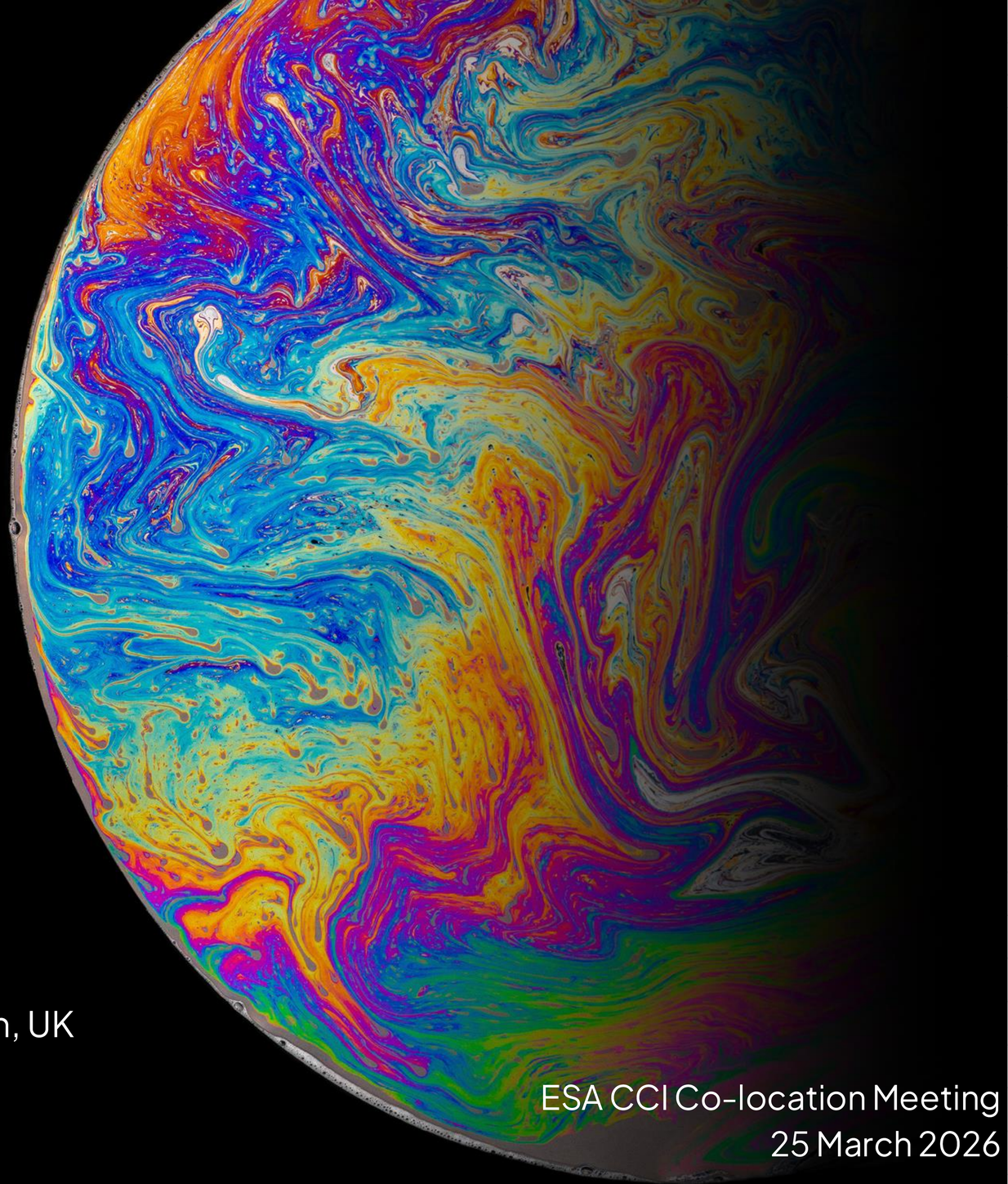
CMIP
Coupled Model Intercomparison Project

