Earth System Model Evaluation and Benchmarking with the *PCMDI Metrics Package (PMP)*

Current Core Team members:

Jiwoo Lee, Ana Ordonez, Peter Gleckler, Paul Ullrich,
Bo Dong, Kristin Chang (LLNL, PCMDI)

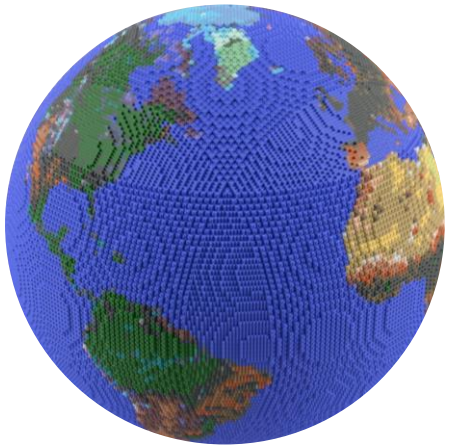
Along with contributors:

Mark Zelinka, Li-Wei Chao, Tom Vo, Paul Durack (LLNL),
Min-Seop Ahn (NASA GSFC), Yann Planton (Monash U),
Michael Wehner (LBNL), Daehyun Kim (SNU),
Elina Valkonen (NASA GSFC), Julie Caron (NCAR),
and many others!

@AOGS, 2024 June 25th

Systematic Evaluation for Diverse Models

There are 1000s simulations from 100s of diverse models in CMIPs!




Different resolution


Dynamics / Physics

Or, something unique!

How can we **objectively evaluate** and **efficiently document** their **performance**?

PCMDI Metrics Package (PMP)

Lee et al. 2024: Systematic and Objective Evaluation of Earth System Models: PCMDI Metrics Package (PMP) version 3. *Geoscientific Model Development*, 17, 3919–3948, doi: 10.5194/gmd-17-3919-2024

PMP is:

Open-source Python package for **objective evaluation and benchmarking** of physical climate as simulated by models

PMP does :

Assess model **performance** using diverse metrics

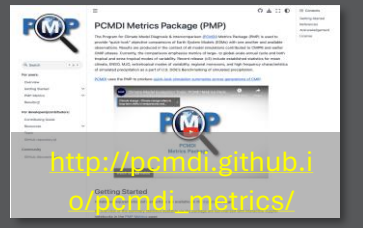
Ensure **reproducibility** with detailed **provenance** and **version** control (codes, data, and operating conditions)

Link to reference datasets from **obs4MIPs** for more robust evaluation and reproducibility

PMP provides:

Reuseable software with documentation

Pre-calculated **database** of statistics and metrics for the CMIP archive



Impacts:

Quantitatively measure the performance evolution across **CMIP** generations

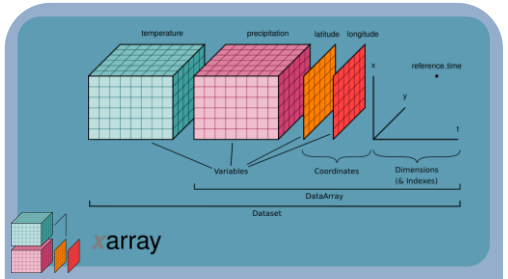
Provide objective goals for **model development**

Downloaded > *33,000* times and used for evaluation of DOE and other agencies' models

PMP's primary building component: xCDAT

T. Vo, S. Po-Chedley, J. Boutte, J. Zhang, J. Lee (LLNL)

xCDAT *encourages reusable code and reproducible science*



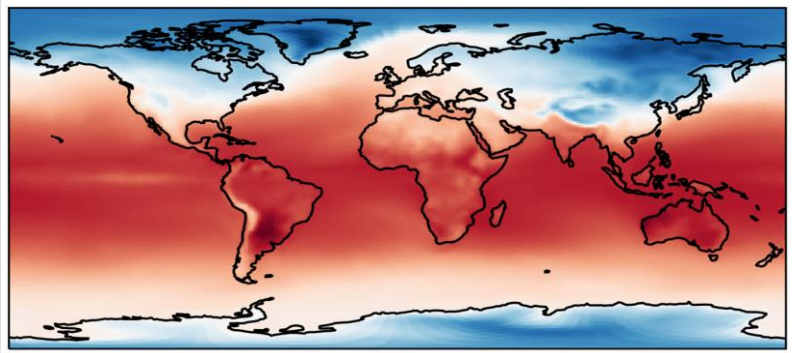
- Geospatial weighted averaging
- Temporal averaging, climatologies, departures
- Dataset bounds and CF metadata handling
- Horizontal and vertical regridding

Pure Xarray

```python
import numpy as np
import xarray as xr

# 1. Open the dataset.
dpath = (
    "/p/user_pub/work/CMIP6/CMIP/E3SM-Project/"
    "E3SM-2-0/historical/r1i1p1f1/Amon/ts/gr/v20220830/"
)
ds = xr.open_mfdataset(dpath + "*.nc")

# 2. Calculate monthly departures.
ts_mon = ds.ts.groupby("time.month")
ts_mon_clim = ts_mon.mean(dim="time")
ts_anom = ts_mon - ts_mon_clim

# 3. Compute global average.
coslat = np.cos(np.deg2rad(ds.lat))
ts_anom_wgt = ts_anom.weighted(coslat)
ts_anom_global = ts_anom_wgt.mean(dim="lat").mean(dim="lon")

# 4. Calculate annual averages.
# ncar.github.io/esds/posts/2021/yearly-averages-xarray/
mon_len = ts_anom_global.time.dt.days_in_month
mon_len_by_year = mon_len.groupby("time.year")
wgts = mon_len_by_year / mon_len_by_year.sum()

temp_sum = ts_anom_global * wgts
temp_sum = temp_sum.resample(time="AS").sum(dim="time")
denom_sum = (wgts).resample(time="AS").sum(dim="time")

ts_anom_global_ann = temp_sum / denom_sum

```

xCDAT

```python
import xcdat as xc

# 1. Open the dataset.
dpath = (
    "/p/user_pub/work/CMIP6/CMIP/E3SM-Project/"
    "E3SM-2-0/historical/r1i1p1f1/Amon/ts/gr/v20220830/"
)
ds = xc.open_mfdataset(dpath)

# 2. Calculate monthly departures.
ds_anom = ds.temporal.departures("ts", freq="month")

# 3. Compute global average.
ds_anom_global = ds_anom.spatial.average("ts")

# 4. Calculate annual averages
ds_anom_global_ann = ds_anom_global.temporal.group_average
("ts", freq="year")

```



A comparison of code to calculate global-mean, monthly anomalies

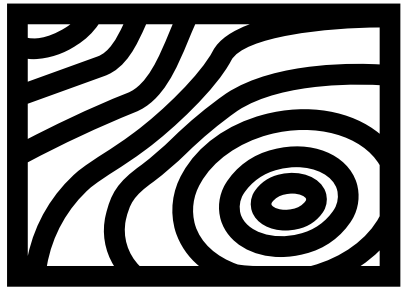
More examples available at https://xcdat.readthedocs.io

Vo et al. (2024) xCDAT: A Python package for simple climate data analysis on structured grids. Journal of Open Source Software. DOI: 10.21105/joss.06426

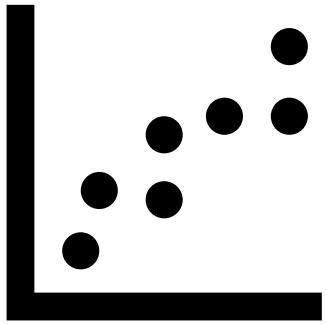
What Do We Evaluate?

