Benchmarking Task Team

- The **Climate Model Benchmarking Task Team** focuses on designing and integrating systematic and comprehensive model evaluation tools into the CMIP project.
- The Climate Model Benchmarking Task Team consists of 12 members and two co-leads from eight countries, as shown in the author list above.

Model Benchmarking Tools – Info "Cards" & Videos

- The main characteristics of open source benchmarking and evaluation tools that are available for analyses of CMIP-style data are summarized in an overview "card" or an informational video.
- Collected information is presented centrally on the CMIP website for easy community access.
- Cards can be filled out for all available open source benchmarking and evaluation tools if they can be used for CMIP data analysis; a pre-defined questionnaire is available on the CMIP website.

For more information, see https://wcrp-cmip.org/tools/model-benchmarking-and-evaluation-tools/

Model Evaluation and Benchmarking

Defined here are key terms relating to model data analysis and model data evaluation by systematically framing the purpose, outcomes and limitations of such analyses:

- **Verification**: the process of assessing model consistency in terms of correct implementation of the included processes as expected by the model design.
- **Validation**: the process of determining the degree to which a model accurately represents processes in the real world, particularly for the intended uses of the model.
- **Evaluation**: the process of assessing simulations against observations; the necessity for observations means it can only be done for the historical period, and only for variables or processes for which observations or reanalysis data are available.
- **Benchmarking**: the process where model simulations are evaluated with observations, reanalysis data or with other models often resulting in a statement made about the "goodness" of the simulation or model based on a predetermined set of standards or criteria.

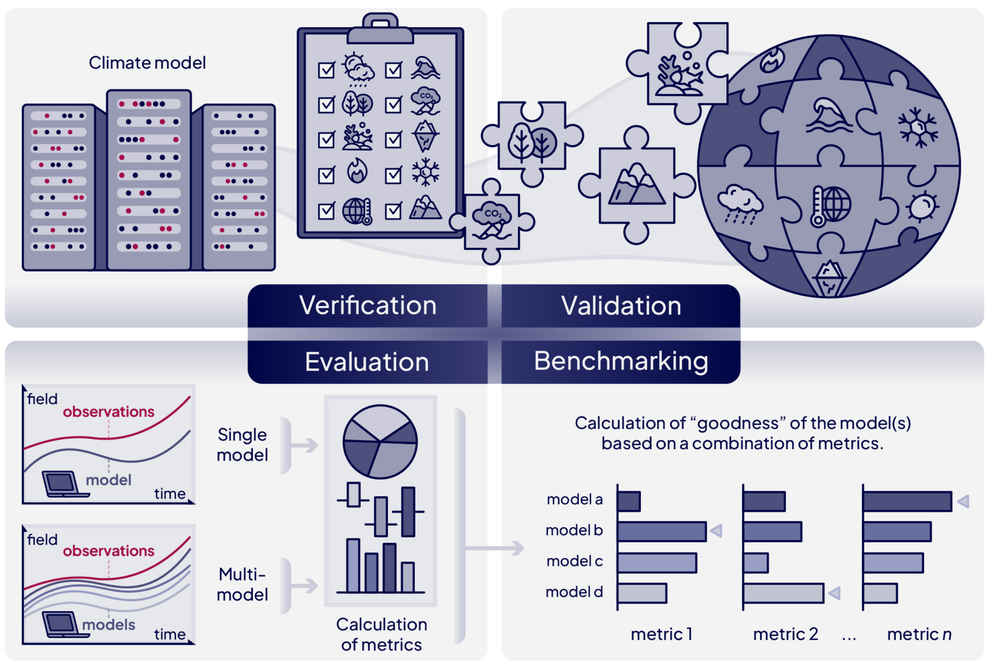


Figure 2: Schematic definition of the terms Verification, Validation, Evaluation and Benchmarking for use in the climate model context. Note that although through benchmarking some kind of ranking can be performed based on the chosen metric and selected observations, this is by far not a generally applicable ranking valid for all metrics, all realms and all possible observational references.