Observations for Climate Model Evaluation



Ranjini Swaminathan[1], Rebecca Beadling[2] and the Model Benchmarking Task Team

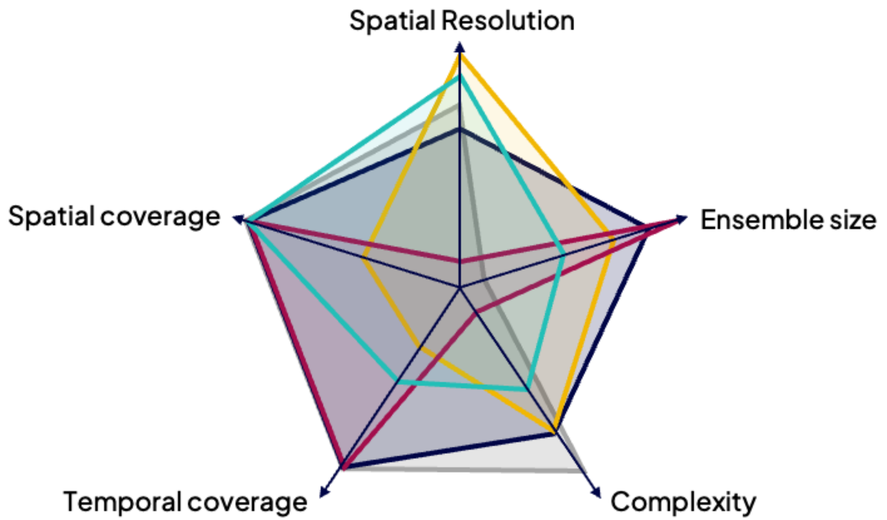
[1] University of Reading and National Centre for Earth Observation, UK

[2] Temple University, USA

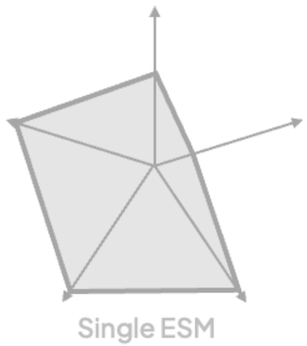
ESA CCI Co-location Meeting
25 March 2026

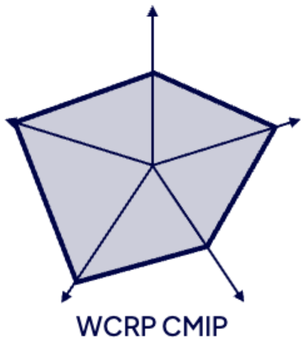
Background

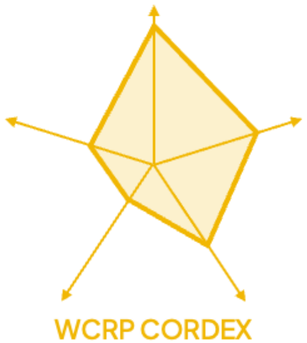
The WCRP Modelling Multiverse

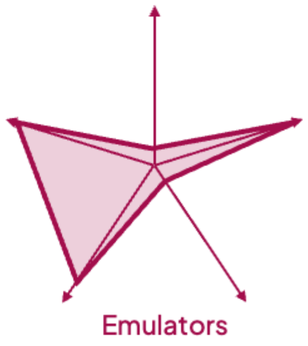


Each type of model or modelling project has a different ability to model over different spatial resolutions, spatial coverage, temporal coverage, model complexities, and ensemble sizes.







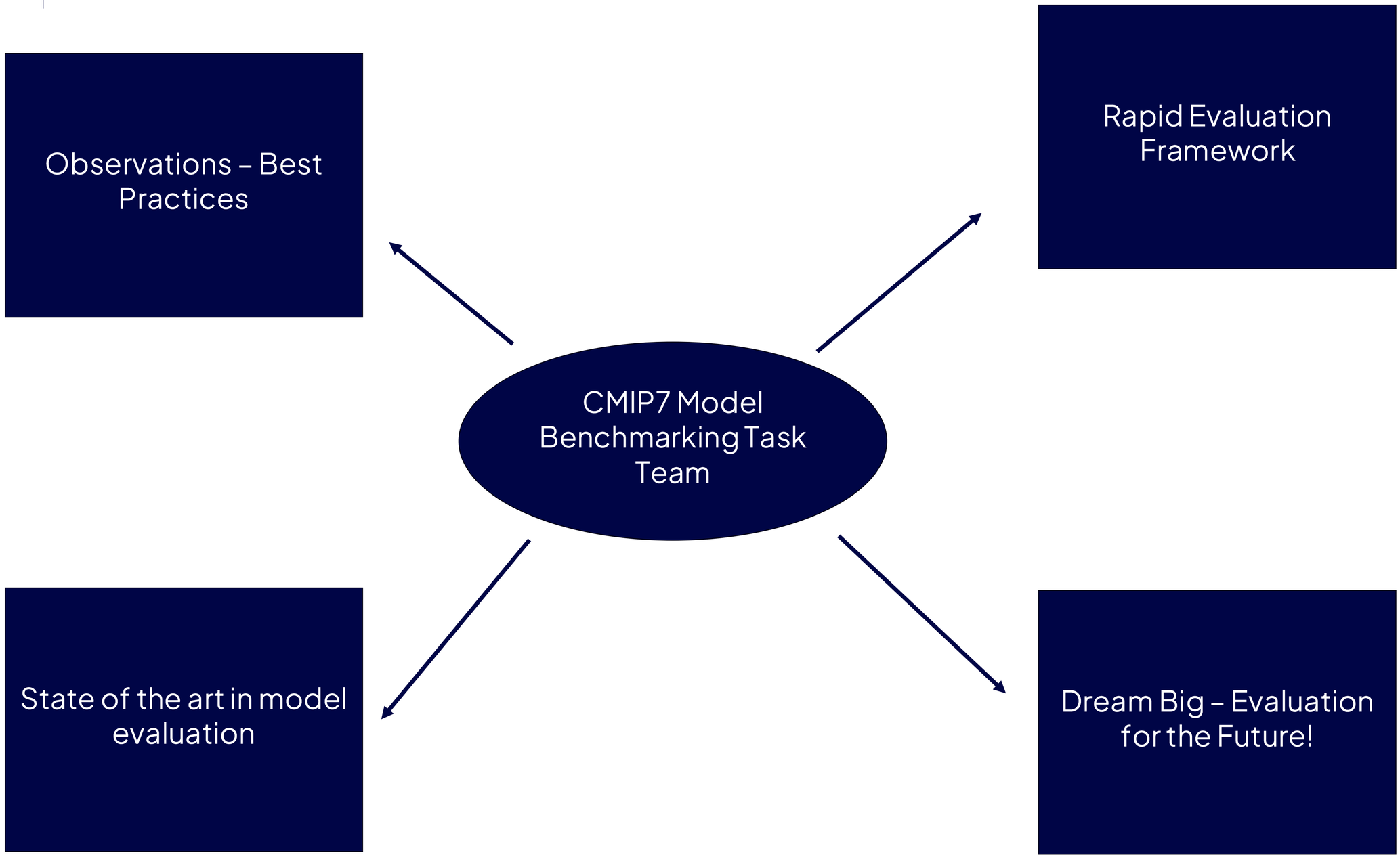


Dingley, B. et al (2023). WCRP Modelling Multiverse. Zenodo. https://doi.org/10.5281/zenodo.8047805

CMIP7 Model Benchmarking Task Team

- CMIP (Coupled Model Intercomparison Project) goal : better understand past, present, and future climate changes in a multi-model context.
- Prerequisite for reliable climate information from climate and Earth system models: understand model capabilities and limitations.
- Essential to evaluate the models systematically and comprehensively with the best available observations and reanalysis data.
- Full integration of routine benchmarking and evaluation of the models into the CMIP publication workflow.
  - Technical challenges: high resolution, memory limits, unstructured grids.
  - Science challenges: innovative diagnostics, novel observations, ML based analysis

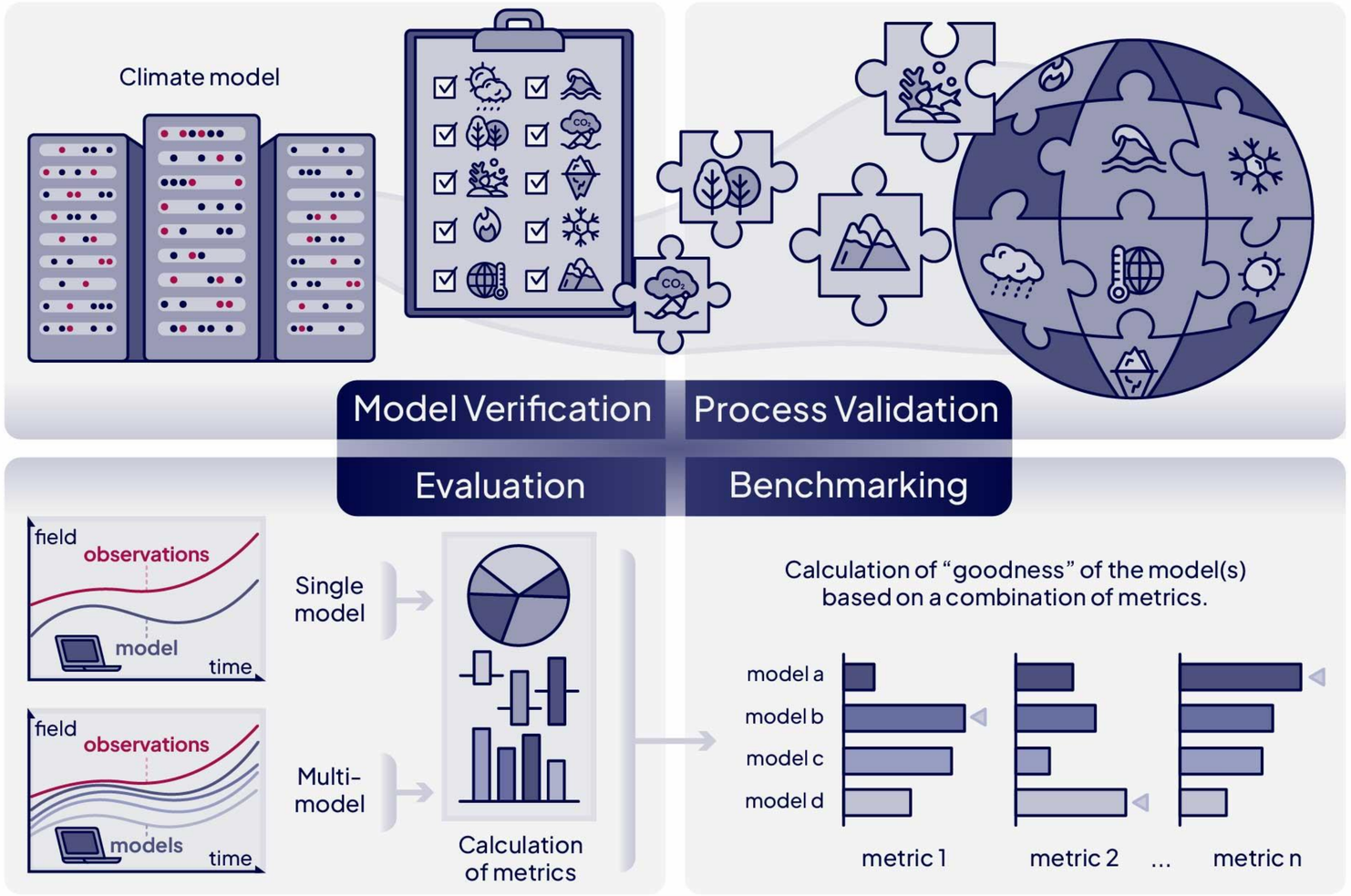
CMIP | WCRP


Observations – Best Practices
Rapid Evaluation Framework
CMIP7 Model Benchmarking Task Team
State of the art in model evaluation
Dream Big – Evaluation for the Future!

Model evaluation

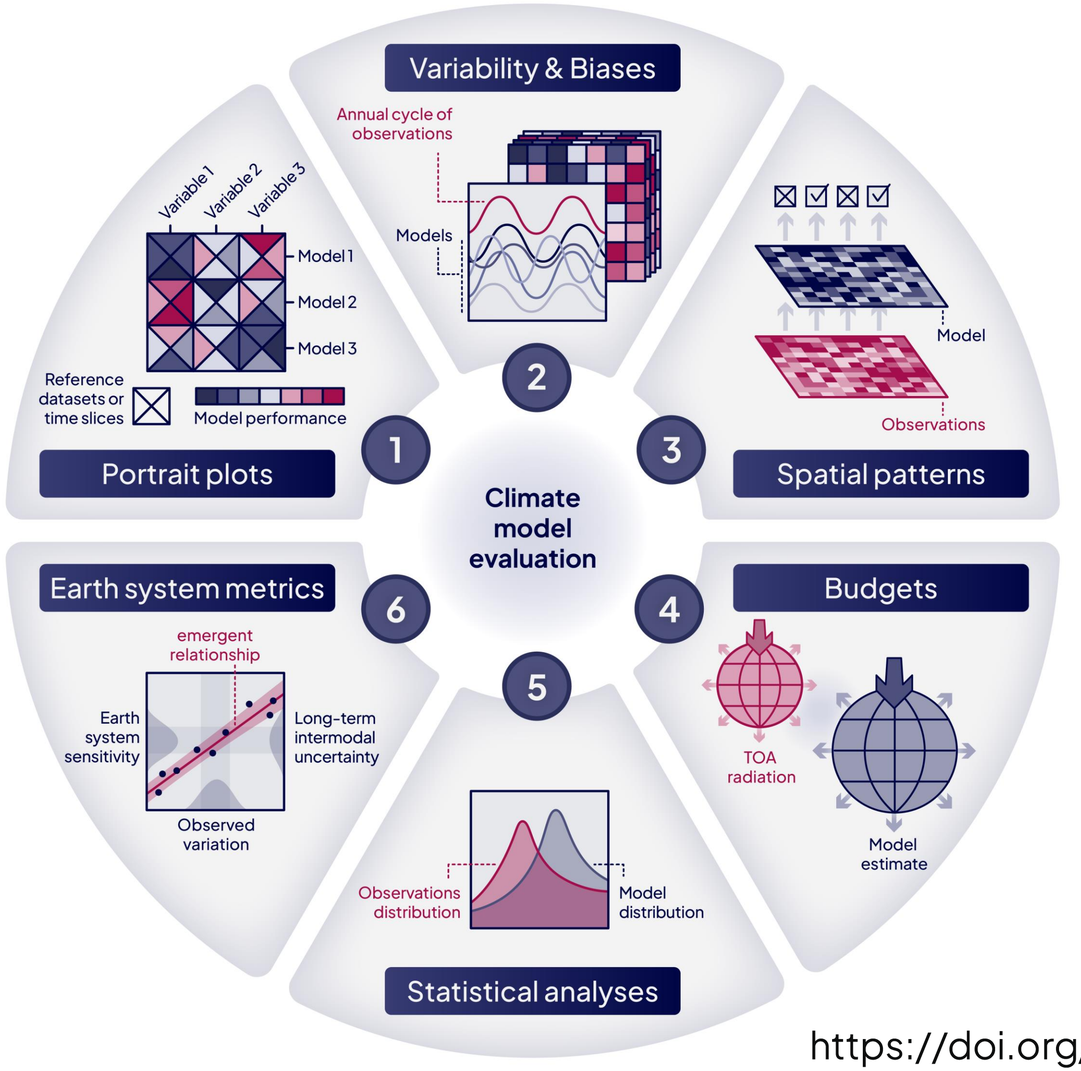
CMIP | WCRP

Definition of model evaluation and benchmarking



https://doi.org/10.5281/zenodo.17649933

Benchmarking and evaluation schemes



- [illegible] grouped according to their underlying evaluation principles
- Portrait plots, Variability & Biases, Spatial patterns, Budgets, Statistical analyses, Earth System metrics
- Most of them can be applied [illegible]

https://doi.org/10.5281/zenodo.13934006

Benchmarking and evaluation tools

Filter Sortieren ...

ILAMB

Category
Evaluation and benchmarking tools

Description
The International Land Model Benchmarking (ILAMB) project is a model-data intercomparison and integration project designed to assess the performance of land ...

Website
https://www.ilamb.org/

ESMValTool

Category
Evaluation and benchmarking tools

Description
ESMValTool is an open-source community-developed diagnostics and performance metrics tool for the evaluation and analysis of Earth System Models.

Website
https://www.esmvaltool.org/

bgcval2

Category
Evaluation and benchmarking tools

Description
Python based Software toolkit for monitoring on-going simulations of the ocean, and the marine component of earth System models.

Website
https://github.com/valeriupredoi/bgcval2

- Open-source
- Useable for CMIP data analyses
- 11 tools in the database so far
- Available at: https://wcrp-cmip.org/tools/model-benchmarking-and-evaluation-tools/

Role of Observations

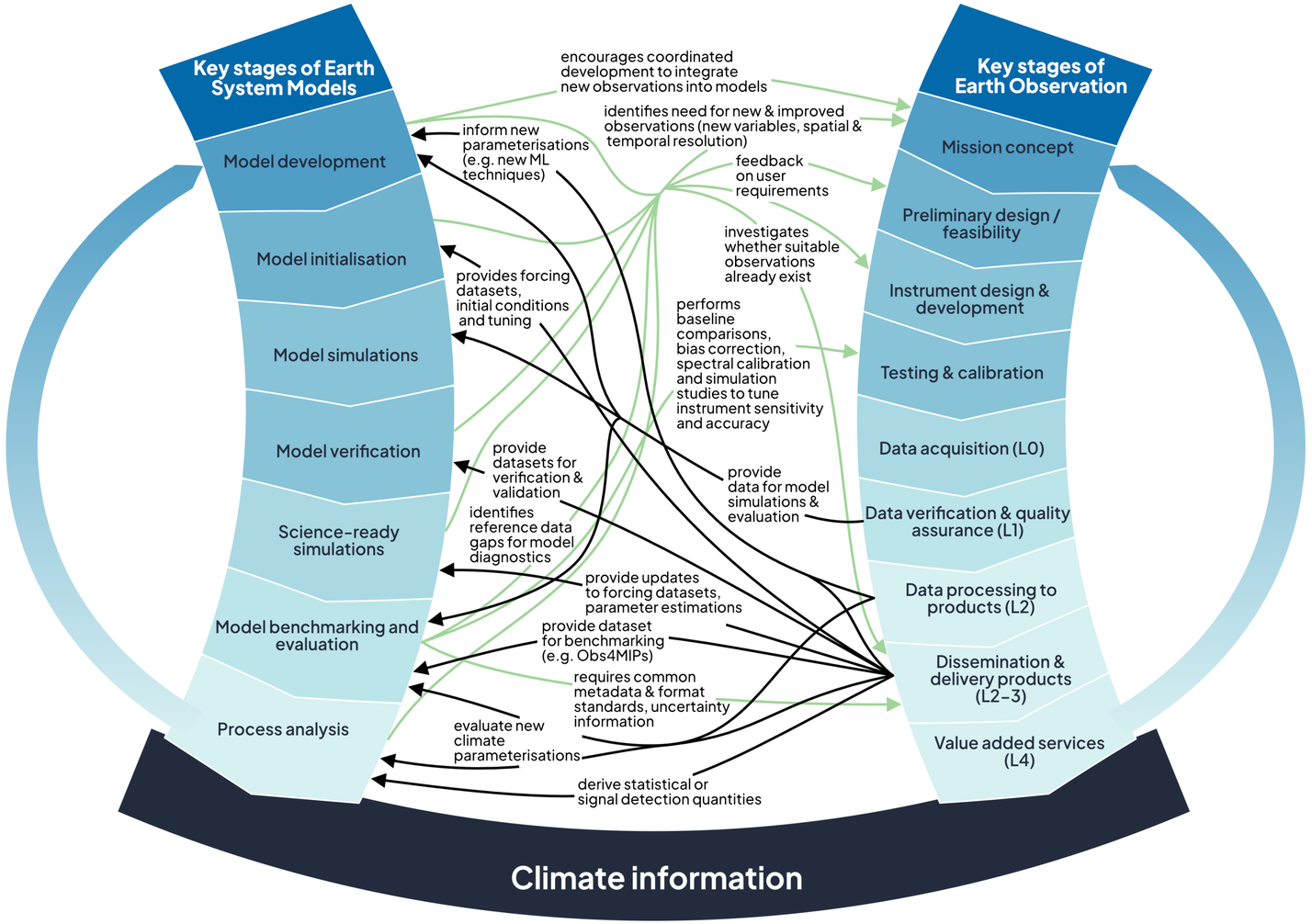
CMIP | WCRP

Community Engagement

- Surveys
- ESA CCI Co-location meeting – Oct 2024
- CMUG engagement
- REF Hackathon – March 2025
- Evaluation tools
- Observation dataset providers
- Expanding uncertainty information – Uncertainty Workshop – Aug 2025

CMIP | WCRP

The climate modelling-observation interface



.org/10.5281/zenodo.18433020

Best Practices

CMIP | WCRP

Data processing

- Spatial resolution and sampling
- Point source data
- Scaling , gap filling and extrapolation
- Point source data

CMIP | WCRP

Model to observation comparison in metrics

- Ensure variables represent same quantities
- Account for different forcings and experiments.
- Process-based evaluation.
- Time series and climatology construction – must be same for model and observations.
- Data and met-data standards - CF, Obs4MIPs.

CMIP | WCRP

Uncertainties

- Vocabulary around quantity represented and uncertainty provided.
- Proxies.
- Documentation on uncertainties.
- Cross-community engagement on updates – Rapid Evaluation Framework

CMIP | WCRP

Next general model evaluation

- AI models – uncertainty, physical consistency.
- Data provenance.
- Under utilized and unutilized datasets.
- Community engagement.
- Funding!!!!

Rapid Evaluation Framework

Diagnostic collections including reference datasets

The CMIP Model Benchmarking Task Team, through consultation with the CMIP modelling community, have collected a set of informative diagnostic collections for five different climate model themes for CMIP Assessment Fast Track model evaluation.

Oceans & Sea ice - 6 diagnostic collections

Atmosphere - 6 diagnostic collections

Land & Land ice - 7 diagnostic collections

Earth system - 5 diagnostic collections

Impacts & adaptation - 2 diagnostic collections

For the full list see: DOI 10.5281/zenodo.14284374

Scientific example

- Sea ice diagnostics 1.1 and 1.6 developed by ESMValTool:
  - Rate of sea ice area (SIA) loss per degree warming (dSIA/dGMST)
  - Antarctic/Arctic sea ice area seasonal cycle
- Observational uncertainty ranges from 3 reference datasets inc. CCI/OSI SAF (Laverne et al. 2019)

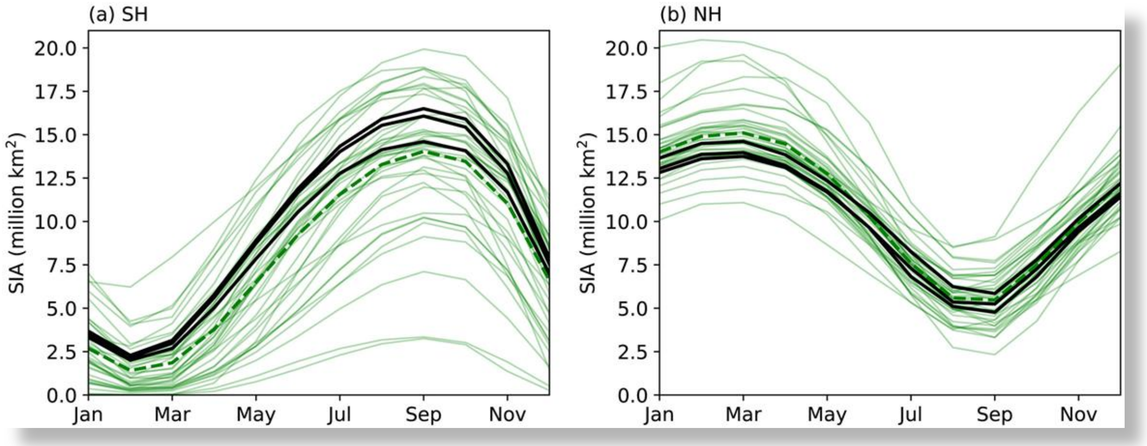
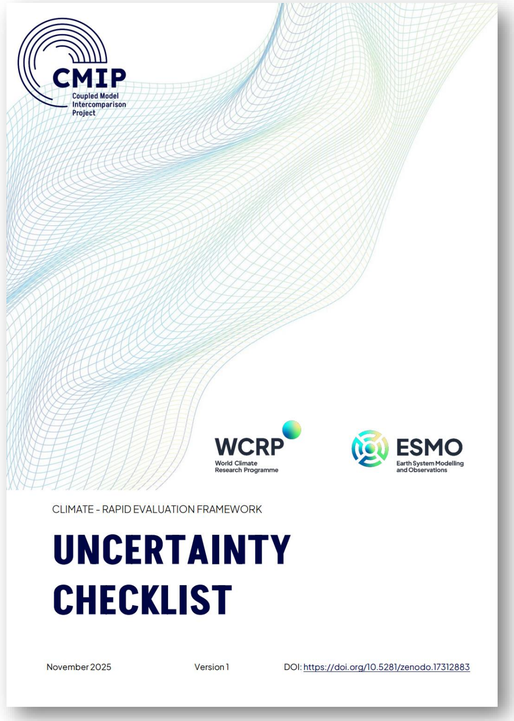


Fig 1 from Roach et al. 2020: The 1979–2014 mean seasonal cycle in sea ice area for the CMIP6, and three observational products (black lines) in (a) the Antarctic and (b) the Arctic.

Treatment of uncertainties



- REF requires uniform approach
- Consultations with obs4MIPs, observation community, metric package developers and CMIP7-CVs Task Team
- Additional global attributes describing if and what uncertainty information is included ("has_auxdata", "aux_variable_id")
- Different approaches for CMOR and CMOR-like datasets
  - CMOR: each uncertainty information is included in separate files
  - CMOR-like: uncertainty information included as ancillary variables following CF conventions
- Participation of REF as case study in NPL led working group on new Uncertainty Metadata/Nomenclature Convention (UNC)

References

- Hassler, B., Hoffman, F. M., et al. (2026). Systematic benchmarking of climate models: Methodologies, applications, and new directions. Reviews of Geophysics, 64, e2025RG000891. https://doi.org/10.1029/2025RG000891 **(Published)**

- Beadling, R. L., Swaminathan, R., et al. (2026). Observational Data for Next Generation Climate Model Evaluation: Requirements, Considerations, and Best Practices. *Bulletin of the American Meteorological Society* (published online ahead of print 2026), BAMS-D-25–0079.1, Article BAMS-D-25–0079.1. https://doi.org/10.1175/BAMS-D-25–0079.1 **(Early online edition).**

- Hoffman, F. M., Hassler, B., Swaminathan, R., et al. Rapid Evaluation Framework for the CMIP7 Assessment Fast Track, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025–2685 , 2025. **(Under Review)**

Thank you!