Average

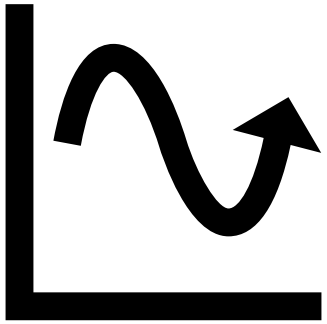
Pattern



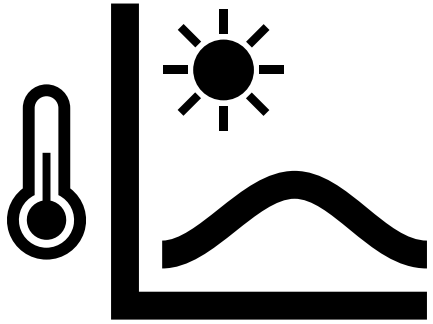
Physical Relationships



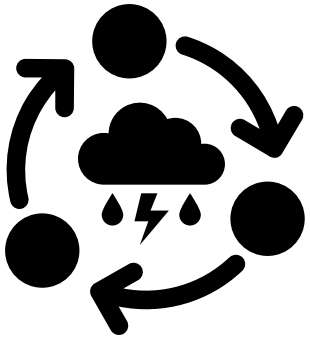
Variability



Extremes



Processes

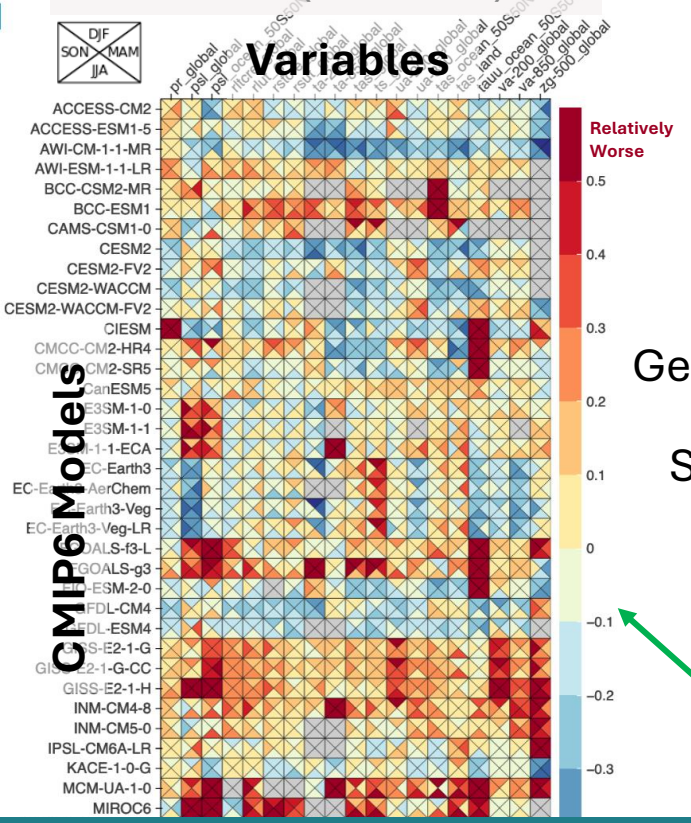


Evaluation needs to include diverse aspects of the simulated physical climate

Evaluation of **Climatology**

High-level Performance Summary

Gleckler Plot (aka Portrait Plot)



Sea-surface Temperature

Precipitation

Precipitable Water

Wind

Geopotential Height

Sea Level Pressure

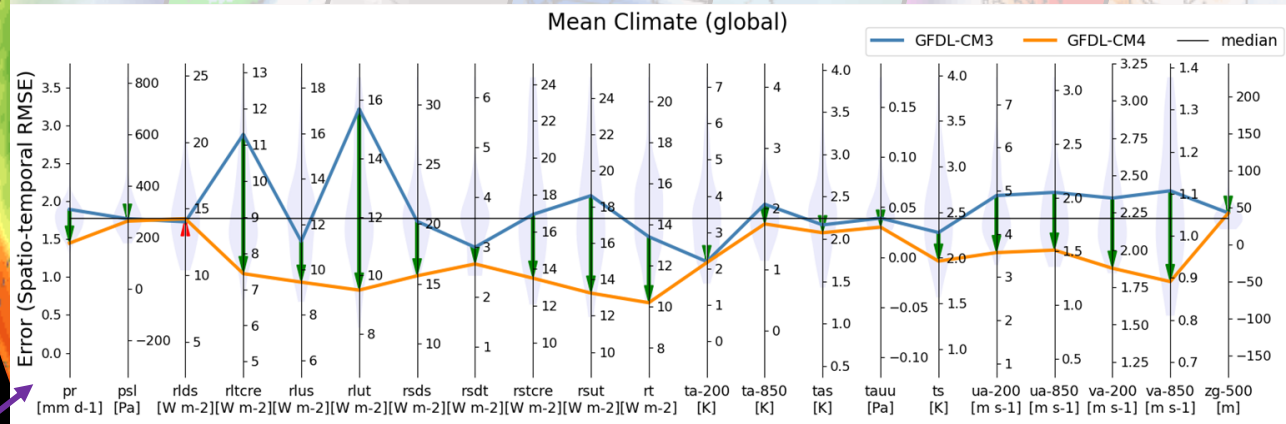
... and many others!

Relative Performance

Absolute error

Objective Performance Tracking during Model Development

Parallel Coordinate Plot



Collective evaluation of multiple climate fields enables objective performance tracking

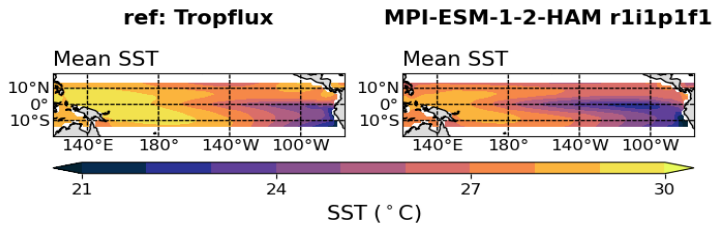
Field image : ClimateReanalyzer.org

Lawrence Livermore National Laboratory

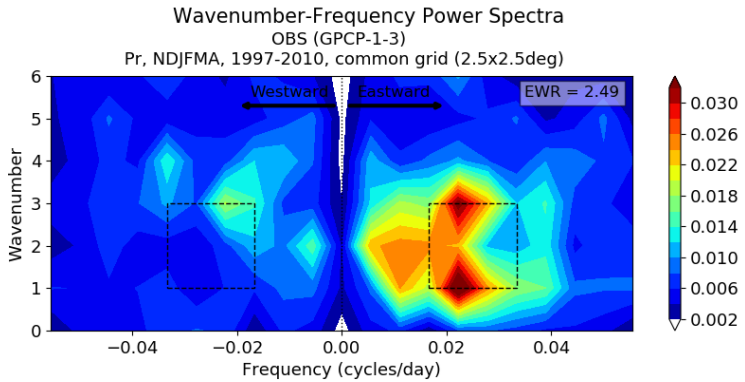
PCMDI

Climate Variability

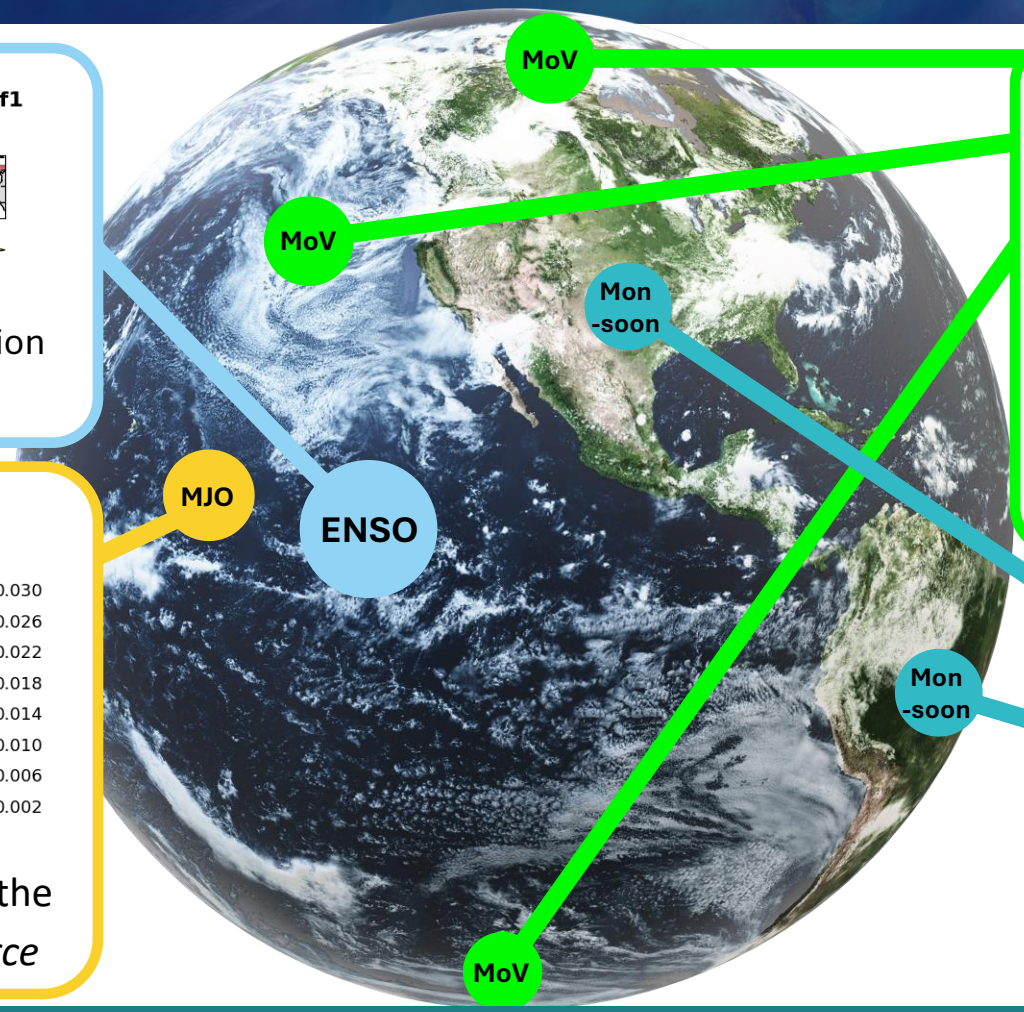
Tropics



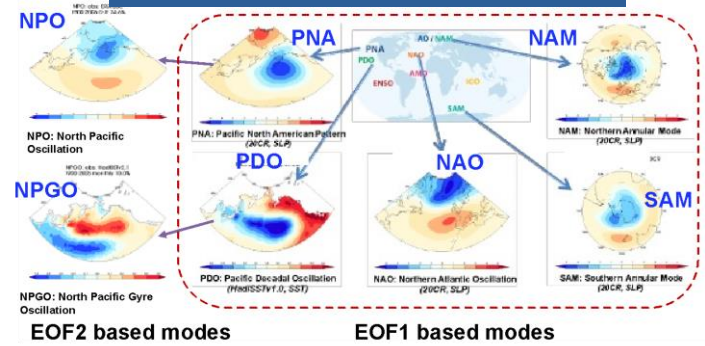
ENSO metrics developed by the collaboration with *CLIVAR Pacific Region Panel*



MJO propagation metrics developed by the collaboration with *CLIVAR MJO Task Force*

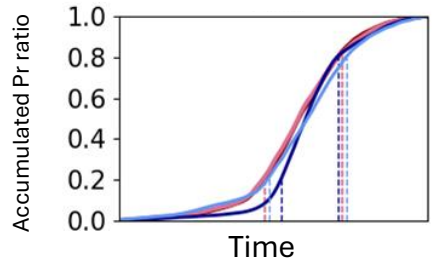


Mid-to-High Latitude



Extratropical **Modes of Variability** *using EOF and CBF*

Monsoon

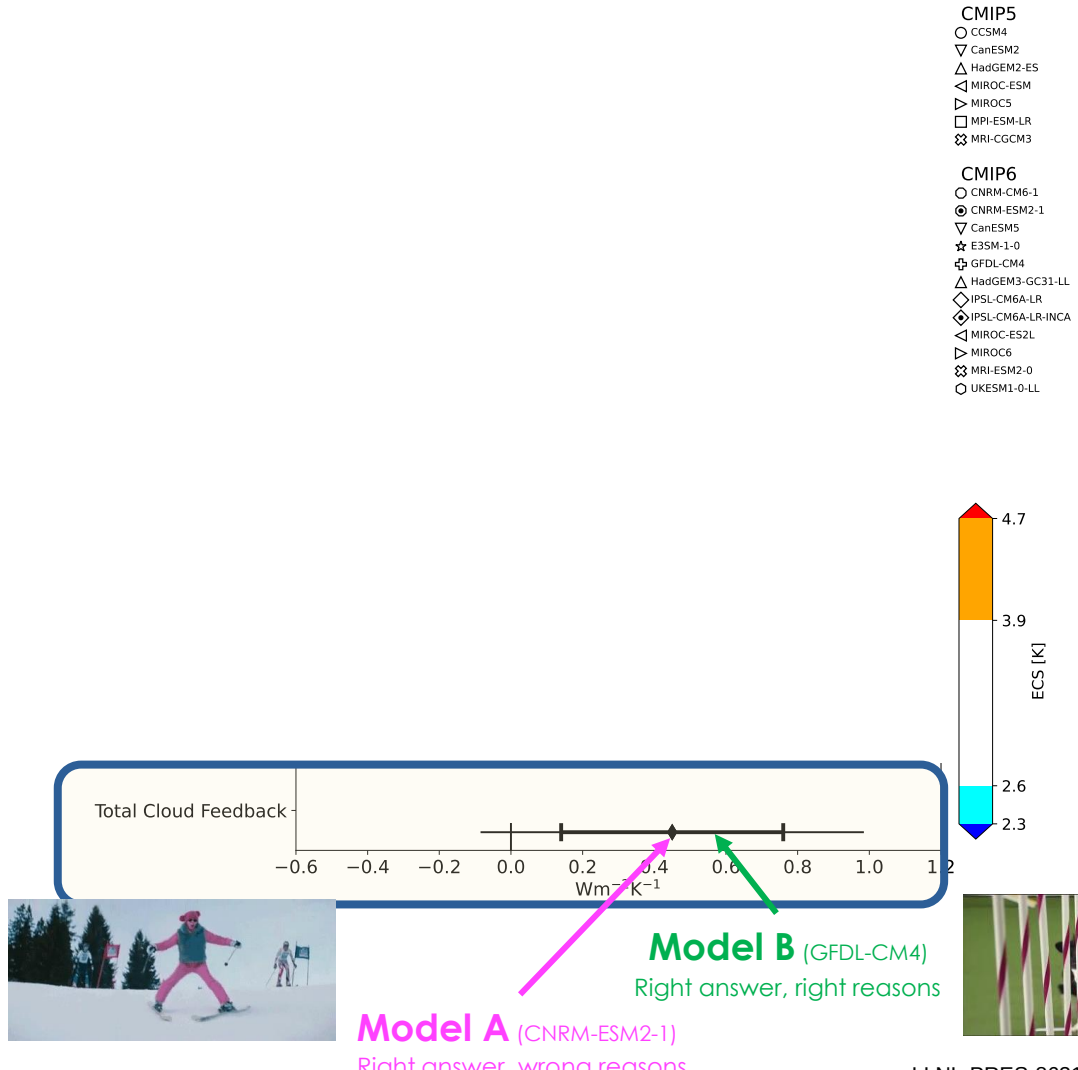


Evaluation of climate variability allows us to explore connectivity in the climate system

Cloud Feedbacks


Implementation contributed by Mark Zelinka and Li-Wei Chao

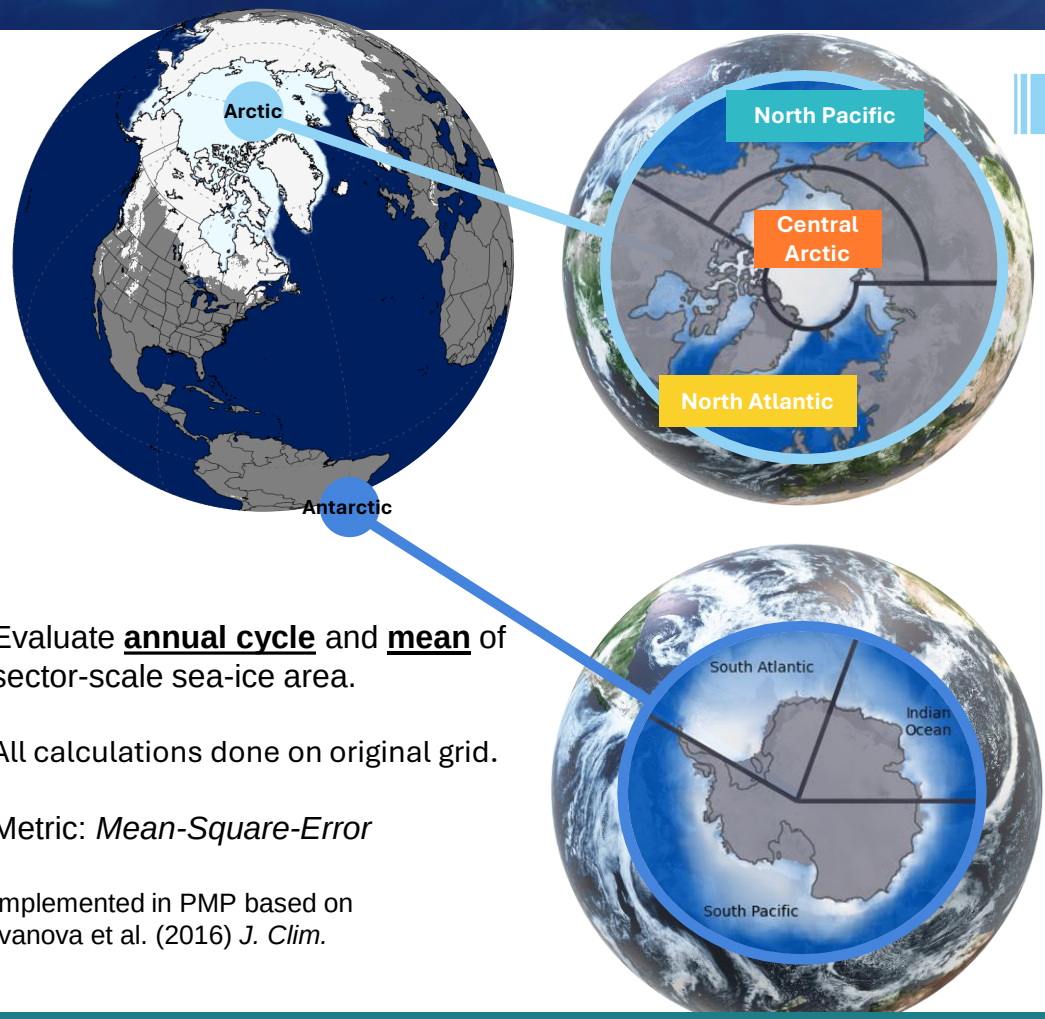
- **Cloud feedbacks** are broken down into individual components quantified in the WCRP Climate Sensitivity assessment. (Sherwood et al. 2020)
- An overall cloud feedback error metric is computed for each model based on the RMSE across the individual cloud feedback components.
- **Mean-state cloud property** error metrics (Klein et al. 2013) are also computed as part of this package.



Recent addition: Sea Ice Area

Sector scale analysis: Moving beyond total hemispheric sea-ice extent

Ana Ordonez, Jiwoo Lee, Paul Durack, Peter Glekler

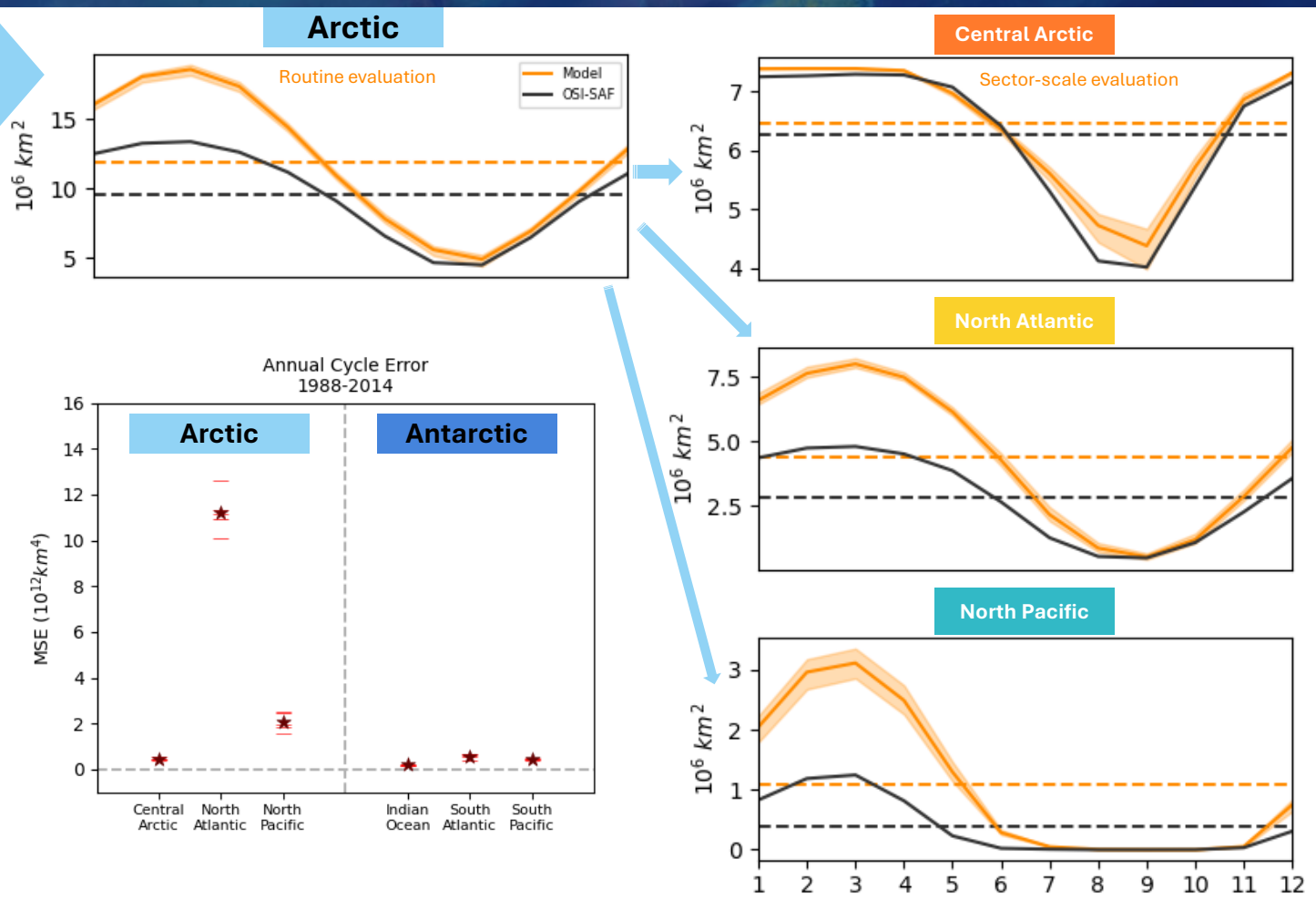


Evaluate **annual cycle** and **mean** of sector-scale sea-ice area.

All calculations done on original grid.

Metric: *Mean-Square-Error*

Implemented in PMP based on Ivanova et al. (2016) *J. Clim.*

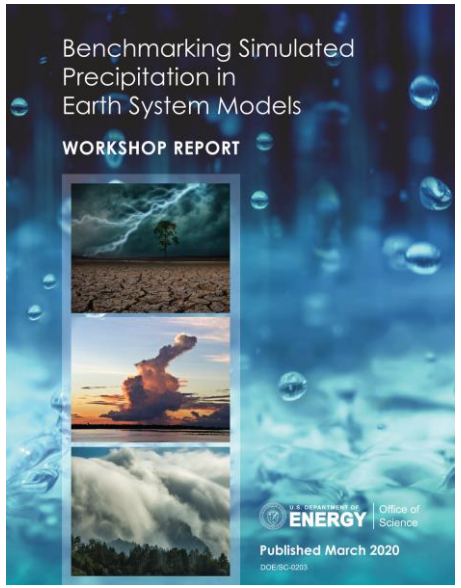


Evaluation efforts have expanded to include more components of the climate system

Please visit Ana Ordonez's poster for details!

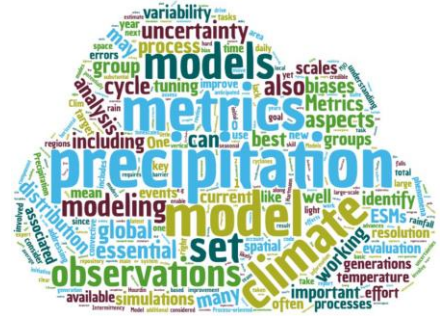
Precipitation Benchmarking (including Extremes)

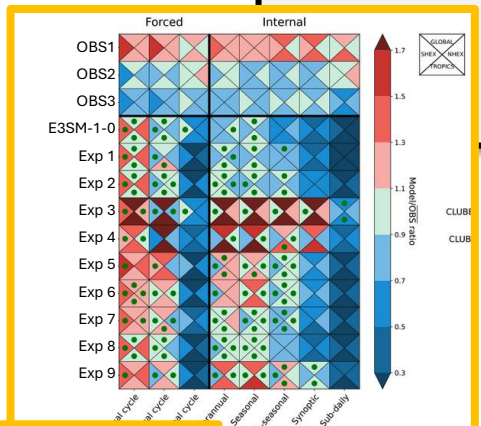
Min-Seop Ahn, P. J. Gleckler, J. Lee, A. G. Pendergrass, C. Jakob, A. Ordonez, P. Ullrich, M. Wehner, and many others



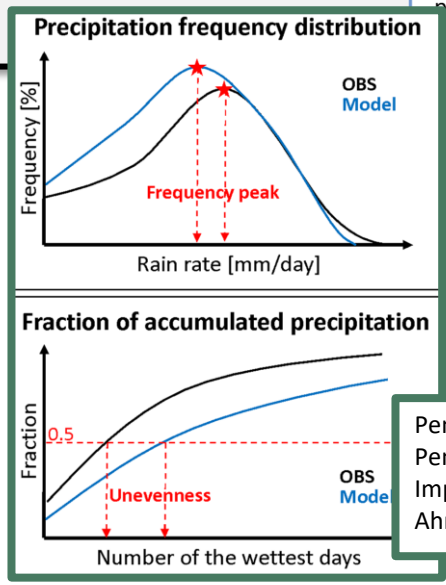
U.S. DEPARTMENT OF ENERGY July 2019 Workshop @DC

Spatial Distribution of Mean State	Seasonal Cycle	Variability Across Timescales	Intensity / Frequency Distributions	Extremes
RMS Error / MAE of mean state	Amplitude + phase of seasonal cycle	Standard deviation at different timescales	Simple Daily Intensity Index (SDII)	Rx1day
Pattern correlation			Unevenness	Rx5day
Monthly mean snow	Monthly score (RMS error)	Diurnal cycle – phase and amplitude	Mean and variance of daily precip.	20y return values (from GEV)
			Perkins score	Rx3h
			Fraction of precipitation days	Seasonal breakdown