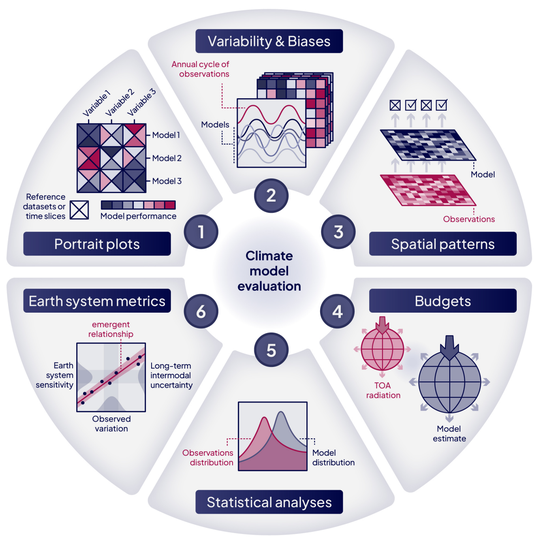


Figure 3: Shown are six different schemes commonly used for evaluation and benchmarking of climate models. Most of the schemes can be applied to different realms, and each can include more than one diagnostic or metric. Scheme 1 includes only the portrait plot as a metric which is very versatile in its application across different domains, analysed variables and a number of included observations or time periods. Scheme 2 represents all diagnostics that are based on analyses of biases and variabilities. Scheme 3 includes all diagnostics that focus on spatial analyses, e.g. spatial correlations or physical connections between neighboring regions/realms. Scheme 4 includes any budget assessments. These diagnostics are commonly applied globally and can also be applied regionally if boundary conditions and fluxes across boundaries are clearly defined. Scheme 5 represents all other statistical approaches for model evaluation, e.g. the analyses of distributions. Scheme 6 includes all diagnostics for describing the Earth System and its interconnections and changes as a whole, e.g. emergent constraints or equilibrium climate sensitivity (ECS).

CMIP Rapid Evaluation Framework (REF)

The CMIP Model Benchmarking Task Team developed a system specification for a **Rapid Evaluation Framework (REF)** that would leverage community benchmarking metrics to evaluate CMIP model output as they are submitted to the Earth System Grid Federation (ESGF).

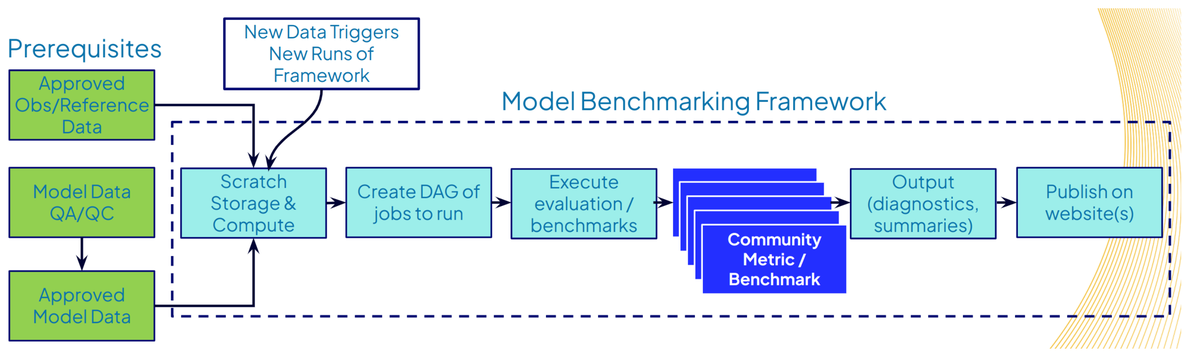
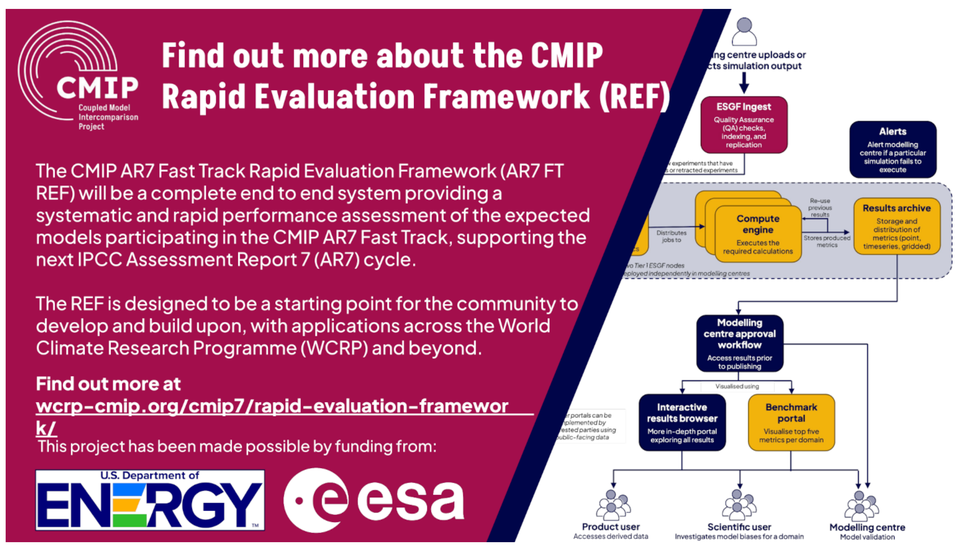