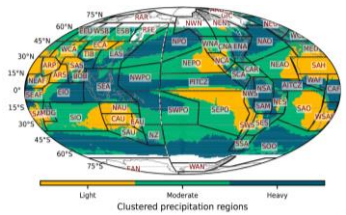




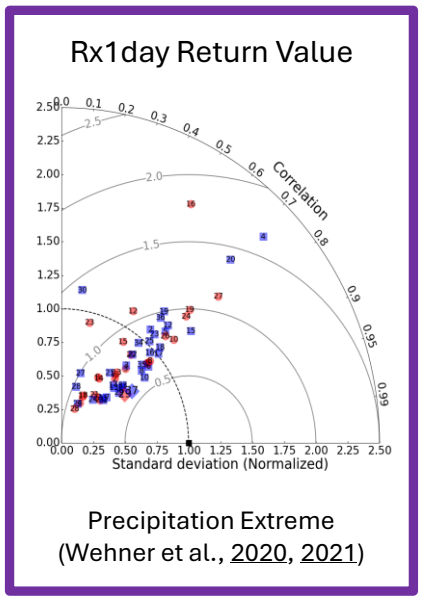
Ahn et al. (2021, J. Clim), also used for examine E3SM simulations.



Pendergrass and Deser (2017), Pendergrass and Knutti (2018) Implementation and application by Ahn et al. (2023, GMD)



Ahn et al. 2023 *GMD*

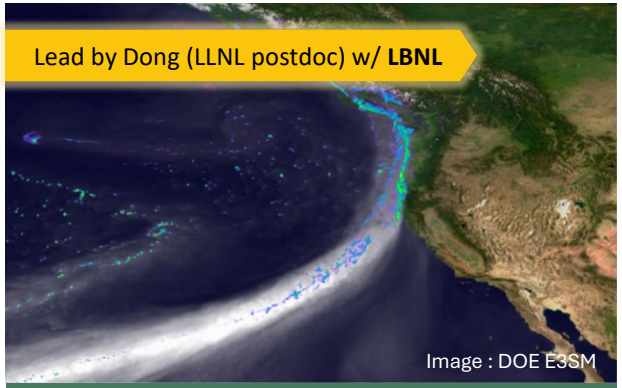


Precipitation Extreme (Wehner et al., 2020, 2021)

Collaborative additions (on-going and planned)

Large-scale Meteorology Driving Extremes

Lead by Dong (LLNL postdoc) w/ **LBNL**



Atmospheric Rivers

Narrow and stretched strong rain band

Landfall frequency,
Length and width, etc.

Collaboration w/ **CSU** (Valkonen , Barnes)

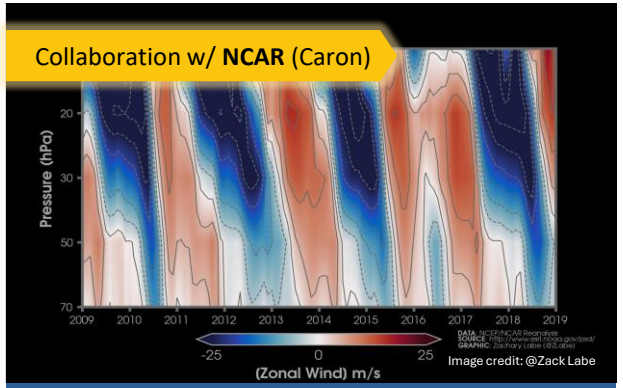
Atmospheric Blocking

Traffic jam in the atmosphere

Frequency, duration, etc.

Stratosphere-troposphere Connections

Collaboration w/ **NCAR** (Caron)



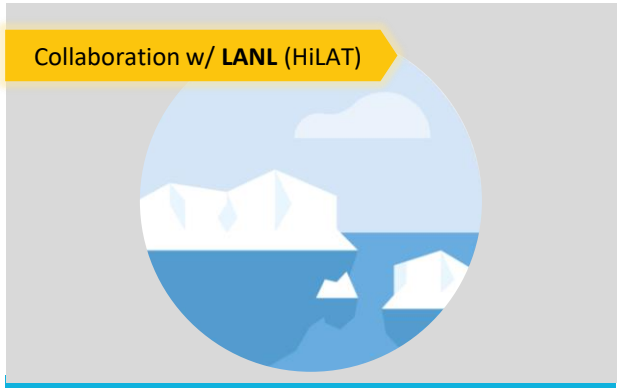
Quasi-Biennial Oscillation (QBO)

Stratosphere oscillation impacting weather

Amplitude, QBO-MJO teleconnection, etc.

Cryosphere

Collaboration w/ **LANL** (HiLAT)



Polar Metrics

Diverse metrics for Arctic/Antarctic regions

We are working on identifying potential candidate metrics

Leveraging TEMPEST EXTREMES

We are leveraging collaborations with the community to incorporate advanced performance measures

Reference datasets

The PMP leverages data products provided by obs4MIPs

- Obs4MIPs accelerates model evaluation, research and development, via:
  - Technical alignment of **observations** and selected **reanalysis** with CMIP
  - Detailed **provenance** including product origins, data preparation, and unified version control
  - Delivery on ESGF side-by-side with CMIP
- PMP uses dozens of obs4MIPs datasets including daily and 3hr products
- A suite of new obs4MIPs compliant products are now staged for **ESGF** publication
- As a WCRP project, obs4MIPs is expected to be a critical resource for **CMIP benchmarking**
- PCMDI, NASA and ESA are providing leadership

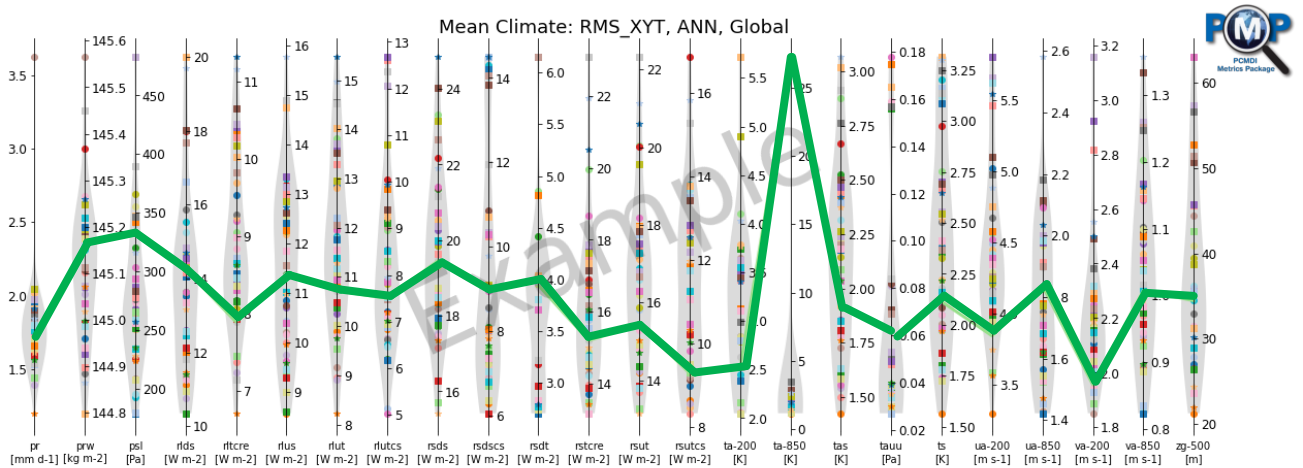
For further information contact Peter Gleckler (gleckler1@llnl.gov)

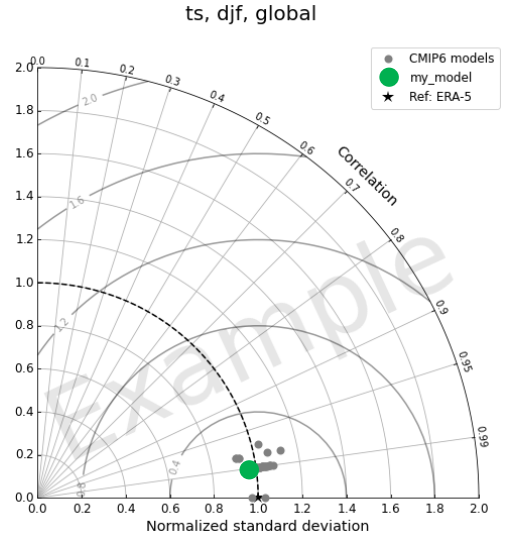
Database of pre-calculated metrics provides a foundation for benchmarking

- We release a **database of pre-calculated PMP metrics for CMIP5/6** https://github.com/PCMDI/pcmdi_metrics_results_archive
- which enables **objective benchmarking** (i.e., comparing the user's model with other CMIP models) *without* massive downloading of CMIP models to calculate the same metrics for them.
- **Provenance** information ensures robust **reproducibility** of the statistical numbers of metrics.

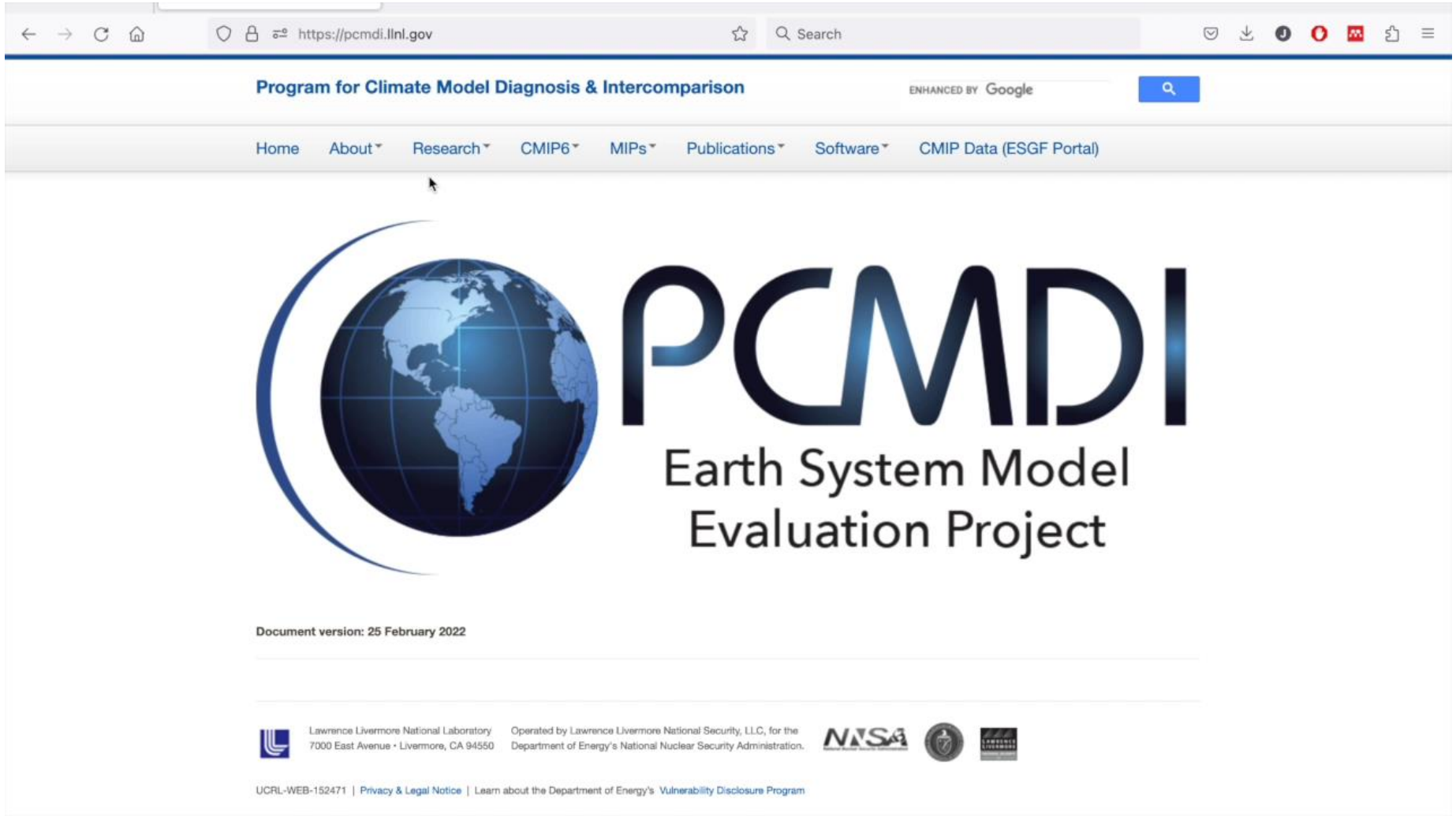

```
"provenance": {
    "platform": {
        "OS": "Linux",
        "Version": "3.10.0-1160.71.1.el7.x86_64",
        "Name": "gates.llnl.gov"
    },
    "userId": "lee1043",
    "osAccess": false,
    "commandLine": "/home/lee1043/.conda/envs/pcmdi_metrics_dev_
    "date": "2023-09-19 01:17:53",
    "conda": {
        "Version": "23.3.1",
        "buildVersion": "not installed",
        "PythonVersion": "3.10.12.final.0",
        "Platform": "linux-64"
    },
    "packages": {
        "python": "3.10.10",
        "cdat_info": "8.2.1",
        "cdms": "3.1.5",
        "cdp": "1.7.0",
        "cdtime": "3.1.4",
        "cdutil": "8.2.1",
        "esmf": "8.4.2",
        "esmpy": "8.4.2",
        "genutil": "8.2.1",
        "blas": "0.3.23",
        "lapack": "3.9.0",
        "mesalib": null,
```


Provenance information





PMP results interactive visualization



https://pcmdi.llnl.gov/metrics/

Thank You!

http://pcmdi.github.io/pcmdi_metrics/

Lee et al. 2024
Systematic and Objective Evaluation of Earth System Models: PCMDI Metrics Package (PMP) version 3. *Geoscientific Model Development*, 17, 3919–3948, doi: 10.5194/gmd-17-3919-2024

Lawrence Livermore National Laboratory

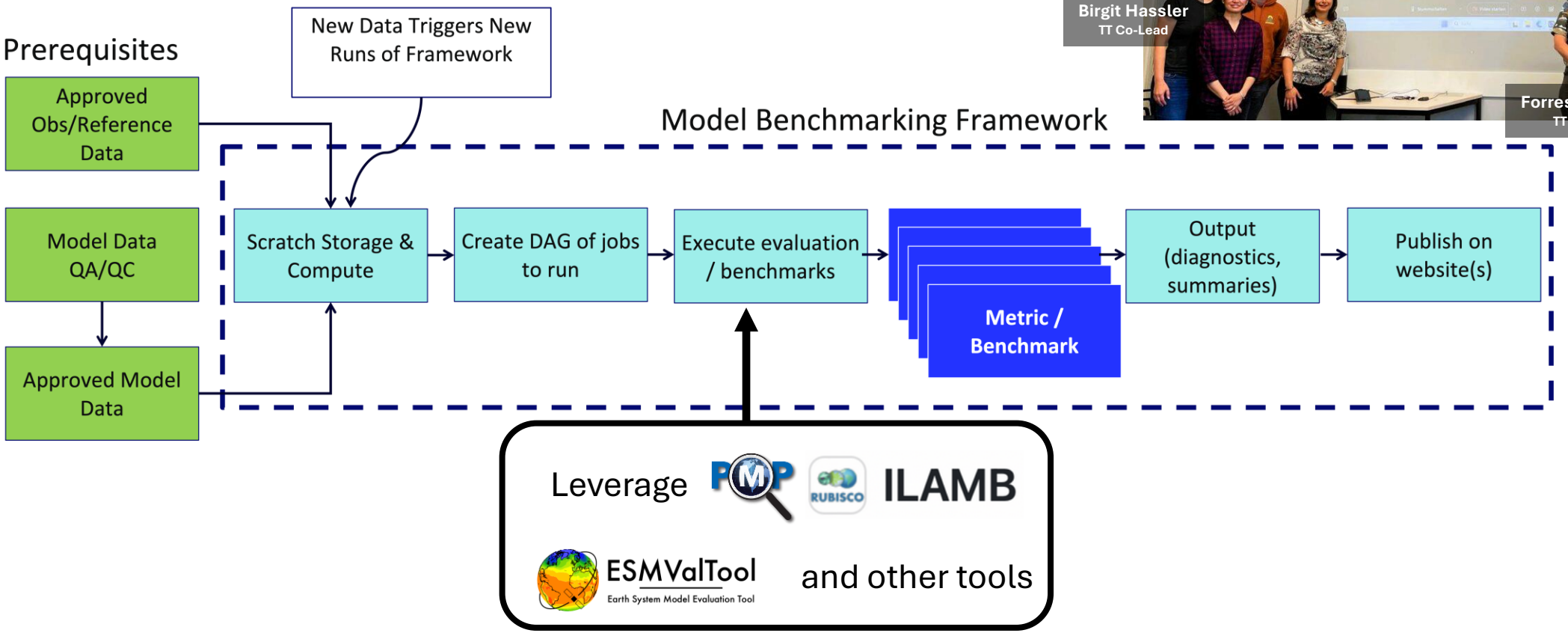
U.S. DEPARTMENT OF ENERGY | Office of Science

PCMDI
Program for Climate Model Diagnosis and Intercomparison

PMP contributes to CMIP Model Benchmarking

Engagement with the CMIP Benchmarking Task Team

Proposal for **CMIP7 Rapid Evaluation Framework**



CMIP Coupled Model Intercomparison Project
Model Benchmarking

Task Team Meeting @Germany, DLR (2024 May)

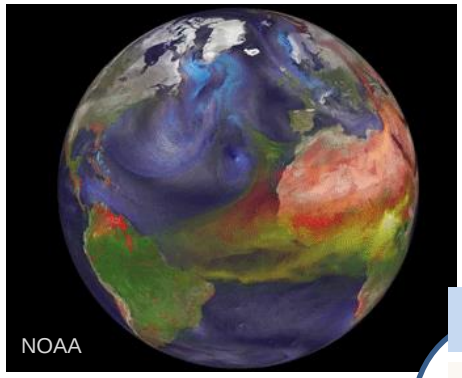
WCRP ESMO

Climate Models are Essential Tools for Understanding Climate Change

Evolutions of Earth System/Climate Models

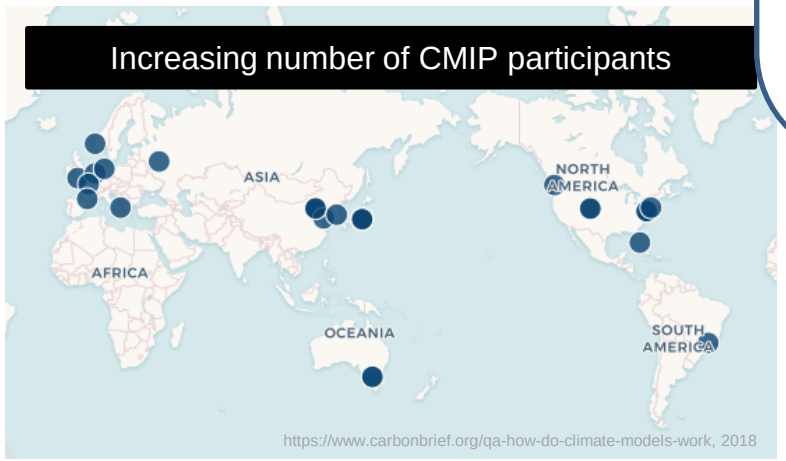
In 1960s

Current

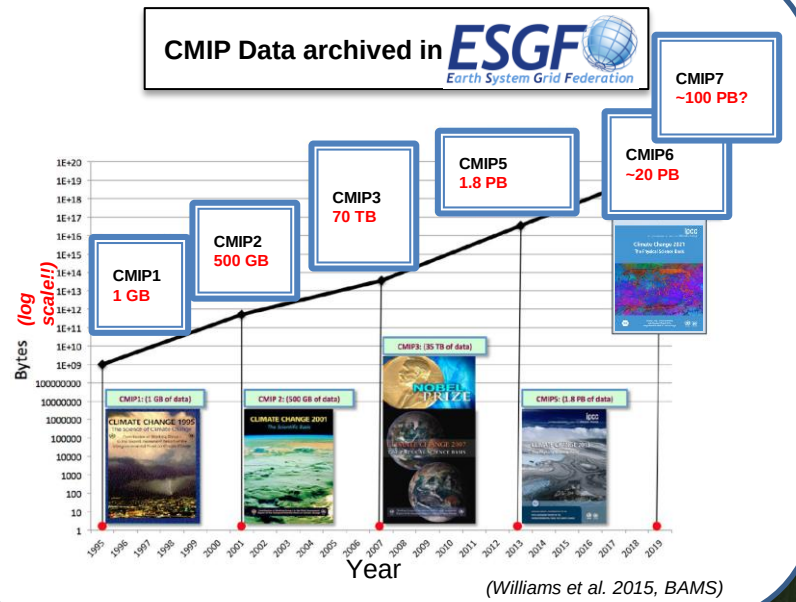


Diversified Earth System/Climate Models

Increasing number of CMIP participants



Outburst of Data



(Williams et al. 2015, BAMS)