Figure 4: New model output data arriving at ESGF will trigger the Rapid Evaluation Framework (REF) to launch a suite of community-developed model evaluation and benchmarking modules. Diagnostics and summaries of comparisons with observational data will be generated and published on websites to assist the research community in selecting appropriate models for their studies.

Initial Diagnostic Selection

The diagnostics listed in the following table were selected for implementation in the initial version of the REF, which is expected to be deployed for evaluating and benchmarking AR7 Fast Track simulations by the middle of 2025.

ID	Diagnostic	Theme
1.1	Antarctic annual mean/Arctic September rate of sea ice area (SIA) loss per degree warming (dSIA/dGMST)	Oceans and sea ice
1.2	Atlantic Meridional Overturning Circulation	Oceans and sea ice
1.3	ENSO diagnostics (lifecycle/ seasonality/ teleconnections)	Oceans and sea ice
1.4	Sea surface temperature (SST)/Sea surface salinity (SSS) bias	Oceans and sea ice
1.5	Ocean heat content (OHC)	Oceans and sea ice
1.6	Antarctic/Arctic sea ice area seasonal cycle	Oceans and sea ice
2.1	Soil carbon	Land and land ice
2.2	Gross primary production	Land and land ice
2.3	Runoff	Land and land ice
2.4	Surface soil moisture	Land and land ice
2.5	Net Ecosystem Exchange	Land and land ice
2.6	Leaf area index	Land and land ice
2.7	Snow cover	Land and land ice
3.1	Annual cycle and seasonal mean of multiple variables	Atmosphere
3.2	Radiative and heat fluxes at the surface and TOA	Atmosphere
3.3	Climate variability modes (e.g., ENSO, MJO, Extratropical modes of variability, monsoon)	Atmosphere
3.4	E-P (evaporation – precipitation)	Atmosphere
3.5	Double ITCZ	Atmosphere
3.6	Cloud Radiative Effects	Atmosphere
3.7	Scatterplots of two cloud-relevant variables (specific regions of the globe: with specific cloud regimes)	Atmosphere
4.1	Equilibrium climate sensitivity (ECS)	Earth System
4.2	Transient climate response (TCR)	Earth System
4.3	Transient climate response to cumulative emissions of carbon dioxide (TCRE)	Earth System
4.4	Zero Emissions Commitment (ZEC)	Earth System
4.5	Historical changes in climate variables (time series, trends)	Earth System
5.1	High Amplitude Rossby Waves	Impacts and Adaptation
5.2	Internal Variability or ensemble spread within individual models (focused on precipitation and surface temperature)	Impacts and Adaptation
5.3	Evaluation of key climate variables at Global Warming Levels	Impacts and Adaptation
5.4	Climate drivers for fire (Fire burnt area, fire weather and fuel continuity)	Impacts and Adaptation



Engagement Opportunities

Acknowledgments

This research uses resources of the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for the US Department of Energy under Contract No. DE-AC05-00OR22725. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP.